Published in final edited form as:

ACS Chem Biol. 2018 September 21; 13(9): 2560–2567. doi:10.1021/acscchembio.8b00511.

Paramagnetic Tag for Glycosylation sites in Glycoproteins: Structural Constraints on Heparan Sulfate Binding to Robo1

Maria J. Moure^{#†}, Alexander Eletsky^{#†}, Qi Gao^{†,‡}, Laura C. Morris[†], Jeong-Yeh Yang[†], Digantkumar Chapla[†], Yuejie Zhao^{†,‡}, Chengli Zong[†], I. Jonathan Amster[‡], Kelley W. Moremen^{†,§}, Geert-Jan Boons^{†,‡,||}, and James H. Prestegard^{*,†,§,‡}

[†]Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602, United States

[‡]Department of Chemistry, University of Georgia, Athens, Georgia 30602, United States

[#]Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia 30602, United States ^{||}Department of Chemical Biology and Drug Discovery, Utrecht Institute for Pharmaceutical Sciences, and Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands

[#] These authors contributed equally to this work.

Abstract

An enzyme- and click chemistry-mediated methodology for the site-specific nitroxide spin labeling of glycoproteins has been developed and applied. The procedure relies on the presence of single N-glycosylation sites that are present natively in proteins or that can be engineered into glycoproteins by mutational elimination of all but one glycosylation site. Recombinantly expressing glycoproteins in HEK293S (GnT1-) cells results in N-glycans with high-mannose structures that can be processed to leave a single GlcNAc residue. This can in turn be modified by enzymatic addition of a GalNAz residue that is subject to reaction with an alkyne-carrying TEMPO moiety using copper (I)-catalyzed Click Chemistry. To illustrate the procedure, we have made an application to a two-domain construct of Robo1, a protein that carries a single N-glycosylation site in its N-terminal domains. The construct has also been labeled with ¹⁵N at amide nitrogens of lysine residues to provide a set of sites that are used to derive an effective location of the paramagnetic nitroxide moiety of the TEMPO group. This, in turn, allowed measurements of paramagnetic perturbations to the spectra of a new high affinity heparan sulfate ligand. Calculation of distance constraints from these data facilitated determination of an atomic level model for the docked complex.

*Corresponding Author: jpresteg@ccrc.uga.edu.

Present Addresses

Q.G; Structure Elucidation Group, Process and Analytical Research and Development, Merck & Co., Inc., 2000 Gal-lop-ing Hill Road, Kenilworth, NJ, 07033, USA.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

ASSOCIATED CONTENT

Supporting Information Available: This material is available free of charge via the Internet: Synthetic protocols, compound characterization, experimental details of NMR data acquisition, MD analysis and calculation of ligand-protein interaction energies (PDF)

The authors declare no competing financial interest.

For table of contents only

Glycosylation sites are selectively tagged for paramagnetic NMR

Glycosylation is an important form of post-translational modification of mammalian proteins with glycans playing functional roles that include altering stability, affecting cell migration, and modulating signaling.^{1, 2} Glycosylation is also very common, with more than half of mammalian proteins being glycosylated.^{3, 4} Yet, glycosylation is often avoided in structural studies, because it can make formation of suitable crystals for X-ray diffraction studies less likely,⁵ and it restricts options for isotopic labeling in NMR studies when expression of proteins in eukaryotic cells is required.⁶ Here we present an NMR approach that, not only allows glycosylation, but takes advantage of glycosylation sites to extend structural studies.

One of the more efficient means of isotopic labeling of glycosylated proteins for NMR structural studies involves supplementing mammalian cell cultures with single or multiple types of isotopically labeled amino acids.^{7–9} The proteins labeled in this way will not provide the numerous specifically assigned NOEs used in most NMR structures, and NOE data need supplementation with other data types. Many NMR studies also focus, not on *de novo* determination of protein structures, but on the ways proteins, or protein domains, interact with other proteins or functionally relevant ligands. In all of these cases it has proven very valuable to add long range structural information from paramagnetic perturbations.¹⁰ This usually requires the addition of a “tag” that carries unpaired electrons in chemical entities such as the nitroxide of a 2,2,6,6-tetramethyl-1-piperidinyloxy (TEMPO) group or in metal ions complexed to a chemically attached chelate. These tags must be added to a specific site with minimal effect on protein structure and function, but still allow sufficient definition of tag geometry to derive distance constraints on placement of structural elements of single domain proteins, interacting domains of multiple domain proteins or bound ligands of a protein complex. In this paper, we demonstrate an ability to meet these requirements by adding a TEMPO group to a glycoprotein via a unique *N*-glycosylation site carrying a single GlcNAc unit.

The use of paramagnetic tags in the structural characterization of proteins has a long history.¹¹ Their use in NMR began with the covalent attachment of the nitroxide-carrying TEMPO moiety to an isolated cysteine residue through a disulfide bond.¹² Paramagnetic relaxation enhancement (PREs), which display a useful $1/r^6$ distance dependence, are easily measured from the reduction in intensity of crosspeaks in typical ^1H - ^{15}N heteronuclear single quantum coherence (HSQC) spectra of the protein or from the simple reduction in intensity of resonances from a bound ligand. Effects can be observed up to distances of 25 to 40 Å,

depending on the paramagnetic species. Application has recently expanded to include attachment of chelates that bind paramagnetic lanthanides; these provide PREs as well as pseudo-contact shifts (PCSs) and field induced alignment.^{13, 14} While modes of attachment have also expanded to include lanthanide binding peptides inserted in the peptide sequence,¹⁵ peptide segments that carry aldehyde groups susceptible to reductive amination,¹⁶ and attachment of tags to ligands that bind proteins,¹⁷ attachment via a disulfide linkage to a single cysteine in the protein still dominates the field. A cysteine can be inserted at a specific site by adding or converting an existing codon to a cysteine codon when cysteines are absent in the native sequence or by converting all but one cysteine to serines when multiple cysteines exist. For many proteins, including that used for illustration here, elimination of cysteines is a problem. Internal disulfide bonds are often needed for structural stability and a non-bonded cysteine is sometimes involved in function. Hence, there is a need for alternate attachment options.

