



Comparative proteomic and transcriptomic profile of
Staphylococcus epidermidis biofilms grown in glucose-enriched
medium. *Talanta* (2015) 132:705–12

Virginia Carvalhais^{a,b,c}, Ângela França^{a,c}, Gerald B Pier^c,
Manuel Vilanova^{d,e}, Nuno Cerca^a, Rui Vitorino^b

^a CEB - Centre of Biological Engineering, University of Minho, Braga, Portugal

^b QOPNA, Mass Spectrometry Center, Department of Chemistry, University of Aveiro, Portugal

^c Division of Infectious diseases, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA

^d ICBAS - Instituto de Ciências Biomédicas de Abel Salazar, Universidade do Porto, Porto, Portugal

^e IBMC - Instituto de Biologia Molecular e Celular, Rua do Campo Alegre 83, Porto, Portugal

Corresponding author email: rvitorino@ua.pt

Abstract

Staphylococcus epidermidis is an important nosocomial agent among carriers of indwelling medical devices, due to its strong ability to form biofilms on inert surfaces. Contrary to some advances made in the transcriptomic field, proteome characterization of *S. epidermidis* biofilms is less developed. To highlight the relation between transcripts and proteins of *S. epidermidis* biofilms, we analyzed the proteomic profile obtained by two mechanical lysis methods (sonication and bead beating), associated with two distinct detergent extraction buffers, namely SDS and CHAPS. Based on gel electrophoresis-LC-MS/MS, we identified a total of 453 proteins. While lysis with glass beads provided greater amounts of protein, CHAPS extraction buffer allowed identification of a higher number of proteins compared to SDS. Our data shows the impact of different protein isolation methods in the characterization of the *S. epidermidis* biofilm proteome. Furthermore, the correlation between proteomic and transcriptomic profiles was evaluated. The results confirmed that proteomic and transcriptomic data should be analyzed simultaneously in order to have a comprehensive understanding of a specific microbiological condition.

Introduction

When *Staphylococcus epidermidis* establishes biofilms on indwelling medical devices, this may cause an early removal of the implant due to the lack of effective antibiotic therapeutics and risk of eventual systemic infection [1,2]. *S. epidermidis* biofilms grown in glucose-enriched medium were previously associated with an increased proportion of dormant cells within a biofilm [3]. Dormancy is a clinically relevant physiological state, since it has been associated with long-term bacterial survival, increased cellular tolerance to antibiotics [4-6] and evasion of the host immune system [3,7].

The availability of the complete genome of *S. epidermidis* strains RP62A (ATCC 35984) [8] and ATCC 12228 [9] has led to the development of proteomic studies, since it has been suggested that the genome sequence is not sufficient to elucidate the biological functions of an organism [10]. Although there have been major advances in the molecular characterization of the pathogenic mechanisms of *S. epidermidis* biofilms [1], much less is known regarding the proteome. A few proteomic studies, however, have been performed with some *Staphylococcus* spp. in order to identify specific features associated with the

pathogenicity and physiology of these microorganisms [11-18]. In *S. epidermidis*, 2-Dimensional Electrophoresis (2-DE) analysis of proteomic patterns showed several differentially expressed proteins when comparing commensal and invasive strains [12]. To ensure high quality and reliable proteomic results, an appropriate sample preparation is fundamental [19,20]. Due to the complex structure of biofilms [21], it is necessary to develop an effective lysis method in order to obtain maximum coverage of the biofilm proteome and minimal protein losses, similar to the approach optimized for total RNA extraction from *S. epidermidis* biofilms [22]. Different protein extraction methods, including enzymatic, chemical, mechanical and other methods available via commercial extraction kits have been tested to obtain the highest number of proteins in *Staphylococcus* spp. [11,14]. Although the majority of these studies were performed with cell suspensions, relatively harsh techniques have been shown to be rapid and efficient to disrupt and lyse biofilms of Gram-positive bacteria, such as mechanical methods like bead beating with glass beads (FastPrep) or sonication [14,23,24]. Often, to optimize protein recovery, enzymes and detergents may be used in conjugation with mechanical lysis. Since lysostaphin efficiently cleaves staphylococcal cell wall peptidoglycan [25,26], it may be used to disrupt staphylococcal biofilms [27]. However, due to high costs of lysostaphin, detergents are frequently used to enhance protein isolation and solubilisation [28].

In an attempt to determine the relation between protein and mRNA levels, several studies have shown that often the correlation is surprisingly low, and differs widely among organisms [29]. Correlation coefficients were found to vary from 0.09 to 0.46 in multi-cellular organisms, from 0.34 to 0.87 in yeasts, whereas in bacteria the correlations ranged from 0.20 to 0.47 (reviewed in [30]). Up to now, no correlation analysis between the transcriptome and proteome of *S. epidermidis* biofilms has been conducted.

Hence, herein, since distinct lysis and extraction methods may yield different protein recoveries, we assessed different lysis methods to obtain proteins from *S. epidermidis* biofilms grown in glucose-enriched medium. Then, we compared the proteomic profile with the gene expression profile obtained by RNA sequencing (RNA-seq) technology. We undertook a gel-based method to determine protein isolation efficiency, using total protein extracted with a detergent-based extraction (SDS or CHAPS) coupled with mechanical lysis (sonication or bead beating). A detailed analysis of proteomic data was performed in each condition. Label-free relative protein abundance index (emPAI) was used for the relative quantitation of the proteome and was compared to transcriptomic profile. The overall goal was to characterize and correlate both proteomic and transcriptomic profiles of *S. epidermidis* biofilm-grown cells.

Material and Methods

Growth conditions

Biofilm forming *S. epidermidis* 9142 strain was used as a model [31]. One colony was inoculated into 1 mL of Tryptic Soy Broth (TSB) (Liofilchem, Roseto degli Abruzzi, Italy) and incubated at 37°C at 120 rpm for 18 hours. The overnight culture was adjusted to an optical density at 640 nm of 0.250 (± 0.05) with TSB and 10 μ L of the suspension was transferred into a 24-well plate (Thermo Fisher Scientific, Waltham, MA, USA) containing 1mL of TSB supplemented with 0.4% glucose (v/v) (TSB 0.4% G) (Thermo Fisher Scientific). The culture plate was then incubated at 37°C at 120 rpm for 24 hours. After this period, the culture medium covering the biofilm was removed and replaced by fresh TSB supplemented with 1% glucose (v/v) (TSB 1% G). Biofilms were allowed to grow in these same conditions for 24 additional hours. Thereafter, biofilm culture medium was removed and biofilms were washed twice with phosphate buffered saline (PBS).

