Proteomic analysis of lipid-protein complexes

Tomas Vaisar

1Department of Medicine, University of Washington, Seattle, Washington, USA, 98109

*Address correspondence to:
Tomas Vaisar
Division of Metabolism, Endocrinology, and Nutrition
University of Washington
815 Mercer Street
Seattle WA 98109
Phone: (206) 543-3470
Fax: (206) 685-3781
E-mail: tvaisar@u.washington.edu
Abstract

There is intense interest in comprehensive proteomic approaches for analyzing integral membrane proteins and lipoproteins. Key features of mass spectrometric analysis center on enriching biological material for proteins of interest, efficiently digesting them, extracting the resulting peptides, and using fractionation methods to comprehensively sample proteins or peptides by MS/MS. However, lipid-associated proteins are generally rich in hydrophobic domains and often low in abundance. These features, together with the associated lipid, make their mass spectrometric analysis technically challenging. In this article, we review analytical strategies for successful proteomic analysis of lipid-associated proteins.

Introduction

Now that the genomes of humans and many model organisms have been fully sequenced, scientists must determine the molecular and cellular functions of the hundreds of thousands of proteins encoded by those genomes. They would also like to explain how the proteins cooperate or otherwise interact in complex physiological systems and how inappropriate interactions trigger human disease. These important biological problems are the central focus of the rapidly emerging field of proteomics—the study of protein expression, structure, and function.

Mass spectrometry (MS) is a powerful tool for proteomics, because it can identify and quantify hundreds or even thousands of proteins in complex biological samples. To study peptides and proteins with MS, it is necessary to use an instrument that vaporizes these normally nonvolatile compounds. Therefore, a key element in the development of MS-based proteomics, recognized by the 2002 Nobel Prize in Chemistry, was the discovery of methods for introducing peptides and proteins into the gas phase.(1, 2)

The two most common ionization techniques are electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI). ESI is typically applied to analytes in the liquid-phase, and is often used in concert with separation techniques such as reverse-phase chromatography. It is therefore well-suited to the analysis of complex mixtures. MALDI, which uses laser pulses to vaporize analytes embedded in a semi-crystalline matrix, is generally applied to relatively simple mixtures of peptides or proteins.

Although protein identification based on tandem mass spectrometric (MS/MS) analysis of intact proteins is an ideal approach, this so call *top-down proteomic* method (3) is still being developed, and is not amenable to large-scale investigations. Most proteomic studies take a *bottom-up* approach (4) by first digesting proteins with a protease (typically trypsin). The resulting small peptides are much more amenable to separation (typically by liquid chromatography, LC) and MS/MS analysis.

Two general approaches have evolved for protein analysis: gel-based proteomics and gel-free proteomics. The first methodology is widely used by biological investigators and involves various forms of electrophoretic gel-based separations, primarily one-dimensional (e.g., SDS-PAGE) or two-dimensional electrophoresis (2DE). Separated protein spots (or digests of spots) are extracted from the gel and identified by mass spectrometry. The second approach, often termed shotgun proteomics, takes advantage of facile coupling of liquid-phase separation techniques, primarily LC, with ESI-MS/MS.(5) Methods for analyzing many types of biological samples, including soluble cellular proteins and body fluids (plasma, urine, cerebrospinal fluid), are well established for both gel-based and gel-free approaches.
Proteins associated with lipids are more difficult to analyze with proteomic methods. These proteins play many essential roles in biological systems. Integral membrane proteins mediate interactions of a cell with its environment as well as interactions among various intracellular compartments, and a wide range of other cellular processes. In order for proteins to embed themselves into membranes and fulfill these functions, they assemble an outer surface that is rich in hydrophobic amino acids. Their rigid transmembrane domains often contain α-helices or β-barrels.

Apolipoproteins, which complex with lipids to form lipoproteins, are another class of lipid-associated proteins. They play essential roles in the biology and metabolism of low-density lipoprotein (LDL) and high-density lipoprotein (HDL), which are of central importance in cholesterol metabolism in vivo. Like integral membrane proteins, apolipoproteins have the ability to tightly associate with lipid and contain α-helices and β-barrels in their secondary structures.(6-10)

The hydrophobic domains of integral membrane proteins and apolipoproteins, together with the associated lipid, create unique problems for proteomic analysis. Moreover, these proteins are found in complex mixtures that are often dominated by a few very abundant proteins. While associated with lipids, proteins are shielded from efficient proteolytic digestion, and the peptides that are generated often have limited solubility and a strong tendency to aggregate due to their highly hydrophobic nature. This review will summarize current approaches to proteomic analysis of lipid-associated proteins.

