Chapter 4

Cell Shaving and False-Positive Control Strategies Coupled to Novel Statistical Tools to Profile Gram-Positive Bacterial Surface Proteomes

Nestor Solis and Stuart J. Cordwell

Abstract

A powerful start to the discovery and design of novel vaccines, and for better understanding of host-pathogen interactions, is to profile bacterial surfaces using the proteolytic digestion of surface-exposed proteins under mild conditions. This "cell shaving" approach has the benefit of both identifying surface proteins and their surface-exposed epitopes, which are those most likely to interact with host cells and/or the immune system, providing a comprehensive overview of bacterial cell topography. An essential requirement for successful cell shaving is to account for (or minimize) cellular lysis that can occur during the shaving procedure and thus generate data that is biased towards non-surface (e.g., cytoplasmic) proteins. This is further complicated by the presence of "moonlighting" proteins, which are proteins predicted to be intracellular but with validated surface or extracellular functions. Here, we describe an optimized cell shaving protocol for Gram-positive bacteria that uses proteolytic digestion and a "false-positive" control to reduce the number of intracellular contaminants in these datasets. Released surface-exposed peptides are analyzed by liquid chromatography (LC) coupled to high-resolution tandem mass spectrometry (MS/MS). Additionally, the probabilities of proteins being surface exposed can be further calculated by applying novel statistical tools.

Key words Cell shaving, Gram-positive bacteria, Mass spectrometry, Surface proteomics, Surfaceome

Abbreviations

CID	Collision-induced dissociation
DTT	Dithiothreitol
ESI	Electrospray ionization
IAA	Iodoacetamide
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
MeCN	Acetonitrile

Hee-Jeon Hong (ed.), *Bacterial Cell Wall Homeostasis: Methods and Protocols*, Methods in Molecular Biology, vol. 1440,
DOI 10.1007/978-1-4939-3676-2_4, © Springer Science+Business Media New York 2016

1 Introduction

1.1 Bacterial Surface Topography

Interactions between an organism and its environment are initially mediated by the interplay of surface-exposed proteins and other macromolecular structures on the exterior face of the cell. As a major example, it is the ability of a pathogen to recognize its environment and respond to it that enables colonization and infection in the host. Interactions between bacterial and host surfaces and extracellular matrix further facilitate pathogenesis. As such, surface-exposed proteins are critical for understanding adherence, colonization, and disease progression caused by bacterial pathogens. Furthermore, surface structures including lipopolysaccharide, capsule, and proteins are the first markers recognized by the human immune response and thus knowledge of the topography of bacterial cells is crucial for vaccine design.

The cell surface in Gram-positive bacteria consists of a thick peptidoglycan wall and an inner cytoplasmic membrane. There are four major groups of cell wall/envelope proteins, including those anchored to the cytoplasmic membrane by hydrophobic domains, lipoproteins, cell wall proteins anchored by sortase via an LPXTG signal, and non-covalently cell wall-associated proteins [1–4]. Some of these proteins may remain buried within the envelope and are thus not truly "surface exposed." Membrane-embedded proteins are characterized by the Ala-X-Ala *N*-terminal signal peptidase I recognition sequence, while lipoproteins are covalently anchored to phospholipid and contain the signal peptidase II recognition sequence, Leu-Ala-Ala-Cys. Cell wall-anchored proteins contain an *N*-terminal *Sec* signal and a *C*-terminal LPXTG motif. Such proteins are retained in the membrane by a hydrophobic *C*-terminal domain and cleaved by sortase.

1.2 Cell Shaving Proteomics

Identification of those proteins representing the true "surfaceome" of an organism represents a rich reservoir of information that can be utilized in the production of novel therapeutics and vaccines, based on either individual proteins or multiple combined peptide epitopes. Substantial technical progress has been made in subcellular proteomics of bacterial pathogens, with several studies describing methods for enriching outer membrane, periplasmic, and secreted proteins. These studies however do not provide specific assignment of surface-exposed proteins, nor those peptide epitopes located outside the cell. Due to their low abundance and hydrophobic transmembrane regions, surface proteins are generally considered very difficult to enrich from among complex protein mixtures. Several methods for analysis of membrane-associated proteins have been proposed, including (1) surface labeling by biotinylation and capture through streptavidin affinity; (2) precipitation, density gradient ultracentrifugation, and detergent extraction; and (3) detergent-phase partitioning. Such methods are useful for enrichment of membrane-associated

proteins prior to separation and analysis using gel-based and gel-free approaches; however false positives may occur in these protocols due to cell lysis or residual biotin resulting in labeling of cytoplasmic contaminants. Additionally, specific analysis of surface proteins and their surface-exposed epitopes has remained challenging.

