ABRF-sPRG2006 Study: A Proteomics Standard

A B R F
Proteomics Standards Research Group

P.C. Andrews[1], D.P. Arnott[2], M.A. Gawinowicz[3], J.A. Kowalak[4], W.S. Lane[5], K.S. Lilley[6], L.T. Martin[7] and S.E. Stein[8]

[1]University of Michigan School of Medicine, Ann Arbor MI; [2]Genentech, Inc, So. San Francisco CA; [3]Columbia University, New York NY; [4]National Institute of Mental Health, Bethesda MD; [5]Harvard University, Cambridge MA; [6]Cambridge University, Cambridge UK; [7]East-West University, Chicago, IL; [8]National Institute of Standards and Technology, Gaithersburg MD

Abstract

A principal task of proteomics laboratories is protein identification, often involving constituents of complex mixtures. The number of laboratories entering the field of proteomics research continues to expand at an impressive rate, while at the same time, advances in instrumentation and bioinformatics tools rapidly move forward. A clear need exists for a reasonably complex, well-defined mixture of proteins to serve as a reference standard for labs of all experience levels to evaluate their performance and to use in methods development.

The Proteomics Standards Research Group (sPRG) developed the following prototype standard protein mixture, sPRG2006:

- **49 human proteins**
- **5 pmoles per protein**

Study statistics:

- **120 laboratories requested the standard**
- **74 laboratories returned data**

The sample was distributed without revealing the exact number and identity of the proteins. The instruction was to use any proteomics platform desired.

Introduction

Why we need a proteomics standard:

- **To objectively evaluate a laboratory's ability to perform this type of analysis.**
- **To provide a way for new laboratories to evaluate their performance relative to laboratories with extensive experience.**
- **To be useful for comparative evaluation of instrument performance and bioinformatic analysis of data.**

The sPRG2006 standard has been developed to assist in the evaluation of:

- **Separations technologies employed in proteomics analysis**
- **Methodologies used to identify proteins**
- **Bioinformatics tools used to consolidate protein identifications**

Methods

- **Human proteins were purified from their biological source or expressed as a recombinant.**
- **Subjected to multiple analytical methodologies, e.g., 1D PAGE, IEF, and RP-HPLC, in order to assess required levels of purity (≥ 95%).**
- **Protein concentration was determined by amino acid analysis.**
- **Five picomole aliquots of each protein were combined and lyophilized in a 1 mL polypropylene tube.**
- **Prior to distribution to participating labs, the sPRG2006 protein standard was distributed to the sPRG committee members for validation.**
- **Three RG laboratories analyzed using a shotgun approach by digestion with trypsin of the entire protein mixture followed by on-line 1D LC-MS/MS, on-line 2D LC-MS/MS, off-line capillary RP-HPLC, TOF/TOF**
- **Two RG laboratories separated the protein mixture by 1D SDS-PAGE, followed by in-gel tryptic digestion and on-line 1D LC-MS/MS.**