**Proteomics of integral membrane proteins**

In biological membranes, transmembrane proteins are embedded in a lipid bilayer. The outer membrane of a cell, the plasma membrane, typically consists of 50% lipid and 50% protein by mass (w/w; molar ratio approximately 50:1), but the composition varies widely, depending on the membrane’s function. Thus, metabolically active mitochondrial membranes are highly enriched in protein (protein: lipid ratio 3:1 w/w), whereas neuronal membranes generally contain less than 20% protein.(11)

Integral membrane proteins typically have three major domains: i) an extracellular domain or ectodomain; ii) a hydrophobic domain embedded in the membrane; and iii) a luminal or intracellular domain. Based on their interaction(s) with the membrane, integral membrane proteins are generally classified as single pass (bitopic) or multipass (polytopic). Bitopic proteins have only one transmembrane domain, and can be oriented with the N-terminus on the outside of the membrane and the C-terminus on the luminal side (type I transmembrane proteins) or the other way around (type II transmembrane proteins). Polytopic proteins have several transmembrane domains, and their termini can be on either side of the membrane.

Transmembrane domains are highly structured sequences of amino acids that can assume two major secondary structures: β-barrels or α-helices. β-barrels are found in only a small fraction of integral membrane proteins, and do not usually present major challenge for proteomic analysis.(12) In contrast, α-helices, which characterize the majority of integral membrane proteins, are problematic because they are often rigid and highly hydrophobic. Furthermore, they are typically present in low abundance. Thus, it is generally critical to use various enrichment strategies to prepare them for proteomic analysis.(13, 14)

**Enrichment strategies.** These strategies make use of either physical-chemical (ultracentrifugation, phase separation) or biochemical (affinity isolation) principles. The first step typically involves subcellular fractionation to enrich for a specific organelle or whole membrane fraction. This is often accomplished by gradient density ultracentrifugation with sucrose or sorbitol gradients.(15, 16)
Plasma membranes of intact cells can be exposed to cationic colloidal silica, which interacts strongly with phospholipid headgroups. The silica selectively coats the extracellular face of the plasma membrane without disrupting the cell. After the cell is lysed, the plasma membrane is harvested by ultracentrifugation. Membranes can also be enriched with a two-phase mixture of polyethyleneglycol (PEG) and dextran. They partition into the hydrophobic PEG phase.

Affinity-based methods can also enrich fractions in plasma membrane. One method exploits one of the strongest noncovalent interactions in nature: the formation of the biotin-avidin complex \((K_d = 10^{-15} \text{ M})\). Cell-surface proteins are selectively labeled with a biotinylation reagent that cannot penetrate cells. When the cells are lysed and passed over avidin- or streptavidin-affinity columns, the labeled cell-surface molecules are specifically pulled down. Caution is needed, however, because samples can easily be contaminated with secreted proteins or cellular debris from the medium. Furthermore, if unreacted reagent persists during lysis, it will label cellular proteins, which therefore will be isolated along with plasma membrane proteins.

Because many integral membrane proteins are decorated with sugars, affinity isolation of glycosylated proteins is another means of enrichment. One approach uses lectin affinity chromatography, while others take advantage of the specific reactivity of oxidized glycosyl groups with hydrazine-activated solid supports. As with biotinylation, rigorous washing is essential, because a number of secreted proteins are also glycosylated.

Successful washing strategies typically include high-pH and high-salt washes to release proteins that associate nonspecifically with integral membrane proteins. At high pH, membrane vesicles unfold and release these contaminants while retaining proteins, such as GPI-anchored proteins, that specifically bind to membrane.

Membrane lipid and hydrophobic transmembrane domains present major challenges for proteomic analysis. The phospholipid bilayer hampers protein extraction, while insolubility in the aqueous phase limits proteolysis. The structurally rigid, hydrophobic transmembrane domains often aggregate and precipitate after they are liberated from lipid. To alleviate these problems, a number of methods have been developed to remove lipid and solubilize transmembrane proteins.

**Delipidation.** The most common delipidation method is chloroform/methanol extraction. Lipid dissolves efficiently in the chloroform phase, while protein precipitates at the chloroform/methanol interface. A simple alternative is to use acetone, which both precipitates proteins and dissolves most lipids. It should be noted that precipitation with trichloroacetic acid (TCA) does not remove lipids and should be used only in concert with acetone or chloroform/methanol delipidation. It is also important to note that some proteins are soluble in organic solvents, which may not quantitatively precipitate low-abundance proteins.