Protein preparation

The same number of biofilms (twelve biofilms for each condition) were directly scraped and suspended in detergent extraction buffers: 25mM Tris-HCl (pH=7.2) (Pharmacia Biotech, Uppsala, Sweden), 10mM CHAPS (Sigma-Aldrich, St. Louis, MO, USA), 0.5M NaCl (VWR, Radnor, PA, USA), 5% glycerol (Sigma-

Aldrich) and 1mM PMSF (Sigma-Aldrich) or 60mM Tris-HCl (pH=6.8), 10% glycerol, 5% SDS (USB Corporation, Cleveland, OH, USA) and 1mM PMSF. Mechanical lyses were performed in a sonicator (Cole-Parmer® 750-Watt Ultrasonic Homogenizer, Cole Parmer, Vernon Hills, IL, USA) (10 min, 30 sec running, 10 sec pause, 40% amplitude) or by bead beating, using glass beads of 0.1 mm (Sigma-Aldrich) in a FastPrep® cell disruptor (BIO 101, ThermoElectron Corporation, Thermo Fisher Scientific) (3 cycles of 30 sec and 6.5 meter/sec). After lyses, cell debris were removed by centrifugation (15,000 g for 15 min at 4°C) and proteins precipitated with cold acetone [32]. Then, protein quantification was performed using the RC-DC assay (Bio-Rad, Hercules, CA, USA), following the manufacturer's instructions.

One-dimensional gel electrophoresis (SDS-PAGE) in gel protein digestion and protein identification

Forty µg of protein were incubated with SDS 10% (w/v), 0.5M Tris-HCl (pH 6.8), glycerol, mercaptoethanol and bromophenol blue (w/v) for 5 min at 100°C. Then, samples were loaded on Novex NuPAGE® 4-12% Bis-Tris gel (Life Technologies, Grand Island, NY, USA) and proteins were separated at a constant voltage (200 V). The gel was stained with colloidal Coomassie G-250 and all the lanes were manually excised into 16 gel slices for in-gel digestion with trypsin TPCK (AB Sciex, Framingham, MA, USA). Peptide extraction was made with 10% formic acid/ 50% acetonitrile. Dried peptides were dissolved in 5% acetonitrile (VWR), 0.1% formic acid (Sigma-Aldrich) and 0.1% trifluoroacetic acid (Sigma-Aldrich). Separation of tryptic peptides by nano-HPLC was performed on the module separation Ultimate 3000 (Dionex, Thermo Fisher Scientific) using a capillary column (Pepmap100 C18; 3 µm particle size, 0.75 µm internal diameter, 15 cm in length). A gradient of solvent A, (water/acetonitrile/trifluoroacetic acid (98:2:0.05, v/v/v)) to solvent B (water/acetonitrile/ trifluoroacetic acid (10:90:0.045, v/v/v)) was used. The separation of 2 µg/µL sample was performed using a linear gradient (5-50 % B for 30 min, 50-70% B for 10 min and 70-5% A for 5 min) with a flow rate of 0.3 µL/min. The eluted peptides were mixed with a continuous flow of CHCA matrix solution (270 nL/min, 2 mg/mL in 70% ACN/0.1% TFA and internal standard Glu-Fib at 15 fmol) and applied directly on a MALDI plate in 20 seconds fractions using an automatic fraction collector Probot (Dionex, Thermo Fisher Scientific). Mass spectra were obtained on a matrix assisted laser desorption/ionization–time of flight MALDI-TOF/TOF mass spectrometer (4800 Proteomics Analyzer, Applied Biosystems, Foster City, CA, USA) in the positive ion reflector mode and obtained over the mass range from 700-4500 Da with 900 laser shots. A fragmentation voltage of 2 kV was used throughout the automated runs. The spectra were processed and analyzed by the T2S (v1.0, Matrix Science Ltd, London, U.K) and submitted in Mascot software (v.2.3.0.2, Matrix Science Ltd) for protein/ peptide identification based on MS/ MS data using the following criteria: trypsin as enzyme; a maximum of two missed cleavages; mass tolerances of 40 ppm for peptide precursors, mass tolerance of 0.6 Da was set for fragment ions. Simultaneously, phosphorylation of Threonine (T), Serine (S) and Tyrosine (Y) were searched as variable modifications. The local false discovery rate (FDR) was calculated by searching the spectra against SwissProt (Firmicutes, release date 06022013) decoy (random) database. Protein identification was considered reliable when the individual ion score for each peptide had a minimum individual score of 95% and a minimum sequence tag of four aminoacids. Relative quantitation was performed using empAI. Gene ontology [33] analysis was performed with proteins identified in each extraction condition using STRING database (Search Tool for the Retrieval of Interacting Genes/Proteins) (version 9.1) [34] (statistically significant specific GO terms, FDR adjust $p < 0.05$). Two independent experiments were performed with pools of *S. epidermidis* biofilms.

RNA sequencing

RNA extraction from *S. epidermidis* biofilms, cDNA library preparation and RNA-seq, were performed as previously described, with some modifications [35]. Briefly, total RNA was extracted using the RNeasy Mini kit (Qiagen, Valencia, CA, USA). To remove genomic DNA, Ambion® TURBO DNA-free™ kit (Life Technologies) was used followed by acid-phenol:chloroform precipitation (Ambion®, Life Technologies), both following the manufacturer's instructions. RNA integrity was determined using the Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA), and samples with RNA Integrity Number (RIN) above 8

were selected for complementary DNA (cDNA) library preparation. Ribosomal RNA was depleted from total RNA samples using the Ribo-Zero™ Magnetic kit for Gram-Positive bacteria (Epicentre, Madison, WI, USA), following the manufacturer's specifications. The Illumina® TruSeq™ RNA Sample Preparation kit (Illumina, San Diego, CA, USA) was used with the already purified mRNA, following the manufacturer's protocol. The construction of the libraries was rigorously validated by quantitative PCR and Hi-Sensitivity D1K TapeStation (Agilent 2200 TapeStation). Sequencing data were generated from paired-end reads (2 × 150bp) on a MiSeq® system (Illumina) using a RNA-seq library of 10 nM. The mapping to reference genome and normalization of gene expression was performed using CLC Genomics Workbench software (V.5.5.1, CLC Bio.). RNA-seq reads were aligned to the reference genome of *S. epidermidis* strain RP62A (GenBank accession number NC_002976). Gene expression was normalized by calculating Reads per Kilobase per Million Mapped Reads (RPKM) using the methods described by Mortazavi and colleagues [36], in which normalization is adjusted by the counts of reads per kilobase per million mapped and gene length. All genes with RPKM value under 1.00 were not considered for analysis. Three independent experiments, each one from a pool of four biofilms, were performed.

Results and Discussion

Proteomic characterization

In the present study, a proteomic and transcriptomic analysis of *S. epidermidis* biofilms grown in glucose-enriched medium was performed using gel-based proteomic separation and LC-MS/MS analysis along with RNA-seq technology, respectively. The proteomic profile obtained is highly dependent on the extraction procedure used, with quickness, handling and robustness of protein extraction methodologies being important criteria [11,37-39]. Before protein concentration, we used SDS-PAGE to compare the lysis efficiency of sonication and glass bead disruption of *S. epidermidis* biofilm cells (Fig. 1). Cellular lysis by the glass bead disruption method produced a higher protein yield compared to the performed sonication cycle. Due to sample and protein properties, precipitation may be required as a treatment to reduce or eliminate substances which could interfere with protein identification, as well as to concentrate proteins [40]. However, it is known that protein precipitation leads to specific changes in the protein composition of proteomic samples. These modifications depend, mainly, on the specific structure of the protein rather than the precipitating agent used [41]. Nevertheless, proteins recovered in each extraction method were precipitated, solubilised and quantified, followed by SDS-PAGE and LC-MS/MS analysis.

