Characterization of the outer membrane subproteome of the virulent strain *Salmonella* Typhimurium SL1344

Ferrer-Navarro M.[1#], Ballesté-Delpierre C.[1#], Vila J.[1,2] , Fàbrega A.[1¥]

[1] ISGlobal, Barcelona Ctr. Int. Health Res. (CRESIB), Hospital Clinic-Universitat de Barcelona, Barcelona Spain.

[2] Department of Clinical Microbiology, Hospital Clínic, School of Medicine, University of Barcelona, Barcelona, Spain

[#] These authors contributed equally.

[¥] Present address: Department of Clinical Microbiology, Hospital Vall d'Hebron, Vall d'Hebron Institut Recerca (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain

Corresponding author:

Dr. Anna Fàbrega

Department of Clinical Microbiology, Hospital Vall d'Hebron, Vall d'Hebron Institut Recerca (VHIR), Passeig de la Vall d'Hebron 119-129, 08035 Barcelona, Spain

Phone: +34 93 274 68 67

Fax: +34 93 274 68 01

E-mail: anna.fabrega@vhir.org

Keywords: *Salmonella* Typhimurium, outer membrane, OM subproteome, LC-MS/MS.

Abstract

Outer membrane proteins (OMPs) play an important role in the interaction of bacterial pathogens with host cells. Indeed, some OMPs from different Gram-negative bacteria have been recognized as important virulence factors for host immune recognition. This scenario has led to the study of the outer membrane (OM) subproteome of pathogenic bacteria as an essential step for gaining insight into the mechanisms of pathogenesis and for the identification of virulence factors. Although progress in the characterization of the OM has recently been reported, detailed protein composition of this subcellular localization has not been clearly defined for most pathogens. *Salmonella enterica* serovar Typhimurium is not only a leading cause of human gastroenteritis in high-income countries but is also one of the main causes of invasive non-typhoidal salmonellosis (iNTS) in middle- and low-income countries. The incidence of non-typhoidal salmonellosis is increasing worldwide, causing millions of infections and deaths among humans each year. Regrettably, antimicrobial resistance to a broad spectrum of antibiotics is common among non-Typhi *Salmonella* strains. Therefore, the development of vaccines targeting this leading invasive pathogen is warranted. In the present study we have identified the OM protein profile of the virulent *S.* Typhimurium strain SL1344 by means of sarkosyl extraction.

1. Introduction

Salmonella enterica serovar Typhimurium is a common facultative intracellular pathogen that causes food-borne gastroenteritis around the world. In high-income countries such gastrointestinal disease is rarely life-threatening and is normally self-limiting. Nausea, vomiting, profuse watery diarrhea, and abdominal pain are the usual clinical manifestations in immune-competent individuals [1]. According to the 2013 annual report of the Rapid Alert System for Food and Feed (RASFF) of the European Union, there were about 120 alerts for *Salmonella* contaminations in food other than poultry in Europe during 2013, while in poultry meat the number of alerts had tripled compared to the previous year [2]. In addition, according to the European Food Safety Authority (EFSA), over 100,000 human cases are reported only in the European Union each year and it has been estimated that the overall economic burden of human salmonellosis could be as high as 3 billion euros a year [3]. In low- and middle-income countries, however, *S.* Typhimurium*,* together with *Salmonella* Enteritidis, are responsible for a more serious manifestation of the disease through a form of invasive illness, invasive non-typhoidal *Salmonella* (iNTS) disease, which is considered a major public health problem in these countries [4]. iNTS disease is a neglected disease that is endemic in sub-Saharan Africa, but is also significantly present in Asia. Data of the global burden of disease estimates are not currently available, although associated case fatality rates range from 20-25% [5]. Clinical manifestations are diverse including fever, hepatosplenomegaly and respiratory symptoms as the most common, whereas typical features of enterocolitis are often absent [6]. In sub-Saharan African countries non-typhoidal *Salmonella* are either the leading or next most common pathogen isolated from blood after pneumococcus [7], for which vaccines are available and currently implemented in the region. Unfortunately no vaccines against iNTS are currently available, although some attempts are ongoing [4]. Moreover, among the most worrisome facts is the increasing resistance to antimicrobials which notably limits the clinical success of the present therapeutic options [8].

S. Typhimurium has evolved to survive in adverse environments. After ingestion of this bacterium by a mammalian host through contaminated food, or other vehicles, it progresses through the diverse environments of the gastrointestinal tract until reaching the intestine. Here it interacts with the wall of the intestine, and the invasion process takes place through expression of specific proteins involved in the translocation of *S.* Typhimurium across the epithelial cell barrier [9]. Once these bacteria reach the basolateral membrane, they are engulfed by phagocytes [10]. In order for the infection to extend beyond the intestinal mucosa, *S.* Typhimurium survives and replicates inside macrophages, a privileged niche that allows this pathogen to elude the adaptive immune response

thereby facilitating dissemination throughout the body, and reaching strategic organs like the liver or spleen [11]. During this process, infected individuals and animals expel this pathogen through their feces. In regions with poor sanitation systems, these bacteria can then contaminate water sources and food, thus spreading and infecting larger populations. From the point of view of induction of antibody production, the ability of *S.* Typhimurium to survive and replicate inside macrophages represents a challenge for vaccinology, since humoral immunity plays a key role in dealing with extracellular bacteria [12, 13].

