

# Molecular BioSystems

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

*Accepted Manuscripts* are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.



[www.rsc.org/molecularbiosystems](http://www.rsc.org/molecularbiosystems)

**A comprehensive analysis of the *Streptococcus pyogenes* and human plasma protein interaction network†**

Kristoffer Sjöholm<sup>a</sup>, Christofer Karlsson<sup>b</sup>, Adam Linder<sup>b</sup> and Johan Malmström<sup>\*b</sup>

a. Department of Immunotechnology, Faculty of Engineering, Lund University, Sweden

b. Division of Infection Medicine, Department of Clinical Sciences, Lund, Lund University, Sweden

†. Electronic supplementary information (ESI) available

## Abstract

*Streptococcus pyogenes* is a major human bacterial pathogen responsible for severe and invasive disease associated with high mortality rates. The bacterium interacts with several human blood plasma proteins and clarifying these interactions and their biological consequences will help to explain the progression from mild to severe infections. In this study, we used a combination of mass spectrometry (MS) based techniques to comprehensively quantify the components of the *S. pyogenes*-plasma protein interaction network. From an initial list of 181 interacting human plasma proteins defined using liquid chromatography (LC)-MS/MS analysis we further subdivided the interacting protein list using selected reaction monitoring (SRM) depending on the level of enrichment and protein concentration on the bacterial surface. The combination of MS methods revealed several previously characterized interactions between the *S. pyogenes* surface and human plasma along with many more, so far uncharacterised, possible plasma protein interactions with *S. pyogenes*. In follow-up experiments, the combination of MS techniques was applied to study differences in protein binding to a *S. pyogenes* wild type strain and an isogenic mutant lacking several important virulence factors, and a unique pair of invasive and non-invasive *S. pyogenes* isolates from the same patient. Comparing the plasma protein-binding properties of the wild type and the mutant and the invasive and non-invasive *S. pyogenes* bacteria revealed considerable differences, underlining the significance of these protein interactions. The results also demonstrate the power of the developed mass spectrometry method to investigate host-microbial relationships with a large proteomics depth and high quantitative accuracy.

## 1. Introduction

Host-microbial relationships are based on a multitude of molecular interactions that play an important role in host-immune evasion and pathogen survival. Specifically, protein interactions at the bacterial surface allow the bacteria to manipulate host defence mechanisms using the host capabilities for their own advantage. *Streptococcus pyogenes* is a Gram-positive bacterium with a complex host-pathogen protein interaction network and is one of the most important human bacterial pathogens. *S. pyogenes* cause more than 700 million cases of superficial infections such as pharyngitis and skin infections each year, and is also responsible for clinically important sequelae such as post-streptococcal glomerulonephritis and acute rheumatic fever. Additionally, *S. pyogenes* is responsible for several severe invasive infections such as toxic shock syndrome (STSS) and necrotizing fasciitis connected with high mortality rates. The number of deaths caused by *S. pyogenes* is estimated to at least 517, 000 annually.<sup>1</sup>

*S. pyogenes* ability to colonize the throat and the skin, and to cause invasive disease requires interaction with host cells and molecules, mediated by a large number of bacterial surface molecules such as M protein,<sup>2-7</sup> protein F,<sup>6, 8, 9</sup> vitronectin-binding protein<sup>10</sup> and the hyaluronate capsule.<sup>11</sup> Several of these surface proteins specifically bind various plasma proteins such as immunoglobulins,<sup>12, 13</sup> albumin,<sup>12, 13</sup> proteins of the complement,<sup>14, 15</sup> coagulation<sup>16</sup> and contact systems,<sup>8, 10, 17, 18</sup> *etc.* and contribute to *S. pyogenes* virulence. Although the interactions between *S. pyogenes* and human plasma have been studied extensively the past decades, a comprehensive and quantitative description of the plasma proteins that bind to the surface of *S. pyogenes* is not available. Additionally the biological consequences of this plethora of protein-protein interactions are far from understood. From a clinical point of view, it would be of considerable interest to know whether changes in the plasma protein-binding pattern correlate with the progression of a superficial and in many cases asymptomatic infection, to severe invasive disease. The starting point for the present investigation was the assumption that mass spectrometry-based proteomics could make it possible to address this issue.

The deciphering of the *S. pyogenes*-plasma protein interaction network requires reproducible and comprehensive protein measurements. In previously published work we have demonstrated that the combination of two mass spectrometry-based

approaches, selected reaction monitoring mass spectrometry (SRM-MS) and tandem mass spectrometry (LC-MS/MS), can provide an absolute quantitative description of a bacterial proteome.<sup>19-21</sup> The absolute quantitative data enables comparisons of different proteins in the same sample and provides information regarding protein abundance that is of particular relevance for host pathogen interactions. In LC-MS/MS, tryptic digests of complex protein mixtures are analysed in data-dependent acquisition mode to generate relative-quantitative information of the proteins in samples with unknown protein composition. In SRM, pre-selected proteotypic peptides are specifically targeted using triple quadrupole mass spectrometers to generate complete data matrices of samples with known protein composition.<sup>22</sup> The ability to target specific proteotypic peptides relies on *a priori* defined information of the selected peptide mass over charge ratio, preferred daughter fragments, retention time and collision energy.<sup>22</sup> This information is collectively referred to as SRM assays that needs to be constructed and used for the quantitative analysis.<sup>23</sup> The targeted analysis of a limited number of peptides provides quantitatively accurate, sensitive and reproducible analysis of the target proteins. The complementary data output from the two mass spectrometry techniques can generate the comprehensive information necessary to capture the dynamic and complex protein-protein interaction network between host and pathogen.

In this study, we demonstrate how the combination LC-MS/MS and SRM enables the quantitative characterization of the human blood plasma-*S. pyogenes* bacterial surface protein interaction network. The combination of MS methods facilitated the categorization of the interacting proteins into groups that specifically bind to the surface of *S. pyogenes* at high concentrations. Applying these methods to a strain that caused both asymptomatic pharyngitis and invasive disease within the same patient,<sup>24</sup> demonstrates the importance of interactions between *S. pyogenes* and human plasma proteins for the transition from mild to severe infection. The results also underline that the combination of mass spectrometry based techniques can improve our current understanding of molecular interactions that contribute to bacterial virulence, a knowledge that has the potential to identify novel opportunities to treat patients with invasive infections.

## 2. Materials and methods

### 2.1 Bacterial strains and growth conditions

*S. pyogenes* strain AP1 (strain 40/58 from the WHO Collaborating Centre for Reference and Research on Streptococci, Prague, Czech Republic) is a *covS* truncated clinical isolate of the M1 serotype. The AP1 isogenic mutant BMJ71 has a transposon insertional inactivation of *mga*,<sup>25</sup> a positive gene regulator of several virulence factors, including the two surface proteins M1 and H that are known to bind to several human plasma proteins. An M1 strain isolated from asymptomatic pharyngitis (in the throat) and from necrotic tissue (in the leg) of the same patient<sup>24</sup> was also investigated. Single colonies were grown to exponential phase in 30g/L Todd-Hewitt broth (BD) and 6g/L yeast extract (Merck, Darmstadt, Germany) (THY), harvested by centrifugation and resuspended in 20mM Tris-HCl (Merck), 150mM NaCl (Sigma-Aldrich, St. Louis, MO, USA), pH 7.6, to a concentration of  $2 \times 10^9$  CFU/ml.

### 2.2 Plasma adsorption, glycine elution and sample preparation

Blood plasma was prepared and pooled from 5 healthy individuals as previously described.<sup>24</sup> Plasma was mixed with bacteria and the samples were incubated for 30 minutes at 37°C, allowing plasma proteins to adsorb to the bacterial surface. Bacteria were harvested after several washes in 20mM Tris-HCl (Merck), 150mM NaCl (Sigma-Aldrich), pH 7.6 using centrifugation ( $5,000 \times g$ ). To elute the adsorbed proteins the final cell pellets were resuspended in 0.1 M glycine (Sigma-Aldrich), pH 2.0, followed by incubation for 10 minutes, the supernatants were neutralized to pH 7-8 with 1M Tris (Ultrapure, Saveen Werner AB). The samples were denatured with urea (Fluka, Steinheim, Germany) before in-solution digestion (see below).

