Comparative Proteomics of *E. Coli* O157:H7: Two-Dimensional Gel Electrophoresis vs. Two-Dimensional Liquid Chromatography Separation

Nereus W. Gunther IV*, Hoan-Jen Pang+, Alberto Nuñez, Gaylen A. Uhlich

Microbial Food Safety Research Unit, Eastern Regional Research Center, Agricultural Research Service, U.S. Department of Agriculture, 600 E. Mermaid Lane, Wyndmoor, PA 19038, USA

+*Current address: Jean Mayer USDA Human Nutrition Research Center on Aging (HNRCA) at Tufts University, Boston, MA, 021111, USA*

Abstract: The method for comparing bacterial proteomes has traditionally been two-dimensional gel electrophoresis (2-D GE); however, in recent years, new procedures for protein separation have been introduced. One of these new procedures utilizes column-based liquid chromatography (2-D LC) separation. The techniques by which these two methods separate proteins differ significantly; however, it is currently unclear to what degree the sets of proteins identified by these different methods will diverge. To address this question we compared the proteomes of *Escherichia coli* O157:H7 strain EDL933 against a naturally occurring variant using both 2-D GE and 2-D LC. Whole protein samples were prepared from the wild type and variant and split in half, with one half analyzed by 2-D GE and the other with the 2-D LC. Differentially regulated proteins were observed in each system and identified by MALD/I-TOF/TOF analysis. The differences in the protein detection sensitivities of Coomassie-blue stain used in 2-D GE and UV detectors used for 2-D LC resulted in different numbers of total proteins visualized in each system and therefore different numbers of matched protein pairs visualized during comparative assays. Despite the differences in visualization the numbers, but not the identities, of the differentially regulated proteins that could be identified by MALD/I analyses were similar for both 2-D GE and 2-D LC. However, a lack of significant redundancy between the sets of proteins identified suggests that these two methods are complimentary and not strictly corroborative.

Keywords: Comparative proteomics, *Escherichia coli*, liquid chromatography.

INTRODUCTION

The ability to determine the protein complement expressed by an organism at specific times and under specific conditions provides insights into the proteins the organism needs to selectively express to survive and thrive. Likewise, comparing the protein complements expressed by an isogenic mutant and its parent can provide vital clues to the function and importance of the genetic portion, which differs between the two strains. In the study of bacteria relevant to food safety, the practice of comparative proteomics holds great promise for helping to determine the way in which food pathogens are able to exist and persist within the food supply resulting in food-borne illness. To compare the protein complements of an organism, it is first necessary to separate, visualize, and measure the relative amounts of the various individual proteins. The techniques, two-dimensional gel electrophoresis (2-D GE) and two-dimensional liquid chromatography (2-D LC) accomplish these tasks by separating the proteins in a sample over two dimensions. The 2-D GE technique separates individual proteins based on the isoelectric point and the mass of each protein [1, 2]. The 2-D LC technique utilizes the isoelectric point and the hydrophobicity of the individual proteins for separation [3]. These two techniques are both well established and described; however, 2-D GE has been in use for a significantly longer period of time compared to 2-D LC. In this study both of these techniques were utilized to compare the protein profile of *E. coli* O157:H7 strain 43895 against the profile of its naturally occurring phase variant 43895OR [4-6]. The variant strain is known to have a point mutation in the promoter region of the *csgD* gene [6]. It has also been characterized with the following phenotypic differences: a red, dry, and rough colony appearance, increased curli expression, strong biofilm formation, increased invasiveness of HEp-2 cells, and increased virulence in a mouse model [5, 6]. A comparison of the protein profiles of the parent and variant strains provides individual protein functions and possible physiological pathways to be further examined in subsequent experiments. However an unfortunate reality of this type of work is a lack of consistency in the data returned from one experiment to the other also known as biological replicates [7]. Through the investigation of two different techniques we had hoped to identify the cause of these inconsistencies and determine a superior technique. In the end, neither technique was found to be superior and returned protein lists that were significantly dissimilar from one another, suggesting that these techniques would be better viewed as complementary. These results comparing prokaryotic proteomes is consistent with similar

*Address correspondence to this author at the Eastern Regional Research Center, Microbial Food Safety Research Unit, 600 E. Mermaid Lane, Wyndmoor, PA 19038; Tel: (215) 233-6503; Fax: (215) 233-6581;
E-mail: jack.gunther@ars.usda.gov

1875-0397/10

research comparing separation techniques used to investigate eukaryotic proteomes [8, 9]. It is suggestive that multiple proteomic techniques should be applied to this type of research in order to increase the likelihood of identifying a larger percentage of the proteins differentially expressed and to better judge the significance of those proteins identified.

MATERIALS AND METHODOLOGY

Bacterial Strains, Growth, and Whole Cell Protein Extraction

Bacterial strains were grown and maintained on yeast extract casamino acids (YESCA) plates (1.5 % agar). *E. coli* strains 43895 and 43895OR were streaked for confluence onto three YESCA plates each and the plates were incubated at 30°C for 48 h. The plates were flooded with 2 ml of PBS, and the cells were removed from the plate surface with the use of a plastic cell spreader. The cells were pelleted by centrifugation at 5500 × g for 5 min. The following protein lysis and extraction procedure was previously described [10] and is as follows. Cell pellets were resuspended in 0.4 ml of 50 mM Tris (pH 8.0), and tubes were placed in an ice bath and sonicated for 30 seconds using a microtip. The sonication process was then repeated two more times. Next, 1.6 ml of lysis buffer (7.5 M urea, 2.5 M thiourea, 12.5 % glycerol, 50 mM Tris, 2.5 % n-octylglucoside, 6.25 mM TCEP [Tris-carboxyethyl phosphine hydrocholine], 1.25 mM protease inhibitor) were added to each bacterial suspension. The solution was centrifuged at 21,000 × g for 60 min, and the resulting supernatant was collected. A PD-10 column (Amersham Biosciences, Sweden) was then used to exchange the total cellular proteins from the lysis mix into a proprietary Start buffer (Beckman Coulter, CA). Finally the total concentration of proteins present in the resulting preparations was measured by means of a BCA protein assay (Pierce, IL).

