

Chemical and Biological Approaches Volume 43

Carbohydrate Chemistry

Chemical and Biological Approaches

Volume 43

Editors

Amelia Pilar Rauter, *Universidade de Lisboa, Portugal*

Thisbe K. Lindhorst, *Christiana Albertina University of Kiel,
Germany*

Yves Queneau, *Université de Lyon, France*

Authors

Ana Luísa Carvalho, *Universidade Nova De Lisboa, Portugal*

Thierry Benvegna, *Ecole Nationale Supérieure de Chimie de Rennes and
Université Bretagne Loire, France*

Richard Daniellou, *University of Orléans, France*

Vincent Ferrières, *Ecole Nationale Supérieure de Chimie de Rennes,
France*

Ravneet Kaur Grewal, *Sri Guru Granth Sahib World University, India*

Yanlong Gu, *Huazhong University of Science and Technology, China*

Pierre Lafite, *University of Orléans, France*

Laurent Legentil, *Ecole Nationale Supérieure de Chimie de Rennes,
France*

Loïc Lemiègre, *Ecole Nationale Supérieure de Chimie de Rennes and
Université de Bretagne Loire, France*

Filipa Marcelo, *Universidade Nova De Lisboa, Portugal*

Alberto Marra, *Institut des Biomolécules Max Mousseron, France*

Laurence A. Mulard, *Institut Pasteur, France*

Angelina S. Palma, *Universidade Nova De Lisboa, Portugal*

Freddy Pessel, *Ecole Nationale Supérieure de Chimie de Rennes and
Université Bretagne Loire, France*

Cédric Peyrot, *University of Orléans, Ecole Nationale Supérieure de
Chimie de Rennes and Université de Bretagne Loire, France*

Benedita A. Pinheiro, *Universidade Nova De Lisboa, Portugal*

Palanisamy Ravichandiran, *Huazhong University of Science and
Technology, China*

Louise Renault, *Ecole Nationale Supérieure de Chimie de Rennes and Université Bretagne Loire, France*

Diana O. Ribeiro, *Universidade Nova De Lisboa, Portugal*

Paula Alexandra Videira, *Universidade Nova De Lisboa, Portugal*

Renaud Zelli, *Université Grenoble Alpes, France*



ISBN: 978-1-78801-003-0
PDF ISBN: 978-1-78801-064-1
EPUB ISBN: 978-1-78801-409-0
ISSN: 0306-0713
DOI: 10.1039/9781788010641

A catalogue record for this book is available from the British Library

© The Royal Society of Chemistry 2018

All rights reserved

Apart from fair dealing for the purposes of research for non-commercial purposes or for private study, criticism or review, as permitted under the Copyright, Designs and Patents Act 1988 and the Copyright and Related Rights Regulations 2003, this publication may not be reproduced, stored or transmitted, in any form or by any means, without the prior permission in writing of The Royal Society of Chemistry or the copyright owner, or in the case of reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency in the UK, or in accordance with the terms of the licences issued by the appropriate Reproduction Rights Organization outside the UK. Enquiries concerning reproduction outside the terms stated here should be sent to The Royal Society of Chemistry at the address printed on this page.

Whilst this material has been produced with all due care, The Royal Society of Chemistry cannot be held responsible or liable for its accuracy and completeness, nor for any consequences arising from any errors or the use of the information contained in this publication. The publication of advertisements does not constitute any endorsement by The Royal Society of Chemistry or Authors of any products advertised. The views and opinions advanced by contributors do not necessarily reflect those of The Royal Society of Chemistry which shall not be liable for any resulting loss or damage arising as a result of reliance upon this material.

The Royal Society of Chemistry is a charity, registered in England and Wales, Number 207890, and a company incorporated in England by Royal Charter (Registered No. RC000524), registered office: Burlington House, Piccadilly, London W1J 0BA, UK, Telephone: +44 (0) 207 4378 6556.

Visit our website at www.rsc.org/books

Printed in the United Kingdom by CPI Group (UK) Ltd, Croydon, CR0 4YY

Preface

DOI: 10.1039/9781788010641-FP007

This volume, 43 of the series Carbohydrate Chemistry, illustrates the diversity of the field of the glycosciences, encompassing chemical and biological approaches towards a wide range of applications, from potential novel therapeutic strategies to novel chemical products and processes for everyday life.

In Chapter 1, Alberto Marra and Renaud Zelli provide a rich and complete overview of iminodisaccharides and iminooligosaccharides. They cover *O*-, *S*-, *N*-, and *C*-linked iminodisaccharides for which they have reviewed the chemical and enzymatic synthetic strategies. These compounds are mostly studied for their glycosidases or glycosyltransferases inhibition properties. The authors finally provide a useful listing of all types of enzymes which have been addressed in the literature related to iminooligosaccharides since 1985.

Laurence Mulard, in Chapter 2, addresses the tremendous structural diversity of surface bacterial polysaccharides and their implications in vaccine development. The chapter focuses on the subtle *O*-acetylation pattern of microbial polysaccharides, in relation to their properties to act as shields against environmental assaults at the primary interface with the host. The chapter is illustrated with structures that are part of pre-clinical programs for vaccine development or components of marketed vaccines.

The regioselectivity of the glycosylation reaction is reviewed in Chapter 3, Vincent Ferrieres and Laurent Legentil look at both enzymatic and chemical strategies. The authors first discuss the benefits of the use of glycosyltransferases, glycosylhydrolases and glycosylphosphorylases as biocatalysts towards defined oligosaccharides, showing how enzyme engineering techniques can improve the efficiency and regioselectivity. With respect to the chemical approaches, the authors address the key issues, namely the reactivity order of various hydroxy groups, the importance of hydrogen bonding, armed/disarmed character of donors or their partial protection, covering the essentials of this fundamental reaction in carbohydrate chemistry.

Chapter 4 gives a complete overview of glycosyltransferases inhibitors, a topic with considerable importance to provide novel strategies for fighting abnormal glycosylation diseases. In their review, Paula Videira, Filipa Marcelo and Ravneet Grewal discuss the conventional specific inhibitors, notably those designed as mimics of donor, acceptor or transition state. Recently proposed alternative chemotype designs, not substrate like, are also discussed, illustrated with promising examples that could lead to therapeutic developments.

Carbohydrate microarrays are high-throughput and sensitive tools which can uncover complex carbohydrate structures and protein-carbohydrate interactions. In Chapter 5, Angelina Palma, together with Diana Ribeiro, Benedita Pinheiro and Ana Luisa Carvalho, beautifully

illustrate the power of these tools for the determination of plant cell wall structure and degradation pathways. The chapter reviews nicely the literature on cell wall structure and degradation, offers a complete overview on glycan microarrays and on the existing platforms and specific chemical modifications required for the grafting of the carbohydrates on the surfaces. The authors illustrate the efficiency of the method with numerous examples, including efficient coupling technologies with mass spectroscopy or high-performance thin layer chromatography.

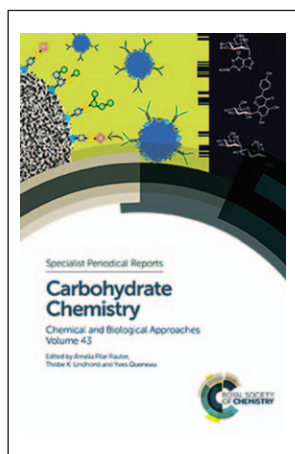
In the search for cleaner chemical processes, the solvent is a critical issue. In Chapter 6, Yanlong Gu and Palanisamy Ravichandiran show that carbohydrates can be useful components of novel media for chemical reactions. They review particularly the properties and uses of combinations of carbohydrates and organic or inorganic salts leading to low-melting mixtures which can then be used as the medium for various organic reactions showing specific chemical activation abilities.

Following the same philosophy which is to develop innovative eco-friendly chemicals and processes from carbohydrates, Thierry Benvegnu, together with Louise Renault and Freddy Pessel, have reviewed in Chapter 7 the field of carbohydrate-based surfactants. After an overview covering the most typical kinds, they focus on compounds constructed with uronic building blocks widely available from marine resources, referring to them as “blue” (and green) novel surfactants. They also address other novel building blocks, such as isosorbide and trehalose, as well as more elaborate systems which comprise a spacer between the hydrophilic and the hydrophobic moieties. Finally, they review recent work on the direct transformation of polysaccharides, such as cellulose, starch, xylan or alginates to various types of surfactants (alkylpolyglycosides or glycuronamides), or even directly from lignocellulose to lignocellulosic hydrolysate fatty esters.

In keeping with the field of physicochemical properties of amphiphilic systems built on carbohydrates, Richard Daniellou, together with Cédric Peyrot, Pierre Lafite and Loïc Lemiègre address the field of formation of hydrogels using low molecular weight hydrogelators. They give an overview of the main types of monosaccharidic or disaccharidic hydrogelators, as well as bolaamphiphilic systems, discussing their syntheses and some of their properties in relation to their structure.

We hope that readers will enjoy this volume, and that the wide scope of contributions, arising from synthetic, biological, structural, and applied motivations, will stimulate transdisciplinary approaches towards novel developments in the glycosciences.

Yves Queneau
Amélia P. Rauter
Thisbe K. Lindhorst



Cover

Front cover image courtesy of Ana Marta de Matos.

 Preface

vii

Synthesis and biological properties of imino-disaccharides and -oligosaccharides **1**

Alberto Marra and Renaud Zelli

1 Introduction	1
2 Synthesis of imino- <i>O</i> -oligosaccharides	3
3 Synthesis of imino- <i>S</i> -disaccharides	30
4 Synthesis of imino- <i>N</i> -oligosaccharides	31
5 Synthesis of imino- <i>C</i> -disaccharides	36
6 Glycosidase inhibition properties	60
7 Conclusions	66
References	66

Bacterial polysaccharides as major surface antigens: interest in O-acetyl substitutions **71**

Laurence A. Mulard

1 Introduction	71
2 PS O-acetylation: a widespread modification	72
3 On the role of CPS O-acetylation on the host-pathogen crosstalk	76

4 O-Acetylated CPSs from pathogenic bacteria: implication in vaccine development	81
5 Conclusion	97
References	97

Regioselective glycosylation: What's new? **104**

Laurent Legentil and Vincent Ferrières

1 Introduction	104
2 Tackling the regioselectivity problem using biocatalysts	105
3 Bioinspired chemical glycosylation	118
4 Conclusion	129
References	130

Glycosyltransferase inhibitors: a promising strategy to pave a path from laboratory to therapy **135**

Paula Alexandra Videira, Filipa Marcelo and Ravneet Kaur Grewal

1 Glycan biosynthesis and glycan processing enzymes	135
2 Small molecule inhibitors to modulate glycan processing enzymes	136
3 Glycosyltransferase inhibitors	138
4 Substrate analogs	139
5 Transition state analogs	143
6 Glycomimetics	145
7 Alternate chemotype analogs	147
8 Metabolic chain terminator	151
9 Conclusion	154
References	154

Targeting protein-carbohydrate interactions in plant cell-wall biodegradation: the power of carbohydrate microarrays **159**

Diana O. Ribeiro, Benedita A. Pinheiro, Ana Luísa Carvalho and Angelina S. Palma

1 Structural diversity of plant cell-wall polysaccharides	159
2 Cellulolytic microorganisms express proteomes highly efficient in plant cell-wall biodegradation	162
3 Carbohydrate microarrays	166
4 Conclusions	172
Acknowledgements	173
References	173

Low melting carbohydrate mixtures and aqueous carbohydrates – an effective green medium for organic synthesis 177

Palanisamy Ravichandiran and Yanlong Gu

1 Introduction	177
2 Organic Transformations in carbohydrate-based low melt	180
3 Organic Transformations in an aqueous Sugar	185
4 Conclusions and Outlook	193
References	193

Surfactants based on green/blue sugars: towards new functionalities in formulations 196

Louise Renault, Freddy Pessel and Thierry Benvegnu

1 Introduction	196
2 Sugar-based ionic surfactants	199
3 Surfactants based on novel carbohydrate motifs	204
4 Sugar-based surfactants including original linkers between polar and lipophilic domains	214
5 Surfactant compositions resulting from the direct transformation of polysaccharides	232
6 Conclusion	240
Abbreviations	240
References	241

Low molecular weight carbohydrate-based hydrogelators 245

Cédric Peyrot, Pierre Lafite, Loïc Lemiègre and Richard Daniellou

1 Introduction	245
2 Monosaccharides	246
3 Disaccharides	255
4 Bipolar hydrogelators	258
5 Rationalisation and structure/property relationships	261
6 Conclusion	261
References	261

Synthesis and biological properties of imino-disaccharides and -oligosaccharides

Alberto Marra*^a and Renaud Zelli^b

DOI: 10.1039/9781788010641-00001

Since 1985, various O-, S-, N- and C-linked imino-disaccharides and imino-oligosaccharides have been designed as potent and more selective glycosidase inhibitors.

1 Introduction

Iminosugars,¹ in the past erroneously called azasugars,² are polyhydroxylated monocyclic (pyrrolidine, piperidine, azepane) and bicyclic (pyrrolizidine, indolizidine, nortropane) nitrogenated compounds that can be considered carbohydrate analogues bearing a basic nitrogen instead of the endocyclic oxygen atom (Fig. 1). These naturally occurring products are strong inhibitors of both glycosidases, the enzymes that catalyse the cleavage of glycosidic bonds in oligosaccharides and glycoconjugates, and glycosyltransferases, the enzymes that catalyse the formation of the glycosidic bond starting from an activated sugar donor. Besides the sterical and stereochemical resemblance to sugars, their inhibition activity arises from the protonated endocyclic nitrogen, at physiological pH, which leads to strong electrostatic interactions with the carboxylate ion located in the active site of the enzyme.

In order to find new treatments for the severe pathologies originated by a malfunction of the above sugar processing enzymes, many synthetic monosaccharidic iminosugar have been prepared over the last four decades.¹ However, also designed and synthesized were less conventional derivatives such as the iminosugar clusters and the imino-disaccharides and -oligosaccharides. While the former class of compound has been extensively reviewed,³ only two review articles and a book's chapter have been dedicated to the latter family of iminosugars, two dealing⁴ exclusively with the carbon-linked disaccharides (imino-C-disaccharides, see section 5), the other focused⁵ mainly on the synthesis and biological properties of the imino-O-disaccharides (see section 2). The interest in the di- and oligosaccharidic iminosugar, *i.e.* carbohydrates constituted of an iminosugar moiety linked to one or more sugar units, resides in their expected stronger and more selective glycosidases inhibition. Indeed, these enzymes are not totally selective for the monosaccharide (*e.g.* D-glucose, D-mannose, *etc.*) and the anomeric linkage (α or β) to be cleaved, and thus also the iminosugar-based inhibitors are poorly

^aInstitut des Biomolécules Max Mousseron (IBMM), UMR 5247, Ecole Nationale Supérieure de Chimie de Montpellier, 8 Rue de l'Ecole Normale, 34296 Montpellier Cedex 5, France. E-mail: Alberto.Marra@umontpellier.fr

^bUniversité Grenoble Alpes, CNRS, DPM UMR 5063, CS40700, 38058 Grenoble, France. E-mail: renaud.zelli@univ-grenoble-alpes.fr

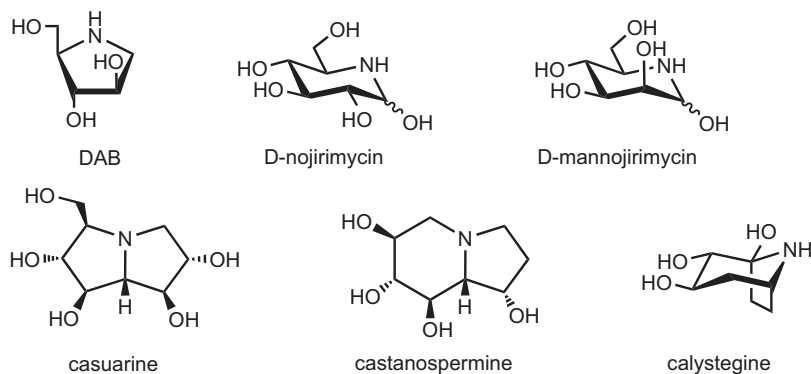


Fig. 1 Structures of some naturally occurring iminosugars.

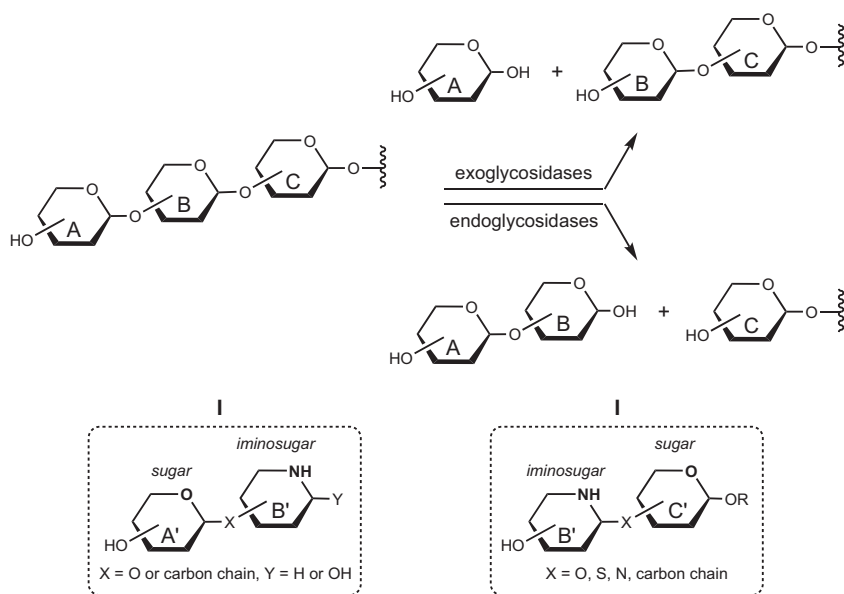


Fig. 2 Oligosaccharide hydrolysis catalysed by glycosidases and their iminodisaccharide-based inhibitors.

selective and scarcely used in therapy. Nevertheless, many glycosidases are endowed with some aglycon specificity, *i.e.* they selectively recognize the sugar(s) linked to the monosaccharide to be hydrolysed. Therefore, compounds bearing a mimic of the glycon hydrolysis intermediate (the iminosugar) and a natural mono- or oligosaccharide are good candidates for the highly selective glycosidase inhibition.

During the last three decades, two series of imino-disaccharides and -oligosaccharides have been synthesized, one featuring the imino-sugar (or 1-deoxy-iminosugar) moiety at the reducing end (**I**, Fig. 2), the other displaying the iminosugar at the non-reducing end (**II**). Interestingly, in the synthetic disaccharides belonging to the latter series, the polyhydroxy-piperidine (or -pyrrolidine) is linked to the sugar unit

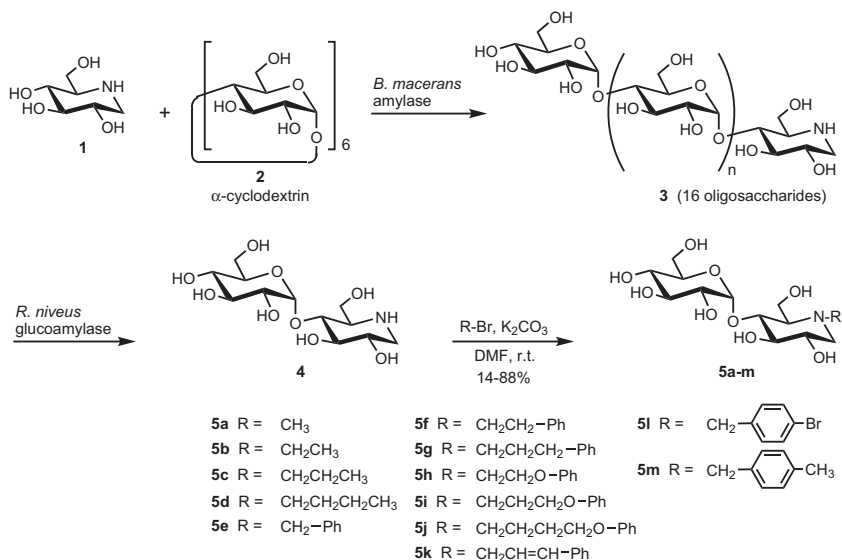
through *O*-, *S*-, *N*- and *C*-(pseudo)glycosidic bonds whereas in the other series, *i.e.* **I**, only oxygen- and carbon-linked iminodisaccharides are known. However, many of these imino-*C*-disaccharides were not genuine isomers, *i.e.* the interglycosidic oxygen atom of the naturally occurring disaccharides was replaced by various carbon chains instead of a single methylene group. Finally, along with actual *N*-glycoside derivatives, a few disaccharide analogues that contain endocyclic nitrogen-tethered moieties have been recently synthesized (see section 4.2).

2 Synthesis of imino-*O*-oligosaccharides

The oxygen-linked glycosides being the first iminodisaccharides reported in the literature, this family of compounds is constituted of a large number of members, mainly prepared by means of conventional glycosylation methods. However, the vast majority of them are di- and oligosaccharides bearing the iminosugar moiety at the reducing end, and only 4 imino-*O*-disaccharides bearing the sugar unit at the reducing end (described in 3 articles) are known to date.

2.1 Imino-*O*-oligosaccharides bearing the iminosugar moiety at the reducing end

2.1.1 Enzymatic synthesis of imino-*O*-oligosaccharides. In 1985, Ezure published⁶ the first synthesis of an iminodisaccharide, the 4-*O*-(α -D-glucopyranosyl)-1-deoxynojirimycin **4** (Scheme 1) obtained in a multigram scale from 1-deoxynojirimycin (**1**) and α -cyclodextrin (**2**) by two glycosidase-promoted reactions, namely an *O*-transglycosylation (catalysed by the *Bacillus macerans* amylase) followed by hydrolysis of the glycosidic bond between two glucose units (catalysed by the *Rhizopus niveus* glucoamylase). The disaccharide **4** was then employed



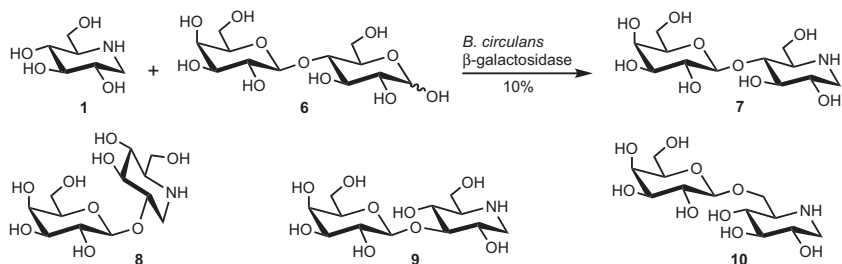
Scheme 1

to prepare^{7,8} a series of 13 *N*-alkylated derivatives **5a–m** which were submitted to various biological assays (see section 6).

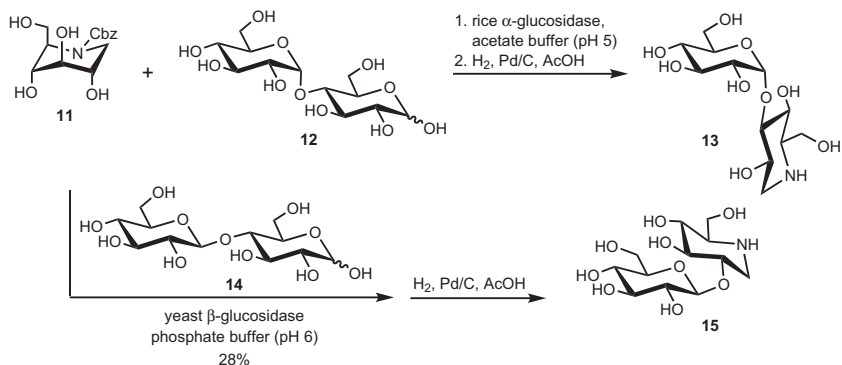
The following year, the iminodisaccharide **4** as well as seven oligosaccharides **3** ($n = 1–7$) were also prepared⁹ by Kainosho and co-workers by transglycosylation of starch with 1-deoxynojirimycin (**1**) in the presence of bacterial saccharifying amylase (pH 5.5, 37 °C, 16 hours).

More than one decade later, Ezure and co-workers reported¹⁰ on the multikilogram scale synthesis of β -D-galactopyranosyl-1-deoxynojirimycin derivatives catalysed by *Bacillus circulans* β -galactosidase (Scheme 2). Thus, the incubation at 40 °C for 18 hours of 1-deoxynojirimycin **1** (5 kg) and D-lactose **6** (50 kg) in the presence of the enzyme led mainly to the 1,4-linked imino-*O*-disaccharide **7**. Also isolated and fully characterized were the 1,2- (**8**), 1,3- (**9**) and 1,6-linked (**10**) regioisomers (yields not given).

