Suppl. Figure 1. Introductory information sent to the 2012 ABRF gPRG study participating laboratories.
2012 Comparative Glycopeptide Mapping Study

Invitation
Laboratories engaged in mass spectrometry-based proteomics, glycoproteomics, or glycomics are invited to participate in a glycoprotein analysis study in 2012. This study was initiated by the Glycoprotein Research Group of the Association for Biomolecular Resource Facilities (ABRF). Those indicating interest will be sent two sources of prostate specific antigen (PSA) for comparative glycopeptide N-glycosylation analysis. Samples will be sent in early March, 2012. The deadline for submission of results will be July 31, 2012. The results will be communicated in a publication that will be written by the study organizers. All those submitting data for the study will be included as co-authors for this publication. The study results will be presented at the ABRF annual conference in 2013 and at other scientific meetings. To participate please email your interest to Joe Zaia (jzaia@bu.edu) and include your shipping address.

Study goals
The goal of this study is to determine the ability of the glycoproteomics community to compare N-glycosylation between two different sources of prostate specific antigen (PSA). The PSA sources have been selected by the study organizers and the differences in glycosylation are known.

Background
Accurate mapping of glycoprotein glycosylation is essential for basic glycosciences, biomarker discovery, and recombinant glycoprotein therapeutic characterization. For these purposes, it is necessary to identify and determine the abundances of glycopeptides derived from the target protein. This glycopeptide mapping insures that data are produced from the target proteins rather than from contaminants. N-Linked glycopeptides are present typically as a series of glycoforms, all containing the chitobiose core. Because proteolytic digestion often results in glycopeptides containing incomplete cleavage sites, the abundances of glycopeptide glycoforms must be reconstructed from those of several different ions. For these reasons, accurate glycopeptide mapping represents an analytical challenge.

Study rationale
This is a glycopeptide mapping study. For the purpose of accurate determination of glycoprotein glycosylation, it is necessary to determine glycosylation on specific peptides. Most glycoprotein samples are not pure. Contamination with other proteins, even at a level of a few percent, may give rise to false positive identifications. Thus, for the purpose of identifying the peptide and glycan parts, it is important to determine the masses of intact glycopeptides. Investigators may also wish to analyze released glycans. The study will compare results obtained using glycopeptides with those from released glycans.

Data reporting
Participants should report their results using an Excel template that will be provided by the organizing committee. To use this template, the data should be converted into glycan compositions (Hex, HexNAc, dHex, NeuAc, NeuGc, Sulfate, Phosphate) that modify the PSA N-glycosylation site.

Study sponsors
PSA was donated generously by Lee Biosolutions. Support for the costs of the study was provided by the Association for Biomolecular Resource Facilities, Thermo-Fisher Scientific, and Bruker Daltonics, Inc.

We look forward to your participation. Sincerely,

Study organizers
Nancy Leymarie (Center for Biomedical Mass Spectrometry, Boston University)
Joseph Zaia (Center for Biomedical Mass Spectrometry, Boston University)
Ron Orlando (Complex Carbohydrate Research Center, University of Georgia)
Daniel Kolarich (Max Plank Institute of Colloids and Interfaces)
Karen Jonscher (University of Colora)
ABRF GlycoProtein Research Group (gPRG)

Joseph Zaia
Boston University School of Medicine
Professor of Biochemistry

Nancy Leymarie
Boston University School of Medicine
Senior Scientist: Glycomics

Karen Joncher
University of Colorado Denver
Proteomics Director Systems Biology Core Facility

Daniel Kolarich
Max Planck Institute of Colloids and Interfaces, Berlin
Group Leader: Glycoproteomics

Ron Orlando
Complexe Carbohydrate Research Center
Professor of Biochemistry and Molecular Biology, and Chemistry

Expansion of the group is needed!
ABRF GlycoProtein Research Group (gPRG)

• Interest of the group
  – site determination (N-, O-linked)
  – structure determination (complex, hybrid,...)
  – quantification (differential, % occupancy)

• Explore methods and technology
  – strategies available for characterization

• Develop resources for education
  – Workshops: ABRF, US-HUPO, HUPO
  – gPRG Study: ABRF 2010, 2011/2012
2012 gPRG study: Criteria for the Ideal Glycoprotein

