



*Proteomics Research
Group*

Qualitative Proteomics Study: Identifying Differences in Primary Structure

The 2008 Proteomics Research Group Study



*Proteomics Research
Group*

PRG Mission

The PRG is a volunteer scientific organization dedicated to sharing knowledge about the analysis of proteins. The PRG aims to assist protein scientists and resource facilities to achieve their highest potential by sponsoring annual research studies that examine current techniques and capabilities. Through the promotion of broad participation and scientific excellence, the PRG aims to raise awareness, knowledge and education about modern methods of protein analysis.



*Proteomics Research
Group*

PRG Members

- Nathan A. Yates (Chair) - Merck & Co. Inc.
- Chris W. Turck (EB Liaison) - Max Planck Institute
- Allis S. Chien - Stanford University
- David B. Friedman - Vanderbilt University
- Jeroen Krijgsveld - Utrecht University
- William S. Lane - Harvard University
- Kathryn S. Lilley - University of Cambridge
- Michael MacCoss - University of Washington
- Nicholas E. Sherman - University of Virginia
- Susan T. Weintraub – Univ. of Texas Health Science Center
- H. Ewa Witkowska – Univ. of California, San Francisco



Past Research Studies

- **PRG2002:** Identification of Proteins in a Simple Mixture
 - Task: Identify components of a 5 protein mixture
- **PRG2003:** Phosphorylation Site Determination
 - Task: Identify 2 phosphopeptides and sites of phosphorylation
- **PRG2004:** Differentiation of Protein Isoforms
 - Task: Discrimination of 3 closely related proteins
- **PRG2005:** Sequencing Unknown Peptides
 - Task: *De novo* sequence analysis of 5 peptide mixture
- **PRG2006:** Quantification of Proteins from a Simple Mixture
 - Task: Relative Abundance of 8 Proteins Between 2 Different Samples
- **PRG2007:** Quantification of Proteins in a Complex Mixture
 - Task: Relative Abundance of 12 in an E. coli background



*Proteomics Research
Group*

PRG2008 Study Objectives

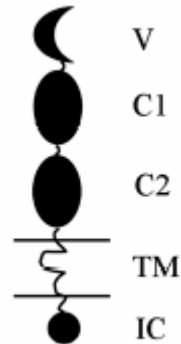
- survey the approaches used for qualitative analysis of proteins
- highlight the type of information obtained by individual approaches
- share best practices by posting the study results on-line and publishing the study
- work to enhance participation from all sectors of the scientific community



Proteomics Research Group

PRG 2008 Study Design

Soluble domains of RAGE_HUMAN (sRAGE) were expressed in vitro as described in Dattilo, Fritz, Leclerc et al, 2007 Biochemistry 46:6957. The PRG2008 study used three expression clones directly from this study.



MGSSHHHHHSSGLVPRGSHMAQNITARIGEPLVLKCKGAPKKPPQRLEWKLNTGRTEAWKV
 LSPQGGGPWDSVARVLPNGSLFLPAVGIQDEGIFRCQAMNRNGKETKSNYRVRVYQIPGKPEI
 VDSASELTAGVPNKVGTVCVSEGSYPAGTLSWHLDGKPLVPNEKGVSVKEQTRRHQPETGLFTL
 QSELMVTPARGGDPRTFSCSFSPGLPRHRALRTAPIQPRVWEPVPLEEVQLVVEPEGGAVA
 PGGTVTLTCEVPAQPSPQIHWMKDGVPLPLPPSPVLILPEIGPQDQGTYSVATHSSHGPQES
 RAVSISIIEPGEEGPTAGSVGGSGGLTALALGILGGLGTAALLIGVILWQRRQRRGEERKAPEN
 QEEEEERAELNQSEEPAGESSTGGP



