iPRG2011: Identification of Electron Transfer Dissociation (ETD) Mass Spectra

Lennart Martens[1] , Karl R. Clauser[2], Robert J Chalkley[3], Paul A. Rudnick[4], Manor Askenazi[5], W. Hayes McDonald[6], Henry H. Lam[7], Nuno Bandeira[8], Eric Deutsch[9] and Thomas Neubert[10]

[1]Ghent University, Ghent, Belgium; [2]The Broad Institute of MIT and Harvard, Cambridge, MA; [3]University of California, San Francisco, CA; [4]National Institute of Standards and Technology, Gaithersburg, MD, [5]Dana-Farber Cancer Institute, Boston, MA; [6]Vanderbilt University School of Medicine, Nashville, TN; [7]Hong Kong University of Science and Technology; [8]University of California, San Diego, CA; [9]Institute for Systems Biology, Seattle, WA; [10]New York University School of Medicine, New York, NY

Proteome Informatics Research Group

A Proteome Informatics Challenge

The field of mass spectrometry based proteomics has seen several key innovations over the last several years, including novel experimental methods, new instruments, and unique fragmentation strategies. The latter, in the form of electron capture dissociation (ECD) and the more widely applicable electron transfer dissociation (ETD) have captured the imaginations of many researchers, expanding their ability to identify and analyze peptides and proteins. However, since ECD/ETD spectra differ substantial from more traditional collision induced dissociation (CID) spectra in both their prominent ion series as well as their preferred bond-breaking characteristics, the (automatic) interpretation of ECD/ETD spectra requires novel algorithm optimizations. Efficient identification of ECD/ETD spectra thus remains an active and exciting field of proteomics informatics research.

Study Goals

1. Evaluate the consistency of reporting peptide identifications from ETD spectra across laboratories
2. Characterize the underlying reasons why result sets differ
3. Produce a benchmark ETD dataset, spectral library and analysis resource

Study Design

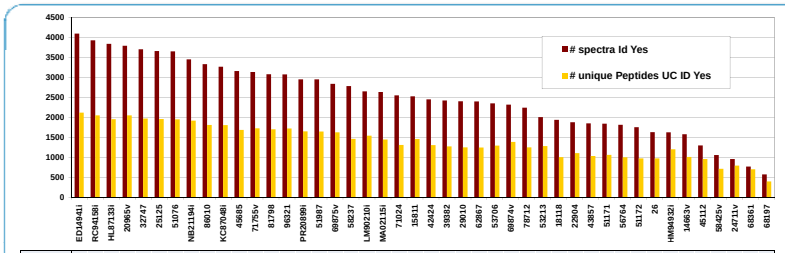
- Use a common dataset
- Use a common sequence database
- Allow participants to use the bioinformatic tools and methods of their choosing
- Use a common reporting template
- Fix the identification confidence (1% FDR)
- Ignore modification localization
- Ignore protein inference

Study Materials

- 1 Orbitrap XL dataset (1 files)
  - RAW, mzML, mzXML, MGF, or dta – conversions by ProteoWizard
- 1 FASTA file (UniProt yeast sequences)
- 1 template (Excel)
- 1 on-line survey (Survey Monkey)

Study Instructions

1. Analyze the dataset
2. Report the peptide spectrum matches in the provided template
3. Complete an on-line survey
4. Attach a 1-2 page description of your methodology



Total Identifications

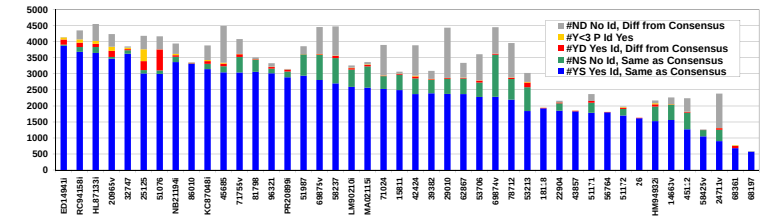
Total identifications reported by each participant. Both **Total Spectra** and **Unique Peptides** are indicated.

