Synthesis of the highly branched hexasaccharide core of chlorella virus *N*-linked glycans

Sicheng Lin and Todd L. Lowary*[a]

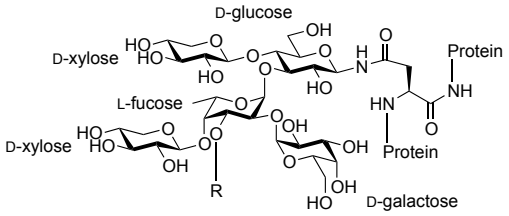
Abstract:

Chlorella viruses produce *N*-linked glycoproteins with carbohydrate moieties that differ in structure from all other *N*-linked glycans. In addition, unlike most viruses, these organisms do not hijack the biosynthetic machinery of the host to make glycocoproteins; instead, they produce their own carbohydrate-processing enzymes. A better understanding of the function and assembly of these fascinating and structurally-unprecedented glycans requires access to probe molecules. We describe here the first synthesis of a chlorella virus *N*-linked glycan, a highly branched hexasaccharide that contains the pentasaccharide present in all of the >15 structures reported to date. The target molecule includes a glucosyl-asparagine linkage and a 'hyperbranched' fucose residue in which all of the hydroxyl groups are glycosylated. Both convergent and linear approaches were investigated with the latter being successful in providing the target in 16 steps and 13% overall yield.

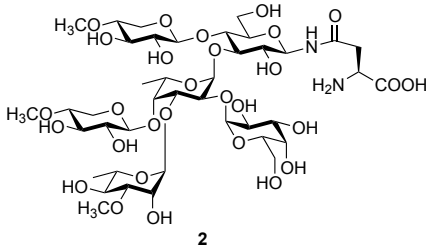
N-Glycosylation of proteins is a ubiquitous post-translational modification found in bacteria, viruses, plants, archaea and mammals.[1] The presence of *N*-linked glycans influences the folding, stability and bioactivity of proteins and, as such, tremendous efforts have focused on understanding, at the molecular-level, these roles. This, in turn, has led to the development of many chemical and enzymatic approaches for synthesizing *N*-linked glycan probes, either as the carbohydrate moieties themselves or as part of intact glycoproteins.[2] To date, the focus of these studies has been primarily on mammalian *N*-linked glycans, which all contain an *N*-acetyl-glucosamine as the sugar attached the protein asparagine residue.[2] Because most viruses use host biosynthetic machinery to make their *N*-linked glycans, such structures are also present in viral glycoproteins. Methods to synthesize *N*-linked glycans from bacteria and archaea, which also contain a 2-acetamidosugar (bacillosamine) linked to asparagine, have also been developed.[3]

Recent reports have detailed the structures of *N*-linked glycans from chlorella viruses.[4] These viruses normally infect algae, but a recent investigation has shown that humans can also harbor these organisms.[5] In the same study, a link between chlorella virus infection and reduced cognitive function in mice was reported. Chlorella *N*-linked glycans and their biosynthesis are remarkable, in that: 1) they the bear no structural resemblance to other *N*-linked glycans reported previously and 2) chlorella viruses encode their own carbohydrate-processing enzymes and hence do not rely on host biosynthetic pathways for their assembly. Understanding the function and biosynthesis of these species requires access to appropriate probe molecules. However, to date, there have been no reports of their synthesis by either chemical or enzymatic means.

The structure of chlorella virus *N*-linked glycans is species specific, but they all share a common pentasaccharide consisting of D-glucose (as the monosaccharide linking the glycan to the protein), L-fucose, D-galactose and two D-xylose moieties (**1**, Figure 1).[4] This motif is further elaborated with either a D-rhamnose or L-rhamnose forming a semi-conserved hexasaccharide core structure. The simplest of the structures reported to date, from *Acanthocystis turfacea* chlorella virus 1 (ATCV-1, **2**), consists of a hexasaccharide (with L-rhamnose) further elaborated with three methyl groups.[4b] We describe here the first synthesis of chlorella virus *N*-linked glycans, using **2** as a target.



