frontiers | Research Topics

Preparation and characterization of glycosylated biomolecules

Edited by
Zhongping Tan, Jian Yin and Xuefei Huang

Published in
Frontiers in Chemistry

ISSN 1664-8714
ISBN 978-2-83251-283-8
DOI 10.3389/978-2-83251-283-8

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Preparation and characterization of glycosylated biomolecules

Topic editors

Zhongping Tan — Chinese Academy of Medical Sciences and Peking Union Medical College, China

Jian Yin — Jiangnan University, China

Xuefei Huang — Michigan State University, United States

Citation

Tan, Z., Yin, J., Huang, X., eds. (2023). *Preparation and characterization of glycosylated biomolecules*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-83251-283-8

Table of contents

frontiers in Chemistry

ORIGINAL RESEARCH
published: 26 October 2021
doi: 10.3389/fchem.2021.766932

Probing Multivalent Carbohydrate-Protein Interactions With On-Chip Synthesized Glycopeptides Using Different Functionalized Surfaces

Alexandra Tsouka[1,2], *Kassandra Hoetzel*[1], *Marco Mende*[1], *Jasmin Heidepriem*[1,2], *Grigori Paris*[1,3], *Stephan Eickelmann*[1], *Peter H. Seeberger*[1,2], *Bernd Lepenies*[4] *and Felix F. Loeffler*[1]*

[1]Department of Biomolecular Systems, Max Planck Institute of Colloids and Interfaces, Potsdam, Germany, [2]Institute of Chemistry and Biochemistry, Freie Universität Berlin, Berlin, Germany, [3]Department of System Dynamics and Friction Physics, Institute of Mechanics, Technical University of Berlin, Berlin, Germany, [4]Institute for Immunology and Research Center for Emerging Infections and Zoonoses, University of Veterinary Medicine Hannover, Hannover, Germany

OPEN ACCESS

Edited by:
Yuan Guo,
University of Leeds, United Kingdom

Reviewed by:
Matthew Robert Pratt,
University of Southern California, United States
Rongsheng (Ross) Wang,
Temple University, United States

***Correspondence:**
Felix F. Loeffler
felix.loeffler@mpikg.mpg.de

Specialty section:
This article was submitted to Chemical Biology, a section of the journal Frontiers in Chemistry

Received: 30 August 2021
Accepted: 11 October 2021
Published: 26 October 2021

Citation:
Tsouka A, Hoetzel K, Mende M, Heidepriem J, Paris G, Eickelmann S, Seeberger PH, Lepenies B and Loeffler FF (2021) Probing Multivalent Carbohydrate-Protein Interactions With On-Chip Synthesized Glycopeptides Using Different Functionalized Surfaces. Front. Chem. 9:766932. doi: 10.3389/fchem.2021.766932

Multivalent ligand–protein interactions are a commonly employed approach by nature in many biological processes. Single glycan–protein interactions are often weak, but their affinity and specificity can be drastically enhanced by engaging multiple binding sites. Microarray technology allows for quick, parallel screening of such interactions. Yet, current glycan microarray methodologies usually neglect defined multivalent presentation. Our laser-based array technology allows for a flexible, cost-efficient, and rapid *in situ* chemical synthesis of peptide scaffolds directly on functionalized glass slides. Using copper(I)-catalyzed azide–alkyne cycloaddition, different monomer sugar azides were attached to the scaffolds, resulting in spatially defined multivalent glycopeptides on the solid support. Studying their interaction with several different lectins showed that not only the spatially defined sugar presentation, but also the surface functionalization and wettability, as well as accessibility and flexibility, play an essential role in such interactions. Therefore, different commercially available functionalized glass slides were equipped with a polyethylene glycol (PEG) linker to demonstrate its effect on glycan–lectin interactions. Moreover, different monomer sugar azides with and without an additional PEG-spacer were attached to the peptide scaffold to increase flexibility and thereby improve binding affinity. A variety of fluorescently labeled lectins were probed, indicating that different lectin–glycan pairs require different surface functionalization and spacers for enhanced binding. This approach allows for rapid screening and evaluation of spacing-, density-, ligand and surface-dependent parameters, to find optimal lectin binders.

Keywords: glycopeptides, glycan binding proteins, lectin—carbohydrate interaction, multivalency, surface functionalization

INTRODUCTION

Glycan-protein interactions exist in many biological processes, such as protein folding, cell-cell interaction, cell-adhesion, and signaling. Thus, their understanding is of fundamental importance (Varki, 2009). Glycan arrays are considered versatile tools for high-throughput screening of such interactions. Immobilization of glycans on solid support by high-precision robotics can be achieved in multiple ways, (Geissner et al., 2014; Geissner et al., 2019; O'Neil et al., 2018; Purohit et al., 2018; Mende et al., 2019) becoming nowadays a dominant methodology for detection of novel interactions in immunological and biomedical research (Varki, 2009; Gao et al., 2019), as well as drug discovery (Geissner and Seeberger, 2016; Rademacher et al., 2019; Tikhonov et al., 2020).

Glycans play a key role in diseases and virulence (*e.g.,* diabetes, inflammation, cancer, infections), rendering scientists to investigate their structural and functional characteristics (Zhou and Cobb, 2021). Their interaction with other cells, and their recognition by glycan binding proteins (GBPs), so called lectins, triggered the investigation of their binding ability, and molecular mechanism (Raman et al., 2016; Valverde et al., 2020). Individual interactions between glycans and their GBPs are relatively weak (*e.g.*, K_d values ≈ μM–mM range). The recognition process that nature has evolved to enhance the binding strength and specificity is called multivalency. This effect enables high binding affinities *via* simultaneous recognition of one or several glycans by GBPs, which have multiple and spatially well-defined glycan binding sites (Fasting et al., 2012; Haag, 2015). For a strong multivalent interaction, not only the type(s) of sugar(s), but also their spatial orientation, their accessibility, and the carrier scaffold are important, to achieve optimum distance with the binding pockets of the multivalent receptor.

Despite the importance of multivalency, it is often neglected on the solid support, since the density and the spacing between sugar moieties is difficult to be determined. Therefore, various studies in the last years focused on the optimum glycan presentation, concentration, flexibility, orientation, and density in the array format (Oyelaran et al., 2009; Müller et al., 2016; Kim et al., 2018; Mende et al., 2019; Valles et al., 2019; Di Maio et al., 2021). In addition, a plethora of multivalent glycan scaffolds have been investigated with diverse size and shape to mimic the natural recognition (Cecioni et al., 2015; Delbianco et al., 2016; Redman and Krauss, 2021). Peptide scaffolds have been widely studied due to their simple synthesis *via* solid phase peptide synthesis, (Merrifield, 1963) offering well-defined monodisperse structures. Introduction of sugars on the peptide moieties can be employed using glycosylated amino acids or, in a concerted fashion onto unnatural, azido modified amino acids for specific conjugation (e.g., Click chemistry or Staudinger Ligation) (Specker and Wittmann, 2006; Freichel et al., 2017; Hill et al., 2018; Camaleño de la Calle et al., 2019).

Yet, the application of this approach in the microarray format remains challenging. Fabrication of natural glycoproteins, (Kilcoyne et al., 2012) neoglycopeptides, (Wang et al., 2002) glycodendrimers, (Laigre et al., 2018) DNA-based glycoconjugates, (Hawkes et al., 2019) glycoclusters, (Moni et al., 2009) and glycopolymers (Godula et al., 2009; Zilio et al., 2015) in the microarray format with multivalent presentation require extensive synthetic work prior to the printing onto glass slides. Unfortunately, printing of these compounds on the microarray ties in with solubility and density fluctuations of the material, printing and humidity inconsistencies during coupling, and the microarray surface functionalization (linker) effect, resulting in insufficient coupling and/or poor morphology of the spotted material (Ruprecht et al., 2019; Temme et al., 2019).

Herein, we report our progress in and deeper understanding of our laser-based method for *in-situ* generation of multivalent glycopeptides in the microarray format with controlled glycan spacing and density (Mende et al., 2020). We expanded our technology, making it compatible with different commercially available microarray surfaces, to probe previously inaccessible glycan interactions. Therefore, we first optimized the synthesis on each microarray surface type and we equipped them with an additional linker to investigate its effect on lectin binding. We demonstrate the importance of surface accessibility and wettability on glycan-GBP interactions, enabling us to study a wide range of plant lectins in a high-throughput manner.

MATERIALS AND METHODS

Donor Slide Preparation

Microscope glass slides (Marienfeld Superior, Germany; size 76 × 26 × 1 mm, ground edges, pure white glass) were covered on one side with self-adhesive polyimide foil (Kapton, DuPont, United States, CMC Klebetechnik GmbH, Germany; thickness of polyimide layer approximately 25 μm, thickness of glue layer approximately 45 μm). A thin layer of the transfer material was placed on top of the polyimide foil by spin coating (80 rps, Schaefer Technologie GmbH, Germany; KLM Spin-Coater SCC-200). Two different spin coating solutions were prepared. Pentafluorophenyl (OPfp)-activated 9-fluorenylmethoxycarbonyl (Fmoc) protected L-glycine, (Fmoc-Gly-OPfp) **1** (3.00 mg), was pre-dissolved in dimethylformamide (DMF) (50 μL), while inert polymer matrix (27 mg) (SLEC PLT 7552, Sekisui Chemical GmbH, Germany) was dissolved in dichloromethane (DCM) (450 μL), resulting in the final spin coating solution (500 μL). The non-activated amino acid, Fmoc-propargyl-glycine (Fmoc-Pra-OH) (3 mg) was pre-dissolved in DMF (50 μL), followed by addition of *N,N′*-diisopropylcarbodiimide (DIC) (1.4 μL) and pentafluorophenol (PfpOH) (1.7 mg) consecutively, while the inert polymer matrix (27 mg) was pre-dissolved in DCM (450 μL), forming the desired Fmoc-Pra-OPfp **2** *in situ* (see **Supplementary Material**).

Acceptor Slide Preparation

Fmoc-NH-β-Ala-PEGMA-co-MMA glass slides (~20 nm thick coating, loading of functional groups according to vendor 1 nmol cm^{-2}, estimated functional group spacing of 7–10 nm) were acquired from PEPperPRINT GmbH (Germany) and the 3D-Amino glass slides (according to vendor 1–5 nmol cm^{-2}) from PolyAn GmbH (Germany). On PolyAn and PEPperPRINT slides,

a hydrophilic PEG ($(EG)_3$) -based spacer (≈17 Å length) was attached (see **Supplementary Material**, Section 3.2), before the synthesis of the desired tetrapeptides. In a variant process, PolyAn slides without PEG-spacer were used directly, without prior spacer functionalization.

Laser Transfer Parameters

For the array synthesis, a spot pitch of 250 μm was used. A laser scanning system with 488 nm wavelength and 120 mW maximum output power was used (Mende et al., 2020), with a laser focus diameter of ~20 μm. *PEPperPRINT slides:* A laser power of 80 mW and a pulse duration of 6 ms per spot was applied. *PolyAn slides:* A laser power of 60 mW with a pulse duration of 6 ms was applied. The final spot diameter was about 150 μm.

General Laser-Based Synthesis Process and Synthesis of Tetrapeptide Scaffolds

General laser-based synthesis process: The laser transfer and peptide synthesis were conducted as reported previously (Eickelmann et al., 2019; Mende et al., 2020; Paris et al., 2020). The process begins with the preparation of different donor slides (*Donor Slide Preparation*) that are easily prepared by spin-coating a solution of polymer matrix and activated amino acid building block onto polyimide foil (Kapton) bearing glass slides. The polymer and amino acid mixture forms an about 200 nm thin layer on the polyimide. For the patterning process, an amino acid containing donor slide is placed on top of an acceptor slide (*Acceptor Slide Preparation*) and a focused laser (*Laser Transfer Parameters*) transfers solid polymer material spotwise from the donor to the acceptor (one pulse of 6 ms transfers one spot). The laser is absorbed by the polyimide foil, which heats up and expands. Eventually, the expanding polyimide contacts the acceptor slide, causing the transfer of nanometer thin and about 150 μm wide polymer material spots. The transfer is repeated with different donor slides until the desired amino acid pattern is completed. Afterwards, the acceptor slide is placed into an oven at 95°C under nitrogen for several minutes to initiate the coupling reaction. In the oven, the polymer spots "melt" while retaining their shape, enabling the reaction of the building blocks according to the transferred pattern. The activated amino acid building blocks couple to the amino groups on the acceptor slide. Next, the acceptor slide is washed, removing unreacted amino acids and residual polymer. Each amino acid coupling step is repeated three times to increase the coupling yield and to minimize deletion sequences. Then, remaining free amino groups on the acceptor surface are acetylated and the Fmoc protecting groups are removed before the next synthesis cycle. Peptide synthesis steps are repeated, until the final peptide length is reached.

Synthesis of tetrapeptide scaffolds: Commercially available slides from PEPperPRINT or PolyAn were used. Before the synthesis of the tetrapeptides, a PEG-based spacer was attached if not indicated otherwise, (see **Supplementary Material**). PEPperPRINT slides require a spacer due to the high protein resistance of the surface. The first layer of OPfp-activated and Fmoc-protected amino acids was transferred *via* laser transfer, using two different donor slides sequentially to create the desired combinatorial pattern on the acceptor slide. The coupling reaction was accomplished under heat in an oven under nitrogen atmosphere at 95°C for 10 min. Subsequently, the slides were washed with acetone twice, initially for 2 min in an ultrasonic bath, and then for another 2 min in a petri dish on a shaker (450 rpm). Then, slides were dried in a jet of air. The laser transfer of the same amino acid pattern, the coupling, and the acetone washing steps were repeated twice, to increase the coupling efficiency. Each time, a new donor slide was used for every transfer and coupling cycle. Free unreacted amino groups on the slides were acetylated with a capping solution twice for 30 min (see **Supplementary Material**). The slides were washed with DMF (3 × 5 min), methanol (MeOH) (1 × 2 min), DCM (1 × 1 min), and dried in a jet of air. Deprotection of the terminal Fmoc-groups was achieved for 20 min with Piperidine (see **Supplementary Material**) on a shaker (450 rpm). The slides were washed with DMF (3 × 5 min), MeOH (1 × 2 min), DCM (1 × 1 min), consecutively, and dried in a jet of air. The whole process was repeated, as needed, for each pattern to synthesize the desired peptides. In the case of terminal amino acids within the peptide chain, the Fmoc removal was accomplished before the acetylation step, capping of the free amino groups.

Sugar Azides

Each sugar azide **3–7** was obtained according to known literature procedures (see **Supplementary Material**, Section 2.1). Two sugar azides, **8** and **9**, were obtained from *Conju-Probe*.

Copper (I)-Catalyzed Alkyne-Azide Cycloaddition (CuAAC)

$CuSO_4$ (530 μg, 3.36 μmol, 2.00 equiv) was dissolved in a mixture of dimethyl sulfoxide (DMSO) and water (1:1, 200 μL). Sodium ascorbate (998 μg, 5.04 μmol, 3.00 equiv) was added and the mixture was thoroughly vortexed. The precipitate was centrifuged for 1 min. The remaining solution was passed through a polypropylene syringe filter (0.2 μm polypropylene filter media with polypropylene housing, 25 mm diameter, Whatman, Global Life Sciences Solutions Operations United Kingdom). The sugar azide (1.68 μmol, 1.00 equiv) was dissolved in this solution and then applied on the acceptor surface (c = 8.4 μmol/ml). For the incubation, a 16-well format incubation chamber was used. The prepared solution (200 μL) was poured in one of the wells and then shaken overnight in the dark. The next day, the slide was washed with water three times for 5 min inside the well and one time for 30 min in a petri dish on a shaker (450 rpm). Finally, the slide was dried in a jet of air.

Plant Lectin Assay

To avoid unspecific binding, the acceptor slides were incubated with a blocking buffer for 40 min (Rockland, United States; blocking buffer for fluorescent western blotting MB-070). Fluorescently labeled plant lectins, *concanavalin A* (*i.e.*, ConA; CF®633 ConA, Biotium, Inc., United States) was diluted to

100 µg/ml in lectin buffer (50 mM HEPES, 100 mM NaCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$, 10% blocking buffer, 0.05% Tween 20, pH 7.5), *ricinus communis* agglutinin I, (RCA-I), *peanut* agglutinin (PNA), *soybean* agglutinin (SBA), *dolichos biflorus* agglutinin (DBA), and *wheat germ* agglutinin (WGA) (Rhodamine labeled, Lectin kit 1, Vector laboratories, United States) were diluted to 10 µg/ml in lectin buffer and incubated for 1 h at room temperature. Subsequently, each stained well was washed with PBS-T (3 × 5 min). Then, the acceptor slide was rinsed with Tris buffer (1 mM Tris-HCl buffer, pH = 7.4) to remove all the remaining salt residues, and dried in a jet of air. Fluorescence scanning was used to detect the lectin binding on the corresponding sugar moieties.

Fluorescence Scan

All fluorescence scans were carried out in a high-resolution microarray GenePix 4000B scanner. CF® ConA labeled glycopeptides were screened with an excitation wavelength of 635 nm and PMT gain of 600. Rhodamine RCA-I, PNA, SBA, DBA, WGA labeled glycopeptides were scanned at an excitation wavelength of 532 nm and PMT gain of 500. Carboxytetramethylrhodamine (TAMRA) labeled tetrapeptides were detected at an excitation wavelength of 532 nm and PMT gain of 400. The laser power was always set to 33% and the pixel resolution to 5 µm. For the analysis of the fluorescence images, the analysis software GenePix Pro 6.0 (Molecular Devices, Sunnyvale/California, United States) was used.

Analysis of Glycopeptides Regarding Multivalency Effects

For each sugar azide, the reaction was performed in a separate cavity of a 16-well format incubation chamber (PEPperPRINT GmbH, Germany). Each well contained three sets of quadruplicates of the same single sugar azide and tetrapeptide, giving twelve glycopeptide replicas of each synthesized structure. The median of the fluorescence intensity of the scanned area was determined with the microarray analysis software GenePix Pro 6.0. For the analysis, the mean value of the twelve spot medians was calculated. Spots (*i.e.,* outlier/artifacts) with more than 40% standard deviation from the mean were excluded from calculations.

RESULTS AND DISCUSSION

We applied our laser transfer technology to generate peptide scaffolds directly in the array format.(Loeffler et al., 2016; Eickelmann et al., 2019; Mende et al., 2020) Therefore, different donor slides were prepared, containing alkyne-functionalized L-propargylglycine (Pra) or L–Glycine (Gly) amino acid building blocks. These donor slides were placed on top of a functionalized acceptor slide and a laser precisely transferred the building blocks in desired patterns. Next, the amino acid pattern was coupled in an oven to the acceptor slide, the surface was washed, capped and Fmoc deprotected. Repeating these *in-situ* solid phase synthesis steps, peptides were generated in the array format on the acceptor. Finally, copper (I)-catalyzed alkyne-azide cycloaddition (CuAAC) was used to attach different azido-functionalized glycan monomers to the alkyne groups of the peptide scaffolds (**Figure 1**).

For our work, we aimed to employ different commercially available amine functionalized acceptor slides from different suppliers to determine the influence of the surface functionalization on glycan binding events. Hydrophobicity and sterical hindrance of the acceptor surface functionalization may lead to lower accessibility of the glycans. Therefore, we had to find new process conditions for the synthesis of the peptides on the different functionalized slides. Then, we studied the interactions of the fluorescently labeled lectins on these substrates and analyzed with fluorescence scanning. To compare high and low- affinity glycan-GBP interactions, we chose to probe the plant lectins *concanavalin A* (ConA; tetramer), *ricinus communis* agglutinin I (RCA-I; tetramer with only two Gal-specific subunits) (Wittmann and Pieters, 2013)), *peanut* agglutinin (PNA; tetramer), *soybean* agglutinin (SBA; tetramer), *dolichos biflorus* agglutinin (DBA; tetramer), and *wheat germ* agglutinin (WGA; dimer) with their corresponding glycans under the same conditions. Furthermore, we screened the CLR-Fc fused C-type lectins mLangerin, mMincle, and mMGL-1 (Maglinao et al., 2014; Artigas et al., 2017; Mayer et al., 2017; Valverde et al., 2020). However, since we did not observe any binding of these three lectins, details are only discussed in the **Supplementary Material** (Section 8).

Synthesis of Glycopeptides

All sixteen possible variants of the peptide tetramers, containing the two derivatives Fmoc-Gly-OPfp **1** and Fmoc-Pra-OPfp **2**, were synthesized in the microarray format (**Figure 2**). Amine functionalized glass slides from PEPperPRINT (PPP) were used with prior functionalization with a PEG-based spacer (Stadler et al., 2008). 3D-amino glass slides from PolyAn were either used with or without prior PEG-spacer functionalization. Before the synthesis, we optimized the transfer and coupling conditions for each solid support (see **Supplementary Material**, Section 4). Subsequently, a pre-patterning of all acceptor slides was performed with two glycines **1**, to further increase the distance between the tetrapeptides and the solid support and, thereby, the accessibility of the glycopeptides. After Fmoc deprotection of the N-terminus, the free amino groups were used for peptide synthesis. Two donor slides were employed to synthesize the sixteen tetrapeptide combinations, Fmoc-Gly-OPfp **1** (**G**) and Fmoc-Pra-OPfp **2** (**B**) **Figure 2A** (conventional synthesis from C-terminus to N-terminus, *e.g.,* N-GBGB-C, 1VII). Coupling and laser transfer of each amino acid layer was repeated three times to achieve high coupling efficiency and prevent deletion sequences while growing the chains. Coupling of the amino acids was conducted in an oven under nitrogen gas atmosphere at 95°C, resulting in three sets of quadruplicates on one array (n = 12 spots; binding intensity is calculated as the mean of the 12 spot replica) (**Figure 2B**). Quality control of the three synthesized arrays was carried out *via* clicking a TAMRA azide dye to the scaffolds and analyzing the fluorescence intensity. On the PEPperPRINT slides, a

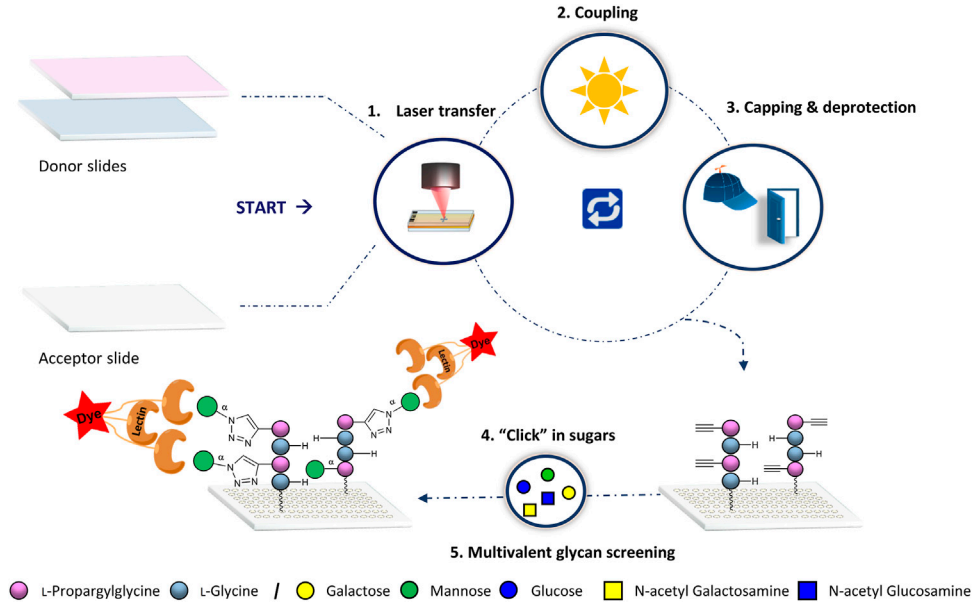


FIGURE 1 | General synthetic approach for glycopeptide generation. Synthesis consists of the 1) laser transfer process, 2) coupling of the amino acids on defined spots, 3) capping to mask unreacted amine groups with subsequent deprotection of the Fmoc protecting group for the next coupling step, and 4) attachment of the sugar azides on the peptide scaffolds by CuAAC. Finally, 5) binding of lectins is screened for multivalent interactions.

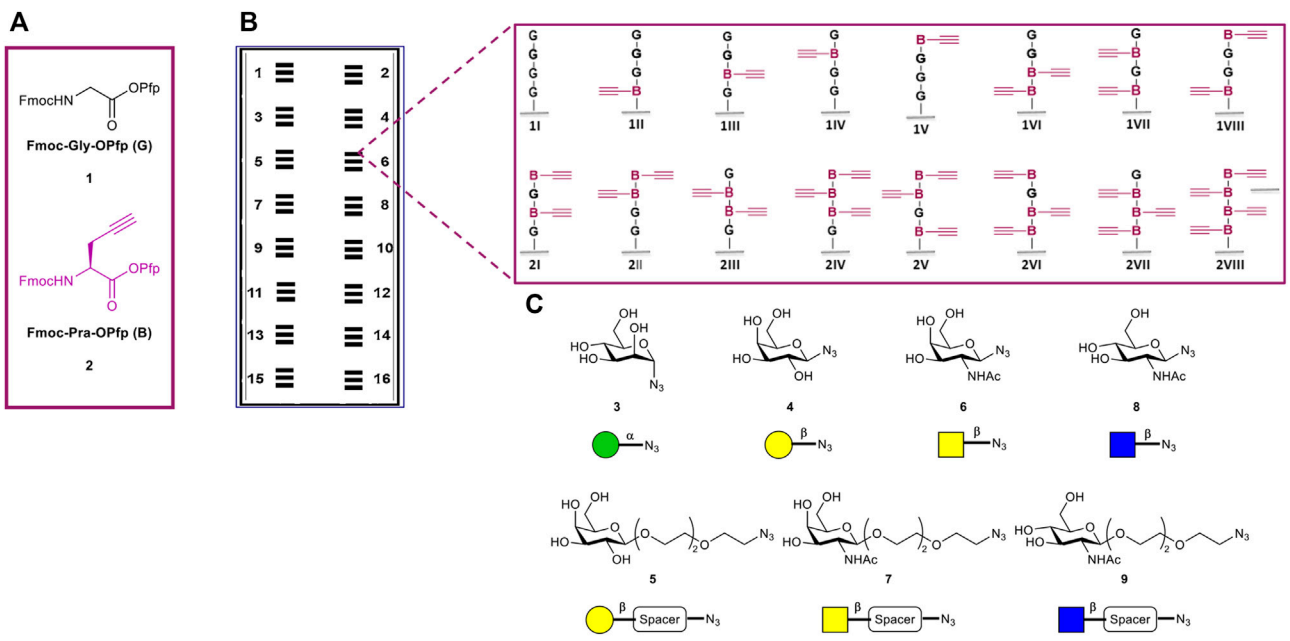


FIGURE 2 | Overview of tetramer peptide scaffolds and sugar azides. **(A)** The two amino acids used for the synthesis of the desired scaffolds. **(B)** Representation of the 16 synthesized microarrays per glass slide, each array containing three copies of the sixteen different tetramers in quadruplicate spots ($n = 3 \times 4 = 12$ spots of each structure per microarray). **(C)** Collection of sugar azides for the generation of the glycopeptides.

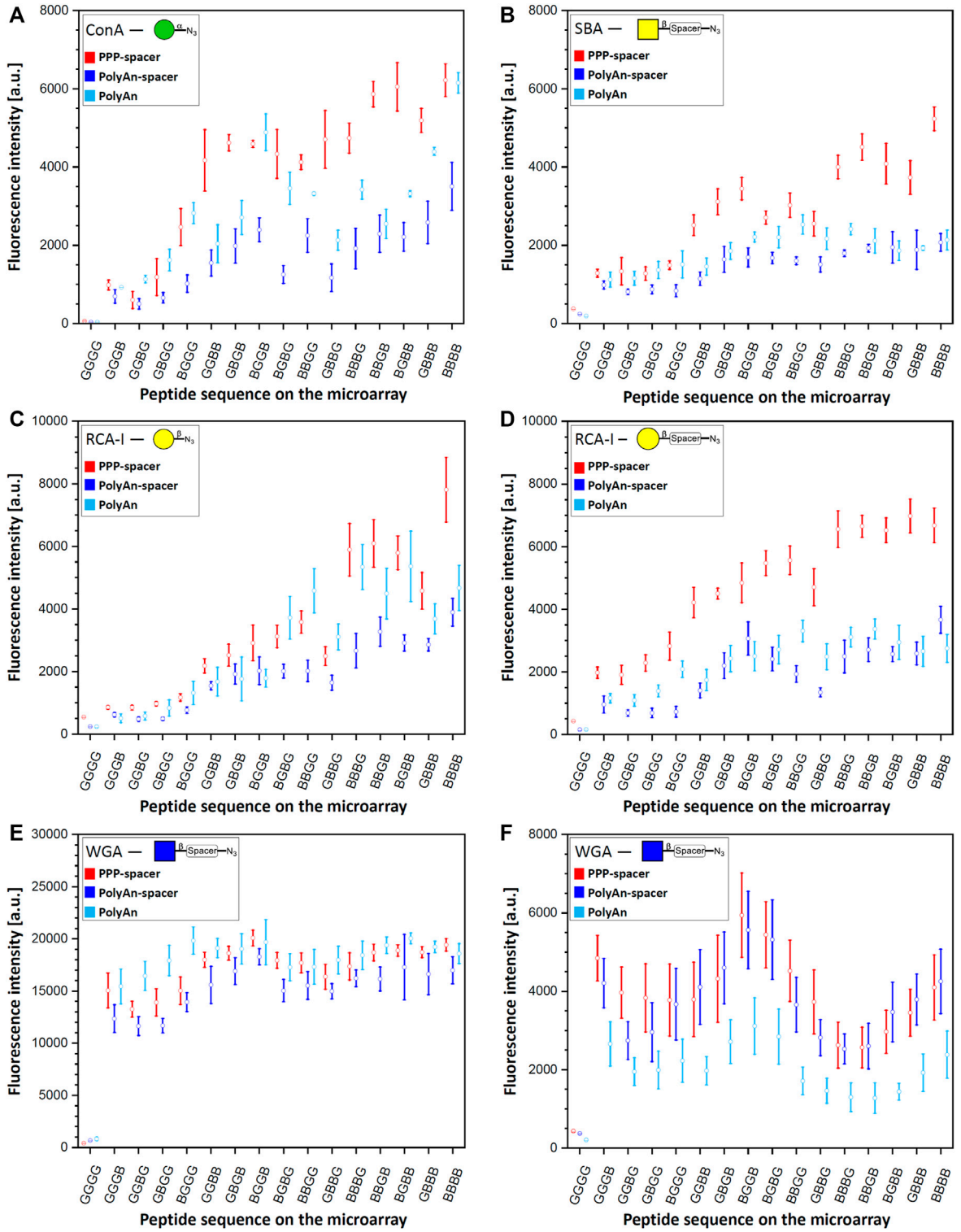


FIGURE 3 | Fluorescence staining intensities of respective sugar azides **3**, **4**, **5**, **7** and **9** with their corresponding lectins: **(A)** α-Man azide **3** with ConA (CF®633 labeled, 100 μg/ml concentration), **(B)** β-GalNAc-PEG3 azide **7** with SBA (rhodamine labeled, 10 μg/ml concentration) **(C)** β-Gal azide **4** with RCA-I (rhodamine labeled, 10 μg/ml concentration), **(D)** β-Gal-PEG3 azide **5** with RCA-I (rhodamine labeled, 10 μg/ml concentration), **(E)** β-GlcNAc-PEG3azide **9** with WGA (rhodamine labeled) with 10 μg/ml concentration, and **(F)** 0.2 μg/ml concentration on PEPperPRINT slides with PEG-spacer (PPP-spacer; red), and on PolyAn functionalized slides with (dark blue) and without PEG-spacer (light blue).

rather constant fluorescence intensity was observed, indicating a quenching effect for higher valencies, as reported previously (Mende et al., 2020). Comparing the results of the two PolyAn slides with and without PEG-spacer, also some quenching could be observed (see **Supplementary Material** for more details, Section 5).

CuAAC of the Sugars for Glycopeptide Synthesis

The copper(I)-catalyzed alkyne-azide cycloaddition (CuAAC) has been widely used in the last years for the synthesis of glycoconjugates on solid support (Freichel et al., 2017; Hill et al., 2018; Camaleño de la Calle et al., 2019; Mende et al., 2020). Herein, we used this approach to attach the following collection of sugar azide monomers to our synthesized peptide scaffolds: α-mannose (α-Man) azide **3**, β-galactose (β-Gal) azide **4**, β-galactose PEG3-spacer (β-Gal-PEG3) azide **5**, *N*-acetyl-β-galactosamine azide (β-GalNAc) **6**, *N*-acetyl-β-galactosamine PEG3-spacer (β-GalNAc-PEG3) azide **7**, *N*-acetyl-β-glucosamine (β-GlcNAc) **8**, and *N*-acetyl-β-glucosamine PEG3-spacer (β-GlcNAc-PEG3) azide **9**. The sugar azides **3**–**6** and **8** were synthesized based on known experimental procedures from their corresponding unmodified building blocks, while compounds **7** and **9** were commercially acquired (**Figure 2C**, see **Supplementary Material**). Each CuAAC reaction with individual sugars was performed in a separate well, reacting all peptide scaffold spots (n = 12) of one array with on sugar.

Glycan-GBP Assays and Fluorescence Evaluation

After the generation of the glycopeptides on the differently functionalized acceptor slides, we probed the synthesized structures with their corresponding fluorescently labeled lectins (**Figure 3**). Tetrapeptides, carrying the α-Man azide **3** were incubated with ConA (100 μg/ml, **Figure 3A**). Structures with β-Gal azide **4**, and β-Gal-PEG3 azide **5**, were probed with fluorescently labeled RCA-I, (**Figure 3C, D**) and PNA (10 μg/ml, **Supplementary Material**, Section 7.2). Tetrapeptides with attached β-GalNAc azide **6** and β-GalNAc-PEG3 azide **7** were incubated with DBA and SBA (10 μg/ml) (see **Supplementary Material**, Sections 7.3 and 7.4), while scaffolds with β-GlcNAc **8** (see **Supplementary Material**, Section 7.6) and β-GlcNAc-PEG3 azide **9** were probed with WGA (**Figure 3E, F**) (10 μg/ml). Since we observed an intensity plateau with WGA already for divalent structures, which was different from all other lectins, a 50-fold decreased WGA concentration (0.2 μg/ml) was screened additionally. We analyzed the spacing, density, and ligand dependent binding, and we could confirm that protein binding is surface dependent. In the case of the multivalent glycan-GBP interactions, similar intensity trends were observed for all used lectins on the microarrays (except for WGA, **Figure 3E, F**), with an increase in binding with an increasing number of sugars on the peptide backbone. Structures with only one attached sugar moiety, *e.g.*, BGGG, GGBG, GBGG, GGGB, showed structure dependent binding, with higher intensity for the N-terminal propargylglycine on all used slides. This could be explained by the higher distance between the sugar and the surface, making it more accessible. The tetra-glycine scaffold (GGGG) was considered as the background control.

In terms of slide functionalization, for all detected interactions, the fluorescence intensities were higher on the PEPperPRINT slides (apart from WGA and DBA). Between the two differently functionalized PolyAn slides, some structure and lectin dependent binding differences were observed.

Binding Studies on PEPperPRINT Slides

On PEPperPRINT slides, which were always equipped with the PEG-spacer, the binding of ConA to α-Man azide **3** increased exponentially with linear increase in the number of sugar moieties Our divalent scaffolds show an up to 10-fold increase in fluorescence signals in comparison to the monovalent ones, while the trivalent show an up to 20-fold, and the tetravalent is in the range of the trivalent system without significant change on the binding ability (**Figure 3A**). This trend agrees with our previous data (Mende et al., 2020). However, with the here introduced optimized synthetic conditions (see **Supplementary Material**, Section 4), the observed intensities are one order of magnitude higher with the same assay protocol. On the same acceptor slides, we screened multiple sugar monomers with and without PEG-spacer at the anomeric center. Binding of PNA, DBA, and SBA to β-Gal azide **4**, β-Gal-PEG3 azide **5**, and β-GalNAc azide **6**, respectively, was not observed. Notably, multivalent binding was only detected for SBA to the β-GalNAc-PEG3 azide **7** (**Figure 3B**). The enhanced flexibility between the anomeric position and the azide moiety given from the spacer allows the SBA to bind to the more flexible β-GalNAc-PEG3 azide **7**, but not to the β-GalNAc azide **6**. The fluorescence intensities of SBA on PEPperPRINT slides follow the same binding trend as ConA, but the binding to the tetravalent *vs.* the monovalent structures only increases about 6-fold. Despite the fact that ConA and SBA differ in their sugar specificity, both have similar orientation of binding sites and ligand recognition mechanism (Sinha et al., 2005). In contrast, RCA-I binds to both, β-Gal azide **4** and β-Gal-PEG3 azide **5** (**Figure 3C, D**). Interestingly, for the more flexible β-Gal-PEG3 azide **5**, the binding intensities of RCA-I are already at least 4-fold higher for the monovalent structures in comparison to the β-Gal azide **4**. Again, the PEG-spacer increases the flexibility of the sugar moiety and increases the distance to the triazole ring, making it more accessible for the lectin. Thus, the multivalent effect is much more pronounced for the β-Gal azide **4** than the β-Gal-PEG3 azide **5**, while the tetravalent structures from both reach a similar maximum (*i.e.,* saturation) intensity at our tested lectin concentration.

Similarly, WGA binds stronger to β-GlcNAc-PEG3-azide **9** structures (**Figure 3E, F**) than to β-GlcNAc azide **8** (see **Supplementary Material**, Section 7.6). All other lectins we studied are tetramers, WGA is the only dimer and its binding was markedly different to all other lectin binding experiments. The intensity is already high for the monovalent structures and seems to reach a plateau/saturation for divalent structures. To assess the potential impact of a lower lectin concentration, we also tested a 50-fold decreased WGA concentration (**Figure 3F**).

While, as expected, with lower concentration the total intensity was lower, a very similar trend as in the higher concentration could still be observed. Yet, a somewhat decreased intensity for trivalent structures was apparent, which seems to be a density or spacing effect. Notably, for WGA, the monovalent structure GGGB has a stronger binding (sugar is close to the surface), while generally for all other lectins, the monovalent structure BGGG (sugar is furthest away from the surface) gives the highest intensity.

Our studies show a spacing dependent binding for the divalent systems. Higher intensities for ConA (to α-Man azide **3**), and RCA-I (to β-Gal-PEG3 azide **5**) are attained for non-adjacent divalent structures (GBGB, BGGB, BGBG). A similar effect is observed for trivalent binders: the intermediate glycine (BGBB, BBGB) increases the binding for ConA, SBA, and RCA-I (β-Gal azide **4**) in comparison to structures with terminal glycines (BBBG, GBBB). For RCA-I, the more flexible β-Gal-PEG3 azide **5** shows a generally higher binding, but especially on the trivalent system with the C-terminal glycine (BBBG). In the case of WGA, the divalent scaffolds with two neighboring Pra moieties (BBGG, GGBB, GBBG) give less binding, while stronger binding is obtained on non-neighboring Pra scaffolds (GBGB, BGBG, BGGB).

Binding Studies on PolyAn Slides

Next, we investigated the impact of a different commercial substrate on the binding of the lectins. Thus, we functionalized the more hydrophilic PolyAn slides with the same PEG-spacer (see **Supplementary Material**, Section 3.2.). We measured the hydrophobicity of all used slides (PEPperPRINT and PolyAn) with and without PEG-spacer, showing that the hydrophilic character of the PoyAn slides does not change after the attachment of the PEG-spacer (**Supplementary Material**, Section 6). Comparing the PolyAn to the PEPperPRINT slides, generally similar interactions were detected, while some distinct differences for multivalency, sugar density, and spacing could be observed. The binding ability of ConA on PolyAn slides bearing the PEG-spacer decreased by a factor of 2 compared to the intensities observed on PEPperPRINT slides (**Figure 3A**). This trend was observed for almost all other lectin interactions. In the case of SBA (**Figure 3B**), the PolyAn slide surface seems to prevent a multivalent effect (*i.e.,* only linear intensity increase), at least for this lectin concentration. For RCA-I (**Figure 3C, D**), the PolyAn slide without PEG-spacer showed a similar trend as the PEPperPRINT slide for the β-Gal azide **4**. For β-Gal-PEG3 azide **5**, again, both PolyAn slides showed a similar trend to the PEPperPRINT slide, but with a much weaker multivalent effect and a generally 2- to 3-fold lower intensity. In case of WGA with β-GlcNAc-PEG3 azide **9** (**Figure 3E, F**), the PolyAn surface without PEG showed a generally higher intensity in the assay with high concentration. For the low concentration WGA assay, PolyAn showed a lower intensity, but still the same trend. Similar binding behavior was also observed for β-GlcNAc azide **8** with WGA (see **Supplementary Material**, Section 7.6).

Interestingly, only on the PolyAn surfaces, DBA showed a weak binding to β-GalNAc azide **6** and β-GalNAc-PEG3-azide **7** (see **Supplementary Material**, Section 7.3). However, in this case, we also observed a high background signal for the GGGG control, which is a hydrophobic structure. In the future, it should be further investigated, whether a more hydrophobic alkyl linker (instead of PEG) on the surface can increase this binding, since DBA is known to have a hydrophobic adenine-binding site in addition to the carbohydrate recognition domain (Hamelryck et al., 1999).

As reported before with the PEPperPRINT slides, no binding could be identified for SBA and PNA with β-Gal azide **4**, β-Gal-PEG3 azide **5**, and β-GalNAc azide **6** on PolyAn slides (see **Supplementary Material**, Sections 7.2, 7.4).

Structure dependent binding was also observed between the different lectins on PolyAn slides. Structures with same theoretical spacing (GBGB and BGBG) do not show the same binding intensities. The strongest binding for WGA on PolyAn slides was detected for the divalent structure BGGB, especially for the lower lectin concentration. Thus, and because the binding sites of WGA are very close to each other (see Conclusion), it indicates cross-linking and chelating binding mode (*i.e.,* two binding sites of WGA bind to one structure). Remarkably reduced binding of WGA was detected on the tri- and tetravalent structures on all substrates compared to the divalent structures, which might be caused by sterical hindrance.

CONCLUSION

We describe a flexible and cost-efficient method for the synthesis of defined multivalent glycopeptide arrays. On each microarray, 16 different tetrapeptides were generated *in situ* by our laser-based technology and seven different azido sugar monomers were attached by CuAAC (resulting in a total of 112 different structures on three different surfaces). To study the impact of different commercial surfaces functionalized with different linkers, we first optimized the solid-phase synthesis conditions (amino acid concentration, lasing parameters, coupling time) for different commercial microarray substrates. These optimizations improved the signal-to-noise ratios for our model lectin ConA by one order of magnitude, and helped to expand the applications for our synthesis platform to include weakly binding lectins (*e.g.,* DBA).

Lectin binding depends on spacing, density, surface functionalization, and concentration. PEG-functionalized PEPperPRINT slides provided generally higher signal intensities than PolyAn slides, with the exception of DBA. Lower binding intensities on PolyAn slides equipped with the PEG-spacer indicate that lectin binding decreases under very hydrophilic conditions for the majority of lectins. For a better understanding, we experimentally determined the (water) contact angle of the different surfaces. PEPperPRINT slides are more hydrophobic, while PolyAn slides maintained their hydrophilic character even after the attachment of a PEG-spacer.

Most lectins showed a multivalent binding effect that mainly depends on the valency with exception of the WGA binding assay. A saturation of binding intensity for divalent structures was detected on all microarrays due to the chelating binding mode, leading to cross-linking. Yet, no binding was observed for PNA

and DBA on PEPperPRINT slides with simple sugar moieties, while weak interaction was obtained on PolyAn slides with DBA. Spacing of the synthetic scaffolds may not fit the binding sites of most lectins, the selection of sugars was not optimal, and the triazole ring might cause sterical problems. Future investigations will require screening of different mono- and disaccharides, such as lactose and the T-antigen with PNA. In case of DBA, an α-N-acetyl galactosamine residue should offer a much higher binding ability than the β-N-acetyl galactosamine residue. Additionally, longer peptide scaffolds should be synthesized, as well as longer linkers (*e.g.,* PEG5) should be introduced between the anomeric position and the peptide backbone, to increase the size and the flexibility of the synthesized structures.

We were unable to detect any binding between the C-type lectins mLangerin, mMGL-1, and mMincle with their corresponding sugar monomers (see **Supplementary Material**, Section 8). Interestingly, Di Maio *et al.* very recently reported a microarray assay with multivalent display of mono- and dimannose, where other C-type lectins (DC-SIGNR ECD, trivalent Langerin ECD, monomeric Dectin-2 ECD) were screened. These lectins selectively and strongly bind to Man-α1,2Man, but almost no binding for α-Man monomer was reported (Di Maio et al., 2021). Future screening of disaccharides such as Man-α1,2Man with high valency and staining with directly fluorescently labeled lectins may provide more information on these lectins.

Notably, on PolyAn slides with and without spacer, most lectins showed a more linear (less multivalent) increase in binding with increasing numbers of sugar PEG3 azides. For the less flexible sugar azides without PEG3, typical multivalent trends could be observed.

The molecular spacing of the sugars on the tetrapeptides had a similar impact on ConA, SBA, and RCA-I. Scaffolds with the same theoretical spacing, such as GBGB and BGBG, showed different binding strengths with the latter typically showing a stronger binding strength. Similarly, divalent structures with larger spacing (BGGB) showed stronger binding than the more adjacent scaffolds (*e.g.,* GBBG).

To our knowledge, this work is the first, showing the synthesis of glycopeptides with defined valencies and spacing *in situ* on different commercially available microarrays to investigate the effect of substrate functionalization. Our technology relies on readily available compounds (Eickelmann et al., 2019) and can be fully automated (Paris et al., 2019). This enables us to screen a diverse collection of glycopeptides with their corresponding lectins. We believe that by using other propargyl amino acids in our process in the future, we should be able to find ideal multivalent glycopeptide binders for different lectins. However, the microarray substrate functionalization plays an important role for glycan-GBP interaction studies and has to be thoroughly considered.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

All authors contributed in the revision of the manuscript and have given their approval to the final version of the manuscript. Synthesis of sugar azides and characterization performed by AT. Synthesis, characterization, and evaluation of all formed glycopeptides were done by AT and KH. Bioassays were performed by AT and KH. AT and MM developed the methodology. SE measured the contact angles. BL provided the C-type lectins. JH, GP, AT, and KH optimized the oven coupling protocol. GP supported the laser system development. All graphs were prepared by AT and KH. FFL devised and supervised the project. AT and FL wrote the manuscript with input from PHS.

FUNDING

This research was supported by the German Federal Ministry of Education and Research (BMBF, grant number 13XP5050A) the MPG-FhG cooperation (Glyco3Display), and the Max Planck Society.

ACKNOWLEDGMENTS

The authors would like to thank the members of the Department of Biomolecular Systems for the help and technical support, specifically Klaus Bienert, Eva Settels, Olaf Niemeyer, and Felix Hentschel.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.766932/full#supplementary-material

REFERENCES

Artigas, G., Monteiro, J. T., Hinou, H., Nishimura, S.-I., Lepenies, B., and Garcia-Martin, F. (2017). Glycopeptides as Targets for Dendritic Cells: Exploring MUC1 Glycopeptides Binding Profile toward Macrophage Galactose-type Lectin (MGL) Orthologs. *J. Med. Chem.* 60, 9012–9021. doi:10.1021/acs.jmedchem.7b01242

Camaleño de la Calle, A., Gerke, C., Chang, X. J., Grafmüller, A., Hartmann, L., Schmidt, S., et al. (2019). Multivalent Interactions of Polyamide Based Sequence-Controlled Glycomacromolecules with Concanavalin A. *Macromol. Biosci.* 19, 1900033. doi:10.1002/mabi.201900033

Cecioni, S., Imberty, A., and Vidal, S. (2015). Glycomimetics versus Multivalent Glycoconjugates for the Design of High Affinity Lectin Ligands. *Chem. Rev.* 115, 525–561. doi:10.1021/cr500303t

Delbianco, M., Bharate, P., Varela-Aramburu, S., and Seeberger, P. H. (2016). Carbohydrates in Supramolecular Chemistry. *Chem. Rev.* 116, 1693–1752. doi:10.1021/acs.chemrev.5b00516

Di Maio, A., Cioce, A., Achilli, S., Thépaut, M., Vivès, C., Fieschi, F., et al. (2021). Controlled Density Glycodendron Microarrays for Studying Carbohydrate-Lectin Interactions. *Org. Biomol. Chem.* 19, 7357–7362. doi:10.1039/D1OB00872B

Eickelmann, S., Tsouka, A., Heidepriem, J., Paris, G., Zhang, J., Molinari, V., et al. (2019). A Low-Cost Laser-Based Nano-3D Polymer Printer for Rapid Surface Patterning and Chemical Synthesis of Peptide and Glycan Microarrays. *Adv. Mater. Technol.* 4, 1900503. doi:10.1002/admt.201900503

Fasting, C., Schalley, C. A., Weber, M., Seitz, O., Hecht, S., Koksch, B., et al. (2012). Multivalency as a Chemical Organization and Action Principle. *Angew. Chem. Int. Ed.* 51, 10472–10498. doi:10.1002/anie.201201114

Freichel, T., Eierhoff, S., Snyder, N. L., and Hartmann, L. (2017). Toward Orthogonal Preparation of Sequence-Defined Monodisperse Heteromultivalent Glycomacromolecules on Solid Support Using Staudinger Ligation and Copper-Catalyzed Click Reactions. *J. Org. Chem.* 82, 9400–9409. doi:10.1021/acs.joc.7b01398

Gao, C., Wei, M., McKitrick, T. R., McQuillan, A. M., Heimburg-Molinaro, J., Cummings, R. D., et al. (2019). Glycan Microarrays as Chemical Tools for Identifying Glycan Recognition by Immune Proteins. *Front. Chem.* 7, 833. doi:10.3389/fchem.2019.00833

Gao, C., Wei, M., McKitrick, T. R., McQuillan, A. M., Heimburg-Molinaro, J., and Cummings, R. D. (2019). Glycan Microarrays as Chemical Tools for Identifying Glycan Recognition by Immune Proteins. *Front. Chem.* 7, 833. doi:10.3389/fchem.2019.00833

Geissner, A., Anish, C., and Seeberger, P. H. (2014). Glycan Arrays as Tools for Infectious Disease Research. *Curr. Opin. Chem. Biol.* 18, 38–45. doi:10.1016/j.cbpa.2013.11.013

Geissner, A., Reinhardt, A., Rademacher, C., Johannssen, T., Monteiro, J., Lepenies, B., et al. (2019). Microbe-focused Glycan Array Screening Platform. *Proc. Natl. Acad. Sci. USA* 116, 1958–1967. doi:10.1073/pnas.1800853116

Geissner, A., and Seeberger, P. H. (2016). Glycan Arrays: From Basic Biochemical Research to Bioanalytical and Biomedical Applications. *Annu. Rev. Anal. Chem.* 9, 223–247. doi:10.1146/annurev-anchem-071015-041641

Godula, K., Rabuka, D., Nam, K. T., and Bertozzi, C. R. (2009). Synthesis and Microcontact Printing of Dual End-Functionalized Mucin-like Glycopolymers for Microarray Applications. *Angew. Chem.* 121, 5073–5076. doi:10.1002/ange.200805756

Haag, R. (2015). Multivalency as a Chemical Organization and Action Principle. *Beilstein J. Org. Chem.* 11, 848–849. doi:10.3762/bjoc.11.94

Hamelryck, T. W., Loris, R., Bouckaert, J., Dao-Thi, M.-H., Strecker, G., Imberty, A., et al. (1999). Carbohydrate Binding, Quaternary Structure and a Novel Hydrophobic Binding Site in Two Legume Lectin Oligomers from Dolichos Biflorus 1 1Edited by R. Huber. *J. Mol. Biol.* 286, 1161–1177. doi:10.1006/jmbi.1998.2534

Hawkes, W., Huang, D., Reynolds, P., Hammond, L., Ward, M., Gadegaard, N., et al. (2019). Probing the Nanoscale Organisation and Multivalency of Cell Surface Receptors: DNA Origami Nanoarrays for Cellular Studies with Single-Molecule Control. *Faraday Discuss.* 219, 203–219. doi:10.1039/C9FD00023B

Hill, S. A., Gerke, C., and Hartmann, L. (2018). Recent Developments in Solid-phase Strategies towards Synthetic, Sequence-Defined Macromolecules. *Chem. Asian J.* 13, 3611–3622. doi:10.1002/asia.201801171

Kilcoyne, M., Gerlach, J. Q., Gough, R., Gallagher, M. E., Kane, M., Carrington, S. D., et al. (2012). Construction of a Natural Mucin Microarray and Interrogation for Biologically Relevant Glyco-Epitopes. *Anal. Chem.* 84, 3330–3338. doi:10.1021/ac203404n

Kim, H. S., Hyun, J. Y., Park, S.-H., and Shin, I. (2018). Analysis of Binding Properties of Pathogens and Toxins Using Multivalent Glycan Microarrays. *RSC Adv.* 8, 14898–14905. doi:10.1039/C8RA01285G

Laigre, E., Tiertant, C., Goyard, D., and Renaudet, O. (2018). Identification of Nanomolar Lectin Ligands by a Glycodendrimer Microarray. *ACS Omega* 3, 14013–14020. doi:10.1021/acsomega.8b01526

Loeffler, F. F., Foertsch, T. C., Popov, R., Mattes, D. S., Schlageter, M., Sedlmayr, M., et al. (2016). High-flexibility Combinatorial Peptide Synthesis with Laser-Based Transfer of Monomers in Solid Matrix Material. *Nat. Commun.* 7, 11844. doi:10.1038/ncomms11844

Maglinao, M., Eriksson, M., Schlegel, M. K., Zimmermann, S., Johannssen, T., Götze, S., et al. (2014). A Platform to Screen for C-type Lectin Receptor-Binding Carbohydrates and Their Potential for Cell-specific Targeting and Immune Modulation. *J. Controlled Release* 175, 36–42. doi:10.1016/j.jconrel.2013.12.011

Mayer, S., Raulf, M.-K., and Lepenies, B. (2017). C-type Lectins: Their Network and Roles in Pathogen Recognition and Immunity. *Histochem. Cel Biol.* 147, 223–237. doi:10.1007/s00418-016-1523-7

Mende, M., Bordoni, V., Tsouka, A., Loeffler, F. F., Delbianco, M., and Seeberger, P. H. (2019). Multivalent Glycan Arrays. *Faraday Discuss.* 219, 9–32. doi:10.1039/c9fd00080a

Mende, M., Tsouka, A., Heidepriem, J., Paris, G., Mattes, D. S., Eickelmann, S., et al. (2020). On-Chip Neo-Glycopeptide Synthesis for Multivalent Glycan Presentation. *Chem. Eur. J.* 26, 9954–9963. doi:10.1002/chem.202001291

Merrifield, R. B. (1963). Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* 85, 2149–2154. doi:10.1021/ja00897a025

Moni, L., Pourceau, G., Zhang, J., Meyer, A., Vidal, S., Souteyrand, E., et al. (2009). Design of Triazole-Tethered Glycoclusters Exhibiting Three Different Spatial Arrangements and Comparative Study of Their Affinities towards PA-IL and RCA 120 by Using a DNA-Based Glycoarray. *ChemBioChem* 10, 1369–1378. doi:10.1002/cbic.200900024

Müller, C., Despras, G., and Lindhorst, T. K. (2016). Organizing Multivalency in Carbohydrate Recognition. *Chem. Soc. Rev.* 45, 3275–3302. doi:10.1039/C6CS00165C

O'Neil, C. L., Stine, K. J., and Demchenko, A. V. (2018). Immobilization of Glycans on Solid Surfaces for Application in Glycomics. *J. Carbohydr. Chem.* 37, 225–249. doi:10.1080/07328303.2018.1462372

Oyelaran, O., Li, Q., Farnsworth, D., and Gildersleeve, J. C. (2009). Microarrays with Varying Carbohydrate Density Reveal Distinct Subpopulations of Serum Antibodies. *J. Proteome Res.* 8, 3529–3538. doi:10.1021/pr9002245

Paris, G., Heidepriem, J., Tsouka, A., Mende, M., Eickelmann, S., and Loeffler, F. F. (2019). "Automated Laser-Assisted Synthesis of Microarrays for Infectious Disease Research," in *Microfluidics, BioMEMS, and Medical Microsystems XVII*. Editors B. L. Gray and H. Becker (San Francisco, CA: SPIE), 11. doi:10.1117/12.2516781

Paris, G., Klinkusch, A., Heidepriem, J., Tsouka, A., Zhang, J., Mende, M., et al. (2020). Laser-induced Forward Transfer of Soft Material Nanolayers with Millisecond Pulses Shows Contact-Based Material Deposition. *Appl. Surf. Sci.* 508, 144973. doi:10.1016/j.apsusc.2019.144973

Purohit, S., Li, T., Guan, W., Song, X., Song, J., Tian, Y., et al. (2018). Multiplex Glycan Bead Array for High Throughput and High Content Analyses of Glycan Binding Proteins. *Nat. Commun.* 9, 258. doi:10.1038/s41467-017-02747-y

Raman, R., Tharakaraman, K., Sasisekharan, V., and Sasisekharan, R. (2016). Glycan-protein Interactions in Viral Pathogenesis. *Curr. Opin. Struct. Biol.* 40, 153–162. doi:10.1016/j.sbi.2016.10.003

Redman, R. L., and Krauss, I. J. (2021). Directed Evolution of 2′-Fluoro-Modified, RNA-Supported Carbohydrate Clusters that Bind Tightly to HIV Antibody 2G12. *J. Am. Chem. Soc.* 143, 8565–8571. doi:10.1021/jacs.1c03194

Ruprecht, C., Geissner, A., Seeberger, P. H., and Pfrengle, F. (2019). Practical Considerations for Printing High-Density Glycan Microarrays to Study Weak Carbohydrate-Protein Interactions. *Carbohydr. Res.* 481, 31–35. doi:10.1016/j.carres.2019.06.006

Sinha, S., Mitra, N., Kumar, G., Bajaj, K., and Surolia, A. (2005). Unfolding Studies on Soybean Agglutinin and Concanavalin A Tetramers: A Comparative Account. *Biophysical J.* 88, 1300–1310. doi:10.1529/biophysj.104.051052

Specker, D., and Wittmann, V. (2006). *"Synthesis and Application of Glycopeptide and Glycoprotein Mimetics," in* Glycopeptides And Glycoproteins. Berlin, Heidelberg: Springer Berlin Heidelberg, 65–107. doi:10.1007/128_2006_104

Stadler, V., Kirmse, R., Beyer, M., Breitling, F., Ludwig, T., and Bischoff, F. R. (2008). PEGMA/MMA Copolymer Graftings: Generation, Protein Resistance, and a Hydrophobic Domain. *Langmuir* 24, 8151–8157. doi:10.1021/la800772m

Temme, J. S., Campbell, C. T., and Gildersleeve, J. C. (2019). Factors Contributing to Variability of Glycan Microarray Binding Profiles. *Faraday Discuss.* 219, 90–111. doi:10.1039/C9FD00021F

Tikhonov, A., Smoldovskaya, O., Feyzkhanova, G., Kushlinskii, N., and Rubina, A. (2020). Glycan-specific Antibodies as Potential Cancer Biomarkers: A Focus on Microarray Applications. *Clin. Chem. Lab. Med.* 58, 1611–1622. doi:10.1515/cclm-2019-1161

Valles, D. J., Naeem, Y., Rozenfeld, A. Y., Aldasooky, R. W., Wong, A. M., Carbonell, C., et al. (2019). Multivalent Binding of Concanavalin A on Variable-Density Mannoside Microarrays. *Faraday Discuss.* 219, 77–89. doi:10.1039/c9fd00028c

Valverde, P., Martínez, J. D., Cañada, F. J., Ardá, A., and Jiménez-Barbero, J. (2020). Molecular Recognition in C-Type Lectins: The Cases of DC-SIGN, Langerin, MGL, and L-Sectin. *ChemBioChem* 21, 2999–3025. doi:10.1002/cbic.202000238

Varki, A. (2009). *Essentials of Glycobiology*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Wang, D., Liu, S., Trummer, B. J., Deng, C., and Wang, A. (2002). Carbohydrate Microarrays for the Recognition of Cross-Reactive Molecular Markers of Microbes and Host Cells. *Nat. Biotechnol.* 20, 275–281. doi:10.1038/nbt0302-275

Wittmann, V., and Pieters, R. J. (2013). Bridging Lectin Binding Sites by Multivalent Carbohydrates. *Chem. Soc. Rev.* 42, 4492. doi:10.1039/c3cs60089k

Zhou, J. Y., and Cobb, B. A. (2021). Glycans in Immunologic Health and Disease. *Annu. Rev. Immunol.* 39, 511–536. doi:10.1146/annurev-immunol-101819-074237

Zilio, C., Bernardi, A., Palmioli, A., Salina, M., Tagliabue, G., Buscaglia, M., et al. (2015). New "Clickable" Polymeric Coating for Glycan Microarrays. *Sensors Actuators B: Chem.* 215, 412–420. doi:10.1016/j.snb.2015.03.079

Conflict of Interest: FFL is named on a patent related to laser-based microarray synthesis.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Chemistry

ORIGINAL RESEARCH
published: 19 November 2021
doi: 10.3389/fchem.2021.767448

Site-Specific Glycosylation Patterns of the SARS-CoV-2 Spike Protein Derived From Recombinant Protein and Viral WA1 and D614G Strains

*Yuan Tian, Lisa M. Parsons, Ewa Jankowska and John F. Cipollo**

Food and Drug Administration, Center for Biologics Evaluation and Research, Division of Bacterial, Parasitic and Allergenic Products, Silver Spring, MD, United States

The SARS-CoV-2 spike protein is heavily glycosylated, having 22 predicted N-glycosylation sites per monomer. It is also O-glycosylated, although the number of O-glycosites is less defined. Recent studies show that spike protein glycans play critical roles in viral entry and infection. The spike monomer has two subdomains, S1 and S2, and a receptor-binding domain (RBD) within the S1 domain. In this study, we have characterized the site-specific glycosylation patterns of the HEK293 recombinant spike RBD and S1 domains as well as the intact spike derived from the whole virus produced in Vero cells. The Vero cell-derived spike from the WA1 strain and a D614G variant was analyzed. All spike proteins, S1, and RBDs were analyzed using hydrophilic interaction chromatography (HILIC) and LC-MS/MS on an Orbitrap Eclipse Tribrid mass spectrometer. N-glycans identified in HEK293-derived S1 were structurally diverse. Those found in the HEK293-derived RBD were highly similar to those in HEK293 S1 where N-glycosites were shared. Comparison of the whole cell-derived WA1 and D614G spike proteins revealed that N-glycosites local to the mutation site appeared to be more readily detected, hinting that these sites are more exposed to glycosylation machinery. Moreover, recombinant HEK293-derived S1 was occupied almost completely with complex glycan, while both WA1 and D614G derived from the Vero E6 cell whole virus were predominantly high-mannose glycans. This stands in stark contrast to glycosylation patterns seen in both CHO- and HEK cell-derived recombinant S1, S2, and the whole spike previously reported. Concerning O-glycosylation, our analyses revealed that HEK293 recombinant proteins possessed a range of O-glycosites with compositions consistent with Core type 1 and 2 glycans. The O-glycosites shared between the S1 and RBD constructs, sites T323 and T523, were occupied by a similar range of Core 1 and 2 type O-glycans. Overall, this study reveals that the sample nature and cell substrate used for production of these proteins can have a dramatic impact on the glycosylation profile. SARS-CoV-2 spike glycans are associated with host ACE2 receptor interaction efficiency. Therefore, understanding such differences will serve to better understand these host–pathogen interactions and inform the choice of cell substrates to suite downstream investigations.

Keywords: SARS-CoV-2, N-glycosylation, O-glycosylation, glycan shield, cell substrate, spike protein, microheterogeneity, D614G variant

OPEN ACCESS

Edited by:
Zhongping Tan, Chinese Academy of Medical Sciences and Peking Union Medical College, China

Reviewed by:
Shisheng Sun, Northwest University, China
Giovanni Signore, Fondazione Pisana per la Scienza Onlus, Italy

****Correspondence:***
John F. Cipollo
john.cipollo@fda.hhs.gov

Specialty section:
This article was submitted to Chemical Biology, a section of the journal Frontiers in Chemistry

Received: *30 August 2021*
Accepted: *14 October 2021*
Published: *19 November 2021*

Citation:
Tian Y, Parsons LM, Jankowska E and Cipollo JF (2021) Site-Specific Glycosylation Patterns of the SARS-CoV-2 Spike Protein Derived From Recombinant Protein and Viral WA1 and D614G Strains. Front. Chem. 9:767448. doi: 10.3389/fchem.2021.767448

INTRODUCTION

SARS-CoV-2 is an enveloped, positive single-stranded RNA beta coronavirus expressing four main structural proteins, which include nucleocapsid, spike, membrane, and envelope proteins (Lu et al., 2020; Singh and Yi, 2021). The trimeric spike protein is the major surface protein of the SARS-CoV-2 virus and serves as an entry protein for host cell infection (Shang et al., 2020). To facilitate the fusion of the viral membrane with the infected cells, the spike proteins are cleaved into S1 and S2 subunits by cellular proteases, such as furin (Hoffmann et al., 2020a; Hoffmann et al., 2020b; Xia, 2021). The S1 subunit contains the N-terminal and the receptor-binding domain (RBD) (Xia, 2021), and recombinant RBD binds to the human angiotensin converting enzyme 2 (ACE2) present as a surface receptor on host cells. The S2 domain serves the function of membrane fusion, which contains a fusion peptide (FP), an internal fusion peptide (IFP), two heptad repeat domains (HR1 and HR2), a transmembrane domain, and a C-terminal domain (Coutard et al., 2020; Xia, 2021).

Enveloped viruses have evolved to take advantage of many host cell processes including glycosylation (Watanabe et al., 2019). Viral protein glycosylation functions in a number of ways in the viral lifestyle including viral particle stability, mediating viral infection, host immune response, and immune evasion (Walls et al., 2016; Watanabe et al., 2019). Viral glycosylation of key envelope glycoproteins can be dynamic over time as the virus propagates through the host population, allowing immune avoidance to evolve over time (Li et al., 2021). Recent cryo-EM studies reported that the recombinant SARS-CoV-2 spike protein is extensively glycosylated (Grant et al., 2020; Wrapp et al., 2020). Using recombinant proteins, earlier studies reported glycosylation of the 22 predicted N-linked glycosites in the spike protein at high occupancy and lower glycosylation occupancy on O-linked glycosites (Watanabe et al., 2020a; Watanabe et al., 2020b; Shajahan et al., 2020; Zhao et al., 2020; Zhou et al., 2021). A recent study reported that glycosylation is essential for SARS-CoV-2 viral entry and infection (Yang et al., 2020). Since glycans are produced through a stochastic process that is dependent upon glycosylation, enzyme expression, location, concentration, and the particular glycoprotein's sequence and structural characteristics, it can be altered under selective pressure. During viral evolution, with passage through the human population, glycosites are added and deleted often, leading to an increased number of sites and glycan complexity. The overall glycosylation characteristics such as composition, subclass, heterogeneity, and density over the surface of the protein can have dramatic effects on viral survival, transmission, and immune evasion (Vigerust and Shepherd, 2007; Watanabe et al., 2019; Li et al., 2021). Spike glycoproteins are often the major target for vaccine design and antivirus drug development. Understanding the glycosylation microheterogeneity of the spike protein can facilitate the process.

Here, we characterize site-specific glycosylation on recombinant RBD and the S1 domain of the spike protein produced in HEK293 cells to understand the glycosylation microheterogeneity produced using this cell substrate. The question remains open: whether the glycosylation of these recombinant proteins differs from that of the native spike produced in the whole virus. Thus, we compare the glycosylation of recombinant RBD and S1 to two intact viruses, the WA1 strain and a D614G variant, both produced in Vero E6 cells. The SARS-CoV-2/USA-WA1/2020 (USA-WA1) viral strain was isolated from the specimen of the first confirmed case in the United States (Harcourt et al., 2020; Wang et al., 2021a). Whole genome sequencing confirmed that this strain contained D614 as the original form of the SARS-CoV-2 virus. SARS-CoV-2/Massachusetts/VPT1/2020 (MA/VPT1), containing the D614G mutation, was isolated in Vero E6 cells from a nasopharyngeal specimen collected in April 2020 (Wang et al., 2021a). The D614G mutation, which appeared in early 2020 (Korber et al., 2020), has become dominant worldwide. The D614G mutation is also carried by the more recent and concerning SARS-CoV-2 variants, including B.1.1.7, B.1.351, P.1, and B.1.617 (https://www.cdc.gov/coronavirus/2019-ncov/variants/). Compared to strains containing the original D614, viruses with the D614G mutation have significantly higher infection titers as well as faster transmission but are less sensitive to spike-based SARS-CoV-2 vaccine sera produced in mice, non-human primates, and humans (Hou et al., 2020; Korber et al., 2020; Yurkovetskiy et al., 2020). In addition, structural analysis demonstrates that the G614 spike is in a more open conformation with extended RBDs (Yurkovetskiy et al., 2020). Given this conformational shift, it is of interest to examine glycosylation for possible changes in the D614G spike compared to its close progenitor, the WA1 strain, while keeping the viral propagation cell platform the same. Therefore, in addition to analysis of recombinant spike constructs, we report the glycosylation patterns of spikes in WA1 and D614G strains produced by the whole virus in Vero E6 cells. Our results may aid in interpretation of experimental data concerning spike interactions with the host and surrogates as well as the development of therapeutics and vaccines against the SARS-CoV-2 virus.

MATERIALS AND METHODS

Recombinant Proteins and Intact Viruses

Recombinant protein SARS-CoV-2 Spike S1 and RBD proteins expressed in HEK293 cells were purchased from Sanyou Bio (China). The whole virus of the WA1 strain was from the first patient of SARS-CoV-2 virus infection in the United States (Harcourt et al., 2020). The virus was isolated from nasopharyngeal and oropharyngeal specimens from this patient, and the viral sequence was confirmed (Wang et al., 2021a). This strain, SARS-CoV-2/USA-WA1/2020 (USA-WA1), is the original form of the SARS-CoV-2 virus without mutation at the 614 amino acid (Li et al., 2021). The D614G variant carrying the spike protein amino acid change at 614D to G, SARS-CoV-2/Massachusetts/VPT1/2020 (MA/VPT1), was isolated from Vero E6 cells from a nasopharyngeal specimen collected in April 2020 (Wang et al., 2021a). Both viruses were grown in Vero E6 cells, and the supernatant of the passage 4 stock of each virus was collected by centrifugation. After the viruses

were frozen to −80°C at least overnight, the viruses were inactivated by gamma irradiation.

Chemicals, reagents, and TSKgel amide 80 particles were purchased from Tosoh Bioscience LLC (Montgomeryville, PA). Sep-Pak C18 cartridges were purchased from Waters (Milford, MA, United States). Sequencing grade-modified trypsin, chymotrypsin, and Glu-C were purchased from Promega Corp. (Madison, WI). PNGase F was purchased from New England BioLabs, Inc (Ipswich, MA). Iodoacetimide, dithiothreitol (DTT), trifluoroacetic acid (TFA) (≥99%), and other chemicals were purchased from Sigma-Aldrich (St. Louis, MO, United States). Solvents were of high-pressure liquid chromatography (HPLC) grade or higher and purchased from Thermo Fisher Scientific (Waltham, MA). All other reagents were of ACS grade or higher.

Protein Digestion

Recombinant proteins (200 μg) were dissolved in 50 mM ammonium bicarbonate at a concentration of 2 μg/μL. DTT was added to reduce the disulfide bonds at a final concentration of 5 mM for 30 min at 60°C. The samples were cooled to room temperature (RT), and iodoacetamide (IAA) was added to alkylate the reduced cysteine residues at a concentration of 15 mM for 30 min in the dark at RT. DTT was added to 25 mM to neutralize the remaining IAA. Trypsin or chymotrypsin was added (enzyme/protein, 1:50, w/w), and the samples were incubated at 37°C overnight.

For intact viruses, approximately 500 μg of the proteins was reduced with 5 mM DTT in 6M urea and 50 mM NH_4HCO_3 for 1 h at 37°C and subsequently alkylated with 15 mM iodoacetamide for 30 min at RT in the dark. To neutralize the remaining IAA, DTT was added to 25 mM. Samples were diluted 6-fold with 50 mM NH_4HCO_3 and 1 mM $CaCl_2$ and digested with sequencing grade-modified trypsin or chymotrypsin at 1:50 (enzyme/protein, w/w) overnight at 37°C. Glu-C was added to the tryptic digest at 1:50 (enzyme/protein, w/w) and incubated overnight at 37°C.

Glycopeptide Enrichment by the HILIC Resin

Intact glycopeptides were enriched by solid-phase extraction using the TSKgel amide 80 hydrophilic interaction chromatography (HILIC) resin according to our previous report (An and Cipollo, 2011). Briefly, 200 mg (400 μL of the wet resin) of the amide 80 resin was placed into a Supelco fritted 1 ml column and washed with 1 ml of 0.1% trifluoroacetic acid (TFA)–water solution. The column was conditioned with 1 ml of 0.1% TFA–80% ACN. The peptides were suspended in 0.1% TFA–80% ACN and slowly loaded to the column. The hydrophobic species were washed away with 3 ml of 0.1% TFA–80% ACN. For recombinant proteins, the glycopeptides were eluted with 1 ml of 0.1% TFA–50% ACN and 1 ml of 0.1% TFA–25% ACN. The eluents were combined and vacuum-dried. For whole viruses, the glycopeptides were eluted sequentially with 1 ml of 0.1% TFA–65% ACN, 0.1% TFA–60% ACN, 0.1% TFA–50% ACN, and 0.1% TFA–25% ACN. Each eluent was vacuum-dried and analyzed by mass spectrometry separately.

Reversed-Phase HPLC Fractionation

The PNGase F-treated WA1 and D614G peptides were dried by speed vacuum and resuspended in 20 μl of 10 mM TEAB. The fractionation of the peptide samples is carried out using an Agilent Poroshell 120 Column (2.7 μm, 2.1 × 150 mm) and an Agilent UHPLC 1290 system. The separation was performed by running a gradient of Solvent B (10 mM TEABC, pH 8.0, 90% ACN) and Solvent A (10 mM TEAB, pH 8.0) at a flow rate of 200 μl/min in a 150 min run. The elute fractions are collected into a 96-well plate using a 1260 series auto-sample fraction collector. The 96 elute fractions were further combined into 12 fractions according to the collection time (combined per column into one fraction, 12 column 12 fractions). Each fraction was dried by rotary evacuation.

Site Occupancy Analysis

Digested peptides were deglycosylated with PNGase F in 50 mM NH_4HCO_3. PNGase F cleaves between the innermost N-linked core GlcNAc and the Asn residue to which it is covalently linked. PNGase F deamidates the N-linked Asn producing an Asp residue, with a resulting increase of 0.984 Da in molecular weight (Gonzalez et al., 1992). PNGase F-treated peptides were desalted by C18 cartridge solid-phase extraction. The percent occupancy for each site is calculated by comparing the extracted chromatographic area under the curve of peptides with Asn to those with Asp using Byonic software (Version 3.10; Protein Metrics Inc.).

Mass Spectrometry Analysis

The peptides were reconstituted in 0.1% formic acid–water solution and analyzed on an Orbitrap Eclipse Tribrid mass spectrometer equipped with a nanospray ion source and connected to a Dionex binary solvent system (Thermo Fisher Scientific). Peptides were separated using an Acclaim™ PepMap™ 100 C18 Column (75 μm × 15 cm). A trapping column (PepMap 100 C18 3 μM 75 μM × 2 cm) was used in line with the LC prior to separation with the analytical column. The solvent system consisted of solvent A (100% water/0.1% formic acid) and solvent B (100% ACN/0.1% formic acid). The LC conditions were as follows: 5–35% of solvent B for 165 min, 90% of solvent B for 5 min, and 1% of solvent B for 5 min. The flow rate was set to 300 nl/min. The spray voltage was set to 2.7 kV, and the ion-transfer tube temperature was set to 275°C. The full MS scan range was 400–2000 m/z. Precursor masses were detected in the Orbitrap at resolution (R) = 120,000 (at m/z 200). Stepped HCD (higher-energy collisional dissociation) spectra (HCD energy at 15, 25, and 35%) were recorded for the top 15 most abundant precursors with the standard mode of the AGC target. Dynamic exclusion was set at 30 s. If at least one typical glycan fragment ion abundance (m/z 204.0867 and 366.1396 Da) was observed within the top 15 most abundant fragments and within a 15 ppm mass accuracy, an EThcD [electron-transfer dissociation (ETD) followed by supplemental HCD collision energy at 25%] spectrum of the same precursor would be recorded in the Orbitrap at R = 15,000. The ETD reaction time was set to use calibrated charge-dependent ETD parameters. The glycopeptides of the intact virus were

analyzed by stepped HCD fragmentation and HCD-triggered EThcD fragmentation to analyze N-linked glycans and O-linked glycans, respectively. Deamidated peptides were analyzed only by stepped HCD fragmentation.

Data Analysis Using Byonic and Manual Verification

The LC-MS/MS spectra were searched against the FASTA sequence of the spike protein of the SARS-CoV-2 original virus or the D614G variant using Byos™ (Version 3.10; Protein Metrics Inc.). The searching parameters were specified as follows: fully specific digestion specificity, 2 missing cleavage sited allowed, carbamidomethyl at cysteine as a fixed modification, and oxidation at methionine as a common modification. The precursor ion mass tolerance was set at 6 ppm, and that for fragment ions was at 20 ppm. A 1% false discovery rate (FDR) was applied. The results were filtered with PEP 2D < 0.01, score ≥100, and Delta Mod. Score ≥10. The glycopeptide fragmentation data were evaluated manually for each glycopeptide; the peptide was confirmed when the b and y fragment ions were observed along with oxonium ions corresponding to the glycan identified. A minimum of 3 b/y ions were required. The relative amounts of each glycan at each site were determined by comparing the extracted chromatographic area under the curve. All charge states for a single glycopeptide were summed.

Glycans were categorized to subtypes according to the composition detected: Hex (4–10)HexNAc(2) was classified as high mannose, NeuAc (0–1)dHex (0–1)Hex (3–7)HexNAc(3) was classified as Hybrid, and NeuAc (0–5)dHex (0–3) Hex (3–8)HexNAc(4–7) was classified as complex-type glycans. Any glycan containing at least one fucose or sialic acid was counted as fucosylated or sialylated, respectively.

Model Construction

Monomeric structural models of N-linked glycan presentation on SARS-CoV-2 were created using electron microscopy structures (PDB ID: 6ZGG), which were visualized with CCP4MG. Glycan cartoon structures are inferred from knowledge of common glycans as identification was done solely by mass. A trimeric structural model of SARS-CoV-2 was created from the electron microscopy structure (PDB ID: 7A96) and visualized with CCP4MG. The antigenic epitopes were predicted using NetCTL-1.2 (Larsen et al., 2007).

RESULTS AND DISCUSSION

Mapping Glycosylation on Recombinant RBD Proteins

Recombinant RBD proteins expressed in HEK293 cells were trypsin-digested. Seventy-five percent of the digest was used for glycopeptide enrichment using HILIC separation, and 25% was deglycosylated in preparation for glycosylation site occupancy analysis. The HILIC-enriched intact glycopeptides were analyzed by LC-MS/MS using HCD-triggered EThcD fragmentation. The deglycosylated peptides were analyzed by LC-MS/MS with HCD fragmentation. The LC-MS/MS raw files were analyzed using Byonic software. The Byonic results were filtered with a 1% false discovery rate and other parameters to achieve high confidence identifications (see the *Method* section). All the spectra were manually verified.

The RBD has two potential N-linked glycosylation sites at amino acid positions 331 and 343 relative to the WA1 spike protein. Our data show that both sites are heavily glycosylated with greater than 99% occupancy (**Figure 1** and **Supplementary Table S1**). We observed a high degree of fucosylation at the two N-glycosites, and Man5GlcNAc2 (Man5) is highly abundant at both sites (**Figure 1** and **Supplementary Table S1**). Glycans identified at N331 included high mannose; short complex, paucimannose; and highly abbreviated forms (**Figure 1**; **Supplementary Table S1** and **Supplementary Figure S1**). The reason is not fully understood but may be related to prompt fragmentation or degradative processes incurred during RBD production and/or purification. Prompt decay is unlikely as no other glycosites demonstrated this pattern.

We also identified two O-glycosylation sites at residues T323 and T523 with a diverse range of glycan compositions. Interestingly, most glycans at the two O-glycosites contain sialic acid (**Figure 1** and **Supplementary Table S2**). Glycosylation of T323, but not T523, has been previously reported. Therefore, we carefully examined the spectra and observed strong evidence of glycosylation at T523 (**Supplementary Figure S2**). Previous studies reported O-glycosylation at T325 (Shajahan et al., 2020; Zhao et al., 2020), although the occupancy was estimated to be low (Zhao et al., 2020). However, our data did not show direct evidence of fragment ions which can confirm that T325 is glycosylated.

Site-Specific Microheterogeneity of Spike Glycosylation in Recombinant S1 Proteins

The recombinant S1 protein expressed by HEK293 cells was treated according to the same protocol as the RBD protein (see above), except that two enzymes were used for digestion to facilitate glycoproteomic coverage of the protein. These two enzymes were trypsin and chymotrypsin, used in separate digestion. Byonic search parameters and filters were also the same as for the RBD protein.

The gene encoding the S1 domain has 13 possible sites of N-glycosylation. Twelve of the 13 predicted N-glycosites were found to be extensively glycosylated (**Supplementary Figure S3** and **Supplementary Table S3**). The one missing glycosite, N17, was detected glycosylated, but it did not meet our criteria due to low confidence scores. Although the scores are low, many hybrid and complex-type glycans were detected at N17 with at least two technical replicates. The site occupancy for 10 glycosites is greater than 90%. Sites N149 and N657 had a site occupancy rate of 25 and 58%, respectively (**Supplementary Figure S3** and **Supplementary Table S3**).

We observed a diverse range of glycan compositions across the N-linked glycosylation sites. Glycosites N331, N343, N603, and N616 had less glycan variety, while those at N122, N165, N234,

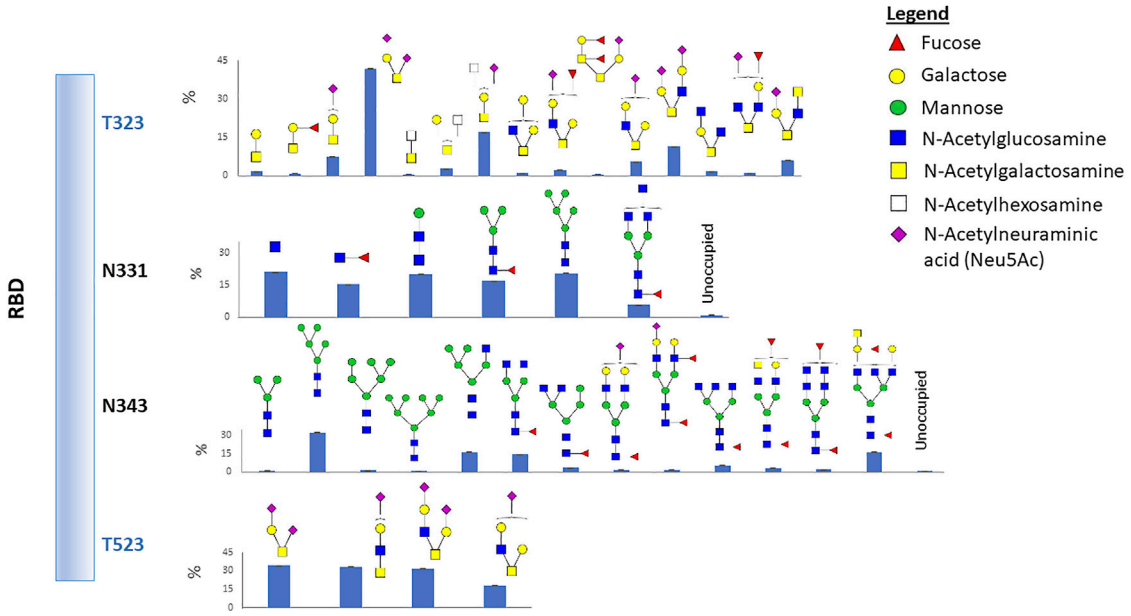


FIGURE 1 | Glycosylation profile on the recombinant RBD protein. Two N-glycosites and two O-glycosites were identified. Glycan cartoon structures are inferred from knowledge of common glycans as identification was done solely by mass. The bar graphs represent the glycan abundance and unoccupied percent based on total ion abundances at each site.

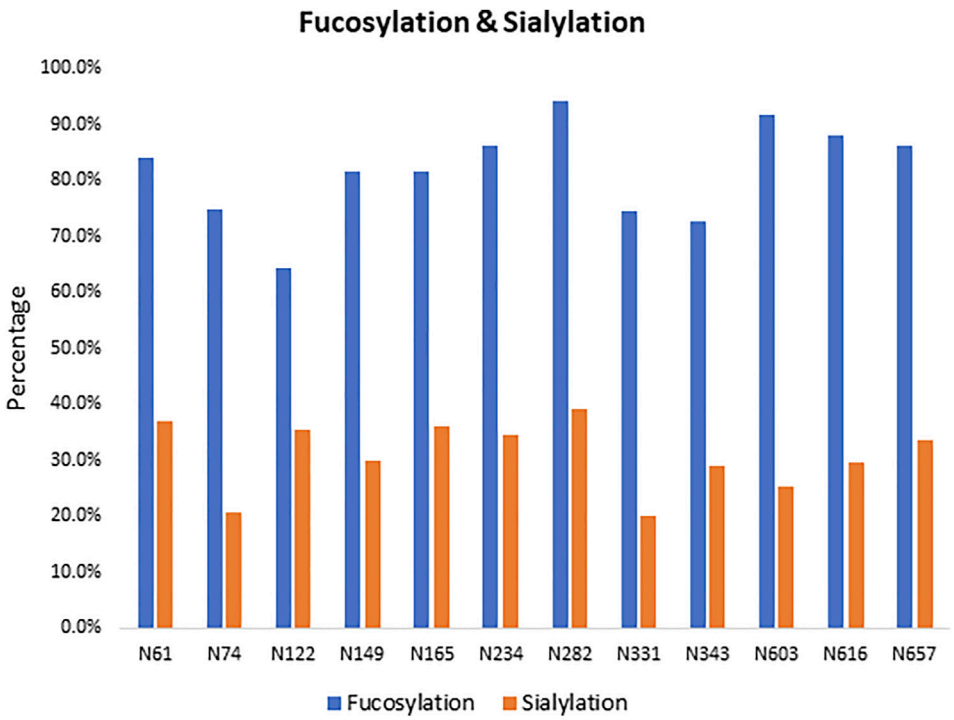


FIGURE 2 | Fucosylated and sialylated N-linked glycosylation of the recombinant S1 protein. Twelve of 13 potential N-glycosylation sites were found occupied, and these N-linked glycans are highly fucosylated.

N282, and N657 were more diverse (**Supplementary Figure S3** and **Supplementary Table S3**). In addition to the site-specific glycan compositions, overall trends in glycosylation across sites were examined. The results revealed that the three most common types of N-glycans were Man5GlcNAc2 (Man5), HexNAc4Hex3Fuc1, and HexNAc5Hex3Fuc1 (**Supplementary**

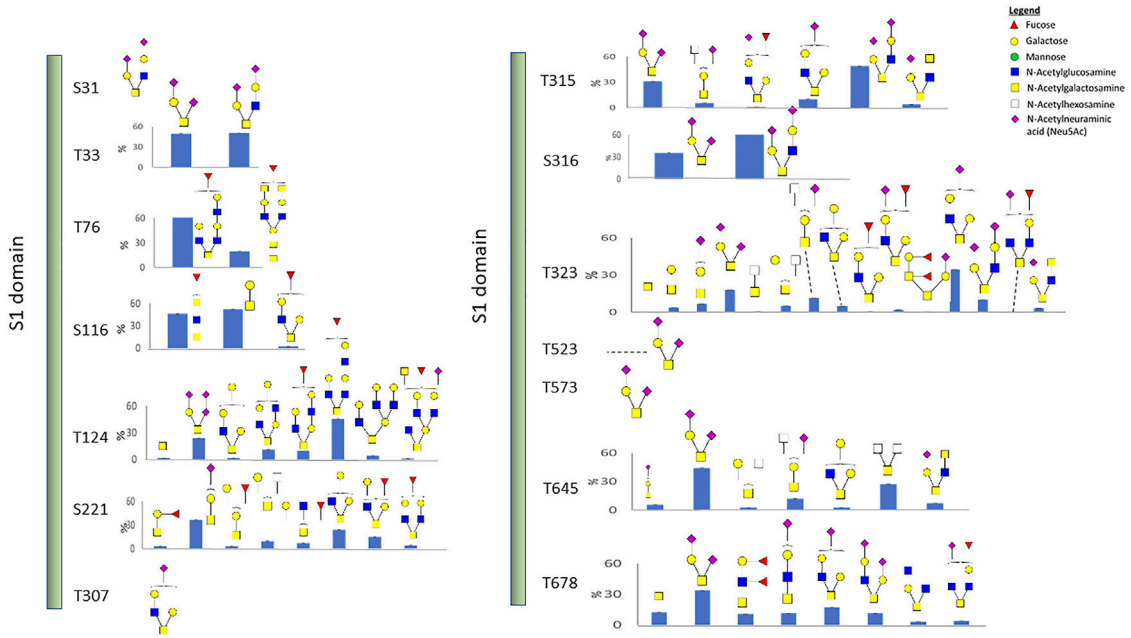


FIGURE 3 | O-linked glycosylation of the recombinant S1 protein. Glycan cartoon structures are inferred from knowledge of common glycans as identification was solely based on mass.

Figure S3). Man5 has also been reported by others as a predominant high-mannose glycan composition across all N-glycosites on the SARS-CoV-2 spike protein when produced in HEK293 cells but interestingly not in CHO cells (Zhao et al., 2020; Wang et al., 2021b). The relative abundance of complex-type glycans and the level of fucosylation and sialylation for each site were examined. As shown in **Figure 2**, the N-linked glycans on the S1 protein were both highly fucosylated (~80%) and sialylated (~30%) with overlap where both substitutions were observed on some glycans.

We also identified 14 O-linked glycosites on the recombinant S1 protein, including the two sites, T323 and T523, which were identified in the recombinant RBD protein. O-glycosylation has been reported to function in immunological shielding, protein stability, and regulation of conformational changes (Casalino et al., 2020a). About half of the 14 sites have not been reported before, and most glycosites display a variety of glycan modifications (**Figure 3** and **Supplementary Table S4**). Three O-glycosites, S673, T678, and S686, are located in the furin cleavage region. Thus, such glycans may modulate the SARS-CoV-2 entry (Andersen et al., 2020). Of the three hypothesized O-glycosites, T678 was identified in this study.

The experimentally observed glycosylation sites are illustrated on the monomeric SARS-CoV-2 spike glycoprotein (PDB code 6ZGG) (**Figure 4**). To convey the main processing features at each site, the abundances of each glycan were summed by glycan subtype and displayed as a pie chart next to each site. We observed a combination of high-mannose, hybrid, and complex-type N-glycans for most of the sites. Overall, all glycosites were dominated by complex-type glycans when tabulated by subtype. N74 displayed more hybrid-type glycans (30%). N343, in the RBD region, displayed a higher relative amount of mannose-type glycans (28%). This observation aligned with our observations in the recombinant RBD protein (see **Figures 1**, **4**).

To illustrate the possible impact of the glycosylation microheterogeneity on the virus antigenicity, we mapped the N-glycosites with antigenic sites and the receptor-binding motifs to the SARS-CoV-2 trimer using a 3D model previously determined by electron microscopy (PDB code 7A96) (**Figure 5**). The data show extensive microheterogeneity across the glycosites. The number of identified glycoforms at each site ranged from 12 to 83. The antigenic epitopes were predicted using NetCTL-1.2 (Larsen et al., 2007) (**Supplementary Table S5**). We found that many occupied glycosites are close to, or even overlap with, the antigenic epitopes. Those that overlapped with antigenic sites included N165, N343, N616, and N657, which display substantial glycan diversity (**Figure 5**). The 3D model has one open RBD bound to the ACE2 protein. The shielding of receptor-binding sites by glycans is a common feature of viral glycoproteins and has been observed for the SARS-CoV spike, the HIV-1 envelope, and influenza hemagglutinin (Bonomelli et al., 2011; Stewart-Jones et al., 2016; An et al., 2019; Zhao et al., 2020).

There are two states of RBD: the "down" conformation and the "up" conformation, corresponding to the receptor-inaccessible state and receptor-accessible state, respectively (Gui et al., 2017; Walls et al., 2019; Wrapp et al., 2020). The modeling reveals that N343, N234, and N165 are near to the receptor-binding motif [limited to amino acids 438–506 (Zhou et al., 2021)]. Previous structure analysis revealed that in the RBD "down" state, the RBD region is shielded by the glycans at N343, N165, and N234

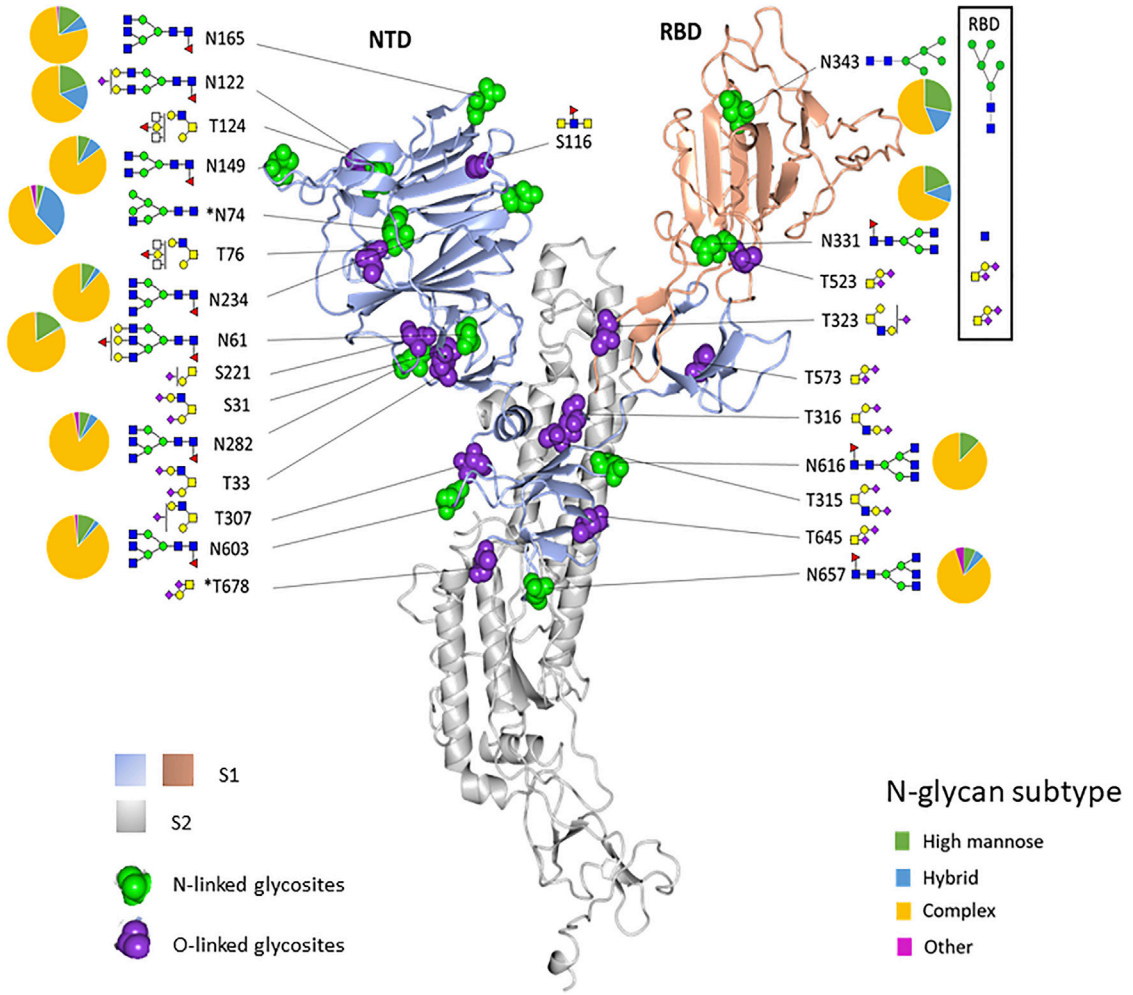


FIGURE 4 | Structure-based mapping of glycans on the recombinant SARS-CoV-2 spike protein. The modeling of experimentally observed glycosylation site compositions is illustrated on the monomeric SARS-CoV-2 spike glycoprotein (PDB code 6ZGG). The S1 subunit is colored light blue and peach. The S2 subunit is gray. N- and O-linked glycosylation sites are indicated by green balls and purple balls, respectively. Most abundant glycans at each site are shown. Pie charts show the percentage of glycan subtypes at each site. The boxed area shows the predominant glycans and the N-linked glycosylation subtype distribution for the glycans identified in the recombinant RBD sample. *N74 and T678 are not in the structure.

