



# LUND UNIVERSITY

## A holistic approach to host-pathogen interactions

### Detecting the large to unravel the small

Sjöholm, Kristoffer

2017

*Document Version:*

Publisher's PDF, also known as Version of record

[Link to publication](#)

*Citation for published version (APA):*

Sjöholm, K. (2017). *A holistic approach to host-pathogen interactions: Detecting the large to unravel the small* (1 ed.). [Doctoral Thesis (compilation), Department of Immunotechnology]. Department of Immunotechnology, Lund University.

*Total number of authors:*

1

#### General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117  
221 00 Lund  
+46 46-222 00 00

# A holistic approach to host-pathogen interactions

Detecting the large to unravel the small

Kristoffer Sjöholm



**LUND**  
UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculty of Engineering, Lund University, Sweden.

To be defended at Belfrage lecture hall, Klinikgatan 32, Lund.

Friday June 2<sup>nd</sup> at 09:15.

*Faculty opponent*

Prof. Dr. Bernd Wollscheid,

Department of Health Sciences and Technology,

ETH Zurich, Zurich, Switzerland

LUND UNIVERSITY  Faculty of Engineering Dep. of Immunotechnology Ideon Medicon Village, Bld. 406 22381, Lund, Sweden  <b>Kristoffer Sjöholm</b>	<b>Document name:</b> Doctoral dissertation	
	<b>Date of issue:</b> 2017-06-02	
	<b>Sponsoring organization:</b>	
<b>Title and subtitle:</b> A holistic approach to host-pathogen interactions - Detecting the large to unravel the small		
<b>Abstract</b> Sepsis is one of the leading causes of mortality and morbidity in the world, and is an overreaction by the immune system due to pathogen invasion of the bloodstream. The interactions, particularly the protein interactions, between the host and pathogen are fundamental for the outcome of the disease. However, some of the protein interactions are unknown and those that are known have been studied as single entities. This thesis focuses on expanding the knowledge of these interactions. In addition, the thesis emphasises the analysis of every interaction at the same time. The identification and quantification of the whole network of interactions is central to determine the impact of single proteins, since the interactions can be co-dependent. The proteins of the human immune system are of particular interest since the pathogens have to avoid the immune system in order to survive in the host. Using state of the art techniques, the human plasma interaction proteome for several pathogens were determined and several novel interactions were discovered. For the gram-positive bacterium <i>Streptococcus pyogenes</i> , the interaction proteome was determined with higher accuracy than ever before. In conclusion, when focusing on the whole interaction pattern the specific details can be explained. The wider understanding of the field of host-pathogen interactions that is established with this thesis aids in future development of diagnostics and therapeutics in pathogen related diseases.		
<b>Key words:</b> Mass spectrometry, host-pathogen interaction, sepsis, proteomics		
<b>Classification system and/or index terms (if any):</b>		
<b>Supplementary bibliographical information:</b>		<b>Language:</b> English
<b>ISSN and key title:</b>		<b>ISBN:</b> 978-91-7753-234-7
<b>Recipient's notes:</b>	<b>Number of pages:</b> 138	<b>Price:</b> -
	<b>Security classification:</b>	

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature:



Date: 2017 – 04 – 24

# A holistic approach to host-pathogen interactions

Detecting the large to unravel the small

Kristoffer Sjöholm



**LUND**  
UNIVERSITY



Cover by Kristoffer Sjöholm

Copyright Kristoffer Sjöholm

Lund University  
Faculty of Engineering  
Department of Immunotechnology

ISBN: 978-91-7753-234-7

ISBN: 978-91-7753-235-4 (electronic)

Printed at Tryckeriet i E-huset, Lund University  
Lund 2017

*To Sara*

# Content

<b>Original papers</b>	<b>8</b>
<b>My contribution to the papers</b>	<b>9</b>
<b>Abbreviations</b>	<b>10</b>
<b>Aim</b>	<b>11</b>
<b>Chapter 1 – Introduction to proteomics</b>	<b>13</b>
Bottom-up vs. top-down	14
<b>Mass spectrometry</b>	<b>15</b>
Ionisation source	15
Mass analysers	16
Detector	17
<b>Mass spectrometry based proteomics</b>	<b>17</b>
Tandem MS	18
Peptide Fragmentation	19
Separation	20
<b>Conclusion</b>	<b>21</b>
<b>Chapter 2 – Applying proteomics in biology</b>	<b>23</b>
<b>Quantification</b>	<b>23</b>
Labelling	24
<b>Mass spectrometer setup</b>	<b>26</b>
Data dependent acquisition	26
Selected reaction monitoring	27
Data independent acquisition	28
<b>Data analysis</b>	<b>29</b>
SRM	29
DDA	30
DIA	31
False discovery rate	32
Peptide to protein	32
<b>Issues with bottom-up proteomics</b>	<b>33</b>

<b>Conclusion</b>	<b>34</b>
<b>Sepsis</b>	<b>35</b>
Blood	37
<b>Innate immune system</b>	<b>37</b>
Cells	38
Coagulation system	38
Complement system	38
<b>Adaptive immune system</b>	<b>39</b>
Cells	39
Immunoglobulins	40
<b>Conclusion</b>	<b>41</b>
<b>Chapter 4 – Pathogen and interactions</b>	<b>43</b>
<b>Pathogen surface</b>	<b>43</b>
Surface proteins	44
Surface interactions	45
<i>Streptococcus pyogenes</i>	45
Virulence factors	46
<b>Host-pathogen interaction proteomics</b>	<b>47</b>
Extend the knowledge	48
<b>Conclusion</b>	<b>49</b>
<b>Chapter 5 – A holistic approach</b>	<b>51</b>
<b>Future</b>	<b>53</b>
<b>The Original papers</b>	<b>54</b>
Paper I	54
Paper II	55
Paper III	56
Paper IV	58
<b>Conclusions and relevance</b>	<b>58</b>
<b>Populärvetenskaplig sammanfattning</b>	<b>61</b>
<b>Acknowledgement</b>	<b>63</b>
<b>References</b>	<b>65</b>

## Original papers

- I. **A comprehensive analysis of the *Streptococcus pyogenes* and human plasma protein interaction network.** Kristoffer Sjöholm, Christofer Karlsson, Adam Linder and Johan Malmström. **2014.** *Molecular BioSystems*, 10, 1698-1708.
- II. **Targeted proteomics and absolute protein quantification for the construction of a stoichiometric host-pathogen surface density model.** Kristoffer Sjöholm, Ola Kilsgård, Johan Teleman, Lotta Happonen, Lars Malmström and Johan Malmström. **2017.** *Molecular & Cellular Proteomics*, 16 (4 suppl 1), S29-S41.
- III. **Pathogen interactions with human blood plasma: A comparative and quantitative study between 12 species.** Kristoffer Sjöholm and Johan Malmström. *Manuscript in preparation.*
- IV. **Development of Phage-Based Antibody Fragment Reagents for Affinity Enrichment of Bacterial Immunoglobulin G Binding Proteins.** Anna Säll, Kristoffer Sjöholm, Sofia Waldemarson, Lotta Happonen, Christofer Karlsson, Helena Persson and Johan Malmström. **2015.** *Journal of proteome research*, 14 (11), 4704–4713.

## **My contribution to the papers**

- I. I designed the experiments, performed the experiments, MS-analysis, the data analysis and wrote the paper.
- II. I designed the experiments, performed the experiments, MS-analysis, the data analysis, constructed the surface density model and wrote the paper.
- III. I designed the experiments, performed the experiments, MS-analysis, the data analysis and wrote the manuscript.
- IV. I performed the MS-analysis, the data analysis and helped in writing the paper.

## Abbreviations

<b>CID</b>	– Collision-induced dissociation
<b>DDA</b>	– Data dependent acquisition
<b>DIA</b>	– Data independent acquisition
<b>DNA</b>	– Deoxyribonucleic acid
<b>ECD</b>	– Electron-capture dissociation
<b>ESI</b>	– Electrospray ionisation
<b>ETD</b>	– Electron-transfer dissociation
<b>FTICR</b>	– Fourier transform ion cyclotron resonance
<b>HCD</b>	– Higher-energy collisional dissociation
<b>iTRAQ</b>	– Isobaric tags for relative and absolute quantitation
<b>LIT</b>	– Linear ion trap
<b>m/z</b>	– Mass to charge ratio
<b>MALDI</b>	– Matrix-assisted laser desorption/ionisation
<b>MS</b>	– Mass spectrometry
<b>MS1</b>	– Precursor measurement by MS
<b>MS2</b>	– Fragment measurement by MS
<b>PRM</b>	– Parallel reaction monitoring
<b>PTM</b>	– Post-translational modification
<b>Q</b>	– Quadrupole
<b>RNA</b>	– Ribonucleic acid
<b>RR</b>	– Respiratory rate
<b>SILAC</b>	– Stable Isotope Labelling by Amino acids in Cell culture
<b>SIRS</b>	– Systemic inflammatory response syndrome
<b>SRM</b>	– Selected reaction monitoring
<b>TMT</b>	– Tandem mass tag
<b>TOF</b>	– Time-of-flight

## **Aim**

*Expand the knowledge of host-pathogen interactions to improve the understanding of the development of sepsis, which will be accomplished via state of the art techniques. This wider understanding for the field of host-pathogen interactions will in turn support the development of hypotheses for diagnostics and therapeutics in pathogen related diseases.*

In the following chapters, I will provide the background information required to understand the aim of this thesis, and in detail explain how and why I have performed my research. The chapters are designed to provide a basic understanding of the **Original papers** and to add them into context.





# Chapter 1 – Introduction to proteomics

Most actions in the cells of all life are carried out by protein molecules. The number of different protein molecules in a cell can differ from a few hundred in prokaryotes [1] to a few ten thousands in eukaryotes [2]. In body fluids, protein concentration can vary up to ten billion times [3]. In addition, protein modifications after translation (post-translational modification, PTM) diversify the functional proteins to an even larger number [4, 5]. The PTM can change the function of a protein, for example turn an enzyme active or inactive. All proteins within a limited subset (see examples below) are termed a proteome [6], and have a high analytical complexity, due to the variation in concentration and the number of different proteins. Examples of subsets are the proteins within a species, a strain or as the case for this thesis, the proteins that interacts with the surface of a specific species. The term proteomics was first established by Peter James in 1997 [7], which refers to the study of proteomes and the term comes from genomics (the study of genomes, DNA) and transcriptomics (the study of transcriptomes, RNA).

There are many methods to measure a proteome, and the technical development since the introduction of proteomics has been tremendous [6, 8-10]. Some of the techniques available are two-dimensional gel electrophoresis (2-DE), antibody based capture of proteins and mass spectrometry (MS) [11-13]. 2-DE is based on separating proteins depending on two different inherent properties, most commonly molecular weight and isoelectric point. This method was mainly used in the beginning of the proteomics field and other techniques have passed in popularity due to the low resolution [13, 14]. Antibody based capture relies on

antibodies (immunoglobulins) that binds proteins of interest, and there are several approaches to obtain a readout if the antigen is bound or not [11], for example via detection with florescent light. The specific nature of the capture and the immunoglobulin variability allows multiplexing in a proteomics fashion [11]. However, for each targeted antigen, a unique immunoglobulin has to be identified and there are concerns that some of the commercially available immunoglobulins may be unspecific [15-17]. MS is the main focus of this thesis and there are several different techniques used within MS, which are covered in **Chapter 2**. In **Paper IV**, a combination of immunoglobulin based capture and MS techniques were used.

### **Bottom-up vs. top-down**

The two major fields within MS based proteomics are top-down and bottom-up, where intact proteins or digested proteins (peptides) are measured, respectively [18, 19]. When analysing proteins it may seem straightforward to measure intact proteins. However, mass analysers are generally better at measuring molecules that are smaller than intact proteins. Hence, bottom-up proteomics covers most of the MS based proteomics [19, 20]. The most common approach to cleave proteins are to use the enzyme trypsin [21], which is a very specific enzyme that generates peptides of a suitable length and easily ionized peptides. However, there are several issues related to bottom-up proteomics, the main two are; mapping the peptides back to the protein, they may match multiple proteins; and the loss of information of PTMs in the proteins [22, 23]. There is an in-between field referred to as middle-down proteomics [23], which attempts to capture the benefits of both top-down and bottom-up. Middle-down utilise limited digestion that generates longer peptides than that of a bottom-up digestion, but not of sufficient length to cause problems in the mass analysers

[23]. In the **Original papers** of this thesis bottom-up proteomics and trypsin cleaved proteins has been used exclusively.

## **Mass spectrometry**

A mass spectrometer is an instrument that can separate ions based on their mass to charge ratio ( $m/z$ ). The mass spectrometer consists of three major parts, an ionisation source, a mass analyser and a detector [12].

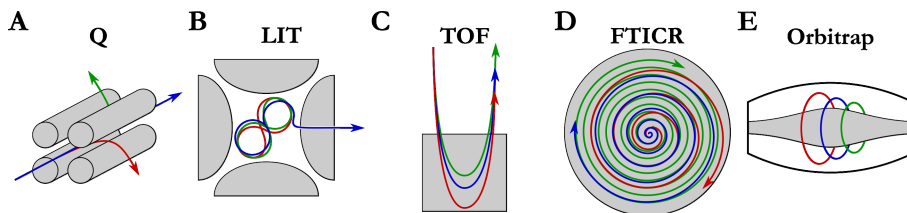
### **Ionisation source**

The most basic components of a mass analyser are electromagnets; hence, the requirement of producing ions since a magnetic field only can manipulate particles that are charged. In addition, the mass analysers can only analyse molecules that are in gas phase. When ionising proteins and peptides there is a demand for “soft ionisation techniques” to avoid breaking the macromolecules apart. Development of the soft ionisation techniques matrix-assisted laser desorption/ionisation (MALDI) and electrospray ionisation (ESI) was awarded the Nobel Prize in Chemistry in 2002. In MALDI, the sample is adsorbed to a matrix, then pulsed by a laser that causes the molecules of the sample to acquire protons from the matrix, which makes them charged [24, 25]. MALDI mainly produces singly (positively) charged ions, but by changing the matrix or the power of the laser, the charge of the molecules can be adapted to suit the scientific question [26]. In addition, the heat produced when the laser hits the matrix generates molecules in gas phase [24]. In ESI, the sample is first ionised by applying high voltage to a liquid [27, 28] and small droplets are obtained with a sharp tip at an emission point of the liquid. These small droplets contain many molecules and by the use of vacuum and temperature, the droplet evaporates until a certain point (the Rayleigh limit) [29]. When the Rayleigh limit is reached, the droplet breaks apart into separate units due to the repulsion forces of the charged

particles [27, 29]. ESI has the benefit of being used with a continuous flow, which is an essential part of bottom-up proteomics and is the only ionisation method used in any of the **Original papers**.