Another study on the transglycosylation catalysed by glycosidases was published¹¹ in 1994 by Asano and co-workers. They used an immobilized rice α -glucosidase to promote the reaction between a large excess of maltose (**12**) and the Cbz-protected 1-deoxynojirimycin **11** which, being a carbamate, did not act as a glycosidase inhibitor (Scheme 3). After 2 hours of incubation at 37 °C, the main product[†] was found to be the α -1,3



Scheme 2



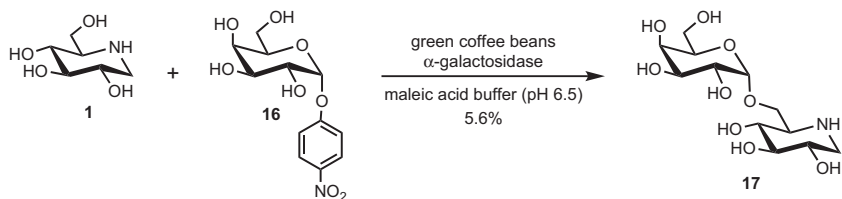
Scheme 3

[†]Unfortunately, due to a series of mistakes in the Experimental section of the article, the isolated yield cannot be determined by the reader.

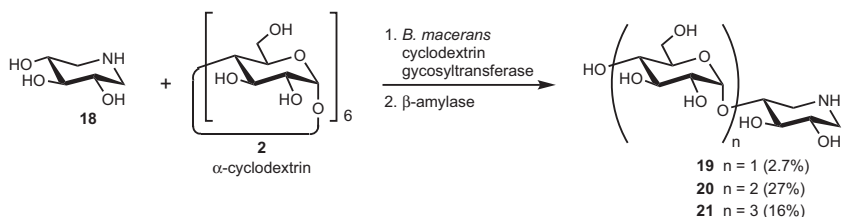
disaccharide that, after hydrogenation (yield not given) afforded the de-protected imino-*O*-disaccharide **13**. The transglycosylation gave also the α -1,4 and α -1,2 regioisomeric disaccharides that could be isolated.[†] The native (*i.e.* not immobilized) enzyme led to similar results. On the other hand, the α -1,2 isomer was the main product when native yeast α -glucosidase was used as the catalyst while the immobilization of this enzyme led to a significant decrease of its activity. The β -D stereoisomers were obtained employing native yeast β -glucosidase and cellobiose (**14**), in large excess, as the glycosyl donor. After 3 hours at 37 °C the β -1,2 imino-*O*-disaccharide was isolated in 28% yield and then hydrogenated to give **15** (yield not given). Also isolated was the β -1,4 regioisomer (4%). Upon immobilization, this enzyme was still reactive and the β -1,2 isomer was recovered as the main product. Interestingly, in any case the α - or β -1,6 imino-*O*-disaccharide was formed, probably due to steric hindrance caused by the Cbz protecting group. The disaccharides **13** and **15** were assayed as glycosidase inhibitors (see section 6).

On the other hand, Paek and co-worker found¹² in 1998 that the α -galactosidase from green coffee beans was able to promote the transglycosylation from *p*-nitrophenyl α -D-galactopyranoside (**16**) to the 6-OH of 1-deoxynojirimycin **1** leading to the corresponding 1,6-linked imino-*O*-disaccharide **17** (Scheme 4) as the main product (5.6%).

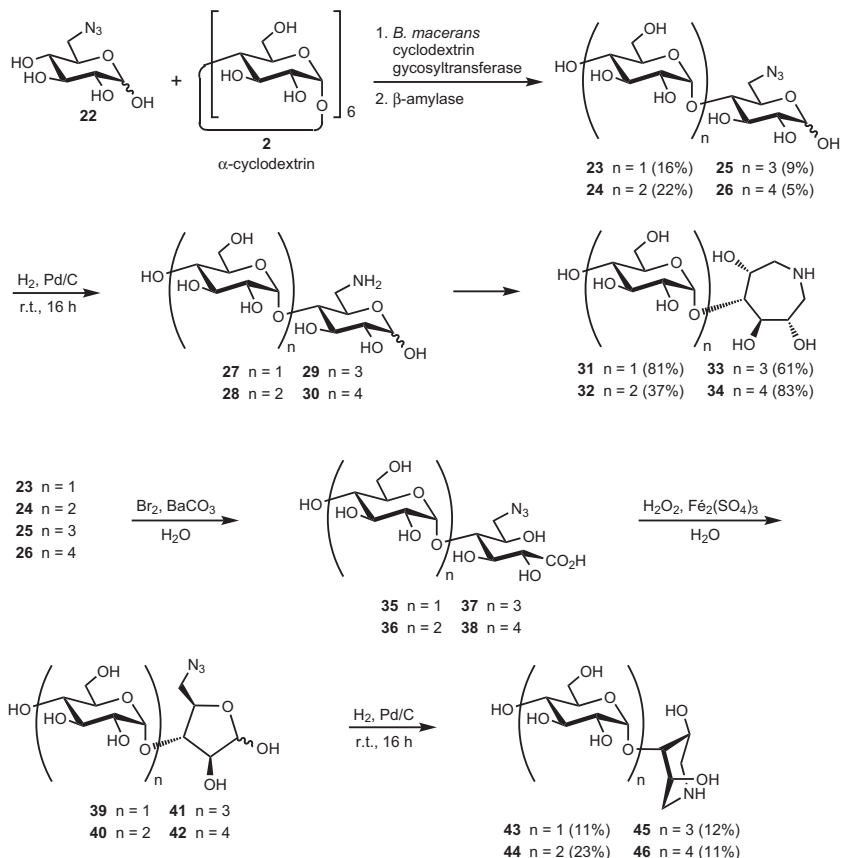
Various di- and oligosaccharides containing three different iminosugar moieties were prepared¹³ from α -cyclodextrin (**2**) by Uchida and co-workers taking advantage of two enzymatic reactions. The first was catalysed by *Bacillus macerans* CGTase (cyclodextrin glycosyltransferase or cyclodextrin glucanotransferase), which is a glycosidase despite the misleading name, whereas the second was promoted by a β -amylase, which catalyses the hydrolysis of the second α -1,4 glucosidic bond releasing β -maltose. Using these two enzymes, the di- (**19**), tri- (**20**) and tetra-saccharides (**21**) were isolated and characterized starting from the known¹⁴ D-xylo configured 1-deoxy-iminosugar **18** (Scheme 5).



Scheme 4



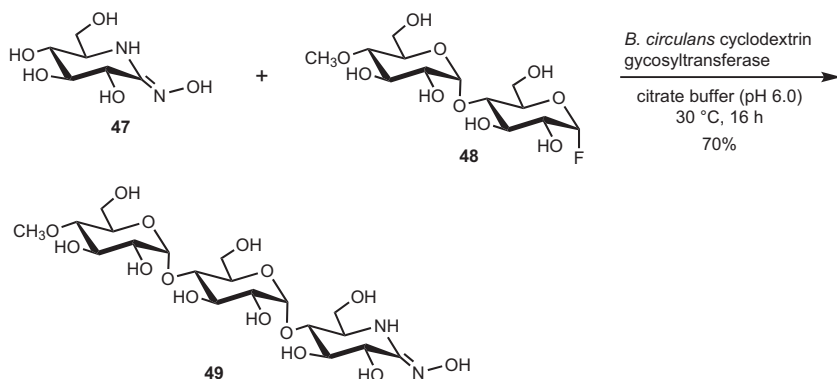
Scheme 5



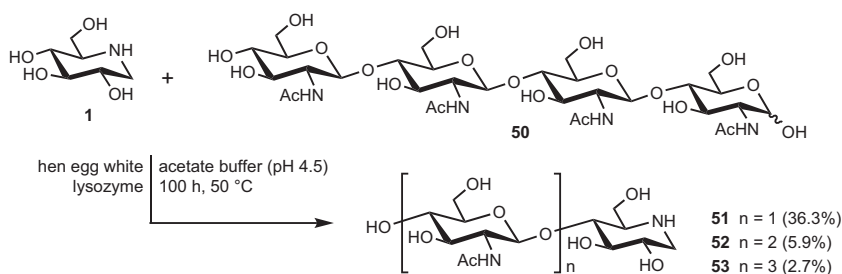
Scheme 6

When the same enzymatic reactions were carried out using 6-azido-6-deoxy-D-glucose (22, Scheme 6) as the glycosyl acceptor, the four compounds 23–26 were isolated.¹³ Catalytic hydrogenation of each of the latter products led to reductive aminocyclisation affording the glucosyl-azepine derivatives 31–34 in good yield. On the other hand, oxidation of 23–26 with bromine in the presence of barium carbonate gave the corresponding aldonic acid intermediates 35–38 which were directly submitted to oxidative decarboxylation (H_2O_2 , $\text{Fe}_2(\text{SO}_4)_3$) to produce the 5-azido-5-deoxy-arabinofuranose derivatives 39–42. These products were individually hydrogenated to afford the disaccharide 43 and the three oligosaccharides 44–46 bearing a D-arabino configured 1-deoxy-iminosugar unit at the reducing end. Most of these compounds were found to be inhibitors of amylases (see section 6).

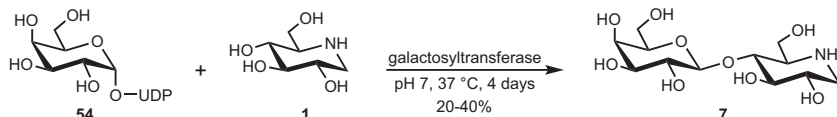
The CGTase enzyme (from *Bacillus circulans*) was also exploited¹⁵ by Whithers and co-workers to prepare the trisaccharide 49 (Scheme 7), containing a hydroximolactam unit at the reducing end, as a new α -amylase inhibitor (see section 6). Incubation of the 4'-O-methyl- α -D-maltosyl fluoride 48 (1.6 mg), obtained¹⁶ in 7 steps from D-maltose, with the known¹⁷ hydroximolactam 47 (1.0 mg) gave, after column chromatography, 49 in 70% yield (1.7 mg).



Scheme 7



Scheme 8



Scheme 9

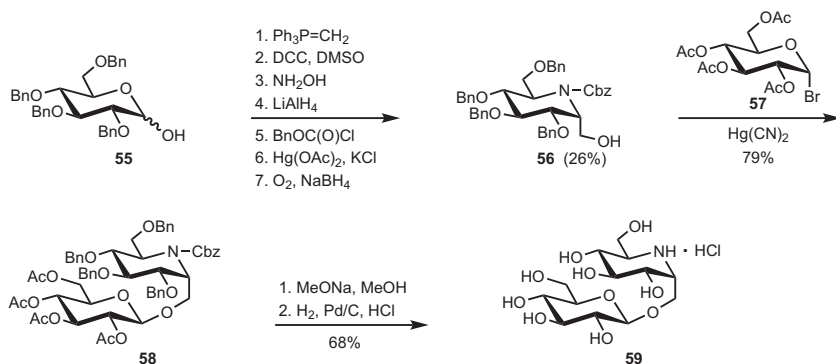
In a recent paper,¹⁸ Usui, Fukamizo and their co-workers described the synthesis of 4-*O*-(2-acetamido-2-deoxy- β -D-glucopyranosyl)-1-deoxynojirimycin (**51**, Scheme 8) *via* transglycosylation of the commercially available tetra-*N*-acetyl-chitotetraose **50** with 1-deoxynojirimycin (**1**) catalysed by hen egg white lysozyme (HEWL). Although the iminodisaccharide **51** was the main reaction product, the corresponding tri- (**52**) and tetra-saccharide (**53**) were also isolated and characterized.

The only example of use of glycosyltransferases, instead of glycosidases, for the synthesis of imino-*O*-disaccharides was reported¹⁹ by Gautheron-Le Narvor and Wong in 1991 (Scheme 9). The coupling of uridine diphosphate D-galactose (UDP-Gal, **54**) with 1-deoxynojirimycin (**1**) in the presence of β -1,4-galactosyltransferase (EC 2.4.1.22) afforded, after 4 days, the disaccharide **7** in 20–40% yield.

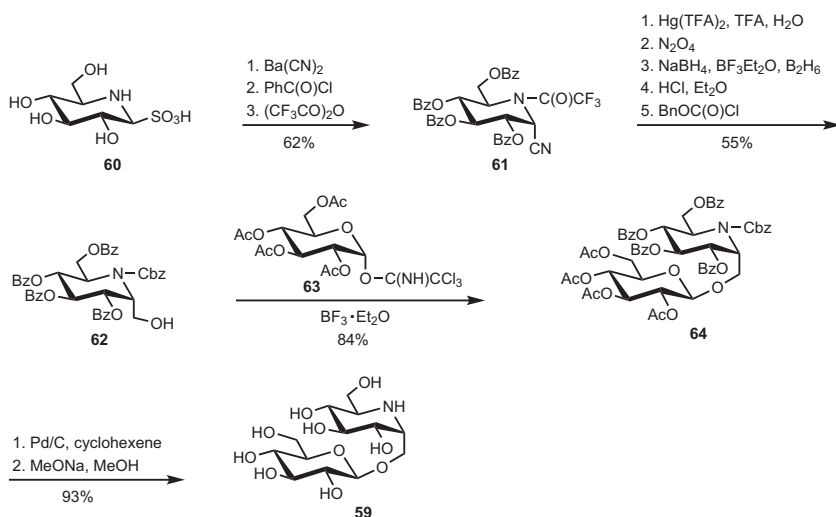
2.1.2 Chemical synthesis of imino-*O*-disaccharides. The first non-enzymatically synthesized imino-*O*-disaccharide was prepared²⁰ by Liu in 1987 by direct glycosylation of the iminoheptitol **56**, obtained from

the tetra-*O*-benzyl-D-glucopyranose **55** (7 steps, 26% overall yield), with the glucopyranosyl bromide **57** (Scheme 10). The presence of a participating group at the C-2 position of the glycosyl donor insured the totally stereoselective synthesis of the protected disaccharide **58** from which the desired imino-*O*-disaccharide **59** was obtained as hydrochloride salt after standard deprotection reactions. The latter compound was found to be a potent inhibitor of various glycosidases (see section 6).

Two years later, the same Author developed²¹ another synthesis of the imino-*O*-disaccharide **59** starting from the known²² 1-deoxynojirimycin-1-sulfonic acid **60** (Scheme 11). The glycosyl acceptor **62** was prepared through a rather long reaction sequence based on the key iminosugar nitrile intermediate **61**. Then, the glycosylation of **62** by the tri-chloroacetimidate **63** was fully stereoselective affording the β -D disaccharide **64** in 84% yield. Finally, hydrogenation and transesterification gave the imino-*O*-disaccharide **59** in 93% total yield.



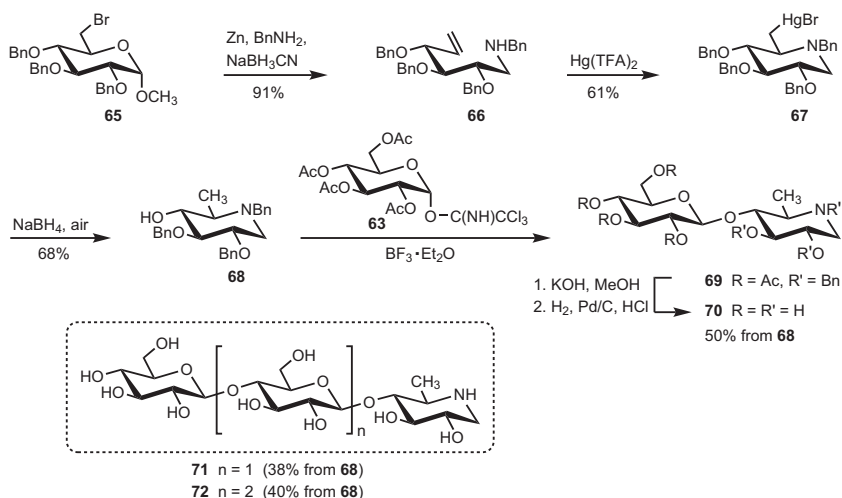
Scheme 10



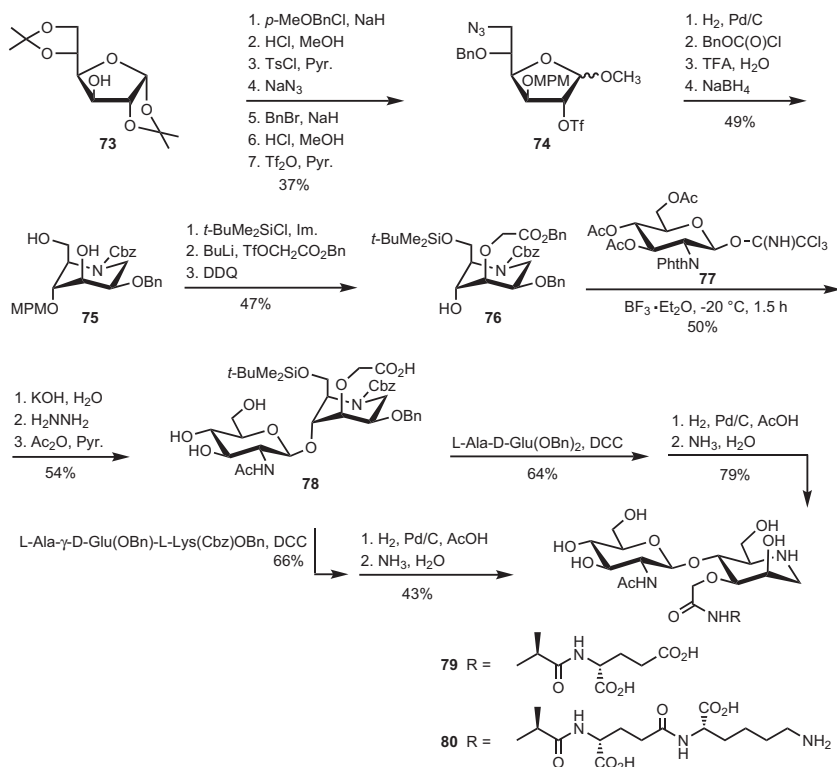
Scheme 11

The trichloroacetimidate-based glycosylation was also employed²³ by Ganem and co-workers to synthesize di-, tri- and tetrasaccharides (**70–72**, Scheme 12) containing a 1,6-dideoxy-nojirimycin moiety. The suitably protected iminosugar alcohol derivative **68** was obtained from the 6-bromo-glucoside **65** through reductive ring opening followed by *in situ* reductive amination to give the aminoalkene²⁴ **66**. The latter was submitted to aminomercuration to afford²⁴ the iminosugar **67** that gave the desired glycosyl acceptor **68** upon carefully optimized²³ reductive oxygenation. The coupling of **68** with the glucosyl imidate **63** afforded the protected imino-*O*-disaccharide **69** from which the target compound **70** was obtained by standard deprotection reactions in 50% total yield (3 steps). Moreover, glycosylation of the iminosugar **68** with cellobiosyl and celotriosyl trichloroacetimidates allowed, after *O*- and *N*-deprotection, to recover the trisaccharide **71** and the tetrasaccharide **72**, respectively, in satisfactory yield (Scheme 12). Compounds **71–72** proved to be efficient glycosidases inhibitors (see section 6).

Another use of trichloroacetimidate derivatives as glycosyl donors was described^{25a} by Moss and Vallance who prepared the iminosugar analogues (**79** and **80**, Scheme 13) of the repeating disaccharide unit found in peptidoglycan, *i.e.* the *N*-acetylglucosamine- β -1,4-*N*-acetylmuramic acid. The 3-OH of the commercially available diacetoneglucose **73** was first protected as *p*-methoxybenzyl ether, then a selective hydrolysis of the 5,6-*O*-isopropylidene group allowed to activate the primary alcohol as tosylate which was replaced by an azido function. Benzylation, followed by methanolysis of the isopropylidene and treatment with triflic anhydride gave the methyl glycoside **74** as an anomeric mixture. Reduction of the azide led to intramolecular cyclisation and, after protection of the amino group, hydrolysis of the acetal and reduction of the aldehyde, to the Cbz-protected 1-deoxymannojirimycin **75**. Silylation of the primary alcohol and installation of the lactate moiety at the 3-OH gave, after removal of the *p*-methoxybenzyl group, to the glycosyl acceptor **76**. Coupling of the



Scheme 12

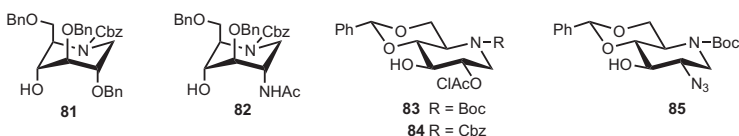


Scheme 13

latter with the 2-phthalimido-glucosyl trichloroacetimidate **77** in the presence of boron trifluoride afforded the corresponding disaccharide in 50% yield. Selective deprotection of this compound gave the free acid **78** which was coupled with a di- and a tri-peptide under standard conditions to afford, after hydrogenolysis and basic treatment, the peptidyl imino-*O*-disaccharides **79** and **80**, respectively. In a following paper, Moss and Southgate reported^{25b} the preparation of the 2-*epi*-acetamido analogues of **79** and **80**, *i.e.* disaccharides bearing a 2-acetamido-1,2-dideoxynojirimycin instead of 1-deoxymannojirimycin unit, using the same synthetic approach. Unfortunately, all compounds did not show antibacterial activity neither were inhibitors of translocases 1 and 2, and transglycosylase, the enzymes involved in peptidoglycan biosynthesis.

The synthesis of six imino-*O*-disaccharides was described by Hasegawa and co-workers.²⁶ They prepared a series of suitably protected 1-deoxynojirimycin glycosyl acceptors having a free OH group at the position 4 (**81** and **82**, Fig. 3) or 3 (**83–85**) that were allowed to react with thioglycoside or glycosyl bromide donors belonging to the *D*-galacto (**86** and **87**, Fig. 3) and *D*-gluco series (**88** and **57**). The Authors proved by NMR analysis that the iminosugar adopts, in both mono- and disaccharide derivatives, a ¹C₄ conformation only when the endocyclic nitrogen is protected as carbamate (Boc or Cbz) and the 4,6-benzylidene protecting group is not present (see Fig. 3).

Glycosyl acceptors:



Glycosyl donors:

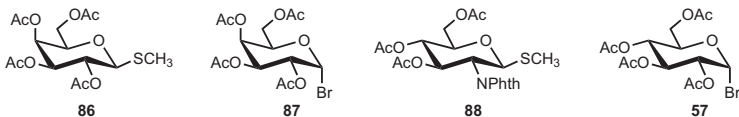
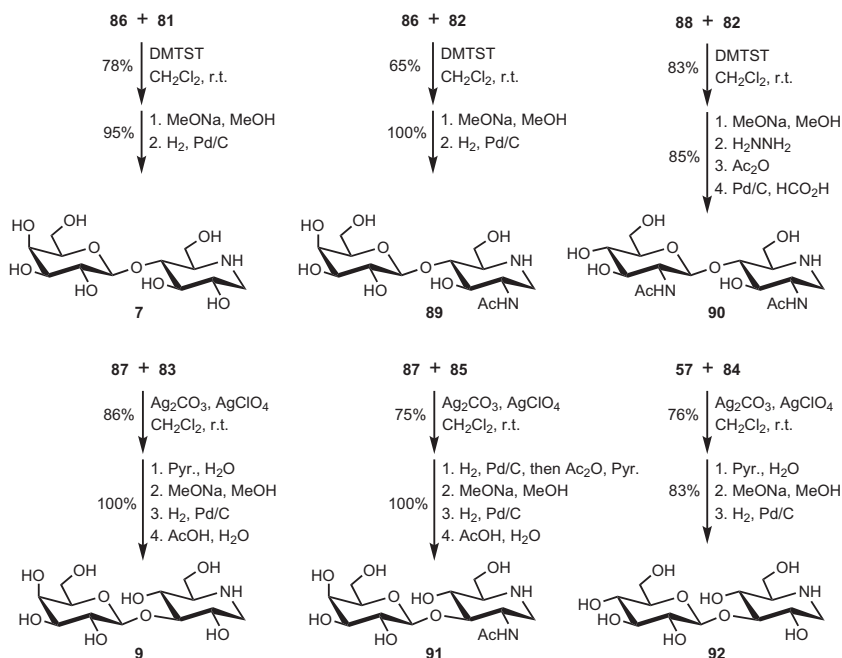


Fig. 3 Glycosyl acceptors and glycosyl donors prepared by Hasegawa and co-workers.²⁶

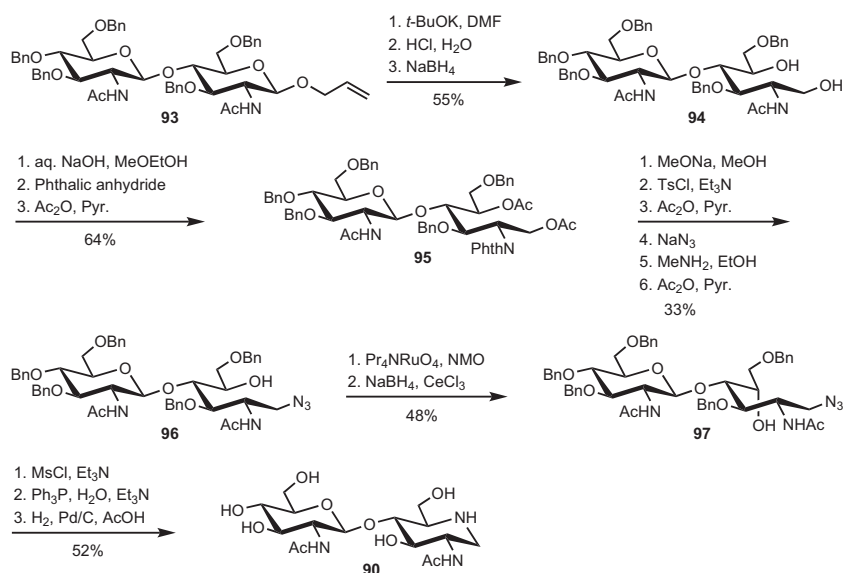


Scheme 14

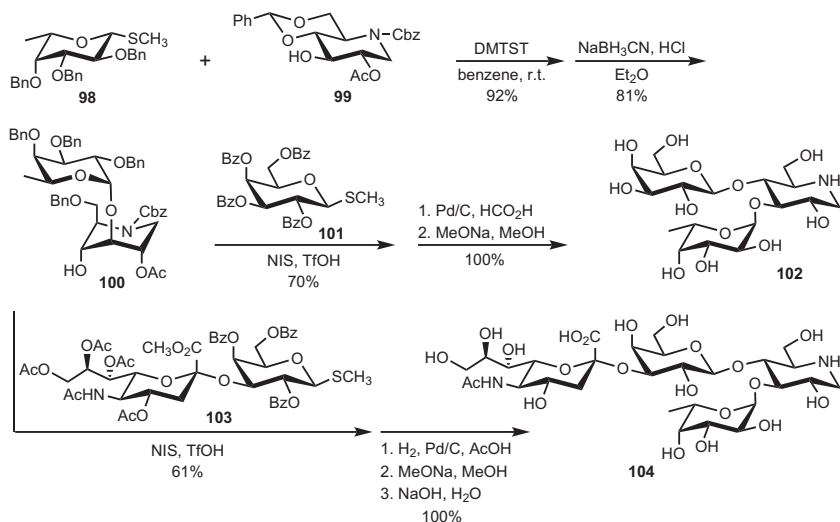
The glycosylation with the thioglycoside donors **86** and **88** were promoted by dimethyl(methylthio)sulfonium triflate (DMTST) whereas the bromide donors were activated by silver carbonate and silver perchlorate (Scheme 14). The *N*- and *O*-protected disaccharides, obtained in good isolated yields (65–86%), were then submitted to classical deprotection reaction sequences in order to remove the carbamate, (chloro)acetate, benzylidene, and phthalimido groups to give the imino-*O*-disaccharides **7**, **9**, and **89–92**. Interestingly, three of them (**89–91**) contained a 2-acetamido-1,2-dideoxy-nojirimycin unit, the bis-aminated disaccharide **90** being an analogue of the naturally occurring chitobiose (GlcNAc-β-1,4-GlcNAc).