A novel method for better understanding bacterial surface protein topology involves cell "shaving," where a proteolytic enzyme is incubated with whole cells to release their exposed peptide epitopes while maintaining cell integrity (Fig. 1) [5–12]. This approach, combined with the resolution of LC-MS/MS, provides a peptide repertoire of surface-exposed epitopes belonging to surface and membrane-associated proteins. The method provides a simple and fast route for the gentle digestion of entire cells, purification of released peptides, and proteomic identification. A false-positive control strategy [9] can also be employed to better control for cell lysis and the release of intracellular proteins. Furthermore, using novel statistical tools and bioinformatic predictions, a probability can be calculated for the likelihood of any identified protein being surface exposed [11]. A final high-confidence list of proteins can then be functionally validated.

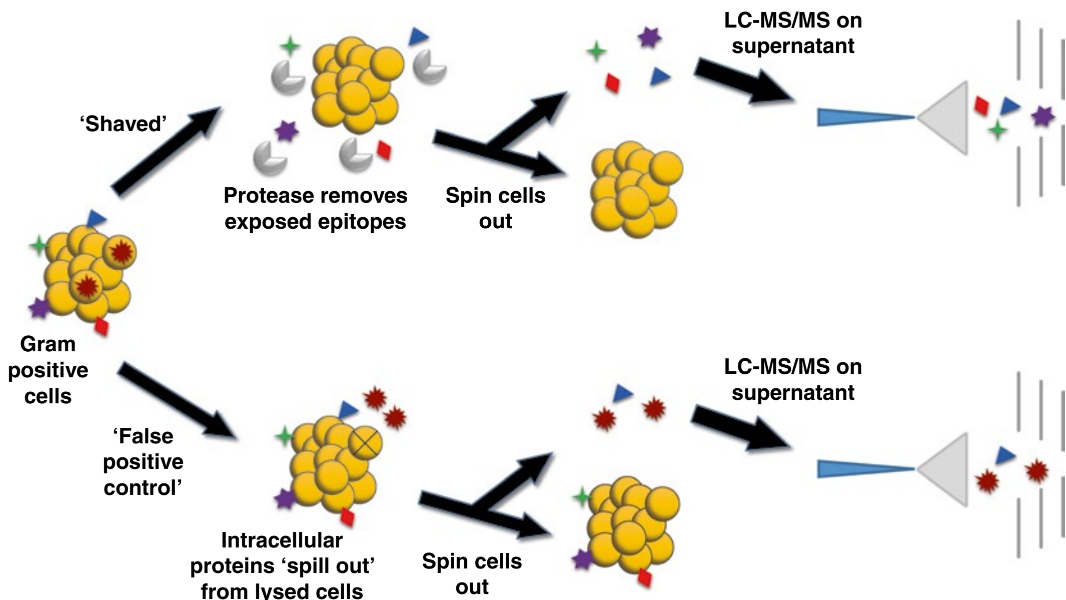

Fig. 1 Cell shaving and false-positive control strategy for Gram-positive cell surfaceomics. (*Upper*) Cell shaving; whole cell-shaved fractions are generated by high concentration, short-duration incubation with a protease (generally trypsin) in isotonic buffer. Released peptides are collected and analyzed by LC-MS/MS. (*Lower*) False-positive control; false-positive control is used to identify proteins released by cell lysis. Cells are incubated as for the shaving protocol but no protease is included. Whole cells are removed by centrifugation and the supernatants then digested with a protease. Any identified peptides are present as a result of lysis and are analyzed by LC-MS/MS

2 Materials

2.1 Growth of Microorganisms

1. Frozen stock of pure culture of Gram-positive organism to be analyzed.
2. Agar plates (12%) with media of choice (e.g., Luria-Bertani Broth, Tryptic Soya Broth).
3. Liquid media of choice (e.g., Luria-Bertani Broth, Tryptic Soya Broth) with supplements as required.