1 R = D-rhamnose or L-rhamnose



2

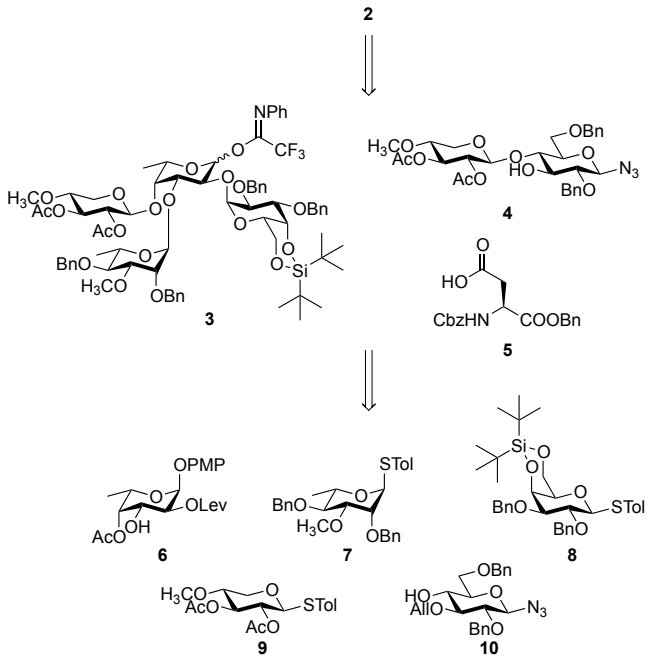
Figure 1. Conserved pentasaccharide of chlorella virus *N*-glycans and the semi-conserved D- or L-rhamnose moiety (**1**) and chlorovirus *N*-glycan from strain ATCV-1 (**2**) used as a synthetic target.

In developing a route to **2**, we envisioned that the primary challenge would be preparation of the 'hyperbranched' fucose residue in which every hydroxyl group is glycosylated. Such heavily branched carbohydrate residues are rare in nature and have thus not been the subject of significant synthetic

[a] S. Lin, Prof. Dr. T. L. Lowary
Alberta Glycomics Centre and Department of Chemistry
The University of Alberta
Edmonton, Alberta Canada T6G 2G2
E-mail: tlowary@ualberta.ca

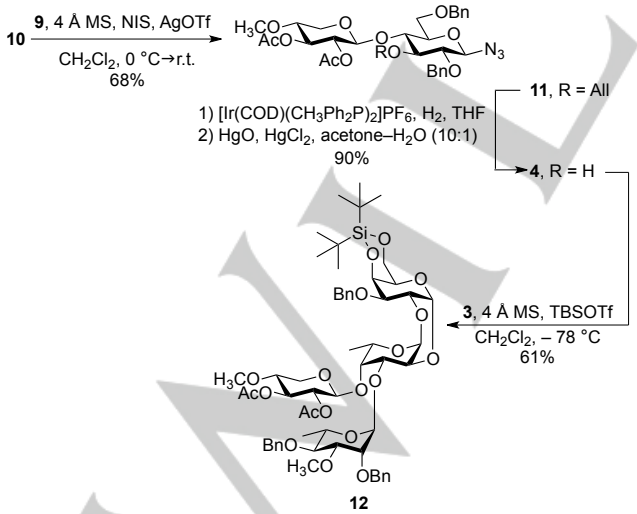
Supporting information for this article is given via a link at the end of the document.

investigations.[6] The first approach we studied was a convergent 4+2 strategy (Scheme 1), in which tetrasaccharide donor **3**



Scheme 1. Retrosynthetic analysis of hexasaccharide **2** via a convergent strategy.