We envision a protein tagging strategy that capitalizes on expression of a target protein in mammalian cells deficient in a particular N-acetylglucosaminyltransferase, N-acetylglucosaminyltransferase I (GnT1).¹⁸ This prevents extension of glycans into complex forms and leaves only oligomannose (Man₅GlcNAc₂-Asn) glycans that can be cleaved to a single GlcNAc by the action of endoglycosidase F1 (EndoF1).¹⁹ It is frequently the case that this single GlcNAc is adequate to maintain structural and functional characteristics of *N*-glycosylated glycoproteins.²⁰ A “clickable” monosaccharide carrying an azide group (GalNAz) will then be installed to these sites enzymatically and the modified glycoprotein will be conjugated to an alkyne terminated TEMPO group using a click chemistry reaction. The absence and reasonable inertness of azides and alkynes in biology has made the copper-catalyzed azide-alkyne cycloaddition (CuAAC) an excellent candidate for undertaking modification of biomolecules.²¹ Indeed, the CuAAC can be performed site-selectively with complete conversion and has been used in many significant applications,²² including those involving carbohydrates.²³ This basic scheme for introducing a paramagnetic tag to Robo1-Ig1-2 is summarized in Figure 1.

We illustrate our tagging procedure with application to a segment of the human Roundabout 1 (Robo1) protein, corresponding to the two N-terminal domains (Robo1-Ig1-2). Robo1 is a glycosylated cell surface signaling molecule which plays an important role in axon guidance during mammalian development.²⁴ Robo1-Ig1-2 contains a single *N*-glycosylation site at N160 and two pairs of disulfide bonded cysteines which are believed essential for stability.

We have previously characterized the heparan sulfate binding properties of this Robo1 construct,²⁵ we have assigned crosspeaks in a ¹H-¹⁵N HSQC spectrum of the protein labeled by supplementing expression media with ¹⁵N-enriched lysine and we have used these assignments in conjunction with chemical shift perturbation data to generate a model for a complex with a ligand of moderate affinity (HS4-1).²⁶ Here we use the new tagging procedure and a novel diffusion editing NMR experiment to collect PREs on an exchanging ligand with higher affinity (IdoA-GlcNS6S-IdoA-GlcNS6S-(CH₂)₅NH₂).²⁷ We refer to this ligand as HS4-2. Interpretation of the PREs in terms of distance constraints is facilitated by generation of a distributed conformation model for the tag, and these constraints are used to generate an atomic level model for the Robo1-Ig1-2 / HS4-2 complex. This study, not only

leads to identification of interactions essential for high affinity in the Robo1 system, but demonstrates the utility of an alternative to procedures which use cysteine disulfides for paramagnetic tagging. We believe that it will be applicable to a significant number of glycosylated protein constructs which natively contain a single glycosylation site, or can be engineered to remove all but a single site.

RESULTS AND DISCUSSION

Syntheses of Tags.

Selectivity in site modification was satisfied by choosing “click-chemistry” involving bio-orthogonal reactive groups, namely an azide and an alkyne.²⁸ The azide moiety is conveniently carried in an analog of GalNAc (GalNAz) and this was added to an existing GlcNAc residue on the protein using a mutated form of a β 1,4-galactosyltransferase (human β 4GALT1 (Y285L)).²⁹ Addition to the GlcNAc removes the tag sufficiently from the protein surface to minimize structural perturbations, yet the glycosidic linkage between the GalNAz and GlcNAc is expected to be sufficiently well-defined to restrict conformational sampling. TEMPO was selected as the paramagnetic moiety for this initial application because of its small size in comparison to lanthanide chelates. The alkyne moiety was easily added to commercially available 4-hydroxy-TEMPO using 3-bromopropyne and sodium hydride. The reaction of alkyne and azide is catalyzed by Cu(I) in the presence of THPTA to maintain Cu(I) levels while suppressing undesired side reactions.³⁰ The product contains a five-membered heterocycle which also contributes to conformational restriction. The chemical structures of key reactants, as well as the resulting reaction product are shown in Figure 2. Details of synthetic procedures are included in Supporting Information.

Preparation of the Protein.

Robo1-Ig1-2 was prepared as described previously²⁶ except that the protein was expressed in HEK293S (GnT1⁻) cells¹⁸ and the resulting glycans, which are predominantly Man₅GlcNAc₂, were cleaved to a single GlcNAc using EndoF1. The glycosylation site is located in an extended three-residue stretch connecting two beta strands (155-157 and 161-168) and is expected to be somewhat conformationally restricted. The protein was also isotopically labeled as previously described²⁶ by including ¹⁵N-lysine in the expression medium in order to provide sites to assess the distribution of conformers adopted by the added tag. The tag was added, monitoring the progress of GalNAz addition as well as the ligation process by ESI-MS spectroscopy. This analysis, as well as the near complete abolition of NMR crosspeaks from residues close to the tag (K103 in Figure 3B), showed tag addition to be greater than 90% complete. Comparison of the position of remaining crosspeaks to crosspeaks in Robo1 constructs with wild-type, Man₅GlcNAc₂, and GlcNAc glycosylation suggest the retention of the protein's structural features.

