universität wien

MASTERARBEIT

Titel der Masterarbeit

„Experimental Evaluation and Optimization of Membrane Proteome Profiling Techniques Employing HPLC and Orbitrap Mass Spectrometry“

verfasst von

Selin HIZAL, BSc.

angestrebter akademischer Grad

Master of Science (MSc)

Wien, 2013

Studienkennzahl lt. Studienblatt:	A 066 862
Studienrichtung lt. Studienblatt:	Masterstudium Chemie
Betreut von:	Univ. Prof. Dr. Christopher Gerner

Abstract

Membrane proteins perform several important functions in the cell, including the transport of ions and molecules, cell signalling, immune response and regulation of enzymes and protein synthesis. Functionally, they can be receptors, ion channels, active transporters or enzymes located in the membrane. Malfunctioning of membrane proteins (MP) has been linked to important human diseases such as cancer, Alzheimer's disease, diabetes, and cardiac disorders. As such, proteomic analyses with a focus on membrane proteins can provide important insight into molecular mechanisms operating in the cell and also allow the discovery of new drug targets.

Proteome profiling using shotgun proteomics involves the steps of isolation of fractions from cultured cells, electrophoretic separation of the proteins in the sample fraction, digestion with a protease such as trypsin, chromatographic separation of the protein digests, and mass spectrometric analysis in the form of nano-flow LC-MS/MS. The "fingerprint" of each peptide's fragmentation mass spectrum is used to identify the protein from which they derive by searching against a sequence database with commercially available software (e.g. MASCOT). In spite of their enormous significance, the analysis of membrane proteins by standard LC-MS experiments is notoriously difficult. With their low abundance added to their hydrophobicity, the coverage of hydrophobic peptides is usually relatively low.

In this master's thesis, various strategies were applied and compared with respect to their ability to capture membrane proteins and optimized. Cytoplasmic and nuclear fractions were isolated from Jurkat cells, which were separated, digested and analysed by an established standard protocol. On the other hand, membrane-enriched fractions were isolated by disrupting the cells with freeze-thaw cycles using liquid N_2 and afterwards ultracentrifugation. Membrane and cytoplasmic samples were also prepared using a Protein Extraction Kit von Thermo Scientific. In addition to trypsin, chymotrypsin was used for the digestion of various samples. A digestion protocol where the digestion buffer contained 20 % ACN was also tested. A four hour gradient was used for the separation of the peptides and a top12 method for the mass spectrometric analysis on a QExactive Orbitrap instrument by Thermo Scientific.

With the membrane-enriched samples, a large number of membrane proteins were identified. Especially the membrane fraction digested with trypsin contained many intrinsic plasma membrane and organelle membrane proteins, that perform important functions in the cell as receptors, transporters or ion channels, and participate in vital biological processes such as cell communication, signal transduction and immune reaction.

Zusammenfassung

Membranproteine vollziehen viele wichtige Funktionen in der Zelle, wie den Transport von Ionen und Molekülen, Zellsignalprozesse, Immunantwort und Regulation der Proteinsynthese. Etwa ein Drittel der offenen Leseraster kodieren Membranproteine und deren Funktionsstörung können mit folgenschweren humanen Krankheiten wie Krebs, Alzheimer, Diabetes und Herzerkrankungen verbunden sein. Daher können proteomische Analysen mit Schwerpunkt auf Membranproteinen wichtige Erkenntnisse über molekulare Mechanismen, die in der Zelle stattfinden, liefern.

„Proteome Profiling“ ist eine gut etablierte Technik, die im Wesentlichen auf der Isolation von Proteinen, der nachfolgenden Trennung proteolytischer Peptide und der Identifikation mittels Peptidfragmentierung in Massenspektrometer beruht. Die Identifikation von Membranproteinen stellt allerdings eine besondere Herausforderung dar, weil sie nicht nur in äußerst geringen Mengen vorliegen, sondern zusätzlich durch ihre Hydrophobizität und die geringe Anzahl tryptischer Schnittstellen nur unzureichend mit Standardmethoden zu erfassen sind. Im Rahmen dieser Diplomarbeit wurden unterschiedliche Strategien bezüglich Ihrer Eignung Membranproteine zu identifizieren getestet, verglichen und optimiert. Zuerst wurden zytoplasmatische und Kernfraktionen von Jurkatzellen isoliert, die mit einem Standardprotokol verdaut und analysiert wurden. Diese wurden dann mit Membran-angereicherten Fraktionen (Membranen, Organellen und Kerne) verglichen, die mittels N_2-Aufschluß und Ultrazentrifugation isoliert wurden. Ferner wurden mit einem Aufschlußkit (Protein Extraction Kit von Thermo Scientific) Membran- und zytoplasmatische Fraktionen isoliert.

Die Proben wurden mit verschiedenen Methoden gereinigt, verdaut und analysiert. Neben Trypsin wurde für den Verdau der Membran-angereicherten Fraktionen auch Chymotrypsin eingesetzt, da dieses Proteine an hydrophoben Aminosäuren (Phe, Tyr, Trp) schneidet, die in Intramembranregionen häufig vorkommen. Alternativ enthielt der Hydrogencarbonatpuffer, in dem Trypsin verdünnt wurde, 20 % Acetonitril, um hydrophobe Domänen für Trypsin besser zugänglich zu machen. Die Trennung erfolgte über einen 4 Stunden-Gradienten auf einer C18 Säule und die Analyse mit einem QExactive Orbitrap Massenspektometer unter Verwendung einer Top12 Methode. Mit den N_2-aufgeschlossenen Fraktionen (Membranen und Organellen) war es möglich eine große Anzahl an Membranproteinen zu identifizieren. Besonders die Membranprobe enthielt viele genuine Membranproteine, die in der Zelle an lebenswichtigen biologischen Prozessen wie zelluläre Kommunikation, Signaltransduktion und Immunrreaktion teilnehmen.

Acknowledgements

I am grateful to my supervisor Prof. Christopher Gerner for giving me the opportunity to work on this intriguing project, for his ever positive attitude, and for his support and encouragement throughout my master's thesis.

I want to thank Besnik Muqaku for his enlightening tutelage and assistance in the lab, Andrea Bileck for her invaluable support with cell culture and Amar Tahir for his lively motivation and friendliness.

Further, I wish to express my appreciation to Dr. Michael Größl and Dr. Astrid Slany for their solid practical knowledge and guidance.

Marietta Maurer, Rupert Mayer and all other members of the group should also be acknowledged for the congenial atmosphere.

Last but not least, I would like to thank my parents, without whose support my studies would not have been possible.

Contents

1. Introduction

1.1. Membrane Proteins

Membrane proteins perform several important functions in the cell, including the transport of ions and molecules, cell signalling, immune response and regulation of enzymes and protein synthesis. Functionally, they can be receptors, ion channels, active transporters or enzymes located in the membrane. About one third of all open reading frames code for membrane proteins. As such, proteomic analyses with a focus on membrane proteins can provide important insight into molecular mechanisms operating in the cell. Malfunctioning of membrane proteins (MP) have been linked to important human diseases such as cancer, Alzheimer's disease, diabetes, and cardiac disorders. Of clinical importance is the possibilty to compare healthy and versus disease samples, or disease versus treated samples. As a result of their vital role in the proper functioning of the cell and of the organism, about 70 % of all pharmaceuticals target membrane proteins, with 25% of these belonging to G-protein coupled receptors (GPCRs)[1]. Proteomic research of membrane proteins may allow the discovery of new drug targets as well as the prediction of response to and side-effects of drugs. On the other hand, because they are processed by membrane proteins, a better understanding of lipids also requires a thorough study of the expression, topology and function of various MPs.

The amount and classes of MP depend on the specific type and function of the cell they belong to. Around 75 % of mitochondrial membranes and 50 % of plasma membranes is protein by mass. With respect to their form of attachment to the lipid bilayer, MP can be either integral MP or peripheral MP[2]. Most integral MPs are transmembrane proteins, with hydrophobic regions passing through the membrane at least once; others, that are known as integral monotopic proteins, are attached to only one side of the membrane by a hydrophobic domain or a lipid chain. Cyclooxygenases (COX-1 and COX-2), which are important enzymes converting arachidonic acid to prostanoids, belong to this second class of integral membrane proteins.

Cell surface receptors, transporters and ion channels are transmembrane proteins because they must communicate the information coming from outside the cell to the intracellular machinery or permit the passage of polar molecules and ions through the hydrophobic lipid bilayer. Some examples are voltage-gated ion channels, ABC transporters and GPCRs like rhodopsin.

Integral membrane proteins can only be solubilized by using detergents that disrupt the hydrophobic interactions within the lipid bilayer. Above a critical concentration (CMC), which depends on pH, temperature and salt concentration, the detergent monomers aggregate to form micelles. A strong ionic detergent that denatures even the most hydrophobic proteins is sodium dodecyl sulfate (SDS), which is also used for their separation by polyacrylamide-gel electrophoresis according to their molecular weight as discussed below.

Peripheral membrane proteins are associated to the membrane only temporarily by interacting with other proteins or directly with the lipid bilayer by means of H-bonds and hydrophilc or electrostatic forces. G-proteins and certain protein kinases interact with transmembrane proteins and the lipid bilayer simultaneously. Another example is the above mentioned phospholipase C, which takes part in lipid signalling. A special class of peripheral membrane proteins are those that are attached on the noncytosolic surface of the membrane by a GPI (glycosylphosphatidylinositol) anchor.

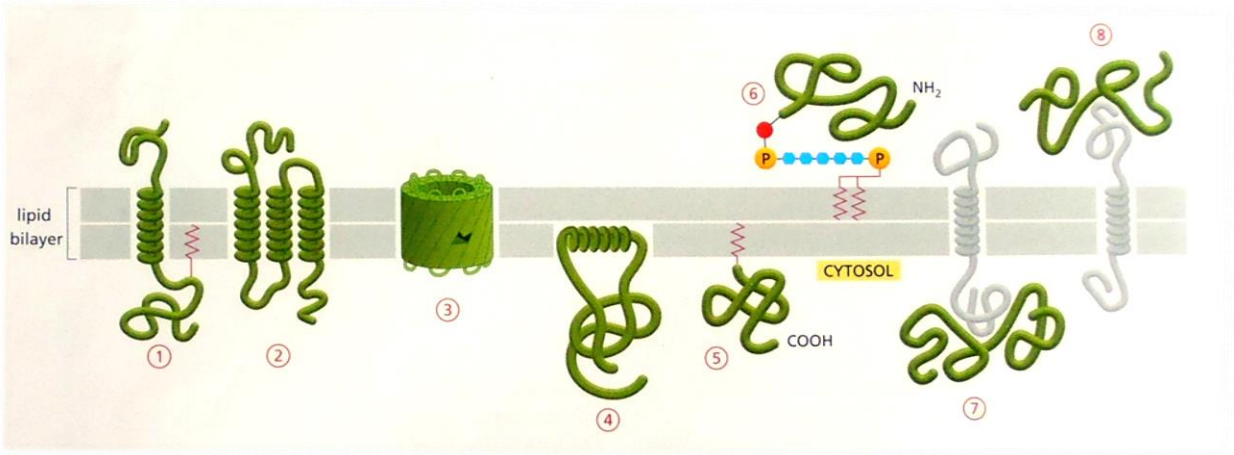


Fig. 1. Various ways in which membrane proteins associate with the lipid bilayer: (1) as a single α helix, (2) as multiple α helices, (3) as a rolled up β sheet, (4) anchored to the cytosolic monolayer of the lipid bilayer by a hydrophobic α helix, (5) attached by a lipid chain, (6) via a GPI anchor that consists of an oligosaccharide and phosphatidylinositol whose fatty acid chains insert into the membrane, (7,8) by non-covalent interactions with other membrane proteins[3].

Cells of an organism must communicate with each other to function as a whole, passing information by signalling a complex body of information concerning the vitality of the cell and the organism, including cell growth, metabolic events, gene expression, immune response or the maintenance of the balance of biologically important molecules[4]. Transmembrane proteins perform as receptors on the target cell receiving the signal. Upon binding of a

ligand, these receptor molecules become activated and trigger an entire signalling cascade (signal transduction).

Most cell surface receptor proteins belong to one of three classes: *transmitter-gated ion channels (TGIC), G-protein-coupled receptors (GPCR)* and *enzyme-coupled receptors (ECR)*[5]. *Ion channels* are a type of transmembrane channel responsible for the passive transport of ions (sodium, potassium, calcium, hydrogen, magnesium and chloride) and small molecules. They consist of pores within the lipid bilayer of the cell membrane formed by integral membrane-protein complexes. Ion channels are not as selective as transporters and pumps, differentiating solutes primarily by size and ionic charge but the rate of ion transport through the channel is very high. Thus, ion channels are key components in a wide variety of biological processes that involve rapid changes in cells, such as cardiac, skeletal, and smooth muscle contraction, epithelial transport of nutrients and ions, T-cell activation and insulin release. They can be voltage gated (ions) or ligand-gated (e.g. GABA-gated receptor) ion channels[6]. Voltage-gated channels are critical to the production of an action potential in neurons resulting in a nerve impulse. Membrane transporters in contrast perform active transport using ATP. Neurotransmitter transporters (e.g. Dopamin transporter) and ion pumps (the Na+/K+ antiporter, an ATPase) that move ions across a plasma membrane against their concentration gradient, are examples of these.

G-protein coupled receptors (GPCR) have seven transmembrane domains and are activated by a variety of specific ligands such as hormones, neurotransmitters, light and olfactory stimulatory molecules, γ-aminobutyric acid (GABA), chemokines, lipid mediators of inflammation (e.g. prostaglandins and platelet-activating factor) or proteins such as endothelin and vasopressin[7]. Signal transduction by GPCR occurs by phosphorylation of a trimeric G-protein (GTP-binding protein), which leads to the activation of the target protein (protein kinases as well as various ion channels). Some GPCRs are coupled to a stimulatory G protein, which activates adenylyl cyclase, a plasma-membrane-bound enzyme converting ATP to cAMP (an intracellular signalling molecule). Regulation of the target proteins is achieved by their phosphorylation by cAMP-dependent protein kinases (PKA). Adrenalin receptors on heart and muscle cells, for example, mediate the effect of the hormone (muscle contraction) in this way. We also owe our sense of smell to GPCR: In olfactory receptor neurons, cAMP causes cAMP-gated ion channels to open and initiates a nerve impulse by depolarization of the neuron. It is also a GPCR, the rhodopsin, that takes part in the sense of vision, but the synaptic signal is created by the decrease in cGMP and the closing of cGMP-gated cation channels.

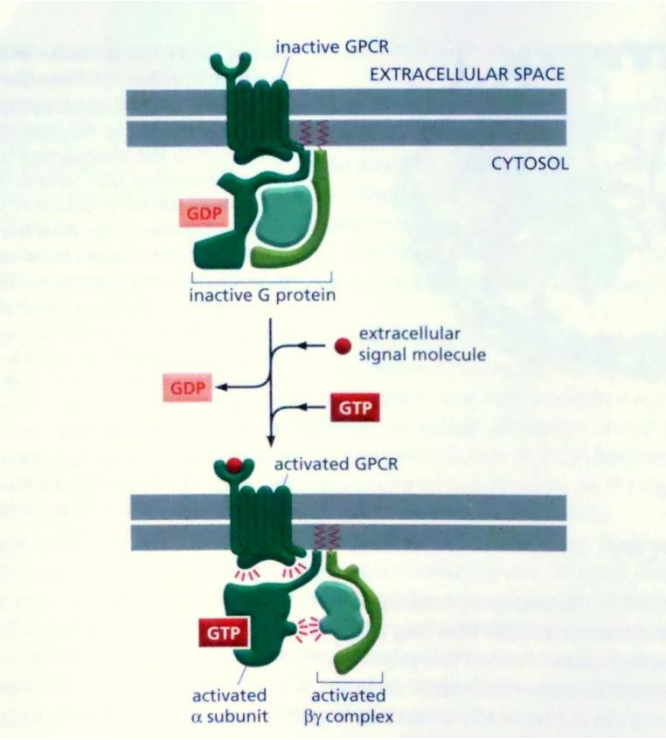


Fig. 2. An extracellular signal binding to a GPCR activates the receptor by causing a conformational change. The activated receptor in turn alters the conformation of the G protein, which exchanges the GDP bound on its α subunit for GTP. In this activated form, the G protein can regulate the activity of several proteins in the plasma membrane[8].

Another group of GPCR exert their effects via a G-protein that activates phospholipase C-β, also a plasma-membrane-bound enzyme, that activates the inositol phospholipid signalling pathway leading to the phosphorylation of proteins by a Ca^{2+} - dependent protein kinases (PKC). As an example, the thrombin receptor on blood platelets that initiates the signal cascade which causes them to aggregate, belongs to this group.

Enzyme-coupled receptors are usually single-pass transmembrane proteins, most of which are protein kinases, whose cytosolic domain has enzyme activity[9]. A major group of enzyme-coupled receptors are the receptor tyrosine kinases (RTK), which bind several growth factors such as the EGF, VEGF, FGF and insulin, that control cell growth and proliferation. The mitogen-activated protein (MAP)-kinase signalling pathway controls several gene regulatory proteins and is initiated by the binding of a ligand on a RTK. Besides stimulating the phospholipid signalling pathway directly by binding of a phospholipase C, an insulin-like growth factor (IGF) binding on a RTK triggers another mechanism mediated by phosphatidylinositides, that leads to cell growth and inhibition of apoptosis.

The second important group of enzyme-coupled receptors are those lacking kinase activity and which depend on cytoplasmic tyrosine kinases for phosphorylation. Antigen and interleukin receptors on lymphocytes, integrins and receptors for various cytokines and the

growth hormone belong to this group. When an interferon (or a growth hormone) binds on its specific receptor, a signalling pathway (JAK-STAT) is mobilized, which brings about inflammatory and immune mechanisms (or stimulates growth by production of IGF)[10].

Among others, an emphasis has to be laid on the observation of certain membrane proteins which are of interest due to their pivotal and decisive role in cell signalling, communication and the determination of the cell fate. Receptor tyrosine-protein phosphatases (PTPs) are signaling molecules with one transmembrane domain that regulate a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation. Beside the named JAK proteins, CD proteins (clusters of differentiation) are important molecules that participate in immune system processes. They act as receptors or ligands and play a role in cell signalling, adhesion, growth and differentiation. Multidrug-Resistance-Protein (MDR1, CD243) is a multipass membrane protein (12 TMDs) that pumps xenobiotics through the membrane and out of the cell by active transport. It belongs to the superfamily of ATP-binding cassettes, which used the energy of ATP hydrolysis to transport ions and various organic molecules (peptides, sugars, lipids etc.) across the plasma membrane as well as intracellular membranes. Tumor necrosis factor receptors bind cytokines known as tumor necrosis factors, which control apoptosis, cell proliferation and immune response. As will be discussed below, representative members and subunits of these vital proteins have been identified.

1.2. Shotgun Proteomics – the Universal Method

Proteomics refers to the study of the sum of proteins expressed in a cell at a given time and under certain conditions. Despite remarkable progress using MS to investigate biological problems, the challenges of proteomics are daunting[11]. Extraction of the proteins as well as their digestion and analysis necessitates solubilizing them, which requires the use of detergents that are incompatible with mass spectrometry. The complexity of biological samples demands a prefractionation step that precedes MS, consisting of a chromatographic (IEC, SEC, RP) or electrophoretic technique (SDS-PAGE). Moreover, for identification, the proteins have to be digested into peptides. However, this process is not uniform and depends on factors such as the peptide sequence and the digestion time. Unlike oligonucleotides, proteins cannot be amplified and therefore sensitivity is of the essence, especially for the analysis of membrane proteins involved in cell communication and signalling, which are effective in very low concentrations.

Because protein expression is affected under environmental, biological, pharmacological, and disease conditions, detection of variations in the diversity and extent of their production provides information about the different mechanisms taking place in the cell under these

conditions. Thus, the most fundamental step in proteome profiling is the culturing of a specific cell type under definite conditions (medium, stimulants, physical environment) and isolation of relevant fractions such as those containing the cytoplasm or plasma membrane from the cultured cells by several centrifugation and solubilization steps.

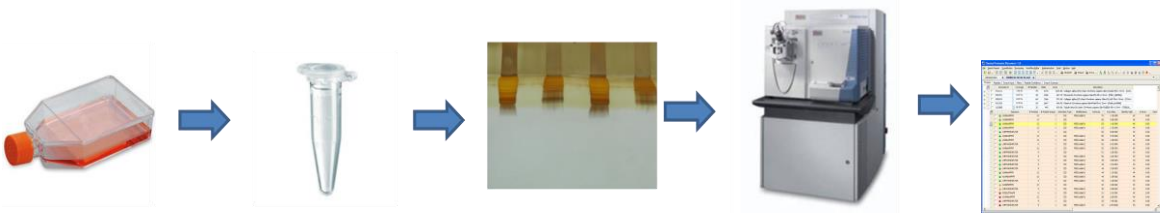


Fig. 3. Proteome profiling using shotgun proteomics involves the steps of isolation of fractions from cultured cells, electrophoretic separation of the proteins in the sample fraction, digestion with a protease such as Trypsin, chromatographic separation of the protein digests, and mass spectrometric analysis in the form of µLC-MS/MS. The "fingerprint" of each peptide's fragmentation mass spectrum is used to identify the protein from which they derive by searching against a sequence database with commercially available software (e.g. MASCOT).

There are two major strategies for converting proteins extracted from biological material to peptides suitable for mass spectrometry-based proteome analysis[12]. The first involves solubilization of proteins with detergents, separation of proteins by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis and digestion of the gel-trapped proteins (“in-gel” digestion). The second method comprises protein extraction with strong chaotropic reagents such as urea, thiourea or guanidinium hydrochloride and digestion of proteins on filter under denaturing conditions (‘in-solution’ digestion). Advantages of in-gel digestion include its robustness against impurities, which interfere with digestion, and the better solubilization of proteins by SDS. Furthermore, the sample complexity is greatly reduced, as it consists a separation step before mass spectrometric analysis. One disadvantage is that, the gel may prevent peptide recovery, meaning a proportion of the peptides not being eluted from the gel pieces. In-solution digestion saves a huge amount of time in sample preparation, but the proteome may be incompletely solubilized, and digestion may be impeded by interfering substances. SDS is indispensible for complete solubilization of proteins, however, detergents, even in small concentrations, can preclude enzymatic digestion and dominate mass spectra owing to their ready ionizability and their great abundance compared to individual peptides. Therefore, in case of in solution digestion, depletion of SDS is necessary for efficient mass-spectrometric analysis in proteomics. An alternative to SDS is the ionic detergent sodium deoxycholate, which can be removed by acid precipitation and is then compatible with tryptic digestion.

Yet another disadvantage of 'in-solution' digestion is that a separation step is forfeited and sample complexity is substantially increased. A single cell population may consist of a complex mixture of up to 30,000 proteins, and the wide dynamic range of protein expression brings in added difficulty to their analysis. Digestion of the proteins with a proteolytic enzyme on the other hand, dramatically increases sample complexity. For this reason, in studies of low-abundant membrane proteome, the separation of sample components, by techniques such as electrophoresis and chromatography prior to mass spectrometric analysis is crucial. By using two or more independent separation mechanisms in a consecutive manner, the peak capacity, which corresponds to the maximum number of components that can be resolved by the chromatographic system, is dramatically increased[13]. Separation methods prior to MS improves ionization efficiency and minimizes ion suppression by highly abundant peptides, which would obscure the detection of other peptides present in the mixture, such as those of the low abundant membrane proteins. A large number of chromatographic and electrophoretic separation modes, offering a wide range of selectivities e.g., size, hydrophobicity, charge etc. of the molecules of interest, may be combined to increase the peak capacity. For example, isoelectric focusing, which separates components on the basis of charge, is orthogonal to RPLC, which separates them on the basis of hydrophobicity.