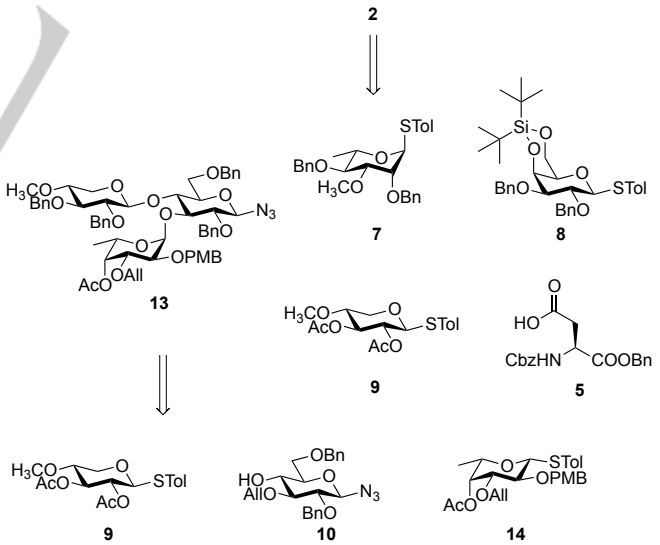
would be coupled with disaccharide acceptor **4**. If successful, this would provide a hexasaccharide that could be elaborated, via azide reduction, coupling with amino acid **5** and deprotection to the glycosyl amino acid **2**. As outlined in the Supporting Information (Scheme S6), it was possible to prepare **3** using four monosaccharides (**6**–**9**).



Scheme 2. Attempted synthesis of hexasaccharide moiety of **2** using a convergent 4+2 approach.