PRE Measurements on the Protein.

To make use of the tag in assessment of bound ligand geometry we must first properly position the tag at the surface of the protein. To do this we will utilize PREs for backbone amide protons of ¹⁵N-labeled lysines. There are 12 lysine sites in the Robo1-Ig1-2 construct. ¹H-¹⁵N HSQC spectra of Robo1-GalNAz with and without TEMPO are shown in Figure 3.

Only 10 crosspeaks are observed in the Robo1-Ig1-2-GalNAz spectrum (Figure 3A) and only 9 are seen with TEMPO added (Figure 3B). The peaks missing in both correspond to residues 137 and 266. The former is a lysine in the binding loop whose resonance is broadened due to motion on the timescale of inverse chemical shift differences and the latter is a lysine very close to the C-terminus that appears to have been proteolytically cleaved during expression. These assignments have been verified by mutagenesis and mass spectrometry analysis.²⁶ The remainder of the crosspeaks were assigned using a sparse-label assignment strategy that has been previously described.^{8, 26} The crosspeak for residue K103 is not seen in the TEMPO spectrum of Figure 3B. However, a residual peak can be seen upon lowering the threshold of the plot. This is approximately 5% of the intensity of the crosspeak in the absence of TEMPO and likely corresponds to unmodified protein. K103 is obviously so close to the nitroxide group that we can only place an upper limit to the distance between the TEMPO nitroxide and the amide proton of K103.

Crosspeaks belonging to K81, K90, and K112 show a significant, but measurable, intensity loss, placing them at intermediate distances. The crosspeaks from other residues are affected less, indicating that they are far from the spin label. To be sure that the losses are purely due to PREs and not changes in dynamics induced by the addition of TEMPO, the nitroxide was reduced to a hydroxyl amine with sodium ascorbate to eliminate paramagnetic effects. The cross peaks all returned to the intensities similar to those seen with GalNAz (See Supplementary Figure S4).

Changes in crosspeak intensities between spin-labeled and non-labeled forms of Robo1 can be related more quantitatively to distances between the nitroxide oxygen and amide protons on lysines using the expression in equation 1. V_{dia} is the crosspeak volume without spin label and V_{para} is the volume with spin label. T is the total time of transfer and refocusing periods in the HSQC pulse sequence (10.8 ms); r is the distance between each amide proton and the effective position of the unpaired electron on the nitroxide, τ_c is taken to be the rotational correlation time for protein tumbling; ω_H is the Larmor frequency for the proton; and K is a sum of constants related to spin properties of the system ($1.23 \times 10^{-44} \text{ m}^6 \text{ s}^{-2}$). The use of this formula with reasonable estimates of a correlation time ($\sim 19 \text{ ns}$) would lead us to expect a greater than 10% reduction in crosspeak volume at $20 \text{ \AA}$ and loss of reliable volume estimates at distances less than $12 \text{ \AA}$. These limits can prove useful in choosing a tagging site when multiple glycosylation sites offer a choice of positions. In our case the single glycosylation site falls at a useful distance from the region known to be involved in ligand binding based on mutational data and chemical shift perturbation.

$$\ln\left(\frac{V_{dia}}{V_{para}}\right) = T \times \frac{K}{r^6} \times \left(4\tau_c + \frac{3\tau_c}{1 + \omega_H^2 \tau_c^2}\right) \quad (1)$$

Table 1 gives a list of measurements for V_{dia}/V_{para} along with predicted r values for lysines in the first domain of Robo1-Ig1-2. There are small reductions in volume for two residues in the second domain (K224 and K206), but because of uncertainty about inter-domain orientation and motion, we will ignore these. The column labeled 100% would correspond to

distances from various lysine amide protons to the nitroxide oxygen on the TEMPO group if the tag were rigidly held in one conformation.

If the tag were, in fact, rigidly held in one conformation, these distances could be used to define the position of the nitroxide group. Attaching the tag to the nitrogen of N160 of the crystal structure (PDB ID 2V9R)³¹ using tools in Chimera³² we unfortunately find it physically impossible for the nitroxide group to reach the point defined by these distances without severe clashes with atoms of residues on the protein surface. Despite our efforts to build a conformationally restricted tag, it must have significant flexibility and sample a number of conformations that contribute to the measured PREs. Ideally a computational procedure for generation of an ensemble of conformers could be pursued.³³ Initially we will adopt a more manual approach.

To generate a set of possible conformers we have assumed the GalNAz-GlcNAc glycosidic bond torsion angle of the tag remains fixed at values predicted for a GalNAc-GlcNAc linkage using GLYCAM-Web tools,³⁴ amide bonds are fixed in trans conformations and the amide N-H bonds are oriented trans relative to the C-H bonds at their attachment sites as observed in most acetylated amino sugars. This leaves a set of seven adjustable torsions that can generate a distribution of possible conformers. Initially we searched for a minimal set of physically realistic conformers that can satisfy PRE constraints; actually a pair of equally populated conformers in which each member satisfied the distance of close approach to K103 (12 Å) plus one or more of the distances to the other three lysines. The search was facilitated by first searching for close approach distances needed for 50% occupation of each conformer (see column 4 of Table 1). The two structures (model 1 and model 2) are shown attached to a ribbon diagram of Robo1-Ig1-2 in Figure 4A with distances to perturbed lysine amide protons shown. These change by 3% or less after minimization. The sum of the PRE contributions from the two conformers predicts volume ratios of crosspeaks quite well (Column 5 in Table 1).

