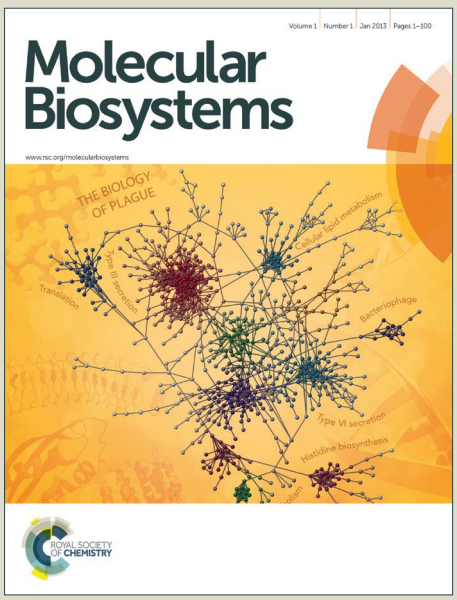
Molecular BioSystems

Accepted Manuscript

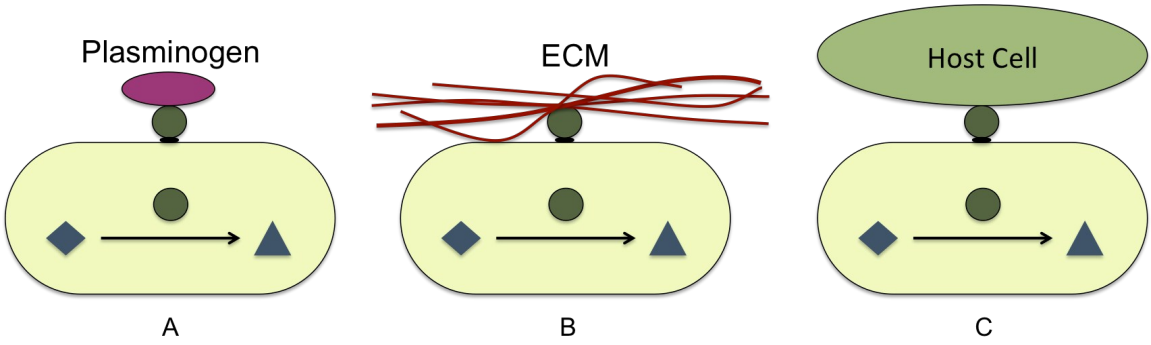


ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

W Wang and CJ Jeffery
Molecular Biosystems 2015
Graphical and textual abstract for the contents pages

Several dozen intracellular proteins are known to have a second function on the cell surface, sometimes referred to as “intracellular/surface moonlighting proteins”. An analysis of the results of 22 cell surface proteomics studies was performed to address whether the hundreds of intracellular enzymes and chaperones found on the cell surface in these studies could be due to experimental artifacts or could be candidates for also being intracellular/surface moonlighting proteins.



An Analysis of Surface Proteomics Results Reveals Novel Candidates for Intracellular/Surface Moonlighting Proteins in Bacteria

Wangfei Wang[a,b] and Constance J. Jeffery[*b]

[a]Department of Bioengineering, University of Illinois at Chicago, Chicago, IL 60607, USA

[b]Department of Biological Sciences, MC567, University of Illinois at Chicago, 900 S. Ashland Ave., Chicago, IL 60607, USA.

*To whom correspondence should be addressed: Constance Jeffery, Tel. 312-996-3168. Email: cjeffery@uic.edu.

Abstract

Proteins expressed on the bacterial cell surface play important roles in infection and virulence and can be targets for vaccine development or used as biomarkers. Surprisingly, an increasing number of surface proteins are being found to be identical to intracellular enzymes and chaperones, and a few dozen intracellular/surface moonlighting proteins have been found that have different functions inside the cell and on the cell surface. The results of twenty-two published bacterial surface proteomics studies were analyzed using bioinformatics tools to consider how many additional intracellular proteins are also found on the cell surface. More than 1,000 out of the 3,677 proteins observed on the cell surface lack the transmembrane alpha-helices or transmembrane beta-barrels found in integral membrane proteins and also lack the signal peptides found in proteins secreted through the Sec pathway. Many of the proteins found on the cell surface are intracellular chaperones or enzymes involved in central metabolic pathways, including some that have previously been shown to have a moonlighting function on the cell surface in at least one species, such as Hsp60/GroEL, DnaK, glyceraldehyde 3-phosphate dehydrogenase, enolase, and fructose 1,6-bisphosphate aldolase. The results of the proteomics studies suggest they could also be moonlighting on the surface of many other species. Hundreds of other intracellular proteins are also found on the cell surface, although a second function on the surface has not yet been demonstrated, for example, glutamine synthetase, gamma-glutamyl phosphate reductase, and cysteine desulfurase. The presence of intracellular proteins on the cell surface is more common than previously expected and suggests that many additional proteins might be candidates for being intracellular/surface moonlighting proteins.

1 Introduction

Cell surface proteins on bacterial pathogens play key roles in invasion of host cells and tissues. In nonpathogenic symbionts, they are important in commensual interactions with host species. Colonization of the host requires adhesion of the bacterium to the host cells and tissues, so some surface proteins bind to proteins in the extracellular matrix or directly to host cells, while others play additional roles in invasion and virulence.

Surface proteomics studies have previously been used to identify the proteins found on the cell surfaces of several bacterial species. Interestingly, the results of these experiments suggested that many bacterial proteins that are known to have a function inside the cell are also localized to the cell surface. Some of these observations might be experimental artifacts, but there are several dozen cases of intracellular proteins that have been experimentally shown to be used for a different function (or multiple functions) on the cell surface (Figure 1). These are referred to as intracellular/surface moonlighting proteins.

In general, moonlighting proteins are a subset of multifunctional proteins in which the multiple functions are not due to gene fusions, families of homologous proteins, promiscuous enzyme activity or pleiotropic effects[1–3]. To date, over 300 moonlighting proteins have been identified[4]. Many are receptors (e.g., neuropilin, which is a VEGF receptor in endothelial cells but the receptor for semaphorin III in nerve axons[5]), metabolic enzymes (e.g., phosphoglucose isomerase is a glycolytic enzyme and also a neurotrophic factor, an autocrine motility factor and a differentiation and maturation

mediator[6,7]), chaperones (e.g., some ATP-dependent proteases, including Clp and FtsH in bacteria also act as chaperones[8]), ribosomal proteins (e.g., human ribosomal protein L13a also binds to mRNA and regulates translation[9]) and transmembrane channels (e.g., cystic fibrosis transmembrane-conductance regulator is a chloride channel and a regulator of the ENaC channel[9]).

After the first discovery of a glycolytic enzyme, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), on the surface of group A streptococci[10,11], GAPDHs in several other species were also found to be intracellular/cell surface moonlighting proteins. On the cell surface, GAPDH can moonlight as an adhesin for cell binding in *Streptococcus suis* serotype 2[12]. In *S. pyogenes*, GAPDH is a cell surface receptor for plasminogen[10,13]. GAPDH has also been found to have multiple functions in other organisms. In gram-negative bacteria such as *E. coli*, GAPDH has been found to be secreted outside the cell where it plays a role in signal transduction[14]. Other glycolytic enzymes, such as enolase, are also intracellular/surface moonlighting proteins. In *Lactobacillus plantarum*, surface located enolase has the ability to bind fibronectin, a major host component of the ECM[15]. In some group A *Streptococci* species, streptococcal surface enolase (SEN) serves as a plasminogen-binding protein[16].

Glycolytic enzymes are not the only group of intracellular proteins that have been found to play a second role on the cell surface. In addition to playing a critical role in protein folding inside the cell, some chaperones can be released onto the cell surface. Hsp70/DnaK from *Bifidobacterium* was observed to be surface exposed by transmission electron microscopy and showed high affinity to human plasminogen[17]. *Legionella pneumophila* Hsp 60/GroEL was identified as an adhesin and was shown to promote

binding of this pathogenic bacterium to HeLa cells by which it can mediate invasion[18]. More examples of intracellular/surface moonlighting proteins are listed in Table 1.

Intracellular/surface moonlighting proteins not only exist in bacteria, but also in fungi, protozoan parasites and mammalian cells. For instance, Hsp60 has been identified by internal microprotein sequencing to be the cell surface binding protein for HDL in human cell lines [19]. The yeast *Candida albicans* uses an alcohol dehydrogenase as a cell surface receptor for human plasminogen [20].

The examples of intracellular/surface moonlighting proteins are by no means limited to the proteins listed above, but these examples prove that the presence of intracellular proteins at "unexpected" locations is not always because of experimental artifacts. In this study, the results of twenty-two published proteomics studies identifying proteins located on the cell surface of nineteen bacterial species were combined to determine what other known intracellular proteins are also located on the cell surface. Bioinformatics tools were used to remove integral membrane proteins and secreted proteins from the analysis so that the remaining proteins were predominantly those that have at least one function in the cytoplasm. Although some of the intracellular proteins might be identified in surface proteomics studies because of a close association with multiprotein complexes that interact tightly with the cytoplasmic side of the cell membrane, it is likely that many of the hundreds of proteins identified in this analysis do have a function on the cell surface. Some might perform the same function inside the cell as on the cell surface, but others are good candidates for being intracellular/surface moonlighting proteins.