• Goal is to characterize glycosylation & target glycopeptides
  – Map sites
  – Characterize glycosylation
  – Differential quantification

• Suitable protein target
  – Biological impact
  – Protein commercially available
  – Inexpensive
  – large amount
  – Mammalian/ human
  – Ideally one site of glycosylation
2012 gPRG study : Human Prostate Specific Antigen

*Biological impact*

- Known diagnostic marker for prostate cancer

- **Statistics of prostate cancer**
  - One of the most common cancers
  - 240,000 new cases every year
  - 28,000 deaths per year
  - second leading cause of cancer death in American men
  - 1/6 men will have cancer in their life
2012 gPRG study: Human Prostate Specific Antigen

• **Mechanism**
  – PSA secreted as a proenzyme into the lumen of prostate gland
  – In disease (cancer, benign prostatic hyperplasia), the membrane can be disrupted
  – PSA will leak into the peripheral circulation
  – In blood PSA is free or complexes with a protease inhibitor as alpha-1-antichymotrypsin

• **Correlation between prostate cancer and [PSA]**
  – Healthy men, [PSA] < 4ng/ml, normal examination: 15% chance of prostate cancer
  – Grey zone: 2.5 to 10 ng/ml: 25% chance of prostate cancer
  – >10ng/ml: 67% chance of prostate cancer
    - PSA is not a perfect Biomarker

• **Improvement of the specificity of prostate cancer detection**
  – Measurements of % complexed, % specific and % free (proPSA or cleaved PSA)
  – Measure glycosylation pattern: know to change in cancer tissue

➤ Glycosylation pattern of PSA and Cancer ??
Human Prostate Specific Antigen & Glycosylation

**PSA Characteristics**
- Function: Hydrolyzes semenogelin-1 thus leading to the liquefaction of the seminal coagulum.
- ≈ 30 kDa, 261 Aa
- 5 disulfide bonds reported
- 1 N-linked glycosylation site: aa 69

**Glycans from seminal fluid from healthy patient**
- Sialylated complex biantennary, mostly core fucosylated
- Minor presence of GalNAc on the antennae, increases with the pI of PSA
- High Mannose and hybrid glycans reported

**Glycans from cancer patient vs healthy patients**
- Glycans: **contradictory**, purity of sample, cell lines vs patient cancer
- Glycopeptides: PSA N-Glycans seminal plasma of healthy patient less sialylated than than PSA N-Glycan for cancer patient
- Linkage: 2-3 linked sialic acid could distinguish malignant from benign
Study Rational

• Goal is to characterize glycosylation & target glycopeptides
  – Map sites
  – Characterize glycosylation
  – Differential quantification

• Ideal scenario
  – Obtain PSA from plasma draw
  – Compare glycosylation healthy controls vs. cancer
  – Identify the biomarker

• In reality
  – “Off the shelf” PSA
  – Purified from seminal fluid
  – Take part in the gPRG 2012 study!
Initial Proteomics Analysis of PSA

**PSA_high Isoform**

<table>
<thead>
<tr>
<th>Score</th>
<th>Mass</th>
<th>Matches</th>
<th>Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>29293</td>
<td>230 (101)</td>
<td>79 (48)</td>
<td></td>
</tr>
<tr>
<td>215</td>
<td>17 (9)</td>
<td>17 (9)</td>
<td></td>
</tr>
<tr>
<td>156</td>
<td>15 (6)</td>
<td>10 (4)</td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>5 (1)</td>
<td>7 (2)</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>9 (2)</td>
<td>5 (1)</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>7 (2)</td>
<td>5 (1)</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>6 (3)</td>
<td>5 (1)</td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>6 (2)</td>
<td>4 (2)</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>7 (5)</td>
<td>6 (5)</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>12 (3)</td>
<td>7 (3)</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>2 (1)</td>
<td>2 (1)</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>2 (1)</td>
<td>2 (1)</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>4 (1)</td>
<td>4 (1)</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>2 (1)</td>
<td>2 (1)</td>
<td></td>
</tr>
</tbody>
</table>