### 2.3 Double digestions

Plasma was mixed with bacteria and washed as described above. After the washes following plasma adsorption, the cells were washed one additional time and resuspended in ice-cold water and transferred to tubes containing 0.1 mm Silica beads (Lysing Matrix tubes, Nordic Biolabs #6911100, Täby, Sweden). The cells were lysed with a cell disruptor (Minibeadbeater-96 Cat#1001, Biospec products, Inc., Bartlesville, OK, USA). To digest proteins associated with bacterial cell wall fragments, 1 $\mu$ g trypsin

(Sequence grade modified trypsin Porcin, Promega, Madison, WI, USA) and urea at a final concentration of 1M was added and samples were incubated for 30min at 37°C.

#### *2.4 SDS-PAGE*

Bound plasma proteins from a total bacterium concentration of  $6 \times 10^6$  CFU per ml plasma was used as the starting material. Precasted gels (Criterion™, 12+2 well comb, 45µl, Bio-Rad laboratories Inc., Hercules, CA, USA) and up to 100µg protein was used for SDS-PAGE. The gel was run at (Criterion™, Bio-Rad laboratories Inc.) 60V until the samples have started to migrate and then the voltage were increased to 160V until the lanes had migrated until about 1cm to the edge of the gel.

#### *2.5 In gel digestion*

The gel was prepared according to previous work with minor modifications<sup>26, 27</sup>. The gel was stained with GelCode® Blue Stain Reagent (Thermo Scientific, Rockford, IL, USA) and each lane of the gel was cut into 10 fractions and 20mM (instead of 10mM) dithiothreitol was used. To increase the peptide concentration the samples were dried in speedvac and solved in 3% ACN and 0.1% formic acid.

#### *2.6 In solution digestion*

The protein sample was reduced using Tris (2-carboxyethyl) phosphine (TCEP) (Sigma-Aldrich), at a final concentration of 5mM and the samples were incubated at 37 °C for 60 minutes. The samples were incubated for 30 minutes in the dark at room temperature with 2-Iodoacetamide (IAA) (AppliChem) at a concentration of 10mM as alkylating agent. Samples were diluted with twice the sample volume in ammonium bicarbonate (ABC) (Sigma-Aldrich) and digested with 2.5ng/µl trypsin (Sequence grade modified trypsin Porcin, Promega) over night. Adding formic acid to pH 2-3 stopped the digestion.

#### *2.7 C18 peptide clean-up*

Vydac UltraMicroSpin® Silica C18 300Å Columns (#SUM SS18V, The Nest Group, Inc., Southborough, MA, USA) was used for sample desalting, clean-up and concentrating peptides according to the manufactures instructions.

### 2.8 Mass spectrometry analysis

The hybrid Orbitrap-LTQ XL mass spectrometer (Thermo Electron, Bremen, Germany) was coupled online to a split-less Eksigent 2D NanoLC system (Eksigent technologies, Dublin, CA, USA). Peptides were loaded with a constant flow rate of 15  $\mu\text{l}/\text{min}$  onto a pre-column (PepMap 100, C18, 5  $\mu\text{m}$ , 0.3 mm x 5 mm, LC Packings, Amsterdam, Netherlands) and subsequently separated on a RP-LC analytical column (10  $\mu\text{m}$  fused silica emitter, 75  $\mu\text{m}$  x 16 cm, PicoTip™ Emitter, New Objective, Inc. Woburn, MA, USA, packed in-house with Reprosil-Pur C18-AQ resin, 3  $\mu\text{m}$ , Dr. Maisch, GmbH) with a flow rate of 300 nl/min. The peptides were eluted with a linear gradient from 95% solvent A (0.1% formic acid in water) and 5% solvent B (0.1% formic acid in acetonitrile) to 35% solvent B over 60 minutes. The data analysis was performed as previously described.<sup>28</sup>

### 2.9 Selected reaction monitoring (SRM)

Selected reaction monitoring (SRM) transition assays were constructed by testing the ten most abundant peptide fragments for selected proteotypic peptides identified with high confidence in the LC-MS/MS experiments. Spiked in the RT-peptides (Biognosys AG, Zurich, Switzerland) allowed normalization of the retention time as previously described.<sup>28</sup> The SRM measurements were performed on a TSQ Vantage triple quadrupole mass spectrometer (Thermo Electron, Bremen, Germany) equipped with a nano electrospray ion source (Thermo Electron). Chromatographic separations of peptides were performed on an Easy-nLC II system (Thermo scientific, San Jose, CA, USA). The data analysis was performed as previously described<sup>28</sup> using a 2.5% false discovery rate. The resulting peptide abundances were exported into a database, where protein abundances were inferred by summing up the abundances for the peptides uniquely mapping to each protein.<sup>29</sup>

### 2.10 Data processing

The initial list of putative interacting proteins was generated by fractionating the proteins prior to LC-MS/MS analysis resulting in high use of instrument time and in complex data outputs. To omit the protein fractionation step and still maintain the same level of sensitivity we turned to SRM. SRM is associated with high sensitivity and specificity enabling quantification of the majority of the target proteins in 1D-LC-SRM-MS.<sup>30</sup> By pooling high, medium and low abundant proteins in separate transitions sets the sample load was optimized for the individual transitions sets. To reduce the number

of target proteins certain proteins were removed prior to the SRM analysis. The specific selection of target proteins reflects why there are additional proteins in the plasma abundance graph compared to the plasma adsorption abundance graph. The criteria we used to exclude proteins were as follows. Firstly, all immunoglobulin proteins were excluded since shotgun MS result in many different immunoglobulins due to their variable regions; secondly, proteins unable to generate suitable SRM assays; thirdly, the proteins not detected in the plasma adsorption samples and fourthly proteins with small pXIC ratio between plasma and plasma adsorption.

### 2.11 Equations

Calculation of the relative enrichment on of proteins on the bacterial surface:

$$\text{Enrichment ratio} = \frac{pXIC_{\text{plasma adsorption}}}{pXIC_{\text{plasma}} + pXIC_{\text{plasma adsorption}}} \quad (1)$$

A more detailed description of materials and methods can be found in the supplemental material, this includes mass spectrometric settings and database search preferences.

### 3. Results

#### 3.1 LC-MS/MS analysis of plasma proteins adsorbed to the surface of *S. pyogenes*

The attachment of specific plasma proteins to *S. pyogenes* surface has been known for decades.<sup>12, 16</sup> However, a comprehensive map of the multitude of protein interactions at the bacterial surface is still missing. To expand the current knowledge of the protein interactions between human plasma and *S. pyogenes*, we incubated *S. pyogenes* strain AP1 cells in human blood plasma, removed unbound proteins and eluted off the interacting proteins using acidic conditions. The starting human blood plasma proteins and the proteins eluted from the *S. pyogenes* surface, referred to as the plasma adsorbed (PA) sample, were separated by SDS-PAGE. The resulting gels were sliced and processed for mass spectrometry followed by LC-MS/MS analysis in data dependent acquisition mode. The obtained data was analysed using X!Tandem, Peptide- and Protein -Prophet with a FDR of 1%, resulting in the identification of 264 human proteins in plasma and 181 human proteins in the PA sample. The extracted ion intensities (XIC) for the individual proteins were estimated using SuperHirn<sup>29</sup> as described in the supplemental material. To quantitatively compare the plasma sample with the plasma adsorbed sample we divided the XIC for the individual proteins with the total ion current (TIC) of the identified proteins from plasma and PA sample respectively to obtain proportional XIC (pXIC) for the individual proteins. By using the pXIC for every protein we could estimate the relative proportion a particular protein consumes of the total protein mass for the two separate samples, ranging from 0.64 to  $8.8 \times 10^{-8}$  in plasma (S-table 1) and 0.22 to  $2.8 \times 10^{-8}$  in the PA sample (S-table 2). In the plasma sample, the vast majority of the identified proteins were, as expected, classical plasma proteins such as transport proteins, complement proteins and coagulation proteins. In contrast, the PA sample contained many non-classical plasma proteins (S-table 2) such as proteins involved in cell adhesion, intracellular proteins, extracellular matrix components and secreted proteins (Fig. 1a). Many of the numerous non-classical plasma proteins were however present in low amounts. The relative abundance of these proteins was less than 5% of the total pXIC.

