Rapid label-free quantitative analysis of the *E. coli* BL21(DE3) inner membrane proteome

Malvina Papanastasiou‡, Georgia Orfanoudaki‡§, Nikos Kountourakis‡, Marina Koukaki‡, Marios Frantzeskos Sardis‡¶, Michalis Aivaliotis‡, Konstantinos C. Tsolis‡§¶, Spyridoula Karamanou‡¶ and Anastassios Economou‡§¶

‡Institute of Molecular Biology and Biotechnology, FORTH, Iraklio, Crete 71110, Greece

§Department of Biology-University of Crete, P.O. Box 1385, Iraklio, Crete 71110, Greece;

¶Laboratory of Molecular Bacteriology, Rega Institute, Dpt of Microbiology and Immunology, KU Leuven, Herrestraat 49, B-3000 Leuven, Belgium

For correspondence:

E-mail: tassos.economou@rega.kuleuven.be

Running title: *E. coli* BL21(DE3) IMP

Keywords: inner membrane proteome, label-free quantitation, LC-MS/MS, SecYEG over-expressed, surface proteolysis

Abbreviations

CE, cell envelope; CEP, cell envelope proteome; CV, coefficient of variation; DDM, *n*-Dodecyl ² -D-maltoside; emPAI, exponentially modified protein abundance index; IAM, iodoacetamide; IM, inner membrane; IMP, inner membrane proteome; IMVs, inverted inner membrane vesicles; NSAF, normalized spectral abundance factors; OM, outer membrane; PIM, peripheral inner membrane; SP, surface proteolysis; TCEP, tris(2-carboxyethyl)phosphine; TM, transmembrane helix

Abstract (151 words)

Biological membranes define cells and cellular compartments and are essential in regulating bidirectional flow of chemicals and signals. Characterizing their protein content therefore is required to determine their function, nevertheless, the comprehensive determination of membrane-embedded sub-proteomes remains challenging. Here, we experimentally characterized the inner membrane proteome (IMP) of the model organism *E. coli* BL21(DE3). We took advantage of the recent extensive re-annotation of the theoretical *E. coli* IMP regarding the sub-cellular localization of all its proteins. Using surface proteolysis of IMVs with variable chemical treatments followed by nanoLC-MS/MS analysis, we experimentally identified ~45% of the expressed IMP in wild type *E. coli* BL21(DE3) with 242 proteins reported here for the first time. Using modified label-free approaches we quantified 220 IM proteins. Finally, we compared protein levels between wild type cells and those over-synthesizing the membrane-embedded translocation channel SecYEG proteins. We propose that this proteomics pipeline will be generally applicable to the determination of IMP from other bacteria.

Statement of significance

A high-throughput method based on surface proteolysis of IMVs followed by nanoLC-MS/MS analysis was used to quantify the inner membrane proteome of *E. coli* BL21(DE3). Our results identify and quantify inner membrane proteins of wild type and SecYEG over-expressing strains. Surface proteolysis of membrane vesicles enables rapid characterization of the IMP. This pipeline is readily applied to large-scale profiling of bacterial IM proteomes, essential for understanding bacterial physiology, pathogenesis and biotechnology.

Introduction

The availability of completed genomes and advances in analytical techniques have ushered in wide scale protein studies, otherwise known as proteomics. *Escherichia coli*, a reference model organism, has served as *proof-of-concept* for systems-based studies and also for validating emerging technologies [1]. *E. coli* BL21(DE3) is the most widely used workhorse strain for the production of recombinant proteins. It has the phage T7 RNA polymerase gene integrated in its chromosome and expresses genes placed behind the strong T7 promoter [2]. BL21(DE3) is more robust than *E. coli* K12 strains for heterologous protein synthesis due to reduced acetate production [3-5].

E. coli belongs to the Gram$^-$ bacteria that have characteristic cell envelopes (hereafter, “CE”) [6]. The CE is a layered structure comprising the inner membrane (IM) that directly surrounds the cytoplasm and the outer membrane (OM), of the cell. To date, only a few studies have targeted the cell envelope proteome (CEP) of BL21(DE3), compared to CE proteomics studies of strain K12. This is partly due to the earlier availability of the complete genome of K12 strains MG1655 [7] and W3110 [8]. Most of the earlier studies employing BL21(DE3), querried the K12 proteome database; these included the interrogation and tuning of membrane protein over-expression [9, 10], the extracellular proteome production [11, 12] and the characterization of the CEP [13, 14] and the IM complexome [14, 15]. These studies used gel-based approaches followed by protein identification by ESI-MS/MS or MALDI-ToF. More recently, the now sequenced BL21(DE3) genome were also exploited; these studies targeted the peripheral inner membrane (PIM) [16], the CE [17] and the OM [18] proteomes using surface proteolysis [16], 2D gels [17] and SDS-PAGE [18].

The IM proteome (IMP) accounts for ~22% of the theoretical cellular proteome of *E.coli* BL21(DE3) and more than half of the CEP [19] and an in-depth IM proteome coverage has yet to be reported. IM proteins are involved in a plethora of cellular

functions including signal transduction, energy conversion, protein secretion, cell division, DNA replication. Further, the IMP is involved in viral trafficking and affects the assembly and release of viruses as well as their entry into cells [20]. Finally, functional membranes are employed in modern biotechnology for the production of biopharmaceuticals including cytokines, antibodies and exopolysaccharides [21]. Structural and mechanistic insights of the IM proteome therefore is essential to determine protein activities, interactions and responses to environmental changes, surface-exposed pathogenic molecules and vaccine targets.

Nevertheless, the low expression levels of integral IM proteins [5, 22] and their hydrophobic properties require demanding sample preparation methods that often entail sub-cellular fractionation followed by additional enrichment steps and the use of detergents for solubilization [23-25]. Furthermore, typical proteases employed in bottom-up proteomics applications, such as trypsin and Lys-C that target basic residues (Arg/Lys), produce large transmembrane (TM) peptides that a) are hydrophobic, b) are poorly extracted from gels and are therefore rare in LC-MS/MS samples compared to soluble peptides and c) have a limited charge-potential, therefore might evade detection in typical *m/z* ranges commonly used in MS analysis. Emerging developments (i.e. liquid preparative isoelectric focusing (OFFGEL) [26] and filter-aided sample preparation (FASP) [27, 28] that improve identification of TM domains have been reported [27, 29].

An alternative sample preparation option for IM protein identification is to target the soluble domains exclusively, a technique known as surface proteolysis (or "membrane shaving" or "surface shaving") [30]. In this approach, intact membranes or membrane vesicles are incubated with specific (i.e. trypsin, Lys-C) or non-specific (i.e. proteinase K, thermolysin) proteases that target the extra-membrane protein domains, generating soluble peptides identified subsequently by MS [31]. The success of the method is heavily dependent on the presence of specific cleavage sites (for specific

proteases) whereas disadvantages of using non-specific proteases include a much larger number of peptides [31] and the generation of internal ions during the MS/MS process~~, that can be overcome by modifying the N-terminus of the peptides~~ [32]. Surface proteolysis has been traditionally used in $Gram^+$ bacteria that have a more rigid and resistant to cell lysis CE than $Gram^-$ bacteria and a single CEP membrane [33]. The technique has been applied for the profiling of OM proteins of uropathogenic [34, 35] and *K12* [36] *E. coli*, the PIM proteome of BL21(DE3) [16] and to the topological analysis of *E. coli* BL21(DE3) IM proteins using thermolysin [32].

