UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova
Sede Consorziata: NOVARTIS VACCINES INSTITUTE FOR GLOBAL HEALTH, Exploratory Program, *Shigella* project, Siena

Dipartimento di Biologia

SCUOLA DI DOTTORATO DI RICERCA IN: BIOSCIENZE E BIOTECNOLOGIE
INDIRIZZO: BIOLOGIA CELLULARE
CICLO XXIV

OUTER MEMBRANE PARTICLES AS *SHIGELLA* VACCINE CANDIDATE: GENERATION, PROTEOMIC AND IMMUNOPROTEOMIC ANALYSIS

Direttore della Scuola: Ch.mo Prof. Giuseppe Zanotti
Coordinatore d'indirizzo: Ch.mo Prof. Cesare Montecucco
Supervisore interno: Ch.mo Prof. Cesare Montecucco
Supervisore esterno: Dott.ssa Christiane Gerke

Dottoranda: Luana Maggiore

INDEX

ABBREVIATIONS

-Fe	Low iron
+Fe	High iron
-p	Lack of virulence plasmid (either pSS or pINV)
+p	Presence of virulence plasmid (either pSS or pINV)
1-D	Mono-dimensional
2-D	Bi-dimensional
aa	aminoacids
Ab	Antibody
Ag	Antigen
BPI	Bactericidal/permeability increasing protein
CFU	Colony forming units
Cyto	Cytoplasm
DM	Defined medium
Δ*msbB*	Deletion of *msbB* gene
Δ*ompA*	Deletion of *ompA* gene
Δ*rfbG*	Deletion of *rfbG* gene
Δ*tolR*	Deletion of *tolR* gene
Δ*wbg*	Deletion of *wbg* gene
ΔOAg	Lack of O-antigen molecule
EC	Epithelial cells
E. coli	Escherichia coli
EDTA	Ethylendiaminetetracetic acid
ELISA	Enzyme-Linked Immuno-Sorbent Assay
emPAI	Exponentially Modified Portein Abundance Index
ETEC	Enterotoxigenic *Escherichia coli*
Ext	Extracellular
GMMA	Generalized Modules for Membrane Antigens
H	Heavy labeling
HTMC	Yeast extract medium
IgA	Immunoglobulin A
IgG	Immunoglobulin G

IgM	Immunoglobulin M
IL	Interleukin
IM	Inner membrane
IP	Immunoprecipitate
L	Light labeling
LB	Luria Broth
LC	Liquid chromatography
LPS	Lipopolysaccharide
M	Intermediate labeling
MALDI	Matrix-Assisted Laser Desorption Ionization
MS	Mass Spectrometry
N/A	Not available
NF-kB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NLRs	Nod-like receptors
NK	Natural killer
OD	Optical density
OM	Outer Membrane
OM lipo	Outer Membrane lipoprotein
OMP	Outer Membrane Protein
OMV	Outer Membrane Vesicles
OAg	O-antigen
PCR	Polymerase Chain Reaction
PEP	Posterior Error Probability
Peri	Periplasm
PMNs	Polymorphonucleates
rEPA	Recombinant enterotoxin A from *Pseudomonas aeruginosa*
RT	Room temperature
SDS-PAGE	Sodium Dodecyl Sulphate – PolyAcrylamide Gel Electrophoresis
Sf	*Shigella flexneri*
SHIB3	*Shigella boydii 18* database
SHIBS	*Shigella boydii 4* database
SHIDY	*Shigella dysenteriae* database
SHIDS	*Shigella dysenteriae* CDC 74-1112 database

SHIFL	*Shigella flexneri* database
ultra@	ultra-centrifugation
SHIF8	*Shigella flexneri 5b* database
SHISO	*Shigella sonnei 53G* database
SHISS	*Shigella sonnei* 046 database
Ss	*Shigella sonnei*
TOF	Time-of-flight
TOT	Total
TLR	Toll-like receptor
TNF-α	Tumor necrosis factor alpha
WB	Western blotting

1. SUMMARY

Shigellae are Gram-negative enteric bacteria representing the major cause of dysenteric diseases in infants and young children in developing countries. The genus *Shigella* comprises 50 different serotypes characterized by the carbohydrate composition of the O-antigen of the lipopolysaccharide. In natural infection the immune response to the O-antigen has been shown to be protective but serotype-specific, and no vaccine against *Shigella* is currently available. Novartis Vaccines Institute for Global Health (NVGH) aims to develop a broadly protective protein-based vaccine using outer membrane particles. These particles, called Generalized Modules for Membrane Antigens (GMMA), also known as outer membrane vesicles (OMV), are naturally shed from the surface of Gram-negative bacteria and expose the host to proteins of the outer membrane in their native orientation. They are different from detergent-extracted OMV which are depleted of lipooligosaccharides and lipoproteins.

As *Shigella* naturally produce low amounts of GMMA, the *tolR* gene was deleted to induce GMMA overproduction. To avoid immuno-dominant immune responses to the O-antigen, the genes of the O-antigen biosynthesis gene cluster were deleted. In addition, the *msbB* gene was deleted to reduce the endotoxicity of the lipopolysaccharide (LPS). GMMA were generated under iron-restricted conditions to induce the expression of highly conserved iron siderophore receptors that have been shown to be protective antigens against *Salmonella* infection.

The proteomics analysis of *S. sonnei 53G* and *S. flexneri 2a 2457T* GMMA confirmed that GMMA are highly enriched in outer membrane and periplasmic proteins. Approximately 65% and 71% of the total predicted outer membrane proteins of *S. flexneri* and *S. sonnei*, respectively, and 50% of the total predicted periplasmic proteins were found to be present in GMMA. In contrast, only 4% of the total predicted cytoplasmic proteins were indentified in GMMA. We identified 50 outer membrane proteins and 2 predicted extracellular proteins that were conserved in *S. sonnei* and *S. flexneri* GMMA. Among these proteins we identified 3 proteins encoded on the virulence plasmid (MxiD, MxiM and IcsA/VirG), and other 3 proteins with a virulence-associated function (SepA, SigA and TieB) which are specific for *Shigella*

and pathogenic *E. coli*. The proteomic results also confirmed the overexpression of 8 iron-regulated proteins in GMMA of which 5 are conserved in *S. sonnei* and *S. flexneri*. Mice were immunized using purified GMMA and immune sera were analyzed by ELISA and bi-dimensional Western blots. All GMMA elicited similarly high levels of total GMMA-specific IgG. By 2 dimensional Western blots (2-D WB) analysis mainly periplasmic proteins and only 3 outer membrane proteins were identified, suggesting that a different approach was needed to identify outer membrane proteins in their native conformation. In this work we present a novel approach to identify immunogenic surface proteins based on immunoprecipitation studies with live bacteria using sera raised against GMMA. We identified 30 immunogenic surface proteins and among those we found OmpA, OmpX and YaeT confirming our previous 2-D WB results. In addition, we identified proteins that were not found by 2-D WB but are known to be immunogenic in *Shigella,* e.g. TolC, supporting the power of this approach. Furthermore, OmpC, FepA, FhuA, and CirA were identified using this approach suggesting that the native conformation of the epitopes is important for antibody binding and that these epitopes are lost by 2-D WB analysis.

A quantitative proteomic analysis of GMMA proteomes, based on stable isotope dimethyl labeling followed by GeLC-MS/MS, revealed that the majority of the 30 immunogenic proteins are the most abundant proteins in GMMA. Interestingly, YiaD, YtfM and one of the iron-regulated proteins, FepA, raised a strong antibody response but are of low abundance in GMMA.

The majority of the immunogenic proteins identified in *S. sonnei* are also present and highly conserved in *S. flexneri* GMMA. Thus, it is likely that these proteins might raise cross-reactive immune responses.

The present work shows for the first time an extensive qualitative and quantitative proteomic analysis of *Shigella* GMMA. In addition, the analysis of immunogenic surface proteins under native conditions identified proteins that are highly conserved among *Shigella* serotypes. This work provides the basis for further characterization of GMMA as a protein-based vaccine against *Shigella*.

2. RIASSUNTO

La *Shigella* è un batterio enterico Gram-negativo e rappresenta la principale causa di dissenteria nei neonati e nei bambini al di sotto dei 5 anni nei paesi in via di sviluppo. Il genere *Shigella* include 50 serotipi diversi distinguibili in base alla composizione saccaridica dell'antigene O del lipopolisaccaride (LPS). Nell'infezione naturale è stato osservato che l'antigene O è in grado di indurre una risposta immunitaria protettiva contro l'infezione da *Shigella*, ma tale risposta risulta essere specifica contro i singoli serotipi. Non ci sono al momento vaccini disponibili per prevenire l'infezione da *Shigella*. L'istituto "Novartis Vaccines Institute for Global Health" (NVGH) sta lavorando su un progetto per lo sviluppo di un vaccino proteico protettivo e ad ampio spettro per la prevenzione di infezioni da *Shigella* basato sull'utilizzo di particelle di membrana esterna. Queste particelle, chiamate Moduli Generalizzati per l'espressione di Antigeni di Membrana (GMMA) e comunemente conosciute in letteratura come vescicole di membrana esterna (OMV), sono naturalmente rilasciate dalla superficie dei batteri Gram-negativi e presentano all'ospite le proteine della membrana esterna nella loro conformazione originaria. Queste particelle si distinguono dalle OMV ottenute mediante estrazione con detergenti che sono prive di lipo-oligosaccaridi e lipoproteine. Poiché *Shigella* rilascia naturalmente piccole quantità di GMMA, è stata eseguita una delezione del gene *tolR* al fine di indurre un maggior rilascio di GMMA. Per abolire la risposta immunitaria serotipo-specifica dell'antigene O è stata eseguita una delezione del cluster di geni codificanti per l'antigene O. Una successiva delezione del gene *msbB* è stata effettuata per ridurre la reattogenicità delle molecole di LPS. Le GMMA sono state ottenute crescendo i ceppi produttori in condizioni limitanti di ferro al fine di indurre l'espressione di recettori del ferro molto conservati in *Shigella*. Tali recettori si sono dimostrati antigeni protettivi contro l'infezione da *Salmonella*. L'analisi delle GMMA di *S. sonnei 53G* e *S. flexneri 2a 2457T* attraverso approccio proteomico ha confermato che le GMMA contengono prevalentemente proteine di membrana esterna e proteine periplasmatiche. Nelle GMMA abbiamo identificato circa il 65% e il 71% delle proteine di membrana esterna predette in totale rispettivamente in *S. flexneri* e in *S. sonnei* e il 50% delle proteine periplasmatiche. Le proteine citoplasmatiche identificate nelle GMMA rappresentano, invece, solamente il 4%

delle proteine citoplasmatiche totali predette. Abbiamo identificato 50 proteine di membrana esterna e 2 proteine con predizione extracellulare conservate nelle GMMA di *S. sonnei* e di *S. flexneri*. Tra le proteine comuni abbiamo identificato 3 proteine codificate dal plasmide di virulenza (MxiD, MxiM and IcsA/VirG) e altre 3 proteine con una funzione associata alla virulenza (SepA, SigA and TieB) che sono specifiche per *Shigella* e ceppi patogeni di *E. coli*. I risultati dell'analisi proteomica hanno inoltre confermato l'overespressione di 8 proteine ferro-regolate nelle GMMA, di cui 5 sono conservate in *S. sonnei* e *S. flexneri*. Le GMMA sono state utilizzate per l'immunizzazione di topi e i sieri immuni sono stati analizzati mediante saggi ELISA e Western blots bi-dimensionali (2-D WB). Tutte le GMMA testate hanno stimolato in modo equiparabile la produzione di alti livelli di immunoglobuline di tipo G (IgG) GMMA-specifiche. Dalle analisi di 2-D WB abbiamo identificato proteine immunogeniche prevalentemente periplasmatiche e solo 3 proteine di membrana esterna. Tali risultati hanno suggerito la necessità di trovare un approccio alternativo per l'identificazione di proteine di membrana esterna nella loro conformazione nativa. In questo studio presentiamo un approccio per l'identificazione di proteine immunogeniche di superficie basato su studi di immunoprecipitazione su batteri vivi usando sieri immuni specifici contro le GMMA. Abbiamo identificato 30 proteine immunogeniche di superificie e tra queste abbiamo trovato OmpA, OmpX e YaeT confermando i risultati precedentemente ottenuti mediante 2-D WB. Inoltre, abbiamo identificato proteine che sono riportate immunogeniche in *Shigella*, come ad esempio TolC, ma che non sono state rilevate mediante 2-D WB supportando l'efficacia di questo approccio. L'ulteriore identificazione delle proteine OmpC, FepA, FhuA e CirA mediante immunoprecipitazione suggerisce che la conformazione nativa degli epitopi è importante per il legame con l'anticorpo e che tali epitopi vengono persi nelle analisi di 2-D WB. Un'analisi proteomica di tipo quantitativa delle GMMA, basata sull'utilizzo di marcatura isotopica stabile con legame dimetilico seguita da GeLC-MS/MS, ha rivelato che la maggior parte delle 30 proteine immunogeniche sono anche le proteine più abbondanti nelle GMMA. E' interessante notare che nonostante YiaD, YtfM e una delle proteine ferro-regolate (FepA) siano tra le proteine meno abbondanti nelle GMMA, tali proteine hanno indotto una forte risposta immunitaria.

La maggior parte delle proteine immunogeniche identificate in *S. sonnei* sono anche presenti e altamente conservate nelle GMMA di *S. flexneri*. Perciò, è molto probabile che queste proteine siano in grado di indurre una risposta immunitaria cross-reattiva.

Il presente studio mostra per la prima volta un'estensiva analisi proteomica di tipo qualitativa e quantitativa delle GMMA ottenute da *Shigella*. Inoltre, l'analisi delle proteine immunogeniche di superficie condotta in condizioni native ha permesso di identificare proteine altamente conservate tra i serotipi di *Shigella*. Questo lavoro fornisce le basi per una successiva caratterizzazione delle GMMA come vaccino a base proteica contro *Shigella*.

3. INTRODUCTION

3.1. SHIGELLA

3.1.1. Distribution and global impact

Shigella are a major global public health concern in infants and young children in developing countries[1], and are also responsible for the traveler's diarrhea in individuals from industrialized countries visiting developing areas[29]. The majority of shigellosis cases occur in regions of Asia, Africa and Latin America where poor hygiene and limited access to clean drinking water promote the spread of enteric diseases[35,137]. In 1999, the World Health Organization (WHO) estimated 165 million cases of shigellosis per year with 1.1 million deaths in developing countries, mostly in children under 5 years[82]. A recent estimate shows that the number of deaths in Asia has been reduced from 88,000 (80% of the previously estimated 1.1 M deaths were projected to occur in Asia) to 14,000[8]. Most of the endemic disease is caused by *S. flexneri* and *S. sonnei* while the epidemic disease is usually due to *S. dysenteriae* type 1. The latter is the only serotype with a chromosomal gene encoding a 70KDa heterodimeric protein known as Shiga toxin, causing the most severe dysentery.

3.1.2. Characteristics

Shigella are Gram-negative capsulated bacteria, non-spore-forming, non-motile, and facultative anaerobic bacilli belonging to the family of Enterobacteriaceae. The genus *Shigella* can be divided into four serogroups that are historically treated as species[42,90]. Each serogroup comprises different serotypes, characterized by the carbohydrate composition of the O-antigen, the polysaccharide part of the lipopolysaccharide (LPS):

Shigella dysenteriae (serogroup A) – 15 serotypes

Shigella flexneri (serogroup B) – 14 serotypes

Shigella boydii (serogroup C) – 20 serotypes

Shigella sonnei (serogroup D) – 1 serotype

The sequencing of housekeeping genes[113,114] together with data from DNA reassociation tests[16] and studies using multilocus enzyme electrophoresis[104,114] revealed that *Shigella* and *E. coli* belong to the same species. According to the sequences of four regions of the chromosome of all *Shigella* serotypes most *Shigella*

fall into 3 clusters (clusters 1, 2, and 3)[114]. *Shigella dysenteriae* serotype 1, 8, and 10, *Shigella sonnei*, and *S. boydii* serotype 13 are not included in any of the clusters[114]. *S. boydii* serotype 13 was shown to diverge from the main *Shigella-E. coli*, even retaining the biochemical identification of the *S. boydii* serogroup[16].

Shigella are human adapted pathogens that are capable of colonizing the intestinal epithelium by exploiting epithelial cell functions and avoiding the host innate immune response. The ability of *Shigella* to invade epithelial cells and cause dysenteric disease is related to the presence of a high molecular weight and low copy-numbers plasmids, collectively termed pINV plasmids (pINV)[56] with minor variations between serotypes. These virulence plasmids are approximately 230Kb large, are non-conjugative[92,121], and encode for the major virulence determinants[107], including invasins, molecular chaperones, motility, regulation, and a specialized type III secretion system[87]. Additional virulence-associated genes are located in pathogenicity islands on the chromosome and contribute directly or indirectly to the pathogenic process[61].

3.1.3. Shigellosis

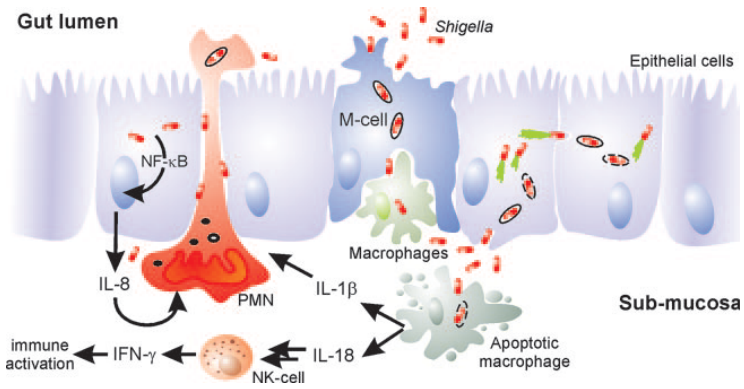
3.1.3.1. Definition and transmission

Shigella are the etiological agent of shigellosis, an acute intestinal infection that is transmitted by person-to-person contact through the fecal-oral route or indirectly through contaminated food or water. Shigellosis is highly infectious as few bacteria (ranging from 10 to 100 for *S. dysenteriae* type I) can be sufficient to cause the disease[35]. It produces a spectrum of clinical symptoms that can range from mild watery diarrhea to classic severe dysentery characterized by high fever, nausea, strong abdominal cramps and diarrhea containing blood and mucus. The disease is usually self-limiting. In rare cases complications and long term effects, such as kidney failure and reactive arthritis, can occur in children. Additionally, shigellosis may become severe in immunocompromised patients, including neonates and children with underlying immune deficiencies (including HIV infections) or malnutrition, or if there is no adequate medical care.

3.1.3.2. Pathogenesis

The low infectious dose of *Shigella* species can be partially related to the ability of *Shigella* to survive in the acidic conditions of the stomach, through specific acid resistance systems[55]. In addition, *Shigella* can decrease the expression of antibacterial peptides that are constantly released from the mucosal surface of the intestinal tract[63]. Once the bacteria pass through the stomach and small intestine they reach the large intestine where they establish the infection. The major part of what is currently known about the mechanisms of *Shigella* pathogenesis is derived from studies of *S. flexneri*[124] (figure 1). The infection is a multi step process that starts with the translocation through M cells. Since *Shigella* are unable to invade epithelial cells through the apical pole they use M cells to pass the epithelium by transcytosis. M cells are specialized epithelial cells that sample the luminal content and deliver it to the underlying mucosal lymphoid tissue. They have a poorly differentiated brush border and show invaginations at the basolateral side containing lymphocytes. After transcytosis the bacteria are taken up by resident macrophages which, instead of destroying the bacteria in the phagosome, succumb to apoptotic death[149]. Free bacteria, released from dead macrophages, invade the epithelial cells from the basolateral side. They escape from the phagosome into the cytoplasm and subsequently spread to adjacent cells in an actin-dependent process through the alteration of the tight-junction proteins composition[120]. Macrophage cell death is accompanied by the release of many pro-inflammatory cytokines, as IL-1β, IL-18, and IL-8 inducing an acute and massive inflammatory response. IL-1β and IL-8 induce the recruitment of polymorphonuclear cells (PMNs), that infiltrate the infected site and destabilize the epithelium[109,110]. The influx of PMN compromises the integrity of epithelial barrier facilitating the invasion of further organisms from the colonic lumen. These processes are essential for the development of the diarrhea. Ultimately, however, the PMNs phagocytose and kill *Shigella* contributing to the resolution of infection. This is aided by the generation of an effective antibacterial response induced by the released IL-18. IL-18 activates natural killer (NK) cells which induce and promote the production of IFN-γ, that is responsible for the activation of macrophages, amplifying innate immune response[88,138]. The alteration and destruction of the tissue in the early stages of the infection causes an impaired

adsorption of solutes, nutrients, and water, which is subsequently followed by watery diarrhea and presence of blood and mucus in stools.



Re-printed from Schroeder and Hilbi
Clin Microbiol Rev. 2008

Figure 1. Cellular pathogenesis of *Shigella* spp.
S. flexneri pass the epithelial cells barrier by transcytosis through M cells and encounters resident macrophages. The bacteria evade degradation in macrophages by inducing an apoptosis-like cell death, which is accompanied by proinflammatory signaling. Free bacteria invade the epithelial cells (EC) from the basolateral side, move into the cytoplasm by vectorial actin polymerization, and spread to adjacent cells. Proinflammatory signaling by macrophages and EC further activates the innate immune response involving natural killer cells and attracts PMN. The influx of PMN disintegrates the EC lining, which initially exacerbates the infection and tissue destruction by facilitating the invasion of more bacteria. Ultimately, PMN phagocytose and kill *Shigella*, thus contributing to the resolution of the infection[124].

3.2. CURRENT STATUS OF VACCINE DEVELOPMENT

The estimate of disease burden and mortality rates caused by shigellosis during the last 50 years, together with the increasing antibiotic resistance, raised the interest of research in identifying vaccines to prevent shigellosis[82]. Epidemiologic studies [45], re-challenge of volunteers [80], and animal[47] studies have shown that protective immunity against natural *Shigella* infection is directed against the O-antigen of the LPS molecule, and is serotype-specific. As *Shigella* comprise 50 different serotypes, the main obstacle in generating a broadly protective vaccine against *Shigella* is the specificity of the O-antigen. Several studies aimed to test vaccine candidates containing O-antigen showed that the O-antigen is highly immunogenic[111,129], but no cross-protection to different serotypes was observed[45]. Currently, no *Shigella*

vaccine is available. Several vaccines using live-attenuated or killed bacteria or bacterial components have been developed and tested in different clinical trails.

3.2.1. Live attenuated vaccines

Many live attenuated vaccines have been developed and tested in different Countries. However, only in China a live bivalent vaccine was licensed in 1997. This is a combination of *S. flexneri* 2a and *S. sonnei* live attenuated strains that were tested in a clinical trial showing a protective efficacy of approximately 60-70% against the homologous strains, and approximately 50% efficacy against heterologous strains (*S. sonnei* and *S. flexneri 2a*). However, due to the need of three immunizations with high dose of bacteria the vaccine was never clinically evaluated outside China[85].

Other live attenuated vaccine candidates that have been clinically evaluated include a live-attenuated *S. flexneri 2a*, SC602, attenuated by impairment of spreading from cell to cell (deletion of *virG/IcsA*) and impairment of iron uptake by inhibition of aerobactin synthesis (deletion of *iuc*), developed at Institut Pasteur. When the vaccine was tested in phase I and phase II trials in North America volunteers it was shown to be well tolerated and immunogenic[74]. A recent trial conducted in healthy adults and children in Bangladesh showed that SC602 vaccine was safe in both adults and children, but a minimal reactogenicity and immunogenicity were observed[115]. The limited immunogenicity was attributed to a possible low uptake of the vaccine in the volunteers, and to the lack of gut colonization in children.

Another vaccine (SC599) containing *S. dysenteriae 1* attenuated by deletion of spreading from cell to cell (*virG/icsA*), iron chelation (*ent*, *fep*), and Shiga toxin A subunit (*stxA*) genes was developed at the Pasteur Institute. The vaccine was shown to be well tolerated in phase I safety trials conducted in UK, but showed modest or absent serum antibody levels[119]. A phase II trials conducted in healthy human volunteers, in which colonization of SC599 was allowed to proceed differently from the phase I trials, showed that one single oral immunization of SC599 vaccine is well tolerated and elicits immune responses (anti-LPS IgA and IgG antibody secreting cells and serum anti-LPS IgA and IgG) that may confer protection against shigellosis[86].

A series of *S. flexneri 2a 2457T* attenuated vaccines (CVD) was developed by the University of Maryland Center for Vaccine Development in USA. CVD1203 vaccine carrying deletions of *aroA* (involved in intracellular proliferation) and *virG* was well tolerated at a dose of 10^6 CFU when tested in a phase I trial in North American adults, but caused unacceptable reactogenicity[81] together with strong immune responses when administered at doses of 10^8 CFU or 10^9 CFU. CVD1207 vaccine carrying the deletions of *virG*, *guaBA* (encoding enzymes involved in guanine-nucleotide biosynthesis), *sen* (encoding ShET2) and *set* (encoding ShET1) genes was shown to be hyper-attenuated and well tolerated in phase I clinical trials with North American volunteers, but was not sufficiently immunogenic[85]. CVD1208 vaccine, harbouring deletions of *guaBA*, *sen* and *set* genes showed a better balance between clinical acceptability and robust immunogenicity than most attenuated strains representing an attractive vaccine candidate[90].

The main difficulty in using the live-attenuated vaccines is mostly related to the lack of a right balance of clinical acceptability and robust immunogenicity. Although this balance was probably achieved with SC602 vaccine in industrialized countries, the same vaccine did not show the same effect in developing countries.

3.2.2. Polysaccharide conjugate vaccines

Different studies have shown that levels of O-antigen specific IgG-antibodies in the serum inversely correlate with the probability immunity against shigellosis in humans[22,117]. Also, natural infection only protects against the homologous serotype, suggesting that the immune response to the O-antigen is important. Based on these observations O-antigen-conjugates were developed as vaccine candidates against *Shigella*. In these vaccines, the O-antigen from a major bacterial serotype is linked to a carrier protein. In particular, the O-antigen from *S. dysenteriae* type 1 was conjugated to tetanus toxoid, and O-antigens from *S. flexneri* and *S. sonnei* were conjugated to the recombinant exotoxin A from *Pseudomonas aeruginosa* (rEPA) or to the CRM9-mutant diphteria toxin[90]. A phase II clinical study in Israeli volunteer soldiers revealed that parenteral vaccines consisting of *S. sonnei* or *S. flexneri 2a* O-antigens conjugated to rEPA induced high concentrations of anti-LPS antibodies, similarly to those present after natural infection. The levels of serum IgG and IgA

anti-LPS remained elevated 2 years after vaccination[22]. In the subsequent phase III trial, vaccination with *S. sonnei* O-antigen-rEPA conjugate protected against *S. sonnei* infection with 74% efficacy [23]. In age de-escalation studies both of the *S. sonnei* and *S. flexneri 2a* rEPA-conjugates were shown to be safe and immunogenic in 4-7 years old children, but only *S. flexneri* conjugate elicited a booster effect after the second vaccination[6]. And a recent phase III trial revealed that *S. sonnei* and *S. flexneri* rEPA-conjugates are safe in 1-4 years old children. However, the immunogenicity and efficacy were age-related, as 71.1% of efficacy was observed in the 3-4 years old, 35.5% in the 2-3 years old, but no efficacy was observed in the 1-2 years old children[108].

An alternative strategy for polysaccharide conjugates is the conjugation of synthetic oligosaccharides onto specific carriers. Three repeating units of the *S. flexneri* 2a O antigen were shown to induce an efficient anti-LPS antibody response[111].

3.2.3. Vaccine for broad-spectrum protection

There is widespread recognition of the need of vaccines cocktails to protect against the broad-spectrum *Shigella* serotypes (*S. dysenteriae* 1, *S. sonnei* and all 14 strains of *S. flexneri* serotypes)[85,90]. A bivalent vaccine containing a combination of attenuated *S. flexneri 2a* and *S. flexneri 3a* was tested in the guinea pig keratoconjunctivitis model[103]. The mixture induced the production of serum and tear antibodies that cross-reacted with LPS from heterologous *Shigella flexneri* serotypes 1a, 1b, 2b, 4b, 5b, Y but not with LPS from serotype 6. The vaccine candidate conferred significant protection against challenge with the homologous serotypes and *S. flexneri* serotypes 1b, 2b, 5b, and Y, but not serotype 6 in accordance with the cross-reactivity data. Interestingly, no protection was observed against serotypes 1a, and 4b despite the confirmation of cross-reactive LPS-antibodies. According to these results, the combination of *S. flexneri 2a* and *S. flexneri 3a* serotypes showed a broad cross-protection, albeit not complete. As may be expected, given the antigenic diversity of type and group factors within the *S. flexneri* serotypes, a high variation of cross-protection was observed[103].

A *Shigella*-ETEC vaccine was developed utilizing live-attenuated *Shigella* strains (*S. dysenteriae 1*, *S. flexneri 2a* and *S. sonnei*) each expressing one or two critical

enterotoxigenic *Escherichia coli* (ETEC) antigens[9]. Intranasal vaccination of guinea pigs with the individual live vectors or with combinations of 2 or all 3 strains elicited strong serum and mucosal responses against each live vector component as well as each ETEC antigen. Further, the induced immune responses were able to confer protection against homologous wild type but not against heterologous *Shigella* challenge.

3.2.4. Other vaccine candidates

Other approaches to develop vaccine to protect against shigellosis are based on the association of LPS or O-antigen with meningococcal proteosomes and ribosomes[85] to amplify the immune response to the carbohydrate. Vaccines containing *S. flexneri 2a* and *S. sonnei* LPS complexed with meningococcal outer membrane protein proteosomes were tested by oral and intranasal vaccination[105]. Immunization studies in mice showed that the proteosomes-LPS complexes induced specific homologous anti-LPS antibodies in serum and mucosal secretions. Further, challenge studies using the keratoconjunctivitis guinea pig model showed that the orally or intranasally administered proteosomes-LPS complexes protected against homologous bacteria. An intranasal immunization study conducted in healthy adults showed that *S. flexneri 2a* proteosome-LPS was well tolerated and elicited *S. flexneri 2a* LPS-specific antibodies (IgA, IgG and IgM)[48]. The level of the antibody responses was similar to those observed with live vaccine candidates associated with protective efficacy in human challenge models.

Shigella ribosome-based vaccines were shown to induce immunogenicity and protection in mice after intranasal vaccination. A formulation using a macromolecular bacterial complex named Invaplex 50, containing LPS and proteins, including invasins IpaB, IpaC, and IpaD, showed a robust and protective immune response in the guinea pig model. The first intranasal vaccination study in adult volunteers revealed that the Invaplex was safe and immunogenic. Studies on dose-optimization and improvement of vaccine uptake are currently ongoing[70,132].

3.2.5. Animal models

The lack of a relevant animal model that mimics human bacillary dysentery is another obstacle for the development of a successful *Shigella* vaccine. However, different animals have been considered as possible models for *Shigella* infection: guinea pigs, rabbits, mice, and macaques[85]. The keratoconjunctivitis assay, known as Sereny test, conducted on guinea pigs was long considered to be the gold standard test for protective immunity upon vaccination[85]. However, the main difficulties in the use of this model is the irrelevance of the target organ and the quantitative description of the inflammatory response[85]. Also, while the model was shown to be useful to assess the protective efficacy of live-attenuated strains, its value for assessing protection after parenteral immunization is unclear.

A currently widely accepted small animal model is the murine model of pulmonary *Shigella* infection. In this model, mice are infected with *Shigella* via the intranasal route resulting in bacterial invasion of bronchial and alveolar epithelia. This leads to the development of acute bronchiolitis and subsequent lethal pneumonia[135]. The characteristics of the pulmonary lesions observed in infected mice are similar to the colitis developed in humans and primates with shigellosis. In addition, it has been used for the characterization of cellular and humoral immune responses against *Shigella* infection[135] and been found to present similar immune responses as those detected in intestinal shigellosis. Thus, the model has been used to assess protective immune response[85,135] and good protection has been achieved using intranasal immunization with subunit vaccines.

More promising models have been developed to study the pathogenesis of *Shigella* but many of these have limitations as models to assess active immunization against *Shigell*a. For examples, recently, a model of intragastric infection in newborn mice has been developed. Although the mice show inflammatory destruction of the mucosa and infiltration of PMNs into the gut, this model cannot be used to evaluate protective active immunity, as the mice must be infected within 4-5 days after birth[43]. Another model uses human intestinal xenografts in severe combined immunodeficient (SCID) mice. The mice expresses high levels of human IL-1β and IL-8 and injection of *Shigella* into the lumen of the intestinal xenograft[145] results in a marked infiltration of neutrophils . Thus, this model is a viable option for studying

the interactions between *Shigella* and the human intestine. But so far, no protection studies using this model have been reported. The Rabbit ileal-ligated loop model is characterized by the ability of *Shigella* to induce bacterial invasion and inflammatory rupture of the epithelial cells within 18-24 hours after infection. Even though this model is sensitive to virulence attenuation, the likelihood to use it for protection studies has still to be determined[85].

The closest model to human infection is a non-human primate model. In Rhesus monkeys typical bacillary dysentery can be induced by oral infection with *S. flexneri* without pretreatment with antibiotics or starvation[85]. However, ethical, financial, and logistical considerations limit the testing of vaccine candidates in Non-human primates. Recently, a new guinea pig model for bacillary dysentery has been investigated. Guinea pigs were infected by intra-rectal route with *S. flexneri 2a* or *S. flexneri 5a* strains and developed severe and acute rectocolitis mimicking human shigellosis[127]. This model has been used for assessing the protective efficacy of *Shigella* vaccines, such as the live attenuated SC602 strain, suggesting that the model may be useful for evaluating protective immunity of other *Shigella* vaccine candidates[85].

3.2.6. Need for alternative approaches

As described in the previous paragraphs all current vaccine candidates have been designed and developed with the aim to stimulate an O-antigen specific immune response. Although attractive vaccines have entered clinical trials, further studies are needed to assess protection in humans and, most importantly, to understand if they can protect pediatric populations in developing countries[90]. Due to the serotype-specificity of the O antigen, the current candidates are unlikely to induce a broadly protective immune response against the main 16 *Shigella* serotypes and, thus, alternative approaches are needed. A combination of different O-antigens might possibly be able to cover a large percentage of *Shigella* strains. However, the recent emergence of non typable strains and the possibility of strain replacement will likely require constant updates of a multivalent O antigen vaccine. An alternative strategy to develop a vaccine with broad-coverage will be to focus on bacterial

structures that are conserved among different serotypes, such as highly conserved outer membrane proteins.

3.3. GENERALIZED MODULES FOR MEMBRANE ANTIGENS

Both pathogenic and non-pathogenic species of Gram-negative bacteria release blebs from the surface of the outer membrane into the extracellular space. These blebs are usually called outer membrane vesicles (OMV). But, frequently this term is also used to describe detergent-extracted vesicles, which are obtained from detergent and denaturing treatments of whole bacteria. Such treatments cause the removal of lipooligosaccharides and lipoproteins, altering the composition of the native vesicles[44]. The native vesicles are, instead, not depleted of any components and present the outer membrane proteins in their native conformation and membrane environment. To distinguish between the native blebs and the detergent-extracted vesicles we named the native blebs GMMA for Generalized Modules for Membrane Antigens.

The release of GMMA has been observed for a wide variety of Gram-negative bacteria including *Escherichia coli*, *Neisseria meningitidis*, *Pseudomonas aeruginosa*, *Shigella*, and *Helicobacter pylori*[31,46,58,67]. This phenomenon represents a constant process that has been observed in all stages of growth and in a variety of growth environments, including liquid cultures, solid culture, and biofilms[14]. McBroom and Kuehn (in 2007[98]) demonstrated that the production of GMMA is strongly related to the bacterial stress response, and that it increases during periods of bacterial stress, such as the colonization of the host tissues.

3.3.1. Characteristics

GMMA are closed spherical particles characterized by a bilayer membrane containing electron-dense luminal content[20]. They have a heterogeneous size ranging from 50 to 250nm in diameter, and reflect the composition of the outer membrane. GMMA contain LPS, glycerophospholipids, outer membrane proteins, and enclosed periplasmic components. Due to their lipid content, GMMA can be separated from soluble proteins by density gradients. They can also be pelleted by

ultra-centrifugation[44,59]. Biochemical analysis of GMMA purified by density gradient confirmed that GMMA contain proteins and lipids of the outer membrane and periplasm[97]. Moreover, GMMA contain adhesins, toxins, invasins, and immunomodulatory compounds[76,83]. Although GMMA are enriched in envelope components, several proteomic studies showed the presence of some cytosolic and inner membrane proteins in the GMMA preparations[12,89]. In many cases, specific proteins have been shown to be enriched or excluded from GMMA suggesting a specific sorting mechanism for these proteins[59,73]. These observations, together with the presence of toxins and virulence factors in GMMA, suggest that they represent a secretory pathway for the transport of bacterial lipids, membrane proteins and virulence factors to host cells and other bacteria. Moreover, GMMA represent a secretion strategy for soluble proteins that can be released in a protective complex and allow the delivery of proteins at high concentration and in close proximity to other factors[84].

3.3.2. Mechanism of bacterial formation

Although the observation that Gram-negative bacteria produce GMMA was made over 40 years ago, very little is known about the mechanism by which they are produced. Some studies suggest that lipoproteins play a role in GMMA formation as they are covalently attached to peptidoglycan and act to cross-link it to the outer membrane. A study conducted in E. coli showed that GMMA contained less lipoprotein than the outer membrane of the bacterial cells, suggesting that GMMA must originate from regions of the membrane characterized by few peptidoglycan-outer membrane linkages[139]. Other studies provide evidence that lipoproteins involved in synthesis/degradation of peptidoglycan are essential for the GMMA formation, indicating that the peptidoglycan turnover is critical for the blebbing process[38]. Based on these results, different models of GMMA formation have been proposed[94] (figure 2). In the model 1, GMMA originate from regions on the cell surface where peptidoglycan-associated lipoproteins are missing, resulting from the faster expansion of the outer membrane than the underlying peptidoglycan layer[139]. In the model 2, peptidoglycan fragments generated during normal turnover are not efficiently transported back into the bacterial cytoplasm, and build up locally in the

periplasmic space. These products exert a turgor pressure on the outer membrane causing its blebbing[147]. Studies from P. aeruginosa propose a new model, model 3, in which quinolone signal molecules sequester the positive charge of Mg2+, resulting in enhanced anionic repulsion between LPS molecules and thereby in membrane blebbing[94]. These three models are not mutually exclusive, and the blebbing process might be a result of all of these phenomena or combinations of them. The basic mechanism of GMMA formation suggests that this process will probably not take place in regions of the outer membrane containing strong linkages between the outer membrane and the peptidoglycan.

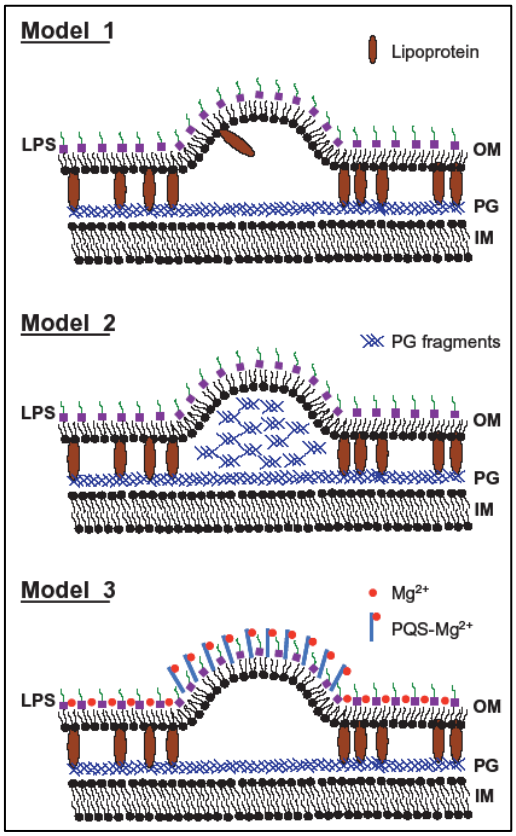


Figure 2: Proposed models for Gram-negative bacterial GMMA formation.
Model 1: GMMA originate from regions on the cell surface where peptidoglycan-associated lipoproteins are missing due to the faster expansion of the outer membrane than the underlying peptidoglycan layer. **Model 2:** accumulation of peptidoglycan fragments in the periplasm causes increased turgor pressure, thereby increasing blebbing of the outer membrane. **Model 3:** ionic interactions between quinolone (PQS) and Mg^{2+} in the *P. aeruginosa* outer membrane enhances anionic repulsion between LPS molecules resulting in membrane blebbing. (Reprinted from Mashburn et al, 2005).

3.3.3. Biological roles of GMMA release

3.3.3.1. Interbacterial activities

Some bacteria use GMMA to gain a survival advantage over other bacteria by their capability to eliminate competing bacterial strains. The presence of periplasmic peptidoglycan hydrolases in GMMA, for example, can lead to the killing of co-cultured bacteria[66,21] . Since GMMA can fuse to the outer membrane of several Gram-

negative bacteria [67] and release their luminal content into the receiving strain, they can allow the receiving bacteria to 'share' the antibiotic resistance proteins[21]. E.g. GMMA containing β-lactamase[21] can confer antibiotic resistance to non-β-lactamase producing bacteria trafficking β-lactamase from one cell to another. GMMA act as bridging factors in biofilms, mediating the coaggreagation of bacteria and producing an environment resistant to antibiotics, and enhance colonization[69]. GMMA seem to play a role in the transfer of beneficial material between bacteria enhancing the genetic diversity and bacterial survival[32]. In addition GMMA are used for communication in environments with mixed pathogen populations, by participating in quorum sensing[94].

3.3.3.2. Delivery of virulence factors

GMMA have been found in human biopsies[75] and infected tissues[142] suggesting that they are produced during the infection. The discovery of GMMA in fluid from infected hosts[33], demonstrates that they are able to disseminate away from the direct site of bacterial colonization. Based on these findings GMMA may play a role in the pathogenic process, and may be beneficial for the bacteria (e.g. mediating the inflammatory damage, delivering toxins and virulence factors) or the host (e.g. being an indicator of infection).

As they are released into the extracellular space, GMMA are able to modify the bacterial environment contributing to the virulence[39]. They represent natural vehicles for the transport of bacterial virulence factors into host cells. Among the virulence factors delivered, toxins have been found associated with GMMA from a variety of Gram-negative bacteria[12,59,95]. In some cases GMMA-associated toxins are more active than the secreted toxin alone[83]. The Shiga toxin[37], for example, is an extracellular secreted toxin expected to be found in the lumen of GMMA. However, the toxin has been found to be partially sensitive and partially resistant to proteases, suggesting a simultaneous localization on the surface and in the lumen[144]. Also, the heath-labile toxin (LT) from enterotoxigenic *E. coli* can be found associated with LPS on the surface of GMMA or in the lumen[60]. Studies conducted on *H. pylori* GMMA revealed that the truncated form of the VacA toxin (without secretion signal) can induce apoptosis in gastric epithelial cells as well as the full-length form[7]. Since the

truncated form contains the active domain but not the secretion signal peptide, it was not expected to be found in GMMA. Thus, it was proposed that GMMA can mediate the toxicity effect of the truncated form of VacA by allowing its presentation to host cells. Together with toxins, GMMA can deliver and secrete other periplasmic and membrane components associated with virulence, for example adhesins, proteases, and signal molecules like quorum-sensing molecules[94].