(Casalino et al., 2020b). Besides shielding, the glycans at N165 and N234 have also been reported to stabilize the RBD in the "up" conformation (Casalino et al., 2020b). Sztain et al. revealed that the receptor-binding motif is consistently shielded by the glycans at N165 and N234, but RBD opening decreases shielding by the glycans at N343 (Sztain et al., 2021). The N343 glycan may play the role of the "glycan gate" by facilitating conformational shift of the RBD from the "down" to the "up" conformation by interacting with the residues of the ACE2-binding motif (Sztain et al., 2021).

Site-Specific Glycosylation of the Spike From the WA1 Strain and the D614G Variant

To determine the differences and similarities in glycosylation between the recombinant S1, produced in HEK293 cells, and that of the spike produced in the virus, we examined the spike derived from the intact virus from two strains, the WA1 strain and D614G, propagated in Vero E6 cells. The WA1 strain was from the first patient in the United States who was diagnosed with SARS-CoV-2 viral infection. This case was declared by the state of Washington and CDC on January 20, 2020 (Harcourt et al., 2020). This viral identity was confirmed by whole genome sequencing (GenBank accession no. MN985325), and it did not have mutation at the 614 amino acid. The D614G variant contains the spike protein amino acid change at 614 from D to G, which is more infectious and transmissible and has become the most prevalent form in the global pandemic since March 2021 (Hou et al., 2020; Korber et al., 2020). Both viruses were grown in Vero E6 cells, and the supernatant of the passage 4 stock of each virus was collected, inactivated by gamma irradiation, and analyzed by our glycoproteomics approach.

Of the 13 predicted N-linked glycosites in the S1 domain, 10 N-glycosites were identified in the WA1 strain (**Figure 6A** and

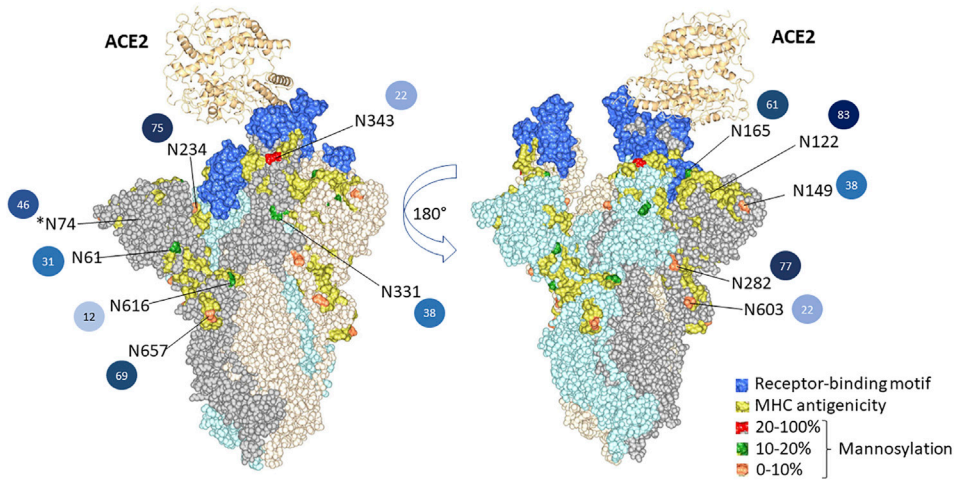


FIGURE 5 | 3D structural modeling of spike glycosylation microheterogeneity. The N-glycosites are mapped on the SARS-CoV-2 trimer structure (PDB code 7A96). Blue indicates the receptor-binding motif in the RBD region. Yellow indicates the predicted MHC antigenic sites. The glycosites are colored according to the mannosylation percentage. The number of glycoforms at each site from less to more heterogeneous glycoforms detected is colored by light to dark, and the number is also listed.

Supplementary Table S6). The two N-glycosites, N603 and N616, were identified with several high-mannose-type glycans (Man7GlcNAc2, Man8GlcNAc2, and Man9GlcNAc2 at N603 and Man8GlcNAc2 at N616) in a single replicate of the WA1 sample, which did not meet our criteria where two replicates were required to achieve confident identification. Therefore, glycosylation at these sites was considered tentative and not considered further. In contrast, 12 of 13 N-linked glycosites were identified in the S1 domain of the D614G variant (**Figure 6A** and **Supplementary Table S7**). Site occupancy identified by the PNGase F deglycosylation methodology revealed that 10 S1 N-glycosites from the WA1 strain (N61, N122, N165, N234, N282, N331, N343, N603, N616, and N567) and 9 N-glycosites from the D614G variant (N61, N149, N165, N234, N331, N343, N603, N616, and N657) were almost 100% glycosylated (**Supplementary Tables S8, S9**). Three N-glycosites, N74, N122, and N282, were only identified in D614G with a single replicate in the site occupancy study; therefore, we were not able to determine site occupancy at these three sites. (**Supplementary Table S9**). We do note, however, that these peptides are at least partially occupied by high-mannose glycans based on our glycopeptide analysis (**Figure 6A** and **Supplementary Table S7**). Likewise, some N-glycosites were identified with highly diverse glycan compositions upon glycoproteomics analysis of intact glycopeptides. However, occupancy analysis at sites such as N149 in the WA1 strain and N122 in D614G strains did not meet our criteria. Conversely, no glycopeptides were identified at N603 and N616 of the WA1 strain, but these two sites were identified as occupied based on detection of Asp in place of Asn subsequent to PNGase F digestion, which supports that the two sites were glycosylated. Estimated occupancies were between 96.4 and 100%, making it unlikely that spontaneous deamination of unoccupied sites was solely responsible for the Asp presence at the site. The most likely reason for the discrepancy is due to the random sampling issue of mass spectrometry or incomplete enrichment for these very complex samples. The samples of intact viruses contained less than 5% of spike protein abundance according to the protein quantitation by Byonic software. High-abundance glycopeptides of host cells were also enriched using the HILIC column, which resulted in the high complexity of the glycopeptide pool in this experiment.

When comparing the mutant form, D614G, with the original form, WA1, we observed a similar glycosylation pattern for most N-linked glycosites in both S1 and S2 domains (**Figures 6A,B**). The most abundant glycoform at each N-glycosite was comparable between WA1 and D614G (**Figure 7**). Man7GlcNAc2 and Man8GlcNAc2 were the most abundant glycoforms for the majority of the sites, except for N343. There are several sites showing different glycan contents between the two strains, such as N331, N343, and N1074. The D614G variant presents more complex-type glycans at N331 but less complex-type glycans at N343 compared to the WA1 strain. As mentioned earlier, the N343 glycans significantly affect the RBD "up" conformation (Sztain et al., 2021). The glycan changes at N343 in D614G compared to WA1 could, at least partially, account for D614G phenotype changes if similar shifts in glycosylation occur in nature.

In addition, four glycosites, N603, N616, N1158, and N1194, were not identified in the WA1 strain, while all four were identified in the D614G variant (**Figures 6A,B**, **Supplementary Tables S6 and S7**). This may not mean that these sites are not glycosylated in WA1. Their absence may have resulted from sample complexity, random sampling, and limitations of our enrichment strategy as discussed above. This hypothesis is supported by the site occupancy analysis where WA1 spike N603, N616, and N1194 sites were clearly occupied (**Supplementary Table S8**).

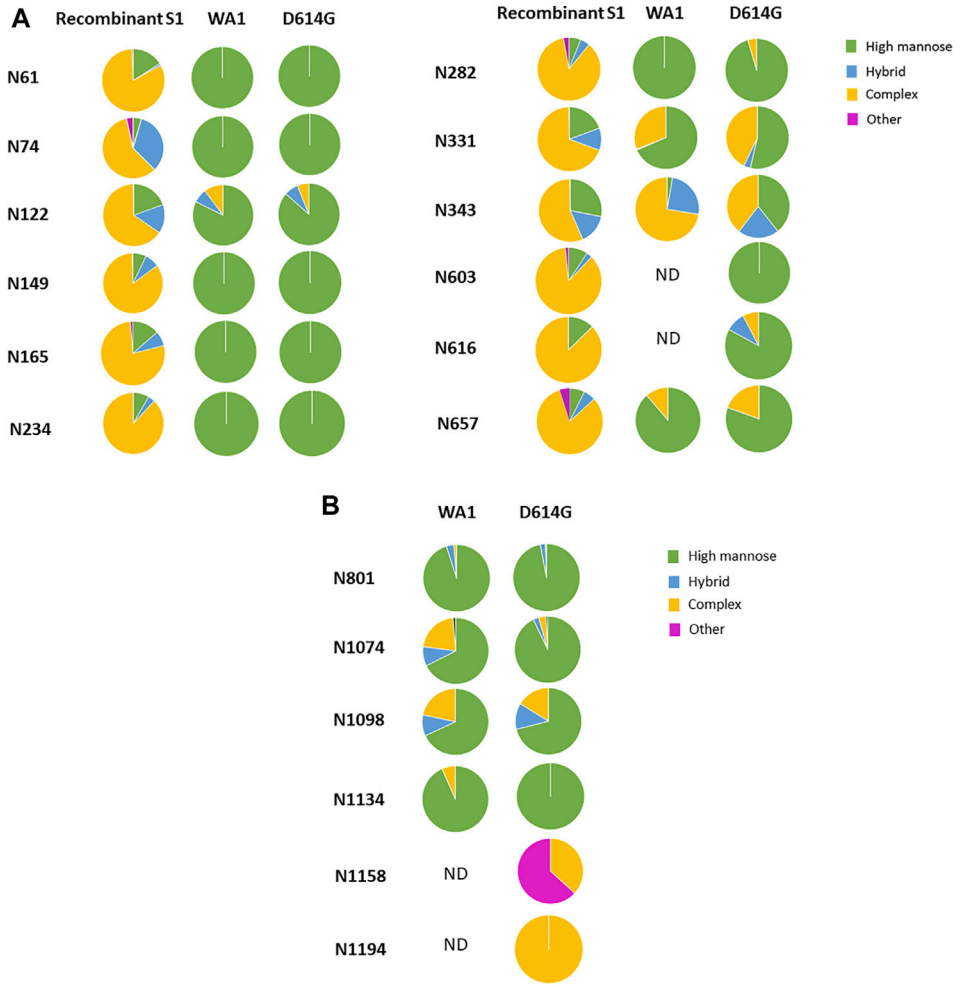


FIGURE 6 | Comparison of the glycosylation pattern on the spike protein from recombinant S1, the WA1 strain, and the D614G variant. The most abundant glycoforms detected in Vero E6 cell-derived WA1 and D614G strain spike N-glycosites were comparable but different from that detected in recombinant S1 produced in HEK293. **(A)** N-linked glycan subtypes in the S1 domain. **(B)** N-linked glycan subtypes in the S2 domain.

Interestingly, in the recombinant S1 protein, most N-linked glycosites are dominated by complex-type N-glycans, while most glycosites in both the WA1 strain and the D614G variant produced in the virus were dominated by high-mannose-type glycans (**Figure 6A**). No O-linked glycosites were identified in the virus-derived spike from WA1 and D614G. The reason for the observed dramatic differences in the glycosylation pattern detected between recombinant spike S1 and the virus-derived spike is not clear. There are several possibilities: first, protein structural differences may influence access to glycosylation machinery. Second, the expression of glycosylation enzymes may differ between cell substrates. Third, the secretory location of glycosylation enzymes may differ between cell substrates under conditions of expression and/or virus propagation. WA1 and D614G strains were grown in Vero E6 cells which are derived from the African green monkey kidney, while the recombinant S1 protein was expressed by HEK293 cells which are derived from human embryonic kidney cells.

It is clear that S1 and the whole spike differ structurally as the former is without the S2 subunit. Several publications reported differences in glycosite occupancy and glycan composition between the intact spike protein and individually expressed S1 and S2 proteins (Watanabe et al., 2020a; Watanabe et al., 2020b; Shajahan et al., 2020; Zhao et al., 2020). The spike expressed in FreeStyle293F cells was found to be partially expressed as the S0 form, without S1/S2 or S2′ cleavage. The S0 form was found to primarily contain high-mannose glycans (Watanabe et al., 2021). We searched our data for evidence of S1/S2 cleavage. We did detect cleavage of the S1/S2 furin cleavage site in the chymotrypsin digest. However, the intensity of these peptides was low, suggesting that significant amounts of the uncleaved S0 forms were present. We must also note that no peptides representing the uncleaved site were detected. It has been reported that Vero E6 cells do not produce high-abundance furin and cleavage of the S glycoprotein in SARS-CoV-2-infected Vero E6 cell lysates was reported to be inefficient (Klimstra et al., 2020). One study, using human serum to detect SARS-CoV-2 proteins produced in infected Vero E6 cell lysates, showed mainly an uncleaved S protein (Haveri et al.,

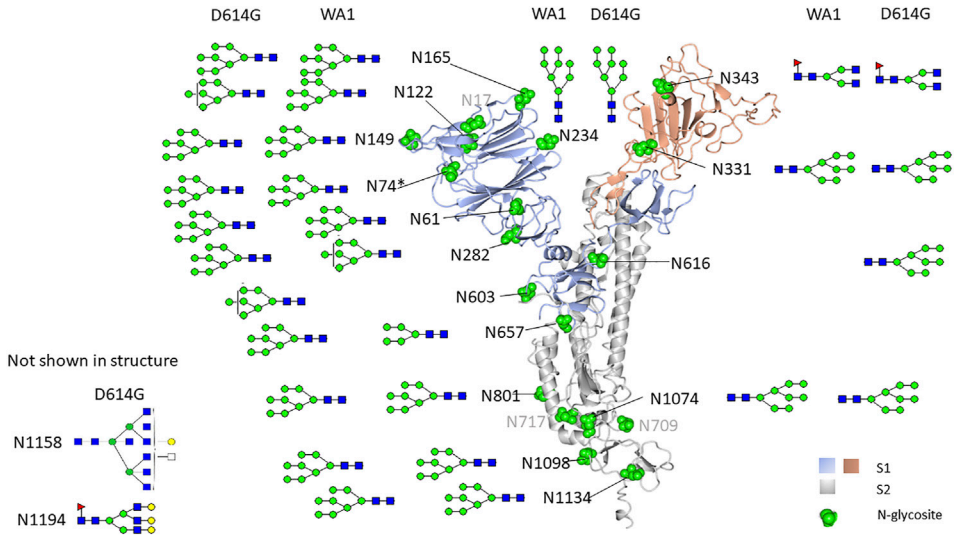


FIGURE 7 | N-glycosylation map of Vero E6-derived viruses. The most abundant glycoform detected in Vero E6 cell-derived WA1 and the D614G strain spike N-glycosite was comparable but different from that detected in recombinant S1 produced in HEK293.

2020). Additionally, in our case, there was also no evidence for TMPRSS2 or cathepsin cleavage. Our observations, however, may have been due to low abundance of the spike in our samples; thus, peptides specifically containing these cleavage sites may not have been detected.

In general, Vero cells are not known to limit glycan processing primarily to high-mannose glycans. Vero cells have been used as a cell substrate for propagation of influenza and recombinant proteins without report of bias toward high-mannose glycans (Gornik et al., 2012; Rödig et al., 2013). The nascent capacity of Vero cell expression and the secretory location of glycosylation enzymes should not be an issue. Therefore, our observations are not likely to be due to inherent limitations of Vero cells in terms of glycosylation processes. However, our data show that high-mannose-type glycans represent a large portion of total glycans displayed on the Vero E6 host's glycoproteins (**Supplementary Figure S4**). A range of complex glycans were also identified, albeit with far less abundance. Therefore, the high percent of high-mannose-type glycans on WA1 and D614G grown in Vero E6 cells was not limited to the SARS-CoV-2 spike.

The proper location of glycosylation enzymes is a complex process involving Rab GTPases, coiled-coil tethers termed golgins, and the multi-subunit tethering complex known as the conserved oligomeric Golgi (COG) complex (Fisher and Ungar, 2016). These factors contribute toward anterograde and retrograde transport of glycosylation active enzymes and other necessary proteins involved in glycoprotein production. Regulation of these processes is essential for appropriate localization and sequential activities of glycosylation active enzymes (Starr et al., 2010). In our studies of the Vero cell-propagated SARS-CoV-2 spike, we noted a low amount of glycosylation processing beyond ER mannosidase I (Moremen et al., 2012) and other mannosidases which are normally present in the cis/medial cisternae of the Golgi (Moremen et al., 2012). This was evidenced by the dominant presence of primarily Man5-8GlcNAc2. There were only low abundances of Man3-5GlcNAc3, also suggesting little processing by cis/medial Golgi located N-acetylglucosaminidase I (Gnt1) (Sztain et al., 2021; Moremen et al., 2012). Therefore, one possibility is that under conditions of viral propagation, the Golgi COG system-mediated anterograde/retrograde system is shifted or viral packaging and routing differs from normal secretion, resulting in an altered distribution of glycosylation active enzymes or proper sequential exposure of these enzymes to nascent glycoproteins. Notably, virus-like particles and the SARS-CoV-2 virus have been localized to the endoplasmic reticulum–Golgi intermediate compartment (ERGIC), a site of secretory sorting between the ER and Golgi, and it has been hypothesized that SARS-CoV-2 exits the cell *via* lysosomal exocytosis, suggesting little exposure to Golgi enzymes (Mendonça et al., 2021; Plescia et al., 2021). We note that among 25 cell lines tested, Vero E6 produced among the highest viral titers including all those expressing the human ACE2 receptor. Therefore, the high-mannose glycan distribution does not appear to significantly negatively affect viral propagation in the Vero E6 cell line compared to alternative cell substrates typically used in the SARS-CoV-2 viral study (Wang et al., 2021a).

Watanabe et al. (2021), who also noted unprocessed glycans on the spike, albeit produced in HEK cells expressing the ChAdOx1 vaccine vector, hypothesized that these high-mannose-bearing spike proteins represented those in transit through the secretory system and suggested that furin protease is located in the later trans-Golgi stacks. In our case, this is unlikely as the majority of the virus isolated formed mature viral particles. Significantly, both HEK293 and Vero cells produce predominantly high-mannose glycosylation patterns on the SARS-CoV-2 spike under certain circumstances. The exact reason for this remains an open question.

Overall, we have found similar glycosylation site-specific N-glycan distributions in S1 and RBD to those previously reported that were produced in HEK293 cell lines. We also report here previously unreported O-glycosylation site occupancy including T523 and confirm the presence of 14 total O-glycosylation site occupancies including T678, which appears in the furin cleavage domain. Significantly, we also report that the native spike produced in SARS-CoV-2/USA-WA1/2020 (USA-WA1) is substituted with primarily high-mannose glycans that do not appear to effect viral propagation in Vero E6 compared to alternative cell substrates (Wang et al., 2021a).

CONCLUSION

In this study, we characterized the site-specific glycosylation of the spike protein from recombinant RBD and S1 domains and from two intact viruses, the WA1 strain and the D614G variant. Glycosylation was found to be of high occupancy in all samples examined and highly heterogeneous across the majority of glycosites in the HEK293-derived S1 and RBD. Glycan modification at most N-glycosites is very similar between WA1 and D614G and primarily high-mannose, with significant differences at N343. Our results also revealed different patterns of glycan modification among the recombinant S1 protein, recombinant RBD, and the WA1 and D614G strains, which implies that these spike proteins may perform differently *in vitro* and *in vivo*. Therefore, the origin of spike glycosylation should be put in consideration for vaccine design and drug development.

AUTHOR SUMMARY

The SARS-CoV-2 virus spike protein binds to host cells, fuses with the host cell membrane, and enters the cell. It is heavily glycosylated, and recent studies revealed that glycan modification is essential to modulate spike conformation and host cell invasion. In this study, we analyzed the glycan modification of recombinant spike protein subunit RBD and the S1 domain, both of which function to bind host receptor ACE 2. We also analyzed the glycan modification of whole viruses, the WA1 strain, and the D614G variant. The WA1 strain was isolated from the first case of COVID-19 in the United States. The D614G variant, carrying the protein amino acid change at 614 from aspartate(D) to glycine(G), is now prevalent in the circulating SARS-CoV-2 virus and is carried by all recently identified and highly concerning SARS-CoV-2 variants. We found different patterns of glycan modification among the recombinant S1 protein, recombinant RBD, and WA1 and D614G strains. Glycan modification at most N-glycosites is very similar between WA1 and D614G, with significant differences at N343. This recombinant S1 and RBD glycosylation patterns differ dramatically from the whole virus produced in Vero cells and implies that these spike proteins may perform differently *in vitro* and *in vivo*, which could have implications for viral studies, vaccine design, and drug development.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD029218.

AUTHOR CONTRIBUTIONS

JC served as the principal investigator and conceived, designed, and directed this research. YT designed and performed mass spectrometry experiments and served as the primary author. LP performed additional mass spectrometry experiments and participated in manuscript construction. EJ contributed toward experimental design and construction of the manuscript.

FUNDING

This work was funded by FDA Program: Z01 BJ 02044-13 LBP to JC.

ACKNOWLEDGMENTS

We thank Natalie Thornburg and Azaibi Tamin (Centers for Disease Control and Prevention) for supply of the WA1 and D614G virus preparations, their insight, comments, and contributions toward this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.767448/full#supplementary-material

Supplementary Figure S1 | Spectra of glycopeptides with short complex glycans identified at N331 from the recombinant RBD protein.

Supplementary Figure S2 | Spectra of O-linked glycopeptides identified at T523 from the recombinant RBD protein.

Supplementary Figure S3 | Relative abundances of all N-linked glycans detected at each site of the recombinant S1 protein. Twelve of 13 potential N-glycosylation sites were found occupied and are displayed. The displayed value is the mean abundance of each glycan detected across three biological replicates normalized to total ion abundance at each site.

Supplementary Figure S4 | Glycan profiles identified from glycoproteins of Vero E6 cells. **(A)** N-glycans identified from WA1 trypsin and the GluC digest. **(B)** N-glycans identified from the WA1 chymotrypsin digest.

REFERENCES

An, Y., Parsons, L. M., Jankowska, E., Melnyk, D., Joshi, M., and Cipollo, J. F. (2019). N-glycosylation of Seasonal Influenza Vaccine Hemagglutinins: Implication for Potency Testing and Immune Processing. *J. Virol.* 93 (2). doi:10.1128/JVI.01693-18

An, Y., and Cipollo, J. F. (2011). An Unbiased Approach for Analysis of Protein Glycosylation and Application to Influenza Vaccine Hemagglutinin. *Anal. Biochem.* 415 (1), 67–80. doi:10.1016/j.ab.2011.04.018

Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., and Garry, R. F. (2020). The Proximal Origin of SARS-CoV-2. *Nat. Med.* 26 (4), 450–452. doi:10.1038/s41591-020-0820-9

Bonomelli, C., Doores, K. J., Dunlop, D. C., Thaney, V., Dwek, R. A., Burton, D. R., et al. (2011). The Glycan Shield of HIV Is Predominantly Oligomannose Independently of Production System or Viral Clade. *PLoS One* 6 (8), e23521. doi:10.1371/journal.pone.0023521

Casalino, L., Gaieb, Z., Goldsmith, J. A., Hjorth, C. K., Dommer, A. C., Harbison, A. M., et al. (2020). Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* doi:10.1021/acscentsci.0c01056

Harcourt, J., Azaibi, T., Xiaoyan, L., Shifaq, K., Senthil, K. S., Queen, M., et al. (2020). Severe acute respiratory syndrome coronavirus 2 from patient with coronavirus disease, United States. *Emerging Infectious Diseases* 26 (6), 1266. doi:10.1021/acscentsci.0c01056

Coutard, B., Valle, C., de Lamballerie, X., Canard, B., Seidah, N. G., and Decroly, E. (2020). The Spike Glycoprotein of the New Coronavirus 2019-nCoV Contains a Furin-like Cleavage Site Absent in CoV of the Same Clade. *Antiviral Res.* 176, 104742. doi:10.1016/j.antiviral.2020.104742

Fisher, P., and Ungar, D. (2016). Bridging the Gap between Glycosylation and Vesicle Traffic. *Front. Cel Dev. Biol.* 4, 15. doi:10.3389/fcell.2016.00015

Gonzalez, J., Takao, T., Hori, H., Besada, V., Rodriguez, R., Padron, G., et al. (1992). A Method for Determination of N-Glycosylation Sites in Glycoproteins by Collision-Induced Dissociation Analysis in Fast Atom Bombardment Mass Spectrometry: Identification of the Positions of Carbohydrate-Linked Asparagine in Recombinant α-amylase by Treatment with Peptide-N-Glycosidase F in 18O-Labeled Water. *Anal. Biochem.* 205 (1), 151–158. doi:10.1016/0003-2697(92)90592-u

Gornik, O., Pavić, T., and Lauc, G. (2012). Alternative Glycosylation Modulates Function of IgG and Other Proteins - Implications on Evolution and Disease. *Biochim. Biophys. Acta (Bba) - Gen. Subjects* 1820 (9), 1318–1326. doi:10.1016/j.bbagen.2011.12.004

Grant, O. C., Montgomery, D., Ito, K., and Woods, R. J. (2020). 3D Models of Glycosylated SARS-CoV-2 Spike Protein Suggest Challenges and Opportunities for Vaccine Development. *bioRxiv*. doi:10.1101/2020.04.07.030445

Gui, M., Song, W., Zhou, H., Xu, J., Chen, S., Xiang, Y., et al. (2017). Cryo-electron Microscopy Structures of the SARS-CoV Spike Glycoprotein Reveal a Prerequisite Conformational State for Receptor Binding. *Cell Res* 27 (1), 119–129. doi:10.1038/cr.2016.152

Harcourt, J., Tamin, A., Lu, X., Kamili, S., Sakthivel, S. K., Murray, J., et al. (2020). Severe Acute Respiratory Syndrome Coronavirus 2 from Patient with Coronavirus Disease, United States. *Emerg. Infect. Dis.* 26 (6), 1266–1273. doi:10.3201/eid2606.200516

Haveri, A., Smura, T., Kuivanen, S., Österlund, P., Hepojoki, J., Ikonen, N., et al. (2020). Serological and Molecular Findings during SARS-CoV-2 Infection: the First Case Study in Finland, January to February 2020. *Euro Surveill.* 25 (11). doi:10.2807/1560-7917.ES.2020.25.11.2000266

Hoffmann, M., Kleine-Weber, H., and Pöhlmann, S. (2020). A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. *Mol. Cel* 78 (4), 779–784. doi:10.1016/j.molcel.2020.04.022

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* 181 (2), 271–280. doi:10.1016/j.cell.2020.02.052

Hou, Y. J., Chiba, S., Halfmann, P., Ehre, C., Kuroda, M., Dinnon, K. H., 3rd, et al. (2020). SARS-CoV-2 D614G Variant Exhibits Enhanced Replication *Ex Vivo* and Earlier Transmission *In Vivo*. *bioRxiv*. doi:10.1101/2020.09.28.317685

Klimstra, W. B., Tilston-Lunel, N. L., Nambulli, S., Boslett, J., McMillen, C. M., Gilliland, T., et al. (2020). SARS-CoV-2 Growth, Furin-Cleavage-Site Adaptation and Neutralization Using Serum from Acutely Infected Hospitalized COVID-19 Patients. *J. Gen. Virol.* 101 (11), 1156–1169. doi:10.1099/jgv.0.001481

Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., et al. (2020). Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* 182 (4), 812–827. doi:10.1016/j.cell.2020.06.043

Larsen, M. V., Lundegaard, C., Lamberth, K., Buus, S., Lund, O., and Nielsen, M. (2007). Large-scale Validation of Methods for Cytotoxic T-Lymphocyte Epitope Prediction. *BMC Bioinformatics* 8, 424. doi:10.1186/1471-2105-8-424

Li, Y., Liu, D., Wang, Y., Su, W., Liu, G., and Dong, W. (2021). The Importance of Glycans of Viral and Host Proteins in Enveloped Virus Infection. *Front. Immunol.* 12, 638573. doi:10.3389/fimmu.2021.638573

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., et al. (2020). Genomic Characterisation and Epidemiology of 2019 Novel Coronavirus: Implications for Virus Origins and Receptor Binding. *The Lancet* 395 (10224), 565–574. doi:10.1016/s0140-6736(20)30251-8

Mendonça, L., Howe, A., Gilchrist, J. B., Sheng, Y., Sun, D., Knight, M. L., et al. (2021). Correlative Multi-Scale Cryo-Imaging Unveils SARS-CoV-2 Assembly and Egress. *Nat. Commun.* 12 (1), 4629. doi:10.1038/s41467-021-24887-y

Moremen, K. W., Tiemeyer, M., and Nairn, A. V. (2012). Vertebrate Protein Glycosylation: Diversity, Synthesis and Function. *Nat. Rev. Mol. Cel Biol* 13 (7), 448–462. doi:10.1038/nrm3383

Sztain, T., Surl-Hee, A., Anthony, T. B., Lorenzo, C., Jory, A. G., Evan, S., and Ryan, S. M. (2021). A glycan gate controls opening of the SARS-CoV-2 spike protein. *Nature Chemistry* 13 (1), 963–968.

Plescia, C. B., David, E. A., Patra, D., Sengupta, R., Amiar, S., Su, Y., et al. (2021). SARS-CoV-2 Viral Budding and Entry Can Be Modeled Using BSL-2 Level Virus-like Particles. *J. Biol. Chem.* 296, 100103. doi:10.1074/jbc.ra120.016148

Rödig, J. V., Rapp, E., Bohne, J., Kampe, M., Kaffka, H., Bock, A., et al. (2013). Impact of Cultivation Conditions onN-Glycosylation of Influenza Virus a Hemagglutinin Produced in MDCK Cell Culture. *Biotechnol. Bioeng.* 110 (6), 1691–1703. doi:10.1002/bit.24834

Shajahan, A., Supekar, N. T., Gleinich, A. S., and Azadi, P. (2020). Deducing the N- and O-Glycosylation Profile of the Spike Protein of Novel Coronavirus SARS-CoV-2. *Glycobiology* 30 (12), 981–988. doi:10.1093/glycob/cwaa042

Shang, J., Wan, Y., Luo, C., Ye, G., Geng, Q., Auerbach, A., et al. (2020). Cell Entry Mechanisms of SARS-CoV-2. *Proc. Natl. Acad. Sci. USA* 117 (21), 11727–11734. doi:10.1073/pnas.2003138117

Singh, D., and Yi, S. V. (2021). On the Origin and Evolution of SARS-CoV-2. *Exp. Mol. Med.* 53 (4), 537–547. doi:10.1038/s12276-021-00604-z

Starr, T., Sun, Y., Wilkins, N., and Storrie, B. (2010). Rab33b and Rab6 Are Functionally Overlapping Regulators of Golgi Homeostasis and Trafficking. *Traffic* 11 (5), 626–636. doi:10.1111/j.1600-0854.2010.01051.x

Stewart-Jones, G. B. E., Soto, C., Lemmin, T., Chuang, G.-Y., Druz, A., Kong, R., et al. (2016). Trimeric HIV-1-Env Structures Define Glycan Shields from Clades A, B, and G. *Cell* 165 (4), 813–826. doi:10.1016/j.cell.2016.04.010

Sztain, T., Ahn, S. H., Bogetti, A. T., Casalino, L., Goldsmith, J. A., McCool, R. S., et al. (2021). A Glycan Gate Controls Opening of the SARS-CoV-2 Spike Protein. *Nat. Chem.* 13, 963–968. doi:10.1038/s41557-021-00758-3

Vigerust, D. J., and Shepherd, V. L. (2007). Virus Glycosylation: Role in Virulence and Immune Interactions. *Trends Microbiol.* 15 (5), 211–218. doi:10.1016/j.tim.2007.03.003

Walls, A. C., Tortorici, M. A., Frenz, B., Snijder, J., Li, W., Rey, F. A., et al. (2016). Glycan Shield and Epitope Masking of a Coronavirus Spike Protein Observed by Cryo-Electron Microscopy. *Nat. Struct. Mol. Biol.* 23 (10), 899–905. doi:10.1038/nsmb.3293

Walls, A. C., Xiong, X., Park, Y.-J., Tortorici, M. A., Snijder, J., Quispe, J., et al. (2019). Unexpected Receptor Functional Mimicry Elucidates Activation of Coronavirus Fusion. *Cell* 176 (5), 1026–1039. doi:10.1016/j.cell.2018.12.028

Wang, L., Fan, X., Bonenfant, G., Cui, D., Hossain, J., Jiang, N., et al. (2021). Susceptibility to SARS-CoV-2 of Cell Lines and Substrates Commonly Used to Diagnose and Isolate Influenza and Other Viruses. *Emerg. Infect. Dis.* 27 (5), 1380–1392. doi:10.3201/eid2705.210023

Wang, Q., Wang, Y., Wang, S., Lin, C., Aliyu, L., Chen, Y., et al. (2021). A Linkage-specific Sialic Acid Labeling Strategy Reveals Different Site-specific Glycosylation Patterns in SARS-CoV-2 Spike Protein Produced in CHO and HEK Cell Substrates. *Front. Chem.* 9 (735558), 735558. doi:10.3389/fchem.2021.735558

Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S., and Crispin, M. (2020). Site-specific Glycan Analysis of the SARS-CoV-2 Spike, *Science*, 369, 330–333. doi:10.1126/science.abb9983

Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S., and Crispin, M. (2020). Site-specific Glycan Analysis of the SARS-CoV-2 Spike. *Science* 369 (6501), 330–333. doi:10.1126/science.abb9983

Watanabe, Y., Bowden, T. A., Wilson, I. A., and Crispin, M. (2019). Exploitation of Glycosylation in Enveloped Virus Pathobiology. *Biochim. Biophys. Acta (Bba) - Gen. Subjects* 1863 (10), 1480–1497. doi:10.1016/j.bbagen.2019.05.012

Watanabe, Y., Mendonça, L., Allen, E. R., Howe, A., Lee, M., Allen, J. D., et al. (2021). Native-like SARS-CoV-2 Spike Glycoprotein Expressed by ChAdOx1 nCoV-19/AZD1222 Vaccine. *ACS Cent. Sci.* 7 (4), 594–602. doi:10.1021/acscentsci.1c00080

Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., et al. (2020). Cryo-EM Structure of the 2019-nCoV Spike in the Prefusion Conformation. *Science* 367 (6483), 1260–1263. doi:10.1126/science.abb2507

Xia, X. (2021). Domains and Functions of Spike Protein in Sars-Cov-2 in the Context of Vaccine Design. *Viruses* 13 (1). doi:10.3390/v13010109

Yang, Q., Hughes, T. A., Kelkar, A., Yu, X., Cheng, K., Park, S., et al. (2020). Inhibition of SARS-CoV-2 Viral Entry upon Blocking N- and O-Glycan Elaboration. *Elife* 9. doi:10.7554/eLife.61552

Yurkovetskiy, L., Wang, X., Pascal, K. E., Tomkins-Tinch, C., Nyalile, T. P., Wang, Y., et al. (2020). Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *Cell* 183 (3), 739–751. doi:10.1016/j.cell.2020.09.032

Zhao, P., Praissman, J. L., Grant, O. C., Cai, Y., Xiao, T., Rosenbalm, K. E., et al. (2020). Virus-Receptor Interactions of Glycosylated SARS-CoV-2 Spike and Human ACE2 Receptor. *Cell Host & Microbe* 28 (4), 586–601. doi:10.1016/j.chom.2020.08.004

Zhou, D., Tian, X., Qi, R., Peng, C., and Zhang, W. (2021). Identification of 22 N-Glycosites on Spike Glycoprotein of SARS-CoV-2 and Accessible Surface Glycopeptide Motifs: Implications for Vaccination and Antibody Therapeutics. *Glycobiology* 31 (1), 69–80. doi:10.1093/glycob/cwaa052

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article or claim that may be made by its manufacturer is not guaranteed or endorsed by the publisher.

frontiers in Chemistry

ORIGINAL RESEARCH
published: 23 December 2021
doi: 10.3389/fchem.2021.796690

Electrochemical Bromination of Glycals

Zhao-Xiang Luo[1†], *Miao Liu*[1†], *Tian Li*[1], *De-Cai Xiong*[1,2*] *and Xin-Shan Ye*[1*]

[1]State Key Laboratory of Natural and Biomimetic Drugs, School of Pharmaceutical Sciences, Peking University, Beijing, China, [2]State Key Laboratory of Pharmaceutical Biotechnology, School of Life Sciences, Nanjing University, Nanjing, China

Herein, the convenient one-step electrochemical bromination of glycals using Bu_4NBr as the brominating source under metal-catalyst-free and oxidant-free reaction conditions was described. A series of 2-bromoglycals bearing different electron-withdrawing or electron-donating protective groups were successfully synthesized in moderate to excellent yields. The coupling of tri-*O*-benzyl-2-bromogalactal with phenylacetylene, potassium phenyltrifluoroborate, or a 6-OH acceptor was achieved to afford 2C-branched carbohydrates and disaccharides via Sonogashira coupling, Suzuki coupling, and Ferrier rearrangement reactions with high efficiency. The radical trapping and cyclic voltammetry experiments indicated that bromine radicals may be involved in the reaction process.

Keywords: electrochemistry, bromination, glycals, 2-bromoglycals, cross-coupling, ferrier rearrangement

OPEN ACCESS

Edited by:
Jian Yin,
Jiangnan University, China

Reviewed by:
Bruno Linclau,
University of Southampton, United Kingdom
Xiaodong Shi,
University of South Florida, United States

****Correspondence:***
De-Cai Xiong
decai@bjmu.edu.cn
Xin-Shan Ye
xinshan@bjmu.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Organic Chemistry, a section of the journal Frontiers in Chemistry

Received: *18 October 2021*
Accepted: *01 December 2021*
Published: *23 December 2021*

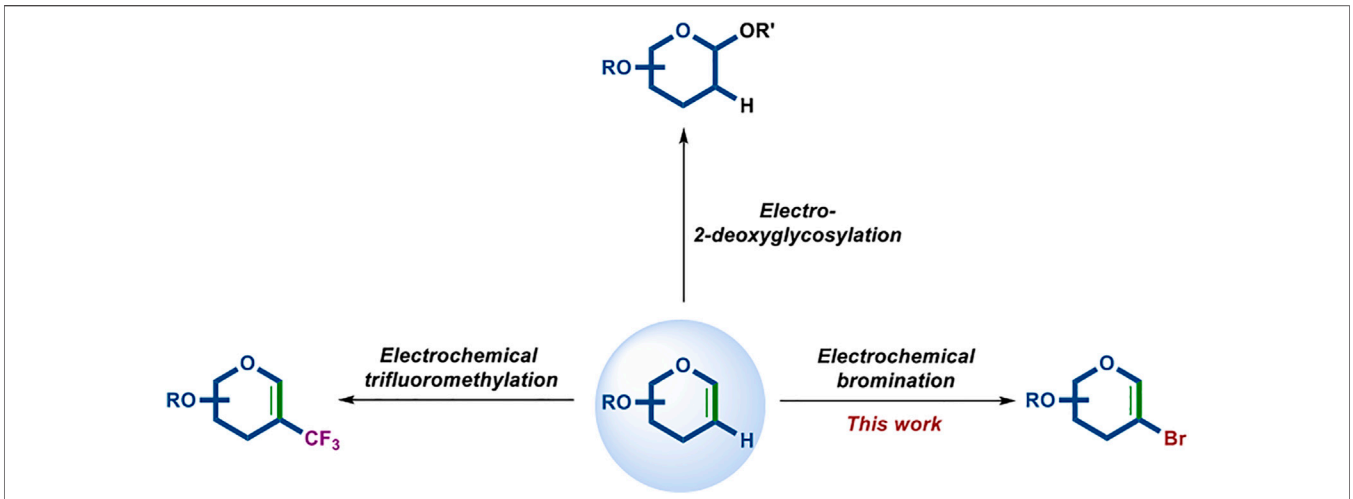
Citation:
Luo Z-X, Liu M, Li T, Xiong D-C and Ye X-S (2021) Electrochemical Bromination of Glycals. Front. Chem. 9:796690. doi: 10.3389/fchem.2021.796690

INTRODUCTION

Carbohydrates mainly exist in the form of glycoconjugates, polysaccharides, oligosaccharides, and monosaccharides and play a pivotal role in a broad range of important biological processes including cell proliferation, host–pathogen interactions, cell adhesion, hormone function, and the immune response (Kiessling and Kraft, 2013; Wang et al., 2020). Chemical synthesis can afford both naturally occurring important carbohydrates and biologically active carbohydrate mimetics in sufficient quantities, providing a powerful tool to understand the biological functions of carbohydrates (Muthana et al., 2009; Panza et al., 2018; Li and Ye, 2020).

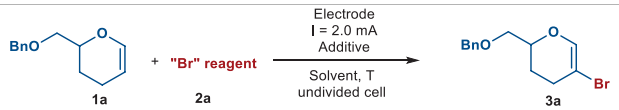
Organic electrosynthesis is of current interest as one of the most promising methods for the efficient, sustainable, and green synthesis of medicinally significant compounds (Francke and Little, 2014; Horn et al., 2016; Liang and Zeng, 2020; Meyer et al., 2020; Yuan et al., 2021). In recent years, the electrochemical synthesis of oligosaccharides has been successfully demonstrated through the activation of different types of glycosyl donors, such as thio-, seleno-, and telluro-glycosides (Nokami et al., 2015; Manmode et al., 2018; Zhang et al., 2020). In addition, our group has been involved in the electrochemical transformation of glycals to obtain significant synthetic carbohydrate compounds (Liu et al., 2020; Liu et al., 2021). Along with the use of $MnBr_2$ as the redox mediator, the electrochemical trifluoromethylation of glycals has been realized (Liu et al., 2021). In the exploration of this reaction, we found that 2-bromoglycals could also be isolated when the equivalent of Bu_4NBr was added. Inspired by this surprising result, we turned our focus to the electrochemical bromination of glycals (**Scheme 1**).

Over the past few decades, 2-bromoglycals have been widely employed as important synthons in combination with metal-catalyzed cross-coupling reactions to access 2C-branched carbohydrates and their analogs (Leibeling et al., 2010a; Leibeling et al., 2010b; Leibeling and Werz, 2012; Dharuman and Vankar, 2014; Martin et al., 2015). Due to the importance of



SCHEME 1 | Electrochemical transformation of glycals.

TABLE 1 | Optimization of reaction conditions[a].

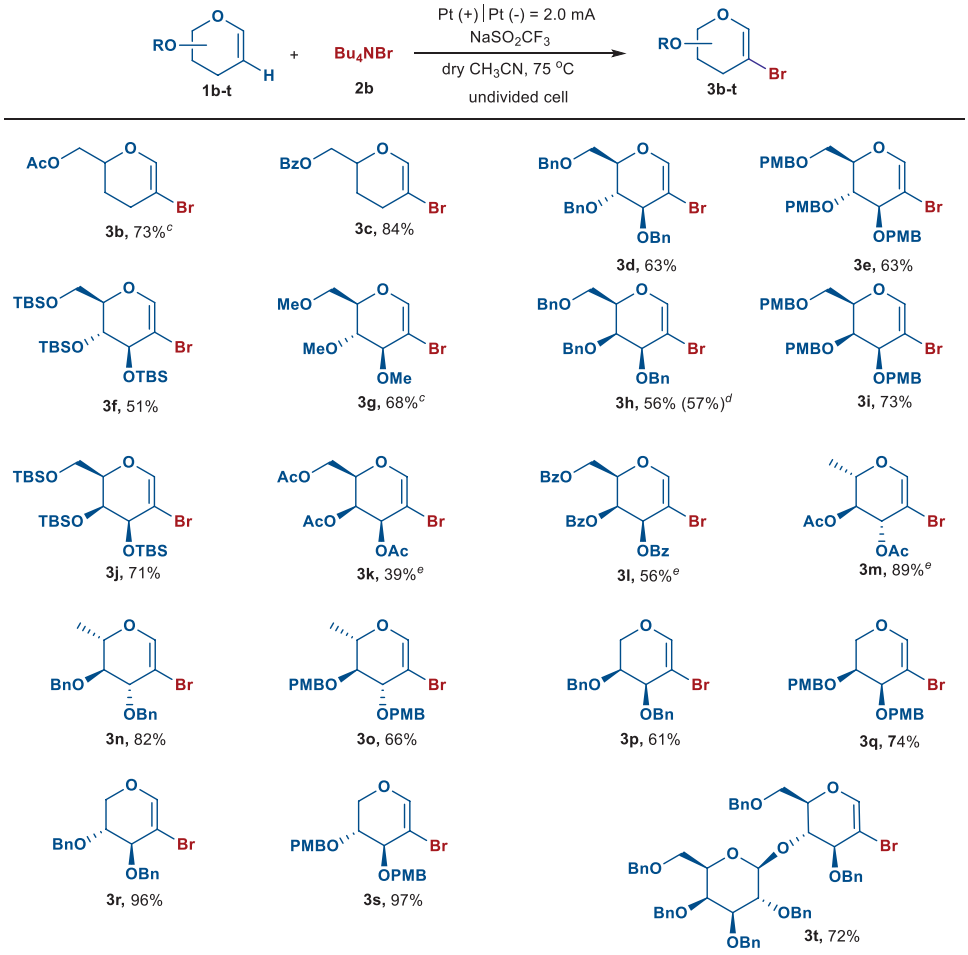


Entry	Electrode	"Br" reagent	Additive	Solvent	T (°C)	Yield (%)[b]
1	Pt (+)/Pt (–)	KBr (1.5 equiv)	—	CH_3CN	Rt	0 (0)
2	Pt (+)/Pt (–)	NaBr (1.5 equiv)	—	CH_3CN	Rt	0 (0)
3	Pt (+)/Pt (–)	Bu_4NBr (1.5 equiv)	—	CH_3CN	Rt	10 (0)
4	Pt (+)/Pt (–)	Bu_4NBr (1.5 equiv)	—	CH_3CN	50°C	18 (14)
5	Pt (+)/Pt (–)	Bu_4NBr (1.5 equiv)	—	CH_3CN	75°C	35 (10)
6	Pt (+)/Pt (–)	Bu_4NBr (2.0 equiv)	—	CH_3CN	75°C	40 (13)
7	Pt (+)/Pt (–)	Bu_4NBr (2.0 equiv)	K_2CO_3 (1.2 equiv)	CH_3CN	75°C	43 (4)
8	Pt (+)/Pt (–)	Bu_4NBr (2.0 equiv)	Na_2CO_3 (1.2 equiv)	CH_3CN	75°C	54 (6)
9	Pt (+)/Pt (–)	Bu_4NBr (2.0 equiv)	$NaSO_2CF_3$ (1.2 equiv)	CH_3CN	75°C	67 (0)
10	Pt (+)/Pt (–)	Bu_4NBr (2.0 equiv)	$NaSO_2CF_3$ (2.0 equiv)	CH_3CN	75°C	82 (0)
11	Pt (+)/Pt (–)	Bu_4NBr (2.0 equiv)	$NaSO_2CF_3$ (2.0 equiv)	CH_3CN/H_2O (3/1)	75°C	15 (0)
12	Pt (+)/Pt (–)	Bu_4NBr (2.0 equiv)	$NaSO_2CF_3$ (2.0 equiv)	1,2-Dimethoxyethane	75°C	46 (7)
13	Pt (+)/Pt (–)	Bu_4NBr (2.0 equiv)	$NaSO_2CF_3$ (2.0 equiv)	$ClCH_2CH_2Cl$	75°C	Trace (0)
14	C (+)/Pt (–)	Bu_4NBr (2.0 equiv)	$NaSO_2CF_3$ (2.0 equiv)	CH_3CN	75°C	71 (0)
15	Pt (+)/C (–)	Bu_4NBr (2.0 equiv)	$NaSO_2CF_3$ (2.0 equiv)	CH_3CN	75°C	32 (3)
16	C (+)/C (–)	Bu_4NBr (2.0 equiv)	$NaSO_2CF_3$ (2.0 equiv)	CH_3CN	75°C	33 (7)
17[c]	Pt (+)/Pt (–)	Bu_4NBr (2.0 equiv)	$NaSO_2CF_3$ (2.0 equiv)	CH_3CN	75°C	58 (0)
18[d]	Pt (+)/Pt (–)	Bu_4NBr (2.0 equiv)	$NaSO_2CF_3$ (2.0 equiv)	CH_3CN	75°C	54 (0)
19[e]	Pt (+)/Pt (–)	Bu_4NBr (2.0 equiv)	$NaSO_2CF_3$ (2.0 equiv)	CH_3CN	75°C	Trace (0)
20[f]	Pt (+)/Pt (–)	Bu_4NBr (2.0 equiv)	$NaSO_2CF_3$ (2.0 equiv)	CH_3CN	75°C	NR (97)

[a]*Reaction conditions:* ***1a*** *(0.05 mmol), "Br" reagent, Additive, Solvent (4.0 ml), Electrode, constant current = 2.0 mA, T, 4 h, in an undivided cell, under an argon atmosphere.*
[b]*Yield of the isolated product, the yield of recovered starting material was represented in the parentheses.*
[c]*I = 1.0 mA*
[d]*I = 3.0 mA*
[e]*Under an air atmosphere.*
[f]*No electricity.*

2-bromoglycals, the development of a novel, practical, and environmentally friendly method for the synthesis of 2-bromoglycals is still of high interest. The most common way to obtain 2-bromoglycals consists of two steps using Br_2, which is toxic and unstable, as the brominating source (Leibeling et al., 2010a). An alternative approach involves the one-step synthesis of 2-bromoglycals from glycals using *N*-bromosuccinimide and silver nitrate (Dharuman and Vankar, 2014). We herein report a one-step electrochemical bromination of glycals using commercially available,

TABLE 2 | Substrate scope of glycals[a,b,c,d,e].



[a]*Reaction conditions: glycals (0.05 mmol), $NaSO_2CF_3$ (0.10 mmol, 2.0 equiv), Bu_4NBr (0.10 mmol, 2.0 equiv), dry CH_3CN (4.0 ml) in an undivided cell with Pt as the anode and cathode, constant current = 2.0 mA, 75°C, under argon atmosphere, 4 h.*

[b]*Yield of the isolated product.*

[c]*glycals (0.10 mmol), $NaSO_2CF_3$ (0.20 mmol, 2.0 equiv), Bu_4NBr (0.30 mmol, 3.0 equiv), dry CH_3CN (5.0 ml) in an undivided cell with Pt as the anode and cathode, constant current = 2.0 mA, 75°C, under argon atmosphere, 6 h.*

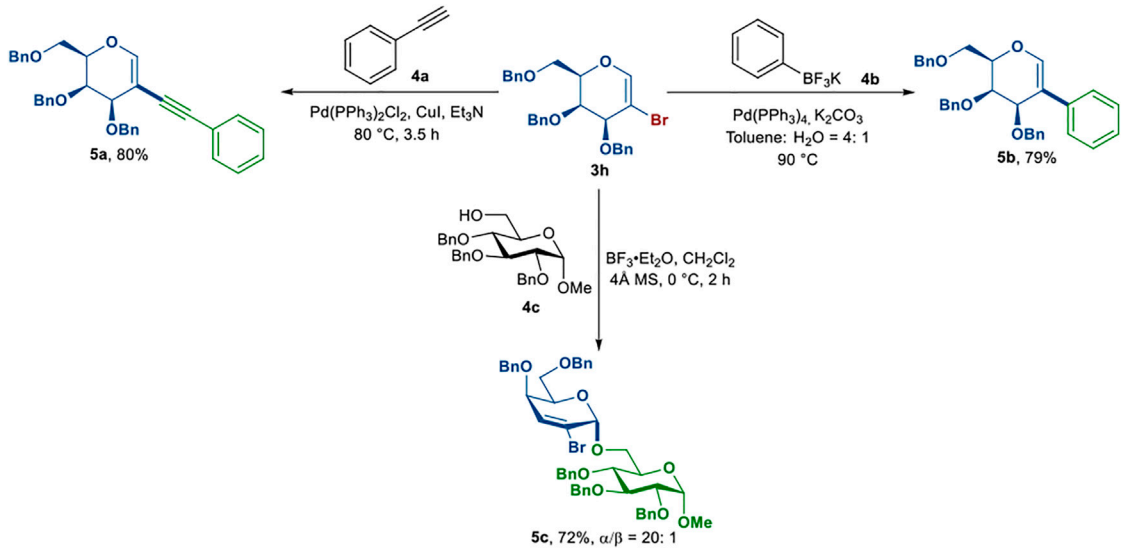
[d]*glycals (0.60 mmol), $NaSO_2CF_3$ (1.20 mmol, 2.0 equiv), Bu_4NBr (1.20 mmol, 2.0 equiv), dry CH_3CN (50.0 ml) in an undivided cell with Pt as the anode and cathode, constant current = 2.0 mA, 75°C, under argon atmosphere, 30 h.*

[e]*glycals (0.05 mmol), $NaSO_2CF_3$ (0.10 mmol, 2.0 equiv), Bu_4NBr (0.15 mmol, 3.0 equiv), dry CH_3CN (4.0 ml) in an undivided cell with Pt as the anode and cathode, constant current = 2.0 mA, 75°C, under argon atmosphere, 6 h.*

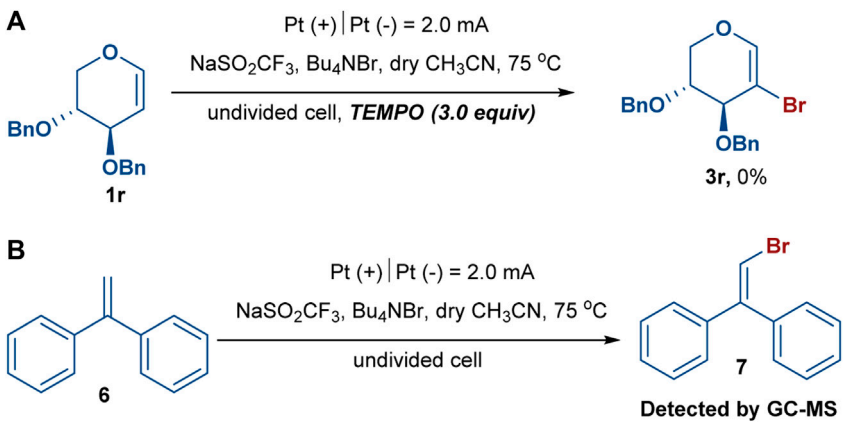
stable, and safe Bu_4NBr as the bromine source in an undivided cell under metal-catalyst-free and oxidant-free reaction conditions.

RESULTS AND DISCUSSION

Initially, we began our investigation with 6-*O*-benzyl-3,4-dideoxy-glycal **1a** as the model substrate for electrochemical bromination using Pt as the anode and cathode in an undivided cell. Unfortunately, no desired 2-bromo-3,4-dideoxy-glycal **3a** was detected using KBr (Ye and Shreeve, 2004; Zhao and Lu, 2018) or NaBr (Alberto et al., 2014) as the "Br" reagent (**Table 1**, entries 1 and 2). To our delight, the product **3a** could be obtained in 10% isolated yield when **1a** was treated with 1.5 equiv of Bu_4NBr (Yoshimitsu et al., 2009; Kamon et al., 2012) in dry CH_3CN at room temperature under a constant electric current of 2 mA (**Table 1**, entry 3). The yield was increased slightly when the reaction temperature was increased to 50°C (**Table 1**, entry 4). Further raising the temperature to 75°C was beneficial to this transformation, leading to the formation of **3a** in 35% yield (**Table 1**, entry 5), and the yield could be increased to 40% when the amount of Bu_4NBr was increased to 2.0 equiv (**Table 1**, entry 6). It was found that the addition of base, such as K_2CO_3 or Na_2CO_3, could further improve the reaction yield (**Table 1**, entries 7 and 8). Surprisingly, when the bromination reaction was conducted with $NaSO_2CF_3$ as the additive, the desired product **3a** was isolated in 67% yield (**Table 1**, entry 9). Altering the amount of $NaSO_2CF_3$ to 2.0 equiv led to an increased yield of 82%



SCHEME 2 | Reaction of 2-bromogalactal **(3h)** with different substrates **(4a–c)**.



SCHEME 3 | In the radical trapping experiments, experiment **(A)** was performed with glycal (0.05 mmol) and TEMPO (0.15 mmol, 3.0 equiv) under standard reaction condition: NaSO2CF3 (0.10 mmol, 2.0 equiv), Bu4NBr (0.10 mmol, 2.0 equiv), dry CH3CN (4.0 ml) in an undivided cell with Pt as the anode and cathode, constant current = 2.0 mA, 75 oC, under argon atmosphere, 4 h. Besides, experiment **(B)** was performed with 1,1-diphenylethylene (0.15 mmol) under standard reaction condition as well.

(**Table 1**, entry 10). Comparatively, when other solvents such as CH_3CN/H_2O (3:1), 1,2-dimethoxyethane, or $ClCH_2CH_2Cl$, were used instead of dry CH_3CN, lower yields were achieved (**Table 1**, entries 11–13). Extensive screening experiments revealed that either changing electrode materials or modifying the reaction current were not effective for improving the yield of **3a** (**Table 1**, entries 14–18). Moreover, the yield of **3a** was decreased drastically when the electrochemical bromination reaction was performed in an air atmosphere (**Table 1**, entry 19). Finally, the control experiment confirmed that the role of electricity was essential, as the reaction could not proceed in the absence of an electric current (**Table 1**, entry 20).

With the optimal reaction conditions in hand, we then evaluated the substrate scope of the electrochemical bromination of various types of glycals with Bu_4NBr (**Table 2**). First, 3,4-dideoxy-glycals with electron-withdrawing groups were examined. Substrates with an acetyl or benzoyl group provided the respective brominated products **3b** and **3c** in good yields. In addition, benzyl (Bn), *p*-methoxybenzyl (PMB), *tert*-butyldimethylsilyl (TBS), and methyl (Me) substituted glucals could also be converted into the corresponding products **3d–g**. Similarly, galactals bearing Bn, PMB, or TBS groups were found to be amenable to the electrochemical reaction, providing the desired products **3h–j** in 56–73% yields. And the scalability of this electrochemical

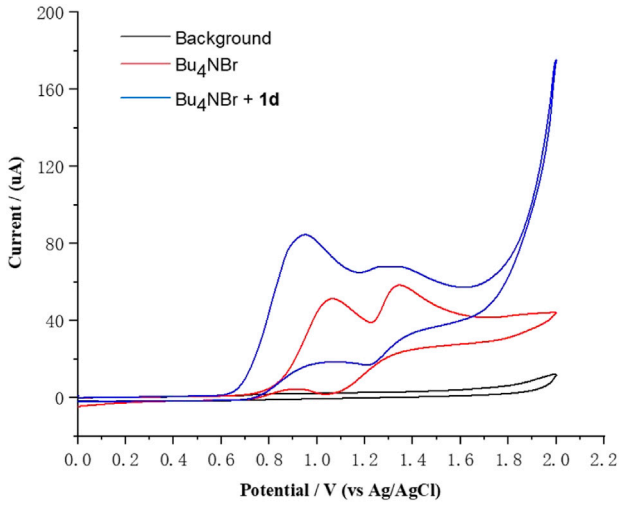
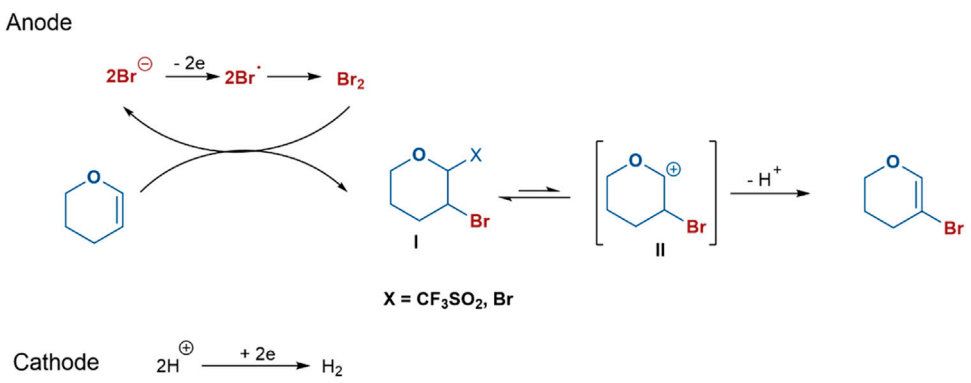


FIGURE 1 | Cyclic voltammetry measurements of Bu_4NBr and **1d**. Conditions: glassy carbon disk electrode (diameter is 3.0 mm, PTFE shroud) as the working electrode, platinum wire as the counter electrode, Ag/AgCl electrode (3.5 M KCl solution) as the reference electrode, Bu_4NOTf (0.10 M in MeCN), under an argon atmosphere, cyclic voltammogram at 0.05 V s^{-1} with Bu_4NBr (5 mM) or Bu_4NBr (5 mM) and **1d** (5 mM).

bromination was further showed by an efficient conversion of compound **1h** on a 250 mg scale in 57% isolated yield. Peracetylated and perbenzoylated galactals could also undergo the electrochemical transformation to afford 2-bromo-galactals **3k–l**, albeit in slightly low yields. Notably, bromination of peracetylated L-rhamnal proceeded smoothly to deliver compound **3m** in 89% yield. Under the same conditions, the L-rhamnal, L-arabinal, and D-xylal equipped with Bn or PMB groups were also able to furnish the corresponding brominated products **3n–s** in moderate to excellent yields. Furthermore, benzylated lactal **1t** underwent this reaction to afford the desired product **3t** in 72% yield.

To demonstrate the potential applicability of 2-bromoglycals, the reaction of 2-bromogalactal **3h** with different substrates was carried out (**Scheme 2**). First, we explored the utility of 2-bromogalactal **3h** in the synthesis of 2C-substituted carbohydrates, which exist in many natural products (Yin and Linker, 2012; Dubbu et al., 2018; Darbem et al., 2020). Compound **3h** reacted with phenylacetylene **4a** in the presence of $Pd(PPh_3)_2Cl_2$, CuI and Et_3N to afford the coupled product **5a** in 80% yield (Koester and Werz, 2012). The reaction of **3h** with potassium phenyltrifluoroborate **4b** also proceeded smoothly to provide the corresponding product **5b** in 79% yield (Molander and Fumagalli, 2006). Moreover, disaccharide **5c** was successfully synthesized in the promotion of $BF_3{\cdot}Et_2O$ in 72% yield with excellent α-selectivity via the Ferrier rearrangement reaction (Dharuman et al., 2013; Wang et al., 2019).

To gain insight into the mechanism of this electrochemical bromination, radical trapping experiments were performed. As expected, the reaction was completely shut down when 3.0 equiv of the radical scavenger 2,2,6,6-tetramethylpiperidine-1-oxyl (TEMPO) was added under the standard reaction conditions, indicating that radical chemistry was likely involved in the reaction (**Scheme 3A**) (Makai et al., 2020; Wang et al., 2019). Furthermore, another experiment using 1,1-diphenylethylene (**6**) was also conducted under the standard reaction conditions, and (2-bromoethene-1,1-diyl) dibenzene (**7**) was successfully detected in the GC–MS, confirming the participation of the bromine radical in the reaction system (**Scheme 3B**) (Chen et al., 2020; Kale et al., 2021). We observed that the reaction mixture gradually turned brown during the reaction process, indicating that Br_2 might be generated. The 2-bromoglycal **3h** was resubmitted to the standard reaction conditions for 20 h with the recovery of **3h** in a 78% yield. In addition, cyclic voltammetry experiments were carried out to investigate the redox behavior of the reaction. The cyclic voltammetry measurements of Bu_4NBr indicated two obvious oxidative peaks at 1.07 and 1.37 V (vs. Ag/AgCl) (**Figure 1**, red curve), which likely corresponded to Br_3^-/Br^- and Br_3^-/Br_2 redox couples, respectively (Damljanovic et al., 2011; Bennett et al., 2016; Kang et al., 2016). The first oxidative peak was associated with the oxidation of Br^- to the bromine radical. The bromine



SCHEME 4 | Plausible reaction mechanism for the electrochemical bromination of glycals.

radical then integrated into Br_2, which could combine with Br^- to form Br_3^-. The second oxidative peak was attributed to the oxidation of Br_3^- to Br_2. An increase in the oxidative peak current was observed when Bu_4NBr and **1d** were combined, which was attributed to a catalytic current, resulting from the chemical reaction of the electrochemically-generated Br_2 and glucal **1d** (**Figure 1**, blue curve).

Mechanism

Based on the above results and previous reports (Yuan et al., 2019; Gou et al., 2021; Wu et al., 2021), a plausible reaction mechanism for the electrochemical bromination of glycals was depicted in **Scheme 4**. A bromine anion was oxidized to the bromine radical on the anode and subsequently molecular Br_2. This was then attacked by glycal to yield the intermediate **I**. Finally, the brominated product was obtained by the deprotonation of **II**, which would be stabilized by acetonitrile. $NaSO_2CF_3$ might be used as the electrolyte to increase the conductivity of the reaction solution and a proton scavenger to generate strong acid for cathode reduction; besides, it may be an anion to stabilize the glycosyl cation. At the same time, H^+ was reduced to produce hydrogen on the cathode.

CONCLUSION

In summary, we performed the one-step electrochemical bromination of various glycals with electron-withdrawing and electron-donating protective groups using commercially available, nontoxic Bu_4NBr as the brominating source under metal-catalyst-free and oxidant-free reaction conditions. The synthetic applicability of 2-bromoglycals has been demonstrated by providing the corresponding 2C-substituted carbohydrates and disaccharides via palladium-catalyzed cross-coupling reactions and the Ferrier rearrangement reaction. The readily available substrates and ease of handling make this methodology a practical tool to access diversified brominating synthons for the preparation of biologically relevant carbohydrates.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

Z-XL and ML contributed equally to this work. Z-XL, ML, and TL performed the experiments. Z-XL and ML carried out the mechanistic studies. Z-XL, ML, D-CX, and X-SY analyzed and discussed the research results. D-CX and X-SY designed the experiments. Z-XL, ML, and D-CX wrote the manuscript.

FUNDING

This research was funded by grants from the National Natural Science Foundation of China (Grant Nos. 21772006, 91853122), the Beijing Outstanding Young Scientist Program (BJJWZYJH01201910001001), the open fund of State Key Laboratory of Pharmaceutical Biotechnology, Nanjing University, China (Grant no. KF-GN-202106), and the Fundamental Research Funds for the Central Universities.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.796690/full#supplementary-material

REFERENCES

Alberto, E. E., Muller, L. M., and Detty, M. R. (2014). Rate Accelerations of Bromination Reactions with NaBr and H_2O_2 via the Addition of Catalytic Quantities of Diaryl Ditellurides. *Organometallics* 33, 5571–5581. doi:10.1021/om500883f

Bennett, B., Chang, J., and Bard, A. J. (2016). Mechanism of the Br $-/Br_2$ Redox Reaction on Platinum and Glassy Carbon Electrodes in Nitrobenzene by Cyclic Voltammetry. *Electrochimica Acta* 219, 1–9. doi:10.1016/j.electacta.2016.09.129

Chen, D., Li, J., Cui, P., Shan, Y., Zhao, Y., and Qiu, G. (2020). Tandem Oxidative Radical Halogenated Addition of Alkynyl Imines: Regioselective Synthesis of 3-Haloquinolines. *Eur. J. Org. Chem.* 2020, 169–175. doi:10.1002/ejoc.201901395

Damljanović, I., Stevanović, D., Vukićević, M., and Vukićević, R. D. (2011). Electrochemical Bromochlorination of Peracetylated Glycals. *Carbohydr. Res.* 346, 2683–2687. doi:10.1016/j.carres.2011.09.016

Darbem, M. P., Esteves, H. A., Oliveira, I. M., and Stefani, H. A. (2020). α,β-Unsaturated 2-Ketoglycosides via Pd-Catalyzed Carbonylative Heck Reaction of 2-Iodoglycals. *Eur. J. Org. Chem.* 2020, 5220–5226. doi:10.1002/ejoc.202000846

Dharuman, S., Gupta, P., Kancharla, P. K., and Vankar, Y. D. (2013). Synthesis of 2-Nitroglycals from Glycals Using the Tetrabutylammonium Nitrate-Trifluoroacetic Anhydride-Triethylamine Reagent System and Base-Catalyzed Ferrier Rearrangement of Acetylated 2-Nitroglycals. *J. Org. Chem.* 78, 8442–8450. doi:10.1021/jo401165y

Dharuman, S., and Vankar, Y. D. (2014). *N*-Halosuccinimide/$AgNO_3$-Efficient Reagent Systems for One-step Synthesis of 2-Haloglycals from Glycals: Application in the Synthesis of 2*C*-Branched Sugars via Heck Coupling Reactions. *Org. Lett.* 16, 1172–1175. doi:10.1021/ol500039s

Dubbu, S., Verma, A. K., Parasuraman, K., and Vankar, Y. D. (2018). Stereoselective Synthesis of 1,2-Annulated-*C*-Aryl Glycosides from Carbohydrate-Derived Terminally Unsubstituted Dienes and Arynes: Application towards Synthesis of Sugar-Fused- or Branched- Naphthalenes, and *C*-Aryl Glycosides. *Carbohydr. Res.* 465, 29–34. doi:10.1016/j.carres.2018.06.001

Francke, R., and Little, R. D. (2014). Redox Catalysis in Organic Electrosynthesis: Basic Principles and Recent Developments. *Chem. Soc. Rev.* 43, 2492–2521. doi:10.1039/c3cs60464k

Gou, X.-Y., Li, Y., Luan, Y.-Y., Shi, W.-Y., Wang, C.-T., An, Y., et al. (2021). Ruthenium-Catalyzed Radical Cyclization/meta-Selective C-H Alkylation of Arenes via σ-Activation Strategy. *ACS Catal.* 11, 4263–4270. doi:10.1021/acscatal.1c00359

Horn, E. J., Rosen, B. R., Chen, Y., Tang, J., Chen, K., Eastgate, M. D., et al. (2016). Scalable and Sustainable Electrochemical Allylic C-H Oxidation. *Nature* 533, 77–81. doi:10.1038/nature17431

Kamon, T., Shigeoka, D., Tanaka, T., and Yoshimitsu, T. (2012). Intramolecular Iron(II)-catalyzed Aminobromination of Allyl *N*-Tosyloxycarbamates. *Org. Biomol. Chem.* 10, 2363–2365. doi:10.1039/c2ob07190h

Kang, L.-S., Luo, M.-H., Lam, C. M., Hu, L.-M., Little, R. D., and Zeng, C.-C. (2016). Electrochemical C-H Functionalization and Subsequent C-S and C-N Bond Formation: Paired Electrosynthesis of 3-Amino-2-Thiocyanato-α,β-Unsaturated Carbonyl Derivatives Mediated by Bromide Ions. *Green. Chem.* 18, 3767–3774. doi:10.1039/c6gc00666c

Kiessling, L. L., and Kraft, M. B. (2013). A Path to Complex Carbohydrates. *Science* 341, 357–358. doi:10.1126/science.1241788

Koester, D. C., and Werz, D. B. (2012). Sonogashira-Hagihara Reactions of Halogenated Glycals. *Beilstein J. Org. Chem.* 8, 675–682. doi:10.3762/bjoc.8.75

Leibeling, M., Koester, D. C., Pawliczek, M., Kratzert, D., Dittrich, B., and Werz, D. B. (2010a). Hybrids of Sugars and Aromatics: A Pd-Catalyzed Modular Approach to Chromans and Isochromans. *Bioorg. Med. Chem.* 18, 3656–3667. doi:10.1016/j.bmc.2010.03.004

Leibeling, M., Koester, D. C., Pawliczek, M., Schild, S. C., and Werz, D. B. (2010b). Domino Access to Highly Substituted Chromans and Isochromans from Carbohydrates. *Nat. Chem. Biol.* 6, 199–201. doi:10.1038/nchembio.302

Leibeling, M., and Werz, D. B. (2012). Winding up Alkynes: A Pd-Catalyzed Tandem-Domino Reaction to Chiral Biphenyls. *Chem. Eur. J.* 18, 6138–6141. doi:10.1002/chem.201200175

Li, B.-H., and Ye, X.-S. (2020). Recent Advances in Glycan Synthesis. *Curr. Opin. Chem. Biol.* 58, 20–27. doi:10.1016/j.cbpa.2020.04.009

Liang, S., and Zeng, C.-C. (2020). Organic Electrochemistry: Anodic Construction of Heterocyclic Structures. *Curr. Opin. Electrochemistry* 24, 31–43. doi:10.1016/j.coelec.2020.06.005

Liu, M., Liu, K. M., Xiong, D. C., Zhang, H., Li, T., Li, B., et al. (2020). Stereoselective Electro-2-deoxyglycosylation from Glycals. *Angew. Chem. Int. Ed.* 59, 15204–15208. doi:10.1002/anie.202006115

Liu, M., Luo, Z.-X., Li, T., Xiong, D.-C., and Ye, X.-S. (2021). Electrochemical Trifluoromethylation of Glycals. *J. Org. Chem.* 86, 16187–16194. doi:10.1021/acs.joc.1c01318

Makai, S., Falk, E., and Morandi, B. (2020). Direct Synthesis of Unprotected 2-Azidoamines from Alkenes via an Iron-Catalyzed Difunctionalization Reaction. *J. Am. Chem. Soc.* 142, 21548–21555. doi:10.1021/jacs.0c11025

Manmode, S., Matsumoto, K., Nokami, T., and Itoh, T. (2018). Electrochemical Methods as Enabling Tools for Glycosylation. *Asian J. Org. Chem.* 7, 1719–1729. doi:10.1002/ajoc.201800302

Martin, A., Arda, A., Désiré, J., Martin-Mingot, A., Probst, N., Sinaÿ, P., et al. (2015). Catching Elusive Glycosyl Cations in a Condensed Phase with HF/SbF_5 Superacid. *Nat. Chem* 8, 186–191. doi:10.1038/nchem.2399

Meyer, T. H., Choi, I., Tian, C., and Ackermann, L. (2020). Powering the Future: How Can Electrochemistry Make a Difference in Organic Synthesis? *Chem* 6, 2484–2496. doi:10.1016/j.chempr.2020.08.025

Molander, G. A., and Fumagalli, T. (2006). Palladium(0)-Catalyzed Suzuki–Miyaura Cross-Coupling Reactions of Potassium Aryl- and Heteroaryltrifluoroborates with Alkenyl Bromides. *J. Org. Chem.* 71, 5743–5747. doi:10.1021/jo0608366

Muthana, S., Cao, H., and Chen, X. (2009). Recent Progress in Chemical and Chemoenzymatic Synthesis of Carbohydrates. *Curr. Opin. Chem. Biol.* 13, 573–581. doi:10.1016/j.cbpa.2009.09.013

Nokami, T., Isoda, Y., Sasaki, N., Takaiso, A., Hayase, S., Itoh, T., et al. (2015). Automated Electrochemical Assembly of the Protected Potential TMG-Chitotriomycin Precursor Based on Rational Optimization of the Carbohydrate Building Block. *Org. Lett.* 17, 1525–1528. doi:10.1021/acs.orglett.5b00406

Panza, M., Pistorio, S. G., Stine, K. J., and Demchenko, A. V. (2018). Automated Chemical Oligosaccharide Synthesis: Novel Approach to Traditional Challenges. *Chem. Rev.* 118, 8105–8150. doi:10.1021/acs.chemrev.8b00051

Prabhakar Kale, A., Nikolaienko, P., Smirnova, K., and Rueping, M. (2021). Intramolecular Electrochemical Oxybromination of Olefins for the Synthesis of Isoxazolines in Batch and Continuous Flow. *Eur. J. Org. Chem.* 2021, 3496–3500. doi:10.1002/ejoc.202100640

Wang, J., Deng, C., Zhang, Q., and Chai, Y. (2019). Tuning the Chemoselectivity of Silyl Protected Rhamnals by Temperature and Brønsted Acidity: Kinetically Controlled 1,2-Addition vs Thermodynamically Controlled Ferrier Rearrangement. *Org. Lett.* 21, 1103–1107. doi:10.1021/acs.orglett.9b00009

Wang, Y., Carder, H. M., and Wendlandt, A. E. (2020). Synthesis of Rare Sugar Isomers through Site-Selective Epimerization. *Nature* 578, 403–408. doi:10.1038/s41586-020-1937-1

Wu, Y., Xu, S., Wang, H., Shao, D., Qi, Q., Lu, Y., et al. (2021). Directing Group Enables Electrochemical Selectively Meta-Bromination of Pyridines under Mild Conditions. *J. Org. Chem.* 86, 16144–16150. doi:10.1021/acs.joc.1c00923

Ye, C., and Shreeve, J. n. M. (2004). Structure-Dependent Oxidative Bromination of Unsaturated C−C Bonds Mediated by Selectfluor. *J. Org. Chem.* 69, 8561–8563. doi:10.1021/jo048383x

Yin, J., and Linker, T. (2012). Recent Advances in the Stereoselective Synthesis of Carbohydrate 2-*C*-Analogs. *Org. Biomol. Chem.* 10, 2351–2362. doi:10.1039/c2ob06529k

Yoshimitsu, T., Ino, T., Futamura, N., Kamon, T., and Tanaka, T. (2009). Total Synthesis of the β-Catenin Inhibitor, (−)-Agelastatin A: A Second-Generation Approach Based on Radical Aminobromination. *Org. Lett.* 11, 3402–3405. doi:10.1021/ol9012684

Yuan, Y., Yang, J., and Lei, A. (2021). Recent Advances in Electrochemical Oxidative Cross-Coupling with Hydrogen Evolution Involving Radicals. *Chem. Soc. Rev.* 50, 10058–10086. doi:10.1039/d1cs00150g

Yuan, Y., Yao, A., Zheng, Y., Gao, M., Zhou, Z., Qiao, J., et al. (2019). Electrochemical Oxidative Clean Halogenation Using HX/NaX with Hydrogen Evolution. *Iscience* 12, 293–303. doi:10.1016/j.isci.2019.01.017

Zhang, H. Y., Liu, M., Wu, X., Liu, M., Xiong, D. C., and Ye, X. S. (2020). Photo-/Electro-Driven Carbohydrate-Based Reactions. *Prog. Chem.* 32, 1804–1823. doi:10.7536/pc200608

Zhao, M., and Lu, W. (2018). Catalytic Bromination of Alkyl sp^3C-H Bonds with KBr/Air under Visible Light. *Org. Lett.* 20, 5264–5267. doi:10.1021/acs.orglett.8b02208

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Chemistry

ORIGINAL RESEARCH
published: 01 February 2022
doi: 10.3389/fchem.2021.816544

In Planta Production of the Receptor-Binding Domain From SARS-CoV-2 With Human Blood Group A Glycan Structures

*Julia König-Beihammer[1], Ulrike Vavra[1], Yun-Ji Shin[1], Christiane Veit[1], Clemens Grünwald-Gruber[2], Yasmin Gillitschka[3], Jasmin Huber[3], Manuela Hofner[3], Klemens Vierlinger[3], Dieter Mitteregger[4], Andreas Weinhäusel[3] and Richard Strasser[1]**

[1]Department of Applied Genetics and Cell Biology, Institute of Plant Biotechnology and Cell Biology, University of Natural Resources and Life Sciences, Vienna, Austria, [2]Core Facility Mass Spectrometry, University of Natural Resources and Life Sciences Vienna, Muthgasse, Austria, [3]Core Facility Mass Spectrometry, University of Natural Resources and Life Sciences, Vienna, Austria, [4]Laboratory Dr. Kosak, Dr. Reckendorfer, Vienna, Austria

OPEN ACCESS

Edited by:
Jian Yin,
Jiangnan University, China

Reviewed by:
Qiong Wang,
Johns Hopkins University, United States
Qinzhe Wang,
University of Utah, United States

***Correspondence:**
Richard Strasser
richard.strasser@boku.ac.at

Specialty section:
This article was submitted to Chemical Biology, a section of the journal Frontiers in Chemistry

Received: 16 November 2021
Accepted: 28 December 2021
Published: 01 February 2022

Citation:
König-Beihammer J, Vavra U, Shin Y-J, Veit C, Grünwald-Gruber C, Gillitschka Y, Huber J, Hofner M, Vierlinger K, Mitteregger D, Weinhäusel A and Strasser R (2022) In Planta Production of the Receptor-Binding Domain From SARS-CoV-2 With Human Blood Group A Glycan Structures. Front. Chem. 9:816544. doi: 10.3389/fchem.2021.816544

Glycosylation of viral envelope proteins is important for infectivity and immune evasion. The SARS-CoV-2 spike protein is heavily glycosylated and host-derived glycan modifications contribute to the formation of specific immunogenic epitopes, enhance the virus-cell interaction or affect virus transmission. On recombinant viral antigens used as subunit vaccines or for serological assays, distinct glycan structures may enhance the immunogenicity and are recognized by naturally occurring antibodies in human sera. Here, we performed an *in vivo* glycoengineering approach to produce recombinant variants of the SARS-CoV-2 receptor-binding domain (RBD) with blood group antigens in *Nicotiana benthamiana* plants. SARS-CoV-2 RBD and human glycosyltransferases for the blood group ABH antigen formation were transiently co-expressed in *N. benthamiana* leaves. Recombinant RBD was purified and the formation of complex N-glycans carrying blood group A antigens was shown by immunoblotting and MS analysis. Binding to the cellular ACE2 receptor and the conformation-dependent CR3022 antibody showed that the RBD glycosylation variants carrying blood group antigens were functional. Analysis of sera from RBD-positive and RBD-negative individuals revealed further that non-infected RBD-negative blood group O individuals have antibodies that strongly bind to RBD modified with blood group A antigen structures. The binding of IgGs derived from sera of non-infected RBD-negative blood group O individuals to blood group A antigens on SARS-CoV-2 RBD suggests that these antibodies could provide some degree of protection from virus infection.

Keywords: blood group antigen, carbohydrate, glycoengineering, glycosylation, posttranslational modification, virus

INTRODUCTION

Processing of glycans on viral proteins depends on the protein conformation and the glycosylation machinery of the expressing cell. Viral spike proteins are heavily glycosylated, and glycosylation is crucial for virus infection and the host immune response (Watanabe et al., 2020a; Li Q et al., 2020). During viral infection, enveloped viruses engage with host cell receptors to initiate uptake by the cells. SARS-CoV-2 cell entry is dependent on the heavily glycosylated spike protein. The spike protein is a type I transmembrane protein with a large N-terminal ectodomain that protrudes from the viral surface (Walls et al., 2020). The spike protein binds to the cell surface receptor angiotensin-converting enzyme 2 (ACE2) which mediates membrane fusion and virus entry (Hoffmann et al., 2020; Walls et al., 2020). The SARS-CoV-2 monomeric spike protein possesses 22 N-glycosylation sites (Asn-X-Ser/Thr, with X any amino acid except proline) and displays a mixture of oligomannosidic, hybrid-type and complex N-glycans on recombinantly produced protein (Watanabe et al., 2020a; Shajahan et al., 2021). Furthermore, recent studies have shown site-specific differences in N-glycan composition between recombinant spike and infectious virions obtained from different human cells (Turoňová et al., 2020; Brun et al., 2021) highlighting the importance of glycoprotein presentation/site accessibility, quaternary protein architecture, and the host glycosylation machinery.

The ABH(O) blood group antigens are specific carbohydrate structures attached to glycoproteins and glycolipids present on the surface of human erythrocytes, epithelial and endothelial cells of different tissues (Yamamoto, 2004; Cooling, 2015). Responsible for blood group A biosynthesis is an N-acetylgalactosaminyltransferase (ABO A enzyme) that transfers an N-acetylgalactosamine (GalNAc) in α1,3-linkage to a precursor structure called the H antigen (Fucα1,2-Galβ-R) resulting in the trisaccharide GalNAcα1-3-(Fucα1,2)-Galβ-R (**Figure 1**). By contrast, the ABO B enzyme transfers a galactose residue in α1,3-linkage to the H antigen resulting in the trisaccharide Galα1-3-(Fucα1,2)-Galβ-R. The ABO A and ABO B glycosyltransferases are encoded by distinct alleles of the ABO gene locus which are codominant to each other. The O allele is a null allele at the ABO locus and lacks the corresponding A or B glycosyltransferase activities, in which case the H antigen remains unmodified. Individuals who have the genotype AA or AO synthesize exclusively the A antigen, BB and BO individuals have blood group B and individuals with genotype OO have the blood group O. Blood group O individuals have high titers of circulating antibodies against A and B antigens (Stussi et al., 2005). Blood group A individuals have anti-B antibodies and blood group B individuals have anti-A antibodies. Blood group AB individuals express both antigens and lack anti-A or anti-B antibodies (Cooling, 2015). While the biological role of these carbohydrates is still poorly understood, the blood group antigens are of clinical relevance and anti-blood group antibodies are critical for blood transfusions and transplantation medicine.

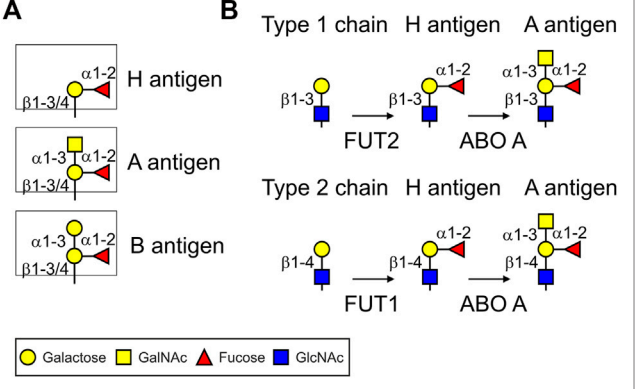


FIGURE 1 | Cartoon illustration of ABH(O) blood group antigens and biosynthetic pathways. **(A)** ABH blood group carbohydrate structures. **(B)** Biosynthetic steps to produce blood group A type 1 and 2 antigens common to N- and O-glycans. FUT1/FUT2: α1,2-fucosyltransferases; ABO A: α1,3-GalNAc-transferase (A enzyme). For the biosynthesis of blood group B structures the ABO A enzyme is replaced by the α1,3-galactosyltransferase (ABO B enzyme).

Blood group frequencies vary among human populations and the specific exposure to pathogens may explain some of the observed variations in infectivity (Goel et al., 2021). Earlier epidemiological observations indicated that ABO blood groups may contribute to susceptibility to SARS-CoV-1 infection (Cheng et al., 2005). In a more recent genome-wide association study of nearly 2000 SARS-CoV-2 infected individuals, a gene cluster carrying the ABO locus was enriched in patients with COVID-19 (Ellinghaus et al., 2020). Furthermore, a study with 2,173 patients from different Chinese hospitals reported that ABO blood groups display different association risks for the infection with SARS-CoV-2 resulting in COVID-19 (Zhao J. et al., 2020). Blood group A was associated with an increased risk whereas blood group O was associated with a decreased risk. A meta-analysis reported that the proportion of blood group A in patients infected with SARS-CoV-2 was significantly increased compared to a control group (Li J et al., 2020). From these and further epidemiological studies using different populations the association between distinct ABO blood groups and COVID-19-linked hospitalization is well established (Leaf et al., 2020; Wu et al., 2020; Miotto et al., 2021). Overall, these studies suggest a role for blood group A glycans and anti-A antibodies in SARS-CoV-2 infection, which could potentially be harnessed for applications aiming to prevent the transmission from individuals to individuals. Still, the underlying mechanisms of increased risks for blood group A individuals are unclear and different hypotheses have been discussed and tested (Arend, 2021; Deleers et al., 2021; Goel et al., 2021; Wu et al., 2021). Potential mechanisms include a protective role of natural antibodies against blood group antigens (Arend, 2021; Deleers et al., 2021) or the presence of a lectin domain in RBD that mediates binding to blood group A structures on the cell surface of respiratory epithelial cells that could promote infection of the cells (Wu et al. 2021). There is an association of variations in angiotensin-converting enzyme-1 (ACE1) activity and ABO blood groups (Goel et al., 2021) and

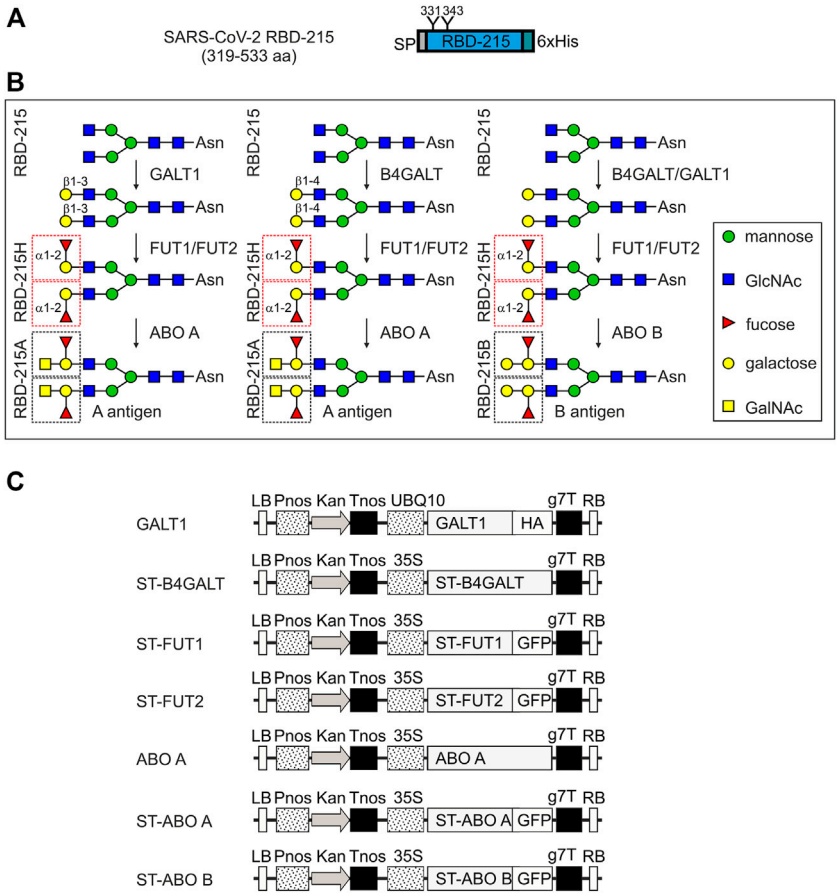


FIGURE 2 | Schematic overview of the glycoengineering strategy and the used expression constructs. **(A)** Illustration of the plant-expressed RBD (amino acid region 319–533 of the SARS-CoV-2 spike protein). **(B)** Glycoengineering steps to produce different ABH blood group antigens in plants. GALT1: *A. thaliana* Lewis-type β1,3-galactosyltransferase 1; B4GALT: human β1,4-galactosyltransferase; FUT1/FUT2: human α1,2-fucosyltransferases. ABO A: human α1,3-GalNAc-transferase; ABO B: human α1,3-galactosyltransferase. **(C)** Schematic presentation of the expression cassettes for the different glycosyltransferases. LB: left border; Pnos: nopaline synthase gene promoter; Kan: neomycin phosphotransferase II gene; Tnos: nopaline synthase gene terminator; UBQ10: *A. thaliana* ubiquitin-10 promoter; ST: N-terminal *trans*-Golgi targeting region from rat α2,6-sialyltransferase (amino acids 1–52); HA: hemagglutinin tag; GFP: green fluorescent protein; g7T: agrobacterium gene 7 terminator; RB: right border.

the half-life of coagulation factors like factor VIII and von Willebrand factor is altered by N-glycans carrying certain ABO blood group structures (Gallinaro et al., 2008). As a consequence, increased levels of von Willebrand factor in blood group A individuals could contribute to thrombosis and adverse outcomes upon SARS-CoV-2 infection. Based on these observations it is possible that more than one mechanism could protect blood group O individuals from infection and severe disease progression.

Here, we performed glycoengineering in *N. benthamiana* to produce betacoronavirus antigens furnished with blood group carbohydrate structures. We transiently expressed the receptor-binding domain (RBD) of the SARS-CoV-2 spike protein (RBD-215) (Shin et al., 2021) and the RBD from the SARS-CoV-1 spike in *N. benthamiana* and characterized the binding to antibodies and the cellular ACE2 receptor. The binding of IgGs derived from sera of blood group O and B donors to blood group A antigens on SARS-CoV-2 RBD suggests that these antibodies could provide some degree of protection from transmission of virus carrying blood group A carbohydrates.

RESULTS

Recombinant RBD With Blood Group A N-Glycans can Be Produced in *N. benthamiana*

Recombinant RBD-215 (amino acids 319–533 of the SARS-CoV-2 spike protein, **Figure 2A**) expressed in glycoengineered ΔXT/FT plants carries mainly $GlcNAc_2Man_3GlcNAc_2$ (GnGn) N-glycans on both N-glycosylation sites (Shin et al., 2021). To see if these complex N-glycans can be modified with blood group carbohydrates, we transiently co-expressed glycosyltransferases for the formation of blood group antigens (**Figure 2B**). To achieve modification of recombinant RBD-215 with blood

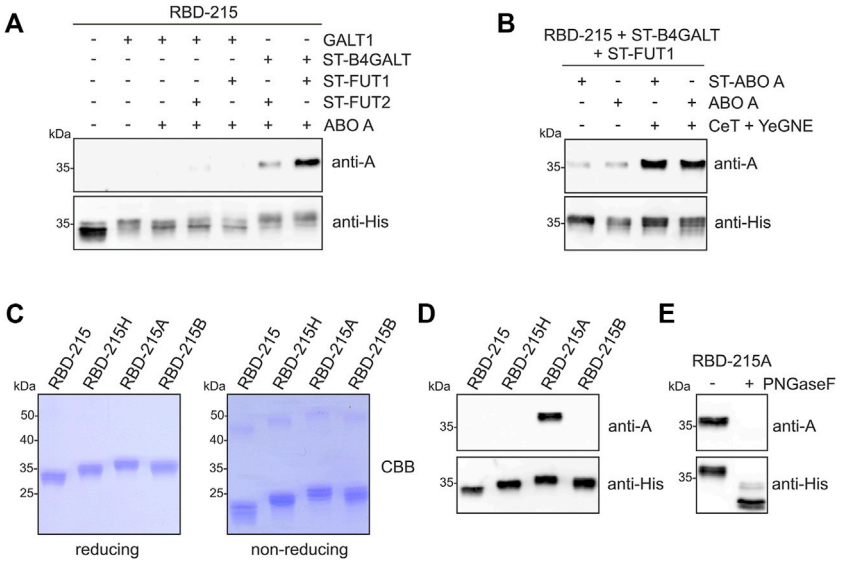


FIGURE 3 | Plant-based production of RBD-215 with type 2 blood group A antigens. **(A)** RBD-215 was transiently expressed in the glycoengineered *N. benthamiana* line ΔXT/FT together with the indicated glycosyltransferases. 3 days-after infiltration, RBD-215 was purified from crude protein extracts using magnetic beads and subjected to SDS-PAGE, and immunoblotting with anti-blood group A (3-3A) or anti-His-tag antibodies. **(B)** Co-expression of *Y. enterocolitica* UDP-GlcNAc 4-epimerase (YeGNE) and *C. elegans* UDP-GlcNAc/UDP-GalNAc transporter (CeT) improves the formation of blood group A antigens. **(C)** RBD-215 variants were IMAC-purified from the apoplastic fluid of infiltrated line ΔXT/FT and subjected to SDS-PAGE under reducing and non-reducing conditions. **(D)** SDS-PAGE and immunoblotting of IMAC-purified RBD-215 variants with anti-blood group A (3-3A) or anti-His-tag antibodies. **(E)** PNGase F digestion of IMAC-purified RBD-215A.

group A structures we co-expressed either human β1,4-galactosyltransferase (B4GALT) (Strasser et al., 2009) or *A. thaliana* β1,3-galactosyltransferase (GALT1) (Strasser et al., 2007), with one human α1,2-fucosyltransferase (FUT1 or FUT2), and the human ABO A enzyme transiently in ΔXT/FT *N. benthamiana*. To avoid interference from other N-glycan processing steps, the catalytic domain of B4GALT (type 2 chain formation) was targeted to the *trans*-Golgi using the N-terminal targeting sequence from rat α2,6-sialyltransferase (ST) (Boevink et al., 1998; Strasser et al., 2009). The catalytic domains of FUT1 and FUT2 were also fused to the ST-region to enable transfer of the fucose to the galactose in the same Golgi compartment. In addition, GFP was attached to the C-terminal end of the glycosyltransferase to enable monitoring of the subcellular localization (**Figure 2C**). *A. thaliana* GALT1 resides in the *trans*-Golgi (Strasser et al., 2007) and therefore the native sequence was expressed to achieve efficient β1,3-galactosylation (type 1 chain formation, **Figure 1**). The human ABO A enzyme was expressed as a chimeric protein with the catalytic domain fused to ST and GFP (ST-ABO A) and as a native enzyme without any foreign targeting signal or tag. ST-FUT1, ST-FUT2 and ST-ABO A expressed well in plants and displayed Golgi localization (**Supplementary Figure S1**).

