



Delft University of Technology

Document Version

Final published version

Citation (APA)

van Ede, J. M. (2026). *Beneath the Sweet Surface: A Deep Dive into Microbial Glycomolecule Precursors and Their Enzymes*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:1e81c778-ba4b-44ad-b21d-8d756f828cc5>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

Beneath the Sweet Surface



A Deep Dive into Microbial
Glycomolecule Precursors and Their
Enzymes

Jitske Marianne van Ede

Beneath the Sweet Surface

A Deep Dive into Microbial Glycomolecule
Precursors and Their Enzymes

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology,
by the authority of the Rector Magnificus,
Prof. dr. ir. H. Bijl,
chair of the Board for Doctorates

to be defended publicly on
Thursday, 15 October 2026, 10:00

by

Jitske Marianne VAN EDE

This dissertation has been approved by the promotor.

Composition of the doctoral committee:

Rector Magnificus

Prof. dr. ir. M.C.M. van Loosdrecht

Dr. M. Pabst

Chairperson

Delft University of Technology, promotor

Delft University of Technology, promotor

Independent members:

Prof. dr.ir. P.A.S. Daran-Lapujade

Prof. dr. S.V. Albers

Prof. dr. M.Ø. Arntzen

Prof. dr. D. Lefeber

Dr. C.E. Paul

Prof.dr. U. Hanefeld

Delft University of Technology

University of Freiburg, Germany

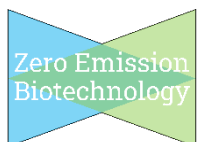
Norwegian University of Life Sciences, Norway

Radboud UMC

Delft University of Technology

Delft University of Technology, reserve member

The research presented in this thesis was performed at the Environmental Biotechnology Section, Department of Biotechnology, Faculty of Applied Sciences, Delft University of Technology, The Netherlands. This work was supported by the Zero Emission Biotechnology Program (ZEB) of the Department of Biotechnology, Delft University of Technology. Finally, chapter 6 of this thesis was supported by the company IFF, Leiden, The Netherlands.



Key words:

mass spectrometry, metabolomics, nucleotide-sugars,
carbohydrate-active enzymes

Printed by:

Proefschriftspecialist

Cover design by:

Jitske van Ede

ISBN:

978-94-6518-436-4

Copyright © 2026 by Jitske Marianne van Ede

An electronic version of this dissertation is available at <https://repository.tudelft.nl/>

Contents

Summary	iv
Samenvatting	vi
1. Introduction.....	1
2. Decoding Sugars: Mass Spectrometric Advances in the Analysis of the Sugar Alphabet	19
3. SugarBase: mapping glycomolecule precursors in microbes	97
4. Measuring Temporal Variations of Nucleotide Pools in Microbial Granular Biofilm	143
5. Microbial extracts as accessible source of CMP-NuLOs for glycoconjugate production, and the demonstration of broad substrate promiscuity for <i>A. baumannii</i> KpsS1	175
6. Discovery of microbial glycoside hydrolases via enrichment and metaproteomics.....	205
7. Outlook	249

Summary

The surface of microorganisms is often decorated with a wide variety of glycans and glycoconjugates. These biomolecules play key roles in structural support, cellular communication, immune recognition, and host–pathogen interactions. Despite their importance, glycan analysis still lags far behind that of DNA and proteins. This is mainly because glycans are not directly encoded in the genome but arise from the combined action of glycosyltransferases and glycosidases. Their characterization is further complicated by extensive structural diversity, as glycans consist of a large and expanding repertoire of monosaccharides. In addition, restricted access to these building blocks hinders functional studies and the exploitation of glycans as antimicrobial targets. This thesis aims to advance the discovery of microbial metabolites, glycomolecule precursors, and carbohydrate-active enzymes using mass spectrometry and bioinformatic approaches.

Chapter 2 reviews advanced mass spectrometry-based methods for the identification and quantification of monosaccharides and their nucleotide-activated forms. Key analytical challenges are highlighted, including the differentiation of stereoisomers and the limited availability of reference standards. Complementary strategies, such as high-performance separation techniques, advanced fragmentation methods, and isotope tracing, are discussed for improved structural characterization and pathway analysis.

Building on these analytical foundations, **Chapter 3** presents SugarBase, a mass spectrometry and bioinformatic pipeline for untargeted exploration of microbial nucleotide sugar networks. SugarBase combines narrow-window DIA fragmentation with a theoretical sugar composition database for parent ion annotation. The platform revealed species-specific nucleotide sugar profiles, including distinct nonulosonic acid profiles in *Campylobacter jejuni* strains and previously unannotated compounds across diverse microbes, including a higher-carbon ulosonic acid in *Magnetospirillum*.

Extending this work to complex microbial systems, **Chapter 4** includes a collaborative study adapting the Chapter 3 methodology into a targeted metabolomics workflow to quantify nucleotides in granular biofilms involved in enhanced biological phosphorus removal. Using fast quenching, boiling-water extraction, and high-resolution mass spectrometry with ^{13}C -labeled internal

standards, the approach enabled measurement of adenylate and uridylate energy charge dynamics. Applied to a lab-scale system, it revealed energy pool fluctuations during acetate uptake and polyphosphate degradation, demonstrating that energy imbalances can underlie metabolic switching in biofilms.

In **Chapter 5**, we demonstrate that microbial metabolite extracts can serve as effective and readily accessible substrate sources for glycosyltransferase assays, without further purification. Using a *Campylobacter jejuni* extract containing diverse nonulosonic acid derivatives, we show substrate promiscuity of the previously identified pseudaminyltransferase from *Acinetobacter baumannii*, currently the only confirmed enzyme of its kind. These findings support the production of glycans and glycoconjugates relevant for antimicrobial and vaccine development.

While glycans and glycoconjugates are central to many biological processes, carbohydrate-active enzymes are of similar interest due to their potential to replace harsh chemical processes and their ability to enable new products. A key challenge remains the targeted identification of enzymes with desired activities and characteristics. In **Chapter 6**, we combine enrichment cultures with metagenomics and metaproteomics to discover glycoside hydrolases. This strategy enables functional enrichment under customized conditions and yields a focused set of enzyme candidates. One of the candidates was subsequently confirmed as a pullulan-degrading enzyme using a mass spectrometry-based activity assay.

Finally, **Chapter 7** summarizes some of the main findings, discusses main challenges and outlines future research directions.

Samenvatting

Het oppervlak van alle cellen, waaronder dat van micro-organismen, is bedekt met diverse glycanen (suiker structuren) en glycoconjugaten. Deze moleculen spelen een fundamentele rol in de celstructuur en in cellulaire communicatie, daarnaast zijn ze vaak ook essentieel voor het wel of niet activeren van een afweerreactie en in interacties tussen gastheer en pathogeen. Ondanks hun belang blijft onze kennis van deze biomoleculen, evenals de beschikbare analysemethoden die nodig zijn om ze te bestuderen, sterk achter ten opzichte van DNA, eiwitten en lipiden. Dit komt voornamelijk doordat glycanen en glycoconjugaten niet direct door het genoom worden gecodeerd, maar ontstaan door de gecoördineerde samenwerking van verschillende enzymen. Bovendien wordt hun karakterisering bemoeilijkt door hun enorme chemische en structurele diversiteit, waaronder variaties in ketenlengte, vertakkingspatronen, stereochemie, bindings-isomerie en modificaties van hydroxylgroepen. Het gebrek aan efficiënte chemische en biochemische synthesesmethoden beperkt zowel fundamenteel onderzoek naar hun functie als de toepassing van deze moleculen in de geneeskunde, waaronder de ontwikkeling van nieuwe antimicrobiële middelen.

Dit proefschrift heeft tot doel de ontdekking van microbiële metabolieten, bouwstenen van suikerstructuren en koolhydraat-actieve enzymen te bevorderen door middel van massaspectrometrie en bio-informatica.

Hoofdstuk 2 biedt een literatuuroverzicht van geavanceerde massaspectrometrische methoden voor de identificatie en kwantificatie van monosachariden, hun derivaten en nucleotide-geactiveerde vormen, waarbij gefocust wordt op micro-organismen. Daarbij worden belangrijke analytische uitdagingen besproken, waaronder het onderscheiden van stereo-isomeren en de beperkte beschikbaarheid van referentiestoffen voor validatie. Complementaire strategieën komen aan bod, zoals hoogwaardige scheidingstechnieken, geavanceerde fragmentatiemethoden, isotopentracerings en computationele benaderingen voor de ontdekking van microbiële suiker structuren, structurele karakterisering en het ontrafelen van bio-synthetische routes.

Voortbouwend op deze analytische ontwikkelingen introduceert **Hoofdstuk 3** SugarBase, een platform dat massaspectrometrie en bio-informatica integreert voor het ongericht in kaart brengen van microbiële nucleotide-suikernetwerken in complexe microbiële monsters. SugarBase combineert 'narrow-window' DIA-

fragmentatie met een uitgebreide theoretische nucleotide-suikerdatabank om zowel bekende als eerder onbeschreven suikermetabolieten te identificeren. De toepassing op fylogenetisch diverse micro-organismen onthulde soort-specifieke nucleotide-suikerprofielen. Hierbij werden onder meer stam-specifieke nonulosonische zuur modificaties waargenomen in de pathogene bacterie *Campylobacter jejuni*, evenals verschillende niet eerder gerapporteerde nucleotide-suikers, waaronder een vermoedelijk hoger-koolstof α -ketozuur dat is geïdentificeerd in *Magnetospirillum*. Samen benadrukken deze bevindingen de opmerkelijke en grotendeels nog onontdekte diversiteit van het microbiële nucleotide-suiker metabolisme.

Hoofdstuk 4 bevat een samenwerkingsstudie waarin de methodologie uit hoofdstuk 3 werd aangepast tot een gerichte metabolomics-workflow voor het kwantificeren van nucleotiden in granulaire biofilms van een microbiële gemeenschap afkomstig uit afvalwater. Met behulp van snelle ‘quenching’ (het inactiveren van metabolische processen), het toevoegen van ^{13}C -gelabelde interne standaarden, een extractie met kokend water, en hoge-resolutie massaspectrometrie, maakte deze aanpak het mogelijk om de dynamiek van de adenylaat en uridylaat energielading te bepalen. Toegepast op een lab-schaal systeem werden fluctuaties in energievoorraden waargenomen tijdens acetaatopname en polyfosfaatafbraak, wat aantoont dat energie-onevenwichtigheden ten grondslag kunnen liggen aan metabole omschakelingen in microbiële ecosystemen.

In **Hoofdstuk 5** wordt aangetoond dat microbiële metabolietextracten kunnen dienen als direct toegankelijke substraatbron voor het testen van glycosyltransferase activiteit, zonder verdere zuivering. Met behulp van een *Campylobacter jejuni*-extract dat diverse nonulosonische zuur derivaten bevat, laten we de substraatpromiscuïteit van de eerder beschreven pseudaminyltransferase uit *Acinetobacter baumannii* zien, momenteel het enige bevestigde enzym van dit type. Deze bevindingen ondersteunen de productie van nieuwe glycoconjugaten, met potentiële relevantie voor antimicrobiële toepassingen, welke momenteel vaak niet toegankelijk zijn via chemische synthese.

Naast het feit dat glycanen en glycoconjugaten een centrale rol spelen in vele biologische processen, hebben koolhydraat-actieve enzymen potentieel om traditionele chemische processen te vervangen en nieuwe producten te genereren.

Een belangrijke uitdaging blijft de gerichte identificatie van enzymen met gewenste activiteiten en eigenschappen. In **Hoofdstuk 6** worden verrijkingculturen gecombineerd met metagenomics en metaproteomics om nieuwe glycosidehydrolasen uit microbiële milieugemeenschappen te ontdekken. Deze strategie maakt functionele verrijking onder gecontroleerde omstandigheden mogelijk voor de identificatie van nieuwe enzymen die onder gedefinieerde condities actief zijn. Eén van de kandidaten werd na recombinante expressie bevestigd als een nieuw pullulan-afbrekend enzym.

Tot slot bespreekt **Hoofdstuk 7** de belangrijkste uitdagingen binnen het onderzoek naar microbiële glycanen en glycoconjugaten en belicht het nieuwe richtingen en recente ontwikkelingen binnen microbiële omics-technologieën.

Introduction

1

1.1. The unexplored microbial diversity

Life on Earth displays extraordinary diversity. Although approximately one million insect species, 11,000 bird species, and 7,000 mammal species have been described to date¹, these visible organisms represent only a small fraction of the planet's biodiversity. Hidden from the naked eye lies a vast microbial world whose diversity and functional potential far exceed that of macroscopic life. The earliest surviving reference to these "invisible little creatures" dates back to the first century BC, when the Roman polymath Marcus Terentius Varro proposed their existence². It was not until the second half of the 17th century that microorganisms were first observed using primitive microscopes by Robert Hooke and Antonie van Leeuwenhoek³. Currently, it is estimated that up to 10^{12} microbial species inhabit Earth, yet despite centuries of research, the vast majority, approximately 99.999%, remain undiscovered⁴.

Microorganisms inhabit nearly every ecological niche on this planet, from soil and oceans to the gastrointestinal tracts of animals, as well as extreme environments such as soda lakes, deserts, acidic mine drainage sites and hydrothermal vents⁵. Adaptation to these diverse habitats has led to microorganisms with distinct physiological requirements; while some thrive in the presence of oxygen, others cannot tolerate it. As a result, microbes display a wide range of optimal growth conditions, including variations in temperature, salinity, pH, and pressure⁶. This diversity underpins their broad metabolic capabilities and survival strategies, offering significant potential to address major societal challenges.

1.2. Societal challenges

The production of everyday goods, including paper, textiles, food, chemicals, and pharmaceuticals, requires large amounts of raw materials and energy and generates substantial waste⁷. Over the past six decades, the global population has grown rapidly, from approximately three billion people in 1960 to more than eight billion today, and, despite an expected slowdown, is projected to approach ten billion by 2100⁸. This demographic growth, combined with rising wealth and shifting consumption patterns, has intensified the demand for goods and services and increased pressure on the planet's finite resources.

This pressure is reflected in rising greenhouse gas emissions, as well as increasing pollution and waste accumulation that threaten human health and ecosystems^{9,10}. Addressing these challenges requires reducing environmental impact while continuing to meet societal needs. Industrial biotechnology offers a promising

approach by complementing or replacing conventional technologies, thereby contributing to more sustainable manufacturing processes.

Another pressing challenge is the rise of antimicrobial resistance, which negatively affects human health and well-being and also carries substantial economic consequences¹¹. This underscores the need to better understand microbial mechanisms and ecosystems to enable the development of alternative antimicrobial strategies.

1

1.3. Exploiting Microbial Potential

Humans have used microorganisms for millennia, with evidence of sugar fermentation into alcohol dating back to 7000 BC¹², but their role was only recognized in the 18th century with the identification of yeast as the agent of fermentation¹³. Since then, microorganisms have been harnessed as “cell factories” and are widely used to produce compounds such as vitamins, organic acids, amino acids, alcohols, flavor compounds, and pharmaceuticals¹³. Compared to conventional chemical processes, biotechnological production typically relies on renewable feedstocks, operates under mild conditions, and reduces the use of toxic reagents, often resulting in a lower environmental impact^{7,14}.

This is exemplified in the pulp and paper industry, where enzymes, including glycoside hydrolases, can reduce environmental impact compared to conventional methods. For instance, thermomechanical pulping is highly energy-intensive, but the addition of cellulases, which soften wood fibers, can lower energy consumption. Similarly, deinking of recycled paper pulp typically relies on large amounts of harsh chemicals. Enzymatic treatment reduces both chemical usage as well as processing time. Overall, the introduction of enzymes in the pulp and paper industry has been shown to decrease its impact on global warming¹⁵.

Despite these advantages, biotechnological processes are not inherently more sustainable, and trade-offs must be considered. While fermentation can reduce greenhouse gas emissions compared to petrochemical routes, comprehensive life-cycle assessments are essential. Improvements in one impact category may be offset by increases in others, such as eutrophication¹⁶ or land use⁷.

Beyond environmental considerations, microbial production enables access to structurally complex molecules that are difficult to synthesize chemically, particularly glycans and glycoconjugates. The formation of glycosidic bonds requires strict control of regio- and stereoselectivity. This complicates chemical synthesis, since it often necessitates multiple activation, protection, and deprotection steps, which reduce overall yields. In contrast, microbial production

exploits enzymes with inherent selectivity to convert simple substrates into complex oligosaccharides in a single biological process, offering a more efficient and scalable alternative¹⁷. It can also improve product homogeneity and enhance safety by reducing the use of toxic reagents¹⁸. Examples include human milk oligosaccharides (HMOs) for infant nutrition and glycosaminoglycans (GAGs), which are used in cosmetics, healthcare, and clinical applications¹⁷⁻¹⁹.

1.4. The diverse world of carbohydrates

Glycans are ubiquitous biomolecules that occur either as free polysaccharides or covalently attached to proteins and lipids, forming a structurally and functionally diverse class collectively referred to as glycoconjugates. Their composition varies markedly across biological systems, reflecting differences in cellular architecture and function²⁰.

In prokaryotes, glycans occur in several glycoconjugate forms, including N- and O-linked glycoproteins. For a long time, protein glycosylation was believed to be restricted to eukaryotes. In the mid-1970s, the first experimental evidence for a prokaryotic glycoprotein was reported in the cell envelope of *Halobacterium salinarum*²¹, and only in the early 2000s was a general N-linked glycosylation system identified in *Campylobacter jejuni*²². N-linked glycoproteins contain glycans attached to the amide nitrogen of asparagine residues, whereas O-linked glycoproteins carry glycans linked to the hydroxyl groups of serine or threonine residues. One of the main differences between eukaryotic and prokaryotic glycosylation is the diversity of monosaccharide building blocks, which is considerably greater in prokaryotes. In bacteria, glycans are also major components of the cell envelope. A key structural element is peptidoglycan, a polymer of repeating disaccharide units cross-linked by short peptides that provides mechanical strength to the cell wall. Many bacteria also produce capsular polysaccharides (CPS) and exopolysaccharides, high-molecular-weight carbohydrate polymers that surround the cell and contribute to protection and environmental interactions^{20,23}.

The glycan composition of the bacterial cell envelope differs between Gram-negative and Gram-positive species. Gram-negative bacteria possess an outer membrane containing lipooligosaccharides (LOS) and lipopolysaccharides (LPS), which are glycolipids consisting of a lipid anchor linked to oligo- or polysaccharide chains, respectively. In contrast, Gram-positive bacteria lack an outer membrane

and instead feature a thicker peptidoglycan layer together with additional glycopolymers such as lipoteichoic and wall teichoic acids^{24,25} (Figure 1.1).

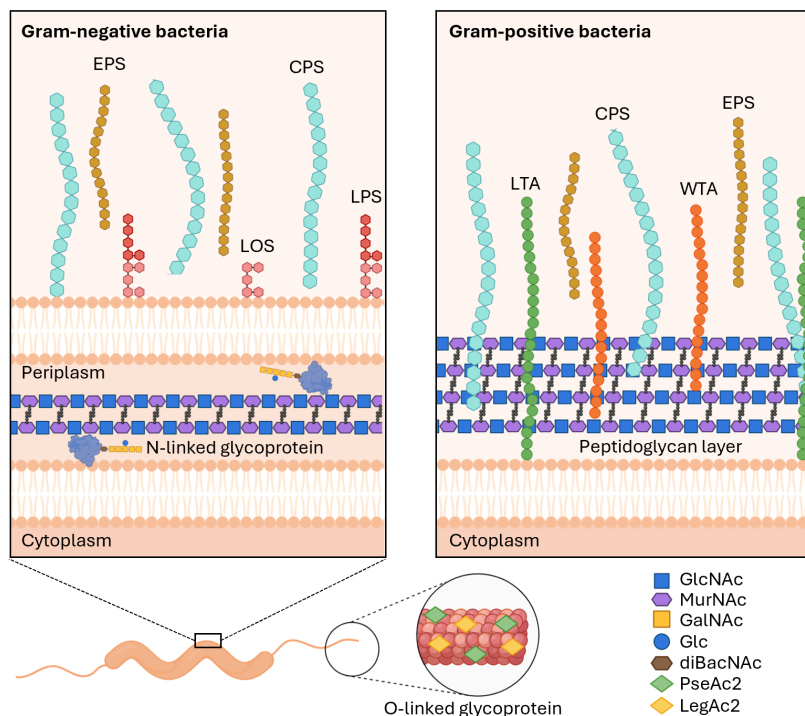


Figure 1.1. The major glycan and glycoconjugate structures in Gram-negative and Gram-positive bacteria, respectively. Abbreviations: **CPS**: capsular polysaccharides; **EPS**: extracellular polysaccharides; **LOS**: lipooligosaccharides; **LPS**: lipopolysaccharides; **LTA**: lipoteichoic acids; **WTA**: wall teichoic acids. **GlcNAc**: N-acetylglucosamine; **MurNAc**: N-acetylmuramic acid; **GalNAc**: N-acetylgalactosamine; **Glc**: glucose; **diBacNAc**: di-N-acetylated bacillosamine; **PseAc2**: di-N-acetylated pseudaminic acid; **LegAc2**: di-N-acetylated legionaminic acid.

1.4.1. Glycans display a variety of biological functions

During the late 19th and early 20th centuries, glycans and glycoconjugates were primarily regarded as structural or metabolic molecules. Glycogen, identified in 1857, was recognized as a glucose-based storage polymer²⁶, while starch was established as the principal energy storage polysaccharide in plants²⁷. Additional structural roles were assigned to polymers such as chitin in invertebrate exoskeletons, and lipopolysaccharides, which contribute to the integrity of the outer membrane of Gram-negative bacteria²⁸.

By the mid-20th century, evidence began to emerge for more intrinsic functional roles of glycans. A notable example was that the incompatibility of certain blood groups in transfusion settings was linked to variation in glycosylation²⁹. In 1988, the

term “glycobiology” was introduced, reflecting advances in analytical techniques and a growing appreciation of glycan structural diversity and functional relevance²⁰. Today, glycans and glycoconjugates are recognized as key mediators of diverse biological processes, extending from structural support and protein folding to immune recognition, signal transduction, and cellular communication²⁸. They can protect proteins from proteolytic degradation, enhance solubility, and regulate protein stability and half-life²⁸. Glycans also play central roles in host–pathogen interactions, functioning as receptors for viral entry and mediating bacterial adhesion and invasion^{30,31}. In addition, many plant and bacterial toxins, as well as bacteriophages, exploit glycan structures for host recognition and targeting³². Glycans further regulate cellular uptake processes, including endocytosis and phagocytosis²⁸. Finally, numerous pathogens evade immune detection through molecular mimicry, in which host-like glycan structures are used to avoid immune recognition³³.

1.4.2. Why glycan analysis lags far behind genomics and proteomics

Despite substantial progress in glycan research, their analysis still lags far behind that of DNA and proteins. This reflects the long-standing view that biological information flows from DNA to RNA to proteins, with functional output largely genome-encoded²⁰. Although the importance of glycans is now well established, their analysis remains more complex because the glycome is not directly encoded in DNA³⁴. Instead, glycans are secondary gene products generated through the combined action of glycosyltransferases and glycosidases. Consequently, they cannot be directly manipulated via mutagenesis, which complicates functional analysis³⁵.

Structural characterization is further complicated by the extensive diversity of glycans. Unlike DNA, RNA, and proteins, which are composed of a limited set of nucleotides or amino acids³⁴, glycans consist of a large and expanding repertoire of monosaccharides. Approximately 100 distinct monosaccharides are currently catalogued in the KEGG database^{36,37}, a number that increases substantially when including modifications such as methylation and acetylation. Moreover, many additional variants are likely yet to be discovered³⁸.

This diversity increases further through variation in linkage positions, α - or β -anomeric configurations, and linear or branched architectures. Consequently, glycans exhibit near-infinite structural variability³⁷; for instance, a reducing hexasaccharide composed solely of hexoses can theoretically form over 10^{12} distinct structures³⁹. Although mammalian systems already display substantial diversity despite a limited monosaccharide repertoire, prokaryotic glycans are even

more structurally diverse³⁸. Glycan analysis is also complicated by microheterogeneity, where a single glycosylation site can be occupied by multiple distinct glycan structures⁴⁰.

1

1.4.3. Nonulosonic acids

Nonulosonic acids (NulOs) are nine-carbon α -keto acid monosaccharides with substantial structural diversity. The first identified member, neuraminic acid, was discovered in salivary mucins and is commonly referred to as sialic acid⁴¹. The most prevalent form is N-acetylneuraminic acid (Neu5Ac), but diversity arises from modifications at the hydroxyl groups at C-4, C-7, C-8, and C-9, including acetyl, methyl, sulfate, and lactyl groups, with more than 50 derivatives described⁴¹.

Sialic acids are typically found at the terminal positions of N- and O-glycans and glycosphingolipids, where they regulate cell–cell recognition, signaling, and immune modulation⁴¹. They can protect underlying proteins from proteolysis and serve as key recognition sites for pathogens and toxins. For example, influenza virus binds to sialylated receptors to initiate infection³⁰. Aberrant sialylation is also a hallmark of many cancer cells⁴¹.

Although sialic acids are abundant in animal tissues, they are also found in prokaryotes, where they can be exploited for immune evasion through molecular mimicry⁴². A notable example is *Campylobacter jejuni*, which expresses sialylated lipooligosaccharides that mimic human gangliosides. While this facilitates immune evasion, it can, in rare cases, trigger autoimmune responses such as the Guillain–Barré syndrome⁴³.

Other NulOs are found exclusively in prokaryotes (Figure 1.2). Among these, legionaminic acid shares the stereochemical backbone of neuraminic acid but contains an N-acetyl group at C7 and lacks the hydroxyl group at C9⁴⁴. Additional stereoisomers include pseudaminic acid, 4- and 8-epilegionaminic acid, acinetaminic acid, 8-epiacinetaminic acid, and fusaminic acid^{45–49}, with further diversity introduced by substitutions such as formyl, glyceryl, hydroxybutyryl, lactoyl, and glutamyl groups⁴¹. These NulOs are commonly associated with surface glycans and proteins, including lipopolysaccharides (LPS), capsular polysaccharides (CPS), pili, and flagella^{44,50}. These sugars structures have been studied primarily in pathogens. While their full functional scope remains unclear, NulOs play important roles in bacterial virulence, promoting adhesion, host colonization, and motility^{31,51}. In addition, as molecular mimics of sialic acids, they contribute to immune evasion by attenuating host immune responses³³. Interestingly, NulOs have also been detected in non-pathogenic prokaryotes,

indicating that these structures likely have roles beyond host–pathogen interactions⁵².

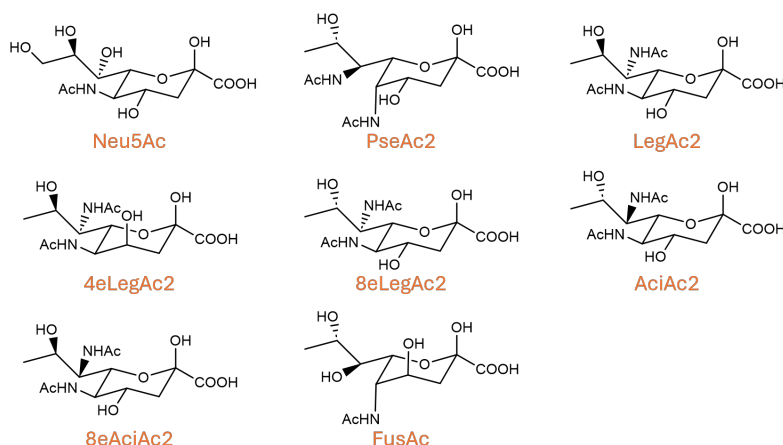


Figure 1.2. The nonulosonic acid structures of the most common derivatives of all identified stereoisomers to date. Abbreviations: **Neu5Ac**: 5-N-acetylneuraminic acid; **PseAc2**: 5,7-di-N-acetyl-pseudaminic acid; **LegAc2**: 5,7-di-N-acetyl-legionaminic acid. **4eLegAc2**: 5,7-di-N-acetyl-4-epilegionaminic acid; **8eLegAc2**: 5,7-di-N-acetyl-8-epilegionaminic acid; **AciAc2**: 5,7-di-N-acetyl-acinetaminic acid; **8eAciAc2**: 5,7-di-N-acetyl-8-epiacinetaminic acid; **FusAc**: 5-N-acetyl fusaminic acid.

1.4.4. Glycans: applications and targets for antimicrobials

While glycan diversity complicates their analysis, it also provides significant potential. Rare sugars have applications across the food, aroma, cosmetics, and pharmaceutical industries⁵³. For example, D-tagatose is used as a low-calorie sweetener⁵⁴, and D-allose has been explored as a protective agent against ischemic injury⁵³. Likewise, glycan polymers are widely applied, such as heparin, which is a standard anticoagulant drug⁵⁵, and human milk oligosaccharides which are used in infant nutrition¹⁷.

Glycans are promising targets for antimicrobial and vaccine development due to their abundant and often strain-specific surface expression on pathogens. With increasing antibiotic resistance, including WHO-classified critical multidrug-resistant pathogens such as *Acinetobacter baumannii*⁵⁶, alternative strategies are urgently required. Glycoconjugate vaccines are already clinically successful, with licensed products targeting *Neisseria meningitidis* and *Streptococcus pneumoniae*⁵⁷. Current research focuses on novel glycoconjugate vaccines and improved production platforms, including recombinant expression in *Escherichia coli* rather than extraction from pathogens^{58,59}.

NulOs are also attractive antimicrobial targets due to their surface exposure and prokaryote specificity⁶⁰. Initial studies have shown promising results for

pseudaminic acid-based vaccines against antibiotic-resistant *Acinetobacter baumannii*^{61,62}. However, progress in both vaccine development and functional studies remains limited by restricted access to NulO-containing glycans and glycoconjugates^{63,64}.

Moreover, the enzymes in the metabolic pathways are promising drug targets as well. For example, 5-(4-chlorophenyl)-2-furoic acid was found to inhibit the enzymes involved in the dTDP-L-rhamnose biosynthesis in several pathogenic bacteria, including *Streptococcus pyogenes* and *Mycobacterium tuberculosis*. dTDP-L-Rhamnose is a critical precursor for the viability and virulence mechanisms of these pathogens, and since dTDP-L-rhamnose is absent in humans, its inhibition is a promising antibacterial approach⁶⁵.

Accordingly, advancing glycan-based biotechnology depends not only on expanding structural knowledge of microbial glycans but also on resolving their biosynthetic pathways. This is essential for understanding glycan function and enabling scalable production for industrial and biomedical applications.

1.5. Integrated omics disciplines provide insight into biological systems

The emergence of high throughput “omics” technologies has transformed biological research by enabling comprehensive analysis of biomolecules across different functional levels. Disciplines such as genomics, proteomics, and metabolomics provide complementary perspectives, from genetic potential to functional activity. Together, they support the discovery of novel biomolecules, the characterization of enzymes, and the elucidation of metabolic pathways⁶⁶. However, each omics layer captures only a part of the biological complexity, underscoring the need for integrated multi-omics approaches (Figure 1.3).

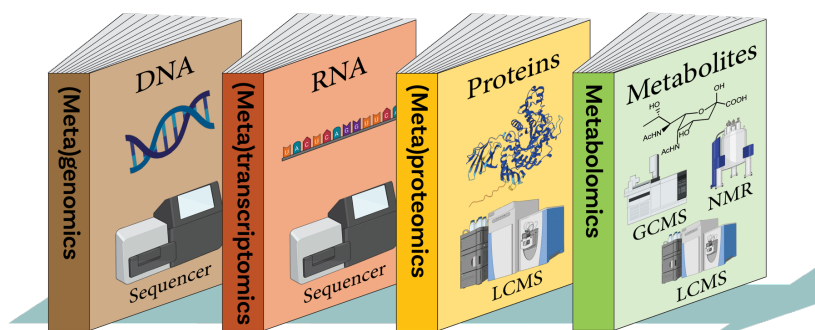


Figure 1.3. Integrating omics technologies enables a more comprehensive view of biological systems. (Meta)genomics, (meta)transcriptomics, (meta)proteomics and metabolomics study an organism’s complete set of DNA, RNA, proteins and metabolites, respectively.

1.5.1. (Meta) genomics: genetic potential of biological systems

Genomics is the study of an organism's complete set of DNA. Key milestones include the elucidation of the DNA double helix by Watson and Crick in 1953, followed by the development of sequencing technologies such as Sanger sequencing^{67,68}. Early approaches were labor-intensive and costly, as exemplified by the Human Genome Project, which required more than a decade to complete⁶⁹. The advent of next-generation sequencing technologies, including Illumina and Oxford Nanopore, has dramatically increased throughput while reducing costs, enabling routine large-scale genome analysis⁷⁰. These advances have facilitated the rise of metagenomics, which analyzes the collective genomes of microbial communities. By enabling direct sequencing of environmental DNA, metagenomics bypasses the need for cultivation, a major advantage given that most microorganisms are not readily culturable under standard laboratory conditions⁷¹.

Metagenomics provides insights into both the taxonomic composition and functional potential of biological systems and has enabled enzyme discovery and metabolic pathway reconstruction through homology-based annotation and functional screening⁷¹. However, challenges remain, including incomplete genome assemblies, uneven sequencing depth, and a high proportion of genes with unknown function^{72,73}. Importantly, gene presence does not necessarily reflect expression or activity, highlighting the need for complementary analyses.

1.5.2. Central role of mass spectrometry in omics disciplines

Mass spectrometry (MS) is a central analytical platform in functional omics. Its origins trace back to J. J. Thomson in 1897, who demonstrated the existence of the electron through measurements of mass-to-charge ratios, leading to the first mass spectrometers⁷⁴. MS gained importance during the Manhattan Project, where it was used for isotope separation, and by the 1940s it had become commercially available with applications in fields such as the petrochemical industry^{74,75}.

Early MS was applied mainly to small molecules, and advances in fragmentation analysis in the 1960s enabled structural elucidation⁷⁴. However, its use in biology remained limited due to extensive fragmentation of large, labile biomolecules. This limitation was overcome in the mid-1980s with the introduction of soft ionization techniques, particularly electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI), developed by John Fenn and Koichi Tanaka^{76,77}. These approaches enabled ionization of biomolecules such as peptides and lipids,

while ESI also allowed direct coupling to liquid-phase separations⁷⁶, marking a major turning point for biological MS applications^{74,75}.

Modern MS employs a range of analyzers with complementary performance characteristics. High-resolution instruments such as time-of-flight (TOF), Fourier-transform ion cyclotron resonance (FT-ICR), and Orbitrap analyzers provide high mass accuracy and resolving power, making them suitable for complex biological samples⁷⁸. In contrast, quadrupoles and ion traps offer lower resolution but higher sensitivity and are widely used for quantitative analysis, with ion traps additionally enabling extensive fragmentation for structural characterization. Tandem mass spectrometry (MS/MS), which enables controlled ion fragmentation and fragment analysis, has become essential for structural omics applications⁷⁹.

The Orbitrap is widely used in metabolomics and proteomics. It consists of a central spindle electrode surrounded by two cup-shaped outer electrodes, which create the ion trapping field and act as receiver plates for image current detection (Figure 1.4)⁸⁰. The introduction of the C-trap, enabling ion accumulation prior to injection, allowed coupling of the discontinuously operated Orbitrap with continuous ion sources such as ESI. Injected ions oscillate around the central electrode of the Orbitrap, and their axial oscillation frequency is measured as an image current and Fourier-transformed into a mass spectrum. The commercialization of Orbitrap-based tandem MS systems in 2005 enabled controlled fragmentation and acquisition of MS/MS spectra, significantly advancing (meta)proteomics⁸⁰.

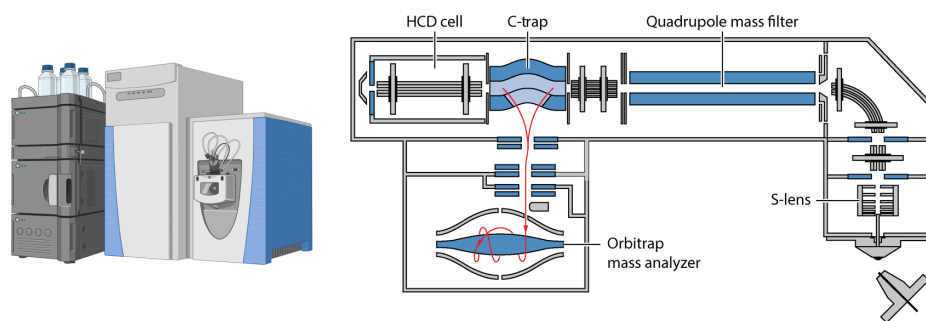


Figure 1.4. Liquid chromatography coupled with a high resolution tandem mass spectrometer.

Schematic overview of the Quadrupole Orbitrap hybrid mass spectrometer used in this thesis. Ions can be filtered by the Quadrupole mass filter and are either directly analyzed by the Orbitrap mass analyzer (MS1), or after fragmentation in the HCD cell (MS2). Figure adapted from Eliuk and Makarov (2015)⁸⁰.

1.5.3. MS-based proteomics: linking genes to proteins

Proteomics aims to characterize the complete set of proteins in a biological system, providing a functional link between genetic information and cellular

activity. A widely used approach is bottom-up proteomics, in which proteins are enzymatically digested, typically with trypsin, into peptides prior to MS analysis. Sample complexity is commonly reduced by liquid chromatography or SDS-PAGE prior to tandem mass spectrometry⁸¹.

Protein identification is based on interpretation of MS/MS spectra, which typically contain b- and y-type ions generated by peptide backbone fragmentation, retaining charges on the N- and C-terminal, respectively⁸¹. Identification can be performed either de novo or via spectral matching. In de novo sequencing, complete or near-complete ion series allow reconstruction of peptide sequences from mass differences between adjacent peaks. In spectral matching, experimental spectra are compared against peptide databases, providing a robust alternative when fragmentation is incomplete⁸¹.

Post-translational modifications (PTMs) add an additional layer of complexity. Glycoproteins are particularly challenging due to glycan heterogeneity⁸². Collision-based fragmentation methods such as collision-induced dissociation (CID) and higher-energy collision dissociation (HCD) mainly generate glycan-derived fragments, which inform on glycan composition, whereas electron-based methods such as electron transfer dissociation (ETD) preferentially fragment the peptide backbone while preserving glycans, enabling site-specific localization of glycosylation⁸². The combination of complementary fragmentation strategies is therefore essential for comprehensive glycoprotein characterization⁸³.

Although proteomics provides detailed information on protein identity and abundance, it does not directly capture biochemical activity. Consequently, it offers limited insight into dynamic metabolic states, motivating the use of metabolomics as a complementary omics layer.

1.5.4. Metabolomics: functional readout of biological systems

Metabolomics focuses on the comprehensive analysis of small molecules (typically <1500 Da) involved in cellular metabolism, such as sugars, amino acids, lipids, and nucleotides. The complete set of these metabolites are known as the metabolome. As it reflects the downstream outcome of gene expression and protein activity, it provides a direct snapshot of the physiological state of a biological system⁸⁴⁻⁸⁶.

The first microbial metabolomics study was reported in 1998, and the field has since expanded rapidly through advances in analytical methods and computational data analysis^{87,88}. Today, metabolomics is widely applied to identify and quantify metabolites, compare metabolic profiles across conditions, and elucidate metabolic pathways and biochemical fluxes⁸⁶.

A typical workflow includes sample preparation, data acquisition, and data analysis. Due to the rapid turnover of metabolites, rapid sampling and quenching are essential to preserve metabolic states, followed by efficient extraction. The choice of extraction method depends on factors such as cell structure and metabolite properties and remains challenging in untargeted approaches due to the broad chemical diversity of metabolites^{85,89}.

The main analytical platforms in metabolomics are mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy^{84,85,88}. MS is typically coupled to chromatographic separation (GC–MS or LC–MS), which reduces ion suppression and improves resolution. GC–MS offers high separation efficiency and robustness but often requires derivatization and is best suited for volatile or low-molecular-weight compounds. LC–MS has become more dominant in the past decade due to its broader applicability, particularly for polar and non-volatile metabolites. Different mass analyzers, including quadrupole, TOF, Orbitrap, and ion traps, provide complementary strengths in sensitivity, resolution, and structural elucidation. Although less sensitive, NMR delivers highly reproducible, quantitative data and is particularly valuable for structural characterization^{84,85,88}.

Analytical strategies can be divided into untargeted approaches, which aim to detect as many metabolites as possible for discovery purposes, and targeted approaches, which focus on the accurate quantification of predefined compounds^{84,85}. Stable isotope labeling further enables metabolic flux analysis, providing insight into pathway activity and regulation⁸⁶. Advances in machine learning and artificial intelligence are increasingly supporting data analysis and metabolite annotation, although their performance depends on the availability of large, high-quality datasets⁹⁰.

The metabolome is highly diverse, with classes such as lipids exhibiting extensive structural variability. This has led to specialized subdisciplines including lipidomics and glycomics. Glycomics addresses the comprehensive analysis of glycans, whether free, protein-bound, or lipid-bound, with further subdivisions such as glycolipidomics and glycoproteomics, the latter of which is considered a branch of proteomics²⁰.

Overall, metabolomics is a powerful tool for pathway elucidation, biomarker discovery, and the characterization of system-level responses to environmental changes. When integrated with genomics and proteomics, it enables a more complete understanding of biological systems by linking genetic potential to functional output.

1.6. Scope and Aim of This Thesis

As discussed above, glycans and glycoconjugates, have important biological functions, but their study is hindered by restricted access to these complex structures. The overarching goal of this thesis was to develop advanced mass spectrometry and data processing approaches, that provide information beyond common genomics methods, and that consequently accelerate the discovery of key microbial metabolites, glycomolecule precursors and carbohydrate-active enzymes.

Chapter 2 reviews advanced mass spectrometric methods for the identification and quantification of monosaccharides, their derivatives, and activated forms, with a focus on microbial systems. It highlights key analytical challenges and discusses complementary strategies, including high-performance separations, advanced fragmentation techniques, and isotope tracing for improved structural characterization and pathway analysis.

Chapter 3 presents a mass spectrometry and bioinformatic pipeline (SugarBase) for the large-scale discovery of microbial nucleotide-sugars, which serve as precursors to glycans and glycoconjugates. SugarBase addresses the metabolic complexity of microbes by combining a narrow-window data independent acquisition fragmentation with a theoretical sugar composition database for parent ion identification.

Chapter 4 describes collaborative work in which the approach developed in Chapter 3 was adapted into a targeted metabolomics method to quantify nucleotides in granular biomass.

The aim of **Chapter 5** is to explore microbial extracts, as analyzed in Chapter 3, as readily accessible sources of nucleotide sugar donors for glycosyltransferase assays, with a particular focus on nonulosonic acid transferases. This strategy circumvents the need for often complex or impractical chemical or enzymatic synthesis of substrates. In addition, the substrate promiscuity of the investigated nonulosonic acid transferase is examined.

Finally, **Chapter 6** describes an approach for identifying novel glycoside hydrolases under defined conditions by integrating microbial enrichment with metagenomics and metaproteomics.

References

- 1 International Union for Conservation of Nature. The IUCN Red List of Threatened Species. Retrieved from: https://nc.iucnredlist.org/redlist/content/attachment_files/2025-1_RL_Table_1a.pdf. (2025).
- 2 Cheesman, D. Varro and the Small Beasts: A Bimillennium for Microbiologists. *Nature* **203**, 911–912 (1964). <https://doi.org/https://doi.org/10.1038/203911a0>
- 3 Gest, H. The discovery of microorganisms by Robert Hooke and Antoni Van Leeuwenhoek, fellows of the Royal Society. *Notes Rec R Soc Lond* **58**, 187–201 (2004). <https://doi.org/10.1098/rsnr.2004.0055>
- 4 Locey, K. J. & Lennon, J. T. Scaling laws predict global microbial diversity. *P Natl Acad Sci USA* **113**, 5970–5975 (2016). <https://doi.org/10.1073/pnas.1521291113>
- 5 Shu, W. S. & Huang, L. N. Microbial diversity in extreme environments. *Nat Rev Microbiol* **20**, 219–235 (2022). <https://doi.org/10.1038/s41579-021-00648-y>
- 6 Wani, A. K., Akhtar, N., Sher, F., Navarrete, A. A. & Américo-Pinheiro, J. H. P. Microbial adaptation to different environmental conditions: molecular perspective of evolved genetic and cellular systems. *Archives of Microbiology* **204** (2022). <https://doi.org/10.1007/s00203-022-02757-5>
- 7 Jegannathan, K. R. & Nielsen, P. H. Environmental assessment of enzyme use in industrial production - a literature review. *J Clean Prod* **42**, 228–240 (2013). <https://doi.org/10.1016/j.jclepro.2012.11.005>
- 8 United Nations, Department of Economic and Social Affairs, Population Division (2024). *World Population Prospects 2024*. Accessed March 4, 2026. <https://population.un.org/wpp/World>
- 9 Ripple, W. J. *et al.* The 2025 state of the climate report: a planet on the brink. *BioScience* **75**, 1016–1027 (2025). <https://doi.org/10.1093/biosci/biaf149>
- 10 Wang, Z., Walker, G. W., Muir, D. C. G. & Nagatani-Yoshida, K. Toward a Global Understanding of Chemical Pollution: A First Comprehensive Analysis of National and Regional Chemical Inventories. *Environ Sci Technol* **54**, 2575–2584 (2020). <https://doi.org/10.1021/acs.est.9b06379>
- 11 Dadgostar, P. Antimicrobial Resistance: Implications and Costs. *Infect Drug Resist* **12**, 3903–3910 (2019). <https://doi.org/10.2147/IDR.S234610>
- 12 McGovern, P. E. *et al.* Fermented beverages of pre- and proto-historic China. *P Natl Acad Sci USA* **101**, 17593–17598 (2004). <https://doi.org/10.1073/pnas.0407921102>
- 13 Buchholz, K. & Collins, J. The roots--a short history of industrial microbiology and biotechnology. *Appl Microbiol Biotechnol* **97**, 3747–3762 (2013). <https://doi.org/10.1007/s00253-013-4768-2>
- 14 Becker, M., Zieminska-Stolarska, A., Markowska, D., Lutz, S. & Rosenthal, K. Comparative Life Cycle Assessment of Chemical and Biocatalytic 2'3'-Cyclic GMP-AMP Synthesis. *ChemSusChem* **16**, e202201629 (2023). <https://doi.org/10.1002/cssc.202201629>
- 15 Skals, P. B., Krabek, A., Nielsen, P. H. & Wenzel, H. Environmental Assessment of Enzyme Assisted Processing in Pulp and Paper Industry. *The International Journal of Life Cycle Assessment* **13**, 124–132 (2008). <https://doi.org/10.1016/j.jclepro.2012.11.005>
- 16 Zuiderveen, E. A. R. *et al.* The potential of emerging bio-based products to reduce environmental impacts. *Nat Commun* **14** (2023). <https://doi.org/10.1038/s41467-023-43797-9>
- 17 Palur, D. S. K., Pressley, S. R. & Atsumi, S. Microbial Production of Human Milk Oligosaccharides. *Molecules* **28** (2023). <https://doi.org/10.3390/molecules28031491>
- 18 Liu, K. F., Guo, L., Chen, X. L., Liu, L. M. & Gao, C. Microbial synthesis of glycosaminoglycans and their oligosaccharides. *Trends Microbiol* **31**, 369–383 (2023). <https://doi.org/10.1016/j.tim.2022.11.003>
- 19 Zhou, W., Jiang, H., Wang, L., Liang, X. & Mao, X. Biotechnological Production of 2'-Fucosyllactose: A Prevalent Fucosylated Human Milk Oligosaccharide. *ACS Synth Biol* **10**, 447–458 (2021). <https://doi.org/10.1021/acssynbio.0c00645>
- 20 Varki, A. & Kornfeld, S. in *Essentials of Glycobiology* Ch. 1 Historical Background and Overview, (2022).
- 21 Mescher, M. F. & Strominger, J. L. Purification and characterization of a prokaryotic glycoprotein from the cell envelope of *Halobacterium salinarum*. *J Biol Chem* **251**, 2005–2014 (1976).
- 22 Wacker, M. *et al.* N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* **298**, 1790–1793 (2002). <https://doi.org/10.1126/science.298.5599.1790>
- 23 Tytgat, H. L. & Lebeer, S. The sweet tooth of bacteria: common themes in bacterial glycoconjugates. *Microbiol Mol Biol Rev* **78**, 372–417 (2014). <https://doi.org/10.1128/MMBR.00007-14>
- 24 Madigan, M. T., Martinko, J., Bender, K. S., Buckley, D. H. & Stahl, D. A. *Brock Biology of Microorganisms*. 14 edn, 65–69 (Pearson, 2015).
- 25 Whitfield, C., Szymanski, C. M., Lewis, A. L. & Aebi, M. *Chapter 21 Eubacteria - Essentials of Glycobiology*. (Cold Spring Harbor, 2022).
- 26 Young, F. G. Bernard, Claude and the Discovery of Glycogen - a Century of Retrospect. *Bmj-Brit Med J* **1**, 1431–1437 (1957). <https://doi.org/DOI10.1136/bmj.1.5033.1431>
- 27 Zeeman, S. C., Kossmann, J. & Smith, A. M. Starch: its metabolism, evolution, and biotechnological modification in plants. *Annu Rev Plant Biol* **61**, 209–234 (2010). <https://doi.org/10.1146/annurev-arplant-042809-112301>
- 28 Varki, A. Biological roles of glycans. *Glycobiology* **27**, 3–49 (2017). <https://doi.org/10.1093/glycob/cww086>
- 29 Watkins, W. M. & Morgan, W. T. J. Inhibition by Simple Sugars of Enzymes Which Decompose the Blood-Group Substances. *Nature* **175**, 676–677 (1955). <https://doi.org/DOI10.1038/175676a0>
- 30 Weis, W. *et al.* Structure of the Influenza-Virus Hemagglutinin Complexed with Its Receptor, Sialic-Acid. *Nature* **333**, 426–431 (1988). <https://doi.org/DOI10.1038/333426a0>
- 31 Smedley, J. I. I. *et al.* Influence of Pilin Glycosylation on *Pseudomonas aeruginosa* 1244 Pilus Function. *Infect Immun* **73**, 7922–7931 (2005). <https://doi.org/10.1128/iai.73.12.7922-7931.2005>
- 32 Sacher, J. C. *et al.* Binding of Phage-Encoded FlaGrab to Motile *Campylobacter jejuni* Flagella Inhibits Growth, Downregulates Energy Metabolism, and Requires Specific Flagellar Glycans. *Front Microbiol* **11**, 397 (2020). <https://doi.org/10.3389/fmicb.2020.00397>
- 33 Stephenson, H. N. *et al.* Pseudaminic acid on *Campylobacter jejuni* flagella modulates dendritic cell IL-10 expression via Siglec-10 receptor: a novel flagellin-host interaction. *J Infect Dis* **210**, 1487–1498 (2014). <https://doi.org/10.1093/infdis/jiu287>

- 34 Marth, J. D. A unified vision of the building blocks of life. *Nat Cell Biol* **10**, 1015–1016 (2008). <https://doi.org/10.1038/ncb0908-1015>
- 35 Zol-Hanlon, M. I. & Schumann, B. Open questions in chemical glycobiology. *Commun Chem* **3** (2020). <https://doi.org/10.1038/s42004-020-00337-6>
- 36 KEGG. <<https://www.genome.jp/kegg/tables/br08082.html>> (2026).
- 37 Cummings, R. D. A periodic table of monosaccharides. *Glycobiology* **34** (2024). <https://doi.org/10.1093/glycob/cwad088>
- 38 Imperiali, B. Bacterial carbohydrate diversity - a Brave New World. *Curr Opin Chem Biol* **53**, 1–8 (2019). <https://doi.org/10.1016/j.cbpa.2019.04.026>
- 39 Laine, R. A. A Calculation of All Possible Oligosaccharide Isomers Both Branched and Linear Yields 1.05×10^{12} Structures for a Reducing Hexasaccharide - the Isomer-Barrier to Development of Single-Method Saccharide Sequencing or Synthesis Systems. *Glycobiology* **4**, 759–767 (1994). <https://doi.org/10.1093/glycob/4.6.759>
- 40 Varki, A. et al. *Essentials of Glycobiology*. 4th edn, (2009).
- 41 Lewis, A. L., Chen, X., Schnaar, R. L. & Varki, A. in *Essentials of Glycobiology* Ch. 51 Sialic Acids and Other Nonulosonic Acids, (2022).
- 42 Jennings, M. P., Day, C. J. & Atack, J. M. How bacteria utilize sialic acid during interactions with the host: snip, snatch, dispatch, match and attach. *Microbiol-Sgm* **168** (2022). <https://doi.org/10.1099/mic.0.001157>
- 43 Godschalk, P. C. et al. Structural characterization of Campylobacter jejuni lipooligosaccharide outer cores associated with Guillain-Barre and Miller Fisher syndromes. *Infect Immun* **75**, 1245–1254 (2007). <https://doi.org/10.1128/IAI.00872-06>
- 44 Knirel, Y. A., Rietschel, E. T., Marre, R. & Zahringer, U. The Structure of the O-Specific Chain of Legionella-Pneumophila Serogroup-1 Lipopolysaccharide. *Eur J Biochem* **221**, 239–245 (1994). <https://doi.org/DOI 10.1111/j.1432-1033.1994.tb18734.x>
- 45 Knirel, Y. A. et al. Sialic Acids of a New Type from the Lipopolysaccharides of Pseudomonas-Aeruginosa and Shigella-Boydii. *Carbohydr Res* **133**, C5–C8 (1984). [https://doi.org/DOI 10.1016/0008-6215\(84\)85213-1](https://doi.org/DOI 10.1016/0008-6215(84)85213-1)
- 46 Kenyon, J. J., Marzaioli, A. M., De Castro, C. & Hall, R. M. 5,7-di-N-acetyl-acinetaminic acid: A novel non-2-ulosonic acid found in the capsule of an Acinetobacter baumannii isolate. *Glycobiology* **25**, 644–654 (2015). <https://doi.org/10.1093/glycob/cww007>
- 47 Vinogradov, E., Michael, F. S. & Cox, A. D. The structure of the LPS O-chain of Fusobacterium nucleatum strain 25586 containing two novel monosaccharides, 2-acetamido-2,6-dideoxy-L-altrose and a 5-acetimidoamino-3,5,9-trideoxy-gluco-non-2-ulosonic acid. *Carbohydr Res* **440–441**, 10–15 (2017). <https://doi.org/10.1016/j.carres.2017.01.002>
- 48 Kenyon, J. J., Notaro, A., Hsu, L. Y., De Castro, C. & Hall, R. M. 5,7-Di-N-acetyl-8-epiacinetaminic acid: A new non-2-ulosonic acid found in the K73 capsule produced by an Acinetobacter baumannii isolate from Singapore. *Sci Rep-Uk* **7** (2017). <https://doi.org/10.1038/s41598-017-11166-4>
- 49 Knirel, Y. A. et al. Identification of a Homopolymer of 5-Acetamidino-7-acetamido-3,5,7,9-tetradeoxy-D-glycero-D-talo-nonulosonic Acid in the Lipopolysaccharides of Legionella pneumophila Non-1 Serogroups. *Biochemistry-Moscow* **66**, 1035–1041 (2001). <https://doi.org/10.1023/A:1012334012605>
- 50 Thibault, P. et al. Identification of the carbohydrate moieties and glycosylation motifs in Campylobacter jejuni flagellin. *J Biol Chem* **276**, 34862–34870 (2001). <https://doi.org/10.1074/jbc.M104529200>
- 51 Guerry, P. et al. Changes in flagellin glycosylation affect Campylobacter autoagglutination and virulence. *Mol Microbiol* **60**, 299–311 (2006). <https://doi.org/10.1111/j.1365-2958.2006.05100.x>
- 52 Kleikamp, H. B. C. et al. Tackling the chemical diversity of microbial nonulosonic acids - a universal large-scale survey approach. *Chem Sci* **11**, 3074–3080 (2020). <https://doi.org/10.1039/c9sc06406k>
- 53 Poonperm, W. et al. Efficient conversion of allitol to D-psicose by Bacillus pallidus Y25. *J Biosci Bioeng* **103**, 282–285 (2007). <https://doi.org/10.1263/jbb.103.282>
- 54 Livesey, G. & Brown, J. C. D-tagatose is a bulk sweetener with zero energy determined in rats. *J Nutr* **126**, 1601–1609 (1996). <https://doi.org/DOI 10.1093/jn/126.6.1601>
- 55 Oduah, E. I., Linhardt, R. J. & Sharfstein, S. T. Heparin: Past, Present, and Future. *Pharmaceuticals (Basel)* **9** (2016). <https://doi.org/10.3390/ph9030038>
- 56 WHO. WHO publishes list of bacteria for which new antibiotics are urgently needed, <<https://www.who.int/publications/i/item/WHO-EMP-IAU-2017.12>> (2017).
- 57 Rappuoli, R. Glycoconjugate vaccines: Principles and mechanisms. *Sci Transl Med* **10** (2018). <https://doi.org/10.1126/scitranslmed.aat4615>
- 58 Kämpf, M. M. et al. n vivo production of a novel glycoconjugate vaccine against Shigella flexneri 2a in recombinant Escherichia coli: identification of stimulating factors for in vivo glycosylation. *Microb Cell Fact* **14** (2015). <https://doi.org/10.1186/s12934-015-0195-7>
- 59 Herbert, J. A. et al. Production and efficacy of a low-cost recombinant pneumococcal protein polysaccharide conjugate vaccine. *Vaccine* **36**, 3809–3819 (2018). <https://doi.org/10.1016/j.vaccine.2018.05.036>
- 60 Zunk, M. K., M. J. The occurrence and biological significance of the α-keto-sugars pseudaminic acid and legionaminic acid within pathogenic bacteria. *Rsc Adv* **4**, 3413–3421 (2014).
- 61 Lee, I. M. et al. Pseudaminic Acid on Exopolysaccharide of Acinetobacter baumannii Plays a Critical Role in Phage-Assisted Preparation of Glycoconjugate Vaccine with High Antigenicity. *J Am Chem Soc* **140**, 8639–8643 (2018). <https://doi.org/10.1021/jacs.8b04078>
- 62 Wei, R. et al. Synthetic Pseudaminic-Acid-Based Antibacterial Vaccine Confers Effective Protection against Acinetobacter baumannii Infection. *ACS Cent Sci* **7**, 1535–1542 (2021). <https://doi.org/10.1021/acscentsci.1c00656>
- 63 Guo, X., Li, P., Liu, H. & Li, X. Recent progress on the chemical synthesis of bacterial non-2-ulosonic acids. *Organic Chemistry Frontiers* **10**, 3429–3446 (2023).
- 64 Flack, E. K. P., Chidwick, H. S., Best, M., Thomas, G. H. & Fascione, M. A. Synthetic Approaches for Accessing Pseudaminic Acid (Pse) Bacterial Glycans. *Chembiochem* **21**, 1397–1407 (2020). <https://doi.org/10.1002/cbic.202000019>
- 65 van der Beek, S. L. et al. Streptococcal dTDP-L-rhamnose biosynthesis enzymes: functional characterization and lead compound identification. *Mol Microbiol* **111**, 951–964 (2019). <https://doi.org/10.1111/mmi.14197>
- 66 Morabito, A., De Simone, G., Pastorelli, R., Brunelli, L. & Ferrario, M. Algorithms and tools for data-driven omics integration to achieve multilayer biological insights: a

- narrative review. *J Transl Med* **23**, 425 (2025). <https://doi.org/10.1186/s12967-025-06446-x>
- 67 Watson, J. D. & Crick, F. H. C. The Structure of DNA. *Cold Spring Harb Sym* **18**, 123–131 (1953). <https://doi.org/Doi10.1101/Sqb.1953.018.01.020>
 - 68 Sanger, F., Nicklen, S. & Coulson, A. R. DNA Sequencing with Chain-Terminating Inhibitors. *P Natl Acad Sci USA* **74**, 5463–5467 (1977). <https://doi.org/DOI10.1073/pnas.74.12.5463>
 - 69 Gibbs, R. A. The Human Genome Project changed everything. *Nat Rev Genet* **21**, 575–576 (2020). <https://doi.org/10.1038/s41576-020-0275-3>
 - 70 Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* **17**, 333–351 (2016). <https://doi.org/10.1038/nrg.2016.49>
 - 71 Uchiyama, T. & Miyazaki, K. Functional metagenomics for enzyme discovery: challenges to efficient screening. *Curr Opin Biotechnol* **20**, 616–622 (2009). <https://doi.org/10.1016/j.copbio.2009.09.010>
 - 72 van der Walt, A. J. *et al.* Assembling metagenomes, one community at a time. *BMC Genomics* **18**, 521 (2017). <https://doi.org/10.1186/s12864-017-3918-9>
 - 73 Sevin, D. C., Fuhrer, T., Zamboni, N. & Sauer, U. Nontargeted in vitro metabolomics for high-throughput identification of novel enzymes in *Escherichia coli*. *Nat Methods* **14**, 187–194 (2017). <https://doi.org/10.1038/nmeth.4103>
 - 74 Griffiths, J. A brief history of mass spectrometry. *Anal Chem* **80**, 5678–5683 (2008). <https://doi.org/10.1021/ac8013065>
 - 75 Smoluch, M. & Silberring, J. A Brief History of Mass Spectrometry. *Wil Ser Mass Spectr*, 5–8 (2019).
 - 76 Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray Ionization for Mass-Spectrometry of Large Biomolecules. *Science* **246**, 64–71 (1989). <https://doi.org/DOI10.1126/science.2675315>
 - 77 Tanaka, K. *et al.* Protein and polymer analyses up to *m/z* 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2**, 151–153 (1988). <https://doi.org/10.1002/rcm.1290020802>Digital Object Identifier (DOI)
 - 78 Xian, F., Hendrickson, C. L. & Marshall, A. G. High Resolution Mass Spectrometry. *Anal Chem* **84**, 708–719 (2012). <https://doi.org/10.1021/ac203191t>
 - 79 Haag, A. M. Mass Analyzers and Mass Spectrometers. *Adv Exp Med Biol* **919**, 157–169 (2016). https://doi.org/10.1007/978-3-319-41448-5_7
 - 80 Eliuk, S. & Makarov, A. Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annu Rev Anal Chem* **8**, 61–80 (2015). <https://doi.org/10.1146/annurev-anchem-071114-040325>
 - 81 Steen, H. & Mann, M. The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Bio* **5**, 699–711 (2004). <https://doi.org/10.1038/nrm1468>
 - 82 Halim, A. & Anonsen, J. H. Microbial glycoproteomics. *Curr Opin Struc Biol* **44**, 143–150 (2017). <https://doi.org/10.1016/j.sbi.2017.03.005>
 - 83 Rudd, P. M. *et al.* in *Essentials of Glycobiology* Ch. 51 Glycomics and Glycoproteomics, (2022).
 - 84 Liu, X. J. & Locasale, J. W. Metabolomics: A Primer. *Trends Biochem Sci* **42**, 274–284 (2017). <https://doi.org/10.1016/j.tibs.2017.01.004>
 - 85 Hajnajafi, K. & Iqbal, M. A. Mass-spectrometry based metabolomics: an overview of workflows, strategies, data analysis and applications. *Proteome Sci* **23** (2025). <https://doi.org/10.1186/s12953-025-00241-8>
 - 86 Jang, C., Chen, L. & Rabinowitz, J. D. Metabolomics and Isotope Tracing. *Cell* **173**, 822–837 (2018). <https://doi.org/10.1016/j.cell.2018.03.055>
 - 87 Tweeddale, H., Notley-McRobb, L. & Ferenci, T. Effect of slow growth on metabolism of *Escherichia coli*, as revealed by global metabolite pool ("metabolome") analysis. *J Bacteriol* **180**, 5109–5116 (1998). <https://doi.org/10.1128/Jb.180.19.5109-5116.1998>
 - 88 Ye, D. Y., Li, X. W., Shen, J. Z. & Xia, X. Microbial metabolomics: From novel technologies to diversified applications. *Trac-Trend Anal Chem* **148** (2022). <https://doi.org/10.1016/j.trac.2022.116540>
 - 89 Putri, S. P. *et al.* Current metabolomics: Practical applications. *J Biosci Bioeng* **115**, 579–589 (2013). <https://doi.org/10.1016/j.jbiosc.2012.12.007>
 - 90 Chi, J. H. *et al.* Artificial intelligence in metabolomics: a current review. *Trac-Trend Anal Chem* **178** (2024). <https://doi.org/10.1016/j.trac.2024.117852>

Decoding Sugars: Mass Spectrometric Advances in the Analysis of the Sugar Alphabet

Jitske M. van Ede, Dinko Soic and Martin Pabst



Abstract

Monosaccharides play a central role in metabolic networks and in the biosynthesis of glycomolecules, which perform essential functions across all domains of life. Thus, identifying and quantifying these building blocks is crucial in both research and industry. Routine methods have been established to facilitate the analysis of common monosaccharides. However, despite the presence of common metabolites, most organisms utilize distinct sets of monosaccharides and derivatives. These molecules therefore display a large diversity, potentially numbering in the hundreds or thousands, with many still unknown. This complexity presents significant challenges in the study of glycomolecules, particularly in microbes, including pathogens and those with the potential to serve as novel model organisms. This review discusses mass spectrometric techniques for the isomer-sensitive analysis of monosaccharides, their derivatives, and activated forms. Although mass spectrometry allows for untargeted analysis and sensitive detection in complex matrices, the presence of stereoisomers and extensive modifications necessitates the integration of advanced chromatographic, electrophoretic, ion mobility, or ion spectroscopic methods. Furthermore, stable-isotope incorporation studies are critical in elucidating biosynthetic routes in novel organisms.

Published as:

van Ede, J. M., Soic, D., & Pabst, M. (2026). Decoding Sugars: Mass Spectrometric Advances in the Analysis of the Sugar Alphabet. *Mass Spectrometry Reviews*, 45(2), 282-333.

2.1 | Introduction

While nucleic acids and proteins have a limited number of building blocks (8 and 22, respectively¹⁻³) carbohydrates can comprise a highly diverse spectrum of monosaccharides⁴⁻⁷. Notably, monosaccharides and their derivatives could number in the hundreds or even thousands, with many yet to be discovered⁸. For example, the Kyoto Encyclopedia of Genes and Genome database (KEGG) lists approximately 100 monosaccharides that have been identified in pathways across all organisms mentioned in the database^{5,9}. One of the most abundant monosaccharides in cells is glucose, which is metabolized to produce ATP, the energy currency of the cell¹⁰. The metabolic intermediates are precursors for a spectrum of other monosaccharides. These sugars are not directly involved in the energy metabolism, but are required to produce a large variety of other glycoconjugates, as well as free oligo- and polysaccharides^{6,11}. Their analysis is challenged by the diversity of carbohydrate building blocks, including numerous connectivity possibilities and stereochemical configurations. Consequently, the study of oligo- and polysaccharides is still less developed than that of nucleic acids and proteins¹².

Carbohydrate oligo- and polymers are abundant and omnipresent, such as cellulose in plant tissues, chitin in the exoskeleton of invertebrates, lactose in milk, and raffinose in legumes^{13,14}. Smaller, but structurally diverse oligomers can be attached to proteins, lipids, and, as recently discovered, to small RNAs forming the so called “glycoRNAs”^{11,15}. These glycans are involved in regulating protein folding and turnover, or are exposed on the cells exterior where they are involved in signaling and immune responses¹¹. The extracellular space is also rich in glycosaminoglycans, large carbohydrate polymers carrying modifications such as phosphorylation and sulfation. Glycosaminoglycans provide structural support to cells and tissues, facilitate cell signaling, and contribute to the integrity of the extracellular matrix¹⁶⁻¹⁸. In addition, free oligosaccharides, such as those found in milk, show probiotic effects, defend against pathogens and support the immune system^{12,14}. In microbes, they are also involved in host/pathogen interactions, protection, and contribute to adaptability and survival in competitive environments¹⁹⁻²¹.

The biocatalytic incorporation of the individual monosaccharides into oligosaccharides and conjugates requires prior activation of the sugars. This is commonly achieved by linking them to nucleotides – forming the so-called nucleotide sugars (glycosyl esters of nucleoside mono- or pyrophosphates) – and in some rare cases to dolichol phosphate (e.g. dol-Man and dol-Glc precursors in the endoplasmic reticulum)¹¹.

Since monosaccharides, their derivatives and activated forms are crucial building blocks in many biological processes, the identification and quantification of these molecules are essential in both research and industry. Quantitative metabolomics studies and ¹³C flux analysis have emerged as powerful techniques, such as for fundamental studies on model organisms, and for evaluating metabolic

engineering efforts in cell factories^{22,23}. Cell factories are promising alternatives to chemical synthesis, in producing precursors for pharmaceuticals and biopolymers^{24,25}. In the pharmaceutical industry, monitoring the monosaccharide composition of recombinant therapeutic glycoproteins is crucial for ensuring their efficacy and safety²⁶. In addition, in medical applications, monosaccharide analysis aids in diagnosing metabolic disorders such as galactosemia and mucopolysaccharidoses²⁷. Furthermore, cellular metabolism is significantly altered in cancer cells compared to normal cells, enabling the former to sustain high rates of proliferation²⁸. Finally, in environmental studies, monosaccharide analysis helps in tracking organic matter sources and biodegradation processes in soil and water samples²⁹.

While routine methods have been established for many of the aforementioned applications over the past decades, expansions to the microbial world beyond a few well-studied model organisms remains a challenge. Especially for uncultured microbes such as those involved in global biogeochemical cycles, members of the human microbiome, and potential symbiotic or pathogenic microbes, the monosaccharide building blocks and their metabolic routes are, to date, poorly understood. Microbes typically produce many species-specific monosaccharide derivatives, likely influenced by their ecological niche and nutrient availability. Using these monosaccharides, glycoconjugates are produced and presented on their cell surface, often forming important virulence factors and components of the cell wall³⁰. However, the composition and the specific activated substrates involved in the production of most microbial glycomolecules remain largely unexplored. The same holds true for plants, where the biosynthetic routes of essential sugars of the plant cell wall, such as 3-deoxy-lyxo-2-heptulosaric acid and aceric acid (3-carboxy-5-deoxy-xylofuranose), have not yet been identified^{31,32}. Additionally, the role of activation with unusual nucleotides, such as thymidine diphosphate, is currently unknown³².

However, the uniqueness of microbial monosaccharide derivatives, makes their metabolic pathways promising drug targets, increasing the interest in these sugars. For example, the enzyme non-hydrolyzing C2 epimerase, which is involved in the production of UDP-Man2NAc and UDP-l-Fuc2NAc, is found in most archaea and bacteria, but is absent in human cells. Inhibitors of this enzyme have shown promising results against methicillin-resistant *Staphylococcus aureus*³³. The identification of novel monosaccharides has also led to the discovery of new antibiotics. For example, the biosynthesis of the antibiotic erythromycin involves a series of glycosylation steps catalyzed by specific glycosyltransferases, a process that is difficult to replicate via chemical synthesis³⁴. In addition, the identification of rare sugars – such as D-tagatose, used as a pharmaceutical and low-calorie sweetener – has led to the discovery of specific enzymes to catalyze their synthesis³⁵. Generally, carbohydrate-active enzymes are of significant interest in the development of biotechnological applications involving carbohydrate molecules³⁶.

A substantial number of reviews have focused on the mass spectrometric characterization of various oligo- and polysaccharides, glycoproteins, glycolipids, and other glycomolecules^{12,37-52}. Therefore, this review focuses on advanced mass spectrometric approaches facilitating the identification and quantification of the fundamental monosaccharide building blocks, their derivatives, and activated forms. Mass spectrometry enables not only untargeted analysis but also the detection of sub-stoichiometric quantities in highly complex matrices. However, the mass spectrometric analysis of sugar molecules is hampered by the presence of many stereoisomers and the extensive molecular diversification, which is particularly prevalent in microbes. This necessitates the integration of high-performance separation techniques such as liquid and gas chromatography or capillary electrophoresis alongside mass spectrometric detection for many applications. On the other hand, sophisticated fragmentation experiments, ion mobility measurements, and infrared spectroscopy can provide additional structural information. Stable isotope tracing and metabolic network analysis have also emerged as powerful tools for exploring the biosynthetic routes of sugars in novel species. Finally, this review highlights emerging medical, biotechnological, and industrial applications.

2.2 | Monosaccharide chemistry and diversity

Simple monosaccharides are typically chains of chiral hydroxymethylene units and follow the general empirical formula $C_n(H_2O)_n$, where n ranges from 3 to 9, or in rare cases, up to 10. They terminate with a hydroxymethyl group at one end and contain either an aldehyde group (aldoses) or an α -hydroxy ketone group (ketoses) at the other end^{6,11,53}. Accordingly, glyceraldehyde is the simplest aldose, and dihydroxyacetone is the simplest ketose. Apart from dihydroxyacetone, all monosaccharides possess a number of asymmetric carbon atoms equal to $n-2$ for aldoses and $n-3$ for ketoses, where n is the number of carbon atoms^{6,11,53}. Consequently, the number of stereoisomers for each monosaccharide is 2 to the power of the number of asymmetric carbon atoms. For example, an aldohexose (with a total of six carbons) has four asymmetric carbon atoms ($n-2$ for aldoses) and, therefore, it has $2^4 = 16$ possible isomeric forms.

Sugars can form cyclic structures called furanose and pyranose rings. Furanose rings are five-membered structures composed of four carbon atoms and one oxygen atom, while pyranose rings are six-membered structures comprising five carbon atoms and one oxygen atom. Hexoses, such as glucose, predominantly form pyranose rings but can also adopt a furanose structure, especially in certain derivatives or under specific conditions. Similarly, heptoses, containing seven carbon atoms, can form both furanose and pyranose rings depending on environmental conditions. In aqueous solutions, sugars exist in equilibrium between their cyclic and open-chain forms. In the case of aldohexoses, the pyranose form is created when the hydroxyl group on C-5 attacks the carbonyl carbon (C-1), forming a hemiacetal linkage. This reaction can occur from either side

of the carbonyl group, leading to the formation of either the α - or β -anomer (Figure 2.1).

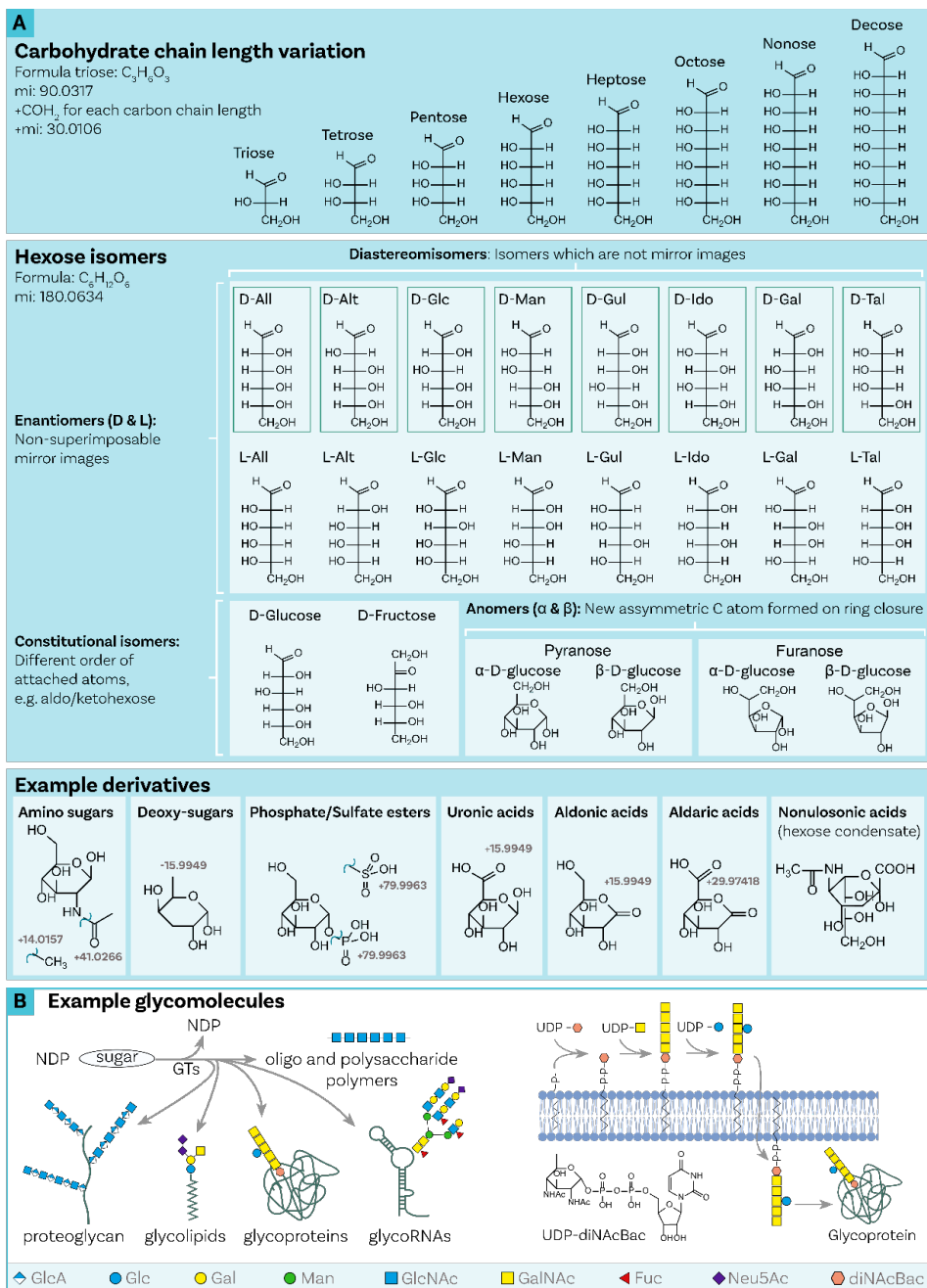


Figure 2.1. Diversity of monosaccharide building blocks. **A)** The diversity of monosaccharides, highlighting variations in carbon atom count, stereochemistry—including diastereomers, enantiomers, and constitutional isomers—the positioning of the carbonyl group, which differentiates aldoses from ketoses, and isomerization resulting from ring formation (α/β pyranose and furanose

forms). **B)** The biocatalytic incorporation of monosaccharides into various glycomolecules. e.g., into proteoglycans, glycolipids, glycoproteins, or glycoRNAs, or free poly and oligosaccharides, by glycosyltransferases (GTs) requires the prior activation of these sugars. This activation is typically achieved by converting them into nucleotide activated sugars (e.g., UDP-diNAcBac), which are glycosyl esters of nucleoside mono- or pyrophosphates (NDPs). As an example, in *Campylobacter jejuni* a heptasaccharide is sequentially assembled on a polyisoprenyl diphosphate carrier from activated monosaccharides and then transferred to the asparagine side chain of acceptor proteins. **All:** allose; **Alt:** altrose; **diNAcBac:** di-N-acetyl-bacillosamine; **Gal:** galactose; **Glc:** glucose; **GalNAc:** N-acetylgalactosamine; **GTs:** glycosyltransferases; **Gul:** gulose; **Ido:** idose; **Man:** mannose; **Tal:** talose; **UDP:** uridine diphosphate.

The diversity of monosaccharides is even further enhanced via modifications of the hydroxyl or carboxyl groups. For example, sugar-specific amino- or amidotransferases introduce an amino functionality to sugar phosphates or sugar nucleotides, forming the so called aminosugars⁵³. Free amine groups in sugars are rare (only found in some glycosaminoglycans), and are commonly acetylated to N-acetamido groups in, for example, the widely found GlcNAc, GalNAc, and NeuNAc monosaccharides. Other frequent modifications include the replacement of hydroxyl groups with hydrogen atoms to form deoxysugars⁵³. Common deoxy sugars include 2-deoxyribose, which is part of the DNA, and 6-deoxy-galactose (fucose), a widely utilized monosaccharides in vertebrates¹¹. In microbes, many dideoxy sugars are found, such as abequose, a 3,6-dideoxyhexose present in the lipopolysaccharides of bacteria including *Citrobacter* and *Pectobacterium*^{54,55}. Other enzymes esterify the hydroxyl groups to produce phosphate esters, esters with acetic acid or fatty acids, and sulfate esters⁵³. The phosphorylation of sugars makes them generally more reactive. Moreover, the addition of phosphate to glucose traps this monosaccharide within the cell⁵⁶. Further alkylation reactions can occur including methylation through a range of dedicated methyltransferases⁵³. This not only introduces hydrophobicity, but also neutralizes charges on amines and carboxylic acids⁵⁷. Finally, the oxidation of neutral sugars can form carboxylic acid groups. For instance, the oxidation of C6 yields uronic acids, the oxidation of the aldehyde at C1 produces aldonic acids, and the oxidation of both C6 and C1 results in aldaric acids^{11,58,59}. A noteworthy case are nonulosonic acids, a class of nine-carbon alpha-keto acids, where the C2 ketone group is oxidized to a carboxylic acid^{11,60}. These sugars are produced in cells via a condensation reaction between specific hexosamine sugars and phosphoenolpyruvate, forming a nine-carbon backbone and a terminal alpha-keto acid moiety^{11,60,61}. Some microbes contain dedicated lyases known as sialic acid aldolases, which, for example, catalyze the reversible aldol cleavage of the nonulosonic acid neuraminic acid (Neu), from pyruvate and ManNAc⁶². The broader group of known nonulosonic acid derivatives contains some 100 members. Neuraminic acid (Neu) and its unmodified variant, keto-deoxy-neuraminic acid (Kdn), are commonly found in animal tissues. Pseudaminic acid (Pse) and its stereoisomers, legionaminic acid (Leg), acinetaminic acid (Aci), and fusaminic acid (Fus), have been exclusively identified in prokaryotes^{60,63}. Prokaryotic nonulosonic

acids are often modified with, for example, formyl, glyceryl, hydroxybutyryl, lactoyl, and glutamyl groups, among many others^{64,65}. Neuraminic acids play a role in processes such as cell–cell interactions, signaling, adhesion, and immune responses. Similarly, Pse/Leg derivatives are found on the bacterial cell surface, including lipopolysaccharides, adhesins, pili, and flagella^{11,60}.

Highly modified and rare monosaccharides are particularly frequent in microbes, where they often play crucial roles in determining their biological activity⁶⁶. However, explanations for the development of modifications and rare monosaccharides are only speculative. For example, while the derivative Bac2Ac4Ac is present in the glycoproteins of *Campylobacter jejuni*, it can also be replaced by other N-acetyl sugars, such as Glc2NAc or Fuc2NAc³³, demonstrating that similar properties can arise from different monosaccharide building blocks. This enhances adaptability and may be particularly crucial for the survival of microbes in competitive environments with fluctuating nutrient availability. The spectrum of monosaccharide building blocks is not dependent on the proteome size and differs greatly between species. Although a common group of monosaccharides is found across all three domains of life, most organisms do not use the exact same set of monosaccharides³³. The development of more specific monosaccharides likely emerged from the neo- or sub-functionalization of enzymes at later stages of evolution³³.

Richard D. Cummings recently introduced the “periodic table of monosaccharides” (Figure 2.2), analogous to the periodic table of elements, which illustrates the diversity and similarities of monosaccharides via their systematic arrangement⁵. Currently, the table contains the 103 monosaccharides listed in the KEGG⁹ database and on the Symbol Nomenclature for Glycans (SNFG) website⁶⁷. The sugars are ordered by the number of carbon atoms in their backbone, and are classified into five major groups: aldoses, ketoses, aminosugars, uronic acids, and deoxysugars. These groups are further organized into periods, starting with 3-carbon trioses and extending to 10-carbon decoses.

The table also lists rare sugars, namely, branched-chain sugars⁶⁸, apart from common monosaccharides. These sugars follow the general empirical formula of monosaccharides, with the exception that the backbone hydroxyl groups are replaced by another carbon bond, thereby forming a branch. Many of these sugars are biologically active and exhibit antimicrobial and antiviral activities^{68,69}. Plants can also contain branched sugars, such as hamamelose (2-C-hydroxymethyl ribose). This sugar is involved in the regulation of ribulose biphosphate carboxylase/oxygenase, an enzyme responsible for CO₂ assimilation in higher plants, and many photosynthetic bacteria^{69,70}. Another branched chain sugar widely found in plants is apiose, which bridges pectin molecules to regulate the stiffness of plant cell walls^{57,71}.

A Periodic Table of Monosaccharides

	Aldoses				Ketoses		Aminosugars				Uronic Acids				Deoxysugars									
Triose	Dihydroxyacetone				DHA																			
	Gly	C ₃ H ₆ O ₃	150.07																					
Tetrose	Threose				Erythrose																			
	Thr	C ₄ H ₈ O ₄	162.07																					
Pentose	Ribose				Xylose																			
	Rib	C ₅ H ₁₀ O ₅	180.15																					
Hexose	Glucose				Fructose																			
	Glc	C ₆ H ₁₂ O ₆	180.15																					
Heptose	Sedoheptulose				Mannose																			
	Sed	C ₇ H ₁₄ O ₇	202.07																					

	Aldoses				Ketoses		Aminosugars				Uronic Acids				Deoxysugars									
Triose	Dihydroxyacetone				DHA																			
	Gly	C ₃ H ₆ O ₃	150.07																					
Tetrose	Threose				Erythrose																			
	Thr	C ₄ H ₈ O ₄	162.07																					
Pentose	Ribose				Xylose																			
	Rib	C ₅ H ₁₀ O ₅	180.15																					
Hexose	Glucose				Fructose																			
	Glc	C ₆ H ₁₂ O ₆	180.15																					
Heptose	Sedoheptulose				Mannose																			
	Sed	C ₇ H ₁₄ O ₇	202.07																					

	Aldoses				Ketoses		Aminosugars				Uronic Acids				Deoxysugars									
Triose	Dihydroxyacetone				DHA																			
	Gly	C ₃ H ₆ O ₃	150.07																					
Tetrose	Threose				Erythrose																			
	Thr	C ₄ H ₈ O ₄	162.07																					
Pentose	Ribose				Xylose																			
	Rib	C ₅ H ₁₀ O ₅	180.15																					
Hexose	Glucose				Fructose																			
	Glc	C ₆ H ₁₂ O ₆	180.15																					
Heptose	Sedoheptulose				Mannose																			
	Sed	C ₇ H ₁₄ O ₇	202.07																					

	Aldoses				Ketoses		Aminosugars				Uronic Acids				Deoxysugars									
Triose	Dihydroxyacetone				DHA																			
	Gly	C ₃ H ₆ O ₃	150.07																					
Tetrose	Threose				Erythrose																			
	Thr	C ₄ H ₈ O ₄	162.07																					
Pentose	Ribose				Xylose																			
	Rib	C ₅ H ₁₀ O ₅	180.15																					
Hexose	Glucose				Fructose																			
	Glc	C ₆ H ₁₂ O ₆	180.15																					
Heptose	Sedoheptulose				Mannose																			
	Sed	C ₇ H ₁₄ O ₇	202.07																					

	Aldoses				Ketoses		Aminosugars				Uronic Acids				Deoxysugars									
Triose	Dihydroxyacetone				DHA																			
	Gly	C ₃ H ₆ O ₃	150.07																					
Tetrose	Threose				Erythrose																			
	Thr	C ₄ H ₈ O ₄	162.07																					
Pentose	Ribose				Xylose																			
	Rib	C ₅ H ₁₀ O ₅	180.15																					
Hexose	Glucose				Fructose																			
	Glc	C ₆ H ₁₂ O ₆	180.15																					
Heptose	Sedoheptulose				Mannose																			
	Sed	C ₇ H ₁₄ O ₇	202.07																					

	Aldoses				Ketoses		Aminosugars				Uronic Acids				Deoxysugars									
Triose	Dihydroxyacetone				DHA																			
	Gly	C ₃ H ₆ O ₃	150.07																					
Tetrose	Threose				Erythrose																			
	Thr	C ₄ H ₈ O ₄	162.07																					
Pentose	Ribose				Xylose																			
	Rib	C ₅ H ₁₀ O ₅	180.15																					
Hexose	Glucose				Fructose																			
	Glc	C ₆ H ₁₂ O ₆	180.15																					
Heptose	Sedoheptulose				Mannose																			
	Sed	C ₇ H ₁₄ O ₇	202.07																					

	Aldoses				Ketoses		Aminosugars				Uronic Acids				Deoxysugars									
Triose	Dihydroxyacetone				DHA																			
	Gly	C ₃ H ₆ O ₃	150.07																					
Tetrose	Threose				Erythrose																			
	Thr	C ₄ H ₈ O ₄	162.07																					
Pentose	Ribose				Xylose																			
	Rib	C ₅ H ₁₀ O ₅	180.15																					
Hexose	Glucose				Fructose																			
	Glc	C ₆ H ₁₂ O ₆	180.15																					
Heptose	Sedoheptulose				Mannose																			
	Sed	C ₇ H ₁₄ O ₇	202.07																					

	Aldoses				Ketoses		Aminosugars				Uronic Acids				Deoxysugars									
Triose	Dihydroxyacetone				DHA																			
	Gly	C ₃ H ₆ O ₃	150.07																					
Tetrose	Threose				Erythrose																			
	Thr	C ₄ H ₈ O ₄	162.07																					
Pentose	Ribose				Xylose																			
	Rib	C ₅ H ₁₀ O ₅	180.15																					
Hexose	Glucose				Fructose																			
	Glc	C ₆ H ₁₂ O ₆	180.15																					
Heptose	Sedoheptulose				Mannose																			
	Sed	C ₇ H ₁₄ O ₇	202.07																					

	Aldoses				Ketoses		Aminosugars				Uronic Acids				Deoxysugars									
Triose	Dihydroxyacetone				DHA																			
	Gly	C ₃ H ₆ O ₃	150.07																					
Tetrose	Threose				Erythrose																			
	Thr	C ₄ H ₈ O ₄	162.07																					
Pentose	Ribose				Xylose																			
	Rib	C ₅ H ₁₀ O ₅	180.15																					
Hexose	Glucose				Fructose																			
	Glc	C ₆ H ₁₂ O ₆	180.15																					
Heptose	Sedoheptulose				Mannose																			
	Sed	C ₇ H ₁₄ O ₇	202.07																					

	Aldoses				Ketoses		Aminosugars				Uronic Acids				Deoxysugars									
Triose	Dihydroxyacetone				DHA																			
	Gly	C ₃ H ₆ O ₃	150.07																					
Tetrose	Threose				Erythrose																			
	Thr	C ₄ H ₈ O ₄	162.07																					
Pentose	Ribose				Xylose																			
	Rib	C ₅ H ₁₀ O ₅	180.15																					
Hexose	Glucose				Fructose																			
	Glc	C ₆ H ₁₂ O ₆	180.15																					
Heptose	Sedoheptulose				Mannose																			
	Sed	C ₇ H ₁₄ O ₇	202.07																					

	Aldoses				Ketoses		Aminosugars				Uronic Acids				Deoxysugars									
Triose	Dihydroxyacetone				DHA																			
	Gly	C ₃ H ₆ O ₃	150.07																					
Tetrose	Threose				Erythrose																			
	Thr	C ₄ H ₈ O ₄	162.07																					
Pentose	Ribose				Xylose																			
	Rib	C ₅ H ₁₀ O ₅	180.15																					
Hexose	Glucose				Fructose																			
	Glc	C ₆ H ₁₂ O ₆	180.15																					
Heptose	Sedoheptulose				Mannose																			
	Sed	C ₇ H ₁₄ O ₇	202.07																					

	Aldoses				Ketoses		Aminosugars				Uronic Acids				Deoxysugars									
Triose	Dihydroxyacetone				DHA																			
	Gly	C ₃ H ₆ O ₃	150.07																					
Tetrose	Threose				Erythrose																			
	Thr	C ₄ H ₈ O ₄	162.07																					
Pentose	Ribose				Xylose																			
	Rib	C ₅ H ₁₀ O ₅	180.15																					
Hexose	Glucose				Fructose																			
	Glc	C ₆ H ₁₂ O ₆	180.15																					
Heptose	Sedoheptulose				Mannose																			
	Sed	C ₇ H ₁₄ O ₇	202.07																					

	Aldoses				Ketoses		Aminosugars				Uronic Acids				Deoxysugars									
Triose	Dihydroxyacetone				DHA																			
	Gly	C ₃ H ₆ O ₃	150.07																					
Tetrose	Threose				Erythrose																			
	Thr	C ₄ H ₈ O ₄	162.07																					
Pentose	Ribose				Xylose																			
	Rib	C ₅ H ₁₀ O ₅	180.15																					
Hexose	Glucose				Fructose																			
	Glc	C ₆ H ₁₂ O ₆	180.15																					
Heptose	Sedoheptulose				Mannose																			
	Sed	C ₇ H ₁₄ O ₇	202.07																					

	Aldoses				Ketoses		Aminosugars				Uronic Acids				Deoxysugars									
Triose	Dihydroxyacetone				DHA																			
	Gly	C ₃ H ₆ O ₃	150.07																					
Tetrose	Threose				Erythrose																			
	Thr	C ₄ H ₈ O ₄	162.07																					
Pentose	Ribose				Xylose																			
	Rib	C ₅ H ₁₀ O ₅	180.15																					
Hexose	Glucose				Fructose																			
	Glc	C ₆ H ₁₂ O ₆	180.15																					
Heptose	Sedoheptulose				Mannose																			
	Sed	C ₇ H ₁₄ O ₇	202.07																					

	Aldoses				Ketoses		Aminosugars				Uronic Acids				Deoxysugars									
Triose	Dihydroxyacetone				DHA																			
	Gly	C ₃ H ₆ O ₃	150.07																					
Tetrose	Threose				Erythrose																			
	Thr	C ₄ H ₈ O ₄	162.07																					

UDP-N-acetylglucosamine and an altrose intermediate (6-deoxy-Altdi-NAC), while the legionaminic analogue is produced from a bacillosamine (a glucosamine derivative) intermediate^{63,77,78}. Interestingly, the known Pse biosynthesis pathway uses UDP precursors, while the known Leg biosynthesis pathway employs GDP-linked precursors⁷⁹. Overall, the nucleotide sugar metabolism involves a branched network of pathways with many interconversions and has only been well-established for a limited number of mammalian model cells, yeast, and a few bacterial pathogens such as *E. coli* and *M. tuberculosis*.

2.3 | Liquid chromatography coupled with mass spectrometry

One of the most widely employed separation techniques combined with mass spectrometry is liquid chromatography, mainly due to its suitability for polar and thermally labile molecules found in biological material that are challenging to analyze using historical gas chromatographic methods. Although the direct coupling of liquid chromatography with mass spectrometry was long considered difficult, the development of the electrospray ionization interface by Fenn et al. provided a solution that facilitated coupling and offered soft ionization of highly polar molecules—such as cellular metabolites, peptides, nucleotides, oligosaccharides, and even intact proteins^{80,81}. Various chromatographic separation phases have been employed for the separation of monosaccharide derivatives before mass spectrometric detection, each having their own advantages and disadvantages. In the following we will discuss and review the most recent separation techniques which have been coupled to mass spectrometry to explore the fundamental sugar building blocks, including reverse phase chromatography, ion exchange chromatography, hydrophilic interaction chromatography, porous graphite carbon chromatography, mixed mode chromatography and supercritical fluid chromatography.

Reverse-phase chromatography. Reverse phase (RP) chromatography employs a non-polar stationary phase, typically composed of long alkyl chains covalently attached to a solid support. The polar mobile phase consists of an aqueous-organic solvent, and organic compounds are separated based on hydrophobicity^{82,83}. RP chromatography is appealing due to its ease of use and widespread application in analytical research laboratories. Moreover, a significant advantage is its robustness, and the possibility to employ volatile buffers that are compatible with the electrospray interface^{84,85}. Nevertheless, the polar and hydrophilic nature of monosaccharides and their derivatives results in poor retention on conventional reverse-phase resins. Additionally, the ionization efficiency of sugars can be low^{85,86}. Therefore, adduct formation between uncharged sugars and ions, such as Cs⁺, Na⁺, and NH₄⁺ for positive ionization or Cl⁻ and CH₃CO⁻ for negative ionization, was introduced⁸⁷⁻⁹¹. The ions can either be introduced post-column through a make-up flow or a triaxial electrospray probe^{89,91}, or derived directly from the solvent buffer⁸⁸. However, in these methods mostly underivatized carbohydrates were analyzed, therefore non-reverse phase separation phases were typically used. While adduct

formation improved detection, it did not enhance the separation of sugar stereoisomers.

To address these challenges, various derivatization methods have been developed to decrease the polarity and increase ionization efficiency of sugars. Historically, different pre- and post-column derivatization methods were developed to increase retention or to introduce UV-absorbing groups to monosaccharides⁹²⁻⁹⁵. However, many of these early methods, such as benzylation of hydroxyl groups, provided anomeric mixtures from reducing carbohydrates and provided poor derivatization yields⁹³. Dansylhydrazine was introduced which reacts specifically with aldehydes and ketones under mild conditions, introducing a fluorescent tag, which enabled reverse phase separation and fluorescence detection⁹⁶. Later, Han et al. employed another hydrazine label, 3-nitrophenylhydrazine, for small sugars. These were analyzed on a pentafluorophenyl-bonded phase column coupled with electrospray ionization mass spectrometry (ESI-MS) detection⁹⁷. Using a ¹³C-labeled internal standard, this method demonstrated excellent separation and quantitative properties. Alternatively, reductive amination of sugars was introduced for oligosaccharides by Hase et al. which provided a single derivative from reducing carbohydrates. Pyridylation not only improved chromatographic separation but also facilitated sensitive analysis through fluorescence detection⁹⁸. However, reductive amination occurs under acidic conditions, which can lead to the loss of acid-labile groups such as O-sulfate groups and sialic acid residues on oligosaccharides⁹⁹. Honda et al. developed derivatization through condensation with an active methylene group, exemplified with the reagent 1-phenyl-3-methyl-5-pyrazolone (PMP) for monosaccharides. Under basic conditions two PMP molecules were added to the reducing end of each monosaccharide¹⁰⁰. The hydrophobic PMP molecules enhanced monosaccharide separation on reverse-phase chromatography and enabled detection by UV. Since then, PMP has been widely used for monosaccharide analysis especially because of its quantitative conversion under mild conditions. However, the specificity of UV detection is generally low, therefore, separation conditions have been optimized to allow coupling to electrospray ionization mass spectrometry^{85,86,101,102}. Combined with product ion scanning, this provides high selectivity and accurate quantification of various monosaccharides from complex biomass samples⁸⁵. Additionally, micro-scale hydrolysis and PMP derivatization were performed in a 96-well plate without extraction or drying steps, followed by ESI-MS analysis. This method demonstrated increased throughput and efficiency¹⁰³.

Xu et al. recently reported an improved method that facilitated the simultaneous detection of PMP-labelled sugars, including underivatized sialic acids, by RP-MS in multiple reaction monitoring (MRM) mode. The authors optimized PMP concentrations to ensure complete derivatization while avoiding ion suppression caused by excess reagent. Separation was achieved in less than 10 minutes while detection limits reached attomole levels for some sugars, with linear ranges spanning up to six orders of magnitude¹⁰⁴.

Amicucci et al. developed a method for the absolute quantitation of 14 monosaccharides. This method involved a 96-well plate hydrolysis and PMP derivatization procedure, followed by a rapid and sensitive 10-minute reverse-phase separation and triple quadrupole mass spectrometric analysis¹⁰⁵. Recently, Salas and colleagues developed an improved PMP derivatization approach followed by RP-ESI-Orbitrap MS analysis. This method enabled the identification and quantification of 18 monosaccharide derivatives, including neutral sugars, aminosugars, N-acetyl amino sugars, and uronic acids. Notably, the analysis of acid-hydrolyzed samples from archaea, bacteria, fungi, and plants distinguished between these taxonomic groups based on their sugar markers¹⁰⁶ (Figure 2.3).

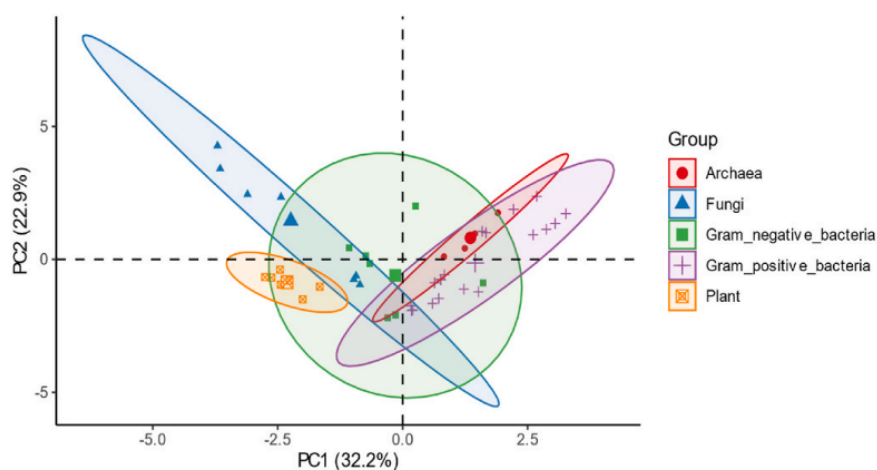


Figure 2.3. A differentiation between archaea, bacteria, fungi, and plants based on their sugar markers¹⁰⁶. The monosaccharides were measured following PMP labeling of the acid hydrolyzed samples. Reprinted with permission from Salas et al., A rapid and sensitive assay to quantify amino sugars, neutral sugars and uronic acid necromass biomarkers using pre-column derivatization, ultra-high-performance liquid chromatography and high-resolution mass spectrometry. *Soil Biology and Biochemistry*, 2023, 177: 108927.

The combination with Orbitrap mass spectrometry demonstrated the potential to characterize rare and species-specific compounds commonly found in microbes. Zhang et al. introduced the alternative label 1-(4-isopropyl) phenyl-3-methyl-5-pyrazolone (PPMP), which produced mono-PPMP labelled monosaccharides under similarly mild conditions⁹⁹. The mono-PPMP derivatives demonstrated better chemical stability than the bis-PMP derivatives. Moreover, excess label was more easily removed. The label was applied to determine the polysaccharide composition of *Spirulina*, dried biomass of the bacteria *Arthospira pletensis*⁹⁹. Anumula introduced reductive amination using the fluorescent label anthranilic acid (2-aminobenzoic acid) in the presence of sodium cyanoborohydride¹⁰⁷. Excess reagent removal was unnecessary because monosaccharide derivatives were separated from the reagent peak on a reverse-phase column. However, performing reductive amination in a methanol-acetate-borate buffer at elevated temperatures

resulted in some epimerization of glucosamine to mannosamine. The separation of anthranilic acid and PMP-derivatized sugars has been optimized for small sample volumes and the separation of pentose and hexose sugars including methylated hexose variants¹⁰⁸. A spectrum of different tags has been developed by then^{109,110}. For comprehensive comparative studies and reviews on different tags, we refer to earlier studies^{109,111-115}. Many applications have relied solely on fluorescence or UV detection and direct comparison of retention times to standards^{108,109,116-118}. However, complex samples matrices, unexpected sugars, derivatives or other conversion products take advantage from additional mass spectrometric detection^{119,120}. Zhu et al. utilized acid hydrolysis and PMP labelling, followed by reverse-phase high-performance liquid chromatography and off-line mass spectrometry, to analyze glycosaminoglycans¹²¹. The mass spectrometric detection enabled the unambiguous identification of acidolysis-resistant components, such as disaccharides, which can be easily misinterpreted when relying solely on fluorescence detection. Additionally, fragmentation provided tags to distinguish between heparin/heparan sulfate and chondroitin sulfate/dermatan sulfate¹²¹. Volatile buffers suitable for direct coupling to electrospray ionization interfaces often do not provide sufficient separation performance to distinguish all pentose and hexose diastereomers¹²⁰. Consequently, mass spectrometric coupling often comes at the cost of suboptimal separation. Windwarder and colleagues, however, developed a method based on permethylation of anthranilic acid-labelled monosaccharides, enabling the chromatographic separation of several pentose and hexose diastereomers¹²⁰. Additionally, their mass spectrometric approach allowed the determination of the degree of methylation of monosaccharides by using deuterated iodomethane in the permethylation step (Figure 2.4). Rath et al. investigated the *Selenomonas sputigena* flagellin O-glycan composition using the anthranilic acid-labelling/permethylation approach, which allowed to assign a methylated deoxysugar as methylated rhamnose¹²². Galermo and colleagues expanded the compositional analysis of oligosaccharides to include glycosidic linkage analysis¹²³. In their method, oligosaccharides were first permethylated, hydrolyzed, and subsequently derivatized with PMP. The monosaccharide derivatives were then separated by reverse-phase chromatography and detected using multiple reaction monitoring (MRM) (Figure 2.5). This approach enabled the rapid determination of terminal, linear, bisecting, and trisecting monosaccharide linkages¹²³. Recently, Wang et al. developed a "paired derivatization approach" employing the label O-(4-methoxybenzyl)hydroxylamine hydrochloride (4-MOBHA·HCl). Samples were labelled with the native form (4-MOBHA·HCl), while reference standards were labelled with a deuterated form (d3-4-MOBHA·HCl)¹²⁴. This mass spectrometric targeting of both variants improved quantification accuracy and sensitivity in complex matrices such as plant and feces samples.

