Automated Glycan Assembly of $^{19}$F-labeled Glycan Probes Enables High-Throughput NMR Studies of Protein–Glycan Interactions

Giulio Fittolani*, Elena Shanina+, Mónica Guberman, Peter H. Seeberger, Christoph Rademacher, and Martina Delbianco*

Abstract: Protein–glycan interactions mediate important biological processes, including pathogen host invasion and cellular communication. Herein, we showcase an expedite approach that integrates automated glycan assembly (AGA) of $^{19}$F-labeled probes and high-throughput NMR methods, enabling the study of protein–glycan interactions. Synthetic Lewis type 2 antigens were screened against seven glycan binding proteins (GBP), including DC-SIGN and BambiL, respectively involved in HIV-1 and lung infections in immunocompromised patients, confirming the preference for fucosylated glycans (Le$^a$, II type 2, Le$^b$). Previously unknown glycan–lectin weak interactions were detected, and thermodynamic data were obtained. Enzymatic reactions were monitored in real-time, delivering kinetic parameters. These results demonstrate the utility of AGA combined with $^{19}$F NMR for the discovery and characterization of glycan–protein interactions, opening up new perspectives for $^{19}$F-labeled complex glycans.

Introduction

Glycans are a highly diverse class of biomolecules involved in several processes such as cellular communication and recognition and play important structural and modulatory roles.[1] Pathogens invade the host by mimicking or exploiting host glycans present on endothelial cells. This process is often mediated by lectins, a class of glycan-binding proteins (GBP) expressed by both pathogens and hosts. Typically, mammalian glycans have low affinity for mammalian receptors, while showing higher affinity for bacterial lectins.[2] Profiling glycan–lectin interactions is a crucial step towards the understanding of the biological functions of glycans. Still, the extreme complexity and diversity of glycans pose a severe bottleneck to the characterization of these generally weak and promiscuous interactions.

Synthetic glycans are valuable probes to dissect glycan–protein interactions. However, lengthy synthetic protocols hampered their systematic and widespread use in glycobiology. Automated glycan assembly (AGA) enables fast access to complex and well-defined glycans.[3,4] With AGA, glycans are typically assembled in an overnight run, permitting the production of broad collections of glycans for systematic screenings.[5]

An additional challenge to the study of glycan–protein interactions is the need for highly sensitive methods able to detect the often inherently low affinities. Several analytical techniques have been developed to quantitatively describe these interactions at the molecular level and in a high-throughput manner.[6–8] Most of these strategies rely on immobilized glycans (e.g., microarray technology)[6,7] or require large amounts of samples and analysis time (ITC,[8] SPR,[12] or X-ray crystallography[13]). In contrast, NMR allows for the detection of protein–glycan interactions in solution in a fast and reliable manner, providing information on the binding mode in a homogenous assay format in absence of immobilization protocols.[14,15]

NMR active labels are commonly introduced to simplify NMR analysis.[13,17] Among all, the $^{19}$F nucleus stands out due to its unique properties such as: i) high sensitivity to local chemical environment, ii) short acquisition times, iii) simple spectra, iv) broad chemical shift range, and v) absence in...
biological systems (no background signal).[18,19] Even though $^{19}$F NMR has enabled the description of peptide (mis)folding, real-time in events,[18–22] protein–ligand interactions, and high-throughput ligand screening,[23,24] the use of fluorinated glycans to investigate protein binding[25] and enzymatic reactions[26–28] is just at the beginning. The labor-intensive multistep synthesis of $^{19}$F-labeled glycans represents the main bottleneck and has limited these studies to small collections of short and relatively simple glycans.[14,29–32] Still, $^{19}$F-labeled glycans have the potential to dissect protein–glycan interactions.[33,34]