2.2 Sample Preparation for Cell Shaving

1. Conical bottom sterile 50 mL volume Falcon tubes.
2. Wash buffer (150 mM NaCl, 20 mM Tris–HCl, pH 7.5).
3. Digestion buffer (150 mM NaCl, 20 mM Tris–HCl, 1 M D-arabinose, 10 mM $CaCl_2$, pH 7.5) (*see* **Note 1**).
4. Sequencing-grade trypsin (vial of 20 μg).
5. Formic acid (HPLC grade).
6. 2 mL Tubes with 1 kDa dialysis membrane cutoffs (e.g., Mini-Dialysis Kit with 1 kDa cutoff, GE Life Sciences, Uppsala, Sweden).
7. 0.22 μm Filters (suitable for use with a handheld 2 mL syringe).
8. 4 L Buckets of cold distilled water for dialysis.
9. Vacuum centrifuge.
10. Dithiothreitol (DTT; 1 M stock).
11. Iodoacetamide (IAA; 0.5 M stock).
12. Acetonitrile (MeCN; 100% stock).
13. C_{18} material for peptide purification (e.g., POROS R2 resin) using home-packed columns.

2.3 Mass Spectrometry and Data Analysis

1. LC-MS/MS system capable of high-speed, high-sensitivity data-dependent acquisition.
2. Nanoflow HPLC system (e.g., Agilent 1100/1200 series or Thermo Scientific EasyLC system).
3. Full genome sequence of organism converted to translated proteome (FASTA format).
4. Database search engine (e.g., MASCOT).

3 Methods

3.1 Cell Shaving and False-Positive Control

1. Inoculate from a stock onto an agar plate to generate a pure culture of the Gram-positive organism to be examined. Grow at the desired temperature until colonies are visible.
2. Inoculate a single colony into the desired broth and grow until $OD_{600} > 1$.

3. For each experiment (shaved and false-positive control), aliquot 200 μL of turbid culture into 19.8 mL of fresh media (1:100 dilution) in a conical shaped sterile tube (50 mL Falcon tube) and grow under desired conditions until mid-log phase. Precool a swing-bucket rotor to 4 °C (*see* **Note 2**).
4. Following growth to mid-log phase, place each tube on ice for 5 min and then centrifuge in a precooled swing-bucket rotor at 1000 × *g* for 15 min at 4 °C.
5. Carefully decant supernatant into waste.
6. Resuspend the cell pellet with ice-cold wash buffer (*see* **Note 3**).
7. Centrifuge in the precooled swing-bucket rotor centrifuge at 1000 × *g* for 15 min at 4 °C and then carefully decant the supernatant into waste.
8. Repeat **steps 6** and **7** another two times, for a total of three washes (*see* **Note 4**). During spin steps prepare fresh 4 mL of digestion buffer. Resuspend 10 μg of sequencing-grade trypsin with 10 μL digestion buffer immediately before the next step and keep on ice.
9. Carefully resuspend the cell pellets for the control and shaved experiments in a total of 4 mL digestion buffer. Slowly invert and keep a homogenous mixture (*see* **Note 3**).
10. Split 2 mL each into two separate large low-protein-binding microfuge (2 mL sized) tubes.
11. To one of the tubes add the ice-cold trypsin—this will be the cell-shaved fraction.
12. Place both tubes on a rotator and spin slowly in a 37 °C controlled room or incubator for 15 min (*see* **Notes 5** and **6**).
13. Immediately after digestion, place the tubes on ice and centrifuge at 1000 × *g* for 10 min at 4 °C. During this centrifugation, prepare new 2 mL low-protein-binding tubes labeled "shaved" and "control," as well as two dialysis filter membranes (1 kDa cutoff) by gently rinsing in water.
14. Remove supernatants carefully by pipetting into fresh microfuge tubes (*see* **Note 7**).
15. Pipette each fraction into separate 2 mL syringes each with a 0.22 μm filter at the end. Pass the solution through the filter directly into separate pre-washed dialysis tubes.
16. Screw the dialysis membranes onto the tubes and dialyze in 4 L of water at 4 °C for 3 h and then replace the water with 4 L fresh cold water and dialyze overnight. Next morning, replace the water one more time for a further 3-h dialysis.
17. Recover samples from dialysis tubes into 2 mL microfuge tubes and concentrate by vacuum centrifugation to 100 μL.