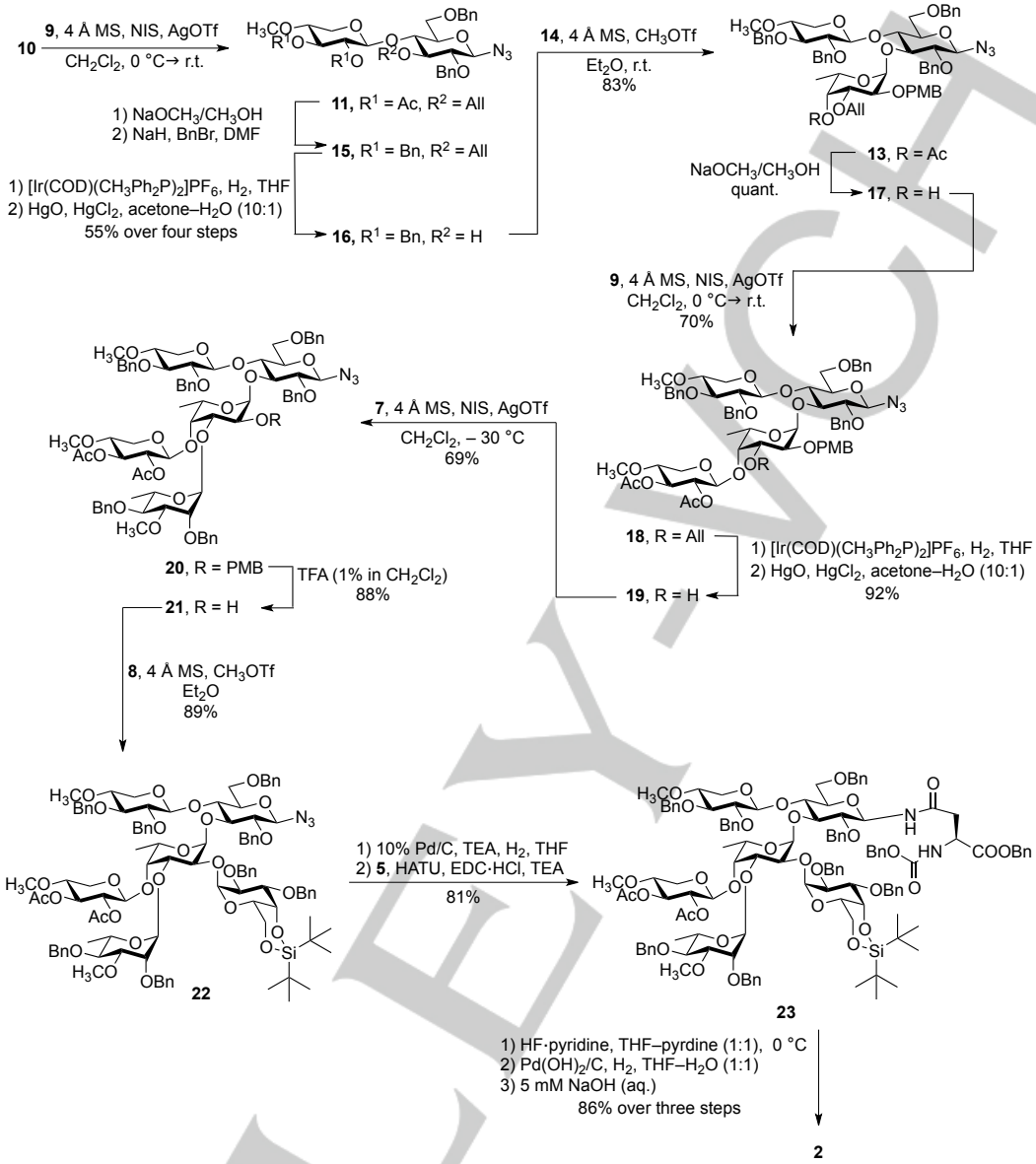
With tetrasaccharide **3** in hand, we proceeded to synthesize disaccharide **4** as illustrated in Scheme 2. An NIS/AgOTf- promoted glycosylation of glycosyl acceptor **10** (Scheme S4) with thioglycoside **9** (Scheme S3) led to the expected disaccharide **11**, which was then subjected to cleavage of the allyl ether affording **4** in 61% yield over two steps. Our first attempt to couple **3** and **4** relied on activation of the *N*-phenyl-trifluoroacetimidate with TBSOTf at –78 °C. Under these conditions, none of the desired hexasaccharide was observed. Instead, anhydride **12** was obtained in 61% yield, together with 15% of the hydrolyzed imidate. We postulated that the failure of the 4+2 glycosylation was due to steric congestion near the alcohol moiety in **3**. Thus, upon activation of the imidate, O-2 of the galactose residue acts as a nucleophile, forming an oxonium ion intermediate, that affords **12** upon loss of a benzyl group.[7] We also explored a 4+1 glycosylation between **3** and a suitably protected glucose derivative (Scheme S7), but was unsuccessful in obtaining the desired glycoside product; instead, **12** was formed as the major product. An inverse glycosylation[8] between **3** and **4** was also attempted, but **12** was again the dominant product.

The failure of a convergent approach prompted us to investigate a linear synthetic strategy to access this highly substituted glycan. To avoid the formation of **12**, we hypothesized that trisaccharide **13**, containing a fucose residue with three orthogonal protecting groups should be first synthesized from **9**, **10** and **14**. From **13**, the elaboration to the target could be achieved by sequential addition of monosaccharide residues **7**–**9** (Scheme 3). Azide reduction, coupling with **5**, and deprotection would provide the target.



Scheme 3. Retrosynthetic analysis of hexasaccharide **2** via a linear strategy.

With this plan in mind, the synthesis of **2** (Scheme 4) started with the glycosylation of glycosyl acceptor **10** with donor **9**. Subsequent Zemplén deacetylation and benzylation, followed



Scheme 4. Synthesis of **2** through a linear approach.

by deprotection of allyl group using $[Ir(COD)(CH_3Ph_2P)]PF_6$[9] and hydrolytic cleavage of the resulting vinyl ether generated the desired disaccharide acceptor **16** in 55% yield over four steps. The purpose of converting the acetyl groups on the xylose moiety to benzyl groups was to allow selective deacylation of the acetate group on the fucose moiety in **13** (see below). It should be noted that we initially tried to avoid this functional group exchange through the use of a levulinate ester to protect O-4 of the fucose residue; however, that approach was unsuccessful.