2 Methods

2.1 Protein information used

The initial lists of proteins were obtained from the published surface proteomics projects listed in Table 2. The UniProtKB[20] and the NCBI protein databases[21] were used to collect protein amino acid sequences using the gene name, locus, UniProt accession number, and/or NCBI accession number provided in the proteomics papers. The amino acid sequences were obtained in FASTA format.

2.2 Prediction of transmembrane alpha-helices

TMHMM Server v.2.0[22] was used in the study to predict the presence of transmembrane alpha-helices (TMHs). FASTA sequences were used as input, and the proteins predicted to contain TMHs were removed from further analysis.

2.3 Prediction of signal peptides

The SignalP 4.1 Server[23] was used to predict the presence and location of signal peptides. Protein sequences in FASTA format were used as input. The program was run with organism group information based on the cell type because the program was trained using different datasets for Gram-positive and Gram-negative proteins. Gram-negative bacteria tend to have shorter signal peptides and different D-cutoff values. D-cutoff values were set to default. SignalP-TM network was used by default for Gram-positive bacteria. SignalP-noTM network was used for Gram-negative bacteria for a slightly better result.

2.4 Functional categories

COG annotation[24] was used to characterize the proteins' functional categories. COG employs seventeen single letter abbreviations to represent different functional categories.

3 Results

3.1 Identification of surface proteins that are expected to be intracellular

Information about 3,677 proteins was collected from the twenty-two surface proteomics papers listed in Table 2. The annotations regarding protein name and/or function were included. 467 of these proteins have annotations such as "hypothetical protein", "uncharacterized protein", "protein of unknown function", "conserved hypothetical protein", etc. After removing those proteins of unknown function that contain transmembrane alpha helices, transmembrane beta-barrels and/or signal peptides, we used BLAST to search for homologous proteins in the NCBI databases. For ninety of the proteins, we were able to predict a function based on a high level of sequence identity to other proteins with known functions. We listed these proteins in Supplementary Table 1 (Table S1). There were still some proteins in the list that did not have significant levels of sequence identity to proteins of known function. These remaining uncharacterized proteins were removed from the list of proteins for analysis because they don't have a known function, and we were looking for proteins with an intracellular function that are also found on the cell surface. In addition, 72 proteins don't have corresponding information in the UniProtKB/NCBI databases and therefore were left out of the study.

The rest of the proteins were then analyzed with TMHMM server v 2.0[22], and 746 were predicted to have TMHs. Proteins predicted to have TMHs were removed from further analysis because the goal was to find intracellular proteins that are observed on the surface, and proteins with TMHs would not be soluble intracellular proteins.

Although proteins with TMHs comprise the majority of transmembrane proteins, the number of proteins containing transmembrane beta-barrels (TMBs), is not negligible, comprising an estimated 1% of all integral membrane proteins. The prediction of TMBs is less accurate than of TMHs due to the nature of the TMB structures compared to TMH structures[25]. The use of algorithms to predict the presence of TMBs did not produce satisfactory results. For example, the FASTA sequence of maltoporin (LamB porins, formed by a TMB) from *E. coli* (gi| 253775338) was used as input into Pred-TMBB[25]. None of the methods, Viterbi, N-best and Posterior Decoding, predicted it as containing a TMB. The prediction score using Viterbi was 2.969, which is higher than the 2.965 threshold, suggesting that the protein does not contain a TMB. Also, this predictor gave some false positive results: a lot of cytoplasmic proteins were predicted to be outer membrane proteins with beta-barrels (data not shown). Therefore, proteins predicted to contain TMBs were manually selected based on SCOP (Structural Classification of Proteins) classification. SCOP classified these proteins into six superfamilies: OmpA-like, OmpT-like, OmpLA, porins, Txs-like channels and autotransporters[26]. Proteins with annotations similar to these superfamilies were selected and compared to the existing literature or the PDB database. 38 were classified as proteins with TMBs and thus left out of the study because we are analyzing intracellular/surface proteins, and most proteins with a TMB would not also be present as soluble intracellular proteins.

The remaining proteins without transmembrane helices or transmembrane beta barrels were analyzed using SignalP 4.1[23]. 431 proteins were predicted to harbor a signal peptide, which suggests they are secreted proteins and do not function inside the cell. SignalP is not able to identify bacterial lipoproteins because the lipoprotein signal peptide is cleaved by signal peptidase II (SPaseII) instead of signal peptide I (SPaseI). In bacteria, lipoproteins are abundant secretory or membrane-attached proteins[27,28]. Therefore, 21 lipoproteins were left out of the study because they usually do not have an intracellular function.

Up to this point, most of the proteins remaining for analysis are cytoplasmic proteins, and some may be candidates for being intracellular/surface moonlighting proteins. But there were still some membrane-associated proteins left. Many of the proteins identified in the surface proteomics studies function as part of transmembrane multiprotein complexes or as peripheral membrane proteins interacting with the cytoplasmic side of the membrane, so they may be tightly associated with the membrane even though they do not contain transmembrane helices or transmembrane beta-barrels. Proteins were removed for the analysis if they belong in any of the following categories: subunits of ABC transporters and other transmembrane transporters, ATP synthase subunits, flagellin and flagellar hook subunits, electron transport chain subunits, peripheral membrane proteins involved in determining cell shape, and proteins taking part in membrane associated complexes involved in cell division. For example, "ABC transporters and other transmembrane transporters" include proteins annotated as: ABC transporter/ABC transporter related, metal ion or small molecule transporter (cobalt, iron, arginine transporter, etc.), efflux pump, and part of a translocation system. Alpha, beta,

gamma, delta and epsilon subunits of the F1/F0 ATP synthase, located in the mitochondria matrix, were considered "ATP synthase subunits". The category of "Electron transport chain subunits" includes NADH dehydrogenase, succinate dehydrogenase, cytochrome c dehydrogenase, etc. FtsE, ftsF, ftsZ, ftsY, ftsA, divIVA are proteins that are related to cell division machinery and considered to be cell membrane associated[29]. 275 membrane-associated proteins including 53 ABC transporters and other transmembrane transporters, 32 ATP synthase subunits, 6 flagellin or flagella subunits, 38 electron transport chain subunits, 25 cell division associated proteins, and 121 other cell membrane-associated proteins were removed from the analysis.

3.2 Intracellular/surface moonlighting protein candidates

Supplementary Table 1 lists all the intracellular proteins that were identified on the cell surface from the surface proteomics studies. The NCBI numbers of the proteins were used to search the COG database[24], and the proteins were classified based on COG annotation. The database covers the majority of genomes of the organisms in this study (Table 2) except *E. coli* BL21, *C. thermocellum*, *R. parkeri*, *L. monocytogenes*, *S. suis, S. aureus*, and GBS, so proteins from those organisms were not included in the accounting of the COG classification, and some individual proteins are also not classified in the COG database.

The largest number of proteins in the study were in functional category J — translation, ribosomal structure and biogenesis[24]. This category consists of ribosomal proteins, tRNA synthetases, translation elongation factors/initiation factors and some other enzymes involved in mRNA degradation or tRNA processing. Another major

functional category is O — post-translational modification, protein turnover, chaperones[24]. This group comprises many molecular chaperones, including GroEL chaperonins, chaperone proteins DnaK and DnaJ, and Clp and Lon family proteases. Functional category G — carbohydrate transport and metabolism[24] is another highly represented group, especially proteins involved in glycolysis and the pentose phosphate pathway. The major functional groups are shown in Table 3.

Not all the proteins in the list are found in the COG database because the species or strain is not included, and some individual proteins have not been classified in the COG database, however, the proteins that aren't included in the COG database fall into similar functional groups: ribosomal/ribosomal associated proteins, molecular chaperones, aminoacyl tRNA synthetases, elongation/initiation/termination factors, enzymes involved in glycolysis, enzymes in replication and transcription, proteins involved in energy production, and proteins in nucleotide transport.

4 Discussion

This study is based on bioinformatics analysis of data from published proteomics studies. Hundreds of proteins that are known to have a function in the cytoplasm have been found in the cell wall or cell surface subproteomes. Some of the proteins may truly have a second localization on the cell surface, but we also must consider that some of the proteins may have been identified in the proteomics studies for other reasons, including artifacts of the methods used combined with tight association with the cytoplasmic side of the membrane or to membrane embedded proteins.

4.1 Comparison of proteomics experimental approaches

The proteomics studies used in this analysis employed variations of three types of experimental approaches to identify cell surface proteins. The first method is based on fractionation of the cells to isolate components of the cell membrane and/or cell wall. In these procedures, the proteins in the cell membrane or cell wall fraction are then separated using 2-dimensional gel electrophoresis (GE)[30,31], 1-dimensional GE[31,32] or chromatography[33] and identified via mass spectrometry. This method is sometimes called the "first generation proteomics" approach[34,35]. Among the papers in this analysis, the cell membrane or cell wall proteomes of *B. quintana*[30], *E. coli* BL21[32], *C. acetobutylicum*[31], *C. thermocellum*[36], *E. coli* B and *E. coli* K-12[37] were identified using these methods. Schaumburg and coworkers[38] also used this method to identify the cell wall-associated proteins of *L. monocytogenes*.

