Mapping glycan-mediated galectin-3 interactions by live cell proximity labeling

Eugene Joeha, Timothy O’Learya, Weichao Liab, Richard Hawkinsa,b, Jonathan R. Hunga, Christopher G. Parkerb, and Mia L. Huanga,b,1

aDepartment of Molecular Medicine, The Scripps Research Institute, Jupiter, FL 33458; and bDepartment of Chemistry, The Scripps Research Institute, Jupiter, FL 33458

Edited by Carolyn R. Bertozzi, Stanford University, Stanford, CA, and approved September 16, 2020 (received for review May 8, 2020)

Galectin-3 is a glycan-binding protein (GBP) that binds β-galactoside glycan structures to orchestrate a variety of important biological events, including the activation of hepatic stellate cells and regulation of immune responses. While the requisite glycan epitopes needed to bind galectin-3 have long been elucidated, the cellular glycoproteins that bear these glycan signatures remain unknown. Given the importance of the three-dimensional (3D) arrangement of glycans in dictating GBP interactions, strategies that allow the identification of GBP receptors in live cells, where the native glycan presentation and glycoprotein expression are preserved, have significant advantages over static and artificial systems. Here we describe the integration of a proximity labeling method and quantitative mass spectrometry to map the glycan and glycoprotein interactors for galectin-3 in live human hepatic stellate cells and peripheral blood mononuclear cells. Understanding the identity of the glycoproteins and defining the structures of the glycans will empower efforts to design and develop selective therapeutics to mitigate galectin-3-mediated biological events.

Significance

Because of the weak interactions between individual glycan-binding proteins (GBPs), such as galectin-3, and glycans, strategies that enable the direct interrogation of these interactions in living cells remain limited. Thus, the glycan and glycoprotein ligands that are physiologically relevant for galectin-3 binding are insufficiently described. Here we used a proximity labeling approach that catalytically tags interactors for galectin-3 and identified its pertinent glycan and glycoprotein counter-receptors in live cells. This study demonstrates that proximity labeling is a powerful tool for mapping GBP complexes in living cells, and when coupled with chemical inhibitors, it can discriminate between protein–protein and protein–glycan interactions.

Author contributions: E.J., C.G.P., and M.L.H. designed research; E.J., T.O., W.L., R.H., J.R.H., and M.L.H. performed research; E.J., T.O., W.L., C.G.P., and M.L.H. analyzed data; and E.J., C.G.P., and M.L.H. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

*To whom correspondence may be addressed. Email: miahuang@scripps.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2009206117/-/DCSupplemental.

First published October 16, 2020.

www.pnas.org/cgi/doi/10.1073/pnas.2009206117

PNAS | November 3, 2020 | vol. 117 | no. 44 | 27329–27338
covalently tag nearby electron-rich residues. The biotinylated proteins can then be enriched and identified using quantitative mass spectrometry (MS)-based proteomics. Because the (glyco) proteins adjacent to the APEX2 fusion protein are preferentially biotinylated, the resulting MS data provide a readout of its immediate environment.

We reasoned that proximity labeling could offer significant advantages over other approaches to determining GBP–glycan interactions, including the opportunity to perform the labeling in live cells and the ability to tag weakly bound glycan-mediated interactors, as the covalent biotinylation reaction ensures that the enrichment step no longer relies on weak GBP–glycan interactions alone. When coupled with inhibitors, the proximity labeling strategy can also distinguish between glycan-mediated and non–glycan-mediated interactors. Integration of this approach with quantitative MS-based proteomics would also expedite the assignment of the interacting proteins and provide calculable measures to distinguish interactors from nonspecific binders.

Here we report that the use of an APEX2 and galectin-3 fusion protein (PX-Gal3) provides a sensitive and comprehensive approach to mapping the proteome-wide glycan-mediated galectin-3 interactome in live human HSCs and peripheral blood mononuclear cells (PBMCs). We found that the exogenous incubation of cells with PX-Gal3 in HSCs leads to glycan-dependent interactions, whereas its cellular overexpression does not. We further validated the interactions between galectin-3 and candidate proteins in vitro and discovered that some proteins are direct glycan-mediated receptors. Using MS-based glycomics, we also examined the glycomes of HSC surfaces, PX-Gal3 tagged glycoproteins, and an individual glycoprotein receptor for galectin-3. Our results highlight the utility of the in situ proximity labeling approach in discovering physiologically relevant GBP interactors in living cells.

Results

Proximity Tagging to Identify Galectin-3 Interactors in HSCs. We genetically fused the peroxidase APEX2 enzyme (20, 21) to either full-length galectin-3 (PX-Gal3) or a mutant galectin-3 lacking the N-terminal domain (PX-Gal3Δ116) (Fig. 2A). This mutant was produced to determine whether the proximity labeling protocol could be used to confirm the role of the N-terminal domain in facilitating homo-oligomerizing interactions to control extracellular binding activities (22, 23). We chose APEX2 over other peroxidase enzymes because of its monomeric nature, small size (28 kDa), and improved activities in various cellular compartments (20), especially for proteomic applications (24). We fused APEX2 to the N terminus of galectin-3, as previous work has shown that such fusion proteins remain active toward binding glycans (25). Both fusion proteins were generated efficiently (SI Appendix, Fig. S1 A–C) in recombinant PX-Gal3 and retained their capacity to bind glycans in an enzyme-linked immunosorbent assay (ELISA) with immobilized asialofetuin (SI Appendix, Fig. S2). Coincubation with a soluble competitor, lactose, efficiently blocked binding (Fig. 2B), indicating that interactions were glycan-dependent. Apparent binding affinity constants determined for PX-Gal3 (∼100 nM) vs. PX-Gal3Δ116 (∼140 nM) were similar, consistent with previous reports of galectin-3 fusion proteins (25).