Electrophoresis is usually performed as a one dimensional SDS-PAGE (polyacrylamide electrophoresis with sodium dodecyl sulfate) by which proteins are separated according to their molecular weight. The forces acting on a charged particle in an electrical field are the electrical force

$$F_e = q.E$$

which accelerates the particle, and the force of friction

$$F_f = f.V$$

with q the charge of the particle, E the electrical field, f the friction coefficient and V the velocity of the particle. The friction coefficient depends on the viscosity of the medium as well as the pore size. These two forces reach an equilibrium so that the particle moves with a constant velocity.

$$F_e = F_f$$

$$q.E = f.V$$

$$\mathrm{V} = \frac{\mathrm{q.E}}{\mathrm{f}}$$

If the term *q* / *f* is substituted as the mobility *u*

$$V = u.E$$

For peptides and proteins, the mobility is roughly equal to

$$u = \frac{q}{M^{2/3}}$$

Because the weakly acidic and basic groups of peptides and proteins are partially dissociated, the effective mobility depends on the degree of dissociation[14]

$$u_{eff} = \sum \alpha . u$$

Hence, the velocity and the resolution can be can be optimized via the pH value of the electrolyte. Certain other factors also influence the quality of separation in gel electrophoresis. At pH > 7, surfaces like fused silica are negatively charged and these are compensated by positive charges in the solvent. Because the electroosmotic flow, which is the migration of these positive charges in the solvent towards the cathode, has a direction opposite to that of sample ions, it may distort zones and lead to a loss of focus. Moreover, dispersion of the sample zone due to diffusion and convection should be avioded as far as possible to maintain a high focus of separation. Discontinuous buffer systems involving a stacking gel prior to the separation gel minimize peak dispersion due to injection. Avoiding a high concentration of salt in the loaded sample as well as using an appropriate buffer contribute also to narrow sample bands. Heat dissipation is another important issue in electrophoresis because of the effect of high temperatures on convection and gains an even higher significance in view of the sensitivity of proteins to heat.

In polyacrylamide gel electrophoresis under native conditions using SDS, protein molecules get a highly negative charge and their own charge become insignificant so that the separation depends on the size of the molecule only. The electrophoresis is performed in a vertical gasket holding the glass plates between which the gel is polymerized. Polyacrylamide gels are stable and transparent, and are attained by copolymerizing acrylamide monomers with the crosslinker N,N'-methylenebisacrylamide. Two parameters, the total acrylamide concentration (T%) and the relative amount of crosslinker (C%) determine the pore size (3-5 nm). After polymerizing the separation gel, the stacking gel is pipetted between the glass plates and a comb inserted on the yet nonpolymerized stacking gel to form the slots where the samples are loaded. These are loaded using micropipettes and molecular weight markers are run in a separate lane to calibrate the gel, determine the

approximate molecular mass of the analytes and to be able to monitor the migration distance.

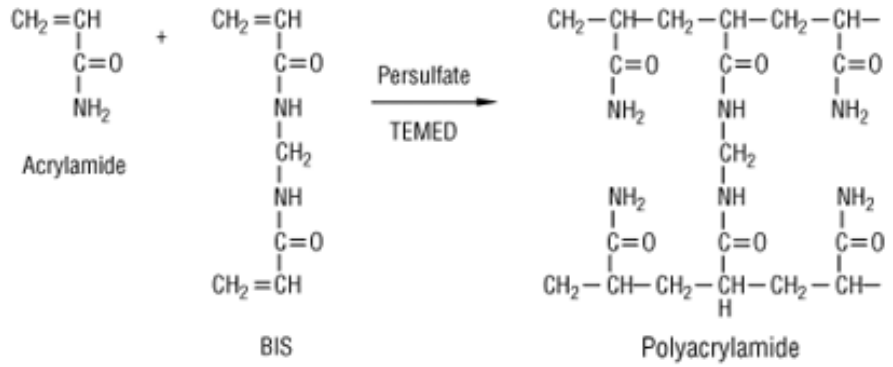


Fig. 4. Structure of a polyacrylamide gel[15].

Several proteomic studies have also made use of two-dimensional gel electrophoresis[16], where proteins are separated according to their isoelectric point on the second dimension. However, for the analysis of membrane proteins, this method has disadvantages over one dimensional GE because of the high pI, hydrophobicity and low-abundance of hydrophobic proteins. Many hydrophobic proteins are not solublilized in the non-detergent isoelectric focusing sample buffer. Secondly, solubilized proteins are prone to precipitation at their isoelectric point. The limitation of the one dimensional approach is the increased protein complexity in each gel band. However, this problem can easily be overcome either by the use of liquid chromatography to resolve the extracted peptides and by increased mass accuracy of the mass detection.

Polyacrylamide (PAA) gels are attained by polymerizing a mixture of acrylamide- and bisacrylamide-monomers; the resulting “pore” size depends on the total monomer concentration (%T) and the content of the cross-linker in the mixture (%C). Usually PAA gels of 12 % T are used. Ammonium peroxodisulfate is used to initiate the polymerization reaction, TEMED stabilizes the forming radicals. Prior to separation, the samples are pre-concentrated on a stacking gel with a smaller mesh than the separation gel. For the stacking gel, a buffer with pH 6,8 is commonly used, whereas in the separation zone the pH is 8,8. Just before loading onto the gel, a certain amount of sample containing about 20 µg/µl protein is mixed with loading buffer. This buffer contains mercaptoethanol as reducing agent for cleaving the disulfide bridges in the proteins, causing protein linearization and facilitating the interaction with SDS, which is also contained in the loading buffer.

After running the gel, the proteins are fixed with a solution of acetic acid and ethanol and a silver staining is performed to render the protein bands visible. Excision of the stained protein bands follows, which are then destained with a solution of potassium ferrocyanide $K_4[Fe(CN)_6]$, washed and buffered up to a pH of about 8 (the optimal range for the activity of trypsin). Reduction of the disulfide bridges by dithiothreitol at 56 °C serves to unfold the proteins and to prevent the oxidation of sulfhydryl groups of cysteine residues, which are then alkylated by iodoacetamide.

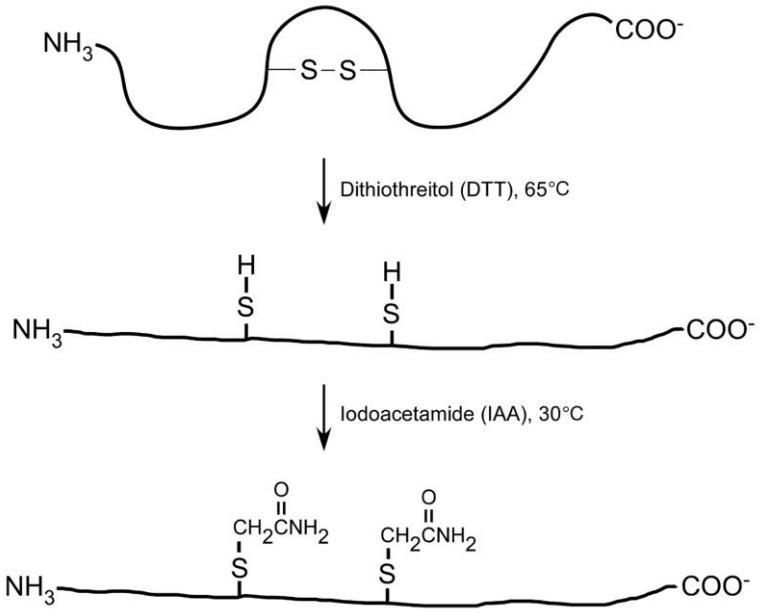


Fig. 5. Reduction and alkylation of the disulfide bridges[17].

The proteins on the gel pieces are digested by trypsin over night and eluted the next day with 50 % acetonitrile. The samples containing the eluted peptides are dried with a SpeedVac concentrator and analyzed by nLC-ESI-MS.

The separation of the peptides by high-pressure reverse phase liquid chromatography (RP-HPLC) is based on their partitioning between the mobile phase of a certain polarity and the lipophilic stationary phase. The sorbent is a granular material made of solid particles (silica) and is derivatized with hydrophobic alkyl chains in the case of reverse phase chromatography. The mobile phase consists of a mixture of solvents (acetonitrile, water or methanol) and its composition plays a major role in the separation of the analytes. The pressurized mobile phase and a sample mixture is pumped through a column filled with a

sorbent, leading to the separation of the sample components. The more hydrophobic a peptide is, the higher its adsorption on the alkyl chains (C18) of the stationary phase and the later it will elute. The time at which a specific analyte elutes is called its retention time and is characteristic of a given analyte under particular conditions. Typically the composition of the mobile phase follows a gradient from hydrophilic to hydrophobic to shorten the retention time and achieve a better separation with sharper peaks. In case of mass spectrometric detection, additives such as formic acid are employed, to prevent loss of analytes during loading on the column and also to assist the ionization of the analytes, forming an ion pair. Commonly a trapping column is used for concentration and cleaning of the sample before it is switched in-line with the analytical column after the loading and washing steps. Several parameters can contribute to selectivity in RPLC, such as pH, ion-pairing agent, type of stationary phase (C4, C18), organic solvent, temperature, pore size, support material.

NanoLC is a miniaturized HPLC technique performed using columns with an internal diameter between 10-150 µm and flow rates in the nanoliter per minute (200-300 nL/min) range. The main advantage of using smaller columns in HPLC is the increased detection sensitivity that can be obtained as a result of reduced sample dilution. For that reason nanoLC is ideally suited for the analysis of low-abundant proteins. In addition, the low flow rate used in nanoLC facilitates its coupling to the ionization source of a mass-spectrometer. When performing automated LC-MS experiments, verification of the pressure curves of the complete cycle, the peak heights, the chromatographic resolution and the retention times of the standard peptides contributes to the reproducibility and accuracy of measurements[18].

Mass spectrometry has become increasingly important for the analysis of proteins in the last decade. It can be used to study the proteins expressed in a cell, determine the members of protein complexes and their structure, and the dynamics of various processes in the cell. It can also focus on the analysis of the post-translational modifications of proteins as well as enabling quantitative analyses by various labeling methods. A mass spectrometer determines the the m/z ratio of molecules by first ionizing them and then subjecting them to an electric field. At a certain combination of direct and alternative fields, only ions of a definite m/z ratio follow a stable trajectory and reach the detector, which either records the image current produced when ions pass by electrodes (as in an Orbitrap) or the charge induced when ions hit a surface. By scanning through a range of electric fields, a mass spectrum can be obtained consisting of m/z values against the signal registered.

In proteomic studies, one of the possibilities to transfer the molecules of interest and simultaneously ionize them is electrospray ionization (ESI), which is termed a soft ionization technique because it induces very little fragmentation, allowing the observation of molecular

ions. Another characteristic of ESI is that, it produces multiply charged ions so that the m/z values of macromolecules of hundreds of kilodaltons fall in the mass range of the analyser. The liquid containing the analytes (the eluent in case of LC-MS) is delivered through a fused silica capillary under high voltage (about +1,5 kV) and forms a Taylor cone at its tip, emitting a jet of charged aerosol toward the capillary (counterelectrode) that leads into the mass spectrometer. Assisted by the nebulizer gas (N_2), the high temperature of the capillary (350 – 450 °C) as well as the volatile additives in the solvent such as formic acid, the solvent evaporates from the aerosol, creating droplets with an increasingly higher ratio of charge to radius. Upon reaching their Rayleigh limit, the droplets burst into smaller ones (Coulomb fission) losing a relatively high proportion of their charge. It is the field desorption of solvated ions as well as further solvent evaporation which contributes to the formation of gas-phase ions.

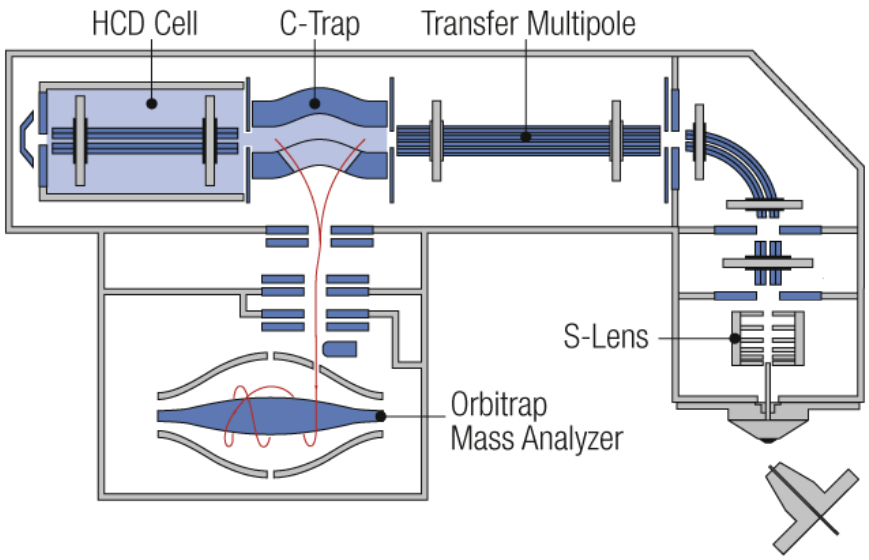


Fig. 6. The structure of the hybrid quadrupole – Orbitrap mass spectrometer *Q Exactive* by Thermo Fisher Scientific (http://planetorbitrap.com).

In an Orbitrap instrument, a type of Fourier transform (FT) ion trap mass analyzer developed by A. Makarov, ions rotate around an spindle-shaped central electrode on elliptical trajectories. They also move back and forth along the axis of the central electrode with a frequency which only depends on the m/z ratios of the ions[19]:

$$\omega = \sqrt{\frac{k}{m/z}}$$

where,

k ..... force constant,

m.....mass,

z.....charge.

The time-domain signal, which is recorded as the image current induced on the outer electrodes as the analyte ions oscillate, is converted into a frequency and then into a m/z spectrum[20]. Beside the m/z ratios, the fragmentation of peptides by electron-transfer dissociation (ETD), collision-induced dissociation (CID), or as in the case of Orbitrap, higher-energy collisional dissociation (HCD yields additional information for protein identification. The peptides which were detected at high intensities are isolated in the quadrupole mass filter, pass through a C-trap and are injected into the collision cell, where dissociation takes place. The ions are then returned to the C-trap before injection into the orbitrap for mass analysis. The resulting spectrum, called an MS/MS (MS^2 or tandem) spectrum, is basically a list of m/z ratios for different fragments, with some of the differences corresponding to the specific mass of one amino acid[21].

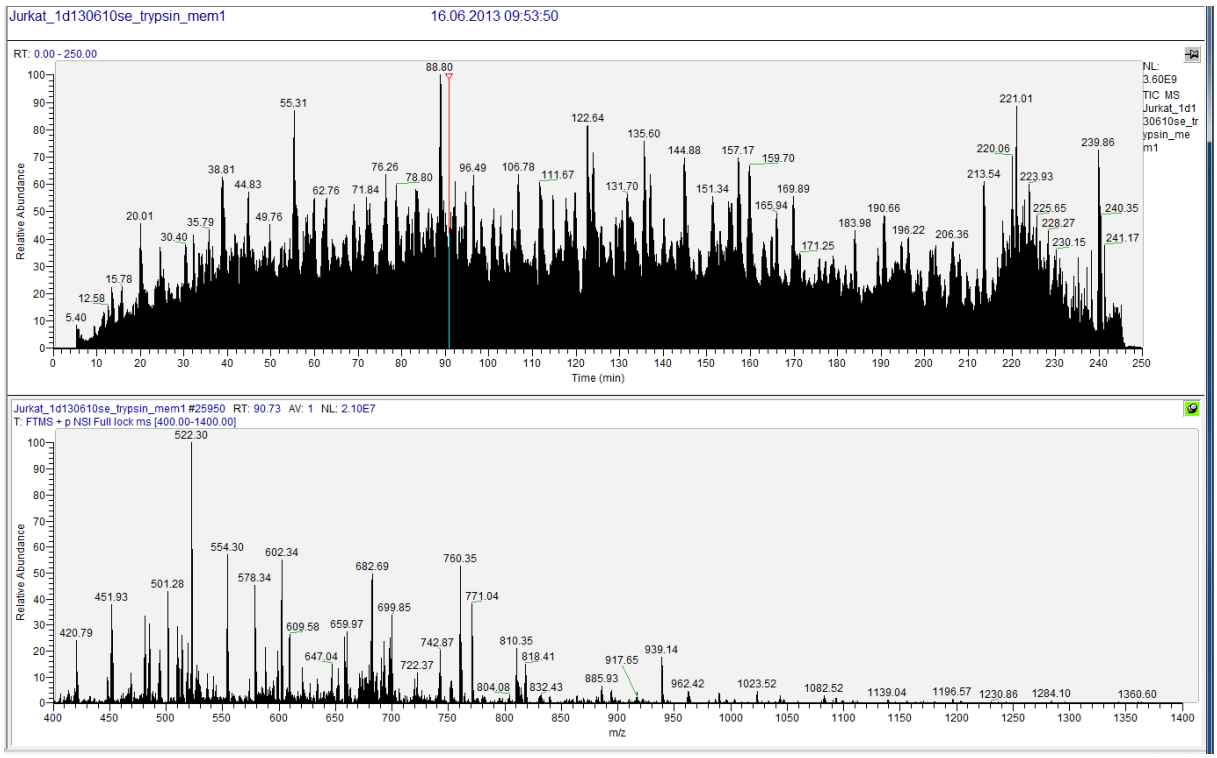


Fig. 7. An example of a mass chromatogram (above) and a spectrum showing detected fragment ions at 90.73 minutes (below).

If the charge is retained on the N terminal fragment, the ion is classified as either a, b or c. If the charge is retained on the C terminal, the ion type is either x, y or z. The subscript indicates the number of residues consisting the fragment. The types of fragment ions observed in an MS/MS spectrum depend on many factors including primary sequence, bond energy, mode of fragmentation, charge state, etc[22].

Resolution, defined as the full width of the peak at half maximum (FWHM) divided by the mass of the peak, is essential in proteomics: high resolution allows the mass spectrometer to distinguish thousands of different peptides from each other with varying ionization efficiency. The resolving power of an orbitrap is proportional to the angular frequency (ω) of harmonic oscillations of the ions; as a result, the resolving power is inversely proportional to the square root of m/z and proportional to acquisition time:

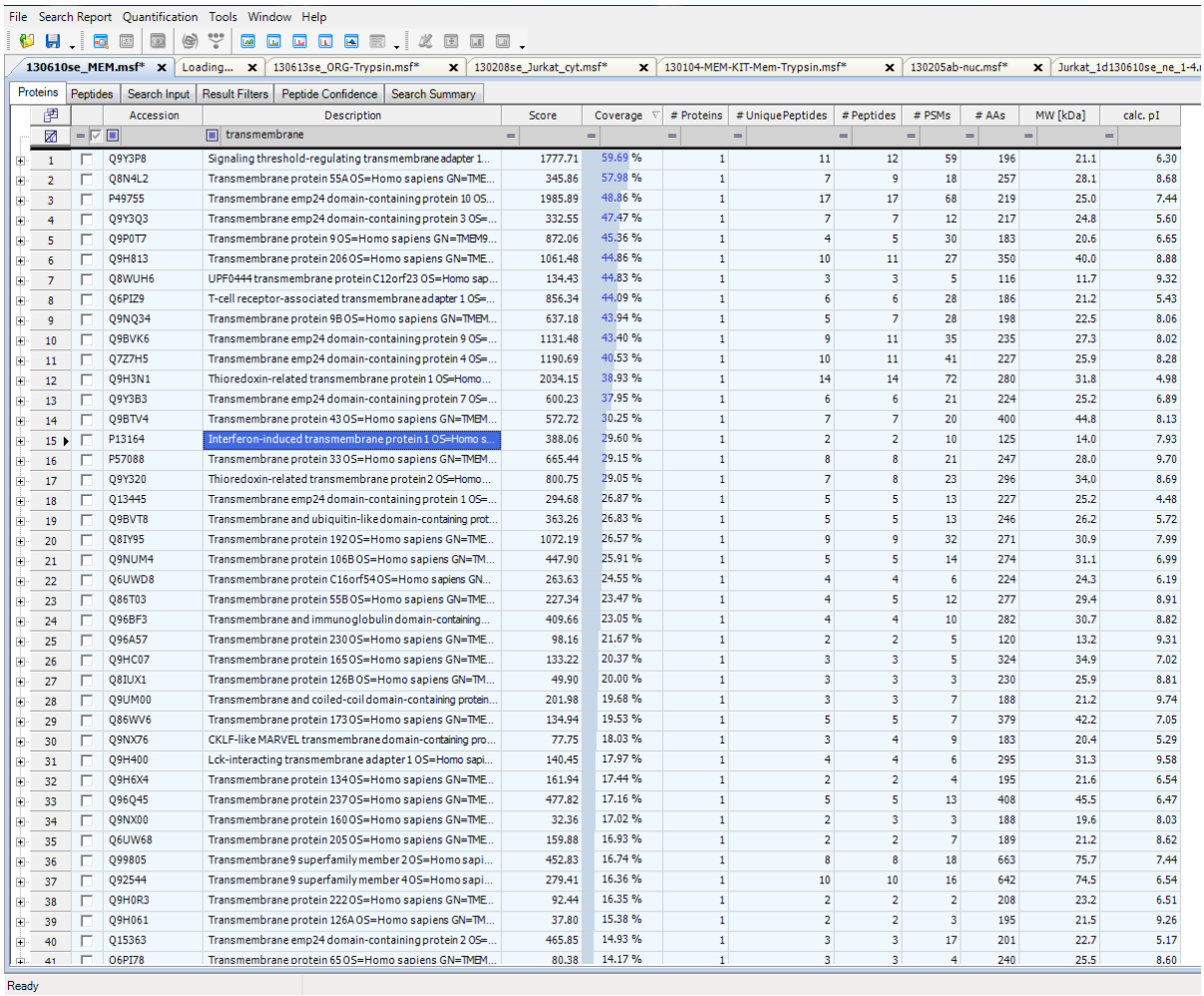
$$R \sim \frac{t}{\sqrt{m/z}}$$

As an example, the maximum resolution of at m/z 200 on the Q Exactive instrument (Thermo Fisher Scientific) is 140,000, whereas the maximum resolution at m/z 800 is 70,000. The time needed for detection of an ion packet at this resolution is 1 – 1,5 seconds; to acquire data at a higher resolution, a longer time period would be necessary. Ongoing developments of the Orbitrap mass analyzer are addressing the trade-off between resolving power and scan speed. In addition to high resolution, orbitrap mass spectrometers also provide a high mass accuracy in the below-ppm range, a high dynamic range and sensitivity, owing to the method of detection based on Fourier transform. By means of internal calibration with ambient contaminants (lock mass option) or based on the deviation of mass values of charge pairs, acquisition of mass accuracy in the ppb range has been demonstrated[23]. High-resolution extracted ion chromatography combined with accurate mass is essential for precise comparison between different functional sates of a biological system and identification of unknowns in metabolomics and lipidomics analyses[24].