Based on the observation that most bacterial vaccines inducing protective antibodies are mainly constituted by highly expressed, surface-exposed antigens and/or secreted toxins [14], the identification of such components in a bacteria can provide invaluable information. Thus, outer-membrane proteins (OMPs) are among the most obvious targets for protective immune response, particularly for subunit vaccine research, primarily because they are surface exposed and can therefore be recognized by the host immune system. It has been estimated that about 2-3% of the genomes of Gram-negative bacteria encode integral OMPs, and a significant proportion of these are expressed ubiquitously [15]. Taking into account all this information, together with the low protein complexity of the outer membrane (OM), it can be defined as an excellent subcellular fraction to be targeted by shotgun proteomics.

With regard to the current methodological approaches for OMPs extraction, several methods are currently available. Nonetheless, purification of OMPs to homogeneity, that is, being free of inner membrane, cell wall and cytoplasmic proteins, is challenging. In this sense, the sarkosyl extraction protocol has shown to be the most effective and selective method when compared with others (glycine extraction, differential detergent extraction using Triton X-100, serial extraction using 1M Tris pH 7, spheroplasting by lysozyme and sonication, and carbonate extraction) [16, 17]. This protocol uses N-lauroyl sarcosinate (sarkosyl) in order to achieve a better enrichment of OMPs in the samples obtained: this compound is an ionic detergent that preferentially solubilizes the inner membrane rather than the OM. Thus, after differential centrifugation, the OM is left as an insoluble pellet [18]. However, dynamic cellular processes and possible protein contamination events during extraction frequently lead to the detection of unexpected proteins belonging to other subcellular locations.

In 2001 Molloy *et al.* performed an analysis of the OMPs of *S*. Typhimurium by means of carbonate extraction, followed by two-dimensional electrophoresis (2DE) and identification of

proteins by PMF [19]. In their work the authors used the *S*. Typhimurium LT2 strain and were able to identify 24 different spots out of 37, corresponding to 23 different ORFs. The LT2 strain is a non-pathogenic reference strain commonly used in most laboratories. In fact, most information on genetic and phenotypic variation in *Salmonella* has derived from studies conducted in this strain, as only a small number of other laboratory strains has been referred elsewhere [20]. Since then, only another study performed by Coldham and Woodward in 2004 analyzed the proteome of the virulent *S*. Typhimurium SL1344 strain [21]. However, to date no exhaustive analysis of the OM subproteome of *S.* Typhimurium has been done, particularly using the most appropriate protocols for the extraction of OMPs. In order to address this gap of knowledge and take advantage of the recently published genome of *S.* Typhimurium SL1344 in 2012 [22], in the present study we have analyzed the OM subproteome of the *S*. Typhimurium SL1344 strain by means of the sarkosyl extraction protocol in combination with MS/MS approaches.

2. Materials and Methods.

2.1. Cultures and media

The *S.* Typhimurium strain ATCC SL1344 was grown in Luria-Bertani (LB) medium at 37 °C with shaking until it reached the exponential phase at an OD_{600} of 0.6. Two independent cultures were obtained and further processed for protein extraction.

2.2. Preparation of outer membrane protein extracts

OMPs extraction was performed by means of the N-lauroyl sarcosinate (also known as sarkosyl) method [18]. Briefly, *Salmonella* cells were harvested from 200 mL of exponential cultures (O.D. = 0.6) by centrifugation at 4 °C and 3500 x g. Cells were cleaned twice in PBS in order to remove any medium residue. Afterwards, cells were resuspended in 6 mL of 10mM Tris, pH 8.0 and NaCl 1%. At this point the cells were disrupted by sonication during 15 minutes (in cycles of 59 seconds on and 59 seconds off). After cell disruption the samples were centrifuged at 4 °C and 3500 x g in order to remove any cell debris. The supernatants were collected and transferred into ultra-centrifugation tubes and samples were centrifuged at 100.000 x g for one hour at 4 °C in a Sorvall MS-150 micro-ultracentrifuge (Thermo Scientific) using a S50-ST rotor. After this centrifugation step, the supernatants were discarded and the pellets were resuspended in 1% of freshly prepared sarkosyl solution and incubated during 60 minutes at room temperature with gentle agitation. After

incubation, the samples were again centrifuged at 100.000 x g for one hour at 4 °C, and the resulting pellets were cleaned twice with 10mM Tris, pH 8.0 and NaCl 1%. Lastly, after the final centrifugation step the pellets were carefully resuspended in 500 μL of milliQ water. The protein concentration was estimated with the 2D-Quant Kit from GE Healthcare (Fairfield, Connecticut, USA).

2.3. SDS-PAGE

Salmonella OMP preparations were analyzed by SDS-PAGE using a gel casting system (Bio-Rad tetra cell) and 12.5% isocratic Laemmli gels. Approximately 75 μg of protein were loaded in each lane. Gels were run at constant amperage (20mA) until the bromophenol blue tracking front had run off the end of the gel. Gels were stained with 0.1% Coomassie blue R-250 dye at room temperature for 30 minutes, and then distained overnight with 10% acetic acid in distilled water. The range of OMP molecular weights was estimated from a standard size marker (Benchmark™ protein ladder, Invitrogen (Chicago, USA)).