5 µg 5 µg

sRAGE full length
33,169 Da, 7.8 pl

GSHMAQNITARIGEPLVLKCKGAPKKPPQRLEWKLNTGRTEAWKVLSPQGGGP
 WDSVARVLPNGSLFLPAVGIQDEGIFRCQAMNRNGKETKSNYRVRVYQIPGKPE
 IVDSASELTAGVPNKVGTVCVSEGSYPAGTLSWHLDGKPLVPNEKGVSVKEQTR
 RHPETGLFTLQSELMVTPARGGDPRTFSCSFSPGLPRHRALRTAPIQPRVWEP
 VPLEEVQLVVEPEGGAVAPGGTVTLTCEVPAQPSPQIHWMKDGVPLPLPPSPVL
 ILPEIGPQDQGTYSVATHSSHGPQESRA**AVSISIIEPGEEG**

sRAGE domain VC1
24,585 Da, 9.5 pl

GSHMAQNITARIGEPLVLKCKGAPKKPPQRLEWKLNTGRTEAWKVLSPQGGGP
 WDSVARVLPNGSLFLPAVGIQDEGIFRCQAMNRNGKETKSNYRVRVYQIPGKPE
 IVDSASELTAGVPNKVGTVCVSEGSYPAGTLSWHLDGKPLVPNEKGVSVKEQTR
 RHPETGLFTLQSELMVTPARGGDPRTFSCSFSPGLPRHRALRTAPIQPR**VWEP**
VPLEEVQLVVE

3 µg

sRAGE domain V
12,524 Da, 9.8 pl

GSHMAQNITARIGEPLVLKCKGAPKKPPQRLEWKLNTGRTEAWKVLSPQGGGP
 WDSVARVLPNGSLFLPAVGIQDEGIFRCQAMNRNGKETKSNYRVRVYQIPGK**PE**
IVDSASE

2 µg

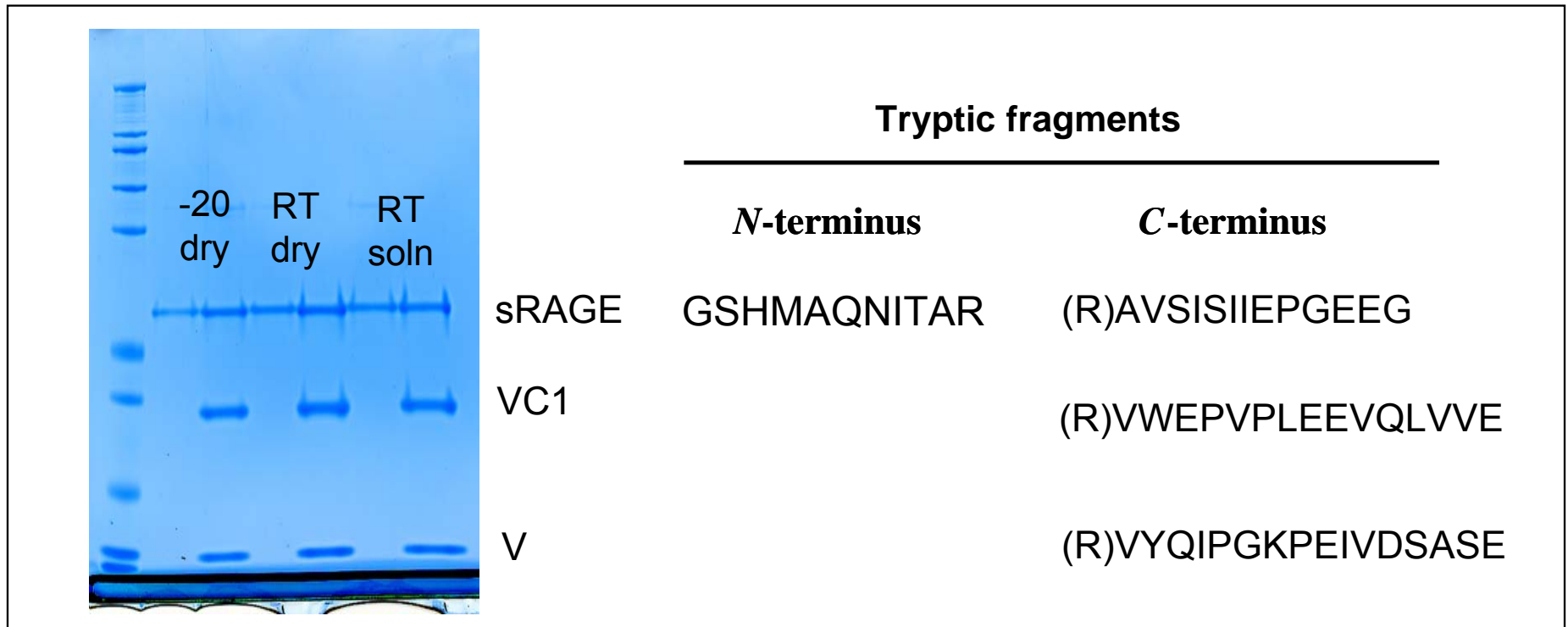


PRG Analysis of Study Samples

- Preliminary analysis of sRAGE sample included:
 - 1D gel/MS (tof/tof)
 - 2D gel/MS (tof/tof)
 - LC/MS/MS (LTQ from 1D gel peptides)
 - Intact ESI/MS
 - P-mod/Monster-Mod search
- 3-day sample stability tests
 - Room temperature dry and in-solution
 - Frozen