The pXIC abundance from the plasma and the PA sample was subsequently used to estimate the relative increase or decrease in protein concentration at the bacterial

surface (Equation 1 in materials and methods) (Fig 1b, S-table 3). Relatively enriched proteins at the bacterial surface are expected to have an enrichment ratio larger than 0.5. In total, 67 proteins (rectangles with rounded edges) were exclusively found in the PA sample indicated by an enrichment ratio of 1 and 39 proteins (stars) displayed an enrichment ratio larger than 0.8 representing proteins likely enriched at bacterial surface. In contrast, 39 proteins (rhombi) displayed an enrichment ratio smaller than 0.2 and 151 proteins (rectangles with sharp edges) were identified exclusively in plasma, with enrichment ratio equal to 0, and indicates proteins with no or weak affinity to the *S. pyogenes* surface. The remaining 36 proteins (triangles) showed neither a small increase or decrease with enrichment ratios between 0.8 and 0.2. The largest difference between the plasma and the PA samples is found among the non-classical plasma proteins such as proteins involved in cell adhesion, intracellular proteins and secreted proteins. These protein groups displayed the largest total increase at the bacterial surface where in particular the proteins involved in cell adhesion were markedly increased (Fig 1c). The average intensity of the classical proteins remained constant between the plasma and the PA sample (Fig. 1c). Although the average intensity value for the classical proteins was even, the classical plasma proteins display a large variation in enrichment at the bacterial surface where specific proteins were strongly elevated at the bacterial surface whereas others were strongly reduced (S-table 3). Two examples are C4bp- $\alpha$  and Serotransferrin that has an enrichment factor of 0.996 and 0.003 respectively.

The use of the pXIC ratio between plasma and the PA sample categorized the proteins into groups depending of their relative changes in concentration at the bacterial surface. In particular the 39 proteins with an enrichment ratio larger than 0.8 and the 67 proteins exclusively found in the PA samples, are of interest for additional studies. These 106 proteins along with the 36 proteins with unclear affinity and a smaller number of selected highly abundant plasma protein with a high pXIC and an enrichment ratio below 0.2, represent the initial target proteins for further examination with SRM.

### **3.2 Confirmation of the enrichment ratios using selected reaction monitoring**

The glycine elution (GE) protocol used above favours the removal of proteins loosely attached to the bacterial surface. Several proteins may however bind tightly to the bacterial surface and could be lost or not fully recovered using the GE protocol. To

maximize the recovery of the surface proteins, we introduced a second elution protocol referred to as Double Digestion (DD) to increase the number of tightly bound plasma proteins. With DD, the bacteria are lysed followed by trypsin incubation to proteolytically digest the surface bound proteins from the bacterial cell lysate followed by the removal of non-protein cell debris and a second tryptic digestion step. The DD protocol enables the extraction of both surface bound proteins and the complete bacterial protein content in one experimental setup. The experimental conditions used in DD release proteins covalently attached to the bacterial surface and intracellular proteins that can be used to compensate for uneven sample loss during the washing steps.

The DD sample preparation protocol considerably increases the complexity of the protein composition in the samples. The increased protein complexity as a result of the DD sample preparation protocol and the need to analyse larger sample sets including biological replicates using the two sample preparation methods, urged us to change mass spectrometry platform, from LC-MS/MS to SRM. We firstly created SRM assays, as previously described,<sup>31</sup> for as many of the 142 target proteins as possible from the list described above, and a few human plasma control proteins resulting in a total of 152 human plasma proteins covered by 406 separate SRM assays and 2030 transitions.

We performed PA on a highly virulent wild type *S. pyogenes* strain, AP1 (same as in LC-MS/MS experiments), in quadruple biological preparations using both GE and DD, and the control plasma sample in duplicates. From the total list of 152 human proteins we could reproducibly quantify 122 human proteins in the samples using SRM (S-table 4), despite that no peptide pre-fractionation was used as in the LC-MS/MS experiments. Among the 122 reproducibly interacting plasma proteins, 35 (29%) were previously described in the literature, whereas 87 (71%) were so far unknown to potentially interact with the *S. pyogenes* surface<sup>28, 32-36</sup> (Supplementary Fig. 1). To visualize the protein abundance patterns we used principal component analysis (PCA) on the z-score of the result. The data set was subsequently subdivided using the PCA results, which were clustered by k-mean clustering into 5 groups (referred to as group 1-5), represented by boxes in Figure 2 as previously described.<sup>32</sup> Group one contains proteins with high intensity values in plasma, group three and five contains proteins with high intensity values in DD and GE respectively. The last two groups, two and four, do not

show as clear differentiation between one of the elution protocols nor plasma and the others, both groups got similar result in the both elution protocols (hence GE=DD) and in group two the plasma value is generally lower than in group four. The vertical axis represents the relative distribution, the individual proteins were divided by the total intensity for the corresponding protein so that the total protein intensity sum equals 1 for all proteins, of the signal between the two different methods and plasma. Enlarged graphs of the different groups are shown in the Supplementary Fig. 2a-e. In particular two groups (three and five) indicates a high release of human plasma proteins bound to *S. pyogenes* surface. In group three (high DD) 14 proteins displayed a stronger signal in the DD protocol. Since this method extracts all surface proteins this indicates that these proteins bind strongly to the bacterial surface, possibly covalently. Previous studies have only identified 1 of the 14 proteins (Vitamin K-dependent protein S) (7%). The relatively low number of previously characterized protein interactions from this group can be attributed to the DD extraction method used here. In contrast the group 5 proteins (high GE) were enriched for known interactions; 8 out of the total 22 (36%) proteins in this group was previously characterize to bind to the *S. pyogenes* surface. The large proportion of known interacting proteins within this group is likely a consequence of the elution protocol. With the low pH elution method, intact proteins are released that can be analysed with classical methods like 2D-gel and Western blot. Group 1 also contains several previously known protein interactions, 18 out of the total 38 proteins (49%) were previously described to bind to the streptococcal surface. The proteins in this group are in general proteins of the complement system and other highly abundant plasma proteins. The relatively high concentration of these proteins in plasma makes it difficult to evaluate the specificity of these interactions.

The combination of sample preparation methods used here followed by SRM analysis identified 90 so far unknown potential interactions between human plasma proteins and *S. pyogenes*. The combination of elution protocols increases the number of visible proteins and is beneficial as both tightly adhering proteins can be released in addition to the complete bacterial proteome. The intracellular proteins are of interest for normalization purposes and comparison between strains. As all five groups contain previously characterized binders we note that no group can be excluded as a source for containing true interacting proteins of biological meaning.

