ABRF-sPRG2006 Study: A Proteomics Standard

P.C. Andrews1, D.P. Annett2, M.A. Gawinowicz3, J.A. Kowalak4, W.S. Lane5, K-S. Lilley6, L.T. Martin7 and S.E. Stein8

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

Abstract

A principal task of proteomics laboratories is protein identification, often involving constraints 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 panels 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 test any proteomics platform desired.

Introduction

Why we need a proteomics standard:

- To objectively evaluate a lab's ability to perform this type of analysis.
- To provide a way for new labs 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 recombinant
  - Subjected to multiple analytical methodologies, e.g., ID PAGE, IEF, and RP-HPLC, in order to ensure required levels of purity (>95%)
  - Protein concentration was determined by amino acid analysis
  - Five peptide aliquots of each protein were combined and lyophilized in a 1 ml polypropylene vial.

Prior to distribution to participating labs, the sPRG2006 protein standard was distributed to the sPRG committee members for validation.

- Three RIU laboratories analyzed using a shotgun approach by digestion with tryptic and non-tryptic protein mixtures followed by either 1D LC-MS/MS, or even 2D LC-MS/MS, offline capillary RP-HPLC, TMT-TOF
- Two RIU laboratories separated the protein mixture by DE-520-PAGE, followed by trypsin digestion and online 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

<table>
<thead>
<tr>
<th>Column: Number of unique peptides</th>
<th>Protein</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A1AT_HUMAN</td>
<td>96%</td>
</tr>
<tr>
<td>2</td>
<td>IGF1B_HUMAN</td>
<td>83%</td>
</tr>
<tr>
<td>3</td>
<td>ALB_HUMAN</td>
<td>80%</td>
</tr>
<tr>
<td>4</td>
<td>A1AT_HUMAN</td>
<td>77%</td>
</tr>
</tbody>
</table>

Table 2. Accuracy of Identifications

| Protein     | TAF5_HUMAN  | LIFR_HUMAN  | C560_HUMAN  | TAGL3_HUMAN  | SCCA1_HUMAN  | VILI_HUMAN  | HBE_HUMAN  | Y0182_HUMAN  | TAF5_HUMAN  | LIFR_HUMAN  | C560_HUMAN  | TAGL3_HUMAN  | SCCA1_HUMAN  | VILI_HUMAN  | HBE_HUMAN  | Y0182_HUMAN  |
|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Accuracy (%)| 98%         | 97%         | 96%         | 95%         | 94%         | 93%         | 92%         | 91%         | 90%         | 89%         | 88%         | 87%         | 86%         | 85%         | 84%         | 83%         |

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 reliably identify a large fraction of proteins at these concentrations with few false positive results.
- The variability expected in a non-replicable 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 mining.
- 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.