Herein, we present a high-throughput NMR-based approach for the screening and characterization of protein–glycan interactions using $^{19}$F-labeled glycans. AGA enabled quick access to a collection of $^{19}$F-labeled Lewis type 2 complex glycans. Lewis type 2 antigens are involved in several physiological and pathological processes, including cancer, where they act as cell adhesion or recognition mediators.[35,36] Subtle differences in the fucosylation pattern strongly impact their interaction with proteins and ultimately can lead to host immune system elusion.[37–40] The $^{19}$F-labeled glycan probes (hereafter F-glycans) were screened against mammalian and bacterial lectins as well as enzymes. Among mammalian lectins, we selected Langerin[41] and the dendritic cell specific ICAM-3 grabbing non-integrin (DC-SIGN)[42] both of which are known to bind high-mannose N-glycans. DC-SIGN also selectively recognizes specific fucosylated glycans,[43] playing a crucial role in the biology of viral pathogens (e.g. HIV). In addition, we screened soluble lectins produced by some opportunistic pathogens responsible for lung infections, such as Pseudomonas (LecA and LecB)[44] and Burkholderia (BambL)[45] species. Finally, we selected two different sialyltransferases and screened their interactions with Lewis antigens, given the importance and widespread occurrence of terminal sialylation in Lewis antigens.[46–47] The labeled glycan probes in combination with $^{19}$F NMR proved to be valuable for detecting binding events in real-time, identifying new weak protein–glycan interactions, and determining affinities ($K_d$) as well as kinetics of enzymatic reactions.

## Results and Discussion

### Automated Synthesis of F-Glycans

Recently, an elegant procedure to access a collection of Lewis type 2 antigens by AGA was reported.[48] We envisioned a similar approach to produce a set of $^{19}$F-labeled analogs to screen protein binding in a simple $^{19}$F NMR assay. Since the position of the $^{19}$F reporter is thought to be crucial to obtain valuable information, F-glycans (F-Lac, F-nLac, F-Le¹, F-H type 2, and F-Le⁰) were designed with the $^{19}$F reporter in the lactose inner core subunit (Figure 1A). This position is distal from the binding site (i.e. non reducing end) to minimize the effect of the fluorine atom during the binding event.[49,50] We hypothesize that labeling of the inner core glucose unit should maintain sensitivity to the binding event due to overall changes in the correlation time of the glycan in the bound state, reporting changes in the $^{19}$F NMR signal.[50]

$^{19}$F-labeled analogs of Lewis type 2 antigens were assembled on a solid support (functionalized Merrifield resin, L1)

---

Figure 1. Integrated approach for the preparation of $^{19}$F-labeled Lewis type 2 glycans by AGA and screening against lectins and enzymes. A) BBs 1–5, including BB 1 bearing the $^{19}$F reporter, were employed for the AGA of a collection $^{19}$F-labeled Lewis type 2 antigen analogs represented following the Symbol Nomenclature For Glycans (SNFG).[16] B) The F-glycans were screened against proteins, including mammalian and bacterial lectins, as well as enzymes. The enzymes were screened in the absence of donor (i.e. CMP-Neu5Ac) to probe binding to the substrate. The binding strength was defined depending on the changes observed in the NMR after addition of the protein (right panel). Strong binding (blue) is defined as a decrease in peak intensity higher than $-25\%$ or a chemical shift perturbation (CSP) bigger than 0.01 ppm in the $^{19}$F NMR. Weak/ medium binding (light blue) is defined as a decrease in peak intensity higher than $-25\%$ in the CPMG-filtered $^{19}$F NMR. No binding (white) is defined as a decrease in peak intensity lower than $-25\%$ in CPMG-filtered $^{19}$F NMR.
using building blocks (BBs) 1–5 (Figure 1A). The BBs are equipped with a thioether or a dibutylphosphate reactive leaving group. Orthogonal cleavage of the 9-fluorenylethoxycarbonyl (Fmoc) and levulinoyl (Lev) temporary protecting groups permits regioselective chain elongation. Benzyl (Bn), benzoyl (Bz), and N-trichloroacetyl (TCA) groups protect the remaining functionalities. β-Stereoselectivity during glycosylation with BBs 1–4 is ensured by anchimeric assistance of the protecting groups at C-2, while α-stereoselectivity with BB 5 was verified in previous studies.[48] BB 1 is labeled with the 19F reporter at the C-3 position.[51] Each oligosaccharide was assembled in an overnight run following previously reported conditions for unlabeled analogs (see SI).[48] Post-AGA manipulations included solid-phase methanalysis,[51] photo cleavage[52] from the solid support, and hydrolysis (see SI). A single final purification step afforded the target F-glycans in overall yields of 5% to 16% over 7 to 15 steps.