3.3.3.3. Interactions with hosts cells

GMMA can interact with host cells through a membrane fusion event or via the interaction adhesin/adhesin-receptor molecules present on the surface of the host cells[39]. In some cases, adherence is followed by GMMA internalization[49]. It has been shown that the adherence process can be mediated by toxins, for example the VacA[46] in *H. pylori*, but it can also depend on common components of GMMA. For example, OmpA enhances the ability of *E. coli* K1 to bind and invade brain microvascular endothelial cells by interacting with the host surface receptor EcgP[112]. Specific host factors as bactericidal/permeability-increasing protein (BPI) in *N. meningitidis* have also been shown contributing to GMMA uptake[126].

3.3.3.4. Modulation of host response

In addition to specific virulence factors and molecule, GMMA contain the immunomodulatory LPS molecule and outer membrane porins that can activate immune cells via Toll-like receptors (TLRs), and other molecules, which are able to activate the host innate and acquired immune response pathways[2,51].

The ability of GMMA to generate inflammatory responses has been observed in several pathogens as *H. pylori* and *P. aeruginosa*. Similarly to the whole cells, *Salmonella* GMMA activate macrophages and dendritic cells inducing production of proinflammatory mediators such as IL-12 and TNF-α[4]. LPS, lipoproteins, and outer membrane porins activate immune cells via Toll-like receptors (TLR) stimulating proinflammatory cytokine secretion and triggering the leucocyte recruitment and migration. GMMA-associated toxins and toxin-free GMMA can stimulate the production of IL-8 enhancing inflammation[64,83]. GMMA can allow the escape of the

bacteria from the immune detection during colonization. The gingipain protease on *Porphyromonas gingivalis* GMMA, for example, degradates the lipopolysaccharide receptor on the surface of human macrophages-like cells deactivating the responsiveness of the immune system in the periodontal disease [36]. The general composition of GMMA seems to modulate the host response in two synergic but opposite processes in the hosts: stimulation of pathogen clearance on one side, and enhancing of the virulence of the infection on the other side[39].

3.3.3.5. Potential application of GMMA in vaccine development

As GMMA originate from the outer membrane of Gram-negative bacteria, their composition reflects the surface composition of the donor bacteria. It has been previously shown that GMMA deliver molecules that are enriched in GMMA compared to the outer membrane of the parental bacteria[73,97]. The presence of virulence-associated factors and outer membrane proteins, together with their biological role in bacterial survival, cell-cell signaling and host interactions, make GMMA good candidates as vaccine agents[14]. However, a major obstacle in using GMMA as vaccine is related to the endotoxic activity of LPS. A detoxified LPS would be needed to administer GMMA by parenteral vaccination, otherwise the use of GMMA would be limited to oral route administration[14].

The development of GMMA-based vaccines has been investigated for a variety of pathogens, showing good potential application in the prevention of the diseases. A multivalent vaccine, based on GMMA derived from 3 genetically modified *N. meningitidis* group B strains, has been shown to be safe and to elicit antibodies against multiple *Neisseria* strains suggesting that the vaccine might be potentially broadly protective. Further clinical trials studies for this vaccine are underway[148]. GMMA from *Vibrio cholerae* showed the capacity to induce protective antibody response in the neonatal mouse model of passive antibody transfer[123].*S.* Typhimurium derived GMMA have been shown to stimulate protective immunity in vivo[4]. They induced both B and T cell responses indicating that GMMA can be used as antigen delivery system to produce effective antibody responses. GMMA derived from *Acinetobacter baumannii* provided protection against clinical isolates in a mouse model of sepsis[99] producing a robust antigen-specific IgG and IgM response.

Several studies revealed, moreover, that GMMA can additionally be used as delivery system for heterologous antigens and can enhance the protective efficacy of other vaccine candidates through their adjuvant activity[101].

GMMA obtained from *Edwardsiella tarda* tested in a vaccine trial were demonstrated to be able to confer protective immunity[106]. Further, when tested in a vaccine trial *E. tarda* GMMA were demonstrated to confer protective immunity. GMMA obtained from *Bordetella pertussis* containing full[118] or mutated lipooligosaccharide[5] were demonstrated to be protective in the murine *B. pertussis* intranasal challenge model.

In conclusion, the presence of virulence factors and native bacterial surface molecules, together with the adjuvanticity effect make GMMA a good vaccine candidate for broad spectrum protection against Gram-negative bacteria.

3.4. AIMS OF THE PHD PROJECT

The present work is part of the program at Novartis Vaccines Institute for Global Health (NVGH) to develop a broadly protective vaccine against *Shigella* using GMMA. Aims of the PhD project are to generate GMMA-overproducing *Shigella* strains and to characterize the protein composition of GMMA and identify immunogenic proteins. Characterizing the protein composition of GMMA obtained from different *Shigella* strains represents the key point for the identification of conserved proteins among *Shigella* serogroups. Further, it is important to identify which proteins within GMMA can induce an immune response in the host in order to evaluate if these antigens are conserved in the majority or in the most important *Shigella* serotypes.

4. MATERIALS AND METHODS

4.1. CONSTRUCTION OF SHIGELLA MUTANTS

Two wild type strains of *Shigella* (*Shigella sonnei 53G* and *Shigella flexneri 2a 2457T*) were chosen as the parent strains for preparing the following genetically modified strains: Ss –p, Ss –p Δ*tolR,* Ss –p Δ*tolR* Δ*msbB,* Ss –p Δ*tolR* Δ*ompA,* Ss Δ*tolR* Δ*wbg,* Sf Δ*tolR* Δ*rfbG*, Sf –p Δ*tolR* Δ*rfbG* (Table 1). All the *S. sonnei* strains are O-antigen deficient. In the *S. sonnei* strains indicated as "-p" the virulence plasmid (pSS) was cured, and in 4 of them the *tolR* gene was also deleted. The difference among Ss –p Δ*tolR,* Ss –p Δ*tolR* Δ*msbB*, and Ss –p Δ*tolR* Δ*ompA* is related to the additional deletions of the *msbB* and *ompA* genes respectively in last two strains. Ss Δ*tolR* Δ*wbg* is an O-antigen deficient strain, but maintaining the virulence plasmid, and is deficient also for the gene tolR. All the *S. flexneri* mutants were obtained deleting the *tolR* gene. In Sf Δ*tolR* Δ*rfbG* and Sf –p Δ*tolR* Δ*rfbG* the *rfbG* gene was also deleted, and the difference between the two is related to the absence of the virulence plasmid (pINV) in the latter strain. The deletions of *tolR*, *msbB*, *ompA*, *wbG* and *rfbG* were obtained by substitution of the gene coding sequence with an antibiotic marker[30]. In particular, kanamycin was used to replace the *tolR* gene, erythromycin for *msbB* and *rfbG* genes, and chloramphenicol for *ompA* gene. For the *tolR* mutants a three step PCR approach[30] was used while for the other mutants a standard cloning procedure.

ABBREVIATION	STRAIN	SOURCE
Ss wt	*S. sonnei53G*	Bibliography[47]
Ss -p	*S. sonnei 53G plasmid cured*	This study
Ss -p Δ*tolR*	*S. sonnei 53G plasmid cured* Δ*tolR*	This study
Ss –p Δ*tolR* Δ*msbB*	*S. sonnei 53G plasmid cured* Δ*tolR* Δ*msbB*	This study
Ss -p Δ*tolR* Δ*ompA*	*S. sonnei 53G plasmid cured* Δ*tolR* Δ*ompA*	This study
Ss Δ*tolR* Δ*wbg*	*S. sonnei 53G* Δ*tolR* Δ*wbg*	Gerke, personal com.
Sf Δ*tolR*	*S. flexneri 2a str. 2457T* Δ*tolR*	Gerke, personal com.
Sf Δ*tolR* Δ*rfbG*	*S. flexneri 2a str. 2457T* Δ*tolR* Δ*rfbG*	Gerke, personal com.
Sf -p Δ*tolR* Δ*rfbG*	*S. flexneri 2a str. 2457T plasmid cured* Δ*tolR* Δ*rfbG*	Gerke, personal com.

Table 1. List of *Shigella* mutant strains.
Ss = *Shigella sonnei 53G*; Sf = *Shigella flexneri 2a str. 2457T*. Δ*tolR* = gene *tolR* replaced by a kanamycin resistance cassette, plasmid cured = virulence plasmid cured of the strain (pSS

for Ss and pINV for Sf), Δ*msbB* = gene *msbB* replaced by erythromycin resistance cassette, Δ*ompA* = *ompA* gene replaced by chloramphenicol resistance cassette, Δ*wbg*= O-antigen cluster (encoded on the virulence plasmid) replaced by chloramphenicol resistance cassette, Δ*rfbG* = *rfbG* gene replaced by erythromycin resistance cassette.

4.1.1. Primers used

The mutants were constructed using combinations of primers listed in the following table:

N°	PRIMER	SEQUENCE
	tolR.Kan.500-5	TCTGGAATCGAACTCTCTCG
	tolR.Kan.L-3	ATTTTGAGACACAACGTGGCTTTCATGGCTTACCCCTTGTTG
	tolR.Kan.L-5	TTCACGAGGCAGACCTCATAAACATCTGCGTTTCCCTTG
	tolR.Kan.500-3	TTGCTTCTGCTTTAACTCGG
	tolR.Kan.ext-5	AGCGGACCCGTATTCTTAAC
	tolR.Kan.ext-3	TTCGCTTTAGCATCTGCC
	Kan.int.For	TCGCGATAATGTCGGGCAATCAG
	Ampli.Kan-5	ATGAGCCATATTCAACGGGAAAC
	Ampli.Kan-3	TTAGAAAAACTCATCGAGCATCAAA
	pS.so53G.*ori*.F	CGTAACCGTAATTACAGCCG
	pS.so53G.*ori*.R	GATTTCACCTTACCCATCCC
	pS.so53G.*wzy*.F	CGTTGAGGTTTCACGTTTCT
	pS.so53G.*wzy*.R	TTACCAATATACCCTCCGCA
1	XbaI.*msbB*.5'.F	CTAGTCTAGAAGTGCTTTCAGTGGGTGACG
2	EcoRV.*msbB*.5'.R	AGCTTGATATCCCATGCTTTTCCAGTTTCGG
3	EcoRV.*msbB*.3'.F	AGCTTGATATCGGCGAAATCCAACCGTATAAG
4	XhoI.*msbB*.3'.R	CCGCTCGAGGGGGAAGTTGTTAAGACAGAC
5	*msbB*.Ery.ext-5	CGCTGAAGATGGCAAATTCTAC
6	*msbB*.Ery.ext-3	GTCGACCTTTCTGATCGAG

Table 2: Primers used to construct *Shigella* mutants.

4.1.2. Δ*tolR* mutant

In the three step PCR approach the upstream and downstream regions of the target gene (about 500 base pairs) are fused to the resistance marker gene. Firstly, the flanking regions of the *tolR* sequence were amplified from *Shigella sonnei 53G* genomic DNA using the following primer pairs: *tolR*.Kan.500-5/*tolR*.Kan.L-3 for the upstream region and *tolR*.Kan.L-5/*tolR*.Kan.500-3 for the downstream region (table 2). Secondly, the kanamycin resistance cassette was amplified from pUC4K[131] using the primers ampli.Kan-5/ampli.Kan-3. 100ng of the three PCR products were mixed together and submitted to a second round of PCR using the *tolR*.Kan.500-5/*tolR*.Kan.500-3 primers. The final product was a linear fragment in which the resistance marker cassette was flanked by upstream and downstream regions homologous to the *tolR* gene. 1µg of the linear fragment purified by QiaQuick PCR purification kit (Qiagen) was used to transform 50µL of lambda-red electrocompetent *Shigella sonnei* cells, carrying the temperature-sensible pAJD434 plasmid[96] to obtain the deletion mutant. The mutant colonies selected on LB + kanamycin plates at 37°C were screened by PCR using the primers *tolR*.Kan.ext-5/*tolR*.Kan.ext-3 and Kan.int.For/*tolR*.Kan.ext-3.

The same mutation was done in *Shigella flexneri* strain transferring the purified Δ*tolR* construct from *S. sonnei* in the electrocompetent *S. flexneri* cells carrying pAJD434.

Electrocompetent *Shigella* cells, carrying the pAJD434, were prepared as follows. The over night culture was diluted at OD_{600nm}= 0.05 in 150mL of LB + trimethoprim and let grow at 30°C, until OD_{600nm}= 0.2. 0.2% Arabinose was added to the culture to allow the expression of the recombinases encoded on pAJD434, and the culture was let grow until OD_{600nm} = 1.0. Bacteria were then recovered through centrifugation at 4000 rpm for 10min at 4°C, and washed 3 times in cold milliQ water. An additional washing step with cold 10% glycerol was performed and bacteria were finally re-suspended in 500µL of 10% glycerol.

4.1.3. Δ*msbB* mutant

Shigella sonnei 53G has 2 copies of the *msbB* gene. One copy is located on the virulence plasmid (*msbB2*) and the other (*msbB1*) on the chromosome. The gene *msbB2* was deleted by curing the plasmid from the *S. sonnei* Δ*tolR* strain. As *S. sonnei* is reported to naturally loose the plasmid during cultivation in vitro, we selected the *S. sonnei* Δ*tolR* strain lacking the plasmid by the white appearance on congo red agar plates. The loose of the plasmid was first confirmed by the changing of the colony phenotype from red to white observed after plating the bacteria on Congo-red plates. Also, the loss of the plasmid was confirmed by the absence of PCR products in the origin of replication (*ori*) and the *wzy* gene using primers pS.so53G.*ori*F/pS.so53G.*ori*R and pS.so53G.*wzy*F/pS.so53G.*wzy*R, respectively, in a selected white colony. The gene *msbB1* was knocked out by cloning. The upstream (496bp) and downstream (597bp) regions of *msbB1* gene were amplified by PCR using the XbaI.*msbB*.5'.F/EcoRV.*msbB*.5'.R and EcoRV.*msbB*.3'.F/XhoI.*msbB*.3'.R primers respectively. The 2 PCR products were digested with XbaI/EcoRV and EcoRV/XhoI respectively, and cloned into the pBluescript plasmid (Stratagen) vector. The ligation was performed according to a molar ratio of 1:6 between the plasmid and each fragment using the T4 DNA-Ligase (Roche). 7.5µL of the reaction were transformed in 50µL of Max Efficiency® DH5αTM-T1R Competent Cells (Invitrogen) according to the manufacturer specificities and the bacteria were selected using ampicillin (100µg/mL). The resistant colonies were screened by PCR using the primers XbaI.msbB.5'.F and XhoI.msbB.3'.R. The positive colony was grown in 3mL of LB + Amp, ON at 37°C to do a mini plasmid preparation (mini-prep Quiagen kit). The purified plasmid was linearized with EcoRV to insert the erythromycin resistance gene. The Ery[133] coding sequence was amplified with primers EcoRV.Ery.F/EcoRV.Ery.R. 69ng of the purified Ery cassette (ratio plasmid : construct = 1:6) and 30ng of the digested vector carrying the flanking regions of msbB1 were fused in a ligation reaction (as described before) to generate the pΔmsbBko::ery (figure 3). The plasmid was transformed in 50µL of Max Efficiency® DH5α^{TM}-T1^{R} Competent Cells (Invitrogen) and bacteria plated on LB + Ery (100µg/mL) plates. The resistant colonies were screened by PCR using the primers XbaI.*msbB*.5'.F/XhoI.*msbB*.3'.R to amplify the Δ*msbB*ko::ery fragment. 1µg of the

purified Δ*msbB*ko::ery fragment was used to transform *S. sonnei*, Δ*tolR,* plasmid-cured electro-competent cells, carrying the pAJD434, to generate the *msbB* knock out mutant. The mutant colony was subsequently controlled by PCR with different combinations of primers.

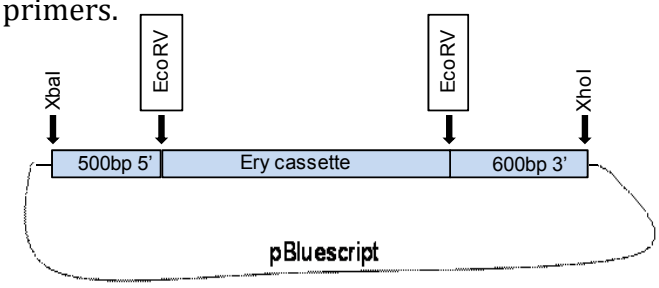


Figure 3. *msbB* mutant cloning.

Schema of the pBluescript (pBS) plasmid carrying the Δ*msbB* construct. The upstream (500bp at 5') and downstream (600bp at 3') regions of the *msbB* gene are ingegnerized in the pBluescript using XbaI/EcoRV and EcoRV/XhoI enzymes respectively. The resistance cassette (Ery) is subsequently cloned in the plasmid carrying the flanking regions using EcoRV enzyme.

4.1.4. Δ*ompA* mutant

Since the flanking regions of *ompA* are conserved between *E. coli* and *Shigella sonnei G53*, the genomic region including Δ*ompA*::cat and flanking regions was amplified from an *E. coli* Δ*ompA* mutant (kindly provided by Guido Grandi's group at Novartis Vaccine & Diagnostics, Siena) and transferred into Ss –p carrying the pAJD434 with lambda-red recombinase functions. In the resulting Ss –p Δ*ompA* strain the *tolR* gene was also deleted to obtain the Ss–p Δ*tolR* Δ*ompA* strain.

4.1.5. O antigen knock out mutants

To generate an O antigen knock out mutation in *S. flexneri* 2a the gene cluster encoding the biosynthesis of the repeating unit was inactivated by deletion of gene *rfbG* and parts of the genes *rfbF* and *rfc*. The resulting mutant was abbreviated as Sf Δ*rfbG* (C. Gerke, personal communication). In *S. sonnei* the O antigen knock out was obtained by complete deletion of the gene cluster encoding the biosynthesis of the repeating unit from gene *wbgZ* to *wzz*. This mutant has been abbreviated as Ss Δ*wbg* (C. Gerke, personal communication).

4.1.6. PCR conditions

All the PCR reactions were performed using the PCR Mix (Invitrogen) with 100 pmol of primers. The standard conditions for the amplification of the products were: denaturing step at 94°C for 30'', annealing at 52°C for 30'', and elongation at 72°C. Only for the amplification of the erythromycine resistance cassette a temperature of 54°C was used in the annealing step.

4.2. BACTERIAL STRAINS AND GROWTH CONDITIONS

E*. coli* strains were grown in LB medium at 37°C. *E. coli* and *Shigella* electrocompetent cells carrying the pAJD434 were grown at 30°C. *Shigella* strains were grown in LB for transformation and in yeast extract medium or defined medium for the production of GMMA. When specified, the mutants were grown in LB medium + 200μM Dipyridyl or at 30°C.

4.2.1. Yeast extract medium (HTMC)

S. sonnei and *S. flexneri* mutants were grown in yeast extract medium (HTMC), containing 200μM dipyridyl (or 100μM when specified), as follows: 30g/L yeast extract, 5g/L KH_2PO_4, 20g/L K_2HPO_4, 1.2g/L $MgSO_4*7H_2O$, 15g/L glycerol, and 0.25g/L PPG. Kanamycin (30μg/mL), chloramphenicol (20μg/mL), trimethoprim (100μg/mL), and erythromycin or ampicillin (100μg/mL) were added to the media as required. All the strains were grown at 37°C at 200rpm.

4.2.2. *Shigella sonnei* defined medium (DM)

Shigella sonnei mutants were also grown in the defined medium as follows: 30 g/L glycerol, 13.3g/L KH_2PO_4, 4g/L $(NH_4)_2HPO_4$, 1.2g/L $MgSO_4*7H_2O$, 1.7g/L citric acid, 2.5mg/L $CoCl_2*6H_2O$, 15mg/L $MnCl_2*4H_2O$, 1.5mg/L $CuCl_2*2H_2O$, 3mg/L H_3BO_3, 2.5mg/L $Na_2MoO4*2H_2O$, 13mg/L $Zn(CH_3COO)_2*2H_2O$, 2μM ferric citrate (unless specified differently in text), 50mg/L thiamine, 10mg/L nicotinic acid, 2.5g/L L-aspartic acid, 0.25g/L polypropylene glycol (PPG). The strains were grown at 37°C (or 30°C when specified). Kanamycin (30μg/mL), chloramphenicol (20μg/mL),

trimethoprim (100µg/mL), erythromycin or ampicillin (100µg/mL) were added to the media as required.

4.3. GMMA PURIFICATION

To produce GMMA, the strains were grown in 50mL of medium supplemented with antibiotics to OD_{600nm} = 0.3. The cultures were then diluted into 300mL of fresh medium to a starting OD_{600nm} of 0.05 and incubated over-night at 37°C or 30°C at 200rpm. All the cultures were stopped at about OD_{600nm} = 8.0 unless specified in the text. Bacteria were removed from the cultures by centrifugation for 10min at 4750rpm at 4°C and the supernatants, containing GMMA, were filtered through a 0.22µm filter unit (Millipore). The filtered samples were concentrated to 60mL in a Stirred Cell Model 8400 (Millipore) through a 100.000Da regenerated cellulose membrane (Millipore). GMMA were then separated from soluble proteins by ultra-centrifugation at 40000rpm for 2h at 4°C (Optima™ L-series, 45Ti rotor, Beckman Instruments). The pellets were washed once with 60mL of cold phosphate buffered saline (PBS) and finally re-suspended in PBS.

4.4. PROTEIN QUANTIFICATION

GMMA re-suspended in PBS were sterilized with 0.22µm filters and the protein concentration was determined by Bradford Assay (BioRad). 100µL of each sample were dilute 1:2 in 6M Guanidine-HCl pH 7.8 and boiled for 10min. The samples were subsequently diluted in the Bradford reagent according to the manufacturer instructions. A standard curve (range 1µg/mL - 9µg/mL in the assay) with BSA standard solution (Pierce) was used as reference to calculate the protein concentration of the samples according to their absorbance at 595nm.

4.5. NEGATIVE STAINING TRANSMISSION ELECTRON MICROSCOPY

GMMA were diluted to 100µg/mL in PBS. 5µL were placed on copper carbon-coated grids and GMMA were absorbed for 5min at room temperature. Grids were then washed with 3 consecutive drops of distilled water and blotted with a filter paper. For the negative staining the grids were treated with 2% uranyl acetate for

50seconds and air-dried. The grids were subsequently analyzed using a TEM FEI Tecnai G2 Spirit operating at 80kV equipped with a 2Kx2K CCD camera Olympus SIS Morada.

4.6. SDS-PAGE

GMMA were analyzed by SDS-polyacrylamide gel electrophoresis using an Xcell SureLock Mini Cell apparatus (Invitrogen). NuPAGER 12% Bis-Tris gels (10 wells, 1mm thick, Invitrogen) were used. All Blue (Invitrogen) was used as protein molecular mass marker, and each gel was run in 3-(N-morpholino)propanesulfonic acid (MOPS) running buffer (Invitrogen) at 30mA per gel.

4.6.1. Mono-dimensional: sample preparation and staining

10µg of GMMA were denatured for 10 min at 95°C in NuPAGE LDS sample buffer (Invitrogen) containing 10mM DL-Dithiothreitol (DTT, Sigma) and loaded into the gel. The gel was stained with Coomassie Brilliant Blue R-250 (40% methanol, 10% acetic acid, 1g/L R-250, Sigma), and distained with 10% acetic acid, 40% ethanol.

4.6.2. Bi-dimensional: sample preparation and staining

200µg of GMMA were separated by 2-dimensional (2-D) gel electrophoresis for Coomassie staining, while 20µg were used for Western Blotting. The proteins were precipitated with 1mL of cold trichloroacetic acid (TCA) with 0.4% DOC (deoxycholic acid) for 30min at 4°C. After centrifugation at 14000rpm for 20min at 4°C, the pellet was re-suspended in 1mL of cold 10% TCA in the ultrasonic bath, and then centrifuged as before. The pellet was washed 3 times with cold ethanol and finally let to dry in a SpeedVac Concentrator (RC1010 ThermoElectron). The proteins were resuspended in 125µL of reswelling buffer containing 7M urea, 2M thiourea, 2% CHAPS, 2% ABS 14, 0.1% DTT, 20mM Tris-base, 2mM TBP (tributylphosphine solution), 2% IPG buffer (GE Healthcare), and Bromo phenol blue. Then the proteins were absorbed overnight onto the Immobiline DryStrips (7-cm; pH gradient 3-11 non linear, GE Healthcare) using the Immobiline Dry-Strip Reswelling Tray (Amersham Biosciences). Isoelectric focusing was performed using an IPGphor IEF

Unit (Amersham Biosciences) setting no re-hydration and 15°C at 60μA/strip. Then were sequentially applied 150V for 95min, 500V for 35min, 1000V for 35min, 2600V for 35min, 3500V for 35min, 4200V for 35min, and finally 5000V to reach 10kV/h. For the second dimension, the IPG strips were equilibrated for 10min in a solution containing 4M urea, 2M thiourea, 2.5% (v/v) acrylamide, 30% (v/v) glycerol, 2% (v/v) SDS, 5mM TBP, 50mM Tris-HCl pH 8.8, and Bromo phenol blue. The IPG strips were then embedded into the gels with MOPS containing 0.5% agarose, 2mM TBP and Bromo phenol blue. Subsequently, the gels were washed in Milli-Q H2O for 10min, fixed for 1h in 200mL of 50% ethanol, 2% phosphoric acid solution, washed 3 times in ultra pure H2O for 30 min, and stained for 48h with colloidal Coomassie G-250 solution (34% methanol, 2% phosphoric acid, 17% ammonium sulphate, 0,65g/L Coomassie G-250). The gels were finally de-stained in Milli-Q H2O till the background was clear.

4.6.1. For LC-MS/MS analysis: sample preparation and staining

10μg of GMMA were mixed with NuPAGE LDS sample buffer (Invitrogen) containing 5mM Tris(2-carboxyethyl)phosphine hydrochloride solution (TCEP, Sigma), heated for 10min at 95°C. Samples were cooled down to room temperature (25°C) and iodoacetamide (IAA, Sigma) was added to final concentration at 10mM. The mixture was incubated for 30min at RT (in dark). The samples were loaded into the gel using SeeBlue Plus 2 Pre-Stained Standard (Invitrogen) as molecular weight marker and MOPS as running buffer. Constant voltage at 200V was applied for about 55min and current was initially at 100-115 mA/gel. The gel was fixed with 40% MeOH, 2% acetic acid for 30min stained overnight with Brillant Blue G-colloidal solution (Sigma), and destained for 1h or more with 30% MeOH until the background was cleared.

4.7. IMAGE ANALYSIS

2-DE gels were scanned with a Personal Densitometer SI (Amersham Biosciences) at 12bits and 50 μm per pixel. 1-D gels were scanned with Epson Perfection V700 PHOTO or Epson 1640XL scanner.

4.8. IN GEL PROTEIN DIGESTION

Proteins separated on 1-D or 2-D SDS-PAGE were excised from the gels and digested with trypsin for mass spectrometry analysis. Two different procedures were applied according to the instruments used for the analysis.

4.8.1. For Matrix-associated laser desorption/ionization time-of-flight (MALDI-TOF)

Proteins spots were excised from the gel, transferred into a polypropylene 96 well microplate (Nunc) and destained with 100µL of 50% 50mM ammonium bicarbonate, 50% acetonitrile at 4°C until required. After a quick spin, the supernatant was decanted from each well and the proteins spots washed with pure acetonitrile for 10min at room temperature. Dried spots were digested over night at 37°C in 10µL of 0.01µg/µL sequencing-grade modified trypsin (Promega), in 5mM ammonium bicarbonate. 0.6µL of the peptides from each sample were loaded into a Score384 target well containing α-cyano-4-hydroxycinnamid acid in 50% acetonitrile, 0.1% TFA, let dry and subsequently resolubilized with 1µL of 0.1% TFA. Mass spectra were acquired with Ultraflex TOF/TOF mass spectrometer (Bruker Daltonics).
Spectra were externally calibrated using a combination of standards pre-spotted on the target, and internally calibrated using enzyme autolysis picks. Mass spectra were analyzed with FlexAnalysis (version 2.4, Bruker Daltonics) and BioTool (version 3.0, Bruker Daltonics). Peptides Mass Fingerprinting (PMFs) were identified using the Mascot program (Mascot server version 2.2.01, Matrix Science) by searching a database containing protein sequences deduced from seven *Shigella* genomes, downloaded from the National Center for Biotechnology Information (NCBInr) or from Sanger Centre. The genomes used were: *Shigella sonnei* 53G, *Shigella flexneri* 2a str. 301, *Shigella flexneri* 2a str. 2457T, *Shigella sonnei* Ss046, *Shigella* boydii Sb227, *Shigella flexneri* 5 str. 8401, *Shigella boydii* CDC 3083-94. Monoisotopic masses were used to search the database, using the following parameters: cleavage by trypsin (cuts C-terminal side of Lysine and Arginine unless next residue is Proline), 1 miss cleavage for trypsin digestion, methionine oxidation as variable modification, and mass tolerance 200ppm.

4.8.2. For Liquid Chromatography Tandem Mass spectrometry (LC-MS/MS)

Each lane of the 1-D gel was excised in 15-16 slices; bands and blank regions were excised separately. Each slice was then cut into small pieces of 1-2 mm2 and transferred into a 96-well pierced plate (Proxeon) stacked on a normal plate. The gel pieces were distained with 150µL of 50% 50mM triethylammonium bicarbonate buffer (TEAB) pH 8.0, 50% acetonitrile for 30min at 37˚C with shaking at 600rpm. The liquid was removed by centrifugation of the Proxeon plate at 700rpm x 1min in a bench top centrifuge 5810R (Eppendorf). The distaining step was repeated until the gel pieces were completely white. The samples were incubated with 150µL of acetonitrile for 30min as before; the liquid was removed as described before, and the plate was air-dried in the laminar flow cabinet for 10min. 150µL of 0.001µg/µL of trypsin (sequencing-grade, Roche) in ice-cold TEAB pH 8.0 was added to each well and digested for 2h at 37°C, followed by 25°C overnight with shaking at 600rpm. For the peptide extraction a clean V bottom plate (NUNC) was stacked as a collection plate under the gel-containing plate. The supernatant, containing the peptides mixture, was collected by centrifugation as before. The gel-pieces were incubated with 80µL of 50% acetonitrile as above, then repeated twice with 100µL of 50% Acetonitrile, 50% of 0.5% Formic Acid and 100µL of Acetonitrile. All supernatant were collected; fractions were pooled by well and the collection plate was completely dry in the SpeedVac.

4.9. DIMETHYL LABELING

To compare the protein content of GMMA obtained from S. sonnei and S. flexneri strains (lacking the O-antigen) carrying or not carrying the virulence plasmid, GMMA were analyzed by LC-MS/MS after dimethyl –labeling based on a combination of Formaldehyde (CH2O) and Cyanoborohydride (NaBH3CN), with their isotopes CD2O, 13CD2O and NaBD3CN. The same approach was used to compare immunogenic proteins obtained in the immunoprecipitation experiments performed in different conditions.

The dried trypsin digested peptides were labeled with different combinations of isotopes. In the “LIGHT” labeling, Formaldehyde CH2O and Sodium

cyanoborohydride NaBH3CN were used; in the "INTERMEDIATE" labeling, Formaldehyde isotope CD2O and Sodium cyanoborohydride NaBH3CN; in the "HEAVY" labeling, Formaldehyde isotope 13CD2O and Sodium cyanoborodeuteride isotope NaBD3CN. The dried samples were re-suspended with 40µL of 100mM TEAB pH 8.0 and mixed for 5min at 750rpm at 20°C. 2µL of 4% (v/v) formaldehyde (CH2O, CD2O or 13CD2O) were added to the samples; after a brief mixing, 2µL of 0.6M Sodium cyanoborohydride (NaBH3CN) or Sodium cyanoborodeuteride (NaBD3CN) were added and the samples were incubated for 1h at RT at 600rpm. The labeling reaction was quenched by adding 1µL of 8% ammonia solution; after a brief mixing the solution was spun down by centrifugation at 600rpm for 5min. 4µL of 5% Formic acid were added to the samples (on ice) to further quench the reaction and to acidify the samples for the consecutive Liquid Chromatography-Mass Spectrometry (LC-MS) analysis. The differentially labeled samples were mixed together and transferred to glass tubes to be analyzed with LC-MS/MS.

4.10. LC/MS-MS

4.10.1. Data acquisition

The dimethyl labelled and mixed samples were analyzed with on-line nano LC-MS/MS on an Ultimate 3000 RSLCnano System (Dionex) coupled to a LTQ Orbitrap Velos (Thermo Fisher) hybrid mass spectrometer equipped with a nanospray source. Samples were first loaded and desalted on a PepMap C18 trap (0.3 mm id x 5 mm, 5µm, Dionex) then separated on a PepMap RSLC column with 2µm particle size (Dionex) at a 75 µm id x 50 cm column, for the initial analysis of GMMA samples, or a 25cm column, for the repeat analysis on the desalted GMMA samples and all IP samples, over a 60 or 90 or 120 or 180min linear gradient of 4–32% CH3CN/0.1% FA. The LTQ Orbitrap Velos mass spectrometer was operated in the "top 10" data-dependant acquisition mode. The 10 most abundant multiply-charged precursor ions in the MS survey scan in the Orbitrap (m/z 400 – 1500, with the lock mass at 445.120025), with a minimal signal above 2000 counts, were dynamically selected for CID fragmentation (MS/MS) in the LTQ Velos ion trap. The preview mode of FT master scan was disabled. The Orbitrap resolution was set at 30,000 at m/z 400 with two microscans (in the first GMMA analysis and all IPs), or 60,000 with one

microscan (in the 2nd GMMA analysis). Use of non-peptide monoisotopic recognition was enabled. The isolation width for the precursor ion was set at 2 Th. The normalized collision energy was set at 35% with activation Q at 0.250 and activation time for 10 msec. The dynamic exclusion mass width was set at ±20ppm and exclusion duration for 45 seconds. To achieve high mass accuracy, the AGC (Automatic Gain Control) were set at $1x10^6$ for the full MS survey in the Orbitrap, and 5000 for the MS/MS in the LTQ Velos, both with a maximum injection time at 200 msec.

4.10.2. Data analysis

The row files were analyzed by the quantitative proteomic MaxQuant software (version 1.1.1.36) for both protein identification and protein quantification. The Andromeda search engine was used to search the MS/MS spectra using the following parameters: trypsin/P with maximum 2 missed cleavages sites; peptide mass tolerance at first search was set at 20ppm; MS/MS fragment mass tolerance at 0.50 Da, and top 6 MS/MS peaks per 100 Da and a minimum peptide length of 6 amino acids were required. The mass accuracy of the precursor ions was improved by the time-dependent recalibration algorithm of MaxQuant. Fixed modification for Carbamidomethyl and variable modification for Deamidated (NQ) and Oxidation (M) were used, and maximum of three (for IP samples) or four (for GMMA samples) labelled amino acids per peptide were allowed. The protein databases were downloaded from UniProt (www.uniprot.org), and composed with four genomes (*Shigella sonnei*, *Shigella flexneri*, *Shigella dysentariae*, *Shigella boydii*) plus an in-house built plasmid database. Common contaminants (e.g. trypsin, bovine serum albumin and keratins) were supplemented by MaxQuant.

4.11. BIOINFORMATICS

The sequences of the identified proteins were analyzed by different bioinformatics tools. The prediction of the sub-cellular protein localization was carried out using PSORTb tool version 2.0 and version 3.0 (http://www.psort.org/psortb/), and LIPO (http://services.cbu.uib.no/tools/lipo) to look for lipoproteins. SMART

(http://smart.embl-heidelberg.de/) and PREDD-TMBB (http://biophysics.biol.uoa.gr/PRED-TMBB/input.jsp) were also used to explore the domain architectures of some of the identified proteins and to predict the presence of trans-membrane domains.

4.12. MOUSE IMMUNIZATION

To prepare antisera against GMMA, 5-6-weeks-old female CD1 mice were immunized with GMMA obtained from Ss –p Δ*tolR* and Ss –p Δ*tolR* Δ*msbB* mutants grown in low and high iron concentration in the medium (2μM and 200μM of ferric citrate). Six mice per group were used. Mice received two injections via the subcutaneous route on days 1 and 21. Each injection contained 2μg of GMMA formulated in PBS, and control mice received PBS alone. Blood samples for analysis were collected on day 0, 20 and 34. Both post-1 and post-2 sera were analyzed, but in this study we report only the data obtained using the final sera (5 weeks).

4.13. ENZYME-LINKED IMMUNOSORBENT ASSAY

To measure *Shigella sonnei* GMMA-specific Immunoglobulin (IgG) in mice serum, sera obtained from the immunization study were analyzed in ELISA. Nunc Maxisorb 96-well plates were coated over night at 4°C with a 0.5 μg/mL suspension of Ss -p Δ*tolR* GMMA diluted in PBS. The plates were washed 3 times with PBS containing 0.05% (v/v) Tween20 with a plate washer (ELx405TM, BioTek), and blocked with PBS containing 1% (w/v) BSA for 90min at 37°C. A serial dilution of Reference (1:1000, 1:2000, ..., 1:1024000) and sample (1:1000, 1:10000, 1:10000) sera were prepared in PBS containing 1% BSA, 0.05% Tween20 in a dilution plate. The ELISA plates were then washed three times as before and each serial dilution was transferred from the dilution plate to the coated plates. After 2h of incubation at 37°C, the plates were washed as described above, and incubated with goat anti-Mouse IgG-alkaline phosphatase, diluted 1:5000 in PBS with 1% BSA and 0.05% Tween20, for 2h at 37°C. The plates were washed as before, and incubated with a solution containing 4-Nitrophenyl phosphate disodium salt hexahydrate substrate (Sigma-Aldrich) dissolved in 1M Diethanolamine pH 9.8 (Sigma-Aldrich). After

60min the optical densities were measured with a plate reader (ELx800, BioTek) at 405 and 490nm wavelength. Absorbance at 490nm was subtracted from the absorbance at 405nm. Only results from sample serum dilutions 1:10000 and 1:100000 were considered. Results are expressed in optical densities related to a standard serum raised against different *S. sonnei* 53G GMMA. All samples were measured in duplicates.

4.14. WESTERN BLOTTING

The proteins resolved by 1-D and 2-D SDS-PAGE were transferred to nitrocellulose membranes using the iBlot® Gel Transfer Device (Invitrogen). Proteins blotted on the membranes were visualized with Ponceau S stain (BioRad) according to the manufacturer specifications. The membranes were then washed in PBS, blocked with 3% (w/v) milk in PBS, washed 3 times for 10min in 3% milk in PBST (0.05% (v/v) Tween20 in PBS), and incubated with mice polyclonal antisera (anti- Ss -p Δ*tolR*, low iron, 37°C; anti- Ss -p Δ*tolR*, high iron, 37°C; anti- Ss -p Δ*tolR* Δ*msbB* low iron, 30°C; anti- Ss -p Δ*tolR* Δ*msbB*, high iron, 30°C) diluted 1:1000 in 3% (w/v) milk in PBS for 2h at room temperature. The membranes were then washed as before, and incubated with sheep anti-mouse IgG secondary antibody conjugated to horseradish peroxidase (GE Healthcare, UK Limited) diluted 1:5000 in 3% (w/v) milk in PBS for 1h at room temperature. Immunoblots were washed three times and SuperSignal West Pico Chemiluminescent Substrate Kit (Pierce, Rockford, U.S.A.) was used for detection, as described by the manufacturer. Western Blotting and Ponceau images were acquired using ImageQuant400 (GE-Healthcare). The WB image was manually aligned with the Ponceau-stained proteins image using multiple hand-made dots on the membranes as references. The immunogenic proteins localized on the Ponceau S image were subsequently compared with the Coomassie image, and identified from the gel-spot.

4.15. IMMUNOPRECIPITATION

To identify immunogenic surface exposed proteins on *Shigella* bacteria, pre-immune serum and 2 batches of serum raised against Ss -p Δ*tolR*, low iron, 37°C GMMA were

used in an immunoprecipitation approach. Ss -p was grown in LB + 200μM Dipyridyl at 37°C until OD_{600nm}= 0.6 (3 x 10^8 bacteria/mL). 10mL of culture for each experiment were washed in 40mL of PBS and recovered by centrifugation at 4000rpm for 15min at 4°C. Bacteria were re-suspended in 400μL of PBS, and incubated with 10μL of GMMA-antiserum (unless specified in the text) for 1h at 37°C with shaking. The samples were washed 3 times with 1mL of PBS by centrifugation at 13000rpm for 2min at 4°C, and lysed for 1h at 4°C or 37°C with 1mL or 5mL of lysis buffer. Two different lysis buffers were tested. Solubilization buffer 1 (Bf 1) containing 10mM Tris-HCl pH 7.8, 150mM NaCl, 10mM EDTA, 1% Triton X-100 (v/v), 0.2% Deoxycholic acid (w/v), 0.1% SDS (v/v), and protease inhibitors (cOmplete, Mini Protease Inhibitor Cocktail Tablets, Roche); solubilization buffer 2 (Bf 2) containing 50mM Tris-HCl pH 8.0, 150mM NaCl, 10% Glycerol (v/v) 1% CHAPS (w/v), 50μg/mL DNasi, 50μg/mL RNase, 250μM $MgCl_2$, and protease inhibitors (cOmplete, Mini, EDTA-free Protease Inhibitor Cocktail Tablets, Roche). Soluble complexes were separated from cellular pellets by centrifugation at 4000rpm for 1h at 20°C. The supernatants were collected and incubated with 50μL of Protein-G MicroBeads (Miltenyi) overnight at 4°C with shaking. Immunopurification was performed through μColumns and μMACS™ Separator (Miltenyi) according to the manufacturer datasheet.

5. RESULTS

5.1. GENERATION OF GMMA-OVERPRODUCING STRAINS

In this study the strains *Shigella sonnei 53G* and *Shigella flexneri 2a 2457T* were chosen for generation of generalized modules for membrane antigens (GMMA). The mutation of the *tolR* gene, that is part of the Tol-Pal system, important for the stability and integrity of the outer membrane, has been shown to increase the release of GMMA from the surface of other Gram-negative bacteria into the culture medium[12,13].

In the current sub-chapter we describe the generation of *Shigella* mutants and the ability of the Δ*tolR* mutant in over-producing GMMA compared to the *wild type* strain.

5.1.1. *Shigella sonnei 53G* mutants

A null mutation in the *tolR* gene was introduced into *Shigella sonnei* by replacing the target gene with a kanamycin resistance cassette. For this mutation, a three step PCR approach was used as described in materials and methods. The mutant colonies were screened by PCR using primers (*tolR*.Kan.ext-5/*tolR*.Kan.ext-3) annealing outside the flanking regions. Using these primers, the PCR should amplify a fragment of 1500bp in the wild type and a fragment of about 2300bp in the Δ*tolR* mutant. As shown in figure 4, the tested colony was a Δ*tolR* mutant as a fragment of 2300bp was amplified.

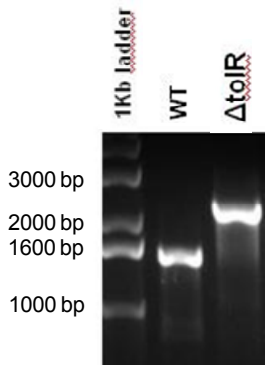


Figure 4. Verification of *tolR* deletion in *Shigella sonnei53G* Δ*tolR* mutants by PCR.
1 *S. sonnei* wild type (WT) and 1 *S. sonnei* Δ*tolR* mutant colonies were compared by PCR using the primers *tolR*.Kan.ext-5 and *tolR*.Kan.ext-3.