RBD-215 co-expressed with the different glycosyltransferases was purified from crude protein extracts and subjected to SDS-PAGE and immunoblotting with blood group A-specific antibodies (**Figure 3A**; **Supplementary Figure S1**). Co-expression of ST-B4GALT, ABO A and either ST-FUT1 or ST-FUT2 (ST-FUT2 was less efficient than ST-FUT1) resulted in reactivity with the blood-group A-specific antibody suggesting the successful formation of type 2 chains. Detection of RBD-215 with the anti-His antibody resulted in a shift in mobility that was consistent with the N-glycan elongation mediated by the co-expressed blood group A-specific glycosyltransferases. Expression of ST-ABO A led to the formation of RBD-215 that reacted with the blood-group A-specific antibody to a similar extent indicating that ST-ABO A and native ABO A are both functional and targeted to a late Golgi compartment when transiently expressed in plants (**Supplementary Figure S1, S2**).

UDP-GalNAc, the nucleotide sugar for the ABO A glycosyltransferase, is not very abundant in plants (Daskalova et al., 2010). However, we have previously shown that O-glycan engineering in plants can be optimized by co-expression of a *Yersinia enterocolitica* UDP-GlcNAc 4-epimerase (YeGNE) capable of converting UDP-GlcNAc to UDP-GalNAc and a *Caenorhabditis elegans* UDP-GlcNAc/UDP-GalNAc transporter (CeT) for increased transport of the donor substrate into the Golgi lumen (Castilho et al., 2012). Therefore, we examined whether these proteins improve the biosynthesis of blood group A type 2 structures. Immunoblotting revealed that co-expression results in a stronger signal with the blood group A-specific antibody (**Figure 3B**). This was further confirmed using a different blood group A-specific antibody (**Supplementary Figure S2**). Our initial glycoengineering approach suggested that blood group A type 1 structures are not efficiently produced on RBD-215 (**Figure 3A**). However, when we expressed GALT1, ST-FUT2 and ST-ABO A together with YeGNE and CeT we could modify

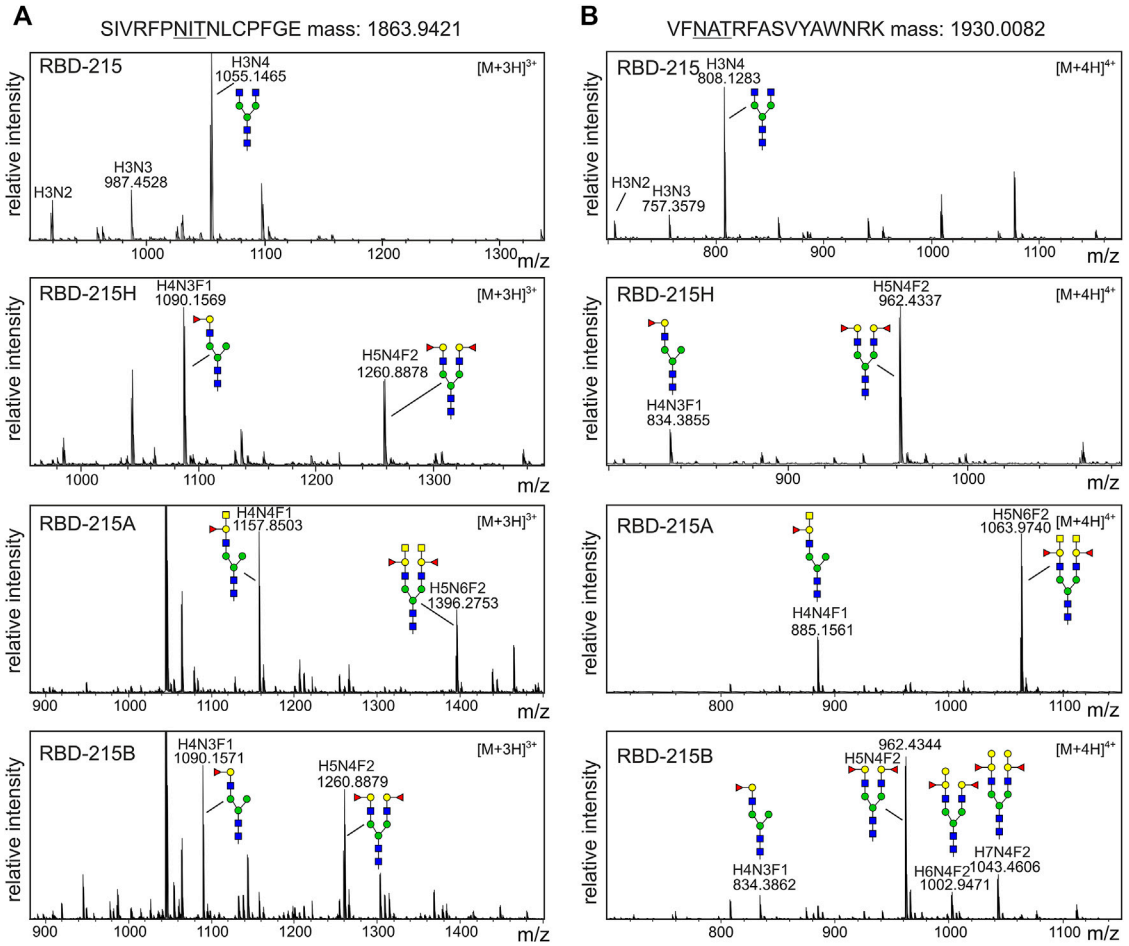


FIGURE 4 | MS spectra of the RBD-215 glycopeptides carrying the N-glycosylation site N331 or N343. IMAC-purified RBD-215 variants were proteolytically digested and analyzed by MS. **(A)** Glycopeptides carrying N331 and **(B)** N343. Major N-glycan peaks are illustrated with a cartoon presentation (see **Figure 2** for details). H: hexose; N: N-acetylhexosamine (HexNAc); F: fucose.

N-glycans on RBD-215 with blood group A type 1 chains (**Supplementary Figure S3**).

Next, we purified RBD-215 variants with different blood group structures from the apoplastic fluid by immobilized metal affinity chromatography (IMAC) and analyzed the purified proteins by SDS-PAGE under reducing and non-reducing conditions (**Figure 3C**). Since the blood group A type 2 structures were more efficiently generated than type 1 structures, we focused only on the characterization of the former. Under reducing conditions, the RBD-215 proteins migrated at the expected positions. Under non-reducing conditions, a faster migration was observed for all variants which is likely caused by the presence of four disulfide bonds leading to a more compact shape. Compared to RBD-215, reduced mobility was detected for all glycoengineered variants and the blood-group A-specific antibody reacted only with the purified RBD-215 protein that was co-expressed with the ABO A glycosyltransferase (RBD-215A, **Figure 3D**), but not with the RBD-215 co-expressed with glycosyltransferases for H (RBD-215H), and B (RBD-215B) antigen formation. Upon PNGase F digestion of RBD-215A, the reactivity with the blood-group A-specific antibody was completely lost showing the presence of the modification on N-glycans (**Figure 3E**).

All four RBD-215 variants purified from the apoplastic fluid were proteolytically digested and the glycopeptides were analyzed by LC-ESI-MS. While RBD-215 carried mainly GnGn structures on both N-glycosylation sites, RBD-215H carried substantial amounts of mono- (major N-glycan at N331) and bi-antennary N-glycans (major N-glycan at N343) corresponding to H-type structures (**Figures 4A,B**; **Supplementary Figure S4**). RBD-215A that was obtained by additional co-expression of ST-ABO A displayed further elongation of the two N-glycan branches with an additional HexNAc residue. In line with the immunoblot data this indicates the successful formation of blood group A carbohydrate structures with terminal GalNAc. On the other hand, modification of H-type structures with galactose by co-expression of Golgi-targeted ST-ABO B (**Supplementary Figure S1**) was less efficient and only detected on the N-glycans at site N343 (**Figure 4B**). This could be either due to differences in catalytic activity of the ST-ABO B enzyme (Letts et al., 2006) or due to the removal of terminal galactose residues

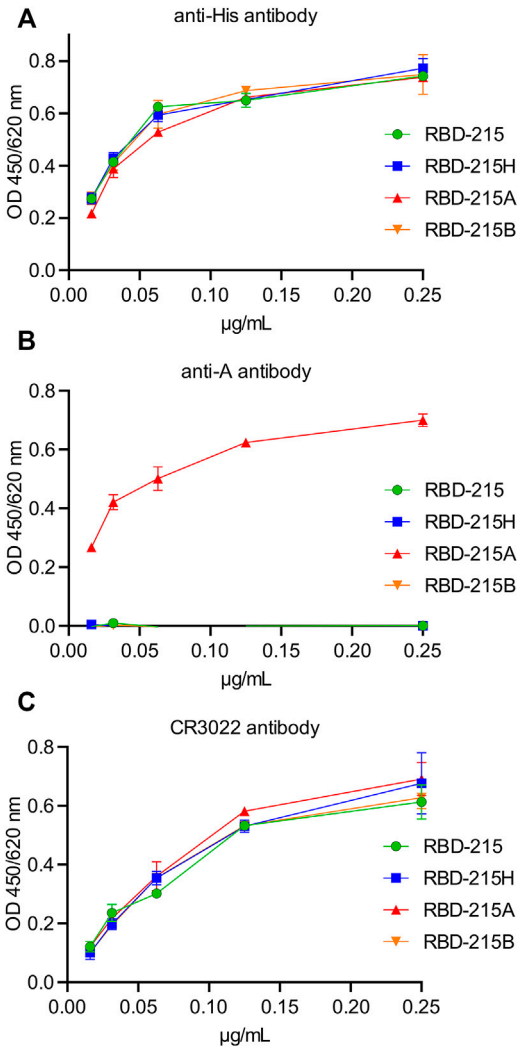


FIGURE 5 | The anti-A antibody specifically reacts with RBD-215 carrying blood group A antigen structures. IMAC-purified RBD-215 variants were coated and ELISA was carried out with different concentrations of **(A)** antibodies against the His-tag, **(B)** the 3-3A antibody against anti-blood group A antigen structures and **(C)** CR3022, a conformation dependent anti-RBD antibody. Values represent the mean ± SD (n = 3).

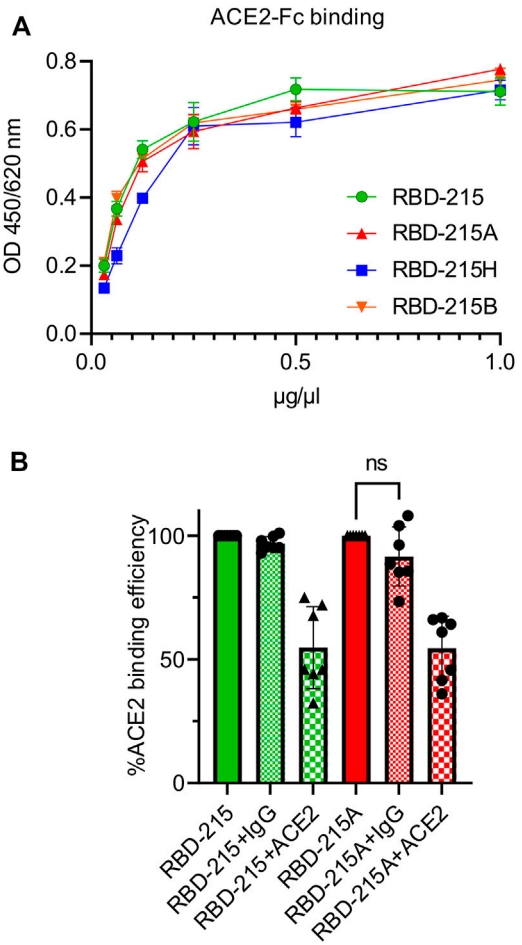


FIGURE 6 | RBD-215 glycoforms display binding to ACE2-Fc. **(A)** ACE2-Fc binding ELISA. Binding curves of different concentrations of IMAC-purified RBD-215 glycoforms to plates coated with ACE2-Fc. Values represent the mean ± SD (n = 3). **(B)** ACE2-Fc competition ELISA. RBD-215 or RBD-215A were premixed with 3-3A anti-blood group A antibody (+IgG) or ACE2-Fc (+ACE2) and added to ELISA plates coated with ACE2-Fc. Detection was done with anti-His antibodies. Values represent the mean ± SD (n = 7). A student's t-test was used to compare the difference in ACE2-Fc binding between RBD-215A and RBD-215A + IgG. "n.s." not significant (p = 0.0870).

by galactosidases present in the apoplast (Kriechbaum et al., 2020). Taken together, this shows that blood group A structures can be efficiently produced on complex N-glycans of recombinant RBD-215 in *N. benthamiana* using a transient glycoengineering approach.

RBD With Blood Group A Antigens Is Functional

The presence of distinct complex N-glycan modifications on RBD-215 may cause structural changes that affect the protein conformation. To assess this possibility ELISA was performed using a conformation-dependent RBD antibody. With the blood group A-specific antibody, binding was only detected with RBD-215A demonstrating the high specificity of the antibody (**Figures 5A,B**). The conformation-dependent RBD antibody CR3022 (Yuan et al., 2020) displayed comparable reactivity with all four variants suggesting that the N-glycan modifications do not have a major impact on the overall conformation of the viral antigen (**Figure 5C**).

Binding of anti-blood group A antibodies to SARS-CoV-2 virus derived from blood group A positive cells could impair the interaction with the cellular receptor and subsequently the cell entry. We carried out an ACE2-binding ELISA to see if the presence of the blood group A-specific antibody interferes with binding of recombinant RBD-215 to immobilised ACE2-Fc. All four RBD-215 glycoforms showed binding to ACE2-Fc suggesting that N-glycan processing and formation of blood

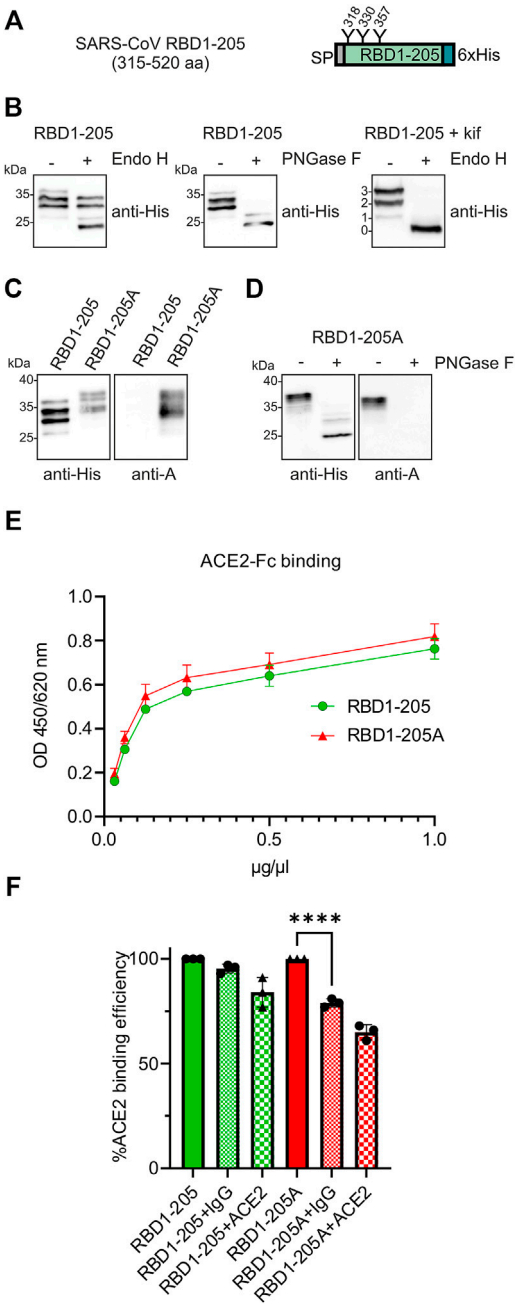


FIGURE 7 | RBD1-205 N-glycans can be engineered to carry blood group A antigens. **(A)** Illustration of the plant-expressed RBD1-205 (amino acid region 315–520 of the SARS-CoV-1 spike protein). **(B)** Immunoblot analysis of IMAC-purified RBD1-205. Deglycosylation was done by digestion with Endo H or PNGase F. To produce RBD1-205 with Endo H-sensitive oligomannosidic N-glycans, RBD1-205 was expressed in the presence of the α-mannosidase inhibitor kifunensine (kif). 0, 1, 2, and 3 indicate the number of N-glycans present on RBD1-205. **(C)** Immunoblot analysis with anti-His and anti-blood group A (3-3A) antibodies **(D)** PNGase F digestion of RBD1-205. **(E)** ACE2-Fc binding ELISA. Binding of different concentrations of IMAC-purified RBD1-205 and RBD1-205A glycoforms to plates coated with ACE2-Fc. Values represent the mean ± SD (n = 3). **(F)** ACE2-Fc competition ELISA. RBD1-205 or RBD1-205A were premixed with the 3-3A anti-blood group A antibody (+IgG) or ACE2-Fc (+ACE2), and added to ELISA plates coated with ACE2-Fc. Detection was done with anti-His antibodies. Values represent the mean ± SD (n = 3). A student's t-test was used to compare the difference in ACE2-Fc binding between RBD1-205A and RBD1-205A + IgG. "****" $p < 0.0001$.

(*Continued*)

group carbohydrates does not interfere with the receptor binding (**Figure 6A**). Next, we carried out a competition ELISA to examine whether anti-blood group A-specific antibodies block interaction of RBD-215A and ACE2-Fc. As a control for the competition ELISA, we added a soluble ACE2-Fc that competes with the immobilized ACE2-Fc for binding (Gattinger et al., 2021). In the presence of soluble ACE2-Fc, reduced binding of RBD-215 and RBD-215A to ACE2-Fc was detected (**Figure 6B**). By contrast, the anti-blood group A antibody had only a minor effect on the RBD-215A ACE2-Fc interaction. This finding is consistent with previous studies reporting that the RBD N-glycans do not clash with the receptor-binding motif (RBM) on the SARS-CoV-2 spike (Pinto et al., 2020).

To examine whether the same finding is observed for the RBD from SARS-CoV-1, we generated a recombinant protein carrying amino acids 315–520 of the SARS-CoV-1 spike protein (RBD1-205). In contrast to RBD-215, RBD1-205 from SARS-CoV-1 carries three N-glycosylation sites (**Figure 7A**). Purified RBD1-205 displayed several bands on immunoblots with the anti-His antibody. Deglycosylation of RBD1-205 with PNGase F or deglycosylation of an RBD1-205 variant carrying oligomannosidic N-glycans with Endo H revealed that the different bands are derived from incomplete occupancy of the three N-glycosylation sites (**Figure 7B**). The RBD1-205 variant was transiently expressed in *N. benthamiana* with or without the machinery (ST-GALT1, ST-FUT1, ST-ABO A, YeGNE, and CeT) for blood group A type 2 antigen formation. Like RBD-215, also RBD1-205 from SARS-CoV-1 could be modified with complex N-glycans that are recognized by the blood-group A-specific antibody on immunoblots (RBD1-205A, **Figures 7C,D**; **Supplementary Figure S5**) and RBD1-205 as well as RBD1-205A interacted with ACE2-Fc (**Figure 7E**). In contrast to RBD-215A, the blood-group A-specific antibody blocked the significantly engagement of RBD1-205A with ACE2-Fc (**Figure 7F**). This could be related to the presence of an additional N-glycan or the overall reduced affinity of the SARS-CoV-1 RBD for the ACE2 receptor.

Blood Group A Antigen Structures Are Specifically Recognized by Natural Antibodies Present in Sera of Blood Group O and B Individuals

To examine whether natural antibodies present in sera from RBD-positive (SARS-CoV-2 exposed or vaccinated) and RBD-negative individuals differentially react with RBD-215 variants produced in plants using our glycoengineering approach, we carried out a multiplex bead-based assay with coupled recombinant viral antigens (Klausberger et al., 2021; Schwestka et al., 2021). In the RBD-positive cohorts, a comparison of the signal intensity did not reveal significant differences among the

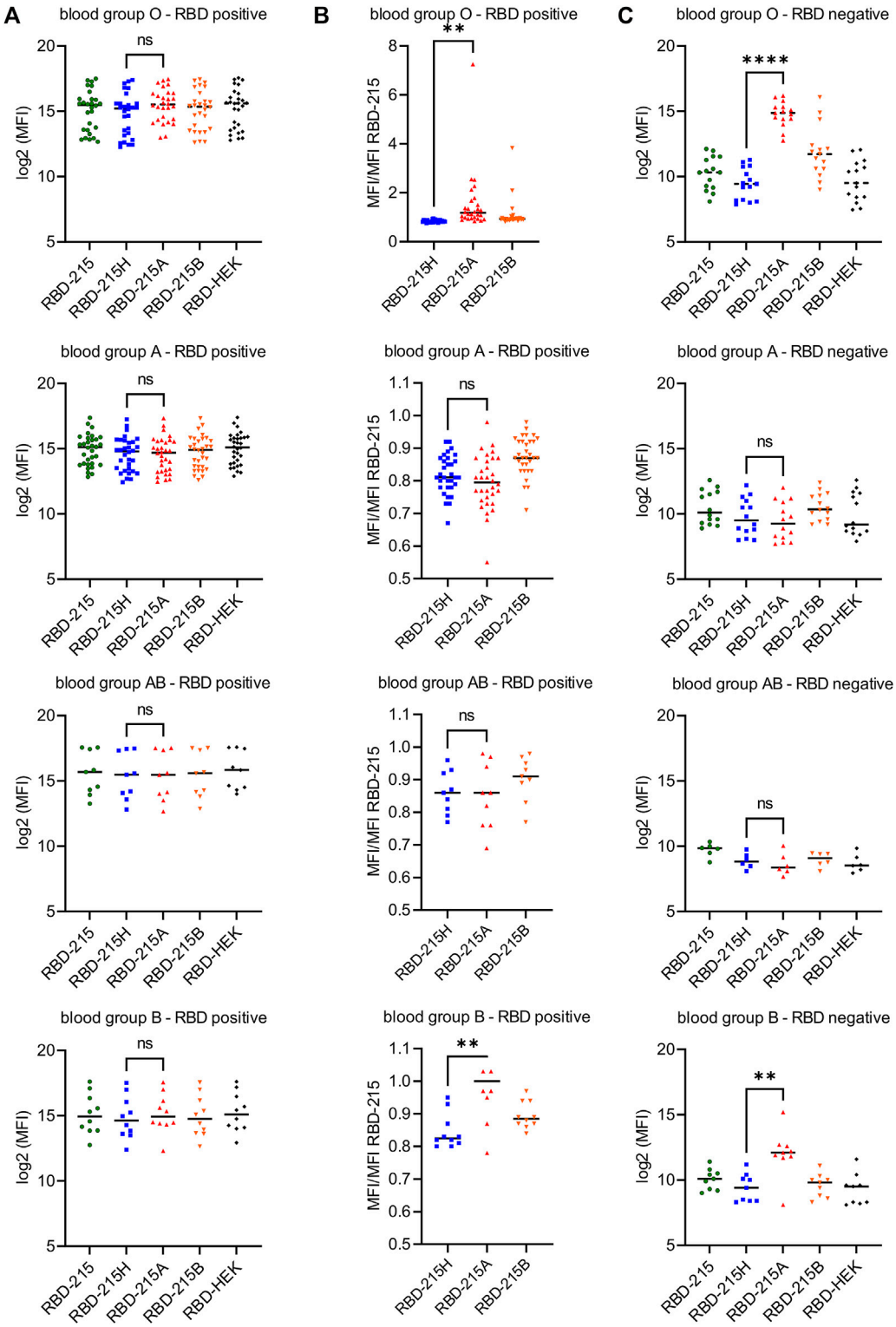


FIGURE 8 | Sera from blood group O and B individuals react strongly with RBD-215A carrying blood group A antigens in a multiplex bead-based assay. **(A)** Reactivity of confirmed RBD positive sera (according to the WHO standard) to different RBD-215 glycoforms. RBD produced in HEK293 cells (RBD-HEK) was included for comparison. The line depicts the median in each group. A student's t-test was used to compare the reactivity to RBD-215H (carrying the H antigen) and RBD-215A (carrying the A type 2 antigen). "n.s." not significant; "**" $p < 0.01$; "****" $p < 0.0001$. Blood group O ($n = 28$), blood group A ($n = 32$), blood group AB ($n = 9$), blood group B ($n = 10$). **(B)** MFI values were normalized to the MFI values for RBD-215. The line depicts the median. A student's t-test was used to compare the normalized values of RBD-215H and RBD-215A. **(C)** Reactivity of confirmed RBD negative sera (according to the WHO standard) to different RBD-215 glycoforms and RBD-HEK. Blood group O ($n = 15$), blood group A ($n = 14$), blood group AB ($n = 6$), blood group B ($n = 9$).

RBD-215 variants (**Figure 8A**). However, when the IgG binding intensity to the RBD-215 variants with blood group antigen structures was normalized to the signal from the unmodified RBD-215, a significant difference was revealed between RBD-215H and RBD-215A in the blood group O and B cohorts (**Figure 8B**; **Supplementary Table S1**). In line with this finding, sera from blood group O and B individuals showed significantly enhanced reactivity in the RBD-negative cohorts. In some RBD-negative sera with blood group O, the IgG binding signal was in the same range as observed for RBD-positive sera (**Figure 8C**) suggesting that those sera could provide enhanced protection for transmission of SARS-CoV-2 or similar coronaviruses from blood group A donors.

DISCUSSION

Glycosylation of viral proteins is not only important for protein quality control and folding (Margolin et al., 2020; Shin et al., 2021), but also for shielding of epitopes (Watanabe et al., 2020b) and interaction with cellular receptors (Hoffmann et al., 2021). Specific interference with receptor binding will lead to novel strategies to prevent infection of cells and tissues. In an earlier study it was shown that human anti-A antibodies inhibited the adhesion of SARS-CoV-1 spike to ACE2-expressing cells (Guillon et al., 2008). The recombinant SARS-CoV-1 spike protein was produced in engineered mammalian cells capable of blood group A biosynthesis and ACE2-expressing Vero cells used in the cell-based assay did not express the blood group A antigen. It is therefore plausible that anti-A antibodies can bind to blood group A antigens present on the heavily glycosylated SARS-CoV-1 spike protein. Our data for recombinant RBD1-205 with blood group A structures are in line with these data. The binding of anti-glycan antibodies could block the spike ACE2 interaction and protect against viral transmission from blood group A individuals to blood group O or B individuals. In analogy to SARS-CoV-1, the increased susceptibility of distinct blood groups to SARS-CoV-2 infection could be related to the presence of naturally circulating anti-A antibodies, which interfere with or even inhibit the virus–cell adhesion process and the subsequent transmission of the virus (Breiman et al., 2020; Goel et al., 2021). Such anti-glycan antibodies are considered part of the innate immune system to combat different bacterial and viral pathogens that carry specific glycan antigens (Cooling, 2015; Galili, 2020). The potential role in anti-viral activity of anti-blood group antibodies is already known for some time, for example, for anti-A antibody-mediated neutralization of HIV (Arendrup et al., 1991). Our data show that the presence of the blood group A antigen on the RBD N-glycans from SARS-CoV-2 does not significantly interfere with ACE2 receptor binding. Overall, this is not surprising because the N-glycosylation sites do not overlap with the amino acids involved in ACE2 binding (Piccoli et al., 2020; Yan et al., 2020). The absence of a major effect in the competition assay does not rule out the possibility that other N- or O-glycans with blood group A antigens on the SARS-CoV-2 spike protein have an impact on ACE2 binding. Alternatively, the binding of the anti-carbohydrate antibodies to viral glycans may cause cross linking or affect the protein conformation and dynamics as has been suggested for the SARS-CoV-2 neutralizing antibody S309 (Pinto et al., 2020). S309 is a non-RBM binding antibody that is targeted to an epitope containing the N-glycan attached at N343. On RBD-215 this N-glycosylation site was also highly modified with blood group A antigen structures and anti-A antibodies binding to the N-glycan at N343 may be implicated in virus neutralization by a similar mechanism as described for S309. Importantly, sotrovimab, a derivative of S309 has recently received an emergency use authorization (EUA) from the FDA for preventing severe COVID-19 disease and sotrovimab/S309 shows neutralization against evolving variants like Omicron which highlights the potential of these group of antibodies (Cameroni et al., 2021; Corti et al., 2021).

The SARS-CoV-2 spike protein monomer harbours 22 N-glycosylation sites that are all to a large extent glycosylated (Watanabe et al., 2020a; Zhao P. et al., 2020; Shajahan et al., 2021). In addition, more than 30 O-glycosylation sites have been predicted and some of them experimentally confirmed on recombinantly expressed spike protein variants including RBD (Zhao P. et al., 2020; Bagdonaite et al., 2021; Brun et al., 2021; Shajahan et al., 2021). Hence, there are numerous potential N- and O-glycosylation sites for the formation of blood group A antigens. So far, the studies focusing on the spike protein glycosylation in human cells did not provide evidence for the presence of H-, A- or B-type structures. However, this is not surprising as the used expression systems for recombinant viral proteins (HEK293, CHO, and insect cells) and host cells used for virus infection assays (e.g., Vero cells) lack the corresponding glycosyltransferase activities (Guillon et al., 2008; Lindberg et al., 2013; El Jellas et al., 2018). Therefore, it is essential to use cell lines or tissue organoids capable of producing a diverse range of glycans to study the contribution of individual modifications to virus infection or immunogenicity. In addition to the native glycosylation capacity of the used cells, virus infection may lead to a transcriptional reprogramming of the host cell resulting in the induction of glycosyltransferase expression that can modulate the cell-surface and viral glycans as has been recently shown for HIV infection (Colomb et al., 2020). All these possibilities must be considered when studying the impact of glycans on viral infection and transmission.

The current SARS-CoV-2 pandemic very drastically shows that we are not well equipped to cope with such an outbreak and numerous efforts are necessary to fight pathogens and increase the preparedness for future emerging or re-emerging viral threats. Recombinant anti-blood group antibodies (e.g., against blood group A antigen structures) could be used as protective drugs that are immediately available to fight newly emerging viruses, especially those that are heavily glycosylated, and carry a glycan shield for immune evasion (Watanabe et al., 2020b). Such recombinant anti-A antibodies could be administered in large quantity via nasal delivery as IgA or IgM formats (Ku et al., 2021) to blood group O or B individuals with low natural antibody levels. This may contribute to prevent initial virus uptake and thus reduce the spread of newly emerging

respiratory viruses in populations until more specific drugs are developed.

MATERIALS AND METHODS

Cloning of Expression Vectors

The generation of the pEAQ-*HT* expression vector for RBD-215 (SARS-CoV-2 RBD) was described previously (Shin et al., 2021). For RBD1-205 (SARS-CoV-1 RBD) expression, a synthetic DNA fragment (GeneArt, Thermo Fisher Scientific) coding for the barley α-amylase signal peptide fused to the N-terminal end of the spike domain (amino acids 315–520), and a 6x-histidine tag at the C-terminal end was cloned into the *Age*I/*Xho*I sites of pEAQ-*HT* (Sainsbury et al., 2009). Expression constructs for GALT1 (p43-GALT1) (Schwestka et al., 2021), ST-B4GALT (ST-GalT) (Strasser et al., 2009), *C. elegans* UDP-GlcNAc/UDP-GalNAc transporter (CeT: pH7WG2:GT), and *Y. enterocolitica* UDP-GlcNAc 4-epimerase (YeGNE: pH7WG2:GE) (Castilho et al., 2012) were available from previous studies. The ST-FUT1 expression vector was generated by cloning of a synthetic DNA fragment coding for a ST-FUT1 chimeric protein (amino acids 1–52 from rat α2,6-sialyltransferase fused to amino acids 37–365 from human FUT1) into p20F (Strasser et al., 2007) to generate p20-ST-FUT1. For ST-FUT2, the FUT2 coding sequence (amino acids 33–343 from human FUT2) was inserted into the *Bam*HI site of p20-ST to generate p20-ST-FUT2. The expression vector (pPT2M-ABO-A) for the full-length untagged ABO A transferase was generated by insertion of a synthetic codon-optimized fragment of the human ABO A coding region into the *Xba*I/*Bam*HI sites of pPT2M (Strasser et al., 2005). For the ST-ABO A expression vector, the coding region for amino acids 47–354 of human ABO A was cloned into the *Bam*HI site of p20-ST and for ST-ABO B the coding region for amino acids 47–354 of human ABO B was cloned in the same manner into p20-ST.

Protein Expression and Purification

Syringe-mediated agroinfiltration of leaves from 5-week-old *N. benthamiana* ΔXT/FT (Strasser et al., 2008) was used for transient expression of RBD-215 variants. For the generation of blood group antigens, the enzymes were co-expressed by mixing of agrobacteria prior to agroinfiltration. For RBD-215 purification from crude extracts, leaves were harvested 3 days after infiltration, frozen in liquid nitrogen and homogenized using metallic beads and a mixer mill (Retsch). The homogenized material was resuspended in 3 volumes of 20 mM Na_2HPO_4, 100 mM NaCl (pH 7.4). Upon centrifugation to remove insoluble material, NaCl and imidazole was added to adjust the buffer to 20 mM Na_2HPO_4, 500 mM NaCl, and 10 mM imidazole (pH 7.4). His-tagged RBD-215 was purified using 100 µl His Mag Sepharose® Ni (Cytiva) according to the manufacturer's instructions. Bound proteins were eluted with 50 µl 20 mM Na_2HPO_4, 500 mM NaCl, and 500 mM imidazole (pH 7.4), mixed with sample loading buffer and used for SDS-PAGE and immunoblotting. For large scale purification of His-tagged proteins via immobilized metal affinity chromatography (IMAC), apoplastic fluid was collected from infiltrated leaves by low-speed centrifugation as described previously (Schwestka et al., 2021). After filtration through a 0.45 µm membrane filter (Merck Millipore), collected apoplastic fluid containing His-tagged RBD-215 in 500 mM NaCl, 20 mM Na_2HPO_4 and 10 mM imidazole (pH 7.4) was purified using a 1 ml HisTrap FF column (Cytiva) and the ÄKTA pure chromatography system (Cytiva). After washing with 15 column volumes and buffer containing 40 mM imidazole, bound proteins were eluted with 20 column volumes and buffer containing 250 mM imidazole. Fractions containing the protein of interest were pooled and dialyzed overnight against phosphate-buffered saline (PBS, pH 7.4) using SnakeSkin dialysis tubing (Thermo Fisher Scientific) with a 10 kDa molecular mass cutoff. Protein samples were then further concentrated using 10 kDa Amicon Ultra centrifugal filters (Merck Millipore). Purification of RBD-His (RBD-HEK) and human soluble ACE2-Fc from HEK293 cells has been described recently (Klausberger et al., 2021).

Immunoblot Analysis

Purified proteins were subjected to SDS-PAGE under reducing or non-reducing conditions. Samples to be analyzed under non-reducing conditions were not boiled prior to loading. Separated proteins were either stained with Coomassie Brilliant Blue (Sigma-Aldrich) or transferred to a nitrocellulose membrane (Cytiva) and detected using anti-His (Thermo Fisher Scientific), anti-blood group A (3-3A—IgG antibody, Novus Biologicals), anti-blood group A (Z2A—IgM antibody, Santa Cruz Biotechnology), and JIM84 (Strasser et al., 2007) antibodies. For deglycosylation, proteins were denatured and incubated with or without Endo H or PNGase F (both from NEB) according to the manufacturer's instructions.

Liquid Chromatography-Electrospray Ionization-Mass Spectrometry (LC-ESI-MS).

Purified RBD-215 proteins were S-alkylated with iodoacetamide and digested in solution with endoproteinases LysC (Roche) and GluC (Promega). Digested samples were analyzed using a maXis 4G QTOF mass spectrometer (Bruker) as described (Klausberger et al., 2021).

RBD and ACE2-Fc Binding ELISA

ELISA was carried out as described in detail recently (Schwestka et al., 2021). Briefly, 96-well plates (Nunc MaxiSorp™, Thermo Fisher Scientific) were coated overnight at 4°C with the indicated concentrations of purified RBD or ACE2-Fc proteins in PBS. For RBD-215 or RBD1-205 detection, plates were incubated for 2 h with the mouse-anti-His antibody (Thermo Fisher Scientific), anti-RBD antibody CR3022 (Klausberger et al., 2021; Shin et al., 2021) or anti-blood group A antibody (3-3A, Novus Biologicals). Plates incubated with anti-His or anti-blood group A antibody were washed again and incubated for 1 h with anti-mouse-HRP antibody (Sigma-Aldrich). CR3022 binding was analyzed using anti-human IgG (H + L)-HRP antibody (Promega). For detection, 150 µl/well of a 3,3′,5,5′-tetramethylbenzidin (TMB, Sigma-Aldrich) solution was added (1:60 of 0.4% TMB, 1:300 of

0.6% H_2O_2 in 50 mM phosphate-citrate buffer pH 5.0) and the reaction was stopped with 2 M H_2SO_4 (50 µl/well). The absorbance was measured at 450 nm using a TECAN Spark plate reader. Background resulting from unspecific binding of detection antibodies was subtracted from the obtained values. Three technical replicates were performed. Data were analyzed using GraphPad Prism Version 9.1.1.

ACE2 Competition ELISA

For the ACE2 competition ELISA assay, 150 ng/well of ACE2-Fc in 50 µl PBS was coated onto F69 MaxiSORP Nunc-Immuno plates (Thermo Fisher Scientific) overnight at 4°C. After washing with PBS supplemented with 0.1% (v/v) Tween-20 (PBST), the plates were blocked with 1% (w/v) BSA in PBST for 1 h at room temperature. For the pre-incubation, 35 µl of 0.25 µg/ml RBD variants were mixed with 35 µl 1% (w/v) BSA in PBST, 35 µl 12.5 µg/ml anti-blood group A antibody (3-3A, Novus Biologicals) or 35 µl 14 µg/ml ACE2-Fc and incubated for 30 min at room temperature. Subsequently, 50 µl of pre-incubated sample mix was transferred onto the ACE2-Fc coated MaxiSORP plates and incubated for 2 h at room temperature. His-tagged RBD samples were detected with biotinylated mouse-anti-His antibody (Invitrogen), followed by incubation with a streptavidin-HRP conjugate (Roche). The chromogenic signal was developed using TMB as a substrate solution and analyzed as described for the binding ELISA.

Luminex Assays

The RBD-215 variants and RBD-HEK were separately coupled to MagPlex carboxylated polystyrene microspheres (Luminex Corporation) and coupled glycoforms were then assayed in parallel with sera from different individuals collected at the AIT (AIT cohort). Coupling as well as Luminex assays were performed according to the manufacturer's instructions with minor modifications as described in detail recently (Klausberger et al., 2021). AIT cohort comprises samples collected for routine SARS-CoV-2 serodiagnosis from 111 SARS-CoV-2 infected, uninfected and/or vaccinated individuals. Seronegativity or seropositivity has been determined via an eight-plex Luminex-based serotest and was based on cut-off values and end-point titers defined according to Frey et al., (1998) on the basis of 160 pre-COVID-19 sera.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethikkommission der Stadt Wien,Thomas-Klestil-Platz 8/2, 1030 Wien. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

JK-B, Y-JS, UV, CV, CG-G, MH, YG, JH and KV conducted the experiments. JK-B, CG-G, MH, AW and RS analyzed the results. DM provided study materials and data. RS conceptualized the study and wrote the paper with support from JK-B All authors have made a substantial and intellectual contribution to the work and approved it for publication.

FUNDING

This work was supported by the Austrian Science Fund (FWF) Project P31920-B32. This project was further supported by the BOKU COVID-19 Initiative and the BOKU Core Facilities Multiscale Imaging and Mass Spectrometry.

ACKNOWLEDGMENTS

We thank Professor George Lomonossoff (John Innes Centre, Norwich, UK) and Plant Bioscience Limited (PBL) (Norwich, UK) for supplying the pEAQ-HT expression vector, Doris Lucyshyn (BOKU) for the kind gift of anti-mouse IgM antibodies and Lukas Mach (BOKU) for providing ACE2-Fc. HEK293 cell-derived RBD protein and antibody CR3022 were kindly provided through the BOKU COVID-19 platform (https://portal.boku-covid19.at/), which enables researcher access to COVID-related protein reagents. The MS equipment was kindly provided by the EQ-BOKU VIBT GmbH and the BOKU Core Facility Mass Spectrometry. We thank Rudolf Figl (BOKU, Core Facility Mass Spectrometry) for technical assistance with the MS analysis and Dr. Silvia Reisner-Reininger, MD, for help with patient sample collection and sample annotations.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.816544/full#supplementary-material

REFERENCES

Arend, P. (2021). Why Blood Group A Individuals Are at Risk whereas Blood Group O Individuals Are Protected from SARS-CoV-2 (COVID-19) Infection: A Hypothesis Regarding How the Virus Invades the Human Body via ABO(H) Blood Group-Determining Carbohydrates. *Immunobiology* 226, 152027. doi:10.1016/j.imbio.2020.152027

Arendrup, M., Hansen, J.-E. S., Clausen, H., Nielsen, C., Mathiesen, L. R., and Nielsen, J. O. (1991). Antibody to Histo-Blood Group A Antigen Neutralizes HIV Produced by Lymphocytes from Blood Group A Donors but Not from Blood Group B or O Donors. *AIDS* 5, 441–444. doi:10.1097/00002030-199104000-00014

Bagdonaite, I., Thompson, A. J., Wang, X., Søgaard, M., Fougeroux, C., Frank, M., et al. (2021). Site-Specific O-Glycosylation Analysis of SARS-CoV-2 Spike Protein Produced in Insect and Human Cells. *Viruses* 13, 551. doi:10.3390/v13040551

Boevink, P., Oparka, K., Cruz, S. S., Martin, B., Betteridge, A., and Hawes, C. (1998). Stacks on Tracks: the Plant Golgi Apparatus Traffics on an Actin/ER Network†. *Plant J.* 15, 441–447. doi:10.1046/j.1365-313x.1998.00208.x

Breiman, A., Ruvën-Clouet, N., and Le Pendu, J. (2020). Harnessing the Natural Anti-glycan Immune Response to Limit the Transmission of Enveloped Viruses Such as SARS-CoV-2. *Plos Pathog.* 16, e1008556. doi:10.1371/journal.ppat.1008556

Brun, J., Vasiljevic, S., Gangadharan, B., Hensen, M., Chandran, A. V., Hill, M. L., et al. (2021). Assessing Antigen Structural Integrity through Glycosylation Analysis of the SARS-CoV-2 Viral Spike. *ACS Cent. Sci.* 7, 586–593. doi:10.1021/acscentsci.1c00058

Cameroni, E., Saliba, C., Bowen, J. E., Rosen, L. E., Culap, K., Pinto, D., et al. (2021). Broadly Neutralizing Antibodies Overcome SARS-CoV-2 Omicron Antigenic Shift. *bioRxiv preprint*.14 472269 doi:10.1101/2021.12.12.472269

Castilho, A., Neumann, L., Daskalova, S., Mason, H. S., Steinkellner, H., Altmann, F., et al. (2012). Engineering of Sialylated Mucin-type O-Glycosylation in Plants. *J. Biol. Chem.* 287, 36518–36526. doi:10.1074/jbc.m112.402685

Cheng, Y., Cheng, Y., Cheng, G., Chui, C. H., Lau, F. Y., Chan, P. K., et al. (2005). ABO Blood Group and Susceptibility to Severe Acute Respiratory Syndrome. *JAMA* 293, 1450–1451. doi:10.1001/jama.293.12.1450-c

Colomb, F., Giron, L. B., Kuri-Cervantes, L., Adeniji, O. S., Ma, T., Dweep, H., et al. (2020). Sialyl-LewisX Glycoantigen Is Enriched on Cells with Persistent HIV Transcription during Therapy. *Cel Rep.* 32, 107991. doi:10.1016/j.celrep.2020.107991

Cooling, L. (2015). Blood Groups in Infection and Host Susceptibility. *Clin. Microbiol. Rev.* 28, 801–870. doi:10.1128/cmr.00109-14

Corti, D., Purcell, L. A., Snell, G., and Veesler, D. (2021). Tackling COVID-19 with Neutralizing Monoclonal Antibodies. *Cell* 184, 4593–4595. doi:10.1016/j.cell.2021.07.027

Daskalova, S. M., Radder, J. E., Cichacz, Z. A., Olsen, S. H., Tsaprailis, G., Mason, H., et al. (2010). Engineering of N. Benthamiana L. Plants for Production of N-Acetylgalactosamine-Glycosylated Proteins - towards Development of a Plant-Based Platform for Production of Protein Therapeutics with Mucin Type O-Glycosylation. *BMC Biotechnol.* 10, 62. doi:10.1186/1472-6750-10-62

Deleers, M., Breiman, A., Daubie, V., Maggetto, C., Barreau, I., Besse, T., et al. (2021). Covid-19 and Blood Groups: ABO Antibody Levels May Also Matter. *Int. J. Infect. Dis.* 104, 242–249. doi:10.1016/j.ijid.2020.12.025

El Jellas, K., Johansson, B. B., Fjeld, K., Antonopoulos, A., Immervoll, H., Choi, M. H., et al. (2018). The Mucinous Domain of Pancreatic Carboxyl-Ester Lipase (CEL) Contains Core 1/core 2 O-Glycans that Can Be Modified by ABO Blood Group Determinants. *J. Biol. Chem.* 293, 19476–19491. doi:10.1074/jbc.ra118.001934

Ellinghaus, D., Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Albillos, A., et al. (2020). Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *N. Engl. J. Med.* 383, 1522–1534. doi:10.1056/NEJMoa2020283

Frey, A., Di Canzio, J., and Zurakowski, D. (1998). A Statistically Defined Endpoint Titer Determination Method for Immunoassays. *J. Immunological Methods* 221, 35–41. doi:10.1016/s0022-1759(98)00170-7

Galili, U. (2020). Human Natural Antibodies to Mammalian Carbohydrate Antigens as Unsung Heroes Protecting against Past, Present, and Future Viral Infections. *Antibodies* 9, 25. doi:10.3390/antib9020025

Gallinaro, L., Cattini, M. G., Sztukowska, M., Padrini, R., Sartorello, F., Pontara, E., et al. (2008). A shorter von Willebrand factor survival in O blood group subjects explains how ABO determinants influence plasma von Willebrand factor. *Blood* 111, 3540–3545. doi:10.1182/blood-2007-11-122945

Gattinger, P., Borochova, K., Dorofeeva, Y., Henning, R., Kiss, R., Kratzer, B., et al. (2021). Antibodies in Serum of Convalescent Patients Following Mild COVID-19 Do Not Always Prevent Virus-receptor Binding. *Allergy* 76, 878–883. doi:10.1111/all.14523

Goel, R., Bloch, E. M., Pirenne, F., Al-Riyami, A. Z., Crowe, E., Dau, L., et al. (2021). ABO Blood Group and COVID-19: a Review on Behalf of the ISBT COVID-19 Working Group. *Vox Sang* 116, 849–861. doi:10.1111/vox.13076

Guillon, P., Clément, M., Sébille, V., Rivain, J.-G., Chou, C.-F., Ruvoën-Clouet, N., et al. (2008). Inhibition of the Interaction between the SARS-CoV Spike Protein and its Cellular Receptor by Anti-histo-blood Group Antibodies. *Glycobiology* 18, 1085–1093. doi:10.1093/glycob/cwn093

Hoffmann, D., Mereiter, S., Jin Oh, Y., Monteil, V., Elder, E., Zhu, R., et al. (2021). Identification of Lectin Receptors for Conserved SARS-CoV-2 Glycosylation Sites. *EMBO J.* 40, e108375. doi:10.15252/embj.2021108375

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* 181, 271–280. doi:10.1016/j.cell.2020.02.052

Klausberger, M., Duerkop, M., Haslacher, H., Wozniak-Knopp, G., Cserjan-Puschmann, M., Perkmann, T., et al. (2021). A Comprehensive Antigen Production and Characterisation Study for Easy-To-Implement, Specific and Quantitative SARS-CoV-2 Serotests. *EBioMedicine* 67, 103348. doi:10.1016/j.ebiom.2021.103348

Kriechbaum, R., Ziaee, E., Grünwald-Gruber, C., Buscaill, P., Hoorn, R. A. L., and Castilho, A. (2020). BGAL1 Depletion Boosts the Level of β-galactosylation of N- and O-glycans in N. Benthamiana. *Plant Biotechnol. J.* 18, 1537–1549. doi:10.1111/pbi.13316

Ku, Z., Xie, X., Hinton, P. R., Liu, X., Ye, X., Muruato, A. E., et al. (2021). Nasal Delivery of an IgM Offers Broad protection from SARS-CoV-2 Variants. *Nature* 595, 718–723. doi:10.1038/s41586-021-03673-2

Leaf, R. K., Al-Samkari, H., Brenner, S. K., Gupta, S., and Leaf, D. E. (2020). ABO Phenotype and Death in Critically Ill Patients with COVID-19. *Br. J. Haematol.* 190, e204–e208. doi:10.1111/bjh.16984

Letts, J. A., Rose, N. L., Fang, Y. R., Barry, C. H., Borisova, S. N., Seto, N. O. L., et al. (2006). Differential Recognition of the Type I and II H Antigen Acceptors by the Human ABO(H) Blood Group A and B Glycosyltransferases. *J. Biol. Chem.* 281, 3625–3632. doi:10.1074/jbc.m507620200

Li J, J., Wang, X., Chen, J., Cai, Y., Deng, A., and Yang, M. (2020). Association between ABO Blood Groups and Risk of SARS-CoV-2 Pneumonia. *Br. J. Haematol.* 190, 24–27. doi:10.1111/bjh.16797

Li Q, Q., Wu, J., Nie, J., Zhang, L., Hao, H., Liu, S., et al. (2020). The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. *Cell* 182, 1284–1294. doi:10.1016/j.cell.2020.07.012

Lindberg, L., Liu, J., Gaunitz, S., Nilsson, A., Johansson, T., Karlsson, N. G., et al. (2013). Mucin-type Fusion Proteins with Blood Group A or B Determinants on Defined O-Glycan Core Chains Produced in Glycoengineered Chinese Hamster Ovary Cells and Their Use as Immunoaffinity Matrices. *Glycobiology* 23, 720–735. doi:10.1093/glycob/cwt011

Margolin, E. A., Strasser, R., Chapman, R., Williamson, A.-L., Rybicki, E. P., and Meyers, A. E. (2020). Engineering the Plant Secretory Pathway for the Production of Next-Generation Pharmaceuticals. *Trends Biotechnol.* 38, 1034–1044. doi:10.1016/j.tibtech.2020.03.004

Miotto, M., Di Rienzo, L., Gosti, G., Milanetti, E., and Ruocco, G. (2021). Does Blood Type Affect the COVID-19 Infection Pattern? *PLoS One* 16, e0251535. doi:10.1371/journal.pone.0251535

Piccoli, L., Park, Y.-J., Tortorici, M. A., Czudnochowski, N., Walls, A. C., Beltramello, M., et al. (2020). Mapping Neutralizing and Immunodominant Sites on the SARS-CoV-2 Spike Receptor-Binding Domain by Structure-Guided High-Resolution Serology. *Cell* 183, 1024–1042. doi:10.1016/j.cell.2020.09.037

Pinto, D., Park, Y.-J., Beltramello, M., Walls, A. C., Tortorici, M. A., Bianchi, S., et al. (2020). Cross-neutralization of SARS-CoV-2 by a Human Monoclonal SARS-CoV Antibody. *Nature* 583, 290–295. doi:10.1038/s41586-020-2349-y

Sainsbury, F., Thuenemann, E. C., and Lomonossoff, G. P. (2009). pEAQ: Versatile Expression Vectors for Easy and Quick Transient Expression of Heterologous Proteins in Plants. *Plant Biotechnol. J.* 7, 682–693. doi:10.1111/j.1467-7652.2009.00434.x

Schwestka, J., König-Beihammer, J., Shin, Y.-J., Vavra, U., Kienzl, N. F., Grünwald-Gruber, C., et al. (2021). Impact of Specific N-Glycan Modifications on the Use of Plant-Produced SARS-CoV-2 Antigens in Serological Assays. *Front. Plant Sci.* 12, 747500. doi:10.3389/fpls.2021.747500

Shajahan, A., Archer-Hartmann, S., Supekar, N. T., Gleinich, A. S., Heiss, C., and Azadi, P. (2021). Comprehensive Characterization of N- and O- Glycosylation of SARS-CoV-2 Human Receptor Angiotensin Converting Enzyme 2. *Glycobiology* 31, 410–424. doi:10.1093/glycob/cwaa101

Shin, Y.-J., König-Beihammer, J., Vavra, U., Schwestka, J., Kienzl, N. F., Klausberger, M., et al. (2021). N-glycosylation of the SARS-CoV-2 Receptor Binding Domain Is Important for Functional Expression in Plants. *Front. Plant Sci.* 12, 689104. doi:10.3389/fpls.2021.689104

Strasser, R., Bondili, J. S., Vavra, U., Schoberer, J., Svoboda, B., Glössl, J., et al. (2007). A Unique β1,3-Galactosyltransferase Is Indispensable for the

Biosynthesis of N-Glycans Containing Lewis a Structures in *Arabidopsis thaliana*. *Plant Cell* 19, 2278–2292. doi:10.1105/tpc.107.052985

Strasser, R., Castilho, A., Stadlmann, J., Kunert, R., Quendler, H., Gattinger, P., et al. (2009). Improved Virus Neutralization by Plant-Produced Anti-HIV Antibodies with a Homogeneous β1,4-Galactosylated N-Glycan Profile. *J. Biol. Chem.* 284, 20479–20485. doi:10.1074/jbc.m109.014126

Strasser, R., Stadlmann, J., Schähs, M., Stiegler, G., Quendler, H., Mach, L., et al. (2008). Generation of Glyco-Engineered *Nicotiana Benthamiana* for the Production of Monoclonal Antibodies with a Homogeneous Human-like N-Glycan Structure. *Plant Biotechnol. J.* 6, 392–402. doi:10.1111/j.1467-7652.2008.00330.x

Strasser, R., Stadlmann, J., Svoboda, B., Altmann, F., Glössl, J., and Mach, L. (2005). Molecular Basis of N-Acetylglucosaminyltransferase I Deficiency in *Arabidopsis thaliana* Plants Lacking Complex N-Glycans. *Biochem. J.* 387, 385–391. doi:10.1042/bj20041686

Stussi, G., Huggel, K., Lutz, H. U., Schanz, U., Rieben, R., and Seebach, J. D. (2005). Isotype-specific Detection of ABO Blood Group Antibodies Using a Novel Flow Cytometric Method. *Br. J. Haematol.* 130, 954–963. doi:10.1111/j.1365-2141.2005.05705.x

Turoňová, B., Sikora, M., Schürmann, C., Hagen, W. J. H., Welsch, S., Blanc, F. E. C., et al. (2020). *In Situ* structural Analysis of SARS-CoV-2 Spike Reveals Flexibility Mediated by Three Hinges. *Science* 370, 203–208. doi:10.1126/science.abd5223

Walls, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., Mcguire, A. T., and Veesler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* 181, 281–292. doi:10.1016/j.cell.2020.02.058

Watanabe, Y., Allen, J. D., Wrapp, D., Mclellan, J. S., and Crispin, M. (2020a). Site-specific Glycan Analysis of the SARS-CoV-2 Spike. *Science* 369, 330–333. doi:10.1126/science.abb9983

Watanabe, Y., Berndsen, Z. T., Raghwani, J., Seabright, G. E., Allen, J. D., Pybus, O. G., et al. (2020b). Vulnerabilities in Coronavirus Glycan Shields Despite Extensive Glycosylation. *Nat. Commun.* 11, 2688. doi:10.1038/s41467-020-16567-0

Wu, S.-C., Arthur, C. M., Wang, J., Verkerke, H., Josephson, C. D., Kalman, D., et al. (2021). The SARS-CoV-2 Receptor-Binding Domain Preferentially Recognizes Blood Group A. *Blood Adv.* 5, 1305–1309. doi:10.1182/bloodadvances.2020003259

Wu, Y., Feng, Z., Li, P., and Yu, Q. (2020). Relationship between ABO Blood Group Distribution and Clinical Characteristics in Patients with COVID-19. *Clinica Chim. Acta* 509, 220–223. doi:10.1016/j.cca.2020.06.026

Yamamoto, F. (2004). Review: ABO Blood Group System--ABH Oligosaccharide Antigens, Anti-A and Anti-B, A and B Glycosyltransferases, and ABO Genes. *Immunohematology* 20, 3–22. doi:10.21307/immunohematology-2019-418

Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., and Zhou, Q. (2020). Structural Basis for the Recognition of SARS-CoV-2 by Full-Length Human ACE2. *Science* 367, 1444–1448. doi:10.1126/science.abb2762

Yuan, M., Wu, N. C., Zhu, X., Lee, C.-C. D., So, R. T. Y., Lv, H., et al. (2020). A Highly Conserved Cryptic Epitope in the Receptor Binding Domains of SARS-CoV-2 and SARS-CoV. *Science* 368, 630–633. doi:10.1126/science.abb7269

Zhao J, J., Yang, Y., Huang, H., Li, D., Gu, D., Lu, X., et al. (2020). Relationship between the ABO Blood Group and the Coronavirus Disease 2019 (COVID-19) Susceptibility. *Clin. Infect. Dis.* 73, 328–331. doi:10.1093/cid/ciaa1150

Zhao P, P., Praissman, J. L., Grant, O. C., Cai, Y., Xiao, T., Rosenbalm, K. E., et al. (2020). Virus-Receptor Interactions of Glycosylated SARS-CoV-2 Spike and Human ACE2 Receptor. *Cell Host & Microbe* 28, 586–601. doi:10.1016/j.chom.2020.08.004

Conflict of Interest: YG, JH, MH, KV, and AW were employed by AIT Austrian Institute of Technology GmbH and DM by Laboratory Dr. Kosak, Dr. Reckendorfer and Partners, Specialists in Pathology LLC. - GmbH, Department of Clinical Microbiology.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Chemistry

MINI REVIEW
published: 24 March 2022
doi: 10.3389/fchem.2022.858708

Genetics Behind the Glycosylation Patterns in the Biosynthesis of Dalbaheptides

Oleksandr Yushchuk[1,2], *Kseniia Zhukrovska*[2], *Francesca Berini*[1], *Victor Fedorenko*[2] *and Flavia Marinelli*[1]*

[1]Department of Biotechnology and Life Sciences, University of Insubria, Varese, Italy, [2]Department of Genetics and Biotechnology, Ivan Franko National University of Lviv, Lviv, Ukraine

Glycopeptide antibiotics are valuable natural metabolites endowed with different pharmacological properties, among them are dalbaheptides used to treat different infections caused by multidrug-resistant Gram-positive pathogens. Dalbaheptides are produced by soil-dwelling high G-C Gram-positive actinobacteria. Their biosynthetic pathways are encoded within large biosynthetic gene clusters. A non-ribosomally synthesized heptapeptide aglycone is the common scaffold for all dalbaheptides. Different enzymatic tailoring steps, including glycosylation, are further involved in decorating it. Glycosylation of dalbaheptides is a crucial step, conferring them specific biological activities. It is achieved by a plethora of glycosyltransferases, encoded within the corresponding biosynthetic gene clusters, able to install different sugar residues. These sugars might originate from the primary metabolism, or, alternatively, their biosynthesis might be encoded within the biosynthetic gene clusters. Already installed monosaccharides might be further enzymatically modified or work as substrates for additional glycosylation. In the current minireview, we cover recent updates concerning the genetics and enzymology behind the glycosylation of dalbaheptides, building a detailed and consecutive picture of this process and of its biological evolution. A thorough understanding of how glycosyltransferases function in dalbaheptide biosynthesis might open new ways to use them in chemo-enzymatic synthesis and/or in combinatorial biosynthesis for building novel glycosylated antibiotics.

Keywords: glycopeptide antibiotics, dalbaheptides, ramoplanin, teicoplanin, A40926, glycosyltransferase, biosynthetic gene cluster

OPEN ACCESS

Edited by:
Zhongping Tan,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

Reviewed by:
Max Julian Cryle,
Monash University, Australia

***Correspondence:**
Flavia Marinelli
flavia.marinelli@uninsubria.it

Specialty section:
This article was submitted to Chemical Biology, a section of the journal Frontiers in Chemistry

Received: *20 January 2022*
Accepted: *21 February 2022*
Published: *24 March 2022*

Citation:
Yushchuk O, Zhukrovska K, Berini F, Fedorenko V and Marinelli F (2022) Genetics Behind the Glycosylation Patterns in the Biosynthesis of Dalbaheptides. Front. Chem. 10:858708. doi: 10.3389/fchem.2022.858708

INTRODUCTION

Among different bacterial phyla, the mycelia-forming members of actinobacteria-broadly known as actinomycetes-remain the best antibiotic providers (Bérdy, 2005; Hutchings et al., 2019). The biosynthesis of antibiotics involves many enzymes, which are encoded by co-localized genes-biosynthetic gene clusters (BGCs) (Medema et al., 2015). BGCs undergo modular evolution, often exchanging operons and single genes coding for biosynthetic and modification enzymes (Medema et al., 2014). Genes for glycosyltransferases (GTs) are one such example, being found in different BGCs, with corresponding proteins having relaxed substrate specificity and consequently being able to modify different natural scaffolds (Salas and Méndez, 2007).

The astonishing variability of glycosylation patterns in one group of antibiotics led to their eponymous description as glycopeptide antibiotics (GPAs, Nicolaou et al., 1999). Natural GPAs amalgamate five types of related compounds, differing in chemical structures, where, paradoxically, only types I-IV-also known as dalbaheptides (Parenti and Cavalleri, 1989)-are glycosylated (Nicolaou et al., 1999). All dalbaheptides possess a non-ribosomal heptapeptide aglycone differing in amino acid (aa) composition, cross-linking, and decoration (Nicolaou et al., 1999); they inhibit the growth of Gram-positive bacteria by blocking cell wall maturation (Binda et al., 2014; Yushchuk et al., 2020a). Glycosylation and acylation (a step depending on glycosylation) of dalbaheptides contribute to their antimicrobial activities, favoring dimerization and membrane localization at the site of action (Gerhard et al., 1993; Mackay et al., 1994; Beauregard et al., 1995; Snyder et al., 1998). On the other hand, excessive glycosylation does not always bring pharmacological benefits: it seemed to induce platelet aggregation in patients treated with ristocetin (Howard and Firkin, 1971; Coller and Gralnick, 1977), which was consequently withdrawn from the clinical use (Gangarosa et al., 1958).

Dalbaheptides are clinically used as drugs of last resort against multidrug-resistant Gram-positive pathogens (Marcone et al., 2018). First-generation GPAs-vancomycin and teicoplanin (produced by different *Amycolatopsis* spp. and *Actinoplanes teichomyceticus* ATCC 31121, respectively)-have a long and reliable history of clinical application (Jovetic et al., 2010; Binda et al., 2014; Marcone et al., 2018). In turn, natural GPAs served as precursors for three second-generation semisynthetic and clinically used GPAs (Binda et al., 2014; Butler et al., 2014): dalbavancin derived from A40926 (produced by *Nonomuraea gerenzanensis* ATCC 39727) (Crotty et al., 2016), telavancin from chloroeremomycin (from *Kibdelosporangium aridum* A82846) (Klinker and Borgert, 2015), and oritavancin from vancomycin (Crotty et al., 2016).

Although the chemical variety of glycosyl groups decorating dalbaheptide aglycones is quite remarkable (Nicolaou et al., 1999), many aspects of the genetics behind their biosynthesis and incorporation remain obscure. In this minireview, we focus on those dalbaheptides whose BGC sequences are nowadays available, and for which some experimental evidence about their glycosylation steps is reported in the literature. The model BGCs are *cep*, *bal*, *tei*, *vcm*, and *dbv*, responsible for the production of chloroeremomycin (in *K. aridum* A82846) (van Wageningen et al., 1998), balhimycin (in *Amycolatopsis balhimycina* DSM 5908) (Shawky et al., 2007), teicoplanin (Li et al., 2004), vancomycin (in *Amycolatopsis orientalis* HCCB10007) (Xu et al., 2014), and A40926 (Sosio et al., 2003), respectively. BGC from *Amycolatopsis* sp. MJM2582 (Truman et al., 2014) represents the ristocetin biosynthetic pathway and was found also in other *Amycolatopsis* spp. (Spohn et al., 2014; Liu et al., 2021). Glycosylation-related genes from more recently described BGCs for UK-68,597 (*auk* from *Actinoplanes* sp. ATCC 53533) (Yim et al., 2014a), pekiskomycin (*pek* from *Streptomyces* sp. WAC1420) (Thaker et al., 2013), keratinimicin (ker from *Amycolatopsis keratiniphila* NRRL B-24117) (Xu et al., 2019), and A50926 (from *Nonomuraea coxensis* DSM 45129) (Yushchuk et al., 2021) are also reviewed. Overall, multiple recent findings on dalbaheptide glycosylation updated the overall picture and merit a proper review, outlining what is known and why it is still worthy of further investigations.

Delineating Steps in Dalbaheptide Glycosylation

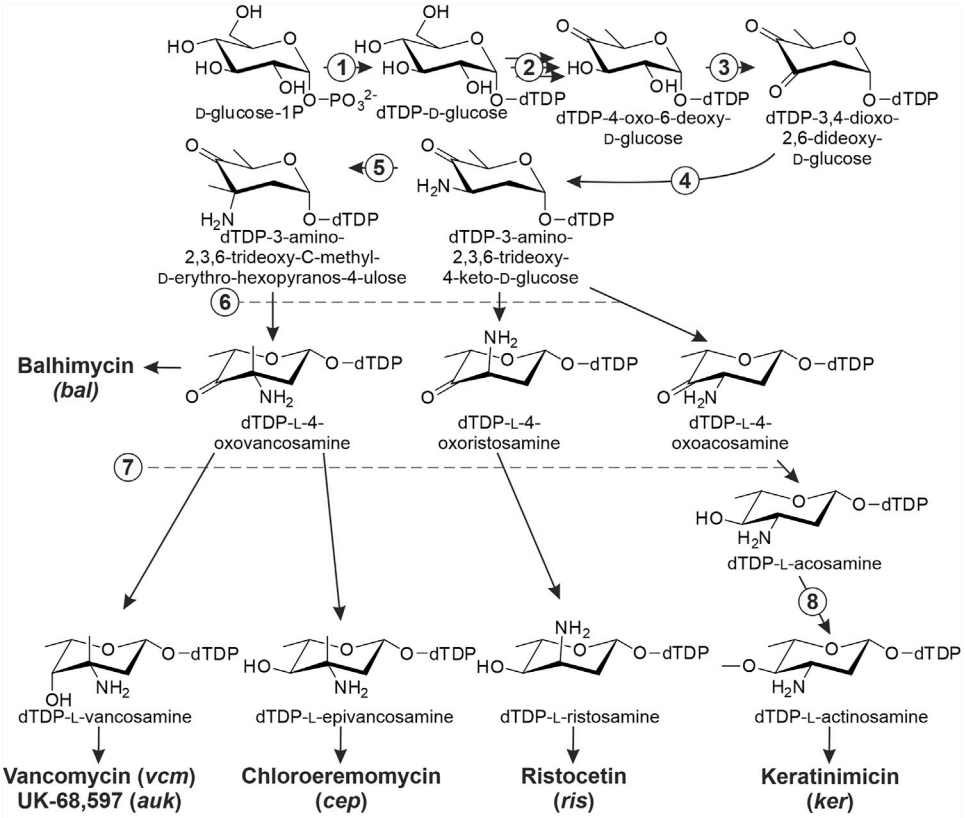
The biosynthesis of dalbaheptides is generally divided into three distinct stages (Yim et al., 2014b, 2016; Yushchuk et al., 2020b), that is, 1) generation of non-proteinogenic aa pool, further utilized in 2) non-ribosomal biosynthesis of the oligopeptide aglycones (coupled with the oxidative cross-linking); fully cross-linked aglycones are further 3) modified in a variety of tailoring steps. All dalbaheptide BGCs encode GTs, tailoring enzymes significantly contributing to the structural variety of these antibiotics (Nicolaou et al., 1999). Non-glycosylated dalbaheptide A47934 (from *Streptomyces toyocaensis* NRRL 15009) is the only exception here; consistently, the corresponding BGC lacks GT genes (Pootoolal et al., 2002).

More in detail, different steps might be defined in the glycosylation process of dalbaheptides, layer by layer "wrapping" the aglycone. The first step includes the biosynthesis of non-conventional sugar donors for aglycone decoration. Indeed, while some dalbaheptides are decorated with sugars deriving from primary metabolism (e.g., *α*-D-mannose and *N*-acetylglucosamine (Glc*N*Ac) in teicoplanin or A40926), aglycones of vancomycin, balhimycin, and chloroeremomycin are decorated with the non-conventional monosaccharides L-vancosamine, L-4-oxovancosamine, and L-epivancosamine, respectively. In a similar manner to the biosynthesis of non-proteinogenic aa, enzymes required for the biosynthesis of such non-conventional monosaccharides are encoded within dalbaheptide BGCs. The second step (often the last one) consists of *O*-glycosylation of the aromatic aa forming the aglycone. In the third one, the installed sugars might be further modified in minor or major ways (e.g., *α*-D-mannose *O*-acetylation and Glc*N*Ac deacetylation in A40926 biosynthesis).

Biosynthesis of Non-Conventional Monosaccharides was Required for the Glycosylation of Dalbaheptides

Conventional sugars in GPAs-from primary metabolism-are D-mannose, D-glucose, D-arabinose, Glc*N*Ac, and L-rhamnose. Non-conventional sugar residues include L-vancosamine, L-epivancosamine, L-4-oxovancosamine, L-ristosamine, and L-actinosamine. In addition to the aforementioned examples, L-vancosamine is present in Substitute with UK-68,597, while L-ristosamine is characteristic for ristocetin, and L-actinosamine for keratinimicin (Xu et al., 2019).

Biosynthesis of L-epivancosamine was initially studied in chloroeremomycin producer (van Wageningen et al., 1998). *In vitro* experiments (Chen et al., 2000; Kirkpatrick et al., 2000) demonstrated how five enzymes encoded within *cep*-namely, EvaA-E-transformed



		vancomycin *vcm*	chloroeremomycin *cep*	balhimycin *bal*	ristocetin *ris*	keratinimicin *ker*	UK-68,597 *auk*
1	glucose-1-phosphate thymidylylransferase	(*abs*)	(*abs*)	(*abs*)	(*abs*)	(*abs*)	Auk7
2	non-BGC-encoded enzymes	(*abs*)	(*abs*)	(*abs*)	(*abs*)	(*abs*)	(*abs*)
3	dTDP-4-dehydro-6-deoxy-α-D-glucopyranose 2,3-dehydratase	VcaA(*is*)	EvaA	DvaA(*is*)	Orf30(*is*)	KraA(*is*)	Auk13(*is*)
4	dTDP-3-amino-2,3,6-trideoxy-4-keto-D-glucose transaminase	VcaB(*is*)	EvaB	DvaB(*is*)	Orf32(*is*)	KraC(*is*)	Auk12(*is*)
5	SAM-dependent C3-methyltransferase	VcaC(*is*)	EvaC	DvaC(*is*)	(*abs*)	(*abs*)	Auk9(*is*)
6	C5-epimerase	VcaD(*is*)	EvaD	DvaD(*is*)	Orf33(*is*)	KraD(*is*)	Auk8(*is*)
7	C4-ketoreductase	VcaE(*is*)	EvaE	DvaE*	Orf31(*is*)	KraB(*is*)	(*abs*)
8	O-methyltransferase	(*abs*)	(*abs*)	(*abs*)	(*abs*)	KerM(?)	(*abs*)

FIGURE 1 | Enzymes involved in the biosynthesis of non-conventional aminosugars, decorating aglycones of some dalbaheptides. Biosynthetic pathway of L-epivancosamine serves as a model, since EvaA-B-C-D-E, coded within *cep*, are the only enzymes that were studied experimentally. Functions of all other enzymes were assigned by *in silico* comparison (*is*); (*abs*) indicates that the corresponding gene is absent from BGCs; (?) indicates that the assigned function is speculative, having no experimentally investigated prototype. Asterisk at DvaE indicates that this protein is mutated. Refer to the main text for more details.

dTDP-4-oxo-6-deoxy-D-glucose into L-epivancosamine (**Figure 1**). Biosynthetic routes for aminosugars decorating other dalbaheptides were deduced from this model pathway. Initially activated substrate dTDP-4-oxo-6-deoxy-D-glucose is commonly derived from D-glucose-1-phosphate by the action of non-BGC-encoded enzymes (**Figure 1**). One notable exception is the UK-68,597 biosynthesis, since *auk* contains a gene for glucose 1-phosphate thymidylylransferase-*auk7* (Yim et al., 2014a)-required for the D-glucose-1-phosphate activation. The presence of Auk7 likely positively contributes to the pool of L-vancosamine precursors in UK-68,597 biosynthesis.

vcm for vancomycin contains the orthologs of *evaA-E* genes, namely, *vcaA-E* (Xu et al., 2014) (**Figure 1**). The two sets of proteins are quite similar, sharing at least 75% of aa sequence identity (aa s.i.) in each pair (Xu et al., 2014). Nevertheless, minor differences in EvaE and VcaE seem to impact the function, making the first to convert dTDP-L-4-

oxovancosamine into dTDP-L-epivancosamine, while the second yields dTDP-L-vancosamine (**Figure 1**). Instead, *auk* contains the orthologs of only *evaA*, *evaB*, *evaC*, and *evaD* (*auk13*, *auk12*, *auk9*, and *auk8*, respectively) (Yim et al., 2014a), lacking the ortholog of *evaE*: probably a functional homolog (or analog) of *evaE* resides outside the *auk* borders, finally contributing to L-vancosamine production. More understandable is the case of *bal* (Shawky et al., 2007), where *evaA-E* orthologs are named *dvaA-E*. Here, *evaE* ortholog *dvaE* is truncated, coding only the 99 aa C-terminal part of C4-ketoreductase (Donadio et al., 2005). Such truncated protein is non-functional, and consequently, the aminosugar biosynthesis terminates at the stage of dTDP-L-4-oxovancosamine (**Figure 1**). Biosynthetic pathways for L-ristosamine and L-actinosamine, coded within *ris* and *ker*, are more diverged. They both missed the EvaC ortholog (Spohn et al., 2014; Truman et al., 2014; Xu et al., 2019; Liu et al., 2021), resulting in the lack of a methyl group at the C3 position (**Figure 1**). Thus, biosynthesis of L-ristosamine might be attributed to the orthologs of EvaA, EvaB, EvaD, and EvaE-Orf30, Orf32, Orf33, and Orf31, respectively (Truman et al., 2014). The same protein set is encoded within *ker*-KraA-D (Xu et al., 2019). The major difference between L-actinosamine and L-ristosamine is *O*-methylation of the C4-position. It is still unknown how this methylation is achieved; however, the annotation of *ker* reveals the presence of a gene coding for an *O*-methyltransferase-*kerM* (Xu et al., 2019). Since the aglycone of keratinimicin lacks any *O*-methylations, it seems plausible that KerM catalyzes the ultimate step of L-actinosamine biosynthesis (**Figure 1**).

In all cases described above the orthologs of *evaA-E-B-D* are most likely co-expressed, forming one operon, while the orthologs of *evaC* belong to a separate transcriptional unit (Shawky et al., 2007; Liu et al., 2021). *auk* is an exception, with *auk8-9* (*evaD-C*) and *auk12-13* (*evaA-B*) probably belonging to different operons (Yim et al., 2014a).

GTs Involved in Dalbaheptide Glycosylation

All GTs decorating aglycones of dalbaheptides belong to two families, according to the Carbohydrate-Active enZYmes Database (CAZy, http://www.cazy.org, Drula et al., 2022). GTs responsible for the installation of non-conventional aminosugars and conventional D-glucose, D-arabinose, Glc*N*Ac, and L-rhamnose, belong to the GT1 family. These GTs require sugar substrates to be either dTDP- or UDP-activated, and share a unique two-domain structure (the so-called GT-B fold, Lairson et al., 2008), having C- and N-terminal Rossmann-like domains connected by a flexible linker region (Zhang et al., 2020). Recognition sites for the donor NDP-activated sugars are located at C-terminal domains (Chang et al., 2011), while N-terminal domains contain the acceptor binding site for dalbaheptide aglycone (Chang et al., 2011; Zhang et al., 2020). GTs of the second-GT39-family are responsible for the installation of D-mannose and require D-mannosyl-1-phosphoundecaprenol as a donor substrate. Large hydrophobic GT39-GTs are predicted as membrane-associated, having a GT-C fold with multiple transmembrane helices and intracellular active sites (Lairson et al., 2008).

In the aforementioned dalbaheptides, GT1-GTs attach sugar residues preferentially at AA-4 (4-hydroxyphenylglycine, Hpg) and AA-6 (*β*-hydroxytyrosine, Bht) of the aglycone or add additional monosaccharides to already existing mono/di/trisaccharides at AA-4. Fully cross-linked aglycones serve as acceptor substrates for GT1-GTs under physiological conditions, albeit some GT1-GTs were able to recognize partially cross-linked aglycones under certain experimental conditions, for instance, in mutasynthesis approaches (Weist et al., 2004; Butz et al., 2008). GT39-GTs attach mannose at AA-7 (3,5-dihydroxyphenylglycine, Dpg). The presence of multiple GTs within one pathway might result in the production of mixtures of related congeners, differing in glycosylation patterns. Dalbaheptides glycosylated at other AA positions were also described (Nicolaou et al., 1999), indicating that glycosylation might also occur at 1) AA-2 (Bht) and AA-1 (Hpg) in type II aglycone, and 2) AA-1 (Hpg) and AA-3 (Bht) in type III aglycone. Unfortunately, we currently lack any genomic information on the producers of these molecules, which would merit further investigations.

Type I dalbaheptides chloroeremomycin, balhimycin, and vancomycin served as the first models for experimental investigation of GT functions. Corresponding BGCs encode slightly different sets of GT1-GTs: GtfA, GtfB, and GtfC in *cem*; BgtfA, BgtfB, and BgtfC in *bal*; and GtfD and GtfE in *vcm*. Orthologous GTs GtfA and BgtfA install L-epivancosamine and L-4-oxovancosamine at AA-6 of chloroeremomycin and balhimycin, respectively. *vcm* does not encode GtfA ortholog, explaining why vancomycin is not glycosylated at AA-6. Peculiarly, this particular difference between vancomycin and chloroeremomycin seems to augment the antimicrobial activity of the latter, implying that L-epivancosamine at AA-6 facilitates cell wall binding (Nagarajan, 1993; Allen et al., 2002). GtfB, BgtfB, and GtfE are orthologs, glucosylating AA-4 in the biosynthesis of all three antibiotics (**Figure 2A**, Pelzer et al., 1999; Losey et al., 2001; Mulichak et al., 2001). Then, GtfC and GtfD attach L-epivancosamine or L-vancosamine to D-glucose at AA-4 in the biosynthesis of chloroeremomycin and vancomycin, respectively (**Figure 2A**, Losey et al., 2001, 2002; Mulichak et al., 2004). GtfC, GtfD, and BgtfC are orthologous proteins, but balhimycin lacks a disaccharide at AA-4, which could be found only in balhimycin V (a congener produced in residual amounts) (Pelzer et al., 1999; Stegmann et al., 2010). This might be due to the low affinity of BgtfC for the donor substrate-L-4-oxovancosamine, produced as a consequence of DvaE mutation. More recently described type I dalbaheptide pekiskomycin is only glucosylated at AA-4, coherently with the only 1 GT encoded in *pek*: Pek28, which is a GtfB ortholog (**Figure 2A**) (Thaker et al., 2013).

So far, little is known about GTs decorating aglycones of type II dalbaheptides. *ker* (single type II BGC sequenced) carries three genes for GT1-GTs and one gene for GT39-GT (Xu et al., 2019). Among *ker*-encoded GT1-GTs, $gtfA_{ker}$, $gtfB_{ker}$, and $gtfC_{ker}$ (*ker* was added to distinguish them from *cep* genes) are orthologs of *gtfA*, *gtfB*, and *gtfC*, respectively (Xu et al., 2019). Thus, $GtfB_{ker}$

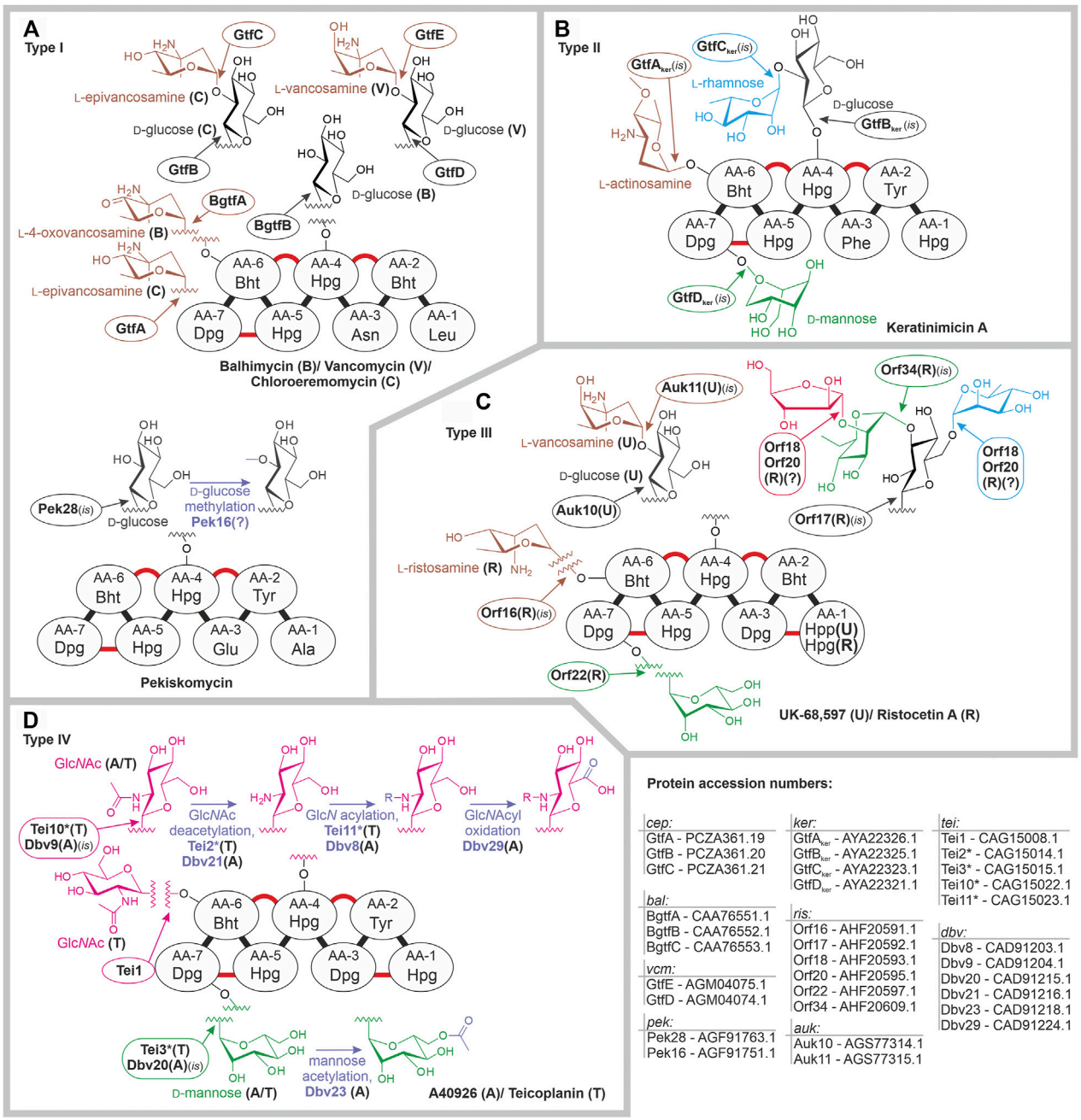


FIGURE 2 | Glycosylation patterns of type I **(A)**, type II **(B)**, type III **(C)**, and type IV **(D)** dalbaheptides and enzymes involved therein. Refer to the main text for more details. Dalbaheptide aglycones are depicted schematically with cross-links shown in red, and chlorination and sulfation sites are not shown. Aglycone amino acid abbreviations mean following: Leu-leucine; Asn-asparagine; Ala-alanine; Glu-glutamine; Phe-phenylalanine; Tyr-tyrosine; Bht-*β*-hydroxytyrosine; Dpg-3,5-dihydroxyphenylglycine; Hpg-4-hydroxyphenylglycine; Hpp-4-hydroxyphenypyruvate; Tyr-. Glc*N* stated for *N*-glucosamine, GlcNAcyl for N-acylglucosamine. All enzymes, whose functions were assigned by *in silico* comparison are marked with (*is*); (?) indicates that function was assigned *in silico* without experimentally investigated prototype or exact function remains unknown. For fast access to protein sequences mentioned in this figure, use following links: *cep*: GtfA-PCZA361.19, GtfB-PCZA361.20, GtfC-PCZA361.21; *bal*: BgtfA-CAA76551.1, BgtfB-CAA76552.1, BgtfC-CAA76553.1; *vcm*: GtfE-AGM04075.1, GtfD-AGM04074.1; *pek*: Pek28-AGF91763.1, Pek16-AGF91751.1; *ker*: $GtfA_{ker}$-AYA22326.1, $GtfB_{ker}$-AYA22325.1, $GtfC_{ker}$-AYA22323.1, $GtfD_{ker}$-AYA22321.1; *ris*: Orf16-AHF20591.1, Orf17-AHF20592.1, Orf18-AHF20593.1, Orf20-AHF20595.1; Orf22-AHF20597.1, Orf34-AHF20609.1; *auk*: Auk10-AGS77314.1, Auk11-AGS77315.1; *tei*: Tei1-CAG15008.1, Tei2*-CAG15014.1, Tei3*-CAG15015.1, Tei10*-CAG15022.1, Tei11*-CAG15023.1; *dbv*: Dbv8-CAD91203.1, Dbv9-CAD91204.1, Dbv20-CAD91215.1, Dbv21-CAD91216.1, Dbv23-CAD91218.1, Dbv29-CAD91224.1.

most likely installs D-glucose at AA-4, and L-rhamnose is then appended to D-glucose by $GtfC_{ker}$; this leaves $GtfA_{ker}$ responsible for the attachment of L-actinosamine at AA-6 (**Figure 2B**). Single *ker*-encoded GT39-GT $GtfD_{ker}$ most probably attaches D-mannose at AA-7 (**Figure 2B**).

More information is available on the GTs involved in the biosynthesis of type III dalbaheptides. Ristocetin BGC encodes 6 GTs, four of GT1 family and two belonging to the GT39 family (Truman et al., 2014). Phylogenetic reconstruction allowed to assign functions to 4 GTs, assuming that Orf16 attaches L-ristosamine at AA-6, Orf17-D-glucose at AA-4, then appended with D-mannose by Orf34; finally, a second mannosyltransferase-Orf22-was expected to act at AA-7 (**Figure 2C**, Truman et al., 2014), as later confirmed by its heterologous expression in *Streptomyces coelicolor* carrying *sta* (Yim et al., 2016). Functions of Orf18 and Orf20 were not assigned, since these proteins were distantly related to known GTs (Truman et al., 2014). Another type III dalbaheptide-UK-68,597-is decorated with 2-L-vancosaminyl-D-glucose disaccharide at AA-4 (Yim et al., 2014a), while *auk* carries three genes for GT1-GTs. *In vitro* assay showed that Auk10 (GtfB ortholog) is responsible for glucosylation; *in silico* analysis then suggested L-vancosamine to be installed by Auk11 (GtfC ortholog, **Figure 2C**) (Yim et al., 2014a). Although the third GT, Auk14, is a GtfA ortholog, UK-68,597 lacks sugar residues at AA-6. *In vitro* assay suggested that Auk14 might be inactive or possess a very low affinity to substrates available in UK-68,597 biosynthesis, as observed with BgtfA (Yim et al., 2014a).

Glycosylation events that take place in type IV dalbaheptide biosynthesis are well defined (**Figure 2D**). *tei* and *dbv* BGCs encode three and two GTs, respectively. Accordingly, both antibiotics carry a Glc*N*Ac moiety at AA-4 and D-mannose at AA-7, while teicoplanin aglycone is also decorated with another Glc*N*Ac at AA-6. GtfB ortholog-Tei10*-installs the Glc*N*Ac moiety at AA-4, as demonstrated from multiple *in vivo* and *in vitro* experiments (Li et al., 2004; Howard-Jones et al., 2007; Truman et al., 2009; Yushchuk et al., 2016, 2020b). Consistently, Dbv9-a Tei10* ortholog-is supposed to play the same role in A40926 glycosylation. Tei1 was shown to attach Glc*N*Ac to teicoplanin aglycone at AA-6 (Li et al., 2004), whereas *dbv* lacks any Tei1 ortholog, explaining why A40926 has no sugars at AA-6. Finally, Tei3* was shown to be responsible for the decoration of teicoplanin aglycone with D-mannose at AA-7 (Yushchuk et al., 2016), implying that Dbv20 (Tei3* ortholog) has the same function in A40926 biosynthesis (**Figure 2D**). Interestingly, ramoplanin BGC (*ramo*), recently shown to be genetically related to *tei* (Waglechner et al., 2019), encodes a homolog of Tei3*-Ramo29 (45% aa s.i., Chen et al., 2013). Ramoplanin is a clinically relevant peptide antibiotic produced by *Actinoplanes ramoplaninifer* ATCC 33076 (Marcone et al., 2017). Unlike dalbaheptides, it carries a 4-D-mannosyl-D-mannose disaccharide, instead of a single D-mannose residue. Ramo29 was shown to install the first D-mannose residue (Chen et al., 2013), but the GT responsible for the second mannosylation remains unknown. This merits further investigation, since such GT looks like a promising tool to further modify mannosylated dalbaheptides.

Concluding this section, it is interesting to report that genes for GTs tend to form one operon in type I-III BGCs (Shawky et al., 2007; Liu et al., 2021), being more scattered in type IV BGCs (Alduina et al., 2007; Yushchuk et al., 2019). It is also notable that genes for mannosyltransferases are present in different types of BGCs coming from distant actinobacterial lineages. This might indicate D-mannose residues at AA-7 to be an ancestral feature for all dalbaheptides.

Further Modification Occurring on Attached Sugar Residues

Some further modifications of attached sugars occur during the biosynthesis of dalbaheptides, although they are quite rare. The first example comes from pekiskomycin, having D-glucose methylated. *pek* encodes two methyltransferases. One of them-Pek30-was experimentally shown to methylate the N-terminus of A47934 aglycone (Yim et al., 2016), leaving the other-Pek16-possibly responsible for D-glucose methylation (**Figure 2A**). Another notable modification is the acylation of AA-4 Glc*N*Ac in type IV dalbaheptides, such as teicoplanin and A40926. To achieve this modification, Glc*N*Ac at AA-4 is first deacetylated with orthologous deacetylases Tei2*/Dbv21 (Ho et al., 2006; Truman et al., 2006), and *N*-glucosamine is then acylated with orthologous acyltransferases Tei11*/Dbv8 (**Figure 2D**) (Li et al., 2004; Kruger et al., 2005; Howard-Jones et al., 2007; Yushchuk et al., 2016). Peculiarly, orthologs of Tei2*/Dbv21 are present in many (if not in all) BGCs for non-acylated type I-III dalbaheptides. The one from *cep*-CepI-was studied *in vitro* and shown to be inactive due to a single aa substitution (Truman et al., 2008). The omnipresence of *tei2** orthologs in type I-III BGCs induces to speculate that the *N*-acylglucosamine moiety at AA-4 is an ancestral feature, lost or modified in many evolutionary lineages of dalbaheptides.

Modifications of A40926 sugars do not end with acylation. *N*-acylglucosamine moiety is further oxidized to *N*-acylaminoglucuronic acid group by Dbv29 (**Figure 2D**) (Li et al., 2007). The biological role of such modification is unclear, although it seems to reduce the A40926 antimicrobial activity (Malabarba et al., 1995). Peculiarly, *noc* BGC in *N. coxensis* DSM 45129 lacks an ortholog for *dbv29*, coding the biosynthesis of non-oxidized A40926 analog-dalbaheptide A50926 (Yushchuk et al., 2021). Finally, the D-mannose residue at AA-7 of A40926 is *O*-acetylated with Dbv23 (**Figure 2D**) (Sosio et al., 2010). This modification is unstable and fades away in the alkaline extraction of A40926 (Alt et al., 2019); once again, its biological role is unclear, although it might be important for the regulation of antibiotic export and self-resistance (Alduina et al., 2020).

CONCLUSION AND OUTLOOK

Further research of dalbaheptide glycosylation is important for several reasons. Understanding of GTs donor- (activated sugar) and acceptor- (aglycone) substrate specificities will allow further chemical derivatization of these scaffolds

using *in vitro* chemo-enzymatic synthesis (Nakayama et al., 2014) or *in vivo* combinatorial biosynthesis (Yim et al., 2016). While the first approach has been widely investigated in the past for generating novel hybrid GPAs by combining natural and synthetic aglycones and sugars (as reviewed in Marcone et al., 2018), *in vivo* combinatorial biosynthesis is promising, but still rather limited in its applications. Alternatively, the two-domain architecture of GT1-GTs might be exploited to create "chimaeras" with an expanded functional repertoire (Truman et al., 2009). Finally, a better comprehension of glycosylation mechanisms will contribute to tracing out a more complete picture of dalbaheptide evolution. Unlike other aspects of glycopeptide biosynthesis (Donadio et al., 2005; Waglechner et al., 2019; Andreo-Vidal et al., 2021), phylogenomics of GTs and sugar modification enzymes has not been studied yet. We believe that such reconstruction might open new scenarios on the evolution of antibiotic biosynthetic pathways.

AUTHOR CONTRIBUTIONS

KZ, FB, and OY collected data and articles and co-wrote the review; OY prepared the figures; and VF and FM supervised the work and co-wrote the review.

FUNDING

This work was supported by grant "Fondo di Ateneo per la Ricerca" to FM and FB and by the BG-09F grant of the Ministry of Education and Science of Ukraine to VF.

ACKNOWLEDGMENTS

We are grateful to Consorzio Interuniversitario per le Biotecnologie for supporting FB congresses attendances.

