ABRF-PRG07: Advanced Quantitative Proteomics Study

A.M. Falick[1], W.S. Lane[2], K.S. Lilley[3], M.J. MacCoss[4], B.S. Phinney[5], N.E. Sherman[6], S.T. Weintraub[7], H.E. Witkowska[8], N.A. Yates[9]

[1]HHMI, Univ. of California, Berkeley CA, [2]Harvard Univ., Cambridge MA, [3]Univ. of Cambridge, Cambridge UK, [4]Univ. of Washington, Seattle WA, [5]Univ. of California, Davis CA, [6]Univ. of Virginia, Charlottesville VA, [7]Univ. of Texas Health Science Center at San Antonio TX, [8]Univ. of California, San Francisco CA, [9]Merck Research Laboratories, Rahway NJ

Goal

The 2007 Proteomics Research Group (PRG) study was designed to assist ABRF members and interested scientists to evaluate their proficiency in identifying and quantifying proteins in a mixture. From a comparison of the results of this study, participating laboratories will be able to gauge their own capabilities and also gain insight into the relative success of the different quantitative approaches used.

Abstract

A major challenge for core facilities today is determining quantitative protein differences across complex biological samples. The determination of relative protein amounts is commonly requested in many core facilities. Although there are a variety of techniques in the literature for determining relative and absolute protein quantitation, most if not all are non-routine and can be very challenging to implement effectively. In addition, it is unknown how accurate these techniques are in practice across multiple laboratories having varied levels of expertise. The PRG began looking at some of these variables in last year's study, but many questions were necessarily left unanswered. In particular, the performance of the methods in the presence of a complex mixture of background proteins was not investigated.

This year's study sample consisted of three tubes, each containing a complex background mixture and up to 15 added known proteins at varying ratios. All of the background proteins were present at the same concentration in each of the three tubes.

The participating laboratories were challenged to identify the proteins and to determine their relative quantities in the three samples.

The primary goal of this study was to allow each laboratory to evaluate its capabilities and approaches with regard to:

- Detection and identification of proteins spiked in different amounts into samples that also contain complex mixtures of background proteins
- Determination of the relative amounts of the spiked proteins in the three samples
- Comparison of the number of correct and incorrect results as an indication of the overall performance of the methodologies used

The PRG also collected and compiled supplementary information about the strategies used by each participant in order to provide an overview of the protocols and techniques used and to aid in the development of optimized protocols for these techniques.

Table 1. Proteins in PRG07 study samples

Protein[2]	Accession Number[3]	M.W. (kDa)	Quantity (pmol)[1] A	B	Ratio (B/A)
Carbonic anhydrase I	69	28.9	2.50	1.14	0.45
Catalase	465	57.5	0.50	0.34	0.67
Cytochrome *c*	3870	13.0	2.50	11.50	4.60
Glucose oxidase	152	80.0	0.50	0.33	0.67
Glycerokinase[4]	904	54.0	2.50	0.78	0.31
Hexokinase	2938	50.0	0.50	0.16	0.31
Horseradish peroxidase	2466	43.3	5.00	11.00	2.20
Lactoperoxidase	2648	77.5	2.50	0.78	0.31
Myoglobin	161	16.5	0.50	5.00	10.00
Serum albumin	1213	66.6	5.00	3.33	0.67
Tryptophanase[4]	2366	51.0	5.00	1.56	0.31
Ubiquitin	4014	8.7	5.00	23.00	4.60

[1]Sample C contained the same quantities of protein as sample B.
[2]Proteins were purchased from Sigma-Aldrich.
[3]Accession number in PRG database
[4]Added *E. coli* proteins

Sample Preparation

The *E. coli* lysate and all added proteins were purchased from Sigma-Aldrich Chemical Co. (St. Louis MO). Stock solutions of the *E. coli* lysate and the individual protein samples were prepared at a concentration of 1 mg/mL in 5% acetic acid/10% acetonitrile in deionized water. Protein purity values supplied by the vendor were used in the concentration calculations. The *E. coli* stock solution was apportioned into two conical tubes (A, 25 mL; B/C, 50 mL) and appropriate volumes of the individual protein stock solutions were added to the two tubes to produce the study test mixtures. One hundred microliter aliquots of A and B/C were transferred to the corresponding polypropylene sample tubes labeled A, B, and C. The mixtures were dried in a vacuum centrifuge and stored at -80 °C prior to mailing.

Results

Data submission. Samples were requested by 87 laboratories; 43 participants (ABRF members, 22; non-members, 21) submitted datasets, corresponding to a 49% return rate. Surveys from 8 of the respondents did not contain any quantitative data.

Database. The PRG provided an anonymized database with the sequences of the added and background proteins as well as several decoy entries. The database was available for download and was on the ProteinProspector, Mascot and X! Tandem websites.

Methods used. Mass spectrometry was used for protein identification for all samples. Gel-based and gel-free approaches were employed for quantification.