**PSA Batch M02015**

<table>
<thead>
<tr>
<th>Score</th>
<th>Mass</th>
<th>Matches</th>
<th>Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>29293</td>
<td>66 (16)</td>
<td>14 (2)</td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>15 (2)</td>
<td>14 (2)</td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>1 (1)</td>
<td>1 (1)</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>2 (1)</td>
<td>2 (1)</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>2 (1)</td>
<td>2 (1)</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>2 (1)</td>
<td>1 (1)</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>2 (1)</td>
<td>2 (1)</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>2 (1)</td>
<td>2 (1)</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>3 (1)</td>
<td>3 (1)</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>1 (1)</td>
<td>1 (1)</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>1 (1)</td>
<td>1 (1)</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>1 (1)</td>
<td>1 (1)</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>1 (1)</td>
<td>1 (1)</td>
<td></td>
</tr>
</tbody>
</table>

Sequence Coverage: ~ 80%

Sequence Coverage: ~ 70%

**Both forms are relatively pure. Initial coverage indicates N-linked at aa.69.**
Guidance: Approach of Experimental design

• **Considerations based upon the gPRG 2012**
  - Enzyme
  - Chromatography: C18 vs HILIC
  - Ionization
  - Mass spectrometry
  - Fragmentation method: CID, ETD, HCD

• **Our approach**
  - Target the glycopeptides directly
  - LC-MS/MS
  - label free differential quantification
Suggested Approach
Digestion to Peptides, Glycopeptides

MWVPVVFLTL SVTWIGAAPL ILSRIVGGWE CEKHSQPWQV LVASRGRAVC 50
GGVLVHPQWV LTAAHCIRNK SVILLGRHSL FHPEDTGQVF QVSHSFPHPL 100
YDMSLLKRNRF LRPGDDSSHD LMLLRLSEP4A ELTDAVKVMD LPTQEPALGT 150
TCYASGWGS1 EPEEFLT PKK LQCVDLHVIS NDVCAQVHPQ KVTKFMLCAG 200
RWTTGGKSTCS GDSGGPLVCN GVLQGITSW G SEPCALPERP SLYTKVVHYR 250
KWIKDTIVAN P 261

Trypsin
NK
NKSVILLGR
NKSVILLGRHSLF...MSLLK
AVCGGVLHPQWVLTAH CIRNK
AVCGGVLHPQWVLTAH CIRN KS VILLGR

ChymoT
TAAHCIRNKSVIL
TAAHCIRNKSVILL
TAAHCIRNKSVI...HSL
VLTAAH CIRN KS VIL
VLTAAH CIRN KS VILL

ArgC
NKSVILLGR
NKSVILLGRH...LLKNR
NKSVILLG...LMLLR
AVCGGVLHPQ...ILLGR

PSA is very stable in solution with multiple di-sulfide bonds.
Location of R/K suggest trypsin approach for digestion could be a eventual problem.
ChymoT difficult to control
gPRG 2012 Study

• Study samples: Lee Biosolutions, PSA from seminal fluid
  – PSA: 100 ug
    • sufficient for protocol development
    • initial characterization
  – PSA high isoform: 20 ug
    • differential characterization & quantification

• Time line
  – Sample distribution end of march March 2012
  – Reports due mid July 2012
  – Presentation at HUPO, September 2012, Boston
  – Presentation at ABRF, March 2013, Palm Springs
References

8. Peracaula R., Tabares G., Royle L., Harvey D.J., Dwek R.A., Rudd P.M., de Llorens R., Aletered glycosilation pattern allows the distinction between prostate-specific antigen (PSA) from normal and tumor origins,
### Supplemental Figure 2.

Conditions used for (a) enzymatic peptide digestion (b) glycoprotein treatment and (c) PNGase F N-glycan release.

