ChemBioChem
Concept
doi.org/10.1002/cbic.202200607
Chemistry Europe
European Chemical Societies Publishing
www.chembiochem.org

VIP *Very Important Paper*

Enabling Technologies in Carbohydrate Chemistry: Automated Glycan Assembly, Flow Chemistry and Data Science

Mei-Huei Lin[+],[a, b] Jakob B. Wolf[+],[a, b] Eric T. Sletten,[a] Dario Cambié,[a] José Danglad-Flores,[a] and Peter H. Seeberger*[a, b]

The synthesis of defined oligosaccharides is a complex task. Several enabling technologies have been introduced in the last two decades to facilitate synthetic access to these valuable biomolecules. In this concept, we describe the technological solutions that have advanced glycochemistry using automated glycan assembly, flow chemistry and data science as examples. We highlight how the synergies between these different technologies can further advance the field, with progress toward the realization of a self-driving lab for glycan synthesis.

Introduction

Carbohydrates are ubiquitous and essential biomolecules. While their role as an energy source and structural material is well established, the study of the biological activities of glycans has historically been lagging.[1] The main reason for the delay in the progress of glycosciences has been the structural complexity of glycans. Polysaccharides are often branched and, compared to linear biopolymers such as polypeptides and polynucleotides, have a wider variety of monomers. Additionally, each glycosidic linkage is a stereogenic centre, further complicating the access to defined structures, which are essential tools for the glycosciences.

Several technological solutions have been developed to offset the many challenges present in glycan synthesis, encompassing both reaction automation platforms and digital tools. In this concept, we will focus on three areas: automated glycan assembly, flow chemistry and digital tools. Together, they have enabled the synthesis of glycans of unprecedented complexity. We discuss the remaining bottlenecks to be addressed in the field and provide an outlook on the future of glycan synthesis in the age of digitalization.

[a] *M.-H. Lin,[+] J. B. Wolf,[+] Dr. E. T. Sletten, Dr. D. Cambié, Dr. J. Danglad-Flores, Prof. P. H. Seeberger*
Department of Biomolecular Systems
Max-Planck Institute of Colloids and Interfaces
Am Mühlenberg 1, 14476 Potsdam (Germany)
E-mail: Peter.Seeberger@mpikg.mpg.de
[b] *M.-H. Lin,[+] J. B. Wolf,[+] Prof. P. H. Seeberger*
Department of Chemistry and Biochemistry
Freie Universität Berlin
Arnimallee 22, 14195 Berlin (Germany)
[+] *These authors contributed equally to this work.*

The Automated Assembly of Glycans

Glycosylations are mechanistically complicated reactions resulting in glycan synthesis being a challenging, time-consuming, and labour-intensive task. To overcome these issues, several automated platforms have been developed for the rapid construction of complex glycans, while simultaneously removing human intervention to increase reproducibility, precision, and safety.

Automated Glycan Assembly (AGA), is a powerful solid phase-based technology for the construction of homogeneous oligosaccharides from reducing end to non-reducing end. Due to the growing oligosaccharide chain being resin-bound, tedious purifications can be bypassed and a multitude of reactions can be performed in sequence in a fully automated way, including the orthogonal removal of several temporary protecting groups and post-synthetic modifications.[2]

Recently, an automated solution-phase device for the multiplicative synthesis of polysaccharides has been developed, taking advantage of a preactivation strategy for the fashioning of glycans from the non-reducing to the reducing end. Due to the preservation of the leaving group on the reducing end, these oligosaccharide fragments could be progressively coupled for the assembly of long linear polysaccharides. Compared to AGA, this system allows for on-line monitoring and increased scale, but still requires extensive manual intervention.[3]

Other platforms have been developed, including a low-cost modified HPLC flow platform,[4] an enzymatic system,[5] a fluorous-tag system[6] and a glycopeptide chemoenzymatic system.[7]

Solid-phase automated glycan assembly

In 2001, the Seeberger laboratory developed the first AGA platform. Over the past two decades, solid phase-based automated glycan synthesizers have served as an enabling technology to provide access to homogeneous glycans.[8] With

one of the first iterations of this instrument, several complex glycans were synthesized, including a phytoalexin elicitor β-glucan dodecamer.[9] Since then, solid-phase AGA has been applied to the synthesis of glycans for glycan arrays,[10] vaccine development,[9b] carbohydrate standards,[11] enzymatic assays,[12] single molecular imaging[13] etc. With this platform, the longest fully automated synthetic polysaccharide was constructed, bearing 100-mer polymannoside in 188 hours.[14]

The solid phase AGA workflow (Figure 1) commences with an acidic wash (step 2) that removes any residual base, and then the resin-bound acceptor is glycosylated (step 3) by an electrophilic donor (building block) in the presence of an activator. The unreacted nucleophiles potentially present are "capped" (step 4) by acetylation and finally, a temporary protecting group (*t*PG) at the intended place of elongation is selectively removed (step 5). The cycle is then repeated with the next desired monosaccharide.

During the glycosylation reaction, several glycosyl donor types have been utilized including glycosyl trichloroacetimidates, phosphodiesters, and thioethers.[9a] Typically, thioglycosides are preferred due to their stability and easy activation under several promoter systems including *N*-iodosuccinimide (NIS) in combination with trifluoromethanesulfonic acid (triflic acid or TfOH) or trimethylsilyl trifluoromethanesulfonate (TMSOTf). For typical glycosylations, the glycosyl donor is initially delivered and subsequently activated at a low temperature (T_1, between $-40\,^{\circ}C$ and $-20\,^{\circ}C$). The reaction temperature is then increased ($T_2 = -10\,^{\circ}C$ to $0\,^{\circ}C$) via a gradual ramp to complete the reaction.

The selection of protecting groups on the building blocks is crucial as they affect the reactivity and the temporary groups need to be placed at the desired points of elongation. In a glycosylation reaction, the protecting groups are known to shift the reaction along the S_N1/S_N2 mechanistic continuum, thus impacting the coupling yield and stereoselectivity. These outcomes can be tuned through the use of permanent protecting groups at positions of no elongation. Benzyl ethers are electron-donating inducing higher reactivity whereas benzoyl esters are electron-withdrawing and decrease the reactivity of the glycosyl donor. C2 benzoyl esters provide anchimeric assistance, ensuring a 1,2-*trans* stereochemical glycosylation outcome.

Temporary protecting groups that can be orthogonally removed are used to guarantee a regioselective outcome. The 9-fluorenylmethyloxycarbonyl carbonate (Fmoc) is one of the most commonly used *t*PGs in AGA, as it is easily removed under mildly basic conditions (e.g. with piperidine or triethylamine). To achieve further branched and complex glycans several other protecting groups have been incorporated, including levulinate ester (Lev), 2-methylnaphthyl ether (NAP), and 2-(azidomethyl)benzoyl (AZMB). Removal of these *t*PGs also provides places for further functionalization (such as sulfation[15] and phosphorylation[16]) upon completion of the glycan backbone. Still, manipulating the inductive effect of glycosyl donors relies on the expertise of the practitioners to select suitable building blocks during the synthesis of the target glycan.

Upon completion of the desired glycan sequence, photocleavage of the commonly utilized photolabile linker is carried out (Figure 1, step 6) to detach the desired glycan from the resin. Photolabile linkers are utilized to withstand the strongly acidic and basic conditions throughout the synthetic cycle. Following purification and global deprotection, characterization is applied to ensure homogeneous glycan with well-defined structure.