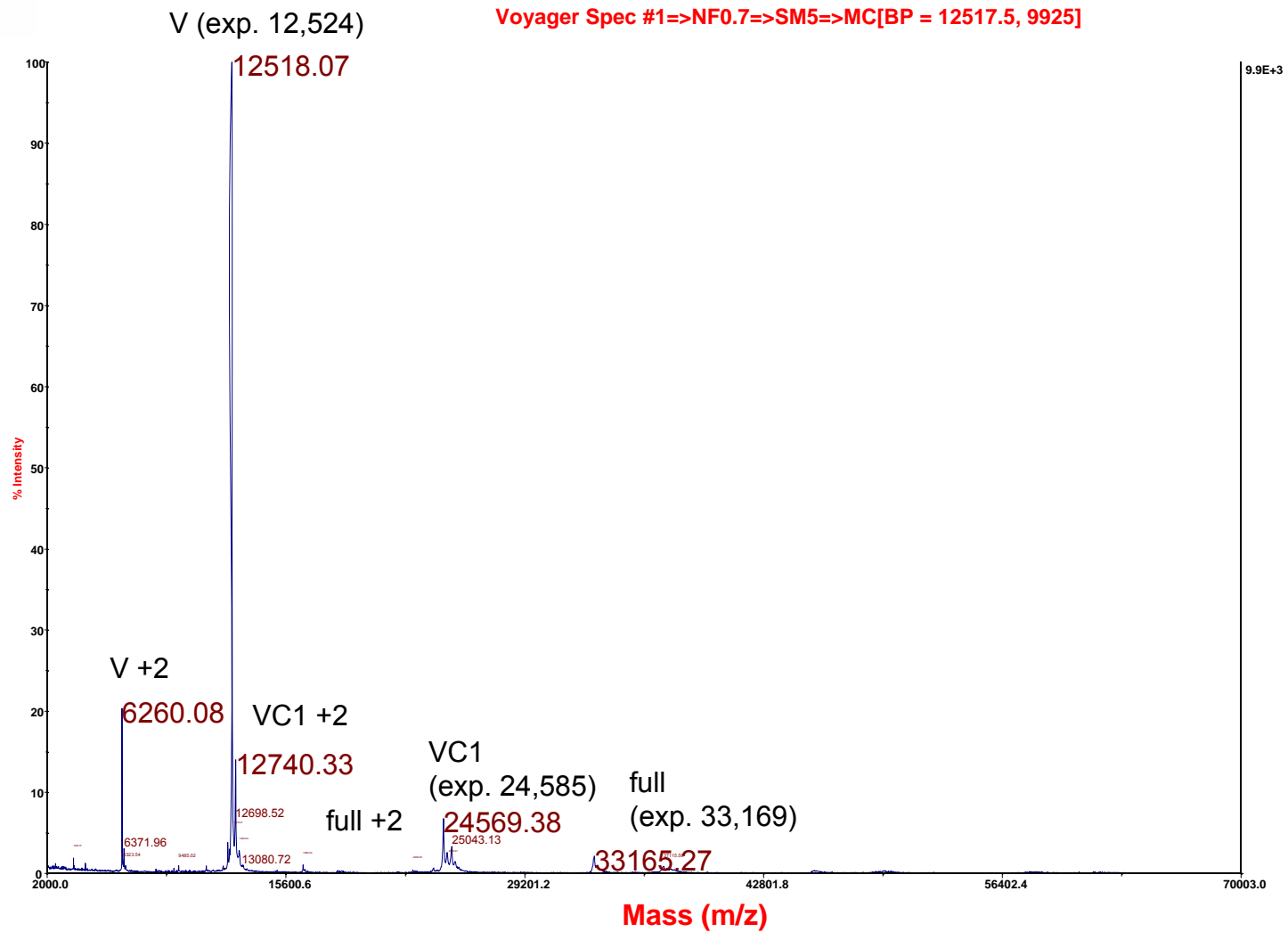
Samples Analyzed by 1D SDS-PAGE





Proteomics Research
Group

Samples Analyzed by MALDI-TOF





Proteomics Research
Group

Participants were asked to:

- analyze two study samples (labeled "A" and "B") that consist of separate preparations of related human proteins that were expressed in *Escherichia coli* and affinity purified
- identify the major protein(s) in each sample
- report any differences in protein composition found between samples A and B
- tabulate the sequence coverage obtained for the major protein(s) in each sample



Proteomics Research
Group

Example Survey Results

Anonymous 5-DIGIT numeric identifier: 57021

Is this protein different or unique? In both A and B

SwissProt accession: Q15109

Entry name: RAGE_HUMAN

Protein name: Advanced glycosylation end product-specific receptor

Degree of confidence? High

N- terminus: GSHMAQNITAR

C-terminus: AVSISIIEPGEEG

Protein Sequence Detail:

GSHMAQNITARIGEPLVLKckgapkkppqrlewklntgrTEAWKVLSPQGGGPWD
SVARVLPNGSLFLPAVGIQDEGIFRcqamnrngketksnyrvrVYQIPGKPEIVDS
ASELTAGVPNKVGTCVSEGSYPAGTLSWHLDGKPLVPNEKgvsvkeqtrRHP
ETGLFTLQSELMVTPARGGDP RPTFSCSFSPGLPRhraIRTAPIQPRVWEPV
PLEEVQLVVEPEGGAVAPGGTVTLTCEVPAQPSPQIHWMKDG VPLPLPPS
PVLILPEIGPQDQGTYS CVATHSSHGPQESRAVSISIIEPGEEG



Proteomics Research
Group

Example Survey Results (Cont.)

Other or clarification of above (please describe): Both have at least one disulfide bond Cys126-Cys190

Intact mass determination (enter M+H): 40,000

Mass accuracy (ppm): 2

Resolution (FWHM): 25,000

Percent AA sequence coverage (%): 65

Mass Spectrometry (select as many as apply): TOF – MALDI, Linear Ion Trap-FT – ESI

Were there other technologies that contributed to your results: 1D SDS-PAGE



Proteomics Research
Group

Example Survey Results (Cont.)

What software tools did you use to identify the proteins? **Bioworks, Mascot, Scaffold**

What software tools did you use to determine the difference between Sample A & Sample B? **Scaffold**

How much time in days did it take you to complete this study? For example, enter 2.5 for two and one half days: **2**

How much experience do you have with the technique(s) used to determine differentially expressed proteins you used in this study?
1-3 years

How difficult do you think this study was? **Moderate**

Would you try and do this type of experiment again? **Yes, it was fun**



PRG Results Table

| | | | | | | Sample A | | Sample B | | | | | | |
|--------|-----|---|---|---|------|-----------|--------|-----------|---------|-----------|---------|-----------|---------|-------------|
| | | | | | | Protein 1 | | Protein 1 | | Protein 2 | | Protein 3 | | |
| ID | Gel | E | T | B | %Cov | N-term | C-term | N-term1 | C-term1 | N-term2 | C-term2 | N-term3 | C-term3 | Addl Incrct |
| 27960v | | | T | B | 100 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 0 |
| 67775 | 1 | E | T | B | 93 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 0 |
| 21094 | 1 | | | B | 91 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 0 |
| 23300 | | | | B | 76 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 0 |
| 14850v | 1 | | | B | 75 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 0 |
| 57021 | 1 | | | B | 65 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 0 |
| 91239 | 1 | | | B | 40 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 0 |
| 12358 | 1 | | | B | 62 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 2 |
| 15987 | 1 | | | B | 76 | 10 | 10 | 10 | 10 | 10 | 10 | | | 0 |
| 47886 | 1 | | | B | 50 | 10 | 10 | 10 | 10 | 10 | 4 | | | 0 |
| 12303 | | | | B | 61 | 10 | 10 | 10 | 10 | 10 | 0 | | | 0 |
| 53104 | | | | B | 98 | 10 | 10 | 10 | 10 | | | | | 0 |
| 12707v | | | | B | 71 | 10 | 10 | 10 | 10 | | | | | 0 |
| 88007v | | | T | B | 92 | 10 | 10 | 10 | 10 | | | | | 3 |
| 14146v | 1 | | T | B | 91 | 10 | 9 | 10 | 9 | | | | | 0 |
| 19327 | 1 | | | B | 61 | 10 | 9 | 10 | 9 | | | | | 0 |
| 46013 | 1,2 | | | B | 79 | 10 | 0 | 10 | 0 | 10 | 0 | 10 | 0 | 0 |
| 93041 | | | | B | 92 | 10 | 0 | 10 | 0 | | | | | 0 |
| 14005 | | | | B | 44 | 10 | 0 | 10 | 0 | | | | | 0 |
| 46011 | | | T | B | 42 | 10 | 0 | 10 | 0 | | | | | 0 |
| 30366 | 1 | | | B | 0 | 10 | 0 | 10 | 0 | | | | | 0 |
| 15119 | 1 | E | | B | 21 | 10 | 0 | | | 10 | 0 | 10 | 0 | 1 |
| 14125 | 1 | | | B | 53 | 0 | 10 | 0 | 10 | 0 | 10 | 0 | 10 | 1 |
| 25058 | | | | B | 29 | 0 | 10 | 0 | 10 | 0 | 0 | | | 0 |
| 12107 | | | | B | 73 | 0 | 10 | 0 | 10 | | | | | 1 |
| 82930 | 1 | | | B | 35 | 0 | 10 | 0 | 10 | | | | | 1 |