An alternative synthesis of the imino-*O*-disaccharide **90** (see Scheme 14), where the 2-acetamido-iminosugar moiety derived from a *N*-acetyl-glucosamine unit, was proposed²⁷ a few years later by Takahashi and co-workers. Starting from the benzylated allyl chitobioside **93** (Scheme 15), easily obtained²⁸ from the *N,N*-diacetylchitobiose disaccharide in turn prepared by partial enzymatic degradation of chitin, the open-chain diol **94** was synthesized *via* allyl ether isomerization, hydrolysis and hemiacetal reduction. Then, the acetamido group at position 2 was selectively removed and replaced by a phthalimido function (**95**) because the tosylation (or mesylation) of the primary alcohol was unsatisfactory. On the other hand, the conversion of **95** into the azido-alcohol **96** was successful (6 steps, 33% overall yield). Oxidation of the 5-OH with tetrapropyl-ammonium perruthenate and *N*-methylmorpholine *N*-oxide (NMO) followed by stereoselective hydride reduction, afforded the epimeric alcohol **97** which, upon mesylation and Staudinger azide reduction, underwent intramolecular aminocyclisation to give, after hydrogenolysis, **90** in 52% total yield.

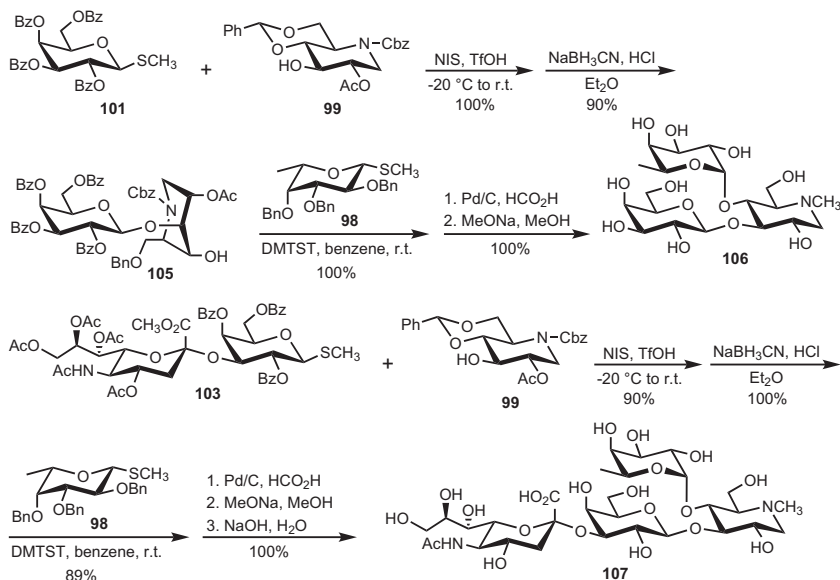
In 1992, Hasegawa and co-workers published²⁹ the synthesis of a trisaccharide and a sialylated tetrasaccharide containing a 1-deoxynojirimycin unit. The iminosugar acceptor **99** (Scheme 16), bearing at the O-2 position an acetate instead of the chloroacetate group present in **84** (see Fig. 3), was glycosylated with the methyl *L*-thiofucoside **98** in the presence of DMTST to give, after benzylidene reductive ring-opening, the disaccharidic acceptor **100**. The latter was glycosylated with the methyl thiogalactoside **101** and the sialyl-thiogalactoside **103** using *N*-iodosuccinimide (NIS) and triflic acid as the promoter to afford, after conventional deprotection steps, **102** and **104**, respectively. The structure of these compounds was related to that of the Lewis X and sialyl-Lewis X,



Scheme 15



Scheme 16



Scheme 17

epitopes recognized by selectins, a pharmacologically important class of lectins.

In another preliminary communications,^{30a} the same Authors described the synthesis of 1-deoxynojirimycin-based tri- (**106**) and tetra-saccharides (**107**) related to the Lewis A and sialyl Lewis A antigens, respectively (Scheme 17). Using the same glycosyl donors (*i.e.* **98**, **101** and **103**) and iminosugar acceptor (**99**) employed for the preparation of the Lewis X and sialyl-Lewis X epitopes (see Scheme 16) but different glycosylation sequences, the target trisaccharide **106** and sialylated

tetrasaccharide **107** were isolated in very high yield. It is worth noting that the catalytic transfer hydrogenolysis (palladium and formic acid in methanol) of the benzyloxycarbonyl group led to the quantitative methylation of the endocyclic nitrogen atom. This alkylation was due to the Pd-catalysed dehydrogenation of methanol to formaldehyde, followed by formation of imine which was hydrogenated to give the *N*-methylated iminosugar.^{30b,31} Therefore, the oligosaccharides **106** and **107** were both recovered as *N*-methyl derivatives. Surprisingly, when the same hydrogenolytic Cbz removal conditions were applied²⁶ to the disaccharide precursor of **90** (see Scheme 14), the 2-acetamido-1,2-dideoxy-nojirimycin moiety of the latter was not methylated.

In the full paper published^{30b} the following year, the Authors described the synthesis of oligosaccharides **102** and **104** (see Scheme 16) as *N*-methyl derivatives. They also reported the preparation of **106** and **107** (see Scheme 17) as well as a new disaccharide (**108**, Fig. 4) and a sialylated trisaccharide (**109**) exploiting the already used glycosyl donors and acceptors.

In 1996, the same Authors prepared³² the sialyl-Lewis X and sialyl-Lewis A epitope analogues **110** and **111** (Fig. 5), *i.e.* the 2-acetamido derivatives of the previously described^{29–31} tetrasaccharides **104** (see Scheme 16) and **107** (see Scheme 17). The syntheses were performed using 2-acetamido-4,6-*O*-benzylidene-*N*-benzyloxycarbonyl-1,2-dideoxy-nojirimycin as the glycosyl acceptor and reaction sequences identical to those displayed in Schemes 16 and 17.

Later on, the *N*-butyl and *N*-decyl derivatives of **110** and **111** (see Fig. 5) were synthesized³³ by Ishida, Kiso and their co-workers. In the same paper, the researchers also described the preparation of the Lewis X analogues **112b–d** (Fig. 6), which are the sulphated, *N*-alkylated derivatives of **102** (see Scheme 16), as well as the Lewis A analogues **113a–c** (Fig. 6), *i.e.* the sulphated, *N*-alkylated derivatives of **106** (see Scheme 17).

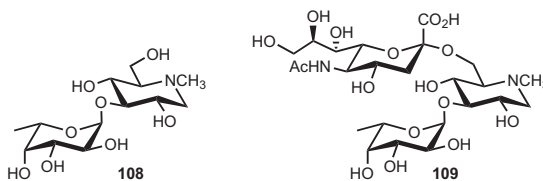


Fig. 4 Imino-disaccharide and -trisaccharide prepared by Hasegawa and co-workers.³¹

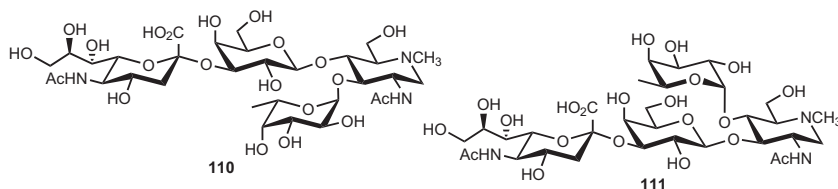


Fig. 5 Sialylated iminotetrasaccharides prepared by Hasegawa and co-workers.³²

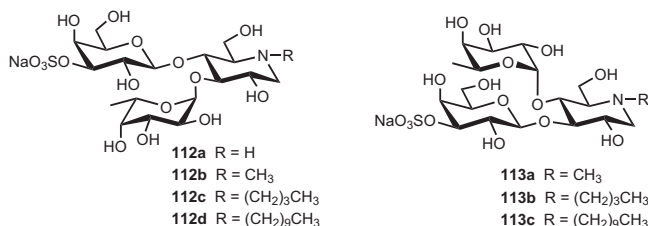
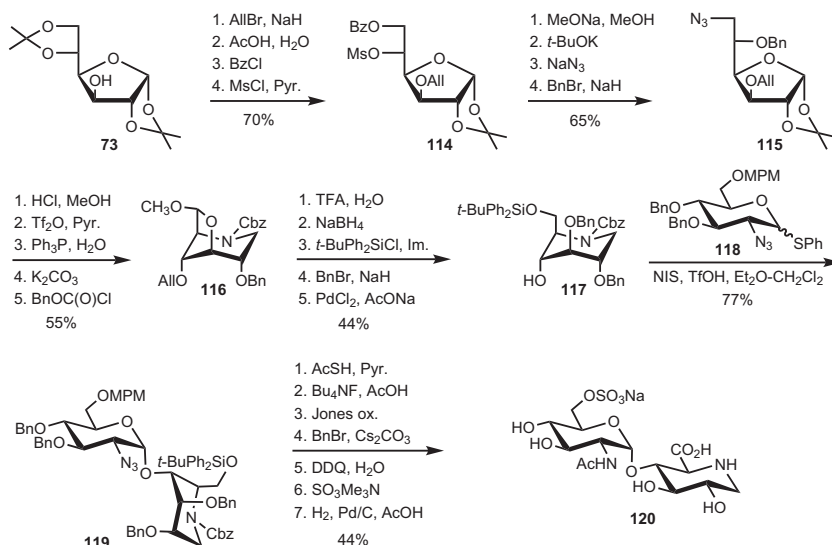


Fig. 6 Sulphated iminotrisaccharides prepared by Ishida, Kiso and co-workers.³³



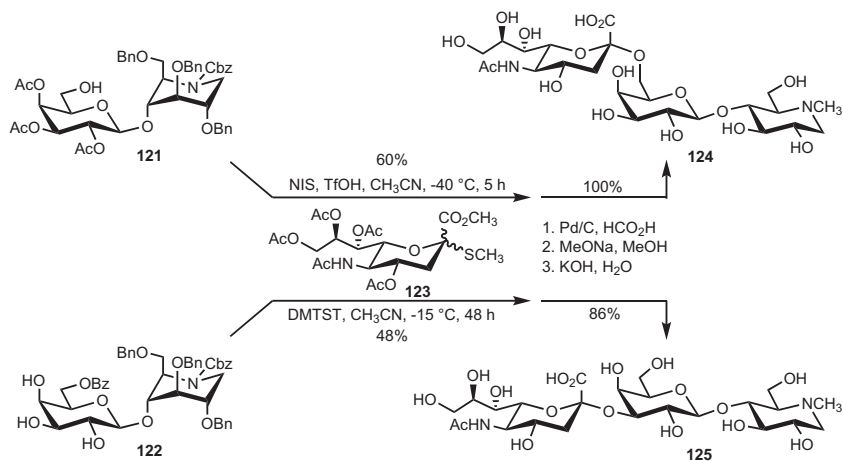
Scheme 18

However, the first sulphated derivative of the trisaccharide **102** was reported³⁴ three years before by Ogawa and co-workers (**112a**, Fig. 6).

A rather unusual sulphated imino-*O*-disaccharide was described by Takahashi and Kuzuhara in a preliminary communication published^{35a} in 1994. Disaccharide **120** (Scheme 18) is indeed constituted of a β -GlcNAc unit α 1,4-linked to a 1-deoxynojirimycin uronic acid derivative. The iminosugar moiety was synthesized from diacetoneglucose **73** by allylation, selective hydrolysis of the 5,6-*O*-isopropylidene group, regioselective benzylation of the primary alcohol, and mesylation of the 5-OH to give **114** in good yield. Upon basic treatment, the latter afforded a 5,6-oxirane intermediate which was reacted with sodium azide to give, after standard benzylation, the *L*-sugar **115** in 65% overall yield. Methanolysis of the 1,2-*O*-isopropylidene group gave an anomeric mixture of methyl furanosides which, however, could be separated and the α -*L* anomer equilibrated to the more stable β -*L* glycoside. Activation of the 2-OH of the latter with triflic anhydride and reduction of the azide to amine allowed, in the presence of potassium

carbonate, the intramolecular aminocyclisation to afford, after treatment with benzyloxycarbonyl chloride, the bicyclic iminosugar **116**. Acidic hydrolysis of the glycoside and aldehyde reduction gave a diol which was regioselectively silylated, then benzylated and the protected derivative submitted to the allyl ether cleavage (PdCl_2 , AcONa , aq. AcOH) to give the glycosyl acceptor **117** in 44% overall yield. This alcohol was glycosylated with the phenyl thioglycoside **118** in the presence of *N*-iodosuccinimide and triflic acid to afford the α -D disaccharide **119** as the main isolated product (77%). After azide reduction and removal of the silyl group, the iminosugar primary alcohol was oxidized to carboxyl group and protected as benzyl ester. Cleavage of the *p*-methoxybenzyl protecting group and sulfation of the GlcNAc 6-OH gave, after catalytic hydrogenolysis, the target imino-*O*-disaccharide **120**. In 2001, the Authors published^{35b} a full paper including all the experimental details as well as an improved synthesis of the same disaccharide **120** based on the oxidation of the primary alcohol of the iminosugar acceptor to the corresponding carboxylic function followed by glycosylation with the same donor **118**.

In 1994, Hasegawa and co-workers reported³⁶ the preparation of a couple of sialylated trisaccharides (**124** and **125**, Scheme 19) featuring structures related to the GM3 ganglioside. The 1-deoxynojirimycin containing glycosyl acceptors **121** and **122** were readily obtained from a disaccharide tetrol precursor²⁶ of **7** (see Scheme 14). The glycosylation of **121** at low temperature with the thiosialoside **123** (mixture of anomers) in the presence of NIS and triflic acid gave the desired 6'-*O*-sialyl-trisaccharide in 60% yield. The corresponding β -D anomer was also isolated (30%). The removal of the Cbz group (palladium in formic acid), followed by transesterification and saponification afforded the fully deprotected trisaccharide **124** in quantitative yield. The regioselective sialylation of the disaccharide triol **122** was carried out using the same glycosyl donor **123** but a different promoter (DMTST). The resulting trisaccharide, isolated in 48% yield, was first peracetylated and then

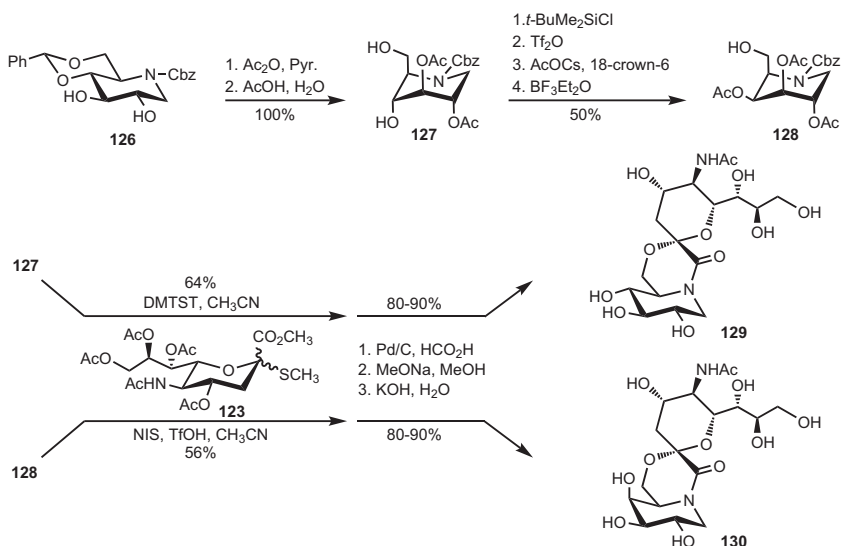


Scheme 19

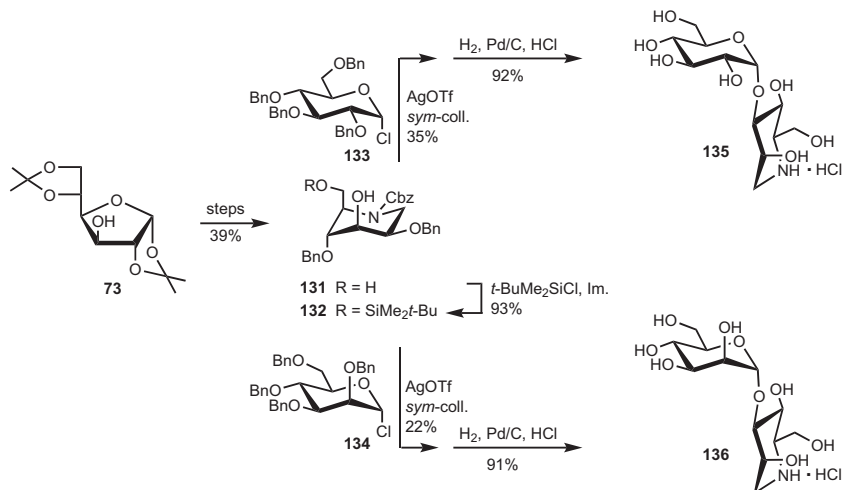
submitted to the above mentioned deprotection sequence to give **125** (86% total yield). Also in this case, the final compounds displayed *N*-methylated iminosugar units.

In a preliminary communication published³⁷ in 1993, the same research team described the formation of uncommon disaccharidic lactams (**129** and **130**, Scheme 20) instead of the expected imino-*O*-disaccharides sialylated at position 6. The glycosyl acceptors were obtained from the protected 1-deoxynojirimycin diol **126** upon acetylation and benzylidene hydrolysis to give first **127** and then, after selective silylation at 6-OH, formation of the 4-O-triflate, inversion of configuration by treatment with cesium acetate and final silyl ether removal, the *D*-galacto configured alcohol **128**. Sialylation of both acceptors with the methyl thiosialoside **123** in the presence of different promoters (DMTST or NIS/TfOH, reaction time and temperature not given) gave the corresponding disaccharides which were separated from the respective β -*D* anomers and deprotected to afford the tricyclic lactams **129** and **130** together with 10–20% of the desired imino-*O*-disaccharides.

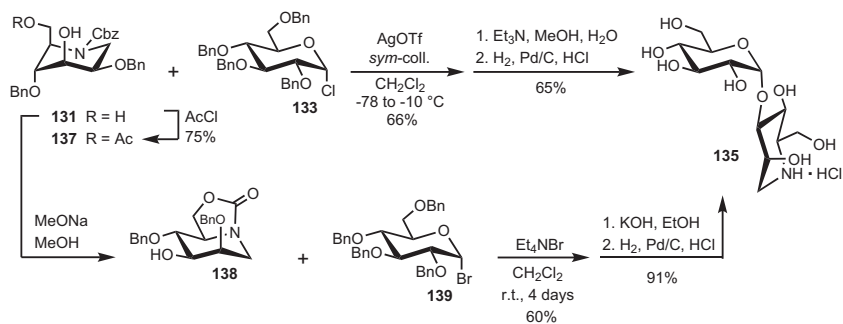
Another imino-*O*-disaccharides bearing 1-deoxy-*D*-mannojirimycin units instead of the more common *gluco* epimer were reported³⁸ by Fleet and co-workers in 1994. The glycosyl acceptor **132** was easily obtained³⁹ from the known diol **131**, in turn prepared on a multigram scale starting from commercially available 1,2:5,6-di-*O*-isopropylidene- α -*D*-glucofuranose **73** (Scheme 21). The glycosylation of **132** with the tetra-benzylated glucosyl (**133**) or mannosyl (**134**) chloride in the presence of silver triflate and *sym*-collidine afforded the corresponding α -*D*-disaccharides in rather poor yields (35 and 22%, respectively). The catalytic hydrogenation in acidic conditions allowed the simultaneous removal of the benzyl, Cbz, and silyl protecting groups to give **135** and **136** in high yields. The evaluation of the biological activity of these compounds



Scheme 20



Scheme 21



Scheme 22

confirmed the results published⁴⁰ the same year by Spiro and co-workers (see section 6).

Indeed, the latter researchers reported⁴¹ a very similar synthesis of the imino-*O*-disaccharide **135** employing the 6-acetate derivative **137** (Scheme 22) obtained by the diol **131** already described³⁹ by Fleet and co-workers. Glycosylation of **137** with the glucosyl chloride **133** stereoselectively gave the α -D-linked disaccharide (66%) that was easily deprotected to afford **135** (hydrochloride salt) in 65% yield. In an alternative approach, the diol **131** was converted into the oxazolidinone **138** which was glycosylated under halide ion catalytic conditions to give the corresponding disaccharide in 60% yield. Basic treatment and hydrogenolysis in acidic conditions afforded the same imino-*O*-disaccharide **135**.

Spoehr and co-workers described⁴¹ also the synthesis of various imino-disaccharides containing *O*-methylated or deoxygenated 1-deoxymannojirimycin units (**140–145**, Fig. 7) as well as, in a following paper⁴² also published in 1993, other disaccharides bearing methylated (**146**, **149**), chlorinated (**147**) or deoxygenated (**148**, **150**, **151**) glucose units, xylose (**152**) and galactose (**153**) moieties (Fig. 7).

