Chemoproteomic mapping of human milk oligosaccharide (HMO) interactions in cells

Abdullah A. Hassan, Jacob M. Wozniak, Zak Vilen, Weichao Li, Appaso Jadhav, Christopher G. Parker and Mia L. Huang

Human milk oligosaccharides (HMOs) are a family of unconjugated soluble glycans found in human breast milk that exhibit a myriad of biological activity. While recent studies have uncovered numerous biological functions for HMOs (antimicrobial, anti-inflammatory & probiotic properties), the receptors and protein binding partners involved in these processes are not well characterized. This can be attributed largely in part to the low affinity and transient nature of soluble glycan–protein interactions, precluding the use of traditional characterization techniques to survey binding partners in live cells. Here, we present the use of synthetic photoactivatable HMO probes to capture, enrich and identify HMO protein targets in live cells using mass spectrometry-based chemoproteomics. Following initial validation studies using purified lectins, we profiled the targets of HMO probes in live mouse macrophages. Using this strategy, we mapped hundreds of HMO binding partners across multiple cellular compartments, including many known glycan-binding proteins as well as numerous proteins previously not known to bind glycans. We expect our findings to inform future investigations of the diverse roles of how HMOs may regulate protein function.

Introduction

The interactions between glycans and proteins mediate a diverse range of biological processes, and aberrant glycosylation can result in a myriad of pathologies. Whereas the vast majority of the annotated glycome exists as either lipid- or protein-modified glycoconjugates, “free” (soluble or unconjugated) glycans represent an under-studied fraction of the glycome. From their initial discovery as plant defense signaling molecules and modulators of host–pathogen interactions, to the recent discovery of a mammalian disaccharide that induces autoimmunity, it is becoming increasingly conspicuous that free glycans are important biological regulators.

Human milk oligosaccharides (HMOs) are a family of free glycans that are abundantly found in breast milk. While initially believed to be mere nutritional agents, increasing evidence has highlighted HMOs as important signaling molecules involved in the activation of protein receptors and the resolution of inflammation. Some HMOs also display antimicrobial activity by acting as soluble decoys for pathogenic microbes or through inhibiting biofilm formation. HMOs share a common lactose (Galβ(1-4)Glc) core structure, which is often decorated with additional modifications, such as fucose or sialic acid residues (Fig. 1A). Importantly, such variants can impart vastly different biological activities. Despite growing evidence demonstrating that HMOs impact diverse biology important to human health, a molecular understanding of the mechanisms and protein partners through which HMOs act remains limited.

Similar to glycoconjugates, the interactions of free glycans with glycan-binding proteins are relatively weak ($K_D \sim 10^{-4\text{–}10^{-7}}$ M), dynamic, and prefer the physiological presentation of proteins in live cells. These inherent characteristics preclude the use of conventional affinity-based techniques, which require harsh sample preparation procedures. Although several methods have been developed to survey the protein-binding partners of glycoconjugates in live cells, these techniques have not been applied directly to the study of free glycans.

Thus, we hypothesized that photoaffinity labeling could be applied towards the study of HMO interactions in live cells. Photoaffinity labeling is able to capture transient and low affinity interactions, and has been previously combined with metabolic oligosaccharide engineering to profile interactions with glycoconjugates. To enable identification of potential HMO-binding proteins, we chose to utilize a ‘fully-functionalized’...
enrichment tag composed of a photoactivatable diazirine group for UV light-induced capture of HMO-bound protein targets and an alkyne handle for conjugation to azide-bearing reporter tags via copper-mediated azide-alkyne (CuAAC) “click” chemistry.26,27 We synthesized four probes composed of the most abundant HMOs in human breast milk (lactose or Lac, 2’fucosyl lactose or 2FL, 3’sialyl lactose or 3SL, and 6’sialyl lactose or 6SL), and first cross-evaluated their engagement with purified glycans in live cells.

