Supporting Information

Rapid Identification of Protein Biomarkers of *E. coli* O157:H7 by Top-Down Proteomics and MALDI-TOF-TOF Mass Spectrometry

Clifton K. Fagerquist[1]*, Brandon R. Garbus[1], William G. Miller[1], Katherine E. Williams[2], Emma Yee[1], Anna H. Bates[1], Síobhán Boyle[1], Leslie A. Harden[1], Michael B. Cooley[1], Robert E. Mandrell[1]

[1]*Western Regional Research Center, Agricultural Research Service, U. S. Department of Agriculture, 800 Buchanan Street, Albany, California 94710.*

[2]*University of California, San Francisco, School of Medicine, Obstetrics, Department of Gynecology and Reproductive Services, 521 Parnassus, San Francisco, California 94143.*

Running title: Protein biomarkers of *E. coli* O157:H7 identified by top-down proteomics

---

**Correspondence to*: C. K. Fagerquist, Western Regional Research Center, Agricultural Research Service, U. S. Department of Agriculture, 800 Buchanan Street, Albany, CA 94710, USA. E-mail: clifton.fagerquist@ars.usda.gov.

Results and Discussion

E. coli O157:H7 strain RM1272 (EDL933)

Table S-5A shows the top identifications of a protein biomarker ion at *m/z* 7272.8 [S1] (and at *m/z* 7271.7 in Figure 1) from *E. coli* O157:H7 strain RM1272 (EDL933) analyzed by MS/MS and top-down proteomics. The top scoring identification is the cold shock-like protein: CspC. The amino acid sequence of CspC sequence is highly conserved across many microorganisms as shown in Table S-5A, in consequence, it is not possible to exploit this protein as a biomarker for microorganism identification. Interestingly, further down the list of protein identifications is the 50S ribosomal protein L29 of *E. coli* O157:H7.

Table S-5B shows the top identifications of the same protein biomarker ion when MS/MS fragment ions are compared to D-, E-, P-specific *in silico* fragment ions. Once again the top scoring identification is CspC, however its identification score is now relatively more significant compared with the non-residue-specific *in silico* fragment ion comparison. In addition, the p-value score for 50S ribosomal protein L29 is now ranked as the top "runner-up". Figure S-3 shows the amino acid sequences of CspC and 50S ribosomal protein L29 of *E. coli* O157:H7 strain EDL933 which shows that the MW of the mature proteins are different by only ~ 2 Da. It is likely that the packet of protein ions which constitute the peak at *m/z* ~ 7273 are comprised of both CspC and 50S ribosomal L29 protein ions. The contribution of fragment ions from each protein is dependent on the relative abundance of the protein (copy number per cell), its ionization efficiency by MALDI and the fragmentation efficiency of the protein ion.

A peak at *m/z* 9947.6 was analyzed by MS/MS and found to produce a spectrum similar to the MS/MS spectrum of the peak at *m/z* 9737.5 in Figure 1.[S1] As previously noted, the peak at *m/z* 9737.5 in Figure 1 was identified as the acid stress chaperone-like protein HdeA (Table 2). The peak at *m/z* 9947.6 appeared when sinapinic acid was used as the matrix but was absent when using HCCA matrix. It was conjectured that sinapinic acid may form a covalent adduct to the HdeA. Proceeding on this assumption, Table S-6A shows the top identifications of the protein biomarker ion at *m/z* 9947.6 analyzed by MS/MS and top-down proteomics using a wider than normal protein MW tolerance of ± 300 Da. The purpose of the wider protein MW tolerance was to include in the comparison *in silico* fragment ions from HdeA (as well as other proteins). As shown in Table S-6A, the top scoring identifications do not include either HdeA or any *E.*

coli strains. The same analysis was then repeated with MS/MS fragment ions compared to D-, E-, P-specific *in silico* fragment ions (Table S-6B). HdeA was the 2nd, 3rd and 4th highest scoring identification by the USDA score whose amino acid sequence is conserved by *E. coli* O157:H7 strain EDL933, *E. coli* strain K-12 and *Shigella flexneri*. Finally, the same analysis was repeated against only D-specific *in silico* fragment ions (Table S-6C). The top scoring identification of both the USDA score and the p-value calculation is HdeA protein although its score is not very "significant". However, this is probably due to a number of the MS/MS fragment ions having a *m/z* that is shifted due to sinapinic acid attachment and, in consequence, not being matched to the correct *in silico* fragment ion. These results strongly suggest that the peak at *m/z* 9947.6 is HdeA with a covalently attached sinapinic acid. Additional evidence and discussion of this hypothesis is provided in another report.[S1]

E. coli O157:H7 strain RM5603

Table S-8A shows the top identification scores of a protein biomarker ion at *m/z* 8323.6 (Figure S-2) from *E. coli* O157:H7 strain RM5603 analyzed by MS/MS and top-down proteomics. A higher than normal intensity threshold cutoff was used (6%) for the analysis to eliminate much of the noise that appeared in the MS/MS spectrum of *m/z* 8323.6. YjbJ of *E. coli* O157:H7 ranked fourth in the identification by the p-value algorithm for the all *in silico* ion comparison. Table S-8B displays the results when MS/MS fragment ions were compared to D-, E-, P-specific *in silico* fragment ions. The top identification now includes YjbJ of *E. coli* O157:H7 strain RM5603 as well as many other *E. coli* and *Shigella* strains. YjbJ is also identified as CsbD family protein. The top USDA and p-value scores are consistent and both are significant compared to the runner-up identifications. Thus, residue-specific analysis assisted in correctly identifying the protein biomarker (and its source microorganism) from MS/MS data having excessive chemical or electronic noise. Table S-8C displays the results when MS/MS fragment ions were compared to only D-specific *in silico* fragment ions. Once again, the top identification is YjbJ (or CsbD family protein), however the relative difference between the top identification and the runner up identification scores has increased even further which indicates the "significance" of the top identification.

Table S-9A shows the top scoring identifications of a protein biomarker ion at *m/z* 9737.8 (Figure S-2) from *E. coli* O157:H7 strain RM5603 analyzed by MS/MS and top-down

proteomics. The top identification of both the USDA peak matching and p-value algorithms is the acid stress chaperone-like protein HdeA of *E. coli* O157:H7 strain RM5603, however the amino acid sequence homology of HdeA is also shared with other *E. coli* serotypes and strains as well as with *Shigella flexnari*. In consequence, the HdeA sequence, by itself, is not sufficiently unique to identify the microorganism. Tables S-9B & 9C compare MS/MS fragment ions to D-, E-, P-specific *in silico* fragment ions and D-specific *in silico* fragment ions, respectively. A relative enhancement of the top identification score is observed as the residue specificity of *in silico* fragment ions is narrowed from all *in silico* fragment ions to D-, E-, P-specific *in silico* fragment ions to D-specific *in silico* fragment ions.

Non-O157:H7 *E. coli* strain RM3061

Table S-11A shows the top scoring identifications of a protein biomarker ion at *m/z* 10461.8 [S1] (and at *m/z* 10458.3 in Figure 4) from non-O157:H7, non-pathogenic *E. coli* strain RM3061 analyzed by MS/MS and top-down analysis. The top scoring identification is YbgS (or homeobox) of *E. coli* strain RM3061. The amino acid sequence of HdeA of RM3061 appears to be sufficiently unique such that it can be used to discriminate RM3061 from other microorganisms. Table S-11B shows the top scoring identifications when MS/MS fragment ions are compared to D-, E-, P-specific *in silico* fragment ions. The top identification of both algorithms is once again the YbgS protein of *E. coli* strain RM3061. The "runner-up" identification score is shared by YbgS of *E. coli* strain K-12, *E. coli* O55 strain RM7208 and *Shigella flexneri*. Table S-11C shows the top identifications when MS/MS fragment ions are compared to D-specific *in silico* fragment ions. Once again, the top scoring identification of both algorithms is YbgS of *E. coli* strain RM3061, and the "runner-up" identification score is shared by YbgS of *E. coli* strain K-12, *E. coli* O55 strain RM7208 and *Shigella flexneri.* There is only a ~ 4 Da MW difference between the YbgS of RM3061 and YbgS of K-12/RM7208/*S. flexneri* which is difficult to detect by MALDI-TOF-MS, but the sequence-specific MS/MS fragment ions easily discriminates between these protein sequences. Finally, there is a relative enhancement of the top identification score as the residue specificity of *in silico* fragment ions is narrowed from non-residue-specific *in silico* fragment ions to D-, E-, P-specific *in silico* fragment ions to D-specific *in silico* fragment ions.

REFERENCES

(S1) Fagerquist, C.K.; Garbus, B.R.; Williams, K.E.; Bates, A.H.; Harden, L.A. *J. Am. Soc. Mass Spectrom. in press.*

(S2) Mandrell, R.E.; Harden, L.A.; Horn, S.T.; Haddon, W.F.; Miller, W.G. *American Society of Microbiology*, Los Angeles, CA, May 21-25, **2000**, Poster C-177.