### 3.3 Integrating the targeted and shotgun proteomics data

We suspected that the relatively large number of proteins found in the PA sample indicate that several proteins are not specific binders to the bacterial surface. The challenge is to select true interacting proteins of biological interest. We hypothesize that biologically significant interactions are likely present at relatively large amounts at the bacterial surface, and that the ratio between the pXIC, the enrichment ratio, is high which indicates specific protein interactions. By merging the output from LC-MS/MS and SRM analysis, a data set containing both relative protein abundances at the surface, the enrichment ratio and the relatively enrichment divided up into five groups was obtained (Fig 3). Firstly, we investigated the relative abundance of the five protein categories described above. The ion intensities for the separate groups were calculated for both plasma (Fig. 3a) and samples following plasma adsorption (Fig. 3b), by summing the ion intensities for all proteins within a given group (S-table 4). The major differences in signal intensities between plasma and plasma adsorption can be found in groups 1, 2, 3 and 5 (Fig. 3, insert pie graphs). For group 1, protein ion intensities in plasma consume 77.7% of the total ion current (TIC) (insert pie graphs in Fig 3a), which is reduced to 7.4% in the plasma adsorption sample (insert pie graphs in Fig 3b). In contrast, groups 2, 3 and 5 display a more than ten-fold increase in intensities in the plasma absorbed sample from  $9.5 \cdot 10^{-3}\%$  to 0.4% for group 2, from 0.1% to 2% for group 3 and from 2.3% to 59.4% for group 5. These results indicate that the group 5 proteins represent high abundance levels at the bacterial surface and which are enriched compared to plasma. It is also within this group where most of the previously characterized protein interactions are found. In group 2 and 3, the proteins show an increase in their relative amount compared to plasma. However, in general these proteins are present at low concentration, underlined by the fact that few of these were previously characterized as interacting proteins.

To visualize the individual proteins, the SRM dependent protein categories were plotted into the protein abundance graph in Figure 3. The coloured markers represents proteins detected with both SRM and shotgun experiments and black pluses represents proteins detected in the shotgun experiments, but are excluded from or not detected in the final SRM experiments. The majority of these proteins are different forms of the light and variable chain of immunoglobulins, notoriously known to be difficult to interpret with

MS. It is evident from the graph that several of the high abundant plasma proteins (group 1, blue circles) shown in Figure 3a, have a considerable lower abundance in the plasma adsorption sample (Fig. 3b), with the exception for albumin, which is still among the most abundant proteins at the bacterial surface. In this experimental setup, the amount of albumin is reduced from 64% pXIC in plasma, to 5.7% in the plasma adsorption sample. Although albumin consumes 5.7% of the signal intensity in this sample the concentration is reduced, compared to plasma, at the surface of bacteria as confirmed by the SRM ratios. In contrast, several well characterized binding proteins are dramatically enriched on the bacterial surface such as fibrinogen and the C4-binding protein (indicated by the orange rhombs, group 5). This is also observed for example for the fibrinogen  $\gamma$  chain, this protein increases from 0.5% pXIC in plasma to 20% following plasma adsorption. In addition, 80% of the SRM signal for fibrinogen  $\gamma$  is confined to the glycine sample eluted, confirming the increase.

This final overview and the supplement tables represent a resource for selecting host-pathogen interactions that can be selected for further analyses. Based on the total amount of proteins at the bacterial surface, and the ratio between plasma proteins and adsorbed proteins, prominent and highly specific interactions can be identified.

### ***3.4 Differences in plasma protein interaction profiles between a virulent wild type *S. pyogenes* and an isogenic *mga*<sup>-</sup> mutant***

To further investigate the relationship between abundant virulence factors belonging to the *mga* regulon and the ability to adhere to plasma proteins, we analysed AP1 (highly virulent wild type (WT) strain) and an isogenic *mga*<sup>-</sup> mutant from the same strain using SRM. The two strains were plasma absorbed and the bound proteins were isolated using both GE and DD in four separate replicates followed by SRM analysis. Figure 4 displays the 28 proteins that are statistically significant different in bound human protein abundances between the WT and *mga*<sup>-</sup> strain, determined using Wilcoxon T-test and x-axis sorted alphabetically. The *mga*<sup>-</sup> mutant lacks, among other proteins, the important cell wall anchored virulence factors M1 protein and the M-like protein, protein H<sup>25</sup> both with several characterized interactions with plasma proteins.<sup>13, 14, 16</sup> Hence we expected large differences in the protein abundance levels of known interacting plasma proteins, which we could confirm in the data (Fig. 4, red bars, left axis). The largest decrease in

protein abundance levels on the *mga*<sup>-</sup> strain were proteins such as C4BP  $\alpha$  and  $\beta$  chain, fibrinogen  $\alpha, \beta, \gamma$  chains and Vitamin K-dependent protein S which were all present more than 35 times higher levels on the surface on the WT strain. All of these proteins are both enriched on the bacterial surface from plasma (Fig. 4, blue bars, left axis) and previously confirmed binders to the protein M1 and/or the H protein.<sup>15, 16</sup> Vitamin K-dependent protein S is described to bind to C4-binding protein potentially explaining why this protein is enriched on the surface of WT strain.<sup>37</sup> The six proteins consumed in total 60% of the total pXIC on the bacterial surface (Fig 4, green bars, right axis) indicating that these proteins are abundantly present at the bacterial surface. Serum albumin, Coagulation factor XIII and Histidine-rich glycoprotein were other classical plasma proteins present more than three times higher on the WT strain surface compared to the *mga*<sup>-</sup> mutant. Conversely, eight complement proteins, among them all of the proteins in the MAC complex, displayed considerable, although not statistically significant, higher levels on the *mga*<sup>-</sup> mutant bacterial surface (data not shown). Lastly, several less characterized proteins were significantly elevated at the bacterial surface, where among the most striking ones were Transglutaminase 3 and Bone morphogenetic protein 1, which were more than 15 times more abundant on the surface of the AP1 strain. Although Transglutaminase 3 and Bone morphogenetic protein 1 display a large fold change comparing the *mga*<sup>-</sup> and WT strain their relative amount on the bacterial surface is estimated to be below 0.001% of the total protein mass, which makes any assumption of their biological role uncertain.

### ***3.5 Differences in plasma protein interaction profiles between invasive and non-invasive S. pyogenes bacteria***

The ability of *S. pyogenes* to bind various human plasma proteins is known since decades and is believed to contribute to its disease-causing potential. However, little information is available regarding changes in the binding profiles between bacteria isolated from patients with invasive and non-invasive infections. To address this question we analysed the plasma protein binding properties, same 152 proteins described above, for two unique *S. pyogenes* isolates obtained simultaneously from the same patient; one from the throat causing no local symptoms and one from a deep invasive infection of the right thigh giving rise to life-threatening necrotizing fasciitis. In both isolates the bacteria were of the M1 serotype and had spread from the throat to the tissue in the thigh via the

blood stream.<sup>24</sup> PA was performed on the two isolates followed by both GE and DD to identify and quantify surface bound plasma proteins. 132 human proteins were detected using SRM with the three transition sets described above (see the full list of proteins in S-table 5).

All samples (including the reference sample of trypsin digested plasma) were prepared in quadruplicates. The results from glycine elution and double digestion were summed up and an average was used for each of the isolates and statistically differentially abundant proteins determined using Wilcoxon test, resulting in 36 proteins differentiating significantly ( $p < 0.05$ ) between the isolates from throat and leg (Figure 5). The majority of the statistically significant differences represent human plasma proteins that are present in higher amounts on the throat isolate. Two of these proteins are fibrinogen and albumin present 4 and 7 times higher respectively. Both fibrinogen and albumin was previously described to bind to the M-protein.<sup>12, 16</sup> A similar difference in fibrinogen and albumin binding was observed in the WT and *mga* mutant comparison in Figure 4, which lead us to suspect that the throat strain have higher amounts of the M1-protein. Via the use of absolute quantification of the M1 protein using 3 proteotypic AQUA peptides in the DD samples we could indeed demonstrate that the isolate from the throat showed a more than 10-fold higher level (data not shown) of M1 protein.