Method[1]	Number
Gel-based (28 %)	
2D Coomassie	1
2D silver-stain	1
2D fluorescence	3
2D DIGE	5
MS/isotope (55 %)	
iTRAQ	16
$^{16}O/^{18}O$	2
ICAT	1
ICPL	1
Label-free (17 %)	6

[1]DIGE, difference gel electrophoresis; iTRAQ, isobaric tags for relative and absolute quantitation; ICAT, isotope coded affinity tag; ICPL, isotope coded protein label

Calculations. As a numerical assessment of how close the submitted results for the relative quantities of a protein in samples A and B were to the expected ratio, the following calculations were made:

***% error of ratio* = [(observed ratio – expected ratio)/expected ratio] x 100**

Conclusions

A summary of observations made from the study results is shown below. It is important to remember that the number of responses for each method was limited. Therefore, no firm conclusions can be drawn regarding the optimal method for assessing differences in protein quantities between different samples.

Results that were close to the expected values were reported by a few of the participants, indicating that quantitative assessment of complex samples is achievable.

Participants using similar techniques did not obtain comparable results, suggesting that experience is a key factor in the overall process.

About one-third of participants were able to identify and detect differences in 5 out of 10 added proteins (not including the *E. coli* proteins).

Differences in the two added *E. coli* proteins were not detectable by the participants, most likely due to the high endogenous levels of these proteins.

Direct comparisons of different approaches cannot be made on the basis of the results of this study because of the limited number of participants and the apparent dependence on experience.

In general the most accurate results were submitted by the most experienced laboratories.

Successful quantitative proteomics requires a combination of instrumentation, experienced personnel and informatics support.

Caveats

This was a model study in which biological variability was not a factor. In "real-life" samples, biological variability among samples would contribute significantly to the difficulty of the analysis.

The only non-*E. coli* proteins in the PRG database other than common contaminants were the proteins included in the mixture spiked into the *E. coli* lysate. It would have been possible, therefore, to deduce the identity of these spikes by careful examination of the database.

Acknowledgements

The PRG gratefully acknowledges the assistance of the following people: Kevin Hakala (UTHSCSA), initial gel analyses; Michelle Salemi (UC Davis), sample management; Dr. Rich Eigenheer (UC Davis), HPLC-ESI-MS/MS analyses; and Ekaterina Deyanova (Merck), HPLC-ESI-MS/MS analyses. The PRG also appreciates the hard work of the participants and their willingness to complete the results survey, even when no proteins were identified.

For additional information, please visit http://www.abrf.org/PRG

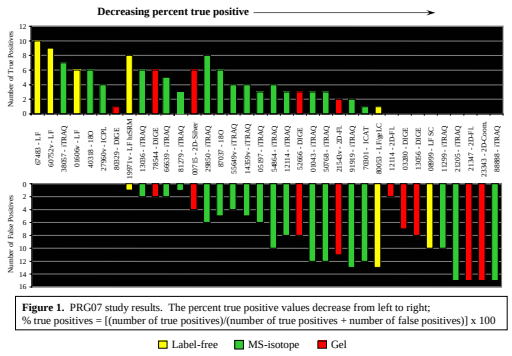

Figure 1. PRG07 study results. The percent true positive values decrease from left to right; % true positives = [(number of true positives)/(number of true positives + number of false positives)] x 100

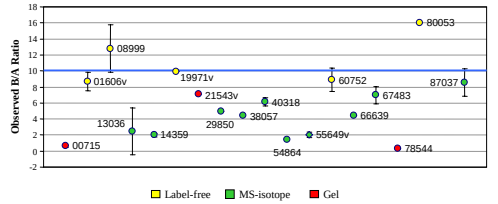

Figure 2. Example of observed B/A ratios for one of the spiked proteins. Shown above are the results for myoglobin; the expected ratio (10:1) is indicated by a blue line. For this study, 12 proteins were added to the *E. coli* lysate.

Table 2. PRG07 Results Compilation

		Added Protein[1] Myoglobin	Ubiquitin	Cyt. *c*	Horseradish peroxidase	Serum albumin	Catalase	Glucose oxidase	Carbonic anhydrase I	Trypto-phanase	Glycero-kinase	Lacto-peroxidase	Hexokinase
	Quantity in sample B (pmol)	5.00	23.00	11.50	11.00	3.33	0.34	0.33	1.14	1.56	0.78	0.78	0.16
	Quantity in sample A (pmol)	0.50	5.00	2.50	5.00	5.00	0.50	0.50	2.50	5.00	2.50	2.50	0.50
	Ratio B/A	10.00	4.60	4.60	2.20	0.67	0.67	0.67	0.45	0.31	0.31	0.31	0.31