Table S-1

Strain_gene	Accession	Strain_gene	Accession	Strain_gene	Accession	Strain_gene	Accession
RM2008_cspC	GU391767	RM2164_hdeA	GU391814	RM2009_yahO	GU391861	RM3061_ybgS	GU391908
RM2009_cspC	GU391768	RM3061_hdeA	GU391815	RM2012_yahO	GU391862	RM3654_ybgS	GU391909
RM2012_cspC	GU391769	RM3654_hdeA	GU391816	RM2023_yahO	GU391863	RM5603_ybgS	GU391910
RM2023_cspC	GU391770	RM5603_hdeA	GU391817	RM2024_yahO	GU391864	RM6087_ybgS	GU391911
RM2024_cspC	GU391771	RM6087_hdeA	GU391818	RM2027_yahO	GU391865	RM6444_ybgS	GU391912
RM2027_cspC	GU391772	RM6444_hdeA	GU391819	RM2038_yahO	GU391866	RM6763_ybgS	GU391913
RM2038_cspC	GU391773	RM6763_hdeA	GU391820	RM2039_yahO	GU391867	RM7208_ybgS	GU391914
RM2039_cspC	GU391774	RM7208_hdeA	GU391821	RM2042_yahO	GU391868	RM7347_ybgS	GU391915
RM2042_cspC	GU391775	RM7347_hdeA	GU391822	RM2054_yahO	GU391869	RM7386_ybgS	GU391916
RM2054_cspC	GU391776	RM7386_hdeA	GU391823	RM2057_yahO	GU391870	RM7416_ybgS	GU391917
RM2057_cspC	GU391777	RM7416_hdeA	GU391824	RM2068_yahO	GU391871	RM7454_ybgS	GU391918
RM2068_cspC	GU391778	RM7454_hdeA	GU391825	RM2069_yahO	GU391872	RM7494_ybgS	GU391919
RM2069_cspC	GU391779	RM7494_hdeA	GU391826	RM2072_yahO	GU391873	RM7495_ybgS	GU391920
RM2072_cspC	GU391780	RM7495_hdeA	GU391827	RM2162_yahO	GU391874	RM7543_ybgS	GU391921
RM2162_cspC	GU391781	RM7543_hdeA	GU391828	RM2163_yahO	GU391875	RM2008_yjbJ	GU391922
RM2163_cspC	GU391782	RM2008_hdeB	GU391829	RM2164_yahO	GU391876	RM2009_yjbJ	GU391923
RM2164_cspC	GU391783	RM2009_hdeB	GU391830	RM3061_yahO	GU391877	RM2012_yjbJ	GU391924
RM3061_cspC	GU391784	RM2012_hdeB	GU391831	RM3654_yahO	GU391878	RM2023_yjbJ	GU391925
RM3654_cspC	GU391785	RM2023_hdeB	GU391832	RM5603_yahO	GU391879	RM2024_yjbJ	GU391926
RM5603_cspC	GU391786	RM2024_hdeB	GU391833	RM6087_yahO	GU391880	RM2027_yjbJ	GU391927
RM6087_cspC	GU391787	RM2027_hdeB	GU391834	RM6444_yahO	GU391881	RM2038_yjbJ	GU391928
RM6444_cspC	GU391788	RM2038_hdeB	GU391835	RM6763_yahO	GU391882	RM2039_yjbJ	GU391929
RM6763_cspC	GU391789	RM2039_hdeB	GU391836	RM7208_yahO	GU391883	RM2042_yjbJ	GU391930
RM7208_cspC	GU391790	RM2042_hdeB	GU391837	RM7347_yahO	GU391884	RM2054_yjbJ	GU391931
RM7347_cspC	GU391791	RM2054_hdeB	GU391838	RM7386_yahO	GU391885	RM2057_yjbJ	GU391932
RM7386_cspC	GU391792	RM2057_hdeB	GU391839	RM7416_yahO	GU391886	RM2068_yjbJ	GU391933
RM7416_cspC	GU391793	RM2068_hdeB	GU391840	RM7454_yahO	GU391887	RM2069_yjbJ	GU391934
RM7454_cspC	GU391794	RM2069_hdeB	GU391841	RM7494_yahO	GU391888	RM2072_yjbJ	GU391935
RM7494_cspC	GU391795	RM2072_hdeB	GU391842	RM7495_yahO	GU391889	RM2162_yjbJ	GU391936
RM7495_cspC	GU391796	RM2162_hdeB	GU391843	RM7543_yahO	GU391890	RM2163_yjbJ	GU391937
RM7543_cspC	GU391797	RM2163_hdeB	GU391844	RM2008_ybgS	GU391891	RM2164_yjbJ	GU391938
RM2008_hdeA	GU391798	RM2164_hdeB	GU391845	RM2009_ybgS	GU391892	RM3061_yjbJ	GU391939
RM2009_hdeA	GU391799	RM3061_hdeB	GU391846	RM2012_ybgS	GU391893	RM3654_yjbJ	GU391940
RM2012_hdeA	GU391800	RM3654_hdeB	GU391847	RM2023_ybgS	GU391894	RM5603_yjbJ	GU391941
RM2023_hdeA	GU391801	RM5603_hdeB	GU391848	RM2024_ybgS	GU391895	RM6087_yjbJ	GU391942
RM2024_hdeA	GU391802	RM6087_hdeB	GU391849	RM2027_ybgS	GU391896	RM6444_yjbJ	GU391943
RM2027_hdeA	GU391803	RM6444_hdeB	GU391850	RM2038_ybgS	GU391897	RM6763_yjbJ	GU391944
RM2038_hdeA	GU391804	RM6763_hdeB	GU391851	RM2039_ybgS	GU391898	RM7208_yjbJ	GU391945
RM2039_hdeA	GU391805	RM7208_hdeB	GU391852	RM2042_ybgS	GU391899	RM7347_yjbJ	GU391946
RM2042_hdeA	GU391806	RM7347_hdeB	GU391853	RM2054_ybgS	GU391900	RM7386_yjbJ	GU391947
RM2054_hdeA	GU391807	RM7386_hdeB	GU391854	RM2057_ybgS	GU391901	RM7416_yjbJ	GU391948
RM2057_hdeA	GU391808	RM7416_hdeB	GU391855	RM2068_ybgS	GU391902	RM7454_yjbJ	GU391949
RM2068_hdeA	GU391809	RM7454_hdeB	GU391856	RM2069_ybgS	GU391903	RM7494_yjbJ	GU391950
RM2069_hdeA	GU391810	RM7494_hdeB	GU391857	RM2072_ybgS	GU391904	RM7495_yjbJ	GU391951
RM2072_hdeA	GU391811	RM7495_hdeB	GU391858	RM2162_ybgS	GU391905	RM7543_yjbJ	GU391952
RM2162_hdeA	GU391812	RM7543_hdeB	GU391859	RM2163_ybgS	GU391906		
RM2163_hdeA	GU391813	RM2008_yahO	GU391860	RM2164_ybgS	GU391907		

Table S-2 *E. coli* strains	Serotype	Source	Location	Year	***hdeA***		***hdeB***		***cspC***		***ybgS***		***yahO***		***yjbJ***	
					NT	**AA**	**NT**	**AA**	**NT**	**AA**	**NT**	**AA**	**NT**	**AA**	**NT**	**AA**
RM1272 (EDL933)	O157:H7	Hamburger	USA	1982		**1**		2*		**1**		**2**		**2**		**1**
RM5603	O157:H7	Water	CA (Monterey County)	2006	1	**1**	2	2*	1	**1**	2	**2**	2	**2**	1	**1**
RM6087	O157:H7	Human (spinach outbreak)	CA	2006	1	1	2	2*	1	1	2	2	2	2	1	1
RM6444	O157:H7	Cow feces/dirt	CA (San Benito County)	2006	1	1	2	2*	1	1	2	2	2	2	1	1
RM6763	O157:H7	Human (iceberg lettuce outbreak)	PA	2006	1	1	2	2*	1	1	2	2	2	2	1	1
RM7347	O157:H7	Cow feces	CA	2008	1	1	2	2*	1	1	2	2	2	2	1	1
RM7386	O157:H7	Human (romaine lettuce outbreak)	WA	2008	1	1	2	2*	1	1	2	2	2	2	1	1
RM7416	O157:H7	Cow feces	CA	2008	1	1	2	2*	1	1	2	2	2	2	1	1
RM7454	O157:H7	Wildlife	CA	2008	1	1	2	2*	1	1	2	2	2	2	1	1
RM7494	O157:H7	Cow feces	CA	2008	1	1	2	2*	1	1	2	2	2	2	1	1
RM7495	O157:H7	Cow feces	CA	2008	1	1	2	2*	1	1	2	2	2	2	1	1
RM7543	O157:H7	Human (lettuce outbreak)	MI	2008	1	1	2	2*	1	1	2	2	2	2	1	1
RM2012	O55:H7	Human	USA (NY)	1950	1	1*	1	1*	1	1	2	2	2	2	1	1
RM2027	O55:H7	Human	USA (FL)	1979	1	1	1	1	1	1	2	2	2	2	1	1
RM2042	O55:H7	Human	USA (NJ)	1966	1	1	1	1	3	**2**	2	2	2	2	1	1
RM2057	O55:H7	Human	Sri Lanka	1965	1	**1***	1	**1***	1	**1**	2	**2**	2	**2**	1	**1**
RM2072	O55:H7	Human	Iran	1963	1	1	1	1	1	1	2	2	2	2	1	1
RM2163	O55:H7	Human	Unknown		1	1*	1	1*	1	1	2	2	2	2	1	1
RM3654	O55:H7	Human	Canada		1	1	1	1	1	1	2	2	2	2	1	1
RM2008	O55:H6	Human	USA (PA)	1956	2	2	3	1	2	1	1	1	1	1	2	2
RM2009	O55:H6	Human	Congo	1962	2	2	3	1	2	1	1	1	1	1	2	2
RM2023	O55:H6	Human	Guyana	1958	2	2	3	1	2	1	1	1	1	1	2	2
RM2038	O55:H6	Human	Germany	1951	2	2	3	1	2	1	1	1	1	1	2	2
RM2039	O55:H6	Human	USA (PA)	1954	2	2	3	1	2	1	1	1	1	1	2	2
RM2054	O55:H6	Human	France	1951	2	2	3	1	2	1	1	1	1	1	3	2
RM2068	O55:H6	Human	Mexico	1986	2	**2**	3	**1**	2	**1**	1	**1**	1	**1**	2	**2**
RM2069	O55:H6	Human	USA (TX)	1977	2	2	3	1	2	1	1	1	1	1	2	2
RM2164	O55:H6		Unknown		2	2	3	1	2	1	1	1	1	1	2	2
RM2024	O55:HN	Human	USA (PA)	1956	2	**2**	3	**1**	2	**1**	1	**1**	1	**1**	2	**2**
RM2162	O55:HN		Unknown		2	2	3	1	2	1	1	1	1	1	2	2
RM7208	O55	Swine	USA (CA)	2008	3	1	4	1	1	1	4	**4**	3	3	1	1
RM3061	non-O157:H7	Romaine lettuce	USA (CA)	2002	4	**1**	5	**1**	4	**1**	3	**3**	4	**3**	1	**1**
K-12	non-O157:H7					**1**		**1**		**1**		**4**		**4**		**1**

Protein sequences in bold were uploaded to the USDA *in silico* database. EDL993 and K-12 protein sequences were obtained from *ExPASy* databases and uploaded to the *in silico* database. An asterisk (*) indicates that the protein(s) is (are) not expressed due either to a mis-sense mutation in the start codon of the gene or a mis-sense mutation in the promoter region of the genes.[S2]

