Published in final edited form as:
J Am Chem Soc. 2007 June 13; 129(23): 7266–7267. doi:10.1021/ja0724083.

Tailored glycoproteomics and glycan site mapping using saccharide-selective bioorthogonal probes

Sarah R. Hanson[1], **Tsui-Ling Hsu**[1,2], **Eranthie Weerapana**[1], **Kuniyuki Kishikawa**[1], **Gabriel M. Simon**[1], **Benjamin F. Cravatt**[1,*], and **Chi-Huey Wong**[1,2,*]

1 *The Department of Chemistry, The Scripps Research Institute, 10550 N. Torrey Pines Rd, La Jolla, CA 92037*

2 *Genomics Research Center, Academia Sinica, 128 Section 2, Academia Road, Nankang, Taipei 115, Taiwan*

Glycosylation is a complex co-/post-translational modification of proteins that serves to govern important biological phenomena ranging from protein stability and folding, to cellular trafficking, signaling, and modulation of macromolecular interactions.[1] Despite an increasing awareness of its importance, much of the basic information about glycosylation and its structural and molecular function is lacking.[2] This is due largely to the complexity of the glycosylation process, which is multifaceted and non-templated, culminating in both variability in glycosylation site occupancy and structural microheterogeneity among glycans (glycoforms) within occupied sites. In addition, there are several classes of glycoproteins that differ by protein linkage (most commonly, N-linked to Asn or O-linked to Ser/Thr), saccharide composition, degree of branching, and size. The remarkable complexity of glycans presents major challenges to deciphering their structures and activities on an individual protein, let alone, proteomic scale.[3] These challenges include identifying glycoproteins, sites of protein modification, and determining information about saccharide composition; in addition to, ultimately, understanding the direct roles of glycans in cellular function and dysfunction.[4] Herein, we outline a glycoproteomic strategy for saccharide-selective glycoprotein identification (ID) and glycan mapping (GIDmap) that generates glycoproteins tailored with bioorthogonally-tagged alkynyl saccharides that can be selectively isolated, allowing for glycoprotein ID and glycan site mapping via liquid chromatography-tandem mass spectrometry (LC-MS2).

GIDmap uses saccharide-selective probes to capture specific secretory glycoprotein subpopulations from proteomes (Figure 1). First, metabolic oligosaccharide engineering (MOE) is employed to insert sugar analogs appended with a bioorthogonal alkyne group in place of their native counterparts via promiscuous glycan synthesis pathways in cells.[5,6] Our previous studies demonstrate that alkynyl sugar derivatives of fucose (Fucyne) and N-acetylmannosamine (ManNAcyne **1**) are incorporated into fucosylated and sialylated proteins, respectively, where they can be selectively labeled using the bioorthogonal Cu(I) catalyzed [3 +2] azide-alkyne cycloaddition, or click chemistry (CC).[6] In GIDmap, alkynyl-glycans are labeled with a biotin-azide CC partner, permitting their enrichment from proteomes via affinity capture with immobilized streptavidin. Tandem protein ID and glycan site mapping are then carried out on-bead, akin to previous strategies.[7,8] First, non-glycosylated peptides are harvested by tryptic digestion, allowing for total protein ID. Analysis of the remaining captured glycopeptides is achieved by treatment with peptide-N-glycosidase F (PNGase), which hydrolyzes the amide bond between the biotinylated glycan and Asn residue of the bound

E-mail: cravatt@scripps.edu; wong@scripps.edu.

peptide. The shift from Asn to Asp at formerly glycosylated sites can be identified by a differential modification (diffmod) of +1 Da on Asn in SEQUEST searches of MS data. [9]

Here, we apply GIDmap to analyze the sialylated N-linked glycoproteome of prostate cancer (PC3) cells. All experiments were performed on 1.5 mg of total cellular protein harvested from PC3 cells grown in the presence of **1**, or untagged ManNAc. Peptides were analyzed by multidimensional nano-LC MS^2 (MudPIT).[9] Notably, manual inspection of peptides with an Asn diffmod showed MS^2 spectra where all b and y ions containing the modification were clearly shifted by +1 Da (supporting information). In glycoproteomes from ManNAcyne-treated cells, very specific enrichment of N-glycopeptides was noted in PNGase-released peptides. Of the 219 unique peptide IDs, over 97% contained a modified Asn within the established N-glycosylation consensus sequence: N-X-T/S, where X is not proline.[10] By comparison, our bioinformatics analysis predicts that only 12.7% of Asn residues within the searched human proteome fall into this consensus sequon, confirming specific enrichment of N-glycopeptides. Negative control glycoproteomes showed negligible IDs after PNGase treatment, further demonstrating selectivity for tagged glycopeptides.