Subcellular fractionation works well for many proteomics studies, but the success of this approach in identifying only cell surface proteins depends in large part on how well the cell membrane is separated from the other subcellular fractions so that cytoplasmic proteins do not "contaminate" the membrane fraction because some separation techniques may not be very selective. In Boonjakuakul's study, nearly 1/3 of the identified surface-associated proteins are also cytoplasmic[30]. In Schaumburg's study[38], about 50% of the identified proteins are cytoplasmic proteins. Similarly, Yu and coworkers[36] and Thein and coworkers[32] also identified significant amount of intracellular proteins in the cell wall subproteome. Schaumburg and coworkers[38] excluded cell lysis as the reason for the presence of cytoplasmic proteins on the cell surface and verified that enolase is present on the cell surface by immunoelectron microscope. Yu and coworkers[36] detected many

typical intracellular proteins in the cell wall proteome and attributed it to protein moonlighting. Although many of the intracellular proteins found on the cell surface in this and other studies are known to be moonlighting proteins in at least one species, we cannot be certain how many of the other intracellular proteins found in the surface proteomes are truly located on the cell surface or if some were observed as artifacts of the cell fractionation methods.

The second proteomics technique, referred to as surface "shaving", is a gel-free method that uses proteolytic enzymes such as trypsin to digest proteins on the cell surface without damaging the cell integrity. Mass spectrometry is then used to identify the peptides resulting from the proteolysis[33,34]. This "second-generation proteomics" method[35] is a valuable tool for identification of cell surface-associated proteins. First described by M. Rodríguez-Ortega and coworkers[33], this method was proved to be able to quickly identify surface-exposed proteins in Group A Streptococcus (GAS) and was soon applied for other proteomics studies. Beside GAS, the shaving method or a modified shaving method was used in identification of the surfeomes of *E. faecalis*[39], *S. suis*[34], *P. aeruginosa*[40], GBS[41], *S. pneumoniae*[35], *L. rhamnosus*[42] and *C. canimorsus*[43] in Table 2.

Shaving has been proven to be more selective than cell fractionation for identifying peripheral membrane proteins[33] because weakly membrane-attached surface proteins may be lost during the chemical or physical separation steps of cell fractionation. Fewer integral membrane proteins are identified because many integral membrane proteins are embedded in the cell membrane and have only small extracellular loops or domains that might not be accessible to the proteases used in the shaving method.

A good shaving technique with appropriate digestion conditions can digest the surface proteins efficiently while not damaging the cell surface, and, theoretically, can result in the isolation of a low amount of cytoplasmic proteins. For example, only 4 intracellular proteins were identified among 72 proteins that were identified on the cell surface of GAS[33]. However, in most studies using the shaving method, intracellular proteins are still a major component of the surfeome. Bøhle and coworkers[39] reported almost half of the total identified proteins were cytoplasmic proteins using the shaving method. Olaya-Abril and coworkers[35] investigated the effect of digestion time on proteins that were identified by the shaving method in *S. pneumoniae*. The lowest percentage of cytoplasmic proteins that were identified constituted 65% of the total identified proteins after 30 min digestion[35]. It is interesting to note the numbers of cytoplasmic proteins that were identified in the surfeome vary greatly with different species. Some cytoplasmic proteins that are identified as being on the cell surface through these proteomics methods may still be artifacts from over-digestion of the cell surface. Intrinsic differences between species may be another factor, for example, it was noted that the cell walls of some Gram-positive species may be more resistant to damage by the shaving method than Gram-negative species. Vecchietti and coworkers[40] used a magneto-capturing method instead of surface shaving to identify the surfeome of Gram-negative bacterium *P. aeruginosa* because of the risk of impacting the outer membrane integrity by shaving method. Therefore, careful select of method is needed for different species. Schaumburg and coworkers[38] confirmed that a large amount of cytoplasmic proteins in the cell wall subproteomes are actually active on the surface, and they also

visualized enolase on the cell surface of *L. monocytogenes*. Enolase is also known to be a cytoplasmic/surface moonlighting protein in many other species.

The third general method involves cell surface labeling before isolation of the proteins for surface proteomics studies. Surface proteins are labeled and then are separated based on the properties of the labeling probes. Mass spectrometry is then used to identify the surface proteins. Pornwiroon and coworkers[44] biotinylated *R. parkeri* surface proteins and then identified the surface proteins with HRP conjugated streptavidin using a chemiluminescent substrate. Gibson and coworkers[45] and Sears and coworkers[46] also applied biotinylation labeling method in their surface proteomics study, but instead of immunoblotting, they purified the biotinylated surface proteins using affinity chromatography. Zhang and coworkers[47] labeled surface proteins with a membrane impermeable ^{18}O probe. In the studies discussed above, cell surface labeling is quantitative and can also be used to show the different abundance of different surface proteins. But this method is fairly time-consuming.

Some studies used several methods and compared the results. For example, Flores-Ramirez and coworkers[48] identified the proteome of *C. burnetii* using Triton X-114 partitioning, liquid-phase IEF and the shaving method. Hempel and coworkers[49] separated surface proteins by biotinylation and prepared extracellular proteins with precipitation of the supernatant. Le Maréchal and coworkers[50] combined shaving and surface labeling methods.

4.2 Reproducibility of different proteomics methods

Among the proteomics results used in this study (Table 2), two of the projects used *L. monocytogenes*. Schuamburg and coworkers[38] identified the cell wall subproteome using MALDI-TOF mass spectrometry (MS). Portillo and coworkers[51] used high resolution MS to study the cell wall proteome of intracellular and extracellular forms of *L. monocytogenes*. Fourteen of the proteins were found in the surface proteomes of both studies, suggesting the reproducibility of these proteomics results: Lmo2653, Lmo2068, Lmo1473, Lmo2459, Lmo2455, Lmo1657, Lom2556, Lmo0223, Lmo2458, Lmo2654, Lmo1055, Lmo1634, Lmo2456 and Lmo1314. They are EF-Tu, GroEL, DnaK, GAPDH, enolase, EF-Ts, fructose 1,6-bisphosphate aldolase, cysteine synthase, phosphoglycerate kinase, EF-G, dihydrolipoamide dehydrogenase, alcohol-acetaldehyde dehydrogenase, PGM and ribosome recycling factors, respectively. EF-Tu, EF-Ts, EF-G and ribosome recycling factors are associated with the ribosome. GroEL and DnaK are molecular chaperones. The rest are mainly metabolic enzymes. Most of these proteins have been discussed above and the identification on the cell surface by two different labs further support the model that the presence of these cytoplasmic proteins in the cell surface proteome is not due to experimental artifacts.

Overall, there is the possibility with each of the methods that some cytoplasmic proteins found in the surface proteomes might be due to experimental artifacts, but the large number of known intracellular/surface moonlighting proteins indicates that at least some of the proteins that are known to function in the cytoplasm were correctly found to be localized to the cell surface.

In the next sections, we discuss which cytoplasmic proteins found in the surface proteomes might have been there due to experimental artifacts and which might be good

candidates for being cytoplasmic/cell surface moonlighting proteins.

4.3 Proteins in DNA binding and replication

Many of the proteins identified in the surface proteomics studies are involved in DNA replication, recombination and repair (Table S1, Group L)[24], including DNA polymerase I (PolA), DNA polymerase III, DNA gyrase, RecA, DNA helicase, and single-strand binding protein ssb. DNA is known to bind to the cell membrane and form a DNA-membrane complex during replication in prokaryotic cells[52]. The interaction between DNA and the cell membrane is important for the transfer of the daughter chromosome during cell division[52,53]. This tight association between the DNA and the cell membrane may be an explanation for some of the proteins that interact with DNA being identified in the cell surface proteomics studies, and they might not be true cell surface proteins.

4.4 Proteins in Protein Biogenesis

Ribosomal proteins comprise the most common functional group found on our list (Table S1, Group J), with many from both the 30S subunit (S2, S3, S4, S5, S6, etc.) and the 50S subunit (L1, L2, L3, L4, etc.) observed on the surface of multiple species. Although a lot of ribosomal proteins are known to be moonlighting proteins and are listed in the MoonProt database[4], none of them are known to have a second function on the cell surface. Ribosomes interact with transmembrane translocons with high affinity during the co-translational assembly of membrane proteins and secreted proteins. This behavior has been proven to be evolutionarily conserved[54]. In addition, ribosomal proteins are often expressed at very high abundance. The combination of these factors suggest that

finding ribosomal proteins in the surface proteomes might be in part due to experimental artifacts, so some of the ribosomal proteins might not be true surface proteins.

Many aminoacyl-tRNA synthetases were also observed on the cell surface, with ones for alanine, glutamine, phenylalanine, serine, lysine, isoleucine, valine, aspartate, proline, glycine, arginine, threonine, tyrosine, and methionine found in the surface proteomes for multiple species (Table S1, Group J and the section containing Candidates not in the GOG Database) Aminoacyl-tRNA synthetases are enzymes that link the proper amino acid to a tRNA that contains the corresponding anticodon. They are important enzymes in protein biogenesis. During protein synthesis of membrane proteins, they also can be membrane associated. Some have been found to perform more than one function inside of the cell, but whether or not they can perform a second function on the cell surface is not clear. It is interesting that Olmedo-Verd and coworkers[55] found that four aminoacyl-tRNA synthetases have putative transmembrane domains in cyanobacteria.