Protein identification is achieved by searching the exact masses of the fragment peptides and the generated tandem mass spectra against a proteome database (peptide mass fingerprinting, PMF). The data analysis software that performs the search translates the known genome of the organism into proteins, then theoretically cuts the proteins into peptides with same protease used in the experiment (e.g. trypsin), and calculates the absolute masses of the peptides from each protein. It then compares the masses of the

peptide fragments of the unknown protein to the known peptide masses of each protein encoded in the genome. The results are statistically analyzed to find the best match. The superiority of this method in terms of the speed of analysis and evaluation of the acquired results, given the huge size of data, which would be impossible to interpret manually, is apparent. Another advantage is that the search is performed against known masses of peptides and de-novo sequencing is not necessary.

With the help of the multidimensional protein identification technology (MudPIT), it is possible to identify a large number of soluble and membrane proteins[25]. The software with which the search is performed also includes algorithms that assign scores to interpreted data, showing the statistical probability of the observation.



	Accession	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs	# AAs	MW [kDa]	calc. pI
		transmembrane									
1	Q9Y3P8	Signaling threshold-regulating transmembrane adapter 1...	1777.71	59.69 %	1	11	12	59	196	21.1	6.30
2	Q8N4L2	Transmembrane protein 55A OS=Homo sapiens GN=TME...	345.86	57.98 %	1	7	9	18	257	28.1	8.68
3	P49755	Transmembrane emp24 domain-containing protein 10 OS...	1985.89	48.86 %	1	17	17	68	219	25.0	7.44
4	Q9Y3Q3	Transmembrane emp24 domain-containing protein 3 OS=...	332.55	47.47 %	1	7	7	12	217	24.8	5.60
5	Q9P0T7	Transmembrane protein 9 OS=Homo sapiens GN=TMEM9...	872.06	45.36 %	1	4	5	30	183	20.6	6.65
6	Q9H813	Transmembrane protein 206 OS=Homo sapiens GN=TME...	1061.48	44.86 %	1	10	11	27	350	40.0	8.88
7	Q8WUH6	UPF0444 transmembrane protein C12orf23 OS=Homo sap...	134.43	44.83 %	1	3	3	5	116	11.7	9.32
8	Q6PIZ9	T-cell receptor-associated transmembrane adapter 1 OS=...	856.34	44.09 %	1	6	6	28	186	21.2	5.43
9	Q9NQ34	Transmembrane protein 9B OS=Homo sapiens GN=TMEM...	637.18	43.94 %	1	5	7	28	198	22.5	8.06
10	Q9BVK6	Transmembrane emp24 domain-containing protein 9 OS=...	1131.48	43.40 %	1	9	11	35	235	27.3	8.02
11	Q7Z7H5	Transmembrane emp24 domain-containing protein 4 OS=...	1190.69	40.53 %	1	10	11	41	227	25.9	8.28
12	Q9H3N1	Thioredoxin-related transmembrane protein 1 OS=Homo...	2034.15	38.93 %	1	14	14	72	280	31.8	4.98
13	Q9Y3B3	Transmembrane emp24 domain-containing protein 7 OS=...	600.23	37.95 %	1	6	6	21	224	25.2	6.89
14	Q9BTV4	Transmembrane protein 43 OS=Homo sapiens GN=TMEM...	572.72	30.25 %	1	7	7	20	400	44.8	8.13
15	P13164	Interferon-induced transmembrane protein 1 OS=Homo s...	388.06	29.60 %	1	2	2	10	125	14.0	7.93
16	P57088	Transmembrane protein 33 OS=Homo sapiens GN=TMEM...	665.44	29.15 %	1	8	8	21	247	28.0	9.70
17	Q9Y320	Thioredoxin-related transmembrane protein 2 OS=Homo...	800.75	29.05 %	1	7	8	23	296	34.0	8.69
18	Q13445	Transmembrane emp24 domain-containing protein 1 OS=...	294.68	26.87 %	1	5	5	13	227	25.2	4.48
19	Q9BVT8	Transmembrane and ubiquitin-like domain-containing prot...	363.26	26.83 %	1	5	5	13	246	26.2	5.72
20	Q8IY95	Transmembrane protein 192 OS=Homo sapiens GN=TME...	1072.19	26.57 %	1	9	9	32	271	30.9	7.99
21	Q9NUM4	Transmembrane protein 106B OS=Homo sapiens GN=TM...	447.90	25.91 %	1	5	5	14	274	31.1	6.99
22	Q6UWD8	Transmembrane protein C16orf54 OS=Homo sapiens GN...	263.63	24.55 %	1	4	4	6	224	24.3	6.19
23	Q86T03	Transmembrane protein 55B OS=Homo sapiens GN=TME...	227.34	23.47 %	1	4	5	12	277	29.4	8.91
24	Q96BF3	Transmembrane and immunoglobulin domain-containing...	409.66	23.05 %	1	4	4	10	282	30.7	8.82
25	Q96A57	Transmembrane protein 230 OS=Homo sapiens GN=TME...	98.16	21.67 %	1	2	2	5	120	13.2	9.31
26	Q9HC07	Transmembrane protein 165 OS=Homo sapiens GN=TME...	133.22	20.37 %	1	3	3	5	324	34.9	7.02
27	Q8IUX1	Transmembrane protein 126B OS=Homo sapiens GN=TM...	49.90	20.00 %	1	3	3	3	230	25.9	8.81
28	Q9UM00	Transmembrane and coiled-coil domain-containing protein...	201.98	19.68 %	1	3	3	7	188	21.2	9.74
29	Q86WV6	Transmembrane protein 173 OS=Homo sapiens GN=TME...	134.94	19.53 %	1	5	5	7	379	42.2	7.05
30	Q9NX76	CKLF-like MARVEL transmembrane domain-containing pro...	77.75	18.03 %	1	3	4	9	183	20.4	5.29
31	Q9H400	Lck-interacting transmembrane adapter 1 OS=Homo sapi...	140.45	17.97 %	1	4	4	6	295	31.3	9.58
32	Q9H6X4	Transmembrane protein 134 OS=Homo sapiens GN=TME...	161.94	17.44 %	1	2	2	4	195	21.6	6.54
33	Q96Q45	Transmembrane protein 237 OS=Homo sapiens GN=TME...	477.82	17.16 %	1	5	5	13	408	45.5	6.47
34	Q9NX00	Transmembrane protein 160 OS=Homo sapiens GN=TME...	32.36	17.02 %	1	2	3	3	188	19.6	8.03
35	Q6UW68	Transmembrane protein 205 OS=Homo sapiens GN=TME...	159.88	16.93 %	1	2	2	7	189	21.2	8.62
36	Q99805	Transmembrane 9 superfamily member 2 OS=Homo sapi...	452.83	16.74 %	1	8	8	18	663	75.7	7.44
37	Q92544	Transmembrane 9 superfamily member 4 OS=Homo sapi...	279.41	16.36 %	1	10	10	16	642	74.5	6.54
38	Q9H0R3	Transmembrane protein 222 OS=Homo sapiens GN=TME...	92.44	16.35 %	1	2	2	2	208	23.2	6.51
39	Q9H061	Transmembrane protein 126A OS=Homo sapiens GN=TM...	37.80	15.38 %	1	2	2	3	195	21.5	9.26
40	Q15363	Transmembrane emp24 domain-containing protein 2 OS=...	465.85	14.93 %	1	3	3	17	201	22.7	5.17
41	O6PI78	Transmembrane protein 65 OS=Homo sapiens GN=TMEM...	80.38	14.17 %	1	3	3	4	240	25.5	8.60

Fig. 8. An example of a list of identified proteins as displayed on Thermo Proteome Discoverer Viewer.

A disadvantage of PMF is that, whereas the attribution of proteotypic sequences to a peptide are definite, the allocation of non-proteotypic peptides, which may belong to different proteins, is ambivalent and leads to uncertainty in the identification of a certain protein.

Another shortcoming of search algorithms is that, although the false discovery rates of peptides are considered, no information is contained as to which peptides are plausible for a certain sample, even for a certain protein. As an example, N-acetylation of an internal peptide is obviously a false result. For a certain protein, the probability of localization in the cell compartment (e.g. organelles) is not considered[26]. A data evaluation software that would take biological possibilites also into account would bring an enhancement to proteomic studies. On the other hand, plausibilities as to mass spectrometric analysis have to be considered as well. An algorithm that designates the likelihood of a certain peptide being ionized and nebulized is currently not available. As an example, the detection of a zinc finger protein is most likely a false discovery.

The uncertainty in the mass of the expressed protein and the deviation from the mass of the sequence in the database leads to the fact that peptide mass fingerprint can only provide the statistically most probable identification[27].

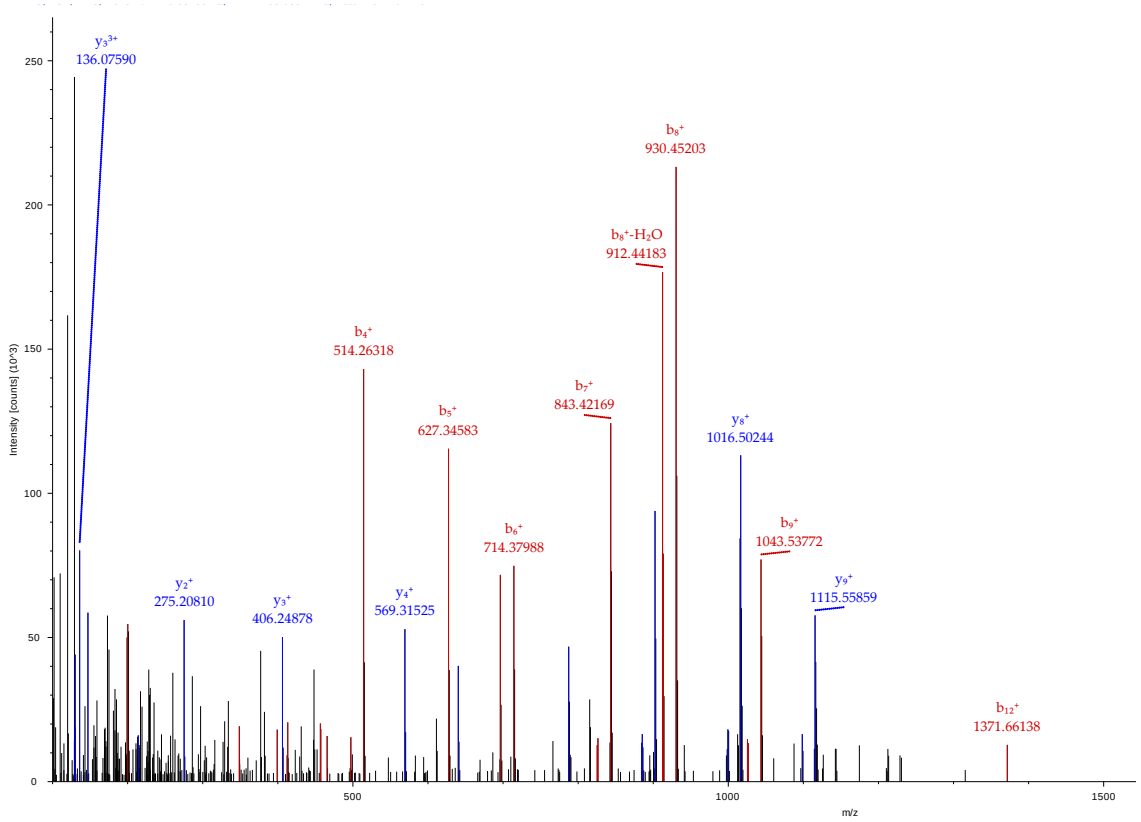


Fig. 9. Fragment match spectrum of the proteotypic peptide RLEDLSESIVNDFAYMKK (m/z = 2158.08) of the protein *Transmembrane emp24 domain-containing protein 10* identified in the membrane fraction of Jurkat cells lyzed by cell disruption with liquid nitrogen.

In addition, the reliability of the spectra must be verified acording to criteria such as the number of peaks assigned, sequence coverage, the type of fragments detected (a spectrum containing only y-series is not as reliable as one with both b- and y-series), and mass accuracy. To sum it up, a manual examination and interpretation is indispensible for every reliable PMF inquiry.

1.3. Shotgun Proteomics for the Analysis of Membrane Proteins

In spite of their enormous significance, the analysis of membrane proteins by standard LC-MS experiments is notoriously difficult. With their low abundance added to their hydrophobicity, the coverage of hydrophobic peptides is usually relatively low.

Because of their poor solubility in aqueous media like those used in standard fractionation and digestion protocols, various membrane solubilization strategies have been developed[28], ranging from solubilization with detergents (SDS), urea, organic solvents, mixed organic-aqueous solvents taking advantage of the amphiphilic nature of membrane proteins, acid labile surfactants (e.g. RapiGest), and organic acids compatible with subsequent proteolytic digestion and analysis by LC/MS, to disruption of membranes by high pH or temperature combined with an organic solvent[29]. The successful use of thiourea[30], zwitterionic surfactants as well as sulfobetaines has also been reported. On the other hand, when mapping protein-protein, protein-drug or protein-ligand interactions or performing activity-based assays, non-denaturing detergents such as n-Dodecyl-D-maltoside (DDM) and Triton X-100 are needed.

Adsorption of viscous membrane-embedded peptides on hydrophobic surfaces such as test tubes and tips during sample preparation also leads to loss of MPs. On the other hand, membrane-spanning segments of proteins are either not readily accessible to proteolytic enzymes or lack the specific proteolytic cleavage sites, which are at the charged amino acids Lysine and Arginine for the commonly used protease trypsin. Reverse phase chromatography with C18 columns also poses a challenge, as the hydrophobic peptides bind on the stationary phase too strongly and elution is problematic. Aggregation of hydrophobic peptides especially at higher temperatures must be avoided as well, which necessitates using urea and a chaotrope (CHAPS) in the sample buffer before loading samples on the gel. The cholic acid derivate CHAPS is a non-denaturing surfactant, especially useful in the purification of membrane proteins that are difficult to sulubilize due to their hydrophobic nature.

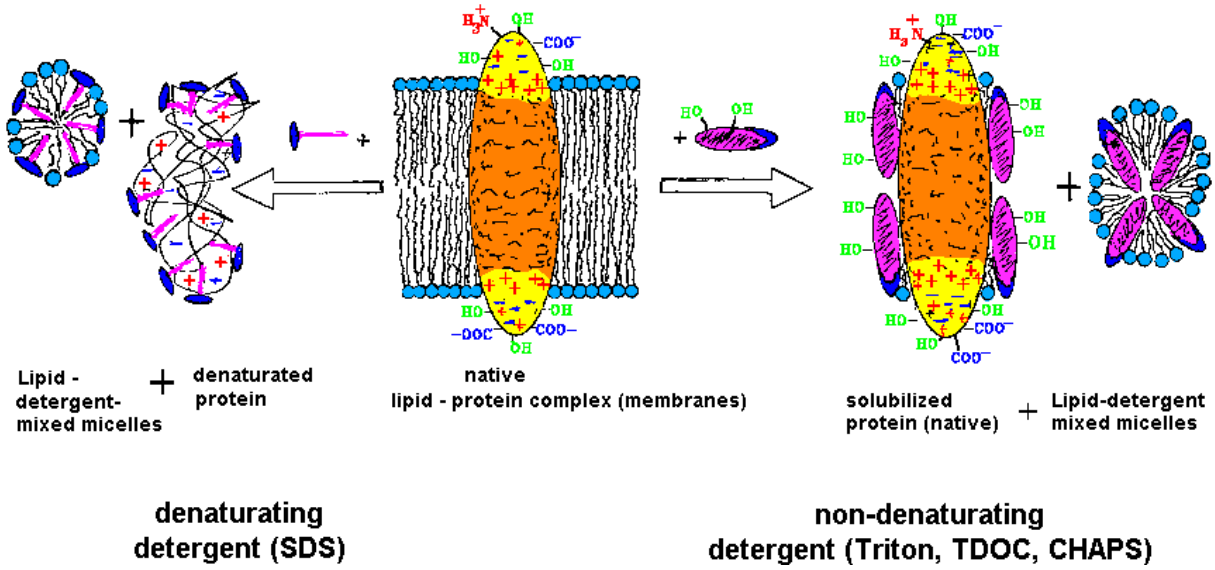


Fig. 10. Solubilization of membrane proteins with denaturing and non-dentauring surfactants[31].

To address these challenges, myriad protocols have been employed in various studies to improve the coverage of the membrane proteins while maintaining compatibility with MS detection. Wu et al. described a proteomic analysis using high pH and proteinase K (hpPK), optimized for the global analysis of both membrane and soluble proteins in unfractionated rat brain homogenizate. Because solubilizing membrane proteins by detergents and organic solvents result in the loss of information about native topology, high pH is employed in this method to disrupt the sealed membrane compartments in the unfractionated sample to produce unsealed membrane sheets with free ends, thus conserving the tertiary structure. Next, the proteins are digested with Proteinase K, analyzed by ESI-MS following 12-step separation on a C18 column. Of the 1610 proteins identified, 454 were predicted to have transmembrane domains. Several membrane proteins such as STX2 and AQP4 were characterized. The majority of the proteins had <20% sequence coverage and the confidence of identification is due to the high redundant sequence coverage (identical sequences identified in a single analysis).

In another study, Wu et al. compared the effect of temperature on chromatographic analysis of plasma membrane fraction digested by trypsin or high pH/proteinase K-CNBr. Especially for the second digestion method, a significantly higher amount of proteins (up to 4-fold) were identified with increasing temperature (20 – 60 °C). However, the problem of elution and bad peak shape persisted. To achieve a notable improvement in the number of integral membrane proteins identified, an increase of temperature up to 90 °C would be necessary, at which the column material degardes irrevocably.

The two-phase on-membrane digestion method employed by Wang et al. did not prove particularly efficient in terms of the total number of membrane proteins, although specific proteins were identified in each of the three regions of the two-phase system, so that to achieve a complete sutudy of membrane proteome, all three fractions need to be digested and analyzed. Although they gave satisfactory results for the analysis of various samples, these methods have had limited success when applied to low-abundant membrane proteins, especially for the identification of those with the function of a transporter (e.g. ABC-transporters) or a receptor (e.g. a GPCR).

In this diploma work, various fractionation, solubilization, digestion, and LC-MS strategies have been tested in an attempt to increase the coverage of membrane proteins with hydrophobic domains and the results were compared with the protein identificatons achieved by the standard protocol for the analysis of cytoplasmic fractions.

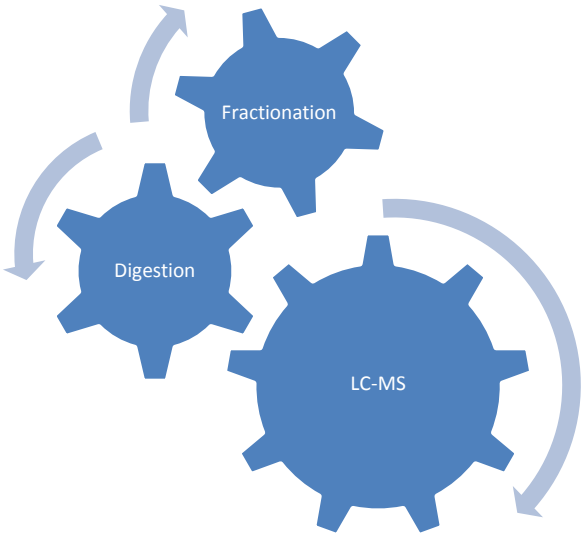


Fig. 11. The various screws that can be adjusted in the shotgun proteomic analysis of membrane proteins.

At the level of fractionation of the cell pellet, the membrane-rich fractions annotated as *organelles,* containing the inner membranes of mitochondria, golgi and endoplasmic reticula, and, *membrane*, comprising the plasma membrane, have been isolated. In contrast to heterogeneous samples such as the cytoplasmic fraction, these organelle and membrane fractions are expected to possess a higher concentration of membrane proteins, which would enable their more sensitive and accurate identification. Hence, the aim is to compare the

number of proteins identified from these fractions to the number captured by the standard protocol (about 4000 proteins) in terms of fractionation and digestion.

SDS gel electrophoresis was performed for the separation of the proteins in the sample fractions and the protein bands cut into equal gel pieces were subsequently digested. In-gel digestion was adopted as the method of choice because of its robustness, reproducibilty and its providing a higher coverage of hydrophobic peptides due to higher solubilization and its effect of reducing the complexity of the sample, which is analyzed by LC-MS, especially considering that the proteins of interest are extremely low-abundant.

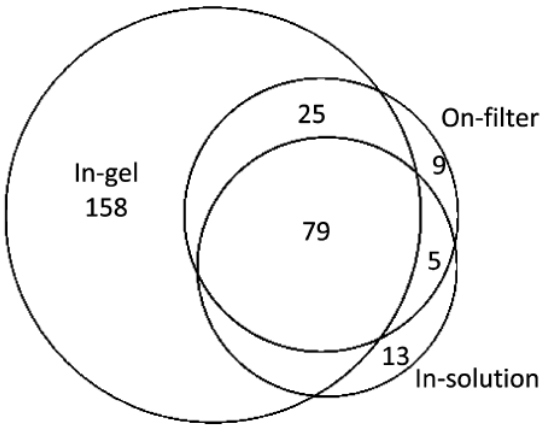


Fig. 12. Venn diagram showing overlap in proteins identified using in-gel, in-solution, and on-filter digestion[32].

As mentioned above, the number of tryptic amino acids especially in the hydrophobic, transmembrane domains of membrane proteins are low and lead to inefficient digestion and unsatisfactory identification. In view of this fact, chymotrypsin was used as a protease in the digestion protocol. Chymotrypsin preferentially cleaves peptide bonds where the carboxyl side of the amide bond is a large hydrophobic amino acid (tyrosine, tryptophan, and phenylalanine), as are common in the transmembrane domains of membrane proteins. Alternatively, digestion of the gel-separated proteins was performed using Trypsin dissolved in buffer containing 40 % acetonitrile to assist the solubility of the hydrophobic peptides in the digestion medium. An optimized, long (4h) gradient was employed for the chromatographic separation of the peptides. A Top 10 method was replaced by a Top 12 method for the analysis of the samples which are rich in membranes. Furthermore, repeated measurement of samples involving exclusion lists of peptides which were fragmented in the first measurement enhanced the scope of identified proteins.

2. Materials and Methods

2.1. The Standard Protocoll for Shot-Gun Proteomics

2.1.1 Cell fractionation

Jurkat cells are immortalized T lymphocytes, established from the peripheral blood of a 14-year-old boy with T cell leukemia. They are especially useful in studying immune response because of their ability to produce interleukin 2, as well as the response of cancer cells to various drugs. Jurkat cells are suspension cells that grow very fast and must be split every 2-3 days to prevent them from dying off due to the high cell density. The cells were grown in RPMI medium 1640 containing 10% fetal calf serum and Penicillin/Streptomycin.

