Analysis of the Membrane Proteins in Human Serum

Pham Duc Dan1, Tran Thai Thuong1, Pham Dinh Minh1, Do Doan Loi2, Nguyen Bich Nhi1 and Phan Van Chi**

1Institute of Biotechnology (IBT), Vietnam Academy of Science and Technology (VAST), 18 Hoang Quoc Viet Rd, Hanoi, Vietnam
2Bach Mai Hospital, 78 Giai Phong Rd, Hanoi, Vietnam

Abstract

Serum and membrane proteins are two of the most attractive targets for proteomic analysis. Previous membrane protein studies tend to focus on tissue samples, while researches on membrane proteins in serum are still limited. In this study, human albumin and IgG depleted sera have been used for the proteome separation by SDS-PAGE. After staining, gels with protein fractions were cut and digested by trypsin. The peptide mixtures extracted from each gel slice were analyzed by nano-liquid chromatography-electrospray ionization mass spectrometry (NanoLC-ESI-MS/MS). The proteins were identified by Mascot v1.8 software and validated by MSQuant v1.5 software. As the result, a database including 1,216 membrane proteins, of which 469 (38.6%) membrane proteins with at least one transmembrane domain (TMD) were identified based on three different transmembrane prediction applications (Phobius, TMHMM, and SOSUI), was established. These proteins were further characterized and classified into different functional categories, such as cellular component, molecular function and biological process, based on their role according to Gene Ontology. Our results suggest that membrane proteomics will allow more pathway study related to diseases and will enable a better understanding of drug targets.

Keywords: Membrane proteins; Proteomics; Human serum; NanoLC-ESI-MS/MS

Introduction

Membrane proteins (MPs) play a central role in cellular and physiological processes. They are essential mediators of material and information between cells, intracellular compartments and organ systems. Their functions include: (i) active and passive transport of molecules into and out of cells and organelles, (ii) transduction of energy among various forms, (iii) reception and transduction of chemical and electrical signals across membranes [1]. Functionally intact MPs are vital to health and specific defects therein are associated with many known human diseases, i.e. heart disease [2,3], type 2 diabetes [4], cystic fibrosis [5]. Beside, MPs are the targets of a large number of pharmacologically and toxicologically active substances, and are directly involved in their uptake, metabolism, and clearance. Hence, study on membrane proteins can help in not only understanding about pathological mechanisms of disease, development of diagnostic methods but also drug discovery.

The human serum is considered one of the most comprehensive samples for proteomic research; many techniques and methods have been utilized and modified to analyze this promising proteome. It’s shown to contain thousands of proteins, including those originating from most, if not all kinds of cells and tissues [6]. Changes in the serum proteome contain vital information for the expression state of disease [7-10].

Proteomics has a great significance in MPs research. However, this method is impeded due to the possible nature of MPs such as high hydrophobicity, low abundance and complex post-translational modifications (PTMs). Nowadays, three promising strategies are used for the identification of MPs: gel-based approach using 1-DE or 2-DE with LC/MS [11], shotgun method coupling with 2D-LC/MS [12] and membrane-shaving method [13]. Each method has its own advantages and disadvantages, depending on the goal of the research. A comparison of the various methods disclosed the one-dimensional gel-LC/MS and the shaving approach to be highly complementary techniques [14]. However, the above-mentioned researches have been applied mostly for cell/tissues samples. Using carbonate extraction, trypsin digestion and NanoLC-MS/MS, a multilaboratory project has been found to profile membrane proteins from mouse liver [15,16]. In the studies on the analysis of human serum proteome, the identification of MPs has been mentioned, however, the given information was very limited [6,17]. In our recent research, by using shotgun method in combination with NanoLC-ESI-MS/MS technologies and bioinformatics tools, a data set of 217 membrane proteins from human serum was identified and characterized [18]. In this study, another approach, including the gel-based fractionation and nano-liquid chromatography-electrospray ionization mass spectrometry (NanoLC-ESI-MS/MS) have been used for human serum MPs identification and characterization. All the obtained results showed that the method yielded better recovery and reliability in the identification of the proteins especially the highly hydrophobic integral membrane proteins, and thus providing a promising tool for the analysis of membrane proteome. A database including 1,216 membrane proteins, of which 469 (38.6%) membrane proteins with one or more transmembrane domains (TMDs), was established. These proteins were further characterized and classified into different functional categories, such as cellular component, biological process, molecular function, based on their role according to Gene Ontology.

Experimental Procedures

Materials

Aurum serum protein mini kit, the Bradford assay kit, dithiothreitol (DTT), iodoacetamide (IAA), ammonium bicarbonate, ammonium

*Corresponding author: Phan Van Chi, Institute of Biotechnology (IBT), Vietnam Academy of Science and Technology (VAST), 18 Hoang Quoc Viet Rd, Hanoi, Vietnam, Tel: 84-91224426; E-mail: chi@ibt.ac.vn

Received November 11, 2013; Accepted December 11, 2013; Published December 16, 2013


Copyright: © 2013 Dan PD, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
acetyls, trypsin (proteomics sequencing grade), acrylamide, bis-acrylamide, 3-(3-cholamidopropyl)dimethylammonio)-1-propanesulfonate (CHAPS), urea, glycine, Tris, and sodium dodecyl sulfate (SDS) were all purchased from Bio-Rad (Bio-Rad Laboratories, Hercules, CA, USA). Formic acid (FA) and trifluoracetate (TFA) were obtained from Fluka (Fluka Chemie GmbH, Buchs, Switzerland). Acetonitrile (ACN, chromatogram grade) and other chemicals (analytical grade) were obtained from Barker (Pittsburgh, USA). All equipment and standard reagents used directly should be clean as necessary.

**Serum sample preparation**

Serum samples from 30 healthy middle-aged individuals were supplied by Bach Mai Hospital (78 Giai Phong Rd, Hanoi, Vietnam). The high abundant albumin and IgG were depleted from the samples by Aurum serum protein mini kit (Bio-Rad Laboratories, Hercules, CA, USA). For each sample, 60 µl of original serum was diluted with 180 µl of serum protein binding buffer, and the depletion procedure was performed at room temperature according to manufacturer’s instructions. The depleted serum, unbound fraction was collected by centrifugation at 10,000xg for 20 sec. The removal of albumin and IgG was evaluated by 12.6% SDS-PAGE.

**Electrophoresis and trypsin digestion**

The depleted serum samples were separated by 12.6% sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) with loading amount of 25 µg of protein for each lane and were visualized by staining with Coomassie Brilliant Blue G-250. Each lane contained different protein fractions was cut into approximately 10 gel pieces, each was put into 1.5 ml tube. The proteins were digested in gel with trypsin as described in our previous study [18]. Briefly, Coomassie-stained gel pieces were washed and destained by using wash solution (50 mM NH₄HCO₃, pH 8.0, 50% ACN). The protein fractions were reduced by incubating with 5 mM DTT solution at 56°C for 45 min and then alkylated for 60 min with 20 mM IAA solution in darkness at room temperature. For the in-gel digestion, trypsin was added and incubated overnight at 37°C. The released peptides were extracted with extraction solution (60% ACN in 1% TFA (v/v)). Finally, extract mixtures were dried and saved for further analysis.

**2DnanoLC-ESI-Q-TOF-MS/MS**

This method was performed as described elsewhere with minor modifications [19,20]. The peptide mixture was resuspended in 30 µl of 0.1% FA. The integrated column included two sections. The first section was a strong cation-exchange column (SCX, 500 µm ID×15 mm, 5 µm, 300 Å) at a flow rate of 30 µl/min. And the second section was a Vydac reversed-phase C18 column (RP, 75 µm×150 mm, 5 µm, 300 Å). The flow rate was maintained at 0.2 µl/min with solvent A containing 0.1% FA. The peptides were eluted from a reverse phase C18 column using the solvent B (85% ACN, 0.1% FA) gradient. Actually, the peptides were separated first by 2D nano-LC, and subsequently were independently analyzed by a QSTAR XL MS/MS mass spectrometer (MDS SCIEX/Applied Biosystems) equipped with a nanoESI source. MS spectra were recorded and processed in IDA (Information Dependent Acquisition) mode controlled by Analyst QS software.

**Protein identification and validation**

For protein identification, the obtained MS and MS/MS spectra were searched against the NCBInr and Swiss-Prot Human protein database using Mascot v1.8 software (Matrix Science Ltd, London, UK). The parameters were set as follows: enzymatic cleavage with trypsin, one potential missed cleavage, and a peptide and fragment mass tolerance of ± 0.25. Carbamidomethyl cysteine was set as a fixed modification; oxidized methionine was set as a variable modification. Protein identifications were performed using a Mowse scoring algorithm with a confidence level of 95% and at least two peptides matched, showing a score higher than 43. For further verification, proteins were validated by MSQuant v1.5 software (http://msquant.sourceforge.net) [21]. Membrane proteins were selected from total identified serum proteins based on UniProt protein database (http://www.uniprot.org) [22] and Gene Ontology annotations (http://www.geneontology.org/GO.ontology.structure.shtml) [23]. Transmembrane Hidden Markov Model (TMHMM) [24], SOSUI [25] and Phobius [26] prediction algorithms were used to predict transmembrane domains of the identified proteins.

**Results**

**Overall identification of MPs in human serum samples**

As mentioned previously, three-dimensional separation method combined SDS-PAGE prefractionation with trypsin digestion of gel slices and continuous two-dimensional LC-MS/MS analysis, was proved to be the best for significantly enhancing protein identifications including single-pass and multiple-pass transmembrane proteins [27,28]. In our study, the albumin and IgG depleted sera were first applied also for SDS-PAGE fractionation. As it was evaluated by SDS-PAGE, a large number of the proteins were enriched and fractionated (data not shown). The SDS-PAGE gel was then cut into approximately 10 bands for in-gel digestion. After desalting and cleaning up with Zip tips, the peptides were applied to 2DnanoLC-MS/MS analyses, as it was shown in the previous publication [19]. The obtained mass spectra were searched using Mascot V1.8 against Homo sapiens sequences in NCBInr. All Mascot results were parsed and selected. All identified proteins were selected based on the information that was sought by UniProt protein database (http://www.uniprot.org). They were also validated using MSQuant software. In total, from 23,243 matched peptides, 3,406 proteins in human serum were identified. Of which, a set of 1,216 MPs being corresponding to ~35.7% of the totally identified proteins were predicted and analyzed. It should be noted that among the identified MPs, 469 membrane proteins were characterized to have at least one or more TMDs based on Phobius, TMHMM, and SOSUI prediction algorithms.

**Distribution of molecular weight of MPs**

The pie chart shown below (Figure 1) provides information about the molecular weight of the identified MPs. This distribution shows the possible diversity and complexity nature of MPs. From the list of 1,216 identified MPs, one can see not only very low molecular weight proteins (<6000 Da), such as glutamate receptor 7 (4388 Da), cytochrome b (5294 Da), but also the very high molecular weight proteins (>300 kDa), i.e., dystonin (862 kDa), hemicentin 1 (623 kDa) and so on. The most outstanding feature of the chart is the high proportion of high molecular weight MPs as compared with low molecular weight ones. The proportion of the MPs with the molecular weight below 50 kDa is only 14.3%, while that of the most identified MPs were in the following ranges: 50 kDa - 100 kDa (31.2%); 100 kDa -150 kDa (27.9%); 150 kDa-200 kDa (11.1%); >200 kDa (15.5%).

**Prediction of transmembrane domains (TMDs)**

TMDs contain interesting features because they show the protein...
is integral or anchored membrane protein and also describe the transmembrane level of that protein. In our study, all 1,216 identified membrane proteins were applied for the prediction by using three different algorithms: SOSUI, Phobius and TMHMM. The Figure 2 shows the results of the prediction: SOSUI found 598 transmembrane proteins, Phobius found 581 proteins, while 517 proteins were also defined by TMHMM. Notably, 12 proteins were removed by SOSUI because their sequences contain more than 5000 amino acids, and they were out of SOSUI calculation ability. It means that almost half of the identified MPs were predicted to have at least one or more TMDs. Of which, 459 MPs were predicted by all three above-mentioned algorithms. This shows that the difference between three methods is so slim and that they provide high reliability and accuracy results. The second half of total 1,216 proteins was predicted to belong to protein groups that adhered or anchored to membrane.