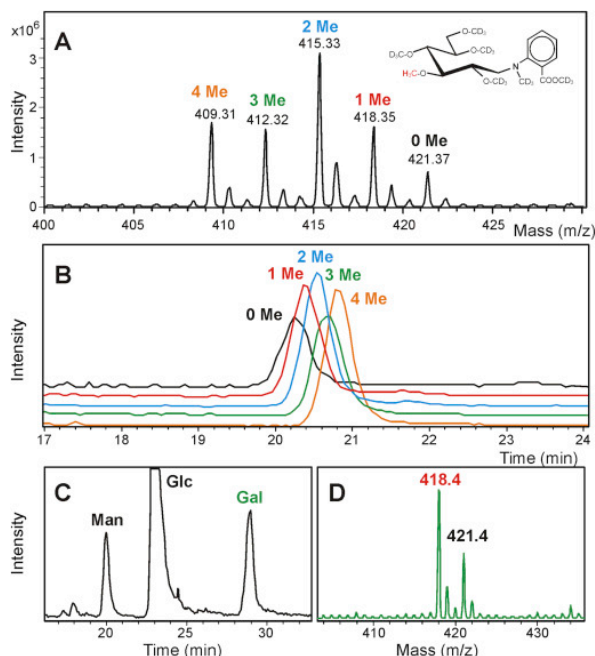


Figure 2.4. Isomeric monosaccharides separation using reverse-phase chromatography followed by mass spectrometric detection of "hyper"-methylated anthranilic acid-labeled sugars¹²⁰. **A)** The mass spectrum of a mannose sample with varying degrees of natural O-methylation, labeled with anthranilic acid (AA) and "hypermethylated" with deuteriomethyl iodide. **B)** Extracted ion chromatograms of the mannose species, illustrating how deuterium incorporation affects retention time. **C)** Extracted ion chromatograms for different hexoses from a glycoprotein preparation obtained from *Achatina fulica*. **D)** Mass spectrum of the galactose peak from panel C, showing varying degrees of methylation in the biological sample. Reprinted from Windwarder, Markus, et al. "Hypermethylation" of anthranilic acid-labeled sugars confers the selectivity required for liquid chromatography-mass spectrometry." *Analytical Biochemistry* 514 (2016): 24-31., Copyright (2016), with permission from Elsevier.

An alternative derivatization method, which also encompassed ketose sugars (e.g., fructose), was achieved through direct acetylation. This increased the hydrophobicity of sugars, enabling analysis by RP-ESI-MS. The method was applied to determine the monosaccharide composition of polysaccharides from various plants and fungi. Characteristic ion transitions were observed for aldopentoses, uronic acids, ketohexoses, aldohexoses, methyl-aldopentoses, and alditols¹²⁵. However, the aforementioned labelling approaches were primarily applied to pentose and hexose derivatives. In contrast, the larger nonulosonic acids (NulOs) were typically analyzed using more specific labelling techniques. Following early colorimetric assays, such as those using thiobarbituric acid¹²⁶ and resorcinol¹²⁷, more sensitive and reliable methods were developed based on fluorescence tags that specifically recognize the alpha-keto acid¹²⁸. In 1987, 1,2-diamino-4,5-methylenedioxybenzene (DMB) was first mentioned as reagent which specifically reacts with the alpha-keto acid of NulOs¹²⁹. Since then, DMB labelling followed by reverse-phase chromatography coupled with fluorescence detection or mass

spectrometry has been widely employed, which has been extensively documented in the literature and several book chapters^{128,130-135}. Furthermore, the mass spectrometric detection allowed to identify NuLOs in complex matrices. For example, Klein and colleagues were the first to demonstrate the separation and mass spectrometric analysis of DMB-labelled NuLOs. The authors identified 28 different sialic acids in the glycolipids of *Lovenia cordiformis*¹³⁴. Furthermore, Leg5Ac7AcAla was identified in *Vibrio vulnificus*, where the NuLO was identified to contribute to membrane integrity¹³⁶. NuLO-deficient mutants exhibited also decreased virulence in the crustacean model *Artemia francisca*, emphasizing the importance of these sugars. By performing an additional reduction step prior to hydrolysis of CMP-activated NuLOs, both free and CMP activated NuLOs could be quantified¹³⁷. However, this method required the removal of any NuLO containing glycoconjugates prior to hydrolysis, to prevent interference of the quantification. Du et al. reported the use of 4,5-dimethylbenzene-1,2-diamine (DMBA) for analyzing NuLOs via reverse-phase chromatography coupled with mass spectrometry¹³⁵. DMBA demonstrated superior chromatographic separation efficiency and comparable mass spectrometric performance to DMB.

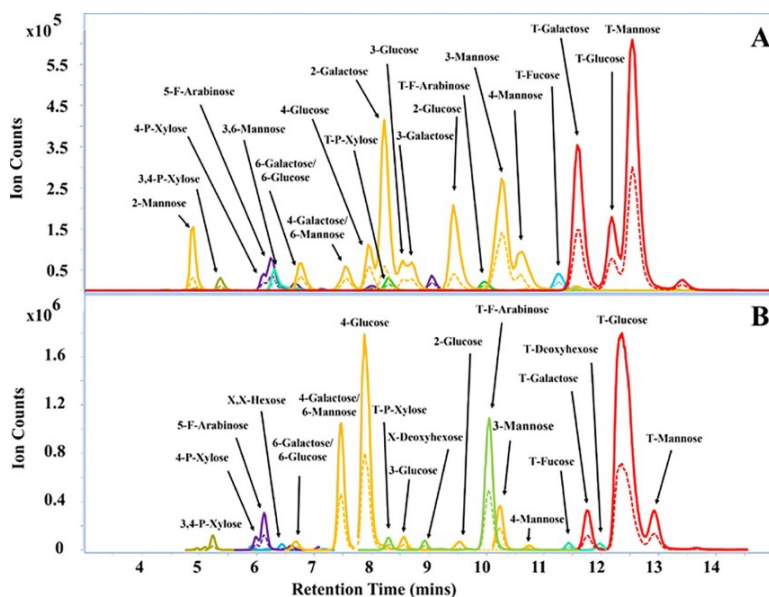


Figure 2.5. Rapid determination of glycosidic linkages in complex biological samples¹²³. The method involves permethylation, hydrolysis, and PMP derivatization of oligosaccharides, followed by reverse-phase chromatography and multiple reaction monitoring. **A)** The MRM chromatogram of commercially available monosaccharide standards. **B)** The MRM monosaccharide profile obtained from a plant biomass sample. Solid lines represent the quantifying transitions, while dashed lines indicate the qualifying transitions for each detected monosaccharide derivative. Reprinted with permission from Galermo, Ace G., et al. "Liquid chromatography–tandem mass spectrometry approach for determining glycosidic linkages." *Analytical chemistry* 90.21 (2018): 13073-13080. Copyright 2018 American Chemical Society

Finally, Kleikamp et al. conducted a fully untargeted mass spectrometric screening for nonulosonic acids (NulOs) in various prokaryotes. DMB labelling proved useful in providing the sensitivity and selectivity necessary for detecting novel NulO derivatives. This labelling introduced conserved ulosonic acid core fragment features distinct from the cell lysate background and altered the double bond equivalents and mass defect, enhancing sensitivity. Additionally, their automated filtering and scoring approach enabled the efficient screening of a large number of prokaryotic samples¹³⁸. Applying this method, NulOs were detected in several environmental microbes, such as *Candidatus Accumulibacter phosphatis* enrichments, although its function remains to be elucidated^{138,139}. However, DMB requires relatively high temperatures and acidic reaction conditions, which can lead to byproducts and degradation of sialic acids and unstable modifications. One such artifact is the spontaneous migration of acetyl groups between hydroxyl groups. Notably, the 8-O-acetyl ester is more labile than the 9-O-acetyl ester, with acetyl groups tending to migrate to the C-9 position even under neutral conditions at room temperature¹⁴⁰. To address this, Zhao et al. developed a method involving the sequential derivatization of the carbonyl and carboxyl groups using 3-NPH. This approach proved compatible with labile O-acetyl modifications, and the derivatives demonstrated good stability and chromatographic separation¹⁴¹.

Ion-pair reverse-phase chromatography. The retention of highly polar monosaccharides, particularly for sugar phosphates and the nucleotide activated monosaccharides can be enhanced by the addition of an ion pair (IP) reagent to the mobile phase. Advantageously, compared to ion exchange chromatography, IP-RP allows for the separation of ionic analytes in parallel to neutral analytes^{142,143}. Initially, it was believed that the IP reagents form neutral ion-pairs with oppositely charged analytes, enabling their partition into the non-polar stationary phase due to their hydrophobic tails. This concept led to the term ion pair chromatography. However, the retention mechanism appeared to be more complex and still remains under debate^{143,144}. In short, modern theories suggest that the IP reagent adsorbs to the surface of the stationary phase via hydrophobic interactions, thereby creating a charged ion layer. The retention of analytes is influenced via at least two factors, namely charge and hydrophobicity. Consequently, IP-RP chromatography allows for separation of isomeric nucleotide-sugars such as UDP-Glc/UDP-Gal and UDP-GlcNAc/UDP-GalNAc¹⁴⁵.

Commonly used IP reagents for phosphate and nucleotide activated sugars are tetra alkyl ammonium salts, like tetrabutylammonium bisulphate^{142,145}. Separation can be enhanced by connecting multiple columns, although at the cost of the analysis time¹⁴². Switching from HPLC to UHPLC mode can further improve separation¹⁴⁶. Significant limitations of IP methods are, however, the high salt concentrations and the often non-volatile IP reagents, which challenge coupling to electrospray ionization mass spectrometry. Nevertheless, different approaches were developed to overcome this limitation, such as post-column removal of non-volatile IP reagents or by replacing them with volatile IP reagents, such as

tributylamine, triethylamine, hexylamine and triethylammonium acetate¹⁴⁶⁻¹⁵⁰. Various experimental parameters were shown to influence the retention and selectivity, such as the type and concentration of the IP reagent, pH of the mobile phase, type and concentration of the organic modifier and the temperature¹⁴³. Although structural isomers often have a similar fragmentation pattern, the fragment ion intensity ratios can differ significantly. In combination with IP chromatographic separation, different hexose monophosphates could be differentiated and quantified based on their fragmentation pattern¹⁴⁶. Van Scherpenzeel et al. optimized the application of IP-RP coupled to mass spectrometry for sugar-phosphates and nucleotide-sugars found in the central metabolic pathways. The authors applied this approach to a range of model organisms including HEK cells and microbes. The authors also traced sugar analogues on sugar metabolism within a cellular context¹⁴⁷. The authors could detect UDP-arabinose as additional low abundant nucleotide sugar, albeit its involvement in glycoconjugate biosynthesis remains unsolved to date. Furthermore, the authors explored the effect of the antitumor compound 3F_{ax}-NeuNAc on sialic acid metabolism. Nevertheless, it was emphasized that IP-RP-ESI-MS is not ideal for multi-purpose instruments due to persistent contamination by IP reagents, necessitating extensive cleaning when switching applications or modes¹⁴⁶⁻¹⁴⁸.

Ion exchange chromatography. Ion exchange chromatography (IEC) relies on the interaction of charged analytes with an oppositely charged stationary phase, a development which dates back to the 1940s¹⁵¹. Originally, the only types of monosaccharides that could be separated by anion-exchange chromatography were acidic monosaccharides, which are negatively charged at physiological pH. Later, high performance (also referred to as 'high-pH') anion exchange (HPAE) chromatography was developed which allows to analyze neutral monosaccharides^{152,153}. Thereby, the hydroxyl groups of monosaccharides are deprotonated in a highly basic solvent (e.g. 0.1M NaOH). The monosaccharide anions then interact with a positively charged stationary phase, which are often quaternary ammonium ions. Differences in the number of hydroxyl groups and pK_a values ultimately facilitate the separation of different monosaccharides^{151,154,155}. This method became employed for the label free analysis of oligo- and monosaccharides from different matrices^{154,156-161}. Elution can be achieved by either a pH gradient or increasing salt concentration¹⁵¹. However, the high concentrations of non-volatile salts make this method incompatible for coupling to mass spectrometry. Therefore, high performance anion exchange chromatography is usually combined with pulsed amperometric detection (PAD), which, however, does not allow the application to complex samples. For example, organic acids in the matrix background can interfere with the accurate determination of sugars^{152,154}. Mass spectrometric analysis is often only achieved off-line, after fractionation¹⁶²⁻¹⁶⁴. HPAE chromatography often also shows poor long-term stability and requires lengthy separation times^{145,165}. Furthermore, hydroxyl modifications,

such as acetylation, often found on NulOs, are base-labile. Therefore, these sugars have not been analyzed using standard HPAE techniques. However, substituted NulOs were successfully separated by anion exchange using solvents at neutral pH¹⁶⁶.

Coupling HPAEC with electrospray ionization mass spectrometry was finally achieved by implementing post-column ion suppressors that, for example, replace Na⁺ ions with H⁺ ions. Additionally, make-up solvents were introduced that include dopants such as LiCl to improve ionization of the monosaccharides¹⁶⁷⁻¹⁷¹. Zhao et al. incorporated an online sample pretreatment using a column-switching technique. In short, the injected sample passed through a reverse-phase pretreatment column, where monosaccharides eluted immediately due to their polarity, eliminating the matrix. The collected sugars were then directed to an ion exchange analytical column for separation. A suppressor column functioned as a desalter, making it MS-compatible, and a T-junction introduced an additional make-up solution containing LiOH, which improved ionization efficiency. The method was then applied to analyze monosaccharides in activated sludge from wastewater treatment plants, crucial for understanding the water treatment mechanism¹⁷⁰. Optimized methods also demonstrated the simultaneous separation of neutral, amino and acidic monosaccharides^{154,168}. For example, Zhang et al., achieved the simultaneous detection of Fuc, Rha, GalN, Ara, GlcN, Gal, Glc, Man, Xyl, Neu5Ac, Neu5GC, GalA, GulA, GlcA, ManA, and IdoA¹⁵⁴.

Some studies also showed the application of HPAE chromatography to analyze nucleotides and nucleotide activated sugars, although this approach has been less widely employed^{172,173}. For example, Alonso and co-workers demonstrated the detection of 16 hexose phosphates and nucleotide activated sugars involved in cell wall biosynthesis, including the separation of isomers such as UDP-GalNAc/UDP-GlcNAc, Glc6P/Fru6P/Man6P, UDP-GalA/UDP-GlcA, UDP-Gal/UDP-Glc and UDP-Ara/UDP-Xyl. The application of ¹³C labelled fructose in growth medium allowed to track the incorporation pattern in cell wall precursors¹⁷³.

Porous graphitic carbon chromatography. The production of porous graphitic carbon (PGC) as chromatographic separation material was developed by Knox and Gilbert and became commercially available in 1988 under the trade name Hypercarb^{TM174}. PGC acts as a strongly retentive alkyl-bonded silica gel for non-polar analytes, yet its retention and selectivity behavior towards polar and structurally related compounds is markedly different¹⁷⁴. The PGC particles are spherical and fully porous, with a porosity of approximately 75% and a crystalline surface. Chemically, they are composed of sheets of hexagonally arranged carbon atoms connected by conjugated 1.5-order bonds, similar to those found in polynuclear aromatic hydrocarbons¹⁷⁴. Consequently, the PGC separation phase can be operated across a wide pH range (1–14) and at high temperatures (250 °C)¹⁷⁴. Multiple studies and reviews have been dedicated to the retention mechanism and applications of PGC¹⁷⁴⁻¹⁸⁶, thus its mechanisms will be discussed only briefly in this review. As mentioned above, although PGC behaves like a strongly retentive alkyl-

bonded silica gel for non-polar analytes, its retention and selectivity towards polar and structurally related compounds are markedly different. In conventional alkyl-bonded silicas, adding a polar group to a molecule usually decreases retention in reverse-phase mode. However, with PGC, this retention reduction is much smaller or may even increase. An important factor contributing to this so-called polar retention effect involves charge-induced interactions between the analyte and the graphite surface. In these interactions, lone-pair or aromatic ring electrons interact with the graphite through electron transfer to the graphite's electron cloud^{174,176}. Furthermore, due to the flat and highly adsorptive surface of graphite, there is enhanced selectivity for structurally related compounds, such as structural isomers and diastereoisomers. This makes PGC well suited to the separation of very polar and ionized molecules such as carbohydrates and other compounds with several hydroxyl, carboxyl, amino and other polar groups. Despite its excellent chromatographic performance, PGC chromatography can suffer from retention time instability. This instability has been primarily attributed to column contamination and redox reactions with buffer components or analyte molecules. For example, the graphite surface has been affected by redox reagents such as hydrogen peroxide or sulphite^{187,188}. To mitigate these issues, regular backflushing and washing procedures with strong acids, bases, and solvents, such as acetone, tetrahydrofuran, hydrochloric acid, sodium hydroxide, and trifluoroacetate, have been employed^{176,189}. Moreover, additives in the mobile phase can potentially adsorb onto the stationary phase, altering separation and retention capabilities. The mobile phase may also affect the ionization state of both the PGC surface as well as the analytes, impacting the retention^{175,189}. Therefore, Bapiro et al. optimized the application of different mobile phase solvents in a gradient program which preserved the retention capacity of PGC columns across multiple runs¹⁸⁹. The authors demonstrated promising results in analyzing gemcitabine (a cytidine derivative with fluorination at the 2-position of the deoxyribose sugar molecule) and its metabolites, achieving good peak shapes and retention without the need for backflushing. In conjunction with electrospray ionization mass spectrometry, separation using PGC has been shown to be influenced by the electrospray voltage, leading to column polarization^{182,190}. This issue can be resolved by introducing electrical grounding points close to the separation column.

The strength of interaction of an analyte with the PGC separation phase depends on both the surface area of the analyte in contact with the graphite and the nature and type of functional groups¹⁷⁴. As a result, the relatively small monosaccharide components exhibit a comparatively weak interaction with the PGC separation phase. Although to a lesser extent, PGC chromatography coupled to ESI-MS has been employed to measure monosaccharides and phosphorylated derivatives. For example, Hammad and colleagues demonstrated a MRM method for the quantification of free and glycoprotein monosaccharides following separation by PGC chromatography⁸⁸. Neutral monosaccharides were detected as their alditol acetate anion adducts, while sialic acids were detected as deprotonated ions. This

method exhibited high sensitivity, with detection limits as low as 1 pg for glucose, galactose, and mannose, and showed linearity over three orders of magnitude⁸⁸. Antonio et al. developed a method to study metabolic sugars and sugar-phosphate intermediates in *A. thaliana*. Although complete separation of sugar isomers was not achieved, distinct fragmentation profiles allowed for differentiation between isomers¹⁹¹. Interestingly, Li et al. developed an improved workflow to quantify glycolytic pathway intermediates, including sugar phosphates. This workflow employed TiO₂ enrichment of phosphorylated metabolites, phosphate methylation to enhance chromatographic performance and incorporated stable isotopes, followed by porous graphitic carbon chromatography and selected reaction monitoring¹⁹². Additionally, a ¹³C-labeled yeast metabolite extract was used as an internal standard. The method achieved limits of quantification ranging between 0.25 and 0.54 pmol on column¹⁹².

PGC chromatography coupled with mass spectrometry has found greater application in the analysis of the larger and more polar activated monosaccharides. For example, Pabst et al. demonstrated the separation of nucleotides and nucleotide-sugars using PGC coupled to ESI-MS^{193,194}. The PGC separation phase allowed the use of volatile buffer components, such as ammonium acetate, thereby eliminating the need for ion-pairing reagents typically required for separating sugar nucleotides in reverse-phase chromatography. Sensitive detection was achieved in negative ionization mode. Retention time and peak shape instabilities were addressed by applying a column regeneration procedure employing hydrochloric acid, achieving satisfactory performance. This method proved effective for separating a spectrum of isomeric sugar nucleotides from various sources, including animal and microbial model cells, as well as plants¹⁹³. Remarkably, the PGC-ESI-MS approach enabled the separation of UDP-activated Xyl, Araf, and Arap variants, and allowed for the detection of the chemically labile CMP-Kdo. In order to tackle retention time variabilities, relative retention times (to UDP-Glc) were used (Figure 2.6). Furthermore, López-Gutiérrez and colleagues employed PGC-ESI-MS to study the nucleotide-sugar profile in the different life stages of *Plasmodium falciparum*, a malaria parasite¹⁹⁵. Behmüller et al. employed porous graphitic carbon chromatography to quantify UDP-sugars in *A. thaliana* extracts, thereby employing a column regeneration step using 80% acetonitrile in water, containing 0.1% trifluoroacetic acid, to limit retention time instability. Although effective, the 20-minute column regeneration, the 90-minute flush with starting conditions, and the 30-minute analysis time, added to a total runtime of 140 minutes per sample¹⁹⁴. Garcia et al. developed a rapid 6-minute MRM method with PGC separation, applying it to various breast cancer cell lines¹⁹⁶. This method enabled the identification and quantification of all seven targeted sugar nucleotides. Other recent applications of PGC chromatography coupled to ESI-MS include quantification of nucleotide sugars in plasma and urine samples¹⁹⁷, studying the impact of antimicrobials on nucleotide sugar levels in biofilm models of the opportunistic pathogen *Candida albicans*¹⁹⁸, as well as demonstrating the in

planta synthesis of CMP-Neu5Ac in transgenic *Arabidopsis thaliana*¹⁹⁹. Profiling methods for sugar nucleotides, including the application of PGC, has also been reviewed by Rejzek and colleagues recently²⁰⁰.

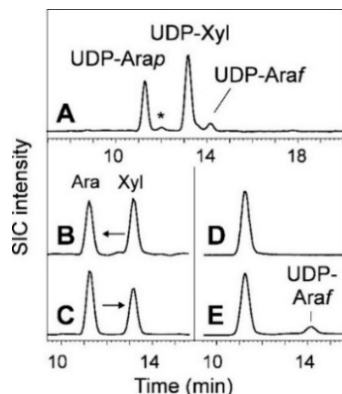


Figure 2.6. Separation of UDP-pentose isomers using PGC-MS⁷⁵. **A)** Extracted ion chromatogram for UDP-pentose sugars (m/z 536.00-526.08) extracted from *Nicotiana benthamiana* leaves, identifying UDP-Arap (pyranose), UDP-Xyl, UDP-Araf, and an unidentified UDP-pentose sugar marked with an asterisk. **B)** Extracted ion chromatogram illustrating the conversion of UDP-Xyl to UDP-Arap using UDP-xylose epimerase from mung bean sprout microsomes. **C)** The same assay as in **B**, but with UDP-Arap as the substrate. **D) and E)** display the conversion of UDP-Arap to UDP-Araf with recombinant UDP-arabinose mutase, where **D)** shows the control without the recombinant enzyme. **SIC:** Selected ion chromatogram. Reprinted with permission from Pabst, et al. Nucleotide and nucleotide sugar analysis by liquid chromatography-electrospray ionization-mass spectrometry on surface-conditioned porous graphitic carbon. Analytical chemistry 82.23 (2010): 9782-9788.

Hydrophilic interaction liquid chromatography. The term hydrophilic interaction chromatography (also referred to as hydrophilic interaction liquid chromatography, HILIC) was first introduced by Alpert in the 1990s²⁰¹. However, HILIC separations for monosaccharide analysis have a much longer history²⁰². Because HILIC separates based on differences in polarity, this technique is particularly useful for the separation of the polar monosaccharides and their derivatives. Typical hydrophilic stationary phases consist of unbound or chemically modified silica (e.g., amino, amide, cyano) or polar polymers. The review by Inoue and Yamamoto summarizes the stationary phases that have been employed for the separation of sugars²⁰³. The mobile phase is usually an organic solvent, such as acetonitrile, with added water, providing the advantage of good compatibility with ESI-MS and low backpressure^{204,205}. The retention of analytes in HILIC occurs mainly via two mechanisms, partitioning and adsorption. Water molecules retained in the hydrophilic stationary phase facilitate the partitioning of polar analytes from the mobile phase to the stationary phase. Additionally, adsorption occurs through interactions, such as ionic interactions, hydrogen bonding, and dipole-dipole forces^{206,207}. As a consequence of this complex retention mechanism, the retention of molecules can vary when different stationary phases are employed²⁰⁵. For a deeper understanding of the retention mechanism of HILIC, we refer to earlier reviews²⁰⁸.

HILIC-MS has been extensively employed in both targeted and untargeted metabolomics studies, as well as an orthogonal approach to applications that employ reverse-phase separation, which have been thoroughly discussed in recent reviews²⁰⁹⁻²²⁰. These also include applications to the analysis of monosaccharides

and derivatives²²¹⁻²²⁹. In the following section, we will cover only recent applications to monosaccharides and their activated sugars.

Su et al. applied HILIC-ESI-MS/MS to quantify phosphate-sugars of the glycolysis and pentose phosphate pathways. Thereby, two HILIC (Atlantis Premier BEH Z-HILIC and Acquity Premier BEH Amide) and a mixed-mode HILIC/strong anion-exchange (SAX) column were compared for its separation performance. Although unable to separate F6P from G1P, the BEH amide column provided the best separation on the tested phosphate sugars²³⁰. Furthermore, Hinterwirth and colleagues introduced a method utilizing a reverse-phase/weak anion-exchanger mixed mode column operated under HILIC conditions²³¹. This approach enabled the separation of sugar phosphates, including hexose and pentose phosphates, glucosamine 1- and 6-phosphate, and 2- and 3-phosphoglycerate. Hammad et al. demonstrated the simultaneous quantification of monosaccharides derived from glycoproteins and blood serum using an aminopropyl column followed by multiple-reaction monitoring²³². Neutral monosaccharides were detected as $[M + \text{CH}_3\text{CO}_2]^-$ ions, while sialic acids were detected as $[M-H]^-$ ions. This method enabled the detection of picogram quantities of the sugars, with a linearity spanning over three orders of magnitude²³². Matějček and Vašíčková utilized HILIC with a zwitterionic ligand to quantify monosaccharide anhydrides, including galactosan, mannosan, and levoglucosan, from environmental samples such as aerosols. The monosaccharide anhydrides were detected as $[M+H\text{COO}]^-$ adduct ions²³³. Furthermore, Yan et al. demonstrated a HILIC method using a BEH amide column, which allowed for the separation of all neutral monosaccharides commonly found in plant-derived oligo- and polysaccharides within 25 minutes²³⁴. Although the authors used charged aerosol detector (CAD) for detection, the method was also suitable for coupling with ESI-MS. The BEH amide column proved advantageous due to its high stability and longer column life.

The separation of isomeric sugars can be a shortcoming of HILIC, which is particularly important for the analysis of samples with unknown derivatives. This issue is critical for the study of rare sugars, which are monosaccharides with very low natural occurrence and abundance^{235,236}. Nevertheless, they may have important functional roles or health effects when consumed²³⁷. For instance, D-psicose and D-allulose have been researched for their effects on glucose tolerance and insulin sensitivity^{238,239}. Therefore, Chung and colleagues developed a HILIC method using a polyethyleneimine separation phase for both reducing and non-reducing monosaccharides, including rare sugars such as D-allose, D-psicose, and D-tagatose²⁴⁰.

Fu and co-workers investigated the isomeric separation of hexoses, pentoses, and disaccharides using the HILIC poly-N-(1H-tetrazole-5-yl)-methacrylamide-bonded stationary phase (DCPak PTZ) combined with tandem mass spectrometry²⁴¹. Low temperatures (e.g. 10°C) and acidic conditions (e.g. 0.1% formic acid) provided selectivity for various isomeric sugars. Additionally, the interconversion between a

and β anomers was studied at different temperatures and mobile phase pH values²⁴¹ (Figure 2.7).

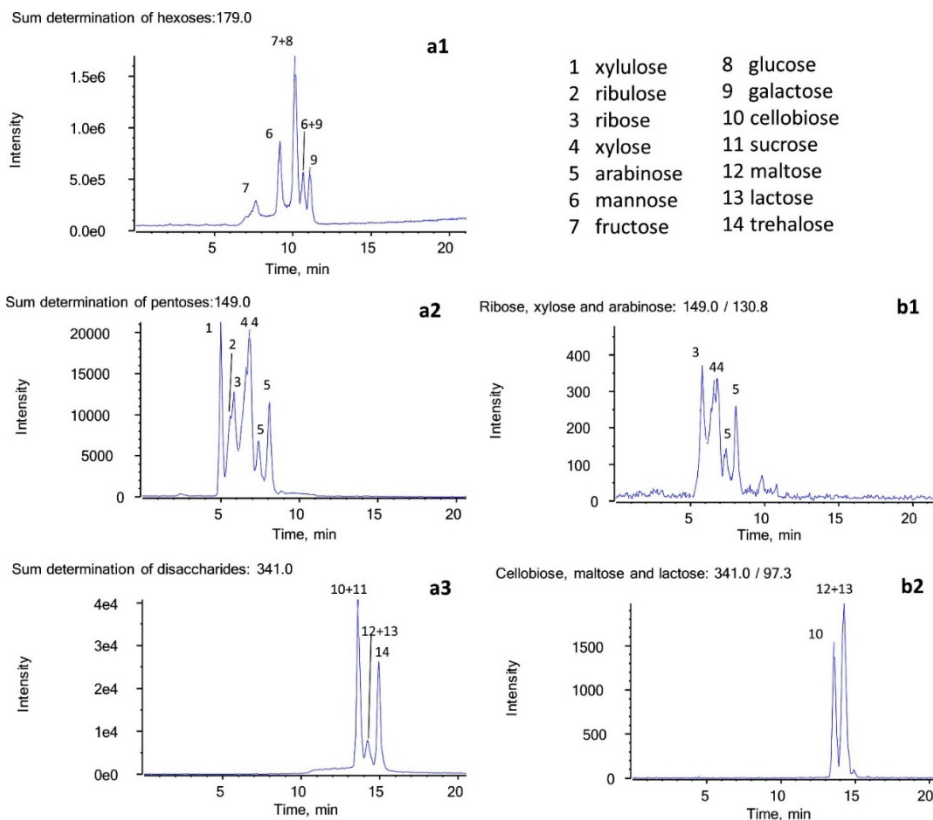


Figure 2.7. Isomeric separation of hexoses, pentoses, and disaccharides using a HILIC DCPak PTZ stationary phase combined with charged aerosol or mass spectrometric detection²⁴¹. **a1)** SIM chromatogram of a hexose mixture. **a2)** SIM chromatogram of a pentose mixture. **a3)** SIM chromatogram of a disaccharide mixture. **b1)** SRM chromatogram for aldopentoses (ribose, xylose, and arabinose). **b2)** SRM chromatogram for the disaccharides cellobiose, lactose, and maltose. Reprinted from Fu, Xiaoqing, et al. "Separation of carbohydrate isomers and anomers on poly-N-(1H-tetrazole-5-yl)-methacrylamide-bonded stationary phase by hydrophilic interaction chromatography as well as determination of anomer interconversion energy barriers." *Journal of Chromatography A* 1620 (2020): 460981., Copyright 2020, with permission from Elsevier.

Although one of the main advantages of HILIC is the avoidance of derivatization prior to chromatographic separation, post-column derivatization methods have been developed to increase MS sensitivity, allowing to decrease the sample volume. Li et al. separated monosaccharides of plant samples by a SeQuant ZIC-HILIC column followed by an online chemical derivatization using 4-benzylamino-benzeneboronic acid (labelling cis-diols) and ESI⁺-MS analysis (MRM mode). The observed increase in sensitivity ranged between a 36 to 42800 fold for the various analyzed monosaccharides²⁴². Heiss et al. eliminated the need of post-column derivatization hardware by developing a modified ESI interface, termed 'contained

electrospray ionization'. This allowed for an in-source, droplet-based derivatization with phenylboronic acid (PBA) forming monosaccharide-bisPBA derivatives, resulting in enhanced sensitivity by MS analysis. Since the low flow rate required for contained-ESI and the large volume of HPLC pre-column components results in a significant gradient delay, either hardware modifications needed to be incorporated or an estimated injection delay has to be programmed²⁴³.

Apart from the separation and detection of monosaccharides, HILIC has been applied to study nucleotide-sugars as well. Using a zwitterionic-silica based stationary phase (ZIC-HILIC, Merck SeQuant), 12 nucleotide-sugars could be detected and quantified in various plant cell wall extracts. Although different LC conditions were tested, the method was unfortunately insufficient to separate the structural isomers UDP-Gal/UDP-Glc and UDP-GalNAc/UDP-GlcNAc²⁰⁴.

Ma et al. (2017) applied HILIC-ESI-IT-TOF (employing an Accucore 150-amide HILIC column) to detect NT-sugars in the fungus *Botrytis cinerea*. They observed an accumulation of UDP-KDG, an intermediate of the UDP-Rhamnose biosynthesis pathway, in a *Botrytis cinerea* mutant, which was associated with a reduced virulence of the fungi²⁴⁴.

HILIC also made its appearance in supporting the discovery of novel biosynthesis pathways in bacteria. By using an Accucore 150-amid HILIC column, Li et al., looked into the intermediates of the CMP-Pse biosynthesis pathway of *Bacillus thuringiensis*. Where gram-negative bacteria, such as *C. jejuni* and *H. pylori* require six enzymes to convert UDP-GlcNAc into CMP-Pse, *B. thuringiensis* required seven enzymes. Using HILIC-ESI-MS complemented with NMR experiments, the enzymes Pen and Pal were identified, converting UDP-GlcNAc into UDP-2-acetamido-6-deoxy-D-xylo-hexopyranose-5,6-ene and UDP-2-acetamido-6-deoxy-D-xylo-hexopyranose-5,6-ene into UDP-2-acet-amido-2,6-dideoxy-L-arabino-hex-4-ulose²⁴⁵. Finally, a rapid HILIC-ESI-MS/MS method using a ZIC-pHILIC stationary phase was developed for the quantification of UDP-glucose and UDP-glucuronic acid. The method was validated and applied to extracts from plant materials²⁴⁶.

Other liquid chromatography separation techniques

Chiral chromatography. Chiral chromatography performs the separation of chiral compounds, such as the D and L enantiomers which determine the absolute configuration in sugars. The chiral stationary phase usually employs a chiral ligand attached to the surface of an achiral support. For a comprehensive overview and mechanism about chiral stationary phases for liquid chromatography we refer to the review of Teixeira et al²⁴⁷.

In sugars, D and L is determined based on the chiral carbon most remote from the carbonyl group, e.g. C-5 in glucose. The chromatographic determination of absolute configuration is particularly challenging. The enantiomeric separation of pentopyranoses and hexopyranoses was achieved after per-O-methylation and conversion into glycosyl chlorides followed by coupling with cesium salt of a fluorescent chiral derivatization reagent. Enantiomeric separation was achieved using a Develosil column²⁴⁸. Recently, boric acid was used as an additive to

enhance the separation of sugars by complexation to vicinal cis hydroxyl groups on the sugar molecules employing a cation-exchange column²⁴⁹. However, D and L sugars did not differ in their complexation abilities, and consequently showed the same retention times. Lopes and Gaspar performed the simultaneous separation of enantiomers (D/L) and anomers (α/β) of several underivatized monosaccharides (arabinose, ribose, mannose, fucose, xylose, lyxose, glucose and fructose) using the chiral separation phase Chiralpak AD-H²⁵⁰. However, their mobile phase consisted of a mixture of hexane–ethanol–TFA and hence was not combined with mass spectrometric detection.

Mixed mode chromatography. Mixed mode chromatography (MMC) combines multiple retention mechanisms within a single chromatographic separation phase²⁵¹. This provides flexibility and versatility for analyzing compounds with diverse physicochemical properties, e.g. ranging from polar to non-polar, making it particularly useful for complex samples²⁵¹⁻²⁵³. The concept of MMC was introduced in 1986, initially by simply combining two separate columns with different retention mechanisms. Later, stationary phases incorporating dual retention mechanisms were developed^{254,255}, which gained popularity, as evidenced by the increasing number of research groups developing MMC stationary phases. For a more comprehensive discussion on advances in mixed-mode chromatographic stationary phases, we refer to the review by Sykora et al.²⁵². The most common MMC columns integrate two or three retention mechanism, predominantly hydrophobic, hydrophilic and electrostatic interactions, although other mechanisms can also be included²⁵¹⁻²⁵³. Stationary phases can be categorized based on their structure, such as silica-based, polymer-based, hybrid and monolithic²⁵². Despite the potential for combining multiple retention mechanisms, in practice, often only one mechanism is used at a time, dictated by the mobile phase conditions²⁵². For instance, IEX columns exhibit electrostatic interactions at low acetonitrile concentrations, while HILIC retention occurs at high acetonitrile concentrations²⁵¹. Another example includes the developed bistrimethoxysilylethane hydrolysis aminopropyltrimethoxysilane (BTME-H-APS) silica stationary, which can be used for normal phase chromatography, IEC, or MMC combining HILIC and RP mechanisms. This column successfully separated different carbohydrates from biological samples, although it did not exploit the advantages of multiple retention mechanisms^{251,256}. Eastwood et al. did exploit these advantages by using the commercial Primsep SB column which allowed them to detect and purify the nucleotide-sugar GDP-trifluoromethyl fucose in an enzymatic reaction mixture, relying on both electrostatic as well as hydrophobic interactions. An ammonium formate buffer was used to ensure compatibility with ESI-MS²⁵⁷.

MMC shows a great potential for untargeted metabolomic studies, where the separation and identification of compounds with diverse characteristics is required. Ammann and co-workers managed by combining HILIC, anion exchange chromatography, cation exchange chromatography and RP to separate a variety of metabolite classes, such as amino acids, sugars, fatty acid derived compounds

and antioxidants all within a single run of 50 min, using an ESI-MS compatible method. However, this method did not resolve sugar isomers²⁵³. Xing et al. utilized a positively charged quaternary amine polyvinyl alcohol stationary phase coupled to ESI-MS, employing varying pH, salt concentrations and organic content to achieve both hydrophilic and electrostatic interactions. Their method covered 65.6% of a metabolite library comprising 607 metabolites within a 60-minute run, missing only the compounds which e.g. required positive mode detection or were unstable in the applied pH range. Moreover, they demonstrated the separation of isomers such as glucose and fructose and four different hexose-monophosphates²⁵⁸.

Another study demonstrated the potential for separating sugar isomers using a reverse-phase/weak anion exchanger operated under HILIC conditions. This method successfully separated six hexose monophosphates. By adjusting the temperature, the authors were able to fine-tune the conditions to achieve either the separation or co-elution of α - and β -anomers²⁵⁹. However, the mobile phase used, included trifluoroacetic acid (TFA), rendering it incompatible with ESI-MS. The choice of the mobile phase is also critical for successful MMC²⁵¹. As mentioned before (HILIC section), Su et al. investigated the separation of isomeric glycolytic phosphorylated monosaccharides. While the HILIC BEH amide column outperformed their MMC HILIC/SAX column (HILICpak VT-50 2D), the latter enabled the separation of F6P from G1P, which the former could not achieve. Consequently, they combined the HILIC BEH amide and MMC HILIC/SAX columns for complete separation²³⁰. More recently, Rahm and colleagues developed a MMC-ESI-MS method to separate and quantify 17 nucleotide sugars commonly found in human cells. They employed a mixed-mode separation phase (Atlantis Premier BEH C18 AX) that combines weak anion-exchange and reverse-phase chromatography. The mobile phase A consisted of water, while mobile phase B was an ammonium acetate buffer in 5% acetonitrile at pH 4.7²⁶⁰. Interestingly, the authors not only demonstrated good separation (including UDP-GlcNAc and UDP-GalNAc) and high robustness of the method, but they also identified distinctive fragment ion patterns for various UDP-hexoses, including UDP-mannose, UDP-glucose, and UDP-galactose, supporting their identification²⁶⁰ (Figure 2.8).

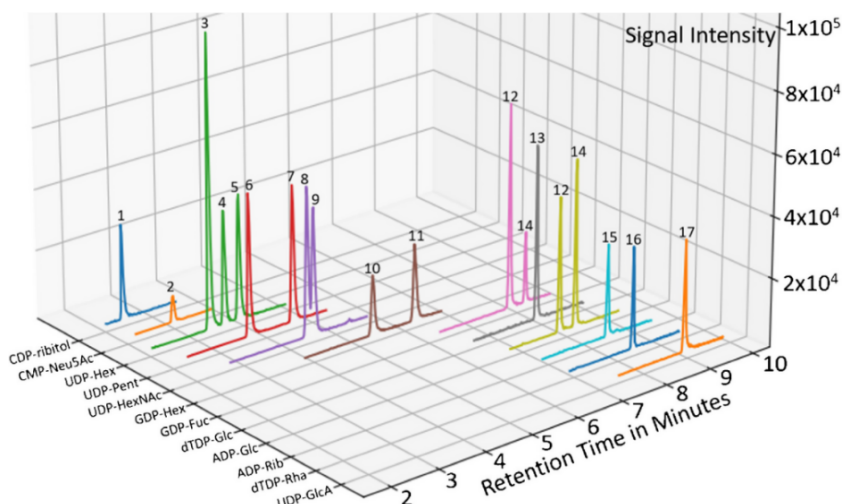


Figure 2.8. The separation of 17 (isomeric) nucleotide sugars commonly found in human cells, using mixed-mode chromatographic separation coupled to ESI-MS²⁶⁰. The graph shows the MRM transitions. for: 1. CDP-ribitol, 2. CMP-Neu5Ac, 3. UDP-Man, 4. UDP-Gal, 5. UDP-Glc, 6. UDP-Ara, 7. UDP-Xyl, 8. UDP-GalNAc, 9. UDP-GlcNAc, 10. GDP-Man, 11. GDP-Glc, 12. GDP-Fuc, 13. dTDP-Glc, 14. ADP-Glc, 15. ADP-Rib, 16. dTDP-Rha, 17. UDP-GlcA. Used with permission of Springer Nature BV, from Rahm, et al. Mixed-phase weak anion-exchange/reverse-phase LC-MS/MS for analysis of nucleotide sugars in human fibroblasts. Analytical and bioanalytical chemistry (2024): 1-10; permission conveyed through Copyright Clearance Center, Inc.

Supercritical fluid chromatography. A more recent approach for the analysis of monosaccharides is supercritical fluid chromatography (SFC) coupled to mass spectrometry^{261,262}. This technique employs a supercritical mobile phase, which typically consist of CO₂ due to its low critical parameters ($T_c = 31\text{ }^{\circ}\text{C}$ and $P_c = 74\text{ bar}$). SFC combines the advantages of both a liquid (i.e. density) and a gas (i.e. viscosity and diffusivity)²⁶³⁻²⁶⁵. The technique gained more attention when more robust instruments allowed an easy hyphenation to ESI-MS^{263,264}. Although SFC was originally used to analyze hydrophobic compounds due to the non-polar characteristics of CO₂, the addition of an organic modifier (frequently methanol), also allowed for the analysis of polar compounds^{263,265,266}. Highly polar compounds (such as monosaccharides and derivatives) can be analyzed by creating a gradient of 100% supercritical CO₂ to 100% organic modifier, thereby shifting the physical state by controlling pressure and temperature, known as unified chromatography²⁶³. Besides the advantage of the simultaneous analysis of both polar and non-polar compounds, the analysis time of SFC-MS is often much faster compared to traditional LC separation methods^{261,266}.

Desfontaine et al. evaluated SFC-MS for metabolomics. Their final optimized method employed a Poroshell HILIC (silica based) column with a linear gradient starting from 98% supercritical CO₂ and 2% organic modifier to 100% organic modifier. The organic modifier consisted of 95/5 MeOH/H₂O (v/v) with 50 mM ammonium formate and 1 mM ammonium fluoride. A make-up solvent of pure MeOH was applied²⁶⁶. Using this method, they later showed a detectability of 66%

from 597 compounds. Compounds with varying chemical characteristics were identified, including carbohydrates. Nevertheless, phosphate containing analytes, including sugar phosphates and nucleotides were poorly detected^{262,266}. The potential of SFC-MS for the analysis of monosaccharides has been demonstrated for several applications. For example, various monosaccharides (including the isomers arabinose/xylose and glucose/galactose) could be separated from plant samples using a HSS C18SB column with an analysis time of only 4.5 minutes²⁶¹. Yang and Zhengjin compared 10 different stationary phases for SFC to separate eight saccharides including fructose, glucose and sucrose²⁶⁷. Overall, SFC-MS shows great potential for metabolomics because it allows the combined analysis of apolar and polar compounds (including monosaccharide derivatives). However, parameters such as type of stationary phase, type and concentration of organic modifier, temperature, additives etc., require optimisation in order to obtain satisfactory performance.

2.4 | Gas chromatography coupled with mass spectrometry

Gas chromatographic separation operates by vaporizing the sample and passing it through a separation column with a stationary phase, separating compounds based on their volatility and interaction with the column material, providing high resolution and sensitivity. Gas chromatography (GC) can also be easily coupled with mass spectrometry, which makes it ideal for the analysis of complex mixtures^{42,268}. However, to make the non-volatile and thermally labile monosaccharide building blocks suitable for GC analysis these compounds require derivatization prior to analysis^{42,223}. This makes the sample preparation more demanding compared to other methods. While monosaccharides are detectable with a simple flame ionization detector the combination with mass spectrometry opened the window to the analysis of unknown derivatives and complex samples²⁶⁹. Nevertheless, because close isomers often show similar fragmentation patterns, the retention time obtained in gas chromatography remains crucial for identification²⁶⁸. The first application of gas chromatography to volatile derivatives of monosaccharides was carried out in the early 60s, followed by coupling GC to electron ionization (EI) MS by DeJongh et al. in 1969²⁶⁹. Since then, GC-EI-MS has become widely employed for the analysis of various glycomolecules. For example, gas chromatography combined with methanolysis – a mild reaction which cleaves oligosaccharide chains into monosaccharide components – evolved as one of the most widely employed approaches for the compositional analysis of carbohydrate-containing molecules^{42,270,271}. This has led to a substantial number of scientific articles, book chapters, and reviews covering the fundamentals of GC-MS and its applications to carbohydrates over the past decades^{42,272-279}. For example, the book “The Art of Carbohydrate Analysis” by Gerwig et al., provides a comprehensive introduction and detailed protocols for the application of GC to monosaccharide analysis⁴². Therefore, this chapter only offers

a concise overview of the basic concepts, recent advancements, and applications to the microbial world.

The preparation of volatile monosaccharide derivatives is challenged by the numerous functional groups, steric hindrance, presence of tautomeric forms, and molecular lability²⁶⁸. Recently, the presence of other compounds such as amino acids has also been shown to alter the derivatization products of monosaccharides²⁸⁰. Common derivatization strategies for mono- and oligosaccharides have been comprehensively reviewed in earlier articles^{44,223,268}, which will therefore only be covered briefly in the following.

Classical derivatization approaches include the formation of methyl ethers, acetates, trifluoroacetates, and trimethylsilyl ethers. These derivatization products offer good volatility and stability. However, sugars with free carbonyl groups can be present as different tautomeric forms, which are stabilized through derivatization. This issue can be partially solved by performing a two-step derivatization of the reducing sugars, consisting of oximation followed by silylation to produce trimethylsilyl oximes. This derivatization procedure results in only two peaks, one for the syn (E) oxime form and one for the anti (Z) oxime form. Moreover, these forms are not observed with non-reducing carbohydrates, because they do not form oximes. The peak splitting could be completely eliminated by reducing aldoses to alditol acetates, resulting in a single peak per sugar. However, different aldoses and ketoses can yield the same alditol upon reduction, making them indistinguishable through this procedure²⁶⁸. Sugar acids, such as aldonic, uronic, and aldaric acids have been derivatized with methods like silylation or acetylation²⁶⁸. The alpha keto acid containing nonulosonic acids have been derivatized using a two-step process: first, methyl esterification of the alpha keto acid using diazomethane, followed by derivatization of the hydroxyl groups with heptafluorobutyric anhydride²⁸³. Heptafluorobutyric anhydride derivatives are highly stable, exhibit excellent chromatographic properties on traditional methyl-siloxane columns, and are very volatile. This allowed the separation and identification of various nonulosonic acid derivatives, including O-acylated variants, methylated and sulfated derivatives, de-N-acetylated neuraminic acid, and sialic acid lactones. The unambiguous identification of nonulosonic acids from complex mixtures was based on distinct fragmentation patterns of the individual sialic acids^{281, 282,283}.

Phosphorylated sugars, such as glucose-6-phosphate, fructose-6-phosphate, and fructose-1,6-bisphosphate, are key intermediates in central metabolic pathways and are frequently studied in metabolic flux analysis of model cells including microbes. Their highly polar nature makes gas chromatographic analysis particularly challenging. However, their analysis has been achieved after derivatization with methoxyamine hydrochloride, which converts carbonyl groups into stable methoxime derivatives – which eliminates tautomeric forms – followed by derivatization using N-methyl-N-trimethylsilyl-trifluoroacetamide (MSTFA) to silylate the hydroxyl groups of the sugar backbone and the phosphate groups²⁸⁴⁻²⁸⁷. Thereby, gas chromatography achieved baseline separation of sugar phosphate

isomers, which allowed discrimination of positional phosphate isomers of metabolic intermediates (e.g. glucose-1P and glucose-6P). The typical ionization method for the GC-MS analysis of carbohydrates is electron ionization. However, when applied to sugar phosphates, this produces primarily fragment ions containing phosphate groups and only parts of the carbon backbone^{288,289}. This is disadvantageous for stable carbon isotope labelling experiments in metabolic flux analysis, as it results in a loss of labelling information. To prevent extensive fragmentation and consequently retain ¹³C labelling information, Okahashi et al. developed a GC-MS method using soft ionization by coupling gas chromatography to negative chemical ionization-mass spectrometry²⁸⁹. This method involves derivatization with pentafluorobenzyloxime (PFBO) and trimethylsilyl (TMS). PFBO acts as an electron-capturing substituent of electrons generated by collision with the reagent gas, making this approach more sensitive and selective²⁸⁹. Koubaa et al. also employed chemical ionization and reported improved isotopomer quantification in metabolic flux analysis using ¹³C-labeled sugars²⁹⁰.

An alternative labelling of monosaccharides has been employed by Faraco and colleagues. They employed methylboronic derivatization followed by acetylation to analyze monosaccharides in plant gum extracts. Baseline separation of all monosaccharides was achieved employing a phosphonium-based ionic liquid capillary column. This method proved to be robust in the presence of inorganic substances and did not necessitate intermediate cleanup or evaporation steps²⁹¹. In addition to determining the monosaccharide composition, methylation analysis has been used to identify glycosidic linkages in glycomolecules. After permethylation of the intact glycomolecules, the linkages are hydrolyzed, and the resulting monosaccharides are reduced and peracetylated. The partially methylated alditol acetates are then separated and analyzed by GC-MS. This not only identifies the monosaccharide constituents but also provides unambiguous information on the individual sugar linkage positions^{275,292,293}. The sample preparation has been further streamlined and miniaturized to make it suitable to small amounts of glycoproteins²⁷⁵. When performing linkage hydrolysis by methanolysis, the different cyclic tautomeric forms – in which monosaccharides exist in free solution – are locked as methyl glycosides. This increases the complexity of the data because it results in multiple peaks in the gas chromatogram²⁶⁸. On the other hand, the peak pattern allows to identify monosaccharides according to their fixed ratio of the α/β furanose and α/β pyranose forms. As mentioned earlier, the peak splitting can be avoided by modification of the anomeric centre²⁶⁸.

Interestingly, a more recent study reevaluated derivatization and separation condition to propose an optimized GC-MS procedure for the analysis of complex monosaccharide mixtures²⁹⁴. For example, the formose reaction – which is studied in the context of the origin of life – produces a remarkably complex mixture, ranging from the C1 building block formaldehyde to longer chain sugars hexoses and heptoses, which occur in different open-chain, furanose, and pyranose forms^{295,296}.

Their optimized method achieved complete separation of all oximated monosaccharides ranging from C2 to C7. However, while the method separated sugars by chain lengths into blocks and differentiated hexoses into aldoses and ketoses, it did not achieve baseline resolution of individual hexoses due to the large number of existing stereoisomers²⁹⁴ (Figure 2.9).

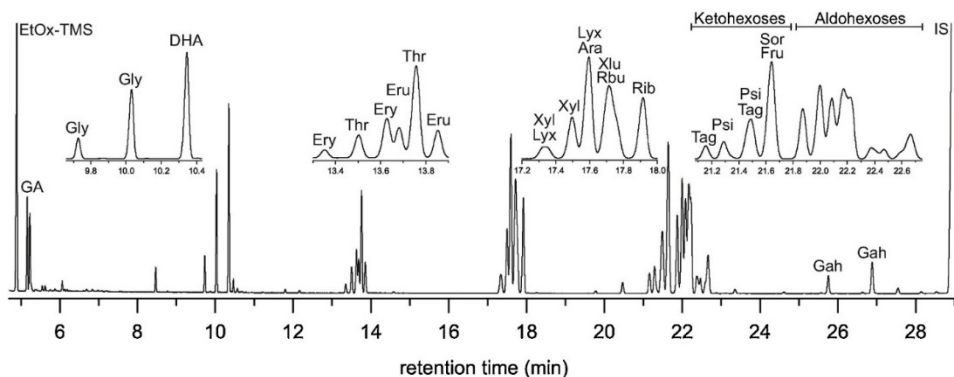


Figure 2.9. An optimized gas chromatographic separation of monosaccharides, ranging from C2 to C7 units, using a SE-52 column²⁹⁴. This was achieved following ethyloxime-trimethylsilyl derivatization and adjustment of the temperature program for monosaccharides. The shorter, unbranched compounds were individually separated, while the hexoses could only be partially separated due to the large number of stereoisomers. **GA:** glycolaldehyde; **Gly:** glyceraldehyde; **DHA:** dihydroxyacetone; **Ery:** erythrose; **Thr:** threose; **Eru:** erythrulose; **Xyl:** xylose; **Lyx:** lyxose; **Ara:** arabinose; **Xlu:** xylulose; **Rbu:** ribulose; **Rib:** ribose; **Tag:** tagatose; **Psi:** psicose; **Sor:** sorbose; **Fru:** fructose; **Gah:** galactoheptose; **IS:** internal standard (phenyl- β -D-glucopyranoside). Reprinted from Haas, Maren, Saskia Lamour, and Oliver Trapp. Development of an advanced derivatization protocol for the unambiguous identification of monosaccharides in complex mixtures by gas and liquid chromatography. *Journal of Chromatography A* 1568 (2018): 160-167, Copyright 2018, with permission from Elsevier.

Albeit these mixtures did not contain any derivatives, this clearly demonstrates that even methods with high resolving power face challenges in separating complex sugar mixtures. Therefore, a truly enantioselective separation of monosaccharides requires the application of either chiral stationary phases or reacting sugars with a chiral reagent to form diastereomers, which are then separated on an achiral column²⁹⁷⁻²⁹⁹. For example, there are four aldopentoses—lyxose, xylose, arabinose, and ribose—which in solution exist in equilibrium between the open-chain form, the α and β furanose forms and the α and β pyranose forms. Considering the two enantiomeric forms (D and L configurations), this accounts for 32 distinct aldopentose molecules. The demand for enantioselective analysis increases with the diversity of monosaccharide types. Myrgorodska et al., demonstrated the enantioresolution of a racemic mixture of aldopentoses following derivatization with trifluoroacetic anhydride and a two-dimensional gas chromatography including separation on a β -cyclodextrin column³⁰⁰. Interestingly, the mass spectrometric fragmentation also showed distinct patterns for pyranose and furanose forms (Figure 2.10).

As mentioned above, carbohydrates are polar and thermally unstable, and their gas chromatographic analysis requires derivatization of the polar groups. Practically, this limits the approach to smaller sugars, such as mono-, di-, and trisaccharides, and hinders the analysis of nucleotide-activated monosaccharides. The latter are thermally labile, which decompose at the high temperatures typically used in gas chromatography³⁰¹.

Finally, apart from optimizing derivatization and column chemistry, advancements include coupling to high resolution mass spectrometers, such as GC-TOF-MS or GC-Orbitrap-MS. This enhances untargeted metabolomics work including the identification of novel monosaccharide derivatives³⁰². For advances in high-resolution GC-MS technologies we refer to the recent review by Misra Biswapriya³⁰³. The high performance and robustness of GC-MS methods have been utilized for the analysis of monosaccharides and their derivatives across various fields, including food analysis^{304,305}, detection of specific sugars as surrogate indicators and biomarkers^{306,307}, for the compositional analysis of plant and fungal cell wall components^{57,308-310}, and for the characterization of different bacterial glycocomolecules^{311,312}.

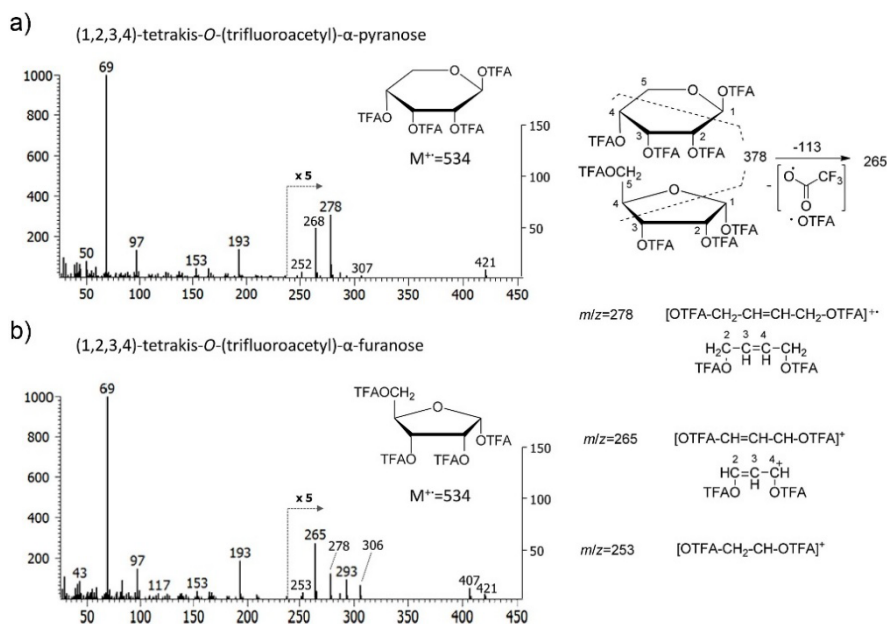


Figure 2.10. Fragmentation analysis of ribopyranose and ribofuranose O-trifluoroacetyl derivatives using two-dimensional gas chromatography coupled to EI-MS³⁰⁰. Distinct fragmentation patterns for the pyranose and furanose form were identified. **a)** Fragmentation pattern of the pyranose form. **b)** Fragmentation pattern of the furanose form. Structural diagrams in the graph illustrate the parent ions and the key fragment ions. Arrows indicate bond cleavages and rearrangements that led to the observed fragments. Reprinted from Myrgorodska et al. Enantioresolution and quantification of monosaccharides by comprehensive two-dimensional gas chromatography. *Journal of Chromatography A* 1487 (2017): 248-253, Copyright 2017, with permission from Elsevier.

2.5 | Capillary electrophoresis coupled with mass spectrometry

Capillary electrophoresis is a highly effective method for separating polar and charged metabolites, particularly for substances like amino acids, nucleotides, small organic acids, sugars and sugar phosphates^{313,314}. In metabolomics studies, capillary zone electrophoresis (CZE) has been the predominant separation mode in CE-MS³¹⁴⁻³¹⁷. Thereby, a narrow capillary is filled with an electrolyte solution, where metabolites are separated based on their charge-to-size ratio when an electric field is applied. The migration rate of molecules through the capillary depends on several factors, including the charge of the analyte, charge-to-hydrodynamic volume ratio, buffer pH and ionic strength, applied voltage, temperature, capillary length and diameter, and on the surface charge of the capillary wall³¹⁷⁻³²³. CE offers several advantages over liquid chromatography, including a higher resolution and a faster analysis times. In addition, due to lower sample and reagent consumption, CE can be more environmental friendly. CE also requires only minimal sample preparation. However, CE has limitations such as challenges in separating neutral compounds and in regard to reproducibility of migration times, which can be affected by variations in the capillary surface and buffer conditions. Furthermore, in CE, the different charge states of metabolites necessitate multiple analytical runs, such as conducting separate runs for cationic and anionic analytes. The broader application of CE-MS to metabolomics studies, including separation modes, capillary coatings, and sample preparation techniques has been extensively reviewed by Ramautar and colleagues^{313-315,324-328}, and will only be discussed briefly here in the context of monosaccharides and nucleotide activated sugars.

CE is often paired with UV, conventional fluorescence, and/or laser-induced fluorescence (LIF) detection^{316,317,323}. For example, very sensitive oligosaccharide analysis with capillary gel electrophoresis (using conventional DNA sequences) has been performed after labelling with 9-aminopyrene-1,4,6-trisulfonic acid (APTS), and coupling to LIF for detection^{329,330}. While this allows for straightforward and reliable compound detection in many applications, additional mass spectrometric detection is often necessary for identifying compounds, especially in the analysis of sugar derivatives, unknown metabolites, and complex cellular extracts such as those from microbes. Despite its high separation performance, CE remained underrepresented in metabolomic studies – including the analysis of monosaccharides – compared to more common LC and GC approaches. This is mainly because CE-MS is considered technically more challenging and less reproducible than other separation methods³¹⁵. Moreover, the lack of convenient interfaces for connecting CE with mass spectrometers has challenged their application^{327,331}. Generally, CE faces challenges with low liquid flow rates suboptimal for electrospray ionization. Sheath flow interfaces mix a sheath liquid with CE effluent to increase flow rates but may reduce sensitivity and may lead to peak distortion³³²⁻³³⁴. Sheathless-flow interfaces struggle with spray stability and reproducibility³³³. Liquid-junction interfaces partially disconnect CE and ESI components and produce a stable electrospray^{331,335}. In addition, bare fused-silica

capillaries compromise separation efficiencies and migration time reproducibility due to potential analyte interactions. Surface coatings are commonly applied to prevent these interactions, stabilize the electroosmotic flow (EOF), and support MS coupling³³⁶. Moreover, commercially available microfluidic CE systems with integrated nano electrospray ionization interfaces have addressed this issue^{331,337}. As mentioned earlier, CE relies on the charge of molecules for separation. Thus, CE has been frequently employed for analyzing charged metabolites, such as organic acids, sugar-phosphates and nucleotides. However, many monosaccharide building blocks are neutral, posing a challenge for their analysis by CE. Consequently, various carbohydrate labelling strategies have been developed to introduce a charge and overcome these issues. Reductive amination of the aldehyde group at the reducing end of sugars is one of the most commonly used derivatization methods³²³. The applied labels often include a fluorophore and are therefore paired with sensitive fluorescence detection. For instance, monosaccharides have been separated by CE following labelling with 1-phenyl-3-methyl-5-pyrazolone, 2-aminobenzoic acid, or 7-aminonaphthalene-1,3-disulfonic acid³²³. The latter method facilitated the separation of ribose enantiomers³³⁸. An updated protocol detailing the analysis of monosaccharides following 2-aminobenzoic acid has also been published recently³³⁹. Another method was developed by Guttman, who performed CE analysis of APTS-labelled monosaccharides and 2-aminoacridone (AMAC) labelled sialic acids coupled with LIF detection³⁴⁰. Szabo and colleagues used 2-AMAC to develop a rapid sialic acid speciation method. The classical reductive amination in the presence of sodium cyanoborohydride in DMSO/acetic acid caused the sialic acids to decarboxylate spontaneously. The resulting neutral sialic acids were analyzed using a neutral-coated capillary (to eliminate the EOF) in reversed polarity mode, with a background electrolyte (BGE) containing 200 mM boric acid³⁴¹. This achieved complete baseline separation between Neu5Gc and Neu5Ac. However, MS was only used to investigate the decarboxylation of sialic acids. Schmitt-Kopplin optimized borate concentration and buffer pH for the separation of acidic monosaccharides and their 1,4-lactones without the need for derivatization³⁴². In fact, many of these methods employed buffers which are incompatible with mass spectrometry³³¹. Similarly, chiral ligand-exchange capillary electrophoresis successfully achieved the enantiomeric separation of monosaccharides, but it was only utilized in conjunction with fluorescence detection^{343,344}. Furthermore, advancements in chiral capillary electrophoresis has been reviewed by Clark et al., recently³⁴⁵.

Khatri and co-workers developed a CE-MS method for the analysis of monosaccharides, oligosaccharides and glycopeptides employing a microfluidic separation system with a nanoESI source as a detachable interface³³¹. The interface minimized gaps, dead volumes, and sample dilution. Aminopropyl silane coating on the capillary surface reduced the EOF. Monosaccharides were labelled with aminoxy TMT reagent to introduce a positive charge, facilitating the

electrophoretic migration of neutral saccharides and enabling multiplexing in tandem-MS experiments (Figure 2.11).

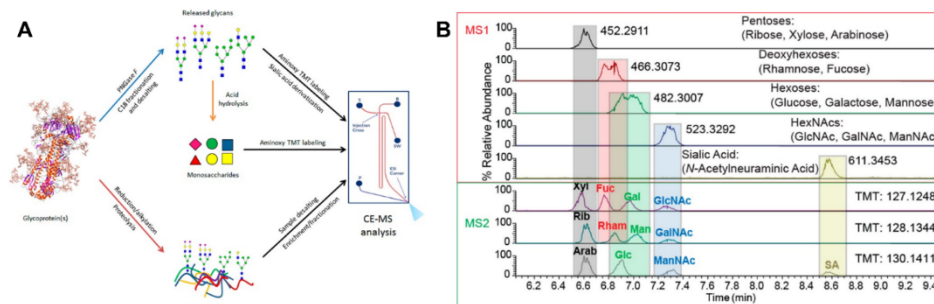


Figure 2.11. Khatri et al., developed a microfluidic capillary electrophoresis–mass spectrometry approach for the analysis of monosaccharides, oligosaccharides, and glycopeptides³³¹. **A)** Workflow for the simultaneous analysis of monosaccharides, oligosaccharides, and glycopeptides using a microfluidic CE-ESI-MS approach, including the CE/ESI chip interface. Neutral monosaccharides are analyzed following aminoxy TMT labeling, while sialic acids are analyzed label-free. **B)** The multiplexed analysis of common monosaccharides, utilizing different aminoxy TMT reagents to distinguish isomers based on migration time and fragment ion profiles. Reprinted with permission from Khatri, Kshitij, et al. Microfluidic capillary electrophoresis–mass spectrometry for analysis of monosaccharides, oligosaccharides, and glycopeptides." *Analytical chemistry* 89.12 (2017): 6645–6655. Copyright 2017, American Chemical Society.

Additionally, a sialic acid derivatization method was applied to prevent interaction between negatively charged carboxyl groups and the positively charged capillary surface coating. Lageveen-Kammeijer and co-workers introduced a uniform positive charge to oligosaccharides by derivatizing them with linkage-specific sialic acid and labelling them with Girard's Reagent P hydrazide³⁴⁶. The authors enhanced labelling efficiency and devised a one-step, cleanup-free, rapid protocol, enabling direct injection of labelled sugars into the CE-MS system. This facilitated ultra-sensitive analysis of oligosaccharides. However, to the best of the authors' knowledge, this approach has not yet been applied to monosaccharide derivatives. Sugar acids including nonulosonic acids can be analyzed by CE without further derivatisation³⁴⁷. Ortner and Buchberger demonstrated this by performing quantitative analysis of the underivatized sugars, extracted from glycoproteins after acid hydrolysis, to determine Neu5Ac/Neu5Gc ratios in various samples³⁴⁸. The same group also investigated the possibility to analyze neutral carbohydrates by CE-MS without derivatization, using strongly alkaline (pH > 12) carrier electrolytes compatible with mass spectrometry³⁴⁹. The high pH favored detection in negative ionization mode³⁴⁹. Daniel and colleagues expanded this method to include tandem mass spectrometry, which facilitated the unambiguous identification of the analyzed monosaccharides³⁵⁰.

Furthermore, phosphate and nucleotide-activated sugars do not require derivatization or elevated buffer pH for their analysis with CE. A review with focus on anionic metabolites has been conducted only recently³⁵¹. Soga et al.

investigated the anionic metabolic profile of *B. subtilis* extracts using CE-MS. They employed a cationic polymer-coated capillary and weakly alkaline ammonia buffers, utilizing reversed CE polarity and negative ionization mode, which allowed them to analyze sugar phosphates and nucleotides from the complete central carbon metabolism. Interestingly, to overcome corona discharge issues in negative mode – electrical discharge which results from the ionization of a fluid surrounding a conductor that is subjected to a high voltage – Zhang et al., introduced a sheathless CE-MS method for efficient and sensitive nucleotide analysis in positive ion mode³⁵².

Capillary electrophoresis has also been combined with simple UV detection (260 nm), as demonstrated recently for the analysis of sugar nucleotides from mammalian cells³⁵³⁻³⁵⁵. However, the identification of the peaks requires comparing migration times to reference sugar nucleotide samples. More sophisticated experiments, such as those involving stable isotope incorporation experiments require mass spectrometric detection. For example, the detection of sugar phosphates and their metabolic turnover rates in the Calvin cycle of tobacco leaves was measured³⁵⁶. The isotopomer abundance changes when switching to ¹³C-labeled CO₂ were obtained by coupling capillary electrophoresis with tandem mass spectrometry, operating in multiple reaction monitoring mode. This provided accurate concentrations of the sugar phosphate intermediates and revealed significantly lower turnover of glucose-1-phosphate compared to glucose-6-phosphate.

Finally, Soo and co-workers developed and applied a novel CE-MS precursor ion scanning approach to detect nucleotide-activated sugars, involved in the glycoconjugate biosynthesis of *Campylobacter jejuni* strains³⁵⁷. The authors employed N-(2-hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid) as a sample stacking buffer, achieving detection limits down to 0.2 pmol/mL and a good linear dynamic range for various nucleotide sugar standards. Their parent ion scanning approach identified CMP-linked nonulosonic acids and potential UDP-linked precursor sugars. Although the authors did not provide a complete description of the sugar nucleotide profile in this article, the approach of comparing a parent strain with isogenic mutants shows great potential for delineating the biosynthesis routes of microbial glycoconjugates.

2.6 | Matrix-assisted laser desorption/ionization mass spectrometry

This chapter offers a concise overview of the applications of matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) to the analysis of monosaccharides, derivatives and activated forms. For comprehensive introduction on the fundamentals of MALDI mass spectrometry, as well as general aspects related to matrices and sample preparation, we would like to refer to earlier reviews^{6,37,358-363}. In MALDI-MS the sample is co-crystallized with a matrix compound that absorbs laser energy, typically from UV lasers such as nitrogen lasers (337 nm) or Nd:YAG lasers (355 nm or 266 nm), resulting in the desorption

and ionization of the analytes³⁶⁴. MALDI typically generates molecular ions as adducts with cations such as protons or sodium ions, resulting in spectra with predominantly singly charged species such as $[M+H]^+$ (protonated) and $[M+Na]^+$ (sodiated), and minimal fragmentation^{6,37,364}. The generally gentle ionization process, allows for the ionization and analysis of fragile and large biomolecules. For carbohydrates, positive-ion MALDI-MS predominantly generates sodiated (or potassiated) rather than protonated species, owing to their strong affinity for these alkali metal ions^{6,365,366}. However, the ionization of carbohydrates as sodiated species is less sensitive compared to the analysis of peptides, where protonated species dominate. As a result, higher laser power is typically required for carbohydrate analysis⁶. A theoretical study on glucose also demonstrated that most protons produced by MALDI attach to matrices rather than to glucose, due to lower proton affinity of glucose and the higher concentration of matrix molecules. Even when protons do attach to glucose, protonated glucose rapidly dissociates into fragments because of the high temperatures during MALDI³⁶⁷. For neutral carbohydrate oligomers, MALDI-TOF has demonstrated consistent ionization efficiency with increasing molecular size, in contrast to ESI, where efficiency declines with increasing molecular weight^{368,369}. Consequently, MALDI-MS is especially useful for profiling mixtures of neutral oligosaccharides. Furthermore, MALDI-MS generally requires minimal sample preparation, and through careful selection and preparation of the matrix it can analyze complex mixtures at high sensitivity^{6,37,362,370,371}. Another advantage is the possibility to provide spatial information of samples, through MALDI imaging mass spectrometry (MSI), which has been reviewed in several articles recently³⁷²⁻³⁷⁵. The high spatial resolution enables detailed mapping of metabolites at cellular and even sub-cellular levels in tissues from clinical samples and microbial biofilm samples³⁷⁶⁻³⁷⁸. However, MALDI-MS faces challenges in shot-to-shot reproducibility and sample heterogeneity due to inhomogeneous sample/matrix co-crystallization, which leads to "hot spots" with high matrix concentrations that significantly increase ionization efficiency compared to nearby areas³⁷⁹. This impacts quantitative reproducibility and sample preparation automation. Nevertheless, the direct application of crude samples and the short analysis times enable high-throughput metabolic profiling, allowing for the analysis of hundreds to thousands of samples per day³⁸⁰⁻³⁸³. Excellent matrices are available for proteins, peptides, and polysaccharides, such as α -cyano-4-hydroxycinnamic acid³⁸⁴ and 2,5-dihydroxybenzoic acid³⁸⁵, including several other organic acids, ionic liquids or nanoparticles^{6,37,386}. Furthermore, MALDI-MS is also a powerful approach to study and monitor glycomolecules for different fields of research. Applications to oligosaccharides and glycoconjugates have been extensively reviewed by Harvey and others^{6,37,45,49,365,387-396}.

However, the analysis of small mass compounds – such as monosaccharides and activated sugars – can be complicated by the spectral interference in the low m/z region, which is produced by conventional organic matrices^{388,397}. Therefore, the

selection of matrices that minimize interference in the low mass range is crucial for the analysis of monosaccharide building blocks. An overview on current matrices and strategies to design matrix systems for low molecular weight compounds has been provided in recent articles^{388,397,398}. Coupling MALDI to very high-resolution mass spectrometers, such as Fourier Transform Ion Cyclotron Resonance also enhances discrimination between matrix and analyte peaks in the low mass range³⁹⁹⁻⁴⁰¹. Furthermore, such very high resolution mass analyzers provide the sum formulae of unknown derivatives (and their fragments), which are commonly encountered in microbial samples.

Efforts have been made to adapt classical matrices for low molecular weight applications through structural modifications, such as polymerization or conversion into ionic liquids^{397,402}. For example, Schmidt De León et al., employed norharmane containing ionic liquid matrices for the analysis of low molecular weight carbohydrates⁴⁰³. Additionally, new high molecular weight matrices, which are potentially "MALDI silent", meaning they do not result in matrix-related peaks in the spectrum, are under investigation. Horatz et al. tested different conjugated polymers which provided comparable profiles to 2, 5-dihydroxybenzoic acid (DHB) in positive mode and 9-aminoacridine in negative mode, making these suitable for metabolic studies⁴⁰⁴. Another higher molecular weight matrix was applied with addition of potassium hydroxide as dopant, which allowed direct analysis of fructose/glucose and sucrose from different beverages at high sensitivity⁴⁰⁵. Finally, mesoporous silica SBA-15, modified with 1,8-naphthalimide, was developed as a matrix for small molecules such as monosaccharides and amino acids. This matrix exhibited minimal background interference in the low mass range⁴⁰⁶.

Other strategies include reactive matrices that not only promote desorption and ionization of the analyte but also derivatize it to enhance detection and characterization³⁹⁸. For example, the aldehydes and ketones of carbohydrates react with amines to form Schiff bases. This reaction has been reported to improve the detection limit when using anthranilic acid or 2-Phenyl-3-(p-aminophenyl) acrylonitrile for various compounds, including oligosaccharides^{407,408}. Other reactive matrices include hydrazines, which condense with carbonyl groups to form hydrazones^{397,398}. While these matrices have rarely been employed for the analysis of monosaccharide components to the best of the authors' knowledge, these have the potential to facilitate the analysis in positive mode with good sensitivity. For example, Han et al. developed an on-tissue derivatization strategy for imaging and quantifying monosaccharides in biological tissues, employing the derivatization reagent 1-naphthaleneacethydrazide. The reactive matrix increased sensitivity and produced sugar-specific diagnostic ions during MALDI-MS/MS analysis⁴⁰⁹. This enabled the authors to distinguish between glucose and fructose. MALDI matrices for positive ion mode typically have acidic functional groups, facilitating proton transfer. In contrast, matrices for negative ion mode need to be moderately strong bases to abstract protons from the analyte, as per the Brønsted-Lowry acid-base theory^{397,398}. Many matrices are designed for positive ion mode, but

negative ion mode matrices often offer the advantage of fewer background signals^{397,398}. They are particularly useful for metabolomic studies, as many sugar intermediates in central metabolic pathways are negatively charged. Examples for negative mode matrices which were used for low mass applications are 9-Aminoacridine and 2-Aminoacridine (9AA and 2AA). The amine group in these matrices accepts protons from acidic analytes, resulting in the formation of deprotonated and negatively charged analyte species $[A-H]^{-410,411}$. 9AA was successfully used in the analysis of low molecular weight metabolites including organic acids, sugar phosphates, and sugar nucleotides⁴¹²⁻⁴¹⁴. Advantageously, sugar nucleotides are negatively charged compounds. While their analysis using MALDI-MS has not been widely employed, Heinrich et al. demonstrated the possibility of analyzing activated sugars using negative MALDI-MS⁴¹⁵. Steinhoff et al. investigated a rapid MALDI-MS method in negative mode for determining the energy charge of mammalian cells⁴¹⁶. The method provided insights not only into nucleotides but also into nucleotide sugars, such as UDP-HexNAc and UDP-Hex. By spiking ¹³C-labeled ATP into the cell lysates, the authors furthermore determined the degree of in-source decay for different negative mode matrices, specifically 9-aminoacridine (9-AA) and 2,4,6-trihydroxyacetophenone (2,4,6-THAP) (Figure 2.12). In-source decay was found to be strongly dependent on the applied laser power and the extraction pulse delay⁴¹⁶. The same MALDI-MS approach was applied to rapid monitoring of metabolites from central metabolic pathways of Chinese hamster ovary cells, including sugar phosphate intermediates, nucleotides and sugar nucleotides^{414,417,418}.

MALDI-MS was also used to monitor the synthetic reactions of monosaccharides, as summarized in earlier articles and reviews^{391,393,395,396,419}. Additionally, advanced spotting and sample target devices enabled the detection of metabolites (e.g. sugar phosphates) from single mammalian and microbial cells^{413,420-422}. Furthermore Gouw et al. introduced a quaternary ammonium center into sugars such as glucose, glucosamine, cellobiose, sucrose, raffinose, and sorbitol using glycidyltrimethylammonium chloride or Girard's reagent T. The modified sugars, when analyzed with DHB, produced intense signals, while the matrix-related signals remained comparatively weak⁴²³. Another challenge when analyzing bacterial cells or biofilms directly is the efficient extraction of analytes from deeper within the bacterial biomass and from the cytoplasm of cells. To overcome this limitation, the authors employed a pulsed infrared laser to disrupt and ablate bacterial cells without matrix, followed by laser-induced post ionization (MALDI-2), facilitating the detection of multiple small molecules. This holds a great promise for the direct analysis of microbial samples.

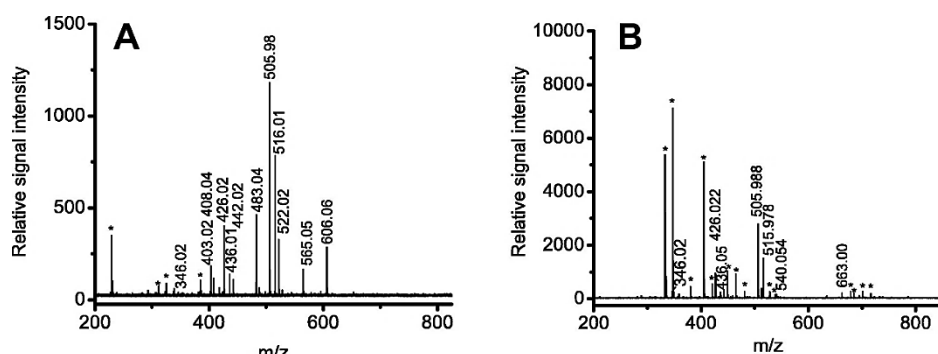


Figure 2.12. Steinhoff et al.⁴¹⁶ present a rapid negative mode MALDI-TOF MS method for analyzing central metabolic metabolites, including nucleotides and sugar nucleotides. The authors compared different matrices, specifically **A)** 9-aminoacridine and **B)** 2,4,6-trihydroxyacetophenone, and assessed the degree of in-source decay using ¹³C-labeled ATP (516.01 m/z). In addition to determining the energy charge, the method also provided strong signals for common nucleotide sugars, e.g. UDP-Hex (565.05 m/z) and UDP-HexNAc (505.06 m/z). Reprinted from Steinhoff, Robert F., et al. "Rapid estimation of the energy charge from cell lysates using matrix-assisted laser desorption/ionization mass spectrometry: Role of in-source fragmentation." *Analytical biochemistry* 447 (2014): 107-113., Copyright (2014), with permission from Elsevier.

Although not yet demonstrated for monosaccharides, innovative sample preparation techniques for MALDI-MS have significantly advanced carbohydrate analysis. For example, methods such as using frozen matrix/sample solutions or the Rapidly Freeze-Drying Droplet (RFDD) approach improve sample morphology homogeneity, lower the laser energy threshold, reduce fragmentation, and resolve the 'sweet spot' issue commonly encountered with traditional preparation methods^{424,425}. Finally, matrix-free strategies which are suitable for small-molecules such as monosaccharides, should also be mentioned. An example includes pulsed laser desorption-ionization from a porous silicon surface⁴²⁶. Nevertheless, MALDI-MS has generally found limited applications in exploring monosaccharide components, primarily due to its inability to discriminate between structural and positional isomers—a critical factor when studying glycomolecules. Therefore, recent approaches have aimed to combine MALDI with ion mobility to provide an additional dimension of selectivity^{389,427}. This demonstrated great potential for the untargeted identification of secondary metabolites, such as carbohydrates, in microbial cells and microbial biofilms^{428,429}. The combined application of laser-based ionization with ion mobility spectrometry has been reviewed previously^{389,430}. The application of ion mobility mass spectrometry to monosaccharide components is reviewed in Chapter 2.8.

2.7 | Flow Injection mass spectrometry

During flow injection mass spectrometry, the samples are injected into the mobile phase, which directly enters the mass spectrometer. This flow injection analysis of biological samples offers high-throughput profiling, making it ideal for studies with large sample numbers. Key advantages over chromatographic methods include

extremely short analysis times (typically under 60 seconds) and simplicity. The increasing application of this approach has been driven by advancements in sensitivity and mass resolving power. This has led to its use in clinical diagnostics, metabolomics, and environmental sciences. Developments and applications of flow injection analysis have been described in recent reviews⁴³¹⁻⁴³⁴. However, like direct MALDI-MS analysis, flow injection MS offers lower selectivity compared with chromatography-based approaches. Compounds such as structural isomers and enantiomers cannot be distinguished by mass alone. Some compounds may also undergo in-source ion fragmentation, which fragments may mimic ions of other analytes⁴³². Consequently, this approach is not particularly popular for studying monosaccharide components, except in the context of high-throughput monitoring of central metabolic pathways⁴³⁵. However, Nagy and Pohl reported the application of a chiral mass spectrometry approach using a fixed ligand kinetic method, which allowed for the discrimination of sugar isomers by mass spectrometry. In this method, sugars form a trimeric ion complex with ligands in the gas phase via electrospray ionization, which is then subjected to collision-induced dissociation, resulting in two fragment ions. The relative intensities of these two fragments are characteristic of each sugar isomer. This approach successfully distinguished the complete set of pentose isomers, as well as 24 aldohexose and 2-ketohexose isomers^{436,437} (Figures 2.13 and 2.14).

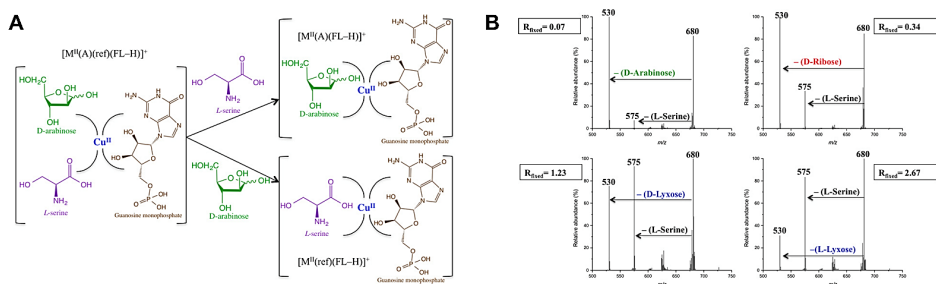


Figure 2.13. Nagy and Pohl demonstrate a mass spectrometric method called fixed ligand kinetic method for distinguishing all 12 pentose isomers, including their absolute configuration⁴³⁷. **A)** Fixed ligand components: D-arabinose as the analyte (A), copper(II) as the metal cation (M^{II}), L-serine as the chiral reference (ref), and guanosine monophosphate as the fixed ligand (FL). These components form a trimeric complex $[M^{II}(A)(ref)(FL-H)]^{2+}$ during electrospray ionization. The trimeric complex is then fragmented by collision-induced dissociation into the diastereomeric fragments $[M^{II}(A)(FL-H)]^{2+}$ and $[M^{II}(ref)(FL-H)]^{2+}$. **B)** Fragmentation spectra of the trimeric complex with the fixed ligand combinations of $Cu^{II}/L-Ser/5'GMP$, for the pentoses D-arabinose, D-ribose, as well as D/L-lyxose. Reprinted with permission from Nagy and Pohl. Monosaccharide identification as a first step toward de novo carbohydrate sequencing: mass spectrometry strategy for the identification and differentiation of diastereomeric and enantiomeric pentose isomers. Analytical chemistry 87.8 (2015): 4566-4571. Copyright 2015, American Chemical Society.

Later, Woocke and colleagues combined this method with prior chromatographic separation using ion-exclusion chromatography⁴³⁸. The eluents from the separation

system were then mixed with a fixed ligand combination before performing mass spectrometric experiments to determine fragment ion ratios. Furthermore, Giorgi and Speranza demonstrated the isomeric discrimination of underivatized monosaccharides as their noncovalent hydrazine complexes based on their fragmentation pattern, after low-energy collision-induced fragmentation using an ion trap mass spectrometer⁴³⁹. Interestingly, Madhusudanan and Srivastava demonstrated that oxocarbenium ions of aldohexoses and ketohexoses produce characteristic tandem mass spectra in the m/z 80 to 120 range⁴⁴⁰. This facilitated the differentiation of a range of diastereomers (Figure 2.15).

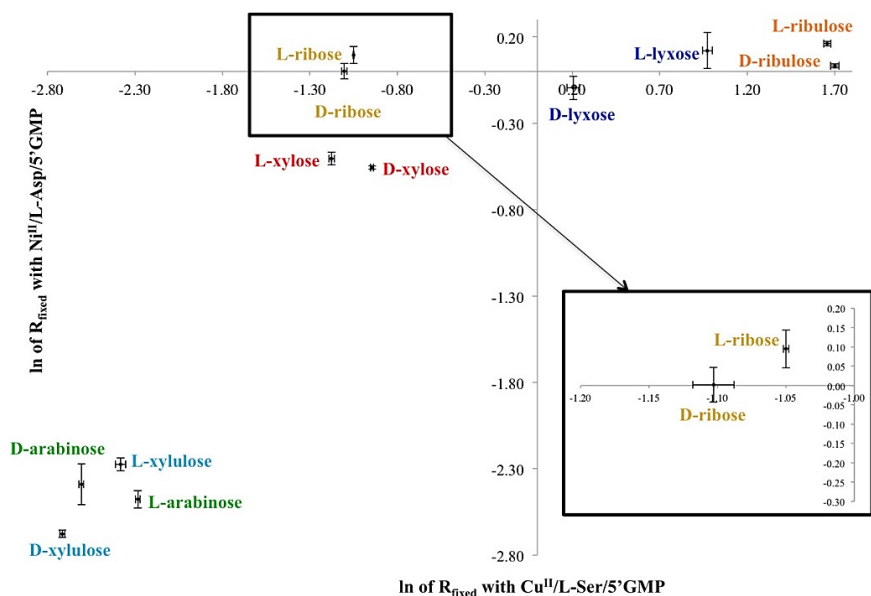


Figure 2.14. Nagy and Pohl demonstrate a mass spectrometric approach “fixed ligand kinetic method” for distinguishing all 12 pentose isomers, including their absolute configuration⁴³⁷. The plot shows the unique R_{fixed} values, with Cu^{II}/L-Ser/5'GMP on the x-axis and Ni^{II}/L-Asp/5'GMP on the y-axis, allowing to differentiate all pentose isomers. The error bars represent one standard deviation. Reprinted with permission from Nagy and Pohl. Monosaccharide identification as a first step toward de novo carbohydrate sequencing: mass spectrometry strategy for the identification and differentiation of diastereomeric and enantiomeric pentose isomers. *Analytical chemistry* 87.8 (2015): 4566-4571. Copyright 2015, American Chemical Society.

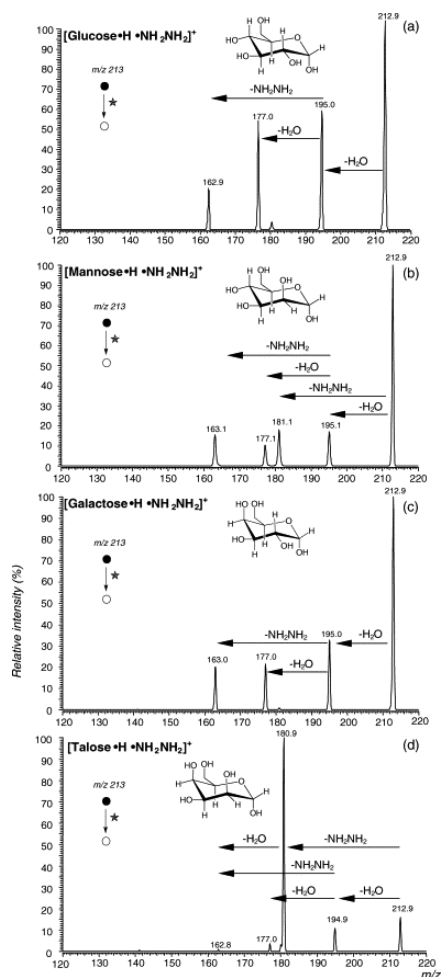


Figure 2.15. Giorgi and Speranza demonstrate how complexes between underivatized monosaccharides and hydrazine exhibit distinct losses of hydrazine and multiple water molecules during collision-induced dissociation⁴³⁹. The spectra, from top to bottom, display selected low-energy MS/MS spectra of noncovalent complexes of **a)** glucose, **b)** mannose, **c)** galactose, and **d)** talose, with a collision energy of 0.7 eV. Reprinted from Giorgi and Speranza. Stereoselective noncovalent interactions of monosaccharides with hydrazine. *International Journal of Mass Spectrometry* 249 (2006): 112-119, Copyright 2006, with permission from Elsevier.

2.8 | Ion-mobility spectrometry

Mass spectrometry allows for the determination of the accurate mass and fragmentation profile of monosaccharides. This enables, to some extent, the differentiation of structural isomers, as described in chapter 7 (flow injection mass spectrometry). However, the highly sensitive analysis of complex isomeric mixtures requires an additional pre-separation step. Due to the narrow polarity range and similar size of monosaccharides and its derivatives, most chromatographic separation techniques do not ensure the complete separation of isomeric structures^{438,441}. Because of their identical elemental composition and similar structures, isomeric monosaccharides tend to elute closely together even during long separation times, and their almost identical fragmentation profiles further complicate the analysis⁴⁴¹. To address these limitations, ion mobility spectrometry in combination with mass spectrometry is being increasingly utilized in various

omics fields⁴⁴¹⁻⁴⁴⁶. For reviews on the application of ion mobility spectrometry in carbohydrate research, we would like to refer to earlier articles⁴⁴⁷⁻⁴⁵¹.

Ion mobility spectrometry (IMS) is a gas-phase electrophoretic technique that enables rapid structural and conformational separations. In IMS, compounds with varying sizes, shapes, and charges are separated as they travel through a gas-filled mobility cell. Under the influence of an electric field, ions collide with inert gas particles. Larger, more diffuse species experience a higher frequency of collisions, resulting in slower electrophoretic motion compared to more compact ions^{452,453}. The measured IMS drift time can then be converted into a rotationally averaged collision cross section (CCS), which describes the three-dimensional gas phase structure of ions^{452,453}.

In addition to providing an orthogonal separation dimension, IMS can effectively improve the signal-to-noise ratio by reducing chemical noise and enhancing peak capacities when coupled with other separation techniques^{452,453}. Following the emergence of the first commercially available IMS-MS instrument, the Synapt HDMS, which utilized travelling wave IMS (TWIMS), several other IMS-MS devices were introduced. These instruments employed IMS devices including drift tube IMS (DTIMS), high-resolution cyclic IMS (cIMS), trapped IMS (TIMS), high-field asymmetric waveform IMS (FAIMS), differential mobility spectrometry (DMS), differential ion mobility spectrometry (DIMS), differential mobility analyzers (DMA), and structures for lossless ion manipulations (SLIM)^{444,454-459}. FAIMS, DMS and DIMS are often discussed together, since they all operate similarly and only differ in the geometry of their electrodes. The differences between all the mentioned techniques are in the nature of the electric field used to propel the ions through the IM cell, variations in gas flow, ion packet distribution, and the ability to measure collision cross-section (CCS)⁴⁶⁰. FAIMS, DMS, and DIMS are typically placed directly behind the ion source, serving as filtering devices. By scanning, only those analytes that respond to the changing electric field and match the applied compensation voltage (CV) can pass through and exit the drift region⁴⁶⁰. For detailed information on various ion mobility devices and recent advancements, we would like to refer to earlier reviews^{443,460-463}. The flagship instruments from different vendors have each established their own niche advantages, use cases, and drawbacks, as detailed by Delafield and co-workers⁴⁶⁴.

Although ion mobility was first paired with mass spectrometry in the 1960s, its application to carbohydrate research emerged much later, in 1997, when Liu and Clemmer demonstrated that different oligosaccharide isomers can be characterized based on their drift times⁴⁶⁵. In 2005, Clowers et al, demonstrated that IMS-MS can be used for unambiguous separation of isobaric carbohydrate mixtures⁴⁶⁶. Historically, ion mobility separations preceded tandem MS. However, recent advancements allowed for fragmentation to occur both before and after mobility separation, providing valuable insights into individual monosaccharide components of oligo- and polysaccharides, thereby facilitating the use of ion mobility for carbohydrate sequencing⁴⁶⁷. In that regard, most of the applications of

IMS-MS focused on the analysis of carbohydrate oligo and polysaccharides^{450,468-480}, with only a few studies focusing on the fundamental monosaccharide building blocks. Typically, isobaric monosaccharides and small oligosaccharides were considered to lack sufficiently distinct CCS values to be separated by commercially available IMS systems. To mitigate this issue, various shift reagents have been utilized to amplify the CCS differences between isobaric sugars, making it easier to obtain specific IMS profiles. These reagents, which can bind either covalently or non-covalently increase the mass of monosaccharides up to several times. Reagents include alkali and alkaline earth metal cations, 1-phenyl-3-methyl-5-pyrazolone (PMP), 3-carboxy-5-nitrophenyl boronic acid, or cyclodextrins, among others^{473,481-484}.

In 2007, one of the first studies which utilized IMS-MS for carbohydrate monomer research showed efficient separation of metal adducts of anomeric methyl glycoside isomers and anomeric forms of reducing monosaccharides⁴⁸⁵. Nevertheless, the extent of separation was significantly affected by the nature of the drift gas and by the nature of an adducting metal ion. Depending on the spatial orientations of the available electron donor groups, different isomers coordinate the metal ions differently, resulting in different overall shapes and compactness of the complex ion. Utilizing different drift gases and metal ions, complex monosaccharide mixtures may be resolved^{485,486}. Several other studies found distinguishable drift times for many, although not all, monosaccharide isomers^{450,481,486-488}. Interestingly, a recent study by Winkler and co-workers focused on the conformational isomers of acetylated sialic acids⁴⁸⁹. However, these isomers could not be completely resolved based solely on their mobilities. More recent high-resolution IMS (HR-IMS) setups, such as the cyclic IMS instruments, advanced the possibilities of separating isomeric monosaccharide mixtures. The IMS resolution scales approximately with the square root of the separation path length. Therefore, extended distances for ion mobility measurements, such as those used in cyclic IMS, provide advantages where complex isomer mixtures need to be analysed⁴⁶⁴.

Earlier IMS-MS setups could not resolve some monosaccharide isomers (e.g., galactose and mannose), even after derivatization^{481,486}. However, using cyclic IMS (cIMS), it became possible to resolve these diastereomers. Initially, this was done with the aid of shift reagents^{483,484}, but a recent study by Ollivier and co-workers, achieved separation even with underivatized monosaccharides⁴⁹⁰ (Figure 2.16). The study successfully generated distinctive fingerprints of underivatized monosaccharides for some of the most common monomers composing plant or bacterial polysaccharides. These fingerprints could be used to identify monomer fragments produced by fragmentation of larger oligo- and polysaccharides, forming the basis for de novo sequencing of carbohydrates using IMS-MS^{490,491}. On top of that, the same platform was shown to be able to separate α/β anomers of common protected monosaccharide building blocks, paving the way for use of cyclic ion mobility MS in high throughput screening platforms for synthetic carbohydrate

reactions, along with screening for anomeric impurities⁴⁹². In metabolomic studies, the application of DMS-MS has largely focused on targeted experiments and the removal of interferences. However, Wernisch and colleagues investigated DMS-MS for untargeted cellular metabolomics. They described the chemical selectivity of the DMS-MS platform for a variety of metabolite classes, including sugar phosphates and activated sugars⁴⁹³. The general application of IMS to metabolomics has been reviewed and discussed in depth only recently⁴⁹⁴⁻⁴⁹⁸.

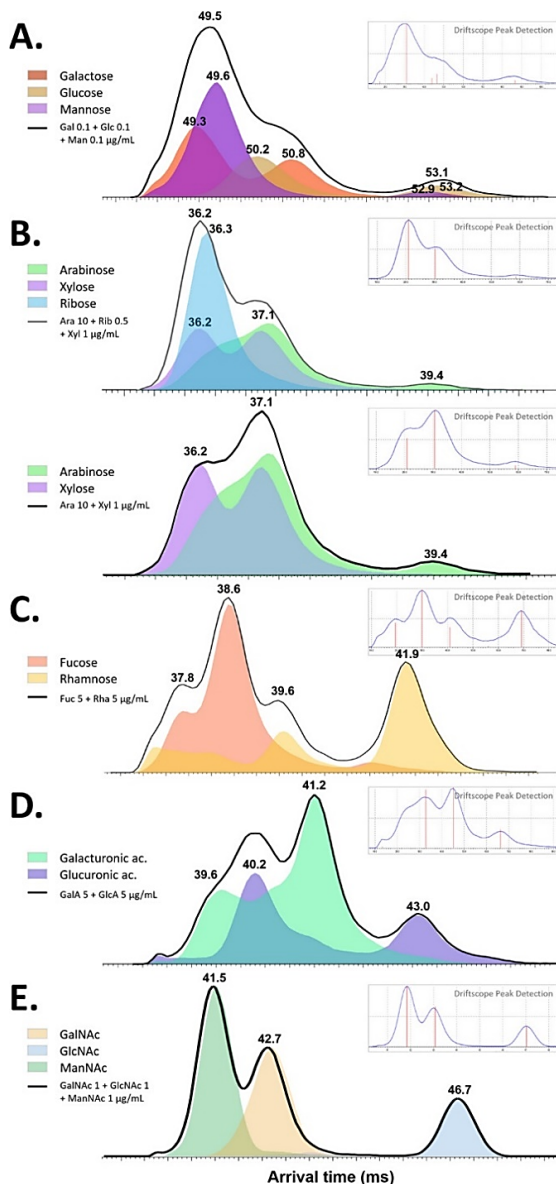


Figure 2.16. Fingerprinting of isomeric, unlabeled monosaccharides using high-resolution ion mobility spectrometry⁴⁹⁰. The graphs display the arrival time distributions for different monosaccharides: **A)** hexoses; **B)** pentoses; **C)** deoxyhexoses; **D)** hexuronic acids; **E)** N-acetyl hexosamines. The arrival times of individual sugars are shown in color, while the arrival time distribution of mixtures within each isomeric family is represented by a solid black line. The insets show the results of the (Driftscope) peak detection for the mixtures. Reprinted from Ollivier et al., Fingerprinting of underivatized monosaccharide stereoisomers using high-resolution ion mobility spectrometry and its implications for carbohydrate sequencing. *Analytical Chemistry* 95.26 (2023): 10087-10095, Copyright 2023, with permission from Elsevier.

2.9 | Infrared ion spectroscopy

Infrared ion spectroscopy (IRIS) overcomes the limitations of traditional mass spectrometry by delivering structural information in addition to mass⁴⁹⁹. This technique involves laser irradiation of mass-selected ions to obtain structural information about gas-phase analytes within the ion trap of a mass spectrometer. Unlike classical infrared spectroscopy, which measures the attenuation of the IR beam after passing through a sample, IRIS cannot obtain these conventional transmission spectra due to the extremely low density of ions in the mass spectrometer. Instead, infrared multiple photon dissociation (IRMPD) is routinely employed, as it can easily be integrated with existing MS platforms. This is commonly performed at room temperature, albeit it has also been employed at cryogenic temperatures to achieve higher-resolution infrared spectra⁴⁹⁹⁻⁵⁰¹. IRMPD detects absorption as a function of wavelength by observing the photoinduced fragmentation of irradiated ions. Resonant absorption of multiple photons induces cleavage of covalent bonds, which is dependent on the molecular structure of the ion. By plotting the photodissociation yield as a function of the wavenumber, the IRMPD spectra are obtained. Therefore, IRIS has been regarded as an additional MSn stage⁵⁰²⁻⁵⁰⁴. Proof-of-concept studies date back to the early 1980s⁵⁰⁵. For recent advancements and a detailed technical description of IRIS, we refer to recent articles and reviews^{499,500,503,506-516}.

Since stereoisomers may exhibit different absorption patterns across a range of wavelengths, IRIS can differentiate subtle structural differences, making it highly useful for studying monosaccharides and their derivatives⁵⁰². In addition, IR spectra of small molecules can be reliably predicted using computational approaches based on modern quantum-chemical calculations. This allows for reference-free identification^{503,517}. The differentiation of isomers and the reference free nature of this method is especially an exciting prospect for the characterization of monosaccharides derived from microbes, as their reference compounds are challenging to obtain. However, direct coupling of IRIS and LC separations, which is commonly necessary for analyzing complex metabolic samples, is complicated by the differing timescales of the two techniques. While an IRIS experiment usually takes minutes, LC peak widths are typically only seconds. To overcome this challenge, van Outersterp and colleagues used two switching valves and two sample loops to bypass direct online coupling. This method enabled them to record multiple IR spectra for two LC features from a single microliter injection⁵¹⁸.

After the first spectroscopic analysis and full structural assignment of a model glycoside in the gas phase by Talbot and Simons, in 2002⁵¹⁹, IRIS was employed to investigate the existence of isomeric structures. For example, the IR fingerprint region (900–1800 cm⁻¹ in most studies) was successfully used for distinguishing monosaccharides with different functional groups. The O–H stretching region (~3000–3700 cm⁻¹) of the abundant hydroxyl groups showed great potential for monosaccharide isomer differentiation^{502,516,520-526} (Figure 2.17).