19F NMR Screening of F-Glycan Library

A 19F and CPMG NMR screening was performed to probe the interactions of five F-glycans (F-Lac, F-n-Lac, F-Le¹, F-H type 2, and F-Le') with mammalian (Langerin, DC-SIGN) and bacterial (LecA, LecB, BambL) lectins and enzymes (α(2,3)-sialyltransferase from Pasteurella multocida (Pm23ST)) and α(2,6)-sialyltransferase from Photobacterium damselae (Pda26ST)). Upon protein binding, the molecular tumbling rate of the glycan is drastically affected resulting in a decrease of the 19F signal intensity.[50] Monitoring 19F chemical shift perturbation (CSP) or change in peak intensity upon addition of protein allowed us to qualitatively evaluate the strength of the interaction. A decrease in peak intensity or a CSP in 19F NMR indicates strong binding. Application of a CPMG-based spin echo filter allows us to detect weak binders. As a result, bacterial (LecA, LecB, and BambL) and mammalian (DC-SIGN ECD) lectins preferred fucosylated glycans (Figures S2A, S2B, S2C, and S2E). No binding to F-glycans was observed in presence of Langerin ECD (Figure S2D), in agreement with previous reports.[53] In contrast, the enzymes showed much weaker interactions and a slight preference for shorter non-branched glycans (Figure S3).

Reporter Position on F-Glycans Does Not Affect Binding to Mammalian and Bacterial Lectins

DC-SIGN recognizes cellular ligands and pathogens that express Lewis antigens. In particular, Le¹ and Le' present on Schistosoma mansoni[54] and Helicobacter pylori[41] or endothelial cells,[55] respectively, are known binding partners for DC-SIGN.[56] The strong preference of DC-SIGN for fucosylated ligands has also been elucidated with the crystal structure of the carbohydrate-binding site of DC-SIGN bound to Le¹.[59] The qualitative CPMG NMR screening of mammalian lectins confirmed the interaction of DC-SIGN with fucosylated glycans F-Le¹, F-H type 2, and F-Le' (Figure 2A), as indicated by changes in the NMR peak intensity of the reporter molecule. This effect is maximized with a protein-to-ligand ratio of 2:1 (Figure S4A).

First, we explored the role of the 19F reporter in F-glycan binding to DC-SIGN. We performed protein-observed 15N HSQC NMR and recorded an HSQC NMR spectrum of DC-SIGN CRD in the presence of F-Le¹ and Le'. Both ligands promoted similar changes in the backbone of DC-SIGN CRD (Figure 2B and S4B). Next, we investigated the effect of the reporter’s position on the ability to reveal binding events. We conjugated a CF₃ moiety to the remote end of the aminopentyl linker on H type 2 (CF₃-H type 2), far from the carbohydrate-binding site, and tested the new ligand in 19F and CPMG NMR. Remarkably, its binding was observed with both mammalian (DC-SIGN, Figure 2A) and bacterial lectins (BambL, Figure S5). These results indicate that the positioning of the 19F reporter on the Glc unit does not affect the binding of F-glycans with proteins. Furthermore, the 19F reporter can be remote to the glycan binding site to avoid any interference with the binding event, while preserving excellent sensitivity. However, the functionalization of the amino linker with a CF₃ moiety prevents any further conjugation of the glycan (e.g. to protein, surface, liposome).