REFERENCES

Alduina, R., Piccolo, L. L., D'Alia, D., Ferraro, C., Gunnarsson, N., Donadio, S., et al. (2007). Phosphate-controlled Regulator for the Biosynthesis of the Dalbavancin Precursor A40926. *J. Bacteriol.* 189, 8120–8129. doi:10.1128/JB.01247-07

Alduina, R., Tocchetti, A., Costa, S., Ferraro, C., Cancemi, P., Sosio, M., et al. (2020). A Two-Component Regulatory System with Opposite Effects on Glycopeptide Antibiotic Biosynthesis and Resistance. *Sci. Rep.* 10, 1–12. doi:10.1038/s41598-020-63257-4

Allen, N. E., LeTourneau, D. L., Hobbs, J. N., and Thompson, R. C. (2002). Hexapeptide Derivatives of Glycopeptide Antibiotics: Tools for Mechanism of Action Studies. *Antimicrob. Agents Chemother.* 46, 2344–2348. doi:10.1128/AAC.46.8.2344-2348.2002

Alt, S., Bernasconi, A., Sosio, M., Brunati, C., Donadio, S., and Maffioli, S. I. (2019). Toward Single-Peak Dalbavancin Analogs through Biology and Chemistry. *ACS Chem. Biol.* 14, 356–360. doi:10.1021/acschembio.9b00050

Andreo-Vidal, A., Binda, E., Fedorenko, V., Marinelli, F., and Yushchuk, O. (2021). Genomic Insights into the Distribution and Phylogeny of Glycopeptide Resistance Determinants within the *Actinobacteria* Phylum. *Antibiotics* 10, 1533–1630. doi:10.3390/antibiotics10121533

Beauregard, D. A., Williams, D. H., Gwynn, M. N., and Knowles, D. J. (1995). Dimerization and Membrane Anchors in Extracellular Targeting of Vancomycin Group Antibiotics. *Antimicrob. Agents Chemother.* 39, 781–785. doi:10.1128/AAC.39.3.781

Bérdy, J. (2005). Bioactive Microbial Metabolites. *J. Antibiot.* 58, 1–26. doi:10.1038/ja.2005.1

Binda, E., Marinelli, F., and Marcone, G. (2014). Old and New Glycopeptide Antibiotics: Action and Resistance. *Antibiotics* 3, 572–594. doi:10.3390/antibiotics3040572

Butler, M. S., Hansford, K. A., Blaskovich, M. A. T., Halai, R., and Cooper, M. A. (2014). Glycopeptide Antibiotics: Back to the Future. *J. Antibiot.* 67, 631–644. doi:10.1038/ja.2014.111

Butz, D., Schmiederer, T., Hadatsch, B., Wohlleben, W., Weber, T., and Süssmuth, R. D. (2008). Module Extension of a Non-ribosomal Peptide Synthetase of the Glycopeptide Antibiotic Balhimycin Produced by *Amycolatopsis balhimycina*. *ChemBioChem* 9, 1195–1200. doi:10.1002/cbic.200800068

Chang, A., Singh, S., Phillips, G. N., and Thorson, J. S. (2011). Glycosyltransferase Structural Biology and its Role in the Design of Catalysts for Glycosylation. *Curr. Opin. Biotechnol.* 22, 800–808. doi:10.1016/j.copbio.2011.04.013

Chen, H., Thomas, M. G., Hubbard, B. K., Losey, H. C., Walsh, C. T., and Burkart, M. D. (2000). Deoxysugars in Glycopeptide Antibiotics: Enzymatic Synthesis of TDP-L-Epivancosamine in Chloroeremomycin Biosynthesis. *Proc. Natl. Acad. Sci.* 97, 11942–11947. doi:10.1073/pnas.210395097

Chen, J.-S., Wang, Y.-X., Shao, L., Pan, H.-X., Li, J.-A., Lin, H.-M., et al. (2013). Functional Identification of the Gene Encoding the Enzyme Involved in Mannosylation in Ramoplanin Biosynthesis in *Actinoplanes* Sp. *Biotechnol. Lett.* 35, 1501–1508. doi:10.1007/s10529-013-1233-3

Coller, B. S., and Gralnick, H. R. (1977). Studies on the Mechanism of Ristocetin-Induced Platelet Agglutination. *J. Clin. Invest.* 60, 302–312. doi:10.1172/JCI108778

Crotty, M. P., Krekel, T., Burnham, C.-A. D., and Ritchie, D. J. (2016). New Gram-Positive Agents: the Next Generation of Oxazolidinones and Lipoglycopeptides. *J. Clin. Microbiol.* 54, 2225–2232. doi:10.1128/JCM.03395-15

Donadio, S., Sosio, M., Stegmann, E., Weber, T., and Wohlleben, W. (2005). Comparative Analysis and Insights into the Evolution of Gene Clusters for Glycopeptide Antibiotic Biosynthesis. *Mol. Genet. Genomics* 274, 40–50. doi:10.1007/s00438-005-1156-3

Drula, E., Garron, M.-L., Dogan, S., Lombard, V., Henrissat, B., and Terrapon, N. (2022). The Carbohydrate-Active Enzyme Database: Functions and Literature. *Nucleic Acids Res.* 50, D571–D577. doi:10.1093/nar/gkab1045

Gangarosa, E. J., Landerman, N. S., Rosch, P. J., and Herndon, E. G. (1958). Hematologic Complications Arising during Ristocetin Therapy. *N. Engl. J. Med.* 259, 156–161. doi:10.1056/nejm195807242590402

Gerhard, U., Mackay, J. P., Maplestone, R. A., and Williams, D. H. (1993). The Role of the Sugar and Chlorine Substituents in the Dimerization of Vancomycin Antibiotics. *J. Am. Chem. Soc.* 115, 232–237. doi:10.1021/ja00054a033

Ho, J.-Y., Huang, Y.-T., Wu, C.-J., Li, Y.-S., Tsai, M.-D., and Li, T.-L. (2006). Glycopeptide Biosynthesis: Dbv21/Orf2* from *dbv/tcp* Gene Clusters Are *N*-Ac-Glm Teicoplanin Pseudoaglycone Deacetylases and Orf15 from *cep* Gene Cluster Is a Glc-1-P Thymidyltransferase. *J. Am. Chem. Soc.* 128, 13694–13695. doi:10.1021/ja0644834

Howard, M. A., and Firkin, B. G. (1971). Ristocetin - A New Tool in the Investigation of Platelet Aggregation. *Thromb. Haemost.* 26, 362–369. doi:10.1055/s-0038-1653684

Howard-Jones, A. R., Kruger, R. G., Lu, W., Tao, J., Leimkuhler, C., Kahne, D., et al. (2007). Kinetic Analysis of Teicoplanin Glycosyltransferases and Acyltransferase Reveal Ordered Tailoring of Aglycone Scaffold to Reconstitute Mature Teicoplanin. *J. Am. Chem. Soc.* 129, 10082–10083. doi:10.1021/ja0735857

Hutchings, M. I., Truman, A. W., and Wilkinson, B. (2019). Antibiotics: Past, Present and Future. *Curr. Opin. Microbiol.* 51, 72–80. doi:10.1016/j.mib.2019.10.008

Jovetic, S., Zhu, Y., Marcone, G. L., Marinelli, F., and Tramper, J. (2010). β-Lactam and Glycopeptide Antibiotics: First and Last Line of Defense? *Trends Biotechnol.* 28, 596–604. doi:10.1016/j.tibtech.2010.09.004

Kirkpatrick, P. N., Scaife, W., Spencer, J. B., Williams, D. H., Hallis, T. M., and Liu, H.-w. (2000). Characterisation of a Sugar Epimerase Enzyme Involved in the

Biosynthesis of a Vancomycin-Group Antibiotic. *Chem. Commun.* 17, 1565–1566. doi:10.1039/b004463f

Klinker, K. P., and Borgert, S. J. (2015). Beyond Vancomycin: the Tail of the Lipoglycopeptides. *Clin. Ther.* 37, 2619–2636. doi:10.1016/j.clinthera.2015.11.007

Kruger, R. G., Lu, W., Oberthür, M., Tao, J., Kahne, D., and Walsh, C. T. (2005). Tailoring of Glycopeptide Scaffolds by the Acyltransferases from the Teicoplanin and A-40,926 Biosynthetic Operons. *Chem. Biol.* 12, 131–140. doi:10.1016/j.chembiol.2004.12.005

Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. (2008). Glycosyltransferases: Structures, Functions, and Mechanisms. *Annu. Rev. Biochem.* 77, 521–555. doi:10.1146/annurev.biochem.76.061005.092322

Li, T., Huang, F., Haydock, S., Mironenko, T., Leadlay, P., and Spencer, J. (2004). Biosynthetic Gene Cluster of the Glycopeptide Antibiotic Teicoplanin Characterization of Two Glycosyltransferases and the Key Acyltransferase. *Chem. Biol.* 11, 107–119. doi:10.1016/S1074-5521(04)00002-X

Li, Y.-S., Ho, J.-Y., Huang, C.-C., Lyu, S.-Y., Lee, C.-Y., Huang, Y.-T., et al. (2007). A Unique Flavin Mononucleotide-Linked Primary Alcohol Oxidase for Glycopeptide A40926 Maturation. *J. Am. Chem. Soc.* 129, 13384–13385. doi:10.1021/ja075748x

Liu, K., Hu, X.-R., Zhao, L.-X., Wang, Y., Deng, Z., and Tao, M. (2021). Enhancing Ristomycin A Production by Overexpression of ParB-like StrR Family Regulators Controlling the Biosynthesis Genes. *Appl. Environ. Microbiol.* 87, 1–19. doi:10.1128/AEM.01066-21

Losey, H. C., Jiang, J., Biggins, J. B., Oberthür, M., Ye, X.-Y., Dong, S. D., et al. (2002). Incorporation of Glucose Analogs by GtfE and GtfD from the Vancomycin Biosynthetic Pathway to Generate Variant Glycopeptides. *Chem. Biol.* 9, 1305–1314. doi:10.1016/S1074-5521(02)00270-3

Losey, H. C., Peczuh, M. W., Chen, Z., Eggert, U. S., Dong, S. D., Pelczer, I., et al. (2001). Tandem Action of Glycosyltransferases in the Maturation of Vancomycin and Teicoplanin Aglycones: Novel Glycopeptides,. *Biochemistry* 40, 4745–4755. doi:10.1021/bi010050w

Mackay, J. P., Gerhard, U., Beauregard, D. A., Maplestone, R. A., and Williams, D. H. (1994). Dissection of the Contributions toward Dimerization of Glycopeptide Antibiotics. *J. Am. Chem. Soc.* 116, 4573–4580. doi:10.1021/ja00090a005

Malabarba, A., Ciabatti, R., Scotti, R., Goldstein, B. P., Ferrari, P., Kurz, M., et al. (1995). New Semisynthetic Glycopeptides MDL 63,246 and MDL 63,042, and Other Amide Derivatives of Antibiotic A-40,926 Active against Highly Glycopeptide-Resistant VanA Enterococci. *J. Antibiot.* 48, 869–883. doi:10.7164/antibiotics.48.869

Marcone, G. L., Binda, E., Berini, F., and Marinelli, F. (2018). Old and New Glycopeptide Antibiotics: from Product to Gene and Back in the post-genomic Era. *Biotechnol. Adv.* 36, 534–554. doi:10.1016/j.biotechadv.2018.02.009

Marcone, G. L., Binda, E., Reguzzoni, M., Gastaldo, L., Dalmastri, C., and Marinelli, F. (2017). Classification of *Actinoplanes* Sp. ATCC 33076, an Actinomycete that Produces the Glycolipodepsipeptide Antibiotic Ramoplanin, as *Actinoplanes ramoplaninifer* Sp. Nov. *Int. J. Syst. Evol. Microbiol.* 67, 4181–4188. doi:10.1099/ijsem.0.002281

Medema, M. H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J. B., Blin, K., et al. (2015). Minimum Information about a Biosynthetic Gene Cluster. *Nat. Chem. Biol.* 11, 625–631. doi:10.1038/nchembio.1890

Medema, M. H., Cimermancic, P., Sali, A., Takano, E., and Fischbach, M. A. (2014). A Systematic Computational Analysis of Biosynthetic Gene Cluster Evolution: Lessons for Engineering Biosynthesis. *Plos Comput. Biol.* 10, e1004016–12. doi:10.1371/journal.pcbi.1004016

Mulichak, A. M., Losey, H. C., Walsh, C. T., and Garavito, R. M. (2001). Structure of the UDP-Glucosyltransferase GtfB that Modifies the Heptapeptide Aglycone in the Biosynthesis of Vancomycin Group Antibiotics. *Structure* 9, 547–557. doi:10.1016/S0969-2126(01)00616-5

Mulichak, A. M., Lu, W., Losey, H. C., Walsh, C. T., and Garavito, R. M. (2004). Crystal Structure of Vancosaminyltransferase GtfD from the Vancomycin Biosynthetic Pathway: Interactions with Acceptor and Nucleotide Ligands,. *Biochemistry* 43, 5170–5180. doi:10.1021/bi036130c

Nagarajan, R. (1993). Structure-activity Relationships of Vancomycin-type Glycopeptide Antibiotics. *J. Antibiot.* 46, 1181–1195. doi:10.7164/antibiotics.46.1181

Nakayama, A., Okano, A., Feng, Y., Collins, J. C., Collins, K. C., Walsh, C. T., et al. (2014). Enzymatic Glycosylation of Vancomycin Aglycon: Completion of a Total Synthesis of Vancomycin and N- and C-Terminus Substituent Effects of the Aglycon Substrate. *Org. Lett.* 16, 3572–3575. doi:10.1021/OL501568T/SUPPL_FILE/OL501568T_SI_001.PDF

Nicolaou, K. C., Boddy, C. N. C., Bräse, S., and Winssinger, N. (1999). Chemistry, Biology, and Medicine of the Glycopeptide Antibiotics. *Angew. Chem. Int. Ed.* 38, 2096–2152. doi:10.1002/(sici)1521-3773(19990802)38:15<2096::aid-anie2096>3.0.co;2-f

Parenti, F., and Cavalleri, B. (1989). Proposal to Name the Vancomycin-Ristocetin like Glycopeptides as Dalbaheptides. *J. Antibiot.* 42, 1882–1883. doi:10.7164/antibiotics.42.1882

Pelzer, S., Süßmuth, R., Heckmann, D., Recktenwald, J., Huber, P., Jung, G., et al. (1999). Identification and Analysis of the Balhimycin Biosynthetic Gene Cluster and its Use for Manipulating Glycopeptide Biosynthesis in *Amycolatopsis mediterranei* DSM5908. *Antimicrob. Agents Chemother.* 43, 1565–1573. doi:10.1128/aac.43.7.1565

Pootoolal, J., Thomas, M. G., Marshall, C. G., Neu, J. M., Hubbard, B. K., Walsh, C. T., et al. (2002). Assembling the Glycopeptide Antibiotic Scaffold: The Biosynthesis of from *Streptomyces toyocaensis* NRRL15009. *Proc. Natl. Acad. Sci.* 99, 8962–8967. doi:10.1073/pnas.102285099

Salas, J. A., and Méndez, C. (2007). Engineering the Glycosylation of Natural Products in Actinomycetes. *Trends Microbiol.* 15, 219–232. doi:10.1016/j.tim.2007.03.004

Shawky, R. M., Puk, O., Wietzorrek, A., Pelzer, S., Takano, E., Wohlleben, W., et al. (2007). The Border Sequence of the Balhimycin Biosynthesis Gene Cluster from *Amycolatopsis balhimycina* Contains *bbr*, Encoding a StrR-like Pathway-specific Regulator. *J. Mol. Microbiol. Biotechnol.* 13, 76–88. doi:10.1159/000103599

Snyder, N. J., Cooper, R. D. G., Briggs, B. S., Zmijewski, M., Mullen, D. L., Kaiser, R. E., et al. (1998). Enzymatic Deacylation of Teicoplanin Followed by Reductive Alkylation: Synthesis and Antibacterial Activity of New Glycopeptides. *J. Antibiot.* 51, 945–951. doi:10.7164/antibiotics.51.945

Sosio, M., Canavesi, A., Stinchi, S., and Donadio, S. (2010). Improved Production of A40926 by *Nonomuraea* Sp. Through Deletion of a Pathway-specific Acetyltransferase. *Appl. Microbiol. Biotechnol.* 87, 1633–1638. doi:10.1007/s00253-010-2579-2

Sosio, M., Stinchi, S., Beltrametti, F., Lazzarini, A., and Donadio, S. (2003). The Gene Cluster for the Biosynthesis of the Glycopeptide Antibiotic A40926 by *Nonomuraea* Species. *Chem. Biol.* 10, 541–549. doi:10.1016/S1074-5521(03)00120-0

Spohn, M., Kirchner, N., Kulik, A., Jochim, A., Wolf, F., Muenzer, P., et al. (2014). Overproduction of Ristomycin a by Activation of a Silent Gene Cluster in *Amycolatopsis japonicum* MG417-CF17. *Antimicrob. Agents Chemother.* 58, 6185–6196. doi:10.1128/AAC.03512-14

Stegmann, E., Frasch, H.-J., and Wohlleben, W. (2010). Glycopeptide Biosynthesis in the Context of Basic Cellular Functions. *Curr. Opin. Microbiol.* 13, 595–602. doi:10.1016/j.mib.2010.08.011

Thaker, M. N., Wang, W., Spanogiannopoulos, P., Waglechner, N., King, A. M., Medina, R., et al. (2013). Identifying Producers of Antibacterial Compounds by Screening for Antibiotic Resistance. *Nat. Biotechnol.* 31, 922–927. doi:10.1038/nbt.2685

Truman, A. W., Dias, M. V. B., Wu, S., Blundell, T. L., Huang, F., and Spencer, J. B. (2009). Chimeric Glycosyltransferases for the Generation of Hybrid Glycopeptides. *Chem. Biol.* 16, 676–685. doi:10.1016/j.chembiol.2009.04.013

Truman, A. W., Fan, Q., Röttgen, M., Stegmann, E., Leadlay, P. F., and Spencer, J. B. (2008). The Role of Cep15 in the Biosynthesis of Chloroeremomycin: Reactivation of an Ancestral Catalytic Function. *Chem. Biol.* 15, 476–484. doi:10.1016/j.chembiol.2008.03.019

Truman, A. W., Kwun, M. J., Cheng, J., Yang, S. H., Suh, J.-W., and Hong, H.-J. (2014). Antibiotic Resistance Mechanisms Inform Discovery: Identification and Characterization of a Novel *Amycolatopsis* Strain Producing Ristocetin. *Antimicrob. Agents Chemother.* 58, 5687–5695. doi:10.1128/AAC.03349-14

Truman, A. W., Robinson, L., and Spencer, J. B. (2006). Identification of a Deacetylase Involved in the Maturation of Teicoplanin. *ChemBioChem* 7, 1670–1675. doi:10.1002/cbic.200600308

Van Wageningen, A. A., Kirkpatrick, P. N., Williams, D. H., Harris, B. R., Kershaw, J. K., Lennard, N. J., et al. (1998). Sequencing and Analysis of Genes Involved in

the Biosynthesis of a Vancomycin Group Antibiotic. *Chem. Biol.* 5, 155–162. doi:10.1016/S1074-5521(98)90060-6

Waglechner, N., McArthur, A. G., and Wright, G. D. (2019). Phylogenetic Reconciliation Reveals the Natural History of Glycopeptide Antibiotic Biosynthesis and Resistance. *Nat. Microbiol.* 4, 1862–1871. doi:10.1038/s41564-019-0531-5

Weist, S., Kittel, C., Bischoff, D., Bister, B., Pfeifer, V., Nicholson, G. J., et al. (2004). Mutasynthesis of Glycopeptide Antibiotics: Variations of Vancomycin's AB-Ring Amino Acid 3,5-Dihydroxyphenylglycine. *J. Am. Chem. Soc.* 126, 5942–5943. doi:10.1021/ja0499389

Xu, F., Wu, Y., Zhang, C., Davis, K. M., Moon, K., Bushin, L. B., et al. (2019). A Genetics-free Method for High-Throughput Discovery of Cryptic Microbial Metabolites. *Nat. Chem. Biol.* 15, 161–168. doi:10.1038/s41589-018-0193-2

Xu, L., Huang, H., Wei, W., Zhong, Y., Tang, B., Yuan, H., et al. (2014). Complete Genome Sequence and Comparative Genomic Analyses of the Vancomycin-Producing *Amycolatopsis orientalis*. *BMC Genomics* 15, 1. doi:10.1186/1471-2164-15-363

Yim, G., Kalan, L., Koteva, K., Thaker, M. N., Waglechner, N., Tang, I., et al. (2014a). Harnessing the Synthetic Capabilities of Glycopeptide Antibiotic Tailoring Enzymes: Characterization of the UK-68,597 Biosynthetic Cluster. *ChemBioChem* 15, 2613–2623. doi:10.1002/cbic.201402179

Yim, G., Thaker, M. N., Koteva, K., and Wright, G. (2014b). Glycopeptide Antibiotic Biosynthesis. *J. Antibiot.* 67, 31–41. doi:10.1038/ja.2013.117

Yim, G., Wang, W., Thaker, M. N., Tan, S., and Wright, G. D. (2016). How to Make a Glycopeptide: a Synthetic Biology Approach to Expand Antibiotic Chemical Diversity. *ACS Infect. Dis.* 2, 642–650. doi:10.1021/acsinfecdis.6b00105

Yushchuk, O., Binda, E., and Marinelli, F. (2020a). Glycopeptide Antibiotic Resistance Genes: Distribution and Function in the Producer Actinomycetes. *Front. Microbiol.* 11, 1–9. doi:10.3389/fmicb.2020.01173

Yushchuk, O., Horbal, L., Ostash, B., Marinelli, F., Wohlleben, W., Stegmann, E., et al. (2019). Regulation of Teicoplanin Biosynthesis: Refining the Roles of Tei Cluster-Situated Regulatory Genes. *Appl. Microbiol. Biotechnol.* 103, 4089–4102. doi:10.1007/s00253-019-09789-w

Yushchuk, O., Ostash, B., Pham, T. H., Luzhetskyy, A., Fedorenko, V., Truman, A. W., et al. (2016). Characterization of the post-assembly Line Tailoring Processes in Teicoplanin Biosynthesis. *ACS Chem. Biol.* 11, 2254–2264. doi:10.1021/acschembio.6b00018

Yushchuk, O., Ostash, B., Truman, A. W., Marinelli, F., and Fedorenko, V. (2020b). Teicoplanin Biosynthesis: Unraveling the Interplay of Structural, Regulatory, and Resistance Genes. *Appl. Microbiol. Biotechnol.* 104, 3279–3291. doi:10.1007/s00253-020-10436-y

Yushchuk, O., Vior, N. M., Andreo-Vidal, A., Berini, F., Rückert, C., Busche, T., et al. (2021). Genomic-led Discovery of a Novel Glycopeptide Antibiotic by *Nonomuraea coxensis* DSM 45129. *ACS Chem. Biol.* 16, 915–928. doi:10.1021/acschembio.1c00170

Zhang, P., Zhang, Z., Zhang, L., Wang, J., and Wu, C. (2020). Glycosyltransferase GT1 Family: Phylogenetic Distribution, Substrates Coverage, and Representative Structural Features. *Comput. Struct. Biotechnol. J.* 18, 1383–1390. doi:10.1016/j.csbj.2020.06.003

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Chemistry

REVIEW
published: 13 April 2022
doi: 10.3389/fchem.2022.863118

Strategies for Glycoengineering Therapeutic Proteins

Kris Dammen-Brower[1,2†], *Paige Epler*[1,2†], *Stanley Zhu*[1,2], *Zachary J. Bernstein*[1,2], *Paul R. Stabach*[3], *Demetrios T. Braddock*[3], *Jamie B. Spangler*[1,2,4,5,6,7,8] *and Kevin J. Yarema*[1,2]*

[1]Translational Tissue Engineering Center, Johns Hopkins School of Medicine, Baltimore, MD, United States, [2]Department of Biomedical Engineering, The Johns Hopkins University, Baltimore, MD, United States, [3]Department of Pathology, Yale University School of Medicine, New Haven, CT, United States, [4]Department of Chemical and Biomolecular Engineering, The Johns Hopkins University, Baltimore, MD, United States, [5]Department of Oncology, Johns Hopkins School of Medicine, Baltimore, MD, United States, [6]Bloomberg-Kimmel Institute for Cancer Immunotherapy, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins School of Medicine, Baltimore, MD, United States, [7]Department of Ophthalmology, Wilmer Eye Institute, Johns Hopkins School of Medicine, Baltimore, MD, United States, [8]Department of Molecular Microbiology and Immunology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

Almost all therapeutic proteins are glycosylated, with the carbohydrate component playing a long-established, substantial role in the safety and pharmacokinetic properties of this dominant category of drugs. In the past few years and moving forward, glycosylation is increasingly being implicated in the pharmacodynamics and therapeutic efficacy of therapeutic proteins. This article provides illustrative examples of drugs that have already been improved through glycoengineering including cytokines exemplified by erythropoietin (EPO), enzymes (ectonucleotide pyrophosphatase 1, ENPP1), and IgG antibodies (e.g., afucosylated Gazyva®, Poteligeo®, Fasenra™, and Uplizna®). In the future, the deliberate modification of therapeutic protein glycosylation will become more prevalent as glycoengineering strategies, including sophisticated computer-aided tools for "building in" glycans sites, acceptance of a broad range of production systems with various glycosylation capabilities, and supplementation methods for introducing non-natural metabolites into glycosylation pathways further develop and become more accessible.

Keywords: glycoengineering, pharmacodynamics, pharmacokinetics, therapeutic, glycosylation, N-glycans, biomanufacturing

OPEN ACCESS

Edited by:
Zhongping Tan, Chinese Academy of Medical Sciences and Peking Union Medical College, China

Reviewed by:
Sander Izaak Van Kasteren, Leiden University, Netherlands
Qinzhe Wang, University of Utah, United States

***Correspondence:**
Kevin J. Yarema
kyarema1@jhu.edu

[†]*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Chemical Biology, a section of the journal Frontiers in Chemistry

Received: *26 January 2022*
Accepted: *25 March 2022*
Published: *13 April 2022*

Citation:
Dammen-Brower K, Epler P, Zhu S, Bernstein ZJ, Stabach PR, Braddock DT, Spangler JB and Yarema KJ (2022) Strategies for Glycoengineering Therapeutic Proteins. Front. Chem. 10:863118. doi: 10.3389/fchem.2022.863118

1 INTRODUCTION

This report describes the impact of glycosylation on the pharmacokinetics, pharmacodynamics, therapeutic activity, and production (biomanufacturing) of therapeutic proteins using several examples that illustrate strategies and methods to glycoengineer this important class of drugs for increased effectiveness. In **Section 2**, we describe how glycosylation affects the pharmacokinetics (PK) of protein-based drugs; defined simply, PK is the study of the effects of the body on a drug including absorption, distribution, metabolism, and excretion. Next, in **Section 3** and **Section 4**, we describe how glycosylation modulates a drug's pharmacodynamic (PD) properties, which are defined as the effects of the drug on the body and the body's biochemical and physiological responses to a drug. More specifically, **Section 3** covers several classes of therapeutic proteins whose PD activities depend on glycosylation, including enzymes, hormones, and blood-acting factors. **Section 4** covers therapeutic antibodies, which constitute the largest class of protein-based drugs and have unique

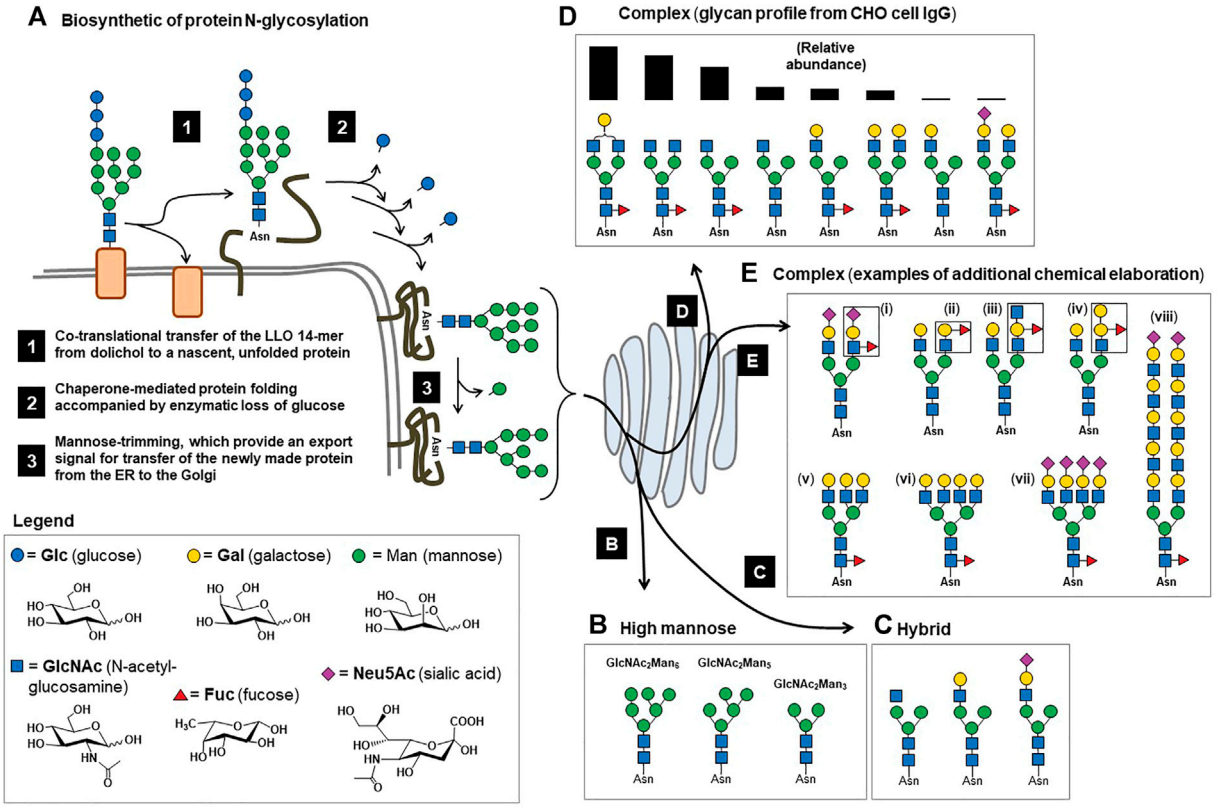


FIGURE 1 | Overview of mammalian N-glycosylation. **(A)** Step 1. The LLO 14-mer structure shown ($GlcNAc_2Man_9Glc_3$) is co-translationally transferred from dolichol phosphate to an asparagine residue of a nascent unfolded protein by oligosaccharyltransferase (OST) in the ER (Breitling & Aebi 2013). Step 2. Chaperone-mediated protein folding occurs concomitant with glucose trimming, generating a (in Step 3) a $GlcNAc_2Man_9$ or $GlcNAc_2Man_8$ structure that functions as an export signal for the transfer of successfully folded proteins to the Golgi (Helenius & Aebi 2001). **(B)** In the Golgi, further trimming of mannose residues occurs to produce a series of $GlcNAc_2Man_n$ structures referred to as "high mannose"-type N-glycans, where n is typically between 3 and 6. **(C)** Also in the Golgi, one, two, or three GlcNAc residues are added to a $GlcNAc_2Man_n$ structures, which can be further elaborated (e.g., with galactose and sialic acid, as shown) producing "hybrid" type N-glycans when a single GlcNAc is added to a $GlcNAc_2Man_n$ structure. Hybrid N-glycans typically are low in abundance and have few known roles in therapeutic proteins. A larger proportion of N-glycans have GlcNAc residues added to both terminal mannose residues of the $GlcNAc_2Man_3$ structure, most frequently resulting in small biantennary structures (Werz et al., 2007) such as those shown in Panel **(D)**, where the glycoprofile of IgG Fc domain N-glycans from one study (del Val et al., 2016) are shown rank ordered by their relative abundance. **(E)** A relatively small proportion (generally 5% or less) of N-glycans are further elaborated, resulting in epitopes such as **(i)** sialyl Lewis x (sLe^x), the H, A, and B blood type antigens **[(ii), (iii)**, and **(iv)**, respectively**]**; **(v)** tri- and **(vi)** tetra-antennary structures that can be unsialylated to fully sialylated **(vii)**; and finally certain N-glycans have extended "LacNAc" repeats (four are shown) that can serve as preferred ligands for certain receptors, such as the hemagglutinin protein of the influenza virus (Ji et al., 2017), whereas glycans from EPO can have single LacNAc repeats (Cowper et al., 2018).

glycosylation features compared to most proteins. Finally, in **Section 5** we provide an overview of methods for controlling and modulating this glycosylation during the design and biomanufacturing of therapeutic proteins. Throughout each section we provide illustrative examples of therapeutic proteins but emphasize that our examples are not complete or exhaustive.

Before covering these topics in detail, here in the Introduction (**Section 1**), we briefly describe key concepts related to the glycosylation of therapeutic proteins (**Figure 1**). With few exceptions (e.g., regulatory peptides and small hormones such as insulin), all therapeutic proteins have at least one, and often several, N-glycans. Overall, approximately 50% of human proteins are glycosylated, which governs their folding, intracellular and extracellular trafficking, stability, circulatory half-life, and immunogenicity (Olden et al., 1982; Breitfeld et al., 1984; Dwek 1996; Willey 1999; Dwek & Butters 2002). Mammalian glycosylation is remarkably complex, consisting of N-linked glycans, O-linked glycans, C-linked glycans, phosphoglycosylation, and glypiation. In this article, we will almost exclusively discuss N-linked glycosylation, because clinical translational glycoengineering efforts have overwhelmingly focused on this type of glycosylation to date.

From a biochemical perspective, virtually all cell surface or secreted proteins (i.e., candidates for drug development) are N-glycosylated, which occurs co-translationally when the lipid-linked oligosaccharide (LLO) $GlcNAc_2Man_9Glc_3$ 14-mer structure is added to a consensus sequon (**Figure 1A**, Step 1). This structure is critical for chaperone-mediated protein folding in the endoplasmic reticulum (ER) (Helenius & Aebi 2001), where the three glucose residues are sequentially trimmed during the folding process (**Figure 1A**, Step 2). Successfully folded proteins with a $GlcNAc_2Man_9$, or a slightly trimmed

$GlcNAc_2Man_8$ structure (**Figure 1A**, Step 3), are exported to the Golgi where mannosidases trim additional mannose residues, ultimately resulting in $GlcNAc_2Man_5$ to $GlcNAcMan_3$ structures (**Figure 1B**). In some cases, these "high mannose" glycans appear on mature proteins without further processing and affect the proteins' distribution and by extension, their bioactivities. In other cases, the resulting $GlcNAc_2Man_n$ glycans are precursor structures for further elaboration in the Golgi, forming hybrid (**Figure 1C**) and complex type N-glycans. In most cases, the ultimate complex type N-glycans are relatively small in size; for perspective, ~90% of mammalian glycans are comprised of 12 or fewer monosaccharides (Werz et al., 2007), which covers the size range for Fc-domain glycans of IgG antibodies (**Figure 1D**). Less frequently, complex type N-glycans can be considerably larger (**Figure 1E**), as found on therapeutic proteins such as erythropoietin (EPO).

2 PHARMACOKINETICS

Historically, the effects of glycosylation on therapeutic proteins were first evident through changes to their pharmacokinetic (PK) properties (Liu 2015; Liu 2018; Boune et al., 2020). Accordingly, we begin by describing the impact of glycosylation on the PK of protein drugs. Definitions of PK include "the movement of drugs through the body" or "the study of what the body does to a drug," and includes a drug's absorption, distribution, metabolism, and excretion; this set of metrics is typically abbreviated "ADME" (Tibbitts et al., 2016).

2.1 Serum Clearance

One of the earliest contexts where glycosylation was recognized to be important for therapeutic proteins was through serum clearance. This endpoint was evident from studies with erythropoietin (EPO), a drug that pioneered the importance of glycoengineering for improving biologics. Specifically, glycoengineering improved the PK properties of EPO by modulating two ways that glycans contribute to serum clearance, and ultimately, drug elimination. These mechanisms are kidney filtration, which can be slowed by increasing the size of a protein by adding N-glycan sites (**Section 2.1.1**) and avoiding receptor-mediated clearance by the asialoglycoprotein receptor [(ASPGR) **Section 2.1.2**] or the mannose receptor [(MR) **Section 2.1.3**].

2.1.1 N-Glycans Add Steric Bulk and Increase Hydrodynamic Radius to Avoid Kidney Filtration

The efficiency of kidney filtration rapidly increases as protein's size falls below ~40 kDa; for example, the glomerular sieving coefficient of the anionic form of horseradish peroxidase (40 kDa) is 0.007, compared to 0.33 for superoxide dismutase (~32 kDa) and 0.75 for myoglobin (16.9 kDa) (Maack et al., 1979; Tsao et al., 1991). Erythropoietin has a molecular weight of ~18.4 kDa based on its amino acid sequence, suggesting that it should experience kidney filtration similar to myoglobin. Although wild-type EPO is cleared from the serum relatively rapidly [its half-life ranges between 5 and 11 h (Elliott et al., 2008)], EPO produced with truncated N-glycans had substantially (~7-fold) faster clearance (Wasley et al., 1991). The glycosylation of EPO has now been thoroughly characterized, with the protein's three N-glycans contributing ~12 kDa of the glycoprotein's total mass of ~30.4 kDa; each N-glycan is typically a tri- or tetra-antennary structure that is highly sialylated and often has LacNAc repeats (**Figure 1E**). These large glycans are particularly effective at avoiding glomerular filtration, because unlike amino acid chains that fold into compact proteins, they are fully extended in the aqueous physiological milieu; furthermore, they are motile, allowing them to "sweep out" space.

These two factors enable glycans to increase the hydrodynamic radius of a protein more effectively than a commensurate increase in peptide mass; for example, RNAse is a ~15 kDa protein whose hydrodynamic radius is doubled through attachment of a small, biantennary N-glycan of ~2 kDa (Dwek 1996). Similarly, the size of glycosylated EPO is dramatically larger than non-glycosylated EPO (**Figure 2A**). Although the larger size of naturally-glycosylated EPO improves its serum longevity by ~7-fold compared to aglycosylated protein (Wasley et al., 1991), its molecular weight of ~30.4 kDa suggested that further improvements were possible because proteins greater than ~40 kDa have even lower glomerular sieving coefficients; for example, the coefficient for superoxide dismutase [32 kDa] of 0.33 is reduced to 0.007 for horseradish peroxidase [40 kDa]. Accordingly, the addition of two N-glycans to EPO to form hyper-glycoengineered darbepoetin alfa (Aranesp®) (**Figure 2A**) increased the drug's molecular weight to ~37–38 kDa, slowing serum clearance from ~8 to ~25 h (Macdougal 2002; Egrie et al., 2003; Elliott et al., 2008).

A limitation to glycoengineering strategies designed to avoid glomerular sieving and concomitant kidney filtration is that they depend on the target protein being appropriately sized. On one hand, if a protein or peptide is too small (e.g., insulin and/or interleukins), it may not be possible to add a sufficient number of N-glycans to enlarge the protein above the ~40 kD size threshold without loss of biological activity. In particular, EPO illustrates how both the natural and glycoengineered glycans are oriented towards one side of the protein (**Figure 2A**). In retrospect, this orientation was critical to avoid steric interference with its binding to its partner proteins; similarly fortuitous submolecular siting of built-in glycans may not be possible for all therapeutic proteins. In other cases, [e.g., ENPP-1 (**Section 2.2.1**) and therapeutic antibodies (**Section 4**)], the proteins are already above the threshold for kidney filtration, and any further increase in steric bulk is unlikely to provide additional improvement in serum longevity. In other words, EPO was ideally situated for glycoengineering due to its size, which was marginally below the threshold where glomerular sieving becomes ineffective. Nevertheless, the "size matters" principle is likely to benefit at least some additional therapeutic proteins. For example, efforts are underway to produce glycoengineered insulin (Guan et al., 2018) and glucagon (Higashiyama et al., 2018; Ichikawa et al., 2018); addition of glycans will substantially increase the

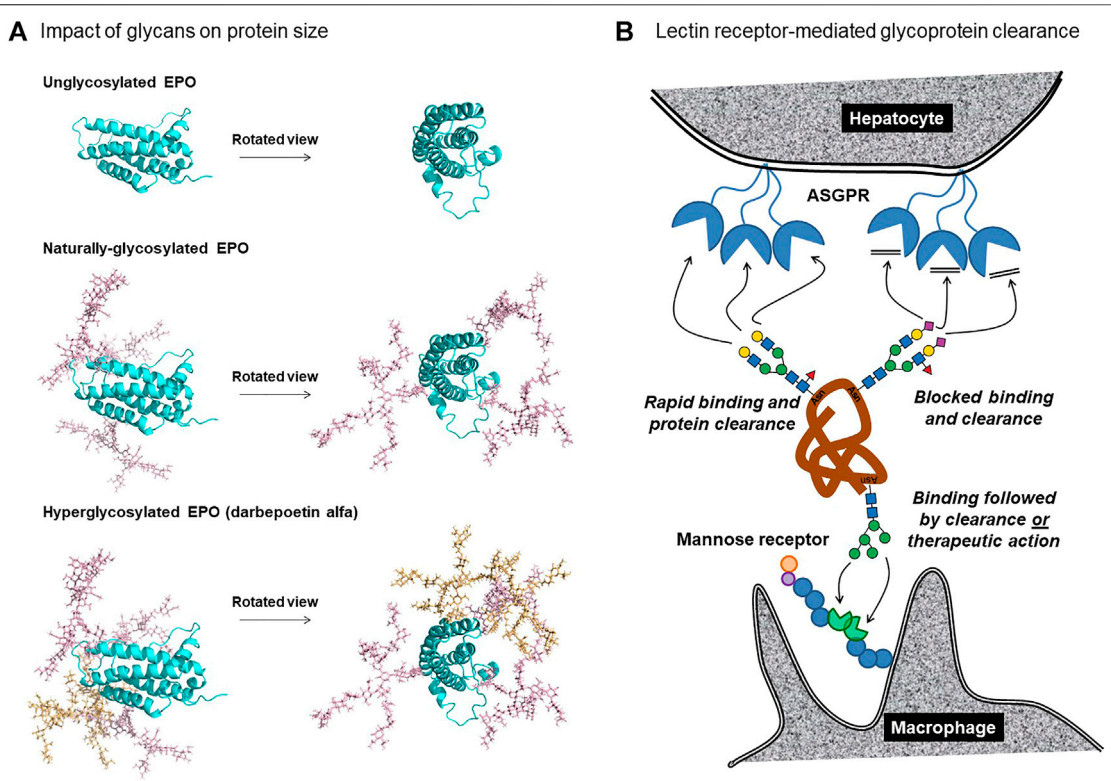


FIGURE 2 | N-glycans influence the clearance of therapeutic proteins based (A) on size and (B) receptor-mediated clearance. **(A)** Unglycosylated EPO (top) is compared with naturally glycosylated EPO, which has three N-glycans at Asn24, N38, and N83 (middle) and with darbepoetin alfa, which has five glycans including those newly-added at Asn30 and N88 (bottom). The glycan structures depicted are representative of the experimentally-determined N-glycan profile of EPO (Cowper et al., 2018), in particular the structure shown in **Figure 1E(vii)**. Protein models were generated using SWSS-MODEL software (Waterhouse et al., 2018) and modified to present N-glycan structures via CHARMM (Jo et al., 2008) and PyMOL (PyMOL Molecular Graphics System, Version 2.0, Schrödinger, LLC). The darbepoetin alfa sequence was obtained from the KEGG Drug database. **(B)** Lectin receptor-mediated clearance removes proteins from circulation through ASPGR binding of galactose-terminated glycans (top left); addition of sialic acid masks the galactose blocking binding and clearance (top right). Mannose-terminated glycans bind to mannose receptors on macrophages, dendritic cells, dermal fibroblasts, and keratinocytes, resulting in clearance or, in some cases, therapeutic activity [for example to treat Gaucher disease (**Section 2.3.3**)].

hydrodynamic radius of these small proteins, potentially slowing kidney filtration.

2.1.2 Sialylation Masks Asialoglycoprotein Receptor-Mediated Clearance

As just discussed, adding steric bulk to a therapeutic protein *via* glycosylation can be an effective albeit limited strategy to improve PK properties. A more general glycan-related clearance mechanism involves receptor-mediated cellular uptake by lectin receptors. The dominant example of this mechanism involves hepatic clearance of serum proteins *via* the asialoglycoprotein receptor [ASGPR (Ashwell & Harford 1982)]. The ASPGR functions by multivalent recognition of the terminal galactose residues of non-sialylated N-glycans, rapidly depleting the host proteins from circulation (Schwartz 1984; Weigel 1994) (**Figure 2B**). The effectiveness of this mechanism for removing "aged" proteins from the serum as they lose their terminal sialic acids over time, thereby exposing their otherwise penultimate galactose moieties, is illustrated by deliberately desialylated EPO, which has a serum half-life of ~10 min. By contrast, normally sialylated EPO has a serum half-life ranging from 5 to 11 h (Elliott et al., 2008). The increased serum longevity of darbepoetin alfa is not only attributed to increased size (**Figure 2A**) but also to hypersialylation, having as many as 22 copies of sialic acid (Elliot et al., 2000), which helps it avoid ASGPR clearance (**Figure 2B**). This pioneering example illustrates the general importance of high sialic site acid occupancy for prolonged *in vivo* circulation of therapeutic proteins. As an aside, sialic acid can improve the safety of therapeutic proteins by a similar masking mechanism where this sugar obscures underlying antigenic epitopes, reducing the generation of neutralizing antibodies (Bork et al., 2009; Li & d'Aniou 2009).

2.1.3 Mannose Receptor-Mediated Glycoprotein Clearance

Glycoproteins also can be recognized by mannose-binding receptors (MRs) on various cell types, including hepatocytes, fibroblasts, and endothelial cells, as well as by immune cells such as macrophages and dendritic cells (Schlesinger et al., 1978; Sheikh et al., 2000). These receptors have multiple functions. One function is to rapidly clear proteins with high mannose-type

glycans (**Figure 1B**), as well as GlcNAc and fucose-containing glycans (Feinberg et al., 2021), such as the glycoprotein hormone lutropin. In general, these receptors help maintain serum glycoprotein homeostasis (Roseman & Baenziger 2000; Lee et al., 2002). A second function of MRs is to facilitate the phagocytosis of pathogens such as *Candida albicans*, *Pneumocystis carnini,* and *Leishmania donovan* whose surfaces are covered with mannose-terminated glycans. These glycans allow the removal of these pathogens from the host by macrophages as well as by non-immune cells that also express mannose receptors such as keratinocytes (Szolnoky et al., 2001; Gazi & Martinez-Pomares 2009). A third and also immunomodulatory function of MRs is to enhance soluble, but not cell-associated antigens, for cross-presentation (Burgdorf et al., 2006).

Another aspect of human immune response to pathogens is the generation of inflammatory glycoproteins such as hydrolases, tissue plasminogen activator, and myeloperoxidase, which can be damaging to host tissues if retained after the infection has been resolved; high-mannose glycans on these glycoproteins provide these conditionally protective factors with quick clearance *via* cells with MRs, helping to avoid post-infection damage to the host (Lee et al., 2002; Gazi & Martinez-Pomares 2009). From a drug development standpoint, the ability of certain cells to internalize mannose-terminated glycans has been exploited to direct therapeutic proteins to cell types such as macrophages, as described for Gaucher's disease in **Section 2.3.3**.

2.1.4 IgG Antibodies: An Exception to Rapid Clearance

Therapeutic antibodies, which to date are almost all IgGs, are outliers compared to other therapeutic proteins because they are not subject to the two "universal" clearance mechanisms just discussed (size-based kidney filtration and glycan-based receptor clearance). First, IgG antibodies are large (~150 kD), well above the size range susceptible for kidney filtration. Second, the N-glycans of commercial IgG antibodies are uniquely oriented inwards, being "buried" between the two Fc region protein domains, making them largely inaccessible to ASGPR clearance despite their low sialylation status (**Figure 1D**). In addition to glycan-based clearance mechanisms, the Fc domain of IgG antibodies binds to the neonatal Fc receptor, which directs intracellular trafficking to avoid proteosomal degradation upon uptake into the cell by re-releasing the antibody into circulation. These factors provide therapeutic antibodies with *in vivo* half-lives ranging from several days to many weeks (Ryman & Meibohm 2017; Liu 2018; Ovacik & LIn 2018) instead of the several hours typical of most other protein-based drugs. For example, the half-life of the commercial anti-HER2 antibody drug trastuzumab is 28 days (Boekhout et al., 2011), even though only ~1.1% of its Fc N-glycans are sialylated (Nakano et al., 2009).

2.2 Absorption and Distribution

Unlike the well-known role for glycosylation in the elimination of therapeutic proteins and in already-approved glycoengineered drugs such as darbepoetin alfa that exploit glycans for improved circulatory half-life, the role of glycoengineering in modulating the absorption and distribution of these drugs throughout the body is in relative infancy. Nevertheless, two case studies (ENPP1-Fc, **Section 2.2.1** and hyaluronidase, **Section 2.2.2**) demonstrate the intriguing potential for exploiting glycoengineering to improve the absorption and biodistribution of therapeutic proteins. In this discussion, we focus on subcutaneous delivery. Subcutaneously injected therapeutics have been popular for their potential convenience for physicians, patients at greater risk for systemic reactions, and those in which constant venous access is difficult to maintain (particularly infants) (Turner & Balu-Iyer 2018). Furthermore, subcutaneous delivery often allows patient self-administration, reducing the cost, stress, and inconvenience of repeated administration at a healthcare center. These benefits have made subcutaneous administration appealing to a growing number of therapeutic proteins, including cytokines, human insulin, and immunoglobulins (Turner & Balu-Iyer 2018).

2.2.1 Absorption of Glycoengineered ENPP-1

Despite the many benefits of subcutaneous administration, this method is limited in the volume that can be infused, and perhaps more importantly, the bioavailability of the therapeutic following injection. In one study, a glycoengineering strategy dramatically improved the bioavailability of subcutaneously delivered ENPP1-Fc. As a brief introduction, ENPP1 is ectonucleotide pyrophosphatase/phosphodiesterase 1, a blood enzyme whose deficiency results in generalized arterial calcification of infancy (GACI), a potentially lethal disease (Ferreira et al., 2021). Wild-type ENPP1 has a short serum half-life of ~5 h when used for enzyme replacement therapy (ERT), necessitating thrice a day dosing in a mouse model of GACI for therapeutic effectiveness. Braddock's research team first took a protein engineering approach by fusing an IgG Fc domain to ENPP1 (Martins et al., 2016). The resulting ENPP1-Fc construct had a substantially improved serum half-life of ~37 h (**Figure 3A**) but nonetheless relatively modest bioavailability when delivered subcutaneously (Albright et al., 2016). By taking a glycoengineering approach and adding a fifth N-glycan site to ENPP1-Fc through an I256T mutation (**Figure 3B**; the methodology for adding N-glycans to therapeutic proteins is outlined in Section *5.1.3*), the serum half-life almost doubled (from 37 to 67 h; **Figure 3F**), while a surrogate measure of bioavailability, the cumulative "area under curve" (AUC) value for enzyme activity in the serum, increased dramatically by 794% from 3,400 to 27,000 units (Stabach et al., 2021).

As a caveat, the biochemical mechanism for the PK improvements for ENPP1-Fc remain incompletely defined; for example, unlike the "size matters" improvement when N-glycans were added to EPO (**Figure 2**), ENPP1-Fc is already a large-sized protein, making it unlikely that avoidance of kidney filtration was involved in its improved serum longevity. A straightforward explanation, such as increased enzyme activity for the I256T glycoform, was ruled out by measurements that show that the enzyme's catalytic activity was affected negligibly (Stabach et al., 2021). Instead, it is plausible (but not experimentally verified) that reduced access of serum proteases to exposed protein surfaces protected by the newly-added glycan reduced degradation (**Section 2.3.2**), concomitantly increasing serum

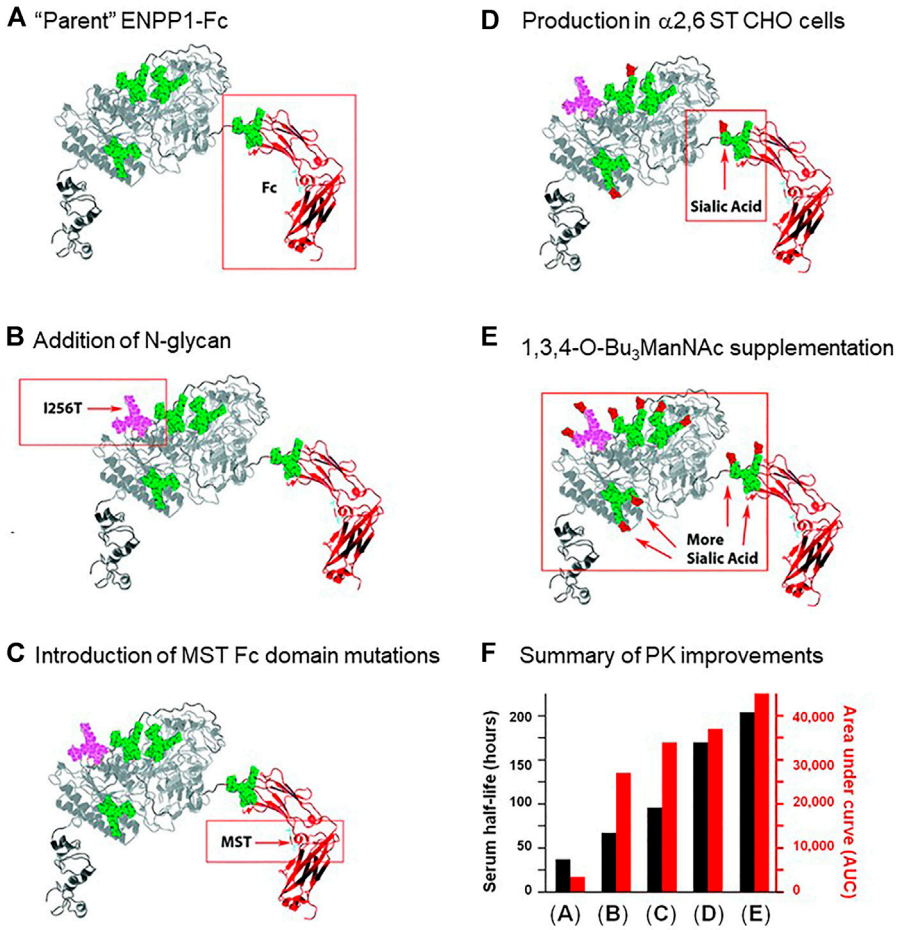


FIGURE 3 | ENPP1 protein and glycoengineering. Improvements made to the pharmacokinetics of ENPP1 as reported by Stabach and coauthors (Stabach et al., 2021) are summarized in this figure. **(A)** First, in previous work (Albright et al., 2016), the enzyme was fused to the immunoglobulin Fc domain to increase protein recycling and serum recirculation through interactions with the neonatal Fc receptor (Albright et al., 2016); this "parent" construct had a serum half-life of 37 h and an AUC of 3,400 as depicted graphically in Panel **(F)**. **(B)** Addition of an N-glycan site was achieved through the I256T mutation to ENPP1 resulting in addition of the glycan to Asn254; this newly added N-glycan approximately doubled serum half-life and octupled the AUC value. **(C)** Mutation of Met, Ser, and Thr (MST) that increase affinity for the neonatal Fc receptor (Vaccaro et al., 2005) were introduced into ENPP1-Fc Fc's domain, further improving both serum half-life and AUC. Finally, two approaches to increase sialylation including **(D)** production of ENPP1-Fc in α2,6-sialyltransferase overexpressing CHO cells and **(E)** supplementation of the culture medium with the sialic acid metabolic precursor 1,3,4-O-Bu_3ManNAc sequentially further increased serum half-life (to a final value of 204 h) and the AUC value (to 45,000).

longevity. The increase in apparent bioavailability evidenced by the I256T glycoform's dramatic AUC increase is also unexplained; a specific structure-activity response appears to be involved insofar as only one of over 50 glycovariants of ENPP1-Fc created in the study gained such a dramatically improved ability to effectively extravasate from the subcutaneous compartment into circulation (Stabach et al., 2021). At present, it is unknown if the mechanisms involved will apply to therapeutic proteins in general or whether they are unique to ENPP1-Fc.

2.2.2 Hyaluronidase-Assisted Subcutaneous Delivery of Therapeutic Proteins

Unlike the addition of an N-glycan to ENPP1-Fc that serendipitously improved its PK properties, hyaluronidase provides a broader approach to facilitate the absorption and bioavailability of subcutaneously-delivered therapeutics. Hyaluronan contributes to inefficient bioavailability of subcutaneously-injected drugs by endowing the hypodermis with viscoelastic properties that prevent bulk fluid flow of liquids or the diffusion of drug molecules, in particular high molecular weight therapeutic proteins (Frost 2007). Recombinant human hyaluronidase (rHuPH20) enzymatically degrades hypodermal hyaluronan, helping to overcome this impediment for subcutaneous drug delivery (Frost 2007; Wasserman 2017; Liu et al., 2021). One example of hyaluronidase's efficacy is for subcutaneous delivery of IgG to treat primary immunodeficiency diseases (PIDDs) where regular and prolonged bioavailability of antibodies is essential (Wasserman 2017). In a second example, hyaluronidase can degrade hyaluronan capsules associated with tumors, increasing the accessibility and

effectiveness of anti-cancer drugs (Shuster et al., 2002; Whatcott et al., 2011; McAtee et al., 2014; Kohi et al., 2016; Maneval et al., 2020).

2.3 Metabolism: Enzymatic Modification and Intracellular Trafficking

The term "metabolism" (i.e., the "M" in ADME) is broadly defined in this article to include any host-mediated enzymatic modification of a therapeutic protein, including catabolic (**Section 2.3.1**) and biosynthetic (**Section 2.3.2**) activities, as well as aspects of intracellular trafficking (**Section 2.3.3**).

2.3.1 Serum Sialylation and Desialylation

As mentioned above, a major determinant of serum longevity is the sialylation status of many types of therapeutic proteins through shielding from ASGP receptor-mediated clearance. Accordingly, efforts are made to fully sialylate therapeutic proteins as practical; for example, EPO produced in CHO cells has sialic acid occupancy of 70–85% up to as high as 99% [i.e., ~22 sialic acids per molecule of darbepoetin alfa (Elliot et al., 2000; Egrie & Browne 2002)]. Once in circulation, sialidase present in the serum stochastically remove sialic acids over time. As proteins become less and less sialylated, the loss of this terminal sugar functions as a molecular clock leading to the clearance of older and damaged proteins by the ASGPR. More recently, the idea has emerged that biosynthetic sialylation can also occur in the serum; in particular, sialic acid is added to the Fc glycans of circulating IgG antibodies. In rodents, the sialylation of IgG N-glycans is linked to secreted ST6Gal1 produced by liver epithelial cells and CMP-sialic acid leached into the serum by degranulating platelets (Jones et al., 2016). This naturally-occurring precedent for post-production modification of immunomodulatory proteins (Johnson et al., 2013), along with the commercial availability of reasonably-priced sialyltransferases, has opened the door for cell-free glycoengineering of protein therapeutics (**Section 5.1.2**).

2.3.2 Glycan Shielding of Protease Activity

Recently, the SARS-CoV-2 virus has provided a dramatic example of how glycans can shield a protease cleavage site. For this virus (Casalino et al., 2020; Gong et al., 2021), and others such as influenza (Tong et al., 2003), heavy glycosylation is advantageous for evading host immunity by shielding underlying immunogenic foreign epitopes of the viruses. Conversely, the furin protease cleavage site that mediates cell infectivity of SARS-CoV-2 is sterically shielded by nearby glycans, providing evolutionary pressure for reduced glycosylation (Zhang et al., 2021). Based on this precedent, the addition of glycans to therapeutic proteins has been considered for protection from proteases that cleave proteins during degradation, although a potential downside is loss of the protein's biological function. Indeed, an original impetus for adding N-glycans to ENPP1-Fc (**Figure 3**) was to protect the enzyme from proteases (Stabach et al., 2021). Guan and others added an O-linked tri-mannose structure to insulin, enhancing its proteolytic stability and decreasing unwanted aggregation while maintaining biological activity (Guan et al., 2018). In addition to protection from protease degradation, N-glycans play multiple auxiliary roles in protein stability by protecting proteins from oxidation, aggregation, pH-induced damage, and thermal degradation (Qun & Qiu 2019).

2.3.3 Intracellular Trafficking

Another way that glycosylation can affect protein degradation, although indirectly, is through intracellular trafficking. A naturally-occurring example is the impact of hybrid and complex N-glycans on the cell surface vs. lysosomal/endosomal targeting of endogenously-produced sodium potassium chloride cotransporter NKCC1 encoded by *SLC12A2* (Singh et al., 2015). A second example is that increased sialylation weakens the galectin lattice and directs the epidermal growth factor receptor (EGFR) for degradation instead of surface recycling (Lajoie et al., 2007; Mathew et al., 2016). The ability to modulate subcellular trafficking through N-glycan composition led to the use of glycoengineering to create successful enzyme replacement therapy for Gaucher disease (GD). For context, initial efforts in the 1970s to use unmodified human β-glucocerebrosidase to treat GD were unsuccessful because macrophages (the target cells in this disease) did not bind and internalize this enzyme when it was isolated from natural sources (Tekoah et al., 2013); it was later discovered that the enzyme's inefficient uptake could be ameliorated through a glycoengineering approach.

Specifically, upon discovery that glycans with exposed terminal mannose residues facilitated macrophage uptake of β-glucocerebrosidase (Friedman et al., 1999; Sato & Beutler 1993), glycoengineered versions of this enzyme were created to treat GD. The first version made was imiglucerase (Cerezyme®) produced in CHO cells and modified enzymatically after production to expose mannose, resulting in ~40–60% of exposed Man_3 structures (**Figure 1B**). A second version, velaglucerase alfa (Vpriv®) is produced in human fibroblast carcinoma cells and achieves ~100% exposed Man_5-Man_9 residues through treatment of the production cells with kifunensine, a mannosidase I inhibitor; this drug has ~2-fold greater internalization into macrophages compared to imiglucerase, showing the importance of glycosylation in therapeutic efficacy (Brumshtein et al., 2010). Taliglucerase alfa (Elelyso®) is a third version of therapeutic β-glucocerebrosidase; it is produced in a carrot cell-based production system and achieves ~100% exposed Man_3 residues without *in vitro* processing or mannosidase inhibitors. Taliglucerase alfa has increased uptake into macrophages compared to imiglucerase (Shaaltiel et al., 2007), presumably because of its completely unshielded terminal Man_3 groups. This example of multiple competing products to treat GD, using alternative methods to control glycosylation towards the common goal of exposed terminal mannose residues, illustrates the benefits of flexible biomanufacturing platforms that tailor glycosylation for individual diseases, as outlined in **Section 5**, below.

3 IMPACT OF GLYCOSYLATION ON PHARMACODYNAMICS AND BIOLOGICAL ACTIVITY

Here, in **Section 3**, we describe how biochemical interactions mediated through glycosylation affects a drug's pharmacodynamic (PD) properties, which are defined as the body's biological response to a drug [i.e., what the drug does to the body; the word comes from the Greek "pharmakon" meaning drug and "dynamikos" meaning power (Marino et al., 2021)]. Pharmacodynamic properties are broad, including receptor, cofactor, and ligand interactions as well as virtually all other biological activities of a protein (Marino et al., 2021). Therapeutic proteins fall into several categories; in this report, we cover therapeutic enzymes in **Section 3.1**, hormones in **Section 3.2**, and blood proteins in **Section 3.3** (therapeutic antibodies are covered in **Section 4**), providing examples illustrating how glycans impact the PD properties of these drugs and how glycoengineering can improve therapeutic efficacy.

3.1 Enzymes

3.1.1 Hyaluronidase

The biological activity of hyaluronidase, the enzyme that facilitates subcutaneous drug delivery through transient solubilization of hyaluronan in the hypodermis (**Section 2.2.2**), depends on glycosylation. Recombinant human hyaluronidase (rHuPH20) is heavily glycosylated with size N-glycan sites (Asn47, Asn131, Asn200, Asn219, Asn333, and Asn358) that are all modified with high mannose type N-glycans (Frost 2007; Liu et al., 2021) (**Figure 1B**). As discussed above (**Section 2.3.3**), high mannose structures target proteins for clearance *via* MRs; hyaluronidase's glycosylation status also affects its biological activity, and by extension its PD properties (Liu et al., 2021). Specifically, PNGase removal of its N-glycans decreased enzymatic activity of rHuPH20 by ~80% in an *in vitro* assay; in a corresponding *in vivo* test, aglycosylated rHuPH20 dramatically reduced trypan blue dispersion (a surrogate measure of drug diffusion) in a mouse model when compared with naturally-glycosylated enzyme (Liu et al., 2021). This study illustrated how N-glycosylation was necessary for rHuPH20 to solubilize host hyaluronan (i.e., a PD effect) for facilitating subcutaneous delivery of a second drug (i.e., a PK effect). As the complex interplay between such PK properties and PD endpoints becomes more widely appreciated, the growing toolkit to glycoengineer therapeutic proteins (**Section 5**) to optimize both endpoints is becoming increasingly important.

3.1.2 Esterases

Esterases are a diverse family of enzymes that have several pharmaceutical roles. In some cases, reminiscent of the role of hyaluronidase in improving subcutaneous drug delivery, esterases augment the effectiveness of a second drug. For example, esterases activate pro-drugs such as the Alzheimer's drug tacrine (Bencharit et al., 2003), doxazolidine carbamates (Burkhart et al., 2006), the breast cancer drug tamoxifen (Fleming et al., 2005), the influenza drug oseltamivir (Shi et al., 2006), and hexosamine analogs used in metabolic glycoengineering (Mathew et al., 2017; Sarkar et al., 1995; Wang et al., 2009) (**Section 5.1.5**). Esterases also detoxify narcotics such as cocaine and heroin (Pindel et al., 1997) as well as chemical warfare agents such as soman and tabun (Fleming et al., 2003). Finally, these enzymes are being investigated for the direct treatment of diseases such as Alzheimer's (Greig et al., 2002; Nordberg et al., 2013; Saez-Valero et al., 2000), Similar to hyaluronidases, glycosylation modulates both the enzymes' PK and PD properties (Kolarich et al., 2008; Schneider et al., 2013; Weikert et al., 1994; Xu et al., 2015). In particular, sialylation is important for prolonging serum circulation (Chitlaru et al., 1998; Fukami & Yokoi 2012) and glycosylation affects the catalytic activity of several esterases including human acetylcholinesterase (Velan et al., 1993), human carboxylesterase 1 (Arena de Souza et al., 2015; Kroetz et al., 1993), and human carboxylesterase 2 (Alves et al., 2016). In one example of how glycoengineering can improve esterases, a metabolic glycoengineering approach (**Section 5.1.5**) using 1,3,4-O-Bu_3ManNAc to sialylation (**Section 5.1.5**) increased sialylation of glycans situated at the interface of trimeric units of carbosylesterase one; *in silico* modeling indicated that these glycans increased the stability of the multimeric, active form of this enzyme (Mathew et al., 2017).

3.2 Hormones: Hypoglycosylated Follitropins

In many cases, gain-of-glycosylation (e.g., increased sialylation or newly-added N-glycans) improve PK or PD properties of therapeutic proteins. In some cases, however, reduced glycosylation can be beneficial, as is illustrated by the follicle stimulating hormone (FSH). This hormone is produced in the anterior pituitary and travels through the circulation to gonodal cells where it interacts with FSH receptors (FSHRs) to promote follicle development in women and spermatogenesis in men (Daya 2004; Davis et al., 2014; Ulloa-Aguirre et al., 2018). Therapeutically, recombinant FSH or follitropins can substitute for naturally-occurring FSH deficiencies to treat infertility (Dias & Ulloa-Aguirre 2021).

Endogenous FSH consists of an alpha and beta subunit; both have two putative sites of N-glycosylation. The alpha subunit is consistently fully glycosylated with the beta subunit occupied with zero, one, or two N-glycans (Davis et al., 2014; Ulloa-Aguirre et al., 2018; Dias & Ulloa-Aguirre 2021). The alpha subunit of FSH plays a pivotal role in receptor interactions by engaging the receptor-ligand interface (Ulloa-Aguirre et al., 2018; Dias & Ulloa-Aguirre 2021). The importance of the glycosylation of the alpha subunit is illustrated by the deletion of one glycosite (at Asn78), which increases FSHR binding, while the removal of its other N-glycan (Asn52) decreases efficacy (Ulloa-Aguirre et al., 2018; Dias & Ulloa-Aguirre 2021). Similarly, removal of the glycosylation sites on the beta subunit of FSH yielded significantly greater bioactivity (Dias & Ulloa-Aguirre 2021). Overall, hypoglycosylated FSH 9- to 26- fold more active than its fully glycosylated variant but also experienced reduced *in vivo* half-life, presumably due to loss of α2,3-siaylation (Ulloa-Aguirre

et al., 2018). These experiments completely removed N-glycans at each site and did not explore microheterogeneity leaving open the intriguing possibility that fucosylation, sialylation, increased glycan branching, or another property could be tuned to optimize the glycosylation profile for FSH to meet the dual but competing PK and PD requirements. Overall, FSH demonstrates how glycosylation can have complex effects on a therapeutic protein by augmenting one endpoint while undermining the other, reinforcing the need for versatile glycoengineering strategies to meet such competing demands.

3.3 Blood-Modulatory Proteins

Overall, therapeutic proteins are dominated by blood-acting or blood-modulatory proteins (e.g., EPO and ENPP1-Fc, discussed above and antibodies that largely function in the blood are the largest class of therapeutic proteins; **Section 4**). Another category of blood-regulatory proteins whose activity critically depends on glycosylation are clotting factors that need to be administered therapeutically for people with deficiencies in these proteins, such as hemophilia patients. Deglycosylation diminishes the conformational stability, activity, and macromolecular interactions of coagulation factor VIII [FVIII (Kosloski et al., 2009)] and decreases the effectiveness of factor XIII-B [FXIII-B (Hurjak et al., 2020)]. Based on the importance of glycosylation in blood clotting, efforts to produce coagulation factors in low-cost hosts (e.g., in plant cells, **Section 5.3.4**) to increase availability for patients are cognizant of the importance of maintaining appropriate glycosylation; this topic is discussed extensively in a review article by Top and coauthors (Top et al., 2019).

4 THERAPEUTIC ANTIBODIES

Monoclonal antibodies are the largest class of biotherapeutics on the clinical market; in April 2021 the FDA approved its 100th monoclonal antibody product, GlaxoSmithKline's PD1 blocker dostarlimab (Mullard 2021). The specificity, signaling versatility, and half-life of antibodies, all of which are modulated by glycosylation (Alter et al., 2018; Buettner et al., 2018; Irvine & Alter 2020), make them potent and highly sought therapeutics against a variety of diseases. Brian Cobb's review article on antibody glycosylation (Cobb 2020) partitions the history of IgG glycosylation into two overlapping eras. The first era began in the 1970s when research uncovered how glycosylation contributed to the pro-inflammatory activities of IgG antibodies. Based on almost half a century of foundational knowledge, pro-inflammatory mAbs are now in clinical practice, mainly designed to destroy cancer cells; these efforts are described in more detail in **Section 4.1**. The second era of IgG glycosylation can be traced roughly to Jeffrey Ravetch's group's discovery that terminal α2,6-sialylation (**Figure 4A**) of IgG's Fc glycans endowed these antibodies with anti-inflammatory properties (Kaneko et al., 2006). Efforts are underway to exploit these antibodies for intravenous immunoglobin (IVIg) and other therapies, as covered in **Section 4.2**. In the body, antibodies typically have either pro- or anti-inflammatory activities but their exquisite ability to bind to select targets—and by doing so inactivate the activity of the marker—has led to the creation of numerous blocking and neutralizing antibodies; as described in **Section 4.3**; to date this class of therapeutics has found great utility in cancer treatment by ablating the activity of oncoproteins and intense efforts are devoted developing broadly neutralizing antibodies for HIV-1. Finally, in **Section 4.4** we outline glycoengineering approaches to increase the potency of antibody-drug conjugates.

4.1 Pro-Inflammatory Antibodies

As depicted in **Figures 1D**, **4A** and described in detail elsewhere (Pereira et al., 2018; Wang et al., 2018; Zafir et al., 2013), glycan patterns on the conserved fragment crystallizable (Fc) region of IgG antibodies have significant effects on an antibody's pro-inflammatory activities [e.g., antibody-dependent cellular cytotoxicity (ADCC), antibody-dependent cellular phagocytosis (ADCP), and complement-dependent cytotoxicity (CDC), discussed here in **Section 4.1**] as well as their anti-inflammatory activity (**Section 4.2**). Briefly stated, increasing elaboration of an Fc N-glycan with galactose, core fucose, and sialic acid increases anti-inflammatory activity, and antibodies designed to induce ADCC, CDC, and ADCP benefit from the absence of these monosaccharides (Buettner et al., 2018).

4.1.1 Mechanism(s)

The majority of pro-inflammatory antibodies now in clinical use are designed to bind to tumor selective antigens and elicit downstream effector responses (Ząbczyńska et al., 2020) that kill the target cancer cells. Specific mechanisms of action include ADCC, ADCP, and CDC where ADCC is a type of immune reaction where the target cell becomes coated with the therapeutic antibodies and then is lysed by immune effector cells that include natural killer (NK) cells, macrophages, neutrophils, and eosinophils. ADCP utilizes a similar process but the effector cells, typically macrophages, phagocytose antibody-opsonized target cells instead of lysing them. CDC is mediated by IgG and IgM antibodies, which trigger the classical complement pathway to lyse the target cells upon binding of the C1q protein to the Fc region of Fcγ receptors. Naturally-occurring unbalanced glycosylation profiles can lead to and/or exacerbate pro-inflammatory ADCC and CDC in disorders, such as the destruction of thyroid tissue in Hashimoto's thyroiditis (Ząbczyńska et al., 2020).

In general, with the effects of ADCC being the most thoroughly studied, sialic acid and core fucose inhibit these pro-inflammatory responses that often are desired in anti-cancer antibodies. Mechanistically, the glycan composition allosterically alters Fcγ receptor interactions, as reviewed in detail elsewhere (Zafir et al., 2013; Pereira et al., 2018; Wang et al., 2018). The IgG glycomes of human-derived antibodies are highly fucosylated, with afucosylated IgG ranging from only ~1.3–19.3% in one study (Pucić et al., 2011); CHO cell-produced IgG has a similarly high fucose occupancy of 90% or more (**Figure 1D**). As discussed below (**Section 4.1.2**), the highly fucosylated glycoprofile of CHO cell produced antibodies has led to glycoengineering efforts to produce afucosylated mAbs to treat cancer *via* ADCC. By contrast, anti-cancer IgG antibodies

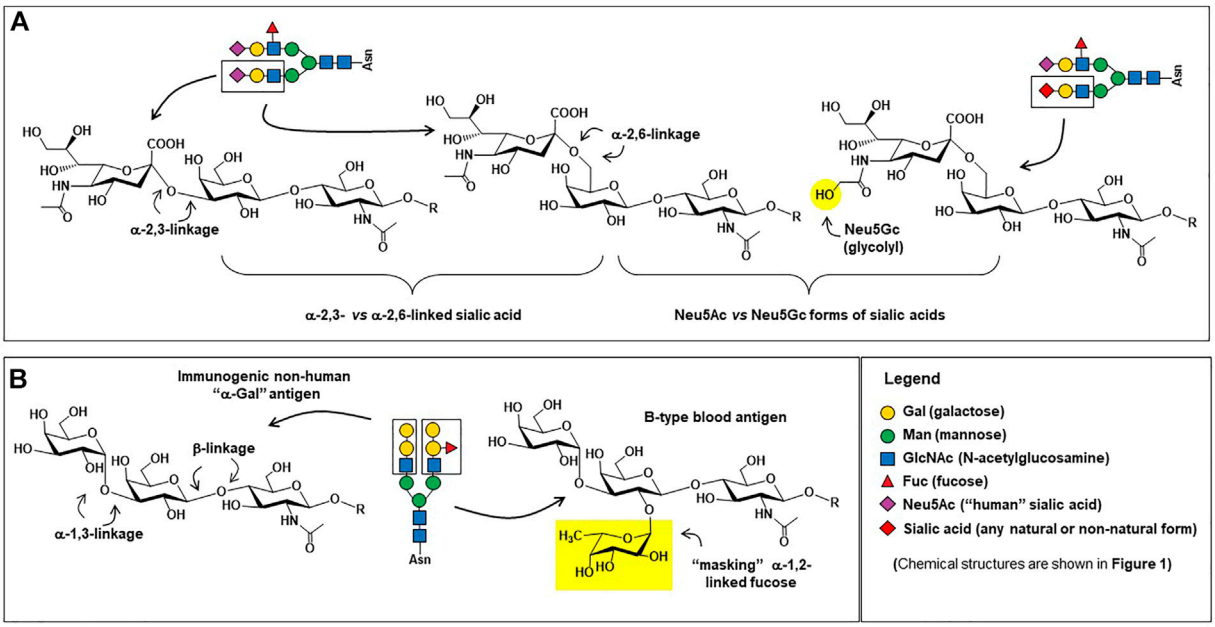


FIGURE 4 | Carbohydrate epitopes relevant to therapeutic antibodies. **(A)** Sialic acid is found in human proteins in both α2,3-linkages (left) and α2,6-linkages (center); α2.6-linked sialic acid is critical for providing IgG antibodies with anti-inflammatory characteristics (Kaneko et al., 2006) whereas α2,3-linked sialic acid are effective at preventing ASPGR clearance (Ellies et al., 2002). The presence of the N-glycolylneuraminic acid (Neu5Gc, right) form of sialic acid on proteins produced in non-human mammalian cells can be pro-inflammatory (Tangvoranuntakul et al., 2003; Samraj et al., 2015), which may or may not be desired in a therapeutic protein. **(B)** The structure of the "α-Gal" trisaccharide epitope (left) is a major safety concern (**Section 5.2.1**); in human cells, the terminal alpha-linked galactose is not added to a glycan until the penultimate masking α1,2-linked fucose (right) is installed, preventing the synthesis of the "naked" immunogenic α-Gal epitope. Incidentally, the tetrasaccharide shown comprises the B-type blood antigen, whose present is a quality control parameter in IVIg therapy (**Section 5.1.2**).

produced in industry-standard CHO cells have an attractive pro-inflammatory profile insofar as 98% or more of the Fc domain N-glycans are asialylated; nevertheless, emerging evidence suggests that even residual levels of 2% or less sialic acid can have a profound anti-inflammatory effects (**Section 4.1.3**).

4.1.2 Afucosylated Clinical Antibodies

Evidence that natural variations in Fc glycosylation impact IgG antibody activity spurred efforts to produce afucosylated therapeutic antibodies; for example, these antibodies have superior anti-HIV-1 activity (Ackerman et al., 2013). As of 2018, there were three FDA-approved afucosylated antibodies: Obinutuzumab (Gazyva®; targets CD20), Mogamulizumab (Poteligeo®; targets CC chemokine receptor 4, and benralizumab (Fasenra™; targets IL-5Rα), with more than 20 in clinical trials (Pereira et al., 2018). Since then, inebilizumab [Uplizna®; targets CD19 to treat neuromyelitis optica spectrum disorder (NMOSD) (Cree et al., 2019)] has been approved and ublituximab, which targets CD20 to treat multiple sclerosis and chronic lymphocytic leukemia is in the final stages of approval (Fox et al., 2021).

Of these afucosylated antibodies, obinutuzumab and mogamulizumab are both anti-cancer drugs where lack of fucosylation increases ADCC, ADCP, or CDC potency against tumor cells. For example, the afucosylated CD20-targeting drug obinutuzmab activates neutrophils and mediates phagocytosis more efficiently than rituximab, which is a normally fucosylated CD20-targeting mAb (Golay et al., 2013). By contrast, benralizumab blocks IL-5R signaling leading to ADCC-mediated depletion of IL-5Rα-expressing eosinophils (Kolbeck et al., 2010); in essence it is an anti-inflammatory mAb by leading to the death of excess immune cells to treat severe eosinophil asthma.

4.1.3 Asialylated Clinical Antibodies

From a practical perspective, the glycan profile of IgG therapeutic antibodies produced in industry-standard CHO cells, superficially at least, has an attractive pro-inflammatory profile insofar as 98% or more of the drug copies are asialylated (**Figure 1D**). As a result, unlike multiple efforts to reduce fucosylation that already have been adopted for commercial biomanufacturing and received regulatory approval, efforts to reduce sialylation have lagged. Nevertheless, the importance of reducing even the residual levels of sialic acid in therapeutic antibodies was illustrated by a study of pertuzumab (Perjeta®, Genentech), a mAb that binds to HER2, blocking its dimerization and subsequent oncogenic signaling.

Although the mechanism of action of pertuzumab was originally described as a conventional blocking/neutralizing antibody (i.e., by blocking HER2 signaling in breast cancer), it also has pro-inflammatory activity *via* ADCC and CDC. To explore whether these activities could be augmented by desialylation, Luo and coworker enzymatically removed sialic acid from pertuzumab using neuraminidase, and observed an

approximately five-fold increase in CDC and almost two-fold increase in ADCC (Luo et al., 2017). These increases were unexpectedly large, considering that the parent material was only ~2.5% sialylated; the most plausible explanation for this result was that these residual levels of sialylation potently inhibit CDC and ADCC and neuraminidase treatment relieves this inhibition.

In theory, as pertuzumab illustrates, the complete removal of sialic acid (and fucose (Luo et al., 2017)) offers a way to improve the efficacy of anti-cancer mAbs by facilitating CDC and ADCC. However, as a counterargument to this strategy, pertuzumab has a substantial number of deleterious side effects, including diarrhea or constipation, hair loss, loss of neutrophils and red blood cells, hypersensitive allergic reactions, decreased appetite, insomnia, distorted taste perception, inflammation of the mouth and lips, rashes, and muscle pain. Therefore, in practice, increasing the pro-inflammatory potency of this drug could exacerbate these side effects, reducing patient tolerance and overall clinical efficacy. The "take-home" lesson is that in principle it could be beneficial to glycoengineer anti-cancer antibodies to increase their pro-inflammatory activities; in practice, however, these efforts must be balanced by the danger of exacerbating off-target side effects. The ability to precisely tune the pro-vs. anti-inflammatory properties of IgG antibodies has been demonstrated using chemoezymatic synthesis; for example, homogeneous glycoforms of cetuximab with Fab N-glycans with two, sialylated antennae and Fc N-glycans with no fucosylation or sialyation have been created. The end result was an antibody with equal binding affinity to EGFR and increased affinity to FcγRIIIa, generating stronger ADCC (Giddens et al., 2018).

4.2 Anti-Inflammatory Antibodies

4.2.1 Mechanisms

As outlined above (**Section 4.1**), the role of fucose and sialic acid in the pro-inflammatory properties of therapeutic antibodies (Pereira et al., 2018)) are now well established. The flip-side to the necessary absence of both fucose and sialic acid for ADCC, ACDP, and CDC is that the presence of these sugars is beneficial–indeed, often required—for anti-inflammatory antibodies. For example, even residual levels of sialylation endow pertuzumab with potent anti-ADCC and anti-CDC properties (**Section 4.1.3** above). To quickly summarize the role of these two sugars [along with galactose, which has a more modest effect (Buettner et al., 2018)], they function as a tunable on/off switch where their presence turns on the anti-inflammatory properties of antibodies.

4.2.2 Immunoglobin G Therapy

Immunoglobulins from human donors are highly sialylated (from 20 to 60% site occupancy) compared to IgG antibodies produced in CHO cells (generally <2% and often <1%); accordingly, they have potent anti-inflammatory properties that can be attributed to their sialylation status (Li D. et al., 2021). As a result, polyclonal immunoglobulin provides a non-steroidal anti-inflammatory treatment safe for vulnerable patients, including children and pregnant women. More generally individuals with a broad range of autoimmune diseases including secondary hypogammaglobulinemia, recurrent infections, idiopathic thrombocytopenia purpura, Kawasaki disease, polyneuropathies, and graft versus host disease following organ transplantation (Barahona Alfonso & João 2016). Therapeutic immunoglobulin is typically administered intravenously as intravenous IgG (i.e., IVIg) therapy at up to 2 g/kg every few weeks to months (or, in rarer cases, subcutanoeus administration anti-inflammatory antibodies is achieved through co-delivery with hyaluronidase (Wasserman 2017)). With the continued growth of IgG therapy (Li D. et al., 2021), donor supply is projected to be insufficient, posing the quandary that CHO cell-produced recombinant IgG is poorly-sialylated (<2% overall and completely lacking in FcγR-modulating α2,6-sialic acids) and therefore lacking anti-inflammatory properties.

4.2.3 Anti-Inflammatory Monoclonal Antibodies

Intravenous immunoglobulin (IVIg) therapy, by using pooled samples from multiple donors contains immunosuppressive antibodies against numerous epitopes and is broadly anti-inflammatory. An alternative approach is the development of anti-inflammatory monoclonal antibodies against single epitopes for the treatment of non-cancerous indications. These efforts began over 30 years ago with the development of the anti-TNFα infliximab to treat rheumatoid arthritis (Semerano & Boissier 2009). Within the next two decades, several anti-inflammatory monoclonal antibodies have been approved to treat not only rheumatoid arthritis but also Crohn's disease, ulcerative colitis, spondyloarthropathies, juvenile arthritis, psoriasis, and psoriatic arthritis (Kotsovilis & Andreakos 2014). Indeed, four of the first five and the first seven of the first 10 FDA-approved mAbs were for anti-inflammatory indications (Lu et al., 2020). Although no longer as prolific as pro-inflammatory anti-cancer antibody drugs, anti-inflammatory monoclonal antibodies still comprise a substantial market share [e.g., including Orencia®, Humira®, Kineret®, Cimzia®, Enbrel®, Simponi®, and Remicade® (Kotsovilis & Andreakos 2014)]. The success of these drugs is exemplified by Humira®, which, in 2018, had a market value of US$ 19.9 billion (Lu et al., 2020). Up to now, the lucky happenstance that industry-standard CHO cell production systems provide monoclonal antibodies with anti-inflammatory properties due to high fucosylation and residual 1–2% sialylation levels has allowed clinical anti-inflammatory antibodies to be successful. In the future, we predict that deliberate efforts to increase the anti-inflammatory nature of these drugs, e.g., through increased sialylation (**Section 5.1**), will make these drugs even more effective.

4.3 Blocking/Neutralizing Antibodies

4.3.1 Mechanism(s)

In the body, the natural function of many antibodies is to have either pro- or anti-inflammatory activity (e.g., as discussed above in **Section 4.1** and **Section 4.2**, respectively); many other antibodies, however, have blocking and neutralizing action (e.g., HIV-neutralizing antibodies). Naturally-occurring neutralizing antibodies typically function by binding to a virus

or microbe, which can, at a minimum, negate the pathogen's infectivity, and ideally target it for immune destruction. These antibodies provide precedent to exploit this class of molecules to, in theory, bind to any receptor and block its activity. These neutralizing antibodies, also commonly referred to as blocking antibodies, are currently the largest class of clinical FDA-approved protein therapeutics; indeed, multiple blocking antibodies exist for PD1/PDL1 (7 FDA approved drugs), CD20 (6), TNF (4), HER2 (4), CGRP/CGRPR (4), IL-7/IL-6R (4), IL23 p19 (3), EGFR (3) and CD19 (3) (Mullard 2021).

4.3.2 Early Cancer Treatment Monoclonal Antibodies Were Blocking Antibodies

Immune checkpoint inhibitors are one of today's most exciting cancer immunotherapies, evidenced by the largest category (7 of 100) FDA-approved monoclonal antibodies being in this category (Mullard 2021). Several of the first immunotherapies, particularly for anticancer mAbs, also were blocking antibodies, including rituximab (1997, CD20); trastuzumab (1998, HER2); alemtuzumab (2001, CD52); and cetuximab (2004, EGFR) (Lu et al., 2020). Of these, cetuximab is a notable example from almost 20 years ago that alerted the pharmaceutical industry and regulatory agencies to the importance of glycosylation when the α-Gal epitope posed a major safety concern, as discussed below (**Section 5.2.1**). Interestingly, despite its early development, cetuximab remains one of the few commercial IgG mAbs that have a non-canonical Fab region N-glycan (Ayoub et al., 2013; Janin-Bussat et al., 2013). In the future as the role of Fab glycans in auto-antibody responses and auto-immune diseases become better defined (van de Bovenkamp et al., 2016; Van de Bovenkamp et al., 2018), we predict that commercial mAb development will revisit this category of mAbs.

4.3.3 HIV Neutralizing Antibodies

Human immunodeficiency virus 1 (HIV-1) remains an elusive and difficult-to-treat pathogen that causes acquired immunodeficiency syndrome (AIDS). The viral envelope's negligible immunogenicity is attributed to its host-derived glycan shield similar to SARS-CoV-2 and influenza (**Section 2.3.2**). Antibodies against the virus primarily target the envelope spike glycoprotein (Env), the only viral protein on the virus' surface, which is expressed in three form: gp120, gp140, and gp160 (Go et al., 2017; Heß et al., 2019; Seabright et al., 2019; Offersen et al., 2020; Wang et al., 2020). The Env protein is displayed sparsely on HIV-1, limiting the ability of antibodies to crosslink and elicit an immunogenic response to this virus. Nevertheless, certain individuals develop broadly neutralizing antibodies (bNAbs) against Env (Go et al., 2017; Seabright et al., 2019; Wang et al., 2020) that, although not providing a complete cure, do suppress most deleterious effects of HIV infection.

The capability of certain AIDS patients to produce bnAbs against HIV-1 spurred interest in mimicking these antibodies to produce effective vaccines. Engaging, or perhaps more precisely thwarting, glycosylation is critical for enhancing the immunogenicity of emerging HIV-1 vaccines. A longstanding difficulty in developing an effective bNAb vaccine is the notorious ability of HIV-1 to shift its glycosylation patterns (Wei et al., 2003), generating entirely new profiles in response to the adaptive immune response (Go et al., 2017; Offersen et al., 2020; Wang et al., 2020); a well-known example involves the N334 position on the Env protein (Seabright et al., 2019). As a counterpoint, bNAbs to Env function by recognizing glycosylation patterns that are conserved across clades of viral proteins, including atypical oligomannose structures (Seabright et al., 2019; Wang et al., 2020). Recent studies have focused on determining highly conserved glycoprofiles across viral strains, metabolic activities, and cell types (Wang et al., 2020) to facilitate bNAb-inducing HIV-1 vaccine development.

The previous two paragraphs laid out challenges facing natural immunity to HIV-1 infection, many of which result from viral glycosylation. To turn the tables on the virus, intriguing glycoengineering strategies have been directed towards treating AIDS. In one pioneering effort, Song and coworkers describe how the addition of an N-glycan to the HIV neutralizing antibody ibalizumab (Trogarzo®) improves its efficacy (Song et al., 2013). The added N-glycan helps fill "empty space" between the antibody and viral epitope, thereby increasing the binding interface and affinity. In this groundbreaking study, the glycan was limited a the Glc_2Man_5 structure (**Figure 1B**); in the future, follow-on glycoengineering efforts can further facilitate ibalizumab-Env binding interactions, resulting in even more potent neutralizing antibodies. (Strategies for attaining improved glycoforms towards these objectives are provided in **Section 5** of this report.)

4.4 Antibody-Drug Conjugates

Antibodies are attractive drug delivery vehicles because their binding specificity allows them to deliver payloads with minimal off-target toxicity. As such, a variety of methods have evolved to directly link a drug of interest to an antibody, thus forming antibody-drug conjugates (ADCs). Conventional chemical conjugation of drug payloads typically utilize the amines of lysine or thiols of cysteine residues present in the amino acid sequence of the antibody (Qasba 2015; Tang et al., 2019). This approach results in heterogeneous ADCs with greater susceptibility to aggregation, decreased antibody stability, or cytotoxicity that together pose barriers to effective clinical use and increase regulatory scrutiny.

These pitfalls have spurred researchers to create active, homogenous ADC populations with one such class of these drugs known as glycosite-specific ADCs (gsADCs) (Tang et al., 2019). These glycoengineering strategies take advantage of the conserved, biantennary N-glycosylation site present at the asparagine 297 residue of the CH2 regions of the Fc domain. One strategy uses metabolic glycoengineering to install thiol-modified fucose in Fc domain glycans (**Figure 6A**), which can be used as a chemical handle for drug conjugation (**Figure 5A**, (Okeley et al., 2013)). Another chemical method for site-specific chemical conjugation to Fc glycans involves mild periodate oxidation (Jourdian et al., 1971; Peters & Aronson Jr 1976), which selectively introduces aldehyde groups into sialic acids (**Figure 5B**); a downside of this approach is the low sialylation of Fc glycans, often 2% or lower. A strategy using non-natural

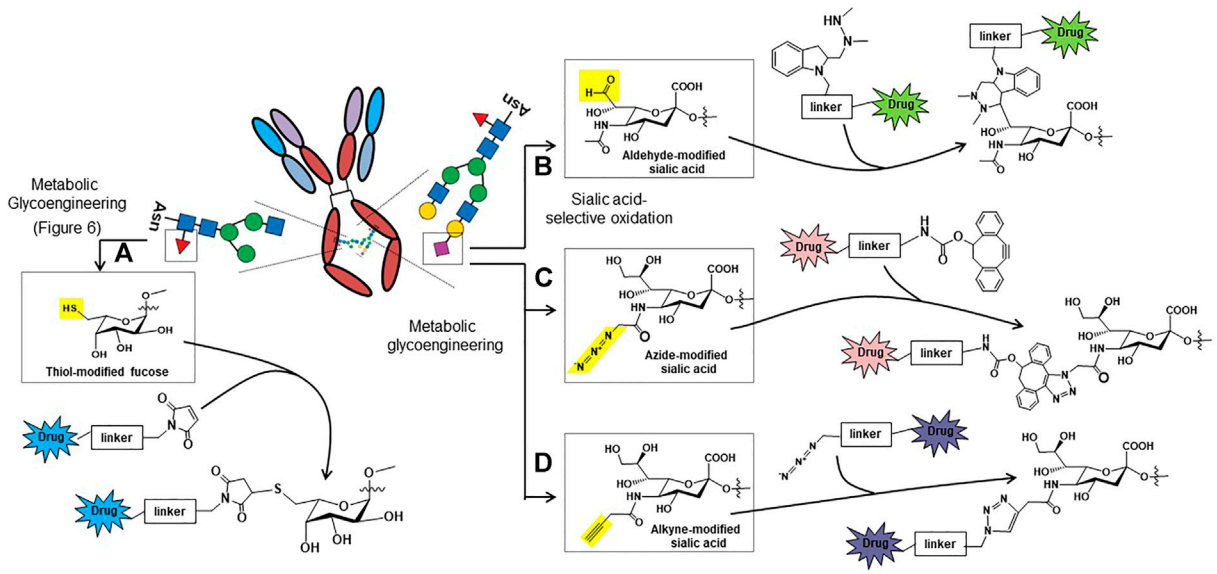


FIGURE 5 | Glycosylation-based antibody-drug conjugate (ADC) ligation strategies based on chemically-modified fucose (A) or sialic acid (B, C, and D). **(A)** Thiols can be installed into non-natural fucose using metabolic glycoengineering and used as "chemical handles" to ligate drug molecules to the Fc domain glycans of antibodies using thiol-reactive maleimides (Okeley et al., 2013). **(B)** Aldehydes can be selectively introduced into sialic acids by oxidizing the C8-OH groups; the aldehyde then can be conjugated to drugs using the hydrazino-*iso*-Pictet-Spengler (HIPS) reaction (Drake et al., 2014). **(C)** Metabolic glycoengineering can be used to install azido-sialic acids into glycans (Saxon & Bertozzi 2000), which can then be used to conjugate drugs to the antibody using dibenzocyclooctyne (DIBO) conjugation reactions (Li et al., 2014). **(D)** Alkyne groups can also be introduced into sialic acids through metabolic glycoengineering, which can then be conjugated using conventional copper catalyzed click chemistry (Du et al., 2009; Hong et al., 2010).