In total, GIDmap identified 219 unique N-glycosylated peptides representing 108 non-redundant glycoproteins. 75 of these proteins had strong IDs in the tryptic and PNGase phases. Analysis of the other 33 (resulting from ID mainly in PNGase phase) strongly indicates that they are also true N-glycopeptides enriched from underrepresented (i.e. low abundance) proteins in the tryptic digest. This set was discriminated by several checks including reproducibility in triplicate runs, coverage by multiple glycopeptides, and/or agreement with experimentally assigned glycosylation sites. The number of N-glycosylation sites found per protein ranged from 1 to 7, with an average of 2. The N-glycosylation site IDs were sorted according to Swiss-Prot database annotation,[11] which indicates if sites have associated experimental evidence, ‘verified’, or whether they have been predicted based on homology and/or computational programs, ‘potential’. As depicted in Figure 2a, out of the 219 mapped sites, only 69 (32%) fell into a verified status. Notably, at least 1/3 of these (23) were only recently found by other glycoproteomic mapping endeavors.[12] The majority of hits represent previously uncharacterized glycosylation sites, 113 (52%) of which were annotated as potential, and 37 (17%) of which are novel sites, previously not annotated (22 are from proteins of unknown function). Consistent with known N-linked glycoprotein distribution, the majority of IDs were membrane-bound receptors, transporters, adhesion molecules, and components of subcellular locations rich in glycoproteins (lysosome, ER, and golgi).[13] Interestingly, around 26% (28) of the protein IDs had known associations with tumor progression and/or metastasis. Comparison with healthy prostate tissue is currently underway to determine the role sialic acid might play in protein dysfunction.

GIDmap contributes important advances to the emerging collection of glycoproteome characterization methods that seek to enrich low abundance glycoproteins as a primary step.[4] Previous isolation strategies for secretory glycoproteins have exploited cis-diol chemistry of saccharide chains to immobilize total glycan populations, or immobilized lectins to enrich subpopulations of N-glycosylated proteins and/or peptides after tryptic digestion.[8] GIDmap offers the combined advantage of covalent immobilization and subpopulation enrichment using chemistry that is non-destructive to peptides and glycans. On the other hand, GIDmap requires MOE in cell systems and is thusly not applicable to certain proteomes (e.g. human plasma). We view this feature as a tradeoff, since MOE also empowers GIDmap to tailor the isolation of specific glycoproteins based on their unique carbohydrate composition. This capability not only adds a precise saccharide-selective dimension to traditional glycoprotein isolation, but also relays specific details regarding glycan content. For example, by using GIDmap, specific glycosylation events, such as sialylation and fucosylation, can be directly compared by analyzing cells treated with ManNAcyne and Fucyne, respectively.[6] Such discrimination

should prove useful for determining how these saccharides are involved in protein dysfunction. Indeed, aberrant terminal sialylation and hyper-fucosylation has been documented in several cancers.[14]

Further on-bead analysis of affinity captured material should also be possible by GIDmap, following protein ID and N-glycan site mapping (Figure 1). Site mapping of O-glycans can be incorporated by established routes, like BEMAD (alkaline induced β-elimination of glycans followed by Michael addition, usually by a thiol).[15] Furthermore, total glycomic analysis may be feasible by chemically eluting affinity-captured saccharide moieties and subjecting them to glycan sequencing technology.[4] Notably, this additional step would not be possible using chemical immobilization strategies, since the carbohydrate structure is destroyed by the covalent immobilization chemistry; lectin affinity methods are also not amenable because glycans must be cleaved from peptides off-resin, requiring a complex separation of peptides and glycans.[8]

Overall, GIDmap represents a powerful and robust method for analyzing distinct facets of secretory glycoproteins on a proteome-wide scale. In this communication, the effectiveness of GIDmap to inventory the glycoproteome was demonstrated by the analysis of sialylated N-linked glycoproteins from PC3 cells. Over 200 N-glycosylation site IDs, 68% of which had no previous experimental documentation, were inventoried and mapped. Future experiments will not only serve to further establish glycosylation sites, glycan linkage, and occupancy by specific saccharides, but will also assist to identify and better understand the role of glycans and glycoproteins in temporal- and stage-specific tissues, especially in healthy versus diseased samples. Such comparative analysis is currently underway.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by the NIH (CA087660, GM44154). SRH and TLH acknowledge fellowships from ARCS and TMS (094-1-B-003), respectively.