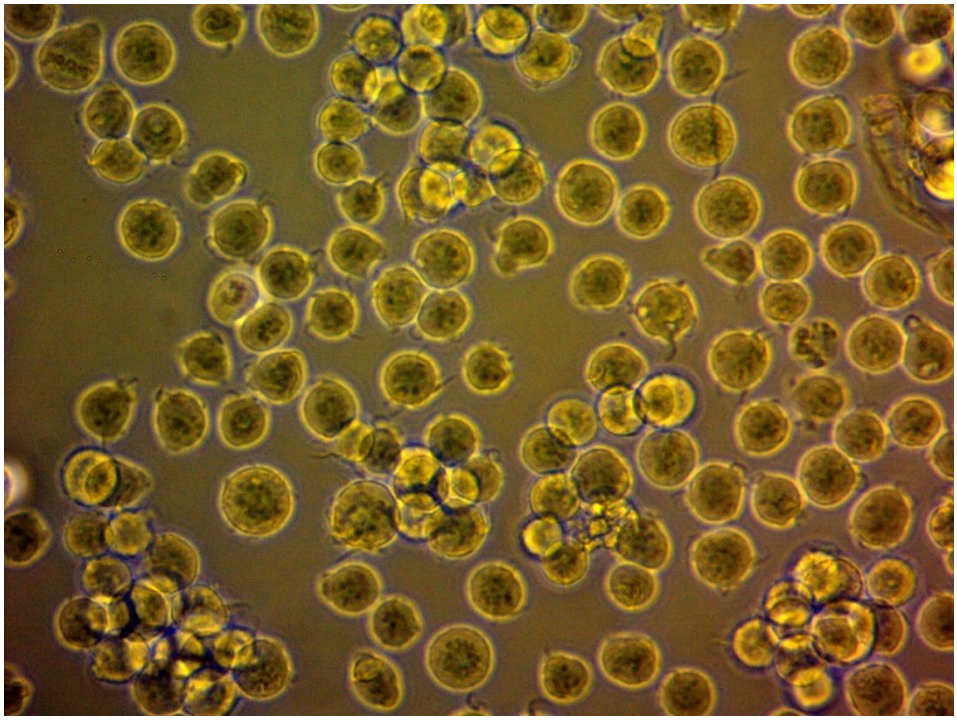


Fig. 13. A microscope photo of Jurkat cells taken with 100x magnification.

Chemicals

- Hepes/NaOH (pH 7,4) (Gerbu 1009.0250)
- NaCl (Merck / VWR ref. Nr 1.06404.1000)
- $MgCl_2$ (VWR 1.05833.0250)
- EGTA (VWR E4378)
- Sucrose (VWR 1.07687.1000)
- Triton X-100 (Roche 10.789.70.001)
- Pepstatin A (Sigma P4265)
- Leupeptin (Sigma L2884)
- Aprotinin (Sigma A1153)
- Urea (Gerbu 1044.1000)
- Thiourea (Sigma-Aldrich T8656)
- CHAPS (Gerbu 1083.0025)
- SDS (Gerbu 1012)
- DTT (Gerbu 1008.0025)
- Phenylmethanesulfonylfluoride (PMSF) (0,1 M in 2-propanol)
- NaCl (Merck / VWR 1.06404.1000)
- KCl (Merck / VWR 1.04936.0500)
- KH_2PO_4 (Merck / VWR 1.04873.0250)
- $Na_2HPO_4 . 2H_2O$ (Merck/VWR 1.06580.1000)

Solutions

Fractionation buffer

1 M Hepes/NaOH (pH 7,4) ad 10 mM

5 M NaCl ad 10 mM

1 M $MgCl_2$ ad 3,5 mM

2 M EGTA ad 1 mM

2 M Sucrose 0,25 mM

20 % Triton X-100 ad 0,5 %

filled up to 100 ml with H_2O, bidest.

Protease inhibitor cocktail (PIC)

100 µl Pepstatin A
10 µl Leupeptin
10 µl Aprotinin
filled up to 1 ml with H_2O, bidest.

PMSF

Phenylmethanesulfonylfluoride (0,1 M in 2-propanol)

Sample Buffer (SB)

4,5 g Urea ad 7,5 M
1,14 g Thiourea ad 1,5 M
0,4 g CHAPS ad 4 %
25 µl 20% SDS (w/w) 0,05 %
1 ml 1M DTT ad 0,1 M
filled up to 50 ml with H_2O, bidest.

Phosphate buffered saline (PBS)

8,2 g NaCl ad 137 mM
0,2 g KCl ad 2,7 mM
0,2 g KH_2PO_4 ad 1,5 mM
1,0 g $Na_2HPO_4 . 2H_2O$ ad 5,6 mM
0,2 g $NaH_2PO_4 . H_2O$ ad 1,1 mM
filled up to 1L with H_2O, bidest.
pH 7,3

TE-NaCl

1M Tris/HCl ad 10 mM
0,5 M EDTA ad 1 mM
5 M NaCl ad 0,5 mM
Fill up to 100 ml with H_2O, bidest.

TE-NP-40

1M Tris/HCl ad 10 mM

0,5 M EDTA ad 1 mM

10 % NP-40 ad 0,5 %

Filled up to 100 ml with H_2O, bidest

The Standard Fractionation Procedure

The suspension containing the Jurkat cells was transferred to a 10 ml falcon and centrifuged 1200 rpm for 5 min at 4 °C. The supernatant (containing the secretome) was discarded after which the cells were washed with PBS by resuspending the pellet and centrifuging as above. After a repeated washing step, the supernatant was discarded again and the pellet dried with a folded paper towel. On ice, 3 ml fractionation buffer with freshly added 30 µl PIC and 30 µl PMSF were put on the pellet, the cells lysed by pipetting up and down five times with a 1 mL pipette and centrifuged at 3500 rpm for 5 min at 4 °C. The supernatant containing the *cytoplasmic* fraction was transferred to a falcon and upon diluting with 12 ml cold ethanol to precipitate the contained protein, stored at – 20 °C overnight.

On the other hand, the pellet was suspended in 100 µl TE-NaCl and incubated on ice for 10 min. Further, 900 µl TE-NP-40 (containing freshly added 30 µl PIC and 30 µl PMSF) were added and the mixture incubated on ice for 15 min before centrifuging for 5 min at 3500 rpm and 4 °C. The supernatant consisting of the *nuclear extract* was transferred to a falcon, precipitated with ethanol by diluting 1:5 and stored at -20 C.

On the second day, the precipitated samples were centrifuged at 5000 rpm for 20 min at 4 °C, the supernatant decanted, the falcons were left to dry for 10 mins and degassed also for 10 mins in a desiccator under vacuum. To mildly denature and solubilize proteins, a tiny amount of urea was added to the sample as well as 50 µl of sample buffer and stored at 4 °C overnight. The next day, the clear solution was transferred to an Eppendorf tube and the protein concentration was measured by a Bradford assay (1 µl sample and 50 µl Bradford solution from Bio-Rad ref. 500-0203 in 199 µl H_2O).

2.1.2 SDS-PAGE

Various samples prepared by the fractionation procedure described above were separated by one-dimensional, vertical SDS-PAGE, loading a maximum of four samples on a gel. After polymerizaion, the gels were loaded and stained on the same day.

Chemicals

- Acrylamide (Gerbu 1001)
- Piperazine di-acrylamide (PDA) (1,4-Bis(acryloyl) piperazine) (Sigma-Aldrich 14470)
- SDS (sodium dodecyl sulfate) (Gerbu 1012)
- APS (ammonium persulfate) (VWR 1.01201.0500)
- TEMED (tetramethylethylenediamine) (Sigma T8133)
- Tris X (Gerbu 1018.1000)
- Glycine (Gerbu 1023.1000)
- Molecular mass marker: *Precision Plus* Protein Dual Color Standards (Bio-Rad 161-0374)
- Methanol, technical (VWR 20903.368) and methanol for analysis (Sigma 34966)
- Acetic acid for analysis (Merck / VWR ref. Nr. 1.00063.2511)
- Bromphenol blue (Merck / VWR ref. Nr. 1.11746.0005)
- β-mercaptoethanol (Sigma-Aldrich M3148)

Solutions

- 30% polyacrylamide (PAA) solution: 1L aqueous solution contains 292 g acrylamide and 8 g piperazine di-acrylamide (PDA) as cross-linker.
- 2 M Tris/HCl (pH 8,8) and 1 M Tris/HCl (pH 6,8)
- 20 % SDS (sodium dodecyl sulfate)
- 10 % APS (ammonium persulfate)
- TEMED (tetramethylethylenediamine)
- 10x Tris/Glycine: 1L aqueous solution contains 30 g Tris X and 144g Glycine
- Electrode buffer: 100 ml 10x Tris/Glycine and 5 ml 20 % SDS filled up to 1 L with H_2O
- Lämmli buffer (5x SDS - loading buffer):

  5 ml 1 M Tris / HCl pH 6,8 ad 300 mM
  2 g SDS ad 10 %
  10 ml Glycine 50 %
  0,05 g Bromphenol blue ad 0,05%
  H_2O ad 17,5 ml.
  62,5 µl β-Mercaptoethanol (12%) was added to 437,5 µl aliquots before use.

Instrumentation

Electrophoresis equipment: Mini-ProteanTM- Cell Tetra System from Bio-Rad, Austria

Desiccator from Kartell (VWR ref. Nr. 467-2120)

Vortex-Genie from Scientific Industries (VWR ref. Nr. 444-5900)

The separation gel (12 % acrylamide) has the following composition in 12 ml (for 2 gels):

4,80 ml 30 % AA / PDA
2,25 ml 2 M Tris / HCl pH 8,8
4,83 ml H_2O

After putting 2 ml of this mixture aside for the stopping gel, the remaining 10 ml for the separation gel are degassed under vacuum in a dessicator. Meanwhile, 20 µl 10 % APS and 5 µl TEMED are added to the 2 ml solution for the stopping gel and the mixture swiftly pipetted between the glass plates. After degassing, 50 µl 20 % SDS, 45 µl 10 % APS and 7,5 µl TEMED are added to the mixture for the separation gel, pipetted up to 2,5 cm below the top edge of the front plate, overlaid with 90 % isopropanol to form an even front and left to polymerize for 45 min.

The stacking gel (4 % acrylamide) has the following composition in 8 ml (2 gels):

1,066 ml 30 % AA / PDA
1,000 ml 1 M Tris / HCl pH 6,8
5,860 ml H_2O

After degassing for 10 minutes, 40 µl 20 % SDS, 40 µl 10 % APS and 8 µl TEMED were added to the mixture, the solution pipetted above the separation gel after removing the isopropanol with a filter paper, the comb placed between the plates and the stacking gel allowed to polymerize for 30 min.

Next, an volume of each sample corresponding to 20 µg of protein were mixed with 3 µl Lämmli buffer, filled up to 15 µl with water, vortexed and loaded onto the gel. In the slots between the samples, a 2:1 mixture of sample buffer and Lämmli buffer was loaded. For following the migration distance, 5 µl of a molecular mass marker were loaded in the last slot.

Table 1. Example of a loading scheme for an SDS-PAGE gel. Samples are in boldface.

Gel position	Sample no	Sample name	Protein concen-tration	Sample vol.	H_2O	Laemmli buffer
links 1		SB10µl		10µl		5µl
2	**1**	**jurkat_130116se_CYT**	**6,0µg/µl**	**3,3µl**	**8,7µl**	**3µl**
3		SB10µl		10µl		5µl
4	**2**					3µl
5		SB10µl		10µl		5µl
6	**3**	**jurkat_130116se_NUC**	**3,0µg/µl**	**6,7µl**	**5,3µl**	**3µl**
7		SB10µl		10µl		5µl
8	**4**					3µl
9		SB10µl		10µl		5µl
10	**M**	MW-Marker		5µl		

Electrophoresis was run at 40 mA for two gels with a 200 V limitation, until each marker band measured about 1,5 cm. Next, the gels were fixed with an aqueous solution of 50% methanol and 10% acetic acid for minimum 20 min and subsequently silver stained. Fixation denatures proteins by interfering with hydrogen bonds, upon which they precipitate in large insoluble aggregates within the gel matrix[33]. This prevents the diffusion of proteins, thus keeping the protein bands sharp and resolved during the staining process. In addition, fixation removes gel buffer components, most importantly SDS, which may interfere in the staining process and in the mass spectrometric analysis.

Chemicals for silver staining

- Methanol for analysis from Merck (Darmstadt, Germany) (VWR ref. Nr. 1.06009.2511)
- Sodium thiosulfate for analysis ($Na_2S_2O_3.5H_2O$) (Merck / VWR ref. Nr. 1.06516.0500)
- Silver nitrate ($AgNO_3$) (Sigma-Aldrich S6506)
- Sodium carbonate (Na_2CO_3) for analysis (Merck / VWR ref. Nr. 1.06392.1000)
- Acetic acid for analysis (Merck / VWR ref. Nr. 1.00063.2511)
- Formaldehyde 37% (Merck / VWR ref. Nr. 1.04003.1000)

Procedure for Silver Staining

The fixed gels were washed with 50 % methanol for 10 min, rinsed with water twice (each 5min), sensitized with 0,02 % $Na_2S_2O_3.5H_2O$, rinsed shortly, stained with 0,1 % $AgNO_3$ for 10 min, rinsed again and developed with 3 % Na_2CO_3 containing 0,05 % formaldehyde, then stopped with 1 % acetic acid.

Silver staining is in principle similar to developing a film. In fact, the original lithographic developer consisted of sulfite/bisulfite and formaldehyde. It is based on the specific chemical

reduction of silver ions (Ag^+), which are complexed by side chains and carboxyl groups of the amino acids, by formaldehyde to metallic silver. Thiosulfate ($H_2S_2O_3$) complexes non-specific bound silver on the gel surface, which is then washed away, eliminating the formation of a dark precipitate in the gel. Increasing the concentration of formaldehyde improves sensitivity but also background staining[34]. Temperatures above 25 °C also lead to a dark background.

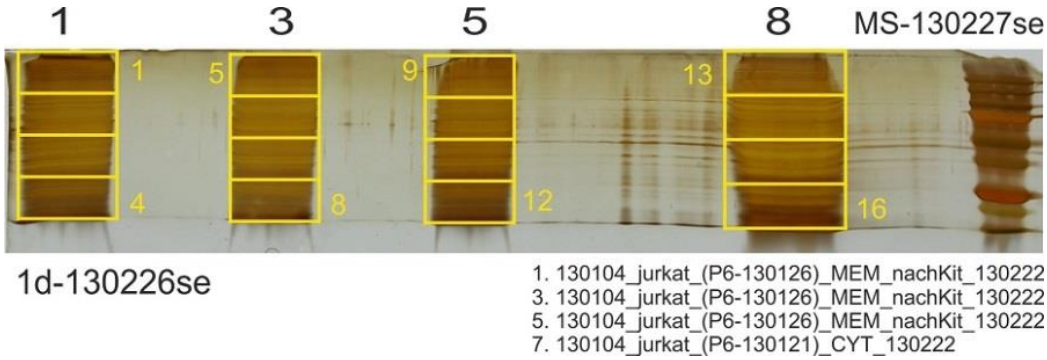


Fig. 14. An example of a stained SDS-PAGE gel with a superposed grid showing how the sample lanes were cut.

After documenting the gel with a photograph, the protein lane of each sample (1,5 cm long) was cut horizontally in four bands by means of a scalpel (Swann-Morton / VWR ref. Nr. 233-5482), which were further cut into approximately one mm^2 pieces to be prepared for in-gel digestion of the proteins comprised therein. The gel pieces cut out of a particular horizontal slice were placed in a single tube so that each sample was contained in four tubes.

2.1.3 Digestion

Prior to digestion of the proteins, the gel pieces have to be destained, washed and buffered up to the optimum pH range of the protease. Then, the sulfur bridges between the cysteine residues of proteins are reduced with dithiothreitol (DTT), disrupting the tertiary structure of proteins and making the internal domains accessible for digestion. To prevent intramolecular and intermolecular disulfide bonds from forming between the reduced sulfhydryl groups, they are alkylated using iodoacetamide (IAA), giving carbamidomethyl modified thiol groups. After washing off the unreacted IAA, the gel pieces are dehydrated with acetonitrile so that they can absorb the trypsin solution fully. Dried gel pieces can be stored several weeks or incubated with trypsin overnight.

Solutions

Destain solution

1 ml 150 mM $K_3[Fe(CN)_6]$ (Sigma-Aldrich P3667)

1 ml 500 mM $Na_2S_2O_3.5H_2O$ (Merck / VWR ref. Nr. 1.06516.0500)

8 ml H_20

Washing solution

50 ml Methanol (Merck / VWR ref. Nr. 1.06009.2511)

10 ml Acetic acid (Merck / VWR ref. Nr. 1.00063.2511)

40 ml H_2O

Buffer

0,197 g NH_4HCO_3 (Sigma-Aldrich A6141)

in 50 ml H_2O (50 mM)

DTT

50 µl 1M DTT (Gerbu 1008.0025)

in 5 ml 50 mM NH_4HCO_3 (10 mM DTT)

IAA

500 µl 500 mM IAA (Sigma-Aldrich I-6125)

in 4,5 ml 50 mM NH_4HCO_3 (50 mM IAA)

Digestion solution

Trypsin from Roche Diagnostics, Germany (Ref. 11418475001)

10 µl aliquot of a 0,1 µg/µl Trypsin solution in 1 mM HCl

mixed with 140 µl 50 mM NH_4HCO_3 buffer (6,6 ng/µl Trypsin)

Elution solution

50 ml Acetonitril HiPerSolv CHROMANORM® gradient grade for HPLC (VWR 20060.320)

5,5 ml Formic acid (Merck / VWR ref. Nr. 1.00253.1000)

44,5 ml H_2O

Instrumentation

Thermomixer Comfort from Eppendorf (VWR ref. Nr. 460-1112)

Ultrasonic bath Sonorex Digital from Bartelt (Gtraz, Austria, ref. Nr. 9877059)

Vacuum centrifuge GeneVac miVac from Bartelt (ref. Nr. DNA23050B00)

The standard digestion Protocol

At each step up to digestion, 200 µl of each solution was used. A freshly mixed destain solution was pipetted on each sample, vortexed 2 – 5 min until every gel piece was colorless and the solution taken up. With the washing solution, the samples were incubated twice for 5 min and twice for 10 min on the thermomixer at 800 rpm. Then, they were buffered for 5 min and thereafter reduced with the above solution of DTT, incubating for 30 min at 56 °C on the thermomixer. After draining and washing off the excess reagent with the buffer solution for 5 min, alkylation was carried out, for which incubation in the dark for 20 min is required. Subsequently the excess reagent was washed again with the buffer solution as above and the gel pieces dehydrated by incubating with acetonitril for 5 min and then vortexing until they were white. Excess acetonitril was drained off and the samples were placed on ice for digestion. A 10 µl aliquot of a 0,1 µg/µl trypsin solution was mixed with 140 µl of the 50 mM NH_4HCO_3 buffer on ice (or double the amount, if more than two samples were loaded on the gel). Of this digestion solution, 15 µl were pipetted in each tube and after 10 – 15 min, 25 µl of the 50 mM NH_4HCO_3 buffer were added. The samples were incubated with trypsin over night (about 15 hours) at 37 °C.

On the next day, another 40 µl of the NH_4HCO_3 buffer were added on the solution containing the peptides, the samples vortexed, sonicated for 15 min and centrifuged. The supernatant containing the peptides were taken up in siliconized tubes. The gel pieces were extracted two times with 40 µl of the extraction solution above (50 % ACN, 5 % FA), sonicating for 10 min, centrifuging and taking up the eluate in the same siliconized tubes as before. Eventually, the samples were evaporated to dryness at 40 °C on the Speed Vac concentrator (the pellet contains the peptides) and stored at - 20 °C. For LC-MS analysis, the samples were diluted to 60 µl by an aqueous solution with 2% ACN and 0.2% FA, then 5 µl of a solution of standard peptides (containing 20 fmol/µl of each peptide) was added in order to check the instrument performance during LC-MS analysis. After vortexing and centrifugation at 1000 rpm for 3 min, 60 µl of this solution was pipetted into a well plate.

2.1.4 HPLC-MS Analysis

First, the peptides were separated by nano-HPLC using the UltiMate 3000 RSLC nano System from Dionex (pre-column: Acclaim PepMap RSLC Nano Trap Column, C18, 100 µm x 2 cm, C18, 5µm, 100Å; analytical column: Acclaim PepMap RSLC, 75 µm x 50 cm, C18, 2µm, 100Å, Dionex, California, USA) operating at a flow rate of 300 nL/min at about 625 bar on the analytical column and about 80 bar on the trap column, with the column oven at 40 °C. A 270 min gradient from 99% A and 1% B to 80 % A and 20 %B was applied for the

analysis of all samples. Solvent composition A: 98% water, 2% ACN, 0,2% FA; Solvent composition B: 20% water, 80% ACN. For each run 5 µl of sample solution were injected.

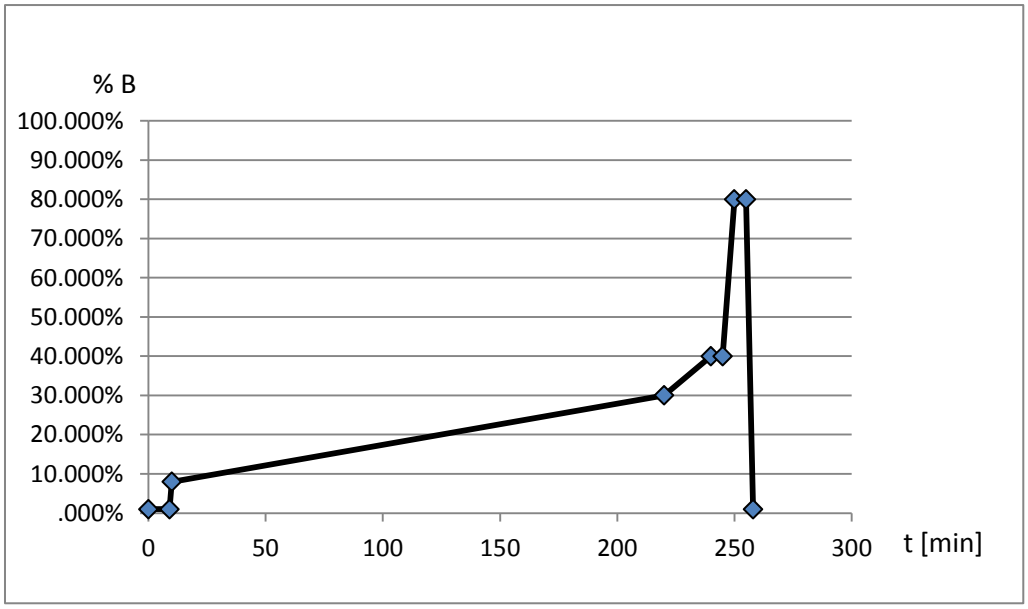


Fig. 15. Schematic representation of the elution gradient used for the chromatographic separation of the samples.

Peptide identification was done via MS/MS fragmentation analysis using a Q Exactive mass spectrometer from Thermo Scientific (Massachusetts, USA) equipped with a nanospray ion source. The MS data were acquired using a data-dependent “top 10 method”, which dynamically chooses the ten most abundant precursor ions from full scan MS (m/z 400-1400), followed by isolation of these within a window of 2 Dalton and fragmentation by “High-energy collision-induced dissociation” (HCD). The full scan MS was acquired at a resolution of 70.000 and the MS/MS scan at 17.500, both at m/z 200. Protein identification was achieved using Proteome Discoverer 1.3 (Thermo) employing the Mascot search algorithm and searching against Swissprot human database with following settings: maximum two missed cleavages and maximum allowed mass deviation of 5 ppm for peptide ions and 20 ppm for fragment ions. Further, carbamidomethylation of cysteines was set as fixed and protein N-terminal acetylation methionine oxidation as variable modification. On the search result the two peptides per protein filter was applied and the relaxed false discovery rate was set at 0,05 and the strict 0,01.

2.2. Shot-Gun Proteomics Experiments with a View to the Analysis of Membrane Proteins

Membrane proteins possess several hydrophobic domains in addition to hydrophilic ones and, as stated above, this is the basis of several problems in their analysis and identification. For this diploma project, several approaches have been implemented independently to overcome the difficulties in their isolation, purification, digestion and separation, starting from the isolation of fractions which are rich in membranes, through separation and digestion of these samples with a view to their relative hydrophobicity, up to their chromatographic analysis and mass spectrometric identification.