**Solubilization.** Membrane-embedded proteins are most efficiently solubilized with strong ionic detergent, SDS or TritonX100, chaotropic agents, urea, thiourea or guanidine, or a combination. However, the high concentrations of detergent (0.5%–4%) or chaotrope (5–8 M) that is typically required inhibit trypsin. Chaotropic agents also suppress ionization, and have to be removed prior to MS analysis. Many commonly used detergents (e.g. TritonX100 and other polyethers) also generate multiple high abundance ions that interfere with MS analysis.

Ionic detergents are incompatible with isoelectric focusing. The commonly used zwitterionic detergent CHAPS, even at high concentrations and in combination with urea and thiourea, does not efficiently solubilize many integral membrane proteins.
2DE analysis, sulphobetaine zwitterionic detergents have been used (e.g. ASB-14).(31, 32) Alone or in combination with CHAPS, they were superior to CHAPS alone for solubilizing membrane proteins.(33-35) In contrast, ionic detergents are fully compatible with SDS-PAGE. Many investigators first fractionate proteins with SDS-PAGE, cut the gel into segments, digest the proteins in the segments, and analyze the resulting peptides with LC-ESI-MS/MS.(20, 36)

Most detergents and chaotropic agents at the concentrations needed for efficient solubilization are also incompatible with gel-free proteomic approaches, because they inhibit trypsin. Moreover, they interfere with LC-ESI-MS by suppressing ionization and compromising chromatographic separation. While dilution can minimize the adverse effects on proteolysis and chaotropes can be removed by solid-phase extraction of the digested peptides, detergents such as SDS, TritonX100, and NP-40 are not easily removed. Therefore, alternative detergents have recently been developed. These acid-labile surfactants can be readily degraded by acidification prior to LC-ESI-MS analysis. Commercially available RapiGest (Waters),(37) an SDS analog, and PPS (Protein Discovery),(38) a zwitterionic detergent, are acid-labile and have been reported to increase sequence coverage for integral membrane proteins (37, 39) without compromising enzymatic activity.

Organic solvents can also solubilize membranes and membrane proteins. Trypsin is surprisingly tolerant of such solvents, which may even enhance its activity.(40-42) Organic-aqueous mixtures (typically acetonitrile or methanol) efficiently dissolve phospholipids, and the released proteins can be digested by trypsin. Concentrations of organic solvent up to 90% have been successfully used for trypsinization. (41, 42) Organic solvent can be readily removed prior to LC-ESI-MS analysis by evaporation or lyophilization. Several studies demonstrated significantly increased coverage of membrane proteins when high concentrations of organic solvents were used for delipidation and protein digestion.(40, 43, 44) Organic solvents have also been combined with acid-labile detergents to obtain better results than with traditional chaotropes.(39)

**Proteolytic digestion.** Proteins can also be predigested with CNBr or chymotrypsin before they are exposed to trypsin. CNBr, which cleaves methionine residues, can be used in concentrated acid solutions under conditions that partly solubilize membranes.(45-47) The resulting peptides are typically too large for direct LC-ESI-MS analysis, but they can be further digested with trypsin if the acid is first removed by lyophilization. The recent development of new ion dissociation, electron capture dissociation (ECD)(48), and electron transfer dissociation (ETD) (49, 50) methods that efficiently fragment large peptides offers new possibilities for MS analysis of membrane proteins that have been digested with CNBr. In contrast to CNBr, chymotrypsin recognizes hydrophobic and bulky amino acids. It can therefore cleave hydrophobic transmembrane domains, which are often underrepresented in tryptic digests.(40)

Finally, the basic conditions used to wash membranes cause their disruption and formation of open sheets.(51) This property can be exploited if strongly basic conditions are combined with digestion with proteinase K. This nonspecific protease rapidly digests proteins to dipeptides under optimal conditions. However, at pH>11, its activity is limited, and it yields peptides of 10–20 residues that are suitable for proteomic analysis.(52) Furthermore, it generates an overlapping series of peptides, increasing the probability of identifications from MS/MS spectra.

