Journal of Visualized Experiments

Identification of antibacterial immunity proteins in Escherichia coli using MALDI-TOF-TOF-MS/MS and top-down proteomic analysis

--Manuscript Draft--

Article Type:	Invited Methods Article - JoVE Produced Video
Manuscript Number:	JoVE62577R2
Full Title:	Identification of antibacterial immunity proteins in Escherichia coli using MALDI-TOF-TOF-MS/MS and top-down proteomic analysis
Corresponding Author:	Clifton Fagerquist USDA-ARS WRRC: USDA-ARS Western Regional Research Center Albany, California UNITED STATES
Corresponding Author's Institution:	USDA-ARS WRRC: USDA-ARS Western Regional Research Center
Corresponding Author E-Mail:	Clifton.Fagerquist@ars.usda.gov
Order of Authors:	Clifton K Fagerquist
	Ernesto Rojas
Additional Information:	
Question	**Response**
Please indicate whether this article will be Standard Access or Open Access.	Standard Access (US$2,400)
Please specify the section of the submitted manuscript.	Chemistry
Please indicate the **city, state/province, and country** where this article will be **filmed**. Please do not use abbreviations.	Albany, California, United States of America
Please confirm that you have read and agree to the terms and conditions of the author license agreement that applies below:	I agree to the Author License Agreement
Please provide any comments to the journal here.	
Please indicate whether this article will be Standard Access or Open Access.	Standard Access ($1400)

TITLE:

Identification of Antibacterial Immunity Proteins in *Escherichia coli* using MALDI-TOF-TOF-MS/MS and Top-Down Proteomic Analysis

AUTHORS AND AFFILIATIONS:

Clifton K. Fagerquist, Ernesto Rojas

Produce Safety & Microbiology, Western Regional Research Center, Agricultural Research Service, U.S. Department of Agriculture, 800 Buchanan Street, Albany, CA.

E-mail Addresses of Co-authors:
Clifton K. Fagerquist (clifton.fagerquist@usda.gov)
Ernesto Rojas (ernesto_rojas@berkeley.edu)

Corresponding Author
Clifton K. Fagerquist (clifton.fagerquist@usda.gov)

SUMMARY:

Here we present a protocol for the rapid identification of proteins produced by genomically sequenced pathogenic bacteria using MALDI-TOF-TOF tandem mass spectrometry and top-down proteomic analysis with software developed in-house. Metastable protein ions fragment because of the aspartic acid effect and this specificity is exploited for protein identification.

ABSTRACT:

This protocol identifies the immunity proteins of the bactericidal enzymes: colicin E3 and bacteriocin, produced by a pathogenic *Escherichia coli* strain using antibiotic induction, and identified by MALDI-TOF-TOF tandem mass spectrometry and top-down proteomic analysis with software developed in-house. The immunity protein of colicin E3 (Im3) and the immunity protein of bacteriocin (Im-Bac) were identified from prominent b- and/or y-type fragment ions generated by the polypeptide backbone cleavage (PBC) on the C-terminal side of aspartic acid, glutamic acid, and asparagine residues by the aspartic acid effect fragmentation mechanism. The software rapidly scans *in silico* protein sequences derived from the whole genome sequencing of the bacterial strain. The software iteratively removes amino acid residues from the N-terminus only if the mature protein sequence can be truncated at the N- or C-termini (or both). A single protein sequence possessed mass and fragment ions consistent with those detected for each immunity protein. The candidate sequence was then manually inspected to confirm that all detected fragment ions could be assigned. The N-terminal methionine of Im3 was post-translationally removed, whereas Im-Bac had the complete sequence. In addition, we found that only two or three non-complementary fragment ions formed by PBC are necessary to identify the correct protein sequence. Finally, a promoter (SOS box) was identified upstream of the antibacterial and immunity genes in a plasmid genome of the bacterial strain.

INTRODUCTION:

Analysis and identification of undigested proteins by mass spectrometry is referred to as the top-down proteomic analysis[1–4]. It is now an established technique that utilizes electrospray ionization (ESI)[5] and high-resolution mass analyzers[6], and sophisticated dissociation techniques, e.g., electron transfer dissociation (ETD), electron capture dissociation (ECD)[7], ultraviolet photo-dissociation (UV-PD)[8], etc.

The other soft ionization technique is matrix-assisted laser desorption/ionization (MALDI)[9–11] that has been less extensively utilized for the top-down analysis, in part because it is primarily coupled to time-of-flight (TOF) mass analyzers, which have limited resolution compared to other mass analyzers. Despite these limitations, MALDI-TOF and MALDI-TOF-TOF instruments have been exploited for the rapid top-down analysis of pure proteins and fractionated and unfractionated mixtures of proteins. For the identification of pure proteins, in-source decay (ISD) is a particularly useful technique because it allows mass spectrometry (MS) analysis of ISD fragment ions, as well as tandem mass spectrometry (MS/MS) of protein ion fragments providing sequence-specific fragment often from the N- and C-termini of the target protein, analogous to Edman sequencing[12,13]. A drawback to the ISD approach is that, as in Edman sequencing, the sample must contain only one protein. The one protein requirement is due to the need for unambiguous attribution of fragment ions to a precursor ion. If two or more proteins are present in a sample, it may be difficult to assign which fragment ions belong to precursor ions.

Fragment ion/precursor ion attribution can be addressed using MALDI-TOF-TOF-MS/MS. As with any classical MS/MS experiment, precursor ions are mass-selected/isolated prior to fragmentation, and the fragment ions detected can be attributed to a specific precursor ion. However, the dissociation techniques available for this approach are restricted to primarily high energy collision-induced dissociation (HE-CID)[14] or post-source decay (PSD)[15,16]. HE-CID and PSD are most effective at fragmenting peptides and small proteins, and the sequence coverage can, in some cases, be limited. In addition, PSD results in polypeptide backbone cleavage (PBC) primarily on the C-terminal side of aspartic and glutamic acid residues by a phenomenon called the aspartic acid effect[17–20].

MALDI-TOF-MS has also found a niche application in the taxonomic identification of microorganisms: bacteria[21], fungi[22], and viruses[23]. For example, MS spectra are used to identify unknown bacteria by comparison to a reference library of MS spectra of known bacteria using pattern recognition algorithms for comparison. This approach has proved highly successful because of its speed and simplicity, although requiring an overnight culturing of the isolate. The protein ions detected by this approach (usually under 20 kDa) comprise a MS fingerprint allowing taxonomic resolution at the genus and species level and in some cases at the sub-species[24] and strain level[25,26]. However, there remains a need to not only taxonomically classify potentially pathogenic microorganisms but also identify specific virulence factors, toxins, and antimicrobial resistance (AMR) factors. To accomplish this, the mass of peptides, proteins, or small molecules are measured by MS and subsequently isolated and fragmented by MS/MS.

Pathogenic bacteria often carry circular pieces of DNA called plasmids. Plasmids, along with prophages, are a major vector of horizontal gene transfer between bacteria and are responsible

for the rapid spread of antimicrobial resistance and other virulence factors across bacteria. Plasmids may also carry antibacterial (AB) genes, e.g., colicin and bacteriocin. When expressed and their secreted proteins act to disable the protein translation machinery of neighboring bacteria occupying the same environmental niche[27]. However, these bactericidal enzymes can also pose a risk to the host that produced them. In consequence, a gene is co-expressed by the host that specifically inhibits the function of an AB enzyme and is referred to as its immunity protein (Im).

DNA-damaging antibiotics such as mitomycin-C and ciprofloxacin are often used to induce the SOS response in Shiga toxin-producing *E. coli* (STEC) whose Shiga toxin gene (*stx)* is found within a prophage genome present in the bacterial genome[28]. We have used antibiotic induction, MALDI-TOF-TOF-MS/MS, and top-down proteomic analysis previously to detect and identify Stx types and subtypes produced by STEC strains[29–32]. In the previous work, STEC O113:H21 strain RM7788 was cultured overnight on agar media supplemented with mitomycin-C. However, instead of detecting the anticipated B-subunit of Stx2a at *m/z* ~7816, a different protein ion was detected at *m/z* ~7839 and identified as a plasmid-encoded hypothetical protein of unknown function[33]. In the current work, we identified two plasmid-encoded AB-Im proteins produced by this strain using antibiotic induction, MALDI-TOF-TOF-MS/MS, and top-down proteomic analysis using standalone software developed to process and scan *in silico* protein sequences derived from whole-genome sequencing (WGS). In addition, the possibility of post-translation modifications (PTM) involving sequence truncation were incorporated into the software. The immunity proteins were identified using this software from the measured mass of the mature protein ion and sequence-specific fragment ions from PBC caused by the aspartic acid effect and detected by MS/MS-PSD. Finally, a promoter was identified upstream of the AB/Im genes in a plasmid genome that may explain the expression of these genes when this strain is exposed to a DNA-damaging antibiotic. Portions of this work were presented at the National American Chemical Society Fall 2020 Virtual Meeting & Expo (August 17-20, 2020)[34].

PROTOCOL:

1. Microbiological sample preparation

1.1 Inoculate 25 mL of Luria broth (LB) in a 50 mL conical tube with *E. coli* O113:H21 strain RM7788 (or another bacterial strain) from a glycerol stock using a sterile 1 µL loop. Cap the tube and pre-culture at 37 °C with shaking (200 rpm) for 4 h.

1.2 Aliquot 100 µL of pre-cultured broth and spread onto a LB agar plate supplemented with 400 or 800 ng/mL of mitomycin-C. Culture agar plates statically overnight in an incubator at 37 °C.

CAUTION: STEC strains are pathogenic microorganisms. Perform all microbiological manipulations, beyond culturing, in a BSL-2 biosafety cabinet.

1.3. Harvest bacterial cells from single visible colonies using a sterile 1 µL loop and transfer to a

2.0 mL O-ring-lined screw-cap microcentrifuge tube containing 300 µL of HPLC-grade water. Cap the tube, vortex briefly, and centrifuge at 11,337 x *g* for 2 min to pellet the cells.

2 Mass spectrometry

2.1 Spot 0.75 µL aliquot of the sample supernatant onto the stainless steel MALDI target and allow it to dry. Overlay the dried sample spot with saturated solution of sinapinic acid (0.75 mL) in 33% acetonitrile, 67% water, and 0.2% trifluoracetic acid. Allow the spot to dry.

2.2 Analyze the dried sample spots using a MALDI-TOF-TOF mass spectrometer.

2.2.1 After loading the MALDI target into the mass spectrometer, click the button for MS linear mode acquisition in the acquisition software. Enter the m/z range to be analyzed by entering the m/z of the lower and upper bounds (e.g., 2 kDa to 20 kDa) into their respective fields in the acquisition method software.

2.2.2 Click on the sample spot to be analyzed on the MALDI target template in the software. Then, depress the left mouse button and drag the mouse cursor over the sample spot to specify the rectangular region to be sampled for laser ablation/ionization. Release the mouse button and the acquisition will initiate. Collect 1,000 laser shots for each sample spot.

NOTE: Data acquisition is displayed in real-time in the software acquisition window.

2.2.3 If no ions are detected, increase the laser intensity by adjusting the **Sliding Scale Bar** under **Laser Intensity** in the software until the protein ion signal is detected. This is referred to as threshold.

NOTE: Prior to the sample spot analysis, externally calibrate the instrument in MS linear mode with protein calibrants whose m/z span the range being analyzed, e.g., the +1 and +2 charge states of protein calibrants: cytochrome-C, lysozyme, and myoglobin cover a mass range of 2 kDa to 20 kDa. An intermediate mass within the specified mass range is used as a focus mass, e.g., 9 kDa. The focus mass is the ion whose m/z is optimally focused for detection by the linear mode detector.

2.2.4 When the MS linear mode acquisition is complete, click the button for MS/MS reflectron mode acquisition in the acquisition software. Enter the precursor mass to be analyzed into the **Precursor Mass** field. Next, enter an isolation width (in Da) into the **Precursor Mass Window** for the low and high mass side of the precursor mass, e.g., ±100 Da.

2.2.5 Click on the **CID Off** button. Click on the **Metastable Suppressor ON** button. Adjust the laser intensity to at least 90% of its maximum value by adjusting the **sliding scale bar** under the **Laser Intensity** in the software.

2.2.6 Click on the sample spot to be analyzed on the MALDI target template in the software.

Then depress the left mouse button and drag the mouse cursor over the sample spot to specify the rectangular region to be sampled for laser ablation/ionization. Release the mouse button, and the acquisition will initiate. Collect 10,000 laser shots for each sample spot.

NOTE: Prior to the sample spot analysis, the instrument should be externally calibrated in MS/MS-reflectron mode using the fragment ions from post-source decay (PSD) of the +1 charge state of alkylated thioredoxin[35].

2.3 Do not process raw MS data. Process MS/MS-PSD raw data using the following sequence of steps in the specified order: advanced baseline correction (32, 0.5, 0.0) followed by noise removal (two standard deviations) followed by Gaussian smoothing (31 points).

2.4 Manually inspect MS/MS-PSD data for the presence of prominent fragment ions generated by PBC[19,20].

2.5 Evaluate MS/MS data with respect to the absolute and relative abundance of fragment ions and their signal-to-noise (S/N). Use only the most abundant fragment ions for protein identification, especially if the MS/MS-PSD data is noisy.

3 *In silico* protein database construction

3.1 Generate a text file containing *in silico* protein sequences of the bacterial strain, which will be scanned by the Protein Biomarker Seeker software for the protein identification. Protein sequences are derived from whole-genome sequencing (WGS) of the strain being analyzed (or a closely related strain).