Here, we investigate how potential biases or advantages of surface proteolysis are anticipated to affect protein identification, coverage and quantification of the *E. coli* BL21(DE3) IM proteome. We present a theoretical analysis of challenges that may hamper detection of IM proteins, with a focus on their physicochemical features. Based on this analysis we estimate the percentage of the theoretical IMP that is 'MS-detectable'. We then employ extensive biochemical treatment of inverted inner membrane vesicles (IMVs) to produce highly pure membrane fractions. IM proteins are successfully identified by their polar segments. Finally, we estimate the relative abundance of IM proteins using label-free quantification with an algorithmic adjustment to the sample preparation approach employed. These tools are applied to detect fine changes to the IMP of cells over-expressing the protein translocase.

Materials and methods

1. Chemicals & Reagents

Chemicals were obtained from Sigma Aldrich, Fisher Scientific and Bioline-Applichem. Iodoacetamide (IAM) was from Merck and trypsin (sequencing grade) from Roche.

2. Cell culture and membrane preparation

E.coli BL21(DE3) cells grown overnight (37ºC; 600ml LB; 100mg/ml ampicillin) were used to inoculate 30L cells grown in LB medium in a closed fermentor setup (Bioengineering AG); 100mg/ml ampicillin. The vector pET610 containing a synthetic *his-secY/secE/secG* operon, under the control of a *trc* promoter operon [37] and containing amp^R allowed for the overproduction of His-tagged SecYEG. Induction was performed with 0.2mM IPTG, for 3h (OD_{600}=1.5). Cells were grown with agitation, at 37ºC, until OD_{600}=0.6. Cell paste was collected at 5000×g (Beckman JLA 8.1000), at 4ºC, for 20min and resuspended in an equal volume of 50mM Tris-HCl; pH:8.0; 20% glycerol; 10mg/ml DNaseI; 50mg/ml RNase; 1mM PMSF, at 4ºC. Cells were lysed in a French press (8000psi; 5 times), diluted 5-10 times with 50mM Tris-HCl; pH:8.0 and centrifuged at 4000×g (swing-out bucket, Sorvall Heraeus 6445; 10min) to remove unbroken cells. The supernatant was centrifuged at 120,000×g, at 4ºC, for 90min in a fixed angle rotor (Sorvall T647.5). Pellets were resuspended in 50mM Tris-HCl (pH:8.0, buffer A), loaded on a 5-step sucrose gradient (1.9; 1.7; 1.5; 1.3; 1.1 M) and centrifuged at 84,000×g in a swing-out bucket rotor (Sorvall AH 629/36; 4ºC, 16h). Following washing steps with 50mM Tris-HCl to remove sucrose, IMVs were resuspended in buffer A (“intact IMVs”). These were further treated with KCl (2M), EDTA (2mM), Na_2CO_3 (100mM) and urea (6M). Finally, IMVs were flotated on a 5-step sucrose gradient (2.4; 2; 1.7; 1.4; 1M) at

200,000×g, at 4ºC, for 16h ("treated IMVs"). Following washing steps, "treated IMVs" were re-suspended in buffer A and were subjected to surface proteolysis.

3. Surface proteolysis

Surface proteolysis was performed as previously described [16]. Briefly, "intact" or "treated" IMVs (10μg total protein) were reduced with TCEP (10-fold molar excess) for 30min at room temperature and alkylated with IAM (5mM) for 45min in the dark. Samples were digested overnight, at 37ºC with trypsin (1:200 enzyme:protein). Following ultra-centrifugation (200,000×*g,* 4ºC, 30min, Sorvall WX Ultra Series 80, TH641), the supernatant was desalted using StageTips (C18) following standard protocol (Thermo Scientific). Peptide mixtures were dried in a SpeedVac concentrator (Savant ISS110, Thermo Scientific) and reconstituted in formic acid (0.5%) prior to nanoLC-MS/MS analysis.

4. Quantitative immunoblotting

For quantitative immunoblotting of SecY, SecE, SecG and SecA, same IMV protein amounts [5μg, estimated by a standard bicinchoninic acid (BCA) assay] were loaded for all samples. SecY was quantified through a standard curve of purified SecY using an α-SecY specific antibody (Davids Biotechnologie, GmbH) raised against the N-terminus of *E.coli* SecY (AKQPGLDFQSAKGGLGE, A274). The relative amounts of the three other proteins tested were quantified using specific antibodies raised against the whole protein [SecA, A56; [38]] or the N-terminus of SecE (A117) and SecG (A7) [39, 40]. For SecE that cannot be detected in wild type cells [39], its fold increase in SecYEG over-synthesizing IMVs was calculated as follows: over-synthesizing IMVs were

analyzed by western immunostaining at increasingly lower concentrations so that the limit of detection of the antibody was reached.

5. nanoLC-MS/MS analysis

An EASY-nLC system (Proxeon, software version 2.7.6 #1) coupled with an LTQ-Orbitrap XL ETD (Thermo Scientific, Bremen Germany) through a nanoES ion source (Proxeon) was used for sample analysis. Data were acquired with Xcalibur (LTQ Tune 2.5.5 sp1, Thermo Scientific). Peptides (0.5% formic acid) were separated on a C18 column packed in-house, using a gradient previously described [16]. MS parameters were: spray 2300V, capillary 35V, tube lens 140V and capillary temperature 180^{o}C. Survey scans were acquired in *m/z* 300-1800 with an AGC MS target value of 10^6 and the 10 most intense precursor ions were subjected to collision-induced dissociation in the ion trap (isolation width 3Da, normalized collision energy 35%, activation *q* 0.25, activation time 30ms).

6. Analysis of MS/MS-derived data

MS raw data were analyzed in Proteome Discoverer 1.1.0.263 (Thermo Scientific) using both Mascot 2.3.01 (Matrix Science) and Sequest (Thermo Scientific). The *Escherichia coli* (strain B / BL21-DE3) theoretical proteome [511693, December 2010] containing 4156 entries was used [41] with a list of common contaminants [42]. Search parameters were: precursor error tolerance 10ppm, fragment ion tolerance 0.8Da, trypsin full specificity, maximum number of missed cleavages 3 and cysteine alkylation as a fixed modification. False discovery rate (FDR) was calculated through a decoy database search with strict criteria set to 0.01 and relaxed criteria to 0.05. Final protein lists were compiled in Scaffold (version 4.4.1.1, Proteome Software; Portland, OR) employing the following criteria: Scaffold peptide probability >80%, Mascot ion score >25 (+2), >35 (+3,

+4), Mascot ion score-identity score >0, Sequest XCorr >2.5 (+2), >3.5 (+3, +4) and DeltaCn >0.1. Proteins identified on the basis of one unique peptide were accepted only if the peptide was identified in more than two replicate experiments. FDR's were calculated in Scaffold through a decoy database search and were <1% for peptide and <3% for protein identifications respectively.

7. Label-free quantitation of proteins

Abundance was calculated using the exponentially modified abundance index [emPAI, [43]] and the normalized spectral abundance factors [NSAF, [44]]. emPAI takes into account the number of tryptic peptides detected divided by the number of theoretically detectable peptides of a protein ('MS-detectable'), whereas NSAF takes into account the number of spectral counts divided by the protein's length normalized to the sum of these values for all proteins identified in the experiment. For calculating the abundance of proteins identified in the surface proteolysis approach by $emPAI_{SP}$, the number of 'MS-detectable' peptides was replaced by the number of trypsin-accessible surface peptides ('MS-detectable'$_{SP}$, Fig. 2C). For the calculation of $NSAF_{SP}$, the length of a protein was replaced by the 'protein length'$_{SP}$, calculated by adding the length of individual 'MS-detectable'$_{SP}$ peptides of a protein (Fig. 2C, 2G). For NSAF and $NSAF_{SP}$, zero spectral counts were replaced with the value of 0.01 as described [44]. Briefly, the distribution of the natural logarithm of spectral counts was tested for fitting to a Gaussian distribution using the Shapiro-Wilk normality test, upon replacement of zero values with a fraction of a unit spectrum. The smallest zero-replacement value not disrupting the normality of the logarithmic spectral counts distribution was selected.