### Mass analysers

There are many different methods to achieve  $m/z$  separation and each with its own advantages and disadvantages, but all relies on magnetic fields to manipulate the flight path of the ions. In **Figure 1**, a schematic overview of the main type of mass analysers is seen. Quadrupoles (Q, **Figure 1A**) consist of four metal rods (the poles) and an electromagnetic wave is applied between the opposing poles. This triggers the quadrupole to function as an  $m/z$  filter, where only the selected  $m/z$  ranges can pass through [12]. The frequency of the electromagnetic wave can be changed to alter the range of analysed ions [30]. Quadrupoles was used in all the **Original papers**. Linear ion traps (LIT, **Figure 1B**) are similar to the quadrupoles with additional capabilities of longitudinal trapping of ions and possibility of gradual ejection through lateral slits to the detector [31]. Time-of-flight (TOF, **Figure 1C**) mass analyser depends on ions being accelerated in an electron field where the velocity depends on the  $m/z$  of the ion [32, 33]. Fourier transform ion cyclotron resonance (FTICR, **Figure 1D**) is the most advanced type of mass analyser [34]. In FTICR, the ions enter a



**Figure 1.** The five most common mass analysers. The red, green and blue arrow represent three different ions (different  $m/z$ ) and their behaviour in the different mass analysers. The grey area represents the electromagnets that alters the flight path of the ions in the mass analyser. The mass analysers visualised in the figure: **A)** Quadrupole (Q), **B)** Linear ion trap (LIT), **C)** Time of flight (TOF), **D)** Fourier transform ion cyclotron resonance (FTICR) and **E)** Orbitrap.

magnetic field and based on the ions movement in the magnetic field the  $m/z$  are determined [35]. FTICR is similar to a particle accelerator and requires a superconducting magnet, which is cooled by liquid helium [34] and is the most expensive of the mass analysers [34]. In an orbital trap (Orbitrap, **Figure 1E**), a voltage ramp is applied to a barrel like structure, which causes the ions to rotate around and to oscillate back and forth the barrel. The speed of the oscillations depends on the  $m/z$  of the ion [36, 37]. An Orbitrap was used in **Paper I** and **III**.

### **Detector**

The detector is the MS part that detects and counts the ions. There are two different types of detectors. The first is an electron multiplier that counts number of the ions hitting the detectors, which initiate an electron cascade to generate a detectable current. Hence, an electron multiplier depends on the mass analysers to separate the ions prior to collision with the detector [33]. The second type of detectors is used in FTICR and Orbitrap mass analysers. In this case, the ions pass near a detection plate, which records the oscillations. Hence, the ions are not hitting the detector and multiple ions can be detected simultaneously. The oscillations that are recorded are translated to  $m/z$  via inverted Fourier transform [38].

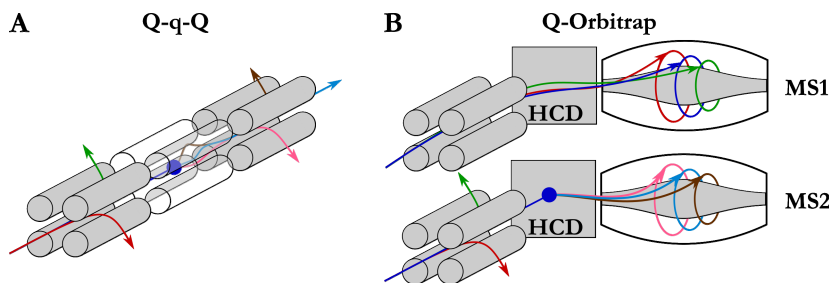
## **Mass spectrometry based proteomics**

The selection of mass analysers depends on the objective of the analysis and come with inherent strengths and weaknesses. The main properties are speed and precision (resolution). If the resolution is sufficient, the  $m/z$  accuracy is sufficient to differentiate the natural isotopes of the elements present in all molecules. These isotopes always have  $[m + 1]$ ,  $[m + 2]$ , ...,  $[m + n]$ , and

the charge determination is possible if the isotopes are differentiated. Hence, mass determination of the peptide is possible.

## Tandem MS

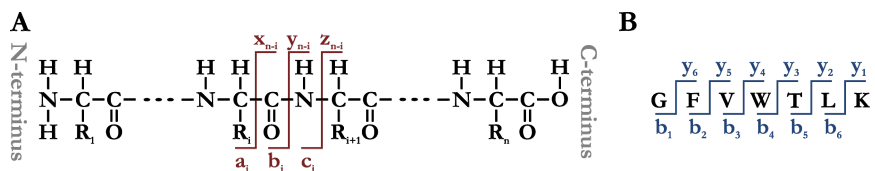
The selection of a mass analyser becomes even more complicated when the analysers are combined in tandem MS. Information of the mass of a peptide is insufficient to distinguish the peptides in most proteome sample. This demands a gathering of more information, which is solved by the addition of a fragmentation cell and another (sometimes more than one) mass analyser [12, 39]. Examples of combinations are, triple quadrupoles (Q-q-Q, **Figure 2A**), Q-Orbitrap (**Figure 2B**), Q-TOF, Q-LIT-Orbitrap. The peptides are in general first measured as peptides without fragmentation. In the next step, the peptides are selected one by one then fragmented and measured. There are some variations of the setups and a few of them is described in detail in **Chapter 2**.



**Figure 2.** Two versions of tandem MS where multiple mass analysers are combined to measure the peptide fragments. **A)** A triple quadrupole (Q-q-Q) mass spectrometer, two quadrupoles act as mass filters. The first (Q1), filters for peptide (precursor) and the second (Q3) filters for the fragment. The third quadrupole (the second one sequentially, q2) functions as the fragmentation cell, in where the ions are fragmented using a collision gas. **B)** A Q-Orbitrap, first measures the mass of all the ions entering (MS1) then a single precursor is selected in the Q and all the fragments are measured (MS2). The peptides are fragmented by a higher-energy collisional dissociation (HCD) cell.

## Peptide Fragmentation

Introduction of a step where the peptides are fragmented provides for instance the possibility to differentiate between peptides with the same precursor  $m/z$  but with different amino acid sequences. Peptides are more likely to fragment in places with a high degree of rotational freedom, which is along the backbone of the peptide (**Figure 3**) [39-43]. A peptide can break in three positions (**Figure 3A**) per amino acid along the backbone. The N-terminal pieces are termed a-, b- and c-ion and are denoted by the number of amino acids in that sequence (**Figure 3B**) [43]. The C-terminal pieces are in the same manner termed z-, y- and x-ion (**Figure 3A**) [43]. There are two major methods to achieve peptide fragmentation, by addition of electrons or colliding with a gas. In electron-capture dissociation (ECD) and electron-transfer dissociation (ETD), electrons are transferred by bombarding the peptide with electrons or with negatively charged radicals, respectively [44, 45]. This method has a tendency to generate mostly c- and z-ions (**Figure 3A**) [46, 47]. ETD and ECD fragmentation is typically more suited for longer peptides than the peptides normally provided by trypsin and is henceforth mostly used in top-down, middle-down and PTM analysis applications [45, 48]. In collision-induced dissociation (CID) and higher-energy collisional dissociation (HCD), the peptides collide with an inert gas to produce fragmentation, which generates mostly b- and y-ions (**Figure 3B**) [46]. In principle, the type of inert gas



**Figure 3.** Peptide fragmentation in tandem MS. **A)** The peptide is fragmented along the backbone of the peptide. The names of the N-terminal fragments are a-, b-, c-ions and the corresponding C-terminal fragments are named z-, y-, x-ions. The fragments are denoted of the length of the fragment in number of amino acid side chains (R). **B)** Shows an example of a peptide with all potential fragments if only b/y-ions occurs. Figure is adapted from Marcotte, 2007.



(size of the gas molecules), the pressure (number of gas molecules) or the energy (speed of the gas molecules) can be altered to change the number of collisions, which will change the amount and the size of fragments [49, 50]. The optimal settings for fragmentation are different for every peptide and can be estimated relatively well [51, 52]. However, if a set of peptides is measured often, optimising the fragmentation can be worthwhile. If multiple peptides are fragmented at the same time, a setting that fits most peptides is used. There is normally no demand for optimal fragmentation, only sufficient fragmentation for detection [53-55]. Exceptionally, during the application of *de novo* peptide sequencing other rules applies (see under, data analysis, **Chapter 2**). CID and HCD was the fragmentation methods used in the **Original papers**.

### **Separation**

Tandem MS and peptide fragmentation is very good at detecting peptides but the technique is not fast nor sensitive for adequate detection of all peptides in a proteome sample derived from cells, tissues or body fluids. Therefore, the sample is simplified (separated) with different methods before being injected into the mass spectrometer [40]. Either the sample is separated prior to MS analysis (off-line separation) or a separation method is coupled directly to the mass spectrometer (on-line separation). There are many methods to separate peptides based on other properties than mass. Some examples are, molecular weight of the protein using SDS-PAGE (off-line) [9, 56] and isoelectric focusing of peptides (off-line) [57]. Generally, reverse phase (RP) liquid chromatography (LC) is used with bottom-up proteomics [40]. In RP-LC, the peptides are loaded to a hydrophobic matrix (column) and a gradient of hydrophobic solvent (for example isopropanol or acetonitrile) is gradually added to the column, eluting peptides depending on their hydrophobicity [58]. The gradient time can be varied depending on the separation required for the scientific

question, for example in **Paper II** the gradient was 30 minutes and increased from 5% acetonitrile to 30% acetonitrile. RP-LC is still not sufficient to cover every single peptide of a proteome [59-61], but it is the most time and cost effective method. For example, in one hour gradient 90% of the yeast proteome (approximately 4500 expressed proteins) can be detected [10]. If extreme coverage is required, additional methods of separation (preferably orthogonal to hydrophobicity) are added to the sample preparation procedure. For example, the proteome of a human cell line (12000 expressed proteins) was determined to 90% using mass spectrometry, but required additional off-line separation and required in total 288 hours to analyse with MS [62]. However, only on-line coupled RP-LC was used in the **Original papers**. In **Paper I**, one additional (off-line) separation method was used to increase the coverage.

## Conclusion

In this chapter, I have introduced the technique used in the **Original papers**, which is the base of this thesis. In the next chapter, I will expand further around the concepts of MS based proteomics and its application to biology.



## Chapter 2 – Applying proteomics in biology

This thesis is focused on the selected subset of host-pathogen interactions, with the overall goal to map out the interactions between human blood plasma and pathogen surfaces (see **Chapter 3**). However, only identifying the interacting proteins are not sufficient. In addition, the frequency of the interaction is required.

### Quantification

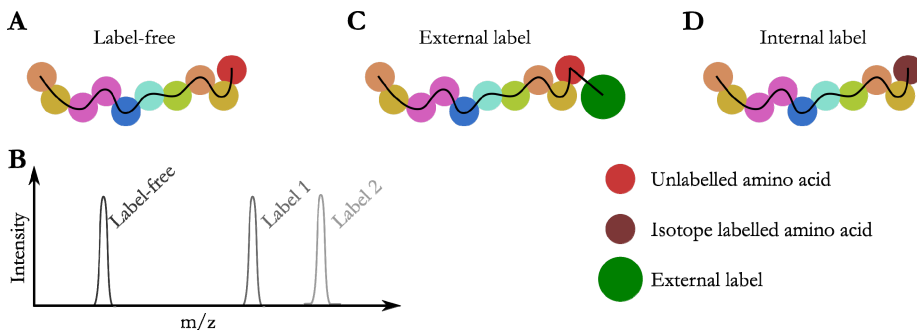
The knowledge whether a protein is present in a sample is relevant, however, the amount of that protein is more relevant. There are numerous approaches to perform quantification in bottom-up MS and several parameters should be considered [63]. Some of the important parameters are the number of samples in the study, the numbers of peptides that are monitored, the accuracy of the quantification and the cost, which all depends on the scientific question. In addition, the MS instrumentation available and the possible settings of those instruments are of great importance [9, 63-65], and are covered later in this chapter. The optimal MS approach depends on the scientific question and the most important is to select a method that answers the scientific question, with as little exertion as possible.

Proteins in a sample that are quantified in relation to the same protein in other samples is known as relative quantification [66, 67], and quantification of a protein independent of other samples is known as absolute quantification [66, 67]. Quantification of peptides without addition of internal standards is termed label-free quantification (**Figure 4A**) [68]. This method has several advantages.

For example, no additional processing steps are introduced in the sample preparation procedure and it is possible to do label-free quantification for any samples [67-70]. However, label-free quantification is less accurate compared to labelled quantification [70] and label-free quantification is associated with higher sample-to-sample variations in the mass spectrometer partially explained by a phenomena termed ion suppression [71]. Ion suppression occurs in ESI where the formation or evaporation of the small droplets are not ideal, possibly explained by the peptides eluting at the specific time point are less volatile [71]. Hence, the amount of peptides that are in gas phase and can enter the instrument is lower than the total amount of peptides [71]. The quality of the label free quantification is dependent on the mass spectrometry setup and the software used for data analysis (see under data analysis, later in the chapter) [63].

### **Labelling**

To help quantification accuracy and the issues mentioned above, it is possible to use labelled peptides. In principle, a mass shift (label) is introduced to differentiate between peptide analogues from different samples (**Figure 4B**) [70]. The analysed samples are pooled prior to injection to the mass spectrometer [70], reducing the technical variation between samples. There are two different approaches to labelling peptides, externally (**Figure 4C**) or internally (**Figure 4D**). In external labels, the peptides react with a reagent that adds a tag on the peptide. Two examples of external labels are isobaric tags for relative and absolute quantitation (iTRAQ) [72] and tandem mass tag (TMT) [73]. Multiple tags are available and there is a possibility to pool many samples. Internal labelling utilises uncommon isotopes of carbon (or nitrogen) in the amino acids [74, 75]. Most commonly in lysine and arginine, which are always present in peptides from tryptic digests [74, 75]. In Stable Isotope Labelling by Amino acids in Cell culture



**Figure 4.** Peptide labelling is important for quantification in some applications. The peptide is normally unlabelled, which is termed label-free (**A**). **B**) A label of a peptide introduces a mass shift to the peptide and there are several reagents, which may introduce different mass shifts (labels). There are two predominant approaches to label peptides. **C**) The first label type is an external label, which introduces an extra molecule to the peptide. **D**) The second label type is an internal label, which introduces another isotope version to one of the amino acids in the peptide chain. The amino acid that is changed is normally an arginine or a lysine, which is the last amino acid in the sequence of a tryptic peptide. The legend contains the colours of the modifications.

(SILAC) the sample is grown in a medium containing heavy labelled arginine and lysine [74]. This means that the samples are prepared the same and the relative quantification is high [74]. However, there are some limitations to this method, like the demand of growing the sample in a culture, which is not always possible. In addition, SILAC and external labels modifies all peptides and therefore creating a more complex sample due to the sample pooling. External labels and SILAC are not used in any of the **Original papers**.