Identifier[2]	No. Prot. identified	TP[3]	FP[4]	Method[5]	Days[6]	Percent error of ratio[7] Myoglobin	Ubiquitin	Cyt. *c*	Horseradish peroxidase	Serum albumin	Catalase	Glucose oxidase	Carbonic anhydrase I	Trypto-phanase	Glycero-kinase	Lacto-peroxidase	Hexokinase
67483	10	10 (6)		LF	10	30.2	0.7	10.9	12.3	7.0	-19.4	-8.0	-1.3			-12.9	-316
19971v	8	8 (6)	1	LF-hrSRM	9	0.3	-10.7	-9.8	-13.6	-26.0	-7.5		-123			-148	
29850	8	8 (7)	6 (2)	iTRAQ	7	50.9	18.7	18.9	2.3	-11.0	-16.3		-16.7			-19.4	
60752v	9	9 (9)		LF	7	10.9	6.5	-3.7	10.0	22.0	29.9	28.0	18.5			6.5	
01606v	6	6 (5)		LF	7	13.3	3.5	-6.7	-1.8	-17.0			-7.9				
87037	7	6 (6)	5 (3)	18O	5	14.4	25.4	27.0	-6.8	-14.0					-63.2	-51.6	
40318	6	6		18O	7	38.4	-9.8	35.4	3.6	11.5			-299				
38057	7	7		iTRAQ	5	55.6	41.7	42.0	24.5	-20.0	4.5		-41.0				
27960v	4	4 (3)		ICPL	10		-17.2		-17.7	-18.5			-23.3				
00715	6	6 (4)	4 (3)	2D-S		93.6	46.1	78.9		-48.5			-1.3			6.5	
13036	6	6 (5)	2	iTRAQ	10	75.4	40.2	42.0	26.8	-15.5			-10.1				
78544	6	6 (6)	2	DIGE	2	97.2	92.2		69.5	-112	-7.5		-294				
66639	5	5 (5)	2	iTRAQ	4	55.8	63.0		36.4	-24.5			-25.6				
52666	3	3 (2)	8 (2)	DIGE	6					13.0			14.1		-156		
12114	3	3	8 (1)	iTRAQ	4			5.0	35.9	13.0							
05197	3	3	6	iTRAQ	8		33.0	29.1	8.2								
01043	3	3	12	iTRAQ	3			54.8	32.3				-25.6				
54864	4	4 (4)	10 (6)	iTRAQ	4	85.4	65.7	68.0		-26.0							
50768	3	3 (2)	12 (5)	iTRAQ	14		82.0		74.5					-47.2			
14359v	4	4 (3)	5	iTRAQ	5	80.0	60.2	60.2	26.8								
91919	2	2 (1)	13	iTRAQ	6				64.5					-732			
55649v	4	4 (3)	4	iTRAQ	4	79.7	58.9	55.9	27.7								
80053	1	1	13	LF-gel/LC	15	-60.0											
80329	1	1		DIGE	6				-31.8								
21543v	2	2	11	2D-FI	21	28.6							29.5				
08999	1		10 (2)	LF-count+	10	-27.9											
70301	1	1 (1)	12	ICAT	8					-295							
21205	1		15 (10)	iTRAQ	4							41.5					
81279		3 (2)	1 (1)	iTRAQ	5												
88888			15	iTRAQ	6												
13056			8	DIGE	10												
23343			15 (14)	2D-C	3												
12114			2	2D-FI	10												
21347			15 (14)	2D-FI	25												
03280			7 (7)	DIGE	10												
11299			10 (3)	iTRAQ	7												

8 respondents did not return quantitative data; those entries were not included in this table

[1]The indicated quantities of proteins were added to an *E. coli* lysate as described in Methods; the quantities shown above for glycerokinase and tryptophanase (both are *E. coli* proteins) represent the amounts added to the mixture. Sample C contained the same quantities of protein as sample B.
[2]v, instrument vendor
[3]True positives, the number of proteins correctly identified as being present at different relative levels in samples A and B; the number of TP results considered to be "high confidence" by the respondent is shown in parenthesis.
[4]False positives, the number of proteins incorrectly identified as being present at different relative levels in samples A and B; the number of FP results considered to be "high confidence" by the respondent is shown in parenthesis.
[5]Methods: LF, label-free; LF-count+, label-free-spectral count+differential peptide ion intensity integration; iTRAQ, isobaric tags for relative and absolute quantitation; ICAT, isotope coded affinity tag; ICPL, isotope coded protein label; 18O, $^{16}O/^{18}O$ labeling; 2D-S, 2D-PAGE/silver stain; 2D-C, 2D PAGE/Coomassie, 2D-FI; 2D PAGE/fluorescent stain
[6]The amount of time required for the study, as provided by participants in the results survey
[7]% error of ratio = [(observed ratio – expected ratio)/expected ratio] x 100; results that were greater than the expected ratio are shown as positive values; observed ratios that are less than expected are negative numbers.

The PRG strongly points out that the data received from the study participants are not intended to promote any particular method or type of equipment. The study and survey were undertaken with the goals of helping member laboratories and others to both improve and expand the range of their own capabilities and to provide them with a means of testing the techniques that they use. Furthermore the number of submitted responses was insufficient to afford a statistically significant measure of the ability of any method to "get the correct answer." We also point out that in some cases it is likely that the results represent the current experience levels of the scientists who performed the analyses and not the absolute capabilities of the methods used because some of the participating laboratories were conducting these analyses for the first time. Any representation to the contrary of what is stated above is the responsibility of the entity making that representation, and the PRG explicitly does not endorse any such representation.