8. Statistical Analysis

For the comparison of protein abundance in wild type (four biological repeats) and SecYEG over-expressed (three biological repeats) cells, the following statistical tests were applied: a) $emPAI_{SP}$: An unpaired two-sided t-test including the null-hypothesis that the variances of the two distributions are equal was employed. The average number of unique peptides was calculated among technical repeats within each biological replicate; a protein's abundance was considered significantly different if its p-value was <0.05. b) $NSAF_{SP}$: A beta-binomial test available in R package [45] was applied. An average number of the spectral counts of a protein across technical repeats within a biological sample was calculated; a protein's abundance was considered significantly different if its p-value was <0.05.

Results

Sub-cellular topologies of the *E.coli* BL21(DE3) theoretical proteome

We follow the comprehensive annotation of the sub-cellular topology of the proteome of *E.coli* BL21(DE3) presented in the STEPdb [19]. The annotation is based on protein sub-cellular location (e.g. IM, OM etc), their mode of interaction with a membrane (IM protein, PIM protein, lipoprotein) and the side of the IM on which the interaction occurs (internal or external side). These classes are relevant to all Gram$^-$ bacteria. The IM proteome (Fig. 1A) comprises ~34% of the theoretical BL21(DE3) proteome (Fig. 1B) [19].

How amenable is the IMP to MS-based detection by surface proteolysis?

We first sought to identify the physicochemical properties of the proteins and peptides of the theoretical IMP and determine to what extend these are 'MS-detectable' by surface proteolysis (Table S1). To this end we analyzed all IM proteins with respect to their length and hydrophobicity [46]; these display a bimodal distribution shifted to elevated hydrophobicity values (Fig. 2A, black line). We also determined the number of their predicted TMs using TMHMM [47] and Phobius [48] (Fig. 2A; contour density lines). IM proteins possessing 1 or 2 TMs (yellow) display a wide distribution of length and hydrophobicity and range from large and hydrophilic (frequently due to the presence of large soluble domains) to small and extremely hydrophobic (commonly due to only short soluble stretches). More TMs per polypeptide result in narrower distributions of length and in enhanced hydrophobicity (green and brown). Proteins possessing $\geq$7 TMs have limited range of hydrophobic values (brown) and largely account for the second peak of the bimodal distribution (black line).

We then performed *in-silico* trypsin digestion and analyzed the physicochemical features of the derived theoretical peptides. IM proteins possess long tryptic peptides and the majority of them have extreme GRAVY values (Fig. 2B, black solid line). We defined 'MS-detectable' peptides (Fig. 2C) those that have theoretical *m/z* values of 300<*m/z*<1800 (experimental range used this study) and satisfy the following criteria: **a.** they are at least 5 amino acids long, **b.** they have a molecular mass >600 Da and **c.** their *m/z* is <1800 Da when obtaining the highest theoretical number of charges possible (Fig. 2B, black dashed curve). ~8% of the IM peptides are anticipated to elude detection as they exhibit a reduced capacity for protonation (Fig. 2B; dark grey area). To investigate whether these are embedded in the lipid bilayer, we mapped their sequences against predicted TMs [49] and calculated their TM amino acid content (Fig. 2D; supporting information). Indeed, the majority (87%) of the 'MS-excluded' peptides has a high TM content >50% (grey bars). The remaining 'MS-detectable' peptide pool still contains hydrophobic and large peptides (Fig. 2B, black dashed curve), however more than half contain a TM content <5% (Fig. 2D, black bars). Given that surface proteolysis targets exposed IM regions (typically hydrophilic), we defined an 'MS-detectable' peptide sub-set specific to the approach, termed hereafter 'MS-detectable'$_{SP}$ in which, only peptides with TM content <10% were included (Fig. 2C). The majority of those were small and more hydrophilic (Fig. 2B, red dashed curve).

Removing these large and hydrophobic peptides from the protein sequence results in an overall decrease of the 'MS-detectable' peptides per IM protein (from an average of 16 to 11, Fig. 2E), that subsequently leads to a smaller chance of identifying these. Excluded hydrophobic peptides also significantly decrease the overall anticipated length coverage of IM proteins from ~80% to 20% (Fig. 2F).

Experimental determination of the IMP of *E.coli* BL21(DE3)

To experimentally characterize the IMP, we carried out surface trypsinolysis of intact IMVs treated in various ways to enrich for protein sub-populations (Fig. 1D, [16]). Resulting peptides were collected after centrifugal removal of membranes and analyzed by MS. Identification of IM proteins by this approach is based largely on their surface-exposed loops that a) possess Lys/Arg residues and b) give rise to tryptic peptides of appropriate size for MS detection (Fig. 2C) [50]. These soluble loops may vary significantly in length (6-1158 residues [51-53]).

Surface proteolysis of “intact IMVs” (sucrose fraction) that had not been subjected to any treatment to remove peripheral membrane and cytoplasmic proteins (Fig. 1D), identified 177 IM, 342 PIM and 298 cytoplasmic proteins (Table S2). PIM and cytoplasmic proteins are washed away upon treatment with chemicals and chaotropes [16]; washing the membranes with high salt concentration, EDTA and sodium carbonate at pH 11 prior to surface proteolysis resulted in a slight increase of the detected IM proteins (Fig. 1E, Table S3). The efficiency of each washing step was monitored by SDS-PAGE (Fig. S1). Urea treatment and sucrose flotation [54] significantly improved detection leading to the identification of 312 IM proteins (Fig. 1E). These IM proteins were previously “masked” by PIM or/and by non-specifically bound cytoplasmic polypeptides (Table S3). Addition of the non-ionic detergent dodecyl maltoside (DDM) to “treated IMVs” at concentrations low enough (0.34 mM, 2 x CMC) to disrupt hydrophobic interactions without extensive solubilization of the bilayer did not result in the identification of additional IM proteins (not shown).

Collectively, 3719 peptides were identified (Table S2) corresponding to 405 IM proteins (Fig. 1C, Table S2). 5-13 peptides per protein were identified for >50% of the detected IMP (Fig. 2E). These have an average length of ~13 AAs and most are rather hydrophilic (Fig. 2B). Sequence coverage of IM proteins in proteomics experiments is rather poor [55, 56], attributed to the presence of TM peptides that may be under-

represented in peptide pools either due to sample preparation limitations or suppressed during the ionization process due to detergent or abundant soluble ions present. Surface proteolysis circumvents this since it targets only soluble peptides, clearly sacrificing sequence coverage (protein coverage was ~15% for half of the detected dataset, Fig. 2F). However, TM content was not a discriminating factor for detection, indicating that the approach is governed by the presence of tryptic cleavage sites of surface-exposed regions of a protein. Single TM proteins were the most abundant class.

Label-free quantification of the IMP

To estimate the abundance of IM and PIM proteins, we performed surface proteolysis of the wild type "treated IMVs" (defined in Fig. 1D) and employed two independent label-free quantitation methods, emPAI and NSAF. The former takes into account the number of tryptic peptides detected divided by the number of theoretically detectable peptides of a protein [43], whereas the latter takes into account the number of spectral counts divided by the protein's length and normalized to the sum of these values for all proteins identified in the experiment [44].