18. Reduce samples with DTT to a final concentration of 10 mM for 1 h at 37 °C and then alkylate with IAA to a final 15 mM concentration at room temperature in the dark. Quench with additional DTT to a final 20 mM.
19. To the false-positive "control" fraction add 1 μg trypsin in digestion buffer and digest overnight at 37 °C.
20. Acidify both samples to a final 0.1 % formic acid.
21. Activate a C_{18} micro-column in 70 % MeCN and 0.1 % formic acid. Equilibrate the column twice with 50 μL 0.1 % formic acid and load with 50 μL sample. Wash twice with 50 μL 0.1 % formic acid and elute peptides with 50 μL 70 % MeCN and 0.1 % formic acid.
22. Lyophilize purified peptides to complete dryness and store at −20 °C until required for mass spectrometric analysis.

3.2 LC-MS/MS (See Note 8)

1. Purified peptide supernatants are separated by reversed-phase nanoflow LC (e.g., using an EASY-nLC [Thermo Scientific, San Jose CA]).
2. Peptides are resolved using a one column reversed-phase (3 μm particle size, 50 cm × 50 μm inner diameter [I.D.], C_{18}) setup over a linear gradient of 0–40 % buffer B (80 % MeCN, 0.1 % formic acid) at 250 nL/min over 103 min (*see* **Note 9**).
3. Peptides are eluted into the mass spectrometer via electrospray ionization (ESI).
4. Operate the mass spectrometer in data-dependent acquisition mode, which automatically switches between MS and MS/MS. Depending on the mass spectrometer, for each MS scan, the 3–30 most intense peptide ions are automatically selected for fragmentation by collision-induced dissociation (CID) (*see* **Note 10**).

3.3 Data Analysis (See Notes 11 and 12)

1. Search raw MS files in a database search engine of choice. RAW files generated by an LTQ Orbitrap XL are searched in the Proteome Discoverer environment using SEQUEST with an MS1 tolerance of 10 ppm and an MS2 tolerance of 0.8 Da. Allow for four missed cleavages as discussed in [7]. Variable modifications should include oxidation of methionine and carbamidomethylation of cysteines. Semi-tryptic protease specificity can also be employed to maximize coverage of surface-exposed peptides (*see* **Note 13**).
2. Determine the predicted localization of all protein hits identified by database searching. This can be done with a variety of tools such as PSORTb [13], SurfG+ [14], or LocateP [15] (Fig. 2) (*see* **Note 14**).