Conversion of **16** into trisaccharide **13** was achieved by a MeOTf-promoted glycosylation with **14**. This reaction proceeded in 83% yield and with excellent α-selectivity. On the other hand, the use of NIS/AgOTf to promote this reaction led to the rapid formation of an *N*-glycosyl succinimide byproduct[10] **S20** again pointing to the low reactivity of the C-3 hydroxyl group in the glucose residue in this disaccharide. The acetyl group on O-4 in **14** may help provide α-selectivity during the glycosylation via remote participation as proposed previously.[11]

Having a robust route to trisaccharide **13** in place, we turned our attention to investigating the appropriate sequence for glycosylating the fucose moiety. In a series of model studies (not shown) we investigated an approach in which the rhamnose

residue was added to O-3 of the fucose, followed by the introduction of xylose on O-4. Although this approach had worked in the synthesis of tetrasaccharide **3** (Scheme S6), it was unsuccessful using **13** as a precursor. This indicated the need to install the xylose before the rhamnose. Thus, the acetyl group in **13** was removed under Zemplén conditions and the resulting alcohol was glycosylated by an NIS/AgOTf-promoted reaction with **9** to give tetrasaccharide **18** in 70% yield over two steps. Moving on from **18**, the allyl ether was first converted into the corresponding vinyl ether using $[Ir(COD)(CH_3Ph_2P)]PF_6$, then subsequent hydrolysis using HgO and $HgCl_2$ in wet acetone provided the desired tetrasaccharide acceptor **19** in 92% yield. This compound was further glycosylated with **7** under NIS/AgOTf-promoted conditions at –30 °C to generate pentasaccharide **20** in 69% yield. Treatment of **20** with 1% TFA in dichloromethane resulted in the removal of the PMB group in 88% yield generating **21**. Glycosylation of **21** with **8**[12] gave the desired hexasaccharide **22** with exclusive α-selectivity[13] in 89% yield. The success of this 'counter-clockwise' approach to introducing the substituents onto the fucose residue can be rationalized by the need to glycosylate the least reactive (axial) alcohol first, followed by OH-3 and finally OH-2. Adding the last two carbohydrate residues onto **18** in the opposite order could be expected to fail in that glycosylation of OH-2 and deprotection of the allyl group would provide an alcohol that would be very sterically-hindered by two flanking carbohydrate residues.

The final key step in the synthesis was introduction of the amino acid. Palladium-catalyzed hydrogenation of **22** was carried out in the presence of a stoichiometric amount of triethylamine. The addition of this quantity of triethylamine was found to be crucial in preventing debenzylation and minimizing anomerization of the product. The resulting glycosyl amine was then coupled with protected amino acid **5** to afford the fully protected hexasaccharide **23** in 81% yield (two steps).

To accomplish the deprotection of **23**, the di-*tert*-butylsilylene group was first cleaved using HF·pyridine at 0 °C. Debenzylation was carried out using 20% palladium hydroxide on carbon in THF–H_2O (1:1). The use of methanol instead of water resulted in methylation of the amino group through reductive amination of trace amounts of formaldehye present in solution.[14] Finally, the acetyl groups were removed, and the methyl ester saponified, with 5% aqueous NaOH to generate the desired glycan **2** in 86% yield over three steps. Overall, **2** was synthesized using this linear synthesis in 16 steps and in 13% overall yield from **10**. A comparison between the NMR data of **2**, with that previously reported for the natural product, confirmed the structure (Figure S1).