We analyzed four biological replicates, each containing three or four technical repeats (Table S3D). IM proteins were the most abundant class followed by PIM proteins (282 and 98 respectively) (Table S3B). Values of protein synthesis satisfying quantification criteria (Materials & Methods) were obtained for 221 integral IM, 50 PIM proteins and 25 ribosomal proteins. Spectral counts of each protein identified were well correlated among biological replicates (Spearman's correlation coefficients >0.873) (Table S3D).

We initially calculated protein abundance values, using for emPAI the 'MS-detectable' peptide set ('MS-observable' in [43]) and for NSAF the total protein length, as reported in Uniprot. However, since only soluble segments of IM proteins are

anticipated to be detected with the sample preparation approach employed in this study, we modified both algorithms by replacing required values: a) for emPAI (hereafter emPAI$_{SP}$), the number of 'MS-detectable' peptides was replaced by the number of those anticipated to be detected only by *surface proteolysis* ('MS-detectable'$_{SP}$) and, b) for NSAF (hereafter NSAF$_{SP}$), the total protein length was replaced by the sum of the length of all 'MS-detectable'$_{SP}$ peptides within a protein ('Protein Length'$_{SP}$; Fig. 2C, Table S3B, Materials & Methods). To then estimate how protein abundance values generated by emPAI$_{SP}$ and NSAF$_{SP}$ were corrected compared to those calculated by emPAI and NSAF, for each protein we calculated the ratio of emPAI$_{SP}$/emPAI and NSAF$_{SP}$/NSAF [Fig. 3A, "corrected abundance (fold)"]. Calculated values were fitted in Lorentzian distributions with coefficients of determination >0.96 for emPAI and >0.98 for NSAF. For the majority of IM proteins (~70%), emPAI$_{SP}$ values were on average >1.23 fold higher than those of emPAI; NSAF$_{SP}$ values were either smaller or larger than respective NSAF ones. Modified emPAI$_{SP}$ and NSAF$_{SP}$ values exhibited improved linearity compared to the original algorithms when plotted against absolute IM and PIM protein concentrations reported in the literature for K12 [22, 57] (Fig. S2).

To illustrate our corrected label-free quantification, we show as an example polypeptides of the Sec translocation system that comprises IM (only SecYEG shown here) and PIM (only SecA shown here) proteins ranging in size from 11-102 kDa (Fig. 3B, Table S3B) [58]. For SecYEG that contain TM peptides, the 'MS-detectable'$_{SP}$ peptides (Fig. 3B, blue) range from 12-40% of the initial 'MS-detectable' peptides. This is depicted in detail for SecY that contains 10 TMs (Fig. 3C); peptides corresponding to TMs 3/4 and 7/8 and 10 do not contain trypsin cleavage sites and due to size are not anticipated to be detected by MS, irrespective of the sample preparation method (Table S1B). The application of the surface proteolysis filtering criteria (Fig. 2C) excludes peptides belonging to TMs 1,2,5,6 and 9 from the remaining 'MS-detectable' set. This

defines the 'MS-detectable'$_{SP}$ peptide pool of SecY to 27% of its total length. Abundance values derived from emPAI$_{SP}$ and NSAF$_{SP}$ were shown to be 1.8 and 1.9 folds higher than those obtained by the unmodified algorithms (Table S3B).

Finally, we isolated "treated IMVs" (Fig. 1D) from BL21(DE3) with over-produced SecYEG [59] and subjected them to surface proteolysis/nanoLC-MS/MS. In this case, three biological replicates with 3-4 technical repeats each were performed; for protein quantification, thresholds similar to those in the IMVs from the strain with wild type SecYEG levels were applied (Table S3C).Of the 265 IM proteins identified, 221 were quantified (Table S3B, S3C). SecY and SecE are in substantially increased amounts in the over-synthesizing IMVs using both quantification tools. SecG is detected at higher quantities with NSAF$_{SP}$ but not with emPAI$_{SP}$. emPAI yields higher errors for proteins with few observable peptides [60]. SecA is also detected at higher quantities on the SecYEG over-expressing IMVs, presumably due to the elevated levels of its receptor SecY. The values calculated using emPAI$_{SP}$ and NSAF$_{SP}$ are in good accordance with those obtained by quantitative immuno-blotting (Fig. 3B). The amounts of 50 IM, 16 PIM proteins and 19 ribosomal proteins (presumably derived from IM-attached ribosomes) were found at significantly altered amounts in the two IMVs (p-value <0.05, Fig. 3D). Of those, 39 and 46 proteins were detected at measurably higher (green) or lower (red) quantities respectively in the IMVs with elevated SecYEG.

Discussion

The detection of IM proteins strongly depends on sample preparation and MS analysis. Thus, depending on the IMP coverage targeted, an escalated effort may be required: a) for simple protein identification IM proteins may easily be identified through the characterization of their polar peptides using shotgun approaches [61], b) aiming for higher protein sequence coverage necessitates biochemical pre-fractionation at a protein or peptide or membrane level and finally c) for label-free quantification tailored methods that by-pass technical challenges and ensure efficient hydrophobic peptide recoveries are required. To date, ~19% of the theoretical *E. coli* BL21(DE3) IMP has been experimentally identified, ~15% of it quantified by the collective proteomics effort [Table S4B, [9-16, 18].

Here, we present a fast and simple pipeline for the determination of the IMP. We combined annotation of the *E. coli* BL21(DE3) proteome [19], an elaborate protocol for the production of IMVs, comprehensive theoretical analysis of ‘MS-detectable’ peptides by surface proteolysis and adjusted algorithms for the label-free quantitation of the IMP. We address three main proteomics goals, protein identification, coverage and quantification.

Given the high demand of fractionation approaches in terms of time, protein amounts, sample processing and cost [62] and considering the effort required in data meta-processing, surface proteolysis is favorable for rapid protein identification. Collectively, we identified 405 IM and 342 PIM proteins (Fig. 1E, Table S3A). Most IM proteins identified (397) fall under the Uniprot categories ‘inferred from homology’ (35%) and ‘predicted’ (63%) (Fig. 4A; Table S2A). Evidence at a protein level existed previously for only 8 BL21(DE3) IM proteins compared to K12 in which the presence of 70% of IM proteins has been experimentally confirmed [41]. Identified IM proteins belong

to the major functional classes: transport of ions and macromolecules (47%), cell structure (12%), metabolism (10%) and all other classes (31%) (Fig. 4B).

Low numbers of 'MS-detectable'$_{SP}$ peptides does not disfavour detection of abundant IM and PIM small proteins (e.g. the small subunit of succinate dehydrogenase cytochrome b556 that has five 'MS-detectable'$_{SP}$ peptides and is present at ~1220 copies per cell [57] was detected Table S2). We could identify and quantify with NSAF$_{SP}$ IM proteins containing few 'MS-detectable'$_{SP}$ peptides (<10) [Table S3B; emPAI is not preferred for proteins with few observable peptides [60]]. For almost half of the detected proteins, abundance values in K12 were not available and NSAF$_{SP}$ values obtained were comparatively low corroborating that a low number of peptides and low abundance may be refractory to detection. The highest NSAF$_{SP}$ values across these were obtained for the single-pass IM protein YajC (6 'MS-detectable'$_{SP}$ peptides and ~1960 copies per cell [57], Table S1) and proteins of the NADH-quinone oxidoreductase complex [NuoA (4 'MS-detectable'$_{SP}$ peptides, ~670 copies per cell), NuoK and NuoJ (2 and 3 'MS-detectable'$_{SP}$ peptides respectively, 215 and 13 copies per cell respectively)] [57].