The pair is, of course, an over simplification of the true distribution. To generate something closer to a true distribution a long (1 μ s) molecular dynamics simulation of the tagged protein in a box of explicit TIP3P water molecules was run. Interestingly, the resulting distribution of structures is not uniform, but is highly clustered (see plot in Supplementary Figure S5). An analysis of nitroxide oxygen positions using 5 clusters and the “kmeans” routine in MATLAB yielded 4 clusters that accounted for approximately 95% of all frames. Three of these, accounting for more than 75% of the frames, had members with significant protein surface contacts, usually involving contacts that could be attributed to hydrophobic contacts between protein residues and the TEMPO group or its triazole linker. One of the three clusters contained structures with nitroxide oxygen positions less than 2.5 Å from that of model 1 and distances to the amide protons of K81 and K103 of approximately 17 and 13 Å respectively. Hence, model 1 can be considered a reasonable representation of a physically accessible state. This is significant as only model 1 displays a close approach of its nitroxide oxygen to K81, the one lysine residue previously implicated in binding of HS oligomers. Below we will use the position of this oxygen as an anchor position to derive distance constraints from PRE effects on bound ligands.²⁶

PRE Measurements on a Bound Ligand.

Collection of PRE data from proton resonances of a rapidly exchanging ligand lacking isotope labels is actually challenging, particularly when the ligand is primarily composed of sugar residues. Most sugar resonances occur in the 3-5 ppm region where substantial protein resonance intensity exists, and, in contrast to standard practice with STD and transferred NOE (trNOE) experiments, a large excess of ligand cannot be used, since PRE effects are scaled down linearly by the ratio of bound to free ligand. We have devised an experiment that uses diffusion ordered spectroscopy (DOSY)³⁵ to deconvolute spectra of a small, rapidly diffusing, ligand from protein resonances, and use these spectra in extraction of PREs. See Supplementary Figure S4 for a comparison of raw and deconvoluted spectra and the chemical structure of HS4-2. Regions showing most glycan resonances from deconvoluted one-dimensional ¹H spectra of the HS4-2 ligand are shown in Figure 5. Spectra labeled A, B and C correspond to Robo1-Ig1-2 with the normally oxidized paramagnetic tag, with the tag reduced by ascorbate to a diamagnetic state and without the tag, respectively. There are clear reductions in amplitudes and areas of peaks from protons H1D, H4D and H5D in the presence of the paramagnetic tag (Figure 5A vs 5C). They recover most of their intensity when the paramagnetic tag is reduced with ascorbate producing a diamagnetic hydroxyl amine (Figure 5B). These strongly perturbed resonances all belong to the hexose ring at the non-reducing terminus (ring D) suggesting a bound geometry very similar to that depicted for the HS4-1 ligand in previous work (see Figure 4B).

It is possible to interpret these perturbations more quantitatively using a modified version of the formula in Equation 1. The DOSY sequences used to collect the spectra have delays for application of pulsed field gradients in which proton magnetization is transverse (time T in the equation). These amount to approximately 10 ms. There are additional delays for the diffusion time, but the magnetization is longitudinal during these periods and the PRE contribution to the T_1 (longitudinal) relaxation is less than 1% of the PRE contribution to the T_2 spin relaxation rate.

Integrals (areas A) of the affected peaks replace the volumes in equation 1 and there is a scaling factor for the fraction of ligand bound to the protein as well as a factor accounting for the 50% occupancy of model 1. Dissociation constants have been determined to be approximately 20 μ M by SPR²² and approximately 45 μ M by NMR.²³ In either case the fraction bound is within 10% of the ratio of protein to ligand (1/2). The A_{para}/A_{dia} values derived from the spectra are given in Table 2 along with distances determined using the modified Equation 1.

Docked model of HS4-2 bound to Robo1-Ig1-2.

The distances labeled r_{50} in Table 2 have been used to produce a docked model of the ligand-protein complex. The initial conformation of the ligand was produced using GLYCAM-Web.³⁴ The starting structure of protein was taken from the crystal structure of human Robo1-Ig1-2 (PDB ID 2V9R).³¹ However, a substantial manual movement of the V133-D141 loop was required to place HS4-2 at distances compatible with those derived from PRE effects. A crosspeak for K137 in the middle of this loop is not, in fact, observed in

our spectra due to chemical exchange broadening of its resonances, supporting the natural flexibility of this loop. Large movements of the loop prohibit direct application of most docking programs to the original crystal structure. Instead, a manually docked structure was generated by positioning HS4-2 with its GLYCAM-Web conformation in a Robo1-Ig1-2 molecule having its tag as positioned as in model 1 and the V133-D141 loop adjusted to allow a close match to r_{50} distances. This structure was refined by subjecting it to minimization and a short (50 ns) molecular dynamics (MD) run.

A structure from the middle of the trajectory is depicted in Figure 6A showing distances from the nitroxide oxygen of model 1 (see Figure 4) to the anomeric protons of the four residues in HS4-2 of 9.5, 14.7, 16.5 and 20.0 Å. Figure 6B shows a stereo figure of the binding site for a structure earlier in the trajectory that depicts all strongly interacting residues. Additional depictions of these structures are included in Supplementary Figure S6. 54% of the structures throughout the trajectory share the seven strongest interactions shown (separation < 1 Å over VDW contact); an additional 32% share 6 of the 7 and the remaining 14% share 5 of the 7. While interactions with the V133-D141 loop use several sequentially equivalent positively charged residues seen in a previous crystal structure of *Drosophila* Robo1-Ig1-2 complexed with a more highly sulfated heparin tetramer,³⁶ the position of the ligand is quite different (see Supplementary Figure S7). The crystal structure has the ligand poised between two Robo1 molecules in the unit cell and interacts with these residues on one side of the loop. Our model interacts with the same residues, but on the other side of the loop. The position of the ligand does share more similarities to that of the 2-O-sulfated tetramer (HS4-1) studied in our previous work and shown in Figure 4B.²⁶

Energy analysis of the Robo1-Ig1-2 – HS4-2 complex.

