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

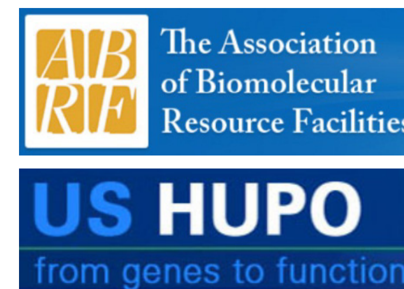
Expansion of the group is needed!

ABRF GlycoProtein Research Group (gPRG)

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

Science, Carbohydrates and Glycobiology, 23 March 2001, Vol. 291 no. 5512.

- **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

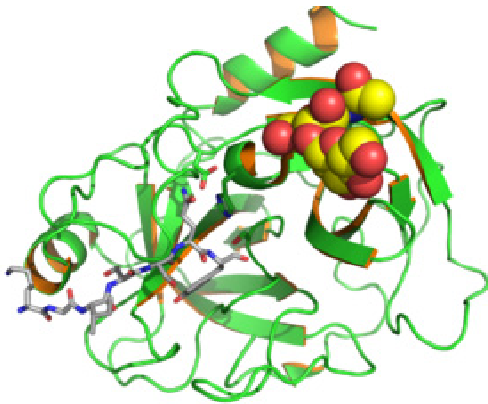
2012 gPRG study : Human Prostate Specific Antigen

- **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

		Score	Mass	Matches	Sequences
1	KLK3_HUMAN	1883	29293	230 (101)	79 (48)
2	FINC_HUMAN	215	266034	17 (9)	17 (9)
3	TPIS_HUMAN	156	26938	15 (6)	10 (4)
4	CTRB1_HUMAN	65	28479	3 (3)	3 (3)
5	MUC6_HUMAN	64	258219	5 (1)	5 (1)
6	MSMB_HUMAN	55	13483	9 (2)	7 (2)
7	TIMP1_HUMAN	49	23840	6 (3)	5 (3)
8	SAP_HUMAN	48	59899	6 (2)	4 (2)
9	CRIS3_HUMAN	43	28524	7 (5)	6 (5)
10	CRIS2_HUMAN	36	28153	12 (3)	7 (3)
11	KLD7B_HUMAN	36	54003	2 (1)	2 (1)
12	THG1_HUMAN	34	35151	2 (1)	2 (1)
13	ACSA_HUMAN	31	79613	4 (1)	4 (1)
14	ECM1_HUMAN	31	62232	2 (1)	2 (1)

```
  1 MWVPVVFLTL SVTWIGAAPL ILSRIVGGWE CEKHSQPWQV LVASRGRAVC
 51 GGVLVHPQWV LTAAHCIRNK SVILLGRHSL FHPEDTGQVF QVSHSFPHPL
101 YDMSLLKNRF LRPGDDSSHD LMLLRLSEPA ELTDAVKVMD LPTQEPALGT
151 TCYASGWGSI EPEEFLTPKK LQCVDLHVIS NDVCAQVHPQ KVTKFMLCAG
201 RWTGGKSTCS GDSGGPLVCN GVLQGITSWG SEPCALPERP SLYTKVVHYR
251 KWIKDTIVAN P
```

Sequence Coverage: ~ **80 %**

PSA Batch M02015

		Score	Mass	Matches	Sequences
1	KLK3_HUMAN	159	29293	66 (16)	33 (10)
2	CTRA_BOVIN	56	26220	15 (2)	14 (2)
3	KTHY_METHJ	51	23197	1 (1)	1 (1)
4	M3K11_RAT	42	93622	2 (1)	2 (1)
5	RSE1_EMENI	41	135146	2 (1)	2 (1)
6	DIF_DROME	39	74630	2 (1)	1 (1)
7	ALDR_BOVIN	35	36296	2 (1)	2 (1)
8	KHSE_SALAR	33	33631	2 (1)	1 (1)
9	TIMP1_HUMAN	32	23840	3 (1)	3 (1)
10	RF3_CYAP7	32	61303	1 (1)	1 (1)
11	RGA_ARATH	29	64565	1 (1)	1 (1)
12	K2C1_HUMAN	28	66170	1 (1)	1 (1)
13	Y3985_DICDI	27	5616	1 (1)	1 (1)
14	AMYG_ASPAW	25	68837	1 (1)	1 (1)

```
MWVPVVFLTL SVTWIGAAPL ILSRIVGGWE CEKHSQPWQV LVASRGRAVC
GGVLVHPQWV LTAAHCIRNK SVILLGRHSL FHPEDTGQVF QVSHSFPHPL
YDMSLLKNRF LRPGDDSSHD LMLLRLSEPA ELTDAVKVMD LPTQEPALGT
TCYASGWGSI EPEEFLTPKK LQCVDLHVIS NDVCAQVHPQ KVTKFMLCAG
RWTGGKSTCS GDSGGPLVCN GVLQGITSWG SEPCALPERP SLYTKVVHYR
KWIKDTIVAN P
```