Another internally labelled method is spike-in labelled peptides. In this method, the peptides are added to the sample just prior to analysis by MS [69, 75]. Hence, spike-in peptides only quantify what is injected into the mass spectrometer and the method does not cover the deviation in sample preparation procedure [70]. The purity and the concentration accuracy of the labelled spike-in peptides are adapted depending on the analytical accuracy required to answer the scientific question. However, the cost is greater the higher the demanded concentration

accuracy of the peptides is and the number of peptides in the study might have to be reduced depending on the demanded accuracy. In **Paper II**, around 70 peptides were absolutely quantified with high analytical accuracy. The quantification methods used in the **Original papers** were in **Paper III** label-free exclusively, in **Paper II** spike-in exclusively, and in **Paper I** and **IV** label-free and spike-in. The quantification method should be selected to answer the specific scientific question and there is not always a demand for the highest analytical accuracy, which result in label-free quantification producing sufficiently accurate results in most cases [76]. The accuracy of the quantification is also dependent of the choice of mass spectrometer, and the settings of the instrument.

## Mass spectrometer setup

There are two main MS techniques available for performing an analysis of peptides, which relies on different instrumentation. The two techniques are targeted and non-targeted [63]. In the targeted method, the ions that are monitored are selected prior to analysis [63, 77]. In non-targeted methods, the instrument is set to monitor as many ions entering the instrument as possible for identification of the ions [37, 63]. Hence, a targeted approach often prerequisite a non-targeted approach. I will go through in detail the three methods (**Figure 5**) used in the **Original papers**.

### Data dependent acquisition

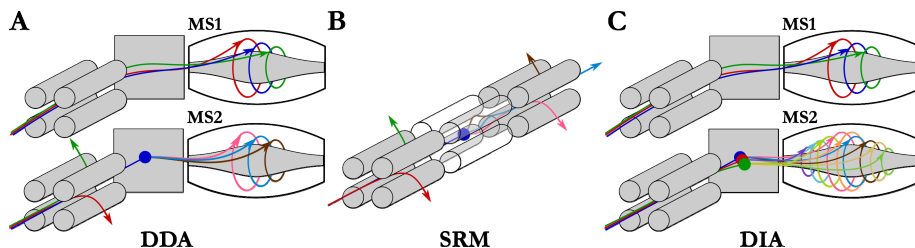
The first MS method, data dependent acquisition (DDA, **Figure 5A**) is a non-targeted MS method, and one of the most commonly used techniques to identify the peptides present in the sample [63, 78]. The method predominantly utilises Q-Orbitrap and Q-TOF mass analysers [76], due to their high mass accuracy and scan speed. The data dependency originates from the automatic picking of peptides from survey scans based on pre-set selection criteria [68]. The

peptide ions are sequentially isolated by a quadrupole, fragmented in a collision cell and then a fragment spectrum is acquired, which is repeated for the whole period of the method. To avoid measuring the same ion repeatedly, its  $m/z$  is added to a list and ignored (dynamic exclusion) for a set time [68], which results in constant measurement of different ions. In general, the method first measures the peptides then automatically picks the most abundant peptides for fragmentation [79], in **Paper III** the 15 most abundant ions were picked in each cycle. Hence, the instrument might pick different peptides in each sample [68], which is a major issue for quantification. This method is mainly used for identification of the detectable peptides in the sample. Due to the dynamic exclusion, quantification in DDA is attained at peptide level (MS1), which is not as accurate as some other MS methods (see, data analysis) that quantify at fragment level (MS2) [65, 70]. In DDA, it is possible to identify up to 4000 proteins in a single run [10]. DDA was used for identifying host-pathogen interactions in **Paper I** and **III**.

### **Selected reaction monitoring**

The second MS method, selected reaction monitoring (SRM, **Figure 5B**) is a targeted MS method used to quantify peptides in a sample. SRM is normally performed using a triple quadrupole, since, the dual filtering step of the instrument [63, 76, 77]. In SRM, a list of peptides and fragments (the combination is termed a transition) are selected before the measurement [77, 80, 81]. The peptides (precursors) are sequentially selected in the Q1, fragmented in q2 and then the fragment is selected in Q3. Normally four to five fragments are selected per peptide [81-83], and then the instrument selects the next peptide to measure. The fragments that are selected in SRM are mainly y-ions and some b-ions, due to the CID fragmentation in a triple quadrupole [46]. This method is sensitive, reproducible and accurate (MS2 quantification) [75, 77, 84]. In addition, absolute





**Figure 5.** A graphical overview of the three types of MS methods used in the **Original papers**. **A)** In data dependent acquisition (DDA), the MS1 of the peptides are measured, which are then selected one at a time and fragmented, sequentially. **B)** Selected reaction monitoring (SRM) is a targeted method and work with a dual filtering principle. The peptide is selected in Q1, which is fragmented in q2 and the fragment is selected in Q3. **C)** In data independent acquisition (DIA), the peptides are measured, and then multiple peptides are fragmented at the same time and the fragments are measured.

quantification is possible when using SRM and the dynamic range is roughly two orders of magnitude larger than other MS methods [61, 66, 85]. However, there are some drawbacks, prior knowledge of the peptides and fragments is required before measurement and the number of peptides measured simultaneously is limited. In SRM, roughly 100 proteins can be monitored in a single run. SRM was used **Paper I, II** and **IV** for quantification and absolute quantification. The version of SRM that is run on a quadrupole-Orbitrap is termed parallel reaction monitoring (PRM), and measures all the generated fragments of the selected precursor. In PRM, only the precursor is selected, but with some trade-off in accuracy and dynamic range.

### Data independent acquisition

The third MS method, data independent acquisition (DIA, **Figure 5C**) utilises aspects of both DDA and SRM. This method is a rather recent addition to the MS toolbox, and most commonly relies on a Q-Orbitrap or a Q-q-TOF [65, 86]. DIA is based on the fragmenting of multiple peptides simultaneously [65, 87-89], instead of one at a time as in DDA. In DIA, an  $m/z$  window size is

selected (normally 25 m/z) and all peptides within that mass window are fragmented [65, 86]. Then the instrument selects the next mass window, which is repeated (normally 32 windows) until the whole m/z range (normally 400 – 1200 m/z) is covered [65, 86]. DIA has the possibility to identify as many proteins as in DDA and the quantification accuracy is high due to MS2 quantification [65]. However, DIA requires a lot more computational power to deconvolute the spectra compared to other MS methods [89-92]. In addition, the deconvolution of spectra prerequisite knowledge of the peptide fragmentation and the retention time (assay), which information is normally obtained from DDA. This means that DIA requires a library of assays for data analysis [90-92]. One of the benefits with DIA in comparison to DDA is that all the detectible peptides are measured with the method, which results in a digitalised sample. [65, 86]. However, only peptides with an existing assay can be analysed, but it is possible to acquire new assays afterwards and reanalyse the data without repeating the sample measurement. DIA was used **Paper III** for proteome wide quantification. The data analysis is different depending on the MS method used and most advanced in DIA.

## Data analysis

The data acquired in a mass spectrometer is mass spectra and several thousands are normally generated in a single experiment using one-hour gradient [9]. Hence, the annotation from mass spectra to peptide is performed computationally and there are numerous software available depending on the type of analysis.

### SRM

In SRM, a set of transitions and intensities are measured over time [81]. The data from SRM is typically smaller than that of DDA and DIA and easier to analyse. Since the peptides are selected before SRM analysis, the only objective is

peptide quantification. The peptide quantification is accomplished by calculating the area under the curve for all the transitions matching to a single peptide [80]. There are several programs available to perform analysis of SRM data, such as Skyline [80], Anubis [82] and mProphet [93]. All SRM data in the **Original papers** were analysed by Skyline, due to the easy to use interface, and the manual correction and validation capabilities.

## DDA

In DDA, the annotation of spectra to peptide is primarily performed by spectral matching to *in silico* generated (theoretical) spectra [39, 40, 53-55]. A theoretical spectra contains all potential fragments generated by the peptide fragmentation rules (**Figure 3**) [94]. It is possible to generate the theoretical spectra for any peptide, but since an entirely random approach generates too many spectra for searching ( $20^n$ , where  $n$  is the length of the peptide), so the search space is reduced. Hence, the genome of the species analysed is used as a template [94] and since the development of genome sequencing has drastically been improved, the time and cost of sequencing [95] has been reduced drastically. This cost reduction results in the requirement of a genome no longer being especially problematic and spectral matching functions well. There are several software tools that performs this matching, such as Mascot [53], MaxQuant [96] and the Trans-Proteomic Pipeline (TPP) [97]. These software tools can also map the peptides to proteins (see, peptide to protein). The software used to process DDA data in the **Original papers** was TPP. The pipeline is composed of several tools, first spectral matching by X!Tandem [98] (and/or several other tools [97]), then PeptideProphet [99] for validation of the search engine results (see below, false discovery rate) and lastly the ProteinProphet [100] to map peptides to proteins. DDA is the best method for identification since the only input required is a digested sample and the theoretical proteome of the sample. The

quantification can be accomplished by counting the number of spectra that match to each protein (spectral counting) [101], which is easy but crude. Another possibility is to calculate the area under the curve for the MS1 peak (MS1 quantification) [102], which is the best option for DDA, but still less accurate than DIA [65].

*De novo* sequencing, the identification of peptides without prior genomic knowledge has big issues with peptide fragments not completely covering the peptide sequence [41, 103-105], and is not common within proteomics. *De novo* sequencing is performed with DDA, and by annotating the fragmentation in spectra to amino acids [104-106]. This is difficult without the complete fragmentation pattern (**Figure 3B**), which is rarely obtained [46]. In addition, the *de novo* identification software are underdeveloped since the spectral matching performs well in most occasions. However, it is an interesting future possibility, an example, to account for mutations not accounted for in the original proteome.

## DIA

In DIA, a mass spectrum contains multiple peptides in the same spectrum. The method is the newest of the method presented in this thesis, due to recent instrumental and software development [65]. One of the main difficulties is to deconvolute the spectra so the peptides can be quantified. Some of the software that solves this problem are DIANA [91], OpenSWATH [90], DIA-Umpire [92] and Skyline [80, 89]. DIA software uses assays to deconvolute and match spectra to peptide [90, 91]. Assays contain information of peptide, modifications, charge, fragments, relative fragment intensity and retention time [90, 91]. The assays are generated from DDA data. Hence, DIA initially requires multiple MS runs, until a suitable assay library is built. The assays from DDA data are stored in the set of assays (assay library) with the possibility of

complementation with new assays afterwards [65]. Over time, the demand for adding assays to a library will decrease and more instrument time can be spent on DIA [65]. DIA quantifies peptides by calculating the area under the assay fragments, in a similar manner as in SRM [91]. DIA obtains some of the missing values between samples that are a big issue in DDA but there are still some missing values. However, there has recently been some development to manage this matter by Röst et al [107].

### **False discovery rate**

To validate whether a spectrum in fact is a true match to a peptide the software tool is required to determine the false discovery rate (FDR) [108]. In DDA, this means that the data is searched with a standard database (the theoretical proteome of the analysed organism) and a decoy database (usually the reverse of the standard) and all spectra are scored based on the best matching peptide [108]. This means that a spectrum can match a standard or a decoy peptide, but should in theory only match the decoy peptide by random. The FDR is calculated as the number of decoy peptides divided by the number of standard peptides at a specific score [108]. The decoy peptides score in general considerably worse (hopefully) compared to the standard peptides and a cut-off in the score is placed at the FDR of 1% (normally) [108]. This means that 1% false peptide identifications are accepted. A similar approach is performed in DIA but with decoy assays [90, 91]. When inferring from peptide to protein the 1% FDR is at protein level [109]. More stringent criteria can be set for avoiding false positives, in **Paper III**, the minimal of 2 peptides per proteins (in addition to 1% protein FDR), resulting in removing roughly 15% of the data.

### **Peptide to protein**

Peptide to protein are trivial if the peptide only match one protein (proteotypic) but this is not always the case [110]. Instead, some peptides match

multiple proteins, which is then a protein group. An example is IgG where there are four subclasses (for more detail, see **Chapter 3**), this means that the potential protein groups are 15 instead of the four subclasses. Hence, the subclasses all have peptides that match the other subclasses, but not always to all the subclasses. In most cases all non-proteotypic peptides are removed [110]. However, sometimes there might be an interest in working with for example all subclasses of IgG rather than the single entities. The only method to perform this is manually, which is unpractical for large datasets.

### **Issues with bottom-up proteomics**

Bottom-up MS is today the best method to identify the components in a proteomes. However, the technique is not without flaws. For example, protein that are naturally cleaved by enzymes before the sample processing might be detected by bottom-up [19], but the information of the cleavage will be lost since the peptides still match the same proteins [19]. Biological context of PTMs is frequently lost, especially in the case when proteins have multiple modifications [111]. Proteins with few cleavage sites and peptides with low solubility are often lost due the different properties than the general peptide and protein [112]. A group of proteins that are underrepresented due to these reasons are membrane proteins [112]. In addition, the quadrupole is supposed to pick one peptide at a time but the quadrupole cannot always differentiate the peptides sufficiently, resulting in multiple peptides per spectrum [113]. Some software accounts for multiple peptide spectra, like DeMix [114], but they are not included in the major search engines. The scientists developing algorithms for analysing MS data are often far away from the biological problem that is supposed to be solved by their algorithms. I think the field would benefit from closing this gap, so that the solutions are aimed for the end goal rather than an intermediate goal.

## Conclusion

In this chapter, I have provided information on how to apply MS in biology and the methods I have performed this with, from sample to protein. I have worked with selecting the optimal parameters and setup for answering my specific scientific questions in each of the **Original papers**. Adapting the question for what instrumentation was available and optimising the settings for answering that question. In the next chapter, I will provide the reasoning behind working with host-pathogen interactions.

## Chapter 3 – Host

A biological system can be defined as a network of biological entities with shared interactions. In infectious medicine, this system is exemplified in the complex network of interactions between host and pathogen, which has shaped human evolution and resulted in several distinct mechanisms, like the development of the human immune system. The focus of this thesis is to unravel the key mechanics of this biological system in the hopes of finding potential therapeutic opportunities. This chapter is dedicated to the host of this biological system, with the pathogens covered in **Chapter 4**. There are several reasons for selecting humans as the host for my scientific questions. These reasons include, the knowledge of human defence mechanisms is extensive, the value for the scientific community of any findings is high and the setup of the experiments is performed *ex vivo*, hence, no host has to suffer for my results.