Comparison of our data with 10 reported datasets [Table S4] indicated that we have identified the highest percentage of the IMP (Fig. 4C). The '*in-silico*' analysis of the 674 IM proteins that eluded detection, revealed that: **a.** 60% of these have few 'MS-detectable'$_{SP}$ peptides (<6) and **b.** 10% have 'MS-detectable' peptides only on the periplasm. Detection of the remaining 30% could be constrained by low expression levels in LB.

emPAI-derived quantification is frequently used in large-scale studies [63]; in BL21(DE3) these have been confined to analysis of the OM proteome [18]. Here, given the 'target group' of surface proteolysis, for the quantification of the IMP we made use of the 'MS-detectable'$_{SP}$ peptide pool to ensure quantitative accuracy. We thus determined the abundance of 301 IM, PIM and membrane-associated ribosomal proteins in wild type

BL21(DE3) growing in LB (Fig. 4D; Table S3B). Quantitative values suggest that the 40 most abundant IM and PIM proteins in BL21(DE3) belong to small molecule and protein transporters (including several drug efflux and multi-drug resistance transporters), the F_1 ATPase, proteases and chaperones, electron transport enzymes, LPS biosynthesis proteins, cell division and stress response proteins (Fig. 4D; dark red).

Finally, we analyzed “treated IMVs” over-expressing the SecYEG translocation channel. The Sec system is conserved in all three domains of life [64] and in bacteria, >95% of synthesized exported proteins utilize it for insertion in the membrane or for crossing to the periplasm and beyond [58, 65]. Those features make SecYEG extremely attractive for heterologous protein expression of biopharmaceuticals and industrial enzymes [66] with the aim of increasing protein secretion rate and consequently, the final yield of proteins of interest. SecYEG operates in oligomeric assemblies [67, 68] with other proteins such as SecA [67], therefore addressing protein composition and abundance at a proteome level allows assessment of the overall physiological effects of SecYEG over-expression on the membrane and dissection of interaction partners and assemblies during translocation.

Protein abundance values of “treated IMVs” over-expressing SecYEG calculated with the modified label-free methods are in good accordance with those obtained by quantitative immuno-blotting (Fig. 3B). Despite over-expression of the SecYEG translocation channel, only a small number of other IM proteins, that all use SecYEG for their membrane integration [69], showed elevated quantities in the membrane (i.e. the trehalose transporter PttBC and the mechanosensitive channel MscS; Fig. 3D, green; Table S3C). This suggests that other Sec pathway proteins necessary for secretion of CE proteins (e.g. YidC, SecF, SecD etc) may be limiting as they are not over-synthesized. Other IM proteins were detected at lower amounts (e.g. the DNA recombinase RmuC, the acyl-coenzyme A dehydrogenase FadE, the cellulose synthase

BcsB and the glutamate/aspartate transporter GltJ; Fig. 3D, red; Table S3C). These may be masked by Sec polypeptides on the membrane or their membrane insertion may be reduced due to stress (see below). The elongation factor Tu2 and many ribosomal proteins were also detected at significantly higher levels, presumably due to more ribosomes bound to more SecYEGs [69] and mediating co-translational translocation [58] (Fig. 4D). Ribosome/membrane associations may be further facilitated by the elevated levels of other ribosome receptors, ElaB and YqjD [70]. SecA that also binds SecYEG to catalyze post-translational protein translocation [71], was also detected at elevated levels (Fig. 3D). Although treatment of the IMVs with the various agents leads to extensive removal of PIM proteins, some, including SecA, are tightly bound and poorly removed [16].

SecYEG over-synthesis must be perceived as cellular stress as it induces the stress response PspB and PspC components of the Psp machinery [72-74], the HflK component of the HflKC (HflA) complex that binds to the FtsH (HflB) and regulates SecY proteolysis [75], the stress sigma factor-regulated ribosome-binding protein YqjD that causes ribosome accumulation to the membrane during stationary phase [70], the EFTu2 that associates with membranes [76] and MreB filaments [77, 78] more so under stress and the IM-docked but periplasm-active chaperone PpiD [79] (Fig. 5D; Table S3C).

Dissection of these events suggests that our approach can elucidate complex membrane-related protein cascades not previously seen at such breadth and resolution. The bird's-eye view snapshot (Fig. 4D) reveals the complexity of the molecular geography of the membrane. Nevertheless, it only offers a mere glimpse at the dynamics. Many of the complexes shown will co-exist in assemblies of different partners (e.g. Sec translocase) and ribosomes will bind to the membrane different receptors [e.g. ElaB, YqjD, YgaM and YhjD [70] [80]].

The IM determination and quantification pipeline presented here is applicable to investigations of the IM of other bacteria, including pathogens. Exciting future studies in *E.coli* and other bacteria will address complexome and interactome dynamics, cross-talk between the cytoplasmome, the CEP, the cytosceletal elements and the chromosome, the stress response and, finally, how proteostasis alters CEP complexes and interactions.

Acknowledgements

Research in our lab was funded by grants (to AE): Nonaco (Aristeia-Excellence #1473 and Sympa 09S¥• -13-832; Operational Programme for Education and Lifelong Learning co-funded by the European Social Fund and National Resources); KUL-Spa (Onderzoekstoelagen 2013; Bijzonder Onderzoeksfonds; KU Leuven); RiMembR (Vlaanderen Onderzoeksprojecten; #G0C6814N; FWO); StrepSynth (FP7 KBBE.2013.3.6-02: Synthetic Biology towards applications; #613877; EU). GO is an Onassis Foundation doctoral fellow.

References

[1] Han, M. J., Lee, S. Y., The *Escherichia coli* proteome: Past, present, and future prospects. *Microbiol. Mol. Biol. Rev.* 2006, *70*, 362-439.
[2] Tabor, S., Richardson, C. C., A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *PNAS* 1985, *82*, 1074-1078.
[3] Li, Z., Nimtz, M., Rinas, U., The metabolic potential of *Escherichia coli* BL21 in defined and rich medium. *Microb. Cell Fact.* 2014, *13*, 45-61.
[4] Shiloach, J., Rinas, U., in: Lee, S. (Ed.), *Systems Biology and Biotechnology of Escherichia coli*, Springer Netherlands 2009, pp. 377-400.
[5] Yoon, S., Han, M.-J., Jeong, H., Lee, C*., et al.*, Comparative multi-omics systems analysis of *Escherichia coli* strains B and K-12. *Genome Biol.* 2012, *13*, R37-49.
[6] Silhavy, T. J., Kahne, D., Walker, S., The bacterial cell envelope. *Cold Spring Harb. Perspect. Biol.* 2010, *2*, 1-16.
[7] Blattner, F., Plunkett, G., Bloch, C., Perna, N*., et al.*, The complete genome sequence of *Escherichia coli* K-12. *Science* 1997, *277*, 1453-1474.
[8] Hayashi, K., Morooka, N., Yamamoto, Y., Fujita, K*., et al.*, Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. *Mol. Syst. Biol.* 2006, *2*, 2006.0007.
[9] Wagner, S., Baars, L., Ytterberg, A. J., Klussmeier, A*., et al.*, Consequences of membrane protein overexpression in *Escherichia coli*. *Mol. Cell. Proteom.* 2007, M600431-MCP600200.
[10] Wagner, S., Klepsch, M. M., Schlegel, S., Appel, A*., et al.*, Tuning *Escherichia coli* for membrane protein overexpression. *PNAS* 2008, *105*, 14371-14376.
[11] Qian, Z.-G., Xia, X.-X., Choi, J. H., Lee, S. Y., Proteome-based identification of fusion partner for high-level extracellular production of recombinant proteins in *Escherichia coli*. *Biotech. Bioeng.* 2008, *101*, 587-601.
[12] Xia, X.-X., Han, M.-J., Lee, S. Y., Yoo, J.-S., Comparison of the extracellular proteomes of *Escherichia coli* B and K-12 strains during high cell density cultivation. *Proteomics* 2008, *8*, 2089-2103.
[13] Stenberg, F., Chovanec, P., Maslen, S. L., Robinson, C. V*., et al.*, Protein complexes of the *Escherichia coli* cell envelope. *J. Biol. Chem.* 2005, *280*, 34409-34419.
[14] Maddalo, G., Stenberg Bruzell, F., Gotzke, H., Toddo, S*., et al.*, A systematic analysis of native membrane protein complexes in *Escherichia coli*. *J. Proteom. Res.* 2011, 1848-1859.
[15] Spelbrink, R. E. J., Kolkman, A., Slijper, M., Killian, J. A., de Kruijff, B., Detection and identification of stable oligomeric protein complexes in *Escherichia coli* inner membranes. *J. Biol. Chem.* 2005, *280*, 28742-28748.
[16] Papanastasiou, M., Orfanoudaki, G., Koukaki, M., Kountourakis, N*., et al.*, The *E. coli* inner membrane peripherome. *Mol. Cel. Proteom.* 2013 *12*, 599-610.
[17] Han, M.-J., Yun, H., Lee, J. W., Lee, Y. H*., et al.*, Genome-wide identification of the subcellular localization of the *Escherichia coli* B proteome using experimental and computational methods. *Proteomics* 2011, *11*, 1213-1227.
[18] Thein, M., Sauer, G., Paramasivam, N., Grin, I., Linke, D., Efficient subfractionation of gram-negative bacteria for proteomics studies. *J. Proteom. Res.* 2010, *9*, 6135-6147.
[19] Orfanoudaki, G., Economou, A., Proteome-wide sub-cellular topologies of *E.coli* polypeptides database (STEPdb). *Mol. Cell. Proteom.* 2014, *13*, 3674-3687.
[20] Brandenburg, B., Zhuang, X., Virus trafficking – learning from single-virus tracking. *Nature Rev. Microbiol.* 2007, *5*, 197-208.