Four elongation factors EF-Tu, EF-Ts, EF-G and EF-P were identified in several of the surface proteomics studies (Table S1, Group J and the section containing Candidates not in the GOG Database)). They are required for protein synthesis in prokaryotes. EF-Ts, EF-G and EF-P are not currently in the MoonProt database[4]. EF-Ts is the exchange factor that induces the release of GDP from EF-Tu[56]. EF- G binds to the ribosome and helps the translocation of tRNA from the A site to P site[57]. EF-P is required in peptide bond synthesis on ribosomes[58]. The close relation of these three elongation factors in protein translation with ribosomes may explain why these elongation factors were identified in the surface proteomics studies, while whether they are moonlighting proteins on the cell surface requires more experimental evidence. However, *L. johnsonii* and *L.*

plantarum EF-Tu have been identified to be intracellular/surface moonlighting proteins that mediate the attachment of the bacteria to human cells[59,60]. *Mycoplasma pneumonia* EF-Tu on the cell surface has been found to bind to host fibronectin[61]. *P. aeruginosa* EF-Tu has also been found to localize to the cell surface and bind to plasminogen[62]. The additional EF-Tu proteins in Table S1 might also be strong candidates for being intracellular/surface moonlighting proteins.

4.5 Molecular Chaperones

Several kinds of intracellular chaperones were observed in surface proteomes, with trigger factor, Hsp60/GroEL, Hsp70/DnaK, Clp protease, and DnaJ, observed on the cell surface for multiple species (Table S1, Group O and the section containing Candidates not in the GOG Database). Intracellular molecular chaperones from many species have been found to be true intracellular/surface moonlighting proteins because a second function on the cell surface has been identified experimentally. Several intracellular/surface moonlighting chaperones are included in the MoonProt database[4]. DnaK and GroEL are stress proteins that play a critical role in protein folding and help protect proteins from stress[63]. On the cell surface, several Hsp70/DnaK proteins show high affinity for plasminogen[17,38,64], and Hsp60/GroEL proteins function as adhesins[65–68] that help the bacteria to interact with host cells[69]. Katakura and coworkers[70] found that DnaK and Hsp60/GroEL on the cell surface of *L. lactis* can bind to mannoprotein of yeast cells and might help reduce the stress on the yeast surface caused by lactic acid. It is likely that many of the chaperones found on the cell surface of additional species in the proteomics studies are similarly performing second functions, but they have not yet been experimentally confirmed.

Several intracellular proteins appear to have a function on the cell surface, although it is not clear if it involves an additional biochemical function or if the protein is making use of its intracellular function in a second location. DnaJ, Clp and Lon family proteases were previously found on the cell surface in several species. Despite the lack of direct evidence of a specific biochemical function on the cell surface, they have been proven to promote intracellular growth of intracellular pathogens and thus enhance the virulence of these pathogens[71,72].

A few additional intracellular chaperones might be found in the surface proteomes because they have a transient interaction with the membrane or with membrane-associated proteins. DjlA, a member of Hsp40/DnaJ family, was found to be important for the growth of *L. dumoffii* in host cells, but it is possible that it was functioning in folding and transporting the membrane protein Dot/Icm, a part of the type IV secretion system (T4SS)[73]. Therefore, it might be only transiently associated with the cell membrane. Trigger factor is a cytoplasmic chaperone protein that can bind to and disassociate from the ribosome[74]. Ribosomal protein L23 has both binding sites for trigger factor and signal recognition particle and helps with protein targeting to the cell membrane[75]. This may explain the appearance of trigger factor in the cell membrane proteomes.

4.6 Metabolic enzymes

Many proteins from central carbohydrate metabolism, such as glycolysis, the citric acid cycle, and the pentose phosphate pathway were also observed in multiple surface proteomes (Table S1, Group G and C and the section containing Candidates not in the

GOG Database), including enolase, triose phosphate isomerase, transketolase, 6-phosphogluconate dehydrogenase, aldolase, glucose 6-phosphate isomerase, malate dehydrogenase, pyruvate kinase, phosphofructokinase, citrate synthase, and glyceraldehyde 3-phosphate dehydrogenase. It is interesting that most of the proteins used in glycolysis in bacteria can perform moonlighting functions on the cell surface in at least one species. They bind to plasminogen, fibronectin, laminin and can also serve as adhesins[69]. For example, enolase and GAPDH from *S. aureus* and *S. suis* have been identified to be moonlighting proteins and are listed in the MoonProt Database[4]. Other cytoplasmic enzymes in central metabolism are also surface moonlighting proteins in various species.

As has been observed for many types of moonlighting proteins, most of the known intracellular/surface moonlighting protein candidates, in particular the chaperones and enzymes in central metabolism, are ubiquitous in bacterial species (and in many cases eukaryotic and archael species as well) and have been present during evolution for a very long time. There are many examples like these of organism reusing proteins with functions that evolved long ago by adding new functional sites while retaining a more ancient function[76].

4.7 Other candidates for intracellular/surface moonlighting proteins

Because so many enzymes in central metabolism have already been found to moonlight on the cell surface in at least one species, it is possible that many of these enzymes identified in the surface proteomics projects are good candidates for being intracellular/surface moonlighting proteins. Similarly, proteins that belong to the

molecular chaperones category and were identified in the surface proteomics studies may be good candidates for being intracellular/surface moonlighting proteins. There are many other cytoplasmic proteins that were identified in the surface proteomics studies that don't belong to these two groups. Several alcohol dehydrogenases, involved in energy production and conversion[77], were found on the cell surface of *E. coli BL21*, *S. oneidensis*, *C. thermocellum*, *L. monocytogenes*, and *S. suis*. Among them, alcohol acetaldehyde dehydrogenase has been identified to act as an adhesin on the cell surface of *L. monocytogenes*[78]. This suggests the other alcohol dehydrogenases identified in the proteomics studies are strong candidates to be intracellular/surface moonlighting proteins as well. Similarly, 2-oxoglutarate dehydrogenase, an important complex in the TCA cycle, was found on the cell surface in several species (Table S1). In addition to the proteins discussed above, intracellular enzymes such as glutamine synthetase, phosphomannomutase, gamma-glutamyl phosphate reductase, ribose-phosphate pyrophosphokinase, adenylosuccinate lyase, cysteine synthase, formate acetyltransferase, acetate kinase, adenylate kinase, aspartate aminotransferase, tagatose 1,6-diphosphate aldolase, uracil phosphoribosyltransferase, phosphoglucosamine mutase, adenylosuccinate synthase, and cysteine desulfurase were also found on the cell surface in multiple species (Table S1, Groups E, F, and the section containing Candidates not in the GOG Database). It's not yet known if they have a function while on the cell surface, and if so if that function is the same or different from the function when inside the cell, but they could be good candidates for being intracellular/surface moonlighting proteins.

4.8 How are intracellular/surface moonlighting proteins secreted?

Questions remain about how these intracellular/cell surface proteins are

transported across the cell membrane. The known intracellular/surface moonlighting proteins do not possess signal peptides or known cell surface anchor motifs. Secretion without typical signal peptides was first found in eukaryotes, and was defined as non-classical secretion[79–81]. A few intracellular/surface proteins have been found to make use of known non-classical secretion pathways, but it is possible that there are additional non-classical secretion pathways that have not yet been identified.

4.8.1 Secretory Pathways in Gram-positive bacteria

Gram-positive bacteria have a single membrane with a thick peptidoglycan layer in the cell wall. In the signal peptide-mediated secretion system, many pre-proteins that are targeted to the cell surface have a signal peptide at the N-terminus and a sorting signal at the C-terminus that can be recognized and cleaved by sortase. The signal peptide directly leads secreted proteins from the cytoplasm into the extracellular environment, and after the cleavage of the sorting signal, the proteins become attached to the cell wall.

However, non-classically secreted proteins don't have signal peptides. Some potential non-classical pathways have been proposed, yet the underlying mechanisms are still unknown for most of the proteins. EAST-6 and CFP-10 are examples of two small proteins secreted by *M. tuberculosis* that lack signal peptides[82]. Pallen[82] used PSI-BLAST to identify many proteins containing amino acid sequence homology with EAST-6 and found that the small antigenic proteins apparently secreted by the same system as EAST-6 share a WXG motif but lack a signal peptide, suggesting a possible novel Gram-positive secretion pathway.

Another example of a non-classical secretion pathway is the SecA2 system. This

accessory Sec system has been identified in several Gram-positive bacterial species. Braunstein and coworkers[83,84] compared secreted proteins in wild-type and secA2 mutant *M. tuberculosis* and found three proteins including SodA that lack a signal peptide but are still secreted in a SecA2-dependent manner. It is interesting that two of the identified SecA2 dependent lipoproteins in *M. smegmatis* still possess a signal peptide[85]. In *L. monocytogenes*, SecA2-dependent proteins also include some proteins proteins with a signal peptide and some without[84,86]. It remains a question as to why the SecA2 secretion system exports proteins both with and without signal peptides.