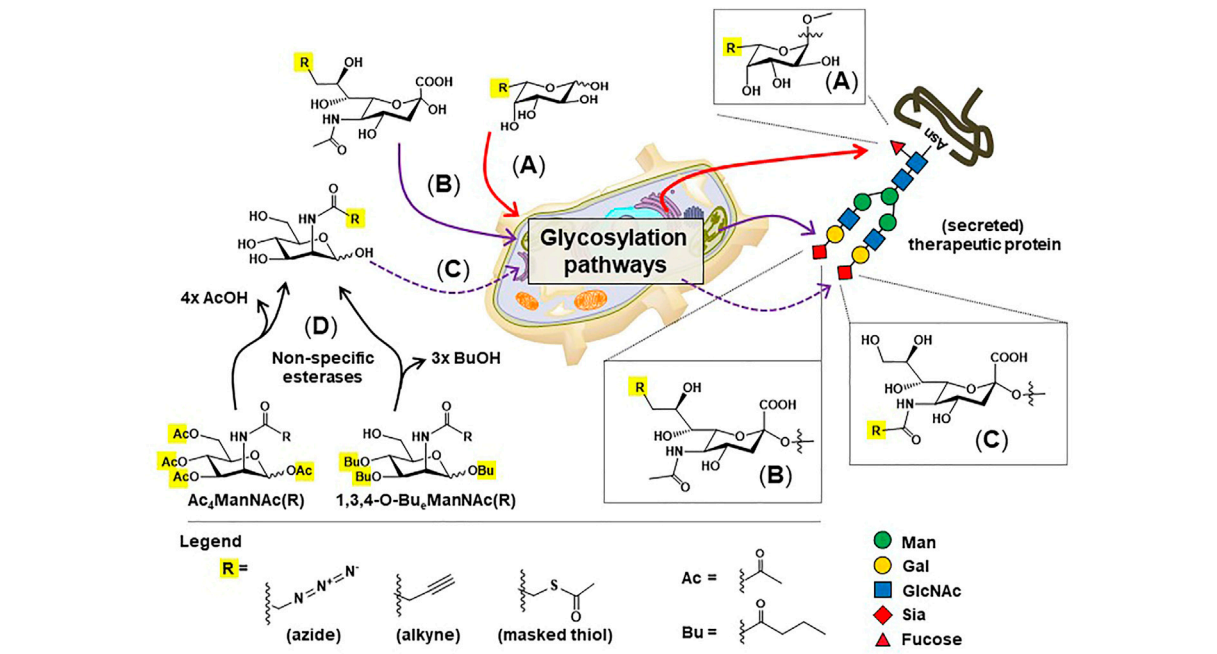


FIGURE 6 | Overview of metabolic glycoengineering (MGE). Non-natural monosaccharide analogs capable of installing "chemical handles" into the N-glycans of therapeutic proteins include: **(A)** C6-modified fucose **(B)** C9-modified sialic acids, and **(C)** C2-modified ManNAc analogs, which are converted to N-acyl (C5) modified sialic acids before installation into N-glycans. **(D)** "High-flux" esterase-protected ManNAc analogs are now widely employed in MGE experiments to increase cell uptake and reduce the concentrations required for media supplementation from 30 to 75 mM (Yarema et al., 1998) to 100 μM or less (Jones et al., 2004; Kim et al., 2004; Almaraz et al., 2012).

ManNAc analogs to increase flux through the sialic acid pathway (**Figure 6**) and simultaneously install bioorthogonal chemical functional groups such as azides (**Figure 5C**) and alkynes (**Figure 5D**) provides additional options to create gdADCs.

5 CONSIDERATIONS FOR THE DESIGN AND PRODUCTION OF GLYCOENGINEERED THERAPEUTIC PROTEINS

So far, this report outlined various ways that glycosylation controls the pharmacokinetics, pharmacodynamics, and overall clinical efficacy of therapeutic proteins. Knowing this, biomedical researchers and the pharmaceutical industry are increasingly aware of the importance of controlling the glycosylation of therapeutic proteins using glycoengineering strategies summarized in **Section 5.1**. Ultimately, the production of glycoengineered proteins depends on glycocompatible production systems, which today are focused on CHO cell biomanufacturing (**Section 5.2**. Finally, arguments that the current industry-standard CHO cell platform is stifling innovation, especially with respect to glycosylation, are leading to the development of alternative cell-based production platforms (**Section 5.3**)

5.1 Glycoengineering–Methods and Approaches

5.1.1 Glycoengineering of Proteins Isolated From Natural Sources to Increase Their Effectiveness

The clinical use of therapeutic proteins pre-dated today's recombinant protein production technologies with early generations of these drugs obtained from natural sources; insulin is a well-known example initially derived from bovine and porcine pancreases. Additional examples from the current report include hyaluronidase obtained from mammalian sperm, β-glucocerebrosidase isolated from human placenta (Deegan & Cox 2012), blood coagulation and clotting factors obtained from human plasma, and FSH prepared from human urine from postmenopausal women. In some cases, exemplified by β-glucocerebrosidase, glycoengineering was a critical enabling technology to turn this enzyme into a useful drug by installing high mannose N-glycans (**Figure 1B**) that enabled macrophage uptake to treat GD. In other cases, illustrated by FSH, the complex role of glycosylation is still being unraveled. For example, certain glycoforms can be beneficial for PD properties while detrimental for PK properties and vice versa; once a fuller understanding is in hand, glycoengineering strategies can be applied to improve this type of therapeutic protein. There is even evidence that the few non-glycosylated therapeutic proteins can benefit from glycoengineering, for example, insulin with three newly-added N-glycans has improved resistance to proteases, potentially opening the door to oral dosing (Guan et al., 2018).

Chemoenzymatic synthesis, which combines chemical synthesis with the use of enzymes such as glycosyltransferases, glycosidases, lipases, and glycosynthases, is a powerful method for the synthesis of complex glycans and glycoproteins (Muthana et al., 2009; Wang et al., 2019; Ma et al., 2020b; Zeng et al., 2022). In addition to building new glycans, chemoenzymatic methods can be used to remodel glycans on antibodies and other glycoproteins, thus improving glycoform homogeneity (Wang et al., 2019). Additionally, this method removes the need for protection and deprotection of peptides that occurs in purely chemical synthesis (Zeng et al., 2022). Because this approach combines both the selectivity of enzymatic reactions and the flexibility of chemical glycan synthesis, it provides a facile method for the synthesis of complex polysaccharides, heparin sulfates, glycoproteins and glycolipids that are difficult to synthesize homogeneously *via* other methods (Muthana et al., 2009). For example, chemoenzymatic glycan remodeling of IgG antibodies can be employed to produce glycosite-specific antibody-drug conjugates (Zeng et al., 2022).

5.1.2 Cell-Free Methods to Modulate Glycosylation: IVIg Therapy as a "Case Study"

Immunoglobulin used in IVIg therapy illustrates how isolation of therapeutic proteins from natural sources (e.g., human blood donors) is a cumbersome and inefficient process. Depending on the manufacturer, 1,000 to 100,000 donor samples are pooled to purify and concentrate IgG to 50–100 mg/ml with preparations typically still containing residual levels of IgE, IgM, and IgA antibodies at ≤ 700 μg/ml. The pooled samples are screened for viral contamination (Hep B, Hep C, and HIV) and monitored for conformance to an acceptable titer of ABO blood type-recognizing antibodies to reduce risk of hemolytic reactions in the recipients (Barahona Alfonso & João 2016). At the end of this cumbersome process, sialylated IgG antibodies can be as low as 15% of the total, resulting in less-than-optimal anti-inflammatory potential.

The enticing prospect of increasing the potency of immunoglobulin therapy by enhancing the α2,6-sialylation (**Figure 4A**) of donor IgG has been apparent for ~15 years (Kaneko et al., 2006). Li and coauthors summarize several attempts to increase sialylation (Li D. et al., 2021), one of the first efforts involved the use of *Sambucus nigra* agglutinin (SNA) affinity chromatography to prepare IVIg to treat rheumatoid arthritis (Sudo et al., 2014). Taking a different approach, chemoenzymatic strategies to improve IgG sialylation incubate pooled IgG samples with α2,6-sialyltransferase in the presence of CMP-sialic acid (the enzyme's co-substrate). In some cases, the IgG is pretreated with neuraminidase to remove non-inflammatory α2,3-sialic acids and, in other cases, the sialylation reaction is done in the presence of β1,4-galactosyltransferase and UDP-galactose to install the penultimate galactose required for terminal sialylation (Anthony et al., 2008; Washburn et al., 2015; Bartsch et al., 2018). Another enticing approach, pioneered by Lai-Xi Wang's group, is to use transglycosidases to remove existing Fc domain N-glycans and enzymatically replace them with homogenously sialylated glycans (Li et al., 2017; Giddens et al., 2018; Wang et al., 2019).

These post-production glycoengineering strategies have successfully improved the efficacy and potency of immunoglobulin therapy; for example, a 0.1 g/kg dose of SNA-

enriched IVIg is as effective as 1.0 g/kg of unfractionated drug (Kaneko et al., 2006). A major pitfall, however, is that these methods can only be performed on the milligram to Gram scales (or optimistically, on a kilogram scale) based on the expense of the lectins, glycosyltransferases, and nucleotide sugar donors involved (Li D. et al., 2021). Considering that worldwide consumption of IVIg is over 100 tons per year, post-production chemoenzymatic glycoengineering strategies remain niche technologies not yet applicable to large scale preparation of this drug. This case study illustrates the need for versatile cell-based production systems for manufacturing of glycoengineered protein therapeutics where cells produce expensive reagents such as glycosyltransferases and nucleotide sugar donors essentially "for free" (i.e., they are produced by cellular metabolism).

5.1.3 Cell-Based Production of Recombinant Glycoproteins

With a few notable exceptions (e.g., IgG antibodies for IVIg therapy that are isolated and purified from natural sources), today's therapeutic proteins are produced in cell-based systems. Production in living cells became an option with the maturation of DNA cloning technologies in the late 1970s and early 1980s that enabled recombinant techniques for protein expression. Benefits for cell-based production of recombinant proteins are numerous including theoretically limitless supplies of the therapeutic, the ability to humanize products by altering the amino acid sequence to avoid immunogenicity and increase productivity, easier purification, and the avoidance of potential pathogens and immunogens from non-human sources. Equally important and most germane to this article, cell-based production systems can be customized to provide beneficial glycosylation patterns as discussed in detail in **Section 5.2** for CHO cells and in **Section 5.3** for emerging alternative production systems.

5.1.4 "Building in" N-Glycan Sites

Natural N-glycosylation machinery recognizes a consensus sequon (Asn–X–Ser/Thr, where X is any amino acid except proline), and initiates glycosylation with the addition of the LLO 14-mer (**Figure 1A**, Step 1) to the nitrogen atom of the asparagine side chain. In theory, the installation of new N-linked glycans into a protein of interest can be achieved by introducing amino acid substitutions that yield this sequon. In practice, this sequon is a necessary, but not sufficient, condition for successful N-glycosylation because, for example, the built-in glycan must not interfere with protein folding. Even if a target protein is successfully glycosylated, the required amino acid substitution(s) or neoglycan may lead to structural alterations that deleteriously affect PK, PD, or therapeutic efficacy. In the past, efforts to add N-glycan sites to therapeutic proteins used a trial-and-error process. For example, two decades ago when darbepoetin alfa was designed, several dozen variants of recombinant human EPO containing one or more new sites for N-glycan attachment were evaluated (Elliot et al., 2000; Egrie & Browne 2002). More recent approaches for glycosylation site installation combine structural information with rational and computational design approaches to more efficiently design functional and efficacious constructs.

To design new glycosylation sites, a script with a sliding window evaluation of every amino acid triplet can be employed to identify all possible sites for insertion of an N-glycan by modifying existing amino acid sequences to the Asn–X–Ser/Thr consensus sequence. This method quickly identifies single and double amino acid substitutions that yield potential sites for N-linked glycosylation. Ideally, the sequence change should be minimal (i.e., a single amino acid mutation is ideal), to offer the highest probability that the protein remains functional. Once potential sites for N-glycans have been identified, further *in silico* evaluation can help guide specific glycovariants to be made experimentally. Online tools such as the NetNGlyc Server, an N-linked glycosylation prediction site (Gupta & Brunak 2002), can be used to estimate the likelihood that each of the possible engineered glycosylation sites will be successfully glycosylated. Sites with low likelihood of glycosylation (<0.55) can be disregarded before proceeding; in our experience, most sites with prediction frequencies of >0.70 or more are successfully glycosylated (Saeui et al., 2020).

Using structure design tools, such as the PyMOL mutagenesis wizard or the Rosetta software package, each neoglycosylation site within a glycoengineered protein can be modeled to ensure that desired features of the protein structure are maintained. First, the glycosylation site should be solvent-exposed and not be buried within the interior of the protein. Second, the glycosylation site should be positioned to avoid steric interference of attached glycans with important domains of the protein. For example, if the therapeutic protein is an enzyme, the glycan should not interfere with substrate access to its active site; this was a design feature for ENPP1-Fc (**Figure 3**), where built-in glycans did not comprise substrate binding or catalysis (Stabach et al., 2021). If the protein is a cytokine, hormone, growth factor, or antibody, the glycan should not interfere with the therapeutic protein's binding to partner proteins. In certain cases, glycan-based steric factors can be advantageous to the protein's function. For example, the increased size resulting from installed glycans in darbepoetin alfa leads to decreased kidney filtration and extended pharmacokinetic half-life (**Section 2.1.1**). As another design feature, also considered for glycoengineered ENPP1-Fc, novel N-glycans can be situated to block protease access to vulnerable surfaces of the enzyme (Stabach et al., 2021). Finally, the addition of new glycans can improve binding affinity through their introduced ionic, van der Waals, or entropic forces as exemplified by improved affinity of an HIV-neutralizing IgG antibody to gp120 upon addition of a non-canonical glycan (Song et al., 2013). Regardless of the glycoengineering objective, candidate proteins must be individually evaluated to ensure that their functional activity is as desired.

The use of *in silico* tools in combination with structural information can be used to rationally design N-linked glycosylation cites with the goal of maintaining, or even enhancing, the activity of the target protein. In cases where a solved structure is unavailable, *in silico* structure prediction tools can be leveraged to generate theoretical protein structures and guide design of theoretical N-linked glycosylation sites. Various computational tools have emerged to generate protein structures using homology-based and/or *de novo* modeling in place of

directly resolving the protein structure (Kuhlman & Bradley 2019; Jumper et al., 2021; Kryshtafovych et al., 2021). Furthermore, recent advances in modeling glycans themselves can be incorporated into the design process, to provide additional information about how the structure and activity of a protein may be impacted by the glycans themselves (Labonte et al., 2016; Li M. et al., 2021). In summary, the ability to predict both protein and glycan structure using computational tools empowers many glycoengineering approaches where structural information is lacking.

5.1.5 Metabolic Glycoengineering: Further Control of Glycan Chemistry

Metabolic glycoengineering (MGE, **Figure 6**) is a method pioneered approximately 40 years ago when Brossmer and others developed chemically-modified sialic acid analogs, including bulky moieties such as fluorophores, that could be enzymatically installed into glycans (Gross & Brossmer 1988; Gross et al., 1989; Gross & Brossmer 1995). Subsequent advances in the 1990s and 2000s include the Reutter group's demonstration of MGE in living cells and animals (Kayser et al., 1992b; Keppler et al., 2001; Wratil et al., 2016); the Bertozzi group's development of analogs with chemical functionalities not normally found on cells, thereby allowing bioorthogonal chemoselective ligation reactions (Mahal et al., 1997; Saxon & Bertozzi 2000; Prescher et al., 2003); the extension of MGE to biosynthetic pathways beyond sialic acid including fucose (Sawa et al., 2006; Okeley et al., 2013), GlcNAc (Vocadlo et al., 2003; Du et al., 2009), and GalNAc (Kayser et al., 1992a; Boyce et al., 2011); as well as efforts to incorporate high-reactivity chemoselective reaction partners including ring-strained cyclooctynes (Baskin et al., 2007; Ning et al., 2010) and tri- or tetrazines (Kamber et al., 2019; Agatemor et al., 2020).

Today, MGE technologies have matured to the point where they comprise an attractive toolkit for cancer treatment (Agatemor et al., 2019; Wang & Mooney 2020), and increasingly, for other conditions such as enhancement of neuronal differentiation for spinal cord and brain regeneration (Sampathkumar et al., 2006; Du et al., 2021; Du et al., 2022) Specific to therapeutic proteins, MGE can be used in various ways. For example, MGE can be used to endow antibodies with "chemical handles" into antibodies by replacing core fucose with thiol-modified residues and terminal sialic acids with their azido-modified counterparts (**Section 4.4**; **Figure 5**). In theory, introduction of non-natural sialic acids into IgG Fc domain glycans can achieve an antibody-to-drug ratio of four if both glycans are fully sialylated, biantennary structures; in practice, however, the low site occupancy of sialic acid on Fc domain glycans hinders the use production of high valency ADCs.

A variation of MGE can help overcome suboptimal levels of sialic acid by improving the sialylation of IgG Fc domain glycans and therapeutic proteins in general. Briefly, "high-flux" MGE analogs began with peracetylation where the ester-linked acetyl groups rendered the sugars more lipophilic, facilitating diffusion into cells (Lemieux et al., 1999; Sarkar et al., 1995; Yarema et al., 2001). Upon entry into a cell, non-specific esterases remove the acetate groups, allowing the "core" monosaccharide to enter its targeted biosynthetic pathway (Mathew et al., 2012; Wang et al., 2009). Our team discovered that tri-butanoylated hexosamines, exemplified by 1,3,4-O-Bu_3ManNAc (**Figure 6D**), provide even higher flux into biosynthetic pathways, increasing sialylation with high efficiency (Aich et al., 2008; Almaraz et al., 2012). This analog increases the sialylation of therapeutic proteins including IgG antibodies (Yin et al., 2017), EPO (Mertz et al., 2020) and ENPP-1 (Stabach et al., 2021). In the case of ENPP1-Fc, production with 1,3,4-O-Bu_3ManNAc increased serum half-life from 170 to 204 h and the AUC from 37,000 to 45,000 (**Figures 3E,F**).

5.2 Current Therapeutic Protein Biomanufacturing Overwhelmingly Uses CHO Cells

Chinese hamster ovary (CHO) cells have become the workhorse biomanufacturing platform for therapeutic proteins over the past 2 decades. Because of the importance of these cells, we describe their safety qualifications (**Section 5.2.1**), limitations (**Section 5.2.2**), and efforts towards overcoming these pitfalls by using genetically modified CHO cell variants with altered glycosylation capacities (**Section 5.2.3**).

5.2.1 Safety Issues–Exemplified by the α-Gal Trisaccharide Immunogenic Epitope

Chinese hamster ovary cells have become the "go-to" cell line for biomanufacturing therapeutic proteins for several reasons, including efficiency, cost-effectiveness, and—historically—for safety reasons. Historically, CHO cells have been used for recombinant protein production since the 1980s based on several advantages, including their ability to produce relatively large amounts of glycoproteins, their lack of human pathogens, and their ability to approximately replicate human glycosylation patterns (Ma et al., 2020a). Over the past decade or so, production has coalesced around CHO cells for safety/regulatory reasons after pioneering anticancer antibodies severely harmed patients in early clinical testing. In particular, in 2004 cetuximab (Erbitux®)—a blocking antibody that inhibits the epidermal growth factor receptor (EGFR) and is used to treat metastatic colorectal cancer and head and neck cancer -- triggered anaphylaxis in cancer patients, resulting in several deaths (Friedman 2008). The affected patients had pre-existing IgE antibodies against galactose-α-1,3-galactose (i.e., "α-Gal" **Figure 4B**) generated by lone star tick bites; subsequent anaphylaxis was elicited by the presence of α-Gal on Erbitux® produced in murine SP2/0 cells (Steinke et al., 2015). This incident raised awareness that CHO cells, which do not make α-Gal are safe host cells for biomanufacturing of therapeutic proteins helping these cells gain widespread regulatory acceptance.

5.2.2 Limitations/Drawbacks of CHO Cells

Chinese hamster ovary cells have glycosylation patterns that are generally regarded as safe (i.e., they lack the hyper-immunogenic α-Gal epitope) but they do have drawbacks. For example, they lack α2,6-sialyltransferase activity, making them inappropriate

production hosts for potently anti-inflammatory antibodies. Another pitfall is that CHO cells produce the Neu5Gc form of sialic acid (**Figure 4B**) (Hokke et al., 1990); although only weakly immunogenic, its presence in therapeutic proteins has raised caution (Ghaderi et al., 2010; Ghaderi et al., 2012). Despite these shortcomings, CHO cells currently produce ~90% of therapeutic antibodies including virtually all newly-approved drugs. One reason why CHO cells are dominant is because of their acceptance by regulatory agencies, which can be regarded as a positive feature but also has its drawbacks. For example, quoting from Burnett and Burnett (Burnett & Burnett 2020):

"As promising as technology may be, drug companies are unwilling to risk the huge sums of money required to get a new product approved by the large drug approval administrations (e.g., the FDA or EMA) if there is already a proven alternative expression system with regulatory approval. This economic constraint has a stagnating effect on the pharmaceutical industry, limiting the scale of progress and development of new drug production technologies."

Despite the stifling influence of regulatory agencies that have helped embed CHO cells as the go-to cell line for biomanufacturing, efforts continue to develop alternative production platforms. These efforts are not primarily driven by glycoengineering concerns but they often represent substantial departures from standard glycosylation patterns inherently produced by CHO cells. As such they face regulatory hurdles but also provide opportunities to tune glycosylation to improve the efficacy of therapeutic proteins.

5.2.3 Genetically Modified CHO Cells

Before describing major departures from CHO cells (e.g., the use of bacteria, plant cells, yeast, and insect cells for biomanufacturing, **Section 5.3**), we cover "baby steps" being taken to rectify glycosylation deficits in CHO cells, or more positively, to endow them with enhanced glycosylation capabilities. Mammalian cells have 250 or more glycogenes, the majority are glycosyltransferases present in the Golgi, the last stage of glycan production (Tariq et al., 2018). The potential for genetic control of glycosylation in CHO cells was demonstrated almost 30 years ago by a library of lectin-selected, mutant sublines developed by Pamela Stanley's research group (Stanley et al., 1996; Stanley & Patnaik 2005).

Today, advances in nucleic acid gene-editing techniques including zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeats with Cas9 protein (CRISPR/Cas9) facilitate precise, stable, and systematic engineering of the glycosylation capabilities of mammalian cells (Narimatsu et al., 2021; Wang et al., 2019). One example is the over-expression of α2,6-sialyltransferase (ST6) in non-human cell lines such as CHO cells (Yin et al., 2015), which have been used to produce EPO and IgG antibodies (Mertz et al., 2020; Yin et al., 2015) as well as ENPP1-Fc with improved sialylation and PK properties [**Figure 3D**, (Stabach et al., 2021)]. ST6 over-expression increases overall sialylation and results in a humanized α2,6-sialylation profile in CHO cells. In addition to over-expression of glycogenes to improve CHO cells as production hosts for therapeutic proteins, it can be advantageous to knock out other glycogenes. Indeed, the first glycogene KO'd for biomanufacturing involved a tour-de-force effort in CHO cells where two rounds of targeted homologous recombination ablated the two allelic copies of the α6-fucosyltransferase (Fut8) gene (Narimatsu et al., 2021). These efforts have reached fruition with several afucosylated therapeutic antibodies now in clinical use (**Section 4.1.2**).

5.3 Additional Cell-Based Options for Biomanufacturing Therapeutic Proteins

The limitations of CHO cells for biomanufacturing (**Section 5.2.1**) have kept alive efforts to develop additional cell lines as production platforms. Here in **Section 5.3** we describe cell systems used to produce at least one, and often several, FDA-approved therapeutic proteins; each is discussed briefly providing a perspective on the system's current use and future prospects with an emphasis each system's glycoengineering capabilities.

5.3.1 Human Cells

By definition, production of therapeutic proteins in human cells provides the drugs with humanized glycosylation, including features such as α2,6-sialylation lacking in CHO cells and, unlike mouse cells, a lack of α-Gal that enhances safety. Downsides of production include the high cost and potential safety concerns of animal products used in production (e.g., fetal bovine serum is generally required for the culture of human cells, opening the door to xenopathogen contamination) to low productivity (typically one to ~100 s mg/L) (Dumont et al., 2016). Nonetheless there are five FDA approved therapeutic proteins produced in human cells: Idursulfase (Hunter syndrome, approved 2006), Velaglucerase alfa (Type 1 Gaucher disease, approved 2010), rFVIIFc (Hemophilia A, approved in 2014), rFIXFc (Hemophilia B, approved 2014), and Dulaglutide (Type 2 diabetes, approved 2014). The lag in the approval of new products over the past several years, however, suggests that production of therapeutic proteins in human cells remains an infrequently used, niche strategy.

5.3.2 Murine Cells

Expression of FDA-approved therapeutic proteins in murine cells began in the 1990s (Lifely et al., 1995) and continues today despite safety issues including the anaphylaxis-inducing α-Gal epitope (**Figure 4B**) and high levels of the mildly immunogenic Neu5Gc form of sialic acid (**Figure 4A**) (Lalonde & Durocher 2017). Although immunogenicity concerns remain for these drugs, safety risks are minimized by pre-screening patients for IgE anti-α-Gal antibodies linked to anaphylaxis, allowing murine-produced mAbs to remain on the market. Mouse myeloma NS0 and Sp2/0 lines are used in biomanufacturing, produceing cetuximab (Erbitux® mentioned earlier) and several mAbs approved up to ~2015, including palivizumab (Synagis®), dinutuximab (Unituxin®), necitumumab (Portrazza®), and elotuzumab (Empliciti®). Similar to human cells, the negligible approval of new products in the past few years suggests that murine cells are

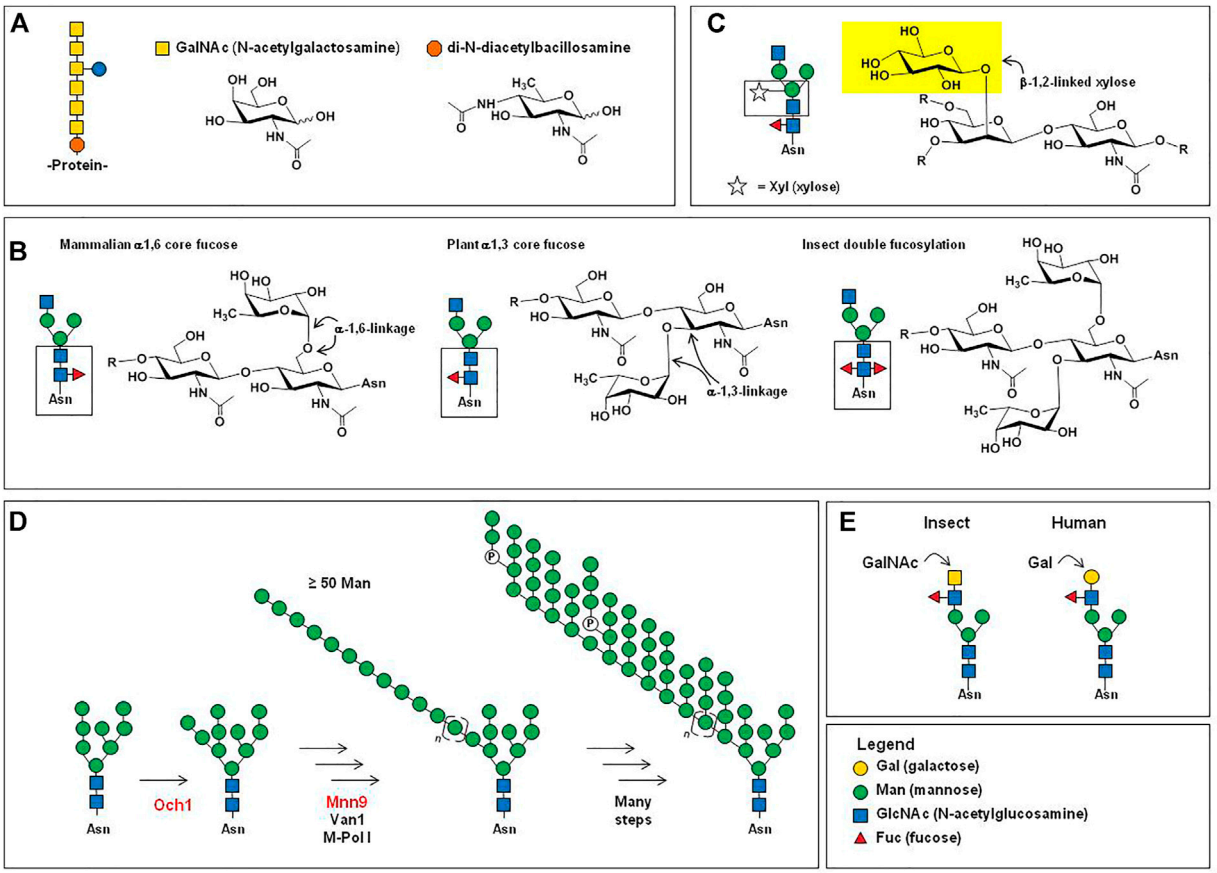


FIGURE 7 | Glycoforms of concern in bacterial, plant, fungal, and insect production systems. **(A)** Efforts to produce glycosylated recombination proteins in bacteria (**Section 5.3.3**) have resulted in the non-human glycan structure shown. **(B)** Mammalian N-glycans have α1,6-linked core fucose (right), which along with sialic acid, endow IgG antibodies with anti-inflammatory properties; plant cells (**Section 5.3.4**) produce N-glycan with α1,6-core fucose (center), and insect cells (**Section 5.3.6**) produce doubly-fucosylated N-glycans (right). **(C)** Xylose, a monosaccharide not present in mammalian N-glycans, is added to plant-produced N-glycans (**Section 5.3.4**). **(D)** Mannan synthesis in fungi (**Section 5.3.5**). **(E)** GalNAc incorporation in insects as compared to human galactose addition (**Section 5.3.6**).

unlikely to figure prominently in future biomanufacturing efforts.

5.3.3 Bacteria

During the early development of recombinant DNA technologies in the 1970s, efforts were focused on producing products in bacteria; for example, *Escherichia coli*, was an attractive low-cost, high-yield (e.g., ~8 g per liter (Menacho-Melgar et al., 2020)) production host. Despite some issues, such as the challenge of purifying recombinant mammalian proteins from bacterial cell components (e.g., the cell wall) and the possibility of endotoxin contaminants (Singh et al., 2016), there were several successfully-produced therapeutic proteins in *E. coli* in the 1980s. These including Humulin® (a recombinant form of insulin), Protropin® and Humatrope® (to treat hGH deficiency), Roferon A® (to treat hairy cell leukemia), and IntronA® (to treat genital warts and hepatitis) (Sanchez-Garcia et al., 2016). In retrospect, the inability of *E. coli* to N-glycosylate proteins, and therefore to take advantage of the concomitant protein folding chaperone system in mammalian cells (Helenius & Aebi 2001), posed a significant challenge during early attempts to express large, difficult-to-fold mammalian proteins in *E. coli* that more often than not resulted in inclusion body formation.

Unlike human and murine cells, which suffer from fundamental limitations for biomanufacturing (e.g., low product yield safety concerns, and high cost), a major detriment of bacterial production systems is their lack of mammalian-type glycosylation. In theory, this pitfall can be overcome, at least in part, by building N-glycosylation capabilities into *E. coli* used for recombinant protein production (Wayman et al., 2019). For example, the protein glycosylation pathway of *Campylobacter jejuni*, a pathogenic bacterium (Szymanski et al., 2002), has been transferred into laboratory strains of *E. coli* (Pandhal & Wright 2010; Wacker et al., 2002). The resulting glycans, however, are distinctly different than human N-glycans (**Figure 7A**) (Abu-Qarn et al., 2008). For example, they are mainly comprised of GalNAc, a mammalian monosaccharide that does not normally appear in mammalian N-glycans. Similarly, the presence of glucose is unusual for mammalian N-glycans, where this monosaccharide

appears in the LLO 14mer precursor structure (**Figure 1A**) but not in mature N-glycans found on glycoconjugates. Finally, di-N-diacetylbacillosamine is a prokaryotic monosaccharide not found in eukaryotes. Despite technologies to install N-glycosylation pathways in bacteria being in their nascent stages, proof-of-principle experiments (**Figure 7A**) coupled with continuing robust efforts to improve prokaryote glycosylation (Ding et al., 2019; Wayman et al., 2019; Pratama et al., 2021; Yates et al., 2021) provide hope that in the future, additional therapeutic proteins will be manufactured in bacterial hosts.

5.3.4 Plant Cells

Therapeutic protein production in plants has several advantages, including that the infrastructure for large scale production of crops is in place and in theory requires only water, sunlight, and cheap fertilizers (Burnett & Burnett 2020; Karki et al., 2021). In practice, current plant-based manufacturing uses cell-based methodology rather than field-grown crops but retains advantages over mammalian cell culture. These advantages include a lack of animal products needed for plant cell culture that reduces the risk of viral contamination; the ability to grow cells in inexpensive polyethylene bags rather than stainless steel bioreactors; and room temperature manufacturing without the need for strict temperature control. Counteracting these advantages are the low productivity of plants (e.g., ~100 mg/kg of plant mass) and the complexity of purifying human proteins from plant matter where cell wall components pose a challenge (Singh et al., 2016; Schillberg et al., 2019).

The use of plant cells for therapeutic protein production is in its infancy, with only one FDA-approved drug. This drug, Elelyso® (i.e., taliglucerase alfa, mentioned above in **Section 2.3.3** as a treatment for GD), was approved in 2012 (Fox 2012). Elelyso® illustrates glycosylation differences between plant N-glycans and their mammalian counterparts; for example, the plant ProCellEx® platform directly produces N-glycans with exposed terminal mannose residues (e.g., Glc_2Man_3 to Glc_2Man_5 structures, **Figure 1B**) needed for macrophage targeting and uptake (Tekoah et al., 2013). The direct production of MR-targeting glycans in plant cells offers simplicity and cost savings compared to the use of chemical modulators of glycosylation (required for velaglucerase alfa production) or enzymatic modification (imiglucerase) as described in **Section 2.3.3**.

On a cautionary note, plant cells produce core structures with α1,3-linked fucose not found in humans (human α1,6- vs. plant α1,3-core fucosylation is shown in **Figure 7B**) and β1,2-linked xylose (**Figure 7C**) not found in mammalian proteins (Castilho & Steinkellner 2012; Montero-Morales & Steinkellner 2018). Initially, concerns were raised that these non-human glycoforms could be immunogenic in a mildly harmful way reminiscent of Neu5Gc, or possibly with the severe effects of α-Gal. Fortunately, only a small fraction of patients had pre-existing antibodies that recognized these glycans on Elelyso®, and those that did experienced no adverse effects (Rup et al., 2017). By contrast, glycan-based immunogenicity of plant-produced blood coagulation factors VIII and XIII remains a substantial impediment to the commercialization of these hemophilia drugs (Top et al., 2019). In some cases, instead of being harmful, the potential immunogenicity of plant glycans has been proposed to enhance cancer vaccines and cancer immunotherapeutics through lectin-based stimulation of antigen-presenting cells (Rosales-Mendoza et al., 2015). Overall, similar to bacterial systems where improved cell hosts are actively being pursued to improve glycosylation, glycoengineering efforts remain underway in plants (Sukenik et al., 2018; Fischer et al., 2021), opening the door for increased use of plants for therapeutic protein production.

5.3.5 Fungi

Several yeast strains, including the widely used production hosts *Saccharomyces cerevisiae* and *Pichia pastoris,* are generally recognized as safe (GRAS) by regulatory agencies. Advantages to fungal production include high yield (up to 12 g/L); cost and safety advantages by avoiding the use of animal products such as FBS during production; and sidestepping the danger of endotoxins from *E. coli* production. A glycosylation-related drawback is the production of hypermannosylated (mannan) N-glycans in yeast that can contain dozens to hundreds of mannose residues [**Figure 7D** (Orlean 2012)]. These extremely large mannose structures clearly are incompatible with therapeutic glycans. Fortunately, a straightforward solution was found by knocking out two early genes (*Och1* and *Mnn9*, **Figure 7D**) in mannan biosynthesis [Hamilton & Zha 2015; De Wachter et al., 2021); this approach was commercialized by GlycoFi to humanize yeast glycosylation (Beck et al., 2010)]. The success of such approaches is evident from fungal production systems being second only to CHO cells in the breadth of commercial products; Kulagina and coauthors summarize the use of fungal cells to produce four hormones (Novolin®, Glucagen®, Valtropin®, and Semglee®); six vaccines (Recombinvax®, Tritanrix-hepB®, Gardasil®, Mosquirix®, Hexacima®, and Heplisalv-B®); four blood-related proteins (Revasc®, Kalbitor®, Novothirteen®, and Jetrea®); one cytokine (Leukine®) and one enzyme (Fasturtec®) (Kulagine et al., 2021). In addition, glycoengineering strategies are being applied to provide humanized glycan profiles of antibody drugs such as trastuzumab (Herceptin®), where fungal Fc domain glycans optimize ADCC (Liu et al., 2018).

5.3.6 Insect Cells

Insect cells, which have been under investigation for recombinant protein production since the 1970s and 1980s (Hollister et al., 1988), represent another low-cost (by using serum-free, chemically defined media), high-yield (~5 g/L) production system. Manufacturing advantages include no requirement to control CO_2 levels, relaxed temperature control allowing production at lower temperatures, and reduced biosafety and contamination concerns (Yee et al., 2018). Insect cell lines used include S2 from *Drosophila melanogaster*, Sf9 from *Spodoptera frugiperda,* and High Five® from *Trichoplusia ni* (Yee et al., 2018). Additionally, insect cells can perform both N- and O-glycosylation, they efficiently secrete proteins and can cleave signaling peptides, giving them an advantage over prokaryotic pathways and making them plausible production hosts for large

glycoproteins like antibodies (Palmberger et al., 2011; Toth et al., 2014).

There are, however, several glycosylation concerns related to the production of therapeutic proteins in insect cells. For example, insect cells produce simpler N-glycans than mammalian cells, which can affect bioactivity and increase immunogenicity (Geisler et al., 2015; Loos & Steinkellner 2012). Another concern is that although there is some evidence of sialylation in insect cells (Joshi et al., 2001; Kim et al., 2002; Marchal et al., 2001; von Bergen Granell et al., 2011), in general, they do not add this therapeutically important sugar to their glycans. Another concern is that insect cells have double core fucosylation (**Figure 7B**), which, if installed in the Fc domain of IgG antibodies likely would impact downstream Fcγ receptor-mediated effector responses. Finally, the presence of GalNAc in the place of galactose on the elaborated antennae of N-glycans produced in insect cells (**Figure 7E**) is another potential concern. Overall, we note that although there has been decades-long interest in producing IgG antibodies in non-mammalian expression hosts including yeast, plants, and insect cells (Palmberger et al., 2011; Loos & Steinkellner 2012), these efforts have not resulted in commercially successful products. We posit that these difficulties stem in part from glycosylation hurdles, and as such, new glycoengineering approaches will be critical for future production of therapeutic antibodies in a wider range of host cells.

Despite these glycosylation concerns, which have thwarted production of IgG antibodies, several other commercial products have been successfully produced in insect cells. One product is Cervarix®, a virus-like particle (VLP) cervical cancer vaccine produced in High Five cells® (Senger et al., 2009). A second is Provenge®, the first immunotherapy for hormone-refractory prostate cancer, which is produced in Sf21 cells (Contreras-Gómez et al., 2014). A third is Glybera®, a now discontinued adeno-associated virus-based gene therapy for lipoprotein lipase deficiency (LPLD) produced in Sf9 cells (Kurusawa et al., 2020). Finally, Flublok® is a hemagglutinin protein used as an influenza vaccine, which is also made in Sf9 cells (Cox & Hollister 2009). Based on the title of Yee and coauthor's review article "*The coming age of insect cells for manufacturing and development of protein therapeutics*" (Yee et al., 2018), there is reason for cautious optimism for continued expansion of insect cells as a production platform based on long-standing glycoengineering efforts (Ailor et al., 2000; Hollister et al., 1988; Lawrence et al., 2001; Toth et al., 2014). For example, Mabashi-Asazuma and coworkers have developed a new baculovirus vector that eliminates core α1,3-fucosylation in insect cells (**Figure 7B**) (Mabashi-Asazuma et al., 2014), decreasing the immunogenicity of glycoproteins produced in these cells.

6 CONCLUDING COMMENTS AND FUTURE DIRECTIONS

As the most abundant and varied post-translational modification in mammals in general and humans in particular, glycosylation offers great potential to improve today's predominant drugs, which are protein therapeutics. As described in this report, which focuses on N-glycans, there are numerous opportunities to glycoengineer current and upcoming proteins to improve their folding, trafficking, ligand interactions, solubility, stability, and to improve the safety, activity, pharmacokinetics, and pharmacodynamics of this increasingly important class of therapeutics. There already are a handful of deliberately glycoengineered products on the market, with prominent examples being afucosylated pro-inflammatory antibodies and β-glucoceraminidase endowed with high mannose-type glycans for macrophage targeting to treat GD. To date, glycoengineered drugs have exploited a single strategy, typically selection of a host cell line capable of biosynthetically producing the desired type of glycosylation. In the future, as already demonstrated pre-clinically by the glycoengineered ENPP1-Fc (**Figure 3**), multiple glycoengineering strategies (installing new N-glycan sites, production in ST6-overexpressing cells, and media supplementation with a sialic acid precursor) can be productively combined for multifaceted improvement.

Also in the future, additional forms of glycosylation including O-, C-, or S- will provide additional avenues to improve therapeutic proteins. Moreover, the glycoengineering "toolkit" described in **Section 5** provides methodology to improve additional biological therapeutics including antimicrobial peptides (AMPs); glycosylated nanoparticles, liposomes, and exosomes for drug delivery and bioimaging; and glycodendrimers (Jain et al., 2012; Grimsey et al., 2020; Torres-Pérez et al., 2020). The safety of these biopharmaceuticals, including toxicity and immunogenicity, are impacted by glycosylation and, similar to antibodies, their glycoprofiles are critical quality control attributes during biomanufacturing (Mastrangeli et al., 2019).

In conclusion, key examples provided for various glycosylation scenarios demonstrate the potential of individualized, targeted glycan modification to improve various therapeutic proteins. As therapeutic proteins advance, the specific adjustment of glycosylation profiles will hold greater importance as biomanufacturers increasingly move from tuning glycosylation to avoid immunogenicity or toxicity to proactively improving drug efficacy.

AUTHOR CONTRIBUTIONS

KD-B, PE, SZ, ZB, PS, DB, KY and JS all contributed to writing the manuscript. KD-B and KY created and formatted the figures. KD-B, PE, SZ, ZB, KY and JS participated in the review and editing of the manuscript.

FUNDING

This work was supported by the National Institutes of Health (R01EB029455, R01CA240339, R21CA249381, and R01CA112314 and the Cohen Translation Fund (JHU).

REFERENCES

Abu-Qarn, M., Eichler, J., and Sharon, N. (2008). Not just for Eukarya Anymore: Protein Glycosylation in Bacteria and Archaea. *Curr. Opin. Struct. Biol.* 18, 544–550. doi:10.1016/j.sbi.2008.06.010

Ackerman, M. E., Crispin, M., Yu, X., Baruah, K., Boesch, A. W., Harvey, D. J., et al. (2013). Natural Variation in Fc Glycosylation of HIV-specific Antibodies Impacts Antiviral Activity. *J. Clin. Invest.* 123, 2183–2192. doi:10.1172/jci65708

Agatemor, C., Buettner, M. J., Ariss, R., Muthiah, K., Saeui, C. T., and Yarema, K. J. (2019). Exploiting Metabolic Glycoengineering to advance Healthcare. *Nat. Rev. Chem.* 3, 605–620. doi:10.1038/s41570-019-0126-y

Agatemor, C., Muthiah, K., Ha, L., Chai, J., Osman, A., Robertson, B. M., et al. (2021). "Imaging Glycans with Metabolic Glycoengineering," in *Chemistry, Molecular Sciences and Chemical Engineering*. Editor J. BarchiJr (Elsevier), 253–274. doi:10.1016/B978-0-12-409547-2.14962-5

Aich, U., Campbell, C. T., Elmouelhi, N., Weier, C. A., Sampathkumar, S.-G., Choi, S. S., et al. (2008). Regioisomeric SCFA Attachment to Hexosamines Separates Metabolic Flux from Cytotoxicity and MUC1 Suppression. *ACS Chem. Biol.* 3, 230–240. doi:10.1021/cb7002708

Ailor, E., Takahashi, N., Tsukamoto, Y., Masuda, K., Rahman, B. A., Jarvis, D. L., et al. (2000). *N*-glycan Patterns of Human Transferrin Produced in *Trichoplusia Ni* Insect Cells: Effects of Mammalian Galactosyltransferase. *Glycobiology* 10, 837–847. doi:10.1093/glycob/10.8.837

Albright, R. A., Stabach, P., Cao, W., Kavanagh, D., Mullen, I., Braddock, A. A., et al. (2015). ENPP1-Fc Prevents Mortality and Vascular Calcifications in Rodent Model of Generalized Arterial Calcification of Infancy. *Nat. Commun.* 6, 10006. doi:10.1038/ncomms10006

Almaraz, R. T., Aich, U., Khanna, H. S., Tan, E., Bhattacharya, R., Shah, S., et al. (2012). Metabolic Oligosaccharide Engineering with N-Acyl Functionalized ManNAc Analogs: Cytotoxicity, Metabolic Flux, and Glycan-Display Considerations. *Biotechnol. Bioeng.* 109, 992–1006. doi:10.1002/bit.24363

Alter, G., Ottenhoff, T. H. M., and Joosten, S. A. (2018). Antibody Glycosylation in Inflammation, Disease and Vaccination. *Semin. Immunol.* 39, 102–110. doi:10.1016/j.smim.2018.05.003

Alves, M., Lamego, J., Bandeiras, T., Castro, R., Tomás, H., Coroadinha, A. S., et al. (2016). Human Carboxylesterase 2: Studies on the Role of Glycosylation for Enzymatic Activity. *Biochem. Biophys. Rep.* 5, 105–110. doi:10.1016/j.bbrep.2015.11.018

Anthony, R. M., Nimmerjahn, F., Ashline, D. J., Reinhold, V. N., Paulson, J. C., and Ravetch, J. V. (2008). Recapitulation of IVIG Anti-inflammatory Activity with a Recombinant IgG Fc. *Science* 320, 373–376. doi:10.1126/science.1154315

Arena de Souza, V., Scott, D. J., Nettleship, J. E., Rahman, N., Charlton, M. H., Walsh, M. A., et al. (2015). Comparison of the Structure and Activity of Glycosylated and Aglycosylated Human Carboxylesterase 1. *PLoS One* 10, e0143919. doi:10.1371/journal.pone.0143919

Ashwell, G., and Harford, J. (1982). Carbohydrate-specific Receptors of the Liver. *Annu. Rev. Biochem.* 51, 531–554. doi:10.1146/annurev.bi.51.070182.002531

Ayoub, D., Jabs, W., Resemann, A., Evers, W., Evans, C., Main, L., et al. (2013). Correct Primary Structure Assessment and Extensive Glyco-Profiling of Cetuximab by a Combination of Intact, Middle-Up, Middle-Down and Bottom-Up ESI and MALDI Mass Spectrometry Techniques. *mAbs* 5, 699–710. doi:10.4161/mabs.25423

Barahona Afonso, A., and João, C. (2016). The Production Processes and Biological Effects of Intravenous Immunoglobulin. *Biomolecules* 6, 15. doi:10.3390/biom6010015

Bartsch, Y. C., Rahmöller, J., Mertes, M. M. M., Eiglmeier, S., Lorenz, F. K. M., Stoehr, A. D., et al. (2018). Sialylated Autoantigen-Reactive IgG Antibodies Attenuate Disease Development in Autoimmune Mouse Models of Lupus Nephritis and Rheumatoid Arthritis. *Front. Immunol.* 9, 1183. doi:10.3389/fimmu.2018.01183

Baskin, J. M., Prescher, J. A., Laughlin, S. T., Agard, N. J., Chang, P. V., Miller, I. A., et al. (2007). Copper-free Click Chemistry for Dynamic *In Vivo* Imaging. *Proc. Natl. Acad. Sci. U.S.A.* 104, 16793–16797. doi:10.1073/pnas.0707090104

Beck, A., Cochet, O., and Wurch, T. (2010). GlycoFi's Technology to Control the Glycosylation of Recombinant Therapeutic Proteins. *Expert Opin. Drug Discov.* 5, 95–111. doi:10.1517/17460440903413504

Bencharit, S., Morton, C. L., Hyatt, J. L., Kuhn, P., Danks, M. K., Potter, P. M., et al. (2003). Crystal Structure of Human Carboxylesterase 1 Complexed with the Alzheimer's Drug Tacrine. *Chem. Biol.* 10, 341–349. doi:10.1016/s1074-5521(03)00071-1

Boekhout, A. H., Beijnen, J. H., and Schellens, J. H. M. (2011). Trastuzumab. *Tratuzumab. Oncologist.* 16, 800–810. doi:10.1634/theoncologist.2010-0035

Bork, K., Horstkorte, R., and Weidemann, W. (2009). Increasing the Sialylation of Therapeutic Glycoproteins: the Potential of the Sialic Acid Biosynthetic Pathway. *J. Pharm. Sci.* 98, 3499–3508. doi:10.1002/jps.21684

Boune, S., Hu, P., Epstein, A. L., and Khawli, L. A. (2020). Principles of N-Linked Glycosylation Variations of IgG-Based Therapeutics: Pharmacokinetic and Functional Considerations. *Antibodies* 9, 22. doi:10.3390/antib9020022

Boyce, M., Carrico, I. S., Ganguli, A. S., Yu, S.-H., Hangauer, M. J., Hubbard, S. C., et al. (2011). Metabolic Cross-Talk Allows Labeling of O-Linked β- N -Acetylglucosamine-Modified Proteins via the N -acetylgalactosamine Salvage Pathway. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3141–3146. doi:10.1073/pnas.1010045108

Breitfeld, P. P., Rup, D., and Schwartz, A. L. (1984). Influence of the N-Linked Oligosaccharides on the Biosynthesis, Intracellular Routing, and Function of the Human Asialoglycoprotein Receptor. *J. Biol. Chem.* 259, 10414–10421. doi:10.1016/s0021-9258(18)90980-4

Breitling, J., and Aebi, M. (2013). N-linked Protein Glycosylation in the Endoplasmic Reticulum. *Cold Spring Harbor Perspect. Biol.* 5, a013359. doi:10.1101/cshperspect.a013359

Brumshtein, B., Salinas, P., Peterson, B., Chan, V., Silman, I., Sussman, J. L., et al. (2010). Characterization of Gene-Activated Human Acid- -glucosidase: Crystal Structure, Glycan Composition, and Internalization into Macrophages. *Glycobiology* 20, 24–32. doi:10.1093/glycob/cwp138

Buettner, M. J., Shah, S. R., Saeui, C. T., Ariss, R., and Yarema, K. J. (2018). Improving Immunotherapy through Glycodesign. *Front. Immunol.* 9, 2485. doi:10.3389/fimmu.2018.02485

Burgdorf, S., Lukacs-Kornek, V., and Kurts, C. (2006). The Mannose Receptor Mediates Uptake of Soluble but Not of Cell-Associated Antigen for Cross-Presentation. *J. Immunol.* 176, 6770–6776. doi:10.4049/jimmunol.176.11.6770

Burkhart, D. J., Barthel, B. L., Post, G. C., Kalet, B. T., Nafie, J. W., Shoemaker, R. K., et al. (2006). Design, Synthesis, and Preliminary Evaluation of Doxazolidine Carbamates as Prodrugs Activated by Carboxylesterases. *J. Med. Chem.* 49, 7002–7012. doi:10.1021/jm060597e

Burnett, M. J. B., and Burnett, A. C. (2020). Therapeutic Recombinant Protein Production in Plants: Challenges and Opportunities. *Plants People Planet.* 2, 121–132. doi:10.1002/ppp3.10073

Casalino, L., Gaieb, Z., Goldsmith, J. A., Hjorth, C. K., Dommer, A. C., Harbison, A. M., et al. (2020). Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* 6, 1722–1734. doi:10.1021/acscentsci.0c01056

Castilho, A., and Steinkellner, H. (2012). Glyco-engineering in Plants to Produce Human-like N-Glycan Structures. *Biotechnol. J.* 7, 1088–1098. doi:10.1002/biot.201200032

Chitlaru, T., Kronman, C., Zeevi, M., Kam, M., Harel, A., Ordentlich, A., et al. (1998). Modulation of Circulatory Residence of Recombinant Acetylcholinesterase through Biochemical or Genetic Manipulation of Sialylation Levels. *Biochem. J* 336, 647–658. doi:10.1042/bj3360647

Cobb, B. A. (2020). The History of IgG Glycosylation and where We Are Now. *Glycobiology* 30, 202–213. doi:10.1093/glycob/cwz065

Contreras-Gómez, A., Sánchez-Mirón, A., García-Camacho, F., Molina-Grima, E., and Chisti, Y. (2014). Protein Production Using the Baculovirus-Insect Cell Expression System. *Biotechnol. Prog.* 30, 1–18. doi:10.1002/btpr.1842

Cowper, B., Li, X., Yu, L., Zhou, Y., Fan, W. H., and Rao, C. M. (2018). Comprehensive Glycan Analysis of Twelve Recombinant Human Erythropoietin Preparations from Manufacturers in China and Japan. *J. Pharm. Biomed. Anal.* 153, 214–220. doi:10.1016/j.jpba.2018.02.043

Cox, M. M. J., and Hollister, J. R. (2009). FluBlok, a Next Generation Influenza Vaccine Manufactured in Insect Cells. *Biologicals* 37, 182–189. doi:10.1016/j.biologicals.2009.02.014

Cox, T., and Cox, T. M. (2012). Imiglucerase in the Treatment of Gaucher Disease: a History and Perspective. *Ddt* 6, 81–106. doi:10.2147/DDDT.S14395

Cree, B. A. C., Bennett, J. L., Kim, H. J., Weinshenker, B. G., Pittock, S. J., Wingerchuk, D. M., et al. (2019). Inebilizumab for the Treatment of Neuromyelitis Optica Spectrum Disorder (N-MOmentum): a Double-Blind,

Randomised Placebo-Controlled Phase 2/3 Trial. *Lancet* 394, 1352–1363. doi:10.1016/S0140-6736(19)31817-3

Davis, J. S., Kumar, T. R., May, J. V., and Bousfield, G. R. (2014). Naturally Occurring Follicle-Stimulating Hormone Glycosylation Variants. *J. Glycomics Lipidomics* 4, e117. doi:10.4172/2153-0637.1000e117

Daya, S. (2004). Follicle-stimulating Hormone in Clinical Practice. *Treatments Endocrinol.* 3, 161–171. doi:10.2165/00024677-200403030-00004

De Wachter, C., Van Landuyt, L., and Callewaert, N. (2018). Engineering of Yeast Glycoprotein Expression. *Adv. Biochem. Eng. Biotechnol.* 175, 93–135. doi:10.1007/10_2018_69

del Val, I. J., Polizzi, K. M., and Kontoravdi, C. (2016). A Theoretical Estimate for Nucleotide Sugar Demand towards Chinese Hamster Ovary Cellular Glycosylation. *Sci. Rep.* 6, 28547. doi:10.1038/srep28547

Dias, J. A., and Ulloa-Aguirre, A. (2021). New Human Follitropin Preparations: How Glycan Structural Differences May Affect Biochemical and Biological Function and Clinical Effect. *Front. Endocrinol.* 12, 636038. doi:10.3389/fendo.2021.636038

Ding, N., Ruan, Y., Fu, X., Lin, Y., Yu, H., Han, L., et al. (2019). Improving Production of N-Glycosylated Recombinant Proteins by Leaky *Escherichia coli*. *3 Biotech.* 9, 302. doi:10.1007/s13205-019-1830-5

Drake, P. M., Albers, A. E., Baker, J., Banas, S., Barfield, R. M., Bhat, A. S., et al. (2014). Aldehyde Tag Coupled with HIPS Chemistry Enables the Production of ADCs Conjugated Site-Specifically to Different Antibody Regions with Distinct *In Vivo* Efficacy and PK Outcomes. *Bioconjug. Chem.* 25, 1331–1341. doi:10.1021/bc500189z

Du, J., Agatemor, C., Saeui, C. T., Bhattacharya, R., Jia, X., and Yarema, K. J. (2021). Glycoengineering Human Neural and Adipose Stem Cells with Novel Thiol-Modified N-Acetylmannosamine (ManNAc) Analogs. *Cells* 10, 377377. doi:10.3390/cells10020377

Du, J., Liu, X., Yarema, K. J., and Jia, X. (2022). Glycoengineering Human Neural Stem Cells (hNSCs) for Adhesion Improvement Using a Novel Thiol-Modified N-Acetylmannosamine (ManNAc) Analog. *Mater. Sci. Eng. C* 2022, 112675. doi:10.1016/j.msec.2022.112675

Du, J., Meledeo, M. A., Wang, Z., Khanna, H. S., Paruchuri, V. D. P., and Yarema, K. J. (2009). Metabolic Glycoengineering: Sialic Acid and beyond. *Glycobiology* 19, 1382–1401. doi:10.1093/glycob/cwp115

Dumont, J., Euwart, D., Mei, B., Estes, S., and Kshirsagar, R. (2016). Human Cell Lines for Biopharmaceutical Manufacturing: History, Status, and Future Perspectives. *Crit. Rev. Biotechnol.* 36, 1110–1122. doi:10.3109/07388551.2015.1084266

Dwek, R. A., and Butters, T. D. (2002). Introduction: GlycobiologyUnderstanding the Language and Meaning of Carbohydrates. *Chem. Rev.* 102, 283–284. doi:10.1021/cr010385j

Dwek, R. A. (1996). Glycobiology: Toward Understanding the Function of Sugars. *Chem. Rev.* 96, 683–720. doi:10.1021/10.1021/cr940283b

Egrie, J. C., and Browne, J. K. (2002). Development and Characterization of Darbepoetin Alfa. *Oncology (Williston Park)* 16, 13–22.

Egrie, J. C., Dwyer, E., Browne, J. K., Hitz, A., and Lykos, M. A. (2003). Darbepoetin Alfa Has a Longer Circulating Half-Life and Greater *In Vivo* Potency Than Recombinant Human Erythropoietin. *Exp. Hematol.* 31, 290–299. doi:10.1016/s0301-472x(03)00006-7

Ellies, L. G., Ditto, D., Levy, G. G., Wahrenbrock, M., Ginsburg, D., Varki, A., et al. (2002). Sialyltransferase ST3Gal-IV Operates as a Dominant Modifier of Hemostasis by Concealing Asialoglycoprotein Receptor Ligands. *Proc. Natl. Acad. Sci. U.S.A.* 99, 10042–10047. doi:10.1073/pnas.142005099

Elliot, S. G., Lorenzini, T., Strickland, T. W., Delorme, E., and Egrie, J. C. (2000). Rational Design of Novel Erythropoiesis Stimulating Protein (Aranesp^{TM}): a Super-sialylated Molecule with Increased Biological Activity. *Blood* 96. 2a (abstract 352).

Elliott, S., Pham, E., and Macdougall, I. C. (2008). Erythropoietins: A Common Mechanism of Action. *Exp. Hematol.* 36, 1573–1584. doi:10.1016/j.exphem.2008.08.003

Feinberg, H., Jégouzo, S. A. F., Lasanajak, Y., Smith, D. F., Drickamer, K., Weis, W. I., et al. (2021). Structural Analysis of Carbohydrate Binding by the Macrophage Mannose Receptor CD206. *J. Biol. Chem.* 296, 100368–100369. doi:10.1016/j.jbc.2021.100368

Ferreira, C. R., Kintzinger, K., Hackbarth, M. E., Botschen, U., Nitschke, Y., Mughal, M. Z., et al. (2021). Ectopic Calcification and Hypophosphatemic Rickets: Natural History of ENPP1 and ABCC6 Deficiencies. *J. Bone Miner. Res.* 36, 2193–2202. doi:10.1002/jbmr.4418

Fischer, R., Holland, T., Sack, M., Schillberg, S., Stoger, E., Twyman, R. M., et al. (2018). Glyco-engineering of Plant-Based Expression Systems. *Adv. Biochem. Eng. Biotechnol.* 175, 137–166. doi:10.1007/10_2018_76

Fleming, C. D., Bencharit, S., Edwards, C. C., Hyatt, J. L., Tsurkan, L., Bai, F., et al. (2005). Structural Insights into Drug Processing by Human Carboxylesterase 1: Tamoxifen, Mevastatin, and Inhibition by Benzil. *J. Mol. Biol.* 352, 165–177. doi:10.1016/j.jmb.2005.07.016

Fleming, C. D., Edwards, C. C., Kirby, S. D., Maxwell, D. M., Potter, P. M., Cerasoli, D. M., et al. (2007). Crystal Structures of Human Carboxylesterase 1 in Covalent Complexes with the Chemical Warfare Agents Soman and Tabun,. *Biochemistry* 46, 5063–5071. doi:10.1021/bi700246n

Fox, E., Lovett-Racke, A. E., Gormley, M., Liu, Y., Petracca, M., Cocozza, S., et al. (2021). A Phase 2 Multicenter Study of Ublituximab, a Novel Glycoengineered Anti-CD20 Monoclonal Antibody, in Patients with Relapsing Forms of Multiple Sclerosis. *Mult. Scler.* 27, 420–429. doi:10.1177/1352458520918375

Fox, J. L. (2012). First Plant-Made Biologic Approved. *Nat. Biotechnol.* 30, 472. doi:10.1038/nbt0612-472

Friedman, B., Vaddi, K., Preston, C., Mahon, E., Cataldo, J. R., and McPherson, J. M. (1999). A Comparison of the Pharmacological Properties of Carbohydrate Remodeled Recombinant and Placental-Derived β-Glucocerebrosidase: Implications for Clinical Efficacy in Treatment of Gaucher Disease. *Blood* 93, 2807–2816. doi:10.1182/blood.v93.9.2807.409k08_2807_2816

Friedman, R. (2008). *A Southern Mystery.* The Scientist, 26526.

Frost, G. I. (2007). Recombinant Human Hyaluronidase (rHuPH20): an Enabling Platform for Subcutaneous Drug and Fluid Administration. *Expert Opin. Drug Deliv.* 4, 427–440. doi:10.1517/17425247.4.4.427

Fukami, T., and Yokoi, T. (2012). The Emerging Role of Human Esterases. *Drug Metab. Pharmacokinet.* 27, 466–477. doi:10.2133/dmpk.dmpk-12-rv-042

Gazi, U., and Martinez-Pomares, L. (2009). Influence of the Mannose Receptor in Host Immune Responses. *Immunobiology* 214, 554–561. doi:10.1016/j.imbio.2008.11.004

Geisler, C., Mabashi-Asazuma, H., and Jarvis, D. L. (2015). "An Overview and History of Glyco-Engineering in Insect Expression Systems," in *Glyco-Engineering. Methods in Molecular Biology*. Editor A. Castilho (New York: Humana Press), 131–152. doi:10.1007/978-1-4939-2760-9_10

Ghaderi, D., Taylor, R. E., Padler-Karavani, V., Diaz, S., and Varki, A. (2010). Implications of the Presence of N-Glycolylneuraminic Acid in Recombinant Therapeutic Glycoproteins. *Nat. Biotechnol.* 28, 863–867. doi:10.1038/nbt.1651

Ghaderi, D., Zhang, M., Hurtado-Ziola, N., and Varki, A. (2012). Production Platforms for Biotherapeutic Glycoproteins. Occurrence, Impact, and Challenges of Non-human Sialylation. *Biotechnol. Genet. Eng. Rev.* 28, 147–176. doi:10.5661/bger-28-147

Giddens, J. P., Lomino, J. V., DiLillo, D. J., Ravetch, J. V., and Wang, L.-X. (2018). Site-selective Chemoenzymatic Glycoengineering of Fab and Fc Glycans of a Therapeutic Antibody. *Proc. Natl. Acad. Sci. U.S.A.* 115, 12023–12027. doi:10.1073/pnas.1812833115

Go, E. P., Ding, H., Zhang, S., Ringe, R. P., Nicely, N., Hua, D., et al. (2017). Glycosylation Benchmark Profile for HIV-1 Envelope Glycoprotein Production Based on Eleven Env Trimers. *J. Virol.* 91, e02428-02416. doi:10.1128/jvi.02428-16

Golay, J., Da Roit, F., Bologna, L., Ferrara, C., Leusen, J. H., Rambaldi, A., et al. (2013). Glycoengineered CD20 Antibody Obinutuzumab Activates Neutrophils and Mediates Phagocytosis through CD16B More Efficiently Than Rituximab. *Blood* 122, 3482–3491. doi:10.1182/blood-2013-05-504043

Gong, Y., Qin, S., Dai, L., and Tian, Z. (2021). The Glycosylation in SARS-CoV-2 and its Receptor ACE2. *Signal. Transduct. Target. Ther.* 6, 396. doi:10.1038/s41392-021-00809-8

Greig, N. H., Lahiri, D. K., and Sambamurti, K. (2002). Butyrylcholinesterase: an Important New Target in Alzheimer's Disease Therapy. *Int. Psychogeriatr* 14 (Suppl. 1), 77–91. doi:10.1017/s1041610203008676

Grimsey, E., Collis, D. W. P., Mikut, R., and Hilpert, K. (2020). The Effect of Lipidation and Glycosylation on Short Cationic Antimicrobial Peptides. *Biochim. Biophys. Acta Biomem.* 1862, 183195. doi:10.1016/j.bbamem.2020.183195

Gross, H. J., and Brossmer, R. (1988). Enzymatic Introduction of a Fluorescent Sialic Acid into Oligosaccharide Chains of Glycoproteins. *Eur. J. Biochem.* 177, 583–589. doi:10.1111/j.1432-1033.1988.tb14410.x

Gross, H. J., and Brossmer, R. (1995). Enzymatic Transfer of Sialic Acids Modified at C-5 Employing Four Different Sialyltransferases. *Glycoconj. J.* 12, 739–746. doi:10.1007/BF00731233

Gross, H. J., Rose, U., Krause, J. M., Paulson, J. C., Schmid, K., Feeny, R. E., et al. (1989). Transfer of Synthetic Sialic Acid Analogues to N- and O-Linked Glycoprotein Glycans Using Four Different Mammalian Sialyltransferases. *Biochemistry* 28, 7386–7392. doi:10.1021/bi00444a036

Guan, X., Chaffey, P. K., Wei, X., Gulbranson, D. R., Ruan, Y., Wang, X., et al. (2018). Chemically Precise Glycoengineering Improves Human Insulin. *ACS Chem. Biol.* 13, 73–81. doi:10.1021/acschembio.7b00794

Gupta, R., and Brunak, S. (2002). Prediction of Glycosylation across the Human Proteome and the Correlation to Protein Function. *Pac. Symp. Biocomputing* 7, 310–322.

Hamilton, S. R., and Zha, D. (2015). "Progress in Yeast Glycoengineering," in *Glyco-Engineering: Methods in Molecular Biology*. Editor A. Castilho (New York: Humana Press), 73–90. doi:10.1007/978-1-4939-2760-9_6

Helenius, A., and Aebi, M. (2001). Intracellular Functions of N-Linked Glycans. *Science* 291, 2364–2369. doi:10.1126/science.291.5512.2364

Heß, R., genannt Bonsmann, M. S., Lapuente, D., Maaske, A., Kirschning, C., Ruland, J., et al. (2019). Glycosylation of HIV Env Inpacts IgG Subtype Responses to Vaccination. *Viruses* 11, 153. doi:10.3390/v11020153

Higashiyama, T., Umekawa, M., Nagao, M., Katoh, T., Ashida, H., and Yamamoto, K. (2018). Chemo-enzymatic Synthesis of the Glucagon Containing N-Linked Oligosaccharide and its Characterization. *Carbohydr. Res.* 455, 92–96. doi:10.1016/j.carres.2017.11.007

Hokke, C. H., Bergwerff, A. A., Vvan Dedem, G. W. K., van Oostrum, J., Kamerling, J. P., and Vliegenthart, J. F. G. (1990). Sialylated Carbohydrate Chains of Recombinant Human Glycoproteins Expressed in Chinese Hamster Ovary Cells Contain Traces of *N*-Glycolylneuraminic Acid. *FEBS Lett.* 275, 9–14. doi:10.1016/0014-5793(90)81427-P

Hollister, J. R., Shaper, J. H., and Jarvis, D. L. (1988). Stable Expression of Mammalian B1,4-Galactosyltransferase Extends the N-Glycosylation Pathway in Insect Cells. *Glycobiology* 8, 473–480. doi:10.1093/glycob/8.5.473

Hong, V., Steinmetz, N. F., Manchester, M., and Finn, M. G. (2010). Labeling Live Cells by Copper-Datalyzed Alkyne-Azide Click Chemistry. *Bioconjug. Chem.* 21, 1912–1916. doi:10.1021/bc100272z

Hurjak, B., Kovács, Z., Dönczó, B., Katona, É., Haramura, G., Erdélyi, F., et al. (2020). N-glycosylation of Blood Coagulation Factor XIII Subunit B and its Functional Consequence. *J. Thromb. Haemost.* 18, 1302–1309. doi:10.1111/jth.14792

Ichikawa, M., Hirayama, T., Fukushima, M., Kitazawa, I., Kojima, K., Sakai, T., et al. (2018). Glycosaminoglycan Conjugation for Improving the Duration of Therapeutic Action of Glucagon-like Peptide-1. *ACS Omega* 3, 5346–5354. doi:10.1021/acsomega.8b00467

Irvine, E. B., and Alter, G. (2020). Understanding the Role of Antibody Glycosylation through the Lens of Severe Viral and Bacterial Diseases. *Glycobiology* 30, 241–253. doi:10.1093/glycob/cwaa018

Jain, K., Kesharwani, P., Gupta, U., and Jain, N. K. (2012). A Review of Glycosylated Carriers for Drug Delivery. *Biomaterials* 33, 4166–4186. doi:10.1016/j.biomaterials.2012.02.033

Janin-Bussat, M.-C., Tonini, L., Huillet, C., Colas, O., Klinguer-Hamour, C., Corvaia, N., et al. (2013). Cetuximab Fab and Fc N-Glycan Fast Characterization Using IdeS Digestion and Liquid Chromatography Coupled to Electrospray Ionization Mass Spectrometry. *Methods Mol. Biol.* 988, 93–113. doi:10.1007/978-1-62703-327-5_7

Ji, Y., White, Y. J. B., Hadden, J. A., Grant, O. C., and Woods, R. J. (2017). New Insights into Influenza A Specificity: an Evolution of Paradigms. *Curr. Opin. Struct. Biol.* 44, 219–231. doi:10.1016/j.sbi.2017.06.001

Jo, S., Kim, T., Iyer, V. G., and Im, W. (2008). CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* 29, 1859–1865. doi:10.1002/jcc.20945

Johnson, J. L., Jones, M. B., Ryan, S. O., and Cobb, B. A. (2013). The Regulatory Power of Glycans and Their Binding Partners in Immunity. *Trends Immunol.* 34, 290–298. doi:10.1016/j.it.2013.01.006

Jones, M. B., Oswald, D. M., Joshi, S., Whiteheart, S. W., Orlando, R., and Cobb, B. A. (2016). B-cell–independent Sialylation of IgG. *Proc. Natl. Acad. Sci. U. S. A.* 11, 7207–7212. doi:10.1073/pnas.1523968113

Jones, M. B., Teng, H., Rhee, J. K., Baskaran, G., Lahar, N., and Yarema, K. J. (2004). Characterization of the Cellular Uptake and Metabolic Conversion of Acetylated *N*-Acetylmannosamine (ManNAc) Analogues to Sialic Acids. *Biotechnol. Bioeng.* 85, 394–405. doi:10.1002/bit.10901

Joshi, L., Shuler, M. L., and Wood, H. A. (2001). Production of a Sialylated *N*-Linked Glycoprotein in Insect Cells. *Biotechnol. Prog.* 17, 822–827. doi:10.1021/bp010071h

Jourdian, G. W., Dean, L., and Roseman, S. (1971). The Sialic Acids. XI. A Periodate-Resorcinol Method for the Quantitative Estimation of Free Sialic Acids and Their Glycosides. *J. Biol. Chem.* 246, 430–435. doi:10.1016/s0021-9258(18)62508-6

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* 596, 583–589. doi:10.1038/s41586-021-03819-2

Kamber, D. N., Nyguyen, S. S., Liu, F., Briggs, J. S., Shih, H.-W., Row, R. D., et al. (2019). Isomeric Triazines Exhibit Unique Profiles of Bioorthogonal Reactivity. *Chem. Sci.* 10, 9109. doi:10.1039/c9sc01427f

Kaneko, Y., Nimmerjahn, F., and Ravetch, J. V. (2006). Anti-inflammatory Activity of Immunoglobulin G Resulting from Fc Sialylation. *Science* 313, 670–673. doi:10.1126/science.1129594

Karki, U., Fang, H., Guo, W., Unnold-Cofre, C., and Xu, J. (2021). Cellular Engineering of Plant Cells for Improved Therapeutic Protein Production. *Plant Cell Rep.* 40, 1087–1099. doi:10.1007/s00299-021-02693-6

Kayser, H., Geilen, C. C., Paul, C., Zeitler, R., and Reutter, W. (1992a). Incorporation of *N*-Acyl-2-Amino-2-Deoxy-Hexoses into Glycosphingolipids of the Pheochromocytoma Cell Line PC 12. *FEBS Lett.* 301, 137–140. doi:10.1016/0014-5793(92)81233-c

Kayser, H., Zeitler, R., Kannicht, C., Grunow, D., Nuck, R., and Reutter, W. (1992b). Biosynthesis of a Nonphysiological Sialic Acid in Different Rat Organs, Using *N*-Propanoyl-D-Hexosamines as Precursors. *J. Biol. Chem.* 267, 16934–16938. doi:10.1016/s0021-9258(18)41874-1

Keppler, O. T., Horstkorte, R., Pawlita, M., Schmidt, C., and Reutter, W. (2001). Biochemical Engineering of the *N*-Acyl Side Chain of Sialic Acid: Biological Implications. *Glycobiology* 11, 11R–18R. doi:10.1093/glycob/11.2.11R

Kim, E. J., Sampathkumar, S.-G., Jones, M. B., Rhee, J. K., Baskaran, G., and Yarema, K. J. (2004). Characterization of the Metabolic Flux and Apoptotic Effects of *O*-Hydroxyl- and *N*-Acetylmannosamine (ManNAc) Analogs in Jurkat (Human T-Lymphoma-Derived) Cells. *J. Biol. Chem.* 279, 18342–18352. doi:10.1074/jbc.M400205200

Kim, K., Lawrence, S. M., Park, J., Pitts, L., Vann, W. F., Betenbaugh, M. J., et al. (2002). Expression of a Functional *Drosophila melanogaster* N-Acetylneuraminic Acid (Neu5Ac) Phosphate Synthase Gene: Evidence for Endogneous Sialic Acid Biosynthetic Ability in Insects. *Glycobiology* 12, 73–83. doi:10.1093/glycob/12.2.73

Kohi, S., Sato, N., Koga, A., Hirata, K., Harunari, E., and Igarashi, Y. (2016). Hyaluromycin, a Novel Hyaluronidase Inhibitor, Attenuates Pancreatic Cancer Cell Migration and Proliferation. *J. Oncol.* 2016, 9063087. doi:10.1155/2016/9063087

Kolarich, D., Weber, A., Pabst, M., Stadlmann, J., Teschner, W., Ehrlich, H., et al. (2008). Glycoproteomic Characterization of Butyrylcholinesterase from Human Plasma. *Proteomics* 8, 254–263. doi:10.1002/pmic.200700720

Kolbeck, R., Kozhich, A., Koike, M., Peng, L., Andersson, C. K., Damschroder, M. M., et al. (2010). MEDI-563, a Humanized anti-IL-5 Receptor Alpha mAb with Enhanced Antibody-dependent Cell-Mediated Cytotoxicity Function. *J. Allergy Clin. Immunol.* 125, 1344–1353e1342. doi:10.1016/j.jaci.2010.04.004

Kosloski, M. P., Miclea, R., and Balu-Iyer, S. V. (2009). Role of Glycosylation in Conformational Stability, Activity, Macromolecular Interaction and Immunogenicity of Recombinant Human Factor VIII. *AAPS J.* 11, 424. doi:10.1208/s12248-009-9119-y

Kotsovilis, S., and Andreakos, E. (2014). Therapeutic Human Monoclonal Antibodies in Inflammatory Diseases. *Methods Mol. Biol.* 1016, 37–59. doi:10.1007/978-1-62703-586-6_3

Kroetz, D. L., McBride, O. W., and Gonzalez, F. J. (1993). Glycosylation-dependent Activity of Baculovirus-Expressed Human Liver Carboxylesterases: cDNA

Cloning and Characterization of Two Highly Similar Enzyme Forms. *Biochemistry* 32, 11606–11617. doi:10.1021/bi00094a018

Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K., and Moult, J. (2021). Critical Assessment of Methods of Protein Structure Prediction (CASP) - Round XIV. *Proteins: Struct. Funct. Bioinf.* 89, 1607–1617. doi:10.1002/prot.26237

Kuhlman, B., and Bradley, P. (2019). Advances in Protein Structure Prediction and Design. *Nat. Rev. Mol. Cell Biol.* 20, 681–697. doi:10.1038/s41580-019-0163-x

Kulagine, N., Besseau, S., Godon, C., Goldman, G. H., Papon, N., and Courdavault, V. (2021). Yeasts as Biopharmaceutical Production Platforms. *Front. Fungal Biol.* 2, 733492. doi:10.3389/ffunb.2021.733492

Kurusawa, J. H., Park, A., Sowers, C. R., Halpin, R. A., Towchigrechko, A., Dobson, C. L., et al. (2020). Chemically Defined, High-Density Insect Cell-Based Expression System for Scalable AAV Vector Production. *Mol. Ther. Methods Clin. Dev.* 19, 330–340. doi:10.1016/j.omtm.2020.09.018

Labonte, J. W., Adolf-Bryfogle, J., Schief, W. R., and Gray, J. J. (2016). Residue-centric Modeling and Design of Saccharide and Glycoconjugate Structures. *J. Comput. Chem.* 38, 276–287. doi:10.1002/jcc.24679

Lajoie, P., Partridge, E. A., Guay, G., Goetz, J. G., Pawling, J., Lagana, A., et al. (2007). Plasma Membrane Domain Organization Regulates EGFR Signaling in Tumor Cells. *J. Cell Biol.* 179, 341–356. doi:10.1083/jcb.200611106

Lalonde, M.-E., and Durocher, Y. (2017). Therapeutic Glycoprotein Production in Mammalian Cells. *J. Biotechnol.* 17, 128–140. doi:10.1016/j.jbiotec.2017.04.028

Lawrence, S. M., Huddleston, K. A., Tomiya, N., Nguyen, N., Lee, Y. C., Vann, W. F., et al. (2001). Cloning and Expression of Human Sialic Acid Pathway Genes to Generate CMP-Sialic Acids in Insect Cells. *Glycoconj. J.* 18, 205–213. doi:10.1023/a:1012452705349

Lee, S. J., Evers, S., Roeder, D., Parlow, A. F., Risteli, J., Risteli, L., et al. (2002). Mannose Receptor-Mediated Regulation of Serum Glycoprotein Homeostasis. *Science* 295, 1898–1901. doi:10.1126/science.1069540

Lemieux, G. A., Yarema, K. J., Jacobs, C. L., and Bertozzi, C. R. (1999). Exploiting Differences in Sialoside Expression for Selective Targeting of MRI Contrast Reagents. *J. Am. Chem. Soc.* 121, 4278–4279. doi:10.1021/ja984228m

Li, D., Lou, Y., Zhang, Y., Liu, S., Li, J., and Tao, J. (2021a). Sialylated Immunoglobulin G: A Promising Diagnostic and Therapeutic Strategy for Autoimmune Diseases. *Theranostics* 11, 5430–5446. doi:10.7150/thno.53961

Li, H., and d'Aniou, M. (2009). Pharmacological Significance of Glycosylation in Therapeutic Proteins. *Curr. Opin. Biotechnol.* 20, 678–684. doi:10.1016/j.copbio.2009.10.009

Li, M., Zheng, X., Shanker, S., Jaroentomeechai, T., Moeller, T. D., Hulbert, S. W., et al. (2021b). Shotgun Scanning Glycomutagenesis: A Simple and Efficient Strategy for Constructing and Characterizing Neoglycoproteins. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2107440118. doi:10.1073/pnas.2107440118

Li, T., DiLillo, D. J., Bournazos, S., Giddens, J. P., Ravetch, J. V., and Wang, L.-X. (2017). Modulating IgG Effector Function by Fc Glycan Engineering. *Proc. Natl. Acad. Sci. U. S. A.* 114, 3485–3490. doi:10.1073/pnas.1702173114

Li, X., Fang, T., and Boons, G.-J. (2014). Preparation of Well-Defined Antibody–Drug Conjugates through Glycan Remodeling and Strain-Promoted Azide–Alkyne Cycloadditions. *Angew. Chem. Int. Ed. Eng.* 53, 7179–7182. doi:10.1002/anie.201402606

Lifely, M. R., Hale, C., Boyce, S., Keen, M. J., and Phillips, J. (1995). Glycosylation and Biological Activity of CAMPATH-1H Expressed in Different Cell Lines and Grown under Different Culture Conditions. *Glycobiology* 5, 813–822. doi:10.1093/glycob/5.8.813

Liu, C.-P., Tsai, T.-I., Cheng, T., Shivatare, V. S., Wu, C.-Y., Wu, C.-Y., et al. (2018). Glycoengineering of Antibody (Herceptin) through Yeast Expression and *In Vitro* Enzymatic Glycosylation. *Proc. Natl. Acad. Sci. U. S. A.* 115, 720–725. doi:10.1073/pnas.1718172115

Liu, L. (2015). Antibody Glycosylation and its Impact on the Pharmacokinetics and Pharmacodynamics of Monoclonal Antibodies and Fc-Fusion Proteins. *J. Pharm. Sci.* 104, 1866–1884. doi:10.1002/jps.24444

Liu, L. (2018). Pharmacokinetics of Monoclonal Antibodies and Fc-Fusion Proteins. *Protein Cell* 9, 15–32. doi:10.1007/s13238-017-0408-4

Liu, T., Li, Y., Xu, J., Guo, Q., Zhang, D., Song, L., et al. (2021). N-glycosylation and Enzymatic Activity of the rHuPH20 Expressed in Chinese Hamster Ovary Cells. *Anal. Biochem.* 632, 114380. doi:10.1016/j.ab.2021.114380

Loos, A., and Steinkellner, H. (2012). IgG-Fc Glycoengineering in Non-mammalian Expression Hosts. *Arch. Biochem. Biophys.* 526, 167–173. doi:10.1016/j.abb.2012.05.011

Lu, R.-M., Hwang, Y.-C., Liu, I.-J., Lee, C.-C., Tsai, H.-Z., Li, H.-J., et al. (2020). Development of Therapeutic Antibodies for the Treatment of Dieases. *J. Biomed. Sci.* 27, 1. doi:10.1186/s12929-019-0592-z

Luo, C., Chen, S., Xu, N., Wang, C., Sai, W. b., Zhao, W., et al. (2017). Glycoengineering of Pertuzumab and its Impact on the Pharmacokinetic/pharmacodynamic Properties. *Sci. Rep.* 7, 46347. doi:10.1038/srep46347

Ma, B., Guan, X., Li, Y., Shang, S., Li, J., and Tan, Z. (2020a). Protein Glycoengineering: An Approach for Improving Protein Properties. *Front. Chem.* 8, 622. doi:10.3389/fchem.2020.00622

Ma, B., Guan, X., Li, Y., Shang, S., Li, J., and Zhongping, T. (2020b). Protein Glycoengineering: An Approach for Improving Protein Properties. *Front. Chem.* 8, 622. doi:10.3389/fchem.2020.00622

Maack, T., Johnson, V., Kau, S. T., Rigueiredo, J., and Sigulem, D. (1979). Renal Filtration, Transport, and Metabolism of Low-Molecular-Weight Proteins: A Review. *Kidney Int.* 16, 251–270. doi:10.1038/ki.1979.128

Mabashi-Asazuma, H., Kuo, C.-W., Khoo, K.-H., and Jarvis, D. L. (2014). A Novel Baculovirus for the Production of Nonfucosylated Recombinant Glycoproteins in Insect Cells. *Glycobiology* 24, 325–340. doi:10.1093/glycob/cwt161

Macdougal, I. C. (2002). Optimizing the Use of Erythropoietic Agents – Pharmacokinetic and Pharmacodynamic Considerations. *Nephrol. Dial. Transpl.* 17, 66–70. doi:10.1093/ndt/17.suppl_5.66

Mahal, L. K., Yarema, K. J., and Bertozzi, C. R. (1997). Engineering Chemical Reactivity on Cell Surfaces through Oligosaccharide Biosynthesis. *Science* 276, 1125–1128. doi:10.1126/science.276.5315.1125

Maneval, D. C., Caster, C. L., Cerunes, C., Locke, K. W., Muhsin, M., Sauter, S., et al. (2020). "Chapter 9 - Pegvorhyaluronidase Alfa: A PEGylated Recombinant Human Hyaluronidase PH20 for the Treatment of Cancers that Accumulate Hyaluronan," in *Polymer-Protein Conjugates*. Editors G. Pasut and S. Zalipsky (Elsevier B.V.).

Marchal, I., Jarvis, D. L., Cacan, R., and Verbert, A. (2001). Glycoproteins from Insect Cells: Sialylated or Not? *Biol. Chem.* 382, 151–159. doi:10.1515/BC.2001.023

Marino, M., Jamal, Z., and Zito, P. M. (2021). Pharmacodynamics. NCBI Bookshelf. Available at: www.ncbi.nlm.nih.gov/books/NBK507791/.

Martins, J. P., Kennedy, P. J., Santos, H. A., Barria, C., and Sarmento, B. (2016). A Comprehensive Review of the Neonatal Fc Receptor and its Application in Drug Delivery. *Pharmacol. Ther.* 161, 22–39. doi:10.1016/j.pharmthera.2016.03.007

Mastrangeli, R., Palinsky, W., and Bierau, H. (2019). Glycoengineered Antibodies: towards the Next-Generation of Immunotherapeutics. *Glycobiology* 29, 199–210. doi:10.1093/glycob/cwy092

Mathew, M. P., Tan, E., Labonte, J. W., Shah, S., Saeui, C. T., Liu, L., et al. (2017). Glycoengineering of Esterase Activity through Metabolic Flux-Based Modulation of Sialic Acid. *ChemBioChem* 18, 1204–1215. doi:10.1002/cbic.201600698

Mathew, M. P., Tan, E., Saeui, C. T., Bovonratwet, P., Sklar, S., Bhattacharya, R., et al. (2016). Metabolic Flux-Driven Sialylation Alters Internalization, Recycling, and Drug Sensitivity of the Epidermal Growth Factor Receptor (EGFR) in SW1990 Pancreatic Cancer Cells. *Oncotarget* 7, 66491–66511. doi:10.18632/oncotarget.11582

Mathew, M. P., Tan, E., Shah, S., Bhattacharya, R., Meledeo, M. A., Huang, J., et al. (2012). Extracellular and Intracellular Esterase Processing of SCFA-Hexosamine Analogs: Implications for Metabolic Glycoengineering and Drug Delivery. *Bioorg. Med. Chem. Lett.* 22, 6929–6933. doi:10.1016/j.bmcl.2012.09.017

McAtee, C., Barycki, J. J., and Simpson, M. A. (2014). Emerging Roles for Hyaluronidase in Cancer Metastasis and Therapy. *Adv. Cancer Res.* 123, 1–34. doi:10.1016/b978-0-12-800092-2.00001-0

Menacho-Melgar, R., Ye, Z., Moreb, E. A., Yang, T., Efromson, J. P., Decker, J. S., et al. (2020). Improved, Scalable, Two-Stage, Autoinduction of Recombinant Protein Expression in *E. coli* Using Phosphate Depletion. *bioRxiv*. doi:10.10023/bit.27440

Mertz, J. L., Sun, S., Yin, B., Hu, Y., Bhattacharya, R., Betenbaugh, M. J., et al. (2020). Comparison of Three Glycoproteomic Methods for the Analysis of the Secretome of CHO Cells Treated with 1,3,4-O-Bu_3ManNAc. *Bioengineering* 7, 144. doi:10.3390/bioengineering7040144

Montero-Morales, L., and Steinkellner, H. (2018). Advanced Plant-Based Glycan Engineering. *Front. Bioeng. Biotechnol.* 6, 81. doi:10.3389/fbioe.2018.00081

Mullard, A. (2021). FDA Approves 100th Monoclonal Antibody Product. *Nat. Rev. Drug Discov.* 20, 491–495. doi:10.1038/d41573-021-00079-7

Muthana, S., Cao, H., and Chen, X. (2009). Recent Progress in Chemical and Chemoenzymatic Synthesis of Carbohydrates. *Curr. Opin. Chem. Biol.* 13, 573–581. doi:10.1016/j.cbpa.2009.09.013

Nakano, M., Higo, D., Arai, E., Nakagawa, T., Kakehi, K., Taniguchi, N., et al. (2009). Capillary Electrophoresis-Electrospray Ionization Mass Spectrometry for Rapid and Sensitive *N*-Glycan Analysis of Glycoproteins as 9-fluorenylmethyl Derivatives. *Glycobiology* 19, 135–143. doi:10.1093/glycob/cwn115

Narimatsu, Y., Büll, C., Chen, Y.-H., Wandall, H. H., Yang, Z., and Clausen, H. (2021). Genetic Glycoengineering in Mammalian Cells. *J. Biol. Chem.* 296, 100448. doi:10.1016/j.jbc.2021.100448

Ning, X., Temming, R. P., Dommerholt, J., Guo, J., Ania, D. B., Debets, M. F., et al. (2010). Protein Modification by Strain-Promoted Alkyne-Nitrone Cycloaddition. *Angew. Chem. Int. Ed. Eng.* 49, 3065–3068. doi:10.1002/anie.201000408

Nordberg, A., Ballard, C., Bullock, R., Darreh-Shori, T., and Somogyi, M. (2013). A Review of Butyrylcholinesterase as a Therapeutic Target in the Treatment of Alzheimer's Disease. *Prim. Care Companion CNS Disord.* 15, 12r01412. pii: PCC. doi:10.4088/PCC.12r01412

Offersen, R., Yu, W.-H., Scully, E. P., Julg, B., Euler, Z., Sadanand, S., et al. (2020). HIV Antibody Fc N-Linked Flycosylation Is Associated with Viral Rebound. *Cell Rep* 33, 103502. doi:10.1016/j.celrep.2020.108502

Okeley, N. M., Toki, B. E., Zhang, X., Jeffrey, S. C., Burke, P. J., Alley, S. C., et al. (2013). Metabolic Engineering of Monoclonal Antibody Carbohydrates for Antibody–drug Conjugation. *Bioconjug. Chem.* 24, 1650–1655. doi:10.1021/bc4002695

Olden, K., Parent, J. B., and White, S. L. (1982). Carbohydrate Moieties of Glycoproteins a Re-evaluation of Their Function. *Biochim. Biophys. Acta* 650, 209–232. doi:10.1016/0304-4157(82)90017-X

Orlean, P. (2012). Architecture and Biosynthesis of the *Saccharomyces cerevisiae* Cell wall. *Genetics* 192, 775–818. doi:10.1534/genetics.112.144485

Ovacik, M., and LIn, K. (2018). Tutorial on Monoclonal Antibody Pharmacokinetics and its Considerations in Early Development. *Clin. Transl. Sci.* 11, 540–552. doi:10.1111/cts.12567

Palmberger, D., Rendić, D., Tauber, P., Krammer, F., Wilson, I. B. H., and Grabherr, R. (2011). Insect Cells for Antibody Production: Evaluation of an Efficient Alternative. *J. Biotechnol.* 153, 160–166. doi:10.1016/j.jbiotec.2011.02.009

Pandhal, J., and Wright, P. C. (2010). N-linked Glycoengineering for Human Therapeutic Proteins in Bacteria. *Biotechnol. Lett.* 32, 1189–1198. doi:10.1007/s10529-010-0289-6

Pereira, N. A., Chan, K. F., Lin, P. C., and Song, Z. (2018). The "Less-Is-More" in Therapeutic Antibodies: Afucosylated Anti-cancer Antibodies with Enhanced Antibody-Depedent Cellular Cytotoxicity. *MAbs* 10, 693–711. doi:10.1080/19420862.2018.1466767

Peters, B. P., and Aronson, N. N., Jr (1976). Reactivity of the Sialic Acid Derivative 5-Acetamido-3,5-Dideoxy-L-*Arabino*-Heptulosonic Acid in the Resorcinol and Thiobarbituric Acid Assays. *Carbohydr. Res.* 47, 345–353. doi:10.1016/s0008-6215(00)84204-4

Pindel, E. V., Kedishvili, N. Y., Abraham, T. L., Brzezinski, M. R., Zhang, J., Dean, R. A., et al. (1997). Purification and Cloning of a Broad Substrate Specificity Human Liver Carboxylesterase that Catalyzes the Hydrolysis of Cocaine and Heroin. *J. Biol. Chem.* 272, 14769–14775. doi:10.1074/jbc.272.23.14769

Pratama, F., Linton, D., and Dixon, N. (2021). Genetic and Process Engineering Strategies for Enhanced Recombinant *N*-Glycoprotein Production in Bacteria. *Microb. Cell Fact.* 20, 198. doi:10.1186/s12934-021-01689-x

Prescher, J. A., Dube, D. H., and Bertozzi, C. R. (2003). Probing Azido Sugar Metabolism *In Vivo* Using the Staudinger Ligation. *Glycobiology* 13, 894.