<table>
<thead>
<tr>
<th>Lab</th>
<th>Enzyme</th>
<th>Qty (μg)</th>
<th>Buffer</th>
<th>pH</th>
<th>Red/Alk</th>
<th>Enz/Prot.</th>
<th>Digest Cond.</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Tryp</td>
<td>[2:4]</td>
<td>100mM Am Bic</td>
<td>8</td>
<td>DTT, IAA</td>
<td>1:50</td>
<td>24h Solution</td>
<td></td>
</tr>
<tr>
<td>2 (a)</td>
<td>Tryp Sequencing Grade, Roche</td>
<td>2.5</td>
<td>50mM Am Bic</td>
<td>8</td>
<td>DTT, IAA</td>
<td>1:50</td>
<td>37C, 15h SDS-PAGE</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Tryp</td>
<td>100-20</td>
<td>6M UREA, exchanged 30 kDa MWC</td>
<td>8</td>
<td>DTT, TCEP</td>
<td>1:50</td>
<td>37C Solution</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Tryp reductively methylated, TPK</td>
<td>[10:3]</td>
<td>1.6M Gdn-&gt; 0.1M Gdn, 50mM Am Bic</td>
<td>8.3</td>
<td>DTT, IAA</td>
<td>1:10</td>
<td>37C, 18h Solution</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Tryp sequencing modified, Promega</td>
<td>3</td>
<td>10 mM Tris, 0.6 M Urea</td>
<td>7.8</td>
<td>-</td>
<td>1:20</td>
<td>OV Solution</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Tryp Gold, Promega</td>
<td>[6:4]</td>
<td>50mM PBS, 150mM NaCl</td>
<td>7-7.5</td>
<td>-</td>
<td>1:25</td>
<td>37.5C, 16h Solution</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Tryp</td>
<td>&lt;1</td>
<td>0.5M Triethyl Am Bic</td>
<td>8.5</td>
<td>TCEP, MTS</td>
<td>1:20</td>
<td>37C, 3h Solution</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Tryp Sequencing grade, Promega</td>
<td>3</td>
<td>50mM Am Bic, 10%ACN</td>
<td>-</td>
<td>-</td>
<td>1:50</td>
<td>37C, 13h SDS-PAGE</td>
<td></td>
</tr>
<tr>
<td>9 (a)</td>
<td>Tryp, Promega</td>
<td>[1:3]</td>
<td>8M UREA -&gt; 1M UREA, 50mM Am Bic</td>
<td>8</td>
<td>DTT, IAA</td>
<td>1:40</td>
<td>37C, 12h Solution</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Tryp</td>
<td>10</td>
<td>1.6M Gdn-&gt; 0.3M Gdn, 50mM Am Bic</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>SDS-PAGE</td>
</tr>
<tr>
<td>11 (31)</td>
<td>ChymoTryp</td>
<td>[10:50]</td>
<td>50 mM Am Bic</td>
<td>8</td>
<td>-</td>
<td>1:50</td>
<td>18h Solution</td>
<td>Immob.bead</td>
</tr>
<tr>
<td>12</td>
<td>ChymoTryp</td>
<td>2</td>
<td>100mM Tris, 10mM CaCl₂</td>
<td>8</td>
<td>-</td>
<td>1:20</td>
<td>50C, 1.5h Barocycler Solution</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Tryp and ChymoTryp</td>
<td>1</td>
<td>50mM Am Bic</td>
<td>8</td>
<td>-</td>
<td>1:100</td>
<td>57C, 30min Solution</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>ArgC</td>
<td>8</td>
<td>90mM Tris, 8.5mM CaCl₂</td>
<td>7.6</td>
<td>DDT, EDTA</td>
<td>1:50</td>
<td>37C, 24h Solution</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>ArgC-Tryp</td>
<td>[4:10]</td>
<td>50 mM Am Bic</td>
<td>TCEP, IAA</td>
<td>Arg 1:50</td>
<td>Tryp 1:100</td>
<td>ArgC , 24h Solution</td>
<td>Tryp :2h</td>
</tr>
<tr>
<td>16</td>
<td>LysC</td>
<td>10</td>
<td>0.1M Am Bic</td>
<td>8</td>
<td>S-CarboMethyl</td>
<td>1:50</td>
<td>24h Solution</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Lysyl Endopeptidase, Wako</td>
<td>7</td>
<td>100mM Tris-HCl</td>
<td>8.5</td>
<td>BABAC</td>
<td>1:20</td>
<td>37C, 16h Solution</td>
<td></td>
</tr>
<tr>
<td>18 (a)</td>
<td>LysC</td>
<td>[14-70]</td>
<td>80mM TEAB, 10% ACN</td>
<td>8.6</td>
<td>-</td>
<td>1:75-1:60</td>
<td>37C, 4h Solution</td>
<td></td>
</tr>
<tr>
<td>19(b)</td>
<td>Arg C</td>
<td></td>
<td>50 mM Am Bic</td>
<td>8.3</td>
<td>yes</td>
<td>1:50</td>
<td>37C, 14h Solution</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lab</th>
<th>Qty (μg)</th>
<th>Red/Alk</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 (b)</td>
<td>[14-70]</td>
<td>Yes</td>
</tr>
<tr>
<td>19 (b)</td>
<td>15</td>
<td>Yes</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>21</td>
<td>30</td>
<td>-</td>
</tr>
</tbody>
</table>

