

Cite this: *Chem. Sci.*, 2017, 8, 7932

## Chemoenzymatic synthesis of heparan sulfate and heparin oligosaccharides and NMR analysis: paving the way to a diverse library for glycobiochemists†

Xing Zhang,<sup>‡a</sup> Vijayakanth Pagadala,<sup>‡b</sup> Hannah M. Jester,<sup>b</sup> Andrew M. Lim,<sup>b</sup> Truong Quang Pham,<sup>c</sup> Anna Marie P. Goulas,<sup>b</sup> Jian Liu<sup>\*c</sup> and Robert J. Linhardt<sup>‡\*a</sup>

Heparan sulfate (HS) is a member of the glycosaminoglycans (GAG) family that plays essential roles in biological processes from animal sources. Heparin, a highly sulfated form of HS, is widely used as anticoagulant drug worldwide. The high diversity and complexity of HS and heparin represent a roadblock for structural characterization and biological activity studies. Access to structurally defined oligosaccharides is critical for the successful development of HS and heparin structure–activity relationships. In this study, a library of 66 HS and heparin oligosaccharides covering different sulfation patterns and sizes was prepared through an efficient method of chemoenzymatic synthesis. A systematic nuclear magnetic resonance spectroscopy study was firstly undertaken for every oligosaccharide in the library. In addition to the availability of different oligosaccharides, this work also provides spectroscopic data helpful for characterizing more complicated polysaccharide structures providing a safeguard to ensure the quality of the drug heparin. This HS/heparin library will be useful for activity screening and facilitate future structure–activity relationship studies.

Received 13th August 2017  
Accepted 20th September 2017

DOI: 10.1039/c7sc03541a

rsc.li/chemical-science

## Introduction

Heparan sulfate (HS) glycosaminoglycan (GAG) is a linear sulfated polysaccharide consists of repeating 1→4 glycosidically linked disaccharide units comprised of alternating uronic acid residues, β-D-glucuronic acid (GlcA) or α-L-iduronic acid (IdoA) and α-D-glucosamine (GlcN) residues.<sup>1</sup> The HS disaccharide units can be substituted with *N*-acetyl (Ac) or *N*-sulfo (S) groups and *O*-sulfo (S) groups (at the 3-*O*- and 6-*O*-positions of the GlcN residue and 2-*O*-position of the uronic acid residues).<sup>2</sup> HS is widely expressed on mammalian cell surfaces and in the extracellular matrix and plays essential roles in a number of biological processes, controlling angiogenesis, embryonic development, inflammatory responses, lipid metabolism, infectious disease, tumor metastasis and blood coagulation.<sup>3</sup> Heparin is a special form of HS, found within specialized granulated cells, having a higher level of sulfation, a higher

IdoA content, and displaying prominent anticoagulant activity.<sup>4</sup> There are an average of 2.6 sulfo groups and 80–90% IdoA per disaccharide unit in heparin, compared with only 0.6 sulfo groups and 20% IdoA per disaccharide unit in a typical HS.<sup>5</sup> A pentasaccharide sequence motif present within HS and heparin polysaccharides including, →4)GlcNS6S(1→4)GlcA(1→4)GlcNS3S6S(1→4)IdoA2S(1→4)GlcNS6S(1→, is responsible for its specific binding to the serine protease inhibitor (serpin) antithrombin III (AT).<sup>6</sup> A repeating trisulfated disaccharide sequence →4)IdoA2S(1→4)GlcNS6S(1→ corresponds to heparin's thrombin (or factor IIa, FIIa) binding site and facilitates the assembly of the ternary heparin–AT–FIIa complex required for heparin's global anticoagulant activity.<sup>7</sup>

Unfractionated heparins (UFHs) and low molecular weight heparins (LMWHs) are polypharmacological agents, complex mixtures of molecules prepared from animal tissues.<sup>8</sup> Due to the poorly regulated supply chain of pharmaceutical heparin, a worldwide distribution of oversulfated chondroitin sulfate (OCSC) contaminated heparin in 2007 was discovered and associated with over 200 deaths in the U.S., adversely affecting the purity and safety of animal-sourced heparins.<sup>9</sup> Another potential threat to the supply chain of heparin is that other bioactive entities, such as prions and viruses, might remain associated with the HS and heparin chains in animal extracts.<sup>10</sup> Thus, the cost-effective preparation of a more structurally defined heparin from non-animal sources is highly desirable.<sup>11</sup> Meanwhile, developing more reliable structural characterization techniques is critically important to safeguard the quality of heparin drugs.<sup>12</sup>

<sup>a</sup>Department of Chemistry and Chemical Biology, Rensselaer Polytechnic Institute, Troy, New York 12180, USA. E-mail: linhar@rpi.edu

<sup>b</sup>Glycan Therapeutics, LLC, Chapel Hill, North Carolina 27599, USA

<sup>c</sup>Division of Chemical Biology and Medicinal Chemistry, Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, North Carolina 27599, USA. E-mail: liuj@email.unc.edu

† Electronic supplementary information (ESI) available: Materials, detailed experimental protocols, supporting schemes and figures, synthetic methods, analytical data including HPLC profiles, mass and NMR spectra of synthesized oligosaccharides. See DOI: 10.1039/c7sc03541a

‡ Contributed equally to this work.



Heparin impurities and contaminants are often other sulfated polysaccharides that are challenging to distinguish as impurities or contaminants when present in the heterogeneous and polydisperse drug heparin. Often there are only relatively subtle structural features, such as sugar epimers, sulfo group position, or differences in linkage position and anomeric configuration, unique to these impurities or contaminants.<sup>13</sup> Nuclear magnetic resonance (NMR) spectroscopy is widely utilized for such sophisticated structure analysis and is considered one of the best methods for determining subtle structural differences in heterogeneous and polydisperse polysaccharides, such as identifying and quantifying the dermatan sulfate impurity and the oversulfated chondroitin sulfate heparin contaminant, and, thus, playing an important role in heparin quality control.<sup>14</sup> Pure HS and heparin oligosaccharide library standards have not been widely available and in the past have been isolated through the partial depolymerization of HS or heparin.<sup>15</sup> This makes NMR signal assignment of HS and heparin polysaccharides, often requiring multi-dimensional techniques, very challenging and time-consuming.<sup>16</sup> The synthesis of a library of structurally defined HS and heparin oligosaccharides in sufficient amounts to serve as NMR standards for the analysis of current heparin drugs, is especially beneficial for the analysis of structurally important or “rare” residues, such as GlcNS3S6S, GlcA2S and IdoA residues.<sup>17</sup> HS and heparin oligosaccharides of moderate size, with 5 to 10 saccharide units (pentasaccharides to decasaccharides), are generally required for high-affinity interactions with heparin-binding proteins and are responsible for many of heparin’s biological activities.<sup>18</sup> Moreover, a library comprised of moderate sized oligosaccharides would also support structure-activity relationship (SAR) studies.<sup>19</sup>