Microwave-assisted solid-phase automated glycan assembly

In a conventional AGA system, the glycosylation occurs at subzero temperature, while the deprotection of *t*PGs occurs at higher temperatures (up to room temperature). A dynamic temperature control system is necessary to achieve those temperatures, though this system is energy and time inefficient over a wide range of temperatures. Recently, the encasement of the jacketed reaction vessel by a microwave generator accelerated the overall cycle time and post-assembly functionalization, while expanding the temporary protecting group portfolio (Figure 2).[2] By combining microwave radiation and constant cooling, it enabled rapid temperature adjustments from $-40\,^{\circ}C$ to $100\,^{\circ}C$, shortening the time for temperature adjustment and

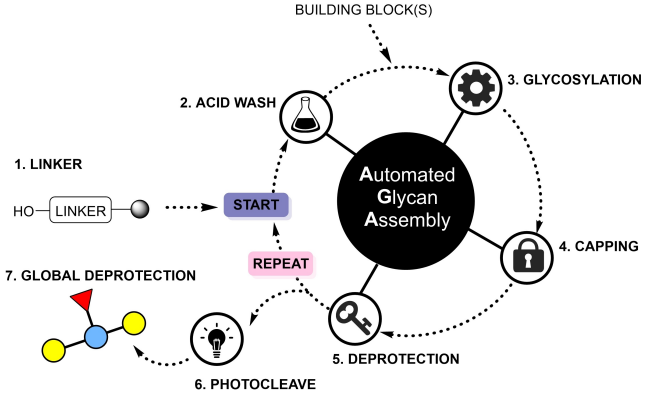


Figure 1. The automated glycan assembly workflow.[14]

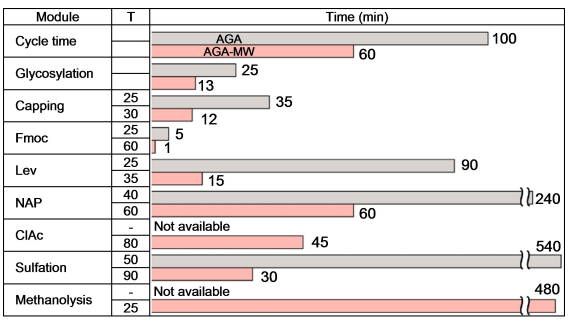


Figure 2. The current state of the art on AGA chemical operations in the AGA and AGA-MW systems.[2]

accelerating deprotection and functionalization. Naphthyl ether cleavage time was shortened from 240 to 60 min in the microwave-AGA (MW-AGA) system. Moreover, increasing the temperature range accelerated the sulfation on resin support at the suitable reaction temperature to 30 min, compared to previously 9 h.

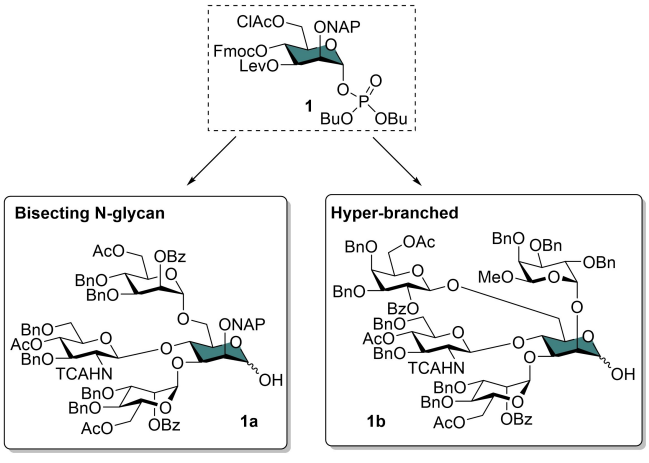
To demonstrate the implications of this work, chloroacetate (ClAc) ester was added as a new temporary protecting group to synthesize highly branched glycans (Scheme 1). Mannoside **1** was protected with four different orthogonal *t*PGs, C2-NAP, C3-Lev, C4-Fmoc, and C6-ClAc. The building block was a critical core intermediate in synthesizing bisecting *N*-glycan **1 a** (28% overall yield) and its hyper-branched derivative **1 b** (32% overall yield) under MW-AGA assistance.

Probing activation temperature of glycosyl building blocks

Glycosylations are highly sensitive to reaction temperature. Therefore, clearly defining the activation temperature of building blocks is essential.

A semi-automated assay to measure glycosyl donor activation temperature was devised.[17] A glycosyl donor in solvent (CH_2Cl_2) is first introduced to the reaction vessel and chilled to the set temperature. Once the desired temperature has been reached, the promoter solution is delivered and mixed with the donor for 5 min. Subsequently, a quench solution (10% pyridine in DMF) is added, and the crude solution is collected in a tube containing 10% aqueous sodium thiosulfate. Then, following the same process, the next temperature was screened with the same glycosyl donor. The temperature in the reaction vessel was monitored by a fibre optic probe and the quenched reaction mixture was subjected to a proton NMR to assess the degree of activation. From the NMR analysis was derived an activation temperature (T_A), this refers to the highest temperature at which the glycosyl donor can be preserved. In contrast, the decomposition temperature (T_D) is the lowest reaction temperature where complete consumption of glycosyl donor is observed. Following this protocol, the activation and decomposition temperatures of 20 building blocks were measured. The outcome of the temperature screening helped to improve the synthesis of a β-1,4-glucose tetramer by replacing the standard temperature ramp with isothermal conditions (−25 °C) resulting in high relative purity (89%).

Although the correlation between activation temperature and glycosyl donor can improve glycan assembly, more than 11 factors including environmental and coupling partners influence the glycosylation reaction.[18] Therefore, future studies will have to address more factors to optimize the conversion of the coupling step. Using machine learning and statistical analysis, the optimal synthesis rules in the AGA system are expected to be established algorithmically.[18–19] Thereby, the current AGA platform will be further improved while maintaining the broad range of operational conditions.



Scheme 1. Orthogonal mannoside phosphate **1** enables the synthesis of highly branched glycan precursors **1 a** and **1 b**.[2]

Flow Chemistry

Flow chemistry relies on the use of channels or tubing to conduct a reaction in a continuous stream rather than in a flask.[20] Compared to conventional reactions carried out in a flask, reactions performed in flow benefit from a larger surface area to volume ratio, thus a more accurate control over temperature. Moreover, the shorter distances involved result in fast mixing reducing concentration gradients, while continuous production (i.e. scaling out) is a trivial approach to reaction scale-up. The possibility of performing different reaction steps in sequence (i.e. to telescope) and the relatively low amount of chemicals reacting at any given time can result in improved safety. Finally, the possibility to control several reaction parameters such as reaction time and stoichiometry by simply changing the flow rate associated with different streams, makes flow chemistry ideal for automated systems. In this section, we illustrate how carbohydrate chemistry can benefit from flow chemistry.