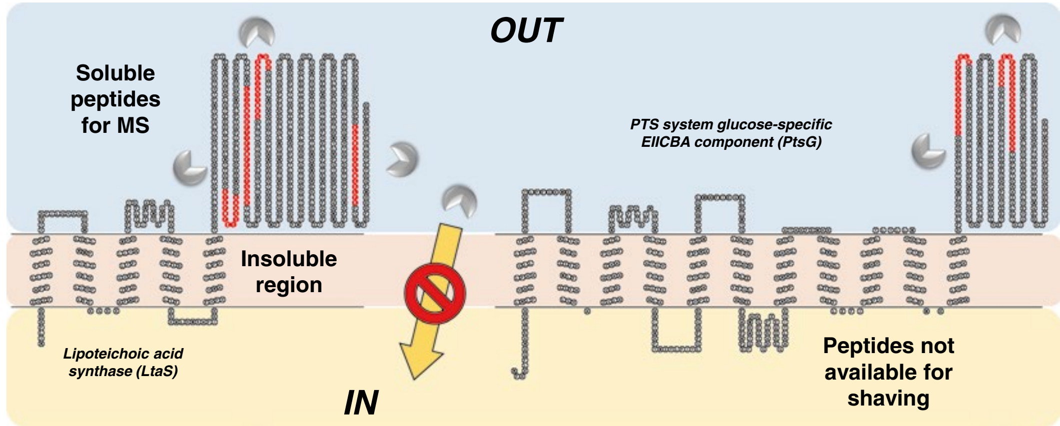

Fig. 2 Cell shaving data and surface-exposed peptide topology prediction (SurfG+) from 2 *Staphylococcus aureus* proteins. (*Left*) Lipoteichoic acid synthase (LtaS). (*Right*) PTS system glucose-specific EIICBA component (PtsG). Amino acid sequences shown in red were identified by cell shaving of *S. aureus* cells. Transmembrane and internal amino acid sequences are intractable to shaving, as they are not exposed

3. Once protein and peptide lists have been compiled for shaved and false-positive control fractions, a statistical methodology can be employed to determine the likelihood of a protein being surface exposed based on the number of peptides identified in the shaved and false-positive control fraction [11] using this equation:

$$p_{\text{experimental}} = 1 - \sum_{k=\frac{m}{2}}^{m} \frac{\binom{n_c}{k}\binom{n_s}{m-k}}{\binom{n_T}{m}}$$

where n_c = number of control peptides, n_s = number of shaved peptides, $n_T = n_c + n_s$, and $m = 0.4 \times n_T$ (to closest higher integer):

$$p_{\text{adjusted}} = \frac{(p_{\text{experimental}})(p_{\text{predicted}})}{(p_{\text{experimental}})(p_{\text{predicted}}) + (1 - p_{\text{experimental}})(1 - p_{\text{predicted}})}$$

4. Protein lists with number of peptides per identified protein should be compiled and the probabilities of each protein being surface exposed calculated from **step 2**. These can be used as an input to calculate (a) experimental probability of the protein being surface exposed and (b) the adjusted probability (accounting for the predictions made in **step 2** and based on sequence predictions) of the protein being surface exposed. These can be run directly on https://github.com/mehwoot/cellshaving to provide a final score for each protein.

4 Notes

1. Arabinose should be added fresh before use.
2. Mid-log phase typically contains cells that are most robust and less prone to lysis due to active cellular division. The conical shape of the Falcon tube is important for pelleting cells at low centrifugal speeds. At this step, one tube of 20 mL culture will be for cell shaving and one tube of 20 mL culture will be for the false-positive control.
3. Do not pipette the buffer directly onto the cell pellet—pipette against the inside of the tube gently to avoid cell lysis. Do not vortex. Gentle inversions to wash the pellet are appropriate.
4. Additional washes ensure removal of media components and any loosely bound, non-surface-specific proteins on the exterior of the cell surface.
5. For optimizing incubation periods, obtain 10 μL aliquots every 5 min for each treatment and perform cell counts using a hemocytometer under a phase-contrast microscope. This will give an estimate of the number of intact cells.
6. Digestion times are preferably short to minimize cell lysis and as such require higher amounts of trypsin to achieve proteolysis. However, lower amounts of trypsin (2–5 μg) may be required for cells with less rigid cell walls.
7. It is best to leave a small volume close to the cell pellet for higher purity. At this stage, it is also optional to acidify with formic acid to a final 0.1 % to stop any proteolytic digestion.
8. Cell shaving protocols are reliant on comprehensive peptide coverage in the relatively simple fractions generated by tryptic digestion of cell surfaces. Therefore, access to a high-speed, high-sensitivity mass spectrometer is essential for proper determination of peptide sequences representing surface-exposed proteins.
9. Any typical nanoflow reversed-phase LC setup will be compatible with these peptide analyses.
10. Settings on the mass spectrometer should be optimized and will be instrument dependent.
11. Raw MS data being analyzed for cell shaving experiments are best searched against the strain of the organism in question to overcome issues associated with point mutations or posttranslational modifications.
12. Data analysis in many cases depends on the acquisition mass spectrometer and proprietary software. Here, we describe an analysis workflow used for cell shaving data acquired on an LTQ Orbitrap XL and the use of Proteome Discoverer (Thermo Scientific) software.