Two Dimensional Gel Electrophoresis

A portion of the whole cell protein preparations was precipitated using 8 % TCA (Trichloroacetic acid) and washed with ice cold acetone. Protein pellets were then suspended in rehydration buffer (8 M Urea, 2 % CHAPS, 0.25 % ampholytes, 0.0002 % bromophenol blue, 3 mM tributylphosphine (TBP) in nanopure water). Next, 17 cm isoelectric focusing (IEF) strips (Bio-Rad, CA) were actively rehydrated at 50V for 12 h. The whole cell protein samples containing roughly 450 μg of total protein each were loaded onto the rehydrated strips and separated following the manufacturer's recommendations. This includes an initial linear ramp at 250 V for 30 min followed by a linear ramp at 10,000 volts for 3 h, and finishing with a rapid ramp at 10,000 volts for 55,000 volt-hours. The fully focused IEF strips were placed in equilibration buffer 1 (6 M urea, 2 % SDS, 50 mM Tris HCl, 30 % glycerol, 1 % TCP) for 10 min followed by 10 min in equilibration buffer 2 (same as buffer 1 but with 2.5 % iodoacetamide substituted for the TBP) and then separated in the second dimension using large format Protean II precast 10-20 % Tris-HCL gels (Bio-Rad. The gels were run at 16 mA/gel for 5 to 6 h at a temperature of 10°C. The resulting Coomassie stained gels were scanned using the Personal Densitometer SI (GE Healthcare, NJ), and the digital images were used to align and compare the protein spot intensities between gels from strain 43895 and strain 43895OR. The Z3 2-D Gel Image Analysis System (Compugen, Isreal) program was used to align replicate gels to create a master gel alignment for each strain and to make the spot comparisons between the master gels of the two strains. The pixel densities were measured for each strain for protein spots determined to be common to both master gels. A protein expression ratio for each of the matched proteins was determined by dividing the pixel concentration of a protein spot for one strain by the pixel concentration for the matching protein spot from the other strain (protein X [43895] / protein X [43895OR]). The log of each ratio value was determined, and the mean value for all log values of protein ratios in the comparison between whole cell preparations was determined. Standard deviations were determined for the mean of all log[ratios], and values that fell outside the standard deviations were recorded as being differentially expressed. Therefore, the protein spots, whose log[ratio] values fell outside of the standard deviations and therefore, determined to be differentially expressed significantly, were excised from the gel using a clean razor and stored at -20ºC.

Two Dimensional Liquid Chromatography

The following 2-D LC separation method was described in previous research [10] and is as follows. The remainders of the whole cell protein preparations were diluted in a proprietary Start buffer to a concentration of 5 mg/ml for separation on the Proteolab PF2D system (Beckman-Coulter). First dimension separation utilized a chromatofocusing column (250 × 2.1 mm i.d.) that generated a pH gradient from 8.5 to 4 using the proprietary Start and Elute buffers; this was accomplished over 185 min at a flow rate 0.2 ml/min. Protein separation was monitored by UV absorbance at 280 nm. First dimension fractions were further separated in the second dimension using a C18 reverse-phase column (4.6 × 33 mm; 1.5 μm particle size, non-porous) utilizing an acetonitrile water gradient at a flow rate of 0.75 ml/min over 45 min and monitored by UV absorbance at 214 nm with relative protein concentrations recorded in terms of absorbance units (AU). Fractions were collected using a FC 204 fraction collector (Gilson) in 96-well plates at 30 sec intervals between minutes 10 and 30. The volumes collected from the first dimension separation allowed for multiple second dimension separations to check for reproducibility. The multiple second dimension separations were very reproducible, however since the multiple second dimension separations were not true technical replicates only one of the separations each were used for making the protein expression comparison between the two strains. The plates containing the collected fractions from the second dimension separation were stored at -20ºC until processed for identification using mass spectrometry. The ProteoVue™ (Eprogen Inc., IL) software application was used to convert chromatographic intensities into two-dimensional protein expression maps representative of the contents of the whole cell protein preparations assayed. Differential analysis of individual protein peaks in matched whole cell protein samples was accomplished using the DeltaVue™ (Eprogen Inc.) software application. The concentrations of individual proteins expressed by *E. coli* 43895 were divided by the concentrations of corresponding individual proteins expressed by 43895OR (protein X

[43895] / protein X [43895OR]). The log of each ratio value was calculated and the mean value for all protein ratios in a comparison between strains 43895 and 43895OR determined. Standard deviations were determined for the mean of all log[ratios], and values that fell outside the standard deviations were recorded as being differentially expressed. Fractions containing proteins selected for analysis were transferred to 0.5 ml volume Eppendorf tubes and stored at -20°C. The Eppendorf tubes were previously washed with 50 % acetonitrile followed by 1 % trifluoroacetic acid and Milli Q water to remove potential contaminants.

