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

Lennart Martens1, Karl R. Clauser2, Robert J Chalkley2, Paul A. Rudnicki3, Manor Askonazi3, W. Hayes McDonald4, Henry H. Lam5, Nuno Bandeira6, Eric Deutsch7 and Thomas Neubert10

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

Proteomics Informatics Research Group

A Proteomics 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 substantially 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 results sets differ.
3. Produce a benchmark ETD dataset, spectral library and analysis resource.

Study Design

• Use a common dataset
• Use a common sequence database
• All participants to use the bioinformatic tools and methods of their choosing
• Use a common reporting template
• File the identification confidence [1% FDR]
• Ignore modification localization
• Ignore protein inference

Study Materials

• 10 Driftgas dataset (1 file)
  - Mono reproducible,可在 or all
  - Individual files (optional)
• 1 FASTA file (appropriate sequences)
• 1 template (Excel)
• 1 on-line survey (SurveyMonkey)

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 2-page description of your methodology

Total identifications

Table: Total identifications as reported by each participant. Both total spectra and unique Peptides are indicated.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Total Spectra</th>
<th>Unique Peptides</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>500</td>
<td>200</td>
</tr>
<tr>
<td>B</td>
<td>400</td>
<td>150</td>
</tr>
<tr>
<td>C</td>
<td>300</td>
<td>75</td>
</tr>
</tbody>
</table>

Distribution of Precursor Charge

<table>
<thead>
<tr>
<th>Charge</th>
<th>Participant A</th>
<th>Participant B</th>
<th>Participant C</th>
</tr>
</thead>
<tbody>
<tr>
<td>2+</td>
<td>15%</td>
<td>20%</td>
<td>30%</td>
</tr>
<tr>
<td>3+</td>
<td>25%</td>
<td>15%</td>
<td>10%</td>
</tr>
<tr>
<td>4+</td>
<td>30%</td>
<td>20%</td>
<td>15%</td>
</tr>
</tbody>
</table>

Preliminary Conclusions

• While many spectra were readily identifiable (2422 were agreed upon by half of the participants), others were much more challenging and resulted in a broad range of results.
• Quality depended less on the particular tools employed and more on the way they were used
• Likewise, both single and multiple algorithms could be successful

Room for Improvement in ID Certainty Thresholds

<table>
<thead>
<tr>
<th>Confidence</th>
<th>Participant A</th>
<th>Participant B</th>
<th>Participant C</th>
</tr>
</thead>
<tbody>
<tr>
<td>80%</td>
<td>30%</td>
<td>25%</td>
<td>15%</td>
</tr>
<tr>
<td>90%</td>
<td>20%</td>
<td>15%</td>
<td>10%</td>
</tr>
<tr>
<td>95%</td>
<td>10%</td>
<td>5%</td>
<td>5%</td>
</tr>
</tbody>
</table>

Table: Extraordinary skill rate or high false discovery rate

<table>
<thead>
<tr>
<th>Participant</th>
<th>Extraordinary skill rate</th>
<th>High false discovery rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>10%</td>
<td>20%</td>
</tr>
<tr>
<td>B</td>
<td>15%</td>
<td>30%</td>
</tr>
<tr>
<td>C</td>
<td>20%</td>
<td>40%</td>
</tr>
</tbody>
</table>

Resource for Inspecting Peptide Id Overlaps

A table containing the consensus match for each spectrum was provided from all of the participant data. The FDR and confidence is calculated for each participant according to the key above. Color codes the bars in the chart to the right

Preliminary Conclusions

• While many spectra were readily identifiable (2422 were agreed upon by half of the participants), others were much more challenging and resulted in a broad range of results.
• Quality depended less on the particular tools employed and more on the way they were used
• Likewise, both single and multiple algorithms could be successful

Acknowledgments: The iPRG is grateful to all of the participants. We would also like to thank Jeremy Cancer (UCSF) for serving as the "Transporter".

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