Sequence Coverage: **~ 70 %**

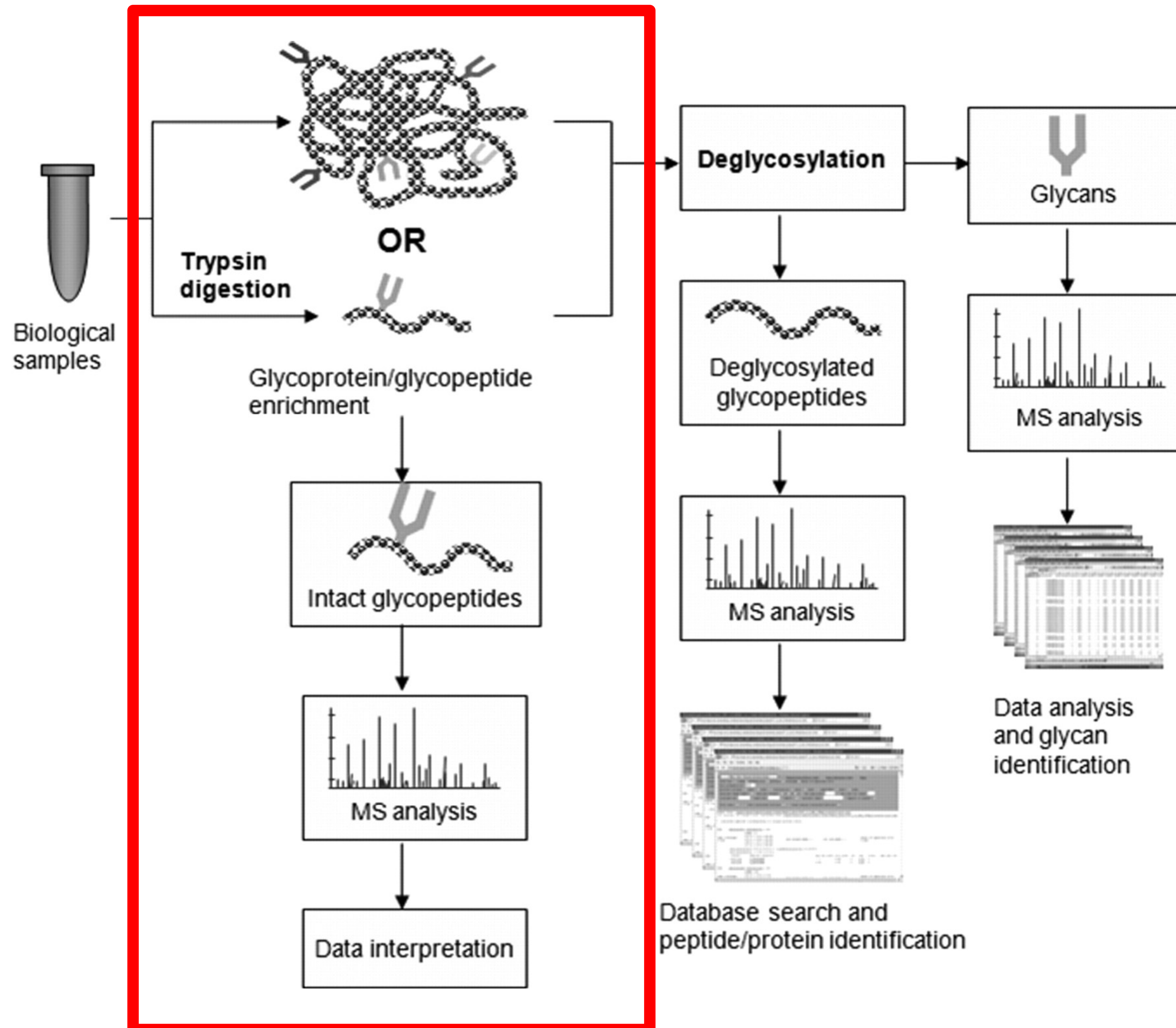
Both forms are relatively pure. Initial coverage indicates N-linked at aa.69.