In summary, we describe here the first synthesis of chlorella virus *N*-linked glycans, which are characterized by an unprecedented structure containing a glucosyl-asparagine residue and a fucose residue in which all three of the hydroxyl groups are glycosylated. Attempts to obtain the hexasaccharide using a convergent 4+2 strategy failed, but a linear approach was successful. A key feature of this route was the 'counterclockwise' introduction of the carbohydrate residues onto the fucose moiety. This is likely to be a general approach required to access any molecules of this type. In that regard, the synthesis of additional, more complex, derivatives of these structurally intriguing molecules is on-going. In addition, the use of **2** and related fragments in biosynthetic and immunological studies is in progress.

Acknowledgements

This work was supported by the Alberta Glycomics Centre, the Canadian Glycomics Network and the Natural Sciences and Engineering Research Council of Canada. SL thanks Alberta Innovates for a studentship award. We thank Dr. Christina De Castro (University of Naples "Federico II") for the NMR data for the natural glycoprotein for purposes of comparing the data for **2**.

Keywords: *N*-linked glycans • total synthesis • chlorella virus • carbohydrates • glycoproteins

[1] a) F. Schwarz, M. Aebi, *Curr. Opin. Struct. Biol.* **2011**, *21*, 576–582; b) D. Calo, L. Kaminski, J. Eichler, *Glycobiology* **2010**, *20*, 1065–1076; c) R. Strasser, *Glycobiology* **2016**, *26*, 926–939; d) L. Bagdonaite, H. H. Wandall, *Glycobiology* **2018**, *28*, 443–467; e) H. Nothaft, C. M. Szymanski, *Nat. Rev. Microbiol.* **2010**, *8*, 765–778.

[2] a) Z. Wang, Z. S. Chinoy, S. G. Ambre, W. Peng, R. McBride, R. P. de Vries, J. Glushka, J. C. Paulson, G.-J. Boons, *Science* **2013**, *341*, 379–383; b) A. J. Fairbanks, *Beilstein J. Org. Chem*. **2018**, *14*, 416–429; c) N. Wang, S.-T. Li, T.-T. Lu, H. Nakanishi, X.-D. Gao, *Chin. Chem. Lett.* **2018**, *29*, 35–39; d) C. Unverzagt, Y. Kajihara, *Chem. Soc. Rev*. **2013**, *42*, 4408–4420.

[3] a) A. Ishiwata, S. Ohta, Y. Ito, *Carbohydr. Res*. **2006**, *341*, 1557–1573; b) M. N. Amin, A. Ishitawa, Y. Ito, *Tetrahedron* **2007**, *63*, 8181–8198; c) Y. J. Lee, A. Ishiwata, Y. Ito, *Tetrahedron* **2009**, *65*, 6310–6319.

[4] a) C. De Castro, A. Molinaro, F. Piacnte, J. R. Gurnon, L. Sturiale, A. Palmigiano, R. Lanzetta, M. Parrilli, D. Garozzo, M. G. Tonetti, J. L. Van Etten, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 13956–13960; b) C. De Castro, I. Speciale, G. Duncan, D. D. Dunigan, I. Agarkova, R. Lanzetta, L. Sturiale, A. Palmigiano, D. Garozzo, A. Molinaro, M. Tonetti, J. L. Van Etten, *Angew. Chem. Int. Ed*. **2016**, *55*, 654–658; c) J. L. Van Etten, I. Agarkova, D. D. Dugigan, M. Tonetti, C. De Castro, G. A. Duncan, *Viruses* **2017**, *9*, 88–110.

[5] R. H. Yolken, L. Jones-Brando, D. D. Dunigan, G. Kannan, F. Dickerson, E. Severance, S. Sabunciyan, C. Conover Talbot Jr., E. Prandovszky, J. R. Gurnon, I. V. Agarkova, F. Leister, K. L. Gressitt, O. Chen, B. Deuber, F. Ma, M. V. Pletnikov, J. L. Van Etten, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 16106–16111.