Example Write-up

Sample Information. The study samples were supplied in two vials (labeled "A" and "B") and consist of separate preparations of related human proteins that were expressed in *Escherichia coli* and affinity purified. The cloning vectors that were used produced an N-terminal His₆-tagged fusion protein that included a thrombin cleavage site in the His-tag. In both preparations, the sequence GSHM was retained with the protein after thrombin cleavage to remove the Histag. Vials A and B contain 5 µg and 10 µg of protein, respectively. The samples were prepared from aqueous solutions that also contained small amounts of salts. To the best of our knowledge, there are no appreciable quantities of interfering substances that contain primary amines and/or free thiols. The samples have been successfully dissolved in 25 – 50 mM ammonium bicarbonate with or without 20% acetonitrile; as well as 0.1% formic acid; we anticipate that other solvents can be used.

Sample Preparation. Both samples were dissolved in water containing 5% acetonitrile to a concentration of 1ug/µL.

In-solution Trypsin Digestion. 1ug of each sample was diluted to 10µL with 50mM ammonium bicarbonate pH=8.0. The samples were reduced with a final concentration of 10mM DTT for 30 minutes (RT) followed by alkylation with a final concentration of 50mM iodoacetamide for 30 minutes (RT). The samples were then digested with 0.1ug of Promega trypsin overnight at 37C in the dark. The resulting sample was acidified with 0.5µL of acetic acid.

LC-MS/MS Analysis of Tryptic Digest. The LC-MS/MS system consisted of a ThermoElectron LTQ-FT mass spectrometer system with a Protana nanospray ion source interfaced to a self-packed 8 cm x 75 µm id Phenomenex Jupiter 10 µm C18 reversed-phase capillary column. 10% of each sample was manually injected and the peptides eluted from the column by an acetonitrile/0.1 M acetic acid gradient at a flow rate of 0.4 µL/min (0-80% acetonitrile over 20 minutes). The nanospray ion source was operated at 2.8 kV. The digest was analyzed using the full data dependency (no repeat, 120s exclusion, 1Da window) of the instrument acquiring full scan mass spectra to determine peptide molecular weights (FT – 100,000 resolution) and product ion spectra to determine amino acid sequence in ten sequential scans (IT). This mode of analysis produced approximately 1000 CAD spectra of ions ranging in abundance over several orders of magnitude.

MALDI-TOF Analysis of Proteins. 1ug of each sample was diluted in sinapinic acid matrix (SA) 1:1. 0.5µL of this solution was spotted onto the plate and allowed to dry. The sample was analyzed on a Bruker Microflex by averaging 300 scans using external calibration and instrument defaults for the 5-50KD range. The matrix was saturated SA in a solution of 70% acetonitrile in 0.1% TFA.