Mass Spectrometry Sample Preparation

Protein spots removed from 2-D gels were thawed, destained (NH_4HCO_3/50% acetonitrile), and dehydrated (100% acetonitrile). The acetonitrile was removed, and the dried gel slice was resuspended in a solution of 40 mM NH_4HCO_3/10 % acetonitrile containing trypsin gold (Promega, WI) at a concentration of 20 µg/ml (previously diluted in 50 mM acetic acid) and incubated overnight at 37ºC. Next the gel slices were extracted into 50 % acetonitrile/5 % TFA, dried, and resuspended in 3 % acetonitrile/1 % TFA.

Fractions collected from the 2-D LC method were thawed and concentrated to approximately 30 µL using a Speedvac concentrator. For tryptic digestion, 10 µL each of NH_4HCO_3 (pH 8.95) and DTT were added to each sample to a final concentration of 100 mM and 1 mM, respectively, heated to 60°C for 10 min on a preheated digital dry bath, and allowed to cool at room temperature. One microliter of Trypsin Gold at a concentration of 50 µg/µl in 50 mM acetic acid was added to each sample and incubated for 4 h at 37°C with gentle agitation. After digestion, the protein samples were treated with 2 % TFA to stop the trypsin activity.

The resulting peptides from both the 2-D GE and the 2-D LC protein samples were extracted and cleaned using OMIX C18 tips (Pierce) following manufacturer recommendations. Peptides were extracted using a pre-cleaned C18 ZipTip, washed with water containing 0.1 % trifluoroacetic acid (TFA), re-extracted with acetonitrile-water- (50:50) 0.1 % TFA, and mixed with a recrystallized α-cyano-4-hydroxycinnamic acid matrix solution (5 mg/ml, acetonitrile-water- (50:50) 0.1 % TFA) to a final concentration between 100 fmol to 1 pmol/µl. Approximately 0.6-0.7 µl of the peptide-matrix solution was spotted on the mass spectrometer target plate.

Mass Spectrometry and Protein(s) Identification

Matrix-Assisted Laser Desorption/Ionization mass spectrometry with automated tandem time of flight fragmentation of selected ions (MALD/I-TOF/TOF) of trypsin digested proteins were acquired with a 4700 Proteomics Analyzer mass spectrometer (Applied Biosystems, MA) in the positive reflectron mode with a 200 Hz Nd-YAG 355 nm laser. Spectra were obtained by averaging 1000 acquired spectra in the MS mode or 2500 in the MS/MS mode. Collision induced dissociation (CID) with air as the collision gas at approximately $1x10^{-6}$ Torr and a 1 keV acceleration voltage was used for obtaining the MS/MS spectra of selected peptides. Conversion of time of flight to mass (Da) for the monoisotopic ions, $[M + H]^+$ was based on calibration of the instrument with a peptide standard calibration kit (Applied Biosystems) that contained the following peptides: des-Arg^1-bradykinin (*m/z* 904.4681), angiotensin I (*m/z* 1,296.6853), Glu^1-fibrinopeptide B (*m/z* 1,570.6774), ACTH (clip 1-17) (*m/z* 2,903.0867), ACTH (clip 18-39) (*m/z* 2.465.1989), and ACTH (clip 7-38) (*m/z* 3,657.9294). The MS/MS time of flight calibration was obtained from the CID produced fragments of Glu^1-fibrinopeptide B. Peptide mass fingerprints and MS/MS of selected peptides were combined and queried against primary sequence databases using the Mascot search engine-associated GPS Explorer program (Applied Biosystems). Reported protein(s) from database searches from putative peptide sequences were within a ≥95% confidence interval.

Construction of 43895Δ*yjbJ* and Biofilm Formation Assays

A gene knockout mutant of Ecs5028 (*yjbJ*) was constructed using strain 43895. The mutant was constructed utilizing a previously well described method [11] that is briefly as follows: the *yjbJ* gene of *E. coli* strain 43895 was interrupted by replacing 97 bp of sequence immediately following the start codon with a *Pst*I restriction site using overlap extension PCR (primers not shown). The interrupted gene and flanking sequences were amplified by primers 5' - aaggaattcgctgttgttgcaactctgtttc / 5' - aagggatccgatggcgctatcagcgcttatg and cloned into the *Eco*RI/*Bam*HI sites of plasmid pBluescriptII SK+. The kanamycin resistance cassette of plasmid pUC4K was transferred to the *Pst*I site of the interrupted *yjbJ* gene to generate plasmid pBSYK. The excised *Eco*RI/*Bam*HI fragment of pBSYK was used as linear template for chromosomal recombination into strain 43895 using the Quick and Easy Gene Deletion Kit (Gene BridgesGmbH, Germany) resulting in the isogenic mutant 43895Δ*yjbJ*.

The newly constructed mutant was used in conjunction with the parent strain 43895 and the naturally occurring variant 43895OR to investigate if the gene *yjbJ* has an effect on *E. coli's* ability to form biofilms. Biofilm assays were performed using LB medium with no added salt (LB-NS) in 96-well polystyrene microtiter plates (TPP, Switzerland) as previously described but without heat fixation [4].