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
YDMSLLKNRF LRPGDDSSHD LMLLRLSEPA ELTDAVKVMD LPTQEPALGT 150
TCYASGWGSI EPEEFLTPKK LQCVDLHVIS NDVCAQVHPQ KVTKFMLCAG 200
RWTGGKSTCS GDSGGPLVCN GVLQGITSWG SEPCALPERP SLYTKVVHYR 250
KWIKDTIVAN P                                           261
```

Trypsin

NK
NKSVILLGR
NKSVILLGRHSLFH…MSLLK
AVCGGVLVHPQWVLTAAHCIRNK
AVCGGVLVHPQWVLTAAHCIRNKSVILLGR

ChymoT

TAAHCIRNKSVIL
TAAHCIRNKSVILL
TAAHCIRNKSVI..HSL
VLTAAHCIRNKSVIL
VLTAAHCIRNKSVILL

ArgC

NKSVILLGR
NKSVILLGRHS..LLKNR
NKSVILLG..LMLLR
AVCGGVLVHPQ..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

1. Ahmedin J. , Siegel R., Ward E., Hao Y., Xu J., Murray T., Thun M.J., Cancer Statistics 2008, *CA Cancer J Clin* , 2008, 58 (2), 71–96.
2. Pan S., Chen R., Aebersold R., Brentnall A.T., *Mass Spectrometry Based Glycoproteomics - From a Proteomics Perspective, Mol. Cell. Prot*., 2011, Vol 10 (1) 1-14.
3. Varki A. , Cummings R.D., Esko J.D., Freeze H.H., Stanley P., Bertozzi C.R., Hart G.W., and Etzler M.E., Essentials of Glycobiology, 2nd edition, Cold Spring Harbor (NY), 2009.
4. Okada T, Sato, Y., Kobayashi N., Sumida K., Satomura S., Matsuura S., Takasaki M., Endo T., Structural characteristics of N-Glycans of two isoforms of prostate-specific antigens purified from human seminal fluid, *Biochim. Biophys. Acta,* 2001, 1525, 149-160
5. Tabares G., Radcliffe C.M., Barrabes S., Ramirez M., Aleixandre R.N., Hoesel W., Dwek R.A., Rudd P.M., Peracaula R., de Llorens R.., Different glycan structures in prostate-specific antigen from prostate cancer sera in relation to seminal plasma PSA, *Glycobiology*, 2006, 16(2), 132-145