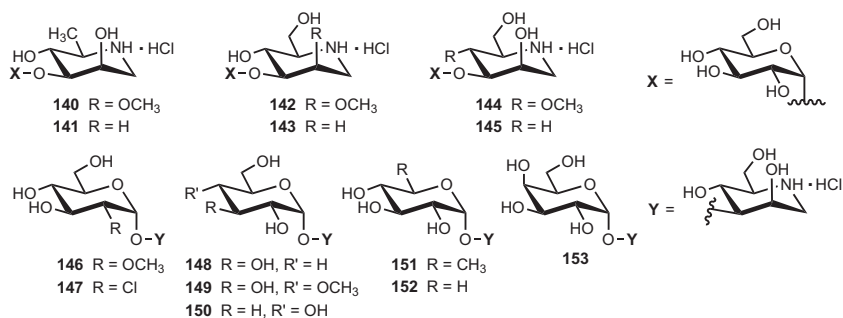
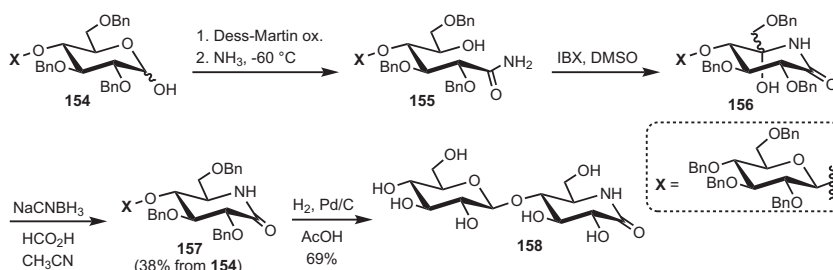


Fig. 7 Series of iminodisaccharides prepared by Spohr and co-workers.^{41,42}

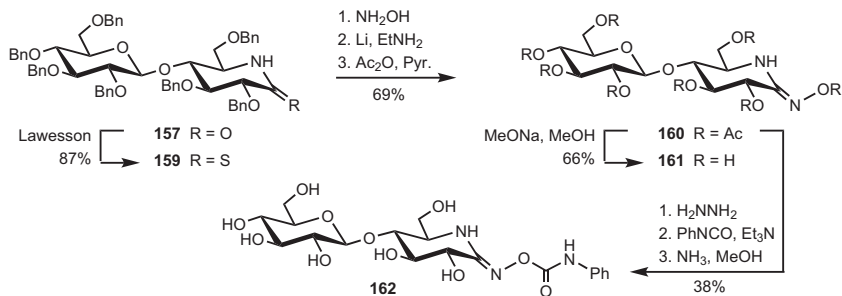


Scheme 23

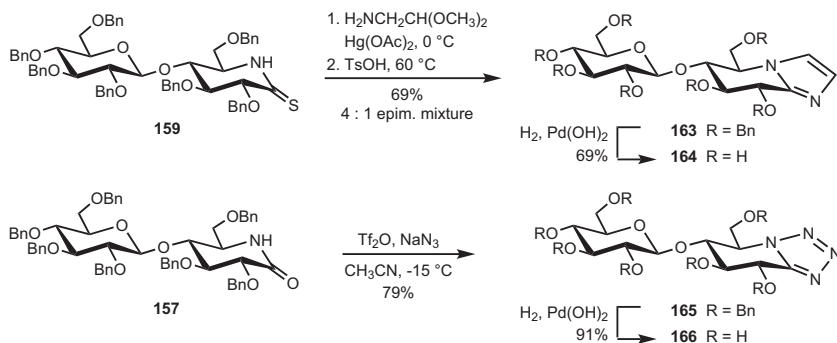
In 1999, Vasella and co-workers described⁴³ the synthesis of disaccharides bearing hydroxylactam and imidazole functionalized imino-sugar moieties. The hepta-*O*-benzyl-*D*-cellobiose **154** (Scheme 23) was oxidized with Dess-Martin periodinane to the corresponding lactone which, upon treatment with liquid ammonia at -60 °C gave the hydroxylamide **155**. Oxidation of crude **155** in the presence of 1-hydroxy-1,2-benziodoxol-3(1*H*)-one 1-oxide (IBX) led to the intermediate oxoamide that cyclized to afford a mixture of C-5 epimeric lactams **156**. Reductive deoxygenation of the latter gave the desired *D*-*gluco* configured product (**157**) in 38% overall yield from **154**. Also isolated was the corresponding *L*-*ido* epimer (7% from **154**). Removal of the benzyl groups by hydrogenolysis afforded the disaccharidic lactam **158** in 69% yield.

The benzylated cellobionolactam **157** was exploited, after conversion into the corresponding thiolactam **159** (Scheme 24), to prepare⁴³ the iminodisaccharidic hydroxylactam **161** and its phenylcarbamate **162**. Treatment of **159** with hydroxylamine, followed by debenzylation with lithium in ethylamine at -70 °C and acetylation, gave the octaacetate **160**, isolated in 69% overall yield. Transesterification of the latter afforded the hydroxylactam **161**, whereas the selective deacetylation of the acetoximo group with hydrazine, followed by reaction with phenyl isocyanate and final transesterification, led to the carbamate **162**.

The thiolactam **159** was also used as starting material for the synthesis⁴³ of a disaccharide bearing a fused iminosugar-imidazole unit (**164**, Scheme 25). The latter was prepared by treatment of **159** with aminoacetaldehyde dimethyl acetal and mercury(II) acetate to give the intermediate



Scheme 24

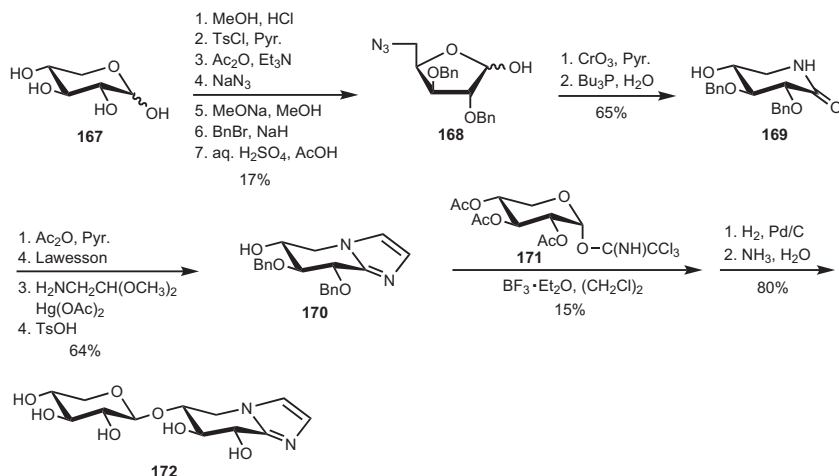


Scheme 25

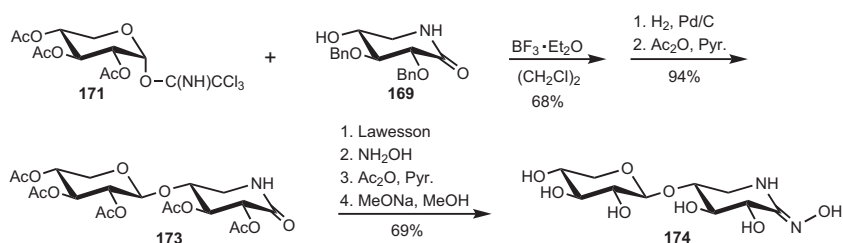
amidines which cyclized in the presence of *p*-toluenesulfonic acid at 60°C to afford an inseparable 4:1 mixture of C-2 epimers (*i.e.* *D*-gluco/*D*-manno) in 69% total yield. However, upon di-iodination of the imidazole ring, column chromatography, and deiodination, the *D*-gluco derivative **163** could be isolated. The removal of the benzyl groups by hydrogenolysis gave the imino-*O*-disaccharide **164** in 69% yield.

A closely related analogue of the latter, *i.e.* the fused iminosugar-tetrazole disaccharide **166** (Scheme 25) was easily obtained⁴⁴ by treatment of the lactam **157** with triflic anhydride and sodium azide and subsequent hydrogenolysis.

In 2000, an alternative approach to iminodisaccharides featuring hydroxylolactam and imidazole functions was proposed⁴⁵ by Withers and co-workers. Indeed, the target disaccharide **172** (Scheme 26) was obtained by conventional glycosylation of the *D*-xylo configured fused iminosugar-imidazole derivative **170** with the xylopyranosyl trichloroacetimidate **171** followed by the benzyl and acetyl groups removal. The acceptor **170** was prepared starting from *D*-xylose **167** through a long reaction sequence in order to synthesize the lactam **169** *via* cyclisation of the 5-aminolactone intermediate formed by oxidation of **168** and subsequent Staudinger reduction of the azide function. The lactam **169** was transformed into the imidazole derivative **170** through the usual reaction sequence involving treatment with the Lawesson reagent, then reaction with aminoacetaldehyde dimethyl acetal and $\text{Hg}(\text{OAc})_2$ and final cyclisation in the presence of *p*-toluenesulfonic acid.



Scheme 26



Scheme 27

Also the hydroximolactam-based *D*-xylo – *D*-xylo iminodisaccharide **174** (Scheme 27) was obtained⁴⁵ by glycosylation instead of transformation of a natural *O*-disaccharide as reported⁴³ by Vasella and co-workers for the *D*-gluco – *D*-gluco analogue **161** (see Scheme 24). Actually, glycosylation of the lactam **169** with the imidate **171** gave the corresponding disaccharide as pure β-*D* anomer in 68% yield. Debenzylation and acetylation of the latter afforded **173** which was transformed into the desired iminodisaccharide **174** by treatment with the Lawesson reagent followed by reaction with hydroxylamine and deprotection.

The same glycosyl donor **171** was also employed for the synthesis⁴⁵ of the imino-*O*-disaccharides **179–181** (Fig. 8) using the *D*-xylo configured acceptor **175**, easily obtained from the lactam **169**, and its 2-deoxy derivatives **176** and **177**. These disaccharides and those showed in the previous Schemes 26 and 27 were found to be inhibitors of xylanases (see section 6). In a following paper, the same Authors described⁴⁶ the preparation of the iminodisaccharide **182**, also obtained by conventional glycosylation of the lactam **178** with the trichloroacetimidate **171**.

Catelani and co-workers published⁴⁷ in 2001 imino-*O*-disaccharides where the iminosugar moiety was not introduced through glycosylation, instead, it derived from the glucose unit of the *D*-lactose (**6**, Scheme 28). The latter was transformed^{48,49} into the tetrol **183**, containing an

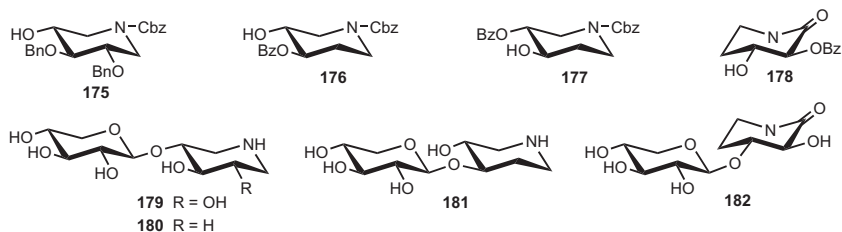
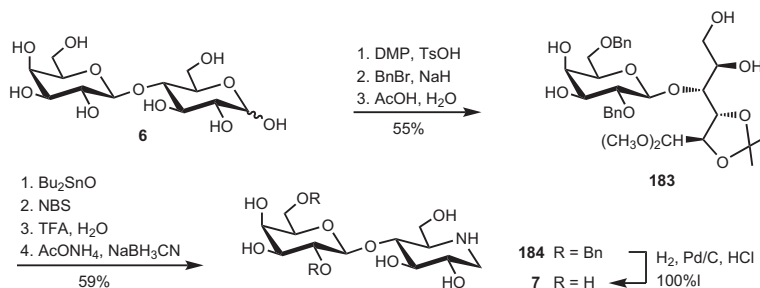


Fig. 8 Glycosyl acceptors and iminodisaccharides prepared by Withers and co-workers.^{45,46}



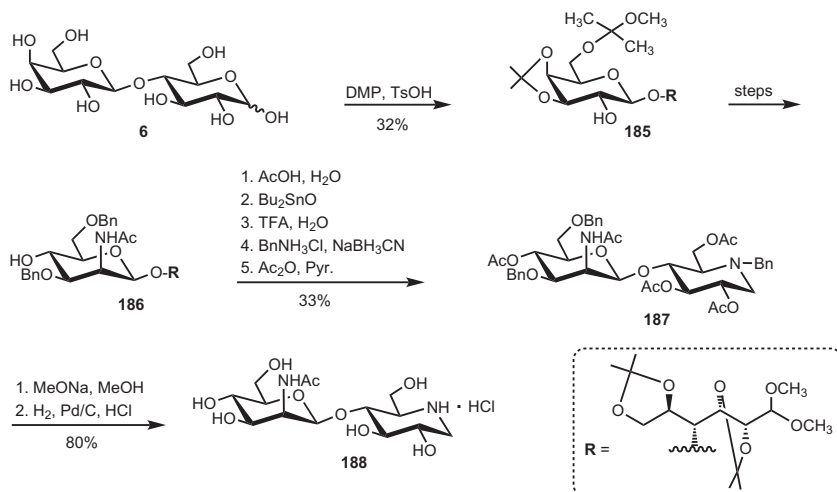
Scheme 28

open-chain glucose moiety, by treatment with 2,2-dimethoxypropane and *p*-toluenesulfonic acid followed by benzylation and selective isopropylidene hydrolysis. *Via* selective C-5 oxidation, removal of acetal groups and double reductive aminocyclisation, the protected disaccharide **184** was isolated in 59% yield. Quantitative hydrogenolysis gave the desired imino-*O*-disaccharide **7**.

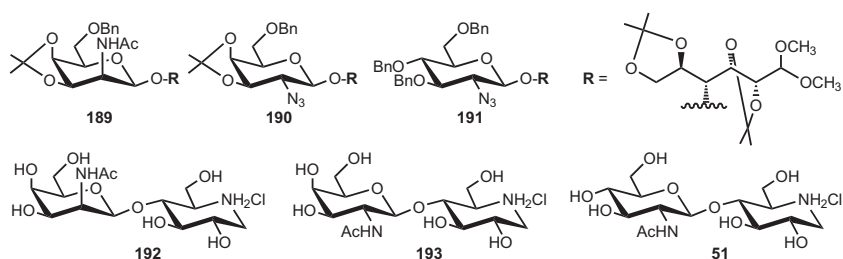
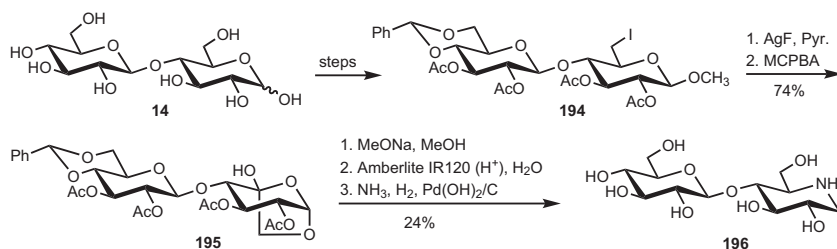
In a more recent paper⁵⁰ the same research team exploited this approach for the synthesis of four imino-*O*-disaccharides constituted of a 2-acetamido-2-deoxy-hexopyranose linked to the 4-OH of the 1-deoxy-nojirimycin. The D-lactose (**6**) was directly converted⁴⁸ into the protected key intermediate **185** (Scheme 29) from which the 2-acetamido-mannopyranoside derivative **186** was obtained by inversion of configuration first⁵¹ at C-2 (with concomitant amination) and then⁵² at C-4 through a very long reaction sequence. Then, selective deprotection of **186** followed by double reductive aminocyclisation of the aldohexos-5-ulose intermediate led to the *N*-benzylated imino-*O*-disaccharide **187**, easily deprotected to give the disaccharide **188**.

Starting from the same open-chain disaccharide **185**, other three aminated or azidated derivatives were prepared⁵⁰ (**189–191**, Fig. 9). *Via* similar reaction sequences, these compounds gave the *D*-*talo* (**192**), *D*-*galacto* (**193**), and *D*-*gluco* (**51**) configured imino-*O*-disaccharides (hydrochloride salts) in good yield.

The above mentioned synthetic approach was also employed⁵³ by Steiner and Stütz in 2004 to prepare two more imino-*O*-disaccharides from D-cellobiose and D-maltose instead of the D-lactose used by Catelani



Scheme 29

**Fig. 9** Iminodisaccharides and the corresponding precursors prepared by D'Andrea and co-workers.⁵⁰

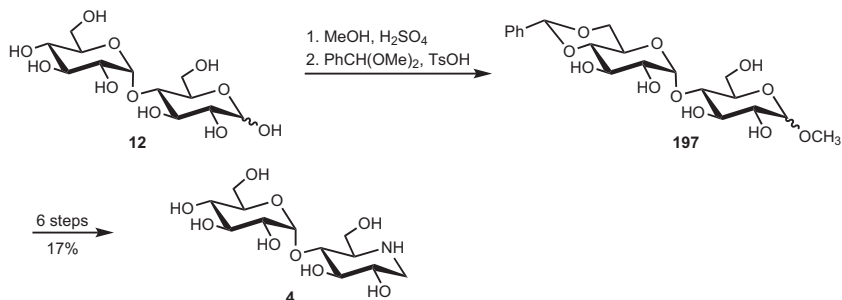
Scheme 30

and co-workers. The cellobiose (**14**) was transformed into the per-acetylated glycosyl chloride and then into the methyl β-D-cellobioside (details and yield not given) from which the selectively protected iodide derivative **194** was obtained (Scheme 30). Elimination of HI with silver fluoride and treatment of the alkene with *m*-chloroperbenzoic acid gave the corresponding unstable epoxide which, upon purification on silica gel column, afforded the anhydrosugar **195**. Transesterification of the latter and double reductive amination of the aldohexos-5-ulose

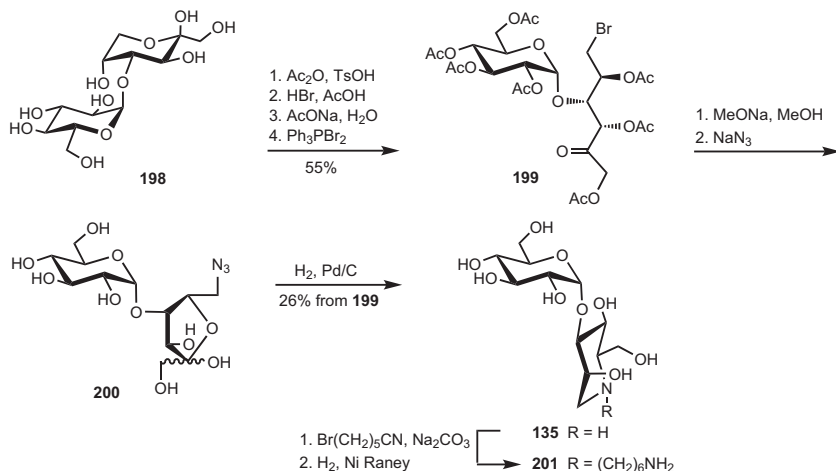
intermediate allowed to obtain the deprotected 1,4-linked imino-*O*-disaccharide **196** in modest yield.

The same reaction sequence was also applied⁵³ to the methyl α,β -D-maltoside derivative **197** (Scheme 31), obtained from D-maltose (**12**) *via* Fischer glycosylation and standard benzylidenation (details and yield not given). The α -D configured imino-*O*-disaccharide **4** was thus synthesized in 17% yield.

The same year, Spreitz and Stütz described⁵⁴ another approach for the synthesis of imino-*O*-disaccharides based on the conversion of a sugar moiety into a 1-deoxy-mannojirimycin unit avoiding the glycosylation step. Peracetylation at 0 °C of D-maltulose (**198**, Scheme 32) followed by treatment with HBr in acetic acid gave the anomeric bromide that was hydrolysed in the presence of aqueous sodium acetate. The corresponding hemiketal was then brominated to afford the open chain product **199** which was carefully transesterified (−30 °C) and reacted with sodium azide to give **200** that was immediately hydrogenated to form the already known 3-*O*-(α -D-glucopyranosyl)-1-deoxy-D-mannojirimycin disaccharide (**135**). Alkylation of the latter with a bromonitrile gave, after hydrogenation, the *N*-hexylamino



Scheme 31



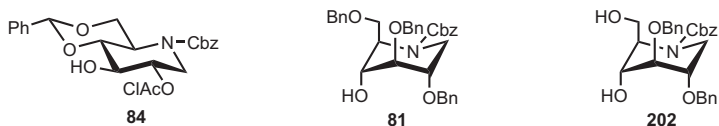
Scheme 32

derivative **201** that could be used for the installation of the imino-*O*-disaccharide on a solid support.

In 2006 Furneaux and co-workers reported⁵⁵ the synthesis of a series of di-, tri- and oligosaccharides containing a single 1-deoxynojirimycin unit at the reducing end to be tested as potential fungicides (see section 6). They were prepared by a classical glycosylation method using per-acetylated trichloroacetimidates (**203–208**, Fig. 10) as the glycosyl donors and the known Cbz-protected iminosugar alcohols **84** and **81** (see Fig. 3) together with the diol **202** as the glycosyl acceptors (Fig. 10).

The glycosylation followed by standard deprotection reactions allowed to prepare a series of β -D-linked imino-*O*-disaccharides (**92**, **196**, **213**, Fig. 11), -trisaccharides (**209**, **214**, **215**), and -oligosaccharides (**210–212**).

Glycosyl acceptors:



Glycosyl donors:

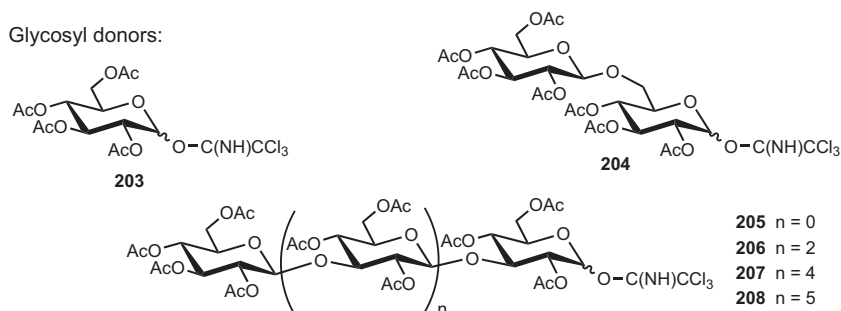


Fig. 10 Glycosyl acceptors and glycosyl donors employed by Furneaux and co-workers for the synthesis of 1-deoxynojirimycin containing di-, tri- and oligosaccharides.⁵⁵

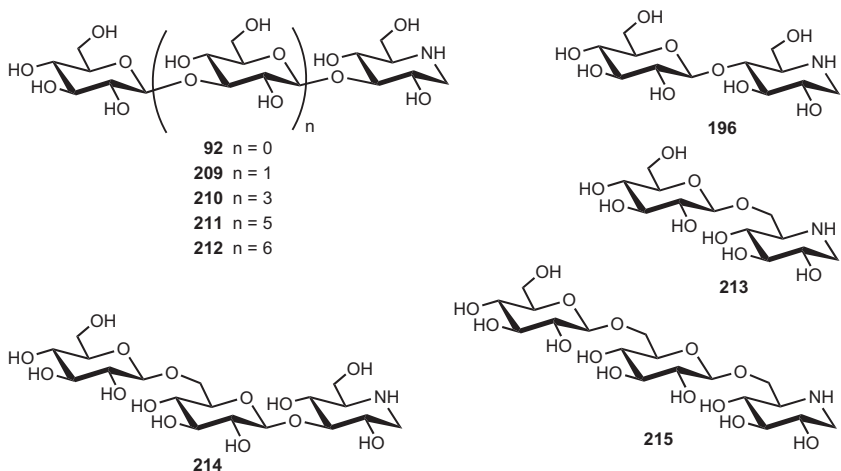
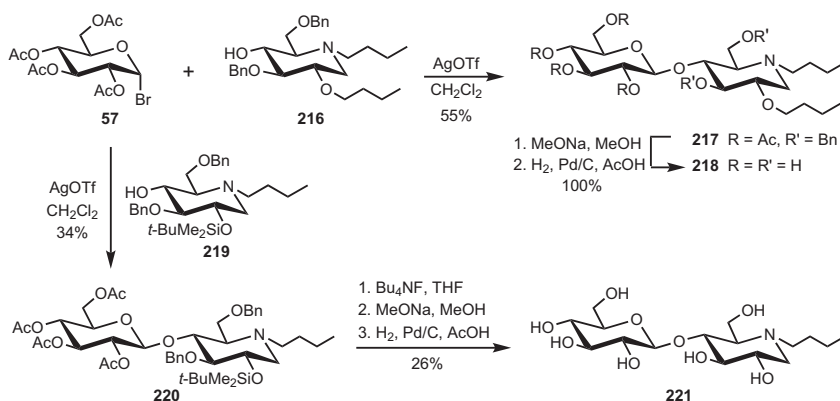


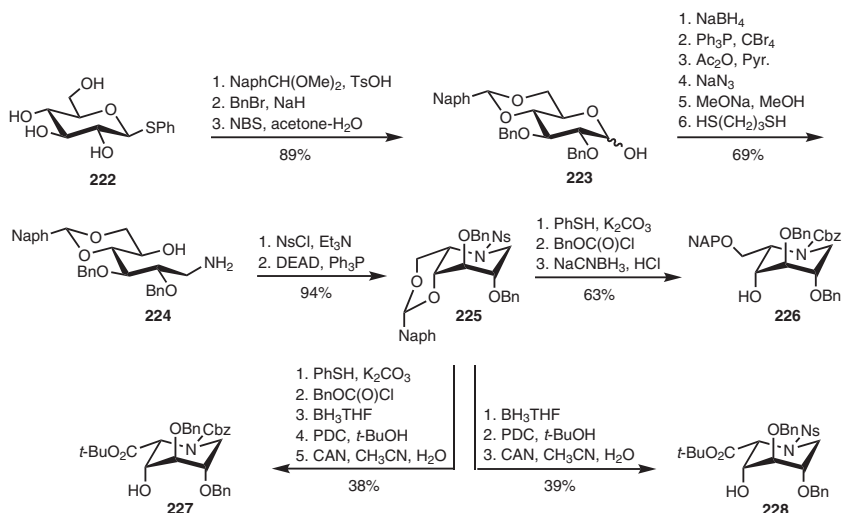
Fig. 11 Series of imino-*O*-disaccharides, -trisaccharides and -oligosaccharides prepared by Furneaux and co-workers.⁵⁵

In order to prepare potential inhibitors of human β -glucosidase, Compain, Martin and their co-workers synthesized⁵⁶ **218** and **221** (Scheme 33), the alkylated derivatives of the known 4-O-(β -D-glucopyranosyl)-1-deoxynojirimycin disaccharide **196**. To this aim, the suitably protected acceptors **216** and **219** were glycosylated with the glucosyl bromide **57** in the presence of silver triflate. The corresponding disaccharides **217** and **220**, isolated in modest yield, were then deprotected to give **218** and **221**.