We further investigated the interactions of DC-SIGN CRD with F-Le¹ and F-H type 2 in 15N HSQC NMR (Figure 3A and S6A). Even though Le¹ is known for its interaction with DC-SIGN, structural data are lacking.[57]
Both ligands promoted CSPs of the residues located in the carbohydrate-binding and remote sites of DC-SIGN CRD. Binding to F-Le6 promoted larger changes in DC-SIGN CRD than F-H type 2 or the monosaccharide positive control α-mannose (Figure 3B and S6B). This result proved that the avidity effect plays a crucial role in the interactions between DC-SIGN and Lewis type 2 antigens, as similarly noted for high-mannose structures. The CSPs observed in remote parts of the protein suggest allosteric binding, a known mechanism for C-type lectins such as DC-SIGN. Cumulatively, we believe these probes are valuable tools for the description of interaction mechanisms between DC-SIGN and fucosylated blood antigens.

**Binding Affinity of F-Glycans to Bacterial Lectins**

Bacterial lectins show a remarkably high affinity for fucosylated blood group antigens. The interaction of BambL from Burkholderia ambifaria with H type 2 has been thoroughly investigated and two binding sites were identified in a crystal structure of the complex (Figure 4A). We set on to verify this interaction for F-glycans in 19F NMR and protein-observed NMR.

First, we performed 19F NMR screening and titration experiments with fucosylated F-glycans. 19F NMR experiments allowed us to confirm the interaction and obtain affinity constants for F-H type 2 (Kd = 9 ± 2 μM, Figure 4B and 3C) and F-Le6 (Kd = 14 ± 2 μM, Figure S7A). Given that BambL has two binding sites available for glycan binding, we applied one- and two-binding site models to derive the affinities for both sites. Both models resulted in matching Kd values, in agreement with values reported by ITC. Even though we did not observe a difference in the affinities between the two sites in 19F NMR, we showed that 19F NMR can be applied reliably to derive affinities while considerably reducing the amount of ligand needed for ITC.

We verified the interaction of F-H type 2 (Figure 4D) and F-Le6 with 14N-labeled BambL in protein-observed 15N TROSY NMR. Changes in protein backbone similar to the one obtained with α-2,6-fucose indicate that the α-2,6-fucose branch was mainly responsible for the binding (Figures S7C and S4D). To derive affinities, we titrated both ligands and followed the changes in peak intensities and CSPs for the peaks in slow (F-H type 2: Kd = 12 ± 8 μM, Figure 4F and F-Le6: Kd = 17 ± 3 μM, Figure S7D), and fast (F-H type 2: Kd = 94 ± 33 μM, Figure 4G and F-Le6: Kd = 245 ± 29 μM, Figure S7E) exchange regimes, respectively. However, protein-observed NMR is not well suitable for the determination of Kd for ligands with high affinities and thus, it hampered the accurate derivation of the Kd. This underscores the advantage of the 19F NMR ligand-observed approach.

In addition to the known strong interactions of LecB and BambL, with fucosylated glycans, CPMG NMR screening revealed weak interactions between LecA and fucosylated F-glycans. To confirm this observation, we performed 19F R-filtered, protein-observed 19F (PrOF) and 15N TROSY NMR experiments. F-H type 2 showed a faster relaxation in presence of protein, indicating a weak interaction with LecA (Figure S8B). Protein-observed NMR experiments with 5-fluorotryptophan (5FW, Figure S8C) and 15N-labeled LecA (Figure S8D and S8E) confirmed that this interaction takes place in the canonical carbohydrate-binding site of LecA, as indicated by perturbation of W42 and CSPs promoted in a similar manner to α-galactose, respectively. To the best of our knowledge this is the first report of such weak binding detected using a biophysical method. These results demonstrate that F-glycans serve as probes for the affinity determination and discovery of new interactions using low amounts of protein and ligand.