6. Barnidge DR, Goodmanson MK, Klee GG, Muddiman DC., Absolute qauntification of the model biomarker PSA in serum by LC/MS using protein cleavage and isotope dilution mass spectrometry, *J Proteome Res.,* 2004, 3(3), 644-652.
7. Ohyama C, Hosono M, Nitta K, Oh-eda M, Yoshikawa K, Habuchi T, Arai Y, Fukuda M., Carbohydrate structure and differential binding of prostate specific antigen to Maackia amurensis lectin between prostate cancer and benign prostate hypertrophy, *Glycobiology,* 2004 , 14(8), 671-679
8. Peracaula R., Tabares G., Royle L., Harvey D.J., Dwek R.A., Rudd P.M., de Llorens R., Aletered glycosilation pattern allows the disctinction between prostate-specific antigen (PSA) from normal and tumor origins*,*
9. Tajiri M., Ohyama C., Wada Y., Oligosaccharide profiles of the PSA in free and complexed forms from the prostate cancer patient serum and seminal plasma : a glycopeptide approach, Glycobiology, 2008, 18(1), 2-8
10. Saldova R, Fan Y, Fitzpatrick JM, Watson RW, Rudd PM., Core fucosylation and alpha 2-3 sialylation in serum N-glycome is significant increased in prostate cancer comparing to benign prostate hyperplasia, *Glycobiology*, 2011,21(2):195-205.
11. Sarrats A, Saldova R, Comet J, O'Donoghue N, de Llorens R, Rudd PM, Peracaula R., Glycan Characterixation of PSA 2.DE subfroms from serum and seminal plasma, *OMICS,* 2010, 14(4):465-74.
12. Yan Li, Yuan Tian, Taha Rezai, Amol Prakash, Mary F. Lopez, Daniel W. Chan, and Hui Zhang, Simultaneous Analysis of Glycosylated and Sialylated Prostate-Specific Antigen Revealing Differential Distribution of Glycosylated Prostate-Specific Antigen Isoforms in Prostate Cancer Tissues, Anal. Chem., 2011, 83 (1), pp 240–245.