3.2 Access the NCBI/PubMed (https://www.ncbi.nlm.nih.gov/protein/) website to download approximately 5,000 protein sequences of the specific bacterial strain (e.g., *Escherichia coli* O113:H21 strain RM7788) being analyzed. The maximum download size is 200 sequences.

3.2.1 In consequence, copy and paste the 25 downloads into a single text file. Select the FASTA (text) format for each download.

4 Operating Protein Biomarker Seeker software

4.1 Double click on the **Protein Biomarker Seeker** executable file. A graphical user interface (GUI) window will appear (**Figure 1**, top panel).

4.2 Enter the mass of the protein biomarker (as measured in MS-linear mode) into the **Mature Protein Mass** field. Next, enter the mass measurement error into the **Mass Tolerance** field. The standard mass measurement error is ±10 Da for a 10,000 Da protein.

4.3 Optionally, click on the **Complementary b/y ion Protein Mass Calculator** button in order to calculate the protein mass from a putative complementary fragment ion pair (CFIP or b/y). A

pop-up window, Protein Mass Calculator Tool, will appear (**Figure 1**, bottom panel).

4.3.1 Enter the **m/z** of the putative CFIP and click on the **Add Pair** button. The calculated protein mass will appear.

4.3.2 Copy and paste this number into the **Mature Protein Mass** field and close the **Protein Mass Calculator Tool** window.

4.4 Select an **N-terminal Signal Peptide Length** by clicking on the **Set Residue Restriction** box. A pop-up with a sliding scale and cursor will appear. Move the cursor to the desired signal peptide length (maximum 50). If no signal peptide length is selected, an unrestricted sequence truncation will be performed by the software.

4.5 Under the **Fragment Ion Condition** in the GUI, select residues for polypeptide backbone cleavage (PBC). Click on the boxes of one or more residues: D, E, N, and/or P.

4.5.1 Click on the **Enter Fragment Ions (+1) To Be Searched** button. A pop-up **Fragment Page** will appear. Next, click on the **Add Fragment Ion** button, which corresponds to the number of fragment ions to be entered, i.e., one click for each fragment ion. A dropdown field will appear for each fragment ion to be entered.

4.5.2 Enter the m/z of the fragment ions and their associated m/z tolerance. When completed, click on the **Save and Close** button.

NOTE: A reasonable m/z tolerance is ±1.5.

4.5.3 Select the minimum number of fragment ions that must be matched for an identification by scrolling to the desired number in the box to the right of **How Many Fragment Ions Need to be Matched**.

NOTE: Three matches should be adequate.

4.5.4 Select cysteine residues to be in their oxidized state by clicking on the corresponding circle. If no protein identifications are found after the search, repeat the search with cysteines in their reduced state. If no identifications are found after the search, widen the fragment ion tolerance to ±3 and repeat the search.

4.6 Under the **File Setup**, click on the **Select FASTA File** button to browse and select the FASTA (text) file containing the *in silico* protein sequences of the bacterial strain previously constructed in protocol steps 3.1 to 3.2. Then select an output folder and create an output file name.

4.7 Click on the **Run Search on File Entries** button. A pop-up window will appear entitled **Confirm Search Parameters** (**Figure 1**, bottom panel), displaying the search parameters before the search is initiated.

4.8 If the search parameters are correct, click on the **Begin Search** button. If the search parameters are not correct, click on the **Cancel** button and re-enter the correct parameters. Once the search is initiated, the parameter window closes, and a new pop-up window with a progress bar appears (**Figure 1**, bottom panel) showing the progress of the search and a running tally of the number of identifications found.

4.9 Upon completion of the search (a few seconds), the progress bar automatically closes, and a summary of the search is displayed in the Log field of the GUI (**Figure 2**, top panel). In addition, a new pop-up window will also appear displaying the protein identification(s) if any (**Figure 2**, bottom panel).

NOTE: *In silico* protein sequences having unrecognized residues, e.g., U or X, are automatically skipped from the analysis and these sequences are subsequently reported with a separate pop-up window to alert the operator as to which (if any) sequences were skipped upon completion of the search.

5 Post-search confirmation of protein sequence

5.1 Confirm the correctness of a candidate sequence by manual analysis.

NOTE: The purpose of the Protein Biomarker Seeker software is to identify a protein sequence with high accuracy by eliminating many obviously incorrect protein sequences from consideration and incorporating sequence truncation as a possible PTM in the mature protein. As the number of possible candidate sequences returned are few, manual confirmation is manageable.

5.2 Generate a table of the average m/z of b- and y-type fragment ions of the candidate sequence using any mass spectrometry or proteomic software having such functionality. Compare the average m/z of *in silico* fragment ions on the C-terminal side of D-, E-, and N-residues (and on the N-terminal side of P-residues) to the m/z of prominent fragment ions from the MS/MS-PSD data.

NOTE: The most prominent MS/MS-PSD fragment ions should be easily matched to D-, E-, and N-associated *in silico* fragment ions. However, the aspartic acid effect fragmentation mechanism is less efficient near the N- or C-termini of a protein sequence[36].

REPRESENTATIVE RESULTS:

Figure 3 (top panel) shows the MS of STEC O113:H21 strain RM7788 cultured overnight on LBA supplemented with 400 ng/mL mitomycin-C. Peaks at m/z 7276, 7337, and 7841 had been identified previously as cold-shock protein C (CspC), cold-shock protein E (CspE), and a plasmid-borne protein of unknown function, respectively[33]. The protein ion at m/z 9780 $[M+H]^+$ was analyzed by MS/MS-PSD as shown in **Figure 3** (bottom panel). The precursor ion was isolated with a TIS window ±100 Da. Fragment ions are identified by their m/z and type/number. The

fragment ion at m/z 2675.9 (highlighted with a star) is spillover from the dissociation of the metastable protein ion at m/z 9655 shown in **Figure 3** (top panel). The theoretical average m/z of each fragment ion is shown in parentheses based on PBC of the sequence of colicin E3 immunity protein (Im3) shown above. Sites of PBC are highlighted with a red asterisk with the corresponding fragment ion(s) produced. The N-terminal methionine is underlined signifying that it is post-translationally removed in the mature protein. The sequence has a single cysteine residue (boxed) and is therefore considered in its reduced state.

Using the mass of the protein biomarker and a few prominent *non-complementary* fragment ions: m/z 1813.8, 2128.9, and 4293.7 (±1.5 tolerance) (**Figure 1**, bottom panel) and restricting PBC to the C-terminal side of D- and E-residues, only one candidate sequence was reported by the software: Im3 protein sequence (without its N-terminal methionine) (**Figure 2**, bottom panel). When selecting fragment ions for a search, it should be emphasized that any group of non-complementary fragment ions assumes that summing the m/z of any two fragment ions in the group (and subtracting two protons) results in a mass sum that do not fall within the biomarker mass and associated mass tolerance (±10 Da). Draft WGS of RM7788 revealed 5008 protein sequences (open reading frames)[37]. Of these ~5,000 full protein sequences, 189,490 full and partial sequences (unrestricted truncation) met the biomarker mass criteria (**Figure 2**, top panel). Those sequences passing the mass criteria then undergo *in silico* PBC on the C-terminal side of D- and/or E-residues. The resulting fragment ions generated are then compared to the observed fragment ions entered. The candidate sequence reported by the software was based solely on its mass and three D- and/or E-specific PBC sites. The specificity achieved by such a small amount of information will be discussed in the next section.

As shown in **Figure 3** (bottom panel), the most abundant fragment ions are the result of PBC on the C-terminal side of D- and E-residues via the aspartic acid effect fragmentation mechanism[19,20]. Two CFIP are observed: b_{67}/y_{17} (m/z 7645.1 / m/z 2128.9) and b_{70}/y_{14} (m/z 7959.4 / m/z 1813.8). These CFIP can be used to more accurately calculate the mass of the protein precursor ion using the simple formula: b (m/z) + y (m/z) – $2H^+$ = protein mass (Da)[33]. Using the two CFIP, we obtain an average mass of the protein: 9771.6 Da, which is closer to its theoretical value of 9772.5 Da than the measured mass of the protein ion in MS-linear mode: 9779 Da (**Figure 3**, top panel). Only a few CFIP were detected because most of the precursor ions having the ionizing proton sequestered at the only arginine residue: R80. The higher gas phase basicity of arginine (237.0 kcal/mol[38]) compared to a lysine residue (K) (221.8 kcal/mol[38]) is likely responsible for preferential sequestration of the ionizing proton at the only R-residue.

Figure 4 (top panel) shows the MS of STEC O113:H21 strain RM7788 cultured overnight on LBA supplemented with 800 ng/mL mitomycin-C. **Figure 4** (top panel) is quite similar to **Figure 3** (top panel), although there are differences in the relative abundance of some protein ions due to the differences in antibiotic concentrations utilized. There are also slight shifts in protein biomarker m/z that reflect differences in external calibration of the instrument on different days. Once again, the protein ions at m/z 7272, 7335, and 7838 are CspC, CspE, and a plasmid-borne protein, respectively. In addition, we detect the Im3 protein ion at m/z 9778 (albeit with less abundance than in **Figure 3**) as well as a protein ion at m/z 9651 $[M+H]^+$. **Figure 4** (bottom panel) shows

MS/MS-PSD of the protein precursor ion at m/z 9651. The precursor ion was isolated using a narrower and asymmetric TIS window of -75/+60 Da to eliminate contributions of adjacent protein ions at m/z 9539 and 9778. Fragment ions are identified by their m/z and type/number. The sequence of the immunity protein of bacteriocin (Im-Bac) is shown above. Sites of PBC are highlighted with a red asterisk with their corresponding fragment ion(s). The theoretical average m/z of each fragment ion is also shown in parentheses in the spectrum. The Im-Bac sequence also has a single cysteine residue (boxed) and is therefore considered in its reduced state.

Using the protein biomarker mass, three prominent non-complementary fragment ions: m/z 2675.4, 3853.5, and 5772.8 (±1.50 tolerance) from **Figure 4** and restricting PBC to only the C-terminal side of D- and/or E- and/or asparagine (N)-residues, only one candidate sequence was reported by the software: Im-Bac protein. The candidate sequence was retrieved after scanning 191,375 full or partial sequences that met the biomarker mass and tolerance (±10 Da) criteria. The candidate sequence was identified by the software-based solely on its mass and three D- and/or E- and/or N-specific PBC sites.

The most prominent fragment ions in **Figure 4** (bottom panel) were, once again, the result of PBC on the C-terminal side of D and/or E-residues and also on the N-terminal side of one of the P-residues[20]. We also observe PBC on the C-terminal side of an N-residues that is also likely to occur by an aspartic acid effect-like fragmentation mechanism[39,40]. The weakness of the protein precursor ion signal results in a limited number of interpretable fragment ions. The accuracy of the fragment ion m/z declines with fragment ion abundance. No CFIP were detected due presumably to the ionizing proton being sequestered at the only arginine residue (R74) of the protein ion sequence. All fragment ions contain the R74 residue, consistent with this hypothesis.

The promoter of antibacterial immunity genes

Figure 5 shows a portion of the 6482 bp contig00100 of *E. coli* strain RM7788 (GenBank: NWVS01000096.1) from whole-genome shotgun sequencing[37]. The coding regions for colicin E3, its immunity protein (Im3), the immunity protein of bacteriocin (Im-Bac), and a lysis protein are highlighted in yellow. Upstream of the coding region for the colicin E3 gene are the -35 region, the Pribnow box (PB), inverted repeat of the SOS box, the Shine-Dalgarno/ribosomal binding site (SD/RBS)[27]. There is a nine base-pair intergenic region between colicin E3 and Im3. LexA (a repressor protein and an autopeptidase) binds to the SOS box blocking the expression of genes downstream. Upon DNA damage (e.g., UV radiation or DNA-damaging antibiotics), LexA undergoes self-cleavage allowing expression of genes downstream[27,28]. Thus, the expression of these two immunity proteins is consistent with exposure of this strain to a DNA-damaging antibiotic.

FIGURE AND TABLE LEGENDS:

Figure 1: **Screen shots of Protein Biomarker Seeker software.** Top panel: Graphical user interface (GUI) of the Protein Biomarker Seeker software. Bottom panels: Pop-up windows of Protein Mass Calculator Tool, Fragment Page, Confirm Search Parameters, and Search progress bar.

Figure 2: **Search results of a protein identification using Protein Biomarker Seeker software.** Top panel: Summary of search results displayed in the Log Field of the software GUI. Bottom panel: A pop-up window displaying a protein identification using the software.

Figure 3: **Mass spectrometry analysis of STEC O113:H21 strain RM7788.** Top panel: MS of STEC O113:H21 strain RM7788 cultured overnight on LBA supplemented with 400 ng/mL mitomycin-C. Bottom panel: MS/MS-PSD of the protein precursor ion at m/z 9780 (top panel). The precursor ion was isolated with a TIS window ±100 Da. Fragment ions are identified by their m/z and ion type. The sequence of the immunity protein for colicin E3 (Im3) is shown. Basic residues (sites of possible charge sequestration) are highlighted in blue. PBC are highlighted with a red asterisk with the corresponding fragment ion(s) generated. The theoretical average m/z of each fragment ions is shown in parentheses.

Figure 4: Mass spectrometry analysis of STEC O113:H21 strain RM7788. Top panel: MS of STEC O113:H21 strain RM7788 cultured overnight on LBA supplemented with 800 ng/mL mitomycin-C. Bottom panel: MS/MS-PSD of the protein precursor ion at m/z 9651 (top panel). The precursor ion was isolated with an asymmetric TIS window of -75 on the low m/z side of the precursor ion and +60 on the high m/z side of the precursor ion. Fragment ions are identified by their m/z and ion type. The sequence of the immunity protein of bacteriocin (Im-Bac) is shown. Basic residues (sites of possible charge sequestration) are highlighted in blue. PBC are highlighted with a red asterisk with the corresponding fragment ion(s) generated. The theoretical average m/z of each fragment ion is shown in parentheses.

Figure 5: **Analysis of a section of the plasmid genome carried by *E. coli* O113:H21 strain RM7788.** A portion of the 6482 bp contig00100 of *E. coli* O113:H21 strain RM7788 (GenBank: NWVS01000096.1) from whole genome shotgun sequencing[37].

Supplementary File 1 (S1 Im3): Results of benchmarking analysis of software using select fragment ions of Im3 (from **Figure 3**, bottom panel).

Supplementary File 2 (S2 ImBac): Results of benchmarking analysis of software using select fragment ions of Im-Bac (from **Figure 4**, bottom panel).

DISCUSSION:

Protocol considerations

The primary strengths of the current protocol are its speed, simplicity of sample preparation, and use of an instrument that is relatively easy to operate, be trained on, and maintain. Although bottom-up and top-down proteomic analysis by liquid chromatography-ESI-HR-MS are ubiquitous and far superior in many respects to top-down by MALDI-TOF-TOF, they require more time, labor, and expertise. Instrument complexity can often affect whether certain instrument platforms are likely to be adopted by scientists not formally trained in mass spectrometry. The top-down approach with MALDI-TOF-TOF is meant to extend the analysis of MALDI-TOF-MS beyond its current use for taxonomic identification of bacteria in clinical microbiology labs while not dramatically increasing the labor, complexity, or expertise required for analysis.

The protocol does not employ any mechanical (or electrical) cell lysis step. Although secreted or extracellular proteins may be detected using the protocol, an earlier version of this method was first developed for detection of Shiga toxin (Stx) from STEC strains wherein antibiotic induction triggers the bacterial SOS response resulting in expression of phage genes, including *stx* as well as late phage genes responsible for bacterial cell lysis[41]. We found that antibiotic-induced cell lysis has certain advantages for the detection of Stx as well as plasmid proteins that have SOS promoters (current work). Certainly, mechanical cell lysis (e.g., bead-beating) can also be used (although not used in the current work). However, mechanical lysis results in all bacterial cells being lysed (not simply induced cells) resulting in the sample being enriched with abundant, highly conserved host proteins that can make detection of phage and plasmid proteins from an unfractionated sample more challenging.

The antibiotic concentrations for a bacterial strain were found to be generally reproducible with respect to the antibiotic-induced proteins detected. We noted variations in the relative protein abundance with respect to the antibiotic-induced proteins detected. Since our analysis is qualitative (not quantitative), protein biomarker abundance need only be sufficient for adequate MS/MS analysis. A putative STEC strain is first cultured with a range of antibiotic concentrations (e.g., 300 ng/mL to 2,000 ng/mL of mitomycin-C) to determine the optimum concentration such that it triggers the bacterial SOS response while still providing enough bacterial cells for harvesting. For the STEC strain RM7788, we found that the optimum antibiotic concentration for detection of the biomarkers identified was 400 to 800 ng/mL of mitomycin-C.

In addition to protein sequence truncation, *E. coli* proteins can have PTMs that involve addition of mass, e.g., phosphorylation, glycosylation, etc. As MS/MS utilizes PSD for dissociation of singly charged metastable protein ions (under 20 kDa in mass) generated by MALDI, such PTMs attached to residue side chains would likely undergo facile dissociative loss because PSD is an ergodic dissociation technique. The presence of such PTMs could be inferred from the appearance of a fragment ion close in mass to the original precursor ion (minus the mass of the PTM) in the MS/MS data. However, neither PSD nor the software would be able to identify where such PTMs are attached. In addition, the software can only identify proteins from fragment ions of PBC and not dissociative loss of small molecules (e.g., water or ammonia) or PTMs attached to the side-chains of residues. However, if fragment ions from PBC are detected, the protein could still be identified using the software by either widening the protein mass tolerance window to include the mass of the PTM or simply entering the mass of the protein fragment ion corresponding to dissociative loss of the suspected PTM. Any identification by the software would be of the protein sequence without the PTM. Interestingly, we have not detected proteins having phosphorylation, glycosylation, etc. in our bacterial work thus far. However, that may be due to: their relative abundance by MALDI, the mass range being used: 2–20 kDa, that such PTMs may be unusually labile and may not survive application of the MALDI matrix, or that such PTMs may undergo very rapid dissociative loss in the source before ions are accelerated from the source.

Currently, the software does not include cysteine alkylation, and our sample protocol does not

include a disulfide reduction step for cysteine residues. The protocol has been clarified to indicate that the search is to be operated with cysteine residues in their Oxidized state, and if no identification is obtained, then to execute the search again with cysteine residues in their Reduced state. If no identifications are found again, widening the fragment ion tolerance to ±2 or ±3 lowers the threshold for fragment ion matching allowing sequences with cysteines to be matched whether they are present in their oxidized and/or reduced states.

Top-down proteomic analysis by MALDI-TOF-TOF mass spectrometry

Most top-down proteomic analysis has been achieved using ESI and high-resolution mass spectrometry platforms. By contrast, fewer top-down proteomic analysis has been conducted using MALDI-TOF-TOF platforms. In consequence, there is very little top-down proteomic software for analysis of singly charged metastable protein ions generated and analyzed by MALDI-TOF-TOF-MS/MS-PSD that exploit the aspartic acid effect for fragmentation[15,42]. There are a number of reasons for this. First, the ionization efficiency of MALDI is biased toward lower molecular weight peptides and proteins, and this bias is particularly apparent with a mixture of proteins as would be found in an unfractionated bacterial cell lysate. Second, MALDI generates low charge states, and there is little or no Coulomb repulsion to facilitate protein ion dissociation. Third, PSD sequence coverage is quite limited unlike other techniques ECD[7], ETD[7], UV-PD[8], etc. Fourth, the fragmentation efficiency of PSD declines with increasing mass of the protein ion. Fifth, ergodic dissociation techniques, such as PSD, tend to result in facile dissociative loss of PTMs attached to residues, e.g., phosphorylation, glycosylation, etc., making it challenging to determine the site of PTM attachment. In spite of these severe limitations, top-down analysis using MALDI-TOF-TOF-MS/MS-PSD has clear advantages, e.g., simplicity of sample preparation, absence of LC separation, isolation of metastable protein ions by MS/MS allowing attribution of fragment ions to precursor ions, identification of PTMs involving sequence truncation and intramolecular disulfide bonds and most importantly the speed of analysis. When combined with *in silico* protein sequences derived from WGS data, this technique can provide rapid information before other more time-consuming and labor-intensive analyses are completed.

The Protein Biomarker Seeker software was developed using IntelliJ and written in Java to efficiently process and search protein amino acid sequences derived from WGS of a bacterial strain. The software was modified from an earlier algorithm that operated as a macro within Excel[33]. We decided to develop a standalone version of the software with a GUI interface to make it more user-friendly as well as provide further improvements.

In the event of PTMs involving protein sequence truncation, the software sequentially removes an amino acid residue from the N-terminus while iteratively adding residues of the sequence until the mass sum meets or exceeds the measured mass of the detected protein biomarker. Although this process can result in a very large number of protein mass fragments (~200,000 from ~5000 full protein sequences), it has the advantage of not excluding any potential protein fragments from the truncation at the N-terminus or C-terminus (or both) however improbable such truncation may be from a biological perspective. This approach is referred to as unrestricted truncation. However, the most common bacterial PTMs involving truncation are removal of the N-terminal methionine or N-terminal signal peptide. In consequence, the software also allows

the operator to select an upper limit (50 residues) for residue truncation from the N-terminus, which results in much fewer protein fragments that meet the protein biomarker mass criteria.

PBC on the C-terminal side of D- and E-residues as well as on the N-terminal side of P-residues are consistent with the aspartic acid effect mechanism, which has been studied extensively both experimentally and theoretically[17–20]. Inclusion of PBC on the C-terminal side of N-residues was included in the software because of an aspartic acid effect-like mechanism that has been observed for a number of metastable protein ions in our laboratory[39,40]. The most abundant fragment ions from the dissociation of singly charged metastable protein ions analyzed by MS/MS-PSD are due to the aspartic acid effect fragmentation mechanism. The operator selects the most prominent fragment ions from the MS/MS-PSD data and enters their m/z into the software as well as an associated fragment ion tolerance (±m/z). The fragment ion tolerance can be adjusted for each fragment ion to reflect its relative abundance. An appropriate fragment ion tolerance may vary between ±1.0 to ±2.5 m/z depending on the absolute abundance of the fragment ion as well as its relative abundance compared to background chemical noise. Typically, the more abundant a fragment ion, the better its mass accuracy, which allows a narrower fragment ion tolerance to be used.

MS/MS-PSD data of metastable protein ions can vary dramatically in terms of their complexity. Some MS/MS-PSD spectra are more easily interpretable than others. There are several reasons for this phenomenon. First, the protein ion may not fragment efficiently on the timescale of the analysis (~10–30 μs) perhaps because it remains folded or partially folded even after solubilization in the MALDI matrix solution. Second, in addition to PBC, metastable protein ions can undergo dissociative loss of small molecules, i.e., ammonia (-17 Da) or water (-18 Da)[15]. A significant contributor to spectral complexity appears to be dissociative loss of ammonia from the side-chain of R-residues[33]. We have observed an increase in spectral complexity of MS/MS-PSD data with the number of R-residues in the protein sequence. Proteins with no R-residues (YahO protein[36] and cold-shock protein CspC[33,43]), with one R-residue (cold-shock protein CspE[33] and B-subunit of Stx2[41]), with two R-residues (hypothetical protein[33]), produce MS/MS-PSD spectra that are relatively uncomplicated and easy to interpret. However, when the number of R-residues increase to three (HU protein[44]), or four (ubiquitin[35] and cold-shock protein CsbD[33,43]), spectral complexity increases significantly. The software compares fragment ions from PBC at residues specific to the aspartic acid effect mechanism *only* as this is the most accessible dissociation channel of singly charged metastable protein ions analyzed by MS/MS-PSD. The software does *not* include fragment ions resulting from dissociative loss (or losses) of small neutral molecule(s). In consequence, it is important that the operator does not select fragment ions that include small neutral dissociative losses. Fragment ions from PBC are typically the most prominent fragment ions; however, when the number of R-residues in a protein increases to three or four, the most abundant fragment ion at a PBC site may be one that includes a small dissociative loss (or losses). If such a cluster of fragment ions (separated by multiples of 17 or 18 m/z) is detected, the fragment ion with the highest m/z within a cluster should be the one entered into the fragment ion search parameters.

It should be emphasized that the software was not designed for operator-free proteomic

identification. The operator must select which fragment ions from MS/MS-PSD data are to be included in the search. However, based on numerous experiments that have confirmed the aspartic acid effect by MS/MS-PSD, the most prominent fragment ions are always the result of PBC on the C-terminal side of D- or E- or N-residues. The utility of the software is that it eliminates many obviously incorrect sequences and retrieves only a few likely candidates. Some candidate sequences may be eliminated based on the absence of a fragment ion where a D-residue in a sequence would be expected to generate a prominent fragment ion. Invariably, D-residues result in prominent fragment ions throughout the polypeptide backbone *except* when they are located within a few residues of the N- or C-termini where the efficiency of the aspartic acid effect declines[36].

Minimum number of PBC sites needed for tentative protein identification

A CFIP is formed from two identical protein precursor ions that dissociate at the same PBC site but have their ionizing proton on opposite sides of the cleavage site. Although a CFIP can be used to calculate the mass of the protein biomarker more accurately (allowing a narrowing of the protein mass tolerance during a search), its utility for sequence-specific identification is less useful than that of two non-complementary fragment ions formed from two different cleavage sites, which provide greater identification specificity. The ease with which the two AB-Im proteins were identified led us to speculate as to the minimum number of fragment ions necessary to tentatively identify the correct protein sequence from thousands of proteins or protein fragment sequences. We quickly determined that it was not the number of fragment ions per se but the number of non-complementary fragment ions that is important because each non-complementary fragment ion represents one PBC site whereas a CFIP represents the same cleavage site. Thus, identification specificity is derived from the number of PBC sites detected not the number of fragment ions.

It is possible that the success in identification with only three fragment ions may have been simply fortuitous. To test this hypothesis and to eliminate bias in the selection of fragment ions, we created a benchmarking module within the software that randomly selects fragment ions from a larger pool of complementary and/or non-complementary fragment ions. The larger fragment ion pool was selected from the 14 prominent fragment ions identified in **Figure 3** (bottom panel) based upon their relative abundance.

The testing protocol was as follows. Using a binary search, three fragment ions were randomly selected from the pool of 14 prominent fragment ions in **Figure 3** (bottom panel) (m/z 1813.8, 2128.9, 3881.3, 4293.7, 5158.0, 6505.0, 6619.9, 6939.4, 7645.1, 7959.4, 8022.7, 8136.2, 8583.3, and 8961.5). A three-fragment ion cohort was compared against *in silico* fragment ions from PBC on the C-terminal side of D- or E- or N-residues as well as a combination of D & E and D & E & N. This comparison was performed for each individual fragment ion of a cohort, for the three fragment ion pairs of a cohort and for the three-fragment ion combination of a cohort. For a comparison to be counted as a match, both fragment ions of a pair and all three fragment ions of a combination must match to *in silico* fragment ions. After completion of the analysis, another three-fragment ion cohort is randomly selected, and the analysis is repeated. Repetition in fragment ion selection was allowed. As there are 364 possible combinations [(n!/r!(n-r)!] of a

three-fragment ion cohort (r) from a pool of 14 fragment ions (n), only 10 analyses were performed as shown in the **S1 Im3** (Supplementary Information).

The three-fragment ion identification requirement appears to be a general phenomenon as shown in column *3_ABC* of **Tables 2-7**, **9-10** (**S1 Im3**). All counts of 1 in the *3_ABC* column correspond to the Im3 sequence (without N-terminal methionine). The only failure in identification occurred because the fragment ion at m/z 8136.2 (shown in **Figure 3**, bottom panel and highlighted in gray in **Tables 1** and **8**) exceeded the fragment ion tolerance (±1.5 m/z) entered for the analysis. Since the testing algorithm requires that all fragment ions of a three-fragment ion cohort be matched, any group that included the m/z 8136.2 fragment ion would fail to identify/count the correct protein sequence.

Table 6 in **S1 Im3** shows that when two of three fragment ions are complementary (highlighted in yellow), more incorrect sequences matched the criteria than that observed when all three fragment ions were non-complementary. As noted previously, this is because a CFIP corresponds to a single PBC site, a threshold that is attainable by many more incorrect *in silico* sequences compared to using two non-complementary fragment ions that correspond to two PBC sites, a more stringent criterion.

A similar analysis was performed on six prominent fragment ions (m/z 2675.4, 2904.5, 3076.2, 3853.5, 5657.5, and 5772.8) of Im-Bac shown in **Figure 4** (bottom panel). Unlike Im3, Im-Bac has no discernable CFIP, therefore the six fragment ions correspond presumably to six PBC sites. As there are 20 possible combinations of a three-fragment ion cohort selected from a pool of six fragment ions, only 10 analyses were performed as shown in the tables of **S2 Im-Bac** (Supplementary Information). The Im-Bac sequence was correctly identified/counted for all three-fragment ion groups in column *3_ABC* in all analyses. In four analyses, one or two incorrect sequences were also matched. However, this small number of incorrect sequences is a manageable number for manual confirmation.

Overall, complementary and/or non-complementary fragment ions that correspond to two or three PBC sites appear to provide enough specificity to retrieve one or two candidate sequences. Of course, the fragment ions selected by the operator should be relatively abundant and have good S/N. One or two fragment ions from a single PBC site does not provide enough specificity to avoid retrieving an unworkable number of incorrect sequences that must be confirmed by the operator. It is not clear why two or three PBC sites are adequate, but a single PBC site is apparently not specific enough. Although unrestricted truncation results in ~200,000 proteins and protein fragment sequences that meet the protein mass criteria, it is probable that the site/residue-specific nature of the cleavage sites, i.e., C-terminal side of D-, E-, and N-residues, contributes to the sharp narrowing of possible sequences during fragment ion comparison. This may be due, in part, to the frequency of D-, E-, and N-residues in bacterial protein sequences as well as their unique locations in protein sequences across the proteome of bacteria. Acidic residues play critical roles in protein structure and solvent interactions. In consequence, their frequency and locations in the primary sequence are critical if not unique for protein function and may explain why only a few PBC sites are necessary to tentatively identify the correct protein

sequence among hundreds of thousands incorrect sequences.

From a gas phase chemistry perspective, the importance of D-, E-, and N-residues stems from their participation in a dissociation channel that is accessible at low internal energies of singly charged metastable protein ions generated by MALDI and decay by PSD[20]. The relatively long timescale (~10–30 μs) of molecular ion fragmentation by PSD means that the internal energy of the protein ion is randomized among all vibrational and rotational degrees-of-freedom of the molecular ion such that dissociation is ergodic and statistical. It should also be pointed out that the mechanism of aspartic acid effect involves a molecular ion rearrangement that occurs by a sequence of steps or a single concerted step involving multiple atoms until a favorable geometry is achieved that lowers the activation barrier of PBC[17–19].

Two plasmid-encoded antibacterial immunity proteins produced by a STEC strain were identified using a protocol involving antibiotic induction, MALDI-TOF-TOF-MS/MS-PSD, and top-down proteomic analysis. These proteins were identified using software developed in-house that incorporates the measured mass of the protein and a relatively small number of sequence-specific fragment ions formed as a result of the aspartic acid effect. The software compares the MS and MS/MS data to *in silico* protein and protein fragment sequences derived from WGS data. Although the software does not provide identification metrics or scoring, it eliminates a very high percentage of incorrect sequences resulting in a very small number of candidate sequences (one or two) that can be easily confirmed by manual inspection. Finally, manual inspection of the WGS data of this bacterial strain revealed a promoter (SOS box) upstream of the AB and Im genes in a plasmid genome, which rationalizes expression of these genes due to exposure of DNA-damaging antibiotics.

ACKNOWLEDGMENTS:

Protein Biomarker Seeker software is freely available (at no cost) by contacting Clifton K. Fagerquist at clifton.fagerquist@usda.gov. We wish to acknowledge support of this research by ARS, USDA, CRIS grant: 2030-42000-051-00-D.

DISCLOSURES:

The authors have no conflicts of interest.

REFERENCES:

1. Fornelli, L. et al. Accurate sequence analysis of a monoclonal antibody by top-down and middle-down orbitrap mass spectrometry applying multiple ion activation techniques. *Analytical Chemistry*. **90** (14), 8421–8429 (2018).
2. Fornelli, L. et al. Top-down proteomics: Where we are, where we are going? *Journal of Proteomics*. **175**, 3–4 (2018).
3. He, L. et al. Top-down proteomics-a near-future technique for clinical diagnosis? *Annals of Translational Medicine*. **8** (4), 136 (2020).
4. Wu, Z. et al. MASH explorer: A universal software environment for top-down proteomics. *Journal of Proteome Research*. **19** (9), 3867–3876 (2020).

5. Konermann, L., Metwally, H., Duez, Q., Peters, I. Charging and supercharging of proteins for mass spectrometry: recent insights into the mechanisms of electrospray ionization. *Analyst*. **144** (21), 6157–6171 (2019).
6. Bourmaud, A., Gallien, S., Domon, B. Parallel reaction monitoring using quadrupole-Orbitrap mass spectrometer: Principle and applications. *Proteomics*. **16** (15–16), 2146–2159 (2016).
7. Hart-Smith, G. A review of electron-capture and electron-transfer dissociation tandem mass spectrometry in polymer chemistry. *Analitica Chimica Acta*. **808**, 44–55 (2014).
8. Brodbelt, J. S., Morrison, L. J., Santos, I. Ultraviolet photodissociation mass spectrometry for analysis of biological molecules. *Chemical Reviews*. **120** (7), 3328–3380 (2020).
9. Karas, M., Bachmann, D., Bahr, U., Hillenkamp, F. Matrix-assisted ultraviolet-laser desorption of nonvolatile compounds. *International Journal of Mass Spectrometry and Ion Processes*. **78**, 53–68 (1987).
10. Karas, M., Bachmann, D., Hillenkamp, F. Influence of the wavelength in high-irradiance ultraviolet-laser desorption mass-spectrometry of organic-molecules. *Analytical Chemistry*. **57** (14), 2935–2939 (1985).
11. Tanaka, K. et al. Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*. **2** (8), 151–153 (1988).
12. Resemann, A. et al. Top-down de Novo protein sequencing of a 13.6 kDa camelid single heavy chain antibody by matrix-assisted laser desorption ionization-time-of-flight/time-of-flight mass spectrometry. *Analytical Chemistry*. **82** (8), 3283–3292 (2010).
13. Suckau, D., Resemann, A. T3-sequencing: targeted characterization of the N- and C-termini of undigested proteins by mass spectrometry. *Analytical Chemistry*. **75** (21), 5817–5824 (2003).
14. Mikhael, A., Jurcic, K., Fridgen, T. D., Delmas, M., Banoub, J. Matrix-assisted laser desorption/ionization time-of-flight/time-of-flight tandem mass spectrometry (negative ion mode) of French Oak Lignin: A Novel Series of Lignin and Tricin Derivatives attached to Carbohydrate and Shikimic acid Moieties. *Rapid Communications in Mass Spectrometry*. **34** (18), e8841 (2020).
15. Demirev, P. A., Feldman, A. B., Kowalski, P., Lin, J. S. Top-down proteomics for rapid identification of intact microorganisms. *Analytical Chemistry*. **77** (22), 7455–7461 (2005).
16. Fagerquist, C. K. Unlocking the proteomic information encoded in MALDI-TOF-MS data used for microbial identification and characterization. *Expert Review of Proteomics*. **14** (1), 97–107 (2017).
17. Gu, C., Tsaprailis, G., Breci, L., Wysocki, V. H. Selective gas-phase cleavage at the peptide bond C-terminal to aspartic acid in fixed-charge derivatives of Asp-containing peptides. *Analytical Chemistry*. **72** (23), 5804–5813 (2000).
18. Herrmann, K. A., Wysocki, V. H., Vorpagel, E. R. Computational investigation and hydrogen/deuterium exchange of the fixed charge derivative tris(2,4,6-trimethoxyphenyl) phosphonium: implications for the aspartic acid cleavage mechanism. *Journal of the American Society for Mass Spectrometry*. **16** (7), 1067–1080 (2005).
19. Rozman, M. Aspartic acid side chain effect-experimental and theoretical insight. *Journal of the American Society for Mass Spectrometry*. **18** (1), 121–127 (2007).

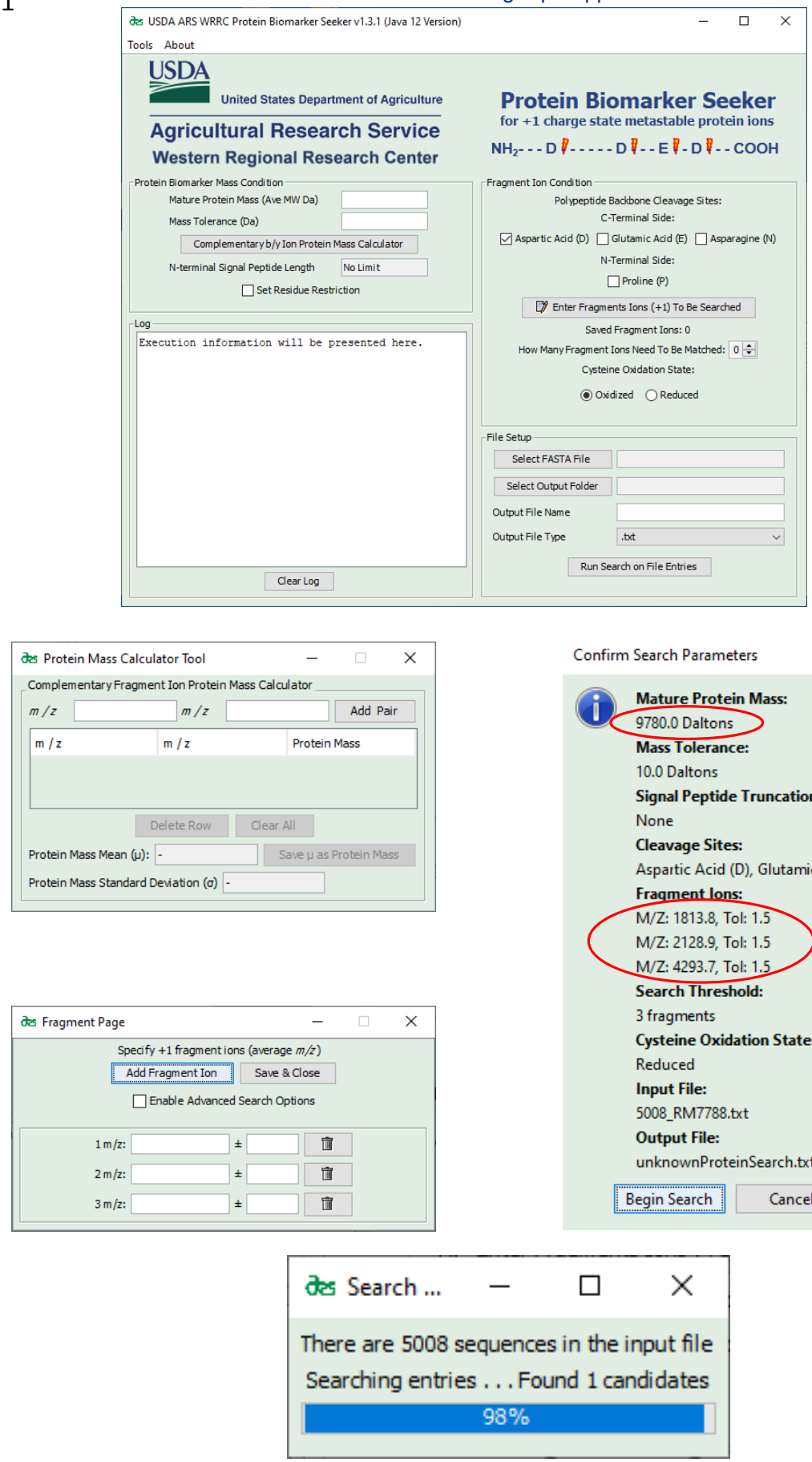
20. Yu, W., Vath, J. E., Huberty, M. C., Martin, S. A. Identification of the facile gas-phase cleavage of the Asp-Pro and Asp-Xxx peptide bonds in matrix-assisted laser desorption time-of-flight mass spectrometry. *Analytical Chemistry*. **65** (21), 3015–3023 (1993).
21. Luethy, P. M., Johnson, J. K. The use of Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS) for the identification of pathogens causing sepsis. *The Journal of Applied Laboratory Medicine*. **3** (4), 675–685 (2019).
22. Knabl, L., Lass-Florl, C. Antifungal susceptibility testing in Candida species: current methods and promising new tools for shortening the turnaround time. *Expert Review of Anti-Infective Therapy*. **18** (8), 779–787 (2020).
23. Gould, O., Ratcliffe, N., Krol, E., de Lacy Costello, B. Breath analysis for detection of viral infection, the current position of the field. *Journal of Breath Research*. **14** (4), 041001 (2020).
24. Fagerquist, C. K. et al. Sub-speciating Campylobacter jejuni by proteomic analysis of its protein biomarkers and their post-translational modifications. *Journal of Proteome Research*. **5** (10), 2527–2538 (2006).
25. Sandrin, T. R., Goldstein, J. E., Schumaker, S. MALDI TOF MS profiling of bacteria at the strain level: a review. *Mass Spectrometry Reviews*. **32** (3), 188–217 (2013).
26. Christner, M. et al. Rapid MALDI-TOF mass spectrometry strain typing during a large outbreak of Shiga-Toxigenic Escherichia coli. *PLoS One*. **9** (7), e101924 (2014).
27. Masaki, H., Ohta, T. Colicin E3 and its immunity genes. *Journal of Molecular Biology*. **182** (2), 217–227 (1985).
28 Michel, B. After 30 years of study, the bacterial SOS response still surprises us. *PLoS Biology*. **3** (7), e255 (2005).
29 Fagerquist, C. K., Sultan, O. Induction and identification of disulfide-intact and disulfide-reduced beta-subunit of Shiga toxin 2 from Escherichia coli O157:H7 using MALDI-TOF-TOF-MS/MS and top-down proteomics. *Analyst*. **136** (8), 1739–1746 (2011).
30. Fagerquist, C. K., Sultan, O. Top-down proteomic identification of furin-cleaved alpha-subunit of Shiga toxin 2 from Escherichia coli O157:H7 using MALDI-TOF-TOF-MS/MS. *Journal of Biomedicine & Biotechnology*. **2010**, 123460 (2010).
31. Fagerquist, C. K. et al. Top-down proteomic identification of Shiga toxin 2 subtypes from Shiga toxin-producing Escherichia coli by matrix-assisted laser desorption ionization-tandem time of flight mass spectrometry. *Applied and Environmental Microbiology*. **80** (9), 2928–2940 (2014).
32. Fagerquist, C. K., Zaragoza, W. J., Lee, B. G., Yambao, J. C., Quiñones, B. Clinically-relevant Shiga toxin 2 subtypes from environmental Shiga toxin-producing Escherichia coli identified by top-down/middle-down proteomics and DNA sequencing. *Clinical Mass Spectrometry*. **11**, 27–36 (2019).
33. Fagerquist, C. K., Lee, B. G., Zaragoza, W. J., Yambao, J. C., Quiñones, B. Software for top-down proteomic identification of a plasmid-borne factor (and other proteins) from genomically sequenced pathogenic bacteria using MALDI-TOF-TOF-MS/MS and post-source decay. *International Journal of Mass Spectrometry*. **438**, 1–12 (2019).
34. Fagerquist, C. K., Rojas, E. in *ACS Fall 2020 Virtual Meeting & Expo.* American Chemical Society, Virtual (2020).
35. Fagerquist, C. K., Sultan, O. A new calibrant for matrix-assisted laser desorption/ionization time-of-flight-time-of-flight post-source decay tandem mass spectrometry of non-digested

proteins for top-down proteomic analysis. *Rapid Communications in Mass Spectrometry*. **26** (10), 1241–1248 (2012).
36. Fagerquist, C. K., Zaragoza, W. J. Complementary b/y fragment ion pairs from post-source decay of metastable YahO for calibration of MALDI-TOF-TOF-MS/MS. *International Journal of Mass Spectrometry*. **415**, 29–37 (2017).
37. Quinones, B., Yambao, J. C., Lee, B. G. Draft genome sequences of Escherichia coli O113:H21 strains recovered from a major produce production region in California. *Genome Announcements*. **5** (44), e01203-17 (2017).
38. Harrison, A. G. The gas-phase basicities and proton affinities of amino acids and peptides. *Mass Spectrometry Reviews*. **16** (4), 201–217 (1997).
39. Fagerquist, C. K. Polypeptide backbone cleavage on the C-terminal side of asparagine residues of metastable protein ions analyzed by MALDI-TOF-TOF-MS/MS and post-source decay. *International Journal of Mass Spectrometry*. **457** (2020).
40. Fagerquist, C. K., Zaragoza, W. J. in *Mass Spectrometry: Application to the Clinical Lab 2019 (MSACL 2019)* Vol. 2019 (ed Mass Spectrometry: Application to the Clinical Lab 2019 (MSACL 2019)) MSACL, Palm Springs, CA (2019).
41. Fagerquist, C. K., Zaragoza, W. J. Bacteriophage cell lysis of Shiga toxin-producing Escherichia coli for top-down proteomic identification of Shiga toxins 1 & 2 using matrix-assisted laser desorption/ionization tandem time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*. **30** (6), 671–680 (2016).
42. Fagerquist, C. K. et al. Web-based software for rapid top-down proteomic identification of protein biomarkers, with implications for bacterial identification. *Applied and Environmental Microbiology*. **75** (13), 4341–4353 (2009).
43. Fagerquist, C. K. et al. Rapid identification of protein biomarkers of Escherichia coli O157:H7 by matrix-assisted laser desorption ionization-time-of-flight-time-of-flight mass spectrometry and top-down proteomics. *Analytical Chemistry*. **82** (7), 2717–2725 (2010).
44. Maus, A., Bisha, B., Fagerquist, C., Basile, F. Detection and identification of a protein biomarker in antibiotic-resistant Escherichia coli using intact protein LC offline MALDI-MS and MS/MS. *Journal of Applied Microbiology*. **128** (3), 697–709 (2020).

Figure

Figure 1

Click here to access/download;Figure;REVISED_Figure 1 Fagerquist.pptx



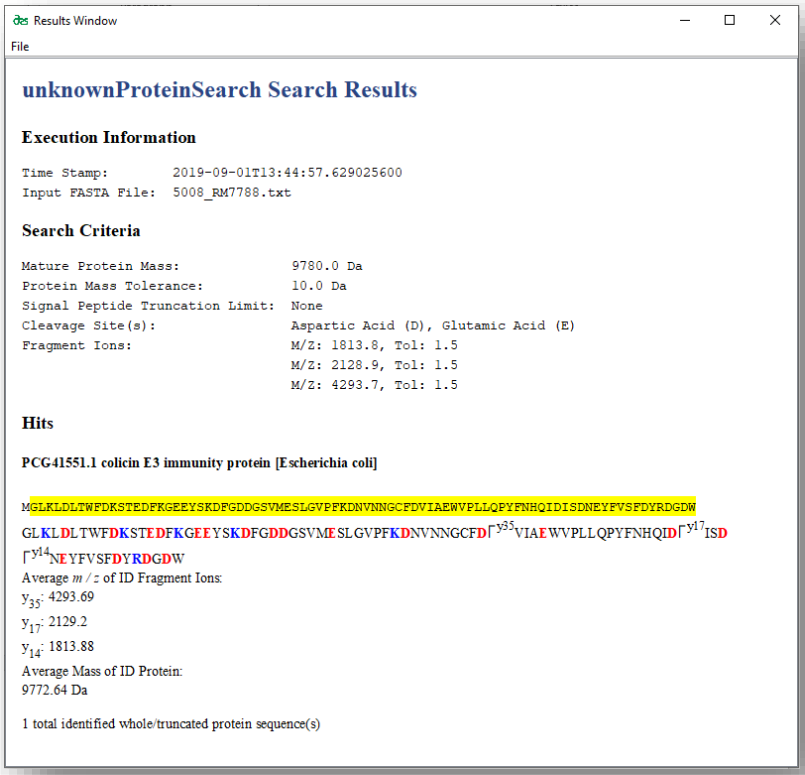
Figure

Click here to access/download;Figure;REVISED_Figure 2 Fagerquist.pptx

Figure 2

Execution information will be presented here.

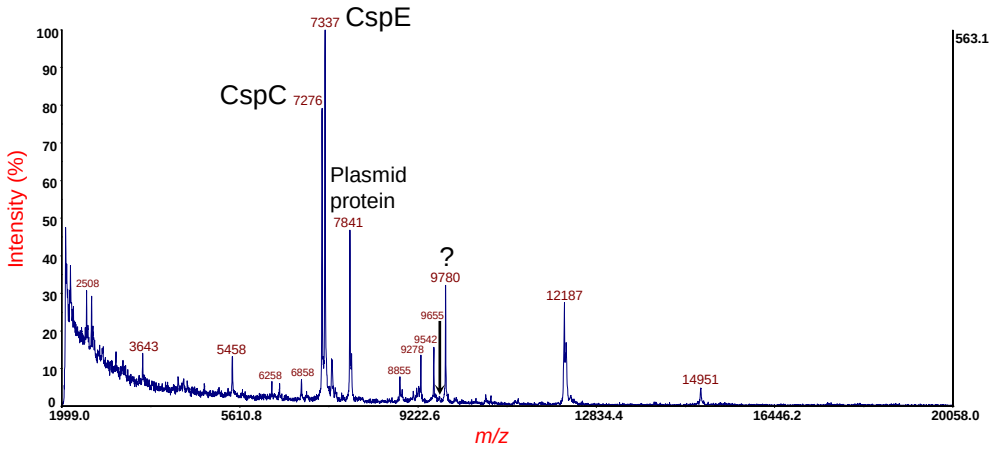
Reading 5008_RM7788.txt
-------SEARCH CRITERIA -------
Protein Biomarker Mass Window:
From 9770.0 to 9790.0 Da
Searching for these 3 fragment ions:
M/Z: 1813.8, Tol: 1.5
M/Z: 2128.9, Tol: 1.5
M/Z: 4293.7, Tol: 1.5
3 fragment ions must be matched
-------SEARCH RESULTS -------
Searched entries:
5008
Time elapsed:
3 second(s)
Protein sequences and/or truncated protein sequences that met the Protein Biomarker Mass condition:
189490
Protein sequences and/or truncated protein sequences that met the fragment ion condition:
1

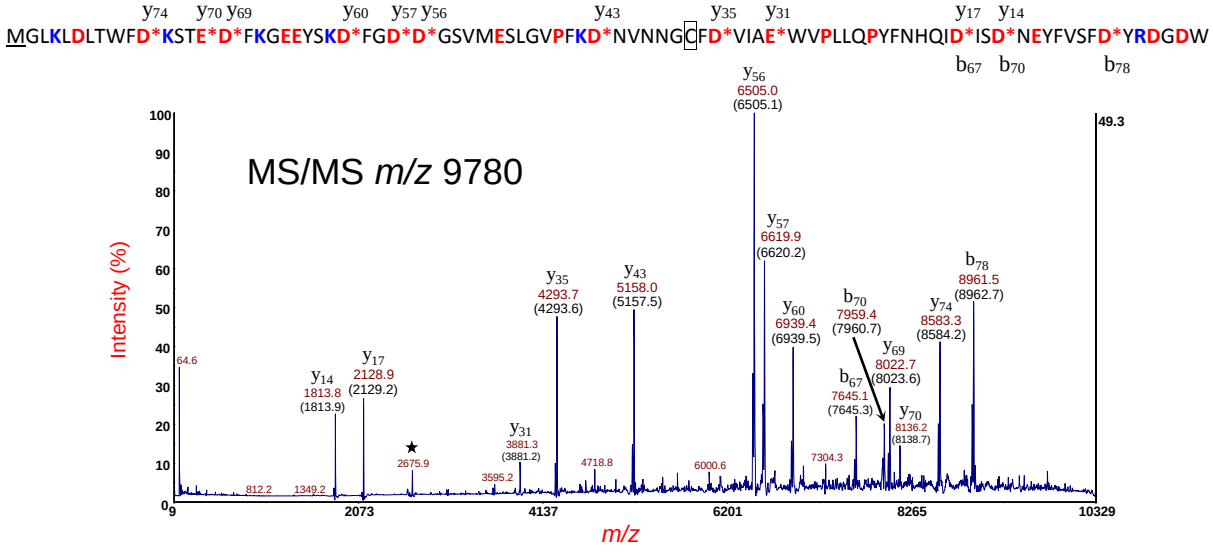


Figure

Figure 3

Click here to access/download;Figure;REVISED_Figure 3 Fagerquist.pptx





Figure

Click here to access/download;Figure;REVISED_Figure 4 Fagerquist.pptx

Figure 4

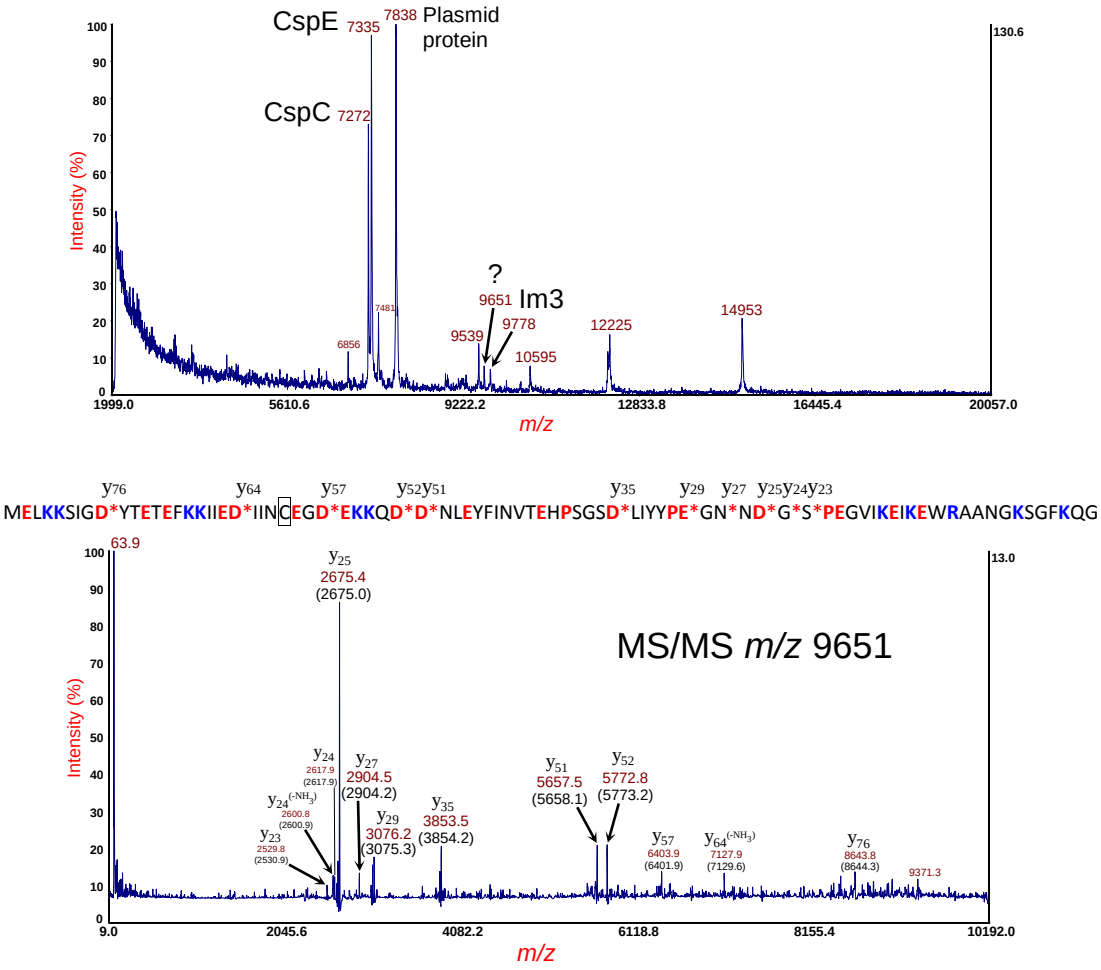


Figure 5

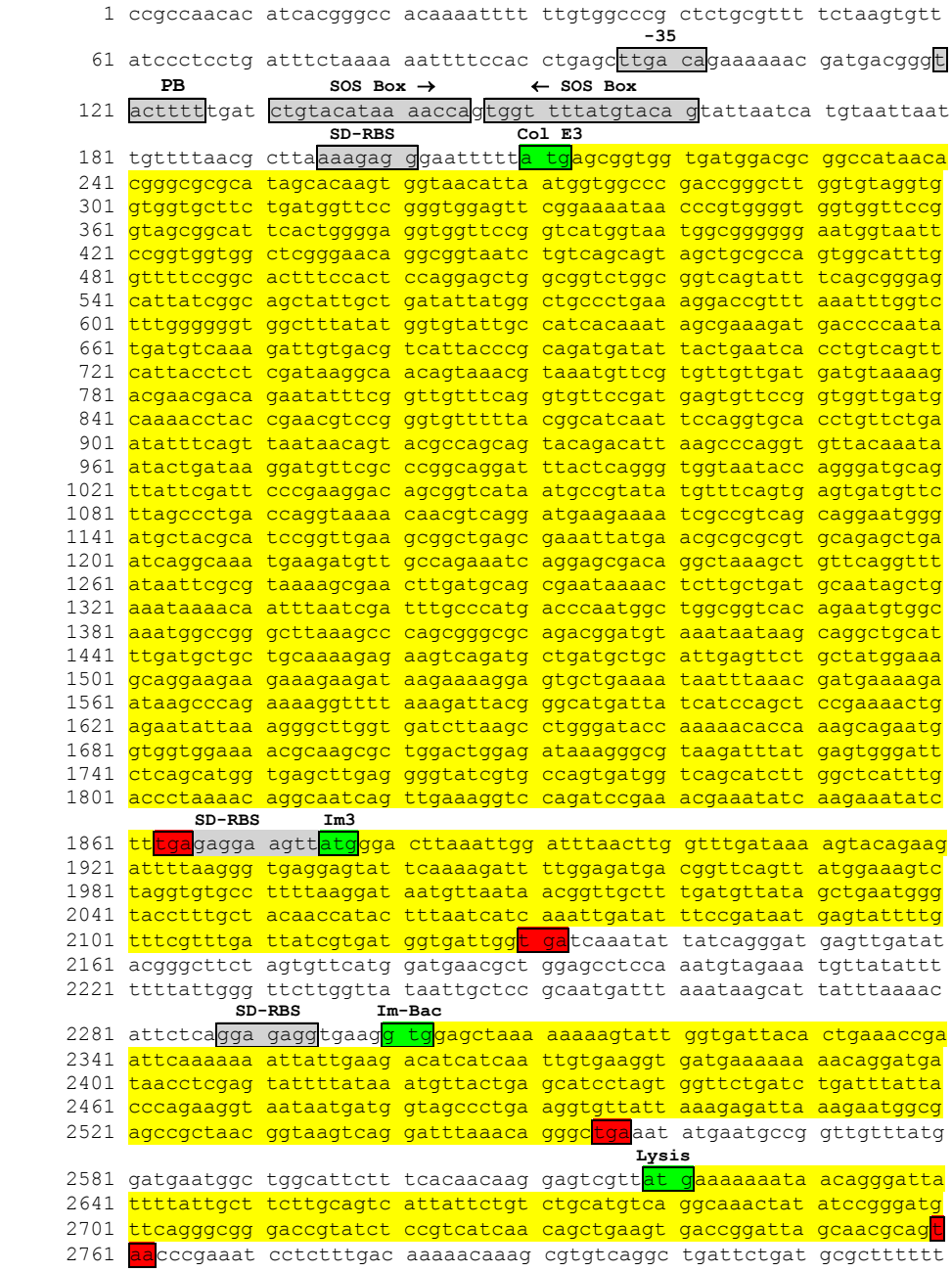


Table of Materials

Click here to access/download;Table of Materials;JoVE_Materials Fagerquist (1).xls

Name of Material/ Equipment	Company	Catalog Number	Comments/Description
4000 Series Explorer software	AB Sciex	Version 3.5.3	
4800 Plus MALDI TOF/TOF Analyzer	AB Sciex		
Acetonitrile Optima LC/MS grade	Fisher Chemical	A996-1	
BSL-2 biohazard cabinet	The Baker Company	SG403A-HE	
Cytochrome-C	Sigma	C2867-10MG	
Data Explorer software	AB Sciex	Version 4.9	
Focus Protein Reduction-Alkylation kit	G-Biosciences	786-231	
GPMAW software	Lighthouse Data	Version 10.0	
Incubator	VWR	9120973	
LB Agar	Invitrogen	22700-025	
Luria Broth	Invitrogen	12795-027	
Lysozyme	Sigma	L4919-1G	
Microcentrifuge Tubes, 2 mL, screw-cap, O-ring	Fisher Scientific	02-681-343	
MiniSpin Plus Centrifuge	Eppendorf	22620207	
Mitomycin-C (from streptomyces)	Sigma-Aldrich	M0440-5MG	
Myoglobin	Sigma	M5696-100MG	
Shaker MaxQ 420HP Model 420	Thermo Scientific	Model 420	
Sinapinic acid	Thermo Scientific	1861580	
Sterile 1 uL loops	Fisher Scientific	22-363-595	
Thioredoxin (E. coli, recombinant)	Sigma	T0910-1MG	
Trifluoroacetic acid	Sigma-Aldrich	299537-100G	
Water Optima LC/MS grade	Fisher Chemical	W6-4	

Editorial comments:

Changes to be made by the Author(s):

1. Please take this opportunity to thoroughly proofread the manuscript to ensure that there are no spelling or grammar issues.

Author Response: The manuscript has been proofread for spelling and grammatical issues.

2. Please format the manuscript as: paragraph Indentation: 0 for both left and right and special: none, Line spacings: single. Please include a single line space between each step, substep and note in the protocol section. Please use Calibri 12 points

Author Response: Manuscript has been formatted for these issues.

3. Please provide an email address for each author.

Author Response: Author e-mails have been provided.

4. Please rephrase the Short Abstract/Summary to clearly describe the protocol and its applications in complete sentences between 10-50 words: “Here we present a protocol …”

Author Response: The Summary has been revised.

5. Please ensure that the long Abstract is within 150-300-word limit and clearly states the goal of the protocol.

Author Response: The abstract was revised to clearly state the goal of the protocol.

6. JoVE cannot publish manuscripts containing commercial language. This includes trademark symbols (™), registered symbols (®), and company names before an instrument or reagent. Please remove all commercial language from your manuscript and use generic terms instead. All commercial products should be sufficiently referenced in the Table of Materials and Reagents. For example: Orbitrap, 4800 MALDI-TOF-TOF, etc.

Author Response: All commercial language has been removed.

7. Please ensure that all text in the protocol section is written in the imperative tense as if telling someone how to do the technique (e.g., “Do this,” “Ensure that,” etc.). The actions should be described in the imperative tense in complete sentences wherever possible. Avoid usage of phrases such as “could be,” “should be,” and “would be” throughout the Protocol. Any text that cannot be written in the imperative tense may be added as a “Note.”

Author Response: The protocol has been revised using the imperative tense.

8. The Protocol should contain only action items that direct the reader to do something.

Author Response: The protocol has been revised as suggested.

9. Please adjust the numbering of the Protocol to follow the JoVE Instructions for Authors. For example, 1 should be followed by 1.1 and then 1.1.1 and 1.1.2 if necessary. Please do not number caution and notes.

Author Response: The protocol has been revised as suggested.

10. The Protocol should be made up almost entirely of discrete steps without large paragraphs of text between sections. Please ensure that the Protocol step contains only 2-3 actions per step and a maximum of 4 sentences per step.

Author Response: The protocol has been revised as suggested.

11. Please convert centrifuge speeds to centrifugal force (x g) instead of revolutions per minute (rpm).

Author Response: Centrifuge speeds have been converted to g however the incubator/shaker is only provided in rpm.

12. Please add more details to your protocol steps. Please ensure you answer the "how" question, i.e., how is the step performed? For this please include mechanical action, button clicks in the software, knob turns, command lines, etc.

Author Response: More detail and clarity has been added to the protocol.

13. There is a 10-page limit for the Protocol, but there is a 3-page limit for filmable content. Please highlight 3 pages or less of the Protocol (including headings and spacing) that identifies the essential steps of the protocol for the video, i.e., the steps that should be visualized to tell the most cohesive story of the Protocol.

Author Response: Filmable content has been highlighted in yellow and is less than 3 pages.

14. Please obtain explicit copyright permission to reuse any figures from a previous publication. Explicit permission can be expressed in the form of a letter from the editor or a link to the editorial policy that allows re-prints. Please upload this information as a .doc or .docx file to your Editorial Manager account.

The Figure must be cited appropriately in the Figure Legend, i.e. “This figure has been modified from [citation].”

Author Response: All data and figures are original.

15. As we are a methods journal, please ensure that the Discussion explicitly cover the following in detail in 3-6 paragraphs with citations:
a) Critical steps within the protocol
b) Any modifications and troubleshooting of the technique
c) Any limitations of the technique
d) The significance with respect to existing methods
e) Any future applications of the technique

Author Response: Additional text has been added to the Discussion covering these topics under **Protocol considerations**.

16. We do not have a separate conclusion section. Please merge this with the discussion instead.

Author Response: Conclusions have been merged with the Discussion.

Reviewers' comments:
Reviewer #1:
Manuscript Summary:
The manuscript describes the protocol for the top-down analysis of extracellular E. coli proteins by MALDI-TOF-TOF-MS/MS. Additionally, the protocol for the analysis of this data using a standalone software, Protein Biomarker Seeker, for preliminary identification of immunity proteins induced by mitomycin-C is also given. Finally, 2 antibacterial immunity proteins were identified and manually confirmed.

Minor Concerns:
Protocol 1.3: I am assuming that extracellular protein are being analyzed here because the protocol does not specify a lysis procedure. If this is the case, perhaps a discussion of the proteins being probed with this method should be included.

Author Response: The reviewer is correct; the protocol does not employ any mechanical or electrical cell lysis step. Although secreted or extracellular proteins may be detected using the protocol, an earlier version of this method was first developed for detection of Shiga toxin (Stx) from Shiga toxin-producing *Escherichia coli* (STEC) wherein antibiotic induction triggers the bacterial SOS response resulting in expression of phage genes including *stx* as well as late phage genes responsible for *bacterial cell lysis* (Reference #41). We found that antibiotic-induced cell lysis has certain advantages for the detection of Stx as well as plasmid proteins that have SOS promoters (current work). Certainly, mechanical cell lysis (e.g. bead-beating) can also be used (although not used in the current work). However, mechanical lysis results in all bacterial cells being lysed (not just induced cells) resulting in the sample being enriched

with abundant, highly conserved host proteins that can make detection of phage and plasmid proteins from an unfractionated sample more challenging.

Protocol 4.7: The PTMs involving sequence truncation are discussed extensively and the response of the software to these modifications is clear. However, while E. coli does have limited PTMs associated with the addition of chemical or complex groups, they exist. How would the software respond to these kinds of modifications?

Author Response: The reviewer raises an important point. In addition to sequence truncation, *E. coli* proteins can have PTMs that involve addition of mass, e.g. phosphorylation, glycosylation, etc. As MS/MS utilizes post-source decay (PSD) for dissociation of singly charged metastable proteins ions (under 20 kDa in mass) generated by MALDI, such PTMs attached to residue side-chains would likely undergo facile dissociative loss because PSD is an ergodic dissociation technique. The presence of such PTMs may be inferred from the appearance of a fragment ion close in mass to the original precursor ion (minus the mass of the PTM) in the MS/MS data. However, neither PSD nor the software would be able to identify *where* such PTMs are attached. In addition, the software can only identify proteins from polypeptide backbone cleavage fragment ions and not dissociative loss of small molecules (water or ammonia) or PTMs attached at residue side-chains. However, if fragment ions from polypeptide backbone cleavage are detected, the protein could still be identified using the software by either widening the protein mass tolerance window to include the mass of the PTM or simply entering the mass of the protein fragment ion corresponding to dissociative loss of the suspected PTM. Any identification by the software would be of the protein sequence but without the additional mass of the PTM.

Interestingly, we have not detected proteins having phosphorylation, glycosylation, etc. in our bacterial work thus far. However, that may be due to 1. their abundance as ionized by MALDI; 2. the MS mass range being used (less than 20 kDa); 3. these PTMs may be unusually labile and may not survive application of MALDI matrix which is a saturated solution of sinapinic acid at low pH; 4. these PTMs may undergo very rapid dissociative loss *in the source* before ions are accelerated from the source.

Protocol 4.8.5: The software allows the user to state the oxidation state of the cysteine residues. If the cysteine residues are reduced, it is common practice to protect the residues from oxidation using alkylating agents. Can the software account for cysteine alkylation? Additionally, if the cysteine residues are not reduced, it would be possible for both reduced and oxidized cysteine residues to exist simultaneously.

Author Response: At the present time, the software does not include cysteine alkylation, and our sample protocol does not include a disulfide reduction step for cysteine residues. The protocol has been clarified to indicate that the search is to be operated with cysteine residues in their Oxidized state, and if no identification is obtained to execute the search again with cysteine residues in their Reduced state. If no identifications are found again, widening the fragment ion tolerance to ± 2 or ± 3 lowers the threshold for fragment ion matching allowing sequences with cysteines to be matched whether they are present in their oxidized and/or reduced states.

Lines 288-291: Where replicate runs conducted to determine the variation in relative protein expression under the individual antibiotic concentration conditions?

Author Response: With respect to the antibiotic-induced proteins detected, the antibiotic concentrations for a bacterial strain were found to be generally reproducible. We noted variations in the relative protein abundance. Since our analysis is qualitative (not quantitative), protein biomarker abundance need only be sufficient for adequate MS/MS analysis. A putative STEC strain is first cultured with a range of antibiotic concentrations (e.g. 300 ng/mL to 2000 ng/mL of mitomycin-C) to determine the optimum concentration such that it triggers the bacterial SOS response while still providing enough bacterial cells for harvesting. For the STEC strain RM7788, we found that the optimum antibiotic concentration for detection of the biomarkers identified was 400 to 800 ng/mL of mitomycin-C.

Overall comment: The author does a good job in the conclusion of discussing the strengths and weaknesses of the use of MALDI-TOF for top-down proteomics analysis. However, I think the manuscript would also benefit from a discussion of why this particular analysis is best performed using top-down techniques generally. Usually the ability of top-down to identify unique proteoforms is the reason given for its use. In this case it seems only protein truncation can be identified, so what is the advantage of top-down over bottom-up techniques?

Author Response: The primary strength of the current protocol is its speed, simplicity of sample preparation and use of an instrument that is relatively easy to operate, be trained on and maintain. A bottom-up approach involving enzymatic digestion of a bacterial cell lysate or supernatant will result in a more complex sample that is best analyzed by LC-ESI using a high-resolution mass spectrometer (HR-MS). Although bottom-up and top-down proteomic analysis by LC-ESI-HR-MS are ubiquitous and far superior in many respects to top-down by MALDI-TOF-TOF, it requires more time, labor, expense, and expertise. Much of the work in our laboratory is interdisciplinary: spanning analytical mass spectrometry to microbiology. Instrument complexity can often affect whether certain instrument platforms are likely to be adopted by scientists not formally trained in mass spectrometry. The top-down approach with MALDI-TOF-TOF is meant to extend the analysis of MALDI-TOF-MS beyond its current use for taxonomic identification of bacteria in clinical microbiology labs while not dramatically increasing the labor, complexity or expertise required for analysis.

Reviewer #2:

The manuscripts titled "Antibacterial immunity proteins produced by pathogenic E. coli are identified using MALDI-TOF-TOF-MS/MS and top-down proteomic analysis", identifies two plasmid encoded AB-Im proteins produced by E. coli. The authors used top town proteomic analysis using standalone software developed to process sequences developed using whole genome sequencing analysis in addition to MALDI TOF -MS/MS. In case post translational modification is detected, the authors also encounter for these modifications in the software they are using. Gene location for the identified proteins had also been suggested.

The manuscript needs some sentence revision and work on the representation of the figures for clarity.

The manuscript is well written however minor sentence edits should be considered.
Line 108 -add LB in brackets

Author Response: (LB) was added.

Line 121-144 needs grammar sentence revision

Author Response: These lines have been revised using the imperative tense and corrected for grammar.

Line 230 ..is not to definitely identify- consider changing to identify with 100% accuracy.

Author Response: The sentence has been rephrased to the following: "The purpose of the *Protein Biomarker Seeker* software is to identify with high accuracy a protein sequence by eliminating from consideration many obviously incorrect protein sequences as well as to incorporate sequence truncation as a possible PTM in the mature protein."

Figures should be presented with better resolution and merging of few figures is suggested.

Author Response: Figures 1-3 have been combined into a single figure (now Figure 1) with some additional screen shots of the software to clarify the protocol. Figures 4-5 have been combined into a single figure (now Figure 2). It was not possible to increase the resolution of screen shots of the software.

S1 Im3: Ten analyses with fragment ions of Im3

```
$ java -jar ProteinMassID.jar --benchmark

Please input the protein mass (Da):
9780

Please input the protein tolerance (Da):
10

Please enter the desired fragment ion tolerance for this benchmark:
1.5

Please enter fragment ion M/Z values, delimited by commas:
1813.8,2128.9,3881.3,4293.7,5158.0,6505.0,6619.9,6939.4,7645.1,7959.4,8022.7,8136.2,8583.3,8961.5

Please enter the desired input fasta file path:
C:\5008_RM7788.txt

Please enter the number of tables desired:
10

Please specify the desired benchmark oxidation state (o/r):
r

Beginning benchmark.
```

Table # 1 of randomly selected fragment ions (m/z A, B, C)

```
The randomly selected fragment ions were:
A: 1813.8
B: 8022.7
C: 8136.2
```

	1_A,B,C	2_AB,BC,AC	3_ABC	(#of fragment ions)
D:	551,544,580	2,19,3	0	
E:	597,569,577	0,1,1	0	
N:	402,432,427	0,12,1	0	
D+E:	1148,1113,1154	3,34,7	0	
D+E+N:	1549,1543,1576	5,99,13	0	

Table # 2 of randomly selected fragment ions (m/z A, B, C)

The randomly selected fragment ions were:
A: 4293.7
B: 8022.7
C: 8583.3

	1_A,B,C	2_AB,BC,AC	3_ABC	(#of fragment ions)
D:	547,544,542	3,5,3	1	
E:	568,569,582	4,1,5	0	
N:	414,432,444	1,0,2	0	
D+E:	1113,1113,1124	13,9,10	1	
D+E+N:	1527,1543,1567	16,13,20	1	

Table # 3 of randomly selected fragment ions (m/z A, B, C)

The randomly selected fragment ions were:
A: 3881.3
B: 6939.4
C: 8961.5

	1_A,B,C	2_AB,BC,AC	3_ABC	(#of fragment ions)
D:	531,509,555	1,2,2	0	
E:	594,544,598	2,2,1	0	
N:	386,458,401	0,1,0	0	
D+E:	1124,1052,1153	8,10,7	1	
D+E+N:	1507,1510,1554	14,17,15	1	

Table # 4 of randomly selected fragment ions (m/z A, B, C)

The randomly selected fragment ions were:
A: 1813.8
B: 2128.9
C: 6619.9

	1_A,B,C	2_AB,BC,AC	3_ABC	(#of fragment ions)
D:	551,595,521	2,1,2	1	
E:	597,570,583	6,2,1	0	
N:	402,417,391	3,0,0	0	
D+E:	1148,1163,1104	12,7,5	1	
D+E+N:	1549,1580,1495	23,13,9	1	

Table # 5 of randomly selected fragment ions (m/z A, B, C)

The randomly selected fragment ions were:
A: 1813.8
B: 3881.3
C: 5158.0

	1_A,B,C	2_AB,BC,AC	3_ABC	(#of fragment ions)
D:	551,531,525	2,1,3	0	
E:	597,594,587	3,3,2	0	
N:	402,386,416	0,1,1	0	
D+E:	1148,1124,1112	11,7,10	1	
D+E+N:	1549,1507,1527	18,9,18	1	

Table # 6 of randomly selected fragment ions (m/z A, B, C)

The randomly selected fragment ions were:
A: 1813.8
B: 6939.4
C: 7959.4

	1_A,B,C	2_AB,BC,AC	3_ABC	(#of fragment ions)
D:	551,509,570	3,2,69	1	
E:	597,544,621	4,0,76	0	
N:	402,458,389	1,1,57	0	
D+E:	1148,1052,1190	7,6,146	1	
D+E+N:	1549,1510,1579	17,11,206	1	

Table # 7 of randomly selected fragment ions (m/z A, B, C)

The randomly selected fragment ions were:
A: 4293.7
B: 5158.0
C: 8961.5

	1_A,B,C	2_AB,BC,AC	3_ABC	(#of fragment ions)
D:	547,525,555	2,3,2	1	
E:	568,587,598	1,1,4	0	
N:	414,416,401	1,2,0	0	
D+E:	1113,1112,1153	7,7,7	1	

D+E+N:	1527,1527,1554	11,16,12	1

Table # 8 of randomly selected fragment ions (m/z A, B, C)

The randomly selected fragment ions were:
A: 6619.9
B: 6939.4
C: 8136.2

	1_A,B,C	2_AB,BC,AC	3_ABC	(#of fragment ions)
D:	521,509,580	1,1,1	0	
E:	583,544,577	1,2,2	0	
N:	391,458,427	2,1,0	0	
D+E:	1104,1052,1154	5,8,6	0	
D+E+N:	1495,1510,1576	10,20,11	0	

Table # 9 of randomly selected fragment ions (m/z A, B, C)

The randomly selected fragment ions were:
A: 1813.8
B: 6939.4
C: 8961.5

	1_A,B,C	2_AB,BC,AC	3_ABC	(#of fragment ions)
D:	551,509,555	3,2,3	1	
E:	597,544,598	4,2,1	0	
N:	402,458,401	1,1,2	0	
D+E:	1148,1052,1153	7,10,5	1	
D+E+N:	1549,1510,1554	17,17,11	1	

Table # 10 of randomly selected fragment ions (m/z A, B, C)

The randomly selected fragment ions were:
A: 1813.8
B: 6619.9
C: 8022.7

	1_A,B,C	2_AB,BC,AC	3_ABC	(#of fragment ions)
D:	551,521,544	2,2,2	1	
E:	597,583,569	1,2,0	0	

```
N:      402,391,432          0,0,0                0
D+E:    1148,1104,1113       5,7,3                1
D+E+N:  1549,1495,1543       9,8,5                1

Benchmark complete.
```

S2 ImBac: Ten analyses with fragment ions of ImBac

```
$ java -jar ProteinMassID.jar --benchmark

Please input the protein mass (Da):
9650

Please input the protein tolerance (Da):
10

Please enter the desired fragment ion tolerance for this benchmark:
1.5

Please enter fragment ion M/Z values, delimited by commas:
2675.4,2904.5,3076.2,3853.5,5657.5,5772.8

Please enter the desired input fasta file path:
C:\5008_RM7788.txt

Please enter the number of tables desired:
10

Please specify the desired benchmark oxidation state (o/r):
r

Beginning benchmark.
```

Table # 1 of randomly selected fragment ions (m/z A, B, C)

```
The randomly selected fragment ions were:
A: 2675.4
B: 3076.2
C: 5657.5
       1_A,B,C               2_AB,BC,AC            3_ABC                  (#of fragment ions)
-------------------------------------------------------------------------------------------
D:     551,505,574           0,1,5                 0
E:     569,619,607           3,2,2                 0
N:     455,389,454           0,2,0                 0
D+E:   1119,1124,1178        6,8,10                1
D+E+N: 1573,1511,1630        10,14,20              1
```

Table # 2 of randomly selected fragment ions (m/z A, B, C)

The randomly selected fragment ions were:
A: 2904.5
B: 3076.2
C: 5657.5

	1_A,B,C	2_AB,BC,AC	3_ABC	(#of fragment ions)
D:	533,505,574	4,1,0	0	
E:	626,619,607	1,2,2	0	
N:	416,389,454	1,2,0	0	
D+E:	1159,1124,1178	8,8,4	0	
D+E+N:	1574,1511,1630	23,14,9	1	

Table # 3 of randomly selected fragment ions (m/z A, B, C)

The randomly selected fragment ions were:
A: 3076.2
B: 5657.5
C: 5772.8

	1_A,B,C	2_AB,BC,AC	3_ABC	(#of fragment ions)
D:	505,574,515	1,28,2	0	
E:	619,607,621	2,1,2	0	
N:	389,454,418	2,17,1	1	
D+E:	1124,1178,1135	8,61,8	1	
D+E+N:	1511,1630,1552	14,134,10	2	

Table # 4 of randomly selected fragment ions (m/z A, B, C)

The randomly selected fragment ions were:
A: 2675.4
B: 2904.5
C: 3853.5

	1_A,B,C	2_AB,BC,AC	3_ABC	(#of fragment ions)
D:	551,533,564	9,0,3	0	
E:	569,626,609	2,5,2	0	
N:	455,416,400	2,1,0	0	
D+E:	1119,1159,1173	22,11,7	0	
D+E+N:	1573,1574,1572	43,16,12	1	

Table # 5 of randomly selected fragment ions (m/z A, B, C)

The randomly selected fragment ions were:
A: 2675.4
B: 5657.5
C: 5772.8

	1_A,B,C	2_AB,BC,AC	3_ABC	(#of fragment ions)
D:	551,574,515	5,28,3	2	
E:	569,607,621	2,1,0	0	
N:	455,454,418	0,17,3	0	
D+E:	1119,1178,1135	10,61,7	2	
D+E+N:	1573,1630,1552	20,134,15	3	

Table # 6 of randomly selected fragment ions (m/z A, B, C)

The randomly selected fragment ions were:
A: 2904.5
B: 3076.2
C: 5772.8

	1_A,B,C	2_AB,BC,AC	3_ABC	(#of fragment ions)
D:	533,505,515	4,2,1	0	
E:	626,619,621	1,2,2	0	
N:	416,389,418	1,1,0	0	
D+E:	1159,1124,1135	8,8,7	0	
D+E+N:	1574,1511,1552	23,10,14	1	

Table # 7 of randomly selected fragment ions (m/z A, B, C)

The randomly selected fragment ions were:
A: 2675.4
B: 2904.5
C: 3853.5

	1_A,B,C	2_AB,BC,AC	3_ABC	(#of fragment ions)
D:	551,533,564	9,0,3	0	
E:	569,626,609	2,5,2	0	
N:	455,416,400	2,1,0	0	
D+E:	1119,1159,1173	22,11,7	0	
D+E+N:	1573,1574,1572	43,16,12	1	

Table # 8 of randomly selected fragment ions (m/z A, B, C)

The randomly selected fragment ions were:
A: 3076.2
B: 5657.5
C: 5772.8

	1_A,B,C	2_AB,BC,AC	3_ABC	(#of fragment ions)
D:	505,574,515	1,28,2	0	
E:	619,607,621	2,1,2	0	
N:	389,454,418	2,17,1	1	
D+E:	1124,1178,1135	8,61,8	1	
D+E+N:	1511,1630,1552	14,134,10	2	

Table # 9 of randomly selected fragment ions (m/z A, B, C)

The randomly selected fragment ions were:
A: 2675.4
B: 2904.5
C: 5772.8

	1_A,B,C	2_AB,BC,AC	3_ABC	(#of fragment ions)
D:	551,533,515	9,1,3	0	
E:	569,626,621	2,2,0	0	
N:	455,416,418	2,0,3	0	
D+E:	1119,1159,1135	22,7,7	0	
D+E+N:	1573,1574,1552	43,14,15	2	

Table # 10 of randomly selected fragment ions (m/z A, B, C)

The randomly selected fragment ions were:
A: 2675.4
B: 2904.5
C: 5657.5

	1_A,B,C	2_AB,BC,AC	3_ABC	(#of fragment ions)
D:	551,533,574	9,0,5	0	
E:	569,626,607	2,2,2	0	
N:	455,416,454	2,0,0	0	
D+E:	1119,1159,1178	22,4,10	0	
D+E+N:	1573,1574,1630	43,9,20	1	

Benchmark complete.

1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

Title of Article:	Antibacterial immunity proteins produced by pathogenic Escherichia coli are identified using MALDI-TOF-TOF-MS/MS and top-down proteomic analysis
Author(s):	Clifton K. Fagerquist* and Ernesto Rojas

Item 1: The Author elects to have the Materials be made available (as described at http://www.jove.com/publish) via:

☑ Standard Access ☐ Open Access

Item 2: Please select one of the following items:

☐ The Author is **NOT** a United States government employee.

☑ The Author is a United States government employee and the Materials were prepared in the course of his or her duties as a United States government employee.

ARTICLE AND VIDEO LICENSE AGREEMENT

1. **Defined Terms.** As used in this Article and Video License Agreement, the following terms shall have the following meanings: "**Agreement**" means this Article and Video License Agreement; "**Article**" means the article specified on the last page of this Agreement, including any associated materials such as texts, figures, tables, artwork, abstracts, or summaries contained therein; "**Author**" means the author who is a signatory to this Agreement; "**Collective Work**" means a work, such as a periodical issue, anthology or encyclopedia, in which the Materials in their entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole; "**CRC License**" means the Creative Commons Attribution-Non Commercial-No Derivs 3.0 Unported Agreement, the terms and conditions of which can be found at: http://creativecommons.org/licenses/by-nc-nd/3.0/legalcode; "**Derivative Work**" means a work based upon the Materials or upon the Materials and other pre-existing works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Materials may be recast, transformed, or adapted; "**Institution**" means the institution, listed on the last page of this Agreement, by which the Author was employed at the time of the creation of the Materials; "**JoVE**" means MyJove Corporation, a Massachusetts corporation and the publisher of The Journal of Visualized Experiments; "**Materials**" means the Article and / or the Video; "**Parties**" means the Author and JoVE; "**Video**" means any video(s) made by the Author, alone or in conjunction with any other parties, or by JoVE or its affiliates or agents, individually or in collaboration with the Author or any other parties, incorporating all or any portion of the Article, and in which the Author may or may not appear.

2. **Background.** The Author, who is the author of the Article, in order to ensure the dissemination and protection of the Article, desires to have the JoVE publish the Article and create and transmit videos based on the Article. In furtherance of such goals, the Parties desire to memorialize in this Agreement the respective rights of each Party in and to the Article and the Video.

3. **Grant of Rights in Article.** In consideration of JoVE agreeing to publish the Article, the Author hereby grants to JoVE, subject to **Sections 4** and **7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Article in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Article into other languages, create adaptations, summaries or extracts of the Article or other Derivative Works (including, without limitation, the Video) or Collective Works based on all or any portion of the Article and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and(c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. If the "Open Access" box has been checked in **Item 1** above, JoVE and the Author hereby grant to the public all such rights in the Article as provided in, but subject to all limitations and requirements set forth in, the CRC License.

4. **Retention of Rights in Article.** Notwithstanding the exclusive license granted to JoVE in **Section 3** above, the Author shall, with respect to the Article, retain the non-exclusive right to use all or part of the Article for the non-commercial purpose of giving lectures, presentations or teaching classes, and to post a copy of the Article on the Institution's website or the Author's personal website, in each case provided that a link to the Article on the JoVE website is provided and notice of JoVE's copyright in the Article is included. All non-copyright intellectual property rights in and to the Article, such as patent rights, shall remain with the Author.

5. **Grant of Rights in Video – Standard Access.** This **Section 5** applies if the "Standard Access" box has been checked in **Item 1** above or if no box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby acknowledges and agrees that, Subject to **Section 7** below, JoVE is and shall be the sole and exclusive owner of all rights of any nature, including, without limitation, all copyrights, in and to the Video. To the extent that, by law, the Author is deemed, now or at any time in the future, to have any rights of any nature in or to the Video, the Author hereby disclaims all such rights and transfers all such rights to JoVE.

6. **Grant of Rights in Video – Open Access.** This **Section 6** applies only if the "Open Access" box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby grants to JoVE, subject to **Section 7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Video in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Video into other languages, create adaptations, summaries or extracts of the Video or other Derivative Works or Collective Works based on all or any portion of the Video and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. For any Video to which this **Section 6** is applicable, JoVE and the Author hereby grant to the public all such rights in the Video as provided in, but subject to all limitations and requirements set forth in, the CRC License.

7. **Government Employees.** If the Author is a United States government employee and the Article was prepared in the course of his or her duties as a United States government employee, as indicated in **Item 2** above, and any of the licenses or grants granted by the Author hereunder exceed the scope of the 17 U.S.C. 403, then the rights granted hereunder shall be limited to the maximum rights permitted under such statute. In such case, all provisions contained herein that are not in conflict with such statute shall remain in full force and effect, and all provisions contained herein that do so conflict shall be deemed to be amended so as to provide to JoVE the maximum rights permissible within such statute.

8. **Protection of the Work.** The Author(s) authorize JoVE to take steps in the Author(s) name and on their behalf if JoVE believes some third party could be infringing or might infringe the copyright of either the Author's Article and/or Video.

9. **Likeness, Privacy, Personality.** The Author hereby grants JoVE the right to use the Author's name, voice, likeness, picture, photograph, image, biography and performance in any way, commercial or otherwise, in connection with the Materials and the sale, promotion and distribution thereof. The Author hereby waives any and all rights he or she may have, relating to his or her appearance in the Video or otherwise relating to the Materials, under all applicable privacy, likeness, personality or similar laws.

10. **Author Warranties.** The Author represents and warrants that the Article is original, that it has not been published, that the copyright interest is owned by the Author (or, if more than one author is listed at the beginning of this Agreement, by such authors collectively) and has not been assigned, licensed, or otherwise transferred to any other party. The Author represents and warrants that the author(s) listed at the top of this Agreement are the only authors of the Materials. If more than one author is listed at the top of this Agreement and if any such author has not entered into a separate Article and Video License Agreement with JoVE relating to the Materials, the Author represents and warrants that the Author has been authorized by each of the other such authors to execute this Agreement on his or her behalf and to bind him or her with respect to the terms of this Agreement as if each of them had been a party hereto as an Author. The Author warrants that the use, reproduction, distribution, public or private performance or display, and/or modification of all or any portion of the Materials does not and will not violate, infringe and/or misappropriate the patent, trademark, intellectual property or other rights of any third party. The Author represents and warrants that it has and will continue to comply with all government, institutional and other regulations, including, without limitation all institutional, laboratory, hospital, ethical, human and animal treatment, privacy, and all other rules, regulations, laws, procedures or guidelines, applicable to the Materials, and that all research involving human and animal subjects has been approved by the Author's relevant institutional review board.

11. **JoVE Discretion.** If the Author requests the assistance of JoVE in producing the Video in the Author's facility, the Author shall ensure that the presence of JoVE employees, agents or independent contractors is in accordance with the relevant regulations of the Author's institution. If more than one author is listed at the beginning of this Agreement, JoVE may, in its sole

discretion, elect not take any action with respect to the Article until such time as it has received complete, executed Article and Video License Agreements from each such author. JoVE reserves the right, in its absolute and sole discretion and without giving any reason therefore, to accept or decline any work submitted to JoVE. JoVE and its employees, agents and independent contractors shall have full, unfettered access to the facilities of the Author or of the Author's institution as necessary to make the Video, whether actually published or not. JoVE has sole discretion as to the method of making and publishing the Materials, including, without limitation, to all decisions regarding editing, lighting, filming, timing of publication, if any, length, quality, content and the like.

12. **Indemnification.** The Author agrees to indemnify JoVE and/or its successors and assigns from and against any and all claims, costs, and expenses, including attorney's fees, arising out of any breach of any warranty or other representations contained herein. The Author further agrees to indemnify and hold harmless JoVE from and against any and all claims, costs, and expenses, including attorney's fees, resulting from the breach by the Author of any representation or warranty contained herein or from allegations or instances of violation of intellectual property rights, damage to the Author's or the Author's institution's facilities, fraud, libel, defamation, research, equipment, experiments, property damage, personal injury, violations of institutional, laboratory, hospital, ethical, human and animal treatment, privacy or other rules, regulations, laws, procedures or guidelines, liabilities and other losses or damages related in any way to the submission of work to JoVE, making of videos by JoVE, or publication in JoVE or elsewhere by JoVE. The Author shall be responsible for, and shall hold JoVE harmless from, damages caused by lack of sterilization, lack of cleanliness or by contamination due to the making of a video by JoVE its employees, agents or independent contractors. All sterilization, cleanliness or decontamination procedures shall be solely the responsibility of the Author and shall be undertaken at the Author's expense. All indemnifications provided herein shall include JoVE's attorney's fees and costs related to said losses or damages. Such indemnification and holding harmless shall include such losses or damages incurred by, or in connection with, acts or omissions of JoVE, its employees, agents or independent contractors.

13. **Fees.** To cover the cost incurred for publication, JoVE must receive payment before production and publication of the Materials. Payment is due in 21 days of invoice. Should the Materials not be published due to an editorial or production decision, these funds will be returned to the Author. Withdrawal by the Author of any submitted Materials after final peer review approval will result in a US$1,200 fee to cover pre-production expenses incurred by JoVE. If payment is not received by the completion of filming, production and publication of the Materials will be suspended until payment is received.

14. **Transfer, Governing Law.** This Agreement may be assigned by JoVE and shall inure to the benefits of any of JoVE's successors and assignees. This Agreement shall be governed and construed by the internal laws of the Commonwealth of Massachusetts without giving effect to any conflict of law provision thereunder. This Agreement may be executed in counterparts, each of which shall be deemed an original, but all of which together shall be deemed to me one and the same agreement. A signed copy of this Agreement delivered by facsimile, e-mail or other means of electronic transmission shall be deemed to have the same legal effect as delivery of an original signed copy of this Agreement.