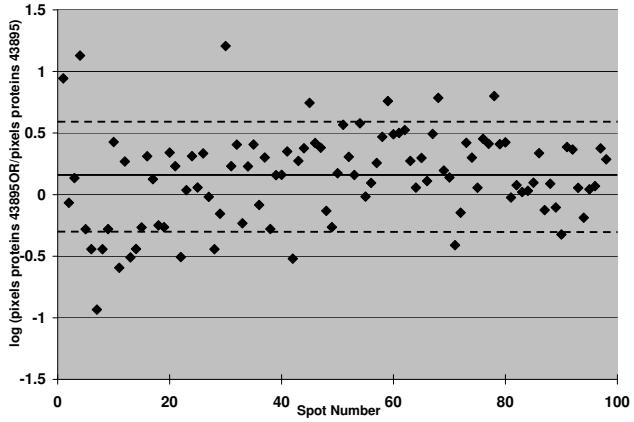
RESULTS

Protein preps #1 for strains 43895 and 43895OR were each initially divided in half with half of the preps being used for 2-D GE and the second half of the preps being used for 2-D LC separation. When the protein preps #1 for both strains were analyzed using 2-D gel electrophoresis, 224 spots were visualized on the gel produced by strain *E. coli* O157:H7 43895 and 225 spots were visualized on the gel produced by strain *E. coli* O157:H7 43895OR. Attempts to overlay and match the spots produced from strain 43895 with the spots produced by strain 43895OR resulted in 98 matching spots between the two strains; with the remaining 126 spots for strain ATCC 43895 and 127 spots for strain ATCC 43895OR unable to be match between the two gels. The log values of the ratio of the pixel concentrations for each of the 98 matched spots were plotted as a scatter graph (Fig. **1A**). Mean and standard deviation values were determined and paired protein ratios falling outside of the two standard deviation lines were established as being differentially regulated. Seven protein spots were identified as being

upregulated in strain 43895 as compared to strain 43895OR with 11 protein spots being identified as being upregulated in strain 43895OR as compared to strain 43895 (Fig. **1A**).

Next, a complete repetition (biological replicate) of the previous protein preparations was made resulting in protein preps #2 for strains 43895 and 43895OR which were again separated and visualized by 2-D GE and 2-D LC. The 2-D gels produced from protein preps #2 resulted in 238 spots being visualized for strain 43895 and 192 spots being visualized for strain 43895OR. The gels were again overlaid and matched resulting in 101 matched protein spots between the two strains (Fig. **1B**). From the analysis 13 protein spots were judged to be upregulated in strain 43895 as compared to strain 43895OR, and 10 protein spots upregulated in strain 43895OR as compared to strain 43895.

A



B

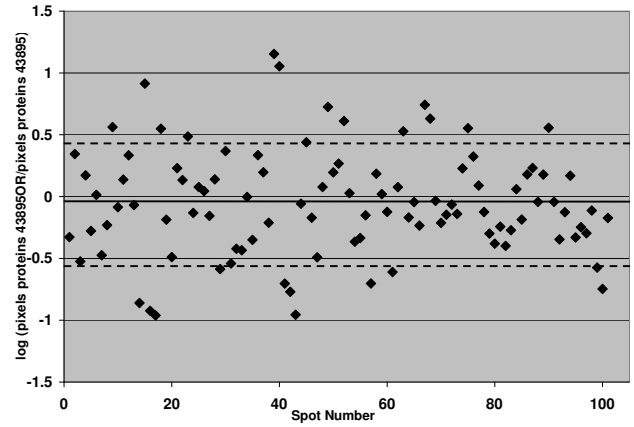
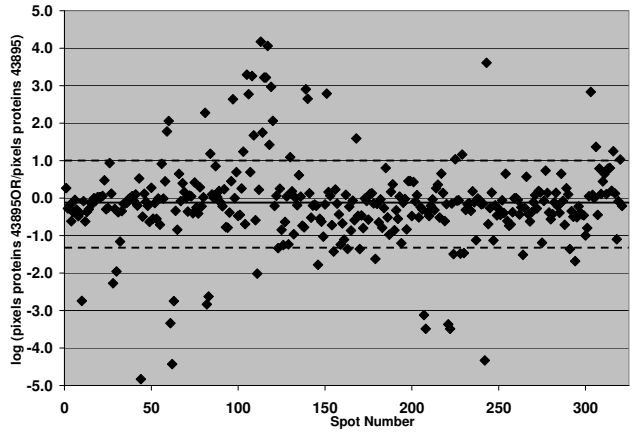


Fig. (1). Graphical analysis of protein expression levels as determined by 2-D GE. (A) A scatter plot of the resulting values for the log (pixel conc. of protein X of strain 43895OR / pixel conc. of protein X of strain 43895) for each of the matched protein pairs derived from protein preparation #1 and analyzed by 2D GE. **(B)** A scatter plot of the resulting values for the log (pixel conc. of protein X of strain 43895OR / pixel conc. of protein X of strain 43895) for each of the matched protein pairs derived from protein preparation #2 and analyzed by 2-D GE. Line "——" = mean value for all of the protein pairs log ratio values. Line "– – – –" = the standard deviations.

The second halves of the protein preps #1 for both strains 43895 and 43895OR were compared using 2-D LC and resulted in the identification of 321 matched protein peaks between strains 43895 and 43895OR. The ratios of the log values in absorbance units (AU) for each matched protein peak along with the mean value and standard deviations were plotted on a scatter graph (Fig. **2A**). Strain 43895 was determined to have 27 protein peaks upregulated as compared to strain 43895OR among the matched peaks. Conversely, strain 43895OR was determined to have 30 upregulated peaks as compared to strain 43895.