Results and discussion

Using a similar experimental protocol previously reported by the Townsend group to prepare the 2FL probe (2),28 the chemical synthesis of our photoactivatable HMO probes commenced with the Kochetkov microwave-assisted conversion of commercially available HMO hemiacetals into their corresponding amine derivatives,29,30 followed by amide bond coupling with a ‘fully-functionalized’ bifunctional diazirine-alkyne acid linker (see ESI†). The probes 1–4 were synthesized in 44% to 53% overall yields (Fig. 1A). The anomeric positions were chosen based on lectin crystallography studies and biological data suggesting that this site would be the most inert position for derivatization.31–33

With probes in hand, we evaluated their engagement with plant lectins in vitro. Plant lectins, such as LTL, MAL II, and SNA, are commonly used reagents to describe glycan patterns in cells. Plant lectins are known to display a degree of selectivity among various glycans, but they also tolerate a wide array of substitutions (Fig. 1B).34–36 Upon incubation of the probes with the plant lectins and UV-light irradiation at 365 nm, the mixtures were reacted with tetramethylrhodamine (TAMRA) azide via CuAAC, separated using denaturing SDS-PAGE, and visualized by in-gel fluorescence scanning and silver staining (Fig. 1C).19 We observed dose-dependent (0–50 μM) photo-crosslinking of probes 2–4 with the lectins (Fig. 1C and Fig. S1, ESI†), and cross-linking was dependent on UV irradiation. Consistent with the multi-subunit nature of plant lectins, we observed several bands in SDS-PAGE corresponding to lectin monomers, dimers, or oligomers (Fig. S1, ESI†).

To further evaluate the selectivity of the probes and determine whether the glycan recognition domains of the lectins were engaged, we performed a cross comparison analysis of different probes against individual lectins and used an array of excess unmodified HMOs as competitors. We observed that LTL was equally captured by probes 1 and 2, and excess 2FL abrogated the interaction with probe 2 but not with probe 1 (Fig. 2A and Fig. S2A, ESI†). MAL II was engaged by all probes, with probe 4 showing maximum cross-linking. However, only the interaction with probe 3 was competed by excess 3SL. Although cross-linking with SNA was observed with probe 2 and probe 4, only the interaction with probe 4 was reduced by 6SL. We also evaluated the ability of unmodified soluble HMOs (Lac, 2FL, 3SL, 6SL) to reduce photo-crosslinking of the probes (Fig. 2B and Fig. S2B, ESI†), and observed that the preferred glycan of

Fig. 1 HMO probes bind to purified plant lectins. (A) Chemical structures of HMO probes 1–4. (B) Plant lectins, such as LTL, MALII, and SNA are oligomeric proteins known to exhibit preferences in glycan binding. (C) Analysis of probe engagement with plant lectins. Photo-crosslinking was found to be UV- and dose-dependent.
each lectin is superior as a competitor molecule compared to other HMOs. We also note, however, that the interaction of probe 1 in the presence of excess Lac reduced photo-crosslinking with LTL. Overall, these experiments demonstrate that the known preferred glycan ligands are required for successful competition, and that, while probes can cross-react with non-binding lectins, incubation with an excess of the corresponding preferred glycan competes for specific interactions.

We next evaluated the ability of our probes to engage proteins in live RAW264.7 murine macrophages (Fig. 3A and Fig. S3A, ESI†), as the immunomodulatory activities of HMOs have previously been described in these cells.11,37 In addition to the previously reported immunomodulatory roles, RAW264.7 cells are widely used in immunology due to their phenotypic and functional stability.38 In brief, cells were treated with each probe, followed by exposure to UV light, harvesting, lysis, coupling of probe-modified proteins to TAMRA-azide and fluorescence visualization. We observed dose-dependent photo-crosslinking with probes 2–4 in cells, with protein targets engaged across a range of molecular weights. Photo-crosslinking was observed to be dependent on the presence of the probe and UV, and probe labeling was significantly blocked by excess soluble HMO (Fig. 3B and Fig. S3B, ESI†), suggesting the selective capture of proteins that recognize the cognate free glycan. We also observed that proteins across multiple cellular compartments, including those that are located within cells were engaged by the probes (Fig. S4A–C, ESI†).