2.4. In-Gel Tryptic Digestion

The Coomassie-stained lanes were cut into 10 equal bands, and immediately distained and digested as described elsewhere [23]. Briefly, the Coomassie-stained lanes were washed twice with water for 20 min and distained with 200 μL of 50 mM ammonium bicarbonate/50% acetonitrile. Before tryptic digestion, reduction and alkylation with DTT/IAA was performed by incubating the samples with 200 μL of 10 mM DTT in 50 mM ammonium bicarbonate for 1 h at 56°C, followed by alkylation with 200 μL of 55 mM IAA in 50 mM ammonium bicarbonate for 30 min at room temperature, protected from light.

Gel pieces were digested overnight with 6 ng/μL trypsin at 37 °C. The peptide extraction was carried out with three consecutive washes with 1% formic acid for ESI analysis. The eluted peptides were dried in a SpeedVac and stored at −20 °C until analysis by mass spectrometry.

2.5. Mass Spectrometry (MS) Analysis

For GeLC–MS/MS analysis, the digests of the SDS-PAGE lanes were analyzed on an AmaZon ETD Ion Trap mass spectrometer (Bruker Daltonics), coupled to a nano-HPLC system (Proxeon). Peptide mixtures were first concentrated on a 300 mm i.d., 1 mm PepMap nanotrapping column and then loaded onto a 75 mm i.d., 15 cm PepMap nanoseparation column (LC Packings). Peptides were then eluted by a 0.1% formic acid/acetonitrile gradient (0–40% in 120 min; flow rate ca. 300 nL/min) through a nanoflow ESI Sprayer (Bruker Daltonics) onto the nanospray ionization source of the Ion

Trap mass spectrometer. MS/MS fragmentation (3 × 0.3 s, 100–2800 *m*/*z*) was performed on three of the most intense ions, as determined from a 0.8 s MS survey scan (310–1500 *m*/*z*), using a dynamic exclusion time of 1 min for precursor selection and excluding single-charged ions. An automated optimization of MS/MS fragmentation amplitude, starting from 0.60 V, was used.

Proteins were identified using Mascot (Matrix Science) to search in the NCBInr (July 2014). The search was restricted to the *Salmonella* taxonomy entry and contained 622.410 sequences. ~~First, common contaminants (tryptic autolytic fragment-, keratin-, and matrix-derived peaks) were removed using the contaminants database available in the Mascot search engine.~~ MS/MS spectra were searched with a precursor mass tolerance of 0.4 Da, fragment tolerance of 0.7 Da, trypsin specificity with a maximum of one missed cleavage and methionine oxidation was set as variable modification. Replicate analyses of all the LC-MS/MS analysis, using independent biological replicas, showed ≈ 80% coincidence indicating a high level of reproducibility. In order to ensure that the data are reliable a decoy database was used. In this sense the searches were repeated using identical search parameters against a database in which the sequences had been reversed or randomized [24]. Every time a protein sequence from the target database is tested during the search in the Mascot search engine a decoy sequence of the same length is automatically generated and tested. The average amino acid composition of the decoy sequences is the same as the average composition of the target database. Protein localization was predicted using the PSORTb v3.0[25]

Complete information on all peptide and protein identifications, including identification probabilities and sequences can be found in Supporting information Tables S1 and S2, respectively. The information shown is only related to the peptides present in both replicas since the peptides present in a single replica were not considered.

3. Results and discussion

3.1. OMP extraction

The cell envelope of *S.* Typhimurium and other Gram-negative enteric bacteria is a complex structure composed of 3 morphologically distinct layers [26, 27]: a cytoplasmic membrane, a rigid peptidoglycan layer external to the cytoplasmic membrane and a second membranous structure, the OM or L-layer, at the outer surface of the cell. The OM contains substantial amounts of proteins and phospholipids and, in addition, most or all of the lipopolysaccharide of the cell envelope. Osborn *et*

al. estimated that the OM of bacteria contains approximately 60% of all the proteins present in the two membranes [28]. However, although some OMPs from *S.* Typhimurium have been reported and characterized, no extensive analysis of the OM subproteome has been done using the most appropriate methodology.

The SL1344 genome contains 4742 protein-coding genes. A total of 4530 of these genes are present on the chromosome, and 212 genes are encoded by three plasmids. Currently, proteomes inferred from genome sequence data are extremely accessible but often remain unverified [29]. Only proteomics can unambiguously determine if the gene expressed is translated into a protein. High-throughput LC-MS/MS-based proteomics approaches measure protein fragments directly, and the resulting peptide sequences confirm the existence of a protein from a specific genome. Peptides that map to genomic regions outside the boundaries of previously annotated genes represent evidence of novel genes or extensions of their predicted termini.

In the present study we used the GeLC-MS/MS approach, widely used in shotgun proteomics, in order to identify the OM subproteome of *S.* Typhimurium SL1344 obtained through the sarkosyl extraction method. In terms of protein analysis, several reports have proven and identified the sarkosyl protocol as the best methodology for OMP extraction in order to obtain the purest samples. Hobb *et al.* [16] analyzed a total of nine methodologies and concluded that glycine extraction, differential detergent extraction using Triton X-100, serial extraction using 1M Tris pH 7, spheroplasting by lysozyme and sonication, and carbonate extraction did not produce pure OM preparations [16]. According to the authors, the extraction of OMPs using sarkosyl produced the purest samples leading to the most reproducible results. In addition, Cao *et al.* extracted the OMPs from the Gram-negative bacterium *Caulobacter crescentus* using the carbonate method and the sarkosyl extraction protocol. Similarly, they also concluded that the sarkosyl protocol gave the purest OMP preparations[17]. The results supporting better enrichment of samples with OMPS rely on the use of sarkosyl, an ionic detergent which allows the recovery of the OM as an insoluble pellet due to a preferential solubilization of the inner membrane[18] .