**Enzyme Binding and Real-Time Kinetics with F-Glycans**

The 19F NMR assay allowed us to monitor the binding of F-glycans (F-Lac and F-nLac) to enzymes. Two sialyltransferases (Pmto23ST[53] and Pdo26ST[54]) were screened in the absence of donor (i.e. CMP-Neu5Ac) and revealed weak binding to the glycan substrate (Figure 1B and S3). This is particularly relevant because binding sites of transferases usually have a very low affinity for the acceptors, making these interactions difficult to detect. Shorter non-branched glycans (F-Lac and F-nLac) showed stronger binding than longer branched ones. F-Le6 did not show any binding with Pmto23ST or Pdo26ST, matching its known poor reactivity as acceptor (Figure S3). In contrast, Pdo26ST showed weak binding to the carbohydrate-binding site of DC-SIGN (unassigned resonances in dashed line) and CSP plot showing that F-Le6-perturbed resonances similarly to α-mannose (red, positive control). The magnitude of F-Le6-promoted CSPs is higher compared to α-mannose. CSPs exceeding the threshold (dashed line at 0.005 ppm) and intensities decreasing by more than 50% were used for mapping the binding site of F-Le6 on a structure of DC-SIGN CRD.
Neu5Ac induced a chemical shift perturbation of 0.2 ppm on the $^{19}$F-labeled acceptor, allowing to track in real-time the enzymatic sialylation process (Figure 5B). When the $^{19}$F reporter was positioned remotely to the reactive site of the acceptor (>3 sugar units away, F-nLac), no chemical shift perturbation was noticed, despite the success of the enzymatic transformation (Figure S10). Thus, in contrast to what is observed for protein binding, the position of the $^{19}$F reporter is key for monitoring enzymatic reactions.

**Conclusion**

AGA enabled the fast assembly of $^{19}$F-labeled Lewis type 2 antigens for the high-throughput screening of protein binding. Mammalian and bacterial lectins as well as enzymes were analyzed. $^{19}$F NMR screening of F-glycans permitted a quick qualitative evaluation as well as a reliable quantification of lectin binding ($K_s$). The assay does not require labeled proteins or complex 2D NMR experiments. All NMR experiments can be performed in an extremely small scale (few nmol of glycan and protein per experiment). Enzymatic reactions, including sialylation, were monitored in real-time, demonstrating that $^{19}$F-labeled glycans hold a great potential as molecular probes to uncover enzymatic processes and for high-throughput screening. Protocols for the selective $^{19}$F-labeling of monosaccharides are available; the implementation of these novel BBs in AGA will fuel the production of new classes of glycan probes. Given the high dispersion of $^{19}$F NMR signals, libraries of F-glycans with diverse chemical shifts can be designed to increase the high throughput of this approach. The ability of $^{19}$F glycan probes to reveal binding or enzymatic transformation in solution and in real-time could open the way to in cell NMR applications, often hampered by high background signals. Overall, these probes are valuable tools for a better molecular understanding of the interactions of complex glycans with protein receptors.
Figure 5. Real-time enzyme kinetics by $^{19}$F NMR using F-glycans. A) $^{19}$F NMR of F-Lac incubated with β-galactosidase. $^{19}$F NMR real-time tracking of product formation (black arrows) upon incubation of F-Lac with β-galactosidase (right). Kinetic data were derived plotting the product formation rate as a function of the substrate concentration. The best fit of the experimental data provides a $K_m$ value of 86.5 ± 10.5 μM according to the Henry-Michaelis–Menten equation (left). B) $^{19}$F NMR of F-Lac incubated with Pm23ST in presence of CMP-Neu5Ac. The formation of F-sLac (black arrows) can be followed by $^{19}$F NMR in real-time. Product formation was confirmed by HPLC (Figure S9).

Acknowledgements

We thank the Max Planck Society for generous financial support and Olaf Niemeyer for technical assistance with the NMR. M.D. and G.F. thank the MPG-FhG Cooperation Project Glyco3Dysplay and C.R. and E.S. thank the ANR/DFG French-German project “Glycomime” (RA1944/7-1). Open access funding enabled and organized by Projekt DEAL.

Conflict of interest

The authors declare no conflict of interest.

Keywords: $^{19}$F NMR spectroscopy · automated glycan assembly · glycans · lectins · protein–glycan interactions


Manuscript received: February 22, 2021
Revised manuscript received: March 19, 2021
Accepted manuscript online: March 30, 2021
Version of record online: May 7, 2021