We next sought to identify the protein targets of photactivatable HMOs using quantitative mass spectrometry-based proteomics. Briefly, following probe treatment and photo-crosslinking, we used CuAAC to append a biotin-azide enrichment handle to crosslinked proteins, which were then enriched with streptavidin beads and digested with trypsin. The resulting
peptides were subsequently labeled with isobaric tandem mass tags (TMT)\textsuperscript{39,40} to facilitate multiplexed, quantitative comparisons. We also included a control probe (5; see ESI\textsuperscript{†}) to account for possible background interactions dependent on the enrichment tag alone.\textsuperscript{19} Overall, we identified 512 proteins in this experiment (Appendix Table S1, ESI\textsuperscript{†}), observed excellent within replicate correlation and replicates clustered together via unbiased hierarchical clustering based on Pearson correlation (Fig. S5A, ESI\textsuperscript{†}). As expected, the median protein abundance (Fig. S5B, ESI\textsuperscript{†}) and number of enriched protein targets (defined as \textgreater 4-fold, \( p < 0.05 \) relative to control probe 5 across two biological replicates; Fig. S4C, ESI\textsuperscript{†}) mirrored the general gel fluorescence intensity for each probe. We identified between 100–350 protein targets for each probe for a total of 411 unique targets across all probes (Fig. 4A). Targets spanned diverse cellular compartments, including the cell surface and intracellular organelles (mitochondria, ER, lysosome, Fig. 4B), consistent with our fractionation and microscopy experiments (Fig. S4A–C, ESI\textsuperscript{†}).

As probe 3 engaged most of the proteins detected and its cognate free glycan (3SL) displays important immunomodulatory properties\textsuperscript{11,41,42} we pursued competition experiments to provide additional evidence that the identified targets bind 3SL. In these experiments, cells were pre-treated with excess 3SL prior to incubation with probe 3 plus 3SL and processed for proteomic analysis as described above. We observed excellent correlation between replicates (Fig. S6A, ESI\textsuperscript{†}), and as expected from gel labeling experiments, we observed a strong, dose-dependent reduction of global protein abundances with co-incubation of the 3SL competitor (Fig. S6B, ESI\textsuperscript{†}) and diverse competition profiles (Fig. S6C and D, ESI\textsuperscript{†}).

Overall, we identified 193 unique targets that were both enriched by 3 and competed with coincubation of excess free 3SL (\textgreater 2-fold, \( p < 0.05 \) at maximum competition; Appendix Table S2, ESI\textsuperscript{†}). Notably, this list encompasses known glycan-binding proteins including Lgals1 (galectin-1), Lgals3 (galectin-3), and Gpnmb. We also observed proteins that have not been reported to exhibit glycan-binding activities, such as Pon3 (paraoxanase 3) and Acp6 (lysophosphatidic acid phosphatase type 6). (Fig. 4C and D) Intriguingly, both proteins are thought to use small polar molecules (organophosphates, aryl esters and lactones for Pon3) or negatively charged molecules (lysophosphatidic acid for Acp6) as substrates.\textsuperscript{43–45} We selected galectin-1 and galectin-3 as targets for validation studies, which entailed over-expressing FLAG(DDK)-tagged galectin-1 or His-tagged galectin-3 in HEK293T cells (Fig. S7A–C, ESI\textsuperscript{†}). We observed robust photo-crosslinking of galectin-1 (Fig. 5A) or...
galectin-3 (Fig. 5B) by probe 3 in live cells, which was abrogated by excess 3SL. While galectins (expressed both intra- and extracellularly) are commonly known to bind galactose-terminated glycans, they are also known to tolerate (2,3)-sialic acid substitutions. Previous shotgun glycan array studies highlighted that galectins can also bind HMOs and may in some instances be receptors of HMOs. Together, these data pinpoint high confidence interactors of probe 3 and demonstrate the capture of known glycan binders as well as unexpected proteins with affinity for HMO probes.