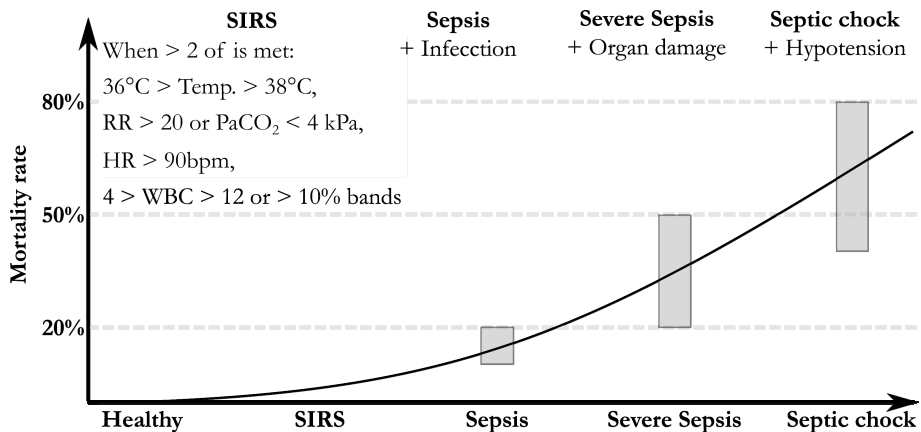
There are many locations that an infection can originate from (entry point), where the most common are the airways, the skin, the abdomen and the urinary tract [115-117]. The entry point is of importance for the entire development of disease and for the patient. However, to limit the area of this thesis, the focus is when the pathogen infiltrates deeper in the body and enters the bloodstream, which may result in sepsis.

### Sepsis

Sepsis is defined as a, “life-threatening organ dysfunction caused by a dysregulated host response to infection” [118]. The clinical parameters for sepsis or the derivatives of sepsis are shown in **Figure 6**. The first stage of the disease



progression is known as systemic inflammatory response syndrome (SIRS), where the patient presents with two out of the four clinical parameters. The first, having a temperature below 36°C or above 38°C. The second, the respiratory rate (RR) is above 20 per minute, or the arterial carbon dioxide tension (PaCO<sub>2</sub>) is lower than 4kPa. The third, having a heart rate above 90bpm. The fourth, having a white blood cell count outside the range 4,000 – 12,000 cells per μL or over 10% immature neutrophils (bands) [117, 119]. Sepsis is the next stage of the disease progression and clinically defined as SIRS plus presumed or confirmed infection. In the western world, sepsis have a mortality rate of 10-20% [116]. However, the stage of sepsis becomes more critical if the patient also develops organ dysfunction (severe sepsis), in which the mortality rate increases to 20-50% [116]. This state can worsen if the patient also develops hypotension despite intravenous fluids (septic shock), resulting in a further increased mortality rate of 40-80% [116]. In addition to the high mortality rates, sequelae are common, which increases the morbidity of sepsis [120-122].



**Figure 6.** The clinical parameters for sepsis and the derivatives of sepsis are listed in the top. The sequence is SIRS → Sepsis → Severe sepsis → Septic shock, and for diagnosis of a later stage all the previous stages are required. The graph shows the mortality rate of the different steps, adapted of data from Martin, 2012.

## **Blood**

Blood have several distinct roles in human physiology including delivering nutrients and oxygen to distant cells, removing waste products from tissues and maintaining system homeostasis. The bloodstream is the circulatory system where blood flows, and is normally a sterile environment without any microbial presence. However, when a pathogen breach into this sterile environment there are several defence systems in place to remove and incapacitate the intruders. Blood is composed of erythrocytes, leucocytes, thrombocytes and plasma. Of these components, leucocytes (white blood cells) and plasma play a role in the defence against pathogens, referred to as the immune system [123]. The blood plasma is the liquid component of the blood, comprising roughly 50% of the total blood volume and contains a high concentration of proteins (70 mg/ml) [124]. There are roughly two thousand different proteins detected in human plasma [125], five times lower than the average proteome of a human cell [126]. Even though there are relatively few proteins in plasma compared to a human cell, these proteins display a large dynamic range [3], which can complicate protein analysis [127]. The high abundance of proteins involved in the immune response, highlights the importance for the host to keep this environment clear from invading pathogens. There are two separate parts of the immune system, the innate and the adaptive.

## **Innate immune system**

This part of the immune system is generic and acts like a first line of defence against invading pathogens [128]. The cells and proteins in this system have developed mechanisms that can identify pathogens based on general pathogenic traits, such as the saccharides on the cell wall [129].

## Cells

There are several leucocytes involved in the innate immune system, such as neutrophils and monocytes [128, 130]. Neutrophils are the most common of the leucocytes and have the ability opsonise or phagocyte the pathogen, in addition to activating the coagulation and complement systems [128]. Monocytes can differentiate into macrophages when leaving the bloodstream and entering tissues [130]. As neutrophils, the macrophages can perform multiple actions but phagocytosis of pathogens is their main action [128]. Monocytes can also differentiate to dendritic cells [130], which migrates to tissues in contact with the outer environment (for example, skin or nose), where they mature [131]. Dendritic cells can phagocyte pathogens and digest their proteins into peptides [130]. These peptides are subsequently presented as antigens for the adaptive immune system, bridging the innate and the adaptive immune system [130, 132].

## Coagulation system

The coagulation system (3 mg/ml plasma [133]) can trap pathogens to stop their movement in combination with stopping a bleeding [134]. The system is composed of several plasma proteins including, fibrinogen, thrombin and coagulation factors, where fibrinogen constitutes for around 95% of the coagulation protein concentration [133, 135]. In the clot formation, the coagulation factors activates a cascade of effects that ends with prothrombin (inactive) being cleaved into thrombin (active), which in turn cleaves fibrinogen into fibrin [133]. Fibrin then cross-links with coagulation factor XIII, which can link multiple fibrin together and forms a network (the clot) [133]. Fibrinogen is a central protein in all the **Original papers**.

## Complement system

The complement system (3 mg/ml plasma [136]) can directly kill pathogens and identify pathogens for phagocytosis [137]. There are nine

complement components (C1-C9) and several complement factors in the system. The central protein in the complement mechanisms (C3) [138] is also the most abundant with 40% [136] of the total concentration of all the complement proteins, and the second most abundant is the regulator protein complement factor H with 20% [136]. The system can recruit and mediate immune cells for phagocytosis [139]. In addition, the system can directly kill pathogens by assembling a membrane attack complex (MAC) with C5-C9 [140]. The MAC creates a pore in the surface of the pathogen, which lyses the pathogen by osmotic pressure [140]. Many proteins within the complement system are analysed in **Paper I, II and III.**

## **Adaptive immune system**

This part of the immune system is specific against a particular pathogen [141] and is more effective to neutralise that particular pathogen [141]. However, the host must have encountered the pathogen before the system is adapted to defend against the specific pathogen, which requires a few days to develop [141].

### **Cells**

The adaptive immune system consists of two kinds of lymphocytes, T-cells and B-cells [132]. These cells develop specific responses against a pathogen depending on the antigens presented to them (for example by dendritic cells) during their maturation process [132]. The specific responses are the development of an antigen specific receptor for T-cells [141] or an antibody (immunoglobulin) for B-cells [141]. Since, the T-cell receptor and immunoglobulins should have the possibility to bind all potential antigens there is a demand for extreme variability [142]. For example, the immunoglobulins can have more than  $10^{16}$  different sequences [142], due to the variable region (**Figure 7**). Cytotoxic T-cells detects virus infected and tumour cells by antigen recognition and destroys them [143]

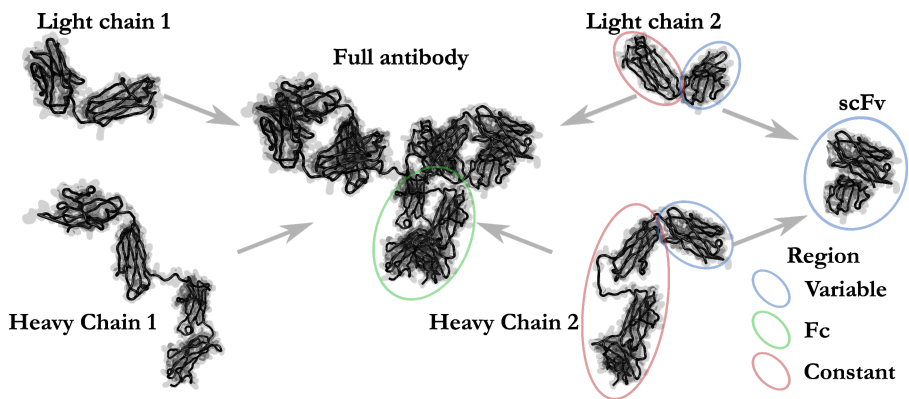
and Helper T-cells aids in the immune response by regulating and directing the response for best effect [144]. B-cells can produce immunoglobulins against their antigen and secrete them into the plasma [145-147]. There are also memory-cells of B-cells and T-cells that stores the adaptive response over time [141].

### **Immunoglobulins**

Immunoglobulins identify pathogens for the recruitment of other parts of the immune system [148, 149]. The concentration of immunoglobulins in plasma is around 15 mg/ml (20% of total plasma concentration) [150]. An immunoglobulin is a tetramer, composed of two identical heavy chains and two identical light chains (**Figure 7**). All the chains have a constant and a variable region. There are several isotypes of immunoglobulins and the most abundant in plasma is IgG followed by IgA and IgM with a concentration of 11.2, 2.6 and 1.5 mg/ml, respectively [150]. The isotypes have different constant regions of the heavy chain, thus they have different fragment crystallisable (Fc) regions (**Figure 7**). The variable region binds the antigen, which triggers a conformational change in the Fc region [151] and allows the Fc region to bind to Fc-receptors on leucocytes [152], such as macrophages or neutrophils. This binding activates the cells for an immune response [152] and the response is dependent on the immunoglobulin isotype and subtype [153]. In addition to the sequence differences of the immunoglobulins, they have different glycosylation (a type of PTM) patterns [154], which ensures optimal binding to the Fc-receptor [155, 156] and maintain conformation [157]. IgG is the most important immunoglobulin and is subdivided into four subtypes (IgG1, IgG2, IgG3 and IgG4) with a relative abundance in plasma of 60%, 32%, 4% and 4%, respectively [158]. Immunoglobulins are a key part in all the **Original papers**.

The antigen binding property of immunoglobulins [159] can also be used for capturing targets of interest in a sample purification process [160, 161].

In **Paper IV**, several single-chain variable fragments (scFv) were produced to capture specific proteins. A scFv is the variable region of a light chain and a heavy chain (**Figure 7**) combined into a single protein. The scFv have the antigen binding property of an immunoglobulin but can be recombinantly expressed. This means that manufacturing the scFv in a simpler organism that is easier to grow is possible [162-164], while recombinant expression of large proteins like immunoglobulins is hard due to incorrect folding of the protein [163, 164], but possible for a scFv.



**Figure 7.** The immunoglobulin is composed of four chains, two heavy chains and two light chains. The chains contains two regions, the constant (in red), and the variable (in blue). The fragment crystallisable (Fc) region (in green) is of importance for cell binding and activation. It is possible to recombinantly express the two variable regions as a single protein, which is termed a single-chain variable fragment (scFv).

## Conclusion

In this chapter, I have provided the first part of the biological system that my thesis revolve around. In the next chapter, I will provide information of the second part of the system, the pathogen. In combination with the pathogen, I will discuss the interactions that can occur when the pathogen invades a host.



## Chapter 4 – Pathogen and interactions

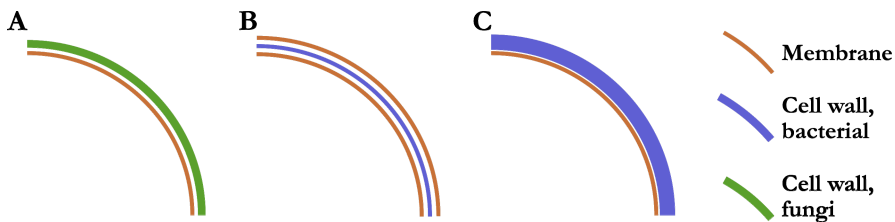
Pathogens are foreign agents that can enter a host and cause disease (an infection). Examples of agents are bacteria, fungi, parasites, virus and prions. In this thesis, I have focused on bacterial and fungal infections, which are the pathogens that are included in term pathogens from here on. Some strains of a pathogen are capable of causing disease, whereas others are less effective. This is termed virulence or pathogenicity and is a very diffuse term and difficult to measure. For example, many pathogens are opportunistic and only cause disease in immunosuppressed hosts [165-168]. How virulent are the strains and species causing disease in these hosts? Virulence should be seen more as a scale that is continuous rather than discrete and there is yet no objective measurement of virulence.

A pathogen can interact with the host in two major approaches [169]. The first is by secreting proteins, which can interact with the host at another location than the pathogen, and these secreted proteins are not covered within this thesis. The second mechanism is by direct interaction with the host via the surface of the pathogen, which is the theme of the **Original Papers**.

### Pathogen surface

In addition to the plasma membrane, most pathogens have extra physical barriers to protect from the environment. Yeasts (a type of fungi) are eukaryotic cells with a cell surface consisting of the obligatory plasma membrane and a fungal cell wall (**Figure 8A**). The fungal cell wall is constructed of chitin and polysaccharides [170]. Bacterial cells are divided in two groups depending on





**Figure 8.** The different pathogen surfaces and the structure of their cell envelope, all cells are enclosed with a plasma membrane to separate inside from outside of the cell. **A)** Fungi surface, are in combination with the membrane protected by a fungal cell wall. **B)** Gram-negative bacteria are enclosed with two membranes a plasma and an outer membrane, with a thin bacterial cell wall between these membranes. **C)** Gram-positive bacteria are in combination with the plasma membrane also enclosed in a thick bacterial cell wall.

the structure of their surface. The first group, gram-negative bacteria (**Figure 8B**) have a thin peptidoglycan layer followed by a second (outer) membrane. The second group, gram-positive bacteria (**Figure 8C**) have a thick peptidoglycan layer, three to ten times thicker than that of a gram-negative bacterium [171]. A peptidoglycan layer consists of polysaccharides fused with peptides that forms a rigid structure [171, 172]. These presented surface structures are the main groups, but the exact surface structure varies slightly depending on the species and the strain. Some pathogens have an extra shield to the environment, in the form of a capsule [173, 174]. However, the capsule also limits the pathogens surface proteins ability to interact with the environment [173, 174].

### Surface proteins

In combination with the membrane and cell wall structures, all cells have proteins on the surface that carry out several important mechanisms for the cell. After synthesis of a surface protein, the protein is transported from the cytosol to the cell envelope [175]. In gram-positive bacteria, this is achieved by a LPXTG motif using a group of proteins termed sortases [176], which covalently attaches the proteins to the cell wall. Examples of functions for surface proteins

are nutrient transportation across cell wall and membrane, and communication [177, 178]. Pathogens have also developed several mechanisms for interactions with the host. The surface proteins are of considerable importance when interacting with the environment, especially the interaction with the host as shown in **Paper III**.