The increase in M protein surface abundance of the throat isolate may also influence the binding of other human plasma proteins than previously shown above. For instance, among the proteins that have the highest difference in binding profile between throat and leg isolates are complement Factor H (8 times) and C1q (13 times). In addition proteins of the membrane attack complex (complement component 5 to 9) were twice as abundant at the throat isolate surface compared to the leg isolate's and were enriched 97-238 times from plasma (white bars in Fig 5). Potentially, this can be correlated to the 3.6 fold higher (data not shown) binding of IgG to the throat isolate, also determined using AQUA peptides as previously described.<sup>24</sup> Hence, higher levels of M1 protein may lead to increased binding of IgG and increased complement deposition. Only three proteins were statistically significantly increased with at least 2 times on the leg isolate, Cytosolic malate dehydrogenase, Hemoglobin and C-factor-B, indicating a higher capacity of the throat isolate to bind human host proteins. Most of the proteins with significant difference between the throat and leg isolates (32 out of 36) were also

statistically significantly enriched on both the throat and leg isolate compared to plasma indicating specific interaction. The proteins that are not statistically significantly different are by \* or \*\* at the top of the Figure 5.

These experiments demonstrates that LC-MS/MS analysis in combination with SRM can confirm previously characterized protein interactions between the bacterial surface and human plasma and at the same time discover proteins that was not previously reported to bind to the bacterial surface. When comparing two unique isolates from a single patient the method revealed large differences in the ability of the two strains to bind specific human plasma proteins, partly explained by different protein abundance level of the M1-protein. The results show that invasive and non-invasive strain used in this work display large differences in their capacity to bind human plasma proteins.

#### 4. Discussion

In this work we used a mass spectrometry based proteomics strategy to investigate protein interactions between host and pathogen for improving the understanding of how *S. pyogenes* transits from mild to severe disease. To increase the number of potential interacting proteins and to demonstrate the validity of the method we made several modifications to previously published protocols. Firstly we expanded the current state-of-the art method by including a double digestion protocol. Secondly we used combinations of mass spectrometry based technologies to demonstrate the abundance level of specific proteins at the bacterial surface and to demonstrate specific interactions indicated by the relative protein enrichment at the bacterial surface from plasma. Overall the method facilitated the division of the proteins into five groups, where in particular three groups (two, three and five) are likely to contain several protein interactions of relevance.

The initial LC-MS/MS analysis of the PA sample resulted in the identification of 181 proteins. It is unlikely that all these proteins have specific interaction with the bacterial surface. Possible sources of contamination could be residual amounts of proteins from plasma present in the PA samples or that some host proteins could be coupled to other host proteins. The differentiation between specific and non-specific interactions is not trivial. In this work we used as a rule of thumb that enriched proteins at the bacterial surface found at relatively high concentrations are likely to have biological consequences. It is apparent that this assumption can only serve as an estimate as some known interactions like albumin are not enriched at the bacterial surface although the albumin-*S. pyogenes* interaction is ascribed to a biological function.<sup>28</sup> However, the general over-representation of previously characterized interactions in the high glycine elution group such as fibrinogen indicates validity of the approach. In the supplement tables we have provided the quantitative values for the individual proteins to also allow others to draw conclusion from the data.

Historically protein interactions between host and pathogen were investigated by eluting the bound protein using acidic conditions. By adding the double digest protocol the number of detectable proteins was expanded on the *S. pyogenes* surface. The combination of this method and mass spectrometry provides the possibility to study

partially degraded proteins, which may be difficult with many other methods. In addition several bacterial surface proteins and intracellular proteins were detected using the double digest protocols. The intracellular proteins could also allow the assessment of differential sample loss of specific strains in the washing steps. On the other hand the presence of complete bacterial proteomes in the sample increased the sample complexity and dynamic range of proteins within the sample. We do however, not observe any attrition in the number of detected proteins comparing the two protocols. Nevertheless, we believe that the optimal solution would be a combination of the two protocols in one experimental setup.

In shotgun MS, normalizing by TIC provides an approximation of the amount of a protein compared to the other proteins. Since only selected proteins are measured in SRM it is not straightforward to normalize by TIC in SRM and therefore hard to relate the proportional amount of a protein within a sample. Hence, in these SRM experiments we could not confidently state that a particular protein was enriched at the bacterial surface as appropriate plasma levels to analyse was difficult to assess. We alleviated this issue by merging the SRM data and the shotgun proteomics data. The combination of SRM and LC-MS/MS data provided a subdivision of the groups compared to the protein plasma levels and an estimation of individual protein amount at the bacterial surface. Several relevant *S. pyogenes*-host interacting proteins may be found in the other groups we predict, due to the high enrichment in one or both of the extraction methods, that protein groups 2, 3 and 5 contains the largest number of relevant proteins for future investigate. Importantly, to fully assess the relevance of the interaction of a particular protein additional investigations are necessary. For example Group 1 mainly contain high abundant proteins, the decrease in abundance when plasma is adsorbed might not make the interaction with the *S. pyogenes* surface biological irrelevant, one example is serum albumin with known interaction.<sup>32</sup> Based on the rationale described above we predict that the most interesting proteins for future studies are the ones with previously uncharacterised interactions with bacterial surface.

By using the method developed above to investigate the binding pattern of a wild type strain and an isogenic mutant lacking the proteins of the *mga*- regulon, like the M-protein and the M-like protein (protein H), we could assess the combined binding of human plasma proteins to the M-protein and the other missing virulence factors. The

removal of the mga proteins resulted in dramatic decrease of the binding of several proteins. The result demonstrates the methods ability to measure both known and putative interactions within one biological experiment. The use of additional isogenic mutants, missing only one virulence factor, may provide a more precise relationship between a surface protein and its interaction partners. The two *S. pyogenes* isolates from the same patient isolated from the throat with no local symptoms and from a necrotising fasciitis in the leg provided us with the unusual possibility to study the mechanism behind *S. pyogenes* transition from superficial to severe disease. The strain isolated from the leg showed significant lower ability to bind plasma proteins known to interact with the M1 protein such as fibrinogen and serum albumin, correlating with the more than ten-fold lower levels of the M1 protein in the leg strain. The throat strain also had a larger number of proteins belonging to the complement MAC complex attached to the surface, which may in turn be linked to the higher levels of IgG and C1q found in the throat isolate. In fact, *in vivo* analysis of the throat/leg strains previously reported that the strain isolated from the leg has a large degree of shedding of the surface proteins.<sup>24</sup> As we can observe large differences in these *in vitro* experiments suggests that the leg strain may have acquired genetic differences resulting in decreased levels of the M1 protein and decreased ability to bind specific plasma proteins and increased ability to avoid immune response.

## 5. Conclusions

The combination of MS methods used here demonstrate a considerable expansion of the human interacting blood plasma proteome. Relying on the quantitative ability and the orthogonal output of two MS methods (S-Figure 3, schematic overview) allowed subdivision of the proteins into groups. The ability to measure several protein interactions at the same time is important as many of the *S. pyogenes* virulence factors have many human plasma protein-binding sites. We anticipate that this method can be extended to include the interaction sub proteomes for other biological fluids such as saliva and also to other pathogens.