There are generally two approaches for the target driven synthesis of a library of structurally defined HS and heparin oligosaccharides. The first approach relies on a purely chemical synthesis method, based on repetitive steps of protection, activation, coupling and de-protection, and is incredibly challenging for the preparation of oligosaccharides larger than pentasaccharides.<sup>20</sup> For example, Arixtra®, an FDA-approved pentasaccharide anticoagulant drug, requires as many as 60 chemical steps, and is produced in a reported overall yield is only 0.1%.<sup>21</sup> In addition, during the preparation, there are many separation steps needed to remove undesirable isomers (especially the incorrect anomers) as well as the unavoidable side products that add cost and decrease overall yields. The second approach, chemoenzymatic synthesis, relies on combined chemical and enzymatic methods, mimicking the biosynthetic pathway of heparin, and represents a promising strategy to address many of these synthetic challenges.<sup>22</sup> The enzymatic component of synthesis is usually performed under mild conditions (20–37 °C) using aqueous “green” solvents catalyzing glycosylation reactions with exquisite stereoselectivity ( $\alpha$ - or  $\beta$ -glycosidic linkages) and sulfonation reactions with excellent regioselectivity, without the need for repetitive protection and deprotection steps.<sup>23</sup> In recent years, a number of reports on the chemoenzymatic synthesis of structurally defined oligosaccharides have demonstrated the feasibility of preparing those oligosaccharides that

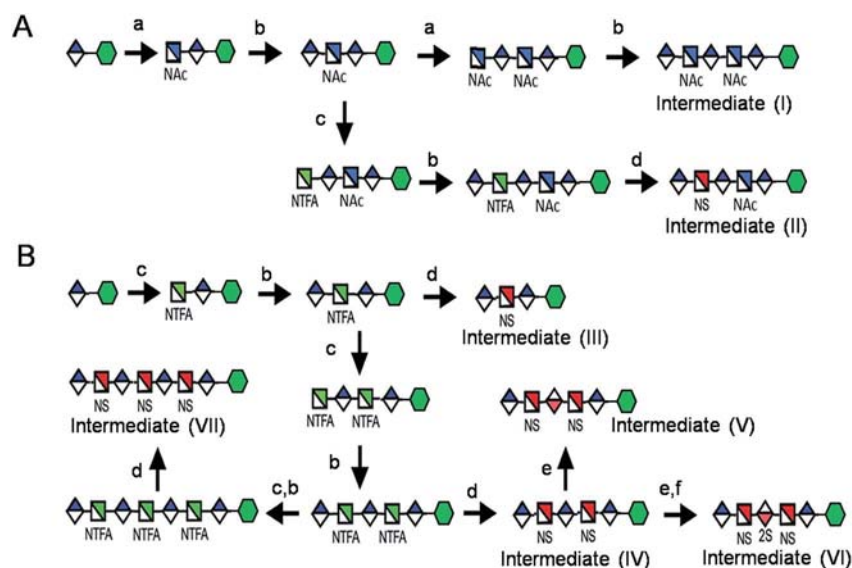
have been difficult to synthesize by traditional organic approaches.<sup>24</sup> Chen and coworkers<sup>24a</sup> have developed an efficient one-pot multi-enzyme (OPME) system utilizing with a series of unnatural sugar nucleotides to synthesize heparin oligosaccharides, where oligosaccharides were prepared without the required isolation or purification of intermediates. Wang and coworkers<sup>24b</sup> have reported a practical chemoenzymatic strategy, core synthesis/enzymatic extension (CSEE), for the rapid production of diverse *N*-glycans. In our previous efforts,<sup>24c-f</sup> we demonstrated that chemoenzymatic synthesis, starting from an inexpensive commercially available acceptor, *p*-nitrophenyl glucuronide (GlcA-*p*NP), which is ultraviolet detectable, is capable of generating a series of heparin oligosaccharides in good overall yield and possessed excellent anticoagulant activity. Herein, we have utilized another ultraviolet detectable, addressable acceptor, *N*-(6-azidohexanamidyl) *p*-aminophenyl glucuronide (GlcA-*p*NA-N<sub>3</sub>), for chemoenzymatic synthesis and a systematic NMR study, and produced the largest library of 66 HS and heparin oligosaccharides to date.

## Results

### Key intermediate synthesis

Seven key intermediates **I–VII** (Scheme 1 and ESI†), of sizes ranging from trisaccharide to heptasaccharide, were designed to prepare the target library of 66 oligosaccharides. The GlcN residues of pentasaccharide intermediate **I** were both GlcNAc and pentasaccharide intermediate **II** contained one GlcNAc and one GlcNS residue. These were primarily used to prepare oligosaccharides having low sulfation levels such as those typically encountered in HS (Scheme 1A). The GlcN residues of intermediates **III–VII** were all GlcNS and represent versatile precursors to access highly sulfated heparin oligosaccharides (Scheme 1B). Oligosaccharide synthesis was initiated on a  $\beta$ -GlcA(1→*p*NA-N<sub>3</sub>) monosaccharide acceptor and extended with UDP-sugar donors using a chemoenzymatic approach. Heparosan synthase-2 (pmHS2) from *Pasteurella multocida*<sup>22a</sup> was used to elongate this monosaccharide acceptor, to appropriate sized backbones **I–VII**. GlcNTFA containing backbones were then subjected to chemical de-*N*-trifluoroacetylation and modification by *N*-sulfo-transferase (NST) resulting in key intermediates **II–IV** and **VII**. Key intermediates **VI** and **VII** were subjected to additional enzymatic modification relying on C<sub>5</sub>-epimerase, 2-*O*-sulfo-transferase (2-OST), 6-*O*-sulfo-transferases (6-OSTs) and 3-*O*-sulfo-transferases (3-OSTs). For example, to synthesize intermediates **IV** through **VI**, GlcA-*p*NA-N<sub>3</sub> was first repetitively elongated using pmHS2 with UDP-*N*-trifluoroacetyl glucosamine (UDP-GlcNTFA) donor and UDP-glucuronic acid (UDP-GlcA), to form pentasaccharide, GlcA(1→4)GlcNTFA(1→4)GlcA(1→4)GlcNTFA(1→4)GlcA-*p*NA-N<sub>3</sub> (Scheme 1B). The resulting GlcNTFA residues were then converted to GlcNS by removing the trifluoroacetyl group under mild alkaline conditions followed by the introduction of *N*-sulfo groups using NST and 3'-phosphoadenosine-5'-phosphosulfate (PAPS) to obtain intermediate **IV**. Treatment of intermediate (**IV**) with C<sub>5</sub>-epimerase and 2-OST afforded IdoA2S-containing intermediate (**VI**) for further modification. C<sub>5</sub>-epimerase is known to be a bi-directional enzyme,





**Scheme 1** Route for the chemoenzymatic synthesis of seven key intermediates. Reagents and conditions: (a) pmHS2, GlcNAc; (b) pmHS2, GlcA; (c) pmHS2, GlcNTFA; (d) de-*N*-trifluoroacetylation, NST, PAPS; (e) C<sub>5</sub>-epimerase; (f) 2-O-ST, PAPS. pmHS2, *Pasteurella multocida* heparosan synthase-2; NST, *N*-sulfotransferase; 2-O-ST, 2-*O*-sulfotransferase; PAPS, 3'-phosphoadenosine-5'-phosphosulfate. Each intermediate was purified to minimum 97% purity by HPLC with a TSKgel DNA-NPR column (ESI).†

catalyzing both the forward and reverse reaction and affording an equilibrium mixture of GlcA and IdoA residues.<sup>25</sup> While the GlcA residue is thermodynamically favored, the IdoA residue can be selectively generated in the presence of 2-O-ST and PAPS to afford IdoA2S, which cannot be converted by C<sub>5</sub>-epimerase to GlcA2S.<sup>17b,c</sup> However, interestingly, we obtained IdoA containing intermediate V by treating with C<sub>5</sub>-epimerase alone. This resulting intermediate V could be applied into the preparation of more structurally complex IdoA-containing heparin oligosaccharides (Fig. 1). The sequence of enzymatic reactions needs to be carefully controlled for the synthesis of each desired oligosaccharide target. For example, C<sub>5</sub>-epimerase can only be used after the introduction of *N*-sulfo groups, as it will only act on a GlcA residue that is between two GlcNS residues. The remaining key intermediates were similarly synthesized (ESI†).