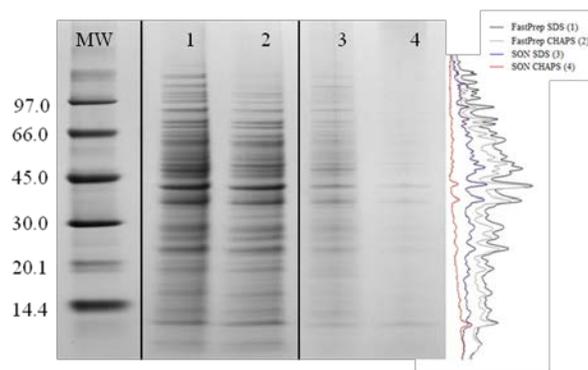


Figure 1 | SDS-PAGE before protein concentration and respective optical density traces. 1 - Protein profile obtained with FastPrep and SDS buffer extraction; 2 - Protein profile obtained with FastPrep and CHAPS buffer extraction; 3 - Protein profile obtained with sonication and SDS buffer extraction; 4 - Protein profile obtained with sonication and CHAPS buffer extraction. MW – molecular weight (kDa).

A total of 453 *S. epidermidis* proteins were identified, which corresponds to 17.96% of the predicted proteome (total of 2522 for *S. epidermidis* RP62A strain in UniProtKB database, at 29112013) with an increased confidence (FDR was below 5.5% among replicates). Venn diagram [42] demonstrated that 158 proteins were common to all extraction conditions (Supplementary Fig.S2). A total of 361 proteins were identified in FastPrep lysates with SDS, 391 in FastPrep with CHAPS, 359 in sonication lysates with SDS and 175 proteins after sonication in the CHAPS extraction buffer. The FastPrep protocol seemed to be the most efficient to lyse cells within *S. epidermidis* biofilms. Additionally, the CHAPS extraction buffer resulted in the identification of a higher number of proteins, compared to the SDS extraction buffer. Interestingly, this last finding, diverges from the work described by Encheva and colleagues [43], where it was shown that SDS gave a higher protein yield as compared to CHAPS, in an analysis of the *Streptococcus pneumoniae* proteome. SDS is an anionic detergent, used as a surfactant in protein denaturation, breaking lipid-lipid interactions and lipid-protein interactions rather than protein-protein interactions [44]. SDS is known to be very effective at solubilising almost all proteins [45,46]. However, it denatures them, causing loss of their native conformations and functions [47]. Alternatively, zwitterionic detergents such as CHAPS minimize protein-protein interactions [28,48]. Differences in extraction efficiency support the importance of extraction buffer to optimize protein yield or to increase yields of groups of proteins.

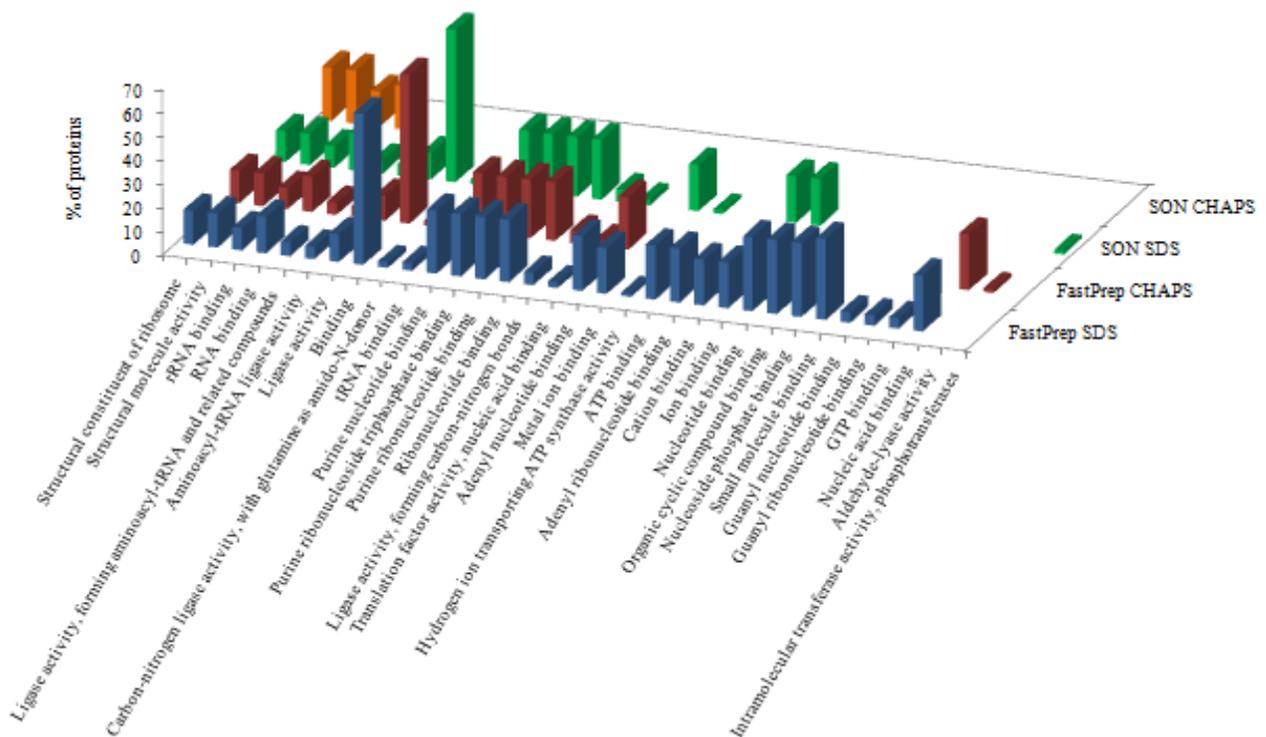


Figure 2 | GO classification. Molecular function specific terms with statistically significant enrichment ($p < 0.05$, FDR-corrected).

Next, Gene Ontology (GO) annotation, incorporating molecular function (Supplementary Table 1), biological processes (Supplementary Table 2) and cellular component (Supplementary Table 3) of proteins recovered from each extraction condition was analyzed. The main differences were observed in molecular function enriched categories (Fig.2). The GO terms adenyl nucleotide binding (GO:0030554) and nucleic acid binding (GO:0003676) were only found in protein extraction performed with FastPrep

lysis. Similarly, the GO terms for metal ion binding (GO:0046872), cation binding (GO:0043169) and ion binding (GO:0043167) were only found enriched in extractions performed with SDS buffer. These findings may be explained by the interaction established by SDS and cationic metal ions, due to the anionic head group of SDS [49]. While some studies showed that mechanical disruption of bacterial cells enhances the identification of cytoplasmic proteins [50,51], a number of cell membrane, secreted and cell surface proteins were also identified here. Membrane proteins are hydrophobic and they are present in low levels, since they are unstable and often aggregate, causing problems in analyzing membrane protein extractions [52]. However, within a heterogeneous protein properties set, it is difficult to find the conditions which will solubilise all proteins, including most membrane proteins, since not all are equally hydrophobic [53].