The mutant and the wild type strains were subsequently grown in LB at 37°C to compare the growth kinetics and GMMA release into the medium. No differences in

the bacterial growth were detected and SDS-PAGE of precipitated culture supernatants revealed a large amount of proteins in the Δ*tolR* sample (figure 5).

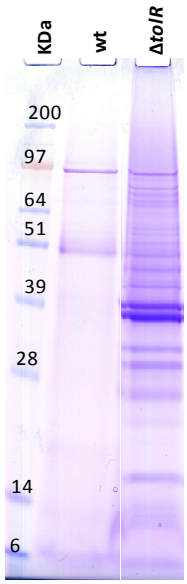


Figure 5. SDS-PAGE of supernatants from *S. sonnei* wt and *S. sonnei* Δ*tolR* strains.
12% acrylamide gel; See Blue Plus marker. Culture supernatants of *S. sonnei* wt and *S. sonnei* Δ*tolR* mutant were recovered from bacterial growth by low speed centrifugation. 30mL of each supernatant were TCA-precipitated and the protein content of each sample was measured by Bradford. 20µg of proteins of the Δ*tolR* mutant sample and the amount corresponding to the equivalent culture volume of the wt sample were loaded on the gel to compare the pattern of proteins present in the same conditions in the two strains.

As the O-antigen (OAg) is an immunodominant molecule on the surface of *Shigella* and is serotype-specific, we generated an *S. sonnei* Δ*tolR* strain lacking the O-antigen (ΔOAg), by curing the virulence plasmid (pSS) carrying the genes for the O-antigen biosynthesis[65]. The loss of the plasmid was confirmed by the lack of PCR amplification of the origin of replication (*ori*) and the *wzy* gene (encoded on the plasmid) in the Ss Δ*tolR* strain (figure 6).

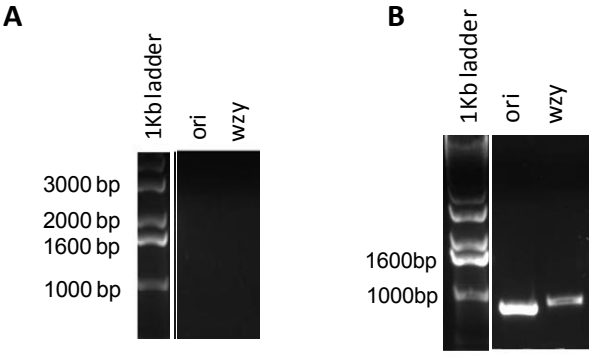
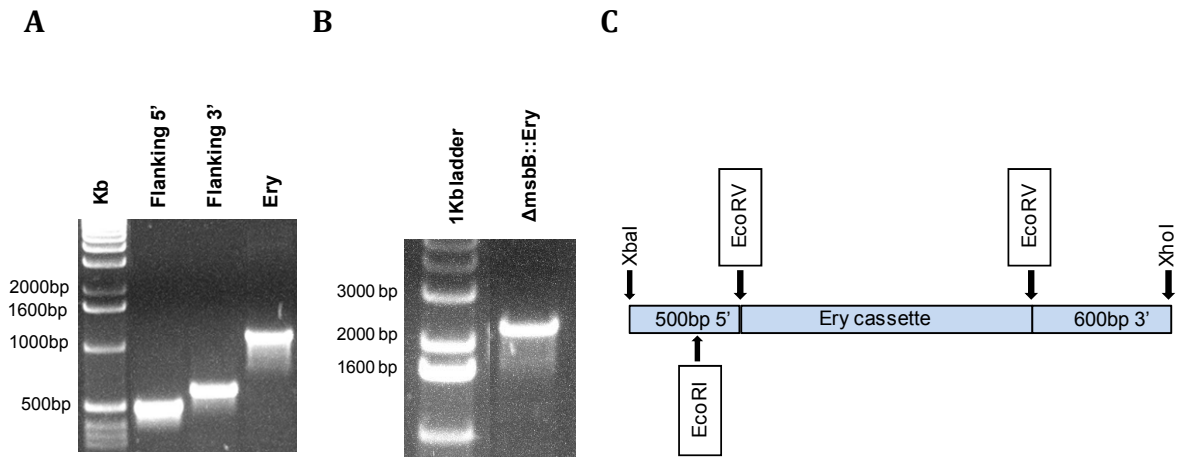
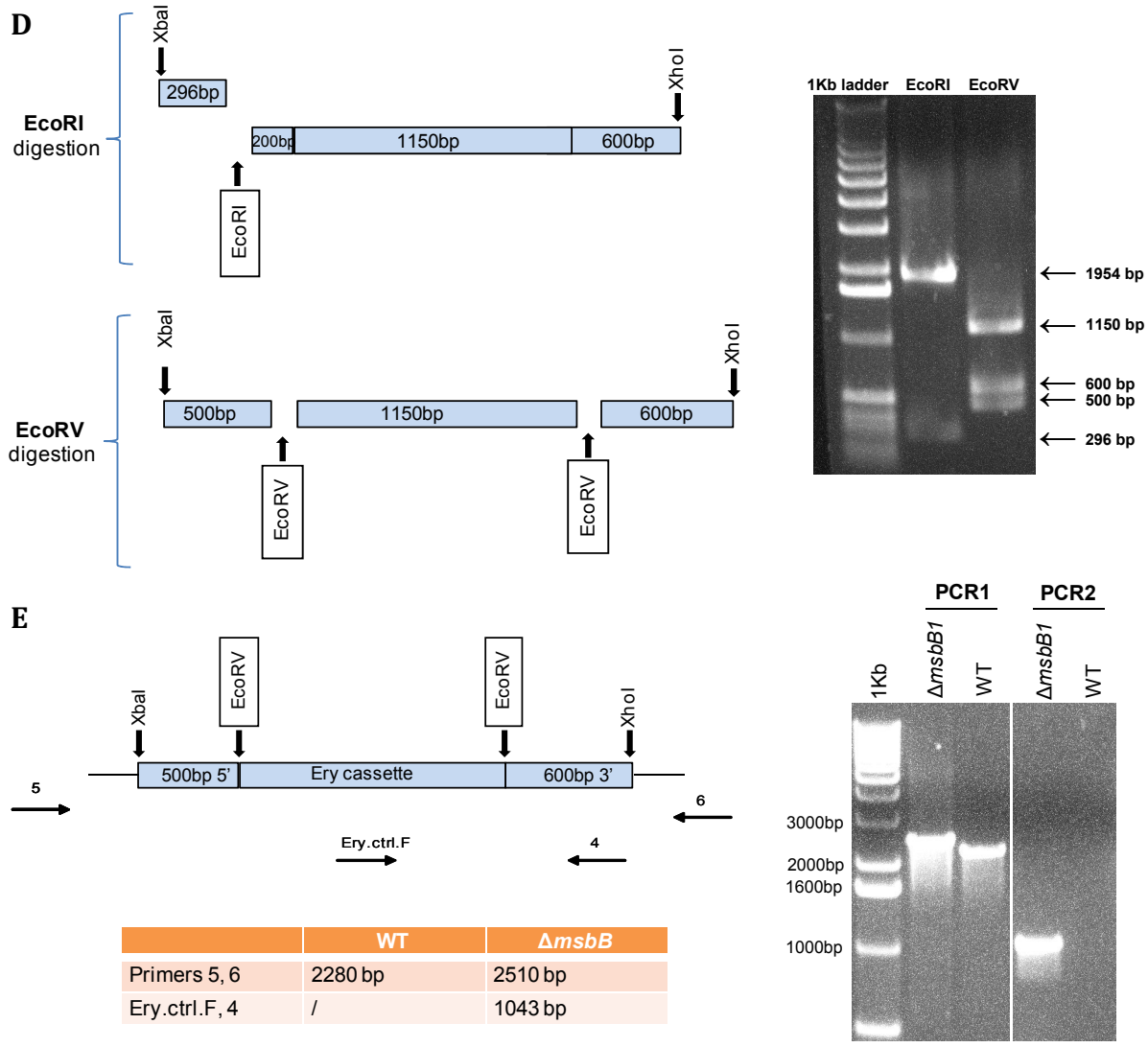


Figure 6. Verification of the absence of the virulence plasmid in the *Shigella sonnei 53G* Δ*tolR* mutant.
A. The *S. sonnei* Δ*tolR* mutant was screened by PCR using primers to amplify the origin of replication (*ori*) and the gene *wzy* carried by the plasmid. **B.** positive controls of *ori* and *wzy* PCR in the parent wild type strain.

A third genetic modification (Δ*msbB*) was subsequently introduced into the *S. sonnei* Δ*tolR*, -*pSS* strain to enhance vaccine safety of GMMA. The inactivation of the *msbB* gene results in a penta-acylated lipid A with reduced endotoxin activity as reported

in the literature[28]. *Shigella* carry two copies of the *msbB* gene (*msbB1* on the chromosome and *msbB2* on the plasmid). The gene *msbB1* was replaced by a null mutation and *msbB2* was lost by curing of the plasmid. In order to delete *msbB1* the *msbB1* flanking regions and the erythromycin cassette were amplified by PCR producing 3 fragments of 496bp, 596bp and 1150bp, respectively (figure 7A). The PCR products of the flanking regions were inserted into the pBS vector and electroporated into *E.coli* yielding plasmid 'pBS-flanking'. The plasmid was purified by mini-prep and the erythromycin cassette was inserted into the EcoRV site of purified 'pBS-flanking' between the 2 flanking regions. The resulting plasmid pΔ*msbB*ko::ery was purified and used as template to amplify by PCR the Δ*msbB*ko::ery fragment containing the erythromycin cassette plus the flanking regions (figure 7B). The purified product was controlled by enzymatic digestions with EcoRI and EcoRV (figure 7C-D). EcoRI produced one fragment of 296bp and another of 1954bp, as expected. EcoRV generated 3 fragments of the expected sizes as well: ~500bp, ~600bp, and 1150bp. The fragment Δ*msbB*ko::ery was subsequently electroporated into *S. sonnei* -pSS Δ*tolR* competent cells expressing the lambda red recombination system encoded on pAJD434. A mutant colony, selected on erythromycin, was controlled by PCR using different combinations of primers (figure 7E-F). All the combinations produced PCR fragments of the expected sizes confirming the *msbB1* knock-out. As *msbB2* gene is gone with the loss of the plasmid, the mutation of *msbB1* gene is indicated as Δ*msbB*.





	WT	Δ*msbB*
Primers 5, 6	2280 bp	2510 bp
Ery.ctrl.F, 4	/	1043 bp

Figure 7. Generation and control of Δ*msbB* mutant.
A. PCR products of *msbB* flanking regions (496bp at 5' and 596bp at 3') and *ery* cassette (1150bp). **B.** PCR of Δ*msbB*ko::ery construct (expected size ~2250bp) from pΔ*msbB*ko::ery. **C-D.** Schematic presentation of Δ*msbB*ko::ery construct and enzymatic digestions with EcoRI and EcoRV to control the construct. **E.** Schema of the primer combinations tested to control the *Shigella sonnei –pSS*, Δ*tolR,* Δ*msbB* colony. PCR1= primers external to the construct (5, 6). PCR 2= control primer internal to *ery* and primer annealing in the flanking region at 3' (Ery.ctrl.F, 4). **F.** PCR products of the Δ*msbB* mutant in comparison with the WT colony. The expected sizes of the wt and mutant are shown in the table.

The outer membrane protein A (OmpA) is known to be one of the most abundant and immunogenic proteins on the surface and in GMMA of many Gram-negative bacteria[100]. In order to be able to identify other proteins, including immunogenic proteins, with similar molecular mass as OmpA that might be masked by OmpA by SDS-PAGE, the *ompA* gene was deleted from Ss –pSS strain. Since the flanking regions

of *ompA* are conserved between *E.coli* and *Shigella sonnei*, the genomic region including Δ*ompA*::cat and flanking regions was amplified from *E.coli* Δ*ompA* mutant (kindly provided by Guido Grandi's group at Novartis Vaccine & Diagnostics, Siena) and transferred into Ss –pSS carrying the pAJD434 with lambda-red recombinase functions. Three mutant colonies were controlled by PCR using the same primers as used for amplification of the regions from *E. coli* and 2 colonies were positive for the Δ*ompA* mutation (figure 8A). Colony n° 3 was subsequently used to make the Δ*tolR* mutation in order to have the Δ*ompA* GMMA-over producing strain. The Ss –pSS Δ*tolR* Δ*ompA* mutant was finally controlled by PCR showing the right size of the Δ*tolR*::kan construct (figure 8B).

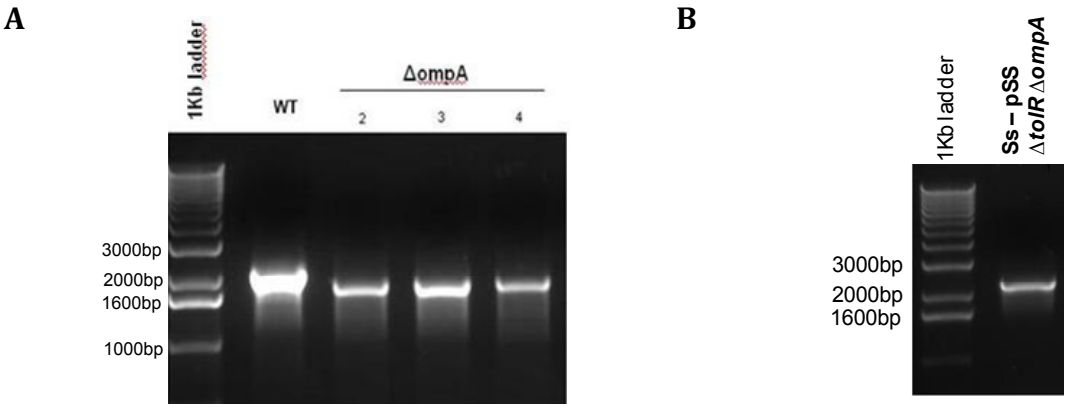


Figure 8. Controls of Ss –pSS Δ*ompA* Δ*tolR* mutant.
A. 3 colonies of Ss –pSS Δ*ompA* Δ*tolR*, were controlled by amplifying the Δ*ompA* construct in comparison to a wt colony. As the *ompA* gene is 200bp longer than the cloramphenicol cassette inserted in the ompA mutant, the slight difference in size of colony n° 4 likely represents a wild type colony. **B.** PCR to control the Δ*tolR* mutation in the colony n° 3.

5.2. PURIFICATION OF GMMA FROM CULTURE SUPERNATANT

In the current sub-chapter we verify the GMMA production from different mutants and compare the protein profiles of GMMA obtained in different conditions.

All GMMA over-producing *Shigella* strains were initially grown in LB and HTMC at 37°C. Subsequently a defined medium was optimized for the fermentation process of Ss –pSS Δ*tolR*. The *S. sonnei* strain Ss –pSS Δ*tolR* Δ*msbB* modified to produce LPS with reduced endotoxicity did only grow to low optical densities (OD_{600nm} 0.5) in the

defined medium at 37°C. At 30°C, the strain grew to high ODs in the defined medium and was thus cultivated at 30°C for the production of GMMA. To produce GMMA all the strains were grown in flasks to an OD_{600nm} of approximately 5-8. Bacteria were removed from the liquid culture by low speed centrifugation and the supernatant containing the GMMA was subsequently filtered though a 0.22µm filter. GMMA were separated from soluble proteins by ultra-centrifugation, washed, and re-suspended in PBS.

As Ss –pSS Δ*tolR and* Ss –pSS Δ*tolR* Δ*msbB* were grown at a different temperature, we were interested in understanding if the different temperature could have altered the protein profile in Ss –pSS Δ*tolR* Δ*msbB* GMMA compared to Ss –pSS Δ*tolR* GMMA. This could present a critical issue for to the production process of a GMMA-based vaccine. If the temperature caused a different expression of the proteins in Ss –pSS Δ*tolR* Δ*msbB* GMMA compared to Ss –pSS Δ*tolR* GMMA, it would be necessary to optimize the defined medium for the *msbB* mutant at 37°C (physiological temperature during the infection process). Therefore, the Ss –pSS Δ*tolR* strain was grown both at 30°C and 37°C, and purified GMMA were compared by mono-dimensional SDS-PAGE. The gel presented in figure 9 shows that the main protein profile was conserved at both temperatures, and only small differences can be detected. In particular, only one visible protein with a molecular weight of roughly 20KDa was less abundant at 30°C. A group of proteins with a molecular weight of approximately 50KDa and other two proteins were, instead, more expressed at 30°C suggesting that the decrease of the temperature from 37°C to 30°C does not cause a significant loss of expressed proteins.

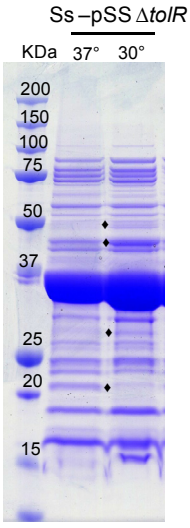


Figure 9 SDS-PAGE of *Shigella sonnei* –pSS Δ*tolR* GMMA obtained at 30°C and 37°C.
10µg of protein of each GMMA preparation were loaded on a 12% acrylamide gel. Few differences are detected in the proteins expression of GMMA obtained at 30°C and 37°C. '♦' indicate proteins that show a temperature-regulated expression. The group of proteins with a molecular weight of 50KDa and other 2 proteins seem to be more expressed at 30°C than at 37°C. On the contrary, 1 protein (~20KDa) is over-expressed at 37°C.

Iron uptake systems are expressed by bacteria in the body where iron limitation is present, and are important for the survival in the host. Since iron siderophore receptors have been shown to be protective antigens against *Salmonella* infection in chickens[72], we hypothesized that they could also be protective against other bacteria and we were interested in including these proteins in GMMA. Different iron concentrations (0.2μM, 2μM, 100μM, and 200μM ferric citrate) in the growth medium were tested to define a concentration that induces the expression of iron-regulated proteins. A concentration of 2μM ferric citrate in the DM was optimal to allow a normal growth of the bacteria and the over-expression of three iron-regulated proteins visible in figure 10. A high concentration of iron (200μM) was, instead, sufficient to reduce their expression. GMMA were purified from Ss –pSS Δ*tolR* and Ss –pSS Δ*tolR* Δ*msb*B strains grown with 2μM and 200μM ferric citrate at 37°C and 30°C, respectively, and compared by SDS-PAGE. As shown in figure 10, GMMA contained many proteins ranging from 15KDa to more than 150KDa. However, there might be proteins with higher molecular weights that are not visible in the gel conditions we used. Two proteins, overlapping in a single band (roughly 36KDa) in the gel, were identified by mass fingerprint as outer membrane protein A (OmpA) and outer membrane protein C (OmpC). OmpA and OmpC are known to be among the most abundant proteins in the outer membrane of Gram-negative bacteria[77,98], and they represented the 40%-50% of the total GMMA protein content in these mutants (data not shown). The protein profiles of Ss –pSS Δ*tolR* and Ss –pSS Δ*tolR* Δ*msb*B GMMA showed specific differences. Three proteins identified by mass fingerprint as FepA (Ferrienterobactin receptor), FhuA (Ferrichrome-iron receptor), and CirA (Colicin I receptor) were up-regulated in both of the strains when the iron concentration was low (2μM). Among these three proteins, CirA seemed to be less expressed in the Δ*msbB* strain than in the Ss –pSS Δ*tolR* strain. In contrast, 5 proteins highlighted with '♦' appeared to be more abundant in the Δ*msbB* mutant. These proteins were not identified. The temperature-regulation of the group of proteins with a molecular weight of approximately 50KDa, already observed in the comparison 30°C/37°C of Ss –pSS Δ*tolR* GMMA, was confirmed also in the Ss –pSS Δ*tolR* Δ*msbB* mutant. The 2 proteins with a molecular weight higher than 150KDa are difficultly visible in Ss –pSS Δ*tolR*, thus, their over-expression in Ss –pSS Δ*tolR* Δ*msbB* GMMA likely result from the *msbB* mutation.

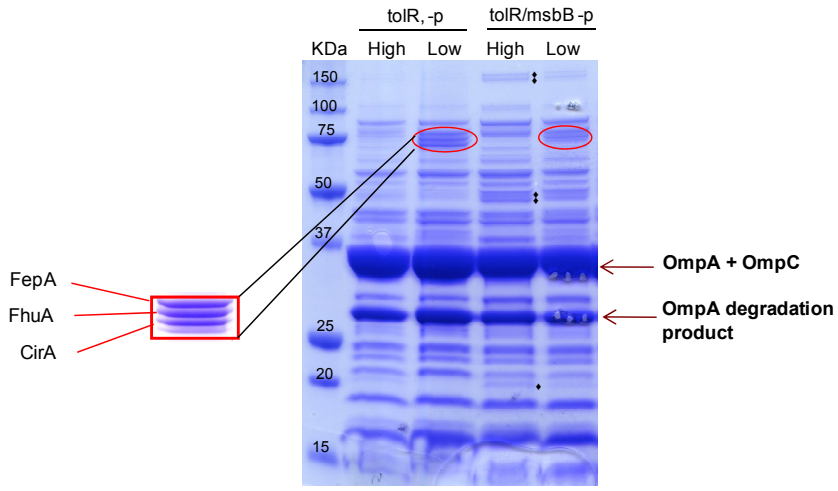


Figure 10. SDS-PAGE of *Shigella sonnei* GMMA.
10µg of protein of each GMMA preparation were loaded on a 12% acrylamide gel. 'Low' and 'high' refer to the iron concentration in the medium (2µM and 200µM*)*. '-p' indicates the plasmid cured strain. The red circles indicate the three iron-regulated proteins. '♦' indicate proteins that are over-expressed in the Δ*tolR* Δ*msbB* mutant compared to the Δ*tolR* strain. OmpA and OmpC represent the major outer membrane proteins in GMMA.

Purified GMMA from different mutants were analyzed by transmission electron microscopy, revealing the presence of round particles with a diameter ranging from 50m to 80nm (figure 11).

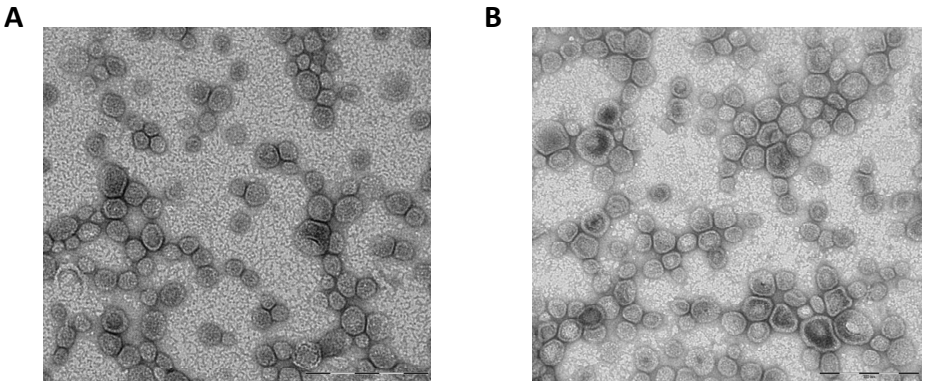


Figure 11. Electron micrographs of *S. sonnei* and *S. flexneri* GMMA.
Representative pictures of *S. sonnei* and *S. flexneri* GMMA. **A.** GMMA from Ss +pSS Δ*tolR* grown in defined medium with 2µM of ferric citrate. **B.** GMMA from Sf -pINV Δ*tolR* Δ*rfbG* (ΔOAg) GMMA grown in HTMC with 200µM dipyridyl. Bar length= 200nm.

5.3. CHARACTERIZATION OF GMMA PROTEIN CONTENT

In the current sub-chapter we describe the protein content of *S. sonnei* and *S. flexneri* GMMA using two different proteomic approaches: MALDI-TOF MS and LC-MS/MS. Subsequently we characterize the proteins conserved in *S. sonnei* and *S. flexneri* GMMA and perform a quantitative analysis to estimate the abundance of these proteins in GMMA.

5.3.1. Characterization of the protein composition of GMMA by bi-dimensional mapping and MALDI-TOF MS analysis

GMMA purified from *S. sonnei* and *S. flexneri* were analyzed by bi-dimensional SDS-PAGE coupled with Matrix-Assisted Laser Desorption-Ionization Time-of-Flight (MALDI-TOF) mass spectrometry analysis, to determine and characterize the protein profiles of GMMA. These analyses were performed in collaboration with the Proteomic Team headed by Nathalie Norais at Novartis Vaccines and Diagnostics in Siena.

5.3.1.1. Shigella sonnei GMMA

A bi-dimensional electrophoresis approach was used to better visualize the protein composition of GMMA. In the 2-D SDS-PAGE, the proteins were separated in two subsequent dimensions: a non linear pH 3-11 gradient in the first dimension, and a 12% acrylamide gel in the second dimension. With this method, each protein is separated in the first dimension according to its isoelectric point, and subsequently in the second dimension according to its molecular weight. The resulting gel contained hundreds of spots, visible after Coomassie Blue staining, each of which represents a single protein that can be excised from the gel and analyzed by mass fingerprint. As shown in figure 10, OmpA and OmpC are major components of the GMMA proteome and overlap in a single band in 1-D SDS-PAGE. Bi-dimensional electrophoresis performed on *E. coli* GMMA[100] showed that OmpA and OmpC are well separated in the 2-D gel. Further, OmpA rather than OmpC can be detected in multiple locations of the gel indicating the presence of isoforms and fragments migrating at different isoelectric point and mass values likely representing

degradation products and modifications of the OmpA protein. As outlined earlier, OmpA could mask other proteins with similar molecular weight limiting the number of detectable proteins, and thus, an *ompA* null mutation was introduced into Ss –pSS. Then, the *tolR* gene was deleted from the resulting Ss –pSS Δ*ompA* strain in order to induce GMMA over-production. GMMA were purified from the Δ*ompA* mutant grown in HTMC under iron-limited conditions (using 100μM of the iron chelator dypirydil), and the protein composition was analyzed by 2-D SDS-PAGE in combination with mass spectrometry analysis. Approximately 120 spots were picked from the gel (figure 12) and analyzed by MALDI-TOF. A total of 78 proteins were identified by mass fingerprint, and the protein sequences were analyzed using PSORTb 2.0 to predict their sub-cellular localization. 55 proteins were predicted to be outer membrane proteins (18), outer membrane lipoproteins (16) and periplasmic proteins (21). A small fraction of cytoplasmic proteins and proteins classified as 'unknown', as no prediction was available, was also identified. The three previously found iron-regulated proteins FepA, FhuA and CirA were also identified by this approach.

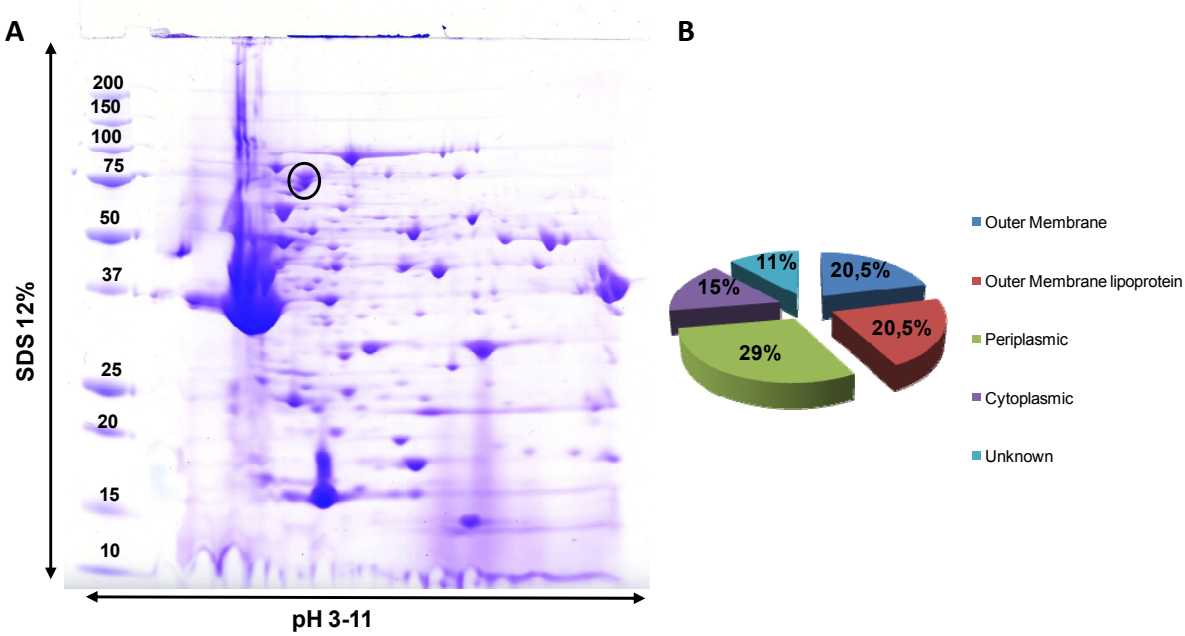


Figure 12. Bi-dimensional SDS-PAGE of Ss –pSS Δ*tolR* Δ*ompA* GMMA.
A. 200µg of protein were separated on a non linear pH 3-11 gradient in the first dimension, then on a 12% acrylamide gel in the second dimension. The circle highlights the iron-regulated proteins FepA, FhuA and CirA. **B.** Subcellular localization of GMMA proteins according to PSORTb 2.0 prediction.

A bi-dimensional electrophoresis approach was subsequently used to better separate and analyze the protein composition of GMMA from Ss –pSS Δ*tolR* and Ss –pSS Δ*tolR* Δ*msbB* with detoxified LPS. The aim was to investigate if the Δ*msbB* mutation affected the protein content and to identify differentially expressed proteins. Ss –pSS Δ*tolR* and Ss –pSS Δ*tolR* Δ*msbB* GMMA showed clear differences in protein profile (figure 13). In particular, 10 spots showed a higher expression in Δ*tolR* than in Δ*tolR* Δ*msbB* GMMA (A-E, G, L-O, S). For example, the iron-regulated proteins FepA, and CirA (H-I) were confirmed to have a lower expression in the detoxified GMMA than in the Ss –pSS Δ*tolR* as previously observed in the 1D SDS-PAGE. In contrast, the protein spots F and R (that was also picked in the Ss –pSS Δ*tolR* and indicated as P) were more expressed in GMMA from the detoxified strain than from Ss –pSS Δ*tolR*. The protein with a molecular weight higher than 150KDa previously observed in the Ss –pSS Δ*tolR* Δ*msbB* GMMA by 1-D gel was not visible in this gel. This might be related to the technical limitation of 2-D gel in performing the first dimension of proteins with a high molecular weight and particularly hydrophobic, as also described in the literature[24]. The proteins that were

differentially expressed in the two samples are listed in table 3. The proteins OmpA and OmpC were identified covering the central area of the gel.

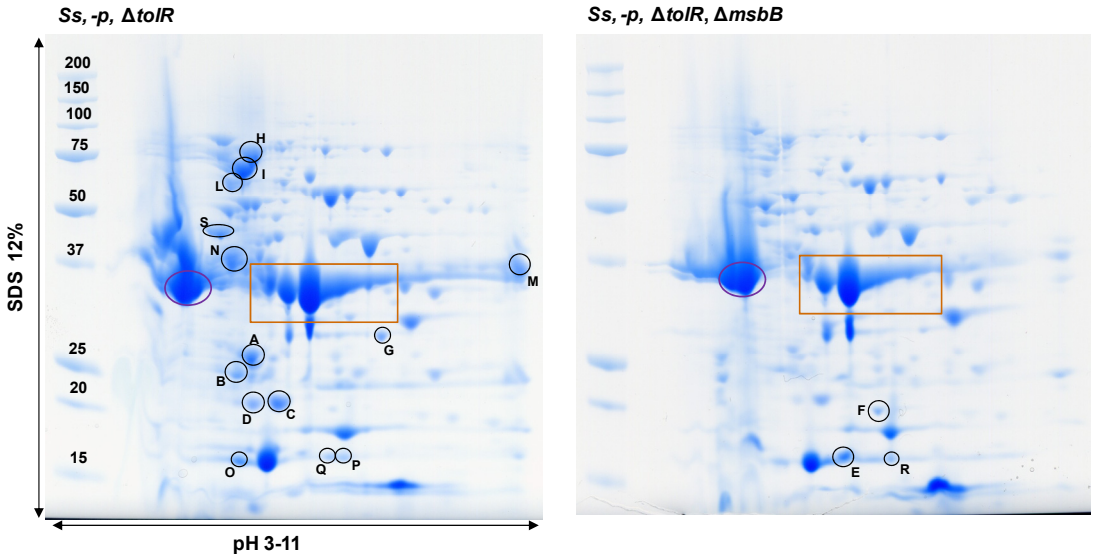


Figure 13. Bi-dimensional SDS-PAGE of Ss –pSS Δ*tolR* and Ss –pSS Δ*tolR* Δ*msbB* GMMA. 200µg of protein were separated in the first dimension on a non linear pH 3-11 gradient and subsequently by SDS-PAGE on a 12% acrylamide gel in the second dimension. Single spots were picked from the Coomassie blue stained gel and proteins were identified by mass fingerprint. The molecular weights (KDa) of the standard are indicated in the gel on the left. Left panel: GMMA from Ss -pSS ΔtolR grown in SSDM with low iron at 37°C, right panel: GMMA from Ss -pSS ΔtolR ΔmsbB grown in SSDM low iron at 30°C. Violet circle indicates OmpC, orange box: OmpA.

Proteins identified by MALDI-TOF MS in Ss-p Δ*tolR* and Ss-p Δ*tolR* Δ*msbB* GMMA (37°C vs 30°C)				
Spot	**Protein name**	**gi**	**gene**	**PSORTb 2.0**
A	Histidine binding periplasmic protein of high-affinity histidine transport system	74312826	hisJ	Peri
B	N/A			
C,D	Osmotically inducibile protein Y	187733898	osmY	Peri
E	Putative lipoprotein	82777619	ybjP	Ukn
F	Outer membrane protein W	74312394	ompX	OM
G	Serine protease	24114232	htrA	OM
H	Outer membrane receptor FepA	74311118	fepA	OM
I	Colicin I receptor	74312677	cirA	OM
L	Chaperonin GroEL	24115498	groEL	cyto
M	Putative receptor Serine protease	74312191 24114232	SSON_1681 htrA	Ukn Peri
N	Hypothetical protein SBO_2040	82544504	ycdO	Peri
O-R	Outer membrane protein X	56479734	ompX	OM lipo
S	Maltoporin	56480532	lamB	OM

Table 3. Differentially expressed proteins in GMMA purified from Ss –pSS Δ*tolR* and Ss –pSS Δ*tolR* Δ*msbB* with detoxified LPS.
The letters in the first column refer to the proteins highlighted in figure 13. The protein name, the accession number (gi), and the localization predicted by PSORTb 2.0 are shown in the following columns.

5.3.1.2. Shigella flexneri GMMA

To characterize the protein composition of *S. flexneri* GMMA, an *S. flexneri* strain overproducing GMMA (Δ*tolR*) and lacking the O-antigen (Δ*rfbG*) was generated. GMMA obtained in complex medium (HTMC) under iron limiting conditions (100µM Dipyridyl) were analyzed by 2-D SDS-PAGE to identify their protein content. 156 spots were excised from the gel (figure 14), and 118 were successfully identified by mass fingerprint corresponding to 65 unique proteins. The protein sequences were subsequently processed by PSORTb 2.0. 40% (26) of the total identified proteins were predicted to be outer membrane and periplasmic proteins. 7% (5) were predicted to be inner membrane proteins, and another 7% (5) were classified as 'unknown'. 38% were predicted to be cytoplasmic. The significant difference with the cytoplasmic fraction observed in *S. sonnei* GMMA could potentially be a result of a higher level of lysis in the HMTC medium at high optical densities compared to the defined medium used for *S. sonnei*, or to the fact that the cytoplasmic proteins are much easier to be identified in a small number of identified spots.

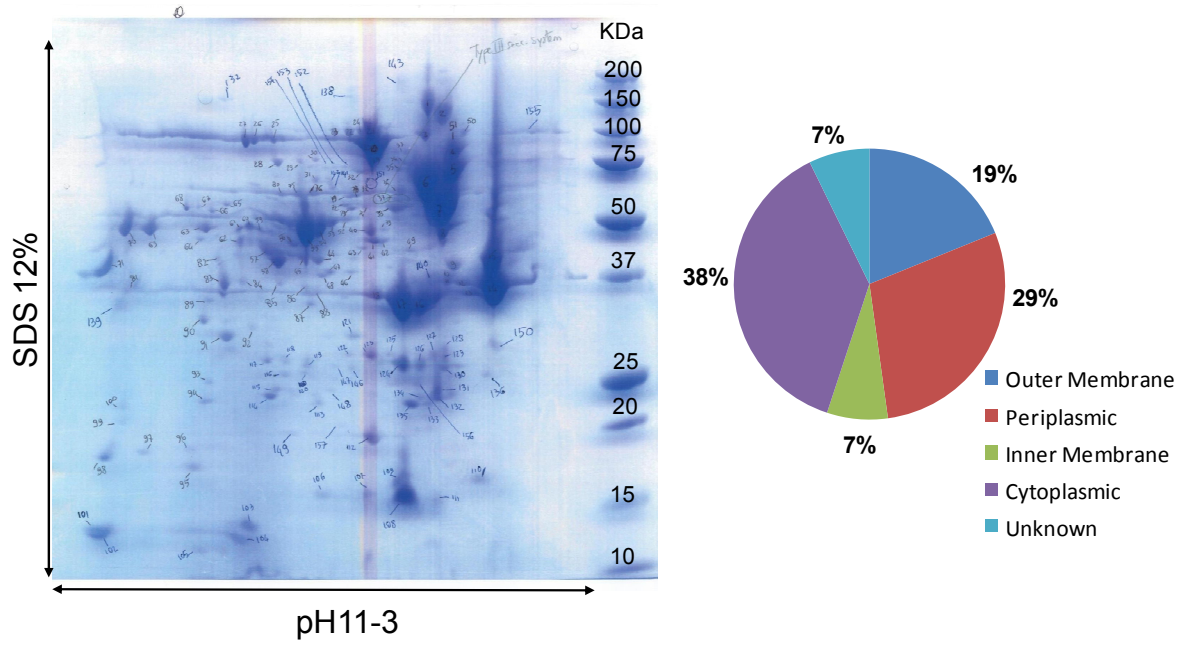


Figure 14. 2-D SDS-PAGE of Sf +pINV Δ*tolR* Δ*rfbG* GMMA.
A. 200µg of protein were separated on a non linear pH 3-11 gradient in the first dimension, then on a 12% acrylamide gel in the second dimension. GMMA were obtained from Sf Δ*tolR* Δ*rfbG* strain with the virulence plasmid (+pINV). Each excised spot is labeled. **B.** Subcellular localization of GMMA proteins according to PSORTb 2.0 prediction.

5.3.1.3. Conserved proteins in S. sonnei and S. flexneri GMMA

In order to use GMMA as a protein-based vaccine against multiple *Shigella* strains, we were interested in identifying conserved proteins expressed in both *S. sonnei* and *S. flexneri* GMMA. Although we expected to identify a high percentage of common proteins, in particular outer membrane proteins, when we compared the two lists of *S. sonnei* and *S. flexneri* proteins identified by 2-D SDS-PAGE we only identified 40% of shared proteins. However, this percentage might be bigger considering that only 78 and 65 proteins were identified respectively in *S. sonnei* and *S. flexneri* GMMA, and of these 28 were identified in both (table 4). Several proteomics studies have shown that the solubility of membrane proteins in 2-D SDS-PAGE buffer is limited, and their abundance is often too low to be detected or reproducibly observed[122].

GI	Gene	Protein Definition	PSORTb 2.0
24114232	sigA	serine protease	Ext/OM
74311859	prc	Carboxy-terminal protease	OM
24112822	yeaF	hypothetical protein SF1441	OM
56480244	tolC	outer membrane channel protein	OM
74312736	ompC	outer membrane porin protein C	OM
74311514	ompA	outer membrane protein A	OM
56479734	ompX	outer membrane protein X	OM
24112608	lolB	outer membrane lipoprotein LolB	OM
24111612	yaeT	outer membrane protein assembly factor YaeT	OM
56479690	pal	peptidoglycan-associated outer membrane lipoprotein	OM lipo
30063856	nlpB	lipoprotein	OM lipo
74311310	ybhC	putative pectinesterase	OM lipo
24114611	fkpA	FKBP-type peptidyl-prolyl cis-trans isomerase	Peri
24111599	htrA	serine endoprotease	Peri
30062097	tolB	translocation protein TolB	Peri
24111968	modA	molybdate transporter periplasmic protein	Peri
24114628	ppiA	peptidyl-prolyl cis-trans isomerase A (rotamase A)	Peri
24111499	surA	peptidyl-prolyl cis-trans isomerase SurA	Peri
74311061	ushA	bifunctional UDP-sugar hydrolase/5'-nucleotidase periplasmic precursor	Peri
110805056	mdoG	glucan biosynthesis protein G	Peri
74312961	cysP	thiosulfate transporter subunit	Peri
24114441	yraP	hypothetical protein SF3191	Peri
110806822	yggE	hypothetical protein SFV_2968	Peri
187734005	bglX	beta-glucosidase, periplasmic	Peri

56479605	lpdA	dihydrolipoamide dehydrogenase	Cyto
74310732	aceE	pyruvate dehydrogenase subunit E1	Cyto
24115158	glnA	glutamine synthetase	Cyto
24115498	groEL	chaperonin GroEL	Cyto

Table 4. 28 common proteins identified in *S. sonnei* and *S. flexneri* GMMA by 2-D SDS-PAGE and MALDI-TOF MS.
OM: outer membrane proteins, OM lipo: outer membrane lipoproteins, Peri: periplasmic proteins, Cyto: cytoplasmic proteins.

5.3.2. Qualitative analysis of GMMA using stable isotope dimethyl labeling in combination with LC/MS-MS

To identify a larger number of proteins and to obtain a more extensive analysis of proteome of *Shigella* GMMA, we focused on an alternative proteomic approach based on GeLC-MS/MS analysis. This approach, when combined to stable isotope dimethyl labeling, allows a simultaneous analysis of 2-3 different samples and the subsequent relative quantitation of the proteins measuring the changes in their expression levels in different samples[15]. In the first analysis we used the stable isotope dimethyl labeling to compare the GMMA protein content obtained from different mutants. In this approach, proteins are separated by mono-dimensional SDS-PAGE, each lane of the gel is cut into slices, and each slice is digested. The peptides extracted from each slice are labeled with a different combination of formaldehyde and cyanoborohydrade isotopes with heavy, medium or light for corresponding to the different protein samples. Subsequently, the labeled samples are mixed and simultaneously analyzed by GeLC-MS/MS. Given that the mass spectrometer can recognize the mass difference among the differentially labeled forms of a peptide (table 5) it is possible to obtain a quantification of the peptides in each combined sample comparing their respective signal intensities. It should be noted that in this approach different proteins with similar molecular weights are digested and processed together.