Glycosylation

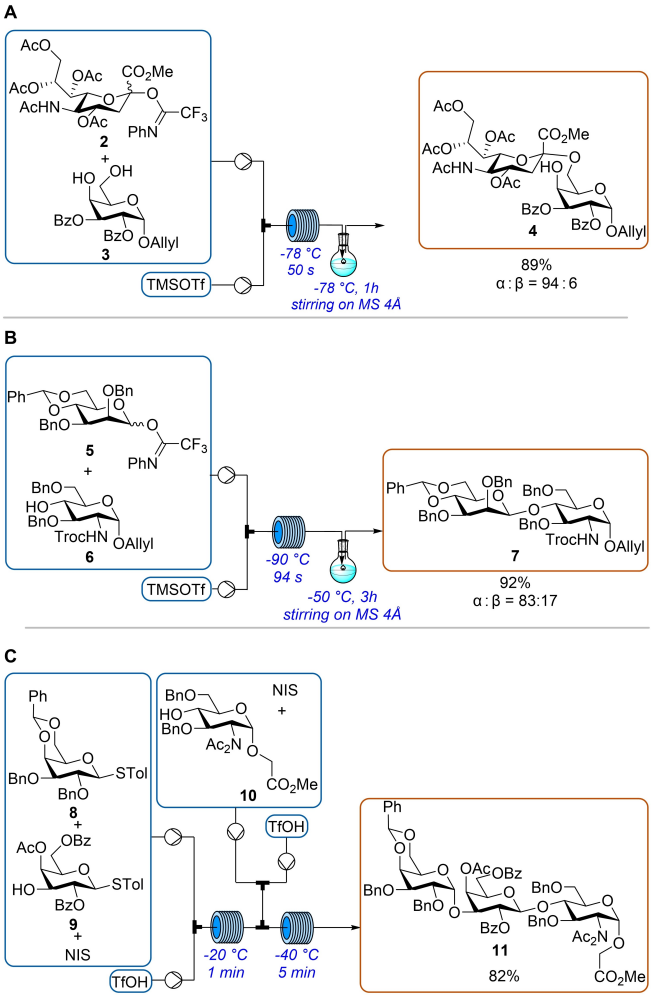
Glycosylations are particularly sensitive to several parameters that can shift the balance between their different simultaneous mechanistic pathways and highly reactive glycoside intermediates.[21] Among these are some environmental factors such as temperature and mixing that are poorly controlled in reactions performed in conventional round-bottom flasks. As a result, it is not always clear whether a poor reaction outcome is caused by environmental factors or by intrinsic stereoelectronic parameters associated with the coupling partners. The glycosylations involving the modified C5-trichloroacetamide sialic acid **2** are one example of such cases. This glycosyl donor was reported to give unreliable or mediocre yields and stereoselectivity, an observation that was initially attributed to its inherent reactivity,[22] or the formation of supramers.[23] Fukase

and co-workers, however, hypothesized poor mixing and lack of precise temperature control in the activator addition as reasons for the poor performance of this donor in conventional batch reactions.[24] By performing the glycosylation of **2** with the galactose acceptor **3** in flow with a Comet X-01 micromixer, they showed that the relatively small change in temperature between −78 °C and −60 °C has a significant detrimental impact on both yield and stereoselectivity. With optimized reaction conditions and a flow setup (Scheme 2A), the corresponding disaccharide **4** was obtained on gram-scale in 89% yield and good stereoselectivity ($\alpha:\beta = 94:6$) in only 30 min. Notably, the high flow rate (2.0 mL/min) ensured good mixing and combined with the fast glycosylation kinetics, resulted in high productivity despite the relatively small reactor volume (1.6 mL).

It is not uncommon in glycan synthesis to observe a significant deterioration in reaction yields upon scale-up. In these cases, the reliable scalability afforded by a flow reactor can help to easily translate optimized conditions on small scale to a larger reaction scale. Fukase and co-workers reported a flow-enabled scalability in the stereoselectively challenging β-mannosylation to obtain Manβ(1,4)GlcNAc disaccharide **7**.[25] The original reaction conditions involved the commercially unavailable TMSB$(C_6F_5)_4$ as the activator.[26] Screening commercially available Lewis acids revealed trimethylsilyl triflate to provide the β-isomer in an encouraging 84% on a 20 mg reaction scale. However, the yield decreased to 27% upon scale-up to 500 mg. By moving to flow and mixing a solution containing donor **5** and acceptor **6** with a solution containing TMSOTf in a micromixer, they found conditions yielding **7** in 92% yield and a $\beta:\alpha$-selectivity of 83:17 on a 1.3 g reaction scale. Mixing was performed at −90 °C with a further 90 s at −90 °C, after which the resulting reaction mixture was collected and stirred in a cooled flask at −50 °C for a further 3 h (Scheme 2B). Despite the lower β-selectivity than that observed with TMSB$(C_6F_5)_4$, the flow procedure compared positively for its scalability and the use of inexpensive and commercially available TMSOTf.

The possibility of directly using the outlet stream of one reactor to feed a further subsequent reaction, i.e. the telescoping of several reactions, is another potential advantage of flow chemistry. Since the process is somewhat akin to multistep one-pot synthesis, it is not surprising that it has been adopted for the iterative formation of glycosidic linkages. In particular, Tsutsui *et al.* pursued the synthesis of α-Gal, an antigenic trisaccharide in a fully telescoped fashion (Scheme 2C).[27] Sequential one-pot reactivity-based oligosaccharide synthesis of α-Gal gave scale-up problems, with yields of 80% on a 20 mg scale and 57% on a 120 mg scale, which was attributed to overreaction (i.e. the activation of the -STol leaving group of the disaccharide formed from **10** and **11**). Since reactions performed in flow benefit from precise control over reaction time, the authors surmised that it would have been possible to prevent the overreaction with better control over the reaction time for the first step. They implemented a flow synthesis by combining a stream with donor **8**, acceptor **9**, and NIS as activator with a stream of triflic acid and after a residence time of 1 min at −20 °C the target disaccharide was obtained in a 94% yield. The flow setup was then expanded for the telescoped production of α-Gal, mixing at −40 °C the disaccharide stream with a second stream containing the GlcN$(Ac)_2$ acceptor **10** and further activator. The resulting reaction mixture was then collected in a cooled flask. Stirring for 5 min at −40 °C yielded 82% of the protected trisaccharide **11** and, after global deprotection, 71% of α-Gal over two steps in a fully telescoped process (Scheme 2C).

Widely varying reactivities of glycosyl donors and acceptors, and the unique nature of each glycosylation linkage are such that no universal reaction conditions exist to reliably furnish the corresponding glycosylated product in good yields and predictable stereoselectivity. Often, particularly when less common coupling partners are used, extensive optimization of the glycosylation reaction conditions is needed. Microreactor technology for optimization is truly an enabling technology since a larger parameter space can be investigated by reducing the amount of material necessary for each reaction. At the same



Scheme 2. Flow chemistry-based approaches to glycosylation. A) α-sialylation of galactose acceptor **3** with sialic acid **2**; B) β-mannosylation of glucosamine derivative **6** with donor **5**; C) telescoped flow synthesis of protected α-Gal.

time, flow reactors are well suited to screen variables that are relevant to glycosylations, such as reaction time and temperature, without any setup reconfiguration.