3.2. Identification of the OMPs

After OMP extraction, proteins were separated by SDS-PAGE. Two biological replicas were processed and each was loaded into a different lane, which was cut into 10 pieces and analyzed by LC-MS/MS. Using this approach about 42315 MS/MS spectra were acquired for replica 1 while 45331 MS/MS spectra were acquired for replica 2. Searches were performed with MASCOT using the

226 *Salmonella* taxonomy entry from the NCBInr database. However, since not all the entries had their
227 corresponding entry to the SL1344 strain, a BLAST search was carried out each time against the
228 SL1344 strain in order to generate a list with the SL1344 entries.

229 The results delivered 34812 peptides for replica 1 while 28571 peptides were obtained for
230 replica 2. The two independent biological replicas showed approximately 78% of coincidence in the
231 peptide spectral matches (PSMs) identified. In order to obtain reliable results we used a quite strict
232 cut-off in order to accept positive identifications. Proteins were accepted only if at least two different
233 peptides had been identified and the p-values were lower than 0.05. After removing duplicates,
234 PSMs identified with p-values higher than 0.05 as well as those PSMs present in only one of the two
235 biological replicas, a total of 1191 different PSMs were identified (Supporting Information Table S1).
236 Then, we grouped together the peptides into proteins and a total of up to 180 different ORFs were
237 identified (Supporting Information Table S2). As exceptions, two proteins were accepted with the
238 identification of only one PSM, these were the Small protein A and Entericidin B (gi numbers:
239 378700596 and 378702179 respectively). These proteins were accepted because the PSM identified
240 represented 18% (Small protein A) and 40% (Entericidin B) of the sequence coverage with very good
241 p-values (6,5E-5 and 7E-12, respectively). The Small protein A is a small outer-membrane lipoprotein
242 that is a component of the essential YaeT outer-membrane protein assembly complex [30].
243 Entericidin B is a small cell-envelope bacteriolytic lipoprotein that can maintain plasmids in bacterial
244 populations by means of post-segregational killing [31]. Fragmentation spectra of these two peptides
245 are provided as Supporting Information Figure 1 and Figure 2.

246 The first and most notable result was the experimental evidence of proteins the existence of
247 which had not previously been demonstrated. Indeed, 21 of the total proteins identified were
248 annotated as hypothetical proteins (Supporting Information Table S2). As the existence of these
249 proteins had been predicted only by means of *in silico* methods, herein we provide the first
250 experimental evidence of their synthesis. Concerning the remaining 160 proteins, these were
251 classified according to their major or putative function. The resulting classification is detailed in
252 Supporting Information Table S1 but Table 1 summarizes the results of the proteins belonging to
253 each category. Most of the proteins (43) were related to transport functions, whereas the second
254 and third most representative groups were composed of proteins related to virulence (29) and
255 ribosomal proteins (23), respectively.

256 On comparing our results with those reported in previous studies several considerations
257 need to be taken into account. In the work conducted in 2001 by Molloy *et al.* it was of note that the

authors used a carbonate extraction method followed by a 2DE analysis, which allowed the identification of only 23 different ORFs in the non-pathogenic LT2 strain [19]. It is well known that one of the limitations of 2DE analysis is the under-representation of membrane proteins due to poor solubility in the buffer required for the isoelectrofocusing separation procedure. This limitation could lead to the limited results obtained. Almost all of the proteins identified by Molloy *et al.*, were also identified in the present study. The proteins not identified in our study were the molecular chaperone DnaK, the 30S ribosomal protein subunit S1, a phosphoglycerate kinase, an enolase and the heat shock protein Hsp90, most of which are known to have an intracellular function. Thus, these differential identifications are likely to be the result of the different methodological approaches used in the two studies. Molloy *et al.* used the carbonate method which, as mentioned previously [16, 17] does not produce pure OM preparations.

Moreover, two different studies have analyzed the proteome of the pathogenic strain *S.* Typhimurium SL1344. In 1996 Qi *et al.* analyzed by 2DE the phosphate buffer-insoluble proteome in the absence of a specific OMP extraction method [32]. Proteins were electroblotted from the gel onto a PVDF membrane and then, N-terminal sequences were determined by sequential Edman degradation. However, in addition to the handicap of the solubilization of membrane proteins in 2DE analysis, Edman degradation presents limitations in terms of sensitivity as well as the problem of identifying the blocked N- terminal residues. The authors identified a total of 5 OMPs with known function at the time of the experiment, representing 10.2% of the total proteins identified. Thus, the well-known under-representation of membrane proteins when using this methodology may explain why only 5 out of the 49 proteins identified by these authors were also found in the present study [32]. Later, in 2004 Coldham and Woodward performed an in-depth analysis of the proteome of *S.* Typhimurium SL1344 [21]. The authors considered the insoluble fraction of a buffer containing urea/thiourea, triton X-100/CHAPS as the OM subproteome. It was then expected that these sample preparations contained all kinds of membrane proteins as well as macromolecular complexes. They identified 34 OMPs from a total of 816 proteins identified, hence representing only 4.2%. Among these 34 proteins, 24 were also found in the present work.