To examine whether the proximity labeling protocol could identify interactors with galectin-3, we incubated the recombinant fusion proteins with live adherent HSCs. After washing and the stepwise application of biotin-phenol and H2O2, biotin-tagged interactors were visualized using a fluorophore-conjugated streptavidin probe and fluorescence microscopy (Fig. 2C). We observed significant fluorescence in cells incubated with PX-Gal3 compared with negative controls and PX-Gal3Δ116 (Fig. 2D), with signals detectable as early as after 5 min of protein incubation, which did not increase significantly even after up to 2 h (SI Appendix, Fig. S3A). We also observed a dose-dependent increase in fluorescence with increasing amounts (25–100 nM) of PX-Gal3 (SI Appendix, Fig. S3B). In contrast to the plate-based ELISA findings, PX-Gal3Δ116 did not elicit a significant signal compared with PX-Gal3. We attribute this observation to the differing glycan density and presentation found in live cells, establishing an important role for the N-terminal oligomerization domain not directly apparent using plate-based methods. This protocol enabled the sensitive detection of interactors with nanomolar concentrations of PX-Gal3, whereas querying for galectin-3 binding to HSCs in simple binding experiments failed to achieve a significant signal even at a concentration of 10 μM and a prolonged incubation period (SI Appendix, Fig. S3C). While fluorescence signals were mostly observed at the cell surface, some intracellular labeling was also observed with increased PX-Gal3 incubation times (SI Appendix, Fig. S4A) and concentrations (SI Appendix, Fig. S4B), presumably due to known endocytic mechanisms that occur upon galectin-3 binding (13, 26).

Proximity labeling conditions in which PX-Gal3 was coincubated with lactose (SI Appendix, Fig. S4A) or where cell surface glycosylation was inhibited, either by tunicamycin (SI Appendix, Fig. S4B) or by PNGase F (SI Appendix, Fig. S4C), showed reduced fluorescence. Furthermore, a PX-Gal3 mutant that included three point mutations (R144S, R186A, and G182A) that abolish glycan-binding activity (27) failed to show significant fluorescence (SI Appendix, Fig. S5). These results indicate that glycan-mediated interactions are highly represented between PX-Gal3 and HSCs. We also compared labeling of live intact cells vs. cellular lysates and found that only a limited subset of potential interactors was captured and substantially blocked with excess lactose (Fig. 2E and SI Appendix, Fig. S6). Taken together, these observations indicate that the proximity labeling strategy can probe for glycan-dependent interactors in live cells with high sensitivity, and that native glycan presentation is important for the selective capture of interactions.

Identification of Biotinylated Interactors with Quantitative MS-Based Proteomics. To identify the putative galectin-3 interactors, we subjected HSCs to proximity labeling, followed by cell lysis,
no proteins were enriched by PX-Gal3

number of enriched proteins using PX-Gal3 (Fig. 3

S7 (99%) were competed with lactose (TMT ratio

analog that displays potent galectin-3 inhibition (Fig. 3

C-terminal domain of galectin-3. PX-Gal3

Δ

B

Δ

B

Δ

116, were constructed. Both proteins include an N-terminal His-tag sequence, followed by the APEX2 enzyme, and either full-length galectin-3 or the C-terminal domain of galectin-3. PX-Gal3 Δ116 lacks the N-terminal domain of galectin-3, which was previously implicated for its homo-oligomerization on binding of glycans at the cell surface. (b) Using an ELISA with asialofetuin as a model glycan-bearing ligand, the fusion constructs retained similar glycan-binding activities, as determined by EC50 values. (c) Application of the proximal labeling method to live adherent HSCs. The recombinant fusion proteins were first incubated with HSCs (30 min, 37 °C). After washing to remove unbound proteins, biotin phenol (yellow circle with a B; 500 μM, 30 min, 37 °C) was added, followed by H2O2 (1 mM, 1 min, RT). Biotinylated interactors (purple) were subsequently probed using Cy5-labeled streptavidin. Conditions were optimized according to reported radical-mediated biotinylation procedures (20, 21). (d) Fluorescence microscopy images of biotin-tagged (purple) HSCs showing that PX-Gal3 (100 nM; 30 min) generates significant labeling over negative controls in which a component (e.g., protein or biotin phenol and H2O2) of the labeling protocol is omitted. Coincubation with lactose (Lac; 100 mM) causes a loss of signal, suggesting that the majority of labeled interactions are glycan-dependent. PX-Gal3 Δ116 fails to significantly label interactors. (e) Western blotting of (10 μg lysate per lane) biotinylated proteins produced from the proximity labeling method applied to intact live cells vs. cell lysates showing that only a subset of interactors is captured (SI Appendix, Fig. S6).

Among the PX-Gal3 targets identified by the proximity labeling method in LX-2 HSCs (SI Appendix, Table S1) are vasorin (VASN), basigin (BSG/CD147), neuroplastin (NPTN), members of the tetraspanin family (CD9, CD81, and TSPAN6), members of the ephrin family (EPHB1 and EPHB2), members of the solute carrier family of transporters (SLC35F2, SLC23A2, and SLC7A5), and members of the integrin family (ITGAI and ITGB1). Although many of the targets are assigned to plasma membrane or cell surface locations, several are also found intracellularly, including members of the human ribonucleoprotein (HNRNP) complex (33), which are found in the nucleus. These results indicate that despite their exogenous application, the fusion proteins traverse pathways that are expected of the dynamic nature of galectin-3, first interacting with cell surface proteins and then eventually traversing the cell membrane to associate with intracellular components (34).