Table S-3

HdeA (acid stress chaperone-like protein)

1. MKKVLGVILGGLLLLPVVSNAADAQKAADNKKPVNSWTCEDFLAVDESFQPTAVGFAEALNNKDKPEDAVLDVQGIATVTPAIVQACTQDKQANF
KDKVKGEWDKIKKDM
2. MKKVLGVILGGLLLLPVVSNAADAQKAADNKKPVNSWTCEDFLAVDESFQPTAVGFAEALNNKDKPEDAVLDVQGIATVTPAIVQACTQDKKASF
KDKVKGEWDKIKKDM

HdeB (acid stress chaperone-like protein)

1. MNISSLRKAFIFMGAVAALSLVNAQSALAANESAKDMTCQEFIDLNPKAMTPVAWWMLHEETVYKGGDTVTLNETDLTQIPKVIEYCKKNPQKNL
YTFKNQASNDLPN
2*. Not expressed due to mis-sense mutation in the start codon of *hdeB*.[S2]

CspC (cold shock-like protein)

1. MAKIKGQVKWFNESKGFGFITPADGSKDVFVHFSAIQGNGFKTLAEGQNVEFEIQDGQKGPAAVNVTAI
2. MAKIKVQVKWFNESKGFGFITPADGSKDVFVHFSAIQGNGFKTLAEGQNVEFEIQDGQKGPAAVNVTAI

YbgS (or homeobox)

1. MKMTKLATLFLTATLSLASGAALAADSGAQSNNGQANAAADAGQVAPDARENVAPNNVDNNGVNTGSGGTMLHSDGSSMNNDGMTKDEEHKN
TMCKDGRCPDINKKVQTGDGINNDVDTKTDGTTQ
2. MKMTKLATLFLTATLSLASGAALAADSGAQTNNGQANAAADAGQVAPDARENVAPNNVDNNGVNTGSGGTMLHPDGSSMNNDGMTKDEEHKN
TMCKDGRCPDINKKVQTGDGINNDVDTKTDGTTQ
3. MKMTKLATLFLTATLSLASGAALAADSGAQSNNGQANAAADAGQVAPDARENVAPNNVDNNGVNTGSGGTMLHPDGSSMNNDGMTKDEEHKN
TMCKDGRCPDINKKVQTGDGINNDVDTKTDGTTQ
4. MKMTKLATLFLTATLSLASGAALAADSGAQTNNGQANAAADAGQVAPDARENVAPNNVDNNGVNTGSGGTMLHSDGSSMNNDGMTKDEEHKN
TMCKDGRCPDINKKVQTGDGINNDVDTKTDGTTQ

YahO (unknown function)

1. MKIISKMLVGALAFAVTNVYAAELMTKAEFEKVASQYEKIGDISTSNEMSTADAKEDLIKKADEKGADVLVLTSGQTDNKIHGTANIYKKK
2. MKIISKMLVGALAFAVTNVYAAELMTKAEFEKVESQYEKIGDISTSNEMSTADAKEDLIKKADEKGADVLVLTSGQTDNKIHGTADIYKKK
3. MKIISKMLVGALAFAVTNVYAAELMTKAEFEKVESQYEKIGDISTSNEMSTADAKEDLIKKADEKGADVLVLTSGQTDNKIHGTANIYKKK
4. MKIISKMLVGALALAVTNVYAAELMTKAEFEKVESQYEKIGDISTSNEMSTADAKEDLIKKADEKGADVLVLTSGQTDNKIHGTANIYKKK

YjbJ (putative stress-response protein) or CsbD family protein

1. MNKDEAGGNWKQFKGKVKEQWGKLTDDDMTIIEGKRDQLVGKIQERYGYQKDQAEKEVVDWETRNEYRW
2. MNKDEAGGNWKQFKGKVKEQWGKLTDDDMTIIEGKRDQLVGKIQERYGYQKDQAEKEVDSWEKRHDYRW

Table S-4 (ID #46) The top seven identification scores of a protein biomarker from *E. coli* O157:H7 strain RM1272 (EDL933) observed at *m/z* 10471.7 (Figure 1) and analyzed by MS/MS and top-down proteomics using a non-residue-specific *in silico* fragment ion comparison.

Table S-4

***In Silico* ID**	**Identifier**	**Sample Name**	**Protein**	**USDA Score**	***P*-value**
36101	**>tr|Q8X948|Q8X948_ECO57**	***Escherichia coli* O157:H7 (strain EDL933)**	**Putative homeobox protein PTM-24SigPep 10473.06**	**41.58**	**2.2E-4**
43990	>◊|WGM|WGM_PSMRU_5B	*Escherichia coli* O157:H7 (strain RM5603)	YbgS (homeobox) protein PTM-24SigPep 10473.06	41.58	2.2E-4
43973	>◊|WGM|WGM_PSMRU_1B	*Escherichia coli* O55:H7 (strain RM2057)	YbgS (or homeobox) protein PTM-24SigPep 10473.06	41.58	2.2E-4
43995	>◊|WGM|WGM_PSMRU_7B	*Escherichia coli* O55 (strain RM7208)	YbgS (or homeobox) protein PTM-24SigPep 10463.02	38.61	
32552	>tr|A6FIF8|A6FIF8_9GAMM	*Moritella* PE36	Putative uncharacterized protein PTM Met 10472.03		1.1E-3
35345	>sp|P0AAV6|YBGS_ECOLI	*Escherichia coli* (strain K-12)	Uncharacterized YbgS protein PTM-24SigPep 10463.02	38.61	1.9E-3
35346	>sp|P0AAV7|YBGS_SHIFL	*Shigella flexneri*	Uncharacterized YbgS protein PTM-24SigPep 10463.02	38.61	
43995	>◊|WGM|WGM_PSMRU_7B	*Escherichia coli* O55 (strain RM7208)	YbgS (or homeobox) protein PTM-24SigPep 10463.02		1.9E-3
32552	>tr|A6FIF8|A6FIF8_9GAMM	*Moritella* PE36	Putative uncharacterized protein PTM Met 10472.03	37.62	
35346	>sp|P0AAV7|YBGS_SHIFL	*Shigella flexneri*	Uncharacterized YbgS protein PTM-24SigPep 10463.02		1.9E-3

MS/MS to *in silico* comparison parameters

Intensity threshold: 8%
Number of MS/MS peaks with intensity ≥ 8%: 101.
m/z range for comparison: 0-14,000 Th.
Fragment ion tolerance: 2.5 Th.
Protein MW 10471 ± 10 Da. Number of bacterial proteins 2041.
All *in silico* fragment ions compared.
"PTM N-Met" indicates that the *in silico* protein sequence was modified to remove the N-terminal methionine.
"PTM #SigPep" indicates that the *in silico* protein sequence was modified to remove a signal peptide.

Algorithm computation times

USDA peak matching algorithm: 71.4 seconds.
P-value: 204.8 seconds.

Table S-5A (ID #44) The top sixteen identification scores of a protein biomarker from *E. coli* O157:H7 strain RM1272 (EDL933) observed at *m/z* 7272.8 [S1] and analyzed by MS/MS and top-down proteomics using a non-residue-specific *in silico* fragment ion comparison.

Table S-5A

In Silico ID	Identifier	Sample Name	Protein	USDA Score	*P*-value
31566	**>sp|P0A9Y8|CSPC_ECO57**	***Escherichia coli* O157:H7**	**Cold shock-like protein cspC PTM-Met 7271.17**	**54.55**	**1.3E-6**
43992	>◊|WGM|WGM_PSMRU_5E	*Escherichia coli* O157:H7 (strain RM5603)	Cold shock-like protein cspC PTM-Met 7271.17	54.55	1.3E-6
43981	>◊|WGM|WGM_PSMRU_1E	*Escherichia coli* O55:H7 (strain RM2057)	Cold shock-like protein cspC PTM-Met 7271.17	54.55	1.3E-6
43970	>◊|WGM|WGM_PSMRU_3	*Escherichia coli* O55:H6 (strain RM2068)	Cold shock-like protein cspC PTM-Met 7271.17	54.55	1.3E-6
43977	>◊|WGM|WGM_PSMRU_4E	*Escherichia coli* O55:HN (strain RM2024)	Cold shock-like protein cspC PTM-Met 7271.17	54.55	1.3E-6
31575	>sp|P0A9Y6|CSPC_ECOLI	*Escherichia coli* (strain K-12)	Cold shock-like protein cspC PTM-Met 7271.17	54.55	1.3E-6
43964	>◊|WGM|WGM_PSMRU_2E	Non-O157:H7 *Escherichia coli* (strain RM3061)	Cold shock-like protein cspC PTM-Met 7271.17	54.55	1.3E-6
		38 other bacterial microorganisms	"	54.55	1.3E-6
27841	>tr|A4Y561|A4Y561_SHEPC	*Shewanella putrefaciens* (strain CN-32 / ATCC BAA-453)	Cold shock DNA-binding protein family, PTM-Met 7273.15	50.91	1.6E-5
28335	>tr|A1RLK9|A1RLK9_SHESW	*Shewanella* (strain W3-18-1)	Cold shock DNA-binding protein family, PTM-Met 7273.15	50.91	1.6E-5
28514	>tr|A2V1K6|A2V1K6_SHEPU	*Shewanella putrefaciens* 200	Cold shock DNA-binding domain protein, PTM-Met 7273.15	50.91	1.6E-5
27594	>tr|Q3IDH5|Q3IDH5_PSEHT	*Pseudoalteromonas haloplanktis* (strain TAC 125)	Putative secreted calmodulin-like protein 7273	43.64	
27410	>tr|B4F1J2|B4F1J2_PROMI	*Proteus mirabilis*	50S ribosomal protein L29 7273.5		1.1E-3
27410	>tr|B4F1J2|B4F1J2_PROMI	*Proteus mirabilis*	50S ribosomal protein L29 7273.5	41.82	
27594	>tr|Q3IDH5|Q3IDH5_PSEHT	*Pseudoalteromonas haloplanktis* (strain TAC 125)	Putative secreted calmodulin-like protein 7273		1.7E-3
27069	**>sp|P0A7M8|RL29_ECO57**	***Escherichia coli* O157:H7**	**50S ribosomal protein L29 7273.46**	**40.00**	**2.8E-3**
27046	>sp|P0A7M6|RL29_ECOLI	*Escherichia coli* (strain K12)	50S ribosomal protein L29 7273.46	40.00	2.8E-3
		24 other bacterial microorganisms	"	40.00	2.8E-3

MS/MS to *in silico* comparison parameters

Intensity threshold: 2%.
Number of MS/MS peaks with intensity ≥ 2%: 55.
m/z range for comparison: 0-14,000 Th.
Fragment ion tolerance: 2.5 Th.
Protein MW 7271 ± 10 Da. Number of bacterial proteins 1588.
All *in silico* fragment ions compared.
"PTM N-Met" indicates that the *in silico* protein sequence was modified to remove the N-terminal methionine.
"PTM #SigPep" indicates that the *in silico* protein sequence was modified to remove a signal peptide.