2.2.1 Fractionation Protocoll for Jurkat Organelles and Membranes

10x Phosphate buffered saline solution was warmed in a bath to ca. 37 °C and diluted to 1x. Two cell flasks were pooled into two 15 ml falcons and the flasks were washed with a little PBS. The falcons containing the Jurkat cell suspension were centrifuged at 12 000 rpm for 5 min and the supernatant containing the secretome was discarded. The pellet was dried genltly with a paper towel and washed twice by solving it in 12 ml PBS, centrifuging at 12 000 rpm for 5 min and discarding the supernatant.

Afterwards, the cells were lysed by a freeze – thaw cycle, in which the falcons containing the cell pellets were held in a N_2 bath for about 30 s and then into a water bath at 37 °C for about 1 min until the pellets were thawed. The cycle was performed three times. Next, 3 ml lysis buffer containing PIC and PMSF (the composition of the buffer is the same as above, except that it does not contain the surfactant Triton X) were pipetted on the pellet and the cells lysed by sucking and ejecting the solution against the wall of the falcon three times with the help of a syringe. Meanwhile, the state of the cells were monitored using the microscope. At this stage, nuclei were visible with some cytoskeleton around them, and organelles were floating in the matrix.

The pellets were centrifuged at 2000 rpm for 5 min and the supernatant was portioned into four 1,5 ml tubes evenly (two tubes per falcon). The pellet containing the nuclei was discarded. At this stage, no nuclei were visible in the supernatant in the microscope. Next, the tubes were centrifuged at 15 000 rpm for 15 mins and the supernatants containing the membranes were transferred with a Pasteur pipette into an ultracentrifuge tube and 10 ml PBS added. Then, the solution was ultracentrifuged at 100 000 g for 2h. Afterwards, the supernatant was decanted and the the tube dried gently. Sample buffer was added on the *membrane* pellet (30 µl) and left at 4°C for 24 hours before running a gel.

On the other hand, each pellet comprising the *organelle* fraction was solved in 300 µl lysis buffer, the whole 1,2 ml solution transferred into a 1,5 ml tube and centrifuged at 15 000 rpm for 15 min. The supernatant was decanted and the tube patted dry while keeping it upside down. 30 µl sample buffer was added on the organelle pellet and the tube was kept for 2 days at 4°C before running a gel.

Because of an absence of peptides in the mass spectrometric analysis, the fractionation was repeated with some differences. Four Jurkat cell pellets, containing in total about 2.10^7 cells, that had been washed with PBS, were pooled after solving all pellets in 1 ml lysis buffer containing 10 µl PIC and 10 µl PMSF. Next, the cells were disrupted by three freeze – thaw cycles as described above. After adding another ml lysis buffer, the cells were lyzed in three steps using a Dounce homogenizer, at each step knocking 10 times with the potter. Between each lysis, the state of the cells were monitored under the microscope. A fine matrix of organelles was visible, scattered around the much larger nuclei. The solution was centrifuged at 2000 rpm for 5 min after which nuclei were still present in the supernatant. Thereupon, 1 ml lysis buffer was added to the supernatant and the 3 ml solution centrifuged again at 3500 rpm for 5 min. The pellet containing the nuclei was worked up as described above for the standard protocol. The supernatant was pooled in two tubes and centrifuged at 20 000 g for 20 min. The pellet containing the *organelles* was washed with 100 µl lysis buffer and after drying, solved in 80 µl sample buffer adding some urea and deep freezed.

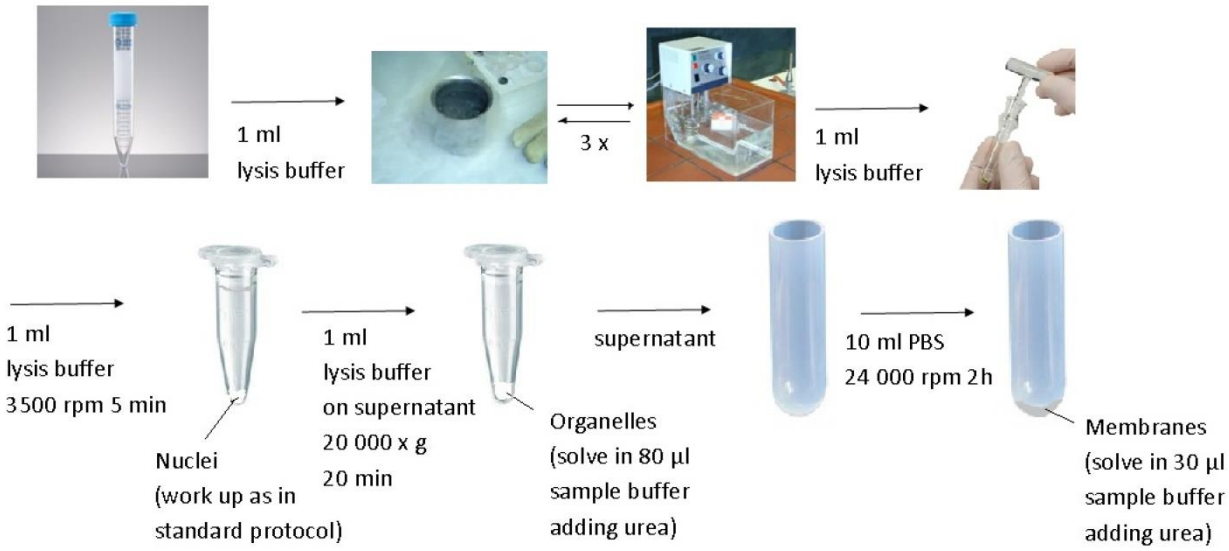


Fig. 16. Isolation of the fractions that are rich in membranes (organelles and membranes) from washed Jurkat cell pellets (represented by the falcon tube).

On the other hand, a *cytosolic* fraction was generated by diluting 0,5 ml of the supernatant of the last centrifugation (containing cytosolic as well as plasma membrane proteins) 1:1 with H_2O and precipitating overnight with 4 ml ethanol at – 20°C. The rest of the supernatant was diluted with 10 ml PBS and ultra-centrifuged at 24 000 rpm for 2 h. A second (main) *cytosolic* fraction was sampled by adding 12 ml ethanol to 3 ml of the supernatant and precipitating overnight at -20 °C. The *membrane* pellet was solved in 30 µl sample buffer adding some urea directly in the ultracentrifuge tube, from which the whole sample was loaded onto the gel after adding 5 µl H_2O and 8 µl Lämmli buffer.

On the next day, the three ethanol precipitated samples (two cytosolic fractions and the nuclei) were centrifuged at 5000 rpm for 20 min at 4 °C, the supernatant decanted, the falcons were left to dry for 10 mins and degassed also for 10 mins in a dessicator under vacuum. To mildly denature and solubilize proteins, a tiny amount of urea was added to the sample as well as 50 µl (cytosol) or 30 µl (nuclei) of sample buffer and stored at 4 °C overnight. The next day, the clear solution was transferred to an Eppendorf tube and the protein concentration was measured by a Bradford assay.

2.2.2 Fractionation using a Kit

Membrane and cytoplasm fractions were isolated from Jurkat cells by using a Thermo Scientific Mem-PER Membrane Protein Extraction Kit (THP Medical Products, GmbH, Wien, Ref. 89842). The enrichment of the proteins in both fractions was performed according to the procedure accompanying the kit for suspension mammalian cells[35]. In this protocol, the cells are permeabilized with a mild detergent to allow the release of soluble cytosolic proteins, after which a second detergent then solubilizes membrane proteins.

First, the cells are harvested by centrifuging for 5 min at 300 g, washed with 3 ml of Cell Wash Solution and centrifuged at 300 × g for 5 minutes. The supernatant is discarded, the cells resuspended in 1,5 ml Cell Wash Solution and transferred to a 2mL centrifuge tube, after which they are centrifuged again as above. The supernatant is discarded and 0.75mL of *Permeabilization Buffer* is added to the cell pellet. Following incubation for 10 min at 4 °C, the permeabilized cells are centrifuged for 15 min at 16 000 g upon which the supernatant contains cytosolic proteins and is transferred to a new tube. The pellet is resuspended by adding 0.5mL of *Solubilization Buffer* and pipetting up and down and incubated at 4°C for 30 minutes. Membrane proteins are thus solubilized and isolated in the supernatant by centrifuging at 16000 g for 15 min at 4 °C.

Table 2. Gel loading scheme of membrane and cytoplasmic samples isolated by the Kit. The membrane fraction was loaded three times and digested by three different methods.

sample no	sample	protein amount	sample	H_2O	5x SDS-PP	protease
	130104_jurkat_MEM_nachKit	**45µg**	**10µl**	**2.0µl**	**5µl**	**Typsin (+40 % ACN)**
1	SB10µl		10µl		3µl	
	130104_jurkat_MEM_nachKit	**45µg**	**10µl**	**2.0µl**	**5µl**	**Chymotrypsin**
2	SB10µl		10µl		3µl	
	130104_jurkat_MEM_nachKit	**45µg**	**10µl**	**2.0µl**	**5µl**	**Trypsin**
3	SB10µl		10µl		3µl	
	SB10µl		10µl		5µl	
4	**130104_jurkat_CYT**	**8.0µg/µl**	**2.5µl**	**9.5µl**	**3µl**	**Trypsin**
	SB10µl		10µl		5µl	
M	MW-Marker		10µl			

2.2.3 Separation and Digestion

Chemicals

Trypsin Sequencing Grade from Roche Diagnostics, Germany (Ref. 11418475001)

Chymotrypsin Protease (TLCK treated), MS Grade from THP Medical Products Vertriebs GmbH (Wien) (Ref. 90056)

10 µl aliquot of a 0,1 µg/µl Trypsin solution in 1 mM HCl
mixed with 140 µl 50 mM NH_4HCO_3 buffer (6,6 ng/µl Trypsin)

10 µl aliquot of a 0,1 µg/µl Trypsin solution in 1 mM HCl
mixed with 140 µl 50 mM NH_4HCO_3 buffer containing 40 % ACN (6,6 ng/µl Trypsin)

10 µl aliquot of a 0,1 µg/µl Trypsin solution in 1 mM HCl
mixed with 70 µl 50 mM NH_4HCO_3 buffer containing 40 % ACN (12,5 ng/µl Trypsin)

10 µl aliquot of a 0,1 µg/µl Chymotrypsin solution in 1 mM HCl
mixed with 140 µl 50 mM NH_4HCO_3 buffer containing 2 mM $CaCl_2$ (6,6 ng/µl Chymotrypsin)

Protocoll

SDS gel electrophoresis was performed as above, loading a volume of *cytoplasm*, *organelles,* or *nuclei* that contained 20 µg protein (e.g. 4 µl of a sample that had a protein concentation of 5 µg/µl). All of the *membrane* fraction prepared by the Mem-PER Membrane Protein Extraction Kit containing 45 µg protein in 30 µl was loaded in three portions of 10 µl (see Table 2). Also the whole *membrane* sample in the ultracentrifuge tube was loaded after adding 5 µl H_2O and 8 µl Lämmli buffer. The gels were fixed and stained, the cut gel pieces destained, washed, buffered, reduced and alkylated as before. After dehydrating the gel pieces, trypsin in NH_4HCO_3 buffer containing 40 % ACN was used for digestion of the membrane samples prepared by the Mem-PER Membrane Protein Extraction Kit. By using ACN, increasing the solubility and accesibility of the proteins in the digestion medium as well as the solubility of the digestion products, which are partially hydrophobic peptides, was aimed.

Table 3. Gel loading scheme of membrane and organelle fractions isolated after cell disruption by a freeze-cycle. The organelle fraction was loaded twice and digested by trypsin or chymotrypsin.

sample no	sample	protein conc	sample	H_2O	5x SDS-PP	protease
	SB10µl		10µl		5µl	
1	**Jurkat_130607se-cg_ORG**	**3.0µg/µl**	**7µl**	**5µl**	**3µl**	**Typsin**
	SB10µl		10µl		5µl	
2	**Jurkat_130607se-cg_ORG**	**3.0µg/µl**	**7µl**	**5µl**	**3µl**	**Chymotrypsin**
	SB10µl		10µl		5µl	
	SB10µl		10µl		5µl	
3	**Jurkat_130607se-cg_MEM**	**2.0µg/µl**	**10.0µl**	**5µl**	**8µl**	**Trypsin**
	SB10µl		10µl		5µl	
M	MW-Marker		10µl			

With a view to be able to digest hydrophobic domains more efficiently, chymotrypsin was used as a second approach to digest the samples. Because trypsin cleaves proteins at the charged, and thus hydrophilic amino acids lysine and arginine, membrane proteins are mostly captured by their hydrophilic domains. Chymotrypsin, on the other hand, cleaves at hydophobic amino acids (phenylalanine, tyrosine, tryptophan) and should avail identification by transmembrane domains. Chymotrypsin is added in NH_4HCO_3 buffer containing 2 mM $CaCl_2$ to enhance its activity. Digestion with chymotrypsin was applied to the organelle membrane and plasma membrane fractions isolated as described above as well as to the membrane sample prepared by the Mem-PER Membrane Protein Extraction Kit.

Because a significant number of larger peptides were observed in the MS analysis of the already measured samples, the concentration of trypsin in the digestion buffer was doubled by mixing the trypsin aliquot with half the volume of buffer. Organelle, membrane, cytosol and nuclear extract samples were digested in this way. Elution of the peptides on the next day as well as evaporation on the Speed Vac were performed as in the standard protocoll above.

2.2.4 HPLC-MS Analysis

Organelle and membrane samples were analyzed twice, the second analysis involving an exclusion list with the peptides already selected for fragmentation in the first analysis. A top 12 method was employed for the mass spectrometric analysis of these samples with the settings mentioned above and the raw files searched using the Proteome Discoverer Software from Thermo Scientific against the Swissprot human database using the Mascot algorithm with the method used for the other samples. The identified proteins were filtered with the two peptides per protein filter and the false discovery rates set at 0,05 (relaxed) and 0,01 (strict).

The resulting proteins were compared with reviewed human proteins on the Uniprot Database (available at http://uniprot.org) by using the compare tool on the GPDE database (available at http://dbserver.proteomics.local/gpde). The membrane proteins in each sample were established by comparing a list identified proteins from the search with the raw data with the proteins on the Uniprot Database which are annotated with certain gene ontologies. Gene ontology (GO) is a bioinformatics project consisting of defined terms annotating gene products as to their *molecular function* (e.g. receptor activity), the *biological process* they participate in (e.g. cell signalling) or the *cell compartment* where they are located in (e.g. membrane). In other words, it is a set of defining characteristics, which is species-neutral and which avails a unified vocabulary in describing the localization and activity of gene products as well as their relationships. Each GO term has a unique numerical identifier, belongs to one of the three cited domains and is related to other GO terms in a definite manner. The ontologies of GO are structured as a direct acyclic graph.

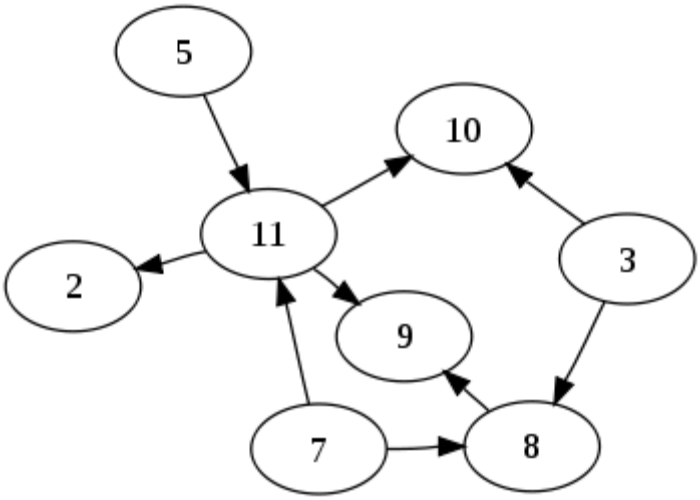


Fig. 17. An example of a directed acyclic graph[36].

Just as each term is exactly defined, the relations between GO terms are also categorized and defined. Some examples are: *Part of*, *is a*, *regulates*, *negatively regulates* and *positively regulates*.

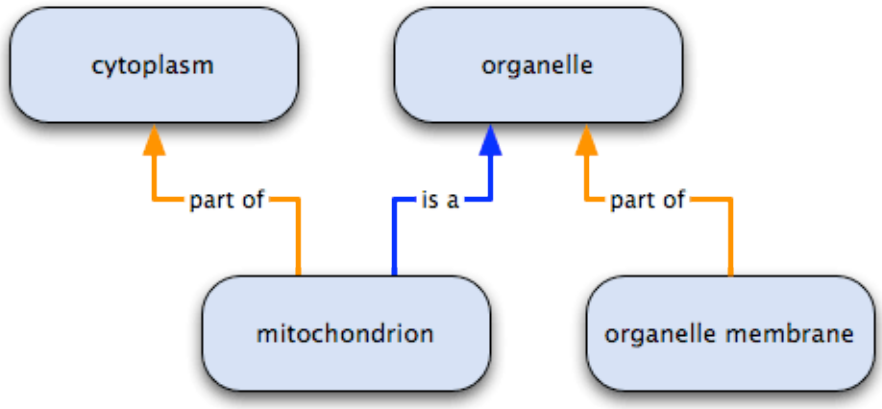


Fig. 18. A node (GO term) may have a "*part of*" relationship to one node, and an "*is a*" relationship to another.

For instance, the number of integral membrane proteins from each experiment were determined by comparing the accession numbers of the search results to those of the proteins on the Uniprot Database with the GO annotations regarding the cell compartment *integral to membrane*, *intrinsic to plasma membrane* and *intrinsic to organelle membrane*. Intrinsic to membrane (GO: 0031224) refers to proteins that have some covalently attached moiety embedded in the membrane, and is further split into integral to membrane (GO: 0016021) and anchored to membrane (GO:0031225). The former refers to proteins in which

some part of the peptide sequence spans all or part of the membrane; the latter refers to proteins tethered to a membrane by a covalently attached anchor, such as a lipid moiety. Each of these terms can have *child terms* referring to specific membranes, for example integral to plasma membrane (GO: 0031226) or extrinsic to vacuolar membrane (GO: 0000306)[37].

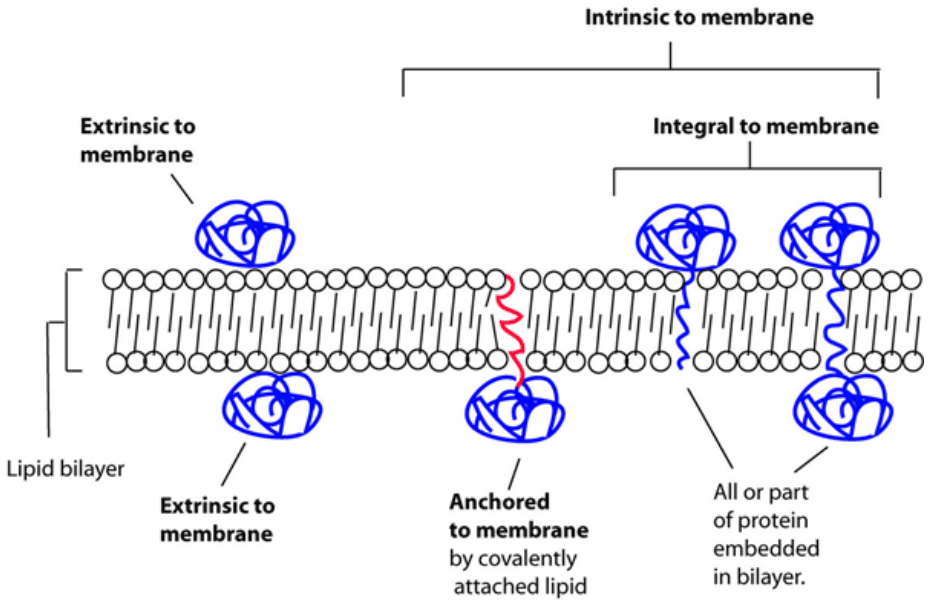


Fig. 19. Whereas *integral to membrane* refers to proteins in which some part of the peptide sequence threads through the membrane, *intrinsic to membrane* also comprises those proteins that are attached to the membrane by a covalently bound moiety.

When comparing with the human proteome with a certain GO annotation in the Uniprot database, only the “reviewed” proteins (corresponding to SwissProt entries comprising the whole human proteome, i.e. 20264 proteins in total) were considered. Various experimental conditions were thus compared with respect to the number of membrane proteins covered.

Considering the fact that most integral membrane proteins are relatively hydrophobic, the grand average of hydrophobicity (GRAVY) scores of the identified proteins as well as of their peptides were calculated and the number of proteins or peptides plotted against the GRAVY score . With the performed experiments that aim to amend the problems which are common in the analysis of membrane proteins on the one hand, and involve the enrichment of membrane proteins in cell fractions such organelles and plasma membrane on the other, a greater number of hydrophobic peptides and proteins are expected to be covered in comparison to the standard protocol employed in the analysis of cytoplasmic proteins.

3. Results and Discussion

3.1 Membrane Proteins in Various Fractions of Jurkat Cells

About 80 % of the identified proteins in all samples had more than 10 % sequence coverage. 85 – 87 % of the proteins were identified by at least 3 peptides and 80 - 84 % by at least 3 *unique* peptides.

130610se_MEM.msf*

Proteins | Peptides | Search Input | Result Filters | Peptide Confidence | Search Summary

	Accession	Description	Score	Coverage	# Proteins	# UniquePeptides	# Peptides	# PSMs	# AAs	MW [kDa]	calc. pI
1	Q13740	CD166 antigen OS=Homo sapiens GN=ALCAM PE=1 SV=...	1310.53	42.88 %	1	19	19	45	583	65.1	6.25

	A2	Sequence	# PSMs	# Proteins	# Protein Groups	Protein Group Accessions	Modifications	ΔCn	IonScore	Exp Value	Charge	MH+ [Da]	ΔM [ppm]	RT [min]	# Missed Cleavages
1		FVcMLVTEDNVFEAPTMK	2	1	1	Q13740	C3(Carbamidomethyl)	0.0000	90	5.7E-009	2	2212.10283	-0.06	217.79	0
2		SmIASTAITVHYLDLSLNPS...	1	1	1	Q13740	M2(Oxidation)	0.0000	90	4.9E-009	3	2691.37168	2.14	180.31	0
3		SMIASTAITVHYLDLSLNPSG...	1	1	1	Q13740		0.0000	82	2.8E-008	3	2675.38401	4.86	201.99	0
4		SSPSFSSLHYQDAGNYVcET...	2	1	1	Q13740	C18(Carbamidomethyl)	0.0000	61	2.3E-006	3	3216.48447	1.67	205.51	0
5		QIGDALPVScTISASR	1	1	1	Q13740	C10(Carbamidomethyl)	0.0000	61	4.0E-006	2	1674.84734	-0.41	105.93	0
6		ADIQMPFTcSVTYYGPSGQK	3	1	1	Q13740	C9(Carbamidomethyl)	0.0000	58	3.9E-006	2	2250.01592	-2.10	163.09	0
7		VLHPLEGAVVIIPKK	4	1	1	Q13740		0.0000	53	5.6E-006	3	1663.03232	1.03	155.05	1
8		VLHPLEGAVVIIPK	4	1	1	Q13740		0.0000	50	1.1E-005	3	1534.93705	0.92	186.49	0
9		VFKQPSKPEIVSK	4	1	1	Q13740		0.0000	49	2.9E-005	2	1486.85942	-2.31	14.24	2
10		SVQYDDVPEYKDR	3	1	1	Q13740		0.0000	49	4.4E-005	2	1613.74016	-2.53	34.98	1
11		EMDPVTQLYTMTSTLEYK	1	1	1	Q13740		0.0000	48	5.1E-005	2	2150.00444	0.55	225.17	0
12		TTKADIQMPFTcSVTYYGPS...	2	1	1	Q13740	C12(Carbamidomethyl)	0.0000	43	2.0E-004	3	2580.21292	0.76	134.47	1
13		ESLTLIVEGKPQIK	4	1	1	Q13740		0.0000	41	1.1E-004	2	1554.91057	0.23	98.58	1
14		LDVPQNLMFGKWK	1	1	1	Q13740		0.0000	39	7.6E-004	3	1575.83512	-0.10	159.39	1
15		ALFLETEQLKK	4	1	1	Q13740		0.0000	36	6.1E-004	2	1319.75725	0.20	79.35	1
16		ALFLETEQLK	2	1	1	Q13740		0.0000	33	2.5E-003	2	1191.66228	0.22	120.42	0
17		LDVPQNLMFGK	2	1	1	Q13740		0.0000	30	5.7E-003	2	1261.66106	0.05	146.79	0
18		TIHSEQAVFDIYYPTEQVTIQ...	1	1	1	Q13740		0.0000	29	4.0E-003	3	3016.56150	-1.73	204.97	0
19		YEKPDGSPVFIAFR	2	1	1	Q13740		0.0000	27	1.3E-002	3	1625.83335	0.65	119.36	1
20		RFVcMLVTEDNVFEAPTMK	1	1	1	Q13740	C4(Carbamidomethyl)	0.0000	21	4.1E-002	3	2368.20700	1.24	191.05	1

Fig. 20. Search result showing the CD166 antigen identified by 19 unique peptides with 43% coverage and a score of 1310.