For more information about this study, please visit

www.abrf.org/sprg

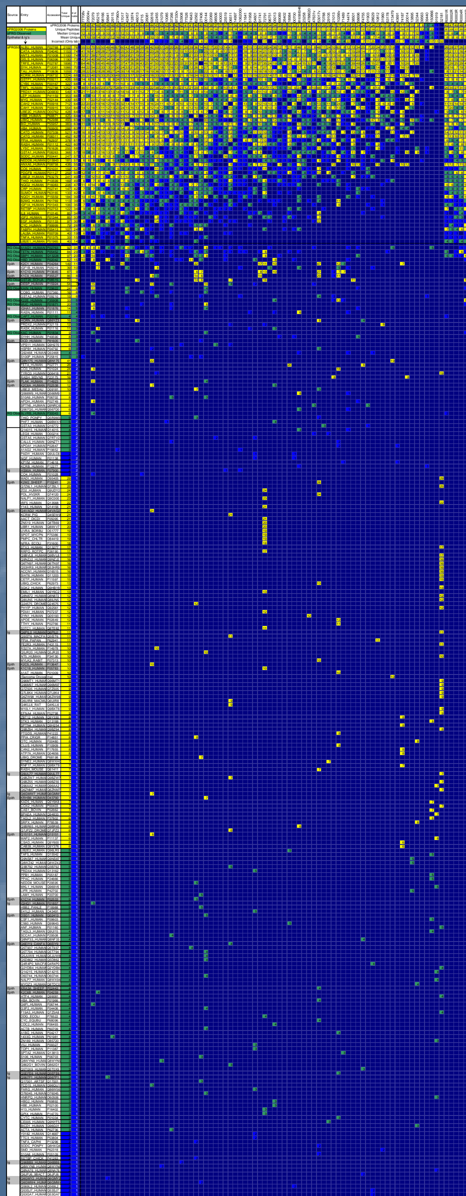


Table 1. Heat Map of Proteins Identified in the sPRG2006 Proteomics Standard

- **Columns: 74 labs, left to right, in descending no. of proteins identified.**
  - **For reference, the five sPRG member labs' results are segregated in the rightmost columns.**
- **Rows: Proteins identified, descending order of number of responses.**
  - **Left-hand source protein columns are color coded as follows:**

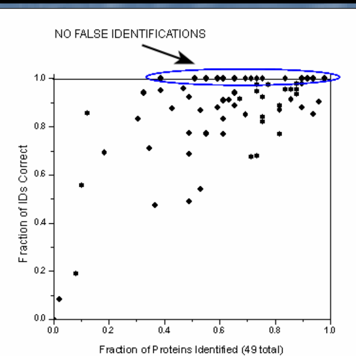
sPRG2006 Proteins	Constituents of sPRG2006
sPRG Observed	Proteins observed by 2 or more sPRG labs
Epithelial & Ig's	Epithelial (keratin, hornerin, etc.) & Ig's

- **Cells: Number of unique peptides:**

1 - 2   3 - 5   > 5

Protein Identification Performance for 74 Labs.

In this ROC curve the performance of each lab is represented by a point defined by the fraction of all known proteins identified and the fraction of all identifications that were correct.



Conclusions

- **A standard protein mixture has been developed that has broad usefulness for a variety of proteomics strategies.**
- **No approach performed better than any other: Success was possibly experience-, or technical ability-dependent.**
- **Good results are achievable by a lab which does not have the latest instruments but optimizes variables within its control.**
- **Many labs can reliably identify a large fraction of proteins at these concentrations with few false positive results.**
- **The variability expected in a semi-complex protein mixture analysis needs to be dealt with in future standards design and analysis.**
- **We highlighted the need for a community standardized method for data reporting.**
- **The study has led to a publicly available set of raw data files for further analysis.**

Acknowledgments

The sPRG thanks Sigma-Aldrich for their significant contribution to this study.

Table 2. Accuracy of Identifications

Accuracy = Correct Identifications / (Correct + Incorrect Identifications).

Overall = Percentage of Correctly Identified Proteins x Accuracy of Identification.

(See ROC curve)

