This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal’s standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.
Efficient Chemoenzymatic Synthesis of an N-glycan Isomer Library

Lei Li,a Yunpeng Liua, Cheng Ma,a Jingyao Qu,a Angie D. Calderon,a Baolin Wub, Na Wei,a Xuan Wang,a Yuxi Guoa, Zhongying Xiaoa, Jing Song,a Go Sugiarto,b Yanhong Lia, Hai Yua, Xi Chena and Peng George Wang*a

Quantification, characterization and biofunctional studies of N-glycans on proteins remain challenging tasks due to complexity, diversity and low abundance of these glycans. The availability of structurally defined N-glycans (especially isomers) libraries is essential to help on solving these tasks. We reported herein an efficient chemoenzymatic strategy, namely Core Synthesis/Enzymatic Extension (CSEE), for rapid production of diverse N-glycans. Starting with 5 chemically prepared building blocks, 8 N-glycan core structures containing one or two terminal N-acetyl-D-glucosamine (GlcNAc) residue(s) were chemically synthesized via consistent use of oligosaccharidic thioethers as glycosylation donors in the convergent fragment coupling strategy. Each of these core structures was then extended to 5 to 15 N-glycan sequences by enzymatic reactions catalyzed by 4 robust glycosyltransferases. Success in synthesizing N-glycans with Neu5Gc and core-fucosylation further expanded the ability of enzymatic extension. High performance liquid chromatography with an amide column enabled rapid and efficient purification (>98% purity) of N-glycans in milligram scales. A total of 73 N-glycans (63 isomers) were successfully prepared and characterized by MS and NMR. The CSEE strategy provides a practical approach for “mass production” of structurally defined N-glycans, which are important standards and probes for Glycoscience.

Introduction

Glycans are ubiquitous and play diverse roles in a wide range of biological processes, such as protein folding and degradation, glycoproteostasis, cell adhesion and trafficking, cell signaling, fertilization and embryogenesis, as well as pathogen recognition and immune responses. Abnormal cell surface glycoforms and/or glycan-profiles are usually related to diseases such as cancer and atherosclerosis. Accordingly, glycan/glycoprotein biomarkers have been developed. Thus, elucidating the structures and functions of glycans is essential for understanding carbohydrate related biological and pathological processes, and for developing diagnostics and therapeutics for human diseases.

N-glycans found in nature possess an inherited complexity and diversity. These are mainly due to the variable and multiple connectivity of glycan building blocks (monosaccharides) and the process that they are assembled in biosystems. In mammalian glycomics, numerous glycan structures can be formed including branched-, regio- and stereo-isomers from only 10 common monosaccharide building blocks. Unlike precise template directed transcription/translation of nucleic acids/proteins, glycan structures are determined by the activities of glycosyltransferases (GTs), glycosidases, and other glycan biosynthetic enzymes, as well as the availability of donor substrates. For example, more than 30 GTs and glycosidases in the Golgi of human cells are involved in processing N-glycans. The expression, activity, substrate specificity, and localization of each enzyme have the potential to influence the assembly of N-glycans. It is thus understandable that N-glycans are extremely micro-heterogeneous even in one particular N-glycosylation site. For example, 58 different complex N-glycan structures were identified at one N-glycan site in mouse zona pellucida glycoprotein 3. As a result, despite decades of efforts in developing novel approaches for glycan analysis, absolute quantification and characterization of complex mixtures of N-glycans remain challenging tasks. At present, the main approach for characterizing N-glycan isomers is ion-trap mass spectrometry (MS) analysis of permethylated glycans, which requires large quantities of samples, therefore not suitable for low abundance glycans and rare biological samples. The availability of libraries of structurally defined N-glycans (especially isomers) provides essential standards and probes for MS-based N-glycan analysis and glycan microarray studies of carbohydrate binding proteins.