## 6. Acknowledgements

This work was supported by the Swedish Research Council (projects 2008:3356 and 621-2012-3559), the Swedish Foundation for Strategic Research (grant FFL4), the

Crafoord Foundation (grant 20100892), the Wallenberg Academy Fellow KAW (2012.0178) and European research council starting grant (ERC-2012-StG-309831)

## 7. References

1. J. R. Carapetis, A. C. Steer, E. K. Mulholland and M. Weber, *The Lancet infectious diseases*, 2005, **5**, 685-694.
2. M. G. Caparon, D. S. Stephens, A. Olsen and J. R. Scott, *Infection and immunity*, 1991, **59**, 1811-1817.
3. J. B. Dale, R. W. Baird, H. S. Courtney, D. L. Hasty and M. S. Bronze, *The Journal of infectious diseases*, 1994, **169**, 319-323.
4. R. P. Ellen and R. J. Gibbons, *Infection and immunity*, 1972, **5**, 826-830.
5. R. P. Ellen and R. J. Gibbons, *Infection and immunity*, 1974, **9**, 85-91.
6. N. Okada, A. P. Pentland, P. Falk and M. G. Caparon, *The Journal of clinical investigation*, 1994, **94**, 965-977.
7. J. R. Wang and M. W. Stinson, *Infection and immunity*, 1994, **62**, 442-448.
8. E. Hanski and M. Caparon, *Proceedings of the National Academy of Sciences of the United States of America*, 1992, **89**, 6172-6176.
9. E. Hanski, P. A. Horwitz and M. G. Caparon, *Infection and immunity*, 1992, **60**, 5119-5125.
10. P. Valentin-Weigand, J. Grulich-Henn, G. S. Chhatwal, G. Muller-Berghaus, H. Blobel and K. T. Preissner, *Infection and immunity*, 1988, **56**, 2851-2855.
11. M. R. Wessels and M. S. Bronze, *Proceedings of the National Academy of Sciences of the United States of America*, 1994, **91**, 12238-12242.
12. G. Kronvall, A. Simmons, E. B. Myhre and S. Jonsson, *Infection and immunity*, 1979, **25**, 1-10.
13. P. Akesson, K. H. Schmidt, J. Cooney and L. Bjorck, *The Biochemical journal*, 1994, **300 ( Pt 3)**, 877-886.
14. R. D. Horstmann, H. J. Sievertsen, J. Knobloch and V. A. Fischetti, *Proceedings of the National Academy of Sciences of the United States of America*, 1988, **85**, 1657-1661.
15. A. Thern, L. Stenberg, B. Dahlback and G. Lindahl, *J Immunol*, 1995, **154**, 375-386.
16. F. S. Kantor, *The Journal of experimental medicine*, 1965, **121**, 849-859.
17. W. A. Simpson and E. H. Beachey, *Infection and immunity*, 1983, **39**, 275-279.
18. B. A. Sanford, V. E. Davison and M. A. Ramsay, *Infection and immunity*, 1982, **38**, 513-520.
19. J. Malmstrom, M. Beck, A. Schmidt, V. Lange, E. W. Deutsch and R. Aebersold, *Nature*, 2009, **460**, 762-765.
20. A. Schmidt, M. Beck, J. Malmstrom, H. Lam, M. Claassen, D. Campbell and R. Aebersold, *Molecular systems biology*, 2011, **7**, 510.
21. S. Kuhner, V. van Noort, M. J. Betts, A. Leo-Macias, C. Batisse, M. Rode, T. Yamada, T. Maier, S. Bader, P. Beltran-Alvarez, D. Castano-Diez, W. H. Chen, D. Devos, M. Guell, T. Norambuena, I. Racke, V. Rybin, A. Schmidt, E. Yus, R. Aebersold, R. Herrmann, B. Bottcher, A. S. Frangakis, R. B. Russell, L. Serrano, P. Bork and A. C. Gavin, *Science*, 2009, **326**, 1235-1240.
22. B. Kuster, M. Schirle, P. Mallick and R. Aebersold, *Nature reviews. Molecular cell biology*, 2005, **6**, 577-583.
23. P. Picotti, H. Lam, D. Campbell, E. W. Deutsch, H. Mirzaei, J. Ranish, B. Domon and R. Aebersold, *Nature methods*, 2008, **5**, 913-914.
24. P. Nordenfelt, S. Waldemarson, A. Linder, M. Morgelin, C. Karlsson, J. Malmstrom and L. Bjorck, *The Journal of experimental medicine*, 2012, **209**, 2367-2381.

25. B. M. Kihlberg, J. Cooney, M. G. Caparon, A. Olsen and L. Bjorck, *Microbial pathogenesis*, 1995, **19**, 299-315.
26. A. Shevchenko, M. Wilm, O. Vorm and M. Mann, *Analytical chemistry*, 1996, **68**, 850-858.
27. M. Wilm, A. Shevchenko, T. Houthaeve, S. Breit, L. Schweigerer, T. Fotsis and M. Mann, *Nature*, 1996, **379**, 466-469.
28. J. N. Cole, T. C. Barnett, V. Nizet and M. J. Walker, *Nature reviews. Microbiology*, 2011, **9**, 724-736.
29. L. N. Mueller, O. Rinner, A. Schmidt, S. Letarte, B. Bodenmiller, M. Y. Brusniak, O. Vitek, R. Aebersold and M. Muller, *Proteomics*, 2007, **7**, 3470-3480.
30. A. Wolf-Yadlin, S. Hautaniemi, D. A. Lauffenburger and F. M. White, *Proceedings of the National Academy of Sciences of the United States of America*, 2007, **104**, 5860-5865.
31. C. Karlsson, L. Malmstrom, R. Aebersold and J. Malmstrom, *Nat Commun*, 2012, **3**, 1301.
32. J. Malmstrom, C. Karlsson, P. Nordenfelt, R. Ossola, H. Weisser, A. Quandt, K. Hansson, R. Aebersold, L. Malmstrom and L. Bjorck, *The Journal of biological chemistry*, 2012, **287**, 1415-1425.
33. C. S. Bates, G. E. Montanez, C. R. Woods, R. M. Vincent and Z. Eichenbaum, *Infection and immunity*, 2003, **71**, 1042-1055.
34. P. Valentin-Weigand, M. Y. Traore, H. Blobel and G. S. Chhatwal, *FEMS microbiology letters*, 1990, **58**, 321-324.
35. B. Akerstrom, A. Lindqvist, C. V. Maelen, A. Grubb, G. Lindahl and J. P. Vaerman, *Molecular immunology*, 1994, **31**, 393-400.
36. M. Nilsson, S. Wasylik, M. Morgelin, A. I. Olin, J. C. Meijers, R. H. Derksen, P. G. de Groot and H. Herwald, *Molecular microbiology*, 2008, **67**, 482-492.
37. B. Dahlback and J. Stenflo, *Proceedings of the National Academy of Sciences of the United States of America*, 1981, **78**, 2512-2516.

## Figures and legends

### Figure 1 (a, b, c): Inventory of human plasma proteins binding to the *S. pyogenes* bacterial surface

Plasma and the plasma-adsorbed proteins were fractionated by SDS-PAGE. The gel was sliced and prepared for MS by in gel trypsin digestion followed by LC-MS/MS analysis and the resulting data was searched against a hybrid *S. pyogenes*-human database using X!Tandem and the transproteomics pipeline (TPP). The features were extracted by SuperHirn to construct the protein extracted ion intensities (XIC). The XIC for the individual proteins were divided by the total ion intensities (TIC) to obtain proportional extracted ion intensities (pXIC). The proteins were divided into groups based on literature searches. **A)** Summary of the number of identified proteins per protein group. **B)** Estimation of the adsorption ratio for the individual proteins at the bacterial surface compared to blood plasma. The adsorption ratio was calculated by dividing the proteins pXIC for plasma-adsorbed protein by the sum of pXIC for plasma and plasma-adsorbed protein, E. g. a protein existing only in plasma-adsorbed sample  $pXIC_{pad}/(pXIC_p+pXIC_{pad}) = pXIC_{pad}/(0+pXIC_{pad}) = 1$ . The adsorption ratio for all 335 proteins found in both plasma and plasma-adsorbed sample can be found in S-table 3. **C)** Average adsorption ratio of the protein groups. The sum of pXIC for all proteins within a group from PA divided by the sum for the same proteins in plasma. The immunoglobulin variable chains were excluded.