Next, a complete repetition (biological replicate) of the first 2-D LC protein separations of protein preps #1 were made using half of the protein preps #2 for the strains 43895 and 43895OR. This matching of the resulting protein peaks from strains 43895 and 43895OR resulted in 205 matched peaks in this series of experiments (Fig. **2B**). Strain 43895

A



B

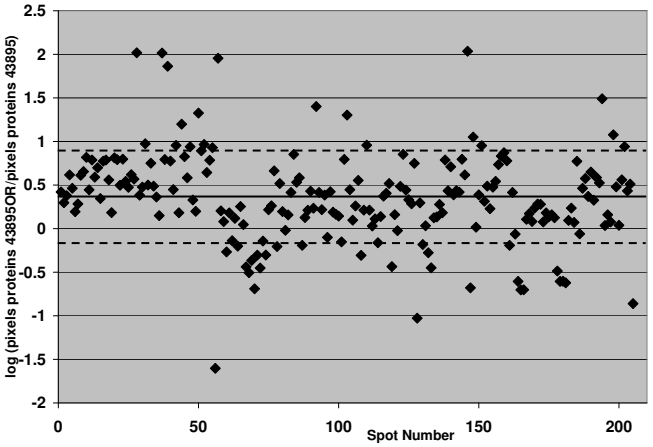


Fig. (2). Graphical analysis of protein expression levels as determined by 2-D LC. (A) A scatter plot of the resulting values for the log (AU protein X of strain 43895OR / AU protein X of strain 43895) for each of the matched protein peaks derived from protein preparation #1 and analyzed by 2-D LC methods. **(B)** A scatter plot of the resulting values for the log (AU protein X of strain 43895OR / AU protein X of strain 43895) for each of the matched protein peaks derived from protein preparation #2 and analyzed by 2-D LC methods. . Line "——" = mean value for all of the protein pairs log ratio values. Line "– – – –" = the standard deviations.

was determined to have 19 upregulated protein peaks as compared to strain 43895OR while strain 43895OR was shown to have 22 protein peaks upregulated as compared to strain 43895OR.

MALD/I-TOF/TOF based identification of the differentially expressed protein spots from the 2-D GE separation and protein peaks from the 2-D LC separation was performed and the results are listed in Tables **1** and **2**. The protein spots from protein preparation #1, which were separated and collected from 2-D GE resulted in 6 proteins being identified out of the 7 proteins judged upregulated in strain 43895 compared to 43895OR. Additionally, 8 proteins were identified from the 11 proteins determined to be upregulated in strain 43895OR compared to 43895 (Table **1**). Likewise the separation of protein preparation #2 by 2-D GE yielded identifications for 9 of the 13 proteins upregulated for strain 43895 and 5 of the 10 proteins upregulated for strain 43895OR (Table **1**). Of the proteins analyzed by 2-D GE, there were 2 proteins observed to be differentially regulated

Table 1. Identities of Differentially Regulated Proteins After Separation by 2-D GE

2-D Gel ID	Expression Characteristic	Spot #	Protein ID	Accession #	Protein Score C.I. %
Protein prep #1	Upregulated in 43895OR	6	Tryptophan synthase alpha	H90857	100
		7	YaeH	S45230	99.757
		8	3-oxoacyl- synthase	AAN42714	97.621
		11	FkbP-type peptidyl-prolyl cis-trans isomerase FkpA	I65035	100
		22	imidazoleglycerol-phosphate synthase	AAN81007	100
		28	H+-transporting two-sector ATPase	CAA23519	99.996
		71	30s ribosomal protein S6	AAU37076	100
		90	Phosphoglyceromutase 1	H85577	99.664
	Upregulated in 43895	1	Glyceraldehyde 3-phosphate dehydrogenase A	AG0711	100
		4	Clp peptidase, chain A	1TYFA	100
		30	Catalase hpii, chain A	1IPHA	100
		45	Alcohol dehydrogenase	A64901	100
		59	UTP-glucose-1-phosphate uridylyltransferase, GalF	A64970	100
		78	ECU 36834	AAC43534	100
Protein prep #2	Upregulated in 43895OR	14	Acetyl-coenzyme A carboxylase	ACCA_ECOLI	99.97
		41	30S ribosomal protein S6	AAU37076	100
		43	Alkyl hydroperoxide reductase C22 subunit	D90709	100
		61	Outer membrane protein HlpA precursor	AAN78707	100
		100	Hypothetical protein Z3776	C85895	100
	Upregulated in 43895	9	Global response regulator	Q5PN99_SALPA	100
		15	Hyperosmotically inducible perisplasmic protein	H86136	100
		18	Triose-phosphate isomerase	1TMHA	100
		39	Outer membrane protein X	AAN79373	100
		40	YciE protein	C85762	100
		49	SpermidinePUTRESCINE-binding protein monomer	1POT	100
		52	Phosphoglycerate kinase	PGK_ECO57	100
		67	Clp peptidase, chain A	1TYFA	100
		75	Hypothetical protein YzzN	Q9R2E3_EC	100

Table 2. Identities of Differentially Regulated Proteins After Separation by 2-D LC

2-D LC ID	Expression Characteristic	Spot #	Protein ID	Accession #	Protein Score C.I. %
Protein prep #1	Upregulated in 43895	44	Hypothetical protein ECs5028 (YjbJ)	D91257	99.98
		146	Probable structural protein YciF	B85762	100
		155	Glyceraldehyde-3-phosphate dehydrogenase	BAA18884	100
		170	Hypothetical protein ECs3154	B91023	99.978
		228	Hypothetical protein ECs 1159	G90773	100
		230	Catalase hpii mutant YES, chain A	1QF7A	100
		264	30S ribosomal protein S2	RS2_ECO57	100
		294	DnaK molecular chaperone	AAN78519	100
Protein prep #2	Upregulated in 43895	28	Hypothetical protein ECs5028 (YjbJ)	D91257	100
		52	Adenylate kinase	GI|15800203	100
		92	50S ribosomal protein L28	GI | 15804178	100
		110	LsrG protein AI-2 modifying protein	GI |15801619	100
		148	Hypothetical protein Z5276	GI |15804356	100
	Upregulated in 43895OR	64	30S ribosomal protein S7	GI |15803854	100
		108	ATP synthase alpha subunit	GI |168779969	100

in both protein preparation #1 and protein preparation #2: 30s ribosomal protein S6 and Clp peptidase, chain A.