Pucić, M., Knezević, A., Vidic, J., Adamczyk, B., Novokmet, M., Polasek, O., et al. (2011). High Throughput Isolation and Glycosylation Analysis of IgG-Variability and Heritability of the IgG Glycome in Three Isolated Human Populations. *Mol. Cell. Proteomics* 10, M111. doi:10.1074/mcp.M111.010090

Qasba, P. K. (2015). Glycans of Antibodies as a Specific Site for Drug Conjugation Using Glycosyltranserases. *Bioconjug. Chem.* 26, 2170–2175. doi:10.1021/acs.bioconjchem.5b00173

Qun, Z., and Qiu, H. (2019). The Mechanistic Impact of N-Glycosylation on Stability, Pharmacokinetics, and Immunogenicity of Therapeutic Proteins. *J. Pharm. Sci.* 108, 1366–1377. doi:10.1016/j.xphs.2018.11.029

Rosales-Mendoza, S., Salazar-González, J., Decker, E. L., and Reski, R. (2015). Implications of Plant Glycans in the Development of Innovative Vaccines. *Expert Rev. Vaccin.* 15, 915–925. doi:10.1586/14760584.2016.1155987

Roseman, D. S., and Baenziger, J. U. (2000). Molecular Basis of Lutropin Recognition by mannose/GalNAc-4-SO_4 Receptor. *Proc. Natl. Acad. Sci. U. S. A.* 94, 9949–9954. doi:10.1073/pnas.170184597

Rup, B., Alon, S., Amit-Cohen, B.-C., Almon, E. B., Chertkoff, R., Tekoah, Y., et al. (2017). Immunogenicity of Glycans on Biotherapeutic Drugs Produced in Plant Expression Systems - the Taliglucerase Alfa story. *PLoS One* 12, e0186211. doi:10.1371/journal.pone.0186211

Ryman, J. T., and Meibohm, B. (2017). Pharmacokinetics of Monoclonal Antibodies. *CPT Pharmacometrics Syst. Pharmacol.* 6, 576–588. doi:10.1002/psp4.12224

Saeui, C. T., Cho, K.-C., Dharmarha, V., Nairn, A. V., Galizzi, M., Shah, S. R., et al. (2020). Cell Line-, Protein-, and Sialoglycosite-specific Control of Flux-Based Sialylation in Human Breast Cells: Implications for Cancer Progression. *Front. Chem.* 8, 13. doi:10.3389/fchem.2020.00013

Saez-Valero, J., Barquero, M., Marcos, A., McLean, C. A., and Small, D. H. (2000). Altered Glycosylation of Acetylcholinesterase in Lumbar Cerebrospinal Fluid of Patients with Alzheimer's Disease. *J. Neurol. Neurosurg. Psych.* 69, 664–667. doi:10.1136/jnnp.69.5.664

Sampathkumar, S.-G., Li, A. V., Jones, M. B., Sun, Z., and Yarema, K. J. (2006). Metabolic Installation of Thiols into Sialic Acid Modulates Adhesion and Stem Cell Biology. *Nat. Chem. Biol.* 2, 149–152. doi:10.1038/nchembio770

Samraj, A. N., Pearce, O. M. T., Läubli, H., Crittenden, A. N., Bergfeld, A. K., Banda, K., et al. (2015). A Red Meat-Derived Glycan Promotes Inflammation and Cancer Progression. *Proc. Natl. Acad. Sci. U. S. A.* 112, 542–547. doi:10.1073/pnas.1417508112

Sanchez-Garcia, L., Martin, L., Magues, R., Ferrer-Miralles, N., Vázquez, E., and Villaverde, A. (2016). Recombinant Pharmaceuticals from Microbial Cells: a 2015 Update. *Microb. Cell Fact.* 15, 33. doi:10.1186/s12934-016-0437-3

Sarkar, A. K., Fritz, T. A., Taylor, W. H., and Esko, J. D. (1995). Disaccharide Uptake and Priming in Animal Cells: Inhibition of Sialyl Lewis X by Acetylated Gal β1,4GalNAc B-Onaphthalenemethanol. *Proc. Natl. Acad. Sci. U. S. A.* 92, 3323–3327. doi:10.1073/pnas.92.8.3323

Sato, Y., and Beutler, E. (1993). Binding, Internalization, and Degradation of Mannose-Terminated Glucocerebrosidase by Macrophages. *J. Clin. Invest.* 91, 1909–1917. doi:10.1172/JCI116409

Sawa, M., Hsu, T.-L., Itoh, T., Sugiyama, M., Hanson, S. R., Vogt, P. K., et al. (2006). Glycoproteomic Probes for Fluorescent Imaging of Fucosylated Glycans *In Vivo*. *Proc. Natl. Acad. Sci. U. S. A.* 103, 12371–12376. doi:10.1073/pnas.0605418103

Saxon, E., and Bertozzi, C. R. (2000). Cell Surface Engineering by a Modified Staudinger Reaction. *Science* 287, 2007–2010. doi:10.1126/science.287.5460.2007

Schillberg, S., Raven, N., Spiegel, H., Rasche, S., and Buntru, M. (2019). Critical Analysis of the Commercial Potential of Plants for the Production of Recombinant Proteins. *Front. Plant Sci.* 10, 720. doi:10.3389/fpls.2019.00720

Schlesinger, P. H., Coebber, T. W., Mandell, B. F., White, R., DeSchryver, C., Rodman, J. S., et al. (1978). Plasma Clearance of Glycoproteins with Terminal Mannose and *N*-Acetylglycosamine by Liver Non-parenchymal Cell. Studies with B-Glucuronidase, *N*-Acetyl-β-D-Glucosaminidase, Ribonuclease B and Agalacto-Orosomucoid. *Biochem. J* 176, 103–109. doi:10.1042/bj1760103

Schneider, J. D., Castilho, A., Neumann, L., Altmann, F., Loos, A., Kannan, L., et al. (2013). Expression of Human Butyrylcholinesterase with an Engineered Glycosylation Profile Resembling the Plasma-Derived Orthologue. *Biotechnol. J.* 9, 501–510. doi:10.1002/biot.201300229

Schwartz, A. L. (1984). The Hepatic Asialoglycoprotein Receptor. *CRC Crit. Rev. Biochem.* 16, 207–233. doi:10.3109/10409238409108716

Seabright, G. E., Doores, K. J., Burton, D. R., and Max, C. (2019). Protein and Glycan Mimicry in HIV Vaccine Design. *J. Mol. Biol.* 431, 2223–2247. doi:10.1016/j.jmb.2019.04.016

Semerano, L., and Boissier, M.-C. (2009). Les anticorps dans les maladies immunes inflammatooires chroniques. *Med. Sci. (Paris).* 25, 1108–1112. doi:10.1051/medsci/200925121108

Senger, T., Schädlich, L., Gissmann, L., and Müller, M. (2009). Enhanced Papillomavirus-like Particle Production in Insect Cells. *Virology* 388, 344–353. doi:10.1016/j.virol.2009.04.004

Shaaltiel, Y., Bartfeld, D., Hashmueli, S., Baum, G., Brill-Almon, E., Galili, G., et al. (2007). Production of Glucocerebrosidase with Terminal Mannose Glycans for Enzyme Replacement Therapy of Gaucher's Disease Using a Plant Cell System. *Plant Biotechnol. J.* 5, 579–590. doi:10.1111/j.1467-7652.2007.00263.x

Sheikh, H., Yarwood, H., Ashworth, A., and Isacke, C. M. (2000). Endo 180, an Endocytic Recycling Glycoportein Related to the Macrophage Mannose Receptor Is Expressed on Fibroblasts, Endothelial Cells and Macrophages and Functions as a Lectin Receptor. *J. Cell Sci.* 113, 1021–1032. doi:10.1242/jcs.113.6.1021

Shi, D., Yang, J., Yang, D., LeCluyse, E. L., Black, C., You, L., et al. (2006). Anti-influenza Prodrug Oseltamivir Is Activated by Carboxylesterase Human Carboxylesterase 1, and the Activation Is Inhibited by Antiplatelet Agent Clopidogrel. *J. Pharmacol. Exp. Ther.* 19, 1477–1484. doi:10.1124/jpet.106.111807

Shuster, S., Frost, G. I., Csoka, A. B., Formby, B., and Stern, R. (2002). Hyaluronidase Reduces Human Breast Cancer Xenografts in SCID Mice. *Int. J. Cancer* 102, 192–197. doi:10.1002/ijc.10668

Singh, R., Almutairi, M. M., Pacheco-Andrade, R., Mahmoud Almiahuob, M. Y., and Di Fulvio, M. (20152015). Impact of Hybrid and Complex N-Glycans on Cell Surface Targeting of the Endogenous Chloride Cotransporter Slc12a2. *Int. J. Cell Biol.*, 505294. doi:10.1155/2015/505294

Singh, R., Kumar, M., Mittal, A., and Mehta, P. K. (2016). Microbial Enzymes: Industrial Progress in 21st century. *3 Biotech.* 6, 174. doi:10.1007/s13205-016-0485-8

Song, R., Oren, D. A., Franco, D., Seaman, M. S., and Ho, D. D. (2013). Strategic Addition of an N-Linked Glycan to a Monoclonal Antibody Improves its HIV-1–Neutralizing Activity. *Nat. Biotechnol.* 31, 1047–1052. doi:10.1038/nbt.2677

Stabach, P. R., Zimmerman, K., Adame, A., Kavanagh, D., Saeui, C. T., Agatemor, C., et al. (2021). Improving the Pharmacodynamics and *In Vivo* Activity of ENPP1-Fc through Protein and Glycosylation Engineering. *Clin. Transl. Sci.* 14, 362–372. doi:10.1111/cts.12887

Stanley, P., and Patnaik, S. K. (2005). "Chinese Hamster Ovary (CHO) Glycosylation Mutants for Glycan Engineering," in *Handbook of Carbohydrate Engineering*. Editor K. J Yarema (Boca Raton, Florida: Francis & Taylor/CRC Press), 371–385.

Stanley, P., Raju, T. S., and Bhaumik, M. (1996). CHO Cells Provide Access to Novel *N*-Glycans and Developmentally Regulated Glycosyltransferases. *Glycobiology* 6, 696–699. doi:10.1093/glycob/6.7.695

Steinke, J. W., Platt-Mills, T. A. E., and Commins, S. P. (2015). The Alpha Gal story: Lessons Learned from Connecting the Dots. *J. Allergy Clin. Immunol.* 135, 589–597. doi:10.1016/j.jaci.2014.12.1947

Sudo, M., Yamaguchi, Y., Späth, P. J., Matsumoto-Morita, K., Ong, B. K., Shahrizala, N., et al. (2014). Different IVIG Glycoforms Affect *In Vitro* Inhibition of Anti-ganglioside Antibody-Mediated Complement Deposition. *PLoS One* 9, e017772. doi:10.1371/journal.pone.0107772

Sukenik, S. C., Karuppanan, K., Li, Q., Lebrilla, C. B., Nandi, S., and McDonald, K. A. (2018). Transient Recombinant Protein Production in Glycoengineered *Nicotiana Benthamiana* Cell Suspension Culture. *Int. J. Mol. Sci.* 19, 1205. doi:10.3390/ijms19041205

Szolnoky, G., Bata-Csörgö, Z., Kenderessy, A. S., Kiss, M., Pivarcsi, A., Novák, Z., et al. (2001). A Mannose-Binding Receptor Is Expressed on Human Keratinocytes and Mediates Killing of Candida Albicans. *J. Investig. Dermatol.* 117, P205–P213. doi:10.1046/j.1523-1747.2001.14071.x

Szymanski, C. M., Yao, R., Ewing, C. P., Trust, T. J., and Guerry, P. (2002). Evidence for a System of General Protein Glycosylation in *Campylobacter Jejuni*. *Mol. Microbiol.* 32, 1022–1030. doi:10.1046/j.1365-2958.1999.01415.x

Tang, F., Shi, W., and Huang, W. (2019). Homogeneous Antibody-Drug Conjugates via Glycoengineering. *Methods Mol. Biol.* 2033, 221–238. doi:10.1007/978-1-4939-9654-4_15

Tangvoranuntakul, P., Gagneux, P., Diaz, S., Bardor, M., Varki, N., Varki, A., et al. (2003). Human Uptake and Incorporation of an Immunogenic Nonhuman Dietary Sialic Acid. *Proc. Natl. Acad. Sci. U. S. A.* 100, 12045–12050. doi:10.1073/pnas.2131556100

Tariq, N., Khan, M. A., Ijaz, B., Ahmed, N., ur Rahman, Z., Latif, M. S., et al. (2018). Glycosylation of Recombinant Anticancer Therapeutics in Different Expression Systems with Emerging Technologies. *Cancer Res.* 78, 2787–2789. doi:10.1158/0008-5472.can-18-0032

Tekoah, Y., Tzaban, S., Kizhner, T., Hainrichson, M., Gantman, A., Golembo, M., et al. (2013). Glycosylation and Functionality of Recombinant B-Glycocerebrosidase from Various Production Systems. *Biosci. Rep.* 33, e00071. doi:10.1042/BSR20130081

Tibbitts, J., Canter, D., Graff, R., Smith, A., and Khawli, L. A. (2016). Key Factors Influencing ADME Properties of Therapeutic Proteins: A Need for ADME Characterization in Drug Discovery and Development. *MAbs* 8, 229–245. doi:10.1080/19420862.2015.1115937

Tong, L., Baskaran, G., Jones, M. B., Rhee, J. K., and Yarema, K. J. (2003). "Glycosylation Changes as Markers for the Diagnosis and Treatment of Human Disease," in *Biochemical and Genetic Engineering Reviews. Andover*. Editor S. Harding (Hampshire, UK: Intercept Limited), 199–244. doi:10.1080/02648725.2003.10648044

Top, O., Geisen, U., Decker, E. L., and Reski, R. (2019). Critical Evaluation of Strategies for the Production of Blood Coagulation Factors in Plant-Based Systems. *Front. Plant Sci.* 10, 261. doi:10.3389/fpls.2019.00261

Torres-Pérez, S. A., Torres-Pérez, C. E., and Pedraza-Escalona, M. (2020). Glycosylated Nanoparticles for Cancer-Targeted Drug Delivery. *Front. Oncol.* 10, 605037. doi:10.3389/fonc.2020.605037

Toth, A. M., Kuo, C.-W., Khoo, K.-H., and Jarvis, D. L. (2014). A New Insect Cell Glycoengineering Approach Provides Baculovirus-Inducible Glycogene Expression and Increases Human-type Glycosylation Efficiency. *J. Biotechnol.* 182-183, 19–29. doi:10.1016/j.jbiotec.2014.04.011

Tsao, C., Green, P., Odlind, B., and Brater, D. C. (1991). Pharmacokinetics of Recombinant Human Superoxide Dismutase in Healthy Volunteers. *Clin. Pharmacol. Ther.* 50, 713–720. doi:10.1038/clpt.1991.211

Turner, M. R., and Balu-Iyer, S. V. (2018). Challenges and Opportunities for the Subcutaneous Delivery of Therapeutic Proteins. *J. Pharm. Sci.* 107, 1247–1260. doi:10.1016/j.xphs.2018.01.007

Ulloa-Aguirre, A., Reiter, E., and Crépieux, P. (2018). FSH Receptor Signaling: Complexity of Interactions and Signal Diversity. *Endocrinology* 159, 3020–3035. doi:10.1210/en.2018-00452

Vaccaro, C., Zhou, J., Ober, R. J., and Ward, E. S. (2005). Engineering the Fc Region of Immunoglobulin G to Modulate *In Vivo* Antibody Levels. *Nat. Biotechnol.* 23, 1283–1288. doi:10.1038/nbt1143

van de Bovenkamp, F. S., Derksen, N. I. L., Ooijevaar-de Heer, P., van Schie, K. A., Kruithof, S., Berkowska, M. A., et al. (2018). Adaptive Antibody Diversification through N-Linked Glycosylation of the Immunoglobulin Variable Region. *Proc. Natl. Acad. Sci. U. S. A.* 115, 1901–1906. doi:10.1073/pnas.1711720115

van de Bovenkamp, F. S., Hafkenscheid, L., Rispens, T., and Rombouts, Y. (2016). The Emerging Importance of IgG Fab Glycosylation in Immunity. *J. Immunol.* 196, 1435–1441. doi:10.4049/jimmunol.1502136

Velan, B., Kronman, C., Ordentlich, A., Flashner, Y., Leitner, M., Cohen, S., et al. (1993). *N*-glycosylation of Human Acetylcholinesterase: Effects on Activity, Stability and Biosynthesis. *Biochem. J* 296, 649–656. doi:10.1042/bj2960649

Vocadlo, D. J., Hang, H. C., Kim, E.-J., Hanover, J. A., and Bertozzi, C. R. (2003). A Chemical Approach for Identifying O-GlcNAc-Modified Proteins in Cells. *Proc. Natl. Acad. Sci. U. S. A.* 100, 9116–9121. doi:10.1073/pnas.1632821100

von Bergen Granell, A. E., Palter, K. B., Akan, I., Aich, U., Yarema, K. J., Betenbaugh, M. J., et al. (2011). DmSAS Is Required for Sialic Acid Biosynthesis in Cultured Drosophila Third Instar Larvae CNS Neurons. *ACS Chem. Biol.* 6, 1287–1295. doi:10.1021/cb200238k

Wacker, M., Linton, D., Hitchen, P. G., Nita-Lazar, M., Haslam, S. M., North, S. J., et al. (2002). N-linked Glycosylation in *Campylobacter Jejuni* and it Functional Transfer into *E. coli*. *Science* 298, 1790–1793. doi:10.1126/science.298.5599.1790

Wang, H., and Mooney, D. J. (2020). Metabolic Glycan Labelling for Cancer-Targeted Therapy. *Nat. Chem.* 12, 1102–1114. doi:10.1038/s41557-020-00587-w

Wang, L.-X., Tong, X., Li, C., Giddens, J. P., and Li, T. (2019). Glycoengineering of Antibodies for Modulating Functions. *Annu. Rev. Biochem.* 88, 433–459. doi:10.1146/annurev-biochem-062917-012911

Wang, Q., Chung, C.-Y., Chough, S., and Betenbaugh, M. J. (2018). Antibody Glycoengineering Strategies in Mammalian Cells. *Biotechnol. Bioeng.* 115, 1378–1393. doi:10.1002/bit.26567

Wang, S., Voronin, Y., Zhao, P., Ishihara, M., Mehta, N., Porterfield, M., et al. (2020). Glycan Profiles of Gp120 Protein Vaccines from Four Major HIV-1 Subtypes Produced from Different Host Cell Lines under Non-GMP oGMP or GMP Conditions. *J. Virol.* 94. doi:10.1128/jvi.01968-19

Wang, Z., Du, J., Che, P.-L., Meledeo, M. A., and Yarema, K. J. (2009). Hexosamine Analogs: from Metabolic Glycoengineering to Drug Discovery. *Curr. Opin. Chem. Biol.* 13, 565–572. doi:10.1016/j.cbpa.2009.08.001

Washburn, N., Schwab, I., Ortiz, D., Bhatnagar, N., Lansing, J. C., Medeiros, A., et al. (2015). Controlled Tetra-Fc Sialylation of IVIG Results in a Drug Candidate with Consistent Anti-inflammatory Activity. *Proc. Natl. Acad. Sci. U. S. A.* 112, E1297–E1306. doi:10.1073/pnas.1422481112

Wasley, L. C., Timony, G., Murtha, P., Stoudemire, J., Dorner, A. J., Caro, J., et al. (1991). The Importance of N- and O-Linked Oligosaccharides for the Biosynthesis and *In Vitro* and *In Vivo* Biologic Activities of Erythropoietin. *Blood* 77, 2624–2632. doi:10.1182/blood.v77.12.2624.bloodjournal77122624

Wasserman, R. L. (2017). Recombinant Human Hyaluronidase-Facilitated Subcutaneous Immunoglobin Infusion in Primary Immundeficiency Disease. *Immunotherapy* 9, 1035–1050. doi:10.2217/imt-2017-0092

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., et al. (2018). SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res.* 46, W296–W303. doi:10.1093/nar/gky427

Wayman, J. A., Glasscock, C., Mansell, T. J., DeLisa, M. P., and Varner, J. D. (2019). Improving Designer Glycan Production in *Escherichia coli* through Model-Guided Metabolic Engineering. *Metab. Eng. Commun.* 9. doi:10.1016/j.mec.2019.e00088

Wei, X., Decker, J. M., Wang, S., Hui, H., Kappes, J. C., Wu, X., et al. (2003). Antibody Neutralization and Escape by HIV-1. *Nature* 422, 307–312. doi:10.1038/nature01470

Weigel, P. H. (1994). Galactosyl and N-Acetylgalactosaminyl Homeostasis: a Function for Mammalian Asialoglycoprotein Receptors. *Bioessays* 16, 519–524. doi:10.1002/bies.950160713

Weikert, T., Ebert, C., Rasched, I., and Layer, P. G. (1994). Novel Inactive and Distinctively Glycosylated Forms of Butyrylcholinesterase from Chicken Serum. *J. Neurochem.* 63, 318–325. doi:10.1046/j.1471-4159.1994.63010318.x

Werz, D. B., Ranzinger, R., Herget, S., Adibekian, A., Von der Lieth, C. W., and Seeberger, P. H. (2007). Exploring the Structural Diversity of Mammalian Carbohydrates ("glycospace") by Statistical Databank Analysis. *ACS Chem. Biol.* 2, 685–691. doi:10.1021/cb700178s

Whatcott, C. J., Han, H., Posner, R. G., Hostetter, G., and Von Hoff, D. (2011). Targeting the Tumor Microenvironment in Cancer: Why Hyaluronidase Deserves a Second Look. *Cancer Disc* 1, 291–296. doi:10.1158/2159-8290.CD-11-0136

Willey, K. P. (1999). An Elusive Role for Glycosylation in the Structure and Function of Reproductive Hormones. *Hum. Reprod. Update* 5, 330–355. doi:10.1093/humupd/5.4.330

Wratil, P. R., Horstkorte, R., and Reutter, W. (2016). Metabolic Glycoengineering with N-Acyl Side Chain Modified Mannosamines. *Angew. Chem. Int. Ed. Eng.* 55, 9482–9512. doi:10.1002/anie.201601123

Xu, M. L., Luk, W. K. W., Lau, K. M., Bi, C. W. C., Cheng, A. W. M., Gong, A. G. W., et al. (2015). Three N-Glycosylation Sites of Human Acetylcholinesterase Shares Similar Glycan Composition. *J. Mol. Neurosci.* 57, 486–491. doi:10.1007/s12031-015-0629-z

Yarema, K. J., Goon, S., and Bertozzi, C. R. (2001). Metabolic Selection of Glycosylation Defects in Human Cells. *Nat. Biotechnol.* 19, 553–558. doi:10.1038/89305

Yarema, K. J., Mahal, L. K., Bruehl, R. E., Rodriguez, E. C., and Bertozzi, C. R. (1998). Metabolic Delivery of Ketone Groups to Sialic Acid Residues. Application to Cell Surface Glycoform Engineering. *J. Biol. Chem.* 273, 31168–31179. doi:10.1074/jbc.273.47.31168

Yates, L. E., Mills, D. C., and DeLisa, M. P. (2021). Bacterial Glycoengineering as a Biosynthetic Route to Customized Glycomolecules. *Adv. Biochem. Eng. Biotechnol.* 175, 167–200. doi:10.1007/10_2018_72

Yee, C. M., Zak, A. J., Hill, B. D., and Wen, F. (2018). The Coming Age of Insect Cells for Manufacturing and Development of Protein Therapeutics. *Ind. Eng. Chem. Res.* 57, 10061–10070. doi:10.1021/acs.iecr.8b00985

Yin, B., Gao, Y., Chung, C. Y., Yang, S., Blake, E., Stuczynski, M. C., et al. (2015). Glycoengineering of Chinese Hamster Ovary Cells for Enhanced Erythropoietin N-Glycan Branching and Sialylation. *Biotechnol. Bioeng.* 112, 2343–2351. doi:10.1002/bit.25650

Yin, B., Wang, Q., Chung, C. Y., Bhattacharya, R., Ren, X., Tang, J., et al. (2017). A Novel Sugar Analog Enhances Sialic Acid Production and Biotherapeutic Sialylation in CHO Cells. *Biotechnol. Bioeng.* 114, 1899–1902. doi:10.1002/bit.26291

Ząbczyńska, M., Polak, K., Kozłowska, K., Sokołowski, G., and Pocheć, E. (2020). The Contribution of IgG Glycosylation to Antibody-dependent Cell-Mediated Cytotoxicity (ADCC) and Complement-dependent Cytotoxicity (CDC) in Hashimoto's Thyroditis. *Biomolecules* 10. doi:10.3390/biom10020171

Zafir, A., Readnower, R., Long, B. W., McCracken, J., Aird, A., Alvarez, A., et al. (2013). Protein O-GlcNAcylation Is a Novel Cytoprotective Signal in Cardiac Stem Cells. *Stem Cells* 31, 765–775. doi:10.1002/stem.1325

Zeng, Y., Tang, F., Shi, W., Dong, Q., and Huang, W. (2022). Recent Advances in Synthetic Glycoengineering for Biological Applications. *Curr. Opin. Biotechnol.* 74, 247–255. doi:10.1016/j.copbio.2021.12.008

Zhang, L., Mann, M., Syed, Z. A., Reynolds, H. M., Tian, E., Samara, N. L., et al. (2021). Furin Cleavage of the SARS-CoV-2 Spike Is Modulated by *O*-Glycosylation. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2109905118. doi:10.1073/pnas.2109905118

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Chemistry

ORIGINAL RESEARCH
published: 27 April 2022
doi: 10.3389/fchem.2022.871509

Novel Insights Into the Sulfated Glucuronic Acid-Based Anti-SARS-CoV-2 Mechanism of Exopolysaccharides From Halophilic Archaeon *Haloarcula hispanica*

Yueqiang Xu[1†], Yan Li[2†], Xin You[3†], Caixia Pei[4†], Zhuo Wang[1], Siming Jiao[1], Xin Zhao[2], Xuan Lin[1], Yang Lü[4], Cheng Jin[4], George Fu Gao[2], Jianjun Li[1], Qi Wang[2]* and Yuguang Du[1]**

[1]State Key Laboratory of Biochemical Engineering, National Engineering Research Center for Biotechnology (Beijing), Key Laboratory of Biopharmaceutical Production & Formulation Engineering, PLA, Institute of Processing and Engineering, Chinese Academy of Sciences, Beijing, China, [2]CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China, [3]Lung Cancer Translational Medicine Center, The Second Affiliated Hospital of Dalian Medical University, Dalian, China, [4]State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China

The pandemic caused by SARS-CoV-2 is the most widely spread disease in the 21st century. Due to the continuous emergence of variants across the world, it is necessary to expand our understanding of host–virus interactions and explore new agents against SARS-CoV-2. In this study, it was found exopolysaccharides (EPSs) from halophilic archaeon *Haloarcula hispanica* ATCC33960 can bind to the spike protein of SARS-CoV-2 with the binding constant K_D of 2.23 nM, block the binding of spike protein to Vero E6 and bronchial epithelial BEAS-2B cells, and inhibit pseudovirus infection. However, EPSs from the gene deletion mutant *△HAH_1206* almost completely lost the antiviral activity against SARS-CoV-2. A significant reduction of glucuronic acid (GlcA) and the sulfation level in EPSs of *△HAH_1206* was clearly observed. Our results indicated that sulfated GlcA in EPSs is possible for a main structural unit in their inhibition of binding of SARS-CoV-2 to host cells, which would provide a novel antiviral mechanism and a guide for designing new agents against SARS-CoV-2.

Keywords: sulfated glucuronic acid, SARS-CoV-2, exopolysaccharide, archaea, *Haloarcula hispanica*

OPEN ACCESS

Edited by:
Zhongping Tan,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

Reviewed by:
Essa M. Saied,
Humboldt University of Berlin, Germany
Decai Xiong,
Peking University, China

***Correspondence:**
Jianjun Li
ygdu@ipe.ac.cn
Qi Wang
wqdlmu@163.com
Yuguang Du
jjll@ipe.ac.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Chemical Biology, a section of the journal Frontiers in Chemistry

Received: *08 February 2022*
Accepted: *25 March 2022*
Published: *27 April 2022*

Citation:
Xu Y, Li Y, You X, Pei C, Wang Z, Jiao S, Zhao X, Lin X, Lü Y, Jin C, Gao GF, Li J, Wang Q and Du Y (2022) Novel Insights Into the Sulfated Glucuronic Acid-Based Anti-SARS-CoV-2 Mechanism of Exopolysaccharides From Halophilic Archaeon Haloarcula hispanica. Front. Chem. 10:871509. doi: 10.3389/fchem.2022.871509

INTRODUCTION

SARS-CoV-2 represents one of the most fast-spreading viruses in the 21st century (Tan et al., 2020; Wang C et al., 2020), and the pandemic has swept across the world. More than 300 million people were infected, and five million were killed by the virus. Several variants of the virus have been designated as variants of concerns (VOCs), including B.1.1.7 (alpha), B.1.351 (beta), P.1 (gamma), B.1.617.2 (delta), and B.1.1.529 (omicron). Extensive studies have focused on vaccines neutralizing antibodies and antiviral chemical compounds. However, these efforts were challenged by the more virulent and easily transmitted SARS-CoV-2 variants. Epecially for the variant of omicron, both vaccines and neutralizing antibodies displayed reduced neutralizing titers (Cao et al., 2021; Lu et al., 2021). In addition to the urgent demand of preventive and therapeutic strategies, it is also necessary to deeply understand the interaction between virus and host cells in nature.

The receptor-binding domain (RBD) of spike protein of SARS-CoV-2 can bind to angiotensin-converting enzyme 2 (ACE2) on the surface of host cells specifically, then the activated proteases such as furin, transmembrane serine protease 2 (TMPRSS2), or cathepsin L cleave the spike protein, and finally the HR1 and HR2 regions in the S2 subunit interact with the cell membrane to mediate fusion, resulting in the release of the viral genome into the cytoplasm (Shang et al., 2020; Saied et al., 2021). The 3D structures of SARS-CoV-2 virus, S protein, and human ACE2 (hACE2) have been determined (Walls et al., 2020; Wang Q et al., 2020). The spike protein of SARS-CoV-2 is heavily glycosylated with 22 *N*-glycosylation sites and 17 *O*-glycosylation sites (Shajahan et al., 2020; Tian et al., 2021), whereas hACE2 employs seven *N*-glycosylation sites and one *O*-glycosylation site (Shajahan et al., 2021). In addition to ACE2, there might be multiple receptors or co-receptors in host cells for SARS-CoV-2 infection, including heparan sulfate (HS) on the host cell surface, and immune mannose receptors of DC-SIGN, L-SIGN, MGL, Siglec-9, and Siglec-10 in cells can also bind with the spike protein (Chiodo et al., 2020; Clausen et al., 2020; Gao et al., 2020). Therefore, an attractive approach to fight against SARS-CoV-2 is to block or interfere with virus attachment and binding to host cells.

Glycans are one of the most important molecules in cells, which play critical roles in virus assembly, attachment, recognition, entry, and immune escape (Watanabe et al., 2019). Viruses can employ glycans as receptors to infect hosts, such as human influenza A viruses recognizing α2, 6-linked sialic acid, and avian influenza A viruses, showing preference for α2, 3-linked sialic acid (Kumlin et al., 2008; Shi et al., 2014; Li et al., 2017). Furthermore, HCoV-OC43, HCoV-HKU1, BCoV, and PHEV can use 9-*O*-acetyl-sialic acid as a receptor (Tortorici et al., 2019). The fact that 9-*O*-acetyl-sialic acid can prevent MERS-CoV from binding to host cells means "glycan inhibitors" might be ideal candidate drugs to fight against virus infection (Li et al., 2017). Several teams have reported that glycans can be used as anti-SARS-CoV-2 agents, such as marine sulfated polysaccharides (Jin et al., 2020; Kwon et al., 2020; Song et al., 2020; Dwivedi et al., 2021; Zhang et al., 2022) and HS (Clausen et al., 2020; Kim et al., 2020; Hao et al., 2021). HS can bind to S protein of SARS-CoV-2 and block binding of the spike protein to hACE2 and can impede the infection by pseudovirus and authentic SARS-CoV-2. HS possesses broad-spectrum activities against a multitude of distinct viruses, including flaviviruses, herpes, influenza, HIV, and Coronaviridae. Recent studies had shown that HS can inhibit the invasion of SARS-CoV-2 depending on its chain length and sulfation pattern (Kim et al., 2020; Mycroft-West et al., 2020; Hao et al., 2021; Liu et al., 2021). For instance, *N*-desulfated HP, 2-*O*-desulfated HP, and 6-*O*-desulfated HP were unable to compete with immobilized HP for binding to SARS-CoV-2 (Kim et al., 2020), while an octasaccharide composed of IdoA2S-GlcNS6S can inhibit spike–heparin interaction with an IC_{50} of 38 nM, and Tris HS hexasaccharide [GlcA (2S)-GlcNS (6S)] also can bind to the trimeric spike protein of SARS-CoV-2 (Liu et al., 2021). Therefore, it is important to elucidate the critical structures responsible for antiviral activity in glycans.

Archaea are one of the most primitive organisms on the Earth and usually live in extreme environments such as saline lakes, Antarctic ecosystems, geothermal springs, and deep sea. Exopolysaccharides (EPSs) from extremophiles can be applied in food, pharmaceutical, and cosmetics industries (Nicolaus et al., 2010). *Haloarcula hispanica* ATCC33960 is an extremely halophilic archaeon isolated from a solar saltern in Spain, which can produce sulfated EPSs (Lü et al., 2017). In this study, our results showed that EPSs from *H. hispanica* ATCC33960 can bind to the spike protein of SARS-CoV-2 inhibit the binding of spike protein to Vero-E6 and bronchial epithelial BEAS-2B cells, and impede the infection of SARS-CoV-2 pseudovirus. As far as we know this is the first discovery that EPSs from archaea can inhibit SARS-CoV-2 infection *in vitro*. Further analysis showed that the GlcA content and the sulfation level of EPSs play essential roles in anti-SARS-CoV-2 activity.

MATERIALS AND METHODS

Strain Culture and Exopolysaccharides Preparation

Haloarcula hispanica ATGG33960 was cultured in AS-168 medium to late stationary phase (5 g/L Bacto casamino acids, 5 g/L Bacto yeast extract, 1 g/L sodium glutamate, 3 g/L trisodium citrate, 20 g/L $MgSO_4.7H_2O$, 2 g/L KCl, 200 g/L NaCl, 50 mg/L $FeSO_4.7H_2O$, 0.36 mg/L $MnCl_2.4H_2O$, and pH 7.0). EPSs were first precipitated from the supernatant by 4-fold volume of ethanol and then dialyzed against water. The dialyzed solution was treated with Benzonase nuclease and protease K subsequently at 37°C for 12 h. After concentrated with the 100 kDa ultrafiltration membrane, the EPSs solution was lyophilized. Crude EPSs were further sequentially purified by a DEAE-Sepharose Fast Flow and Sephacryl S-400/HR column as described. The concentration of EPSs was measured by the phenol–sulfuric acid method and determined at A490 (Lü et al., 2017).

Homogeneity and Molecular Weight

For molecular weight (MW) measurement of EPSs, the samples were analyzed by high-performance gel permeation chromatography (HPGPC) with a TSK GEL GMPWXL column, and the polysaccharides were eluted with a mobile phase containing ddH_2O at a flow rate of 0.5 mL/min and detected by using an evaporative light-scattering detector (ELSD).

Sulfate Content Comparison

To evaluate the sulfate content of the polysaccharides, 10 μg of EPSs was run in 7.5% (w/v) SDS-PAGE, then the gel was stained with 0.5% (w/v) methylene blue in 3% (v/v) acetic acid, and SO_4^{2-} in EPSs can be stained with methylene blue (Lü et al., 2017).

Monosaccharide Composition Analysis

For monosaccharide analysis, EPSs (5 mg) were hydrolyzed in 2 M trifluoroacetic acid at 120°C for 2 h, and then the solution was

evaporated to dryness by using a rotary evaporator after adding 2-fold volume of methanol. Hydrolyzed EPSs were dissolved in 1 mL ddH_2O, and then the samples were analyzed in HPAEC-PAD with a CarboPac PA-10 column. For analysis of neutral sugars, the elution condition is 18 mM NaOH at a flow rate of 1.0 mL/min, while the acidic sugars were analyzed by 100 mM NaOH and 100 mM CH_3COONa at a flow rate of 1.0 mL/min. Mannose (Man), galactose (Gal), glucose (Glc), D-glucuronic acid (GlcA), and D-galacturonic acid (GalA) were used as standards.

Binding Assay of S Protein and Exopolysaccharides

The recombinant spike RBD protein of SARS-CoV-2 expressed in HEK-293 cells was purchased from BioRobust (Shenzhen, China). The binding of glycans to RBD of spike protein was first evaluated by Monolith NT.115 (Nanotemper): 50 μg of RBD was labeled with a Monolith RED-NHS protein Labeling Kit, then a series of EPSs solutions of *H. hispanica* were prepared by a 2-fold serial-dilution method, and the binding affinity between RBD protein and EPSs was evaluated in Monolith NT.115 after co-incubated for 10 min. When the binding capability of glycan was observed from Monolith NT.115, the binding kinetics between glycan and the RBD protein was further determined by biolayer interferometry (BLI)-based assay with Octet R8 (Sartorius): The Ni-NTA sensor was coated with 5 μg/mL RBD for 10 min, then the EPSs solution was diluted by 2-fold with buffer [10 mM PBS +0.02% Tween 20 (w/w)], and the diluted EPSs solution was incubated with the sensors coated with the RBD protein for 2 min. After dissociated for another 2 min, the binding constant K_D between EPSs and RBD was measured.

Cell Culture and Cell Viability Assay

The human bronchial epithelial BEAS-2B cells were cultured in RPMI 1640 with L-glutamine (Corning, 10-040-CV) supplemented with 10% FBS, 200 mg/mL streptomycin and 200 IU/mL penicillin at 37°C, 5% CO_2. The African green monkey kidney Vero E6 cells were maintained in DMEM (Gibco, 11965092) supplemented with 10% fetal bovine sera (FBS), 200 mg/mL streptomycin, and 200 IU/mL penicillin at 37°C, 5% CO_2. For cell viability assay, the cells were seeded in a 96-well plate with 1×10^4 cells/well and cultured for 24 h, then the supernatants were discarded, and 150 μl of serial-diluted EPSs in culture medium was added to the cells and incubated for another 24 h. Subsequently, 15 μl of MTT [3-(4,5-dimethyl-2-thiazolyl)-2,5-diphenyl-2-H-tetrazolium bromide, 5 mg/mL] was added to each well and incubated for 4 h, then the supernatants were discarded and 100 μl of DMSO (dimethyl sulfoxide) was added to dissolve the purple precipitate. Finally, the 96-well plate was scanned with Infinite M200 Pro (TECAN) at 405 nm. Data were expressed as the means ± standard errors of the means (SEM). *p* values were analyzed by unpaired *t* test with GraphPad 5.

Cell-Binding Assay Against S Protein

The cells of BEAS-2B and Vero E6 were seeded into 24-well glass bottom plates (Cellvis) with 1×10^5 cells/well and cultured for 48 h. The cells were washed with PBS three times, and then fixed with 4% (w/v) of paraformaldehyde (PFA). Then, the cells were incubated with 200 μl solution containing EPSs (2 μg/well) and RBD of spike protein (1 μg/well) for 2 h at 37°C. The RBD protein binding to the cells can be detected by SARS-CoV-2 spike-neutralizing antibody, mouse mAb (SinoBiological, 40592-MM57), and Alexa Fluor 488™ goat anti-mouse IgG (H + L) (Invitrogen, A-11001). The nuclear DNA of cells was stained using 4',6-diamidino-2-phenylindole (DAPI, 1 μg/mL). Images were captured with a Leica TCS SP8 STED confocal microscope, and data were analyzed using LAS X software (Leica).

Preparation of Pseudotyped Virus and Neutralization Assay

The construction of VSV-ΔG-GFP-based SARS-CoV-2 pseudotyped virus was mentioned in previous work with slight modifications (Zhang Z et al., 2021; Zhao et al., 2021). The codon-optimized wild-type SARS-CoV-2 (Wuhan-1 reference strain) was constructed into the pCAGGS vector. The construct (30 μg) was transfected into HEK 293T cells. VSV-ΔG-G-GFP pseudovirus was added 24 h after the transfection and removed after 1-h incubation. Media were replaced with fresh complete DMEM medium supplemented with anti-VSV-G antibody (I1HybridomaATCC® CRL2700™). Supernatants were collected after another 30-h incubation, passed through a 0.45-μm filter (Millipore, SLHP033RB), aliquoted, and stored at −80°C.

Neutralization was measured by the reduction in GFP expression as described previously (Zhang S et al., 2021). One day before neutralization assay, Vero E6 cells were seeded into 48-well plates with 1×10^5 cells/well and incubated at 37°C. Pseudovirus was incubated with 3-fold serially diluted EPSs for 1 h in advance, together with the virus control and cell control. Then, pseudovirus was transferred to pre-plated Vero E6 cells washed by fresh DMEM without FBS, followed by incubation at 37°C for 24 h. After lysed by trypsin, the GFP positive cells were measured with an FACSCanto II flow cytometer (BD Biosciences, United States).

RESULTS

Binding of Exopolysaccharides With the Receptor-Binding Domain of Spike Protein

It has been found that several sulfated polysaccharides could bind with RBD of spike protein including HS, fucoidan, carrageenan, and sulfated polysaccharide from sea cucumber, leading to their interference with binding of RBD to ACE2 at different extents (Clausen et al., 2020; Kim et al., 2020; Song et al., 2020; Hao et al., 2021). Therefore, sulfated polysaccharides were potential candidates against SARS-CoV-2. EPSs from *H. hispanica* were sulfated too, so their binding with RBD was investigated.

Binding of EPSs from *H. hispanica* to SARS-CoV-2 RBD protein was first analyzed by Microscale thermophoresis

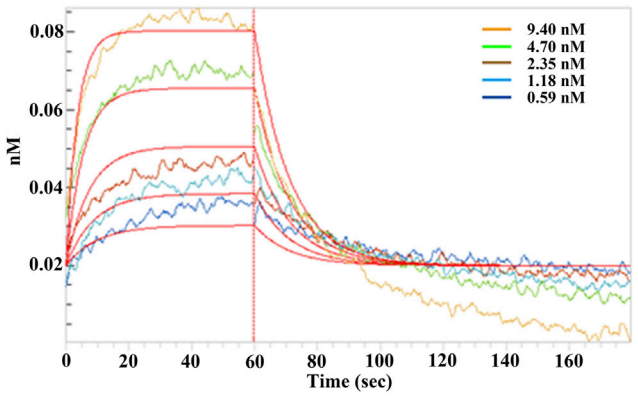


FIGURE 1 | Binding affinity assay between SARS-CoV-2 RBD and EPSs from *H. hispanica*. Serially diluted EPSs solutions were incubated with RBD (50 μg), which was coated on a Ni-NTA sensor for 2 min, and the affinity kinetics were analyzed after being dissociated for another 2 min. The binding kinetics between EPSs and RBD were determined by using the biolayer interferometry (BLI) method.

(MST). Noticeably, EPSs can bind with RBD protein well (**Supplementary Figure S1**). The binding kinetics between EPSs and RBD of spike protein were further determined by biolayer interferometry (BLI)-based assay with Octet R8 (Sartorius). The result showed that EPSs from *H. hispanica* can bind to RBD with high affinity with the calculated K_D as 2.23×10^{-9} M (**Figure 1**). EPSs from *H. hispanica* displayed good affinity to the RBD protein of SARS-CoV-2.

Cell Viability Assay

In order to assess the toxicity of EPSs from wild-type of *H. hispanica* and △*HAH_1206*, the cell viabilities were checked after EPSs were incubated with BEAS-2B or Vero E6 cells. The results showed the viabilities of BEAS-2B and Vero E6 cells were affected when the concentrations of EPSs from WT exceeded 12.5 μg/mL (**Supplementary Figure S2A**). But for EPSs from △*HAH_1206*, the viabilities of BEAS-2B and Vero E6 cells seemed not affected by serially diluted EPSs, only Vero E6 cells were slightly affected at EPSs concentration of l00 μg/mL, which meant that EPSs from △*HAH_1206* nearly lost toxicity. Therefore, the EPSs concentration was set below 12.5 μg/mL in following experiments.

Interference of Exopolysaccharides With Binding of the Receptor-Binding Domain to Cells

Interference of EPSs with the binding of RBD to cells was further investigated *via* immunofluorescence. It was observed that RBD can bind to BEAS-2B and Vero E6 cells. However, the signals from Alexa Fluor 488 were blurry when EPSs were incubated with RBD in advance. These results clearly demonstrated that EPSs could block the interaction between RBD and cells expressing hACE2 (**Figure 2**).

Anti-Infection of Pseudovirus by Exopolysaccharides

To investigate the anti-SARS-CoV-2 activity of EPSs, the inhibition effects of EPSs were determined using pseudovirus

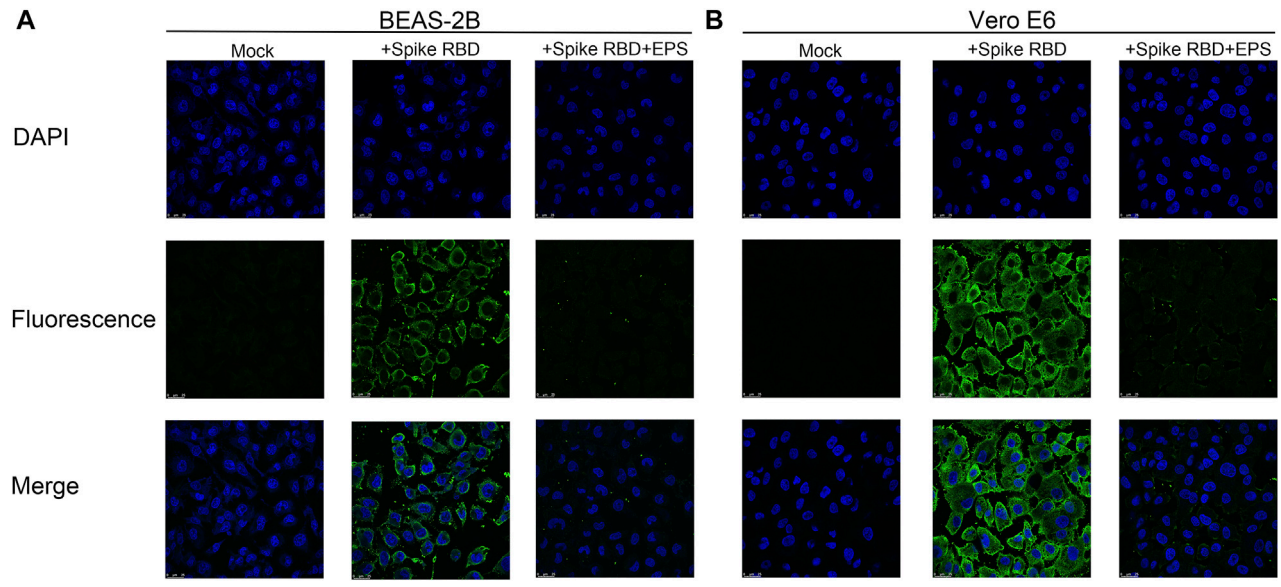


FIGURE 2 | Immunofluorescence assay of EPSs from *H. hispanica*-blocking RBD to bind with BEAS-2B and Vero E6. **(A)** Binding of RBD to BEAS-2B cells inhibited by EPSs from *H. hispanica*. **(B)** Binding of RBD to Vero E6 cells inhibited by EPSs from *H. hispanica*. **Mock**: cells detected in the absence of RBD and EPSs; + **Spike RBD**: cells detected in the presence of RBD (1 μg/well); and + **Spike RBD** + **EPSs**: cells detected in the presence of RBD (1 μg/well) and EPSs (2 μg/well). All wells were detected by immunofluorescence using SARS-CoV-2 (2019-nCoV) spike-neutralizing antibody and Alexa Fluor 488 goat anti-mouse IgG (H + L) by using a confocal microscope. The fluorescence signals were captured with an FITC channel, and the nuclear DNAs of cell were stained with DAPI.

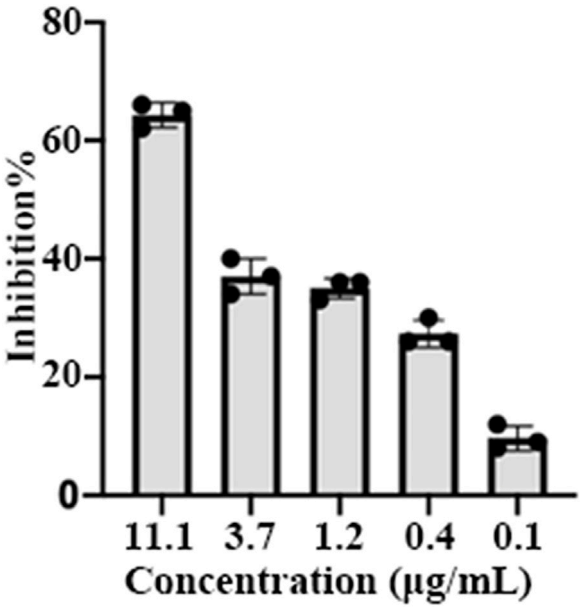


FIGURE 3 | Inhibition of pseudovirus by EPSs from *H. hispanica*. Pseudoviruses of SARS-CoV-2 were incubated with 3-fold serially diluted EPSs for 1 h and then transferred into pre-plated Vero E6 cells, followed by incubation for 24 h. Finally, the cells expressing GFP were measured by using a flow cytometer after lysed by trypsin.

in Vero cells. The results showed EPSs from *H. hispanica* can efficiently inhibit infection of pseudovirus to Vero E6 cells. The inhibition rate of EPSs toward pseudovirus reached 65% at 11.1 μg/mL (**Figure 3**).

Evaluation of the Anti-SARS-CoV-2 Activity of Exopolysaccharides From △*HAH_1206*

EPSs from *H. hispanica* are the first identified anti-SARS-CoV-2 polysaccharides from archaea, and their antiviral mechanisms are worth investigating. A serial of mutants related to glycosylation in *H. hispanica* were constructed and tested (Lu et al., 2020). Compared to the wild-type of *H. hispanica*, the △*HAH_1206* mutant was significantly affected (data not shown). Interestingly, it was found that EPSs from △*HAH_1206* could not bind with RBD (**Supplementary Figure S3**) and did not inhibit the binding of RBD to BEAS-2B and Vero E6 cells either (**Figure 4**). The results indicated that EPSs from △*HAH_1206* almost completely lost anti-SARS-CoV-2 activity.

Structural Comparison of Exopolysaccharides From Wild-Type *H. hispanica* and △*HAH_1206*

To investigate the reasons underlining the difference in antiviral activities of EPSs from wild-type *H. hispanica* and △*HAH_1206*, their molecular weights (MW), monosaccharide compositions, and sulfation levels were analyzed. The results from HPGPC showed that the MW of EPSs from WT and △*HAH_1206* were 2.126×10^7 Da and 2.007×10^7 Da, respectively, and their MW were nearly identical (**Supplementary Figure S4**). The sugar compositions were analyzed by HPAEC-PAD. The molar ratio of monosaccharides in EPSs from wild-type *H. hispanica* was GlcA:Man:Glc:Gal = 4.3:3.9:1.6:1, and that for △*HAH_1206* was GlcA:Man:Glc:Gal = 0.8:3:1.6:1. These two kinds of EPSs have the same molar ratio of glucose and galactose, and their molar ratios of mannose were slightly different. However, the contents of

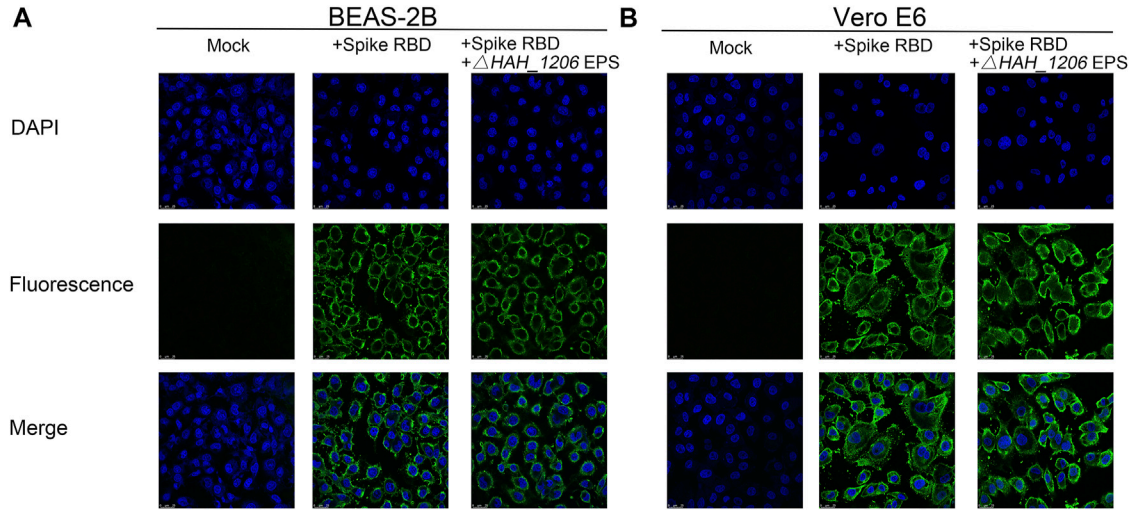


FIGURE 4 | Immunofluorescence assay of EPSs from △*HAH_1206* toward binding of RBD to BEAS-2B and Vero E6. **(A)** Influence of EPSs from △*HAH_1206* on binding of RBD to BEAS-2B cells. **(B)** Influence of EPSs from △*HAH_1206* on binding of RBD to Vero E6. **Mock**: cells detected in the absence of RBD and EPSs; + **Spike RBD**: cells detected in the presence of RBD (1 μg/well); and **+ Spike RBD** + **EPSs from △*HAH_1206***: cells detected in the presence of RBD (1 μg/well) and EPSs (2 μg/well). All wells were detected by immunofluorescence using SARS-CoV-2 (2019-nCoV) spike-neutralizing antibody and Alexa Fluor 488 goat anti-mouse IgG (H + L) by using a confocal microscope. The fluorescence signals were captured with an FITC channel, and the nuclear DNAs of cell were stained with DAPI.

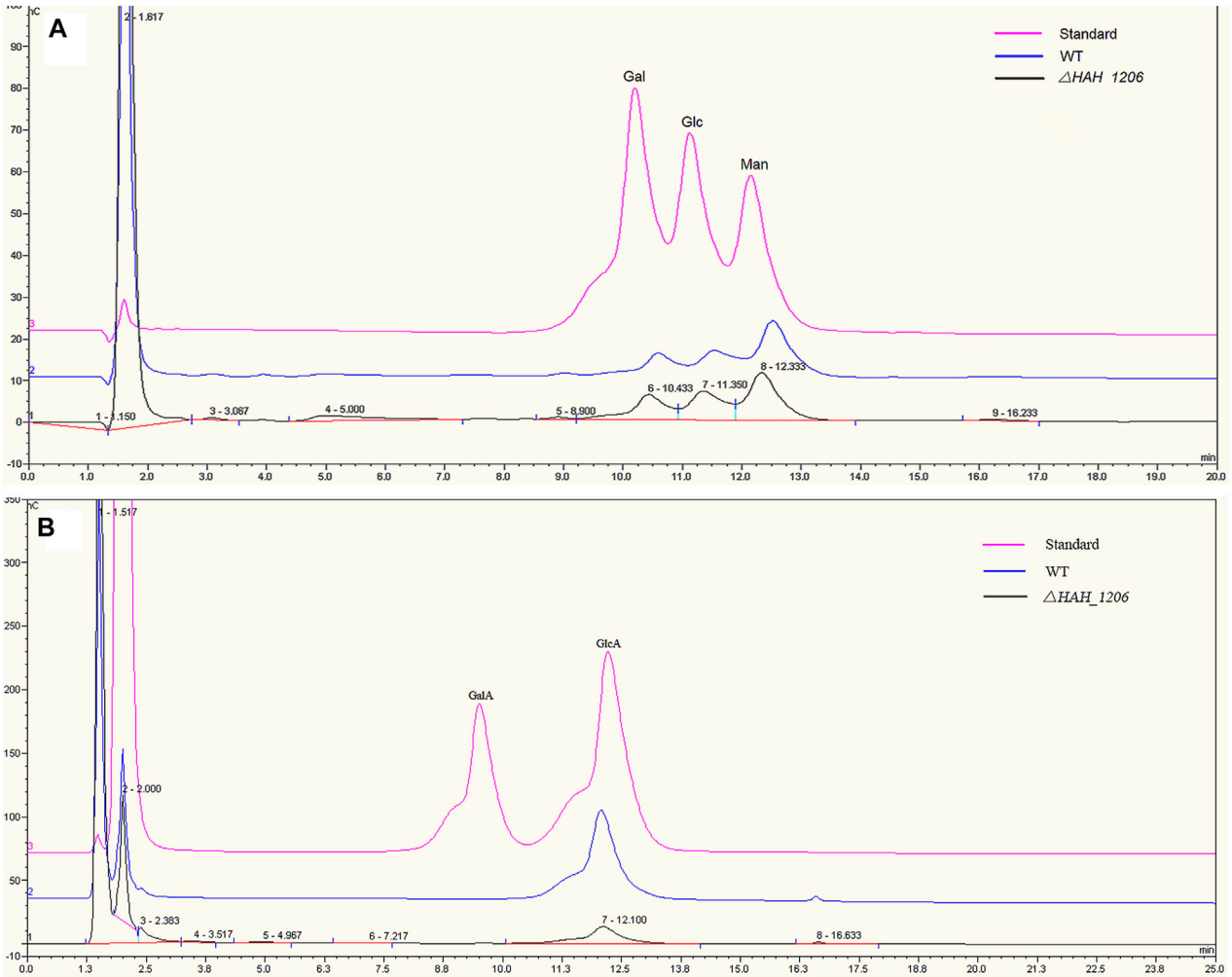


FIGURE 5 | Comparative analysis of monosaccharides in EPSs from *H. hispanica* and △*HAH_1206* with HPAEC-PAD. **(A)** Analysis of neutral monosaccharides of EPSs. The chromatomap in pink is the standard mixture of galactose (Gal), glucose (Glc), and mannose (Man) at 2 mM, whereas the one in blue refers to the hydrolyzed product of EPSs from *H. hispanica* and the black one is the hydrolyzed product of EPSs from △*HAH_1206*. **(B)** Analysis of acidic monosaccharides of EPSs. The chromatomap in pink is the standard mixture of D-galacturonic acid (GalA) and D-glucuronic acid (GlcA) at 2 mM, whereas the blue one represents the hydrolyzed product of EPSs from *H. hispanica*, and the black one is the hydrolyzed product of EPSs from △*HAH_1206*.

glucuronic acid in EPSs between two strains were quite different. GlcA in EPSs of △*HAH_1206* was only one-fifth of the wild-type (**Figure 5**). A significant reduction of GlcA in △*HAH_1206* was clearly observed, suggesting that GlcA might play critical roles in resistance against SARS-CoV-2. Because some sulfated polysaccharides displayed anti-SARS-CoV-2 activity (Kwon et al., 2020; Song et al., 2020), and EPSs from *H. hispanica* were also sulfated, the sulfation levels of EPSs from wild-type *H. hispanica* and △*HAH_1206* were also compared. An apparent reduction of the sulfation level was found in EPSs from △*HAH_1206* (**Figure 6**). Combined results of changes in monosaccharide composition and the sulfation level of EPSs from △*HAH_1206* meant GlcA of EPS from wild-type *H. hispanica* was heavily sulfated, which was essential for the anti-SARS-CoV-2 activity, and the decline of the sulfation level in △*HAH_1206* might result from reduction of GlcA.

DISCUSSION

Glycosylation is essential for assembly, recognition, and entry of SARS-CoV-2. The glycosylation sites of N165 and N234 in S protein are important for maintaining the "open" state, and these sites are also necessary for the binding of S protein with hACE2 (Casalino et al., 2020). In addition, the N90, N322, and N546 glycosylation sites in hACE2 can promote the binding of S protein with hACE2 (Zhao et al., 2020). These observations promoted scientists in the field of glycoscience to wonder whether there is a glycan recognition mechanism between SARS-CoV-2 and host cells. Accordingly, they are striving to find an answer and searching for glycan inhibitors against SARS-CoV-2 infection. Recent study showed the spike protein of SARS-CoV-2 cannot recognize sialic acid, but can specifically bind with HS in a sulfation-dependent manner (Hao et al., 2021). Furthermore, the spike protein can

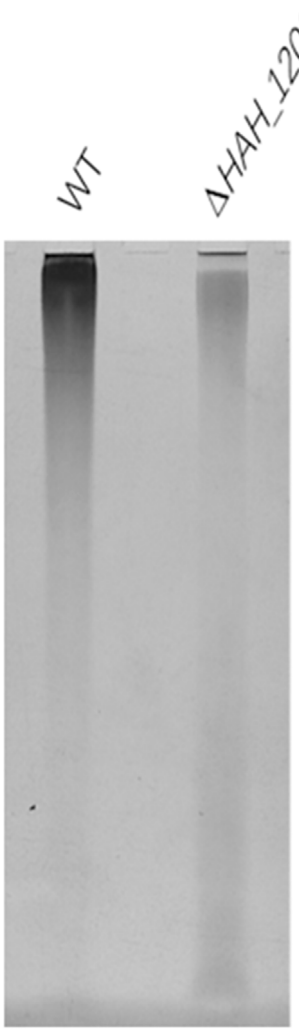


FIGURE 6 | Detection of sulfate in EPSs using the methyl blue method. EPSs (10 μg) from wild-type *H. hispanica* and *△HAH_1206* were analyzed in 7.5% (w/v) SDS-PAGE, and sulfates in EPSs were stained with methyl blue.

bind with hACE2 and HS through different domains, and HS on the surface of host cells can promote the entry of SARS-CoV-2 into host, which can be inhibited by exogenous heparin (Clausen et al., 2020). Moreover, researchers found other sulfated polysaccharides were also potential candidates against SARS-CoV-2, such as fucoidan, carrageenan, and sulfated polysaccharide from sea cucumber. In addition, β-chitosan, xylitol, capsular polysaccharides from *S. pneumoniae*, and LPS from *P. aeruginosa* can also bind with S protein (Alitongbieke et al., 2020; Chiodo et al., 2020; Song et al., 2020; Bansal et al., 2021). These examples clearly demonstrated that glycans with different structural characteristics could prevent SARS-CoV-2 infection, possibly using different mechanisms.

Archaea are one of the most mysterious parts of global ecosystem, which have developed various adaptations under extreme conditions, and the interaction between archaea and virus is less understood. Therefore, our understanding of antiviral activities of archaea far lags behind our knowledge of those in bacteria, mammals, and plants. Currently, more than 100 archaea have been discovered, and over 90 viruses were described as halophilic archaeal viruses (Atanasova et al., 2015; Snyder et al., 2015). Members of the family *Pleolipoviridae* (termed pleolipoviruses) belong to pseudo-spherical and pleomorphic archaeal viruses, which have a narrow host range as extremely halophilic archaea in the class *Halobacteria* (Bamford et al., 2017). EPSs are biomacromolecules with high molecular weights secreted by microbes. Most of EPSs are heteropolysaccharides containing three to four types of monosaccharides arranged in groups of 10 to form the repeating units (Poli et al., 2010). EPSs around cells can provide effective protection against severe environment and pathogen. Moreover, EPSs have displayed multi-function as anti-microbe, immunomodulator, anti-inflammation, antioxidant, anticancer, hypocholestrolemia, antidiabetes, and antivirus (Angelin and Kavitha, 2020; Abdalla et al., 2021). In this study, EPSs of *H. hispanica* can bind to SARS-CoV-2 RBD of spike protein with a high affinity of K_D as 2.23 nM and can inhibit the binding of RBD to BEAS-2B and Vero E6 cells. Importantly, the infection of SARS-CoV-2 pseudovirus to Vero E6 cells was effectively blocked by EPSs from *H. hispanica*. This is the first report that glycans from archaea can inhibit the infection of SARS-CoV-2.

Another important finding is the anti-SARS-CoV-2 activity of EPSs from *H. hispanica* is possibly related to sulfated GlcA. This means sulfation on GlcA may play critical roles in EPSs of *H. hispanica* against SARS-CoV-2. As we know, GlcA is the basic monosaccharide of glycosaminoglycans (GAGs), which are ubiquitously present on almost all mammalian cells and considered to be the first interface between a host cell and various bacterial, parasitic, and viral pathogens. GAGs and their derivatives, some of which lack significant anticoagulant activity, are under-exploited antiviral candidate drugs as they possess broad-spectrum activity against a multitude of distinct viruses (Mycroft-West et al., 2020). The repeating disaccharide units of GAGs, comprising a hexosamine and uronic acid or a galactose residue, are often sulfated. The anti-SARS-CoV-2 activities of GAGs have been confirmed in previous reports, along with other sulfated polysaccharides. Therefore, many people presumed that this kind of antiviral activity was possibly related to sulfation (Clausen et al., 2020; Kwon et al., 2020; Song et al., 2020; Hao et al., 2021). However, the exact structural unit of sulfated GAG, which contributes to their anti-SARS-CoV-2 activities, is still unknown. It is likely that the role of sulfated uronic acid in GAGs in anti-SARS-CoV-2 is similar to that of sulfated GlcA in EPSs of *H. hispanica*. The current study identified sulfated GlcA in EPSs is important for their anti-SARS-CoV-2 activity, which prompted us to propose that sulfated GlcA in HS is an important structural unit for their anti-SARS-CoV-2 activity. For EPSs biosynthesis in *H. hispanica*, a polysaccharide biosynthesis gene cluster has been annotated (Liu et al., 2011), which contained seven genes from *HAH_1661* to *HAH_1667*. The biosynthesis pathway of EPSs in *H. hispanica* has not been fully elucidated, and the details of the process are still less understood. Because mannose is the major composition of EPSs, two genes, *HAH_1662* and *HAH_1667*, which were considered coding for mannosyltransferase were deleted in *H. hipanica* respectively, the mutants of *△HAH_1662* and *△HAH_1667* almost lost acidic EPSs.

To confirm the structure and modification of EPSs, we need to characterize the function of each gene in the polysaccharide biosynthesis gene cluster of *H. hispanica* in future.

This study identified that sulfated GlcA in EPSs is important for the anti-SARS-CoV-2 activity, which has deepened our understanding of the key structural unit of glycans containing sulfated GlcA toward the anti-SARS-CoV-2 activity. Nowadays, the study about antiviral mechanisms of archaea focused on the nucleic acid level, including the CRISPR-Cas system and DNA phosphorothioation. The unusual metabolic pathways of archaeal cells can produce unique biomacromolecules and metabolites with novel characteristics. EPSs from extremophiles are quite different in composition and characteristics from those in other microbes. EPSs from most mesophilic microbes are toxic, whereas extremophilic microorganisms produce non-pathogenic EPSs, which can be applied in food, pharmaceutical, and cosmetic industries (Nicolaus et al., 2010). Although EPSs from *H. hispanica* displayed slight toxicity to cells, they are still good candidates to be developed into antiviral reagents, which would provide a new strategy against SARS-CoV-2.

CONCLUSION

In this study, we found that EPSs from halophilic archaeon *Haloarcula hispanica* displayed activities against SARS-CoV-2; it is the first discovery that EPSs from archaea can effectively inhibit SARS-CoV-2 *in vitro*. Compared to EPSs from deletion mutants of △*HAH_1206*, which lost anti-SARS-CoV-2 activity, it is likely that sulfated GlcA in EPSs from wild-type *H. hispanica* contribute to anti-SARS-CoV-2 activities. Our findings will provide a novel antiviral mechanism and a guide for designing new agents against SARS-CoV-2.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, and further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

YD, QW, and JL conceived and designed the study. YX, YLi, CP, SJ, and XZ carried out the laboratory tests. XL and YLü provided instructions to the binding test and EPSs preparation, respectively. YX, YLi, XY, ZW, and JL led the data collection and analysis. YX and YLi drafted the manuscript. ZW, JL, YD, QW, CJ, and GG provided critical review of the manuscript. The corresponding authors attested that all listed authors meet authorship criteria, and no others meeting the criteria have been omitted.

FUNDING

This work was supported by the National Key Research and Development Program of China (2019YFD0902000) and the special fund for "Respiratory Tract Protection Engineering Center in Dalian" from Dalian Development and Reform Commission.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2022.871509/full#supplementary-material

REFERENCES

Abdalla, A. K., Ayyash, M. M., Olaimat, A. N., Osaili, T. M., Al-Nabulsi, A. A., Shah, N. P., et al. (2021). Exopolysaccharides as Antimicrobial Agents: Mechanism and Spectrum of Activity. *Front. Microbiol.* 12, 664395. doi:10.3389/fmicb.2021.664395

Alitongbieke, G., Li, X.-m., Wu, Q.-C., Lin, Z.-C., Huang, J.-F., Xue, Y., et al. (2020). Effect of β-chitosan on the Binding Interaction between SARS-CoV-2 S-RBD and ACE2. *bioRxiv [Preprint]*. doi:10.1101/2020.07.31.229781

Angelin, J., and Kavitha, M. (2020). Exopolysaccharides from Probiotic Bacteria and Their Health Potential. *Int. J. Biol. Macromolecules* 162, 853–865. doi:10.1016/j.ijbiomac.2020.06.190

Atanasova, N. S., Bamford, D. H., and Oksanen, H. M. (2015). Haloarchaeal Virus Morphotypes. *Biochimie* 118, 333–343. doi:10.1016/j.biochi.2015.07.002

Bamford, D. H., Pietilä, M. K., Roine, E., Atanasova, N. S., Dienstbier, A., and Oksanen, H. M.Ictv Report Consortium (2017). ICTV Virus Taxonomy Profile: Pleolipoviridae. *J. Gen. Virol.* 98 (12), 2916–2917. doi:10.1099/jgv.0.000972

Bansal, S., Jonsson, C. B., Taylor, S. L., Figueroa, J. M., Dugour, A. V., Palacios, C., et al. (2021). Iota-carrageenan and Xylitol Inhibit SARS-CoV-2 in Vero Cell Culture. *PLoS One* 16 (11), e0259943. doi:10.1371/journal.pone.0259943

Cao, Y., Wang, J., Jian, F., Xiao, T., Song, W., Yisimayi, A., et al. (2021). Omicron Escapes the Majority of Existing SARS-CoV-2 Neutralizing Antibodies. *Nature* 602, 657–663. doi:10.1038/s41586-021-04385-3

Casalino, L., Gaieb, Z., Goldsmith, J. A., Hjorth, C. K., Dommer, A. C., Harbison, A. M., et al. (2020). Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* 6 (10), 1722–1734. doi:10.1021/acscentsci.0c01056

Chiodo, F., Bruijns, S. C. M., Rodriguez, E., Li, R. J. E., Molinaro, A., Silipo, A., et al. (2020). Novel ACE2-independent Carbohydrate-Binding of SARS-CoV-2 Spike Protein to Host Lectins and Lung Microbiota. *bioRxiv [Preprint]*. doi:10.1101/2020.05.13.092478

Clausen, T. M., Sandoval, D. R., Spliid, C. B., Pihl, J., Perrett, H. R., Painter, C. D., et al. (2020). SARS-CoV-2 Infection Depends on Cellular Heparan Sulfate and ACE2. *Cell* 183 (4), 1043–1057. doi:10.1016/j.cell.2020.09.033

Dwivedi, R., Samanta, P., Sharma, P., Zhang, F., Mishra, S. K., Kucheryavy, P., et al. (2021). Structural and Kinetic Analyses of Holothurian Sulfated Glycans Suggest Potential Treatment for SARS-CoV-2 Infection. *J. Biol. Chem.* 297 (4), 101207. doi:10.1016/j.jbc.2021.101207

Gao, C., Zeng, J., Jia, N., Stavenhagen, K., Matsumoto, Y., Zhang, H., et al. (2020). SARS-CoV-2 Spike Protein Interacts with Multiple Innate Immune Receptors. *bioRxiv [Preprint]*. doi:10.1101/2020.07.29.227462

Hao, W., Ma, B., Li, Z., Wang, X., Gao, X., Li, Y., et al. (2021). Binding of the SARS-CoV-2 Spike Protein to Glycans. *Sci. Bull.* 66 (12), 1205–1214. doi:10.1016/j.scib.2021.01.010

Jin, W., Zhang, W., Mitra, D., McCandless, M. G., Sharma, P., Tandon, R., et al. (2020). The Structure-Activity Relationship of the Interactions of SARS-CoV-2 Spike Glycoproteins with Glucuronomannan and Sulfated Galactofucan from *Saccharina Japonica*. *Int. J. Biol. Macromolecules* 163, 1649–1658. doi:10.1016/j.ijbiomac.2020.09.184

Kim, S. Y., Jin, W., Sood, A., Montgomery, D. W., Grant, O. C., Fuster, M. M., et al. (2020). Characterization of Heparin and Severe Acute Respiratory Syndrome-Related Coronavirus 2 (SARS-CoV-2) Spike Glycoprotein Binding Interactions. *Antiviral Res.* 181, 104873. doi:10.1016/j.antiviral.2020.104873

Kumlin, U., Olofsson, S., Dimock, K., and Arnberg, N. (2008). Sialic Acid Tissue Distribution and Influenza Virus Tropism. *Influenza Other Respir. Viruses* 2 (5), 147–154. doi:10.1111/j.1750-2659.2008.00051.x

Kwon, P. S., Oh, H., Kwon, S.-J., Jin, W., Zhang, F., Fraser, K., et al. (2020). Sulfated Polysaccharides Effectively Inhibit SARS-CoV-2 *In Vitro*. *Cell Discov* 6 (1), 50. doi:10.1038/s41421-020-00192-8

Li, W., Hulswit, R. J. G., Widjaja, I., Raj, V. S., McBride, R., Peng, W., et al. (2017). Identification of Sialic Acid-Binding Function for the Middle East Respiratory Syndrome Coronavirus Spike Glycoprotein. *Proc. Natl. Acad. Sci. U.S.A.* 114 (40), E8508–E8517. doi:10.1073/pnas.1712592114

Liu, H., Wu, Z., Li, M., Zhang, F., Zheng, H., Han, J., et al. (2011). Complete Genome Sequence of *Haloarcula Hispanica*, a Model Haloarchaeon for Studying Genetics, Metabolism, and Virus-Host Interaction. *J. Bacteriol.* 193 (21), 6086–6087. doi:10.1128/JB.05953-11

Liu, L., Chopra, P., Li, X., Bouwman, K. M., Tompkins, S. M., Wolfert, M. A., et al. (2021). Heparan Sulfate Proteoglycans as Attachment Factor for SARS-CoV-2. *ACS Cent. Sci.* 7 (6), 1009–1018. doi:10.1021/acscentsci.1c00010

Lu, H., Pei, C., Zhou, H., Lü, Y., He, Y., Li, Y., et al. (2020). Agl22 and Agl23 Are Involved in the Synthesis and Utilization of the Lipid-linked Intermediates in the Glycosylation Pathways of the Halophilic archaeaonHaloarcula Hispanica. *Mol. Microbiol.* 114 (5), 762–774. doi:10.1111/mmi.14577

Lu, L., Mok, B. W.-Y., Chen, L.-L., Chan, J. M.-C., Tsang, O. T.-Y., Lam, B. H.-S., et al. (2021). Neutralization of Severe Acute Respiratory Syndrome Coronavirus 2 Omicron Variant by Sera from BNT162b2 or CoronaVac Vaccine Recipients. *Clin. Infect. Dis.*, ciab1041. doi:10.1093/cid/ciab1041

Lü, Y., Lu, H., Wang, S., Han, J., Xiang, H., and Jin, C. (2017). An Acidic Exopolysaccharide fromHaloarcula hispanicaATCC33960 and Two Genes Responsible for its Synthesis. *Archaea* 2017, 1–12. doi:10.1155/2017/5842958

Mycroft-West, C. J., Su, D., Pagani, I., Rudd, T. R., Elli, S., Gandhi, N. S., et al. (2020). Heparin Inhibits Cellular Invasion by SARS-CoV-2: Structural Dependence of the Interaction of the Spike S1 Receptor-Binding Domain with Heparin. *Thromb. Haemost.* 120 (12), 1700–1715. doi:10.1055/s-0040-1721319

Nicolaus, B., Kambourova, M., and Oner, E. T. (2010). Exopolysaccharides from Extremophiles: from Fundamentals to Biotechnology. *Environ. Tech.* 31 (10), 1145–1158. doi:10.1080/09593330903552094

Poli, A., Anzelmo, G., and Nicolaus, B. (2010). Bacterial Exopolysaccharides from Extreme marine Habitats: Production, Characterization and Biological Activities. *Mar. Drugs* 8 (6), 1779–1802. doi:10.3390/md8061779

Saied, E. M., El-Maradny, Y. A., Osman, A. A., Darwish, A. M. G., Abo Nahas, H. H., Niedbała, G., et al. (2021). A Comprehensive Review about the Molecular Structure of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): Insights into Natural Products against COVID-19. *Pharmaceutics* 13 (11), 1759. doi:10.3390/pharmaceutics13111759

Shajahan, A., Archer-Hartmann, S., Supekar, N. T., Gleinich, A. S., Heiss, C., and Azadi, P. (2021). Comprehensive Characterization of N- and O- Glycosylation of SARS-CoV-2 Human Receptor Angiotensin Converting Enzyme 2. *Glycobiology* 31 (4), 410–424. doi:10.1093/glycob/cwaa101

Shajahan, A., Supekar, N. T., Gleinich, A. S., and Azadi, P. (2020). Deducing the N- and O-Glycosylation Profile of the Spike Protein of Novel Coronavirus SARS-CoV-2. *Glycobiology* 30 (12), 981–988. doi:10.1093/glycob/cwaa042

Shang, J., Wan, Y., Luo, C., Ye, G., Geng, Q., Auerbach, A., et al. (2020). Cell Entry Mechanisms of SARS-CoV-2. *Proc. Natl. Acad. Sci. U.S.A.* 117 (21), 11727–11734. doi:10.1073/pnas.2003138117

Shi, Y., Wu, Y., Zhang, W., Qi, J., and Gao, G. F. (2014). Enabling the 'host Jump': Structural Determinants of Receptor-Binding Specificity in Influenza A Viruses. *Nat. Rev. Microbiol.* 12 (12), 822–831. doi:10.1038/nrmicro3362

Snyder, J. C., Bolduc, B., and Young, M. J. (2015). 40 Years of Archaeal Virology: Expanding Viral Diversity. *Virology* 479-480, 369–378. doi:10.1016/j.virol.2015.03.031

Song, S., Peng, H., Wang, Q., Liu, Z., Dong, X., Wen, C., et al. (2020). Inhibitory Activities of marine Sulfated Polysaccharides against SARS-CoV-2. *Food Funct.* 11 (9), 7415–7420. doi:10.1039/d0fo02017f

Tan, W., Zhao, X., Zhao, X., Ma, X., Wang, W., Niu, P., et al. (2020). A Novel Coronavirus Genome Identified in a Cluster of Pneumonia Cases - Wuhan, China 2019–2020. *China CDC Wkly* 2 (4), 61–62. doi:10.46234/ccdcw2020.017

Tian, W., Li, D., Zhang, N., Bai, G., Yuan, K., Xiao, H., et al. (2021). O-glycosylation Pattern of the SARS-CoV-2 Spike Protein Reveals an "O-Follow-N" Rule. *Cell Res* 31 (10), 1123–1125. doi:10.1038/s41422-021-00545-2

Tortorici, M. A., Walls, A. C., Lang, Y., Wang, C., Li, Z., Koerhuis, D., et al. (2019). Structural Basis for Human Coronavirus Attachment to Sialic Acid Receptors. *Nat. Struct. Mol. Biol.* 26 (6), 481–489. doi:10.1038/s41594-019-0233-y

Walls, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., McGuire, A. T., and Veesler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* 181 (2), 281–292. doi:10.1016/j.cell.2020.02.058

Wang, C., Horby, P. W., Hayden, F. G., and Gao, G. F. (2020). A Novel Coronavirus Outbreak of Global Health Concern. *The Lancet* 395 (10223), 470–473. doi:10.1016/S0140-6736(20)30185-9

Wang, Q., Zhang, Y., Wu, L., Niu, S., Song, C., Zhang, Z., et al. (2020). Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell* 181 (4), 894–904. doi:10.1016/j.cell.2020.03.045

Watanabe, Y., Bowden, T. A., Wilson, I. A., and Crispin, M. (2019). Exploitation of Glycosylation in Enveloped Virus Pathobiology. *Biochim. Biophys. Acta (Bba) - Gen. Subjects* 1863 (10), 1480–1497. doi:10.1016/j.bbagen.2019.05.012

Zhang, S., Pei, R., Li, M., Su, H., Sun, H., Ding, Y., et al. (2022). Cocktail Polysaccharides Isolated from *Ecklonia Kurome* against the SARS-CoV-2 Infection. *Carbohydr. Polym.* 275, 118779. doi:10.1016/j.carbpol.2021.118779

Zhang, S., Qiao, Z., Zhao, Z., Guo, J., Lu, K., Mayo, K. H., et al. (2021). Comparative Study on the Structures of Intra- and Extra-cellular Polysaccharides from *Penicillium oxalicum* and Their Inhibitory Effects on Galectins. *Int. J. Biol. Macromolecules* 181, 793–800. doi:10.1016/j.ijbiomac.2021.04.042

Zhang, Z., Zhang, Y., Liu, K., Li, Y., Lu, Q., Wang, Q., et al. (2021). The Molecular Basis for SARS-CoV-2 Binding to Dog ACE2. *Nat. Commun.* 12 (1), 4195. doi:10.1038/s41467-021-24326-y

Zhao, P., Praissman, J. L., Grant, O. C., Cai, Y., Xiao, T., Rosenbalm, K. E., et al. (2020). Virus-Receptor Interactions of Glycosylated SARS-CoV-2 Spike and Human ACE2 Receptor. *Cell Host & Microbe* 28 (4), 586–601. doi:10.1016/j.chom.2020.08.004

Zhao, X., Zheng, A., Li, D., Zhang, R., Sun, H., Wang, Q., et al. (2021). Neutralisation of ZF2001-Elicited Antisera to SARS-CoV-2 Variants. *The Lancet Microbe* 2 (10), e494. doi:10.1016/S2666-5247(21)00217-2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Chemistry

REVIEW
published: 05 May 2022
doi: 10.3389/fchem.2022.896098

Automated Peptide Synthesizers and Glycoprotein Synthesis

Jiekang Tian[1†], *Yaohao Li*[2†], *Bo Ma*[2], *Zhongping Tan*[2*] *and Shiying Shang*[1*]

[1]*Center of Pharmaceutical Technology, School of Pharmaceutical Sciences, Tsinghua University, Beijing, China,* [2]*State Key Laboratory of Bioactive Substance and Function of Natural Medicines, Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China*

The development and application of commercially available automated peptide synthesizers has played an essential role in almost all areas of peptide and protein research. Recent advances in peptide synthesis method and solid-phase chemistry provide new opportunities for optimizing synthetic efficiency of peptide synthesizers. The efforts in this direction have led to the successful preparation of peptides up to more than 150 amino acid residues in length. Such success is particularly useful for addressing the challenges associated with the chemical synthesis of glycoproteins. The purpose of this review is to provide a brief overview of the evolution of peptide synthesizer and glycoprotein synthesis. The discussions in this article include the principles underlying the representative synthesizers, the strengths and weaknesses of different synthesizers in light of their principles, and how to further improve the applicability of peptide synthesizers in glycoprotein synthesis.

Keywords: peptide synthesis, glycoprotein synthesis, automated synthesizer, solid phase peptide synthesis, batch synthesis, continuous flow, high throughput synthesis

OPEN ACCESS

Edited by:
Xuechen Li,
The University of Hong Kong, Hong Kong SAR, China

Reviewed by:
Ryo Okamoto,
Osaka University, Japan
Suwei Dong,
Peking University, China
Mattan Hurevich,
Hebrew University of Jerusalem, Israel

***Correspondence:**
Zhongping Tan
zhongping.tan@imm.pumc.edu.cn
Shiying Shang
shangshiying@tsinghua.edu.cn

[†]*These authors have contributed equally to this work*

Received: *14 March 2022*
Accepted: *20 April 2022*
Published: *05 May 2022*

Citation:
Tian J, Li Y, Ma B, Tan Z and Shang S (2022) Automated Peptide Synthesizers and Glycoprotein Synthesis. Front. Chem. 10:896098. doi: 10.3389/fchem.2022.896098

INTRODUCTION

The synthesis of glycoproteins with well-defined protein and carbohydrate structures is essential for the study of their structures, properties and functions (Davis, 2002). To meet with this requirement, many different methods have been investigated for their use in the synthesis of homogeneous glycoforms, *i.e.*, glycoprotein isoforms with the same protein amino acid sequence but different glycosylation patterns. Based on the way of how to incorporate glycans into glycoproteins, these methods can be roughly classified into two categories, biochemical methods and chemical methods (Rich and Withers, 2009). No matter which type of methods is selected, the synthesis is more or less associated with glycopeptide and peptide synthesis (**Figure 1**).

The biochemical methods include two main subtypes, glycoprotein remodeling and enzymatic glycosylation of proteins. Glycoprotein remodeling involves first enzymatic trimming of the heterogenous glycans on a recombinant glycoprotein to produce a homogeneous glycoform or chemical synthesis of a homogeneous glycoform, and subsequent modification of the glycoform containing a core glycan or single N-acetylglucosamine (GlcNAc) residue by the glycosyltransferase- or endoglycosidase-catalyzed stepwise glycan synthesis or one-step transglycosylation (Li and Wang, 2018), while enzymatic glycosylation of proteins uses glycosyltransferases to directly build glycans on recombinant or chemically synthesized proteins (**Figure 1**). In the chemical methods of glycoprotein synthesis, the desired products are generated by covalently joining the peptide and glycopeptide fragments together using native chemical ligation (NCL) (Dawson et al., 1994) (**Figure 2**).

Compared with chemical methods, biochemical methods are more convenient to use and more practical in the preparation of large glycoproteins and large-scale preparation of glycoproteins.

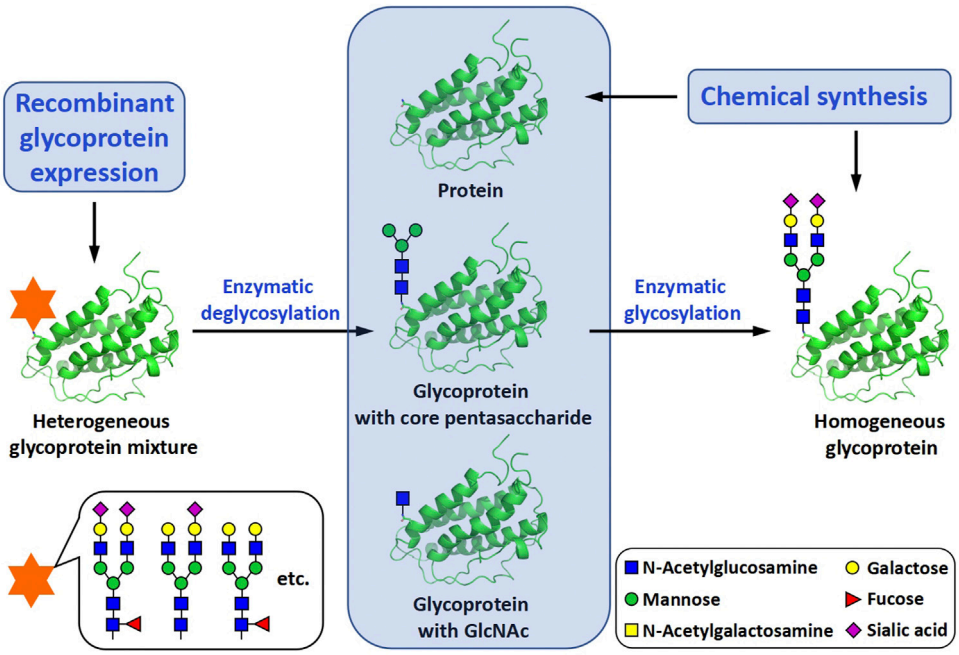


FIGURE 1 | Biochemical and chemical methods for the preparation of homogeneous glycoproteins.

However, due to the substrate specificity of the enzymatic reactions, the number of glycosylation patterns that can be generated by biochemical methods are limited (Ma et al., 2020). Although tedious and labor-intensive to perform, chemical methods are more flexible and precise, and in theory, can be used to prepare glycoforms with any structures. This advantage is very helpful to gain a more comprehensive and in-depth understanding of the role of protein glycosylation (Chaffey et al., 2018).

Chemical methods typically use NCL reactions to join glycopeptides to peptides to produce homogeneous glycoproteins (**Figure 2**). NCL reactions are relatively simple to set up and perform, and can be completed quickly, usually in a day. The difficulties associated with the chemical synthesis of glycoproteins are mainly due to two reasons: the high synthetic complexity and the low synthetic efficiency. Most of the complexity comes from the preparation and purification of glycopeptides and the cause of the low efficiency can be attributed largely to the use of relatively short peptides as the synthetic fragments. The use of short peptides increases the number of required synthetic steps and thus leads to decrease in synthetic efficiency. Based on these facts, it is reasonable to expect that small glycopeptide and long peptide fragments would reduce the difficulties in glycoprotein synthesis (**Figure 2**). Small glycopeptides only contain a small number of amino acids and can be prepared by manual synthesis. Manual synthesis is more likely to make better use of precious synthetic glycans, thereby reducing the time required for repetitive glycan and glycopeptide synthesis. The use of larger peptide fragments, on the other hand, would reduce the number of ligation and purification steps, thereby improving the efficiency of the assembly of glycoproteins. With the continuous improvement of automated peptide synthesizers, the problems associated with the synthesis of large peptides are expected to be slowly addressed. In this review, we provide a brief overview of the development history of peptide synthesizers, with a focus on the progress achieved at each stage of development.

EARLY PEPTIDE SYNTHESIZERS

Before the 1960s, peptide synthesis was an almost impossible task. In 1954, Vincent Du Vigneaud used the strategy of tetraethyl pyrophosphate-mediated solution-phase coupling reaction and sodium/liquid ammonia deprotection to synthesize oxytocin (Du Vigneaud et al., 1954). It took him many years and much effort to complete the synthesis of this octapeptide hormone. In 1963, Bruce Merrifield turned peptide synthesis from nearly impossible to possible. After more than 4 years of exploration, he developed a new technology called Solid-Phase Peptide Synthesis (SPPS) (Merrifield, 1963). This revolutionary technology allows all steps, including coupling, washing and deprotection, to be carried out in the same reaction vessel without isolation and purification of reaction intermediates, thus greatly simplifying the peptide synthesis process.

In the process of developing SPPS, Merrifield and his coworkers conducted many optimization studies and established most of the basic principles that still apply today: 1) optimal resins are critical for efficient synthesis of peptides; 2) attaching the C-terminal amino acid to the solid support and extending the peptide chain in the C to N direction to reduce racemization; 3) using orthogonal protecting groups for the α-

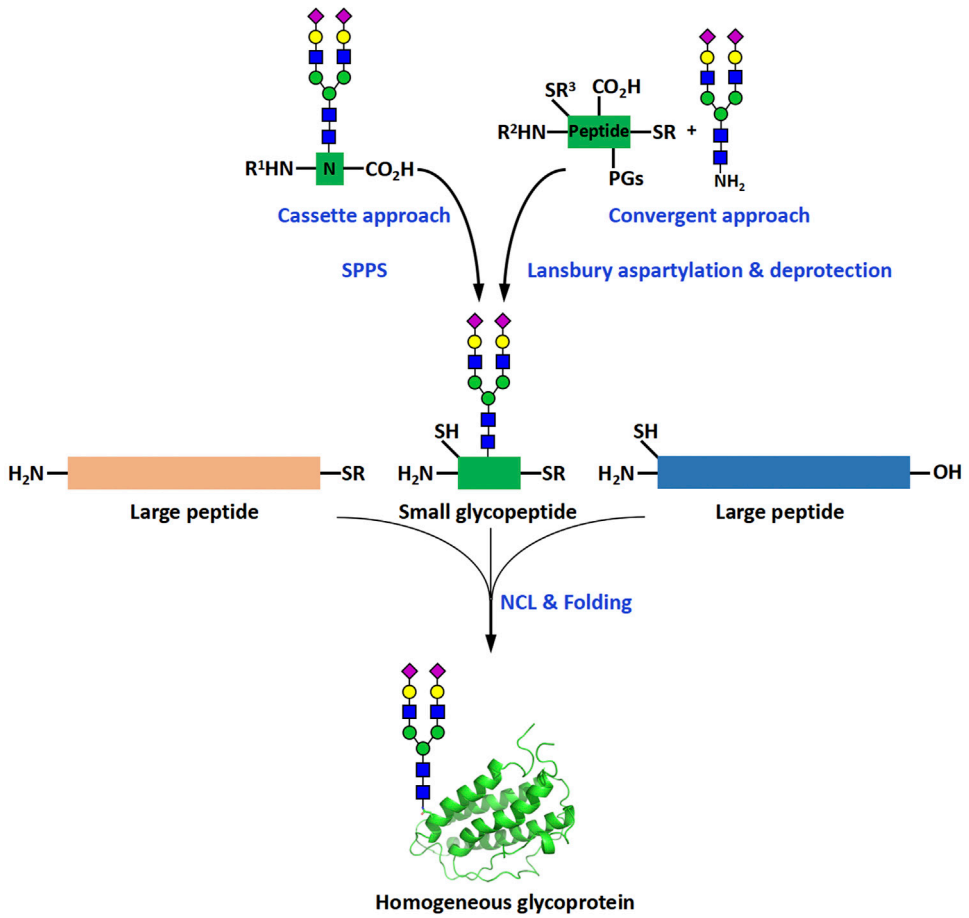


FIGURE 2 | Chemical methods for the synthesis of homogeneous glycoproteins. R, alkyl or aryl groups; R^1, R^2, R^3, protecting groups of amino or thiol groups; PGs, protecting groups of peptide side chains. NCL is the abbreviation of "native chemical ligation", which is a reaction that is commonly used in assembly of proteins and glycoproteins. It normally involves the chemoselective ligation between a C-terminal thioester and an N-terminal thiol amino acid-containing fragment.

amines of amino acids to enable the selective formation of amide bonds; 4) increasing the coupling efficiency to be higher than 99%; 5) using strong acids to simultaneously cleave peptides from resins and remove side-chain protecting groups (**Figure 3A**) (Merrifield, 1965).

In accordance with these principles, Merrifield and coworkers optimized the Boc chemistry for SPPS and established the following conditions as their optimal conditions: SPPS was carried out on a resin obtained by the copolymerization of styrene and divinylbenzene (98% and 2% respectively). Boc was used as the α-amine protecting group, DCC as the coupling reagent, DMF as the solvent, and the solution of HBr or HF in TFA as the cleavage cocktail. Under the aforementioned optimized conditions, SPPS became more routine and easier to perform. In 1964, Merrifield reported the total synthesis of Bradykinin (sequence: RPPGFSPFR), a nonapeptide plasma kinin, using optimized Boc chemistry (Merrifield, 1964). The manual synthesis was carried out on a Boc-L-Arg (NO_2)-polystyrene-divinylbenzene resin and was completed in 8 days. The yield of the crude peptide was 93% and after purification, the overall yield of the chromatographically pure peptide was 68%. This result clearly demonstrated that the efficiency of solid-phase synthesis of peptides is much higher than that of solution-phase synthesis.

Automated Peptide Synthesizers Using Boc Chemistry

Although the manual SPPS gave satisfactory results in the synthesis of short peptides, this technology became less practical when applied to the synthesis of large peptides due to the considerable amount of time demanded by the synthesis task. To overcome this issue, in 1965, Merrifield and his collaborators designed and constructed the first automated solid-phase peptide synthesizer based on the use of Boc chemistry (Merrifield and Stewart, 1965). The application of automation to the SPPS process greatly reduced the need for manpower for performing peptide synthesis and human error, and thus significantly improved and simplified the synthesis of long peptides. As shown in **Figure 3B**, in the automated synthesizer, the resin beads were kept in the same reaction vessel during the entire synthesis process. Metering pumps

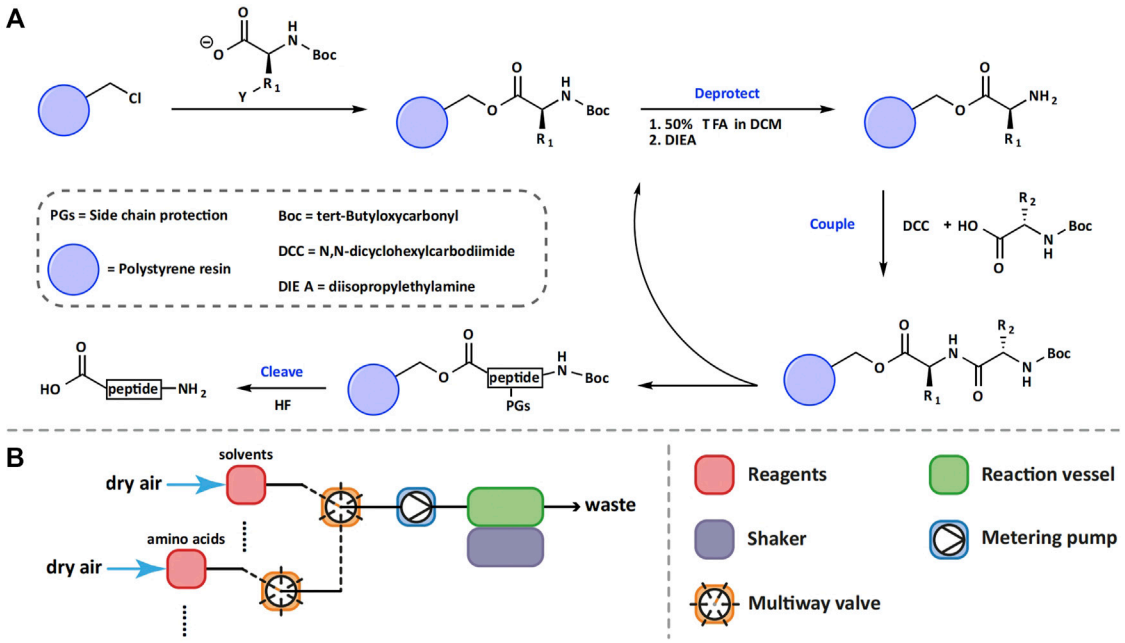


FIGURE 3 | Schematic representation of Boc-SPPS **(A)** and the first automated solid-phase peptide synthesizer **(B)**.

were used to transfer an appropriate amount of solvent or reagents into the reaction vessel according to the pre-calculated volume and concentration.