### Figure 2 Subdivision of interacting proteins using SRM.

*S. pyogenes* strain AP1 was incubated in plasma and interacting proteins isolated using two different extraction protocols, glycine elution (GE) and double digestions (DD) in quadruplicates. The plasma adsorbed samples and the starting blood-plasma was analyzed using SRM targeting 152 proteins with 2030 transitions, resulting in the detection of 122 proteins divided into five groups. The orange lines in the Figure are based on PCA and k-mean clustering. The average intensities for the glycine elution are displayed as blue circles, double digestion as red squares and plasma as green triangles. The individual proteins were divided by the total intensity for the corresponding protein, so that the total protein intensity sum equals 1 for all proteins, displayed on the y-axis. The x-axis represents proteins sorted by groups and the decreasing plasma intensity

value within each group. The full list of protein can be found in S-table 4 and enlarged figures (including protein name) for each group can be found in S-figure 2a-e.

**Figure 3: Merged of LC-MS/MS and SRM results.**

The shotgun data was obtained by fractionating plasma and plasma adsorbed to bacterial surface using SDS-PAGE followed by LC-MS/MS analysis. From the LC-MS/MS analysis pXIC was estimated by dividing the protein XIC by TIC. The SRM data facilitated the subdivision of the proteins into five groups depending on protein intensity differences between plasma and plasma absorption. The resulting data was merged to estimate protein individual and group intensity difference between plasma and plasma absorption. Not measured proteins in SRM are shown as black pluses, group 1 (High Plasma) in blue circles, group 2 (GE=DD>Plasma) in red squares, group 3 (High DD) in green triangles, group 4 (GE=DD<Plasma) in purple stars and group 5 (High GE) in orange rhombs. The graph displays the pXIC intensity distribution for all proteins detected in plasma **A**) and plasma adsorption **B**) and the summed up total intensity for the groups defined by SRM shown in the insert pie graphs. The colors of the pie graphs are according to the same color scheme as the intensity distribution, group 1 is blue, group 2 is red, group 3 is green, group 4 is purple, group 5 is orange and the proteins that were not measured with SRM is black. The protein order is found in S-table 1 and 2.

**Figure 4: Statistically significant protein abundance changes comparing a wild type strain and an isogenic mga<sup>-</sup> mutant strain**

Plasma adsorption (PA) was performed with a WT strain and an mga<sup>-</sup> isogenic mutant strain in quadruplicates followed by SRM analysis. Statistically significant protein differences was determined using Wilcoxon T-test,  $p \leq 0.05$  and are visualized in the Figure. Ratio between PA and Plasma (P), blue bars left axis, represents data from the LC-MS/MS experiment. Ratio between WT and mga<sup>-</sup> mutant, red bars left axis, represents data from the LC-SRM-MS experiment. Value of 1000 (logarithmic scale axis) indicates that the protein was not detected in plasma sample (only blue bars). The pXIC (relative amount of protein) from PA, green bars right axis, data from LC-MS/MS experiment.

**Figure 5: Statistically significant protein abundance changes comparing two M1 isolates originating from two infection sites within the same individual.**

Bacteria were incubated with plasma, unbound protein washed away and measured with SRM. Statistical significance calculated using Wilcoxon signed-rank test and a p-

value of 0.05. Proteins sorted alphabetically on the x-axis. The relative sum for each protein is represented on the y-axis. All proteins in the graph are statistically significant different between the throat and leg isolate. In addition most of the proteins are also statistically significant different between plasma and both isolates with a few exceptions. Proteins marked with a \* are not significant between the throat isolate and plasma and the proteins marked with \*\* are not significant between the leg isolate and plasma. Data for proteins in figure are available in S-table 5 together with the data for the proteins with non-significant change between the two isolates.

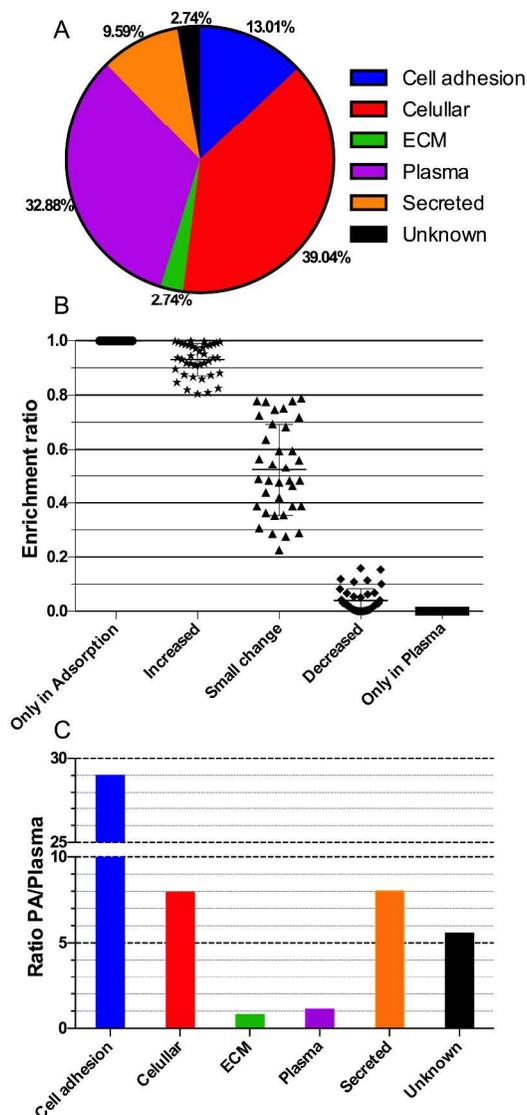


Figure 1. Inventory of human plasma proteins binding to the *S. pyogenes* bacterial surface. Plasma and the plasma-adsorbed proteins were fractionated by SDS-PAGE. The gel was sliced and prepared for MS by in gel trypsin digestion followed by LC-MS/MS analysis and the resulting data was searched against a hybrid *S. pyogenes*-human database using X!Tandem and the transproteomics pipeline (TPP). The features were extracted by SuperHirn to construct the protein extracted ion intensities (XIC). The XIC for the individual proteins were divided by the total ion intensities (TIC) to obtain proportional extracted ion intensities (pXIC). The proteins were divided into groups based on literature searches. A) Summary of the number of identified proteins per protein group. B) Estimation of the adsorption ratio for the individual proteins at the bacterial surface compared to blood plasma. The adsorption ratio was calculated by dividing the proteins pXIC for plasma-adsorbed protein by the sum of pXIC for plasma and plasma-adsorbed protein, E. g. a protein existing only in plasma-adsorbed sample  $pXIC_{pad} / (pXIC_p + pXIC_{pad}) = pXIC_{pad} / (0 + pXIC_{pad}) = 1$ . The adsorption ratio for all 335 proteins found in both plasma and plasma-adsorbed sample can be found in S-table 3. C) Average adsorption ratio of the protein groups. The sum of pXIC for all proteins within a group

from PA divided by the sum for the same proteins in plasma. The immunoglobulin variable chains were excluded.  
177x376mm (600 x 600 DPI)

## Unable to Convert Image

The dimensions of this image (in pixels) are too large to be converted. For this image to convert, the total number of pixels (height x width) must be less than 40,000,000 (40 megapixels).

Figure 2. Subdivision of interacting proteins using SRM. *S. pyogenes* strain AP1 was incubated in plasma and interacting proteins isolated using two different extraction protocols, glycine elution (GE) and double digestions (DD) in quadruplicates. The plasma adsorbed samples and the starting blood-plasma was analyzed using SRM targeting 152 proteins with 2030 transitions, resulting in the detection of 122 proteins divided into five groups. The orange lines in the Figure are based on PCA and k-mean clustering. The average intensities for the glycine elution are displayed as blue circles, double digestion as red squares and plasma as green triangles. The individual proteins were divided by the total intensity for the corresponding protein, so that the total protein intensity sum equals 1 for all proteins, displayed on the y-axis. The x-axis represents proteins sorted by groups and the decreasing plasma intensity value within each group. The full list of protein can be found in S-table 4 and enlarged figures (including protein name) for each group can be found in S-figure 2a-e.