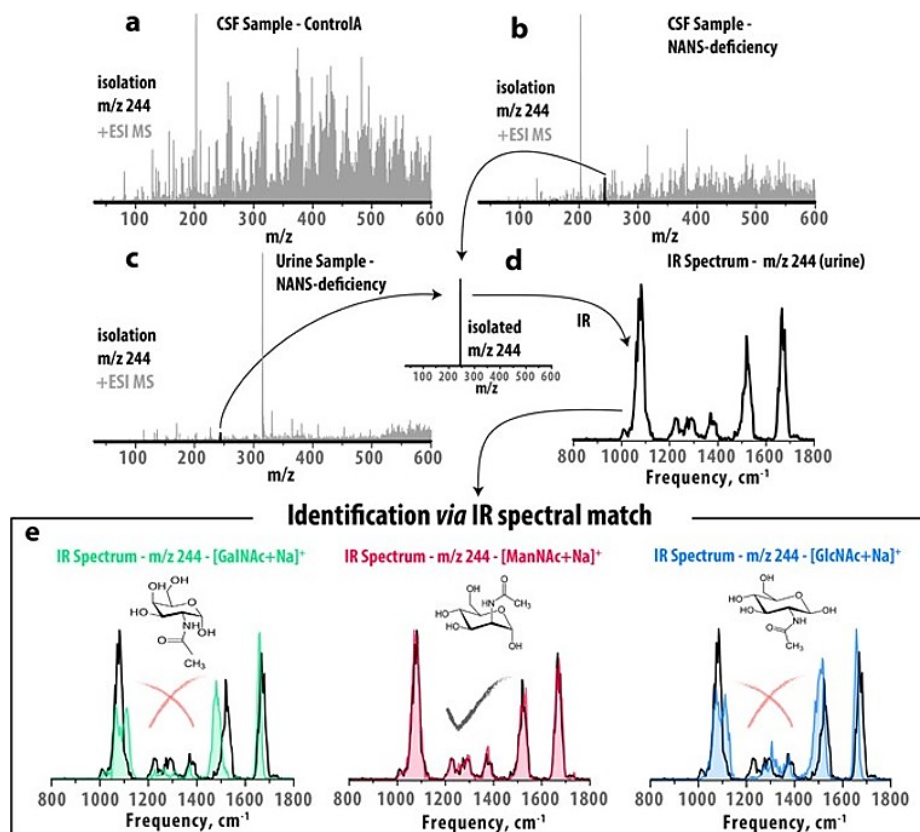


Figure 2.17. Experimental workflow for combining mass spectrometry with infrared spectroscopy for metabolomics applications⁵¹⁶. **a)** and **b)** show the isolation of a metabolite ion peak at m/z 244 (HexNAc) from a CSF control and a patient with NANS deficiency. **c)** A clear m/z 244 peak (ManNAc, $[M + Na]^+$ m/z 244), a precursor for sialic acid (Neu5Ac), identified in the patient sample. **d)** The isolation of ions at m/z 244, followed by subsequent IR analysis. **e)** Measured IR spectrum of the isolated HexNAc ion from the patient sample is matched against reference spectra of various N-acetylhexosamines, confirming that the spectrum exclusively matches ManNAc. Reprinted with permission from Martens et al, J., Berden, G., van Outersterp, R.E. et al. Molecular identification in metabolomics using infrared ion spectroscopy. *Sci Rep* 7, 3363 (2017).

It was shown that also the formation of adducts reveal unambiguous structural information. For example, a study of methyl galactosides concluded that ammonium adducts are diagnostic of the ring size (i.e. pyranose/furanose), while lithium adducts can differentiate stereochemistry (such as α and β anomers)⁵⁰². Identification of glucosamine epimers was achieved even without any sample preparation, from complex mixtures like urine and cerebrospinal fluid⁵¹⁶. Anomeric configuration and epimer discrimination was also demonstrated for protonated hexosamines⁵²⁷. Unlike the O–H region, the long-wavelength region of the IR spectrum (~ 300 – 1000 cm^{-1}), which includes more delocalized vibrations, has largely been neglected in saccharide analysis, but is now gaining momentum. In this longer wavelength IR range, spectra of many cation-saccharide complexes

have shown distinct and highly characteristic vibrational features granting their straightforward identification⁵²⁵. The enhanced IRIS spectral resolution when cooling the ions to cryogenic temperatures was used to study conformational heterogeneity of protonated glucosamine⁵²⁸⁻⁵³⁰. In addition, the application of infrared ion spectroscopy enabled the elucidation of the glycosylation mechanism and the structure of short-lived glycosyl cations^{506,531 532,533}. Finally, Hernandez and coworkers demonstrated the combination of ion mobility separation with infrared ion spectroscopy⁵³⁴. This allowed infrared characterization of both mobility- and mass-selected ions, facilitating carbohydrate isomer discrimination.

2.10 | Choice of mass analyzer guided by the analytical goal

A wide variety of mass spectrometers exists, with the mass analyzer being one of the main components, responsible for determining the mass-to-charge ratio. A distinction is made between low- and high-resolution mass spectrometry (LRMS and HRMS, respectively). In order to qualify as HRMS, a resolving power of 10,000 or higher is required⁵³⁵. Among the five most commonly used mass analyzers, the quadrupole mass filters and ion traps are categorized as LRMS, while time-of-flight (TOF), Fourier-transform ion cyclotron resonance (FT-ICR), and Orbitrap analyzers are considered HRMS⁵³⁶. The mechanisms of these mass analyzers are nicely explained by Haag (2016)⁵³⁷. Recent developments in HRMS are reviewed by Li et al. (2021)⁵³⁸.

When selecting an appropriate mass analyzer, several key factors should be considered, with the primary one being the analytical goal. HRMS is particularly well suited for detailed structural characterization and for the analysis of complex mixtures. Thanks to their high resolving power and mass accuracy, HRMS instruments can distinguish molecules with nearly identical masses. In addition, determining the exact mass enables the elucidation of the molecular formula⁵³⁵. Although ideally the choice of mass spectrometer and analyzer solely depends on the analytical goal, practical considerations such as budget and laboratory space also play critical roles. HRMS is associated with higher costs compared to LRMS. Between HRMS systems, variation in performance and expenses also exist. For example, when HRMS is required, FT-ICR is often preferred for its superior resolving power and mass accuracy, especially in the low molecular weight range. However, its reliance on a strong magnetic field necessitates a large superconducting magnet, posing two significant challenges: the need for substantial lab space and high operating costs due to the use of liquid helium as a coolant⁵³⁷. Consequently, while FT-ICR offers the highest resolving power, the Orbitrap represents a practical alternative, providing high resolving power with comparatively lower operating costs⁵³⁷.

As mentioned before, tandem mass spectrometry (MS/MS or MS²) is essential for obtaining fragmentation profiles, which are valuable for tasks such as the structural characterization of saccharides. Multistage MS or MSⁿ in time can be achieved by

mass spectrometers based on ion trapping⁵³⁹. MSⁿ, using a linear ion trap, has been demonstrated advantageous in the structural analysis of oligosaccharides⁵⁴⁰.

For routine monitoring or quantification in simpler matrices, LRMS instruments are often sufficient. As an advantage, both the quadrupole and ion trap are compact in size and less expensive compared to HRMS systems⁵³⁷. A common low resolution tandem MS instrument includes the triple quadrupole, in which three quadrupoles (Q1, Q2, and Q3, respectively) are connected in sequence, allowing various scan modes. Q1 and Q3 function as mass filter, while Q2 functions as collision cell. These instruments, when operated in multiple reaction monitoring (MRM) mode, are popular for the quantifications of compounds, including monosaccharides^{105,537}.

In principle, all of the above discussed MS-based methods and analyzers could also allow for the quantification of monosaccharides, provided that some additional requirements are met. Absolute quantification can be achieved through external or internal calibration^{106,541,542}. Both approaches depend on pure standards to create calibration curves with high linearity (R^2 close to 1) within a specified concentration range. Internal calibration is typically more accurate due to the use of internal standards (IS), which are often isotopically labelled variants of the target analytes, which share similar chemical and physical properties. Spiking consistent amounts of IS into every sample, including calibration standards, accounts for differences in ionization efficiency, matrix effects, injection volume, and sample handling⁵⁴³. When no internal standards are available, external calibration could be performed⁵⁴¹, although it requires additional experiments to evaluate potential matrix effects^{104,196,543}. Wang et al. (2022) introduced a “paired derivatization” technique that involves labelling samples with O-(4-methoxybenzyl)hydroxylamine hydrochloride (4-MOBHA·HCl) and reference standards with its deuterated form (d_3 -4-MOBHA·HCl)⁵⁴⁴. This method allows the use of an internal standard even in the absence of commercially available isotopically labelled sugars. However, reference standards remain a prerequisite, and the derivatization step limits correction for sample handling effects, as spiking occurs after labelling. The availability of pure standards represents a major limitation for absolute quantification. While common monosaccharides such as glucose and galactose are readily available, rarer ones like 4-epi-legionaminic acid or acinetaminic acid, are not.

Although relative quantification does not require pure standards, the obtained information is limited to an analyte increase or decrease when comparing multiple samples. Moreover, as Millette et al. (2023) point out, such data could be misleading when expressed as a fraction of the total polysaccharide or sample amount, since a decrease in one monosaccharide will result in a proportional increase in others and vice versa⁵⁴⁵. In glycomics, relative quantification has been achieved using isotopic or isobaric labelling. For a detailed overview of these techniques and recent advancements, readers are referred to Wang et al. (2024)⁵⁴⁶.

Another key requirement for quantification is the baseline separation of isomers¹⁰⁴. Various approaches have been explored to address this challenge as discussed in this review, but complete separation of all isomers remains difficult. The limited availability of reference standards for uncommon monosaccharides and their derivatives further complicates the process. Consequently, most methods rely on a restricted set of monosaccharide reference standards, and comprehensive isomer separation has been investigated in only a few studies^{547,548}.

Targeted quantification typically employs multiple reaction monitoring (MRM) or parallel reaction monitoring (PRM) modes. MRM targets specific transitions to ensure high specificity¹⁰⁴, while PRM, a targeted MS/MS approach, captures all fragment ions of a precursor, reducing interference and improving quantification accuracy. High-resolution mass spectrometry (HRMS) further enhances accuracy by providing superior mass precision, enabling the discrimination of closely related monosaccharides, and increasing confidence in results¹⁰⁶. These advances collectively support robust quantification of monosaccharides and nucleotide sugars in complex biological samples.

2.11 | Databases and bioinformatic approaches

The unambiguous interpretation of chromatographic and mass spectrometric data commonly requires data from reference molecules. These enable identification through comparison of elution/migration times, fragmentation profiles or collision cross section values. However, especially for microbes, reference molecules for many monosaccharides and its derivatives are unavailable. Furthermore, sufficient sample quantities for NMR experiments are not always achievable, resulting in unassigned stereochemistry in many cases. Nevertheless, mass spectrometric experiments and additional spectroscopic and chemical analyses can help to classify these new compounds based on chain length, presence of an aldose/ketose, acidic/basic nature, oxidation state, and type of nucleotide activation. Comparative spectroscopic data for a variety of sugars and their derivatives can also be obtained from numerous metabolomics databases. This provides information on sugar conversion products, especially sugar phosphates, from major metabolic pathways. For example, a comprehensive summary of general metabolomics databases, repositories, and tools supporting the annotation of mass spectrometric data and related biological pathways has been made available via the Metabolomics Society website⁵⁴⁹. Comprehensive metabolomics databases supporting mass spectrometric identification include (among others) HMDB⁵⁵⁰, MiMeDB⁵⁵¹, GMD⁵⁵², MoNA⁵⁵³ and mzCloud⁵⁵⁴. MiMeDB (Microbial Metabolites Database) is dedicated to storing information about metabolites of the human microbiome. Another general compound information resource is the NIST Chemistry WebBook⁵⁵⁵, which is an extensive chemical and physical property database particularly useful for identifying compounds based on electron ionization (EI) mass spectrometric data⁵⁵⁶. The Natural Products Atlas is another comprehensive resource containing information about more than 25,000

compounds (2024) found in microbes. Next to compound names, it provides structural data, source organisms, and synthesis information⁵⁵⁷.

Li et al. have reviewed databases dedicated to glycans and monosaccharides, along with related bioinformatic tools⁵⁵⁸. For example, UniCarb-DB is a LC-MS/MS library for N- and O-linked glycans including a data submission pipeline⁵⁵⁹. Another database dedicated to glycans is the Glycan Mass Spectral DataBase (GMDB)⁵⁶⁰. The GMDB stores MSⁿ spectra of glycans released from proteins and lipids, providing structural information. Additionally, the Carbohydrate Structure Database (CSDB) is a repository of carbohydrate structures that includes detailed information about monosaccharide building blocks and derivatives from all domains of life, including bacteria and archaea⁵⁶¹. The CSDB obtains data from scientific publications through automatic and manual curation. Furthermore, for ion mobility spectrometry, the METLIN-CCS database contains collision cross-section values derived from over 27,000 molecular standards⁵⁶². For infrared ion spectroscopy, Houthuijs and colleagues developed an *in silico* library of vibrational spectra. They created an automated workflow to generate IR spectra of molecular ions, resulting in over 75,000 calculated vibrational spectra for 4,640 metabolites from the HMDB⁵¹⁷. Moreover, the authors developed a scoring algorithm to identify molecular structures that closely match user-supplied experimental infrared ion spectra.

After identification, the biological interpretation requires integrating the data into metabolic networks. A useful resource to explore possible reaction pathways is the KEGG PATHWAY database which maps molecular interaction and reaction networks, including metabolic and regulatory pathways⁹. Additionally, KEGG GLYCAN collects experimentally determined glycan structures, including their biosynthesis pathways⁵⁶³. MetaCyc is another comprehensive resource of metabolic pathways and enzymes from all domains of life⁵⁶⁴. These databases are widely employed to facilitate pathway-based research. The CAZy database has been established to enable a categorization, for studying their structure, function, and evolutionary relationships⁵⁶⁵. Nevertheless, the pathway and enzyme database are still incomplete in terms of the strains represented, as many microbial genomes and their associated pathways have yet to be fully mapped and included⁵⁶⁶. This is particularly evident with the numerous enzymes involved in the production, conversion, and further modification of monosaccharides. Such limitations can bias the interpretation of biosynthetic routes and hinder understanding of metabolic capabilities. A more recent tool, MetaPathPredict, aims to address incomplete metabolic annotations by using deep neural networks, to predict the presence or absence of KEGG modules based on annotated features in the genome⁵⁶⁷.

The large-scale interpretation of metabolomics data also requires sophisticated bioinformatics tools, many of which have been developed in recent years. Specifically, these tools support the processing of mass spectrometric data, identification of metabolites, statistical analyses, and the correlation of

metabolites with metabolic pathways. For a comprehensive discussion of bioinformatics tools dedicated to metabolomics studies, we refer to recently published reviews⁵⁶⁸⁻⁵⁷⁴. In untargeted experiments thousands of features and molecular compounds are obtained, of which many may not provide library matches. Interestingly, Giera and colleagues reported that in source fragmentation is responsible for many of the ESI mass peaks instead of native molecules. The authors investigated ESI in source fragmentation across the METLIN MS/MS database of molecular standards^{575,576}. The authors concluded that a substantial fraction of the 'dark metabolome' may derive from in source fragmentation⁵⁷⁵.

Over the past two decades, collision induced dissociation (CID) has proven to be a valuable tool for obtaining structural insights into carbohydrates. However, the CID profiles of monosaccharides alone often fail to distinguish between both different anomeric forms and stereochemical variants. The CID profiles of glycopyranosyl-glycolaldehydes ions (m/z 221), generated using a Paul trap, did allow to differentiate between all 16 possible structural variants of unmodified hexoses. These glycopyranosyl-glycolaldehydes, produced through the fragmentation of disaccharides, consist of the intact non-reducing monosaccharide linked to a glycolaldehyde group⁵⁷⁷.

Despite these advancements, such methods still rely on chemical standards to construct CID profile databases. Recent progress has introduced promising database-independent approaches. In 2018, Huynh et al. investigated the CID behavior of sodiated glucose, galactose, and mannose to identify their anomeric configurations (α/β). Using a low-pressure linear ion trap combined with computational modelling, they demonstrated that the ratio of dehydration to cross-ring dissociation serves as a reliable indicator of anomeric configuration. Specifically, cis-anomers favored dehydration, while trans-anomers primarily underwent cross-ring cleavage. This rule-based approach has potential for broader application to other sugars, although it requires substantial computational effort⁵⁷⁸. Building on these insights, a logically derived sequence (LODES) tandem mass spectrometry (MS^n) method was developed for the structural determination of underivatized oligosaccharides. This technique enables the assignment of both anomeric configuration, linkages as well as branch locations. A key advantage is that it eliminates the need for a complete oligosaccharide database; a disaccharide database suffices. The method was validated by commercial oligosaccharides, and recently has been shown effective for N-glycans^{540,579}. LODES/ MS^n shows potential for extension to oligosaccharides containing other monosaccharides, provided their dissociation mechanisms are elucidated and the corresponding disaccharide databases are available⁵⁸⁰. Nevertheless, expanding this method to include all possible monosaccharides may introduce considerable complexity.

Alternatively, in silico-generated MS2 spectra may be used to infer molecular structures, which alone, however, are often insufficient to infer the true chemical structure⁵⁸¹. This is particularly true for compounds like monosaccharides which undergo extensive levels of molecular diversification. Kleikamp et al. developed a

hybrid approach for nonulosonic acids, that involved fully untargeted screening of cell lysates using a data-independent fragmentation, combined with a carbohydrate chemistry-guided bioinformatics pipeline⁶⁴. The pipeline identified compound peaks as ulosonic acids using an established fragmentation tree for labelled alpha-keto acids, and a theoretical sugar composition and modification database (Figure 2.18).

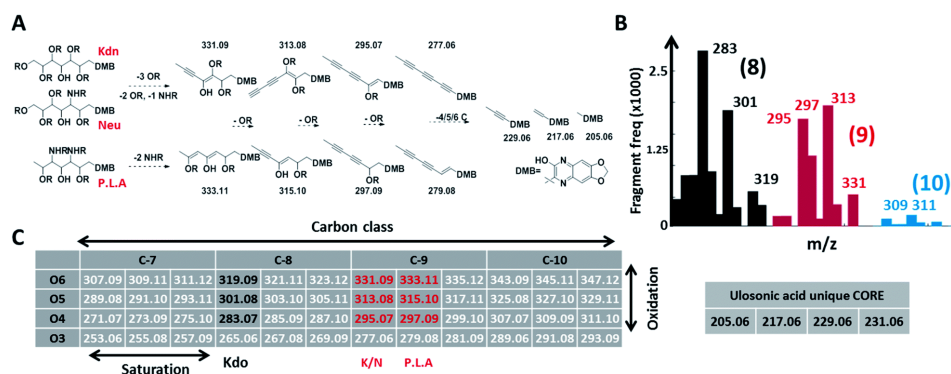


Figure 2.18. Kleikamp et al.¹³⁸ characterized the conserved fragment ion space for various nonulosonic acids (NuLOs), taking into account different degrees of oxidation, saturation, and carbon chain lengths¹³⁸. **A)** The universal fragmentation routes for Kdn, Neu and Pse/Leg/Aci. **B)** The frequency of ulosonic acid related compounds observed in a screening study, analyzing a large spectrum of microbial samples. **C)** The alpha-keto acid specific fragment ions such as the quinoxaline-based ulosonic acid core fragments and carbon chain length markers. Reprinted with permission from Kleikamp et al. Tackling the chemical diversity of microbial nonulosonic acids—a universal large-scale survey approach. Chemical Science 11.11 (2020): 3074–3080.

Recent advances in artificial intelligence (AI), particularly in machine learning (ML), offer promising opportunities to overcome the limitations of existing spectral libraries and reference databases discussed above. The remarkable success of AI in glycan analysis, demonstrated by tools like CandyCrunch which enable accurate prediction and classification of glycan structure from tandem mass spectra⁵⁸², suggests similar potential for characterization of monosaccharides and sugar metabolites. Machine learning algorithms now support various aspects of data pre-processing, including automated peak detection, spectral denoising, and retention time alignment, among others⁵⁸³. Moreover, neural networks, trained on existing spectral libraries, could help characterize structural features of atypical sugar metabolites by predicting fragmentation patterns when reference data is unavailable⁵⁸⁴. These models could help recognize subtle spectral features that distinguish stereoisomeric monosaccharides, a task that traditionally requires extensive manual interpretation. Furthermore, deep learning has shown considerable promise in predicting liquid chromatography retention times of sugar metabolites, improving nontargeted identification. This information is particularly valuable for isobaric molecules, as it can be used to filter false positive hits and

increase identification accuracy⁵⁸⁵. In the realm of metabolic pathway elucidation, graph neural networks have emerged as powerful tools for predicting biosynthetic routes of novel monosaccharide derivatives based on known enzymatic functions and genomic data. For example, MetaPathPredict can reconstruct complete metabolic pathways in microbes, potentially aiding in the exploration of the functions of uncharacterized enzymes involved in monosaccharide modifications⁵⁶⁷. Additionally, machine learning has demonstrated the ability to utilize variation in monosaccharide composition among different tissue samples to accurately classify their tissue origins⁵⁸⁶. Nevertheless, success of AI approaches depends not only on the quality and quantity of training data available but also on the careful validation of predictions against experimental results. The limited availability of validated training datasets for rare monosaccharides presents a substantial obstacle, particularly for microbial systems where novel derivatives are frequently encountered.

2.12 | Summary & outlook

Carbohydrates not only constitute a source of energy but also mediate a wide range of functions essential to cellular processes across all domains of life. Furthermore, notably in microbes, highly modified and uncommon monosaccharides are prevalent, often playing crucial roles in determining biological activity. Despite many fundamental questions about the utilization and development of individual glycomolecules, numerous medical and industrial applications have emerged for these carbohydrates. For example, the unique metabolic pathways of microbial monosaccharide derivatives make them promising and highly specific drug targets, which are urgently needed as alternatives in the face of rising antimicrobial resistance. Furthermore, carbohydrate-active enzymes hold significant potential for biotechnological applications, including the production of carbohydrate molecules that are challenging to generate via chemical synthesis.

While the protein sequences can be implied from the genetic code alone, the sequence and isomeric composition of glycomolecules cannot, highlighting the importance of sensitive and selective methods. The mass spectrometric analysis of monosaccharide derivatives is complicated by the presence of stereoisomers and extensive molecular diversification. Therefore, while sensitive detection in complex matrices can be achieved, the primary challenge in analyzing monosaccharide building blocks is the identification of new derivatives and the discrimination of isomeric structures. This necessitates the application of high-performance separation techniques alongside mass spectrometric detection for most applications.

Among the diverse analytical approaches discussed, several methods have gained widespread adoption in monosaccharide analysis. Reverse-phase liquid chromatography coupled with mass spectrometry employing a derivatization step remains one of the most widely used approaches for routine analysis. Derivatization agents such as PMP increase ionization efficiency of sugars, offering both reliable

separation and sensitive detection of common monosaccharides, including stereoisomers. Retention of highly polar sugars could be further enhanced by the addition of an ion-pairing reagent to the mobile phase. Hydrophilic interaction liquid chromatography (HILIC) has also gained prominence due to its natural affinity for polar molecules like monosaccharides and their activated sugars. For the analysis of nucleotide-activated sugars, porous graphitic carbon chromatography coupled to mass spectrometry has become the method of choice, as it provides excellent separation of isomers while avoiding the contamination issues associated with ion-pairing reagents. More recently, mixed-mode chromatography combining weak anion-exchange and reverse-phase separation has emerged as a promising alternative. In cases where high-throughput screening is required, such as in biotechnology applications, negative mode MALDI-MS has proven effective for rapid profiling of sugar nucleotides and phosphorylated intermediates. While traditionally more suited for larger molecules, recent advances in matrix development and high-resolution mass analyzers have improved its application to smaller sugar molecules. Meanwhile, capillary electrophoresis-mass spectrometry has established itself as valuable for analyzing charged sugar species, particularly sugar phosphates and nucleotides, offering high separation efficiency with minimal sample requirements and short analysis times. Gas chromatography-mass spectrometry, historically significant technique for sugar analysis, remains crucial for detailed stereochemical analysis of monosaccharides and their derivatives. Nevertheless, it requires extensive sample preparation as it is restricted to volatile and thermally stable compounds. For metabolic studies and pathway elucidation, approaches combining stable isotope incorporation and chemical labelling with advanced separation techniques have proven particularly powerful. These methods, especially when paired with high-resolution mass spectrometry, enable tracking of novel biosynthetic routes and verification of predicted pathways, proving indispensable when traditional genomic predictions may be insufficient. It is worth noting that analytical approaches discussed often complement each other to provide comprehensive characterization of sugar compounds. For instance, high-throughput capabilities of MALDI-MS can be used for initial screening of samples, followed by more detailed analysis using chromatographic methods for structural confirmation. In addition, MALDI imaging can be used for the determination of spatial resolution in tissues and biofilms.

Looking toward emerging technologies, ion mobility spectrometry (IMS) and infrared ion spectroscopy show exceptional promise for resolving previously inseparable monosaccharide isomers without requiring derivatization, complementing cases where standard separation techniques fall short. The development of high-resolution IMS and its recent applications to the separation of isomeric monosaccharides have introduced a new dimension to the analysis of complex samples. Notably, platforms such as cyclic IMS enable the investigation of isomeric monosaccharides in glycoconjugates after the initial fragmentation of glycomolecules. Similarly, the recent integration of infrared ion spectroscopy with

mass spectrometry offers a promising approach for the stereoselective analysis of carbohydrate molecules, including monosaccharides. An advantage of this methodology is that the infrared spectra of small molecules can be predicted using computational methods, making it independent of the availability of reference monosaccharides. Its ability to provide detailed structural information through computational prediction of vibrational spectra offers unique advantages for identification of unknown compounds.

The limited availability of reference standards for less common monosaccharides and their derivatives poses a significant challenge in the development and evaluation of novel analytical methods. As a result, most methods have been calibrated using only a limited set of monosaccharide reference standards, and the comprehensive separation of all possible isomers of a given monosaccharide has been explored in only a few cases. Additionally, the highly polar nature of activated monosaccharides—universal substrates for glycomolecule biosynthesis—further complicates the development of sensitive and selective analytical techniques. Recent developments include alternative approaches to traditional reverse-phase ion-pairing methods, such as porous graphitic carbon chromatography and mixed-mode separation coupled with tandem mass spectrometry. In this context, stable isotope incorporation and chemical labelling proved useful in investigating metabolic and biosynthetic pathways.

Artificial intelligence is emerging as a promising player in addressing current analytical challenges. The integration of AI approaches with established analytical methods promises to accelerate the discovery and characterization of novel monosaccharide building blocks, particularly in understudied organisms. Machine learning approaches show particular potential for predicting chromatographic behavior of modified monosaccharides, identifying novel derivatives from complex mass spectra, and elucidating biosynthetic pathways. However, the success of these AI approaches heavily depends on the quality and quantity of available training data, which remains limited for rare and modified monosaccharides, particularly from microbial sources. While the past decades have seen the establishment of routine methods for the qualitative and quantitative analysis of common monosaccharides, the stereoselective resolution and fully untargeted identification of novel monosaccharide components remain ongoing challenges. The continued advancement of high-resolution mass spectrometers, coupled with sophisticated separation techniques and AI tools, promises to further expand our understanding of monosaccharide diversity and function across all domains of life.

Abbreviations

2AA	2-aminoacridine
3-NPH	3-nitrophenylhydrazine
4-OBHA·HCl	O-(4-methoxybenzyl)hydroxylamine hydrochloride
9AA	9-aminoacridine
Aci	acinetaminic acid
AMAC	2-aminoacridone

Api	apiose
APTS	9-aminopyrene-1,4,6-trisulfonic acid
Ara	arabinose
Araf	arabinofuranose
Arap	arabinopyranose
ATP	adenosine triphosphate
Bac2Ac4Ac	2,4-diacetyl-bacillosamine
BEH	Ethylene Bridged Hybrid
BGE	background electrolyte
BTME-H-APS	bistrimethoxysilylethane hydrolysis aminopropyltrimethoxysilane
CAD	charged aerosol detector
CCS	collision cross-section
CE	capillary electrophoresis
cIMS	cyclic ion mobility spectrometry
CMP	cytidine monophosphate
CSDB	Carbohydrate Structure Database
CV	compensation voltage
CZE	capillary zone electrophoresis
DHB	2, 5-dihydroxybenzoic acid
DIMS	differential ion mobility spectroscopy
DMA	differential mobility analysers
DMB	1,2-diamino-4,5-methylenedioxybenzene
DMBA	4,5-dimethylbenzene-1,2-diamine
DMS	differential mobility spectrometry
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DTIMS	drift tube ion mobility spectrometry
EI	electron ionisation
EOF	electroosmotic flow
ESI	electrospray ionization
FAIMS	high-field asymmetric waveform ion mobility spectrometry
Fru6P	fructose-6-phosphate
Fuc2NAc	2-acetamido-2,6-dideoxy-D-galactose
Fus	fusaminic acid
Gal	galactose
GalA	galacturonic acid
GalN	galactosamine
GalNAc	N-acetyl-galactosamine
GC	gas chromatography
GDP	guanosine diphosphate
Glc	glucose
Glc1P	glucose-1-phosphate
Glc6P	glucose-6-phosphate
GlcA	glucuronic acid
GlcN	glucosamine
GlcNAc	N-acetyl-glucosamine
GMD	golm metabolome database
GMDB	glycan Mass Spectral DataBase
GuIA	guluronic acid
HDMS	high definition mass spectrometer
HEK	human embryonic kidney
HILIC	hydrophilic interaction liquid chromatography
HMDB	human Metabolome Database

HPAE	high performance anion exchange
HPAEC	high performance anion exchange chromatography
HPLC	high-performance liquid chromatography
HR	high-resolution
IdoA	iduronic acid
IEC	ion exchange chromatography
IEX	ion exchange
IMS	ion mobility spectrometry
IP	ion pair
IR	infrared
IRIS	infrared ion spectroscopy
IRMPD	infrared multiple photon dissociation
IT	ion trap
KDG	2-keto-3-deoxygluconate
KDN	keto-deoxy-neuraminic acid
KDO	2-keto-3-Deoxy-D-Mannooctanoic Acid
KEGG	Kyoto Encyclopedia of Gene and Genome database
LC	liquid chromatography
Leg	legionaminic acid
Leg5Ac7AcAla	5-N-acetyl-7-N-acetyl-d-alanyl-legionaminic acid
LIF	laser-induced fluorescence
MALDI	matrix-assisted laser desorption/ionization
Man6P	mannose-6-phosphate
ManA	mannuronic acid
ManNAc	N-acetyl-mannosamine
MMC	mixed mode chromatography
MRM	multiple reaction monitoring
MS	mass spectrometry
MSI	imaging mass spectrometry
MSTFA	N-methyl-N-trimethylsilyl-trifluoroacetamide
NDP	nucleotide diphosphate
Neu	neuraminic acid
Neu5Ac	5-N-acetyl-neuraminic acid
Neu5Gc	5-N-glycolyl-neuraminic acid
NeuNAc	N-acetyl-neuraminic acid
NIST	National Institute of Standards and Technology
NMR	nuclear magnetic resonance
NT	nucleotide
NuLO	nonulosonic acids
PAD	pulsed amperometric detection
PBA	phenylboronic acid
PFBO	pentafluorobenzoyloxime
PGC	porous graphitic carbon
PMP	1-phenyl-3-methyl-5-pyrazolone
PPMP	1-(4-isopropyl) phenyl-3-methyl-5-pyrazolone
Pse	pseudaminic acid
Rha	rhamnose
RNA	ribonucleic acid
RP	reverse phase
SAX	strong anion exchange
SFC	supercritical fluid chromatography
SIA	sialic acid
SLIM	structures for lossless ion manipulations

SNFG	Symbol Nomenclature For Glycans
SPE	solid phase extraction
TFA	trifluoroacetic acid
TIMS	trapped ion mobility spectrometry
TMS	trimethylsilyl
TMT	tandem mass tag
TOF	time of flight
TWIMS	traveling wave ion mobility spectrometry
UDP	uridine diphosphate
UHPLC	Ultra-High-Performance Liquid Chromatography
UTP	uridine triphosphate
UV	ultraviolet
Xyl	xylose

References

- 1 Sanger, F. Determination of nucleotide sequences in DNA. *Science* **214**, 1205-1210 (1981).
- 2 Chaudhuri, B. N. & Yeates, T. O. A computational method to predict genetically encoded rare amino acids in proteins. *Genome biology* **6**, 1-15 (2005).
- 3 Alberts, B. *et al.* *Essential cell biology*. (Garland Science, 2015).
- 4 Herget, S. *et al.* Statistical analysis of the Bacterial Carbohydrate Structure Data Base (BCSDB): characteristics and diversity of bacterial carbohydrates in comparison with mammalian glycans. *BMC structural biology* **8**, 1-20 (2008).
- 5 Cummings, R. D. A periodic table of monosaccharides. *Glycobiology* **34**, cwad088 (2024).
- 6 Harvey, D. J. Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates. *Mass spectrometry reviews* **18**, 349-450 (1999).
- 7 Pabst, M. *et al.* A general approach to explore prokaryotic protein glycosylation reveals the unique surface layer modulation of an anammox bacterium. *The ISME Journal* **16**, 346-357 (2022).
- 8 Imperiali, B. Bacterial carbohydrate diversity—A brave new world. *Current opinion in chemical biology* **53**, 1-8 (2019).
- 9 Kanehisa, M. in *'In silico' simulation of biological processes: Novartis Foundation Symposium 247*. 91-103 (Wiley Online Library).
- 10 Chandel, N. S. Glycolysis. *Cold Spring Harbor Perspectives in Biology* **13**, a040535 (2021).
- 11 Varki, A. *et al.* *Essentials of Glycobiology* [internet]. (2015).
- 12 Grabarics, M. *et al.* Mass spectrometry-based techniques to elucidate the sugar code. *Chemical Reviews* **122**, 7840-7908 (2021).
- 13 Tharanathan, R. N. Food-derived carbohydrates—structural complexity and functional diversity. *Critical reviews in biotechnology* **22**, 65-84 (2002).
- 14 Bode, L. The functional biology of human milk oligosaccharides. *Early human development* **91**, 619-622 (2015).
- 15 Flynn, R. A. *et al.* Small RNAs are modified with N-glycans and displayed on the surface of living cells. *Cell* **184**, 3109-3124. e3122 (2021).
- 16 Bülow, H. E. & Hobert, O. The molecular diversity of glycosaminoglycans shapes animal development. *Annu. Rev. Cell Dev. Biol.* **22**, 375-407 (2006).
- 17 Ricard-Blum, S. Vol. 34 275-276 (Springer, 2017).
- 18 Zappe, A., Miller, R. L., Struwe, W. B. & Pagel, K. State-of-the-art glycosaminoglycan characterization. *Mass Spectrometry Reviews* **41**, 1040-1071 (2022).
- 19 Hooper, L. V. & Gordon, J. I. Glycans as legislators of host-microbial interactions: spanning the spectrum from symbiosis to pathogenicity. *Glycobiology* **11**, 1R-10R (2001).
- 20 Lee, S. *et al.* Glycan-mediated molecular interactions in bacterial pathogenesis. *Trends in microbiology* **30**, 254-267 (2022).
- 21 Nothaft, H. & Szymanski, C. M. Protein glycosylation in bacteria: sweeter than ever. *Nature Reviews Microbiology* **8**, 765-778 (2010).
- 22 Van Gulik, W., De Laat, W., Vinke, J. & Heijnen, J. Application of metabolic flux analysis for the identification of metabolic bottlenecks in the biosynthesis of penicillin-G. *Biotechnology and Bioengineering* **68**, 602-618 (2000).
- 23 Hackett, S. R. *et al.* Systems-level analysis of mechanisms regulating yeast metabolic flux. *Science* **354**, aaf2786 (2016).
- 24 Koster, C. C., Postma, E. D., Knibbe, E., Cleij, C. & Daran-Lapujade, P. Synthetic genomics from a yeast perspective. *Frontiers in Bioengineering and Biotechnology* **10**, 869486 (2022).
- 25 Ko, Y.-S. *et al.* Tools and strategies of systems metabolic engineering for the development of microbial cell factories for chemical production. *Chemical Society Reviews* **49**, 4615-4636 (2020).
- 26 Edwards, E. *et al.* Strategies to control therapeutic antibody glycosylation during bioprocessing: Synthesis and separation. *Biotechnology and Bioengineering* **119**, 1343-1358 (2022).
- 27 Gilbert-Barness, E. F., Philip. Approach to diagnosis of metabolic diseases *Translational Science of Rare Diseases* **1**, 3-22 (2016). <https://doi.org/10.3233/TRD-160001>
- 28 Cantor, J. R. & Sabatini, D. M. Cancer cell metabolism: one hallmark, many faces. *Cancer discovery* **2**, 881-898 (2012).
- 29 Otto, A. & Simpson, M. J. Analysis of soil organic matter biomarkers by sequential chemical degradation and gas chromatography – mass spectrometry. *Journal of Separation Science* **30**, 272-282 (2007). <https://doi.org/https://doi.org/10.1002/jssc.200600243>
- 30 Samuel, G. & Reeves, P. Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydrate research* **338**, 2503-2519 (2003).
- 31 Seifert, G. J. Nucleotide sugar interconversions and cell wall biosynthesis: how to bring the inside to the outside. *Current opinion in plant biology* **7**, 277-284 (2004).
- 32 Mikkola, S. Nucleotide sugars in chemistry and biology. *Molecules* **25**, 5755 (2020).
- 33 Srivastava, J., Sunthar, P. & Balaji, P. V. The glycan alphabet is not universal: a hypothesis. *Microbial Genomics* **6**, e000452 (2020).
- 34 Moncrieffe, M. C. *et al.* Structure of the glycosyltransferase EryCIII in complex with its activating P450 homologue EryCII. *Journal of molecular biology* **415**, 92-101 (2012).
- 35 Parıldı, E., Kola, O., Özcan, B. D., Akkaya, M. R. & Dikkaya, E. Recombinant D-tagatose 3-epimerase production and converting fructose into allulose. *Journal of Food Processing and Preservation* **46**, e15508 (2022).
- 36 Paliya, B. S. *et al.* Bacterial glycobiotechnology: A biosynthetic route for the production of biopharmaceutical glycans. *Biotechnology Advances*, 108180 (2023).
- 37 Harvey, D. J. Analysis of carbohydrates and glycoconjugates by matrix-assisted laser desorption/ionization mass spectrometry: An update for 2019–2020. *Mass spectrometry reviews* **42**, 1984-2206 (2023).
- 38 Lee, C. & Ni, C.-K. Soft matrix-assisted laser desorption/ionization for labile Glycoconjugates. *Journal of The American Society for Mass Spectrometry* **30**, 1455-1463 (2019).
- 39 Girgis, M. *et al.* Analysis of N- and O-linked site-specific glycosylation by ion mobility mass spectrometry: State of the art and future directions. *Proteomics*, 2300281 (2024).

- 40 Trbojević-Akmačić, I. *et al.* High-throughput glycomic methods. *Chemical reviews* **122**, 15865-15913 (2022).
- 41 de Haan, N., Yang, S., Cipollo, J. & Wührer, M. Glycomics studies using sialic acid derivatization and mass spectrometry. *Nature Reviews Chemistry* **4**, 229-242 (2020).
- 42 Gerwig, G. J. *The art of carbohydrate analysis*. (Springer, 2021).
- 43 Kailemia, M. J., Ruhaak, L. R., Lebrilla, C. B. & Amster, I. J. Oligosaccharide analysis by mass spectrometry: a review of recent developments. *Analytical chemistry* **86**, 196-212 (2014).
- 44 Harvey, D. J. Derivatization of carbohydrates for analysis by chromatography; electrophoresis and mass spectrometry. *Journal of Chromatography B* **879**, 1196-1225 (2011).
- 45 Wang, J., Zhao, J., Nie, S., Xie, M. & Li, S. MALDI mass spectrometry in food carbohydrates analysis: A review of recent researches. *Food Chemistry* **399**, 133968 (2023).
- 46 Wang, J., Zhao, J., Nie, S., Xie, M. & Li, S. Mass spectrometry for structural elucidation and sequencing of carbohydrates. *TrAC Trends in Analytical Chemistry* **144**, 116436 (2021).
- 47 Pabst, M. & Altmann, F. Glycan analysis by modern instrumental methods. *Proteomics* **11**, 631-643 (2011).
- 48 Sarbu, M. & Zamfir, A. D. Modern separation techniques coupled to high performance mass spectrometry for glycolipid analysis. *Electrophoresis* **39**, 1155-1170 (2018).
- 49 Hykollari, A., Paschinger, K. & Wilson, I. B. Negative-mode mass spectrometry in the analysis of invertebrate, fungal, and protist N-glycans. *Mass Spectrometry Reviews* **41**, 945-963 (2022).
- 50 Manz, C. & Pagel, K. Glycan analysis by ion mobility-mass spectrometry and gas-phase spectroscopy. *Current Opinion in Chemical Biology* **42**, 16-24 (2018).
- 51 Wang, W., Wang, W. & Lageveen-Kammeijer, G. S. CE-MS Approaches for Glyco (proteo) mic Analysis. *Capillary Electrophoresis-Mass Spectrometry for Proteomics and Metabolomics: Principles and Applications*, 335-375 (2022).
- 52 Lageveen-Kammeijer, G. S., Kuster, B., Reusch, D. & Wührer, M. High sensitivity glycomics in biomedicine. *Mass spectrometry reviews* **41**, 1014-1039 (2022).
- 53 Seeberger, P. H. Monosaccharide diversity. (2017).
- 54 Katzenellenbogen, E. *et al.* Structure of an abequose-containing O-polysaccharide from *Citrobacter freundii* O22 strain PCM 1555. *Carbohydrate research* **344**, 1724-1728 (2009).
- 55 Kowalczyk, A. *et al.* The structure of an abequose-containing O-polysaccharide isolated from *Pectobacterium aquaticum* IFB5637. *Carbohydrate Research* **522**, 108696 (2022).
- 56 K, A. I., Rajagopal, T., Tan, J. *Metabolism - Sugars. In Biochemistry Free For All* (6.1). (Biology LibreTexts, 2021).
- 57 Pabst, M. *et al.* Rhamnogalacturonan II structure shows variation in the side chains monosaccharide composition and methylation status within and across different plant species. *The Plant Journal* **76**, 61-72 (2013).
- 58 Bhagavan, N. & Ha, C. Carbohydrate Metabolism II: Gluconeogenesis, Glycogen Synthesis and Breakdown, and Alternative Pathways. *Medical biochemistry* **4** (2001).
- 59 TOUSTER, O. in *Comprehensive Biochemistry* Vol. 17 219-240 (Elsevier, 1969).
- 60 Schauer, R. & Kamerling, J. P. Exploration of the sialic acid world. *Advances in carbohydrate chemistry and biochemistry* **75**, 1-213 (2018).
- 61 Rangarajan, E. S. *et al.* Structural and functional analysis of *Campylobacter jejuni* PseG: a udp-sugar hydrolase from the pseudaminic acid biosynthetic pathway. *Journal of biological chemistry* **284**, 20989-21000 (2009).
- 62 Romero-Rivera, A., Iglesias-Fernández, J. & Osuna, S. Exploring the Conversion of ad-Sialic Acid Aldolase into al-KDO Aldolase. *European Journal of Organic Chemistry* **2018**, 2603-2608 (2018).
- 63 Lewis, A. L. *et al.* Innovations in host and microbial sialic acid biosynthesis revealed by phylogenomic prediction of nonulosonic acid structure. *Proceedings of the National Academy of Sciences* **106**, 13552-13557 (2009).
- 64 Kleikamp, H. B. *et al.* Tackling the chemical diversity of microbial nonulosonic acids—a universal large-scale survey approach. *Chemical Science* **11**, 3074-3080 (2020).
- 65 Angata, T. & Varki, A. Chemical diversity in the sialic acids and related α -keto acids: an evolutionary perspective. *Chemical reviews* **102**, 439-470 (2002).
- 66 He, X. M. & Liu, H.-w. Formation of unusual sugars: mechanistic studies and biosynthetic applications. *Annual review of biochemistry* **71**, 701-754 (2002).
- 67 Varki, A. *et al.* Symbol nomenclature for graphical representations of glycans. *Glycobiology* **25**, 1323-1324 (2015).
- 68 Shafizadeh, F. in *Advances in Carbohydrate Chemistry* Vol. 11 263-283 (Elsevier, 1956).
- 69 Kim, W. H. *et al.* Efficient and practical synthesis of L-hamamelose. *Carbohydrate research* **344**, 2317-2321 (2009).
- 70 Pičmanová, M. & Møller, B. L. Apiose: one of nature's witty games. *Glycobiology* **26**, 430-442 (2016).
- 71 O'Neill, M. A. *et al.* Rhamnogalacturonan-II, a pectic polysaccharide in the walls of growing plant cell, forms a dimer that is covalently cross-linked by a borate ester: in vitro conditions for the formation and hydrolysis of the dimer. *Journal of Biological chemistry* **271**, 22923-22930 (1996).
- 72 Cardini, C. E., Paladini, A. C., Caputto, R. & Leloir, L. F. Uridine diphosphate glucose: the coenzyme of the galactose-glucose phosphate isomerization. *Nature* **165**, 191-192 (1950).
- 73 Ashihara, H., Crozier, A. & Ludwig, I. A. *Plant nucleotide metabolism: Biosynthesis, degradation, and alkaloid formation*. (John Wiley & Sons, 2020).
- 74 Frey, P. A. The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose. *The FASEB Journal* **10**, 461-470 (1996).
- 75 Pabst, M. *et al.* Nucleotide and nucleotide sugar analysis by liquid chromatography-electrospray ionization-mass spectrometry on surface-conditioned porous graphitic carbon. *Analytical chemistry* **82**, 9782-9788 (2010).
- 76 Konishi, T. *et al.* A plant mutase that interconverts UDP-arabinofuranose and UDP-arabinopyranose. *Glycobiology* **17**, 345-354 (2007).
- 77 Zamora, C. Y., Schocker, N. S., Chang, M. M. & Imperiali, B. in *Methods in enzymology* Vol. 597 145-186 (Elsevier, 2017).
- 78 Schoenhofen, I. C., Young, N. M. & Gilbert, M. in *Methods in Enzymology* Vol. 597 187-207 (Elsevier, 2017).
- 79 Ricaldi N, J., Matthias, M. A., Vinetz, J. M. & Lewis, A. L. Expression of sialic acids and other nonulosonic acids in *Leptospira*. *BMC microbiology* **12**, 1-12 (2012).

- 80 Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64-71 (1989).
- 81 Whitehouse, C. M., Dreyer, R. N., Yamashita, M. & Fenn, J. B. Electrospray interface for liquid chromatographs and mass spectrometers. *Analytical chemistry* **57**, 675-679 (1985).
- 82 Rassi, Z. E. Reversed-phase and hydrophobic interaction chromatography of carbohydrates and glycoconjugates. *Carbohydrate Analysis by Modern Liquid Phase Separation Techniques*, 35-124 (2021). <https://doi.org/10.1016/B978-0-12-821447-3.00017-2>
- 83 Poole, C. F. Reversed-phase liquid chromatography. *Liquid Chromatography: Fundamentals and Instrumentation: Volume 1, Third Edition* **1**, 89-119 (2023). <https://doi.org/10.1016/B978-0-323-99968-7.00015-1>
- 84 Vreeker, G. C. M. & Wührer, M. Reversed-phase separation methods for glycan analysis. *Analytical and Bioanalytical Chemistry* **2016** 409:2 **409**, 359-378 (2016). <https://doi.org/10.1007/S00216-016-0073-0>
- 85 Fan, B., Li, T., Song, X., Wu, C. & Qian, C. A rapid, accurate and sensitive method for determination of monosaccharides in different varieties of *Osmanthus fragrans* Lour by pre-column derivatization with HPLC-MS/MS. *International Journal of Biological Macromolecules* **125**, 221-231 (2019). <https://doi.org/10.1016/J.IJBIOMAC.2018.12.033>
- 86 Wu, X., Jiang, W., Lu, J., Yu, Y. & Wu, B. Analysis of the monosaccharide composition of water-soluble polysaccharides from *Sargassum fusiforme* by high performance liquid chromatography/electrospray ionisation mass spectrometry. *Food Chemistry* **145**, 976-983 (2014). <https://doi.org/10.1016/J.FOODCHEM.2013.09.019>
- 87 Cai, Y. & Cole, R. B. Stabilization of anionic adducts in negative ion electrospray mass spectrometry. *Analytical chemistry* **74**, 985-991 (2002).
- 88 Hammad, L. A., Saleh, M. M., Novotny, M. V. & Mechref, Y. Multiple-reaction monitoring liquid chromatography mass spectrometry for monosaccharide compositional analysis of glycoproteins. *Journal of the American Society for Mass Spectrometry* **20**, 1224-1234 (2009).
- 89 Kohler, M. & Leary, J. A. LC/MS/MS of carbohydrates with postcolumn addition of metal chlorides using triaxial electrospray probe. *Analytical chemistry* **67**, 3501-3508 (1995).
- 90 Harvey, D. J. Fragmentation of negative ions from carbohydrates: part 1. Use of nitrate and other anionic adducts for the production of negative ion electrospray spectra from N-linked carbohydrates. *Journal of the American Society for Mass Spectrometry* **16**, 622-630 (2005).
- 91 Rogatsky, E., Jayatilake, H., Goswami, G., Tomuta, V. & Stein, D. Sensitive LC MS quantitative analysis of carbohydrates by Cs⁺ attachment. *Journal of the American Society for Mass Spectrometry* **16**, 1805-1811 (2005).
- 92 Simatupang, M. & Dietrichs, H. Detection of reducing sugars by neocuproin following ion-exchange chromatography. *Chromatographia* **11**, 89-95 (1978).
- 93 White, C. A., Kennedy, J. F. & Golding, B. T. Analysis of derivatives of carbohydrates by high-pressure liquid chromatography. *Carbohydrate Research* **76**, 1-10 (1979).
- 94 Hase, S. in *Journal of Chromatography Library* Vol. 58 555-575 (Elsevier, 1995).
- 95 Gao, X. *et al.* Progresses of derivatization techniques for analyses of carbohydrates. *Analytical letters* **36**, 1281-1310 (2003).
- 96 Alpenfels, W. F. A rapid and sensitive method for the determination of monosaccharides as their dansyl hydrazones by high-performance liquid chromatography. *Analytical Biochemistry* **114**, 153-157 (1981).
- 97 Han, J., Lin, K., Sequira, C., Yang, J. & Borchers, C. H. Quantitation of low molecular weight sugars by chemical derivatization-liquid chromatography/multiple reaction monitoring/mass spectrometry. *Electrophoresis* **37**, 1851-1860 (2016).
- 98 Hase, S., Ikenaka, T. & Matsushima, Y. A highly sensitive method for determination of molecular weights of neutral oligosaccharides by fluorescence labeling. *The Journal of Biochemistry* **90**, 1275-1279 (1981).
- 99 Zhang, P., Wang, Z., Xie, M., Nie, W. & Huang, L. Detection of carbohydrates using a pre-column derivatization reagent 1-(4-isopropyl) phenyl-3-methyl-5-pyrazolone by high-performance liquid chromatography coupled with electrospray ionization mass spectrometry. *Journal of Chromatography B* **878**, 1135-1144 (2010). <https://doi.org/10.1016/J.JCHROMB.2010.03.027>
- 100 Honda, S. *et al.* High-performance liquid chromatography of reducing carbohydrates as strongly ultraviolet-absorbing and electrochemically sensitive 1-phenyl-3-methyl-5-pyrazolone derivatives. *Analytical Biochemistry* **180**, 351-357 (1989). [https://doi.org/10.1016/0003-2697\(89\)90444-2](https://doi.org/10.1016/0003-2697(89)90444-2)
- 101 Durán-Álvarez, J. C., Rodríguez-Varela, M., Verdeja-Muñoz, E. J. & Córdova-Aguilar, M. S. Determination of the monosaccharide composition in mucilage of *Opuntia ficus indica* by HPLC-ESI-MS-validation of the sample preparation and the analytical method. *Journal of Food Measurement and Characterization* **15**, 4233-4244 (2021).
- 102 Guo, N. *et al.* Quantitative Analysis of Polysaccharide Composition in *Polyporus umbellatus* by HPLC-ESI-TOF-MS. *Molecules* **2019**, Vol. 24, Page 2526 **24**, 2526-2526 (2019). <https://doi.org/10.3390/MOLECULES24142526>
- 103 Rühmann, B., Schmid, J. & Sieber, V. Fast carbohydrate analysis via liquid chromatography coupled with ultra violet and electrospray ionization ion trap detection in 96-well format. *Journal of Chromatography A* **1350**, 44-50 (2014).
- 104 Xu, G., Amicucci, M. J., Cheng, Z., Galermo, A. G. & Lebrilla, C. B. Revisiting monosaccharide analysis-quantitation of a comprehensive set of monosaccharides using dynamic multiple reaction monitoring. *Analyst* **143**, 200-207 (2018).
- 105 Amicucci, M. J. *et al.* A rapid-throughput adaptable method for determining the monosaccharide composition of polysaccharides. *International Journal of Mass Spectrometry* **438**, 22-28 (2019).
- 106 Salas, E. *et al.* A rapid and sensitive assay to quantify amino sugars, neutral sugars and uronic acid necromass biomarkers using pre-column derivatization, ultra-high-performance liquid chromatography and high-resolution mass spectrometry. *Soil Biology and Biochemistry* **177**, 108927-108927 (2023). <https://doi.org/10.1016/J.SOILBIO.2022.108927>
- 107 Anumula, K. Quantitative determination of monosaccharides in glycoproteins by high-performance liquid chromatography with highly sensitive fluorescence detection. *Analytical biochemistry* **220**, 275-283 (1994).
- 108 Stepan, H. & Staudacher, E. Optimization of monosaccharide determination using anthranilic acid

- and 1-phenyl-3-methyl-5-pyrazolone for gastropod analysis. *Analytical biochemistry* **418**, 24-29 (2011).
- 109 Gerwig, G. J. & Gerwig, G. J. Analytical Techniques to Study Carbohydrates. *The Art of Carbohydrate Analysis*, 89-126 (2021).
 - 110 Harazono, A. *et al.* A comparative study of monosaccharide composition analysis as a carbohydrate test for biopharmaceuticals. *Biologicals* **39**, 171-180 (2011).
 - 111 Royle, L. in *Liquid Chromatography* 183-200 (Elsevier, 2017).
 - 112 Pabst, M. *et al.* Comparison of fluorescent labels for oligosaccharides and introduction of a new postlabeling purification method. *Analytical biochemistry* **384**, 263-273 (2009).
 - 113 Ruhaak, L. R. *et al.* Glycan labeling strategies and their use in identification and quantification. *Analytical and bioanalytical chemistry* **397**, 3457-3481 (2010).
 - 114 Keser, T., Pavić, T., Lauc, G. & Gornik, O. Comparison of 2-Aminobenzamide, Procainamide and Rapi Fluor-MS as Derivatizing Agents for High-Throughput HILIC-UPLC-FLR-MS N-glycan Analysis. *Frontiers in Chemistry* **6**, 324 (2018).
 - 115 Šoić, D. *et al.* In a pursuit of optimal glycan fluorescent label for negative MS mode for high-throughput N-glycan analysis. *Frontiers in chemistry* **10**, 999770 (2022).
 - 116 Dhume, S. T., Saddic, G. N. & Anumula, K. R. Monitoring glycosylation of therapeutic glycoproteins for consistency by HPLC using highly fluorescent anthranilic acid (AA) tag. *Post-translational Modifications of Proteins: Tools for Functional Proteomics*, 317-331 (2008).
 - 117 Grass, J. *et al.* Discovery and structural characterization of fucosylated oligomannosidic N-glycans in mushrooms. *Journal of Biological Chemistry* **286**, 5977-5984 (2011).
 - 118 Hamiot, A., Lemy, C., Krzewinski, F., Faille, C. & Dubois, T. Sporulation conditions influence the surface and adhesion properties of *Bacillus subtilis* spores. *Frontiers in Microbiology* **14**, 1219581 (2023).
 - 119 Wang, T., Nam, P. K.-s., Shi, H. & Ma, Y. Compositional Monosaccharide Analysis of Transgenic Corn Glycoproteins by HPLC with Fluorescence Detection and LC-MS with Sonic Spray Ionization. *Journal of chromatographic science* **45**, 200-206 (2007).
 - 120 Windwarder, M. *et al.* "Hypermethylation" of anthranilic acid-labeled sugars confers the selectivity required for liquid chromatography-mass spectrometry. *Analytical Biochemistry* **514**, 24-31 (2016).
 - 121 Zhu, H. *et al.* Acidolysis-based component mapping of glycosaminoglycans by reversed-phase high-performance liquid chromatography with off-line electrospray ionization-tandem mass spectrometry: Evidence and tags to distinguish different glycosaminoglycans. *Analytical biochemistry* **465**, 63-69 (2014).
 - 122 Rath, C. B. *et al.* Flagellin glycoproteomics of the periodontitis associated pathogen *Selenomonas sputigena* reveals previously not described O-glycans and rhamnose fragment rearrangement occurring on the glycopeptides. *Molecular & Cellular Proteomics* **17**, 721-736 (2018).
 - 123 Galermo, A. G. *et al.* Liquid chromatography-tandem mass spectrometry approach for determining glycosidic linkages. *Analytical chemistry* **90**, 13073-13080 (2018).
 - 124 Wang, S.-L. *et al.* Paired derivatization approach with H/D-labeled hydroxylamine reagents for sensitive and accurate analysis of monosaccharides by liquid chromatography tandem mass spectrometry. *Analytical Chemistry* **94**, 3590-3599 (2022).
 - 125 Gao, J. N. *et al.* Direct acetylation for full analysis of polysaccharides in edible plants and fungi using reverse phase liquid chromatography-multiple reaction monitoring mass spectrometry. *Journal of Pharmaceutical and Biomedical Analysis* **222**, 115083-115083 (2023). <https://doi.org/10.1016/J.JPBA.2022.115083>
 - 126 Aminoff, D. Methods for the quantitative estimation of N-acetylneuraminic acid and their application to hydrolysates of sialomucoids. *Biochemical Journal* **81**, 384 (1961).
 - 127 Uchida, Y., Tsukada, Y. & Sugimori, T. Distribution of neuraminidase in *Arthrobacter* and its purification by affinity chromatography. *The Journal of Biochemistry* **82**, 1425-1433 (1977).
 - 128 Kitajima, K., Varki, N. & Sato, C. Advanced technologies in sialic acid and sialoglycoconjugate analysis. *SialoGlyco Chemistry and Biology II: Tools and Techniques to Identify and Capture Sialoglycans*, 75-103 (2015).
 - 129 Hara, S., Takemori, Y., Yamaguchi, M., Nakamura, M. & Ohkura, Y. Fluorometric high-performance liquid chromatography of N-acetyl- and N-glycolylneuraminic acids and its application to their microdetermination in human and animal sera, glycoproteins, and glycolipids. *Analytical Biochemistry* **164**, 138-145 (1987). [https://doi.org/10.1016/0003-2697\(87\)90377-0](https://doi.org/10.1016/0003-2697(87)90377-0)
 - 130 Gerwig, G. J. & Gerwig, G. J. Analysis of Sialic Acids. *The Art of Carbohydrate Analysis*, 229-243 (2021).
 - 131 Spichtig, V., Michaud, J. & Austin, S. Determination of sialic acids in milks and milk-based products. *Analytical Biochemistry* **405**, 28-40 (2010).
 - 132 Zeleny, R., Kolarich, D., Strasser, R. & Altmann, F. Sialic acid concentrations in plants are in the range of inadvertent contamination. *Planta* **224**, 222-227 (2006).
 - 133 Zhang, Q., Wang, Y., Zheng, Q. & Li, J. Analysis of O-acetylated sialic acids in dried blood spots. *Analytical chemistry* **91**, 2744-2751 (2019).
 - 134 Klein, A. *et al.* New sialic acids from biological sources identified by a comprehensive and sensitive approach: liquid chromatography-electrospray ionization-mass spectrometry (LC-ESI-MS) of SIA quinoxalinones. *Glycobiology* **7**, 421-432 (1997).
 - 135 Cheeseman, J., Kuhnle, G., Spencer, D. I. & Osborn, H. M. Assays for the identification and quantification of sialic acids: challenges, opportunities and future perspectives. *Bioorganic & Medicinal Chemistry* **30**, 115882 (2021).
 - 136 McDonald, N. D., DeMeester, K. E., Lewis, A. L., Grimes, C. L. & Fidelma Boyd, E. Structural and functional characterization of a modified legionaminic acid involved in glycosylation of a bacterial lipopolysaccharide. *Journal of Biological Chemistry* **293**, 19113-19126 (2018). <https://doi.org/10.1074/JBC.RA118.004966>
 - 137 Galuska, S. P. *et al.* Quantification of nucleotide-activated sialic acids by a combination of reduction and fluorescent labeling. *Analytical Chemistry* **82**, 4591-4598 (2010). https://doi.org/10.1021/AC100627E/SUPPL_FILE/AC100627E_SI_001.PDF
 - 138 Kleikamp, H. B. C. *et al.* Tackling the chemical diversity of microbial nonulosonic acids – a universal large-scale survey approach. *Chemical Science* **11**, 3074-3080 (2020). <https://doi.org/10.1039/C9SC06406K>
 - 139 Tomás-Martínez, S. *et al.* Production of nonulosonic acids in the extracellular polymeric substances of "Candidatus Accumulibacter phosphatis". *Applied*

- Microbiology and Biotechnology* **105**, 3327-3338 (2021). <https://doi.org/10.1007/S00253-021-11249-3/TABLES/3>
- 140 Vliegthart, J. *et al.* Migration of O-acetyl groups in N, O-acetylneuraminic acids. *European journal of biochemistry* **162**, 301-607 (1987).
 - 141 Zhao, C. *et al.* Analysis of O-acetylated sialic acids by 3-nitrophenylhydrazine derivatization combined with LC-MS/MS. *Analytical Methods* **16**, 2472-2477 (2024). <https://doi.org/10.1039/D4AY00330F>
 - 142 Sha, S., Handelman, G., Agarabi, C. & Yoon, S. A high-resolution measurement of nucleotide sugars by using ion-pair reverse chromatography and tandem columns. *Analytical and Bioanalytical Chemistry* **412**, 3683-3693 (2020). <https://doi.org/10.1007/S00216-020-02608-6/FIGURES/8>
 - 143 García-Alvarez-Coque, M. C., Ramis-Ramos, G. & Ruiz-Angel, M. J. in *Encyclopedia of Analytical Science* 117-126 (2019).
 - 144 Ståhlberg, J. in *Encyclopedia of Separation Science* (ed Ian D. Wilson) 676-684 (Academic Press, 2000).
 - 145 Kochanowski, N. *et al.* Intracellular nucleotide and nucleotide sugar contents of cultured CHO cells determined by a fast, sensitive, and high-resolution ion-pair RP-HPLC. *Analytical Biochemistry* **348**, 243-251 (2006). <https://doi.org/10.1016/J.AB.2005.10.027>
 - 146 Buescher, J. M., Moco, S., Sauer, U. & Zamboni, N. Ultrahigh performance liquid chromatography-tandem mass spectrometry method for fast and robust quantification of anionic and aromatic metabolites. *Analytical Chemistry* **82**, 4403-4412 (2010). https://doi.org/10.1021/AC100101D/SUPPL_FILE/AC100101D_SI_001.PDF
 - 147 Van Scherpenzeel, M. *et al.* Dynamic tracing of sugar metabolism reveals the mechanisms of action of synthetic sugar analogs. *Glycobiology* **32**, 239-250 (2022). <https://doi.org/10.1093/GLYCOB/CWAB106>
 - 148 Coulier, L. *et al.* Simultaneous quantitative analysis of metabolites using ion-pair liquid chromatography-electrospray ionization mass spectrometry. *Analytical Chemistry* **78**, 6573-6582 (2006). <https://doi.org/10.1021/AC0607616/ASSET/IMAGES/LARGE/AC0607616F00005.JPEG>
 - 149 Rábiná, J. *et al.* Analysis of nucleotide sugars from cell lysates by ion-pair solid-phase extraction and reversed-phase high-performance liquid chromatography. *Glycoconjugate Journal* **18**, 799-805 (2001). <https://doi.org/10.1023/A:1021107602535/METRICS>
 - 150 Forngren, B. H. *et al.* Reversed-phase ion-pair chromatography coupled to electrospray ionisation mass spectrometry by on-line removal of the counter-ions. *Journal of Chromatography A* **854**, 155-162 (1999). [https://doi.org/10.1016/S0021-9673\(99\)00526-9](https://doi.org/10.1016/S0021-9673(99)00526-9)
 - 151 Cummins, P. M., Rochfort, K. D. & O'Connor, B. F. Ion-exchange chromatography: Basic principles and application. *Methods in Molecular Biology* **1485**, 209-223 (2017). https://doi.org/10.1007/978-1-4939-6412-3_11/TABLES/2
 - 152 Hardy, M. R., Townsend, R. R. & Lee, Y. C. Monosaccharide analysis of glycoconjugates by anion exchange chromatography with pulsed amperometric detection. *Analytical biochemistry* **170**, 54-62 (1988).
 - 153 Hardy, M. R. & Townsend, R. R. Separation of positional isomers of oligosaccharides and glycopeptides by high-performance anion-exchange chromatography with pulsed amperometric detection. *Proceedings of the National Academy of Sciences* **85**, 3289-3293 (1988).
 - 154 Zhang, Z., Khan, N. M., Nunez, K. M., Chess, E. K. & Szabo, C. M. Complete monosaccharide analysis by high-performance anion-exchange chromatography with pulsed amperometric detection. *Analytical Chemistry* **84**, 4104-4110 (2012). https://doi.org/10.1021/AC300176Z/SUPPL_FILE/AC300176Z_SI_001.PDF
 - 155 Lee, Y.-C. Carbohydrate analyses with high-performance anion-exchange chromatography. *Journal of chromatography A* **720**, 137-149 (1996).
 - 156 Karlsson, N. G. & Hansson, G. C. Analysis of monosaccharide composition of mucin oligosaccharide alditols by high-performance anion-exchange chromatography. *Analytical biochemistry* **224**, 538-541 (1995).
 - 157 Behan, J. L. & Smith, K. D. The analysis of glycosylation: a continued need for high pH anion exchange chromatography. *Biomedical Chromatography* **25**, 39-46 (2011).
 - 158 Townsend, R. R. & Hardy, M. R. Analysis of glycoprotein oligosaccharides using high-pH anion exchange chromatography. *Glycobiology* **1**, 139-147 (1991).
 - 159 Mechelke, M. *et al.* HPAEC-PAD for oligosaccharide analysis—novel insights into analyte sensitivity and response stability. *Analytical and bioanalytical chemistry* **409**, 7169-7181 (2017).
 - 160 Cataldi, T. R., Campa, C. & De Benedetto, G. E. Carbohydrate analysis by high-performance anion-exchange chromatography with pulsed amperometric detection: the potential is still growing. *Fresenius' journal of analytical chemistry* **368**, 739-758 (2000).
 - 161 Li, J., Hao, X., Persiani, R., van Loosdrecht, M. C. & Lin, Y. Reinvestigating the Composition of Alginate-Like Exopolymers Extracted from Activated Sludge. *ACS ES&T Water* (2024).
 - 162 Barr, J. R., Anumula, K. R., Vettese, M. B., Taylor, P. B. & Carr, S. A. Structural classification of carbohydrates in glycoproteins by mass spectrometry and high-performance anion-exchange chromatography. *Analytical biochemistry* **192**, 181-192 (1991).
 - 163 Okinaga, T., Ohashi, Y. & Hoshi, M. A novel saccharide structure, Xyl 1→3 Gal 1→(SO₃⁻) 3, 4 Fuc→, is present in acrosome reaction-inducing substance of the starfish, *Asterias amurensis*. *Biochemical and biophysical research communications* **186**, 405-410 (1992).
 - 164 Walters, J. J. *et al.* Ion-exchange chromatography followed by ESI-MS for quantitative analysis of sugar monophosphates from glucose catabolism. *Journal of the American Society for Mass Spectrometry* **17**, 104-107 (2006).
 - 165 Ramm, M., Wolfender, J.-L., Queiroz, E. F., Hostettmann, K. & Hamburger, M. Rapid analysis of nucleotide-activated sugars by high-performance liquid chromatography coupled with diode-array detection, electrospray ionization mass spectrometry and nuclear magnetic resonance. *Journal of Chromatography A* **1034**, 139-148 (2004). <https://doi.org/10.1016/j.chroma.2004.02.023>
 - 166 Manzi, A. E., Diaz, S. & Varki, A. High-pressure liquid chromatography of sialic acids on a pellicular resin anion-exchange column with pulsed amperometric detection: a comparison with six other systems. *Analytical biochemistry* **188**, 20-32 (1990).
 - 167 Bruggink, C., Maurer, R., Herrmann, H., Cavalli, S. & Hoefler, F. Analysis of carbohydrates by anion exchange chromatography and mass spectrometry. *Journal of Chromatography A* **1085**, 104-109 (2005). <https://doi.org/10.1016/J.CHROMA.2005.03.108>
 - 168 Wunschel, D. S. *et al.* Quantitative analysis of neutral and acidic sugars in whole bacterial cell hydrolysates

- using high-performance anion-exchange liquid chromatography–electrospray ionization tandem mass spectrometry. *Journal of Chromatography A* **776**, 205–219 (1997). [https://doi.org/10.1016/S0021-9673\(97\)00356-7](https://doi.org/10.1016/S0021-9673(97)00356-7)
- 169 Tedesco, R. *et al.* Carbohydrate determination in honey samples by ion chromatography–mass spectrometry (HPAEC–MS). *Analytical and Bioanalytical Chemistry* **412**, 5217–5227 (2020). <https://doi.org/10.1007/S00216-020-02732-3/FIGURES/5>
 - 170 Zhao, Z., Lin, Q., Feng, Y., Zhou, Y. & Wang, X. Determination of monosaccharides hydrolyzed from polysaccharides in activated sludge by ion chromatography–mass spectrometry with online pretreatment of column switching technique. *Analytical and Bioanalytical Chemistry* **412**, 8061–8071 (2020). <https://doi.org/10.1007/S00216-020-02955-4/TABLES/4>
 - 171 Bruggink, C. *et al.* Oligosaccharide analysis by capillary-scale high-pH anion-exchange chromatography with on-line ion-trap mass spectrometry. *Journal of Chromatography B* **829**, 136–143 (2005).
 - 172 Tomiya, N., Ailor, E., Lawrence, S. M., Betenbaugh, M. J. & Lee, Y. C. Determination of nucleotides and sugar nucleotides involved in protein glycosylation by high-performance anion-exchange chromatography: Sugar nucleotide contents in cultured insect cells and mammalian cells. *Analytical Biochemistry* **293**, 129–137 (2001). <https://doi.org/10.1006/ABIO.2001.5091>
 - 173 Alonso, A. P., Piasecki, R. J., Wang, Y., LaClair, R. W. & Shachar-Hill, Y. Quantifying the Labeling and the Levels of Plant Cell Wall Precursors Using Ion Chromatography Tandem Mass Spectrometry. *Plant Physiology* **153**, 915–924 (2010). <https://doi.org/10.1104/pp.110.155713>
 - 174 Pereira, L. Porous Graphitic Carbon as a Stationary Phase in HPLC: Theory and Applications. *Journal of Liquid Chromatography & Related Technologies* **31**, 1687–1731 (2008). <https://doi.org/10.1080/10826070802126429>
 - 175 West, C., Elfakir, C. & Lafosse, M. Porous graphitic carbon: A versatile stationary phase for liquid chromatography. *Journal of Chromatography A* **1217**, 3201–3216 (2010). <https://doi.org/10.1016/J.CHROMA.2009.09.052>
 - 176 Russo, M. *et al.* Principles and applications of porous graphitic carbon stationary phase in liquid chromatography: An update. *Journal of Chromatography A* **1719**, 464728–464728 (2024). <https://doi.org/10.1016/J.CHROMA.2024.464728>
 - 177 De Matteis, C. I., Simpson, D. A., Euerby, M. R., Shaw, P. N. & Barrett, D. A. Chromatographic retention behaviour of monosubstituted benzene derivatives on porous graphitic carbon and octadecyl-bonded silica studied using molecular modelling and quantitative structure–retention relationships. *Journal of Chromatography A* **1229**, 95–106 (2012).
 - 178 Lepont, C., Gunatillaka, A. D. & Poole, C. F. Retention characteristics of porous graphitic carbon in reversed-phase liquid chromatography with methanol–water mobile phases. *Analyst* **126**, 1318–1325 (2001).
 - 179 Knox, J. H. & Ross, P. Carbon-based packing materials for liquid chromatography: structure, performance, and retention mechanisms []. *Advances in chromatography* **37** (1997).
 - 180 Ross, P. & Knox, J. H. Carbon-based packing materials for liquid chromatography: applications. *Advances in chromatography* **37**, 121–162 (1997).
 - 181 Knox, J. H., Kaur, B. & Millward, G. Structure and performance of porous graphitic carbon in liquid chromatography. *Journal of Chromatography A* **352**, 3–25 (1986).
 - 182 Pabst, M. & Altmann, F. Influence of electrosorption, solvent, temperature, and ion polarity on the performance of LC–ESI–MS using graphitic carbon for acidic oligosaccharides. *Analytical chemistry* **80**, 7534–7542 (2008).
 - 183 Corman, C., Muraco, C., Ye, M., Frantz, C. & Maule, W. Applications of Newly Developed 2.7 µm Porous Graphitic Carbon Particles in U/HPLC. *Sci. & Techn. Innov. Analytix Report* **10**, 3–7 (2021).
 - 184 Corman, C., Muraco, C., Ye, M. & Michel, F. Insights into the Shape-Selective Retention Properties of Porous Graphitic Carbon Stationary Phases and Applicability for Polar Compounds. (2023).
 - 185 Chaïmbault, P., Elfakir, C. & Lafosse, M. Comparison of the retention behavior of polyethoxylated alcohols on porous graphitic carbon and polar as well as apolar bonded-silica phases. *Journal of Chromatography A* **797**, 83–91 (1998).
 - 186 Wan, Q. H., Shaw, P. N., Davies, M. C. & Barrett, D. A. Chromatographic behaviour of positional isomers on porous graphitic carbon. *Journal of Chromatography A* **697**, 219–227 (1995).
 - 187 Shibukawa, M., Terashima, H., Nakajima, H. & Saitoh, K. Evaluation of the surface charge properties of porous graphitic carbon stationary phases treated with redox agents. *Analyst* **129**, 623–628 (2004).
 - 188 Törnkvist, A., Markides, K. E. & Nyholm, L. Chromatographic behaviour of oxidised porous graphitic carbon columns. *Analyst* **128**, 844–848 (2003).
 - 189 Bapiro, T. E., Richards, F. M. & Jodrell, D. I. Understanding the Complexity of Porous Graphitic Carbon (PGC) Chromatography: Modulation of Mobile-Stationary Phase Interactions Overcomes Loss of Retention and Reduces Variability. *Analytical Chemistry* **88**, 6190–6194 (2016). https://doi.org/10.1021/ACS.ANALCHEM.6B01167/ASSET/IMAGES/LARGE/AC-2016-01167F_0003.JPEG
 - 190 Törnkvist, A., Nilsson, S., Amirkhani, A., Nyholm, L. M. & Nyholm, L. Interference of the electrospray voltage on chromatographic separations using porous graphitic carbon columns. *Journal of mass spectrometry* **39**, 216–222 (2004).
 - 191 Antonio, C. *et al.* Quantification of sugars and sugar phosphates in Arabidopsis thaliana tissues using porous graphitic carbon liquid chromatography–electrospray ionization mass spectrometry. *Journal of Chromatography A* **1172**, 170–178 (2007). <https://doi.org/10.1016/J.CHROMA.2007.10.011>
 - 192 Li, P., Su, M., Chatterjee, M. & Lämmerhofer, M. Targeted analysis of sugar phosphates from glycolysis pathway by phosphate methylation with liquid chromatography coupled to tandem mass spectrometry. *Analytica Chimica Acta* **1221**, 340099 (2022).
 - 193 Pabst, M. *et al.* Nucleotide and nucleotide sugar analysis by liquid chromatography–electrospray ionization–mass spectrometry on surface-conditioned porous graphitic carbon. *Analytical Chemistry* **82**, 9782–9788 (2010). https://doi.org/10.1021/AC101975K/SUPPL_FILE/AC101975K_SI_001.PDF
 - 194 Behmüller, R., Forstenlehner, I. C., Tenhaken, R. & Huber, C. G. Quantitative HPLC–MS analysis of nucleotide sugars in plant cells following off-line SPE sample preparation. *Analytical and Bioanalytical Chemistry* **406**, 3229–3237 (2014). <https://doi.org/10.1007/S00216-014-7746-3/FIGURES/5>
 - 195 López-Gutiérrez, B., Dinglasan, R. R. & Izquierdo, L. Sugar nucleotide quantification by liquid chromatography

- tandem mass spectrometry reveals a distinct profile in *Plasmodium falciparum* sexual stage parasites. *Biochemical Journal* **474**, 897-905 (2017). <https://doi.org/10.1042/BCJ20161030>
- 196 Garcia, A. D., Chavez, J. L. & Mechref, Y. Sugar nucleotide quantification using multiple reaction monitoring liquid chromatography/tandem mass spectrometry. *Rapid Communications in Mass Spectrometry* **27**, 1794-1800 (2013).
 - 197 Caron, P. *et al.* A liquid chromatography-mass spectrometry assay for the quantification of nucleotide sugars in human plasma and urine specimens and its clinical application. *Journal of Chromatography A* **1677**, 463296 (2022).
 - 198 Wang, T. *et al.* Characterization of nucleotides and nucleotide sugars in *Candida albicans* by high performance liquid chromatography-mass spectrometry with a porous graphite carbon column. *Analytical Letters* **47**, 234-249 (2014).
 - 199 Castilho, A. *et al.* Construction of a functional CMP-sialic acid biosynthesis pathway in *Arabidopsis*. *Plant physiology* **147**, 331-339 (2008).
 - 200 Rejzek, M. *et al.* in *Methods in enzymology* Vol. 597 209-238 (Elsevier, 2017).
 - 201 Alpert, A. J. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *Journal of Chromatography A* **499**, 177-196 (1990). [https://doi.org/10.1016/S0021-9673\(00\)96972-3](https://doi.org/10.1016/S0021-9673(00)96972-3)
 - 202 Linden, J. C. & Lawhead, C. L. Liquid chromatography of saccharides. *Journal of Chromatography A* **105**, 125-133 (1975). [https://doi.org/10.1016/S0021-9673\(01\)81096-7](https://doi.org/10.1016/S0021-9673(01)81096-7)
 - 203 INOUE, Y. & YAMAMOTO, A. Stationary phases for the separation of reducing sugars by normal-phase partition chromatography. *Chromatography* **35**, 63-72 (2014).
 - 204 Ito, J. *et al.* Analysis of plant nucleotide sugars by hydrophilic interaction liquid chromatography and tandem mass spectrometry. *Analytical Biochemistry* **448**, 14-22 (2014). <https://doi.org/10.1016/J.AB.2013.11.026>
 - 205 McCalley, D. V. Study of the selectivity, retention mechanisms and performance of alternative silica-based stationary phases for separation of ionised solutes in hydrophilic interaction chromatography. *Journal of Chromatography A* **1217**, 3408-3417 (2010). <https://doi.org/10.1016/J.CHROMA.2010.03.011>
 - 206 McCalley, D. V. Separation mechanisms in hydrophilic interaction chromatography. *Hydrophilic Interaction Chromatography: A Guide for Practitioners*, 1-41 (2013).
 - 207 Cubbon, S., Antonio, C., Wilson, J. & Thomas-Oates, J. Metabolomic applications of HILIC-LC-MS. *Mass Spectrometry Reviews* **29**, 671-684 (2010). <https://doi.org/10.1002/MAS.20252>
 - 208 Guo, Y. Recent progress in the fundamental understanding of hydrophilic interaction chromatography (HILIC). *Analyst* **140**, 6452-6466 (2015).
 - 209 Tang, D. Q., Zou, L., Yin, X. X. & Ong, C. N. HILIC-MS for metabolomics: An attractive and complementary approach to RPLC-MS. *Mass spectrometry reviews* **35**, 574-600 (2016).
 - 210 Cubbon, S., Antonio, C., Wilson, J. & Thomas-Oates, J. Metabolomic applications of hilic-lc-ms. *Mass spectrometry reviews* **29**, 671-684 (2010).
 - 211 Spagou, K. *et al.* Hydrophilic interaction chromatography coupled to MS for metabolomic/metabolomic studies. *Journal of separation science* **33**, 716-727 (2010).
 - 212 Kohler, I., Verhoeven, M., Haselberg, R. & Gargano, A. F. Hydrophilic interaction chromatography-mass spectrometry for metabolomics and proteomics: state-of-the-art and current trends. *Microchemical Journal* **175**, 106986 (2022).
 - 213 Segers, K., Declerck, S., Mangelings, D., Heyden, Y. V. & Eeckhaut, A. V. Analytical techniques for metabolomic studies: A review. *Bioanalysis* **11**, 2297-2318 (2019).
 - 214 Patti, G. J. Separation strategies for untargeted metabolomics. *Journal of separation science* **34**, 3460-3469 (2011).
 - 215 Kohler, I., Giera, M. & Derks, R. J. The rise of hydrophilic interaction chromatography in untargeted clinical metabolomics. (2016).
 - 216 Rojo, D., Barbas, C. & Rupérez, F. J. LC-MS metabolomics of polar compounds. *Bioanalysis* **4**, 1235-1243 (2012).
 - 217 Hosseinkhani, F. *et al.* Systematic evaluation of HILIC stationary phases for global metabolomics of human plasma. *Metabolites* **12**, 165 (2022).
 - 218 Kohler, I. & Giera, M. Recent advances in liquid-phase separations for clinical metabolomics. *Journal of separation science* **40**, 93-108 (2017).
 - 219 Wernisch, S. & Pennathur, S. Evaluation of coverage, retention patterns, and selectivity of seven liquid chromatographic methods for metabolomics. *Analytical and bioanalytical chemistry* **408**, 6079-6091 (2016).
 - 220 Buszewski, B. & Noga, S. Hydrophilic interaction liquid chromatography (HILIC)—a powerful separation technique. *Analytical and bioanalytical chemistry* **402**, 231-247 (2012).
 - 221 Ikegami, T. *et al.* Highly efficient analysis of underivatized carbohydrates using monolithic-silica-based capillary hydrophilic interaction (HILIC) HPLC. *Analytical and Bioanalytical Chemistry* **391**, 2533-2542 (2008).
 - 222 Gao, Q., Nilsson, U., Ilag, L. L. & Leck, C. Monosaccharide compositional analysis of marine polysaccharides by hydrophilic interaction liquid chromatography-tandem mass spectrometry. *Analytical and bioanalytical chemistry* **399**, 2517-2529 (2011).
 - 223 Meyer, M., Montero, L., Meckelmann, S. W. & Schmitz, O. J. Comparative study for analysis of carbohydrates in biological samples. *Analytical and Bioanalytical Chemistry* **414**, 2117-2130 (2022).
 - 224 Mathon, C., Barding Jr, G. A. & Larive, C. K. Separation of ten phosphorylated mono- and disaccharides using HILIC and ion-pairing interactions. *Analytica chimica acta* **972**, 102-110 (2017).
 - 225 Meriö-Talvio, H., Dou, J., Vuorinen, T. & Pitkänen, L. Fast HILIC method for separation and quantification of non-volatile aromatic compounds and monosaccharides from willow (*Salix* sp.) bark extract. *Applied Sciences* **11**, 3808 (2021).
 - 226 Pismennõ, D., Kiritsenko, V., Marhivka, J., Kütt, M.-L. & Vilu, R. Development and optimisation of HILIC-LC-MS method for determination of carbohydrates in fermentation samples. *Molecules* **26**, 3669 (2021).
 - 227 Yan, X. High-performance liquid chromatography for carbohydrate analysis. *High-Performance Liquid Chromatography (HPLC): Principles, Practices and Procedures*; Zhuo, Y., Ed (2014).
 - 228 Lowenthal, M. S., Kilpatrick, E. L. & Phinney, K. W. Separation of monosaccharides hydrolyzed from glycoproteins without the need for derivatization. *Analytical and bioanalytical chemistry* **407**, 5453-5462 (2015).
 - 229 Li, Y., He, Z., Zou, P., Ning, Y. & Zhu, X. Determination of seventeen sugars and sugar alcohols in fruit juice samples using hydrophilic interaction liquid chromatography-tandem mass spectrometry combining response surface methodology design. *Microchemical*

- Journal* **193**, 109136-109136 (2023).
<https://doi.org/10.1016/J.MICROC.2023.109136>
- 230 Su, M., Serafimov, K., Li, P., Knappe, C. & Lämmerhofer, M. Isomer selectivity of one- and two-dimensional approaches of mixed-mode and hydrophilic interaction liquid chromatography coupled to tandem mass spectrometry for sugar phosphates of glycolysis and pentose phosphate pathways. *Journal of Chromatography A* **1688**, 463727-463727 (2023).
<https://doi.org/10.1016/J.CHROMA.2022.463727>
 - 231 Hinterwirth, H. et al. Selectivity issues in targeted metabolomics: Separation of phosphorylated carbohydrate isomers by mixed-mode hydrophilic interaction/weak anion exchange chromatography. *Journal of separation science* **33**, 3273-3282 (2010).
 - 232 Hammad, L. A., Derryberry, D. Z., Jmeian, Y. R. & Mechref, Y. Quantification of monosaccharides through multiple-reaction monitoring liquid chromatography/mass spectrometry using an aminopropyl column. *Rapid Communications in Mass Spectrometry* **24**, 1565-1574 (2010).
 - 233 Matějčák, D. & Vašíčková, J. Hydrophilic interaction chromatography with Fourier transform mass spectrometry of monosaccharide anhydrides. *Journal of Separation Science* **39**, 4748-4755 (2016).
 - 234 Yan, J. et al. Neutral monosaccharide composition analysis of plant-derived oligo- and polysaccharides by high performance liquid chromatography. *Carbohydrate polymers* **136**, 1273-1280 (2016).
 - 235 Roca, C., Alves, V. D., Freitas, F. & Reis, M. A. Exopolysaccharides enriched in rare sugars: bacterial sources, production, and applications. *Frontiers in microbiology* **6**, 288 (2015).
 - 236 Zhang, W., Zhang, T., Jiang, B. & Mu, W. Enzymatic approaches to rare sugar production. *Biotechnology advances* **35**, 267-274 (2017).
 - 237 Ahmed, A., Khan, T. A., Dan Ramdath, D., Kendall, C. W. & Sievenpiper, J. L. Rare sugars and their health effects in humans: a systematic review and narrative synthesis of the evidence from human trials. *Nutrition reviews* **80**, 255-270 (2022).
 - 238 Franchi, F. et al. Effects of D-allulose on glucose tolerance and insulin response to a standard oral sucrose load: results of a prospective, randomized, crossover study. *BMJ Open Diabetes Research and Care* **9**, e001939 (2021).
 - 239 Hossain, M. A. et al. Rare sugar D-psicose improves insulin sensitivity and glucose tolerance in type 2 diabetes Otsuka Long-Evans Tokushima Fatty (OLETF) rats. *Biochemical and biophysical research communications* **405**, 7-12 (2011).
 - 240 Chung, H., Shimura, A. & Matsui, T. Discriminant and simultaneous HPLC analysis of reducing and non-reducing monosaccharides on a polyethyleneimine-attached hydrophilic interaction liquid chromatography column. *Food Science and Technology Research* **24**, 501-508 (2018).
 - 241 Fu, X., Cebo, M., Ikegami, T. & Lämmerhofer, M. Separation of carbohydrate isomers and anomers on poly-N-(1H-tetrazole-5-yl)-methacrylamide-bonded stationary phase by hydrophilic interaction chromatography as well as determination of anomer interconversion energy barriers. *Journal of Chromatography A* **1620**, 460981 (2020).
 - 242 Li, S., Cai, W. J., Wang, W., Sun, M. X. & Feng, Y. Q. Rapid Analysis of Monosaccharides in Sub-milligram Plant Samples Using Liquid Chromatography-Mass Spectrometry Assisted by Post-column Derivatization. *Journal of Agricultural and Food Chemistry* **68**, 2588-2596 (2020).
https://doi.org/10.1021/ACS.JAFC.9B07623/ASSET/IMAGES/LARGE/JF9B07623_0005.JPEG
 - 243 Heise, T. et al. Potent Metabolic Sialylation Inhibitors Based on C-5-Modified Fluorinated Sialic Acids. *Journal of Medicinal Chemistry* **62**, 1014-1021 (2019).
https://doi.org/10.1021/ACS.JMEDCHEM.8B01757/SUPPL_FILE/JM8B01757_SI_002.CSV
 - 244 Ma, L. et al. Genetic alteration of UDP-rhamnose metabolism in *Botrytis cinerea* leads to the accumulation of UDP-KDG that adversely affects development and pathogenicity. *Molecular Plant Pathology* **18**, 263-275 (2017).
<https://doi.org/10.1111/MPP.12398>
 - 245 Li, Z., Hwang, S., Ericson, J., Bowler, K. & Bar-Peled, M. Pen and pal are nucleotide-sugar dehydratases that convert UDP-GlcNAc to UDP-6-deoxy-D-GlcNAc-5,6-ene and then to UDP-4-keto-6-deoxy-L-AltNAc for CMP-pseudaminic acid synthesis in *Bacillus thuringiensis*. *Journal of Biological Chemistry* **290**, 691-704 (2015).
<https://doi.org/10.1074/jbc.M114.612747>
 - 246 Warth, B. et al. Hydrophilic interaction liquid chromatography coupled with tandem mass spectrometry for the quantification of uridine diphosphate-glucose, uridine diphosphate-glucuronic acid, deoxynivalenol and its glucoside: In-house validation and application to wheat. *Journal of Chromatography A* **1423**, 183-189 (2015).
<https://doi.org/10.1016/J.CHROMA.2015.10.070>
 - 247 Teixeira, J., Tiritan, M. E., Pinto, M. M. & Fernandes, C. Chiral stationary phases for liquid chromatography: recent developments. *Molecules* **24**, 865 (2019).
 - 248 Bai, C., Ohri, H., Nishida, Y. & Meguro, H. A Fluorometric Determination Method for the Configurations of Per-O-Methylated Monosaccharides by Anomeric 2-Methyl-2-β-naphthyl-1, 3-benzodioxole 4-Carboxylation and High-Performance Liquid Chromatography. *Analytical biochemistry* **246**, 246-252 (1997).
 - 249 De Muynck, C., Beauprez, J., Soetaert, W. & Vandamme, E. J. Boric acid as a mobile phase additive for high performance liquid chromatography separation of ribose, arabinose and ribulose. *Journal of Chromatography A* **1101**, 115-121 (2006).
 - 250 Lopes, J. F. & Gaspar, E. M. Simultaneous chromatographic separation of enantiomers, anomers and structural isomers of some biologically relevant monosaccharides. *Journal of Chromatography A* **1188**, 34-42 (2008).
 - 251 Wang, L., Wei, W., Xia, Z., Jie, X. & Xia, Z. Z. Recent advances in materials for stationary phases of mixed-mode high-performance liquid chromatography. *TrAC Trends in Analytical Chemistry* **80**, 495-506 (2016).
<https://doi.org/10.1016/J.TRAC.2016.04.001>
 - 252 Sýkora, D., Řezanka, P., Záruba, K. & Král, V. Recent advances in mixed-mode chromatographic stationary phases. *Journal of Separation Science* **42**, 89-129 (2019).
<https://doi.org/10.1002/JSSC.201801048>
 - 253 Ammann, A. A. & Suter, M. J. F. Multimode gradient high performance liquid chromatography mass spectrometry method applicable to metabolomics and environmental monitoring. *Journal of chromatography. A* **1456**, 145-151 (2016).
<https://doi.org/10.1016/J.CHROMA.2016.06.001>
 - 254 Halfpenny, A. P. & Brown, P. R. Mixed mode chromatography via column switching for the simultaneous HPLC analysis of ionic and non-ionic nucleic acid constituents. *Chromatographia* **21**, 317-320 (1986).
<https://doi.org/10.1007/BF02311602/METRICS>

- 255 Kennedy, L. A., Kopaciewicz, W. & Regnier, F. E. Multimodal liquid chromatography columns for the separation of proteins in either the anion-exchange or hydrophobic-interaction mode. *Journal of Chromatography A* **359**, 73-84 (1986). [https://doi.org/10.1016/0021-9673\(86\)80063-2](https://doi.org/10.1016/0021-9673(86)80063-2)
- 256 Chen, M. & Chen, L. A novel high hydrothermal stability amino-functionalized stationary phase prepared by a vapour deposition method. *Analytical Methods* **10**, 1538-1546 (2018). <https://doi.org/10.1039/C8AY00133B>
- 257 Eastwood, H. et al. Development of a nucleotide sugar purification method using a mixed mode column & mass spectrometry detection. *Journal of Pharmaceutical and Biomedical Analysis* **115**, 402-409 (2015). <https://doi.org/10.1016/J.JPBA.2015.08.001>
- 258 Xing, G., Sresht, V., Sun, Z., Shi, Y. & Clasquin, M. F. Coupling mixed mode chromatography/esi negative ms detection with message-passing neural network modeling for enhanced metabolome coverage and structural identification. *Metabolites* **11**, 772-772 (2021). <https://doi.org/10.3390/METABO11110772/S1>
- 259 Hinterwirth, H. et al. Selectivity issues in targeted metabolomics: Separation of phosphorylated carbohydrate isomers by mixed-mode hydrophilic interaction/weak anion exchange chromatography. *Journal of Separation Science* **33**, 3273-3282 (2010). <https://doi.org/10.1002/JSSC.201000412>
- 260 Rahm, M., Kwast, H., Wessels, H. J., Noga, M. J. & Lefebvre, D. J. Mixed-phase weak anion-exchange/reversed-phase LC-MS/MS for analysis of nucleotide sugars in human fibroblasts. *Analytical and bioanalytical chemistry*, 1-10 (2024).
- 261 Pauk, V., Pluháček, T., Havlíček, V. & Lemr, K. Ultra-high performance supercritical fluid chromatography-mass spectrometry procedure for analysis of monosaccharides from plant gum binders. *Analytica Chimica Acta* **989**, 112-120 (2017). <https://doi.org/10.1016/J.ACA.2017.07.036>
- 262 Losacco, G. L. et al. Applicability of Supercritical fluid chromatography–Mass spectrometry to metabolomics. II–Assessment of a comprehensive library of metabolites and evaluation of biological matrices. *Journal of Chromatography A* **1620**, 461021-461021 (2020). <https://doi.org/10.1016/J.CHROMA.2020.461021>
- 263 van de Velde, B., Guillaume, D. & Kohler, I. Supercritical fluid chromatography – Mass spectrometry in metabolomics: Past, present, and future perspectives. *Journal of Chromatography B* **1161**, 122444-122444 (2020). <https://doi.org/10.1016/J.JCHROMB.2020.122444>
- 264 Losacco, G. L., Veuthey, J. L. & Guillaume, D. Supercritical fluid chromatography – Mass spectrometry: Recent evolution and current trends. *TrAC Trends in Analytical Chemistry* **118**, 731-738 (2019). <https://doi.org/10.1016/J.TRAC.2019.07.005>
- 265 Gazárková, T., Plachká, K., Svec, F. & Nováková, L. Current state of supercritical fluid chromatography-mass spectrometry. *TrAC Trends in Analytical Chemistry* **149**, 116544-116544 (2022). <https://doi.org/10.1016/J.TRAC.2022.116544>
- 266 Desfontaine, V. et al. Applicability of supercritical fluid chromatography – mass spectrometry to metabolomics. I – Optimization of separation conditions for the simultaneous analysis of hydrophilic and lipophilic substances. *Journal of Chromatography A* **1562**, 96-107 (2018). <https://doi.org/10.1016/J.CHROMA.2018.05.055>
- 267 Huang, Y. & Jiang, Z. Determination of Saccharides in Honey Using Supercritical Fluid Chromatography Coupled with Single Quadrupole Mass Spectrometry. *Journal of Chromatographic Science* **60**, 705-711 (2022). <https://doi.org/10.1093/CHROMSCI/BMAB104>
- 268 Ruiz-Matute, A. I., Hernández-Hernández, O., Rodríguez-Sánchez, S., Sanz, M. L. & Martínez-Castro, I. Derivatization of carbohydrates for GC and GC–MS analyses. *Journal of chromatography B* **879**, 1226-1240 (2011).
- 269 DeJongh, D. C. et al. Analysis of trimethylsilyl derivatives of carbohydrates by gas chromatography and mass spectrometry. *Journal of the American Chemical Society* **91**, 1728-1740 (1969).
- 270 Merkle, R. K. & Poppe, I. in *Methods in enzymology* Vol. 230 1-15 (Elsevier, 1994).
- 271 McInnes, A., Gall, D., Cooper, F. & Bishop, C. Separation of carbohydrate derivatives by gas-liquid partition chromatography: J.Chromatogr. *J.Chromatogr.* **1**, 556-557 (1958).
- 272 Fox, A. in *Journal of chromatography library* Vol. 66 829-843 (Elsevier, 2002).
- 273 Jousse, C. & Pujos-Guillot, E. in *Advances in Botanical Research* Vol. 67 303-329 (Elsevier, 2013).
- 274 Obendorf, R. L., Horbowicz, M. & Lahuta, L. B. Characterization of sugars, cyclitols and galactosyl cyclitols in seeds by GC. (2012).
- 275 Geyer, R. & Geyer, H. in *Methods in enzymology* Vol. 230 86-108 (Elsevier, 1994).
- 276 BeMiller, J. N. Carbohydrate analysis. *Food analysis*, 333-360 (2017).
- 277 Biermann, C. J. in *Analysis of carbohydrates by GLC and MS* 1-17 (CRC Press, 2021).
- 278 Liu, D., Tang, W., Yin, J.-Y., Nie, S.-P. & Xie, M.-Y. Monosaccharide composition analysis of polysaccharides from natural sources: Hydrolysis condition and detection method development. *Food Hydrocolloids* **116**, 106641 (2021).
- 279 Sweeley, C. C., Wells, W. W. & Bentley, R. in *Methods in enzymology* Vol. 8 95-108 (Elsevier, 1966).
- 280 Engel, B., Suralik, P. & Marchetti-Deschmann, M. Critical considerations for trimethylsilyl derivatives of 24 primary metabolites measured by gas chromatography–tandem mass spectrometry. *Separation Science Plus* **3**, 407-418 (2020).
- 281 Pons, A. et al. Sequential GC/MS analysis of sialic acids, monosaccharides, and amino acids of glycoproteins on a single sample as heptafluorobutyrate derivatives. *Biochemistry* **42**, 8342-8353 (2003).
- 282 Zanetta, J.-P. et al. Diversity of sialic acids revealed using gas chromatography/mass spectrometry of heptafluorobutyrate derivatives. *Glycobiology* **11**, 663-676 (2001).
- 283 Kamerling, J. P. & Gerwig, G. J. Structural analysis of naturally occurring sialic acids. *Glycobiology protocols*, 69-91 (2007).
- 284 Roessner, U., Wagner, C., Kopka, J., Trethewey, R. N. & Willmitzer, L. Simultaneous analysis of metabolites in potato tuber by gas chromatography–mass spectrometry. *the plant journal* **23**, 131-142 (2000).
- 285 Cipollina, C. et al. A comprehensive method for the quantification of the non-oxidative pentose phosphate pathway intermediates in *Saccharomyces cerevisiae* by GC–IDMS. *Journal of Chromatography B* **877**, 3231-3236 (2009).
- 286 Koek, M. M., Muijtijk, B., van der Werf, M. J. & Hankemeier, T. Microbial metabolomics with gas chromatography/mass spectrometry. *Analytical chemistry* **78**, 1272-1281 (2006).
- 287 de Jonge, L. P., Douma, R. D., Heijnen, J. J. & van Gulik, W. M. Optimization of cold methanol quenching for

- quantitative metabolomics of *Penicillium chrysogenum*. *Metabolomics* **8**, 727-735 (2012).
- 288 Chu, D. B. *et al.* Isotopologue analysis of sugar phosphates in yeast cell extracts by gas chromatography chemical ionization time-of-flight mass spectrometry. *Analytical and bioanalytical chemistry* **407**, 2865-2875 (2015).
 - 289 Okahashi, N. *et al.* Sugar phosphate analysis with baseline separation and soft ionization by gas chromatography-negative chemical ionization-mass spectrometry improves flux estimation of bidirectional reactions in cancer cells. *Metabolic engineering* **51**, 43-49 (2019).
 - 290 Koubaa, M., Mghaieth, S., Thomasset, B. & Roscher, A. Gas chromatography-mass spectrometry analysis of ¹³C labeling in sugars for metabolic flux analysis. *Analytical biochemistry* **425**, 183-188 (2012).
 - 291 Faraco, M., Fico, D., Pennetta, A. & De Benedetto, G. E. New evidences on efficacy of boronic acid-based derivatization method to identify sugars in plant material by gas chromatography-mass spectrometry. *Talanta* **159**, 40-46 (2016).
 - 292 Corey, E. & Chaykovsky, M. Methylsulfinylcarbanion. *Journal of the American Chemical Society* **84**, 866-867 (1962).
 - 293 HAKOMORI, S.-I. A rapid permethylation of glycolipid, and polysaccharide catalyzed by methylsulfinyl carbanion in dimethyl sulfoxide. *The Journal of Biochemistry* **55**, 205-208 (1964).
 - 294 Haas, M., Lamour, S. & Trapp, O. Development of an advanced derivatization protocol for the unambiguous identification of monosaccharides in complex mixtures by gas and liquid chromatography. *Journal of Chromatography A* **1568**, 160-167 (2018).
 - 295 Harsch, G., Harsch, M., Bauer, H. & Voelter, W. Produktverteilung und Mechanismus der Gesamtreaktion der Formose-Reaktion/Product Analysis und Mechanismus der Formose Reaction. *Zeitschrift für Naturforschung B* **38**, 1269-1280 (1983).
 - 296 Pallmann, S. *et al.* Schreibersite: an effective catalyst in the formose reaction network. *New Journal of Physics* **20**, 055003 (2018).
 - 297 König, W. A., Mischnick-Lübbecke, P., Brassat, B., Lutz, S. & Wenz, G. Improved gas chromatographic separation of enantiomeric carbohydrate derivatives using a new chiral stationary phase. *Carbohydrate research* **183**, 11-17 (1988).
 - 298 Gervig, G. J., Kamerling, J. P. & Vliegthart, J. F. Determination of the D and L configuration of neutral monosaccharides by high-resolution capillary GLC. *Carbohydrate Research* **62**, 349-357 (1978).
 - 299 Leontin, K., Lindberg, B. & Lönngren, J. Assignment of absolute configuration of sugars by glc of their acetylated glycosides formed from chiral alcohols. *Carbohydrate Research* **62**, 359-362 (1978).
 - 300 Myrgorodska, I., Javelle, T., Meinert, C. & Meierhenrich, U. J. Enantioresolution and quantification of monosaccharides by comprehensive two-dimensional gas chromatography. *Journal of Chromatography A* **1487**, 248-253 (2017).
 - 301 Eisenberg Jr, F. & Bolden, A. H. Gas chromatography of sugar phosphates and sugar nucleotides. *Analytical Biochemistry* **29**, 284-292 (1969).
 - 302 Duangkumpha, K. *et al.* GC× GC-TOFMS metabolomics analysis identifies elevated levels of plasma sugars and sugar alcohols in diabetic mellitus patients with kidney failure. *Journal of Biological Chemistry* **298** (2022).
 - 303 Misra, B. B. Advances in high resolution GC-MS technology: a focus on the application of GC-Orbitrap-MS in metabolomics and exposomics for FAIR practices. *Analytical Methods* **13**, 2265-2282 (2021).
 - 304 Adebo, O. A. *et al.* Application of gas chromatography-mass spectrometry (GC-MS)-based metabolomics for the study of fermented cereal and legume foods: A review. *International Journal of Food Science & Technology* **56**, 1514-1534 (2021).
 - 305 Wang, X. *et al.* Improvement of simultaneous determination of neutral monosaccharides and uronic acids by gas chromatography. *Food chemistry* **220**, 198-207 (2017).
 - 306 De, P. *et al.* Estimation of D-arabinose by gas chromatography/mass spectrometry as surrogate for mycobacterial liparabinomannan in human urine. *PLoS One* **10**, e0144088 (2015).
 - 307 Mairinger, T., Weiner, M., Hann, S. & Troyer, C. Selective and accurate quantification of N-acetylglucosamine in biotechnological cell samples via GC-MS/MS and GC-TOFMS. *Analytical chemistry* **92**, 4875-4883 (2020).
 - 308 Millette, P. G., Chabot, J., Sheppard, D. C. & Le Mauff, F. Identification and Quantification of Monosaccharides from Fungal Cell Walls and Exopolysaccharides by Gas Chromatography Coupled to Mass Spectrometry. *Current Protocols* **3**, e853 (2023).
 - 309 Zweckmair, T., Schiehsner, S., Rosenau, T. & Potthast, A. Improved quantification of monosaccharides in complex lignocellulosic biomass matrices: A gas chromatography-mass spectrometry based approach. *Carbohydrate research* **446**, 7-12 (2017).
 - 310 Niemi, C., Takahashi, J., Gorzsás, A. & Gentili, F. G. Quantitative and qualitative saccharide analysis of North Atlantic brown seaweed by gas chromatography/mass spectrometry and infrared spectroscopy. *International Journal of Biological Macromolecules* **254**, 127870 (2024).
 - 311 Leker, K., Lozano-Pope, I., Bandyopadhyay, K., Choudhury, B. P. & Obonyo, M. Comparison of lipopolysaccharides composition of two different strains of *Helicobacter pylori*. *BMC microbiology* **17**, 1-9 (2017).
 - 312 Saad, M. H., Sidkey, N. M. & El-Fakharany, E. M. Identification and statistical optimization of a novel alginate polymer extracted from newly isolated *Synechocystis algin* MNE ON864447 with antibacterial activity. *Microbial Cell Factories* **22**, 229 (2023).
 - 313 Ramautar, R. & Chen, D. D. *Capillary Electrophoresis-Mass Spectrometry for Proteomics and Metabolomics: Principles and Applications*. (John Wiley & Sons, 2022).
 - 314 Ramautar, R., Somsen, G. W. & de Jong, G. J. CE-MS in metabolomics. *Electrophoresis* **30**, 276-291 (2009).
 - 315 Zhang, W. & Ramautar, R. CE-MS for metabolomics: Developments and applications in the period 2018–2020. *Electrophoresis* **42**, 381-401 (2021).
 - 316 Zaia, J. Capillary electrophoresis-mass spectrometry of carbohydrates. *Capillary Electrophoresis of Biomolecules: Methods and Protocols*, 13-25 (2013).
 - 317 do Lago, C. L., Daniel, D., Lopes, F. S. & Cieslarova, Z. In *Carbohydrate Analysis by Modern Liquid Phase Separation Techniques* 443-484 (Elsevier, 2021).
 - 318 Jorgenson, J. W. & Lukacs, K. D. Capillary zone electrophoresis. *Science* **222**, 266-272 (1983).
 - 319 Voeten, R. L., Ventouri, I. K., Haselberg, R. & Somsen, G. W. Capillary electrophoresis: trends and recent advances. *Analytical chemistry* **90**, 1464-1481 (2018).
 - 320 Monnig, C. A. & Kennedy, R. T. Capillary electrophoresis. *Analytical chemistry* **66**, 280-314 (1994).