Given the difficulties in separating structurally defined glycans from natural resources, chemical or chemoenzymatic approaches have been developed for the synthesis of mostly symmetric N-glycans in the last two decades. Among chemically synthesized N-glycans, only few contains terminal sialic acid (Sia) due to difficulties in sialic acid chemistry, which was later overcome by enzymatic glycosylation using sialyltransferases. Most recently, Boons developed a strategy for chemoenzymatic synthesis of asymmetrical N-glycans, and 14 tri-antennary complex N-glycans were obtained. Nevertheless, only a few N-glycan structures were prepared in each report, mainly due to their complexity and diversity. A simple and robust strategy for efficient production of large numbers of N-glycan structures is still highly desirable.
Another roadblock in the rapid access of glycans in high purity is the purification strategy, which now largely relies on gel filtration chromatography (usually Sephadex G-25 or Bio-gel P2). However, even though gel filtration has been applied for decades in purifying glycans, it is time-consuming, less efficient, and may waste significant portion of products in preparing small quantities of precious N-glycans. Therefore, a more reliable and rapid N-glycan purification approach is yet to be developed.

LewisX [LeX, Galβ1,4-(Fucα1,3-)GlcNAc] and sialyl Lewis X [SLeX, Siaα2,3-Galβ1,4-(Fucα1,3-)GlcNAc] are among the most biologically significant glycan epitopes. For example, also known as CD15 antigen or SSEA-1 trisaccharide, LeX plays a role in the development of central nervous systems in vertebrates, and interferes with pathogen transfer in breastfed infants. SLeX is a specific ligand on human leukocytes for E-, L-, and P-selectins, and was shown to mediate leukocyte recruitment. SLeX (on both N- and O-glycans of glycoprotein zona pellucida) was also shown to mediate human sperm binding during fertilization. In addition, LeX and SLeX are usually overexpressed on the surface of cancer cells. Granted all the significances, N-glycans containing these epitopes were not synthesized until recently. In this study, we described an efficient Core Synthesis/Enzymatic Extension (CSEE) strategy and a HPLC based purification approach for rapid preparation of N-glycans with/without (S)LeX epitopes. In this strategy, 8 N-glycan core structures with GlcNAc residue(s) at the non-reducing terminal were firstly synthesized by convergent assembly of 5 building blocks. A set of robust GTs were then used to elongate these cores to yield a library of 73 N-glycans (Fig. 1 & 3). The development of an HPLC based approach using an amide column enabled rapid purification of milligrams of the chemoenzymatically synthesized N-glycans to minimum 98% purity. In addition, MS analysis of selected N-glycans yielded unique fragmentation patterns that may be used for distinguishing certain isomers.
Results

Convergent Core Synthesis. A good amount of work had been reported in chemically synthesizing N-glycan structures. For example, by Danishefsky and Unverzagt synthesized multi-antennary complex type N-glycans with acetylated Schmidt’s trichloroacetimidate donor. After enzymatic extension of the unprotected branch, the acetyl groups can be removed easily for chemical glycosylation using a diverse set of trifluoroacetimidate donors. In these cases, to prepare various glycolipids, a temporary anionic protecting group was used and then transformed into trichloroacetimidates, fluorides or trifluoroacetimidates, depending on the choice of glycosylation reaction. In this study, we developed an efficient convergent strategy that utilized oligosaccharidyl thioether as a versatile donor for glycosylation, facilitating assembly in just one or two glycosylation step(s) with excellent yield and good stereoselectivity.

We envisaged that trisaccharide 1 (Fig. 2) containing a crucial β-mannoside would be a versatile precursor for the synthesis of core structures, where the C4, C6-hydroxyl groups (OH) of the β-Man are protected with benzylidene and the C3-OH is unprotected to allow further chemical glycosylation. Installation of β-mannoside, the most challenging task in N-glycan synthesis, was accomplished using Crich-Kahne conditions with satisfying yield and stereocontrol, which was then glycosylated with 4 to yield pentasaccharide 16 in 91% yield and a satisfactory stereoselectivity (α/β = 3.5:1). Compound 16 was further deprotected to yield N040 as previously described. The core structures and stereosemistry of all glycosidic linkages was confirmed by NMR (ESI†).

Enzymatic Extension of N-glycans. The core structure N010 was produced in a total yield of 63% over the three steps.

Similarly, cores N000, N020, N030, N050, N110, and N210 were synthesized by installing 2, 3 or 4 onto C3-hydroxy of β-man of 1, then followed by installation of corresponding building blocks onto the α1,6Man branch. For the synthesis of N040, simple 3-O-benzylaion and controlled reductive cleavage of the benzylidene acetal of 1 was performed to afford acceptor 15, which was then glycosylated with 4 to yield pentasaccharide 16 in 91% yield and a satisfactory stereoselectivity (α/β = 3.5:1). Compound 16 was further deprotected to yield N040 as previously described. The structures and stereosemistry of all glycosidic linkages was confirmed by NMR (ESI†).