Aiming to gain access to inhibitors of heparanase, Csiki and Fügedi synthesized⁵⁷ disaccharides bearing a GlcNAc unit α -D-linked to *L*-ido configured iminosugar moieties featuring a primary alcohol function or a carboxyl group. Moreover, they prepared also a couple of sulphated disaccharide derivatives. The glycosyl acceptors **226**–**228** (Scheme 34) were obtained by the key intermediate **225** displaying the endocyclic nitrogen atom protected by a 4-nitrobenzenesulfonyl (nosyl, Ns) group,



Scheme 33

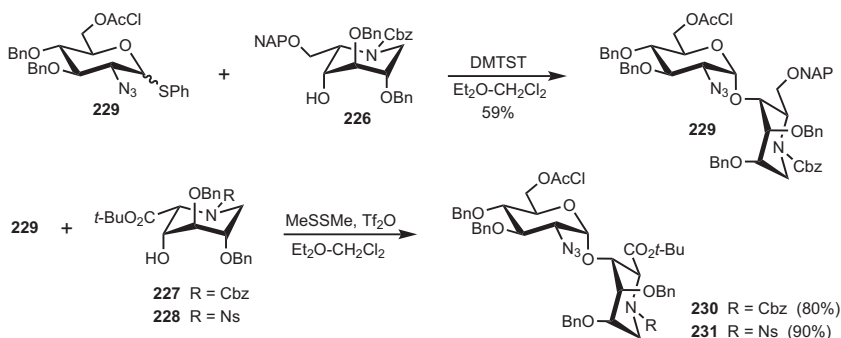


Scheme 34

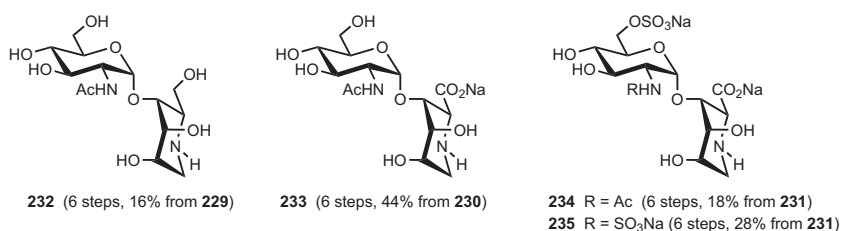
which was proved to be superior to the more common Cbz protecting group. In particular, the nosyl group allowed to avoid the formation of the carbamate rotamers that lead to complex NMR spectra when registered at room temperature. The compound **225** was obtained starting from the phenyl thioglucoside **222** by protection with 1-naphthaldehyde dimethyl acetal, benzylation, and NBS-mediated hydrolysis to give **223** in 89% yield. Conversion of the latter into the open-chain aminoalcohol **224** was followed by the installation of the nosyl group and subsequent ring closure by intramolecular Mitsunobu reaction. A series of reactions afforded the Cbz-protected acceptors **226–227** and the Ns-protected acceptor **228** (Scheme 34).

The three alcohols **226–228** were glycosylated with the same donor **229** (Scheme 35) in the presence of different promoters, the dimethyl(methylthio)sulfonium triflate (DMTST) for **226**, dimethyl disulfide and triflic anhydride in the case of **227–228**. The corresponding α -D-linked disaccharides were isolated in good yield when latter promoter system was used (**230**: 80%; **231**: 90%), whereas a lower yield was observed using DMTST (**229**: 59%).

The three iminodisaccharides **229–231** were then submitted to deprotection reaction sequences involving, in the case of **231**, also a sulfation step to give the target compounds **232–234** (Fig. 12). Moreover, as described in a preliminary communication published⁵⁸ the same year by the same Authors, the protected disaccharide **231** could be also converted into the O- and N-sulfated derivative **235**.



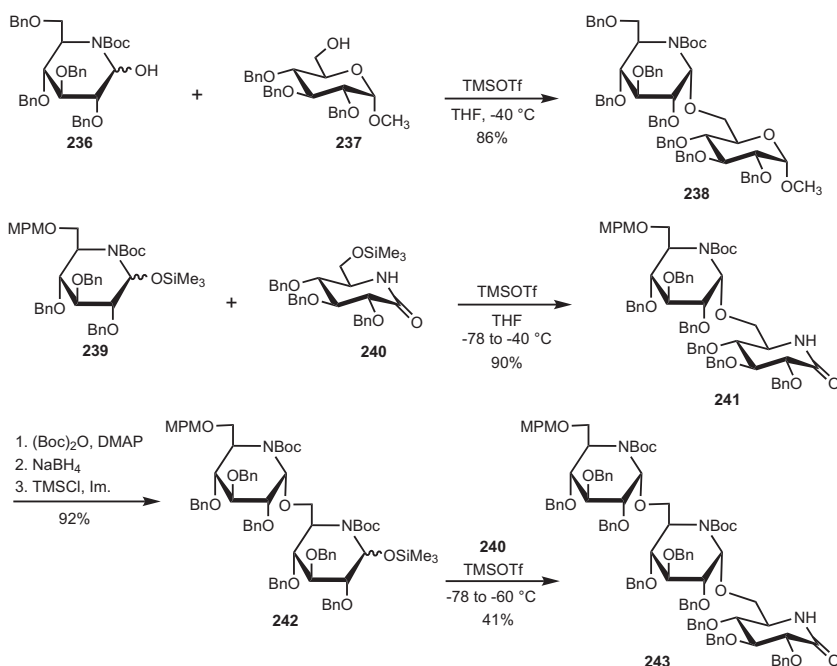
Scheme 35

Fig. 12 Iminodisaccharides prepared by Csiki and Fügedi.^{57,58}

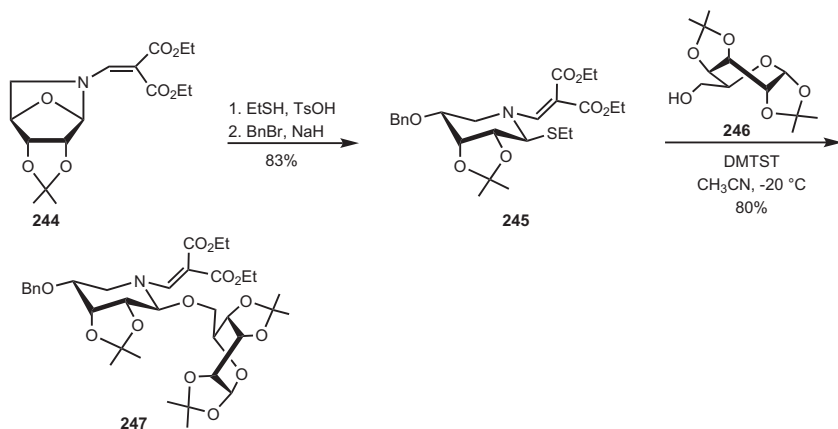
2.2 Imino-*O*-disaccharides bearing the sugar unit at the reducing end

In a 2005 preliminary communication, Ikegami and co-workers claimed⁵⁹ the synthesis of an imino-*O*-disaccharide by trimethylsilyl triflate-promoted glycosylation of the 6-OH glucoside **237** (Scheme 36) with the known,⁶⁰ protected hemiaminal **236**. Contrary to the ⁴C₁ representation in the paper,⁵⁹ it was shown⁶⁰ by Granier and Vasella that the two anomers of **236** adopt two different boat conformations. The anomeric configuration of the isolated disaccharide **238** was not proved, moreover, no physical data were given. The *p*-methoxybenzylated iminosugar **239**, silylated at the anomeric position, was used to glycosylate the silylated lactam **240**. The obtained disaccharide **241** was then transformed into the glycosyl donor **242** and reacted with the same acceptor **240** to give the trisaccharide **243**, which, however, was not deprotected. Unfortunately, due to the lack of spectral and physical data, the configuration and conformation of **241** and **243** cannot be proved.

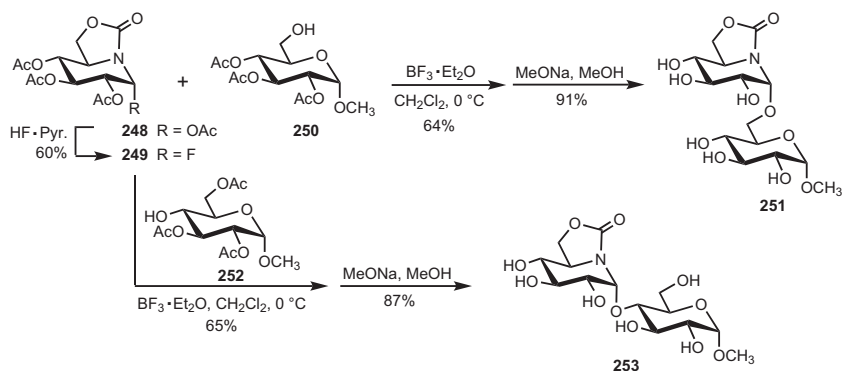
Another imino-*O*-disaccharide bearing at the non-reducing end a sugar unit, was reported⁶¹ by Fuentes and co-workers in a preliminary communication appeared in 2008. The 1,5-anhydro-ribofuranosylamine derivative **244** (Scheme 37), obtained^{62,63} from 2,3-*O*-isopropylidene-*D*-ribofuranosylamine *p*-toluenesulfonate (3 steps, 26% overall yield), was opened⁶³ in the presence of *p*-toluenesulfonic acid and ethanethiol to give, after benzylation, the thioethyl glycoside **245**. Coupling of the latter with the commercially available glycosyl acceptor **246** in the presence of DMTST afforded the disaccharide **247** in 80% yield as pure β-*D* (*i.e.* equatorial) anomer. The use of acetonitrile at low temperature as the



Scheme 36



Scheme 37



Scheme 38

reaction solvent insured the stereochemical control of the glycosylation when donors carrying non-participating groups at the C-2 position were employed (formation of an anomeric acetonitrilium ion intermediate).⁶⁴ Unfortunately, the disaccharide **247** was not deprotected and thus its biological properties could not be evaluated. In the same article the Authors claimed the synthesis of two other imino-*O*-disaccharides, however, no physical or spectral data were given.

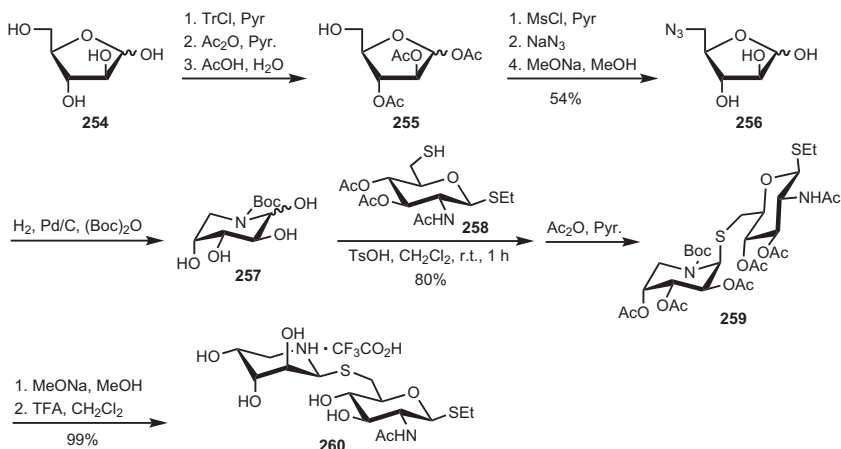
To date the only imino-*O*-disaccharides, fully characterized and suitable for the biological assays, bearing the sugar unit at the reducing end, were described⁶⁵ by Ortiz Mellet, Angulo and their co-workers in 2012. Aiming at studying the glycosidase inhibition activity of configurationally and conformationally stable iminosugar analogues of natural α -D-linked disaccharides, the two regioisomeric iminodisaccharides **251** and **253** (Scheme 38) were prepared from the same fluoride **249**. Reaction of the latter with the primary (**250**) or secondary (**252**) sugar alcohol in the presence of boron trifluoride etherate gave the corresponding disaccharides which, after standard transesterification, afforded the free-OH products **251** and **253**. The exclusive formation of axial glycosidic bonds

despite the presence of a participating group (OAc) at position 2, was ascribed to the very strong anomeric effect exerted by the nitrogen atom located in the cyclic carbamate due to its substantial sp^2 -character. In the same paper⁶⁵ the Authors described also the synthesis of the *N*- and *S*-glycosides analogues of the above mentioned iminodisaccharides. All compounds were then submitted to the glycosidase inhibition assays (see section 6).

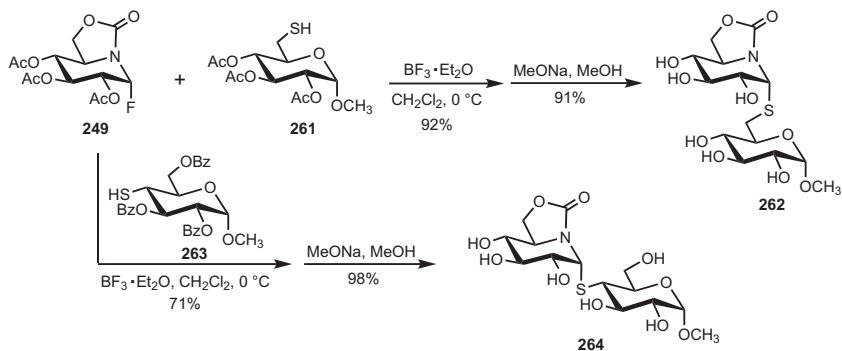
3 Synthesis of imino-*S*-disaccharides

A very limited number of iminosugar-based thiodisaccharides have been described to date, and all of them feature the sugar moiety at the reducing end. The first imino-*S*-disaccharide was reported^{66a} in 1994 by Suzuki and Hashimoto in a preliminary communication followed, six years later, by the corresponding full paper.^{66b} The Boc-protected hemiaminal **257** (Scheme 39) was prepared from the triacetyl *D*-arabinofuranose **255**, in turn obtained from *D*-arabinose **254** by regioselective tritylation, acetylation and acidic removal of the trityl group (yield not given). Mesylation of **255**, followed by treatment with sodium azide and transesterification, gave the 5-azido derivative **256** which was directly hydrogenated in the presence of $(\text{Boc})_2\text{O}$ to afford the acceptor **257** (details and yield not given). Coupling of the latter with the thioglycoside 6-thiol **258** in the presence of *p*-toluenesulfonic acid gave the axial (β -D) thiodisaccharide **259** isolated in 80% yield after acetylation. Transesterification and removal of the Boc group afforded the deprotected compound **260** (as trifluoroacetate salt) in almost quantitative yield. This compound proved^{66b} to be stable in acidic aqueous media whereas it underwent interglycosidic bond hydrolysis at pH values higher than 5.

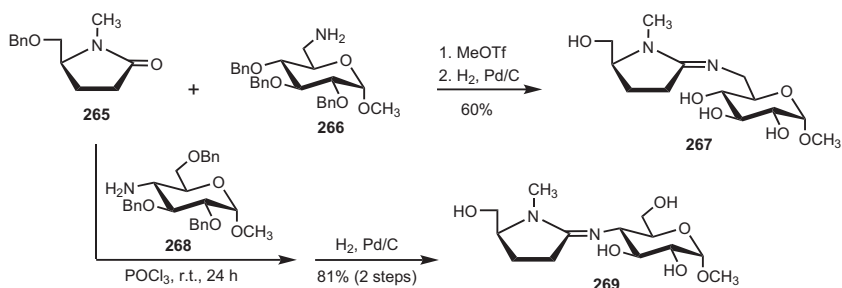
As mentioned in section 2.2, Ortiz Mellet, Angulo and their co-workers described in their 2012 paper⁶⁵ also the synthesis of two imino-*S*-disaccharides, namely **262** and **264** (Scheme 40). They were prepared taking advantage of the same glycosyl donor, the fluoride **249**,



Scheme 39



Scheme 40



Scheme 41

and promoter ($\text{BF}_3\text{Et}_2\text{O}$) but using the acetylated 6-SH glucoside **261** and the benzoylated 4-SH glucoside **263** as the acceptors. The corresponding α -D-linked disaccharides were isolated in good yield after acetylation of the crude reaction mixture and then submitted to standard transesterification to give the desired compounds **262** and **264**.

4 Synthesis of imino-*N*-oligosaccharides

Various imino-*N*-disaccharides have been prepared, all bearing the sugar unit at the reducing end. Moreover, some *N*-linked analogues of imino-disaccharides have also been synthesized, *i.e.* compounds where the sugar unit is linked to the endocyclic nitrogen atom instead of the oxygen or carbon atoms of the iminosugar moiety (see section 4.2).

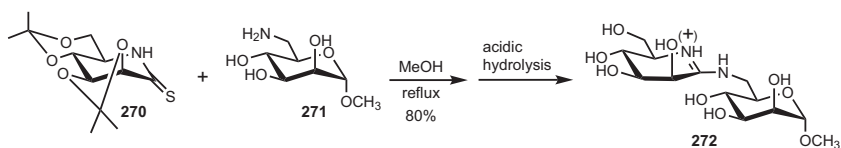
4.1 Imino-*N*-disaccharides bearing the sugar unit at the reducing end

The first, yet very simple, imino-*N*-disaccharides were reported⁶⁷ in 1993 by Knapp and co-workers. The amidine-based compounds **257** and **269** (Scheme 41) were obtained by coupling the *N*-methyl-lactam **265** (prepared in 52% yield, 4 steps, from L-pyrroglutamic acid) with the sugar amines **266** and **268** under different conditions. Indeed, in the first case the coupling was promoted by methyl triflate, whereas in the other one Vilsmeier activation (POCl_3) was used. After hydrogenation the disaccharides **267** and **269** were isolated in 60% and 81% yield, respectively.

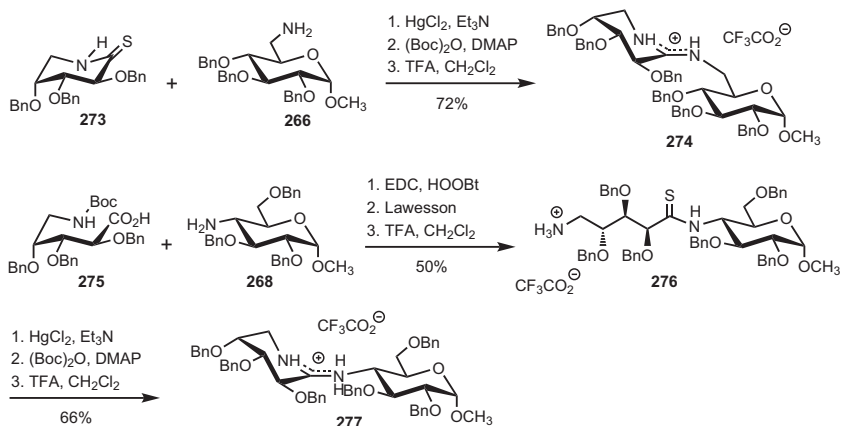
The Authors described also the synthesis of the demethylated analogue of **269** obtained in 68% yield starting from the *N*-benzyl derivative of the lactam **265**.

The synthesis of another iminodisaccharide featuring the amidine function was described⁶⁸ in 1995 by Tellier and co-workers. The protonated product **272** (Scheme 42) was easily obtained by heating in methanol the *D*-manno configured thiolactam **270** with the free-OH 6-amino-mannoside **271**. The resulting disaccharide, recovered in 80% yield, was then submitted to acidic hydrolysis (reagents, conditions and yield not given) to afford **272**, an inhibitor of α -mannosidases (see section 6).

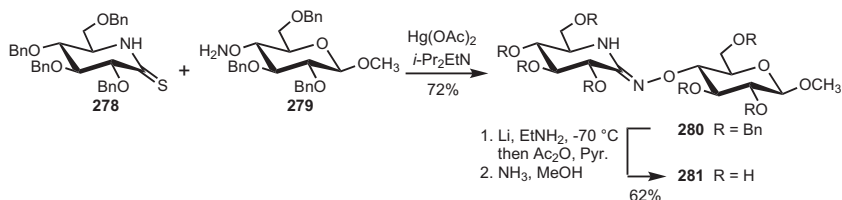
The coupling between a thiolactam and a primary amine was exploited also by Hashimoto and co-workers to prepare⁶⁹ the amidine-based imino-*N*-disaccharide **274** (Scheme 43). However, they used mercury(II) chloride and triethylamine to promote the coupling, an activating system that totally failed when the thiolactam **273** was reacted with the 4-amino-glucoside **268**. Therefore, a different strategy was adopted. The *N*-Boc arabinonic acid **275**, easily obtained from the lactam precursor of **273**, was coupled with the amine **268** under classical conditions (EDC, HOObt) to give, after treatment with the Lawesson reagent and acidic removal of the Boc group, the open-chain thioamide **276** (50%, 3 steps). Treatment of the latter with HgCl_2 , *N*-protection and acidic hydrolysis, afforded **277** in 66% overall yield. Unfortunately, both disaccharides were not debenzylated and thus their biological properties could not be evaluated.



Scheme 42



Scheme 43



Scheme 44

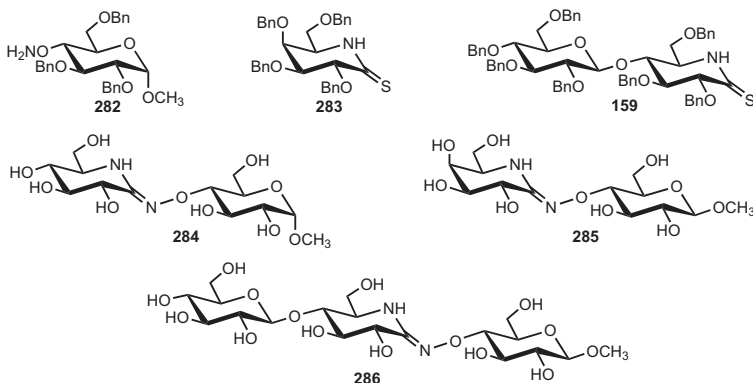


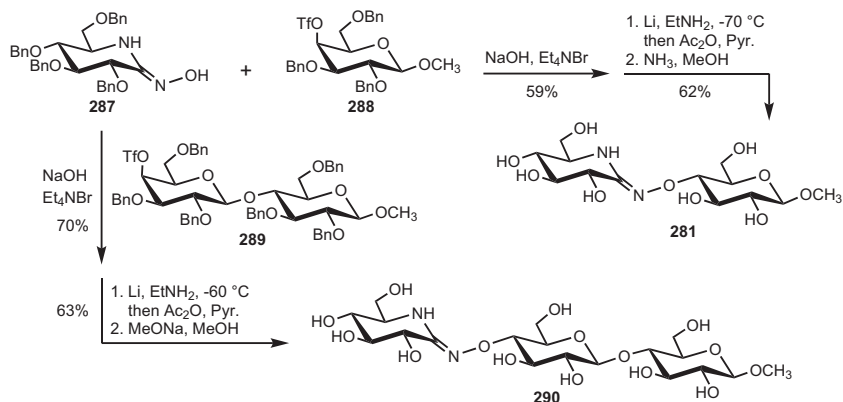
Fig. 13 Imino-disaccharides and -trisaccharide and their precursors prepared by Vasella and co-workers.^{43,70}

Vasella and co-workers prepared^{43,70} a series of di- and trisaccharides featuring hydroximolactam-linked iminosugar moieties. The coupling of the thiolactam **278** (Scheme 44) with the sugar hydroxylamine **279** in the presence of mercury(II) acetate and Hünig base afforded⁷⁰ the disaccharide **280** (72%) which was debenzylated with lithium in ethylamine at low temperature, then acetylated to allow the isolation and finally treated with ammonia in methanol to give **281** in 62% overall yield.