We derived effective galectin-3 binding affinity constants (EC50) for selected proteins (BSG, CD47, ITGAI, JAM3, EPHB2, and VASN) using the proximity labeling protocol and TMT quantitation (SI Appendix, Fig. S7F). Such binding constants are reflective of the native cellular glycan and protein expression levels. While it is difficult to extrapolate these constants as a comparative measure of binding of galectin-3 against individual proteins, it is gratifying to see that the proximity labeling method can achieve robust nanomolar binding against these proteins in cells. We used integrated pathway analysis to derive putative binding relationships and observed that BSG lies at a nexus to other proteins enriched by PX-Gal3, such as ITGAI, SLC7A5, JAM3, and SLC23A2 (SI Appendix, Fig. S7G).

Considering that we identified intracellular protein targets with exogenously administered PX-Gal3, we set out to evaluate whether proximity labeling could also identify interactors for streptavidin bead enrichment, and on-bead trypsinization (Fig. 3A). We then subjected the resulting peptides to tandem mass tag (TMT) labeling, which chemically tags the N termini and free amines of the resulting peptides from each condition with a unique isobaric tag that can be distinguished at the MS3 level. TMT analysis of multiplexed samples in a single run, further ensuring the accuracy of the comparison.

We identified a total of 248 proteins across two replicates (Fig. 3B, Table 1, and SI Appendix, Table S1). These proteins were consistently detected with three or more unique peptides and were highly enriched (TMT ratio ≥10) by PX-Gal3 (100 nM) over the negative control, cells that were not treated with any fusion protein but were similarly treated with biotin phenol and H2O2 (Fig. 3B). We observed a dose-dependent increase in the number of enriched proteins using PX-Gal3 (Fig. 3C), whereas no proteins were enriched by PX-Gal3 Δ2116 (SI Appendix, Fig. S7A) or by the triple-point mutant protein (SI Appendix, Fig. S7B). A majority (71 to 90%) of enriched proteins in each dose of PX-Gal3 was previously reported to be N- or O-glycosylated (Fig. 3C) (30). Among the 248 proteins, 209 (84%) were defined as statistically significant (P < 0.05; Fig. 3D); among these, 208 (99%) were corrected with lactose (TMT ratio ≥4; Fig. 3E). Proteins that were enriched generally correlated with those that were competed with lactose (Fig. 3F) and TD139 (31), a lactose analog that displays potent galectin-3 inhibition (SI Appendix, Fig. S7C). The proteins enriched by PX-Gal3 span a wide range of abundance, as determined using peptide spectral matches from unenriched proteomics in LX-2 HSCs (SI Appendix, Fig. S7D and Table S2) and reported iBAQ (intensity-based absolute quantification) values (SI Appendix, Fig. S7E) (32).
galectin-3 when it is expressed intracellularly. To do so, we overexpressed PX-Gal3 in HSCs by transient transfection (SI Appendix, Fig. S8 A and B). Following transfection, some cells were further treated with tunicamycin, an N-glycosylation inhibitor, to assess whether N-linked glycosylation plays a role (lactose is impermeable to cells). A total of 498 proteins (filtered by three or more unique peptides and with a TMT ratio (PX-Gal3/Neg) ≥ 10) were found to be enriched across two replicates, defined as proteins with three or more unique peptides and with a TMT ratio (PX-Gal3/Neg) ≥ 10. Neg indicates conditions in which the cells were not incubated with PX-Gal3 but were still treated with biotin phenol and H2O2. (C) Analysis of proteins enriched by PX-Gal3 by dosage, glycosylation status (assigned by UniProt), and competition with lactose (100 nM). (D) Statistically significant (P < 0.05) and enriched proteins found in cells treated with PX-Gal3 (100 nM, 30 min). (E) Proteins that were statistically significant (P < 0.05) and enriched proteins found in cells treated with PX-Gal3 (100 nM, 30 min). (F) Proteins that were statistically significant (P < 0.05) and competed (TMT ratio of PX-Gal3/PX-Gal3+Lac ≥ 4) on coincubation with exogenous lactose (100 mM). (F) There is high linear correlation between proteins that were enriched by and those that were competed with lactose. (G) A total of 498 proteins were found to be enriched across two replicates when PX-Gal3 was transiently overexpressed in HSCs. (H) A total of 431 proteins were found to be significantly enriched (P < 0.05). (I) There were 122 overlapping proteins between the exogenous and transfection protocols (SI Appendix, Table S1).

Validation of Interactions between Galectin-3 and Identified Proteins. As the proximity labeling strategy only tags proteins that occur within the vicinity (<20 nm) of galectin-3, and that these might not necessarily represent direct receptors, we next examined whether selected proteins identified by the proteomic experiments are direct receptors for galectin-3. We first confirmed that BSG, CD9, EPHB1, CD47, and VASN are prominently expressed in LX-2 cells (SI Appendix, Fig. S9). Then, using immobilized human recombinant proteins (SI Appendix, Table S1) of BSG, CD9, CD47, CD81, EPHB1, NPTN, and VASN, we found that galectin-3 binds in a dose-dependent manner (Fig. 4A and SI Appendix, Table S2). Apparent binding affinities, with EC50 values ranging from 0.7 μM (CD47), were observed (SI Appendix, Table S2). Notably, we found that the majority of the proteins were competed with the presence of either lactose or TD139, CD81 binding was not blocked by either competitor, and CD9 interactions could only be competed with TD139 (Fig. 4B). Human CD81 is not known to be glycosylated, suggesting that its binding to galectin-3 might be mediated by secondary protein–protein interactions. Human CD9 (UniProt P21926) is predicted to have two N-linked glycosylation sites, implicating potential glycan-mediated interactions with galectin-3, but these interactions may only partially contribute to galectin-3 binding, as competition with TD139 was still observed. Taken together, these results indicate that BSG, CD9, CD47, CD81,
complex galactose-terminating N-glycans (with some tolerance and glycosaminoglycans (36) can serve as ligands for galectin-3, groups have reported that O-linked glycans from mucins (35) structures of the glycans found in LX-2 HSCs. While several interactions (Figs. 2 and 3), we next sought to identify the importance of cell surface glycosylation in mediating galectin-3 extracellular region or secreted; Nuc, nucleus; Endo, endosome; Lyso, lysosome.

EHF-B1, VASN, and NPTN are potential receptors for galectin-3 in cells.

**MS-Based Characterization of Glycomes from LX-2 Cell Surfaces, PX-Gal3–Enriched Interactors, and Endogenous BSG.** Given the importance of cell surface glycosylation in mediating galectin-3 interactions (Figs. 2 and 3), we next sought to identify the structures of the glycans found in LX-2 HSCs. While several groups have reported that O-linked glycans from mucins (35) and glycosaminoglycans (36) can serve as ligands for galectin-3, complex galactose-terminating N-glycans (with some tolerance of sialylation) have been definitively established as the prominent high-affinity binding galectin-3 epitopes in cells (37, 38). Thus, we next focused our investigation on profiling N-glycans, with particular attention on assessing the abundance of N-glycans bearing the high-affinity terminal galactose monosaccharides (39). We further evaluated whether certain N-glycan motifs are enriched by PX-Gal3 and whether they are present on the identified glycoprotein receptors for galectin-3 (Figs. 3 and 4).

We harvested N-glycans from intact LX-2 cells by brief treatment with trypsin and PNGase F. The free N-glycans were then subjected to reduction and permethylation before MS analysis. Using the extracted ion intensities acquired during the full MS scan, relative quantitation of the most abundant N-glycans was determined (40). The most abundant N-glycans found on LX-2 cell surfaces consisted mostly of complex and oligomannose, with the most abundant being the di-sialylated FA2G2, composing ∼20% of the total population (Fig. 5 and SI Appendix, Figs. S10 and S11). LX-2 cells stained positively with *Sambucus nigra* and *Maackia amurensis* lectins, indicating the presence of α(2,3)- and α(2,6)-linked terminal Neu5Ac sialic acids, respectively (SI Appendix, Fig. S12). Oligomannose N-glycans (M7, M6, and M8) were the next most abundant type, representing a cumulative total of ∼27% of the cell surface N-glycome. Terminal mono- and di-galactosylated N-glycans, FA2G1 and FA2G2, composed only ∼3% and ∼1% of the population, respectively.

We also identified N-glycans from the resulting mixture of peptides and glycopeptides captured from proximity labeling with PX-Gal3 (Fig. 24) by harvesting them with PNGase F. The most abundant N-glycans from these samples were overall quite similar to those found on LX-2 cell surfaces (Fig. 5 and SI Appendix, Fig. S12). FA2G1 and FA2G2 composed ∼5% and ∼1% of the total N-glycome, respectively. We observed a modest increase in the amount of terminal galactosylated N-glycans, with the di-galactosylated FA3G2 showing an ∼57% increase compared with the cell surface (SI Appendix, Fig. S13).

Given the central role of BSG as a glycan-dependent binding receptor for galectin-3 (Fig. 4) and its ability to bind other enriched proteins (SI Appendix, Fig. S7G), we sought to determine its native N-glycosylation pattern. To this end, we immunoprecipitated

### Table 1. Abbreviated list of galectin-3 interactors determined by proximity labeling and quantitative MS

| Gene name | Protein name | UniProt ID | Glycosylation state | Subcellular location |
|-----------|--------------|------------|---------------------|----------------------|
| BSG       | BSG          | P35613     | 3 N-linked sites    | PM; ERS              |
| NPTN      | Neuroplastin | Q9Y639     | 6 N-linked sites    | PM                   |
| JAM3      | Junctional adhesion molecule C | Q9BX67 | 2 N-linked sites | PM, ERS |
| SLC25F2   | Solute carrier family 35 member F2 | Q8XKU6 | Not known | PM  |
| SLC7A5    | Large neutral amino acids transporter small subunit 1 | Q01650 | Not known | PM; Lyso; ERS |
| SLC23A2   | Solute carrier family 23 member 2 | Q9UGH3 | 2 N-linked sites | PM |
| ITGA1     | Integrin-a-1 | P56199     | 26 N-linked sites   | PM; ERS              |
| ITGB1     | Integrin-b-1 | P05556     | 4 N-linked sites    | PM; Endo; ERS        |
| CD44      | CD44 antigen | P16070     | 9 N-linked sites    | PM; ERS              |
| CD47      | Leukocyte surface antigen CD47 | Q08722 | 6 N-linked sites | PM; ERS |
| CD81      | CD81 antigen | P60033     | Not known           | PM; ERS              |
| CD9       | CD9 antigen  | P21926     | 2 N-linked sites    | PM; ERS              |
| VASP      | Vasin        | P63172     | 1 N-linked site     | PM; Lyso; ERS        |
| VASP      | Vasin        | P52799     | 2 N-linked sites    | PM; ERS              |
| VASN      | Vasorin      | P63172     | 1 N-linked site     | PM; Lyso; ERS        |
| HNRNPH3   | Heterogeneous nuclear ribonucleoprotein H3 | P31942 | Not known | Nuc |
| HNRNPD    | Heterogeneous nuclear ribonucleoprotein D0 | Q14103 | Not known | Nuc |
| HNRNPF    | Heterogeneous nuclear ribonucleoprotein F | P52597 | Not known | Nuc |

An expanded list is available in SI Appendix, Table S1. The glycosylation states are assigned according to UniProt (30). PM, plasma membrane; ERS, extracellular region or secreted; Nuc, nucleus; Endo, endosome; Lyso, lysosome.
proteomics identified 248 and 333 enriched proteins, respectively (Fig. 6). Integration of this method with quantitative MS-based proteomics (9–14), and our dataset extends the number of previously known interactors for galectin-3. These enriched proteins rely exclusively on glycan-dependent interactions for catalytic tagging. In contrast to in vitro experiments and tagging procedures on lysed cells, we showed that the N-terminal oligomerization domain (23) is critical for productive cell surface interactions (Fig. 2D), and despite many other possible interactors, only a subset form favorable interactions with galectin-3 (Fig. 2E).

While other labeling strategies use horseradish peroxidase fusion proteins (41) that require oxidative conditions, our method uses the APEX2 enzyme for covalent tagging. APEX2 is active in both oxidative extracellular conditions and the reducing environment of the intracellular milieu (42), permitting the identification of 498 interactors for PX-Gal3 when it is intracellularly overexpressed (Fig. 3G).

Consistent with previous reports, we also found that intracellular interactions with galectin-3 are not glycan-dependent (34). Some of the 122 proteins commonly enriched between the exogenously and the overexpressed conditions are transmembrane proteins (e.g., CD47, CD81) with the opportunity to be tagged from either side of the membrane. Others have also been found at the cell surface and intracellularly (e.g., GPC4, SLC3A2). These two points highlight the potential of galectin-3 to act via different mechanisms depending on its entry, providing a rationale for why the same proteins were tagged in a glycan-dependent manner exogenously and also in a non–glycan-dependent manner in the overexpressed model. However, we cannot exclude the possibility that tunicamycin treatment adversely affects glycoprotein expression and structural conformation.

**Proximity Tagging to Identify Galectin-3 Interactors in PBMCs.** Encouraged by the results with LX-2 HSCs, we sought to evaluate whether the proximity tagging approach could be extended to other cell types. Given the importance of galectin-3 in the immune system, we chose to identify interactors in PBMCs, which are composed of mixtures of myeloid and lymphoid cells. We identified a total of 333 proteins that were consistently and significantly enriched across three replicates (Fig. 6A, Table 2, and SI Appendix, Table S3). We observed that 332 proteins (99.7%) interacted in a glycan-dependent manner (Fig. 6B), and that there was similarly a general correlation between enriched proteins and proteins competed with lactose (Fig. 6C). Among these 332 proteins, 39 overlapped with proteins identified in HSCs, and the remainder were expressed exclusively in PBMCs (Fig. 6D).

**Discussion**

Our overall goal was to determine whether a proximity labeling approach would be suitable for mapping the glycan-mediated interactors for galectin-3 in live cells (Fig. 1). Here we have demonstrated that a 1-min live labeling step with an exogenously applied fusion protein of APEX2 and galectin-3 covalently tagged protein interactors in HSCs (Figs. 2 and 3) and in PBMCs (Fig. 6). Integration of this method with quantitative MS-based proteomics identified 248 and 333 enriched proteins, respectively (Figs. 3 and 6). Before this study, only 100 to 185 proteins were found in complex with galectin-3 using standard affinity techniques (9–14), and our dataset extends the number of previously known interactors for galectin-3. These enriched proteins rely exclusively on glycan-dependent interactions for catalytic tagging. In contrast to in vitro experiments and tagging procedures on lysed cells, we showed that the N-terminal oligomerization domain (23) is critical for productive cell surface interactions (Fig. 2D), and despite many other possible interactors, only a subset form favorable interactions with galectin-3 (Fig. 2E).
Along with galectin-3, many of its 248 putative interactors are known to be important in HSC activation. For example, BSG and EPHB2 expression have been correlated to HSC activation (43, 44), and knockout of BSG has been shown to attenuate fibrosis (45). However, additional investigation is needed to confirm that the interactions between galectin-3 and these proteins are relevant in HSC activation. We selected seven proteins to confirm direct interactions with galectin-3 and found that recombinant BSG, CD9, CD47, CD81, EPHB1, NPTN, and VASN bound galectin-3, and that binding to CD9 and CD81 were not glycan-mediated in vitro (Fig. 4). It must be noted, however, that these recombinant proteins might not necessarily reflect the native protein sequence or glycosylation patterns however, that these recombinant proteins might not necessarily reflect the native protein sequence or glycosylation patterns. The recombinant BSG glycoprotein used for the in vitro binding experiments has only two sites for N-glycosylation (18) required for the galectin-3-mediated function, although BSG has not previously been shown to display terminal galactosides (51, 52), such differences may reflect the natural abundance of glycan biosynthetic enzymes and sugar nucleotide precursors between different tissues. Human BSG naturally occurs in four natural isoforms (53), and our proteomics data suggest that either BSG-1 or BSG-2 is expressed in LX-2s. Endogenous BSG-2 differs from the canonical BSG-1 by missing regions 24 to 139 (SI Appendix, Fig. S17) and similarly bears three N-linked glycosylation sites. The recombinant BSG glycoprotein used for the in vitro binding experiments has only two sites for N-glycosylation (SI Appendix, Fig. S17), suggesting that tighter interactions may be achieved by the endogenous glycoprotein.

Given the importance of galectin-3 in regulating immunity, we also applied the PX-Gal3 fusion protein exogenously to PBMCs and identified glycan-mediated interactors. Of note, these proteins are highly important in both the innate and adaptive immune responses. For example, CD3G and CD3E are two of four subunits of the CD3 complex present on T cell surfaces that mediate T cell activation (54, 55). TLR-2 is one of a family of TLRs that are key for the innate immune response to various microbial agents, responsible for recognizing pathogen-associated molecular patterns (56–58). TLR-2 has also been associated with formation of a complex with CD36, another glycoprotein that we captured with high enrichment and with successful competition (59).

We have demonstrated that proximity labeling is a powerful approach to capturing glycan-mediated binding events in live cells. We show that this method is generally applicable in a cell line and freshly isolated primary immune cells. We expect that it could also be extended to studies in other biological samples, such as tissues, and could also be used to track the spatiotemporal dynamics of galectin-3 interactions (60) or to visualize interactions using electron microscopy (21). Empowered by this approach, we now have generated a priority list of proteins that will be the subject of future work to further reveal the “professional glycoprotein ligands” (18) required for the galectin-3-mediated functional activation of HSCs and regulation of immunity.

Along with galectin-3, many of its 248 putative interactors are known to be important in HSC activation. For example, BSG and EPHB2 expression have been correlated to HSC activation (43, 44), and knockout of BSG has been shown to attenuate fibrosis (45). However, additional investigation is needed to confirm that the interactions between galectin-3 and these proteins are relevant in HSC activation. We selected seven proteins to confirm direct interactions with galectin-3 and found that recombinant BSG, CD9, CD47, CD81, EPHB1, NPTN, and VASN bound galectin-3, and that binding to CD9 and CD81 were not glycan-mediated in vitro (Fig. 4). It must be noted, however, that these recombinant proteins might not necessarily reflect the native protein sequence or glycosylation patterns however, that these recombinant proteins might not necessarily reflect the native protein sequence or glycosylation patterns. The recombinant BSG glycoprotein used for the in vitro binding experiments has only two sites for N-glycosylation (18) required for the galectin-3-mediated function, although BSG has not previously been shown to display terminal galactosides (51, 52), such differences may reflect the natural abundance of glycan biosynthetic enzymes and sugar nucleotide precursors between different tissues. Human BSG naturally occurs in four natural isoforms (53), and our proteomics data suggest that either BSG-1 or BSG-2 is expressed in LX-2s. Endogenous BSG-2 differs from the canonical BSG-1 by missing regions 24 to 139 (SI Appendix, Fig. S17) and similarly bears three N-linked glycosylation sites. The recombinant BSG glycoprotein used for the in vitro binding experiments has only two sites for N-glycosylation (SI Appendix, Fig. S17), suggesting that tighter interactions may be achieved by the endogenous glycoprotein.

Given the importance of galectin-3 in regulating immunity, we also applied the PX-Gal3 fusion protein exogenously to PBMCs and identified glycan-mediated interactors. Of note, these proteins are highly important in both the innate and adaptive immune responses. For example, CD3G and CD3E are two of four subunits of the CD3 complex present on T cell surfaces that mediate T cell activation (54, 55). TLR-2 is one of a family of TLRs that are key for the innate immune response to various microbial agents, responsible for recognizing pathogen-associated molecular patterns (56–58). TLR-2 has also been associated with formation of a complex with CD36, another glycoprotein that we captured with high enrichment and with successful competition (59).

We have demonstrated that proximity labeling is a powerful approach to capturing glycan-mediated binding events in live cells. We show that this method is generally applicable in a cell line and freshly isolated primary immune cells. We expect that it could also be extended to studies in other biological samples, such as tissues, and could also be used to track the spatiotemporal dynamics of galectin-3 interactions (60) or to visualize interactions using electron microscopy (21). Empowered by this approach, we now have generated a priority list of proteins that will be the subject of future work to further reveal the “professional glycoprotein ligands” (18) required for the galectin-3-mediated functional activation of HSCs and regulation of immunity.

Table 2. Abbreviated list of galectin-3 interactors in PBMCs determined by proximity labeling and quantitative MS

| Gene name       | Protein name                          | UniProt ID | Glycosylation state | Subcellular location |
|-----------------|---------------------------------------|------------|---------------------|----------------------|
| ESAM            | Endothelial cell-selective adhesion molecule | Q96AP7    | 4 N-linked sites    | PM                   |
| NPTN            | Neuroplastin                          | Q9Y639     | 6 N-linked sites    | PM                   |
| PBP             | Platelet basic protein                 | P02775     | Not known           | ERS                  |
| GP9             | Platelet glycoprotein IX              | P14770     | 1 N-linked site     | PM                   |
| CD36            | Platelet glycoprotein 4               | P16671     | 10 N-linked sites   | ERS; PM              |
| SLC3A2          | 4F2 cell-surface antigen heavy chain   | P08195     | 4 N-linked sites    | PM; Lyso             |
| FBLN1           | Fibulin-1                             | P23142     | 3 N-linked sites    | ERS                  |
| CCL5            | C-C motif chemokine 5                 | P13501     | 2 O-linked sites    | ERS                  |
| CD44            | CD44 antigen                          | P16070     | 9 N-linked sites    | PM; ERS              |
| CD47            | Leukocyte surface antigen CD47        | Q08722     | 6 N-linked sites    | PM; ERS              |
| TLR2            | Toll-like receptor 2                  | O60603     | 4 N-linked sites    | PM                   |
| CD9             | CD9 antigen                           | P21926     | 2 N-linked sites    | PM; ERS              |
| CD3G            | T-cell surface glycoprotein CD3 gamma chain | P09693 | 2 N-linked sites | PM                   |
| FBLN1           | FBLN1                                 | P23142     | 3 N-linked sites    | ERS                  |
| CD3E            | T cell surface glycoprotein CD3 epsilon chain | P07766 | Not known           | PM                   |
| SLC2A3          | Solute carrier family 2, facilitated glucose transporter member 3 | P11169 | 1 N-linked site | PM; ERS              |
| MS4A1           | B-lymphocyte antigen CD20             | P11836     | Not known           | PM; ERS              |
| GP1B A          | Platelet glycoprotein Ib alpha chain  | P07359     | 2 N-linked sites and 1 O-linked site | PM; ERS |
| TLT-1           | Trem-like transcript 1 protein        | Q86YWS     | Not known           | PM                   |

An expanded list is available in SI Appendix, Table S3. The glycosylation states are assigned according to UniProt (30). PM, plasma membrane; ERS, extracellular region or secreted; Nuc, nucleus; Endo, endosome; Lyso, lysosome.
Conclusions

GBP-glycan interactions choreograph many biological events, yet they are often interrogated in nonnative environments without the identification of the glycan-bearing glycoprotein receptors. Herein, we used a proximity labeling method consisting of fusion proteins of galectin-3 and the APEX2 enzyme to map glycan-mediated galectin-3 interactions in live cells. This method is sensitive, enabling detection of galectin-3 binding at nanomolar concentrations, and we have produced an inventory of proteins that occur in complex with galectin-3 and confirmed some of these proteins as direct glycan-mediated glycoprotein receptors. Consequently, our results provide higher-resolution insight into the glycan and protein determinants that govern galectin-3 interactions in HSCs and with individual glycoproteins. Critically, although we have mapped the galectin-3 interactome with live cells, the functional consequences imparted by these discrete interactions remain unknown. Informed by our results, we are now poised to formulate new hypotheses that define the contributions of individual proteins and glycosylation sites. Importantly, we expect that the proximity labeling approach will be generally applicable to other GBPcs and glycoproteins, and that it will be a powerful tool to advance our understanding of glycan-mediated cell biology.

Materials and Methods

Growth and Maintenance of HSCs. LX-2 human HSCs (61) (Millipore) were cultured in DMEM supplemented with 10% fetal bovine serum at 37 °C under 5% CO2. For fluorescence microscopy imaging experiments, cells were grown on plastic 24-well culture plates. To improve the adherence of LX-2 cells, dishes were precoated with poly-o-lysine and washed with Dulbecco’s phosphate-buffered saline (DPBS) before cell plating.

In Situ Proximity Tagging in Live LX-2 Cells and Imaging. The LX-2 cells were plated ( precoated with poly-o-lysine) at a density of 1.5 × 105 cells per well on a plastic 24-well plate overnight at 37 °C in 5% CO2. The next day, the cells were gently washed with DPBS and then incubated with DMEM supplemented with the fusion protein of interest for defined periods (usually 30 min, unless indicated otherwise) at 37 °C in 5% CO2. After two washes with DPBS to remove unbound proteins, labeling was initiated by replacing the medium with prewarmed 10% FBS/DMEM containing 500 μM biotin-phenol. This solution was further incubated with the cells at 37 °C in 5% CO2. After washing twice with DPBS, the cells were gently washed with DPBS and then incubated with DMEM supplemented with a cold mixture of MeOH:CHCl3 (4:1) added to the cells, washed by 1 ml of cold DPBS. The resulting mixture was vortexed and centrifuged (5,000 rpm, 10 min, 4 °C). The organic and aqueous layers were aspirated, and the remaining protein disk was further washed via sonication in 2 ml of cold MeOH:CHCl3 (4:1) and pelleted by centrifugation (5,000 rpm, 10 min, 4 °C). The protein pellet was aspirated and dissolved in 500 μL of freshly prepared proteomics-grade urea (6 M in DPBS) with 150 μL of 10% SDS by sonication. Disulfide bonds were reduced by adding 50 μL of freshly prepared 1:1 solution of TCEP (200 mM in DPBS) and K2CO3 (600 mM in DPBS) for 30 min at 37 °C on a shaker. Free thiols were alkylated by adding 70 μL of freshly prepared iodoacetamide (400 mM in DPBS) at RT in the dark. To each solution, 130 μL of 10% SDS in DPBS was added, and each sample was diluted with DPBS (5.5 ml) and incubated with prepeptidilated streptavidin-agarose beads (100 μL of 50% slurry; Pierce) for 1.5 h at RT while rotating. The streptavidin beads were pelleted by centrifugation (2,000 rpm, 2 min, 4 °C) and sequentially washed with 0.2% SDS in DPBS (1 × 5 ml), DPBS (2 × 5 ml), and 200 mM EPPS [4-(2-hydroxyethyl)-1-piperazinepropanesulfonic acid], pH 8.4 (1 × 5 ml). The beads were transferred into LoBind microcentrifuge tubes (Eppendorf; 022431081), and bound proteins were digested for ~14 h at 37 °C in 200 μL of trypsin premix (~200 μg)+containing sequence-grade trypsin (2 μg; Promega), urea (0.5 M), and a final concentration of 1 mM, and the plate was gently agitated for 1 min at room temperature. The digested peptides were transferred to new LoBind microcentrifuge tubes. Each digested sample was labeled with TMT (Thermo Fisher Scientific; 90406). In general, for each sample, 8 μL of the 20 μg/L stock of TMT reagent was added along with dry MS-grade acetonitrile to dissolve a final acetonitrile concentration of ~9% v/v. The digested peptides were digested for ~1.5 h at RT while rotating. The streptavidin beads were pelleted by centrifugation (2,000 rpm, 2 min, 4 °C) and the supernatant containing the digested peptides were transferred to new LoBind microcentrifuge tubes. Each digested sample was dried via vacuum centrifugation, and the samples were combined by redissolving each sample with 400 μL of buffer A (5% MeCN in H2O, 0.1% formic acid) and transferring the solution into each sample tube until all samples were redissolved (final volume 600 μL). pH was adjusted to ~8.0 with 10 mM HCl. The combined sample was desalted using two C-18 columns (Thermo Fisher Scientific; 89870) according to the manufacturer’s instructions. The combined sample was dried via vacuum centrifugation and stored in the −80 °C until ready for injection.

LC-MS Analysis for Proteomics. Here, 10-plex samples were redissolved in 20 μL of buffer A (99.9% H2O, 0.1% formic acid) for MS analysis. In brief, 3 μL of each sample was loaded onto an Acclaim PepMap 100 precolumn (75 μm × 2 mm) and eluted on an Acclaim PepMap RSLC analytical column (75 μm × 15 cm) using the UltiMate 3000 RSLCnano system (Thermo Fisher Scientific). Buffer A (0.1% formic acid in H2O) and buffer B (0.1% formic acid in acetonic) were used to establish the 220 min gradient comprising 10 min of 2% buffer B, 192 min of 2% to 30% buffer B, 8 min of 30% to 60% buffer B, 1 min of 60% to 95% buffer B, 5 min of 95% buffer B, and 1 min of 95% to 2% buffer B, followed by reequilibration at 2% buffer B for 6 min. The flow rate was 0.3 mL/min at 35 °C.

Peptides were then analyzed with a Thermo Fisher Scientific Orbitrap Fusion Lumos proteomic mass spectrometer in a data-dependent manner with a cycle time of 3 s. Voltage applied to the nano-LC electrospray ionization source was 2.0 kV. MS1 spectra were acquired at a resolution of 120,000 with an automatic gain control (AGC) target value of 1 × 106 ions and a maximum injection time of 50 ms (dynamic exclusion enabled; repeat count 1, duration 20s). The scan range was limited from 375 to 1,500 m/z. Peptide fragmentation for MS2 was performed via collision-induced dissociation (CID), tandem quadrupole ion trap analysis, AGC 1.8E4, CID collision energy 30%, maximum injection time 120 ms, and isolation window at 1.6. MS3 precursor was fragmented via high-energy collision-induced dissociation (HCD) with a collision energy of 65%. Synchronous precursor selection (SPS) was enabled to include up to 10 MS2 fragment ions for the MS3 spectrum. The detector type was an Orbitrap with a resolution of 50,000, an AGC target value of 1.5 × 106, and a maximum injection time of 120 ms.

Proteomics Data Analysis. Data processing was performed using Proteome Discoverer 2.4 software (Thermo Fisher Scientific), and peptide sequences were determined by matching protein databases with the acquired fragment pattern using the SEQUEST HT algorithm. The precursor mass tolerance was set to 10 ppm, and fragment ion mass tolerance was set to 0.6 Da. One missed cleavage site of trypsin was allowed. Carbamidomethyl
proteomics datasets have been deposited to the ProteomeXchange Consortium via the PRIDE (62) partner repository with the dataset identifier PXD021622.

ACKNOWLEDGMENTS. E.J. is supported by a Skaggs Graduate Fellowship, enabled by the Henry and Jennifer Luttrell Foundation. M.L.H. is supported by an NIH K99/R00 Pathway to Independence Award (R00HD0292-03). The M.L.H. lab is supported by startup funds from Scripps Research and the Joe W. and Dorothy Dorsett Brown Foundation.

Data Availability. All MS raw files have been deposited to the PRIDE archive (https://www.ebi.ac.uk/pride/archive) with the identifier PXD021622.
53. C. G. Liao et al., Characterization of basigin isoforms and the inhibitory function of basigin-3 in human hepatocellular carcinoma proliferation and invasion. Mol. Cell. Biol. 31, 2591–2604 (2011).

54. J. Dietrich et al., Role of CD3 gamma in T cell receptor assembly. J. Cell Biol. 132, 299–310 (1996).

55. J. Dietrich, X. Hou, A. M. Wegener, C. Geisler, CD3 gamma contains a phosphoserine-dependent di-leucine motif involved in down-regulation of the T cell receptor. EMBO J. 13, 2156–2166 (1994).

56. Y. Bulut, E. Faure, L. Thomas, O. Equils, M. Arditi, Cooperation of Toll-like receptor 2 and 6 for cellular activation by soluble tuberculosis factor and Borrelia burgdorferi outer surface protein A lipoprotein: Role of Toll-interacting protein and IL-1 receptor signaling molecules in Toll-like receptor 2 signaling. J. Immunol. 167, 987–994 (2001).

57. A. O. Aliprantis et al., Cell activation and apoptosis by bacterial lipoproteins through Toll-like receptor 2. Science 285, 736–739 (1999).

58. H. D. Brightbill et al., Host defense mechanisms triggered by microbial lipoproteins through Toll-like receptors. Science 285, 732–736 (1999).

59. M. Triantafilou et al., Membrane sorting of Toll-like receptor (TLR)-2/6 and TLR2/1 heterodimers at the cell surface determines heterotypic associations with CD36 and intracellular targeting. J. Biol. Chem. 281, 31002–31011 (2006).

60. J. Paek et al., Multidimensional tracking of GPCR signaling via peroxidase-catalyzed proximity labeling. Cell 169, 338–349.e11 (2017).

61. L. Xu et al., Human hepatic stellate cell lines, LX-1 and LX-2: New tools for analysis of hepatic fibrosis. Gut 54, 142–151 (2005).

62. Y. Perez-Riverol et al., The PRIDE database and related tools and resources in 2019: Improving support for quantification data. Nucleic Acids Res. 47, D442–D450 (2019).