- 321 Harstad, R. K., Johnson, A. C., Weisenberger, M. M. & Bowser, M. T. Capillary electrophoresis. *Analytical chemistry* **88**, 299-319 (2016).
- 322 Lindhardt, R. J. & Pervin, A. Separation of negatively charged carbohydrates by capillary electrophoresis. *Journal of Chromatography A* **720**, 323-335 (1996).
- 323 Mantovani, V., Galeotti, F., Maccari, F. & Volpi, N. Recent advances in capillary electrophoresis separation of monosaccharides, oligosaccharides, and polysaccharides. *Electrophoresis* **39**, 179-189 (2018).
- 324 Ramautar, R., Mayboroda, O. A., Somsen, G. W. & de Jong, G. J. CE-MS for metabolomics: Developments and applications in the period 2008–2010. *Electrophoresis* **32**, 52-65 (2011).
- 325 Ramautar, R., Somsen, G. W. & de Jong, G. J. CE-MS for metabolomics: Developments and applications in the period 2010–2012. *Electrophoresis* **34**, 86-98 (2013).
- 326 Ramautar, R., Somsen, G. W. & de Jong, G. J. CE-MS for metabolomics: developments and applications in the period 2012–2014. *Electrophoresis* **36**, 212-224 (2015).
- 327 Ramautar, R., Somsen, G. W. & de Jong, G. J. CE-MS for metabolomics: developments and applications in the period 2014–2016. *Electrophoresis* **38**, 190-202 (2017).
- 328 Ramautar, R., Somsen, G. W. & de Jong, G. J. CE-MS for metabolomics: Developments and applications in the period 2016–2018. *Electrophoresis* **40**, 165-179 (2019).
- 329 Callewaert, N., Geysens, S., Molemans, F. & Contreras, R. Ultrasensitive profiling and sequencing of N-linked oligosaccharides using standard DNA-sequencing equipment. *Glycobiology* **11**, 275-281 (2001).
- 330 Ruhaak, L. R. *et al.* Optimized workflow for preparation of APTS-labeled N-glycans allowing high-throughput analysis of human plasma glycomics using 48-channel multiplexed CGE-LIF. *Journal of proteome research* **9**, 6655-6664 (2010).
- 331 Khatiri, K. *et al.* Microfluidic capillary electrophoresis–mass spectrometry for analysis of monosaccharides, oligosaccharides, and glycopeptides. *Analytical chemistry* **89**, 6645-6655 (2017).
- 332 Fang, P., Pan, J.-Z. & Fang, Q. A robust and extendable sheath flow interface with minimal dead volume for coupling CE with ESI-MS. *Talanta* **180**, 376-382 (2018).
- 333 Höcker, O., Montealegre, C. & Neusüß, C. Characterization of a nanoflow sheath liquid interface and comparison to a sheath liquid and a sheathless porous-tip interface for CE-ESI-MS in positive and negative ionization. *Analytical and bioanalytical chemistry* **410**, 5265-5275 (2018).
- 334 Sauer, F., Sydow, C. & Trapp, O. A robust sheath-flow CE-MS interface for hyphenation with Orbitrap MS. *Electrophoresis* **41**, 1280-1286 (2020).
- 335 Fanali, S., D'Orazio, G., Foret, F., Kleparnik, K. & Aturki, Z. On-line CE-MS using pressurized liquid junction nanoflow electrospray interface and surface-coated capillaries. *Electrophoresis* **27**, 4666-4673 (2006).
- 336 Huhn, C., Ramautar, R., Wuhler, M. & Somsen, G. Relevance and use of capillary coatings in capillary electrophoresis–mass spectrometry. *Analytical and bioanalytical chemistry* **396**, 297-314 (2010).
- 337 Mellors, J., Gorbounov, V., Ramsey, R. & Ramsey, J. Fully integrated glass microfluidic device for performing high-efficiency capillary electrophoresis and electrospray ionization mass spectrometry. *Analytical chemistry* **80**, 6881-6887 (2008).
- 338 Sun, B. *et al.* Analytical method development for directed enzyme evolution research: a high throughput high-performance liquid chromatography method for analysis of ribose and ribitol and a capillary electrophoresis method for the separation of ribose enantiomers. *Journal of Chromatography A* **1271**, 163-169 (2013).
- 339 Abo, M., He, L.-p., Sato, K. & Okubo, A. Determination of monosaccharides derivatized with 2-aminobenzoic acid by capillary electrophoresis. *Capillary Electrophoresis of Biomolecules: Methods and Protocols*, 45-50 (2013).
- 340 Guttman, A. Analysis of monosaccharide composition by capillary electrophoresis. *Journal of Chromatography A* **763**, 271-277 (1997).
- 341 Szabo, Z., Bones, J., Guttman, A., Glick, J. & Karger, B. L. Sialic acid speciation using capillary electrophoresis: optimization of analyte derivatization and separation. *Analytical chemistry* **84**, 7638-7642 (2012).
- 342 Schmitt-Kopplin, P., Fischer, K., Freitag, D. & Kettrup, A. Capillary electrophoresis for the simultaneous separation of selected carboxylated carbohydrates and their related 1, 4-lactones. *Journal of Chromatography A* **807**, 89-100 (1998).
- 343 Kodama, S., Aizawa, S. i., Taga, A., Yamashita, T. & Yamamoto, A. Chiral resolution of monosaccharides as 1-phenyl-3-methyl-5-pyrazolone derivatives by ligand-exchange CE using borate anion as a central ion of the chiral selector. *Electrophoresis* **27**, 4730-4734 (2006).
- 344 Kodama, S. *et al.* Simultaneous chiral resolution of monosaccharides as 8-aminonaphthalene-1, 3, 6-trisulfonate derivatives by ligand-exchange CE using borate as a central ion of the chiral selector. *Electrophoresis* **28**, 3930-3933 (2007).
- 345 Clark, C., Somsen, G. W. & Kohler, I. Chiral Separation by Capillary Electrophoresis and Capillary Electrophoresis–Mass Spectrometry: Fundamentals, Recent Developments, and Applications. *Chiral Separations and Stereochemical Elucidation: Fundamentals, Methods, and Applications*, 103-142 (2023).
- 346 Lageveen-Kammeijer, G. S. *et al.* Highly sensitive CE-ESI-MS analysis of N-glycans from complex biological samples. *Nature communications* **10**, 2137 (2019).
- 347 Campa, C., Baiutti, E. & Flamigni, A. Capillary electrophoresis of sugar acids. *Capillary Electrophoresis: Methods and Protocols*, 307-355 (2008).
- 348 Ortner, K. & Buchberger, W. Determination of sialic acids released from glycoproteins using capillary zone electrophoresis/electrospray ionization mass spectrometry. *Electrophoresis* **29**, 2233-2237 (2008).
- 349 Klampfl, C. W. & Buchberger, W. Determination of carbohydrates by capillary electrophoresis with electrospray-mass spectrometric detection. *Electrophoresis* **22**, 2737-2742 (2001).
- 350 Daniel, D., Lopes, F. S., Dos Santos, V. B. & do Lago, C. L. Detection of coffee adulteration with soybean and corn by capillary electrophoresis-tandem mass spectrometry. *Food chemistry* **243**, 305-310 (2018).
- 351 van Mever, M., Hankemeier, T. & Ramautar, R. CE-MS for anionic metabolic profiling: an overview of methodological developments. *Electrophoresis* **40**, 2349-2359 (2019).
- 352 Zhang, W., Guled, F., Hankemeier, T. & Ramautar, R. Profiling nucleotides in low numbers of mammalian cells by sheathless CE-MS in positive ion mode: circumventing corona discharge. *Electrophoresis* **41**, 360-369 (2020).
- 353 Bucsella, B., Fornage, A., Le Denmat, C. & Kálmán, F. Nucleotide and Nucleotide Sugar Analysis in Cell Extracts by Capillary Electrophoresis: FH-HES Universities of Applied Sciences. *Chimia* **70**, 732-732 (2016).
- 354 Lehmann, R. *et al.* Simultaneous, quantitative analysis of UDP-N-acetylglucosamine, UDP-N-

- acetylgalactosamine, UDP-glucose and UDP-galactose in human peripheral blood cells, muscle biopsies and cultured mesangial cells by capillary zone electrophoresis. *ELECTROPHORESIS: An International Journal* **21**, 3010-3015 (2000).
- 355 Feng, H.-T., Wong, N., Wee, S. & Lee, M. M. Simultaneous determination of 19 intracellular nucleotides and nucleotide sugars in Chinese Hamster ovary cells by capillary electrophoresis. *Journal of Chromatography B* **870**, 131-134 (2008).
- 356 Hasunuma, T. *et al.* Metabolic turnover analysis by a combination of in vivo ^{13}C -labelling from $^{13}\text{CO}_2$ and metabolic profiling with CE-MS/MS reveals rate-limiting steps of the C3 photosynthetic pathway in *Nicotiana tabacum* leaves. *Journal of Experimental Botany* **61**, 1041-1051 (2010).
- 357 Soo, E. C. *et al.* Selective detection and identification of sugar nucleotides by CE- electrospray-MS and its application to bacterial metabolomics. *Analytical chemistry* **76**, 619-626 (2004).
- 358 Karas, M. & Krüger, R. Ion formation in MALDI: the cluster ionization mechanism. *Chemical reviews* **103**, 427-440 (2003).
- 359 Knochenmuss, R. & Zenobi, R. MALDI ionization: the role of in-plume processes. *Chemical reviews* **103**, 441-452 (2003).
- 360 Hillenkamp, F. & Peter-Katalinic, J. *MALDI MS: a practical guide to instrumentation, methods and applications*. (John Wiley & Sons, 2013).
- 361 El-Aneel, A., Cohen, A. & Banoub, J. Mass spectrometry, review of the basics: electrospray, MALDI, and commonly used mass analyzers. *Applied spectroscopy reviews* **44**, 210-230 (2009).
- 362 Dreisewerd, K. Recent methodological advances in MALDI mass spectrometry. *Analytical and bioanalytical chemistry* **406**, 2261-2278 (2014).
- 363 Zenobi, R. & Knochenmuss, R. Ion formation in MALDI mass spectrometry. *Mass spectrometry reviews* **17**, 337-366 (1998).
- 364 Lu, I.-C., Lee, C., Lee, Y.-T. & Ni, C.-K. Ionization mechanism of matrix-assisted laser desorption/ionization. *Annual Review of Analytical Chemistry* **8**, 21-39 (2015).
- 365 Harvey, D. J. Analysis of carbohydrates and glycoconjugates by matrix-assisted laser desorption/ionization mass spectrometry: An update covering the period 1999–2000. *Mass spectrometry reviews* **25**, 595-662 (2006).
- 366 Lee, C. *et al.* Formation of metal-related ions in matrix-assisted laser desorption ionization. *Journal of The American Society for Mass Spectrometry* **27**, 1491-1498 (2016).
- 367 Chen, J. L. *et al.* Theoretical investigation of low detection sensitivity for underivatized carbohydrates in ESI and MALDI. *Journal of Mass Spectrometry* **51**, 1180-1186 (2016).
- 368 Zaia, J. Mass spectrometry of oligosaccharides. *Mass spectrometry reviews* **23**, 161-227 (2004).
- 369 Harvey, D. Quantitative aspects of the matrix-assisted laser desorption mass spectrometry of complex oligosaccharides. *Rapid Communications in Mass Spectrometry* **7**, 614-619 (1993).
- 370 Batoy, S., Akhmetova, E., Miladinovic, S., Smeal, J. & Wilkins, C. Developments in MALDI mass spectrometry: the quest for the perfect matrix. *Applied Spectroscopy Reviews* **43**, 485-550 (2008).
- 371 Dreisewerd, K. The desorption process in MALDI. *Chemical reviews* **103**, 395-426 (2003).
- 372 McDonnell, L. A. & Heeren, R. M. Imaging mass spectrometry. *Mass spectrometry reviews* **26**, 606-643 (2007).
- 373 Gessel, M. M., Norris, J. L. & Caprioli, R. M. MALDI imaging mass spectrometry: spatial molecular analysis to enable a new age of discovery. *Journal of proteomics* **107**, 71-82 (2014).
- 374 van Hove, E. R. A., Smith, D. F. & Heeren, R. M. A concise review of mass spectrometry imaging. *Journal of chromatography A* **1217**, 3946-3954 (2010).
- 375 Rzagalinski, I. & Volmer, D. A. Quantification of low molecular weight compounds by MALDI imaging mass spectrometry—A tutorial review. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **1865**, 726-739 (2017).
- 376 Seeley, E. H. & Caprioli, R. M. MALDI imaging mass spectrometry of human tissue: method challenges and clinical perspectives. *Trends in biotechnology* **29**, 136-143 (2011).
- 377 Ucal, Y. *et al.* Clinical applications of MALDI imaging technologies in cancer and neurodegenerative diseases. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **1865**, 795-816 (2017).
- 378 Cazares, L. H., Troyer, D. A., Wang, B., Drake, R. R. & John Semmes, O. MALDI tissue imaging: from biomarker discovery to clinical applications. *Analytical and bioanalytical chemistry* **401**, 17-27 (2011).
- 379 O'Rourke, M. B., Djordjevic, S. P. & Padula, M. P. The quest for improved reproducibility in MALDI mass spectrometry. *Mass Spectrometry Reviews* **37**, 217-228 (2018).
- 380 Haslam, C. *et al.* The evolution of MALDI-TOF mass spectrometry toward ultra-high-throughput screening: 1536-well format and beyond. *Journal of biomolecular screening* **21**, 176-186 (2016).
- 381 Urban, P. L. *et al.* High-density micro-arrays for mass spectrometry. *Lab on a Chip* **10**, 3206-3209 (2010).
- 382 Pabst, M. *et al.* Self-aliquoting microarray plates for accurate quantitative matrix-assisted laser desorption/ionization mass spectrometry. *Analytical chemistry* **85**, 9771-9776 (2013).
- 383 Yukihira, D., Miura, D., Saito, K., Takahashi, K. & Wariishi, H. MALDI- MS-based high-throughput metabolite analysis for intracellular metabolic dynamics. *Analytical chemistry* **82**, 4278-4282 (2010).
- 384 Beavis, R., Chaudhary, T. & Chait, B. α -Cyano-4-hydroxycinnamic acid as a matrix for matrix-assisted laser desorption mass spectrometry. *Organic mass spectrometry* **27**, 156-158 (1992).
- 385 Bourcier, S., Bouchonnet, S. & Hoppilliard, Y. Ionization of 2, 5-dihydroxybenzoic acid (DHB) matrix-assisted laser desorption ionization experiments and theoretical study. *International Journal of Mass Spectrometry* **210**, 59-69 (2001).
- 386 Abdelhamid, H. N. Organic matrices, ionic liquids, and organic matrices@ nanoparticles assisted laser desorption/ionization mass spectrometry. *TrAC Trends in Analytical Chemistry* **89**, 68-98 (2017).
- 387 Harvey, D. J. Analysis of carbohydrates and glycoconjugates by matrix-assisted laser desorption/ionization mass spectrometry: An update for 2013–2014. *Mass Spectrometry Reviews* **37**, 353-491 (2018).
<https://doi.org/https://doi.org/10.1002/mas.21530>
- 388 Harvey, D. J. Analysis of carbohydrates and glycoconjugates by matrix-assisted laser desorption/ionization mass spectrometry: An update for

- 2015–2016. *Mass Spectrometry Reviews* **40**, 408–565 (2021).
- 389 Harvey, D. J. Analysis of carbohydrates and glycoconjugates by matrix-assisted laser desorption/ionization mass spectrometry: An update for 2017–2018. *Mass Spectrometry Reviews* **42**, 227–431 (2023).
<https://doi.org/https://doi.org/10.1002/mas.21721>
- 390 Harvey, D. J. Analysis of carbohydrates and glycoconjugates by matrix-assisted laser desorption/ionization mass spectrometry: An update covering the period 2001–2002. *Mass spectrometry reviews* **27**, 125–201 (2008).
- 391 Harvey, D. J. Analysis of carbohydrates and glycoconjugates by matrix-assisted laser desorption/ionization mass spectrometry: An update for 2003–2004. *Mass spectrometry reviews* **28**, 273–361 (2009).
- 392 Harvey, D. J. Analysis of carbohydrates and glycoconjugates by matrix-assisted laser desorption/ionization mass spectrometry: An update for the period 2005–2006. *Mass Spectrometry Reviews* **30**, 1–100 (2011).
- 393 Harvey, D. J. Analysis of carbohydrates and glycoconjugates by matrix-assisted laser desorption/ionization mass spectrometry: An update for 2007–2008. *Mass spectrometry reviews* **31**, 183–311 (2012).
- 394 Harvey, D. J. Analysis of carbohydrates and glycoconjugates by matrix-assisted laser desorption/ionization mass spectrometry: An update for 2009–2010. *Mass spectrometry reviews* **34**, 268–422 (2015).
- 395 Harvey, D. J. Analysis of carbohydrates and glycoconjugates by matrix-assisted laser desorption/ionization mass spectrometry: An update covering the period 2011–2012. *Mass spectrometry reviews* **36**, 255–422 (2015).
- 396 Harvey, D. J. Analysis of carbohydrates and glycoconjugates by matrix-assisted laser desorption/ionization mass spectrometry: An update for 2013–2014. *Mass Spectrometry Reviews* **37**, 353–491 (2018).
- 397 Qiao, Z. & Lissel, F. MALDI matrices for the analysis of low molecular weight compounds: rational design, challenges and perspectives. *Chemistry–An Asian Journal* **16**, 868–878 (2021).
- 398 Calvano, C. D., Monopoli, A., Cataldi, T. R. & Palmisano, F. MALDI matrices for low molecular weight compounds: an endless story? *Analytical and bioanalytical chemistry* **410**, 4015–4038 (2018).
- 399 Pabst, M. *et al.* Quantification of saquinavir from lysates of peripheral blood mononuclear cells using microarrays and standard MALDI-TOF-MS. *Journal of The American Society for Mass Spectrometry* **25**, 1083–1086 (2014).
- 400 Krismer, J. *et al.* A MALDI-MS methodology for studying metabolic heterogeneity of single cells in a population. *Single Cell Metabolism: Methods and Protocols*, 113–124 (2020).
- 401 Wang, H.-Y., Chu, X., Zhao, Z.-X., He, X.-S. & Guo, Y.-L. Analysis of low molecular weight compounds by MALDI-FTICR-MS. *Journal of Chromatography B* **879**, 1166–1179 (2011).
- 402 Armstrong, D. W., Zhang, L.-K., He, L. & Gross, M. L. Ionic liquids as matrixes for matrix-assisted laser desorption/ionization mass spectrometry. *Analytical chemistry* **73**, 3679–3686 (2001).
- 403 Schmidt De León, T., Salum, M. L. & Erra-Balsells, R. norHarmane containing ionic liquid matrices for low molecular weight MALDI-MS carbohydrate analysis: The perfect couple with α -cyano-4-hydroxycinnamic acid. *Journal of mass spectrometry* **54**, 643–654 (2019).
- 404 Horatz, K. *et al.* Conjugated polymers as a new class of dual-mode matrices for MALDI mass spectrometry and imaging. *Journal of the American Chemical Society* **140**, 11416–11423 (2018).
- 405 Ayorinde, F. O., Bezabeh, D. Z. & Delves, I. Preliminary investigation of the simultaneous detection of sugars, ascorbic acid, citric acid, and sodium benzoate in non-alcoholic beverages by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid communications in mass spectrometry* **17**, 1735–1742 (2003).
- 406 Mei-Hua, D., Shi-Mei, S. & Biao, J. Application of modified mesoporous silica materials for analysis of small molecules by MALDI-TOF-MS. *Chinese Journal of Analytical Chemistry* **42**, 1235–1239 (2014).
- 407 Zhang, L.-K. & Gross, M. L. Location of abasic sites in oligodeoxynucleotides by tandem mass spectrometry and by a chemical cleavage initiated by an unusual reaction of the ODN with MALDI matrix. *Journal of the American Society for Mass Spectrometry* **13**, 1418–1426 (2002).
- 408 Ling, L. *et al.* 2-Phenyl-3-(p-aminophenyl) acrylonitrile: a reactive matrix for sensitive and selective analysis of glycans by MALDI-MS. *Analytical chemistry* **91**, 8801–8807 (2019).
- 409 Han, Y. *et al.* On-tissue derivatization for isomer-specific mass spectrometry imaging and relative quantification of monosaccharides in biological tissues. *Analytica Chimica Acta* **1225**, 340241 (2022).
- 410 Vermillion-Salsbury, R. L. & Hercules, D. M. 9-Aminoacridine as a matrix for negative mode matrix-assisted laser desorption/ionization. *Rapid communications in mass spectrometry* **16**, 1575–1581 (2002).
- 411 Teearu, A. *et al.* Method development for the analysis of resinous materials with MALDI-FT-ICR-MS: novel internal standards and a new matrix material for negative ion mode. *Journal of Mass Spectrometry* **52**, 603–617 (2017).
- 412 Shroff, R., Muck, A. & Svatoš, A. Analysis of low molecular weight acids by negative mode matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry: An International Journal Devoted to the Rapid Dissemination of Up-to-the-Minute Research in Mass Spectrometry* **21**, 3295–3300 (2007).
- 413 Ibáñez, A. J. *et al.* Mass spectrometry-based metabolomics of single yeast cells. *Proceedings of the National Academy of Sciences* **110**, 8790–8794 (2013).
- 414 Karst, D. J. *et al.* Intracellular CHO cell metabolite profiling reveals steady-state dependent metabolic fingerprints in perfusion culture. *Biotechnology progress* **33**, 879–890 (2017).
- 415 Heinrich, M., Schäffer, C., Messner, P. & Allmaier, G. Negative ion ultraviolet matrix-assisted laser desorption ionization mass spectrometry and post source decay of glycosyl esters of nucleoside pyrophosphates. *European Journal of Mass Spectrometry* **14**, 401–409 (2008).
- 416 Steinhoff, R. F. *et al.* Rapid estimation of the energy charge from cell lysates using matrix-assisted laser desorption/ionization mass spectrometry: Role of in-source fragmentation. *Analytical biochemistry* **447**, 107–113 (2014).

- 417 Karst, D. J. *et al.* Isotope labeling to determine the dynamics of metabolic response in CHO cell perfusion bioreactors using MALDI-TOF-MS. *Biotechnology progress* **33**, 1630-1639 (2017).
- 418 Villiger, T. K. *et al.* High-throughput profiling of nucleotides and nucleotide sugars to evaluate their impact on antibody N-glycosylation. *Journal of biotechnology* **229**, 3-12 (2016).
- 419 Sato, Y. *et al.* Matrix-assisted ultraviolet laser desorption/ionization time-of-flight (UV-MALDI-TOF) mass spectra of N-acylated and N, O-acylated glycosylamines. *Carbohydrate research* **342**, 2567-2574 (2007).
- 420 Hu, R., Li, Y., Yang, Y. & Liu, M. Mass spectrometry-based strategies for single-cell metabolomics. *Mass Spectrometry Reviews* **42**, 67-94 (2023).
- 421 Zenobi, R. Single-cell metabolomics: analytical and biological perspectives. *Science* **342**, 1243259 (2013).
- 422 Guillaume-Gentil, O. *et al.* Single-cell mass spectrometry of metabolites extracted from live cells by fluidic force microscopy. *Analytical Chemistry* **89**, 5017-5023 (2017).
- 423 Gouw, J. W., Burgers, P. C., Trikoupi, M. A. & Terlouw, J. K. Derivatization of small oligosaccharides prior to analysis by matrix-assisted laser desorption/ionization using glycidyltrimethylammonium chloride and Girard's reagent T. *Rapid communications in mass spectrometry* **16**, 905-912 (2002).
- 424 Wu, Y.-C., Zhang, X.-W., Huang, Y.-C. & Lu, I.-C. Advancing carbohydrate quantification in MALDI mass spectrometry by the rapidly freeze-drying droplet (RFDD) method. *Analyst* **149**, 1766-1773 (2024).
- 425 Liang, C.-W., Chang, P.-J., Lin, Y.-J., Lee, Y.-T. & Ni, C.-K. High ion yields of carbohydrates from frozen solution by UV-MALDI. *Analytical chemistry* **84**, 3493-3499 (2012).
- 426 Wei, J., Buriak, J. M. & Siuzdak, G. Desorption-ionization mass spectrometry on porous silicon. *Nature* **399**, 243-246 (1999).
- 427 Djambazova, K. V. *et al.* MALDI TIMS IMS of Disialoganglioside Isomers—GD1a and GD1b in Murine Brain Tissue. *Analytical Chemistry* **95**, 1176-1183 (2022).
- 428 Kuik, C., van Hoogstraten, S. W., Arts, J. J., Honing, M. & Cillero-Pastor, B. Matrix-assisted laser desorption/ionization mass spectrometry imaging for quorum sensing. *AMB Express* **14**, 45 (2024).
- 429 Feucherolles, M. & Frache, G. MALDI mass spectrometry imaging: a potential game-changer in a modern microbiology. *Cells* **11**, 3900 (2022).
- 430 Kiss, A. & Hopfgartner, G. Laser-based methods for the analysis of low molecular weight compounds in biological matrices. *Methods* **104**, 142-153 (2016).
- 431 Ruiz-Capillas, C., Herrero, A. M. & Jiménez-Colmenero, F. Flow injection analysis-tandem mass spectrometry. *Fingerprinting Techniques in Food Authentication and Traceability*; Siddiqi, KS, Nollet, LML, Eds, 79-95 (2018).
- 432 Nanita, S. C. & Kaldon, L. G. Emerging flow injection mass spectrometry methods for high-throughput quantitative analysis. *Analytical and bioanalytical chemistry* **408**, 23-33 (2016).
- 433 Draper, J., Lloyd, A. J., Goodacre, R. & Beckmann, M. Flow infusion electrospray ionisation mass spectrometry for high throughput, non-targeted metabolite fingerprinting: a review. *Metabolomics* **9**, 4-29 (2013).
- 434 González-Domínguez, R., Sayago, A. & Fernández-Recamales, Á. Direct infusion mass spectrometry for metabolomic phenotyping of diseases. *Bioanalysis* **9**, 131-148 (2017).
- 435 Taki, K. *et al.* A preliminary study of rapid-fire high-throughput metabolite analysis using nano-flow injection/Q-TOFMS. *Analytical and Bioanalytical Chemistry* **412**, 4127-4134 (2020).
- 436 Nagy, G. & Pohl, N. L. Complete hexose isomer identification with mass spectrometry. *Journal of the American Society for Mass Spectrometry* **26**, 677-685 (2015).
- 437 Nagy, G. & Pohl, N. L. Monosaccharide identification as a first step toward de novo carbohydrate sequencing: mass spectrometry strategy for the identification and differentiation of diastereomeric and enantiomeric pentose isomers. *Analytical chemistry* **87**, 4566-4571 (2015).
- 438 Wooke, Z., Nagy, G., Barnes, L. F. & Pohl, N. L. B. Development of a Post-Column Liquid Chromatographic Chiral Addition Method for the Separation and Resolution of Common Mammalian Monosaccharides. *J Am Soc Mass Spectrom* **1**;30(3):419-25 (2019).
- 439 Giorgi, G. & Speranza, M. Stereoselective noncovalent interactions of monosaccharides with hydrazine. *International Journal of Mass Spectrometry* **249**, 112-119 (2006).
- 440 Madhusudanan, K. & Srivastava, M. Identification of hexose diastereomers by means of tandem mass spectrometry of oxocarbenium ions followed by neural networks analysis. *Journal of mass spectrometry* **43**, 126-131 (2008).
- 441 Burnum-Johnson, K. E. *et al.* in *Ion Mobility Spectrometry and the Omics: Distinguishing Isomers, Molecular Classes and Contaminant Ions in Complex Samples*. *Trends Anal Chem* TRAC Vol. 116 292-299 (2019).
- 442 Te Brinke, E., Arrizabalaga-Larrañaga, A. & Blokland, M. H. Insights of ion mobility spectrometry and its application on food safety and authenticity: A review. *Analytica chimica acta* **1222**, 340039 (2022).
- 443 Christofi, E. & Barran, P. Ion mobility mass spectrometry (IM-MS) for structural biology: insights gained by measuring mass, charge, and collision cross section. *Chemical Reviews* **123**, 2902-2949 (2023).
- 444 Ibrahim, Y. M. *et al.* New frontiers for mass spectrometry based upon structures for lossless ion manipulations. *Analyst* **142**, 1010-1021 (2017).
- 445 Ortmayr, K., Causon, T. J., Hann, S. & Koellensperger, G. Increasing selectivity and coverage in LC-MS based metabolome analysis. *TrAC Trends in Analytical Chemistry* **82**, 358-366 (2016).
- 446 Causon, T. J., Kurulugama, R. T. & Hann, S. Drift-tube ion mobility-mass spectrometry for nontargeted' omics. *Ion Mobility-Mass Spectrometry: Methods and Protocols*, 79-94 (2020).
- 447 Zhu, M., Bendiak, B., Clowers, B. & Hill, H. H. Ion mobility-mass spectrometry analysis of isomeric carbohydrate precursor ions. *Anal Bioanal Chem* **1**;394(7):1853-67 (2009).
- 448 Mu, Y., Schulz, B. L. & Ferro, V. Applications of Ion Mobility-Mass Spectrometry in Carbohydrate Chemistry and Glycobiology. *Molecules* **Oct**;23(10):2557 (2018).
- 449 Hofmann, J. & Pagel, K. Glycan Analysis by Ion Mobility-Mass Spectrometry. *Angew Chem Int Ed* **10**;56(29):8342-9 (2017).
- 450 Both, P. *et al.* Discrimination of epimeric glycans and glycopeptides using IM-MS and its potential for carbohydrate sequencing. *Nat Chem Jan*;6(1):65-74 (2014).
- 451 Gabryelski, W. & Froese, K. L. Rapid and sensitive differentiation of anomers, linkage, and position isomers of disaccharides using High-Field Asymmetric Waveform

- Ion Mobility Spectrometry (FAIMS). *J Am Soc Mass Spectrom* **1**;14(3):265–77 (2003).
- 452 Morris, C. B., Poland, J. C., May, J. C. & McLean, J. A. Fundamentals of Ion Mobility-Mass Spectrometry for the Analysis of Biomolecules. *Methods Mol Biol Clifton NJ* **2084**, 1–31 (2020).
 - 453 Zhou, Z., Tu, J. & Zhu, Z. J. Advancing the large-scale CCS database for metabolomics and lipidomics at the machine-learning era. *Curr Opin Chem Biol* **42**, 34–41 (2018).
 - 454 Pringle, S. D. *et al.* An investigation of the mobility separation of some peptide and protein ions using a new hybrid quadrupole/travelling wave IMS/oa-ToF instrument. *International Journal of Mass Spectrometry* **261**, 1–12 (2007).
 - 455 May, J. C. *et al.* Conformational ordering of biomolecules in the gas phase: nitrogen collision cross sections measured on a prototype high resolution drift tube ion mobility-mass spectrometer. *Analytical chemistry* **86**, 2107–2116 (2014).
 - 456 Groessel, M., Graf, S. & Knochenmuss, R. High resolution ion mobility-mass spectrometry for separation and identification of isomeric lipids. *Analyst* **140**, 6904–6911 (2015).
 - 457 Michelmann, K., Silveira, J. A., Ridgeway, M. E. & Park, M. A. Fundamentals of trapped ion mobility spectrometry. *Journal of the American Society for Mass Spectrometry* **26**, 14–24 (2014).
 - 458 Guevremont, R. High-field asymmetric waveform ion mobility spectrometry: a new tool for mass spectrometry. *Journal of Chromatography A* **1058**, 3–19 (2004).
 - 459 Kolakowski, B. M. & Mester, Z. Review of applications of high-field asymmetric waveform ion mobility spectrometry (FAIMS) and differential mobility spectrometry (DMS). *Analyst* **132**, 842–864 (2007).
 - 460 Dodds, J. N. & Baker, E. S. Ion Mobility Spectrometry: Fundamental Concepts, Instrumentation, Applications, and the Road Ahead. *J Am Soc Mass Spectrom* (2019).
 - 461 Cumeras, R., Figueras, E., Davis, C., Baumbach, J. I. & Gracia, I. Review on ion mobility spectrometry. Part 1: current instrumentation. *Analyst* **140**, 1376–1390 (2015).
 - 462 Olajide, O. E., Kartowikromo, K. Y. & Hamid, A. M. in *Electron Microscopes, Spectroscopy and Their Applications* (IntechOpen, 2023).
 - 463 Naylor, C. N. & Nagy, G. Recent advances in high-resolution traveling wave-based ion mobility separations coupled to mass spectrometry. *Mass Spectrometry Reviews* (2024).
 - 464 Delafield, D. G., Lu, G., Kaminsky, C. J. & Li, L. High-end ion mobility mass spectrometry: A current review of analytical capacity in omics applications and structural investigations. *TrAC Trends Anal Chem* **157** (2022).
 - 465 Liu, Y. & Clemmer, D. E. Characterizing oligosaccharides using injected-ion mobility/mass spectrometry. *Anal Chem* **1**;69(13):2504–9 (1997).
 - 466 Clowers, B. H., Dwivedi, P., Steiner, W. E., Hill, H. H. & Bendiak, B. Separation of sodiated isobaric disaccharides and trisaccharides using electrospray ionization-atmospheric pressure ion mobility-time of flight mass spectrometry. *J Am Soc Mass Spectrom* **1**;16(5):660–9 (2005).
 - 467 Morrison, K. A. & Clowers, B. H. Contemporary glycomic approaches using ion mobility-mass spectrometry. *Current opinion in chemical biology* **42**, 119–129 (2018).
 - 468 Li, H., Bendiak, B., Siems, W. F., Gang, D. R. & Hill, H. H. Carbohydrate structure characterization by tandem ion mobility mass spectrometry (IMMS)2. *Anal Chem* **5**;85(5):2760–9 (2013).
 - 469 Harvey, D. J. Negative Ion Mass Spectrometry for the Analysis of N-Linked Glycans. *Mass Spectrom Rev* **39**, 586–679 (2020).
 - 470 Seo, Y., Andaya, A. & Leary, J. A. Preparation, separation, and conformational analysis of differentially sulfated heparin octasaccharide isomers using ion mobility mass spectrometry. *Anal Chem* **6**;84(5):2416–23 (2012).
 - 471 Manabe, N. *et al.* A Data Set of Ion Mobility Collision Cross Sections and Liquid Chromatography Retention Times from 71 Pyridylaminated N-Linked Oligosaccharides. *J Am Soc Mass Spectrom* **7**;33(9):1772–83 (2022).
 - 472 Bohrer, B. C. & Clemmer, D. E. Biologically-inspired peptide reagents for enhancing IMS-MS analysis of carbohydrates. *J Am Soc Mass Spectrom* **22**(9):1602–9 (2011).
 - 473 Huang, Y. & Dodds, E. D. Ion mobility studies of carbohydrates as group I adducts: isomer specific collisional cross section dependence on metal ion radius. *Anal Chem* **15**;85(20):9728–35 (2013).
 - 474 Fenn, L. S. & McLean, J. A. Structural resolution of carbohydrate positional and structural isomers based on gas-phase ion mobility-mass spectrometry. *Phys Chem Chem Phys PCCP* **14**;13(6):2196–205 (2011).
 - 475 Fasciotti, M. *et al.* Separation of isomeric disaccharides by traveling wave ion mobility mass spectrometry using CO₂ as drift gas. *J Mass Spectrom JMS* **Dec**;47(12):1643–7 (2012).
 - 476 Li, H., Bendiak, B., Siems, W. F., Gang, D. R. & Hill, H. H. Determining the isomeric heterogeneity of neutral oligosaccharide-alditols of bovine submaxillary mucin using negative ion traveling wave ion mobility mass spectrometry. *Anal Chem* **17**;87(4):2228–35 (2015).
 - 477 Hofmann, J., Hahm, H. S., Seeburger, P. H. & Pagel, K. Identification of carbohydrate anomers using ion mobility-mass spectrometry. *Nature* **8**;526(7572):241–4 (2015).
 - 478 Williams, J. P. *et al.* Characterization of simple isomeric oligosaccharides and the rapid separation of glycan mixtures by ion mobility mass spectrometry. *Int J MASS Spectrom* **1**;298(1–3):119–27 (2010).
 - 479 J, S. T. *et al.* Identification of Isomeric N-Glycans by Conformer Distribution Fingerprinting using Ion Mobility Mass Spectrometry. *Chem Weinh Bergstr Ger* **26**;27(6):2149–54 (2021).
 - 480 Miller, R. L. *et al.* Shotgun ion mobility mass spectrometry sequencing of heparan sulfate saccharides. *Nat Commun* **20**;11(1):1481 (2020).
 - 481 Yang, H. *et al.* Identification of structurally closely related monosaccharide and disaccharide isomers by PMP labeling in conjunction with IM-MS/MS. *Sci Rep* **6** (2016).
 - 482 Huang, Y. & Dodds, E. D. Discrimination of Isomeric Carbohydrates as the Electron Transfer Products of Group II Cation Adducts by Ion Mobility Spectrometry and Tandem Mass Spectrometry. *Anal Chem* **2**;87(11):5664–8 (2015).
 - 483 Williamson, D. L. & Nagy, G. Evaluating the Utility of Temporal Compression in High-Resolution Traveling Wave-Based Cyclic Ion Mobility Separations. *ACS Meas Sci Au* **17**;2(4):361–9 (2022).
 - 484 McKenna, K. R. *et al.* Carbohydrate isomer resolution via multi-site derivatization cyclic ion mobility-mass spectrometry. *Analyst* **2**;144(24):7220–6 (2019).
 - 485 Dwivedi, P., Bendiak, B., Clowers, B. H. & Hill, H. H. Rapid resolution of carbohydrate isomers by electrospray ionization ambient pressure ion mobility spectrometry-

- time-of-flight mass spectrometry (ESI-APIMS-TOFMS). *J Am Soc Mass Spectrom* **1**;18(7):1163–75 (2007).
- 486 Li, H. *et al.* Resolving Structural Isomers of Monosaccharide Methyl Glycosides Using Drift Tube and Traveling Wave Ion Mobility Mass Spectrometry. *Anal Chem* **3**;84(7):3231–9 (2012).
 - 487 Gaye, M. M., Kurulugama, R. & Clemmer, D. E. Investigating carbohydrate isomers by IMS-CID-IMS-MS: precursor and fragment ion cross-sections. *Analyst* **28**;140(20):6922–32 (2015).
 - 488 Gaye, M. M., Nagy, G., Clemmer, D. E. & Pohl, N. L. B. Multidimensional Analysis of 16 Glucose Isomers by Ion Mobility Spectrometry. *Anal Chem* **16**;88(4):2335–44 (2016).
 - 489 Winkler, W., Huber, W., Vlasak, R. & Allmaier, G. Positive and negative electrospray ionisation travelling wave ion mobility mass spectrometry and low-energy collision-induced dissociation of sialic acid derivatives. *Rapid Commun Mass Spectrom* **25**, 3235–3244 (2011).
 - 490 Olivier, S., Ropartz, D., Fanuel, M. & Rogniaux, H. Fingerprinting of Underivatized Monosaccharide Stereoisomers Using High-Resolution Ion Mobility Spectrometry and Its Implications for Carbohydrate Sequencing. *Anal Chem* **4**;95(26):10087–95 (2023).
 - 491 B, P., WJC, B. & HA, S. Structural Characterization of Disaccharides Using Cyclic Ion Mobility Spectrometry and Monosaccharide Standards. *J Am Soc Mass Spectrom* **1**;35(5):1012–20 (2024).
 - 492 Peterson, T. L. & Nagy, G. Rapid cyclic ion mobility separations of monosaccharide building blocks as a first step toward a high-throughput reaction screening platform for carbohydrate syntheses. *RSC Adv* **13**;11(63):39742–7 (2021).
 - 493 Wernisch, S., Afshinnia, F., Rajendiran, T. & Pennathur, S. Probing the application range and selectivity of a differential mobility spectrometry–mass spectrometry platform for metabolomics. *Analytical and bioanalytical chemistry* **410**, 2865–2877 (2018).
 - 494 Paglia, G., Smith, A. J. & Astarita, G. Ion mobility mass spectrometry in the omics era: Challenges and opportunities for metabolomics and lipidomics. *Mass Spectrometry Reviews* **41**, 722–765 (2022).
 - 495 Delvaux, A., Rathahao-Paris, E. & Alves, S. Different ion mobility-mass spectrometry coupling techniques to promote metabolomics. *Mass Spectrometry Reviews* **41**, 695–721 (2022).
 - 496 Levy, A. J. *et al.* Recent progress in metabolomics using ion mobility-mass spectrometry. *TrAC Trends in Analytical Chemistry* **116**, 274–281 (2019).
 - 497 Mairinger, T., Causon, T. J. & Hann, S. The potential of ion mobility–mass spectrometry for non-targeted metabolomics. *Current Opinion in Chemical Biology* **42**, 9–15 (2018).
 - 498 Zhang, X., Quinn, K., Cruickshank-Quinn, C., Reisdorph, R. & Reisdorph, N. The application of ion mobility mass spectrometry to metabolomics. *Current Opinion in Chemical Biology* **42**, 60–66 (2018).
 - 499 Cismesia, A. P., Bell, M. R., Tesler, L. F., Alves, M. & Polfer, N. C. Infrared ion spectroscopy: an analytical tool for the study of metabolites. *Analyst* **143**, 1615–1623 (2018).
 - 500 Cismesia, A. P., Bailey, L. S., Bell, M. R., Tesler, L. F. & Polfer, N. C. Making mass spectrometry see the light: the promises and challenges of cryogenic infrared ion spectroscopy as a bioanalytical technique. *Journal of The American Society for Mass Spectrometry* **27**, 757–766 (2016).
 - 501 Cismesia, A. P., Tesler, L. F., Bell, M. R., Bailey, L. S. & Polfer, N. C. Infrared ion spectroscopy inside a mass-selective cryogenic 2D linear ion trap. *Journal of Mass Spectrometry* **52**, 720–727 (2017).
 - 502 Ho, J. S. *et al.* Distinguishing Galactoside Isomers with Mass Spectrometry and Gas-Phase Infrared Spectroscopy. *J Am Chem Soc* **21**;143(28):10509–13 (2021).
 - 503 Martens, J. *et al.* Infrared ion spectroscopy: New opportunities for small-molecule identification in mass spectrometry - A tutorial perspective. *Anal Chim Acta* **1093**, 1–15 (2020).
 - 504 van, O. R. E. *et al.* Evaluation of table-top lasers for routine infrared ion spectroscopy in the analytical laboratory. *Analyst* **146**, 7218–7229 (2021).
 - 505 Wight, C. A. & Beauchamp, J. Infrared spectra of gas-phase ions and their use in elucidating reaction mechanisms. Identification of C₇H₇-structural isomers by multiphoton electron detachment using a low-power infrared laser. *Journal of the American Chemical Society* **103**, 6499–6501 (1981).
 - 506 Braak, F. t. *et al.* Characterization of Elusive Reaction Intermediates Using Infrared Ion Spectroscopy: Application to the Experimental Characterization of Glycosyl Cations. *Accounts of Chemical Research* **55**, 1669–1679 (2022).
 - 507 Maitre, P. *et al.* Applications of infrared multiple photon dissociation (IRMPD) to the detection of posttranslational modifications. *Chemical reviews* **120**, 3261–3295 (2019).
 - 508 Kranenburg, R. F. *et al.* Mass-spectrometry-based identification of synthetic drug isomers using infrared ion spectroscopy. *Analytical Chemistry* **92**, 7282–7288 (2020).
 - 509 Stroganova, I. & Rijs, A. M. Ion Spectroscopy Coupled to Ion Mobility–Mass Spectrometry. (2021).
 - 510 Gray, C. J., Compagnon, I. & Flitsch, S. L. Mass spectrometry hybridized with gas-phase InfraRed spectroscopy for glycan sequencing. *Current Opinion in Structural Biology* **62**, 121–131 (2020).
 - 511 Greis, K., Kirschbaum, C., von Helden, G. & Pagel, K. Gas-phase infrared spectroscopy of glycans and glycoconjugates. *Current Opinion in Structural Biology* **72**, 194–202 (2022).
 - 512 Rijs, A. M. & Oomens, J. IR spectroscopic techniques to study isolated biomolecules. *Gas-Phase IR Spectroscopy and Structure of Biological Molecules*, 1–42 (2015).
 - 513 Eyler, J. R. Infrared multiple photon dissociation spectroscopy of ions in Penning traps. *Mass spectrometry reviews* **28**, 448–467 (2009).
 - 514 van Outersterp, R. E. *et al.* Distinguishing Oligosaccharide Isomers Using Far-Infrared Ion Spectroscopy: Identification of Biomarkers for Inborn Errors of Metabolism. *Analytical Chemistry* **95**, 9787–9796 (2023).
 - 515 Carlo, M. J. & Patrick, A. L. Infrared multiple photon dissociation (IRMPD) spectroscopy and its potential for the clinical laboratory. *Journal of Mass Spectrometry and Advances in the Clinical lab* **23**, 14–25 (2022).
 - 516 Martens, J. *et al.* Molecular identification in metabolomics using infrared ion spectroscopy. *Sci Rep* **7** (2017).
 - 517 Houthuijs, K. J. *et al.* An in silico infrared spectral library of molecular ions for metabolite identification. *Analytical Chemistry* **95**, 8998–9005 (2023).
 - 518 van Outersterp, R. E. *et al.* Targeted Small-Molecule Identification Using Heartcutting Liquid Chromatography–Infrared Ion Spectroscopy. *Analytical Chemistry* **95**, 3406–3413 (2023).

- 519 Talbot, F. O. & Simons, J. P. Sugars in the gas phase: the spectroscopy and structure of jet-cooled phenyl β -D-glucopyranoside. *Phys Chem Chem Phys* **18**;4(15):3562–5 (2002).
- 520 Contreras, C. S. *et al.* On the path to glycan conformer identification: Gas-phase study of the anomers of methyl glycosides of N-acetyl-D-glucosamine and N-acetyl-D-galactosamine. *Int J Mass Spectrom* **332**, 285–294 (2012).
- 521 Martens, J. *et al.* Unraveling the unknown areas of the human metabolome: the role of infrared ion spectroscopy. *J Inherit Metab Dis* **41**, 367–377 (2018).
- 522 Pearson, W. L. *et al.* Differentiation of Rubidiated Methyl-D-Glycoside Stereoisomers by Infrared Multiple-Photon Dissociation Spectroscopy in the O–H and C–H Stretching Regions. *J Phys Chem B* **15**;119(41):12970–81 (2015).
- 523 Tan, Y. *et al.* Vibrational Signatures of Isomeric Lithiated N-acetyl-D-hexosamines by Gas-Phase Infrared Multiple-Photon Dissociation (IRMPD) Spectroscopy. *J Am Soc Mass Spectrom* **1**;28(3):539–50 (2017).
- 524 Cagmat, E. B. *et al.* Vibrational signatures of metal-chelated monosaccharide epimers: gas-phase infrared spectroscopy of Rb⁺-tagged glucuronic and iduronic acid. *Phys Chem Chem Phys* **12** (2010).
- 525 RE, O. *et al.* Distinguishing Oligosaccharide Isomers Using Far-Infrared Ion Spectroscopy: Identification of Biomarkers for Inborn Errors of Metabolism. *Anal Chem* **21**;95(26):9787–96 (2023).
- 526 Stefan, S. E. & Eyler, J. Differentiation of Methyl-glucopyranoside Anomers by Infrared Multiple Photon Dissociation with a Tunable CO₂ Laser. *Anal Chem* **1**;81(3):1224–7 (2009).
- 527 Barnes, L., Schindler, B., Chambert, S., Allouche, A. R. & Compagnon, I. Conformational preferences of protonated N-acetylated hexosamines probed by InfraRed Multiple Photon Dissociation (IRMPD) spectroscopy and ab initio calculations. *Int J Mass Spectrom* **421**, 116–123 (2017).
- 528 Pellegrinelli, R. P. *et al.* A New Strategy Coupling Ion-Mobility-Selective CID and Cryogenic IR Spectroscopy to Identify Glycan Anomers. *J Am Soc Mass Spectrom* **4**;33(5):859–64 (2022).
- 529 Scutelnic, V. & Rizzo, T. R. Cryogenic Ion Spectroscopy for Identification of Monosaccharide Anomers. *J Phys Chem A* **4**;123(13):2815–9 (2019).
- 530 Bansal, P. *et al.* Using SLIM-Based IMS-IMS Together with Cryogenic Infrared Spectroscopy for Glycan Analysis. *Anal Chem* **7**;92(13):9079–85 (2020).
- 531 Elferink, H. *et al.* Direct experimental characterization of glycosyl cations by infrared ion spectroscopy. *Journal of the American Chemical Society* **140**, 6034–6038 (2018).
- 532 Mucha, E. *et al.* Unravelling the structure of glycosyl cations via cold-ion infrared spectroscopy. *Nature Communications* **9**, 4174 (2018).
- 533 Elferink, H. *et al.* The Glycosylation Mechanisms of 6, 3-Uronic Acid Lactones. *Angewandte Chemie International Edition* **58**, 8746–8751 (2019).
- 534 Hernandez, O., Isenberg, S., Steinmetz, V., Glish, G. L. & Maitre, P. Probing Mobility-Selected Saccharide Isomers: Selective Ion-Molecule Reactions and Wavelength-Specific IR Activation. *J Phys Chem A* **11**;119(23):6057–64 (2015).
- 535 Xian, F., Hendrickson, C. L. & Marshall, A. G. High resolution mass spectrometry. *Analytical Chemistry* **84**, 708–719 (2012).
https://doi.org/10.1021/AC203191T/ASSET/IMAGES/LARGE/AC-2011-03191T_0007.JPEG
- 536 Rathahao-Paris, E., Alves, S., Junot, C. & Tabet, J. C. High resolution mass spectrometry for structural identification of metabolites in metabolomics. *Metabolomics* **12** (2016).
- 537 Haag, A. M. Mass analyzers and mass spectrometers. *Advances in Experimental Medicine and Biology* **919**, 157–169 (2016). https://doi.org/10.1007/978-3-319-41448-5_7/FIGURES/8
- 538 Li, C. *et al.* Towards Higher Sensitivity of Mass Spectrometry: A Perspective From the Mass Analyzers. *Frontiers in Chemistry* **9**, 813359–813359 (2021).
<https://doi.org/10.3389/FCHEM.2021.813359/BIBTEX>
- 539 Grabarics, M. *et al.* Mass Spectrometry-Based Techniques to Elucidate the Sugar Code. *Chemical Reviews* **122**, 7840–7908 (2022).
https://doi.org/10.1021/ACS.CHEMREV.1C00380/ASSET/IMAGES/LARGE/CR1C00380_0042.JPEG
- 540 Hsu, H. C., Liew, C. Y., Huang, S. P., Tsai, S. T. & Ni, C. K. Simple Method for De Novo Structural Determination of Underivatized Glucose Oligosaccharides. *Scientific Reports* **2018** **8**, 1–12 (2018).
<https://doi.org/10.1038/s41598-018-23903-4>
- 541 Han, J., Lin, K., Sequria, C., Yang, J. & Borchers, C. H. Quantitation of low molecular weight sugars by chemical derivatization-liquid chromatography/multiple reaction monitoring/mass spectrometry. *ELECTROPHORESIS* **37**, 1851–1860 (2016).
<https://doi.org/10.1002/ELPS.201600150>
- 542 Zhang, M. J. *et al.* Quantification of carbohydrates in human serum using gas chromatography–mass spectrometry with the stable isotope-labeled internal standard method. *New Journal of Chemistry* **46**, 11357–11367 (2022). <https://doi.org/10.1039/D2NJ01243J>
- 543 Alseekh, S. *et al.* Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. *Nature Methods* **2021** **18**;7 **18**, 747–756 (2021). <https://doi.org/10.1038/s41592-021-01197-1>
- 544 Wang, S. L. *et al.* Paired Derivatization Approach with H/D-Labeled Hydroxylamine Reagents for Sensitive and Accurate Analysis of Monosaccharides by Liquid Chromatography Tandem Mass Spectrometry. *Analytical Chemistry* **94**, 3590–3599 (2022).
https://doi.org/10.1021/ACS.ANALCHEM.1C04924/ASSET/IMAGES/LARGE/AC1C04924_0005.JPEG
- 545 Millette, P. G., Chabot, J., Sheppard, D. C. & Le Mauff, F. Identification and Quantification of Monosaccharides from Fungal Cell Walls and Exopolysaccharides by Gas Chromatography Coupled to Mass Spectrometry. *Current Protocols* **3**, e853–e853 (2023).
<https://doi.org/10.1002/CPZ1.853>
- 546 Wang, Z., Zhang, J. & Li, L. Recent Advances in Labeling-Based Quantitative Glycomics: From High-Throughput Quantification to Structural Elucidation. *PROTEOMICS*, e202400057–e202400057 (2024).
<https://doi.org/10.1002/PMIC.202400057>
- 547 Nagy, G. & Pohl, N. L. B. Complete Hexose isomer identification with mass spectrometry. *Journal of the American Society for Mass Spectrometry* **26**, 677–685 (2015). https://doi.org/10.1007/S13361-014-1072-Z/SUPPL_FILE/JS8B05005_SI_001.PDF
- 548 Fang, T. T., Zirrolli, J. & Bendiak, B. Differentiation of the anomeric configuration and ring form of glucosyl-glycolaldehyde anions in the gas phase by mass spectrometry: isomeric discrimination between m/z 221 anions derived from disaccharides and chemical synthesis of m/z 221 standards. *Carbohydrate Research*

- 342, 217-235 (2007).
<https://doi.org/10.1016/J.CARRES.2006.11.021>
- 549 Metabolomics-Society.
<https://wiki.metabolomicssociety.org/index.php/Databases> (2024).
- 550 Wishart, D. S. *et al.* HMDB 5.0: the human metabolome database for 2022. *Nucleic acids research* **50**, D622-D631 (2022).
- 551 Wishart, D. S. *et al.* MiMeDB: the human microbial metabolome database. *Nucleic Acids Research* **51**, D611-D620 (2023).
- 552 Hummel, J. *et al.* Mass spectral search and analysis using the golm metabolome database. *The handbook of plant metabolomics*, 321-343 (2013).
- 553 Fiehn, O. <https://mona.fiehnlab.ucdavis.edu/> (2024).
- 554 HighChemLLC. <https://www.mzcloud.org/> (2024).
- 555 <http://webbook.nist.gov/> (2024).
- 556 Linstrom, P. J. & Mallard, W. G. The NIST Chemistry WebBook: A chemical data resource on the internet. *Journal of Chemical & Engineering Data* **46**, 1059-1063 (2001).
- 557 Van Santen, J. A. *et al.* The natural products atlas: an open access knowledge base for microbial natural products discovery. *ACS central science* **5**, 1824-1833 (2019).
- 558 Li, X., Xu, Z., Hong, X., Zhang, Y. & Zou, X. Databases and bioinformatic tools for glycobiology and glycoproteomics. *International Journal of Molecular Sciences* **21**, 6727 (2020).
- 559 Campbell, M. P. *et al.* Validation of the curation pipeline of UniCarb-DB: building a global glycan reference MS/MS repository. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **1844**, 108-116 (2014).
- 560 Kameyama, A. *et al.* A strategy for identification of oligosaccharide structures using observational multistage mass spectral library. *Analytical chemistry* **77**, 4719-4725 (2005).
- 561 Toukach, P. V. & Egorova, K. S. Carbohydrate structure database merged from bacterial, archaeal, plant and fungal parts. *Nucleic acids research* **44**, D1229-D1236 (2016).
- 562 Baker, E. S. *et al.* METLIN-CCS: an ion mobility spectrometry collision cross section database. *Nature methods* **20**, 1836-1837 (2023).
- 563 Aoki-Kinoshita, K. F. & Kanehisa, M. Glycomic analysis using KEGG GLYCAN. *Glycoinformatics*, 97-107 (2015).
- 564 Caspi, R. *et al.* The MetaCyc database of metabolic pathways and enzymes-a 2019 update. *Nucleic acids research* **48**, D445-D453 (2020).
- 565 Cantarel, B. L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic acids research* **37**, D233-D238 (2009).
- 566 Altman, T., Travers, M., Kothari, A., Caspi, R. & Karp, P. D. A systematic comparison of the MetaCyc and KEGG pathway databases. *BMC bioinformatics* **14**, 1-15 (2013).
- 567 Geller-McGrath, D. *et al.* Predicting metabolic modules in incomplete bacterial genomes with MetaPathPredict. *Elife* **13**, e85749 (2024).
- 568 Gardinassi, L. G., Xia, J., Safo, S. E. & Li, S. Bioinformatics tools for the interpretation of metabolomics data. *Current Pharmacology Reports* **3**, 374-383 (2017).
- 569 Misra, B. B. New software tools, databases, and resources in metabolomics: Updates from 2020. *Metabolomics* **17**, 49 (2021).
- 570 O'Shea, K. & Misra, B. B. Software tools, databases and resources in metabolomics: updates from 2018 to 2019. *Metabolomics* **16**, 36 (2020).
- 571 Misra, B. B. & Mohapatra, S. Tools and resources for metabolomics research community: A 2017–2018 update. *Electrophoresis* **40**, 227-246 (2019).
- 572 Misra, B. B. New tools and resources in metabolomics: 2016–2017. *Electrophoresis* **39**, 909-923 (2018).
- 573 Misra, B. B., Fahrman, J. F. & Grapov, D. Review of emerging metabolomic tools and resources: 2015–2016. *Electrophoresis* **38**, 2257-2274 (2017).
- 574 Banimfeg, B. H., Shamayleh, A. & Alshraideh, H. Survey for computer-aided tools and databases in metabolomics. *Metabolites* **12**, 1002 (2022).
- 575 Giera, M., Aisporna, A., Uritboonthai, W. & Siuzdak, G. The hidden impact of in-source fragmentation in metabolic and chemical mass spectrometry data interpretation. *Nature Metabolism*, 1-2 (2024).
- 576 Smith, C. A. *et al.* METLIN: a metabolite mass spectral database. *Therapeutic drug monitoring* **27**, 747-751 (2005).
- 577 Fang, T. T. & Bendiak, B. The stereochemical dependence of unimolecular dissociation of monosaccharide-glycolaldehyde anions in the gas phase: A basis for assignment of the stereochemistry and anomeric configuration of monosaccharides in oligosaccharides by mass spectrometry via a key discriminatory product ion of disaccharide fragmentation, m/z 221. *Journal of the American Chemical Society* **129**, 9721-9736 (2007).
https://doi.org/10.1021/JA0717313/SUPPL_FILE/JA0717313SI20070709_124325.PDF
- 578 Huynh, H. T. *et al.* Collision-induced dissociation of sodiated glucose, galactose, and mannose, and the identification of anomeric configurations. *Physical Chemistry Chemical Physics* **20**, 19614-19624 (2018).
<https://doi.org/10.1039/C8CP03753A>
- 579 Liew, C. Y. *et al.* Structural identification of N-glycan isomers using logically derived sequence tandem mass spectrometry. *Communications Chemistry* **2021 4:1 4**, 1-11 (2021). <https://doi.org/10.1038/s42004-021-00532-z>
- 580 Huang, S. P., Hsu, H. C., Liew, C. Y., Tsai, S. T. & Ni, C. K. Logically derived sequence tandem mass spectrometry for structural determination of Galactose oligosaccharides. *Glycoconjugate Journal* **38**, 177-189 (2021). <https://doi.org/10.1007/S10719-020-09915-1/FIGURES/9>
- 581 van Tetering, L. *et al.* A spectroscopic test suggests that fragment ion structure annotations in MS/MS libraries are frequently incorrect. *Communications Chemistry* **7**, 30 (2024).
- 582 Urban, J. *et al.* Predicting glycan structure from tandem mass spectrometry via deep learning. *Nature Methods* **21** (2024). <https://doi.org/10.1038/s41592-024-02314-6>
- 583 Sirocchi, C. *et al.* Exploring machine learning for untargeted metabolomics using molecular fingerprints. *Comput Meth Prog Bio* **250** (2024).
- 584 Park, J., Jo, J. & Yoon, S. Mass spectra prediction with structural motif-based graph neural networks. *Sci Rep-Uk* **14** (2024).
- 585 Xue, J., Wang, B. Y., Ji, H. C. & Li, W. H. RT-Transformer: retention time prediction for metabolite annotation to assist in metabolite identification. *Bioinformatics* **40** (2024).
- 586 Hsiung, S. Y. *et al.* Machine learning-based monosaccharide profiling for tissue-specific classification of *Wolffiporia extensa* samples. *Carbohydrate Polymers* **322** (2023).

SugarBase: Mapping Glycomolecule Precursors in Microbes

Jitske M. van Ede, Martine C. Holst Sørensen, Mark C.M.
van Loosdrecht, Martin Pabst



Abstract

Glycan biosynthesis relies on nucleotide-activated sugars, essential metabolites across all domains of life, yet their usage in microbes is poorly understood. Here we present SugarBase, a mass spectrometry and bioinformatic pipeline for untargeted exploration of microbial nucleotide sugar networks. SugarBase resolves the chemical complexity of microbial metabolism by combining narrow-window DIA fragmentation with a chemistry-informed parent ion identification algorithm.

Applying SugarBase across a broad phylogenetic range of microbes revealed extensive, species-specific nucleotide sugar profiles, including many candidates with no existing annotation, generating the most comprehensive inventory of nucleotide sugars to date.

SugarBase guided identification of gene clusters and allowed discrimination between pseudaminic- and legionaminic acid-producing strains, where genomic and proteomic data provided only ambiguous information. We resolved distinct nonulosonic acid profiles in several *Campylobacter jejuni* strains, sugars which may alter susceptibility towards distinct flagellotropic phages. We further identify previously undescribed CMP-activated higher-carbon ulosonic acids in *Magnetospirillum*, expanding the known chemical space in glycan biosynthesis. In summary, SugarBase supports scalable discovery of microbial nucleotide sugar pathways and enzymes, expanding access to chemically complex glycans and providing new targets for antimicrobial development.

3.1 | Introduction

Nucleotide-activated sugars are central precursors for the diverse range of glycomolecules found across all domains of life¹. They form storage polymers, structural components, and surface glycans that mediate protection and environmental interactions. In pathogenic bacteria, these glycans often act as virulence factors; for example, O-linked glycans on the flagella of *Campylobacter jejuni* are essential for motility²⁻⁵. As a result, glycan biosynthesis pathways represent attractive targets for the development of antimicrobials and vaccines⁶⁻⁹. Since the discovery of UDP-glucose in the 1950s, many monosaccharides have been identified as UDP, ADP, GDP, CMP, and dTDP conjugates¹⁰. Their biosynthesis often follows defined biochemical rules, such as the preferential use of specific nucleotides for particular sugars. For example, glucose is predominantly activated in the form of UDP-Glc, which links central metabolism with glycan biosynthesis pathways¹¹. While some nucleotide sugars are conserved across taxa, their diversity can vary substantially. Mammalians typically rely on a relatively limited and well-defined pool of nine nucleotide-sugars^{1,12}. This diversity expands to more than 30 in plants¹³. In contrast, the prokaryotic repertoire is still largely unexplored and likely includes hundreds to thousands of distinct derivatives.

For example, O-antigen biosynthesis in *E. coli* and *Salmonella* alone involves some 30 different nucleotide sugar intermediates¹⁴. Furthermore, sugars can be activated by different nucleotides. For example, next to UDP-Glc, also ADP-Glc, GDP-Glc and TDP-Glc are commonly found in microbes, where the different nucleotides likely provide a strategy to partition flux into distinct biosynthetic pathways or classes of glycans¹⁵.

A few sugar classes are activated exclusively by a single nucleotide. For example, ulosonic acids, including sialic acids and the bacterial analogues pseudaminic and legionaminic acids, have thus far only been observed in CMP-activated form. Sialic acids and their bacterial analogues are important functional sugars involved in recognition and immune evasion and are known virulence factors in pathogenic bacteria. Their exclusive activation is likely dictated by molecular stability, as the α -keto acid group allows a stable linkage to monophosphates, but not to diphosphates¹⁶. Currently, only a few complete biosynthetic pathways for nucleotide sugars and their glycans are known in microbes, while the vast majority remain unexplored. Moreover, their chemical synthesis is often infeasible due to stereochemical complexity and limited stability^{17,18}, and the lack of commercial standards presents a major bottleneck for developing novel glycoconjugates and biomedical assays. In particular, non-pathogenic microbes represent an attractive source of nucleotide sugars, required to accelerate scalable production of

glycoconjugates. For example, CMP-activated nonulosonic acids are particularly challenging to synthesise^{1,19,20}, and the biosynthetic routes from *H. pylori* and *C. jejuni* are the most effective ways to produce these compounds^{21,22}.

Different analytical strategies are used to identify nucleotide sugars from cell lysates. However, these approaches generally do not support untargeted, large-scale exploration of microbial species. Chromatographic methods, including reverse-phase separations employing ion-pairing reagents and ion-exchange chromatography, are either incompatible with mass spectrometric detection or heavily dependent on the availability of reference standards, limiting their application to complex mixtures of unknown sugars^{12,23-26}. By contrast, hydrophilic interaction liquid chromatography and porous graphitic carbon chromatography are compatible with mass spectrometric detection, where the latter is chemically more stable and provides excellent separation of isomeric structures^{24,27,28}. Alternatively, capillary electrophoresis was coupled to mass spectrometric precursor ion scanning which allowed an untargeted exploration of nucleotide sugars²⁹. However, coupling capillary electrophoresis to mass spectrometry is technically challenging, and high-resolution mass spectrometric analysis of complex cell lysates yields thousands of features, making manual exploration impossible.

Therefore, we developed the SugarBase workflow, which enables untargeted mapping of nucleotide sugar networks in complex, previously uncharacterized microbial species. The workflow combines porous graphitic carbon separation with high-resolution, narrow-window fragmentation. Sugar nucleotides are identified using the SugarBase algorithm, which leverages a large empirical database to link diagnostic fragments to precursor ions, allowing detection of novel derivatives. To demonstrate its performance, we screened a broad taxonomic range of microbes, including mammalian and plant controls, generating the most comprehensive inventory of nucleotide sugars to date.

Interestingly, in several environmental microbes with biotechnological potential, such as *Ca. Accumulibacter phosphatis* and *M. gryphiswaldense*, we identified abundant levels of CMP-activated nonulosonic acids (CMP-NulOs). In addition, we established the CMP-NulO profiles of several pathogenic *C. jejuni* strains, revealing over a dozen different pseudaminic acid (Pse) and legionaminic acid (Leg) derivatives, with the potential to modulate protein and cell surface hydrophobicity, net charge, and susceptibility towards flagellotropic phages. Finally, SugarBase also provides the first evidence of a CMP-activated higher-carbon ulosonic acid discovered in *M. gryphiswaldense*. While differentiation between Leg and Pse is poorly reported in literature, the SugarBase workflow guides identification of

biosynthetic gene clusters and allows discrimination between pseudaminic- and legionaminic acid-producing strains, which was validated with orthogonal labelling experiments and synthetic reference standards.

The SugarBase pipeline is freely available as Python code and as a standalone desktop executable with a browser-based graphical user interface that enables interactive exploration of nucleotide sugar profiles: <https://sourceforge.net/projects/sugarbase-x/files>. All acquired data are publicly accessible through the 4TU.ResearchData platform (<https://data.4tu.nl/project/SugarBase>) and can be reanalysed using the SugarBase workflow.

3

3.2 | Materials and Methods

Sources of microbial biomass and control cells. A table detailing the analyzed biomass, including sources, growth conditions, and suppliers, is provided in SI Excel 1. Specifically, *Campylobacter jejuni*, *Magnetospirillum gryphiswaldense*, and *Ca. Accumulibacter phosphatis* were cultured as described in the following.

Culturing of *Campylobacter jejuni*. *C. jejuni* strains NCTC12658 and NCTC12662 were grown in liquid Brain Heart Infusion broth under microaerophilic conditions (6% CO₂, 6% O₂, 88% N₂) at 41.6°C. Biomass was harvested at an OD₆₀₀ of 0.25 by centrifuging at 6500 rpm, for 7 min, in multiple rounds. *Campylobacter jejuni* strains 11168 (DSM 27585) and 81116 (DSM 24189) were cultured on Mueller–Hinton agar supplemented with Oxoid™ *Campylobacter* Growth Supplement (Thermo Fisher Scientific). Microaerophilic conditions were generated in a 2.5 L anaerobic jar using Anaerocult™ C (Merck). Plates were incubated for 48 h at 37°C. Biomass was harvested with sterile cotton swabs into sterile 1× PBS, centrifuged (10 min, 14,000 rpm, 4°C), and stored at –80°C until use.

Culturing of *M. gryphiswaldense*. *Magnetospirillum gryphiswaldense* DSM 6361 was cultured in a lactate-containing medium under microaerophilic conditions at 30 °C. 5 L culture in 20 L bottle was initiated under the gas phase containing 2% O₂ in argon to support growth onset. Upon visible initiation of growth, microaerophilic conditions were relaxed by opening the culture flasks to atmospheric air through a sterile filter and the bottle was placed on a rotary shaker at 100 rpm, allowing continued growth under oxic conditions. ***Ca. Accumulibacter* enrichment.** A laboratory-scale enrichment of *Candidatus Accumulibacter* was maintained in sequencing batch reactor mode under alternating anoxic/aerobic feast–famine cycles following methods described recently³⁰. Briefly, the enrichment was maintained at 20°C and pH 7.0 and fed with acetate as the primary carbon source. Anoxic phases were maintained by N₂ sparging, followed by aerobic phases with air supply. Biomass was harvested at the end of the aerobic phase.

Metabolite extraction from cell materials. 500 μL pre-cooled extraction solvent (-20°C , 40% acetonitrile, 40% methanol, 20% MS-grade water) and 0.1 g acid-washed glass beads (500 μm , Sigma Aldrich) were added to 100–200 mg of biomass, wet weight. Cells were disrupted by three cycles of vortexing for 1.5 min, with 1 min on ice between cycles. Plant tissues were ground in liquid nitrogen using a mortar and pestle prior to solvent addition. The lysates were centrifuged at 14000 rpm, 4°C for 15 min in an Eppendorf container (5424R), and supernatants were collected. The extracts were purified and concentrated using HyperSep™ Hypercarb™ SPE 96-well plates (Thermo Scientific, Germany). In short, the columns were conditioned with 600 μL 10 mM ammonium bicarbonate (ABC) in 60% acetonitrile (ACN) and equilibrated with $2 \times 600 \mu\text{L}$ MS-grade H_2O . Samples were diluted 1:10 with MS-grade H_2O , loaded onto the columns, washed with 1 mL MS-grade H_2O , and eluted with $2 \times 200 \mu\text{L}$ 10 mM ABC in 60% ACN. Eluates were divided into three equal aliquots and dried in a SpeedVac concentrator at room temperature. Dried extracts were stored at -80°C and reconstituted in 15 μL MS-grade H_2O prior to mass spectrometric analysis.

SugarBase workflow.

A. PGC chromatography coupled to narrow-mass window DIA. Data independent mass spectrometric analysis was performed using an Acquity Ultra Performance Liquid Chromatograph (Waters, UK) coupled to a Q Exactive™ Focus Hybrid Quadrupole-Orbitrap™ Mass Spectrometer (Thermo Scientific, Germany). Chromatographic separation was performed on a 100 x 1 mm Hypercarb™ Porous Graphitic Carbon Column (Thermo Scientific, No. 35005-101030) at a flow rate of 100 $\mu\text{L}/\text{min}$. Additional electrical grounding between the separation column and the electrospray ionization source was applied to prevent interference of the electrospray voltage with chromatographic separation³¹. Mobile phase A consisted of 15 mM ammonium bicarbonate in MS-grade H_2O and mobile phase B of 100% acetonitrile. A linear gradient from 2.5% B to 35% B was applied over 18 min, followed by a linear gradient to 60% B over another 6 minutes. Each sample was analysed in duplicate followed by two blanks. The mass spectrometer was operated in parallel reaction monitoring mode, where a mass range of 500–612 Da (Part 1) was fragmented continuously in 8 Da steps, using an isolation window of 10 m/z. Each window was measured twice, first applying a NCE of 10 (pseudo MS1) followed by a NCE of 28 (MS2). Electrospray ionization was performed in negative ionisation mode. MS analysis was performed at a resolution of 35 K, with a AGC target of $2.00\text{E}5$ and a maximum injection time of 75 ms. Data were acquired in the centroid mode, with microscans set to 1. Every sample was injected a second time using an extended mass range, covering 596–708 m/z (Part 2). Raw data were

analysed using XCalibur 4.1 (Thermo Fisher Scientific, Germany), Matlab R2023b (MathWorks) and Python 3.9. The mass spectrometer was calibrated using the Pierce™ LTQ ESI negative ion calibration solution (Thermo Fisher Scientific, Germany).

B. Nucleotide sugar data processing pipeline. Potential nucleotide-sugars were identified using the Python SugarBase pipeline. Briefly, raw mass spectrometry files were converted to “.mzXML” format using msConvert³². Peak lists were imported from which mass lists were extracted. The odd and even scans were separated, representing pseudo-MS1 and fragmented MS2 data, hereafter referred to as “MS1” and “MS2,” respectively. “MS1” data was filtered to exclude masses outside the designated isolation window (“channel”) to remove potential in-source fragments. Peaks with an intensity <1000 or peaks outside the empirical retention time window (4.5–15.0 min) were not further processed. Doubly charged peaks were removed, and deisotoping was performed considering 10 ppm mass error. Up to three isotopes peaks with an abundance ratio threshold compared to M were considered ($M+1 < 0.75$, $M+2 < 0.33$, $M+3 < 0.15$). A cell array was generated for each “MS2” channel (10 Da window), containing scan index, mass channel, retention time, fragment peaks, and peak intensities. MS2 arrays containing potential nucleotide fragment markers (CMP, UMP, UDP, GMP, GDP, AMP, ADP, CDP, dTMP, dTDP, TMP, and TDP; masses 322.0446, 323.0286, 402.9949, 362.0507, 442.0171, 346.0558, 426.0221, 402.0109, 321.0493, 401.0157, 337.0442, and 417.0106, respectively) were extracted considering a mass error of 10 ppm. Corresponding MS1 scans were then examined for candidate nucleotide sugar peaks containing the identified nucleotide marker. Discrimination between true nucleotide sugars and background metabolites was achieved using a sugar composition database based on work published by Pabst and co-workers (2021)³³. All possible combinations of $C_{5-16}H_{8-27}N_{0-4}O_{3-9}S_{0-1}$ were generated and constrained using a series of physicochemical filters. Briefly, compositions with monoisotopic masses below 100 Da or above 400 Da were excluded. Mass defect thresholds were set to exclude values below 0.01 or above 0.19, with additional constraints applied for lower mass species. Compositions with masses between 100–200 Da and mass defects greater than 0.09, and those between 100–250 Da with mass defects greater than 0.13, were excluded. Double bond equivalents (DBE) were restricted to values between 1.0 and 6.0, with compositions in the 100–150 Da mass range excluded if DBE exceeded 3.0. Elemental ratio filters were applied by excluding compositions with carbon-to-hydrogen (C/H) ratios below 0.40 or above 1.25, and carbon-to-(oxygen plus nitrogen) [C/(O+N)] ratios below 0.90 or above 2.10, accounting for potential sulfonic acid formation. Finally, nitrogen and sulfur content was limited

by excluding compositions with more than one nitrogen atom in the 100–200 Da mass range, more than three nitrogen atoms in the 100–275 Da range, or any compositions containing both nitrogen and sulfur atoms. The resulting sugar compositions were then combined with the nucleotide masses of CMP, UDP, GDP, ADP, CDP, dTDP, and TDP, to generate a theoretical nucleotide sugar database. All MS1 mass peaks matching the corresponding nucleotide sugar space within 10 ppm mass window were retained for further processing. Precursors adjacent to CMP marker fragments were matched to CMP and CDP sugars. Moreover, to increase robustness, only precursor ions detected in at least three consecutive scans within the same channel and exhibiting the corresponding nucleotide marker fragment in MS2 spectra were retained for further processing. Finally, to provide confidence in the nucleotide sugar candidate outputs, candidates were scored using multiple criteria. Briefly, a rule-based scoring system was employed, where hits detected in at least four consecutive scans were assigned a score of +10. Signal consistency across scans was further evaluated using Pearson correlation coefficients calculated over three consecutive scans. The “MS1” scan with the highest intensity for a given nucleotide-sugar was designated the “main hit” provided that both the precursor and fragment masses were present in the adjacent spectra of the same channel. These three spectra were used to calculate the Pearson correlation coefficients, where correlations greater than 0.80 were rewarded (+10), whereas correlations below 0.70 and 0.50 were penalized (–10 and –50, respectively). For non-CMP-activated nucleotide sugars, the ratio of precursor ion intensity in the MS1 scan to fragment ion intensity in the corresponding MS2 scan was used as an additional criterion. Ratios below 100 were rewarded (+10), ratios below 1 were penalized (–10), and ratios exceeding 110 were strongly penalized (–50). For CMP-activated nucleotide sugars, scoring criteria included the ratio of precursor intensity in the MS2 scan relative to the CMP fragment intensity in the same MS2 scan, with values ≤ 25 rewarded (+10) and values > 25 penalized (–50). High precursor abundance in MS1 ($> 100K$ intensity units) and presence of a characteristic CO_2 neutral loss in MS2 spectra ($-\text{CO}_2$, -43.98983) contributed positively (+10). In addition, an elemental parity (“N-rule”) was applied, requiring both nitrogen and hydrogen atom counts to be either even or odd. Deviations from this rule resulted in a penalty (–50). Finally, for each nucleotide sugar candidate, an extracted ion chromatogram was generated together with the corresponding precursor and fragmentation spectra. These results are visualized via the browser-based interface and exported as PNG files, while a summary table containing scores and assigned chemical compositions is reported as an Excel file. Additional targeted nucleotide sugar analyses from acquired data were performed by

specifically screening for common pentose, hexose, and heptose monosaccharide derivatives, including amino sugars and N-acetylated, deoxy, dideoxy, uronic acid, and N-acetyluronic acid variants. These monosaccharides were combined *in silico* with the common activating nucleotides ADP, UDP, GDP, dTDP, TDP, CDP, and CMP. In addition, CMP-, UMP-, GMP-, dTMP-, and TMP-activated NeuAc, NulOAc₂, NulOAcAm, and NulOAcAla, were included (SI Excel file, targeted analysis). The same filtering and scoring criteria as used for untargeted discovery were applied, except that no MS1 intensity threshold was used and only candidates with scores ≥ 0 were retained for further analysis.

C. SugarBase code availability. The SugarBase Python code and the standalone desktop executable with graphical user interface are provided via SourceForge: <https://sourceforge.net/projects/sugarbase-x/>. The provided pipeline was established and tested using Q Exactive Orbitrap data (Thermo Fisher Scientific).

Shotgun proteomic analysis. Approximately 25 mg (wet weight) of each cell biomass material was resuspended in 175 μ L 50 mM TEAB containing 1% (w/w) NaDOC, adjusted to pH 8 and 175 μ L Bacterial Protein Extraction Reagent (B-PER; Thermo Scientific, cat. 78243). After addition of 0.15 g acid-washed glass beads (150–212 μ m; Sigma-Aldrich), samples underwent three cycles of bead beating (1.5 min) using a bench top vortex, with 1 min breaks on ice in between. Subsequently, the samples were centrifuged at 14000 rcf for 10 min at 4 °C. For *C. jejuni*, the samples were additionally incubated at 80°C for 3 min at 1000 rpm, centrifuged, and supernatants were passed through a 0.22 μ m filter before further processing. For all samples, proteins were precipitated from the supernatants by adding TCA at a 1:4 ratio, vortexed, and incubating at 4°C, followed by centrifugation to collect the protein pellets. Pellets were washed with ice-cold acetone and resuspended in 200 μ L 6 M urea (*C. jejuni*) or 100 μ L 6 M urea (*E. coli* and *Ca. Accumolibacter*) and incubated at 37°C, 1000 rpm until solubilized. Reduction was performed by adding 10 mM dithiothreitol in 200 mM ammonium bicarbonate to a final DTT concentration of 2.3 mM and incubating for 1 h at 37°C with shaking at 300 rpm. Alkylation was carried out by adding 20 mM iodoacetamide in 200 mM ammonium bicarbonate to a final concentration of 3.75 mM and incubating for 30 min at room temperature in the dark. Samples were diluted with ammonium bicarbonate buffer to bring the sample to below <1 M urea. A 100 μ L aliquot for each sample was digested with 5 μ L of 0.1 μ g/ μ L trypsin (Promega, cat. V5111) overnight at 37°C, 300 rpm. On the following day, samples were cleaned using Oasis HLB 96-well μ Elution plates (2 mg sorbent, 30 μ m, Waters, cat. 186001828BA). Columns were conditioned with 750 μ L methanol, equilibrated with 2 $\times$ 500 μ L MS-grade water, and loaded with samples. After washing with 350 μ L 5% methanol in MS-grade

water, peptides were eluted with 200 μ L 2% formic acid in 80% methanol followed by 200 μ L 1 mM ABC in 80% methanol. Eluates were dried in a SpeedVac concentrator and reconstituted in 3% acetonitrile containing 0.1% formic acid to a final concentration of approximately 500 ng/ μ L. 1 μ L of proteolytic digest, as estimated by NanoDrop UV spectrophotometer at 280 nm, was used for shotgun proteomic analysis, using an EASY nano-LC 1200 system coupled to a Q Exactive Plus Orbitrap mass spectrometer (Thermo Scientific, Germany). Peptides were separated on a 0.05 $\times$ 150 mm C18 analytical column (Thermo Scientific, catalog no. 164943). Mobile phase A consisted of 0.1% formic acid and 1% acetonitrile in MS-grade water, while mobile phase B contained 0.1% formic acid and 80% acetonitrile in MS-grade water. The LC gradient started with a 5 min hold at 5% mobile phase B, followed by a linear increase to 30% B over 90 min, then ramped to 75% B over an additional 25 min, maintaining a constant flow rate of 300 nL/min throughout the run. Full MS scans were acquired at a resolution of 70K with an automatic gain control target of 3E6 and a maximum injection time of 100 ms. Data-dependent acquisition was employed, where top 10 precursor ions were selected across a mass range of 550–1500 m/z using an isolation window of 1.7 m/z. Fragmentation was performed using a normalized collision energy of 28%. Fragmentation spectra were recorded at a resolution of 17.5K with an AGC target of 5E5 and a max injection time of 100 ms. Protein identification was carried out using PEAKS Studio X (Bioinformatics Solutions Inc., Canada). Database searches were performed against protein sequence databases obtained from UniProt for *Campylobacter jejuni* (UP000000799) and *Escherichia coli* K-12 (UP000000625), together with the GPM cRAP contaminant database. The reference proteome for *Ca. Accumulibacter phosphatis* was derived from whole-metagenome sequencing, as described previously³⁴. Carbamidomethylation was specified as a fixed modification, whereas oxidation and deamidation were included as variable modifications. Trypsin was selected as the digestion enzyme, permitting up to three missed cleavage sites. Mass tolerance settings were defined as 20 ppm for precursors and 0.02 Da for fragment ions. Peptide and protein identifications were filtered at a false discovery rate (FDR) of 1%, and protein identifications were considered significant when supported by at least two unique peptides. The NovoBridge and NovoLign pipelines were used to confirm the taxonomic purity of the samples^{34,35}.

DMB labelling and mass spectrometric analysis of free NulOs. Approximately 50 mg (wet weight) of biomass was suspended in 250 μ L of 2 M acetic acid, followed by the addition of 0.15 g acid-washed glass beads (150–212 μ m). Cells were disrupted by three bead-beating cycles (1.5 min each), with 1 min incubation on ice

and 2 min centrifugation (14,000 rpm, 4°C) between cycles. Acid hydrolysis was performed at 80°C for 2 h at 300 rpm using an Eppendorf heating block. After centrifugation (10 min, 14,000 rpm, 4°C), supernatants were filtered through a 0.22 µm membrane to remove residual cells and dried in a vacuum concentrator at 60°C. Alternatively, dried nucleotide-sugar extracts, obtained as described in the *Metabolite extraction* section, were hydrolyzed in 200 µL 2 M acetic acid under identical conditions and dried at 60°C. For DMB derivatization, dried hydrolysates were resuspended in DMB labeling solution (1.4 M acetic acid, 0.75 M 2-mercaptoethanol, 18 mM sodium dithionite, 7 mM DMB) and incubated for 2.5 h at 50°C (600 rpm), using 50 µL for biomass-derived samples and 25 µL for nucleotide-sugar extracts. DMB-labeled samples were analyzed using the protocol described in Pabst et al.³³ Briefly, labeled lysates were separated on a Waters Acquity UPLC system equipped with a C18 BEH column (1.7 µm particle size) and coupled to a Q Exactive Focus Orbitrap mass spectrometer. The instrument was operated in positive electrospray ionization mode with alternating full MS scans and narrow-mass window fragmentation scans covering the *m/z* range 390–525. Mass windows of 5.5 Da were sequentially isolated and fragmented using a normalized collision energy of 26. Raw mass spectrometric data were processed using MATLAB scripts as described by Kleikamp et al.³⁶ Additional MS3 fragmentation experiments were performed on DMB-labeled ulosonic acids extracted from *Magnetospirillum gryphiswaldense* and on DMB-labeled synthetic standards of legionaminic acid and Neu5Ac. Compounds were separated using an UltiMate 3000 LC system equipped with an HT µPAC Neo Plus C18 column (Thermo Fisher Scientific) and coupled to an Orbitrap Eclipse Tribrid mass spectrometer, operated in positive ionization mode. The flow rate was set to 1.5 µL/min, with water plus 0.1% formic acid as solvent A, and 80% acetonitrile, 0.1% formic acid as solvent B. A gradient from 7.5% to 45% solvent B was maintained over 15 min. The mass spectrometer continuously targeted the intact DMB-XuLO, DMB-LegAc₂, and DMB-Neu5Ac precursor ions, followed by targeted fragmentation of the corresponding C9 and C10 marker fragments *m/z* 295.0716, 297.0872, and 311.1056, respectively. MS2 isolation was performed in the quadrupole with a 1.5 *m/z* window using a normalized collision energy of 25 and Orbitrap detection at 500K resolution. MS3 isolation was carried out in the linear ion trap using a 2 *m/z* window and a normalized collision energy of 32 or 40 and Orbitrap detection at 240K resolution. Fragmentation spectra were manually interpreted using Xcalibur, ChemDraw, and ChemCalc³⁷. **Sequence alignment, and investigation of biosynthetic gene clusters (BGCs).** Reference protein sequences representing the legionaminic acid (Leg) and pseudaminic acid (Pse) biosynthetic pathways were compiled to identify homologous genes across

other microbial genomes. For the legionaminic acid pathway, reference proteins included LegB (Q0P8T8), LegC (Q0P8T7), LegH (Q0P8V9), LegG (Q0P8T0), LegI (Q0P8T1), LegF (Q0P8S7), Maf4/1 (Q0P8S3), PtmG (Q0P8T3), PtmH (Q0P8T2), and GImU (Q9PPA2). UDP-GlcNAc-linked pathway components included PtmE (Q0P8S9), PgmL (Q0P8K7), PtmA (Q0P8S6), and PtmF (Q0P8S8). For the pseudaminic acid pathway, reference sequences comprised PseB (Q0P8W4), PseC (Q0P8W3), PseH (Q0P8U4), PseG (Q0P8U5), PseI (Q0P8U0), PseF (Q0P8U6), and PseE (Q0P8S2). Additional bacterial nonulosonic acid biosynthesis references included PglF (Q0P9D4), PglE (UniRef100_Q0P9D3), PglD (UniRef100_Q0P9D1), NeuC/Lpg0753 (Q5ZXH8), NeuB (Q5ZXH9), and NeuA (A0A0P0LXX4). These reference sequences were used to construct a custom database for homology searches using DIAMOND. Predicted proteomes from the analyzed microbial genomes were aligned against this reference set using DIAMOND with default parameters to identify candidate pseudaminic acid and legionaminic acid-associated biosynthetic genes and homologous pathway components. Alignments within a 5% range of the top alignment score for a query were reported ('--top 5'). DIAMOND alignment results were processed using a custom Python script to retain high confidence homologs. For each alignment, query coverage was approximated from the aligned query span, and hits were filtered using stringent criteria to exclude spurious or local matches. Only alignments covering at least 70% of the query sequence were retained to avoid partial or domain-level hits. In addition, a minimum sequence identity of 25% was required, together with a minimum bitscore of 80 and an E-value threshold of $\leq 1 \times 10^{-10}$ to ensure high statistical significance. Analysis of biosynthetic gene clusters was performed using the NCBI Genome Viewer. For *Ca. Kuenenia stuttgartiensis*, analyses were based on the NCBI reference sequence NZ_LT934425.1, and for *Campylobacter jejuni* subsp. *jejuni* NCTC 11168 (ATCC 700819), on the reference sequence NC_002163.1.

Synthetic sugar standards and ^{13}C labelled metabolite extract. ^{13}C -labeled *Penicillium chrysogenum* extract was produced in-house as described previously³⁸. UDP-Glc was obtained from Sigma-Aldrich (Cat. No. U4625-100MG), UDP-GlcNAc from Sigma-Aldrich (Cat. No. U4375-100MG), CMP-Neu5Ac from BioSynth (Cat. No. MC04391), Pse5,7Ac₂ from BioSynth (Cat. No. MP07650), and Leg5,7Ac₂ from Synvenio (Cat. No. SV7243). The AdvanceBio Sialic Acid Reference Panel was obtained from Agilent (Cat. No. GKR-2503), and KDN from Sigma-Aldrich (Cat. No. 60714).

3.3 | Results and Discussion

3.3.1. SugarBase: untargeted exploration of nucleotide sugar networks in microbes.

The complexity of crude cell lysates, the chemical diversity of microbial nucleotide sugars, and the high data density of next-generation mass spectrometers require open searches and advanced data mining strategies. To address this, we developed a mass spectrometry workflow that combines porous graphitic carbon (PGC) separation, continuous narrow-window DIA scanning, and a nucleotide sugar discovery software (Figure 3.1). PGC allows analysis of crude lysates while providing excellent isomer separation of polar compounds. Narrow mass windows (over an m/z range of 500–708, negative mode) are sequentially scanned to maximize sensitivity for phosphate-containing nucleotide sugars. Each window is acquired twice: first at low collision energy to generate a focused pseudo-MS1 spectrum of precursor ions, and then at higher collision energy to obtain corresponding fragmentation spectra (“true MS2”). Narrow isolation windows in fast mass spectrometers reduce spectral complexity by limiting the number of co-fragmented features. In microbial systems, the nucleotide moiety is known which can be either CMP, UDP, GDP, ADP, CDP, dTDP or TDP, whereas the attached sugar is often unknown. To identify such precursors, we developed an algorithm that links nucleotide marker fragments to the correct MS1 precursor ions by matching against a large theoretical sugar composition database containing >2200 entries, allowing assignment of the sugar and nucleotide components (Figure 3.1, SI Excel 2, SI DOC). To avoid false positives, we implemented a range of inclusion and scoring criteria as detailed in the SI DOC. For example, nucleotide sugars and their marker fragments were expected to be present in at least three consecutive scans within the same isolation window, and co-elution of precursor and marker fragment was confirmed using Pearson correlation. Additionally, precursor intensity, and fragment-to-precursor abundance ratios were expected to fall within empirical ranges. For CMP-nonulosonic acids, additional scoring criteria were included, including the presence of CO_2 loss (from the carboxyl group) and consistency with the nitrogen rule, where the nominal mass reflects the expected number of nitrogen atoms.

The SugarBase tool visualizes identified nucleotide sugars in a browser-based dashboard, which enables interactive exploration of candidate features, scores, mass traces, and extracted ion chromatograms. The SugarBase open-source Python code, standalone desktop application, and documentation are freely available via SourceForge. While the current pipeline was developed and tested using Q Exactive Orbitrap (Thermo Fisher Scientific) data, all parameters are adjustable to support other acquisition methods and vendors, provided raw files can be converted to mzXML format and meet the required input structure.



Figure 3.1. SugarBase: mass spectrometry and bioinformatics workflow for untargeted exploration of microbial nucleotide sugar networks. a) Metabolite extracts are analyzed by porous graphitic carbon separation coupled to narrow-window DIA mass spectrometry. Each window is first acquired at low collision energy to record intact precursor ions (pseudo-MS1), followed by higher energy to obtain fragment spectra (MS2). Pseudo-MS1 spectra are deisotoped and filtered, while MS2 spectra are screened for nucleotide marker fragments. Matching scans are linked to identify nucleotide sugars, and precursor ions are assigned using a sugar composition database. Candidates are evaluated based on consecutive detection, co-elution of precursor and marker ions, realistic intensity ratios, and for CMP-linked sugars, CO_2 neutral loss and the nitrogen rule. Identifications can

be explored via a browser-based Dash interface, of which a screenshot as illustration is displayed in the figure. Structures of CMP, CDP, UDP, dTDP, TDP, ADP, and GDP are shown (reported as $[M-H]^-$ ions). The histogram below shows the theoretical composition space of nucleotide-linked sugars, generated from all plausible $C_{5-16}H_{8-27}N_{0-4}O_{3-9}S_{0-1}$ combinations, yielding >2200 candidates. These are used to annotate precursor ions when a nucleotide marker is detected. Finally, an example of a nucleotide-sugar as precursor for glycoconjugate is illustrated, showing the transfer of Legionaminic acid onto the flagella of *Campylobacter jejuni*.

3.3.2. Highly divergent nucleotide sugar profiles across microbes.

To demonstrate the SugarBase workflow we mapped the sugar profiles across a broad phylogenetic spectrum of microbes, encompassing environmental, pathogenic, and biotechnologically relevant taxa (SI Excel 1), including *Campylobacter*, *Pseudomonas*, *Escherichia*, *Streptomyces*, *Magnetospirillum*, *Legionella*, anammox bacteria, archaea, and multiple enrichment cultures. We furthermore included well-studied eukaryotic samples as controls, for example, mammalian cells were included as these produce a defined set of nucleotide sugars, and plant species were included as these are devoid of CMP-NuLOs.

Interestingly, for prokaryotes, the principles that govern sugar-type usage and nucleotide activation preferences remain largely unknown³⁹. Therefore, we first systematically screened these microbes for the presence of a systematic matrix of nucleotide/sugar combinations. We generated this matrix of nucleotide sugars by combining common sugar chain lengths with typical modifications (oxidation, deoxygenation, and amination) across all major nucleotides (UDP, ADP, GDP, CDP, CMP, dTDP, and TDP). Nonulosonic acids were paired exclusively with monophosphate nucleotides (UMP, AMP, GMP, CMP, dTMP, and TMP), as these form only stable bonds with monophosphates¹⁶. This produced a matrix of 203 nucleotide sugar combinations (SI Excel 2), which was used for targeted searches with the SugarBase pipeline, applying default filtering and scoring parameters without abundance cutoffs.

This provided a view on nucleotide sugar usage and combination across a broad phylogenetic spectrum. Of the 203 theoretical nucleotide sugars, 33 were detected in more than one sample (>50 when including single detections, Figure 3.2, SI Excel 3 & 4, SI DOC). The mammalian and yeast controls showed the expected set of sugars and activating nucleotides, such as abundant UDP-hexose, UDP-HexNAc, GDP-hexose, GDP-deoxyhexose, dTDP-hexose, ADP-pentose, ADP-deoxyhexose, and CMP-Neu5Ac in mammalian cells (right columns, Figure 3.2a). The abundant ADP-pentose peak likely represents ADP-ribose derived from nicotinamide cofactors, that is involved in ADP-ribosylation and DNA repair rather than glycomolecule biosynthesis.

In contrast, prokaryotes displayed a diverse spectrum of nucleotide sugars with many species-specific features³⁶. Generally, ADP- and UDP-linked sugars dominated, dTDP-linked sugars were also frequently observed, but CDP activation was only observed in a few cases (Figure 3.2). Pentoses were predominantly ADP-linked, while hexoses showed greater heterogeneity and occurred as UDP-, ADP-, GDP-, and dTDP-linked sugars (Figure 3.2c). The different nucleotides may partition the carbon flux into different biosynthetic glycan pathways, as glycosyltransferases are selective for the linked nucleotide^{15,40}. Nonulosonic acids were found exclusively linked to CMP, with no significant evidence for linkage to other nucleotides (Figure 3.2c). While UDP-Hex, UDP-HexNAc, and GDP-Hex are expected to be universal metabolic precursors, in prokaryotes they were not always the most abundant and in some cases were not detected at all. Compared to eukaryotes, this variability likely reflects the greater metabolic flexibility of microbes and the diversity of carbon sources across their environments.

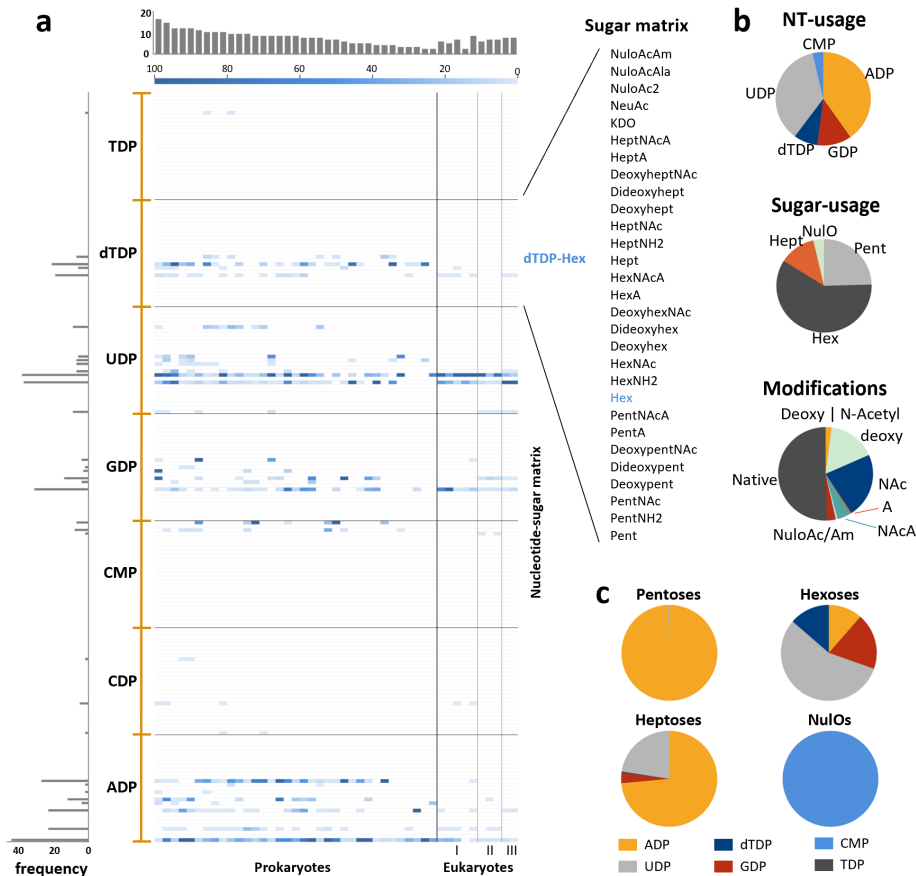


Figure 3.2. Systematic survey for nucleotide sugar combinations across a broad phylogenetic spectrum. a) The heatmap shows the outcome of a targeted search using a systematic matrix of

nucleotide sugar combinations across a broad phylogenetic spectrum of microbes, with mammalian and yeast samples included as controls. A matrix of 203 nucleotide sugars combined common pentose, hexose, heptose, and nonulosonic acid sugars and their modifications to the seven activating nucleotides UDP, ADP, GDP, CDP, CMP, dTDP, and TDP. The monosaccharide classes of the detected nucleotide sugar candidates are displayed on the y-axis, and the color scale represents normalized peak intensities. The bar graph on the left of the heatmap shows the detection frequency of individual nucleotide sugars across all microbes, while the bar graph on the top shows the number of identified nucleotide-sugars per organism. I: yeast; II: mammals; III: plants. B) The pie charts show the nucleotide activation preference, the sugar class usage and the modification usage across all analyzed samples. c) The pie charts show nucleotide activation preferences for each sugar type (pentoses, hexoses, heptoses, and nonulosonic acids) across all analyzed samples.

3.3.3. Deep untargeted mapping of nucleotide sugars in pathogens and microbial model organisms.

Next, we employed SugarBase to perform a fully untargeted profiling to also capture previously uncharacterized nucleotide sugars and to explore CMP-NuLO modification profiles in common pathogens. Thereby we identified 87 recurring nucleotide sugars (using a score cutoff of ≥ 20 and detection in at least two species, Figure 3.3a, SI Excel 5). Notably, 10 of the 25 most frequent features could not be assigned to known nucleotide sugars. However, we also observed many recurring background features, including AMP and AMP-pentose related fragments, likely originating from in-source fragmentation or degradation of NAD- and CoA-derived metabolites. All commonly observed background feature are summarized in SI Excel 3. Overall, the most frequent nucleotide sugars (detected in >5 organisms) included central precursors for glycan and cell envelope biosynthesis, such as UDP-HexNAc, UDP-Hex, GDP-Hex, UDP-MurNAc, ADP-heptose, ADP-deoxypentose, ADP-hexose, dTDP-deoxyhexose, GDP-deoxyhexose, dTDP-hexose, UDP-pentose, and various CMP-NuLOs. Notably, several less studied but abundant species including CDP-Hex, UDP-deoxyHexNAc, UDP-Hexdi/triNAc, and HexNAc derivatives linked to dTDP, ADP, or GDP, were also observed, which shows the metabolic flexibility in microbes. Furthermore, Figure 3.3b zooms in on microbes with biotechnological relevance. The model unicellular eukaryote and cell factory *S. cerevisiae*, showed mostly UDP-HexNAc, followed by UDP-Hex, GDP-Hex, ADP-ribose, and some lower abundant dTDP-hexose derivatives, in line with known fungal metabolic routes (Figure 3.3b). Similarly, the embryonic kidney (HEK) cells, a well-characterized mammalian model, showed nucleotide sugars consistent with the conserved eukaryotic glycosylation pathways (SI Excel 5)^{12,41}. In contrast, bacterial samples showed much greater heterogeneity, both within and between species. The pathogen *Campylobacter jejuni* produces abundant GDP-heptose and

multiple CMP-linked sugar derivatives alongside UDP-HexNAc and UDP-dHexNAc, consistent with bacillosamine-containing N-glycan and lipooligosaccharide biosynthesis pathways^{42,43}. *Ca. Kuenenia stuttgartiensis*, an anammox bacterium involved in global nitrogen cycling and wastewater treatment³³ displayed multiple CMP-linked sugars, UDP-glucuronic acid, UDP-HexNAcA, UDP-pentose, ADP- and GDP-linked heptoses, and UDP-MurNAc. Interestingly, *Ca. Kuenenia stuttgartiensis* activates heptose with both ADP and GDP, where the ADP-heptose is used in lipopolysaccharide biosynthesis⁴⁴, and GDP-heptose likely relates to the recently reported oligoheptosidic glycans modifying the surface layer³³. The magnetotactic bacterium *Magnetospirillum gryphiswaldense*, a model platform in nanobiotechnology⁴⁵, showed one of the most distinctive nucleotide sugar profiles, with several previously uncharacterized CMP-linked sugars further described in the following section. A particularly complex spectrum was observed in the *Ca. Accumulibacter phosphatis* enrichment, which next to its mixed composition is also known for its metabolic versatility⁴⁶, producing unique compounds such as UDP-HexNAc2A, ADP-dHex, and dTDP-dHex along with UDP-HexNAcA and ADP-hexose. In many species we also detected abundant UDP-MurNAc and downstream UDP-muramyl peptide intermediates from peptidoglycan and cell wall assembly⁴⁷. Interestingly, the analysed microbes frequently showed CMP-, GMP-, or ADP-linked sugars that were more abundant than the otherwise central UDP-Hex and UDP-HexNAc derivatives. Several nucleotide sugars with high confidence scores could not be annotated, highlighting the unexplored diversity of microbial metabolism.

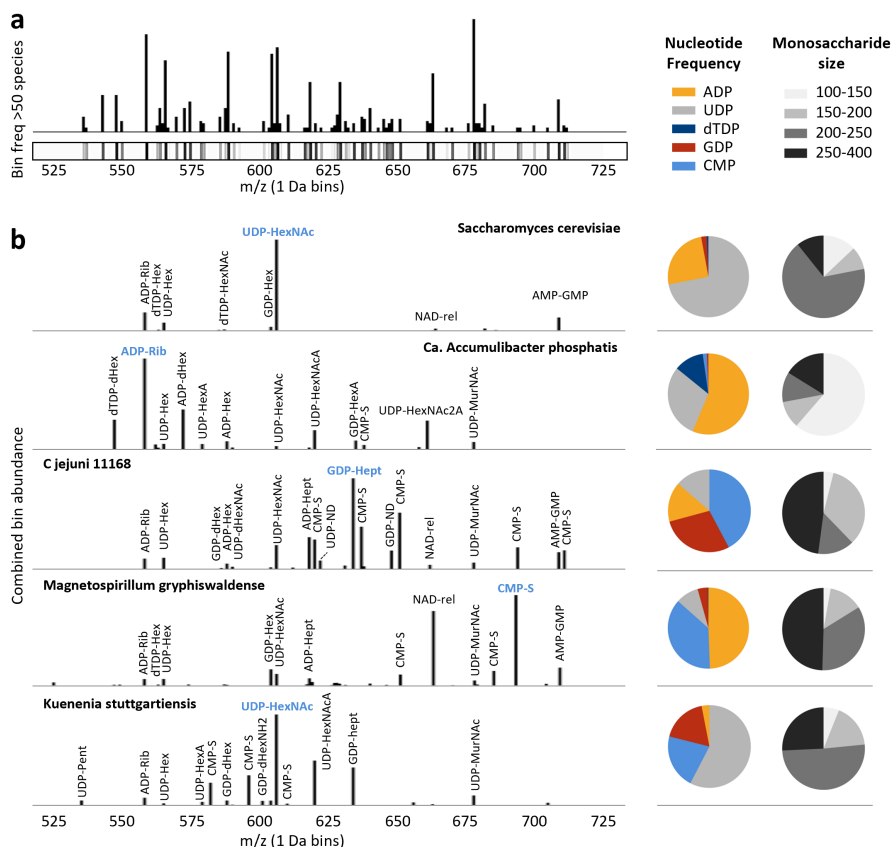


Figure 3.3. Fully untargeted SugarBase workflow shows diverse nucleotide sugar profiles across pathogens and biotech strains. a) The frequency histogram and heatmap show all nucleotide sugars identified across the analyzed microbes using the fully untargeted SugarBase workflow. b) The histograms below show nucleotide sugar profiles for individual species, including *Saccharomyces cerevisiae*, *Ca. Accumulibacter phosphatis*, *Campylobacter jejuni* NCTC 11168, *Magnetospirillum gryphiswaldense*, and *Ca. Kuenenia stuttgartiensis*. The pie charts next to each spectrum indicate the relative abundance of nucleotide types: ADP, UDP, CDP, (d)TDP, GDP, CMP and monosaccharide size classes: 100–150, 150–200, 200–250, and 250–400 Da. A summary table of all identified nucleotide sugar candidates across species, along with detailed tables for each organism, is provided in SI Excel 5. Abbreviations: CMP-S, CMP-sugar; dHex, deoxyhexose; dHexNAc, N-acetyl-deoxyhexose; Hept, heptose; Hex, hexose; HexA, hexuronic acid; HexNAc, N-acetyl-hexose; HexNAcA, N-acetyl-hexuronic acid; HexNAc2A, di-N-acetyl-hexuronic acid; MurNAc, N-acetyl-muramic acid; NAD-rel, NAD-related compound; ND, not defined; Pent, pentose; Rib, ribose.

3.3.4. Mapping CMP-NuLO modification profiles in pathogens and biotechnologically relevant microbes.

Nonulosonic acids (NuLOs) are large nine-carbon sugars containing a highly acidic α -keto acid moiety. Beyond their distinctive chemistry, they often cap glycoconjugates and fulfil key biological functions. In humans, NuLOs regulate

protein half-life, and immunity⁴⁸, while in microbes they can interfere with phage binding⁴⁹ and are important virulence factors in pathogens, such as *Campylobacter jejuni* and *Helicobacter pylori*^{5,50}. Their analysis is challenged by extensive chemical diversification and instability. However, understanding their diversity and biosynthetic pathways provides promising targets for the development of novel antimicrobials.

SugarBase identifies CMP-NuLO derivatives based on diagnostic features: the CMP fragment, characteristic CO₂ loss from the acid group, strong CMP dissociation due to the labile monophosphate ester, and a distinct nitrogen-rich composition¹⁶. To further explore the identified NuLOs, another fraction of the metabolite extract was subjected to mild acid hydrolysis, labeled with 1,2-diamino-4,5-methylenedioxybenzene (DMB), and analyzed by DIA-MS as described by Kleikamp et al.³⁶. This confirmed the presence of the α -keto acid, the carbon backbone length (e.g., C8, C9, or larger), the redox state (e.g., Neu versus Leg/Pse), and the nature of N- and O-linked modifications.

As these sugars are key modulators of cellular function, we investigated the CMP-NuLO profiles of *Ca. Accumolibacter phosphatis*, *Magnetospirillum gryphiswaldense*, *Ca. Kuenenia stuttgartiensis*, and four *Campylobacter jejuni* strains. In *C. jejuni*, NuLOs are O-linked to the FlaA and FlaB proteins and are essential for flagellar assembly and motility, with specific modifications influencing host cell invasion²⁻⁵. The strain NCTC 11168 is a well-characterized, highly invasive reference strain and a common cause of human gastroenteritis, whereas strain 81116 is less invasive and was first isolated during a waterborne outbreak in the UK⁵¹⁻⁵³. NCTC 11168 produces both pseudaminic and legionaminic acid derivatives with a well-defined modification profile², whereas strain 81116 lacks the legionaminic acid biosynthesis pathway⁵⁴ and its NuLO structures have not been characterized before. Across all these strains we identified over twenty distinct CMP-linked sugars (Figure 3.4a, SI Excel 6), of which all but two could be confidently assigned to ulosonic acid derivatives. *C. jejuni* 11168 showed the greatest CMP-NuLO diversity, with at least eight distinct variants. Identified structures included NuLOAcAc and NuLOAcAm species, additional O-acetylation and methylation, 2,3-di-O-methylglyceroyl and methyl acetimidoyl modified (the latter has so far only been identified for Leg), consistent with previous reports for this strain². Notably, the acetimidoylamino (Am; CH₃C(=NH)-NH-) modification generated by the amidotransferase PtmG (Cj1324) predominated over the acetamido modification (Ac; CH₃C(=O)-NH-) (SI Excel 6). This likely has functional implications, as protonation of the imine at physiological pH yields a net-neutral sugar, whereas the acetamido form is negatively charged, and contributes to a

negative surface charge on proteins or cells. Interestingly, in addition to Pse/Leg acetimidoylamino modifications, we also observed peaks consistent with acetimidoylamino modified Neu. Notably, the nucleotide sugar pool predominantly contained the putative Neu5Am form, whereas after acid hydrolysis and DMB labeling, Neu5Ac was the predominant species. A similar shift from Am to Ac was also observed for Pse and Leg derivatives. This shift likely originates from reacylation during acetic acid treatment and the amidotransferase may also preferably act on nucleotide sugars. Interestingly, we observed that acetimidoyl modifications may also affect other N-acetylated sugars, such as UDP-HexNAc, which has not been reported previously. However, their potential incorporation into glycoconjugates was not further investigated in this study.

For *C. jejuni* 81116, the CMP-NulO modification profile has not been established previously. Using SugarBase, we observed a comparable structural diversity as for NCTC 11168, but here a single structure (m/z 680.1934, $[M-H]^-$) accounted for approx. 85% of the total signal (Figure 3.4a & b; SI DOC). However, the inferred NulO moiety mass and sum formula (358.1488 Da, $C_{14}H_{22}N_4O_7$) cannot be explained by simple acetylation or methylation of the canonical NulO backbone. The orthogonal DMB-NulO fragmentation revealed C9 Leg/Pse marker ions, an N-acetyl or acetimidoyl group, and an additional neutral loss of 102.04 Da, consistent with an N-modification of formula $C_3H_6N_2O_2$. Given the close genomic relatedness to NCTC 11168 and the observed mass and elemental composition, this modification likely represents a glyceroyl substituent on an acetimidoyl group, as present in NCTC 11168 (possibly lacking O-methylation and with one hydroxyl oxidized to an aldehyde, or a closely related variant). Further minor CMP-NulOs that were detected correspond to canonical NulOAcAm, O-acetylated, and dehydrated forms. Interestingly, two low-abundance species (<5% combined) were consistent with NulOAmNH₂, featuring a free amine, including one variant lacking the N-acetyl group, likely arising from incomplete N-acetyltransferase activity or degradation. NCTC12658 showed a CMP-NulO profile similar to 11168, including both smaller NulOAcAm and larger more hydrophobic 2,3-di-O-methylglyceroyl modified structures that were absent in NCTC12662. Strain 12662 showed a previously identified NulO monoisotopic mass of 330.154, likely corresponding to substitution of a hydrogen atom by an amino group ($-NH_2$)⁴⁹. Strain 12658 displayed an additional NulO mass of 333.106, which could not be annotated based on known structures, nor structurally elucidated from the available data. Both *C. jejuni* strains have also been investigated for their susceptibility to different phages⁵⁵. Flagellotropic phages target the flagellum and require bacterial motility, where receptor recognition depends on flagellin O-glycosylation and thus may be

sensitive to presence and type of NulOs^{49,55-57}. Flagellotropic phages are of particular interest because by binding they attenuate motility and consequently virulence, making these phages especially promising antibacterial agents⁵⁷.

The CMP-sugar profile of *Ca. Accumulibacter* was dominated by a single CMP-NulOAc2 species, whereas *M. gryphiswaldense* showed a surprisingly broad CMP-sugar profile (Figure 3.4a). In addition to low-abundant NulOAcAm and acetylated/methylated derivatives, we detected several large CMP-sugars ($[M-H]^-$ m/z 679.1983, 693.2139, and 707.2296). These masses were initially consistent with additional acetylation and methylation, and efficient fragmentation with characteristic CO_2 loss confirmed their identity as CMP-linked α -keto acids. However, fragmentation of the DMB-labeled compounds showed a different image. While the NulOAcAm and NulOAcAmOMe derivatives showed the expected C9 Pse/Leg marker fragment m/z 297.09, the larger species yielded the m/z 311.10 marker fragment, indicative of an extended carbon backbone (Figure 3.4a & b, SI DOC). This observation is consistent with earlier reports by Kleikamp et al., who described C10 marker fragments in *M. gryphiswaldense* ulosonic acids that could not be assigned to known structures³⁶. Therefore, we established a fragmentation tree for the observed marker fragments (SI DOC), which showed sequential loss of water and ammonia, followed by one O-acetyl and two N-acetyl groups, ultimately yielding the m/z 311.10 carbon length marker fragment. Further isolation and MS³ fragmentation of m/z 311.10 revealed an initial CO/H_2O loss, followed by stepwise cleavage of four carbon atoms to the m/z 229.06 core fragment (DMB-keto acid). In comparison, MS³ fragmentation of the DMB-LegAc₂ m/z 297.09 marker and the DMB-Neu5Ac marker m/z 295.09 showed the same initial CO/H_2O loss, followed by cleavage of only three carbon atoms to the same m/z 229.06 core fragment as observed in *M. gryphiswaldense*. This suggests the presence of an elongated carbon backbone and a previously unreported α -keto acid in *M. gryphiswaldense*. Moreover, O-methylation would be lost early during fragmentation, before extensive C–C bond cleavage. Interestingly, lower-abundance CMP-linked species with unusually low masses (m/z 203 and 217), consistent with HexNAc-related compounds, were also detected in *M. gryphiswaldense*, which furthermore were not detected in DMB-labelling experiments confirming the lack of an α -keto acid moiety.

Finally, we investigated *Ca. Kuenenia stuttgartiensis*, which was recently shown to express complex O-glycans containing NulOs with a free amino group. Like acetimidoyl modifications, free amines are protonated at physiological pH, yielding a net-neutral sugar despite the strong acidic α -keto group. Free amines have previously been reported only in certain cancer cells, where they influence cell

surface charge and aggregation⁵⁸. Consistent with these observations, the SugarBase workflow identified NulOs containing one free amine and a second amine that was formylated (rather than acetylated), and CMP-NulOs with +14.01 Da consistent with O-methylation (Figure 3.4a & b, SI DOC). The presence of these modifications in CMP-activated sugars show that deacetylation and formylation occur at the nucleotide sugar level and not after glycan biosynthesis. This prompted a closer examination of the *Ca. Kuenenia stuttgartiensis* genome to identify the gene cluster that enables the production of CMP-NulOs with an unsubstituted amine. We searched the genome for pseudaminic and legionaminic acid biosynthetic genes in proximity to enzymes such as deacetylases and methyl- or formyl-transferases. Indeed, we found a legionaminic acid-related biosynthetic gene cluster with putative legB, legF, and legC homologs, together with genes encoding a deacetylase and a formyltransferase, followed by putative neuB, legF, and legG genes (Figure 3.4c). This biosynthetic pathway intrinsically incorporates deacetylation and formylation steps. Although overall sequence homology is low, a similar genomic architecture is present in *C. jejuni* NCTC 11168, where ptmG and ptmH between legC and neuB mediate conversion of N-acetyl groups to acetimidoyl moieties and subsequent methylation (with the N-acetyltransferase legH located near the pseudaminic acid locus).

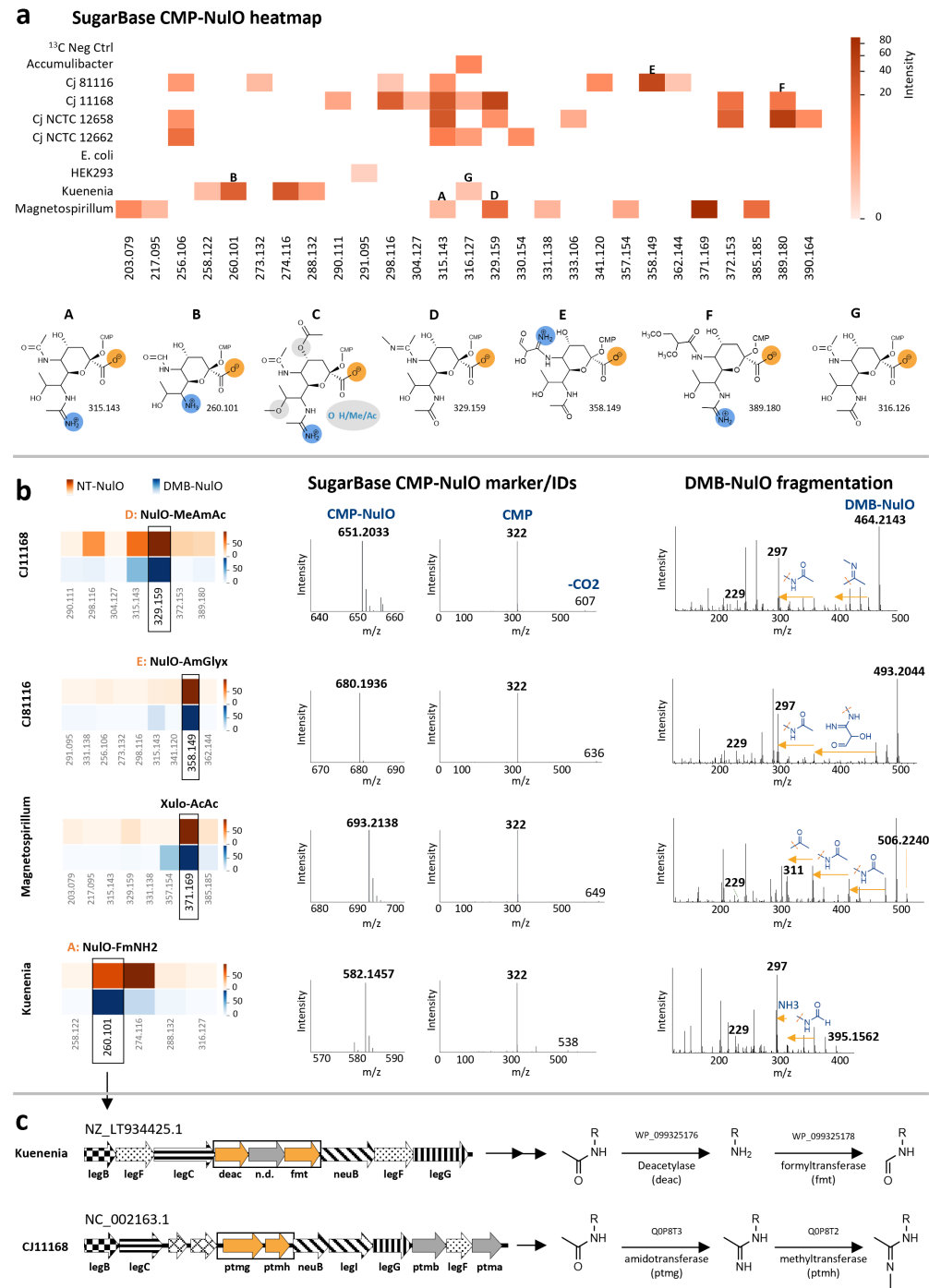


Figure 3.4. SugarBase CMP-NuLO profile in pathogens and environmental microbes. **a)** The heatmap shows the SugarBase CMP-NuLO profile across pathogenic *C. jejuni* and environmental microbes. The monoisotopic mass of identified NuLOs are shown on the x-axis, and colors indicate normalized peak intensity. The schemes below present selected CMP-NuLOs with their putative

structures; no structure could be assigned for m/z 333. **b)** The graphs show SugarBase CMP-NuLO identification and structural characterization using DMB labeling and fragmentation for the dominant NuLO structures in *Campylobacter jejuni* 11168, *Campylobacter jejuni* 81116, *Magnetospirillum gryphiswaldense*, and *Ca. Kuenenia stuttgartiensis*. The middle panel shows SugarBase CMP-NuLO pseudo-MS1 signals with characteristic CMP and CO₂-loss fragments in MS2. The right panels show DMB labeling and fragmentation, confirming the ketoacid group, carbon-chain length and oxidation state, and modifications, with putative structures annotated. The left heatmaps compare the abundance of the CMP-NuLOs and corresponding DMB-NuLOs, with color indicating normalized intensity. **c)** The identified NuLO type and modification in *Ca. Kuenenia stuttgartiensis* enabled identification of the corresponding biosynthetic gene cluster (genome NZ_LT934425.1). Notably, genes encoding a deacetylase and a formyltransferase, likely responsible for its unique modifications, are located between homologs of legB, legF, and legC, and putative neuB, legF, and legG genes. For comparison, the Leg biosynthetic pathway and methylation step from *C. jejuni* NCTC 11168 (genome NC_002163.1) are shown.

3.3.5. SugarBase enables discrimination between strains that produce pseudaminic or legionaminic acid.

Distinguishing nonulosonic acid (NuLO) analogues with identical composition but different stereochemistry remains a major challenge. Among known NuLOs, pseudaminic and legionaminic acids are the most common stereoisomers, while other variants are rarely reported⁴⁸. Recently developed monoclonal antibodies specifically recognize Pse-containing glycoconjugates which allowed staining and glycoproteomic detection of Pse containing glycans⁵⁹.

However, broader studies of this group of NuLO molecules still require NMR and predictions based on genomic context. This is particularly challenging for hard-to-culture organisms and species with poorly annotated genomes.

Pseudaminic and legionaminic acids differ in stereochemistry at multiple chiral centers and are synthesized via distinct biosynthetic pathways^{21,22,60}. Pseudaminic acid is derived from UDP-GlcNAc through sequential dehydration, epimerization, transamination, and acetylation steps. Nucleotidase activity then produces 2,4-diacetamido-2,4,6-trideoxy-L-altropyranose, which is condensed with phosphoenolpyruvate by Pse synthase to form Pse5Ac7Ac. This product is subsequently activated by the cytidyltransferase PseF to generate CMP-Pse5Ac7Ac (Figure 3.5b, SI DOC). Legionaminic acid biosynthesis follows a similar pathway but begins with GDP-GlcNAc (Figure 3.5b), enabling metabolic partitioning and independent regulation of the two pathways^{21,22,60}.

We used SugarBase to investigate the presence of Pse- and Leg-specific nucleotide sugar intermediates, which would indicate which of these compounds are produced. *C. jejuni* strain 11168 produces both Pse and Leg, whereas strain 81116 produces only Pse. Consistent with this, we observed nucleotide sugar

intermediates for both Pse and Leg in *C. jejuni* NCTC 11168, while we only detected Pse-related intermediates in strain 81116 (Figure 3.5a, upper heatmap). We next examined *Ca. Accumulibacter phosphatis*, for which homology-based predictions and proteomics data provide only ambiguous evidence regarding the presence of Pse, Leg, or both pathways (Figure 3.5a, lower heatmap). This ambiguity largely reflects the low sequence homology of the associated enzymes across taxa. In contrast, nucleotide sugar intermediates show uniform chemical structures because modifications occur at the NulO product level. When investigating *Ca. Accumulibacter* with SugarBase, we detected only key intermediates of the Pse pathway (Figure 3.5a).

To support this annotation, we performed acid hydrolysis and DMB labeling of *Ca. Accumulibacter* biomass, followed by separation and fragmentation analysis together with commercial Pse5Ac7Ac and Leg5Ac7Ac standards. Pse and Leg showed partial chromatographic separation and distinct fragment ion ratios, which confirmed that the DMB-NulO peak in *Ca. Accumulibacter* was Pse5Ac7Ac (Figure 3.5c, SI DOC). Here, separation and fragmentation of labelled NulOs were possible because only doubly acetylated forms were present and standards were available. In summary, SugarBase enables differentiation between Pse- and Leg-producing strains based on key nucleotide sugar intermediates, without relying on genomic predictions or the need for large-scale fractionation for NMR studies. This provides a new strategy to profile microbial systems for Pse and Leg production.

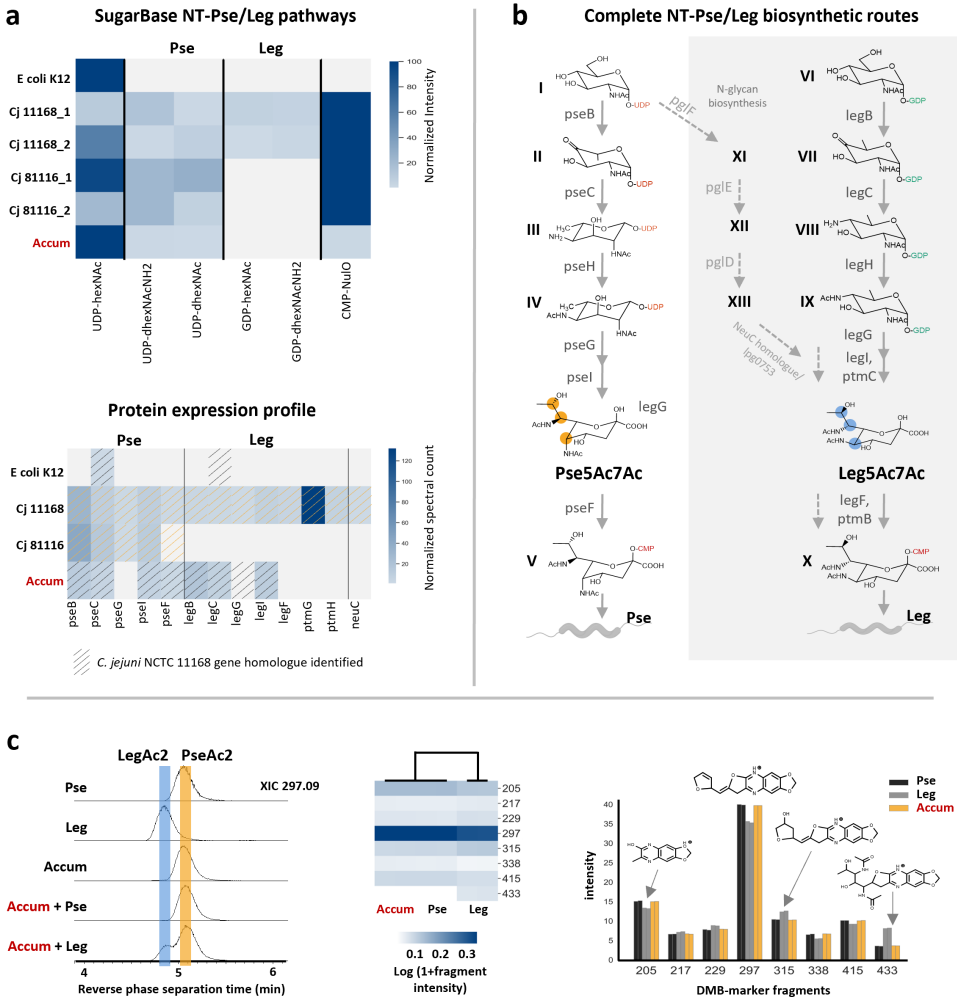


Figure 3.5. SugarBase discriminates between Pse and Leg producing strains. a) The upper heatmap shows SugarBase screening for key nucleotide sugar intermediates that distinguish Pse- and Leg-producing strains. For Pse, these include UDP-dHexNAcNH₂ and UDP-dHexNAc; for Leg, GDP-HexNAc and GDP-dHexNAcNH₂, shown together with the UDP-HexNAc precursor and CMP-NulO products. From top to bottom: *E. coli* K12, *C. jejuni* 11168 (duplicates 1–2), *C. jejuni* 81116 (duplicates 1–2), and *Ca. Accumulibacter phosphatis*. *E. coli* serves as a negative control, *C. jejuni* 81116 as a Pse-positive control, and *C. jejuni* 11168 as a Pse/Leg-positive control. SugarBase detects key intermediates in *Ca. Accumulibacter phosphatis* only for the Pse pathway. The lower heatmap shows proteomics data for the same strains, confirming expression of enzymes involved in Pse5Ac7Ac and Leg5Ac7Ac biosynthesis. Abundance is indicated by normalized spectral counts. The orange hatched pattern marks genes supported by genome annotation and literature^{21,22,54}, while grey hatching indicates genes identified by sequence alignment to *C. jejuni* NCTC 11168 pathways. b) The schemes show biosynthesis pathways for Pse5Ac7Ac (from UDP-GlcNAc) and Leg5Ac7Ac (from GDP-GlcNAc). Intermediates are labeled I–XIII, covering key UDP- and GDP-linked sugar derivatives and final CMP-activated products. c) The left graph shows reverse-phase separation of DMB-Pse5Ac7Ac (Pse) and DMB-Leg5Ac7Ac (Leg), monitored by the C9 fragment (m/z 297.09). DMB-labeled *Ca. Accumulibacter*

phosphatis (Accum) was analyzed with and without spiked standards. The Accum NulOAc2 co-elutes with the Pse marker. The heatmap (cosine similarity clustering) groups Accum fragments with the Pse standard. The bar plot shows diagnostic fragments that clearly distinguish Pse from Leg.

3.4 | Conclusion

SugarBase enables fully untargeted mapping of microbial nucleotide sugar networks by screening crude cell extracts using porous graphitic carbon separation, narrow-window DIA, and a new algorithm with a large composition database comprising thousands of theoretical entries. This approach links nucleotide marker fragments to their corresponding nucleotide sugar-parent compounds. We also provide a comparative overview of monosaccharide classes and nucleotide activation preferences across a broad phylogenetic range of microbes, revealing extensive species-specific features and many high-confidence identifications that remain unannotated.

This study provides a quantitative map of the *C. jejuni* NulO repertoire and supports the identification of biosynthetic gene clusters, for example in *Ca. Kuenenia stuttgartiensis*, which produces NulOs with a free amine. Integration of orthogonal DMB-based fragmentation confirmed different nonulosonic acid classes, including N- and O-linked modifications that influence physicochemical properties relevant to microbial biology. Distinct NulO profiles in *Campylobacter* species may also affect susceptibility to flagellotropic phages. Notably, we also report the first evidence of previously undescribed CMP-activated higher-carbon ulosonic acids in *Magnetospirillum*. Finally, pathway-specific nucleotide sugar intermediates enabled discrimination between pseudaminic- and legionaminic-acid-producing microbes without requiring genomic information. This was supported by comparative chromatographic separation and fragmentation experiments using synthetic Pse and Leg standards.

Together, SugarBase provides a scalable approach to discover new sugar pathways and enzymes, accelerating glycobiology in non-model microbes and enabling access to chemically challenging nucleotide sugars for antimicrobial and vaccine development. The SugarBase pipeline is freely available as Python code and as a standalone desktop executable.

3.5 | Data and code availability

The SugarBase pipeline is freely available as Python code and as a standalone desktop executable with a browser-based graphical user interface: <https://sourceforge.net/projects/sugarbase-x/files>. All acquired data are publicly

accessible through the 4TU.ResearchData platform (<https://data.4tu.nl> project “SugarBase”). MATLAB functions are available upon request.

3.6 | Supplementary information

3.6.1. SugarBase QuickStart and Overview

SugarBase is designed to process LC-MS/MS data (in .mzXML format) to identify nucleotide-activated sugars (NT-sugars) based on diagnostic nucleotide fragments and a predefined chemical composition space. The workflow combines pseudo-MS1 precursor information with MS2 fragment evidence, filters candidate masses using theoretical chemical compositions, and assigns a confidence score to each detected hit.

The pipeline supports both a discovery mode (screening a broad chemical space of possible nucleotide-sugars) and a targeted mode (searching only predefined compositions from an input list). The desktop executable allows to explore the results interactively, by providing the extracted ion chromatograms as well as the pseudo-MS1 and MS2 scans in which the potential NT-sugar was detected for each nucleotide-sugar hit. In addition, results can be exported as an Excel table.

Key analytical steps include MS1 preprocessing (deisotoping and filtering), MS2 fragment extraction, MS1-MS2 alignment, precursor matching against a theoretical chemical sugar space, detection of consistent signals across consecutive scans, and multi-parameter scoring to rank candidate nucleotide-sugar hits. SugarBase can be used to discover the nucleotide-sugar profile in microbial extracts without requiring any prior knowledge.

QuickStart

1. Download the desktop executable:
https://sourceforge.net/projects/sugarbase-x/files/SugarBase_v08032026.zip/download
2. Open the desktop executable, see Figure S3.1.
3. Download the test data:
https://sourceforge.net/projects/sugarbase-x/files/Test_data_SugarBase.zip/download
4. Select your input folder (Figure S3.1 → 1). A maximum of three .mzXML files is allowed per analysis.
5. Select your output folder (Figure S3.1 → 2).
6. Enter a sample name (Figure S3.1 → 4).
7. Select the analysis mode (Figure S3.1 → 5).
 - a. 'Targeted' -> requires a '.xlsx' file which should include the chemical formulas of the sugars you want to target in the first column, e.g. C₆H₁₂O₆ (Figure S3.1 → 3).
 - b. 'Discovery' -> no Target.xlsx file is required.

8. To run the analysis with default parameters, click "Run Analysis". For a description of the input parameters, the reader is referred to the caption of Figure S3.1 or to the ReadMe file as provided on SourceForge (<https://sourceforge.net/projects/sugarbase-x/files/>).

Figure S3.1. Screenshot of the SugarBase desktop executable. For all parameters the default values are as shown in the Figure. Analysis mode: can be set to discovery (screen a broad chemical space of possible nucleotide sugars) or Targeted (only search for predefined compositions from an input list as defined in Select Targets ('xlsx')). GENERAL PARAMETERS. Consecutive Scans represents a filter parameter: hits representing potential nucleotide-sugars are only kept when the identified hit is present in at least X consecutive scans of the MS data. Min MS1 Intensity: a minimum intensity threshold for peaks identified in pseudo-MS1 spectra. Min Retention Time (min): a lower retention time cutoff (in minutes) for pseudo-MS1 data. Max Retention Time (min): an upper retention time cutoff (in minutes) for pseudo-MS1 data. Max Fragment Error (ppm): the mass accuracy (in ppm) for fragment ion searches. Max Sugar Mass Error (ppm): mass accuracy (in ppm) for MS1 data matching to the

chemical sugar space. DIA Scan Cycle: Represents one PRM cycle. An 'DIA Scan Cycle' of 30 means that the same mass is targeted every 30th scan, with consecutive scans at i , $i+30$, $i+2*30$, etc. ADVANCED PARAMETERS. Enable Deisotoping: whether to perform deisotoping of the pseudo-MS1 spectra. Mass Error (ppm): mass accuracy (in ppm) used for deisotoping. Charge States: charge state to consider when removing isotopic envelopes (e.g. 2 = remove up to doubly charged envelopes). Cons Score Threshold: a scoring parameter. A hit (potential nucleotide-sugar) scores +10 when at least 'Cons Score Threshold' consecutive scans are present. Otherwise, the score is 0. Pearson High threshold: a scoring parameter. A hit scores +10 if the Pearson correlation of the precursors and fragments across three consecutive scans > 'Pearson High Threshold'. Pearson Low Threshold: a scoring parameter. A hit scores -10 if the Pearson correlation of the precursors and fragments across three consecutive scans < 'Pearson Low Threshold'. Note, the 'Pearson Low Threshold' should be ≥ 0.50 , since a score of -50 is assigned when the Pearson correlation < 0.50. Moreover, 'Pearson High Threshold' cannot be set smaller than 'Pearson Low Threshold'. MS2 Peak Fraction Score: a scoring parameter. A hit scores +10 if the intensity ratio of the MS2 fragment compared to the "MS1" precursor > 'MS2 Peak Fraction Score' AND < 100. A score of -50 is assigned when the ratio > 110. MS1 Intensity Score: a scoring parameter. A hit scores +10 if the MS1 Intensity of the precursor > 'MS1 Intensity Score'. Otherwise, the score is 0. Apply Nitrogen Rule: Enables filtering of CMP-hit data using the nitrogen rule. According to the nitrogen rule, a valid CMP-NuLO hit must have monosaccharides with both N and H atoms either even or odd in number; all others are excluded. Precursor Ratio Cutoff (%): Enables filtering of CMP-hit data based on the intensity ratio of the precursor identified in the MS2 scan compared to the fragment intensity in the MS2 scan. The red outlined numbers 1-5, correspond to references in the text section 'QuickStart'.

3.6.2. SugarBase interactive dashboard interpreter

After running an analysis in SugarBase a tabular output will appear (Figure S3.2), presenting the identified nucleotide-sugar hits with various output parameters, such as total score, NT-Sugar and Sugar Mass. A maximum total score of 50 can be obtained, with 50 being very likely that the candidate is indeed a nucleotide-sugar. For a full list of output parameters, the reader is referred to the caption of Figure S3.2 or the ReadMe file provided on SourceForge (<https://sourceforge.net/projects/sugarbase-x/files/>). The tabular output can be exported as an Excel file by using the 'Export' function. By selecting one of the identified hits in the table, four graphs will appear, which allow for interactive exploration of the data. These graphs include: i) The extracted ion chromatograms of the precursor ion and the corresponding nucleotide fragment ion; ii) The three consecutive scans used to identify a putative nucleotide-sugar. Both the ion intensity of the precursor ion and of the fragment ion are displayed; iii) The MS2 spectrum of the identified hit; iv) The pseudo-MS1 spectrum of the identified hit. The graphs can be explored by using the cursor to zoom in and out.

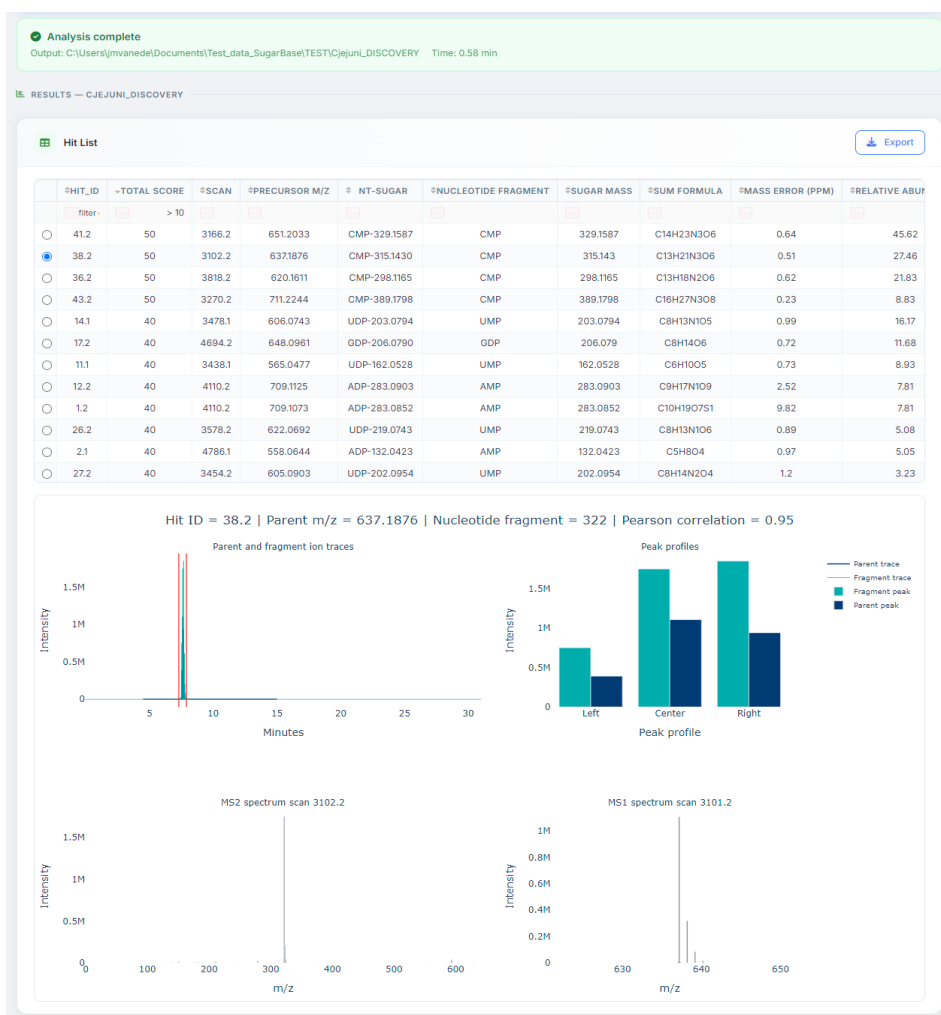


Figure S3.2. Screenshot of the results of a *C. jejuni* 11168 sample analyzed by the SugarBase desktop executable in discovery mode using default parameters. The Hit List provides a tabular overview of all the identified nucleotide-sugar candidates, which can be exported as an excel file using the 'Export' function. The table can be explored interactively by selecting one of the hits. This will result in four graphs appearing underneath the table output. The top left graph shows the extracted ion chromatograms of the precursor ion in the pseudo-MS1 scan and the nucleotide fragment ion in the MS2 scan. The top right graph, shows the three consecutive scans on which the hit is identified thereby displaying the peak intensity of the precursor ion of the pseudo MS1-scan (dark blue) and the fragment ion of the MS2 scan (turquoise). The bottom left graph shows the MS2 spectrum of the identified hit. The bottom right graph shows the pseudo-MS1 spectrum of the identified hit. The graphs can be explored by using the cursor to zoom in and out. The following parameters are displayed in the table: HIT_ID: A unique identifier assigned to each potential nucleotide-sugar hit. Total score: The overall score assigned to the identified hit. Scan: The scan number associated with the identified hit. Precursor m/z: The m/z value of the theoretical nucleotide-sugar, derived from the chemical composition database. NT-Sugar: The nucleotide combined with the monoisotopic mass of the corresponding monosaccharide. Nucleotide Fragment: The nucleotide fragment detected for the given hit. Sugar Mass: The monoisotopic mass of the identified potential monosaccharide. Sum

Formula: The chemical sum formula of the potential monosaccharide. Mass Error (ppm): The difference (in parts per million) between the mass of the identified hit and the theoretical mass of the corresponding nucleotide-sugar from the chemical composition database. Relative abundance: The relative intensity (%) of the hit, normalized to the most intense hit detected within the same sample. MS1 Intensity: The MS1 signal intensity corresponding to the identified hit. RT (min): The chromatographic retention time of the hit, expressed in minutes. Isolation centre: The m/z center of the isolation window used during acquisition. Consecutive Score: Assigned +10 if the hit is detected in at least X consecutive scans; otherwise, the score is 0. Pearson Score: Assigned +10 if the Pearson correlation of the precursor and corresponding fragment intensities across three consecutive scans exceeds 'Pearson High Threshold'. A score of -10 is assigned if < 'Pearson Low Threshold' and a score of -50 when <0.50. Frag Score: Assigned +10 if the intensity ratio of the MS2 fragment relative to the "MS1" precursor is <100, a score of -10 is assigned when <'MS2 Peak Fraction Score' and a score of -50 when >110. Parent Score: Assigned +10 if the MS1 Intensity of the identified hit exceeds 'MS1 Intensity Score'; otherwise, the score is 0. CO2 Score: Assigned +10 if a CO2 loss peak is present in the corresponding "MS2" spectrum of the identified hit; otherwise, the score is 0. MS1/MS2 Correlation: the Pearson correlation of the precursors and fragments across three consecutive scans. % Fragmentation: The MS2 fragment intensity/MS1 precursor intensity (or MS2 precursor intensity for CMP-sugars)*100%

3.6.3. Theoretical nucleotide sugar chemical composition space.

Table S3.1. Filter criteria for the theoretical nucleotide-sugar chemical composition space.

	Remove compositions with:	
Mass	A mass	< 100
	A mass	> 400
Mass defect	A mass defect	< 0.01
	A mass defect	> 0.19
	A mass between 100-200 and a mass defect	> 0.09
	A mass between 100-250 and a mass defect	> 0.13
Double Bond Equivalent (DBE)	A DBE	< 1.0
	A DBE	> 6.0
	A mass between 100-150 and a DBE	> 3.0
C/H Ratio	A C/H ratio	< 0.40
	A C/H ratio	> 1.25
C/(O+N) Ratio	A C/(O+N) ratio	< 0.90
	A C/(O+N) ratio	> 2.10
Nr. of N- and S atoms	A mass between 100-200 and nr. of N-atoms	> 1
	A mass between 100-275 and nr. of N-atoms	> 3
	Nr. of S- and N-atoms	> 0

3.6.4. SugarBase scoring and filtering criteria.

Table S3.2. Scoring criteria for the identified NT-sugar hits.

Scoring criteria	Threshold	Score
Number of consecutive scans in which the hit is present	≥ 4	+10
Pearson correlation (based on three scans)	> 0.80	+10
	< 0.70	-10
	< 0.50	-50
For non-CMP-sugar hits:	< 100	+10
Percentage of the precursor intensity in the “MS1” scan compared to the fragment intensity in the “MS2” scan.	< 1	-10
	> 110	-50
For CMP-sugar hits:	≤ 25	+10
Percentage of the precursor intensity in the “MS2” scan compared to its intensity in the “MS1” scan	> 25	-50
Precursor intensity in the “MS1” scan	> 100000	+10
Presence of a CO ₂ loss peak in the “MS2” scan	Pass	+10
For CMP-sugar hits:		
N-rule: contains both an even or both an odd number of N- and H-atoms	Fail	-50

3

3.6.5. Retention time variability of the porous graphitic carbon column.

Retention time accuracy of the porous graphitic carbon separation was monitored by injecting a ^{13}C -labelled *Penicillium chrysogenum* reference metabolite extract prior to every sample batch (each batch containing a maximum of six samples).

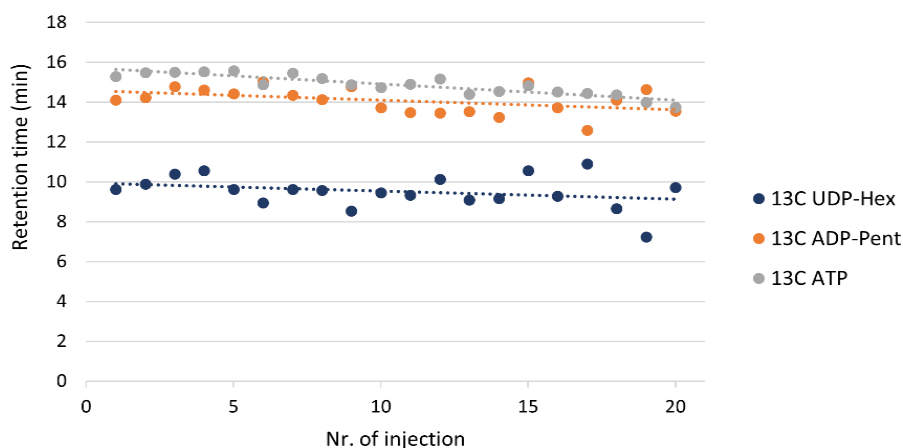


Figure S3.3. Retention time variability of the porous graphite carbon chromatographic separation. The retention time of ^{13}C UDP-hexose, ^{13}C ADP-Pentose and ^{13}C ATP is plotted against the injection number, where each injection number represents a different day.

3.6.6. Nucleotide sugar profile across analyzed taxonomic space of microbes.

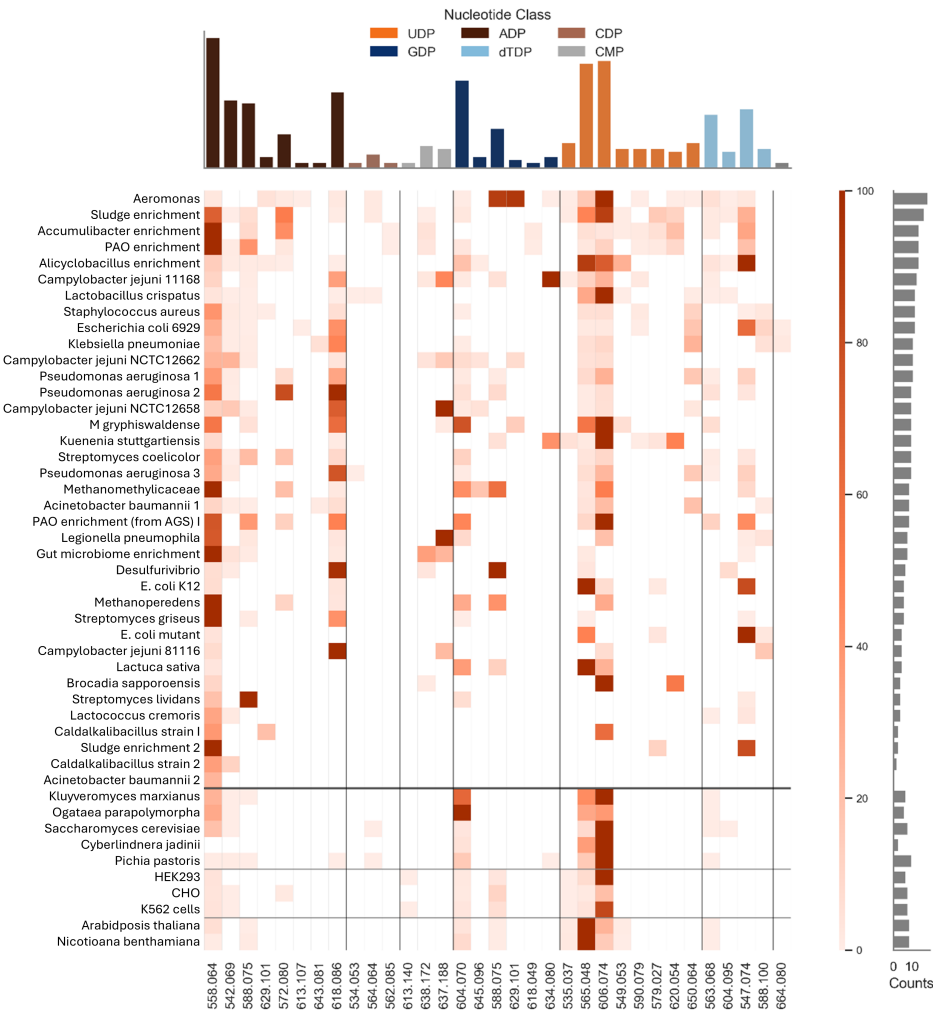


Figure S3.4. Heatmap showing identified nucleotide-sugar candidates across diverse microorganisms using the established “SugarBase” pipeline in targeted mode. The pipeline was applied without an abundance threshold, but with the default filtering and scoring parameters. Nucleotide sugar candidates identified in >1 sample are displayed. The m/z values of the identified nucleotide sugar candidates ($[M-H]^-$) are displayed on the x-axis, and the color scale represents normalized peak intensities. The bar graph above the heatmap indicates the number of analyzed microorganisms in which each nucleotide sugar candidate was identified, while the bar graph on the right shows the number of nucleotide sugar candidates detected per microorganism.



Figure S3.5. The distribution of common sugar modifications across the different nucleotides-sugars identified over a broad phylogenetic spectrum of microbes. A matrix of 174 nucleotide sugars, linking common pentose, hexose, heptose, and nonulosonic acid sugars and their modifications to the seven activating nucleotides UDP, ADP, GDP, CDP, CMP, dTDP, and TDP, was queried using the SugarBase pipeline. The analyzed organisms are illustrated in Figure S3.4.

3.6.7. DMB-labelling reaction of α -keto acid group in ulosonic acids.

3

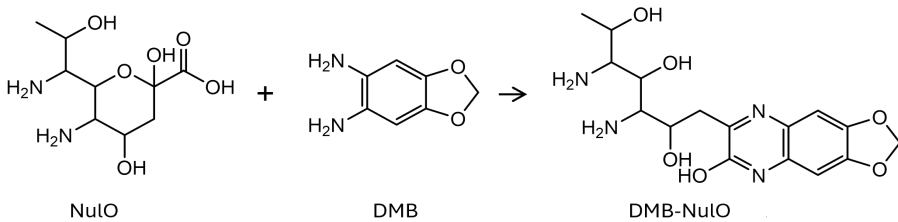


Figure S3.6. Chemical DMB-labelling reaction of α -keto acids.

3.6.8. Fragmentation spectra of hydrolyzed and DMB-labelled nucleotide-sugar extracts.

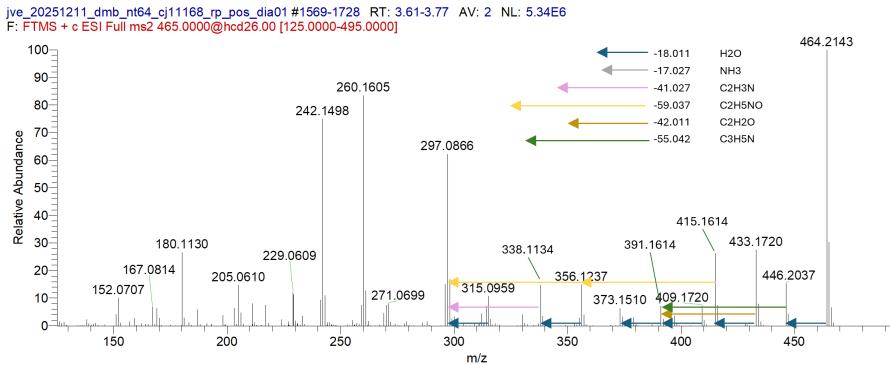


Figure S3.7. Annotated fragmentation spectrum of m/z 464.2143 of a DMB-labelled nucleotide-sugar extract from *Campylobacter jejuni* NCTC 11168.

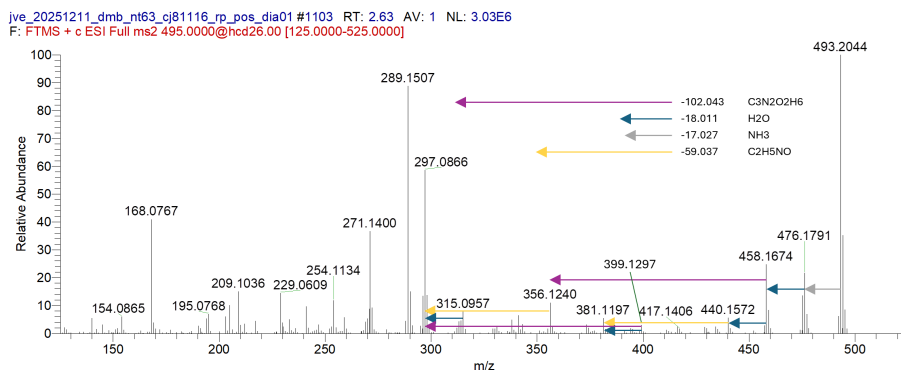


Figure S3.8. Annotated fragmentation spectrum of m/z 493.2044 of a DMB-labelled nucleotide-extract from *Campylobacter jejuni* NCTC 81116.

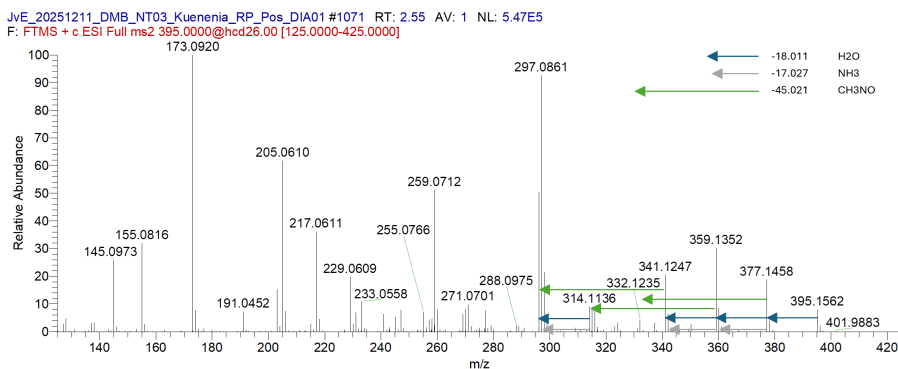


Figure S3.9. Annotated fragmentation spectrum of m/z 395.1562 of a DMB-labelled nucleotide-extract from *Candidatus Kuenenia stuttgartiensis*.

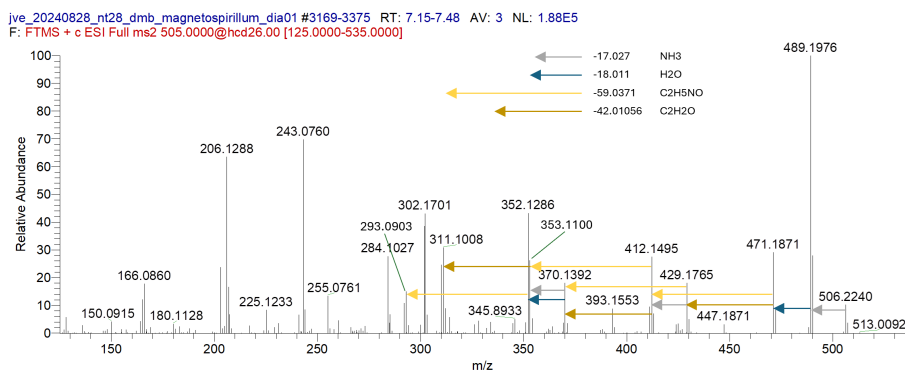


Figure S3.10. Annotated fragmentation spectrum of m/z 506.2240 of a DMB-labelled nucleotide-extract from *Magnetospirillum gryphiswaldense*.

3.6.9. MSn fragmentation tree of DMB-labelled higher carbon (C10) and C9 ulosonic acids

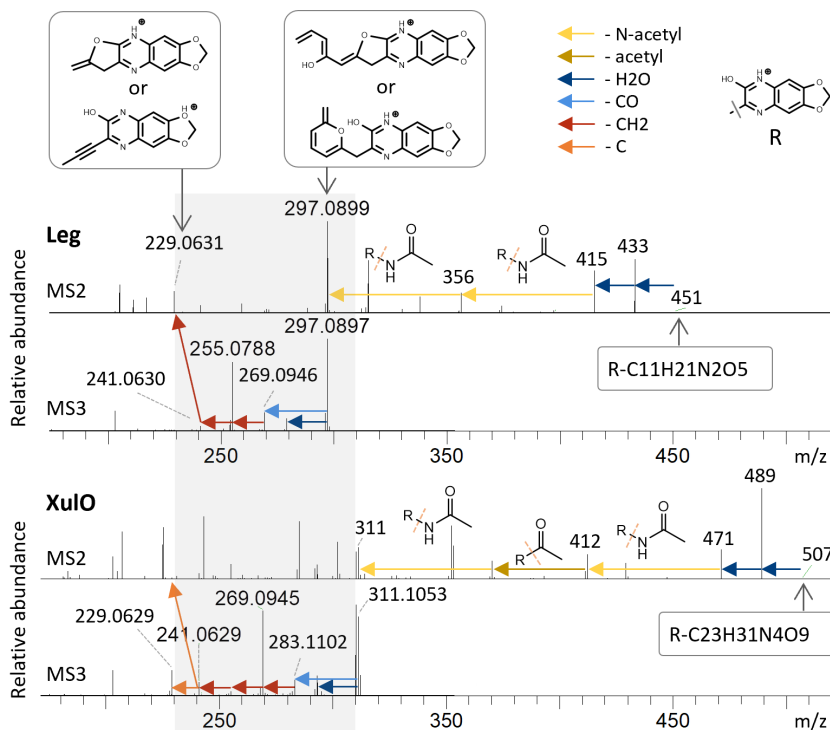


Figure S3.11. Fragmentation tree of the CMP-activated (C10) ulosonic acid in *M. gryphiswaldense*. The graphs show the fragmentation trees of DMB-Leg5Ac7Ac (Leg) and the DMB-labeled higher-carbon ulosonic acid from *M. gryphiswaldense*, termed “XuLO” (deculosonic acid). For Leg, fragmentation of the precursor ion (m/z 451) produced the characteristic C9 marker fragment at m/z 297.089, which was further isolated and fragmented to yield a fragmentation tree terminating in the conserved core fragment at m/z 229.06. Similarly, fragmentation of the XuLO precursor ion (m/z 507) generated a C10 marker fragment at m/z 311.105, which upon isolation and fragmentation showed the same fragmentation tree, except fragmentation of one additional CH₂ unit before converging on the same m/z 229.06 core fragment. Fragment losses are annotated by coloured arrows.

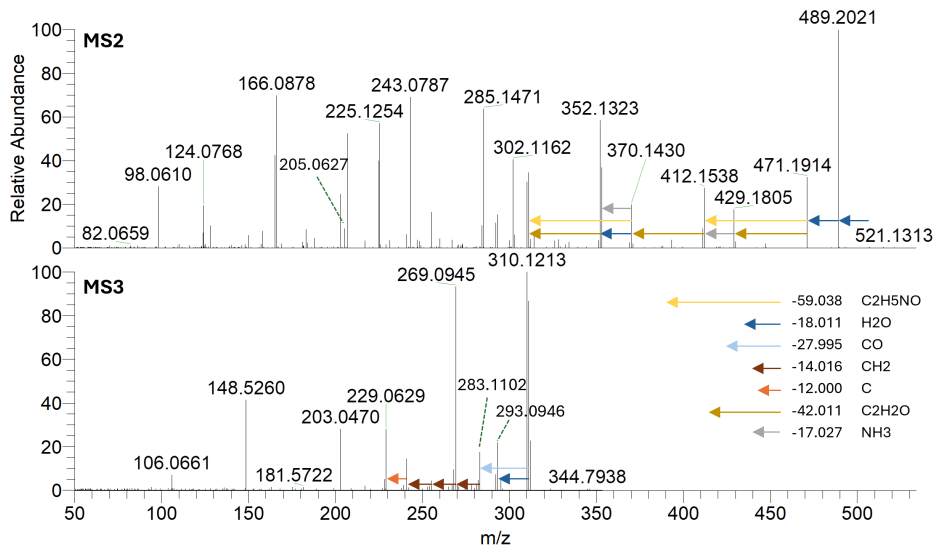


Figure S3.12. Annotated fragmentation spectrum of DMB labelled *Magnetospirillum gryphiswaldense*. Top: full MS2 spectrum of m/z 507.2292 at an HCD of 25. Bottom: full MS3 spectrum of m/z 311.1053 at an HCD of 32.

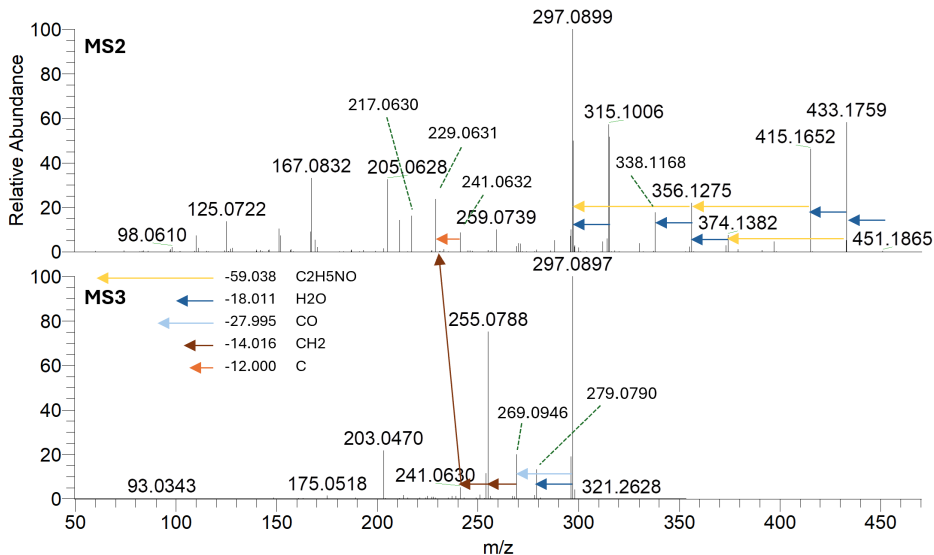


Figure S3.13. Annotated fragmentation spectrum of DMB labelled Leg5Ac7Ac. Top: full MS2 spectrum of m/z 451.1650 at an HCD of 25. Bottom: full MS3 spectrum of m/z 297.0895 at an HCD of 32.

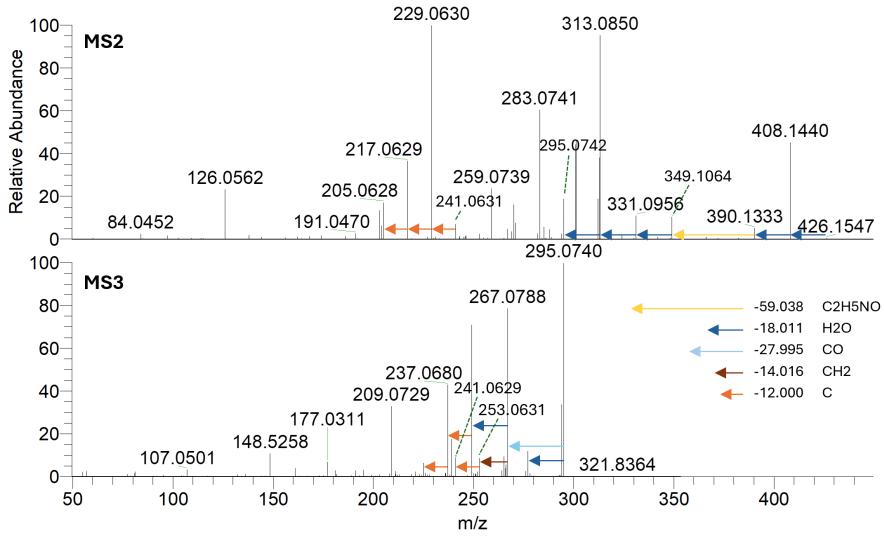


Figure S3.14. Annotated fragmentation spectrum of DMB labelled Neu5Ac. Top: full MS2 spectrum of m/z 426.1398 at an HCD of 25. Bottom: full MS3 spectrum of m/z 295.0739 at an HCD of 32.

3.6.10. Fragmentation of DMB-labelled *Ca. Accum phosphatis* biomass.

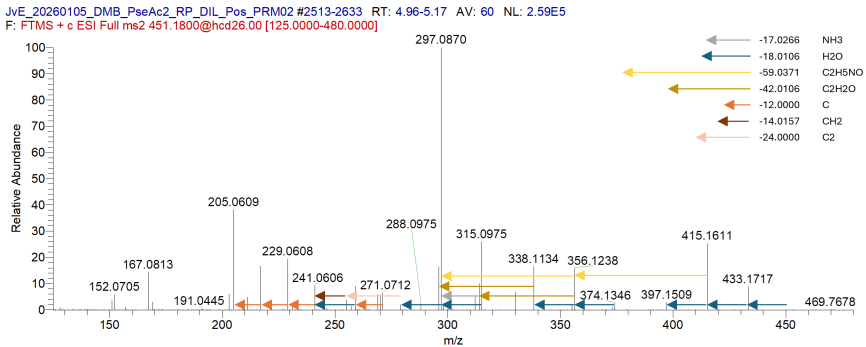


Figure S3.15. Annotated fragmentation spectrum of m/z 451.180 of DMB-Pse5Ac7Ac.

3.6.11. DMB-Pse5Ac7Ac and DMB-Leg5Ac7Ac fragmentation profiling

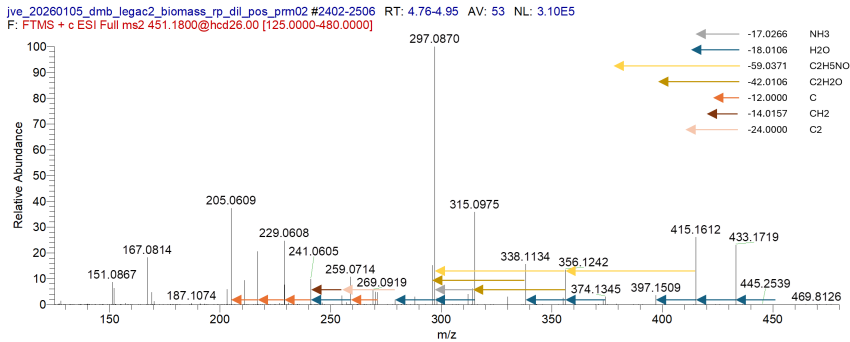


Figure S3.16. Annotated fragmentation spectrum of m/z 451.180 of DMB-Leg5Ac7Ac.

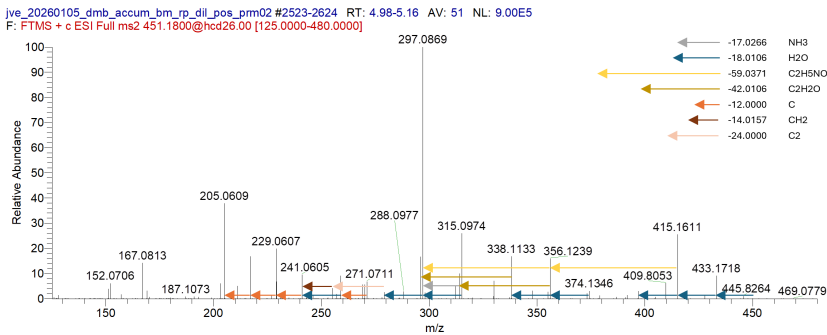


Figure S3.17. Annotated fragmentation spectrum of m/z 451.180 from DMB-labelled *Ca. Accumulibacter phosphatis* biomass.

Table S3.3. Integrated area of DMB-Pse5Ac7Ac and DMB-Leg5Ac7Ac fragments with a 10 ppm error marge.

	PSE1	PSE2	LEG1	LEG2	Accum1	Accum2
205.1	103600	106660	120670	120119	351222	342889
217.1	45539	46630	64196	66200	157291	150912
229.1	53729	53653	80018	79774	185671	179444
297.1	275079	279266	322017	320514	930397	904737
315.1	71617	72628	111823	114555	240041	234355
338.1	44584	46575	49206	49927	157403	152379
415.2	69643	71003	83184	83620	235644	231251
433.2	24595	24548	73030	74634	85645	82553

Figure S3.18. Legionaminic acid and Pseudaminic acid biosynthesis pathways, including putative transferases^{4,22,60-63}.



References

- 1 Mikkola, S. Nucleotide Sugars in Chemistry and Biology. *Molecules* **25** (2020). <https://doi.org/ARTN575510.3390/molecules25235755>
- 2 Logan, S. M. *et al.* Identification of novel carbohydrate modifications on 11168 flagellin using metabolomics-based approaches. *Febs J* **276**, 1014–1023 (2009). <https://doi.org/10.1111/j.1742-4658.2008.06840.x>
- 3 Thibault, P. *et al.* Identification of the carbohydrate moieties and glycosylation motifs in flagellin. *J Biol Chem* **276**, 34862–34870 (2001). <https://doi.org/DOI10.1074/jbc.M104529200>
- 4 Zebian, N. *et al.* Comprehensive analysis of flagellin glycosylation in NCTC 11168 reveals incorporation of legionaminic acid and its importance for host colonization. *Glycobiology* **26**, 386–397 (2016). <https://doi.org/10.1093/glycob/cwv104>
- 5 Guerry, P. *Campylobacter* flagella: not just for motility. *Trends in Microbiology* **15**, 456–461 (2007). <https://doi.org/https://doi.org/10.1016/j.tim.2007.09.006>
- 6 Kay, E., Cuccui, J. & Wren, B. W. Recent advances in the production of recombinant glycoconjugate vaccines. *Npj Vaccines* **4** (2019). <https://doi.org/10.1038/s41541-019-0110-z>
- 7 Poly, F., Noll, A. J., Riddle, M. S. & Porter, C. K. Update on *Campylobacter* vaccine development. *Hum Vacc Immunother* **15**, 1389–1400 (2019). <https://doi.org/10.1080/21645515.2018.1528410>
- 8 Wei, R. H. *et al.* Synthetic Pseudaminic-Acid-Based Antibacterial Vaccine Confers Effective Protection against Infection. *Acs Central Sci* **7**, 1535–1542 (2021). <https://doi.org/10.1021/acscentsci.1c00656>
- 9 Micoli, F., Costantino, P. & Adamo, R. Potential targets for next generation antimicrobial glycoconjugate vaccines. *Fems Microbiol Rev* **42**, 388–423 (2018). <https://doi.org/ARTNfuy01110.1093/femsre/fuy011>
- 10 Caputto, R., Leloir, L. F., Cardini, C. E. & Paladini, A. C. Isolation of the Coenzyme of the Galactose Phosphate-Glucose Phosphate Transformation. *J Biol Chem* **184**, 333–350 (1950).
- 11 Berbis, M. A., Sánchez-Puelles, J. M., Cañada, F. J. & Jiménez-Barbero, J. Structure and Function of Prokaryotic UDP-Glucose Pyrophosphorylase, A Drug Target Candidate. *Curr Med Chem* **22**, 1687–1697 (2015). <https://doi.org/DOI10.2174/0929867322666150114151248>
- 12 del Val, I. J., Kyriakopoulos, S., Polizzi, K. M. & Kontoravdi, C. An optimized method for extraction and quantification of nucleotides and nucleotide sugars from mammalian cells. *Anal Biochem* **443**, 172–180 (2013). <https://doi.org/10.1016/j.ab.2013.09.005>
- 13 Yin, Y. B., Huang, J. L., Gu, X. G., Bar-Peled, M. & Xu, Y. Evolution of Plant Nucleotide-Sugar Interconversion Enzymes. *Plos One* **6** (2011). <https://doi.org/ARTNe2799510.1371/journal.pone.0027995>
- 14 Samuel, G. & Reeves, P. Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydr Res* **338**, 2503–2519 (2003). <https://doi.org/10.1016/j.carres.2003.07.009>
- 15 Bar-Peled, M. & O'Neill, M. A. Plant Nucleotide Sugar Formation, Interconversion, and Salvage by Sugar Recycling. *Annu Rev Plant Biol* **62**, 127–155 (2011). <https://doi.org/10.1146/annurev-arplant-042110-103918>
- 16 Kajihara, Y. *et al.* Unique Self-Anhydride Formation in the Degradation of Cytidine-5'-monophosphosialic Acid (CMP-Neu5Ac) and Cytidine-5'-diphosphosialic Acid (CDP-Neu5Ac) and its Application in CMP-sialic Acid Analogue Synthesis. *Chem-Eur J* **17**, 7645–7655 (2011). <https://doi.org/10.1002/chem.201003387>
- 17 Bülter, T. & Elling, L. Enzymatic synthesis of nucleotide sugars. *Glycoconjugate J* **16**, 147–159 (1999). <https://doi.org/DOI10.1023/A:1026444726698>
- 18 Frohnemeyer, H. & Elling, L. Enzyme cascades for the synthesis of nucleotide sugars: Updates to recent production strategies. *Carbohydr Res* **523** (2023). <https://doi.org/ARTN10872710.1016/j.carres.2022.108727>
- 19 Flack, E. K. P. *et al.* Biocatalytic Transfer of Pseudaminic Acid (Pse5Ac7Ac) Using Promiscuous Sialyltransferases in a Chemoenzymatic Approach to Pse5Ac7Ac-Containing Glycosides. *Acs Catal* **10**, 9986–9993 (2020). <https://doi.org/10.1021/acscatal.0c02189>
- 20 Nidetzky, B., Gutmann, A. & Zhong, C. Leloir Glycosyltransferases as Biocatalysts for Chemical Production. *Acs Catal* **8**, 6283–6300 (2018). <https://doi.org/10.1021/acscatal.8b00710>
- 21 Schoenhofen, I. C., McNally, D. J., Brisson, J. R. & Logan, S. M. Elucidation of the CMP-pseudaminic acid pathway in *Helicobacter pylori*: synthesis from UDP-acetylglucosamine by a single enzymatic reaction. *Glycobiology* **16**, 8c–14c (2006). <https://doi.org/10.1093/glycob/cwl010>
- 22 Schoenhofen, I. C., Vinogradov, E., Whitfield, D. M., Brisson, J. R. & Logan, S. M. The CMP-legionaminic acid pathway in *Campylobacter*: Biosynthesis involving novel GDP-linked precursors. *Glycobiology* **19**, 715–725 (2009). <https://doi.org/10.1093/glycob/cwp039>
- 23 Kochanowski, N. *et al.* Intracellular nucleotide and nucleotide sugar contents of cultured CHO cells determined by a fast, sensitive, and high-resolution ion-pair RP-HPLC. *Anal Biochem* **348**, 243–251 (2006). <https://doi.org/10.1016/j.ab.2005.10.027>
- 24 Pabst, M. *et al.* Nucleotide and Nucleotide Sugar Analysis by Liquid Chromatography-Electrospray Ionization-Mass Spectrometry on Surface-Conditioned Porous Graphitic Carbon. *Anal Chem* **82**, 9782–9788 (2010). <https://doi.org/10.1021/ac101975k>
- 25 Răbină, J. *et al.* Analysis of nucleotide sugars from cell lysates by ion-pair solid-phase extraction and reversed-phase high-performance liquid chromatography. *Glycoconjugate J* **18**, 799–805 (2001). <https://doi.org/DOI10.1023/A:1021107602535>
- 26 Ramm, M., Wolfender, J. L., Queiroz, E. F., Hostettmann, K. & Hamburger, M. Rapid analysis of nucleotide-activated sugars by high-performance liquid chromatography coupled with diode-array detection, electrospray ionization mass spectrometry and nuclear magnetic resonance. *J Chromatogr A* **1034**, 139–148 (2004). <https://doi.org/10.1016/j.chroma.2004.02.023>
- 27 Ito, J. *et al.* Analysis of plant nucleotide sugars by hydrophilic interaction liquid chromatography and tandem mass spectrometry. *Anal Biochem* **448**, 14–22 (2014). <https://doi.org/10.1016/j.ab.2013.11.026>
- 28 Rautengarten, C., Heazlewood, J. L., & Ebert, B. . Profiling Cell Wall Monosaccharides and Nucleotide-Sugars from Plants. *Current Protocols in Plant Biology* **4** (2019). <https://doi.org/https://doi.org/10.1002/CPPB.20092>
- 29 Soo, E. C. *et al.* Selective detection and identification of sugar nucleotides by CE-electrospray-MS and its

- application to bacterial metabolomics. *Anal Chem* **76**, 619–626 (2004). <https://doi.org/10.1021/ac034875i>
- 30 Páez-Watson, T. *et al.* Sweet Secrets: Exploring Novel Glycans and Glycoconjugates in the Extracellular Polymeric Substances of “Candidatus Accumulibacter”. *ACS Es&t Water* **4**, 3391–3399 (2024).
 - 31 Pabst, M. & Altmann, F. Influence of electrosorption, solvent, temperature, and ion polarity on the performance of LC-ESI-MS using graphitic carbon for acidic oligosaccharides. *Anal Chem* **80**, 7534–7542 (2008).
 - 32 Chambers, M. C. *et al.* A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* **30**, 918–920 (2012). <https://doi.org/10.1038/nbt.2377>
 - 33 Pabst, M. *et al.* A general approach to explore prokaryotic protein glycosylation reveals the unique surface layer modulation of an anammox bacterium. *Isme J* **16**, 346–357 (2022). <https://doi.org/10.1038/s41396-021-01073-y>
 - 34 Kleikamp, H. B. *et al.* Database-independent de novo metaproteomics of complex microbial communities. *Cell Systems* **12**, 375–383. e375 (2021).
 - 35 Kleikamp, H. B. *et al.* NovoLign: metaproteomics by sequence alignment. *ISME communications* **4**, ycae121 (2024).
 - 36 Kleikamp, H. B. C. *et al.* Tackling the chemical diversity of microbial nonulosonic acids - a universal large-scale survey approach. *Chem Sci* **11**, 3074–3080 (2020). <https://doi.org/10.1039/c9sc06406k>
 - 37 Patiny, L. & Borel, A. ChemCalc: A Building Block for Tomorrow's Chemical Infrastructure. *J Chem Inf Model* **53**, 1223–1228 (2013). <https://doi.org/10.1021/ci300563h>
 - 38 Janoska, A. *et al.* Influence of oxygen concentration on the metabolism of *Penicillium chrysogenum*. *Engineering in Life Sciences* **23**, e2100139 (2023).
 - 39 Schäffer, C. & Messner, P. Emerging facets of prokaryotic glycosylation. *Fems Microbiol Rev* **41**, 49–91 (2017). <https://doi.org/10.1093/femsre/fuw036>
 - 40 Palcic, M. M. Glycosyltransferases as biocatalysts. *Curr Opin Chem Biol* **15**, 226–233 (2011). <https://doi.org/10.1016/j.cbpa.2010.11.022>
 - 41 van Scherpenzeel, M. *et al.* Dynamic tracing of sugar metabolism reveals the mechanisms of action of synthetic sugar analogs. *Glycobiology* **32**, 239–250 (2022). <https://doi.org/10.1093/glycob/cwab106>
 - 42 Nothaft, H. & Szymanski, C. M. Protein glycosylation in bacteria: sweeter than ever. *Nat Rev Microbiol* **8**, 765–778 (2010). <https://doi.org/10.1038/nrmicro2383>
 - 43 Hameed, A., Woodacre, A., Machado, L. R. & Marsden, G. L. An Updated Classification System and Review of the Lipooligosaccharide Biosynthesis Gene Locus in. *Front Microbiol* **11** (2020).
 - 44 Kneidinger, B. *et al.* Biosynthesis Pathway of ADP-l-glycero- β -d-manno-Heptose in *Escherichia coli*. *J Bacteriol* **184**, 363–369 (2002). <https://doi.org/10.1128/jb.184.2.363-369.2002>
 - 45 Losse, A. *et al.* Functional Analysis of the Magnetosome Island in: The Operon Is Sufficient for Magnetite Biomineralization. *Plos One* **6** (2011). <https://doi.org/ARTN e25561> 10.1371/journal.pone.0025561
 - 46 Eerden, S. A., Abeel, T., van Loosdrecht, M. C. M. & Roy, S. Phylogenetic analysis reveals diversity in glycan biosynthesis in “Candidatus Accumulibacter”. *Biofilm* **11** (2026). <https://doi.org/https://doi.org/10.1016/j.biofilm.2026.100350>
 - 47 van Heijenoort, J. Formation of the glycan chains in the synthesis of bacterial peptidoglycan. *Glycobiology* **11**, 25r–36r (2001).
 - 48 McDonald, N. D. & Boyd, E. F. Structural and Biosynthetic Diversity of Nonulosonic Acids (NulOs) That Decorate Surface Structures in Bacteria. *Trends Microbiol* **29**, 142–157 (2021). <https://doi.org/10.1016/j.tim.2020.08.002>
 - 49 Sacher, J. C. *et al.* Binding of Phage-Encoded FlaGrab to Motile *Campylobacter jejuni* Flagella Inhibits Growth, Downregulates Energy Metabolism, and Requires Specific Flagellar Glycans. *Front Microbiol* **11**, 397 (2020). <https://doi.org/10.3389/fmicb.2020.00397>
 - 50 Schirm, M. *et al.* Structural, genetic and functional characterization of the flagellin glycosylation process in. *Mol Microbiol* **48**, 1579–1592 (2003). <https://doi.org/10.1046/j.1365-2958.2003.03527.x>
 - 51 Parkhill, J. *et al.* The genome sequence of the food-borne pathogen reveals hypervariable sequences *Nature* **403**, 665–668 (2000). <https://doi.org/Doi 10.1038/35001088>
 - 52 Pearson, B. M. *et al.* The complete genome sequence of strain 81116 (NCTC11828). *J Bacteriol* **189**, 8402–8403 (2007). <https://doi.org/10.1128/Jb.01404-07>
 - 53 Palmer, S. R. *et al.* Water-Borne Outbreak of *Campylobacter Gastroenteritis*. *Lancet* **1**, 287–290 (1983). [https://doi.org/Doi 10.1016/S0140-6736\(83\)91698-7](https://doi.org/Doi 10.1016/S0140-6736(83)91698-7)
 - 54 Meng, X., Boons, G.-J., Węsten, M. M. S. M., & Wennekes, T. . Glycosylation Probes Metabolic Labeling of Legionaminic Acid in Flagellin Glycosylation of *Campylobacter jejuni* Identifies Maf4 as a Putative Legionaminyl Transferase. *Angew. Chem* **60**, 24811–24816 (2021). <https://doi.org/https://doi.org/10.1002/anie.202107181>
 - 55 Ostenfeld, L. J. *et al.* A hybrid receptor binding protein enables phage F341 infection of *Campylobacter* by binding to flagella and lipooligosaccharides. *Frontiers in Microbiology* **15**, 1358909 (2024).
 - 56 Holst Sørensen, M. C. *et al.* Primary isolation strain determines both phage type and receptors recognised by *Campylobacter jejuni* bacteriophages. *Viruses of microbes: Structure and function, from molecules to communities. Book of Abstracts* (2014).
 - 57 Javed, M. A. *et al.* A receptor-binding protein of *Campylobacter jejuni* bacteriophage NCTC 12673 recognizes flagellin glycosylated with acetamidino-modified pseudaminic acid. *Mol Microbiol* **95**, 101–115 (2015). <https://doi.org/10.1111/mmi.12849>
 - 58 Wang, J. *et al.* Chemical Remodeling of Cell-Surface Sialic Acids through a Palladium-Triggered Bioorthogonal Elimination Reaction. *Angew Chem Int Edit* **54**, 5364–5368 (2015). <https://doi.org/10.1002/anie.201409145>
 - 59 Tang, A. H. *et al.* Uncovering bacterial pseudaminylation with pan-specific antibody tools. *Nature Chemical Biology*, 1–12 (2026).
 - 60 McNally, D. J. *et al.* Targeted metabolomics analysis of VC167 reveals legionaminic acid derivatives as novel flagellar glycans. *J Biol Chem* **282**, 14463–14475 (2007). <https://doi.org/10.1074/jbc.M611027200>
 - 61 Glaze, P. A., Watson, D. C., Young, N. M. & Tanner, M. E. Biosynthesis of CMP-N,N'-Diacytyllegionaminic acid from UDP-N,N'-Diacytylbacillosamine in. *Biochemistry-U S* **47**, 3272–3282 (2008). <https://doi.org/10.1021/bi702364s>
 - 62 Ménard, R. *et al.* Small-Molecule Inhibitors of the Pseudaminic Acid Biosynthetic Pathway: Targeting Motility as a Key Bacterial Virulence Factor. *Antimicrob Agents Ch* **58**, 7430–7440 (2014). <https://doi.org/10.1128/Aac.03858-14>

- 63 Meng, X. K., Boons, G. J., Wösten, M. M. S. M. & Wennekes, T. Metabolic Labeling of Legionaminic Acid in Flagellin Glycosylation of *Campylobacter jejuni* Identifies

Maf4 as a Putative Legionaminyl Transferase. *Angew Chem Int Edit* **60**, 24811–24816 (2021).
<https://doi.org/10.1002/anie.202107181>

Measuring Temporal Variations of Nucleotide Pools in Microbial Granular Biofilm

Jitske M. van Ede, Timothy Páez-Watson, Mark C.M. van
Loosdrecht and Martin Pabst



4.1 | Introduction

Microorganisms in natural and engineered environments continuously face fluctuations in physicochemical conditions, including changes in substrate availability¹, pH², temperature³, and redox potential⁴. To survive and adapt, cells employ regulatory responses that span multiple layers. The timescale of the environmental change relative to the cell's response capabilities determines which layer of regulation dominates^{5,6}. Transcriptomic responses and proteome remodelling typically require minutes to hours, especially considering that proteome remodelling is dependent on protein dilution by growth^{7,8}. However, environmental perturbations can occur rapidly, and are often faster compared to proteome level changes^{5,6}. Metabolic fluxes can also be regulated through allosteric regulation, post translational modifications and changes in substrate/product concentrations. Among these metabolites, nucleotides such as ATP, ADP, AMP, and their guanylate (GxP), uridylate (UxP), and cytidylate (CxP) counterparts are central. Their concentrations not only reflect the energy and redox status of the cell but also influence flux distributions and enzyme activity through thermodynamic and kinetic effects. Accordingly, metrics such as the adenylate energy charge (AEC)⁹ have long served as proxies for cellular energetic health, often maintained within a narrow physiological range of 0.7–0.95 across different species and conditions^{10–14}.

An example where fast regulation clearly dominates is that of Enhanced Biological Phosphorous Removal (EBPR) processes⁴. The microbial community in this system experience rapid anaerobic–aerobic transitions during wastewater treatment cycles. Under anaerobic conditions, members of the community rapidly take up acetate and store it as polyhydroxyalkanoates (PHAs), fuelled by glycogen and polyphosphate degradation. Upon the introduction of oxygen, this metabolic strategy is abruptly reversed, with PHA oxidation supporting polyphosphate and glycogen replenishment¹⁵. These dynamic transitions occur within minutes, orders of magnitude faster than the imposed growth rates of the biomass (often referred to as sludge retention time) which span days. Consequently, proteome remodelling is too slow to explain the observed flux changes.

Recent proteomic studies revealed protein abundance changes of only < 3% during the EBPR cycle¹⁶, remarkably small considering the dramatic flux inversions observed. These observations have led to the hypothesis that core metabolic regulation underpinning EBPR transformations is driven by shifts in key metabolite pools within the cells of the community, particularly those associated with energy and redox metabolism^{17,18}. Detailed metabolomic studies of EBPR communities have not been conducted to date, mainly due to two challenges: lack of pure

cultures and the growth of the EBPR community as dense granular biofilms embedded in extracellular polymeric substances hampering metabolite extraction. Various extraction methods have been explored for planktonic cells, including high-temperature extractions, freeze/thaw cycles, chloroform-methanol lysis and acid or alkaline treatments¹⁹⁻²⁴. Biofilms often require an additional mechanical disruption such as bead beating and sonication to enhance metabolite recovery^{25,26}. The choice of solvent also depends on the chemical characteristics of the target metabolites, including polarity and solubility²⁴. As a result, there is no universal extraction method; protocols must be tailored to both the microbial system and metabolite class of interest. Moreover, metabolites like nucleotides turn over rapidly and require instant quenching to preserve their *in vivo* state^{21,27}. Quenching strategies typically include a solvent solution combined with extreme temperatures or pH²¹, such as cold methanol^{22,28}, liquid nitrogen^{24,29} or acidic organic solvents¹⁹.

Two key factors are commonly evaluated when testing the extraction efficiency: efficacy and stability²¹. The efficacy is assessed by applying different extraction methods to the same biological sample and comparing their ability to extract different metabolites. On the other hand, stability can be evaluated by spiking metabolite standards and measuring recovery rates²¹. While metabolite degradation should generally be prevented, minor changes can be corrected for by using internal standards²². Isotope dilution mass spectrometry uses a uniformly ¹³C-labeled cell extract as an internal standard, which allows to correct for degradation during sample preparation, injection variability, ion suppression effects in electrospray ionisation, and instrumental fluctuations^{24,30}.

Metabolomics is predominantly analysed via liquid chromatography coupled to mass spectrometry via electrospray ionisation (ESI), due to its high sensitivity, selectivity and dynamic range compared to enzymatic assays or UV and fluorescence detection³¹. Moreover, it allows for untargeted analysis of metabolites and requires only minute amounts of sample. Earlier techniques to separate nucleotides include ion pair reverse phase (IPRP) and ion exchange chromatography (IEC)³². While the latter is only poorly compatible with ESI-MS due to high salt concentrations, IPRP is not ideal due to persistent contamination by IP reagents³³⁻³⁵. On the other hand, hydrophilic interaction liquid chromatography (HILIC) is compatible with ESI-MS and effective for separating nucleotides. However, it requires a high proportion of organic solvent, which can hinder the solubility of some metabolite extracts. HILIC also often suffers from peak broadening and retention time drift, affecting method robustness and reproducibility, though certain columns have been shown to diminish these

issues^{4,24,25}. A promising alternative is porous graphitic carbon (PGC) chromatography, which shows great separation performance for polar metabolites and is compatible with ESI-MS³⁵⁻³⁸. Nevertheless, PGC can suffer from retention time instability and column polarization, which, however, can be largely circumvented by using internal standards, thorough sample cleanup, choice of running buffers and additives, and electrical grounding^{36,39,40}.

Here, we demonstrate a robust method for quantifying nucleotide pools in a granular biofilm microbial community performing EBPR. By combining rapid quenching with liquid nitrogen, boiling water extraction, PGC chromatographic separation followed by high resolution mass spectrometric detection, we were able to capture energy nucleotide dynamics during a typical EBPR anaerobic-aerobic cycle. Our findings provide evidence for metabolite-level regulation and highlight the potential of energy charge modulation in driving the EBPR transformations.

4

4.2 | Materials and Methods

EBPR granular sludge enrichment. A lab scale EBPR system was run in a 1L (0,8 L working volume) sequencing batch reactor (SBR), following conditions described in Páez-Watson, et al.⁴¹ with some adaptations. The reactor was inoculated using enriched sludge from the work of Páez-Watson, et al.⁴¹, which was previously inoculated with activated sludge from a municipal wastewater treatment plant (Harnaschpolder, The Netherlands). Each SBR cycle lasted 6 hours, consisting of 10 minutes of settling, 70 minutes of effluent removal, 10 minutes of N₂ sparging, 5 minutes of anaerobic feeding, 130 minutes of anaerobic phase and 135 minutes of aerobic phase. N₂ gas and compressed air were sparged at 500 ml/min into the reactor broth to maintain anaerobic and aerobic conditions respectively. The hydraulic retention time (HRT) was 12 hours (removal of 400 ml of broth per cycle, each cycle of 6 hours). The average solids retention time (SRT) was controlled to 8 days by the removal of 25 ml of mixed broth at the end of the mixed aerobic phase in each cycle (total 100 ml per day). The pH was controlled at 7.3 ± 0.1 by dosing 0.5 M HCl or 0.5 M NaOH. The temperature was maintained at 20 ± 1 °C.

The reactor was fed with three separate media components diluted in demineralized water: a concentrated COD medium (400 mg COD/L) of acetate (17 g/L sodium acetate $\times 3\text{H}_2\text{O}$); a concentrated mineral medium (2.51 g/L $\text{NaH}_2\text{PO}_4 \times \text{H}_2\text{O}$, 1.53 g/L NH_4Cl , 1.59 g/L $\text{MgSO}_4 \times 7\text{H}_2\text{O}$, 0.4 g/L $\text{CaCl}_2 \times 2\text{H}_2\text{O}$, 0.48 g/L KCl, 0.4 g/L N-allylthiourea (ATU), and 6 mL/L of trace element solution prepared following Smolders, et al.¹⁵); and demineralized water containing 5 mg/L of yeast extract. In each cycle, 40 mL of COD medium, 40 mL of mineral medium and 320

mL of water with yeast extract were added to the reactor during the 5 minutes of feeding. The final feed contained 400 mg COD/L of acetate.

Reactor and biomass analyses. Extracellular concentrations of phosphate and ammonium were measured with a Gallery Discrete Analyzer (Thermo Fisher Scientific, Waltham, MA). Acetate was measured by high performance liquid chromatography (HPLC) with an Aminex HPX-87H column (Bio-Rad, Hercules, CA), coupled to RI and UV detectors (Waters, Milford, MA), using 0.0015 M phosphoric acid as eluent supplied at a flowrate of 1 mL/min.

The biomass concentration (total and volatile suspended solids – TSS and VSS) was measured in accordance with Standard Methods as described in Smolders, et al.¹⁵ with some modifications: 10 mL of mixed broth were obtained at the end of the aerobic phase, centrifuged at 3600*g for 3 minutes and washed twice with demineralized water to remove salts. The sludge was then dried at 100 °C for 24 hours and weighed on a microbalance to determine the dry content – TSS. The ash content was determined by incinerating the dry material in an oven at 550 °C, and the difference used to calculate the VSS.

For glycogen and PHA determination, biomass samples (10 mL mixed broth) were collected throughout the batch test and stored in 15 mL conical tubes containing 0.3 mL of 37 % formaldehyde to stop biological activity. After each batch test, the biomass tubes were pottered to break the granular structure of the biomass, centrifuged at 3700 *g for 5 minutes and washed twice. The pellet was then frozen at -80 °C for at least 3 hours and freeze dried. For glycogen analysis the method described by Smolders, et al.¹⁵ was used: 5 mg of dry biomass was digested in 0.9 M HCl solutions in glass tubes at 100 °C for 5 hours. After this time, tubes were cooled at room temperature, filtered with 0.45 µm Whatman disk filters and neutralized with equal volumes of 0.9 M NaOH. The glucose resulting from digestion was quantified using the D-Glucose Assay Kit (GOPOD Format) from Megazyme (Bray, Ireland).

For PHA extraction the method described by Oehmen, et al.⁴² was used. Approximately 30 mg of freeze-dried biomass was placed in a tube with 1.5 mL of 10% sulfuric acid in methanol and 1.5 mL of chloroform. To each tube, 100 µL of an internal standard solution (0.01 g/mL benzoic acid in methanol) was added. The samples were then hydrolyzed and esterified at 100 °C for 5 hours with periodic manual vortexing. Following this, 3 mL of ultrapure water was added to facilitate phase separation. The organic phase containing the esters was filtered and analyzed using a gas chromatograph (GC 6890 N, Agilent, USA). Quantifications of PHB, PH2MV, and PHV were performed using commercial standards, including 3-

hydroxybutyrate, 2-hydroxyhexanoate, and a synthetic copolymer of (R)-3-hydroxybutyrate-(R)-3-hydroxyvalerate (Sigma-Aldrich, USA).

Whole metagenome sequencing. For metagenomics, DNA from the biomass samples was extracted using the DNeasy PowerSoil Pro-Kit (Qiagen, Germany) following the manufacturer's protocol. Shotgun sequencing was performed by Hologenomix (Delft, The Netherlands). Paired-end sequencing with a read length of 150 bp was conducted using the Illumina NovaSeq X sequencing system. Library preparation was carried out using the Nextera XT DNA Library Preparation Kit. Approximately 10 Gbp of sequencing data were generated per sample. The quality of raw sequenced reads was assessed using FastQC (version 0.11.7) with default parameters⁴³, and results were visualized with MultiQC (version 1.19). Low-quality paired-end reads were trimmed and filtered using Fastp (version 0.23.4) in paired-end mode⁴⁴.

Taxonomic profiling from raw reads was performed by Hologenomix (Delft, The Netherlands) using the standard Kraken2 (version 2) database (using all complete bacterial, archaeal, and viral genomes in NCBI Refseq database, complemented with a curated wastewater database) with default parameters⁴⁵. The taxonomic classification outcomes from Kraken2.0 were converted into abundance tables using the biom file converter tool to explore metagenomics classification datasets. Figures were generated with the R package “phyloseq”⁴⁶. Clean reads were further assembled into contigs using MetaSPAdes (version 3.15.5) with default parameters⁴⁷. From there, protein-coding sequences were predicted from contigs using Prodigal (<https://github.com/hyatt/Prodigal>). The resulting fasta file was further used for metaproteomic analysis. Sequencing data are available through the BioProject database <https://www.ncbi.nlm.nih.gov/bioproject/> under the BioProject no. PRJNA1305385.

Metaproteomics. Samples were taken during the aerobic phase of the anaerobic-aerobic cycle as described in the section “Sampling and quenching for Metabolomics”. The crushed sample was transferred to a LoBind Eppendorf tube, freeze-dried overnight and subsequently processed as described in van Olst et al. (2025)⁴⁸. In short, approximately 3.5 mg of biomass (dry weight) was lysed by three rounds of vortexing at max speed with acid washed glass beads after the addition of B-PER reagent and 50 mM TEAB pH 8. The biomass was heated to 80 °C in the Thermomixer with Thermotop (1000 rpm), and sonicated for 10 min in an ultrasonic bath. After pelleting the biomass, the supernatant was collected in a clean LoBind Eppendorf tube and proteins were precipitated by the addition of TCA (ratio 4:1) at

4 ° for 30 min. The sample was centrifuged at 14000 rcf, 4 °C for 15 min and the protein pellet was washed with ice cold acetone, centrifuged for 3 min, 14000 rcf, and dissolved in 100 µL 6 M urea. Reduction was performed by the addition of 30 µL 10 mM dithiothreitol at 37 °C for 60 min in the Thermomixer with Thermotop (300 rpm), and alkylation was performed by adding 30 µL 20 mM iodoacetamide at room temperature in the dark for 30 min. Subsequently, 100 mM ammonium bicarbonate was added to dilute the Urea concentration to <1 M and the sample was digested with 15 µL 0.1 µg/µL Trypsin for approximately 18 hours at 37 °C, 300 rpm. A solid phase extraction was performed as described by van Olst et al. (2025)⁴⁸. Finally, the sample was speed vac dried and dissolved in 3% acetonitrile, 0.1% formic acid in water to reach a concentration of 0.5 mg/mL, where approximately 500 ng was analysed by shotgun proteomics.

Shotgun proteomics was performed on an EASY 1200 nano-LC connected to a Q Exactive plus Orbitrap Mass Spectrometer (Thermo Scientific, Germany). Peptides were separated on a C18 PepMap column (0.05 x 150 mm column, Thermo Scientific) using a constant flow rate of 350 nL/min, with mobile phase A (0.1% formic acid (FA), 1% acetonitrile (ACN) in Optima LC/MS-grade H₂O) and mobile phase B (0.1% FA, 80% ACN in Optima LC/MS-grade H₂O). The LC gradient included an initial 2 min at 5% B, followed by a linear increase to 25% B over 90 min, a linear increase to 55% B over 60 min, then a return to 5% B within 3 min, and an equilibration for an additional 20 min. Electrospray ionization was carried out in positive mode. MS1 spectra were acquired at a resolution of 70,000, an automatic gain control (AGC) of 3e6 and a maximum injection time of 75 ms. Data-dependent acquisition (DDA) was applied, isolating precursors using a 2.0 m/z isolation window over a scan range of 385-1250 m/z. Fragmentation was performed with a normalized collision energy of 28%. MS2 spectra were acquired at a resolution of 17,500, an AGC of 2e5 and an maximum injection time of 75 ms. Database searching was performed in PEAKS Studio 10.5 (Bioinformatics Solutions Inc.) using a reference sequence database obtained from whole metagenome sequencing, with a parent mass error tolerance set to 20.0 ppm, a fragment mass error tolerance set to 0.02 Da and Trypsin as the proteolytic enzyme. Carbamidomethylation was specified as a fixed modification, while oxidation and deamidation were set as variable modifications, allowing up to three variable modifications per peptide. The identified peptide spectrum match list was furthermore taxonomically annotated using the NovoBridge pipeline⁴⁹, which retrieves a taxonomic lineage from the Unipept database (<https://unipept.ugent.be/>) for every peptide sequence, and which are then grouped into a taxonomic composition for every taxonomic rank. Raw shotgun

metaproteomics data, the reference sequence database and the search files are available in the PRIDE database under project accession number PXD067586.

Sampling and quenching for Metabolomics. Samples were collected from the stirred broth of the bioreactor using a 5 mL syringe attached to the sampling port. For method optimization, samples were taken at the end of the aerobic phase, whereas for dynamic profiling, samples were collected throughout the anaerobic–aerobic cycle. Each sample, containing both broth and biomass granules, was immediately transferred into a pre-chilled mortar containing liquid nitrogen, ensuring submersion within 3 seconds of collection. A precooled pestle was then used to crush the frozen sample into a fine powder in the presence of excess liquid nitrogen. The resulting powder was transferred into 2 mL Eppendorf tubes with manually pre-cut lids to allow them to fit inside 15 mL tubes used during heat-based extraction methods, and 50 μL of uniformly ^{13}C -labeled yeast extract was added as an internal standard. The ^{13}C -labeled yeast extract was prepared in house as published by Wu Liang et al., 2005³⁰. The tubes were stored in liquid nitrogen until further processing. Depending on the quenching and extraction protocol, one of the following procedures was used to inactivate metabolic enzymes and prepare the samples:

Acidic Acetonitrile/Methanol: 500 μL of pre-cooled (at $-20\text{ }^{\circ}\text{C}$) acetonitrile:methanol:water (40:40:20, v/v/v) with 0.1 M formic acid was added to the sample. After vortexing, samples were placed on ice for 3 minutes. Subsequently, 43.5 μL of pre-cooled 15% (w/v) ammonium bicarbonate in MS-grade water was added to neutralize the extract. Samples were then stored at $-80\text{ }^{\circ}\text{C}$ until further processing by solid-phase extraction (SPE) and LC-MS analysis.

Boiling Ethanol: 2 mL of molecular-grade ethanol was preheated in 15 mL conical tubes using a $100\text{ }^{\circ}\text{C}$ water bath. One by one, frozen samples were removed from liquid nitrogen, lids were discarded, and tubes were immediately submerged in boiling ethanol. After brief mixing by inversion, samples were incubated in the water bath for 3 minutes, then cooled on ice and transferred to fresh 2 mL Eppendorf tubes for further processing.

Boiling Water: 2 mL of miliQ water was added to 15 mL conical tubes and preheated in a $100\text{ }^{\circ}\text{C}$ water bath. Frozen samples were retrieved individually from liquid nitrogen, the pre-cut lids were removed, and the tubes were immediately submerged in the boiling water. After brief mixing by inversion, the samples were incubated at $100\text{ }^{\circ}\text{C}$ for 3 minutes to ensure enzyme inactivation. Tubes were then cooled on ice and transferred to fresh 2 mL Eppendorf tubes for subsequent processing.

Solid phase extraction of metabolites. Samples were centrifuged for 15 minutes at 14000 rcf and 4 °C using the Eppendorf centrifuge 5424R, after which the supernatant was transferred to a clean Eppendorf tube. The samples were subsequently washed and concentrated with a HyperSep™ Hypercarb™ SPE 96-well plate (25/1 ml Plate, Thermo Scientific). In short, the columns were preconditioned by washing with 600 µL 10 mM ammonium bicarbonate (ABC) in 60% acetonitrile (ACN), followed by an equilibration with 2 x 600 µL MS-grade H₂O. The samples were loaded onto the columns and a wash step with 1 ml MS-grade H₂O was performed. The samples were eluted using 2 x 200 µL 10 mM ABC in 60 % ACN. The samples were collected in clean LoBind Eppendorf tubes and dried using the SpeedVac concentrator plus (Eppendorf) at room temperature. The dried samples were redissolved in 20 µL MS-grade H₂O prior to analysis by LC-MS.

PGC-HRMS metabolomics. Liquid chromatographic separation coupled to high resolution mass spectrometric detection was conducted using an Acquity Ultra Performance Liquid Chromatograph (Waters, UK) coupled to a Q Exactive™ Focus Hybrid Quadrupole-Orbitrap™ Mass Spectrometer (Thermo Scientific, Germany). Chromatographic separation was achieved on a 100 x 1 mm Hypercarb™ Porous Graphitic Carbon HPLC Column (Thermo Scientific, No. 35005-101030) using a constant flow rate of 100 µL/min. The mobile phases consisted of 50 mM ammonium bicarbonate in MS-grade H₂O (Phase A) and 100% MS-grade acetonitrile (Phase B). A linear gradient from 2.5% to 35% Phase B was applied over 18 minutes, followed by a gradient increase to 60% B over the next 6 minutes. The system was then returned to the initial conditions and equilibrated for 4 minutes. Each sample was analysed in duplicate using 6 µL injections, followed by two blank runs. High resolution mass spectrometric detection was performed in negative ion mode, scanning across a mass range of 100–850 m/z. The mass spectrometer was set to a resolution of 35K, with an AGC target of 1.0E6 and a maximum injection time of 100 ms. Raw data were processed using XCalibur 4.1 (Thermo Fisher Scientific, Germany). In addition, raw MS files were converted to mzXML files using MSConvertGUI, and the integration of nucleotides was performed in Matlab R2022a (MathWorks). Calibration of the mass spectrometer was performed using the Pierce™ LTQ ESI negative ion calibration solution (Thermo Fisher Scientific, Germany).

4.3 | Results

4.3.1 Establishing a metabolomics pipeline for nucleotide quantification in EBPR granular biomass.

A stable isotope dilution mass spectrometry pipeline was developed for quantifying the nucleotides AMP, ADP, ATP, UMP, UDP, UTP, GMP, GDP, GTP, CMP, CDP and CTP in a granular biofilm from a lab scale EBPR system. This included homogeneous biomass sampling, selecting an effective quenching and metabolite extraction method, establishing an effective chromatographic separation of nucleotides and validating possible interference from extracellular nucleotides (Figure 4.1A). Finally, we demonstrated the method by monitoring nucleotide concentrations throughout an EBPR anaerobic-aerobic cycle.

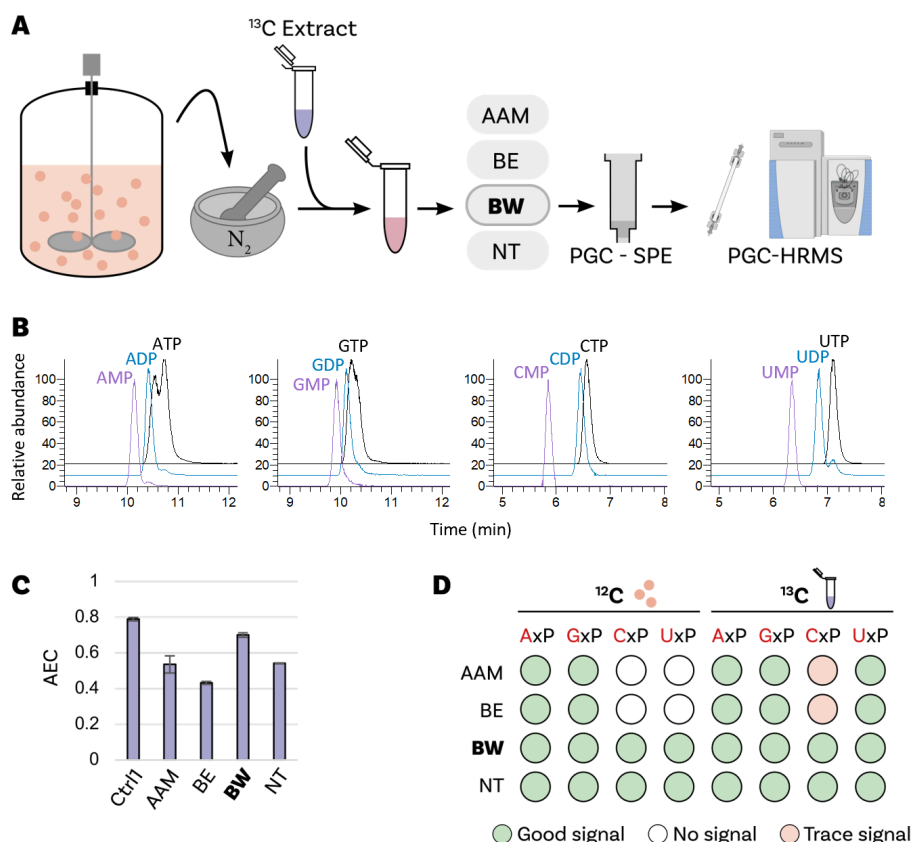


Figure 4.1. A) Schematic overview of the workflow to quantify nucleotides in a granular biofilm microbial community performing EBPR. Broth samples of the lab scale EBPR system were frozen in liquid nitrogen and crushed with a mortar and pestle. The collected samples were spiked with uniformly ^{13}C -labeled yeast extract followed by either an acidic acetonitrile methanol extraction (AAM), a boiling ethanol extraction (BE), a boiling water extraction (BW) or no treatment (NT). A solid

phase extraction (SPE) was performed followed by a chromatographic separation on porous graphite carbon (PGC) and high resolution mass spectrometric analysis. **B)** Extracted ion chromatograms of the uniformly ^{13}C -labelled nucleotides from the analysed ^{13}C yeast extract. **C)** Adenylate energy charge (AEC) calculated from the ^{13}C nucleotide PGC-HRMS signals of the spiked ^{13}C yeast extract, after performing different extraction approaches. $\text{AEC} = ([\text{ATP}] + 0.5[\text{ADP}]) / ([\text{ATP}] + [\text{ADP}] + [\text{AMP}])$. **D)** Detectability of the ^{12}C nucleotides and the ^{13}C isotope variants after performing various extraction methods. **Ctrl1:** ^{13}C yeast extract.

The separation prior to mass spectrometric detection was performed using porous graphitic carbon (PGC) as the chromatographic separation phase combined with ammonium bicarbonate as mobile phase. PGC was chosen for its previously demonstrated effectiveness in retaining and separating nucleotides using volatile buffers³⁶⁻³⁸. To address common issues such as retention time instability and column polarization^{39,40}, the ammonium bicarbonate concentration in the mobile phase was increased to 50 mM, and electrical grounding points were installed before and after the separation column. As a result, interday retention time shifts were limited to ± 1 minute (see SI DOC). Moreover, the PGC column enabled the separation of nucleotides in the order of mono-, di-, and tri-phosphates, allowing to estimate the contribution of in-source fragmentation to the abundance of analysed metabolites (Figure 4.1B).

Calibration curves of all nucleotides were established using synthetic reference standards spiked with a ^{13}C -labeled yeast extract³⁰, to allow determination of absolute nucleotide concentrations. The $^{12}\text{C}/^{13}\text{C}$ signal intensities of each nucleotide was plotted against the known ^{12}C nucleotide concentrations. All calibration curves demonstrated strong linearity (R^2 approaching 1) within the concentration range of 0.7-2150 μM (Table 4.1, SI DOC Figure S4.4).

Table 4.1. Signal response curves for quantified metabolites by plotting the $^{12}\text{C}/^{13}\text{C}$ signal intensity of each metabolite against its ^{12}C metabolite concentration. Linear fits were forced through zero, covering a concentration range of 0.7–2150 μM , where the squared correlation coefficient (R^2) is provided, confirming excellent linearity. In addition, intra- and extracellular metabolite concentrations of the granular biofilm microbial community are provided. The concentrations are averages of two biological duplicates, which were each analysed twice (technical duplicates). The standard deviation is provided (+/- standard deviation). For all nucleotides, the extracellular concentration was < 1%.

	R^2 for metabolite range: 0.7–2150 μM	Broth nucleotides (μM)	Extracellular nucleotides (μM)	Extracellular nucleotides (%)	Average nucleotide concentration (mmol/mg X)
ATP	0.9931	58.1 +/- 3.6	0.48 +/- 0.07	0.83	1.51 (+/- 0.30)
ADP	0.9919	63.0 +/- 1.8	0.48 +/- 0.06	0.76	0.78 (+/- 0.11)
AMP	0.9905	84.5 +/- 2.1	0.66 +/- 0.08	0.78	0.45 (+/- 0.13)
GTP	0.9889	39.3 +/- 3.6	0.14 +/- 0.02	0.35	0.67 (+/- 0.12)
GDP	0.9904	21.3 +/- 0.5	0.09 +/- 0.02	0.40	0.20 (+/- 0.02)
GMP	0.9880	26.5 +/- 0.8	0.17 +/- 0.03	0.64	0.28 (+/- 0.06)
CTP	0.9841	134.7 +/- 0.4	0.10 +/- 0.06	0.08	2.25 (+/- 0.54)
CDP	0.8784	9.9 +/- 0.2	0.00 +/- 0.003	0.03	0.09 (+/- 0.02)
CMP	0.9926	7.5 +/- 0.3	0.03 +/- 0.001	0.40	0.06 (+/- 0.02)
UTP	0.9932	64.2 +/- 0.2	0.34 +/- 0.04	0.53	1.44 (+/- 0.31)
UDP	0.9658	37.0 +/- 0.9	0.17 +/- 0.003	0.47	0.37 (+/- 0.08)
UMP	0.9804	24.9 +/- 2.0	0.00 +/- 0.00	0.00	0.17 (+/- 0.06)

In order to assess the sampling procedure, stirred-tank bioreactor samples were immediately immersed in liquid nitrogen and crushed into a fine powder, both to quench the metabolism as well as to facilitate metabolite extraction. This approach was evaluated and shown to ensure homogeneous biomass sampling from the bioreactor, yielding an average biomass concentration of 6.77 ± 0.06 g/L (triplicate samples).

Subsequently, four metabolite extraction methods were evaluated, including an acidic acetonitrile-methanol extraction followed by neutralization (AAM), a boiling ethanol extraction (BE), a boiling water extraction (BW), and one without additional treatment (NT). To assess metabolite conversion and degradation, all samples were spiked with ^{13}C -labeled yeast extract³⁰. The adenylate energy charge (AEC =

$\frac{[ATP]+0.5*[ADP]}{[ATP]+[ADP]+[AMP]})$ of the ^{13}C -labeled metabolites in the samples following the different extraction procedures was compared to that of an untreated ^{13}C -labeled yeast extract (control). Among the extraction methods, the BW method yielded an AEC of 0.70 ± 0.01 , which was closest to the control value of 0.79 ± 0.01 (Figure 4.1C). In contrast, the AEC for samples extracted using AAM, BE, and NT showed a more pronounced decrease, with values of 0.54 ± 0.05 , 0.43 ± 0.01 , and 0.54 ± 0.002 , respectively. In addition, all possible ^{13}C nucleotide ratios (NTP/NDP, NDP/NMP, and NTP/NMP) were compared against the control sample. The NTP/NDP ratios for all nucleotides followed a similar pattern to that seen for the AEC, with the BW method producing values closest to the control (SI DOC Figure S4.2). However, although the BW method also yielded NTP/NMP and NDP/NMP ratios that were generally closest to the control, these ratios were often significantly lower than those of the control sample (SI DOC Figure S4.2). This indicates that still some conversion/degradation of the cellular metabolome was ongoing, likely increasing the levels of NMPs. Further, the ^{12}C nucleotides were analysed for the different extraction procedures (SI DOC Figure S4.3). Only trace signals were detected for C_xP and U_xP (x representing mono-, di-, or tri-phosphate) in samples extracted using the AAM and BE method (Figure 4.1D, SI DOC Figure S4.3), while a good signal was obtained for the BW method. Since the BW extraction introduced the least bias in the ^{13}C AEC and ^{13}C nucleotide ratios, while also providing ^{12}C signals for all nucleotides of interest, this method was chosen for the remainder of the study.

Furthermore, to assess potential metabolite degradation during sample storage in the autosampler prior to analysis, ^{13}C -labeled yeast extract was injected in regular time frames throughout the analysis period. Both the AEC and nucleotide ratios were monitored, revealing that the AEC and most nucleotide ratios remained stable over time, indicating minimal to no nucleotide degradation. This control was performed during all analyses (SI DOC Figure S4.1.1, S4.1.2, and S4.1.3).

Finally, since samples are immediately frozen and pulverized upon collection, no distinction is made between intra- and extracellular nucleotides. This method was chosen to minimize sampling time and capture the metabolome as accurately as possible at a specific time point. Therefore, no separation of cell biomass from the supernatant was performed. To support this choice, an additional evaluation was conducted to assess whether extracellular nucleotide levels interfere with the quantification. Alongside the metabolite extraction of a regular broth sample (biomass in medium), an extraction was performed on a filtered broth sample, representing the extracellular environment. Extracellular nucleotide

concentrations were found to contribute to less than 1% to the total broth concentration (Table 4.1), confirming that extracellular metabolite levels are negligible.

4.3.2. Variations in the nucleotide pool of PAOs in response to the environment.

The established metabolomics method was employed to a typical anaerobic / aerobic EBPR cycle enriched with granular biomass. The microbial community contained various bacterial groups, with *Candidatus* Contendobacter and *Dechloromonas* showing the highest relative abundance based on taxonomic profiling using metaproteomics (Figure 4.2C). A characteristic EBPR profile was observed during the cycle¹⁵. In the initial anaerobic phase (first 30 minutes), acetate uptake coincided with the degradation of polyphosphate and glycogen, resulting in a P/C ratio of 0.376 P-mol released per C-mol uptake. Simultaneously, polyhydroxyalkanoate (PHA) accumulation occurred, predominantly as polyhydroxybutyrate (PHB), with a smaller fraction of polyhydroxyvalerate (PHV). During the aerobic phase, these PHAs were consumed, leading to the uptake of phosphate from the media and the replenishment of polyphosphate and glycogen (Figure 4.2B).

Quantification of nucleotide energy charges revealed distinct dynamics across the cycle (Figure 4.2A). The adenylate energy charge (AEC), calculated as $([ATP] + 0.5[ADP])/([ATP] + [ADP] + [AMP])$, increased during the first hour of the anaerobic phase, reaching a peak of approximately 0.72. This increase coincided with acetate uptake and phosphate release. Following acetate depletion, the AEC plateaued for the remainder of the anaerobic phase. Upon transition to aerobic conditions, a minor transient increase was observed, followed by a gradual decline in AEC, returning to the initial value of ~0.6 by the end of the cycle.

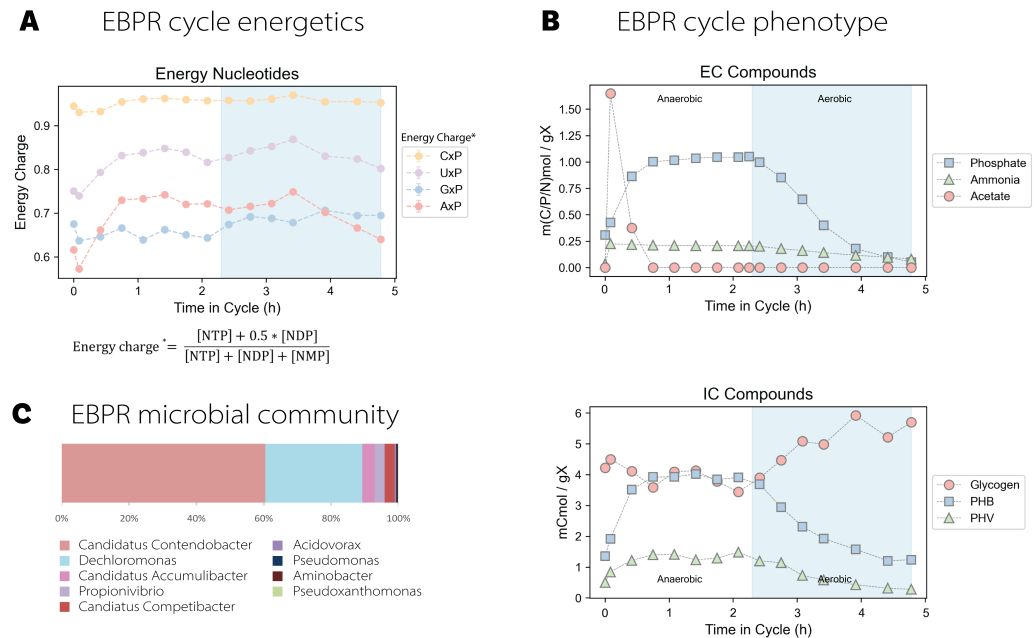


Figure 4.2. Anaerobic/aerobic cycle of a granular biofilm microbial community performing EBPR fed with acetate. **A)** The energy balance for AxP, GxP, CxP and UxP was calculated by: energy balance = $\frac{([NTP] + 0.5[NDP])}{([NTP] + [NDP] + [NMP])}$. **B)** The phosphate, ammonia and acetate concentrations are presented in m(C/P/N)mol/gX, and the glycogen, PHB and PHV concentrations in mCmol/gX. **EC:** extracellular; **IC:** intracellular. **C)** The composition of the microbial community as determined by metaproteomics using NovoBridge⁴⁹.

The uridylate energy charge (UEC), calculated analogously as $\frac{([UTP] + 0.5[UDP])}{([UTP] + [UDP] + [UMP])}$, followed a similar trend to AEC, with a peak during the anaerobic phase and a gradual decline during the aerobic phase. The guanylate energy charge (GEC) showed modest variation, with a slight increase during the aerobic phase. In contrast, the cytidylate energy charge (CEC) remained consistently high and stable throughout the entire cycle, with minimal differences between the anaerobic and aerobic phases. Together, the data show that nucleotide energy charge varies over the EBPR cycle, with adenylate and uridylate pools exhibiting more pronounced temporal dynamics than guanylate and cytidylate pools.

4.4 | Discussion

In this study, we established a targeted and quantitative workflow to quantify nucleotide pools in granular biofilms, as exemplified for a microbial community performing EBPR. In the following, we elaborate on the strengths and potential limitations of the established pipeline. Subsequently, we discuss the observed

nucleotide dynamics in relation to EBPR transformations during anaerobic–aerobic cycling.

4.4.1. A quantitative nucleotide metabolomics pipeline for granular biofilms.

Small changes in energy nucleotides can significantly impact metabolic flows, thereby holding valuable information. Although numerous metabolomics studies have been conducted on planktonic cells, limited method development has been dedicated to granules or biofilms. Most of these efforts predominantly focused on untargeted methods to maximize metabolite identification⁵⁰. While useful for comparing planktonic and biofilm states or distinguishing bacterial species^{25,26,51}, these approaches lack the precision required for quantitative metabolome analyses in granules and biofilms. Moreover, common pre-quenching sample handling steps, including centrifugation and washing, can distort the metabolite ratios and should be avoided^{27,29}.

The presented method enables rapid, homogeneous biomass freezing. Due to the speed of metabolite conversions, minor environmental shifts may still alter the metabolome; such conversions could be reduced by even faster sampling. In yeast research, rapid sampling devices have reduced sampling times to <1 second⁵². Nevertheless, the tubes and sampling ports required for high speed are too small for working with larger aggregates of biomass such as granules.

Among the tested extraction methods, the BE extraction proved inadequate (Figure 4.1C), possibly due to a poor recovery of phosphorylated compounds, as previously reported⁵³. Both the NT and AAM extractions resulted in altered nucleotide profiles compared to the control, indicating potential degradation or residual enzymatic activity. Although other studies have reported improved triphosphate recovery using similar cold acidic²⁹ or acidic acetonitrile/methanol/water (40:40:20)¹⁹ extractions, those analyses were performed on planktonic *E. coli* cells and did also not include an extraction with boiling water. The structurally distinct biomass used in this study (granules) may account for differences in extraction efficiency and possible degradation processes.

The BW extraction method proved to be the most effective (Figure 4.1C), yielding ¹³C AEC and NTP/NDP ratios closest to the control. A hot water extraction has previously enabled the quantification of over 50 hydrophilic metabolites in planktonic *E. coli*⁵⁴, including nucleotides, yet that study did include a centrifugation step, which may alter metabolite ratios and should be avoided. Despite the effectiveness of the BW method, deviations in the NDP/NMP and NTP/NMP ratios were observed compared to the control, suggesting the presence of residual enzymatic activity, metabolite degradation and/or matrix effects during

MS analysis. Although the spiked ^{13}C yeast extract can account for some of these factors, it cannot correct for enzymatic interconversion between nucleotides, unless the nucleotide concentrations in the spiked standard closely match those in the actual sample. This limitation could be addressed by generating a uniformly ^{13}C -labeled extract from the enrichment culture of interest, thereby replacing the ^{13}C yeast extract in the experiments. As such, the current method requires cautious interpretation of the NDP/NMP and NTP/NMP ratios.

Previous research has shown that metabolite profiles can differ between bacteria growing in planktonic form and those in a biofilm state^{25,26,51}. An intriguing question that remains is whether bacteria also exhibit distinct metabolic profiles based on their spatial location, such as the surface versus the core of a granule. These differences could arise from variations in access to nutrients and oxygen, influenced by factors like granule size, density, and the diffusion coefficient of solutes⁵⁵. Smaller granules are likely to have a more uniform metabolome. The current method averages measurements across the entire bacterial population, potentially masking spatial variations. Future advancements might help to address these type of questions. Nevertheless, the employed approach reveals how the energy state changes when varying the environmental conditions.

4.4.2. Insights into the metabolism of EBPR microbial community under dynamic conditions.

Our results show that the adenylate energy charge (AEC) changes dynamically throughout the EBPR cycle and may drive rapid acetate uptake and polyphosphate degradation. These findings support the hypothesis, backed by *in vitro* studies, that polyphosphate degradation is triggered by imbalances in the adenylate pool (ATP, ADP and AMP)^{18,56}. Interestingly, the anaerobic phase begins with a very low AEC (Figure 4.2A), suggesting the cells are primed for substrate uptake and energy mobilization. The magnitude of AEC fluctuation is remarkable—previous studies on *E. coli*¹³ and *Synechococcus elongatus*⁵⁷ report minimal changes in energy charge under varying environmental conditions. Although transient drops in AEC have been observed in *E. coli* under sudden starvation⁵⁸, cells adapted to feast–famine regimes maintain a stable AEC^{13,59}. In contrast, our data indicate that even in an adapted PAO community, AEC remains highly responsive to environmental transitions.

Notably, the most pronounced changes in AEC occurred during anaerobic substrate uptake and toward the end of the aerobic phase, while the transition between anaerobic and aerobic conditions had little immediate effect. The uridylate energy charge (UEC) showed a gradual increase under aerobic conditions,

which may reflect increased biosynthetic activity, particularly in protein translation⁶⁰. This aligns with UTP's role in anabolic processes and supports previous reports of oxygen-induced biosynthetic activation⁶¹. Remarkably, despite the major metabolic shift that occurs upon oxygen introduction—reversing PHA synthesis, and glycogen and polyphosphate degradation—the energy nucleotide pools remained relatively stable. This suggests that the switch is regulated by factors outside the adenylate and uridylate pools, possibly involving redox cofactors such as NAD(P)H/NAD(P)⁺. While these metabolites are likely central to this transition, their known instability^{62,63} poses significant challenges for accurate quantification. These findings highlight the importance of quantitative metabolomics in capturing rapid microbial responses and provide a foundation for further studies into the regulation of energy and redox metabolism in complex biofilm systems.

4.5 | Supplementary information

4.5.1. Evaluating metabolite stability during sample analysis.

The sample stability in terms of degradation was determined using the adenylate energy charge (AEC) and the nucleotide ratios of the same ^{13}C -labeled yeast extract as proxy. For this, the ^{13}C -labeled yeast extract was injected throughout the sample analysis at regular time intervals (Figure S4.1.1, S4.1.2, and S4.1.3).

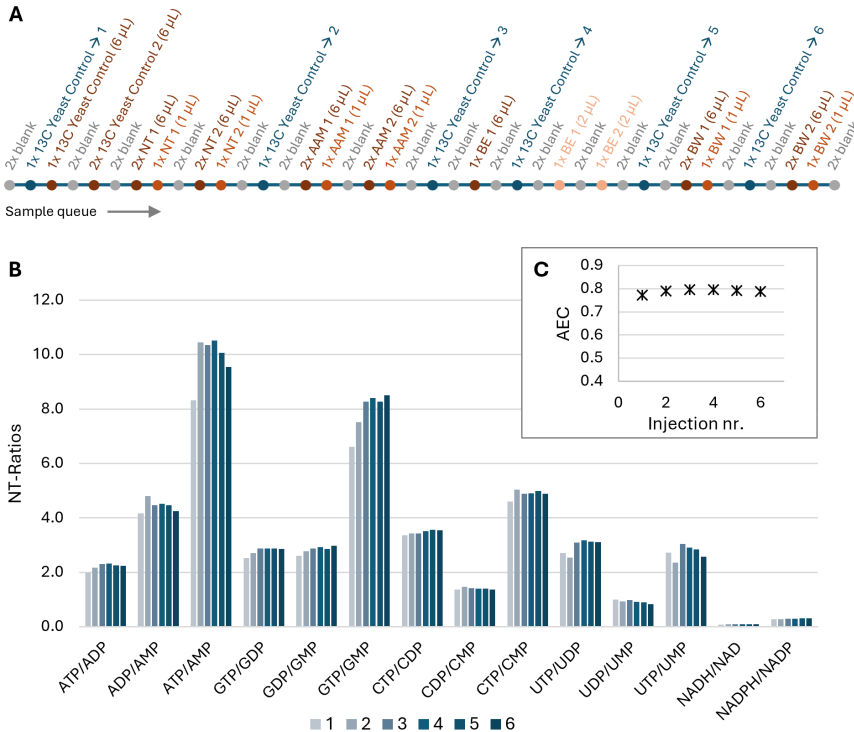


Figure S4.1.1. **A)** PGC-HRMS sample queue. Nucleotides from a granular biofilm enrichment performing Enhanced Biological Phosphorous Removal (EBPR) were extracted employing different extraction methods: NT: no treatment, AAM: acidic acetonitrile-methanol extraction, BE: boiling ethanol extraction, BW: boiling water extraction. **B)** Nucleotide ratios of the ^{13}C yeast extract controls. **C)** AEC (adenylate energy charge) of the ^{13}C yeast extract controls. $\text{AEC} = ([\text{ATP}] + 0.5 \cdot [\text{ADP}]) / ([\text{ATP}] + [\text{ADP}] + [\text{AMP}])$. The energy charge and nucleotide ratios are calculated based on the ^{13}C signals of the metabolites. 1-6 corresponds to the ^{13}C yeast extract control injections as indicated in **A)**.

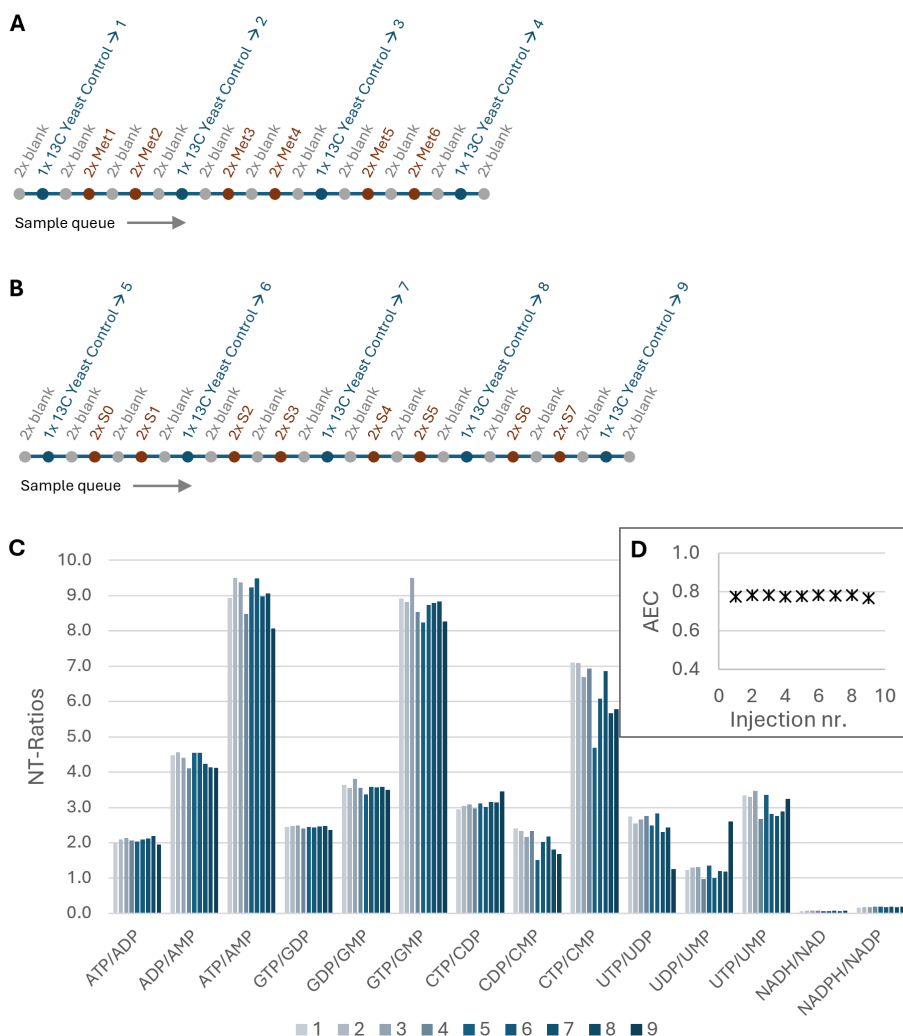


Figure S4.1.2. A) PGC-HRMS sample queue. Met1, Met2: Biomass from the lab-scale EBPR system was crushed in liquid nitrogen after which ^{13}C yeast extract was added. Metabolites were extracted with the boiling water method. Met3, Met4: Biomass from the lab-scale EBPR system was crushed in liquid nitrogen, ^{13}C yeast extract was added before the crushing. Metabolites were extracted with the boiling water method. Met5, Met6: Biomass from the lab-scale EBPR system was filtered using a 0.42 μm filter. The filtrate was subsequently further processed, representing the extracellular space. **B)** PGC-HRMS sample queue. S0-S7: calibration samples, mixtures of ^{12}C metabolites of different concentrations spiked with 25 μL ^{13}C yeast extract. **C)** Nucleotide ratios of the ^{13}C yeast extract controls. **D)** AEC (adenylate energy charge) of the ^{13}C yeast extract controls. $\text{AEC} = ([\text{ATP}] + 0.5 \cdot [\text{ADP}]) / ([\text{ATP}] + [\text{ADP}] + [\text{AMP}])$. The energy charge and nucleotide ratios are calculated based on the ^{13}C signals of the metabolites. 1-9 corresponds to the injected ^{13}C yeast extract control as indicated in **A)** and **B)**.

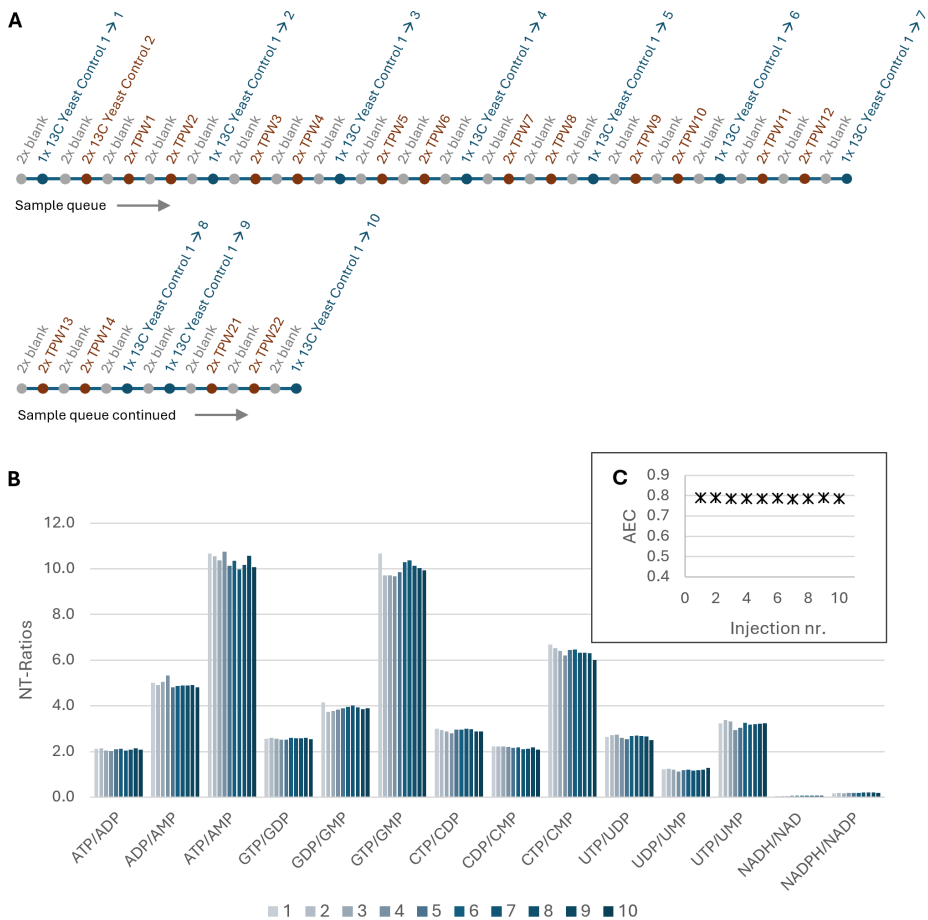


Figure S4.1.3. A) PGC-HRMS sample queue. TPW1-TPW14, TPW21-TPW22: metabolite extracts of the granular biofilm microbial community performing Enhanced Biological Phosphorous Removal (EBPR). **B)** Nucleotide ratios of the ¹³C yeast extract controls. **C)** AEC (adenylate energy charge) of the ¹³C yeast extract controls. $AEC = ([ATP] + 0.5 \cdot [ADP]) / ([ATP] + [ADP] + [AMP])$. The energy charge and nucleotide ratios are calculated based on the ¹³C signals of the metabolites. 1-10 correspond to the injected ¹³C yeast extract control as indicated in **A)**.

4.5.2. Nucleotide ratios of spiked ^{13}C yeast after employing different extraction protocols.

The nucleotide ratios of the ^{13}C -labeled yeast extract that was spiked to the granules from the lab-scale EBPR system, and which were extracted with different extraction protocols, were compared to the ^{13}C -labeled yeast extract control (Figure S4.2).

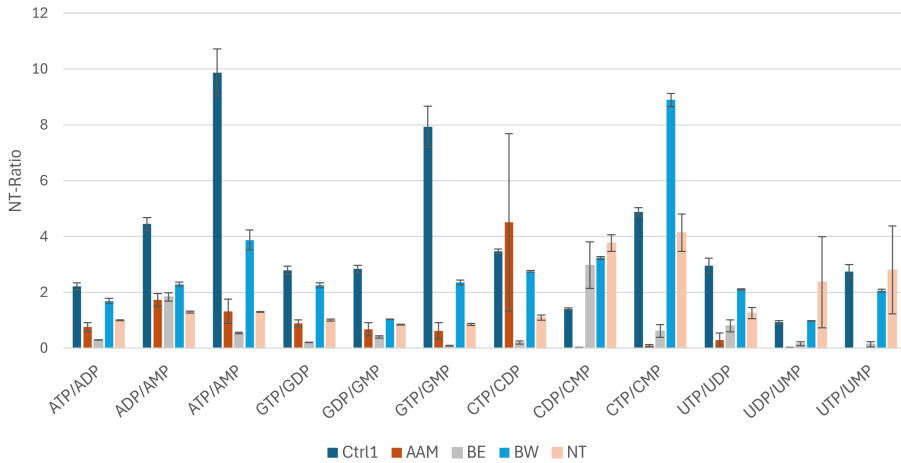


Figure S4.2. Nucleotide ratios of the spiked ^{13}C yeast extract after employing different extraction protocols to granules from the lab-scale EBPR system. AAM: acidic acetonitrile-methanol extraction, BE: boiling ethanol extraction, BW: boiling water extraction, NT: no treatment. Ctrl1: untreated ^{13}C yeast extract.

4.5.3. Nucleotide areas of extracted granules from lab-scale EBPR system.

The nucleotide signals of the extracted granules from the lab-scale EBPR system were compared to determine the extraction efficiency of different extraction protocols (Figure S4.3). No significant amount of ^{12}C metabolites were observed in the uniformly ^{13}C -labeled yeast extract control, which confirms, that all extracted metabolite signals originated from the EBPR microbes. Interestingly, only trace signals were observed for CTP, CDP, CMP, UTP, UDP, and UMP in the samples extracted with the acidic acetonitrile-methanol (AAM) method and the boiling ethanol (BE) method.

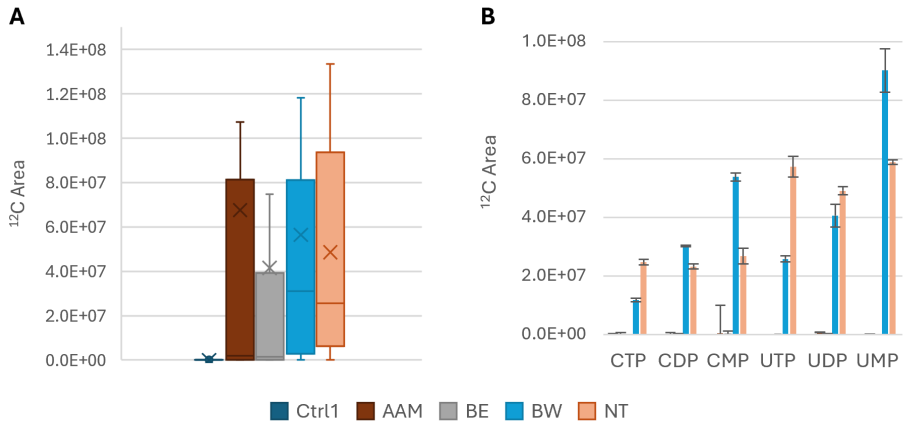


Figure S4.3. Differently extracted granules from the lab-scale EBPR system. AAM: acidic acetonitrile-methanol extraction, BE: boiling ethanol extraction, BW: boiling water extraction, NT: no treatment. Ctrl1: uniformly ^{13}C -labeled yeast extract. **A)** Boxplot of the summed ^{12}C nucleotides. **B)** Uridine and cytidine nucleotide signal intensities.

4.5.4. Nucleotide calibration curves.

Calibration curves were established for all nucleotides by plotting the $^{12}\text{C}/^{13}\text{C}$ HRMS signal against the known ^{12}C nucleotide concentrations (Figure S4.4) to allow absolute quantification of the extracted nucleotides from the microbes in the EBPR system.

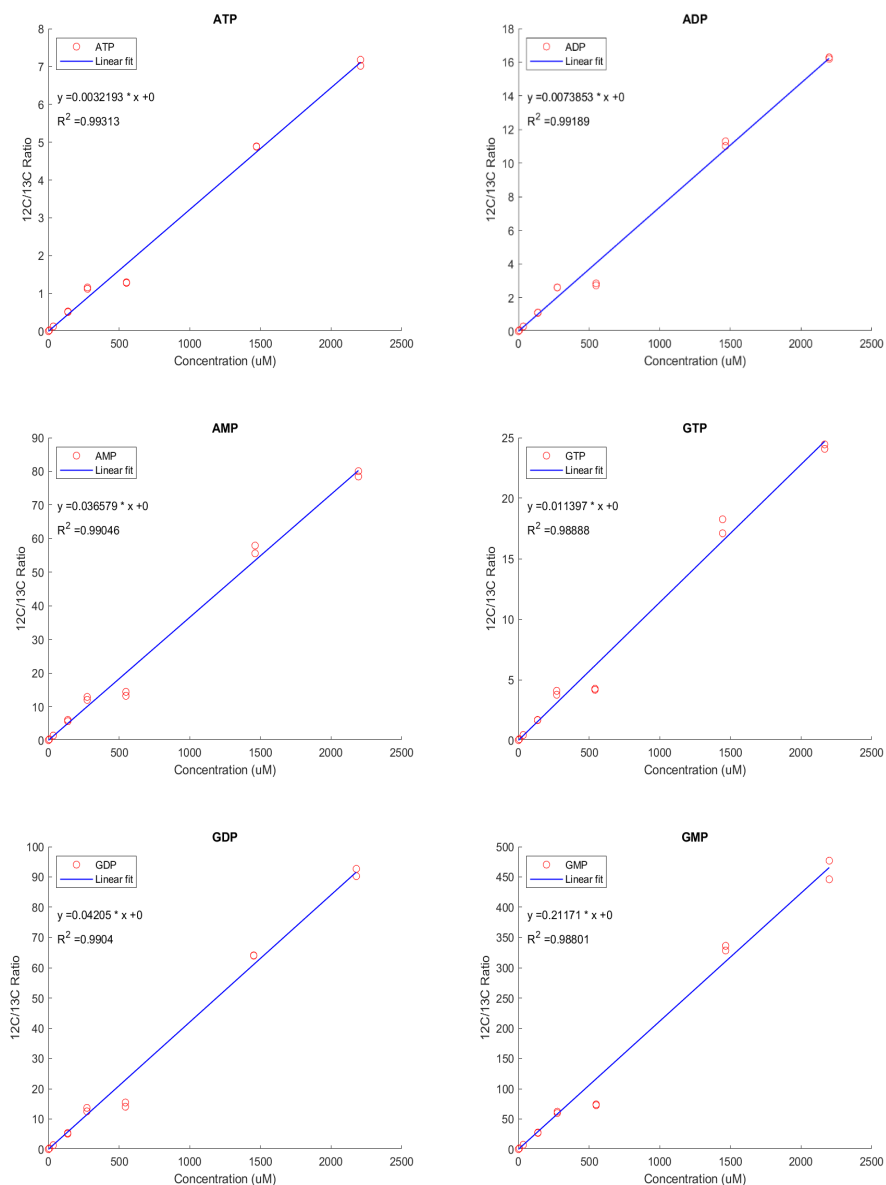


Figure S4.4. Calibration curves of the nucleotides AxP, GxP, CxP and UxP. The $^{12}\text{C}/^{13}\text{C}$ HRMS signal is plotted against the known ^{12}C nucleotide concentrations.

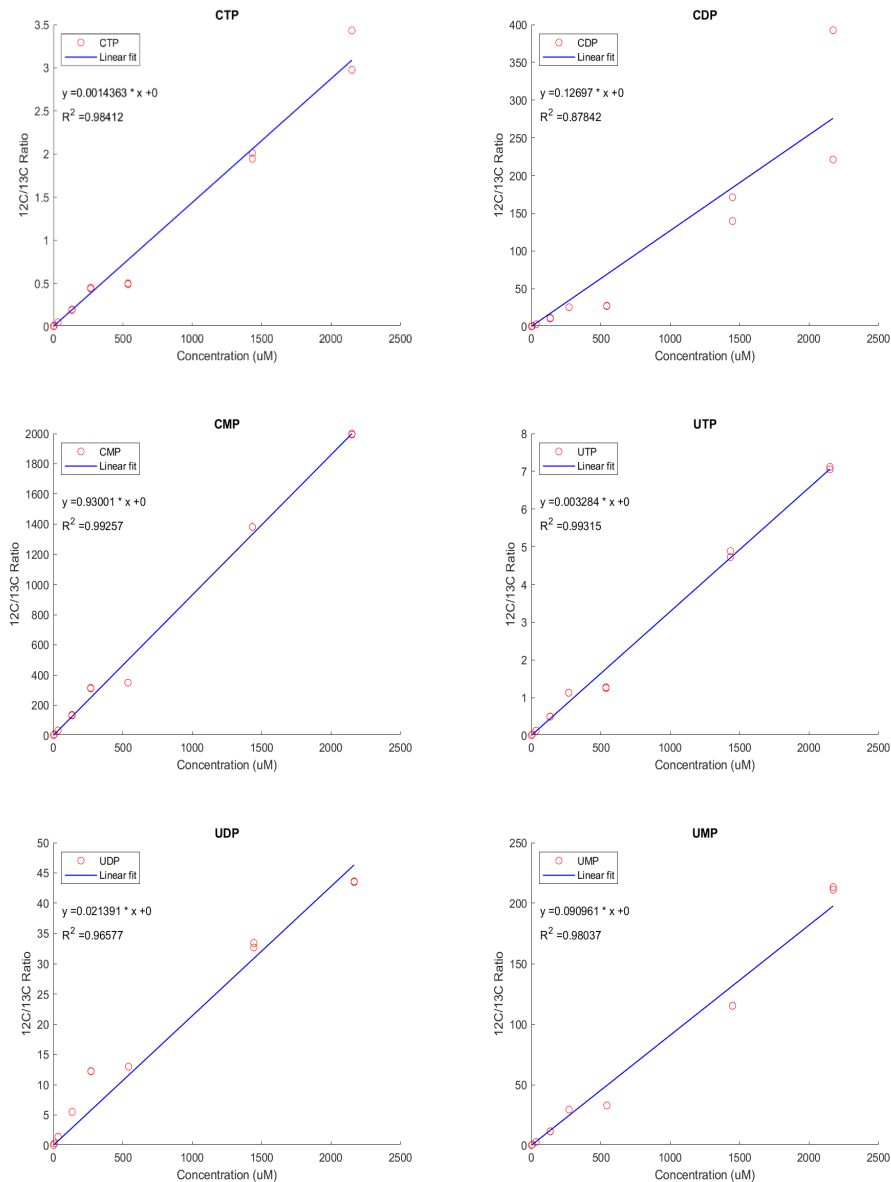


Figure S4.4 continued. Calibration curves of the nucleotides AxP, GxP, CxP and UxP. The $^{12}\text{C}/^{13}\text{C}$ HRMS signal is plotted against the known ^{12}C nucleotide concentrations.

4.5.5. Nucleotide elution profiles on porous graphitic carbon (PGC).

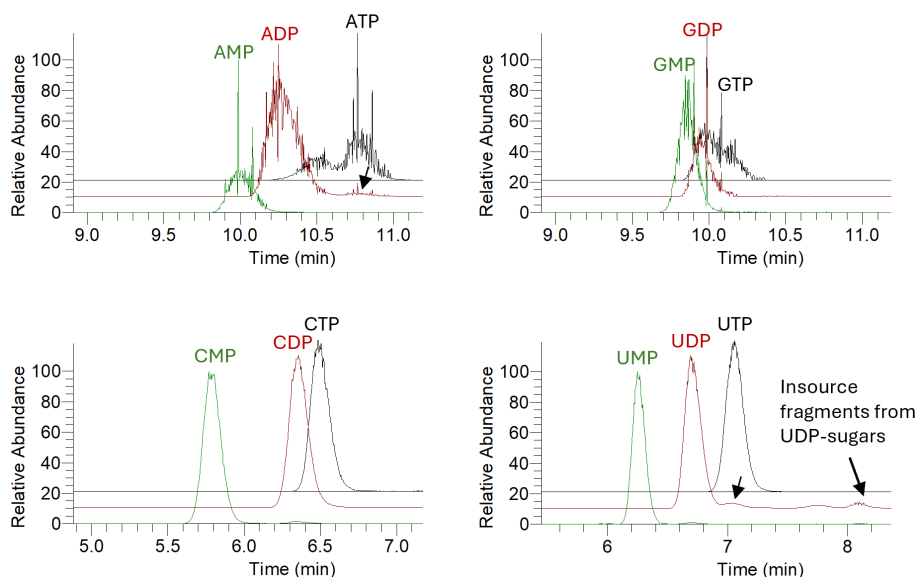


Figure S4.5.1. Extracted ion chromatograms (XIC) of the ^{12}C nucleotides from a metabolite extract of a granular biofilm microbial community performing enhanced biological phosphorus removal (EBPR). Analysis was performed by high resolution mass spectrometry in negative mode, using porous graphitic carbon (PGC) as separation phase. The masses were extracted with a 5 ppm error range. Arrows indicate insource fragments.

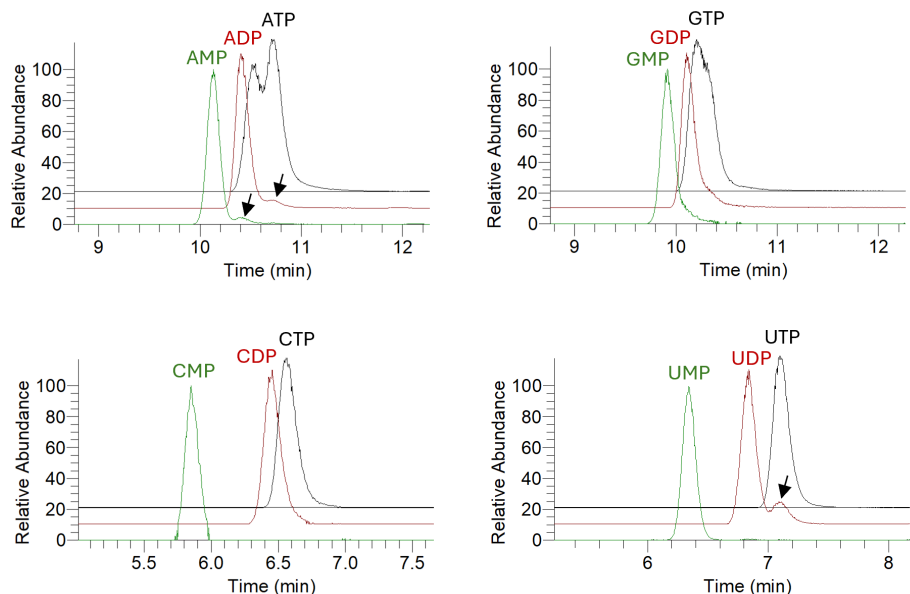


Figure S4.5.2. Extracted ion chromatograms (XIC) of the ^{13}C -labelled nucleotides in a uniformly ^{13}C -labelled yeast extract measured by high resolution mass spectrometry in negative mode, using porous graphitic carbon (PGC) as separation phase. The masses were extracted with a 5 ppm error range. Arrows indicate in-source fragments.

4.5.6. Metagenome taxonomic profiling.

Whole metagenome sequencing was performed followed by taxonomic classification of raw sequencing reads using Kraken2 (Figure S4.6).



Figure S4.6. Microbial composition of the granular biofilm performing enhanced biological phosphorus removal (EBPR) based on taxonomic classification of raw sequencing reads obtained from whole metagenome sequencing.

References

- 1 Song, W. *et al.* Functional traits resolve mechanisms governing the assembly and distribution of nitrogen-cycling microbial communities in the global ocean. *MBio* **13**, e03832–03821 (2022).
- 2 Wang, C.-y. *et al.* Soil pH is the primary factor driving the distribution and function of microorganisms in farmland soils in northeastern China. *Annals of Microbiology* **69**, 1461–1473 (2019).
- 3 Dell, A. I., Pawar, S. & Savage, V. M. Temperature dependence of trophic interactions are driven by asymmetry of species responses and foraging strategy. *Journal of Animal Ecology* **83**, 70–84 (2014).
- 4 Oehmen, A., Carvalho, G., Lopez-Vazquez, C., Van Loosdrecht, M. & Reis, M. Incorporating microbial ecology into the metabolic modelling of polyphosphate accumulating organisms and glycogen accumulating organisms. *Water research* **44**, 4992–5004 (2010).
- 5 Ma, H.-W., Buer, J. & Zeng, A.-P. Hierarchical structure and modules in the *Escherichia coli* transcriptional regulatory network revealed by a new top-down approach. *BMC bioinformatics* **5**, 199 (2004).
- 6 Liebermeister, W. *et al.* Visual account of protein investment in cellular functions. *Proceedings of the National Academy of Sciences* **111**, 8488–8493 (2014).
- 7 Milo, R. & Phillips, R. *Cell biology by the numbers*. (Garland Science, 2015).
- 8 den Ridder, M., Daran-Lapujade, P. & Pabst, M. Shot-gun proteomics: why thousands of unidentified signals matter. *Fems Yeast Res* **20** (2020). <https://doi.org/ARTNfoz08810.1093/femsyr/foz088>
- 9 Atkinson, D. E. & Walton, G. M. Adenosine Triphosphate Conservation in Metabolic Regulation - Rat Liver Citrate Cleavage Enzyme. *Journal of Biological Chemistry* **242**, 3239–+ (1967).
- 10 Swedes, J. S., Sedo, R. J. & Atkinson, D. E. Relation of growth and protein synthesis to the adenylate energy charge in an adenine-requiring mutant of *Escherichia coli*. *Journal of Biological Chemistry* **250**, 6930–6938 (1975).
- 11 Andersen, K. B. & von Meyenburg, K. Charges of nicotinamide adenine nucleotides and adenylate energy charge as regulatory parameters of the metabolism in *Escherichia coli*. *Journal of Biological Chemistry* **252**, 4151–4156 (1977).
- 12 De la Fuente, I. M. *et al.* On the dynamics of the adenylate energy system: homeorhesis vs homeostasis. *PLoS one* **9**, e108676 (2014).
- 13 Vasilakou, E., Van Loosdrecht, M. C. & Wahl, S. A. *Escherichia coli* metabolism under short-term repetitive substrate dynamics: adaptation and trade-offs. *Microbial Cell Factories* **19**, 116 (2020).
- 14 Steinhoff, R. F. *et al.* Rapid estimation of the energy charge from cell lysates using matrix-assisted laser desorption/ionization mass spectrometry: Role of in-source fragmentation. *Anal Biochem* **447**, 107–113 (2014). <https://doi.org/10.1016/j.ab.2013.11.013>
- 15 Smolders, G., Van der Meij, J., Van Loosdrecht, M. & Heijnen, J. A structured metabolic model for anaerobic and aerobic stoichiometry and kinetics of the biological phosphorus removal process. *Biotechnology and Bioengineering* **47**, 277–287 (1995).
- 16 Wilmes, P. *et al.* Community proteogenomics highlights microbial strain-variant protein expression within activated sludge performing enhanced biological phosphorus removal. *The ISME journal* **2**, 853–864 (2008).
- 17 Páez-Watson, T., van Loosdrecht, M. C. & Wahl, S. A. From metagenomes to metabolism: Systematically assessing the metabolic flux feasibilities for “*Candidatus Accumulibacter*” species during anaerobic substrate uptake. *Water Research* **250**, 121028 (2024).
- 18 Comeau, Y., Hall, K. J., Hancock, R. E. & Oldham, W. K. Biochemical model for enhanced biological phosphorus removal. *Water Research* **20**, 1511–1521 (1986).
- 19 Rabinowitz, J. D. & Kimball, E. Acidic Acetonitrile for Cellular Metabolome Extraction from *Escherichia coli*. *Analytical Chemistry* **79**, 6167–6173 (2007). <https://doi.org/DOI10.1021/ac070470c>
- 20 Maharjan, R. P. & Ferenci, T. Global metabolite analysis: the influence of extraction methodology on metabolome profiles of *Escherichia coli*. *Analytical Biochemistry* **313**, 145–154 (2003).
- 21 Pinu, F. R., Villas-Boas, S. G. & Aggio, R. Analysis of Intracellular Metabolites from Microorganisms: Quenching and Extraction Protocols. *Metabolites* **7** (2017). <https://doi.org/ARTN5310.3390/metabo7040053>
- 22 Canelas, A. B. *et al.* Quantitative Evaluation of Intracellular Metabolite Extraction Techniques for Yeast Metabolomics. *Analytical Chemistry* **81**, 7379–7389 (2009). <https://doi.org/10.1021/ac900999t>
- 23 Hiller, J., Franco-Lara, E. & Weuster-Botz, D. Metabolic profiling of *Escherichia coli* cultivations: evaluation of extraction and metabolite analysis procedures. *Biotechnology Letters* **29**, 1169–1178 (2007). <https://doi.org/10.1007/s10529-007-9384-8>
- 24 Rost, L. M. *et al.* Absolute Quantification of the Central Carbon Metabolome in Eight Commonly Applied Prokaryotic and Eukaryotic Model Systems. *Metabolites* **10** (2020). <https://doi.org/https://doi.org/10.3390/metabo10020074>
- 25 Booth, S. C. *et al.* Differences in Metabolism between the Biofilm and Planktonic Response to Metal Stress. *Journal of Proteome Research* **10**, 3190–3199 (2011). <https://doi.org/10.1021/pr2002353>
- 26 Stipetic, L. H. *et al.* A novel metabolomic approach used for the comparison of planktonic cells and biofilm samples. *Metabolomics* **12** (2016). <https://doi.org/ARTN7510.1007/s11306-016-1002-0>
- 27 Buckstein, M. H., He, J. & Rubin, H. Characterization of nucleotide pools as a function of physiological state in. *Journal of Bacteriology* **190**, 718–726 (2008). <https://doi.org/10.1128/Jb.01020-07>
- 28 Dekoning, W. & Vandam, K. A Method for the Determination of Changes of Glycolytic Metabolites in Yeast on a Subsecond Time Scale Using Extraction at Neutral Ph. *Analytical Biochemistry* **204**, 118–123 (1992). [https://doi.org/Doi10.1016/0003-2697\(92\)90149-2](https://doi.org/Doi10.1016/0003-2697(92)90149-2)
- 29 Varik, V., Oliveira, S. R. A., Haurlyuk, V. & Tenson, T. HPLC-based quantification of bacterial housekeeping nucleotides and alarmone messengers ppGpp and pppGpp. *Scientific Reports* **7** (2017). <https://doi.org/ARTN1102210.1038/s41598-017-10988-6>
- 30 Wu, L. *et al.* Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly C-labeled cell extracts as internal standards. *Analytical Biochemistry* **336**, 164–171 (2005). <https://doi.org/10.1016/j.ab.2004.09.001>

- 31 Seifar, R. M. *et al.* Quantitative analysis of intracellular coenzymes in using ion pair reversed phase ultra high performance liquid chromatography tandem mass spectrometry. *Journal of Chromatography A* **1311**, 115–120 (2013). <https://doi.org/10.1016/j.chroma.2013.08.076>
- 32 Tomiya, N., Ailor, E., Lawrence, S. M., Betenbaugh, M. J. & Lee, Y. C. Determination of nucleotides and sugar nucleotides involved in protein glycosylation by high-performance anion-exchange chromatography: Sugar nucleotide contents in cultured insect cells and mammalian cells. *Analytical Biochemistry* **293**, 129–137 (2001). <https://doi.org/10.1006/abio.2001.5091>
- 33 van Scherpenzeel, M. *et al.* Dynamic tracing of sugar metabolism reveals the mechanisms of action of synthetic sugar analogs. *Glycobiology* **32**, 239–250 (2022). <https://doi.org/10.1093/glycob/cwab106>
- 34 Coulier, L. *et al.* Simultaneous quantitative analysis of metabolites using ion-pair liquid chromatography - Electrospray ionization mass spectrometry. *Analytical Chemistry* **78**, 6573–6582 (2006). <https://doi.org/10.1021/ac0607616>
- 35 van Ede, J. M., Soic, D. & Pabst, M. Decoding Sugars: Mass Spectrometric Advances in the Analysis of the Sugar Alphabet. *Mass Spectrom Rev* (2025). <https://doi.org/10.1002/mas.21927>
- 36 Pabst, M. *et al.* Nucleotide and Nucleotide Sugar Analysis by Liquid Chromatography-Electrospray Ionization-Mass Spectrometry on Surface-Conditioned Porous Graphitic Carbon. *Analytical Chemistry* **82**, 9782–9788 (2010). <https://doi.org/10.1021/ac101975k>
- 37 Xing, J. S., Apedo, A., Tymiak, A. & Zhao, N. Liquid chromatographic analysis of nucleosides and their mono-, di- and triphosphates using porous graphitic carbon stationary phase coupled with electrospray mass spectrometry. *Rapid Communications in Mass Spectrometry* **18**, 1599–1606 (2004). <https://doi.org/10.1002/rcm.1524>
- 38 Bustamante, S., Gilchrist, R. B. & Richani, D. A sensitive method for the separation and quantification of low-level adenine nucleotides using porous graphitic carbon-based liquid chromatography and tandem mass spectrometry. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **1061**, 445–451 (2017). <https://doi.org/10.1016/j.jchromb.2017.07.044>
- 39 Bapiro, T. E., Richards, F. M. & Jodrell, D. I. Understanding the Complexity of Porous Graphitic Carbon (PGC) Chromatography: Modulation of Mobile-Stationary Phase Interactions Overcomes Loss of Retention and Reduces Variability. *Analytical Chemistry* **88**, 6190–6194 (2016). <https://doi.org/10.1021/acs.analchem.6b01167>
- 40 Pabst, M. & Altmann, F. Influence of electrosorption, solvent, temperature, and ion polarity on the performance of LC-ESI-MS using graphitic carbon for acidic oligosaccharides. *Analytical Chemistry* **80**, 7534–7542 (2008). <https://doi.org/10.1021/ac801024r>
- 41 Pérez-Watson, T. *et al.* Sweet Secrets: Exploring Novel Glycans and Glycoconjugates in the Extracellular Polymeric Substances of “Candidatus Accumulibacter”. *ACS Es&t Water* **4**, 3391–3399 (2024).
- 42 Oehmen, A., Keller-Lehmann, B., Zeng, R. J., Yuan, Z. & Keller, J. Optimisation of poly- β -hydroxyalkanoate analysis using gas chromatography for enhanced biological phosphorus removal systems. *Journal of Chromatography A* **1070**, 131–136 (2005).
- 43 Wingett, S. W. & Andrews, S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research* **7**, 1338 (2018).
- 44 Chen, S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *Imeta* **2**, e107 (2023).
- 45 Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome biology* **20**, 257 (2019).
- 46 McMurdie, P. J. & Holmes, S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *Plos One* **8** (2013). <https://doi.org/10.1371/journal.pone.0061217>
- 47 Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome research* **27**, 824–834 (2017).
- 48 van Olst, B. *et al.* Metaproteomic Profiling of the Secretome of a Granule-forming *Ca. Accumulibacter* Enrichment. *Proteomics* **25** (2025). <https://doi.org/10.1002/pmic.202400189>
- 49 Kleikamp, H. B. C. *et al.* Database-independent metaproteomics of complex microbial communities. *Cell Syst* **12**, 375–+ (2021). <https://doi.org/10.1016/j.cels.2021.04.003>
- 50 Favre, L. *et al.* Discrimination of Four Marine Biofilm-Forming Bacteria by LC-MS Metabolomics and Influence of Culture Parameters. *Journal of Proteome Research* **16**, 1962–1975 (2017). <https://doi.org/10.1021/acs.jproteome.6b01027>
- 51 Tang, T. S. *et al.* Evaluation of the differences between biofilm and planktonic via metabolomics and proteomics. *Functional & Integrative Genomics* **21**, 421–433 (2021). <https://doi.org/10.1007/s10142-021-00788-7>
- 52 van Gulik, W. M. *et al.* Fast sampling of the cellular metabolome. *Microbial systems biology: methods and protocols* **881**, 27 (2012).
- 53 Villas-Bôas, S. G., Hojer-Pedersen, J., Åkesson, M., Smedsgaard, J. & Nielsen, J. Global metabolite analysis of yeast: evaluation of sample preparation methods. *Yeast* **22**, 1155–1169 (2005). <https://doi.org/10.1002/yea.1308>
- 54 Teleki, A., Sánchez-Kopper, A. & Takors, R. Alkaline conditions in hydrophilic interaction liquid chromatography for intracellular metabolite quantification using tandem mass spectrometry. *Analytical Biochemistry* **475**, 4–13 (2015). <https://doi.org/10.1016/j.ab.2015.01.002>
- 55 van den Berg, L., Ortega, S. T., van Loosdrecht, M. C. M. & de Kreuk, M. K. Diffusion of soluble organic substrates in aerobic granular sludge: Effect of molecular weight. *Water Research X* **16** (2022). <https://doi.org/ARTN 100148> 10.1016/j.wroa.2022.100148
- 56 McMahon, K. D., Dojka, M. A., Pace, N. R., Jenkins, D. & Keasling, J. D. Polyphosphate kinase from activated sludge performing enhanced biological phosphorus removal. *Applied and environmental microbiology* **68**, 4971–4978 (2002).
- 57 Jaiswal, D. & Wangikar, P. P. Dynamic inventory of intermediate metabolites of cyanobacteria in a diurnal cycle. *IScience* **23** (2020).
- 58 Löffler, M. *et al.* Engineering *E. coli* for large-scale production—strategies considering ATP expenses and transcriptional responses. *Metabolic engineering* **38**, 73–85 (2016).
- 59 Link, H., Anselment, B. & Weuster-Botz, D. Rapid media transition: an experimental approach for steady state analysis of metabolic pathways. *Biotechnology progress* **26**, 1–10 (2010).

- 60 Rodnina, M. V. *et al.* GTPase mechanisms and functions of translation factors on the ribosome. (2000).
- 61 Páez-Watson, T., van Loosdrecht, M. C. & Wahl, S. A. Predicting the impact of temperature on metabolic fluxes using resource allocation modelling: Application to polyphosphate accumulating organisms. *Water Research* **228**, 119365 (2023).
- 62 Ortmayr, K. *et al.* Sample preparation workflow for the liquid chromatography tandem mass spectrometry based analysis of nicotinamide adenine dinucleotide phosphate cofactors in yeast. *J Sep Sci* **37**, 2185–2191 (2014). <https://doi.org/10.1002/jssc.201400290>
- 63 Lowry, O. H., Passonneau, J. V. & Rock, M. K. The Stability of Pyridine Nucleotides. *Journal of Biological Chemistry* **236**, 2756–2759 (1961).

5.1 | Introduction

Nonulosonic acids (NulOs) are a diverse family of nine-carbon α -keto acids. The first identified member, neuraminic acid, commonly known as sialic acid, plays key roles in eukaryotic systems, mediating processes such as cell–cell recognition, signal transduction, and immune modulation¹. In 1984, the first prokaryote-specific NulO, pseudaminic acid, was discovered². Along with legionaminic acid, it represents one of the most prevalent NulO types in bacteria, although additional variants and epimers have since been identified, including acinetaminic acid, 8-epi-acinetaminic acid and fusaminic acid^{3,4}.

Many pathogenic bacteria are decorated with NulOs, such as *Legionella pneumophila*⁵, *Tannerella forsythia*⁶, *Pseudomonas aeruginosa*⁷, *Helicobacter pylori*⁸, *Acinetobacter baumannii*³, and *Campylobacter jejuni*^{9,10}. These NulOs frequently modify key surface glycans and proteins, including lipopolysaccharides (LPS), capsular polysaccharides (CPS), pili, and flagella, where they play critical roles in bacterial virulence. NulOs have been shown to enhance bacterial colonization of host tissues, such as the lung epithelium by *Pseudomonas aeruginosa*⁷. In *Campylobacter jejuni*, a major cause of gastroenteritis worldwide, NulOs are essential for proper flagellar assembly, thereby conferring motility¹¹. Moreover, these NulOs also serve as recognition points for flagellotropic bacteriophages, where NulO variety may affect the susceptibility to these phages^{12,13}. Variations in flagellin glycosylation among strains may further contribute to antigenic diversity¹⁴. Since NulOs are molecular mimics of sialic acids, they can also facilitate immune evasion by dampening the host immune response^{6,15}. Given the increasing prevalence of antibiotic resistance among pathogens, including WHO-classified critical multidrug-resistant bacteria such as *Acinetobacter baumannii*¹⁶, there is an urgent need for innovative antimicrobial and vaccine strategies.

Since NulOs are typically surface-exposed and unique to prokaryotes¹⁷, they represent attractive targets for antimicrobial and vaccine development^{18,19}. The growing recognition of the potential of carbohydrate-based vaccines further supports this approach²⁰. Initial studies demonstrated promising results for Pseudaminic acid based vaccine approaches against antibiotic resistant *A. baumannii*, including both a Pseudaminic acid containing exopolysaccharide glycoconjugate vaccine²¹ and the direct conjugation of Pseudaminic acid to a carrier protein²².

However, these developments, as well as our understanding of the biofunctions of prokaryote specific NulOs are hindered by the limited access to NulO-containing glycans and glycoconjugates^{19,23}. Their chemical synthesis remains challenging due

to variable stereochemistry, a limited substrate scope and various N-modifications¹⁸. The total synthesis often requires tens of steps²³, leading to low yields and restrained scalability. While more efficient chemical strategies are being developed, these issues have not yet been entirely resolved¹⁸.

An alternative approach is biosynthesis, which exploits the inherent regio- and stereoselectivity of enzymes²³. The biosynthetic pathways of both Pseudaminic acid²⁴ and Legionaminic acid²⁵ have been elucidated. For CMP-Pse5Ac7Ac, a one pot, multi-enzyme synthesis has been established²⁴ and the overexpression of Leg5Ac7Ac has been achieved in *E. coli*, which can subsequently be converted to its CMP-activated form²⁶. Nevertheless, many NulO biosynthetic pathways have yet to be discovered¹ and the enzymatic synthesis of NulO-containing glycoconjugates remains hindered by the lack of characterized NulO transferases.

To date, only a few bacterial NulO transferases have been confirmed experimentally, including the pseudaminyltransferase KpsS1 in *A. baumannii*, which catalyzes the formation of the Pse5Ac7Ac- α (2,6)-GlcP linkage²⁷. Several other candidates have been described in literature, including the putative pseudaminyltransferase FlmG in *Caulobacter crescentus*²⁸, Tanf_01245 in *Tannerella forsythia*²⁹, maf in *Magnetospirillum magneticum*³⁰ and *Clostridium botulinum*^{31,32}, maf1 in *Aeromonas caviae*³³, maf2 (pseD) in *C. jejuni*¹¹, maf5 (pseE) in *C. jejuni*^{13,34,35} and pseE in *H. pylori*³⁶. Moreover, maf4 in *C. jejuni*³⁷ was identified as a putative Legionaminic acid transferase. However, direct biochemical evidence confirming these functions is still lacking. The absence of activity-based validation may be due to the limited availability of CMP-NulO substrates. Since these donors are not commercially available, they must be synthesized in-house, either chemically or enzymatically.

Metabolite extracts from microorganisms offer a promising source for a variety of activated-NulO derivatives, as identified recently³⁸. In this study, we demonstrate with small scale screening experiments that a crude metabolite extract of *C. jejuni* can serve as substrate for KpsS1 (a pseudaminyltransferase from *A. baumannii*) to produce a variety of Pse PNP-Glc conjugates. Interestingly, we show that KpsS1 has a considerable substrate donor promiscuity producing NulO-conjugates beyond the classical NAc modifications. This removes a major bottleneck in producing NulO-glycoconjugates which enable functional studies and the development of novel antimicrobial strategies.

5.2 | Materials and Methods

Plasmids. pET28a(+) plasmids encoding either the CMP-pseudaminic acid synthase (pseF) from *A. baumannii* or the glycosyltransferase from *A. baumannii*

fused to the Im9 protein (Im9-KpsS1) were obtained from Synbio Technologies (sequences in SI DOC). Both proteins contain an N-terminal His₆-tag.

Recombinant protein expression. Chemically competent *E. coli* GOLD BL21 cells were transformed with 100 ng of pET28(+) plasmid carrying either *A. baumannii* pseF or Im9-KpsS1. After incubation on ice (30 min), cells were heat-shocked at 45 °C for 2 min, chilled on ice (2 min), and recovered in 500 µL LB for 1 h at 37 °C, 500 rpm. Transformed cells were plated on LB agar containing 50 µg/mL kanamycin and incubated overnight. A single colony was used to inoculate 15 mL LB (50 µg/mL kanamycin) and grown overnight at 37 °C, 180 rpm. Two 1 L TB cultures (30 µg/mL kanamycin) in 5L flasks were inoculated with 5 mL preculture and grown at 37 °C, 180 rpm to OD₆₀₀ = 0.8. Protein expression was then induced with 0.1 mM IPTG, and cultures were incubated overnight at 25 °C. Cells were harvested by using a Sorvall centrifuge at 6000xg for 15 min at 20 °C, washed with 50 mM potassium phosphate buffer (pH 7.4), and stored at -80 °C.

Protein purification. Cell pellets (~15 g) were resuspended in 30 ml buffer A (50 mM potassium phosphate buffer pH 7.4, 300 mM NaCl, 20 mM imidazole), containing additionally 0.1 mM MgCl₂, 1 mg/ml lysozyme, a spatula tip of DNase and three cOMplete mini Protease Inhibitor Cocktail tables (EDTA-free). For *A. baumannii* Im9-KpsS1, buffer A contained an additional 1 mM DTT. The cells were incubated for 20 minutes on a magnetic stir plate at RT. The cells were lysed by one passage using the cell disrupter (Constant Systems) operating at 1.6 kDa. The collected sample was centrifuged for 15 min at 16000 x g and 4 °C, after which the supernatant was collected and filtered using a 0.22 µm filter and kept on ice. Protein purification from the cell-free extract (CFE) was performed by immobilized metal affinity chromatography (IMAC), exploiting the N-terminal His₆-tag of the recombinant protein. A 1 mL Ni Sepharose 6 Fast Flow column (Cytiva) was equilibrated with 10 column volumes (CV) of buffer A using an NGC chromatography system (Bio-Rad). The sample was loaded, washed with 10 CVs of buffer A, and eluted with a linear gradient of 0–100% buffer B (50 mM potassium phosphate, pH 7.4, 300 mM NaCl, 500 mM imidazole) over 40 CVs at a flow rate of 1 mL/min. For *A. baumannii* Im9-KpsS1, 1 mM DTT was included in buffer B. Fractions of 0.5 mL were collected. Fractions containing the protein of interest, as identified by SDS-PAGE, were pooled and concentrated to ~1 mL using a 10 kDa molecular weight cut-off centrifugal filter (Merck) at 3214 rpm and 4 °C. Buffer exchange into 50 mM potassium phosphate (pH 7.4) was performed using a PD-10 desalting column (GE Healthcare) according to the manufacturer's instructions. Protein concentrations were estimated using the Nanodrop at 280 nm.

SDS-PAGE. 14.5 μ L 2x Laemmli buffer and 0.5 μ L 100 mM DTT were added to 15 μ L of sample. The mixture was heated at 95 °C for 5 min and loaded onto a 4–15% Mini-PROTEAN® TGX Stain-Free™ Protein Gel (Bio-Rad). 5 μ L Precision Plus Protein Standards Unstained (Bio-Rad) was included as molecular weight marker. Electrophoresis was performed in 10x TGS buffer at 200 V for approximately 35 min.

***Campylobacter jejuni* nucleotide sugar extract.** Nucleotide sugar extracts of *Campylobacter jejuni* NCTC 11168 (DSM 27585) were prepared as described in van Ede et al. (2026)³⁸. Briefly, *C. jejuni* was cultured on Mueller–Hinton agar supplemented with Oxoid™ *Campylobacter* Growth Supplement (Thermo Fisher Scientific) under microaerophilic conditions generated using Anaerocult™ C (Merck) in a 2.5 L anaerobic jar. Plates were incubated at 37 °C for 48 h. Biomass was collected with sterile cotton swabs into 1× PBS, centrifuged (10 min, 14,000 rpm, 4 °C), and stored at –80 °C until extraction. For metabolite extraction, 100–200 mg of cell pellet was mixed with 500 μ L pre-cooled extraction solvent (–20 °C; 40% acetonitrile, 40% methanol, 20% MS-grade water) and 0.1 g acid-washed glass beads (500 μ m, Sigma-Aldrich). Cells were disrupted by three 1.5 min vortexing cycles with 1 min cooling on ice between cycles. Lysates were centrifuged (15 min, 14,000 rpm, 4 °C; Eppendorf 5424R), and supernatants were collected. Extracts were diluted 1:10 with MS-grade water and purified using HyperSep™ Hypercarb™ SPE 96-well plates (Thermo Scientific, Germany) as described below. Eluates were divided into three aliquots, dried in a SpeedVac concentrator at room temperature, and stored at –80 °C. Dried samples were reconstituted in 10 μ L MS-grade water for use in the KpsS1 activity assay.

Activity assays. CMP-Pseudaminic acid synthase (pseF) assay. 50 μ L reaction mixtures were prepared in 0.6 ml Eppendorf tubes, containing 50 mM sodium phosphate pH 7.5, 0.33 mg/ml pse5Ac7Ac, 5 mM CTP, 5 mM MgCl₂ and 10 μ L purified pseF. The alternative substrates 0.33 mg/ml Leg5Ac7Ac and 0.33 mg/ml Neu5Ac were evaluated as well. Negative controls, lacking either the enzyme, the substrate (PseAc2) or the acceptor (CTP) were included. Reactions were incubated at 37 °C and 300 rpm, and samples were collected after 0, 1, and 13 h. Samples were purified by solid-phase extraction using a HyperSep PGC 96-well plate (Hypercarb 25/1 mL, Thermo Scientific). Columns were washed with 600 μ L 10 mM ammonium bicarbonate in 60% acetonitrile, equilibrated twice with 600 μ L MS-grade water, and loaded with the samples. After washing with 1 mL MS-grade water, analytes were eluted twice with 200 μ L 10 mM ammonium bicarbonate in 60%

acetonitrile. Eluates were dried in a SpeedVac at room temperature and reconstituted in 15 μ L MS-grade water. 1 μ L injections were made on the LC-MS.

Glycosyltransferase (Im9-KpsS1) assay. 50 μ L reaction mixtures were prepared in 0.6 ml Eppendorf tubes, containing 50 mM sodium phosphate pH 7.5, 10 mM $MgCl_2$, 25 mM NaCl, 10 μ L CMP-PseAc₂ (produced as described above) or 10 μ L *C. jejuni* 11168 metabolite extract and 10 μ L purified Im9-KpsS enzyme or 10 μ L CFE containing Im9-KpsS1. The added acceptor was either 6 mM PNP-Glc or 6 mM PNP-Gal. Moreover, instead of CMP-PseAc₂, 0.2 mg/ml CMP-Neu5Ac was also tested as a substrate. Controls lacking either the enzyme or the substrate (CMP-Pse5Ac7Ac) were included. Reaction mixtures were incubated at 37 °C (300 rpm), and samples were taken after 1 and 18 hours. The samples were cleaned by a solid phase extraction using Oasis HLB 96-well μ Elution plates (2 mg sorbent, 30 μ m; Waters, cat. 186001828BA). In short, the columns were conditioned using 750 μ L methanol, equilibrated with 2 $\times$ 500 μ L MS-grade water, after which the sample was loaded. The columns were washed with 350 μ L of 5% methanol in water, and elution was performed with 200 μ L of 2% formic acid in 80% methanol and 200 μ L of 1 mM ammonium bicarbonate in 80% methanol. Eluates were dried in a SpeedVac concentrator (45 °C) and reconstituted in 15 μ L MS-grade H₂O prior to LC-MS/MS analysis.

LC-MS/MS analysis of reaction products. Reaction products were analyzed using a Q Exactive™ Focus Hybrid Quadrupole-Orbitrap™ mass spectrometer (Thermo Scientific) coupled to an Acquity UPLC system (Waters). For the CMP-Pseudaminic acid synthase (pseF) reaction, chromatographic separation was achieved on a 100 $\times$ 1 mm Hypercarb column (Thermo Scientific) using 50 mM ammonium bicarbonate in MS-grade water (solvent A) and 100% acetonitrile (solvent B). The gradient started at 2.5% B for 2 min, increased linearly to 35% B over 18 min, and then to 60% B over 6 min, at a constant flow rate of 0.1 mL/min. For the glycosyltransferase (KpsS1) reaction, chromatographic separation was performed on a C18 BEH column (1.7 μ m particle size) using 0.1% formic acid in MS-grade water (solvent A) and 0.1% formic acid in acetonitrile (solvent B). The gradient started at 9% B for 2.5 min, increased linearly to 35% B over 10 min, and then to 65% B over an additional 2.5 min, at a constant flow rate of 0.05 mL/min. For both analyses, Full MS (MS1) spectra were acquired at a resolution of 35,000 over a scan range of 250–850 m/z, with an AGC target of 2×10^5 and a maximum injection time of 50 ms. Targeted MS2 scans were performed using a 2.0 m/z isolation window and a normalized collision energy (NCE) of 26. For the pseF reaction, ions at m/z 333.1303, 481.9772, 638.1716, 308.0987, and 613.1399 were targeted, while for

the KpsS1 reaction, ions at m/z 640.1862, 302.0870, and 618.2141 were targeted. When the metabolite extract *C. jejuni* served as the donor substrate source the additional ions at m/z 617.2301, 631.2457, 691.2669 and 692.2509 were targeted. MS2 spectra were acquired at a resolution of 17,500 with an identical AGC and maximum injection time. Data were processed in Xcalibur (Thermo Scientific). Extracted ion chromatograms were generated for relevant ions with a ± 10 ppm error marge.

Shotgun proteomic analysis. Protein extraction, proteolytic digestion and shotgun proteomic analysis has been performed as described by van Ede et al., 2026³⁸. Briefly, about 25 mg cell pellet was resuspended in 175 μ L 50 mM TEAB containing 1% (w/w) NaDOC, adjusted to pH 8 and 175 μ L Bacterial Protein Extraction Reagent (B-PER; Thermo Scientific, cat. 78243). After adding 0.15 g acid-washed glass beads (150–212 μ m; Sigma-Aldrich), samples underwent 3 cycles of bead beating (1.5 min) using a bench top vortex, with short breaks on ice (1 min) in between. The samples were incubated at 80°C for 3 min at 1000 rpm and placed in an ultrasonic bath for 10 min. After centrifuging for 10 min. at 14000 rcf at 4°C the supernatant was collected in a clean LoBind Eppendorf tube. To precipitate the proteins from the supernatant TCA was added in a 1:4 ratio, followed by vortexing and a 30 min incubation at 4 °C. Precipitates were collected by centrifugation (14,000xg, 15 min, 4 °C), washed with ice-cold acetone, and resuspended in 50 μ L of 6 M urea. Samples were incubated at 37 °C (1000 rpm) until fully solubilized. Reduction was performed by the addition of 10 mM dithiothreitol (DTT) in 200 mM ammonium bicarbonate (ABC), reaching a final concentration of 2.3 mM, followed by incubation for 1 h at 37 °C, 300 rpm. Alkylation was carried out by adding 20 mM iodoacetamide in 200 mM ABC to a final concentration of 3.75 mM and incubating for 30 min at room temperature in the dark. The urea concentration was reduced to <1 M by diluting in 200 mM ABC. A 100 μ L aliquot was digested overnight at 37 °C (300 rpm) with 5 μ L of trypsin (0.1 μ g/ μ L; Promega, cat. V5111). The next day, peptides were purified using Oasis HLB 96-well μ Elution plates (2 mg sorbent, 30 μ m; Waters, cat. 186001828BA). Columns were conditioned with 750 μ L methanol, equilibrated with 2 $\times$ 500 μ L MS-grade water, and loaded with the digested samples. After washing with 350 μ L of 5% methanol in water, peptides were eluted sequentially with 200 μ L of 2% formic acid in 80% methanol and 200 μ L of 1 mM ammonium bicarbonate in 80% methanol. Eluates were dried in a SpeedVac concentrator (45 °C) and reconstituted in 3% acetonitrile containing 0.1% formic acid to a final concentration of approximately 500 ng/ μ L, as estimated by UV absorbance at 280 nm (NanoDrop).

For shotgun proteomic analysis, 1 μ L of digest was injected into an EASY-nLC 1200 system coupled to a Q Exactive Plus Orbitrap mass spectrometer (Thermo Scientific, Germany). Chromatographic separation was performed on a 0.05 $\times$ 150 mm C18 analytical column (Thermo Scientific, catalog no. 164943) with 0.1% formic acid, 1% acetonitrile in MS-grade water (mobile phase A) and 0.1% formic acid, 80% acetonitrile in MS-grade water (mobile phase B). Applying a constant flow rate of 300 nL/min, the LC gradient started with a 5 min hold at 5% mobile phase B, followed by a linear gradient to 30% B over 90 min and a subsequent linear gradient to 75% B over an additional 25 min. Full MS scans were acquired in positive ion mode at 70,000 resolution, with an automatic gain control (AGC) target of 3×10^6 and a maximum injection time of 100 ms. Data-dependent acquisition (DDA) was performed on the top 10 precursor ions within a 550–1500 m/z range, using an isolation window of 1.7 m/z. Fragmentation was achieved with a normalized collision energy of 28%. MS/MS spectra were acquired at 17,500 resolution with an AGC target of 5×10^5 and a 100 ms maximum injection time.

Protein identification was performed using PEAKS Studio X (Bioinformatics Solutions Inc., Canada). Database searches were conducted against the *E. coli* (UP000002032) UniProt proteome, supplemented with the *A. baumannii* pseF sequence and the Im9-KpsS1 sequence. Carbamidomethylation was set as a fixed modification, while oxidation and deamidation were included as variable modifications. Trypsin was selected as the digestion enzyme, permitting up to three missed cleavage sites. Mass tolerances were set to 20 ppm for precursor ions and 0.02 Da for fragments. Identifications were filtered to a 1% false discovery rate (FDR), and proteins supported by at least two unique peptides were considered confidently identified.

DMB-labelling of a *C. jejuni* 11168 metabolite extract. LC-MS/MS data of the 1,2-diamino-4,5-methylenedioxybenzene (DMB) labelled *C. jejuni* 11168 metabolite extract was acquired from van Ede et al., 2026 (BioRxiv)³⁸. In short, the *C. jejuni* 11168 nucleotide sugar extract was hydrolyzed in 200 μ L 2 M acetic acid at 80 °C for 2 hours at 300 rpm using an Eppendorf heating block and dried using a vacuum concentrator at 60 °C. Subsequently, 25 μ L DMB-solution (1.4 M acetic acid, 0.75 M 2-mercaptoethanol, 18 mM sodium dithionite, 7 mM DMB) was added and the sample was incubated at 50 °C, 600 rpm for 2.5 h. LC-MS/MS analysis was performed on a Waters Acquity UPLC system equipped with a C18 BEH column (1.7 μ m particle size), which was connected to a Q Exactive Focus Orbitrap mass spectrometer (Thermo Scientific). LC separations were performed using a linear gradient from 9% B (0.1% formic acid in acetonitrile) to 35% B over 9.5 min, followed

by a linear gradient to 65% B over 2.5 min, using a constant flow rate of 0.05 mL/min. Mobile phase A consisted of 0.1% formic acid in H₂O. The MS instrument was operated in positive electrospray ionization mode. Full MS scans were acquired with a resolution of 35,000 and the AGC set to 2E5 and a maximum injection time of 50 ms. Fragmentation spectra were acquired by isolating 5.5 Da mass windows over a *m/z* range of 340–530. The resolution was set to 17,500 with a AGC of 1E6 and a maximum injection time of 150 ms. Raw data were analyzed in Xcalibur 4.1 (Thermo Fisher Scientific, Germany).

5.3 | Results

5.3.1. Production of recombinant *A. baumannii* KpsS1 and CMP-Pse5Ac7Ac.

CMP-activated nonulosonic acids (NulOs), and many non-activated forms, are not commercially available, limiting functional studies and antimicrobial development^{19,23}. To enable in-house production of the activated donor, the CMP-Pse5Ac7Ac synthetase PseF from *Acinetobacter baumannii* (molecular weight ~27 kDa) was recombinantly expressed in *Escherichia coli*. The enzyme carried an N-terminal His-tag, allowing purification by nickel-affinity chromatography. Enzymatic conversion of CTP and Pse5Ac7Ac into CMP-Pse5Ac7Ac was confirmed by an LC–MS/MS–based activity assay (Figure 5.1; SI DOC). To assess substrate specificity, Leg5Ac7Ac and Neu5Ac were also tested as potential substrates. However, no conversion was detected (data not shown), which is consistent with their differing stereochemistry compared to Pse5Ac7Ac.

To date, the only pseudaminyltransferase with confirmed activity is KpsS1 from *Acinetobacter baumannii*²⁷. The His-tagged fusion protein Im9-KpsS1 (~69 kDa) was recombinantly expressed in *E. coli* as previously described by Walklett *et al.* (2024). After nickel-affinity chromatography, the SDS-PAGE analysis combined with proteomics showed the presence of low abundant KpsS1 (Figure 5.1; SI DOC). With CMP-Pse5Ac7Ac as the substrate and PNP-Glc as the acceptor, pseudaminyltransferase activity of KpsS1 was confirmed by LC-MS/MS using both a cell-free extract and the purified protein fraction (Figure 5.1; SI DOC).

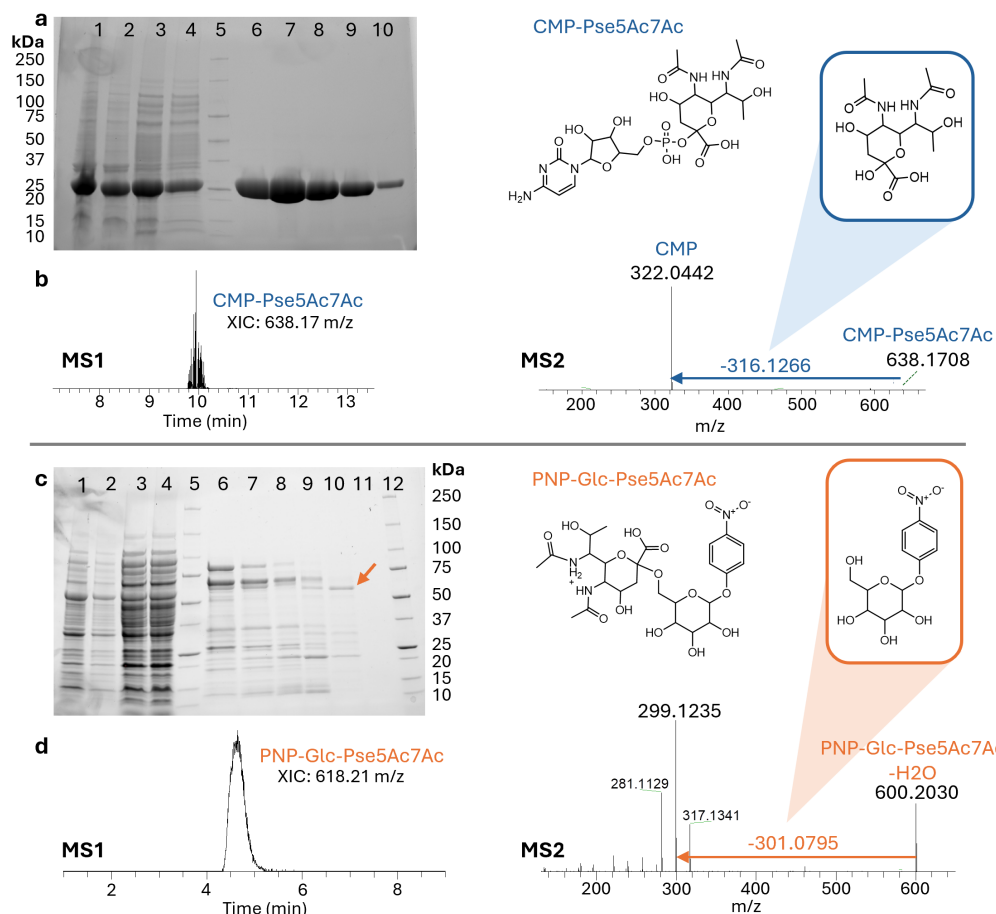


Figure 5.1. Recombinantly expressed *A. baumannii* KpsS1 shows pseudaminyltransferase activity. **a)** Purification of the *A. baumannii* CMP-Pse5Ac7Ac synthetase *pseF*. SDS-PAGE analysis of the N-tagged *pseF* purification by Nickel-affinity chromatography. **1:** Biomass pellet after cell lysis; **2:** 2x diluted biomass pellet after cell lysis; **3:** Cell free extract; **4:** Flow through; **5:** Protein ladder; **6-10:** His-rich protein fractions. A prominent band appeared just above 25 kDa, consistent with the expected molecular weight of the recombinantly expressed *A. baumannii* *PseF* protein carrying an N-terminal His-tag (27 kDa). **b)** CMP-Pseudaminic acid synthase (*A. baumannii* *pseF*) activity assay (13-hour incubation). Left panel: Extracted ion chromatograms of $m/z = 638.1716$ (representing CMP-Pse5Ac7Ac), using a 10 ppm error margin. Right panel: Fragmentation spectrum of $m/z 638.1716$. The chemical structures of the reaction product CMP-Pse5Ac7Ac and the neutral loss during fragmentation (Pse5Ac7Ac) are provided. **c)** *A. baumannii* KpsS1 purification. SDS-PAGE analysis of the N-tagged *Im9-KpsS1* purification by Nickel-affinity chromatography. **1:** Biomass pellet after cell lysis; **2:** 2x diluted biomass pellet after cell lysis; **3:** Cell free extract; **4:** Flow through; **5:** Protein ladder; **6-11:** His-rich protein fractions (A11, A14, A17, A22, A27 and A71, respectively). See SI DOC for purification chromatogram; **12:** Protein ladder. **d)** Pseudaminyltransferase (*A. baumannii* KpsS1) activity assay. Left panel: Extracted ion chromatogram of $m/z 618.21$ (representing PNP-Glc-Pse5Ac7A), using a 10 ppm error margin. Right panel: Fragmentation spectrum of $m/z 618.21$. The chemical structures of the KpsS1 reaction product PNP-Glc-Pse5Ac7Ac and the neutral loss during fragmentation (PNP-Glc) are provided.

5.3.2. Confirming suitability of crude cell lysates as CMP-NulO substrate donors, and showing broad donor substrate promiscuity in KpsS1.

While Pse5Ac7Ac is commercially available, other nonulosonic acid derivatives (NulOs) are not. Moreover, the CMP-activated forms must still be produced either chemically or enzymatically. To facilitate functional tests of glycomolecules, alternative substrate sources are needed. For example, crude metabolite extracts from microorganisms that naturally produce these sugars would be an ideal source of these sugars. One such organism is *Campylobacter jejuni*, a well-studied pathogen known to extensively glycosylate its flagella with NulOs that contribute to virulence¹¹. In the reference strain 11168, several CMP-activated NulOs have been characterized, including CMP-Leg5Ac7Ac, CMP-Leg5Am7Ac, CMP-Leg5MeAm7Ac, CMP-Pse5Ac7Ac, CMP-Pse5Ac7Am, CMP-Pse5AcGriMe7Ac, CMP-Pse5AcGriMe7Acm, and CMP-Neu5Ac^{10,38} (Figure 5.2; SI DOC). Given that CMP-Pse5Ac7Ac is known to be present in *C. jejuni* 11168, we used a metabolite extract from this organism as the substrate source for the KpsS1 activity assay. Using this extract, KpsS1 activity was confirmed by LC-MS/MS through identification of the reaction product PNP-Glc-Pse5Ac7Ac (Figure 5.2), demonstrating that a small-scale metabolite extract can serve as a substrate source without the need to fractionate individual CMP-sugars.

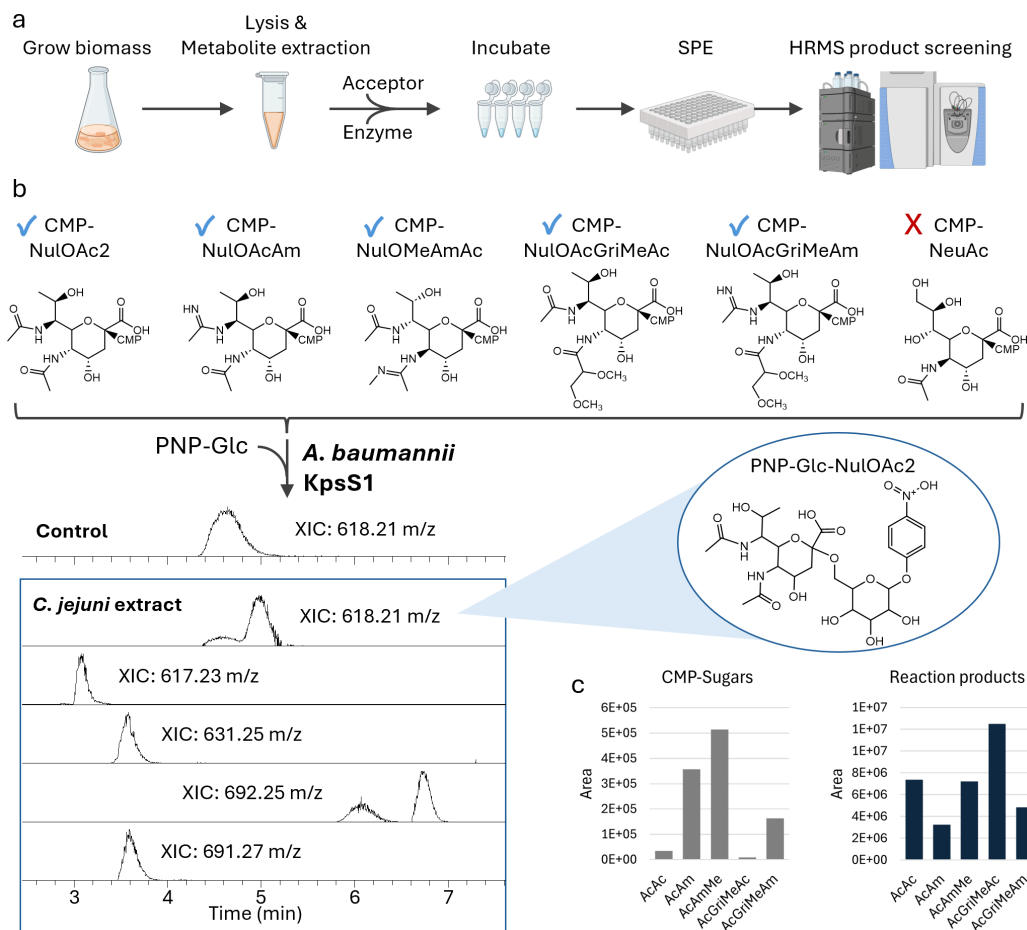


Figure 5.2. Small-scale production of NuLO-conjugates using crude microbial cell extracts as CMP-NuLO donors, and substrate promiscuity of the *A. baumannii* pseudaminyltransferase KpsS1. **a)** Workflow of the activity assay. *C. jejuni* 11168 cells were cultured, after which the cells were lysed and the metabolites were extracted. The acceptor PNP-Glc and the enzyme *A. baumannii* KpsS1 were added to the metabolite extract, and the reaction mixture was incubated overnight at 37°C. The sample was subsequently cleaned by solid phase extraction and analyzed by LC-MS/MS in positive mode. **b)** CMP-NuLO substrate structures and extracted ion chromatograms (MS1 spectra) of the identified corresponding reaction products. The chemical structure of one representative reaction product, PNP-Glc-NuOAc2, is shown. Corresponding fragmentation spectra and the putative chemical structures of the products are provided in the SI DOC. **c)** LC-MS/MS abundance profile of the CMP-activated sugars in the *C. jejuni* 11168 extract prior to the activity assay and of the reaction products formed after the KpsS1 activity assay. The modifications are indicated on the x-axis. For CMP-sugars this corresponds to CMP-NuLO-“modification”, whereas for the reaction products it corresponds to PNP-Glc-NuLO-“modification”.

We next investigated whether other NuLO derivatives present in *C. jejuni* 11168 could serve as donor substrates for the *A. baumannii* KpsS1 transferase. LC-MS/MS analyses revealed products corresponding to all previously identified

NulOs, including NulOAcAc, NulOAcAm, NulOAcGriMeAm, and NulOAcGriMeAc (Figure 5.2; SI DOC), indicating considerable substrate promiscuity. Interestingly, a reaction product with a m/z of 631.2457 (positive mode) was detected. Fragmentation analysis revealed a neutral loss of 301.08 (PNP-Glc), producing a fragment at m/z 330.1646 ($[M-H]^+$), potentially corresponding to NulOMeAmAc (SI DOC). This modification has so far been reported only as a Legionaminic acid (Leg) derivative in *C. jejuni* 11168^{10,35}. This suggests that KpsS1 may also transfer Leg, despite stereochemical differences. However, no conversion was detected when CMP-Neu5Ac, which shares Leg's stereochemistry, was used as a substrate²⁷, rendering this possibility unlikely though not impossible. Alternatively, the observed product may correspond to a Pseudaminic acid derivative bearing the same modification or an isomer with the same molecular formula but a different modification pattern, for example, with the methyl group attached to a hydroxyl group. To explore this possibility, data from van Ede et al., 2026 (BioRxiv)³⁸ was reanalyzed, where the *C. jejuni* 11168 extract was chemically derivatized with 1,2-diamino-4,5-methylenedioxybenzene (DMB) and analyzed by LC-MS/MS. Fragmentation spectra showed no evidence of an O-methyl modification and were consistent with the MeAm modification (Figure 5.3).

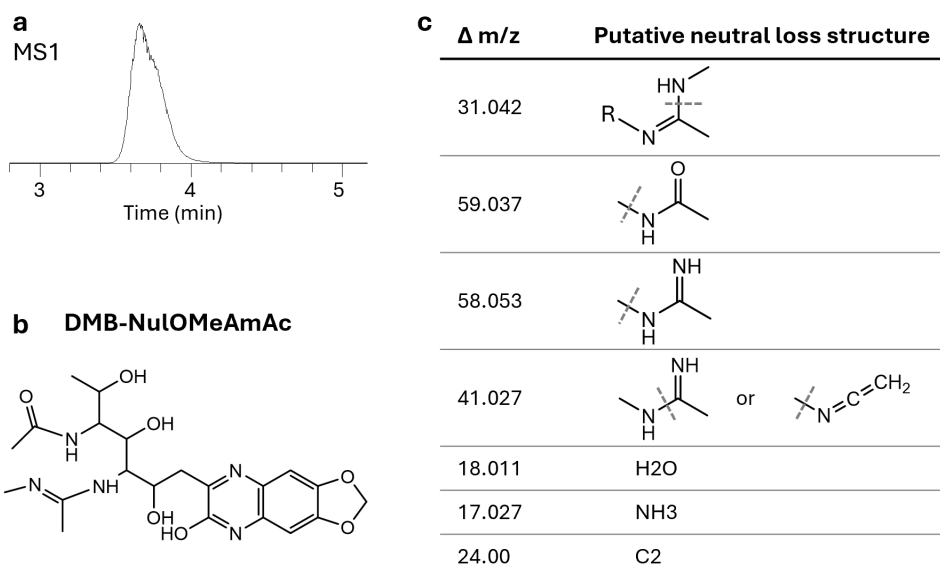


Figure 5.3. Structure elucidation of the KpsS1 reaction product with m/z 631.25. The LC-MS/MS data of the *C. jejuni* 11168 metabolite extract, which was chemically labelled with 1,2-diamino-4,5-methylenedioxybenzene (DMB), was acquired and reanalyzed from van Ede et al., 2026 (BioRxiv)³⁸. The targeted m/z of 464.216 corresponds to the KpsS1 reaction product with a m/z of 631.25. **a)** Extracted ion chromatogram of m/z 464.214, representing DMB-NulOMeAmAc. **b)** Putative structure of DMB-NulOMeAmAc. **c)** In van Ede et al., 2026 (BioRxiv)³⁸ the m/z 464.216, representing DMB-NulOMeAmAc,

was targeted and the main neutral loss fragments are depicted in the table as well as their corresponding putative structures. No O-methyl fragments were observed.

5.4 | Discussion

Antibacterial resistance is becoming a major global health threat, with some bacterial species showing resistance to all available antibiotics^{16,39}. Although the broad activity of antibiotics against multiple bacterial species is often advantageous, it also promotes the wider spread of antimicrobial resistance and can negatively affect the host microbiome. Alternative antimicrobial strategies are therefore urgently needed, with species-specific or narrow-spectrum antibacterial agents representing a promising approach³⁹. Nonulosonic acids containing molecules have emerged as promising targets for the development of novel antimicrobials and vaccines^{21,22}. Nevertheless, progress in these areas, as well as our understanding of the biological roles of nonulosonic acids, is limited by the challenge to produce NuO-containing glycans and glycoconjugates^{19,23}. Furthermore, from the bacterial NuO transferase candidates which have been proposed in the literature^{13,28,30,33,36,37}, to date only one, the *A. baumannii* KpsS1 transferase²⁷, has been confirmed experimentally. This lack of experimental validation can partly be attributed to the limited availability of CMP-activated donor substrates. While the common derivative Pse5Ac7Ac is commercially available, other NuO derivatives are not, and even small quantities remain prohibitively expensive. Moreover, Pse5Ac7Ac must still be converted to its CMP-activated form. Therefore, to facilitate the discovery and validation of novel glycosyltransferases, alternative substrate sources are desirable. In this study, we demonstrated that a microbial metabolite extract can serve as a donor substrate source without requiring further purification of the CMP-sugars.

As a control, the pseudaminyltransferase *A. baumannii* KpsS1 was recombinantly expressed, and its activity was confirmed using commercially available Pse5Ac7Ac. While Walklett and colleagues²⁷ demonstrated pseudaminyltransferase activity using cell-free extracts with PNP-Gal as the acceptor, we observed substantial β -galactosidase activity in our extracts, resulting in hydrolysis of the PNP-Gal acceptor. This was indicated by the development of a yellow coloration due to the release of p-nitrophenol. As this effect was significantly reduced when using PNP-Glc as the acceptor (SI DOC), PNP-Glc was employed in all subsequent assays.

While demonstrating that a crude metabolite extract can serve as a substrate source in small scale reactions, we also revealed considerable donor substrate

promiscuity when using a *Campylobacter jejuni* 11168 metabolite extract with the KpsS1 transferase. Promiscuity has previously been reported for bacterial sialyltransferases, some of which even transfer Legionaminic acid (Leg) derivatives⁴⁰⁻⁴². The reaction products NulOAcAc, NulOAcAm, NulOAcGriMeAm, and NulOAcGriMeAc, originating from the KpsS1 activity assay, are likely Pseudaminic acid (Pse) derivatives, as KpsS1 was originally identified as a pseudaminyltransferase. Moreover, CMP-NulOAcGriMeAm and CMP-NulOAcGriMeAc have previously only been reported as Pse derivatives in the *C. jejuni* 11168 strain¹⁰.

A. baumannii has been extensively studied for its diversity of nonulosonic acids (NulOs), since they are part of its capsular polysaccharide (CPS), which is a key virulence factor⁴³. Identified NulOs include, Pse5Ac7Ac, Pse5Ac7Ac4OAc, Pse5Ac7Hb (Hb: 3-hydroxybutanoyl), 8-epi-Pse5Ac7Ac, Leg5Ac7Ac, Leg5Ac7Ac4OAc, LegAcHb, 8-epi-Leg5Ac7Ac, Aci5Ac7Ac, and 8-epi-Aci5Ac7Ac^{3,44-46}. The *A. baumannii* strain ACICU, from which the pseudaminyltransferase KpsS1 was confirmed, is only known to contain Pse5Ac7Ac²⁷. Possibly, the enzyme KpsS1 may operate without strict substrate specificity, as no other NulOs have been detected in this strain and thus no pathway partitioning would be required.

Our results indicate that KpsS1 tolerates bulky C5 modifications, such as dimethyl glyceric groups. In contrast, the KL42 gene cluster of *A. baumannii*, encodes a predicted Pse5Ac7AcRHb-specific glycosyltransferase^{47,48}, which would suggest that there are also dedicated transferases, although this has yet to be experimentally verified.

The identified reaction product with m/z 631.2457 likely corresponds to NulOMeAmAc. To date, this modification has only been identified as a Leg derivative in *C. jejuni* 11168^{10,35}. This observation suggests two possibilities. First, KpsS1 may also transfer Leg derivatives despite their distinct stereochemistry, as some sialyltransferases have been reported to transfer both Neu and Pse derivatives⁴⁹. However, no transferase activity was observed when CMP-Neu5Ac, which shares the same stereochemistry as Leg, was used as a substrate for KpsS1, making this explanation unlikely though not impossible. Interestingly, the abundance profiles of the CMP-sugars present before the activity assay and the reaction products formed afterward do not align. While the CMP-NulOs were dominated by the AcAm and MeAmAc derivatives, the reaction products were most abundant for the larger glyceric derivative, AcGriMeAc. If the MeAmAc modification indeed corresponds to a Leg derivative, this observation may indicate that KpsS1 transfers Leg inefficiently. Alternatively, the reaction product (m/z 631) may originate from a Pse derivative with the same modification, or from a compound

with the same molecular formula but a different modification pattern. To further investigate this, data from van Ede et al., 2026 (BioRxiv)³⁸ was reanalyzed, where the *C. jejuni* extract was subjected to DMB labeling. The extracted ion chromatogram from the corresponding m/z 464.21 allows one to speculate that two closely eluting species are present, indicated by a small peak shoulder (Figure 5.3), which could reflect an additional NulO besides Leg5MeAm7Ac. Nonetheless, improved chromatographic separation would be required for confirmation. No evidence of O-Me losses was observed in the fragmentation spectra of the DMB-labeled variant, suggesting that the detected NulO is unlikely to differ in its methylation pattern. This could indicate the presence of a Pse derivative bearing the same modification. Nevertheless, MS analysis alone does not allow unambiguous assignment of the modification positions. While the MeAm group has previously been reported at the C5 position in Leg, it could be located at C7 in a potential Pse derivative. In summary, the data suggests that KpsS1 either enables the transfer of Leg derivatives or the presence of a Pse variant with a similar modification in the *C. jejuni* extract.

Interestingly, two product peaks were observed for PNP-Glc-NulOAc2 and PNP-Glc-NulOAcGriMeAc. Although the fragmentation profiles differed, both showed the loss of PNP-Glc (Figure 5.2; SI DOC). Differences were noted in the presence of the parent ion and in the relative abundance of fragment peaks. The initial hypothesis, that the two peaks represented both Pse5Ac7Ac and Leg5AcAc, was excluded, as previous DMB-labeling of Pse and Leg standards showed that Leg elutes before Pse³⁸. Moreover, it remains unclear whether KpsS1 can transfer Leg. While both CMP-Pse5Ac7Ac and CMP-Leg5Ac7Ac have been identified in *C. jejuni* 11168, the larger dimethyl glyceric derivatives have so far only been reported as Pse derivatives¹⁰. Another possibility is that the second peak arises from in-source fragmentation, although no clear evidence has been identified yet.

Finally, a metabolite extract of *C. jejuni* 11168 was used as a substrate source, as its NulOs are well characterized¹⁰. It should be noted, however, that *C. jejuni* is a pathogenic organism. Therefore, for the large scale production other environmental bacteria could be used, as identified recently⁵⁰. A promising alternative substrate source is *Ca. Accumulibacter phosphatis*, which has been shown to also produce CMP-Pse5Ac7Ac³⁸, but no other N or O-modified derivatives.

5.5 | Conclusion

We established a workflow that uses crude microbial metabolite extracts from organisms such as *Campylobacter jejuni*, as CMP-activated nonulosonic acids (NulOs) source for transferases. This approach eliminates the need for chemical or

enzymatic synthesis of CMP-sugars, which are often commercially not available or cannot be produced chemically.

Interestingly, we show that the KpsS1 pseudaminyltransferase shows activity toward all Pse derivatives naturally present in *C. jejuni*, indicating large substrate promiscuity. LC–MS/MS analyses further suggest that KpsS1 may transfer a methylamidated derivative previously associated only with Legionaminic acid, implying that this modification might not be restricted to Leg molecules.

Using microbial extracts as substrate sources removes a major bottleneck in producing bacterial NulO-glycomolecules at small and large scales. This enables studies regarding their biological function, as well as development of novel antimicrobials and vaccines.

5.6 | Supplementary information

5.6.1. Recombinant expression and purification of *Acinetobacter baumannii* pseF.

pseF sequence (*Acinetobacter baumannii*)⁵¹:

MQLAVIPARGGSKRIPRKNIKDFYGKPMIAWSIEAALKSECFDEVVWSTDDQEIADIAVKFGAKVPF
LRPAHLSDDYATTADV MQHATQAFALQGQSPDFVCCLYATAPFVQVDDLKTGLQLLQSHPLDY
AFSATTFAFPIQRAIKLNEDGNVEMFQPENFNKRSQDFEEAWHDAGQFYWGTAKEWLEKSVIFS
NRSRIVELPRTRVQDIDTPEDWHRAELMAQLLLD

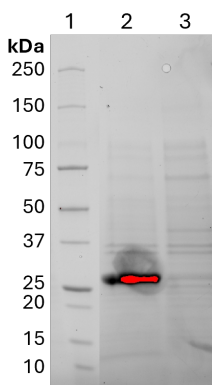


Figure S5.1. SDS-PAGE analysis of *A. baumannii* pseF protein expression in *E. coli* GOLD BL21. The His-tagged pseF protein has a molecular weight of ~27 kDa. **1:** Protein ladder. **2:** Post-induction. **3:** Pre-induction.

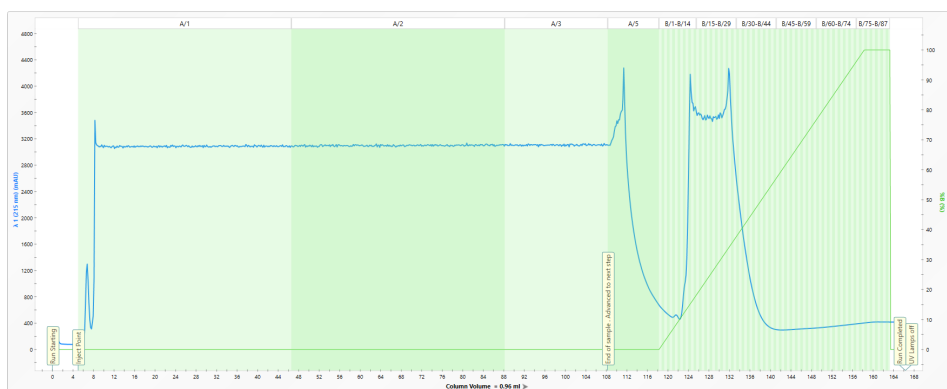


Figure S5.2. Purification chromatogram of the *E. coli* GOLD BL21 cell free extract containing His₆-tagged *A. baumannii* pseF on a Ni-column.

5.6.2. Recombinant expression, purification and activity testing of *Acinetobacter baumannii* KpsS1.

Im9-KpsS1 sequence (*Acinetobacter baumannii*)²⁷:

MGSSHHHHHHSSMELKHSISDYTEAEFLQLVTTICNADTSSEEELVKLVTHFEEMTEHPSGSDLIY
YPKEGDDDDSPSGIVNTVKQWRAANGKSGFKQGLEVLFGQPAMNFLILINSAPNYKYFFYELAKEIE
SRGHNIYFAIDSHRSKYLEPLPELDNNQNSFFFDSYLEKNFDKNLSVSHNNNQEYWGDYFYSDY
DRFLTHDFNLNKDKNYWLNVKVSLDSFFEDIKDKQIDFVLYENISNSFAYAAYLQCTKLGGKYIGL
MGSRLPNHFEIQNSIVEEELKKLEILAQKPITQDEMEWFENYKKSIVDIQPDYMKQNGLDNVAISRI
VKLNKFLKALRLLTIGFKYKHYYDYQFGNPFMVPIKAIRVNIKRYLNTKKSQKFYINNDELEICSSKER
FYIYPIHFHPESSTVLAPEYTNEYSNIINIANNLPFGTYLYVKDHKSAKGVQSYEFYKKVSSLPNVRL
VNFDVNIKRLILKSLGVITVNSTAGYEALLGKPVYLLGRVYENFNNVYNLKSFRDIRDIRDILDFQF
LDVKKDFIAYKKYVYKGVIFIDYGNRVNDKKRYFSELVDSIFLNINTDIE

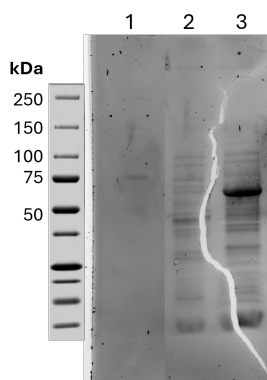


Figure S5.3. SDS-PAGE analysis of *A. baumannii* Im9-KpsS1 protein expression in *E. coli* GOLD BL21. The His-tagged Im9-KpsS1 protein has a molecular weight of ~69 kDa. **1:** Protein ladder. **2:** Pre-induction. **3:** Post-induction. Since the protein ladder was poorly visible, a reference protein ladder has been added.

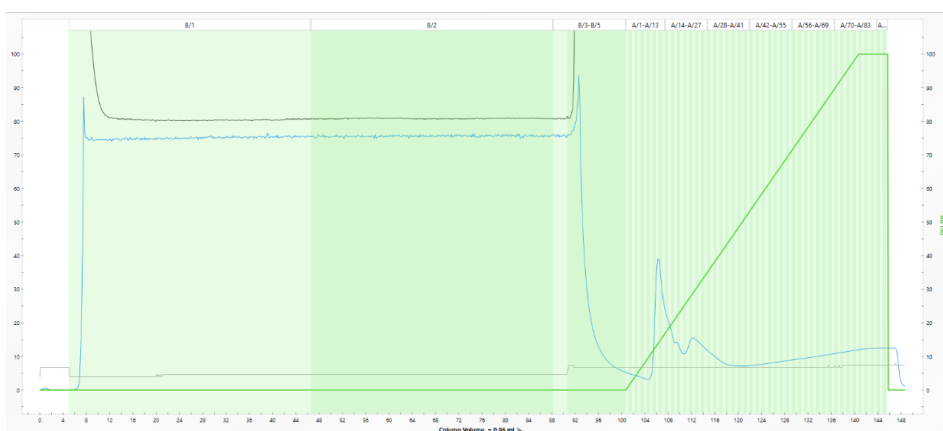


Figure S5.4. Purification chromatogram of the *E. coli* GOLD BL21 cell free extract containing His6-tagged *A. baumannii* Im9-KpsS1 on a Ni-column.

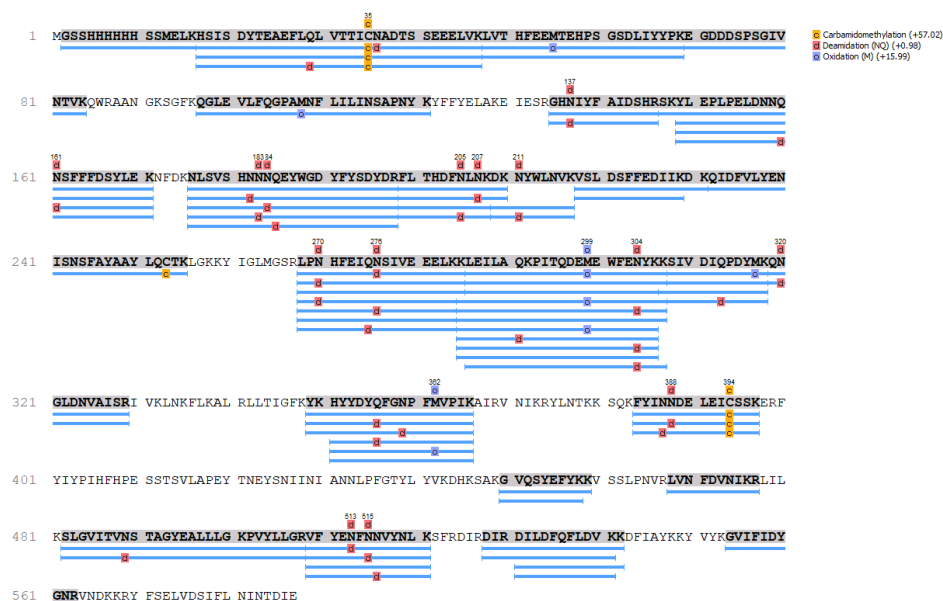


Figure S5.5. Sequence coverage of purified *A. baumannii* Im9-KpsS1 after shotgun proteomics analysis.

5.6.3. Abundance profile of the *A. baumannii* KpsS1 reaction product PNP-Glc-Pse5Ac7Ac.

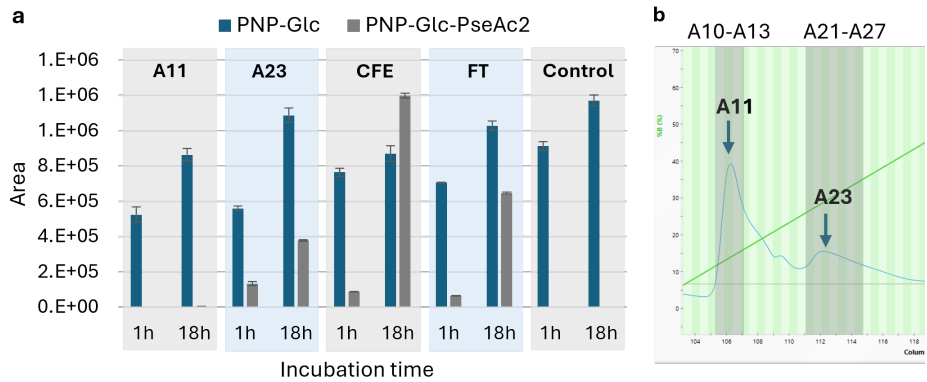


Figure S5.6. a) *Acinetobacter baumannii* KpsS1 activity assay, using CMP-Pse5Ac7Ac as produced with *A. baumannii* pseF as substrate. Reaction mixtures were created with either purified fraction **A11**, purified fraction **A23**, the cell free extract (**CFE**) after cell lysis or the flow through (**FT**) after Ni-column purification, serving as the enzyme (KpsS1) source. The reaction products were analyzed by LC-MS/MS and the peak area of the extracted ion m/z 618.214 was integrated with a 10 ppm error marge. The reaction products were analyzed after 1 or 18 hours of incubation at 37 °C. **b)** Zoomed in on the purification chromatogram of the *E. coli* GOLD BL21 cell free extract containing His6-tagged *A. baumannii* Im9-KpsS1 on a Ni-column. The purified fractions A11 and A23 are highlighted with an arrow. The full purification chromatogram is provided in Figure S5.4.

5.6.4. β -galactosidase activity observed in cell free extract of *E. coli*.

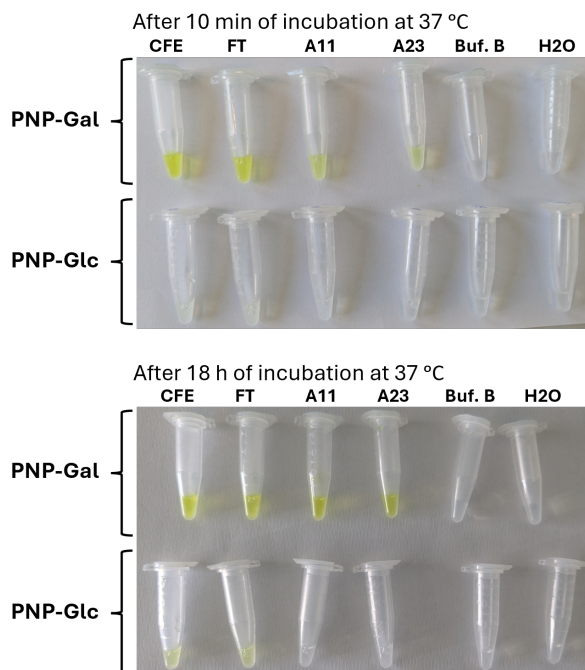


Figure S5.7. Beta-galactosidase activity in *E. coli* cell free extracts. When the acceptor PNP-Gal was added to *E. coli* cell-free extracts containing the *A. baumannii* KpsS1 enzyme, the samples turned yellow within 10 min of incubation at 37 °C, indicating β -galactosidase activity. This enzyme hydrolyzed the acceptor, releasing PNP and causing the yellow coloration. In contrast, samples containing PNP-Glc developed only a faint yellow color, even after 18 h of incubation, and were much less intense than the PNP-Gal samples after just 10 min. Therefore, PNP-Glc was selected as the acceptor for subsequent experiments.

5.6.5. Fragmentation spectra of KpsS1 reaction products.

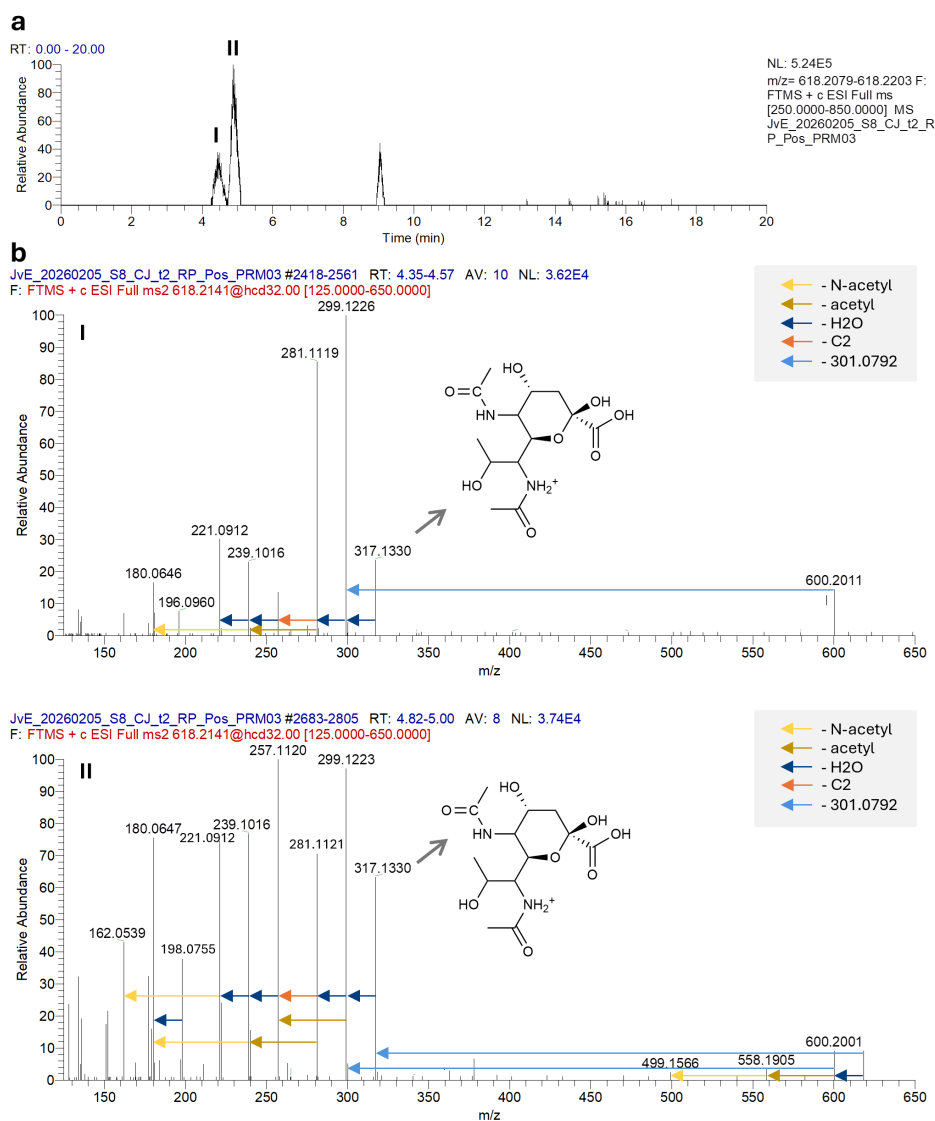


Figure S5.8. *Acinetobacter baumannii* KpsS1 activity assay with a *C. jejuni* metabolite extract serving as substrate source. a) Extracted ion chromatogram of m/z 618.214, representing PNP-Glc-NulOAc2. **b)** Fragmentation spectra of m/z 618.214 for peak I and peak II, respectively, as identified in **a)**. The putative structure of the nonulosonic acid is inserted in the fragmentation spectra based on exact mass and previous reports for NulOs present in *C. jejuni* 11168¹⁰.

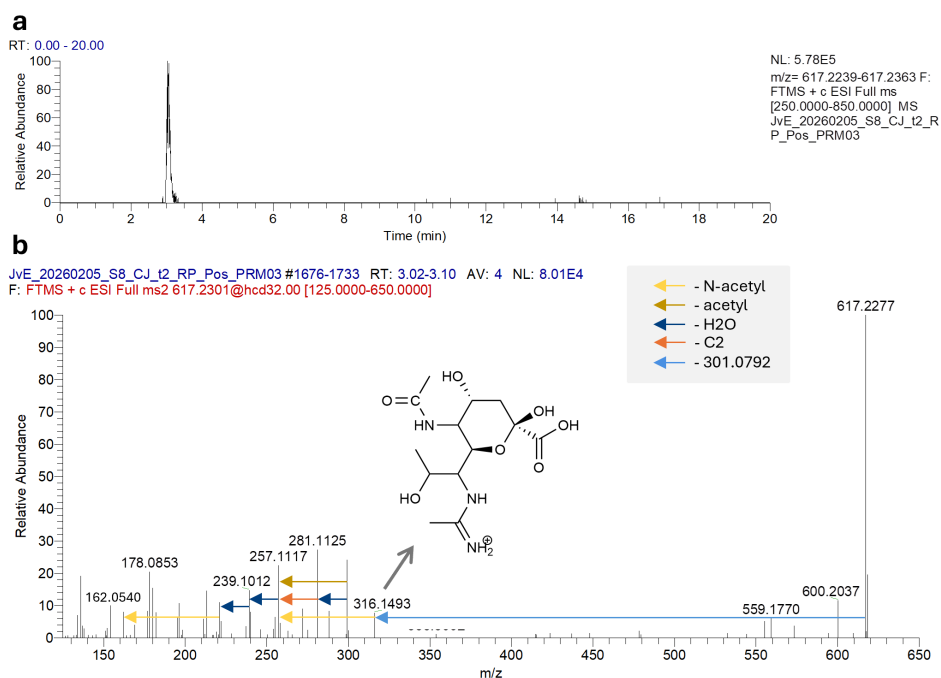


Figure S5.9. *Acinetobacter baumannii* KpsS1 activity assay with a *C. jejuni* metabolite extract serving as substrate source. a) Extracted ion chromatogram of m/z 617.231, representing PNP-Glc-NuLOAcAm. **b)** Fragmentation spectra of m/z 617.231. The putative structure of the nonulosonic acid is inserted in the fragmentation spectrum based on exact mass and previous reports for NuLOs present in *C. jejuni* 11168¹⁰.

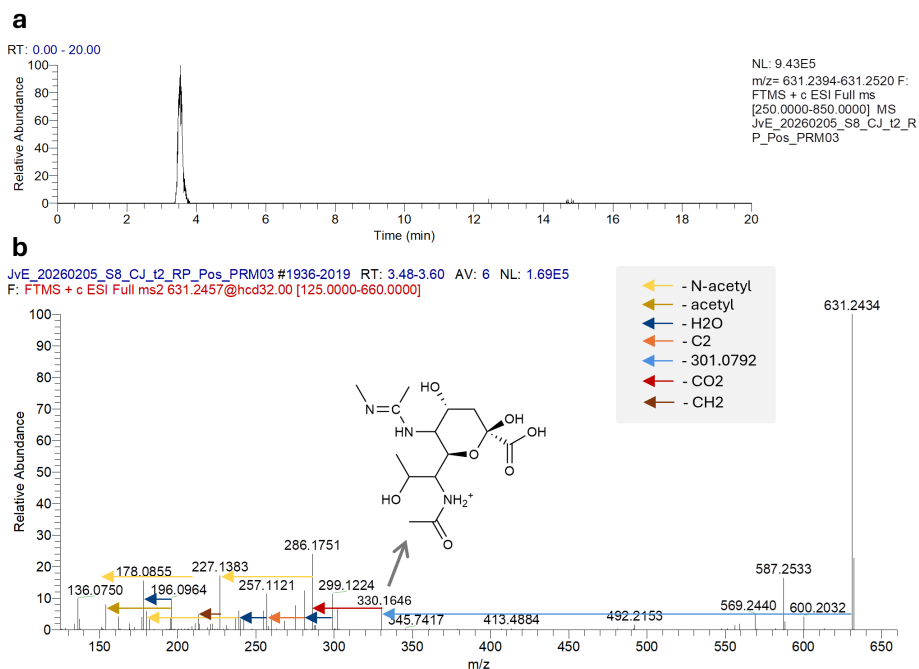


Figure S5.10. *Acinetobacter baumannii* KpsS1 activity assay with a *C. jejuni* metabolite extract serving as substrate source. a) Extracted ion chromatogram of m/z 631.246, representing PNP-Glc-NulOAcAmMe. **b)** Fragmentation spectra of m/z 631.246. The putative structure of the nonulosonic acid is inserted in the fragmentation spectrum based on exact mass and previous reports for NulOs present in *C. jejuni* 11168¹⁰.

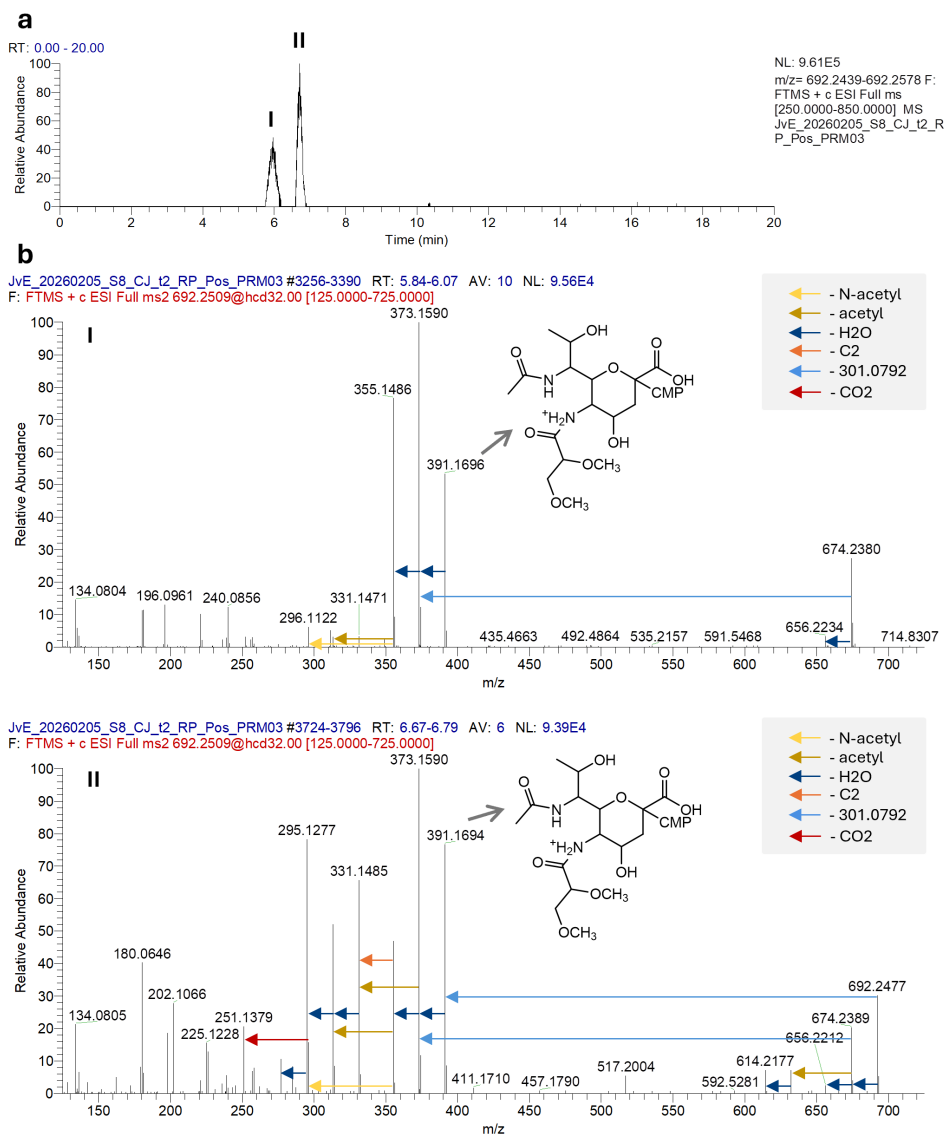


Figure S5.11. *Acinetobacter baumannii* KpsS1 activity assay with a *C. jejuni* metabolite extract serving as substrate source. a) Extracted ion chromatogram of m/z 692.251, representing PNP-Glc-NuLOAcAc-2,3-di-O-methylglycerol. **b)** Fragmentation spectra of m/z 692.251 for peak I and peak II, respectively, as identified in a). The putative structure of the nonulosonic acid is inserted in the fragmentation spectra based on exact mass and previous reports for NuLOs present in *C. jejuni* 11168¹⁰.

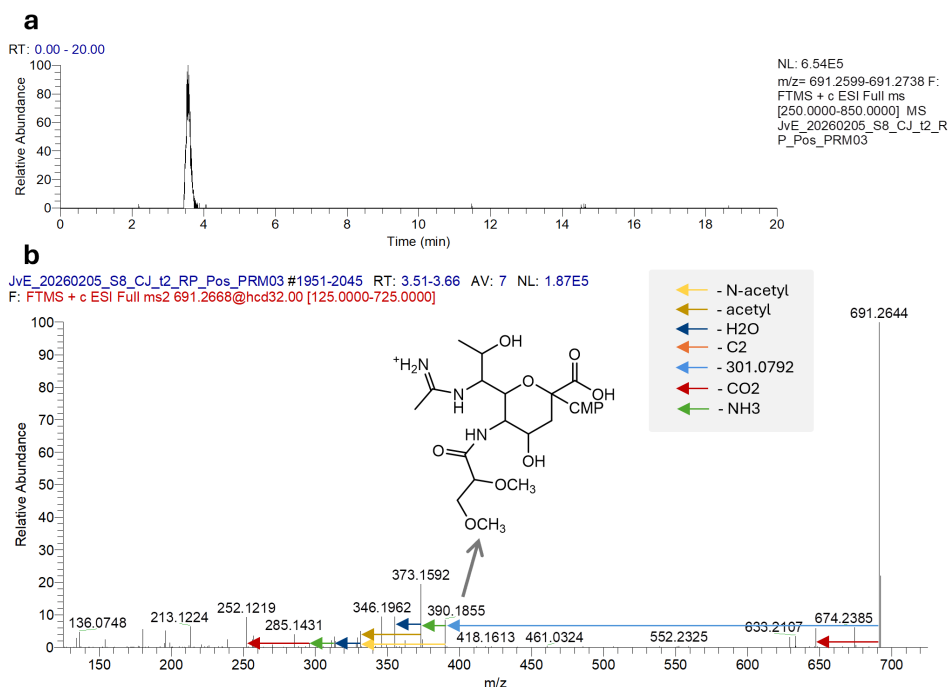


Figure S5.12. *Acinetobacter baumannii* KpsS1 activity assay with a *C. jejuni* metabolite extract serving as substrate source. a) Extracted ion chromatogram of m/z 691.267, representing PNP-Glc-NulOAcAm-2,3-di-O-methylglyceroyl. **b)** Fragmentation spectra of m/z 691.267. The putative structure of the nonulosonic acid is inserted in the fragmentation spectrum based on exact mass and previous reports for NulOs present in *C. jejuni* 11168¹⁰.

References

- McDonald, N. D. & Boyd, E. F. Structural and Biosynthetic Diversity of Nonulosonic Acids (NuLOs) That Decorate Surface Structures in Bacteria. *Trends Microbiol* **29**, 142–157 (2021). <https://doi.org/10.1016/j.tim.2020.08.002>
- Knirel, Y. A. *et al.* Sialic Acids of a New Type from the Lipopolysaccharides of *Pseudomonas-Aeruginosa* and *Shigella-Boydii*. *Carbohydr Res* **133**, C5–C8 (1984). [https://doi.org/10.1016/0008-6215\(84\)85213-1](https://doi.org/10.1016/0008-6215(84)85213-1)
- Kenyon, J. J., Marzaioli, A. M., De Castro, C. & Hall, R. M. 5,7-di-N-acetyl-acinetaminic acid: A novel non-2-ulosonic acid found in the capsule of an *Acinetobacter baumannii* isolate. *Glycobiology* **25**, 644–654 (2015). <https://doi.org/10.1093/glycob/cwv007>
- Vinogradov, E., Michael, F. S. & Cox, A. D. The structure of the LPS O-chain of strain 25586 containing two novel monosaccharides, 2-acetamido-2,6-dideoxy-altrose and a 5-acetimidoylamino-3,5,9-trideoxy-gluconon-2-ulosonic acid. *Carbohydr Res* **440–441**, 10–15 (2017). <https://doi.org/10.1016/j.carres.2017.01.002>
- Knirel, Y. A., Rietschel, E. T., Marre, R. & Zahring, U. The Structure of the O-Specific Chain of *Legionella-Pneumophila* Serogroup-1 Lipopolysaccharide. *Eur J Biochem* **221**, 239–245 (1994). <https://doi.org/10.1111/j.1432-1033.1994.tb18734.x>
- Bloch, S. *et al.* Immune response profiling of primary monocytes and oral keratinocytes to different *Tannerella forsythia* strains and their cell surface mutants. *Mol Oral Microbiol* **33**, 155–167 (2018). <https://doi.org/10.1111/omi.12208>
- Smedley, J. I. *et al.* Influence of Pilin Glycosylation on *Pseudomonas aeruginosa* 1244 Pilus Function. *Infect Immun* **73**, 7922–7931 (2005). <https://doi.org/10.1128/iai.73.12.7922-7931.2005>
- Schirm, M. *et al.* Structural, genetic and functional characterization of the flagellin glycosylation process in. *Mol Microbiol* **48**, 1579–1592 (2003). <https://doi.org/10.1046/j.1365-2958.2003.03527.x>
- Thibault, P. *et al.* Identification of the carbohydrate moieties and glycosylation motifs in flagellin. *J Biol Chem* **276**, 34862–34870 (2001). <https://doi.org/10.1074/jbc.M104529200>
- Logan, S. M. *et al.* Identification of novel carbohydrate modifications on C. jejuni 11168 flagellin using metabolomics-based approaches. *Febs J* **276**, 1014–1023 (2009). <https://doi.org/10.1111/j.1742-4658.2008.06840.x>
- Guerry, P. *et al.* Changes in flagellin glycosylation affect autoagglutination and virulence. *Mol Microbiol* **60**, 299–311 (2006). <https://doi.org/10.1111/j.1365-2958.2006.05100.x>
- Sacher, J. C. *et al.* Binding of Phage-Encoded FlaGrab to Motile *Campylobacter jejuni* Flagella Inhibits Growth, Downregulates Energy Metabolism, and Requires Specific Flagellar Glycans. *Front Microbiol* **11**, 397 (2020). <https://doi.org/10.3389/fmicb.2020.00397>
- Javed, M. A. *et al.* A receptor-binding protein of *Campylobacter jejuni* bacteriophage NCTC 12673 recognizes flagellin glycosylated with acetamidino-modified pseudaminic acid. *Mol Microbiol* **95**, 101–115 (2015). <https://doi.org/10.1111/mpi.12849>
- Guerry, P. *Campylobacter* flagella: not just for motility. *Trends in Microbiology* **15**, 456–461 (2007). <https://doi.org/10.1016/j.tim.2007.09.006>
- Stephenson, H. N. *et al.* Pseudaminic acid on *Campylobacter jejuni* flagella modulates dendritic cell IL-10 expression via Siglec-10 receptor: a novel flagellin-host interaction. *J Infect Dis* **210**, 1487–1498 (2014). <https://doi.org/10.1093/infdis/jiu287>
- WHO bacterial priority pathogens list, 2024: Bacterial pathogens of public health importance to guide research, development and strategies to prevent and control antimicrobial resistance. *Geneva: World Health Organization* (2024).
- Zunk, M. & Kiefel, M. J. The occurrence and biological significance of the α -keto-sugars pseudaminic acid and legionaminic acid within pathogenic bacteria. *Rsc Adv* **4**, 3413–3421 (2014). <https://doi.org/10.1039/c3ra44924f>
- Jurys, A. & Pedersen, C. M. Convergent and stereoselective synthesis of nonulosonic acid (NuLO) scaffolds via imine/aldehyde reductive coupling. *Org Chem Front* (2026). <https://doi.org/10.1039/d5qo01524c>
- Guo, X., Li, P. F., Liu, H. & Li, X. C. Recent progress on the chemical synthesis of bacterial non-2-ulosonic acids. *Org Chem Front* **10**, 3429–3446 (2023). <https://doi.org/10.1039/d3qo00637a>
- Kay, E., Cuccui, J. & Wren, B. W. Recent advances in the production of recombinant glycoconjugate vaccines. *Npj Vaccines* **4** (2019). <https://doi.org/10.1038/s41541-019-0110-z>
- Lee, I. M. *et al.* Pseudaminic Acid on Exopolysaccharide of *Acinetobacter baumannii* Plays a Critical Role in Phage-Assisted Preparation of Glycoconjugate Vaccine with High Antigenicity. *J Am Chem Soc* **140**, 8639–8643 (2018). <https://doi.org/10.1021/jacs.8b04078>
- Wei, R. H. *et al.* Synthetic Pseudaminic-Acid-Based Antibacterial Vaccine Confers Effective Protection against *Acinetobacter baumannii* Infection. *Acs Central Sci* **7**, 1535–1542 (2021). <https://doi.org/10.1021/acscentsci.1c00656>
- Flack, E. K. P., Chidwick, H. S., Best, M., Thomas, G. H. & Fascione, M. A. Synthetic Approaches for Accessing Pseudaminic Acid (Pse) Bacterial Glycans. *ChemBiochem* **21**, 1397–1407 (2020). <https://doi.org/10.1002/cbic.202000019>
- Schoenhofen, I. C., McNally, D. J., Brisson, J. R. & Logan, S. M. Elucidation of the CMP-pseudaminic acid pathway in : synthesis from UDP-acetylglucosamine by a single enzymatic reaction. *Glycobiology* **16**, 8c–14c (2006). <https://doi.org/10.1093/glycob/cwl010>
- Schoenhofen, I. C., Vinogradov, E., Whitfield, D. M., Brisson, J. R. & Logan, S. M. The CMP-legionaminic acid pathway in *Campylobacter*: Biosynthesis involving novel GDP-linked precursors. *Glycobiology* **19**, 715–725 (2009). <https://doi.org/10.1093/glycob/cwp039>
- Hassan, M. I. *et al.* Total Biosynthesis of Legionaminic Acid, a Bacterial Sialic Acid Analogue. *Angew Chem Int Edit* **55**, 12018–12021 (2016). <https://doi.org/10.1002/anie.201606006>
- Walklett, A. J. *et al.* The Retaining Pse5Ac7Ac Pseudaminyltransferase KpsS1 Defines a Previously Unreported glycosyltransferase family (GT118). *Angew Chem Int Edit* **63** (2024). <https://doi.org/10.1002/anie.202318523>
- Ardissone, S., Kint, N. & Viollier, P. H. Specificity in glycosylation of multiple flagellins by the modular and cell cycle regulated glycosyltransferase FlmG. *Elife* **9** (2020). <https://doi.org/10.7554/eLife.60488>
- Tomek, M. B. *et al.* A General Protein-Glycosylation Gene Cluster Encodes the Species-Specific Glycan of the Oral Pathogen: Glycan Biosynthesis and Immunological

- Implications. *Front Microbiol* **9** (2018). <https://doi.org/10.3389/fmicb.2018.02008>
- 30 Sulzenbacher, G. *et al.* Glycosylate and move! The glycosyltransferase Maf is involved in bacterial flagella formation. *Environ Microbiol* **20**, 228–240 (2018). <https://doi.org/10.1111/1462-2920.13975>
 - 31 Khairnar, A., Sunsunwal, S., Babu, P. & Ramya, T. N. C. Novel serine/threonine-O-glycosylation with N-acetylneuraminic acid and 3-deoxy-D-manno-octulosonic acid by bacterial flagellin glycosyltransferases. *Glycobiology* **31**, 288–306 (2021). <https://doi.org/10.1093/glycob/cwaa084>
 - 32 Sunsunwal, S., Khairnar, A., Subramanian, S. & Ramya, T. N. C. Harnessing the acceptor substrate promiscuity of Maf glycosyltransferase to glyco-engineer mini-flagellin protein chimeras. *Commun Biol* **7** (2024). <https://doi.org/10.1038/s42003-024-06736-y>
 - 33 Parker, J. L., Day-Williams, M. J., Tomas, J. M., Stafford, G. P. & Shaw, J. G. Identification of a putative glycosyltransferase responsible for the transfer of pseudaminic acid onto the polar flagellin of *Aeromonas caviae* Sch3N. *Microbiologyopen* **1**, 149–160 (2012). <https://doi.org/10.1002/mbo3.19>
 - 34 Karlyshev, A. V. & Wren, B. W. Development and application of an Insertional system for gene delivery and expression in. *Appl Environ Microb* **71**, 4004–4013 (2005). <https://doi.org/10.1128/Aem.71.7.4004-4013.2005>
 - 35 McNally, D. J. *et al.* Functional characterization of the flagellar glycosylation locus in 81-176 using a focused metabolomics approach. *J Biol Chem* **281**, 18489–18498 (2006). <https://doi.org/10.1074/jbc.M603777200>
 - 36 Yang, K. Y. *et al.* Glycosyltransferase Jhp0106 (PseE) contributes to flagellin maturation in *Helicobacter pylori*. *Helicobacter* **26** (2021). <https://doi.org/10.1111/hel.12787>
 - 37 Meng, X. K., Boons, G. J., Wösten, M. M. S. M. & Wennekes, T. Metabolic Labeling of Legionaminic Acid in Flagellin Glycosylation of *Campylobacter jejuni* Identifies Maf4 as a Putative Legionaminyl Transferase. *Angew Chem Int Edit* **60**, 24811–24816 (2021). <https://doi.org/10.1002/anie.202107181>
 - 38 van Ede, J. M., Sorensen, M. C. H., van Loosdrecht, M. C. M. & Pabst, M. SugarBase: mapping glycomolecule precursors in microbes. *bioRxiv* (2026).
 - 39 Melander, R. J., Zurawski, D. V. & Melander, C. Narrow-Spectrum Antibacterial Agents. *Medchemcomm* **9**, 12–21 (2018). <https://doi.org/10.1039/C7MD00528H>
 - 40 Yu, H. *et al.* A Multifunctional *Pasteurella multocida* Sialyltransferase: A Powerful Tool for the Synthesis of Sialoside Libraries. *J Am Chem Soc* **127**, 17618–17619 (2005). <https://doi.org/10.1021/ja0561690>
 - 41 Watson, D. C. *et al.* Sialyltransferases with enhanced legionaminic acid transferase activity for the preparation of analogs of sialoglycoconjugates. *Glycobiology* **25**, 767–773 (2015). <https://doi.org/10.1093/glycob/cwv017>
 - 42 Cheng, J. S. *et al.* Multifunctionality of *Campylobacter jejuni* sialyltransferase CstII: Characterization of GD3/GT3 oligosaccharide synthase, GD3 oligosaccharide sialidase, and trans-sialidase activities. *Glycobiology* **18**, 686–697 (2008). <https://doi.org/10.1093/glycob/cwn047>
 - 43 Singh, J. K., Adams, F. G. & Brown, M. H. Diversity and Function of Capsular Polysaccharide in *Acinetobacter baumannii*. *Front Microbiol* **9** (2019). <https://doi.org/10.3389/fmicb.2018.03301>
 - 44 Knirel, Y. A. *et al.* 5,7-Diamino-3,5,7,9-tetra-deoxynon-2-ulosonic Acids in the Capsular Polysaccharides of *Acinetobacter baumannii*. *Biochemistry-Moscow* **88**, 202–210 (2023). <https://doi.org/10.1134/S0006297923020049>
 - 45 Kenyon, J. J. *et al.* The K46 and K5 capsular polysaccharides produced by NIPH 329 and SDF have related structures and the side-chain non-ulosonic acids are 4-O-acetylated by phage-encoded O-acetyltransferases. *Plos One* **14** (2019). <https://doi.org/10.1371/journal.pone.0218461>
 - 46 Kenyon, J. J., Notaro, A., Hsu, L. Y., De Castro, C. & Hall, R. M. 5,7-Di-acetyl-8-epiacinetaminic acid: A new non-2-ulosonic acid found in the K73 capsule produced by an isolate from Singapore. *Sci Rep-Uk* **7** (2017). <https://doi.org/10.1038/s41598-017-11166-4>
 - 47 Senchenkova, S. N. *et al.* Structure of a new pseudaminic acid-containing capsular polysaccharide of *Acinetobacter baumannii* LUH5550 having the KL42 capsule biosynthesis locus. *Carbohydr Res* **407**, 154–157 (2015). <https://doi.org/10.1016/j.carres.2015.02.006>
 - 48 Kenyon, J. J., Marzaioli, A. M., Hall, R. M. & De Castro, C. Structure of the K2 capsule associated with the KL2 gene cluster of. *Glycobiology* **24**, 554–563 (2014). <https://doi.org/10.1093/glycob/cwu024>
 - 49 Flack, E. K. P. *et al.* Biocatalytic Transfer of Pseudaminic Acid (Pse5Ac7Ac) Using Promiscuous Sialyltransferases in a Chemoenzymatic Approach to Pse5Ac7Ac-Containing Glycosides. *Acs Catal* **10**, 9986–9993 (2020). <https://doi.org/10.1021/acscatal.0c02189>
 - 50 Kleikamp, H. B. C. *et al.* Tackling the chemical diversity of microbial nonulosonic acids - a universal large-scale survey approach. *Chem Sci* **11**, 3074–3080 (2020). <https://doi.org/10.1039/c9sc06406k>
 - 51 Guo, X. *et al.* Investigation on the substrate specificity and N-substitution tolerance of PseF in catalytic transformation of pseudaminic acids to CMP-Pse derivatives. *Chem Sci* **15**, 5950–5956 (2024). <https://doi.org/10.1039/d4sc00758a>

Discovery of Microbial Glycoside Hydrolases via Enrichment and Metaproteomics

Jitske M. van Ede, Suzanne van der Steen, Geert M. van
der Kraan, Mark C.M. van Loosdrecht and Martin Pabst



Abstract

The immense microbial diversity on Earth represents a vast genomic resource, yet discovering novel enzymes from complex environments remains challenging. Here, we combine a microbial enrichment with metagenomics and metaproteomics to facilitate the identification of microbial glycoside hydrolases that operate under defined conditions. We enriched microbial communities on the carbohydrate polymer pullulan at elevated temperatures and under acidic conditions. Pullulan is a natural polysaccharide composed of maltotriose units linked by α -1,6 glycosidic bonds. Along with its hydrolyzing enzymes, it has broad applications across various industries. The enrichment inocula were sampled from thermophilic compost and from soil from the bank of a pond. In both cases, *Alicyclobacillus* emerged as the dominant microorganism. Metaproteomic analysis of the enriched biomass and secretome enabled the identification of several pullulan-degrading enzyme candidates from this organism. These enzymes were absent in the metagenomic analysis of the initial inoculum, which is highly complex with a wide diversity of species. This underscores the effectiveness of combining microbial enrichment with multi-omics for uncovering novel enzymes and sequence variants, that operate under defined conditions from complex microbial environments.

Published as:

van Ede, J. M., van der Steen, S., van der Kraan, G. M., van Loosdrecht, M. C., & Pabst, M. (2025). Discovery of Microbial Glycoside Hydrolases via Enrichment and Metaproteomics. *RSC Chemical Biology*, 6(11), 1809-1821.

6.1 | Introduction

Enzymes are natural biocatalysts with great potential to advance sustainable processes in biotechnology and industry. Not only are they biodegradable, enzymes also have the potential to conduct conversions under ambient conditions, reduce the use of fossil fuels and chemical waste. Moreover, enzymes provide complex molecules that are challenging to produce via chemical synthesis alone^{1,2}. Enzymes are already extensively employed in industry, with hydrolases being the most prevalent class of enzymes in use³. Hydrolases can modify biopolymers, altering their properties, or break them down into smaller fragments. A vast array of applications exists for (modified) biopolymers. For example, carbohydrate biopolymers are widely applied in the food industry, nutraceuticals, and care products. Cellulose, obtained from plant cell walls, is used as a stabilizer and thickener in various food formulations⁴. Chitosan, derived from crustacean shells, exhibits antimicrobial properties, making it a valuable component in dietary supplements⁵. Another example is pullulan, a natural polysaccharide produced by the fungus *Aureobasidium pullulans*, which is highly valued in the food and pharmaceutical industries for its ability to form edible, non-toxic films and coatings^{6,7}. Its applications include encapsulating food flavours and fragrances, improving product shelf life, and acting as a binder in orally dissolvable films for pharmaceuticals, demonstrating its versatility and biocompatibility⁸. Furthermore, also the hydrolysing enzymes are used in industrial applications. For example, pullulanases are essential in the starch-processing industry, where they are used to hydrolyze pullulan into maltotriose units, thereby enhancing the efficiency of starch liquefaction and saccharification processes. Their use significantly improves the production of glucose and maltose syrups, providing a higher yield and specificity compared to conventional methods⁹. Hydrolases are also employed in cleaning processes, such as laundry. These enzymes facilitate the breakdown of polymers, thereby boosting cleaning efficiency. Additionally, their application offers environmental benefits by enabling lower washing temperatures, shorter wash times, and reduced water usage^{10,11}.

Pullulan represents a polysaccharide which consists of maltotriose units, in which three glucoses are linked via an α -1,4 linkage, and these maltotriose units are linked via α -1,6 glycosidic bonds¹². Pullulan degrading enzymes can be divided into two main groups, pullulanases (type I and type II) and pullulan hydrolases (type I, II and III). Pullulanases cleave the α -1,6 bond, thereby converting pullulan into maltotriose. In addition, pullulanases are also active against other sugar polymers, including starch and glycogen. While pullulanase type I can only cleave the α -1,6 linkages of these sugar polymers, pullulanase type II can cleave both the α -1,6 and

α -1,4 linkages. Pullulan hydrolases are further divided in three types. Type I (neopullulanase) and type II (isopullulanase) cleave the α -1,4 linkage of pullulan, thereby forming panose and isopanose, respectively. Type III has the ability to cleave both the α -1,4 and α -1,6 linkage of pullulan, forming panose, maltotriose, maltose and glucose¹³. In general, the optimal pH of pullulan-degrading enzymes varies widely, ranging from pH 3 to 11, although the majority exhibit optimal activity between pH 5 and 8. Similarly, their optimal temperatures span a broad range from 30°C to 105°C (SI document, Figure S6.5)¹³. Certain enzyme groups, such as pullulanase type I and II, include members active across this entire temperature and pH spectrum. The so far identified pullulan hydrolases type I and II seem to occupy a more narrow space, with pH optima in the range of 5 to 8 for type I and 3.5 to 5 for type II. Pullulan hydrolases type III have only been identified in two archaeal species, both exhibiting high optimal temperatures of 95°C and 97.5°C, and optimal pH values of 6.5 and 3.5, respectively (SI document, Figure S6.5)¹³.

Although a wide variety of enzymes are utilized in various industries, the quest for new enzymes—including pullulanases—with improved features is ongoing¹³. There is currently a focused effort on employing advanced protein/enzyme optimization techniques, such as artificial intelligence, directed evolution, and intelligent design, to improve enzymes¹⁴⁻²⁰. However, despite progress in protein engineering and other modification strategies, these approaches don't always yield the anticipated results due to the complex relationship between the structure and function of proteins²¹. Fortunately, a remarkable diversity of microorganisms inhabit our planet, providing us with a vast genetic, and therefore biocatalytic resource². Exploring this natural diversity with sophisticated enzyme discovery methods presents a promising alternative for finding enzymes with the desired traits or for identifying templates for further engineering.

Conservative methods to mine genetic resources use cultivation of a microorganism and to clean streak it on a plate under certain conditions. Subsequently, a colony is isolated of which the DNA is extracted and sequenced^{22,23}. After the genome is assembled it can be mined for genetic targets of interest. Nevertheless, this approach has limitations, namely, most microorganisms cannot be grown as a pure culture under conditions which prevail in laboratory environments. This limitation is better known as “the great plate anomaly”. It is expected that only about 1% of the microorganisms in a sampled environment can be found in a lab after enrichment²⁴⁻²⁶. A way to avoid this problem and become culture independent is by using metagenomics²⁷. Over the past years, whole metagenome sequencing became a popular method to sequence the entire metagenome from natural microbial communities²⁸. Currently, two main

metagenomic strategies can be distinguished, the sequence based and the function based approach²⁹. The function-based approach involves constructing a genomic library from environmental DNA, followed by screening individual genes in a cultivable host cell for desired activities²⁸⁻³⁰. While this method enables the discovery of entirely novel enzymes, it is limited by high costs and the requirement of high-throughput activity assays². In contrast, the sequence based approach depends on sequence similarity to known templates^{29,30}, which restricts its ability to identify enzymes with novel functions².

More recent strategies include microbial enrichment cultures followed by multi-omics analysis of the cultures (genomics and proteomics, supported by advanced bioinformatics tools)³¹. These strategies first enrich for certain organisms which possess an enzyme with a desired function. This route is only feasible when the enzyme plays a crucial role in the growth of the organism, e.g. is involved in the carbon or nitrogen source utilization³². For instance, in pursuit of hydrolytic enzymes which cleave specific polymers, an enrichment with the polymer as the sole carbon source can be performed. This ensures that only microorganisms that can hydrolyse and utilize the polymer hydrolysis products will grow³³. Moreover, the conditions for enriching microbes can be tuned towards the conditions in which the enzymes should be active (e.g. temperature, pH and salinity). Subsequently, instead of only employing metagenomics to mine for the enzymes of interest, a more streamlined approach is to also integrate proteomics into the workflow^{31,34}. Advantageously, metaproteomics allows to identify the actually expressed enzymes, as well as their cellular location, such as cytosolic, membrane-bound, or secreted^{35,36}. This greatly simplifies genomic data and it supports identification of novel enzyme candidates or those with enhanced properties. It also provides template sequences for further engineering to optimize desired properties³⁷. In summary, this study demonstrates the discovery of novel pullulan-degrading enzymes that operate at defined conditions, by combining microbial enrichment with metagenomics and metaproteomics.

6.2 | Materials and Methods

Microbial enrichment. Organic matter from compost was obtained from a compost pile in Waddinxveen (further referred to as “compost”), The Netherlands. Soil was collected from two locations in Delft, The Netherlands (at approximately 51°59'36.3"N 4°22'23.6"E and 51°59'37.5"N 4°22'34.0"E) and combined (further referred to as “soil”). The enrichments were performed under sterile conditions (to ensure only microbial diversity from the sampled sources are enriched) in 250 mL shake flasks with screw caps. The growth medium contained 0.5 g/L KH₂PO₄, 3 g/L

$(\text{NH}_4)_2\text{SO}_4$, 0.2 g/L $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 10 ml/L MD-VS ATCC vitamin supplement and 2.5 ml/L trace element solution (15 g/L EDTA (Titriplex III), 4.5 g/L $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 0.84 g/L $\text{MnCl}_2 \cdot 2\text{H}_2\text{O}$, 0.30 $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$, 0.20 g/L $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, 0.40 g/L $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$, 4.5 g/L $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 3.0 g/L $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 1.0 g/L H_3BO_3 and 0.10 g/L KI). Pullulan was added to a final concentration of 10 g/L and the pH of the medium was adjusted to 4.5 with 2 M HCl. Finally, the medium was sterilized in the autoclave at 110 °C for 30 min. and subsequently stored at 4°C until usage. 50 mL of growth medium was transferred to a 250 mL sterile flask. The medium was inoculated by either 0.5 g of compost further referred to as “compost enrichment”, or by 0.5 g of soil, further referred to as “soil enrichment”. In addition, a control culture without inoculum was included. The enrichments were kept at 57°C, 100 rpm in the dark in an incubator hood (TH 15, Edmund Buhler GmbH). The pH was monitored twice a day using a pH probe (826 pH mobile, Metrohm) and if necessary adjusted back to pH 4.5 using 0.2 M KOH or 0.2 M HCl.

No organic buffer was used to prevent it from becoming a potential substrate for microbial growth. All steps were performed next to a Bunsen burner to maintain sterile conditions. The enrichments were transferred twice in a 1:1000 (v/v) ratio to fresh medium when almost all pullulan was consumed, after 3 and 4 days, respectively. After the second transfer, the OD_{600} (using the cell density meter Ultrospec 10, Biochrom) and the pullulan consumption (measured by HPLC analysis) were monitored daily, while the pH was still measured twice a day. The microbial enrichments were analyzed once all pullulan was consumed, after 8 and 14 days for the compost and soil enrichment, respectively. Images of the enrichments were created using the Zeiss Axio microscope with a 100 x magnification. In addition, a 250 mL flask was inoculated with 0.25 g compost combined with 0.25 g soil collected on the TUD campus and incubated at 75°C in a water bath. The enrichment was monitored as described above.

Monitoring degradation of pullulan by HPLC. Approximately 1 mL of culture broth was centrifuged for 10 min. at 4 °C, 14000 rcf, after which 667 μL of the supernatant was transferred to a clean screw cap vial. Trifluoroacetic acid (TFA) was added to reach a final concentration of 4 M TFA. The samples were incubated for 4 hours at 100°C in the ThermoMixer with Thermotop (Eppendorf, Thermomixer C). After a centrifugation step of 15 min. at 4°C, 14000 rcf, the supernatant was collected in a clean Eppendorf tube. The samples were measured on a Vanquish HPLC system (Thermo Scientific, Germany) with a Aminex HPX-87H separation column. A constant flow rate of 0.750 mL/min was maintained over a total run time of 45 minutes, using 1.5 mM phosphoric acid in milli-Q water as eluent. The column

chamber temperature was kept at 50°C and compounds were detected using the RI detector (ERC, RefractoMax520). Data analysis was performed with Chromeleon 7 (Thermo Scientific, Germany).

Whole metagenome sequencing. Whole metagenome sequencing of the starting biomass and enrichments was performed by BaseClear B.V. (Leiden, The Netherlands). Briefly, DNA was extracted using standard molecular biology kits from Zymo Research, and DNA concentration was confirmed using a nanodrop spectrophotometer. Sequence libraries were generated using Nextera XT DNA Library Preparation Kit. Paired-end sequence reads were generated using the Illumina NovaSeq PE150 system. The sequences generated were performed under accreditation according to the scope of BaseClear B.V. (L457; NEN-EN-ISO/IEC 17025). FASTQ read sequence files were generated using bcl2fastq version 2.20 (Illumina). Initial quality assessment was based on data passing the Illumina Chastity filtering. Subsequently, reads containing PhiX control signal were removed using an in-house filtering protocol. In addition, reads containing (partial) adapters were clipped (up to a minimum read length of 50 bp). The second quality assessment was based on the remaining reads using the FASTQC quality control tool version 0.11.8. Metagenome assembly has been performed using MEGAHIT 1.2.9. Contigs smaller than 1000 base pairs have been removed from final assembly. PROKKA was used to locate open reading frames (ORFs) on contigs, and to translate the ORFs to protein sequences. These sequences were used as reference database for the metaproteomic analysis. Taxonomic classification of ORFs was performed using the DIAMOND sequence aligner and the NCBI sequence reference database. A consensus lineage for every ORF was obtained by using the bit score approach from the top 25 alignments described earlier^{35,38}. Functional classification was performed as described below. In addition, a sequence alignment of the *CdaA* (Uniprot: Q9WX32) and *AmyA* (Uniprot: Q06307) gene of *Alicyclobacillus acidocaldarius* was performed against the metagenomics data of both the enrichment cultures and the inocula using the DIAMOND sequence aligner.

Metaproteomics on microbial enrichments. Biomass of the enrichment cultures were collected in a clean LoBind Eppendorf tube (approx. 25 mg wet weight). Subsequently, 0.175 mL of TEAB resuspension buffer (50 mM TEAB, 1% (w/w) sodium deoxycholate, pH 8), 0.175 mL of B-PER reagent (Bacterial Protein Extraction Reagent, Thermo Scientific) and 0.15 g of acid washed glass beads (150–212 µm) were added. The samples were homogenized using the bead beater for 1.5

min., followed by a 1-minute incubation on ice and a centrifugation for 2 min. at 4 °C, 14000 rcf. This cycle was repeated twice. Afterwards, the samples were subjected to an ultrasonic bath for 10 min. and centrifuged for 15 min. at 4 °C, 14000 rcf. All centrifugation steps were conducted at 4 °C and 14000 rcf unless stated otherwise. The supernatant was collected in a LoBind Eppendorf tube. Trichloroacetic acid (TCA) was added in a 1:4 ratio to the supernatant. The samples were vortexed and incubated for 30 min. at 4 °C, followed by centrifugation for 15 min. The supernatant was discarded, and the protein pellet was washed with 200 µL of ice-cold acetone, vortexed and centrifuged for 15 min. The supernatant was removed and 100 µL of 6 M Urea was added to the protein pellets. To re-dissolve the pellets, the samples were vortexed thoroughly and incubated at 37 °C, 300 rpm in the ThermoMixer with Thermotop (Eppendorf, Thermomixer). If necessary, an additional 100 µL of 6 M Urea was added to dissolve the pellets. The samples were reduced by incubating at 37 °C for 60 min. after adding 30 µL 10 mM dithiothreitol, and subsequently alkylated for 30 minutes, at RT in the dark, by adding 30 µL of 20 mM iodoacetamide. The samples were diluted with 200 mM ammonium bicarbonate to achieve a Urea concentration < 1 M. To 100 µL of the final sample volume, 5 µL of 0.1 µg/µL trypsin solution was added. After gently shaking, the samples were incubated at 37 °C overnight. The samples were cleaned and concentrated using the Oasis HLB 96-well µElution Plate with 2 mg sorbent per well, 30 µm (Waters, UK). In short, the columns were conditioned with 750 µL MeOH, equilibrated with 2 x 500 µL MS-H₂O and then loaded with the samples. Two washing steps with 350 µL 5% MeOH in MS-H₂O were performed, followed by elution with 200 µL 2% formic acid in 80% MeOH and 200 µL 1 mM ammonium bicarbonate (ABC) in 80% MeOH. The eluates were collected in a LoBind Eppendorf tube. The samples were dried using the SpeedVac concentrator (Thermo Scientific) at 50 °C and stored at -20 °C until analysis by LC-MS. For LC-MS analysis, the dried samples were dissolved in 15 µL 3% ACN plus 0.1% trifluoroacetic acid (TFA) in MS-H₂O, incubated for 30 min. at RT, and vortexed thoroughly. The samples were diluted to a final protein concentration of approximately 0.5 mg/ml, estimated at 280 nm using the NanoDrop ND-1000 spectrophotometer (Thermo Scientific), before proteomic analysis.

Extracellular proteins from the secretome of the enrichment cultures were prepared as described in the following. Samples from the enrichment culture were centrifuged for 15 min. and the supernatant was collected. Several protein precipitation methods were employed including TCA precipitation as described above, an acetone precipitation, an acetone/salt precipitation and a filter assisted sample preparation (FASP). The latter three are described below. For the acetone

precipitation 250 μ L sample was mixed with 1250 μ L ice-cold acetone, vortexed and stored at -20°C for 30 min. The samples were centrifuged for 15 min. and the supernatant was removed. For the acetone/salt precipitation 50 μ L of 3 M NaCl solution was added to 250 μ L sample. Subsequently, 1200 μ L acetone (RT) was added and the samples were mixed gently. The samples were incubated for 30 min. at RT and centrifuged for 15 min. The supernatant was carefully removed, and the protein pellet was washed with 400 μ L acetone (RT). Another 15-minute centrifuging was performed, and the supernatant was removed. After protein precipitation, the protein pellet was re-dissolved in 100 μ L of 6 M Urea and processed as described above. For the filter assisted sample preparation 200 μ L 6 M Urea was added to a 10 kDa Microcon filter (Merk-Millipore), followed by a centrifugation of 30 min. at 14000 rcf and 20°C . After discarding the flow-through, this step was repeated once. Unless stated otherwise, all centrifugations during FASP were conducted at 14000 rcf and 20°C . Either 250 μ L sample or 4 x 500 μ L sample was loaded onto the filter followed by centrifuging for 30 min. The samples were then reduced by adding 30 μ L of 10 mM DTT and 70 μ L of 200 mM ammonium bicarbonate. After vortexing and incubating at 37°C for 60 min., 30 μ L of 20 mM IAA was added to alkylate the samples. Following a 30 min. incubation in the dark at RT, the samples were centrifuged for 30 min. Subsequently a wash and equilibration step was performed by first adding 100 μ L of 6M Urea and centrifuging for 30 min. followed by adding 100 μ L of 200 mM ammonium bicarbonate twice and centrifuging for 30 min. The filter was transferred to a clean collection tube and 5 μ L 0.1 $\mu\text{g}/\mu\text{L}$ trypsin solution together with 95 μ L 200 mM ammonium bicarbonate was added to the filter. The filter was incubated overnight in the ThermoMixer with thermotop at 37°C . Sample collection was performed by centrifuging for 30 min. Subsequently, 150 μ L 200 mM ammonium bicarbonate was added to the filter and a centrifugation of 30 min. followed. The samples were again centrifuged for 30 min. after the addition of 150 μ L of 10% acetonitrile (ACN), 0.1% formic acid (FA) in MS- H_2O . Finally, the samples were cleaned and concentrated using solid phase extraction, as described previously. The results from the different sample preparation methods applied to identify extracellular proteins were combined before further analysis for glycosyl hydrolase enzyme candidates.

Shotgun proteomics was performed on an EASY 1200 nano-LC coupled to a Q Exactive plus Orbitrap Mass Spectrometer (Thermo Scientific, Germany). Chromatographic separation was carried out on a 0.05 x 150 mm C18 column (Thermo Scientific, catalogue number 164943), with mobile phase A consisting of 0.1% formic acid (FA), 1% acetonitrile (ACN) in MS- H_2O and mobile phase B

consisting of 0.1% FA and 80% ACN in MS-H₂O. The chromatographic profile included an initial 2-minute run with 5% B, followed by a linear gradient from 5% to 25% B over 90 min, then a gradient to 55% B over 60 min. The gradient was subsequently returned to 5% B within 3 min. and equilibrated for an additional 20 min., all at a constant flow rate of 350 nL/min. Sample injections of 2 μ L were performed with blank runs in between each sample. Electrospray ionization operated in the positive mode, and MS1 analysis was conducted at a resolution of 70,000, with an AGC target of 3e6 and a maximum injection time of 75 ms. Applying the data dependent acquisition mode, precursors for fragmentation were isolated using a 2.0 m/z window over a scan range of 385–1250 m/z. A normalized collision energy (NCE) was set at 28%. MS2 spectra were collected at a resolution of 17,500, an AGC target of 2e5 and a maximum injection time of 75 ms. Database searching was performed using PEAKS Studio 10.5 (Bioinformatics Solutions Inc., Canada) using the sequence database obtained from whole metagenome sequencing experiments (described above) and the GPM crap contaminant database. Carbamidomethylation was considered a fixed modification, while oxidation and deamidation were included as variable modifications. Trypsin was set as proteolytic enzyme, allowing a maximum of three missed cleavages. The parent mass error tolerance was set to 20.0 ppm and the fragment mass error to 0.02 Da. Finally, filtering criteria, included a 1% false discovery rate (FDR) for peptide-spectrum matches and a requirement of at least 2 unique peptides for protein identification.

Functional annotation pipeline. To facilitate the identification of the target glycoside hydrolase, a streamlined bioinformatics pipeline was established. For this, a Virtual Box (Oracle VM Virtual Box, version 7.0.2) was set up with Linux Mint 21 (cinnamon 64 bit) and the pipeline was operated using Python 3.10.12. The functional annotation pipeline was initiated within the python module. The input metagenomics fasta file is first split into smaller files containing max. 1000 proteins each. The program then loops through these files and performs an annotation using HMM sequence models, by using InterProScan (version 5.59-91.0). InterProScan is run from within the python program by making use of the module 'subprocess'. The split fasta files serve as an input, generating an output in xml format. The precalculated match lookup service of InterProScan has been disabled to run everything locally. After the annotation with InterProScan, a keyword search was performed to mine for sequences that contain elements matching the target functions. Lastly, the data were combined with the metaproteomics results (identified proteins, and abundances).

Expression of enzyme candidate MMBJNONL_14124. Expression of the enzyme candidate MMBJNONL_14124 was carried out by Trenzyme GmbH (Germany). In short, a codon-optimized gene sequence, including a C-terminal His-tag, was synthesized and cloned into the expression vector pTZ_E02_(1)_XX124. Protein expression was performed in the host strain *Rosetta2 (DE3)*. A 100 mL culture in FM medium was grown at 37°C; during autoinduction, the temperature was shifted to 23°C. The total cultivation lasted 43 h, reaching a final OD₆₀₀ of 14.1 and yielding 3.0 g of biomass.

Cell lysis. Approximately 1.5 g of biomass was resuspended in 7.5 ml of 50 mM potassium phosphate pH 7.4, containing 1 mM MgCl₂, a spatula tip of DNase and one cOmplete™ mini Protease Inhibitor Cocktail tablet. The cells were lysed using Lysin Matrix B (MP Biomedicals) in the FastPrep-24 Classic instrument (MP Biomedicals) for 2 cycles of 45 sec at 6.5 m/s with a 5 min break on ice in between. After centrifuging for 15 min, 14000 rcf at 4°C, the SN was filtered using a 0.22 µm filter. Purification of the His-tagged protein from the cell free extract (CFE) was performed on a Bio-Rad NGC Chromatography System using a HisTrap ff crude 1 mL column. The column was equilibrated with 5 column volumes (CVs) of 100% mobile phase A (50 mM potassium phosphate pH 7.4, 300 mM NaCl, 20 mM imidazole and 1 mM DTT), after which the CFE was loaded on the column. A wash phase of 10 CVs mobile phase A was performed, followed by a linear gradient to 100% mobile phase B (50 mM potassium phosphate pH 7.4, 300 mM NaCl, 500 mM imidazole and 1 mM DTT) over 40 CVs, while collecting 0.5 mL fractions. Fractions containing the purified protein (A16-A35) were pooled and concentrated using an Amicon® Ultra Centrifugal Filter, 10 kDa MWCO. The buffer was exchanged to a 50 mM potassium phosphate buffer pH 7.4 using a PD10 desalting column (GE healthcare), according to the manufacturer's protocol. After eluting with 2.5 mL 50 mM potassium phosphate buffer pH 7.4, the sample was again concentrated to ~0.5 mL. **SDS-PAGE.** Proteins were separated by SDS-PAGE on a 4–15% Mini-PROTEAN® TGX Stain-Free™ Protein Gel. As a reference the Precision Plus Protein™ Unstained Protein Standards (Strep-tagged recombinant) was loaded on the gel. 15 µL sample was denatured in 14.5 µL 2 x Laemmli buffer (65.8 mM Tris-HCl, pH 6.8, 2.1% SDS, 26.3% (w/v) glycerol, 0.01% bromophenol blue) and 0.5 µL 100 mM DTT at 95°C for 5 min before loading onto the gel. Electrophoresis was performed at 120 V in 10x Tris/Glycine/SDS buffer until the dye front reached the bottom of the gel (~30 min). Proteins were visualised using the ChemiDoc MP Imaging system of Bio-Rad, using optimal auto exposure for 45 sec.

Activity assay confirming pullulan hydrolytic activity of MMBJNONL_14124.

Reactions with a final volume of 100 μL were prepared in 0.6 mL tubes, each containing 1 mg/mL pullulan, 5 mM CaCl_2 and 100 mM sodium acetate pH 4.5. Activity assays were performed first with the cell free extract and subsequently with the purified enzyme. The CFE was five times diluted in 125 mM sodium acetate pH 4.5, and subsequently diluted three times in H_2O . 50 μL of the diluted CFE was added to the 50 μL reaction mixture (final volume of 100 μL). 50 μL of purified enzyme (~ 1.5 mg/mL in 50 mM potassium phosphate pH 7.4) was diluted in 100 μL Optima LC/MS-grade H_2O . Subsequently 50 μL of this dilution was added to the 50 μL reaction mixture. The reactions were incubated overnight at 60°C. Negative controls lacking either the cell free extract/enzyme or the substrate (pullulan) were included. The samples were washed and concentrated using a HyperSep™ Hypercarb™ SPE 96-well plate (25/1 mL Plate, Thermo Scientific). Briefly, the columns were washed with 750 μL 10 mM ammonium bicarbonate (ABC) in 60% Optima LC/MS-grade acetonitrile (ACN) and equilibrated with 2 x 500 μL Optima LC/MS-grade H_2O . Each sample was diluted to 200 μL and centrifuged at 14K rpm for 5 minutes to remove insoluble material. The samples were then applied to the PGC cartridges, washed twice with 200 μL H_2O and the carbohydrate mono and oligomer fraction was eluted with 300 μL 10 mM ABC in 60% ACN. Eluates were collected in Eppendorf tubes, dried in the SpeedVac concentrator at 60°C and re-dissolved in 50 μL H_2O . For PGC-MS/MS analysis, 2 μL of this sample was injected to the LC-MS system. **PGC-MS/MS analysis.** Liquid chromatographic separation coupled to high resolution mass spectrometric detection was conducted using an Acquity Ultra Performance Liquid Chromatograph (Waters, UK) coupled to a Q Exactive™ Focus Hybrid Quadrupole-Orbitrap™ Mass Spectrometer (Thermo Scientific, Germany). Chromatographic separation was achieved on a 100 x 1 mm Hypercarb™ Porous Graphitic Carbon HPLC Column (Thermo Scientific, No. 35005-101030) using a constant flow rate of 100 $\mu\text{L}/\text{min}$ and an electrical grounding between the ESI source and the column inlet. Mobile phase A consisted of 0.1% formic acid (FA) in Optima LC/MS-grade H_2O and mobile phase B of 0.1% FA in Optima LC/MS-grade acetonitrile. The column was equilibrated with 1% B for 5.1 min, followed by a linear gradient to 40% B over 4.9 min. The system was returned to the initial conditions and equilibrated for another 4 min. 2 μL injections were made, analysing each sample in duplicate with a wash in between each sample. MS1 spectra were acquired in positive mode at a resolution of 70,000 over scan range of 150–1250 m/z. The automatic gain control (AGC) was set to 1e6 and the maximum injection time (IT) to auto. Precursors for product hexose mono and

oligomers were included for continuous isolation and fragmentation in PRM mode. The MS2 spectra were acquired at a resolution of 35,000, with the AGC set at $2e5$ and the maximum IT at auto. Fragmentation occurred with a 25% normalized collision energy. Raw data were processed using XCalibur 4.1 (Thermo Fisher Scientific, Germany). Calibration of the mass spectrometer was performed using the Pierce™ LTQ ESI negative ion calibration solution (Thermo Fisher Scientific, Germany).

6.3 | Results

6.3.1. Enrichment on pullulan as the sole carbon source.

To enrich for microbes capable of degrading pullulan under thermophilic (57°C) and acidic (pH 4.5) conditions, pullulan was supplied as the sole carbon source. The impact of the inoculum on the enrichment of the microbial communities, and therefore the enzymes obtained, was evaluated by establishing two enrichment cultures. The first, referred to as the 'compost enrichment', was inoculated with organic matter from a compost pile. This microbiome was already exposed to elevated temperatures during composting. The second enrichment, called 'soil enrichment', was inoculated with soil from the bank of a pond, which was exposed to environmental temperatures (5–20°C). A control culture without inoculum was included, and all experiments were conducted aseptically to prevent cross-contamination (Figure 6.1). In pursuit of enzymes capable of functioning at even higher temperatures, one additional enrichment was operated at 75°C.

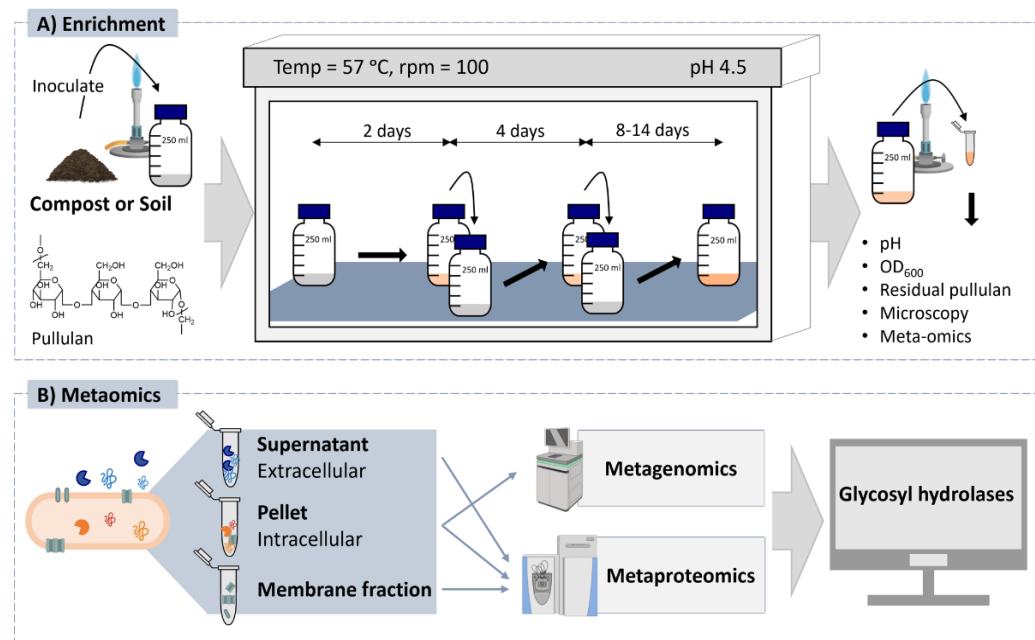
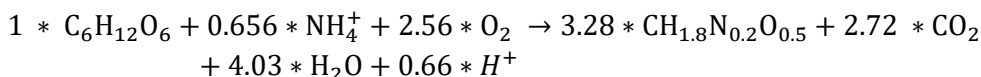


Figure 6.1. Employed workflow for the discovery of novel glycoside hydrolases from microbial enrichments. A) Enrichment cultures at 57°C and pH 4.5 were established by inoculating either with compost (elevated temperature microbial source) or soil (temperature source control), using pullulan as the sole carbon substrate. In addition, growth medium without inoculum served as a control for cross contamination. The pH, OD₆₀₀ and residual pullulan were monitored daily. The enrichments were studied by microscopy and once all pullulan was depleted the biomass and the secretome were analysed by metagenomics and metaproteomics. **B)** The enrichment biomass was subjected to whole metagenome sequencing. Additionally, metaproteomics analysis was performed on both the cell biomass and supernatant fractions (secretome). While not utilized in this study, future metaproteomic investigations could include a separate analysis of the membrane fraction, offering deeper insights into substrate degradation and uptake pathways, by identifying dedicated transporters. The subsequent functional annotation pipeline was tailored to identify potential glycoside hydrolase candidates.

Two days after inoculation, microbial growth was confirmed by microscopy and an increase in OD₆₀₀ in both the ‘compost’ and ‘soil’ enrichment, while absent in the control (Figure 6.2A & 6.2B). In addition, no growth was identified for the enrichment at higher temperature (75°C), which was subsequently terminated. Both the ‘compost’ and ‘soil’ enrichments were observed to acidify over time (Figure 6.2B). This required daily pH adjustments with KOH, since no organic buffer was used to prevent it from becoming a substrate for the microorganisms. Assuming the following biomass growth stoichiometry,



acidification was likely the result of ammonium uptake (see SI document). Moreover, HPLC-UV analysis using an Aminex organic acid analysis column showed no additional peaks indicative of organic acids, excluding the possibility of acid production during the enrichment (SI DOC Figure S6.1). To confirm pullulan utilization, the residual pullulan in the medium was hydrolysed to glucose and analysed by HPLC. The glucose levels decreased over time in the 'compost' and 'soil' enrichment, but remained constant in the control, verifying the consumption of pullulan by microorganism (SI DOC Figure S6.2).

6.3.2. *Alicyclobacillus* identified as the dominant genus in both enrichments.

The inocula and the biomass from the enrichments were analyzed using whole metagenome sequencing and metaproteomics. Interestingly, the compost inoculum was found to contain a complex spectrum of microorganism based on a taxonomic classification of the metagenomics data. In total, the classification of obtained open reading frames (ORFs) yielded more than 100 different genera, including *Geofilum*, *Capillibacterium*, *Methanoculleus*, and *Methanosarcina* (SI DOC Figure S6.3). Over 80% of the identified genes could not be classified using this method, pointing to a significant proportion of yet undescribed microbes. This finding highlights the vast diversity and novelty of microorganisms in the inoculum. After enrichment on pullulan at 57°C and pH 4.5, the complexity in both enrichments reduced to one dominant genus, namely *Alicyclobacillus* (Figure 6.2C and SI DOC Figure S6.4). Thereby indicating that despite the different origin of the inocula, the imposed conditions could select for the same microbe.

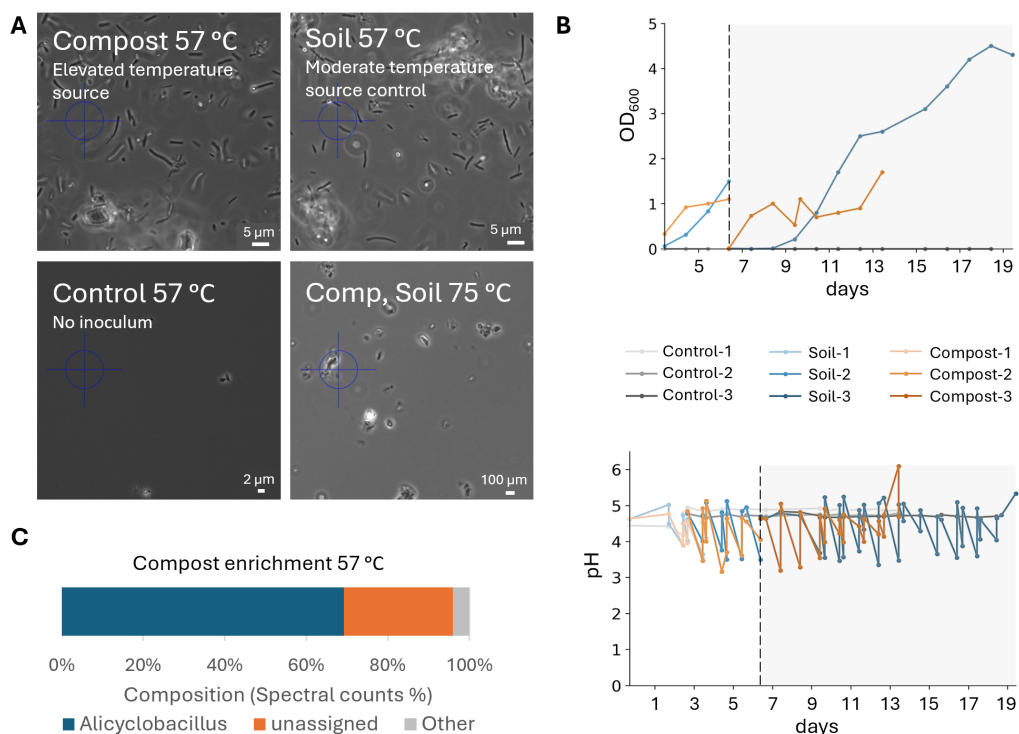


Figure 6.2. A) Microscopy images of the enrichment cultures with a 100 x magnification. Compost (elevated temperature microbial source): inoculated with organic matter of a compost pile. Soil (temperature source control): inoculated with sediments from a pond. Control: No inoculum. Comp, Soil 75°C: inoculated with both compost and soil, incubated at 75°C. All images were taken two days after inoculation. **B)** pH and OD₆₀₀ measurements during the compost, soil and control enrichment cultures. Grey area: period after the second transfer to fresh medium. **C)** Taxonomic composition (genus level) based on spectral counts (metaproteomics) of the 'Compost' enrichment biomass, normalized to 100%. Approx. 70% of the counts could be assigned to the genus *Alicyclobacillus*. The legend shows genera with >10 spectral counts (for full annotation see SI DOC Figure S6.4).

6.3.3. Identification of new glycoside hydrolases from *Alicyclobacillus*.

To identify potential glycoside hydrolases which enable *Alicyclobacillus* to grow efficiently on pullulan, we established a pipeline which combines metagenomic functional annotation with metaproteomic expression and localization data (Figure 6.3A). In short, the pipeline annotates proteins using the InterPro database, with a local InterProScan installation, providing an overview of protein families, domains, and sites for each gene. The annotated data was subjected to a keyword search to identify potential enzymes related to hydrolytic activity and pullulan degradation. Keywords included "family 13," "GH13," "family 57," "GH57," "amylase," "glycosidase," "glycoside hydrolase," and "pullulan". Unannotated genes were also retained in the final list to allow identification of completely novel sequences. The

subsequent integration of metaproteomics data provided information on the expression levels and localization of these proteins.

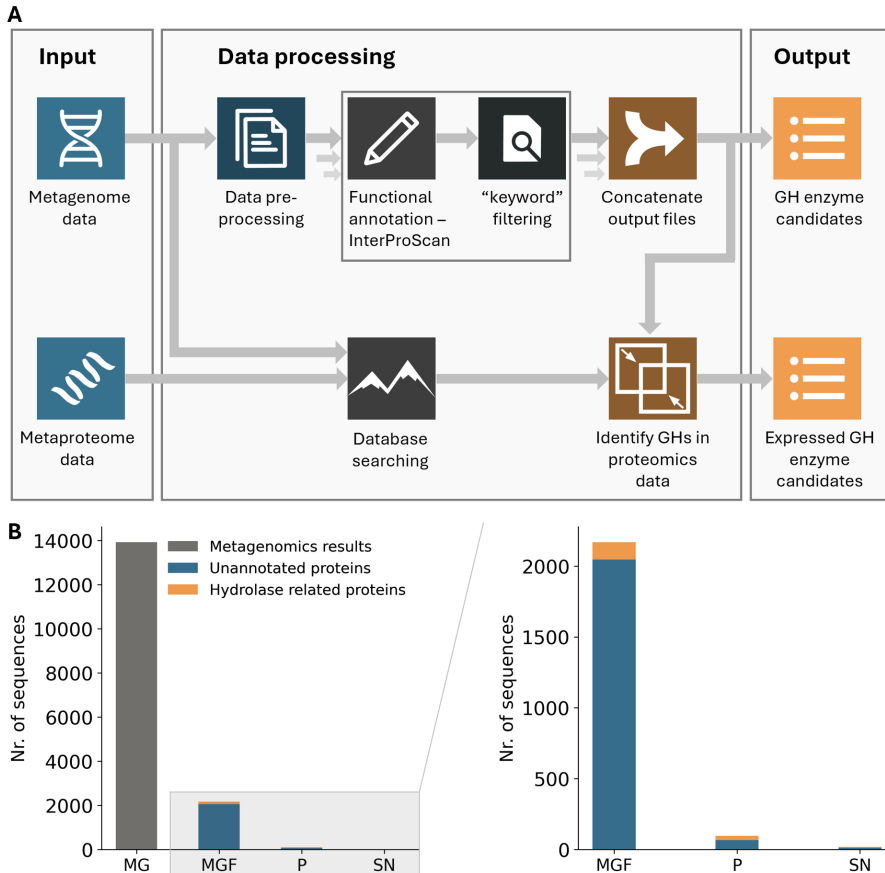


Figure 6.3. A) Schematic overview of the employed bioinformatics pipeline. Functional annotation of the whole metagenome sequencing data was performed using InterProScan. Functional entries containing hydrolase-related keywords were flagged as potential glycoside hydrolases. The annotated database was subsequently employed to analyse the metaproteomics data by database searching, which ultimately identified a focused number of glycoside hydrolase (GH) candidates that were confirmed to be expressed. **B)** The number of pullulan-degrading enzyme candidates obtained from the metaproteomics approach is compared to the total number of sequences obtained from the whole metagenome sequencing data. While metagenomic datasets typically encompass thousands of genes, combining this data with comprehensive functional classification and metaproteomics experiments on both cell pellet and supernatant fractions narrows the focus to a specific set of glycoside hydrolase candidates. In the case of pullulan, this approach also confirmed a two-step degradation process occurring both extracellularly and intracellularly. **MG:** number of genes obtained from the ‘compost’ enrichment by whole metagenome sequencing, **MGF:** number of GH candidates based on comprehensive functional classification of metagenomics data, **P:** number of GH candidates based on metaproteomics on the cell pellet, **SN:** number of GH candidates based on metaproteomics on the supernatant fraction (secreted proteins).

Whole metagenomic sequencing of the ‘compost’ enrichment yielded ~14,000 genes. Applying the functional annotation and keyword filtering this could be narrowed down to ~2,000 genes. However, validating the activity of each gene would remain a costly and time-consuming process. Advantageously, integrating metaproteomics data reveals which enzymes are actively expressed. Because pullulan degradation is essential for growth, relevant enzymes were expected to be highly abundant. Moreover, protein localization was inferred by analyzing supernatant and biomass pellet fractions separately, thereby distinguishing secreted from intracellular enzymes. This especially supports the identification of enzymes that are involved in the degradation of the carbon source before cellular uptake³⁹. Remarkably, for the compost enrichment, integrating the metaproteomics data reduced the number of enzyme candidates to 96 intracellular and 17 extracellular (Table 6.1). Therefore, employing metaproteomics significantly narrowed down the pool of target enzyme candidates. This greatly streamlines subsequent functional and structural investigations, as well as biocatalytic testing after recombinant expression.

6

Table 6.1. Identified pullulan degrading enzyme candidates in the supernatant (secretome) of the compost enrichment using metaproteomics. InterProScan annotations; IPR06047: Glycoside hydrolase, family 13, catalytic domain; IPR004185: Glycoside hydrolase, family 13, N-terminal Ig-like domain; IPR017853: Glycoside hydrolase superfamily; IPR015020: Rv2525c-like, glycoside hydrolase-like domain; IPR015955: Lactate dehydrogenase/glycoside hydrolase, family 4, C-terminal (homologous superfamily). The bars show the spectral counts based abundance of the identified proteins in the supernatant fraction (secreted proteins).

Protein group	Accession nr.	MW	InterPro	Spectral counts (abundance)
Glycoside hydrolase domain	MMBJNONL_14124	140550	IPR06047, IPR004185, IPR017853	28
	MMBJNONL_07072	91271	IPR06047, IPR004185, IPR017853	22
Glycoside hydrolase superfamily	MMBJNONL_03000	75891	IPR015020, IPR017853	2
	MMBJNONL_05960	58192	IPR017853	2
Lactate dehydrogenase/glycoside hydrolase superfamily	MMBJNONL_09916	33232	IPR015955	2
	MMBJNONL_07048	33418	IPR015955	2
Proteins without any functional annotation by InterProScan	MMBJNONL_05225	21160	-	8
	MMBJNONL_03006	18307	-	7
	MMBJNONL_05223	27924	-	4
	MMBJNONL_09075	32173	-	2
	MMBJNONL_01486	47258	-	2
	MMBJNONL_05876	18086	-	4
	MMBJNONL_03029	18064	-	4
	MMBJNONL_02580	15965	-	2
	MMBJNONL_02999	13289	-	2
	MMBJNONL_07823	21122	-	2
	MMBJNONL_01632	28355	-	3



From the 17 extracellular candidates, six were annotated by InterPro, of which two (MMBJNONL_14124 and MMBJNONL_07072) were identified as containing a domain from the glycoside hydrolase family 13 (GH13). These are the most likely candidates possessing pullulan degrading activity, since this family includes enzymes such as alpha-amylases and neopullulanases. Some expressed proteins lacked a functional annotations and would also require further activity testing. The InterProScan annotation of the final extracellular pullulan-degrading enzyme candidates was further validated by dbCAN3⁴⁰ (SI DOC Table S6.2, Figure S6.6 and SI Excel). Both GH13 candidates were confirmed, and MMBJNONL_05960 was further classified by dbCAN3 as belonging to glycoside hydrolase family 53. Finally, MMBJNONL_09916 and MMBJNONL_07048 were annotated by InterProScan, but not recognized by dbCAN3. Both were assigned the domain IPR015955 and annotated as "lactate dehydrogenase/glycoside hydrolase". If these enzymes primarily function as lactate dehydrogenase, it might explain their absence in the dbCAN3 annotation. Overall, the annotation results are largely consistent between the two tools and support the hypothesis that MMBJNONL_14124 and MMBJNONL_07072 are likely responsible for pullulan degradation.

Matzke and co-workers previously identified two pullulan-degrading enzymes in *Alicyclobacillus*: the cytoplasmic protein cyclomaltodextrinase (CdaA), which hydrolyzes 1,4-linkages of pullulan to form panose, and amylopullulanase (AmyA), which hydrolyzes 1,6-linkages of pullulan to form maltotriose⁴¹. AmyA likely hydrolyses pullulan into maltotriose extracellular, which can be transported over the membrane and is further degraded by CdaA (Figure 6.4A). A sequence alignment of AmyA and CdaA against the metagenomics data of the 'compost' enrichment resulted in six matches, the best showing only 80% sequence identity (Figure 6.4B). Moreover, the matches with high sequence identity to AmyA (MMBJNONL_07072 and MMBJNONL_14124, Figure 6.4B) were identified as extracellular by the enzyme discovery approach, while the best match with CdaA (MMBJNONL_12997 in Figure 6.4B) was classified as intracellular. These cellular localizations were further supported by SignalP-6.0⁴² and DeepLocPro⁴³ predictions (SI DOC Table S6.1). Besides the 80% sequence identity, the same domains were identified at similar locations, being CBM34 and GH13_39 for MMBJNONL_07072, MMBJNONL_14124 and AmyA and GH13_20 for MMBJNONL_12997 and CdaA (SI DOC Figure S6.6). Pullulan degrading enzymes are known to contain four conserved sequence regions^{44,45}, all of these have been identified in MMBJNONL_12997, while three have been identified in MMBJNONL_07072 and MMBJNONL_14124 (SI DOC Figure S6.7 and S6.8), further supporting the pullulan degrading activity.

Interestingly, sequence alignment of the AmyA and CdaA genes against the metagenomic data of the compost inoculum yielded matches with only ~45% sequence identity, while the enzymes identified through the employed enrichment metaproteomics approach were not detected at all (Figure 6.4). Although deeper genomic sequencing might detect these enzymes, this underscores the effectiveness of combining microbial enrichment with metaproteomics.

To validate the results of our established meta-omics approach and the associated functional predictions, the top extracellular pullulan-degrading enzyme candidate, MMBJNONL_14124, was recombinantly expressed in *E. coli* with a C-terminal His-tag and purified by affinity chromatography. The purified enzyme was subsequently incubated with 1 mg/mL pullulan at 60°C and pH 4.5, and pullulan hydrolysis was monitored using high-resolution tandem mass spectrometry. Thereby we observed predominant release of triose oligomers (Figure 6.4D), which is aligned with the cleavage of alpha-1,6 glycosidic bonds and release of maltotriose units from pullulan. This confirms the AmyA like pullulan-hydrolytic activity of MMBJNONL_14124 under elevated temperature and acidic pH conditions.

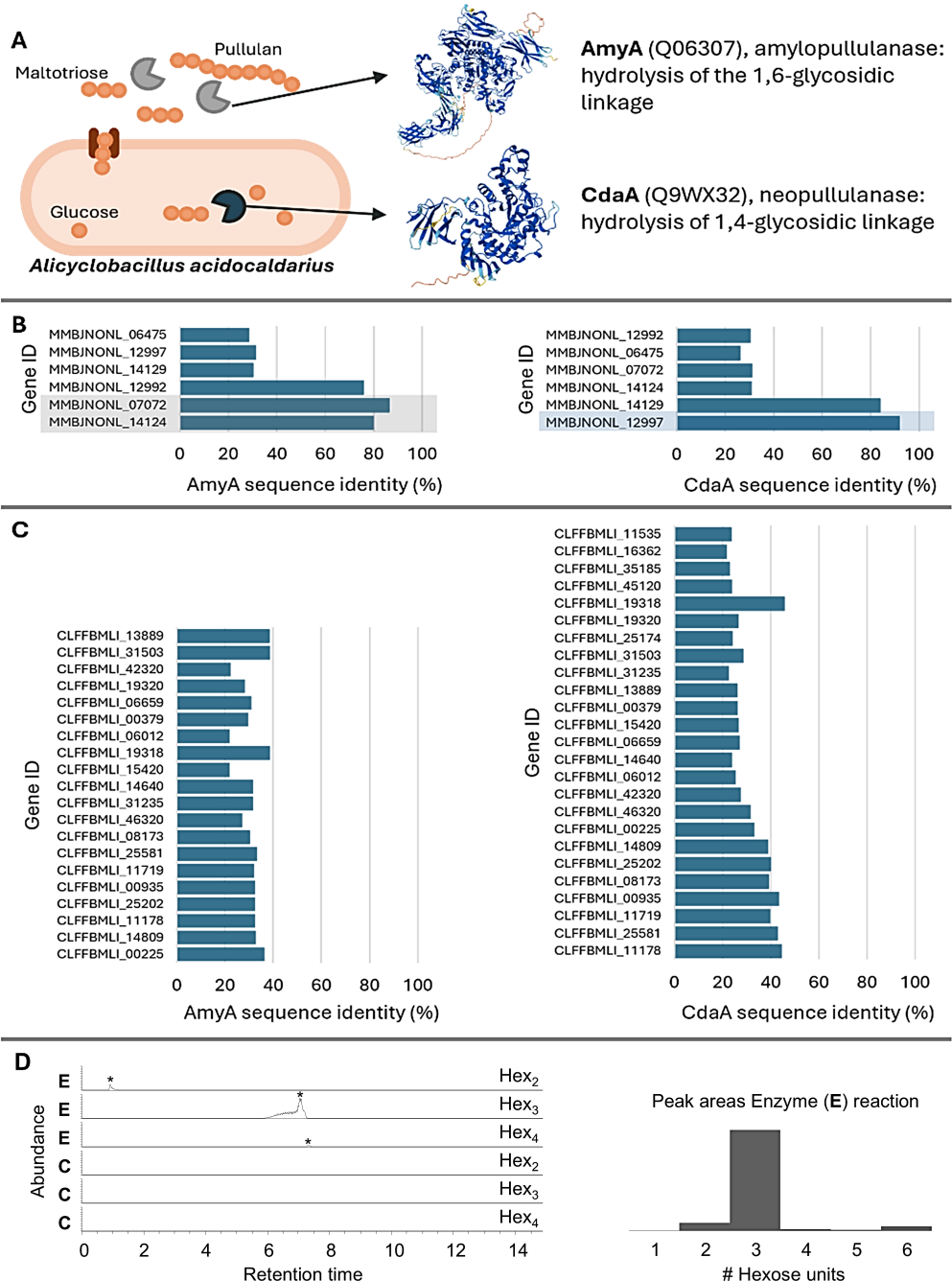


Figure 6.4. A) Proposed pullulan degrading pathway in *Alicyclobacillus* based on the identified enzymes in this study and recent work performed by Matzke et al. (2000) and the enzyme activity assay performed in this study. AmyA hydrolyses the 1,6-glycosidic linkage of pullulan extracellularly. The maltotriose units are then imported, after which CdaA hydrolyses the 1,4-glycosidic linkage to form glucose, which can enter the glycolytic pathway. **B)** Sequence alignment of the AmyA (Q06307) and

CdaA (Q9WX32) gene, respectively, against the metagenomics data of the compost enrichment culture. Genes showing high sequence identity and which were identified with the glycoside hydrolase bioinformatics pipeline are highlighted in grey (extracellular), or in blue (intracellular). The identified genes with high sequence identity (MMBJNONL_12992, MBJNONL_07072, MMBJNONL_14124, MMBJNONL_12997 and MMBJNONL_14129) were not identified in the metagenomics data of the inoculum, which demonstrates the effectiveness of performing microbial enrichment combined with metaproteomics. **C)** Sequence alignment of the AmyA (Q06307) and CdaA (Q9WX32) gene, respectively, against the metagenomics data of the compost inoculum. **D)** Enzyme activity assay using purified MMBJNONL_14124 which was recombinantly expressed in E coli. The release of hexose oligomers from pullulan after incubation with MMBJNONL_14124 at 60°C and pH 4.5 was monitored with PGC chromatographic separation and high resolution mass spectrometric detection. The left traces show the extracted ion chromatograms for Hex₂, Hex₃, and Hex₄ units (*), where the upper three traces show the release after incubation with MMBJNONL_14124 (enzyme reaction, E), and the lower three traces show the control reaction where no enzyme was present (control reaction, C). The control reaction (buffer and pullulan) showed only trace amounts of the respective sugars oligomers. The right histogram shows the peak area distribution for the different hexose oligomer products (Hex₁ to Hex₆), where Hex₃ products dominated.