Protein Identification Summary

Total Correct / (Correct + Incorrect) → Accuracy

% Identified x Accuracy → Overall

Lab	sPRG2006 Identified: Proteins	%	Incorrect	Accuracy	Overall	Unique Peptides: Peptides	Median	Mean
sPRG								
R0001R	49					381	6	8
R0002R	47					408	6	9
R0003R	49					317	4	6
R0004R	48					390	6	8
R0005R	48					645	8	13
77708v	48	98%	0	100%	98%	808	11.5	16.8
29504v	48	98%	0	100%	98%	497	6.0	10.4
72079	47	96%	5	90%	87%	883	13.0	18.8
65215	46	94%	0	100%	94%	646	9.0	14.0
98166	46	94%	0	100%	94%	576	9.0	12.5
58941	46	94%	0	100%	94%	568	9.5	12.3
29115	46	94%	8	85%	80%	562	8.5	12.2
27960v	45	92%	0	100%	92%	712	11.0	15.8
17017	45	92%	0	100%	92%	702	12.0	15.6
42457	45	92%	0	100%	92%	547	10.0	12.2
12874	45	92%	0	100%	92%	442	7.0	9.8
46013	44	90%	0	100%	90%	520	8.5	11.8
0715	44	90%	0	100%	90%	477	8.0	10.8
26081	44	90%	1	98%	88%	349	5.5	7.9
92616	44	90%	6	88%	79%	419	5.5	9.5
22455	43	88%	1	98%	86%	631	10.0	14.7
21079	43	88%	2	96%	84%	347	5.0	8.1
10085	43	88%	2	96%	84%	222	4.0	5.2
37266	43	88%	3	93%	82%	306	5.0	7.1
97330v	42	86%	2	95%	82%	36	5.0	6.2
53908	42	86%	4	91%	78%	340	5.5	8.1
77526	41	84%	0	100%	84%	185	3.0	4.5
74643	41	84%	2	95%	80%	381	6.0	9.3
62562	40	82%	5	89%	73%	462	8.0	11.6
27406	40	82%	6	87%	71%	219	4.5	5.5
00144	40	82%	12	77%	63%	570	7.5	14.2
22069	38	78%	1	97%	76%	195	4.0	5.1
98506	37	76%	0	100%	76%	160	3.0	4.3
40003	37	76%	3	93%	70%	118	2.0	3.2
25519	37	76%	7	84%	63%	177	4.0	4.8
06511	37	76%	8	82%	62%	338	7.0	9.1
14997	36	73%	0	100%	73%	190	4.0	5.3
1062000	36	73%	0	100%	73%	36	1.0	1.0
11641	36	73%	1	97%	71%	221	4.0	6.1
51958	36	73%	2	95%	70%	198	4.0	5.5
54601	36	73%	17	68%	50%	290	5.0	8.1
91741	35	71%	0	100%	71%	328	8.0	9.4
72791	35	71%	17	67%	48%	410	8.0	11.7
96751	34	69%	0	100%	69%	224	4.5	6.6
05013	34	69%	6	85%	59%	235	4.5	6.9
69089	33	67%	3	92%	62%	109	2.0	3.3
26402	32	65%	0	100%	65%	80	1.0	2.5
18984	32	65%	2	94%	61%	253	7.0	7.9
00700	32	65%	4	89%	58%	100	2.5	3.1
8165883487	31	63%	3	91%	58%	337	6.0	10.9
23258	30	61%	0	100%	61%	156	4.0	5.2
4788620	30	61%	3	91%	56%	158	3.5	5.3
47631v	30	61%	6	83%	51%	122	3.5	4.1
53017v	30	61%	9	77%	47%	371	7.0	12.4
70788	29	59%	0	100%	59%	111	3.0	3.8
29105	29	59%	4	88%	52%	213	6.0	7.3
21013	27	55%	0	100%	55%	222	7.0	8.2
10266v	27	55%	8	77%	43%	383	8.0	14.2
71489	26	53%	4	87%	46%	128	4.0	4.9
14474	26	53%	22	54%	29%	95	2.0	3.7
38501	25	51%	0	100%	51%	283	8.0	11.3
58696	24	49%	2	92%	45%	317	5.0	13.2
F4755	24	49%	7	77%	38%	191	6.0	8.0
01903	24	49%	11	69%	34%	70	2.0	2.9
28629	24	49%	25	49%	24%	162	3.5	6.7
10812	23	47%	1	96%	45%	311	12.0	13.5
31960	21	43%	3	88%	38%	77	3.0	3.7
53178	19	39%	0	100%	39%	137	6.0	7.2
213479	19	39%	1	95%	37%	179	9.0	9.4
43691	18	37%	20	47%	17%	18	1.0	1.0
74187	17	35%	7	71%	25%	156	9.0	9.2
20014	16	33%	1	94%	31%	193	11.0	12.1
13895	15	31%	3	83%	26%	117	7.0	7.8
32344	9	18%	4	69%	13%	101	12.0	11.2
39563	6	12%	1	86%	10%	13	1.5	2.2
39440	5	10%	4	56%	6%	5	1.0	1.0
150866	4	8%	17	19%	1.6%	23	5.5	5.8
91919	1	2%	11	8%	0.2%	1	1.0	1.0
32181	0	0%	32	0%	0%	0	0.0	0.0