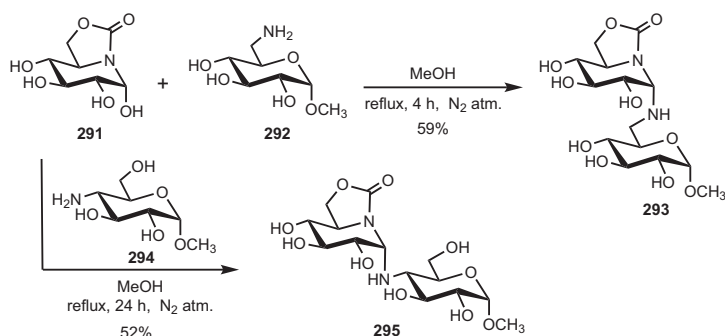
When the same thiolactam **278** was reacted with the hydroxylamine **282** (Fig. 13), *i.e.* the anomer of **279**, the corresponding free-OH disaccharide **284** was recovered⁷⁰ in 28% overall yield. The use of the *D*-galacto-thiolactam **283** and the disaccharidic thiolactam **159** in the coupling with the β -*D* configured hydroxylamine **279** gave, after debenzylation, **285** (66%)⁷⁰ and **286** (30%),⁴³ respectively.

These Authors described also another approach toward the synthesis of the same class of di- and oligosaccharides. The coupling of the hydroximolactam **287** (Scheme 45) with mono- (**288**) and disaccharidic (**289**) sugar triflates in the presence of NaOH led to the formation of the corresponding di- (59%) and trisaccharide (70%) with inversion of configuration. Thus, after deprotection, the cellobiose (**281**)⁷⁰ and cellotriose (**290**)⁴³ mimics were isolated in 62–63% yield.

In a preliminary communication published⁷¹ in 2009 and then in their 2012 full paper,⁶⁵ Ortiz Mellet, Garcia Fernandez and their co-workers



Scheme 45



Scheme 46

exploited the bicyclic carbamate iminosugar donors to prepare imino-*N*-disaccharides. However, contrary to the efficient couplings of the anomeric fluoride **249** (see Schemes 38 and 40) with sugar alcohols or thiols as the glycosyl acceptors, the reaction of **249** with sugar primary amines did not take place. Fortunately, the reaction between the iminosugar carbamate **291** (Scheme 46) and the 6-amino- (**292**) or 4-amino-glucoside (**294**) occurred in refluxing methanol to give the disaccharides **293** and **295**, respectively. The reactions were carried out under a nitrogen atmosphere to avoid the oxidation and intramolecular glycosylation of the 1-amino-1-deoxy-*D*-fructose intermediate formed by Amadori rearrangement of the target compounds.

4.2 *N*-linked analogues of imino-oligosaccharides

In 1996, Bols, Sierks and their co-workers synthesized⁷² a series of neutral and charged isofagomine derivatives *N*-linked to methyl glucopyranoside (**296**–**299**, Fig. 14). These compounds proved to act as potent inhibitors of various glycosidases (see section 6).

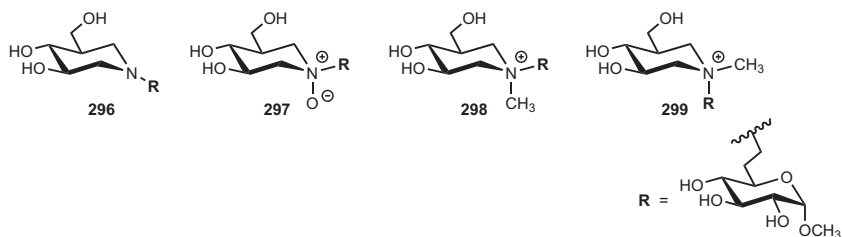


Fig. 14 Isofagomine derivatives *N*-linked to methyl glucopyranoside prepared by Bols, Sierks and their co-workers.⁷²

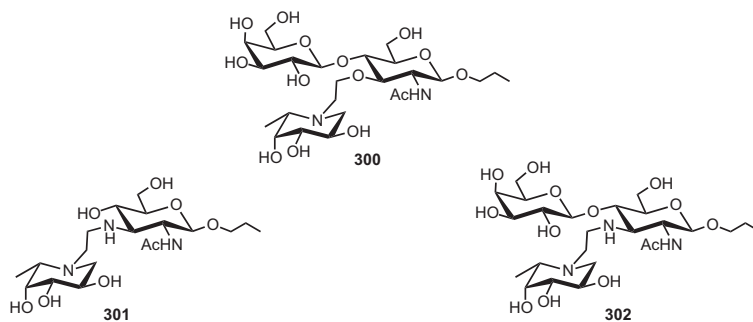


Fig. 15 *N*-linked iminotrisaccharides prepared by Wong and co-workers.^{73,74}

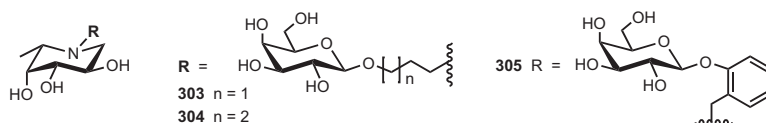


Fig. 16 *N*-linked iminodisaccharides prepared by Jefferies and Bowen.⁷⁵

The same year, Wong and co-workers described⁷³ the synthesis and the α -1,3-fucosyltransferase inhibition activity (see section 6) of the *N*-linked iminotrisaccharide **300** (Fig. 15). Two years later, they prepared⁷⁴ the trisaccharide **302**, *i.e.* the 2,3-diamino-glucoside analogue of **300**, as well as the disaccharide derivative **301** (Fig. 15).

Three galactose-containing derivatives of the 1-deoxyfuconojirimycin (**303–305**, Fig. 16) were prepared⁷⁵ by Jefferies and Bowen in 1997 as inhibitors of the α -1,3-fucosyltransferase (see section 6).

The following year, the hydroxylamino-linked pseudodisaccharides **306** and **307** (Fig. 17) were reported⁷⁶ by Zhao and co-workers, however, their biological properties were not evaluated.

In 2002, Rutters and Van der Eychen described⁷⁷ the solid-phase synthesis of a new class of oligosaccharides analogues (**308–310**, Fig. 18) bearing *L*-ido configured iminosugars linked through a carbamate bond.

Two potential inhibitors of α -glucosidase, the *N*-linked disaccharides **311** and **312** (Fig. 19) bearing a 1,3-dideoxy-3-methyl-iminosugar unit, were reported⁷⁸ in 2011 by Pistarà and co-workers.

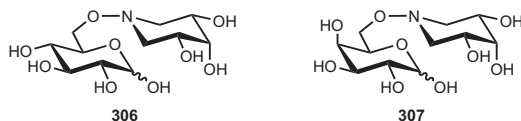


Fig. 17 Hydroxylamino-linked pseudodisaccharides prepared by Zhao and co-workers.⁷⁶

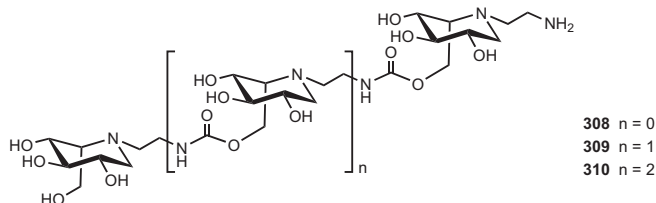


Fig. 18 Carbamate-linked iminooligosaccharides prepared by Ruttens and Van der Eychen.⁷⁷

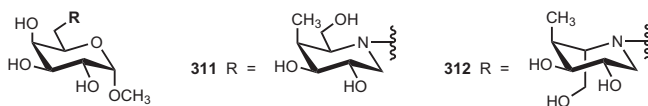


Fig. 19 *N*-linked iminodisaccharides prepared by Pistarà and co-workers.⁷⁸

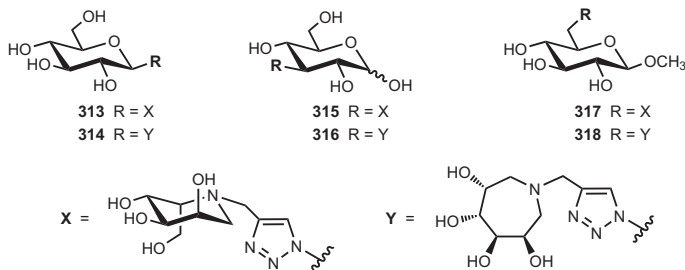


Fig. 20 Triazole-linked iminodisaccharides prepared by Carvalho and co-workers.⁷⁹

In a very recent article, Carvalho and co-workers prepared⁷⁹ *via* copper(I)-mediated azide-alkyne cycloaddition (CuAAC), a series of pseudo-disaccharides containing piperidine (**313**, **315**, **317**, Fig. 20) or azepane (**314**, **316**, **318**) iminosugars *N*-linked to the position 1, 3 or 6 of a glucopyranose unit through a triazole spacer. These compounds were evaluated as glycosidase inhibitors (see section 6).

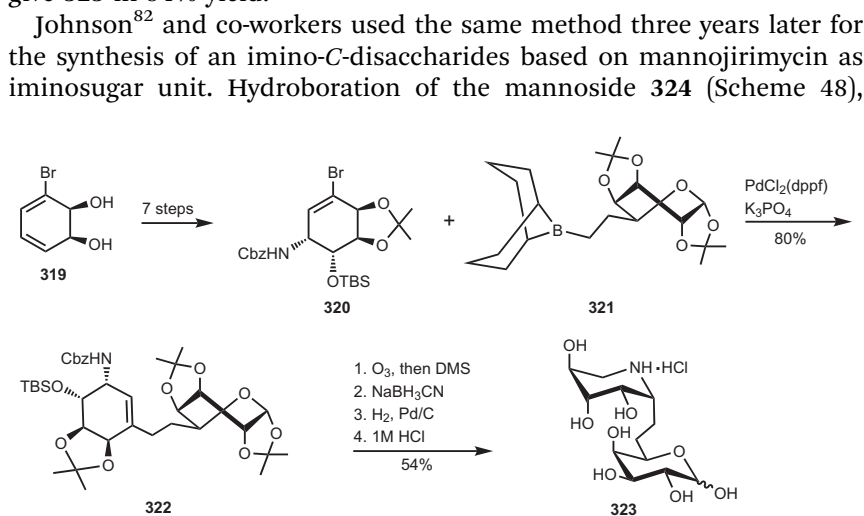
5 Synthesis of imino-*C*-disaccharides

Despite the difficulty of their preparation, imino-*C*-disaccharides represent a large family of imino-disaccharides. Indeed, they often require

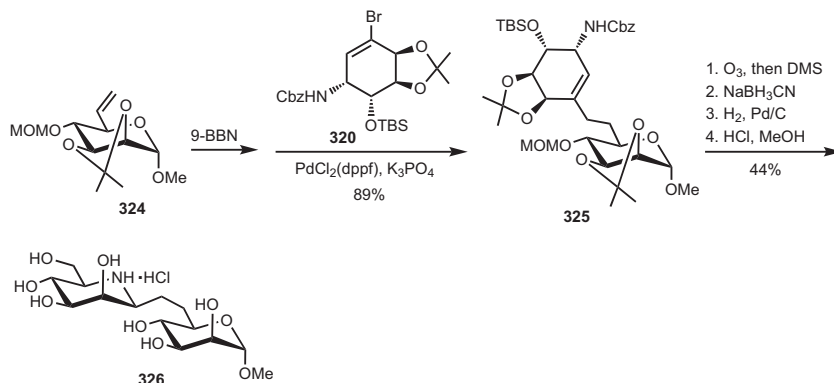
long, multistep syntheses using different C–C bond formation methods, leading mainly to disaccharides bearing the sugar unit at the reducing end. On the other hand, a fair amount of non-anomerically linked disaccharides have been synthesized. In any case, imino-*C*-disaccharides have to be considered as good candidates for glycosidases inhibition because of their stability towards acid and base catalysed hydrolysis.

5.1 Imino-*C*-disaccharides bearing the sugar unit at the reducing end

In 1994, Johnson⁸⁰ and co-workers described a new class of sugar mimics as they synthesized the first example of imino-*C*-disaccharide using Suzuki coupling as key reaction. Intermediate **322** was obtained by coupling vinyl bromide **320** (derived⁸¹ in seven steps from enantiopure diol **319**) with alkyl boron **321** (derived in four steps from *D*-galactose) in 80% yield (Scheme 47). Oxidative cleavage (O₃, DMS) followed by hydrogenolysis of the carbamate moiety led to the iminosugar unit that was deprotected to give **323** in 54% yield.



Scheme 47



Scheme 48

followed by Suzuki coupling with bromide **320**, afforded **325** in very good yield (89%). Oxidative cleavage (O_3 , DMS) followed by reduction of the freshly formed aldehyde ($NaBH_3CN$) and intramolecular reductive amination, induced by the cleavage of the carbamate moiety (H_2 , Pd/C), led to the protected imino-*C*-disaccharide that underwent full deprotection in acidic media providing **326** as hydrochloride salt in 44% yield.

Using the same methodology on different sugar alkenes, the imino-*C*-disaccharides **327–329** were obtained in good overall yields (39–50%, Fig. 21).

Two years later, van Boom⁸³ and co-workers suggested a new strategy for the synthesis of **328** allowing them to easily diversify the iminosugar unit (Scheme 49). In that case, a modified Corey–Fuchs reaction was used as key step in order to link the sugar and iminosugar units. Conversion of dibromoalkene **330** with *n*-BuLi into the corresponding acetylenic anion followed by the addition of 2,3,4,6-tetra-*O*-benzyl-D-gluconolactone **331** afforded the ketose **332** in 88% yield. Then, a classical reduction, oxidation and reductive amination sequence was performed to afford the fully protected disaccharide **333** (38%, 3 steps). Finally, **334** was obtained by hydrogenolysis in 80% yield. The same strategy was applied to obtain the already known imino-*C*-disaccharide **328** (31% yield, 5 steps) and the galactonojirimycin derivative **335** (29%, 5 steps).

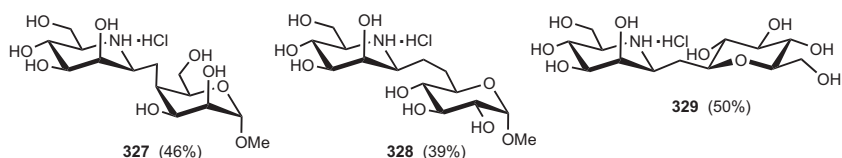
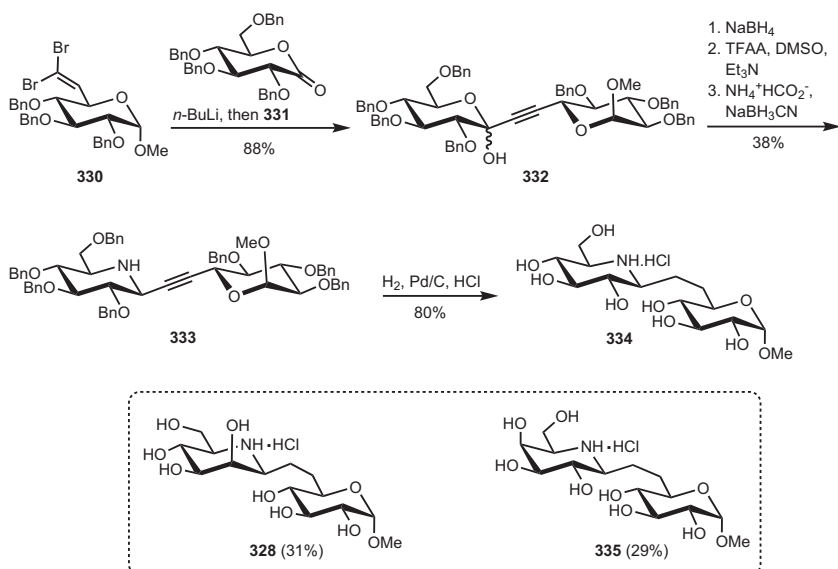


Fig. 21 Imino-*C*-disaccharides prepared by Johnson and co-workers.⁸²

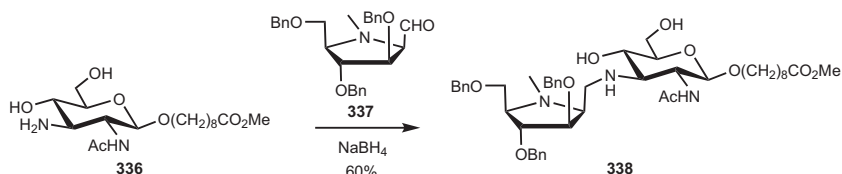


Scheme 49

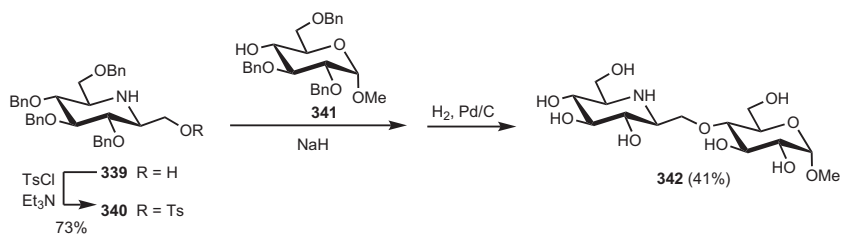
In 1995, Wong⁸⁴ and co-workers envisaged the reductive amination as the key step to join the sugar and the iminosugar units. In presence of NaBH_4 , the reaction between the aldehyde **337**, prepared in 5 steps from the corresponding homoiminosugar, and the 2,3-diamino-glucose derivative **336** gave the protected disaccharide **338** in 60% yield (Scheme 50). Unfortunately, **338** was not deprotected and thus its biological properties could not be evaluated.

One year later, Saavedra and Martin⁸⁵ used the newly synthesized β -homonojirimycin **339** to obtain a homo-analogue of an azacellobioside (**342**, Scheme 51). To this end, methyl 2,3,6-tri-*O*-benzyl- α -D-glucopyranoside **341** was reacted with tosylate **340** to give the corresponding benzylated disaccharide. Hydrogenolysis of the benzyl ethers afforded **342** (41%, 2 steps), the first example of an homoaza-analogue of a natural product.

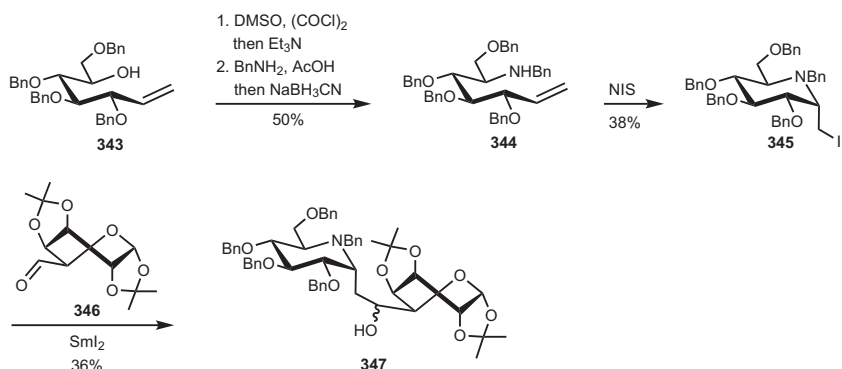
The same year, Martin⁸⁶ and co-workers used the iodo analogue of the α -homonojirimycin **345** as intermediate for the synthesis of an imino-*C*-disaccharide containing a galactose unit (Scheme 52). To this aim,



Scheme 50



Scheme 51

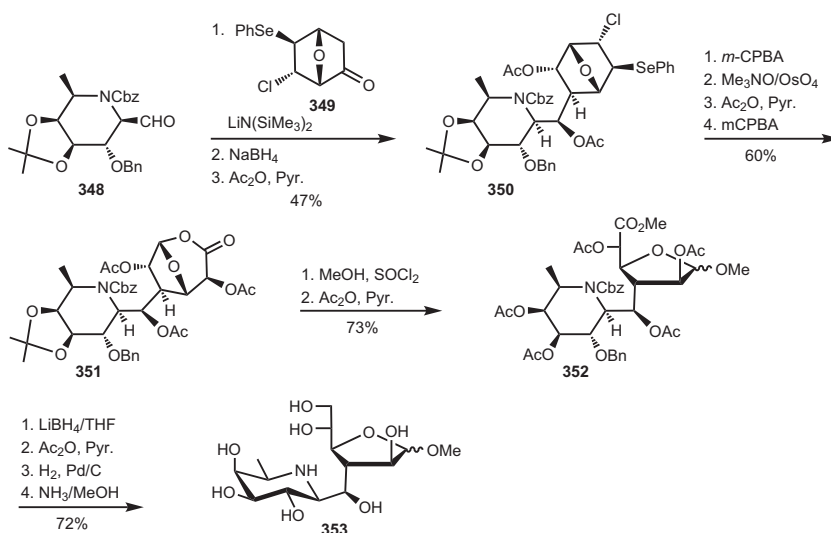


Scheme 52

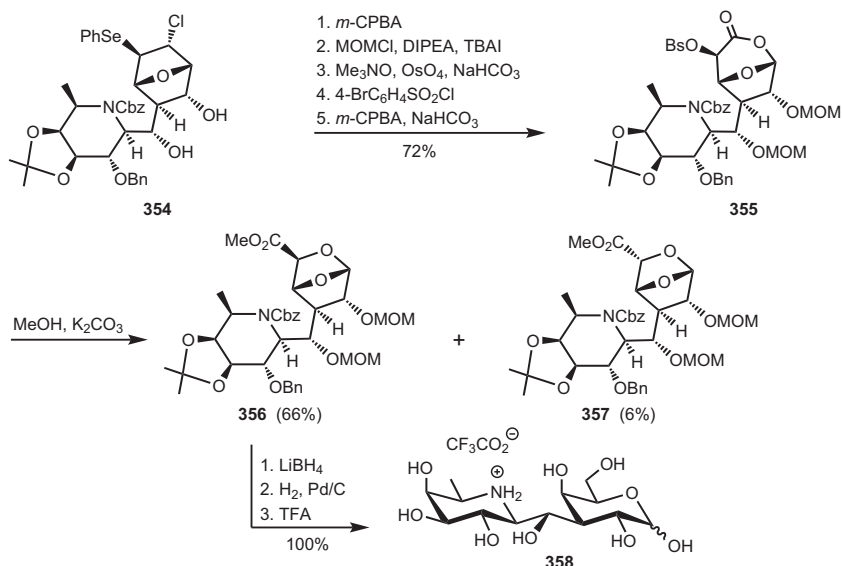
oxidation of the alkene **343** followed by reductive amination in presence of benzylamine led to *D*-gluco-aminoheptenitol **344** (50%) along with the *L*-ido derivative in a 5 : 2 ratio. Cyclisation of **344** mediated by NIS gave the intermediate **345** in good yield (80%). The reaction of the latter with aldehyde **346** under samarium Barbier conditions afforded the desired disaccharide **347** in a mixture of easily separable stereoisomers (36%). Unfortunately, the deprotection of **347** was not attempted by the Authors leaving this new derivative as a fully protected imino-*C*-disaccharide.

Meanwhile, Baudat and Vogel reported⁸⁷ a different approach towards new imino-*C*-disaccharides in a preliminary communication. In fact, their strategy included the modification of a bicyclic moiety already linked to the protected 6-deoxygalactonojirimycin (Scheme 53). Firstly, condensation of the lithium enolate of **349** with aldehyde **348** followed by reduction (NaBH_4) and acetylation (Ac_2O , pyridine, DMAP) steps afforded the intermediate **350** (47%). Oxidation of **350** with *m*-CPBA gave the corresponding vinyl chloride which was hydroxylated (Me_3NO , OsO_4) and acetylated. Baeyer-Villiger oxidation of the resulting acetate afforded the uronolactone **351** (60%). Treatment with MeOH and SOCl_2 followed by acetylation gave an anomeric mixture of methyl furanosides **352** (73%). The α : β ratio of **352** was determined one year later⁸⁸ together with its deprotection procedure to afford the free-OH imino-*C*-disaccharide **353** (72%).

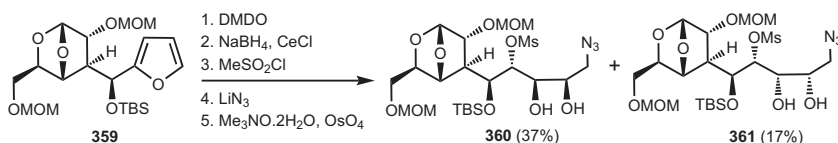
In completion of the preliminary communication, Baudat and Vogel were able to synthesize,⁸⁸ using the same methodology, an analogue of **353** bearing a galactose moiety (Scheme 54). It is worth noting that the product resulting from the ring condensation step (K_2CO_3) was obtained as a mixture of diastereoisomers **356** (66%) and **357** (6%) probably arising from the basic epimerization of **356**. Finally, after reduction of the methyl ester group followed by complete deprotection, **358** was obtained in quantitative yield as a TFA salt.



