Proteomic characterization of host response to viral infection

Niina Lietzén

ACADEMIC DISSERTATION

To be presented for public examination with the permission of the Faculty of Biological and Environmental Sciences of the University of Helsinki in the Auditorium 2402 (Telkänpönttö) at Viikki Biocenter 3, Viikinkaari 1, Helsinki, on May 16th at 12 o’clock noon.

Helsinki 2012
Supervisors

Docent Tuula Nyman  
Institute of Biotechnology  
University of Helsinki  
Helsinki, Finland

Docent Sampsa Matikainen  
Unit of Excellence in Immunotoxicology  
Finnish Institute of Occupational Health  
Helsinki, Finland

Thesis committee

Docent Jaana Vesterinen  
Institute of Biomedicine  
Biochemistry and Developmental Biology  
University of Helsinki  
Helsinki, Finland

Professor Juho Rousu  
Department of Information and Computer Science  
Aalto University  
Espoo, Finland

Reviewers

Docent Jaana Vesterinen  
Institute of Biomedicine  
Biochemistry and Developmental Biology  
University of Helsinki  
Helsinki, Finland

Docent Sampsa Hautaniemi  
Institute of Biomedicine and Genome-Scale Biology Research Program  
Faculty of Medicine  
University of Helsinki  
Helsinki, Finland

Opponent

Professor Kris Gevaert  
VIB Department of Medical Protein Research and Department of Biochemistry  
University of Ghent  
Ghent, Belgium

Custos

Professor Kari Keinänen  
Department of Biosciences  
Faculty of Biological and Environmental Sciences  
University of Helsinki  
Helsinki, Finland

Cover figure: Cytoscape protein interaction network (Niina Lietzén)

ISBN 978-952-10-7955-9 (paperback)  
ISBN 978-952-10-7956-6 (PDF)  
ISSN 1799-7372  
http://ethesis.helsinki.fi

Unigrafia  
Helsinki 2012
1. INTRODUCTION

1.1 Proteins and proteomics

1.2 Methods in proteomics

1.2.1 Sample prefractionation

1.2.2 Protein and peptide separation

1.2.2.1 Gel-based methods

1.2.1.2 Chromatographic methods

1.2.2 Mass spectrometry in proteomics

1.2.2.1 Protein identification by mass spectrometry

1.2.3 Database search engines in protein identification

1.2.4 Quantitative proteomics

1.2.5 Data analysis

1.3 Innate immune system

1.3.1 Innate immune recognition of pathogens

1.3.2 Innate immune responses against viral infection

1.3.3 Influenza A virus

2. AIMS OF THE STUDY

3. MATERIALS AND METHODS

3.1 Cells and stimulations

3.2 Subcellular fractionation and secretome analysis

3.3 Gel-based methods used in proteomic experiments

3.4 Quantitative analysis using iTRAQ

3.5 Mass spectrometry
3.6 Database searches .................................................................................................................. 30
3.7 Protein classification, interaction networks and clustering analysis ................................. 31
3.8 Immunological analyses ....................................................................................................... 32
3.9 Reagents .................................................................................................................................. 32

4. RESULTS ...................................................................................................................................... 34

4.1 Proteomics is an efficient method to study innate immune responses in human keratinocytes and macrophages .......................................................................................... 34

4.2 Mascot and Paragon give comparable protein identification results ................................ 36

4.3 Virus-induced responses in HaCaT keratinocytes and human primary macrophages .......... 37

4.3.1 Several inflammatory pathways are activated in human primary macrophages .......... 37

4.3.2 Viral infection triggers caspase-dependent apoptosis in human macrophages and keratinocytes ................................................................................................................. 39

4.3.2.1 Prediction and identification of potential caspase cleavage targets from proteomic data .......................................................................................................................... 39

4.3.3 Influenza A virus infection and polyI:C transfection trigger significant protein secretion from human primary macrophages ......................................................................... 41

5. DISCUSSION ............................................................................................................................ 43

6. CONCLUSIONS AND FUTURE PERSPECTIVES .................................................................. 49

ACKNOWLEDGEMENTS ............................................................................................................... 50

REFERENCES ............................................................................................................................... 51
LIST OF ORIGINAL PAPERS

This thesis is based on the following original articles that are referred to in the text by their Roman numerals I-V.


- NL did the image analysis and comparison of 2D gels and participated in protein identification, computational data analysis and writing the manuscript.


- NL was responsible of proteomic work, data analysis and writing the manuscript.


- NL did the protein identifications and database searches for proteomic part of the work and participated in making the manuscript.


- NL was responsible of testing the program and writing the manuscript. NL also participated in developing the concept of the tool.


- NL was responsible of testing the tool and analyzing proteomic data with the tool. NL participated in writing the manuscript.

*, # Authors with equal contribution

The original articles were reprinted with the permission of the original copyright holders.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-DE</td>
<td>two-dimensional gel electrophoresis</td>
</tr>
<tr>
<td>2D DIGE</td>
<td>two-dimensional differential gel electrophoresis</td>
</tr>
<tr>
<td>ESI</td>
<td>electrospray ionization</td>
</tr>
<tr>
<td>FDR</td>
<td>false discovery rate</td>
</tr>
<tr>
<td>FT-ICR</td>
<td>fourier transform-ion cyclotron resonance</td>
</tr>
<tr>
<td>GO</td>
<td>Gene Ontology</td>
</tr>
<tr>
<td>HILIC</td>
<td>hydrophilic interaction liquid chromatography</td>
</tr>
<tr>
<td>IFN</td>
<td>interferon</td>
</tr>
<tr>
<td>ICAT</td>
<td>isotope coded affinity tags</td>
</tr>
<tr>
<td>iTRAQ</td>
<td>isobaric tag for relative and absolute quantitation</td>
</tr>
<tr>
<td>LC</td>
<td>liquid chromatography</td>
</tr>
<tr>
<td>LIT</td>
<td>linear ion trap</td>
</tr>
<tr>
<td>MALDI</td>
<td>matrix assisted laser desorption ionization</td>
</tr>
<tr>
<td>MS</td>
<td>mass spectrometry</td>
</tr>
<tr>
<td>MS/MS</td>
<td>tandem mass spectrometry</td>
</tr>
<tr>
<td>m/z</td>
<td>mass-to-charge ratio</td>
</tr>
<tr>
<td>NLR</td>
<td>NOD-like receptor</td>
</tr>
<tr>
<td>PAMP</td>
<td>pathogen-associated molecular pattern</td>
</tr>
<tr>
<td>pI</td>
<td>isoelectric point</td>
</tr>
<tr>
<td>PMF</td>
<td>peptide mass fingerprint</td>
</tr>
<tr>
<td>polyI:C</td>
<td>polyinosic-polycytidylic acid</td>
</tr>
<tr>
<td>PRR</td>
<td>pattern recognition receptor</td>
</tr>
<tr>
<td>Q</td>
<td>quadrupole mass analyzer</td>
</tr>
<tr>
<td>RLR</td>
<td>RIG-I-like receptor</td>
</tr>
<tr>
<td>ROS</td>
<td>reactive oxygen species</td>
</tr>
<tr>
<td>RPLC</td>
<td>reversed-phase liquid chromatography</td>
</tr>
<tr>
<td>SCX</td>
<td>strong cation exchange chromatography</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>sodium dodecyl sulfate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SEC</td>
<td>size exclusion chromatography</td>
</tr>
<tr>
<td>SILAC</td>
<td>stable isotope labeling of amino acids in cell culture</td>
</tr>
<tr>
<td>TLR</td>
<td>toll-like receptor</td>
</tr>
<tr>
<td>TMT</td>
<td>tandem mass tag</td>
</tr>
<tr>
<td>TOF</td>
<td>time-of-flight mass analyzer</td>
</tr>
</tbody>
</table>
ABSTRACT

Proteomics is defined as large-scale study of proteins, and with current proteomic methods thousands of proteins can be identified and quantified from a single experiment. Efficient methods have also been developed for protein localization, posttranslational modification and interaction studies. Mass spectrometry has an important role in proteomics and it is currently used in almost all proteomic experiments to detect and characterize the proteins or peptides in a sample. In addition, various bioinformatics tools have become increasingly important for proteomics by improving data analysis and by helping in the biological interpretation of complex proteomic data. The combination of proteomics and bioinformatics is nowadays an important tool to study cellular signaling mechanisms under different conditions, for example viral infection.

Viruses entering a host cell are first recognized by host’s innate immune receptors. This recognition activates multiple signaling cascades resulting in antiviral immune responses, inflammation and finally programmed cell death, apoptosis, of the infected cell. The detailed mechanisms of host cell defense responses activated after viral infection are still partially unknown. The aim of this project was to develop and utilize proteomic and bioinformatic methods to characterize host responses to viral infection.

Three different proteomic approaches were used in this project to study virus-induced changes in the proteomes of human epithelial cells and macrophages. First, cytosolic viral RNA-induced responses in HaCaT keratinocytes were studied using cell fractionation, two-dimensional gel electrophoresis and mass spectrometry (MS). Second, influenza A virus-induced changes in the mitochondrial, cytoplasmic and nuclear cell fractions as well as in the secretomes of human primary macrophages were characterized using iTRAQ labeling-based quantitative proteomics. Third, cytosolic viral RNA-triggered protein secretion from human primary macrophages was studied using qualitative high-throughput proteomics utilizing SDS-PAGE and liquid chromatographic separations and MS. Various bioinformatics tools were also used to analyze the protein identification and quantitation data. In addition, two computational tools, Compid and Pripper, were developed to simplify the analysis of our proteomic data. Compid simplifies the comparison of protein identification results from different database search engines and Pripper enables the large-scale prediction of caspase cleavage products and their identification from the collected MS data.
Our studies showed that both influenza A virus and cytosolic viral RNA trigger significant changes in the proteomes of human primary macrophages and HaCaT keratinocytes. Virus-induced changes in the expression of 14-3-3 signaling proteins as well as rearrangement of host cell cytoskeleton were detected in HaCaT keratinocytes. Caspase-3-dependent apoptosis was detected in polyI:C transfected HaCaT keratinocytes as well as in influenza A virus infected and polyI:C transfected human primary macrophages. Our studies with human primary macrophages also showed that several inflammatory pathways, and especially the NLRP3 inflammasome, are activated as a result of viral cytosolic RNA and influenza A virus infection. Additionally, we showed that cathepsins, src tyrosine kinase and P2X7 receptor were involved in the inflammasome activation. Finally, we showed that influenza A virus infection and polyI:C transfection triggered extensive secretion of various different proteins. In conclusion, our proteomic experiments have given an extensive view of cellular events activated in human macrophages and keratinocytes after viral infection.
1. INTRODUCTION

Host cell defense responses against viruses are initiated immediately after viruses’ invasion to the cell. Innate immune system is a complex network of interconnected biological pathways that are responsible of organism’s first defence responses against viruses. Therefore, the study of individual molecules or pathways is usually not sufficient to describe the effects of virus on host cells (Gardy et al. 2009). Proteomics can be used to study protein expression levels, localizations, posttranslational modifications and interactions in a cell at certain conditions (Fields 2001). Various different methods have been developed for these purposes, and especially with mass spectrometry (MS)-based proteomics, thousands of proteins can be characterized in a single experiment (Geiger et al. 2012, Boisvert et al. 2012, Phanstiel et al. 2011). Proteomics has been used, for example, to study the expression and functions of viral proteins (Shaw et al. 2008). Additionally, proteomics has been used to study interactions between host and virus proteins (Naji et al. 2012) as well as virus-induced changes in host cell proteomes (Vogels et al. 2011, Emmott et al. 2010b). Proteomic studies can give important information about the cellular mechanisms activated by viral infection and could be utilized for example to evaluate the pathogeneity of different viruses (Rasheed et al. 2009) or to develop drugs and vaccines against viruses. In this Ph.D. project, different proteomic approaches were utilized to study virus-induced events in human primary macrophages and epithelial cells.

1.1 PROTEINS AND PROTEOMICS

Proteins are the workhorses of a cell. They are the molecular instruments expressing genetic information stored in DNA or RNA. Proteins are involved in almost all biological processes of a cell. Gene expression and thus the level of proteins in a cell is constantly regulated by several different processes. Rate of transcription, posttranscriptional processing and degradation of mRNA as well as rate of translation, posttranslational modification, degradation and transport of proteins all affect on the protein contents of a cell in certain conditions (Figure 1). The most important factor affecting protein levels in a cell is the rate of translation (Schwanhäusser et al. 2011). Therefore, the study of protein contents of a cell gives the most information about the biological processes active in a cell at certain conditions.
Proteome is the entire set of proteins expressed by a cell, tissue or organism at given time under certain conditions (Wilkins et al. 1996). The concept of proteomics evolved in 1990s to describe large-scale studies of proteins. In addition to protein levels, protein-protein interactions, protein localization and posttranslational modifications influence the physiological state of a cell. Therefore, the aim of proteomics is to identify all the proteins present in a sample, to quantify them and to study their localizations, posttranslational modifications and interactions (Fields 2001). At present, thousands of proteins can be identified and quantified in a single proteomic experiment (Geiger et al. 2012, Boisvert et al. 2012, Luber et al. 2010). Efficient methods have also been developed for extensive studies of protein posttranslational modifications and interactions (Phanstiel et al. 2011, Kim et al. 2011, Li et al. 2011, Rees et al. 2011). Thus, proteomics can be used to study molecular mechanisms active in a variety of biological systems.