Label	**Light (L)**	**Intermediate (M)**	**Heavy (H)**
Mass increase per label	+28.0313 Da	+32.0564 Da	+36.0757 Da
Formaldehyde isotope	CH_2O	CD_2O	$^{13}CD_2O$
Cyanoborohydride isotope	$NaBH_3CN$	$NaBH_3CN$	$NaBD_3CN$

Table 5. Differential stable isotope dimethyl labeling and resulting mass shifts

To analyze if proteins encoded on the virulence plasmid are present in GMMA, we compared OAg-deficient GMMA obtained from the Ss –pSS Δ*tolR* strain with GMMA from an OAg-deficient mutant carrying the virulence plasmid but lacking the OAg biosynthesis genes Ss +pSS Δ*tolR* Δ*wbg*. Proteins from the 2 samples were differentially labeled and mixed for the comparison. For *S. flexneri* we compared 2 OAg-deficient GMMA preparations obtained from Sf –pINV Δ*tolR* Δ*rfbG* and Sf +pINV Δ*tolR* Δ*rfbG* strains. The 4 OAg-deficient mutants used for the production of GMMA and the growth conditions are shown in the following table:

Mutants	Final OD600nm	Growth conditions
Ss, -pSS, Δ*tolR* (ΔOAg)	1.5	HTMC, 200µM dipyridyl, 37°C
Ss, +pSS, Δ*tolR*, Δwbg(ΔOAg)	1.3	
Sf, -pINV, Δ*tolR*, ΔrfbG (ΔOAg)	0.99	
Sf, +pINV, Δ*tolR*, ΔrfbG (ΔOAg)	0.97	

Table 6: Strains and growth conditions for production of GMMA analyzed by LC-MS/MS

GMMA from the 4 mutants were separated by 1D SDS-PAGE (figure 15). Each lane of the gel (1a, 1b, 1c, 1d) was subsequently cut into 15 or 16 slices, and each slice was digested with trypsin. The peptides extracted from each slice were labeled according to table 5:

- Ss –pSS Δ*tolR* and Sf -pINV Δ*tolR* Δ*rfbG*: intermediate labeling (M)
- Ss +pSS Δ*tolR*, Δ*wbg* and Sf +pINV Δ*tolR* Δ*rfbG*: heavy labeling (H)

The 'intermediate' and 'heavy' labeled peptides from equivalent slices with the same number of *S. sonnei* and *S. flexneri* GMMA, respectively, were then combined generating 16 labeled *S. sonnei* GMMA samples (sample Ss1) and 15 labeled *S. flexneri* GMMA samples (sample Sf1). Ss1 and Sf1 were subsequently analyzed by HPLC-MS/MS.

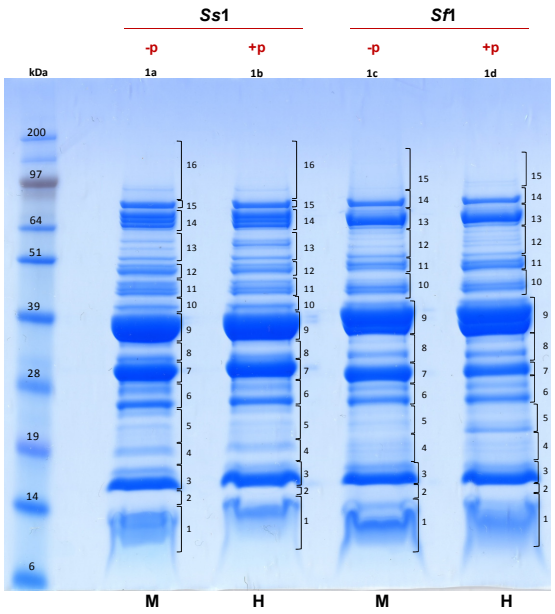


Figure 15. SDS-PAGE of *S. sonnei* (Ss) and *S. flexneri* (Sf) GMMA obtained from O antigen deficient strains.
SDS-PAGE, 12% acrylamide gel. 1a = Ss –pSS Δ*tolR*. 1b= Ss +pSS Δ*tolR* Δ*wbg*. 1c = Sf –INV Δ*tolR* Δ*rfbG*. 1d = Sf +INV Δ*tolR* Δ*rfbG*. M= Intermediate labeling (CD_2O, $NaBH_3CN$). H= heavy labeling ($13CD_2O$, $NaBD_3C$).

The first MS analysis highlighted the presence of a recurrent spectrum profile that was attributed to a polymeric contamination of the sample. In order to decrease this background, all of the labeled samples of Ss1 and Sf1 were subjected to a desalting process and re-analyzed by GeLC-MS/MS, generating a technical replicate of each sample. To distinguish between the 2 replicates, the original samples were labeled as Ss1a and Sf1a and the desalted samples as Ss1b and Sf1b. A schematic representation of the experimental procedure is shown in figure 16.

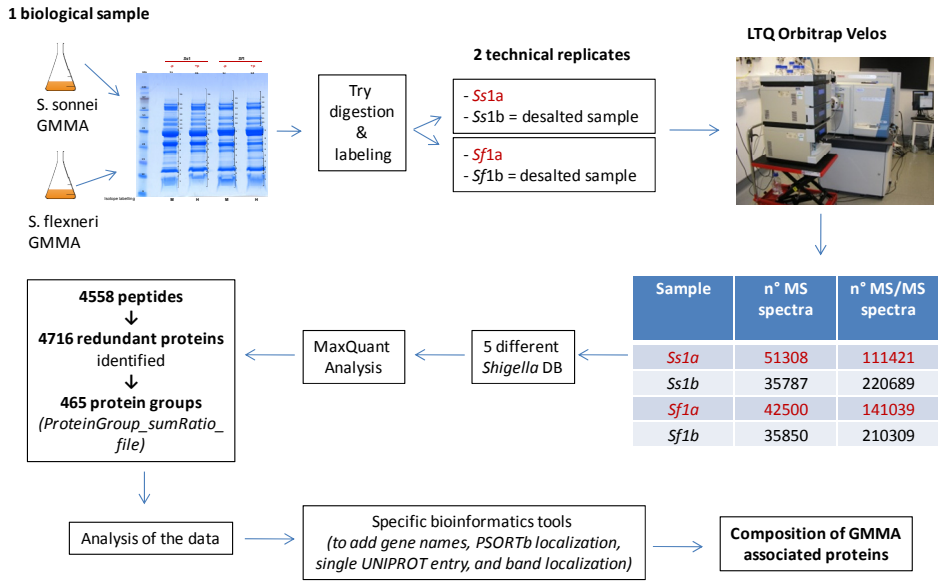


Figure 16. Experimental flow of proteomic analysis of *S. sonnei* and *S. flexneri* GMMA with two technical replicates each.
Shigella sonnei and *Shigella flexneri* GMMA were separated by SDS-PAGE and each lane of the gel was cut into slices followed by digestion with trypsin and isotope labeling. The 'heavy'

and 'intermediate' labeled samples of the same strain were combined producing the samples Ss1 and Sf1. Each sample was processed twice using the LTQ Orbitrap Velos: the first time without additional processing (a), the second time after a desalting step (b). More than 35000 MS spectra and 111000 MS/MS spectra were obtained for each sample. The spectra were first analyzed against 5 different databases, and subsequently processed by MaxQuant software. The final file containing 465 non-redundant Protein Groups was manually checked eliminating Protein Groups with a posterior error probability (PEP) >0.01 and Protein Groups indicated as 'contaminants' and 'reverse'. Gene names, PSORTb 3.0 localization, and number of slice of the gel in which the proteins were identified, were added to the 'Protein Group' table.

More than 35000 MS spectra and 111000 MS/MS spectra were obtained from each sample. The MS spectra contain the information on the intact peptide masses and intensities and the MS/MS spectra contain fragment ions that reveal the sequence information. A database search algorithm is used to correlate spectra against a sequence database and identify the corresponding peptide[27]. The MS/MS spectra were searched with Andromeda engine against 4 different *Shigella* databases (*Shigella sonnei, Shigella flexneri, Shigella boydii, Shigella dysenteriae)* plus an in-house built plasmid database[27]. Subsequently, all the sequences identified in a defined range of confidence are combined to a list of proteins with specific scores of identification. The Andromeda results are then analyzed by the quantitative proteomic software MaxQuant[26]. This software is designed to analyze large mass spectrometry data sets from proteomics studies. It performs several data analysis: validates peptides identifications, assembles the peptides to protein identifications, quantifies the differentially labeled peptides, and calculates protein ratios. MaxQuant can perform multi-sample comparisons in a single MaxQuant project favoring a direct comparison of quantitative information. In particular, it measures the relative ion intensity of the 'H' and 'M' labeled peptide ions for each protein. To understand the assembling process of peptides hits to protein hits, is important to define the concept of 'Protein Group', 'unique' peptides and 'razor' peptides. Whenever the group of peptides identified in one protein is identical to or completely contained in the group of peptides identified in a different protein these two proteins are associated in a Protein Group[25]. 'Unique' peptides are peptides found only in one Protein Group. 'Razor'[102] peptides are, instead, peptides with sequences that are present in multiple Protein Groups (shared peptides) and are parsimoniously associated with the Protein Group showing the highest number of total identified

peptides, but remain in all the Protein Group where they occur. The main output of MaxQuant analysis is represented by 2 tables:

- 'Evidence' table containing all the information at the peptide level: peptide sequences, post-translational modifications, labeling state, Protein Group the peptide belongs to;
- 'Protein Group' table, containing all the information at the protein level: list of the identified proteins (Protein Groups IDs), number of 'unique' and 'razor' peptides identified for each protein, and cross-reference to several databases (e.g. Gene Ontology term).

4558 peptides were matched to 4716 partially redundant proteins. The redundancy could result from incomplete clustering resulting from the use of multiple serogroups databases (table 7). All the identifications attributed to the plasmid database were subsumable to the 4 main *Shigella* databases. The redundant proteins were automatically combined and a list of 465 'Protein Groups' was identified. All the Protein Group rows were manually analyzed: the Protein Group indicated as 'reverse' (proteins derived from the reversed part of the decoy database[26]) and 'contaminants' (e.g. human keratin and trypsin) in the 'Protein Group' table were eliminated. A cut off score of the posterior error probability (PEP, used to calculate the 'false discovery rate' FDR for each spectrum) >0.01 was used as well to eliminate Protein Groups from the original table. Gene names of the proteins, PSORTb 3.0 localization, and number of slice of the gel in which the proteins were identified were inserted into 'Protein Group' table in order to facilitate the identification of the exact protein composition of GMMA.

To calculate how many proteins were identified in each sample (Ss1a, Ss1b, Sf1a, Sf1b), we considered the number of 'razor' peptide and 'unique' peptides identified in each sample. As by definition the 'razor' peptides can be found in different Protein Groups, they represent a critical issue to define if a protein is present or not in a sample as they could result in false positive protein identifications. Therefore, we used a cutoff of >= 3 'razor' peptides and at least 1 'unique' peptide to call a Protein Group present. Looking at the table 4, showing the number of the proteins identified in the samples Ss1 and Sf1 within each database, we can observe that SHISO and SHISS databases are completely overlapped as if we only searched in one of the two *S. sonnei* databases we would have not missed any protein. On the contrary, if we

only searched against SHIFL database, we would have missed 3 proteins identified in SHIF8. The databases within *S. dysenteriae* and *S. boydii* are, instead, completely overlapped. However, analyzing the 465 proteins identified, we still found some redundancy as several proteins (annotated differently in different databases) were found more than once in this list.

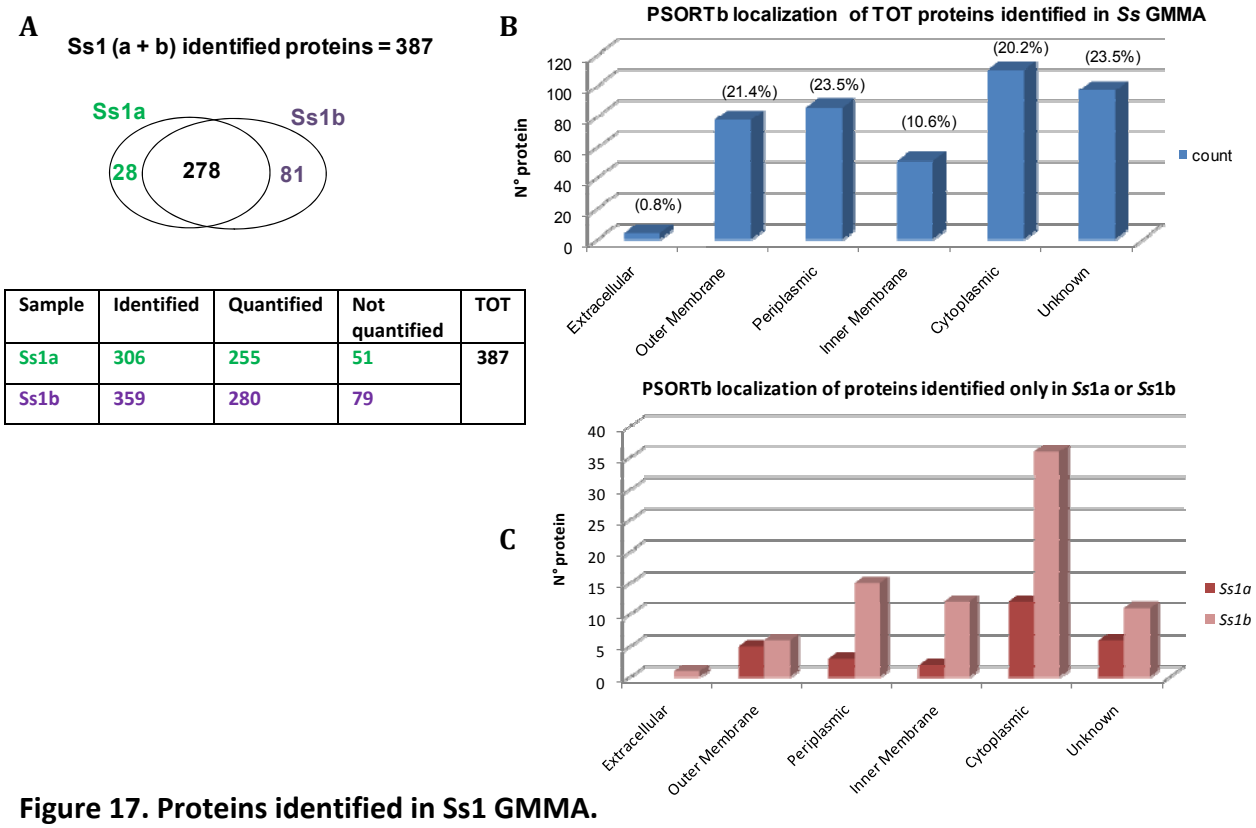
Combined Database	Individual Database	TOT proteins	n° proteins identified in GMMA	TOT proteins identified in the combined databases					
S. sonnei	SHISO	10147	371	380	449	452	463	465	465
	SHISS		366						
S. flexneri	SHIFL	28189	415						
	SHIF8		339						
S. dysenteriae	SHIDS	3918	317						
	SHIDY		327						
S. boydii	SHIBS	8085	336						
	SHIB3		340						

Table 7. GMMA-associated proteins identified using different *Shigella* databases. GMMA associated proteins have been identified using different combined databases: *S. sonnei* (**SHISO=** *Ss53G* + *Ss*, **SHISS**= *Ss046*), *S. flexneri* **(SHIFL**= *Sf1a*, *Sf1b*, *Sf2a*, *Sf3b*, *Sf4a*, *Sf5*, *Sf5a*, *Sf6*, *SfY*, *Sf CDC 796-83*; **SHIF8**= *Sf5b strain 8401*), *S. dysentariae* (**SHIDY**= *Sd*, **SHIDS=** *Sd CDC 74-1112*), and *S. boydii* (**SHIBS=** *Sb4*, **SHIB3=** *Sb18*). The results from databases within the same serogroup were combined and the total number of proteins identified in each combined database is shown in column 'TOT proteins'. In the next column is shown how many proteins (corresponding to individual Protein Groups) were identified in GMMA against each individual database. In the last column, are shown how many new proteins were identified in GMMA using progressively each database.

5.3.2.1. Proteome of Shigella sonnei GMMA

As shown in figure 17A, in the *S. sonnei* Ss1a not desalted GMMA sample 306 Protein Groups were identified of which 255 were also quantified. In *S. sonnei* Ss1b desalted GMMA 359 Protein Groups were identified and 280 quantified. Combining the two lists of Protein Groups, 387 different proteins were identified in sample Ss1 of which 278 proteins were identified in both of the replicates. According to the PSORTb 3.0 prediction, the 387 proteins identified in *S. sonnei* GMMA were predicted as follows: 0.8% extracellular, 21.4% outer membrane, 23.5% periplasmic, 10.6% inner membrane, 20.2% cytoplasmic, and 23.5% unknown (figure 17B). The predicted localization of proteins identified only in Ss1a or Ss1b were distributed among all subcellular compartments indicating that the desalting process did not enhance the recovery of proteins of a specific compartment (figure 17C). Among the proteins, we

identified the iron regulated proteins FepA, FhuA and CirA previously identified with the MALDI-TOF approach and 5 new proteins: Iron(III) dicitrate transport protein fecA (FecA), TonB-dependent Receptor Plug domain protein (CjrC), putative iron-regulated protein (CjrA), ferric aerobactin receptor (IutA), and outer membrane receptor for ferric iron uptake (FhuE). In addition, we identified 3 proteins encoded on the virulence plasmids that were not identified with the MALDI-TOF approach: outer membrane autotransporter barrel domain protein (IcsA/VirG) involved in actin polymerization, outer membrane protein MxiD involved in the ring formation of type III secretion system, and lipoprotein MxiM that interacts with MxiD affecting its stability and envelope association. Additional proteins associated with *Shigella* virulence, but encoded on the chromosome were identified: Serine protease autotransporter SepA (required for actin assembling during infection), Serine protease (SigA), TieB protein (enterotoxin encoded by *senB*) and Entry exclusion protein 2 (Exc).

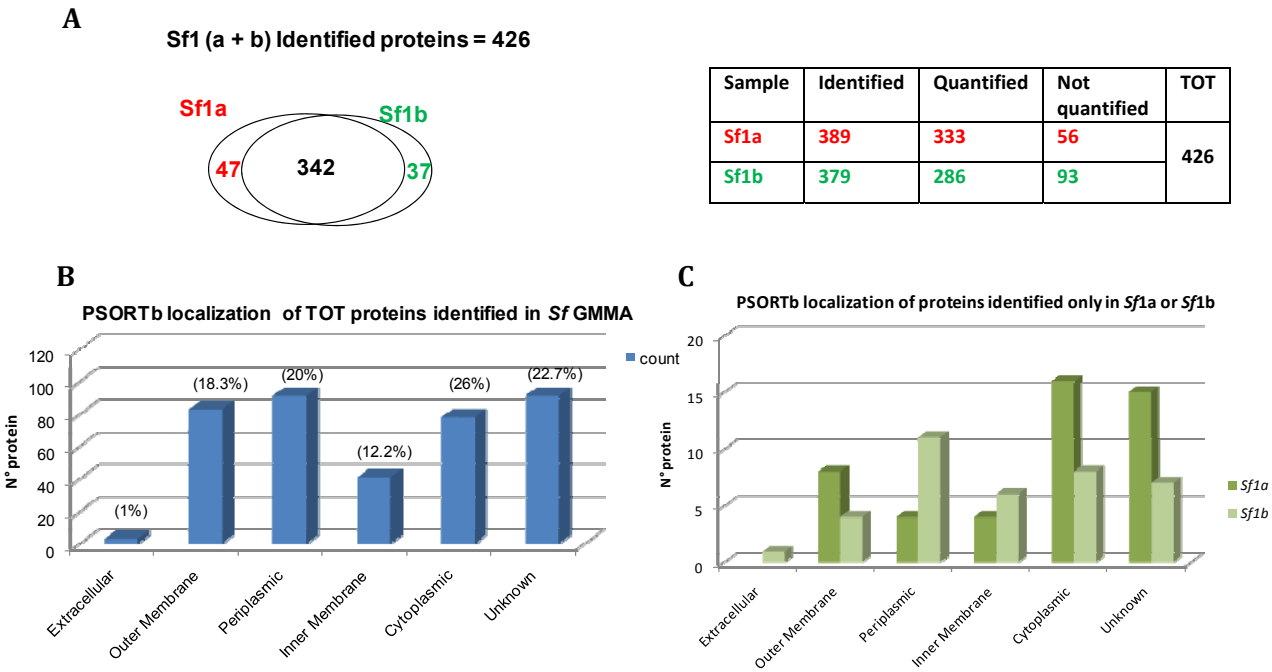


Sample	Identified	Quantified	Not quantified	TOT
Ss1a	306	255	51	387
Ss1b	359	280	79	

Figure 17. Proteins identified in Ss1 GMMA.
A. Number of proteins identified and quantified in 2 different runs of *S. sonnei* sample in the LTQ-Orbitrap. Ss1a = 1st run. Ss1b = 2nd run, after desalting of the samples. **B.** Predicted sub-cellular localizations of total proteins (387) identified in Ss1a + Ss1b using PSORTb 3.0. **C.** number and predicted localization of proteins identified only in Ss1a or Ss1b.

5.3.2.2. Proteome of Shigella flexneri GMMA

As show in figure 18A, in the *S. flexneri* Sf1a not desalted GMMA sample 389 Protein Groups were identified of which 333 were also quantified. In *S. flexneri* Sf1b desalted sample 379 Protein Groups were identified and 286 quantified. Combining the 2 lists of Protein Groups, 426 different Protein Groups were identified in sample Sf1 of which 342 were identified in both of the replicates. According to the PSORTb 3.0 prediction, the 426 proteins identified in *S. flexneri* GMMA were predicted as follows: 1% extracellular, 18.3% outer membrane, 20% periplasmic, 12.2% inner membrane, 26% cytoplasmic, and 22.7% unknown (figure 18B). Several proteins already identified in *S. sonnei* GMMA were found in *S. flexneri* GMMA as well: 5 iron regulated proteins FhuE, IutA, FhuA, CirA and FepA, the plasmid encoded proteins MxiM and MxiD, and the virulence associated proteins SepA, SigA and TieB. As observed for *S. sonnei*, the proteins found only in Sf1a or Sf1b showed a random predicted localization and we did not observe a recovery of proteins from a specific compartment after the desalting process (figure 18C). Overall, the two proteomes appear to show similar localization distribution.



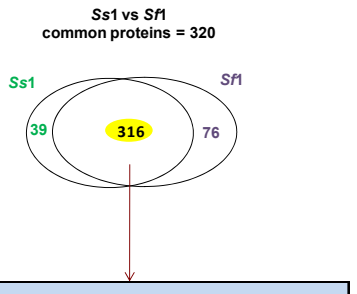
Sample	Identified	Quantified	Not quantified	TOT
Sf1a	389	333	56	426
Sf1b	379	286	93	

Figure 18. Proteins identified in Sf1 GMMA.
A. Number of proteins identified and quantified in 2 different runs of *S. flexneri* sample in the LTQ-Orbitrap. Sf1a = 1st run. Sf1b = 2nd run, after desalting of the samples. **B.** Predicted sub-cellular localizations of total proteins (426) identified in Sf1a + Sf1b using PSORTb 3.0. **C.** number and predicted localization of proteins identified only in Sf1a or Sf1b.

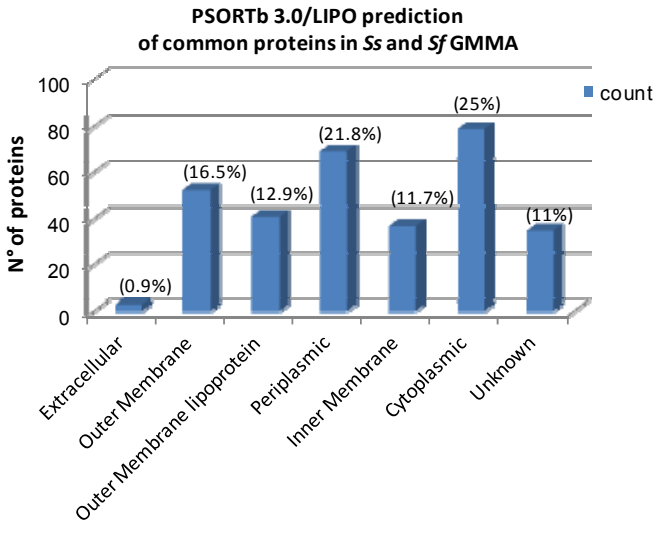
5.3.2.3. Conserved proteins in S. sonnei and S. flexneri GMMA

To identify how many and which proteins are conserved in *S. sonnei* and *S. flexneri* GMMA, the Ss1 and Sf1 lists of Protein Groups IDs were compared. From the comparison we obtained a list of 350 Protein Groups shared between *S. sonnei* and *S. flexneri* GMMA. In order to eliminate redundant proteins within the list of shared proteins, we manually checked each Protein Group identified in *S. sonnei* and *S. flexneri* GMMA looking specifically for SHISO and SHIFL (*Sf2a* when possible) entries. Whenever different Protein Groups contained the same gene name, we considered the ones with SHISO and SHIFL entries. The whole list was checked to exclude these 'redundant' proteins and the final list contained 316 proteins in common between *S. sonnei* and *S. flexneri* GMMA (figure 19A). According to the PSORTb 3.0 prediction the 316 proteins were classified as: 0.9% extracellular, 16.5% outer membrane, 21.8% periplasmic, 11.7% inner membrane, 25% cytoplasmic and 23.9% unknown. The sequences that were predicted as 'unknown' by PSORTb 3.0 were subsequently analyzed using the LIPO program to identify potential lipoproteins. More than half of the unknown proteins were predicted by LIPO to be outer membrane lipoproteins accounting for 12.9% of all identified proteins. For 11% the subcellular localization remained unknown (figure 19B-C). The list of the OM and Extracellular proteins in common is shown in table 8.

A



B



C

PSORTb localization	count	%
Extracellular	3	0,9
Outer Membrane	52	16,5
Outer Membrane lipoprotein	41	12,9
Periplasmic	69	21,8
Inner Membrane	37	11,7
Cytoplasmic	79	25,0
Unknown	35	11,0
TOT	**316**	**100**

Figure 19. Common proteins found in *S. sonnei* and *S. flexneri* GMMA.
A. Number of proteins found in *S. sonnei* or *S. flexneri* GMMA. The number highlighted in yellow represents the common proteins in *S. sonnei* and *S. flexneri* GMMA, after the

exclusion of 'redundant' proteins present more than once in the original Ss1 and Sf1 lists. **B, C.** Sub-cellular localizations of the common proteins predicted using PSORTb 3.0 and LIPO tools.

ID in GMMA	protein name	common annotation	annotation in *Ss53G*	PSORTb 3.0
261	outer membrane lipoprotein Blc	blc	SS53G_2967	OM
178	vitamin B12/cobalamin outer membrane transporter	btuB		OM
310	lipoprotein involved with copper homeostasis and adhesion	cutF	SS53G_3812	OM
165	long-chain fatty acid outer membrane transporter	fadL	SS53G_0810	OM
353	outer membrane receptor FepA	fepA	SS53G_5666	OM
400	ferrichrome outer membrane transporter	fhuA	SS53G_3761	OM
464	ferric-rhodotorulic acid outer membrane transporter	fhuE	SS53G_4297	OM
443/361	putative ferric siderophore receptor	iutA	SS53G_5072	OM
176	maltoporin	lamB		OM
173	murein lipoprotein	lpp	SS53G_0415	OM
271	murein transglycosylase A	mltA	SS53G_1862	OM
175	outer membrane protein mxiD	mxiD		OM
100	lipoprotein	nlpB	SS53G_0626	OM
129	lipoprotein NlpD	nlpD	SS53G_2538	OM
419/397	putative outer membrane porin protein C precurser	nmpC	SS53G_0760	OM
225/481	outer membrane protein A	ompA	SS53G_4120	OM
421/376	outer membrane porin protein C	ompC	SS53G_0957	OM
228/363	outer membrane protein F	ompF	SS53G_4087	OM
105	outer membrane protein X	ompX	SS53G_3941	OM
103	peptidoglycan-associated outer membrane lipoprotein	pal		OM
106	phospholipase A	pldA	SS53G_2160	OM
402	outer membrane lipoprotein	rcsF		OM
19	LPS-assembly lipoprotein RlpB	rlpB	lptE	OM
332	outer membrane protein induced after carbon starvation	slp	SS53G_4970	OM
101	putative outer membrane protein	slyB	SS53G_0370	OM
109	Small protein A	smpA	SS53G_1935	OM/Ukn
423/408	outer membrane channel protein	tolC	SS53G_1603	OM
446/107	Tsx Nucleoside-specific channel-forming protein	tsx	SS53G_5490	OM
163	lipoprotein precursor	vacJ	SS53G_0804	OM
265	putative polysaccharide export protein	wza	SS53G_1145	OM
84/198	outer membrane protein assembly factor YaeT	yaeT		OM
240/362	putative pectinesterase	ybhC	SS53G_3886	OM
108	exopolysaccharide export protein	yccZ	SS53G_4155	OM
120	hypothetical protein S1194	ycfJ	SS53G_4305	OM
274	hypothetical protein S1307	ychP	SS53G_1321	OM

277/104	outer membrane protein W	yciD	SS53G_1360	OM
444	hypothetical protein S1861	yddB	SS53G_5268	OM
102	hypothetical protein S1556	yeaF	SS53G_3445	OM
393	putative outer membrane lipoprotein	yfiB	SS53G_1923	OM
338	putative outer membrane lipoprotein	yiaD	SS53G_2796	OM
406	hypothetical protein S3555	yjbH	SS53G_2454	OM
438	hypothetical protein S1053	ymcA	SS53G_4156	OM
270	hypothetical protein S4531	ytfM	SS53G_3045	OM
279	putative membrane protein precursor	SSON_1762	SS53G_1409	OM
283	putative lipoprotein	SSON_3016		OM
288	Outer membrane protein N	SSON_1748	SS53G_0050	OM
346	outer membrane fluffing protein	flu		OM
387	outer membrane protein assembly complex subunit YfgL	yfgL	SSON_2594	OM
392	outer membrane protein assembly complex subunit YfiO	yfiO	SSON_2721	OM
194/212	organic solvent tolerance protein	imp		OM
205/216	Outer membrane protein IcsA autotransporter	virG/icsA		OM
475	colicin I receptor	cirA	SS53G_1033	OM
424	Serine protease	sigA		OM/Ext
358	Rare lipoprotein A	rlpA		Ext/Ukn
213	Serine protease sepA autotransporter	sepA		Ext
455	TieB protein	senB		Ukn
94	Lipoprotein mxiM	mxiM		OM lipo

Table 8. OM/Ext proteins and proteins with a virulence-associated function conserved in *S. sonnei* and *S. flexneri* GMMA.
52 OM proteins and 3 potentially extracellular proteins are shown in the table. In the last 2 rows are shown 2 proteins with a virulence-associated function with a different predicted localization.

52% of proteins identified in GMMA were found to be associated with the outer membrane, extracellular and periplasmic space. Proteins of these compartments were expected to be present in GMMA. However, also 24.9% cytoplasmic proteins and 11.4% inner membrane proteins were identified. In order to analyze if outer membrane and periplasmic proteins were enriched in GMMA or if the distribution of identified proteins reflected the distribution of proteins in the bacterial cells, the *in silico* proteomes of *SSs046* and *Sf2a 2457T* were processed by PSORTb 3.0. As the complete genome of *Ss53G* was not publicly available, the strain *Ss046* was used as reference for *S. sonnei* GMMA. Subsequently, the predictions of the *in silico* proteomes were compared with the PSORTb 3.0 localization predictions of the proteins identified in *S. sonnei* and *S. flexneri* GMMA. As shown in figure 20, all of the

percentages of predicted proteins in the *in silico* proteomes that were identified in GMMA were very similar for *S. sonnei* and *S. flexneri*. The most striking result was that the outer membrane proteins identified in GMMA represented 71.5% and 65%, respectively, of the total predicted OM proteins in the bacterial proteomes. The remaining proteins of the predicted proteomes that were not identified in GMMA were mainly proteins defined as: 'outer membrane proteins' (e.g. multidrug resistance outer membrane protein MdtQ in *S. sonnei*, and copper/silver efflux system outer membrane protein CusC in *S. flexneri*), 'outer membrane porin' (e.g. outer membrane phosphoporin E and outer membrane porin HofQ, both predicted in *S. sonnei* and *S. flexneri*), fimbrial-related proteins, and hypothetical proteins. The periplasmic proteins indentified in GMMA represented roughly 50% of the total predicted periplasmic proteins. The cytoplasmic and inner membrane proteins, instead, represented only approximately 4%. These results confirmed that GMMA were enriched in outer membrane and periplasmic proteins, and that only small fraction of cytoplasmic and inner membrane proteins were present.

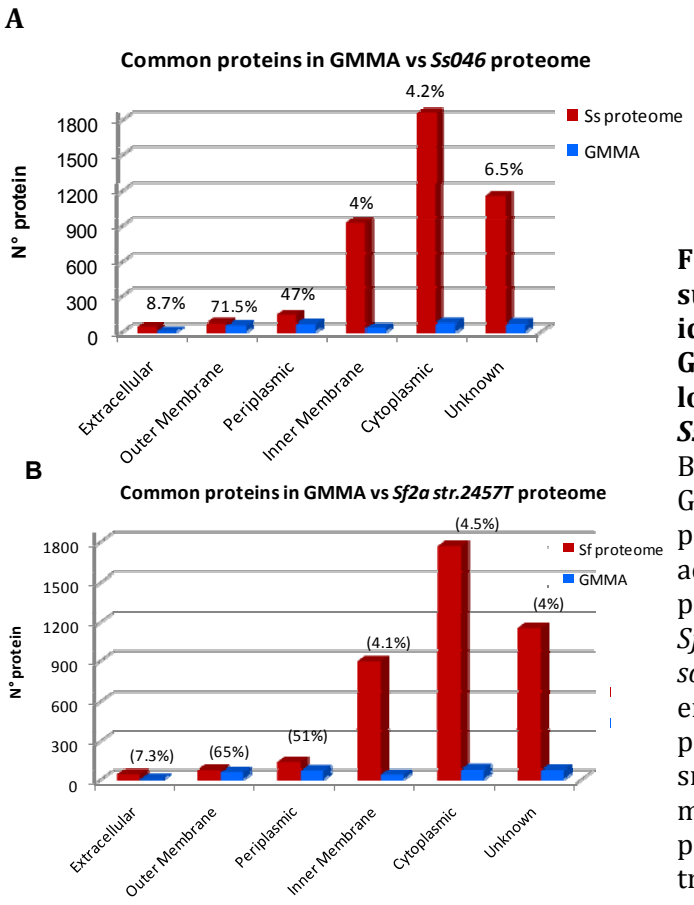


Figure 20. Comparison of the subcellular localization of proteins identified in *S. sonnei* and *S. flexneri* GMMA with the predicted localization of *in silico* proteomes of *Ss046* and *Sf2a 2457T*.
Blue: number of conserved proteins GMMA, red: number of proteins predicted in each translated genome according to PSORTb 3.0 prediction. **A.** proteome of *Ss046*, **B.** proteome of *Sf2a 2457T*. The 2 graphs show that *S. sonnei* and *S. flexneri* GMMA are enriched in outer membrane and periplasmic proteins, and contain only small fractions of cytoplasmic, inner membrane and 'unknown' proteins predicted in the respective total translated genomes.

5.3.3. Quantitative analysis

Knowing that the composition of GMMA was enriched in outer membrane and periplasmic proteins, we were interested in analyzing if this enrichment was also seen in protein amounts. As *S. sonnei* and *S. flexneri* GMMA were labeled and processed separately, the dimethyl labeling cannot be used to perform a quantitative analysis. To provide an estimate of the quantification of a specific protein in Ss1 and Sf1 GMMA, the Exponentially Modified Protein Abundance Index (emPAI) was used[62]. The emPAI is defined as:

$$emPAI = 10^{PAI} - 1$$

where PAI (PAI= N_{obvd}/N_{obvl}) is defined as the number of observed peptides divided by the number of theoretically observable peptides per proteins, respectively[116].

The observed peptides per protein (N_{obvd}) where calculated as follows:

- Identical peptide sequences with different modifications and/or the different charges were counted as separate observed peptides, while the miss-cleavages of the same peptide were excluded;
- All of the peptides of each ProteinGroup ID (razor + unique) were searched against the *Ss53G* or *Sf2a str.2457T* databases. If no results were obtained the search was firstly extended to *Ss046* and *S. flexneri* databases, respectively, secondly to all *Shigella* databases except *S. flexneri* databases for *S. sonnei* GMMA, and *S. sonnei* databases for *S. flexneri* GMMA;
- A list of proteins matching all the peptides of each ProteinGroup was created;
- Within the list, the protein with the highest number of peptide matches was chosen;
- As the longest protein could be preset in different strains, and within each strain there could be many proteins with the same n° of matches, but different length, the longest protein was chosen. This approach is based on accurate sequence of proteins, which may not be there for all the proteins. For this reason, the lower the number of peptide is the more inaccurate the emPAI is if the sequence is not correct.

To calculate the number of observable peptides per protein (N_{obvl}), each protein sequence was digested *in silico* with trypsin, and the obtained masses were filtered

according to the mass range 700-5000 Da. The number of peptides falling in this range was considered as the number of 'observable' peptides. 'emPAI' values >6, >3 and <0.1 were chosen to define a protein as extremely abundant, abundant, and not abundant.

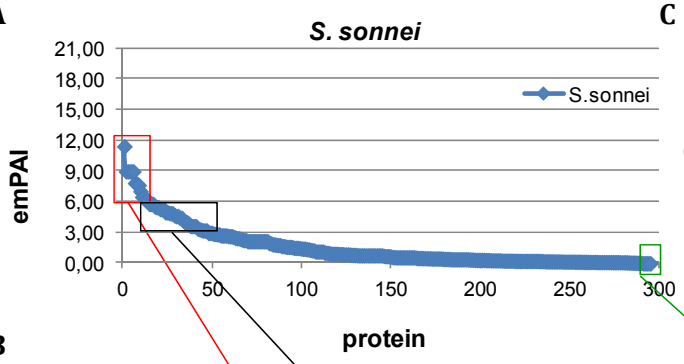
5.3.3.1. Most abundant proteins in S. sonnei GMMA

Analyzing the identified proteins in *S. sonnei* GMMA approximately one third of the 316 common proteins had an emPAI value >1 (figure 21A). For 52 proteins a score higher than 3 was obtained. These proteins were predominantly OM and periplasmic proteins. Only 1 IM and 6 unknown proteins were found in this group (figure 21B). 14 of the 52 proteins with an emPAI score >3 showed a score between 6 and 11.5. These were predominantly OM proteins and included the major porins OmpA and OmpC and the iron regulated protein CirA. The iron regulated proteins FepA and IutA showed an emPAI value of roughly 5, while proteins encoded on the virulence plasmid or with virulence associated function showed an emPAI< 0.2. The proteins with the lowest emPAI scores were, instead, predominantly cytoplasmic. The protein contents were subsequently expressed in molar fraction percentage as:

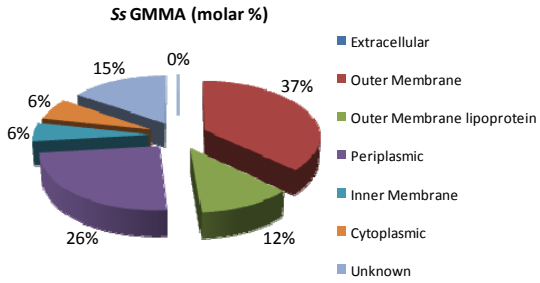
$$Protein\ content\ (mol\ \%) = (emPAI/\sum(emPAI)) \times 100$$

where the ∑(emPAI) is the sum of the emPAI values for all of the identified proteins. According to this quantification the OM, OM lipo and periplasmic proteins represented the 75% of the total protein content in *S. sonnei* GMMA (figure 21C), while cytoplasmic and IM proteins only accounted for 6% each. In particular, OmpA and OmpC showed the highest emPAI scores indicating that they represent the most abundant proteins in GMMA as already expected from the 1-D and 2-D gels previously discussed in this chapter. CirA and the periplasmic proteins YraP and TolB were shown to be abundant as well.

A



C



B

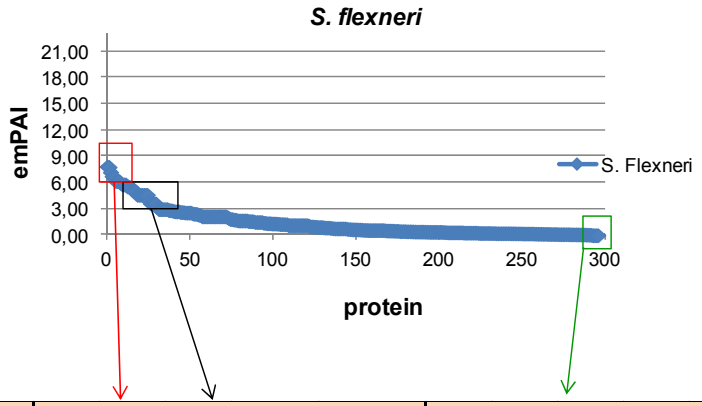
Protein n°	Protein Group ID	emPAI value >6	emPAI value >3	Localization	gene	Protein n°	Portein Group ID	emPAI value <0.1	Localizatic	gene
1	421	11,45		OM	ompC	302	86	0,10	Cyto	clpX
2	130	9,00		OM lipo	osmE	303	157	0,10	Cyto	rpsA
3	173	9,00		OM	lpp	304	269	0,09	Cyto	SS53G_3059
4	225	9,00		OM	ompA	305	366	0,09	Cyto	tnpA
5	303	9,00		Peri	yraP	306	48	0,08	Cyto	pyrG
6	419	9,00		OM	SS53G_0760	307	55	0,07	Cyto	rnb
7	100	7,86		OM lipo	SS53G_0626	308	30	0,07	Cyto	nuoD
8	20	7,80		Peri	tolB	309	296	0,06	Cyto	SS53G_4498
9	473	7,66		Ukn	SDY_1953	310	186	0,06	Cyto	mukB
10	475	7,06		OM	cirA	311	65	0,05	IM	secA
11	105	6,50		OM	SS53G_3941	312	182	0,05	Cyto	alaS
12	294	6,50		Ukn	SS53G_0128	313	204	0,04	Ext	pic
13	125	6,20		Peri	surA	314	410	0,04	Ukn	yhjL
14	147	6,20		Peri	skp	315	52	0,00	Cyto	deoD
16	134		5,81	Peri	yggE	316	62	0,00	Cyto	rpmA
17	284		5,81	OM lipo	SS53G_1727	317	168	0,00	OM lipo	SS53G_0056
18	291		5,81	Ukn	yggN	318	369	0,00	Cyto	icdA
19	84		5,58	OM	yaeT	320	449	0,00	Ukn	SS53G_5644
20	283		5,49	OM	SSON_3016					
21	392		5,49	OM	yfiO					
22	408		5,31	OM	tolC					
23	411		5,31	OM	SS53G_0127					
24	437		5,31	OM	SS53G_3761					
25	484		5,31	Peri	htrA					
26	353		5,14	OM	fepA					
27	103		4,99	OM	pal					
28	108		4,99	OM	SS53G_4155					
29	441		4,93	Peri	SS53G_4658					
30	200		4,91	IM	lpoA					
31	358		4,88	OM lipo	rlpA					
32	476		4,70	Peri	bglX					
33	127		4,62	Ukn	yfhG					
34	349		4,62	OM lipo	SS53G_5491					
35	361		4,53	OM	iutA					
36	387		4,46	OM	yfgL					
37	415		4,46	OM	SS53G_0290					
38	420		4,18	Ukn	SS53G_0912					
39	19		4,01	OM	lptE					
40	382		4,01	OM lipo	ydgH					
41	106		3,87	OM	pldA					
42	101		3,64	OM	SS53G_0370					
43	256		3,64	OM lipo	SS53G_1294					
44	289		3,64	OM lipo	SS53G_0053					
45	357		3,64	Peri	ybgF					
46	442		3,39	Peri	SS53G_4771					
47	118		3,33	Peri	SS53G_4200					
48	194		3,27	OM	lptD					
49	201		3,22	Peri	eco					
50	240		3,22	OM	SS53G_3886					
51	273		3,22	Peri	SS53G_3546					
52	360		3,22	Ukn	ybjP					

Figure 21. emPAI quantification and most abundant proteins in *S. sonnei* GMMA.
A. Distribution of emPAI scores, **B.** list of the most (emPAI >3) and the least abundant proteins (emPAI <0.1) in *S. sonnei* GMMA. Roughly one third of the proteins has an emPAI value <0.1, but in the table are only reported the last 18 proteins of the graph, **C.** relative quantification (molar %) of proteins for each sub-cellular compartment.