### Library oligosaccharides synthesis

Starting from pentasaccharide intermediate VI, GlcA(1→4)GlcNS(1→4)IdoA2S(1→4)GlcNS(1→4)GlcA-*p*NA-N<sub>3</sub> (where the residues are labeled alphabetically with the A residue GlcA at the non-reducing end and the E residue GlcA at the reducing end) three different GlcNS3S6S-containing oligosaccharides 14, 28, 46, were prepared (Fig. 2A and B). Intermediate VI was first elongated with pmHS2 and UDP-GlcNTFA to afford the hexasaccharide. After NTFA group deprotection, and subsequent enzymatic *N*-sulfonation and *O*-sulfonation, catalyzed by *N*-sulfotransferase, 2-*O*-sulfotransferase (2-O-ST), 6-*O*-sulfotransferase (6-O-ST-1 and 6-O-ST-3) and 3-*O*-sulfotransferase-1 (3-O-ST-1), GlcNS3S6S-containing hexasaccharide 14 was obtained (Fig. 2B and 3A). Following the same approach, pmHS2 and UDP-GlcA were employed to construct heptasaccharide 28 and octasaccharide 46. Chemoenzymatic synthesis requires careful reaction scheme design, especially the selection of an appropriate modification

sequence. For example, as residue C is flanked by a GlcA residue at its non-reducing end, 3-O-ST-1 selectively adds a 3-*O*-sulfo group to residue C rather than residue E to generate compound 14 (Fig. 3A). Oligosaccharides 14, 28 and 46, contain an AT-binding pentasaccharide sequence, but also provide another option to access Arixtra®-like compounds with anticoagulant activity, or even ones in which GlcA-*p*NA-N<sub>3</sub> has been removed using Smith or alkaline degradation methods.<sup>26</sup>

The 2-*O*-sulfo glucuronic acid (GlcA2S) residue represents <1–5% of total uronic acid residues in HS and heparin isolated from natural sources and only limited information on the biological role of these rare saccharide residues has been reported.<sup>17b</sup> The availability of pure structurally defined GlcA2S-containing oligosaccharides should promote investigations into the SAR of this rare functional group.<sup>17b</sup> Pentasaccharide intermediate IV, GlcA(1→4)GlcNS(1→4)GlcA(1→4)GlcNS(1→4)GlcA-*p*NA-N<sub>3</sub>, was treated with 2-O-ST<sup>17d</sup> and PAPS to introduce a GlcA2S residue (Fig. 2C). Surprisingly, two oligosaccharides containing one GlcA2S residue, 65, or two GlcA2S residues, 66, respectively, were obtained with good regioselectivity and their structures were confirmed by ESI-MS and NMR analyses (ESI).

Using a similar approach, all the other compounds presented in Fig. 1 were obtained. Their purity was >97% as assessed by HPLC and their structures were confirmed by ESI-MS and NMR analyses (ESI).

### A HPLC based approach for rapid access to pure HS and heparin oligosaccharides

Gel filtration chromatography relying in Bio-gel P2 columns represents an excellent way of removing salts but is inefficient for oligosaccharide purification, since reaction byproducts, often introduced through incomplete glycosylation, have molecular weights similar to the target compounds. An HPLC





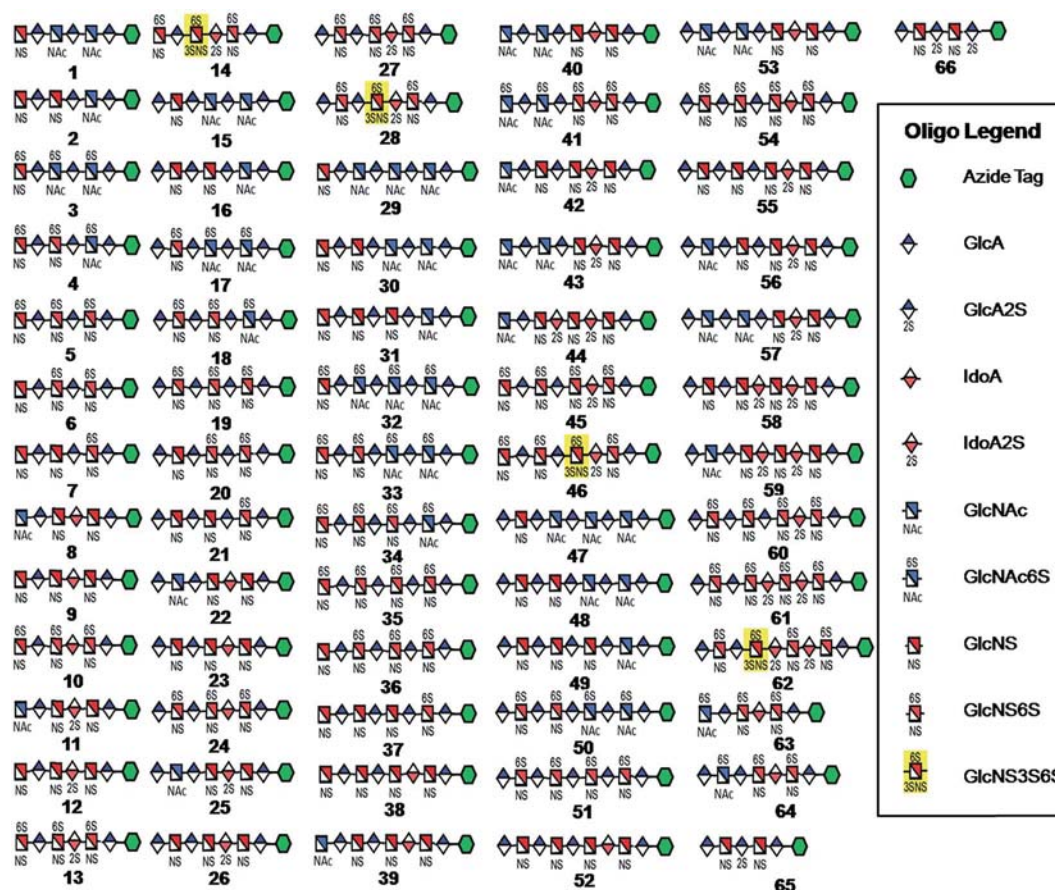


Fig. 1 HS and heparin oligosaccharides 1–66 were synthesized using a chemoenzymatic approach. The azide tag is *N*-(6-azidohexanamidyl) *p*-aminophenyl (*p*NA- $N_3$ ) and the chemical structure is presented in Fig. 2A.