4.8.2 Secretory Pathways in Gram-negative bacteria

Gram-negative bacteria have an inner plasma membrane surrounded by a single layer of peptidoglycan, which is surrounded by an outer membrane. The inner membrane is similar to the plasma membrane of other species, containing phospholipids and alpha-helical transmembrane proteins. The outer membrane contains beta-barrel transmembrane proteins, phospholipids, and lipopolysaccharide (LPS), which is only found in Gram-negative bacteria. Secreted or cell-surface proteins must be transported through the inner membrane, periplasmic space and the outer membrane. So far, at least six types of secretory pathways have been found in Gram-negative bacteria, named the type I-VI secretion systems (T1SS to T6SS). They can be classified as Sec dependent and Sec independent pathways. The Sec dependent secretion system generally relies on a signal peptide on the protein's N-terminus that leads to co-translational transport of proteins across the inner membrane to the periplasm[87]. In an alternative Sec independent secretion system, the twin-arginine translocation pathway (Tat), the secreted proteins are

targeted to the secretion pathway with highly conserved Tat signal peptides in a folded manner[88]. These secretion pathways cannot explain non-classical secretion behavior because of the requirement of a signal peptide, which is lacking in the cytoplasmic/cell surface proteins.

Fewer proteins have been identified that utilize non-classical pathways in Gram-negative bacteria. The type III secretion system (T3SS) is a Sec-independent pathway responsible for export of various proteins in many Gram-negative pathogens. Aguilera and coworkers[89] first reported the export of the house-keeping protein GAPDH through T3SS in enteropathogenic *E. coli* (EPEC) involving an interaction between GAPDH and CesT, the general chaperone of this secretion system. But more effort is needed to investigate why GAPDH can only be secreted in pathogenic *E. coli* strains like EPEC and enterohemorrhagic *E. coli* (EHEC), but not in non-pathogenic *E. coli* strains. Before Aguilera's report, Egea and coworkers[90] demonstrated that the medium condition is a key factor for GAPDH secretion by EPEC. This result shows that there are several secretion pathways for GAPDH in *E. coli,* and it points to the complexity of secretion pathways in Gram-negative bacteria.

The type IV secretion system (T4SS) has been found to be responsible for the translocation of effector proteins in Gram-negative pathogen *C. burnetii*[91]. Samoilis and coworkers [91] identified at least six possible effector molecules with no signal peptide secreted by *C. burneti*, including ribonuclease R, 30S ribosomal protein S2, alcohol dehydrogenase, and DNA repair protein RecN, which are also abundant in Table S1 in our study. Wu and coworkers [92] carried out a shotgun proteome study for a soil Gram-negative bacterium *Agrobacterium tumefaciens* and identified a Sec-independent secreted

unknown protein (Atu4345) that was designated as Hcp in their study. Hcp was expressed but not secreted with the deletion of T6SS operon[92].

All of these studies illustrate the complexity of the secretion pathways of Gram-negative bacteria. More studies are needed to understand how proteins with no signal peptides get secreted, and, in the case of intracellular/surface proteins, to identify the factors and potential regulatory mechanisms that enable most of the protein to remain in the cytoplasm while a portion is partitioned to the cell surface.

4.9 How do intracellular/surface proteins adhere to the cell surface?

After secretion of these intracellular proteins, they become anchored to the surface of the bacterial cells. There are several known mechanisms for the attachment of proteins to the cell surface, although these mechanisms generally require that the protein contains a signal sequence, which is not present in the intracellular/surface proteins that are the focus of this study. Some pre-proteins in Gram-positive bacteria, including bacilli, listeria and staphylococci, harbor a signal peptide at the N-terminus and a sorting signal, the LPXTG motif, at the C-terminus[93]. These pre-proteins are targeted to the secretion pathway via the signal peptide and become anchored to the cell wall after the cleavage of the LPXTG sorting signal by sortase A in these bacteria, albeit the cleavage mechanisms are different in different species[94]. *S. aureus* can code a protein with an NPQTN sorting signal at the C-terminus that can be recognized and cleaved by sortase B, thus anchoring the protein to murein[95]. Other species also have sortase B homologues, but the mechanisms are somewhat different[96]. A different anchor mechanism was found in *S. pneumonia*. Pneumococcal surface protein (PspA) has a N-terminal signal peptide

and C-terminal choline-binding repeats, anchoring to the cell surface via a choline-mediated interaction[93,97]. Almost all of these systems share a common characteristic that the proteins are secreted through use of signal peptides at the N-terminus and then anchored on cell wall via a motif at the C-terminus. A few additional motifs have been identified that are found in a small number of proteins (reviewed in 94, 98, 99), including the GW repeat, the choline binding motif, and the LysM domain, but these are not found in the vast majority of the proteins in Table S1.

Accompanied by the discovery of more and more non-classically secreted proteins, it has been found that some surface proteins perform their functions on the cell surface without LPXTG or NPQTG motifs or choline-binding repeats, sometimes referred to as an "anchorless" mechanism[93]. Chhatwal[93] listed several proteins, including *S. pneumonia* PavA, Eno and *S. pyogenes* FBP54, SEN and SDH that lack signal peptides and known membrane anchors, indicating that the re-association of these secreted proteins could be a general but as-yet-unknown mechanism.

There are not as many published studies about how intracellular proteins in Gram-negative bacteria become attached on the cell surface after secretion, but it is possible that these intracellular/surface moonlighting proteins adhere to the cell surface by re-association in both Gram-positive and Gram-negative bacteria. Using electron microscopy, it was shown that the association behavior of 60 kDa Hsp in *H. ducreyi*[66] shares similarity with the re-association of *S. pneumonia* alpha-enolase on the cell surface[100]. Similar transmission electron microscopy studies of *N. meningitides* also manifested the association of enolase, DnaK and peroxoredoxin with the outer membrane after secretion[64]. All of these results suggest that many of the intracellular/surface

moonlighting proteins re-associate on cell surface of pathogens after secretion, although it is not known to which components of the cell surface, proteins, lipids, etc., the proteins bind.

5 Conclusions

This analysis of the results of twenty-two surface proteomics studies identified hundreds of proteins that are known to have a function inside the cell as being located on the cell surface. While some of the proteins may have been identified in the surface proteomics studies because of a close association with other proteins or structures that interact tightly with the cytoplasmic side of the cell membrane, several dozen of the intracellular proteins are likely to play roles on the cell surface. Some may be performing the same biochemical function on the cell surface as they do inside the cell, but others may be performing a second function, such as the known intracellular/surface moonlighting enzymes and chaperones that bind to plasminogen or extracellular matrix, or directly to host cells as adhesions. Whether or not these proteins are also true intracellular/surface moonlighting proteins will require experimental verification. Future work also includes searching for common characteristics among the identified intracellular/surface proteins that might provide clues to their mechanisms of secretion or membrane anchoring.

Acknowledgements

We would like to thank Michael Chen, Ahlam Shaabneh and Shivani Patel for assisting in gathering protein data from published surface proteomics studies.

Figures

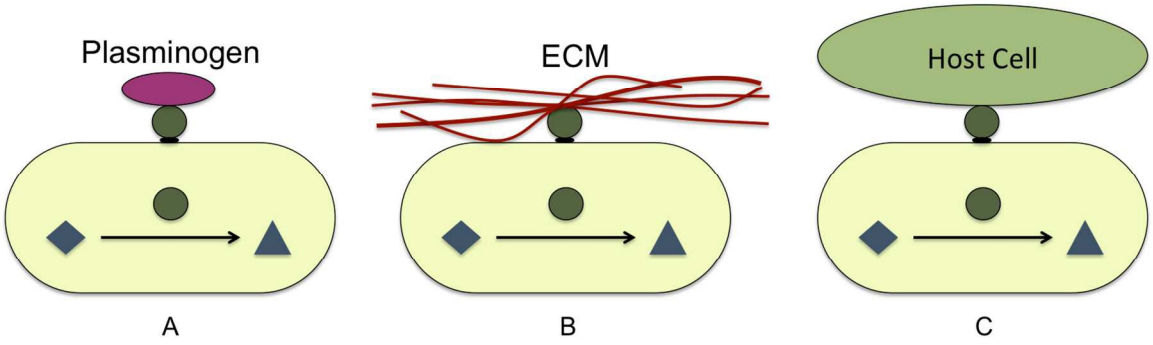


Fig. 1 Intracellular proteins that perform another function on the cell surface. An intracellular/surface moonlighting protein can function as an enzyme inside of the cell, converting a substrate (diamond) to a product (triangle) (A-C) and can also be present on the cell surface. Some of them can bind to plasminogen and convert plasminogen to plasmin (A), or bind to extracellular matrix (ECM) proteins such as fibronectin, collagen and laminin, or interact with host cell surface proteins (C). These intracellular proteins on the cell surface often play an important role in invasion, infection and virulence.