In the cytoplasmic sample, which was prepared and analyzed by the standard protocol described above, a relatively high amount of membrane proteins were identified. With the various protocols that have been tested as described above, the number of different identified membrane proteins has been remarkably increased. The cell pellets that were disrupted by the freeze-thaw cycle with liquid nitrogen and consisted about 2.10^7 cells yielded 1,2 mg of solubilized protein in total (of which 480 µg were contained in the organelles and 60 µg in the membrane pellet from ultracentrifugation). Because the designation of identified proteins in the sample as *membrane proteins* is based on a comparison between these and the proteins on the Uniprot database which are annotated with a certain gene ontology, the number of membrane proteins in a specific cell fraction observed with a particular digestion and analysis protocol depends on the gene ontology selected.

A comparison of the total number of proteins in various samples shows that the fractionation protocol applied for the isolation of the samples *organelle, membrane, nuclei and cytosole* (shown in red), involving cell disruption with a freeze-thaw cycle and ultracentrifugation, enabled the identification of up to twice as many proteins as compared to the standard protocol with the samples *cytoplasm* and *nuclei* (shown in blue).

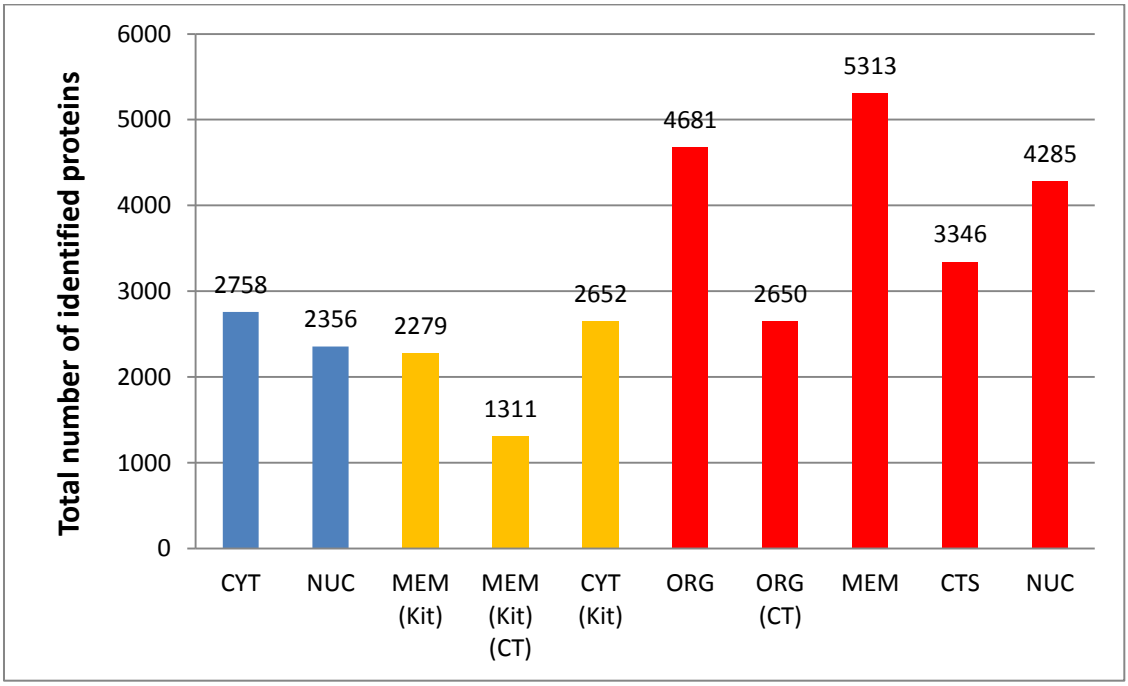


Fig. 21. Number of identified proteins in analyzed samples. Blue: cytoplasm and nuclei analysed by the standard protocol; Yellow: membrane and cytoplasmic samples isolated by the Protein Extraction Kit; Red: the membrane-enriched samples organelles, membranes, cytosol and nuclei. All samples were digested with trypsin unless otherwise stated (CT: chymotrypsin).

The *total* amount of all identified proteins in the cytoplasm analyzed by the standard protocol number 2758, which slightly exceeds the total number in the samples prepared by the Protein Extraction Kit (2652 in the cytoplasm and 2279 in the membranes). However, that particularly the amount of identified *membrane proteins* in the samples isolated by the Kit surmounts the number in the cytoplasm as stated above demonstrates that these proteins of interest were specifically enriched by the Kit, even if at the expense of cytoplasmic proteins. As an example, the cytoplasmic fraction analysed by the standard protocol was found to contain 838 proteins with the GO annotation *membrane*, 517 proteins with the GO annotation *plasma membrane* and 622 proteins with the GO annotation *organelle membrane*. Although the Protein Extraction Kit yielded a less number of total protein, it led to a comparable amount of proteins having these GO annotations (840, 474 and 599 respectively).

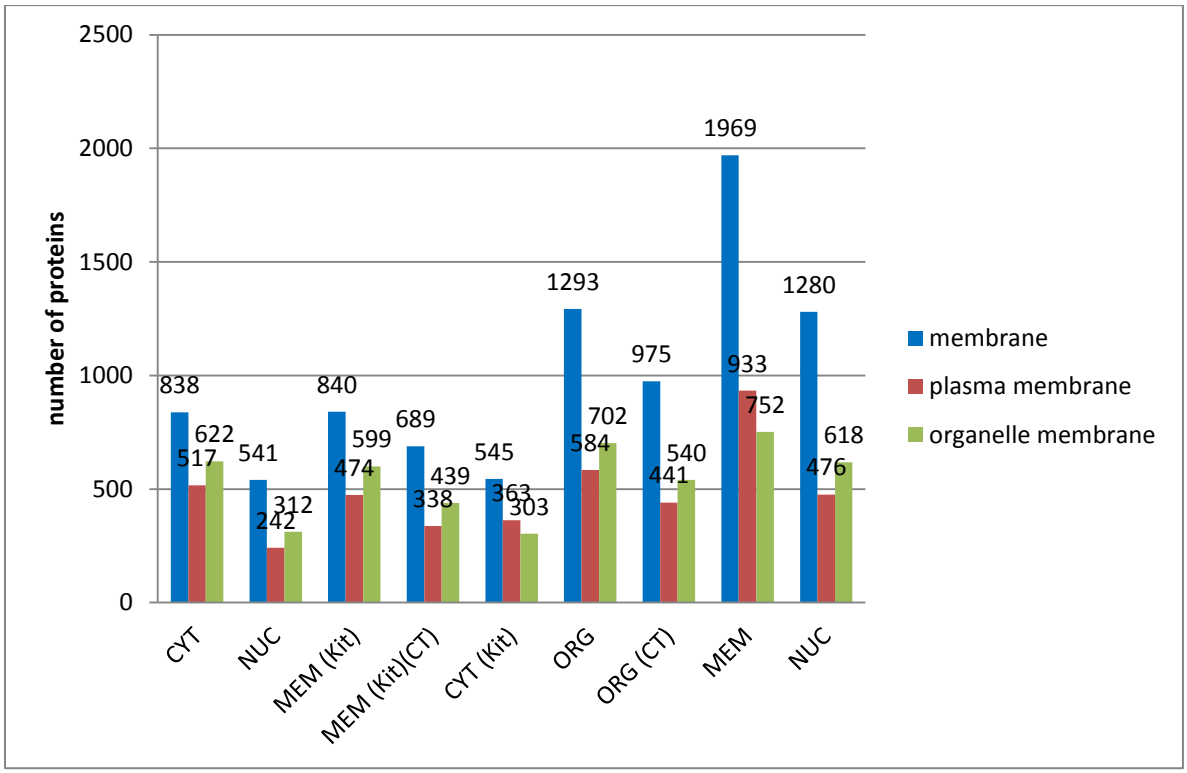


Fig. 22. Number of identified proteins with the GO annotations (cell compartment) *membrane, plasma membrane and organelle membrane.*

In terms of membrane proteins, the samples *organelles, membranes and nuclear extract* (N_2 disruption) again enabled twice as many proteins to be identified as in the cytoplasm. As many as 1969 membrane proteins were identified in the *membrane* prepared by ultracentrifugation and digested with trypsin. This number corresponds to 24 % of all proteins with this annotation. Membrane proteins that have a vital role in cell cycle and homeostasis are among the identified proteins: various chains of the T-cell surface glycoprotein CD3 (T-cell receptor T3), the subunit 6B1 of Cytochrome c oxidase, the various subunits of Sodium/potassium-transporting ATPase, CD59 glycoprotein (Membrane attack complex inhibition factor - MACIF) , Nuclear factor of activated T-cells, cytoplasmic 2 (NF-ATc2), subunits of Calcium/calmodulin-dependent protein kinase type II, various members of the Tumor necrosis factor receptor superfamily, Integrin beta-7, G protein-coupled receptor kinases 5 and 6, Interleukin-1 receptor-associated kinase, several types of the receptor-type tyrosine-protein phosphatase, 1-phosphatidylinositol 3-phosphate 5-kinase, Integral membrane protein 2B (Transmembrane protein BRI), Probable phospholipid-transporting ATPase IF (with 10 transmembrane regions), etc.

A significant increase in the identification of proteins with the GO annotations *plasma membrane* and *organelle membrane* was also achieved with this protocol. Especially with the

membrane fraction, a 20 % coverage of the plasma membrane proteins and 30 % of the organelle membrane proteins were realized.

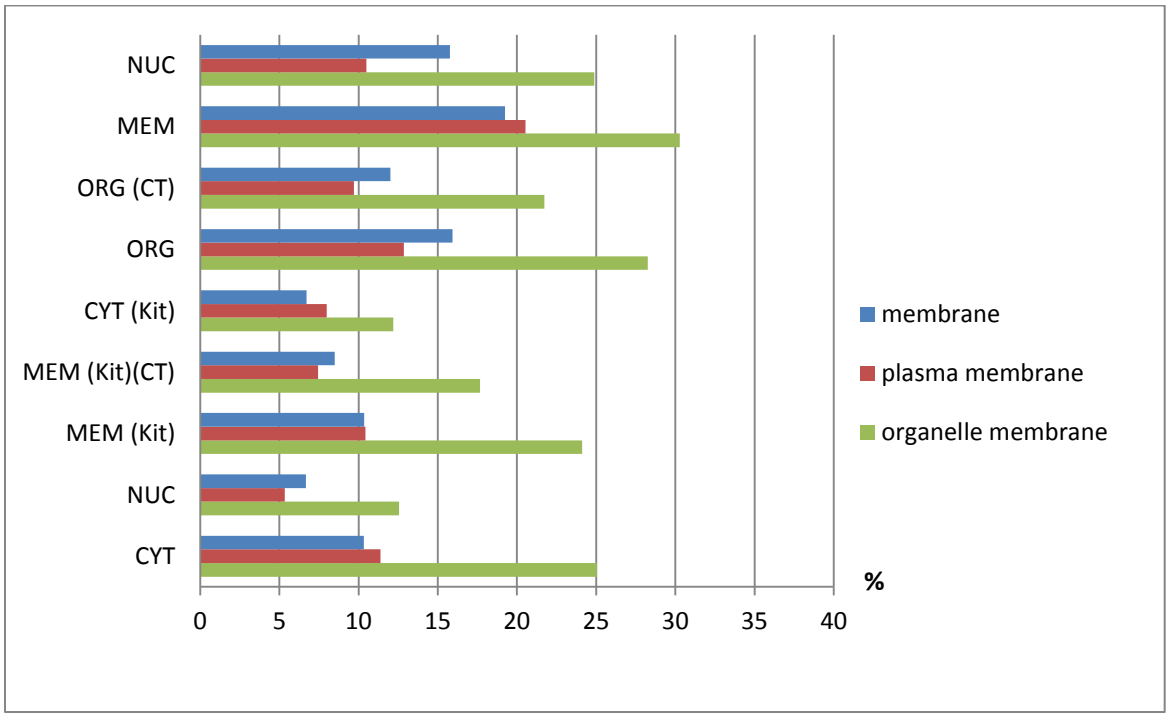


Fig. 23. Coverage of proteins with the above annotations in various samples in percentage.

Digestion and elution of the samples fractionated using the Protein Extraction Kit with 20 % ACN revealed that this method led to a loss of proteins and so it was not further investigated. Digestion of the *membrane* sample produced using the Protein Extraction Kit with chymotrypsin resulted in a small amount of proteins identified (689 with the GO annotation *membrane* and 244 with the annotation *integral to membrane*). Also the organelle fraction from the cell disruption with liquid nitrogen was digested with both proteases, which revealed that a relatively low amount of proteins were identified through digestion with chymotrypsin.

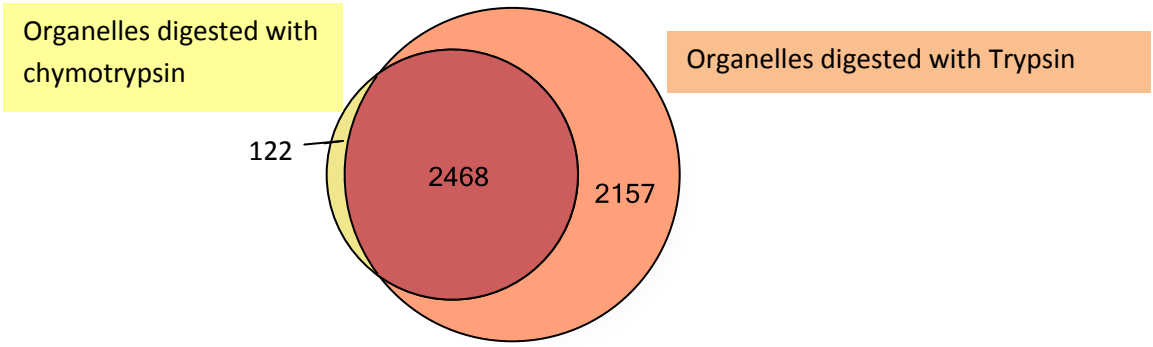


Fig. 24. Comparison of the number of identified proteins in the *organelles* digested with trypsin and chymotrypsin. Most proteins identified in the chymotrypsin-digested sample were also identified upon trypsin digestion.

Hence trypsin was used as the protease of choice for all following samples. In addition, gel pieces which were digested with trypsin and stored at 4°C were subjected to digestion with chymotrypsin, however the analysed samples contained few to no protein at all. This also showed that trypsin actually cleaved all the proteins present in the sample and there were almost no proteins that were cleaved by chymotrypsin exclusively.

The proteins with the annotation *membrane* do not strictly include integral membrane proteins but also peripheral membrane proteins as well as those attached to the membrane, for example by a glycolipid anchor or by interaction with another protein. Integral membrane proteins, which are the main focus of this work, were identified by comparing the list of proteins identified in each sample with the proteins on the Uniprot database annotated as *integral to membrane, intrinsic to plasma membrane* or *intrinsic to organelle membrane*. The number of proteins with these GO annotations in the cytoplasmic fraction were 253, 52 and 33 respectively. This corresponds roughly to 4%, 4% and 11% of all the proteins in these cell compartments.

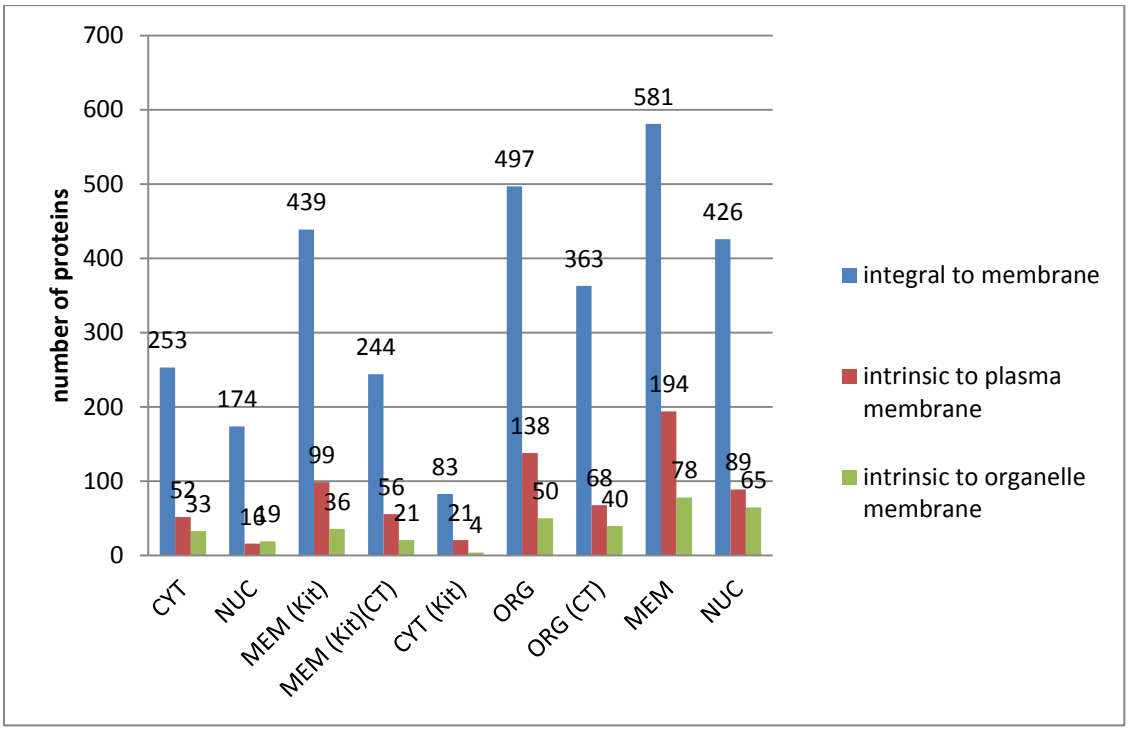


Fig. 25. Number of identified proteins with the GO annotations (cell compartment) integral to membrane, intrinsic to plasma membrane and intrinsic to organelle membrane.

In contrast, the number *of integral membrane proteins* in the trypsin-digested membrane fraction isolated using the Mem-PER Membrane Protein Extraction Kit amount to 439, whereas those in the organelle fraction isolated from cells disrupted by liquid nitrogen count 498, and in the membrane as many as 581 (10 % coverage of all proteins with this annotation). These include several transmembrane proteins such as the Thioredoxin-related transmembrane protein 1 and 2, Transmembrane emp24 domain-containing proteins 1-5, 7, 9 and 10, the Interferon-induced transmembrane protein 1 as well as 42 proteins with the phrase *Transmembrane protein* in the description such as the *Transmembrane protein 230.*

The improvement in the coverage of proteins with the highly specific GO annotation *intrinsic to plasma membrane* is also significant: Whereas 56 proteins had this annotation in the cytoplasmic fraction analyzed by the standard protocol, 99 proteins were found in the membrane fraction isolated using the Protein Extraction Kit, 138 in the organelles and 194 in the membranes (all digested with trypsin). For the membrane sample, this corresponds to 16 % of all proteins with this annotation. Moreover, 27 % of all proteins with the annotation *intrinsic to organelle membrane* were identified in the membrane sample and 17 % in the organelles.

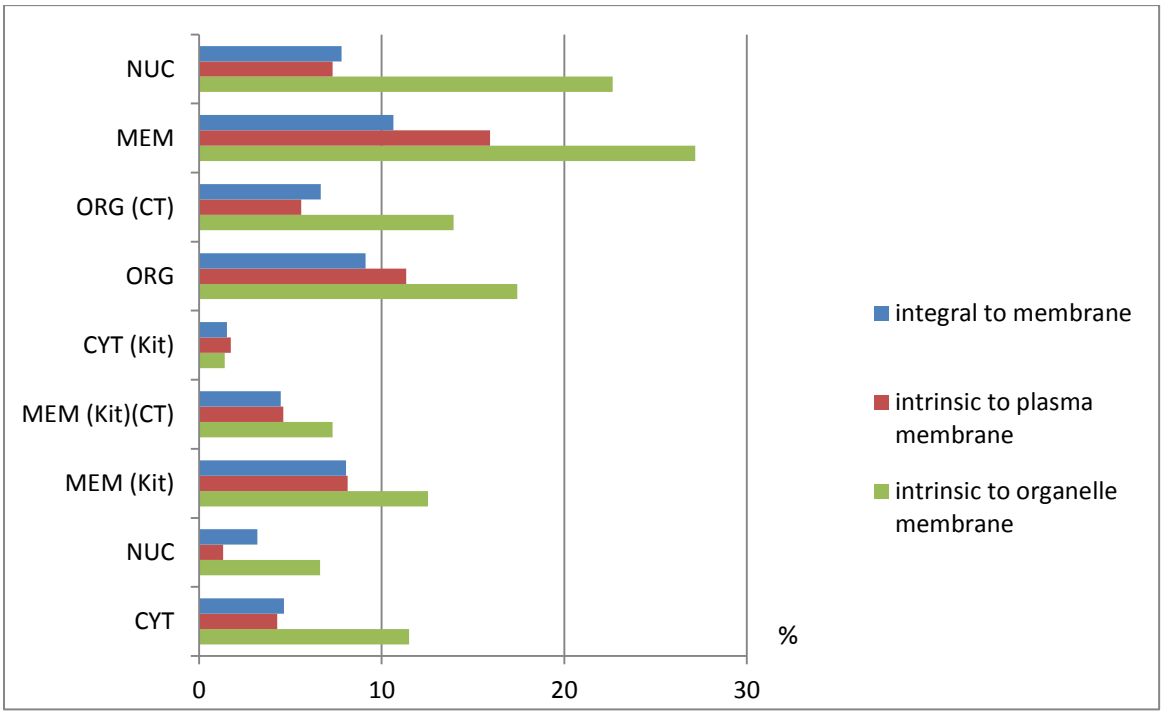


Fig. 26. Coverage of proteins with the above annotations in various samples as a percentage.