based approach for rapid access to pure HS and heparin oligosaccharides was developed. The elongated products of the HS backbone prior to sulfation reactions were purified using a C-18 column, while all the sulfated products of the oligosaccharides were purified using Q-Sepharose fast flow column. Purity and structural analysis of the entire oligosaccharide library (Fig. 1) was conducted (ESI<sup>†</sup>). Representative data for the analysis of a GlcNS3S6S-containing hexasaccharide, 14 GlcNS6S (1 → 4)GlcA(1 → 4)GlcNS3S6S(1 → 4)IdoA2S(1 → 4)GlcNS6S(1 → 4)GlcA-*p*NA- $N_3$ , is presented in Fig. 3. Hexasaccharide 14 eluted as a single peak from high-resolution anion exchange HPLC column, indicating that this compound was pure (97% purity, Fig. 3B). The molecular mass of hexasaccharide 14 was determined to be  $1901.1 \pm 0.8$  by electrospray ionization mass spectrometry (ESI-MS), identical to the calculated molecular mass of 1900.6. During chemoenzymatic synthesis the strong UV absorbance of the *p*NA- $N_3$  aglycone facilitates the detection and purification of the synthetic intermediates.

### NMR characterization and study

The representative characterization of oligosaccharide 14 by NMR spectroscopy is presented in Fig. 4. In the <sup>1</sup>H NMR spectrum, six anomeric proton signals are assigned confirming that this compound is a hexasaccharide (Fig. 4A and 3A). The signals

corresponding to the anomeric protons of the three GlcNS residues (residues A, C and E) resonate from 5.4 to 5.6 ppm and the signals at 4.56 and 5.07 ppm correspond to GlcA residues B and F. The signal at 5.07 ppm of residue F-1 is due to the deshielding effect of the nearby aromatic ring. The <sup>3</sup>J<sub>HH</sub> coupling constants of A-1, C-1 and E-1 were 4 Hz while B-1 and F-1 were 8 Hz, suggesting α-linkages and β-linkages, respectively. Interestingly, when conducting NMR analysis on our oligosaccharide library, we discovered that the anomeric proton of IdoA2S residues usually appeared as broad peaks in the <sup>1</sup>H NMR spectra, and correlation signals in 2D spectra were often of very low intensity. For example, the anomeric proton of the IdoA2S residue (D-1) of hexasaccharide 14 showed a broad peak at 5.18 ppm and the (D-1, D-2) cross-peak was so weak that it was not observed in the <sup>1</sup>H-<sup>1</sup>H COSY spectrum. However, a relatively weak correlation in the IdoA2S residue between D-5 and D-4 was detected (4.76, 4.08). The broad anomeric signal for the IdoA2S observed in the <sup>1</sup>H NMR spectrum is attributed to the inter-conversion of the multiple conformational forms of the IdoA2S pyranose ring D, <sup>1</sup>C<sub>4</sub>-chair, <sup>2</sup>S<sub>0</sub>-skew boat and <sup>4</sup>C<sub>1</sub>-chair. The addition of ethylenediaminetetraacetic acid (EDTA) significantly sharpened the IdoA2S signals and often resulted in significant chemical shift variations, possibly due to the chelation of trace amounts of divalent cations.<sup>17c</sup>



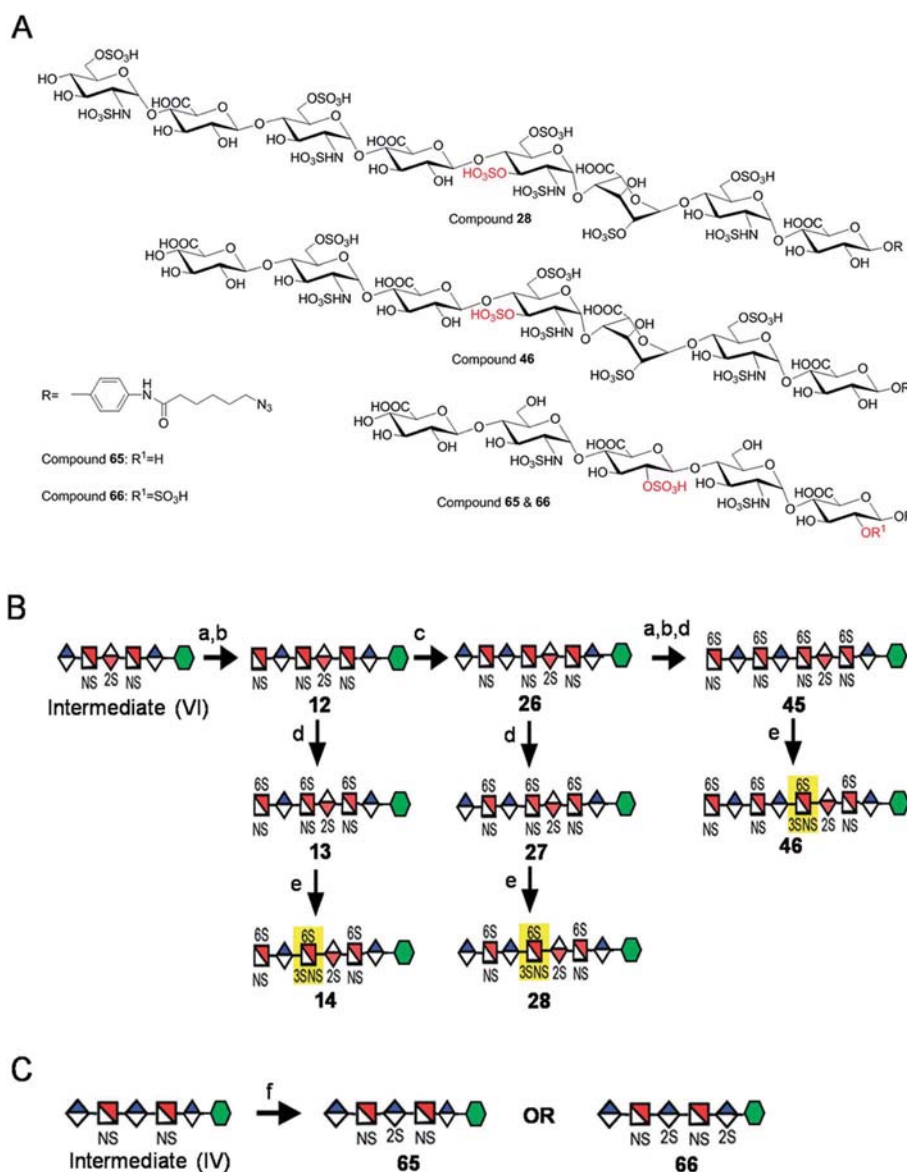


Fig. 2 Structures and synthetic schemes for different HS and heparin oligosaccharides. Panel A shows the structure of the compounds 28, 46, 65 and 66. Panel B shows the scheme for the synthesis of GlcNS3S6S containing compounds and panel C for the GlcA2S containing compounds. Reagents and conditions: (a) pmHS2, GlcNTFA; (b) deacetylation, NST, PAPS; (c) pmHS2, GlcA; (d) 6-OST, PAPS; (e) 3-OST-1, PAPS; (f) 2-OST, PAPS.