Enzymatic Extension of N-glycans. We intend to develop an enzymatic strategy that can efficiently generate 6 glycans (including Le X and SLe X) starting with any terminal GlcNAc residue. To this end, several robust GTs were chosen for the proof-of-concept experiment: β1,4-galactosyltransferase from bovine milk (B4GALT1); α2,3-sialyltransferase 1 mutant E271F/R313Y of Pasteurella multocida (PmST1m) with reduced α2,3-sialidase activity; b) α2,6-sialyltransferase from Photobacterium damsela (Pd2,6ST); c) C-terminal 66 amino acids truncated α1,3-fucosyltransferase from Helicobacter pylori (Hpc1,3FT). For the commercially available B4GALT1, all GTs were from bacteria and have high expression levels in Escherichia coli, high activity, and relatively relaxed substrate specificities. Using GlcNAc-OBn as a starting material, we validated the activities of these GTs (Scheme 2). Same as previously reported, the enzymes could efficiently catalyze the formation of LacNAc (Galβ1,4-GlcNAc), Sialyl-LacNAc (Siaβ2,3-Galβ1,4-GlcNAc), Sialyl-LacNAc (Siaβ2,3-Galβ1,4-GlcNAc), Le X, and SLe X-OBn. For reactions catalyzed by B4GALT1, Pd2,6ST and Hpc1,3FT, no product hydrolysis was observed (analyzed by MS) even with excessive

![Scheme 2](Image)

**Scheme 2**: Proof-of-concept experiment for the proposed enzymatic extension strategy. Reagents and conditions: a) UDP-Gal, Mn2+ and β1,4-galactosyltransferase from bovine milk (B4GALT1); b) CMP-Sia and double mutant E271F/R313Y of α2,3-sialyltransferase 1 from Pasteurella multocida (PmST1m); c) CMP-Sia and α2,6-sialyltransferase from Photobacterium damsela (Pd2,6ST); d) GDP-Fuc, Mn2+ and C-terminal 66 amino acids truncated α1,3-fucosyltransferase from Helicobacter pylori (Hpc1,3FT).
amounts of enzymes and up to 24 h of incubation. However, product hydrolysis was observed for PmST1m with an excess amount of enzyme and an extended reaction time (>1 h), even the α2,3-sialidase activity has been reduced for 6333-fold by mutation of two amino acid residues. 22 Therefore, controlling the amount of PmST1m and the reaction time is still important for the synthesis of α2,3-sialosides. The attempt to synthesize unnatural tetrasaccharide S6Le^3 [Sia^α2,6-Galβ1,4-(Fuc^α1,3-)GlcNAc] failed utilizing either Hpa1,3FT or Pd2,6ST, with the formation of only MS detectable product after 24 h of Pd2,6ST incubation (Fig. S3). In addition, PmST1m can hardly catalyze the sialylation of Le^a to form SLe^a. Such information is important for designing sequential enzymatic N-glycan synthesis schemes.

Utilizing the enzymatic extension strategy, N011 - N015 were prepared starting with chemically prepared cores N010 (Fig. 3A).