1.2 METHODS IN PROTEOMICS

There is a large variety of methods available for proteomic experiments nowadays. The method of choice is often determined by the biological question, sample material, costs and instrumentation available for the experiments. However, most of the modern proteomic experiments utilize MS in the analysis of complex protein samples. MS-based proteomic analyses can be roughly divided into two classes: bottom-up and top-down proteomics (Kelleher et al. 1999). The most common approach at the moment is bottom-up proteomics which relies on MS analysis of proteolytic peptides followed by protein inference using computational methods. Despite the wide range of methods available, there is a common workflow for almost all bottom-up proteomic experiments (Figure 2). In top-down proteomics, intact proteins are
injected into a mass spectrometer and fragmented there to characterize them. Top-down proteomic experiments are still quite rare because of both technical challenges and the difficulties in data analysis (Zhou et al. 2012).

**Figure 2.** General workflow for bottom-up proteomic experiments.

### 1.2.1 Sample prefractionation

Proteomic samples originate from various sources: for example from cell cultures, tissues or biological fluids. There are often thousands or tens of thousands of distinct proteins in one sample and concentration range between low- and high-abundant proteins can be several orders of magnitude. Therefore, various prefractionation and enrichment methods are often needed to improve the analysis of such complex samples.

Subcellular fractionation is often used in proteomics to simplify complex samples. Mitochondrial, cytoplasmic, and nuclear fractions as well as other cell compartments can be extracted from intact cells and studied separately (Andreyev et al. 2010, Qattan et al. 2010, Du et al. 2010). Enrichment of proteins into different cell fractions may facilitate the detection of low-abundant proteins (Du et al. 2010). Additionally, subcellular fractionation can give important insights into cellular events since protein localization is often important for its function (Qattan et al. 2010). Sucrose gradient density centrifugation, immunoaffinity purification and free-flow electrophoresis are often used for cell fractionation in proteomic experiments (Hartwig et al. 2009, Lee et al. 2010). In addition, different commercial kits have been developed for the enrichment of specific subcellular organelles (Hartwig et al. 2009). Regardless of the method used for cell fractionation, the enriched fractions usually contain impurities from other cell compartments. On the other hand, proteins may also exist in multiple subcellular compartments and those localizations may vary between different conditions (Qattan et al. 2010, Boisvert et al. 2010, Lee et al. 2010). Also, database annotations of proteins’ subcellular locations are still incomplete and often show only one location per protein. Therefore, it can be difficult to evaluate the quality and results of cell fractionation experiments and care must be taken when reporting these results.
In addition to cell fractionation, several different affinity-based methods can be used to extract a certain group of proteins from complex biological samples. In phosphoproteomics, several different affinity enrichment methods like immunoaffinity enrichments, titanium dioxide chromatography and immobilized metal ion affinity chromatography are used to extract phosphoproteins or –peptides from complex protein mixtures (Thingholm et al. 2009). Affinity enrichments are also used to study other types of posttranslational modifications like protein glycosylation (Vandenborre et al. 2010) and ubiquitination (Hjerpe et al. 2009) as well as protein-protein interactions (Rees et al. 2011).

1.2.2 Protein and peptide separation

After the initial prefractionation or enrichment steps, proteomic samples might still contain hundreds or thousands of distinct proteins making direct MS analysis of these complex samples challenging. When multiple peptides are introduced to the mass spectrometer simultaneously, the instrument may not have enough time to fragment and analyze all of them. Additionally, simultaneous ionization of multiple different peptides may result in signal suppression based on different ionization properties of peptides causing signal losses for some of the peptides (Horvatovich et al. 2010). Therefore, efficient separation of peptides prior to MS is important to decrease the number of different peptides entering the mass spectrometer simultaneously and to increase the dynamic range of analysis. Increased separation efficiency can also result in more peptide identifications and thus more and better quality protein identifications per sample.

The most common separation methods used in proteomics are two-dimensional gel electrophoresis (2-DE), sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and liquid chromatography (LC). Complexity of samples introduced to the mass spectrometer varies clearly between these methods (Figure 3). Simple peptide mixtures originating from one or few proteins can be extracted from 2D gels and introduced to the mass spectrometer. SDS-PAGE and LC separations, on the other hand, result in highly complex peptide mixtures where links between original proteins and resulting peptides are mostly or completely lost prior MS analyses. Therefore, each of these separation methods has its own requirements for MS analysis and the following protein identification.
Figure 3. Protein/peptide separation methods commonly used in proteomics. A) 2-DE separation of proteins results in protein spots containing only one or a few proteins. The spots are excised from the gel individually followed by MS analysis. B) SDS-PAGE gives rough separation of proteins and is often followed by LC separation prior MS analysis. C) In LC separation-based approaches, the whole sample is digested into peptides prior separation. Thus, all LC fractions analyzed by MS contain peptides from multiple proteins.

1.2.2.1 Gel-based methods

In the early days of proteomics, two-dimensional gel electrophoresis was the most important method for protein separation (Wilkins et al. 1996, Görg et al. 2004). In 2-DE, proteins are first separated based on their isoelectric points (pI) using isoelectric focusing followed by size-based separation using SDS-PAGE. As a result, a two-dimensional map of protein spots is created in the gel. The protein spots are then visualized using different staining methods, for example fluorescent dyes (Ünlü et al. 1997, Berggren et al. 2000) or silver staining (O’Connel et al. 1997). Quantitation of protein spots is done based on the intensity of staining in each spot. Finally, the protein spots of interest can be picked from the gel for identification, in-gel digested, and the resulting peptides can be analyzed by MS.

2-DE aims at complete separation of proteins resulting in gels where each spot represents only one protein. In a routine experiment, even 2000 protein spots can be separated in a single gel providing adequate resolution for many proteomic experiments (Görg et al. 2004). 2-DE is a valuable method to study for example protein degradation since protein fragments can be easily
detected based on their vertical positions in a gel (Bredemeyer et al. 2004). 2-DE can also be used to visualize changes in protein posttranslational modifications since, for example, different phosphorylation and glycosylation states of a protein can often be seen as a horizontal series of spots in a 2D gel (Koponen et al. 2011, Di Michele et al. 2010). However, 2-DE has limitations in the analysis of certain protein classes: the limited solubility of hydrophobic proteins hinders their analysis with 2-DE (Rabilloud et al. 2010). Also, proteins with extreme pIs can often not be detected in a 2D gel because of the limited pI range in isoelectric focusing.

Another method for protein separation in a gel is one-dimensional SDS-PAGE. It is a robust method for protein separation and is more universal than 2-DE because it does not suffer from limited solubility of hydrophobic proteins or limited protein pI range in isoelectric focusing. In SDS-PAGE, proteins are separated in a gel based on their size, and only partial separation of proteins in a complex mixture is achieved. SDS-PAGE is often used as a first separation step for complex protein samples and is usually followed by slicing of the gel into pieces, in-gel digestion of proteins in each gel piece and finally identification of the extracted peptides by LC-MS/MS (Fang et al. 2010, Boisvert et al. 2012, Savijoki et al. 2011). For quantitative analysis, SDS-PAGE separations can also be easily combined with protein labeling using stable isotopes (Ong et al. 2003, Boisvert et al. 2012).

1.2.1.2 Chromatographic methods

In addition to gel-based separation, liquid chromatography (LC) is often used in proteomics to simplify complex samples prior MS analysis. Since proteins are very diverse in their chemical properties, it is usually easier to optimize an LC separation for a set of peptides with more uniform characteristics (Motoyama and Yates 2008). Several different LC methods can be used for peptide separation. Reversed-phase liquid chromatography (RPLC) separates peptides based on their hydrophobicity, strong cation exchange chromatography (SCX) based on their ionic character, hydrophilic interaction liquid chromatography (HILIC) based on peptide hydrophilicity and size exclusion chromatography (SEC) based on peptide size. Unlike many other LC methods, SEC separations can also be easily optimized for a heterogeneous mixture of intact proteins and it is therefore used in proteomics to separate both peptides and proteins (Lee et al. 2011, Wisniewski et al. 2010). Finally, affinity chromatography is used for peptide or protein separation based on certain specific characteristics like posttranslational modifications (Thingholm et al. 2009).
Digestion of protein mixtures into peptides increases sample complexity significantly resulting in tens or hundreds of thousands of different peptides with a concentration range over several orders of magnitude. With such complex samples, a single separation step does usually not give sufficient separation. The power of multidimensional liquid chromatography in proteomics was recognized in the beginning of 21st century when Washburn et al. used a combination of RPLC and SCX to identify almost 1500 yeast proteins (Washburn et al. 2001). Since then, multidimensional LC has been used extensively for various proteomic purposes (Table 1). For maximal separation efficiency, retention mechanisms of each LC separation step should be as independent from each other as possible. Orthogonality studies of Gilar et al. showed that fairly good orthogonality for peptide separation can be achieved by various combinations of LC methods (Gilar et al. 2005). However, RPLC is usually chosen as the last separation method due to the compatibility of RPLC eluents with MS analysis. An alternative approach for multidimensional chromatographic analyses is COFRADIC (combined fractional diagonal chromatography), where peptides fractionated in the first dimension are enzymatically modified based on a specific characteristic (e.g. N-terminal peptides) followed by second separation step and peptide sorting (Gevaert et al. 2005). Here, the properties of peptides are modified instead of chromatographic conditions to yield a good separation for a specific peptide class.

During the last 10 years, liquid chromatography has become an increasingly popular separation technique in proteomics. Due to the very limited amount of sample available for most experiments, development of efficient nanoscale LC separation methods has been important for proteomics (Shen et al. 2002). Additionally, multidimensional LC separations are easily automated, minimizing the amount of manual work in each analysis. Ultra high pressure LC systems have shown an increase in separation performance as well as a decrease in separation time, both features being important for high-throughput proteomic experiments (Nagaraj et al. 2012). Flexibility, variability and separation power of LC and its compatibility with many other primary separation methods such as SDS-PAGE and peptide isoelectric focusing (Boisvert et al. 2012, Martins-de-Souza et al. 2009) have made LC a central method in proteomics.
Table 1. Examples of different multidimensional LC separations used in proteomic experiments. pepSEC = size exclusion chromatography for peptides.

<table>
<thead>
<tr>
<th>1st dimension</th>
<th>2nd dimension</th>
<th>study</th>
<th>reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>HILIC</td>
<td>RPLC</td>
<td>phosphoproteomics for estrogen-induced transcriptional regulation</td>
<td>Wu et al. 2011</td>
</tr>
<tr>
<td>pepSEC</td>
<td>RPLC</td>
<td>biomarkers for hepatocellular carcinoma</td>
<td>Lee et al. 2011</td>
</tr>
<tr>
<td>RPLC (pH 10)</td>
<td>RPLC (pH 2)</td>
<td>human NK cell proteome</td>
<td>Dwivedi et al. 2008</td>
</tr>
<tr>
<td>SCX</td>
<td>RPLC</td>
<td>iTRAQ quantitation of Saccharomyces cerevisiae proteome</td>
<td>Ross et al. 2004</td>
</tr>
<tr>
<td>SCX</td>
<td>RPLC</td>
<td>yeast proteome</td>
<td>Washburn et al. 2001</td>
</tr>
</tbody>
</table>

1.2.2 Mass spectrometry in proteomics

Mass spectrometry is a technique used to measure mass-to-charge ratios (m/z) and abundancies of gas phase ions that are introduced into a mass spectrometer. The technique can be used to analyze any molecule that can be converted into a sufficiently stable gas phase ion. In 1980s, development of electrospray ionization (ESI) (Yamashita and Fenn 1984) and matrix assisted laser desorption ionization (MALDI) (Karas et al. 1985) enabled the efficient ionization and subsequent MS analysis of large biomolecules like proteins and peptides (Fenn et al. 1989, Hillenkamp et al. 1990). A few years later, nanoESI was developed for the efficient analysis of small sample amounts (Wilm & Mann 1996). These developments in biological mass spectrometry started a new era of protein analysis and were also crucial for the development of modern MS-based proteomics.

Sensitivity, variability and resolution of MS are nowadays utilized in many fields of proteomics, for example in protein identification and quantitation (Boisvert et al. 2012, Bluemlein et al. 2011, Luber et al. 2010), characterization of posttranslational modifications (Kim et al. 2011, Rajimakers et al. 2010) and study of protein complexes and protein-protein interactions (Li et al. 2011, Rees et al. 2011). Different types of instruments have been built to meet the requirements for different types of proteomic experiments (Table 2). Quadrupole mass analyzers (Q), time-of-
flight (TOF) analyzers, Fourier transform ion cyclotron resonance instruments (FT-ICR), ion traps and various hybrid instruments have been utilized in proteomics, each of them having their own advantages and disadvantages (Domon and Aebersold 2006). One of the newest inventions, Orbitrap mass analyzer with high mass accuracy and high resolution (Makarov 2000, Hu et al. 2005), has pushed the boundaries of proteomics by increasing the amount of information that can be collected from the complex samples.