Table Key:

Bioworks = Bw
DTA Generator = Dg
Excel = Ex
Extract_msn = Ems
IDPicker = idp
Inspect = Ins
Ma Distiller = MaD
Mascot = Ma
MS-GFDB = M-G
MSQuant = MQ
MyriMatch = My
ProteoWizard = PW
Protein Prospector = PP
Peptide/ProteinProphet = P/PP
TransProteomic Pipeline (TPP) = TPP
In-house software (freely available) = IH(P)
In-house software (not public) = IH(np)

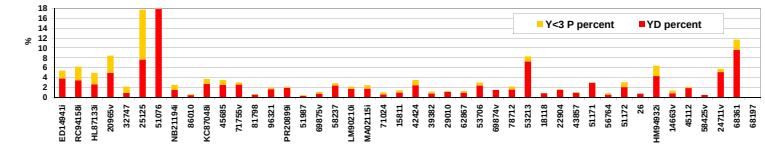
Xcalibur = Xc
X!Tandem = X!
Spectrum Mill = SM
SpectraST = SpST
OMSSA = OM
Other = Other
PEAKS = PK
Percolator = Per
pFind = pF
Phenyx = Ph
ReAdW = Re
SEQUEST = SQ

Modifications: pyroGlu, pyro-carbamidomethyl, M(+16), N(+0.98)
N-terminal: Carbamyl, Acetyl, Carbamidomethyl
All other modifications



Room for Improvement in ID Certainty Thresholds

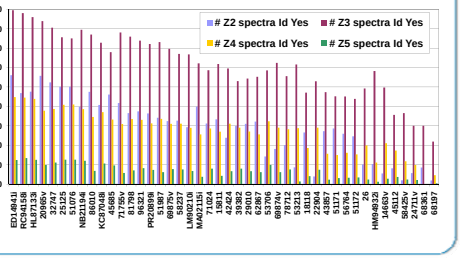
Identifications reported as: **Yes that matched the consensus**; **No, but still matching the consensus**; **Yes, but a different answer than the consensus**; **Yes, < 3 consensus**; No, that disagreed with consensus



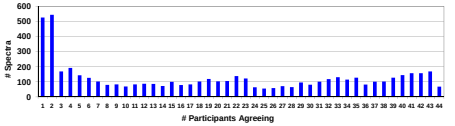
Extraordinary Skill Rate or High False Discovery Rate

ESR + FDR = 100* (Y1P+YD)/total ids. **Yes, but a different answer than the consensus**; **Yes, as unique**

Distribution of Precursor Charge

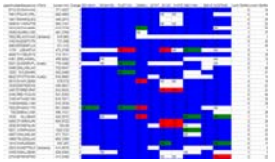


How Much Do the Identifications Overlap?



Resource for Inspecting Peptide Id Overlaps

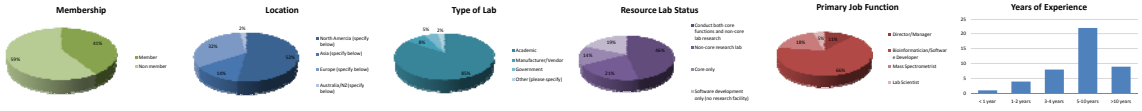
YS: Y – identification, and top sequence same as consensus
NS: N – identification, but top sequence same as consensus
YD: Y – identification, and top sequence different than consensus



A table containing the consensus match for each spectrum was generated from all of the participant data. The ID and confidence is indicated for each participant according to the key above. Color codes match the bars in the charts to the right.

Who Participated:

35 submissions were returned .Additionally, 9 iPRG members completed submissions. Participation was international and covered a wide range of experience level.



Preliminary Conclusions

- While many spectra were readily identifiable (2462 were agreed upon by half of the participants), others were much more challenging and resulted in a broad range of results
- Success depended less on the particular tools employed and more on the way they were used
- Likewise, both single and multiple algorithm strategies could be successful

• The iPRG2011 are in the process of preparing the data for publication. If you participated and would like to help out, contact the iPRG through anonymous.iPRG2011@gmail.com.

For more information on the iPRG and for copies of this poster and the talk (available March 1st) please visit: http://www.abrf.org/iPRG

Acknowledgements: The iPRG are grateful to *all of the participants*. We would also like to thank Jeremy Carver (UCSD) for serving as the "Anonymizer",.