Scheme 1 Convergent synthesis of N-glycan core structures

| Reagents and conditions: | (a) NIS, AgOTf, Et2O, 0 °C, 6: 93%; 8: 85%; 9: 83%; 13: 90% (α:β = 5:1); 16: 91% (α:β = 3:5:1); 17: 94%; 19: 86% (α:β = 4:1); (b) PhBCl, Et3SiH, DCM, -78 °C, 6: 96%; 7: 92%; 12: 95%; 15: 95%; 18: 93%; (c) 1) ethylenediamine, n-butanol, 90 °C; 2) Ac2O, pyridine, rt; 3) Pd(OH)_2, H_2, MeOH/H_2O (10:1), over three steps, N010: 63%; N020: 67%; N040: 61%; N011: 69%; N110: 65%; (d) NIS, AgOTf, DCM, 0 °C, 90 °C; 10: 91%; 11: 95%; (e) 1) ethylenediamine, n-butanol, 90 °C; 2) Ac2O, pyridine, rt; (3) NaOMe, MeOH; (4) Pd(OH)_2, H_2, MeOH/H_2O (10:1), over four steps, N030, 53%; N050: 55%; (f) BnBr, NaH, DMF, 90%; (g) 30% NH_3OHH_2O (1:10), quant. |
Firstly, in a 1.5 mL reaction system, 9 mg of N010 (4 mM) was incubated with UDP-Gal (8 mM), MnCl2 (5 mM), and B4GALT1 (20 µM/µmole acceptor). One microliter of the reaction mixture was aliquoted every hour for analysis. MS analysis showed a peak at m/z = 719.7645, corresponding to N011 [M+2H]2+. Meanwhile, on the HPLC-ELSD (Evaporative Light Scattering Detector) profile, a new peak (TR = 14.86 min) was observed, the area of which grew while the peak corresponding to N010 (TR = 14.86 min) became smaller. After 6 h of incubation, the reaction was freeze-quenched at -80 °C for 30 min, and condensed into 300 µl for HPLC purification using a water/acetonitrile gradient elution, yielding 9.4 mg of N011 (94 % yield). The purified N011 (99% pure) was then utilized for the synthesis of N012, N013, and N014 (Fig 3A) catalyzed by PmST1m, Pd2,6ST, and Hpx1,3FT, respectively (see ESI† for details). It is worth noting that the reaction for the synthesis of N012 was only allowed to proceed for 30 min due to the sialidase activity of PmST1m. N015 was then synthesized from N012 using Hpx1,3FT.

The reaction took 20 h to achieve complete conversion (Fig. S4). Similarly, starting with other chemically synthesized cores (N000, N020, N030, N040 and N050), N-glycans N001 - N005, N021 - N025, N031 - N035, N041 - N045 and N051 - N055 were prepared in a manner analogous to that described above. All prepared N-glycans were analyzed by HPLC-ELSD, ESI/MALDI-MS, and NMR to confirm purity and structures (ESI†).

The synthesis of asymmetric bi-antennary N-glycans N1xx and N2xx (Fig. 1) was carried out by enzymatic extension of the unprotected antenna first and then the other. The synthesis of N1xx was illustrated in Figure 3B. Firstly, Gal was added by B4GALT1 to the GlcNAc residue in the α1,3Man branch of N110 to form N110a, galactosylation on the α1,6Man branch was avoided by peracetylation of corresponding GlcNAc residue. It should be noted that partial de-acetylation was observed when the reaction was incubated for over 12 h. After HPLC purification, N110a was de-acetylated using 30% of ammonium hydroxide: H2O (1:10) to achieve full conversion (Fig. S2). After de-acetylation, N110a was purified into minimum 98% pure by HPLC with a semi-preparative amide column (10 × 250 mm) before being used for the synthesis of N1xx. HPLC analysis of purified N-glycans were performed using an analytical amide column (4.6 × 250 mm) under a gradient condition (solution A: 100 mM ammonium formate, pH 3.4; solution B: acetonitrile; flow rate: 1 mL/min; B%: 65 - 50% within 25 min) monitored by ELSD. The adjacent two peaks on HPLC profiles compound corresponding to α and β anomers of N-glycans. MS data for purified N-glycans were obtained by ESI-MS.
afford N111, which was then used as a substrate for synthesizing other N1xx glycans in a strictly controlled sequential manner. For example, to obtain N155, the α1,3Man branch was firstly extended by PnST1m (Step 1) and Hpo1,3FT (Step 2) to yield N115, the α1,6Man branch was then extended by B4GALT1 (Step 3) and Hpo1,3FT (Step 4) (Fig. 3B). Such synthetic routes were designed according to the substrate specificities of GTs to avoid undesirable glycosylation. Particularly, N144 was not designed to be synthesized from N124 to avoid potential sialylation on the α1,3Man branch by Pd2,6ST. Instead, N-glycan N244 was synthesized from N123 catalyzed by Hpo1,3FT (Fig. 3B). Similarly, N-glycans N2xx and N144 were synthesized from N210 (Fig. S5).