References

1. Varki, A.; Cummings, R.; Esko, JD.; Freeze, H.; Hart, GW.; Marth, J. Essentials of Glycobiology. Cold Spring Harbor Press; Cold Spring Harbor, NY: 1999.
2. Lowe JB, Marth JD. Annu Rev Biochem 2003;72:643–91. [PubMed: 12676797]
3. (a) Sears P, Wong CH. Science 2001;291:2344–50. [PubMed: 11269314] (b) Grogan MJ, Hanson S, Best M, Bryan MC, Wong CH. Trend Biochem Sci 2004;29:656–63. [PubMed: 15544952] (c) Brik A, Ficht S, Wong CH. Cur Opin Chem Biol 2006;10:638–44.
4. (a) Bond MR, Kohler JJ. Curr Opin Chem Biol 2007;11:52–8. [PubMed: 17174139] (b) Morelle W, Canis K, Chirat F, Faid V, Michalski JC. Proteomics 2006;6:3993–4015. [PubMed: 16786490]
5. (a) Prescher JA, Bertozzi CR. Cell 2006;126:851. [PubMed: 16959565]Laughlin, ST.; Agard, NJ.; Baskin, JM.; Carrico, IS.; Chang, PV.; Ganguli, AS.; Hangauer, MJ.; Lo, A.; Prescher, JA.; Bertozzi, CR.; Minoru, F. Meth Enzym. 415. Academic Press; 2006. p. 230-250.
6. Hsu TL, Hanson SR, Kishikawa K, Wang SK, Sawa M, Wong CH. Proc Natl Acad Sci 2007;104:2614–9. [PubMed: 17296930]
7. Speers AE, Cravatt BF. J Am Chem Soc 2005;127:10018–9. [PubMed: 16011363]
8. (a) Zhang H, Li XJ, Martin DB, Aebersold R. Nat Biotech 2003;21:660–6. (b) Kaji H, Saito H, Yamauchi Y, Shinkawa T, Taoka M, Hirabayashi J, Kasai K, Takahashi N, Isobe T. Nat Biotech 2003;21:667–72. (c) Kaji H, Isobe T. Trend Glycosci Glycotech 2006;18:313–22.

9. (a) Eng JK, McCormack AL, Yates JR. J Amer Soc Mass Spec 1994;5:976–89. (b) Washburn MP, Wolters D, Yates JR 3rd. Nat Biotech 2001;19:242–7.

10. Apweiler R, Hermjakob H, Sharon. Biochim Biophys 1999;1473:4.

11. According to Swiss-Prot database (www.expasy.org)

12. (a) Lewandrowski U, Moebius J, Walter U, Sickmann A. Mol Cell Proteomics 2006;5:226. [PubMed: 16263699] (b) Ramachandran P, Boontheung P, Xie YM, Sondej M, Wong DT, Loo JA. J Proteome Res 2006;5:1493. [PubMed: 16740002] (c) Liu T, Qian WJ, Gritsenko MA, CampIi DG, Monroe ME, Moore RJ, Smith RD. J Prot Res 2005;4:2070.

13. Roth J. Chem Rev 2002;102:285–304. [PubMed: 11841244]

14. (a) Dube DH, Bertozzi CR. Nat Rev Drug Discov 2005;4:477–88. [PubMed: 15931257] (b) Sell S. Hum Pathol 1990;21:1003–19. [PubMed: 2210723] (c) Shiraki K, Takase K, Tameda Y, Hamada M, Kosaka Y, Nakano T. Hepatology 1995;22:802–7. [PubMed: 7544756] (d) Comunale MA, Lowman M, Long RE, Krakover J, Philip R, Seeholzer S, Evans AA, Hann HWL, Block TM, Mehta AS. J Proteome Res 2006;5:3108–15.

15. (a) Wells L, Vosseller K, Cole RN, Cronshaw JM, Matunis MJ, Hart GW. Mol Cell Proteomics 2002;1:791–804. [PubMed: 12438562] (b) Khidekel N, Ficarro SB, Peters EC, Hsieh-Wilson LC. Proc Natl Acad Sci 2004;101:13132–7. [PubMed: 15340146] (c) Vosseller K, Trinidad JC, Chalkley RJ, Specht CG, Thalhammer A, Lynn AJ, Snedecor JO, Guan S, Medzihradszky KF, Maltby DA, Schoepfer R, Burlingame AL. Mol Cell Proteomics 2006;5:923–34. [PubMed: 16452088]