To the contrary, we used the sarkosyl extraction method, described as the purest and most reproducible methodology, and from the total of 180 proteins we identified 50 which were categorized as OMPs with the PSORTb, as described in the next subsection “3.3 Prediction of the subcellular location”. This fraction corresponds to 27.8%, a much higher value in comparison with the previous findings. Thus, neither of the previous studies used specific accurate methods for OMP

extraction, hence justifying the important differences observed between their findings and our results. Moreover, the data of these previous reports could not be checked with the genome of *S.* Typhimurium SL1344 since it had not been published until 2012 [22].

In order to specify which proteins could be considered as OMPs, we conducted a search of the bibliography regarding each of the 180 proteins identified in the present study. The information available regarding their function and previous analysis revealed that 61 proteins (33.9%) should be considered as OMPs (Table 1). Thus, in comparison with the abovementioned reports, this percentage represents an increased value and indicates the greater appropriateness of the sarkosyl methodology.

3.3. Prediction of the subcellular location

To further complement the information obtained in the present study, we also investigated the subcellular location of all the 180 proteins identified by two different approaches: we used the PSORTb v3.0 prediction server and, next, we also searched for the annotated subcellular location, if available, from the Uniprot database (Table 2). A total of 158 proteins were classified in a particular bacterial location using the PSORTb prediction software *versus* the 141 classified according to the information reported in the Uniprot database. Details of the predictive results obtained from the analysis of the PSORTb software can be found in Supplementary Information Table S2. According to the prediction, 65 proteins (36.1%) were located in the inner membrane, 50 (27.7%) in the OM, 30 (16.6%) were cytoplasmic, 10 (5.5%) extracellular, 2 (1.1%) periplasmic and the remaining 23 (12.7%) proteins had an unknown location (Figure 1). Among the latter group 8 proteins were predicted to be non-cytoplasmic, although PSORTb was unable to specify the subcellular location.

Thereafter, we aimed to compare the results obtained from the predictive software with the information available in the Uniprot database in order to determine their potential contribution (Table 2). In terms of the 65 inner membrane proteins predicted by PSORTb, only 14 were equally annotated in their corresponding entries in the Uniprot database. Of the remaining 51 proteins, 38 showed membrane location although no distinction between inner or outer membrane was stated, 2 were related to the flagellum basal body, 1 corresponded to the nitrate reductase complex, and 10 did not have any specified location. The PSORTb predicted that 50 proteins are located in the OM *versus* 40 proteins annotated as OMPs according to the Uniprot Database (35 were equally assigned as OMPs whereas the other 5 were of unknown location by PSORTb). For the 30 proteins predicted by PSORTb to be located in the cytoplasm, according to Uniprot 3 they were similarly classified, and

23 corresponded to ribosomal proteins. The other 4 proteins were two integral components of the membrane (Uniprot entries: A0A0H3NL53 and A0A0H3NVV1) and two proteins with no information about subcellular location in the Uniprot database (Uniprot entries: A0A0H3NX93 and A0A0H3NUC4). Additionally, according to the PSORTb annotation, 10 proteins were classified as being extracellularly located whereas only 4 were so classified by the Uniprot database. The remaining 6 proteins were assigned in the bacterial-type flagellum subgroup. Concerning the two proteins predicted to be located in the periplasm by PSORTb, one was a protein involved in the formation of diffusion channels in the OM during phage adsorption (Uniprot entry A0A0H3NIT6) and the other was an OMP with unknown function (Uniprot entry A0A0H3NM65). Neither of these two proteins could be classified in any subgroup in the Uniprot database.

The comparison between the two different databases used (PSORTb and Uniprot) for classification into distinct bacterial locations showed that PSORTb was able to assign a greater number of proteins in a specific subgroup (87.3% with PSORTb *vs* 77.9% with Uniprot). Moreover, proteins located in the membrane, without further identification of inner or outer membrane, according to Uniprot, showed a more precise location by means of the PSORTb software. The number of proteins assigned in the different categories did not fully coincide between the two databases, being the proteins assigned in the inner membrane subgroup the least similar. Only 22.7% of the proteins identified as being located in the inner membrane with PSORTb were equally grouped by Uniprot. However, it is of note that both databases have different categories of bacterial location, therefore the differences seen between the classification of the proteins obtained in the present study could, in part, be due to this fact.