A wide variety of mass spectrometric approaches have been used to analyze integral membrane protein digests, including direct LC-ESI-MS analysis, MALDI-TOF-MS, and multidimensional protein identification technology (MUDPIT). A detailed discussion of these methods can be found elsewhere.(13, 14)
Proteomics of plasma lipoproteins

Lipoproteins are lipid-protein complexes that circulate in blood. Although their primary function is thought to be transport of cholesterol and other lipids, many lines of evidence strongly link them to the immune system and macrophage biology. (53-56) Traditionally, lipoproteins are characterized by buoyancy in density gradient ultracentrifugation. (57) They are classified into four major types: very low density lipoprotein (VLDL), intermediate density lipoprotein (IDL), LDL, and HDL. (57) The protein content by weight of lipoproteins ranges from ~10% (VLDL) to ~50% (HDL). Although lipoproteins are commonly quantified by their cholesterol content, they contain a rich mixture of other lipids, including sphingolipids, cholesteryl esters, triglycerides, and phospholipids.

Lipoprotein particles are spherical. Their nonpolar core of cholesteryl esters and triglycerides is surrounded by an amphipathic layer of phospholipids and free cholesterol. They contain two major classes of apolipoproteins (apo): apoB100 (a large protein found on VLDL, IDL, and LDL) and exchangeable apolipoproteins (found on all classes of lipoprotein particles).

The major protein component of HDL is apoA-I, a 30 kDa exchangeable apolipoprotein. When apoA-I is associated with lipid, almost 70% of its primary sequence is coiled into amphipathic \( \alpha \)-helices. Most current models of HDL place apoA-I molecules in an anti-parallel double-belt wrapping around the lipid core. The phospholipid's hydrophobic fatty acids and polar head groups interact with the protein's amphipathic helices. The several distinct models of apoA-I (9, 10) agree that a single HDL particle contains two to four molecules of apoA-I, depending on its size.

Recent proteomic analyses have revealed that HDL contains 48 or more proteins. (56) Many of these are not apolipoproteins, and they are generally present at relatively low abundance. While some of the non-apolipoprotein proteins contain hydrophobic domains, it is unclear how most interact with lipids or each other. While some also have a high \( \alpha \)-helical content and may interact with the surface lipids, others may form specific protein–protein complexes.

In contrast to HDL, LDL particles are larger and much richer in lipid (80% by weight). Their major protein is apoB100, one of the biggest known single-chain proteins (550 kDa). In each LDL particle, one molecule of apoB100 wraps around the lipid core and stabilizes the complex. (58) Although the precise structure is unclear, it is apparent that most of apoB-100 consists of \( \beta \)-sheets and amphipathic helices, some of which protrude into the lipid core. (58) Due to its size and hydrophobicity, lipid-free apoB-100 is insoluble in aqueous solution. To date, proteomic studies have revealed relatively few other proteins in LDL particles. (59)

Isolation of lipoproteins. For proteomic analysis, lipoproteins have to be isolated from plasma. Because of their lipid content, they are more buoyant than other plasma proteins, and this property has been used to separate them from both plasma and other classes of lipoproteins. Traditional methods, established by Havel et al. (57) use density gradient ultracentrifugation. Alternatively, lipoproteins can be isolated by immunosorption onto antibodies that are specific for dominant proteins of each class. Thus, anti-apoA-I antibodies can be used to isolate HDL particles (60), while anti-apoB100 antibodies yield predominantly LDL particles. (61) However, particles isolated by affinity approaches are not identical to those isolated by density ultracentrifugation. (62)

While ultracentrifugation modifies lipoprotein particles due to sheer stress and the medium's high ionic strength, affinity isolation tends to co-purify nonspecifically associated proteins. Thus, it is critical to include proper controls in proteomic analyses of affinity isolated particles.

Delipidation. Delipidation by the widely used Folch extraction, (63) which involves chloroform-methanol and aqueous phases, has been successful in several studies. Karlsson et
al (59) found that extracting lipids from LDL did not increase the number of spots on 2DE gels; however, it did improve resolution. Due to apolipoproteins' high hydrophobicity, however, they could be lost during lipid extraction. Given the limited dynamic range of protein stains in gel-based approaches and the high abundance of single proteins (apoA-I in HDL; apoB100 in LDL, IDL and VLDL), losses of less abundant proteins would be hard to detect in gel-based analyses. The additional sample preparation step also introduces variability, which makes protein quantification more difficult.