### **Surface interactions**

Host-pathogen interactions occur in all infections and differ between most diseases and how the pathogen interacts with the host is crucial for the outcome of the disease. A pathogen can interact with the host in a multitude of approaches in addition to the aforementioned reasons of the immune system, and the reason for these interactions is wide. One example of the interactions is adherence, which enables the ability to disseminate within the host [179]. A second example is the ability to avoid the host's defences; this can be performed, for example by binding extensive amounts of host proteins, in order to hide in plain sight, making the pathogen not to look like a pathogen (**Paper II**), or by changing the appearance of the surface [180]. A third example of possible interaction is to hijack energy resources from the host to promote survival [8, 181]. For example, several pathogens have generated mechanisms to bind the immunoglobulin Fc region, hence, avoiding proper activation of the immune system [182-184]. Many interaction mechanisms are specific for every pathogen, and I will introduce some of the interactions with *Streptococcus pyogenes*, which is the central pathogen of this thesis.

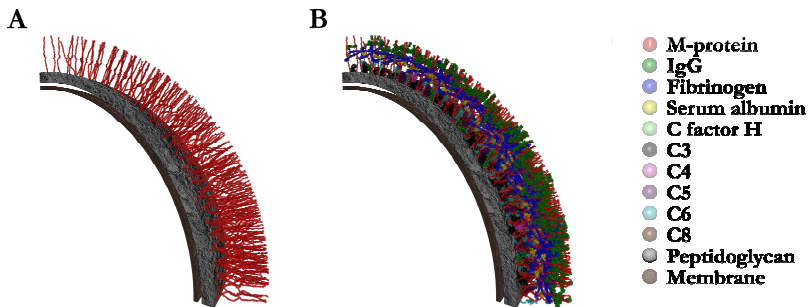
### ***Streptococcus pyogenes***

All the **Original papers** use the gram-positive (**Figure 8C**) bacterium *S. pyogenes*, which is exclusively used as pathogen in **Papers I, II** and **IV**. The reasons for selection of *S. pyogenes* are many. *S. pyogenes* is a human exclusive

pathogen and can cause a wide spectrum of diseases, from mild infections like impetigo and pharyngitis, to severe infections like necrotizing fasciitis and sepsis [185]. M-protein is one of the most abundant *S. pyogenes* proteins. In addition, M-protein has a hypervariable region, which is used to differentiate strains from each other [186-188] and strains of the M1 serotype are the most common *S. pyogenes* strains in invasive disease [185].

### Virulence factors

*S. pyogenes* has developed a multitude of virulence factors that aid in their invasion and survival in the host. The most abundant virulence factor on the surface of *S. pyogenes* is the M-protein [189]. M-protein, a coiled-coil dimer, is a 453 (for M1) amino acid long protein, which is anchored in the bacterial cell wall [190] and can bind several host proteins for immune evasion. For example, M-protein is able to bind fibrinogen and immunoglobulins (**Figure 9**) [182, 190]. The M-protein and fibrinogen binding has been studied in detail and the exact binding site of fibrinogen is in the middle of the protruding structure of M-protein [190]. One of the effects for the fibrinogen binding is complement activation interference [191], resulting in avoidance of phagocytosis. M-protein can also bind to the Fc region (**Figure 7**) of IgG, hindering the IgG to activate the immune



**Figure 9.** A three-dimensional model of the surface of *S. pyogenes*, without (A) and with plasma (B) adsorbed to the surface. The figure is constructed on MS data with absolute quantification of each of the present proteins in the legend, figure from **Paper II**. The density is determined by counting the number of bacteria using bottom-up MS.

system [182]. However, this seems to only manifest when the IgG concentration is low (e.g. in saliva) and not in plasma [192]. As several other *S. pyogenes* proteins, the M-protein also aids in the adherence and the colonisation of the host [185]. In **Paper II**, a model of *S. pyogenes* surface was created with the assumption that the M-protein was the only protein on the surface (**Figure 9**). There are several secreted *S. pyogenes* virulence factors. An example is EndoS, a glycosidase that specifically cleaves the glycosylation of IgG [193], interfering with the binding between the Fc-receptor and IgG. Another secreted protein is IdeS, a protease that can specifically cleave IgG, which divides it into two pieces [194] and leaves it capable to bind the target, but not having any Fc region to activate the immune system with [194]. The list of host-pathogen interaction is long, which is why I observe all of them at the same time.

## Host-pathogen interaction proteomics

By detecting what interacts with the surface, the function for those interactions can be devised. I believe it is only when the whole picture of the interactors is determined that the reason for the individual interaction can be resolved. The choice of mass spectrometer and the method of that instrument should be selected for answering the scientific question. Hence, it is important to know what is expected in the study. DIA provides more accurate information than DDA, but require considerably more instrument and analysis time. Consequently, if DDA would answer the question to an adequate level there is no requirement for DIA. The number of human plasma interactions for a pathogen ranges from 50 proteins to 150 (**Paper III**), which means that in most cases SRM is the method of choice for quantification. However, DDA and DIA are used in research questions where both interacting host proteins and the pathogen proteome are of interest. For example, in **Paper I**, DDA was used to identify the

proteins interacting with the host and the proteins that were enriched on the surface were selected for further analysis by SRM. In **Paper II** and **IV**, the proteins that were studied were selected based on **Paper I** and were absolutely quantified using SRM. In **Paper III**, the pathogen and the interacting proteins were both identified and quantified in a proteome wide fashion using DDA and DIA, respectively, covering around 1000 to 3000 proteins per pathogen.

### **Extend the knowledge**

Some components of the host's immune system act at the surface of the pathogen and have well-established functions and mechanisms. However, these functions and mechanisms have mainly been studied as single entities. I believe that it is of the outmost importance that these interactions are studied simultaneously to find covariance, and this information is currently lacking. This additional information will help in drawing the correct conclusions about the single interactions, by seeing the whole picture. These conclusions will in turn help the hypothesis development for diagnostics and therapeutics of pathogen related diseases. In addition, the focus of the **Original papers** is how the human blood plasma interacts with the pathogen, rather than the other way around. Hence, future development of therapeutics that manipulate the interactions between the specific plasma proteins and the pathogen, which might alter the outcome of the disease beneficially might not be possible without knowledge of these interactions.

In **Paper III**, the difference in surface binding of plasma between 12 different pathogens was analysed. These pathogens cover a wide range of pathogens that cause sepsis, including yeast, gram-positive and gram-negative bacteria. This study demonstrates the interactions dependence on specific strains, species or pathogen group, which was accomplished by the holistic approach of

proteomics (for examples, see **Chapter 5**). In addition, a multitude of mechanisms that pathogens can be virulent to the host was shown.

## **Conclusion**

In this chapter, I have established the second part of the biological system in this thesis. In addition, I have provided the reasoning for working with proteomics, in combination with how and why I used the different MS methods in the **Original Papers**. In the next and last chapter, I will present how I work with big data and some of my thoughts on the proteomics field.



## Chapter 5 – A holistic approach

The postulated question in the **Original papers** was analysed using proteomics, the measurement of all proteins present in a sample. Proteomics is a holistic approach to protein analysis. One of the main differences of a holistic approach in comparison to classical approaches is the hypothesis. Instead of having a hypothesis for a specific case and executing experiments to find whether the hypothesis is true or not, the hypothesis is more open in proteomics. Hence, the hypothesis is formed based on resulting information in the experiment and further developed based the gained information. I prefer this approach of creating and developing hypotheses, since it is biology guided. However, this might guide us astray from the original objective of the research. Clearly, the scientific community benefits from and requires both holistic and classical approaches. A holistic approach to biology is known as systems biology [195], for example by observing everything in a system, the single parts and their relation with each other. The **Original Papers** focuses on detecting all the protein interactions between human blood plasma and pathogens, with the aim to determine the few that affects the disease.

Systems biology research strategies commonly generates large amount of data and the data must be reduced for presentation. The question arises of why so much is measured, if not all data is used. The answer is because biology is complex and it is not always possible to simplify biology. System biology is a holistic approach to biology but it still reduces the biology to a system. This system is used as a model to study a larger system. In the case of this thesis, the major aim was to investigate the molecular mechanisms of blood plasma



interacting with pathogen, as a model for sepsis after pathogen invasion of the bloodstream. The system of the thesis is a much larger system than looking at a particular protein-protein interaction. However, sepsis is considerably more complicated than the model presented here, but it is a part of sepsis. The exact molecular mechanisms of sepsis as a whole are far too many and too complicated for the techniques currently available. The reason for the particular focus of blood plasma interactions was based on the assumption that this particular model system falls within the realm of the technical capabilities. I believe it is important to expand the system as soon as the technology allows it. However, I have progressed as far as the technology available allows it.

### **Data reduction and visualisation**

The standard MS analysis pipeline attempts to deduce proteins from spectra. However, the data analysis is far from finalised at that time point. There is still a demand of analysing and presenting the proteins with a relation to biology. The systems biology data is complicated and the reduction of data is essential, since the presentation of data should be practical and visually attractive. The reduction of data can (and should) be helped by statistical analysis of replicates and sample comparison. There are several software available to help with reduction and visualisation of systems biology data [196]. One example that was developed for MS data is Protter [197], a software that integrates the biological annotation and visualises the data. However, the majority of the data reduction in the **Original papers** was performed using statistical tests to focus on the significant biological differences.

Reducing the data without losing any information is impossible. Nevertheless, choosing the data relevant for presentation is the object of the writer. The writer is however, subjective to the study and has to show that the

presented data is true. The readers of a paper should still be supplied with the possibility to view the original data.

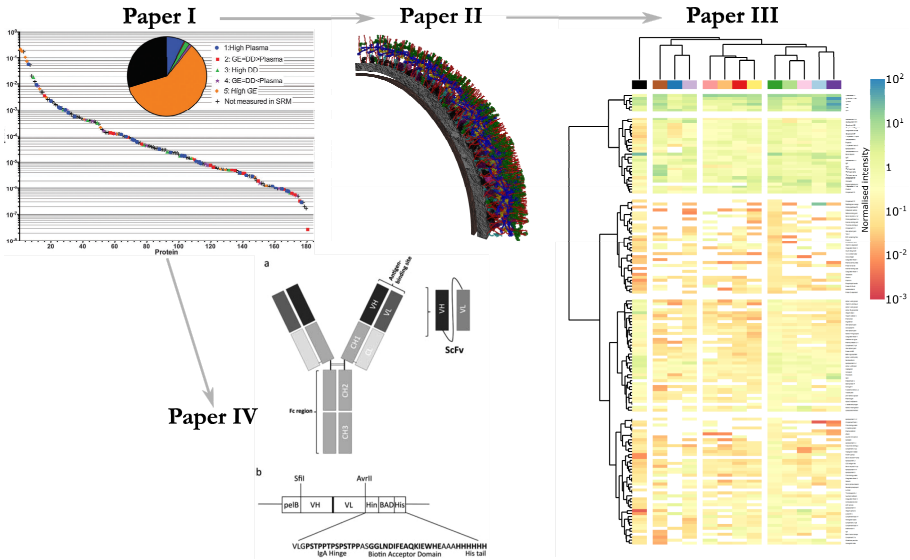
## Future

There are some interesting aspects to tackle for the field of proteomics in the future. The techniques that are available enables the possibility to measure a large number of proteins, this number can however be increased by developments in technology and informatics. In addition, the peptide coverage of proteins can also be increased by further development of ionisation and usage of more ions. The dynamic range of samples, in particular samples deriving from plasma, remains an issue. There are constant developments in instrumentation and bioinformatics, which will continue, so that the scientist can postulate proper hypotheses. I believe that the field of top-down proteomics will emerge as a competitive proteomics field, as the field does not have issues with mapping peptide to protein. In addition, the problem with multiple PTM is not possible to solve in any other MS approach. However, the top-down of today is quite far from being proteome wide.

The greater goal for most research is to be implemented in life. In my research, the ultimate goal is to develop diagnostics and therapeutics for pathogen related diseases. My presented research is not close of being implemented in diagnostics and therapeutics, which was never the aim. The aim was to *support the development of hypotheses*, and I hope that in the future my research will help scientist to develop diagnostics and therapeutics. However, there are several novelties and interesting findings in the **Original papers**.

# The Original papers

This thesis is based on four **Original papers**, which summarises my research. The relations between the four papers are displayed in **Figure 10**. All of the **Original papers** have their own story, but they also connected each other and the scientific question has progressed from the previous findings.



**Figure 10.** Overview of relations between the **Original papers** that are included in my thesis. **Paper I** is the start of the thesis, which resulted in assays that were further used and analysed in **Paper II** and **IV**. In **Paper II**, the assays were refined and proteins were absolutely quantified to create a stoichiometric density model. In **Paper IV**, the assays were used in a different system to evaluate the binding properties of several scFv. These papers focused on *S. pyogenes* as model organism for the pathogen, which was expanded in **Paper III** to 12 pathogens.

## Paper I

The aim of this study was to determine all the interactions between *S. pyogenes* and human blood plasma. DDA was used to identify the proteins and label-free quantification (MS1 quantification) of was performed for relative quantification to decide the importance of the interactions. To obtain a

comprehensive list of interactions, the samples were separated off-line before MS measurement, which resulted in defining 181 interactions between plasma and *S. pyogenes*. From this list of interactions, surface enriched interactions were further characterised. The enrichment was calculated depending on the intensity at the surface and the intensity in plasma. The proteins with high enrichment and/or high concentration at the bacterial surface were subsequently measured using SRM, to quantify the enrichment at the surface more accurately than with DDA. In combination with quantifying most of the previously known interactions between plasma and *S. pyogenes*, 87 unknown interactions were validated.

In follow-up experiments, the differences in plasma protein binding to a *S. pyogenes* wild-type strain and an isogenic mutant lacking several important virulence factors were analysed. In addition, a unique pair of invasive and non-invasive *S. pyogenes* isolates from the same patient were analysed. When comparing the plasma protein-binding properties of the wild type with the mutant and the invasive with non-invasive considerable differences were revealed. The result underlines the significance of these protein interactions and the demand of studying them simultaneously.

## **Paper II**

In **Paper II**, a method was developed for visualising the interaction between *S. pyogenes* and human blood plasma at a molecular level. The aim of **Paper II** was to replicate a visualisation strategy (molecular anatomy) described by Takamori et al in 2006 [198], with host-pathogen interaction data. A molecular anatomy is an accurate three-dimensional model of the surface. The molecular anatomy includes stoichiometric relations between the host proteins and the surface proteins of the pathogen, and includes the density of the surface proteins. The accuracy in the measurements for creating a model demanded absolute quantification, which was accomplished with SRM. In addition, a novel MS-based

technique for counting the number of bacteria using SRM was developed, which enabled determination of protein density at the surface. The main plasma interactions (16 proteins) at the surface of *S. pyogenes* in combination with the M-protein were absolutely quantified. Hence, the stoichiometric ratios between the pathogen and the surface were determined.

This resulted in the first molecular model of the surface of a pathogen interacting with a host in a global scale (**Figure 9**), which I named a “stoichiometric surface density model”. This is one method of presenting proteomics data for visualising of surface interactions. I believe that the development of novel approaches to better present data is of major importance for conveying the message of large-scale proteomics. Furthermore, visualisation is an important step to enhance the understanding of host-pathogen interactions.