[21] Baeshen, M. N., Al-Hejin, A. M., Bora, R. S., Ahmed, M. M. M*., et al.*, Production of biopharmaceuticals in *E. coli*: Current scenario and future perspectives. *J. Microbiol. Biotechnol.* 2015, *25*, 953-962.
[22] Taniguchi, Y., Choi, P. J., Li, G. W., Chen, H*., et al.*, Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 2010, *329*, 533-538.
[23] Speers, A. E., Blackler, A. R., Wu, C. C., Shotgun analysis of integral membrane proteins facilitated by elevated temperature. *Anal. Chem.* 2007, *79*, 4613-4620.
[24] Wagner, S., Bader, M. L., Drew, D., de Gier, J.-W., Rationalizing membrane protein overexpression. *Trends Biotech.* 2006, *24*, 364-371.
[25] Rabilloud, T., Membrane proteins and proteomics: Love is possible, but so difficult. *Electrophoresis* 2009, *30*, S174-S180.
[26] Michel, P. E., Reymond, F., Arnaud, I. L., Josserand, J*., et al.*, Protein fractionation in a multicompartment device using Off-Gel™ isoelectric focusing. *Electrophoresis* 2003, *24*, 3-11.
[27] Wisniewski, J. R., Zougman, A., Mann, M., Combination of FASP and stagetip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *J. Proteom. Res.* 2009, *8*, 5674-5678.
[28] Erde, J., Loo, R. R. O., Loo, J. A., Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments. *J. Proteom. Res.* 2014, *13*, 1885-1895.
[29] Manadas, B., English, J. A., Wynne, K. J., Cotter, D. R., Dunn, M. J., Comparative analysis of OFFGel, strong cation exchange with pH gradient, and RP at high pH for first-dimensional separation of peptides from a membrane-enriched protein fraction. *Proteomics* 2009, *9*, 1-5.
[30] Wu, C. C., MacCoss, M. J., Howell, K. E., Yates, J. R., A method for the comprehensive proteomic analysis of membrane proteins. *Nat. Biotech.* 2003, *21*, 532-538.
[31] Speers, A. E., Wu, C. C., Proteomics of integral membrane proteins: Theory and application. *Chem. Rev.* 2007, *107*, 3687-3714.
[32] Bendz, M., Skwark, M., Nilsson, D., Granholm, V*., et al.*, Membrane protein shaving with thermolysin can be used to evaluate topology predictors. 2013, *13*, 1467-1480.
[33] Solis, N., Cordwell, S. J., Current methodologies for proteomics of bacterial surface-exposed and cell envelope proteins. *Proteomics* 2011, *11*, 3169-3189.
[34] Walters, M. S., Mobley, H. L. T., Identification of uropathogenic *Escherichia coli* surface proteins by shotgun proteomics. *J. Microbiol. Meth.* 2009, *78*, 131-135.
[35] Wurpel, D. J., Moriel, D. G., Totsika, M., Easton, D. M., Schembri, M. A., Comparative analysis of the uropathogenic *Escherichia coli* surface proteome by tandem mass-spectrometry of artificially induced outer membrane vesicles. *J. Proteomics* 2015, *115*, 93-106.
[36] Cirulli, C., Marino, G., Amoresano, A., Membrane proteome in *Escherichia coli* probed by MS3 mass spectrometry: a preliminary report. *Rapid Commun Mass Sp* 2007, *21*, 2389-2397.
[37] Van Der Does, C., Den Blaauwen, T., De Wit, J. G., Manting, E. H*., et al.*, SecA is an intrinsic subunit of the *Escherichia coli* preprotein translocase and exposes its carboxyl terminus to the periplasm. *Mol. Microbiol.* 1996, *22*, 619-629.
[38] Karamanou, S., Vrontou, E., Sianidis, G., Baud, C*., et al.*, A molecular switch in SecA protein couples ATP hydrolysis to protein translocation. *Mol. Microbiol.* 1999, *34*, 1133-1145.
[39] Douville, K., Price, A., Eichler, J., Economou, A., Wickner, W., SecYEG and SecA are the stoichiometric components of preprotein translocase. *J. Biol. Chem.* 1995, *270*, 20106-20111.
[40] Brundage, L., Hendrick, J. P., Schiebel, E., Driessen, A. J. M., Wickner, W., The purified *E. coli* integral membrane protein SecY E is sufficient for reconstitution of SecA-dependent precursor protein translocation. *Cell* 1990, *62*, 649-657.
[41] UniProt, The Universal Protein Resource (UniProt). *Nucleic Acids Res.* 2010, *38*, D142-D148.