On considering the differential classification of the proteins shown in the present study, independently of the database considered, it can be seen that, even though we used the best methodology described so far, which shows better OMP enrichment in the final samples, we have identified proteins whose function is known to be intracellular or related to other subcellular locations (inner membrane, periplasm, extracellular milieu). Regarding the presence of cytoplasmic proteins, although we cannot completely exclude the possibility of cytoplasmic protein contamination, a feasible explanation can be the formation of budding vesicles. It is well known that bacteria constitutively secrete native outer membrane vesicles (OMVs) into the extracellular milieu containing cytoplasmic proteins [33]. It has also been demonstrated that such vesicles carry DNA and RNA molecules. Since translation of OMPs may occur simultaneously with their integration into the membrane, the presence of ribosomal proteins, chaperones or other cytoplasmic proteins may not

be surprising, suggesting that they may play a role in OMVs hence justifying their presence in OMP extraction [34-37]. Accordingly, the information available in the Uniprot database identified 23 proteins as ribosomal proteins, whereas only 3 additional proteins were considered as cytoplasmic proteins. In terms of the presence of inner membrane proteins a more difficult interpretation can be concluded. As mentioned above, the proteins assigned in this subgroup were the least similar: 66 inner membrane proteins identified by PSORTb, in contrast with the information annotated in the Uniprot database, with 16 proteins as components of the inner membrane and 45 additional proteins related to the concept "membrane" without further identification. In view of these findings, it is difficult to conclude which amount of inner membrane proteins is certainly identified in the present study. Moreover, to our knowledge the use of the sarkosyl method has never been used for the study of the *S*. Typhimurium proteome. Therefore, we cannot rule out the possibility that using a different microorganism than those reported in the previous studies may lead to suboptimal results in terms of purification to homogeneity. Alternatively, a certain degree of contamination can always occur with a given technique. In addition, despite the efficiency of any method in the enrichment or extraction of membrane proteins, it is remarkable to indicate that identification of cytoplasmic proteins, or inner membrane proteins in this subproteome should not be surprising as many biological processes require the transit of proteins through membranes, as in vesicle formation, transport or LPS synthesis among others, and the data obtained always represents a picture obtained at a specific time under specific conditions.

4. Conclusions

In conclusion, in the present work we describe for the first time a detailed analysis of the OM subproteome of the pathogenic strain *S*. Typhimurium SL1344 using the most appropriate methodology. We report the efficiency of the sarkosyl extraction method in characterizing a large number of proteins of this bacterial compartment. This methodology together with the GeLC-MS/MS approach allowed the identification of up to 180 proteins whereas previous studies have reported less than 53 proteins. Moreover, further information regarding the intracellular location of these proteins has also been considered. The most important results that may be highlighted in the present work are the identification of 61 proteins which correspond to OMPs, representing 33.9% of the total number of proteins identified. In addition, we have also reported experimental evidence of the existence of 21 proteins annotated as hypothetical proteins. In view of these results, this study

384 provides new information about the proteome of SL1344. According to the current clinical situation
385 of increasing trends of antibiotic resistance and the lack of an effective vaccine, this new insight will
386 be very helpful for future studies on *S*. Typhimurium, particularly those focused on identifying new
387 targets for the development of novel tools to fight against this important pathogen, such as subunit
388 vaccines or inhibitory drugs.

389

390 Supporting information available: This material is available free of charge via the Internet at
391 http://pubs.acs.org

392

393 ## Acknowledgments

394 The authors thank the Ministerio de Economía y Competitividad, Instituto de Salud Carlos III
395 – co-financed by the European Development Regional Fund 'A way to achieve Europe' ERDF, Spanish
396 Network for the Research in Infectious Diseases [REIPI RD12/0015] and FIS 14/0755. This study was
397 also supported by grant 2014SGR0653 from the Departament d'Universitats, Recerca i Societat de la
398 Informació, of the Generalitat de Catalunya and by funding from the Innovative Medicines Initiative
399 (Translocation, contract IMI-JU-6-2012-115525) The authors thank Dr. S. Bronsoms and Dr. S. Trejo
400 from Proteomics Facility from IBB-UAB for technical advice.

401

References

[1] Hohmann EL. Nontyphoidal salmonellosis. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2001;32:263-9.

[2] Commission E. The rapid alert system for food and feed. 2013 Annual report. RASFF Anual reports. 2014.

[3] Authority EFS. EFSA explains zoonotic diseases: Salmonella. EFSA: Corporate Publications. 2014.

[4] MacLennan CA, Martin LB, Micoli F. Vaccines against invasive Salmonella disease: Current status and future directions. Human Vaccines & Immunotherapeutics. 2014;10:1478-93.

[5] Feasey NA, Dougan G, Kingsley RA, Heyderman RS, Gordon MA. Invasive non-typhoidal salmonella disease: an emerging and neglected tropical disease in Africa. Lancet. 2012;379:2489-99.

[6] Gordon MA, Banda HT, Gondwe M, Gordon SB, Boeree MJ, Walsh AL, et al. Non-typhoidal salmonella bacteraemia among HIV-infected Malawian adults: high mortality and frequent recrudescence. AIDS. 2002;16:1633-41.

[7] Reddy EA, Shaw AV, Crump JA. Community-acquired bloodstream infections in Africa: a systematic review and meta-analysis. The Lancet Infectious diseases. 2010;10:417-32.

[8] Nicholas AF, Amy KC, Chisomo LM, Derek P, Maaike A, Martin A, et al. Drug Resistance in Salmonella enterica ser. Typhimurium Bloodstream Infection, Malawi. Emerging Infectious Disease journal. 2014;20:1957.

[9] Fàbrega A, Vila J. Salmonella enterica Serovar Typhimurium Skills To Succeed in the Host: Virulence and Regulation. Clinical Microbiology Reviews. 2013;26:308-41.

[10] Jones BD, Ghori N, Falkow S. Salmonella typhimurium initiates murine infection by penetrating and destroying the specialized epithelial M cells of the Peyer's patches. The Journal of Experimental Medicine. 1994;180:15-23.