13. Sarrats A., Comet J., Tabarés G., Ramírez M., Aleixandre R.N., de Llorens R., Peracaula R., Differential percentage of Serum PSA subfroms suggests a new way to improve prostate cancer diagnosis, Prostate, 2010, 70, 1-9.

Supplemental Figure 2. Conditions used for (a) enzymatic peptide digestion (b) glycoprotein treatment and (c) PNGase F *N*-glycan release.

(a)

Lab	Enzyme	Qty (µg)	Buffer	pH	Red/Alk	Enz/Prot.	Digest Cond.	Note
1	Tryp	[2;4]	100mM Am Bic	8	DTT, IAA	1:50	24h	Solution
2 (a)	Tryp Sequencing Grade, Roche	2.5	50mM Am Bic	8	DTT, IAA	1:50	37C, 15h	SDS-PAGE
3	Tryp	100-20	6M UREA, exchanged 30 kDa MWC	8	DDT, TCEP	1:50	37C	Solution
4	Tryp reductuvely methylated, TPCK	[10;3]	1.6M Gdn-> 0.1M Gdn, 50mM Am Bic	8.3	DDT, IAA	1:10	37C, 18h	Solution
5	Tryp sequencing modified, Promega	3	10 mM Tris, 0.6 M Urea	7.8	-	1:20	OV	Solution
6	Tryp Gold, Promega	[.6;.4]	50mM PBS, 150mM NaCl	7-7.5	-	1:25	37.5C, 16h	Solution
7	Tryp	<1	0.5M Triethyl Am Bic	8.5	TCEP, MTS	1:20	37C, 3h	Solution
8	Tryp Sequencing grade, Promega	3	50mM Am Bic, 10%ACN	-	-	1:50	37C, 13h	SDS-PAGE
9 (a)	Tryp, Promega	[1;3]	8M UREA -> 1M UREA, 50 mM Am Bic	8	DDT, IAA	1:40	37C, 12h	Solution
10	Tryp	10	1.6M Gdn-> 0.3M Gnd, 50mM Am Bic	-	-	-	-	SDS-PAGE
11 (31)	ChymoTryp	[10;50]	50 mM Am Bic	8	-	1:50	18h	Immob.bead
12	ChymoTryp	2	100mM Tris, 10mM $CaCL_2$	8	-	1:20	50C, 1.5h Barocycler	Solution
13	Tryp and ChymoTryp	1	50mM Am Bic	8	-	1:100	57C, 30min	Solution
14	ArgC	8	90mM Tris, 8.5mM $CaCL_2$	7.6	DDT, EDTA	1:50	37C, 24h	Solution
15	ArgC-Tryp	[4;10]	50 mM Am Bic		TCEP, IAA	Arg 1:50 Tryp 1:100	ArgC , 24h Tryp :2h	Solution
16	LysC	10	0.1M Am Bic	8	S-CarboMethyl	1:50	24h	Solution