[42] Rappsilber, J., Ryder, U., Lamond, A. I., Mann, M., Large-scale proteomic analysis of the human spliceosome. *Gen. Res.* 2002, *12*, 1231-1245.
[43] Ishihama, Y., Oda, Y., Tabata, T., Sato, T.*, et al.*, Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell. Proteomics* 2005, *4*, 1265 - 1272.
[44] Zybailov, B., Mosley, A. L., Sardiu, M. E., Coleman, M. K.*, et al.*, Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J. Proteom. Res.* 2006, *5*, 2339-2347.
[45] Pham, T. V., Piersma, S. R., Warmoes, M., Jimenez, C. R., On the beta-binomial model for analysis of spectral count data in label-free tandem mass spectrometry-based proteomics. 2010, *26*, 363-369.
[46] Kyte, J., Doolittle, R., A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 1982, *157*, 105 - 132.
[47] Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E. L., Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 2001, *305*, 567-580.
[48] Käll, L., Krogh, A., Sonnhammer, E. L. L., A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* 2004, *338*, 1027-1036.
[49] Käll, L., Krogh, A., Sonnhammer, E. L., Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. *Nucleic Acids Res.* 2007, *35*, W429-432.
[50] Griffin, N. M., Schnitzer, J. E., Overcoming key technological challenges in using mass spectrometry for mapping cell surfaces in tissues. *Mol. Cell. Proteom.* 2011, *10*, R110.000935.
[51] Fozo, E. M., Kawano, M., Fontaine, F., Kaya, Y.*, et al.*, Repression of small toxic protein synthesis by the Sib and OhsC small RNAs. *Mol. Microbiol.* 2008, *70*, 1076-1093.
[52] Hemm, M. R., Paul, B. J., Schneider, T. D., Storz, G., Rudd, K. E., Small membrane proteins found by comparative genomics and ribosome binding site models. *Mol. Microbiol.* 2008, *70*, 1487-1501.
[53] Dorazi, R., Dewar, S. J., Membrane topology of the N-terminus of the *Escherichia coli* FtsK division protein. *FEBS Lett.* 2000, *478*, 13-18.
[54] Gibeaut, D. M., Carpita, N. C., Separation of membranes by flotation centrifugation for in vitro synthesis of plant cell wall polysaccharides. *Protoplasma* 1990, *156*, 82-93.
[55] Eichacker, L. A., Granvogl, B., Mirus, O., Müller, B. C.*, et al.*, Hiding behind hydrophobicity. *J. Biol. Chem.* 2004, *279*, 50915-50922.
[56] Bernsel, A., Daley , D. O., Exploring the inner membrane proteome of *Escherichia coli*: which proteins are eluding detection and why? *Trends Microbiol.* 2009, *17*, 444-449.
[57] Wisniewski, J. R., Rakus, D., Multi-enzyme digestion FASP and the 'Total Protein Approach'-based absolute quantification of the *Escherichia coli* proteome. *J. Proteomics* 2014, *109*, 322-331.
[58] Chatzi, K., Sardis, M. F., Karamanou, S., Economou, A., Breaking on through to the other side: protein export through the bacterial Sec system. *Biochem. J.* 2013, *449*, 25-37.
[59] Karamanou, S., Bariami, V., Papanikou, E., Kalodimos, C. G., Economou, A., Assembly of the translocase motor onto the preprotein-conducting channel. *Mol. Microbiol.* 2008, *70*, 311-322.
[60] Ishihama, Y., Schmidt, T., Rappsilber, J., Mann, M.*, et al.*, Protein abundance profiling of the *Escherichia coli* cytosol. *BMC Genomics* 2008, *9*, 102.
[61] Yates, J. R., Mass spectrometry and the age of the proteome. *J. Mass Spectrom.* 1998, *33*, 1-19.

[62] Pirmoradian, M., Budamgunta, H., Chingin, K., Zhang, B.*, et al.*, Rapid and deep human proteome analysis by single-dimension shotgun proteomics. *Mol. Cel. Proteom.* 2013, *12*, 3330-3338.
[63] Neilson, K. A., Ali, N. A., Muralidharan, S., Mirzaei, M.*, et al.*, Less label, more free: Approaches in label-free quantitative mass spectrometry. *Proteomics* 2011, *11*, 535-553.
[64] Van den Berg, B., Clemons, W. M., Collinson, I., Modis, Y.*, et al.*, X-ray structure of a protein-conducting channel. *Nature* 2004, *427*, 36-44.
[65] Papanikou, E., Karamanou, S., Economou, A., Bacterial protein secretion through the translocase nanomachine. *Nat. Rev. Microbiol.* 2007, *5*, 839-851.
[66] Makrides, S. C., Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol. Rev.* 1996, *60*, 512-538.
[67] Gouridis, G., Karamanou, S., Sardis, M. F., Schärer, M. A.*, et al.*, Quaternary dynamics of the SecA motor drive translocase catalysis. 2013.
[68] Deville, K., Gold, V. A. M., Robson, A., Whitehouse, S.*, et al.*, The Oligomeric State and Arrangement of the Active Bacterial Translocon. *J Biol Chem* 2011, *286*, 4659-4669.
[69] Saraogi, I., Akopian, D., Shan, S.-o., Regulation of cargo recognition, commitment, and unloading drives cotranslational protein targeting. *J. Cell Biol.* 2014, *205*, 693-706.
[70] Yoshida, H., Maki, Y., Furuike, S., Sakai, A.*, et al.*, YqjD is an inner membrane protein associated with stationary-phase ribosomes in *Escherichia coli*. *J. Bacteriol.* 2012, *194*, 4178-4183.
[71] Chatzi, K. E., Sardis, M. F., Economou, A., Karamanou, S., SecA-mediated targeting and translocation of secretory proteins. *Biochim. Biophys. Acta* 2014, *1843*, 1466-1474.
[72] Sikdar, R., Simmons, A. R., Doerrler, W. T., Multiple envelope stress response pathways are activated in an *Escherichia coli* strain with mutations in two members of the DedA membrane protein Family. *J. Bacteriol.* 2013, *195*, 12-24.
[73] Darwin, A. J., Stress relief during host infection: The phage shock protein response supports bacterial virulence in various ways. *PLoS Pathog.* 2013, *9*, e1003388.
[74] Joly, N., Engl, C., Jovanovic, G., Huvet, M.*, et al.*, Managing membrane stress: the phage shock protein (Psp) response, from molecular mechanisms to physiology. *FEMS Microbiol. Rev.* 2010, *34*, 797-827.
[75] Kihara, A., Akiyama, Y., Ito, K., A protease complex in the *Escherichia coli* plasma membrane: HflKC (HflA) forms a complex with FtsH (HflB), regulating its proteolytic activity against SecY. *EMBO J.* 1996, *15*, 6122-6131.
[76] Jacobson, G. R., Rosenbusch, J. P., Abundance and membrane association of elongation factor Tu in *E. coli*. *Nature* 1976, *261*, 23-26.
[77] Defeu Soufo, H. J., Reimold, C., Linne, U., Knust, T.*, et al.*, Bacterial translation elongation factor EF-Tu interacts and colocalizes with actin-like MreB protein. *PNAS* 2010, *107*, 3163-3168.
[78] Liu, Z., Xing, D., Su, Q. P., Zhu, Y.*, et al.*, Super-resolution imaging and tracking of protein–protein interactions in sub-diffraction cellular space. *Nat. Commun.* 2014, *5*.
[79] Matern, Y., Barion, B., Behrens-Kneip, S., PpiD is a player in the network of periplasmic chaperones in *Escherichia coli*. *BMC Microbiol.* 2010, *10*, 251.
[80] Babu, M., Aoki, H., Chowdhury, W. Q., Gagarinova, A.*, et al.*, Ribosome-dependent ATPase interacts with conserved membrane protein in *Escherichia coli* to modulate protein synthesis and oxidative phosphorylation. *PLoS ONE* 2011, *6*, e18510.
[81] Cox, J., Mann, M., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotech.* 2008, *26*, 1367-1372.

[82] Papanikolau, Y., Papadovasilaki, M., Ravelli, R. B. G., McCarthy, A. A*., et al.*, Structure of dimeric SecA, the *Escherichia coli* preprotein translocase motor. *J. Mol. Biol.* 2007, *366*, 1545-1557.
[83] Serres, M. H., Riley, M., MultiFun, a multifunctional classification scheme for *Escherichia coli* K-12 gene products. *Microb. Compar. Gen.* 2000, *5*, 205-222.
[84] Keseler, I. M., Bonavides-Martínez, C., Collado-Vides, J., Gama-Castro, S*., et al.*, EcoCyc: A comprehensive view of *Escherichia coli* biology. *Nucleic Acids Res.* 2009, *37*, D464-D470.
[85] Aranda, B., Achuthan, P., Alam-Faruque, Y., Armean, I*., et al.*, The IntAct molecular interaction database in 2010. *Nucleic Acids Res.* 2010, *38*, D525-D531.
[86] Shannon, P., Markiel, A., Ozier, O., Baliga, N. S*., et al.*, Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003, *13*, 2498-2504.
[87] Sachelaru, I., Petriman, N.-A., Kudva, R., Koch, H.-G., Dynamic interaction of the Sec translocon with the chaperone PpiD. *J. Biol. Chem.* 2014, *289*, 21706-21715.