[11] Ibarra JA, Steele-Mortimer O. Salmonella--the ultimate insider. Salmonella virulence factors that modulate intracellular survival. Cellular microbiology. 2009;11:1579-86.

[12] Mastroeni P, Villarreal-Ramos B, Hormaeche CE. Adoptive transfer of immunity to oral challenge with virulent salmonellae in innately susceptible BALB/c mice requires both immune serum and T cells. Infection and immunity. 1993;61:3981-4.

[13] Salerno-Goncalves R, Pasetti MF, Sztein MB. Characterization of CD8(+) effector T cell responses in volunteers immunized with Salmonella enterica serovar Typhi strain Ty21a typhoid vaccine. Journal of immunology. 2002;169:2196-203.

[14] Grandi G. Bacterial surface proteins and vaccines. F1000 biology reports. 2010;2.

[15] Wimley WC. The versatile β-barrel membrane protein. Current Opinion in Structural Biology. 2003;13:404-11.

[16] Hobb RI, Fields JA, Burns CM, Thompson SA. Evaluation of procedures for outer membrane isolation from Campylobacter jejuni. Microbiology. 2009;155:979-88.

[17] Cao Y, Johnson HM, Bazemore-Walker CR. Improved enrichment and proteomic identification of outer membrane proteins from a Gram-negative bacterium: Focus on Caulobacter crescentus. Proteomics. 2012;12:251-62.

[18] Filip C, Fletcher G, Wulff JL, Earhart CF. Solubilization of the Cytoplasmic Membrane of Escherichia coli by the Ionic Detergent Sodium-Lauryl Sarcosinate. Journal of bacteriology. 1973;115:717-22.

[19] Molloy MP, Phadke ND, Maddock JR, Andrews PC. Two-dimensional electrophoresis and peptide mass fingerprinting of bacterial outer membrane proteins. Electrophoresis. 2001;22:1686-96.

[20] Beltran P, Plock SA, Smith NH, Whittam TS, Old DC, Selander RK. Reference collection of strains of the Salmonella typhimurium complex from natural populations. Journal of General Microbiology. 1991;137:601-6.

449 [21] Coldham NG, Woodward MJ. Characterization of the Salmonella typhimurium proteome by
450 semi-automated two dimensional HPLC-mass spectrometry: detection of proteins implicated in
451 multiple antibiotic resistance. Journal of proteome research. 2004;3:595-603.
452 [22] Kröger C, Dillon SC, Cameron ADS, Papenfort K, Sivasankaran SK, Hokamp K, et al. The
453 transcriptional landscape and small RNAs of Salmonella enterica serovar Typhimurium. Proceedings
454 of the National Academy of Sciences. 2012;109:E1277–E86.
455 [23] Shevchenko A, Wilm M, Vorm O, Mann M. Mass Spectrometric Sequencing of Proteins from
456 Silver-Stained Polyacrylamide Gels. Analytical Chemistry. 1996;68:850-8.
457 [24] Elias JE, Haas W, Faherty BK, Gygi SP. Comparative evaluation of mass spectrometry platforms
458 used in large-scale proteomics investigations. Nat Methods. 2005;2:667-75.
459 [25] Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, et al. PSORTb 3.0: improved protein subcellular
460 localization prediction with refined localization subcategories and predictive capabilities for all
461 prokaryotes. Bioinformatics. 2010;26:1608-15.
462 [26] De Petris S. Ultrastructure of the cell wall of Escherichia coli and chemical nature of its
463 constituent layers. Journal of ultrastructure research. 1967;19:45-83.
464 [27] Murray RG, Steed P, Elson HE. The Location of the Mucopeptide in Sections of the Cell Wall of
465 Escherichia Coli and Other Gram-Negative Bacteria. Canadian journal of microbiology. 1965;11:547-
466 60.
467 [28] Osborn MJ, Gander JE, Parisi E, Carson J. Mechanism of assembly of the outer membrane of
468 Salmonella typhimurium. Isolation and characterization of cytoplasmic and outer membrane. The
469 Journal of biological chemistry. 1972;247:3962-72.
470 [29] Ansong C, Purvine SO, Adkins JN, Lipton MS, Smith RD. Proteogenomics: needs and roles to be
471 filled by proteomics in genome annotation. Briefings in Functional Genomics & Proteomics.
472 2008;7:50-62.
473 [30] Lewis C, Skovierova H, Rowley G, Rezuchova B, Homerova D, Stevenson A, et al. Small outer-
474 membrane lipoprotein, SmpA, is regulated by σE and has a role in cell envelope integrity and
475 virulence of Salmonella enterica serovar Typhimurium. Microbiology. 2008;154:979-88.
476 [31] Bishop RE, Leskiw BK, Hodges RS, Kay CM, Weiner JH. The entericidin locus of Escherichia coli
477 and its implications for programmed bacterial cell death. Journal of Molecular Biology. 1998;280:583-
478 96.
479 [32] Qi SY, Moir A, O'Connor CD. Proteome of Salmonella typhimurium SL1344: identification of novel
480 abundant cell envelope proteins and assignment to a two-dimensional reference map. Journal of
481 bacteriology. 1996;178:5032-8.
482 [33] Lee EY, Bang JY, Park GW, Choi DS, Kang JS, Kim HJ, et al. Global proteomic profiling of native
483 outer membrane vesicles derived from Escherichia coli. Proteomics. 2007;7:3143-53.
484 [34] Dorward DW, Garon CF, Judd RC. Export and intercellular transfer of DNA via membrane blebs of
485 Neisseria gonorrhoeae. Journal of bacteriology. 1989;171:2499-505.
486 [35] Kadurugamuwa JL, Beveridge TJ. Virulence factors are released from Pseudomonas aeruginosa in
487 association with membrane vesicles during normal growth and exposure to gentamicin: a novel
488 mechanism of enzyme secretion. Journal of bacteriology. 1995;177:3998-4008.
489 [36] Kolling GL, Matthews KR. Export of virulence genes and Shiga toxin by membrane vesicles of
490 Escherichia coli O157:H7. Applied and environmental microbiology. 1999;65:1843-8.
491 [37] Yaron S, Kolling GL, Simon L, Matthews KR. Vesicle-mediated transfer of virulence genes from
492 Escherichia coli O157:H7 to other enteric bacteria. Applied and environmental microbiology.
493 2000;66:4414-20.