The relevance of microreactors for the optimization of glycosylation reaction conditions has been investigated using a custom-made 78 μL silicon microreactor (Figure 3).[28] The reactor features four inlets and one outlet. Three inlets for activator, acceptor and donor are followed by mixing and reaction zones, terminated by the stream coming from the quencher inlet used to precisely control the reaction time before leaving the chip reactor. Initially, the microreactor was used to study the impact of reaction time and temperature on the mannosylation of mannoside acceptor **13** with a mannosyl trichloroacetimidate donor **12** to give the disaccharide **14** (Scheme 3A). The reaction outcome was evaluated by HPLC-DAD, requiring only 2 mg of donor per each reaction to determine relative yield and stereoselectivity. In this case, orthoester **15** was found to be the major reaction product at −70 °C. However, higher product selectivity was obtained at higher temperatures. Notably, for optimal selectivity towards **14**, 213 seconds of residence time were necessary at −60 °C while 26 seconds were sufficient at −35 °C, resulting in ten-fold increased productivity. The possibility of performing a large set of reactions at a precisely defined temperature enables to clearly define the trade-off between product yield and productivity.

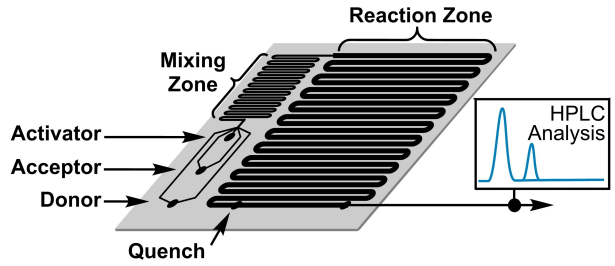
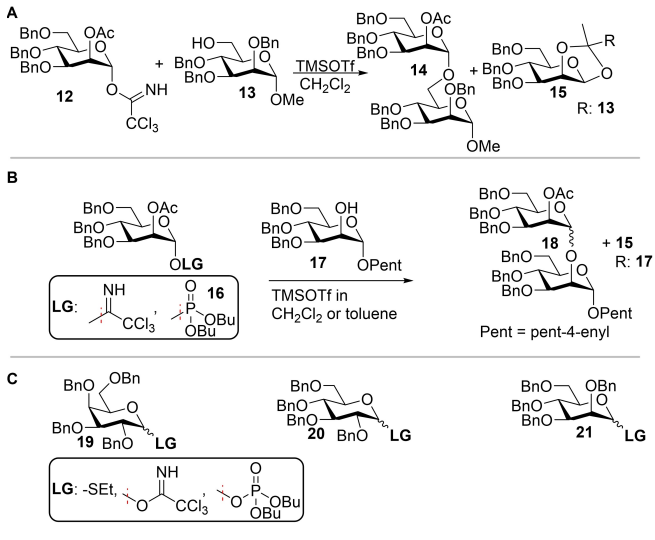


Figure 3. Schematic of the microreactor used to screen glycosylations. The silicon chip microreactor has three primary inlets, a mixing and reaction zone, a secondary inlet for the quenching line, and an outlet for collection and/or analysis by HPLC-DAD.[21]



Scheme 3. Microreactor-based screening of glycosylation conditions. A) glycosylation of mannoside **13** with trichloroacetimidate mannoside donor **12** with orthoester by-product **15**. B) glycosylation of mannoside **17** with trichloroacetimidate mannoside donor **12** or phosphate mannoside donor **16** with orthoester by-product **15**. C) Perbenzylated galactoside **19**, glucoside **20** and mannoside **21** thioether, trichloroacetimidate and phosphate donors. (LG = leaving group).

Expanding investigation in the custom-made microreactor to different solvents and different leaving groups (Scheme 3B), α-1,2-mannosylations with the acceptor **17** and two different mannosyl donors, **12** and **16**, were investigated.[29] By switching solvent from CH_2Cl_2 to toluene, the formation of the orthoester **15** could be almost completely suppressed, and the more reactive phosphate donor **16** gave results similar to the glycosyl imidate **12** but with shorter reaction times. Most notably, the optimized flow conditions directly translated to a batch process with similar results, proving the relevance of the optimization results obtained with the microreactor.

A further expansion of the chemical space under investigation with the same reactor was achieved by introducing in-line HPLC and a control PC to automate the screening, Comparing perbenzylated galactoside **19**, glucoside **20** and mannoside **21** building blocks (Scheme 3C) varying both continuous (temperature, acceptor stoichiometry and presence of water) and discrete (solvent, type and stereochemistry of leaving group, type of activator and acceptor nucleophilicity) parameters.[18] Temperature, solvent and activator had the most profound impact on the stereoselectivity in a series of reactions with simple alcohols as model acceptors. By solely changing the activator, solvent and temperature, the stereoselectivity of the reaction between the perbenzylated glucosyl α-trichloroacetimidate donor **20** and isopropanol could be completely reversed from an α:β-ratio of 9:1 to 1:11. The experiments were automatically performed with restocking solvents and reagents as only required human input.[21] The results were organized in a structured way and reused for statistical analysis. The findings will be discussed in-depth in the data science section.

Synthesis of Glycan Precursors

The limited availability of commercial building blocks further complicates the synthesis of complex glycans. This barrier, combined with advancements in automated glycan assembly, means that currently, the synthesis of differently protected monosaccharides is a bottleneck to oligosaccharide synthesis rather than the glycosylation itself. To ensure success in the multistep synthetic campaign, glycan precursors are generally synthesized in a succession of time-consuming synthetic steps leveraging well-established and robust transformations. However, the limited diversity of reactions employed in their synthesis makes them ideal targets for the application of flow chemistry and reaction automation. In this section, we will highlight some literature examples of the application of flow chemistry to glycan precursor synthesis.

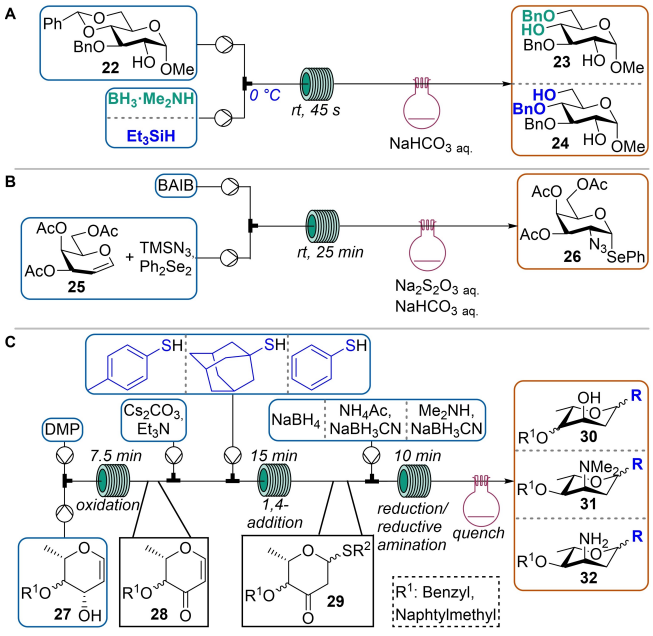
The reductive opening of glycosyl 4,6-*O*-benzylidene acetals such as **22** is a common and very useful reaction in carbohydrate chemistry, affording C4 or C6 unprotected hydroxyl groups depending on the reagents used (Scheme 4A, **23** and **24**). This characteristic can be leveraged for subsequent orthogonal protection or the direct use of the product as a glycosyl acceptor. However, performing this deprotection on large scale often yields significant amounts of 4,6-diols as side products due to the lack of precise temperature control and the formation of concentration gradients. By performing the reaction in flow on an array of 10 different substrates, Tanaka and Fukase obtained yields between 91% and 100%, which were consistently higher than those obtained in batch under otherwise identical conditions.[17] Thanks to the use of a micromixer, a reaction time of 45 seconds was sufficient to achieve full conversion, which translated to a productivity of 8.5 g/h. This result compares favourably with the batch procedure where the slow addition of acid at 0 °C is usually followed by continuous stirring at room temperature for several hours. Another common deprotection in carbohydrate chemistry is the cleavage of benzyl ethers. Recently, a visible-light-mediated debenzylation protocol that features shorter reaction times when performed in flow, thanks to the more efficient light penetration obtained in the microreactor system was reported.[30]