13. It is important to consider that the identified proteins may not necessarily generate a large sequence coverage (as typically found in many proteins during shotgun/bottom-up proteomics experiments), since only their surface-exposed regions should be cleaved and these may not have amino acid sequences flanked by lysine and arginine residues suitable for tryptic cleavage and MS analysis. Hence, one peptide per protein is the minimum requirement for identification and we suggest manual verification of these "one-hit wonders" to ensure correct sequence assignation.
14. It is important to either keep these localizations as a reference for later validation or preferably have a probability score associated with their surface localization that can be used in the next step.

References

1. Schneewind O, Missiakas D (2014) Sec-secretion and sortase-mediated anchoring of proteins in Gram-positive bacteria. Biochim Biophys Acta 1843:1687–1697
2. Buist G, Steen A, Kok J, Kuipers OP (2008) LysM, a widely distributed protein motif for binding to (peptido)glycans. Mol Microbiol 68:838–847
3. Scott JR, Barnett TC (2006) Surface proteins of gram-positive bacteria and how they get there. Annu Rev Microbiol 60:397–423
4. Navarre W, Schneewind O (1999) Surface proteins of gram positive bacteria and mechanisms of their targeting to the cell wall envelope. Microbiol Mol Biol Rev 63: 174–229
5. Dreisbach A, van der Kooi-Pol MM, Otto A, Gronau K, Bonarius HP, Westra H, Groen H, Becher D, Hecker M, van Dijl JM (2011) Surface shaving as a versatile tool to profile global interactions between human serum proteins and the *Staphylococcus aureus* cell surface. Proteomics 11:2921–2930
6. Olaya-Abril A, Jimenez-Munguia I, Gomez-Gascon L, Rodriguez-Ortega MJ (2013) Surfomics: shaving live organisms for a fast proteomic identification of surface proteins. J Proteomics 97:164–176
7. Rodriguez-Ortega MJ, Norais N, Bensi G, Liberatori S, Capo S, Mora M, Scarselli M, Doro F, Ferrari G, Garaguso I, Maggi T, Neumann A, Covre A, Telford JL, Grandi G (2006) Characterization and identification of vaccine candidate proteins through analysis of the group A *Streptococcus* surface proteome. Nat Biotechnol 24:191–197
8. Solis N, Cordwell SJ (2011) Current methodologies for proteomics of bacterial surface-exposed and cell envelope proteins. Proteomics 11:3169–3189
9. Solis N, Larsen MR, Cordwell SJ (2010) Improved accuracy of cell surface shaving proteomics in *Staphylococcus aureus* using a false-positive control. Proteomics 10:2037–2049
10. Tjalsma H, Lambooy L, Hermans PW, Swinkels DW (2008) Shedding & shaving: disclosure of proteomic expressions on a bacterial face. Proteomics 8:1415–1428
11. Solis N, Parker BL, Kwong SM, Robinson G, Firth N, Cordwell SJ (2014) *Staphylococcus aureus* surface proteins involved in adaptation to oxacillin identified using a novel cell shaving approach. J Proteome Res 13:2954–2972
12. Severin A, Nickbarg E, Wooters J, Quazi SA, Matsuka YV, Murphy E, Moutsatsos IK, Zagursky RJ, Olmsted SB (2007) Proteomic analysis and identification of *Streptococcus pyogenes* surface-associated proteins. J Bacteriol 189:1514–1522
13. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FS (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26:1608–1615
14. Barinov A, Loux V, Hammani A, Nicolas P, Langella P, Ehrlich D, Maguin E, van de Guchte M (2009) Prediction of surface exposed proteins in *Streptococcus pyogenes*, with a potential application to other Gram-positive bacteria. Proteomics 9:61–73
15. Zhou M, Boekhorst J, Francke C, Siezen RJ (2008) LocateP: genome-scale subcellular-location predictor for bacterial proteins. BMC Bioinformatics 9:173