The proton signals from the saccharide residues in **14** are highly overlapped, requiring that we conduct 2D NMR experiments, including, <sup>1</sup>H-<sup>1</sup>H TOCSY, NOESY and <sup>1</sup>H-<sup>13</sup>C HSQC, to fully assign chemical shifts (Fig. 4C and ESI<sup>†</sup>). By combining TOCSY and COSY experiments, we were able to identify the proton signals of A-3 and A-4 were 3.56 and 3.51 ppm, respectively, which suggests that residue A is at the non-reducing end of the carbohydrate chain. The NOESY signals between F-1, F-5 protons and aromatic aglycone demonstrated that residue F was located at the reducing end of the chain (ESI<sup>†</sup>). The cross peak (3.39, 4.33) in COSY spectrum is corresponding to the correlation between H-2 and H-3 from residue C, supporting the presence of its important 3-O-sulfo group.

Since anomeric signals in an oligosaccharide are unique and are critical for making NMR assignments, a summary of the anomeric proton and carbon chemical shifts for various glucosamine and uronic acid residues, based on the NMR analysis of our library is presented in Table 1. It was not surprising that the anomeric chemical shift was relatively sensitive to the flanking residue on its reducing end, presumably due to the steric interactions. In contrast, the neighboring residue that is on the non-reducing end has a minimal impact on chemical shift values. These observations inspired us to summarize the anomeric chemical shift variation according to the neighboring residue on the reducing end.

The shift of the anomeric <sup>1</sup>H/<sup>13</sup>C is 5.52/97.0 ppm for GlcNS and 5.52/97.1 ppm for GlcNS6S, in the presence of a GlcA



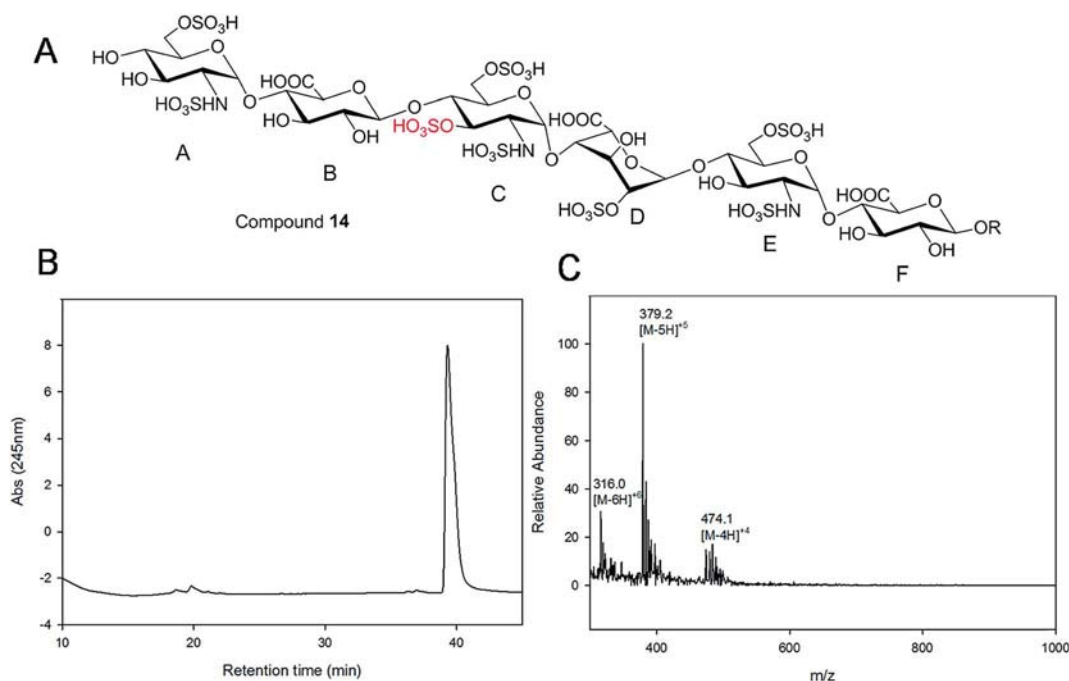


Fig. 3 Purity and structure analysis of hexasaccharide **14**. Panel A shows the chemical structure of hexasaccharide **14**. Panel B shows the chromatogram of TSKgel DNA-NPR HPLC analysis. Panel C shows the ESI-MS spectrum of hexasaccharide **14**. The molecular ions carrying 4, 5 and 6 negative charges are indicated.

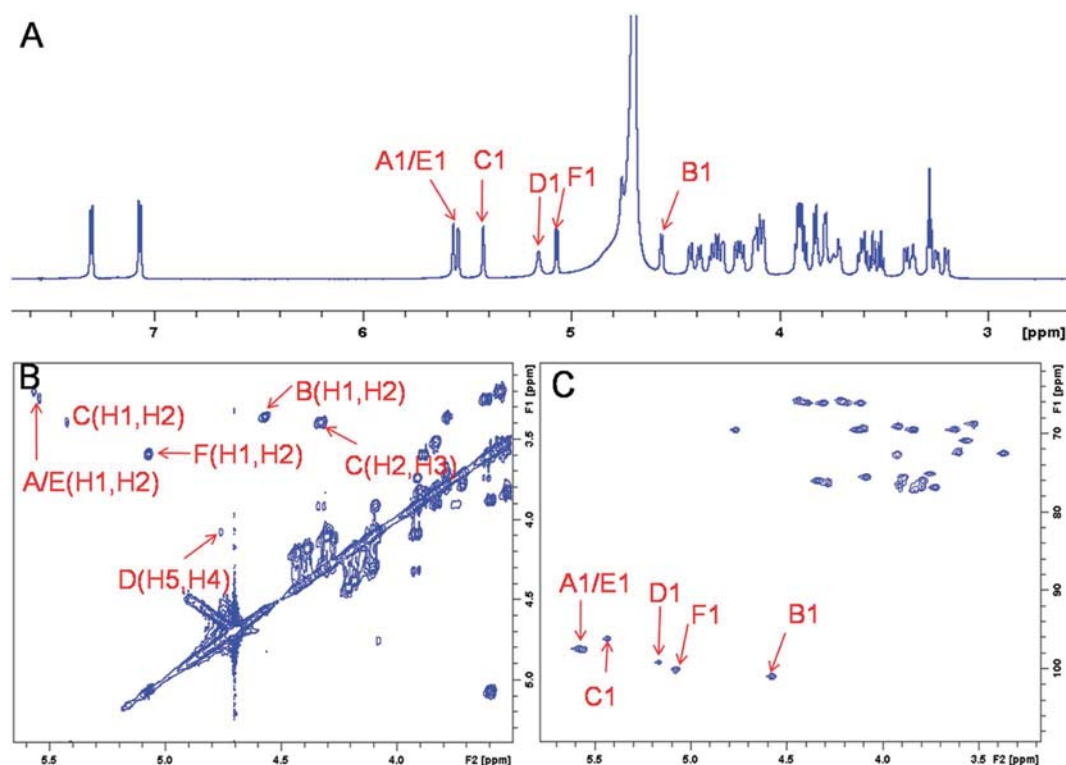


Fig. 4 NMR characterization of hexasaccharide **14** (structure see Fig. 3A). Panel A, B, C show the 1D  $^1\text{H}$  NMR, 2D  $^1\text{H}$ - $^1\text{H}$  COSY and  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectra, respectively. Peaks corresponding to the anomeric protons of these compounds can be clearly identified.