Galactosamine is a common monomer found in glycans. However, the availability of galactosamine-based building blocks for chemical glycosylation is hindered by the price of this monosaccharide. Methods installing a nitro-group at C2 in glycals offer a solution to this problem. The azidophenylselenylation (APS) of glycals is particularly attractive as it installs in a single step a phenylseleno leaving group on the anomeric carbon and a protected amine on C2. However, the highly exothermic nature of this transformation results in erratic yields in batch, aggravated by the safety concerns associated with the use of azides. Therefore, performing APS on large scale, while attractive in principle, is generally discouraged. To solve these problems, a flow procedure for APS at room temperature (Scheme 4B), relying on the improved heat dissipation in flow was developed. Safety concerns are reduced due to the low amounts of hazardous azide reacting at any point in time.[31] On top of a significant acceleration to the reaction (25 min in flow vs. 4 to 16 h in batch for the APS of **25**) and the elimination of cooling, the yield of **26** reported (79%) was better than the average yield obtained in batch (35%).

A synthesis of 2,6-dideoxysugars and 3-amino-2,3,6-trideoxysugars leveraged both continuous-flow and reaction automation.[32] Using commercially available starting materials afforded glycals *via* a telescoped acetylation and bromination followed by a reductive elimination with zinc. Based on these glycals, different telescoped flow procedures helped to install orthogonal protecting groups on C3 and C4. Selective deprotection to **27** and oxidation of C3-OH yielding **28**, followed by 1,4-addition (**29**) and reduction gave access to thioglycoside donors **30**. By adding dimethyl amine or ammonium acetate to the reductive mix after the 1,4-addition, the corresponding 3-amine derivatives could be obtained as well, either protected or unprotected (**31**, **32**) (Scheme 4C). The telescoped flow protocol was controlled by a python script that was made available.

The adoption of flow chemistry in glycan synthesis is still limited, despite several reports highlighting the potential advantages of microreactors. Concerning glycosylations, microreactors have been used as a screening platform,[18,28,29] for mechanistic studies[23] or to tackle challenging glycosidic linkages such as β-mannosylation[25] and α-sialylation.[24] For the synthesis of glycan precursors, flow chemistry has mostly been used to address scale-up[17] or safety issues.[31]

The application of automated flow platforms for the synthesis of glycan precursors, holds significant potential as a way to streamline the labour-intensive synthesis of complex building blocks. The labour-intensive nature of the multi-step synthesis of the glycosyl donors is ideally suited to automation given the relatively limited diversity of reactions commonly used and their general robustness.



Scheme 4. Flow synthesis of glycan precursors. A) reductive opening of glycosyl 4,6-*O*-benzylidene acetals, B) azidophenylselenylation of glycals, C) telescoped synthesis 2,6-dideoxysugars and 3-amino-2,3,6-trideoxysugars (BAIB = bisacetoxy iodobenzene, $TMSN_3$ = trimethylsilyl azide, DMP = Dess-Martin periodinane).

Data Science

The field of carbohydrate chemistry can be intimidating due to its complexity at different levels. In the last two decades, a series of digital tools have been developed to handle this complexity and empower the chemist to make informed choices in a time-efficient way. From the pioneering work of Wong and co-workers in 1999[33] to the more recent application of transfer learning to glycochemistry,[34] a lot of progress has been made. The uptake by practitioners of these tools is increasing as they prove progressively more valuable and as the

digitalization of chemistry advances. While carbohydrate chemists have been early adopters, in recent years several tools have emerged in the chemistry community at large that blurred the boundaries between chemical synthesis and data science.[35] In this section, we will describe some milestones in this area and highlight the potential synergies with both automated reaction systems and flow chemistry.

Relative reactivity values (RRV): a quantitative metric of glycosyl donor reactivity

One-pot sequential glycosylation is a convenient glycosylation strategy whose chemoselectivity is based on the different reactivities of the glycosyl donors used rather than on protective group chemistry.[36] While in principle appealing, the success of approaches based on one-pot sequential glycosylation relies heavily on the relative reactivity of different donors. Such reactivities are non-trivially influenced by many structural parameters, including the nature of the monosaccharide unit and of its protecting groups. Consequently, the design of synthetic routes based on this strategy becomes harder with larger oligosaccharide targets and is restricted to domain experts. To address this issue, Wong and co-workers introduced in 1999 the concept of "relative reactivity value" (RRV) to quantify the reactivity of a glycosyl donor via a competitive HPLC-based experiment (Figure 4).[33] An array of glycosyl donors were equipped with a *p*-methylphenyl thiol aglycon to serve as the standard leaving group because of its high UV absorbency, stability, and activation by several promoter systems. In an RRV experiment, one equivalent of each, the experimental donor and per-*O*-acetylated thiomannoside **33** (reference donor, RRV = 1), were mixed in CH_2Cl_2 at 0 °C. Subsequently, methanol was added as acceptor. The donors were then activated by one equivalent of NIS/TfOH. Due to the use of the promoter reagent as the limiting reagent, a mixture of methyl glycoside and the remaining *p*-methylphenyl thioglycoside is observed in HPLC traces. Then, following a first-order kinetic equation (Equation (1)):

$$\frac{RRV_x}{RRV_{ref}} = \frac{k_x}{k_{ref}} = \frac{\ln(A_x)_t - ln(A_x)_0}{ln(A_{ref})_t - ln(A_{ref})_0} \quad (1)$$

the decreasing proportion among two glycosyl donors was recorded and further converted to an RRV value. After a database of thioglycosides RRV was created, a digital tool named Optimer was developed to search the database and aid the chemist in the selection of glycosyl donor building blocks for the one-pot assembly of oligosaccharides. A significant limitation of Optimer software tool is the need for experimental values of RRV, because the target glycosyl donors have to be synthesized to measure their RRV, meaning that comparing competing synthetic strategies is impracticable when the relevant building blocks are not already part of the database. To address this issue, the Wong group later developed Auto-CHO[37] as a resource to predict relative reactivity values and suggest candidate donors for the synthesis of an oligosaccharide. A glycosyl donor library, composed of 154 experimentally validated building blocks and more than 50,000 virtual building blocks with predicted RRVs was compiled. For the prediction, the correlation between RRV and a large array of molecular descriptors was leveraged.