Table 2. Recent examples of proteomic experiments utilizing different mass analyzers.

<table>
<thead>
<tr>
<th>mass analyzer</th>
<th>large-scale proteome characterization</th>
<th>quantitative proteomics</th>
<th>posttranslational modification analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>FT-ICR</td>
<td>Pounds et al. 2008</td>
<td>Collier et al. 2010</td>
<td>Wang et al. 2011</td>
</tr>
<tr>
<td>QQQ</td>
<td></td>
<td>Bluemlein et al. 2011</td>
<td></td>
</tr>
</tbody>
</table>

1.2.2.1 Protein identification by mass spectrometry

In bottom-up proteomics, mass spectrometer is used to detect and identify peptides rather than proteins (Kelleher et al. 1999). Protein identifications are then retrieved based on peptide identification data using different computational tools. MS-based peptide identification and protein inference can be done by peptide mass fingerprinting (PMF) or by utilizing tandem mass spectrometry (MS/MS).
In PMF analyses, all peptides from one or a few proteins are ionized and introduced to the mass spectrometer simultaneously. Mass spectrometer measures all the masses of ionized peptides, and this combination of peptide masses is considered to be a characteristic “fingerprint” of a protein that can be searched for (Figure 4A) (Gevaert and Vandekerckhove 2000). In PMF, protein identifications are based solely on unique sets of peptide masses and no information about peptide sequences is collected. Therefore, several peptide masses have to be detected for each protein to be able to uniquely assign these masses to a certain protein. PMF analyses are typically used for samples that contain peptides from only one or a few proteins. Therefore, 2-DE separation of proteins followed by MALDI-TOF analyses of protein spots is the most common workflow for PMF analyses.

In MS/MS-based analyses, m/z-values of the ionized peptides are measured first, followed by fragmentation of selected m/z-values and detection of the resulting fragment ions (Figure 4B) (Domon and Aebersold 2006). Fragment ion masses contain information about peptide sequence and thus, both intact peptide masses and sequence information can be retrieved resulting in more reliable protein identifications. MS/MS analyses are used especially in shotgun proteomic experiments where complex proteomic samples are first digested followed by LC separation and ESI-MS/MS analysis of the resulting peptides.
1.2.3 Database search engines in protein identification

High-throughput LC-MS/MS experiments can produce even one million mass spectra per experiment making manual data interpretation impossible. Therefore, various database search engines have been developed to process raw MS data. Sequence searching is the most common method for MS-based protein identification in proteomics. Mascot (Perkins et al. 1999), Sequest (Eng et al. 1994), X!Tandem (Craig et al. 2004) and Paragon (Shilov et al. 2007) are commonly used sequence database search engines for proteomic purposes. In sequence searching, data analysis consists of two consecutive steps: peptide identification and protein inference (Deutsch et al. 2008, Nesvizhskii 2007). In the first step, in silico digestion of all proteins in the protein sequence database is performed. The peptides are created and studied based on user-defined criteria like enzyme specificity, mass tolerance and potential posttranslational modifications. Most database search engines try first to find matches between in silico peptides and experimental data based on intact peptide masses. In silico fragment ion spectra are then created for candidate peptides. After this, a list of potential peptide-spectrum matches is created and qualities of each match are evaluated based on different scoring schemes. In the second step of analysis, peptide identifications are grouped to yield protein identifications. Most database search engines have their own grouping algorithms that handle peptide identification data in slightly different ways resulting in partially different protein identification results for the same set of data. However, a common principle in most search engines is to try to find a minimum set of proteins that can explain all the identified peptides. Sequence searching is a suitable method to study already sequenced organisms since only peptides and proteins whose sequence is present in a database can be detected (Nesvizhskii 2007).

Another type of search engines identifies peptides based on spectral matching of previously observed and identified MS/MS spectra with the collected MS/MS spectra (Lam 2011, Craig et al. 2006, Lam et al. 2007). At present, vast numbers of MS/MS spectra are stored in different data repositories, and it is possible to build extensive spectral libraries from these data (Lam et al. 2011). Comparison between spectral matching search engine SpectraST and sequence database search engine Sequest showed that spectral matching can be a faster and more accurate method for peptide identification than traditional sequence database search engines (Lam et al. 2007). However, quality of MS/MS spectra included in libraries and limitations in library coverage have to be considered when identifying peptides based on spectral matching.
A third type of search engines are *de novo* sequencing algorithms which try to read the peptide sequence directly from MS/MS spectra (Deutsch et al. 2008, Nesvizhskii 2007). This method requires no prior knowledge of peptide sequences that are identified. However, *de novo* sequencing is a computationally heavy process and requires good-quality MS/MS spectra for peptide identification. Additionally, problems with protein inference may appear with complex samples.

Since protein identifications can be performed in numerous different ways using various tools, it is important to be able to evaluate the quality of protein identifications retrieved from a search engine. Most database search engines like Mascot, Sequest and Paragon use a statistical scoring mechanism to assess the reliability of protein identifications. In addition, false discovery rates (FDRs) based on for example target-decoy searches are often used to evaluate the reliability of identifications (Elias and Gygi 2007). In target-decoy strategies, database searches are performed against a composite database of target protein sequences and decoy sequences, the reversed or randomized counterparts of the target protein sequences. Based on the assumption that an incorrect peptide assignment is equally likely to originate from a target or a decoy database, the number of decoy identifications can be used to estimate the total number of incorrect assignments. However, the identities of these false positive assignments can not be determined.

### 1.2.4 Quantitative proteomics

In proteomics, it is often necessary to study changes in protein levels in different conditions. Although proteomic methods for both relative and absolute quantitation have been developed, absolute quantitation is rarely performed in proteomic experiments (Elliott et al. 2009). Peptide’s physicochemical properties affect its ionization efficiency and thus the detected signal intensity in a mass spectrometer. Therefore, a reference ion with known concentration is always required to determine the absolute amount of interesting peptides in a sample. This requirement of reference compounds limits the use of absolute quantitation in proteomics. Instead, most proteomic experiments utilize relative quantitation to study changes in protein levels between different samples. This relative quantitation of proteins can be achieved using either gel-based or MS-based quantitation methods.
In gel-based quantitative proteomics, proteins are usually separated using 2-DE and quantitation is performed based on protein spots detected from the gel. Gels from different samples can be matched and quantitation done based on the intensities of corresponding protein spots in distinct gels. Here, fluorescent dyes (Berggren et al. 2000, Ünlü et al. 1997) or silver staining (Chevallet et al. 2006) are often used for spot detection. Two samples can also be labeled with different fluorophores and separated in the same gel using two-dimensional differential gel electrophoresis (2D DIGE) (Ünlü et al. 1997). Then, relative quantitation is done based on the intensities of the different fluorophores in the same spot. The development of 2D DIGE has improved the quantitation accuracy as well as sensitivity of gel-based quantitative proteomics (Marouga et al. 2005). However, dynamic range of different staining methods used in 2-DE is limited compared to the huge differences in protein abundances in real biological samples (Rabilloud et al. 2010). Sensitivity of staining methods is also sometimes limited hindering the detection of low abundance proteins from the gel and thus their identification and quantitation. In addition, complete separation of proteins is required for quantitative analysis because if more than one protein is present in a single spot, quantitation data cannot be assigned to either of them.

In MS-based quantitative proteomics, different labeling methods or label-free approaches can be used for protein quantitation. Some of the most common labeling strategies in proteomics are stable isotope labeling of amino acids in cell culture (SILAC) (Ong et al. 2002), isotope-coded affinity tags (ICAT) (Gygi et al. 1999), isobaric tag for relative and absolute quantitation (iTRAQ) (Ross et al. 2004) and tandem mass tags (TMT) (Dayon et al. 2008).

In SILAC-based quantitation, protein labeling occurs in cell culture when heavy or light isotopes of common amino acids like arginine and lysine are incorporated metabolically into proteins (Figure 5) (Ong et al. 2002). The labeling is done prior any treatment of the samples minimizing technical variations in sample preparation and analysis. If both lysine and arginine are used in labeling, at least one amino acid in each tryptic peptide should be labeled resulting in the detection of multiple labeled peptides per protein. A mass shift of a few Daltons is detected between the differentially labeled forms of a peptide and relative quantitation of peptides is then performed by comparing MS peak areas of these differentially labeled forms of each peptide. Although SILAC is often used to compare only two or three parallel samples, 5plex SILAC experiments have also been published (Molina et al. 2009). However, multiplexing increases the SILAC sample complexity significantly making the MS analysis more difficult. Finally, for adequate incorporation of labels, viable cell lines that can be cultured long enough are required.
Figure 5. SILAC-, ICAT- and iTRAQ-based quantitation in proteomics. In each of these methods, labeling of proteins/peptides is done at a different stage (first coloured boxes). In SILAC and ICAT experiments, quantitation is based on MS data whereas in iTRAQ, quantitation is based on MS/MS data.

ICAT is a protein labeling method where “heavy” or “light” biotinylated tags are attached to cysteine residues of proteins (Figure 5) (Gygi et al. 1999, Hansen et al. 2003). The proteins are then digested and labeled peptides are enriched using affinity chromatography. Since cysteine is a rare amino acid, enrichment of labeled peptides simplifies the sample mixture significantly. In traditional ICAT method, differentially labeled forms of each peptide show an 8 Da mass difference in MS spectra (Gygi et al. 1999) whereas in the newer, cleavable ICAT method, the corresponding mass difference is 9 Da (Hansen et al. 2003). Quantitation in ICAT can be performed by comparing the peak areas of differentially labeled peptides. Due to the low number of cysteine-containing peptides, ICAT quantitation of a protein is often based on only one or two peptides making the results prone to errors.

iTRAQ and TMT are chemical labeling methods where isobaric tags are attached to peptides after protein digestion (Figure 5)(Ross et al. 2004, Dayon et al. 2008). Both tags are structurally very similar and react with free amino groups of peptides, i.e. N-termini and lysine residues. Each isobaric tag contains a cleavable reporter ion group with a specific mass. Differentially labelled forms of each peptide can be distinguished from each other only after peptide
fragmentation in mass spectrometer when the reporter ion groups are cleaved from the peptide. The quantitation is then performed based on the reporter ion peak areas in MS/MS spectra. Both iTRAQ and TMT are multiplexed methods allowing the analysis of four (4plex iTRAQ), six (6plex TMT) or even eight (8plex iTRAQ) samples in parallel (Ross et al. 2004, Dayon et al. 2008, Pierce et al. 2008). The structure of an iTRAQ label and the principle of iTRAQ-based quantitation are shown in Figure 6.

![Figure 6](image)

**Figure 6.** 4plex iTRAQ labeling. A) Structure of an iTRAQ label and attachment of the label into peptide. B) Differentially labelled forms of one peptide elute simultaneously from RPLC and have the same total mass but can be separated based on MS/MS spectra.

Label-free quantitation can be performed based on peptide peak areas in MS spectra (Bondarenko et al. 2002) or based on spectral counts at MS/MS level (Liu et al. 2004). In both approaches, all the samples are analyzed individually and data analysis and comparison is done computationally after MS analyses. Thus, an unlimited number of samples can be compared with each other. In signal intensity-based measurements, peptide peaks from different runs are matched based on retention times and peptide masses (Bondarenko et al. 2002). Relative quantitation of peptides and subsequently proteins is then performed based on differences in peptide peak areas between runs. Technical reproducibility of LC-MS/MS analyses and minimal overlap of peptides are extremely important because quantitation relies completely on matching the MS data between runs. Spectral counting, on the other hand, is based on the idea that in MS/MS experiments performed using data-dependent acquisition more abundant peptides will
be selected for fragmentation more often (Liu et al. 2004). The method can be refined by taking into consideration for example the number of detectable tryptic peptides for a protein (Rappsilber et al. 2002) and the properties of these peptides (Lu et al. 2007). Both spectral counting- and peptide peak intensity-based label-free quantitation methods have been successfully applied in large-scale proteomic studies (Luber et al. 2010, Mosley et al. 2009, Old et al. 2005).

1.2.5 Data analysis

Database searches of large proteomic datasets result in lists containing thousands of protein identifications. In addition, quantitative data and information about protein posttranslational modifications are often included in these lists. It is extremely difficult to deduce potentially relevant biological processes by the manual inspection of collected data. Thus, numerous bioinformatics tools have been developed to help data interpretation.