[6] For representative examples see: a) W. Zhao, F. Kong, *Carbohydr. Res*. **2005**, *340*, 1673–1681; b) L. Guazzelli, O. McCabe, S. Oscarson, *Carbohydr. Res*. **2016**, *433*, 5-13; c) A.-L. Chauvin, S. A. Nepogodiev, R. A. Field, *J. Org. Chem*. **2005**, *70*, 960–966; d) M. Mönnich, S. Eller, T. Karagiannis, L. Perkams, T. Luber, D. Ott, M. Niemietz, J. Hoffman, J. Walcher, L. Berger, M. Pischl, M. Weishaupt, C. Wirkner, R. G. Lichtenstein, C. Unverzagt, *Angew. Chem. Int. Ed.* **2016**, *55*, 10487–10492.

[7] a) E. P. Dubois, A. Neszmélyi, H. Lotter, V. Pozsgay, *Tetrahedron Lett.* **1996**, *37*, 3627–3630; b) V. Pozsgay, E. P. Dubois, H. Lotter, A. Neszmélyi, *Carbohydr. Res*. **1997**, *303*, 165–173; c) M. Ludewig, D. Lazarević, J. Kopf, J. Theim, J. *Chem. Soc., Perkin Trans. 1*, **1998**, 1751–1752; d) K.-H. Jung, M. Müller, R. R. Schmidt, *Chem. Rev*. **2000**, *100*, 4423–4442.

[8] a) R. R. Schimidt, A. Toepfer, *Tetrahedron Lett*. **1991**, *32*, 3353–3356; b) P.-C. Gao, S.-Y. Zhou, H. Cao, J.-S. Yang, *J. Am. Chem. Soc*. **2016**, *138*, 1684–1688.

[9] K. Sahloul, T. L. Lowary, *J. Org. Chem*. **2015**, *80*, 11417–11434.

[10] L. Liao, F.-I. Auzanneau, *J. Org. Chem*. **2005**, *70*, 6265–6273.

[11] a) M. Dejter-Juszynski, H. M. Flowers, *Carbohydr. Res*. **1972**, *23*, 41–45; b) H. J. Vermeer, C. M. Van Dijk, J. P. Kamerling, J. F. G. Vliegenthart, *Eur. J. Org. Chem*. **2001**, 193–203.

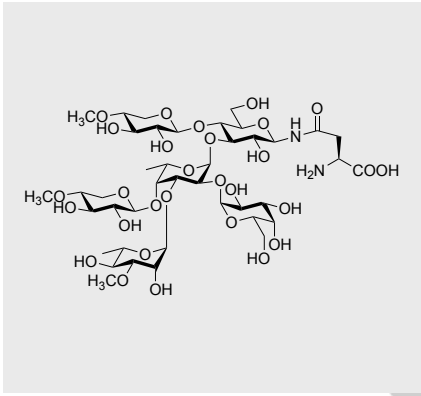
[12] M. Shiozaki, T. Tashiro, H. Koshino, T. Shigeura, H. Watarai, M. Taniguchi, K. Mori, *Carbohydr. Res*. **2013**, *370*, 46–66.

[13] A. Imamura, N. Matsuzawa, S. Sakai, T. Udagawa, S. Nakashima, H. Ando, H. Ishida, M. Kiso, *J. Org. Chem*. **2016**, *81*, 9086–9104.

[14] a) F. Filira, L. Biondi, M. Gobbo, R. Rocchi, *Tetrahedron Lett*. **1991**, *32*, 7463–7464; b) J.-P. Mazaleyrat, J. Xie, M. Wakselman, *Tetrahedron Lett.* **1992**, *33*, 4301–4302.

COMMUNICATION

A linear approach yields the first total synthesis of an *N*-linked glycan from chlorella virus glycoproteins. Probes of this type will be indispensable tools in understanding the function and biosynthesis of these structurally fascinating molecules.



*Sicheng Lin, Todd L. Lowary**

Page No. – Page No.

Synthesis of the highly branched hexasaccharide core of chlorella virus *N*-linked glycans