The Figure 3 illustrates the interrelation between the numbers of proteins and the amount of TMDs in three different predicting applications. In spite of the meaningfully similar number of predicting transmembrane proteins, the used algorithms offered significantly different results about the numbers of TMDs, which are possessed by each protein. The below chart shows that most of the proteins contain one or two TMDs. However, predictions of the applications are unequal. For instance, Phobius calculated 41 proteins that possess 2 TMDs, while SOSUI found 196 proteins that contain the same number of TMDs. The less difference can be seen in the prediction of proteins with 4 or more TMDs. It should be noted that the proteins possessing more than 10 TMDs have a significant proportion. Furthermore, all of SOSUI, Phobius and TMHMM show similar predictions about the number of those proteins.

**Classification of the identified MPs via Gene Ontology**

The identified MPs were classified into different functional categories based on their role according to universal Gene Ontology (GO) annotation [23]. As shown in the Figure 4a, the distribution of the MPs was identified according to their cellular components. It should be noted from the pie chart that the highest portion of MPs (53.4%) was annotated with cell/plasma membrane terms. The intracellular and other organelles make up the lower portion (46.6%).
Of which, GO terms have been found related to intracellular MPs that can be divided into two subgroups: specified and unspecified. The first specified subgroup of intracellular MPs includes endoplasmic reticulum (7%, 91 proteins), and Golgi (7.5%, 98 proteins), nucleus (3.4%, 45 proteins), mitochondrion (3%, 39 proteins), lysosome (0.8%, 11 proteins) and peroxisome MPs (0.8%, 11 proteins). The portion others (24.1%, 317 proteins) means the second subgroup containing unspecified intracellular MPs, such as Dnj3/Cpr3, FAM162B protein, pleckstrin 2, and so on.

According to Gene Ontology annotations, molecular function of verified serum MPs were characterized and categorized as indicated in Figure 4b. The GO distribution function showed several special groups: (i) enzymes that make up the highest rate with 24.8% (361 proteins); (ii) the binding MPs, including metal ion binding (11.5%, 167 proteins), NTP binding (4.9%, 71 proteins), phospholipid binding (2.7%, 39 proteins), cytoskeleton binding (1.5%, 22 proteins) and other binding (7.8%, 113 proteins); (iii) MPs with specific functions were analyzed as following: receptor (10.6%, 155 proteins), structure (2.9%, 43 proteins), transporter (3.8%, 55 proteins), immune response (2.7%, 39 proteins), transcription (2.1%, 30 proteins), enzyme activator 3.4%, 50 proteins), and enzyme inhibitor (2.5%, 36 proteins); (iv) group of MPs with unknown function (13.3%, 194 proteins).

Figure 4c with the pie graph indicates the percentage of the GO terms in the biological process category. There were 17 classified subgroups, such as immune response, cell differentiation, proteolysis, cell adhesion, catabolism, regulation, lipid metabolism, cell proliferation, apoptosis and so on. Of which, regulation constitute 21% (519 proteins), biosynthesis come second with 14.4% (357 proteins) of total GO terms in the biological process category. Follow up are transport (10%, 247 proteins), metabolism (7.8%, 194 proteins), cell adhesion (7.4%, 183 proteins), immune response (6.7%, 165 proteins), cell differentiation (5.5%, 136 proteins), cell proliferation (5.3%, 130 proteins) and so on. The results show that the distribution of biological process category of identified MPs is so diverse, and they participate in various important processes in the cell.

