ABRF-sPRG2009 Study: Development of Quantitative Proteomics Standards

D Arnott[1], JG Farmar[2], AR Ivanov[3], JA Kowalak[4], WS Lane[5], K Mechtler[6], BS Phinney[7], RR Ogorzalek Loo[8], MR Raida[9], and ST Weintraub[10]

[1]Genentech, Inc., So. San Francisco, CA; [2] Barnard College, Dept. of Chemistry, New York, NY; [3]Harvard University School of Public Health, Cambridge, MA, [4]National Institute of Mental Health, Bethesda, MD; [5]Harvard University FAS Center for Systems Biology, Cambridge, MA; [6]Research Institute of Molecular Pathology, Vienna, AT; [7]Proteomics Core UC Davis Genome Center, Davis, CA; [8]UCLA Molecular Biology Institute, Los Angeles, CA; [9]Experimental Therapeutics Centre, SG; and [10]Univ. of Texas Health Science Center at San Antonio, San Antonio, TX

Background

In 2007, the Proteomics Standards Research Group (sPRG) initiated a study design toward the goal of producing a quantitative standard for plasma proteomics utilizing stable isotope labeled (SIL) peptides. The sPRG assembled a reference table of approximately 350 human plasma proteins for which quantitative information was available (1-3). The sPRG's objective was to identify approximately 50 proteins distributed over five orders of magnitude in concentration from this reference table. A further objective was to select a minimum of three peptides per protein from which SIL analogs would be synthesized.

Theoretical tryptic digests from all 350 plasma proteins were compared against a database of theoretical proteotypic peptide constituents of proteins in the human IPI database (4). In addition, the theoretical tryptic peptides were screened using a list of 10 criteria to eliminate sequences that could lead to variability in quantitation.

The sPRG presented an interim status report at ABRF2008 in Salt Lake City, UT. The consensus of the ABRF community was that, while the goal of producing a quantitative standard for plasma proteomics was laudable, the sPRG's study proposal would be technically quite challenging, and that sample analysis would require extensive fractionation leading to an excessive investment of analysis time. Based on comments from the ABRF community, the sPRG concluded that the study proposal was beyond the scope of traditional ABRF studies. In this presentation, we describe the sPRG's efforts to develop a simplified standard for quantitative proteomics.

Current sPRG Activity

sPRG has selected a set of 10 recombinant human proteins derived from the constituents found in Sigma's Universal Proteomics Standard, which is modeled after the sPRG2006 Study sample (5). These proteins will be examined by the sPRG for suitability for the sPRG2009 Study sample. The decision to select recombinant proteins was based on our desire to remove any ambiguities resulting from biological heterogeneity. These ten proteins are currently being purified to a minimum of 90% purity (as assessed by SDS-PAGE and reversed phase HPLC). Presumably, any remaining contaminant proteins will be derived from the recombinant host. After purification, the candidate proteins will be quantified by amino acid analysis (in triplicate).

The sPRG has examined 19 of 30 publically contributed data sets corresponding to the analysis of the sPRG2006 Study sample by a variety of proteomic workflows and mass spectrometric platforms (5-7) . We have tabulated a list of empirical proteotypic peptides, which we define as peptides that are consistently observed in a majority (i.e., > 50%) of the examined data sets. The sPRG has selected a set of 50 candidate peptides which correspond to constituent sequences found in the 10 recombinant human proteins. The peptides are modeled as complete tryptic peptides with no missed cleavages. All 50 candidate peptides are empirical proteotypic peptides; 98% of the sequences are found in the GPM database (8) while 62% are found in PeptideAtlas (9). All candidate peptides conform to all criteria listed in the next panel "Criteria for Selecting Peptides for Standards". All 50 peptides are currently being synthesized as their SIL analogs using the *C*-terminal amino acid (Lys or Arg) as the site of labeling. After RP-HPLC purification, the candidate SIL peptides will also be quantified by amino acid analysis (in triplicate).

Criteria for Selecting Peptides for Standards

Synthesis Difficulty:
- Peptide Length: 8 < N < 20 amino acid residues
- Avoid Imide Formation: [DN].[GASN] (10)
- Avoid DKP Formation: *C*-terminal peptide with G or P at C2 position.
- Avoid Multiple Amino Acid Repeats: e.g., XXX
- Avoid Overly Hydrophobic Sequences: multiple aliphatic aa's e.g., A, L, I, V
- Minimize Incomplete Side Chain Removal: tBu and Trt

Peptide Stability:
- No Potential Imide Forming Sequences [DN].[GASN] (10)
- No Potential Deamidating Residues: [NG]
- No Acid Labile Residues: [DP]
- No Oxidation Sensitive: Residues - Met, multiple Trp (camCys allowed)
- No N-terminal Gln Residues (can convert to pyroglutamic acid)