17	Lysyl Endopeptidase, Wako	7	100mM Tris-HCl	8.5	BABAC	1:20	37C, 16h	Solution
18 (a)	LysC	[14-70]	80mM TEAB, 10% ACN	8.6	-	1:75-1:60	37C, 4h	Solution
19(b)	Arg C		50 mM Am Bic	8.3	yes	1:50	37C, 14h	Solution

(b)

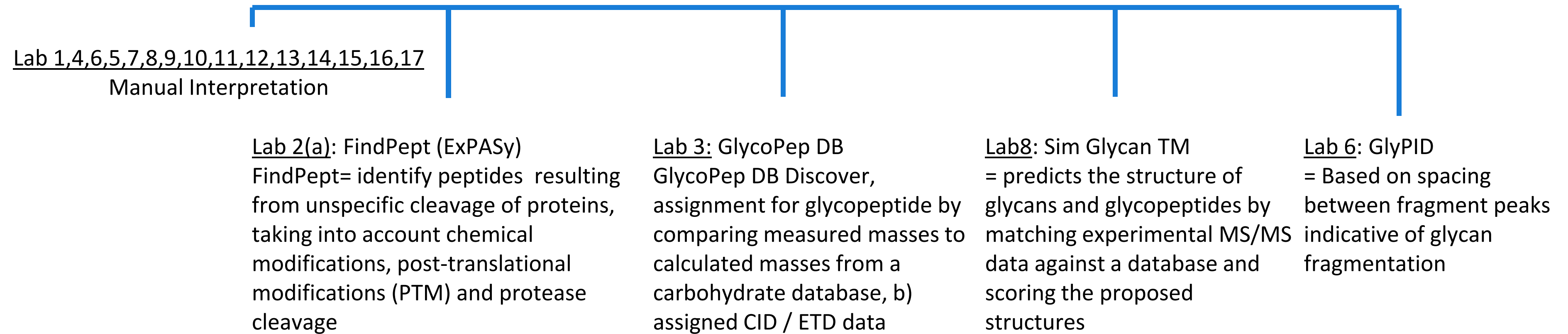
Lab	Qty (µg)	Red/Alk
18 (b)	[14-70]	Yes
19 (b)	15	Yes
20	1	-
21	30	-

(c)

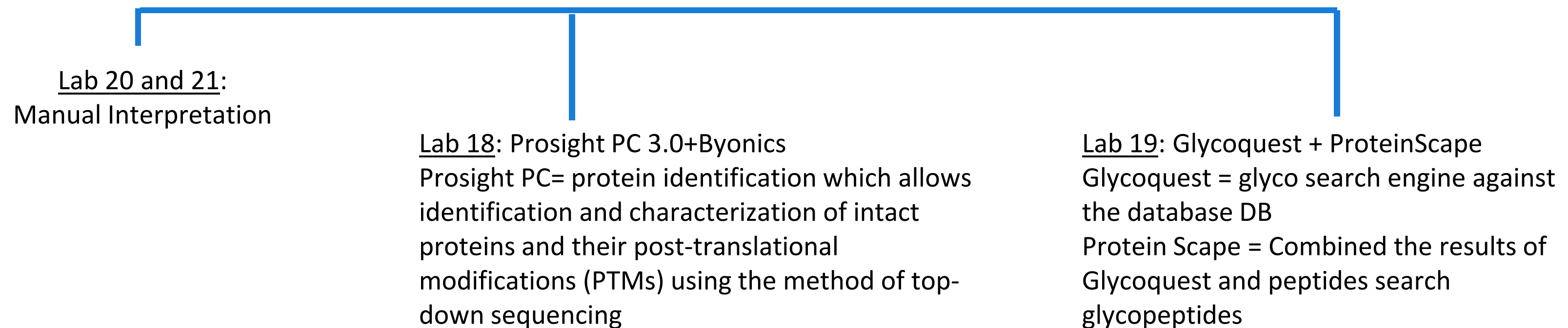
Labs	Qty (µg)	Enzyme	Digest Cond.	Note	Treatment of Released Glycans	Purification
9 (21)	1	2Unit	37C, 24h	Immobilized on PVDFMembrane	Red. 3h 50°C, 1 M NaBH4 50 mM KOH	SCX SPE PGC SPE
2 (21)	2.5	-	37C, 24h	Immobilized on PVDFMembrane	Red. 3h 50°C, 1 M NaBH4 50 mM KOH	SCX SPE PGC SPE
22	10	5ug	37C, 18h	Solution, pH 7.5, 50mM PBS	-	-
23	[60-12]	4000 U/ml	37, 16h	pH 8, 50mM Am Bi and ME, 0.1% SDS, 0.75% NP-40	Permethylated	Methanol precipitation
24	12	0.7 Unit	37, 20h	pH 7.5, 5mM PBS, SDS , 0.5mM DDT, NP-40	-	-

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)



(b). Glycoprotein (Top-down approach)

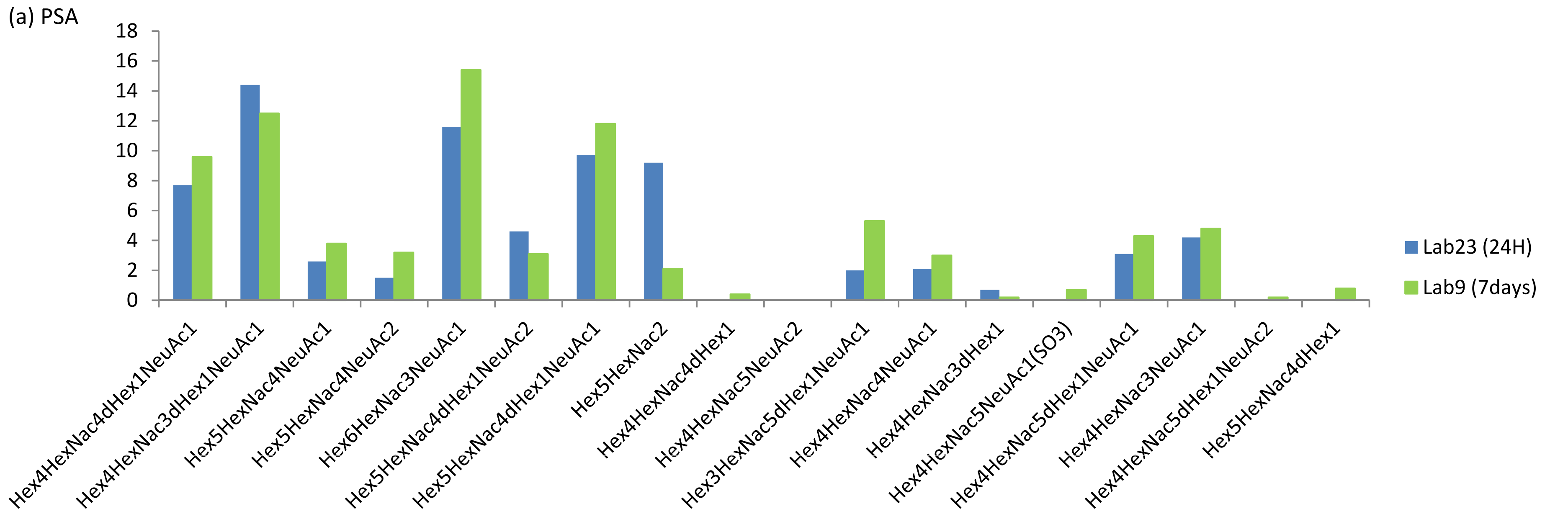


(c) . Released *N*-Glycans (PNGase F approach)

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.

(a) PSA



(b) PSA high isoform

18
16
14
12
10
8
6
4
2
0

Hex4HexNac4dHex1NeuAc1
Hex4HexNac3dHex1NeuAc1
Hex5HexNac4NeuAc1
Hex5HexNac4NeuAc2
Hex6HexNac3NeuAc1
Hex5HexNac4dHex1NeuAc2
Hex5HexNac4dHex1NeuAc1
Hex5HexNac2
Hex4HexNac4dHex1
Hex4HexNac5NeuAc2
Hex3HexNac5dHex1NeuAc1
Hex4HexNac4NeuAc1
Hex4HexNac3dHex1
Hex4HexNac5NeuAc1(SO3)
Hex4HexNac5dHex1NeuAc1
Hex4HexNac3NeuAc1
Hex4HexNac5dHex1NeuAc2
Hex5HexNac4dHex1

Lab23 (24H)
Lab9 (7days)

Supplemental Figure 5. Average relative intensity of each glycan composition for the consensus cluster C (a) major & (b) intermediate glycans.