Data Analysis. The LC-MS/MS data was analyzed using Bioworks 3.3.1 (Sequest) for the database search. The search was performed against the current version of IPI Human and NR. The parameters used were 10ppm parent mass, 1Da fragment mass, full tryptic, carboxyamidomethyl Cys, and variable oxidized Met. The search results were loaded

Sample A

IPI: IPI00014810.3 | SWISS-PROT: Q15109-1 | AGER Isoform 1 of Advanced glycosylation end product-specific receptor precursor

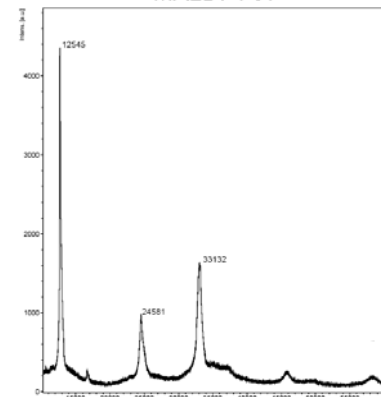
GSHMAQNITARIGEFLVLKCKGAPKPKFPQRLEWKLNTGRTEAWKVLSPQGGGPPWDSVAR
 VLPNGSLFLPAVGIQDEGIFRCQAMNRRNGKETKSNYRVRVYQIPGKPEIVDSASELTAG
 VPNKVGTCSVEGSPYAGTLSWHLDGKPLVPNEKGVSVKEQTRRHPEGLFTLQSELMVT
 FARGGDPREFTFCSPSPGLFRHRALRTAPIQPRVWEPVPLEEVQLVVEPEGGAVAPGGT
 VTLTCEVPAQPSFQIHWMKDGVPFLPFPSPVLILPEIGPQDQGTYSVATHSSHGPKES
RAVSIISIEPGEEG

Sample B

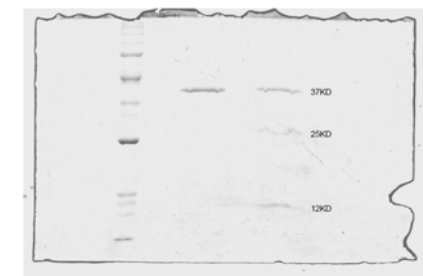
IPI: IPI00014810.3 | SWISS-PROT: Q15109-1 | AGER Isoform 1 of Advanced glycosylation end product-specific receptor precursor

GSHMAQNITARIGEFLVLKCKGAPKPKFPQRLEWKLNTGRTEAWKVLSPQGGGPPWDSVAR
 VLPNGSLFLPAVGIQDEGIFRCQAMNRRNGKETKSNYRVRVYQIPGKPEIVDSASELTAG
 VPNKVGTCSVEGSPYAGTLSWHLDGKPLVPNEKGVSVKEQTRRHPEGLFTLQSELMVT
 FARGGDPREFTFCSPSPGLFRHRALRTAPIQPRVWEPVPLEEVQLVVEPEGGAVAPGGT
 VTLTCEVPAQPSFQIHWMKDGVPFLPFPSPVLILPEIGPQDQGTYSVATHSSHGPKES
RAVSIISIEPGEEG