Beside several *single pass* membrane proteins, many integral transmembrane proteins are *multipass* MPs. Because they span the plasma membrane several times, they contain highly hydrophobic regions which thread through the membrane and constitute an appropriate benchmark for the capacity of the analysis to capture hydrophobic peptides and proteins. These contain among others the ATP binding cassettes, GPCRs, various transporters, ATPases and tumor suppressor proteins. In the ultracentrifuge membranes, a large number of multipass transmembrane proteins were identified (324), in contrast to only 62 proteins in the cytoplasmic sample of the standard protocol.

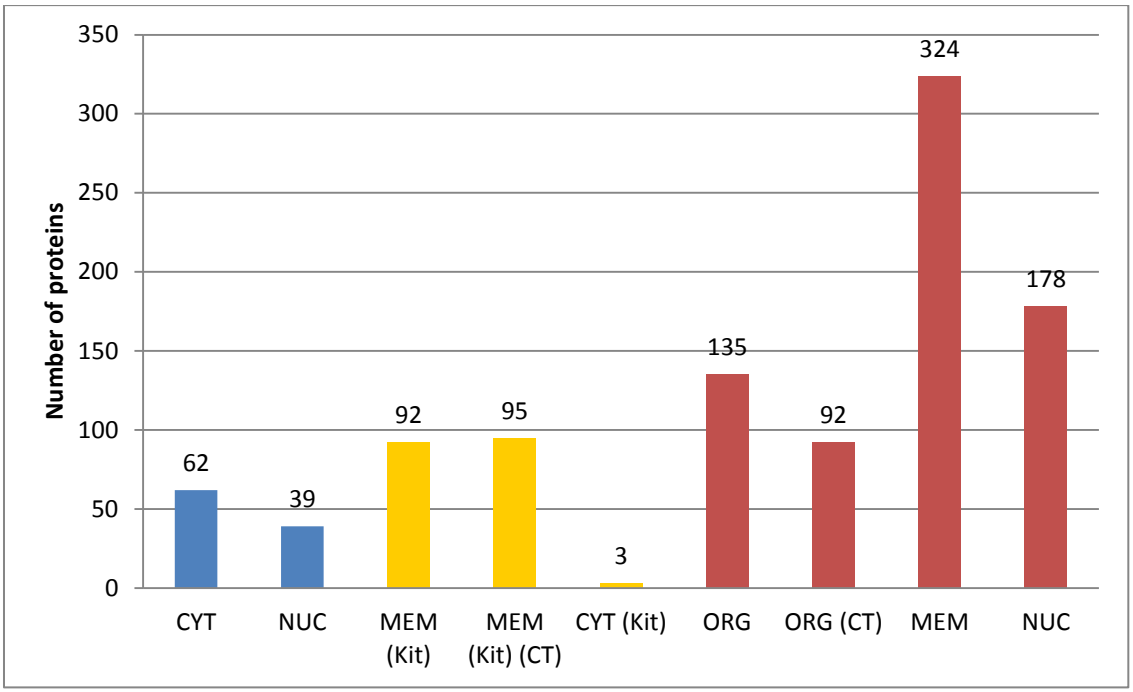


Fig. 27. Number of multipass membrane proteins in various samples.

Compared on the basis of the biological process in which they participate, the membrane and organelle fractions contained a substantial number of proteins that are involved in cell signalling. As an example, 547 proteins were identified in the cytoplasm, that play a role in *signal transduction*, in contrast to the 1105 proteins identified in the membrane, which correspond roughly to a quarter of proteins with this annotation. The CD109 antigen, Regulator of G-protein signaling 12, Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit, Proto-oncogene C-crk (p38), various kinds of the Ras-related protein Ras, Lymphocyte transmembrane adapter 1, various Serine/threonine-protein kinases and Mitogen-activated protein kinase kinase kinase 7 are among these.

Of the 5112 proteins that take part in *cell communication*, 1276 proteins were identified in the membrane sample, corresponding also to a quarter of these proteins. This number was only 604 in the cytoplasm isolated with the standard fractionation and 527 in the cytoplasm prepared by the Protein Extraction Kit. In the organelles digested with trypsin, 17 % of proteins with these two annotations were identified (783 and 864 proteins, respectively). Among the significant proteins identified are the Tumor necrosis factor receptor superfamily member 10B, NF-kappa-B essential modulator and the G protein-coupled receptor kinase 5.

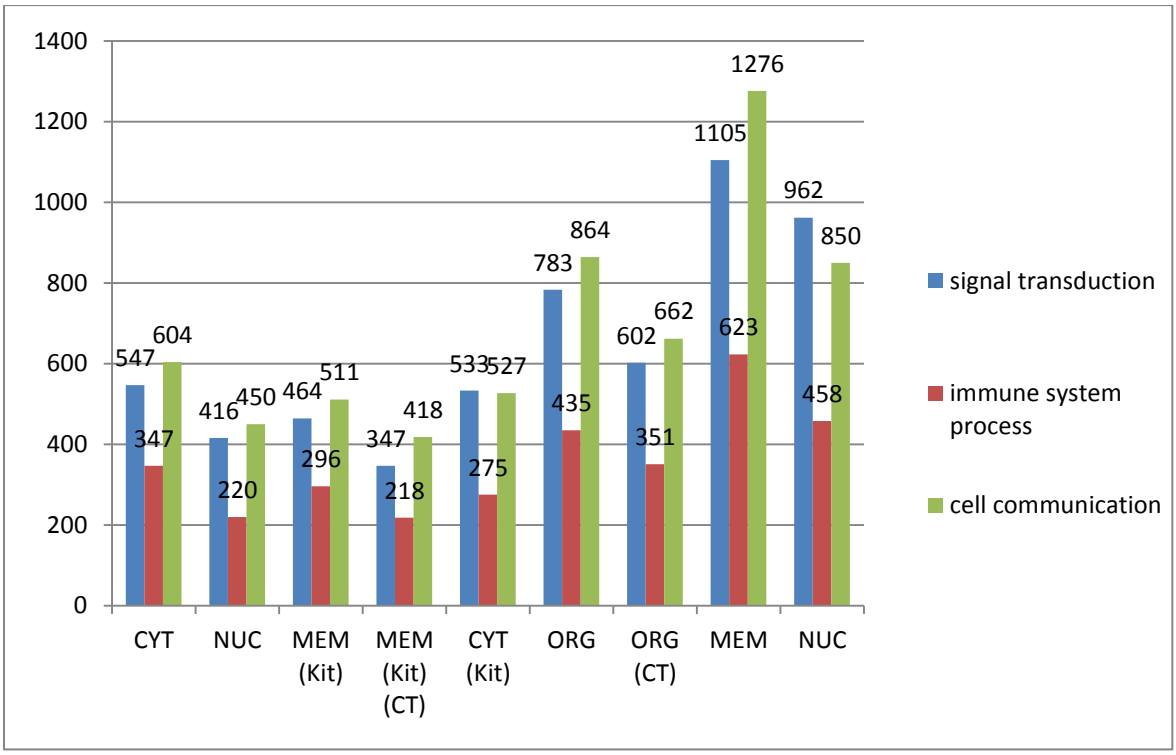


Fig. 28. Number of identified proteins with the GO annotations (biological process) signal transduction, immune system process and cell communication.

As opposed to the 17 % identified in the cytoplasm and 11 % in the nuclei prepared by the standard protocol, almost one third (32 %) of all proteins involved in an *immune system process* were identified in the membrane (623 proteins). These include several vital proteins such as the Interferon gamma receptor 1, Tyrosine-protein kinase JAK1 (Janus kinase 1), Tyrosine-protein kinase JAK3 (Janus kinase 3), the alpha subunit of cAMP-dependent protein kinase type II, the Tyrosine-protein kinases ZAP-70, CSK and ITK/TSK, Proto-oncogene tyrosine-protein kinases LCK, Fyn, Yes, Src, Mitogen-activated protein kinase 8 (MAP kinase 8), the Inositol-trisphosphate 3-kinase B. On the other hand, the organelle and nuclear fractions led to the identification of 23 % of the proteins with this GO annotation.

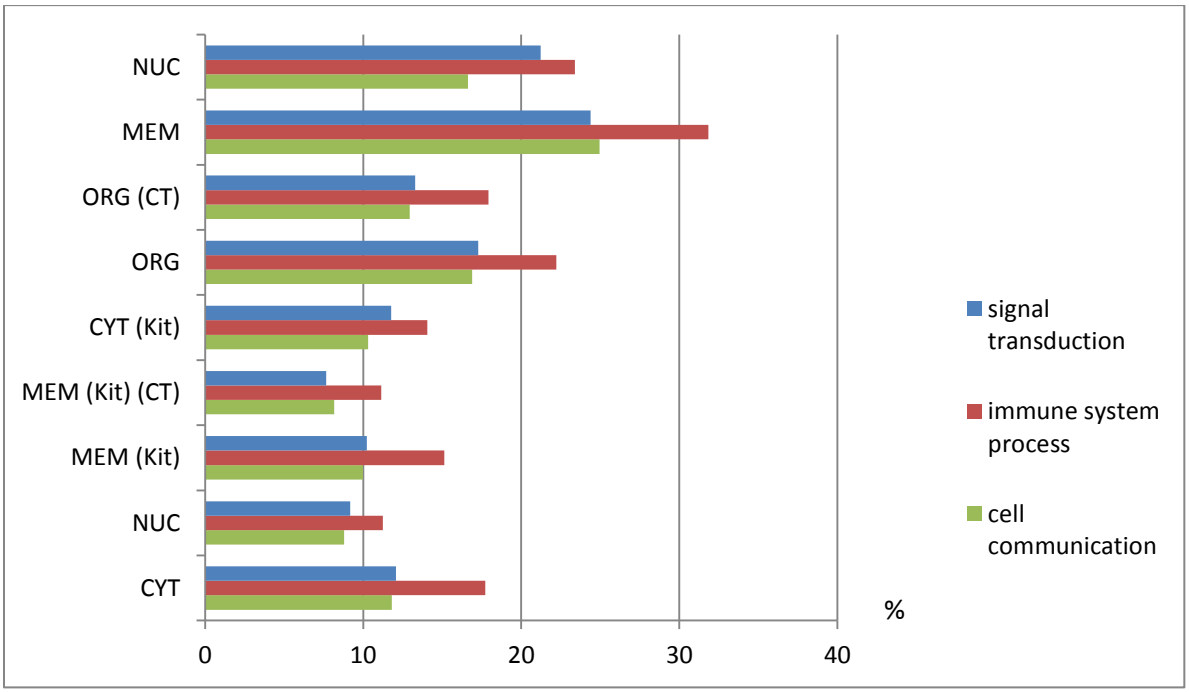


Fig. 29. Coverage of proteins with the above annotations in various samples as a percentage.

As to the molecular function they fulfill, 282 proteins were identified in the membranes that have *transporter activity*, corresponding to 24 % of all proteins with this GO annotation. In the organelles, 222 proteins were identified, and in the nuclear fraction extracted from cells disrupted with the freeze-thaw cycle 191 proteins, whereas the cytoplasm of the standard protocol contained 119 transporter proteins.

Among the proteins identified in the membrane fraction are the Sodium/hydrogen exchanger 1 (Na^+/H^+ antiporter), subunits of the Sodium/potassium-transporting ATPase, ADP/ATP translocase 2, Solute carrier family 12 member 4 and 7 (Electroneutral potassium-chloride cotransporter 1 and 4), Sodium/sialic acid cotransporter Sialin, Membrane magnesium transporter 1 (Transmembrane protein 32), Sodium-dependent phosphate transporter 2, ATP-binding cassette sub-family B member 2, Multidrug resistance-associated protein 4 (ATP-binding cassette sub-family C member 4), Sodium/myo-inositol cotransporter, Thiamine transporter 1, ATP-binding cassette sub-family D member 4 (Peroxisomal membrane protein 69).

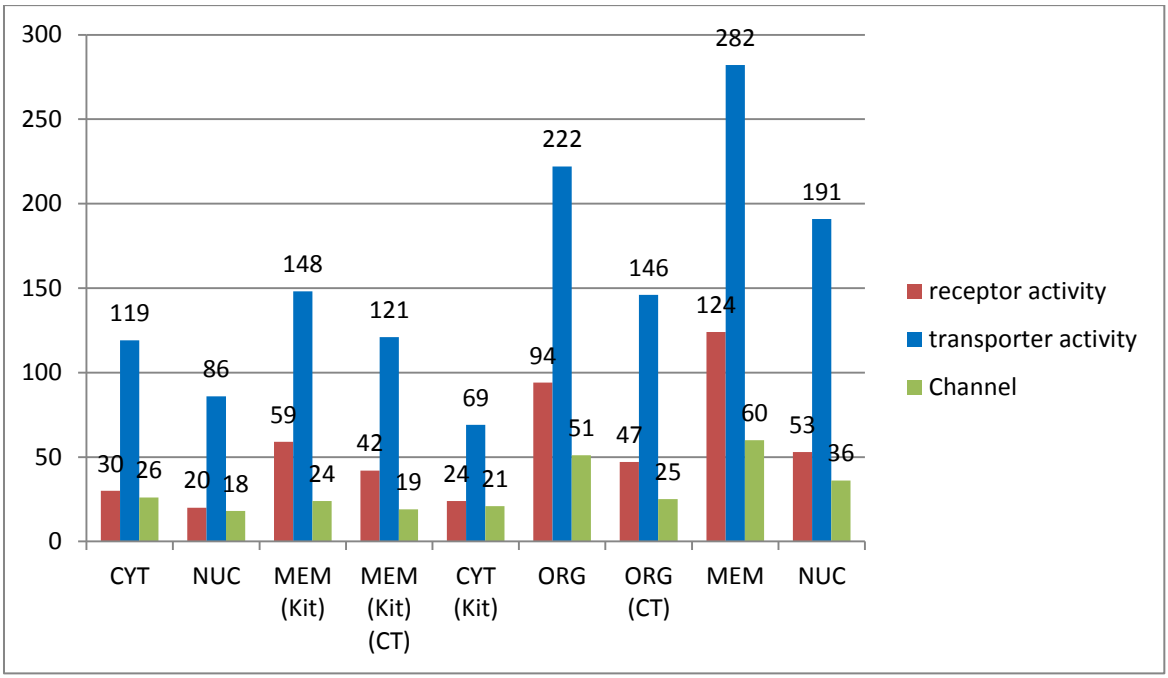


Fig. 30. Number of identified proteins with the GO annotations (molecular function) receptor activity, transporter activity and channel.

124 proteins with *receptor activity* have been identified in the membrane sample, in contrast to the mere 30 proteins in the cytoplasm. Some receptor proteins crucial in the reguation of the cell fate are the CD166 antigen (Activated leukocyte cell adhesion molecule), CD97 antigen, tumor necrosis factor receptor superfamily member 8, T-cell surface antigen CD2 (rosette receptor), transferrin receptor protein 1, receptor-type tyrosine-protein phosphatase F and receptor-type tyrosine-protein phosphatase S. The identified CD97 antigen is an adhesion G protein-coupled receptor, specifically, a member of the EGF-TM7 subfamily, which plays a role in immune function and is upregulated in various cancers.

60 *channel* proteins in the membrane (11 % coverage) as well as 51 in the organelles, whereas only 26 were found in the cytoplasm. Among the channel proteins identified in the membranes are voltage-dependent anion-selective channel protein 2, inhibitor of nuclear factor kappa-B kinase subunit beta, calcium-binding protein p22, voltage-dependent anion-selective channel protein 3 and calcium release-activated calcium channel protein 1 (transmembrane protein 142A).

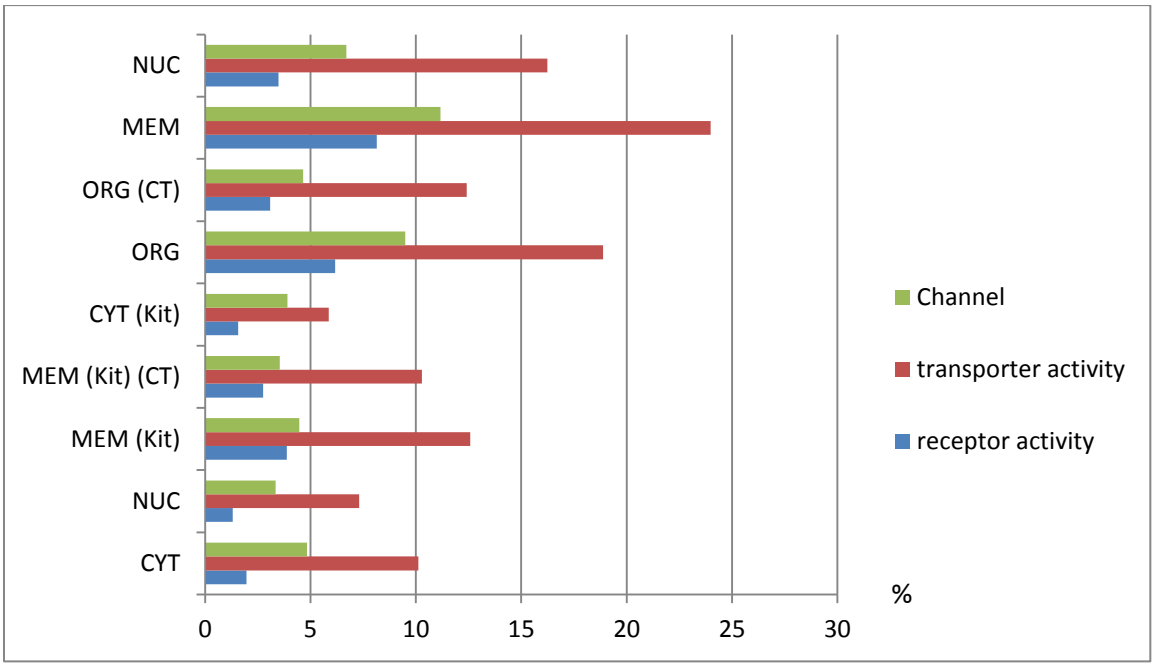


Fig. 31. Coverage of proteins with the above annotations in various samples as a percentage.

A comparison of the proteins identified in the membrane and cytoplasm (standard protocol) fractions shows that the latter is quite reduntant, especially for the analysis of membrane proteins, because the proteins identified in the cytoplasm is almost a subset of those in the membrane, a few as 29 proteins being found only in the cytoplasm.

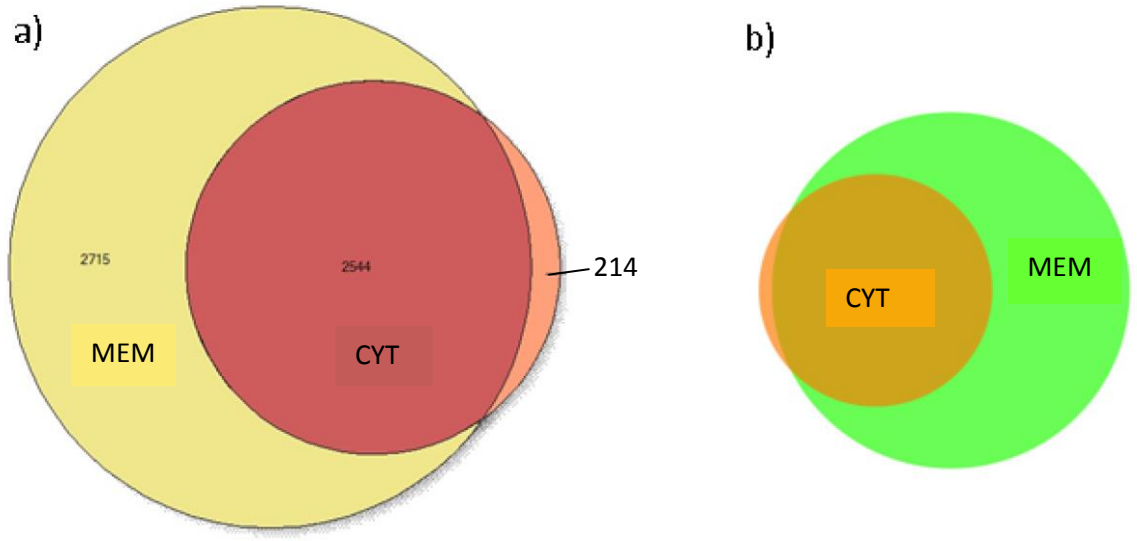


Fig. 32. a) the total number of proteins identified by the protocol applied for the analysis of membrane proteins involving cell disruption by liquid N_2 and ultracentrifugation (*membrane* fraction) versus those identified by the standard protocol (*cytoplasm*). b) a comparison of the number of membrane proteins identified in both samples shows that the proteins identified in the cytoplasmic fraction are almost a subset of those in the membrane.

Inevitably, the organelle fraction contains cytoplasmic proteins which were pelleted with the organelles. In fact, 1202 of the identified 4624 proteins had the GO annotation cytosol. On the other hand, the membrane had relatively fewer cytosolic proteins: 1094 of the identified 5253 proteins had this GO annotation. Because of its high content in plasma membrane proteins and cytosolic proteins, it is not surprising that the samples organelles and cytoplasm overlap a great deal.

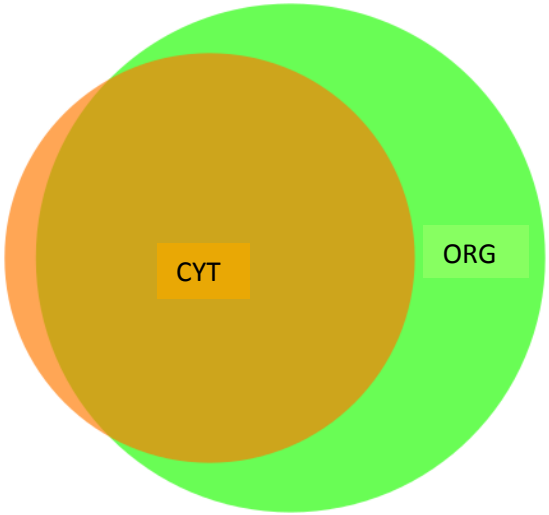


Fig. 33. A comparison of the *membrane proteins* identified in the samples *organelles* and *cytoplasm*. Of the 1292 proteins identified in the organelles, 784 were also found in the cytoplasmic fraction isolated by the standard protocol and 54 were found in the cytoplasm only. That leaves 508 membrane proteins which were identified in the *organelles only* and not in the cytoplasm.

On the other hand, the membrane fraction contains many organelle membrane proteins because these were not completely pelleted during the 20 min centrifugation at 20 000 g, but during ultracentrifugation.

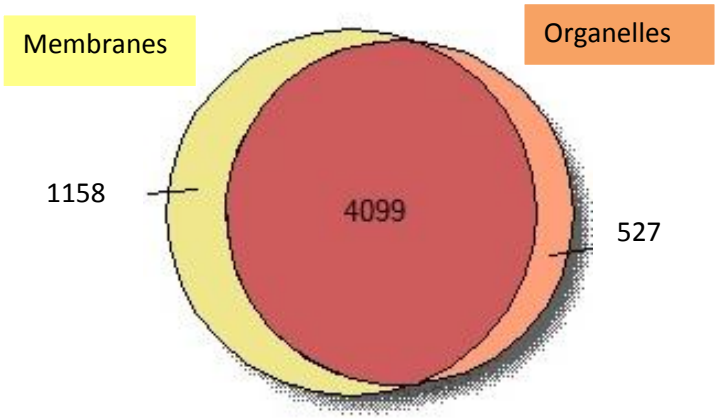


Fig. 34. A comparison of all proteins identified in the membranes and organelles.