In an attempt to search for a methodological impact in the intrinsic properties of identified proteins, GRAVY score (grand average of hydropathy), molecular weight and pI were analyzed (Fig.3). Overall, pI , protein molecular weight and GRAVY score presented a similar tendency, among the tested conditions, with a few exceptions. Most of the identified proteins had a GRAVY score between -0.5 and -0.25 (Fig.3A). CHAPS extraction buffer allowed separation of a higher number of hydrophilic proteins (GRAVY score lower than -1.00) compared to SDS extraction buffer. Although gel-based approaches, e.g. SDS-PAGE, may allow recovery of a significantly higher number of hydrophobic proteins (positive GRAVY score), these extraction methods seemed to be less efficient at isolating hydrophobic proteins (positive GRAVY score). However, a few cell membrane proteins were identified, despite having high hydrophobicity. Our approach allowed us to identify only 3.3% of hydrophobic proteins. In proteins inferred from homology or with evidence at protein level (total of 997 proteins in UniProtKB/Swiss-Prot at 29112013), the number of hydrophobic proteins corresponds to 13.04%. Concerning pI , most of the isolated proteins showed a value between 4-6 or 9-10, similar to proteins inferred from homology or with evidence at protein level. More than 60% of proteins had a pI between 4 and 6 (Fig.3B).

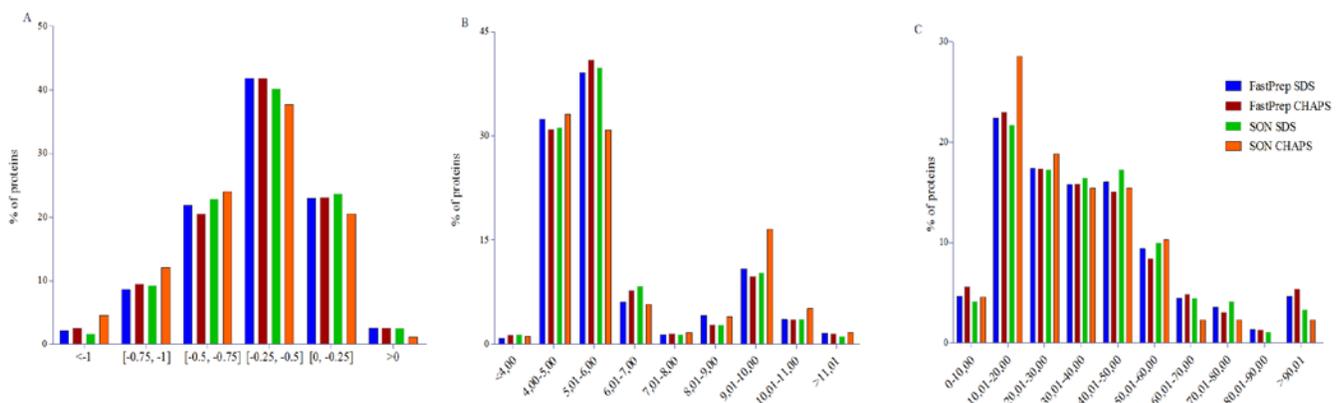


Figure 3 | Intrinsic properties of proteins. GRAVY (A), pI (B) and molecular weight (C) for each condition.

Gumber *et al.* [54] showed that non-ionic detergents, such as Tween20 and Triton X-100 were more efficient at extracting proteins with high pI . Conversely, CHAPS associated with urea was shown to extract more proteins with low pI [54]. Regarding molecular weight, most of the proteins presented a mass between 10 and 50 KDa (Fig.3C), which is also the most common range among the proteins inferred from homology or with evidence at the protein level. The FastPrep protocol allowed separation of higher numbers of proteins with more than 90 KDa compared to sonication. Moreover, the CHAPS extraction buffer combined with FastPrep seemed to be more efficient than SDS in obtaining proteins of less than 10 KDa. Independent of the methodological approach, a predominance of hydrophilic proteins, with 10-50 kDa and a pI between 4 and 6 was observed. Overall, major differences were not clearly

found in molecular weight, GRAVY and *pI*, and only the GO analysis showed evident differences between SDS and CHAPS detergent extraction buffers.

Transcriptomic characterization

Whole transcriptomic profile of *S. epidermidis* biofilm cells was obtained by RNA-seq. This methodology represents a valuable method to measure mRNA expression levels, since RPKM normalization allows comparison of transcript levels among and within samples [36]. A total of 2069 genes were found with a RPKM value above 1.00, which corresponds to 77.72% of the *S. epidermidis* genome (total of 2662 for RP62A strain in GenBank). More than 50% of transcripts presented RPKM values up to 250. The most represented GO slim terms in each RPKM group (RPKM < 150, 151>RPKM<250 and RPKM > 250) are represented in Fig.4.