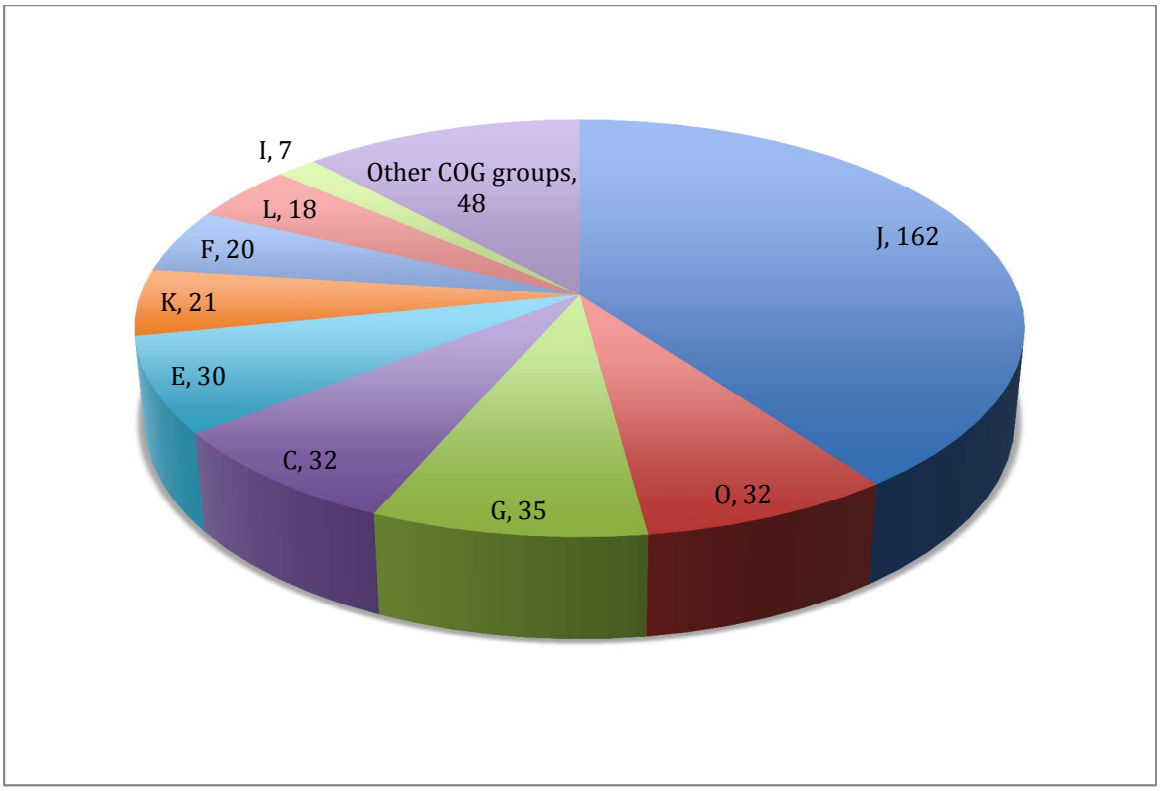


Fig. 2 Intracellular functions of intracellular/surface proteins grouped by COG classification. Each single letter corresponds to a COG code representing a functional group. The number following the comma indicates the number of proteins observed in the surface proteomics studies within that functional group. (The COG database includes information about proteins for the majority of genomes of the organisms in this study except *E. coli* BL21, *C. thermocellum*, *R. parkeri*, *L. monocytogenes*, *S. suis*, *S. aureus*, and GBS, so proteins from those organisms were not included in the accounting of the COG classification, and some individual proteins are also not classified in the COG database.)

Tables

Table 1 Examples of previously identified bacterial intracellular/surface moonlighting proteins in the literature

Intracellular function	Surface function	Species	Gram type
6-phospho	Binds plasminogen	*Streptococcus oralis*[101]	Positive
Enolase	Binds plasminogen and laminin	*Bacillus anthracis*[102]	Positive
GAPDH, Elongation factor Tu (EF-Tu) and triosephosphate isomerase (TPI)	Bind Caco-2 cells	*Lactobacillus plantarum*[103]	Positive
EF-Tu and GroEL	Adhere to mucin and human epithelial cells	*Lactobacillus Johnsonii*[59,65]	Positive
Hsp60, GroEL	Adheres to eukaryotic host cells, binds glycosphingolipids	*Haemophilus ducreyi*[66,67]	Negative
Hsp60, GroEL	Cell surface protein controlling bacterial growth	*Helicobacter pylori*[104]	Negative
Hsp60, GroEL	Cell adhesion	*Chlamydia pneumoniae*[68]	Negative
Hsp65, GroEL2	Cell surface protein that inhibits bacterial association with macrophages	*Mycobacterium tuberculosis*[105]	Acid-fast Gram-positive bacterium
Peroxiredoxin, DnaK and enolase	Bind plasminogen	*Neisseria meningitides*[64]	Negative
Ornithine carbamoyltransferase	Binds fibronectin	*Staphylococcus epidermidis*[106]	Positive
Glucosyltransferase	Adhesin to human endothelial cells	*Streptococcus gordonii*[107]	Positive
Malate synthase	Binds fibronectin and laminin	*Mycobacterium tuberculosis*[108]	Acid-fast Gram-positive bacterium
Glutamine synthetase	Binds laminin, fibronectin, collagen I, plasminogen	*Lactobacillus crispatus*[109]	Positive
TPI	Adhesion to fungal pathogen	*Staphylococcus aureus*[110]	Positive

Pyruvate kinase	Recognizes and binds to yeast mannan	*Lactococcus lactis*[70]	Positive

Table 2. **Surface Proteomics Studies Used in the Analysis**

Species	Gram type	Table Used	References
Bartonella quintana	Negative	Table 2	30
Clostridium acetobutylicum	Positive	Table S1	31
Streptococcus suis	Positive	Table S2 and S3	34
Escherichia coli	Negative	Table S2	32
Capnocytophaga canimorsus	Negative	Table 1	43
Clostridium thermocellum	Positive	Table S2	36
Enterococcus faecalis	Positive	Table 1	39
Pseudomonas aeruginosa	Negative	Table 1 and 2	40
Staphylococcus aureus	Positive	Table S1	49
Streptococcus pneumoniae	Positive	Table 3	35
Shewanella oneidensis	Gram-negative	Table S1	47
Neorickettsia sennetsu	Gram-negative	Table 1	45
Listeria monocytogenes	Gram-positive	Table S1	51
Listeria monocytogenes	Gram-positive	Table 1	38
Escherichia Coli	Gram-negative	Table S1	37
Rickettsia typhi	Gram-negative	Table S3	46
Rickettsia parkeri	Gram-negative	Table 2	44
Group A Streptococcus	Gram-positive	Table S1	33
Group B Streptococcus	Gram-positive	Table S1	41
Coxiella burnetii	Gram-negative	Table S2	48
Propionibacterium freudenreichii	Gram-positive	Table 1	50
Lactobacillus rhamnosus	Gram-positive	Table S2	42

Table 3. **Major functional groups of the intracellular/surface moonlighting protein candidates based on COG[24] database classifications**

Function description	Examples
Translation, ribosomal structure and biogenesis (J)	30S/50S ribosomal protein; EF-G, EF-Ts, EF-Tu, EF-P, IF; aminoacyl tRNA synthetases
Post-translational modification, protein turnover, chaperones (O)	Chaperonin GroEL; chaperone proteins DnaJ, DnaK; Lon/Clp family proteases
Carbohydrate transport and metabolism (G)	Aldolase, TPI, GAPDH, phosphoglycerate kinase, phosphoglycerate mutase (PGM), enolase, pyruvate kinase; transketolase
Amino acid transport and metabolism (E)	Acetylglutamate kinase, glutamine synthetase
Energy production and conversion (C)	Alcohol dehydrogenase, acetate kinase, isocitrate lyase
Transcription (K)	Transcription termination factor rho, transcription termination/antitermination protein NusA, DNA-directed RNA polymerase
Nucleotide transport and metabolism (F)	GMP synthase, adenylate kinase, Phosphoribosylformylglycinamidine synthase
Replication, recombination and repair (L)	DNA polymerase I/III, ATP-dependent RNA helicase, DNA gyrase
Lipid transport and metabolism (I)	Acetyl-CoA acetyltransferase, acetyl-CoA carboxylase

References

1 C. J. Jeffery, *Trends Genet.*, 1999, **24**, 8–11.

2 C. J. Jeffery, *Trends Genet.*, 2003, **19**, 415–417.

3 C. J. Jeffery, *Mol. Biosyst.*, 2009, **5**, 345–350.

4 M. Mani, C. Chen, V. Amblee, H. Liu, T. Mathur, G. Zwicke, S. Zabad, B. Patel, J. Thakkar and C. J. Jeffery, *Nucleic Acids Res.*, 2014, **43**, D277–D282.

5 S. Soker, S. Takashima, H. Q. Miao, G. Neufeld and M. Klagsbrun, *Cell*, 1998, **92**, 735–745.

6 W. Xu, K. Seiter, E. Feldman, T. Ahmed and J. W. Chiao, *Blood*, 1996, **87**, 4502–4506.

7 M. Chaput, V. Claes, D. Portetelle, I. Cludts, A. Cravador, A. Burny, H. Gras and A. Tartar, *Nature*, 1988, **332**, 454–455.