residue attached to its reducing end. The anomeric shift values are relatively unaffected by the presence of a 6-*O*-sulfo group. In contrast, when an IdoA ( $\pm 2\text{S}$ ) residue is on the reducing end of

a hexosamine residue, both proton and carbon anomeric signals for GlcNS and GlcNS6S are dramatically shifted to 5.24/97.3 and 5.33/96.2 ppm, respectively. Similarly, the anomeric





Table 1 Anomeric  $^1\text{H}$  and  $^{13}\text{C}$  chemical shifts ( $\delta$ ) of glucosamine and uronic acid residues in  $\text{D}_2\text{O}$  at 298 K

	$^1\text{H}$ , $\delta$ (ppm)		$^{13}\text{C}$ , $\delta$ (ppm)	
	GlcA <sup>a</sup>	IdoA( $\pm 2\text{S}$ ) <sup>a</sup>	GlcA <sup>a</sup>	IdoA( $\pm 2\text{S}$ ) <sup>a</sup>
Glucosamine				
GlcNS	5.52( $\pm 0.04$ )	5.24( $\pm 0.02$ )	97.0( $\pm 0.2$ )	97.3( $\pm 0.2$ )
GlcNS6S	5.52( $\pm 0.04$ )	5.33( $\pm 0.02$ )	97.1( $\pm 0.2$ )	96.2( $\pm 0.2$ )
GlcNS3S6S	— <sup>b</sup>	5.43( $\pm 0.04$ )	— <sup>b</sup>	96.0( $\pm 0.2$ )
GlcNAc	5.32( $\pm 0.03$ )	— <sup>b</sup>	96.8( $\pm 0.3$ )	— <sup>b</sup>
GlcNAc6S	5.33( $\pm 0.02$ )	— <sup>b</sup>	96.7( $\pm 0.3$ )	— <sup>b</sup>
Uronic acid				
	GlcNX <sup>c</sup>	GlcNX6S <sup>c</sup>	GlcNX <sup>c</sup>	GlcNX6S <sup>c</sup>
GlcA	4.43( $\pm 0.03$ )	4.54( $\pm 0.03$ )	102.0( $\pm 0.2$ )	101.5( $\pm 0.04$ )
GlcA2S	4.65	— <sup>b</sup>	100.1	— <sup>b</sup>
IdoA	4.95( $\pm 0.01$ )	4.95( $\pm 0.02$ )	101.5( $\pm 0.1$ )	101.6( $\pm 0.2$ )
IdoA2S	5.18( $\pm 0.02$ )	5.18( $\pm 0.02$ )	99.0( $\pm 0.1$ )	99.0( $\pm 0.1$ )

<sup>a</sup> Uronic residues are adjacent to the reducing end of glucosamine residues. <sup>b</sup> Not present. <sup>c</sup> X =  $\text{SO}_3$  or Ac; glucosamine residues are adjacent to the reducing end of uronic acid residues.

signal of GlcNAc and GlcNAc6S have a comparable chemical shift of 5.32/96.8 ppm if a GlcA residue is present on the reducing end side of this residue. The shift of the H-1 signal is very similar for GlcNAc(6S) and GlcNS6S, only a COSY spectrum could easily distinguish these due to their large differences in the shift values of their H-2 signals. The 3-*O*-sulfo group also exhibits an effect on GlcNS3S6S with an apparent downfield shift signal to  $\sim 5.43$  ppm, compared to the GlcNS and GlcNS6S residues. This observation is helpful in identifying the  $\rightarrow 4$ ) GlcNS3S6S(1 $\rightarrow$ 4)IdoA2S(1 $\rightarrow$  disaccharide unit, which is an important component of the AT-binding site.

Uronic acid residues show characteristic anomeric  $^1\text{H}/^{13}\text{C}$  shifts for GlcA of 4.43/102.0 ppm and GlcA2S of 4.65/100.1 ppm, respectively, in the presence of a neighboring GlcNX (X = Ac or  $\text{SO}_3\text{H}$ ) on its reducing end. The proton chemical shift for GlcA notably shifts to 4.54 ppm when the reducing end GlcNX residue carries a 6-*O*-sulfo group. It is noteworthy that the 6-*O*-sulfo group has a greater impact than *N*-sulfo group of the neighboring glucosamine residue on the reducing end. The anomeric position of the IdoA2S residue exhibits consistent chemical shifts of 5.18/99.0 ppm independent of whether it has a GlcNX or GlcNX6S at its reducing end. The anomeric position of the IdoA residue has very different chemical shifts 4.95/101.5 ppm.

## Discussion

Heparin is a widely used clinical anticoagulant for the prevention and treatment of arterial and venous thrombosis.<sup>7</sup> Due to the vulnerability of animal-sourced heparin towards the presence of impurities and contamination, developing a method to prepare heparin from non-animal sources represents a safer approach to the development of anticoagulant and antithrombotic therapeutics.<sup>24</sup> New synthetic versions of heparin might also make it possible to develop non-anticoagulant

heparins that could be exploited as antiviral and anticancer agents.<sup>27</sup> As our previous studies have already demonstrated, chemoenzymatically synthesized oligosaccharides exhibit excellent anticoagulant activity and improved pharmacological properties.<sup>24c,e</sup> The control the placement of uronic acid epimers and sulfo groups are essential for obtaining heparin and HS chains having specific biological functions. The diverse HS and heparin oligosaccharide library prepared in the current study serve as a valuable set of tools to perform SAR studies and also might contain potential drug candidates.

Introducing chemically active functional groups is expected to expand the use of the library. In the current library, each oligosaccharide possesses an azido group at the reducing end. The azide functionality is well known to be inert within biological systems but can be readily tagged covalently with imaging probes or modified using an azide-specific reaction, such as Staudinger ligation with phosphines and the [3 + 2] cycloaddition with alkynes.<sup>28</sup> The library of oligosaccharides synthesized in the current study are azide-functionalized, which permits their detection or selective capture in cells or in living organisms.<sup>29</sup> Moreover, azide can be reduced to a free amino group, which would be subsequently biotinylated using carbodiimide chemistry or through reductive amination.<sup>30</sup> The oligosaccharide library can be coupled to the surface of a microarray for high-throughput screening of heparin-binding proteins.<sup>31</sup>

Although sulfation at the 3-*OH* position of glucosamine occurs infrequently in HS, this sulfation pattern is closely related to its biological functions, facilitating the entry of herpes simplex virus into host cells to establish infection,<sup>32</sup> regulating axon guidance and growth of neurons,<sup>33</sup> controlling follicular properties in ovulation,<sup>34</sup> and enabling HS and heparin to bind to AT promoting blood anticoagulation.<sup>35</sup> It is still not entirely clear how the 3-*O*-sulfoglucosamine (GlcNS3S $\pm$ 6S) effects these important roles in contributing to HS biological activity. The ready availability of multi-milligram quantities of oligosaccharide structural variants containing a GlcNS3S6S residue should facilitate further biological investigation.