**Solubilization and Digestion.** Solubilization and digestion present the same difficulties for lipoproteins as for integral membrane proteins. While complexed with lipid, proteins are readily soluble. Once delipidated, many precipitate and aggregate unless high concentrations of chaotropes or strong detergents are used.

The approach to solubilization largely depends on the lipoprotein class and analytical approach. For one-dimensional SDS-PAGE separation of HDL proteins, Laemmli buffer was apparently sufficient to denature and dissociate the lipoprotein particles. (64) In several studies, isoelectric focusing on immobilized pH gradients followed by second dimension SDS-PAGE (64) or by direct protein analysis (65-67). Various combinations of urea (9 M urea or 7M urea with 2 M thiourea) and nonionic (2.5% Triton-X100) or zwitterionic (2% CHAPS) detergents have also been used as solvents. With SDS-PAGE and 2DE, proteins are generally digested in-gel and identified by MALDI-TOF-MS.

To elute intact apolipoproteins from immobilized pH gradient gel strips, one approach used a combination of formic acid, acetonitrile, 2-propanol, and water. (65) About 25% of intact HDL protein mass was recovered, whereas conventional methods proved unsuccessful (65). This approach identified seven proteins in HDL, including multiple forms of apoA-I and apoC-III. HDL has similarly been subjected to direct MALDI-TOF-MS analysis without eluting proteins from an immobilized pH gradient strip. The protein cargo of HDL solubilized with 9 M urea/2% CHAPS has also been investigated by MALDI-TOF-MS. (64) Collectively, these studies demonstrate that gel-based approaches can detect isoforms of apolipoproteins but have difficulty identifying low-abundance proteins in HDL.

LDL poses a significant problem to gel-based proteomics due to its high molecular weight and the hydrophobicity of apoB100. In contrast to HDL, it requires stronger solubilizing agents to prevent protein precipitation. To separate LDL proteins on SDS-PAGE, Karlsson et al (59) used a combination of 4% SDS and 20% sucrose. Two-dimensional electrophoretic separation of apoB100 is also problematic. During isoelectric focusing, poorly soluble apoB-100 tends to precipitate near its isoelectric point, which can also compromise separation of other proteins. Furthermore the size of apoB100 prevents it from entering the gel efficiently and separating in the second dimension.

To avoid these issues, several groups have used gel-free analysis of lipoproteins without prior delipidation and separation. Surface enhance laser desorption ionization (SELDI) was used to analyze LDL (68) or HDL (67, 69) that were directly adsorbed onto the SELDI plate. Although this approach successfully identified apolipoproteins, protein identification in SELDI is based on low mass resolution data, which has many limitations. Thus, its applicability to proteomics is limited.

VLDL contains an even smaller proportion of protein than LDL. When VLDL (70) was delipidated and precipitated proteins resolubilized in 4% CHAPS, 2DE analysis identified several exchangeable apolipoproteins including apoE, apoA-I, apoA-IV, apoM, apoC-I, and apoC-III. Interestingly this study also identified fibrinogen-γ and actin.
To circumvent the complications associated with gel-based approaches, direct analysis of HDL by shotgun proteomics has been attempted. For example, HDL isolated by ultracentrifugation was added directly into ammonium bicarbonate without any delipidation, solubilization or denaturation agents. (71, 72) Using this approach, 14 proteins were identified in HDL.(71)

Shotgun proteomics has been applied to HDL proteins precipitated from HDL with 10% TCA.(56) After precipitation, proteins were resolubilized in 6 M urea and digested with trypsin. The peptides were then separated by reverse-phase or 2DE liquid chromatography and analyzed with ESI-MS/MS. This study identified 48 proteins in HDL. A similar approach used the acid-labile detergent RapiGest.(74)

In contrast to HDL, no published accounts have analyzed LDL by shotgun proteomics.

Conclusions

Lipid-associated proteins present a unique set of challenges to proteomic analysis. Key features of any mass spectrometric analysis center on enriching biological material for proteins of interest, efficiently digesting them, extracting the resulting peptides, and using fractionation methods to comprehensively sample proteins or peptides by MS/MS. Integral membrane proteins have been extensively investigated by a number of well-validated protocols. In contrast, analysis of lipoprotein particles has received limited attention, and optimal approaches for comprehensive proteomic analysis are still being developed. Although the lessons learned while analyzing integral membrane proteins will undoubtedly benefit proteomic investigations of lipoproteins, unique features, especially the dominance of a few proteins in each particle type, will require innovative methods for detecting low abundance proteins in these protein-lipid complexes.
References