Functional classification of the identified proteins is often one of the first data analysis steps after database searches. Gene Ontology (GO) database comprises of a well-standardized set of biological processes, molecular functions and cellular compartments associated to different gene products (Ashburner et al. 2000). Currently, more than 500 000 gene products from several different organisms are annotated in the database. These annotations are often utilized in the initial characterizations of proteomic datasets. AmiGO is the official GO database browsing tool that can be used for example to retrieve GO annotations for a single protein or for simple visualizations of the database’s hierarchical structure (Carbon et al. 2009). GO analyses of large proteomic datasets can be performed using several different bioinformatic tools, such as GeneTrail (Backes et al. 2007) and GOMiner (Zeeberg et al. 2003). These tools utilize different statistical methods to find GO categories that are over- or underrepresented in the dataset of interest compared to a reference dataset such as the genome of the selected organism (Backes et al. 2007, Zeeberg et al. 2003). GO database is manually annotated, and each annotation has an evidence code describing the type of evidence supporting the annotation (Dimmer et al. 2008). This allows the user to evaluate the reliability of the GO analysis results. Although GO classification is a valuable tool in proteomics, it does not provide detailed mechanistic information about the cellular events. Therefore, pathway analyses and protein-protein interaction analyses are used to retrieve more detailed information about the cellular events.
Pathway analyses can be used to study the role of the identified proteins in well-defined biomolecular reactions. The Kyoto Encyclopedia of Genes and Genomes (KEGG) PATHWAY database contains several maps of biochemical pathways, especially metabolic pathways (Ogata et al. 1999). Reactome is another large biological pathway database, which is focused especially on human proteins (Joshi-Tope et al. 2005). Pathway databases contain information about physical and functional interactions between proteins and the annotations in these databases are usually manually curated (Malik et al. 2010). KEGG and REACTOME pathway databases can be mined directly using protein identification data from proteomic experiments. Additionally, some of the GO classification tools, such as GeneTrail and PANTHER, can be used for over- and underrepresentation analysis of biological pathways (Backes et al. 2007, Thomas et al. 2003). Also some additional bioinformatics tools, such as ExPlain, have been developed for biological pathway analyses (Zubarev et al. 2008).

Protein-protein interactions can provide important information about the functional complexes and intermolecular associations in a cell. Protein-protein interaction databases contain both experimentally determined and computationally predicted information about physical and functional interactions between proteins (Malik et al. 2010). Various tools such as String (Jensen et al. 2009), PINA (Wu et al. 2009) and Cytoscape (Shannon et al. 2003) mine the existing protein interaction data stored in these databases and can be used to create protein-protein interaction networks based on these data. PINA and Cytoscape are also capable of integrating functional data into protein-protein interaction networks making them extremely useful and efficient tools in proteomic data analyses (Wu et al. 2009, Shannon et al. 2003). Cytoscape can also be used to incorporate quantitative data into networks and the numerous visualization and analysis possibilities available with Cytoscape make it the most comprehensive visualization tool available for complex proteomic data (Shannon et al. 2003). Since the reliability of annotations in protein-protein interaction databases varies widely, it is important to be aware of the quality of interaction data used in the networks. Filtering out low-quality interactions prior analyses can also simplify the interpretation of the data.

Combinations of Gene Ontology classification, pathway analyses and protein-protein interaction networks are often used when analyzing complex proteomic data. These computational analyses can help building biological hypotheses based on proteomic data. However, different functional experiments are required to verify these hypotheses.
1.3 INNATE IMMUNE SYSTEM

Immune system comprises of various cells and molecules that work to protect an organism against pathogens. It can be divided into two categories: innate immune system and adaptive immune system. Innate immune system is organism’s first line of defence against pathogens. It is responsible for the early detection of invading pathogens and launching of the first immune reactions to eliminate the pathogen (Murphy et al. 2008). The conserved defence mechanisms of innate immune system are triggered immediately after the pathogen has been detected and they protect the organism until the adaptive immune responses against the pathogen have been developed. In addition, innate immune system is required for the development of adaptive immune responses that are fully activated only several days after infection. Adaptive immune responses are targeted specifically against the detected pathogen and are thus more efficient than the unspecific innate immune responses. In addition, adaptive immune system is capable of generating immunological memory that ensures faster and more efficient adaptive immune responses if the same pathogen is re-encountered later.

When organisms encounter with a pathogen, epithelial cells of skin and mucous form the first physical barrier between the pathogen and internal parts of an organism (Medzhitov 2007, Kupper and Fuhlbrigge 2004). Infections can occur only when the pathogen passes this physical barrier. Epithelial cells recognize invading pathogens with their pattern-recognition receptors (PRRs) resulting in the secretion of antimicrobial peptides, chemokines and cytokines. These molecules function as signals of infection to immune cells. Different types of immune cells are activated as a result of infection (Janeway and Medzhitov 2002, Murphy et al. 2008). Macrophages and dendritic cells are important phagocytes present in most tissues. They are activated at early stages of infection when pathogens cross the epithelial barrier. One of their important tasks is to engulf and digest pathogens. In addition, they orchestrate immune responses by inducing inflammation and secreting cytokines and chemokines that activate other immune cells and recruit them to the site of infection. Finally, macrophages and dendritic cells can also function as antigen presenting cells helping the development of adaptive immune responses. Other types of phagocytes working in the innate immune system are neutrophils, eosinophils and basophils that can be activated by inflammatory cytokines and chemokines (Janeway and Medzhitov 2002).
1.3.1 Innate immune recognition of pathogens

Host defence against invading pathogens is initiated when host cells recognize specific microbial components called pathogen-associated molecular patterns (PAMPs) (Janeway and Medzhitov 2002, Akira et al. 2006). PAMPs contain various structures that are essential for microbes, for example viral RNA, DNA and bacterial lipopolysaccharide. Host PRRs are germline-encoded receptors designed to differentiate between self and non-self structures and thus to detect foreign microbial structures invading the cell. Several distinct PRRs are found in mammals, each with their own specificities and roles in the innate immune recognition of pathogens.

Toll-like receptors (TLRs) are the best characterized group of PRRs. They are integral membrane glycoproteins with extracellular domains for the recognition of PAMPs and intracellular domains for signaling (Kumar et al. 2011). Human TLRs 1, 2, 4, 5 and 6 are expressed on cell surface whereas TLRs 3, 7, 8 and 9 are found from endolysosomal membranes. Endolysosomal TLRs are specialized for the recognition of viral nucleic acids (TLR3 for viral dsRNA, TLR7/8 for viral ssRNA and TLR9 for viral DNA) whereas cell surface TLRs recognize mostly bacterial and fungal structures. Recognition of PAMPs by TLRs results in the recruitment of different adaptor molecules such as MyD88 and TRIF. This initiates signaling events resulting in the activation of transcription factors, for example NF-κB, IRF3/7 and MAP kinases, and the production of proinflammatory cytokines and type I interferons (Figure 7).

Another group of PRRs is RIG-I-like receptors (RLRs) found in the cytoplasm of host cells (Kumar et al. 2011). These receptors recognize viral RNA in the cytoplasm of infected cells. There are three receptors belonging to this family: RIG-I and MDA-5, both recognizing different types of viral RNA, and LGP2 which is a positive regulator of RIG-I- and MDA-5-mediated signaling. When RIG-I or MDA-5 are activated with viral RNA, they interact with mitochondrial-antiviral signaling protein (MAVS) located on mitochondrial outer membrane and peroxisomes (Figure 7) (Dixit et al. 2010). Interactions with peroxisomal MAVS result in rapid interferon-independent expression of interferon-stimulated genes whereas mitochondrial MAVS activates type I interferon production with slower kinetics. In addition, activation of RLRs results in the production of pro-inflammatory cytokines.
NOD-like receptors (NLRs) are the third main group of PRRs sensing a wide range of ligands in the cytoplasm (Kumar et al. 2011). NOD1 and 2, for example, are cytoplasmic receptors for bacterial cell wall structures. NOD1- and NOD2-mediated recognition of PAMPS results in the activation of NF-κB or MAP kinases inducing the production of proinflammatory cytokines. Another group of NLRs are the inflammasome components, whose activation results in the assembly of a protein complex called the inflammasome (Martinon et al. 2002). Inflammasome activation results in caspase-1 cleavage followed by the activation of proinflammatory cytokines IL-1β and IL-18 (Figure 7). One of the inflammasome components, NLRP3, is activated by bacterial and viral RNA as well as some endogenous danger signals (e.g. danger signal proteins) and environmental pollutants like asbestos. In addition, NLRP1 and NLRC4 are known inflammasome components recognizing various structures.

**Figure 7.** Innate immune recognition of viral nucleic acids results in the production of interferons and pro-inflammatory cytokines. MAPK = MAP kinases.
1.3.2 Innate immune responses against viral infection

Innate immune responses against viral infection are initiated when host PRRs recognize the invading virus. The most important innate immune responses against viral infection are antiviral responses, inflammation and apoptosis of the infected cells.

Interferons (IFNs) are proteins produced and secreted by virus infected cells. They are antiviral agents helping infected cells to fight against invading viruses. From the three different classes of IFNs (type I, II and III interferons) type I IFNs (interferon α and β) are the most central in innate immune responses against viral infection (Randall and Goodbourn 2008). Production of type I IFNs is initiated after host cell’s TLRs and RLRs have recognized the invading virus. TLRs 3, 7 and 9 as well as RIG-I and MDA-5 each activate type I IFN production via different signaling pathways resulting in the activation of interferon regulatory factors IRF3 and/or IRF7 (Akira et al. 2006).

IFNα/β produced by the infected cells are secreted and can be detected by interferon receptors on the surface of infected and neighboring cells (Randall and Goodbourn 2008). Activation of these IFN receptors initiates the production of interferon-stimulated genes. There are hundreds of interferon-stimulated genes involved in for example host cell transcription and translation, immune modulation, signaling and apoptosis (de Veer et al. 2001). The proteins encoded by these genes are the primary effectors of antiviral immune responses.

Inflammatory responses triggered by viral infection aim at recruitment of leukocytes to the sites of infection and to the elimination of infectious agents. Macrophages residing in infected tissues are important triggerers of inflammatory responses (Medzhitov 2008). Inflammatory responses are initially triggered by PRRs when they recognize PAMPs or specific virulence factors. Additionally, at later phases of infection, inflammation can be induced by endogenous danger signal proteins like HMGB1 and S100A9 secreted by infected cells (Bianchi 2007). Activation of inflammatory pathways results in the production of proinflammatory cytokines (e.g. TNF-α and IL-1β) and chemokines that activate other immune cells and attract them to the site of infection as well as production of proteolytic enzymes like caspases and matrix metalloproteinases for host defence (Medzhitov et al. 2008).
Transcriptional and MAP kinase-mediated activation of pro-inflammatory cytokines can occur via various different pathways depending on the virus (Takeuchi and Akira 2010). TLRs 7 and 9 use MyD88-dependent pathways and TLR3 a TRIF-dependent pathway to activate NF-κB transcription factor and thus the production of proinflammatory cytokines. RIG-I/MAVS interaction activates pathways resulting in, for example, NF-κB activation.

To complete the work of TLRs and RLRs in activating transcription of pro-inflammatory cytokines, a multiprotein complex called the inflammasome functions to activate proinflammatory cytokines IL-1β and IL-18 (Martinon et al. 2002). Inflammasomes consist of a cytoplasmic receptor, an adaptor protein ASC (apoptosis-associated speck-like protein containing a CARD) and caspase-1. There are different types of inflammasomes recognizing different viruses, for example NLRP3 inflammasome recognizing RNA viruses like influenza A virus and AIM2 inflammasome recognizing DNA viruses (Martinon et al. 2009, Hornung et al. 2009). These inflammasomes work as caspase-1 activating platforms (Martinon et al. 2002, 2009). The active caspase-1 can then cleave inactive pro-IL-1β and pro-IL-18 into their active forms which are secreted from the cell to induce inflammation (Pirhonen et al. 1999, 2001).

If viral infection cannot be resolved through antiviral and inflammatory immune responses, programmed cell death, apoptosis, is activated to eliminate the infected cell (Lamkanfi and Dixit 2010). Apoptosis is a caspase-dependent, non-inflammatory form of programmed cell death (Zimmermann et al. 2001, Ting et al. 2008). It can be initiated intracellularly by the release of cytochrome c or other apoptogenic proteins from mitochondrial intermembrane space into the cytosol or via cell-death receptors on cell surface. Initiation of apoptotic events results in the activation of several apoptotic caspases.

Caspases are the most central effector proteins activated during apoptosis. They are a group of cysteine proteases that are synthetized as inactive zymogens and can be activated by proteolytic cleavage of the protein (Crawford and Wells 2011). Caspases mediate their effects via aspartate-specific cleavage of their target proteins. Caspase-3 is one of the central molecules in apoptosis having several hundreds of known target proteins, and its activation is often held as a hallmark for apoptosis. In addition to caspase-3, several other caspases (caspase-2, -6, -7, -8, -9 and -10) are involved in apoptotic signalling. In addition to the apoptotic caspases, the human caspase family also includes the inflammatory caspases-1, -4 and -5 as well as caspase-14 involved in cellular remodelling.
Programmed cell death is an innate immune response that host cells can use to inhibit viral replication and thus to prevent the spread of virus in the infected organism (Best 2008, Lamkanfi and Dixit 2010). However, several viruses have evolved mechanisms to interfere with host cell death pathways (Lamkanfi and Dixit 2010, Kaminskyy and Zhivotovsky 2010). Some viruses have found ways to modulate the activity of caspases, central molecules in apoptosis. Other viruses, such as some herpes viruses, can inhibit apoptosis by encoding proteins that are homologous to cellular anti-apoptotic proteins of Bcl-2 family. Finally, apoptosis is not always beneficial for host. HIV-1 virus, for example, triggers apoptosis in infected host immune cells like dendritic cells and macrophages hindering development of proper immune responses (Kaminskyy and Zhivotovsky 2010).

In addition to apoptosis, necrosis and pyroptosis can also be considered as forms of programmed cell death (Lamkanfi and Dixit 2010). However, the mechanisms related with these pro-inflammatory modes of cell death are still rather unclear.

### 1.3.3 Influenza A virus

Influenza A viruses are negative-stranded RNA viruses belonging to the *Orthomyxovirus* family. They are highly pathogenic respiratory viruses capable of infecting avian and mammalian species. Annual epidemics of influenza A virus cause severe illnesses in millions of people worldwide. The severity of the infections varies from mild symptoms to severe illness and even death. During the last century, influenza A viruses have triggered four pandemics causing morbidity and mortality around the world.

Influenza A virus genome consists of 8 RNA segments that encode 11 distinct proteins (Ludwig et al. 2003) (Figure 8). Each of these proteins has been studied extensively to elucidate their roles in influenza A virus pathogeneity. NS1 protein of influenza A virus has been often associated with viruses pathogenicity. This protein interferes with RIG-I-mediated type I IFN production in many different ways. NS1 protein, for example, blocks RIG-I ubiquitination which is important for RIG-I/MAVS interaction (Gack et al. 2009). Another example of influenza A virus protein interfering with host immune system is matrix protein 2 (M2) which blocks autophagosome fusion in the infected cells (Gannagé et al. 2009). Finally, extensive interactions between influenza A virus and host cell proteins are likely to affect the
consequences of infection. Influenza-induced innate immune responses in host cells can be triggered by endosomal TLR3 and TLR7 recognizing viral dsRNA and ssRNA as well as by RIG-I recognizing viruses cytosolic RNA (Wu et al. 2011). The virus replicates in host cell nucleus and for replication, it has to hijack several host nuclear factors (Josset et al. 2008, König et al. 2010, Karlas et al. 2010). Therefore, several influenza A virus proteins interact with host cell nuclear machinery affecting for example nuclear structure and host cell splicing machinery. Several other interactions between influenza A virus and host cells have also been reported and they might have an impact on the state of host cells after viral infection (Shapira et al. 2009).

Figure 8. Structure of influenza A virus particle. PB1 = polymerase basic protein 1, PB2 = polymerase basic protein 2, PA = polymerase acidic protein. Viruses NS1 protein is important for replication but is not included in virus particles. (Adjusted from Ludwig et al. 2003)

Only few studies have utilized proteomic methods to study influenza A virus-induced changes in host cell proteomes. In 2006, Baas et al. utilized MS-based proteomics to study influenza A virus-induced changes in macaque lung tissues resulting in 3548 protein identifications (Baas et al. 2006). Although the proteomic study was not quantitative, comparisons between proteomic data and parallel mRNA-level studies indicated inconsistencies between protein and mRNA levels of some proteins in the lung tissues. This shows the importance of protein level data in studies of cells physiological state. Later on, few 2-DE-based (Liu et al. 2008, Vester et al. 2009, van Diepen et al. 2010) and MS-based (Coombs et al. 2010, Emmott et al. 2010a) quantitative proteomic studies of host responses against influenza virus infection have been performed. However, most of these studies have resulted in the identification of only few differentially expressed proteins and the large-scale analyses about influenza-host-interplay have been performed using genome-wide screening techniques (e.g. Hao et al. 2008, Shapira et
Therefore, more in-depth quantitative proteomic analyses and data interpretation would be needed to find cellular signalling pathways that are potentially activated by influenza A virus infection. Successful proteomic studies of host-virus interplay have already been published for other viruses showing the potential of this technique (Emmott et al. 2010b, Naji et al. 2012).
2. AIMS OF THE STUDY

Macrophages and keratinocytes have an important role in the activation innate immune responses after viral infection. The aim of this study was to develop and utilize proteomic and bioinformatic methods to characterize host responses to viral infection. The more detailed aims were:

- To characterize cytosolic viral RNA-induced innate immune responses in HaCaT keratinocytes (I)
- To set up a quantitative subcellular proteomics workflow and to utilize it for global characterization of influenza A virus-induced changes in human primary macrophages (II)
- To study virus-induced protein secretion from human primary macrophages (II, III)
- To develop a computational tool to help the comparison and analysis of protein identification results from different database search engines (IV)
- To develop a computational tool for large-scale caspase cleavage site predictions and to utilize it for high-throughput mapping of potential caspase targets based on mass spectral data (V)
3. MATERIALS AND METHODS

3.1 CELLS AND STIMULATIONS

In this project, two different types of human cells, HaCaT keratinocytes (I, IV) and human primary macrophages (II, III), were used. HaCaT keratinocytes (American Type Culture Collection) were cultured in DMEM (Dulbecco’s Modified Eagle Medium) supplemented with 10% Fetal Calf Serum, L-glutamate and antibiotics (I, IV). Human primary macrophages were differentiated from blood monocytes obtained from leukocyte-rich buffy coats of healthy blood donors (Pirhonen et al. 1999). Differentiation was done by maintaining the monocytes in Macrophage serum-free medium supplemented with 10 ng/ml of Granulocyte-macrophage colony-stimulating factor (GM-CSF) and antibiotics. After five days of culturing the macrophages were used in the experiments. Each macrophage sample was a pool of separately cultured and stimulated cells from three different blood donors.

To study viral RNA triggered immune responses in macrophages and HaCaT keratinocytes, cells were transfected with polyinosic-polycytidylic acid (polyI:C) (I, III) or infected with influenza A virus (II). PolyI:C is a mimic of dsRNA and it has been used to mimic RNA virus infections. Transfections were done using 10 μg/ml of polyI:C (Sigma-Aldrich). Lipofectamin 2000 (Invitrogen) was used as a transfection reagent. PolyI:C transfected cells were studied at different timepoints between 1h and 18h. For viral infections of human macrophages, human pathogenic influenza A virus strain Udorn/72/H3N2 was used with viral dose of 2,56 hemagglutination U/ml. The cells were studied at 6h, 9h, 12h and 18h post-infection. In addition, other RNA viruses, influenza A virus strain Beijing/353/89/H3N2 (II), vesicular stomatitis virus (I, III) and encephalomyocarditis virus (I, III) were used for infections to confirm parts of the results.

3.2 SUBCELLULAR FRACTIONATION AND SECRETOME ANALYSIS

For subcellular fractionations, approximately 10 million cells were used. Mitochondrial and cytoplasmic fractions of HaCaT keratinocytes (I, IV) and macrophages (II) were isolated by QProteome Mitochondria Isolation Kit (Qiagen). After isolation, cytoplasmic fractions were further purified using 2-D Clean-Up Kit (GE Healthcare). Nuclear fractions of macrophages
were isolated using QProteome Nuclear Protein Isolation Kit (Qiagen) (II). The resulting soluble and insoluble nuclear protein fractions were combined before analysis. The enrichment of mitochondrial, cytoplasmic and nuclear proteins in corresponding fractions was confirmed using Western blots (II).

To study influenza A virus or polyI:C induced changes in protein secretion, macrophage growth media were collected and analyzed (II, III). The cells grown in complete Macrophage-SFM medium were washed three times with PBS after which the cells were stimulated in RPMI growth media supplemented with 1 mM HEPES, L-glutamine and antibiotics (GIBCO). The growth media were collected and concentrated with Amicon Ultra centrifugal filter devices (Millipore) with 10 000 nominal molecular weight cutoff. The concentrated media were either used directly for western blot analyses or purified with 2-D Clean-Up Kit (GE Healthcare) for proteomic analyses.

3.3 GEL-BASED METHODS USED IN PROTEOMIC EXPERIMENTS

Two-dimensional gel electrophoresis (2-DE) was used for the separation of mitochondrial and cytoplasmic protein fractions of HaCaT keratinocytes (I). 11 cm pl 4-7 IPG-strips (Bio-Rad) were used as the first dimension and Criterion Tris-HCl 8–16% precast gels (Bio-Rad) as the second dimension. The gels were stained using SYPRO Ruby protein gel stain (Bio-Rad or Sigma-Aldrich) according to manufacturer’s instructions. Spot detection, matching and intensity-based quantitation were done using Image Master 2D Platinum version 6.0 (GE Healthcare). Spots with at least 2-fold difference in expression between control and polyI:C transfected samples were considered differentially expressed and were picked for mass spectral analysis. Finally, protein spots in the gels were visualized using silver staining (O’Connel et al. 1997).

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was used for protein separation in the analysis of polyI:C transfected macrophage growth media (III) and mitochondrial protein fraction of HaCaT keratinocytes (IV). In addition, SDS-PAGE was used as a protein separation method in western blot analyses (I, II, III). With intracellular fractions, equal amounts of protein were loaded on the gel. For secretomes, equal amounts of media were taken for the analyses.
3.4 QUANTITATIVE ANALYSIS USING iTRAQ

4plex iTRAQ (AB Sciex) labeling was used for relative quantitation of proteins in influenza A virus infected macrophages (II). Changes in mitochondrial, cytoplasmic and nuclear proteomes and secretomes of influenza A virus infected macrophages were studied as a function of time. The infected cells were studied at three different timepoints and protein amounts in these samples were compared with uninfected control cells (Figure 9). With intracellular fractions, equal amounts of protein from each sample were taken for the analyses based on silver stained gels. For secretome analyses, equal amounts of cells were taken for the analyses and the whole samples were labeled. Cysteine reduction, alkylation and protein in-solution digestion was done for each sample followed by iTRAQ labeling of the resulting peptides. Digestion and labeling were done according to manufacturer’s instructions.

![Figure 9](image.png)

**Figure 9.** Labeling of the samples for iTRAQ analyses. Two biological and two technical replicates of each sample were analyzed.

After labeling, the peptide mixtures were prefractonated by SCX. The SCX separations were performed with Ettan HPLC system (Amersham Biosciences) using a PolySULFOETHYL A column (200 x 2,1 mm, PolyLC). The LC was operated at 0,2 ml/min and 20 mM KH₂PO₄ buffer (pH 3) was used with a gradient of 0-0,4 M KCl in 35 min. The eluting sample was collected in 1 min fractions and the fractions containing peptides were analyzed using nanoLC-ESI-MS/MS.
3.5 MASS SPECTROMETRY

NanoLC-ESI-MS/MS analyses of tryptic peptides were performed with Ultimate 3000 nanoLC system (Dionex) combined with QStar Elite hybrid quadrupole time-of-flight mass spectrometer (AB Sciex) (I, II, III, IV). In the analyses, the sample was first injected into ProteCol C18 trapping column (0,15x10mm; 3 μm; 120Å)(SGE) and the column was washed with 0.1% trifluoroacetic acid (TFA) for 10 min. After that, the peptides were separated in a PepMap100 C18 analytical column (0,075x150mm; 5μm; 100Å) (LC Packings/Dionex) using a linear gradient of 0-40% acetonitrile in 0,1% formic acid. The length of the gradient varied between 20 min and 120 min depending on the sample. Mass spectral analyses were performed on positive ion mode. MS data were acquired using Analyst 2.0 software (AB Sciex) and an information-dependent acquisition method. The method consisted of a 0,5 s TOF-MS survey scan of m/z 400-1400 followed by MS/MS scans of two most abundant ions with charge states +2 to +4. Once an ion was selected for fragmentation, it was put on an exclusion list for 60 s.

MALDI-MS was used for the identification of the most intense protein spots from 2D gels (I). Prior MALDI analyses proteins from each spot were in-gel digested with trypsin and the resulting peptide samples were desalted using μC18 Zip Tips (Millipore). α-cyano-4-hydroxycinnamic acid was used as the crystallization matrix. Mass spectra for MALDI analyses were acquired using an Ultraflex TOF/TOF instrument (Bruker Daltonik) in positive ion reflector mode. The collected spectra were processed using FlexAnalysis version 3.0 (Bruker Daltonik).

3.6 DATABASE SEARCHES

Two different database search engines, Mascot (Matrix Science) (PMF: publicly available Mascot; MS/MS: in-house Mascot version 2.2) (I, III, IV, V) and Paragon from ProteinPilot version 2.0.1 (AB Sciex) (II, IV), were used for protein identifications based on MS data. Searches were done against NCBI (I, IV) or UniProt/SwissProt (II, III, IV, V) databases. Similar search parameters were used in all the searches. False discovery rates for the identifications were calculated using a target-decoy strategy based on searches against concatenated normal and reversed protein sequence databases (Elias & Gygi 2007).
For iTRAQ-based quantitative analyses, only Paragon was used (II). Raw MS data from both technical replicates of an iTRAQ sample set were processed together to improve the quality of quantitation data. Quality of quantitations was also improved by manually excluding all the MS/MS spectra with all reporter ion peak heights below 10 counts from quantitation. Finally, proteins with more than 1.5-fold difference (intracellular fractions) or more than 3-fold difference (secretomes) between control and infected sample were considered differentially expressed.