Due to the low abundance of 2-*O*-sulfoglucuronic acid (GlcA2S) in heparin and HS very little research has been focused on the SAR of this residue. Moreover, while 2-OST has been shown to act, albeit poorly, on GlcA in a polysaccharide substrate<sup>17d</sup> it had not been shown to act on oligosaccharide substrates, nor had the regioselectivity of this enzyme been established. IdoA is another less frequently occurring residue that has been difficult to prepare due to the reversible C5-epimerization. In this study, we successfully synthesized and fully characterized a series of GlcA2S- or IdoA-containing oligosaccharides. Ever since the 2007 heparin contamination heparin crisis, NMR spectroscopy has played an increasingly important role in heparin drug analysis. A library of oligosaccharides, including many possible and often rare structures is in demand but the availability of various oligosaccharides is still a roadblock for such studies. This is the first report of a library covering diverse heparin and HS oligosaccharides and



a systematic NMR study that should pave the way for improving our understanding of the SAR of HS and heparin.

## Conclusions

We have successfully developed a chemoenzymatic approach for the efficient preparation of 66 structurally defined HS and heparin oligosaccharides, including “rare” and critical moieties, such as GlcA2S, IdoA, GlcNS3S6S and AT-binding penta-saccharide sequence. The ultraviolet detectable tag and an HPLC-based purification method facilitate the rapid access to a diverse library containing tens of milligrams of high purity HS and heparin oligosaccharides. These oligosaccharides are not only potentially important for bioactivity evaluation but also good substrates to construct glycoconjugates or microarrays through azide chemistry. This work also describes a systematic NMR study on these HS and heparin oligosaccharides, paving the way to more complicated HS and heparin polysaccharide characterization and helping to safeguard the quality of the critically important drug, heparin.

## Author contribution

R. J. L. and J. L. designed experiments, analyzed data and wrote the manuscript. X. Z. and V. P., H. M. J., A. M. L. and A. M. P. G. performed all the synthesis and purification of oligosaccharides described within this manuscript. T. Q. P. prepared the enzymes and cofactors for the synthesis.

## Conflicts of interest

J. L. is a founder of Glycan Therapeutics. V. P., H. M. J., A. M. L. and A. M. P. G. are employees of Glycan Therapeutics. X. Z., T. Q. P. and R. J. L. declare no competing financial interests.

## Acknowledgements

This work is supported by NIH SBIR contract HHSN261201500019C and grants from National Institutes of Health (GM102137, HL62244 and HL094463), and Eshelman Innovation Institute.

## References

- 1 F. E. Poulain and H. J. Yost, *Development*, 2015, **142**, 3456–3467.
- 2 L. Fu, M. Sufita and R. J. Linhardt, *Adv. Drug Delivery Rev.*, 2016, **97**, 237–249.
- 3 E. I. Oduah, R. J. Linhardt and S. T. Sharfstein, *Pharmaceuticals*, 2016, **9**, 38.
- 4 R. J. Linhardt, *J. Med. Chem.*, 2003, **46**, 2551–2564.
- 5 C. C. Griffin, R. J. Linhardt, C. L. Van Gorp, T. Toida, R. E. Hileman, R. L. Schubert II and S. E. Brown, *Carbohydr. Res.*, 1995, **276**, 183–197.
- 6 Y. Chen, L. Lin, I. Agyekum, X. Zhang, K. St Ange, Y. Yu, F. Zhang, J. Liu, I. J. Amster and R. J. Linhardt, *J. Pharm. Sci.*, 2017, **106**, 973–981.
- 7 A. Onishi, K. St Ange, J. S. Dordick and R. J. Linhardt, *Front. Biosci.*, 2016, **21**, 1372–1392.
- 8 U. Bhaskar, E. Sterner, A. M. Hickey, A. Onishi, F. Zhang, J. S. Dordick and R. J. Linhardt, *Appl. Microbiol. Biotechnol.*, 2012, **93**, 1–16.
- 9 M. Guerrini, D. Beccati, Z. Shriver, A. Naggi, K. Viswanathan, A. Bisio, I. Capila, J. C. Lansing, S. Guglieri, B. Fraser, A. Al-Hakim, N. S. Gunay, Z. Zhang, L. Robinson, L. Buhse, M. Nasr, J. Woodcock, R. Langer, G. Venkataraman, R. J. Linhardt, B. Casu, G. Torri and R. Sasisekharan, *Nat. Biotechnol.*, 2008, **26**, 669–675.
- 10 D. Keire, B. Mulloy, C. Chase, A. Al-Hakim, D. Cairatti, E. Gray, J. Hogwood, T. Morris, P. Mourão, M. D. C. Soares and A. Szajek, *BioPharm Int.*, 2015, **28**, 36–42.
- 11 J. Liu and R. J. Linhardt, *Nat. Prod. Rep.*, 2014, **31**, 1676–1685.
- 12 A. Y. Szajek, E. Chess, K. Johansen, G. Gratzl, E. Gray, D. Keire, R. J. Linhardt, J. Liu, T. Morris, B. Mulloy, M. Nasr, Z. Shriver, P. Torralba, C. Viskov, R. Williams, J. Woodcock, W. Workman and A. Al-Hakim, *Nat. Biotechnol.*, 2016, **34**, 625–630.
- 13 Z. Zhang, B. Li, J. Suwan, F. Zhang, Z. Wang, H. Liu, B. Mulloy and R. J. Linhardt, *J. Pharm. Sci.*, 2009, **98**, 4017–4126.
- 14 T. R. Rudd, D. Gaudesi, M. A. Lima, M. A. Skidmore, B. Mulloy, G. Torri, H. B. Nader, M. Guerrini and E. A. Yates, *Analyst*, 2011, **136**, 1390–1398.
- 15 A. Pervin, C. Gallo, K. A. Jandik, X. J. Han and R. J. Linhardt, *Glycobiology*, 1995, **5**, 83–95.
- 16 B. Casu, M. Guerrini, A. Naggi, G. Torri, L. De-Ambrosi, G. Boveri, S. Gonella, A. Cedro, L. Ferró, E. Lanzarotti, M. Paterno, M. Attolini and M. G. Valle, *Arzneimittelforschung*, 1996, **46**, 472–477.
- 17 (a) Z. Wang, P. Hsieh, Y. Xu, D. Thieker, E. J. E. Chai, S. Xie, B. Cooley, R. J. Woods, L. Chi and J. Liu, *J. Am. Chem. Soc.*, 2017, **139**, 5249–5256; (b) P. Hsieh, Y. Xu, D. A. Keire and J. Liu, *Glycobiology*, 2014, **24**, 681–692; (c) P. Hsieh, D. F. Thieker, M. Guerrini, R. J. Woods and J. Liu, *Sci. Rep.*, 2016, **6**, 29602; (d) J. Rong, H. Habuchi, K. Kimata, U. Lindahl and M. Kusche-Gullberg, *Biochemistry*, 2001, **40**, 5548–5555.
- 18 J. Yin and P. H. Seeberger, *Methods Enzymol.*, 2010, **478**, 197–218.
- 19 E. A. Yates, S. E. Guimond and J. E. Turnbull, *J. Med. Chem.*, 2004, **47**, 277–280.
- 20 (a) R. Lucas, J. Angulo, P. M. Nieto and M. Martín-Lomas, *Org. Biomol. Chem.*, 2003, **1**, 2253–2266; (b) S. U. Hansen, G. J. Miller, M. J. Cliff, G. C. Jayson and J. M. Gardiner, *Chem. Sci.*, 2015, **6**, 6158–6164; (c) G. J. Miller, S. U. Hansen, E. Avizienyte, G. Rushton, C. Cole, G. C. Jayson and J. M. Gardiner, *Chem. Sci.*, 2013, **4**, 3218–3222.
- 21 (a) M. Petitou and C. A. van Boeckel, *Angew. Chem., Int. Ed.*, 2004, **43**, 3118–3133; (b) M. Petitou, J. C. Jacquinot, J. Choay, J. C. Lormeau and M. Nassr, *U.S. Pat.*, 4,818,816, 1989.
- 22 (a) P. L. DeAngelis, J. Liu and R. J. Linhardt, *Glycobiology*, 2013, **23**, 764–777; (b) H. J. Gijzen, L. Qiao, W. Fitz and C. H. Wong, *Chem. Rev.*, 1996, **96**, 443–474.