MALDI-TOF



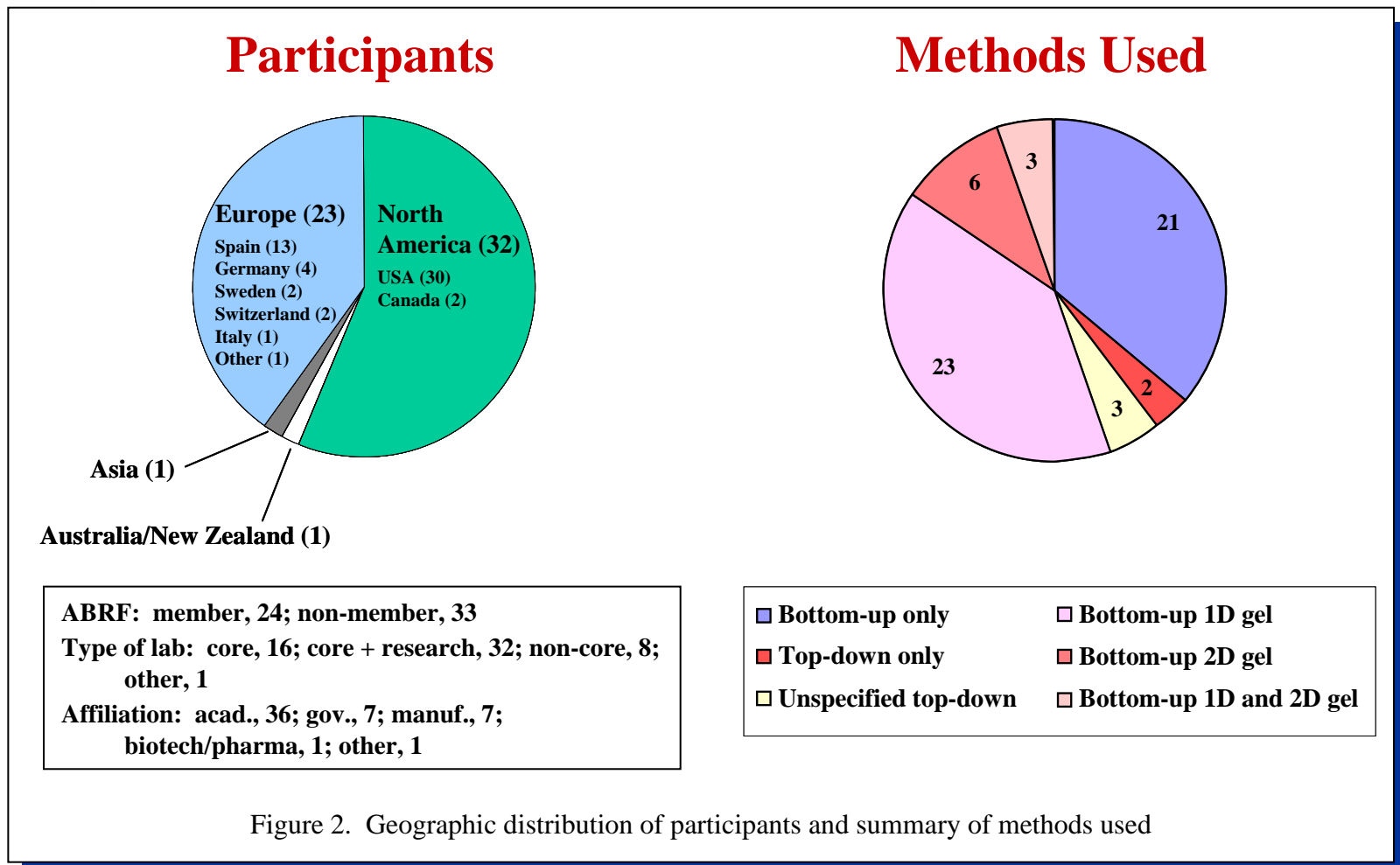
SDS-PAGE



Participation

- Specialized labs encouraged to participate
- Number of sample sets requested: 114
- Number of survey responses: 57 (↑33%)
- Number of participant write-ups: 33

Participation





Biggest Challenges

- “Software analysis tools problems, limited time”
- “Determination of N-terminus had to be done with de novo sequencing”
- “...to find out whether the several masses we got in the intact mass analysis of sample A corresponded to protein fragments actually added in the sample, or were minor fragments originated in the analysis”
- “Data analysis, and the inability to detect and fragment intact proteins. ... I would have liked to have the possibility of performing top-down”
- “Sample amount and coverage”
- “Too little time to spend on it.”
- “...the biggest issue was that nobody from the lab people involved in the study read the sample letter carefully.”



Selected Comments

- “Very interesting and challenging, frustrating also...”
- “This study is a great example for how dreadful little you learn about protein structure if you run mere shotgun approaches. Its even a lot worse in proteomics cases.”
- “I would have been nice if there was a explicit description of the amount and type of data we would be asked for.”
- “Make the instructions clearer so the study is about methods and techniques, not reading comprehension. This was a super sample, it just needed better explanation so we did not waste so much time on it.”
- “The topic and motives for this project are great, but I feel the samples were poor. The samples received were contaminated and degraded
- “More difficult than I thought.”



Conclusions

- Proteomics offers a wide arrange of approaches for the qualitative analysis of proteins
- Many methods and approaches were used successfully identify and sequence the truncated sites
- In many cases, the combination of two complementary approaches (e.g., 1D SDS PAGE to resolve protein components followed by LC-MS/MS for sequence information) gave a higher success rate than use of a single experimental approach.
- As expected, experience remains a key factor in this study
- Best practices were shared by positn results on-line
- Participation increased by a third compared to 2007



*Proteomics Research
Group*

Acknowledgments

Brain Dattilo, Susan Meyn and Walter Chazin (Vanderbilt Center for Structural Biology) for supplying the RAGE samples;

Evelin Szakal (UCSF) who served as the anonymizer.

All of this years study participants