Acceptor nucleophilic constants (Aka): a quantitative metric of glycosyl acceptor reactivity

While the RRVs are designed to quantify the glycosyl donor reactivity,[38] the success of a glycosylation reaction depends on all reaction partners involved (i.e. including activator and acceptor) on top of the actual experimental parameters, as highlighted in the flow chemistry section.[39] In the attempt to better quantify the role of the nucleophilic acceptor on glycosylation yield and stereoselectivity, Codee and co-workers systematically studied three major effects: steric, inductive, and conformational.[40] However, all these factors are highly intertwined and cannot be studied individually. In an attempt to find a way to reduce steric, electronic and structural effects to a single metric of acceptor nucleophilicity, the acceptor nucleophilic constant (Aka) was introduced. Similarly, to RRVs, Akas

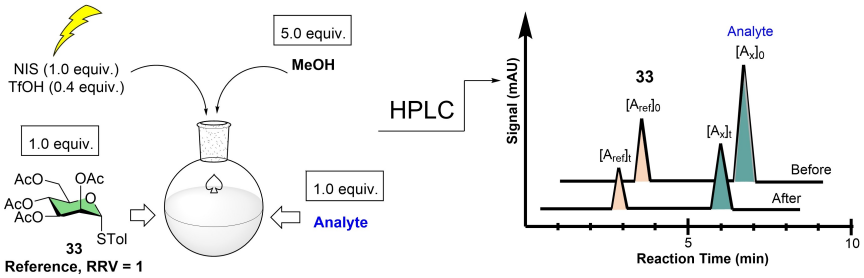


Figure 4. Schematic representation of the experimental procedure for RRV determination. An equimolar mixture of reference mannoside **33** and the donor whose RRV is to be measured (analyte) are activated with one equivalent of NIS and TfOH and reacted with methanol. The ratio of remaining unreacted donors (**33** and analyte) measured by HPLC is used to calculate their relative reactivity according to Equation (1).

were designed to provide a connection between experimental results and a computable property with the final aim of obtaining a software program (GlycoComputer) for the prediction of glycosylation reactions without recourse to computationally expensive ab-initio methods.[41] The determination of Aka is comparable to the process for RRV: in an HPLC-based competition reaction, 3,4-dihydropyran (DHP) served as model electrophile and two competing nucleophilic acceptors including one experimental nucleophile and 4-OH of galactoside **34** (reference nucleophile, Aka = 1) were used (see Figure 5). Based on the reaction outcome assessed from HPLC chromatograms, the Aka value can be calculated as described by Equation (2):

$$Aka = \frac{k_a}{k_{ref}} = \frac{\ln[A]_t - \ln[A]_0}{ln[A_{ref}]_t - ln[A_{ref}]_0} \quad (2)$$

Aka served as a descriptor for the statistical analysis of the outcome of a dataset of 230 glycosylations. In particular, a random forest machine learning algorithm (i.e. an ensemble of decision trees involving different descriptors) was used. The sum of donor-related effects (59%) was found to be more important than acceptor effects (24%) by including 20 different descriptors. Notably, by combining RRV and Aka, stereoselectivity prediction can be performed: the combination of coupling partner with high RRV and low Aka preferentially results in α-selectivity, while coupling of low RRV and high Aka prefers β-selectivity.

Machine-learning based glycosylation models

Language-inspired transformer neural networks that model chemical reactions based on text (e.g. SMILES) notation have recently been shown to be able of capturing chemical reactivity.[42] Since the regio- and stereoselectivity of glycosylations are hard to predict even for expert chemists,[39] the application of these models could be beneficial to the field. However, the accuracy is dependent on the dataset used to train it, and the only freely available dataset of chemical reactivity whose size is sufficient for the training of transformer models is currently based on patent literature.[43] Since the dimension of the carbohydrate chemistry field is not large enough to provide data to train a specialized model from scratch, Reymond and co-workers applied transfer learning to refine the parameters of a previously reported chemistry model (the sequence-2-sequence *Molecular Transformer* trained on the USPTO database) for applications in carbohydrate chemistry.[34] The original *Molecular Transformer* performed poorly in glycochemistry, with an accuracy of ~40%, mostly due to the low abundance of glycosylations in training data. Training the same architecture on a small set of 25k glycosylation reactions (extracted from the commercial database Reaxys) only achieves an accuracy of ~30%. However, fine-tuning the *Molecular Transformer* on 20k glycochemistry reactions, yielding a model called *Carbohydrate Transformer*, improved the accuracy to ~70%. The trained model was tested on syntheses absent from training data and predicted 77% of 13 synthetic steps in the preparation of a lipid-linked oligosaccharide correctly. In the synthesis of a trisaccharide with five challenging regioselective protections and four difficult regio- and stereoselective glycosylations, 68% of steps were correctly predicted, as opposed to only 39% by the *Molecular Transformer*. The authors linked most of the wrong predictions to shortcomings in the training data, e.g. absence of stoichiometry. This observation suggests that to enable better models of glycosylation, it is necessary to create an open database of reaction data with rich machine-readable metadata associated with each transformation.

Another solution for scarcity and heterogeneity in glycosylation data used a dataset obtained from the automated microreactor with inline HPLC-DAD measurement previously described to train a random-forest model.[18] This approach benefits from a highly controlled and homogenous dataset, as all results are obtained under the same environmental and analytical conditions. Another benefit lies in also reporting negative results, mostly absent from published data but beneficial for model training. Manually selected descriptors (e.g. ^{13}C chemical shift of the anomeric carbon to encode the leaving group or binary encoding of the orientation of hydroxyl/protecting groups around the pyrane core, etc.) allow for model creation despite the relatively low number of experimental data points (268). This approach is comparably simple to interpret. From the random forest model they derive the most influential factors on stereoselectivities in the glycosylation, expressed in per cent of influence on stereoselectivity in the following. In agreement with the results of Wang and co-workers previously described,[41] they found the electrophile (27%) more influential than nucleophile (20%) and most of the electrophile influence (17%) is attributed to the C2 group orientation.[18] Thanks to their inclusion in the dataset, environmental factors were also investigated and found to

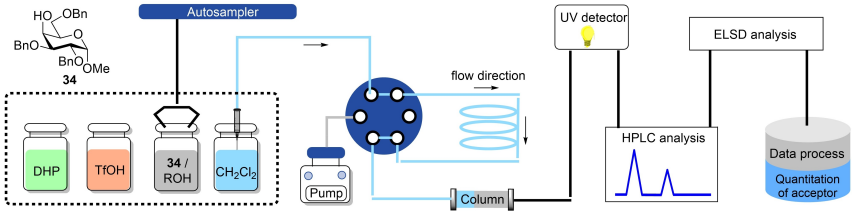


Figure 5. Schematic representation of the automated experimental procedure for Aka determination. An equimolar mixture of reference galactoside **34** and the acceptor whose Aka is to be measured are reacted in the presence of TfOH with 1 equivalent of DHP. The different conversion of reference and analyte to the corresponding 2-tetrahydropyranyl ethers, measured by HPLC, is used to calculate the Aka values.[41]

contribute to stereoselectivity (27% for solvent and 19% for temperature). Unknown glycosylations were predicted with good correspondence to experimental data.