- 23 A. Farrán, C. Cai, M. Sandoval, Y. Xu, J. Liu, M. J. Hernáiz and R. J. Linhardt, *Chem. Rev.*, 2015, **115**, 6811–6853.
- 24 (a) Y. Chen, Y. Li, H. Yu, G. Sugiarto, V. Thon, J. Hwang, L. Ding, L. Hie and X. Chen, *Angew. Chem., Int. Ed.*, 2013, **52**, 11852–11856; (b) L. Li, Y. Liu, C. Ma, J. Qu, A. D. Calderon, B. Wu, N. Wei, X. Wang, Y. Guo, Z. Xiao, J. Song, G. Sugiarto, Y. Li, H. Yu, X. Chen and P. G. Wang, *Chem. Sci.*, 2015, **6**, 5652–5661; (c) Y. Xu, C. Cai, K. Chandarajoti, P. H. Hsieh, L. Li, T. Q. Pham, E. M. Sparkenbaugh, J. Sheng, N. S. Key, R. Pawlinski, E. N. Harris, R. J. Linhardt and J. Liu, *Nat. Chem. Biol.*, 2014, **10**, 248–250; (d) R. J. Linhardt and J. Liu, *Curr. Opin. Pharmacol.*, 2012, **12**, 217–219; (e) Y. Xu, S. Masuko, M. Takiuddin, H. Xu, R. Liu, J. Jing, S. A. Mousa, R. J. Linhardt and J. Liu, *Science*, 2011, **334**, 498–501; (f) R. Liu, Y. Xu, M. Chen, M. Weïwer, X. Zhou, A. S. Bridges, P. L. DeAngelis, Q. Zhang, R. J. Linhardt and J. Liu, *J. Biol. Chem.*, 2010, **285**, 34240–34249.
- 25 J. Sheng, Y. Xu, S. B. Dulaney, X. Huang and J. Liu, *J. Biol. Chem.*, 2012, **287**, 20996–21002.
- 26 X. Zhang, Y. Xu, P. H. Hsieh, J. Liu, L. Lin, E. P. Schmidt and R. J. Linhardt, *Org. Biomol. Chem.*, 2017, **15**, 1222–1227.
- 27 H. H. Hwang and D. Y. Lee, *Macromol. Res.*, 2016, **24**, 767–772.
- 28 (a) A. Borrmann and J. C. M. van Hest, *Chem. Sci.*, 2014, **5**, 2123–2134; (b) R. Adamo, A. Nilo, B. Castagner, O. Boutoureira, F. Berti and G. J. L. Bernardes, *Chem. Sci.*, 2013, **4**, 2995–3008.
- 29 (a) S. Li, L. Wang, F. Yu, Z. Zhu, D. Shobaki, H. Chen, W. Wang, J. Wang, G. Qin, U. J. Erasquin, L. Ren, Y. Wang and C. Cai, *Chem. Sci.*, 2017, **8**, 2107–2114; (b) Q. Hu, M. Allan, R. Adamo, D. Quinn, H. Zhai, G. Wu, K. Clark, J. Zhou, S. Ortiz, B. Wang, E. Danieli, S. Crotti, S. Crotti, M. Tontini, G. Brogioni and F. Berti, *Chem. Sci.*, 2013, **4**, 3827–3832.
- 30 (a) J. Zhao, X. Liu, A. Malhotra, Q. Li, F. Zhang and R. J. Linhardt, *Anal. Biochem.*, 2017, **526**, 39–42; (b) C. H. Grün, S. J. van Vliet, W. E. C. M. Schiphorst, C. M. C. Bank, S. Meyer, I. van Die and Y. van Kooyk, *Anal. Biochem.*, 2006, **354**, 54–63.
- 31 (a) J. Yang, P. H. Hsieh, X. Liu, W. Zhou, X. Zhang, J. Zhao, Y. Xu, F. Zhang, R. J. Linhardt and J. Liu, *Chem. Commun.*, 2017, **53**, 1743–1746; (b) J. Pai, J. Y. Hyun, J. Jeong, S. Loh, E. Cho, Y. Kang and I. Shin, *Chem. Sci.*, 2016, **7**, 2084–2093.
- 32 D. Shukla, J. Liu, P. Blaiklock, N. W. Shworak, X. Bai, J. D. Esko, G. H. Cohen, R. J. Eisenberg, R. D. Rosenberg and P. G. Spear, *Cell*, 1999, **99**, 13–22.
- 33 B. E. Thacker, E. Seamen, R. Lawrence, M. W. Parker, Y. Xu, J. Liu, C. W. Vander Kooi and J. D. Esko, *ACS Chem. Biol.*, 2016, **11**, 971–980.
- 34 A. I. de Agostini, J. C. Dong, C. de Vantéry Arrighi, M. A. Ramus, I. Dentand-Quadri, S. Thalmann, P. Ventura, V. Ibecheole, F. Monge, A. M. Fischer, S. HajMohammadi, N. W. Shworak, L. Zhang, Z. Zhang and R. J. Linhardt, *J. Biol. Chem.*, 2008, **283**, 28115–28124.
- 35 U. Lindahl, G. Backstrom, L. Thunberg and I. G. Leder, *Proc. Natl. Acad. Sci. U. S. A.*, 1980, **77**, 6551–6555.