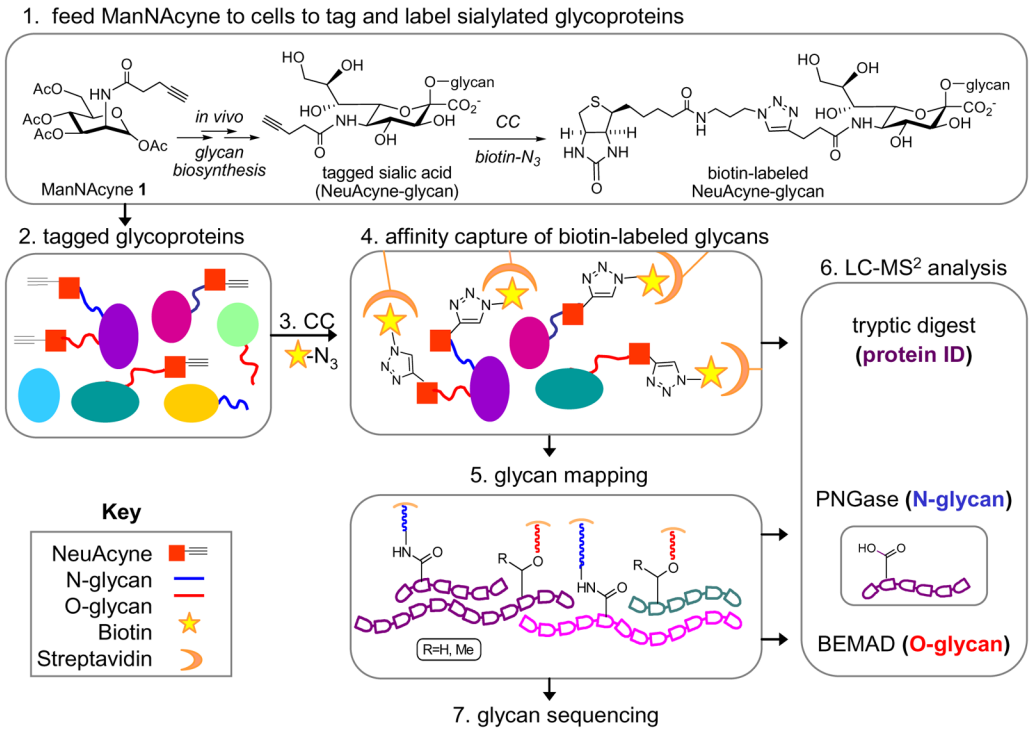

Figure 1.
Saccharide-selective glycoproteomic ID and glycan mapping.

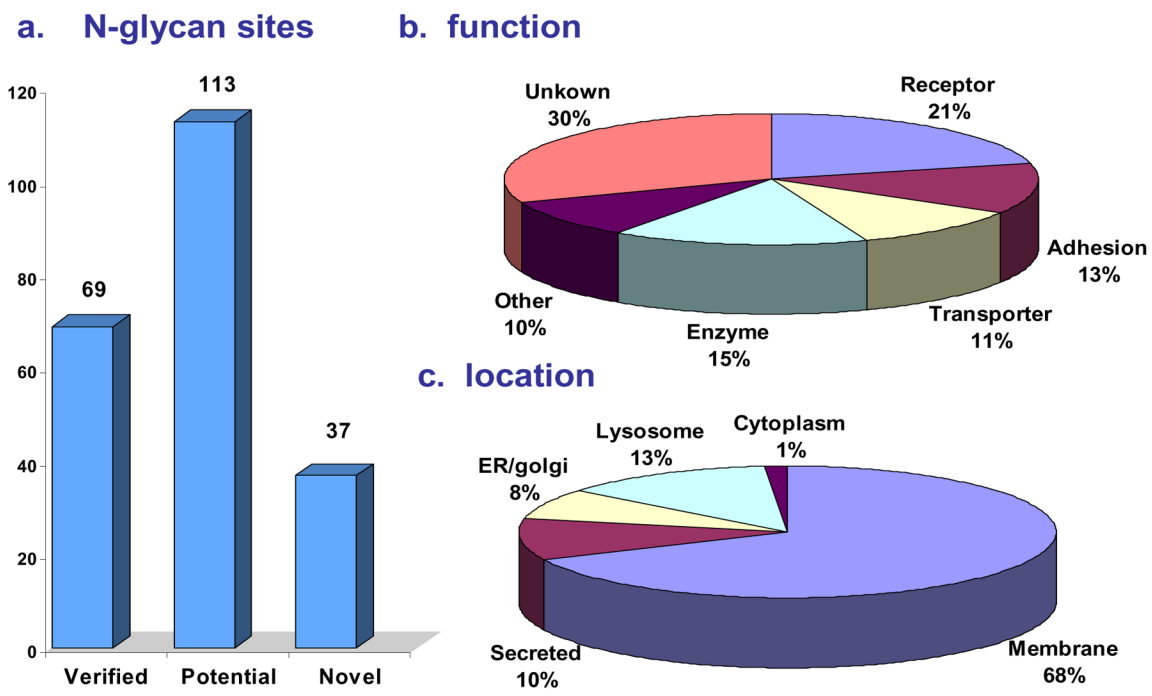

Figure 2.
Breakdown of N-sialylated glycoprotein IDs by GIDmap.