Of the 1293 *membrane proteins* identified in the organelles, only 74 were identified exclusively in this fraction; the remaining 1219 were in common with the membrane fraction. Specifically the proteins with the GO annotations *organelle membrane* and *intrinsic to organelle membrane* were found to be more abundant in the membranes than in the

organelles (752 versus 702, and 78 versus 50). As can be expected in light of this information, the organelle fraction contained several proteins with the annotation plasma membrane (584) and intrinsic to plasma membrane (138) (see figures 19 and 22).

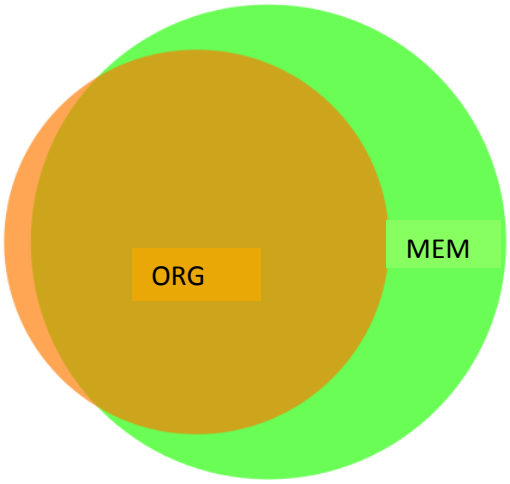


Fig. 35. A comparison of the *membrane proteins* in the membranes and organelles shows that most MP identified in the organelles were also identified in the membranes, 74 proteins were identified in organelles only and 751 MP were identified in the membrane fraction only, 1218 proteins being common to both samples.

3.2. Comparison on the Basis of Hydropathy

Because membrane-spanning domains of integral membrane proteins consist of hydrophobic peptides, the distribution of the Grand Average of Hydropathy (GRAVY) scores of the identified proteins in Kit membranes was analysed. The GRAVY score of a protein is a measure of its hydrophobicity, in other words of its solubility in an aqueous medium. As an example, a positive GRAVY score for a certain domain of a protein indicates its lipophilicity and low solubility in water. As discussed in the above sections in the context of membrane proteins, hydrophobicity leads to the problems of low solubility in the common media used in sample preparation as well as adsorbtion on surfaces and thus loss of MPs. Despite these problems, a remarkable number of essential membrane proteins including numerous transporters and receptors have been identified in the membrane-enriched samples. This shows that the isolation of hydrophobic proteins as well as their separation in SDS-PAGE was performed successfully, without a considerable loss due to the challenges mentioned. Chromatographic separation of the tryptic peptides constitutes the limiting factor at this point,

with long retention times due to the lipophilicity of the C18 column as well as the limited hydrophobicity of the eluent. Hydrophobic proteins, on the other hand, can still be identified through their hydrophilic peptides. Hence, to evaluate the systematic limitation of HPLC, the GRAVY scores of the identified peptides and proteins have been plotted.

Yet another reason why the GRAVY score of a protein is relevant for the purpose of this study, is that information about its possible structure can be inferred. Algorithms predicting transmembrane and cytosolic domains of a protein from its sequence make use of the hydrophobicity scores of its various regions and then plots the predicted topology.

To calculate the GRAVY score of a protein according to Kyte and Doolittle (1982), each amino acid is assigned a hydrophobicity score, the most hydrophobic being isoleucine (+4,5) and the most hydrophilic arginine (-4,5). The hydrophobicity scores of amino acids in a certain window (e.g. of 9 amino acids) are averaged and assigned to the first amino acid in the window. Then, the window is shifted by one amino acid and the average of the amino acids in that window is again assigned to the first amino acid in the window. Thus, the GRAVY score of a protein is determined as the average hydropathy score for all the amino acids in that protein.

Because membrane proteins are typically more hydrophobic than globular proteins[38], a shift of the curve showing the frequency of proteins versus the GRAVY score to the more hydrophobic range would be expected in the membrane-enriched samples. On the other hand, considering that chymotrypsin cleaves at the large hydrophobic amino acids tyrosin, tryptophan and phenylalanine, an increase in hydrophobicity of proteins identified in the chymotrypsin-digested samples would be predicted.

The plot of the Kit membranes did not exhibit a remarkable number of proteins in the hydrophobic range. Digestion with chymotrypsin also did not result in the identification of a higher proportion of hydrophobic proteins than the nuclear fraction prepared by the standard protocol.

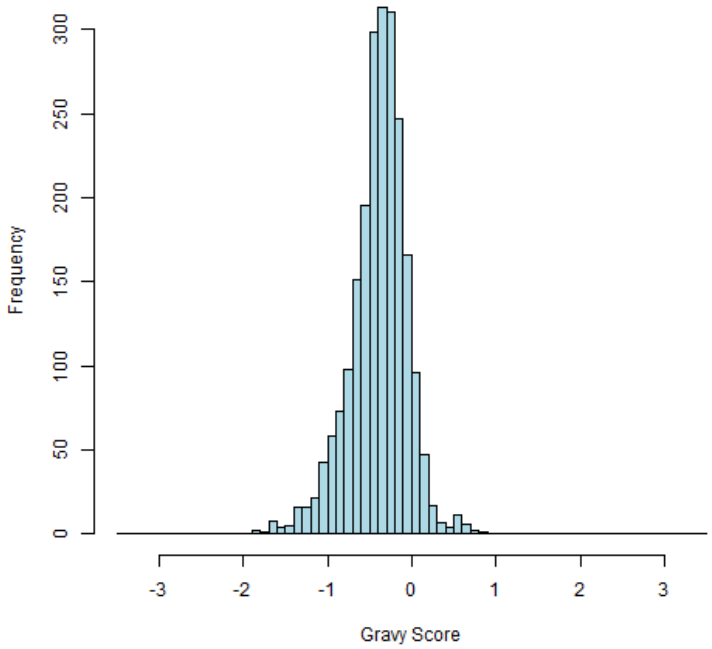


Fig. 36. Distribution of the GRAVY scores of identified proteins in the Kit membranes digested with trypsin.

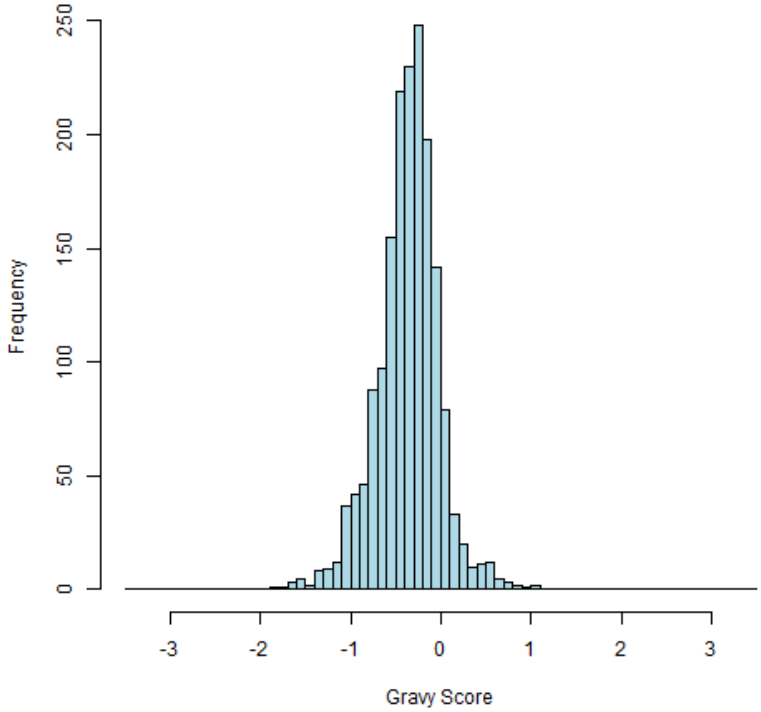


Fig. 37. Distribution of the GRAVY scores of identified proteins in the Kit membranes digested with chymotrypsin.

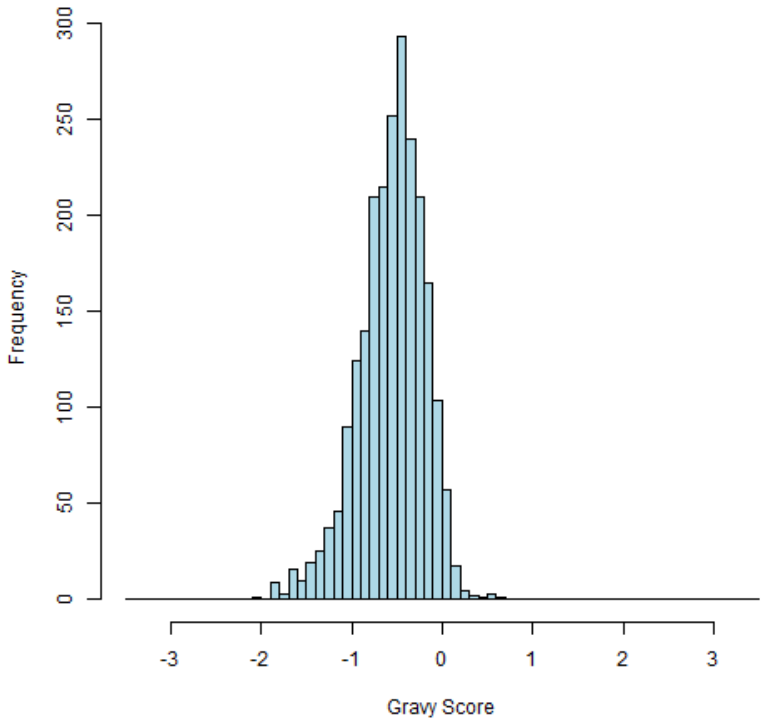


Fig. 38. Distribution of the GRAVY scores of identified proteins in the nuclear extract isolated and analysed by the standard protocol.

A shift in the hydrophobicity was also not observed on the basis of identified *peptides* in the Kit membranes (trypsin or chymotrypsin) compared to the standard protocol.

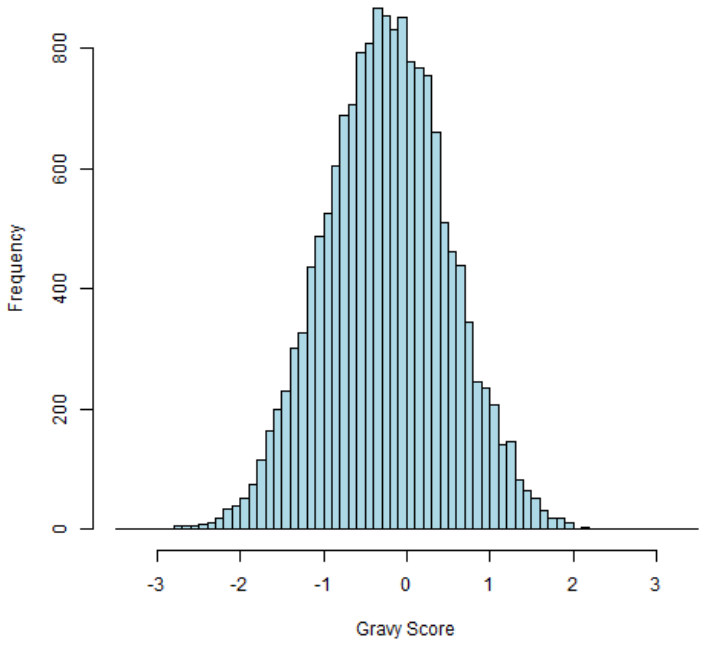


Fig. 39. Distribution of the GRAVY scores of identified pepetides in the Kit membranes digested with trypsin.

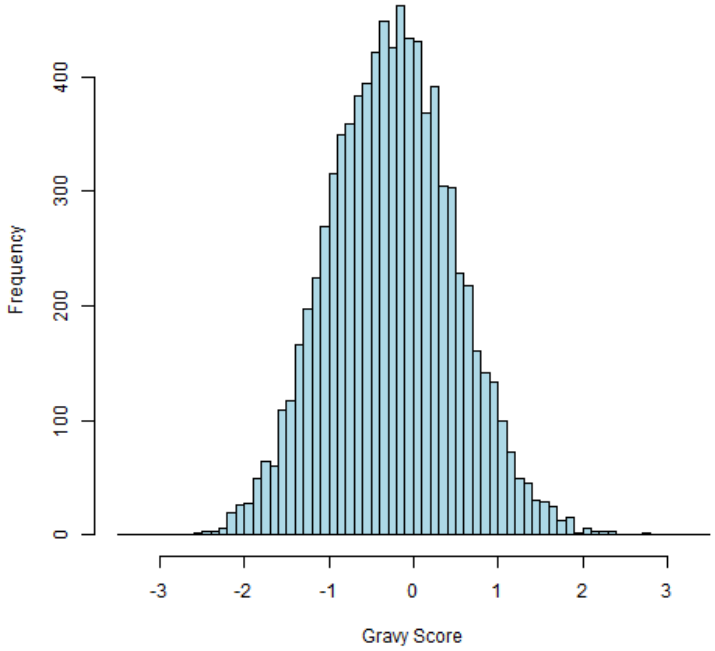


Fig. 40. Distribution of the GRAVY scores of identified proteins in the Kit membranes digested with chymotrypsin.

Thus, the GRAVY score does not relate immediately to the proportion of identified membrane proteins, and the distribution on the protein level may remain well in the hydrophilic range, while the number of identified MPs increases. In fact, several transmembrane (even multipass) proteins have a GRAVY score below zero and a small portion of membrane protein have a score greater than or equal to 1.

On the peptide level, transmembrane regions with a score around 1 constitute a noticeable portion of the curve. Specifically for these peptides, calculating the GRAVY score using a window size of 19 instead of the standard size 9 might give even better results.

4. Conclusion

By the proteomic analysis of the membrane-enriched fractions of Jurkat cells (organelles, membranes and nuclei), identification of an uncommonly high number of different membrane proteins was achieved, the membranes enriched by ultracentrifugation yielding the best results and the trypsin-digested organelles coming closest to these. This and especially the extraordinarily high number of identified *multipass* transmembrane proteins might lead one to expect a shift in the hydropobicity of the observed peptides and proteins. However, the distribution of GRAVY scores of various samples was similar. In fact, membrane proteins rarely have a GRAVY score more positive than 0,3 and peptides greater than 1[39].

By the experiments involving in-gel digestion performed for this master's thesis, it was possible to identify a number of membrane proteins which is comparable to those reported from in-solution experiments[40]. In addition, the fractionation procedure combined with in-gel digestion with trypsin yielded even better results than several membrane proteomics studies involving various detergents, the use of high temperature or high pH and proteinase K[41]. Yet, although the protocol applied in this work for shotgun proteomics of membrane proteins is competent in terms of qualitative analysis, it would hardly be applicable to quantitative measurements due to the limited reproducibility of membrane isolation.

With the samples generated by the Membrane Protein Extraction Kit, it was possible to characterize more *integral* membrane proteins than with the standard protocol, in spite of the fact that the number of membrane proteins in general was comparable with those identified in the cytoplasm. For all GO annotations, the Kit samples performed roughly between the standard protocol for cytoplasmic proteins and the membrane-enriched samples. On the other hand, digestion with chymotrypsin led to the identification of a lower number of total and membrane proteins in the membrane fraction isolated with the Kit as well as in the N_2-extracted organelles. Adding acetonitril to the digestion buffer also proved to be unfavorable and was no longer tested.

With the applied isolation, digestion and analysis protocols, it was possible to capture a significantly higher number of membrane proteins compared to the cytoplasmic and nuclear fractions digested and analysed by the standard protocol. Of the 5313 proteins identified in the ultracentrifuged membrane fraction, 1969 (37 %) had the annotation *membrane. Integral* membrane proteins counted 581, of which 317 were *multipass* transmembrane proteins. 1276 proteins were identified in the ultracentrifuge-membranes which were involved in *cell communication*, 1105 in *signal transduction* and 623 in *immune response*. These numbers are comparable with the number of proteins identified in other cells and tissues (e.g. brain) which are known to have a higher content of protein[42]. With respect to their molecular

function 282 proteins had transporter activity, 124 were receptors and 60 channel proteins. As Jurkat cells do not possess an untypically large number of various receptor and channel proteins, the number of those listed with these two GO annotations is actually larger than commonly reported[43]. In conclusion, the acquired data represent the currently most comprehensive study of membrane proteins of human Jurkat cells, a model cell line used for many proteome studies. The membrane isolation kit provided limited advantage, most important was the use of an optimised LC-MS method for the identification of low membrane proteins.

5. List of Figures

6. References

[1] Vuckovic, D., Dagley, L.F., Purcell, A.W., Emili, A. Membrane proteomics by high performance liquid chromatography - tandem mass spectrometry: analytical approaches and challenges. Proteomics. 13(3-4):404-23 (2013).

[2] Alberts et al. Molecular Biology of the Cell. (2008) Fifth Edition, Taylor and Francis, p. 630.

[3] İbid.

[4] İbid. p. 879.

[5] İbid. p. 892

[6] İbid. p. 667ff.

[7] İbid. p. 904ff

[8] Ibid. p. 906.

[9] İbid. p.921ff.

[10] İbid. p.937ff.

[11] Cox, J. And Mann, M. Is Proteomics the New Genomics? Cell 130, Aug 10 395-398 (2007).

[12] Wis´niewski, J. R., Zougman, A., Nagaraj, N, Mann, M. Universal sample preparation method for proteome analysis. Nature Methods 6, 359 - 362 (2009).

[13] Donato, P., Cacciola, F., Mondello, L., Dugo, P., Comprehensive chromatographic separations in proteomics. Journal of Chromatography A. 1218(49): 8777–8790 (2011).

[14] Lottspeich, F., Engels, J.W. Bioanalytik. (2012) 3. edition, Springer, p. 271.

[15] http://www.piercenet.com/browse.cfm?fldID=21518847-2D72-475F-A5B9-B236EC5B641E

[16] Köcher, T., Pichler, P., Mechtler, K., Swart, R. Analysis of protein mixtures from whole-cell extracts by single-run nanoLC-MS/MS using ultralong gradients. Nat Protoc. Apr 12;7(5):882-90 (2012)

[17] Cshprotocols.chslp.org

[18] Köcher, T., Pichler, P., Mechtler, K., Swart, R. Analysis of protein mixtures from whole-cell extracts by single-run nanoLC-MS/MS using ultralong gradients. Nat Protoc. Apr 12;7(5):882-90 (2012)

[19] Makarov, A. et al.The Orbitrap: a new mass spectrometer. J. Mass Spectrom. 40: 430–443 (2005)

[20] Makarov, A., Scigelova, M. Coupling liquid chromatography to Orbitrap mass spectrometry. Journal of Chromatography A, 1217:3938–3945 (2010)

[21] Mann, M. And Walther, T.C. Mass spectrometry–based proteomics in cell biology. J. Cell Biol. Vol. Aug;190(4):491–500 (2010)

[22] N.N. http://www.matrixscience.com/help/fragmentation_help.html (2013)

[23] J. Cox, M. Mann, Computational Principles of Determining and Improving Mass Precision and Accuracy for Proteome Measurements in an Orbitrap. J. Am. Soc. Mass Spectrom. 20(8): 1477 (2009)

[24] Makarov, A., Scigelova, M. (2010) Ibid.

[25] Wollscheid, B., Watts, J.D., Aebersold, R. Proteomics/genomics and signaling in lymphocytes. Current opinion in immunology. 16(3), 337–344 (2004)

[26] Mann, M. And Walther, T.C. (2010) Ibid.

[27] http://www.matrixscience.com/help/pmf_help.html (2013)

[28] Wu, C.C., Yates, J. R. The application of mass spectrometry to membrane proteomics. Nat Biotechnol. Mar;21(3):262-7 (2003)

[29] Blonder, J., Goshe, M.B., Moore, R.J., Pasa-Tolic, L., Masselon, C.D., Lipton, M.S., Smith, R.D. Enrichment of integral membrane proteins for proteomic analysis using liquid chromatography-tandem mass spectrometry. J. Prot. Res. 1, 351–360 (2002).

[30] Westermeier, R., Naven, T. Proteomics in Practice. (2002) Wiley-VCH , Weinheim.

[31] MPSD Project, Priv.Doz. Dr. T. Nawroth, Universität Mainz. http://mpsd.de/MPSD_MPro.html

[32] Choksawangkarn, W., Edwards, N., Wang, Y., Gutierrez, P. and Fenselau, C. Comparative Study of Workflows Optimized for In-gel, In-solution, and On-filter Proteolysis in the Analysis of Plasma Membrane Proteins. J Proteome Res. May 4;11(5):3030-4 (2012).

[33] https://www.nationaldiagnostics.com/electrophoresis/article/protein-fixation-gels

[34] www.protocol-online.org/files/protocol_4142.pdf

[35] http://www.piercenet.com/instructions/2162519.pdf

[36] Johannes Rössel, http://en.wikipedia.org/wiki/Directed_acyclic_graph

[37] http://www.geneontology.org/GO.doc.cc-membrane-prots.shtml

[38] http://gcat.davidson.edu/rakarnik/kyte-doolittle-background.htm

[39] Zhou J, Lin Y, et al. Development and application of a two-phase, on-membrane digestion method in the analysis of membrane proteome. J Proteome Res. Apr;7(4):1778-83 (2008)

[40] Wis´niewski, J. R., Zougman, A., Mann, M. Combination of FASP and StageTip-Based Fractionation Allows In-Depth Analysis of the Hippocampal Membrane Proteome. J. Prot. Res. 8: 5674–5678 (2009)

[41] Wu, C.C., Speers, A. E., Blackler, A. Shotgun analysis of integral membrane proteins facilitated by elevated temperature. Anal Chem. Jun 15;79(12):4613-20 (2007)

[42] Wollscheid, B., Proteomic Analysis Reveals Drug Accessible Cell Surface N-Glycoproteins of Primary and Established Glioblastoma Cell Lines. J Proteome Res. 11:4885−4893 (2012)

[43] Zhou J, Lin Y, et al. Ibid.

Lebenslauf

Persönliche Daten:

Vor- und Zuname:	Selin Hizal
Geburtsdatum:	23.06.1979
Geburtsort:	Istanbul
Wohnort:	Gentzgasse 52/2/3, A-1180 Wien
Email Adresse:	a0608966@unet.univie.ac.at
Staatsangehörigkeit:	Türkei
Familienstand:	ledig

Universitäres Studium:

10/2012 – 06/2013	Diplomarbeit am Institut für Analytische Chemie der Universität Wien
10/2010	Beginn Masterstudium der Chemie
	Schwerpunkte: Bioorganische Chemie, Bioanorganische Chemie,
	Bioanalytik
10/2006 - 06/2010	Bachelorstudium (BSc.) der Chemie an der Universität Wien
06/2005	Universitäre Zulassungsprüfung zum Studium der Chemie,Istanbul

Berufspraxis:

03/2003-03/2004	Eczacibasi Holding, Verantwortliche der Kunstsammlung

Universitäre Weiterbildung:

10/1997 - 09/2003	Kunst- und Kulturmanagement an der Universität Yildiz, Istanbul

Schulausbildung:

09/1985 - 06/1997 Isik Lisesi, Istanbul

Interessen:

Medizinische Chemie, Biologische Chemie, Pharmazie

Besondere Kenntnisse:

Fremdsprachen	Deutsch (sehr gut)
	Englisch (sehr gut)
IT-Kenntnisse	MS-Office (Excel, Word, Powerpoint)
	Chemische Software
Hobbys	Fotografieren, Schwimmen, Handarbeit