Calculation of interaction energies and decomposition of those energies on a per-residue basis is useful in understanding the specific interactions that drive binding and contribute to specificities for particular ligands. There are several possible approaches to this decomposition.³⁷ We have chosen a Generalized Born-Surface Area (GBSA) approach for its computational efficiency,³⁸ and used it on the 50 ns trajectory discussed above. The approach approximates solvation contributions to binding free energies, but does not include conformational or vibrational contributions to binding entropies. Neglecting these contributions results in a substantial overestimation of total binding free energies, but it is not likely to result in reordering of per-residue effects.

The contributions to energies for HS4-2 residues and the protein residues contributing significantly to interactions are listed in Table 3. The sum of these interactions is strongly negative, and importantly, it is more negative than the sum of a similar range of interactions from an analysis conducted on the same ligand on the outside of the V133-D141 loop. For the ligand, 6-O-sulfate groups have been considered separately, as these have previously been suggested as important groups for Robo1 interactions.³⁹ The first point to note is that electrostatic and polar solvation contributions dominate in these groups. In most cases these two contributions nearly cancel one another. For the 6-O-sulfate groups this results in a small net positive contribution to binding free energies. Therefore, these groups clearly don't drive association, but strong interactions with positively charged protein residues do

compensate for the large sulfate de-solvation energies. If these interactions were absent, total binding energies would be strongly positive rather than negative. In particular, interactions between R249 and R136 for O6S-A and K137 for O6S-C are important.

A second point to note is that the largest contributions to interaction energies are from the non-sulfated iduronic acid groups, particularly IdoA B, which has strong interactions between its carboxylate group and R169 and sometimes H134. This is significant as the tetramer lacking 2-O-sulfation is known to bind more strongly to Robo1-Ig1-2 than a tetramer having this group sulfated.²⁵ In the geometry shown in Figure 6B sulfating the 2 position of IdoA C would have strong repulsive interactions with other sulfates on the ligand and would likely cause a conformation change, negating a number of other favorable interactions.

One final point to note is that residues on both the first Ig domain (K137, K81, H134, R136 and R139) and second Ig domain (R249, H197, R195, and R169) interact with the ligand. The relative orientation of the two domains is highly mobile, as evidenced in our 1 μ s MD trajectory. However, even this trajectory may not adequately sample all possible domain orientations. It is likely that additional experimental information on relative domain orientations in the presence of ligand will be needed to fully elucidate ligand interactions with Robo1-Ig1-2. There are also additional positive residues on both domains (R131 and R173) that may be important for interactions with larger ligands. Interactions with longer ligands may be needed to understand the influence of native HS oligomers on inter-domain orientation.

Conclusion.

A methodology for the site-selective nitroxide spin labeling of glycoproteins having a single N-linked glycosylation site has been developed and its utility in the structural characterization of ligand-glycoprotein complexes has been illustrated using the Robo1-Ig1-2-heparan sulfate system. In this initial application, a heparan sulfate binding site that is distinctly different from one seen in the crystal structure of a homologous protein has been identified along with key interactions that may drive specificity. While the methodology takes advantage of a single glycosylation site natively present in the Robo1 construct studied, it should be applicable to other glycoprotein constructs that can be engineered to have just a single glycosylation site. It is also applicable to glycoproteins having cysteines required for stability or function which may not be amenable to modification by tags dependent on disulfide bond formation. The DOSY method used to separate ligand spectra from protein spectra also has potential application well beyond that illustrated here.

As to future developments and applications, the chemical basis of the tagging strategy should be applicable to chelates carrying lanthanide ions as well as TEMPO groups. Lanthanide ions such as Gd³⁺ produce measurable PREs at longer distances than TEMPO, and other lanthanides provide additional angular dependent data through pseudo-contact shifts (PCSs). The general strategy described for the determination of a ligand protein complex is equally applicable to the determination of inter-domain geometry of multiple domain proteins and the structure of protein-protein complexes. We anticipate pursuing some of these applications in the future.

METHODS

A detailed description of the methods is provided in the Supporting Information.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENT

This work was supported by grants from the National Institute of General Medical Sciences, P41-GM103390 and P01-GM107012, as well as a grant in support of NMR instrumentation, S10 RR027097. Manuscript content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We thank C-W Chou (Proteomics, UGA) for assistance with Mass Spectrometry measurements and C-Y Chen for assistance with NMR in the early stages of this work.

ABBREVIATIONS

PRE	paramagnetic relaxation enhancement
DOSY	diffusion ordered spectroscopy
CuAAC	copper-catalyzed azide-alkyne cycloaddition
EndoF1	endoglycosidase F1
GnT1	N-acetylglucosaminyltransferase I
HS4-1	IdoA-GlcNS6S-IdoA2S-GlcNS6S-(CH ₂) ₅ NH ₂
HS4-2	IdoA-GlcNS6S-IdoA-GlcNS6S-(CH ₂) ₅ NH ₂
HS	heparan sulfate
HSQC	heteronuclear single quantum coherence
MM/GBSA	molecular mechanics generalized Born surface area
PCS	pseudo-contact shifts
Robo1-Ig1-2	two N-terminal domains of Roundabout 1
SPR	surface plasmon resonance
STD	saturation-transfer difference
TEMPO	2,2,6,6-tetramethyl-1-piperidinyloxy
THPTA	Tris(3-hydroxypropyltriazolylmethyl)amine
trNOE	transfer nuclear Overhauser effect

REFERENCES

- [1]. Varki A, and Gagneux P (2017) Biological Functions of Glycans, In Essentials of Glycobiology (Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Darvill AG, Kinoshita T, Packer

NH, Prestegard JH, Schnaar RL, and Seeberger PH, Eds.) 3rd ed., pp 77–88, Cold Spring Harbor Laboratory Press Copyright 2015-2017 by The Consortium of Glycobiology Editors, La Jolla, California. All rights reserved., Cold Spring Harbor (NY).