The protein peaks from protein preparation #1 separated by 2-D LC yielded protein identifications for 9 of the 27 protein peaks that were upregulated in strain 43895 compared to strain 43895OR. However, none of the 30 proteins upregulated in strain 43895OR return identifications, by MALD/I based techniques that had sufficient confidence levels. A second protein preparation yielded identifications of 5 of 19 proteins upregulated in strain 43895, and identified 2 of 22 proteins upregulated in strain 43895OR. A comparison of the identified differentially regulated proteins between the two sample series resulted in only one protein in common between the two preps: hypothetical protein Ecs5028(YjbJ).

The two different separation techniques, 2-D GE and 2-D LC yielded two identifiable proteins in common during the analysis of protein preparation #1: Glyceraldehydes-3-phosphate dehydrogenase and Catalase *hpii*, chain A. However, the two different separation techniques produced no identifiable proteins in common when analyzing the follow up protein preparations #2. The overlaps of the successfully identified differentially regulated proteins from the two different experimental series and from the two different separation techniques are summarized by means of a Venn diagram (Fig. **3**).

The hypothetical protein Ecs5028(YjbJ) was shown to be differentially regulated between 43895 and 43895OR by a great degree during multiple experiments using the 2-D LC technique. In order to see if the *yjbJ* gene was responsible for one of the primary phenotypic differences between strains 43895 and 438895OR, namely biofilm formation, a gene knockout of *yjbJ* was constructed in *E. coli* O157:H7 strain 43895. When the strains were compared in a biofilm assay, the inactivation of the *yjbJ* gene appeared to have no effect on the poor biofilm forming ability of 43895 as compared to its strong biofilm forming naturally occurring variant 43895OR (Fig. **4**).

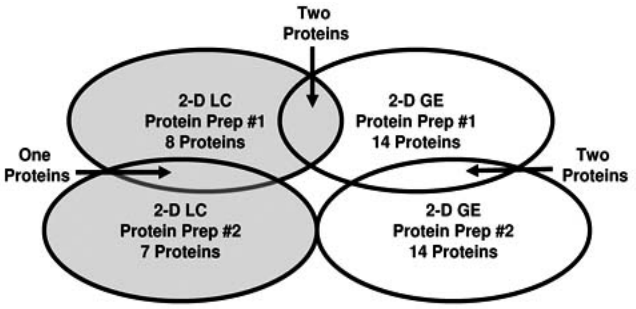


Fig. (3). Overview of the differentially regulated proteins identified. A Venn diagram detailing the overlaps in identified differentially regulated proteins between the two different separation methods and between the two separate experiment series.

DISCUSSION

Two dimensional gel electrophoresis and two dimensional liquid chromatography are two methods used for separating whole proteomes for comparison and identification. The technique of 2-D GE has been in use for a considerably longer time as compared to 2-D LC and is therefore more

widely utilized [1, 2]. However as the newer technique of 2-D LC begins to increase in popularity, it may be advantageous to compare and contrast these two methods in service to a bacterial research project. Therefore, our research detailed in this manuscript employing both of these methods illustrates some similarities, as well as some significant differences in the functionality of the two techniques and the results generated.

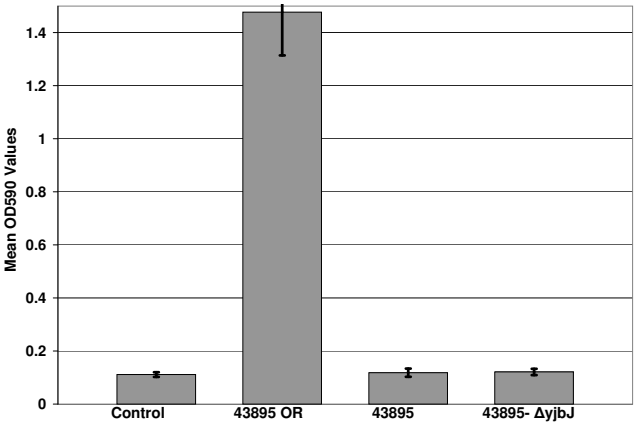


Fig. (4). The effect of *yjbJ* on biofilm formation. A comparison of biofilm formation by *E. coli* O157:H7 strains: 43895OR, 43895, and the isogenic mutant 43895 Δ*yjbJ*. Biofilms formed in LB-NS (no salt) media on polystyrene plates. The values for each strain are the average of three separate experiments with error bars representing the standard deviations for the multiple experiments.

The initial observation made when comparing the results of these two methods is that there is little overlap in terms of proteins identified from one technique as compared to the other technique. This is partially to blame on the fact that in this type of research there is limited overlap in proteins identified from one biological replicate to the next. The technical replicates for the 2-D GE could be combined to make a representative master gel and partial technical replicates demonstrated that the second dimension separations to be very reproducible for 2-D LC. However this did little to improve the reproducibility of the biological replicates. Until it is possible to increase the biological replicate consistency within a 2-D separation technique, it would be unreasonable to expect consistency between different separation techniques [7]. Additionally, the differences in the separation techniques employed by the two different methods make it unlikely that there will be considerable overlap in the proteins identified. Both techniques use the isoelectric point to separate the proteins in the first dimension [1-3]. However, 2-D gels separate proteins in the second dimension based on molecular weight, while 2-D LC uses protein hydrophobicity for separation [1-3]. An excellent example of the lack of consistency that the different separation techniques can cause is best exemplified by the hypothetical protein ECs5028 (YjbJ). The protein YjbJ was repeatedly identified as being differentially regulated by the 2-D LC technique but was never seen in any of the differentially regulated proteins visualized by 2-D GE. This occurs because the YjbJ protein has a relatively small molecular weight (8325 daltons). Proteins in that molecular weight range tend to run off of the bottom of a 10-20 % Tris-HCl gel during a standard run [12, 13]. The 2-D LC method, which utilizes the protein's hydrophobicity, tends to maintain these proteins as part of the proteome separation and is one additional potential reason for the lack of correlation between proteins identified by the two methods. Because of the significant differences in the identities of the proteins separated by the two methods it is unlikely that 2-D GE and 2-D LC should be thought of as competing techniques, but rather as complementary techniques. Similar conclusions were drawn by two other research groups focusing exclusively on eukaryotic systems [8, 9].