Using their newly invented instrument, Merrifield and coworkers did a comparative synthesis of bradykinin. It was found that the synthesis can be completed in 32 h, which is only 20% of the time required in the manual synthesis (Merrifield, 1964). In the following years, they had continuously improved the performance of their synthesizer and was eventually able to produce ribonuclease A, a 124 amino acid-long peptide, in one-shot synthesis (Gutte and Merrifield, 1971).

In the 1980s, the commercialization of the automated peptide synthesizer ABI 430 developed by Applied Biosystems further promoted the synthesis of large peptides. However, as more and more studies were conducted, researchers gradually realized that it is not always possible to synthesize large peptides like ribonuclease A and the length of peptides that could be reliably prepared using automated peptide synthesizers is generally limited to be less than 50 amino acids. This is likely to be related to the batch-wise design of the system, which has the disadvantages of relatively low efficiency of reagent mixing, slow mass transfer and long coupling time (Breen et al., 2021).

Automated Peptide Synthesizers Using Fmoc Chemistry

During the rapid development of peptide synthesizers based on the Boc chemistry, in 1970, Carpino and Han introduced the Fmoc group for the protection of the α-amines of amino acids (Carpino and Han, 1970). Fmoc can be efficiently removed under mild basic conditions like 20% piperidine in DMF. When Fmoc is chosen for the protection of the amine function of the amino acid, the TFA-labile groups like Boc, Trt, and Pbf can be used to protect the side chains, thus avoiding the need for the highly corrosive and toxic acid HF in the final cleavage step of SPPS. In addition, the fluorescence property of the dibenzofulvene adducts formed after treatment of the Fmoc resin with piperidine enables the estimation of the efficiency of each peptide coupling step.

In the early 1970s, a method for thoroughly mixing resin beads with solvent and reagents, the continuous flow SPPS method, was developed by Bayer et al. (1970). Unlike the batch-wise SPPS method adopted by the peptide synthesizers like ABI 430, which maintains the reaction suspension by gas bubbling or mechanical stirring, the continuous flow method uses a pump to provides a rapid and continuous flow through the reaction vessel, thus increasing the efficiency of mixing and mass transfer, and decreasing the coupling time (Lukas et al., 1981).

Using Fmoc chemistry and continuous flow technology, PerSeptive Biosystems developed the Pioneer Peptide Synthesis System, an automated synthesizer capable of performing the peptide synthesis in a simple and straightforward manner. This instrument uses glass column with filters at the top and bottom as the reaction vessel. Pumps and valves are used to control and regulate the flow of solvent and reagent solutions. A distinguishing feature of this peptide synthesizer is that it has a UV detector. The coupling efficiency of each step of SPPS can be assessed based on the intensity of the UV signal, which in turn can direct the optimization process to look for solutions to improve the quality of peptide products.

Slightly different from the methods used for Boc SPPS, pseudoproline dipeptides, Dmb/Hmb-protected dipeptides and isoacyl dipeptides are often required for peptide synthesis using

Fmoc chemistry (Behrendt et al., 2016). This is mainly due to the different aggregation behaviors of the growing peptide chains on resin in the presence or absence of TFA. During Boc SPPS, the TFA-protonated N-terminal amino group can generate electrostatic repulsion between peptides and consequently expose the reactive groups for the following coupling reactions. During Fmoc SPPS, the lack of electrostatic repulsion can increase the risk of aggregation and reduce the synthesis efficiency. The use of dipeptides can alleviate this issue by inhibiting aggregation of peptides. However, because most peptides do not contain enough sites that are suitable for the incorporation of dipeptides, the length of peptides that can be successfully synthesized by the Pioneer Peptide Synthesizer is also limited to less than 50 amino acids (most commonly 30–40 residues in length).

High-Throughput Automated Peptide Synthesizers

Many studies such as mimotope screening require the simultaneous preparation of a large number, but not large quantities of peptides. When tens to hundreds of peptides are needed, it becomes impractical to use 1-channel (ABI 430) or 2-channel (Pioneer) peptide synthesizers for their synthesis. To meet with the high demand in the peptide research community and pharmaceutical companies, new technologies like parallel and combinatorial synthesis of peptide libraries were developed. Among these technologies, the "tea-bag" and the split-and-pool methods have received much attention.

The "tea-bag" method for parallel synthesis of a large number of peptides was developed by Houghten and co-workers (Pinilla et al., 1996). In this method, the sealed polypropylene mesh packets are filled with resin beads and are labeled. Depending on the sequences of the peptides, the packets are placed in different reaction vessels for coupling with the desired amino acids. In this way, each packet can be made to contain only one peptide, whose identity can be determined by the label.

The solid-phase split-and-pool combinatorial peptide synthesis method was reported by Furka et al. (1991). In this method, the resin beads are split into several portions, each reacting with a different amino acid. After the coupling reaction, all the beads are pooled together and resplit into a new set of subgroups to react with different amino acids. As this cycle repeats, the number of the generated peptides increases exponentially. Using the split-and-pool method, it is possible to synthesize millions of peptides in a relatively short period of time, however, the identity of each peptide is unknown.

The complexity of the "tea-bag" and the split-and-pool methods makes it almost impossible to develop automated peptide synthesizers to realize these concepts. Most of the high-throughput peptide synthesizers were designed based on simple parallel synthesis methods. For example, the Advanced ChemTech APEX 396 Automated Multipeptide Synthesizer is capable of the parallel synthesis of 96 peptides. The delivery of reagents and solvents to 96 reaction vessels is accomplished using pipetting robotic arms. The waste generated in the reaction and washing steps is removed from the bottom of the reaction vessels by pressure from the top. Upon completion of the synthesis, the resin needs to be transferred from the reaction vessel of the automated synthesizer to another container for cleavage. Although this type of synthesizers can be well applied in the synthesis of large number of short peptides, they are generally not suitable for the preparation of large peptides.

APPLICATION OF AUTOMATED PEPTIDE SYNTHESIZERS IN GLYCOPROTEIN SYNTHESIS

Fmoc chemistry was more widely used in the synthesis of glycoproteins. The most important reason behind this preference is the fact that glycans are not stable under strongly acidic conditions. If Boc chemistry is used, the repetitive TFA acidolysis employed for the Boc-group deprotection and HF employed for final cleavage can cause deleterious side reactions to complex glycans on glycopeptides. In addition, the reaction conditions in Fmoc SPPS are much milder than those in Boc SPPS and the liquid waste generated in Fmoc SPPS is environmentally more friendly. Therefore, Fmoc chemistry is generally used for the synthesis of glycoproteins.

In the process of previous glycoprotein chemical synthesis, large peptide fragments were generally directly prepared using automated peptide synthesizers. The synthesis of glycopeptides was more complicated. Depending on the types of glycans and how glycans were made, the methods for glycopeptide synthesis can be divided into two categories. The most common method for preparing O-linked glycopeptides was to directly use Fmoc-protected glycosylated amino acids as building blocks for automated SPPS (**Figure 2**). This "cassette approach" was also applied to N-linked glycoamino acids that are isolated from natural sources like chicken eggs (Seko et al., 1997; Sun et al., 2014; Liu et al., 2017). The preparation of most N-linked glycans requires time- and labor-intensive chemical synthesis. The glycopeptides bearing the chemically synthesized glycans were mostly synthesized by convergent approaches like the Lansbury aspartylation reaction, which involves the direct coupling of a glycosylamine with the side-chain carboxyl group of the Asp residue in the peptide (**Figure 2**). After the synthesis of the large peptide fragments and glycopeptide fragments, they were joined together by NCL to afford the desired glycoprotein products (Li et al., 2017).

O-Linked Glycoprotein Synthesis

In the past 20 years, a few O-linked glycoproteins have been prepared by chemical synthesis (Marcaurelle et al., 2001; Asahina et al., 2019; Wang et al., 2021; Zhao et al., 2022) and a representative one is lymphotactin (Lptn), a 93-amino acid chemokine that contains eight O-glycosylation sites at its C-terminus. In 2001, Bertozzi and coworkers reported its synthesis for the first time (Marcaurelle et al., 2001).

According to the retrosynthesis strategy based on NCL, Lptn can be generated by combining two fragments: one fragment is a 47-amino acid long peptide Lptn (1-47), and the other fragment is a 46-amino acid long glycopeptide Lptn

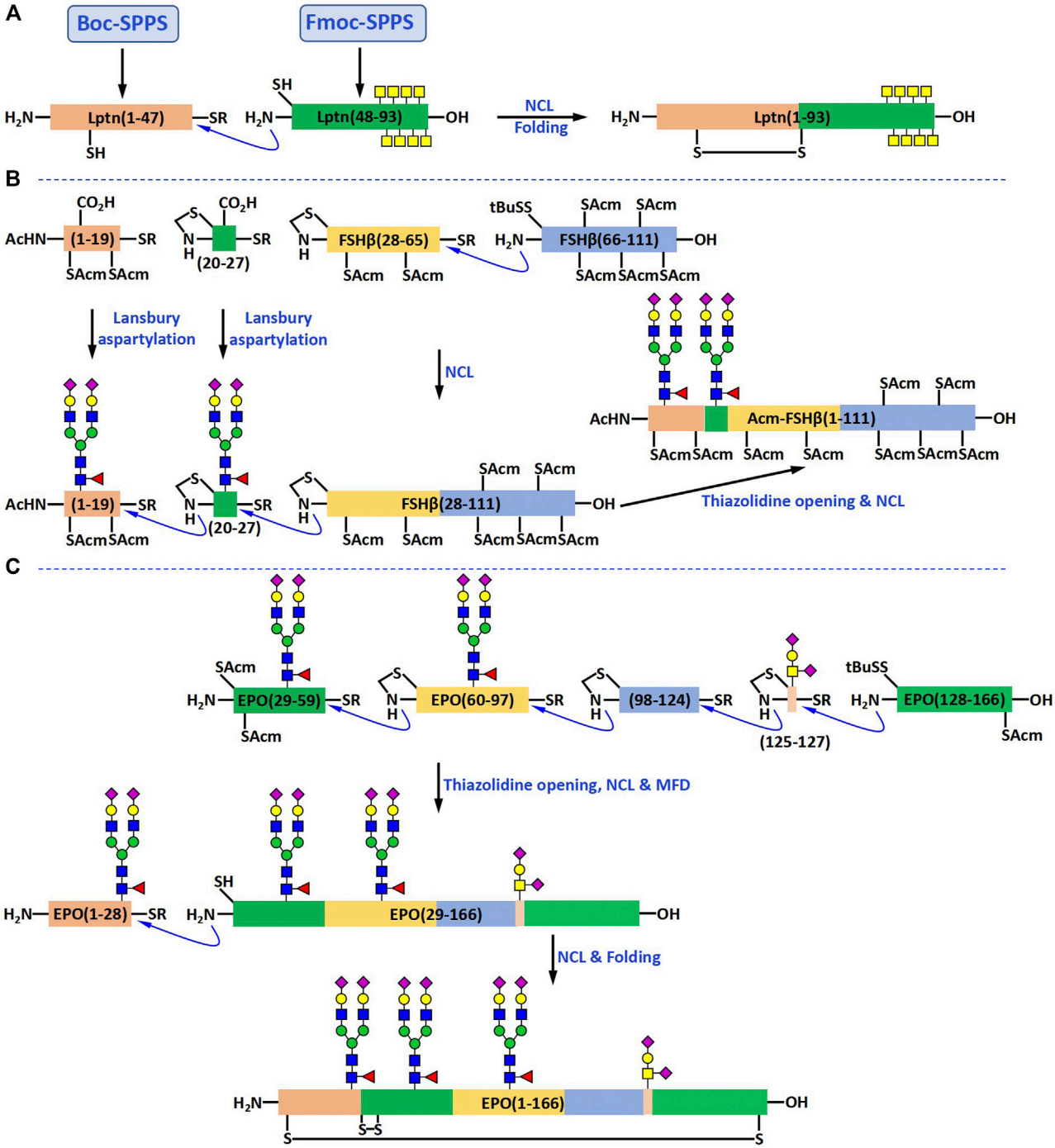


FIGURE 4 | Chemical syntheses of Lptn bearing O-glycans **(A)**, FSHβ bearing N-glycans **(B)** and EPO bearing both O- and N-glycans **(C)**.

(48-93), which contains eight O-linked GalNAc residues. They first attempted to prepare Lptn (1-47) with a C-terminal thioester using Fmoc SPPS (**Figure 4A**). However, their experimental results showed that the desired product could not be detected. To solve this problem, they turned to the Boc SPPS, which enabled them to obtain the peptide thioester directly.

The glycopeptide fragment Lptn (48-93) was prepared using the cassette approach mentioned above. Fmoc-Gly Wang resin and Fmoc chemistry were used in the synthesis. In order to achieve a reasonable coupling efficiency of the sterically hindered glycosylated amino acid building blocks at the glycosylation sites Thr 76, Thr 79, Thr 81, Ser 84, Thr 85, Thr 87, Thr 90, and Thr 92, manual synthesis was carried out for the first 20 amino acids.

HBTU and HOBt were used as coupling agents, and the Kaiser test was used to verify the completion of each coupling step. After the synthesis of the first half of the glycopeptide, the resin beads were transferred to the reaction vessel of an ABI 431A automated peptide synthesizer to complete the second half. The coupling agents used on the peptide synthesizer was DCC and HOBt. The glycopeptide fragment was cleaved from the resin using Reagent K and the acetyl groups on GalNAc residues were removed with the aqueous solution of hydrazine hydrate (10%). RP-HPLC purification afforded the glycopeptide product Lptn (48-93) in 24% yield.

The NCL of the large peptide Lptn (1-47) with the glycopeptide Lptn (48-93) was carried out by mixing these two fragments at a ratio of approximately 1:1 in the ligation buffer. They were slowly ligated together to form Lptn (1-93) as the major product, which was isolated and purified by RP-HPLC in 38% yield. The successful folding of Lptn (1-93) was realized by dissolving this large glycopeptide in a cysteine/cystine redox buffer. After another RP-HPLC purification, the desired glycoprotein product was obtained in 49% yield (**Figure 4A**).

N-Linked Glycoprotein Synthesis

Many N-linked glycoproteins have also been synthesized in the past decade (Piontek et al., 2009a; Piontek et al., 2009b; Izumi et al., 2012; Sakamoto et al., 2012; Okamoto et al., 2014; Reif et al., 2014; Wang et al., 2020; Li et al., 2021). The synthetic strategies for the preparation of most of them are very similar to that used for generating O-glycosylated Lptn, with the only difference being that the building blocks become the N-glycoamino acids isolated from chicken egg yolk powder or soybean powder (Izumi et al., 2012; Sakamoto et al., 2012). The number of N-glycopeptides that were prepared by the direct use of the Lansbury aspartylation reaction is relatively smaller than that by the "cassette approach". In 2012, Danishefsky and coworkers reported the synthesis of the β-subunit of follicle-stimulating hormone (FSHβ) in which the Lansbury aspartylation reaction was employed for the preparation of the N-glycopeptide fragments (Nagorny et al., 2012). FSHβ consists of 111 amino acids and two N-linked glycans at residues Asn7 and Asn24. According to their retrosynthetic analysis, FSHβ could be obtained from four fragments, two glycopeptide fragments FSHβ (1-19) and FSHβ (20-27), and two peptide fragments FSHβ (28-65) and FSH β (66-111) (**Figure 4B**).

The first step in the synthesis of glycopeptide FSHβ (1-19) was the SPPS of Fmoc-FSHβ (1-18). The Fmoc-Arg (Pbf)-TGT resin and the Pioneer Peptide Synthesizer were employed for the synthesis of this peptide. After Fmoc SPPS and resin cleavage by acetic acid, the Phe phenylthioester was attached to the C-terminus of the protected-peptide under the condition optimized by Sakakibara et al. Convergent coupling of the synthetic N-glycan with the peptide thioester *via* the Lansbury aspartylation reaction gave the desired glycopeptide product in 17% yield. The glycopeptide FSHβ (20-27) bearing the same N-linked dodecasaccharide was prepared in a similar way and obtained in 27% yield. The two peptide fragments FSHβ (28-65) and FSHβ (66-111) were synthesized using the Fmoc-Gly-TGT resin and Fmoc-Glu (OtBu)-TGT, respectively. The overall yield for FSHβ (28-65) was 48% and for FSHβ (66-111) 31%.

Under NCL conditions, the two peptide fragments FSHβ (28-65) and FSHβ (66-111) were first ligated together to form the large peptide FSHβ (28-111) in 38% yield. Using a similar ligation approach, the glycopeptide fragment FSHβ (20-27) was added to the N-terminus of FSHβ (28-111) in 26% yield, and the glycopeptide FSHβ (1-19) to the N-terminus of FSHβ (20-111) in 27% yield. The resulting large glycopeptide FSHβ (1-111) was not folded to generate the final glycoprotein product.

Synthesis of Glycoproteins Containing Both O- and N-Linked Glycans

Many natural glycoproteins contain both O- and N-glycans. However, due to its difficulty, there are only a few reports on the synthesis of these glycoproteins in previous studies (Wang et al., 2013; Fernandez-Tejada et al., 2014; Ye et al., 2021). In 2013, Danishefsky and coworkers completed the synthesis of a glycoprotein with such complexity, human erythropoietin (EPO) (Wang et al., 2013). EPO has 166 amino acids and four glycosylation sites, three N-glycosylation sites at Asn24, Asn38, and Asn83, and one O-glycosylation site at Ser126. Four glycopeptide fragments EPO (1-28), EPO (29-59), EPO (60-97), and EPO (125-166), and one peptide fragment EPO (98-124) were used for the synthesis of this glycoprotein (**Figure 4C**). The peptide fragments were prepared *via* Fmoc SPPS using a Pioneer Peptide Synthesizer. The N-glycopeptides were generated by coupling the synthetic dodecasaccharide anomeric amine with different peptide thioesters by the Lansbury aspartylation reaction. The O-glycopeptide was obtained by ligating the glycosylated EPO (125-127) with EPO (128-166).

After the synthesis of all five fragments, four of them were joined together in the C to N direction using NCL to generate the glycosylated EPO (29-166). The four non-native Cys residues in this glycopeptide, which were introduced to facilitate NCL reactions, were quantitatively desulfurized to Ala by the metal-free desulfurization (MFD) reaction. The Acm groups on the native Cys residues were then removed exposing the N-terminal reactive site, which underwent another NCL reaction to generate the large glycopeptide EPO (1-166). The folding was carried out in a cysteine/cystine redox buffer and the correctly folded glycoprotein product was purified by RP-HPLC.

PEPTIDE SYNTHESIZERS WITH HEATING

Most of the peptide synthesizers used in the reported synthesis of glycoproteins were developed more than 20 years ago. Although the synthesis could be achieved using these instruments, the synthetic efficiency is not high enough and the synthetic yields for peptides longer than 30 amino acids were generally lower than 50%. Such low yields led to overall low efficiency in glycoprotein synthesis. In order to improve the synthesis of glycoproteins, as we suggested above, it is necessary to not only significantly optimize the methods for the synthesis of glycans and glycopeptides (*i.e.*, development of more efficient manual

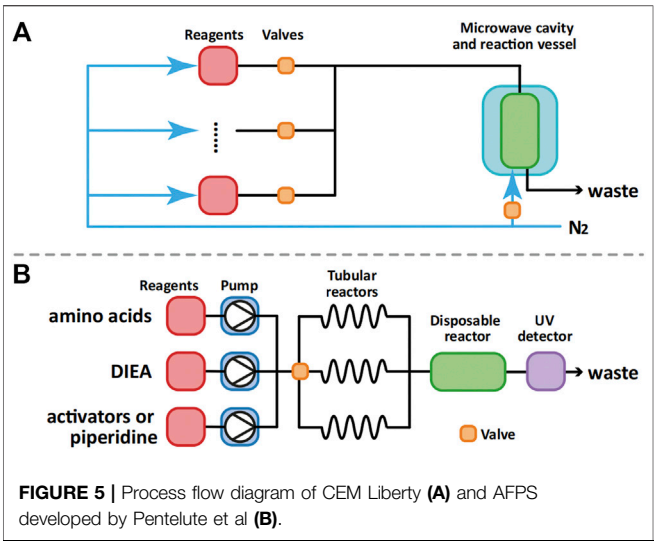


FIGURE 5 | Process flow diagram of CEM Liberty **(A)** and AFPS developed by Pentelute et al **(B)**.

synthetic methods), but also the methods for the synthesis of large peptides.

In recent years, many different methods have been explored to further improve the performance of automated peptide synthesizers in the preparation of large peptides. One simple and reliable approach is to heat the reactions in SPPS. The effect of an elevated temperature during SPPS has been controversial since it was originally proposed. On one hand, peptides are complex and may undergo side reactions at high temperature. On the other hand, as the length of the peptide chain increases, the intra- and intermolecular aggregation can significantly reduces the reaction efficiency, and increasing temperature may be a solution to this issue.

The microwave-assisted peptide synthesis has attracted much attention in the past decade. At the beginning of the development of this technology, it has been speculated that there are so-called non-thermal microwave effects on the synthesis because of the observed acceleration in rate and alteration in product distribution as compared with conventional heating (De La Hoz et al., 2005). However, more and more experiments indicate that there is no essential difference between different heating methods. The observed changes may be the results of thermal/kinetic effects (Bacsa et al., 2008).

Currently, the two major manufacturers of fully automated microwave-assisted peptide synthesizers are CEM and Biotage. Possibly due to the limitations of current microwave technology, the more efficient continuous flow method was not used in their synthesizers. Instead, they used nitrogen bubbling (CEM Liberty, **Figure 5A**) and vortexing (Biotage Syro Wave) to mix resin beads with solvent and reagents during SPPS. Another crucial limitation of microwave-assisted peptide synthesizers is their inability to monitor the progress of the synthesis visually.

Compared with the batch mode synthesis used in the microwave-assisted peptide synthesizers, the flow mode synthesis is more efficient in heat transfer, more economical and more flexible (Plutschack et al., 2017). Recently, Pentelute et al. demonstrated the advantages of flow peptide synthesis. Using their automated fast-flow synthesizer (AFPS, **Figure 5B**), they were able to successfully prepare a series of long peptides such as sortase A (59–206), which is 164 amino acids long (Hartrampf et al., 2020). Although the requirement of the use of large excess of amino acid building blocks (>40 eq) in the synthesis limits the application of such a peptide synthesizer in the synthesis of glycopeptides, its advantage in the preparation of large peptides is expected to have a positive impact on glycoprotein chemical synthesis.

CONCLUSION AND PERSPECTIVES

Although several decades of efforts and development have led to great achievements in glycoprotein synthesis, many challenges still exist in this research area. Currently, the chemical synthesis of glycoproteins is a complex, expensive and time-consuming process. In order to address this issue, optimization of the automated peptide synthesis is required. A possible solution is to apply the newly developed automated fast-flow peptide synthesizers to perform the synthesis of large peptide fragments. It is expected that if such synthesis is reliably realized, it should greatly promote the advancement of glycoprotein synthesis in the future.

AUTHOR CONTRIBUTIONS

JT, YL, BM, ZT, and SS wrote the paper.

FUNDING

The National Key R & D Program of China (Grant No. 2018YFE0111400), the CAMS Innovation Fund for Medical Sciences (CIFMS, 2021-I2M-1-026), the Training Program of the Major Research Plan of National Natural Science Foundation of China (Grant No. 91853120), the National Major Scientific and Technological Special Project of China (Grant Nos. 2018ZX09711001-005 and 2018ZX09711001-013), and the NIH Research Project Grant Program (R01 EB025892).

ACKNOWLEDGMENTS

We would like to thank the Ministry of Science of Technology of China, the Chinese Academy of Medical Sciences and Peking Union Medical College, the National Natural Science Foundation of China, the Division of Laboratory and Instrument Management of Tsinghua University, the State Key Laboratory of Bioactive Substance and Function of Natural Medicines, Institute of Materia Medica, and the National Institute of Health of the United States for funding.

REFERENCES

Asahina, Y., Kawakami, T., and Hojo, H. (2019). Glycopeptide Synthesis Based on a TFA-Labile Protection Strategy and One-Pot Four-Segment Ligation for the Synthesis of O-Glycosylated Histone H2A. *Eur. J. Org. Chem.* 2019, 1915–1920. doi:10.1002/ejoc.201801885

Bacsa, B., Horváti, K., Bõsze, S., Andreae, F., and Kappe, C. O. (2008). Solid-phase Synthesis of Difficult Peptide Sequences at Elevated Temperatures: a Critical Comparison of Microwave and Conventional Heating Technologies. *J. Org. Chem.* 73, 7532–7542. doi:10.1021/jo8013897

Bayer, E., Jun, G., Halász, I., and Sebestian, I. (1970). A New Support for Polypeptide Synthesis in Columns. *Tetrahedron Lett.* 51, 4503–4505. doi:10.1016/s0040-4039(01)83961-7

Behrendt, R., White, P., and Offer, J. (2016). Advances in Fmoc Solid-phase Peptide Synthesis. *J. Pept. Sci.* 22, 4–27. doi:10.1002/psc.2836

Breen, C. P., Nambiar, A. M. K., Jamison, T. F., and Jensen, K. F. (2021). Ready, Set, Flow! Automated Continuous Synthesis and Optimization. *Trends Chem.* 3, 373–386. doi:10.1016/j.trechm.2021.02.005

Carpino, L. A., and Han, G. Y. (1970). 9-Fluorenylmethoxycarbonyl Function, a New Base-Sensitive Amino-Protecting Group. *J. Am. Chem. Soc.* 92, 5748–5749. doi:10.1021/ja00722a043

Chaffey, P. K., Guan, X., Li, Y., and Tan, Z. (2018). Using Chemical Synthesis to Study and Apply Protein Glycosylation. *Biochemistry* 57, 413–428. doi:10.1021/acs.biochem.7b01055

Davis, B. G. (2002). Synthesis of Glycoproteins. *Chem. Rev.* 102, 579–602. doi:10.1021/cr0004310

Dawson, P. E., Muir, T. W., Clark-Lewis, I., and Kent, S. B. H. (1994). Synthesis of Proteins by Native Chemical Ligation. *Science* 266, 776–779. doi:10.1126/science.7973629

De La Hoz, A., Díaz-Ortiz, Á., and Moreno, A. (2005). Microwaves in Organic Synthesis. Thermal and Non-thermal Microwave Effects. *Chem. Soc. Rev.* 34, 164–178. doi:10.1039/b411438h

Du Vigneaud, V., Ressler, C., Swan, J. M., Roberts, C. W., and Katsoyannis, P. G. (1954). The Synthesis of Oxytocin1. *J. Am. Chem. Soc.* 76, 3115–3121. doi:10.1021/ja01641a004

Fernández-Tejada, A., Vadola, P. A., and Danishefsky, S. J. (2014). Chemical Synthesis of the β-Subunit of Human Luteinizing (hLH) and Chorionic Gonadotropin (hCG) Glycoprotein Hormones. *J. Am. Chem. Soc.* 136, 8450–8458. doi:10.1021/ja503545r

Furka, A., Sebestyén, F., Asgedom, M., and Dibó, G. (1991). General Method for Rapid Synthesis of Multicomponent Peptide Mixtures. *Int. J. Pept. Protein Res.* 37, 487–493. doi:10.1111/j.1399-3011.1991.tb00765.x

Gutte, B., and Merrifield, R. B. (1971). The Synthesis of Ribonuclease A. *J. Biol. Chem.* 246, 1922–1941. doi:10.1016/s0021-9258(18)62396-8

Hartrampf, N., Saebi, A., Poskus, M., Gates, Z. P., Callahan, A. J., Cowfer, A. E., et al. (2020). Synthesis of Proteins by Automated Flow Chemistry. *Science* 368, 980–987. doi:10.1126/science.abb2491

Izumi, M., Makimura, Y., Dedola, S., Seko, A., Kanamori, A., Sakono, M., et al. (2012). Chemical Synthesis of Intentionally Misfolded Homogeneous Glycoprotein: A Unique Approach for the Study of Glycoprotein Quality Control. *J. Am. Chem. Soc.* 134, 7238–7241. doi:10.1021/ja3013177

Li, C., and Wang, L.-X. (2018). Chemoenzymatic Methods for the Synthesis of Glycoproteins. *Chem. Rev.* 118, 8359–8413. doi:10.1021/acs.chemrev.8b00238

Li, H., Dao, Y., and Dong, S. (2017). "CHAPTER 5. Chemical Synthesis and Engineering of N-Linked Glycoproteins," in *Chemical Biology of Glycoproteins* (London: The Royal Society of Chemistry), 150–187. doi:10.1039/9781782623823-00150

Li, H., Zhang, J., An, C., and Dong, S. (2021). Probing N-Glycan Functions in Human interleukin-17A Based on Chemically Synthesized Homogeneous Glycoforms. *J. Am. Chem. Soc.* 143, 2846–2856. doi:10.1021/jacs.0c12448

Liu, L., Prudden, A. R., Bosman, G. P., and Boons, G.-J. (2017). Improved Isolation and Characterization Procedure of Sialylglycopeptide from Egg Yolk Powder. *Carbohydr. Res.* 452, 122–128. doi:10.1016/j.carres.2017.10.001

Lukas, T. J., Prystowsky, M. B., and Erickson, B. W. (1981). Solid-phase Peptide Synthesis under Continuous-Flow Conditions. *Proc. Natl. Acad. Sci. U.S.A.* 78, 2791–2795. doi:10.1073/pnas.78.5.2791

Ma, B., Guan, X., Li, Y., Shang, S., Li, J., and Tan, Z. (2020). Protein Glycoengineering: An Approach for Improving Protein Properties. *Front. Chem.* 8, 622. doi:10.3389/fchem.2020.00622

Marcaurelle, L. A., Mizoue, L. S., Wilken, J., Oldham, L., Kent, S. B. H., Handel, T. M., et al. (2001). Chemical Synthesis of Lymphotactin: A Glycosylated Chemokine with a C-Terminal Mucin-like Domain. *Chem. Eur. J.* 7, 1129–1132. doi:10.1002/1521-3765(20010302)7:5<1129::aid-chem1129>3.0.co;2-w

Merrifield, R. B. (1965). Automated Synthesis of Peptides. *Science* 150, 178–185. doi:10.1126/science.150.3693.178

Merrifield, R. B. (1964). Solid Phase Peptide Synthesis. II. The Synthesis of Bradykinin. *J. Am. Chem. Soc.* 86, 304–305. doi:10.1021/ja01056a056

Merrifield, R. B. (1963). Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* 85, 2149–2154. doi:10.1021/ja00897a025

Merrifield, R. B., and Stewart, J. M. (1965). Automated Peptide Synthesis. *Nature* 207, 522–523. doi:10.1038/207522a0

Nagorny, P., Sane, N., Fasching, B., Aussedat, B., and Danishefsky, S. J. (2012). Probing the Frontiers of Glycoprotein Synthesis: The Fully Elaborated β-Subunit of the Human Follicle-Stimulating Hormone. *Angew. Chem. Int. Ed.* 51, 975–979. doi:10.1002/anie.201107482

Okamoto, R., Kimura, M., Ishimizu, T., Izumi, M., and Kajihara, Y. (2014). Semisynthesis of a Post-translationally Modified Protein by Using Chemical Cleavage and Activation of an Expressed Fusion Polypeptide. *Chem. Eur. J.* 20, 10425–10430. doi:10.1002/chem.201403035

Pinilla, C., Appel, J. R., and Houghten, R. A. (1996). Tea Bag Synthesis of Positional Scanning Synthetic Combinatorial Libraries and Their Use for Mapping Antigenic Determinants. *Methods Mol. Biol.* 66, 171–179. doi:10.1385/0-89603-375-9:171

Piontek, C., Ring, P., Harjes, O., Heinlein, C., Mezzato, S., Lombana, N., et al. (2009a). Semisynthesis of a Homogeneous Glycoprotein Enzyme: Ribonuclease C: Part 1. *Angew. Chem. Int. Ed.* 48, 1936–1940. doi:10.1002/anie.200804734

Piontek, C., Varón Silva, D., Heinlein, C., Pöhner, C., Mezzato, S., Ring, P., et al. (2009b). Semisynthesis of a Homogeneous Glycoprotein Enzyme: Ribonuclease C: Part 2. *Angew. Chem. Int. Ed.* 48, 1941–1945. doi:10.1002/anie.200804735

Plutschack, M. B., Pieber, B., Gilmore, K., and Seeberger, P. H. (2017). The Hitchhiker's Guide to Flow Chemistry. *Chem. Rev.* 117, 11796–11893. doi:10.1021/acs.chemrev.7b00183

Reif, A., Siebenhaar, S., Tröster, A., Schmälzlein, M., Lechner, C., Velisetty, P., et al. (2014). Semisynthesis of Biologically Active Glycoforms of the Human Cytokine Interleukin 6. *Angew. Chem. Int. Ed.* 53, 12125–12131. doi:10.1002/anie.201407160

Rich, J. R., and Withers, S. G. (2009). Emerging Methods for the Production of Homogeneous Human Glycoproteins. *Nat. Chem. Biol.* 5, 206–215. doi:10.1038/nchembio.148

Sakamoto, I., Tezuka, K., Fukae, K., Ishii, K., Taduru, K., Maeda, M., et al. (2012). Chemical Synthesis of Homogeneous Human Glycosyl-Interferon-β that Exhibits Potent Antitumor Activity *In Vivo*. *J. Am. Chem. Soc.* 134, 5428–5431. doi:10.1021/ja2109079

Seko, A., Koketsu, M., Nishizono, M., Enoki, Y., Ibrahim, H. R., Juneja, L. R., et al. (1997). Occurrence of a Sialylglycopeptide and Free Sialylglycans in Hen's Egg Yolk. *Biochimica Biophysica Acta (BBA) - General Subj.* 1335, 23–32. doi:10.1016/s0304-4165(96)00118-3

Sun, B., Bao, W., Tian, X., Li, M., Liu, H., Dong, J., et al. (2014). A Simplified Procedure for Gram-Scale Production of Sialylglycopeptide (SGP) from Egg Yolks and Subsequent Semi-synthesis of Man3GlcNAc Oxazoline. *Carbohydr. Res.* 396, 62–69. doi:10.1016/j.carres.2014.07.013

Wang, P., Dong, S., Shieh, J.-H., Peguero, E., Hendrickson, R., Moore, M. A. S., et al. (2013). Erythropoietin Derived by Chemical Synthesis. *Science* 342, 1357–1360. doi:10.1126/science.1245095

Wang, S., Foster, S. R., Sanchez, J., Corcilius, L., Larance, M., Canals, M., et al. (2021). Glycosylation Regulates N-Terminal Proteolysis and Activity of the Chemokine CCL14. *ACS Chem. Biol.* 16, 973–981. doi:10.1021/acschembio.1c00006

Wang, X., Ashhurst, A. S., Dowman, L. J., Watson, E. E., Li, H. Y., Fairbanks, A. J., et al. (2020). Total Synthesis of Glycosylated Human Interferon-γ. *Org. Lett.* 22, 6863–6867. doi:10.1021/acs.orglett.0c02401

Ye, F., Zhao, J., Xu, P., Liu, X., Yu, J., Shangguan, W., et al. (2021). Synthetic Homogeneous Glycoforms of the SARS-CoV-2 Spike Receptor-Binding Domain Reveals Different Binding Profiles of Monoclonal Antibodies. *Angew. Chem. Int. Ed.* 60, 12904–12910. doi:10.1002/anie.202100543

Zhao, J., Liu, J., Liu, X., Cao, Q., Zhao, H., Liu, L., et al. (2022). Revealing Functional Significance of Interleukin-2 Glycoproteoforms Enabled by Expressed Serine Ligation. *Chin. J. Chem.* 40, 787–793. doi:10.1002/cjoc.202100914

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Chemistry

ORIGINAL RESEARCH
published: 15 June 2022
doi: 10.3389/fchem.2022.914698

Biochemical Characterization and Synthetic Application of WciN and Its Mutants From *Streptococcus pneumoniae* Serotype 6B

Wei Gong[1,2], *Min Liang*[1,3], *Jielin Zhao*[1,3], *Hong Wang*[1,3], *Zonggang Chen*[1,3], *Fengshan Wang*[1,2,3] *and Guofeng Gu*[1,3]*

[1]National Glycoengineering Research Center, Shandong Provincial Key Laboratory of Carbohydrate Chemistry and Glycobiology, Shandong University, Qingdao, China, [2]School of Pharmaceutical Science, Shandong University, Jinan, China, [3]NMPA Key Laboratory for Quality Research and Evaluation of Carbohydrate-based Medicine, Shandong University, Qingdao, China

The biochemical properties of α-1,3-galactosyltransferase WciN from *Streptococcus pneumoniae* serotype 6B were systemically characterized with the chemically synthesized Glcα-PP-$(CH_2)_{11}$-OPh as an acceptor substrate. The *in vitro* site-directed mutation of D38 and A150 residues of WciN was further investigated, and the enzymatic activities of those WciN mutants revealed that A150 residue was the pivotal residue responsible for nucleotide donor recognition and the single-site mutation could completely cause pneumococcus serotype switch. Using $WciN_{A150P}$ and $WciN_{A150D}$ mutants as useful tool enzymes, the disaccharides Galα1,3Glcα-PP-$(CH_2)_{11}$-OPh and Glcα1,3Glcα-PP-$(CH_2)_{11}$-OPh were successfully prepared in multi-milligram scale in high yields.

Keywords: ***Streptococcus pneumoniae*** **serotype 6B, capsular polysaccharides, galactosyltransferase, site mutation, enzymatic synthesis**

OPEN ACCESS

Edited by:
Jian Yin,
Jiangnan University, China

Reviewed by:
Jungui Dai,
Chinese Academy of Medical Sciences and Peking Union Medical College, China
Irina Bakunina,
Far Eastern Branch (RAS), Russia

***Correspondence:**
Guofeng Gu
guofenggu@sdu.edu.cn

Specialty section:
This article was submitted to Chemical Biology, a section of the journal Frontiers in Chemistry

Received: *07 April 2022*
Accepted: *02 May 2022*
Published: *15 June 2022*

Citation:
Gong W, Liang M, Zhao J, Wang H, Chen Z, Wang F and Gu G (2022) Biochemical Characterization and Synthetic Application of WciN and Its Mutants From Streptococcus pneumoniae Serotype 6B. Front. Chem. 10:914698. doi: 10.3389/fchem.2022.914698

INTRODUCTION

Pneumonia caused by the Gram-positive pathogen *Streptococcus pneumoniae* is a highly fatal infectious disease worldwide. The extracellular capsular polysaccharides (CPSs) abundantly coating the pneumococcal cell surfaces are recognized as one of the predominant causative virulence factors owing to their enhancing resistance to the complement-mediated opsonophagocytosis (Alonsodevelasco et al., 1995; Neeleman et al., 1999). Furthermore, these CPSs are identified as effective antigenic epitopes for the development of pneumococcal vaccines because of their inducing serotype-specific immunoprotection (Alonsodevelasco et al., 1995). Based on the confirmed serological profiles and unique CPS structures, more than 100 individual pneumococcal serotypes have been characterized and identified thus far (Henrichsen, 1995; Ganaie et al., 2020; Pimenta et al., 2021). Such a large diversity of CPS among pneumococcal serotypes has made it a huge challenge in CPS-related pneumococcal vaccine development.

Each capsular polysaccharide is programmatically synthesized by a series of enzymes encoded by the *cps* locus genes (Muñoz et al., 1997). It has been disclosed that the generation of antigenic diversity of CPS mainly contributed to elevated recombination and substitution rates of the *cps* locus (Mostowy et al., 2017). Although recombination within the *cps* locus has been assumed as the underlying cause for serotype evolution (Joshi et al., 2020), several studies revealed that serotype switching in pneumococcus was also implicated with the site-mutation of glycosyltransferase genes

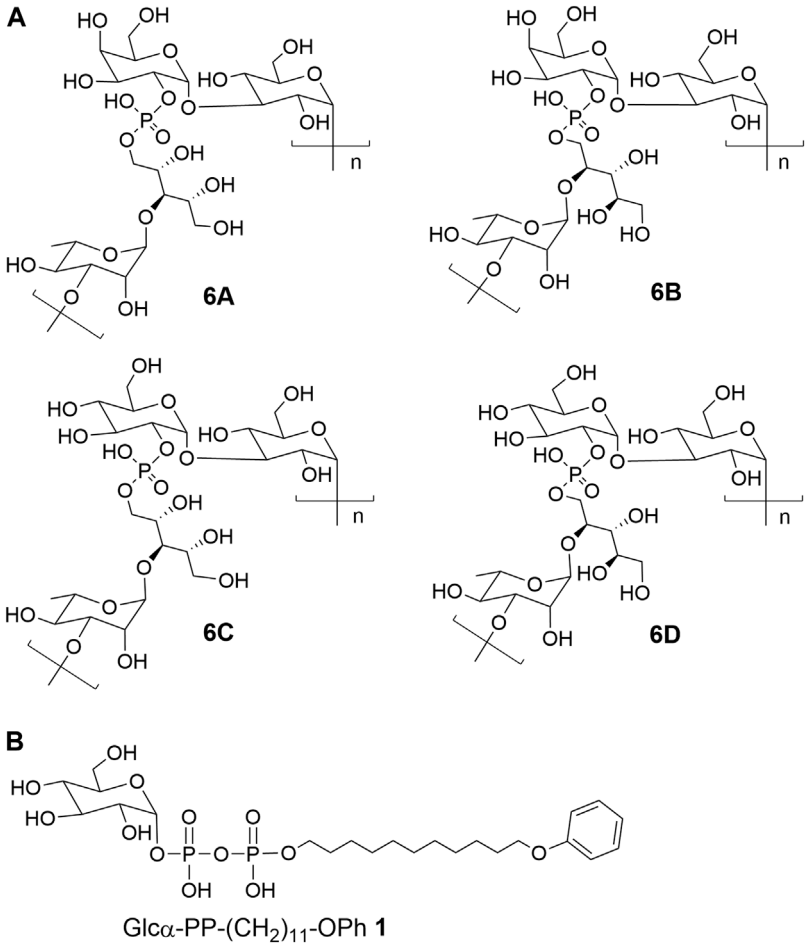


FIGURE 1 | The chemical structures of **(A)** CPS repeat units of pneumococcal serotypes **6A–D** and **(B)** the glycolipid Glcα-PP-$(CH_2)_{11}$-OPh **1**.

located in the *cps* locus (van Selm et al., 2003; Mavroidi et al., 2004; Mavroidi et al., 2007; Sheppard et al., 2010; Oliver et al., 2013). Serogroup six of *S. pneumoniae* has been characterized to contain eight serotypes including 6A-H (Park et al., 2015; van Tonder et al., 2015). Among them, the identified chemical structures of CPS 6A-D were shown in **Figure 1A**. The serotypes 6A and 6B have identical structures of capsular polysaccharides repeat unit (RU) only with a difference in the rhamnosidic linkage, which is α-1,3 linkage in 6A and α-1,4 linkage in 6B (Rebers and Heidelberger, 1961). Further studies revealed that the substitution of the catalytic triad residues, Ala192-Ser195-Arg254, in rhamnosyltransferase (WciP) could completely cause a serotype switch between 6A and 6B (Mavroidi et al., 2004; Sheppard et al., 2010). The CPS RUs of serotypes 6C and 6D have the glucose residue in place of the galactose residue in serotypes 6A and 6B through entire gene replacement of glycosyltransferase WciN, which has been respectively recognized as galactosyltransferase in 6A and 6B and as glucosyltransferase in 6C and 6D (Park et al., 2007a; Park et al., 2007b; Jin et al., 2009; Bratcher et al., 2010). Nevertheless, it has been disclosed that mutagenesis of A150 and/or D38 residues of WciN in serotype 6A or 6B resulted in a novel hybrid serotype 6F or 6G, which was identified as a different mixture ratio of 6A/6C for 6F, or 6B/6D for 6G, respectively (Oliver et al., 2013).

As mentioned above, WciN from pneumococcus type 6B strain is presumed as α-1,3-galactosyltransferase responsible for the RU assembly in the biosynthesis of type 6B CPS. Its galactosylation activity has been preliminarily investigated and identified *in vitro* using a chemosynthetic Und-PP-Glc surrogate, that is, Glcα-PP-$(CH_2)_{10}CH_3$, as an acceptor substrate (Han et al., 2012). However, its detailed biochemical properties have not been reported yet. Moreover, the *in vivo* allelic exchange study has disclosed that the mutagenesis of aspartic acid residue at position 38 and alanine residue at position 150 of WciN could trigger pneumococcus serotype switch. Thus, the more *in vitro* evidence to confirm such serotype evolution are further worthy of exploring. In this study, we presented in detail the biochemical properties of WciN derived from pneumococcus type 6B using the synthesized glycolipid Glcα-PP-$(CH_2)_{11}$-OPh **1** (Wang et al., 2019; Wang et al., 2021; Liang et al., 2022) (**Figure 1B**) as acceptor substrate, and carried out the single or dual amino acid substitution of D38 and/or A150 residue of

WciN to verify the different glycosylation functions of the resultant glycosyltransferase mutants.

MATERIALS AND METHODS

Materials

Sugar nucleotide donors UDP-Gal and UDP-Glc were prepared as previously described (Li et al., 2020). Enzymatic acceptor substrate Glcα-PP-$(CH_2)_{11}$-OPh **1** was synthesized followed the protocol reported previously (Liang et al., 2022). Ni^{2+} Sepharose high performance was the product of GE healthcare. Menthol HPLC grade was purchased from Thermo Fisher Scientific. Other chemicals and solvents used were of analytical grade.

Overexpression and Purification of WciN

The complete *wciN* gene of pneumococcus type 6B without terminator codon (GenBank: KT907353.1, 5253–6194) was synthesized and inserted into *Nde* I and *Xho* I of expression plasmid pET-28b by Sangon Biotech. The resulting recombinant plasmid pET-28b-*wciN* was then transformed into *E. coli* BL21 (DE3) competent cell for overexpression. The proper transformants were first grown at 37°C and 200 rpm in Luria-Bertani (LB) medium containing kanamycin (100 μg ml^{-1}). When the cell density reached an OD value of 0.6–0.8 at 600 nm, isopropyl 1-thio-β-D-galactopyranoside (IPTG) (0.3 mM) was added to the cell culture for recombinant protein induction. After subsequent cultivation at 16°C for another 20 h, cells were harvested and disrupted by ultrasonic treatment. The resulting lysate was centrifuged and the supernatant was subjected to a nickel affinity chromatography for enzyme purification with three different buffers: equilibration buffer (50 mM Tris, 500 mM NaCl, 10 mM imidazole, and pH 7.5), washing buffer (50 mM Tris, 500 mM NaCl, 50 mM imidazole, and pH 7.5), and elution buffer (50 mM Tris, 500 mM NaCl, 200 mM imidazole, and pH 7.5). The purity and homogeneity of WciN protein were analyzed by 12.5% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Its concentration was measured by using a Thermo Scientific™ NanoDrop One spectrometer that was calibrated with the extinction coefficient predicted by ExPASy (http://web.expasy.org/protparam/). Finally, the purified enzyme was stored at -80°C containing 20% glycerol (*v/v*).

Biochemical Characterization of WciN

The catalytic activity of the purified WciN was determined in a solution system as follows: 50 μg ml^{-1} of WciN, 5 mM $MgCl_2$, 1 mM UDP-Gal, and 1 mM Glcα-PP-$(CH_2)_{11}$-OPh **1** in 50 mM buffer. Reaction mixtures were performed for 10 min and then terminated by boiling at 100°C for 30 s. After centrifuging for 10 min under 12,000 rpm, the supernatant was analyzed with HPLC (DionexCarboPac™ PA-100 column, 4 × 250 mm, 0–1 M ammonium acetate buffer eluent). The byproduct UDP was monitored to assess the reaction process owing to its strong UV absorption at 260 nm and its convenience to be quantitated by HPLC (Wang et al., 2021; Liang et al., 2022).

The pH effect on enzyme activity was determined at pH values ranging from 6.0 to 10.5 with three different buffer systems at 37°C. The employed buffers included Bis-Tris-HCl (50 mM, pH 6.0, 6.5, and 7.0), Tris-HCl (50 mM, pH 7.0, 7.5, 8.0, 8.5, and 9.0), and Gly-NaOH (50 mM, pH 9.0, 9.5, 10.0, and 10.5). The optimal temperature for enzyme reaction was assessed at different temperatures (10, 16, 20, 25, 30, 37, 42, and 50°C) in 50 mM Gly-NaOH buffer (pH 9.0). To investigate the influence of metallic ions, enzyme activities were assayed in Gly-NaOH buffer (pH 9.0) with the presence of 5 mM following metal salts including ethylenediamine tetraacetic acid (EDTA), $MgCl_2$, $MnCl_2$, $CaCl_2$, $NiSO_4$, $CoSO_4$, $FeSO_4$, $CuSO_4$, and $ZnSO_4$. To obtain the optimized Mg^{2+} concentration, the enzymatic reactions were carried out under varying concentrations of Mg^{2+} (0.3125–80 mM). Heat-treated WciN was served as a negative control. The Relative activity concluded from the pH and temperature test was defined as the relative value to the maximum enzyme activity, and the effect of metal ions on enzyme activity was determined using the activity measured without adding ions as the reference value.

For acceptor substrate specificity study, Glcα-PP-$(CH_2)_{11}$-OPh, Glcα-PP-$(CH_2)_7$-CH_3, Glcα-P-$(CH_2)_{11}$-ONap, Glcβ-$(CH_2)_7$-CH_3, and Glcβ-$(CH_2)_{11}$-CH_3 were examined with UDP-Gal as nucleotide donor, respectively. For donor substrate specificity study, UDP-Gal, UDP-Glc, UDP-GalNAc, UDP-GlcNAc, and UDP-GlcA were examined with Glcα-PP-$(CH_2)_{11}$-OPh as acceptor substrate, respectively. The reaction was performed in the optimized conditions for WciN.

Site-Directed Mutagenesis of Key Amino Acids of WciN

Three single site mutated enzymes, namely $WciN_{D38N}$, $WciN_{A150T}$, and $WciN_{A150S}$, were obtained using pET-28b-*wciN* plasmid as a template. Another two dual sites mutated enzymes, $WciN_{D38N/A150S}$ and $WciN_{D38N/A150T}$ were created using pET-28b-$wciN_{D38N}$ as a template. All of the site mutations were carried out by the Fast Mutagenesis System (TransGen Biotech). Primers designed for corresponding amino acid substitution were listed in **Supplementary Table S1**. The mutant enzymes were overexpressed and purified following the similar protocols described earlier. Thereafter, UDP-Gal and UDP-Glc were applied to assay the donor recognition of the mutant enzymes, respectively. The reaction progress was monitored by thin-layer chromatography (TLC) or matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS). The developing solvent of TLC was a mixture of EtOAc/CH_3OH/H_2O/AcOH (*v/v/v/v*, 10/3/2/0.5), and the components on TLC were visualized by incubation at 180°C with a chromogenic solvent containing 93% ethyl alcohol, 3.5% sulfuric acid, 1% acetic acid, and 2.5% anisaldehyde.

To further investigate the effect of amino acid substitution at the position 150 of WciN on donor recognition, saturated mutation at A150 was systematically performed, and the corresponding oligonucleotide primers were shown in **Supplementary Table S1**. The catalytic activities of purified mutant enzymes were measured as described earlier. The molecular modeling of enzymes was conducted by PHYRE2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id = index).

Enzyme Kinetics of WciN Mutants

The enzymatic reactions were carried out under the aforementioned optimized conditions, that is, in Gly-NaOH buffer (50 mM, pH 9.0) containing varying concentrations UDP-Gal/UDP-Glc and Glcα-PP-$(CH_2)_{11}$-OPh **1** with 5 mM $MgCl_2$ at 37°C. Then, enzyme reactions using saturated UDP-sugar (4.0 mM) and varying concentrations of Glcα-PP-$(CH_2)_{11}$-OPh **1** (0.0625–4.0 mM) or Glcα-PP-$(CH_2)_{11}$-OPh **1** (1.0 mM) and varying concentrations of UDP-sugar (0.125–4.0 mM) were performed for 10 min. The Michaelis constant (K_m) and maximal velocity (V_{max}) values were graphed using the initial reaction velocities calculated from experimental data by GraphPad Prism 6.04 program.

Milligram-Scale Enzymatic Synthesis of Disaccharides Galα1,3-Glcα-PP-$(CH_2)_{11}$-OPh 2 and Glcα1,3-Glcα-PP-$(CH_2)_{11}$-OPh 3

A 10 ml reaction system containing 2.4 mM UDP-Gal, 2 mM Glcα-PP-$(CH_2)_{11}$-OPh **1**, 5 mM $MgCl_2$, and 100 μg ml^{-1} purified $WciN_{A150P}$ in 50 mM Gly-NaOH buffer (pH 9.0) was incubated at 37°C for 1 h with gently shaking. After acceptor substrate **1** was completely converted into disaccharide product as monitored by TLC and MADI-TOF-MS analysis, the reaction was then terminated by boiling for 30 s. The reaction mixture was centrifuged at 12,000 rpm to remove the formed precipitate, and the resulting supernatant was freeze-dried, resuspended in methanol, and then filtered for further purification. The filtrate was purified by the reversed phase HPLC using a C18 column (10 × 250 mm) and gradient eluent (10–100% methanol in water containing 10 mM NH_4HCO_3). The fractions containing the desired product were pooled and concentrated to afford Galα1,3-Glcα-PP-$(CH_2)_{11}$-OPh **2** (12.7 mg, 85%) as a white solid. ^{1}H NMR (600 MHz, CD_3OD): δ 7.22 (t, 2H, J = 7.8 Hz, Ph), 6.89–6.84 (m, 3H, Ph), 5.65 (dd, 1H, J = 7.8, 3.6 Hz, H-1^{Glc}), 5.34 (d, 1H, J = 2.4 Hz, H-1^{Gal}), 4.24 (t, 1H, J = 6.0 Hz, H-5^{Gal}), 4.00–3.56 (m, 7H, H-3,5^{Glc}, H-4^{Gal}, -OCH_2CH_2-, -CH$_2$CH_2OPh), 3.82 (dd, 1H, J = 12.0, 2.4 Hz, H-6a^{Glc}), 3.80–3.75 (m, 2H, H-2,3^{Gal}), 3.74 (dd, 1H, J = 11.4, 7.2 Hz, H-6a^{Gal}), 3.66 (dd, 1H, J = 11.4, 3.6 Hz, H-6b^{Gal}), 3.64 (dd, 1H, J = 12.0, 3.6 Hz, H-6b^{Glc}), 3.52 (t, 1H, J = 9.6 Hz, H-4^{Glc}), 3.47 (br d, 1H, J = 9.6 Hz, H-2^{Glc}), 1.77–1.71 (m, 2H, -CH_2CH$_2$-), 1.67–1.61 (m, 2H, -CH_2CH$_2$-), 1.49–1.42 (m, 2H, -CH$_2$CH_2-), 1.41–1.27 (m, 15H, -CH$_2$CH_2-); ^{13}C NMR (150 MHz, CD_3OD): δ 159.16 (Ph), 128.93 (2C, Ph), 119.99 (Ph), 114.04 (2C, Ph), 99.52 (C-1^{Gal}), 96.00 (d, $J_{C,P}$ = 6.0 Hz, C-1^{Glc}), 79.94 (C-3^{Glc}), 73.19 (C-5^{Glc}), 70.15 (d, $J_{C,P}$ = 7.5 Hz, C-2^{Glc}), 70.99 (C-5^{Gal}), 70.46 (C-4^{Glc}), 70.12 (C-3^{Gal}), 69.79 (C-4^{Gal}), 69.47 (C-2^{Gal}), 67.14 (-CH$_2$CH$_2$OPh), 65.94 (d, $J_{C,P}$ = 6.0 Hz, -OCH$_2$CH$_2$-), 61.35 (C-6^{Gal}), 61.11 (C-6^{Glc}), 30.34 (d, J = 7.5 Hz, -OCH$_2$CH$_2$CH$_2$-), 29.34, 29.32, 29.28, 29.14, 29.12, 29.02, 25.76, 25.46 (8 C, -OCH$_2$CH$_2$(CH$_2$)$_8$CH$_2$OPh); ^{31}P NMR (243 MHz, CD_3OD): δ -10.47 (d, J = 20.8 Hz) and -12.72 (d, J = 20.8 Hz); ESI-(-)-TOF HRMS m/z: calculated for $C_{29}H_{49}O_{18}P_2$ 747.2400 [M-H]$^-$; found 747.2391.

A 10 ml reaction mixture of 50 mM Gly-NaOH buffer (pH 9.0) containing 2.4 mM UDP-Glc, 2 mM Glcα-PP-$(CH_2)_{11}$-OPh **1**, 5 mM $MgCl_2$ and 200 μg ml^{-1} $WciN_{A150D}$ was incubated at 37°C for 4 h. The reaction was then worked up following the same protocol as described earlier, yielding Glcα1,3-Glcα-PP-$(CH_2)_{11}$-OPh **3** (12.4 mg, 83%) as a white solid. ^{1}H NMR (600 MHz, CD_3OD): δ 7.22 (t, 2H, J = 7.8 Hz, Ph), 6.89–6.85 (m, 3H, Ph), 5.65 (dd, 1H, J = 7.8, 3.6 Hz, H-1^{Glc}), 5.23 (d, 1H, J = 3.6 Hz, H-1$^{Glc'}$), 4.01–3.90 (m, 6H, H-5^{Glc}, H-5$^{Glc'}$, -OCH_2CH$_2$-, -CH$_2$CH_2OPh), 3.88–3.80 (m, 3H, H-3^{Glc,} H-6a^{Glc}, H-6a$^{Glc'}$), 3.67 (t, 1H, J = 9.6 Hz, H-3$^{Glc'}$), 3.65–3.58 (m, 2H, H-6b^{Glc}, H-6b$^{Glc'}$), 3.51 (t, 1H, J = 9.6 Hz, H-4^{Glc}), 3.47 (br d, 1H, J = 9.6 Hz, H-2^{Glc}), 3.40 (dd, 1H, J = 9.6, 3.6 Hz, H-2$^{Glc'}$), 3.23 (t, 1H, J = 9.6 Hz, H-4$^{Glc'}$), 1.77–1.71 (m, 2H, -CH_2CH$_2$-), 1.67–1.60 (m, 2H, -CH_2CH$_2$-), 1.49–1.42 (m, 2H, -CH$_2$CH_2-), 1.41–1.28 (m, 15H, -CH$_2$CH_2-); ^{13}C NMR (150 MHz, CD_3OD): δ 159.15 (Ph), 128.93 (2C, Ph), 119.99 (Ph), 114.04 (2C, Ph), 99.96 (C-1$^{Glc'}$), 95.95 (d, $J_{C,P}$ = 6.0 Hz, C-1^{Glc}), 81.75 (C-3^{Glc}), 73.78 (C-3$^{Glc'}$), 73.06 (C-5^{Glc}), 72.86 (C-2$^{Glc'}$), 72.31 (C-5$^{Glc'}$), 71.13 (d, $J_{C,P}$ = 7.5 Hz, C-2^{Glc}), 70.62 (C-4$^{Glc'}$), 70.27 (C-4^{Glc}), 67.41 (-CH_2CH$_2$OPh), 65.86 (d, $J_{C,P}$ = 6.0 Hz, -OCH$_2$CH$_2$-), 61.49 (C-6$^{Glc'}$), 61.14 (C-6^{Glc}), 30.36 (d, J = 7.5 Hz, -OCH$_2$CH$_2$CH$_2$-), 29.35, 29.32, 29.28, 29.15, 29.13, 29.02, 25.76, 25.48 (8 C, -OCH$_2$CH$_2$(CH$_2$)$_8$CH$_2$OPh); ^{31}P NMR (243 MHz, CD_3OD): δ -10.40 (d, J = 20.8 Hz) and -12.68 (d, J = 20.8 Hz); ESI-(-)-TOF HRMS m/z: calculated for $C_{29}H_{49}O_{18}P_2$ 747.2400 [M-H]$^-$; found 747.2394.

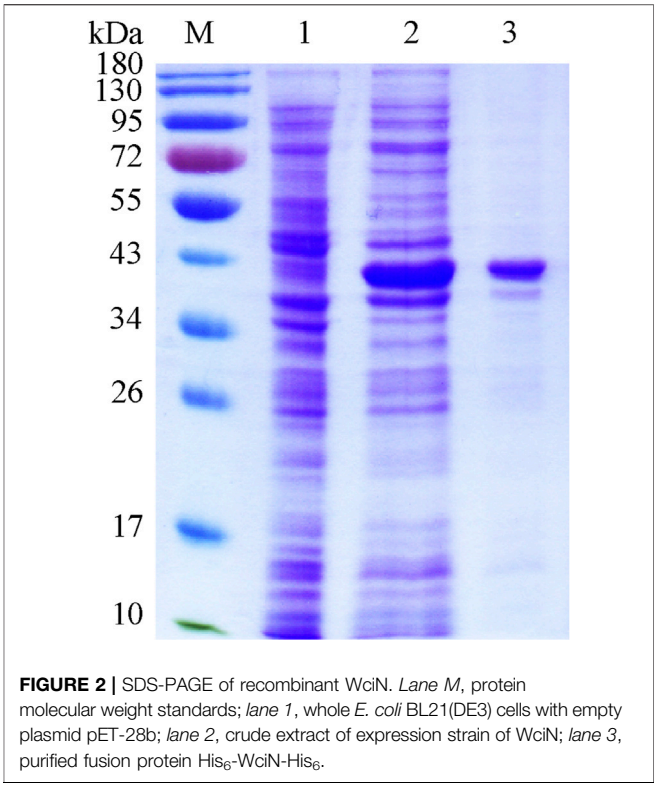


FIGURE 2 | SDS-PAGE of recombinant WciN. *Lane M*, protein molecular weight standards; *lane 1*, whole *E. coli* BL21(DE3) cells with empty plasmid pET-28b; *lane 2*, crude extract of expression strain of WciN; *lane 3*, purified fusion protein His_6-WciN-His_6.

RESULTS AND DISCUSSION

Overexpression and Purification of WciN

The recombinant plasmid pET-28b-*wciN* was designed to encode the full length of WciN with two His_6 tags at its both N- and C-terminus for later convenient protein purification. The His_6-WciN-His_6 fusion protein was overexpressed and purified readily to homogeneity *via* Nickel-chelation affinity chromatography.

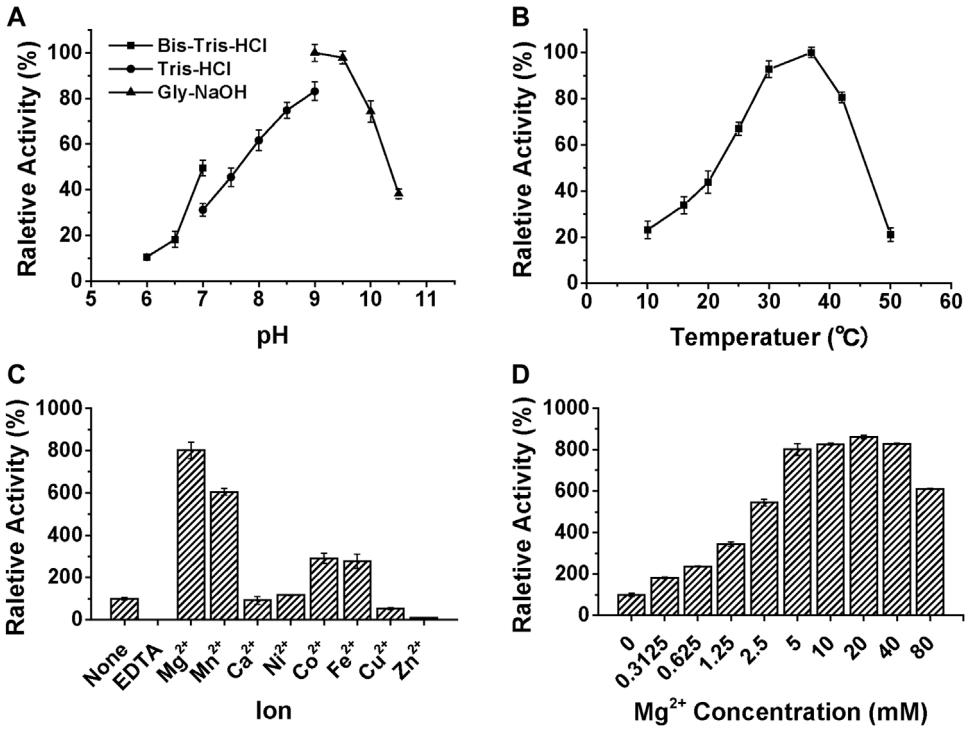


FIGURE 3 | The influences of pH **(A)**, temperature **(B)**, ions **(C)**, and Mg^{2+} concentration **(D)** on the relative activity of WciN to catalyze galactosylation of Glcα-PP-$(CH_2)_{11}$-OPh **1** with UDP-Gal.

The SDS-PAGE depicted in **Figure 2** showed a distinct band at ~40 kDa that was coincident with the theoretically calculated molecular weight (40.01 kDa) of recombinant WciN. In addition, its expression level was determined as 15.2 mg per liter using a NanoDrop One spectrophotometer.

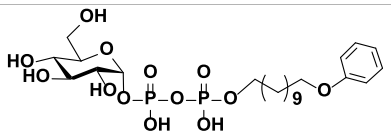
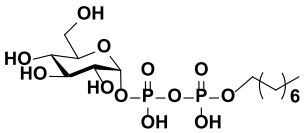
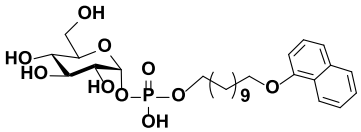
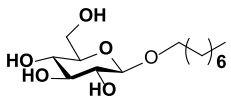
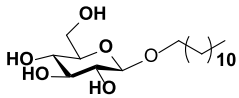
Biochemical Properties of the Recombinant WciN

We have recently reported the optimized synthesis of the glycolipid Glcα-PP-$(CH_2)_{11}$-OPh **1** (Liang et al., 2022) and utilized it as an acceptor substrate to characterize several bacterial glycosyltransferases related to CPS RU biosynthesis (Wang et al., 2019; Wang et al., 2021; Liang et al., 2022). Therefore, using it as an enzymatic substrate, the detailed biochemical properties of WciN were then investigated. The enzymatic activities were analyzed by means of spectrophotometric analysis of the by-product UDP, and the results were shown in **Figure 3**. The better activity (>60%) of WciN was observed under weak alkali conditions (pH 8.0–10.0), and the optimal pH value for its activity was 9.0 (**Figure 3A**). WciN enzyme was highly active (>80%) from 30 to 42°C, and the best glycosylation activity was determined at 37°C (**Figure 3B**). Furthermore, the existence of EDTA could completely inhibit enzyme activity (**Figure 3C**), indicating WciN might belong to GT-A glycosyltransferase family (Lairson et al., 2008; Han et al., 2012). Among eight tested divalent cation ions, Cu^{2+} and Zn^{2+} obviously reduced enzyme activity, and Ca^{2+} and Ni^{2+} slightly affected its activity, whilst Mg^{2+}, Mn^{2+}, Co^{2+}, and Fe^{2+} exhibited remarkable promoting ability on enzyme activity (**Figure 3C**). It has been disclosed that a 5~8 fold improvement of WciN activity was achieved in the presence of Mn^{2+} or Mg^{2+}. In addition, the influence of Mg^{2+} concentrations on WciN activity was also examined. As shown in **Figure 3D**, the activity of WciN improved sharply under a broad range of 5–40 mM of Mg^{2+} concentrations but without any significant difference. Collectively, the optimal reaction conditions for WciN enzyme were established to be 5 mM Mg^{2+} in 50 mM Gly-NaOH buffer with pH 9.0 at 37°C.

Acceptor Substrate Specificity of WciN

The specificity of WciN toward five acceptor substrates was investigated with the earlier optimized reaction conditions using UDP-Gal as the nucleotide donor (**Table 1**). The enzymatic reactions were monitored by TLC and HRMS (**Supplementary Figure S1**). Among five sugar acceptors, only Glcα-PP-$(CH_2)_{11}$-OPh and Glcα-PP-$(CH_2)_7$-CH_3, which had diphosphate moiety in structure, could be well recognized by WciN, whereas Glcα-P-$(CH_2)_{11}$-ONap with monophosphate moiety and the other two acceptors, Glcβ-$(CH_2)_{11}$-CH_3 and Glcβ-$(CH_2)_7$-CH_3, without any phosphate moiety exhibited none detectable activity. These results indicated that the diphosphate moiety in the acceptor substrate played an important role in the acceptor recognition of WciN.

TABLE 1 | Investigation of acceptor recognition of WciN.

No.	Acceptor	Structure	Activity
1	Glcα-PP-$(CH_2)_{11}$-OPh		+
2	Glcα-PP-$(CH_2)_7$-CH_3		+
3	Glcα-P-$(CH_2)_{11}$-ONap		–
4	Glcβ-$(CH_2)_7$-CH_3		–
5	Glcβ-$(CH_2)_{11}$-CH_3		–

+, detectable activity by TLC and HRMS; –, no detectable activity.

Nucleotide Donor Recognition of WciN Enzyme and Its Mutants

In a previous study, Nahm and co-workers have disclosed that the amino acid replacement at the position 150 and/or 38 of WciN could alter donor substrate specificities, resulting in the emergence of two new hybrid serotypes 6F and 6G (Oliver et al., 2013). To further confirm this conclusion, five mutant enzymes including $WciN_{D38N}$, $WciN_{A150T}$, $WciN_{A150S}$, $WciN_{D38N/A150T}$, and $WciN_{D38N/A150S}$ were accordingly designed, overexpressed, and purified to homogeneity. Using UDP-Gal and UDP-Glc as sugar nucleotide donors, the activities of these mutant enzymes were detected with TLC (**Supplementary Figures S2A,B**) and then analyzed by HPLC (**Figure 4**). Compared to wild-type WciN, all mutated enzymes could well recognize UDP-Gal donor (**Supplementary Figure S2A**) but exhibit reduced galactosylation activities in different degrees (40–80% relative activities) (**Figure 4**). Interestingly, except $WciN_{D38N}$ mutant, all other four WciN mutants could also accept UDP-Glc donor and showed weak to the good catalytic ability for glucosylation (**Supplementary Figure S2B**). Moreover, the glucosylation abilities of $WciN_{A150S}$ and $WciN_{D38N/A150S}$ mutants were significantly higher than those exerted by $WciN_{A150T}$ and $WciN_{D38N/A150T}$ mutants (**Figure 4**). These aforementioned results disclosed that the A150 residue of WciN was the pivotal residue responsible for nucleotide donor recognition and its mutation could alter nucleotide donor recognition, whereas mutation of the D38 residue could only decrease the enzymatic activity but not affect its donor specificity. All these findings almost coincided with the results reported previously (Oliver et al., 2013).

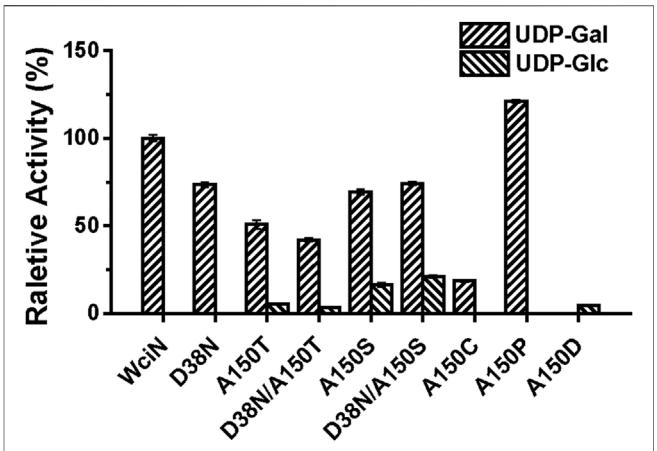


FIGURE 4 | Relative activities of WciN mutant enzymes toward UDP-Gal and UDP-Glc donors.

Thereafter, saturated mutation on A150 residue of WciN was executed using the designed primers listed in **Supplementary Table S1**. Except $WciN_{A150T}$ and $WciN_{A150S}$, other seventeen mutants were obtained and their enzymatic activities were accordingly examined using the same protocol described earlier. As shown in **Supplementary Figures S2C,D**, $WciN_{A150C}$ and $WciN_{A150P}$ mutants exhibited the catalytic activity to only recognize UDP-Gal as donor substrate with increased activity for $WciN_{A150P}$ (121% relative activity) and a decreased activity for $WciN_{A150C}$ (19% relative activity), whereas $WciN_{A150D}$ showed the capability to only accept UDP-Glc as donor substrate even with a lower glucosylation activity (~5% relative activity, **Figure 4**). Nevertheless, the rest mutated enzymes did not exert any detectable enzymatic activities toward UDP-Gal or UDP-Glc. All these results suggested that rational residue replacement at position 150 of WciN could affect the recognizable capability against donor substrate or change its donor specificity. In addition, WciN and its active mutants could not recognize other UDP-sugars, such as UDP-GlcNAc, UDP-GalNAc, and UDP-GlcA, indicating their relative donor specificity. Incidentally, the acceptor specificity of these mutants coincided well with that of wild type WciN enzyme.

In order to explore how the residue replacement at Ala150 of WciN affected its enzyme activity, molecular modeling of WciN and its mutants $WciN_{A150T}$ and $WciN_{A150P}$ were conducted using LgtC, a retaining galactosyltransferase from *Neisseria meningitides*, as the modeling template (Persson et al., 2001). The Gln189 residue of LgtC was located at the catalytic center and interacted with nucleotide donor through *van der* Waals, whilst interacted with acceptor substrate with the assistant of Ala154 residue through several hydrogen bonds (Persson et al., 2001). Alignment of the amino acid

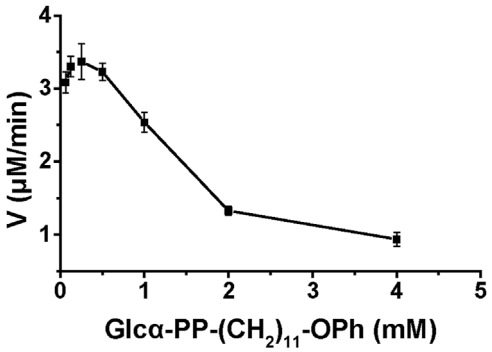


FIGURE 5 | Influence of acceptor concentration on enzyme reaction velocity.

TABLE 2 | Enzyme kinetics using UDP-Gal and UDP-Glc as donor.

Enzyme	UDP-Gal		UDP-Glc	
	K_m (mM)	V_{max} (μM/min)	K_m (mM)	V_{max} (μM/min)
WciN	5.5 ± 0.33	133.5 ± 5.28	—	—
$WciN_{A150P}$	7.06 ± 0.78	151.8 ± 11.79	—	—
$WciN_{A150D}$	—	—	3.52 ± 1.18	0.39 ± 0.08

—, *no data.*

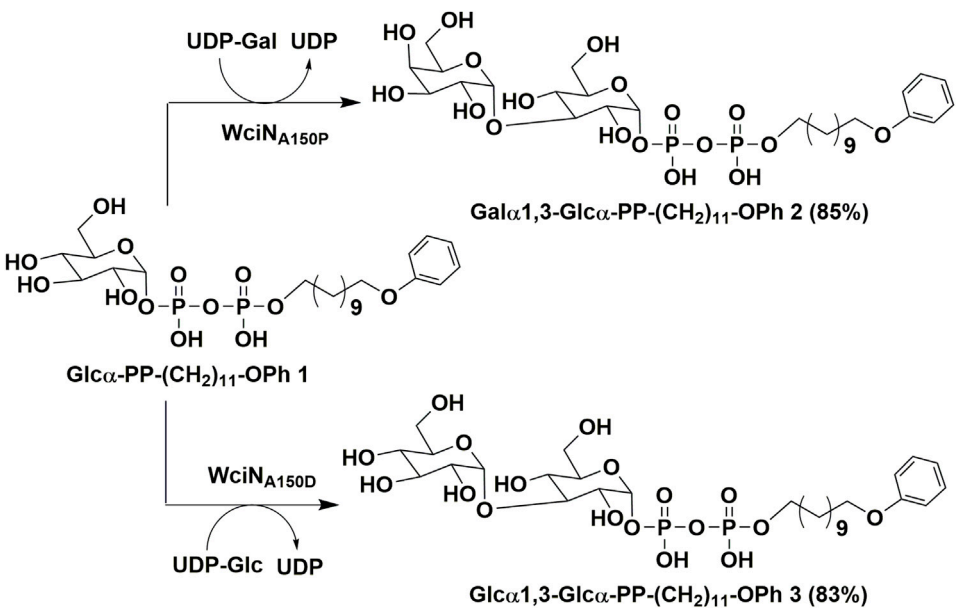
sequence of WciN with that of LgtC indicated that Ala150 and Leu185 residues of WciN corresponded to Ala154 and Gln189 residues of LgtC, respectively (**Supplementary Figure S3**). Therefore, it indicated that the residues Ala150 of WciN indirectly influenced substrate recognition by affecting its Leu185 residue.

Enzyme Kinetics of WciN Enzyme and Its Mutants

The enzyme kinetics of WciN and its mutants were examined at optimal reaction conditions established earlier. The influence of acceptor substrate concentration on the enzyme activity of WciN was explored first. As depicted in **Figure 5**, the reaction velocity catalyzed by WciN was dramatically reduced as the concentration of Glcα-PP-$(CH_2)_{11}$-OPh **1** was greater than 0.25 mM, indicating the activity of WciN could be easily inhibited even by a slightly higher concentration of acceptor substrate. This finding was radically different from those reported glycosyltransferases that could well accept Glcα-PP-$(CH_2)_{11}$-OPh **1** as substrate acceptor (Wang et al., 2019; Liang et al., 2022). Bioinformatics analysis of the amino acid sequence of WciN revealed that there was no transmembrane domain in WciN, lacking a domain that interacts with the membrane. Thus, the great difference in recognizing acceptor substrates by WciN might be remarkably affected by the variation of substrate micelles formed in aqueous solution due to the amphipathic character of Glcα-PP-$(CH_2)_{11}$-OPh **1**. Therefore, the enzyme kinetics of WciN and its mutants, $WciN_{A150P}$ and $WciN_{A150D}$, toward nucleotide donor (UDP-Gal or UDP-Glc) were then briefly measured using 0.125–4.0 mM UDP-Gal/UDP-Glc and Glcα-PP-$(CH_2)_{11}$-OPh **1** at suitable concentration (1.0 mM). The K_m and V_{max} values were calculated from Michaelis–Menten plots and listed in **Table 2**.

Preparation of Disaccharide Products 2 and 3

As outlined in **Scheme 1**, using Glcα-PP-$(CH_2)_{11}$-OPh **1** as substrate acceptor, the disaccharide products Galα1,3-Glcα-PP-$(CH_2)_{11}$-OPh **2** and Glcα1,3-Glcα-PP-$(CH_2)_{11}$-OPh **3** were



SCHEME 1 | Enzymatic synthesis of Galα1,3-Glcα-PP-$(CH_2)_{11}$-OPh **2** and Glcα1,3-Glcα-PP-$(CH_2)_{11}$-OPh **3**.

efficiently prepared under the earlier optimized reaction conditions by $WciN_{A150P}$ and $WciN_{A150D}$ mutants, respectively. Each enzymatic reaction proceeded smoothly and was monitored timely by TLC analysis, and terminated within 1–4 h after the full consumption of **1**. The disaccharides **2** and **3** were then obtained in milligram quantities and high yields of 83–85% after semi-preparative HPLC purification. Furthermore, the correct stereo-/regio-selectivity of each disaccharide product functioned by the WciN mutant was well verified with the assistance of the 1D and 2D NMR spectra. The small $^3J_{1,2}$ coupling constants (2.4 Hz for **2**; 3.6 Hz for **3**) of doublet peaks of the H-$1^{Gal/Glc'}$ signals in their ^{1}H NMR spectra indicated the new formation of α-glyosidic bonds, whilst the observed correlation signals of C-1^{Gal}/H-3^{Glc} and H-$1^{Glc'}$/C-3^{Glc} in their gHMBC spectra confirmed the regioselective formation of the 1,3-glycosidic linkages.

CONCLUSION

In this study, the detailed biological characterization of WciN from pneumococcus type 6B strain was investigated *in vitro*. The single or dual site-directed mutagenesis with D38 and A150 residues of WciN followed by comparison of glycosylation activities of the resultant mutant enzymes revealed that A150 residue played the pivotal role in altering donor recognition (Oliver et al., 2013). Accordingly, saturated mutation at position 150 of WciN was implemented, and only $WciN_{A150T}$, $WciN_{A150S}$, $WciN_{A150C}$, $WciN_{A150P}$, and $WciN_{A150D}$ mutants exhibited the catalytic activity in different degree. Among them, $WciN_{A150T}$ and $WciN_{A150S}$ mutants were recognized as bi-specific glycosyltransferases that could catalyze both galactosylation and glucosylation. Furthermore, $WciN_{A150P}$ mutant showed remarkably increased capability (121% relative activity) for recognition on UDP-Gal with comparison to that of wild WciN enzyme, whereas $WciN_{A150D}$ mutant completely abolished the recognition capability toward UDP-Gal, but could accept UDP-Glc as sole donor substrate even with very low activity (~5% relative activity). All these findings indicated that the single site-mutation of galactosyltransferase WciN at A150 residue could cause the different recognition toward nucleotide donor and thus trigger complete pneumococcus serotype switch. Finally, using $WciN_{A150P}$ and $WciN_{A150D}$ mutants as useful tool enzymes, disaccharide products Galα1,3Glcα-PP-$(CH_2)_{11}$-OPh **2** and Glcα1,3Glcα-PP-$(CH_2)_{11}$-OPh **3** were successfully achieved in multi-milligram scale.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

GG conceived the project and supervised the study. WG and ML conducted the experiments and completed the manuscript. All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This work was supported by grants from the National Natural Science Foundation of China (Grant Numbers 21877074 and 21672129) and the Science and Technology Development Plan of Shandong Province (Nos. 2019GSF108018 and 2016GGH4502).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2022.914698/full#supplementary-material

REFERENCES

Alonsodevelasco, E., Verheul, A. F., Verhoef, J., and Snippe, H. (1995). *Streptococcus pneumoniae*: Virulence Factors, Pathogenesis, and Vaccines. *Microbiol. Rev.* 59 (4), 591–603. doi:10.1128/mr.59.4.591-603.1995

Bratcher, P. E., Kim, K.-H., Kang, J. H., Hong, J. Y., and Nahm, M. H. (2010). Identification of Natural Pneumococcal Isolates Expressing Serotype 6D by Genetic, Biochemical and Serological Characterization. *Microbiol* 156 (2), 555–560. doi:10.1099/mic.0.034116-0

Ganaie, F., Saad, J. S., McGee, L., van Tonder, A. J., Bentley, S. D., Lo, S. W., et al. (2020). A New Pneumococcal Capsule Type, 10D, Is the 100th Serotype and Has a Large Cps Fragment from an Oral Streptococcus. *mBio* 11 (3), e00937–20. doi:10.1128/mBio.00937-20

Han, W., Cai, L., Wu, B., Li, L., Xiao, Z., Cheng, J., et al. (2012). The wciN Gene Encodes an α-1,3-Galactosyltransferase Involved in the Biosynthesis of the Capsule Repeating Unit of *Streptococcus pneumoniae* Serotype 6B. *Biochemistry* 51 (29), 5804–5810. doi:10.1021/bi300640b

Henrichsen, J. (1995). Six Newly Recognized Types of *Streptococcus pneumoniae*. *J. Clin. Microbiol.* 33 (10), 2759–2762. doi:10.1128/jcm.33.10.2759-2762.1995

Jin, P., Kong, F., Xiao, M., Oftadeh, S., Zhou, F., Liu, C., et al. (2009). First Report of Putative *Streptococcus pneumoniae* Serotype 6D Among Nasopharyngeal Isolates from Fijian Children. *J. Infect. Dis.* 200 (9), 1375–1380. doi:10.1086/606118

Joshi, S. S., Al-Mamun, M. A., and Weinberger, D. M. (2020). Correlates of Nonrandom Patterns of Serotype Switching in Pneumococcus. *J. Infect. Dis.* 221 (10), 1669–1676. doi:10.1093/infdis/jiz687

Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. (2008). Glycosyltransferases: Structures, Functions, and Mechanisms. *Annu. Rev. Biochem.* 77 (1), 521–555. doi:10.1146/annurev.biochem.76.061005.092322

Li, S., Wang, H., Jin, G., Chen, Z., and Gu, G. (2020). Exploring the Broad Nucleotide Triphosphate and Sugar-1-Phosphate Specificity of Thymidylyltransferase Cps23FL from *Streptococcus pneumoniae* Serotype 23F. *RSC Adv.* 10, 30110–30114. doi:10.1039/d0ra05799a

Liang, M., Gong, W., Sun, C., Zhao, J., Wang, H., Chen, Z., et al. (2022). Sequential One-Pot Three-Enzyme Synthesis of the Tetrasaccharide Repeating Unit of

Group B *Streptococcus* Serotype VIII Capsular Polysaccharide. *Chin. J. Chem.* 40, 1039–1044. doi:10.1002/cjoc.202100822

Mavroidi, A., Aanensen, D. M., Godoy, D., Skovsted, I. C., Kaltoft, M. S., Reeves, P. R., et al. (2007). Genetic Relatedness of the *Streptococcus pneumoniae* Capsular Biosynthetic Loci. *J. Bacteriol.* 189 (21), 7841–7855. doi:10.1128/jb.00836-07

Mavroidi, A., Godoy, D., Aanensen, D. M., Robinson, D. A., Hollingshead, S. K., and Spratt, B. G. (2004). Evolutionary Genetics of the Capsular Locus of Serogroup 6 Pneumococci. *J. Bacteriol.* 186 (24), 8181–8192. doi:10.1128/jb.186.24.8181-8192.2004

Mostowy, R. J., Croucher, N. J., De Maio, N., Chewapreecha, C., Salter, S. J., Turner, P., et al. (2017). Pneumococcal Capsule Synthesis Locus Cps as Evolutionary Hotspot with Potential to Generate Novel Serotypes by Recombination. *Mol. Biol. Evol.* 34 (10), 2537–2554. doi:10.1093/molbev/msx173

Muñoz, R., Mollerach, M., López, R., and García, E. (1997). Molecular Organization of the Genes Required for the Synthesis of Type 1 Capsular Polysaccharide of *Streptococcus pneumoniae*: Formation of Binary Encapsulated Pneumococci and Identification of Cryptic dTDP-Rhamnose Biosynthesis Genes. *Mol. Microbiol.* 25 (1), 79–92.

Neeleman, C., Geelen, S. P. M., Aerts, P. C., Daha, M. R., Mollnes, T. E., Roord, J. J., et al. (1999). Resistance to Both Complement Activation and Phagocytosis in Type 3 Pneumococci Is Mediated by the Binding of Complement Regulatory Protein Factor H. *Infect. Immun.* 67 (9), 4517–4524. doi:10.1128/iai.67.9.4517-4524.1999

Oliver, M. B., van der Linden, M. P. G., Küntzel, S. A., Saad, J. S., and Nahm, M. H. (2013). Discovery of *Streptococcus pneumoniae* Serotype 6 Variants with Glycosyltransferases Synthesizing Two Differing Repeating Units. *J. Biol. Chem.* 288 (36), 25976–25985. doi:10.1074/jbc.m113.480152

Park, I. H., Geno, K. A., Yu, J., Oliver, M. B., Kim, K.-H., and Nahm, M. H. (2015). Genetic, Biochemical, and Serological Characterization of a New Pneumococcal Serotype, 6H, and Generation of a Pneumococcal Strain Producing Three Different Capsular Repeat Units. *Clin. Vaccine Immunol.* 22 (3), 313–318. doi:10.1128/cvi.00647-14

Park, I. H., Park, S., Hollingshead, S. K., and Nahm, M. H. (2007a). Genetic Basis for the New Pneumococcal Serotype, 6C. *Infect. Immun.* 75 (9), 4482–4489. doi:10.1128/iai.00510-07

Park, I. H., Pritchard, D. G., Cartee, R., Brandao, A., Brandileone, M. C. C., and Nahm, M. H. (2007b). Discovery of a New Capsular Serotype (6C) within Serogroup 6 of *Streptococcus pneumoniae*. *J. Clin. Microbiol.* 45 (4), 1225–1233. doi:10.1128/jcm.02199-06

Persson, K., Ly, H. D., Dieckelmann, M., Wakarchuk, W. W., Withers, S. G., and Strynadka, N. C. J. (2001). Crystal Structure of the Retaining Galactosyltransferase LgtC from *Neisseria meningitidis* in Complex with Donor and Acceptor Sugar Analogs. *Nat. Struct. Biol.* 8 (2), 166–175. doi:10.1038/84168

Pimenta, F., Moiane, B., Gertz, R. E., Jr, Chochua, S., Snippes Vagnone, P. M., Lynfield, R., et al. (2021). New Pneumococcal Serotype 15D. *J. Clin. Microbiol.* 59 (5), e00329–21. doi:10.1128/JCM.00329-21

Rebers, P. A., and Heidelberger, M. (1961). The Specific Polysaccharide of Type VI Pneumococcus. II[1]. The Repeating Unit[2]. *J. Am. Chem. Soc.* 83 (14), 3056–3059. doi:10.1021/ja01475a021

Sheppard, C. L., Pichon, B., George, R. C., and Hall, L. M. C. (2010). *Streptococcus pneumoniae* Isolates Expressing a Capsule with Epitopes of Both Serotypes 6A and 6B. *Clin. Vaccine Immunol.* 17 (11), 1820–1822. doi:10.1128/cvi.00335-10

van Selm, S., van Cann, L. M., Kolkman, M. A. B., van der Zeijst, B. A. M., and van Putten, J. P. M. (2003). Genetic Basis for the Structural Difference between *Streptococcus pneumoniae* Serotype 15B and 15C Capsular Polysaccharides. *Infect. Immun.* 71 (11), 6192–6198. doi:10.1128/iai.71.11.6192-6198.2003

van Tonder, A. J., Bray, J. E., Roalfe, L., White, R., Zancolli, M., Quirk, S. J., et al. (2015). Genomics Reveals the Worldwide Distribution of Multidrug-Resistant Serotype 6E Pneumococci. *J. Clin. Microbiol.* 53 (7), 2271–2285. doi:10.1128/jcm.00744-15

Wang, H., Li, S., Xiong, C., Jin, G., Chen, Z., Gu, G., et al. (2019). Biochemical Studies of a β-1,4-rhamnoslytransferase from *Streptococcus pneumoniae* Serotype 23F. *Org. Biomol. Chem.* 17 (5), 1071–1075. doi:10.1039/c8ob02795a

Wang, H., Sun, C., Sun, X., Zhang, L., Zhao, J., Liang, M., et al. (2021). Biochemical Characterization and Synthetic Application of α-1,3-Glucosyltransferase from Pneumococcus Serotype 18C. *ChemCatChem* 13 (14), 3350–3356. doi:10.1002/cctc.202100507

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Chemistry

ORIGINAL RESEARCH
published: 20 June 2022
doi: 10.3389/fchem.2022.920009

Comprehensive Plasma N-Glycoproteome Profiling Based on EThcD-sceHCD-MS/MS

Yonghong Mao[1], Tao Su[2], Tianhai Lin[3], Hao Yang[2], Yang Zhao[4], Yong Zhang[2]* and Xinhua Dai[4]**

[1]Institute of Thoracic Oncology, West China Hospital, Sichuan University, Chengdu, China, [2]Institutes for Systems Genetics, West China Hospital, Sichuan University, Chengdu, China, [3]Department of Urology, Institute of Urology, West China Hospital, Sichuan University, Chengdu, China, [4]Mass Spectrometry Engineering Technology Research Center, Center for Advanced Measurement Science, National Institute of Metrology, Beijing, China

Glycoproteins are involved in a variety of biological processes. More than one-third of the plasma protein biomarkers of tumors approved by the FDA are glycoproteins, and could improve the diagnostic specificity and/or sensitivity. Therefore, it is of great significance to perform the systematic characterization of plasma N-glycoproteome. In previous studies, we developed an integrated method based on the combinatorial peptide ligand library (CPLL) and stepped collision energy/higher energy collisional dissociation (sceHCD) for comprehensive plasma N-glycoproteome profiling. Recently, we presented a new fragmentation method, EThcD-sceHCD, which outperformed sceHCD in the accuracy of identification. Herein, we integrated the combinatorial peptide ligand library (CPLL) into EThcD-sceHCD and compared the performance of different mass spectrometry dissociation methods (EThcD-sceHCD, EThcD, and sceHCD) in the intact N-glycopeptide analysis of prostate cancer plasma. The results illustrated that EThcD-sceHCD was better than EThcD and sceHCD in the number of identified intact N-glycopeptides (two-folds). A combination of sceHCD and EThcD-sceHCD methods can cover almost all glycoproteins (96.4%) and intact N-glycopeptides (93.6%), indicating good complementarity between the two. Our study has great potential for medium- and low-abundance plasma glycoprotein biomarker discovery.

Keywords: mass spectrometry, plasma, N-glycoproteomics, combinatorial peptide ligand library, EThcD-sceHCD

OPEN ACCESS

Edited by:
Zhongping Tan,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

Reviewed by:
Yuan Tian,
United States Food and Drug Administration, United States
Ying Zhang,
Fudan University, China

****Correspondence:***
Yang Zhao
zhaoy@nim.ac.cn
Yong Zhang
nankai1989@foxmail.com
Xinhua Dai
daixh@nim.ac.cn

Specialty section:
This article was submitted to Chemical Biology, a section of the journal Frontiers in Chemistry

Received: *14 April 2022*
Accepted: *09 May 2022*
Published: *20 June 2022*

Citation:
Mao Y, Su T, Lin T, Yang H, Zhao Y, Zhang Y and Dai X (2022) Comprehensive Plasma N-Glycoproteome Profiling Based on EThcD-sceHCD-MS/MS. Front. Chem. 10:920009. doi: 10.3389/fchem.2022.920009

1 INTRODUCTION

Human plasma is a critical area of clinical and fundamental research, as it contains a large number of disease candidate biomarkers (Jacobs et al., 2005). Protein biomarkers in the plasma change in concentration or state associated with a biological status or disease, offering great potential for patient diagnosis, risk stratification, and disease prevention (Moremen et al., 2012; Silsirivanit, 2019). More than one-third of plasma tumor protein biomarkers approved by the FDA are glycoproteins (Silsirivanit, 2019). Glycosylation has been recognized as one of the most multifunctional protein modifications (Moremen et al., 2012). It plays a vital role in the progression of various cancers (Ohtsubo and Marth, 2006). However, the dynamic range of plasma proteins can exceed 10^9, and many potential biomarkers are low-abundance proteins (Jacobs et al., 2005). Moreover, glycoproteomic analysis is difficult due to the microheterogeneity and macroheterogeneity of glycosylation, and other special properties (Pujic and Perreault, 2021). Hence, comprehensive

identification of human plasma glycoproteome (including intact glycopeptides, glycoproteins, glycosites, and glycans) is an important way to discover new biomarkers.