N-glycolyneuraminic acid (Neu5Gc), often found on mammalian glycans, is another common Sialic acid molecule besides N-acetylneuraminic acid (Neu5Ac).25 Even though human cells cannot produce Neu5Gc because of the inactivation of gene encoding CMP-Neu5Ac hydroxylase,26 it was frequently detected on cells.27 Beside N-acetylneuraminic acid (Neu5Ac), Neu5Gc. Theoretically, another set of 57 N-glycans can be easily synthesized by simply replacing Neu5Gc of the glycans in N011, TR = 19.09 min; 6

The results showed that FUT8 was highly active in using all 4 N-glycans as substrates for core-fucosylation of N-glycans, and for preparing core-fucosylated N-glycan library. Specifically, 4 N-glycans with an identical α1,3Man branch but a different α1,6Man branch were selected for FUT8-catalyzed core-fucosylation (Fig. 3D). The results showed that FUT8 was highly active in using all 4 N-glycans as acceptors. Corresponding core-fucosylated N-glycans (N6030, N6000, N6211, N6212) (0.5 mg each) were synthesized accordingly. Further substrate specificity study showed that FUT8 may have a stricter requirement for structures on the α1,3Man branch than that of α1,6Man branch (detailed study is undergoing).

A HPLC based approach for rapid access of pure N-glycans. Gel filtration (Bio-gel P2, 1 × 110 cm) was first applied in the purification of synthesized N-glycans. Taking the separation of N001 (synthesized from N000 by B4GALT1 catalyzed reaction) as an example, MS of the P2 purified product showed a major peak at m/z = 821.2992 corresponding to N001 [M+2H]²⁺, and a minor peak at m/z = 740.2735 corresponding to N001 minus a Gal residue [M+2H]²⁺, possibly comes from incomplete galactosylation of N000. This result again indicated that gel filtration is inefficient for purifying N-glycans into high purities, especially when incomplete glycosylation occurred. As a consequence, previously reported N-glycan analysis usually employed excessive enzymes and long incubation times to push GT-catalyzed reactions towards completion.9

Hydrophilic interaction liquid chromatography (HILIC) provides a rapid and effective strategy for separating small polar compounds, and has been used extensively in glycan analysis.30 In these cases, N-glycans from biological samples were usually fluorescent labeled via reductive amination and then detected in picomole scales by UPLC-HILIC. However, HILIC has not been applied in milligram scale N-glycan purification. Using an analytical HILIC column (XBridge BEH amide column, 5 μm, 4.6 mm × 250 mm, Waters) under a gradient running condition (solvent A: 100 mM ammonium formate, pH 3.4; solvent B: acetonitrile; flow rate: 1 mL/min; B%: 70 - 50% within 50 min), the abovementioned Bio-gel P2-purified products were analyzed. Four peaks were observed in the HPLC profile using an evaporative light scattering detector (Fig. S1). Peaks 1 (TR = 21.68 min) and 2 (TR = 22.16 min) were next to each other and partially overlapped. The same observation was found for peaks 3 (TR = 24.51 min) and 4 (TR = 25.04 min). These peaks were collected in a parallel run monitored at A210 nm and subjected to MS analysis. Same m/z values were observed for peaks 3 (821.2997) and 4 (821.2991) (Fig. S1), implying that both peaks represented N-glycan N001 [M+2H]²⁺, possibly for α and β anomers, which is common for free glycans due to the process of mutarotation in water. This was confirmed by 'H NMR analysis that showed chemical shifts of both α and β anomers protons (ESI†). Similarly, peaks 1 and 2 represented to α and β anomers of N001 minus a Gal residue [M+2H]²⁺.

These results encouraged us to purify N001 using a semi-preparative HILIC column (10 × 250 mm). Under a similar gradient running condition (solvent A: 100 mM ammonium formate; solvent B: acetonitrile; flow rate: 4 mL/min; B%: 70 - 50% within 50 min; monitored at A210 nm), 10.5 mg of N001 were separated by 3 injections (Fig. S2) with a purity of higher than 98% as analyzed by HPLC-ELSD (ESI†). Different solvent combinations were later tested for N-glycan purification (Fig. S6). Results showed that 100 mM ammonium formate/acetonitrile gradient elution gave the best separation of all N-glycans tested. In addition, N-glycans without Sia residues were able to be separated to similar level using water/acetonitrile gradient elution, where sialylated N-glycans were eluted rapidly (TR < 3 min). Furthermore, it was found that a shorter running time with a narrower B% gradient (65 - 50% in 25 min) was able to achieve a similarly good separation level. Such running conditions were applied to separate enzymatically synthesized N-glycan into 98% (ESI†).