### **Paper III**

The result in **Paper II** was obtained using MS techniques to determine host-pathogen interactions in a proteomics setup for one pathogen with extreme detail. However, there has never been a multi-pathogen approach to host-pathogen interaction using proteomics. In the **Paper III** manuscript, twelve different sepsis-causing pathogens were selected to account for the majority of sepsis causing pathogens. The pathogens biological differences were wide including yeasts, gram-positive and gram-negative bacteria. In addition, four strains of each pathogen were selected, two invasive and two non-invasive. The selection of pathogens was based on the aim of finding pathogen and strain specific interactions. The aim was especially related to immune system interactions and the mechanisms different pathogens have developed to avoid the effects of the immune system.

The interactions of plasma proteins to the surface of *S. pyogenes* was known from **Paper I**. However, the interactions for the other pathogens are not known. In addition, the sample preparation protocol developed in **Paper II** resulted in a sample containing both the pathogens proteins and the interacting plasma proteins. The protocol was specifically developed for bottom-up proteomics with host-pathogen interactions, and resulted in a sample where both the pathogen's proteome and the interacting plasma proteins could be quantified. DDA was used to define the proteomes of the pathogens and the plasma proteins interacting with the pathogen's surfaces. Assay libraries were created in several steps, the samples was measured using DDA, the spectra were scored using X!Tandem, assay libraries were created and filtered using multilevel FDR. Assay libraries were created for each of the pathogens and the interacting plasma proteins. Quantification of the whole proteomes were performed using DIA. This resulted in quantification of 17'128 pathogen protein and 148 human proteins that interacted with the pathogens. The 148 human proteins were manually curated from over 700 protein groups to avoid issues with peptides matching multiple proteins.

This study is currently the first and the largest host-pathogen interaction study performed on multiple species. In this large dataset, several important discoveries were attained that would be impossible to find if the dataset was smaller. One example is the binding of fibrinogen to the surface of *S. pyogenes*, which has been documented previously (**Paper I, II**, and others). This particular binding is a known defence mechanism of *S. pyogenes* and fibrinogen binds extensively to the M-protein. However, this has not been defined in the context of other pathogens. We could clearly verify interactions with fibrinogen as a *S. pyogenes* specific defence mechanism. This is only one of the results in the paper

that shows the power of working with holistic approaches in both experimental setup and in proteomics.

### **Paper IV**

The aim of the study in **Paper IV** was to present an antibody format suited for detection and analysis of IgG-binding virulence factors. As proof of concept, scFv's were produced that specifically target the IgG-binding surface proteins of *S. pyogenes* (M1-protein and protein H). The scFv are particularly well suited for IgG-binding proteins since the scFv do not have issues with Fc binding. In total, four scFv (two against each protein) were selected for MS analysis. Assays generated in **Paper I** was used for validating the binding properties of the scFv, which were subsequently measured in the absence and presence of plasma. In addition, the epitope of the scFv against M1-protein were determined.

The binding ability of the scFv was demonstrated with both recombinant protein and at the surface of *S. pyogenes*. The scFv was also able to enrich M1-protein in the presence of plasma. A potential clinical application for the scFv is to enrich virulence factors or whole bacteria from environments that are too complex for direct analysis. An application example is direct identification of bacteria in blood without the demand of growing the bacteria on an agar plate.

### **Conclusions and relevance**

The presented work in this thesis highlights the demand of proteomics in host-pathogen interactions. The research accomplished in the **Original papers** has defined many new interactions between pathogen and plasma. In addition, both novel interactions and previously known interactions have been quantified. The take-home message from the **Original papers** is that they demonstrate that the large-scale approach can define specific interactions that would be impossible

to define without large-scale analysis. In addition, there are several conclusions of the separate **Original papers** that can be drawn when combining the **Original papers**. I will provide one example below.

In **Paper II**, IgG3 was defined as a specific binder of M1-protein, and enriched with a higher extent at the surface of *S. pyogenes* than other immunoglobulins. In **Paper III**, the IgG3 binding phenomenon was observed for several other M-protein strains of *S. pyogenes*, determining this as a general M-protein binding mechanism. The extreme IgG3 binding was also determined as *S. pyogenes* specific, since none of the other eleven pathogens showed this ability. This phenomenon would not be possible without the specificity and accuracy of MS, or without the holistic approach that is possible within bottom-up MS, or the holistic approach to the scientific question that I have in my **Original papers**.

The work within the **Original Papers** provides the next step towards a holistic view of host-pathogen interactions, which are far from being complete. In **Paper I**, only *S. pyogenes* represented the pathogen, this was later expanded in **Paper III**, but there are still several pathogens left to investigate. In addition, the interactions between the pathogen and host cells are lacking a proteomics perspective, possibly due to the lack of tools available for this analysis. The demand for experimental and technical development is constant and the more tools that are available the more scientific questions have the possibility of being answered. However, scientists should not refrain from using and developing new tools to analyse data. Biology is complex and the quest of seeking an answer to how biology works might not always be easy, but the complexity should not hinder us from trying. I hope that the tools and approaches I have developed in the **Original Papers** will benefit the scientific community within host-pathogen interactions, so that my research will indirectly benefit society as a whole.





# Populärvetenskaplig sammanfattning

Alla organismer är uppbyggda av små beståndsdelar som kallas för celler, vilka består av en mängd molekyler med olika funktioner. I min avhandling fokuserar jag på en grupp av molekyler, proteiner, vilket är de som utför större delen av arbetet i cellen. Det finns många typer av celler, i avhandlingen arbetar jag med bakterier och svampar som kan orsaka en infektion när de invaderar en människa (värden), dessa celler kallas då patogener. Infektioner är normalt begränsade till en lokal plats i värden som till exempel i ett sår, men när en patogen lyckas invadera blodsystemet så kan en systemisk infektion utvecklas. Detta tillstånd kallas för sepsis och är en extremt allvarlig sjukdom med en dödlighet på ca 20%. Den exakta orsaken till varför vissa patogener kan invadera blodsystemet och andra inte är i nuläget okänd. Målet med denna avhandling är att belysa detta problem samt undersöka de bakomliggande mekanismerna.

För att en patogen ska kunna infektera en värd krävs att proteiner på ytan av patogenen interagerar med proteiner från värden. I denna avhandling så fokuserar jag på värdens blodproteiner och dess interaktion med olika patogener. Målet med avhandlingen är att mäta alla blodproteiner som interagerar med patogenen. Kärnan i avhandlingen är proteomik, vilket är studien om alla proteiner i ett system.

Inom proteomik används en maskin vilken fungerar som en väldigt exakt och snabb våg, en så kallad masspektrometer. Genom att mäta vikten (massan) på många proteiner från ett prov kan vi avgöra vilka proteiner som fanns i ett prov. Jag har utvecklat tekniker för att bestämma mängden av varje specifikt protein i provet, vilket ger oss ett inbindningsmönster för den patogenen.

Eftersom jag mäter flera tusen proteiner är att det är svårt att visa all information. Avhandlingen har ett stort fokus på hur information kan visualiseras för att underlätta hur den förmedlas på bästa sätt. Jag har bland annat visualiserat inbindningsmönstret genom att skapa den första datormodellen av ytan av en patogen när den interagerar med blodproteiner.

Genom att studera olika patogener så kan inbindningsmönstret av blodproteiner för varje patogen analyseras för skillnader. Det kan vara mellan en patogen som orsakar en mild sjukdom såsom faryngit (halsfluss) och en annan som orsakar sepsis. Inbindningsmönstret av värd-patogen interaktioner har i denna avhandling studerats i mer detalj än tidigare. Hypotesen för min forskning är att titta på inbindningsmönstret så kan man se individuella skillnader, mellan specifika proteiner, vilket kan vara av stor vikt för att definiera sjukdomsförloppet. Det viktiga här är att hela systemet måste mätas för att det är endast när alla delar studeras som man kan se hur varje del varierar i relation till varandra.

## Acknowledgement

Many persons have helped me throughout this thesis and it is not possible to mention all in this small acknowledgement. To those that are not mentioned, thanks!

One person has meant more to me than any other for my understanding of science is my supervisor **Johan**. Tack för att du övertalade mig att doktorera efter min master, ett beslut som jag aldrig kommer ångra. För mig är din ledarstil perfekt och jag kommer aldrig ha en lika bra chef igen.

Till min biträdande handledare **Christofer**, som alltid föregår med gott exempel när det kommer till att bedriva utsökt forskning. Tack för allt du lärt mig under åren och för att din dörr alltid varit öppen för att få hjälp när jag har haft någon fråga.

To the amazing people in **IMP**, you are the best! Thanks for all the fantastic discussions at lunch (especially the non-scientific). Till **Ola** som jag har följt under alla mina år i gruppen, tack för ditt stöd och för alla fantastiska stunder utanför labbet. Tack för att du såg till att en av de lataste personerna jag vet (mig själv) började träna och sluta aldrig att äta upp resterna av min mat! **Simon** that taught me about mass spectrometry and how to care for an instrument, otherwise I would still be fixing leaks or calibrate Harry. Thanks for showing me how all of Switzerland looks like, even though we were in the states. **Lotta** and **Anahita**, thanks for being amazing office mates and thanks for switching to Finnish when you talk *shit* about me. **Lotta** dina kunskaper inom vetenskap har hjälpt mig minst lika mycket som mina datorkunskaper har hjälpt dig! **Anahita, Ola** and **Simon**

thanks for all the GoT evenings, but when are those dragons going to be used properly?

To all the people of **Immunotechnology**, even though I have not been around much in the department the last couple of years, it is always fantastic when I am. The social events are always wonderful. To the prefects **Carl** and **Sara** thanks for creating the amazing atmosphere in the department, and **Marianne**, **Cornelia** and **Tommie** for simplifying the administrative work for everyone else. Tack **Malin** och **Cornelia** för hjälpen med det administrativa runt avhandlingen. **Brofelth** när ska du sluta stalka mig i yrkeslivet? Helst aldrig om du frågar mig. **Andy** och **Tim** som aldrig skippar att lägga en stund på att snacka fotboll. To everyone whom got a tackle during football, sorry, if I had the chance I would do it again.

**Johan**, **Ola** and **Ufuk**, vad sägs om att minnas tillbaka alla trevliga stunder i Japan med en okonomiyaki kväll?

Alla mina vänner har jag mycket att tacka för mina val vilket fört mig dit där jag är idag! **Roos** och **Marris** tack för att ni har tagit mitt sinne från vetenskapen under doktorandtiden och dragit in mig i den förtrollade värden av bondgårdsbyggande.

Till min familj, tack för att ni i alla fall försöker förstå vad jag sysslar med. **Mamma** och **Pappa**, ert stöd är oerhört viktigt, även om jag kanske inte alltid visar det. **Emmeli** du är en fantastisk syster, din familj med **Johan** och **Astrid** är alltid fantastiska att träffa.

**Sara** som ständigt får mig att minnas att vetenskapen är sekundär, du är alltid först.

## References

1. Bennett, G.M. and N.A. Moran, *Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a Phloem-feeding insect*. *Genome Biol Evol*, 2013. **5**(9): p. 1675-88.
2. Brenchley, R., et al., *Analysis of the bread wheat genome using whole-genome shotgun sequencing*. *Nature*, 2012. **491**(7426): p. 705-10.
3. Geyer, P.E., et al., *Plasma Proteome Profiling to Assess Human Health and Disease*. *Cell Syst*, 2016. **2**(3): p. 185-95.
4. Khoury, G.A., R.C. Baliban, and C.A. Floudas, *Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database*. *Sci Rep*, 2011. **1**.
5. Ponomarenko, E.A., et al., *The Size of the Human Proteome: The Width and Depth*. *Int J Anal Chem*, 2016. **2016**: p. 7436849.
6. Wasinger, V.C., et al., *Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium*. *Electrophoresis*, 1995. **16**(7): p. 1090-4.
7. James, P., *Protein identification in the post-genome era: the rapid rise of proteomics*. *Q Rev Biophys*, 1997. **30**(4): p. 279-331.
8. Malmström, J., et al., *Streptococcus pyogenes in human plasma: adaptive mechanisms analyzed by mass spectrometry-based proteomics*. *J Biol Chem*, 2012. **287**(2): p. 1415-25.
9. Malmstrom, E., et al., *Large-scale inference of protein tissue origin in gram-positive sepsis plasma using quantitative targeted proteomics*. *Nat Commun*, 2016. **7**: p. 10261.
10. Hebert, A.S., et al., *The one hour yeast proteome*. *Mol Cell Proteomics*, 2014. **13**(1): p. 339-47.
11. Wingren, C., *Antibody-Based Proteomics*. *Adv Exp Med Biol*, 2016. **926**: p. 163-179.
12. Aebersold, R. and M. Mann, *Mass spectrometry-based proteomics*. *Nature*, 2003. **422**(6928): p. 198-207.
13. Rabilloud, T. and C. Lelong, *Two-dimensional gel electrophoresis in proteomics: a tutorial*. *J Proteomics*, 2011. **74**(10): p. 1829-41.
14. Aebersold, R., A.L. Burlingame, and R.A. Bradshaw, *Western blots versus selected reaction monitoring assays: time to turn the tables?* *Mol Cell Proteomics*, 2013. **12**(9): p. 2381-2.
15. Pradidarcheep, W., et al., *Lack of specificity of commercially available antisera: better specifications needed*. *J Histochem Cytochem*, 2008. **56**(12): p. 1099-111.

16. Beermann, S., R. Seifert, and D. Neumann, *Commercially available antibodies against human and murine histamine H(4)-receptor lack specificity*. Naunyn Schmiedebergs Arch Pharmacol, 2012. **385**(2): p. 125-35.
17. Elliott, K.J., K. Kimura, and S. Eguchi, *Lack of specificity of commercial antibodies leads to misidentification of angiotensin type-1 receptor protein*. Hypertension, 2013. **61**(4): p. e31.
18. Chait, B.T., *Chemistry. Mass spectrometry: bottom-up or top-down?* Science, 2006. **314**(5796): p. 65-6.
19. Yates, J.R., C.I. Ruse, and A. Nakorchevsky, *Proteomics by mass spectrometry: approaches, advances, and applications*. Annu Rev Biomed Eng, 2009. **11**: p. 49-79.
20. Gillet, L.C., A. Leitner, and R. Aebersold, *Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing*. Annu Rev Anal Chem (Palo Alto Calif), 2016. **9**(1): p. 449-72.
21. Vandermarliere, E., M. Mueller, and L. Martens, *Getting intimate with trypsin, the leading protease in proteomics*. Mass Spectrom Rev, 2013. **32**(6): p. 453-65.
22. Siuti, N. and N.L. Kelleher, *Decoding protein modifications using top-down mass spectrometry*. Nat Methods, 2007. **4**(10): p. 817-21.
23. Moradian, A., et al., *The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications*. Proteomics, 2014. **14**(4-5): p. 489-97.
24. Bae, Y.J. and M.S. Kim, *A Thermal Mechanism of Ion Formation in MALDI*. Annu Rev Anal Chem (Palo Alto Calif), 2015. **8**: p. 41-60.
25. Karas, M. and F. Hillenkamp, *Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons*. Anal Chem, 1988. **60**(20): p. 2299-301.
26. Trimpin, S., et al., *Matrix-assisted laser desorption/ionization mass spectrometry method for selectively producing either singly or multiply charged molecular ions*. Anal Chem, 2010. **82**(1): p. 11-5.
27. Kebarle, P., *A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry*. J Mass Spectrom, 2000. **35**(7): p. 804-17.
28. Fenn, J.B., et al., *Electrospray ionization for mass spectrometry of large biomolecules*. Science, 1989. **246**(4926): p. 64-71.
29. Konermann, L., et al., *Unraveling the mechanism of electrospray ionization*. Anal Chem, 2013. **85**(1): p. 2-9.
30. Medley, S.S., *Energetic ion mass analysis using a radio-frequency quadrupole filter*. Rev Sci Instrum, 1978. **49**(6): p. 698.
31. Schwartz, J.C., M.W. Senko, and J.E. Syka, *A two-dimensional quadrupole ion trap mass spectrometer*. J Am Soc Mass Spectrom, 2002. **13**(6): p. 659-69.
32. Vestal, M.L. and J.M. Campbell, *Tandem time-of-flight mass spectrometry*. Methods Enzymol, 2005. **402**: p. 79-108.

33. Guilhaus, M., *Principles and Instrumentation in Time-of-flight Mass Spectrometry*. JOURNAL OF MASS SPECTROMETRY, 1995. **30**: p. 1519-1532.
34. Nicolardi, S., et al., *Developments in FTICR-MS and Its Potential for Body Fluid Signatures*. Int J Mol Sci, 2015. **16**(11): p. 27133-44.
35. Marshall, A.G., C.L. Hendrickson, and G.S. Jackson, *Fourier transform ion cyclotron resonance mass spectrometry: a primer*. Mass Spectrom Rev, 1998. **17**(1): p. 1-35.
36. Hu, Q., et al., *The Orbitrap: a new mass spectrometer*. J Mass Spectrom, 2005. **40**(4): p. 430-43.
37. Zubarev, R.A. and A. Makarov, *Orbitrap mass spectrometry*. Anal Chem, 2013. **85**(11): p. 5288-96.
38. Guan, S. and A.G. Marshall, *Stored waveform inverse Fourier transform axial excitation/ejection for quadrupole ion trap mass spectrometry*. Anal Chem, 1993. **65**(9): p. 1288-94.
39. Aebersold, R. and D.R. Goodlett, *Mass spectrometry in proteomics*. Chem Rev, 2001. **101**(2): p. 269-95.
40. Mann, M., R.C. Hendrickson, and A. Pandey, *Analysis of proteins and proteomes by mass spectrometry*. Annu Rev Biochem, 2001. **70**: p. 437-73.
41. Harrison, A.G., *To b or not to b: the ongoing saga of peptide b ions*. Mass Spectrom Rev, 2009. **28**(4): p. 640-54.
42. Sleno, L. and D.A. Volmer, *Ion activation methods for tandem mass spectrometry*. J Mass Spectrom, 2004. **39**(10): p. 1091-112.
43. Roepstorff, P. and J. Fohlman, *Proposal for a common nomenclature for sequence ions in mass spectra of peptides*. Biomed Mass Spectrom, 1984. **11**(11): p. 601.
44. Good, D.M., et al., *Performance characteristics of electron transfer dissociation mass spectrometry*. Mol Cell Proteomics, 2007. **6**(11): p. 1942-51.
45. Zubarev, R.A., et al., *Electron capture dissociation for structural characterization of multiply charged protein cations*. Anal Chem, 2000. **72**(3): p. 563-73.
46. Sobott, F., et al., *Comparison of CID versus ETD based MS/MS fragmentation for the analysis of protein ubiquitination*. J Am Soc Mass Spectrom, 2009. **20**(9): p. 1652-9.
47. Li, X., et al., *The effect of fixed charge modifications on electron capture dissociation*. J Am Soc Mass Spectrom, 2008. **19**(10): p. 1514-26.
48. McLafferty, F.W., et al., *Electron capture dissociation of gaseous multiply charged ions by Fourier-transform ion cyclotron resonance*. J Am Soc Mass Spectrom, 2001. **12**(3): p. 245-9.
49. Wells, J.M. and S.A. McLuckey, *Collision-induced dissociation (CID) of peptides and proteins*. Methods Enzymol, 2005. **402**: p. 148-85.
50. Olsen, J.V., et al., *Higher-energy C-trap dissociation for peptide modification analysis*. Nat Methods, 2007. **4**(9): p. 709-12.
51. McLuckey, S.A., *Principles of collisional activation in analytical mass spectrometry*. J Am Soc Mass Spectrom, 1992. **3**(6): p. 599-614.



52. Shukla, A.K. and J.H. Futrell, *Tandem mass spectrometry: dissociation of ions by collisional activation*. *J Mass Spectrom*, 2000. **35**(9): p. 1069-90.
53. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data*. *Electrophoresis*, 1999. **20**(18): p. 3551-67.
54. Sadygov, R.G., D. Cociorva, and J.R. Yates, 3rd, *Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book*. *Nat Methods*, 2004. **1**(3): p. 195-202.
55. Choi, H., D. Ghosh, and A.I. Nesvizhskii, *Statistical validation of peptide identifications in large-scale proteomics using the target-decoy database search strategy and flexible mixture modeling*. *J Proteome Res*, 2008. **7**(1): p. 286-92.
56. Shevchenko, A., et al., *Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels*. *Anal Chem*, 1996. **68**(5): p. 850-8.
57. de Godoy, L.M., et al., *Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast*. *Nature*, 2008. **455**(7217): p. 1251-4.
58. Krokhin, O.V., et al., *An improved model for prediction of retention times of tryptic peptides in ion pair reversed-phase HPLC: its application to protein peptide mapping by off-line HPLC-MALDI MS*. *Mol Cell Proteomics*, 2004. **3**(9): p. 908-19.
59. Mann, M., et al., *The coming age of complete, accurate, and ubiquitous proteomes*. *Mol Cell*, 2013. **49**(4): p. 583-90.
60. Richards, A.L., A.E. Merrill, and J.J. Coon, *Proteome sequencing goes deep*. *Curr Opin Chem Biol*, 2015. **24**: p. 11-7.
61. Zubarev, R.A., *The challenge of the proteome dynamic range and its implications for in-depth proteomics*. *Proteomics*, 2013. **13**(5): p. 723-6.
62. Nagaraj, N., et al., *Deep proteome and transcriptome mapping of a human cancer cell line*. *Mol Syst Biol*, 2011. **7**: p. 548.
63. Aebersold, R. and M. Mann, *Mass-spectrometric exploration of proteome structure and function*. *Nature*, 2016. **537**(7620): p. 347-55.
64. Picotti, P., et al., *A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis*. *Nature*, 2013. **494**(7436): p. 266-70.
65. Gillet, L.C., et al., *Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis*. *Mol Cell Proteomics*, 2012. **11**(6): p. O111 016717.
66. Domon, B. and R. Aebersold, *Mass spectrometry and protein analysis*. *Science*, 2006. **312**(5771): p. 212-7.
67. Ong, S.E. and M. Mann, *Mass spectrometry-based proteomics turns quantitative*. *Nat Chem Biol*, 2005. **1**(5): p. 252-62.
68. Old, W.M., et al., *Comparison of label-free methods for quantifying human proteins by shotgun proteomics*. *Mol Cell Proteomics*, 2005. **4**(10): p. 1487-502.
69. Brun, V., et al., *Isotope-labeled protein standards: toward absolute quantitative proteomics*. *Mol Cell Proteomics*, 2007. **6**(12): p. 2139-49.

70. Bantscheff, M., et al., *Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present*. Anal Bioanal Chem, 2012. **404**(4): p. 939-65.
71. Annesley, T.M., *Ion suppression in mass spectrometry*. Clin Chem, 2003. **49**(7): p. 1041-4.
72. Ross, P.L., et al., *Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents*. Mol Cell Proteomics, 2004. **3**(12): p. 1154-69.
73. Thompson, A., et al., *Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS*. Anal Chem, 2003. **75**(8): p. 1895-904.
74. Ong, S.E., et al., *Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics*. Mol Cell Proteomics, 2002. **1**(5): p. 376-86.
75. Gerber, S.A., et al., *Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS*. Proc Natl Acad Sci U S A, 2003. **100**(12): p. 6940-5.
76. Domon, B. and R. Aebersold, *Options and considerations when selecting a quantitative proteomics strategy*. Nat Biotechnol, 2010. **28**(7): p. 710-21.
77. Picotti, P. and R. Aebersold, *Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions*. Nat Methods, 2012. **9**(6): p. 555-66.
78. McCormack, A.L., et al., *Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level*. Anal Chem, 1997. **69**(4): p. 767-76.
79. Meissner, F. and M. Mann, *Quantitative shotgun proteomics: considerations for a high-quality workflow in immunology*. Nat Immunol, 2014. **15**(2): p. 112-7.
80. MacLean, B., et al., *Skyline: an open source document editor for creating and analyzing targeted proteomics experiments*. Bioinformatics, 2010. **26**(7): p. 966-8.
81. Lange, V., et al., *Selected reaction monitoring for quantitative proteomics: a tutorial*. Mol Syst Biol, 2008. **4**: p. 222.
82. Teleman, J., et al., *Automated selected reaction monitoring software for accurate label-free protein quantification*. J Proteome Res, 2012. **11**(7): p. 3766-73.
83. Rost, H., L. Malmstrom, and R. Aebersold, *A computational tool to detect and avoid redundancy in selected reaction monitoring*. Mol Cell Proteomics, 2012. **11**(8): p. 540-9.
84. Anderson, L. and C.L. Hunter, *Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins*. Mol Cell Proteomics, 2006. **5**(4): p. 573-88.
85. Picotti, P., et al., *Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics*. Cell, 2009. **138**(4): p. 795-806.
86. Malmstrom, L., et al., *Quantitative proteogenomics of human pathogens using DLA-MS*. J Proteomics, 2015. **129**: p. 98-107.

87. Geiger, T., J. Cox, and M. Mann, *Proteomics on an Orbitrap benchtop mass spectrometer using all-ion fragmentation*. Mol Cell Proteomics, 2010. **9**(10): p. 2252-61.
88. Venable, J.D., et al., *Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra*. Nat Methods, 2004. **1**(1): p. 39-45.
89. Egertson, J.D., et al., *Multiplexed peptide analysis using data-independent acquisition and Skyline*. Nat Protoc, 2015. **10**(6): p. 887-903.
90. Rost, H.L., et al., *OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data*. Nat Biotechnol, 2014. **32**(3): p. 219-23.
91. Teleman, J., et al., *DLANA--algorithmic improvements for analysis of data-independent acquisition MS data*. Bioinformatics, 2015. **31**(4): p. 555-62.
92. Tsou, C.C., et al., *DLA-Umpire: comprehensive computational framework for data-independent acquisition proteomics*. Nat Methods, 2015. **12**(3): p. 258-64, 7 p following 264.
93. Reiter, L., et al., *mProphet: automated data processing and statistical validation for large-scale SRM experiments*. Nat Methods, 2011. **8**(5): p. 430-5.
94. Marcotte, E.M., *How do shotgun proteomics algorithms identify proteins?* Nat Biotechnol, 2007. **25**(7): p. 755-7.
95. Mardis, E.R., *A decade's perspective on DNA sequencing technology*. Nature, 2011. **470**(7333): p. 198-203.
96. Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification*. Nat Biotechnol, 2008. **26**(12): p. 1367-72.
97. Deutsch, E.W., et al., *A guided tour of the Trans-Proteomic Pipeline*. Proteomics, 2010. **10**(6): p. 1150-9.
98. Craig, R. and R.C. Beavis, *TANDEM: matching proteins with tandem mass spectra*. Bioinformatics, 2004. **20**(9): p. 1466-7.
99. Keller, A., et al., *Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search*. Anal Chem, 2002. **74**(20): p. 5383-92.
100. Nesvizhskii, A.I., et al., *A statistical model for identifying proteins by tandem mass spectrometry*. Anal Chem, 2003. **75**(17): p. 4646-58.
101. Liu, H., R.G. Sadygov, and J.R. Yates, 3rd, *A model for random sampling and estimation of relative protein abundance in shotgun proteomics*. Anal Chem, 2004. **76**(14): p. 4193-201.
102. Neilson, K.A., et al., *Less label, more free: approaches in label-free quantitative mass spectrometry*. Proteomics, 2011. **11**(4): p. 535-53.
103. Taylor, J.A. and R.S. Johnson, *Sequence database searches via de novo peptide sequencing by tandem mass spectrometry*. Rapid Commun Mass Spectrom, 1997. **11**(9): p. 1067-75.
104. Dancik, V., et al., *De novo peptide sequencing via tandem mass spectrometry*. J Comput Biol, 1999. **6**(3-4): p. 327-42.

105. Mo, L., et al., *MSNovo: a dynamic programming algorithm for de novo peptide sequencing via tandem mass spectrometry*. *Anal Chem*, 2007. **79**(13): p. 4870-8.
106. Ma, B., et al., *PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry*. *Rapid Commun Mass Spectrom*, 2003. **17**(20): p. 2337-42.
107. Rost, H.L., et al., *TRIC: an automated alignment strategy for reproducible protein quantification in targeted proteomics*. *Nat Methods*, 2016. **13**(9): p. 777-83.
108. Elias, J.E. and S.P. Gygi, *Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry*. *Nat Methods*, 2007. **4**(3): p. 207-14.
109. Rosenberger, G., et al., *A repository of assays to quantify 10,000 human proteins by SWATH-MS*. *Sci Data*, 2014. **1**: p. 140031.
110. Kuster, B., et al., *Scoring proteomes with proteotypic peptide probes*. *Nat Rev Mol Cell Biol*, 2005. **6**(7): p. 577-83.
111. Zhang, Y., et al., *Protein analysis by shotgun/ bottom-up proteomics*. *Chem Rev*, 2013. **113**(4): p. 2343-94.
112. Prokhorova, T.A., et al., *Stable isotope labeling by amino acids in cell culture (SILAC) and quantitative comparison of the membrane proteomes of self-renewing and differentiating human embryonic stem cells*. *Mol Cell Proteomics*, 2009. **8**(5): p. 959-70.
113. Houel, S., et al., *Quantifying the impact of chimera MS/MS spectra on peptide identification in large-scale proteomics studies*. *J Proteome Res*, 2010. **9**(8): p. 4152-60.
114. Zhang, B., et al., *DeMix workflow for efficient identification of cofragmented peptides in high resolution data-dependent tandem mass spectrometry*. *Mol Cell Proteomics*, 2014. **13**(11): p. 3211-23.
115. Brun-Buisson, C., *The epidemiology of the systemic inflammatory response*. *Intensive Care Med*, 2000. **26 Suppl 1**: p. S64-74.
116. Martin, G.S., *Sepsis, severe sepsis and septic shock: changes in incidence, pathogens and outcomes*. *Expert Rev Anti Infect Ther*, 2012. **10**(6): p. 701-6.
117. Mayr, F.B., S. Yende, and D.C. Angus, *Epidemiology of severe sepsis*. *Virulence*, 2014. **5**(1): p. 4-11.
118. Singer, M., et al., *The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)*. *JAMA*, 2016. **315**(8): p. 801-10.
119. Jawad, I., I. Luksic, and S.B. Rafnsson, *Assessing available information on the burden of sepsis: global estimates of incidence, prevalence and mortality*. *J Glob Health*, 2012. **2**(1): p. 010404.
120. Heyland, D.K., et al., *Long-term health-related quality of life in survivors of sepsis. Short Form 36: a valid and reliable measure of health-related quality of life*. *Crit Care Med*, 2000. **28**(11): p. 3599-605.
121. Dick, A., et al., *Long-term survival and healthcare utilization outcomes attributable to sepsis and pneumonia*. *BMC Health Serv Res*, 2012. **12**: p. 432.