In the past few years, some researchers have made contributions to study plasma proteome and glycoproteome. In 2005, the Human Plasma Proteome Project identified a total of 7518 proteins and isoforms in plasma (Muthusamy et al., 2005). As a result of improvements in proteomic technologies, 10,546 plasma proteins were included in this database in 2014 (Nanjappa et al., 2014). However, the database of human plasma glycoproteome is still immature. This may be due to difficulties in sample processing, mass spectrometry analysis, and data processing. Still, experts in the field are doing their best to overcome these difficulties. For example, many enrichment materials or methods were developed to remove non-glycoproteins or non-glycopeptides in plasma (Palmisano et al., 2010; Sun et al., 2017; Yang et al., 2017; Zhang et al., 2017; Wu et al., 2018; Chen et al., 2019; Sun et al., 2019; Zhang et al., 2019; Zhang et al., 2020a). Moreover, in order to deplete high-abundance proteins, immunodepletion technologies and ProteoMiner protein enrichment methods were used to remove the abundant plasma proteins (Luque-Garcia and Neubert, 2007; Sennels et al., 2007).

In recent years, various fragmentation techniques (EThcD, sceHCD, EThcd-sceHCD, etc.) have appeared as valuable approaches for plasma glycoproteomics. EThcD means ETD followed by supplemental HCD, which fragments parent ions via ETD first and then the products ions are fragmented via HCD (Zhang et al., 2018), while sceHCD means stepped collision energy HCD (beam-type collisional activation) on Orbitrap systems (Riley et al., 2020). Many intact glycopeptide search engines, such as MSFragger-Glyco, Byonic, pGlyco, Glyco-Decipher, and StrucGP, have improved the accuracy of intact glycopeptide identification (Bern et al., 2012; Liu et al., 2017; Lu et al., 2020; Polasky et al., 2020; Shen et al., 2021; Fang et al., 2022). Kawahara et al. evaluated several search strategies and provided valuable information for serum or plasma glycoproteomic studies (Kawahara et al., 2021).

In previous studies, we developed Glyco-CPLL for human plasma N-glycoproteome profiling based on sceHCD and established a large database (Zhang et al., 2020b). CPLL means combinatorial peptide ligand library (a diverse library of hexapeptides that act as binders for proteins), which is a plasma sample preparation tool used for the compression of the dynamic range of the protein concentration and maintaining representatives of all proteins. When plasma samples were applied to the CPLL beads, the high-abundance proteins saturated their high-affinity ligands, and excess proteins were washed away. In contrast, the medium- and low-abundance proteins were concentrated on their specific affinity ligands (Sennels et al., 2007). Recently, we integrated EThcD and sceHCD into a glycoproteomic workflow (Zhang et al., 2021a; Zhang et al., 2021b). The results clearly showed that EThcD-sceHCD can improve the intact glycopeptide analysis performance of HIV-1 gp120, IgG subclasses, and complex clinical samples (Zhang et al., 2021a; Zhang et al., 2021b; Zeng et al., 2022).

Herein, we aim to improve the accuracy and depth of human plasma intact glycopeptide identification based on the new analysis method. More precisely, we integrated Glyco-CPLL into EThcD-sceHCD, compared the performance of different dissociation methods (EThcD-sceHCD, EThcD, and sceHCD), and determined its superiority for plasma N-glycoproteomic studies.

2 EXPERIMENTAL SECTION

2.1 Materials

Chemical reagents, such as dithiothreitol (DTT), iodoacetamide (IAA), formic acid (FA), trifluoroacetic acid (TFA), Tris base, and urea, were purchased from Sigma (St. Louis, MO, United States). Acetonitrile (ACN), ethanol (EtOH), methanol (MeOH), and acetic acid (HAc) were purchased from Merck (Darmstadt, Germany). Sequencing grade trypsin and Lys-C were obtained from Promega (Madison, WI, United States). ProteoMiner column was purchased from Bio-Rad Laboratories (Hercules, CA). Zwitterionic HILIC (Zic-HILIC) was purchased from Fresh Bioscience (Shanghai, China). The C8 extraction disks were purchased from 3M Empore (St. Paul, MN, United States). All other materials were purchased from Sigma-Aldrich or Thermo Fisher Scientific.

2.2 Biospecimen Collection

Diagnosis and confirmation of PCa patients were performed in the Department of Urology, West China Hospital of Sichuan University, Chengdu, China. Blood was collected into EDTA anticoagulant tubes. After centrifuging at 300 g for 10 min at 4°C, plasma was collected and stored at −80°C. Written informed consents were collected. The experiment was performed in accordance with the guidelines of the Chinese Medical Ethics Committee. The experiment was approved by the Ethics Committee at West China Hospital.

2.3 Sample Pretreatment

Plasma samples from PCa patients were pooled before analysis. The plasma proteins were prepared using the CPLL method as described before. Specifically, 200 μl pooled plasma was loaded into the ProteoMiner column and incubated at room temperature for 2 h. After centrifuging at 1,000 g for 30 s at 4°C, the column was washed with 200 μl of PBS and ddH_2O, and the CPLL bound proteins were eluted with 20 μl of elution buffer (8 M urea, 2% CHAPS).

2.4 Reduction, Alkylation, and Digestion

The eluent was proteolyzed following the filter-aided sample preparation (FASP) protocol. Briefly, after diluting 20 times with UA solution (8 M urea in 0.1 M Tris-HCl, pH 8.5), the eluent was added to a 30-kDa filter. After carrying out reduction reaction by adding 20 mM DTT for 4 h at 37°C, alkylation reaction was carried out by adding 50 mM IAA and incubating the mixture in the dark for 1 h. Proteins were digested by adding trypsin/Lys-C (1:50) to each filter tube. The peptides were collected by washing three times with 100 μl of water.

2.5 Intact Glycopeptide Enrichment

Intact N-glycopeptides were enriched using Zic-HILIC materials. Specifically, 200 μg of tryptic peptides and 10 mg of Zic-HILIC materials were mixed in 70% ACN/0.2% TFA solution. Then, the mixture was transferred to a pipette tip packed with a C8 membrane. Hydrophobic peptides were washed with 70% ACN/0.2% TFA, and intact N-glycopeptides were eluted with 70 μl of 0.1% TFA and collected in a 1.5-ml tube. The eluent was dried using a SpeedVac for further analysis.

2.6 LC-MS/MS Analysis

The dried intact N-glycopeptides were resuspended in 20 μL of 0.1% FA individually. Then 5 μL of samples was taken for analysis on an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher, United States). All intact N-glycopeptides were separated on a column (ReproSil-Pur C18-AQ, 1.9 μm, 75 μm inner diameter, length 20 cm; Dr Maisch) over a 78-min gradient at a flow rate of 300 nL/min. Three different fragmentation modes (EThcD, sceHCD, and EThcD-sceHCD) were used for intact N-glycopeptide analysis.

For EThcD-MS/MS and sceHCD-MS/MS, the parameters were as follows: MS1 was analyzed in the range of 800–2000 m/*z* at an Orbitrap resolution of 60,000. The RF lens, AGC target, MIT, exclusion duration, and cycle time were 40%, custom, 50 ms, 15 s, and 3 s, respectively. The precursor ion in MS2 experiment was performed at 2 m/*z* and acquired at an Orbitrap resolution of 30,000. The AGC target and MIT were custom and 150 ms, respectively. EThcD collision energy was 35%, while the sceHCD mode was turned on with an energy difference of ±10% (20-30-40%).

For EThcD-sceHCD-MS/MS, the analysis was performed using an alternative fragmentation between the EThcD and sceHCD modes in a duty cycle. In the EThcD duty cycle, MS1 was analyzed in the range 800–2000 m/*z* at an Orbitrap resolution of 60,000. The RF lens, AGC target, MIT, and exclusion duration were 40%, 2.0 e^5, 50 ms, and 15 s, respectively. MS2 was analyzed at 2 m/*z* at an Orbitrap resolution of 30,000. The AGC target, MIT, and EThcD type were standard, 150 ms, and 35%, respectively. In the sceHCD duty cycle, MS1 was analyzed in the range 800–2000 m/*z* at an Orbitrap resolution of 60,000. The RF lens, AGC target, MIT, exclusion duration, and cycle time were 40%, standard, auto, 15 s, and 1 s, respectively. The precursor ion in the MS2 experiment was selected at 1.6 m/*z* and acquired at an Orbitrap resolution of 30,000. The AGC target, MIT, and HCD collision energy were 200%, auto, and 30%, respectively. Moreover, the sceHCD mode was turned on with an energy difference of ±10% (20-30-40%). Although the data-dependent mode cycle time of each method was set as 3 s, there were differences in scanning speed between different modes (ETD has a slower scan speed). Therefore, the number of precursors selected for MS/MS for each method may be different.

2.7 Data Analysis

The data files were searched against the Human UniProt database (version 2015_03, 20,410 entries) using Byonic software (version 3.10.10, Protein Metrics, Inc.). Mass tolerance for precursors and fragment ions were set as ± 6 ppm and ± 20 ppm, respectively. Two missed cleavage sites were allowed. Carbamidomethyl (C) was set as fixed modification. Variable modifications contained oxidation (M) and acetyl (protein N-term). Additionally, the "182 human N-glycans" was set as the N-glycan modification. Protein groups were filtered to 1% FDR. Quality control methods for intact N-glycopeptide identification included a Byonic score of over 200, a logProb value of over 2, and at least five amino acids. ANOVA was used for statistical comparison among three groups, and Student's *t*-test was used for statistical comparison between two groups (SPSS Statistics 19.0). The homogeneity test was performed. The error bar denotes SD. p-value < 0.01 was considered significant. The raw data can be obtained *via* ProteomeXchange with identifier PXD030622.

3 RESULTS AND DISCUSSION

Recently, we proved that EThcD-sceHCD has better performance in the intact glycopeptide analysis of HIV-1 gp120 and IgG subclasses (Zhang et al., 2021a; Zhang et al., 2021b). However, whether this method can be applied to complex plasma samples is unknown. To evaluate the efficiency of the method on human plasma samples and systematically characterize plasma N-glycoproteome in PCa patients, the following experiment was designed (**Figure 1**). In detail, low-abundance proteins (LAPs) were extracted from pooled PCa patients' plasma using the ProteoMiner protein enrichment system which contains the combinatorial peptide ligand library (CPLL) (Bandow, 2010; Moggridge et al., 2019; Zhang et al., 2020b; Palstrøm et al., 2020). After digestion by trypsin and Lys-C, Zic-HILIC materials were used to enriched intact N-glycopeptides (Di Palma et al., 2011). Then, the same number of samples were analyzed using three fragmentation methods (EThcD, sceHCD, and EThcD-sceHCD) (**Figure 1**). All raw data files were searched using Byonic software. The search results were analyzed statistically and compared systematically (**Supplementary Table S1**).

Using the same quality control standards as described earlier, the identified N-glycoPSMs, N-glycans, intact N-glycopeptides, and N-glycoproteins from pooled PCa patients' plasma were analyzed and compared. EThcD-sceHCD outperformed EThcD and sceHCD in every number of identification (p < 0.001, ANOVA) (**Figure 2**). In particular, EThcD-sceHCD identified nearly twice (232) as many intact N-glycopeptides as EThcD (129) and sceHCD (126) (**Figure 2**). It has been reported that sceHCD can provide more information on fragment ions than other modes (Riley et al., 2020; Klein and Zaia, 2020). However, it cannot provide the accurate glycosite location and glycan composition information when one sequence contains more than one glycosite (Zhang et al., 2021b). EThcD can produce a greater proportion of fragment ions via both ETD and HCD, and thereby provide more key information for the unambiguous identification of both glycosites and glycans (Yu et al., 2017; Saba et al., 2012; Ma et al., 2016). Nevertheless, ETD has limited dissociation efficiency. Hence, we proposed EThcD-sceHCD, which alternatively fragment samples between EThcD and sceHCD modes in a duty cycle (Zhang et al., 2021b). That is, EThcD-sceHCD has better spectra quality and higher dissociation efficiency. For example, the three methods can

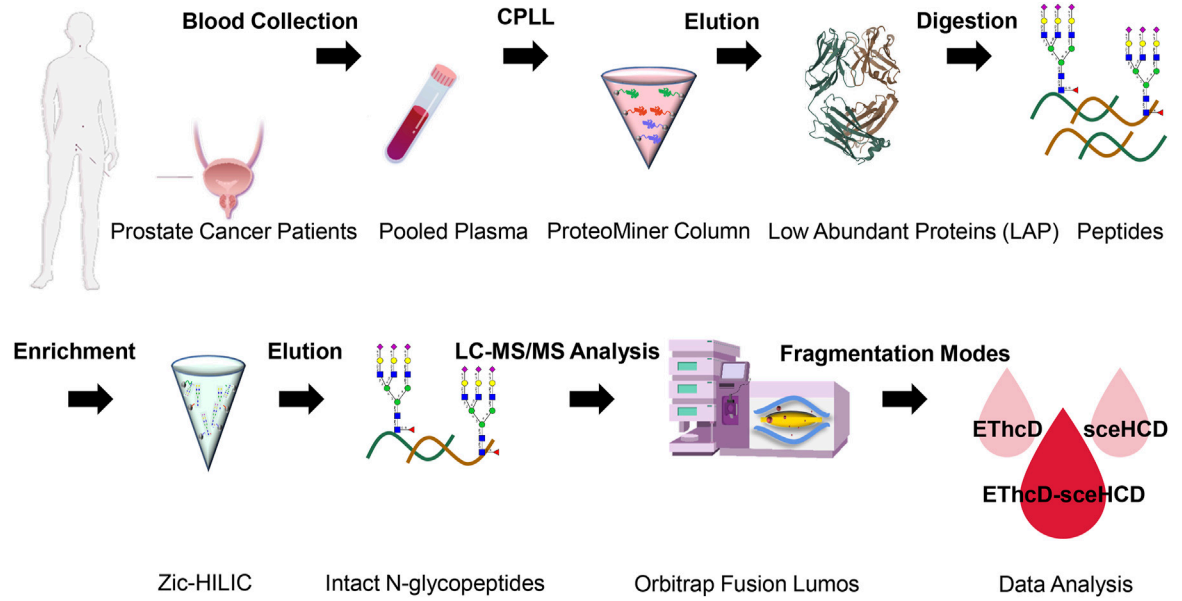


FIGURE 1 | Schematic representation of the workflow for human plasma LAP intact N-glycopeptide analysis using different dissociation methods (EThcD, sceHCD, and EThcD-sceHCD).

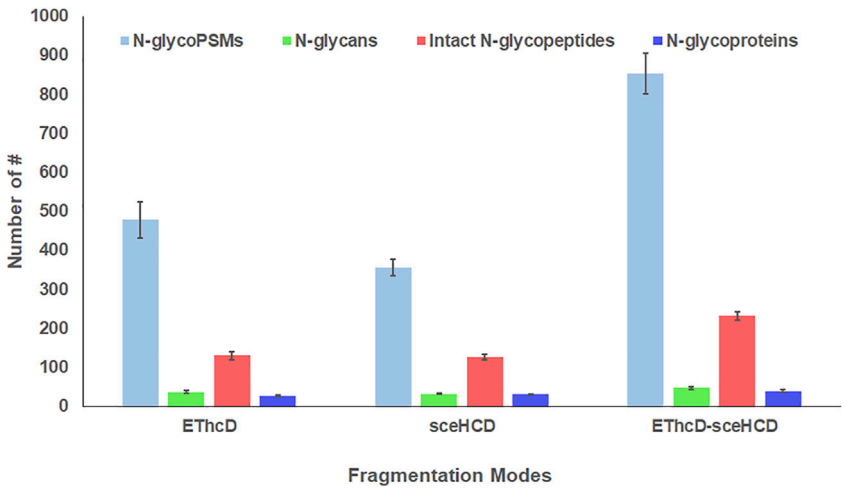


FIGURE 2 | Comparison of the number of N-glycoPSMs, N-glycans, intact N-glycopeptides, and N-glycoproteins from pooled PCa patients' plasma using different dissociation methods. (ANOVA was used for the statistical comparison among three groups, and Student's *t*-test was used for the statistical comparison between two groups. The homogeneity test was performed. The error bar denotes SD.)

provide abundant information about N-glycosite (N143) localization and the N-glycan composition (HexNAc(4)Hex (5)NeuAc (1)) of one N-glycopeptide from prothrombin (**Figure 3**). Both EThcD and EThcD-sceHCD can provide complex and informative fragment ions (glycan fragments, b/y/c/z ions, and Y ions) (**Figures 3A**, **C**). However, sceHCD provided less information (glycan fragments, b/y ions, and Y ions) (**Figure 3B**). In other words, EThcD and EThcD-sceHCD can provide more accurate N-glycosylation modification information than sceHCD, although they have lower intensity than sceHCD because ETD has limited scanning speed (**Figure 3**). In addition, we analyzed the largest Byonic score distribution of intact N-glycopeptide spectra from different fragmentation modes (**Supplementary Table S2**). Byonic score is the "raw" indicator of PSM correctness, reflecting the absolute quality of the PSM (Bern et al., 2012). As expected, the largest Byonic scores of the 69.8% (252/361) intact N-glycopeptides were obtained by EThcD-sceHCD, and the other 17.5% (63/361) and 12.7% (46/361) intact N-glycopeptides were obtained from EThcD and sceHCD, respectively (**Supplementary Table S2**).

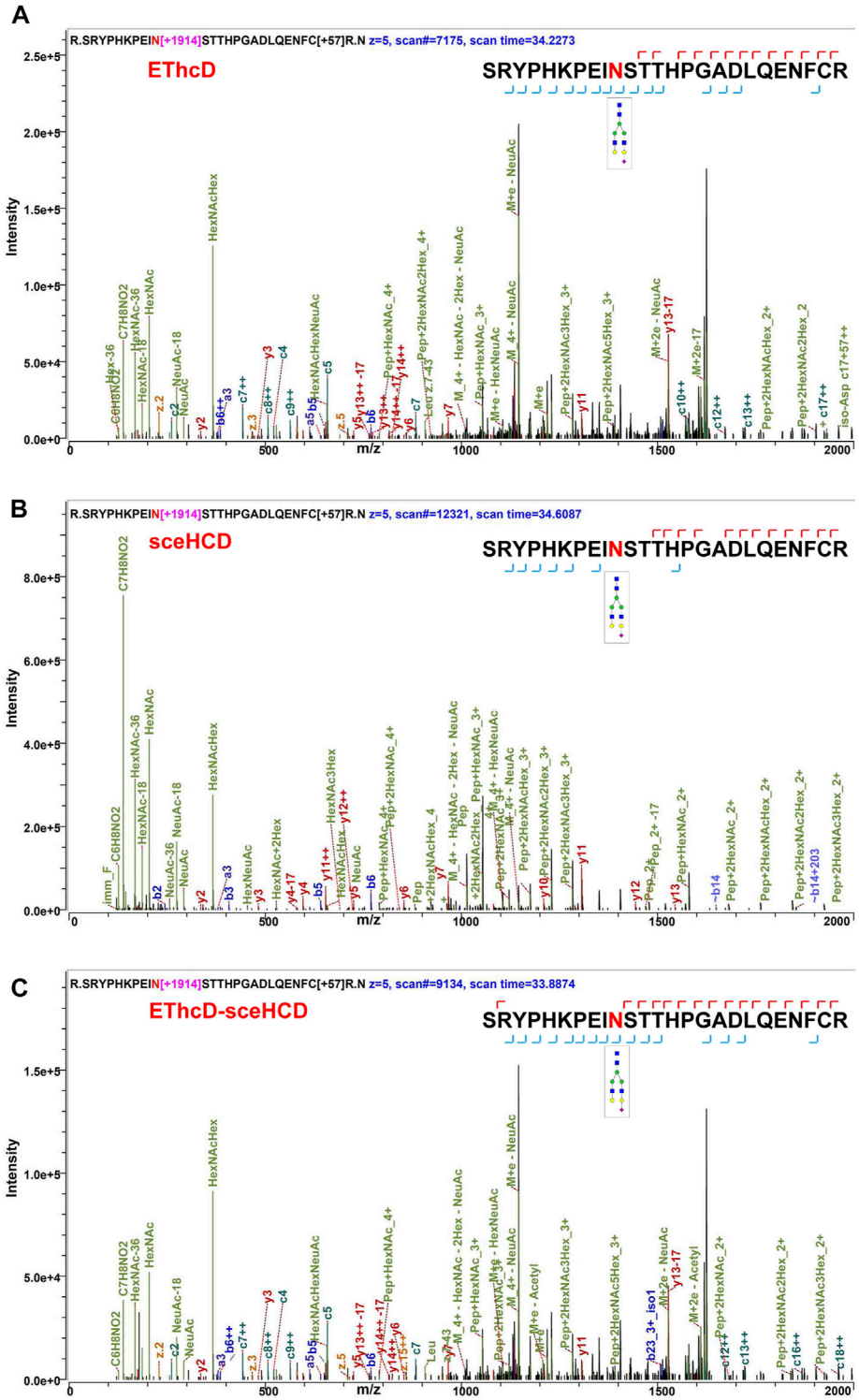


FIGURE 3 | Comparison of the EThcD **(A)**, secHCD **(B)**, and EThcD-sceHCD **(C)** spectra of prothrombin intact N-glycopeptides (N143) from human plasma.

The results showed that EThcD-sceHCD can fragment peptide backbone more extensively than EThcD and sceHCD. The reason may be that different peptide backbones obtained better fragmentation ions under different fragmentation modes. Bertozzi et al. compared multiple mass spectrometry dissociation methods and concluded that sceHCD significantly outperformed EThcD for N-glycopeptide identification from HEK293 whole cell lysates (Riley et al., 2020). Our results

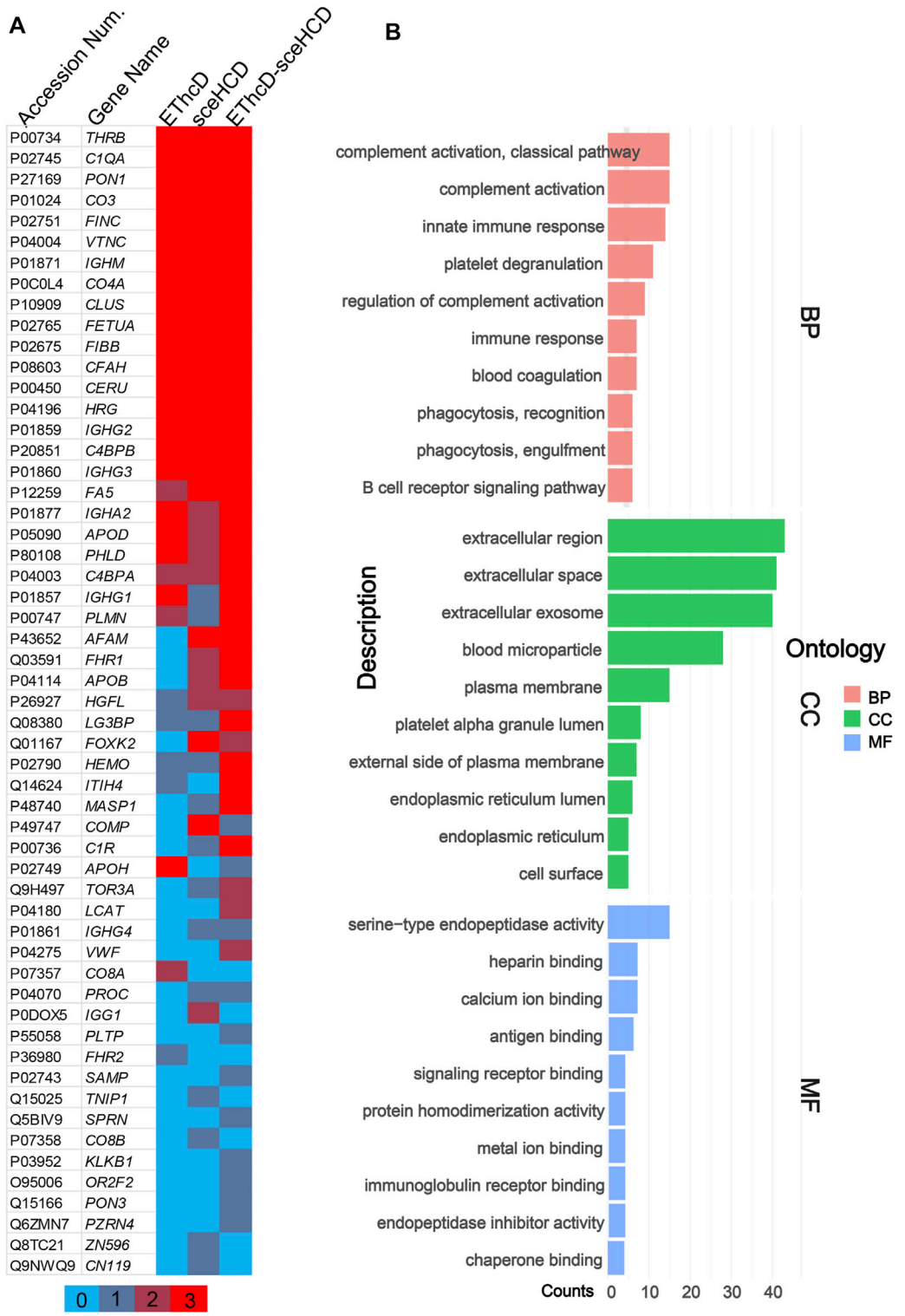


FIGURE 4 | Global analysis of identified plasma N-glycoproteins. **(A)** Heat map of identified plasma N-glycoproteins using different fragmentation modes. Colored lines indicate the number of repetitions identified. **(B)** Gene ontology (GO) biological process (BP), cellular component (CC), and molecular function (MF) enrichment analysis.

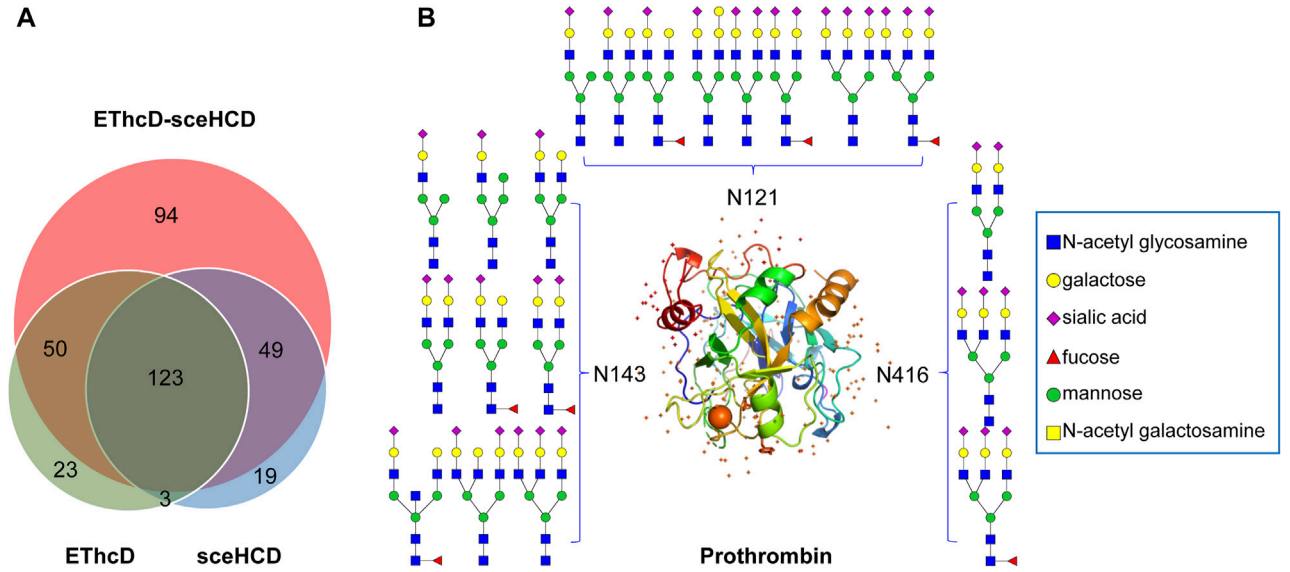


FIGURE 5 | Global analysis of identified plasma intact N-glycopeptides. **(A)** Comparison of identified intact N-glycopeptides using EThcD, secHCD, and EThcD-sceHCD. **(B)** N-glycosites (N121, N143, and N416) and deduced N-glycans were demonstrated in the three-dimensional structure of the prothrombin (PDB code: 1A2C) from pooled PCa patients' plasma.

support this conclusion (Zeng et al., 2022). However, sceHCD outperformed the EThcD-sceHCD method for N-glycopeptide identifications from the urine of IgAN patients, HepG2 cells, and thyroid cancer tissues. And the interesting thing is that EThcD-sceHCD outperformed the sceHCD method for N-glycopeptide identification from plasma treated with/without CPLL. All of these revealed that EThcD-sceHCD can outperform EThcD and sceHCD in the number of identified intact N-glycopeptides and accuracy of plasma intact N-glycopeptide identification.

Meanwhile, we analyzed the identified plasma N-glycoproteins (**Figure 4**). EThcD-sceHCD identified most N-glycoproteins (48/56), and 77.1% N-glycoproteins can be identified at least twice. This result is obviously better than that obtained using EThcD (26/56) and sceHCD (29/56). Integrating sceHCD and EThcD-sceHCD, 96.4% (54/56) N-glycoproteins can be identified (**Figure 4A**). Moreover, using the Wukong platform, gene ontology (GO) enrichment analysis was performed to determine the biological process (BP), cellular component (CC), and molecular function (MF) of these plasma glycoproteins in PCa patients. As shown in **Figure 4B**, they were involved in many biological processes, such as complement activation and immune response, mainly localized in the extracellular space. The major molecular functions were catalytic activities and binding (**Figure 4B**). These results suggested that the function of N-glycoproteins in PCa patients' circulatory system may be dysfunctional (Zhang et al., 2020b).

Based on the comprehensive PCa patients' plasma N-glycoproteomic information obtained in this work, we further compared these intact N-glycopeptides. EThcD-sceHCD can identify an additional 94 intact N-glycopeptides, while EThcD and sceHCD can identify only 23 and 19 intact N-glycopeptides, respectively (**Figure 5A**). Hence, using different fragmentation modes to analyze the same sample would produce complementary results. Based on these results, a PCa patient plasma N-glycoprotein database was established (**Supplementary Table S1**). For example, we found that prothrombin in PCa patients' plasma is a completely sialylated glycoprotein, which plays a key role in blood homeostasis, wound healing, and inflammation. All of its reported N-glycosites (N121, N143, and N416) were identified in this work. The three N-glycosites were occupied by different amounts and types of N-glycans (**Figure 5B**). These results implied that our methods can decipher the site-specific glycosylation of human plasma glycoproteins. It is worth noting that we did not include healthy controls in this study because we were evaluating the usability and superiority of the EThcD-sceHCD technique in this work. We will apply this technique to a large cohort of clinical samples in our future research.

4 CONCLUSION

Herein, we once again proved the reliability and superiority of EThcD-sceHCD. By integrating CPLL into EThcD-sceHCD, we systematically compared the performance of different fragment methods in human plasma intact N-glycopeptide analysis. EThcD-sceHCD performed better in the accuracy and depth of intact N-glycopeptide identification. This finding would drive clinical plasma N-glycoproteomic methodological development and promote related application research.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and

accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by West China Hospital, Sichuan University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YOZ and XD designed the research; YM and YAZ conducted the data analysis; TS, TL, and HY coordinated the acquisition of clinical samples; YOZ and XD wrote the manuscript. All authors have approved the final version of the manuscript.

FUNDING

We are grateful for financial support from the National Natural Science Foundation of China (31901038) and the Department of Science and Technology of Sichuan Province (2021YJ0479).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2022.920009/full#supplementary-material

REFERENCES

Bandow, J. E. (2010). Comparison of Protein Enrichment Strategies for Proteome Analysis of Plasma. *Proteomics* 10 (7), 1416–1425. doi:10.1002/pmic.200900431

Bern, M., Kil, Y. J., and Becker, C. (2012). Byonic: Advanced Peptide and Protein Identification Software. *Curr. Protoc. Bioinforma.* 13, 20. doi:10.1002/0471250953.bi1320s40

Chen, Y., Sheng, Q., Hong, Y., and Lan, M. (2019). Hydrophilic Nanocomposite Functionalized by Carrageenan for the Specific Enrichment of Glycopeptides. *Anal. Chem.* 91 (6), 4047–4054. doi:10.1021/acs.analchem.8b05578

Di Palma, S., Boersema, P. J., Heck, A. J. R., and Mohammed, S. (2011). Zwitterionic Hydrophilic Interaction Liquid Chromatography (ZIC-HILIC and ZIC-cHILIC) Provide High Resolution Separation and Increase Sensitivity in Proteome Analysis. *Anal. Chem.* 83 (9), 3440–3447. doi:10.1021/ac103312e

Fang, Z., Qin, H., Mao, J., Wang, Z., Zhang, N., Wang, Y., et al. (2022). GlycoDecipher Enables Glycan Database-independent Peptide Matching and In-Depth Characterization of Site-specific N-Glycosylation. *J. Nat. Commun.* 13 (1), 1900. doi:10.1038/s41467-022-29530-y

Jacobs, J. M., Adkins, J. N., Qian, W.-J., Liu, T., Shen, Y., Camp, D. G., 2nd, et al. (2005). Utilizing Human Blood Plasma for Proteomic Biomarker Discovery. *J. Proteome Res.* 4 (4), 1073–1085. doi:10.1021/pr0500657

Kawahara, R., Chernykh, A., Alagesan, K., Bern, M., Cao, W., Chalkley, R. J., et al. (2021). Community Evaluation of Glycoproteomics Informatics Solutions Reveals High-Performance Search Strategies for Serum Glycopeptide Analysis. *Nat. Methods* 18 (11), 1304–1316. doi:10.1038/s41592-021-01309-x

Klein, J. A., and Zaia, J. (2020). A Perspective on the Confident Comparison of Glycoprotein Site-specific Glycosylation in Sample Cohorts. *Biochemistry* 59 (34), 3089–3097. doi:10.1021/acs.biochem.9b00730

Liu, M.-Q., Zeng, W.-F., Fang, P., Cao, W.-Q., Liu, C., Yan, G.-Q., et al. (2017). pGlyco 2.0 Enables Precision N-Glycoproteomics with Comprehensive Quality Control and One-step Mass Spectrometry for Intact Glycopeptide Identification. *Nat. Commun.* 8 (1), 438. doi:10.1038/s41467-017-00535-2

Lu, L., Riley, N. M., Shortreed, M. R., Bertozzi, C. R., and Smith, L. M. (2020). O-pair Search with MetaMorpheus for O-Glycopeptide Characterization. *Nat. Methods* 17 (11), 1133–1138. doi:10.1038/s41592-020-00985-5

Luque-Garcia, J. L., and Neubert, T. A. (2007). Sample Preparation for Serum/plasma Profiling and Biomarker Identification by Mass Spectrometry. *J. Chromatogr. A* 1153 (1-2), 259–276. doi:10.1016/j.chroma.2006.11.054

Ma, C., Qu, J., Li, X., Zhao, X., Li, L., Xiao, C., et al. (2016). Improvement of Core-Fucosylated Glycoproteome Coverage via Alternating HCD and ETD Fragmentation. *J. Proteomics* 146, 90–98. doi:10.1016/j.jprot.2016.06.003

Moggridge, S., Fulton, K. M., and Twine, S. M. (2019). Enriching for Low-Abundance Serum Proteins Using ProteoMiner and Protein-Level HPLC. *Methods Mol. Biol.* 2024, 103–117. doi:10.1007/978-1-4939-9597-4_6

Moremen, K. W., Tiemeyer, M., and Nairn, A. V. (2012). Vertebrate Protein Glycosylation: Diversity, Synthesis and Function. *Nat. Rev. Mol. Cell Biol.* 13 (7), 448–462. doi:10.1038/nrm3383

Muthusamy, B., Hanumanthu, G., Suresh, S., Rekha, B., Srinivas, D., Karthick, L., et al. (2005). Plasma Proteome Database as a Resource for Proteomics Research. *Proteomics* 5 (13), 3531–3536. doi:10.1002/pmic.200401335

Nanjappa, V., Thomas, J. K., Marimuthu, A., Muthusamy, B., Radhakrishnan, A., Sharma, R., et al. (2014). Plasma Proteome Database as a Resource for Proteomics Research: 2014 Update. *Nucl. Acids Res.* 42, D959–D965. doi:10.1093/nar/gkt1251

Ohtsubo, K., and Marth, J. D. (2006). Glycosylation in Cellular Mechanisms of Health and Disease. *Cell* 126 (5), 855–867. doi:10.1016/j.cell.2006.08.019

Palmisano, G., Lendal, S. E., Engholm-Keller, K., Leth-Larsen, R., Parker, B. L., and Larsen, M. R. (2010). Selective Enrichment of Sialic Acid-Containing Glycopeptides Using Titanium Dioxide Chromatography with Analysis by HILIC and Mass Spectrometry. *Nat. Protoc.* 5 (12), 1974–1982. doi:10.1038/nprot.2010.167

Palstrøm, N. B., Rasmussen, L. M., and Beck, H. C. (2020). Affinity Capture Enrichment versus Affinity Depletion: A Comparison of Strategies for Increasing Coverage of Low-Abundant Human Plasma Proteins. *Int. J. Mol. Sci.* 21 (16), 5903. doi:10.3390/ijms21165903

Polasky, D. A., Yu, F., Teo, G. C., and Nesvizhskii, A. I. (2020). Fast and Comprehensive N- and O-Glycoproteomics Analysis with MSFragger-Glyco. *Nat. Methods* 17 (11), 1125–1132. doi:10.1038/s41592-020-0967-9

Pujic, I., and Perreault, H. (2021). Recent Advancements in Glycoproteomic Studies: Glycopeptide Enrichment and Derivatization, Characterization of Glycosylation in SARS CoV2, and Interacting Glycoproteins. *Mass Spectrom. Rev.* 41, 21679. doi:10.1002/mas.21679

Riley, N. M., Malaker, S. A., Driessen, M. D., and Bertozzi, C. R. (2020). Optimal Dissociation Methods Differ for N- and O-Glycopeptides. *J. Proteome Res.* 19 (8), 3286–3301. doi:10.1021/acs.jproteome.0c00218

Saba, J., Dutta, S., Hemenway, E., and Viner, R. (2012). Increasing the Productivity of Glycopeptides Analysis by Using Higher-Energy Collision Dissociation-Accurate Mass-product-dependent Electron Transfer Dissociation. *Int. J. Proteomics* 2012, 560391. doi:10.1155/2012/560391

Sennels, L., Salek, M., Lomas, L., Boschetti, E., Righetti, P. G., and Rappsilber, J. (2007). Proteomic Analysis of Human Blood Serum Using Peptide Library Beads. *J. Proteome Res.* 6 (10), 4055–4062. doi:10.1021/pr070339l

Shen, J., Jia, L., Dang, L., Su, Y., Zhang, J., Xu, Y., et al. (2021). StrucGP: De Novo Structural Sequencing of Site-specific N-Glycan on Glycoproteins Using a Modularization Strategy. *Nat. Methods* 18 (8), 921–929. doi:10.1038/s41592-021-01209-0

Silsirivanit, A. (2019). Glycosylation Markers in Cancer. *Adv. Clin. Chem.* 89, 189–213. doi:10.1016/bs.acc.2018.12.005

Sun, N., Wang, J., Yao, J., and Deng, C. (2017). Hydrophilic Mesoporous Silica Materials for Highly Specific Enrichment of N-Linked Glycopeptide. *Anal. Chem.* 89 (3), 1764–1771. doi:10.1021/acs.analchem.6b04054

Sun, N., Wu, H., Chen, H., Shen, X., and Deng, C. (2019). Advances in Hydrophilic Nanomaterials for Glycoproteomics. *Chem. Commun.* 55 (70), 10359–10375. doi:10.1039/c9cc04124a

Wu, Q., Jiang, B., Weng, Y., Liu, J., Li, S., Hu, Y., et al. (2018). 3-Carboxybenzoboroxole Functionalized Polyethylenimine Modified Magnetic Graphene Oxide Nanocomposites for Human Plasma Glycoproteins Enrichment under Physiological Conditions. *Anal. Chem.* 90 (4), 2671–2677. doi:10.1021/acs.analchem.7b04451

Yang, W., Shah, P., Hu, Y., Toghi Eshghi, S., Sun, S., Liu, Y., et al. (2017). Comparison of Enrichment Methods for Intact N- and O-Linked Glycopeptides Using Strong Anion Exchange and Hydrophilic Interaction Liquid Chromatography. *Anal. Chem.* 89 (21), 11193–11197. doi:10.1021/acs.analchem.7b03641

Yu, Q., Wang, B., Chen, Z., Urabe, G., Glover, M. S., Shi, X., et al. (2017). Electron-Transfer/Higher-Energy Collision Dissociation (EThcD)-Enabled Intact Glycopeptide/Glycoproteome Characterization. *J. Am. Soc. Mass Spectrom.* 28 (9), 1751–1764. doi:10.1007/s13361-017-1701-4

Zeng, W., Zheng, S., Su, T., Cheng, J., Mao, Y., Zhong, Y., et al. (2022). Comparative N-Glycoproteomics Analysis of Clinical Samples via Different Mass Spectrometry Dissociation Methods. *Front. Chem.* 10, 839470. doi:10.3389/fchem.2022.839470

Zhang, M., Liu, Y., Zhang, D., Chen, T., and Li, Z. (2017). Facile and Selective Enrichment of Intact Sialoglycopeptides Using Graphitic Carbon Nitride. *Anal. Chem.* 89 (15), 8064–8069. doi:10.1021/acs.analchem.7b01556

Zhang, Y., Jing, H., Meng, B., Qian, X., and Ying, W. (2020a). L-cysteine Functionalized Straticulate C_3N_4 for the Selective Enrichment of Glycopeptides. *J. Chromatogr. A* 1610, 460545. doi:10.1016/j.chroma.2019.460545

Zhang, Y., Jing, H., Wen, T., Wang, Y., Zhao, Y., Wang, X., et al. (2019). Phenylboronic Acid Functionalized C_3N_4 Facultative Hydrophilic Materials for Enhanced Enrichment of Glycopeptides. *Talanta* 191, 509–518. doi:10.1016/j.talanta.2018.09.016

Zhang, Y., Mao, Y., Zhao, W., Su, T., Zhong, Y., Fu, L., et al. (2020b). Glyco-CPLL: An Integrated Method for In-Depth and Comprehensive N-Glycoproteome Profiling of Human Plasma. *J. Proteome Res.* 19 (2), 655–666. doi:10.1021/acs.jproteome.9b00557

Zhang, Y., Xie, X., Zhao, X., Tian, F., Lv, J., Ying, W., et al. (2018). Systems Analysis of Singly and Multiply O -glycosylated Peptides in the Human Serum Glycoproteome via EThcD and HCD Mass Spectrometry. *J. Proteomics* 170, 14–27. doi:10.1016/j.jprot.2017.09.014

Zhang, Y., Zheng, S., Mao, Y., Cao, W., Zhao, L., Wu, C., et al. (2021). Systems Analysis of Plasma IgG Intact N-Glycopeptides from Patients with Chronic Kidney Diseases via EThcD-sceHCD-MS/MS. *Analyst* 146 (23), 7274–7283. doi:10.1039/d1an01657a

Zhang, Y., Zheng, S., Zhao, W., Mao, Y., Cao, W., Zeng, W., et al. (2021). Sequential Analysis of the N/O-Glycosylation of Heavily Glycosylated HIV-1 Gp120 Using EThcD-sceHCD-MS/MS. *Front. Immunol.* 12, 755568. doi:10.3389/fimmu.2021.755568

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Chemistry

TYPE Mini Review
PUBLISHED 09 September 2022
DOI 10.3389/fchem.2022.926353

OPEN ACCESS

EDITED BY
Marta De Zotti,
University of Padua, Italy

REVIEWED BY
Xuefei Huang,
Michigan State University, United States

*CORRESPONDENCE
Yujing Zhang,
zhangyujing@hunnu.edu.cn
Xing Yu,
xingyu@hunnu.edu.cn

SPECIALTY SECTION
This article was submitted to Chemical Biology,
a section of the journal
Frontiers in Chemistry

RECEIVED 22 April 2022
ACCEPTED 22 August 2022
PUBLISHED 09 September 2022

CITATION
Yuan F, Yang Y, Zhou H, Quan J, Liu C, Wang Y, Zhang Y and Yu X (2022), Heparanase in cancer progression: Structure, substrate recognition and therapeutic potential.
Front. Chem. 10:926353.
doi: 10.3389/fchem.2022.926353

Heparanase in cancer progression: Structure, substrate recognition and therapeutic potential

Fengyan Yuan, Yiyuan Yang, Huiqin Zhou, Jing Quan, Chongyang Liu, Yi Wang, Yujing Zhang* and Xing Yu*

Key Laboratory of Model Animals and Stem Cell Biology of Hunan Province, School of Medicine, Hunan Normal University, Changsha, China

Heparanase, a member of the carbohydrate-active enzyme (CAZy) GH79 family, is an endo-β-glucuronidase capable of degrading the carbohydrate moiety of heparan sulphate proteoglycans, thus modulating and facilitating remodeling of the extracellular matrix. Heparanase activity is strongly associated with major human pathological complications, including but not limited to tumour progress, angiogenesis and inflammation, which make heparanase a valuable therapeutic target. Long-due crystallographic structures of human and bacterial heparanases have been recently determined. Though the overall architecture of human heparanase is generally comparable to that of bacterial glucuronidases, remarkable differences exist in their substrate recognition mode. Better understanding of regulatory mechanisms of heparanase in substrate recognition would provide novel insight into the anti-heparanase inhibitor development as well as potential clinical applications.

KEYWORDS

glycosaminoglycan (GAG), heparanase, structure, substrate recognition, cancer

Introduction

As a key component of the extracellular matrix (ECM), heparan sulfate proteoglycans (HSPGs) comprise of a transmembrane or secreted protein core to which one or more heparan sulfate (HS) chains are covalently attached (Iozzo, 2005; Iozzo and Schaefer, 2015). HSPGs are one of the most highly negatively charged biopolymers occurred naturally, collaborating with other ECM components to orchestrate the ECM remodeling and structural integrity (Belting, 2003; Lindahl and Kjellen, 2013). Significantly, HS chains of HSPGs act as a storage depot, providing binding sites for a wide variety of bioactive molecules, such as growth factors, chemokines, lipoproteins and enzymes, which enables HSPGs to play essential roles in regulation of numerous physiological and pathological activities (Varki et al., 2009; Iozzo and Schaefer, 2015).

Heparanase (HPSE; HPSE-1), a member of the glycoside hydrolase (GH) 79 family, has been defined as the only known endo-β-D-glucuronidase that catalyzes HS hydrolysis to date (Parish et al., 2001). Since the cloning and expression of HPSE in 1999, emerging

evidence highlighted the involvement of HPSE in cancer progression, inflammation and angiogenesis (Fairbanks et al., 1999; Hulett et al., 1999; Kussie et al., 1999; Toyoshima and Nakajima, 1999; Vlodavsky et al., 1999). Of interest is that HPSE expression is elevated virtually in all major types of cancers, and this up-regulation is positively correlated with metastatic potential of tumor and poor prognosis, which makes HPSE a valuable therapeutic target. It has to be noted that another HPSE isoform, HPSE-2 that lacks enzymatic activity, was reported in 2000 (McKenzie et al., 2000). HPSE-2 appears not only to be able to inhibit HPSE activity but also regulate a multitude of signaling pathways that mediate cell differentiation, apoptosis and tumor vascularity, leading to tumor suppression.

Interpretation of the substrate specificity of HPSE has been complicated, partly if not all, by the nature of HS structural heterogeneity, which is derived from the extent of the sulfation, deacetylation and epimerisation in HS biosynthesis (Ringvall et al., 2000; Li et al., 2003; Pallerla et al., 2008). The structural features of HPSE also critically contribute to the plasticity in its substrate specificity, which is central to the proper biological function of HPSE. There are three members with available crystallographic structures to date in the GH79 family: Acidobacterium capsulatum β-glucuronidase (AcaGH79) (Michikawa et al., 2012), Burkholderia pseudomallei HPSE (BpHPSE) (Bohlmann et al., 2015) and human HPSE (hHPSE) (Wu et al., 2015) as well as its pro-form HPSE (hproHPSE) (Wu et al., 2017). The structure of the *exo*-acting AcaGH79 was characterized first, followed by the recent structural determination of *endo*-acting bpHPSE and hHPSE. Though hHPSE represents an overall similar folding to that of two bacterial GH79 members, structural variations within the substrate binding canyon fine-tune the distinct substrate specificities. Compelling evidence suggest that HPSE is a multifaceted protein participating in multiple biological processes, some excellent reviews are available pertaining to the engagement of HPSE in cancer progression, inflammation and angiogenesis (Fux et al., 2009b; Vlodavsky et al., 2012; Peterson and Liu, 2013; Pisano et al., 2014; Rivara et al., 2016; Masola et al., 2018; Mohan et al., 2019). In this minireview, we firstly provide an insight into the structure-based rationale of HPSE substrate recognition. Next, we briefly review the pro-tumorigenic effects of HPSE, which may highlight its therapeutic potential against cancer.

Heparan sulfate proteoglycan

HSPGs consist of variable HS chains that covalently attach to core proteins depending on the context of source and growing conditions (Karamanos et al., 2018). HSPGs not only are present as crucial components of the ECM and basement membrane (Sertie et al., 2000; Arikawa-Hirasawa et al., 2001; Campos-Xavier et al., 2009), also are found in secreted vesicles regulating various biological activities after secretion (Zernichow et al., 2006), including membrane-bound syndecans, glypicans, betaglycan, neuropinlin and CD44v3, ECM components perlecan, agrin and collagen XVIII, and secreted serglycin. After the attachment of xylose to specific serine residues in core proteins of HSPGs, HS biosynthesis is commenced by synthesizing a linkage tetrasaccharide, glucuronic acid (GlcA)-galactose-galactose-xylose.

Structurally, HS is a glycosaminoglycan (GAG) chain with potential modifications of sulfation, epimerization and deacetylation, comprising of a linear repeating disaccharide unit constituted by acetylated hexosamines (N-acetyl-glucosamine, GlcNAc or N-sulfo-glucosamine, GlcNS) and uronic acids (GlcA or its C5 epimer L-iduronic acid, IdoA) (Kjellen and Lindahl, 1991). Further O-sulfation can take place at O2 of the uronic acid (2-O sulfation) and O3 and O6 of the hexosamine (3-O and 6-O sulfation). Numerous combinations of the low sulfation and high sulfation domains along HS chains as well as the specific sulfation pattern within each domain complicate the recognition of HSPGs by HPSE. Early studies suggest that the minimum recognition backbone of HSPGs by HPSE is a trisaccharide, and the cleavage occurs at the internal β(1,4)-linked glycosidic bond between GlcA and GlcNS (Figure 1A) (Matsuno et al., 2002; Peterson and Liu, 2013). Further investigation revealed that HPSE cleavage of HSPG is dependent on sulfation types rather than a defined saccharide sequence, and the cleavage by HPSE is regulated by specific sulfation contexts around the cleavage site (Peterson and Liu, 2010).

The number of attached HS chains, together with the sulfation distribution along the HS chains, leads to high structural heterogeneity of HSPGs. In addition, modifications of HS occurred during its biosynthesis appear to be non-template, context-specific and in response to stimuli, therefore resulting in remarkable variations in HS chains (Armistead et al., 2011; Sarrazin et al., 2011; Shi et al., 2011). Of relevance is that the structural heterogeneity of HS facilitates its capability of accommodating a variety of binding partners, which is essential to the diverse biological roles of HSPGs upon HPSE breakdown, leading to activation of downstream signal cascades and promotion of cell proliferation, tumor cell dissemination, inflammation and angiogenesis (Bernfield et al., 1999; Elkin et al., 2001; Iozzo and San Antonio, 2001; Sasaki et al., 2004; Barash et al., 2010b; Goodall et al., 2014). Previous studies have demonstrated that hHPSE is able to act in either a consecutive or a gapped cleavage mode depending on the saccharide sequences released from its initial cleavage (Peterson and Liu, 2013), which allows the efficient release of distinct bioactive molecules from regions with different sulfation patterns along HS chains.

HPSE

Overview

The gene coding for HPSE consisting of 14 exons and 13 introns is located on chromosome 4q21.3 and expressed as two mRNAs (5 and 1.7 kb) by alternative splicing containing the

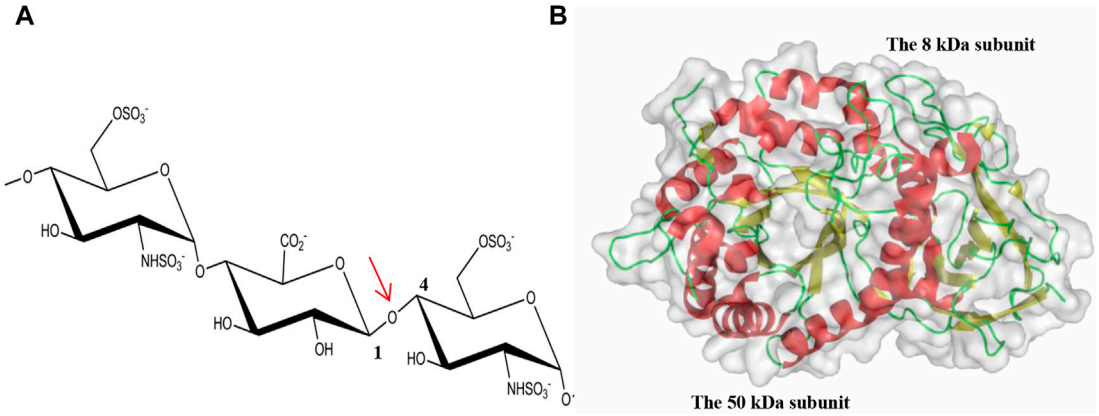


FIGURE 1
Cleavage of the glycosidic bond by HPSE. **(A)** The internal β(1,4)-linked glycosidic bond between GlcA and GlcNS is highlighted by an arrow in red; **(B)** The overall fold of hHPSE is illustrated in ribbon representation with α-helix in red, β-strand in yellow and loop in green.

same open reading frame (Dong et al., 2000). HPSE is initially synthesized as a preproenzyme of 68 kDa containing a signal sequence spanning Met1–Ala35, which is then processed into a proHPSE form after cleaving the signal sequence by signal peptidase. Lysosomal activation by cathepsin L excises a linker domain of Ser110–Gln157, giving rise to the mature HPSE as a non-covalent heterodimer containing an N-terminal 8 kDa (Gln36-Glu109) and a C-terminal 50 kDa (Lys158-Ile543) subunits.

β-glucuronidases are categorized into three GH families, GH1, GH2, and GH79, on the basis of their amino acid sequences (Henrissat and Davies, 1997; Cantarel et al., 2009). There are four characterized β-glucuronidase members in the GH79 family, including heparanase (EC 3.2.1.166), baicalin-β-D-glucuronidase (EC 3.2.1.167), 4-O-methyl-β-glucuronidase, and β-glucuronidase (Sasaki et al., 2000; Parish et al., 2001; Eudes et al., 2008; Konishi et al., 2008). Though both GH2 and GH79 belong to the GH-A clan, the GH79 family is composed of enzymes of both *endo*-acting HPSE and *exo*-acting β-glucuronidase, which contrasts that the GH2 family only consists of *exo*-acting β-glucuronidase. Folding prediction as well as multiple sequence alignment has predicted HPSE being a member of GH-A clan, proposing An $(\beta/\alpha)_8$-TIM barrel as the key folding feature of HPSE (Nardella et al., 2004). This was confirmed after the recent determination of the long-anticipated hHPSE and hproHPSE structures.

HPSE structure and substrate recognition

The structure of apo hHPSE consists of a heterodimer formed by the 8-kDa subunit (residues Gln36–Glu109) and the 50-kDa subunit (Lys159–Ile543) (Wu et al., 2015), with the domain architecture comprising a catalytic (β/α)8-TIM barrel domain flanked by a β-sandwich domain (Figure 1B). Both the 8-kDa subunit and the 50-kDa subunit contribute to the formation of the catalytic $(\beta/\alpha)_8$-TIM barrel and the β-sandwich domain. Though the β-sandwich domain was reported to facilitate secretion and activation, cellular trafficking, enzymatic and nonenzymatic activities of HPSE, its function demands to be further characterized (Simizu et al., 2007; Lai et al., 2008; Fux et al., 2009a). In addition, there are six putative N-glycosylation sites identified in the 50-kDa subunit of hHPSE. After the deglycosylation treatment of Endo-H during the protein preparation, N-linked GlcNAc residues were visible in the apo hHPSE structure at Asn162, Asn200, Asn217, Asn238 and Asn459, respectively (Wu et al., 2015). Intriguingly, glycosylation regulates HPSE secretion and endoplasmic reticulum-to-Golgi transport, but it is not required for enzymatic activity of HPSE (Simizu et al., 2004).

A binding groove of approximately 10 Å in the catalytic $(\beta/\alpha)_8$-TIM barrel domain was recognized in the hHPSE structure. This binding groove contains residues Glu343 and Glu225, which are conserved in the GH79 family and have been previously identified as the catalytic nucleophile and acid-base pair of HPSE, suggesting that the HS-binding site is contained within this groove (Hulett et al., 2000). As shown in Figure 2A, the HPSE binding canyon is lined by side chains of basic residues, which correlates well with the negatively charged nature of HS substrates. Of interest is that the orientation of two subunits in the solved hHPSE structure implicates that the excised Ser110–Gln157 linker of the proHPSE could locate very close in space to the HS binding groove, which would physically clash the HS substrate. This is consistent with the reported hproHPSE structure in 2017, showing the restricted access to the active site cleft for oligosaccharide HS substrates due to the presence of the 6-kDa linker loop (Figure 2B) (Wu et al., 2017).

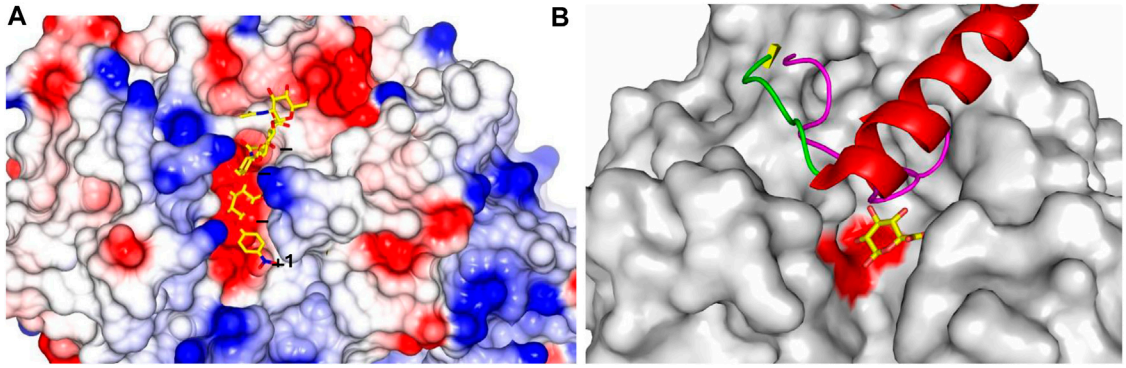


FIGURE 2
HPSE substrate binding site and structural superimposition. **(A)** hHPSE (electrostatic surface) in complex with a bound tetrasaccharide (our unpublished work) in stick presentation (carbon in yellow, nitrogen in blue and oxygen in red) spanning through binding subsites +1, −1, −2, −3 (Davies et al., 1997); **(B)** BpHPSE is illustrated in surface representation with two conserved catalytic glutamate highlighted in red; the loop that forms part of the substrate-binding pocket in AcaGH79 is colored in magenta; the 6-kDa linker of hproHPSE is illustrated in ribbon representation with α-helix in red, β-strand in yellow and loop in green; the bound GlcA are colored with carbon in yellow and oxygen in red.

Structures of hHPSE in complex with HS analogues provides a structural rationale, clearly demonstrating that hHPSE recognizes a trisaccharide spanning the −2, −1 and +1 subsites, with the identical binding of GlcA at the −1 subsite in all bound HS analogues (Figure 2A). This conserved binding of GlcA is also observed in the GH79 bacterial members, suggesting a key GH79 structural motif that has been fine tuned to recognize GlcA (Michikawa et al., 2012; Wu et al., 2015). Significantly, N-sulfate at the −2 subsite and 6O-sulfate at the +1 subsite appear to be the main determinants for recognition of HS analogues due to their direct engagement in the hydrogen-bonding interactions with HPSE. 6O-sulfate at the −2 subsite and N-sulfate at the +1 subsite also contribute to the anchorage of HS substrates through electrostatic interactions to basic residues lining the active site cleft. Overall, structural information gained from the complexes of hHPSE with its HS substrate analogs is consistent with the findings of previous studies that HS sulfation patterns are essential for hHPSE enzymatic activity. Furthermore, sulfation contexts of HS substrates appear to act as a molecular signal that guides the precise cleavage of designated glycan sites. In addition, sulfate groups on the −2 and +1 moieties are implicated to aid hHPSE to unwind the substrate HS helix for a better access of the catalytic residues to facilitate the cleavage of the glycosidic bond.

Structure-based rationale for the *exo*- and *endo*-acting modes of GH79 β-glucuronidases

When compared in the sequence alignment, a loop of 40 amino acids (Gly78–Thr117), which forms part of the *exo*-acting substrate-binding pocket identified in the AcaGH79 structure, corresponds to its counterpart that is substantially reduced in size of *endo*-acting BpHPSE (24 amino acids, Gly67–Pro90). Structurally, this shorter loop of BpHPSE allows the transition of the binding pocket into an open-end binding groove capable of accommodating elongated HS chains, which hereby provides a well-explained structure-based rationale of the discrepancy in acting modes of the substrate cleavage between the *exo*-AcaGH79 and *endo*-BpHPSE enzymes (Figure 2B).

The overall folding of hHPSE is comparable to that of two characterized bacterial GH79 members (Michikawa et al., 2012; Bohlmann et al., 2015), with Cα r.m.s. differences of 2.35 Å and 2.59 Å for AcaGH79 and BpHPSE, respectively. In particular, the 6 kDa linker peptide of hproHPSE that is proteolytically cleaved to enable the activation of hHPSE also corresponds to the AcaGH79 loop. Intriguingly, structural observations revealed that the physical presence of the hproHPSE linker peptide create a binding pocket on the protein surface containing those two highly conserved glutamate, resembling some structural characteristics of the *exo*-acting active site of AcGH79 (Wu et al., 2017). Detailed data indicate that this hproHPSE pocket is not involved in the HS interactions and GlcA occupation of the proHPSE pocket does not inhibit HPSE maturation, suggesting extra subsite interactions may be required for anchoring single GlcA molecules to proHPSE. Further investigations are thus required to determine whether this proHPSE pocket possesses any substrate specificity.

Whereas both hHPSE and BpHPSE demonstrate the *endo*-acting mode of substrate cleavage, differences exist between their substrates recognition. BpHPSE has preference for cleaving HS-containing GlcNAc residues (low sulfation), contrasting that GlcNS is preferably recognized by hHPSE. Moreover, sequence alignment of several eukaryotic HPSEs with BpHPSE

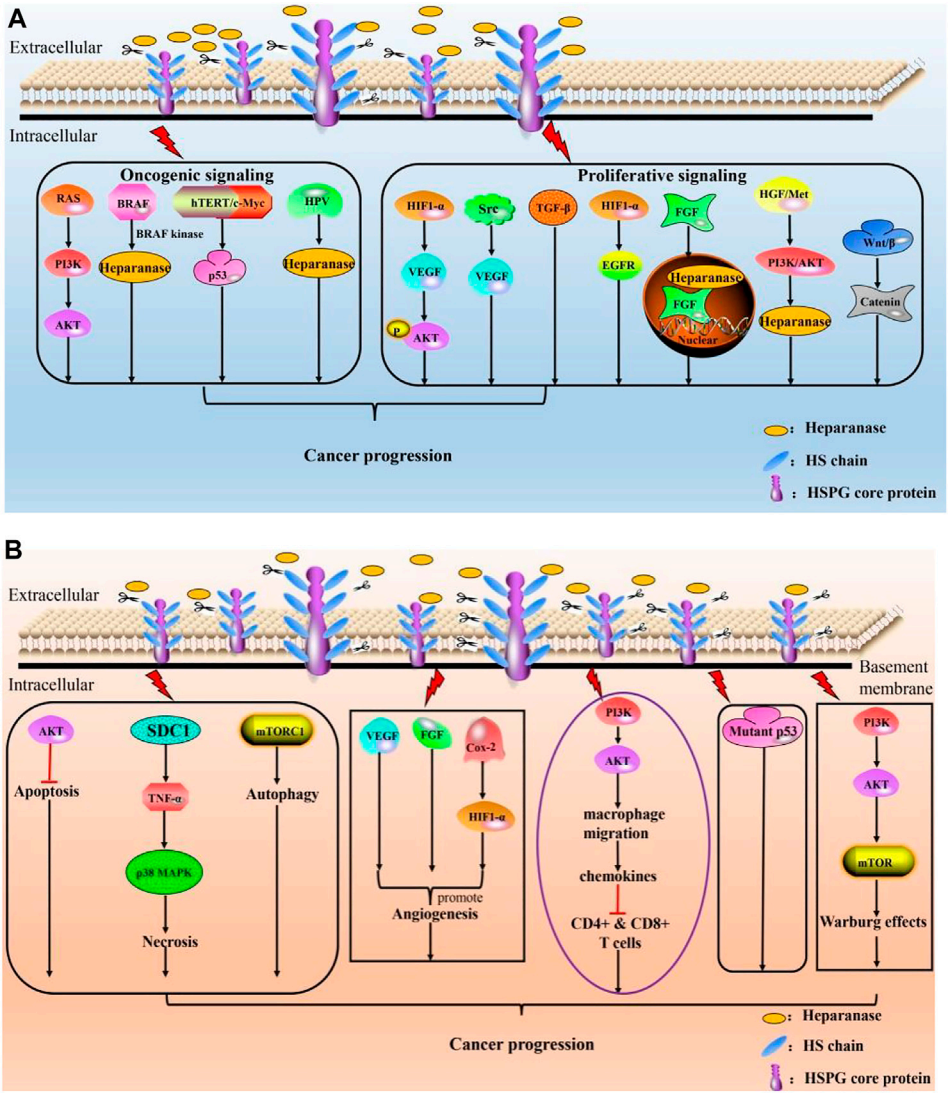


FIGURE 3
Roles of HPSE in cancer progression. **(A)** HPSE modulates cancer progression by mediating oncogenic signaling and proliferative signaling; **(B)** HPSE promotes cancer by resisting cell death, initiating angiogenesis, contributing to anti-immunity failure, circumventing growth inhibition and reprogramming energy metabolism.

and AcaGH79 revealed a remarkable conservation of key residues for accommodating GlcA at the −1 subsite of hHPSE, demonstrated by absolute conservation of residue Asn224, Glu225, Glu343, Gly350 and Tyr391 and evolutionary conservation of residue 62 (Asp of hHPSE vs Glu of BpHPSE and AcaGH79), 97 (Thr of hHPSE vs. Asn of BpHPSE and AcaGH79), 349 (Gly of hHPSE and BpHPSE vs Gln of AcaGH79), whilst residues at the −2 and +1 subsites show much poorer conservation in the BpHPSE and AcaGH79, thus providing a structure-based rationale for distinct substrate specificity amongst those GH79 enzymes (Wu et al., 2015).

Transport and function of HPSE in nucleus and extracellular

It is generally known that mature HPSE is the only known endo-β-D-glucuronidase in mammals, which can cleave the HS chain of HSPG to release growth factors, chemokines, lipoproteins and enzymes, play a role in promoting tumors outside the cell. Many studies have shown that the function of HPSE is regulated by histones. Histones mainly exist in the nucleus and extracellular regions and secreted proteins. Histones (H1, H2A, H2B, H3, and H4) contain a large number of basic R-based amino acids, which are positively

charged in aqueous solutions. While HSPG as one of the most negatively charged biopolymers, HSPG and histones bind to the GAG chain of HSPG through charge interaction to play a regulatory role. For example, extracellular histone H4 induces HS degradation by activating HPSE in chlorine (Cl2)-induced acute respiratory distress syndrome (ARDS). Knockdown of HPSE by RNAi demonstrated that histone h4-induced HS degradation requires HPSE and is dependent on the enzymatic activity of HPSE (Zhang et al., 2022). In cells expressing high levels of HPSE, reduction of nuclear syndecan-1 results in increased histone acetyltransferase (HAT) activity, which stimulates protein transcription and transcriptional upregulation of multiple genes that drive aggressive tumor phenotypes (Purushothaman et al., 2011). In lymphangiosarcoma (SS), histone deacetylase inhibitors (HDACi) upregulate HPSE by inducing the expression of the positive regulator EGR1 and inhibit the negative regulation of p53 by acetylation. By co-treatment with MEK inhibitor (trametinib) or HPSE inhibitor (SST0001/rooneparstat), blocking HDACi-induced erk-egr1-HPSE pathway enhanced antiproliferative and proapoptotic effects (Lanzi et al., 2021).

With the deepening of research, it was found that HPSE can inhibit tumor after entering the nucleus. Human HPSE sequences contain two potential nuclear localization signals (residues 271-277; PRRKTAK and residues 427-430; KRRK), which mediate nuclear localization of enzymes. Secondly, HPSE nuclear translocation can be promoted by its heparin binding domain, using HS as its carrier (Nadav et al., 2002). Yang et al. (2015) using atomic force microscopy and co-precipitation methods, found a direct molecular interaction between HPSE and DNA driven by charge, indicating that HPSE has dual functions in malignant melanoma, with primary extracellular activity and tumor-suppressive nuclear effect. In type 1 diabetes, heparin and HS can be transported to the nucleus and directly or indirectly affect gene transcription. Based on Chip-on-chip studies, heparin interacts with promoters and transcription regions of hundreds of genes and micro-RNAs in activated Jurkat T cells and upregulates transcription at the molecular level. Nuclear HPSE appears to regulate methylation of histone 3 lysine 4 (H3K4) by influencing demethylase recruitment of transcription-active genes (Parish et al., 2013).

HPSE in cancer and its therapeutic potential

As aforementioned, quite a few excellent reviews are available pertaining to activities of HPSE in different physiological and pathological contexts, we thus briefly summarize the pleiotropic actions of HPSE herein and will not go into detailed discussion (Figure 3). Function of HPSE is strongly associated with major human pathological complications, evidenced by that various literatures have linked overexpression of HPSE to enhanced tumor growth, metastasis and poor prognosis. Further, silencing of HPSE or treatment of tumor with compounds that block HPSE activity is shown to remarkably attenuate tumor progression. Therefore, targeting HPSE is considered as a promising therapeutic strategy for cancer treatment. Several classes of inhibitors have been developed, ranging from nucleic acid-based inhibitor, vaccines, MicroRNAs, anti-HPSE monoclonal antibodies, poly-sulfated saccharides to small-molecule inhibitors (Rivara et al., 2016). Though MicroRNAs and anti-HPSE antibodies are demonstrated to have high specificity, none of those so-called biological drugs, such as vaccines, antibodies, and antisense RNAs, have ever passed the clinical trials. Further, small molecule drugs also failed to enter clinical studies.

To the best of our knowledge, only few polysaccharide-based candidates synthesized by either semi-synthetic or total synthesis methods are currently clinical tested by competitively targeting the substrate binding site of HPSE. Irrespective of distinct mechanisms of action, those polysaccharide-based inhibitors, such as PI-88, M-402, PG545, and SST0001, appear to be the most promising anti-tumor agents due to their specificity and reasonable druggability. Significantly, the development of HPSE inhibitors still exist several drawbacks, among of which are structural uncertainty, per-sulfation, *in vivo* instability, poor bioavailability and apparent side effects (Rosenthal et al., 2002; Levidiotis et al., 2004; Hossain et al., 2010). As a result, novel strategies are emerging to develop HPSE inhibitors with higher specificity and greater selectivity (Sletten et al., 2017). Intriguingly, recent findings disclose that HPSE-2, a close homolog of HPSE but lacks enzymatic activity, can regulate antitumor mechanisms. However, this theme is not the main focus of this minireview, therefore it will not be further discussed.

Conclusions and future perspective

Intensive studies have demonstrated that increased levels of HPSE expression are strongly associated with a multiplicity of hematological and solid malignancies. To this end, HPSE has become a promising target for fighting cancer. The therapeutic potential of HS mimetics, due to their ability to bind and modulate the function of HPSE, has therefore been exploited. Although several HS mimetics have advanced into clinical trials, unforeseen adverse effects are documented due to the heterogeneous nature and nonspecific or pleiotropic effects of those HS mimetics (Kudchadkar et al., 2008; Zhou et al., 2011).

Further, HPSE is a multifaceted protein having both enzymatic and non-enzymatic activities. To the best of our knowledge, all HPSE inhibitors under development are

predominately targeting on the enzymatic inhibition of HPSE. Therefore, one main question raised in the development of anti-HPSE inhibitors is whether the enzymatic activity of HPSE is the critical determinant of its pro-tumor and pro-metastasis effects, given the fact that the T5 splice variant of HPSE lacking its enzymatic activity exerts roles in promotion of tumor progress (Barash et al., 2010a; Barash et al., 2019). Intensive studies are thus required to further explore non-enzymatic activities of HPSE attributed to its physiological and pathological function.

Recent determination of crystallographic structures of human and bacterial HPSE could offer an improved understanding of mechanisms of action of HPSE at the atomic level, which will greatly aid the design of HPSE inhibitors. Given the anti-tumor action of HS mimetics appears to be context-dependent and in response to external stimuli, it is advisable to develop HS mimetics as inhibitors in a system where appropriate malignancies and patient population are rationally selected for clinical trials. In addition, HS mimetics are characterized by good safety and tolerability profiles, which make them highly suitable for inclusion in combined therapies with other drugs to enhance anti-tumor efficacy of conventional treatments.

Author contributions

XY and YZ proposed the study. FY wrote the first draft, FY and YY contributed equally to this work. All authors contributed to the interpretation of the study and revised further drafts.

Funding

This work was supported by Grants from the National Nature Science Foundation of China (81874193 and 81803272), the High Level Talents Converging Program of Hunan Province (2019RS1041), National Science of Foundation of Hunan Province (2021JJ30465 and 2020JJ5385), Education Fund Item of Hunan Province (19C1155), Key Grant of Research and Development in Hunan Province (2020DK 2002), graduate innovative program of Hunan Normal University School of Medicine (KF2021025 and KF2021037), Construction Project of Advantages and Characteristics of Colleges and Universities in Hunan Province (2022XKQ0205), and Innovative Program of Hunan Normal University School of Medicine (KF2022002 and KF2022032).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Arikawa-Hirasawa, E., Wilcox, W. R., Le, A. H., Silverman, N., Govindraj, P., Hassell, J. R., et al. (2001). Dyssegmental dysplasia, Silverman-Handmaker type, is caused by functional null mutations of the perlecan gene. *Nat. Genet.* 27 (4), 431–434. doi:10.1038/86941

Armistead, J. S., Wilson, I. B., Van Kuppevelt, T. H., and Dinglasan, R. R. (2011). A role for heparan sulfate proteoglycans in Plasmodium falciparum sporozoite invasion of anopheline mosquito salivary glands. *Biochem. J.* 438 (3), 475–483. doi:10.1042/bj20110694

Barash, U., Cohen-Kaplan, V., Arvatz, G., Gingis-Velitski, S., Levy-Adam, F., Nativ, O., et al. (2010a). A novel human heparanase splice variant, T5, endowed with protumorigenic characteristics. *FASEB J.* 24 (4), 1239–1248. doi:10.1096/fj.09-147074

Barash, U., Cohen-Kaplan, V., Dowek, I., Sanderson, R. D., Ilan, N., and Vlodavsky, I. (2010b). Proteoglycans in health and disease: New concepts for heparanase function in tumor progression and metastasis. *Febs J.* 277 (19), 3890–3903. doi:10.1111/j.1742-4658.2010.07799.x

Barash, U., Spyrou, A., Liu, P., Vlodavsky, E., Zhu, C., Luo, J., et al. (2019). Heparanase promotes glioma progression via enhancing CD24 expression. *Int. J. Cancer* 145 (6), 1596–1608. doi:10.1002/ijc.32375

Belting, M. (2003). Heparan sulfate proteoglycan as a plasma membrane carrier. *Trends biochem. Sci.* 28 (3), 145–151. doi:10.1016/s0968-0004(03)00031-8

Bernfield, M., Gotte, M., Park, P. W., Reizes, O., Fitzgerald, M. L., Lincecum, J., et al. (1999). Functions of cell surface heparan sulfate proteoglycans. *Annu. Rev. Biochem.* 68, 729–777. doi:10.1146/annurev.biochem.68.1.729

Bohlmann, L., Tredwell, G. D., Yu, X., Chang, C. W., Haselhorst, T., Winger, M., et al. (2015). Functional and structural characterization of a heparanase. *Nat. Chem. Biol.* 11 (12), 955–957. doi:10.1038/nchembio.1956

Campos-Xavier, A. B., Martinet, D., Bateman, J., Belluoccio, D., Rowley, L., Tan, T. Y., et al. (2009). Mutations in the heparan-sulfate proteoglycan glypican 6 (GPC6) impair endochondral ossification and cause recessive omodysplasia. *Am. J. Hum. Genet.* 84 (6), 760–770. doi:10.1016/j.ajhg.2009.05.002

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The carbohydrate-active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Res.* 37, D233–D238. doi:10.1093/nar/gkn663

Davies, G. J., Wilson, K. S., and Henrissat, B. (1997). Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem. J.* 321 (2), 557–559. doi:10.1042/bj3210557

Dong, J., Kukula, A. K., Toyoshima, M., and Nakajima, M. (2000). Genomic organization and chromosome localization of the newly identified human heparanase gene. *Gene* 253 (2), 171–178. doi:10.1016/s0378-1119(00)00251-1

Elkin, M., Ilan, N., Ishai-Michaeli, R., Friedmann, Y., Papo, O., Pecker, I., et al. (2001). Heparanase as mediator of angiogenesis: Mode of action. *FASEB J.* 15 (9), 1661–1663. doi:10.1096/fj.00-0895fje

Eudes, A., Mouille, G., Thevenin, J., Goyallon, A., Minic, Z., and Jouanin, L. (2008). Purification, cloning and functional characterization of an endogenous beta-glucuronidase in *Arabidopsis thaliana*. *Plant Cell Physiol.* 49 (9), 1331–1341. doi:10.1093/pcp/pcn108

Fairbanks, M. B., Mildner, A. M., Leone, J. W., Cavey, G. S., Mathews, W. R., Drong, R. F., et al. (1999). Processing of the human heparanase precursor and evidence that the active enzyme is a heterodimer. *J. Biol. Chem.* 274 (42), 29587–29590. doi:10.1074/jbc.274.42.29587

Fux, L., Feibish, N., Cohen-Kaplan, V., Gingis-Velitski, S., Feld, S., Geffen, C., et al. (2009a). Structure-function approach identifies a COOH-terminal domain that mediates heparanase signaling. *Cancer Res.* 69 (5), 1758–1767. doi:10.1158/0008-5472.can-08-1837

Fux, L., Ilan, N., Sanderson, R. D., and Vlodavsky, I. (2009b). Heparanase: Busy at the cell surface. *Trends biochem. Sci.* 34 (10), 511–519. doi:10.1016/j.tibs.2009.06.005

Goodall, K. J., Poon, I. K., Phipps, S., and Hulett, M. D. (2014). Soluble heparan sulfate fragments generated by heparanase trigger the release of pro-inflammatory cytokines through TLR-4. *PLoS One* 9 (10), e109596. doi:10.1371/journal.pone.0109596

Henrissat, B., and Davies, G. (1997). Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* 7 (5), 637–644. doi:10.1016/s0959-440x(97)80072-3

Hossain, M. M., Hosono-Fukao, T., Tang, R., Sugaya, N., Van Kuppevelt, T. H., et al. (2010). Direct detection of HSulf-1 and HSulf-2 activities on extracellular heparan sulfate and their inhibition by PI-88. *Glycobiology* 20 (2), 175–186. doi:10.1093/glycob/cwp159

Hulett, M. D., Freeman, C., Hamdorf, B. J., Baker, R. T., Harris, M. J., and Parish, C. R. (1999). Cloning of mammalian heparanase, an important enzyme in tumor invasion and metastasis. *Nat. Med.* 5 (7), 803–809. doi:10.1038/10525

Hulett, M. D., Hornby, J. R., Ohms, S. J., Zuegg, J., Freeman, C., Gready, J. E., et al. (2000). Identification of active-site residues of the pro-metastatic endoglycosidase heparanase. *Biochemistry* 39 (51), 15659–15667. doi:10.1021/bi002080p

Iozzo, R. V. (2005). Basement membrane proteoglycans: From cellar to ceiling. *Nat. Rev. Mol. Cell Biol.* 6 (8), 646–656. doi:10.1038/nrm1702

Iozzo, R. V., and San Antonio, J. D. (2001). Heparan sulfate proteoglycans: Heavy hitters in the angiogenesis arena. *J. Clin. Invest.* 108 (3), 349–355. doi:10.1172/jci200113738

Iozzo, R. V., and Schaefer, L. (2015). Proteoglycan form and function: A comprehensive nomenclature of proteoglycans. *Matrix Biol.* 42, 11–55. doi:10.1016/j.matbio.2015.02.003

Karamanos, N. K., Piperigkou, Z., Theocharis, A. D., Watanabe, H., Franchi, M., Baud, S., et al. (2018). Proteoglycan chemical diversity drives multifunctional cell regulation and therapeutics. *Chem. Rev.* 118 (18), 9152–9232. doi:10.1021/acs.chemrev.8b00354

Kjellen, L., and Lindahl, U. (1991). Proteoglycans: Structures and interactions. *Annu. Rev. Biochem.* 60, 443–475. doi:10.1146/annurev.bi.60.070191.002303

Konishi, T., Kotake, T., Soraya, D., Matsuoka, K., Koyama, T., Kaneko, S., et al. (2008). Properties of family 79 beta-glucuronidases that hydrolyze beta-glucuronosyl and 4-O-methyl-beta-glucuronosyl residues of arabinogalactan-protein. *Carbohydr. Res.* 343 (7), 1191–1201. doi:10.1016/j.carres.2008.03.004

Kudchadkar, R., Gonzalez, R., and Lewis, K. D. (2008). PI-88: A novel inhibitor of angiogenesis. *Expert Opin. Investig. Drugs* 17 (11), 1769–1776. doi:10.1517/13543784.17.11.1769

Kussie, P. H., Hulmes, J. D., Ludwig, D. L., Patel, S., Navarro, E. C., Seddon, A. P., et al. (1999). Cloning and functional expression of a human heparanase gene. *Biochem. Biophys. Res. Commun.* 261 (1), 183–187. doi:10.1006/bbrc.1999.0962

Lai, N. S., Simizu, S., Morisaki, D., Muroi, M., and Osada, H. (2008). Requirement of the conserved, hydrophobic C-terminus region for the activation of heparanase. *Exp. Cell Res.* 314 (15), 2834–2845. doi:10.1016/j.yexcr.2008.07.004

Lanzi, C., Favini, E., Dal Bo, L., Tortoreto, M., Arrighetti, N., Zaffaroni, N., et al. (2021). Upregulation of ERK-EGR1-heparanase axis by HDAC inhibitors provides targets for rational therapeutic intervention in synovial sarcoma. *J. Exp. Clin. Cancer Res.* 40 (1), 381. doi:10.1186/s13046-021-02150-y

Levidiotis, V., Freeman, C., Punler, M., Martinello, P., Creese, B., Ferro, V., et al. (2004). A synthetic heparanase inhibitor reduces proteinuria in passive Heymann nephritis. *J. Am. Soc. Nephrol.* 15 (11), 2882–2892. doi:10.1097/01.asn.0000142426.55612.6d

Li, J. P., Gong, F., Hagner-Mcwhirter, A., Forsberg, E., Abrink, M., Kisilevsky, R., et al. (2003). Targeted disruption of a murine glucuronyl C5-epimerase gene results in heparan sulfate lacking L-iduronic acid and in neonatal lethality. *J. Biol. Chem.* 278 (31), 28363–28366. doi:10.1074/jbc.c300219200

Lindahl, U., and Kjellen, L. (2013). Pathophysiology of heparan sulphate: Many diseases, few drugs. *J. Intern. Med.* 273 (6), 555–571. doi:10.1111/joim.12061

Masola, V., Bellin, G., Gambaro, G., and Onisto, M. (2018). Heparanase: A multitasking protein involved in extracellular matrix (ECM) remodeling and intracellular events. *Cells* 7 (12), 236. doi:10.3390/cells7120236

Matsuno, H., Kozawa, O., Okada, K., Ueshima, S., Matsuo, O., and Uematsu, T. (2002). Plasmin generation plays different roles in the formation and removal of arterial and venous thrombus in mice. *Thromb. Haemost.* 87 (1), 98–104. doi:10.1055/s-0037-1612950

Mckenzie, E., Tyson, K., Stamps, A., Smith, P., Turner, P., Barry, R., et al. (2000). Cloning and expression profiling of Hpa2, a novel mammalian heparanase family member. *Biochem. Biophys. Res. Commun.* 276 (3), 1170–1177. doi:10.1006/bbrc.2000.3586

Michikawa, M., Ichinose, H., Momma, M., Biely, P., Jongkees, S., Yoshida, M., et al. (2012). Structural and biochemical characterization of glycoside hydrolase family 79 beta-glucuronidase from Acidobacterium capsulatum. *J. Biol. Chem.* 287 (17), 14069–14077. doi:10.1074/jbc.m112.346288

Mohan, C. D., Hari, S., Preetham, H. D., Rangappa, S., Barash, U., Ilan, N., et al. (2019). Targeting heparanase in cancer: Inhibition by synthetic, chemically modified, and natural compounds. *iScience* 15, 360–390. doi:10.1016/j.isci.2019.04.034

Nadav, L., Eldor, A., Yacoby-Zeevi, O., Zamir, E., Pecker, I., Ilan, N., et al. (2002). Activation, processing and trafficking of extracellular heparanase by primary human fibroblasts. *J. Cell Sci.* 115 (10), 2179–2187. doi:10.1242/jcs.115.10.2179

Nardella, C., Lahm, A., Pallaoro, M., Brunetti, M., Vannini, A., and Steinkuhler, C. (2004). Mechanism of activation of human heparanase investigated by protein engineering. *Biochemistry* 43 (7), 1862–1873. doi:10.1021/bi030203a

Pallerla, S. R., Lawrence, R., Lewejohann, L., Pan, Y., Fischer, T., Schlomann, U., et al. (2008). Altered heparan sulfate structure in mice with deleted NDST3 gene function. *J. Biol. Chem.* 283 (24), 16885–16894. doi:10.1074/jbc.m709774200

Parish, C. R., Freeman, C., and Hulett, M. D. (2001). Heparanase: A key enzyme involved in cell invasion. *Biochimica Biophysica Acta - Rev. Cancer* 1471 (3), M99–M108. doi:10.1016/s0304-419x(01)00017-8

Parish, C. R., Freeman, C., Ziolkowski, A. F., He, Y. Q., Sutcliffe, E. L., Zafar, A., et al. (2013). Unexpected new roles for heparanase in Type 1 diabetes and immune gene regulation. *Matrix Biol.* 32 (5), 228–233. doi:10.1016/j.matbio.2013.02.007

Peterson, S. B., and Liu, J. (2013). Multi-faceted substrate specificity of heparanase. *Matrix Biol.* 32 (5), 223–227. doi:10.1016/j.matbio.2013.02.006

Peterson, S. B., and Liu, J. (2010). Unraveling the specificity of heparanase utilizing synthetic substrates. *J. Biol. Chem.* 285 (19), 14504–14513. doi:10.1074/jbc.m110.104166

Pisano, C., Vlodavsky, I., Ilan, N., and Zunino, F. (2014). The potential of heparanase as a therapeutic target in cancer. *Biochem. Pharmacol.* 89 (1), 12–19. doi:10.1016/j.bcp.2014.02.010

Purushothaman, A., Hurst, D. R., Pisano, C., Mizumoto, S., Sugahara, K., and Sanderson, R. D. (2011). Heparanase-mediated loss of nuclear syndecan-1 enhances histone acetyltransferase (HAT) activity to promote expression of genes that drive an aggressive tumor phenotype. *J. Biol. Chem.* 286 (35), 30377–30383. doi:10.1074/jbc.m111.254789

Ringvall, M., Ledin, J., Holmborn, K., Van Kuppevelt, T., Ellin, F., Eriksson, I., et al. (2000). Defective heparan sulfate biosynthesis and neonatal lethality in mice lacking N-deacetylase/N-sulfotransferase-1. *J. Biol. Chem.* 275 (34), 25926–25930. doi:10.1074/jbc.c000359200

Rivara, S., Milazzo, F. M., and Giannini, G. (2016). Heparanase: A rainbow pharmacological target associated to multiple pathologies including rare diseases. *Future Med. Chem.* 8 (6), 647–680. doi:10.4155/fmc-2016-0012

Rosenthal, M. A., Rischin, D., Mcarthur, G., Ribbons, K., Chong, B., Fareed, J., et al. (2002). Treatment with the novel anti-angiogenic agent PI-88 is associated with immune-mediated thrombocytopenia. *Ann. Oncol.* 13 (5), 770–776. doi:10.1093/annonc/mdf117

Sarrazin, S., Lamanna, W. C., and Esko, J. D. (2011). Heparan sulfate proteoglycans. *Cold Spring Harb. Perspect. Biol.* 3 (7), a004952. doi:10.1101/cshperspect.a004952

Sasaki, K., Taura, F., Shoyama, Y., and Morimoto, S. (2000). Molecular characterization of a novel beta-glucuronidase from Scutellaria baicalensis georgi. *J. Biol. Chem.* 275 (35), 27466–27472. doi:10.1016/s0021-9258(19)61531-0

Sasaki, N., Higashi, N., Taka, T., Nakajima, M., and Irimura, T. (2004). Cell surface localization of heparanase on macrophages regulates degradation of extracellular matrix heparan sulfate. *J. Immunol.* 172 (6), 3830–3835. doi:10.4049/jimmunol.172.6.3830

Sertie, A. L., Sossi, V., Camargo, A. A., Zatz, M., Brahe, C., and Passos-Bueno, M. R. (2000). Collagen XVIII, containing an endogenous inhibitor of angiogenesis and tumor growth, plays a critical role in the maintenance of retinal structure and in neural tube closure (Knobloch syndrome). *Hum. Mol. Genet.* 9 (13), 2051–2058. doi:10.1093/hmg/9.13.2051

Shi, Z. D., Wang, H., and Tarbell, J. M. (2011). Heparan sulfate proteoglycans mediate interstitial flow mechanotransduction regulating MMP-13 expression and

cell motility via FAK-ERK in 3D collagen. *PLoS One* 6 (1), e15956. doi:10.1371/journal.pone.0015956

Simizu, S., Ishida, K., Wierzba, M. K., and Osada, H. (2004). Secretion of heparanase protein is regulated by glycosylation in human tumor cell lines. *J. Biol. Chem.* 279 (4), 2697–2703. doi:10.1074/jbc.m300541200

Simizu, S., Suzuki, T., Muroi, M., Lai, N. S., Takagi, S., Dohmae, N., et al. (2007). Involvement of disulfide bond formation in the activation of heparanase. *Cancer Res.* 67 (16), 7841–7849. doi:10.1158/0008-5472.can-07-1053

Sletten, E. T., Loka, R. S., Yu, F., and Nguyen, H. M. (2017). Glycosidase inhibition by multivalent presentation of heparan sulfate saccharides on bottlebrush polymers. *Biomacromolecules* 18 (10), 3387–3399. doi:10.1021/acs.biomac.7b01049

Toyoshima, M., and Nakajima, M. (1999). Human heparanase. *J. Biol. Chem.* 274 (34), 24153–24160. doi:10.1074/jbc.274.34.24153

Varki, A., Etzler, M. E., Cummings, R. D., and Esko, J. D. (2009). *Essentials of glycobiology*. Editors A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, et al. (NY: Cold Spring Harbor).

Vlodavsky, I., Beckhove, P., Lerner, I., Pisano, C., Meirovitz, A., Ilan, N., et al. (2012). Significance of heparanase in cancer and inflammation. *Cancer Microenviron.* 5 (2), 115–132. doi:10.1007/s12307-011-0082-7

Vlodavsky, I., Friedmann, Y., Elkin, M., Aingorn, H., Atzmon, R., Ishai-Michaeli, R., et al. (1999). Mammalian heparanase: Gene cloning, expression and function in tumor progression and metastasis. *Nat. Med.* 5 (7), 793–802. doi:10.1038/10518

Wu, L., Jiang, J., Jin, Y., Kallemeijn, W. W., Kuo, C. L., Artola, M., et al. (2017). Activity-based probes for functional interrogation of retaining beta-glucuronidases. *Nat. Chem. Biol.* 13 (8), 867–873. doi:10.1038/nchembio.2395

Wu, L., Viola, C. M., Brzozowski, A. M., and Davies, G. J. (2015). Structural characterization of human heparanase reveals insights into substrate recognition. *Nat. Struct. Mol. Biol.* 22 (12), 1016–1022. doi:10.1038/nsmb.3136

Yang, Y., Gorzelanny, C., Bauer, A. T., Halter, N., Komljenovic, D., Bauerle, T., et al. (2015). Nuclear heparanase-1 activity suppresses melanoma progression via its DNA-binding affinity. *Oncogene* 34 (47), 5832–5842. doi:10.1038/onc.2015.40

Zernichow, L., Abrink, M., Hallgren, J., Grujic, M., Pejler, G., and Kolset, S. O. (2006). Serglycin is the major secreted proteoglycan in macrophages and has a role in the regulation of macrophage tumor necrosis factor-alpha secretion in response to lipopolysaccharide. *J. Biol. Chem.* 281 (37), 26792–26801. doi:10.1074/jbc.m512889200

Zhang, Y., Xu, F., Guan, L., Chen, M., Zhao, Y., Guo, L., et al. (2022). Histone H4 induces heparan sulfate degradation by activating heparanase in chlorine gas-induced acute respiratory distress syndrome. *Respir. Res.* 23 (1), 14. doi:10.1186/s12931-022-01932-y

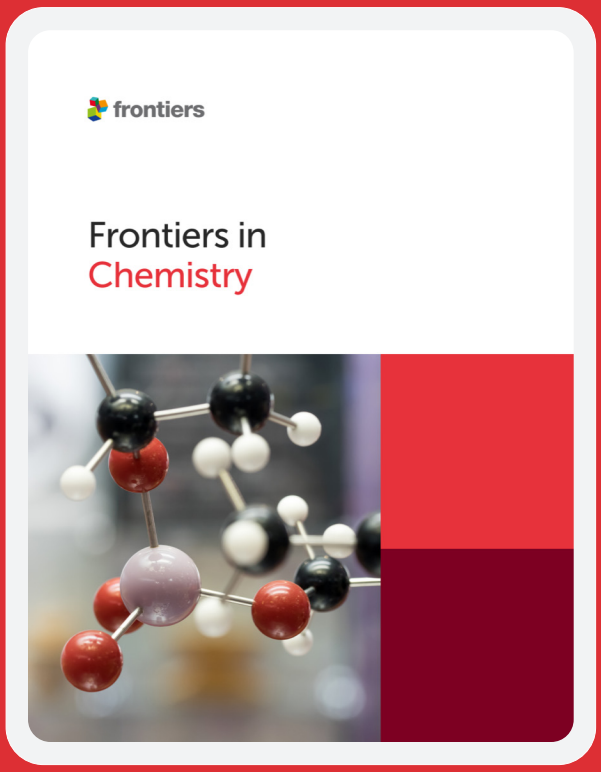
Zhou, H., Roy, S., Cochran, E., Zouaoui, R., Chu, C. L., Duffner, J., et al. (2011). M402, a novel heparan sulfate mimetic, targets multiple pathways implicated in tumor progression and metastasis. *PLoS One* 6 (6), e21106. doi:10.1371/journal.pone.0021106

Frontiers in Chemistry

Explores all fields of chemical science across the periodic table Advances our understanding of how atoms, ions, and molecules come together and come apart. It explores the role of chemistry in our everyday lives - from electronic devices to health and wellbeing.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact