Carbohydrate Synthesis is Entering the Data-Driven Digital Era

Eric T. Sletten,[a] Jakob B. Wolf,[a, b] José Danglad-Flores,[a] and Peter H. Seeberger*[a, b]

Glycans are vital in biological processes, but their nontemplated, heterogeneous structures complicate structure-function analyses. Glycosylation, the key reaction in synthetic glycochemistry, remains not entirely predictable due to its complex mechanism and the need for protecting groups that impact reaction outcomes. This concept highlights recent advancements in glycochemistry and emphasizes the integration of digital tools, including automation, computational modelling, and data management, to improve carbohydrate synthesis and support further progress in the field.

1. Introduction

Glycans are essential for a wide range of biological processes.[1] From serving as an energy source to playing roles in pathogen interactions, biomaterial scaffold formation, and cell-cell signaling, carbohydrates are integral to nearly all aspects of biology.[2] Understanding the structure-function relationship of glycans is complicated by their inherent nontemplated nature.

Advances in synthesis and modification of homogenous glycan motifs must keep pace with recent breakthroughs in glycobiology in order to provide tools for detailed investigations.[3–9] The key chemical reaction of the glycosciences, the glycosylation, suffers from a long-standing reputation for unpredictability and irreproducibility.[10] The reaction mechanism of glycosylations follow paths along a S_N1 to S_N2 continuum depending on the influence of various continuous and discrete conditional factors that influence the stereochemical outcome (Figure 1).[11] The presence of multiple hydroxyl groups in carbohydrates provides different positions of elongation, necessitating the use of protecting groups to achieve regioselectivity. The protecting groups are not merely passive bystanders in the reaction but have significant impact on the stereoelectronic properties of the molecule. These intrinsic problems of chemical glycosylation can be circumvented by enzymatically or chemoenzymatically constructing complex glycans without the need of protecting groups.[12–20] However, these enzymatic methods are limited in substrate scope.

Here, we describe the current state of glycochemistry and highlight the application of new digital tools to the glycosciences. We provide suggestions as for a better integration of automation, computational tools, and data management to further enhance carbohydrate synthesis.

2. Current State of Glycochemistry

Before applying digital tools to glycochemistry, it is essential to identify the existing challenges in glycosynthesis.[21] Assembling the glycosynthesis puzzle with the help of digital tools requires systematic assignment of the pieces and identifying the missing parts of the grand picture. Recent developments in chemical synthesis methodologies for glycan construction have focused on the synthesis of *C*-glycosides, encompassing both alkyl and aryl derivatives.[22–29] Despite numerous syntheses of complex, bespoke glycans,[30–33] significant challenges in chemical *O*-glycoside methodology remain unresolved.[21] Without a centralized, comprehensive glycosylation database, information on chemical glycosynthesis is mostly scattered throughout the scientific literature, shared anecdotally, or exists merely as tacit knowledge of versed practitioners. Novices to the carbohydrate field face challenges that result in reproducibility issues.[34] Fundamental mechanistic insights, essential for predicting novel glycosylation reactions, are obscured by incomplete data.[11,35] Without a standard approach, each individual glycosylation is treated as a unique coupling, with the choice of protecting groups, leaving groups, and reaction conditions such as reaction temperature, time, solvent choice, and concentration based on experience. The lack of uniformity and completeness in data reporting is a significant barrier to data-driven models. Arbitrary parameters, such as ill-defined temperature ramps or vague reaction times (e.g., "overnight"), are often used, making it difficult to identify precise conditions.[34]

Crich illustrated the extensive variety of glycosylation conditions in 2011, referencing Hindsgaul's 1995 tabulation of over 700 glycosylation reactions, promoters, solvents, and temperatures involved, highlighting the significant empiricism and the complex terminology of the field.[36,37] An updated tabulation of 2010 data was considered to show similar diversity, a situation that likely has not changed until today. However, new automated platforms and digital tools to handle large chemical

[a] E. T. Sletten, J. B. Wolf, J. Danglad-Flores, P. H. Seeberger
Max Planck Institute of Colloids and Interfaces, Potsdam Science Park, Am Mühlenberg 1 14476, Potsdam, Germany
E-mail: peter.seeberger@mpikg.mpg.de

[b] J. B. Wolf, P. H. Seeberger
Institut für Chemie, Biochemie und Pharmazie, Freie Universität Berlin, Takusstraße 3 14195, Berlin, Germany

Chem. Eur. J. 2025, *31*, e202500289 (1 of 9)

© 2025 The Author(s). Chemistry – A European Journal published by Wiley-VCH GmbH
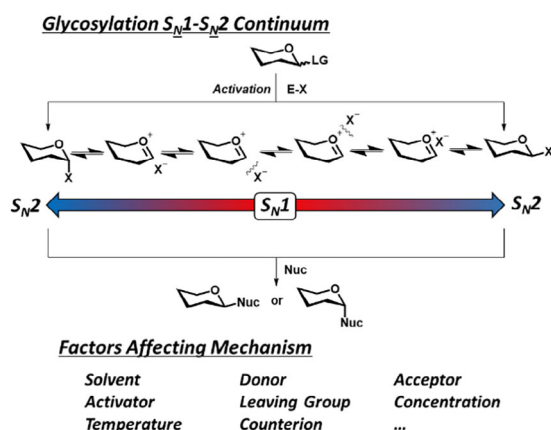
Figure 1. Pathway of glycosylation mechanism along a S_N1 to S_N2 continuum and contributing factors.

datasets can leverage the empirical nature of glycosylation research. While these algorithms and tools have existed for decades,[38–40] accessibility and ease of use have improved recently. Widespread application has demonstrated the value of addressing current synthetic challenges using algorithms that blur the boundaries between chemical synthesis and data science.[41–45]

3. Prerequisites for Digital Glycoscience

Digital chemistry applies computational tools, algorithms, and digital technologies to solve chemical challenges by simulating, analyzing, accelerating, and predicting chemical processes and molecular behavior. It enables deeper insights into molecular interactions, reaction mechanisms, and material properties through data-driven modeling and automation, enhancing both research efficiency and discovery potential across fields such as drug development, materials science, and sustainable chemistry (Figure 2).[44,46] Key focus areas include quantum mechanical calculations, molecular modeling, cheminformatics, virtual screening, machine learning, and artificial intelligence applications, as well as specialized tools and software like electronic lab notebooks, searchable online repositories, retrosynthetic planning applications, and robotics.[44,46]

While some of these methods, such as simulations or automated screenings, create large amounts of data, others, such as machine learning for retrosynthetic planning, require large amounts of data. For data to be usable or understandable for different people or applications, it has to contain metadata, to record and specify what was measured under which conditions. This data needs to be structured in such a way that it is easily reusable. FAIR data is the core of digitalization and scientific

Eric T. Sletten completed his undergraduate studies in chemistry at Wartburg College (IA). In 2013, he joined the University of Iowa Department of Chemistry and obtained his Ph.D. in 2018. In 2019, he subsequently moved to the Max-Planck-Institute of Colloids and Interfaces in Potsdam (Germany) as a postdoctoral researcher and was awarded Alexander-von-Humboldt postdoctoral research fellowship. His research focuses on the development of methodologies for glycan synthesis and automated synthesis.

Jakob B. Wolf studied chemistry at the Humboldt-Universität Berlin where he obtained his M.S. in 2019 on research conducted at University of Glasgow with Prof Cronin. After a short stay at BAM, Berlin to enrich automated synthesis with process control and automatically obtain standard materials, he moved to the Max-Planck-Insitute of Colloids and Interfaces (Potsdam) for his doctoral research, where he focuses on automated screening platforms for glycosylation chemistry and creation of large, structured datasets leveraging flow chemistry.

José Danglad-Flores got his Chemical Engineering degree in 2010 at the Universidad de Oriente, Venezuela. From 2011 to 2014 worked as Research Engineer at the steel making industry. Then completed his PhD in Physical Chemistry in 2018, working at MPIKG (Theory and Bio-system Dept.), TU Berlin, and North Carolina State University. He moved to the MPIKG Bimolecular System Department working as a Postdoc 2019-2020, and from 2021 is an Automation Project Leader. His work focuses on automated reactor development and physicochemical aspects of solid phase synthesis.

Peter H. Seeberger studied chemistry in Erlangen (Germany), received a Ph.D. in biochemistry in at the University of Colorado at Boulder and performed postdoctoral research at the Sloan-Kettering Cancer Center in New York (95-97). After positions as Firmenich Associate Professor with tenure at MIT (98-03) and Professor at ETH Zurich (03-09), he is now a Director at the Max-Planck-Institute in Potsdam and Professor at Freie Universität Berlin. His research interests include chemistry, biology, and engineering.
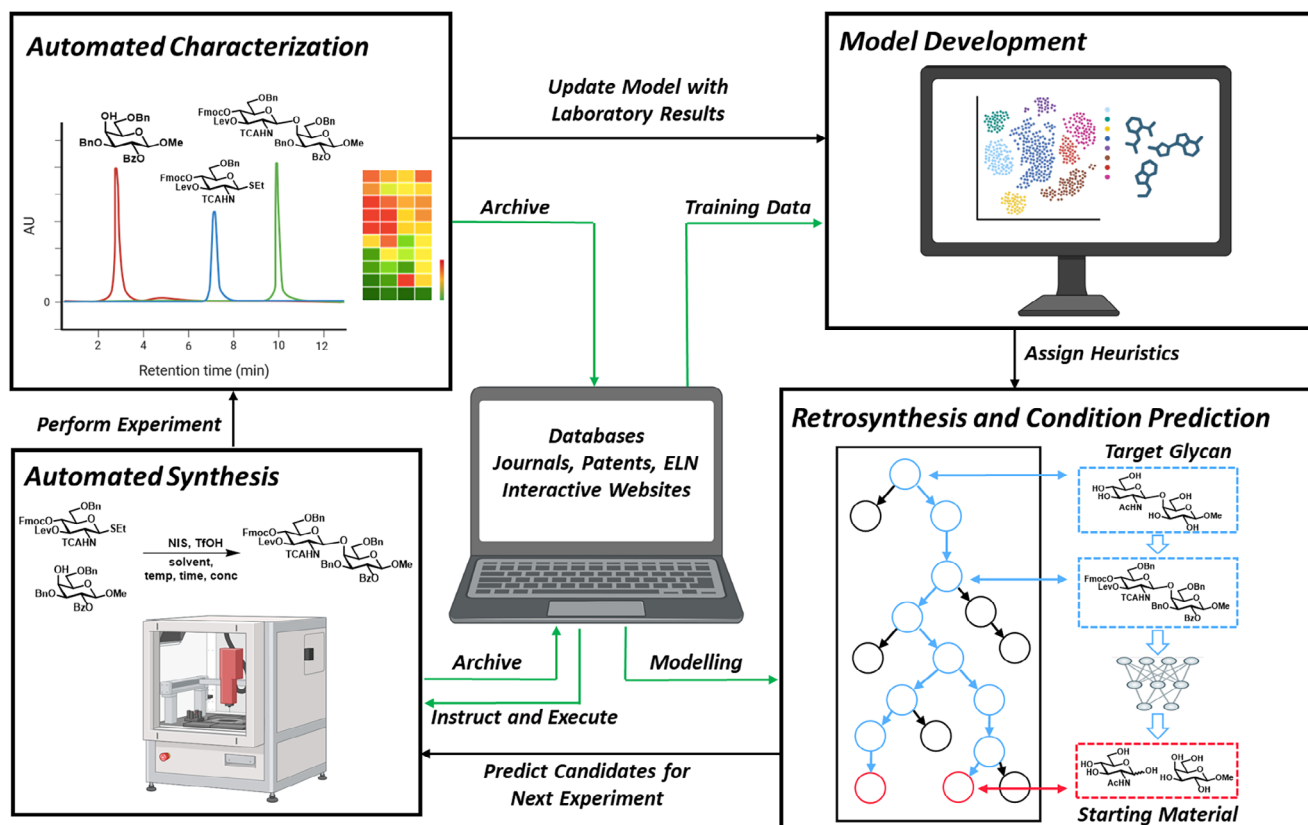
Chemistry–A European Journal

Concept

doi.org/10.1002/chem.202500289

Chemistry Europe
European Chemical
Societies Publishing

Figure 2. Potential workflow for the advancement of glycosylation chemistry using digital chemistry tools.
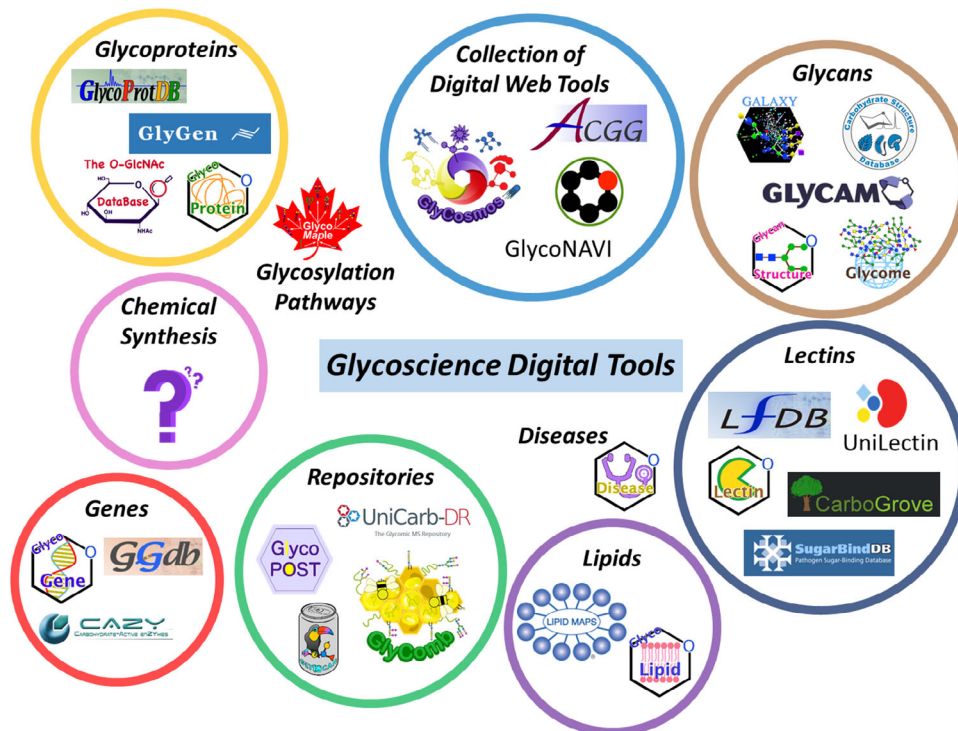

Figure 3. Different categories of glycoscience digital tools utilized to accelerate discovery and understanding.

data management as data stewardship should be based on four foundational values: Findability, Accessibility, Interoperability, and Reusability.[47] Reporting FAIR data facilitates and simplifies the processes of discovery, evaluation, and reuse in subsequent studies. Widespread implementation of the FAIR principles will help bridge the information gap by data standardization, reproducibility, and evaluation.[48] In the past, limited page counts in print publications prevented most scientific data from being published, but even today, despite the capacity of digitally capturing extensive amounts of data, much data remain unpublished or difficult to reuse.[49] To advance data-intensive glycoscience research and address reproducibility issues, we must change how experimental data are collected and reported, ensuring it is both structured and open, and complemented by tools like semantic web technologies to enhance understanding and usability.[49] Implementation of Laboratory Information Management Systems (LIMS) and Electronic Lab Notebooks (ELN), represented by open-source solutions such as Open-BIS or Chemotion, can help this objective.[49–51] Automation and real-time reaction monitoring enable data-rich experimentation, crucial for navigating the complexities of chemical synthesis.[45,52–54] Flexible, automated, and high-throughput experimentation (HTE) platforms help minimize human error and provide an unbiased representation of both positive and negative results.[48,55–57] These platforms are expensive and experimentation can be performed manually in well-plates as well. However all metadata, including the reaction setup, must be reported to be useful.[10,48,58] Publishing the data in searchable and machine-actionable online repositories such as the Open Reaction Database (ORD) or Chemotion, ensures that information is findable and interpretable by chemists and their computational agents.[49,50,59,60] Standardized data structuring paves the way for data-driven techniques, like machine learning (ML) and artificial intelligence (AI), to become indispensable tools for scientific advancement.[44,61]

Mid-sized context-sensitive databases and ontologies were developed early for the glycosciences (Figure 3). CarbBank, a database project for glycan structures was established in 1987 and contained over 23000 glycan sequences.[62] Multiple other database projects followed, including the Kyoto Encyclopedia of Genes and Genomes (KEGG) Glycan,[63] the database of the US Consortium for Functional Glycomics (CFG),[64] GLYCOSCIENCES.de,[65] GlycoSuiteDB,[66] UniCarbKB,[67] Carbohydrate Structure Database (CSDB),[68] GlycomeDB,[69] and EUROCarbDB.[70] By implementing Semantic Web technologies, GlyTouCan linked glycan structures across multiple glycomic databases by a unique accession number.[71] GlycoRDF, a Resource Description Framework (RDF) for a standard representation of storing glycomics data (glycan structures, publication information, biological source information, and experimental data) simplified this process.[72] Glycobiology repositories and databases are compiled on GlyCosmos (https://glycosmos.org/) for structure-property relationship queries.[73] Leveraging the technologies summarized in Figure 3, ensured that glycan data is accessible, interoperable, and useful for researchers and machines.
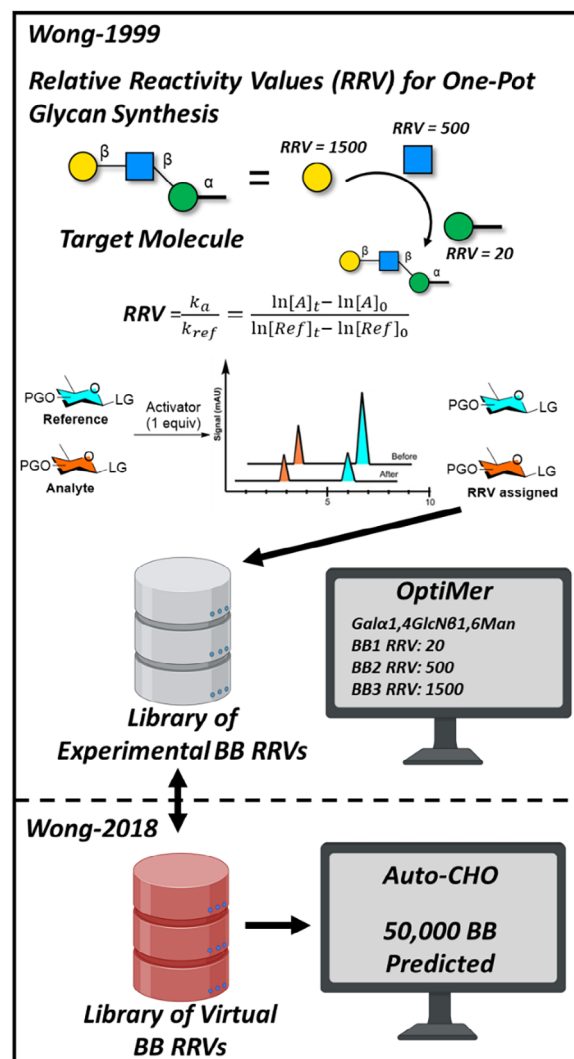

Figure 4. Database software tools for one-pot glycan construction retrosynthesis based on RRVs.

4. Toward Digitizing Chemical Glycan Synthesis

Databases describing the chemical synthesis of glycans are missing (Figure 3). The key is standardization of the discrete input data (reagents, donor, acceptor, activator, solvent, reaction equipment, and setup), and continuous reaction parameters (time, temperature, concentration), mapped to the raw (chromatograms and spectra containing sufficient metadata) and interpreted output data (yield, selectivity, conversion). The databases should follow the guidelines put forth by Andreana and Crich.[34] A glycosylation-specific ontology would facilitate the development of a semantic web-based repository for the curation of chemical species, their properties, and reactions. Thereby, it will be possible to examine the breadth and coverage of the glycosylation space[74] and correlate yield, stereoselectivity, and conversion with a specific structural (steric or electronic) or underlying property using quantitative structure–activity relationship models (QSAR models).[75] Stereoselectivity, solvent effects, armed/disarmed theory, and long-range participation
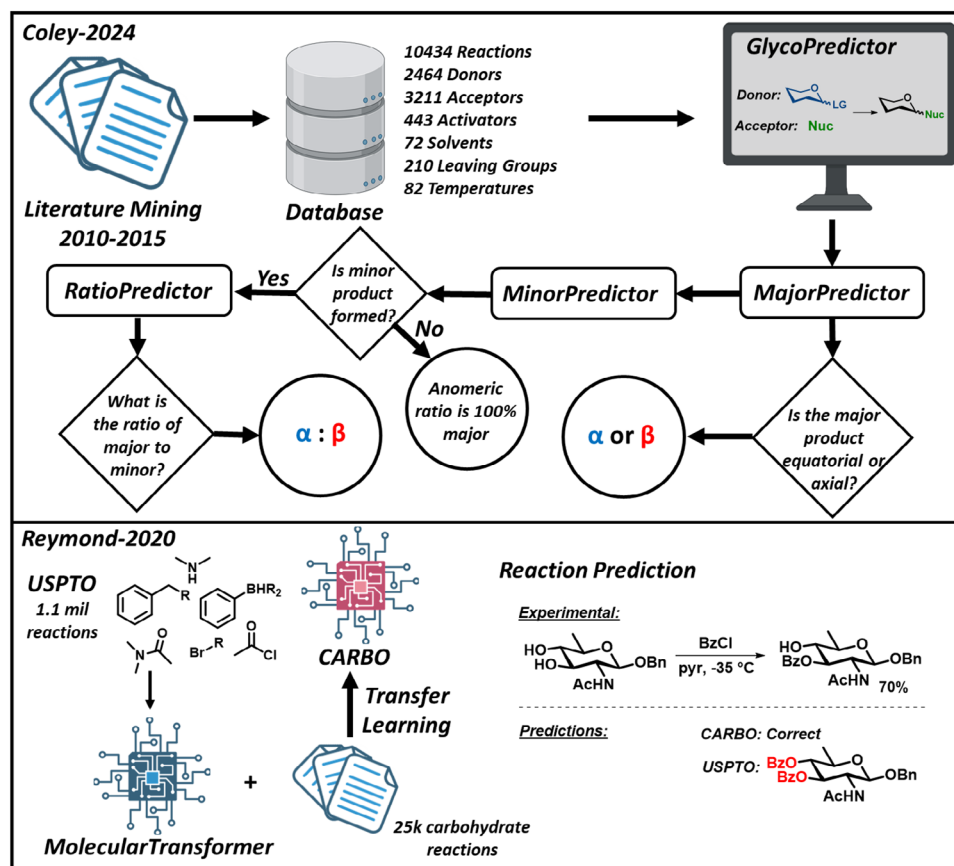
Chemistry–A European Journal

Concept
doi.org/10.1002/chem.202500289

Chemistry Europe
European Chemical
Societies Publishing

Figure 5. Carbohydrate reaction prediction mining literature data.

could be quantified.[11,35,75–78] Since the majority of practitioners invest the majority of their time in the construction of monosaccharide building blocks, detailed synthetic information should encompass protecting group manipulations on glycosyl donors and acceptors. The information will provide valuable insights into retrosynthetic routes to afford monosaccharides and protecting group orthogonality.[79] Within the realm of automated synthesis, a consistent format for reporting operational protocols is still missing, in particular for carbohydrate synthesis;[13,32,80–86] beyond a detailed setup description, effective sharing of the steps and parameters is pivotal for ensuring reproducibility and standardization of the procedures through the field.

Automated experimentation to rapidly and uniformly explore chemical space has been the basis to standardize data formats.[74,87] HTE datasets are available for few reaction types, most commonly Buchwald–Hartwig aminations and Suzuki couplings, however more diverse datasets are emerging.[42,43,88] Extensive reaction datasets as the Chemical Abstracts Service (CAS) Content Collection, Pistachio, USPTO, and Reaxys contain numerous reactions but are proprietary, have varied protocols, and/or underrepresent negative results and can be skewed toward popular reactions and/or methodologies.[48,55–57,89–92] Reaction-specific, rigorous, well-defined repositories, such as those developed by AbbVie for C—C cross-couplings, help mitigate the impact of reaction condition variations.[91,92] Similarly, in computer-aided retrosynthesis, large reaction datasets are signif-

icantly enhanced by integrating synthesis expert knowledge.[93] These initiatives underscore the importance of developing comprehensive and standardized datasets and digital tools to advance the field effectively.

The application of digital tools for glycochemistry such as data visualization or AI/ML for reaction condition prediction and retrosynthetic models is challenging. Building glycosylation reaction databases bottom-up requires numerous experiments. A community effort is needed as reagent procurement is time-consuming. Structural configuration, protecting groups, and chain length of carbohydrate building blocks profoundly impact sterics, electronics, and thereby the reactivity.[94] Although there are only ten types of monosaccharides in mammalian systems,[12] the inclusion of all the permutations of protecting groups at each position results in near-infinite combinations. The incorporation of bacterial motifs would result in a significant increase in the number of monosaccharide types and required protected building blocks. However, the sets of glycosyl pairs trialed can be minimized to effectively and efficiently traverse the glycosylation chemical space by balancing exploration and exploitation.[74] The representation of protected glycan structures in both a human-readable, typically pictorial chair representation, and a machine-readable form (e.g., SMILES) simultaneously needs development. Currently, no grammar-based language exists that can be intuitively and rapidly written and understood by both human and machine, such as the IUPAC condensed form used by GlyLES
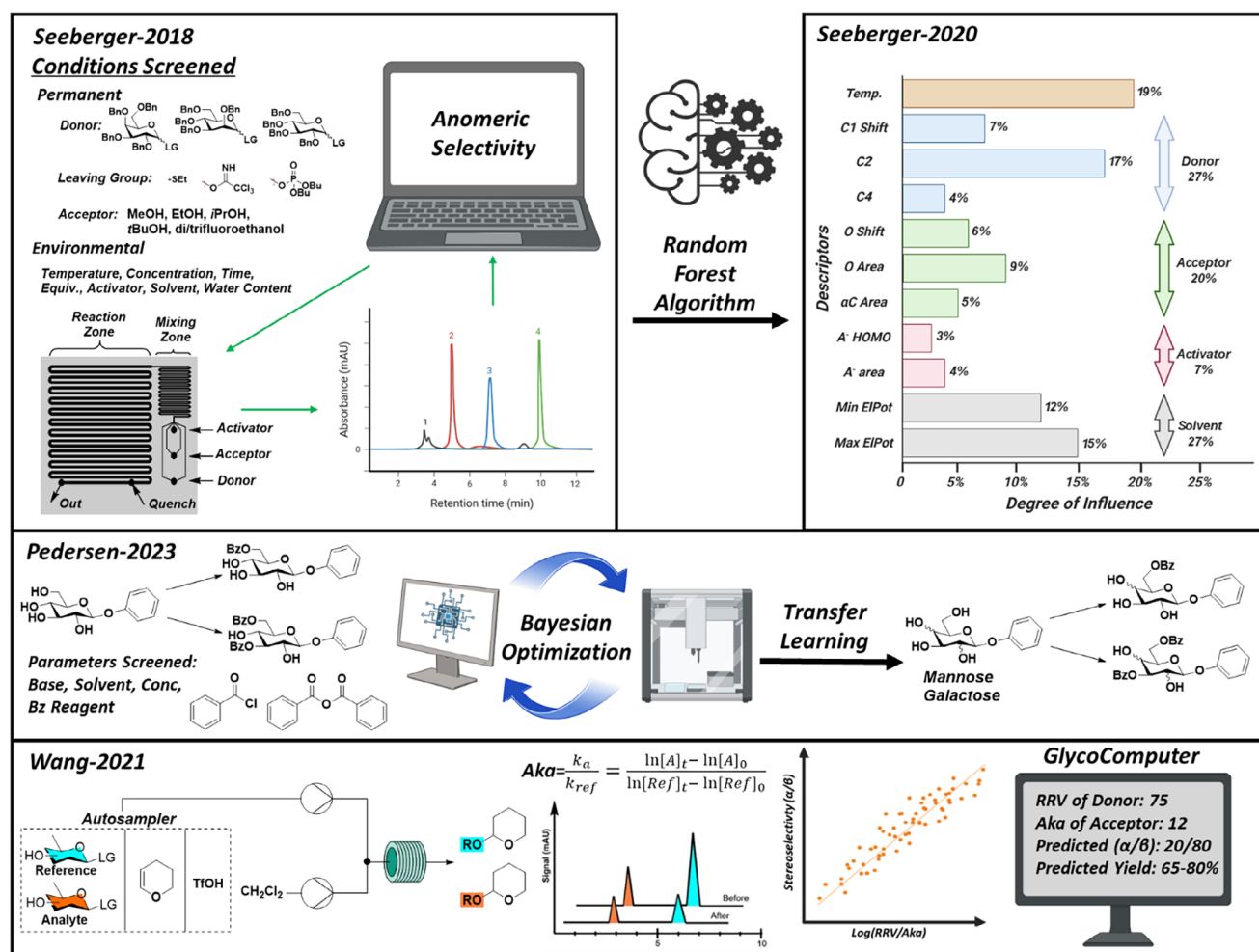
Figure 6. Carbohydrate reaction prediction using automated systems for the collection of homogenous data.

which is limited to unprotected structures.[95] The experimental means of data acquisition and the platform used are important. Traditional reaction conditions for glycosylations are not directly amenable to commercial HTE systems, due to the necessity for anhydrous conditions, sub-zero temperatures, and the use of volatile solvents such as methylene chloride. Rapid reaction rates further complicate the acquisition of kinetic data, all those points requiring construction of bespoke in-house systems.[78]

5. Digitalizing Glycochemistry: Current State of the Art

Sophisticated digital tools and technologies (Figure 4) have been developed to address some of the challenges described above and discussed by others in the field.[82] In 1999, the OptiMer software for programmable one-pot oligosaccharide synthesis was developed by Wong.[96] The OptiMer software filters all the donors in a database for sugar types, glycosylation sites, and stereochemical properties to predict the best donor sets for a target sequence by looking at their respective relative reactivity values (RRV). Using deterministic and Monte Carlo methods,

suitable donor combinations and synthetic routes that maximize RRV differences were identified. In contrast, experts consider practical factors like availability and cost when selecting the final route. The limitations of the OptiMer software were addressed with the introduction of Auto-CHO.[97] The new tool employs ML to predict the RRVs of 50000 virtual building blocks (BBL) via generation of 1D and 2D molecular descriptors from the original dataset.[94,98]

In an attempt to train models on literature datasets, Coley extracted over 10000 glycosylation reactions from the CAS Content Collection and the Hindsgaul dataset with the objective of creating machine-learnt models (Figure 5).[36,99] These models predict the predominantly formed anomer in a glycosylation reaction, determine whether the other anomer also forms and if so, estimate the anomeric ratio. The resulting models have been incorporated into a publicly accessible tool, named GlycoPredictor.[99] Similarly, Reymond applied transfer learning to the *Molecular Transformer*, developed earlier by Schwaller et. al. for general reaction prediction based on the USPTO database[100] by refining it with 25,000 manually extracted carbohydrate reactions from the Reaxys database.[101] The refined CARBO model was able to successfully predict the synthesis of a trisaccharide with five challenging regioselective protections

Chemistry–A European Journal

Concept

doi.org/10.1002/chem.202500289

Chemistry Europe
European Chemical
Societies Publishing

Future Directions
- *Predictable glycosylations: selectivity and yield*
- *Reducing learning curve for new glycochemists*

Key Challenges
- *Glycan complexity*
- *Unstandardized reporting*
- *Sporadic data*
- *Lack of understanding glycosylation mechanism*
  - *Kinetics, thermodynamics, activation barriers*
  - *Discrete and continuous parameter influence*

Proposed Solutions
- *Use of automated technologies*
- *Machine- and human readable nomenclature*
- *Template for experimental reporting*
- *Large and curated datasets*
- *Models for reaction condition prediction*
- *Retrosynthetic models for glycan synthesis*
- *Searchable semantic web databases*
- *Computational studies on protected glycans*

Figure 7. Current challenges and proposed solutions for the digitialization of glychemistry.

and four difficult regio- and stereoselective glycosylations, with 68% of steps correctly predicted, as opposed to only 39% by *Molecular Transformer*. The authors attributed the majority of erroneous predictions of CARBO to deficiencies in the training data, including the absence of stoichiometry. The development of more accurate glycosylation models hinges on the creation of an open database of reaction data, accompanied by comprehensive machine-readable metadata associated with each transformation. Similarly, the Coley model, yielding an R^2 of 0.58 for anomeric prediction would benefit from curated datasets. Accurate glycosylation models depend on accurate data, calling for the creation of a machine-readable, open database of glycosylation data, annotated by comprehensive metadata for each transformation.

Automated HTE platforms are a means to overcome the scarcity and heterogeneity in glycosylation data (Figure 6). Our laboratory developed an automated microfluidic-based screening platform to compare the effects of 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 on the stereoselectivity of several glycosylations.[75] The data obtained in this way were used to train a random-forest model on manually selected structural descriptors to derive the most influential factors on stereoselectivity.[75] Wang introduced an HPLC-based automated platform to establish a single nucleophilicity metric for an array of acceptors.[102] The GlycoComputer software successfully predicts glycosylation yields and stereoselectivity by statistically analyzing the properties of donors, acceptors, activation systems, and solvents, specifically comparing the RRV of donors and their introduced nucleophilic constant of acceptors.[103] Pedersen in a closed-loop Bayesian optimization improved protecting group manipulations in regioselective benzoylations of unprotected glycosides and a

transfer learning method was applied to accelerate the optimization process of regioselectively protecting new substrates.[104]

6. Summary and Outlook

Quick access to defined complex glycans is needed to procure substrates that enable molecular glycobiology (Figure 7). Digital tools can only be employed efficiently when the current experimental practices and reporting are revised. Reporting standards need to be decided upon and then implemented, similar to MIRAGE (Minimum Information Required for A Glycomics Experiment) standards set up by the glycomics community.[105] As mentioned by Crich and Andreana, big data approaches need to include concentration, and glycosylations should be conducted at a well-defined temperature.[34] Along with structured data generation from new experiments and mining the literature, fundamental work concerning reaction order, rate, and barriers of activation for glycosylations and auxiliary chemical manipulations is pivotal for designing efficient automated assembly protocols and developing accurate retrosynthetic predictor models.

Conflicts of Interest

The authors declare no conflict of interest.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

[1] K. T. Schjoldager, Y. Narimatsu, H. J. Joshi, H. Clausen, *Nat. Rev. Mol. Cell Biol.* **2020**, *21*, 729.

[2] A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, A. G. Darvill, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar, P. H. Seeberger, (Eds.) in *Essentials of Glycobiology* Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), **2015**.

[3] J. Hu, D. T. Huynh, M. Boyce, *Biochemistry* **2024**, *63*, 2937

[4] Y. Watanabe, J. D. Allen, D. Wrapp, J. S. McLellan, M. Crispin, *Science* **2020**, *369*, 330.

[5] P. Zhao, J. L. Praissman, O. C. Grant, Y. Cai, T. Xiao, K. E. Rosenbalm, K. Aoki, B. P. Kellman, R. Bridger, D. H. Barouch, M. A. Brindley, N. E. Lewis, M. Tiemeyer, B. Chen, R. J. Woods, L. Wells, *Cell Host Microbe* **2020**, *28*, 586.

[6] S. M. Banik, K. Pedram, S. Wisnovsky, G. Ahn, N. M. Riley, C. R. Bertozzi, *Nature* **2020**, *584*, 291.

[7] W. Huang, J. Giddens, S.-Q. Fan, C. Toonstra, L.-X. Wang, *J. Am. Chem. Soc.* **2012**, *134*, 12308.

Chem. Eur. J. **2025**, *31*, e202500289 (7 of 9)

© 2025 The Author(s). Chemistry – A European Journal published by Wiley-VCH GmbH

Chemistry–A European Journal | Concept

doi.org/10.1002/chem.202500289

Chemistry
Europe

European Chemical
Societies Publishing

[8] F. L. Kearns, M. A. Rosenfeld, R. E. Amaro, *J. Chem. Inf. Model.* **2024**, *64*, 7949.

[9] J. M. Hershewe, K. F. Warfel, S. M. Iyer, J. A. Peruzzi, C. J. Sullivan, E. W. Roth, M. P. DeLisa, N. P. Kamat, M. C. Jewett, *Nat. Comm.* **2021**, *12*, 2363.

[10] D. Crich, *J. Am. Chem. Soc.* **2021**, *143*, 17.

[11] P. O. Adero, H. Amarasekara, P. Wen, L. Bohé, D. Crich, *Chem. Rev.* **2018**, *118*, 8242.

[12] C.-C. Liu, J. Ye, H. Cao, *Acc. Chem. Res.* **2024**, *57*, 636.

[13] T. Li, L. Liu, N. Wei, J.-Y. Yang, D. G. Chapla, K. W. Moremen, G.-J. Boons, *Nat. Chem.* **2019**, *11*, 229.

[14] T. Rexer, D. Laaf, J. Gottschalk, H. Frohnmeyer, E. Rapp, L. Elling, in *Advances in Glycobiotechnology* **2021**, (Eds.: E. Rapp, U. Reichl), Springer International Publishing, Cham, pp. 231.

[15] E. T. Sletten, G. Fittolani, N. Hribernik, M. C. S. Dal Colle, P. H. Seeberger, M. Delbianco, *ACS Cent. Sci.* **2024**, *10*, 138.

[16] N. Lu, Y. Li, H. Xia, K. Zhong, C. Jia, J. Ye, X. Liu, C.-C. Liu, H. Cao, *Angew. Chem. Int. Ed.* **2022**, *61*, e202211032.

[17] J. Ye, H. Xia, N. Sun, C.-C. Liu, A. Sheng, L. Chi, X.-W. Liu, G. Gu, S.-Q. Wang, J. Zhao, P. Wang, M. Xiao, F. Wang, H. Cao, *Nat. Catal.* **2019**, *2*, 514.

[18] N. Lu, J. Ye, J. Cheng, A. Sasmal, C.-C. Liu, W. Yao, J. Yan, N. Khan, W. Yi, A. Varki, H. Cao, *J. Am. Chem. Soc.* **2019**, *141*, 4547.

[19] S. Ma, J. Gao, Y. Tian, L. Wen, *Org. Biomol. Chem.* **2024**, *22*, 7767.

[20] Z.-F. Hu, K. Zhong, H. Cao, *Curr. Opin. Chem. Biol.* **2024**, *78*, 102417.

[21] X. S. Streety, J. C. Obike, S. D. Townsend, *ACS Cent. Sci.* **2023**, *9*, 1285.

[22] C. Zhang, S.-Y. Xu, H. Zuo, X. Zhang, Q.-D. Dang, D. Niu, *Nat. Synth.* **2023**, *2*, 251.

[23] Y. Jiang, Y. Wei, Q.-Y. Zhou, G.-Q. Sun, X.-P. Fu, N. Levin, Y. Zhang, W.-Q. Liu, N. Song, S. Mohammed, B. G. Davis, M. J. Koh, *Nature* **2024**, *631*, 319.

[24] M.-Y. Lyu, S. A. Jacobo, M. K. Brown, *J. Am. Chem. Soc.* **2024**, *146*, 18866.

[25] X. Wu, S. Li, L. Chen, S. Ma, B. Ma, L. Song, D. Qian, *J. Am. Chem. Soc.* **2024**, *146*, 22413.

[26] J. Wu, R. Purushothaman, F. Kallert, S. L. Homölle, L. Ackermann, *ACS Catal.* **2024**, *14*, 11532.

[27] G. Cheng, B. Yang, Y. Han, W. Lin, S. Tao, Y. Nian, Y. Li, M. A. Walczak, F. Zhu, *Precis. Chem.* **2024**,*2* 587.

[28] D. Xie, W. Zeng, J. Yang, X. Ma, *Nat. Comm.* **2024**, *15*, 9187.

[29] Y. Yang, B. Yu, *Chem. Rev.* **2017**, *117*, 12281.

[30] B. Dhakal, A. Mandhapati, P. Eradi, S. Park, K. Fibben, K. Li, A. DeYong, S. Escopy, G. Karki, D. D. Park, C. A. Haller, E. Dai, L. Sun, W. A. Lam, E. L. Chaikof, *J. Am. Chem. Soc.* **2024**, *146*, 17414.

[31] Q. Zhu, Z. Shen, F. Chiodo, S. Nicolardi, A. Molinaro, A. Silipo, B. Yu, *Nat. Comm.* **2020**, *11*, 4142.

[32] 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.

[33] A. A. Joseph, A. Pardo-Vargas, P. H. Seeberger, *J. Am. Chem. Soc.* **2020**, *142*, 8561.

[34] P. R. Andreana, D. Crich, *ACS Cent. Sci.* **2021**, *7*, 1454.

[35] A. Hettikankanamalage, R. Lassfolk, F. Ekholm, R. Leino, D. Crich, *Chem. Rev.* **2020**, *120*, 7104.

[36] F. Barresi, O. Hindsgaul, *J. Carbohydr. Chem.* **1995**, *14*, 1043.

[37] D. Crich, *J. Org. Chem.* **2011**, *76*, 9193.

[38] B. R. Kowalski, C. F. Bender, *J. Am. Chem. Soc.* **1972**, *94*, 5632.

[39] H. Gelernter, J. R. Rose, C. Chen, *J. Chem. Inf. Comp. Sci.* **1990**, *30*, 492.

[40] R. J. Spinrad, *Science* **1967**, *158*, 55.

[41] B. J. Shields, J. Stevens, J. Li, M. Parasram, F. Damani, J. I. M. Alvarado, J. M. Janey, R. P. Adams, A. G. Doyle, *Nature* **2021**, *590*, 89.

[42] D. T. Ahneman, J. G. Estrada, S. Lin, S. D. Dreher, A. G. Doyle, *Science* **2018**, *360*, 186.

[43] E. King-Smith, *Chem. Sci.* **2024**, *15*, 5143.

[44] J. Yano, K. J. Gaffney, J. Gregoire, L. Hung, A. Ourmazd, J. Schrier, J. A. Sethian, F. M. Toma, *Nat. Rev. Chem.* **2022**, *6*, 357.

[45] Y. Shen, J. E. Borowski, M. A. Hardy, R. Sarpong, A. G. Doyle, T. Cernak, *Nat. Rev. Methods Primers* **2021**, *1*, 23.

[46] S. Bräse, *Digital Discovery* **2024**, *3*, 1923.

[47] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P.

[48] M. L. Schrader, F. R. Schäfer, F. Schäfers, F. Glorius, *Nat. Chem.* **2024**, *16*, 491.

[49] K. M. Jablonka, L. Patiny, B. Smit, *Nat. Chem.* **2022**, *14*, 365.

[50] P. Tremouilhac, A. Nguyen, Y.-C. Huang, S. Kotov, D. S. Lütjohann, F. Hübsch, N. Jung, S. Bräse, *J. Cheminf.* **2017**, *9*, 54.

[51] C. Barillari, D. S. M. Ottoz, J. M. Fuentes-Serna, C. Ramakrishnan, B. Rinn, F. Rudolf, *Bioinformatics* **2015**, *32*, 638.

[52] J. Liu, J. E. Hein, *Nat. Synth.* **2023**, *2*, 464.

[53] M. Christensen, L. P. E. Yunker, F. Adedeji, F. Häse, L. M. Roch, T. Gensch, G. dos Passos Gomes, T. Zepel, M. S. Sigman, A. Aspuru-Guzik, J. E. Hein, *Commun. Chem.* **2021**, *4*, 112.

[54] Y. Shi, P. L. Prieto, T. Zepel, S. Grunert, J. E. Hein, *Acc. Chem. Res.* **2021**, *54*, 546.

[55] M. P. Maloney, C. W. Coley, S. Genheden, N. Carson, P. Helquist, P.-O. Norrby, O. Wiest, *Org. Lett.* **2023**, *25*, 2945.

[56] F. Strieth-Kalthoff, M. Sandfort, M. Kühnemund, F. R. Schäfer, H. Kuchen, F. Glorius, *Angew. Chem., Int. Ed.* **2022**, *61*, e202204647.

[57] W. Beker, R. Roszak, A. Wołos, N. H. Angello, V. Rathore, M. D. Burke, B. A. Grzybowski, *J. Am. Chem. Soc.* **2022**, *144*, 4819.

[58] C. A. Hone, C. O. Kappe, *Chemistry–Methods* **2021**, *1*, 454.

[59] S. M. Kearnes, M. R. Maser, M. Wleklinski, A. Kast, A. G. Doyle, S. D. Dreher, J. M. Hawkins, K. F. Jensen, C. W. Coley, *J. Am. Chem. Soc.* **2021**, *143*, 18820.

[60] D. S. Wigh, J. Arrowsmith, A. Pomberger, K. C. Felton, A. A. Lapkin, *J. Chem. Inf. Model.* **2024**, *64*, 3790.

[61] N. H. Angello, V. Rathore, W. Beker, A. Wołos, E. R. Jira, R. Roszak, T. C. Wu, C. M. Schroeder, A. Aspuru-Guzik, B. A. Grzybowski, M. D. Burke, *Science* **2022**, *378*, 399.

[62] S. Doubet, P. Albersheim, *Glycobiology* **1992**, *2*, 505.

[63] K. Hashimoto, S. Goto, S. Kawano, K. F. Aoki-Kinoshita, N. Ueda, M. Hamajima, T. Kawasaki, M. Kanehisa, *Glycobiology* **2006**, *16*, 63R.

[64] R. Raman, M. Venkataraman, S. Ramakrishnan, W. Lang, S. Raguram, R. Sasisekharan, *Glycobiology* **2006**, *16*, 82R.

[65] T. Lütteke, A. Bohne-Lang, A. Loss, T. Goetz, M. Frank, C.-W. von der Lieth, *Glycobiology* **2006**, *16*, 71R.

[66] C. A. Cooper, H. J. Joshi, M. J. Harrison, M. R. Wilkins, N. H. Packer, *Nucleic Acids Res.* **2003**, *31*, 511.

[67] M. P. Campbell, R. Peterson, J. Mariethoz, E. Gasteiger, Y. Akune, K. F. Aoki-Kinoshita, F. Lisacek, N. H. Packer, *Nucleic Acids Res.* **2013**, *42*, D215.

[68] P. V. Toukach, K. S. Egorova, *Sci. Data* **2022**, *9*, 131.

[69] R. Ranzinger, S. Herget, C.-W. von der Lieth, M. Frank, *Nucleic Acids Res.* **2010**, *39*, D373.

[70] C.-W. von der Lieth, A. A. Freire, D. Blank, M. P. Campbell, A. Ceroni, D. R. Damerell, A. Dell, R. A. Dwek, B. Ernst, R. Fogh, M. Frank, H. Geyer, R. Geyer, M. J. Harrison, K. Henrick, S. Herget, W. E. Hull, J. Ionides, H. J. Joshi, J. P. Kamerling, B. R. Leeflang, T. Lütteke, K. Lundborg, K. Maass, A. Merry, R. Ranzinger, J. Rosen, L. Royle, P. M. Rudd, S. Schloissnig et al., *Glycobiology* **2010**, *21*, 493.

[71] M. Tiemeyer, K. Aoki, J. Paulson, R. D. Cummings, W. S. York, N. G. Karlsson, F. Lisacek, N. H. Packer, M. P. Campbell, N. P. Aoki, A. Fujita, M. Matsubara, D. Shinmachi, S. Tsuchiya, I. Yamada, M. Pierce, R. Ranzinger, H. Narimatsu, K. F. Aoki-Kinoshita, *Glycobiology* **2017**, *27*, 915.

[72] R. Ranzinger, K. F. Aoki-Kinoshita, M. P. Campbell, S. Kawano, T. Lütteke, S. Okuda, D. Shinmachi, T. Shikanai, H. Sawaki, P. Toukach, M. Matsubara, I. Yamada, H. Narimatsu, *Bioinformatics* **2014**, *31*, 919.

[73] I. Yamada, M. Shiota, D. Shinmachi, T. Ono, S. Tsuchiya, M. Hosoda, A. Fujita, N. P. Aoki, Y. Watanabe, N. Fujita, K. Angata, H. Kaji, H. Narimatsu, S. Okuda, K. F. Aoki-Kinoshita, *Nat. Methods* **2020**, *17*, 649.

[74] D. Rana, P. M. Pflüger, N. P. Hölter, G. Tan, F. Glorius, *ACS Cent. Sci.* **2024**, *10*, 899.

[75] S. Moon, S. Chatterjee, P. H. Seeberger, K. Gilmore, *Chem. Sci.* **2021**, *12*, 2931.

[76] B. Fraser-Reid, J. C. López, in *Reactivity Tuning in Oligosaccharide Assembly* (Eds.: B. Fraser-Reid, J. Cristóbal López), Springer Berlin Heidelberg, Berlin, Heidelberg, **2011**, p. 1.

[77] M. Marianski, E. Mucha, K. Greis, S. Moon, A. Pardo, C. Kirschbaum, D. A. Thomas, G. Meijer, G. von Helden, K. Gilmore, P. H. Seeberger, K. Pagel, *Angew. Chem. Int. Ed. Engl.* **2020**, *59*, 6166.

Chemistry–A European Journal

Concept

doi.org/10.1002/chem.202500289

Chemistry Europe
European Chemical
Societies Publishing

[78] S. Chatterjee, S. Moon, F. Hentschel, K. Gilmore, P. H. Seeberger, *J. Am. Chem. Soc.* **2018**, *140*, 11942.

[79] K. D. Collins, A. Rühling, F. Lied, F. Glorius, *Chem. – Eur. J.* **2014**, *20*, 3800.

[80] 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.

[81] J. Danglad-Flores, E. T. Sletten, E. E. Reuber, K. Bienert, H. Riegler, P. H. Seeberger, *Device* **2024**,*2* 100499.

[82] M.-H. Lin, J. B. Wolf, E. T. Sletten, D. Cambié, J. Danglad-Flores, P. H. Seeberger, *ChemBioChem* **2023**, *24*, e202200607.

[83] T. Nokami, R. Hayashi, Y. Saigusa, A. Shimizu, C. Y. Liu, K. K. Mong, J. Yoshida, *Org. Lett.* **2013**, *15*, 4520.

[84] N. V. Ganesh, K. Fujikawa, Y. H. Tan, K. J. Stine, A. V. Demchenko, *Org. Lett.* **2012**, *14*, 3036.

[85] S.-L. Tang, N. L. B. Pohl, *Org. Lett.* **2015**, *17*, 2642.

[86] J. Zhang, C. Chen, M. R. Gadi, C. Gibbons, Y. Guo, X. Cao, G. Edmunds, S. Wang, D. Liu, J. Yu, L. Wen, P. G. Wang, *Angew. Chem. Int. Ed.* **2018**, *57*, 16638.

[87] P. Raghavan, B. C. Haas, M. E. Ruos, J. Schleinitz, A. G. Doyle, S. E. Reisman, M. S. Sigman, C. W. Coley, *ACS Cent. Sci.* **2023**, *9*, 2196.

[88] M. Saebi, B. Nan, J. E. Herr, J. Wahlers, Z. Guo, A. M. Zurański, T. Kogej, P.-O. Norrby, A. G. Doyle, N. V. Chawla, O. Wiest, *Chem. Sci.* **2023**, *14*, 4997.

[89] P. Schwaller, A. C. Vaucher, T. Laino, J.-L. Reymond, *Mach. Learn.: Sci. Technol.* **2021**, *2*, 015016.

[90] D. Lowe, **2017**.

[91] P. Raghavan, A. J. Rago, P. Verma, M. M. Hassan, G. M. Goshu, A. W. Dombrowski, A. Pandey, C. W. Coley, Y. Wang, *J. Am. Chem. Soc.* **2024**, *146*, 15070.

[92] R. Duke, R. McCoy, C. Risko, J. R. S. Bursten, *J. Am. Chem. Soc.* **2024**, *146*, 11579.

[93] F. Strieth-Kalthoff, S. Szymkuć, K. Molga, A. Aspuru-Guzik, F. Glorius, B. A. Grzybowski, *J. Am. Chem. Soc.* **2024**, *146*, 11005.

[94] O. T. Tuck, E. T. Sletten, J. Danglad-Flores, P. H. Seeberger, *Angew. Chem., Int. Ed.* **2022**, *61*, e202115433.

[95] R. Joeres, D. Bojar, O. V. Kalinina, *J. Cheminf.* **2023**, *15*, 37.

[96] Z. Zhang, I. R. Ollmann, X.-S. Ye, R. Wischnat, T. Baasov, C.-H. Wong, *J. Am. Chem. Soc.* **1999**, *121*, 734.

[97] C. W. Cheng, Y. Zhou, W. H. Pan, S. Dey, C. Y. Wu, W. L. Hsu, C. H. Wong, *Nat. Comm.* **2018**, *9*, 5202.

[98] M.-H. Lin, Y.-T. Kuo, J. Danglad-Flores, E. T. Sletten, P. H. Seeberger, *Chem. – Eur. J.* **2024**, *30*, e202400479.

[99] N. V. Faurschou, V. Friis, P. Raghavan, C. M. Pedersen, C. W. Coley, *ChemRxiv* **2024**.

[100] P. Schwaller, T. Laino, T. Gaudin, P. Bolgar, C. A. Hunter, C. Bekas, A. A. Lee, *ACS Cent. Sci.* **2019**, *5*, 1572.

[101] G. Pesciullesi, P. Schwaller, T. Laino, J.-L. Reymond, *Nat. Comm.* **2020**, *11*, 4874.

[102] 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.

[103] 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.

[104] N. V. Faurschou, R. H. Taaning, C. M. Pedersen, *Chem. Sci.* **2023**, *14*, 6319.

[105] W. S. York, S. Agravat, K. F. Aoki-Kinoshita, R. McBride, M. P. Campbell, C. E. Costello, A. Dell, T. Feizi, S. M. Haslam, N. Karlsson, K.-H. Khoo, D. Kolarich, Y. Liu, M. Novotny, N. H. Packer, J. C. Paulson, E. Rapp, R. Ranzinger, P. M. Rudd, D. F. Smith, W. B. Struwe, M. Tiemeyer, L. Wells, J. Zaia, C. Kettner, *Glycobiology* **2014**, *24*, 402.