- [2]. Krall N, da Cruz FP, Boutureira O, and Bernardes GJ (2016) Site-selective protein-modification chemistry for basic biology and drug development, *Nat Chem* 8, 103–113. [PubMed: 26791892]
- [3]. Apweiler R, Hermjakob H, and Sharon N (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database, *Biochimica Et Biophysica Acta-General Subjects* 1473, 4–8.
- [4]. Spiro RG (2002) Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds, *Glycobiology* 12, 43R–56R.
- [5]. Shirouzono T, Chirifu M, Nakamura C, Yamagata Y, and Ikemizu S (2012) Preparation, crystallization and preliminary X-ray diffraction studies of the glycosylated form of human interleukin-23, *Acta Crystallogr Sect F Struct Biol Cryst Commun* 68, 432–435.
- [6]. Verardi R, Traaseth NJ, Masterson LR, Vostrikov VV, and Veglia G (2012) Isotope labeling for solution and solid-state NMR spectroscopy of membrane proteins, *Adv Exp Med Biol* 992, 35–62. [PubMed: 23076578]
- [7]. Prestegard JH, Agard DA, Moremen KW, Lavery LA, Morris LC, and Pederson K (2014) Sparse labeling of proteins: Structural characterization from long range constraints, *Journal of Magnetic Resonance* 241, 32–40. [PubMed: 24656078]
- [8]. Gao Q, Chalmers GR, Moremen KW, and Prestegard JH (2017) NMR assignments of sparsely labeled proteins using a genetic algorithm, *Journal of Biomolecular Nmr* 67, 283–294. [PubMed: 28289927]
- [9]. Subedi GP, Johnson RW, Moniz HA, Moremen KW, and Barb A (2015) High Yield Expression of Recombinant Human Proteins with the Transient Transfection of HEK293 Cells in Suspension, *Jove-Journal of Visualized Experiments*.
- [10]. Pederson K, Mitchell DA, and Prestegard JH (2014) Structural Characterization of the DC-SIGN-Lewis(X) Complex, *Biochemistry* 53, 5700–5709. [PubMed: 25121780]
- [11]. Cornish VW, Benson DR, Altenbach CA, Hideg K, Hubbell WL, and Schultz PG (1994) Site-specific incorporation of biophysical probes into proteins, *Proc Natl Acad Sci U S A* 91, 2910–2914. [PubMed: 8159678]
- [12]. Battiste JL, and Wagner G (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data, *Biochemistry* 39, 5355–5365. [PubMed: 10820006]
- [13]. Su XC, Huber T, Dixon NE, and Otting G (2006) Site-specific labelling of proteins with a rigid lanthanide-binding tag, *Chembiochem* 7, 1599–1604. [PubMed: 16927254]
- [14]. Nitsche C, and Otting G (2017) Pseudocontact shifts in biomolecular NMR using paramagnetic metal tags, *Progress in Nuclear Magnetic Resonance Spectroscopy* 98-99, 20–49. [PubMed: 28283085]
- [15]. Franz KJ, Nitz M, and Imperiali B (2003) Lanthanide binding tags as versatile protein coexpression probes, *Chembiochem* 4, 265–271. [PubMed: 12672105]
- [16]. Agarwal P, van der Weijden J, Sletten EM, Rabuka D, and Bertozzi CR (2013) A Pictet-Spengler ligation for protein chemical modification, *Proc Natl Acad Sci U S A* 110, 46–51. [PubMed: 23237853]
- [17]. Deshauer C, Morgan AM, Ryan EO, Handel TM, Prestegard JH, and Wang X (2015) Interactions of the Chemokine CCL5/RANTES with Medium-Sized Chondroitin Sulfate Ligands, *Structure* 23, 1066–1077. [PubMed: 25982530]
- [18]. Reeves PJ, Callewaert N, Contreras R, and Khorana HG (2002) Structure and function in rhodopsin: High-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line, *Proceedings of the National Academy of Sciences of the United States of America* 99, 13419–13424. [PubMed: 12370423]
- [19]. Freeze HH, and Kranz C (2010) Endoglycosidase and glycoamidase release of N-linked glycans, *Curr Protoc Mol Biol Chapter* 17, Unit 17 13A.