Discussion

Membrane proteomics is a highly focused branch of proteomics, because the high percentage of proteins encoded by the mammalian genome is transmembrane proteins [29]. Research on MPs from tissues/cells usually requires multiple steps to be performed: solubilization, fractionation, enrichment, purification, and identification. The strong point of these methods is the isolation of membrane proteins into a separated part from the original sample, so they restrict the influence of other proteins, which were depleted. However, some important MPs might be lost in extracting process, especially, the MPs with low amount. Furthermore, there are many kinds of MPs with different physico-chemical properties. There is no ideal kit, which satisfies the extracting all of MPs in a sample. The approach of using total serum sample for MPs analysis, which was applied in this study, has surmounted significantly above drawback. Instead of extracting out of the sample, membrane proteins were separated/enriched by SDS-PAGE and identified directly by two-dimensional nano liquid chromatography connecting mass spectrometry system. Thus, the loss of MPs with low amount was considerably reduced. Actually, the combination of SDS-PAGE and 2D nano-LC-ESI-MS/MS can create a strong tool to analyze/ identify plasma membrane proteins. The advantages of SDS, which is widely used to solublize hydrophobic proteins, have been reported [15, 16, 19, 30, 31]. These methods help to specify directly some kinds of proteins, instead of extracting them from total protein sample. It saved not only expenses but also time and does not lose accuracy.

Surprisingly that previous membrane protein studies tend to focus mostly on tissue/cell samples, while the researches on MPs in serum are still limited. Furthermore, the presented data on MPs in serum still contain contradictions, both in their number and characterization. Anderson and his colleagues have analyzed the human plasma proteome based on different methodologies. The obtained results...
showed only 18% of 1175 proteins in human plasma in all four datasets that contained transmembrane segments [6]. At the same time, the investigation of Chan et al. [17] using LC-MS/MS resulted in the identification of 1444 unique proteins, including all functional classes, cellular localization, and abundance levels in human serum, of which 32.69% were membrane/membrane associated proteins. However, little information was found on the characterization of the detected MPs, which were considered as not commonly associated with serum. Recently, our research group has launched the membrane proteins in human serum proteome project by using the shotgun method with 2D nano-LC-ESI-MS/MS. Unfortunately, only 217 identified membrane proteins were identified and characterized [18]. Subsequently, in this study, the combination of SDS-PAGE with 2Dnano-LC-ESI-MS/MS was applied to analyze MPs in human serum and yielded in much more number of MPs. As it was shown, in total, from 23,243 matched peptides, 3,406 proteins in human serum were identified. Of which, a data set of 1,216 proteins being corresponding to ~35.7% of the totally identified ones, has been found and validated as MPs in human serum. MPs with molecular weights from 50-100 kDa occupy the highest proportion (31.2%) of total identified MPs. MPs with molecular weights in the range of 100-150 kDa come second with 27.9%. The proportion of the MPs with the molecular weight below 50 kDa is only 14.3%.

Gene Ontology distributions for subcellular location, function and biological process of the identified MPs in human serum showed quite remarkable points. Regarding subcellular location of MPs, 53.4% of MPs belongs to cell/plasma membrane, while 22.5% of MPs are related to intracellular and 24.1% belong to other group/unspecified membrane proteins. GO distribution function showed several special groups, while the binding MPs and enzymes make up the highest rate with 27.9% and 24.8% respectively. It should be noted that there is still a high percentage of MPs with unknown function (13.3%). For the biological processes, there were 17 subgroups of MPs classified based on GO annotation with the following order: regulation (21%), biosynthesis (14.4%), transport (10%), metabolism (7.8%), cell adhesion (7.4%), immune response (6.7%), cell differentiation (5.5%), cell proliferation (5.3%) and so on. Actually, the data on the subcellular location, function and biological process of MPs shows the real their diversity in the serum and that they participate in various important processes. It should be noted the fact that tissues are continuously perfused by serum, their histopathology, i.e. necrosis, apoptosis, and hemolysis, may cause cellular proteins and peptides to be released into the blood stream. Obviously, this potentially pathophysiological information is reflected in serum proteomic patterns. The obtained results may suggest that membrane proteomics will allow more pathway study related to diseases and will enable a better understanding of drug targets.

Conclusion

In conclusion, this study presented a dataset of membrane proteins identified by using a gel based approach (SDS-PAGE) combined with 2DnanoLC-ESI-MS/MS technologies and bioinformatics tools. In total, 1,216 MPs in human serum were identified, of which 469 (~38.6%) proteins were predicted to have at least one TMD based on three不得已. Refs. (Phobiuss, TMHMM, and SOSUI) (Supplementary Tables). According to Gene Ontology, these proteins were classified into different functional categories, such as cellular component, molecular function and biological process. The results of the study may provide a valuable resource for understanding the human plasma proteome.