3.7 PROTEIN CLASSIFICATION, INTERACTION NETWORKS AND CLUSTERING ANALYSIS

After retrieving a list of protein identifications (and quantitation data) from a database search engine, additional data analysis steps were performed to help interpreting the results. Three different types of data analyses were done: protein classification based on their functions and localizations, protein-protein interaction network studies and expression profile-based clustering analyses. Functional classification and subcellular localization studies of proteins were done mainly based on their Gene Ontology annotations using GeneTrail (Backes et al. 2007) (II). Protein expression profiles were studied and clustering of similarly behaving proteins was performed using Chipster (Kallio et al. 2011). Protein-protein interaction networks for interesting groups of proteins were created using String (Jensen et al. 2009) (II).

In more detailed analyses of the published secretome data (II, III), different forms of protein secretion were studied using ExoCarta (Mathivanan et al. 2012) and SignalP (Petersen et al. 2011). ExoCarta is a manually curated database of exosomal proteins, lipids and RNA (Mathaiavan et al. 2009, 2012). Protein entries in ExoCarta database were compared with protein identifications from the secretomes of human primary macrophages (II, III) to detect potentially exosomal proteins from our datasets. SignalP is a prediction tool for signal peptides based on protein sequences (Nielsen et al. 1997, Petersen et al. 2011). It was used to find potential signal peptide containing proteins from our secretome data (II, III) and thus to estimate the contribution of classical, signal peptide-based protein secretion in our secretome datasets.
3.8 IMMUNOLOGICAL ANALYSES

Immunological methods used in this study are listed in Table 3 and introduced in more detail in the original publications.

Table 3. Immunological methods used in the publications.

<table>
<thead>
<tr>
<th>Method</th>
<th>Publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>APO Percentage apoptosis assay</td>
<td>II, III</td>
</tr>
<tr>
<td>cathepsin B inhibition</td>
<td>II, III</td>
</tr>
<tr>
<td>cathepsin D siRNA</td>
<td>III</td>
</tr>
<tr>
<td>confocal microscopy</td>
<td>I</td>
</tr>
<tr>
<td>ELISA</td>
<td>I, II, III</td>
</tr>
<tr>
<td>P2X7 receptor inhibition</td>
<td>II</td>
</tr>
<tr>
<td>P2X7 receptor siRNA</td>
<td>II</td>
</tr>
<tr>
<td>quantitative RT-PCR</td>
<td>I, III</td>
</tr>
<tr>
<td>src tyrosine kinase inhibition</td>
<td>II</td>
</tr>
<tr>
<td>western blotting</td>
<td>I, II, III</td>
</tr>
</tbody>
</table>

3.9 REAGENTS

Several different antibodies were used to verify the proteomic data and to further characterize the innate immune responses in the experiments. All the antibodies used in the publications are listed in Table 4.
<table>
<thead>
<tr>
<th>Antibody</th>
<th>Manufacturer</th>
<th>Publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>14-3-3 σ</td>
<td>Santa Cruz Biotechnology</td>
<td>I</td>
</tr>
<tr>
<td>α/β-tubulin</td>
<td>Cell Signaling</td>
<td>I</td>
</tr>
<tr>
<td>β-amyloid</td>
<td>Santa Cruz Biotechnology</td>
<td>II</td>
</tr>
<tr>
<td>Actin</td>
<td>Santa Cruz Biotechnology</td>
<td>II</td>
</tr>
<tr>
<td>Annexin A1</td>
<td>Santa Cruz Biotechnology</td>
<td>II</td>
</tr>
<tr>
<td>APOE</td>
<td>Santa Cruz Biotechnology</td>
<td>II</td>
</tr>
<tr>
<td>ASC</td>
<td>Millipore</td>
<td>III</td>
</tr>
<tr>
<td>Bid</td>
<td>Cell Signaling</td>
<td>III</td>
</tr>
<tr>
<td>Caspase-1 p10</td>
<td>Santa Cruz Biotechnology</td>
<td>II, III</td>
</tr>
<tr>
<td>Caspase-1 p20</td>
<td>Sigma-Aldrich</td>
<td>III</td>
</tr>
<tr>
<td>Caspase-3 (H-277)</td>
<td>Santa Cruz Biotechnology</td>
<td>I</td>
</tr>
<tr>
<td>Caspase-3, cleaved (#9661)</td>
<td>Cell Signaling</td>
<td>II, III</td>
</tr>
<tr>
<td>Cathepsin B</td>
<td>Calbiochem</td>
<td>II, III</td>
</tr>
<tr>
<td>Cathepsin D</td>
<td>Santa Cruz Biotechnology</td>
<td>II, III</td>
</tr>
<tr>
<td>Cathepsin Z</td>
<td>Santa Cruz Biotechnology</td>
<td>II</td>
</tr>
<tr>
<td>Cytochrome c</td>
<td>Santa Cruz Biotechnology</td>
<td>II</td>
</tr>
<tr>
<td>Cytokeratin 18</td>
<td>Santa Cruz Biotechnology</td>
<td>I</td>
</tr>
<tr>
<td>eIF4A</td>
<td>Santa Cruz Biotechnology</td>
<td>I</td>
</tr>
<tr>
<td>Galectin-3</td>
<td>Santa Cruz Biotechnology</td>
<td>II</td>
</tr>
<tr>
<td>GAPDH</td>
<td>Santa Cruz Biotechnology</td>
<td>II</td>
</tr>
<tr>
<td>Histone H1</td>
<td>Santa Cruz Biotechnology</td>
<td>II</td>
</tr>
<tr>
<td>HMGB1</td>
<td>Santa Cruz Biotechnology</td>
<td>II</td>
</tr>
<tr>
<td>HSP27</td>
<td>Santa Cruz Biotechnology</td>
<td>I</td>
</tr>
<tr>
<td>HSP90</td>
<td>Cell Signaling</td>
<td>I, II</td>
</tr>
<tr>
<td>IFIT3</td>
<td>BD Transduction Laboratories</td>
<td>II</td>
</tr>
<tr>
<td>IL-18</td>
<td>Pirhonen et al. 1999</td>
<td>II, III</td>
</tr>
<tr>
<td>Influenza A virus (H3N2)</td>
<td>provided by prof. Ilkka Julkunen</td>
<td>II</td>
</tr>
<tr>
<td>LAMP-1</td>
<td>Santa Cruz Biotechnology</td>
<td>II</td>
</tr>
<tr>
<td>P2X7, receptor</td>
<td>Santa Cruz Biotechnology</td>
<td>II</td>
</tr>
<tr>
<td>pan14-3-3</td>
<td>Santa Cruz Biotechnology</td>
<td>I</td>
</tr>
<tr>
<td>phospho-Ser 14-3-3 binding motif</td>
<td>Cell Signaling</td>
<td>I</td>
</tr>
<tr>
<td>S100-A9</td>
<td>Santa Cruz Biotechnology</td>
<td>II</td>
</tr>
<tr>
<td>VDAC1</td>
<td>Santa Cruz Biotechnology</td>
<td>II</td>
</tr>
</tbody>
</table>
4. RESULTS

4.1 PROTEOMICS IS AN EFFICIENT METHOD TO STUDY INNATE IMMUNE RESPONSES IN HUMAN KERATINOCYTES AND MACROPHAGES

Combination of proteomics and bioinformatics has emerged as a valuable approach to study cellular signaling mechanisms (Choudhary et al. 2010). In this project, three different proteomic methods were utilized to study virus-induced changes in human primary macrophages and HaCaT keratinocytes (Figure 10).

In our simplest proteomic approach, we used SDS-PAGE combined with LC-MS/MS analyses to study secretomes of polyI:C transfected human primary macrophages (III). PolyI:C is a mimic of viral dsRNA and can be used to model viral infections. This study resulted in the
identification of 595 distinct proteins secreted from human primary macrophages. Although the method was not quantitative, clear differences between protein identifications from the secretomes of Lipofectamine-treated control cells and polyI:C transfected cells were detected. For example, several Ras-related proteins and vacuolar ATPases were identified only from the secretomes of polyI:C transfected cells indicating that the detection of cytoplasmic dsRNA affects vesicular trafficking in host cells. In addition, the secretion of cysteine protease inhibitors cystatin-A and B were detected only from the secretomes of polyI:C transfected macrophages.

The effects of cytoplasmic dsRNA recognition on intracellular host proteomes was studied using quantitative 2-DE combined with mass spectrometry (I). Mitochondrial and cytoplasmic protein fractions of untreated and polyI:C transfected human HaCaT keratinocytes were separated using 2-DE and differential expression of proteins was studied based on these gels. This study resulted in protein identifications from 176 differentially expressed protein spots. The value of subcellular fractionation was clearly seen since only 15% of protein identifications were common for mitochondrial and cytoplasmic fractions. Comparisons of 2D gels from control and transfected cells showed clear polyI:C-induced changes in 14-3-3 protein expression. In addition, 2D gel analysis showed that polyI:C transfection triggered significant fragmentation of cytoskeletal proteins such as cytokeratins. Although similar cytoskeleton fragmentation has probably been present in influenza A virus infected human primary macrophages (Öhman et al. 2009) it could not be detect in our iTRAQ analyses using non-gel-based separation methods.

In publication (II), we used non-gel-based quantitative subcellular proteomics to study influenza A virus-induced changes in mitochondrial, cytoplasmic and nuclear fractions and secretomes of human primary macrophages. Our quantitative subcellular proteomic experiments resulted in altogether 3477 protein identifications including 1321 differentially expressed proteins in the intracellular fractions and 544 upregulated proteins in the secretomes. Non-gel-based proteomics is considered to be the most universal method for proteomic analyses. Accordingly, comparisons between protein identifications from mitochondrial and cytoplasmic fractions in publications (I) and (II) show that 22% of proteins identified in the non-gel-based experiment fall outside the pI and molecular weight range of our 2D gels and would therefore not have been identified using 2-DE. On the other hand, comparisons of proteomic data from SDS-PAGE (III) and iTRAQ (II) secretome experiments showed that these two approaches give comparable results and could be used to support and complement each other. Bioinformatics analyses of our
iTRAQ data indicated that influenza A virus infection results in extensive changes in host cell nuclear proteomes and distracts the mitochondrial structure. Finally, our data analyses revealed several signs about activation of antiviral responses, apoptosis and various inflammatory mechanisms in influenza A virus infected cells.

In conclusion, our experiments show that quantitative subcellular proteomics and secretome analysis is an efficient approach to study virus-induced protein-level changes in host cells. Spatial (I, II, III) and temporal (II, III) information about multiple proteins can be collected in a single experiment. Finally, analysis of proteomic data using various bioinformatics tools helps creating hypotheses about biological mechanisms activated as a result of viral infection (I, II).

4.2 MASCOT AND PARAGON GIVE COMPARABLE PROTEIN IDENTIFICATION RESULTS

In MS-based proteomic experiments, protein identification and quantitation relies on database search engines. Several different database search engines have been developed for proteomic purposes, providing both confirmatory and complementary information (Searle et al. 2008, Yu et al. 2010). Two different database search engines, Mascot (I, III, V) and Paragon (II), were used in our proteomic experiments. A computational tool, Compid, was developed to simplify the comparison of Mascot and Paragon protein identification results and it was used to confirm that the results from both database search engines were comparable (IV).

Compid can be used for peptide and protein level comparisons between two sets of raw data from database search engines and also to regroup these protein identification results. Our comparisons of Mascot and Paragon protein identification results with Compid showed that there are some differences in the results from these two database search engines. The biggest differences were detected on peptide level where only 20% of the peptide identifications were common for both search algorithms. Here, the most important reasons for low overlap were probably the low-quality MS/MS spectra resulting in low-confidence peptide spectrum matches. The overlap between Mascot and Paragon protein identifications was significantly better, 41-65%. The differences in protein identifications were partially caused by the more efficient protein grouping algorithm in Paragon. In addition, the size and redundancy of protein sequence database had a clear impact on the overlap between Mascot and Paragon database search results.
However, most of the good quality protein identifications were common for Mascot and Paragon showing that these database search algorithms gave comparable protein identification results.

### 4.3 VIRUS-INDUCED RESPONSES IN HACAT KERATINOCYTES AND HUMAN PRIMARY MACROPHAGES

Our studies showed that both influenza A virus infection and polyI:C transfection trigger significant changes in the proteomes of human primary macrophages (II, III) and HaCaT keratinocytes (I). Activation of antiviral immune responses was seen as the upregulation of several interferon-inducible proteins in influenza A virus infected human primary macrophages (II). Different mechanisms of inflammation, such as inflammasome activation and secretion of inflammatory danger signal molecules, were also detected (II, III). Viral stimulation clearly activated caspase-mediated apoptosis in HaCaT keratinocytes (I) and human primary macrophages (II, III). Finally, influenza A virus and polyI:C triggered significant protein secretion from human primary macrophages (II, III).