Additionally, when contrasting the two techniques in regards to a comparative proteomics assay it is obvious that the 2-D LC technique returns significantly more matched protein pairs than does the 2-D GE method. We believe this occurs because the 2-D LC method demonstrated greater sensitivity compared to the Commassie blue-stained 2-D gels. Additionally, we found it easier to align the protein peaks of discrete pH range delimited fractions from 2-D LC as compared to the multi-directional alignments necessitated by 2-D GE. Between pH fractions, protein peaks can be adjusted for alignment in only two directions, but 2-D gels have to be pulled in a multitude of directions to achieve a reasonable alignment between gels. This often results in a disproportionately larger number of unmatched proteins for 2-D GE. Problems with inter-gel variation and overlaying of 2-D GE gel images can be addressed to some degree by use of DIGE (differential in gel electrophoresis) 2-D GE techniques. However, only the basic techniques of 2-D GE and 2D-LC were compared in this study.

Despite the differences in protein pair numbers the resulting numbers of differentially expressed proteins successfully identified by the two methods was considerably closer to being equal. The use of the 2-D GE method results in 14 of 18 (78 %) differentially regulated proteins being identified for protein preparation #1 and 14 proteins of the 23 (61 %) differentially regulated proteins being identified for protein preparation #2. The 2-D LC method resulted in 8 proteins of the 56 differentially regulated proteins being identified from preparation #1 (14 %) and 8 of the 41 differentially regulated proteins for preparation #2 (15 %). The low rate of MALD/I-based protein identification of differentially regulated proteins separated by the 2-D LC method is probably the result of the 2-D LC system being able to detect proteins present in lower relative concentration compared to the 2-D GE method. The sensitivity of the Comassie blue-stain is such that those proteins visualized with it are generally plentiful enough to be identified by current MALD/I-based techniques [14]. The UV absorbance method used to visualize proteins in 2-D LC is more sensitive and therefore identifies a number of proteins that are not present in sufficient concentrations to be identified by MALD/I techniques. If we were to only consider the proteins separated by the 2-D LC method that were plentiful in concentration and therefore most likely to be identified by the MALD/I system we would see the number of differentially regulated proteins separated by this method reduced to 37 differentially regulated proteins for preparation #1 and 29 differentially regulated proteins for protein preparation #2. Additionally protein peaks in 2-D LC techniques can span multiple pH fractions resulting in the multiple protein peaks identified as differentially regulated

but which are in the end, all identified as the same protein. When one takes into account proteins that span multiple pH fractions, the numbers of unique differentially-regulated proteins are probably closer to 28 for protein preparation #1 and 23 unique differentially-regulated proteins for preparation #2.

The 2-D LC method exhibits another characteristic, which makes identifying proteins less successful. The proteins separated by the 2-D LC technique are collected in timed liquid fractions. Therefore, proteins falling relatively close together can easily be collected into the same fraction. The mixed protein fractions result in mixed peptide samples which can be confounding to MALD/I based identification since the competing peptides from multiple proteins make for peptide fingerprints that cannot be reasonably assigned to a single protein with sufficient confidence. Additionally it is also possible to identify the secondary protein and not the protein of interest. This can result in assigning characteristics (differential regulation under a specific condition) to the protein that was not exhibiting these traits and was merely collected in a fraction with the responsible protein. This does not happen as frequently in 2-D gels given that protein spots can generally be excised from a gel in a manner, which results in only one protein in the sample. Thus, further separation techniques or different peptide sequence analysis methods for identification of proteins from the 2-D LC system are often required.

Since 2-D GE and 2-D LC should be viewed as complementary techniques, the proteins identified in both protein preparations by either or both of the separation techniques are the targets of further investigations of protein expression differences of *E. coli* O157:H7 strains 43895 and 43895OR. The 30S ribosomal protein S6 was upregulated in strain 43895OR compared to strain 43895. While the Clp peptidase, chain A and hypothetical protein ECs5028 (YjbJ) proteins were unregulated in strain 43895 compared to 43895OR. A knockout of the gene *yjbJ* was constructed in the hope of identifying the role of the gene product in the phenotypic differences between 43895 and 43895OR. Since the poor biofilm forming 43895 made a large amount of YjbJ and the strong biofilm forming, red, dry, rough strain 43895OR made little or no YjbJ it was reasonable to believe that interrupting the *yjbJ* gene in 43895 might increase the strain's ability to form biofilms and shift the colony appearance towards red, rough, and dry. Surprisingly, when the isogenic mutant 43895 Δ*yjbJ* was constructed, it resembled its parent strain 43895 in both colony appearance and biofilm forming ability. The lack of an obvious phenotypic change has resulted in a failure to identify the function of *yjbJ* and the role it might play in the development of the naturally occurring variant 43895OR.