Figure legends

Figure 1: **Curated sub-cellular location of *E.coli* BL21(DE3) IM proteins and experimental proteome analysis.**

A, *E. coli* cell representation entailing sub-cellular topologies of its ribosomal and IM proteomes. Protein topology is: r, ribosomal; PIM, peripherally associated with the IM, facing the cytoplasm; IM, integral inner membrane protein.

B, Distribution of the *E. coli* BL21(DE3) annotated proteome across the sub-cellular topologies described in 1*A*. Annotation as in [19].

C, Distribution of experimentally detected proteins by nanoLC-MS/MS.

D, Experimental workflow employed in this study for IM protein enrichment.

E, IM protein enrichment at different sample preparation conditions employing surface proteolysis. “Intact IMVs”, fractionated in a sucrose gradient, were treated with chemicals, sodium carbonate at high pH and urea and were subsequently floated in a second sucrose gradient (termed “Treated IMVs). IM protein identification and quantity is increased across treatments (Table S3A) due to the removal of cytoplasmic and PIM proteins that occupy the membrane surface and mask IM proteins. Perseus (version 1.2.0.16), a part of the MaxQuant bioinformatics platform, was used for the heat map [81].

Figure 2: Physicochemical properties of integral IM proteins and peptides.

A, Hydrophobicity vs. average length plot of the IM proteins (Table S1A, supporting information). Contour density lines of integral IM proteins’ TMs predicted in TMHMM [47] and Phobius [48] are depicted.

B, Hydrophobicity vs. average length plots of theoretical and ‘MS-detectable’ tryptic peptides. The grey area corresponds to integral IM peptides excluded upon the application of peptide detection criteria (‘MS-excluded’, Fig. 2C, Table S1B)

C, Workflow for the determination of peptides anticipated to be detected by the surface proteolysis approach. *In-silico* trypsin generated peptides are filtered based on length, size and *m/z* ('MS-detectable'). 'MS-detectable' peptides with transmembrane content smaller than 10% of the total peptide length are likely to be detected by the surface proteolysis approach ('MS-detectable'$_{SP}$). 'MS-detectable'$_{SP}$ peptides and respective calculated protein length ('Protein Length'$_{SP}$) are used for the quantitation of the IM fraction using label-free approaches (see text for details).

D, 'MS-excluded', 'MS-detectable' and detected tryptic peptides containing predicted transmembrane regions (% of total peptide length) is shown here as a bar graph (supporting information).

E, F, G, Distributions of the number of peptides/protein, anticipated protein coverage and protein length for the theoretical, detectable (Fig. 2C) and detected peptide sets.

H, Number of TMs, as predicted in TMHMM [47] and Phobius [48] contained within proteins identified in this study.

Figure 3: **Quantitative analysis of the CEP proteome**

A, Abundance correction (fold) of IM proteins identified by surface proteolysis employing modified algorithms emPAI$_{SP}$ and NSAF$_{SP}$ (Table S3B). Fold change was calculated as the ratio of NSAF$_{SP}$ to NSAF and emPAI$_{SP}$ to emPAI. A value of 1 indicates no change in abundance.

B, Cartoon representations of SecY, SecE, SecG and SecA (PDB: 2FSF, [82]) topologies (top). Color annotation as in C. Peptides that fall outside *m/z* instrument experimental range are noted as 'MS-excluded' (TMs 3/4, 7/8, and 10 in SecY). Fold change of protein amounts as calculated by modified algorithms emPAI$_{SP}$ and NSAF$_{SP}$ of *E. coli* wild type and over-expressed SecYEG plasmid cells are shown. Statistically significant up-regulated proteins in the over-synthesized SecYEG strain are indicated.

Counts obtained by quantitative immune-blotting are also given. At bottom, quantitative immunostaining of SecY, SecE, SecG and SecA using appropriate antisera, in “treated IMVs” with wt and elevated levels of SecYEG (indicated as "+++").

C, Schematic map of SecY showing: transmembrane regions (Topology; TM; black); the theoretical peptides that satisfy instrument criteria (‘MS-detectable’; dark grey); the ‘MS-detectable’$_{SP}$ peptides specific to the surface proteolysis approach (blue); for SecY, ‘MS-detectable’$_{SP}$ peptides correspond to a maximum theoretical coverage of 27%; peptides experimentally identified by the surface proteolysis approach (orange). SecY trypsin cleavage sites are indicated with red triangles at the top. Accepted numbering for TMs at the bottom.

D, Comparison of emPAI$_{SP}$ and NSAF$_{SP}$ values for IM, PIM and ribosomal proteins detected in wild type and SecYEG over-synthesizing cells. Proteins detected with statistically significant higher or lower amount in the latter are in green and red contours respectively.

Figure 4: **Features of the IMP**

A, Protein existence as reported in Uniprot (469008, June 2014) for the *E. coli* BL21(DE3) IM proteome. The number of the experimentally identified proteins in this study is noted with a dashed contour per category.

B, Functional assignment of IM proteins detected in this study. Distribution is based on the MultiFun classification system for cellular functions [83]. MultiFun annotation for *E.coli* K12 proteins was retrieved from EcoCyc [84]. IM proteins are classified in ten MultiFun classes of which ‘Cell processes’ and ‘Information transfer’ are divided further in three subcategories [DNA-, RNA- and protein-related regulation for ‘Cell processes’ and Cell division, Adaptation to stress and Motility for ‘Information transfer’). Many IM proteins (~19% of the theoretical IMP) have not been assigned a function yet.

C, Venn diagram of theoretical, identified in this study and reported in MS-based BL21(DE3) proteomics studies to date IM proteins (Table S4) [9-18]. Collectively, 443 IM proteins have been identified, corresponding to 49% of the expressed BL21(DE3) IM proteome (supporting information).

D, Schematic longitudinal section of an *E. coli* cell. The cytoplasm surrounded by the IM is depicted; the rest of the cell envelope and the nucleoid are omitted for simplicity. Only 169 IM, PIM and ribosomal proteins (circles) are shown out of the 301 that were quantified (Table S3B). Proteins are organized by cellular function (MultiFun; as in Panel B). Proteins are colored based on $NSAF_{SP}$ values from less (light red) to more (dark red) abundant. The three levels of abundance were selected by following the Gaussian distribution of the $\log_2(NSAF_{SP})$ values, defining as medium level (red) the range $\pm\sigma$ around the mean point (Table S3B). Proteins are grouped according to known, biochemically isolated complexes [19, 84] (gray shapes) indicated with gray letters. Complexes are only shown for visualization; known stoichiometries are not depicted. Binary protein–protein interactions derived from tandem-affinity purification experiments [85] are also mapped (where available; black connecting lines). For visualization, protein–protein interactions were studied, integrated and loaded in Cytoscape [86] prior to integration. Ribosomal proteins, arranged in the 50S/30S subunits are depicted in proximity to the membrane recognized by the FtsY receptor during co-translational translocation of a chain through the SecYEG channel [58]. Auxiliary subunits of the SecYEG complex such as YidC and PpiD [87] are also depicted. N= N-terminus of nascent polypeptide chain.