Scheme 53



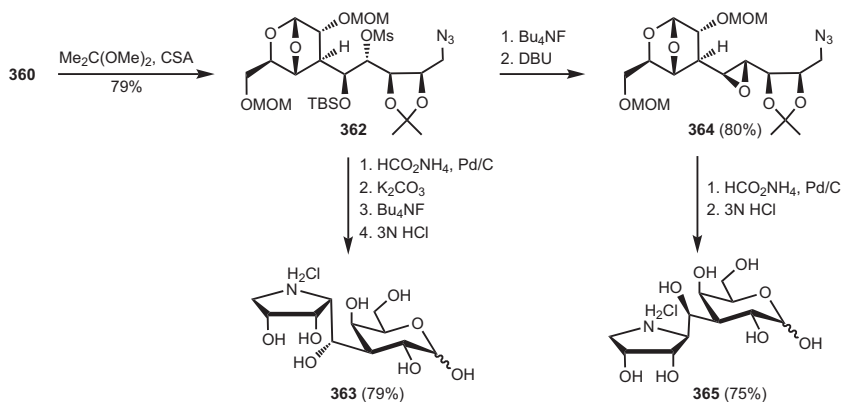
Scheme 54



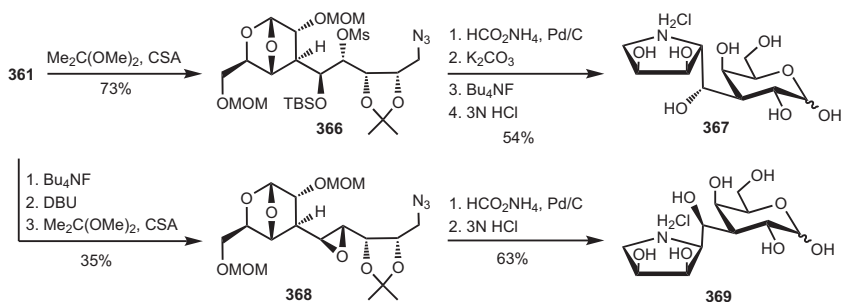
Scheme 55

Simultaneously, Vogel and co-workers reported in a preliminary communication⁸⁹ followed by a full paper⁹⁰ the total synthesis of racemic imino-*C*-disaccharides (11 compounds) in which both iminosugar and sugar units came from the same intermediate ($\pm$) – 359 (Scheme 55). The latter was obtained from ketone 349 (Scheme 53) using furfural as aldehyde followed by the same steps employed for the synthesis of 351. Opening of the furan heterocycle was done with DMDO followed by reduction (NaBH₄) of the previously obtained aldehyde. Both hydroxyls were mesylated and the primary mesylate was substituted by an azide using LiN₃. Finally, the unstable azide was treated with Me₃NO·2H₂O in presence of a catalytic amount of OsO₄ to give a 2 : 1 mixture of diols 360 and 361 (respectively 37% and 17% over 5 steps).

The major diol 360 was protected with Me₂C(OMe)₂ to give the acetonide 362 (79%, Scheme 56). Reduction of the azide (HCO₂NH₄, Pd/C) followed by K₂CO₃ mediated cyclisation and classic deprotection steps afforded racemic 363 (79%, 4 steps) as a 44 : 44 : 12 mixture of α -D,L-pyranose, α -D,L-furanose and β -D,L-pyranose. Direct desilylation of 362 followed by DBU-mediated intramolecular S_N2 reaction afforded the epoxide 364 in 80% yield. Reduction of azide 364 triggered the cyclisation through epoxide opening. Acidic hydrolysis gave racemic 365 (75%,



Scheme 56



Scheme 57

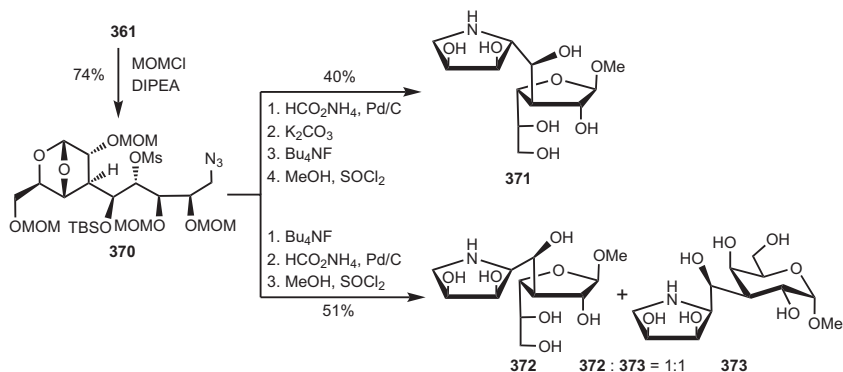
2 steps, 33:33:17:17 anomeric mixture of pyranose and furanose forms), which is an epimer of **363**.

Similarly, **361** was used as precursor for the synthesis of both epimers ($\pm$) – **367** and ($\pm$) – **369** in 54% and 63% yield, respectively (Scheme 57).

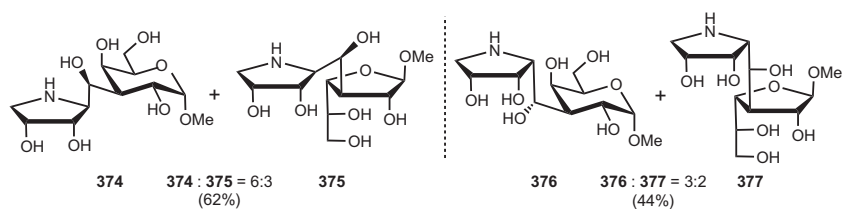
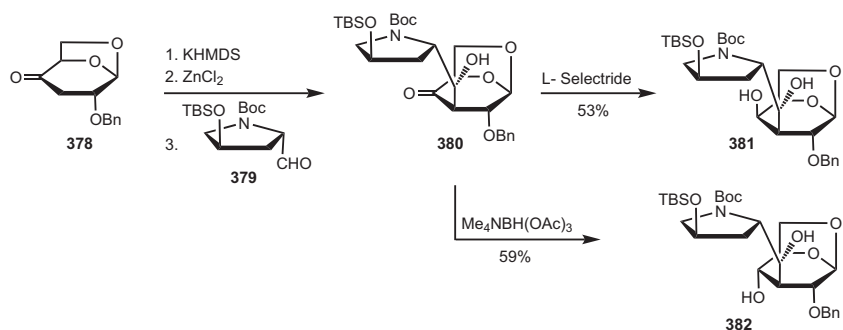
Then, the Authors envisaged the synthesis of methyl pyranosides and furanosides but the classical hydrolysis conditions (dry MeOH with acid catalyst) failed in all cases. It was thus decided to replace the acetonide protecting group by the methoxymethyl groups. For example, azide **370** was readily obtained from **361** (MOMCl, DIPEA) in 74% yield (Scheme 58). As seen before, going through epoxidation or not, and ending with methanolysis and glycosidation (MeOH, SOCl_2) afforded racemic **371** (4 steps, 40%, contaminated by 30% of its anomer and the corresponding pyranoside derivatives) or a 1:1 mixture of ($\pm$) – **372**–**373** (3 steps, 51%), respectively.

Compounds ($\pm$) – **374**–**375** (4 steps, 62%, 6:3 ratio) and ($\pm$) – **376**–**377** (3 steps, 44%, 3:2 ratio) were obtained by the same strategy starting from the azide **360** (Fig. 22). It is worth noting that all these imino-*C*-disaccharides were tested as inhibitors against several glycosidases and that none showed satisfactory inhibitory activity.

In order to shorten the quite complex syntheses of imino-*C*-disaccharides, Zhu and Vogel⁹¹ envisaged a new strategy based on the



Scheme 58

Fig. 22 Imino-C-disaccharides prepared by Vogel and co-workers.^{89,90}

Scheme 59

condensation of the dideoxy iminoaldose **379** with an isolevoglucosenone derived enolate (Scheme 59). Conjugate addition of benzyl alcohol to isolevoglucosenone afforded **378** with high stereoselectivity. After formation of the enolate of **378** (KHMDS, then ZnCl_2), iminoaldose **379** was added leading to a single addition product **380** that was found to be unstable over purification conditions. Therefore crude **380** was reduced using L-selectride to afford the D-galactose derivative **381** (53%, 4 steps). The use of $\text{Me}_4\text{NBH}(\text{OAc})_3$ as reducing agent on crude **380** gave the D-glucose derivative **64** (59%, 4 steps).

The disaccharides **381** and **382** were then submitted to classical deprotection reaction sequences in order to remove the silyl and benzyl groups. Cleavage of the carbamate function was done simultaneously

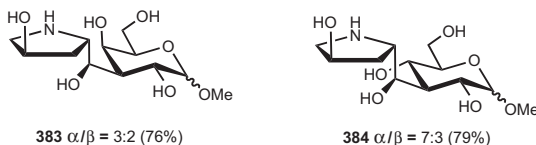
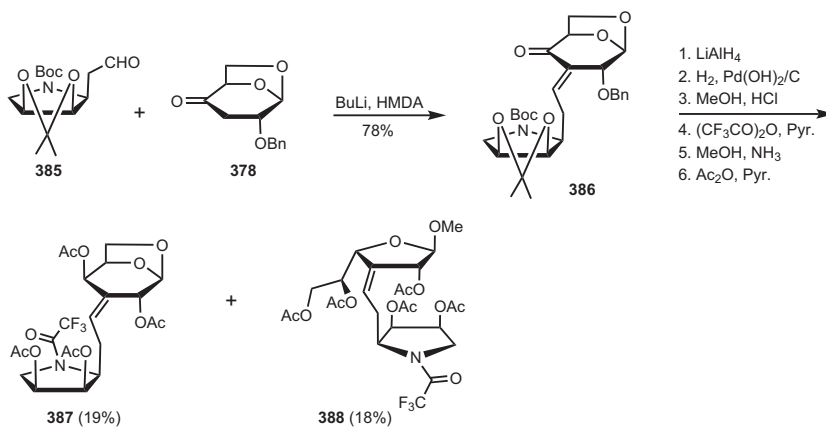


Fig. 23 Imino-C-disaccharides prepared by Zhu and Vogel.⁹¹



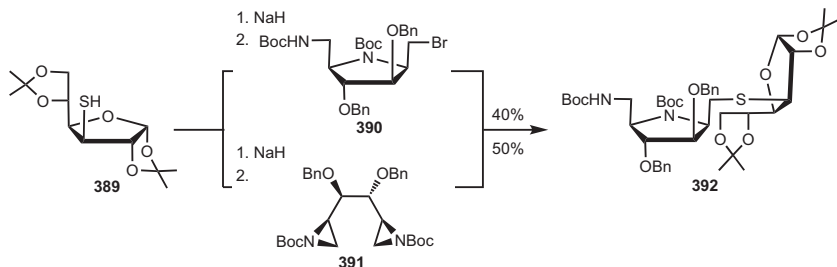
Scheme 60

with the glycosylation reaction (HCl, MeOH) in order to give **383** (76%, 3 steps) and **384** (79%, 3 steps) as mixtures of anomers (Fig. 23).

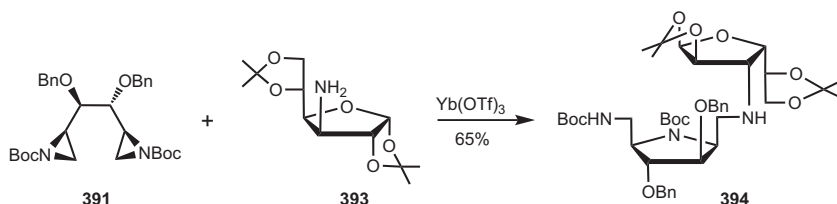
Interestingly, as shown in a paper published⁹² in 2000, the use of a lithium enolate of **378** induced β -elimination of the hydroxyl group resulting from the condensation of **378** with the protected iminosugar **385** (Scheme 60). In that case, aldolisation product **386** was obtained in 78% yield. Later, even if the reduction of the ketone (LiAlH_4) followed by the selective hydrogenolysis of the benzyl ether went smoothly, the methanolysis step proved more difficult than expected. In fact a series of protection and deprotection steps were necessary to afford fully protected imino-C-disaccharide **388** in very low yield (18%, 6 steps) together with the non hydrolysed compound **387** (19%, 6 steps).

Back in 1996, Depeyaz⁹³ and co-workers described the one-step synthesis of *S*- or *N*-linked imino-C-disaccharides (Scheme 61). 3-Deoxy-3-thio-D-glucose **389** and 3-deoxy-3-amino-D-glucose **393** (see Scheme 62) were prepared^{94,95} in three steps starting from diacetone-D-allose. Thiol **389** was reacted with the protected bromo-iminosugar **390** in the presence of NaH to give the desired disaccharide **392** (40%). To obtain the same product, the substitution reaction was also performed with the bis-aziridine **391** instead of **390**. Nucleophilic opening of **391** followed by intramolecular aminocyclization gave **392** in 50% yield.

The second strategy was used to synthesize the *N*-linked imino-C-disaccharide **394** (Scheme 62). The bis-aziridine **391** was reacted with the



Scheme 61



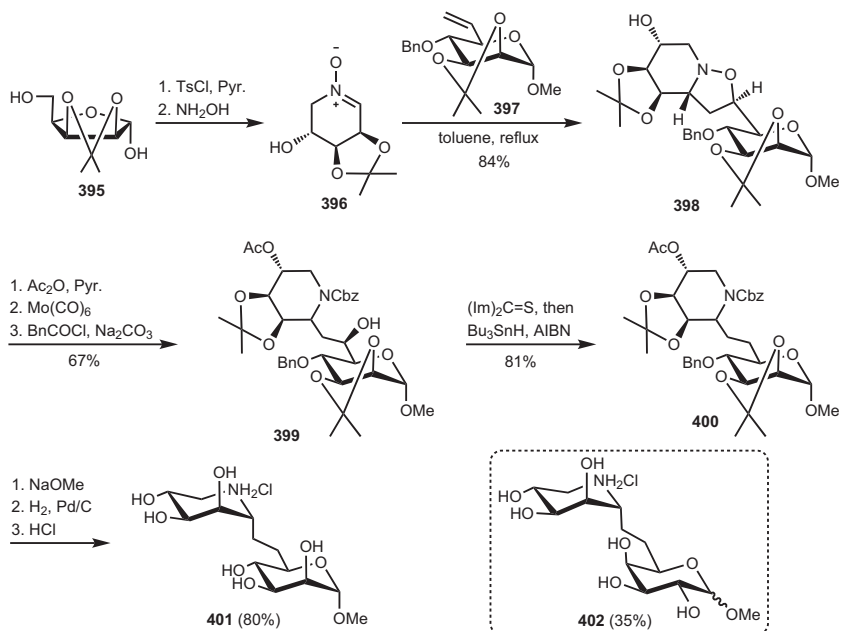
Scheme 62

amine **393** in the presence of $\text{Yb}(\text{OTf})_3$ to give **394** in good yield (65%). Unfortunately, the deprotection of **392** and **394** was not attempted by the Authors leaving this new derivatives as fully protected imino-*C*-disaccharides.

In 2000, Wightman and co-workers exploited⁹⁶ stereoselective cycloaddition reactions between functionalized cyclic nitrones and sugar alkenes for the synthesis of imino-*C*-disaccharides. In this work, the cyclic nitrone **396** was synthesized in 2 steps from 2,3-*O*-isopropylidene-*D*-xylose **395** (Scheme 63). Then, nitrone **396** was reacted with the vinyl mannoside **397** in refluxing toluene to afford the cycloadduct **398** (84%). After acetylation, the reductive cleavage of the N–O bond was performed using $\text{Mo}(\text{CO})_6$ to give, after protection of the resulting free amine, the disaccharide **399** (67%, 3 steps). Deoxygenation of the latter took place using an excess of thiocarbonyldiimidazole at high concentration in order to obtain **400** in good yield (81%). Routine deprotection steps led to **401** (80%, 3 steps). The same synthesis was then carried using a *D*-galactopyranoside alkene to obtain the imino-*C*-disaccharide **402** (35%, 8 steps).

In 2008, using two different glycoside alkenes (**403**, **404**), as well as two different benzylated cyclic nitrones (**405**, **406**), Argyropoulos and co-workers synthesized⁹⁷ four imino-*C*-disaccharides (**407–410**) bearing an hydroxyl group onto the ethyl linker (72–81%, Fig. 24). In this case, the reductive cleavage of the N–O bond was performed using hydrogenolysis (H_2 , Raney Ni). Unfortunately, none of them were deprotected to afford free-OH disaccharides that could be assayed as glycosidases inhibitors.

The same year, Martin and co-workers performed⁹⁸ the same reaction sequence using water as the solvent (Scheme 64). The deprotected nitrone **411** was reacted with methyl 6-*O*-allyl- β -*D*-galactofuranoside **412** (H_2O , 60 °C) to give the cycloaddition product **413** in 44% yield. It is worth noting that at this temperature the reaction takes 10 days to reach



Scheme 63

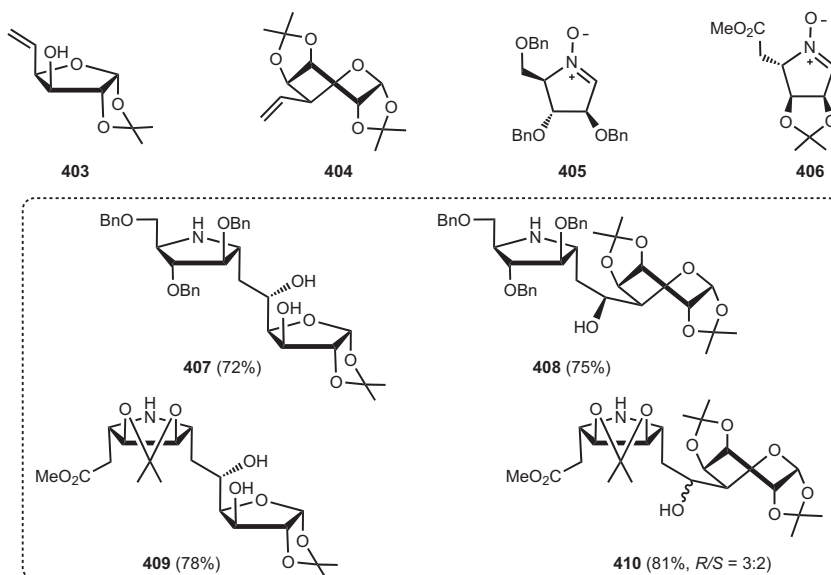
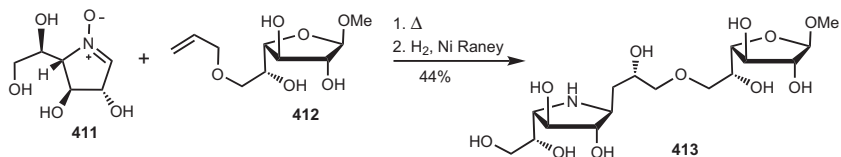


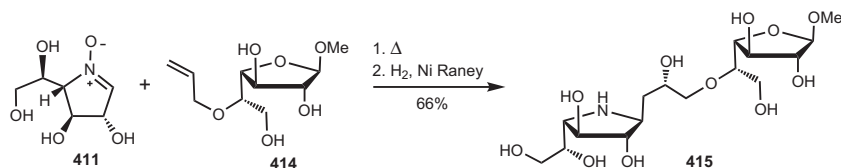
Fig. 24 Imino-C-disaccharides and their precursors prepared by Argyropoulos and co-workers.⁹⁷

completion. The use of water as solvent allowed to carry out the reaction with deprotected compounds to afford directly a free-OH imino-C-disaccharide.

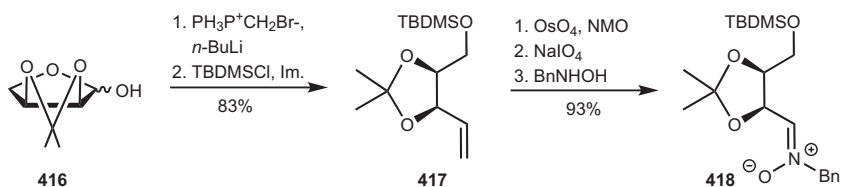
The 5-O-allyl- β -D-galactofuranoside **414** was coupled with the same nitrone **411** to give, after reductive cleavage (H_2 , Raney Ni), the GalF-disaccharide mimic **415** (66%, 2 steps, Scheme 65).



Scheme 64



Scheme 65

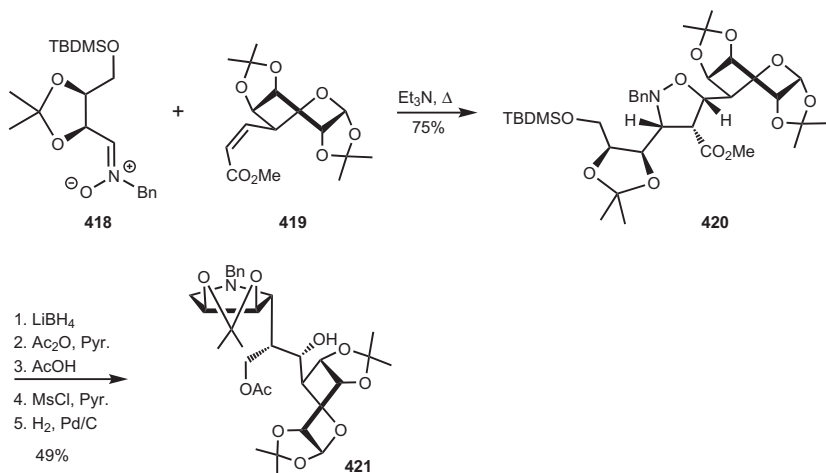


Scheme 66

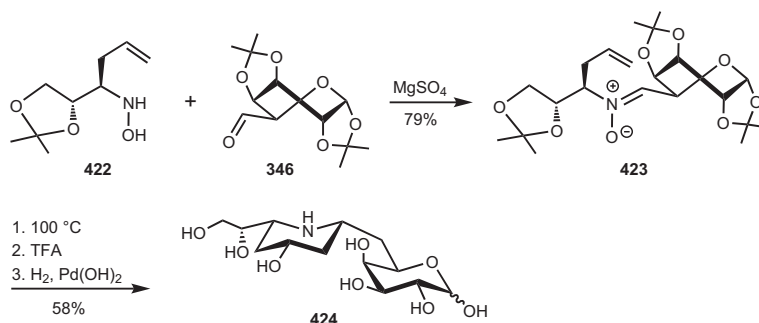
In 2004, Argyropoulos and Sarli used⁹⁹ an open chain nitron as precursor for the iminosugar unit. The synthesis of the nitron **418** (Scheme 66) started with the Wittig olefination ($\text{Ph}_3\text{P}^+\text{CH}_2\text{Br}^-$, $n\text{-BuLi}$) of the L -erythrose monoacetonide **416** followed by the silylation of the newly formed hydroxyl (TBDMSCl, imidazole) to give alkene **417** (83%, 2 steps). The latter was then transformed into the desired nitron **418** (93%, 3 steps).

As for the previous cycloadditions, **418** was heated in refluxing benzene with the galactoside derivative **419** and Et_3N to give the cycloadduct **420** (Scheme 67). After reduction of the methyl ester (LiBH_4) and acetylation of the resulting hydroxyl group, the silyl ether was cleaved and replaced by a mesylate triggering an intramolecular cyclization to form the desired pyrrolidine ring. Reductive cleavage of the N–O bond performed by catalytic hydrogenolysis led to the imino- C -disaccharide **421** (49%, 5 steps) that was not submitted to deprotection.

Following these results, Merino and co-workers described¹⁰⁰ in 2012 a direct high-yielding approach towards imino- C -disaccharides using an intramolecular cycloaddition of alkenyl glycosyl nitrones, the same strategy exploited three years before for the synthesis of imino- C -disaccharide bearing the iminosugar unit at the reducing end (see section 5.2, Scheme 74). In this approach, the nitron is directly formed through reductive amination between an alkenyl hydroxylamine and a sugar bearing an aldehyde function. For instance, hydroxylamine **422** was condensed with the galactose-derived aldehyde **346** in presence of MgSO_4



Scheme 67



Scheme 68

to give the alkenyl nitron **423** in 79% yield (Scheme 68). Then, intramolecular cycloaddition took place at 100 °C followed by deprotection of hydroxyl groups (TFA) and reductive cleavage of the N–O bond to give the 1,6-linked imino-*C*-disaccharide **424** in good yield (58%, 3 steps).

Thereafter, the same approach was used to obtain the disaccharides **425**–**427** in good yields (44–51%) as well as the trisaccharide **428** (33%, Fig. 25).

In 2000 and later in 2002, Dondoni and co-workers described¹⁰¹ the use of the Wittig reaction as a privileged route towards imino-*C*-disaccharides justified by the ready preparation of the relevant reagents, *i.e.* the sugar aldehyde and phosphorous ylide. In a first example (Scheme 69), the ylide generated *in situ* from the *D*-galactopyranose phosphonium iodide **429** was reacted with the aldehyde **430** to give the corresponding alkene **431** (64%). The reduction of the double bond was carried out in presence of *p*-toluenesulfonylhydrazide, followed by hydrogenolysis and treatment with Amberlite IR120 gave the deprotected 1,6-linked imino-*C*-disaccharide **432** (72%, 3 steps).