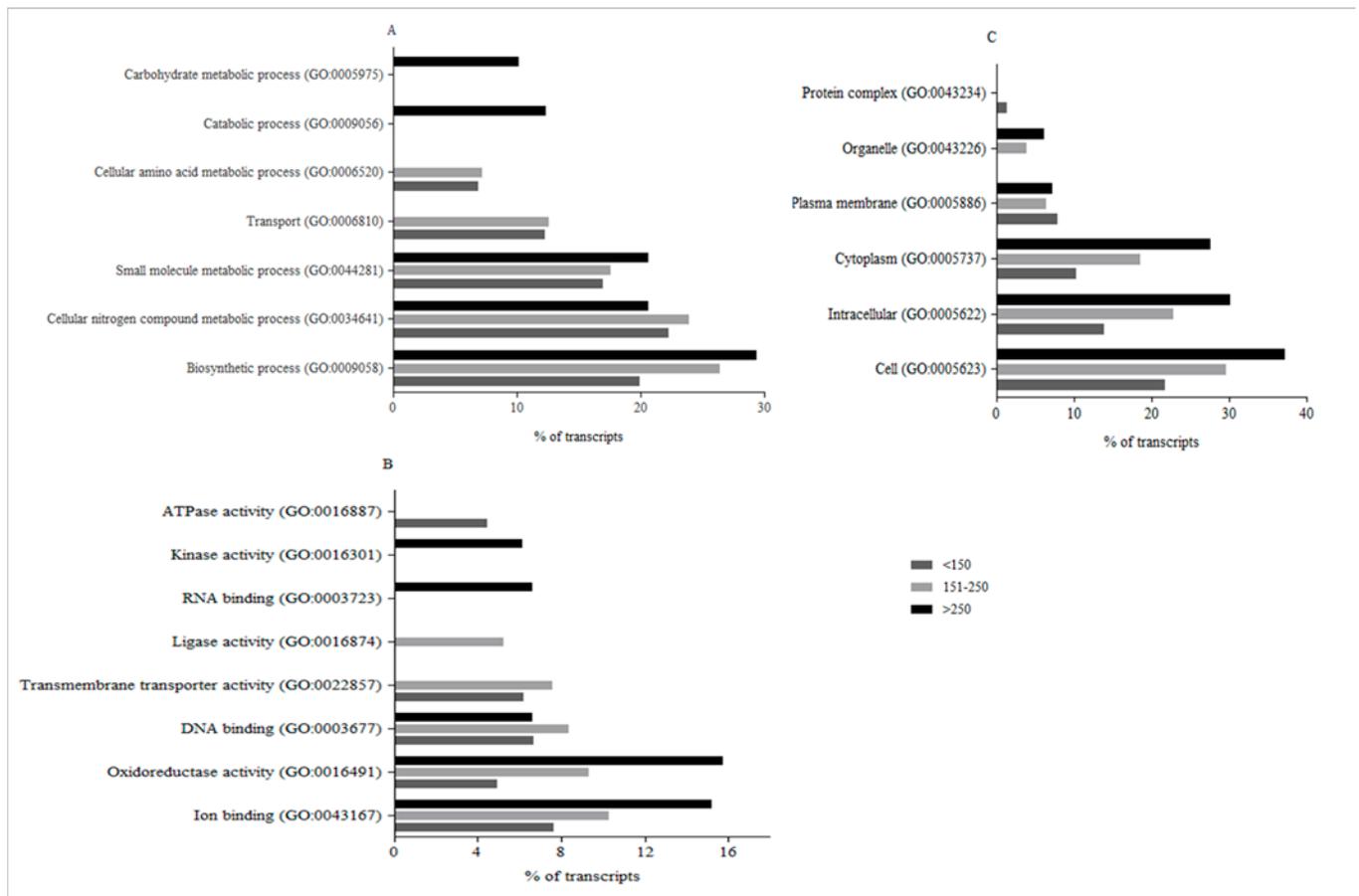


Figure 4 | GO classification of RNA-seq data. Five most represented GO slim terms of biological processes (A), molecular functions (B) and cellular components (C).

Additionally to common GO terms, RPKM values above 250 showed a predominance of catabolic (GO:0009056) and carbohydrate metabolic processes (GO:0005975) within the biological process GO category (Fig.4A). In contrast, transcripts with RPKM lower than 250 were associated with transport (GO:0006810) and cellular amino acid metabolic processes (GO:0006520), in addition to common GO terms. Molecular functions related to RNA binding (GO:0003723) and kinase activity (GO:0016301) were only found in transcripts with RPKM higher than 250 (Fig.4B). Regarding cellular components (Fig.4C),

protein complex category (GO:0043234) was amongst the most represented transcripts with RPKM lower than 150. Amongst the 10 most expressed transcripts are genes involved in response to stimulus, translation, catabolic processes and non-coding RNAs (Table 1). These were expected results since during growth, bacteria need to obtain energy. Moreover, increased expression of certain classes of genes may be related to an adaptive response [55-59]. It is well known that glucose has a key role in biofilm physiology since is generally used to promote biofilm growth *in vitro* [7,31]. However, high concentrations of glucose affect biofilm gene expression [31].

Table 1 | List of the 10 genes with higher RPKM values found in transcriptomic analysis and respective function.

Gene	RPKM	Gene function
<i>SERP_SeSRP1</i>	299510.4	Non-coding RNA involved in protein export
<i>SERP_SetmRNA1</i>	372738.7	Transfer-messenger RNA
<i>SERP1273</i>	28468.9	Universal stress protein
<i>SERP2379</i>	27311.7	Diacetyl reductase [(S)-acetoin forming] which catalyzes the irreversible reduction of 2,3-butanediol to (S)-acetoin in the presence of NADH
<i>asp23</i>	24019.2	Alkaline shock protein 23 which may play a key role in alkaline pH tolerance
<i>arcC</i>	11927.2	Carbamate kinase
<i>SERP2220</i>	13204.3	Universal stress protein
<i>SERP_SernpB1</i>	15829.8	Non-coding RNA
<i>spoVG</i>	13016.0	Septation protein
<i>SERP2343</i>	11865.1	Uncharacterized protein

Comparative proteomic and transcriptomic analysis

In order to compare proteomic and transcriptomic profiles, mass spectrometry and RNA-seq data were compared. Large scale proteomic experiments have been conducted using emPAI to determine, in a single LC-MS/MS experiment, the relative quantitation of proteins based on protein coverage by the peptide matches in a database search result [60-62]. Since the emPAI value has a linear correlation with protein concentration, it allows a more accurate estimation of protein abundance, as compared to simple peptide or spectral count [62,63]. In order to obtain an estimation of protein abundance, two independent experiments were performed. Then, emPAI values were compared within each condition (Supplementary Fig.S3). Importantly, emPAI presented the same tendency among proteins found in each extraction condition. Furthermore, reproducibility of experiments was also high, as indicated by Pearson correlation (Pearson's *r* correlation), which varied between 0.81 and 0.99. Based on these results, we chose to calculate the emPAI average for further analysis. Extremely abundant proteins may affect the efficiency of protein identification because of ionization suppression and detector saturation as well as the limited loading capacity of LC columns [62]. We observed that bead beating lysis led to a lower percentage of identified proteins with low emPAI (0-0.20) compared to those identified following sonication (Supplementary Fig.S4). Moreover, a combination of bead beating and CHAPS extraction buffer showed more proteins with higher emPAI values (>5.00).

Protein separation by SDS-PAGE and in-gel digestion decreased sample complexity and, consequently, contributed to avoiding suppression by extremely abundant proteins. Since proteins with similar abundances may functionally be correlated, proteins common to all conditions were further analyzed using CLUSTER 3.0 [64] (Fig.5). We found four main clusters which resulted from the hierarchical clustering (two main biological processes are represented in Fig.5). Proteins with lower emPAI tended to be involved in biosynthetic and cellular nitrogen compound metabolic processes (cluster A and D). Clusters B and C include mainly proteins with higher emPAI values. Protein cluster B is mainly related to biosynthetic processes and translation. Protein cluster C is related to biosynthetic and catabolic processes. Catabolic process was also found as one of the most prevalent GO term in transcripts with higher RPKM (Fig.4A). Interestingly, proteins with the highest emPAI values presented translation among the main biological processes. In accordance, molecular functions, such as structural constituent of ribosome and RNA binding, were found among these proteins. However, these biological functions were not evidenced in transcripts with the highest RPKM. More precisely, instead of translation, response to stress was a biological process found significant in these transcripts, among others. In relation to molecular functions, only oxidoreductase activity was a significant class. Despite a biological condition is not characterized by the most expressed transcripts or proteins, it is of utmost importance to undertake a whole bacteriome analysis using multi-approaches. RPKM obtained by RNA-seq was next compared to emPAI values obtained from protein extraction experiments performed with FastPrep and sonication (Fig.6). Correlations between mRNA transcripts and protein abundance obtained with FastPrep lyses were 0.37 and 0.36 (Pearson coefficient) in SDS and CHAPS extraction buffer, respectively. A somewhat smaller correlation coefficient was observed in sonication lyses with SDS ($r = 0.30$) and CHAPS extraction buffer ($r = 0.22$). Although this is not a strong correlation, our results are in agreement with correlation data previously described in bacteria (reviewed by [30]). When we zoomed Fig.6A and Fig.6B, we found that several proteins did not correlate with mRNA expression level. The main biological processes of these proteins are translation (when emPAI is higher than RPKM) and response to stress (when emPAI is lower than RPKM). Interestingly, dormancy is associated with low metabolic activity, leading to a decrease expression of transcripts involved in translation [65]. This suggests that these biological processes, in particular, may contribute to a lower correlation between transcriptomic and proteomic data. In almost every organism, transcript abundances only partially predict protein abundances, which suggests that other modes of regulation must be involved to explain how the levels of proteins are set within cells [29]. Also, the localization of proteins may display different correlations when compared to global correlation [66]. Despite direct correlation between proteins and mRNA is lower than 0.50, besides noise and experimental errors [67], several levels of regulation of transcripts and proteins may cause variability, such as transcription, post-transcription and post-translation regulation. Protein folding and post-translational modifications as phosphorylation, glycosylation, acetylation, proteolysis, methylation, lipidation, among others, are the most significant modifications [68-71]. Phosphorylation, such as found in LRTQKIVFS(Phospho)IGLCYSILMALSR peptide (Q5HLC3), corroborate that protein modifications increase phenotypic variability. These regulation mechanisms may explain why transcripts and proteins levels have a low correlation, among bacteria. Previously, it was shown that, in *Staphylococcus aureus*, serine and threonine phosphorylations was associated with metabolic activity, virulence and translation [72]. This addresses the biological importance of regulation processes and modifications over proteins.

Conclusions

Similar to other studies, we demonstrated that protein extraction methods affect the proteome characterization of *S. epidermidis* biofilms. It is assumed that an efficient protein extraction method is crucial to ensure successful proteome analysis. A combination of mechanical and chemical lyses or enzymatic lysis provides a better methodology to extract proteins from *S. epidermidis* biofilms. The ideal choice of a detergent is still empirical, although, optimization of sample preparation is fundamental to correctly analyze the cell proteome. Our data suggest that a higher number of total proteins from *S. epidermidis* biofilm can be extracted and recovered by a combination of glass beads mechanical lysis with CHAPS extraction buffer. However, only mildly hydrophobic proteins can be solubilised with SDS or

CHAPS. Also, the use of SDS extraction buffer led to an enhancement of metal ion binding proteins. A few lyse buffers have been used to enhance protein extraction, but one buffer is not sufficient for all purposes. This study emphasizes the need for a detailed study over extraction buffers that should be used for each experiment in order to identify the most efficient extraction methodology.

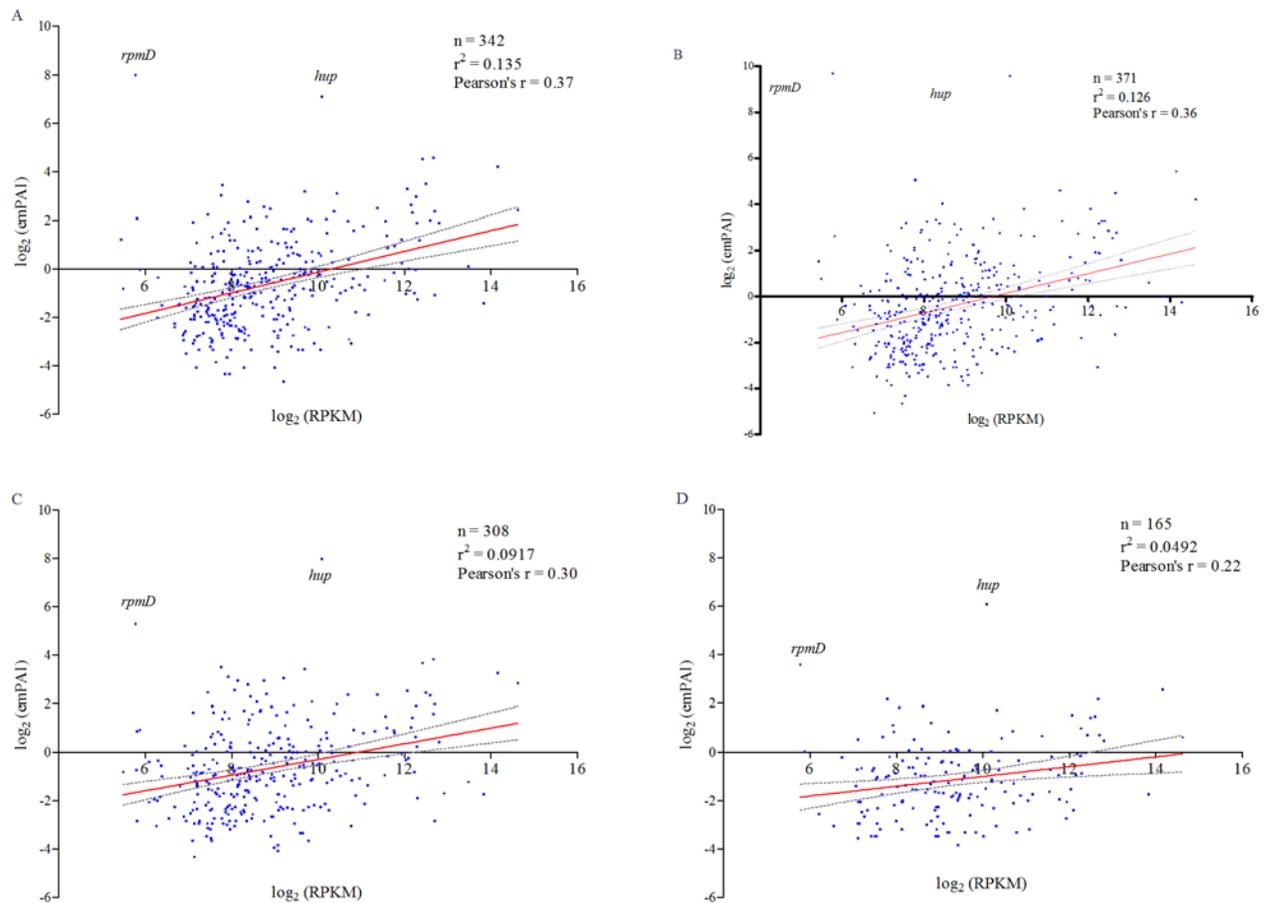


Figure 6 | Protein and transcript correlation in *S. epidermidis* biofilms grown in glucose enriched medium. FastPrep SDS (A), FastPrep CHAPS (B), SON SDS (C), SON CHAPS (D). Linear regression (R^2), Pearson correlation (Pearson's r) and number of proteins (n) are shown. The 95% confidence interval is shown in each condition, represented in grey dashed line. All correlations were significant ($p < 0.01$). SON - sonication.

More important, we provide evidences that transcriptomic and proteomic data present a similar correlation to other bacteria, despite the detergent extraction buffer used. Interestingly, proteins involved in translation or response to stress appear to be those which less correlate with the correspondent transcript. Several "omics" may provide useful information instead of individual transcriptome or proteome analysis, since it gives an integrative view of a specific physiological state.

Acknowledgments

The authors thank Stephen Lorry at Harvard Medical School for providing CLC Genomics software. This work was funded by Fundação para a Ciência e a Tecnologia (FCT) and COMPETE grants PTDC/BIA-MIC/113450/2009, FCOMP-01-0124-FEDER-014309, QOPNA research unit (project PEst-C/UI/UI0062/2013), RNEM (National Mass Spectrometry Network) and CENTRO-07-ST24-FEDER-002034. The following authors had an individual FCT fellowship: VC (SFRH/BD/78235/2011); AF (SFRH/BD/62359/2009). The authors also thank the FCT Strategic Project PEst-OE/EQB/LA0023/2013 and the Project "BioHealth - Biotechnology and Bioengineering approaches to improve health quality", Ref. NORTE-07-0124-FEDER-000027, co-funded by the Programa Operacional Regional do Norte (ON.2 – O Novo Norte), QREN, FEDER. The authors also acknowledge the project "Consolidating Research Expertise and Resources on Cellular and Molecular Biotechnology at CEB/IBB", Ref. FCOMP-01-0124-FEDER-027462.

Conflict of interests

The author(s) declare that they have no conflict of interests.

References

- [1] M. Otto, *Semin. Immunopathol.*, 34 (2012) 201-214.
- [2] L.G. Harris and R.G. Richards, *Injury*, 37 Suppl 2 (2006) S3-14.
- [3] F. Cerca, F. Andrade, A. Franca, E.B. Andrade, A. Ribeiro, A.A. Almeida, N. Cerca, G. Pier, J. Azeredo, and M. Vilanova, *J. Med. Microbiol.*, 60 (2011) 1717-1724.
- [4] K.S. Williamson, L.A. Richards, A.C. Perez-Osorio, B. Pitts, K. McInerney, P.S. Stewart, and M.J. Franklin, *J. Bacteriol.*, 194 (2012) 2062-2073.
- [5] J. Kim, J.S. Hahn, M.J. Franklin, P.S. Stewart, and J. Yoon, *J. Antimicrob. Chemother.*, 63 (2009) 129-135.
- [6] J.A. Shapiro, V.L. Nguyen, and N.R. Chamberlain, *J. Med. Microbiol.*, 60 (2011) 950-960.
- [7] Y. Yao, D.E. Sturdevant, and M. Otto, *J. Infect. Dis.*, 191 (2005) 289-298.
- [8] S.R. Gill, D.E. Fouts, G.L. Archer, E.F. Mongodin, R.T. Deboy, J. Ravel, I.T. Paulsen, J.F. Kolonay, L. Brinkac, M. Beanan, R.J. Dodson, S.C. Daugherty, R. Madupu, S.V. Angiuoli, A.S. Durkin, D.H. Haft, J. Vamathevan, H. Khouri, T. Utterback, C. Lee, G. Dimitrov, L. Jiang, H. Qin, J. Weidman, K. Tran, K. Kang, I.R. Hance, K.E. Nelson, and C.M. Fraser, *J. Bacteriol.*, 187 (2005) 2426-2438.
- [9] Y.Q. Zhang, S.X. Ren, H.L. Li, Y.X. Wang, G. Fu, J. Yang, Z.Q. Qin, Y.G. Miao, W.Y. Wang, R.S. Chen, Y. Shen, Z. Chen, Z.H. Yuan, G.P. Zhao, D. Qu, A. Danchin, and Y.M. Wen, *Mol. Microbiol.*, 49 (2003) 1577-1593.
- [10] K. Aggarwal and K.H. Lee, *Brief. Funct. Genomic. Proteomic.*, 2 (2003) 175-184.
- [11] R. Nandakumar, M.P. Nandakumar, M.R. Marten, and J.M. Ross, *J. Proteome. Res.*, 4 (2005) 250-257.
- [12] X.M. Yang, N. Li, J.M. Chen, Y.Z. Ou, H. Jin, H.J. Lu, Y.L. Zhu, Z.Q. Qin, D. Qu, and P.Y. Yang, *FEMS Microbiol. Lett.*, 261 (2006) 32-40.
- [13] C.L. Gatlin, R. Pieper, S.T. Huang, E. Mongodin, E. Gebregeorgis, P.P. Parmar, D.J. Clark, H. Alami, L. Papazisi, R.D. Fleischmann, S.R. Gill, and S.N. Peterson, *Proteomics.*, 6 (2006) 1530-1549.
- [14] A. Resch, S. Leicht, M. Saric, L. Pasztor, A. Jakob, F. Gotz, and A. Nordheim, *Proteomics.*, 6 (2006) 1867-1877.
- [15] D. Becher, K. Hempel, S. Sievers, D. Zuhlke, J. Pane-Farre, A. Otto, S. Fuchs, D. Albrecht, J. Bernhardt, S. Engelmann, U. Volker, J.M. van Dijl, and M. Hecker, *PLoS. One.*, 4 (2009) e8176.
- [16] C. Kohler, S. Wolff, D. Albrecht, S. Fuchs, D. Becher, K. Buttner, S. Engelmann, and M. Hecker, *Int. J. Med. Microbiol.*, 295 (2005) 547-565.

- [17] R. Monteiro, R. Vitorino, P. Domingues, H. Radhouani, C. Carvalho, P. Poeta, C. Torres, and G. Igrejas, *J. Proteomics.*, 75 (2012) 2892-2915.
- [18] S. Planchon, M. Desvaux, I. Chafsey, C. Chambon, S. Leroy, M. Hebraud, and R. Talon, *J. Proteome. Res.*, 8 (2009) 1797-1809.
- [19] M.M. Shaw and B.M. Riederer, *Proteomics.*, 3 (2003) 1408-1417.
- [20] M. Fountoulakis and B. Takacs, *Methods Enzymol.*, 358 (2002) 288-306.
- [21] H.C. Flemming and J. Wingender, *Nat. Rev. Microbiol.*, 8 (2010) 623-633.
- [22] A. Franca, A.I. Freitas, A.F. Henriques, and N. Cerca, *PLoS. One.*, 7 (2012) e37480.
- [23] S. Vilain and V.S. Brozel, *J. Proteome. Res.*, 5 (2006) 1924-1930.
- [24] C. Rathsam, R.E. Eaton, C.L. Simpson, G.V. Browne, V.A. Valova, D.W. Harty, and N.A. Jacques, *J. Proteome. Res.*, 4 (2005) 2161-2173.
- [25] B.L. de Jonge, Y.S. Chang, D. Gage, and A. Tomasz, *J. Biol. Chem.*, 267 (1992) 11255-11259.
- [26] J.K. Kumar, *Appl. Microbiol. Biotechnol.*, 80 (2008) 555-561.
- [27] J.A. Wu, C. Kusuma, J.J. Mond, and J.F. Kokai-Kun, *Antimicrob. Agents Chemother.*, 47 (2003) 3407-3414.
- [28] J.M. Neugebauer, *Methods Enzymol.*, 182 (1990) 239-253.
- [29] C. Vogel and E.M. Marcotte, *Nat. Rev. Genet.*, 13 (2012) 227-232.
- [30] A.R. de Sousa, L.O. Penalva, E.M. Marcotte, and C. Vogel, *Mol. Biosyst.*, 5 (2009) 1512-1526.
- [31] D. Mack, N. Siemssen, and R. Laufs, *Infect. Immun.*, 60 (1992) 2048-2057.
- [32] L. Jiang, L. He, and M. Fountoulakis, *J. Chromatogr. A*, 1023 (2004) 317-320.
- [33] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, and G. Sherlock, *Nat. Genet.*, 25 (2000) 25-29.
- [34] A. Franceschini, D. Szklarczyk, S. Frankild, M. Kuhn, M. Simonovic, A. Roth, J. Lin, P. Minguez, P. Bork, C. von Mering, and L.J. Jensen, *Nucleic Acids Res.*, 41 (2013) D808-D815.
- [35] A. Franca, V. Carvalhais, T. Maira-Litran, M. Vilanova, N. Cerca, and G. Pier, *Pathog. Dis.*, 70 (2014) 444-448.
- [36] A. Mortazavi, B.A. Williams, K. McCue, L. Schaeffer, and B. Wold, *Nat. Methods*, 5 (2008) 621-628.
- [37] J. Blonder, M.B. Goshe, R.J. Moore, L. Pasa-Tolic, C.D. Masselon, M.S. Lipton, and R.D. Smith, *J. Proteome. Res.*, 1 (2002) 351-360.
- [38] M.P. Washburn, D. Wolters, and J.R. Yates, III, *Nat. Biotechnol.*, 19 (2001) 242-247.
- [39] A. Tanca, G. Biosa, D. Pagnozzi, M.F. Addis, and S. Uzzau, *Proteomics.*, 13 (2013) 2597-2607.
- [40] A. Bodzon-Kulakowska, A. Bierzczynska-Krzysik, T. Dylag, A. Drabik, P. Suder, M. Noga, J. Jarzebinska, and J. Silberring, *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.*, 849 (2007) 1-31.
- [41] M. Zellner, W. Winkler, H. Hayden, M. Diestinger, M. Eliassen, B. Gesslbauer, I. Miller, M. Chang, A. Kungl, E. Roth, and R. Oehler, *Electrophoresis*, 26 (2005) 2481-2489.
- [42] J.C. Oliveros. VENNY. An interactive tool for comparing lists with Venn Diagrams. <http://bioinfogp.cnb.csic.es/tools/venny/>. 2007.
- [43] V. Encheva, S.E. Gharbia, R. Wait, S. Begum, and H.N. Shah, *Proteomics.*, 6 (2006) 3306-3317.
- [44] A.M. Seddon, P. Curnow, and P.J. Booth, *Biochim. Biophys. Acta*, 1666 (2004) 105-117.
- [45] M.P. Molloy, *Anal. Biochem.*, 280 (2000) 1-10.
- [46] S. Luche, V. Santoni, and T. Rabilloud, *Proteomics.*, 3 (2003) 249-253.
- [47] B. Hames, in: D. Rickwood and B. Hames (Eds.), *Gel Electrophoresis of Proteins. A Practical Approach*, New York, 1998, pp. 35-40.
- [48] Bollag DM, Edelstein SJ, and Rozycki MD, in: Bollag DM, Edelstein SJ, and Rozycki MD (Eds.), *Protein methods*, Wiley-Liss, New-York, 1996, pp. 1-26.
- [49] S. Sahu, A. Mishra, and G. Krishnamoorthy, *Analyst*, 138 (2013) 5942-5948.
- [50] K. Sauer, A.K. Camper, G.D. Ehrlich, J.W. Costerton, and D.G. Davies, *J. Bacteriol.*, 184 (2002) 1140-1154.
- [51] K. Sauer and A. Camper. *Journal of Bacteriology* 183, 6579-6589. 2001.
- [52] M.A. Churchward, R.H. Butt, J.C. Lang, K.K. Hsu, and J.R. Coorssen, *Proteome. Sci.*, 3 (2005) 5.

- [53] A.U. Freiburghaus, *Mol. Biotechnol.*, 2 (1994) 281-293.
- [54] S. Gumber, D.L. Taylor, and R.J. Whittington, *J. Microbiol. Methods*, 68 (2007) 115-127.
- [55] S.E. Cramton, M. Ulrich, F. Gotz, and G. Doring, *Infect. Immun.*, 69 (2001) 4079-4085.
- [56] J.K. Knobloch, K. Bartscht, A. Sabottke, H. Rohde, H.H. Feucht, and D. Mack, *J. Bacteriol.*, 183 (2001) 2624-2633.
- [57] S. Rachid, K. Ohlsen, W. Witte, J. Hacker, and W. Zieburh. *Antimicrobial Agents and Chemotherapy* 40, 3357-3363. 2000.
- [58] S. Dobinsky, K. Kiel, H. Rohde, K. Bartscht, J.K. Knobloch, M.A. Horstkotte, and D. Mack, *J. Bacteriol.*, 185 (2003) 2879-2886.
- [59] T. Dorr, M. Vulic, and K. Lewis, *PLoS. Biol.*, 8 (2010) e1000317.
- [60] S.E. Ong and M. Mann, *Nat. Chem. Biol.*, 1 (2005) 252-262.
- [61] A. Kudlicki, *PLoS. One.*, 7 (2012) e32339.
- [62] Y. Ishihama, Y. Oda, T. Tabata, T. Sato, T. Nagasu, J. Rappsilber, and M. Mann, *Mol. Cell Proteomics.*, 4 (2005) 1265-1272.
- [63] K. Shinoda, M. Tomita, and Y. Ishihama, *Bioinformatics.*, 26 (2010) 576-577.
- [64] M.J. de Hoon, S. Imoto, J. Nolan, and S. Miyano, *Bioinformatics.*, 20 (2004) 1453-1454.
- [65] K. Lewis, *Nat. Rev. Microbiol.*, 5 (2007) 48-56.
- [66] D. Greenbaum, C. Colangelo, K. Williams, and M. Gerstein, *Genome Biol.*, 4 (2003) 117.
- [67] S. Haider and R. Pal, *Curr. Genomics*, 14 (2013) 91-110.
- [68] L. Straub, *PLoS. Biol.*, 9 (2011) e1001146.
- [69] J.A. Cain, N. Solis, and S.J. Cordwell, *J. Proteomics.*, 97 (2013) 265-286.
- [70] T. Maier, M. Guell, and L. Serrano, *FEBS Lett.*, 583 (2009) 3966-3973.
- [71] M. Guell, E. Yus, M. Lluch-Senar, and L. Serrano, *Nat. Rev Microbiol.*, 9 (2011) 658-669.
- [72] K. Ohlsen and S. Donat, *Int. J. Med. Microbiol.*, 300 (2010) 137-141.