(c)

<table>
<thead>
<tr>
<th>Labs</th>
<th>Qty (μg)</th>
<th>Enzyme</th>
<th>Digest Cond.</th>
<th>Note</th>
<th>Treatment of Released Glycans</th>
<th>Purification</th>
</tr>
</thead>
<tbody>
<tr>
<td>9 (21)</td>
<td>1</td>
<td>2Unit</td>
<td>37C, 24h</td>
<td>Immobilized on PVDFMembrane</td>
<td>Red. 3h 50°C, 1 M NaBH4 50 mM KOH</td>
<td>SCX SPE PGC SPE</td>
</tr>
<tr>
<td>2 (21)</td>
<td>2.5</td>
<td>-</td>
<td>37C, 24h</td>
<td>Immobilized on PVDFMembrane</td>
<td>Red. 3h 50°C, 1 M NaBH4 50 mM KOH</td>
<td>SCX SPE PGC SPE</td>
</tr>
<tr>
<td>22</td>
<td>10</td>
<td>5ug</td>
<td>37C, 18h</td>
<td>Solution, pH 7.5, 50mM PBS</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>23 [60-12]</td>
<td>4000 U/ml</td>
<td>37, 16h</td>
<td>pH 8, 50mM Am Bi and ME, 0.1% SDS, 0.75% NP-40</td>
<td>Permethylated</td>
<td>Methanol precipitation</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>12</td>
<td>0.7 Unit</td>
<td>37, 20h</td>
<td>pH 7.5, 5mM PBS, SDS , 0.5mM DDT, NP-40</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
**Supplemental Figure 3.** Bioinformatics tools used by participating laboratories to interpret tandem data acquired on (a) glycopeptides (bottom-up method), (b) glycoprotein (top-down approach), (c) N-glycans (PNGase F release method).

(a). Glycopeptides (Bottom-up approach)

- Lab 1, 4, 6, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17: Manual Interpretation

  - Lab 2(a): FindPept (ExPASy)
    - FindPept = identify peptides resulting from unspecific cleavage of proteins, taking into account chemical modifications, post-translational modifications (PTM) and protease cleavage

  - Lab 3: GlycoPep DB
    - GlycoPep DB Discover, assignment for glycopeptide by comparing measured masses to calculated masses from a carbohydrate database, b) assigned CID / ETD data

  - Lab 8: Sim Glycan TM
    - predicts the structure of glycans and glycopeptides by matching experimental MS/MS data against a database and scoring the proposed structures

  - Lab 6: GlyPID
    - Based on spacing between fragment peaks indicative of glycan fragmentation

(b). Glycoprotein (Top-down approach)

- Lab 20 and 21: Manual Interpretation

  - Lab 18: Prosight PC 3.0+Byonics
    - Prosight PC = protein identification which allows identification and characterization of intact proteins and their post-translational modifications (PTMs) using the method of top-down sequencing

  - Lab 19: Glycoquest + ProteinScape
    - Glycoquest = glyco search engine against the database DB
    - Protein Scape = Combined the results of Glycoquest and peptides search glycopeptides

(c). Released N-Glycans (PNGase F approach)

- Lab 9(b), 22, 23: manually interpreted the tandem MS/MS data, Glycomod was used by lab 2(b)
Supplemental Figure 4. Sample Integrity: N-glycan profile of (a) PSA and (b) PSA high isoform presented by laboratories 23 and 9.
Supplemental Figure 5. Average relative intensity of each glycan composition for the consensus cluster C (a) major & (b) intermediate glycans.