- [20]. Arnold JN, Wormald MR, Sim RB, Rudd PM, and Dwek RA (2007) The impact of glycosylation on the biological function and structure of human immunoglobulins, *Annu Rev Immunol* 25, 21–50. [PubMed: 17029568]
- [21]. Meldal M, and Tornøe CW (2008) Cu-catalyzed azide-alkyne cycloaddition, *Chem Rev* 108, 2952–3015. [PubMed: 18698735]
- [22]. Sun T, Yu SH, Zhao P, Meng L, Moremen KW, Wells L, Steet R, and Boons GJ (2016) One-Step Selective Exoenzymatic Labeling (SEEL) Strategy for the Biotinylation and Identification of Glycoproteins of Living Cells, *J Am Chem Soc* 138, 11575–11582. [PubMed: 27541995]
- [23]. Clark PM, Dweck JF, Mason DE, Hart CR, Buck SB, Peters EC, Agnew BJ, and Hsieh-Wilson LC (2008) Direct in-gel fluorescence detection and cellular imaging of O-GlcNAc-modified proteins, *Journal of the American Chemical Society* 130, 11576–11577. [PubMed: 18683930]
- [24]. Andrews W, Barber M, Hernandez-Miranda LR, Xian J, Rakic S, Sundaresan V, Rabbitts TH, Pannell R, Rabbitts P, Thompson H, Erskine L, Murakami F, and Parnavelas JG (2008) The role of Slit-Robo signaling in the generation, migration and morphological differentiation of cortical interneurons, *Dev Biol* 313, 648–658. [PubMed: 18054781]
- [25]. Zong CL, Huang RR, Condac E, Chiu YL, Xiao WY, Li XR, Lu WG, Ishihara M, Wang S, Ramiah A, Stickney M, Azadi P, Amster IJ, Moremen KW, Wang LC, Sharp JS, and Boons GJ (2016) Integrated Approach to Identify Heparan Sulfate Ligand Requirements of Robo1, *Journal of the American Chemical Society* 138, 13059–13067. [PubMed: 27611601]
- [26]. Gao Q, Chen CY, Zong C, Wang S, Ramiah A, Prabhakar P, Morris LC, Boons GJ, Moremen KW, and Prestegard JH (2016) Structural Aspects of Heparan Sulfate Binding to Robo1-Ig1-2, *ACS Chem Biol* 11, 3106–3113. [PubMed: 27653286]
- [27]. Zong C, Venot A, Li X, Lu W, Xiao W, Wilkes JL, Salanga CL, Handel TM, Wang L, Wolfert MA, and Boons GJ (2017) Heparan Sulfate Microarray Reveals That Heparan Sulfate-Protein Binding Exhibits Different Ligand Requirements, *J Am Chem Soc* 139, 9534–9543. [PubMed: 28651046]
- [28]. van Geel R, Wijdeven MA, Heesbeen R, Verkade JM, Wasielec AA, van Berkel SS, and van Delft FL (2015) Chemoenzymatic Conjugation of Toxic Payloads to the Globally Conserved N-Glycan of Native mAbs Provides Homogeneous and Highly Efficacious Antibody-Drug Conjugates, *Bioconjug Chem* 26, 2233–2242. [PubMed: 26061183]
- [29]. Mercer N, Ramakrishnan B, Boeggeman E, Verdi L, and Qasba PK (2013) Use of Novel Mutant Galactosyltransferase for the Bioconjugation of Terminal N-Acetylglucosamine (GlcNAc) Residues on Live Cell Surface, *Bioconjugate Chem.* 24, 144–152.
- [30]. Berg R, and Straub BF (2013) Advancements in the mechanistic understanding of the copper-catalyzed azide-alkyne cycloaddition, *Beilstein J Org Chem* 9, 2715–2750. [PubMed: 24367437]
- [31]. Morlot C, Thielens NM, Ravelli RB, Hemrika W, Romijn RA, Gros P, Cusack S, and McCarthy AA (2007) Structural insights into the Slit-Robo complex, *Proc Natl Acad Sci U S A* 104, 14923–14928. [PubMed: 17848514]
- [32]. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, and Ferrin TE (2004) UCSF Chimera--a visualization system for exploratory research and analysis, *J Comput Chem* 25, 1605–1612. [PubMed: 15264254]
- [33]. Iwahara J, Schwieters CD, and Clore GM (2004) Ensemble approach for NMR structure refinement against H-1 paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule, *Journal of the American Chemical Society* 126, 5879–5896. [PubMed: 15125681]
- [34]. Singh A, Tessier MB, Pederson K, Wang XC, Venot AP, Boons GJ, Prestegard JH, and Woods RJ (2016) Extension and validation of the GLYCAM force field parameters for modeling glycosaminoglycans, *Canadian Journal of Chemistry* 94, 927–935. [PubMed: 28603292]
- [35]. Wu DH, Chen AD, and Johnson CS (1995) An improved diffusion-ordered spectroscopy experiment incorporating bipolar-gradient pulses, *Journal of Magnetic Resonance Series A* 115, 260–264.
- [36]. Fukuhara N, Howitt JA, Hussain SA, and Hohenester E (2008) Structural and functional analysis of Slit and heparin binding to immunoglobulin-like domains 1 and 2 of Drosophila Robo, *Journal of Biological Chemistry* 283, 16226–16234. [PubMed: 18359766]

- [37]. Hadden JA, Tessier MB, Fadda E, and Woods RJ (2015) Calculating binding free energies for protein-carbohydrate complexes, *Methods in molecular biology* (Clifton, N.J.) 1273, 431–465.
- [38]. Chen F, Liu H, Sun HY, Pan PC, Li YY, Li D, and Hou TJ (2016) Assessing the performance of the MM/PBSA and MM/GBSA methods. 6. Capability to predict protein-protein binding free energies and re-rank binding poses generated by protein-protein docking, *Physical Chemistry Chemical Physics* 18, 22129–22139. [PubMed: 27444142]
- [39]. Zhang F, Moniz HA, Walcott B, Moremen KW, Linhardt RJ, and Wang L (2013) Characterization of the interaction between Robo1 and heparin and other glycosaminoglycans, *Biochimie* 95, 2345–2353. [PubMed: 23994753]

Figure 1.
Procedure for nitroxide spin-label attachment to an *N*-glycosylation site.