The examples included in this section highlight attempts to shift part of the complexity associated with glycan synthesis to chemical informer tools/prediction systems. The increasing utility and accuracy of these digital tools is the result of a) more complex systems, from RRV to RRV + Aka to ML models, combined with b) more robust datasets of experimental outcomes. However, the high sensitivity of glycosylations to the experimental conditions means that the reproducibility of the results across different labs is only ensured if significant attention is paid to replicating experimental setups as well. From this point of view, a novel algorithm recently reported by the Aspuru-Guzik laboratory for the identification of robust reaction conditions[40] could be used to optimize glycosylation conditions across different labs and/or setups (both in batch and flow). This would ensure a higher level of reproducibility and provide a solid set of data to train future machine learnt models.

Summary and Outlook

Here, we present examples of enabling technologies in glycan synthesis from automation platforms, flow chemistry and digital tools for carbohydrate chemistry. These different technologies have been often investigated in isolation, but many synergies are possible by combining them. For example, the datasets to train the GlycoComputer and the ML model reported by Seeberger and Gilmore were both obtained in an automated fashion. [18,41] This is particularly beneficial as an automated platform can ensure higher reproducibility and lower error than manually-generated experiments.

Combining automated reaction platforms, flow chemistry and digital tools needs to be explored. In analogy with what has been observed in solid-phase peptide synthesis,[45] flow chemistry could be used to further accelerate the automated glycan assembly cycle time. The number of glycosylations performed in AGA could constitute a valuable source of data[46] if interfaced with an analytical technique (or a surrogate metric[47]) capable of providing information on the degree of success of each cycle.

To enable the field of glycoscience to grow, the synthetic accessibility of complex oligosaccharide structures has to be further streamlined. With automated glycan assembly increasingly consolidating as a mature technology, the synthetic bottlenecks are shifting to the synthesis of building blocks and to the prediction of the building blocks to be used for uncommon linkages. Flow chemistry and data science-based solutions can help tackle these challenges, thus realizing the vision of a self-driving laboratory[48] for oligosaccharide synthesis. To this end, despite the intrinsic complexity associated with carbohydrate chemistry, the field benefits from a head start in the form of pioneering work started more than 20 years ago both for automated glycan assembly[9a] and glycosylation predictions.[33]

Acknowledgements

We thank the Max-Planck Society and the DFG InCHeM (FOR 2177) for generous financial support. E.T.S. acknowledges financial support from the Alexander von Humboldt Foundation. Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: carbohydrates · flow chemistry · glycosylation · oligosaccharides · reaction automation

[1] *Essentials of Glycobiology*, 4th ed. (Eds.: A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar, P. H. Seeberger), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, **2022**.
[2] J. Danglad-Flores, S. Leichnitz, E. T. Sletten, A. Abragam Joseph, K. Bienert, K. Le Mai Hoang, P. H. Seeberger, *J. Am. Chem. Soc.* **2021**, *143*, 8893–8901.
[3] W. Yao, D.-C. Xiong, Y. Yang, C. Geng, Z. Cong, F. Li, B.-H. Li, X. Qin, L.-N. Wang, W.-Y. Xue, N. Yu, H. Zhang, X. Wu, M. Liu, X.-S. Ye, *Nat. Synth.* **2022**, *1*, 854–863.
[4] N. V. Ganesh, K. Fujikawa, Y. H. Tan, K. J. Stine, A. V. Demchenko, *Org. Lett.* **2012**, *14*, 3036–3039.
[5] T. Li, L. Liu, N. Wei, J.-Y. Yang, D. G. Chapla, K. W. Moremen, G.-J. Boons, *Nat. Chem.* **2019**, *11*, 229–236.
[6] R. Roychoudhury, N. L. B. Pohl, *Modern Synthetic Methods in Carbohydrate Chemistry* (Eds.: D. B. Werz, S. Vidal), Wiley-VCH, Weinheim, **2013**, pp. 221–239.
[7] J. Zhang, D. Liu, V. Saikam, M. R. Gadi, C. Gibbons, X. Fu, H. Song, J. Yu, S. M. Kondengaden, P. G. Wang, L. Wen, *Angew. Chem. Int. Ed.* **2020**, *59*, 19825–19829.
[8] P. H. Seeberger, *Nat. Chem. Biol.* **2009**, *5*, 368–372.
[9] a) O. J. Plante, E. R. Palmacci, P. H. Seeberger, *Science* **2001**, *291*, 1523–1527; b) M. Guberman, P. H. Seeberger, *J. Am. Chem. Soc.* **2019**, *141*, 5581–5592.
[10] H. S. Hahm, F. Broecker, F. Kawasaki, M. Mietzsch, R. Heilbronn, M. Fukuda, P. H. Seeberger, *Chem* **2017**, *2*, 114–124.
[11] J. Hofmann, H. S. Hahm, P. H. Seeberger, K. Pagel, *Nature* **2015**, *526*, 241–244.
[12] D. Senf, C. Ruprecht, G. H. M. de Kruijff, S. O. Simonetti, F. Schuhmacher, P. H. Seeberger, F. Pfrengle, *Chem. Eur. J.* **2017**, *23*, 3197–3205.
[13] X. Wu, M. Delbianco, K. Anggara, T. Michnowicz, A. Pardo-Vargas, P. Bharate, S. Sen, M. Pristl, S. Rauschenbach, U. Schlickum, S. Abb, P. H. Seeberger, K. Kern, *Nature* **2020**, *582*, 375–378.
[14] A. A. Joseph, A. Pardo-Vargas, P. H. Seeberger, *J. Am. Chem. Soc.* **2020**, *142*, 8561–8564.
[15] T. Tyrikos-Ergas, E. T. Sletten, J.-Y. Huang, P. H. Seeberger, M. Delbianco, *Chem. Sci.* **2022**, *13*, 2115–2120.
[16] E. T. Sletten, J. Danglad-Flores, S. Leichnitz, A. Abragam Joseph, P. H. Seeberger, *Carbohydr. Res.* **2022**, *511*, 1084–1089.
[17] O. T. Tuck, E. T. Sletten, J. Danglad-Flores, P. H. Seeberger, *Angew. Chem. Int. Ed.* **2022**, *61*, e202115433.
[18] S. Moon, S. Chatterjee, P. H. Seeberger, K. Gilmore, *Chem. Sci.* **2021**, *12*, 2931–2939.
[19] a) H. Grimberg, V. S. Tiwari, B. Tam, L. Gur-Arie, D. Gingold, L. Polachek, B. Akabayov, *J. Cheminf.* **2022**, *14*, 4; b) S. Mohapatra, N. Hartrampf, M. Poskus, A. Loas, R. Gómez-Bombarelli, B. L. Pentelute, *ACS Cent. Sci.* **2020**, *6*, 2277–2286.
[20] M. B. Plutschack, B. Pieber, K. Gilmore, P. H. Seeberger, *Chem. Rev.* **2017**, *117*, 11796–11893.
[21] S. Chatterjee, S. Moon, F. Hentschel, K. Gilmore, P. H. Seeberger, *J. Am. Chem. Soc.* **2018**, *140*, 11942–11953.
[22] a) R. R. Schmidt, W. Kinzy, *Adv. Carbohydr. Chem. Biochem.* **1994**, *50*, 21–123; b) K. Okamoto, T. Goto, *Tetrahedron* **1990**, *46*, 5835–5857; c) K. Fukase, K. Tanaka, T. Goi, *Synlett* **2005**, *2005*, 2958–2962; d) G. J. Boons,