8 C. K. Suzuki, M. Rep, J. M. Van Dijl, K. Suda, L. A. Grivell and G. Schatz, *Trends Biochem. Sci.*, 1997, **22**, 118–123.

9 M. J. Stutts, C. M. Canessa, J. C. Olsen, M. Hamrick, J. A. Cohn, B. C. Rossier and R. C. Boucher, *Science (80-.).*, 1995, **269**, 847–850.

10 R. Lottenberg, C. C. Broder, M. D. Boyle, S. J. Kain, B. L. Schroeder and R. Curtiss, *J. Bacteriol.*, 1992, **174**, 5204–5210.

11 V. Pancholi and V. A. Fischetti, *J. Exp. Med.*, 1992, **176**, 415–426.

12 J. Brassard, M. Gottschalk and S. Quessy, *Vet. Microbiol.*, 2004, **102**, 87–94.

13 S. B. Winram and R. Lottenberg, *Microbiology*, 1996, **142**, 2311–2320.

14 B. Kenny and B. B. Finlay, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 7991–7995.

15 C. Castaldo, V. Vastano, R. A. Siciliano, M. Candela, M. Vici, L. Muscariello, R. Marasco and M. Sacco, *Microb. Cell Fact.*, 2009, **8**.

16 V. Pancholi and V. a Fischetti, *J. Biol. Chem.*, 1998, **273**, 14503–14515.

17 M. Candela, M. Centanni, J. Fiori, E. Biagi, S. Turroni, C. Orrico, S. Bergmann, S. Hammerschmidt and P. Brigidi, *Microbiology*, 2010, **156**, 1609–1618.

18 R. A. Garduño, E. Garduño, S. Paul, E. Gardun and P. S. Hoffman, *Infect. Immun.*, 1998, **66**, 4602–4610.

19 J. L. Wampler, J. L. Wampler, K. Kim, K. Kim, Z. Jaradat, Z. Jaradat, A. K. Bhunia and A. K. Bhunia, *Infect. Immun.*, 2009, **72**, 931–936.

20 M. Magrane and UniProt Consortium, *Database (Oxford)*, 2011, **2011**, bar009.

21 National Center for Biotechnology Information, .

22 A. Krogh, B. Larsson, G. von Heijne and E. L. Sonnhammer, *J. Mol. Biol.*, 2001, **305**, 567–580.

23 T. N. Petersen, S. Brunak, G. von Heijne and H. Nielsen, *Nat. Methods*, 2011, **8**, 785–786.

24 R. L. Tatusov, M. Y. Galperin, D. A. Natale and E. V Koonin, *Nucleic Acids Res.*, 2000, **28**, 33–36.

25 P. G. Bagos, T. D. Liakopoulos, I. C. Spyropoulos and S. J. Hamodrakas, *Nucleic Acids Res.*, 2004, **32**, W400–W404.

26 A. G. Murzin, S. E. Brenner, T. Hubbard and C. Chothia, *J. Mol. Biol.*, 1995, **247**, 536–540.

27 A. P. Pugsley, *Microbiol. Rev.*, 1993, **57**, 50–108.

28 H. Tokuda and S. I. Matsuyama, *Biochim. Biophys. Acta, Mol. Cell Res.*, 2004, **1693**, 5–13.

29 B. D. Corbin, Y. Wang, T. K. Beuria and W. Margolin, *J. Bacteriol.*, 2007, **189**, 3026–3035.

30 J. K. Boonjakuakul, H. L. Gerns, Y. T. Chen, L. D. Hicks, M. F. Minnick, S. E. Dixon, S. C. Hall and J. E. Koehler, *Infect. Immun.*, 2007, **75**, 2548–2561.

31 S. Mao, Y. Luo, G. Bao, Y. Zhang, Y. Li and Y. Ma, *Mol. Biosyst.*, 2011, **7**, 1660–1677.

32 M. Thein, G. Sauer, N. Paramasivam, I. Grin and D. Linke, *J. Proteome Res.*, 2010, **9**, 6135–6147.

33 M. J. Rodríguez-Ortega, N. Norais, G. Bensi, S. Liberatori, S. Capo, M. Mora, M. Scarselli, F. Doro, G. Ferrari, I. Garaguso, T. Maggi, A. Neumann, A. Covre, J. L. Telford and G. Grandi, *Nat. Biotechnol.*, 2006, **24**, 191–197.

34 L. Gómez-Gascón, I. Luque, A. Olaya-Abril, I. Jiménez-Munguía, R. A. Orbegozo-Medina, E. Peralbo, C. Tarradas and M. J. Rodríguez-Ortega, *J. Proteomics*, 2012, **75**, 5654–5666.

35 A. Olaya-Abril, L. Gómez-Gascón, I. Jiménez-Munguía, I. Obando and M. J. Rodríguez-Ortega, *J. Proteomics*, 2012, **75**, 3733–3746.

36 T. Yu, X. Xu, Y. Peng, Y. Luo and K. Yang, *Microbiol. Res.*, 2012, **167**, 364–371.

37 M. J. Han, S. Y. Lee and S. H. Hong, *J. Microbiol. Biotechnol.*, 2012, **22**, 470–478.

38 J. Schaumburg, O. Diekmann, P. Hagendorff, S. Bergmann, M. Rohde, S. Hammerschmidt, L. Jänsch, J. Wehland and U. Kärst, *Proteomics*, 2004, **4**, 2991–3006.

39 L. A. Bøhle, T. Riaz, W. Egge-Jacobsen, M. Skaugen, Ø. L. Busk, V. G. H. Eijsink and G. Mathiesen, *BMC Genomics*, 2011, **12**.

40 D. Vecchietti, D. Di Silvestre, M. Miriani, F. Bonomi, M. Marengo, A. Bragonzi, L.

Cova, E. Franceschi, P. Mauri and G. Bertoni, *PLoS One*, 2012, **7**, e51062.

41 F. Doro, S. Liberatori, M. J. Rodríguez-Ortega, C. D. Rinaudo, R. Rosini, M. Mora, M. Scarselli, E. Altindis, R. D'Aurizio, M. Stella, I. Margarit, D. Maione, J. L. Telford, N. Norais and G. Grandi, *Mol. Cell. Proteomics*, 2009, **8**, 1728–1737.

42 E. Espino, K. Koskenniemi, L. Mato-rodriguez, T. A. Nyman, J. Reunanen, J. Koponen, O. Tiina and P. Siljama, *J. Proteome Res.*, 2014, **14**, 1010–1024.

43 P. Manfredi, F. Renzi, M. Mally, L. Sauteur, M. Schmaler, S. Moes, P. Jenö and G. R. Cornelis, *Mol. Microbiol.*, 2011, **81**, 1050–1060.

44 W. Pornwiroon, A. Bourchookarn, C. D. Paddock and K. R. Macaluso, *Infect. Immun.*, 2009, **77**, 5262–5271.

45 K. Gibson, Y. Kumagai and Y. Rikihisa, *J. Bacteriol.*, 2010, **192**, 5898–5905.

46 K. T. Sears, S. M. Ceraul, J. J. Gillespie, E. D. Allen, V. L. Popov, N. C. Ammerman, M. S. Rahman and A. F. Azad, *PLoS Pathog.*, 2012, **8**, e1002856.

47 H. Zhang, R. N. Brown, W. Qian, M. E. Monroe, S. O. Purvine, R. J. Moore, M. A. Gritsenko, L. Shi, M. F. Romine, J. K. Fredrickson, R. D. Smith and M. S. Lipton, *J. Proteome Res.*, 2010, **9**, 2160–2169.

48 G. Flores-Ramirez, B. Jankovicova, Z. Bilkova, J. a Miernyk and L. Skultety, *Proteomics*, 2014, **14**, 1868–1881.

49 K. Hempel, F. A. Herbst, M. Moche, M. Hecker and D. Becher, *J. Proteome Res.*, 2011, **10**, 1657–1666.

50 C. Le Maréchal, V. Peton, C. Plé, C. Vroland, J. Jardin, V. Briard-Bion, G. Durant, V. Chuat, V. Loux, B. Foligné, S.-M. Deutsch, H. Falentin and G. Jan, *J. Proteomics*, 2014, **113**, 447–461.

51 F. G. Del Portillo, E. Calvo, V. D'Orazio and M. G. Pucciarelli, *J. Biol. Chem.*, 2011, **286**, 34675–34689.

52 W. Firshein, *Annu. Rev. Microbiol.*, 1989, **43**, 89–120.

53 K. Boeneman and E. Crooke, *Curr. Opin. Microbiol.*, 2005, **8**, 143–148.

54 A. Prinz, C. Behrens, T. A. Rapoport, E. Hartmann and K. U. Kalies, *EMBO J.*, 2000, **19**, 1900–1906.

55 E. Olmedo-Verd, J. Santamaría-Gómez, J. A. G. Ochoa De Alda, L. R. De Pouplana and I. Luque, *J. Biol. Chem.*, 2011, **286**, 41057–41068.

56 Y. Wang, Y. Jiang, M. Meyering-Voss, M. Sprinzl and P. B. Sigler, *Nat. Struct. Biol.*, 1997, **4**, 650–656.

57 R. K. Agrawal, P. Penczek, R. A. Grassucci and J. Frank, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 6134–6138.

58 H. Aoki, S. L. Adams, M. A. Turner and M. C. Ganoza, *Biochimie*, 1997, **79**, 7–11.

59 D. Granato, G. E. Bergonzelli, R. D. Pridmore, L. Marvin, M. Rouvet and E. Corthe, *Infect. Immun.*, 2004, **72**, 2160–2169.

60 J. Glenting, H. C. Beck, A. Vrang, H. Riemann, P. Ravn, A. M. Hansen, M. Antonsson, S. Ahrné, H. Israelsen and S. Madsen, *Microbiol. Res.*, 2013, **168**, 245–253.

61 S. F. Dallo, T. R. Kannan, M. W. Blaylock and J. B. Baseman, *Mol. Microbiol.*, 2002, **46**, 1041–1051.

62 A. Kunert, J. Losse, C. Gruszin, M. Hühn, K. Kaendler, S. Mikkat, D. Volke, R. Hoffmann, T. S. Jokiranta, H. Seeberger, U. Moellmann, J. Hellwage and P. F. Zipfel, *J. Immunol.*, 2007, **179**, 2979–2988.

63 T. Langer, C. Lu, H. Echols, J. Flanagan, M. K. Hayer and F. U. Hartl, *Nature*, 1992, **356**, 683–689.

64 A. Knaust, M. V. R. Weber, S. Hammerschmidt, S. Bergmann, M. Frosch and O. Kurzai, *J. Bacteriol.*, 2007, **189**, 3246–3255.

65 G. E. Bergonzelli, D. Granato, R. D. Pridmore, L. F. Marvin-Guy, D. Donnicola and I. E. Corthésy-Theulaz, *Infect. Immun.*, 2006, **74**, 425–434.

66 A. Frisk, C. A. Ison and T. Lagergård, *Infect. Immun.*, 1998, **66**, 1252–1257.

67 M. Pantzar, S. Teneberg and T. Lagergård, *Microbes Infect.*, 2006, **8**, 2452–2458.

68 F. N. Wuppermann, K. Mölleken, M. Julien, C. A. Jantos and J. H. Hegemann, *J. Bacteriol.*, 2008, **190**, 3757–3767.

69 B. Henderson and A. Martin, *Infect. Immun.*, 2011, **79**, 3476–3491.

70 Y. Katakura, R. Sano, T. Hashimoto, K. Ninomiya and S. Shioya, *Appl. Microbiol. Biotechnol.*, 2010, **86**, 319–326.

71 G. T. Robertson, M. E. Kovach, C. A. Allen, T. A. Ficht and M. Roop II, *Mol. Microbiol.*, 2000, **35**, 577–588.

72 C. Rouquette, C. De Chastellier, S. Nair and P. Berche, *Mol. Microbiol.*, 1998, **27**, 1235–1245.

73 H. Ohnishi, Y. Mizunoe, A. Takade, Y. Tanaka, H. Miyamoto, M. Harada and S. I. Yoshida, *Infect. Immun.*, 2004, **72**, 3592–3603.

74 R. Lill, E. Crooke, B. Guthrie and W. Wickner, *Cell*, 1988, **54**, 1013–1018.

75 A. Hoffmann, B. Bukau and G. Kramer, *Biochim. Biophys. Acta*, 2010, **1803**, 650–661.

76 V. Amblee and C. J. Jeffery, *PLoS One*, 2015, **10**, e0130575.

77 M. Y. Galperin, K. S. Makarova, Y. I. Wolf and E. V Koonin, *Nucleic Acids Res.*, 2015, **43**, D261–D269.

78 B. Jagadeesan, O. K. Koo, K.-P. Kim, K. M. Burkholder, K. K. Mishra, A. Aroonnual and A. K. Bhunia, *Microbiology*, 2010, **156**, 2782–2795.

79 J. D. Bendtsen, L. Kiemer, A. Fausbøll and S. Brunak, *BMC Microbiol.*, 2005, **5**.

80 G. Wang, H. Chen, Y. Xia, J. Cui, Z. Gu, Y. Song, Y. Q. Chen, H. Zhang and W. Chen, *Curr. Microbiol.*, 2013, **67**, 688–695.

81 W. Nickel, *Eur. J. Biochem.*, 2003, **270**, 2109–2119.

82 M. J. Pallen, *Trends Microbiol.*, 2002, **10**, 209–212.

83 M. Braunstein, B. J. Espinosa, J. Chan, J. T. Belisle, W. R. Jacobs and W. R. Jacobs, *Mol. Microbiol.*, 2003, **48**, 453–464.

84 N. W. Rigel and M. Braunstein, *Mol. Microbiol.*, 2008, **69**, 291–302.

85 H. S. Gibbons, F. Wolschendorf, M. Abshire, M. Niederweis and M. Braunstein, *J. Bacteriol.*, 2007, **189**, 5090–5100.

86 L. L. Lenz, S. Mohammadi, A. Geissler and D. A. Portnoy, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 12432–12437.

87 M. Kostakioti, C. L. Newman, D. G. Thanassi and C. Stathopoulos, *J. Bacteriol.*, 2005, **187**, 4306–4314.

88 K. Dilks, R. W. Rose, E. Hartmann and M. Pohlschröder, *J. Bacteriol.*, 2003, **185**, 1478–1483.

89 L. Aguilera, E. Ferreira, R. Giménez, F. J. Fernández, M. Taulés, J. Aguilar, M. C. Vega, J. Badia and L. Baldomà, *Int. J. Biochem. Cell Biol.*, 2012, **44**, 955–962.

90 L. Egea, L. Aguilera, R. Giménez, M. A. Sorolla, J. Aguilar, J. Badía and L. Baldoma, *Int. J. Biochem. Cell Biol.*, 2007, **39**, 1190–1203.

91 G. Samoilis, M. Aivaliotis, I. Vranakis, A. Papadioti, Y. Tselentis, G. Tsiotis and A. Psaroulaki, *J. Proteome Res.*, 2010, **9**, 1619–1626.

92 H. Y. Wu, P. C. Chung, H. W. Shih, S. R. Wen and E. M. Lai, *J. Bacteriol.*, 2008, **190**, 2841–2850.

93 G. S. Chhatwal, *Trends Microbiol.*, 2002, **10**, 205–208.

94 O. Schneewind and D. M. Missiakas, *Philos. Trans. R. Soc., B*, 2012, **367**, 1123–1139.

95 S. K. Mazmanian, E. P. Skaar, A. H. Gaspar, M. Humayun, P. Gornicki, J. Jelenska, A. Joachmiak, D. M. Missiakas and O. Schneewind, *Science (80-.).*, 2003, **299**, 906–909.

96 A. W. Maresso, T. J. Chapa and O. Schneewind, *J. Bacteriol.*, 2006, **188**, 8145–8152.

97 J. Yother and J. M. White, *J. Bacteriol.*, 1994, **176**, 2976–2985.

98 J. R. Scott and T. C. Barnett, *Annu. Rev. Microbiol.*, 2006, **60**, 397–423.

99 M. Desvaux, E. Dumas, I. Chafsey and M. Hebraud, *FEMS Microbiol. Lett.*, 2006, **256**, 1–15.

100 S. Bergmann, M. Rohde, S. Gursharan and S. Hammerschmidt, *Mol. Microbiol.*, 2001, **40**, 1273–1287.

101 J. Welin, J. C. Wilkins, D. Beighton and G. Svensater, *Appl. Environ. Microbiol.*, 2004, **70**, 3736–3741.

102 S. Agarwal, P. Kulshreshtha, D. Bambah Mukku and R. Bhatnagar, *Biochim. Biophys. Acta, Proteins Proteomics*, 2008, **1784**, 986–994.

103 K. Ramiah, C. a. van Reenen and L. M. T. Dicks, *Res. Microbiol.*, 2008, **159**, 470–475.

104 H. Yamaguchi, T. Osaki, H. Taguchi, T. Hanawa, T. Yamamoto, M. Fukuda, H. Kawakami, H. Hirano and S. Kamiya, *Microbiol. Immunol.*, 1997, **41**, 909–916.

105 T. B. M. Hickey, L. M. Thorson, D. P. Speert, M. Daffé and R. W. Stokes, *Infect. Immun.*, 2009, **77**, 3389–3401.

106 M. Hussain, G. Peters, G. S. Chhatwal and M. Herrmann, *Infect. Immun.*, 1999, **67**, 6688–6690.

107 A. M. Vacca-Smith, C. A. Jones, M. J. Levine and M. W. Stinson, *Infect. Immun.*, 1994, **62**, 2187–2194.

108 A. G. Kinhikar, D. Vargas, H. Li, S. B. Mahaffey, L. Hinds, J. T. Belisle and S. Laal, *Mol. Microbiol.*, 2006, **60**, 999–1013.

109 V. Kainulainen, V. Loimaranta, A. Pekkala, S. Edelman, J. Antikainen, R. Kylväjä, M. Laaksonen, L. Laakkonen, J. Finne and T. K. Korhonen, *J. Bacteriol.*, 2012, **194**, 2509–2519.

110 H. Furuya and R. Ikeda, *Microbiology*, 2009, **155**, 2707–2713.