5.3.3.2. Most abundant proteins in S. flexneri GMMA

In *S. flexneri* GMMA, 126 of the 316 conserved proteins had an emPAI value >=1 (figure 22A). 35 proteins obtained a score >3 and among them 8 had a score between 6 and 8. Also in *S. flexneri* GMMA the majority of the proteins with high emPAI values were OM (e.g. IutA, OmpA and OmpC) and periplasmic proteins (e.g. TolB and IraP). The only two cytoplasmic proteins Tuf and RpsE showed a score of 3.6 and the IM LpoA showed a score of 4.9. The proteins with the lowest emPAI scores were predominantly cytoplasmic (figure 22B). The molar percentages of the proteins of each sub-cellular localization are shown in figure 22C. In particular the OM, OM lipoproteins, and periplasmic proteins counted for the 66% of the total proteins.

A



B

Protein n°	Protein Group ID	emPAI value >6	emPAI value >3	Localization	gene	Protein n°	Portein Group ID	emPAI value <0.1	Localizatic	gene
1	100	7,86		OM lipo	nlpB	304	459	0,10	Cyto	SGF_02692
2	20	7,80		Peri	tolB	305	86	0,10	Cyto	clpX
3	361	7,21		OM	iutA	306	182	0,10	Cyto	alaS
4	303	6,74		Peri	yraP	308	190	0,09	IM	SF2457T_3419
5	408	6,59		OM	tolC	309	403	0,09	Peri	SF2457T_4880
6	358	6,20		OM lipo	rlpA	310	428	0,09	IM	SGF_04083
7	125	6,20		Peri	surA	311	300	0,09	IM	SF2457T_2545
8	147	6,20		Peri	skp	312	281	0,08	Peri	SF2457T_0921
9	225		5,95	OM	ompA	313	55	0,07	Cyto	rnb
10	101		5,81	OM	slyB	314	296	0,06	Cyto	SF2457T_3149
11	134		5,81	Peri	yggE	315	28	0,06	Peri	napA
12	357		5,81	Peri	ygbF	316	366	0,04	Cyto	tnpA
13	84		5,58	OM	yaeT	317	52	0,00	Cyto	deoD
14	283		5,49	OM	SF2457T_3516	318	62	0,00	Cyto	rpmA
15	392		5,49	OM	yfiO	319	369	0,00	Cyto	icd
16	130		5,31	OM lipo	osmE	320	168	0,00	OM lipo	SF2457T_0957
17	484		5,31	Peri	htrA					
18	387		5,16	OM	yfgL					
19	103		4,99	OM	pal					
20	200		4,91	IM	lpoA					
21	421		4,78	OM	ompC					
22	105		4,62	OM	ompX					
23	349		4,62	OM lipo	SF2457T_4755					
24	382		4,62	OM lipo	SF2457T_2532					
25	284		4,62	OM lipo	SGF_03448					
26	127		4,62	Ukn	yfhG					
27	360		4,62	Ukn	ybjP					
28	386		4,62	OM lipo	SF2457T_0515					
29	129		3,92	OM	SF2457T_3403					
30	98		3,64	Cyto	rpsE					
31	177		3,64	Cyto	tuf					
32	173		3,64	OM	lpp					
33	291		3,64	Ukn	yggN					
34	476		3,31	Peri	bglX					
35	106		3,22	OM	pldA					

C

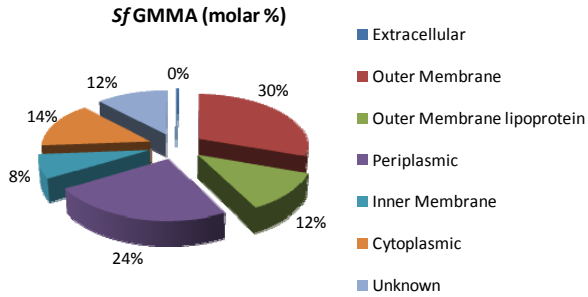


Figure 22. emPAI quantification and most abundant proteins in *S. flexneri* GMMA.
A. Distribution of emPAI scores, **B.** list of the most (emPAI >6) and the least abundant proteins (emPAI <0.1) in *S. flexneri* GMMA. emPAI value >3 has been chosen as threshold for the abundant proteins, and 0.1 as the not abundant ones. **C.** relative quantification (molar %) of proteins for each subcellular compartment.

5.3.3.3. Comparison of the abundant proteins in S. sonnei and S. flexneri GMMA

The lists of the most abundant proteins in Ss1 and Sf1 were compared to analyze if most proteins showed similar abundance in GMMA from both strains or if there were particular differences in the position of the proteins in the emPAI graphs. As shown in table 9, the proteins reported in the first 5 rows were estimated to be among the most abundant once in both *S. sonnei* and *S. flexneri* GMMA. OmpA and OmpC represented the 4th and the 1st proteins in the *S. sonnei* list, while in *S. flexneri* they were in position 9 and 21, respectively. For some proteins no comparison was possible since no quantification was available (N/A). Only few examples showed a significant difference between the two samples: the protein ID 118 (annotated as Imelysin family protein), ID 415 (TonB-dependent Receptor Plug domain protein), and ID 294 (PQQ enzyme repeat family protein) were more abundant in *S. sonnei* than in *S. flexneri* GMMA showing a difference of 100, 163, and 217 positions, respectively.

	position		emPAI > 3					position		emPAI > 3			
Protein Group ID	*Ss*	*Sf*	*Ss*	*Sf*	gene	Localization	Protein Group ID	*Ss*	*Sf*	*Ss*	*Sf*	gene	Localization
303	5	4	9,00	6,74	yraP	Peri	441	28	83	**4,93**	N/A	SS53G_4658	Peri
100	7	1	7,86	7,86	SS53G_0626	OM lipo	200	29	20	**4,91**	N/A	lpoA	IM
20	8	2	7,80	7,80	tolB	Peri	476	31	34	**4,70**	N/A	bglX	Peri
125	13	7	6,20	6,20	surA	Peri	127	32	26	**4,62**	**4,62**	yfhG	Ukn
147	14	8	6,20	6,20	skp	Peri	349	33	23	**4,62**	**4,62**	SS53G_5491	OM lipo
421	1	21	11,45	**4,78**	ompC	OM	387	35	18	**4,46**	**5,16**	yfgL	OM
130	2	16	9,00	**5,31**	osmE	OM lipo	415	36	199	**4,46**	0,44	SS53G_0290	OM
173	3	32	9,00	**3,64**	lpp	OM	420	37	217	**4,18**	N/A	SS53G_0912	Ukn
225	4	9	9,00	**5,95**	ompA	OM	19	38	40	**4,01**	2,98	lptE	OM
419	6	69	9,00	N/A	SS53G_0760	OM	382	39	24	**4,01**	**4,62**	ydgH	OM lipo
475	10	134	7,06	N/A	cirA	OM	106	40	35	**3,87**	**3,22**	pldA	OM
105	11	22	6,50	**4,62**	SS53G_3941	OM	101	41	10	**3,64**	**5,81**	SS53G_0370	OM
294	12	229	6,50	0,33	SS53G_0128	Ukn	256	42	56	**3,64**	2,59	SS53G_1294	OM lipo
408	21	5	**5,31**	6,59	tolC	OM	289	43	123	**3,64**	1,15	SS53G_0053	OM lipo
358	30	6	**4,88**	6,20	rlpA	OM lipo	357	44	12	**3,64**	**5,81**	ybgF	Peri
134	15	11	**5,81**	**5,81**	yggE	Peri	442	45	49	**3,39**	2,73	SS53G_4771	Peri
284	16	25	**5,81**	**4,62**	SS53G_1727	OM lipo	118	46	258	**3,33**	0,23	SS53G_4200	Peri
291	17	33	**5,81**	**3,64**	yggN	Ukn	194	47	43	**3,27**	2,86	lptD	OM
84	18	13	**5,58**	**5,58**	yaeT	OM	201	48	75	**3,22**	N/A	eco	Peri
283	19	14	**5,49**	**5,49**	SSON_3016	OM	240	49	70	**3,22**	2,16	SS53G_3886	OM
392	20	15	**5,49**	**5,49**	yfiO	OM	273	50	107	**3,22**	1,37	SS53G_3546	Peri
411	22	233	**5,31**	N/A	SS53G_0127	OM	360	N/A	27	N/A	**4,62**	ybjP	Ukn
437	23	136	**5,31**	N/A	SS53G_3761	OM	386	61	28	2,73	**4,62**	SF2457T_0515	OM lipo
484	24	17	**5,31**	N/A	htrA	Peri	129	68	29	2,46	**3,92**	SF2457T_3403	OM
103	26	19	**4,99**	**4,99**	pal	OM	98	91	30	1,78	**3,64**	rpsE	Cyto
108	27	54	**4,99**	2,59	SS53G_4155	OM	177	110	31	1,31	**3,64**	tuf	Cyto

Table9. Comparison of the most abundant proteins in *S. sonnei* and *S. flexneri* GMMA. The proteins are sorted firstly according to emPAI value >6 in *S. sonnei* and secondly in *S. flexneri*. **Red**= emPAI> 6; **Black**= emPAI> 3. For each protein its position in the emPAI graph of *S. sonnei* and *S. flexneri* GMMA, respectively, is also indicated. N/A= not available. The proteins showing the major differences in the position in *S. sonnei* and *S. flexneri* are highlighted in green.

To visualize the abundance of the 316 conserved proteins in *S. sonnei* and *S. flexneri* GMMA the positions of the proteins in the *S. sonnei* and *S. flexneri* GMMA emPAI were plotted against each other. Approximately two thirds of the proteins had comparable emPAI values and occupied a similar position in *S. sonnei* and *S. flexneri* GMMA, as seen by the diagonal distribution in the graph shown in figure 23. For example, Protein Groups 134, 284, 291, 84, 283, and 392 listed in table 9 showed positions between 15-20 in *S. sonnei* GMMA and 11-33 in *S. flexneri* GMMA being distributed at the beginning of the diagonal in figure 23. Many other proteins, instead, showed a difference in abundance indicated by different position in the 2 lists. The proteins with the largest difference in position in the emPAI list were the OM protein Exc, the OM lipoprotein LamB, and the cytoplasmic protein RpsC. Exc and LamB showed an

intermediate abundance in *S. sonnei* GMMA (emPAI values between 2 and 3) while RpsC showed an intermediate abundance in *S. flexneri* GMMA.

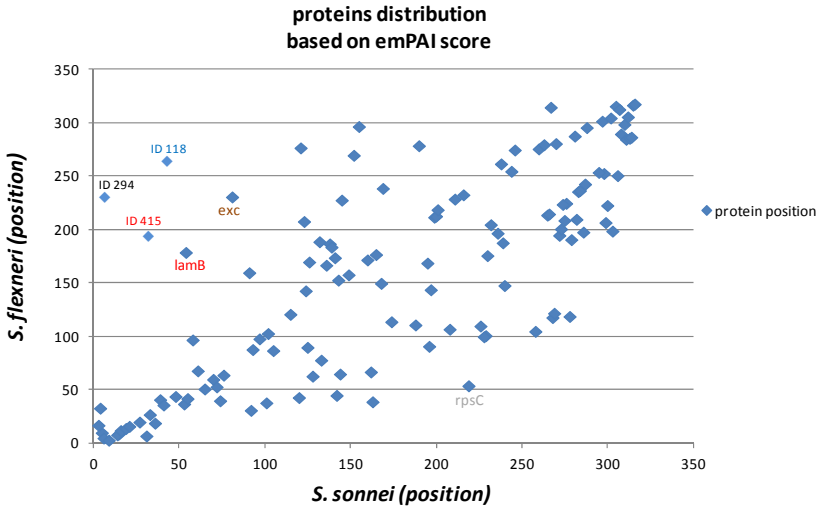
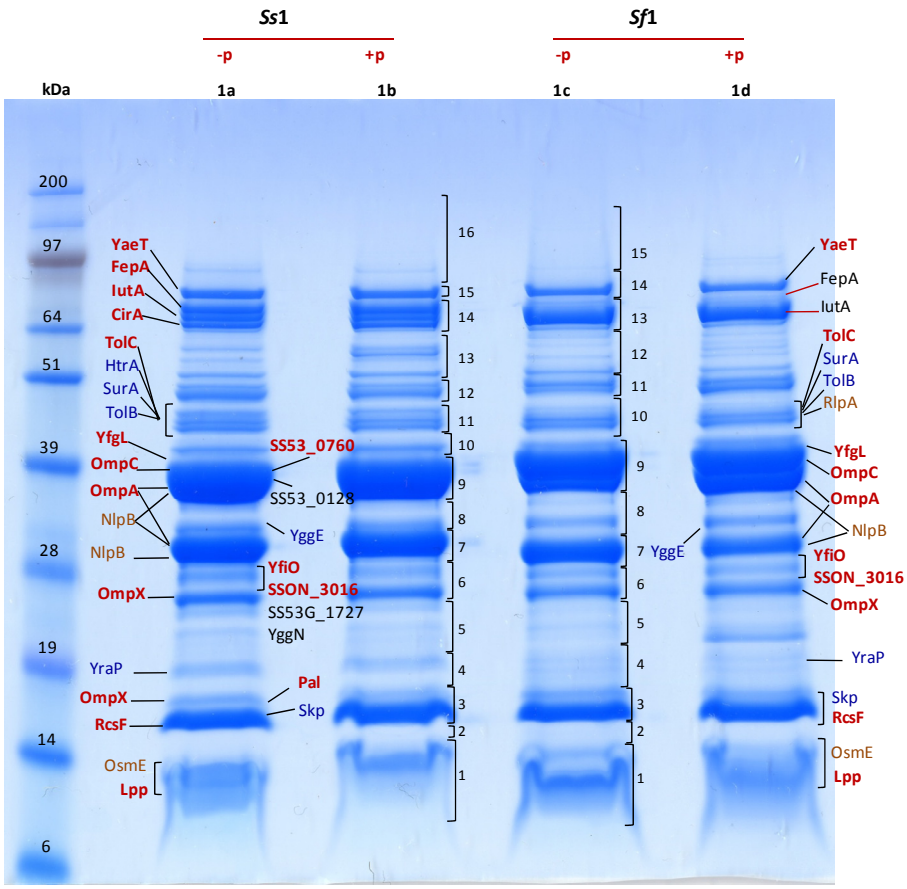


Figure 23. Comparison of abundance of conserved proteins in GMMA *S. sonnei* and *S. flexneri* based on emPAI scores.
The position of each conserved protein in *S. flexneri* GMMA is indicated on the Y axis, while the position in *S. sonnei* GMMA is indicated on the X axis. The protein names highlighted show a significant difference in the abundance of these proteins in the two type of GMMA: LamB and Exc seem to be more abundant in *S. sonnei* than in *S. flexneri* GMMA. RpsC, instead, was more abundant in *S. flexneri* than in *S. sonnei* GMMA.

We subsequently compared the results obtained from the quantitative analysis based on the emPAI score to the gel picture of *S. sonnei* and *S. flexneri* GMMA. The names of the most abundant proteins were assigned to the bands according to the predicted molecular weight of the proteins and the slice number in which they were identified. As shown in figure 24 the most abundant proteins matched quite well with the most visible bands in *S. sonnei* and *S. flexneri* GMMA. A schematic representation of the Gram-negative outer membrane and the localization of the most abundant proteins in *Shigella* GMMA are shown in figure 25.

A



B

Protein gene	MW (KDa)	Protein gene	MW (KDa)
cirA	74	SS53G_0760	39.6
htrA	51	SS53G_1727	31
iutA	81	SSON_3016	27.5
Lpp	8	surA	47
nlpB	37	tolB	46
ompA	37	tolC	54
ompC	40	taeT	90
ompX	18	ybgF	28
osmE	12	yfgL	41.8
rcsF	14	yfiO	27.8
rlpA	37.5	yggE	26.6
Skp	17.6	yggN	26
SS53G_0128	38.8	yraP	20

Figure 24. Gel representation of the most abundant proteins identified in GMMA.
A. Protein names were assigned to bands visible in the gel image of *S. sonnei* and *S. flexneri* GMMA based on the information obtained from the emPAI scores, the slice of the gel in which each protein was identified, and the molecular weight of the proteins. **Red:** outer membrane proteins, brown: outer membrane lipoproteins, blue: periplasmic proteins, black=: proteins with unknown subcellular localization. **B.** molecular weight of the proteins indicated in the gel.

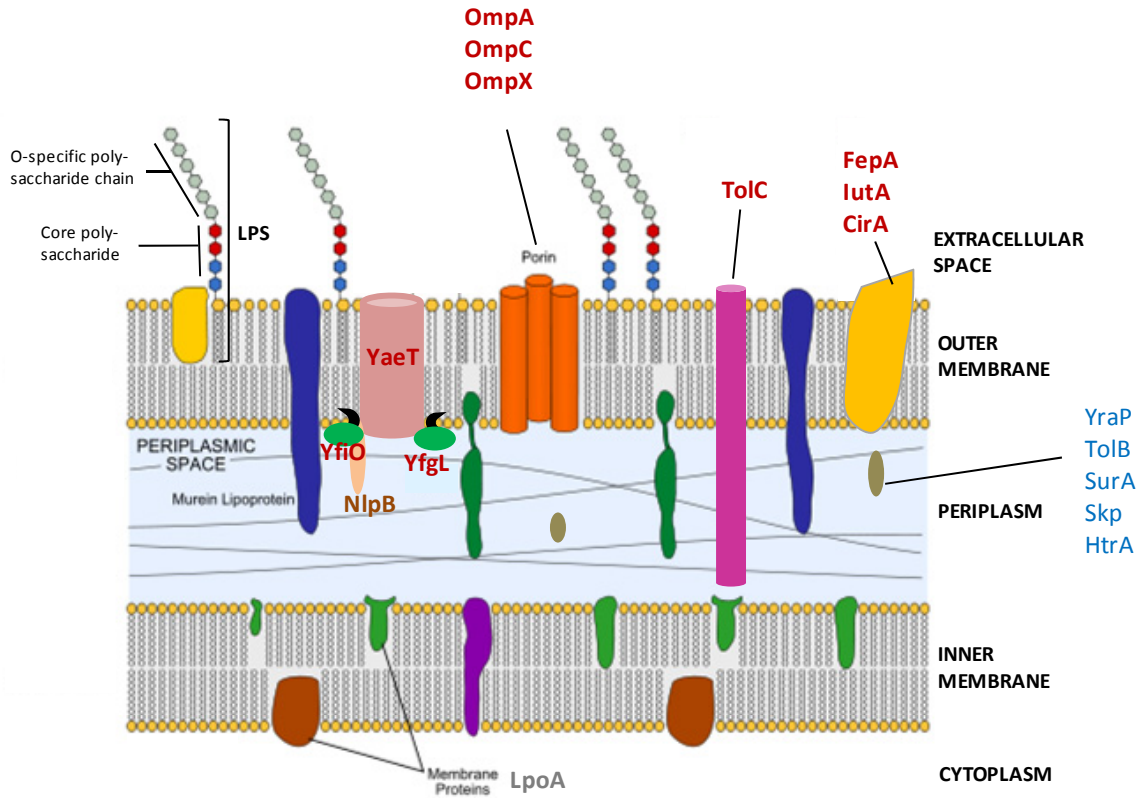


Figure 25. Schematic representation of the outer membrane of Gram-negative bacteria and main proteins identified in *Shigella* GMMA.
The bacterial compartments shown are: extracellular space, outer membrane (OM), periplasm, inner membrane (IM) and cytoplasm. *Shigella* GMMA and OM contain predominantly: porins as OmpA, OmpC and OmpX, trans-periplasmic channel proteins, such as Tsx and TolC interacting with IM proteins, β-barrel proteins such as YaeT, OM lipoproteins as NlpB, YfgL and YfiO interacting with YaeT. The main periplasmic proteins identified are YraP, TolB, SurA, Skp, HtrA. Only the IM protein LpoA showed a quite high amount in both types of GMMA.

5.4. IDENTIFICATION OF IMMUNOGENIC PROTEINS

Following identification and quantification of the protein composition of *Shigella* GMMA, our interest focused on finding out which of the GMMA-associated proteins were immunogenic. In the last sub-chapter, we characterize the immunogenic proteins in *S. sonnei* GMMA by (i) 2-D Western blot and (ii) immunoprecipitation of *Shigella* surface proteins using sera raised against *S. sonnei* GMMA followed by a quantitative analysis using stable isopote dimethyl labeling combined to GeLC-MS/MS.

5.4.1. Immunogenicity of GMMA

GMMA were generated from *tolR* and *tolR msbB S. sonnei* mutants in defined medium with high and low iron. As outlined before Δ*tolR* was grown at 37°C and Δ*tolR* Δ*msbB* at 30°C. In detail, the 4 types of GMMA and the conditions used for their production are:

- Ss -pSS Δ*tolR* grown in DM, low iron, 37°C (tolR, -Fe, 37C)
- Ss -pSS Δ*tolR* grown in DM, high iron, 37°C (tolR, +Fe, 37C)
- Ss -pSS Δ*tolR* Δ*msbB* grown in DM, low iron, 30°C (tolR/msbB, -Fe, 30C)
- Ss -pSS Δ*tolR* Δ*msbB* grown in DM, high iron, 30°C C (tolR/msbB, +Fe, 30C)

To assess the immunogenicity, GMMA were used to perform an immunogenicity study in mice (Sh03). Six weeks-old CD1 mice were immunized subcutaneously on day 1 and 22 with 2µg of GMMA. Antisera were collected on day 36, and GMMA-specific IgG levels were measured by ELISA. Mice immunized with GMMA showed vey high levels of total IgG. No differences were detectable between sera raised against GMMA with full or mutated LPS or sera against GMMA obtained in low and high iron concentration (figure 26). Control mice immunized with PBS and pre-immune serum had no detectable anti-GMMA antibodies.

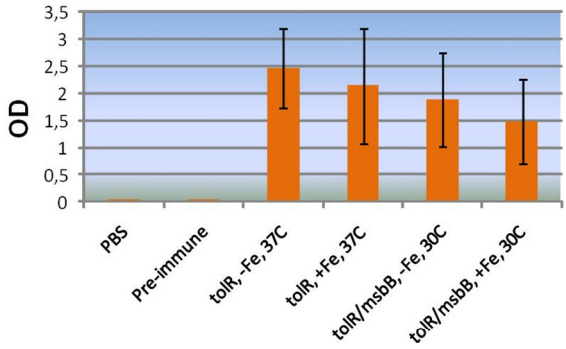


Figure 26. IgG response to different *S. sonnei* GMMA.
CD1 mice were immunized subcutaneously with 2ug of GMMA from (i) Ss -pSS Δ*tolR* grown in DM, low iron, 37°C, (ii) Ss -pSS Δ*tolR* grown in DM, high iron, 37°C, (iii) Ss -pSS Δ*tolR* Δ*msbB* grown in DM, low iron, 37°C, (iiii) Ss -pSS Δ*tolR* Δ*msbB* grown in DM, high iron, 37°C or (iv) PBS alone on day 1 and day 22. Sera were collected on day 36 and pooled sera were tested 1:10000 by ELISA using GMMA from Ss -pSS Δ*tolR* grown in DM with low iron at 37°C coating. Preimmune serum and PBS were used as negative control. '-Fe' and '+Fe' indicate low and high iron concentration, respectively.

5.4.2. Immunogenic proteins identified by 2-D gel Western blots

To investigate which proteins in GMMA raised IgG responses, bi-dimensional western blots were performed. As mentioned above, OmpA has previously been found to be detected by a strong signal in Western blots[11(submitted)] possibly obscuring other potential immunogenic proteins. Thus, GMMA lacking *ompA* were used for the Western blot analysis. They were separated by 2-D gel electrophoresis and probed with the different sera. Several proteins were detected by the sera and a subset of these was identified by mass fingerprint (figure 27). While the overall pattern of recognized proteins is similar for the different sera, specific differences could be observed. Comparing the 2-D WB with sera generated against GMMA from Ss –pSS Δ*tolR*, -Fe^{2+} vs GMMA from Ss –pSS Δ*tolR*, +Fe^{2+} a group of spots (1-7) were missing or less abundant using the serum raised against GMMA produced in high iron conditions, suggesting an iron-regulation of those proteins. The same group of proteins was also not detected in the other two WB with sera raised against Ss –pSS Δ*tolR* Δ*msbB*, low iron (-Fe^{2+}) and Ss –pSS Δ*tolR* Δ*msbB*, high iron (+Fe^{2+}) GMMA obtained at 30°C, independently from the iron concentration. This suggests that the proteins might additionally be regulated by temperature or show effects associated with *msbB* mutation. In paragraph 5.3.1.1 we explored the effects of the temperature and LPS mutation on the expression of GMMA proteins, and we observed that the Serine endoprotease (HtrA), indicated as spot 'M' in figure 13, showed a lower expression in the Ss –pSS Δ*tolR* Δ*msbB* than in the Ss –pSS Δ*tolR* GMMA. As GMMA from Ss-pSS Δ*tolR* and Ss-pSS Δ*tolR* Δ*msbB* mutants were generated under different temperatures and we did not perform a comparison of Ss –pSS Δ*tolR* at 37°C *vs* 37°C by 2-D gel, we cannot directly link the expression of HtrA to a temperature-dependent regulation or LPS modification. However, the temperature likely represents the key factor as already reported in the literature[146]. In the WB shown in figure 27, HtrA corresponds unequivocally to the spots 2, 4, 5, and 6 (indicating the presence of different isoforms of the protein) but it is also mixed to the 'putative receptor' in spots 1 and 3.

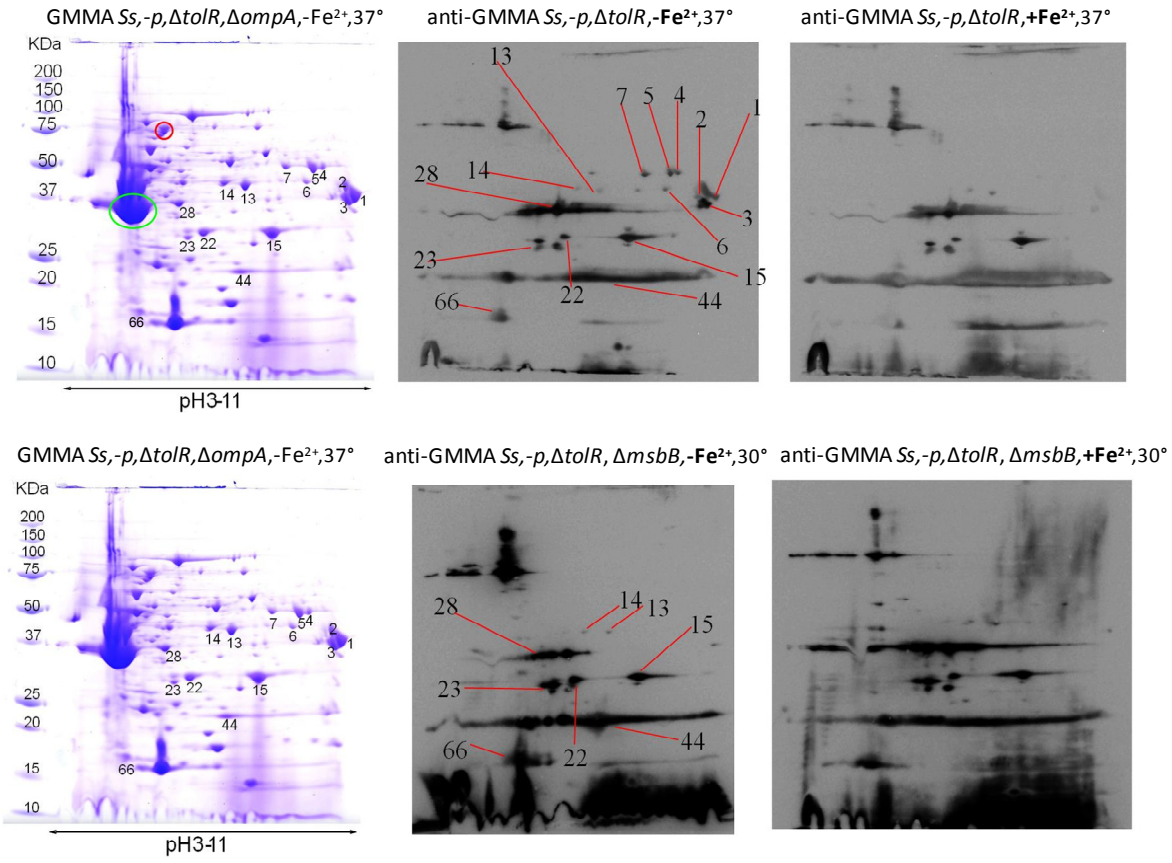


Figure 27. Comparitive 2-D Western blot analysis of sera raised against different GMMA.

GMMA separated by 2D SDS-PAGE: Ss -pSS Δ*tolR* Δ*ompA* GMMA obtained in iron limited conditions. 200 ug of these GMMA were used to visualize the proteins by Coomassie Blue staining and 20µg were used for Western blots. Sera were diluted 1:1000. Numbers refer to proteins identified by mass fingerprint and showed in Table 6. Red circle= iron regulated proteins FepA, FhuA and CirA. Green circle= OmpC.

The results of the WB are summarizes in table 10. The majority of the identified immunogenic proteins were predicted to be periplasmic and two of them (Serine endoprotease and FKBP-type peptidyl-prolyl cis-trans isomerase) were found also in a previous study[11]. So far[11] only few immunogenic outer membrane proteins were identified: OmpA, OmpX, and YaeT. The three iron-regulated proteins were not detected by WB. Interestingly, also the very abundant OmpC protein was not detected by 2-D WB (figure 27).

				ΔtolR (37°C)		ΔtolR ΔmsbB (30°C)	
n°	Name	Gene	PSORTb 2.0	+ Fe^{2+}	- Fe^{2+}	+ Fe^{2+}	- Fe^{2+}
1	serine endoprotease / putative receptor	htrA/SSON_1681	Peri/Ukn	?	x		
2, 4-6	**serine endoprotease**	**htrA**	Peri	?	x		
3	serine endoprotease / putative receptor	htrA/SSON_1681	Peri/Ukn	?	x		
7	glucan biosynthesis protein G	mdoG	Peri	?	x		
13	translocation protein TolB	tolB	Peri	x	x	x	x
14	peptidyl-prolyl cis-trans isomerase SurA	surA	Peri	x	x	x	x
15 (1)	outer membrane protein assembly complex subunit YfiO	yfiO	OM Lipo	x	x	x	x
15 (2)	**FKBP-type peptidyl-prolyl cis-trans isomerase**	**fkpA**	Peri	x	x	x	x
22	hypothetical protein SFV_2968	yggE	Ukn	x	x	x	x
23	tbd	/	?	x	x	x	x
28	glycerophosphodiester phosphodiesterase	glpQ	Peri	x	x	x	x
44	glycoprotein-polysaccharide metabolism	ybaY	Peri	x	x	x	x
66	peptidoglycan-associated outer membrane lipoprotein	pal	OM lipo	?	x	x	x

Table 10. List of immunogenic proteins identified by 2-D WB.
Numbers: immunogenic proteins highlighted in figure 27, "x": protein detected by strong signal in the WB. "?": protein detected by weak signal in the WB. Fields in pink: potentially Fe-regulated proteins. Tbd: to be determined, not identified protein. Protein names highlighted in bold: proteins found also in a previous study.

5.4.3. Immunogenic surface proteins: immunoprecipitation approach

As mentioned in chapter 5.2 we hypothesize that iron regulated proteins might contribute to protection. But so far we have not been able to detect antibodies against these proteins in sera of mice immunized with GMMA generated under iron-limiting conditions. Moreover only few OM proteins have been identified by 2-D WB. As the native conformation of the proteins might be essential for antibody binding, some of the proteins that are exposed on the surface might not be detectable by Western blot after SDS-PAGE. Therefore, we focused on a different method to identify immunogenic surface proteins. We chose an immunoprecipitation (IP) approach using whole bacteria incubated with the sera raised against GMMA. If the sera contain conformational antibodies against surface exposed proteins the corresponding proteins cannot be visualized after denaturing SDS-PAGE. To overcome this, native binding conditions would be required. Bacteria were incubated with the serum and unbound antibodies washed away. Then, the bacteria were lysed in conditions that should not affect antibody/antigen (Ab/Ag) binding. The Ab/Ag complexes were incubated with magnetic proteinG-micro-beads and the Ab/Ag/beads complexes were loaded onto magnetic micro-columns. Finally, the antigens and the antibodies were eluted from the columns and separated by SDS-PAGE and analyzed by mass spectrometry.

5.4.3.1. Initial set up of the IP approach

Two different lysis conditions (containing the same base of buffer but different detergents) were tested for Ss -pSS bacteria grown at low OD:

- Solubilization buffer 1 (containing 1% Triton X-100, lysis at 37°C)
- Solubilization buffer 2 (containing 1% CHAPS, lysis at 4°C)

As shown in table 11, solubilization buffer 1 resulted in lysis of 58% of the bacteria compared to the negative control (bacteria incubated with PBS), while the solubilization buffer 2 resulted in 99.5% lysis.

	Negative control				Sample			
Buffer	**Dilution**	**CFU On plate**	**CFU In sample**	**Average**	**CFU On plate**	**CFU In sample**	**Average**	**% lysis**
Sol. buffer 1	10^{-4}				layer			**58%**
	10^{-5}	450	$4.5x10^8$	$6.4x10^8$	247	$2.5x10^8$	$2.7x10^8$	
	10^{-6}	57	$5.7x10^8$		30	$3x10^8$		
	10^{-7}	9	$9x10^8$					
Sol. buffer 2	10^{-3}				13	$1x10^5$	$1x10^5$	**99.5%**
	10^{-4}				/			
	10^{-5}	226	$2.3x10^8$	$2.4x10^8$	/			
	10^{-6}	29	$2.9x10^8$		/			
	10^{-7}	2	$2x10^8$					

Table 11. Lysis test on Ss -pSS bacteria with different solubilization buffers.
Sol. Buffer 1: lysis buffer containing 1% Triton X-100, Sol. buffer 2: lysis buffer containing 1% CHAPS. $3x10^8$ bacteria were lysed in 500µL of sol. buffer 1 for 1h at 37°C or 1mL of sol. buffer 2 for 1h at 4°C. Serial dilutions of the 2 samples were prepared in 1mL and 1/10 of each dilution was plated on LB plates and incubated ON at 37°C.

Both of the lysis conditions were subsequently compared in the IP experiment using 10µl of a second batch of serum against GMMA from Ss –pSS Δ*tolR*, low iron, DM, 37°C (Sh07). To control for non-specific binding of proteins to the micro-beads, cell lysates (obtained with buffer 1 and buffer 2) were incubated with micro-beads without antibodies and processed in the same way as the samples with antiserum. As shown in the silver stained gel in figure 28, a similar low background was found in the two negative controls while many proteins were visible in both of the IP

experiments. All of the bands visible in the IP lanes were manually picked, trypsin-digested, and the peptides analyzed by MALDI-TOF MS. Proteins that were successfully identified are indicated in figure 28. The two major bands (at approximately 52KDa and 25KDa) correspond to the heavy and light chains of the antibodies chains. Three outer membrane proteins were identified: OmpA, OmpC and YaeT. As OmpC was not detected by 2D-WB, its detection in the IP approach suggests that the antibodies bind to native epitopes that are lost in the denaturing conditions of the SDS-PAGE. The proteins with the lowest molecular weights (<15KDa) were predominantly identified as ribosomal proteins and their presence may result from non-specific binding of these proteins to the micro-beads since they are also visible at low level in the negative control sample. Several high molecular weight (75-150KDa) bands were visible but not identified in IP experiment performed with buffer 2, but not in the IP performed with buffer 1. Considering these data and the results of the lysis test previously described, the solubilization buffer 2 was chosen to optimize the IP protocol, although the ribosomal proteins seemed to be more visible in the IP with buffer 2 than with buffer 1.

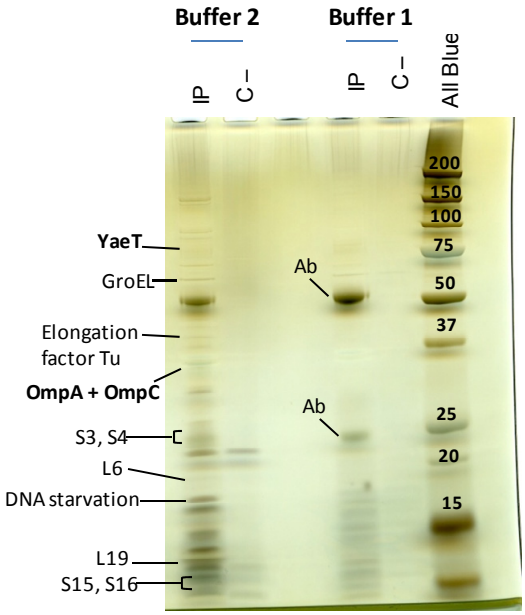


Figure 28. First immunoprecipitation experiments and negative controls.
IP experiments were performed using Ss –pSS bacteria and 10µl of serum against GMMA obtained from Ss –pSS Δ*tolR,* low iron, 37°C mutant (Sh07). Buffer 1: lysis buffer containing 1% Triton X-100. Buffer 2: lysis buffer containing 1% CHAPS. Negative controls (C) were obtained by performing the procedure without sera. Proteins obtained in IP samples and negative controls were separated by SDS-PAGE using a 12% acrylamide gel and visualized by silver staining. Proteins identified by MALDI-TOF MS are indicated. Ab: antibodies chains. S3, S4, S15, S16: 30S ribosomal proteins. L6, L19: 50S ribosomal proteins.

Several conditions in the IP protocol were subsequently tested (figure 29). For all experiments serum raised against GMMA from Ss -pSS Δ*tolR*, -Fe^{2+}, 37°C is used (in samples 10 and 11 serum batch 1 is used, while in all the others serum batch 2):

- two different amounts of serum (1µl and 10µl) *versus* pre-immune serum (negative control) to understand if the IP experiment was performed with saturating antibody concentrations (samples 1, 2, and 3);
- buffer 2 *versus* buffer 2 without EDTA plus RNase, DNase, and $MgCl_2$: to understand if the presence of the ribonucleases could decrease the unspecific binding of the ribosomal proteins to the micro-beads in the IP sample (samples 11 and 10);
- previous procedure *versus* method with additional ultra-centrifugation step of the supernatant (containing Ab/Ag complexes) after lysis: to test if ribosomal complexes can be removed (samples 4 and 7);
- high salt washing step (stringent wash) before elution from the micro-columns of samples in EDTA-free lysis buffer containing ribonucleases *versus* samples in buffer containing EDTA: evaluate if the presence of ribonucleases in the EDTA-free buffer allow to eliminate unspecific binding to the micro-beads when combined with a stringent wash (samples 5 and 6). Results from sample 5 will be indirectly compared to the ones obtained from sample 9 to evaluate the specific effect of the stringent wash;
- negative control (no serum) *versus* IP performed with serum in EDTA-free buffer: to evaluate the background of the experiment (unspecific binding to the micro-beads, sample 8 and 9). The result of sample 9 will be compared indirectly with the results of sample 10 to evaluate if different batch of sera recognize the same proteins.

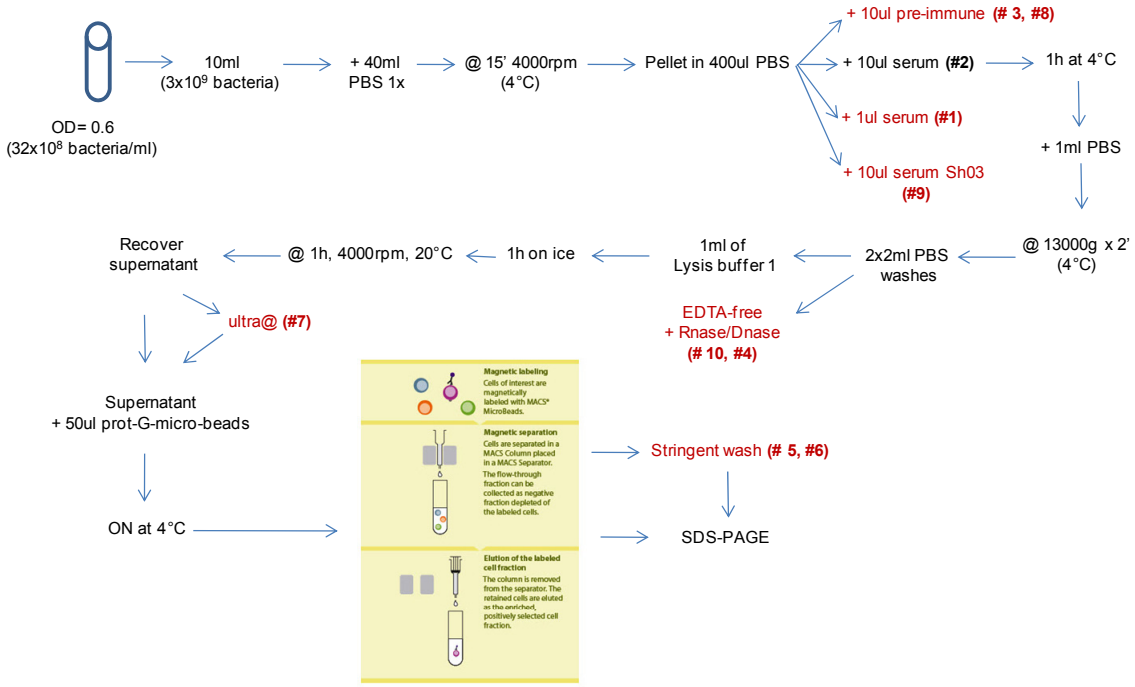
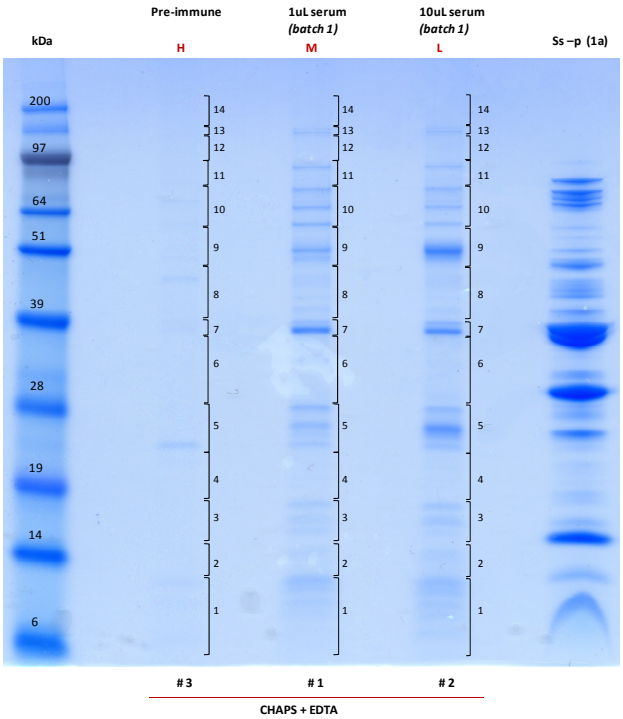


Figure 29. Schematic representation of the immunoprecipitation approach.
10mL of Ss –pSS bacteria grown at OD=0.6 were washed in PBS and collected by centrifgation. The pellet was resuspended in 400µl of PBS and incubated with 10µL of serum raised against Ss -pSS Δ*tolR*, low iron, DM, 37°C GMMA for 1h at 4°C. The unbound antibodies were removed by washing the bacteria with PBS. The bacteria were lysed for 1h at 4°C with 1ml of lysis buffer and subsequently centrifuged at low speed for 1h. The supernatant was incubated ON with proteinG-micro-beads and subsequently processed through the micro-columns according to the manufacturer's specifications. The eluted samples (containing antibodies and antigens) were separated by 12% SDS-PAGE. The samples highlighted in red indicate the different conditions tested in the IP protocol.

5.4.3.2. Quantitative proteomic analysis of IP samples using different amounts of antiserum in comparison to pre-immune serum (sample 1-3)

The quantitative proteomic approach based on stable isotope dimethyl-labeling, described in paragraph 5.3.2, was used to analyze the IP samples. Here, we present the results obtained from the IP experiment performed with two different amount of serum (1µl and 10µl) and with pre-immune serum (negative control). The focus of this comparison was to evaluate if the IP experiment was performed in saturating conditions of antibodies binding. The eluted samples 1 (1µl serum), 2 (10µl serum), and 3 (pre-immune serum) were separated by SDS-PAGE (figure 30), each lane was cut into slices, and each slice was trypsin-digested. The peptides extracted from each

slice were labeled with 'intermediate', 'light', and 'heavy' labeling respectively. The 'intermediate', 'light', and 'heavy' labeled peptides from slices with the same number were mixed together and processed through the LTQ-Orbitrap Velos. The data were subsequently analyzed by Andromeda search engine against protein databases of *S. sonnei*, *S. flexneri*, *S. dysenteriae,* and *S. boydii*, and processed by MaxQuant software (as described in paragraph 5.3.2).



Bacteria: *Ss, -p,* 37°C, LB
Serum: anti *Ss, -p*, Δ*tolR*, low iron, 37°C, DM GMMA

Figure 30. Immunoprecipitate using whole Ss -pSS bacteria with sera raised against GMMA.
Immunoprecipitate using 1µL (#1) and 10µL (#2) of serum against GMMA from Ss -pSS Δ*tolR*, low iron, DM, 37°C (batch1=Sh07), and 10µl of pre-immune serum (#3) as negative control. The lysis condition of all the samples is buffer 2 indicated as 'CHAPS + EDTA'. 5µg of GMMA obtained from Ss -pSS Δ*tolR* low iron, DM, 37°C indicated as 'Ss–p (1a)' (already shown in figure 15) were loaded as a reference for the pattern of proteins already identified. Samples were labeled with Heavy (H), Intermediate (M) and Light (L) isotopes and processed together.

As shown in figure 30 similar protein patterns were observed in the IP experiments using 1µL or 10µL of serum and most of the protein bands showed comparable intensities. The antibodies chains were, instead, detected in different amounts as expected. This result suggested that 1µL of serum was sufficient to reach saturating antibody binding to most of the reactive proteins. In the negative control few bands

were visible on the Coomassie stained gel suggesting non-specific binding of proteins to the micro-beads.

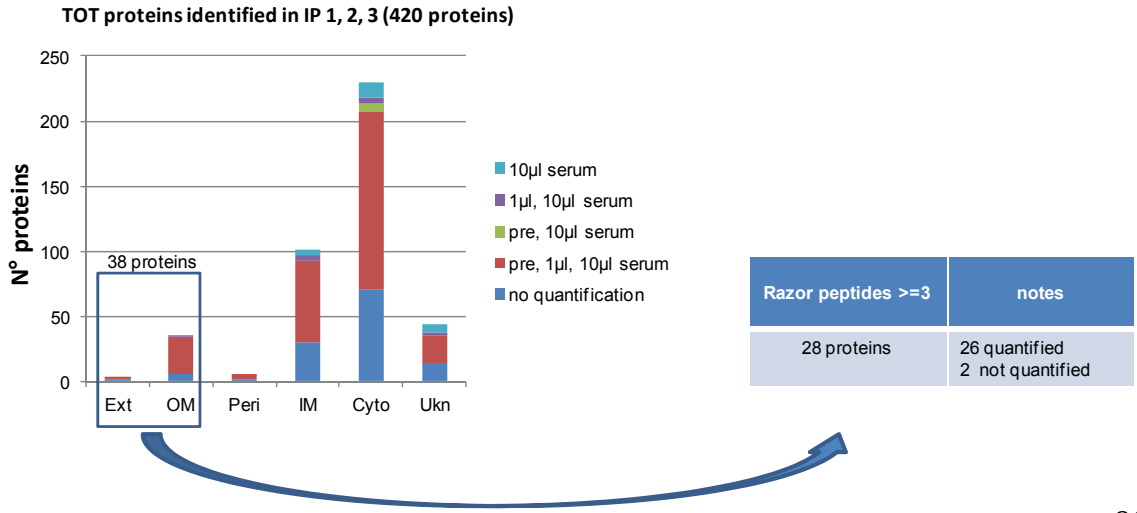
The original 'Protein Group' table produced by MaxQuant was manually analyzed to eliminate 'reverse', 'contaminants' and protein groups with a PEP>0.01, as already described. A total of 420 Protein Groups was identified in the combined samples 1, 2, and 3 (figure 31A). According to PSORTb 3.0 prediction, 3 proteins were predicted extracellular and 35 outer membrane. In addition, 6 periplasmic proteins were identified as Chaperone SurA (*surA*), Serine protease Do (*htrA*), Putative periplasmic protein (*yraP*), Putative uncharacterized protein (SS53G_0584), Formate dehydrogenases alfa subunit (SS53G_2246), Putative periplasmic binding transport protein (*ybeJ*). According to their function (outer membrane protein transit and folding, integrity of cell envelope, and biogenesis of β-barrel outer membrane proteins) the first 3 proteins might be expected to be found by this approach as they could complex with immunogenic OM proteins. For the remaining 3 proteins no evidence of OM protein interaction is available. Surprisingly, 101 inner membrane proteins, 230 cytoplasmic proteins and 45 unknown proteins, which were not expected to be precipitated by this approach were also recovered. The majority of these proteins were present in all the 3 samples, indicating the presence of unspecific binding to the micro-beads used to recover the antibody/antigen complexes. To analyze the results in more details we focused our attention on the 38 identified OM and extracellular proteins. To ensure that a protein was correctly called present in the sample we applied the cutoff of at least 3 identified razor peptides and 1 unique peptide as previously discussed. The final list contained 28 Protein Groups of which 26 were quantified (figure 31B). Unexpectedly, these proteins were found in all of the samples, including the negative control.

In order to analyze if the proteins were enriched in the IP samples compared to the negative control we considered the total intensities of the 'unique peptides quantified' for each protein in sample 1, sample 2 and sample 3. The 'SumRatio M/H' and the 'SumRatio L/H' values represent the fold increase of the protein amount in the IP performed with the serum in comparison to the negative control. These values represent the ratio between the total intensities of the 'unique peptides quantified' in the 'intermediate' and 'heavy' samples, or in the 'light' and 'heavy' samples, respectively. A 'SumRatio' value >1 indicates that the protein is more abundant in

the intermediate (1µl) or light samples (10µl) than in the heavy sample (control). A 'SumRatio' <1 indicates, instead, that the protein is predominantly quantified in the negative control. About half of the proteins showed a fold increase in the 'SumRatio' between 10 and 28 in the IP performed with the serum compared to the negative control (e.g. OmpA and OmpC). Interestingly, OmpC and the iron regulated proteins FepA, FhuA, and IutA were identified as immunogenic by this approach in contrast to the observation in the 2-D WB approach, suggesting that these proteins are immunogenic and the antibodies recognize native epitopes that are lost in the denaturing conditions of 2-D WB. Additional iron regulated proteins were identified as immunogenic in this approach: FecA and CjrC with 'SumRatios' between 3 and 10, and FhuA and CirA with 'SumRatios' only slightly higher than 1, suggesting a poor or absent immunogenicity of these proteins. The OM proteins OmpX and YaeT, already shown to be immunogenic by 2-D WB, were also identified. One extracellular protein (RnfC) and a protein classified as potential extracellular or OM (SigA) were identified with 'SumRatios' of approximately 16 and 2 respectively.

Although from a first analysis all the proteins were identified also in the negative control, the quantitative analysis based on the isotope labeling ('SumRatio') showed that when the serum is present there is an increase in the amount of proteins detected compared with the negative control (figure 31B). Thus, this suggests a specific binding of the proteins to the antibodies. Moreover, we think that a 'SumRatio' >=5 likely suggest a specific antibody/antigen interaction and, therefore, that 19 of these 28 proteins are immunogenic surface proteins.

A



B

Protein Group ID	Protein Descriptions	PSORTb localization	Common annotation	SUM RATIO M/H	SUM RATIO L/H
139	**Electron transport complex protein rnfC**	**Ext**	**rnfC**	**17,23**	**14,34**
245	Serine protease	Ext/OM	sigA	2,50	1,97
282	**Putative outer membrane protein**	**OM**	**yiaD**	**28,09**	**25,05**
218	**Outer membrane protein C**	**OM**	**ompC**	**22,66**	**17,15**
209	**NlpB/DapX lipofamily protein**	**OM**	**nlpB**	**22,12**	**19,26**
291	**Surface antigen family protein**	**OM**	**ytfM**	**20,34**	**27,44**
333	**Outer membrane protein A**	**OM**	**ompA**	**16,79**	**16,85**
206	**Outer membrane assembly lipoprotein YfgL**	**OM**	**yfgL**	**15,19**	**22,63**
34	**Maltoporin**	**OM**	**lamB**	**11,87**	**11,88**
312	**Iron(III) dicitrate transport protein fecA**	**OM**	**fecA**	**10,57**	**7,58**
456	**Ferrienterobactin receptor**	**OM**	**fepA**	**10,40**	**40,00**
341	**Putative polysaccharide export protein gfcE**	**OM**	**yccZ**	**9,00**	**5,47**
214	**Outer membrane porin protein LC**	**OM**	**nmpC**	**8,74**	**21,23**
170	**Outer membrane protein assembly factor yaeT**	**OM**	**yaeT**	**7,18**	**5,64**
241	**Putative uncharacterized protein**	**OM**	**tolC**	**5,69**	**4,51**
382	**Ferric aerobactin receptor**	**OM**	**iutA**	**4,04**	**6,24**
196	**TonB-dependent Receptor Plug domain protein**	**OM**	**cjrC**	**3,42**	**5,23**
183	LPS-assembly protein lptD	OM	lptD	2,52	1,70
253	Outer membrane assembly lipoprotein YfiO	OM	yfiO	2,36	3,60
326	Ferrichrome-iron receptor	OM	fhuA	1,40	3,10
448	Colicin I receptor	OM	cirA	1,21	1,12
225	Long-chain fatty acid outer membrane transporter	OM	fadL	0,96	0,62
160	LPS-assembly lipoprotein lptE	OM	lptE	N/A	N/A
199	**Outer membrane lipoprotein pcp**	**OM**	**slyB**	**19,75**	**20,54**
342	**Putative uncharacterized protein**	**OM**	**ymcA**	**14,04**	**11,91**
430	**Outer membrane protein X**	**OM**	**ompX**	**8,12**	**5,82**
201	Repeated sequence found in lipoLPP family protein	OM	lpp	1,02	1,89
43	Vitamin B12 transporter BtuB	OM	btuB	N/A	N/A

Figure 31. Extracellular and outer membrane proteins identified in IP 1-3
A. Total proteins identified in IP 1-3 and OM/Ext proteins identified with >=3 'razor peptides'. **B.** Details of the 28 OM/Ext proteins identified. The 'Protein Group ID' indicates the progressive number associated to each protein group in the original table generated from MaxQuant. The 'protein descriptions' lists the SHISO annotation of each protein group. PSORTb 3.0 predicted localization is shown for each protein. The Serine protease is classified as 'unknown' but it shows 50% of probability to be extracellular and 50% to be outer membrane protein. For each protein are given: annotation in the original strain *Ss53G*, common annotation in other *Shigella* strains, and number of unique peptides quantified. The "SumRatio" columns show the abundance of each protein (in fold increase) in the IP performed with serum compared with the negative control. M= IP performed with 1µl of serum. L= IP performed with 10µl of serum. H= IP performed with pre-immune serum (negative control). The quantification of each protein in the 3 samples is based on the measure of total unique peptides intensities. In bold are shown the proteins having a SumRatio values >=5, that likely suggest a specific antibody/antigen interaction. N/A= not available. The first 23 proteins were identified also in the second group of IP experiments (IP 4-11) presented in the next paragraph.

Comparing the 'SumRatio M/H' and 'SumRatio L/H' of each protein listed in figure 31B, we observed quite similar values for the majority of the proteins, suggesting that 1µl of serum is sufficient to work in saturating conditions for most proteins.

However, for the proteins FepA and NmpC we observed a significant increase of the 'SumRatio' value when using 10µl of serum compared with 1µl. These results might be explained with a low amount of anti-FepA and anti-NmpC antibodies in the serum, suggesting that more serum is needed to saturate the binding to all the molecules present on *Shigella*.

To evaluate how abundant the immunogenic proteins were in *S. sonnei* GMMA, we compared the immunogenicity results with the protein quantification in GMMA. As the 'Protein Group IDs' are different in the 2 experiments, we linked the results using the gene names corresponding to the proteins. Most of the immunogenic proteins were found in *S. sonnei* GMMA as expected. Surprisingly, the predicted extracellular protein RnfC was not present in the list of identified proteins in GMMA. Since different batches of GMMA were used for the proteomic analysis and the immunization study, RfnC might have been associated to the GMMA used for the mice immunization, allowing the production of anti-RnfC antibodies, but might have been absent in the batch of *S. sonnei* GMMA used for the complete proteome analysis. Another explanation might be that this protein is very low abundant in GMMA but highly immunogenic. Maybe if RnfC has a very immunogenic structure or large protein surface exposed on the external membrane, it might be enriched by the presence of antibodies. Each protein identified in the IP experiment was correlated to its localization in *S. sonnei* GMMA and its relative emPAI values observed (table 12). The majority of the immunogenic proteins were identified among the most abundant proteins in GMMA, and 11 of them (OmpC, OmpA, NmpC, Lpp, OmpX, CirA, NlpB, YaeT, TolC, YfiO, FhuA) were found within the first 30 positions of the *S. sonnei* emPAI graph (paragraph 5.3.3.1). The proteins YiaD, YtfM, and FecA were, instead, not very abundant in GMMA, but were shown to be very immunogenic according the 'SumRatio' values. All the iron-regulated proteins identified in GMMA were also present in the final list of the immunogenic proteins: they were quantified with a high number of peptides, and according to the 'SumRatio' values FecA, FepA, IutA, and CjrC showed a significant difference between the control and the IP sample. These results suggested a specific interaction of the proteins with the antibodies.

Protein Group ID	Protein Descriptions	PSORTb localization	Common annotation	SUM RATIO M/H	SUM RATIO L/H	ID GMMA	n° protein in GMMA	emPAI GMMA
139	**Electron transport complex protein rnfC**	**Ext**	**rnfC**	**17,23**	**14,34**	N/A	N/A	N/A
245	Serine protease eatA	Ext/OM	sigA	2,50	1,97	424	161	0,6
282	**Putative outer membrane protein**	**OM**	**yiaD**	**28,09**	**25,05**	338	110	1,37
218	**Outer membrane protein C**	**OM**	**ompC**	**22,66**	**17,15**	421	2	11,45
209	**NlpB/DapX lipofamily protein**	**OM**	**nlpB**	**22,12**	**19,26**	100	8	7,86
291	**Surface antigen family protein**	**OM**	**ytfM**	**20,34**	**27,44**	270	134	0,85
333	**Outer membrane protein A**	**OM**	**ompA**	**16,79**	**16,85**	225	5	9,00
206	**Outer membrane assembly lipoprotein YfgL**	**OM**	**yfgL**	**15,19**	**22,63**	387	36	**4,46**
34	**Maltoporin**	**OM**	**lamB**	**11,87**	**11,88**	176	54	2,98
312	**Iron(III) dicitrate transport protein fecA**	**OM**	**fecA**	**10,57**	**7,58**	434	104	1,48
456	**Ferrienterobactin receptor**	**OM**	**fepA**	**10,40**	**40,00**	478	N/A	N/A
341	**Putative polysaccharide export protein gfcE**	**OM**	**yccZ**	**9,00**	**5,47**	108	28	**4,99**
214	**Outer membrane porin protein LC**	**OM**	**nmpC**	**8,74**	**21,23**	419	7	9,00
170	**Outer membrane protein assembly factor yaeT**	**OM**	**yaeT**	**7,18**	**5,64**	84	19	**5,58**
241	**Putative uncharacterized protein**	**OM**	**tolC**	**5,69**	**4,51**	408	22	**5,31**
382	**Ferric aerobactin receptor**	**OM**	**iutA**	**4,04**	**6,24**	443	N/A	N/A
196	**TonB-dependent Receptor Plug domain protein**	**OM**	**cjrC**	**3,42**	**5,23**	415	37	**4,46**
183	LPS-assembly protein lptD	OM	lptD	2,52	1,70	194	48	**3,27**
253	Outer membrane assembly lipoprotein YfiO	OM	yfiO	2,36	3,60	392	21	**5,49**
326	Ferrichrome-iron receptor	OM	fhuA	1,40	3,10	437	24	**5,31**
448	Colicin I receptor	OM	cirA	1,21	1,12	475	11	7,06
225	Long-chain fatty acid outer membrane transporter	OM	fadL	0,96	0,62	165	71	2,38
160	LPS-assembly lipoprotein lptE	OM	lptE	N/A	N/A	19	39	**4,01**
199	**Outer membrane lipoprotein pcp**	**OM**	**slyB**	**19,75**	**20,54**	101	42	**3,64**
342	**Putative uncharacterized protein**	**OM**	**ymcA**	**14,04**	**11,91**	438	73	2,27
430	**Outer membrane protein X**	**OM**	**ompX**	**8,12**	**5,82**	105	12	6,50
201	Repeated sequence found in lipoLPP family protein	OM	lpp	1,02	1,89	173	4	9,00
43	Vitamin B12 transporter BtuB	OM	btuB	N/A	N/A	178	123	0,97

Table 12. Abundance of the immunogenic extracellular and outer membrane proteins identified in IP 1-3 in *S. sonnei* GMMA.
The immunogenic proteins identified in the immunoprecipitation experiments 1-3 were linked to the information obtained in the *S. sonnei* GMMA proteome analysis. In the left part of the table is shown: 'Protein Group ID' in IP 1-3, PSORTb 3.0 predicted localization, and 'SumRatio' values showing the abundance of each protein (in terms of fold increase) in the IP performed with serum compared with the negative control. In bold are shown the proteins with a 'SumRatio' >=5 that likely suggest a specific antibody/antigen interaction. Proteins highlighted in pink: proteins identified in all the IP performed with serum of the group 4-11. In the right part of the table are shown the information related to the *S. sonnei* GMMA proteome analysis. For each protein is indicated the respective 'Protein Group ID', the position in the list of abundant proteins, and the calculated emPAI value in GMMA. Red= emPAI>6; **bold**= emPAI>3. 11 of the likely immunogenic proteins are among the most abundant proteins in GMMA, while 3 show a low abundance indicated by an emPAI <1 or roughly equal to 1.

5.4.3.3. Quantitative proteomic analysis of IP 4-11

The immunoprecipitated samples 4-11, already described in paragraph 5.4.3.1 and summarized in table 13, were analyzed by SDS-PAGE and labeled with "intermediate" and "heavy" isotopes (figure 32).

Table 13: lysis conditions and serum (10µL) used in IP 4-11 samples

IP sample	Labeling	EDTA	RNAse/ DNAse/ $MgCl_2$	Batch of serum	High salt washing step	ultra@ after lysis
4	M	-	+	2	-	-
7	H	-	+	2	-	+
6	M	+	-	2	+	-
5	H	-	+	2	+	-
9	M	-	+	2	-	-
8	H	-	+	-	-	-
10	M	-	+	1	-	-
11	H	+	-	1	-	-

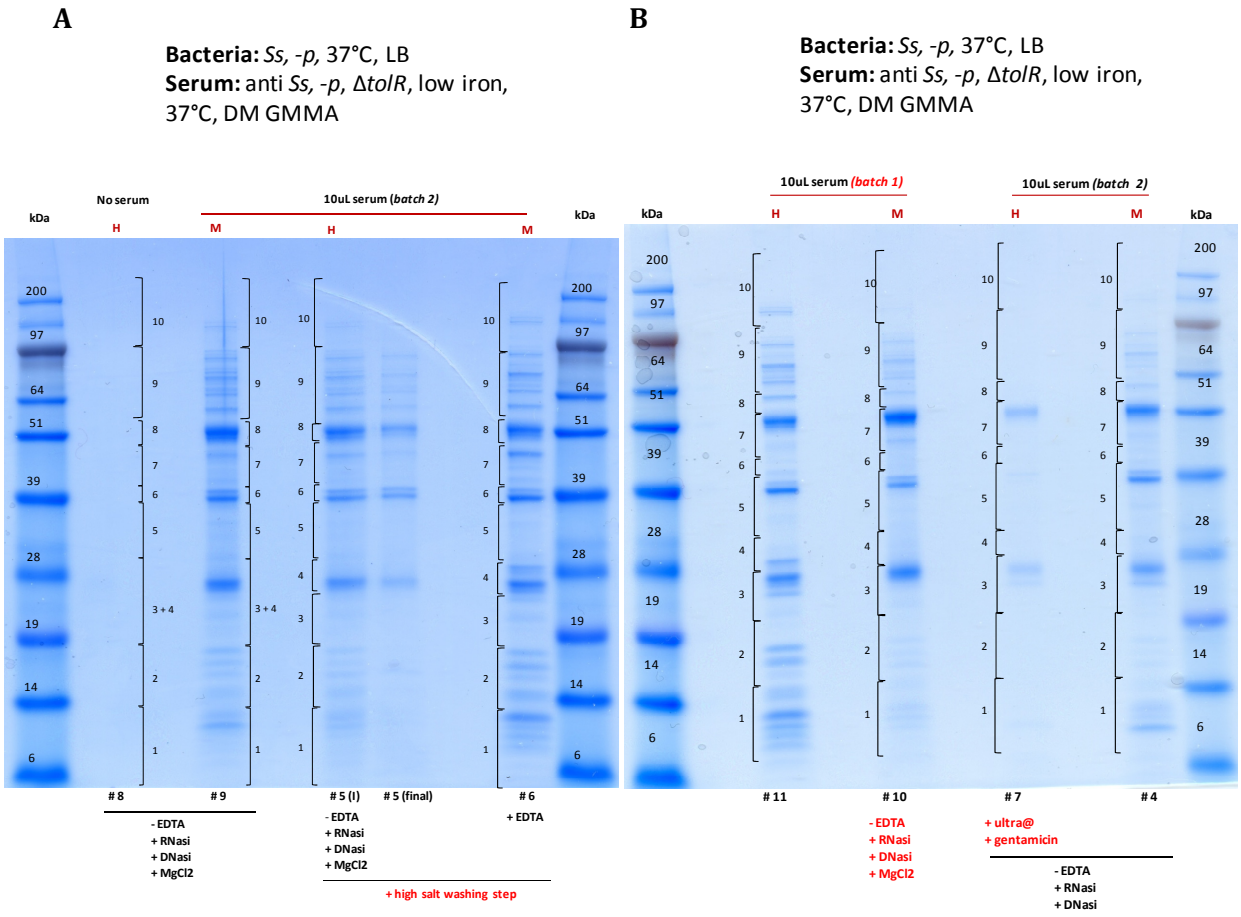


Figure 32. Immunoprecipitate using whole Ss -pSS bacteria with sera raised against GMMA.

In all the IP 10µL of serum were used. The lysis conditions of each IP are summarized in table 13. Serum batch 1 (Sh03): serum raised against Ss –pSS Δ*tolR*, DM, low iron, 37°C GMMA. Serum batch 2 (Sh07): serum raised against Ss –pSS Δ*tolR*, DM, low iron, 37°C GMMA obtained in a second immunization study. **A.** Tryptic peptides from sample 8 (negative control) and 9 (IP) were labeled with 'heavy' (H) and 'intermediate' (M) isotopes and subsequently processed together. Tryptic peptides from IP samples 5 (I) and 6 were labeled with 'H' and 'M' isotopes and subsequently processed together. Sample 5 (I) was incorrectly eluted from the column before the last washing step (low salt), and it should have probably been combined with sample 5 (final). However, only sample 5 (I) was processed and compared with sample 6. **B.** Tryptic peptides from sample 7 and 4 were labeled with 'heavy' (H) and 'intermediate' (M) isotopes respectively, and subsequently processed together. After the lysis step, in sample 7 was added gentamicin (to ensure bacterial death) and the sample

was processed by ultra-centrifugation (ultra@) in order to separate antigen/antibodies complexes from membrane pieces and hypothetical ribosomal complexes. Samples 4 and 7 were labeled with 'H' and 'M' isotopes respectively and processed together.

The LTQ-Orbitrap data obtained from each comparison were analyzed as previously described. The data obtained from the 4 different comparisons were then processed all together in a single MaxQuant experiment. In this way, all the proteins identified in each comparison were classified according to the same 'Protein Group IDs', facilitating the cross-reference of the proteins among the different samples. The original 'Protein Group' table produced by MaxQuant was manually analyzed to eliminate 'reverse', 'contaminants' and protein groups with a PEP>0.01, as described in paragraph 5.3.2.. 731 proteins were identified with at least 1 razor peptide. As mentioned before, the identification or at least 3 razor peptides and 1 'unique' peptide were needed to call a protein present. After applying a razor peptides >=3 filter 335 Protein Groups remained including quantified and not quantified proteins (figure 33A). To calculate the 'quantified' proteins for each couple of comparison we considered the 'unique peptides total intensity' obtained summarizing the 'unique peptide intensity H' and the 'unique peptide intensity M' of each pair. Whenever the total intensity was higher than 0, the protein was considered quantified within the comparison, while no intensity values meant absence of quantification. We manually checked how many proteins were identified in the different comparisons shown in figure 32. In the obtained 'Protein Group' table each comparison (4-7, 5-6, 8-9, and 10-11) is considered as a single experiment, therefore, the number of 'razor' and 'unique' peptides of each protein group refers to the comparisons (sum of the 2 samples) and not to the individual samples. Whenever the 'razor' or 'razor' and 'unique' peptides were equal to 0, we considered that the protein was not identified in that particular comparison. As shown in figure 33B, 29 OM and extracellular proteins were identified in total in the IP 4-11 experiments, and 23 of them were present in all of the comparisons. Similar to IP 1-3, we detected 88 cytoplasmic proteins, 19 inner membrane proteins, 2 periplasmic and 8 unknown proteins in all the IP 4-11 comparisons.

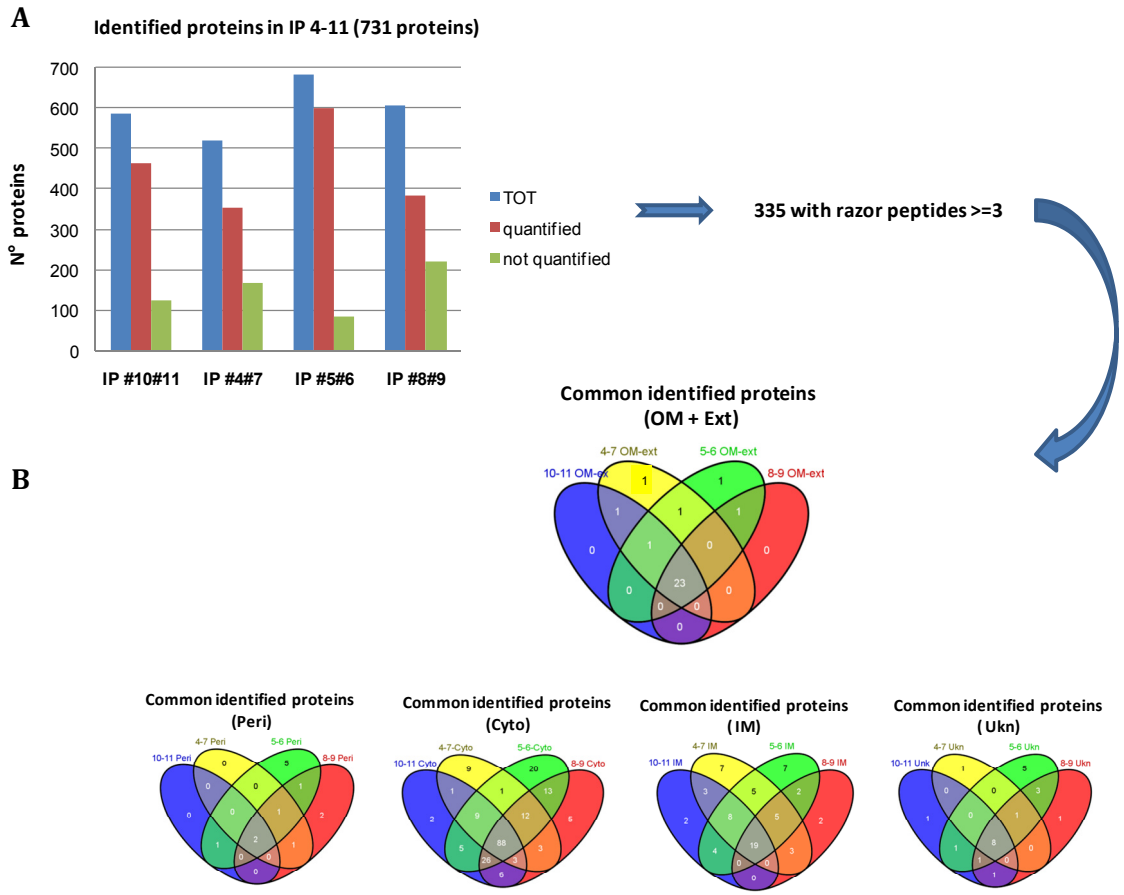


Figure 33. Identified proteins in the immunoprecipitation experiments 4-7, 5-6, 8-9, and 10-11.
A. 731 proteins in total were identified from the IP samples 4-11. A filter of razor peptides >= 3 was applied and the final list was shortened to 335 proteins. **B.** PSORTb 3.0 localization of proteins present in all IP 4-11: OM = 21, Ext = 2, Peri = 2, Cyto = 88, IM = 19, Ukn = 8.

The list of the 29 OM and extracellular proteins identified is shown in table 14. Among them, 23 proteins were found in all the IP 4-11 comparisons confirming the majority of the immunogenic proteins already identified in IP 2-3 (top 23 proteins figure 31B). In particular, all the identified iron regulated proteins and the major immunogenic OM proteins identified in the IP 2-3 were confirmed. The 6 proteins only identified in some of the IP 4-11 comparisons were: YmcA, OmpX, Lpp, BtuB, SSON_0127, and YddB. In particular, YmcA was found in all the samples except 8 (no serum) and 9 and it was also identified in IP 1-3, but only 1 out of 4 peptide was quantified. BtuB was found only in sample 5 and 6 (in which the lysis was performed without and with EDTA, respectively, and characterized by the addition of high salt washing step) and in IP 1-3 no quantification was available. Lpp was found in samples 5 and 6 and IP 1-3. OmpX was found in samples 4, 5, 6, and 7. SSON_0127

was only found in sample 4 and 10 (in which ribonucleases were present), but it was recognized by both the batches of sera. YddB was found only in sample 4, while SlyB was only found in IP 1-3.

Protein Group ID IP 4-11	Protein Descriptions	PSORTb	*Ss53G* annotation	common annotation
186	Electron transport complex protein rnfC	Ext	SS53G_0355	rnfC
553	Serine protease eatA	Ext/OM	SS53G_1669	sigA
395	Vitamin B12 transporter BtuB	OM	btuB	btuB
770	Colicin I receptor	OM	SS53G_1033	cirA
720	TonB-dependent Receptor Plug domain protein	OM	SS53G_0290	cjrC
375	Long-chain fatty acid outer membrane transporter	OM	SS53G_0810	fadL
738	Iron(III) dicitrate transport protein fecA	OM	SS53G_3463	fecA
772	Ferrienterobactin receptor	OM	SS53G_5666	fepA
740	Ferrichrome-iron receptor	OM	SS53G_3761	fhuA
542	Ferric aerobactin receptor	OM	SS53G_5072	iutA
393	Maltoporin	OM	lamB	lamB
386	Repeated sequence found in lipoLPP family protein	OM	SS53G_0415	lpp
422	LPS-assembly protein lptD	OM	lptD	lptD
263	NlpB/DapX lipofamily protein	OM	SS53G_0626	nlpB
723	Outer membrane porin protein LC	OM	SS53G_0760	nmpC
475	Outer membrane protein A	OM	SS53G_4120	ompA
725	Outer membrane protein C	OM	SS53G_0957	ompC
267	Outer membrane protein X	OM	SS53G_3941	ompX
548	Outer membrane assembly lipoprotein YfgL	OM	yfgL	SSON_2594
729	Putative uncharacterized protein	OM	SS53G_1603	tolC
222	Outer membrane protein assembly factor yaeT	OM	yaeT	yaeT
268	Putative polysaccharide export protein gfcE	OM	SS53G_4155	yccZ
749	TonB-dependent Receptor Plug domain protein	OM	SS53G_5268	yddB
561	Outer membrane assembly lipoprotein YfiO	OM	yfiO	yfiO
679	Putative outer membrane protein	OM	SS53G_2796	yiaD
741	Putative uncharacterized protein	OM	SS53G_4156	ymcA
575	Surface antigen family protein	OM	SS53G_3045	ytfM
52	LPS-assembly lipoprotein lptE	OM	lptE	
609	Putative tonB-dependent receptor yncD	OM	SS53G_0127	

Table 14. Extracellular and outer membrane proteins identified in IP 4-11
'Protein Group ID', PSORTB 3.0 localization and gene annotation is given for each protein. 23 proteins, including the iron-regulated proteins, were confirmed also in IP 2-3 samples. The remaining 6 proteins (YmcA, OmpX, Lpp, BtuB, SSON_0127, and YddB) were identified only in some comparisons of the IP group 4-11 and in IP 2-3.

This type of experiments, in which different comparisons of samples are processed in a single MaxQuant experiment, generates a complex dataset to be analyzed. The majority of the information contained in the 'Protein Group' table refers to the comparisons and not to the individual samples. The only information related to the individual samples are the 'number of unique peptides quantified in H' and 'number

of unique peptides quantified in M' for each Protein Group. To understand in which of the samples the 29 immunogenic OM and extracellular proteins were identified we compared the numbers of quantified peptides for each Protein Group. Moreover, we combined these results with the results obtained from the IP 1-3 (table 15). For Protein Groups that were not identified in all of the IP this analysis was used to evaluate which conditions allowed the identification of a specific protein.

OM proteind identified in IP 4-11			unique peptides quantified														IP 1-3	unique peptides quantified							
			serum batch 1			serum batch 2							no serum					serum batch 2							
			10 M	11 H		4 M		7 H	6 M	5 H	9 M		8 H					H	M	L					SumRatio
Protein Group ID IP 4-11	gene	PSORTb localization	- EDTA + Rnasi/Dnasi +MgCl2	+ EDTA		- EDTA + Rnasi/Dnasi		- EDTA + Rnasi/Dnasi + Ultra@	+ EDTA + high salt wash	- EDTA + Rnasi/Dnasi + MgCl2 + high salt wash	- EDTA + Rnasi/Dnasi +MgCl2 10UL SERUM		- EDTA + Rnasi/Dnasi +MgCl2 NO SERUM	SumRatio H/M #8H#9M	Protein Group ID IP 1-3	+ EDTA pre-immune	+ EDTA 1uL	+ EDTA 10uL	M/H	L/H			
186	rnfC	Ext	5	5		3		0	3	3	3		1	**0,58**	139	2	3	4	17,23	14,34					
553	slgA	Ext/OM	9	9		7		4	10	10	5		5	**0,82**	245	13	14	18	2,50	1,97					
772	fepA	OM	21	21		8		2	17	17	2		0		456	5	5	5	10,40	40,00					
575	ytfM	OM	6	6		1		0	2	2	1		0		291	1	1	1	20,34	27,44					
679	yiaD	OM	8	8		6		0	12	12	6		0		282	5	6	6	28,09	25,05					
548	yfgL	OM	7	6		11		4	10	10	10		0		206	4	4	4	15,19	22,63					
723	nmpC	OM	8	8		6		4	10	10	1		1	1,35	214	3	3	4	8,74	21,23					
263	nlpB	OM	13	12		8		0	21	21	9		0		209	4	4	4	22,12	19,26					
725	ompC	OM	8	8		14		3	17	13	3		1	**0,06**	218	10	10	15	22,66	17,15					
475	ompA	OM	22	21		20		3	26	23	13		0		333	81	85	94	16,79	16,85					
393	lamB	OM	8	8		6		1	10	9	5		0		34	6	6	5	11,87	11,88					
738	fecA	OM	3	3		2		1	4	3	5		0		312	17	20	20	10,57	7,58					
542	iutA	OM	10	9		15		1	8	6	3		0		382	9	9	13	4,04	6,24					
222	yaeT	OM	34	34		30		9	22	20	8		0		170	9	12	15	7,18	5,64					
268	yccZ	OM	6	5		4		0	3	2	5		1	**0,46**	341	3	3	6	9,00	5,47					
720	cjrC	OM	9	9		2		0	6	6	1		0		196	2	2	3	3,42	5,23					
729	tolC	OM	12	12		3		2	15	15	1		1	0,71	241	4	4	4	5,69	4,51					
561	yrfO	OM	2	2		1		1	2	2					253	5	5	8	2,36	3,60					
740	fhuA	OM	10	9		6		1	4	4	1		0		326	3	3	3	1,40	3,10					
422	lptD	OM	14	14		4		1	8	8	0		0		183	3	3	4	2,52	1,70					
770	clrA	OM	25	26		9		0	13	10	1		0		448	4	7	8	1,21	1,12					
375	fadL	OM	5	4		1		0	5	5	3		0		225	11	11	11	0,96	0,62					
52	lptE	OM	4	4					9	9	0		0		160	0	0	0	N/A	N/A					
395	btuB	OM							2	2					43	0	0	0	N/A	N/A					
386	lpp	OM							7	7	2		1	0,16	201	16	16	17	1	1,89					
267	ompX	OM				8		9	6	5					430	1	2	2	8,30	5,80					
609	SSS3G_0127	OM	3	1		3		0																	
749	yddB	OM				2		0																	
741	ymcA	OM	3	3		3		0	3	3					342	1	1	1	14,20	11,90					
	slyB	OM													199	1	1	4	20,00	20,54					

Table 15. 30 total proteins (Ext + OM) identified in total in IP 1-11
First 3 columns of the table: details of IP experiments 4-11. Right part: IP 1-3 experiments. For each protein are shown the gene name and the PSORTb 3.0 localization. Each

comparison is visualized with the same font color and the lysis conditions are indicated. The 'number of unique peptides quantified' indicates if a protein was quantified, and therefore detected, in the individual samples. Sample 8 is the only negative control in the IP 4-11 and the majority of the proteins were absent in this sample. 6 proteins were only quantified with 1 peptide and 1 protein was quantified with 5 peptides as in sample 9. The 'SumRatio H/M' of the proteins with a confident quantification (peptides>=3) in sample 9 shows that these few proteins are mainly present in sample 9 than in the control. All the peptides listed in grey indicate a not confident quantification with less than 3 peptides. The 2 batches of sera recognize a similar pattern of proteins and only a slight variability is detected in the identification of the last 6 proteins of IP 4-11 list. With the ultra-centrifugation step (sample 7) the majority of the proteins present in sample 4 are lost into the pellet (as not detected into the supernatant). In IP 1-3,the 'SumRatio M/H' and 'SumRatio L/H' indicate the abundance of each protein (in terms of fold increase) in the IP performed with serum compared with the negative control. N/A= not available. Although few proteins (blue) were quantified with only 1 peptide, the intensity of this peptide in 'L' or 'M' was quite high in all the cases producing high 'SumRatio' values with the negative control. As these proteins, highlighted in blue, were found also in other conditions, it is likely that they were present in the samples 2 and 3.

Analyzing the information contained in table 15 we observed that the two batches of sera recognized 24 proteins, and only a slight variability was detected in the identification of the last 6 proteins of IP 4-11 list (left part of the table). Another important result was obtained from the comparison of sample 8 (negative control) with sample 9 (IP with serum). Sample 8 is the only negative control in the IP 4-11 and the majority of the proteins were absent in this sample. 6 proteins were only quantified with 1 peptide and 1 protein (SigA) was quantified with 5 peptides in sample 9 suggesting a non-specific interaction with the microbeads during the procedure. Moreover, considering the 'SumRatio' values of the 6 proteins which were quantified with only 1 peptide, all were predominantly present in the IP sample compared to the control. As the lysis conditions of the IP 5 and 6 are quite similar to the ones used for sample 9 and the negative control (in sample 5 and 6 a high salt washing step was performed and in 6 the ribonucleases were absent), and the proteins identified in sample 9 were confirmed in samples 5 and 6, we can conclude that the proteins quantified in samples 5, 6 and 9 were immunoprecipitated. Another interesting result was observed in the comparison 4-7 based on the addition (in sample 7) of an ultra-centrifugation step after the lysis of the bacteria. This step was performed to test if it allowed the removal of ribosomal complexes as potential contaminants. However, we observed that in sample 7

approximately half of the proteins did not show any quantified peptides and for several of the remaining proteins no confident quantification was obtained as only a single peptide was quantified. These results suggest that in sample 7 the antibody/antigen complexes were lost during the ultra-centrifugation into the pellet, as indicated by the absence of most of the protein bands in the gel shown in figure 32B. In the comparison of samples 11-10 and 5-6 we observed that the presence of ribonucleases and the addition of the high salt washing step reduced, the unspecific binding of cytoplasmic proteins which was also reflected by a reduced number of protein bands visible by SDS-PAGE (samples 10 and 5, figure 32). The precipitation of OM and extracellular proteins was not affected.

OmpX, YmcA and YtfM quantified with only 1 or 2 peptides in IP 1-3, as previously discussed, were also identified in sample 4, 7, 5, and 6. As more peptides were quantified in the latter samples, the data suggest a real identification of the protein also in samples 2 and 3.

In conclusion, among all the conditions tested, conditions 5 and 6 equally allowed the identification of the highest number of OM and Ext proteins and might be considered as the starting point to improve the procedure of the IP approach.

One of our primary interests was to find the conditions to reduce the precipitation of ribosomal proteins (already detected in the first IP experiments) and, to understand if DNA sticks to the Ab/Ag complexes and to the ribosomal complexes connecting everything. In this case the presence of ribonucleases in the lysis buffer might avoid the formation of such complexes. To address these questions, we analyzed the list of cytoplasmic proteins identified in all the IP experiments (table 16). A list of 49 cytoplasmic proteins was identified in all the conditions and it included ribosomal proteins, chaperones, DNA-polymerase subunits, RNA-polymerase subunits, and proteins involved in the metabolism. Looking at the 'unique peptides quantified' for each proteins, we observed that the protein GroL is the most abundant cytoplasmic protein identified in all the samples, and it showed a quite high 'SumRatio' values in IP 1-3. The risobomal proteins (Rpl- and Rps- proteins) showed high numbers of peptides quantified and 'SumRatio' values as well. Interestingly, most of the proteins detected in samples 9 were absent in the negative control (sample 8), with the exception of few cases, suggesting an interaction of these proteins with the antibody or with the antibody/antigens complexes.

		unique peptides quantified										unique peptides quantified							
		serum batch 2		serum batch 1					no serum			serum batch 1			IP 1-3				
		11	10	4	7	5	6	9	8			3	1	2					
Protein Group ID IP 4-11	gene	+EDTA	- EDTA + Rnasi/Dnasi +MgCl2	- EDTA + Rnasi/Dnasi	- EDTA + Rnasi/Dnasi +Ultra@	- EDTA + Rnasi/Dnasi + MgCl2 + high salt wash	+ EDTA + high salt wash	- EDTA + Rnasi/Dnasi +MgCl2 SERUM	- EDTA + Rnasi/Dnasi +MgCl2 NO SERUM	Protein Group ID IP 1-3	pre-immune H	1µl serum M	10µl serum L	SUM RATIO L/M	SUM RATIO L/H	SUM RATIO M/H
89	accD	13	13	4	1	20	21	5	0										
739	aceF	66	66	52	3	39	41	25	2	320	33	37	43	0,98	0,66	0,68			
635	acnB	0	1	4	0	7	7	3	0										
413	alaS	15	16	12	0	17	18	7	0	116	3	3	3	2,23	3,22	1,44			
65	atpA	33	33	34	5	42	44	17	1	57	24	28	31	0,70	22,02	31,41			
64	atpH	4	4	5	0	10	10	5	0	58	3	3	3	0,59	15,73	26,85			
173	dnaJ	3	3	4	1	2	3	3	0										
407	dnaK	35	41	27	3	33	34	31	5	187	30	30	35	0,99	1,49	1,50			
137	eno	5	6	8	0	8	11	13	1	112	8	8	8	0,39	2,24	5,77			
212	fusA	4	20	12	5	20	19	15	2										
482	gltA	5	6	6	1	11	11	7	2	419	3	3	5	0,82	18,59	22,76			
121	groL	112	118	93	2	105	129	66	3	31	86	100	112	0,85	18,60	21,84			
324	guaB	9	9	8	0	5	5	3	0	207	6	6	7	0,81	13,21	16,23			
340	gyrA	8	8	4	0	8	10	7	0	216	7	8	10	1,50	12,98	8,55			
297	hfIC	7	7	7	0	12	14	6	0	287	7	7	7	0,59	4,83	8,17			
420	htpG	2	10	6	2	13	14	7	0										
46	ilvC	5	5	5	0	6	6	2	0	53	5	5	6	1,09	1,23	1,13			
398	infB	12	12	5	0	7	8	4	0	101	3	3	3	2,13	2,59	1,21			
279	lpdA	5	5	25	0	18	18	6	1	321	14	17	17	0,61	8,25	13,38			
442	mukB	12	12	1	0	12	12	0	0	153	3	4	6	0,82	1,04	1,27			
78	pepQ	2	4	2	0	3	3	3	0										
439	pheS	11	11	7	1	17	18	4	1	143	3	4	5	0,74	8,28	11,18			
418	pheT	15	15	3	1	11	11	2	0	142	2	2	5	1,27	18,14	14,29			
433	pnp	34	35	16	1	21	24	14	0	103	10	15	17	0,67	18,57	27,03			
584	prs	21	21	10	0	33	36	10	0	230	7	9	9	0,85	11,77	13,80			
576	purB	14	15	7	4	19	19	11	1	350	2	2	5	0,87	2,98	3,42			
97	rplA	42	39	8	7	40	58	13	2	40	26	30	38	1,34	78,05	58,37			
98	rplJ	11	11	3	0	22	23	7	0	39	2	2	4	2,80	64,70	23,14			
168	rplU	8	8	5	0	11	11	8	4	98	6	6	7	2,16	3,68	1,70			
256	rpoA	31	32	29	0	36	37	17	1	93	9	9	10	0,76	41,24	54,08			
99	rpoB	49	51	18	0	47	51	14	0	37	48	55	64	0,71	19,66	28,21			
100	rpoC	38	38	8	5	28	30	6	1	36	23	29	41	1,25	23,26	17,71			
616	rpoD	11	11	5	0	12	12	8	2	239	5	6	9	0,81	4,46	5,50			
359	rpsA	52	52	3	0	15	21	7	1	337	13	16	25	1,20	8,53	6,84			
205	rpsC	11	11	3	1	14	18	1	0	79	3	3	5	0,59	3,59	6,08			
193	rpsD	20	20	19	14	27	29	10	1	92	1	2	3	0,53	91,91	171,65			
255	rpsE	64	62	7	1	52	74	32	7	87	24	29	45	1,52	24,33	15,89			
122	rpsF	5	6	2	0	8	9	2	0	29	5	5	7	2,90	18,64	6,43			
213	rpsG	21	20	7	0	20	26	19	0	72	19	20	25	1,20	58,41	48,45			
165	rpsI	5	5	0	0	20	20	0	0										
214	rpsL	4	4	2	0	10	15	2	0	71	3	3	5	1,15	2,40	2,08			
174	rpsT	23	23	31	20	21	21	15	9	185	14	14	19	2,09	6,64	3,18			
162	rpsU	6	6	3	0	14	16	0	0	105	4	4	4	1,00	3,68	3,67			
248	SS53G_1336	6	6	5	3	17	18	7	3	233	15	15	16	1,03	1,38	1,34			
274	SS53G_5762	9	9	4	0	12	12	5	0	454	5	6	6						
485	sucA	8	8	0	0	7	8	1	0	422	1	1	4	0,96	4,93	5,14			
755	sucB	20	20	11	0	28	28	11	0	423	10	11	15	1,62	10,20	9,23			
240	tig	0	4	2	2	16	15	6	0	164	6	5	7	0,25	0,06	0,24			
590	topA	6	6	5	1	6	7	6	3	235	5	5	8	0,81	1,56	1,92			

Table 16. 49 cytoplasmic proteins identified in all the IP experiments.
Left part of the table: details of IP experiments 4-11. Right part: IP 1-3 experiments. The 'Protein Group ID' and gene name are shown for each protein. The 'unique peptides quantified' and the IP conditions are also shown for each sample. Samples 8 and 3 are negative controls.

Analyzing the IM proteins found in each condition (table 17) we observed that the majority of proteins were not detected in the negative control (sample 8). Roughly half of the proteins showed inconsistent quantification with sample 9, as only 1 'unique' peptide was quantified, and the same low detection was observed in some of

the other conditions. For 23 proteins highlighted in red in table 13, instead, we observed quantification based on a high number of 'unique peptides quantified' in all the samples. These data, in combination with the absence of peptides quantified in sample 7, suggest an interaction also for the inner membrane proteins with the antigen/antibody/cytoplasmic proteins complexes. Most of the proteins highlighted in IP 4-11 showed high 'SumRatio' values in IP 1-3, in many cases >20, confirming that these proteins interact with these complexes likely after the lysis of the bacteria or through an indirect interaction with the outer membrane proteins. Studies conducted in *E. coli*, showing the network interactions among cytoplasmic and inner membrane proteins, support our explanation about the identification of these proteins in the IP approach as most of the cytoplasmic and some of the IM proteins we found are present in this network-interactions map[17]. Stronger conditions of lysis might be used to disrupt these interactions considering that they may also affect or remodel the structure of the real targets.

All these results suggest that antibodies contained in GMMA antisera bind to the outer membrane proteins on *S. sonnei* bacteria, but also to some extracellular proteins that after the secretion might associate on the surface of the bacteria.

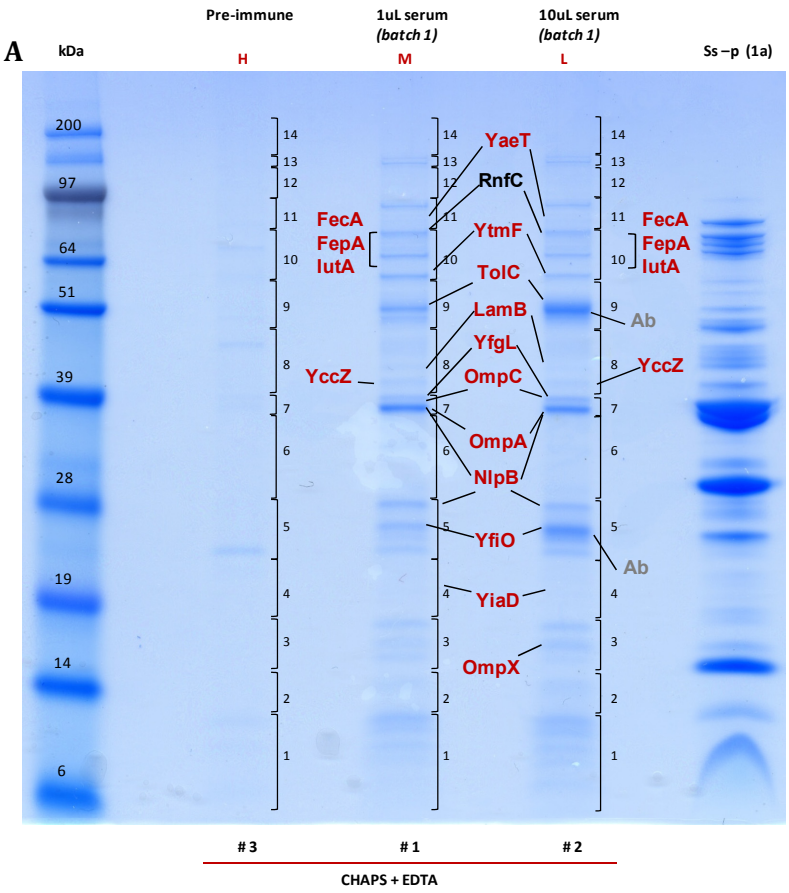
IM proteins		unique peptides quantified												unique peptides quantified					
		serum batch 2		serum batch 1					no serum										
		11	10	4	7	5	6	9	8			3	2	1					
ID IP 4-11	gene	+ EDTA	- EDTA + Rnasi/Dnasi +MgCl2	- EDTA + Rnasi/Dnasi	- EDTA + Rnasi/Dnasi + Ultra@	- EDTA + Rnasi/Dnasi + MgCl2 + high salt wash	+ EDTA + high salt wash	- EDTA + Rnasi/Dnasi +MgCl2 SERUM	- EDTA + Rnasi/Dnasi +MgCl2 NO SERUM	SumRatio H/M ##8H#9M	ID IP 4-11	pre-immune H	1µl serum M	10µl serum L	SUM RATIO L/H	SUM RATIO M/H
67	atpD	53	59	48	2	63	70	34	1	1,02	55	26	29	34	12,48	12,40
63	atpF	12	11	14	0	25	25	18	0		59	16	18	20	7,60	10,94
434	lpoA	9	9	1	0	5	6	3	0		104	2	2	2	4,03	2,06
182	secA	11	11	1	0	6	6	2	0		177	0	0	1	N/A	N/A
646	SS53G_0535	3	3	5	0	3	3	3	1	0,88	202	2	2	2	11,58	24,69
371	wzzB	1	1	6	1	6	6	5	0		222	5	6	7	2,19	5,35
701	narG	2	5	3	0	1	3	3	1	10,25	231	0	0	0	N/A	N/A
606	exbB	5	5	4	0	8	8	5	0		243	2	4	4	10,60	18,12
313	hemY	4	4	5	1	6	6	2	0		261	2	2	2	9,67	15,39
734	emrA	4	4	2	0	6	7	3	0		274	1	1	2	8,48	50,90
570	treB	4	4	3	0	3	3	3	0		293	0	0	2	N/A	N/A
656	manZ	3	4	5	0	7	7	4	0		301	2	2	3	11,70	16,22
288	hflB	29	30	27	0	19	22	14	0		359	14	14	18	5,72	6,83
746	yhil	3	5	2	0	7	7	3	0		378	4	5	5	7,74	29,11
748	proP	2	2	9	1	3	3	3	0		384	4	5	5	12,92	22,56
293	pntB	1	2	3	0	1	2	3	0		392	1	1	1	15,02	25,40
360	secD	13	13	5	0	15	14	2	0		400	5	6	6	9,34	23,44
664	secF	8	9	13	0	13	17	11	0		401	8	8	8	10,42	20,59
671	acrA	19	19	16	0	23	26	14	0		407	8	8	8	15,45	34,35
528	ftsN	8	9	7	0	9	9	3	0		438	3	3	3	14,51	28,74
66	atpG	9	9	11	0	18	18	7	0							
669	ppiD	27	28	20	6	21	22	9	3	1,03						
689	cadA	10	10	14	14	12	12	10	10	0,69						
384	tatB	2	2	3	0	2	2	3	0							
760	glnQ	0	0	2	0	2	2	3	0							
323	ecnB	2	2	2	2	4	6	3	1	0,18						
337	fabI	0	2	1	0	3	3	4	0							
423	nuoB	2	2	0	0	6	6	2	2	21,73	133	0	0	0	N/A	N/A
416	lolD	1	1	0	0	4	5	1	0		150	1	1	1	3,53	4,09
443	ybhG	2	2	1	0	5	6	1	0		157	4	4	4	1,61	1,38
181	murG	3	3	2	1	5	7	1	0		179	3	3	4	7,11	4,39
517	fruA	0	0	8	0	3	3	1	0		220	3	4	5	6,10	9,48
552	lepB	1	1	4	0	5	5	2	0		257	0	0	1	N/A	N/A
385	tatA	0	0	0	0	1	2	1	0		264	1	1	2	21,36	34,02
530	yijP	2	2	2	0	2	2	1	0		271	2	2	2	2,87	11,75
566	cycA	2	2	2	0	3	3	1	0		289	0	0	0	N/A	N/A
568	mgtA	4	4	2	0	4	4	1	0		295	5	5	6	2,72	2,98
567	yjgQ	1	2	1	0	1	2	1	0		297	0	1	1	N/A	N/A
711	gcd	3	3	1	0	3	3	1	0		323	1	2	4	12,26	20,90
712	rseP	2	2	1	0	2	2	1	0		328	2	2	3	5,19	13,42
624	secG	2	2	4	1	2	2	2	0		358	2	2	3	3,18	4,00
373	gadC	0	0	3	1	0	0	1	0		388	2	2	2	6,66	8,59
332	yajC	1	1	1	0	2	2	1	0		399	1	1	1	29,76	24,13
667	cyoA	2	2	4	0	3	4	2	0		402	2	2	2	3,18	6,53
686	ybaL	1	1	2	0	2	2	2	0		409	0	1	1	N/A	N/A
688	cstA	0	2	2	0	3	3	1	0		412	3	4	4	3,00	7,71
484	sdhB	1	1	1	0	4	4	1	2	11,68	421	1	1	3	0,74	13,19
300	cydA	1	2	4	0	4	5	1	0		424	3	3	3	7,33	17,11
304	tolQ	1	1	1	0	1	1	1	0		425	2	2	2	21,75	15,88
705	frdA	0	0	1	0	2	2	1	0		437	2	2	2	4,85	4,42
763	hemX	3	3	3	0	0	0	2	0		440	8	8	12	12,23	14,81
458	sdaC	2	2	7	0	7	7	1	0		444	5	5	6	9,48	23,56
473	etk1	11	11	7	3	9	9	0	0							

Table 17. Inner membrane proteins identified in IP 1-11.
Left part of the table: details of IP experiments 4-11. Right part: IP 1-3 experiments. The 'Protein Group ID' and gene name are shown for each protein. The 'unique peptides quantified' and the IP conditions are also shown for each sample.

5.4.3.4. Summary of immunogenic proteins in S. sonnei GMMA

To evaluate if the results of the immunoprecipitation experiments correlate with the protein bands visible in the SDS-PAGE of IP 1-3, we combined the gel picture with the information and results obtained from the immunoprecipitation analysis. In particular, the names of the proteins were assigned to the gel based on the predicted molecular weight of the identified proteins and the slice number in which the proteins were identified. From this analysis we observed a good match between the major immunogenic proteins identified in the IP experiments and the gel picture. Many of the most visible bands in the gel, in fact, corresponded exactly with the band-number, indicated in MaxQuant data, in witch the proteins were identified (figure 34).

A



B

Protein gene	MW (KDa)
fecA	85
fepA	82
iutA	81
lamB	50
nmpC	39
ompA	37
ompC	40
ompX	18
rnfC	80
tolC	56
yaeT	90
yiaD	22
yccZ	41
yfiO	28
ytfM	64

Figure 34. Gel representation of the major OM and Ext immunogenic proteins identified.
A. The SDS-PAGE gel image of the IP 1-3 experiment is used as reference to show the immunogenic proteins (red). Ab: antibodies chains. Black: protein identified in all the immunoprecipitation experiments but not in GMMA. **B.** molecular weights of the identified proteins.

6. DISCUSSION

The outer membrane of Gram-negative bacteria represents an important interface between the bacterial cell and its environment. Bacterial membrane-associated proteins play an essential role in the uptake of nutrients and other compounds, such as small ions, antibiotics, and inhibitors, from the extracellular space. They are involved in interactions with host cells, in cell-cell communication, in promoting bacterial resistance to host responses (e.g. complement-mediated killing), and in drug resistance. For these reasons membrane-associated proteins represent ideal candidates for vaccine development and provide new targets for antimicrobial agents (e.g. Lpp protein[19]). Several studies have been performed to characterize the outer membrane proteome of various Gram-negative bacteria[68,100], starting from outer membrane preparation. However, a poor enrichment of outer membrane proteins and cytosolic contaminations has frequently been observed. In the last years, an increased interest has been observed in the characterization of the proteome of Generalized Modules for Membrane Antigens (GMMA), also called outer membrane vesicles (OMV), obtained from various Gram-negative bacteria including *Escherichia coli*[12], *Neisseria meningitidis*[44], *Pseudomonas aeruginosa*[66], *Helicobacter pylori*[46], and *Salmonella*[4]. GMMA are closed spherical particles naturally released from the surface of Gram-negative bacteria into the environment. They are characterized by a bilayer membrane containing LPS, outer membrane proteins, glycerophospholipids and enclosed periplasmic components, reflecting the composition of the outer membrane. Moreover, GMMA have also been found to have immunostimulatory effects[40,44,106] and to elicit protection against bacterial infections[4,123].

NVGH's aim is to develop a broadly protective protein-based vaccine against *Shigella* using GMMA.

In the current work we present the generation of GMMA over-producing *Shigella* strains, the proteomics characterization of *Shigella* GMMA, and the identification of immunogenic proteins in GMMA. In particular, we illustrate two different proteomic approaches to characterize the GMMA protein composition: (i) bi-dimensional mapping in combination with MALDI-TOF MS, and (ii) total GMMA proteome using separation by GeLC-MS/MS in combination with stable isotope dimethyl labeling

enabling a quantitative analysis. We performed a qualitative and quantitative analysis of *S. sonnei* and *S. flexneri* GMMA proteomes and we identified a list of potentially conserved proteins shared between *S. sonnei* and *S. flexneri* GMMA. Finally, we present (i) the identification of immunogenic proteins in GMMA by bi-dimensional Western blotting, and (ii) a quantitative analysis of surface immunogenic proteins in *Shigella*.

6.1. GMMA GENERATION AND EFFECT OF TEMPERATURE AND LPS-MODIFICATION ON GMMA COMPOSITION

We initially focused on the generation of GMMA-overproducing *Shigella* strains and the identification of the protein composition of the GMMA. In a first approach *S. sonnei 53G* was chosen to introduce a null mutation of the *tolR* gene to overproduce GMMA (Ss Δ*tolR*), as previously described for *E. coli*[12]. To avoid the immunodominant response to the O-antigen the virulence plasmid carrying the genes for the O-antigen biosynthesis[65] was cured from the Δ*tolR* strain, resulting in strain Ss –p Δ*tolR*. To enhance vaccine safety of GMMA, a null mutation to produce LPS with reduced endotoxicity (Δ*msbB*)[28] was inserted into the Ss –p Δ*tolR* strain, resulting in a mutant lacking the *msbB1* and *msbB2* genes located on the chromosome and on the plasmid, respectively (Ss –p Δ*tolR* Δ*msbB*).

A chemically defined medium was optimized for the fermentation of Ss –pSS Δ*tolR*. The Ss –p Δ*tolR* Δ*msbB* strain was not able to grow to high optical densities in the defined medium at 37°, while it reached high OD at 30°C. Therefore Ss –pSS Δ*tolR* Δ*msbB* was cultivated at 30°C for the production of GMMA. The *msbB* mutation has not been reported to induce any growth defects in *Shigella flexneri 5a*[28] or in *E. coli K-12*[128] cultivated at 37°C. The impaired growth of the Ss –pSS Δ*tolR* Δ*msbB* mutant at 37°C could be a result of the combination of the *tolR* and the *msbB* mutations or a consequence of the medium composition that could potentially be optimized for growth at 37°C. However, comparing the protein profiles of GMMA generated from Ss –p Δ*tolR* at 30°C and 37°C by SDS-PAGE, we observed only minor differences in the protein profiles. This indicates that switching of the temperature from 37°C to 30°C does not induce a significant alteration of the GMMA protein profile.

A first proteomic approach based on mono- and bi- dimensional electrophoresis, followed by picking of proteins spots and MALDI-TOF MS analysis, revealed that GMMA from Ss –pSS Δ*tolR* and Ss –pSS Δ*tolR* Δ*msbB* overall contain a similar protein composition but show some differentially expressed proteins. 10 proteins showed a general lower expression in GMMA from Ss –pSS Δ*tolR* Δ*msbB* than in in GMMA from Ss –pSS Δ*tolR*. On the contrary, outer membrane protein W, outer membrane protein X, and 5 not identified proteins were more expressed in the Ss –pSS Δ*tolR* Δ*msbB* than in the Ss –pSS Δ*tolR* GMMA. Among these, 3 proteins were temperature-regulated, as their over-expression at 30°C was also confirmed in the comparison 30°C/37°C of Ss –pSS Δ*tolR* GMMA. Thus, it is likely that other differences detected in the Ss –pSS Δ*tolR* Δ*msbB* GMMA are linked to the LPS-modification. The *msbB* mutation has been reported to activate the sigmaE pathway which is involved in the regulation of the extracytoplasmic stress response and regulates the transcription of 10 genes[130]. Among these genes, *degP*[91] encodes for a periplasmic protease also known as HtrA, which we found to be conserved in *S. sonnei* and *S. flexneri* GMMA. Surprisingly, by 2-D gel analysis we found that HtrA is less expressed in Ss –pSS Δ*tolR* Δ*msbB* than in Ss –pSS Δ*tolR* GMMA. As we have not verified that the *msbB* mutation triggers expression of the sigmaE regulon in GMMA producing *Shigella* strains, it is not clear if the *msbB* mutation does not enhance expression of *htrA* in *Shigella* or if expression of *htrA* is dominantly regulated by temperature.

6.2. GMMA PROTEOME

OmpA and OmpC were shown to be the major outer membrane proteins in *Shigella* GMMA, confirming existing data regarding the predominance of these molecules in outer membrane preparations and GMMA obtained from different Gram-negative bacteria[79,123,143]. An Ss –pSS Δ*tolR* strain lacking the *ompA* gene was generated to allow the identification of proteins with similar molecular weight of OmpA. The proteomic analysis of Ss –pSS Δ*tolR* Δ*ompA* GMMA by 2D SDS-PAGE and MALDI-TOF of selected spots revealed, as expected, that the majority of detected proteins found in GMMA were outer membrane proteins, outer membrane lipoproteins, and periplasmic proteins. A fraction of predicted cytoplasmic components (15%) was also identified. However, this proportion was low compared to published studies

performed on GMMA proteomes[44,52,140]. The identification of cytoplasmic proteins might have likely resulted from limited cell lysis in the cell culture and "sticking" of some of the released proteins to the surface of the bacteria or the GMMA.

To compare the protein composition of GMMA from different serotypes of *Shigella* we analyzed GMMA obtained from an O-antigen deficient *Shigella flexneri* 2a 2457T strain (*Sf* Δ*tolR* ΔOAg). We identified 65 proteins of which 28 were in common with the proteins identified in Ss –p Δ*tolR* Δ*ompA* GMMA. Surprisingly, in *Sf* Δ*tolR* ΔOAg GMMA we found a small percentage (5 proteins) of inner membrane components which has previously been described in *Legionella* GMMA[52]. In addition, a higher percentage of predicted cytoplasmic proteins (38%) compared to *S. sonnei* GMMA were identified. These differences might be related to the fact that the final optical density for *S. flexneri* GMMA was higher than for *S. sonnei* and, thus, more lysis could have potentially occurred in the *S. flexneri* culture than in the *S. sonnei* culture. However, as approximately twice as many proteins were identified in *S. sonnei* GMMA than *S. flexneri* GMMA the differences might also partially result from a technical bias that cytoplasmic proteins might be easier to identify due to their solubility. Also, the deletion of the major outer membrane protein OmpA in *S. sonnei* might enhance the identification of outer membrane proteins with lower abundance especially of similar size as OmpA.

To overcome the technical limitations of 2-D gel, such as identification of low abundant proteins, difficulty in solubilizing membrane proteins, and identification of large molecular mass proteins which are commonly missing from 2-D maps[24], we compared *S. sonnei* and *S. flexneri* GMMA from strains with homologous mutations using an alternative proteomic approach. An extensive analysis of the GMMA protein content, including a quantitative analysis of the GMMA proteome and an estimate of the enrichment of the outer membrane proteins in GMMA, has never been shown. In the present work we aim to fill this gap by analyzing and comparing the protein content of *S. sonnei* and *S. flexneri* GMMA obtained from strains lacking the O-antigen and carrying or not carrying the virulence plasmid (Ss +p Δ*tolR* ΔOAg, Ss -p Δ*tolR* (ΔOAg), Sf +p Δ*tolR* ΔOAg, Sf -p Δ*tolR* ΔOAg). To achieve this aim we used a proteomic approach based on GeLC-MS/MS analysis. We identified in total 387 proteins in *S. sonnei* GMMA and 426 proteins in *S. flexneri* GMMA. We found that approximately 46% of the identified proteins in *S. sonnei* (178 proteins) and 40% of

the identified proteins in *S. flexneri* (170 proteins) were predicted to be extracellular, outer membrane, and periplasmic proteins. Unlike the 2-D gel approach, in these experiments we identified a quite high and similar percentage of inner membrane proteins (10-12%) and cytoplasmic proteins (20-26%) in GMMA from *S. sonnei* and *S. flexneri*. This difference is likely related to the different methods and instruments used in the two approaches. As discussed above, in the 2-D gel experiments we started with a selection of protein spots visible on the gel and we identified only a subset of the total proteins present in the samples. Most of the low abundance proteins might not have been visible and thus not been picked for identification. In the second approach, instead, the entire protein content of each sample was analyzed and likely identified. Moreover, the sensitivity of the LTQ-Orbitrap Velos instrument used in the second approach is about 10 fold higher than the Ultraflex TOF-TOF instrument used for the 2-D analysis. This allowed us to identify also low abundant proteins..

Although inner membrane and cytoplasmic proteins were identified with percentages of 10% and 20-26%, respectively, we showed that GMMA are enriched in OM and periplasmic proteins, while inner membrane and cytoplasmic proteins represent only a small fraction of the GMMA protein amount. Comparing the predicted localization of the GMMA proteomes with the *in silico* proteomes of *S. sonnei* and *S. flexneri* strains we observed that the outer membrane proteins identified in GMMA represented 71.5% and 65%, respectively, of the total predicted OM proteins in *S. sonnei* and *S. flexneri*. The periplasmic proteins represented approximately 50%, while cytoplasmic and inner membrane proteins represented only a small percentage (4% each). Furthermore, the enrichment of GMMA in OM and periplasmic proteins was confirmed when we analyzed the abundance of the 316 proteins found in common between *S. sonnei* and *S. flexneri* GMMA. As an estimate for the amount of the common proteins in *S. sonnei* and *S. flexneri* GMMA we used the Exponentially Modified Protein Abundance Index (emPAI)[62]. The most abundant proteins were predicted to be OM and periplasmic proteins, while the less abundant proteins where predominantly inner membrane and cytoplasmic proteins. Among the most abundant proteins we found the outer membrane proteins OmpA and OmpC, having the highest emPAI values in *S. sonnei* GMMA, in accordance with what expected from SDS-PAGE picture, and a quite high (but not the highest) emPAI

values in *S. flexneri* GMMA. YaeT, YfiO and TolC showed high emPAI values but not as high as OmpA and OmpC in *S. sonnei*, while in *S. flexneri* they showed similar values with the exception of TolC having a slightly higher emPAI value. Several periplasmic proteins were shown to be abundant as well: YraP, TolB, SurA, and HtrA. Moreover, we observed a good match between the abundance results and the major protein bands visible by SDS-PAGE of GMMA, confirming the quantitative analysis. Although MaxQuant data only indicate the slice in which each protein was identified, we assigned the protein names to the majority of the bands comparing the predicted molecular weights of the most abundant proteins with the most visible bands in the gel. When available, we compared these data with previous identification of proteins by 1-D gel and found a good correlation between the protein name and the band association. Comparing the list of identified proteins in *S. sonnei* and *S. flexneri* GMMA we identified 55 conserved outer membrane and extracellular proteins, and in *S. sonnei* GMMA a further 6 additional proteins were identified. In particular, one protein annotated as 'Entry exclusion protein 2' in *S. sonnei 046* was identified in *S. sonnei* but not in *S. flexneri* GMMA, and it showed 90%-100% identity with several pathogenic *E.* coli isolates. Among the conserved proteins, we found 3 proteins encoded on the virulence plasmids (MxiM, MxiD and IcsA/VirG). MxiD is one of the 4 predicted OM proteins encoded on the virulence plasmid, forming the outer membrane rings of the Mxi-Spa type III secretion system of *Shigella*. MxiM is an outer membrane lipoprotein anchored to the inner face of the outer membrane mediating membrane insertion and stabilization of the MxiD ring[125]. IcsA/VirG, instead, is one of the 2 proteins predicted to have multiple localization sites (OM and Ext), but according to the literature is found secreted[3] and is involved in actin polymerization allowing intra- and intercellular spread of *Shigella*. In addition, we identified other 3 proteins with a virulence-associated function but encoded on the chromosome: serine protease SepA (required for actin assembling of the host cell during infection), serine protease SigA (autotransporter showing a cytopathic effect on epithelial cells) and the TieB protein enterotoxin ShET-2. These 3 proteins were also found with high identity in pathogenic *E. coli*. In particular SepA shows 90% identity with a protein in enterohemorragic *E. coli* (encoded on the plasmid p08641) and in Shiga-toxin producing *E. coli*. SigA shows 90% identity with a toxin in enterohemorragic *E. coli* and in enteroaggregative *E. coli*. TieB shows 100% identity

with a toxin in extraintestinal pathogenic *E. coli*. Our findings confirm the recent identification of IcsA/VirG by Camacho et al.[18] in *S. flexneri* GMMA, and the identification of the enterotoxin SigA by Al-Hasani et al.[3]

6.3. GMMA IMMUNOPROTEOME

In accordance with previous reports[4,123], GMMA elicited high levels of GMMA-specific IgG in mice after application of 2µg of GMMA. Sera raised against GMMA from Ss –pSS Δ*tolR* and Ss –pSS Δ*tolR* Δ*msbB* strains were used to identify immunogenic proteins by 2-dimensional Western Blots using the Ss –pSS Δ*tolR* Δ*ompA* as bait. The pattern of recognized proteins differed slightly for the different sera. In particular, the immunogenic serine protease HtrA was detected with a weaker signal with sera raised against Ss –pSS Δ*tolR* Δ*msbB* GMMA than with sera raised against Ss –pSS Δ*tolR* GMMA. These observations are in accordance with the 2-D SDS-PAGE results that showed the HtrA is less present in the Ss –pSS Δ*tolR* Δ*msbB* GMMA. The majority of the 14 immunogenic proteins identified were predicted to be periplasmic proteins. We only identified 3 immunogenic outer membrane proteins by this approach: OmpA, OmpX and YaeT[11(submitted)]. Interestingly, the abundant protein OmpC was not detected by WB. Also in a previous immunoproteomic analysis of isolated outer membranes of *S. flexneri 2a*[143] OmpC was not detected as immunogenic indicating that either OmpC is not immunogenic or that the epitope of potential antibodies are not maintained after denaturing SDS-PAGE. These hypotheses might also apply to other proteins that were not detected by this approach.

For these reasons, we focused on a different method to identify immunogenic proteins that are exposed on the surface, but might not be detectable by Western blot after SDS-PAGE. We evaluated a novel approach based on immunoprecipitation of *Shigella* surface proteins using sera raised against GMMA. Moreover, we present a quantitative proteomic analysis of the immunogenic proteins based on stable isotope dimethyl labeling followed by LC-MS/MS. We performed the immunoprecipitation (IP) under 11 different experimental conditions. The samples of IP 1-3 were analyzed in one experiment to evaluate the base level of identification of proteins using pre-immune serum and to determine the amount of serum sufficient to achieve

(mostly) saturated conditions for antibody binding to surface proteins. The remaining 8 samples (IP 4-11) were analyzed in pairs to directly compare the panel of immunogenic proteins identified under different conditions, and to identify the most promising experimental conditions for future experiments. By this approach we were able to identify 30 immunogenic proteins predicted to be outer membrane proteins (28) and extracellular proteins (2). Among the immunogenic proteins identified using the IP approach we found OmpA, OmpX, and YaeT confirming our previous 2-D WB results and studies performed on *Shigella flexneri* outer membrane proteome[143]. Interestingly, we identified proteins that were not identified in the 2-D approach, but that were reported to be immunogenic in *Shigella*[143] (e.g. TolC) supporting the enhance power of the IP approach over the 2-D WB blot approach. In addition, using the IP approach we identified OmpC as immunogenic, indicating that the antibodies recognize native epitopes that are missing under the denaturing conditions of 2-D WB.

Among the predicted outer membrane proteins we identified YfiO, NlpB and YfgL that have been described in *E. coli*[93] as OM lipoproteins interacting with YaeT. According to this interaction and the hypothetical localization of the 3 proteins on the inner side of the outer membrane, it is likely that these three proteins were not directly recognized by the antibodies, but were immunoprecipitated as a complex with YaeT. The identification of the extracellular protein RnfC and SigA (predicted as extracellular/outer membrane) might result from the association of these proteins on the surface of the bacteria after their secretion. Among the 30 immunogenic OM/Ext proteins, 23 were identified in all the samples, 4 (Lpp, YmcA, OmpX and SS53G_0127) were found in 4-5 samples, while 3 proteins (SlyB, YddB and BtuB) were found in 3 individual experiments.

Unexpectedly, the proteins indentified in IP 1-3 were present in all of the three experiments, meaning we did not find proteins that were present in the IP performed with sera and absent in the IP performed with pre-immune serum. In the negative control performed using micro-beads alone (sample 8) the majority of the proteins present in IP 4-11 were absent while only 6 proteins were detected. This suggested the presence of unspecific binding to the micro-beads used to recover the antibody/antigen complexes during the experimental procedure. However, we observed that when serum was present an increase in the amount of the identified

proteins was observed suggesting a specific binding of these proteins to the antibodies. In particular, we observed 'SumRatio' values (indicating the fold increase of the protein amount in the IP performed with serum in comparison with the negative control) >=5 for 19 proteins in IP 1-3 likely suggesting a specific antibody/antigen interaction. Moreover, we did not observe significant differences in the amount of proteins detected with two different amounts of serum, suggesting that the IP was performed under saturating conditions of antibody binding already at the lower amount of serum. However, for 2 proteins (FepA and NmpC) we observed a significant increase in the 'SumRatio' values when using 10-fold more serum. This indicates lower antibodies titer against these proteins as the lower amount of serum apparently did not contain sufficient antibodies to saturate the binding to all the FepA and NmpC molecules present on *Shigella*.

Due to experimental limitations not all the IP conditions were compared with the negative control. Only 1 sample in IP 4-11 was performed as negative control (no serum) and it was compared to only 1 sample performed with serum. The negative control showed that for the majority of the proteins no peptides were detected confirming the specificity of the antibody/antigen binding. According to the total number of identified proteins in each sample, the 2 IP conditions allowing the identification of the largest number of surface proteins were: condition 5, containing ribonucleases in EDTA-free lysis buffer, and condition 6, containing EDTA but not ribonucleases, both performed with a high salt washing step after the recovery of Ab/Ag complexes. However, in these samples we also observed the highest background suggesting that likely the washing step does not have an effect on the identification of the immunogenic proteins and on the reduction of the background.

A second batch of serum from a different immunization study with the same type of GMMA recognized a similar patter of proteins, with the exception of 3 proteins, indicating a slight variability of the antibody response in the 2 studies. As the experiments were performed with pooled sera and comparisons of the individual sera were not included it is not clear if the observed difference is a common difference between the 2 studies or if it could have resulted from the response in e.g. only 1 mouse.

Unexpectedly, we also detected immunogenic proteins present in all the conditions with predicted cytoplasmic and inner membrane localization. Among these,

approximately 30 cytoplasmic proteins and 23 inner membrane proteins showed an increase in abundance with the presence of the serum, suggesting a specific interaction. The mechanism for such an interaction is not clear. One hypothesis is that some of the cytoplasmic proteins identified in GMMA that could raise an antibody response, are located to a small percentage also at the surface of live bacteria. The concept that proteins without predicted hydrophobic/trans-membrane domains might become membrane-associated due to function in that compartment is gradually increasing[134]. Documented cases supporting this idea are ATP synthases[50] and heat shock proteins[34]. E.g. the identification of 'ATP synthase subunit beta' and 'ATP synthase subunit b' in our experiments, is consistent with reports that these proteins have been identified as surface located[50]. Also, GroEL which was identified as one of the most abundant cytoplasmic immunogenic proteins in GMMA by 'SumRatio' values has been indentified as surface localized in studies performed on numerous bacteria, such as *Salmonella typhimurium*[41], *Helicobacter pylori*[141], *Legionella pneumophila*[53,54], and *Borrelia burgdorferi*[71] and has been shown to be involved in adhesion or invasion of several type of cells (e.g. human gastric carcinoma cells)[141]. The mechanism by which GroEL and other predominantly cytoplasmic proteins are transported to the bacterial surface is currently unknown, since the protein sequence does not contain a classical signal peptide. Different hypotheses have been made to explain the surface association[57]. The secretion of proteins across the inner membrane of Gram-negative bacteria is mostly mediated by translocation machinery, such as Sec secretion pathway and two-arginine translocation (Tat) secretion pathway, recognizing a signal peptide[10]. Surprisingly, several bacterial proteins have shown to be secreted without a classical signal peptide (e.g. GroEL homolog HspB has been shown to be actively secreted in the stationary phase by *Helicobacter pylori*[136]), and this phenomenon is called 'non-classical secretion'[10]. In addition, although the abundant cytoplasmic protein 'Elongation factor Tu' was identified in GMMA, in the IP we observed a similar low detection of this protein both in the negative control and in the sample with serum, suggesting that the binding of cytoplasmic proteins is not completely random.

The glycolytic enzyme enolase (Eno) identified as immunogenic as well, and already described in *Neisseria*[44] is another example of documented case of membrane association. No evidence of a surface association in other bacteria is currently

available for the majority of the remaining cytoplasmic proteins identified as immunogenic. Although we showed that the cytoplasmic proteins are present in small amount in GMMA, they might be more immunogenic than outer membrane proteins. 15 of the cytoplasmic proteins identified in the IP were found in protein-protein interaction studies conducted on *E. coli*[17], showing that these proteins interact with each other and with an inner membrane protein also identified in the IP. However, it is not clear how these interactions might happen through the inner and outer membranes. One hypothesis is that these proteins interact with some of the well-documented surface-associated cytoplasmic proteins, once released after the lysis of the bacteria, allowing their detection in the IP experiments.

Among the few periplasmic proteins identified, only one protein defined as 'Putative uncharacterized protein' showed an increase in the amount detected in the IP performed with serum, suggesting an interaction with the antibodies or antibody/antigen complexes. After the lysis of the bacteria, it might happen that some of the inner membrane proteins, interacting with periplasmic and cytoplasmic components, form complexes that remain linked to the OM immunogenic proteins. The binding of the antibody to the OM protein might, therefore, allow the recovery of unspecific proteins forming these complexes that are not directly recognized by the antibodies.

We observed that the majority of the cytoplasmic, inner membrane and outer membrane proteins are lost into the pellet after the ultra-centrifugation of the sample containing the Ab/Ag complexes. However, the presence of ribonucleases and the washing step does not seem to decrease the number of cytoplasmic and inner membrane proteins identified. As we did not check if the DNA/RNA were completely digested in the IP samples, we cannot conclude if the nucleic acids are not the key elements connecting the Ab/Ag complexes with hypothetical ribosomal complexes.

Iron siderophore receptors are expressed by bacteria in the body where iron limitation is present and have been shown to be important protective antigens against *Salmonella* infection[72]. We hypothesized that iron siderophore receptors might be protective also against other bacteria. Therefore, we were interested in increased expression of iron siderophore receptors in *Shigella* GMMA. Accordingly, we optimized the iron-concentration in the medium to allow optimal growth and

upregulation of iron-regulated proteins. 3 iron regulated proteins identified as FepA, FhuA, and CirA in 1-D and 2-D gel were found in both Ss –pSS Δ*tolR* and Ss –pSS Δ*tolR* Δ*msbB* GMMA. All showed a general lower expression in the Ss –pSS Δ*tolR* Δ*msbB* than in the Ss –pSS Δ*tolR* GMMA, in particular CirA. When GMMA were analyzed by GeLC-MS/MS we identified additional iron siderophore receptors (FecA, CjrC, CjrA, IutA, FhuE). Five of the iron-regulated proteins were conserved between *S. sonnei* and *S. flexneri* GMMA (FhuA, FepA, IutA, CirA, and FhuE). Interestingly, FepA, FhuA, and CirA were not detected in the 2-D WB but were identified in the immunoprecipitation approach. This indicates that, similarly to OmpC, the antibodies recognize native epitopes in these proteins that are destroyed by the denaturing conditions of 2-D WB. Interestingly, FepA was identified as low abundant in GMMA, but showed high amounts in the IP indicating that is highly immunogenic. Other iron-regulated proteins CjrC, FecA and IutA were also detected as immunogenic and were identified among the most abundant proteins in GMMA.

In summary, our results confirm that GMMA are enriched in outer membrane and periplasmic proteins and contain immunogenic surface proteins. We show that the immunoprecipitation approach can allow the identification of outer membrane proteins not detectable by 2-D WB, and that the iron regulated proteins are immunogenic. We identified experimental conditions that might be used for the optimization of the IP approach. Moreover, our results agree with previous studies showing membrane-association of several cytoplasmic proteins. They are consistent with the hypothesis that non-classically secreted proteins can be exported from the intact bacteria by alternative secretion models[10], and are in agreement with the interaction network containing conserved and essential protein complexes described in E*. coli*.