**Conclusions**

We have demonstrated that fully-functionalized HMO probes can be used to capture and identify potential HMO protein targets in live cells. Our finding that 3SL exhibits maximal interactions with proteins found in murine macrophages is intriguing, given recent findings that it induces transcriptional activation to resolve inflammatory responses, however further investigations are needed to elucidate the underlying molecular mechanisms. Overall, these probes augment the existing toolkit to study of glycan–protein interactions in solution and in equilibrium, and we expect the datasets herein to facilitate our understanding HMO biological activity as well as other untapped free glycans.

**Methods**

Detailed information regarding the availability and sources of all reagents used as well as experimental procedures used are listed in the ESL. All raw mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD035729.

**Conflicts of interest**

There are no conflicts to declare.

**Acknowledgements**

This work was supported by the NIGMS award to MLH (R35GM142462). JW is supported by NIAID T32 AI007244. We are grateful to Lars Bode and Philip Gordts for helpful discussions.

**References**

1. A. Varki, *Essentials of glycobiology*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 3rd edn, 2017.
2. A. Varki, *Glycobiology*, 2017, 27, 3–49.
3. B. A. H. Smith and C. R. Bertozzi, *Nat. Rev. Drug Discovery*, 2021, 20, 217–243.
4. C. Reilly, T. J. Stewart, M. B. Renfrow and J. Novak, *Nat. Rev. Nephrol.*, 2019, 15, 346–366.
5. A. Darvill, C. Augur, C. Bergmann, R. W. Carlson, J. J. Cheong, S. Eberhard, M. G. Hahn, V. M. Lo, V. Marfa and B. Meyer, *et al.*, *Glycobiology*, 1992, 2, 181–198.
6. E. C. Carroll, L. Jin, A. Moli, N. Muñoz-Wolf, E. Oleszycka, H. B. T. Moran, S. Mansouri, C. P. McEntee, E. Lambe, E. M. Agger, P. Andersen, C. Cunningham, P. Hertzog, K. A. Fitzgerald, A. G. Bowie and E. C. Lavelle, *Immunity*, 2016, 44, 597–608.
7. C. S. Fermaintt, K. Sano, Z. Liu, N. Ishii, J. Seino, N. Dobbs, T. Suzuki, Y.-X. Fu, M. A. Lehrman, I. Matsuo and N. Yan, *Nat. Commun.*, 2019, 10, 2377.
8. L. Bode, *Glycobiology*, 2012, 22, 1147–1162.
9. L. Bode, *Early Hum. Dev.*, 2015, 91, 619–622.
10. F. Foata, N. Sprenger, F. Rochat and E. C. Lavell, *Immunity*, 2016, 44, 597–608.
11. A. Pessentheiner, N. Spann, C. Collins, B. Ramms, A. Chiang, Y. Wang, A. Quach, L. Booshehri, A. Hammond, C. Tognaccini, J. Latasiewicz, J. L. Witztum, H. Hoffman, N. Lewis, C. Glass, L. Bode and P. Gordts, *Atherosclerosis*, 2021, 331, e31.
12. C. P. Soda, P. Wipf, Y. Yamaguchi, W. B. Fulton, M. Kovler, D. F. Niño, Q. Zhou, E. Banfield, A. D. Werts, M. R. Ladd, R. H. Buck, K. C. Goehringer, T. Prindle, S. Wang, H. Jia, P. Lu and D. J. Hackam, *Pediatr. Res.*, 2021, 89, 91–101.
13. W. Zhang, J. He-Yang, W. Tu and X. Zhou, *Nutr. Metab.*, 2021, 18, 5.
14. D. L. Ackerman, R. S. Doster, J. H. Weitkamp, D. M. Aronoff, J. A. Gaddy and S. D. Townsend, *ACS Infect. Dis.*, 2017, 3, 595–605.