Under a standard running condition (solvent A: 100 mM ammonium formate; solvent B: acetonitrile; flow rate: 1 mL/min; B%: 65 - 50% within 25 min), all purified N-glycans were analyzed by HPLC-ELSD (ESI†). It was found that when different sugar residues were added to N-glycans, the retention time shifts of peaks on HPLC chromatograms generally decrease in the following order: Neu5Gcα2,3 with Fucα1,3 > Neu5Gcα2,6 > Neu5Acα2,3 with Fucα1,3 > Neu5Gcα2,3 > Neu5Acα2,6 > Galβ1,4 > Fucα1,3 > Neu5Acα2,3 > Fucα1,6. For example, the retention time of N015G, N013G, N012G, N013, N014, N012 are 20.66, 20.16, 19.39, 19.13, 19.09, 18.59, 17.93 min respectively. Such regularity may be found useful in HILIC-based profiling and identification of N-glycans.
were also found and may be used in distinguishing isomers in the enzymatic approaches to synthesize N-glycans, however, none was effort has been paid for developing chemical methodologies and demand for studies in Glycobiology and Glyomedicine. Decades of processes, libraries of structurally well-defined glycans are in urgent glycans, as well as significant roles they played in biological phenomena were also observed in other Neu5Acα2,3, Neu5Acα2,6 isomer pairs, such as N042 and N043, N212 and N213, N222 and N223. Therefore, signal intensity of fragmentation ions can be reliable evidence in distinguishing linkage patterns between Gal and Neu5Ac residues. Interestingly, this phenomenon was only observed for the structures on the α1,3Man branch, for example, while changing linkage on the α1,5Man branch (isomers N002 and N133, or N003 and N233) resulted in similar changes in spectra as described above, no significant changes were observed between N002 and N233 or N003 and N133, which differ only on the α1,6Man branch (Fig. S8). This phenomenon is interesting yet hard to comprehend. Several other unique MS² fragmentation patterns were also found and may be used in distinguishing isomers in the further.

Discussion

Given the diversity and micro-heterogeneity nature of complex glycans, as well as significant roles they played in biological processes, libraries of structurally well-defined glycans are in urgent demand for studies in Glycobiology and Glyomedicine. Decades of efforts has been paid for developing chemical methodologies and enzymatic approaches to synthesize N-glycans, however, none was able to (cost-)efficiently generate large numbers of N-glycans. This is mainly due to: (1) Chemical methodologies developed so far are not cost-effective and rapid enough to prepare numerous complex N-glycans. Generally, each methodology was developed specifically for certain N-glycan structures, thus not suitable for efficient synthesis of other complex N-glycans. (2) GTs applied in the enzymatic or chemoenzymatic synthesis approaches were mostly from eukaryotes,⁴⁶, ⁹, ³¹ which are typically hard to access and exhibit narrow substrate specificity towards glycan acceptors. Most recently, several N- and O-glycans with multisialylated poly-N-acetyllactosamine extensions were successfully synthesized using a single bacterial α2,6-sialyltransferase (Pd2,6ST).⁶ This revealed the power of bacterial GTs in diversifying glycans. (3) A highly efficient and rapid N-glycan purification approach is lacking. So far the only reported approach is gel-filtration, which takes many hours to separate each target, and is not able to separate complex glycans with only one monosaccharide difference.