6.4. IMPLICATION FOR VACCINE DEVELOPMENT

These results show that *S. sonnei* and *S. flexneri* GMMA present a similar pattern of surface proteins, and the majority of the immunogenic proteins in *S. sonnei* have been identified among the common ones. This suggests that the proteins identified in *S. sonnei* might likely be immunogenic also in *S. flexneri*. However, further work is needed to analyze the conservation of these proteins at a gene level, to understand

how conserved are these proteins in other *Shigella* strains and how different they are from the commensal *E. coli*. The analysis of single nucleotide polymorphisms of the identified proteins will address this point and will be useful to predict if these differences might occur in epitopes exposed on the surface, thus affecting the antibody binding. Results from our lab, not presented in this work, showed that *S. flexneri* O-antigen negative GMMA protected against homologous challenge, and *S. sonnei* O-antigen negative GMMA showed potential for heterologous protection against *S. flexneri*. As the sera from survived mice recognized a similar patter of protein bands by 1-D WB (data not shown), it is likely that the immunogenic proteins we identified in *S. sonnei* might confer a protective immune response against homologous and heterologous infection. Therefore, the present work gives an essential contribution in the evaluation of the potentiality of GMMA as a broadly protective protein-based vaccine against *Shigella*.

In the future *Shigella* strains might be engineered to produce GMMA over-expressing specific protective antigens in order to obtain protection against different *Shigella* strains, as already described for *Neisseria meningitidis*[78]. In addition, multiple protective proteins might be over-expressed in *Shigella* strains in order to use GMMA as a broadly protective vaccine against *Shigella* and e.g. pathogenic *E. coli*. Alternatively, GMMA with O-antigen obtained from selected *Shigella* serotypes might be combined in order to confer broad protection in endemic regions against both O-antigen and conserved proteins. Most importantly, the potential use of GMMA as vaccine candidate will also requires further work to assess if the LPS molecules show reduced endotoxicity effect to levels acceptable tolerated in humans.

7. ACKNOWLEDGMENTS

I am the recipient of a Novartis Vaccines Institute for Global Health (NVGH) fellowship from the PhD Program in Bioscience and Biotechnologies of the University of Padua.

I would like to show my gratitude to all the people that supported this work within the last 3 years:

- **Christiane Gerke**, my supervisor and day by day mentor, for her helpfulness, for all the time we spent discussing the experimental results and the most appropriate strategies to achieve the objectives. Thanks Chris for your great efforts in explaining things simply, for your encouragements, for your suggestions during the writing of the thesis and for revising the English of this manuscript
- **Francesco**, for arousing me the interest in the proteomic field, for teaching me basic concepts of proteomic approaches and how to perform 2-dimensional electrophoresis, and for following my work during the first months of the PhD
- **Silvia**, for teaching me how to work in the BSL-2 lab, for pointing out every day the roles and correct behaviors to work in a safe and healthy enviroinment, and for explaining how the work in a Company looks like
- **Anna Maria**, for her work in optimizing fermentation process and media. Thanks Anna Maria for your helpfulness and advices, for taking care of me during the last months of my thesis writing, and for the good time we spent together during the last 3 years in Siena
- **Mariaelena**, for the good time we spent together during the last 3 years and the funny situations we had in the lab (and not only!) during the first months of our experience in NVGH. Thanks Mariaelena for your helpfulness, for the chats we had about the most and the less exiting results of our experiments and last, but not least, for sitting next to my desk
- **Isabella**, for her help in organizing and performing ELISA assay

- **Sara**, for the discussions we had about our results, for all the chats we had during some difficult moments of our life as PhD students, and for the time we spent together at WTSI
- **Eleonora,** for providing a fun environment in which to learn and to grow
- **Vito** and **Oliver**, for their suggestions and ideas in approaching the immunoprecipitation experiments, and for their availability in discussing these experiments
- **Calman MacLennan,** for providing a stimulating enviroinment within the Exploratory program at NVGH
- **Allan Saul**, for giving me the possibility to carry the PhD project in the '*Shigella* group', for sharing his knowledge and experience in different fields, and for the ideas and suggestions in approaching unexpected results
- **Massimiliano** and **Chiara**, for teaching me how to prepare samples for MALDI-TOF analysis and how to analyze the out coming data
- **Nathalie** and **Enzo**, for taking part into my committee meeting, for their availability and suggestions in developing the most interesting aspects of my thesis
- **Ilaria**, for taking pictures at the TEM and for being the link between the PhD students and the Universities
- **Jyoti Choudary**, for her helpfulness and for following my work day by day during the 3 months I spent at the Wellcome Trust Sanger Institute (WTSI). Thanks Jyoti for the time you spent in reading and correcting my thesis and for the good words you spent on my own and my way of working
- **Lu**, for teaching me how to prepare and process samples for LC-MS/MS and for her extreme helpfulness in showing me how Excel can facilitate the analysis of the complex MaxQuant data. Thanks Lu, for your help during my experience at the WTSI
- **Ulrich**, for the great helpfulness and the time he spent in writing bioinformatics scripts to help me in analyzing the MaxQuant proteomic data
- **Gordon Dougan**, for giving me the possibility to collaborate for 3 months with the Wellcome Trust Sanger Institute and obtain exiting results for my thesis
- **Prof. Cesare Montecucco**, my supervisor at the University of Padua

- *A special thank goes to Francesco, for supporting my choice to move to Siena and for his comprehension in the difficult moments during the course of this work and the thesis-writing. To my sister, for the enthusiasm with which she supported my choice in carrying the PhD project within NVGH. To my grandmother, for the loving support she always demonstrated during these 3 years away from my home-town. Lastly, and most importantly, I would like to thank my parents for supporting me during this experience, for their encouragements and especially for taking care of me throughout my thesis-writing period. To them I dedicate this thesis.*

8. REFERENCES

[1] F. Ahmed, M. Ansaruzzaman, E. Haque, M. R. Rao, and J. D. Clemens. Epidemiology of postshigellosis persistent diarrhea in young children. (2001). **20**: 525-530.

[2] S. Akira, K. Takeda, and T. Kaisho. Toll-like receptors: Critical proteins linking innate and acquired immunity. (2001). **2**: 675-680.

[3] K. Al-Hasani, F. Navarro-Garcia, J. Huerta, H. Sakellaris, and B. Adler. The immunogenic SigA enterotoxin of *Shigella flexneri 2a* binds to HEp-2 cells and induces fodrin redistribution in intoxicated epithelial cells. (2011). **4**: e8223.

[4] R. C. Alaniz, B. L. Deatherage, J. C. Lara, and B. T. Cookson. Membrane vesicles are immunogenic facsimiles of *Salmonella typhimurium* that potently activate dendritic cells, prime B and T cell responses, and stimulate protective immunity in vivo. (2007). J. Immunol. **179**: 7692-7701.

[5] C. J. A. Asensio, M. E. Gaillard, G. Moreno, D. Bottero, E. Zurita, M. Rumbo, P. van der Ley, A. van der Ark, and D. Hozbor. Outer membrane vesicles obtained from *Bordetella pertussis* Tohoma expressing the lipid A deacylase PagL as a novel acellular vaccine candidate. (2011). **29**: 1649-1656.

[6] S. Ashkenazi, J. H. Passwell, E. Harlev, D. Miron, R. Dagan, N. Farzan, R. Ramon, F. Majadly, D. A. Bryla, A. B. Karpas, J. B. Robbins, and R. Schneerson. Safety and immunogenicity of *Shigella sonnei* and *Shigella flexneri* 2a O-specific polysaccharide conjugates in children. (1999). **179**: 1565-1568.

[7] G. Ayala, L. Torres, M. Espinosa, G. Fierros-Zarate, V. Maldonado, and J. Melendez-Zajgla. External membrane vesicles from Helicobacter pylori induce apoptosis in gastric epithelial cells. (2006). FEMS Microbiol. Lett. **260**: 178-185.

[8] P. Bardhan, A. S. G. Faruque, A. Naheed, and D. A. Sack. Decrease in shigellosis-related deaths without *Shigella* spp.-specific interventions, Asia. (2010). **16**: 1718-1723.

[9] E. M. Barry, J. Wang, T. Wu, T. Davis, and M. M. Levine. Immunogenicity of multivalent *Shigella-ETEC* candidate vaccine strains in a guinea pig model. (2006). **24**: 3727-3734.

[10] J. D. Bendtsen, L. Kiemer, A. Fausboll, and S. Brunak. Non-classical proteins secretion in bacteria. (2005). **5**.

[11] F. Berlanda Scorza, A. M. Colucci, L. Maggiore, S. Sanzone, I. Ferlenghi, I. Pesce, M. Caboni, O. Rossi, N. Norais, V. Di Cioccio, A. Saul, and C. Gerke. High yield production process for *Shigella* Outer Membrane Particles. (2012).

[12] F. Berlanda Scorza, F. Doro, M. J. Rodriguez-Ortega, M. Stella, S. Liberatori, A. R. Taddei, L. Serino, D. M. Gomes, B. Nesta, M. R. Fontana, A. Spagnuolo, M. Pizza, N. Norais, and G. Grandi. Proteomics characterization of outer membrane vesicles from the extraintestinal pathogenic *Escherichia coli* Δ*tolR* IHE3034 mutant. (2008). Mol. Cell Proteomics. **7**: 473-485.

[13] A. Bernadac, M. Gavioli, J. C. Lazzaroni, S. Raina, and R. Lloubes. *Escherichia coli* tol-pal mutants form outer membrane vesicles. (1998). J. Bacteriol **180**: 4872-4878.

[14] T. J. Beveridge. Structures of gram-negative cell walls and their derived membrane vesicles. (1999). **181**: 4725-4733.

[15] P. J Boersema, E. Raijmakers, S. Lemeer, S. Mohammed, and A. J. R. Heck. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. (2009). **4**: 484-494.

[16] D. J. Brenner, G. R. Fanninig, G. V. Miklos, and A. G. Stigerwalt. Polynucleotide sequence relatedness among *Shigella* species. (1973). **23**: 1-7.

[17] G. Butland, J. M. Peregrin-Alvarez, J. Li, W. Yang, X. Yang, V. Canadien, A. Starostine, D. Richards, B. Beattle, N. Krogan, M. Davey, J. Parkinson, J Greenblatt, and A. Emili. Interaction network containing conserved and essential protein complexes in *Escherichia coli*. (2005). **433**: 531-537.

[18] A. I. Camacho, J. de Souza, S. Sanchez-Gomez, M. Pardo-Ros, J. M. Irache, and C. Gamazo. Mucosal immunization with *Shigella flexneri* outer membrane vesicles induced protection in mice. (2011). **29**: 8222-8229.

[19] T-W. Chang, Y-M. Lin, C-F. Wang, and Y-d. Liao. Outer membrane lipoprotein Lpp is Gram-negative bacterial cell surface receptor for cation antimicrobial peptides. (2012). **287**: 418-428.

[20] S. N. Chatterjee and J. Das. Electron microscopic observations on the excretion of cell-wall material by Vibrio cholerae. (1967). **49**: 1-11.

[21] O. Ciofu, J. Beveridge, J. Kadurugamuwa, J. Walther-Rasmussen, and N. Hoiby. Chromosomal beta-lactamase is packaged into membrane vesicles and secreted from *Pseudomonas aeruginosa*. (2000). **45**: 9-13.

[22] D. Cohen, S. Ashkenazi, M. Green, Y. Lerman, R. Slepon, G. Robin, N. Orr, D. N. Taylor, J. C. Sadoff, C. Chu, J. Shiloach, R. Schneerson, and J. B. Robbins. Safety and immunogenicity of investigational Shigella conjugate vaccines in Israeli volunteers. (1996). Infect. Immun. **64**: 4074-4077.

[23] D. Cohen, S. Ashkenazi, M. S. Green, M. Gdalevich, G. Robin, R. Slepon, M. Yavzori, N. Orr, C. Block, I. Ashkenazi, J. Shemer, D. N. Taylor, T. L. Hale, J.

C. Sadoff, *et al.* Double-blind vaccine-controlled randomised efficacy trial of an investigational *Shigella sonnei* conjugate vaccine in young adults. (1997). Lancet **349**: 155-159.

[24] S. J Cordwell. Technologies for bacterial surface proteomics. (2006). **9**: 320-329.

[25] J. Cox and M. Mann. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. (2008). **26**: 1367-1372.

[26] J. Cox, I. Matic, M. Hilger, N. Nagaraj, M Selbach, J. V. Olsen, and M. Mann. A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. (2009). **4**: 698-705.

[27] J. Cox, N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen, and M. Mann. Andromeda: a peptide search engine integrated into the MaxQuant environment. (2011). **10**: 1794-1805.

[28] H D'Hauteville, S. Khan, D. J. Maskell, A. Kussak, A. Weintraub, J. Mathison, R. J. Ulevitch, N. Wuscher, C. Parsot, and P. J. Sansonetti. Two *msbB* genes encoding maximal acylation of lipid A are required for invasive *Shigella flexneri* to mediate inflammatory rupture and destruction of the intestinal epithelium. (2002). **168**: 5240-5251.

[29] J. de la Cabada Bauche and H. L. Du Pont. New developments in traveler's diarrhea. (2011). **7**: 88-95.

[30] A. Derbise, B. Lesic, D. Dacheux, J. M. Ghigo, and E. Carniel. A rapid and simple method for inactivating chromosomal genes in *Yersinia*. (2003). **38**: 113-116.

[31] I. W. Devoe and J. E. Gilchrist. Release of endotoxin in the form of cell wall blebs during in vitro growth of *Neisseria meningitidis*. (1973). **138**: 1156-1167.

[32] D. W. Dorward and C. F. Garon. DNA-binding proteins in cells and membrane blebs of *Neisseria gonorrhoeae*. (1989). **171**: 2499-2505.

[33] D. W. Dorward, T. G. Schwan, and C. F. Garon. Immune capture and detection of Borrelia burgdorferi antigens in urine, blood, or tissues from infected ticks, mice, dogs, and humans. (1991). **29**: 1162-1170.

[34] A. J. M. Driessen, E. H. Manting, and van der Does.C. The structural basis of protein targeting and translocation in bacteria. (2001). **8**: 492-498.

[35] H. L. Du Pont, M. M. Lewin, R. B. Hornick, and S. B. Formal. Inoculum size in shigellosis and implications for expected mode of transmission. (1989). **159**: 1126-1128.

[36] L. Duncan, M. Yoshioka, F. Chandad, and D. Grenier. Loss of lipopolysaccharide receptor CD14 from the surface of human macrophage-like cells mediated by *Porphyromonas gingivalis* outer membrane vesicles. (2004). **36**: 319-325.

[37] S. Dutta, K. Iida, A. Takade, Y. Meno, G. B. Nair, and S. Yoshida. Release of Shiga toxin by membrane vesicles in Shigella dysenteriae serotype 1 strains and in vitro effects of antimicrobials on toxin production and release. (2004). Microbiol. Immunol. **48**: 965-969.

[38] U. S. Eggert, N. Ruiz, B. V. Falcone, A. A. Branstrom, R. C. Goldman, T. J. Silhavy, and D. Kahne. Genetic basis for activity differences between vancomycin and glycolipid derivatives of vancomycin. (2001). **294**: 361-364.

[39] T. N. Ellis and M. J. Kuehn. Virulence and immunomodulatory roles of bacterial outer membrane vesicles. (2010). **74**: 81-97.

[40] T. N. Ellis, S. A. Leiman, and M. J. Kuehn. Naturally produced outer membrane vesicles from *Pseudomonas aeruginosa* elicit a potent innate immune response via combined sensing of both LPS and protein components. (2010).

[41] M. Ensgraber and M. Loos. A 66-kilodalton heat shock protein of *Salmonella typhimurium* is responsible for binding of the bacterium to intestinal mucus. (1992). **60**: 3072-3078.

[42] W. H. Ewing. *Shigella* nomenclature. (1949). **57**: 633-638.

[43] M. I. Fernandez, A. Thuizat, T. Pedron, M. Neutra, A. Phalipon, and P. J. Sansonetti. A newborn mouse model for the study of intestinal pathogenesis of shigellosis. (2003). **5**: 481-491.

[44] G. Ferrari, I. Garaguso, J. Adu-Bobie, F. Doro, A. R. Taddei, A. Biolchi, B. Brunelli, M. M. Giuliani, M. Pizza, N. Norais, and G. Grandi. Outer membrane vesicles from group B *Neisseria meningitidis* Δ*gna33* mutant: proteomic and immunological comparison with detergent-derived outer membrane vesicles. (2006). Proteomics. **6**: 1856-1866.

[45] C. Ferreccio, V. Prado, A. Ojeda, M. Cayyazo, P. Abrego, L. Guers, and M. M. Levine. Epidemiologic patterns of acute diarrea and endemic *Shigella* infections in children in a poor periurban settino in Santiago, Chile. (1991). **134**: 614-627.

[46] R. Fiocca, V. Necchi, P. Sommi, V. Ricci, J. Telford, T. L. Cover, and E. Solcia. Release of Helicobacter pylori vacuolating cytotoxin by both a specific secretion pathway and budding of outer membrane vesicles. Uptake of released toxin and vesicles by gastric epithelium. (1999). **188**: 220-226.

[47] S. B. Formal, T. H. Kent, H. C. May, A. Palmer, S. Falkow, and E. H. LaBrec. Protection of monkeys against experimental shigellosis with a living

attenuated oral polyvalent dysentery vaccine. (1966). J. Bacteriol. **92**: 17-22.

[48] L. F. Fries, A. D. Montemanaro, C. P. Mallett, D. N. Taylor, T. L. Hale, and G. H. Lowell. Safety and immunogenicity of a proteasome-*Shigella flexneri 2a* lipopolysaccharide vaccine administered intranasally to healthy adults. (2001). **97**: 4545-4553.

[49] N. Furuta, H. Takeuchi, and A. Amano. Entry of *Porphyromonas gingivalis* outer membrane vesicles into epithelial cells causes cellular functional impairment. (2009). **77**: 4761-4770.

[50] M. Futai, T. Noumi, and M. Maeda. ATP synthase (H+-ATPase): results by combined biochemical and molecular biological approaches. (1989). **58**: 111-136.

[51] M. Galdiero, A. Folgore, M. Molitierno, and R. Greco. Porins and lipopolysaccharide (LPS) from *Salmonella typhimurium* induce leucocyte transmigration though human endothelial cells in vitro. (1999). **116**: 453-461.

[52] F. Galka, S. N. Wai, H. Kusch, S. Engelmann, M. Hecker, B. Schmeck, S. Hippenstiel, B. E. Uhlin, and M. Steinert. Proteomic characterization of the whole secretome of *Legionella pneumophila* and functional analysis of Outer Membrane Vesicles. (2008). **76**: 1825-1836.

[53] R. A. Garduno, G. Faulkner, M. A. Trevors, N. Vats, and P. S. Hoffman. Immunolocalization of Hsp60 in *Legionella pneumophila*. (1998). **180**: 505-513.

[54] R. A. Garduno, E. Garduno, and P. S. Hoffman. Surface-associated Hsp60 chaperonin of *Legionella pneumophila* mediates invasion in HeLa cell model. (1998). **66**: 4602-4610.

[55] J. Gorden and P. L. Small. acid resistance in enteric bacteria. (1993). **61**: 364-367.

[56] T. L. Hale. Genetic basis of virulence in *Shigella* species. (1991). **55**: 206-224.

[57] C Hennequin, F. Porcheray, Waligora-Dupriet A.J., A. ollignon, M. C. arc, P. ourlioux, and T. Karjalainen. GroEL (Hsp60) of *Clostridium difficile* is involved in cell adherence. (2001). **147**: 87-96.

[58] D. Hoekstra, J. W. van der Laan, L. de Leij, and B. Witholt. Release of outer membrane fragments from normally growing *Escherichia coli*. (1976). **455**: 889-899.

[59] A. L. Horstman and M. J. Kuehn. Enterotoxigenic Escherichia coli secretes active heat-labile enterotoxin via outer membrane vesicles. (2000). J. Biol. Chem. **275**: 12489-12496.

[60] A. L. Horstman and M. J. Kuehn. Bacterial surface association of heat-labile enterotoxin through lipopolysaccharide after secretion via the general secretory pathway. (2002). J. Biol. Chem. **277**: 32538-32545.

[61] M. Ingersoll, E. A. Groisman, and A. Zychlinsky. Pathogenicity island of *Shigella*. (2002). **264**: 49-65.

[62] Y. Ishihama, Y. Oda, T. Tabata, T. Sato, T. Nagasu, J. Rappsilber, and M. Mann. Exponentially Modified Protein Abundance Index (emPAI) for extimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. (2005). **4**: 1265-1272.

[63] D. Islam, L. Bandholtz, J. Nilsson, H. Wigzell, B. Christensson, B. Agerberth, and G. Gudmundsson. Downregulation of bactericidal peptides in enteric infections: a novel immune escape mechanism with bacterial DNA as a potential regulator. (2001). **7**: 180-185.

[64] S. Ismail, M. B. Hampton, and J. I. Keenan. *Helicobacter pylori* outer membrane vesicles modulate proliferation and interleukin-8 production by gastric epithelial cells. (2003). **71**: 5670-5675.

[65] Y. Jiang, F. Yang, X. Zhang, J. Yang, L. Chen, Y. Yan, H. Nie, Z. Xiong, J. Wang, J. Dong, Y. Xue, X. Xu, Y. Zhu, S. Chen*, et al.* The complete sequence and analysis of the large virulence plasmid pSS of *Shigella sonnei*. (2005). **54**: 149-159.

[66] J. L. Kadurugamuwa and T. J. Beveridge. Virulence factors are released from Pseudomonas aeruginosa in association with membrane vesicles during normal growth and exposure to gentamicin: a novel mechanism of enzyme secretion. (1995). J. Bacteriol. **177**: 3998-4008.

[67] J. L. Kadurugamuwa and T. J. Beveridge. Membrane vesicles derived from Pseudomonas aeruginosa and Shigella flexneri can be integrated into the surfaces of other gram-negative bacteria. (1999). Microbiology **145 (Pt 8)**: 2051-2060.

[68] J. Kahnt, K. Aguiluz, J. Koch, A. Treuner-Lange, A. Konovalova, S. Huntley, M. Hoppert, L. Sogaard-Andersen, and R. Hedderich. Profiling the outer membrane proteome during growth and development of the social bacterium *Myxococcus xanthus* by selective biotinylation and analyses of outer membrane vesicles. (2010). **9**: 5197-5208.

[69] A. Kamaguchi, K. Nakayama, S. Ichiyama, R. Nakamura, T. Watanabe, M. Ohya, H. Baba, and T. Ohyama. Effect of Porphyromonas gingivalis vesicles on coaggregation of Staphylococcus aureus to oral microorganisms. (2003). **47**: 485-491.

[70] R. Kaminski, K. R. Turbyfill, and E. V. Oaks. Mucosal adjuvant properties of the *Shigella* invasin complex. (2006). **74**: 2856-2866.

[71] K. Kaneda, T. Masuzawa, K. Yasugami, T. Suzuki, Y. Suzuki, and Y. Yanagihara. Glycosphingolipid-binding protein of *Borrelia burgdorferi sensu lato*. (1997). **65**: 3180-3185.

[72] T. Kaneshige, K. Yaguchi, and T. Ohgitani. Siderophore receptor IroN is an important protective antigen against *Salmonella* infection in chickens. (2009). **53**: 563-567.

[73] S. Kato, Y. Kowashi, and D. R. Demuth. Outer membrane-like vesicles secreted by *Actinobacillus actinomycetemcomitans* are enriched in leukotoxin. (2002). **32**: 1-13.

[74] D. E. Katz, T. S. Coster, M. K. Wolf, F. C. Trespalacios, D. Cohen, G. Robins, B. Hartman, M. M. Venkatesan, D. N. Taylor, and T. L. Hale. Two studies evaluating the safety and mmunogenicity of a live, attenuated Shigella flexneri 2a vaccine (SC602) and excretion of vaccine organisms in North American volunteers. (2004). **72**: 923-930.

[75] J. Keenan, T. Day, S. Neal, B. Cook, G. Perez-Perez, R. Allardyce, and P. Bagshaw. A role for the bacterial outer membrane in the pathogenesis of Helicobacter pylori infection. (2000). FEMS Microbiol. Lett. **182**: 259-264.

[76] N. C. Kesty and M. J. Kuehn. Incorporation of heterologous outer membrane and periplasmic proteins into *Escherichia coli* outer membrane vesicles. (2004). J. Biol. Chem. **279**: 2069-2076.

[77] R. Kitagawa, A. Takaya, M. Ohya, Y. Mizunoe, A. Takade, S. Yoshida, E. Isogai, and T. Yamamoto. Biogenesis of *Salmonella enterica* serovar Typhimurium membrane vesicles provoked by induction of PagC. (2011). J. Bacteriol **192**: 5645-5656.

[78] O. Koeberling, A. Seubert, and D. M. Granoff. Bactericidal antibody responses elicited by a meningococcal outer membrane vesicle vaccine with overexpressed factor H-binding protein and genetically attenuated endotoxin. (2008). **198**: 262-270.

[79] R. Koebnik, K. P. Locher, and P. Van Gelder. Structure and function of bacterial outer membrane proteins: barrels in a nutshell. (2000). **37**: 239-253.

[80] K. Kotloff, J. P. Nataro, G. A. Losonsky, S. S. Wasserman, T. L. Hale, D. N. Taylor, J. C. Sadoff, and M. M. Levine. A modified *Shigella* volunteer challenge model in which the inoculum is administered with bicarbonate buffer: clinical experience and implications for *Shigella* infectivity. (1995). **13**: 1488-1494.

[81] K. L. Kotloff, F. R Noriega, G. A. Losonsky, M. B. Sztein, S. S. Wasserman, J. P. Nataro, and M. M. Lewin. Safety, immunogenicity, and transmissibility in humans of CVD 1203, a live oral Shigella flexneri 2a vaccine candidate attenuated by deletions in aroA and virG. (1996). **64**: 4542-4548.

[82] K. L. Kotloff, J. P. Winickoff, B. Ivanoff, J. D. Clemens, D. L. Swerdlow, P. J. Sansonetti, G. K. Adak, and M. M. Levine. Global burden of *Shigella* infections: implications for vaccine development and implementation of control strategies. (1999). Bull. World Health Organ **77**: 651-666.

[83] M. J. Kuehn and N. C. Kesty. Bacterial outer membrane vesicles and the host-pathogen interaction. (2005). Genes Dev. **19**: 2645-2655.

[84] A. Kulp and M. J. Kuehn. Biological functions and biogenesis of secreted bacterial outer membrane vesicles. (2010). Annu. Rev. Microbiol. **64**: 163-184.

[85] M.-N. Kweon. Shigellosis: the current status of vaccine development. (2008). Curr. Opin. Infect. Dis. **21**: 313-318.

[86] O. Launay, C. Sadorge, N. Jolly, B. Poirier, S. Béchet, D. van der Vliet, V. Seffer, N. Fenner, K. Dowling, R. Giemza, J. Johnson, A. Ndiaye, M. Vray, P. J. Sansonetti, *et al.* Safety and immunogenicity of SC599, an oral live attenuated *Shigella dysenteriae* type-1 vaccine in healthy volunteers: Results of a phase 2, randomized, double-blind placebo-controlled trial. (2009). **27**: 1184-1191.

[87] T. Le Gall, M. Mavris, M. C. Martino, M. L. Bernardini, E. Denamur, and C. Parsot. Analysis of virulence plasmid gene expression defines three classes of effectors in the type III secretion system of *Shigella flexneri*. (2005). **151**: 951-962.

[88] K. Le-Barillec, J. G. Magalheaes, E. Corcuff, A. Thuizat, P. J. Sansonetti, A. Phalipon, and J. P. Di Santo. Roles for T and NK cells in the innate immune response to *Shigella flexneri*. (2005). **175**: 1735-1740.

[89] E. Y. Lee, J. Y. Bang, G. W. Park, D. S. Choi, J. S. Kang, H. J. Kim, K. S. Park, J. O. Lee, Y. K. Kim, K. H. Kwon, K. P. Kim, and Y. S. Gho. Global proteomic profiling of native outer membrane vesicles derived from Escherichia coli. (2007). Proteomics. **7**: 3143-3153.

[90] M. M. Levine, K. L. Kotloff, E. M. Barry, M. F. Pasetti, and M. B. Sztein. Clinical trials of *Shigella* vaccines: two steps forward and one step back on a long, hard road. (2007). Nat. Rev. Microbiol. **5**: 540-553.

[91] B. Lipinska, M. Zylicz, and C. Georgopoulos. The HtrA (DegP) protein, essential for *Escherichia coli* survival at high temperatures, is an endopeptidase. (1990). **172**: 1791-1797.

[92] S. Makino, C. Sasakawa, and M. Yoshikawa. Genetic relatedness of the basic replicon of the virulence plasmid in shigellae and enteroinvasive *Escherichia coli*. (1988). **5**: 267-274.

[93] J. C. Malinverni, J. Werner, S. Kim, Sklar.J.G., D. Kahne, R. Misra, and T. J. Silhavy. YfiO stabilizes the YaeT complex and is essential for outer membrane protein assembly in *Escherichia coli*. (2006). **61**: 151-164.

[94] L. M. Mashburn and M. Whiteley. Membrane vesicles traffic signals and facilitate group activities in a prokaryote. (2005). **437**: 422-425.

[95] V. Masignani, E. Balducci, F. Di Marcello, S. Serruto D. Savino, D. Veggi, S. Bambini, M. Scarselli, B. Arico, M. Comanducci, J. Adu-Bobie, M. M. Giuliani, R. Rappuoli, and M. Pizza. NarE:a novel ADP-ribosyl-transferase from *Neisseria meningitidis*. (2003). **50**: 1055-1067.

[96] M. E. Maxson and A. J. Darwin. Identification of inducers of the *Yersinia enterocolitica* phage shock protein system and comparison to the regulation of the RpoE and Cpx extracytoplasmic stress responses. (2004). **186**: 4199-4208.

[97] A. J. McBroom and M. J. Kuehn, "Outer Membrane Vesicles,"in *EcoSal-Escherichia coli and Salmonella:Cellular and molecular biology,* R.Curtiss III et al. ed.(ASM Press, Washington, DC., 2005).

[98] A. J. McBroom and M. J. Kuehn. Release of outer membrane vesicles by Gram-negative bacteria is a novel envelope stress response. (2007). Mol. Microbiol. **63**: 545-558.

[99] M. McConnel, C. Rumbo, G. Bou, and J. Pachon. Outer membrane vesicles as an acellular vaccine against *Acinetobacter baumannii*. (2011). **29**: 5705-5710.

[100] M. P. Molloy, B. R. Herbert, M. B. Slade, T. Rabilloud, A. S. Nouwens, K. L. Williams, and A. A. Gooley. Proteomic analysis of the *Escherichia coli* outer membrane. (2000). Eur. J. Biochem **267**: 2871-2881.

[101] M. Muralinath, M. J. Kuehn, K. L. Roland, and R. Curtiss III. Immunization with *Salmonella enterica* serovar *typhimurium*-derived outer membrane vesicles delivering the pneumococcal protein SpA confer protection against challenge with *Streptococcus pneumoniae*. (2011). **79**: 887-894.

[102] A. I. Nesvizhskii and R. Aebersold. Interpretation of shotgun proteomic data: the protein inference problem. (2005). **4**: 1419-1440.

[103] F. R Noriega, F. M. Liao, D. R. Maneval, S. Ren, S. B. Formal, and M. M. Levine. Strategy for cross-protection among *Shigella flexneri* serotypes. (1999). **67**: 782-788.

[104] H. Ochman, T. S. Whittam, D. A. Caugant, and R. K. Selander. Enzyme polymorphism and genetic population structure in *Escherichia coli* and *Shigella*. (1983). **129**: 2715-2726.

[105] N. Orr, G. Robin, D. Cohen, R. Arnon, and G. H. Lowell. Immunogenicity and efficacy of oral or intranasal *Shigella flexneri 2a* and *Shigella sonnei*

proteosome-lipopolysaccharide vaccines in animal models. (1993). **61**: 2390-2395.

[106] S. B. Park, H. B. Jang, S. W. Nho, I. S. Cha, J. Hikima, M. Ohtani, T. Aoki, and T. S. Jung. Outer membrane vesicles as a candidate vaccine against Edwardsiellosis. (2011). **6**: e17629.

[107] C. Parsot and P. J. Sansonetti. Invasion and the pathogenesis of *Shigella* infections. (1996). **209**: 25-42.

[108] J. H. Passwell, S. Ashkenzi, Y. Banet-Levi, R. Ramon-Saraf, N. Farzam, L. Lerner-Geva, H. Even-Nir, B. Yerushalmi, C. Chu, J. Shiloach, J. B. Robbins, R. Schneerson, and The Israeli Shigella Study Group. Age-related efficacy of *Shigella* O-specific polysaccharide conjugates in 1-4-year-old Israeli children. (2010). Vaccine **28**: 2231-2235.

[109] J. J. Perdomo, J. M. Cavaillon, M. Huerre, H. Ohayon, P. Gounon, and P. J. Sansonetti. Acute inflammation causes epithelial invasion and mucosal destruction in experimental shigellosis. (1994). **180**: 1307-1319.

[110] J. J. Perdomo, P. Gounon, and P. J. Sansonetti. Polymorphonuclear leukocyte transmigration promotes invasion of colonic epithelial monolayer by *Shigella flexneri*. (1994). **93**: 633-643.

[111] A. Phalipon, C. Costachel, C. Grandjean, A. Thuizat, C. Guerriero, M. Tanguy, F. Nato, B. Vulliea-Le Normand, F. Bélot, K. Wright, V. Marcel-Peyre, P. J. Sansonetti, and L. A. Mulard. Characterization of functional oligosaccharide mimics of the *Shigella flexneri* serotype 2a O-antigen: implications for the development of a chemically defined glycoconjugate vaccine. (2006). **176**: 1686-1694.

[112] N. V. Prasadarao. Identification of *Escherichia coli* outer membrane protein A receptor on human brain microvascular endothelial cells. (2002). **70**: 4556-4563.

[113] G. M. Pupo, D. K. R. Karaolis, R. Lan, and P. R. Reeves. Evolutionary relationships among pathogenic and nonpathogenic Escherichia coli strains inferred from multilocus enzyme electrophoresis and mdh sequence studies. (1997). **65**: 2685-2692.

[114] G. M. Pupo, R. Lan, and P. R. Reeves. Multiple independent origins of *Shigella* clones of *Escherichia col*i and convergent evolution of many of their characteristics. (2000). **97**: 10567-10572.

[115] K. M. Rahman, S. El Arifeen, K. Zaman, M. Rahman, R. Raqib, M. Yunus, N. Begum, Md. S. Islam, B. M Sohel, M. Rahman, M. Venkatesan, T. L. Hale, D. W. Isenbarger, P. J. Sansonetti*, et al.* Safety, dose, immunogenicity, and transmissibility of an oral attenuated *Shigella flexneri 2a* vaccine candidate (SC602) among healthy adults and school children in Matlab, Bangladesh. (2011). **29**: 1347-1354.

[116] J. Rappsilber, U. Ryder, A. I. Lamond, and M. Mann. Large-scale proteomic analysis of the human spliceosome. (2002). **12**: 1231-1245.

[117] J. B. Robbins, C. Chu, and R. Schnnerson. Hypothesis for vaccine development:protective immunity to enteric deseases caused by nontyphoidal salmonellae and shigellae may be conferred by serum IgG to the O-specific polysaccharide of their lipopolysaccharide. (1992). **15**: 346-361.

[118] R. Roberts, G. Moreno, D. Bottero, M. E. Gaillard, M. Fingermann, A. Graieb, M. Rumbo, and D. Hozbor. Outer membrane vesicles as acellular vaccine against pertussis. (2008). **26**: 4639-4646.

[119] C. Sadorge, A. Ndiaye, N. Beveridge, S. Frazer, R. Giemza, N. Jolly, J. Johnson, H. Liddy, C. A. Cosgrove, P. Allavena, A. Mantovani, S. Béchet, A. Fontaine-Thompson, G. E. Griffin*, et al.* Phase 1 clinical trial of live attenuated *Shigella dysenteriae* type-1 deltaicsA deltaent deltafep deltastxA:HgR oral vaccine SC599 in healthy human adult volunteers. (2008). **26**: 978-987.

[120] T. Sakaguchi, H. Johler, X. Gu, B. A. McCormick, and H. C. Reinecker. *Shigella flexneri r*egulates tight junctions-associated proteins in human intestinal epithelial cells. (2002). **4**: 367-381.

[121] P. J. Sansonetti, D. J. Kopecko, and S. B. Formal. Involvement of a plasmid in the invasive ability of *Shigella flexneri*. (1982). **35**: 852-860.

[122] V. Santoni, M. Molloy, and T. Rabilloud. Membrane proteins and proteomics: un amour impossible? (2000). **21**: 1054-1070.

[123] S. Schild, E. J. Nelson, and A. Camilli. Immunization with *Vibrio cholerae* outer membrane vesicles induces protective immunity in mice. (2008). Infect. Immun. **76**: 4554-4563.

[124] G. N. Schroeder and H. Hilbi. Molecular pathogenesis of *Shigella* spp.: controlling host cell signaling, invasion, and death by tipe III secretion. (2008). **21**: 134-156.

[125] R. Schuch and A. T. Maurelli. MxiM and MxiJ, base elements of the Mxi-Spa type III secretion system of *Shigella*, interact with and stabilize the MxiD secretin in the cell envelope. (2001). J. Bacteriol **183**: 6991-6998.

[126] H. Schultz, J. Hume, D. S. Zhang, T. L. Gioannini, and J. P. Weiss. A novel role for the bactericidal/permeability increasing protein in interactions of Gram-negative bacterial outer membrane blebs with dendritic cells. (2007). **179**: 2477-2484.

[127] D. H. Shim, T. Suzuki, S. Y. Chang, S. M. Park, P. J. Sansonetti, C. Sasakawa, and M. N. Kweon. New animal model of shigellosis in the Guinea pig: its usefulness for protective efficacy studies. (2007). **178**: 2476-2482.

[128] J. E. Somerville, Jr., L. Cassiano, B. Bainbridge, M. D. Cunningham, and R. P. Darveau. A novel *Escherichia coli* lipid a mutant that produces an antiinflammatory lipopolysaccharide. (1996). **97**: 359-365.

[129] R. Szu*, et al.*, "Polysaccharide-Based Conjugate Vaccines for Enteric Bacterial Infections: Typhoid Fever, Notypoidal Salmonellosis, Shigellosis, Cholera, *Escherichia Coli 0157*,"in *New Generation Vaccines,* Fourth ed.(Levine,M.M., New York: Marcel Dekker, 2009), pp.489-496.

[130] C. Tam and D. Missiakas. Changes in lipopolysaccharide structure induce the sigma^{E}-dependent response of *Escherichia coli*. (2005). **55**: 1403-1412.

[131] L. A. Taylor and R. E. Rose. A correction in the nucleotide sequence of the Tn903 kanamycin resistance determinant in pUC4K. (1988). **16**: 358.

[132] D. Tribble, R. Kaminski, J. Cantrell, M. Nelson, C. Porter, S. Baqar, C. Williams, R. Arora, J. Saunders, M. Ananthakrishnan, J. Sanders, G. Zaucha, R. Turbyfill, and E. Oaks. Safety and immunogenicity of a Shigella flexneri 2a Invaplex 50 intranasal vaccine in adult volunteers. (2010). Vaccine **28**: 6076-6085.

[133] P. Trieu-Cuot, C. Poyart-Salmeron, C. Carlier, and P. Courvalin. Nucleotide sequence of the erythromycin resistance gene of the conjugative transposon Tn1545. (1990). **18**: 3660.

[134] L. Uli, L. CAstellanos-Serra, L. Betancourt, F. Dominguez, R. Barbera, F. Sotolongo, G. Guillen, and R. Pajon. Outer membrane vesicles of the VA-MENGOC-BC vaccine^{R} against serogroup B of *Neisseria meningitidis*: analysis of protein components by two-dimensional gel electrophoresis and mass spectrometry. (2006). **6**: 3389-3399.

[135] L. L. Van De Verg, C. P. Mallett, H. H. Collins, T. Larsen, C. Hammack, and T. Hale. Antibody and cytokine responses in a mouse pulmonary model of *Shigella flexneri* serotype 2a infection. (1995). **63**: 1947-1954.

[136] A. Vanet and A. Labigne. Evidence for specific secretion rather than autolysis in the release of some *Helicobacter pylori* proteins. (1998). **66**: 1023-1027.

[137] L. von Seidlein, D. R. Kim, M. Ali, H. Lee, X. Wang, V. D. Thiem, D. G. Canh, W. Chaicumpa, M. D. Agtini, A. Hossain, Z. A. Bhutta, C. Mason, O. Sethabutr, K. Talukder*, et al.* A multicentre study of *Shigella* diarrhoea in six Asian countries: disease burden, clinical manifestations, and microbiology. (2006). PLoS. Med. **3**: e353.

[138] S. S. Way, A. C. Borczuk, R. Dominitz, and M. B. Goldberg. An essential role for gamma interferon in innate resistance to *Shigella flexneri* infection. (1998). **66**: 1342-1348.

[139] J. Wensink and B. Witholt. Outer-membrane vesicles released by normally growing Escherichia coli contain very little lipoprotein. (1981). Eur. J. Biochem. **116**: 331-335.

[140] J. N. Williams, P. J. Skipp, H. E Humphries, M. Christodoulides, C. D. O'Connor, and J. E. Heckels. Proteomic analysis of outer membranes and vesicles from wild-type serogroup B *Neisseria meningitidis* and a lipopolysaccharide-deficient mutant. (2007). **75**: 1364-1372.

[141] H. Yamaguchi, T. Osaki, H. Taguchi, T. Hanawa, T. Yamamoto, and S. Kamiya. Flow cytometric analysis of the heat shock protein Hsp60 expressed on the cell surface of *Helicobacter pylori*. (1996). **45**: 270-277.

[142] R. C. Yashroy. Mechanism of infection of a human isolate Salmonella (3,10:r:-) in chicken ileum: ultrastructural study. (2007). **126**: 558-566.

[143] T. Ying, H. Wang, M. Li, J. Wang, J. Wang, Z. Shi, E. Feng, X. Liu, G. Su, K. Wei, X. Zhang, P. Huang, and L. Huang. Immunoproteomics of outer membrane proteins and extracellular proteins of *Shigella flexneri* 2a 2457T. (2005). **5**: 4777-4793.

[144] K. Yokoyama, T. Horii, T. Yamashino, S. Hashikawa, S. Barua, T. Hasegawa, H. Watanabe, and M. Ohta. Production of shiga toxin by Escherichia coli measured with reference to the membrane vesicle-associated toxins. (2000). **192**: 139-144.

[145] Z. Zhang, L. Jin, G. Champion, K. B. Seydel, and S. L. Jr Stanley. *Shigella* infection in a SCID mouse-human intestinal xenograft model: role for neutrophils in containing bacterial dissemination in human intestine. (2001). **69**: 3240-3247.

[146] G. Zhao, L. Zhu, E. Feng, X. Cao, N. Shang, X. Liu, X. Liao, T. Ying, J. Wang, H. Chen, and H. Wang. A novel anti-virulence gene revealed by proteomic analysis in *Shigella flexneri 2a*. (2010). **8**: 1-9.

[147] L. Zhou, D. E. Srisatjaluk, and R. J Doyle. On the origin of membrane vesicles in gram-negative bacteria. (1998). **163**: 223-228.

[148] W. D. Zollinger, M. A. Donets, D. H. Schmiel, V. B. Pinto, J. E. Labrie III, E. E. Moran, B. L. Brandt, B. Ionin, R. Marques, M. Wu, P. Chen, M. B. Sroddard, and P. B. Keiser. Design and evaluation in mice of a broadly protective meningococcal group B native outer membrane vesicle vaccine. (2010). **28**: 5057-5067.

[149] A. Zychlinsky, M. C. Prévost, and P. J. Sansonetti. *Shigella flexneri* induces apoptosis in infected macrophages. (1992). **358**: 167-168.