A. V. Demchenko, *Chem. Rev.* **2000**, *100*, 4539–4566; e) M. Weishaupt, S. Eller, P. H. Seeberger, *Methods Enzymol.* **2010**, *478*, 463–484; f) M. P. Deninno, *Synthesis* **1991**, 583–593.
[23] I. V. Myachin, Z. Z. Mamirgova, E. V. Stepanova, A. I. Zinin, A. O. Chizhov, L. O. Kononov, *Eur. J. Org. Chem.* **2022**, e202101377.
[24] Y. Uchinashi, M. Nagasaki, J. Zhou, K. Tanaka, K. Fukase, *Org. Biomol. Chem.* **2011**, *9*, 7243–7248.
[25] K. Tanaka, Y. Mori, K. Fukase, *J. Carbohydr. Chem.* **2009**, *28*, 1–11.
[26] K. Fukase, S.-I. Tanaka, M. Takashina, H. Tokimoto, Y. Fujimoto, K. Tanaka, *Synlett* **2005**, *15*, 2325–2328.
[27] M. Tsutsui, J. Sianturi, S. Masui, K. Tokunaga, Y. Manabe, K. Fukase, *Eur. J. Org. Chem.* **2020**, 1802–1810.
[28] D. M. Ratner, E. R. Murphy, M. Jhunjhunwala, D. A. Snyder, K. F. Jensen, P. H. Seeberger, *Chem. Commun.* **2005**, *5*, 578–580.
[29] K. Geyer, P. H. Seeberger, *Helv. Chim. Acta* **2007**, *90*, 395–403.
[30] C. Cavedon, E. T. Sletten, A. Madani, O. Niemeyer, P. H. Seeberger, B. Pieber, *Org. Lett.* **2021**, *23*, 514–518.
[31] M. Guberman, B. Pieber, P. H. Seeberger, *Org. Process Res. Dev.* **2019**, *23*, 2764–2770.
[32] S. Yalamanchili, T. A. Nguyen, A. Zsikla, G. Stamper, A. E. DeYong, J. Florek, O. Vasquez, N. L. B. Pohl, C. S. Bennett, *Angew. Chem. Int. Ed.* **2021**, *60*, 23171–23175.
[33] Z. Zhang, I. R. Ollmann, X.-S. Ye, R. Wischnat, T. Baasov, C.-H. Wong, *J. Am. Chem. Soc.* **1999**, *121*, 734–753.
[34] G. Pesciullesi, P. Schwaller, T. Laino, J. L. Reymond, *Nat. Commun.* **2020**, *11*, 4874.
[35] Y. Shen, J. E. Borowski, M. A. Hardy, R. Sarpong, A. G. Doyle, T. Cernak, *Nat. Rev. Methods Primers* **2021**, *1*.
[36] a) K. M. Koeller, C.-H. Wong, *Chem. Rev.* **2000**, *100*, 4465–4494; b) K.-K T. Mong, C.-H. Wong, *Angew. Chem. Int. Ed.* **2002**, *41*, 4087–4090; c) C.-W. Cheng, C.-W. Cheng, C.-Y. Wu, W.-L. Hsu, C.-H. Wong, *Biochemistry* **2020**, *59*, 3078–3088.
[37] C. W. Cheng, Y. Zhou, W. H. Pan, S. Dey, C. Y. Wu, W. L. Hsu, C. H. Wong, *Nat. Commun.* **2018**, *9*, 5202.
[38] a) C.-W. Chang, C.-H. Wu, M.-H. Lin, P.-H. Liao, C.-C. Chang, H.-H. Chuang, S.-C. Lin, S. Lam, V. P. Verma, C.-P. Hsu, C.-C. Wang, *Angew. Chem. Int. Ed.* **2019**, *58*, 16775–16779; b) M.-H. Lin, C.-W. Chang, T.-Y. Chiang, V. M. Dhurandhare, C.-C. Wang, *Org. Lett.* **2021**, *23*, 7313–7318; c) K. H. Asressu, C.-W. Chang, S. Lam, C.-C. Wang, *Eur. J. Org. Chem.* **2021**, *2021*, 4525-4530;d) C.-W. Chang, M.-H. Lin, C.-C. Wang, *Chem.— Eur. J.* **2021**, *27*, 2556-2568.
[39] S. van der Vorm, T. Hansen, J. M. A. van Hengst, H. S. Overkleeft, G. A. van der Marel, J. D. C. Codée, *Chem. Soc. Rev.* **2019**, *48*, 4688–4706.
[40] a) S. van der Vorm, J. M. A. van Hengst, M. Bakker, H. S. Overkleeft, G. A. van der Marel, J. D. C. Codée, *Angew. Chem. Int. Ed.* **2018**, *57*, 8240–8244; b) S. van der Vorm, T. Hansen, H. S. Overkleeft, G. A. van der Marel, J. D. C. Codée, *Chem. Sci.* **2017**, *8*, 1867–1875; c) T. Hansen, T. P. Ofman, J. G. C. Vlaming, I. A. Gagarinov, J. van Beek, T. A. Goté, J. M. Tichem, G. Ruijgrok, H. S. Overkleeft, D. V. Filippov, G. A. van der Marel, J. D. C. Codée, *Angew. Chem. Int. Ed.* **2021**, *60*, 937–945.
[41] C. W. Chang, M. H. Lin, C. K. Chan, K. Y. Su, C. H. Wu, W. C. Lo, S. Lam, Y. T. Cheng, P. H. Liao, C. H. Wong, C.-C. Wang, *Angew. Chem. Int. Ed.* **2021**, *60*, 12413–12423.
[42] P. Schwaller, T. Laino, T. Gaudin, P. Bolgar, C. A. Hunter, C. Bekas, A. A. Lee, *ACS Cent. Sci.* **2019**, *5*, 1572–1583.
[43] D. M. Lowe, *Extraction of Chemical Structures and Reactions from the Literature*, PhD thesis, University of Cambridge (Cambridge), **2012**.
[44] M. Aldeghi, F. Häse, R. J. Hickman, I. Tamblyn, A. Aspuru-Guzik, *Chem. Sci.* **2021**, *12*, 14792–14807.
[45] N. Hartrampf, A. Saebi, M. Poskus, Z. P. Gates, A. J. Callahan, A. E. Cowfer, S. Hanna, S. Antilla, C. K. Schissel, A. J. Quartararo, X. Ye, A. J. Mijalis, M. D. Simon, A. Loas, S. Liu, C. Jessen, T. E. Nielsen, B. L. Pentelute, *Science* **2020**, *368*, 980–987.
[46] Y. Shi, P. L. Prieto, T. Zepel, S. Grunert, J. E. Hein, *Acc. Chem. Res.* **2021**, *54*, 546–555.
[47] E. T. Sletten, J. Danglad-Flores, M. Nuño, D. Guthrie, P. H. Seeberger, *Org. Lett.* **2020**, *22*, 4213–4216.
[48] M. Seifrid, R. Pollice, A. Aguilar-Granda, Z. Morgan Chan, K. Hotta, C. T. Ser, J. Vestfrid, T. C. Wu, A. Aspuru-Guzik, *Acc. Chem. Res.* **2022**, *55*, 2454–2466.

Manuscript received: October 24, 2022
Revised manuscript received: November 15, 2022
Accepted manuscript online: November 16, 2022
Version of record online: December 3, 2022