6.4 | Discussion and Conclusion

Advancements in next-generation sequencing have opened up new avenues for the discovery of novel enzymes from microbial sources²⁸. However, a purely sequence-based approach is time consuming and commonly overlooks enzymes with new sequences, and the alternative function-based approach is costly and requires the availability of high-throughput activity assays². Therefore, methods that narrow down the pool of potential enzyme candidates without losing the ability to find completely novel enzymes are of great advantage. The combined application of enrichments with multi-omics allows to search for completely novel enzymes operating at defined conditions. We exemplified this approach by searching for new pullulan-degrading enzymes capable of operating at elevated temperatures and acidic pH, from two complex microbial sources (i.e. organic matter from compost, and soil from a pond). The enrichment used pullulan as the sole carbon source, ensuring that only microorganism with pullulan-degrading and utilizing capabilities would thrive. In addition to selecting for a desired function, this process allows customization of conditions to match industrial requirements, such as temperature, pH, and salinity. Nevertheless, the enrichment approach is limited to enzymes that are required for cellular growth or survival, such as enzymes involved in carbon or nitrogen source utilization. Following enrichment, the biomass can be analyzed by whole metagenome sequencing and metaproteomics. Whole metagenome sequencing alone identified approximately 14,000 genes in the compost enrichment, which was narrowed down to around 2,000 candidates after functional classification. Integrating metaproteomics data further allowed to focus

on actively expressed enzymes and their specific locations, such as intracellular, membrane-bound, or extracellular. For the compost enrichment, this approach reduced the number of enzyme candidates to 96 intracellular and only 17 secreted enzymes.

Nevertheless, the presence of dead and consequently lysed cells, can result in the identification of intracellular proteins in the supernatant. To assess the amount of lysed cells we compared the identified proteins in the supernatant with the 50 most abundant proteins in the biomass pellet (determined by spectral counts). 18 of these proteins were not at all identified in the secretome, and a significant cytoplasmic contamination is therefore unlikely. In addition, the supernatant only showed traces of a few of the enzymes of the abundant glycolytic pathway (glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate mutase and glucose-6-phosphate isomerase), which further confirms that there was no significant cell lysis taking place. Finally, some proteins (e.g. membrane proteins) are difficult to detect by conventional proteomics approaches, and might be missed without additional optimisation of the sample preparation⁴⁶. To complement experimental data, we used SignalP-6.0⁴² and DeepLocPro⁴³ to predict the cellular localization of the secretome. Of these proteins, 27% was predicted to contain a signal peptide, and 18% contained a signal peptide according to SignalP-6.0 and was predicted to be extracellular by DeepLocPro (SI Excel). Signal peptides not necessarily imply secretion, proteins may be membrane bound, or, in case of Gram-positive bacteria, be retained in the periplasm. Similarly, proteins without a signal peptide can still end up in the secretome⁴². Nevertheless, these predictive tools are useful in validating whether the final pullulan degrading enzyme candidates are indeed extracellular (SI DOC Table S6.1).

Despite *Alicyclobacillus* being identified as the dominant genus, other coexisting organisms, though present at lower abundances, were identified. Therefore, it cannot be excluded that also these organisms are involved in hydrolyzing pullulan, which subsequently can be consumed by multiple microbes. Intriguingly, a third enrichment was initiated at 75°C, but no growth was observed. A review by Kahar et al. highlights diverse pullulan-degrading enzymes identified from microbial sources, demonstrating activity at high temperature and low pH¹³. For instance, pullulanases from organisms like *Fervidobacterium nodosum* Rt17-B1 (optimal temperature, T_{opt} of 80°C and an optimal pH, pH_{opt} of 5), *Desulfurococcus mucosus* DSM 2162T (T_{opt} = 85°C, pH_{opt} = 5), and *Thermoanaerobacter* sp. B6A (T_{opt} = 75°C, pH_{opt} = 5) are some examples. However, enzyme activity at 75°C and pH 4.5 doesn't guarantee that the organism can grow under these conditions. For instance, for *Fervidobacterium nodosum* the growth range is restricted to pH 6.0–8.0. Moreover,

this bacterium, as well as *Desulfurococcus mucosus* and *Thermoanaerobacter*, are obligate anaerobes, preventing growth under the conditions set in this study⁴⁷⁻⁴⁹.

One enrichment was inoculated with compost (“compost” enrichment), which typically has an elevated temperature and is expected to host thermotolerant microbes. To assess whether such conditions are essential, a second enrichment was inoculated using soil from a pond (‘soil’ enrichment). Despite the complex and diverse starting communities, *Alicyclobacillus* dominated in both cases. This bacterium was first isolated from an acid thermal environment in Yellowstone national park by Darland and Brock⁵⁰. It is a spore-forming, rod-shaped bacterium that typically grows at temperatures of 45–70°C (optimum 60–65°C) and pH values of 2–6 (optimum 3–4), closely matching the applied enrichment conditions. Cross-contamination between the enrichments is unlikely, as no growth was observed in the control experiment. Thus, similar organisms thrived under identical enrichment conditions regardless of the inoculum’s initial temperature. The similarity may also be attributed to the proximity of the sampling sites (~20 km), and their shared soil environment. Both pullulanase types I and II commonly degrade starch, a natural plant polymer found in soil, which might support the growth of similar organisms. Although previous studies identified two pullulan-degrading enzymes in *Alicyclobacillus*, AmyA and CdaA, our enrichment metaproteomics approach revealed distinct enzyme sequences with potentially altered properties (Figure 6.4). The four conserved sequence regions of alpha-amylases^{44,45} were identified in the enzyme candidates, along with the same glycosyl hydrolase domains found in AmyA and CdaA. These similarities support the hypothesis that the enzymes exhibit pullulan degrading activity. The identified enzymes were not detected in the whole metagenome sequencing data of the inoculum, highlighting the effectiveness of our approach. Metaproteomics confirmed that the putative AmyA is extracellular, while the putative CdaA is intracellular. This suggests that AmyA hydrolyzes pullulan into maltotriose units extracellularly, which are then imported into the cell and further broken down by CdaA into glucose. Supporting this, we also found a maltose/maltodextrin-binding periplasmic protein and the maltose/maltodextrin transport system MalG being expressed. This aligns with previous findings of a high-affinity, binding-protein-dependent ABC transport system specific for maltose and maltodextrins in *Alicyclobacillus*⁵¹.

Finally, the pullulan-degrading activity of the extracellular enzymes MMBJNONL_14124 was confirmed by recombinant expression followed by activity testing at 60°C and pH 4.5. High-resolution mass spectrometry confirmed the efficient release of predominantly hexotrioses, consistent with the predicted

cleavage of the alpha-1,6 glycosidic bonds and release of maltotriose units. Nevertheless, some larger and smaller oligomers were also detected, which may arise from some promiscuous activity under the tested conditions.

In summary, we demonstrated the combination of a microbial enrichment and metaproteomics for discovering novel microbial glycoside hydrolases, as exemplified by a pullulan enrichment from two different microbial sources. The integration of metaproteomics significantly accelerated enzyme identification in these complex environments, reducing the number of genes requiring subsequent biochemical validation.

6.5 | Data availability

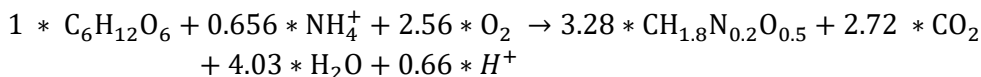
Due to legal confidentiality requirements the genomic and mass spectrometric raw files cannot be made publicly available, but may be provided by the corresponding author upon reasonable request.

6.6 | Supplementary information

6.6.1. Culture acidification.

Over time, the pH of both the ‘compost’ and ‘soil’ enrichments decreased, necessitating daily adjustments with KOH (Figure 6.2B). This acidification was likely a result of ammonium uptake for biomass growth. Initially, each enrichment contained 0.5 g of pullulan (which equals 3.09 mmol of glucose). Throughout the enrichment process, approximately 2.61 ml of 0.2 M KOH (0.52 mmol KOH) was added to maintain the pH at 4.5. Thus, for every mole of glucose consumed, 0.17 mol of KOH was needed to neutralize the produced H^+ , indicating that 0.17 mol of H^+ was generated during growth.

The following biomass growth stoichiometry (normalized to the consumption of 1 mol glucose) is assumed:



6

According to this equation, the consumption of 1 mol of glucose could produce up to 0.66 mol of H^+ . The observed acidification in the culture falls within the theoretically possible range of acidification.

6.6.2. HPLC chromatograms showing the Pullulan depletion.

To confirm that the microorganisms were growing on pullulan, the pullulan depletion was monitored over time by HPLC, both before (Figure S6.1) and after complete hydrolysis (Figure S6.2).

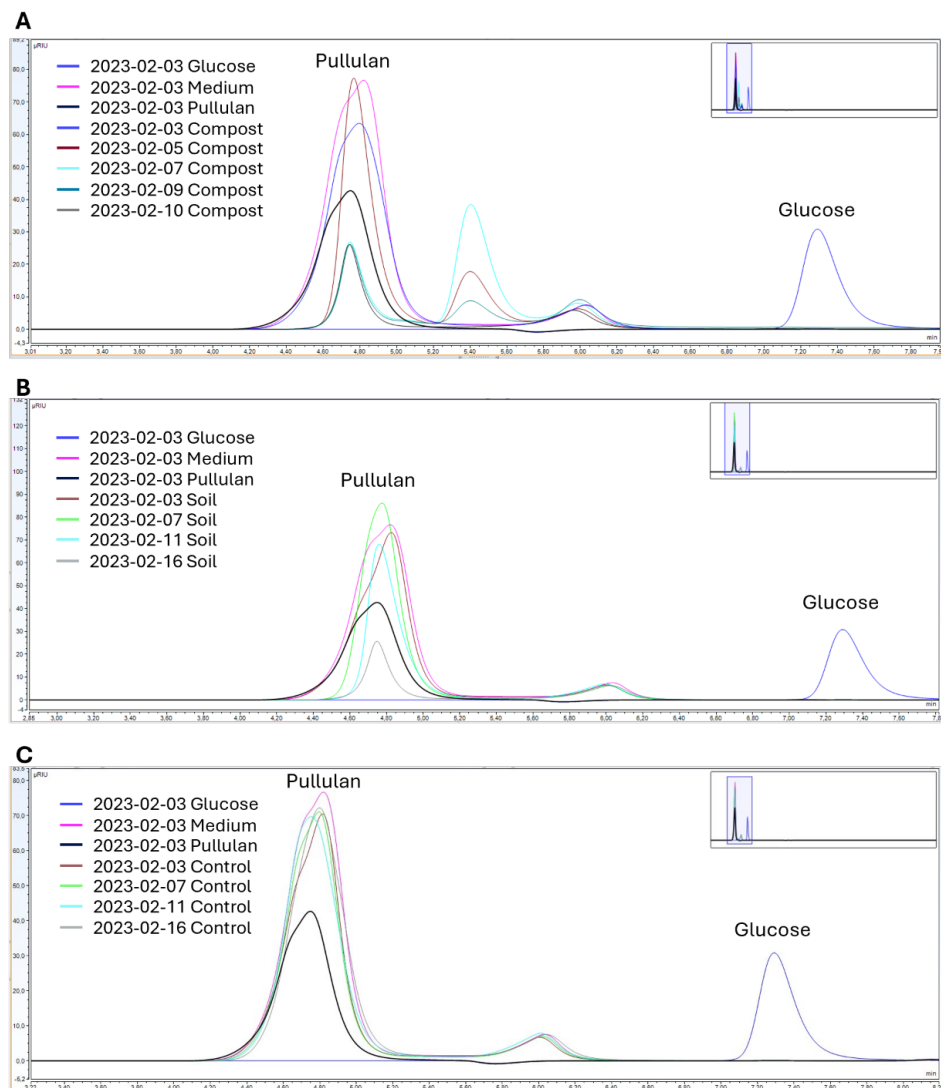


Figure S6.1. Pullulan depletion. Overlaid HPLC chromatograms of the analysed enrichment cultures to follow the pullulan depletion over time. Whole broth samples were centrifuged after which the supernatant was analysed. HPLC conditions are described in the method section. **A)** Compost enrichment (elevated temperature microbial source): inoculated with organic matter of a compost pile. **B)** Soil enrichment (temperature source control): inoculated with sediments from a cold water pond. **C)** Control: No inoculum.

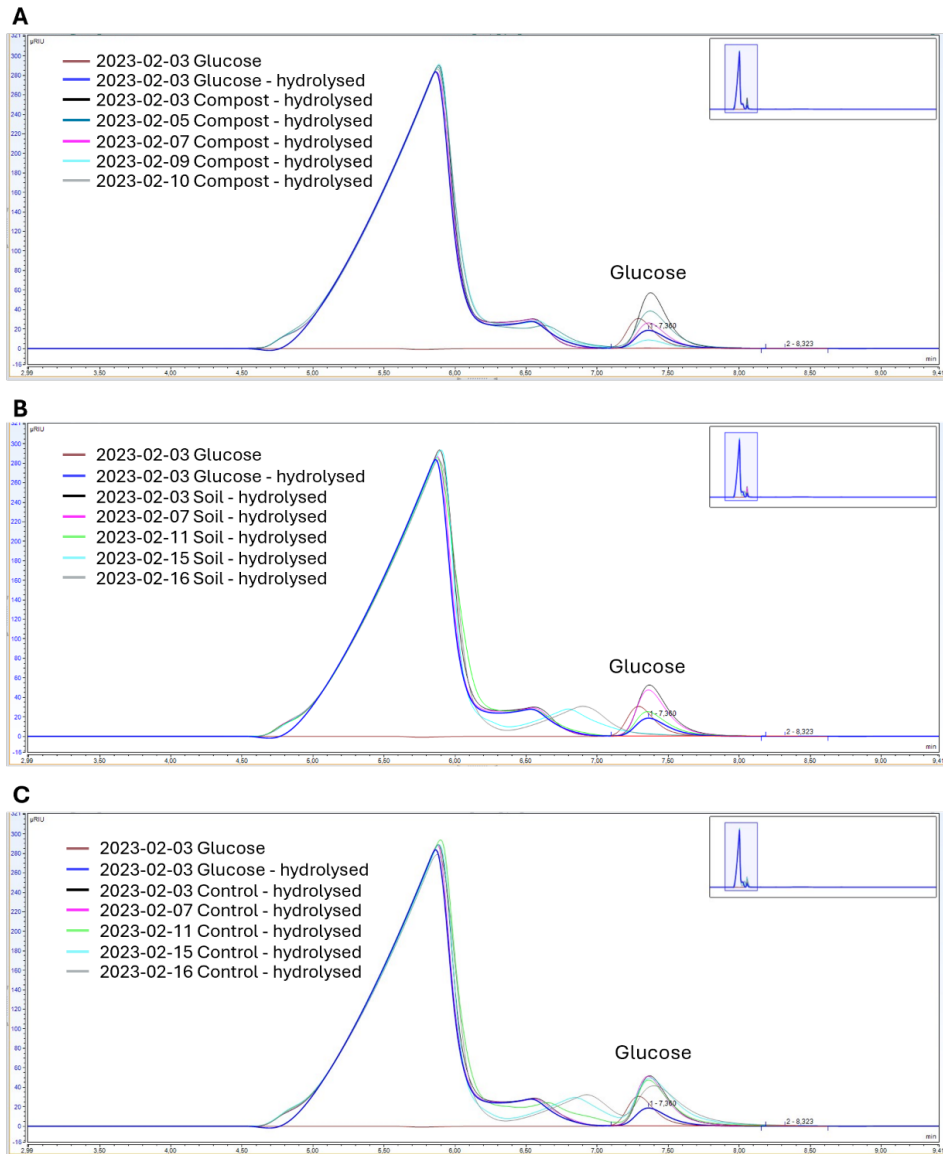


Figure S6.2. Glucose depletion. Overlaid HPLC chromatograms of the analysed enrichment cultures to follow the glucose depletion over time. Whole broth samples were centrifuged, after which the remaining pullulan in the supernatant was hydrolysed by adding trifluoroacetic acid (TFA) to a final concentration of 4M. Samples were incubated at 100 °C for 4 hours. HPLC conditions are described in the method section. **A)** Compost enrichment (elevated temperature microbial source): inoculated with organic matter of a compost pile. **B)** Soil enrichment (temperature source control): inoculated with sediments from a cold water pond. **C)** Control: No inoculum.

6.6.3. Taxonomic composition of the enrichment cultures and the inoculum.

A taxonomic classification of the compost inoculum (Figure S6.3) based on the metagenomics data has been performed. For illustration purposes, only the genera with ≥ 50 gene counts are depicted. The total amount of genera identified by taxonomic profiling equals 421, emphasizing the enormous variety of microorganisms present. A significant number of sequences remains unassigned. Therefore the taxonomic distribution might not be the best representation of the reality and needs to be further optimized. Nevertheless, it does show the enormous variety of microorganisms present in the inoculum.

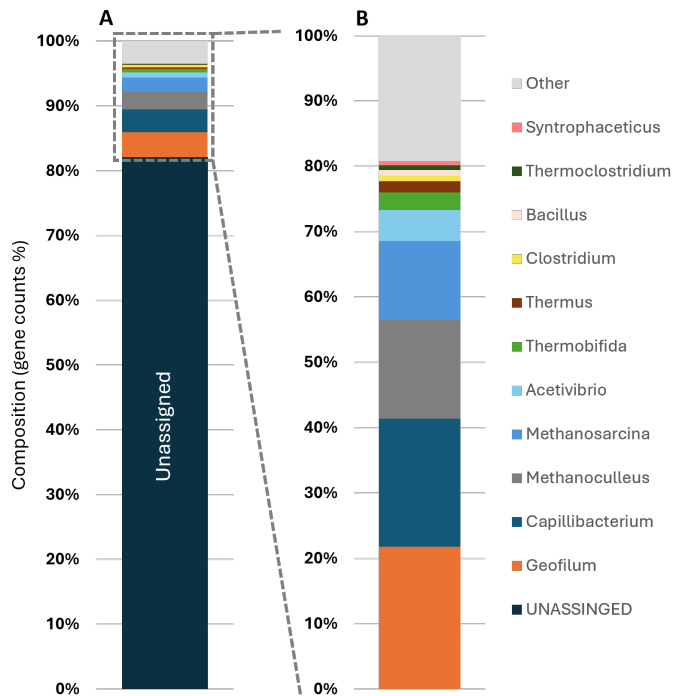


Figure S6.3. Taxonomic composition of the compost inoculum based on gene count (metagenomics data). Taxonomic classification was performed on gene level. **A)** Including the unassigned sequences; **B)** Excluding the unassigned sequences.

In addition, a taxonomic classification of the enrichment cultures was performed based on the metaproteomics data (Figure S6.4).

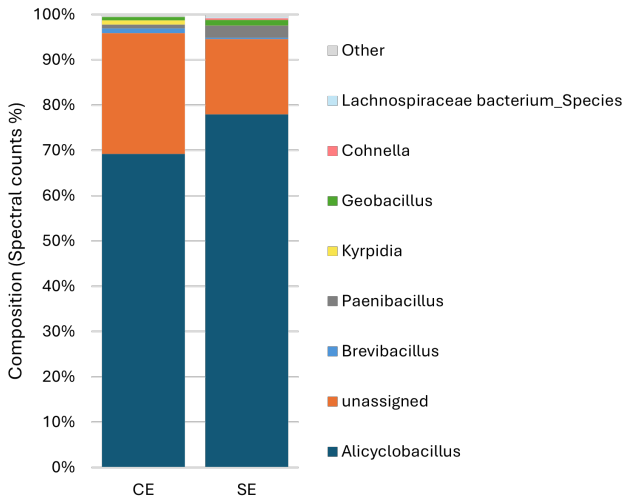


Figure S6.4. Taxonomic composition (genus level) of the ‘Compost’ and ‘Soil’ enrichments based on spectral counts (metaproteomics), normalized to 100%. **CE:** Compost enrichment; **SE:** soil enrichment.

6.6.4. Optimal temperature and pH of pullulan degrading enzymes.

To provide an overview of the optimal temperature and pH of known pullulan degrading enzymes, the data obtained from the review by Kahar et al. (2022)¹³ has been plotted in Figure S6.5.

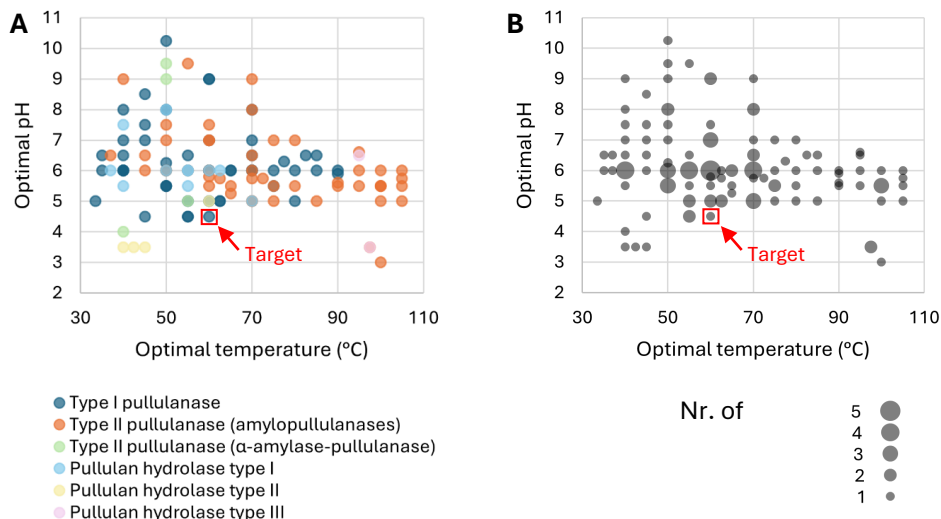


Figure S6.5. Optimal temperature and pH of pullulan degrading enzymes. The data is obtained from the review by Kahar et al. (2022)¹³. **A)** Optimal temperature and pH of pullulan degrading enzymes categorized per enzyme group. Each dot represents a single enzyme. **B)** Optimal temperature and pH of pullulan degrading enzymes illustrated as a dot plot. The size of the dots represent the number of enzymes with a specified optimal temperature and pH. The target space represent the enrichment culture conditions of this work (60 °C at pH 4.5).

6.6.5. Predicting cellular localization using SignalP and DeepLocPro.

The final pullulan degrading enzyme candidates MMBJNONL_14124, MMBJNONL_07072 and MMBJNONL_12997 (Figure 6.4) were subjected to SignalP-6.0^{42,45} and DeepLocPro⁴³ to predict the cellular localization of the enzymes and validate the cellular localization determined with proteomics (Table S6.1).

Table S6.1. Prediction of the cellular localization of the final pullulan degrading enzyme candidates using the bioinformatics tools SignalP-6.0⁴² and DeepLocPro⁴³.

Accession number	DeepLocPro localization (probability)	SignalP-6.0 (probability)
MMBJNONL_14124	Extracellular (0.9240)	Yes (0.8539)
MMBJNONL_07072	Extracellular (0.8952)	Yes (0.7402)
MMBJNONL_12997	Cytoplasmic (0.9973)	No (0.9998)

6

6.6.6. dbCAN3 annotation of the final pullulan degrading enzyme candidates.

The final pullulan degrading enzyme candidates of the compost enrichment have been subjected to an annotation by dbCAN3⁴⁰ to validate the identification as CAZymes (Table S6.2, Figure S6.6). The detailed output of the analysis is provided in the SI Excel file 1.

Table S6.2. Overview dbCAN3 results. dbCAN3 annotation⁴⁰ has been performed on the final pullulan degrading enzyme candidates identified in the supernatant of the compost enrichment (see Table 6.1). EC 3.2.1.1: alpha-amylase; EC 3.2.1.41: Pullulanase; CMB34: Carbohydrate-Binding Module Family 34; GH13_39: Glycoside Hydrolase Family / Subfamily 39; GH53: Glycoside Hydrolase Family 53.

Gene ID	EC#	HMMER	dbCAN_sub	DIAMOND	SignalP	# of Tools
MMBJNONL_14124	3.2.1.1 3.2.1.41	CBM34(395-538)+ GH13_39(621-970)	CBM34_e14+CBM34_e14+ GH13_e172	CBM34+GH13_39	Y(1-53)	3
MMBJNONL_07072	3.2.1.1 3.2.1.41	CBM34(394-538)+ GH13_39(621-842)	CBM34_e14+CBM34_e14+ GH13_e172	CBM34+GH13_39	Y(1-52)	3
MMBJNONL_05960	-	GH53(187-524)	GH53_e21	N	Y(1-33)	2

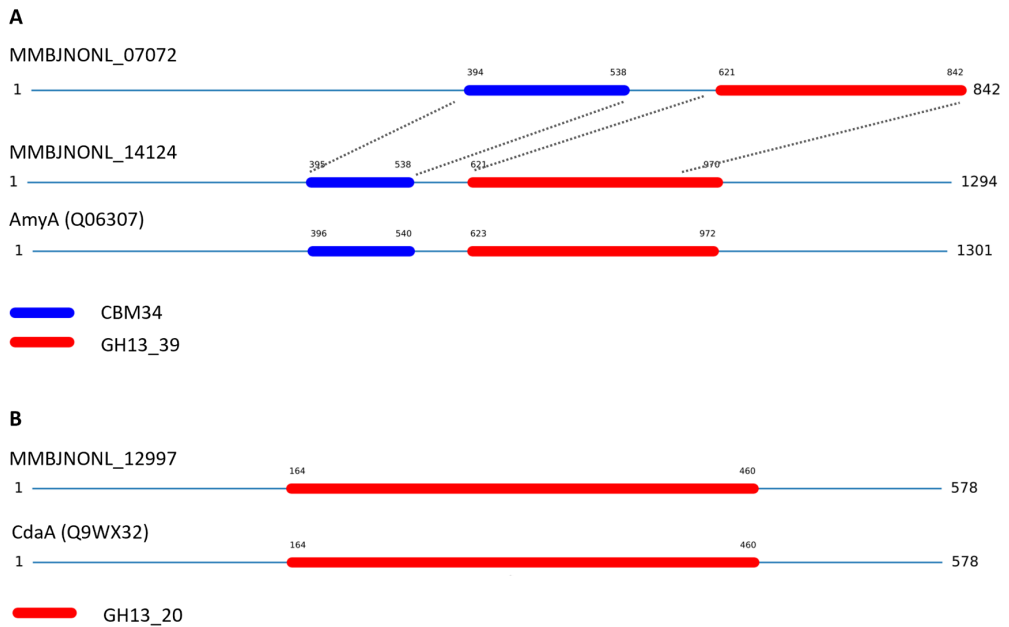


Figure S6.6. Sequence annotation created with dbCAN3⁴⁰. **A)** Sequence annotation of AmyA and the predicted extracellular pullulan degrading enzymes MMBJNONL_07072 and MMBJNONL_14124. **B)** Sequence annotation of CdaA and the predicted intracellular pullulan degrading enzyme MMBJNONL_12997.

6.6.7. Multiple sequence alignment.

A multiple sequence alignment was performed both with a set of neopullulanases and the predicted intracellular pullulan degrading enzyme of this study and with a set of amylopullulanases and the predicted extracellular pullulan degrading enzymes of this study. The alignment was performed using the Constraint-based Multiple Alignment Tool (COBALT) of NCBI⁵². α -amylases are known to contain four conserved sequence regions, which are annotated in the alignments (Figure S6.7 and S6.8)^{44,45}.

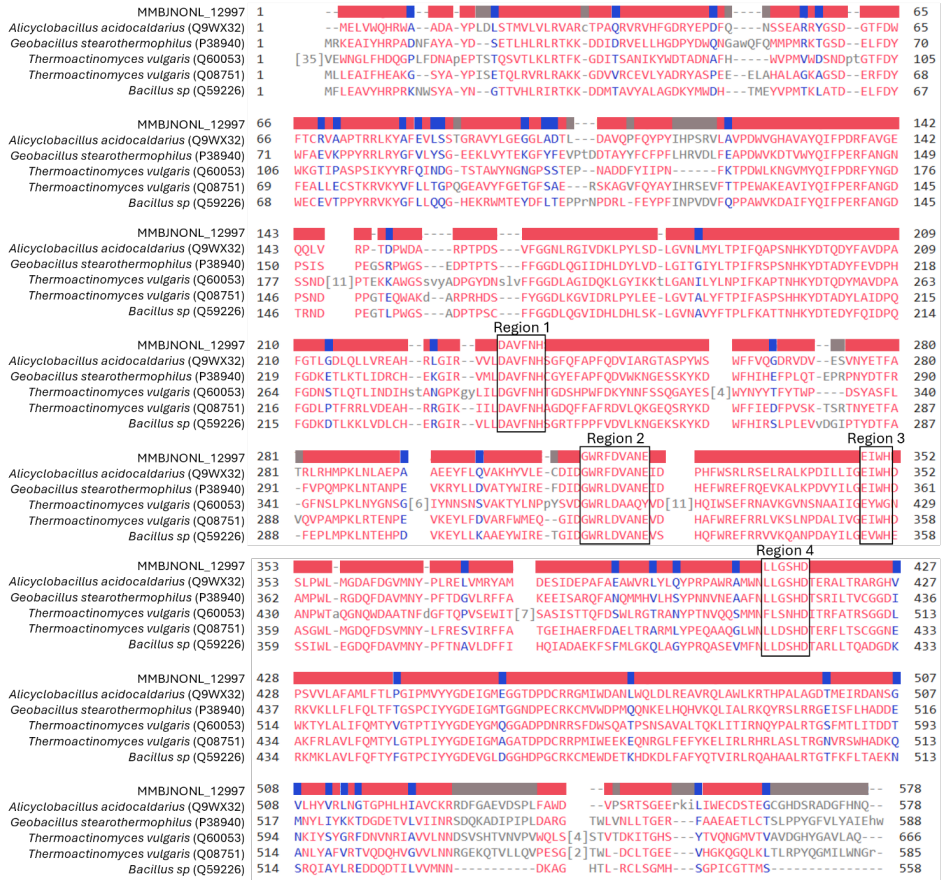


Figure S6.7. Multiple sequence alignment of neopullulanases performed with the Constraint-based Multiple Alignment Tool (COBALT) from NCBI⁵², using the default parameters. In red are the highly conserved amino acids, while in blue the lesser conserved amino acids. The top four conserved regions of alpha-amylases are outlined with a black box^{44,45}. Uniprot identifiers of the used sequences are provided between brackets.

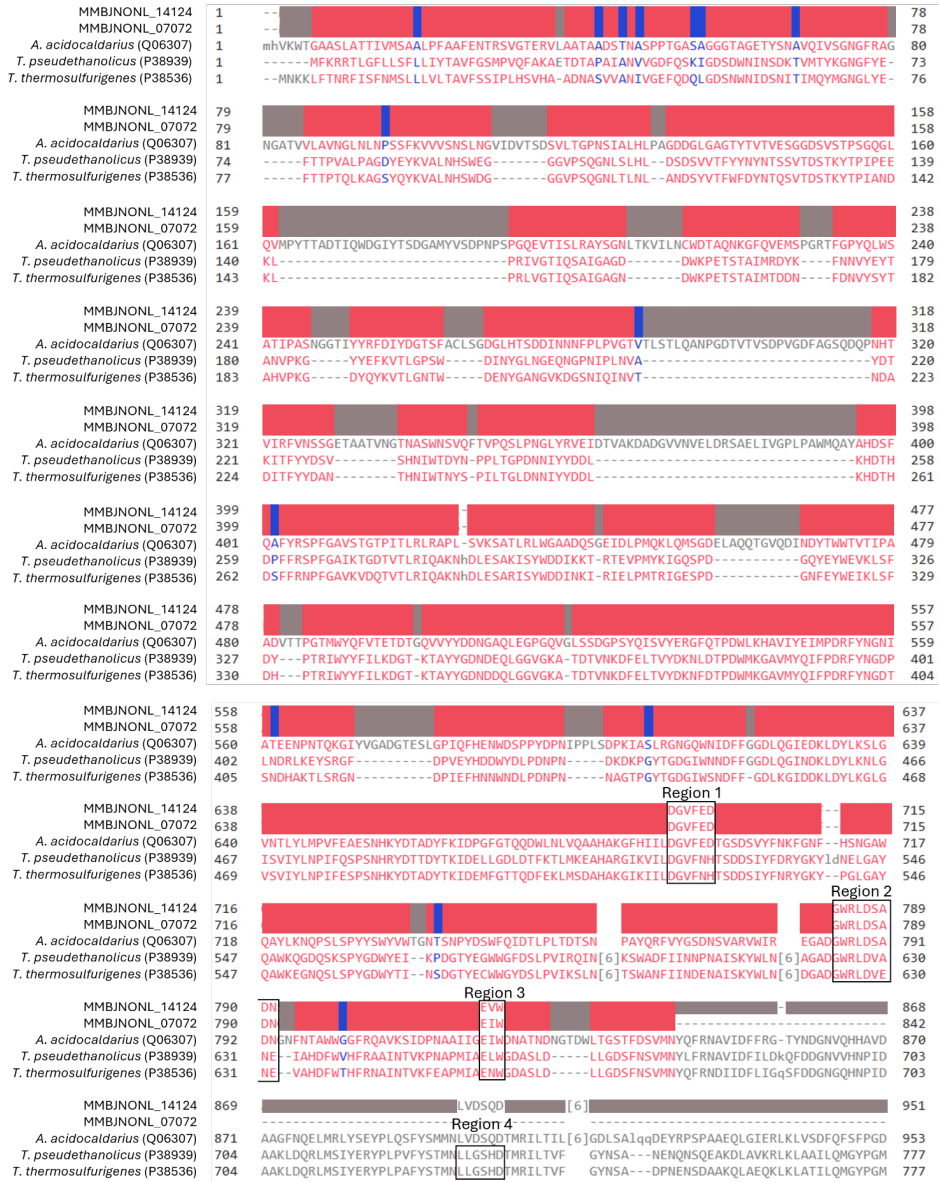


Figure S6.8. Multiple sequence alignment of amylopullulanases performed with the Constraint-based Multiple Alignment Tool (COBALT) from NCBI⁵², using the default parameters. In red are the highly conserved amino acids, while in blue the lesser conserved amino acids. The known four conserved regions of alpha-amylases are outlined with a black box^{44,45}. Uniprot identifiers of the used sequences are provided between brackets. *A. acidocaldarius* = *Alicyclobacillus acidocaldarius*; *T. pseudethanolicus* = *Thermoanaerobacter pseudethanolicus*; *T. thermosulfurigenes* = *Thermoanaerobacterium thermosulfurigenes*.

Figure S6.8 Continued

MMBJNONL_14124	952		1031
MMBJNONL_07072			
<i>A. acidocaldarius</i> (Q06307)	954	PTIFYGDEAGLTGYSDPLNRRTPWDNQNLDDLHNYRKLGAIRNANPVLQTGDFTPLYAQGMVYAFARTIRNGRDVFGVP	1033
<i>T. pseudethanolicus</i> (P38939)	778	PSIYYGDEAGQSGGKDPNRRTFSWGREDKLDQDFKKVVNIIRNENQVLKTGDLETLYANGDVYAFGRRIINGKDVFGNS	857
<i>T. thermosulfurigenes</i> (P38536)	778	ADIYYGDEAGVSGGKDPDRRTFPWGNEDTTLQDFFKNISSIRNNNQVLKTGDLETLYAQNDDVYAGRRRIINGKDAFGTS	857
MMBJNONL_14124	1032		1107
MMBJNONL_07072			
<i>A. acidocaldarius</i> (Q06307)	1034	AEDATAIVAINNQQAITVTIPTDGTVDGSTMLDELNNQ-WYKVQNGGITLTLQSYQGAILVT	1109
<i>T. pseudethanolicus</i> (P38939)	858	YPSVVAIVVINKGEA-KSVQIDTTKFVRDGVAFDALS GK-TYTVRDGQIVVEVVALDGAILIS[8]PQPITDLKAVSGN	940
<i>T. thermosulfurigenes</i> (P38536)	858	YPSAAIVAINRSKSDKQIAIDTTKFLRDGVTFKDLINNvSYSISNGQIVIDVPAMSGVMLIS[8]PQAPSINVVTSGN	942
MMBJNONL_14124	1108		1186
MMBJNONL_07072			
<i>A. acidocaldarius</i> (Q06307)	1110	SQNETAWTPVQGAIGYRVWRQNPNGQWVP[4]LPATDL SVTVERDAYaQTFAVQALFSASDHAQSPVSApkTVSLAVDVP	1190
<i>T. pseudethanolicus</i> (P38939)	941	GQVDLSWSA VDRAVSYNIYRSTVKGLYE	1015
<i>T. thermosulfurigenes</i> (P38536)	943	GKVDLSWLQSDGATGYNIYRSSVEGLYE	1017
MMBJNONL_14124	1187		1266
MMBJNONL_07072			
<i>A. acidocaldarius</i> (Q06307)	1191	AVRLSQPIVSGRVVGDRAVSIPTVSGATQYVIYQRQGDGSYAPVATVTS GDSAAIGVPAQGPANS PHATIRVTvpVP	1270
<i>T. pseudethanolicus</i> (P38939)	1016	AFSIGWAGNMNQV-DTHVIGVNNPVE-VYAEIWAELTDKPGQGENMIAQLGYRYIG---DGGQDATRNKVEGVE--IN	1087
<i>T. thermosulfurigenes</i> (P38536)	1018	AYPIGWVGNLTQVSDNHIIGVDKPTEDIYAEVWADGLTNSTGQGNMIAQLGKYKYVSGTVYDVS VYNSVYGV D---D	1094
MMBJNONL_14124	1267		1294
MMBJNONL_07072			
<i>A. acidocaldarius</i> (Q06307)	1271	AGFSSVTYRVAAQNEDGQAVTNPLTLSLSKK	1301
<i>T. pseudethanolicus</i> (P38939)	1088	KDWTWVDARYVGD SGNNDKYMAKFVPMVGT[363]	1481
<i>T. thermosulfurigenes</i> (P38536)	1095	SGFTWVNAQYVGDIGNNDQYKASFTPKIGQ[736]	1861

6.6.8. Expression and purification of MMBJNONL_14124 in *Rosetta2 DE3*.

MMBJNONL_14124 was expressed with a C-terminal His-tag in *Rosetta2 DE3* by Trenzylase. SDS-PAGE and western blot confirmed the successful soluble expression of the enzyme (Figure S6.9).

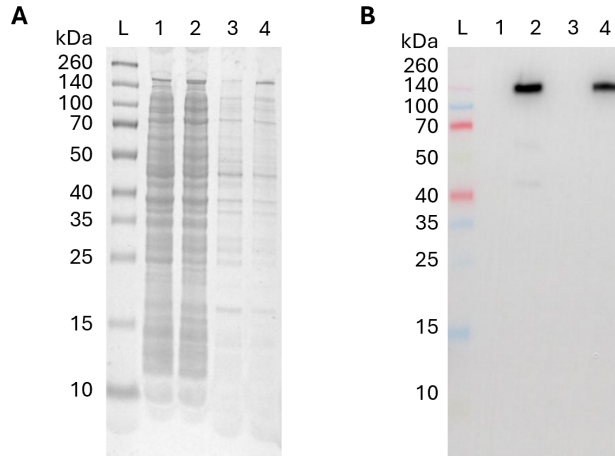


Figure S6.9. **A)** 12% SDS-PAGE, Coomassie staining. **B)** 12% SDS-PAGE and anti-His-HRP (1:2000) Western blot (Standard detection; exposure time: 2 sec). The lanes for both gels were loaded: **L)** protein ladder; **1)** pre induction, soluble; **2)** post induction, soluble; **3)** pre induction, insoluble; **4)** post induction, insoluble.

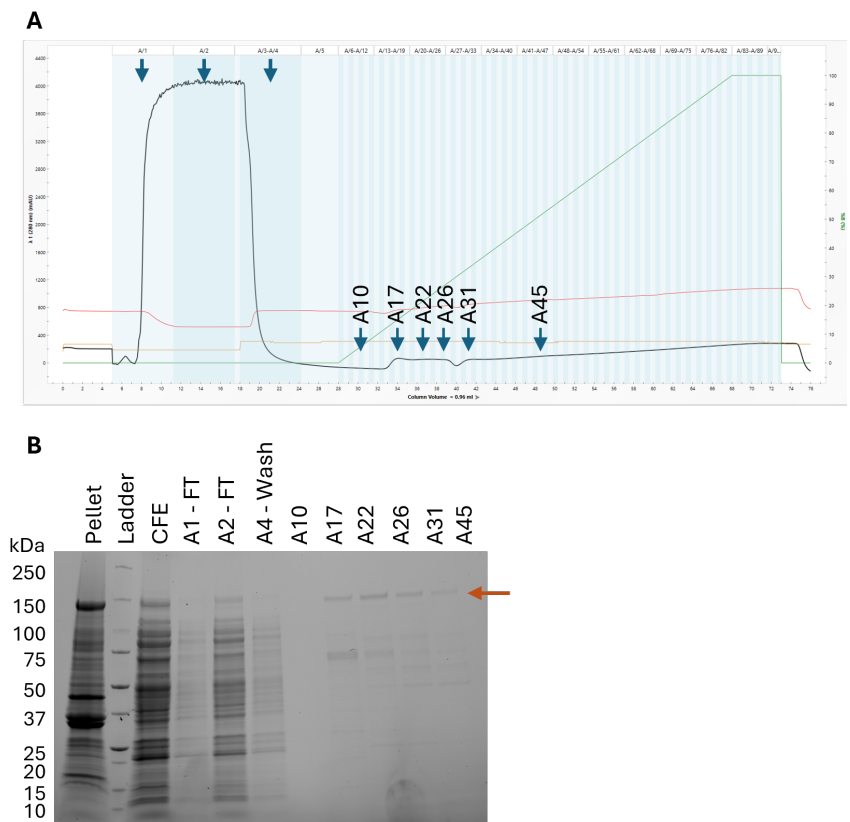


Figure S6.10. A) UV chromatogram trace from MMBJNONL_14124 Ni²⁺ affinity purification. **B)** SDS-PAGE (stain free) of Ni²⁺ affinity purification fractions of MMBJNONL_14124. The lanes were loaded: **1)** Cell pellet after lysis; **2)** protein ladder; **3)** cell free extract; **4)** Flow through fraction A1; **5)** Flow through fraction A2; **6)** Wash fraction A4; **7)** Fraction A10; **8)** Fraction A17; **9)** Fraction A22; **10)** Fraction A26; **11)** Fraction A31; **12)** Fraction A45. The fraction correspond with the arrows in **A)**. The molecular weight of MMBJNONL_14124 equals 138.9 kDa and has been identified in the purified fractions A17, A22, A26 and A31. The orange arrow indicates the protein MMBJNONL_14124.

6.6.9. PGC-MS/MS based pullulan degrading activity assay.

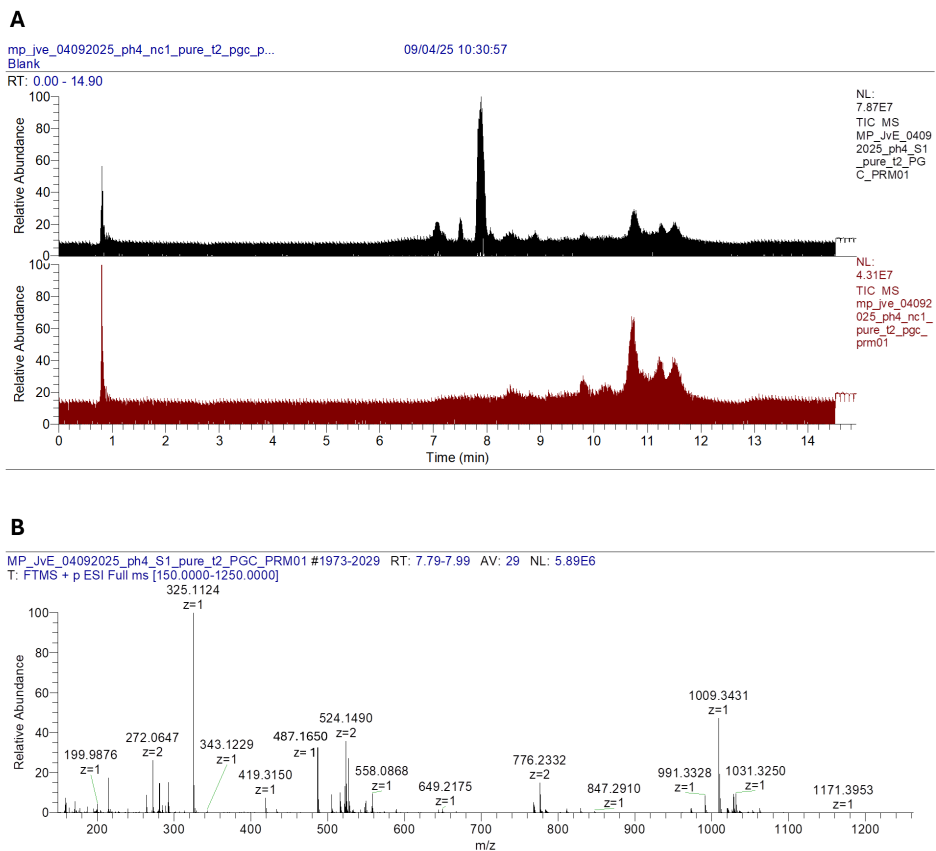


Figure S6.11. A) Total Ion Chromatogram (TIC) of the activity assay with MMBJNONL_14124 (upper panel, black) and the control without MMBJNONL_14124 (lower panel, red). **B)** Full mass spectrum over the retention time range 7.79-7.99 min of the activity assay with MMBJNONL_14124.

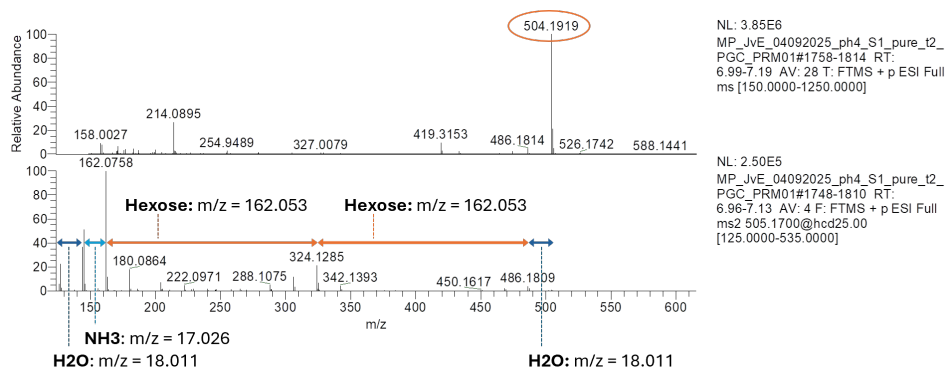


Figure S6.12. Upper panel) Full MS spectrum over the retention time 6.99-7.19 for the activity assay with MMBJNONL_14124. **Lower panel)** Corresponding MS2 spectrum of the identified ion with a m/z of 504.1919. The fragmentation pattern reveals successive losses of water groups, hexoses, and an amino group, indicating that m/z 504.1919 represents an amino-tri-hexose.

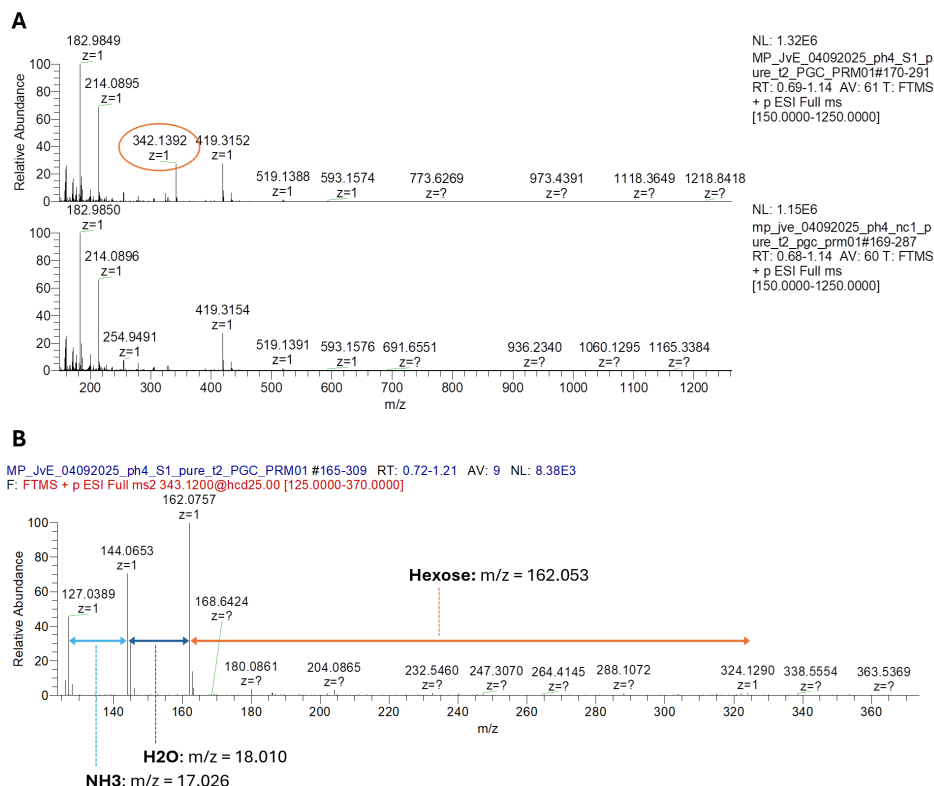


Figure S6.13. A) Full MS spectrum over the retention time 0.68-1.14 for the activity assay with MMBJNONL_14124 (upper panel) and the control without MMBJNONL_14124 (lower panel). **B)** Corresponding MS2 spectrum of the identified ion with a m/z of 342.1392. The fragmentation pattern reveals successive losses of a hexose, a water group and an amino group, indicating that m/z 342.1392 represents an amino-di-hexose.

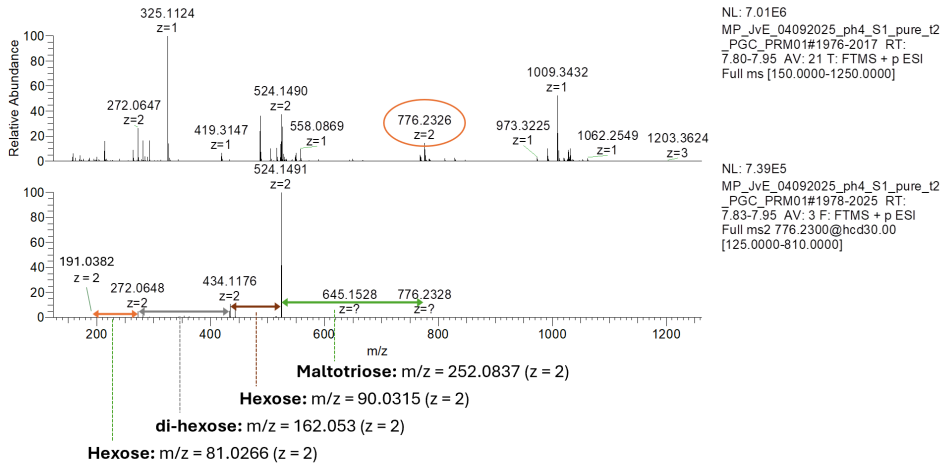


Figure S6.14. Upper panel) Full MS spectrum over the retention time range of 7.80-7.95 min for the activity assay with MMBJNONL_14124. **Lower panel)** Fragmentation spectrum of the ion with m/z = 776.2326 (z = 2).

References

- 1 Liu, X. & Kokare, C. in *Biotechnology of microbial enzymes* 405–444 (Elsevier, 2023).
- 2 Madhuri, R. J. *et al.* Recent approaches in the production of novel enzymes from environmental samples by enrichment culture and metagenomic approach. *Recent developments in applied microbiology and biochemistry*, 251–262 (2019).
- 3 Kocabaş, D. S., Lyne, J. & Ustunol, Z. Hydrolytic enzymes in the dairy industry: Applications, market and future perspectives. *Trends in Food Science & Technology* **119**, 467–475 (2022).
- 4 Klemm, D., Heublein, B., Fink, H. P. & Bohn, A. Cellulose: fascinating biopolymer and sustainable raw material. *Angewandte chemie international edition* **44**, 3358–3393 (2005).
- 5 Rinaudo, M. Chitin and chitosan: Properties and applications. *Progress in polymer science* **31**, 603–632 (2006).
- 6 Singh, R. S., Saini, G. K. & Kennedy, J. F. Pullulan: microbial sources, production and applications. *Carbohydrate polymers* **73**, 515–531 (2008).
- 7 Leathers, T. D. Biotechnological production and applications of pullulan. *Applied microbiology and biotechnology* **62**, 468–473 (2003).
- 8 Oğuzhan, P. & Yangilar, F. Pullulan: Production and usage in food industry. *African Journal of Food Science and Technology* **4**, 2141–5455 (2013).
- 9 Bertoldo, C. & Antranikian, G. Starch-hydrolyzing enzymes from thermophilic archaea and bacteria. *Current opinion in chemical biology* **6**, 151–160 (2002).
- 10 Olsen, H. S. & Falholt, P. The role of enzymes in modern detergency. *Journal of Surfactants and Detergents* **1**, 555–567 (1998).
- 11 Naik, B. *et al.* Pullulanase: unleashing the power of enzyme with a promising future in the food industry. *Frontiers in Bioengineering and Biotechnology* **11**, 1139611 (2023).
- 12 Singh, R. S., Singh, T. & Pandey, A. Microbial enzymes—an overview. *Advances in enzyme technology*, 1–40 (2019).
- 13 Kahar, U. M., Latif, N. A., Amran, S. I., Liew, K. J. & Goh, K. M. A bibliometric analysis and review of pullulan-degrading enzymes—past and current trends. *Catalysts* **12**, 143 (2022).
- 14 Yang, K. K., Wu, Z. & Arnold, F. H. Machine-learning-guided directed evolution for protein engineering. *Nature methods* **16**, 687–694 (2019).
- 15 Lutz, S. Beyond directed evolution—semi-rational protein engineering and design. *Current opinion in biotechnology* **21**, 734–743 (2010).
- 16 Eijssink, V. G., Gåseidnes, S., Borchert, T. V. & Van Den Burg, B. Directed evolution of enzyme stability. *Biomolecular engineering* **22**, 21–30 (2005).
- 17 Porter, J. L., Rusli, R. A. & Ollis, D. L. Directed evolution of enzymes for industrial biocatalysis. *ChemBioChem* **17**, 197–203 (2016).
- 18 Böttcher, D. & Bornscheuer, U. T. Protein engineering of microbial enzymes. *Current opinion in microbiology* **13**, 274–282 (2010).
- 19 Woodley, J. M. Protein engineering of enzymes for process applications. *Current opinion in chemical biology* **17**, 310–316 (2013).
- 20 Yang, H. *et al.* Molecular engineering of industrial enzymes: recent advances and future prospects. *Applied microbiology and biotechnology* **98**, 23–29 (2014).
- 21 Brindha, J., Balamurali, M. & Chanda, K. Evolutionary approaches in protein engineering towards biomaterial construction. *RSC advances* **9**, 34720–34734 (2019).
- 22 Marmur, J. A procedure for the isolation of deoxyribonucleic acid from micro-organisms. *Journal of molecular biology* **3**, 208–IN201 (1961).
- 23 Sambrook, J. & Russell, D. Molecular cloning: a laboratory manual 1. 3rd Cold Spring Harbor. New York, NY, 6.4–6.12 (2001).
- 24 Reguera, G. in *In the Company of Microbes: Ten Years of Small Things Considered*. Ch. 71: The great plate count anomaly” that is no more, 288–290 (2017).
- 25 Staley, J. T. & Konopka, A. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual review of microbiology* **39**, 321–346 (1985).
- 26 Dickson, I. Sequencing the Unculturable Majority. (2021).
- 27 Liu, S. *et al.* Opportunities and challenges of using metagenomic data to bring uncultured microbes into cultivation. *Microbiome* **10**, 76 (2022).
- 28 Uchiyama, T. & Miyazaki, K. Functional metagenomics for enzyme discovery: challenges to efficient screening. *Current opinion in biotechnology* **20**, 616–622 (2009).
- 29 Robinson, S. L., Piel, J. & Sunagawa, S. A roadmap for metagenomic enzyme discovery. *Natural Product Reports* **38**, 1994–2023 (2021).
- 30 Lam, K. N., Cheng, J., Engel, K., Neufeld, J. D. & Charles, T. C. Current and future resources for functional metagenomics. *Frontiers in microbiology* **6**, 165114 (2015).
- 31 Johansson, M. *Metaproteogenomics-guided enzyme discovery: Targeted identification of novel proteases in microbial communities*, Linköping University Electronic Press, (2018).
- 32 Van Den Berg, E. M., Van Dongen, U., Abbas, B. & Van Loosdrecht, M. C. Enrichment of DNRA bacteria in a continuous culture. *The ISME journal* **9**, 2153–2161 (2015).
- 33 Janse, I., Zwart, G., van der Maarel, M. J. & Gottschal, J. C. Composition of the bacterial community degrading *Phaeocystis* mucopolysaccharides in enrichment cultures. *Aquatic Microbial Ecology* **22**, 119–133 (2000).
- 34 Sukul, P. *et al.* Simple discovery of bacterial biocatalysts from environmental samples through functional metaproteomics. *Microbiome* **5**, 1–5 (2017).
- 35 Kleikamp, H. B. *et al.* Metaproteomics, metagenomics and 16S rRNA sequencing provide different perspectives on the aerobic granular sludge microbiome. *Water Research* **246**, 120700 (2023).
- 36 Kleiner, M. Metaproteomics: much more than measuring gene expression in microbial communities. *Msystems* **4**, 10.1128/msystems. 00115–00119 (2019).
- 37 Sharma, A., Gupta, G., Ahmad, T., Mansoor, S. & Kaur, B. Enzyme engineering: current trends and future perspectives. *Food Reviews International* **37**, 121–154 (2021).
- 38 von Meijenfeldt, F. A. B., Arkhipova, K., Cambuy, D. D., Coutinho, F. H. & Dutilh, B. E. Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. *Genome Biol* **20** (2019).
- 39 Burns, R. G. Microbial Extracellular Enzymes and the Degradation of Natural and Synthetic Polymers in Soil. *Progr Soil Sci*, 27–47 (2013). https://doi.org/10.1007/978-94-007-4177-5_2

- 40 Zheng, J. F. *et al.* dbCAN3: automated carbohydrate-active enzyme and substrate annotation. *Nucleic Acids Research* **51**, W115–W121 (2023). <https://doi.org/10.1093/nar/gkad328>
- 41 Matzke, J., Herrmann, A., Schneider, E. & Bakker, E. P. Gene cloning, nucleotide sequence and biochemical properties of a cytoplasmic cyclomaltodextrinase (neopullulanase) from *Alicyclobacillus acidocaldarius*, reclassification of a group of enzymes. *FEMS Microbiology Letters* **183**, 55–61 (2000).
- 42 Teufel, F. *et al.* SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol* **40**, 1023–+ (2022). <https://doi.org/10.1038/s41587-021-01156-3>
- 43 Moreno, J., Nielsen, H., Winther, O. & Teufel, F. Predicting the subcellular location of prokaryotic proteins with DeepLocPro. *Bioinformatics* **40** (2024). <https://doi.org/10.1093/bioinformatics/btae677>
- 44 Lee, S. P., Morikawa, M., Takagi, M. & Imanaka, T. Cloning of the Aapt Gene and Characterization of Its Product, Alpha-Amylase-Pullulanase (Aapt), from Thermophilic and Alkaliphilic *Bacillus* Sp Strain Xal601. *Appl Environ Microb* **60**, 3764–3773 (1994). <https://doi.org/Doi10.1128/Aem.60.10.3764-3773.1994>
- 45 Nakajima, R., Imanaka, T. & Aiba, S. Comparison of amino acid sequences of eleven different α -amylases. *Applied Microbiology and Biotechnology* **23**, 355–360 (1986). <https://doi.org/https://doi.org/10.1007/BF00257032>
- 46 De Jong, S. I., Sorokin, D. Y., van Loosdrecht, M., Pabst, M. & McMillan, D. G. Membrane proteome of the thermoalkaliphile *Caldalkalibacillus thermarum* TA2. A1. *Frontiers in Microbiology* **14**, 1228266 (2023).
- 47 Patel, B., Morgan, H. W. & Daniel, R. M. *Fervidobacterium nodosum* gen. nov. and spec. nov., a new chemoorganotrophic, caldoactive, anaerobic bacterium. *Archives of Microbiology* **141**, 63–69 (1985).
- 48 Duffner, F., Bertoldo, C., Andersen, J. T., Wagner, K. & Antranikian, G. A new thermoactive pullulanase from *Desulfurococcus mucosus*: cloning, sequencing, purification, and characterization of the recombinant enzyme after expression in *Bacillus subtilis*. *Journal of Bacteriology* **182**, 6331–6338 (2000).
- 49 Weimer, P. J., Wagner, L. W., Knowlton, S. & Ng, T. K. Thermophilic anaerobic bacteria which ferment hemicellulose: characterization of organisms and identification of plasmids. *Archives of microbiology* **138**, 31–36 (1984).
- 50 Darland, G. & Brock, T. D. *Bacillus acidocaldarius* sp. nov., an acidophilic thermophilic spore-forming bacterium. *Microbiology* **67**, 9–15 (1971).
- 51 Hulsman, A., Lurz, R., Scheffel, F. & Schneider, E. Maltose and maltodextrin transport in the thermoacidophilic gram-positive bacterium *Alicyclobacillus acidocaldarius* is mediated by a high-affinity transport system that includes a maltose binding protein tolerant to low pH. *Journal of Bacteriology* **182**, 6292–6301 (2000). <https://doi.org/10.1128/jb.182.22.6292-6301.2000>
- 52 Papadopoulos, J. S. & Agarwala, R. COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* **23**, 1073–1079 (2007). <https://doi.org/10.1093/bioinformatics/btm076>

Outlook



Carbohydrates show remarkable structural diversity and are essential to fundamental biological functions across all domains of life. However, their analysis remains complex and lags far behind that of the other fundamental cellular molecules, DNA and proteins. This is a consequence of glycans not being directly encoded in DNA, but instead being secondary gene products formed through the combined action of various enzymes¹. Moreover, their pool of building blocks is far more diverse, with many derivatives yet to be discovered², and their ability to form highly branched, non-linear structures further complicates their analyses. While genomic sequencing has become routine, the function of many gene products remains unknown, leaving true metabolic capabilities uncertain³. Proteomics can complement genomics by revealing which enzymes are expressed. Still, it does not necessarily reflect their activity, as this may be regulated by post-translational modifications⁴. Moreover, low-abundance proteins may escape detection, and membrane proteins remain challenging to analyze using standard preparation methods⁵. Consequently, direct metabolomic analysis of the sugar derivatives produced is essential to understand the glycophenotype of microorganisms.

Nucleotide sugars are the universal precursors of glycomolecules, yet their biochemical diversity has only been investigated to a limited extent to date⁶. Moreover, many nucleotide-sugar derivatives are often present at low abundance, further complicating their detection. Comprehensive characterization of these compounds requires exceptional analytical sensitivity, mass accuracy and resolution. Recent advancements in mass spectrometry instrumentation, including the introduction of the Orbitrap Astral⁷ and the Orbitrap Excedion Pro⁸, have substantially expanded the capabilities of mass spectrometry based omics research. These machines combine markedly increased sensitivity with very high acquisition speeds while maintaining high mass resolution. The application of such next-generation instrumentation will enable deeper coverage of nucleotide sugars, including rapidly turned-over metabolic intermediates, which is required for mapping of metabolic pathways in novel microbes.

SugarBase was developed to serve as a screening platform for nucleotide-sugars and can detect both known nucleotide-sugars as well as discover novel candidates by using a theoretical sugar composition database. It provides a bases to form hypotheses and serves as an entry point for deeper structural and functional analyses. While it provides a putative chemical composition for even entirely novel monosaccharides, it does not provide structural information. For nonulosonic acids (NulOs), orthogonal chemical labelling provided insights into N and O-

modifications and allowed differentiation between Pseudaminic and Legionaminic acid. However, further advancements in biosynthetic approaches are required to produce standards for the other stereoisomers, including 4- and 8-epilegionaminic acid, acinetaminic acid and 8-epiacinetaminic acid, which are currently commercially not available. Interestingly, recent studies have already reported the enzymatic or chemical synthesis of several of these compounds⁹⁻¹¹. These developments will facilitate the discrimination between additional isomers and improve understanding of their occurrence and biological functions, which currently remain largely unexplored.

Most work still depends on NMR spectroscopy which enables robust structural elucidation without the need of a reference standard. Nevertheless, it requires substantially larger amounts of material than MS-based methods, which is often a limiting factor for microbial samples¹². Recent developments in advanced fragmentation strategies are highly promising to provide more insights into the occurrence of sugar stereoisomers. A promising example is demonstrated by Nagy and Pohl (2015) who developed a fixed-ligand kinetic model that generates specific fragment ion pairs, enabling the discrimination of all 24 hexose isomers¹³.

7

Emerging techniques with high potential for the detection and separation of stereoisomers are ion mobility spectrometry (IMS) and infrared ion spectroscopy. IMS-MS separates ions based on size, shape and charge, yielding collision cross section (CCS) values, which reflect their three dimensional structure. This enables the discrimination of isomers and has already shown considerable promise in glycan analysis^{14,15}. Initially, studies indicated that isobaric monosaccharides often exhibit insufficient CCS differences for effective separation, resulting in complex formation strategies^{16,17}. The development of high-resolution instruments such as cyclic ion mobility spectrometry (cIMS) significantly enhanced separation potential and has shown potential in the differentiation of underivatized monosaccharides¹⁷. Advantageously, CCS values can be predicted computationally, which is especially interesting for microbial sugars for which standards are often unavailable. Nevertheless, correlating a predicted single CCS value to an experimental one remains often challenging¹⁸.

Another emerging approach is infrared ion spectroscopy (IRIS), which identifies molecules by inducing ion fragmentation through resonant IR absorption, producing spectra that function as structural fingerprints¹⁹. A key advantage is that these spectra can be accurately predicted using quantum-chemical calculations.

This enables reference-free identification, although it remains computationally intensive¹⁹. IRIS also provides direct functional group information, supporting the structural elucidation of unknown compounds and narrowing down the list of candidate structures²⁰. Its ability to resolve subtle structural differences makes it particularly useful for distinguishing monosaccharides and their derivatives^{19,21}. It has been used to differentiate monosaccharide stereoisomers, as well as structural features like ring size (pyranose vs furanose) and stereochemistry (α vs β anomer) using different adducts²¹. Cryogenic ion spectroscopy has further improved resolution but currently relies on home-built instruments^{18,22}. Commercialization of these instruments might advance structural elucidation of glycans. Further improvements can be obtained in the coupling of IRIS with LC-MS, which remains challenging due to mismatched timescales. Nevertheless, it has been demonstrated for the identification of isomeric species in complex samples¹⁹. Promising separations have also been shown by combining IRIS with ion mobility techniques^{19,22}.

These technical advances substantially increased the quantity of data as well as its complexity. As a result, more sophisticated computational approaches are required for processing and understanding these data. Machine learning and AI have been demonstrating its potential in metabolomics. They already support data pre-processing tasks such as peak picking, spectral denoising and retention time alignment^{23,24}. Deep learning has been used for predicting LC retention times, thereby improving non-targeted identification, although variability between instruments and laboratories remains a limitation²⁵. Emerging tools such as CandyCrunch²⁶ highlight the potential of AI for glycan structure prediction from MS/MS data, suggesting similar opportunities for monosaccharides and activated sugars. However, the effectiveness of these approaches depends on the availability of high-quality training data, which is currently limited for rare microbial monosaccharides, underscoring the importance of experimental data.

Advanced analytical and data processing approaches will help us to determine what microbes produce. As a result, the next question would be “what is it doing” and “how can it be used to answer biological questions or develop novel medical or biotechnological applications”? For this, high resolution stable isotope tracing could help to elucidate novel biosynthetic routes and networks²⁷. Such data, together with the development of metabolic models, can allow us to understand how these pathways are used and regulated. This can provide the foundation for engineering microbial hosts for the overproduction of sugar molecules, for the

discovery of novel carbohydrate active enzymes, and finally to produce these glycomolecules at larger scales so that they can be used to explore their functional roles.

The direct use of cell extracts as source of sugar precursors opens new doors in screening for novel glycosyltransferases and for producing such glycomolecules²⁸⁻³¹. Future work on non-pathogenic organisms will allow a safer and more scalable application. Moreover, identifying microbes which produce a single NulO stereoisomer would help to better assess enzyme specificity and promiscuity. While direct experimental evidence of glycosyltransferase activity is likely hindered by the limited availability of sugar precursors, difficulties in obtaining properly folded, catalytically active recombinant enzymes likely contribute to the lack of such evidence as well.

Advances in native MS have the potential to enable broader screening of glycosyltransferase candidates, thereby facilitating the identification of novel enzymes. High-throughput approaches for the reverse screening, that is screening many ligands against a single protein, already exist³². Initial studies also showed the feasibility of detecting protein-ligand interactions directly from complex mixtures using a reactive desorption electrospray ionization MS based approach³³. Additional developments might expand this approach to a proteome-wide screening, which could provide completely novel glycosyltransferase candidates.

The discovery of glycosyltransferases, together with access to their corresponding nucleotide-sugar donors, enables the biosynthesis of glycans and glycoconjugates that are difficult or infeasible to obtain via chemical synthesis. These glycomolecules can be used to answer biological questions as well as enable novel medical and biotechnological applications. An example includes the production of glycoconjugate vaccines, these consist of glycans covalently linked to a protein carrier to enhance immunogenicity^{34,35}. Another promising area is the use of glycan-based binding assays. For example, flagellotropic bacteriophages have been reported to bind the flagella of the pathogenic bacteria *Campylobacter jejuni* to initiate infection^{36,37}. The flagellin proteins of *C. jejuni* are densely covered with various nonulosonic acid derivatives, which serve as receptors for these phages^{38,39}. Variety in their NulO profiles may alter susceptibility to phage binding³⁸.

Besides glycosyltransferases, enzymes in general are of great interest and are already widely applied in industry, with hydrolases being the most prevalent class

of enzymes in use. As natural, biodegradable catalysts, they have the potential to enable sustainable biotechnological and industrial processes by operating under mild conditions, reducing fossil resource use and limiting chemical waste⁴⁰. While the importance of enzymes is widely recognized, the discovery of novel enzymes with improved characteristics remains challenging.

Due to their diversity, microorganisms provide a vast reservoir of biocatalysts. Metagenomics has advanced enzyme discovery by isolating DNA directly from environmental samples, thereby circumventing the need to grow pure cultures in the lab⁴¹. Future improvements should focus on sequencing depth as well as functional annotations. Currently, many genes remain unannotated, but advances in AI-based approaches, including structural prediction^{42,43}, may help to fill this gap (at least partially). This progress will also depend on the continued availability of high-performance computational resources. The rapid adoption of large-scale AI systems has increased the demand for specialized hardware, including high-performance GPUs, while also intensifying energy consumption and placing growing pressure on data storage and transfer infrastructures⁴⁴⁻⁴⁶. The growing use of digital biological data in bioinformatics is also raising important questions with respect to the Nagoya Protocol, particularly regarding access, ownership, and fair benefit-sharing of genetic resources and associated digital sequence information^{47,48}. In this context, the use of genetic resources is becoming an increasing challenge for industry. One solution could be the development of secure local or national databases that allow countries and companies to identify, manage, and valorize their own biological resources. The integration of advanced omics technologies into microbiological approaches, such as enrichments based on desired functions (which is particularly useful for carbohydrate active enzymes), is an emerging strategy that could be further scaled into multiple parallel screening platforms. Streamlined data processing pipelines for the meta-omics readouts can facilitate the subsequent targeted discovery of enzymes.

Overall, recent instrumental advances in spectrometric and spectroscopic techniques (MS, IMS, IRIS) tremendously increased the amount of information that can be obtained from complex microbial ecosystems. Additional advances in computational methods (e.g. machine learning and metaomics data integration) and functional screening strategies will collectively enable deeper understanding of microbial glycans and carbohydrate-active enzymes, facilitating their applications across fundamental research, medical applications, and biotechnology.

References

- 1 Zol-Hanlon, M. I. & Schumann, B. Open questions in chemical glycobiology. *Commun Chem* **3** (2020). <https://doi.org/10.1038/s42004-020-00337-6>
- 2 Cummings, R. D. A periodic table of monosaccharides. *Glycobiology* **34** (2024). <https://doi.org/10.1093/glycob/cwad088>
- 3 Sevin, D. C., Fuhrer, T., Zamboni, N. & Sauer, U. Nontargeted in vitro metabolomics for high-throughput identification of novel enzymes in *Escherichia coli*. *Nat Methods* **14**, 187–194 (2017). <https://doi.org/10.1038/nmeth.4103>
- 4 Cappelletti, V. *et al.* Dynamic 3D proteomes reveal protein functional alterations at high resolution in situ. *Cell* **184**, 545–559 e522 (2021). <https://doi.org/10.1016/j.cell.2020.12.021>
- 5 Carpenter, E. P., Beis, K., Cameron, A. D. & Iwata, S. Overcoming the challenges of membrane protein crystallography. *Curr Opin Struct Biol* **18**, 581–586 (2008). <https://doi.org/10.1016/j.sbi.2008.07.001>
- 6 Mikkola, S. Nucleotide Sugars in Chemistry and Biology. *Molecules* **25** (2020). <https://doi.org/ARTN575510.3390/molecules25235755>
- 7 Zabrouskov, V. *et al.* Rethink what is possible with the Orbitrap Astral mass spectrometer, <<https://documents.thermofisher.com/TFS-Assets/CMD/Reference-Materials/wp-001800-ms-orbitrap-astral-mass-spectrometer-wp001800-en.pdf>> (Discover. Innovate. Exceed. Orbitrap Excedion Pro hybrid mass spectrometer, <<https://documents.thermofisher.com/TFS-Assets/CMD/Specification-Sheets/ps-003701-orbitrap-excedion-pro-specification-sheet-ps003701-na-en.pdf>> (
- 8 Schoenhofen, I. C., McNally, D. J., Brisson, J. R. & Logan, S. M. Elucidation of the CMP-pseudaminic acid pathway in *Helicobacter pylori*: synthesis from UDP-N-acetylglucosamine by a single enzymatic reaction. *Glycobiology* **16**, 8C–14C (2006). <https://doi.org/10.1093/glycob/cwl010>
- 10 Hassan, M. I. *et al.* Total Biosynthesis of Legionaminic Acid, a Bacterial Sialic Acid Analogue. *Angew Chem Int Ed Engl* **55**, 12018–12021 (2016). <https://doi.org/10.1002/anie.201606006>
- 11 Guo, X., Li, P., Liu, H. & Li, X. Recent progress on the chemical synthesis of bacterial non-2-ulonic acids. *Organic Chemistry Frontiers* **10**, 3429–3446 (2023).
- 12 Ye, D. Y., Li, X. W., Shen, J. Z. & Xia, X. Microbial metabolomics: From novel technologies to diversified applications. *Trac-Trend Anal Chem* **148** (2022). <https://doi.org/10.1007/s13361-022-11654-0>
- 13 Nagy, G. & Pohl, N. L. Complete hexose isomer identification with mass spectrometry. *J Am Soc Mass Spectrom* **26**, 677–685 (2015). <https://doi.org/10.1007/s13361-014-1072-z>
- 14 Sastre Torano, J. *et al.* De novo sequencing of glycans by ion mobility-mass spectrometry using a self-expanding database. *Nat Commun* **17**, 382 (2025). <https://doi.org/10.1038/s41467-025-67069-w>
- 15 Hofmann, J. & Pagel, K. Glycan Analysis by Ion Mobility-Mass Spectrometry. *Angew Chem Int Ed Engl* **56**, 8342–8349 (2017). <https://doi.org/10.1002/anie.201701309>
- 16 Gaye, M. M., Nagy, G., Clemmer, D. E. & Pohl, N. L. Multidimensional Analysis of 16 Glucose Isomers by Ion Mobility Spectrometry. *Anal Chem* **88**, 2335–2344 (2016). <https://doi.org/10.1021/acs.analchem.5b04280>
- 17 Ollivier, S., Ropartz, D., Fanuel, M. & Rogniaux, H. Fingerprinting of Underivatized Monosaccharide Stereoisomers Using High-Resolution Ion Mobility Spectrometry and Its Implications for Carbohydrate Sequencing. *Anal Chem* **95**, 10087–10095 (2023). <https://doi.org/10.1021/acs.analchem.3c01531>
- 18 van Outersterp, R. E. *et al.* Distinguishing Oligosaccharide Isomers Using Far-Infrared Ion Spectroscopy: Identification of Biomarkers for Inborn Errors of Metabolism. *Anal Chem* **95**, 9787–9796 (2023). <https://doi.org/10.1021/acs.analchem.3c00363>
- 19 Martens, J. *et al.* Infrared ion spectroscopy: New opportunities for small-molecule identification in mass spectrometry - A tutorial perspective. *Anal Chim Acta* **1093**, 1–15 (2020). <https://doi.org/10.1016/j.aca.2019.10.043>
- 20 Martens, J. *et al.* Molecular identification in metabolomics using infrared ion spectroscopy. *Sci Rep* **7**, 3363 (2017). <https://doi.org/10.1038/s41598-017-03387-4>
- 21 Ho, J. S. *et al.* Distinguishing Galactoside Isomers with Mass Spectrometry and Gas-Phase Infrared Spectroscopy. *J Am Chem Soc* **143**, 10509–10513 (2021). <https://doi.org/10.1021/jacs.0c11919>
- 22 Warnke, S., Ben Faleh, A. & Rizzo, T. R. Toward High-Throughput Cryogenic IR Fingerprinting of Mobility-Separated Glycan Isomers. *ACS Meas Sci Au* **1**, 157–164 (2021). <https://doi.org/10.1021/acsmeasuresciau.1c00018>
- 23 Sirocchi, C. *et al.* Exploring machine learning for untargeted metabolomics using molecular fingerprints. *Comput Methods Programs Biomed* **250**, 108163 (2024). <https://doi.org/10.1016/j.cmpb.2024.108163>
- 24 Liebal, U. W., Phan, A. N. T., Sudhakar, M., Raman, K. & Blank, L. M. Machine Learning Applications for Mass Spectrometry-Based Metabolomics. *Metabolites* **10** (2020). <https://doi.org/10.3390/metabo10060243>
- 25 Xue, J., Wang, B., Ji, H. & Li, W. RT-Transformer: retention time prediction for metabolite annotation to assist in metabolite identification. *Bioinformatics* **40** (2024). <https://doi.org/10.1093/bioinformatics/btae084>
- 26 Urban, J. *et al.* Predicting glycan structure from tandem mass spectrometry via deep learning. *Nat Methods* **21**, 1206–1215 (2024). <https://doi.org/10.1038/s41592-024-02314-6>
- 27 Gao, Y. *et al.* Charting unknown metabolic reactions by mass spectrometry-resolved stable-isotope tracing metabolomics. *Nat Commun* **16**, 5059 (2025). <https://doi.org/10.1038/s41467-025-60258-7>
- 28 Meng, X., Boons, G. J., Westen, M. & Wennekes, T. Metabolic Labeling of Legionaminic Acid in Flagellin Glycosylation of *Campylobacter jejuni* Identifies Maf4 as a Putative Legionaminyl Transferase. *Angew Chem Int Ed Engl* **60**, 24811–24816 (2021). <https://doi.org/10.1002/anie.202107181>
- 29 Yang, K. Y. *et al.* Glycosyltransferase Jhp0106 (PseE) contributes to flagellin maturation in *Helicobacter pylori*. *Helicobacter* **26**, e12787 (2021). <https://doi.org/10.1111/hel.12787>
- 30 Ardisson, S., Kint, N. & Viollier, P. H. Specificity in glycosylation of multiple flagellins by the modular and cell cycle regulated glycosyltransferase FimG. *Life* **9** (2020). <https://doi.org/10.7554/eLife.60488>
- 31 Sulzenbacher, G. *et al.* Glycosylate and move! The glycosyltransferase Maf is involved in bacterial flagella

- formation. *Environ Microbiol* **20**, 228–240 (2018). <https://doi.org/10.1111/1462-2920.13975>
- 32 Nguyen, G. T. H. *et al.* Multiplexed Screening of Thousands of Natural Products for Protein-Ligand Binding in Native Mass Spectrometry. *J Am Chem Soc* **143**, 21379–21387 (2021). <https://doi.org/10.1021/jacs.1c10408>
 - 33 Zheng, Q. *et al.* Ligand-protein target screening from cell matrices using reactive desorption electrospray ionization-mass spectrometry via a native-denatured exchange approach. *Analyst* **144**, 512–520 (2019). <https://doi.org/10.1039/c8an01708e>
 - 34 Kay, E., Cuccui, J. & Wren, B. W. Recent advances in the production of recombinant glycoconjugate vaccines. *Npj Vaccines* **4** (2019). <https://doi.org/10.1038/s41541-019-0110-z>
 - 35 Wei, R. *et al.* Synthetic Pseudaminic-Acid-Based Antibacterial Vaccine Confers Effective Protection against *Acinetobacter baumannii* Infection. *ACS Cent Sci* **7**, 1535–1542 (2021). <https://doi.org/10.1021/acscentsci.1c00656>
 - 36 Baldvinsson, S. B., Sorensen, M. C., Vegge, C. S., Clokie, M. R. & Brondsted, L. Campylobacter jejuni motility is required for infection of the flagellotropic bacteriophage F341. *Appl Environ Microbiol* **80**, 7096–7106 (2014). <https://doi.org/10.1128/AEM.02057-14>
 - 37 Ostenfeld, L. J. *et al.* A hybrid receptor binding protein enables phage F341 infection of Campylobacter by binding to flagella and lipooligosaccharides. *Front Microbiol* **15**, 1358909 (2024). <https://doi.org/10.3389/fmicb.2024.1358909>
 - 38 Sacher, J. C. *et al.* Binding of Phage-Encoded FlaGrab to Motile Campylobacter jejuni Flagella Inhibits Growth, Downregulates Energy Metabolism, and Requires Specific Flagellar Glycans. *Front Microbiol* **11**, 397 (2020). <https://doi.org/10.3389/fmicb.2020.00397>
 - 39 Javed, M. A. *et al.* A receptor-binding protein of Campylobacter jejuni bacteriophage NCTC 12673 recognizes flagellin glycosylated with acetamidino-modified pseudaminic acid. *Mol Microbiol* **95**, 101–115 (2015). <https://doi.org/10.1111/mmi.12849>
 - 40 Jegannathan, K. R. & Nielsen, P. H. Environmental assessment of enzyme use in industrial production - a literature review. *J Clean Prod* **42**, 228–240 (2013). <https://doi.org/10.1016/j.jclepro.2012.11.005>
 - 41 Uchiyama, T. & Miyazaki, K. Functional metagenomics for enzyme discovery: challenges to efficient screening. *Curr Opin Biotechnol* **20**, 616–622 (2009). <https://doi.org/10.1016/j.copbio.2009.09.010>
 - 42 Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* **50**, D439–D444 (2022). <https://doi.org/10.1093/nar/gkab1061>
 - 43 van Kempen, M. *et al.* Fast and accurate protein structure search with Foldseek. *Nat Biotechnol* **42**, 243–246 (2024). <https://doi.org/10.1038/s41587-023-01773-0>
 - 44 Nobile, M. S., Cazzaniga, P., Tangherloni, A. & Besozzi, D. Graphics processing units in bioinformatics, computational biology and systems biology. *Brief Bioinform* **18**, 870–885 (2017). <https://doi.org/10.1093/bib/bbw058>
 - 45 Stephens, Z. D. *et al.* Big Data: Astronomical or Genomical? *PLoS Biol* **13**, e1002195 (2015). <https://doi.org/10.1371/journal.pbio.1002195>
 - 46 Iman, N. The Hidden Costs of Intelligence: Artificial Intelligence and Machine Learning Adoption and the Paradox of Exponential Digital Growth. *Sustainability and Climate Change* **18**, 225–241 (2025).
 - 47 Smith, D., da Silva, M., Jackson, J. & Lyal, C. Explanation of the Nagoya Protocol on Access and Benefit Sharing and its implication for microbiology. *Microbiology (Reading)* **163**, 289–296 (2017). <https://doi.org/10.1099/mic.0.000425>
 - 48 Watanabe, M. E. The Nagoya Protocol: The Conundrum of Defining Digital Sequence Information. *BioScience* **69**, 480 (2019). <https://doi.org/10.1093/biosci/biz034>

Acknowledgments

This thesis would not have been possible without the wonderful support of the many people who surrounded me throughout the past four years.

First of all, **Martin**, thank you so much for giving me the opportunity to pursue my PhD with you. I can truly say that I learned an incredible amount from you, and I am still amazed by the breadth of your knowledge and your seemingly endless stream of ideas for future experiments. Your never-ending enthusiasm and positivity were a constant source of motivation throughout my PhD and despite your busy schedule, you always made time for our meetings. Thank you for creating such a safe and open environment, where everything could be discussed freely. I also really enjoyed the Christmas dinners and summer BBQs with the MS group, which were great opportunities to connect outside of work. I am also very grateful for your continued guidance as I take the next steps in my career, and for how involved and supportive you remain, always willing to share your knowledge and help wherever you can. I could not have asked for a better supervisor.

Thank you, **Mark**, for bringing together such an amazing group and for creating such a positive and enjoyable working atmosphere. I am grateful for your guidance throughout my PhD. Your endless curiosity encouraged me to look at my research from different perspectives and to keep asking new questions. Thank you as well for supporting the connection within EBT by making sure we had the opportunity to spend time together and connect outside of work during the annual EBT retreats.

Thank you to everyone who contributed to this thesis and with whom I had the opportunity to collaborate. **Caroline**, thank you for your advice and for being always available to answer any questions. Thank you, **Timmy**, for the great collaboration that ultimately resulted in Chapter 4. I really appreciated the warm welcome you gave me at EBT. Your enthusiasm and natural ability to connect with people are truly inspiring, and you always made sure that our lunches were lively, fun, and full of good conversation! **Dinco**, thank you for the great collaboration on the review that became Chapter 2. It is always nice to have someone to struggle through the process with! I also really enjoyed our lunch walks, ice-skating adventures, and, of course, sharing our fear on the roller coaster at the Efteling. You brought a lot of fun and joy to my time at EBT. Thank you, **Miriam**, for giving me the opportunity to

contribute to your work. Your kindness and warmth always made our meetings very enjoyable, and it has been a real pleasure working with you!

Geert, thank you for the opportunity to work together on a project with IFF. Your enthusiasm and curiosity made it a real pleasure to collaborate with you, ultimately resulting in Chapter 6. This would also not have been possible without my students. Thank you **Suzanne**, for your valuable contribution to this project. It was a pleasure supervising you, and I am very grateful that you shared your bioinformatics skills and expertise with me. I also enjoyed hearing about all your creative projects. **Wael**, thank you for choosing to do your bachelor project with me working on transferase assays (Chapter 5). Your curiosity made it a pleasure to work with you! **Louise**, thank you for your enthusiasm, curiosity, and motivation. It has been wonderful to talk with someone who is just as excited about your research as you are. You made valuable contributions to the research presented in Chapter 5, and although I could not include it anymore in this thesis, they will be included in a future manuscript.

A big thank you to **Dita**. You are such a warm-hearted and caring person, always willing to help wherever you can. You made sure that the MS lab was always ready for us to work in, but even more importantly, you contributed so much to the wonderful atmosphere through your kindness, genuine interest, and cheerfulness. Thank you as well for organizing all the fun social activities, from ice skating to the BBQs at your place. You introduced me to climbing, and although I can now confidently say that it is even more painful than rowing (haha), it has nevertheless brought me a lot of joy. Thank you for making both the work and everything around it so much more enjoyable.

Thank you to **Zita, Ben, Dirk, Patricia**, and **Kevin**, who were always ready to lend a helping hand. Without you, the labs would not have run as smoothly as they did. **Marc** and **Laura**, thank you for warmly welcoming me to the BOC labs and for sharing your knowledge and experience with me. **Marijke**, thank you as well for your guidance and support at the IMB labs. **Anka, Kawieta, Miranda, Gea, Jannie, Apilena**, and **Astrid**, thank you for all your assistance and for everything you did to keep things running smoothly in and outside of the labs. A special thank you to **Rita**, who taught me how to work in the ML2 lab and supported me throughout my experiments and research. **Stef**, thank you for your positivity and kindness. I really enjoyed our climbing evenings together and our chats in the office.

A special thank you to my paranymphs. **Berdien**, thank you for all your support throughout the last years. You were always there to listen and offer your advice, both science and non-science related. Your kindness and positivity have been greatly appreciated. I enjoyed all our talks, lunch walks, ice skating adventures, tea breaks and all other activities. **Puck**, thank you for sharing the PhD journey with me. Your creativity and kind-hearted nature are truly inspiring, and I always appreciated hearing your perspective on whatever came up in our conversations. It was also comforting to have someone with whom I could share the ups and downs and struggles of doing a PhD. Thank you for letting me babysit Pepper! She convinced me even more that having two cats of my own was a good idea, and they brought so much joy, love, and many much-needed cuddles during the final year of my PhD.

A big thank you to my office mates. **Marloes, David, Steven, Simon, Marieken** and **Ramon**, it has been a pleasure to share the office with you. Thank you for all the nice conversations both work and non-work related and the fun you all brought to the office. A special thank you to **Marieken**, you made our crazy Ringvaart adventure possible! Thank you, **Ramon**, for also sharing the PhD journey with me. You brought a lot of laughter and fun to my time as a PhD student. Your humor in the office, and especially at the coffee table, made our breaks even more enjoyable. I will never forget the importance of a good coffee or tea, and, of course, the equally important matter of using exactly the right water temperature! Thank you as well for all the lunch walks, decorating the office (making sure it is always Christmas time) and organizing a great day at the Efteling. It was also great fun learning how to climb and getting our climbing certificates together with Siem and Stef.

Thank you, **Beatriz, Marit, Natalia**, and **Lemin**, for all the joy you brought, for sharing the ups and downs of our PhD journeys, and for your support and kindness throughout the years. I enjoyed our lovely dinners together. **Beatriz**, you are such a caring person and an amazing baker. Your creations are always delicious! **Marit**, thank you for your humor and kindness. **Natalia**, thank you for organizing the EBT retreat with me; your creativity and positive spirit are inspiring. **Lemin**, thank you for all your support. You are such a warm and caring person, always ready to help. It has been a real pleasure sharing these past years with all of you!

A big thank you to all the EBTers! **Claudia, Stefan, Timmy, David, Sam, Sergio, Francesc, Chris, Maxim, Hugo, Ali, Nina**, and **Rodoula**, thank you for giving me a warm welcome within EBT. Thank you to **Siem, Yubo, Jelle, Martijn, Sarah, João, Timo, Emine, Dieke, Linghang, Ji, Adarsh, Bogdans, Kyle, Mengqing, Carla**,

Christoff, Ismail, Samara, Alba, Luz, Jisk, Goncalo, Sebastian, Jelmer, Pim, Karel, Mario, Phillip, Dimitry, Sirous, Heike, Rebeca, Robbert, Michele, Samarpita, and Yuemei for providing a wonderful environment and enjoyable time in EBT, for all the good times during lunch and the many EBT cakes!

Aan mijn lieve studievrienden, **Dieuwke, Maaïke B, Maaïke S, Lily, Maureen, Agnes, Samira, Wouter, en Thijs**. Bedankt voor alle leuke momenten en activiteiten samen! Van weekendjes weg tot samen pannenkoeken eten en veel spelletjes spelen. Jullie brengen altijd veel gezelligheid en plezier! Het is mooi om te zien hoe iedereen inmiddels zijn eigen weg heeft gevonden, en ik ben heel benieuwd naar wat de toekomst voor jullie brengt.

A special thank you to all my rowing friends. **Sarah, Maaïke, Daphne, and José**, thank you so much for all the joy you have brought into my life. From all our crazy ideas, rowing competitions, marathons, and the Ringvaart, to lovely dinners, ice creams, and our cycling weekend. I have enjoyed every bit of it. I am incredibly grateful for all the fun, encouragement, and support you bring. Sarah, a special thank you as well for all our runs together and for the many conversations we have shared along the way. It is truly wonderful to be part of such a warm, supportive, and fun rowing crew! **Astrid**, bedankt dat je begon als mijn roeicoach en daarna gewoon samen met mij bent blijven roeien! Je bent een fantastisch en ontzettend sportief persoon en brengt met je humor altijd veel gezelligheid en plezier. Bedankt voor alle mooie, sportieve en vooral leuke momenten samen! **Ronald, Jaap Jan, Maarten, Patricia, en Bert**, bedankt voor alle mooie roeimomenten en marathons die ik met jullie samen heb mogen meemaken. De PhD was blijkbaar nog niet genoeg uitdaging, want jullie introduceerde mij bij de Elfstedentocht. En hoewel het 's nachts afzien was, kijk ik altijd terug op hele mooie en gezellige tochten. **Ronald**, het leverde veel "oh oh" reacties op, maar bedankt voor alle roeitochten samen en het uitstippelen van nieuwe avonturen op ander water.

Dankjewel, **Suze en Jackelien**, voor al jullie gezelligheid! We wonen dan niet allemaal in hetzelfde land, maar dankzij onze 'Netflix-avondjes' op de woensdagavond hebben we genoeg tijd om bij te kletsen en met elkaar te lachen. Ik geniet enorm van onze oud en nieuw vieringen en outdoorvakanties in de Alpen en Scandinavië. Jullie zorgen altijd voor veel plezier, avontuur en de nodige gekkigheid! Bedankt voor al jullie steun de afgelopen jaren en voor een luisterend oor. Op naar nog heel veel mooie momenten, avonturen en herinneringen samen!

Lieber **Georg**, ich kenne dich erst seit etwas mehr als einem Jahr, aber ich bin unglaublich froh, dich in meinem Leben zu haben. Dein Optimismus und deine Begeisterung sind ansteckend und inspirierend. Danke für all deine Verrücktheit, deine Liebe und deine Geduld. Du hast dieses letzte Jahr für mich noch so viel schöner gemacht. Ich bin sehr dankbar, dich an meiner Seite zu haben. Tut mir leid, dass im Büchlein keine Kängurus zu finden sind haha.

Lieve **papa** en **mama**, bedankt voor jullie onvoorwaardelijke steun! Jullie hebben mij altijd aangemoedigd om te doen wat ik leuk vind en staan altijd voor me klaar met hulp, advies en een luisterend oor. Zonder jullie steun was ik nooit gekomen waar ik nu ben. Ik ben ontzettend dankbaar dat ik altijd op jullie kan rekenen. Lieve **Matthijs**, bedankt voor al het plezier dat we samen beleven en voor alle topjes die we beklimmen. Bedankt dat ik altijd op jouw advies kan rekenen, van tips over een Python-code tot welke racefiets een goede keuze is. Lieve **Meeke**, een betere zus had ik me niet kunnen wensen. Je bent een ontzettend lief en warm persoon. Ik kan altijd alles met je bespreken en je staat altijd voor me klaar. Bedankt voor al onze gesprekken en leuke activiteiten. Ik kijk uit naar nog veel meer mooie momenten met jullie samen, bedankt voor al jullie liefde!



Curriculum Vitae

Jitske Marianne van Ede was born on 21 May 1997 in Hellevoetsluis, the Netherlands. Driven by curiosity in biology and chemistry, she enrolled in the joint Bachelor's programme in Life Science and Technology at Leiden University and Delft University of Technology. There she received the Young Talent Incentive Prize from the Royal Holland Society of Sciences and Humanities (KHMW).

In 2021, Jitske obtained her Master's degree in Life Science and Technology from Delft University of Technology. During that time, she developed a strong interest in mass spectrometry-based omics technologies that allow to explore biomolecular processes in microbes. Through a company internship with Aplagon Oy, a Finnish clinical-stage biopharmaceutical company, Jitske also performed molecular level characterization of a novel glycoprotein therapeutic that is currently in clinical phase trials. In 2022, Jitske started her PhD research in microbial omics at the Department of Biotechnology at Delft University of Technology, with Martin Pabst and Mark van Loosdrecht as promoters. Her PhD project focused on mass spectrometry-based omics to uncover novel microbial glycomolecule biosynthetic pathways and carbohydrate-active enzymes, key molecules in the fight against pathogenic microbes. In her free time, Jitske enjoys rowing, painting, and spending time outdoors, including hiking in the Alps with family and friends.

Scientific contributions

Peer reviewed articles and preprints:

1. **van Ede, J. M.**, Sorensen, M. C. H., van Loosdrecht, M., & Pabst, M. (2026). SugarBase: mapping glycomolecule precursors in microbes. *bioRxiv*, 2026-04.
2. Cammaert, M.*, Wouters, R. I.*, **van Ede, J. M.***, de Hulster, E. A., Mooiman, C. M., van Dam, P. T., Pabst, M., van Gulik, W., & Daran-Lapujade, P. (2026). A protocol for lab-scale production of ¹³C yeast extract as internal standard for metabolomics and quantification of intracellular metabolites. *bioRxiv*, 2026-06. (*equal first)
3. **van Ede, J. M.**, Soic, D., & Pabst, M. (2026). Decoding sugars: mass spectrometric advances in the analysis of the sugar alphabet. *Mass Spectrometry Reviews*, 45(2), 282-333.
4. **van Ede, J. M.**, van der Steen, S., van der Kraan, G. M., van Loosdrecht, M. C., & Pabst, M. (2025). Discovery of microbial glycoside hydrolases via enrichment and metaproteomics. *RSC Chemical Biology*, 6(11), 1809-1821.
5. **van Ede, J.***, Pérez-Watson, T.*, van Loosdrecht, M. C., & Pabst, M. (2025). Measuring Temporal Variations of Nucleotide Pools in Microbial Granular Biofilm Performing Enhanced Biological Phosphorous Removal. *bioRxiv*, 2025-08. (*equal first)

6. Chen, L. M., Beck, P., **van Ede, J. M.**, Pronk, M., van Loosdrecht, M. C., & Lin, Y. (2024). Anionic extracellular polymeric substances extracted from seawater-adapted aerobic granular sludge. *Applied Microbiology and Biotechnology*, 108(1), 144.
7. Echeveste Medrano, M. J., Leu, A. O., Pabst, M., Lin, Y., McIlroy, S. J., Tyson, G. W., **van Ede, J. M.**, Sánchez-Andrea, I., Jetten, M. S. M., Jansen, R. & Welte, C. U. (2024). Osmoregulation in freshwater anaerobic methane-oxidizing archaea under salt stress. *The ISME Journal*, 18(1), wræ137.
8. Hu, Z. X., Zhang, J. Y., **Van Ede, J.**, Zhang, Y. Y., Li, Y. Q., Ghirardello, M., Galan, M.C., Pabst, M., Liu, L. & Voglmeir, J. (2026). Identification of a distinct sialic acid (KDN) and a KDN-specific aldolase in Pacific oyster. *Glycoconjugate Journal*, 43(1), 1.
9. Kleikamp, H. B., Van der Zwaan, R., Van Valderen, R., **Van Ede, J. M.**, Pronk, M., Schaasberg, P., Allaart, M. T., van Loosdrecht, M. C. M. & Pabst, M. (2024). NovoLign: metaproteomics by sequence alignment. *ISME communications*, 4(1), ycae121.
10. Chen, L. M., Keisham, S., Tateno, H., **van Ede, J.**, Pronk, M., van Loosdrecht, M. C., & Lin, Y. (2023). Alterations of Glycan Composition in Aerobic Granular Sludge during the Adaptation to Seawater Conditions. *ACS Es&t Water*, 4(1), 279-286.

Conference proceedings

1. **van Ede, J. M.**, van der Kraan, G. M., van Loosdrecht, M. C. M., Pabst, M. Mining complex microbial ecosystems for novel glycoside hydrolases using enrichment and multi-omics. 16th Carbohydrate Bioengineering Meeting, Elsinore (Denmark) 2026. Oral presentation.
2. **van Ede, J. M.**, van Loosdrecht, M. C. M., Pabst, M. Profiling the activated sugar repertoire in microbes with narrow-mass-window DIA. NVMS Fall Meeting, Groningen (The Netherlands) 2026. Poster presentation.
3. **van Ede, J. M.**, van Loosdrecht, M. C. M., Pabst, M. Profiling the molecular diversity of activated nonulosonic acids in microbes. SialoGlyco 2024, Lille (France) 2024. Oral presentation.
4. **van Ede, J. M.**, van Loosdrecht, M. C. M., Pabst, M. Exploring the diversity of nonulosonic acids in pathogenic microbes: targets for the development of novel antimicrobials. 2024 5th NVMS/BSMS Conference, Kerkrade (The Netherlands) 2024. Oral presentation.
5. **van Ede, J. M.**, Saha, C., van Loosdrecht, M. C. M., Louwen, R., Pabst, M. The chemical diversity of the nonulosonic acids of a clinical *C. jejuni* isolate and a related Δ Cas9 mutant strain. SialoGlyco 2022, Nagoya (Japan) 2022. Oral presentation.
6. **van Ede, J. M.**, Saha, C., van Loosdrecht, M. C. M., Louwen, R., Pabst, M. The flagellum sugar modifications of the pathogen *Campylobacter jejuni*. NWO CHAINS 2022, Veldhoven (The Netherlands) 2022. Oral presentation.
7. **van Ede, J. M.**, Revealing the Physicochemical Dynamics in *Campylobacter Jejuni* Flagellum Sialylation. Society for Glycobiology, San Diego (US) 2021. Oral presentation (selected poster talks).