#### 4.3.1 Several inflammatory pathways are activated in human primary macrophages

NLRP3 inflammasome is a caspase-1-activating platform that has an important role in inflammatory responses triggered by influenza A virus infection (Allen et al. 2009, Thomas et al. 2009). Although inflammasome activation is currently under extensive research, the exact mechanisms leading to its activation have remained unresolved. Two out of three NLRP3 inflammasome components, caspase-1 and ASC, were identified from the cytoplasmic fraction of our iTRAQ data (II). In addition, caspase-1 activation and secretion of mature IL-18 were detected in influenza A virus infected and polyI:C transfected human primary macrophages indicating virus-induced activation of the inflammasome.

Lysosome destabilization and especially leakage of lysosomal protease cathepsin B has often been associated with NLRP3 inflammasome activation (Hornung et al. 2008, Allen et al. 2009). At 6h post-infection, our proteomic data showed the upregulation of several lysosomal proteins in the cytoplasmic fraction (II). Since this could be a sign of lysosomal rupture in the infected cells (II), we studied whether lysosomal cathepsins are involved in inflammasome activation.
Both ELISA (II, III) and western blot (II) analyses of IL-18 showed that cathepsin B inhibitor Ca-074 Me abolished IL-18 secretion from the stimulated cells almost completely. Additionally, cathepsin D siRNA decreased the secretion of IL-18 from polyI:C transfected cells (III). Finally, western blot analyses showed that secretion of caspase-1 p10 fragment was abolished by cathepsin B inhibition (II) indicating a role for cathepsins in inflammasome activation upstream of caspase-1 activation.

In addition to cytosolic cathepsin B, several other intracellular signals have been suggested to trigger inflammasome activation. To study other possible routes for inflammasome activation, a protein-protein interaction network was created from all inflammation-related proteins identified from the intracellular fractions of influenza A virus infected macrophages (II). This protein-protein interaction network showed that plasma membrane ATP receptor P2X7 is directly associated with all NLRP3 inflammasome components. Therefore, we studied the role of P2X7 receptor in influenza A virus infection triggered inflammasome activation. Inhibition of P2X7 receptor with a specific inhibitor, AZ11645373, as well as P2X7 receptor siRNA experiments resulted in a clear decrease in IL-18 secretion from human primary macrophages implying that P2X7 receptor has a regulatory role in inflammasome activation after influenza A virus infection.

NADPH oxidase complex has also been previously related with NLRP3 inflammasome activity (Dostert et al. 2008). Two subunits of NADPH oxidase complex, CYBA and CYBB, were overexpressed in cytoplasmic fraction of influenza A virus infected human primary macrophages at 6 h post-infection (II). In addition, src tyrosine kinase, an upstream regulator of NADPH complex (Giannoni et al. 2010), was identified from influenza A virus infected macrophages (II). Our functional studies showed that IL-18 secretion from influenza A virus infected macrophages was completely abolished by the inhibition of src tyrosine kinase with PP2 (II). Thus, in addition to cytosolic cathepsin B and P2X7 receptor activity, also src tyrosin kinase activity has a role in the regulation of inflammasome activity in influenza A virus induced human primary macrophages.

In addition to inflammasome activation, also other signs of proinflammatory responses were detected in influenza A virus infected and polyI:C transfected human primary macrophages (II, III). Several inflammatory proteins like macrophage migration inhibitory factor (MIF), S100-A9 and HMGB1 were secreted from influenza A virus infected or polyI:C transfected macrophages (II, III). Finally, multiple inflammation-related proteins such as allograft inflammatory factor,
S100-A8 and S100-A9 were differentially expressed in the intracellular fractions of influenza A virus infected macrophages (II).

### 4.3.2 Viral infection triggers caspase-dependent apoptosis in human macrophages and keratinocytes

Caspase-3 activation is a classical sign of apoptosis. Our studies showed several different signs of caspase-dependent apoptosis in the influenza A virus and polyI:C stimulated macrophages and keratinocytes (I, II, III). Western blot analyses of influenza A virus infected (II) and polyI:C transfected (III) macrophages showed the activation of caspase-3 in stimulated cells. In addition, classical signs of mitochondrial apoptosis: translocation of Bax from cytosol onto mitochondria as well as leakage of cytochrome c from mitochondria into cytosol, were clearly seen in our iTRAQ data (II). Cathepsin B inhibition experiments showed that caspase-3 activation in influenza A virus infected and polyI:C transfected macrophages was largely dependent on cathepsin B activity (II, III). Finally, based on our kinetics experiments, caspase-3 activation was shown to precede inflammasome activation and it occurred already few hours after infection (III).

Initiation of apoptosis causes clear changes in the morphology of dying cells (Crawford and Wells 2011). A clear rearrangement of cytoskeletal proteins was seen in polyI:C transfected HaCaT keratinocytes (I). Several fragments of cytokeratins and other structural proteins were identified from the stimulated cells, especially from the mitochondrial fraction. In addition, polyI:C transfection-induced changes in the subcellular distribution of cytokeratin 18 were seen in the confocal microscopy images of HaCaT keratinocytes. Caspase-3 inhibition experiments showed that the cleavage of cytokeratin-18 was caspase-dependent and therefore related with apoptosis of the transfected cells.

### 4.3.2.1 Prediction and identification of potential caspase cleavage targets from proteomic data

Determination of caspase target proteins can give important insights into the execution mechanisms of programmed cell death. Therefore, a computational tool, Pripper, was developed to predict potential caspase cleavage motifs based on protein sequences and known caspase
cleavage sites (V). Pripper was the first prediction tool capable of processing multiple protein sequences or even entire proteomes simultaneously making it extremely useful for proteomic applications. In addition, Pripper can create caspase cleavage product databases based on the predictions. These databases can be downloaded directly into a database search engine and utilized to identify potential caspase targets from MS-based proteomic experiments in a high-throughput way. Workflow for caspase target prediction based on Pripper and the following database searches is described in Figure 11.

**Figure 11.** Workflow for caspase cleavage site and caspase target prediction in proteomics by using Pripper.

We utilized Pripper to search our proteomic data from publications (I) and (II) for potential caspase cleavage targets. The analysis of MS data from polyI:C transfected HaCaT keratinocytes resulted in the identification of one cytokeratin-18 fragment as a potential caspase cleavage product (V). This was an additional confirmation for caspase-dependent cytokeratin fragmentation in polyI:C transfected HaCaT cells. In addition, Pripper analysis of proteomic data from publication (II) resulted in the identification of several cytoskeleton-associated proteins (e.g. actin) as potential caspase targets (V). This indicates that influenza A virus infection of human primary macrophages also resulted in caspase-dependent cytoskeletal rearrangement, although it was not detected from the proteomic data. Finally, the analysis of our iTRAQ data using a caspase cleavage product database resulted in the identification of a potential novel caspase cleavage motif in leukosialin. However, additional biological experiments are required to confirm these caspase cleavage sites.
**4.3.3 Influenza A virus infection and polyI:C transfection trigger significant protein secretion from human primary macrophages**

Proteomic methods were used to study changes in protein secretion followed by influenza A virus infection and polyI:C transfection of human primary macrophages (II, III). Our analyses showed that both stimulations trigger significant protein secretion. GeneTrail analysis of the identified proteins showed that the secreted proteins are associated with several different cellular processes including gene expression, signaling and cell death (data not shown). Comparison of protein identifications in publication (III) and secretome data in publication (II) showed that only 30% of the protein identifications were common for both experiments. However, the functional roles of proteins identified in both experiments were very similar (Figure 12A). DAMP (damage-associated molecular pattern) proteins are endogenous proteins with various intracellular functions, acting as danger signals when secreted out of cells after stress or injury (Bianchi 2007). Several DAMPs (e.g. S100-A8, S100-A9, Galectin-3, HMGB1, HMGB2) were secreted from polyI:C transfected (III) and especially from influenza A virus infected (II) human primary macrophages. Secretion of several Ras-related proteins as well as other proteins involved in vesicle trafficking was also triggered by both stimulations. Finally, several lysosomal proteins, for example cathepsins and lysosomal hydrolases were identified from the secretomes of influenza A virus infected (II) and polyI:C transfected (III) macrophages.

Classically, proteins are directed for secretion via ER/Golgi-pathway by signal peptides included in protein sequences (Nickel et al. 2010). However, based on SignalP (Petersen et al. 2011) analyses, only 16% of proteins identified from our secretomes contain a signal sequence (Figure 12B). This indicates that virus-induced human macrophages use other secretory routes to direct proteins out of cells. In addition to signal peptide-mediated protein secretion, proteins can be directed out of the cells via unconventional secretion mechanisms including plasma membrane-located channels, budding of plasma membrane microvesicles, exocytosis of secretory lysosomes and release of exosomes from multivesicular bodies (Théry et al. 2009, Nickel et al. 2010). Exosomes have been associated with cell-to-cell signaling related with, for example, immune responses. We used ExoCarta to determine the contribution of potentially exosomal proteins for our secretome data. More than 50% of the identified proteins were included in ExoCarta database (Figure 12B) suggesting that virus infection induces significant protein secretion via exosomes. Additionally, comparison of our secretome data with commonly identified exosomal proteins showed that most of the proteins that are generally identified from exosomes were also present in our data (Figure 13).
Figure 12. Functionally similar proteins are secreted from influenza A virus infected and polyI:C transfected human primary macrophages. A) Protein abundances in selected Gene Ontology classes are shown for secretomes of Influenza A virus- (II) and polyI:C-stimulated (III) macrophages. B) ExoCarta and SignalP classifications of proteins identified from secretomes of influenza A virus infected (II) and polyI:C transfected (III) macrophages.

Figure 13. Proteins that are typically identified from purified exosomes (Adjusted from Théry et al. 2009 and Mathivanan et al. 2010). All the proteins that were identified from secretomes of influenza A virus infected or polyI:C transfected human primary macrophages are marked with bold letters.
5. DISCUSSION

Proteomics is a powerful approach to characterize system-wide changes in protein expression. Several different methods have been developed for protein localization studies (Hartwig et al. 2009, Lee et al. 2010), quantitative analyses (Coombs 2011, Neilson et al. 2011), characterization of posttranslational modifications (Thingholm et al. 2009, Kim et al. 2011, Hjerpe et al. 2009) and mapping of protein-protein interactions (Rees et al. 2011, Li et al. 2011). Methods used in proteomic experiments have developed enormously during the last decades. Mass spectrometry is one of the fastest developing areas of proteomics. The new orbitrap instruments (Makarov 2000, Hu et al. 2005) have increased the amounts of data that can be collected in proteomic experiments significantly and introduction of new fragmentation techniques such as electron transfer dissociation has improved for example the analysis of protein posttranslational modifications (Zubarev et al. 1998, Kim and Pandey 2012). These new developments enable wider applicability of proteomic methods in solving biological questions thereby opening new possibilities for proteomic studies. The expectations for proteomic experiments have therefore constantly increased making it essential for scientists to keep up with the latest inventions. In this project, two different quantitative subcellular proteomic approaches and quantitative and qualitative secretome analyses were utilized to study virus-induced changes in human primary macrophages and HaCaT keratinocytes.

With modern mass spectral techniques, huge amounts of raw data can be collected from a single proteomic experiment making manual interpretation and validation of these data impossible. Therefore, different database search engines are used to identify and quantify proteins based the collected MS data (Nesvizhskii 2007). However, differences in database search algorithms often result in slightly different protein identification results from the same dataset (Elias et al. 2005, Searle et al. 2008, Savijoki et al. 2011). Especially low quality MS/MS spectra with low signal-to-noise ratio, minor peptide fragmentation or overlapping fragments from multiple distinct peptides are often problematic for database search engines (Salmi et al. 2009). The resulting low-confidence peptide-spectrum matches often differ between database search engines and have a higher probability for being false positive identifications. Large differences in low confidence peptide identifications were also seen in our comparisons between Mascot and Paragon. To overcome this problem, several strategies for filtering out low quality MS/MS spectra and low confidence peptide identifications have been developed (e.g. Keller et al. 2002, Helsens et al. 2008, Salmi et al. 2006). Additionally, parallel use of different database search engines can be used to increase the confidence of protein identifications as well as to increase
the number of proteins identified (Yu et al. 2010, Searle et al. 2008). For example, we have utilized Compid to combine Mascot and Paragon protein identifications resulting in a slight increase in the number of protein identifications and therefore higher coverage of *Lactobacillus rhamnosus* proteomes (Savijoki et al. 2011).