122. Davis, J.S., et al., *Long term outcomes following hospital admission for sepsis using relative survival analysis: a prospective cohort study of 1,092 patients with 5 year follow up*. PLoS One, 2014. **9**(12): p. e112224.
123. Ley, K., et al., *Getting to the site of inflammation: the leukocyte adhesion cascade updated*. Nat Rev Immunol, 2007. **7**(9): p. 678-89.
124. Tate, J.R., et al., *Harmonising adult and paediatric reference intervals in australia and new zealand: an evidence-based approach for establishing a first panel of chemistry analytes*. Clin Biochem Rev, 2014. **35**(4): p. 213-35.
125. Farrah, T., et al., *A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas*. Mol Cell Proteomics, 2011. **10**(9): p. M110 006353.
126. Beck, M., et al., *The quantitative proteome of a human cell line*. Mol Syst Biol, 2011. **7**: p. 549.
127. Hortin, G.L. and D. Sviridov, *The dynamic range problem in the analysis of the plasma proteome*. J Proteomics, 2010. **73**(3): p. 629-36.
128. Janeway, C.A., Jr. and R. Medzhitov, *Innate immune recognition*. Annu Rev Immunol, 2002. **20**: p. 197-216.
129. Poltorak, A., et al., *Physical contact between lipopolysaccharide and toll-like receptor 4 revealed by genetic complementation*. Proc Natl Acad Sci U S A, 2000. **97**(5): p. 2163-7.
130. Geissmann, F., et al., *Development of monocytes, macrophages, and dendritic cells*. Science, 2010. **327**(5966): p. 656-61.
131. Ziegler-Heitbrock, L., et al., *Nomenclature of monocytes and dendritic cells in blood*. Blood, 2010. **116**(16): p. e74-80.
132. Iwasaki, A. and R. Medzhitov, *Regulation of adaptive immunity by the innate immune system*. Science, 2010. **327**(5963): p. 291-5.
133. Palta, S., R. Saroa, and A. Palta, *Overview of the coagulation system*. Indian J Anaesth, 2014. **58**(5): p. 515-23.
134. Furie, B. and B.C. Furie, *Thrombus formation in vivo*. J Clin Invest, 2005. **115**(12): p. 3355-62.
135. Ogston, C.M. and D. Ogston, *Plasma fibrinogen and plasminogen levels in health and in ischaemic heart disease*. J Clin Pathol, 1966. **19**(4): p. 352-6.
136. Glovsky, M.M., P.A. Ward, and K.J. Johnson, *Complement determinations in human disease*. Ann Allergy Asthma Immunol, 2004. **93**(6): p. 513-22; quiz 523-5, 605.
137. Nesargikar, P.N., B. Spiller, and R. Chavez, *The complement system: history, pathways, cascade and inhibitors*. Eur J Microbiol Immunol (Bp), 2012. **2**(2): p. 103-11.
138. Janssen, B.J., et al., *Structures of complement component C3 provide insights into the function and evolution of immunity*. Nature, 2005. **437**(7058): p. 505-11.
139. Carroll, M.C., *The complement system in regulation of adaptive immunity*. Nat Immunol, 2004. **5**(10): p. 981-6.

140. Nauta, A.J., et al., *The membrane attack complex of complement induces caspase activation and apoptosis*. Eur J Immunol, 2002. **32**(3): p. 783-92.
141. Flajnik, M.F. and M. Kasahara, *Origin and evolution of the adaptive immune system: genetic events and selective pressures*. Nat Rev Genet, 2010. **11**(1): p. 47-59.
142. Schroeder, H.W., Jr. and L. Cavacini, *Structure and function of immunoglobulins*. J Allergy Clin Immunol, 2010. **125**(2 Suppl 2): p. S41-52.
143. Harty, J.T., A.R. Tivnnerheim, and D.W. White, *CD8+ T cell effector mechanisms in resistance to infection*. Annu Rev Immunol, 2000. **18**: p. 275-308.
144. Zhu, J. and W.E. Paul, *CD4 T cells: fates, functions, and faults*. Blood, 2008. **112**(5): p. 1557-69.
145. Harwood, N.E. and F.D. Batista, *Early events in B cell activation*. Annu Rev Immunol, 2010. **28**: p. 185-210.
146. Yuseff, M.I., et al., *How B cells capture, process and present antigens: a crucial role for cell polarity*. Nat Rev Immunol, 2013. **13**(7): p. 475-86.
147. Nutt, S.L., et al., *The generation of antibody-secreting plasma cells*. Nat Rev Immunol, 2015. **15**(3): p. 160-71.
148. Fridman, W.H., *Fc receptors and immunoglobulin binding factors*. FASEB J, 1991. **5**(12): p. 2684-90.
149. Indik, Z.K., et al., *The molecular dissection of Fc gamma receptor mediated phagocytosis*. Blood, 1995. **86**(12): p. 4389-99.
150. Gonzalez-Quintela, A., et al., *Serum levels of immunoglobulins (IgG, IgA, IgM) in a general adult population and their relationship with alcohol consumption, smoking and common metabolic abnormalities*. Clin Exp Immunol, 2008. **151**(1): p. 42-50.
151. Woolf, J.M. and D.R. Burton, *Human antibody-Fc receptor interactions illuminated by crystal structures*. Nat Rev Immunol, 2004. **4**(2): p. 89-99.
152. Daeron, M., *Fc receptor biology*. Annu Rev Immunol, 1997. **15**: p. 203-34.
153. Nimmerjahn, F. and J.V. Ravetch, *Divergent immunoglobulin g subclass activity through selective Fc receptor binding*. Science, 2005. **310**(5753): p. 1510-2.
154. Arnold, J.N., et al., *The impact of glycosylation on the biological function and structure of human immunoglobulins*. Annu Rev Immunol, 2007. **25**: p. 21-50.
155. Mimura, Y., et al., *Role of oligosaccharide residues of IgG1-Fc in Fc gamma RIIb binding*. J Biol Chem, 2001. **276**(49): p. 45539-47.
156. Mattu, T.S., et al., *The glycosylation and structure of human serum IgA1, Fab, and Fc regions and the role of N-glycosylation on Fcalpha receptor interactions*. J Biol Chem, 1998. **273**(4): p. 2260-72.
157. Mimura, Y., et al., *The influence of glycosylation on the thermal stability and effector function expression of human IgG1-Fc: properties of a series of truncated glycoforms*. Mol Immunol, 2000. **37**(12-13): p. 697-706.
158. Vidarsson, G., G. Dekkers, and T. Rispens, *IgG subclasses and allotypes: from structure to effector functions*. Front Immunol, 2014. **5**: p. 520.

159. Davies, D.R., E.A. Padlan, and S. Sheriff, *Antibody-antigen complexes*. Annu Rev Biochem, 1990. **59**: p. 439-73.
160. Byrne, B., et al., *Antibody-based sensors: principles, problems and potential for detection of pathogens and associated toxins*. Sensors (Basel), 2009. **9**(6): p. 4407-45.
161. Hennion, M.C. and V. Pichon, *Immuno-based sample preparation for trace analysis*. J Chromatogr A, 2003. **1000**(1-2): p. 29-52.
162. Miller, K.D., et al., *Production, purification, and characterization of human scFv antibodies expressed in Saccharomyces cerevisiae, Pichia pastoris, and Escherichia coli*. Protein Expr Purif, 2005. **42**(2): p. 255-67.
163. Structural Genomics, C., et al., *Protein production and purification*. Nat Methods, 2008. **5**(2): p. 135-46.
164. Rosano, G.L. and E.A. Ceccarelli, *Recombinant protein expression in Escherichia coli: advances and challenges*. Front Microbiol, 2014. **5**: p. 172.
165. Stover, C.K., et al., *Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen*. Nature, 2000. **406**(6799): p. 959-64.
166. Datta, K.K., et al., *Proteogenomics of Candida tropicalis--An Opportunistic Pathogen with Importance for Global Health*. OMICS, 2016. **20**(4): p. 239-47.
167. Hoefsloot, W., et al., *Mycobacterium genavense in the Netherlands: an opportunistic pathogen in HIV and non-HIV immunocompromised patients. An observational study in 14 cases*. Clin Microbiol Infect, 2013. **19**(5): p. 432-7.
168. Lopes, J.P., et al., *Opportunistic pathogen Candida albicans elicits a temporal response in primary human mast cells*. Sci Rep, 2015. **5**: p. 12287.
169. van der Poll, T. and S.M. Opal, *Host-pathogen interactions in sepsis*. Lancet Infect Dis, 2008. **8**(1): p. 32-43.
170. Erwig, L.P. and N.A. Gow, *Interactions of fungal pathogens with phagocytes*. Nat Rev Microbiol, 2016. **14**(3): p. 163-76.
171. Vollmer, W., D. Blanot, and M.A. de Pedro, *Peptidoglycan structure and architecture*. FEMS Microbiol Rev, 2008. **32**(2): p. 149-67.
172. Demchick, P. and A.L. Koch, *The permeability of the wall fabric of Escherichia coli and Bacillus subtilis*. J Bacteriol, 1996. **178**(3): p. 768-73.
173. Bose, I., et al., *A yeast under cover: the capsule of Cryptococcus neoformans*. Eukaryot Cell, 2003. **2**(4): p. 655-63.
174. Yother, J., *Capsules of Streptococcus pneumoniae and other bacteria: paradigms for polysaccharide biosynthesis and regulation*. Annu Rev Microbiol, 2011. **65**: p. 563-81.
175. Silhavy, T.J., D. Kahne, and S. Walker, *The bacterial cell envelope*. Cold Spring Harb Perspect Biol, 2010. **2**(5): p. a000414.
176. Fischetti, V.A., V. Pancholi, and O. Schneewind, *Conservation of a hexapeptide sequence in the anchor region of surface proteins from gram-positive cocci*. Mol Microbiol, 1990. **4**(9): p. 1603-5.
177. Dills, S.S., et al., *Carbohydrate transport in bacteria*. Microbiol Rev, 1980. **44**(3): p. 385-418.

178. Fuqua, C. and E.P. Greenberg, *Listening in on bacteria: acyl-homoserine lactone signalling*. Nat Rev Mol Cell Biol, 2002. **3**(9): p. 685-95.
179. Pizarro-Cerda, J. and P. Cossart, *Bacterial adhesion and entry into host cells*. Cell, 2006. **124**(4): p. 715-27.
180. Herwald, H., et al., *M protein, a classical bacterial virulence determinant, forms complexes with fibrinogen that induce vascular leakage*. Cell, 2004. **116**(3): p. 367-79.
181. Kim, M., et al., *Bacterial interactions with the host epithelium*. Cell Host Microbe, 2010. **8**(1): p. 20-35.
182. Akesson, P., et al., *M1 protein and protein H: IgGFc- and albumin-binding streptococcal surface proteins encoded by adjacent genes*. Biochem J, 1994. **300** ( Pt 3): p. 877-86.
183. Bjorck, L. and G. Kronvall, *Purification and some properties of streptococcal protein G, a novel IgG-binding reagent*. J Immunol, 1984. **133**(2): p. 969-74.
184. Hjelm, H., K. Hjelm, and J. Sjoquist, *Protein A from Staphylococcus aureus. Its isolation by affinity chromatography and its use as an immunosorbent for isolation of immunoglobulins*. FEBS Lett, 1972. **28**(1): p. 73-6.
185. Cunningham, M.W., *Pathogenesis of group A streptococcal infections*. Clin Microbiol Rev, 2000. **13**(3): p. 470-511.
186. Lancefield, R.C., *Persistence of type-specific antibodies in man following infection with group A streptococci*. J Exp Med, 1959. **110**(2): p. 271-92.
187. Beall, B., R. Facklam, and T. Thompson, *Sequencing emm-specific PCR products for routine and accurate typing of group A streptococci*. J Clin Microbiol, 1996. **34**(4): p. 953-8.
188. Cole, J.N., et al., *Molecular insight into invasive group A streptococcal disease*. Nat Rev Microbiol, 2011. **9**(10): p. 724-36.
189. Whitnack, E. and E.H. Beachey, *Biochemical and biological properties of the binding of human fibrinogen to M protein in group A streptococci*. J Bacteriol, 1985. **164**(1): p. 350-8.
190. Macheboeuf, P., et al., *Streptococcal M1 protein constructs a pathological host fibrinogen network*. Nature, 2011. **472**(7341): p. 64-8.
191. Carlsson, F., C. Sandin, and G. Lindahl, *Human fibrinogen bound to Streptococcus pyogenes M protein inhibits complement deposition via the classical pathway*. Mol Microbiol, 2005. **56**(1): p. 28-39.
192. Nordenfelt, P., et al., *Antibody orientation at bacterial surfaces is related to invasive infection*. J Exp Med, 2012. **209**(13): p. 2367-81.
193. Collin, M. and A. Olsen, *EndoS, a novel secreted protein from Streptococcus pyogenes with endoglycosidase activity on human IgG*. EMBO J, 2001. **20**(12): p. 3046-55.
194. von Pawel-Rammingen, U., B.P. Johansson, and L. Bjorck, *IdeS, a novel streptococcal cysteine proteinase with unique specificity for immunoglobulin G*. EMBO J, 2002. **21**(7): p. 1607-15.



195. Ideker, T., T. Galitski, and L. Hood, *A new approach to decoding life: systems biology*. *Annu Rev Genomics Hum Genet*, 2001. **2**: p. 343-72.
196. Gehlenborg, N., et al., *Visualization of omics data for systems biology*. *Nat Methods*, 2010. **7**(3 Suppl): p. S56-68.
197. Omasits, U., et al., *Protter: interactive protein feature visualization and integration with experimental proteomic data*. *Bioinformatics*, 2014. **30**(6): p. 884-6.
198. Takamori, S., et al., *Molecular anatomy of a trafficking organelle*. *Cell*, 2006. **127**(4): p. 831-46.