494

495

Tables

Table 1. Classification of all the proteins characterized into functional categories.

Functional categories (Nº)	Proteins						
Transport (43)							
	AcrA	AcrB	**BtuB**[a]	**CirA**	CycA	ExbB	**FadL**
	FeoB	**FepA**	**FhuA**	FocA	FruA	GlpF	GlpT
	GltS	**HemM**	LysP	ManZ	MetQ	MsbA	MtlA
	NagE	NupC	PotE	PtsG	SdaC	SecD	SecF
	SecY	**TamA**	**TolC**	**Tsx**	UlaA	**VacJ**	YajC
	YajR	YbjY	YdjN	YedE	YeeF	**YrbD**	YrbK
	N.A.[b]						
Porins (9)							
	OmpA	**OmpC**	**OmpD**	**OmpF**	**OmpN**	**OmpS**	**OmpW**
	KdgM	**LamB**					
Cell envelope integrity and biogenesis (16)							
	BamA	**BamB**	**BamC**	**BamD**	**BamE**	FtsH	**Lpp**
	LppB	**MipA**	NlpD	**Pal**	**SlyB**	**YbhC**	YbjR
	YiaD	YidC					
LPS (5)							
	LptD	**LptE**	RfaL	WzzB	WzzE		
Electron transport (13)							
	AtpF	CydA	CydB	CyoA	CyoB	DmsB	NarG
	NuoA	NuoH	NuoL	PntA	PntB	TrxA	
Virulence (29)							
	FimD	FlgE	FlgG	**FlgH**	FlgK	FliC	FliD
	FliF	FliL	FljB	InvA	**InvG**	**Mce**	**OmpX**
	PagC	**PagN**	**PagP**	**PagP**	PrgH	PrgK	**RcK**
	SafC	SiiB	**SiiC**	SipA	SipB	SipC	SopB
	SopE						
Cellular responses (7)							
	CstAa	CstAb	Dps	EcnB	GroEL	**RcsF**	**YeaY**
Metabolism (8)							
	ApeE	Cdh	DgkA	DmsA	FrdC	Gcd	**PldA**
	Psd						
Ribosome (23)							
	RplA	RplB	RplE	RplF	RplJ	RplM	RplN
	RplO	RplP	RplQ	RplR	RplT	RplU	RplV
	RplX	RpsB	RpsC	RpsD	RpsE	RpsI	RpsK
	RpsP	RpsS					
Other (13)							
	DcrB	Ftn	HflC	HflK	HybC	**OsmE**	PqiB
	RlpA	SppA	TraF	TraI	**TraT**	YeiU	

Unknown function (15)	LpxR	YajG	YbjP	YdgA	YdgH	**YdiY**	YecR-like
	YfaZ	**YfeY**	**YgiB**	YhcB	YibN	YijP	YnfB
	YraP						

498

499 [a] Boldface type indicates OMPs according to published information.

500 [b] N.A., Not Available.

501

502 **Table 2.** Comparison of the predicted location by the PSORTb software and the information
503 published in the Uniprot database.

Bacterial location[a]	PSORTb v3.0		Uniprot		Common results[b]	
	Nº proteins	%	Nº proteins	%	Nº proteins	%
Inner membrane	65	36.1	15	8.3	14	7.7
Outer membrane	50	27.7	40	22.2	35	19.4
Membrane			45	25		
Cytoplasmic	30	16.6	3	1.6	26	14.4
Ribosome			23	12.7		
Extracellular	10	5.5	4	2.2	4	2.2
Periplasmic	2	1.1				
Cell wall			1	0.6		
Flagellum			8	4.4		
Nitrate reductase complex			1	0.5		
Unknown (N.A.) [c]	23	12.7	40	22.2	16	8.8

504

505 [a] Different bacterial locations are referred for each classification according to the information
506 supplied by each approach.

507 [b] Common results have only been specified for those common locations. Cytoplasmic and ribosome
508 categories have been unified in the comparison.

509 [c] N.A., Not Available.

510

Figure

Figure 1. Subcellular distribution of the proteins identified in the present study in different categories according to: Left, P-SORT 3.0 prediction; Right, annotations obtained from Uniprot.