After retrieving a list of protein identification, quantitation and modification data, further bioinformatics tools are required to perceive the potentially underlying molecular mechanisms. The repertoire of data analysis tools has expanded enormously and several tools for protein classification, pathway analysis and interaction studies as well as for various specific applications are currently available. In this project, we have used Gene Ontology-based protein classification, protein expression profiling and protein-protein interaction studies to characterize the collected data. Computationally built networks of proteins and their physical and functional relationships are useful when associating proteomic data with cell-wide processes (Goh et al. 2012, Albert 2005). However, one of the problems with proteomics is the incompleteness of the collected data. For example low abundance regulatory proteins present in the samples often remain undetected creating challenges for data analysis. Zubarev et al. have tried to solve this problem with a method called key-node analysis where the presence of undetected regulatory proteins in a pathway can be hypothesized based on the identification of their target proteins (Zubarev et al. 2008). Another challenge in proteomic data analysis is the incompleteness of several biological databases (Goh et al. 2012). Here, the best results can probably be achieved by combinig data from several parallel databases. Finally, the analysis of proteomic data usually results in extremely complex networks that are difficult to interpret. Therefore, the users often focus only on a certain part of the network and filter out the rest of the data limiting the results. Filtering out low quality data and incorporation of different types of data into one network can result in more easily interpretable networks and help, for example, in finding novel associations between protein clusters.

In addition to using general classification and network tools, we created a new tool, Pripper, to study caspase-mediated protein fragmentation. Caspases are a family of proteins involved in the regulation and execution of various immune processes like apoptosis and inflammation (Crawford and Wells 2011). Caspase cleavage of a target protein may result in direct activation or deactivation of the target protein or may alter protein functions by for example affecting their localization. Therefore, identification of caspase targets may give important information about executionary mechanisms in for example apoptosis. Although there are several tools available for the prediction of caspase cleavage sites (Backes et al. 2005, Wee et al. 2009, Ayyash et al.
Pripper was the first tool enabling the direct utilization of the collected MS data for large-scale caspase target predictions. These analyses resulted in the identification of several potential caspase targets, especially structural proteins. Together with our proteomic and immunological data, this indicates that caspases have an important role in virus-induced modulation of cell cytoskeleton.

Combination of proteomic and bioinformatic analyses has been used to study host defense responses against different viruses (Zheng et al. 2011). Influenza A virus is a respiratory virus infecting annually millions of people worldwide. Although influenza A viruses have been studied widely for several decades, only few large-scale proteomic experiments of influenza-induced host responses have been published (Baas et al. 2006, Coombs et al. 2010, Emmott et al. 2010a). Virus infection usually results in the activation of various antiviral and inflammatory pathways and finally death of the infected host cell (Akira et al. 2006, Lamkanfi and Dixit 2010). We detected clear upregulation of several interferon-inducible proteins in the cytoplasmic fraction of influenza A virus infected human primary macrophages showing the activation of antiviral immune responses. Activation of inflammatory responses and caspase-mediated apoptosis were also clearly seen in our data. One of the inflammatory responses triggered by influenza A virus, as well as some other RNA viruses and cytosolic RNA, is the activation of the NLRP3 inflammasome in the infected cells (Kanneganti et al. 2006). This inflammasome activation is crucial for the development of proper innate immune responses after influenza A virus infection (Allen et al. 2009, Thomas et al. 2009). However, the exact mechanisms leading to NLRP3 inflammasome activation are still unknown.

Lysosomal rupture (Hornung et al. 2008, Allen et al. 2009), cytosolic cation imbalance (Qu et al. 2007, Ichinohe et al. 2010) and ROS production (Dostert et al. 2008, Allen et al. 2009) have often been related with inflammasome activation. Our data indicated that influenza A virus infection causes lysosomal damage in influenza A virus infected and polyI:C transfected human primary macrophages. Inhibition of lysosomal cathepsin B has been shown to block inflammasome activation and IL-1β secretion to some extent (Hornung et al. 2008, Allen et al. 2009). This was also seen in our data, since both caspase-1 activation as well as IL-18 cleavage and secretion were clearly blocked in influenza A virus and polyI:C stimulated macrophages treated with cathepsin B inhibitor. Additionally, cathepsin D siRNA decreased the secretion of IL-18 significantly. Cathepsins B and D can also cause significant increase in mitochondrial ROS production (Zhao et al. 2003), providing another possible link between lysosome rupture and inflammasome activation. Additionally, caspase-3 activation in human primary
macrophages was shown to be partially cathepsin dependent indicating that lysosomal rupture might also result in activation of apoptotic pathways at early stages of infection.

Phagocytosis of microbial components and particles results in ROS generation and has been associated with, for example, monosodium urate- and asbestos-triggered inflammasome activation (Dostert et al. 2008). NADPH oxidase mediates ROS production in phagocytosis, and knockdown of NADPH oxidase subunit p22phox (CYBA) has been shown to decrease IL-1β secretion. Src tyrosine kinases regulate NADPH oxidase-mediated ROS generation (Giannoni et al. 2010) and SRC has been shown to localize into dsRNA containing endosomes and associate with TLR3 (Johnsen et al. 2006). In addition, lysosomal protein LAMP-1 was localized in these endosomes suggesting a potential link between src tyrosine kinase activity and lysosomes. We showed that inhibition of src tyrosine kinase blocked IL-18 secretion from influenza A virus infected human primary macrophages indicating that src tyrosine kinase- and NADPH oxidase-mediated ROS has an important role in NLRP3 inflammasome activation. A recent publication showed that src tyrosine kinase might regulate extracellular vesicle formation (Choi et al. 2012) indicating that src tyrosine kinase-mediated inflammasome activation might also be involved in exosomal protein secretion. Finally, ROS-dependent interaction between thioredoxin interacting protein TXNIP and NLRP3 has been shown to result in NLRP3 inflammasome activation creating a potential link between ROS generation and inflammasome activation (Zhou et al. 2010).

Additionally, activation of P2X7 receptor by ATP has been associated with ROS production, and it also increases K+ efflux out of the cells (Pfeiffer et al. 2007). In our studies, inhibition of P2X7 receptor as well as silencing of P2X7 receptor expression decreased the amount of IL-18 secretion from influenza A virus infected human primary macrophages. Similar signs of of P2X7 receptor-related inflammasome activation after influenza A virus infection have also been shown previously (Ichinohe et al. 2010). Influenza A viruses M2 ion channel has an important role in influenza-induced inflammasome activation. M2 ion channel is capable of transporting H+ out of the vesicles in trans-Golgi network affecting the intracellular cation balance. These data indicate that also intracellular cation imbalance might have an important role in inflammasome activation.

Extensive inflammatory responses can be destructive for host organism. A recent publication showed that NLRP3 inflammasome activity can be regulated via autophagosomes (Shi et al.
Autophagy is a specialized system for degradation of intracellular substances. Inflammasome activation triggers autophagosome formation via RalB protein activation. The formed autophagosomes limit inflammasome activity by engulfing inflammasomes. Proteins and organelles that are sequestered into autophagosomes can be delivered to lysosomes for degradation (Behrends et al. 2010) or autophagosomes can fuse with the plasma membrane and release their contents into extracellular space in membrane-enclosed exosomes (Manjithaya and Subramani 2011). In a recent study, proteomics was used to study autophagosome-associated proteins in human breast cancer cells after various stimuli (Dengjel et al. 2012). 80 proteins identified in this study were also present in our secretome datasets implying that autophagosome-related protein secretion is one of the secretory mechanisms activated in human primary macrophages after viral infection. Based on current data, inflammasome activating and regulating processes are speculated in Figure 14.

NLRP3 inflammasome activity has been related with unconventional protein secretion (Qu et al. 2007, Keller et al. 2008). Keller et al. have shown that caspase-1 activation affects the secretion of several proteins, for example galectin-3 and peroxiredoxin-1, that are secreted via nonclassical route (Keller et al. 2008). In addition, P2X7 receptor can stimulate exosome release and IL-1β secretion in an inflammasome-depentent manner (Qu et al. 2007). Finally, inflammasome components as well as active forms of IL-1β and IL-18 are secreted from activated cells via nonclassical mechanisms (Martinon et al. 2002, Qu et al. 2007). Microvesicles have been shown to transfer functional, signaling competent receptors from one cell to another (Mack et al. 2000). Exosomes are probably the most studied group of microvesicles that can be used to transmit signals between cells. Our secretome data showed that several exosome-related proteins, lysosomal proteins and vesicle-related proteins as well as inflammasome components Caspase-1 and ASC are secreted from influenza A virus infected and polyI:C transfected human primary macrophages. The unconventionally secreted proteins also included several DAMPs, for example galectin-3, S100-A8 and S100-A9. This indicates that inflammasome activation-related unconventional protein secretion and especially secretion of inflammasome components might have an important role in cell-to-cell signaling after viral infection.
Figure 14. Cellular processes potentially related with influenza A virus-induced inflammasome activity. 1) Src tyrosine kinase- and NADPH oxidase-mediated ROS is involved in inflammasome activation during phagocytosis. 2) Lysosomal rupture and cytosolic cathepsins B and D are associated with ROS and inflammasome activation. 3) P2X7 receptor activation results in K+ efflux and participates in inflammasome activation. 4) Influenza A virus protein M2 facilitates H+ efflux from Golgi to cytosol activating the inflammasome. 5) Inflammasome activation triggers autophagocytosis of intracellular proteins (e.g. inflammasome components) therefore inhibiting inflammasome activation. Autophagocytosis might be followed by either degradation of inflammasome components or by their secretion from autophagosomes.
6. CONCLUSIONS AND FUTURE PERSPECTIVES

In this project we have combined proteomics, bioinformatics and immunological studies to characterize virus-induced changes in human primary macrophages and HaCaT keratinocytes. We show that influenza A virus infection triggers significant changes in intracellular proteomes of human primary macrophages. The production of several antiviral proteins is initiated as a result of infection. In addition, several inflammatory pathways, especially NLRP3 inflammasome activation, are activated in influenza A virus infected and polyI:C transfected human primary macrophages. Inflammasome activation in human macrophages was regulated by cathepsins, src tyrosine kinase and P2X7 receptor activities and was followed by the initiation of apoptotic events only few hours post stimulation. Caspase-3-dependent apoptosis was detected in polyI:C transfected HaCaT keratinocytes as well as in influenza A virus infected and polyI:C transfected human primary macrophages. Caspase activation resulted in, for example, cytokeratin-18 cleavage and changes in 14-3-3 protein expression in HaCaT keratinocytes. Finally, we show that influenza A virus infection and polyI:C transfection trigger extensive secretion of various different proteins such as inflammatory, vesicle-related and danger signal proteins from human primary macrophages which might have an important role in cell-to-cell signaling after viral infection.

In conclusion, our studies show that quantitative subcellular proteomics and secretome analysis is a powerful tool for system-wide studies of cellular antiviral defense responses. Cellular signaling events are often regulated by protein posttranslational modifications such as phosphorylation and ubiquitination and especially protein ubiquitination has been recently associated with various innate immune signaling events. Therefore, large-scale studies of protein phosphorylation and ubiquitination could give deeper insights into regulation of host cell innate immune responses. Finally, to be able to utilize the full potential of large-scale proteomic studies, close collaboration between proteomics, bioinformatics and biology experts is required.
ACKNOWLEDGEMENTS

This project was carried out at the Institute of Biotechnology, University of Helsinki in close collaboration with the Unit of Excellence in Immunotoxicology, Finnish Institute of Occupational Health and the Department of Information Technology, University of Turku. The Institute of Biotechnology is acknowledged for providing excellent facilities for the research. The work was funded by the Academy of Finland and the Helsinki Graduate Program in Biotechnology and Molecular Biology (GPBM). GPBM is also thanked for providing a variety of interesting courses for PhD studies.

I am extremely grateful to my supervisors Docent Tuula Nyman and Docent Sampsa Matikainen for all the guidance and help during these years. It has been a pleasure to work under your supervision. Your research enthusiasm and expertise have been important for the progress of this project.

I would like to thank the reviewers of this thesis, Docent Jaana Vesterinen and Docent Sampsa Hautaniemi, for their valuable comments. Docent Jaana Vesterinen and Professor Juho Rousu are thanked for the advises and discussions in the thesis committee meetings.

I am grateful to Docent Nisse Kalkkinen, the former head of the Protein Chemistry Research Group, for the opportunity to work in his well-equipped laboratory.

I express my gratitude to all the coauthors of the publications. I would especially like to thank Tiina Öhman: I couldn’t have hoped for a better teacher and lab partner. I am also thankful to Johanna Kerminen for all the hard work in immunology as well as all the encouraging words. Olli Nevalainen and Jussi Salmi are thanked for the fruitful collaboration and for sharing their expertise in bioinformatics with us. Tero Aittokallio is thanked for the valuable help in the analysis of complex proteomic data.

I would also like to thank all the former and current colleagues in the Protein Chemistry Group and at FIOH for creating such a pleasant working environment. I would especially like to thank the other occupants of “Lastenhuone”, Pia, Juho and Wojciech, for all the humour and refreshing discussions.

Finally, I would like to thank my family, friends and Kalle for their love, patience and support.

Helsinki, April 2012
REFERENCES


51
Bianchi ME (2007) DAMPs, PAMPs and alarmins: all we need to know about danger. *J. Leukoc. Biol.* 81, 1-5.


63