CONCLUSIONS

In the current study both 2-D LC and 2-D GE were lacking in biological replicate reproducibility. The 2-D LC technique appears to be more sensitive than 2-D GE Commassie blue stained gels and demonstrates greater ease in the alignment of protein data between comparative separations resulting in more matched protein pairs for investigation. However, we were able to identify by the MALD/I technique a slightly greater number of proteins separated using 2-D GE compared to 2-D LC. This was primarily the result of the detection sensitivity of the 2-D LC system not translating directly to sufficient material for MALD/I based identification. Additionally, the 2-D LC collection system frequently produced mixed protein samples resulting in mixed spectra that could not be sufficiently deconvoluted for identification. The mixture of positive and negative features for both the 2-D GE and 2-D LC methods along with the differences in the differentially regulated proteins identified by each method suggests that the two techniques should be used as complimentary systems and are not suitable replacements for one another.

ABBREVIATIONS

2-D GE	=	Two-dimensional gel electrophoresis
2-D LC	=	Two-dimensional liquid chromatography
MALD/I	=	Matrix-assisted laser desorption/ionization
TOF	=	Time of flight
YESCA	=	Yeast extract casamino acids
TCEP	=	Tris-carboxyethyl phosphine hydrocholine
TCA	=	Trichloroacetic acid
TBP	=	Tributylphosphine
IEF	=	Isoelectric focusing
MS	=	Mass spectra
LB-NS	=	LB growth medium with no salt added.

CONFLICT OF INTERESTS

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

ACKNOWLEDGEMENTS

We would like to thank Jonnee Almond, Laurie Fortis and Bryan Cottrell for technical assistance. Additionally, we would like to thank Dr. Pina Fratamico for reviewing this manuscript.

REFERENCES

[1] O'Farrell, P. H. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.*, **1975**, *250* (*10*), 4007-21.

[2] Gorg, A.; Obermaier, C.; Boguth, G.; Harder, A.; Scheibe, B.; Wildgruber, R.; Weiss, W. The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis*, **2000**, *21* (*6*), 1037-53.

[3] Lubman, D. M.; Kachman, M. T.; Wang, H.; Gong, S.; Yan, F.; Hamler, R. L.; O'Neil, K. A.; Zhu, K.; Buchanan, N. S.; Barder, T. J. Two-dimensional liquid separations-mass mapping of proteins from human cancer cell lysates. *J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci.*, **2002**, *782* (*1-2*), 183-96.

[4] Uhlich, G. A.; Cooke, P. H.; Solomon, E. B. Analyses of the red-dry-rough phenotype of an Escherichia coli O157:H7 strain and its role in biofilm formation and resistance to antibacterial agents. *Appl. Environ. Microbiol.*, **2006**, *72* (*4*), 2564-72.

[5] Uhlich, G. A.; Keen, J. E.; Elder, R. O. Variations in the csgD promoter of Escherichia coli O157:H7 associated with increased virulence in mice and increased invasion of HEp-2 cells. *Infect Immun.*, **2002**, *70* (*1*), 395-9.

[6] Uhlich, G. A.; Keen, J. E.; Elder, R. O. Mutations in the csgD promoter associated with variations in curli expression in certain strains of Escherichia coli O157:H7. *Appl. Environ. Microbiol.*, **2001**, *67* (*5*), 2367-70.

[7] Yan, F.; Subramanian, B.; Nakeff, A.; Barder, T. J.; Parus, S. J.; Lubman, D. M. A comparison of drug-treated and untreated HCT-116 human colon adenocarcinoma cells using a 2-D liquid separa-

tion mapping method based upon chromatofocusing PI fractionation. *Anal. Chem.*, **2003**, *75* (*10*), 2299-308.

[8] McDonald, T.; Sheng, S.; Stanley, B.; Chen, D.; Ko, Y.; Cole, R. N.; Pedersen, P.; Van Eyk, J. E. Expanding the subproteome of the inner mitochondria using protein separation technologies: one- and two-dimensional liquid chromatography and two-dimensional gel electrophoresis. *Mol. Cell Proteomics*, **2006**, *5* (*12*), 2392-411.

[9] Komatsu, S.; Zang, X.; Tanaka, N. Comparison of two proteomics techniques used to identify proteins regulated by gibberellin in rice. *J. Proteome Res.*, **2006**, *5* (*2*), 270-6.

[10] Gunther, N. W. 4th; Nunez, A.; Fortis, L.; Solaiman, D. K. Proteomic based investigation of rhamnolipid production by Pseudomonas chlororaphis strain NRRL B-30761. *J. Ind. Microbiol. Biotechnol.*, **2006**, *33* (*11*), 914-20.

[11] Horton, R. M.; Hunt, H. D.; Ho, S. N.; Pullen, J. K.; Pease, L. R. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene*, **1989**, *77* (*1*), 61-8.

[12] Langen, H.; Roder, D.; Juranville, J. F.; Fountoulakis, M. Effect of protein application mode and acrylamide concentration on the resolution of protein spots separated by two-dimensional gel electrophoresis. *Electrophoresis*, **1997**, *18* (*11*), 2085-90.

[13] Fountoulakis, M.; Juranville, J. F.; Roder, D.; Evers, S.; Berndt, P.; Langen, H. Reference map of the low molecular mass proteins of Haemophilus influenzae. *Electrophoresis*, **1998**, *19* (*10*), 1819-27.

[14] Scheler, C.; Lamer, S.; Pan, Z.; Li, X. P.; Salnikow, J.; Jungblut, P. Peptide mass fingerprint sequence coverage from differently stained proteins on two-dimensional electrophoresis patterns by matrix assisted laser desorption/ionization-mass spectrometry (MALDI-MS). *Electrophoresis*, **1998**, *19* (*6*), 918-27.

Received: March 15, 2010 Revised: June 08, 2010 Accepted: June 09, 2010