Figure 2. GalNAz attachment to the residual GlcNAc of Robo1-Ig1-2 using UDPGalNAz and α 1,4-galactosyl-transferase (Y285L) (I) and copper click conjugation of the TEMPO-alkyne (II).

Figure 3. Comparison of 600 MHz HSQC spectra of ^{15}N Lys labeled Robo1-Ig1-2 without spin label (GlcNAc β 1-4GalNAz, A) and with spin label (GlcNAc β 1-4GalNAz β 1-4TEMPO, B). Crosspeak assignments are indicated. The samples are 200 μM in PBS buffer pH = 7.2. Acquisition times are 3.5 hrs and 24 hrs, respectively.

Figure 4.

Models of Robo1-Ig1-2-GalNAz-TEMPO. (A) Two superimposed models showing conformations of the glycan tag and the proximity of their TEMPO groups to lysine residues (Model 1 in cyan and model 2 in magenta). Distances are from the nitroxide oxygen of TEMPO to lysine amides. Lysines with amides showing measurable PREs are labeled. (B) A model showing the conformer with the closest nitroxide approach to the binding site of the 2-O-sulfated HS tetramer from previous work.²³ The distance is from the nitroxide oxygen of TEMPO to the H1 proton of the IdoA at the non-reducing terminus of the HS tetramer (ring D). The Robo1-Ig1-2 models are based on the crystal structure 2V9R.²⁵ The GalNAz-TEMPO conformations were modeled using tools in Chimera²⁷ and are shown in stick representation. All the Lys residues and the HS tetramer are shown with colors keyed to elements.

Figure 5.

1D 800 MHz proton spectra of the HS4-2 ligand produced by deconvolution of DOSY spectra. (A) In the presence of 150 μ M Robo1-Ig1-2 with the TEMPO tag attached at a 2:1 ligand to protein ratio. (B) After the reduction of the sample in A with a 13-fold excess of sodium ascorbate. (C) In the presence of 150 μ M Robo1-Ig1-2 without the TEMPO tag at a 2:1 ligand to protein ratio.

Figure 6. Structures of the Robo1-Ig1-2 – HS4-2 complex. (A) Structure from frame 601 of the 5000 frame MD trajectory. (B) Stereo image of frame 2201 of the trajectory.

Table 1.

PRE measurements and derived nitroxide to lysine amide proton distances for domain 1.

Residue	$V_{\text{para}}/V_{\text{dia}}$	r with 100% occupancy	r with 50% occupancy	Predicted $V_{\text{para}}/V_{\text{dia}}$ 2 conformers ^a
K81	0.68 +/- 0.09	17.2 +/- 1.2/0.9	15.3	0.67
K90	0.73 +/- 0.10	17.8 +/- 1.6/1.1	15.9	0.68
K103	0.03 +/- 0.03	12.0 +/- 12.0/0.4	10.7	0.03
K112	0.89 +/- 0.11	20.9 +/- 12.3/2.3	18.6	0.86

^aThe two equally populated conformers (model 1 and model 2) have distances from the paramagnetic center to K81, K90, K103 and K112 of 15.3, 21.7, 12.0 and 23.7 and 31.9, 15.7, 12.1, and 18.7 respectively.

Table 2.

Perturbations of proton resonances from bound tetrasaccharide, HS4-2.

Proton	H1C	H1A	H1B	H1D	H5D	H4D	H2A/C
$A_{\text{para}}/A_{\text{dia}}$	0.86	0.88	0.80	0.59	0.74	0.26	0.84
r_{100}	17.6	18.1	16.5	14.3	15.8	12.2	17.3
σ^{+}/σ^{-}	1.8/4.7	1.9/7.0	1.2/2.1	0.6/0.8	1.2/2.1	0.3/0.3	1.5/3.4
r_{50}^a	15.6	16.2	14.7	12.8	14.0	10.9	15.4
r^b	15.7	19.5	17.3	11.7	11.1	10.0	15.0

^a r_{50} represents distances corrected for 50% occupancy of the tag conformer close to the binding site.

^b r denotes distances from the nitroxide oxygen of model 1 to various ligand protons as seen in a frame 6ns into the 50 ns MD simulation.

Table 3.

Per-residue decomposition of interaction energies.

Residue	Van der Waals	Electrostatic	Polar Solv.	Non-Polar Solv.	Total
OMe	−0.6	−13.8	13.8	−0.1	−0.8
GlcNS A	−5.8	−90.3	92.8	−0.9	−4.2
IdoA B	−2.5	−103.3	95.8	−0.8	−10.7
GlcNS C	−4.3	−34.0	39.1	−0.6	0.2
IdoA D	−2.0	−51.0	52.2	−0.8	−1.6
O6S-A	−0.4	−34.3	35.6	−0.1	0.8
O6S-C	−0.3	−24.8	26.0	−0.2	0.7
ARG 249	−0.1	−76.0	75.1	−0.0	−0.9
LYS 137	−0.8	−94.3	93.8	−0.1	−1.5
HIP 197	0.2	−107.6	103.9	−0.2	−3.8
LYS 81	−0.4	−110.8	107.1	−0.2	−4.3
ARG 195	−1.0	−119.2	115.0	−0.4	−5.6
HIP 134	−1.0	−127.5	122.3	−0.1	−6.3
ARG 136	−4.7	−111.3	109.3	−0.9	−7.5
ARG 76	−0.7	−134.3	127.6	−0.3	−7.8
ARG 106	−3.2	−161.6	155.3	−0.7	−10.2