Heterogeneity:
- Problems with synthesis and stability may lead to two or more species.
- No missed cleavages allowed (will require effective digestion conditions!)
- No *N*- or *C*-terminal K↓K, K↓R, R↓K, or R↓R cleavages; no K↓P or R↓P cleavages (11)
- No Pro-Pro; cis-trans isomers may separate during chromatography
- No known PTMs or consensus recognition sequences (e.g., *N*-glycosylation)

Detectability:
- Minimum of 3 proteotypic peptides per protein
- Chromatography: Peptides retained by C18 column and eluted in a reasonable percentage of CH_3CN buffer
- Peptide Length: 8 < N < 20 amino acid residues

Potential Approaches to Peptide Identification and Quantitation

Non-exclusive survey of methods using stable isotope labeled (SIL) internal standards

	Non-targeted detection methods	Targeted detection methods
Based primarily on MS/MS	• Perform undirected MS/MS on as many peptides as possible and identify them by database search. • Inspect protein list for those of interest. • Extract quantitative information from MS scans of peptide and SIL standard corresponding to the identified MS/MS spectra. • Example: data-dependent LC-MS/MS or MudPIT with an ion trap instrument. Advantages: simplicity, generality. Disadvantages: Biased towards high-abundance proteins; element of chance in peptides selected for MS/MS; potential interference of isobaric peptides in quantitation.	Perform MS/MS on all ions with precursor masses of selected peptides and their SIL standards. *Either* detect all product ions (e.g. ion trap or Q-TOF) *or* detect only selected product ions (triple quadrupole SRM – discussed in detail below) Extract quantitative information from ion chromatograms of selected product ions. Advantages: sensitivity, specificity, wide dynamic range for SRM method. Disadvantages: More work to create experiment definition; limits on number of ions monitored in one LC-MS experiment; dynamic range limitations in trapping instruments.
Based primarily on high resolution MS	Perform LC-MS or LC-MS/MS, collecting high resolution MS scans Find desired peptides and SIL standards by "exact" mass and retention times. Extract quantitative information from MS scans. Example: LC-MS or MS/MS on FT-ICR or LTQ-Orbitrap. Advantages: simplicity, specificity from ppm mass accuracy. Disadvantages: potential isobaric interferences remain complex samples; dynamic range limitations in trapping instruments; might still need MS/MS to confirm identifications.	SIM scans of masses corresponding to peptides of interest. Find desired peptides and SIL standards by "exact" mass and retention times. Extract quantitative information from MS scans. Example: LC-MS or MS/MS on FT-ICR or orbitrap. Advantages: higher sensitivity/dynamic range than full scan MS experiment. Disadvantages: potential isobaric interferences remain in highly complex sample; dynamic range limitations in trapping instruments; might still need MS/MS to confirm identifications.

The sPRG Study2009 Sample

The Study2009 sample will contain three to six proteins at three concentration levels spanning approximately two orders of magnitude. Protein concentrations will be determined by amino acid analysis.

Proteins will be reduced, alkylated and sequentially digested with Lys-C and trypsin.

A minimum of three SIL peptides will be spiked into the digest at a level that is within an order of magnitude of the protein concentration. Peptide concentrations will be determined by amino acid analysis.

Sufficient material will be provided so that participants can perform a number of replicate analyses .

Participants will be invited to use the LC-MS proteomics workflow of choice to determine the absolute concentration of constituent proteins.

Participants will be able to report their results via a web-based survey tool.

Interested in Participating?

The sPRG will make a formal study announcement in August 2008 but will accept sample requests immediately.
Send an e-mail with your name, affiliation, and complete shipping address to sPRG2009.Study@gmail.com and type the words "sample request" in the subject line. The sPRG2009 Study is open to the public however priority will be given to ABRF members.

References

1. Haab, B.B. et al., (2005) *Proteomics*, **5:** 3278–3291.
2. Anderson, L., (2005) *J. Physiol.*, **263:** 23-69.
3. Polanski, M., & Anderson, L., (2006) *Biomarker Insights*, **2:** 1-48.
4. Mallick et al., (2007) *Nature Biotechnol.* **25:** 125-131.
5. See www.abrf.org/sPRG for information about the sPRG2006 Study.
6. See www.proteomecommons.org/data for access to the sPRG2006 datasets.
7. See www.abrf.org/iPRG for information about the sPRG-BIC2007 Study.
8. See gpmdb.thegpm.org for access to the gpmDB.
9. See www.peptideatlas.org for access to PeptideAtlas.
10. Yang, Y., et al., (1994) *Techniques in Protein Chemistry VI*, pp.555-564.
11. Rodriguez, J., et al., (2007) *J. Proteome Res.*, **7:** 300-305.

Acknowledgments

The sPRG gratefully acknowledges Sigma-Aldrich Company for their generous support with this study.