This work found answers to the above three obstacles. Firstly, a highly efficient strategy was developed based on the consistent use of oligosaccharyl thioucer for the convergent installation of branched GlcNAc-terminated antennae to achieve high stereoselectivity with excellent yields. This approach minimized synthetic steps and maximized yield, which proceeded very efficiently with less glycosyl donor (1.3 equivalents) and mild conditions (at 0 °C). Notably, when the (Ac₃)GlcNAcβ1,2-Man disaccharide thioucer 3 and Bn-GlcNAcβ1,2-Man disaccharide thioucer 4 were used as donors, installations on 3-OH of tor 1 were achieved with excellent yield and high stereoselectivity. We were also able to install the Man₃ thioucer donor 5 on the 6-OH of the β-Man in good yield and high stereoselectivity, as seen before in our previous report.⁸, ³³ Using this strategy, 8 N-glycan core structures with 5 - 8 monosaccharide residues were convergently synthesized. We expect this strategy would allow us to prepare more N-glycans with various glycoforms for enzymatic extension. Secondly, a general enzymatic extension strategy is developed that can extend any GlcNAc terminated glycans to 5 more glycans (including Leα and SLeα) using B4GALT1 and three robust bacterial GTs. Such a strategy enabled generation of 5 - 15 more N-glycans.
from each chemically synthesized core structure, the synthesis of these N-glycans, each of the GTs was tested towards 10 to 21 N-glycan acceptors. For example, PmST1m showed comparable high activities towards N001, N011, N021, N031, N041, N051, N111, N123, N124, N125, N211, N223, N224 and N225 (which share a common Galβ1,4GalNAc motif), and efficiently catalyzed the formation of corresponding α2,3Sialylated N-glycans. In addition, the successful synthesis of Neu5Gc-terminated N-glycan N012G indicated that PmST1m is also promiscuous towards sugar donors. Furthermore, substrate specificity study revealed that Hpo1,3Ft can well accept various N-glycans terminated with LacNAc or Siaα2,3LacNAc (ESI†, Table S1). Similar relaxed substrate specificities were also found for B4GALT1 and Pd2,6ST towards various N-glycan acceptors. These results clearly indicated that: a) the 4 robust GTs only recognize the most adjacent one or two monosaccharide residues in glycosylation reactions, thus has a great potential to extend various N-glycans; b) the promiscuity of the bacterial GTs towards sugar donors does not affect by acceptors, no matter simple oligosaccharides or complex N-glycans were used, thus has a great potential to synthesize N-glycan derivatives.

Thirdly, instead of generally used gel-filtration, each N-glycan was purified to >98% by HPLC utilizing a HILIC column in milligram scales (up to 4 mg per run). This HPLC-based approach could well separate complex N-glycans with only one monosaccharide difference, and takes only 30 min per injection.

Among the synthesized structures, only a few (e.g. N011, N001, N002, N003, N6000) were previously synthesized via chemical approaches or chemoenzymatic approaches. This library covers a number of low molecular weight N-glycans which have or have not been identified, including most common hybrid and bi-antennary complex types. More importantly, this work represents the first report in preparing high pure N-glycan isomers. This N-glycan library contains 21 groups of isomers (Fig. S7) (2 to 6 distinct structures in each group), e.g. glycans N125, N134, N144, N225, N234 and N244 are isomers with the same molecular weight of 2077.7455. These groups of isomers are valuable standards that may be applied in absolute quantification and structural identification of N-glycans by MS.

Conclusions

We have successfully developed a Core Synthesis/Enzymatic Extension (CSEE) strategy for efficient synthesis of structurally defined N-glycans and a HPLC-based approach for rapid purification of these compounds. The combination of CSEE and HPLC purification allowed rapid access of 0.5 - 2 mg of 73 homogenous N-glycans (21 groups). These N-glycans are valuable materials for glycans analysis and bioactivity evaluation. In this work, oligosaccharyl triether was used as a chemical glycosylation donor for the convergent installation of branched GlcNAc-terminated antennae, this general and efficient approach produced 8 N-glycan core structures with high stereochemistry and excellent overall yields. This work also demonstrated that any GlcNAc-terminated glycans can be enzymatically extended to 5 or more longer glycans (including Leβ and SLeβ) using several robust glycosyltransferases. Moreover, this work showed that complex N-glycans are best purified using HPLC utilizing a HILIC column. In summary, the CSEE strategy described here provides a practical approach for rapid production of structurally defined N-glycans, and has the potential to become a general approach to solve the complexity and diversity of glycomes. This may mark a beginning of “mass production” of glycomes.

Acknowledgements

This work was supported by National Institute of Health (R01GM085267 to P.G.W., R01HD065122 to X.C.) and Federal funds from National Cancer Institute, National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN261201300040C.

Author Contribution

L.L., Y.L., C.M. and P.G.W. designed experiments, analyzed data and wrote the manuscript. L.L. performed enzymatic synthesis, purification and HPLC identification of glycans; Y.L., B.W., and Z.X. performed chemical synthesis and NMR analysis; C.M., J. Q. and N.W. performed MS analysis. All other authors expressed and purified enzymes.

Competing financial interests

The authors declare no competing financial interests.

References