Algorithm computation times

USDA peak matching algorithm: 28.0 seconds.
P-value: 43.8 seconds.

Table S-5B (ID #44) The top fifteen identification scores of a protein biomarker from *E. coli* O157:H7 strain RM1272 (EDL933) observed at *m/z* 7272.8 [S1] and analyzed by MS/MS and top-down proteomics using a D-, E-, P-specific *in silico* fragment ion comparison.

Table S-5B

In Silico ID	Identifier	Sample Name	Protein	USDA Score	*P*-value
31566	**>sp|P0A9Y8|CSPC_ECO57**	***Escherichia coli* O157:H7**	**Cold shock-like protein cspC PTM-Met 7271.17**	**34.55**	**3.9E-9**
43992	>◊|WGM|WGM_PSMRU_5E	*Escherichia coli* O157:H7 (strain RM5603)	Cold shock-like protein cspC PTM-Met 7271.17	34.55	3.9E-9
43981	>◊|WGM|WGM_PSMRU_1E	*Escherichia coli* O55:H7 (strain RM2057)	Cold shock-like protein cspC PTM-Met 7271.17	34.55	3.9E-9
43970	>◊|WGM|WGM_PSMRU_3	*Escherichia coli* O55:H6 (strain RM2068)	Cold shock-like protein cspC PTM-Met 7271.17	34.55	3.9E-9
43977	>◊|WGM|WGM_PSMRU_4E	*Escherichia coli* O55:HN (strain RM2024)	Cold shock-like protein cspC PTM-Met 7271.17	34.55	3.9E-9
31575	>sp|P0A9Y6|CSPC_ECOLI	*Escherichia coli* (strain K-12)	Cold shock-like protein cspC PTM-Met 7271.17	34.55	3.9E-9
43964	>◊|WGM|WGM_PSMRU_2E	Non-O157:H7 *Escherichia coli* (strain RM3061)	Cold shock-like protein cspC PTM-Met 7271.17	34.55	3.9E-9
		38 other bacterial microorganisms	"	34.55	3.9E-9
27598	>tr|Q3K9W5|Q3K9W5_PSEPF	*Pseudomonas fluorescens* (strain PfO-1)	Putative uncharacterized protein, PTM-Met 7271.85	25.45	
27069	**>sp|P0A7M8|RL29_ECO57**	***Escherichia coli* O157:H7**	**50S ribosomal protein L29 7273.46**		**3.7E-4**
27618	>tr|Q4KH92|Q4KH92_PSEF5	*Pseudomonas fluorescens* (strain Pf-5 / ATCC BAA-477)	Putative uncharacterized protein PTM-Met 7276.32	21.82	
27046	>sp|P0A7M6|RL29_ECOLI	*Escherichia coli* (strain K12)	50S ribosomal protein L29 7273.46		3.7E-4
28089	sp|Q9KPE1|Y2429_VIBCH	*Vibrio cholerae*	UPF0243 zinc-binding protein VC-2429, PTM-Met 7262.04	21.82	
		23 other bacterial microorganisms	50S ribosomal protein L29 7273.46		3.7E-4
		3 other *Vibrio cholerae* strains	Putative uncharacterized protein 7262.04	21.82	
27937	>sp|Q0BS01|RL35_GRABC	*Granulibacter bethesdensis* (strain ATCC BAA_1260 / CGDNIH1)	50S ribosomal protein L35 PTM-Met 7267.51		9.3E-4
27069	**>sp|P0A7M8|RL29_ECO57**	***Escherichia coli* O157:H7**	**50S ribosomal protein L29 7273.46**	**20.00**	
26960	>sp|A7MPH1|RL29_ENTS8	*Enterobacter sakazakii* (strain ATCC BAA-894)	50S ribosomal protein L29 7274.49		1.5E-3
27046	>sp|P0A7M6|RL29_ECOLI	*Escherichia coli* (strain K12)	50S ribosomal protein L29 7273.46	20.00	
26949	>sp|A7FNM7|RL29_YERP3	*Yersinia pseudotuberculosis* serotype O:1b (strain IP 31758)	50S ribosomal protein L29 7273.46		2.1E-3
		23 other bacterial microorganisms	"	20.00	
27051	>tr|A6BVS7|A6BVS7_YERPE	*Yersinia pestis* CA88-4125	50S ribosomal protein L29 7273.46		2.1E-3

MS/MS to *in silico* comparison parameters

Intensity threshold: 2%.
Number of MS/MS peaks with intensity ≥ 2%: 55.
m/z range for comparison: 0-14,000 Th.
Fragment ion tolerance: 2.5 Th.
Protein MW 7271 ± 10 Da. Number of bacterial proteins 1588.
D-, E-, P-specific *in silico* fragment ions compared.
"PTM N-Met" indicates that the *in silico* protein sequence was modified to remove the N-terminal methionine.
"PTM #SigPep" indicates that the *in silico* protein sequence was modified to remove a signal peptide.

Algorithm computation times

USDA peak matching algorithm: 13.8 seconds.
P-value: 42.0 seconds.

Table S-6A (ID #45) The top six identification scores of a protein biomarker from *E. coli* O157:H7 strain RM1272 (EDL933) observed at *m/z* 9947.6 [S1] and analyzed by MS/MS and top-down proteomics using a non-residue-specific *in silico* fragment ion comparison and a protein MW tolerance of ± 300 Da.

Table S-6A

In Silico ID	Identifier	Sample Name	Protein	USDA Score	*P*-value
30847	>tr|Q2J3U4|Q2J3U4_RHOP2	*Rhodopseudomonas palustris* (strain HaA2)	Putative uncharacterized protein PTM-23SigPep 9950.95	57.58	5.0E-5
24947	>tr|A7IJA2|A7IJA2_XANP2	*Xanthobacter autotrophicus* (strain ATCC BAA-1158 / Py2)	Putative uncharacterized protein PTM-Met 9732.14	51.52	
24983	>tr|A8F4V8|A8F4V8_THELT	Th*ermotoga lettingae* (strain ATCC BAA_301 / DSM 14385 / TMO)	Flagellar biosynthetic protein FliQ PTM-Met 9731.9		3.1E-4
24983	>tr|A8F4V8|A8F4V8_THELT	Th*ermotoga lettingae* (strain ATCC BAA_301 / DSM 14385 / TMO)	Flagellar biosynthetic protein FliQ PTM-Met 9731.9	51.52	
24947	>tr|A7IJA2|A7IJA2_XANP2	*Xanthobacter autotrophicus* (strain ATCC BAA-1158 / Py2)	Putative uncharacterized protein PTM-Met 9732.14		4.5E-4
24619	>tr|Q8CKW9|Q8CKW9_YERPE	*Yersinia pestis*	Putative uncharacterized protein PTM-Met 9729.38	48.48	8.6E-4
25452	>tr|A0LVV2|A0LVV2_ACIC1	*Acidothermus cellulolyticus* (strain ATCC 43068 / 11B)	Putative uncharacterized protein PTM-Met 9729.33	48.48	9.7E-4
28961	>tr|Q0ASU8|Q0ASU8_MARMM	*Maricaulis maris* (strain MCS10)	Putative uncharacterized protein 9942.93	48.48	1.7E-3

MS/MS to *in silico* comparison parameters

Intensity threshold: 2%.
Number of MS/MS peaks with intensity ≥ 2%: 33.
m/z range for comparison: 0-14,000 Th.
Fragment ion tolerance: 2.5 Th.
Protein MW 9947 ± 300 Da. Number of bacterial proteins 6509.
All *in silico* fragment ions compared.
"PTM N-Met" indicates that the *in silico* protein sequence was modified to remove the N-terminal methionine.
"PTM #SigPep" indicates that the *in silico* protein sequence was modified to remove a signal peptide.

Algorithm computation times

USDA peak matching algorithm: 128.4 seconds.
P-value: 143.3 seconds.

Table S-6B (ID #45) The top six identification scores of a protein biomarker from *E. coli* O157:H7 strain RM1272 (EDL933) observed at *m/z* 9947.6 [S1] and analyzed by MS/MS and top-down proteomics using a D-, E-, P-specific *in silico* fragment comparison and a protein MW tolerance of ± 300 Da.

Table S-6B

***In Silico* ID**	**Identifier**	**Sample Name**	**Protein**	**USDA Score**	***P*-value**
23630	>sp|A1JLK7|Y1782_YERE8	*Yersinia enterocolitica* serotype O:8 / biotype 1B (strain 8081)	UPF0181 protein YE1782 9742.38	33.33	
24207	>tr|Q2BRI9|Q2BRI9_9GAMM	*Neptuniibacter caesariensis*	Putative uncharacterized protein 9730.71		1.8E-4
24512	**>sp|P0AET0|HDEA_ECO57**	***Escherichia coli* O157:H7**	**Chaperone-like protein HdeA PTM-21SigPep 9738.91**	**30.30**	
24947	>tr|A7IJA2|A7IJA2_XANP2	*Xanthobacter autotrophicus* (strain ATCC BAA-1158 / Py2)	Putative uncharacterized protein PTM-Met 9732.14		2.4E-4
24511	>sp|P0AES9|HDEA_ECOLI	*Escherichia coli* (strain K-12)	Chaperone-like protein HdeA PTM-21SigPep 9738.91	30.30	
30776	>tr|Q8GFX0|Q8GFX0_CITFR	*Citrobacter freundii*	Putative uncharacterized protein orf8 9953.22		3.4E-4
24513	>sp|P0AET1|HDEA_SHIFL	*Shigella flexneri*	Chaperone-like protein HdeA PTM-21SigPep 9738.91	30.30	
28561	>sp|Q0VME7|HFQ_ALCBS	*Alcanivorax borkumensis* (strain SK2 / ATCC 700651 / DSM 11573)	Protein hfq 9951.44		9.4E-4
14484	>tr|A3TGL5|A3TGL5_9MICO	*Janibacter* HTCC2649	Putative WhiB-family transcriptional regulator PTM-Met 10206.56	30.30	
28918	>tr|B4WY86|B4WY86_9GAMM	*Alcanivorax* DG881	RNA chaperone Hfq, putative 9951.44		9.4E-4
14841	>tr|B3EEL3|B3EEL3_CHLLI	*Chlorobium limicola* DSM 245	Hydrogenase assembly chaperone hypC/hupF 10210.59	30.30	
30094	>sp|Q0VME7|HFQ_ALCBS	*Alcanivorax borkumensis* (strain SK2 / ATCC 700651 / DSM 11573)	Protein hfq 9951.44		9.4E-4

MS/MS to *in silico* comparison parameters

Intensity threshold: 2%.
Number of MS/MS peaks with intensity ≥ 2%: 33.
m/z range for comparison: 0-14,000 Th.
Fragment ion tolerance: 2.5 Th.
Protein MW 9947 ± 300 Da. Number of bacterial proteins 6509.
D-, E-, P-specific *in silico* fragment ions compared.
"PTM N-Met" indicates that the *in silico* protein sequence was modified to remove the N-terminal methionine.
"PTM #SigPep" indicates that the *in silico* protein sequence was modified to remove a signal peptide.

Algorithm computation times

USDA peak matching algorithm: 79.0 seconds.
P-value: 102.0 seconds.

Table S-6C (ID #45) The top seven identification scores of a protein biomarker from *E. coli* O157:H7 strain RM1272 (EDL933) observed at *m/z* 9947.6 [S1] and analyzed by MS/MS and top-down proteomics using a D-specific *in silico* fragment ion comparison and a protein MW tolerance of ± 300 Da.

Table S-6C

***In Silico* ID**	**Identifier**	**Sample Name**	**Protein**	**USDA Score**	***P*-value**
24512	**>sp|P0AET0|HDEA_ECO57**	***Escherichia coli* O157:H7**	**Chaperone-like protein HdeA PTM-21SigPep 9738.91**	**30.30**	**2.8E-5**
43991	>◊|WGM|WGM_PSMRU_5C	*Escherichia coli* O157:H7 (strain RM5603)	HdeA acid stress chaperone-like protein PTM_21SigPep 9738.91	30.30	2.8E-5
24511	>sp|P0AES9|HDEA_ECOLI	*Escherichia coli* (strain K-12)	Chaperone-like protein HdeA PTM-21SigPep 9738.91	30.30	2.8E-5
43985	>◊|WGM|WGM_PSMRU_2C	Non-O157:H7 *Escherichia coli* (strain RM3061)	HdeA acid stress chaperone-like protein PTM_21SigPep 9738.91	30.30	2.8E-5
43979	>◊|WGM|WGM_PSMRU_1C	*Escherichia coli* O55:H7 (strain RM2057)	HdeA acid stress chaperone-like protein PTM_21SigPep 9738.91	30.30	2.8E-5
24513	>sp|P0AET1|HDEA_SHIFL	*Shigella flexneri*	Chaperone-like protein HdeA PTM-21SigPep 9738.91	30.30	2.8E-5
25406	>tr|Q0S9N3|Q0S9N3_RHOSR	*Rhodococcus* (strain RHA1)	Putative uncharacterized protein PTM-Met 9742.51	24.24	
13938	tr|A8GE15|A8GE15_SERP5	*Serratia proteamaculans* (strain 568)	Putative uncharacterized protein 10210.85		4.4E-5

MS/MS to *in silico* comparison parameters

Intensity threshold: 2%.
Number of MS/MS peaks with intensity ≥ 2%: 33.
m/z range for comparison: 0-14,000 Th.
Fragment ion tolerance: 2.5 Th.
Protein MW 9947 ± 300 Da. Number of bacterial proteins 6509.
D-specific *in silico* fragment ions compared.
"PTM N-Met" indicates that the *in silico* protein sequence was modified to remove the N-terminal methionine.
"PTM #SigPep" indicates that the *in silico* protein sequence was modified to remove a signal peptide.

Algorithm computation times

USDA peak matching algorithm: 61.0 seconds.
P-value: 88.9 seconds.

Table S-7A (ID #48) The top six identification scores of a protein biomarker from *E. coli* O157:H7 strain RM5603 observed at *m/z* 7706.1 (Figure S-2) and analyzed by MS/MS and top-down proteomics using a non-residue-specific *in silico* fragment ion comparison.

Table S-7A

***In Silico* ID**	**Identifier**	**Sample Name**	**Protein**	**USDA Score**	***P*-value**
43989	**>◊|WGM|WGM_PSMRU_5A**	***Escherichia coli* O157:H7 (strain RM5603)**	**YahO protein PTM-21SigPep 7707.62**	**61.67**	**6.7E-19**
26947	>tr|Q8X699|Q8X699_ECO57	*Escherichia coli* O157:H7 (strain EDL933)	Putative uncharacterized protein YahO PTM-21SigPep 7707.62	61.67	6.7E-19
43962	>◊|WGM|WGM_PSMRU_1A	*Escherichia coli* O55:H7 (strain RM2057)	YahO protein PTM-21SigPep 7707.62	61.67	6.7E-19
26281	>sp|P75694|YAHO_ECOLI	*Escherichia coli* (strain K-12)	UPF0379 protein YahO PTM-21SigPep 7706.64	57.50	1.7E-15
43983	>◊|WGM|WGM_PSMRU_2A	Non-O157:H7 *Escherichia coli* (strain RM3061)	YahO protein PTM-21SigPep 7706.64	57.50	1.7E-15
25669	>tr|A5V415|A5V415_SPHWW	*Sphingomonas wittichii* (strain RW1 / DSM 6014 / JCM 10273)	DNA binding domain, excisionase family 7703.11	35.00	
25710	>tr|A6Y3Y9|A6Y3Y9_VIBCH	*Vibrio cholerae* RC385	Transcriptional regulator 7704.84		3.8E-3

MS/MS to *in silico* comparison parameters

Intensity threshold: 2%
Number of MS/MS peaks with intensity ≥ 2%: 120.
m/z range for comparison: 0-14,000 Th.
Fragment ion tolerance: 2.5 Th.
Protein MW 7705 ± 10 Da. Number of bacterial proteins 1323.
All *in silico* fragment ions compared.
"PTM N-Met" indicates that the *in silico* protein sequence was modified to remove the N-terminal methionine.
"PTM #SigPep" indicates that the *in silico* protein sequence was modified to remove a signal peptide.

Algorithm computation times

USDA peak matching algorithm: 43.0 seconds.
P-value: 189.6 seconds.

Table S-7B (ID #48) The top six identification scores for a protein biomarker from *E. coli* O157:H7 strain RM5603 observed at *m/z* 7706.1 (Figure S-2) and analyzed by MS/MS and top-down proteomics using a D-, E-, P-specific *in silico* fragment ion comparison.

Table S-7B

***In Silico* ID**	**Identifier**	**Sample Name**	**Protein**	**USDA Score**	***P*-value**
43989	**>◊|WGM|WGM_PSMRU_5A**	***Escherichia coli* O157:H7 (strain RM5603)**	**YahO protein PTM-21SigPep 7707.62**	**41.67**	**5.3E-19**
26947	>tr|Q8X699|Q8X699_ECO57	*Escherichia coli* O157:H7 (strain EDL933)	Putative uncharacterized protein YahO PTM-21SigPep 7707.62	41.67	5.3E-19
43962	>◊|WGM|WGM_PSMRU_1A	*Escherichia coli* O55:H7 (strain RM2057)	YahO protein PTM-21SigPep 7707.62	41.67	5.3E-19
26281	>sp|P75694|YAHO_ECOLI	*Escherichia coli* (strain K-12)	UPF0379 protein YahO PTM-21SigPep 7706.64	34.17	2.0E-13
43983	>◊|WGM|WGM_PSMRU_2A	Non-O157:H7 *Escherichia coli* (strain RM3061)	YahO protein PTM-21SigPep 7706.64	34.17	2.0E-13
25880	>tr|B1SUG8|B1SUG8_9BACI	*Geobacillus* WCH70	Putative uncharacterized protein 7707.91	22.50	1.4E-4

MS/MS to *in silico* comparison parameters

Intensity threshold: 2%
Number of MS/MS peaks with intensity ≥ 2%: 120.
m/z range for comparison: 0-14,000 Th.
Fragment ion tolerance: 2.5 Th.
Protein MW 7705 ± 10 Da. Number of bacterial proteins 1323.
D-, E-, P-specific *in silico* fragment ions compared.
"PTM N-Met" indicates that the *in silico* protein sequence was modified to remove the N-terminal methionine.
"PTM #SigPep" indicates that the *in silico* protein sequence was modified to remove a signal peptide.

Algorithm computation times

USDA peak matching algorithm: 15.6 seconds.
P-value: 274.0 seconds.

Table S-8A (ID #57) The top five identification scores of a protein biomarker from *E. coli* O157:H7 strain RM5603 observed at *m/z* 8323.6 (Figure S-2) and analyzed by MS/MS and top-down proteomics using a non-residue-specific *in silico* fragment ion comparison.

Table S-8A

***In Silico* ID**	**Identifier**	**Sample Name**	**Protein**	**USDA Score**	***P*-value**
7647	>A5EUR7|A5EUR7_BRASB	*Bradyrhizobium* (strain BTAi1 / ATCC BAA-1182)	Putative uncharacterized protein PTM-Met 8327.41	44.57	4.8E-5
7339	Q3AJX5|Q3AJX5_SYNSC	*Synechococcus* (strain CC9605).	Putative uncharacterized protein 8327	38.04	1.5E-3
39169	>tr|A1IT41|A1IT41_NEIMA	*Neisseria meningitidis* serogroup A	Putative uncharacaterized protein 8318.28	36.96	
39258	>tr|Q64Y50|Q64Y50_BACFR	*Bacteroides fragilis*	Acyl carrier protein 8320.61		2.1E-3
7232	>Q03QZ4|Q03QZ4_LACBA	*Lactobacillus brevis* (strain ATCC 367 / JCM 1170).	Putative uncharacterized protein 8325.36	35.87	
43993	**>◊|WGM|WGM_PSMRU_5F**	***Escherichia coli* O157:H7 (strain RM5603)**	**YjbJ protein 8325.26**		**2.8E-3**
7615	>A4JQ21|A4JQ21_BURVG	*Burkholderia vietnamiensis* (strain G4 / LMG 22486)	Putative uncharacterized protein PTM-Met 8323.28	35.87	
6967	>P68207|YJBJ_ECO57	*Escherichia coli* O157:H7	UPF0337 protein yjbJ 8325.26		2.8E-3

MS/MS to *in silico* comparison parameters

Intensity threshold: 6%
Number of MS/MS peaks with intensity ≥ 6%: 92.
m/z range for comparison: 0-14,000 Th.
Fragment ion tolerance: 2.5 Th.
Protein MW 8323 ± 10 Da. Number of bacterial proteins 2556.
All *in silico* fragment ions compared.
"PTM N-Met" indicates that the *in silico* protein sequence was modified to remove the N-terminal methionine.
"PTM #SigPep" indicates that the *in silico* protein sequence was modified to remove a signal peptide.

Algorithm computation times

USDA peak matching algorithm: 79.7 seconds.
P-value: 201.6 seconds.

Table S-8B (ID #57) The top nine identification scores of a protein biomarker from *E. coli* O157:H7 strain RM5603 observed at *m/z* 8323.6 (Figure S-2) and analyzed by MS/MS and top-down proteomics using a D-, E-, P-specific *in silico* fragment ion comparison.

Table S-8B

***In Silico* ID**	**Identifier**	**Sample Name**	**Protein**	**USDA Score**	***P*-value**
43993	**>◊|WGM|WGM_PSMRU_5F**	***Escherichia coli* O157:H7 (strain RM5603)**	**YjbJ protein 8325.26**	**27.17**	**6.6E-7**
6967	>P68207|YJBJ_ECO57	*Escherichia coli* O157:H7	UPF0337 protein yjbJ 8325.26	27.17	6.6E-7
43982	>◊|WGM|WGM_PSMRU_1F	*Escherichia coli* O55:H7 (strain RM2057)	YjbJ protein 8325.26	27.17	6.6E-7
6956	>P68206|YJBJ_ECOI	*Escherichia coli* (strain K12).	UPF0337 protein yjbJ 8325.26	27.17	6.6E-7
43965	>◊|WGM|WGM_PSMRU_2F	Non-O157:H7 *Escherichia coli* (strain RM3061)	YjbJ protein 8325.26	27.17	6.6E-7
6978	>P68208|YJBJ_SHIFL	*Shigella flexneri*	UPF0337 protein yjbJ 8325.26	27.17	6.6E-7
7027	>A7ZUR7|A7ZUR7_ECO24	*Escherichia coli* O139:H28 (strain E24377A / ETEC).	CsbD family protein 8325.26	27.17	6.6E-7
		39 other *E. coli* or *Shigella* strains	UPF0337 protein yjbJ or CsbD family protein 8325.26	27.17	6.6E-7
7555	>A3M433|A3M433_ACIBT	*Acinetobacter baumannii* (strain ATCC 17978 / NCDC KC 755)	7-Fe ferredoxin PTM-Met 8329.19	21.74	
7419	>Q7UDX6|Q7UDX6_RHOBA	*Rhodopirellula baltica*	Putative uncharacterized protein 8324.58		1.8E-4

MS/MS to *in silico* comparison parameters

Intensity threshold: 6%
Number of MS/MS peaks with intensity ≥ 6%: 92.
m/z range for comparison: 0-14,000 Th.
Fragment ion tolerance: 2.5 Th.
Protein MW 8323 ± 10 Da. Number of bacterial proteins 2556.
D-, E-, P-specific *in silico* fragment ions compared.
"PTM N-Met" indicates that the *in silico* protein sequence was modified to remove the N-terminal methionine.
"PTM #SigPep" indicates that the *in silico* protein sequence was modified to remove a signal peptide.

Algorithm computation times

USDA peak matching algorithm: 31.3 seconds.
P-value: 242.5 seconds.

Table S-8C (ID #57) The top nine identification scores of a protein biomarker from *E. coli* O157:H7 strain RM5603 observed at *m/z* 8323.6 (Figure S-2) and analyzed by MS/MS and top-down proteomics using a D-specific *in silico* fragment ion comparison.

Table S-8C

***In Silico* ID**	**Identifier**	**Sample Name**	**Protein**	**USDA Score**	***P*-value**
43993	**>◊|WGM|WGM_PSMRU_5F**	***Escherichia coli* O157:H7 (strain RM5603)**	**YjbJ protein 8325.26**	**23.91**	**3.2E-11**
6967	>P68207|YJBJ_ECO57	*Escherichia coli* O157:H7	UPF0337 protein yjbJ 8325.26	23.91	3.2E-11
43982	>◊|WGM|WGM_PSMRU_1F	*Escherichia coli* O55:H7 (strain RM2057)	YjbJ protein 8325.26	23.91	3.2E-11
6956	>P68206|YJBJ_ECOI	*Escherichia coli* (strain K12)	UPF0337 protein yjbJ 8325.26	23.91	3.2E-11
43965	>◊|WGM|WGM_PSMRU_2F	Non-O157:H7 *Escherichia coli* (strain RM3061)	YjbJ protein 8325.26	23.91	3.2E-11
6978	>P68208|YJBJ_SHIFL	*Shigella flexneri*	UPF0337 protein yjbJ 8325.26	23.91	3.2E-11
7027	>A7ZUR7|A7ZUR7_ECO24	*Escherichia coli* O139:H28 (strain E24377A / ETEC).	CsbD family protein 8325.26	23.91	3.2E-11
		39 other *E. coli* strains and *Shigella* species/strains	UPF0337 protein yjbJ or CsbD family protein 8325.26	23.91	3.2E-11
39084	>tr|B4V013|B4V013_9ACTO	*Streptomyces* Mg1	Conserved hypothetical protein 8324.35	17.39	
38837	>tr|A8YD15|A8YD15_MICAE	*Microcystis aeruginosa* PCC 7806	Similar to tr|P72616|P72616 8318.67		2.2E-6

MS/MS to *in silico* comparison parameters

Intensity threshold: 6%
Number of MS/MS peaks with intensity ≥ 6%: 92.
m/z range for comparison: 0-14,000 Th.
Fragment ion tolerance: 2.5 Th.
Protein MW 8323 ± 10 Da. Number of bacterial proteins 2556.
D-specific *in silico* fragment ions compared.
"PTM N-Met" indicates that the *in silico* protein sequence was modified to remove the N-terminal methionine.
"PTM #SigPep" indicates that the *in silico* protein sequence was modified to remove a signal peptide.

Algorithm computation times

USDA peak matching algorithm: 19.3 seconds.
P-value: 261.0 seconds.

Table S-9A (ID #49) The top seven identification scores of a protein biomarker from *E. coli* O157:H7 strain RM5603 observed at *m/z* 9737.8 (Figure S-2) and analyzed by MS/MS and top-down proteomics using a non-residue-specific *in silico* fragment ion comparison.

Table S-9A

***In Silico* ID**	**Identifier**	**Sample Name**	**Protein**	**USDA Score**	***P*-value**
43991	**>◊|WGM|WGM_PSMRU_5C**	***Escherichia coli* O157:H7 (strain RM5603)**	**HdeA acid stress chaperone-like protein PTM-21SigPep 9738.91**	**62.07**	**1.2E-5**
24512	>sp|P0AET0|HDEA_ECO57	*Escherichia coli* O157:H7 (strain EDL-933)	Chaperone-like protein HdeA PTM-21SigPep 9738.91	62.07	1.2E-5
43979	>◊|WGM|WGM_PSMRU_1C	*Escherichia coli* O55:H7 (strain RM2057)	Chaperone-like protein HdeA PTM-21SigPep 9738.91	62.07	1.2E-5
24511	>sp|P0AES9|HDEA_ECOLI	*Escherichia coli* (strain K-12)	Chaperone-like protein HdeA PTM-21SigPep 9738.91	62.07	1.2E-5
43985	>◊|WGM|WGM_PSMRU_2C	Non-O157:H7 *Escherichia coli* (strain RM3061)	HdeA acid stress chaperone-like protein PTM-21SigPep 9738.91	62.07	1.2E-5
24513	>sp|P0AET1|HDEA_SHIFL	*Shigella flexneri*	Chaperone-like protein hdeA PTM-21SigPep 9738.91	62.07	1.2E-5
24117	>tr|Q12GM9|Q12GM9_POLSJ	*Polaromonas* (strain JS666 / ATCC BAA-500)	YCII-related 9736.26	51.72	
24674	>tr|Q9ZB08|Q9ZB08_9LACT	*Lactococcus lactis*	Putative uncharacterized protein PTM-N-Met 9743.04		8.1E-4

MS/MS to *in silico* comparison parameters

Intensity threshold: 4%.
Number of MS/MS peaks with intensity ≥ 4%: 29.
m/z range for comparison: 0-14,000 Th.
Fragment ion tolerance: 2.5 Th.
Protein MW 9737 ± 10 Da. Number of bacterial proteins 2017.
All *in silico* fragment ions compared.
"PTM N-Met" indicates that the *in silico* protein sequence was modified to remove the N-terminal methionine.
"PTM #SigPep" indicates that the *in silico* protein sequence was modified to remove a signal peptide.

Algorithm computation times

USDA peak matching algorithm: 32.3 seconds.
P-value: 36.4 seconds.

Table S-9B (ID #49) The top seven identification scores of a protein biomarker from *E. coli* O157:H7 strain RM5603 observed at *m/z* 9737.8 (Figure S-2) and analyzed by MS/MS and top-down proteomics using a D-, E-, P-specific *in silico* fragment ion comparison.

Table S-9B

***In Silico* ID**	**Identifier**	**Sample Name**	**Protein**	**USDA Score**	***P*-value**
43991	**>◊|WGM|WGM_PSMRU_5C**	***Escherichia coli* O157:H7 (strain RM5603)**	**HdeA acid stress chaperone-like protein PTM-21SigPep 9738.91**	**48.28**	**3.3E-7**
24512	>sp|P0AET0|HDEA_ECO57	*Escherichia coli* O157:H7 (strain EDL-933)	Chaperone-like protein HdeA PTM-21SigPep 9738.91	48.28	3.3E-7
43979	>◊|WGM|WGM_PSMRU_1C	*Escherichia coli* O55:H7 (strain RM2057)	HdeA acid stress chaperone-like protein PTM-21SigPep 9738.91	48.28	3.3E-7
24511	>sp|P0AES9|HDEA_ECOLI	*Escherichia coli* (strain K-12)	Chaperone-like protein HdeA PTM-21SigPep 9738.91	48.28	3.3E-7
43985	>◊|WGM|WGM_PSMRU_2C	Non-O157:H7 *Escherichia coli* (strain RM3061)	HdeA acid stress chaperone-like protein PTM-21SigPep 9738.91	48.28	3.3E-7
24513	>sp|P0AET1|HDEA_SHIFL	*Shigella flexneri*	Chaperone-like protein HdeA PTM-21SigPep 9738.91	48.28	3.3E-7
24117	>tr|Q12GM9|Q12GM9_POLSJ	*Polaromonas* (strain JS666 / ATCC BAA-500)	YCII-related 9736.26	27.59	9.5E-4

MS/MS to *in silico* comparison parameters

Intensity threshold: 4%.
Number of MS/MS peaks with intensity ≥ 4%: 29.
m/z range for comparison: 0-14,000 Th.
Fragment ion tolerance: 2.5 Th.
Protein MW 9737 ± 10 Da. Number of bacterial proteins 2017.
D-, E-, P-specific *in silico* fragment ions compared.
"PTM N-Met" indicates that the *in silico* protein sequence was modified to remove the N-terminal methionine.
"PTM #SigPep" indicates that the *in silico* protein sequence was modified to remove a signal peptide.

Algorithm computation times

USDA peak matching algorithm:17.7 seconds.
P-value: 24.1 seconds.

Table S-9C (ID #49) The top seven identification scores of a protein biomarker from *E. coli* O157:H7 strain RM5603 observed at *m/z* 9737.8 (Figure S-2) and analyzed by MS/MS and top-down proteomics using a D-specific *in silico* fragment ion comparison.

Table S-9C

***In Silico* ID**	**Identifier**	**Sample Name**	**Protein**	**USDA Score**	***P*-value**
43991	**>◊|WGM|WGM_PSMRU_5C**	***Escherichia coli* O157:H7 (strain RM5603)**	**HdeA acid stress chaperone-like protein PTM-21SigPep 9738.91**	**44.83**	**1.0E-8**
24512	>sp|P0AET0|HDEA_ECO57	*Escherichia coli* O157:H7 (strain EDL-933)	Chaperone-like protein HdeA PTM-21SigPep 9738.91	44.83	1.0E-8
43979	>◊|WGM|WGM_PSMRU_1C	*Escherichia coli* O55:H7 (strain RM2057)	HdeA acid stress chaperone-like protein PMT-21SigPep 9738.91	44.83	1.0E-8
24511	>sp|P0AES9|HDEA_ECOLI	*Escherichia coli* (strain K-12)	Chaperone-like protein HdeA PTM-21SigPep 9738.91	44.83	1.0E-8
43985	>◊|WGM|WGM_PSMRU_2C	Non-O157:H7 *Escherichia coli* (strain RM3061)	HdeA acid stress chaperone-like protein PMT-21SigPep 9738.91	44.83	1.0E-8
24513	>sp|P0AET1|HDEA_SHIFL	*Shigella flexneri*	Chaperone-like protein HdeA PTM-21SigPep 9738.91	44.83	1.0E-8
25134	>tr|B1F9I0|B1F9I0_9BURK	*Burkholderia ambifaria* IOP40-10	ThiamineS protein PTM-Met 9740.88	20.69	
24095	>tr|A0LMA8|A0LMA8_SYNFM	*Syntrophobacter fumaroxidans* (strain DSM 10017 / MPOB)	Putative uncharacterized protein 9730.07		1.7E-4

MS/MS to *in silico* comparison parameters

Intensity threshold: 4%.
Number of MS/MS peaks with intensity ≥ 4%: 29.
m/z range for comparison: 0-14,000 Th.
Fragment ion tolerance: 2.5 Th.
Protein MW 9737 ± 10 Da. Number of bacterial proteins 2017.
D-specific *in silico* fragment ions compared.
"PTM N-Met" indicates that the *in silico* protein sequence was modified to remove the N-terminal methionine.
"PTM #SigPep" indicates that the *in silico* protein sequence was modified to remove a signal peptide.

Algorithm computation times

USDA peak matching algorithm: 12.6 seconds.
P-value: 19.8 seconds.

Table S-10 (ID #52) The top seven identification scores of a protein biomarker from a non-O157:H7, non-pathogenic *E. coli* strain RM3061 observed at *m/z* 9738.4 (Figure 4) and analyzed by MS/MS and top-down proteomics using a non-residue-specific *in silico* fragment ion comparison.

Table S-10

***In Silico* ID**	**Identifier**	**Sample Name**	**Protein**	**USDA Score**	***P*-value**
43985	**>◊|WGM|WGM_PSMRU_2C**	**Non-O157:H7 *Escherichia coli* (strain RM3061)**	**HdeA acid stress chaperone protein PTM-21SigPep 9738.91**	**72.73**	**3.9E-9**
24511	>sp|P0AES9|HDEA_ECOLI	*Escherichia coli* (strain K-12)	Chaperone-like protein HdeA PTM-21SigPep 9738.91	72.73	3.9E-9
24512	>sp|P0AET0|HDEA_ECO57	*Escherichia coli* O157:H7 (strain EDL933)	Chaperone-like protein HdeA PTM-21SigPep 9738.91	72.73	3.9E-9
43991	>◊|WGM|WGM_PSMRU_5C	*Escherichia coli* O157:H7 (strain RM5603)	HdeA acid stress chaperone protein PTM-21SigPep 9738.91	72.73	3.9E-9
43979	>◊|WGM|WGM_PSMRU_1C	*Escherichia coli* O55:H7 (strain RM2057)	HdeA acid stress chaperone protein PTM-21SigPep 9738.91	72.73	3.9E-9
24513	>sp|P0AET1|HDEA_SHIFL	*Shigella flexneri*	Chaperone-like protein HdeA PTM-21SigPep 9738.91	72.73	3.9E-9
25300	>tr|B3QEW6|B3QEW6_RHOPA	*Rhodopseudomonas palustris* TIE-1	Transcriptional regulator/antitoxin, MazE PTM-Met 9739.97	51.52	2.3E-4

MS/MS to *in silico* comparison parameters

Intensity threshold: 4%.
Number of MS/MS peaks with intensity ≥ 4%: 33.
m/z range for comparison: 0-14,000 Th.
Fragment ion tolerance: 2.5 Th.
Protein MW 9738 ± 10 Da. Number of bacterial proteins 1954.
All *in silico* fragment ions compared.
"PTM N-Met" indicates that the *in silico* protein sequence was modified to remove the N-terminal methionine.
"PTM #SigPep" indicates that the *in silico* protein sequence was modified to remove a signal peptide.

Algorithm computation times

USDA peak matching algorithm: 32.7 seconds.
P-value: 38.3 seconds.

Table S-11A (ID #53) The top five identification scores of a protein biomarker from a non-O157:H7, non-pathogenic *E. coli* strain RM3061 observed at *m/z* 10461.8 [S1] (and at *m/z* 10458.3 in Figure 4) and analyzed by MS/MS and top-down proteomics using a non-residue-specific *in silico* fragment ion comparison.

Table S-11A

In Silico ID	Identifier	Sample Name	Protein	USDA Score	*P*-value
43984	**>◊|WGM|WGM_PSMRU_2B**	**Non-O157:H7 *Escherichia coli* (strain RM3061)**	**YbgS protein (or homeobox protein) PTM-24SigPep 10459.03**	**50.91**	**3.6E-5**
34747	>tr|A5VJP9|A5VJP9_LACRF	*Lactobacillus reuteri* (strain ATCC 23272 / DSM 20016 / F275)	Putative transcriptional regulator, XRE family 10463.18	41.82	1.1E-3
32201	>tr|Q3SNL3|Q3SNL3_NITWN	*Nitrobacter winogradskyi* (strain Nb-255 / ATCC 25391)	Putative uncharacterized protein 10465.9	40.00	2.7E-3
32385	>tr|B0SZR9|B0SZR9_CAUSK	*Caulobacter* (strain K31)	Antibiotic biosynthesis monooxygenase, PTM-34SigPep 10469.03	38.18	
35398	>sp|P19729|DMPK_PSEUF	*Pseudomonas* (strain CF600)	Phenol hydroxylase P0 protein PTM-Met 10453.96		7.2E-3
32595	>sp|Q1D771|RS19_MYXXD	*Myxococcus xanthus* (strain DK 1622)	30S ribosomal protein S19 PTM-Met 10465.32	38.18	
35490	>tr|A6UEN0|A6UEN0_SINMW	*Sinorhizobium medicae* (strain WSM419)	Sarcosine oxidase, delta subunit family, PTM-Met 10455.61		7.2E-3

MS/MS to *in silico* comparison parameters

Intensity threshold: 4%.
Number of MS/MS peaks with intensity ≥ 4%: 55.
m/z range for comparison: 0-14,000 Th.
Fragment ion tolerance: 2.5 Th.
Protein MW 10461 ± 10 Da. Number of bacterial proteins 2406.
All *in silico* fragment ions compared.
"PTM N-Met" indicates that the *in silico* protein sequence was modified to remove the N-terminal methionine.
"PTM #SigPep" indicates that the *in silico* protein sequence was modified to remove a signal peptide.

Algorithm computation times

USDA peak matching algorithm: 60.9 seconds.
P-value: 79.7 seconds.

Table S-11B (ID #53) The top six identification scores of a protein biomarker from non-O157:H7, non-pathogenic *E. coli* strain RM3061 observed at *m/z* 10461.8 [S1] (and at *m/z* 10458.3 in Figure 4) and analyzed by MS/MS and top-down proteomics using a D-, E-, P-specific *in silico* fragment ion comparison.

Table S-11B

***In Silico* ID**	**Identifier**	**Sample Name**	**Protein**	**USDA Score**	***P*-value**
43984	**>◊|WGM|WGM_PSMRU_2B**	**Non-O157:H7 *Escherichia coli* (strain RM3061)**	**YbgS protein (or homeobox protein) PTM-24SigPep 10459.03**	**38.18**	**1.3E-8**
35345	>sp|P0AAV6|YBGS_ECOLI	*Escherichia coli* (strain K12)	Uncharacterized protein ybgS PTM-24SigPep 10463.02	25.45	3.9E-4
43995	>◊|WGM|WGM_PSMRU_7B	*Escherichia coli* O55 (strain RM7208)	YbgS protein (or homeobox) PTM-24SigPep 10463.02	25.45	3.9E-4
35346	>sp|P0AAV7|YBGS_SHIFL	*Shigella flexneri*	Uncharacterized protein ybgS PTM-24SigPep 10463.02	25.45	3.9E-4
32421	>tr|A1W973|A1W973_ACISJ	*Acidovorax* (strain JS42)	Transcriptional regulator, XRE family PTM-Met 10469.12	25.45	5.2E-4
35413	tr|A4C0Q7|A4C0Q7_9FLAO	*Polaribacter irgensii* 23-P	Putative uncharacterized protein PTM-Met 10454.78	23.64	
32910	>tr|Q1CZ11|Q1CZ11_MYXXD	*Myxococcus xanthus* (strain DK 1622)	Transcriptional regulator, ArsR family PTM-Met 10468.24		6.2E-4

MS/MS to *in silico* comparison parameters

Intensity threshold: 4%.
Number of MS/MS peaks with intensity ≥ 4%: 55.
m/z range for comparison: 0-14,000 Th.
Fragment ion tolerance: 2.5 Th.
Protein MW 10461 ± 10 Da. Number of bacterial proteins 2406.
D-, E-, P-specific *in silico* fragment ions compared.
"PTM N-Met" indicates that the *in silico* protein sequence was modified to remove the N-terminal methionine.
"PTM #SigPep" indicates that the *in silico* protein sequence was modified to remove a signal peptide.

Algorithm computation times

USDA peak matching algorithm: 26.1 seconds.
P-value: 70.4 seconds.

Table S-11C (ID #53) The top five identification scores of a protein biomarker from non-O157:H7, non-pathogenic *E. coli* strain RM3061 observed at *m/z* 10461.8 [S1] (and at *m/z* 10458.3 in Figure 4) and analyzed by MS/MS and top-down proteomics using a D-specific *in silico* fragment ion comparison.

Table S-11C

***In Silico* ID**	**Identifier**	**Sample Name**	**Protein**	**USDA Score**	***P*-value**
43984	**>◊|WGM|WGM_PSMRU_2B**	**Non-O157:H7 *Escherichia coli* (strain RM3061)**	**YbgS protein (or homeobox protein) PTM-24SigPep 10459.03**	**38.18**	**7.3E-11**
35345	>sp|P0AAV6|YBGS_ECOLI	*Escherichia coli* (strain K12)	Uncharacterized protein ybgS PTM-24SigPep 10463.02	25.45	2.5E-5
43995	>◊|WGM|WGM_PSMRU_7B	*Escherichia coli* O55 (strain RM7208)	YbgS protein (or homeobox) PTM-24SigPep 10463.02	25.45	2.5E-5
35346	>sp|P0AAV7|YBGS_SHIFL	*Shigella flexneri*	Uncharacterized protein ybgS PTM-24SigPep 10463.02	25.45	2.5E-5
32298	>tr|Q7V6Q0|Q7V6Q0_PROMM	*Prochlorococcus marinus* (strain MIT 9313)	Putative uncharacterized protein 10469.69	12.73	
34647	>tr|A2SKG0|A2SKG0_METPP	*Methylibium petroleiphilum* (strain PM1)	Putative signal peptide protein 10459.18		4.8E-4

MS/MS to *in silico* comparison parameters

Intensity threshold: 4%.
Number of MS/MS peaks with intensity ≥ 4%: 55.
m/z range for comparison: 0-14,000 Th.
Fragment ion tolerance: 2.5 Th.
Protein MW 10461 ± 10 Da. Number of bacterial proteins 2406.
D-specific *in silico* fragment ions compared.
"PTM N-Met" indicates that the *in silico* protein sequence was modified to remove the N-terminal methionine.
"PTM #SigPep" indicates that the *in silico* protein sequence was modified to remove a signal peptide.

Algorithm computation times

USDA peak matching algorithm: 17.5 seconds.
P-value: 66.9 seconds.

Figure S-1

Amino acid sequence of the homeobox protein of the *E. coli* O157:H7 strain EDL933 and *E. coli* O55:H7 strain RM2057, and YbgS of the non-O157:H7, non-pathogenic *E. coli* strain K-12 and *Shigella flexneri*. The protein sequences are post-translationally modified with a 24-residue signal peptide (in outline). *S····S* symbolizes disulfide bridge between the two cysteine residues (boxed). Amino acid variations between the sequences are also boxed.

***E. coli* O157:H7 strain EDL933 and *E. coli* O55:H7 strain RM2057**

MKMTKLATLFLTATLSLASGAALAADSGAQTNNGQANAAADAGQVAPDARENVAP

S·····*S*

NNVDNNGVNTGSGGTMLHPDGSSMNNDGMTKDEEHKNTMCKDGRCPDINKKV
QTGDGINNDVDTKTDGTTQ

MW = 12882.06 Da

Mature protein MW = 10473.06 Da

E. coli* strain K-12 and *Shigella flexneri

MKMTKLATLFLTATLSLASGAALAADSGAQTNNGQANAAADAGQVAPDARENVAP

S·····*S*

NNVDNNGVNTGSGGTMLHSDGSSMNNDGMTKDEEHKNTMCKDGRCPDINKKV
QTGDGINNDVDTKTDGTTQ

MW = 12872.02 Da

Mature protein MW = 10463.02 Da

Non-O157:H7, non-pathogenic *E. coli* strain RM3061

MKMTKLATLFLTATLSLASGAALAADSGAQSNNGQANAAADAGQVAPDARENVAP

S·····*S*

NNVDNNGVNTGSGGTMLHPDGSSMNNDGMTKDEEHKNTMCKDGRCPDINKKV
QTGDGINNDVDTKTDGTTQ

MW = 12866.01 Da

Mature protein MW = 10459.03 Da

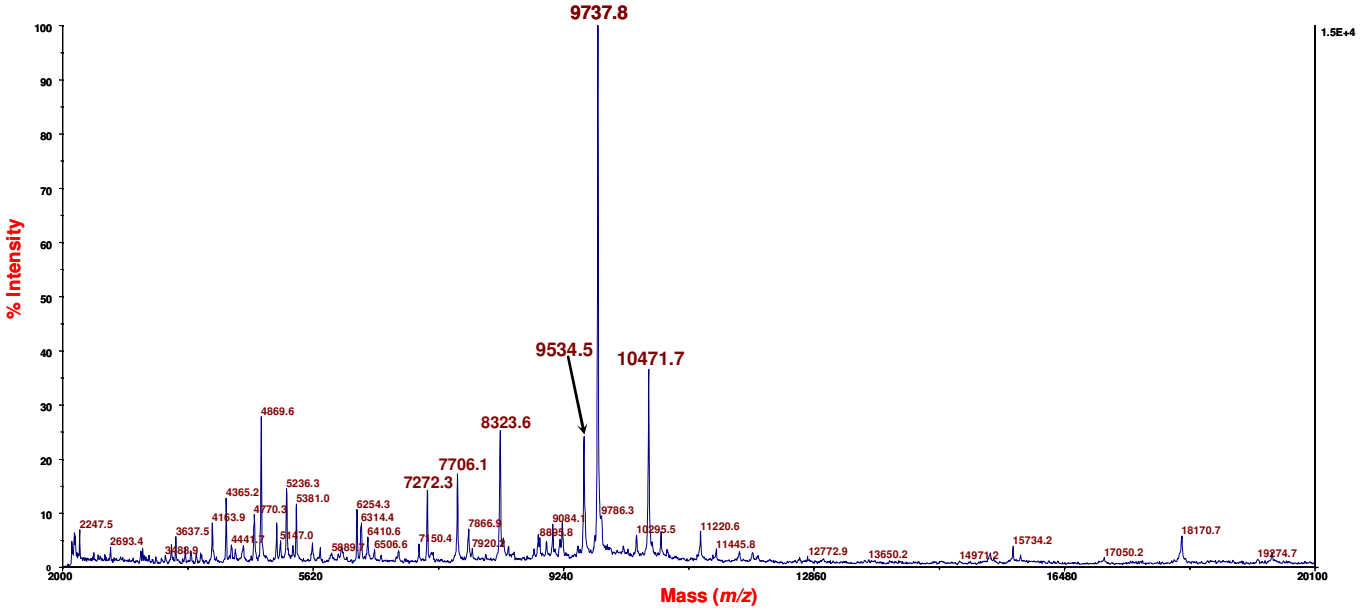

Figure S-2

MS spectrum of the extracted cell lysate of *E. coli* O157:H7 strain RM5603 using HCCA matrix.

Figure S-3

Amino acid sequence of cold shock-like protein CspC and 50S ribosomal protein L29 of *Escherichia coli* O157:H7 strain EDL933. The molecular weight of the post-translationally modified (removal of N-terminal methionine outlined) cold shock-like protein CspC and 50S ribosomal protein L29 are different by only ~2 Da. The sequence of these two proteins is conserved across numerous microorganisms (Tables S-5A,1B).

Cold shock-like protein CspC of *E. coli* O157:H7 strain EDL933
MAKIKGQVKWFNESKGFGFITPADGSKDVFVHFSAIQGNGFKTLAEGQNVEFEIQD GQKGPAAVNVTAI
MW = 7402.37
Mature protein MW = 7271.17

50S ribosomal protein L29 of *E. coli* O157:H7 strain EDL933
MKAKELREKSVEELNTELLNLLREQFNLRMQAASGQLQQSHLLKQVRRDVARVK TLLNEKAGA
MW = 7273.46