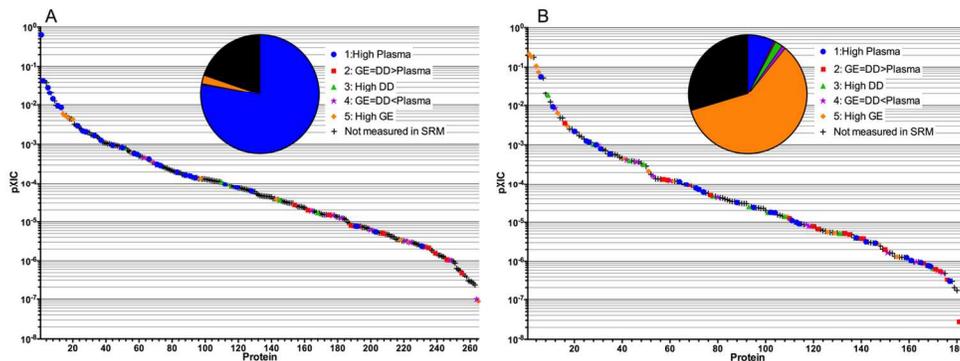


Figure 3. Merged of LC-MS/MS and SRM results. The shotgun data was obtained by fractionating plasma and plasma adsorbed to bacterial surface using SDS-PAGE followed by LC-MS/MS analysis. From the LC-MS/MS analysis pXIC was estimated by dividing the protein XIC by TIC. The SRM data facilitated the subdivision of the proteins into five groups depending on protein intensity differences between plasma and plasma absorption. The resulting data was merged to estimate protein individual and group intensity difference between plasma and plasma absorption. Not measured proteins in SRM are shown as black pluses, group 1 (High Plasma) in blue circles, group 2 (GE=DD>Plasma) in red squares, group 3 (High DD) in green triangles, group 4 (GE=DD<Plasma) in purple stars and group 5 (High GE) in orange rhombs. The graph displays the pXIC intensity distribution for all proteins detected in plasma A) and plasma adsorption B) and the summed up total intensity for the groups defined by SRM shown in the insert pie graphs. The colors of the pie graphs are according to the same color scheme as the intensity distribution, group 1 is blue, group 2 is red, group 3 is green, group 4 is purple, group 5 is orange and the proteins that were not measured with SRM is black. The protein order is found in S-table 1 and 2.

64x24mm (600 x 600 DPI)

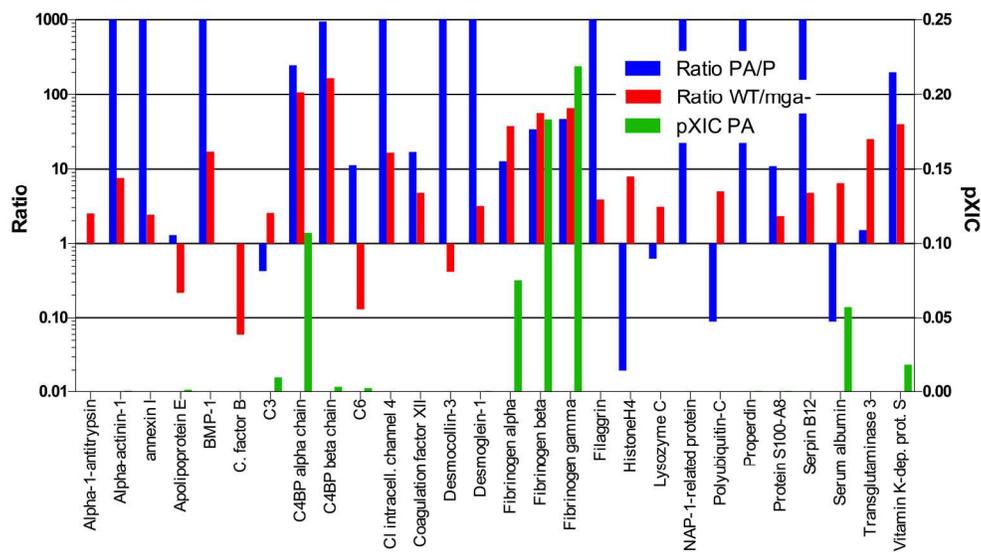


Figure 4. Statistically significant protein abundance changes comparing a wild type strain and an isogenic *mga-* mutant strain. Plasma adsorption (PA) was performed with a WT strain and an *mga-* isogenic mutant strain in quadruplicates followed by SRM analysis. Statistically significant protein differences were determined using Wilcoxon T-test,  $p \leq 0.05$  and are visualized in the Figure. Ratio between PA and Plasma (P), blue bars left axis, represents data from the LC-MS/MS experiment. Ratio between WT and *mga-* mutant, red bars left axis, represents data from the LC-SRM-MS experiment. Value of 1000 (logarithmic scale axis) indicates that the protein was not detected in plasma sample (only blue bars). The pXIC (relative amount of protein) from PA, green bars right axis, data from LC-MS/MS experiment.

96x54mm (600 x 600 DPI)

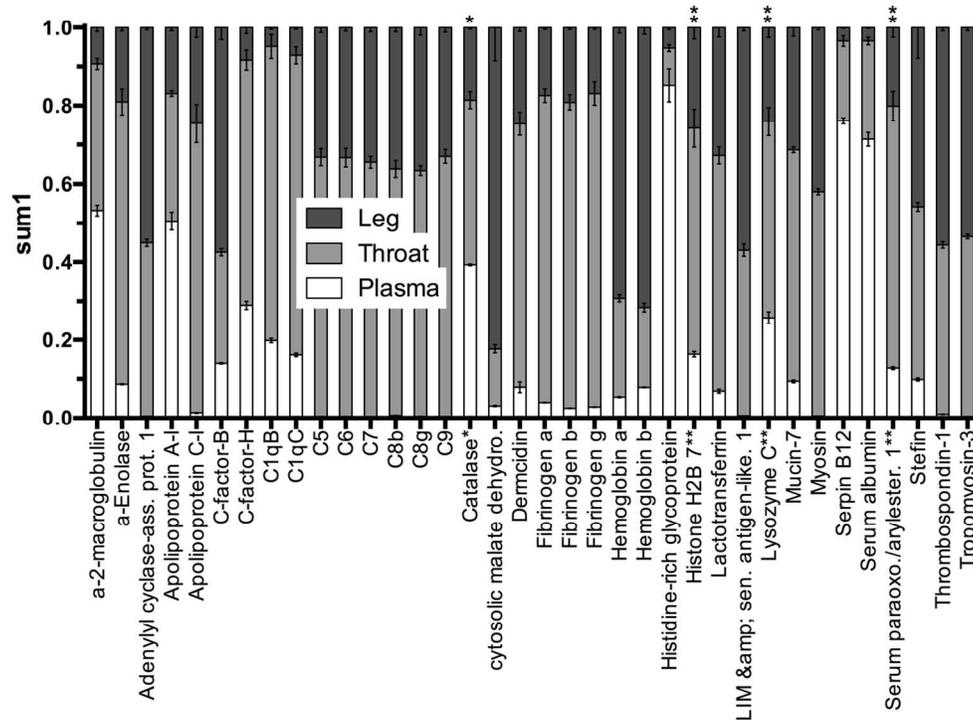


Figure 5. Statistically significant protein abundance changes comparing two M1 isolates originating from two infection sites within the same individual. Bacteria were incubated with plasma, unbound protein washed away and measured with SRM. Statistical significance calculated using Wilcoxon signed-rank test and a p-value of 0.05. Proteins sorted alphabetically on the x-axis. The relative sum for each protein is represented on the y-axis. All proteins in the graph are statistically significant different between the throat and leg isolate. In addition most of the proteins are also statistically significant different between plasma and both isolates with a few exceptions. Proteins marked with a \* are not significant between the throat isolate and plasma and the proteins marked with \*\* are not significant between the leg isolate and plasma. Data for proteins in figure are available in S-table 5 together with the data for the proteins with non-significant change between the two isolates.

67x54mm (600 x 600 DPI)