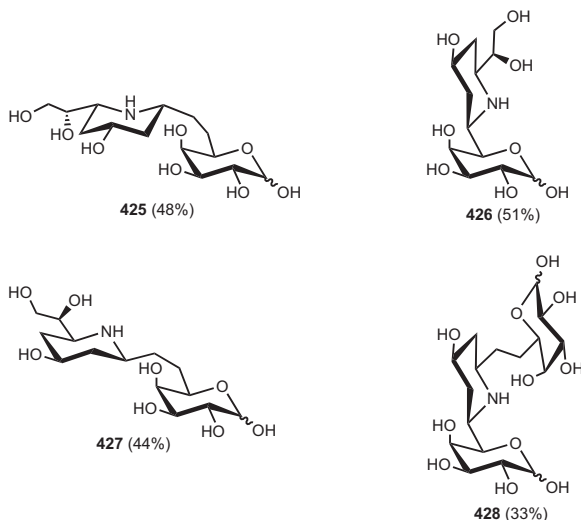
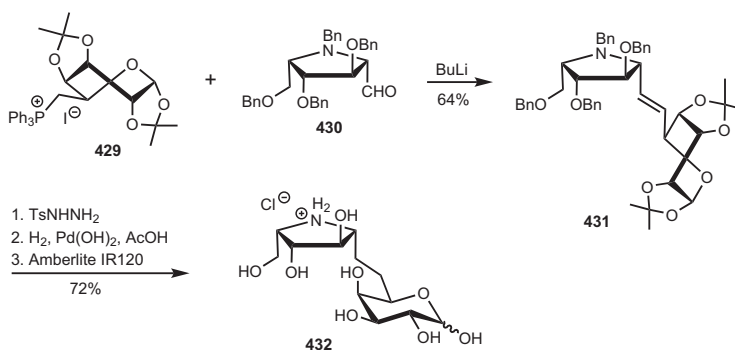


Fig. 25 Imino-C-di- and -trisaccharides prepared by Merino and co-workers.¹⁰⁰



Scheme 69

Then, a ribofuranosyl phosphonium iodide and other sugar aldehydes were used for the preparation of five more imino-C-disaccharides (**433–437**, Fig. 26) in moderate yields (23–35%). However, the original D-ribo configuration was not retained in the disaccharides **434**, **435** and **437**, which featured an L-lyxo configuration.

In 2004, Sharma and co-workers proposed¹⁰² the use of C-linked carbo-β-amino acids as precursors for the synthesis of imino-C-disaccharides (Scheme 70). Hydrogenolysis of the known¹⁰³ ester **438**, followed by reduction, protection ((Boc)₂O, Et₃N) and oxidation (IBX) steps, provided a terminal aldehyde that was subjected to Wittig olefination giving the alkene **439** (43%, *Z/E* = 1.5 : 1, 5 steps). After the Boc removal, only the *cis* isomer was converted into the lactam **440** (DMAP, 51%) while a *N*-Boc protection was necessary to get rid of the remaining *trans* isomer. Finally, di-hydroxylation of **440** in the presence of OsO₄ and NMO gave **441** in 54% yield. Following the same pathway, the imino-C-disaccharide **443**

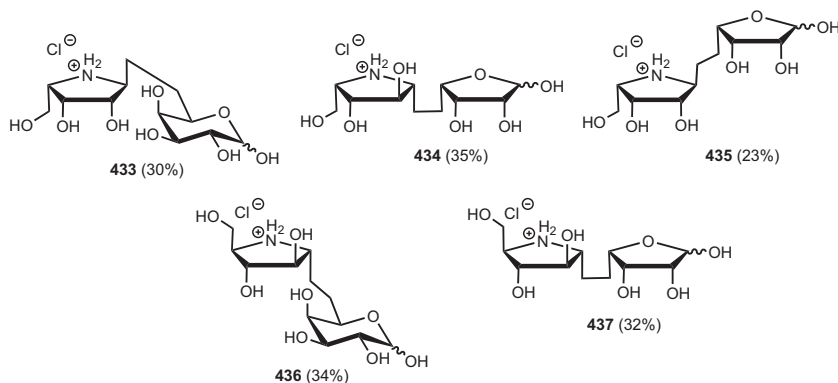
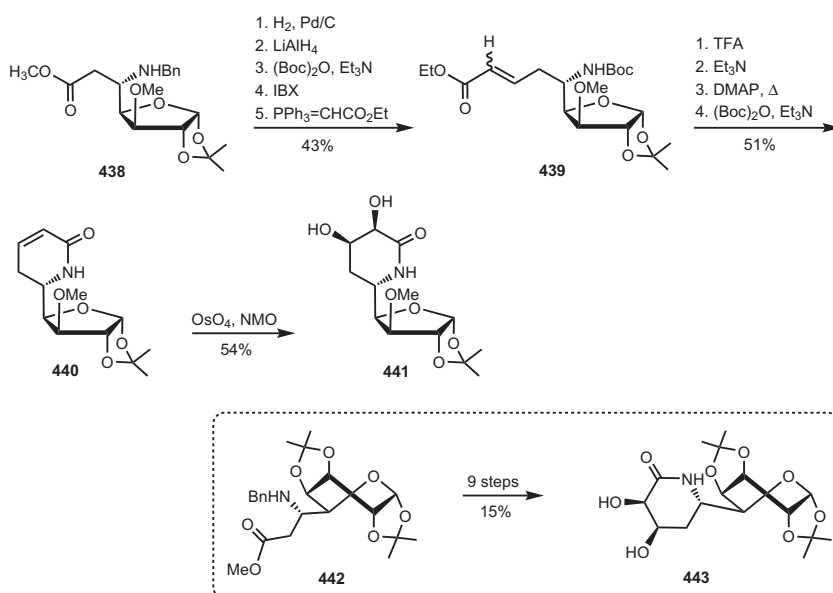


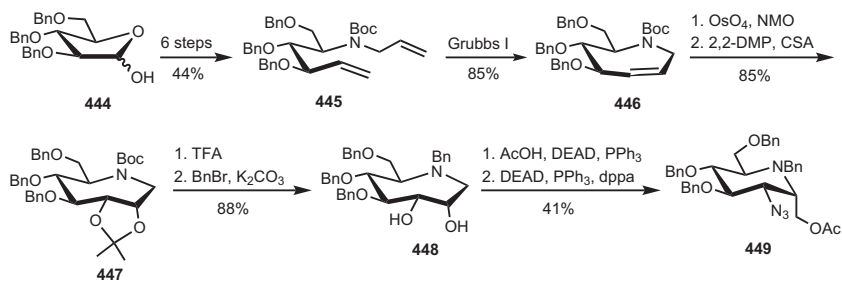
Fig. 26 Imino-*C*-disaccharides prepared by Dondoni and co-workers.¹⁰¹



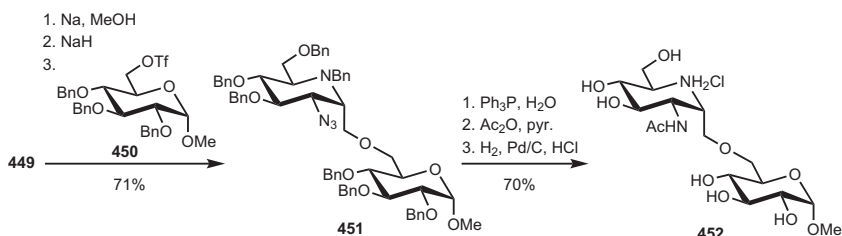
Scheme 70

was obtained starting from ester **442** in 15% yield over 9 steps. Once again, the deprotection of **441** and **443** was not attempted by the Authors.

The last example of imino-*C*-disaccharide bearing the sugar moiety at the reducing end was published¹⁰⁴ by Blériot, Sollogoub and their co-workers in 2014. The key step was the coupling between a 1,2-*cis*-homoinminosugar and an activated methyl glucoside derivative. To this aim, homonojirimycin **449** (Scheme 71) was synthesized starting with the formation of di-alkene **445** from the *D*-arabinofuranose derivative **444** (44%, 6 steps). A ring closing metathesis was then performed (Grubbs I gen.) to get the cyclic alkene **446** (85%) which was oxidized to the fully protected azepane **447** (85%, 2 steps). An acidic treatment ($\text{CF}_3\text{CO}_2\text{H}$) followed by the benzylation of the resulting secondary amine led to the



Scheme 71



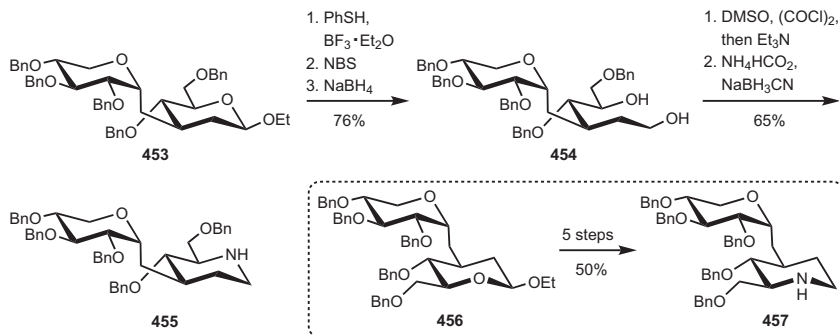
Scheme 72

diol **448** in 88% overall yield. The ring contraction occurred by treatment of **448** with DEAD and triphenylphosphine in acetic acid at 0 °C to give, after azidation at the C-2 position (DEAD, PPh₃, dppa), the 2-azidopiperidine **449** (41%, 2 steps) displaying an α -D-*gluco* configuration confirmed by NMR analysis.

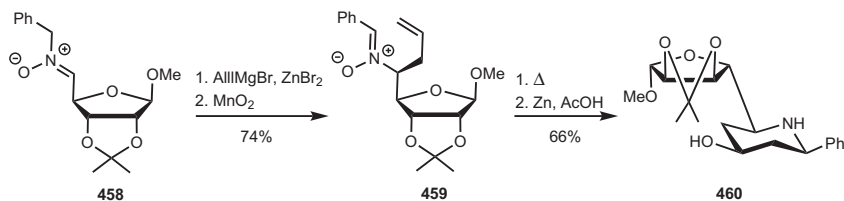
After deacetylation of **449**, the resulting alcohol was treated with NaH and reacted with the known¹⁰⁵ activated sugar **450** (Scheme 72) to give the corresponding imino-*C*-disaccharide **451** (71%, 3 steps). Generation of the acetamido moiety (PPh₃, H₂O, then Ac₂O) followed by hydrogenolysis of the benzyl ethers led to the fully deprotected α -1,6-imino-*C*-disaccharide **452** in 70% yield.

5.2 Imino-*C*-disaccharides bearing the iminosugar moiety at the reducing end

Very few examples of imino-*C*-disaccharides bearing the iminosugar moiety at the reducing end were reported in the literature. The first one was published¹⁰⁶ in 2007 by Kniesz and co-workers in order to synthesize analogues of naturally occurring 3-*O*-(β -D-glucopyranosyl)-fagomine. Their strategy was to modify an already synthesized α -1,3-linked *C*-disaccharide (Scheme 73). In a previous work, the same research team prepared D- (**453**) and L-disaccharides (**456**) that served as starting material for the synthesis of the target imino-*C*-disaccharides. First, hydrolysis of the ethyl glycoside **453** was performed in two steps (PhSH, BF₃·Et₂O, then NBS) followed by reduction of the resulting hemiacetal to give the diol **454** (76%, 3 steps). Classical oxydation (DMSO, (COCl)₂ then Et₃N) and double reductive amination steps (NH₄⁺HCO₂⁻) gave the perbenzylated imino-*C*-disaccharide **455** (65%, 2 steps). Starting from



Scheme 73



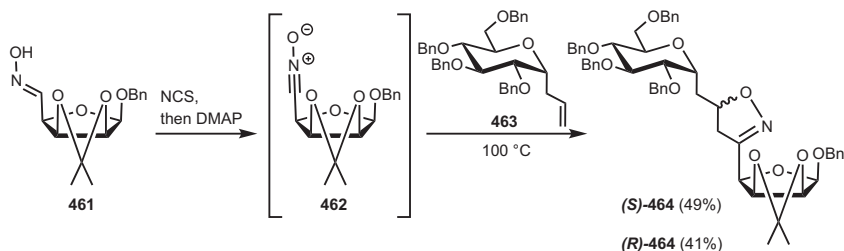
Scheme 74

ethyl glycoside **456** and following the same strategy, the *L*-configured imino-*C*-disaccharide **457** was obtained in 50% yield (5 steps). Unfortunately, the resulting disaccharides were not deprotected thus preventing any biological evaluation on glycosidase inhibition.

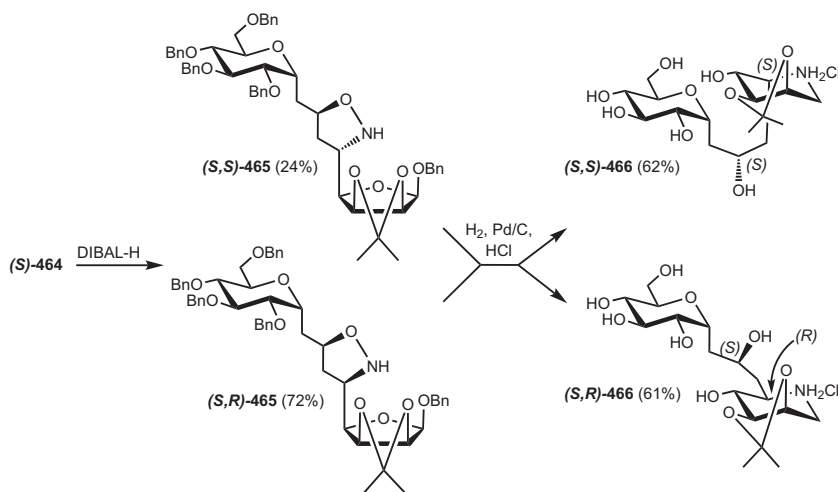
As discussed above (section 5.1, Scheme 68), Merino and co-workers took advantage of the intramolecular 1,3-dipolar cycloaddition of *N*-alkenyl nitrones to obtain¹⁰⁷ a fully protected α -1,6-linked imino-*C*-disaccharide bearing the iminosugar at the reducing end. To this aim, *D*-ribose-nitron **458** (Scheme 74) was allylated using allyl Grignard and ZnBr_2 as additive, followed by oxidation of the resulting hydroxylamine (MnO_2) to obtain the *N*-alkenyl nitron **459** in 74% yield (2 steps). Heating the latter at 100 °C in toluene afforded the corresponding cycloadduct from which the newly formed N-O bond was cleaved (Zn , AcOH) to give the fully protected imino-*C*-disaccharide **460** (66%, 2 steps). It is worth of note that the all-*cis* configuration of **460** was determined after resolution of its crystalline structure.

Finally, the last example of an imino-*C*-disaccharide bearing the iminosugar unit at the reducing end was published¹⁰⁸ in 2014 by Li and co-workers. The unstable sugar derived nitrile oxide **462** (Scheme 75) was prepared by chlorination of the oxime¹⁰⁹ **462** using *N*-chlorosuccinimide followed by DMAP. The allyl *C*-glucopyranoside **463** was then added and the resulting mixture was heated at 100 °C to give a mixture of easily isolated isomers (*S*)-**464** and (*R*)-**464** in 49 and 41% yield, respectively.¹¹⁰

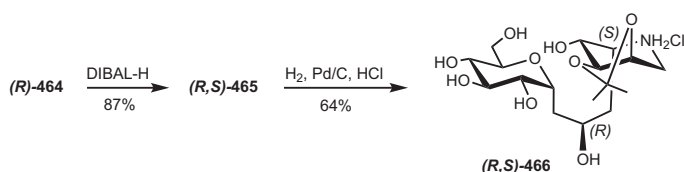
Isoxazoline (*S*)-**464** was then reduced with DIBAL-H to give a 3:1 mixture of intermediates (*S,R*)-**465** and (*R,R*)-**465** in 72 and 24% yield, respectively (Scheme 76). The imino-*C*-disaccharides (*S,R*)-**466** and



Scheme 75



Scheme 76



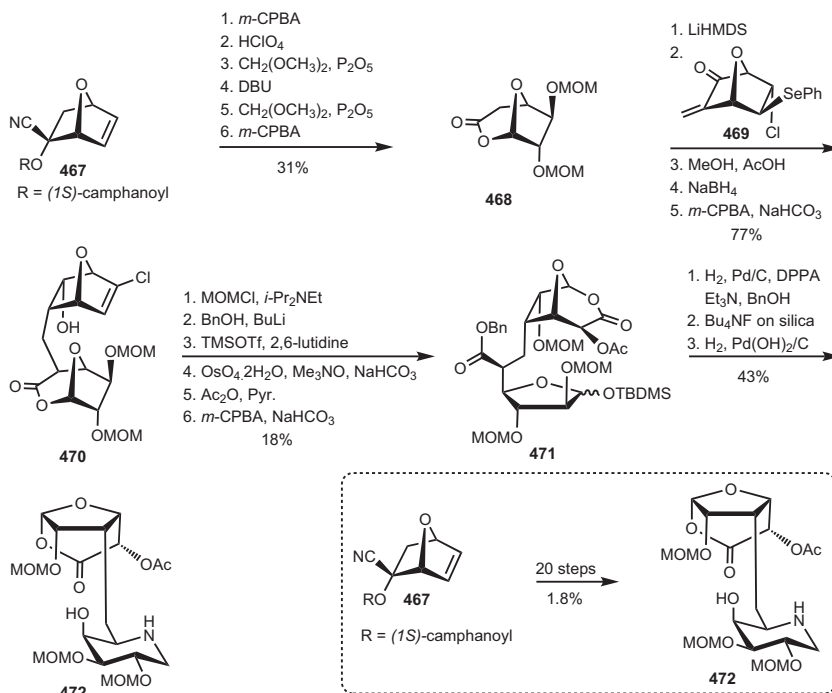
Scheme 77

(R,R)-466 were then obtained (61 and 62% yield, respectively) *via* tandem multi-step reactions including debenzoylation and N–O bond cleavage to form the free amine thus allowing the intramolecular condensation-cyclization step.

Using the same methodology, the disaccharide **(R,S)-466** was obtained from **(R)-464** in 56% yield over 2 steps (Scheme 77).

5.3 Non-anomerically linked imino-*C*-disaccharide analogues

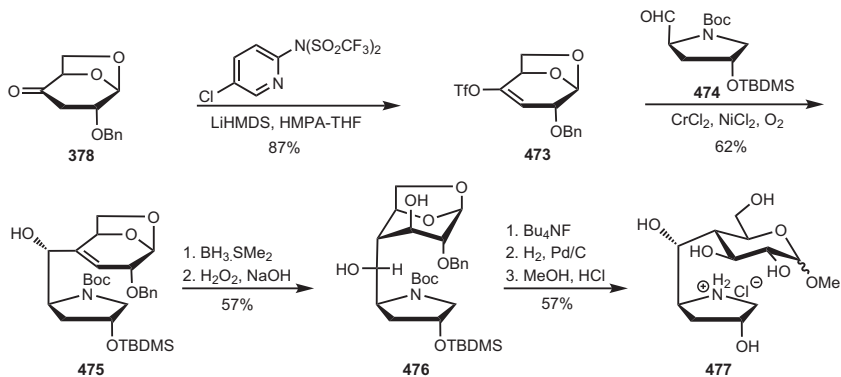
Non-anomerically linked disaccharides can be considered as the last category of imino-*C*-disaccharides. This section includes all imino-*C*-disaccharides where none of the anomeric positions are involved in the link between the iminosugar and the sugar moiety. The first example of



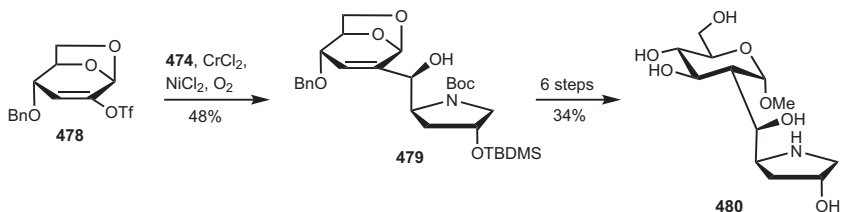
Scheme 78

such disaccharides was reported¹¹¹ by Vogel and co-workers in 1996. This article described the 20 steps synthesis of the partially protected imino-*C*-disaccharide **472** (1.8% total yield) starting from the optically pure 7-oxabicyclo[2.2.1]hept-5-en-2-yl derivative **467** (Scheme 78). The key step of this long total synthesis was the Michael addition of the lithium enolate of lactone¹¹² **468** to the enone¹¹³ **469** followed by reduction and oxidation steps to give the lactone **470** in 77% yield. Further protection, ring opening and oxidation steps allowed to obtain lactone **471** (18%) from which the iminosugar moiety was generated. Indeed, debenzoylation (H₂, Pd/C) followed by reaction of the freshly formed carboxylic acid with DPPA and Et₃N, led to an isocyanate that was treated without purification with benzyl alcohol to give the corresponding benzyl carbamate. Then, cleavage of the silyl group (Bu₄NF) followed by the deprotection of the carbamate (H₂, Pd(OH)₂/C), afforded the imino-*C*-disaccharide **472** (43%) resulting from the formation of an imine that was then hydrogenated.

In their effort to achieve a more convergent synthesis of imino-*C*-disaccharides, Zhu and Vogel planned¹¹⁴ to take advantage of the Nozaki-Kishi coupling (Scheme 79). First, the already used isolevoglucosenone adduct **378** (see Scheme 59) was transformed into its enol triflate **473** using LiHMDS and 2-[bis(tri-fluoromethylsulfonyl)amino]-5-chloropyridine (87%). Nozaki-Kishi coupling of **473** with aldehyde **474** led to the formation of the allylic alcohol **475** as the major product of the coupling reaction (62%). Hydroboration with BH₃·SMe₂ followed by H₂O₂/NaOH



Scheme 79



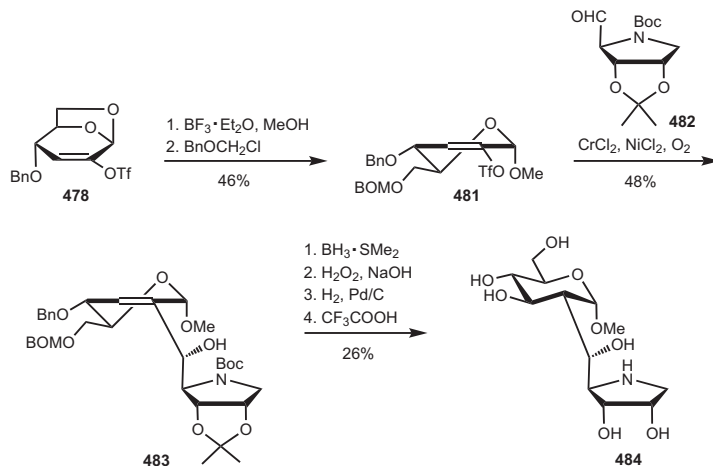
Scheme 80

work-up led to the 1,6-anhydro-glucose derivative **476** (57%) that was then desilylated, debenzylated and submitted to acidic methanolysis to give the 4,4-linked imino-*C*-disaccharide **477** as a 2 : 1 mixture of anomers.

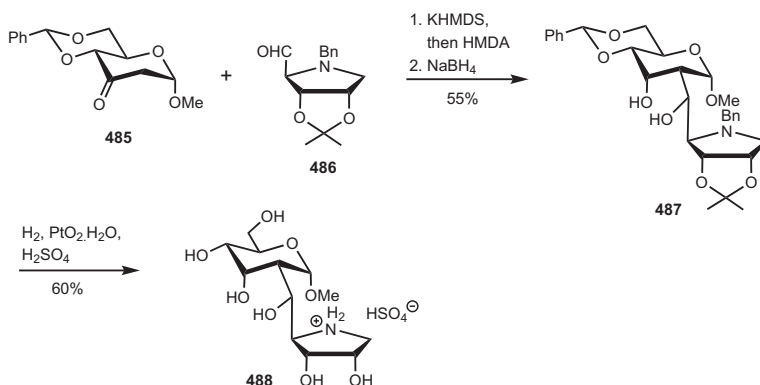
Using the same method, the disaccharide **480** was obtained as a single anomer (16%, 7 steps) starting from the enol triflate **478** (Scheme 80).

Three years later, the same team applied¹¹⁵ the above mentioned method for the synthesis of another imino-*C*-disaccharide in which a pyrrolidine-3,4-diol is attached at the C-2 position of an α -D-glucopyranoside moiety (Scheme 81). In this case, the direct Nozaki-Kishi coupling between enol triflate **478** and aldehyde **482** took place in very low yield. Therefore, it was decided to convert the bicyclic triflate **478** into the monocyclic derivative **481** (46%) through acidic methanolysis (BF₃·Et₂O, MeOH) followed by protection of the newly formed primary alcohol (BOMCl). Then, the Nozaki-Kishi coupling between enol **481** and aldehyde **482** went smoothly providing allylic alcohol **483** in 48% yield. It should be noted that the reaction between **481** and **482** generated the (*S*)-alcohol in contrast with the reaction between **478** and **474** where the (*R*)-alcohol was obtained. Finally, the imino-*C*-disaccharide **484** was isolated in 26% overall yield *via* the same hydroboration/deprotection sequence employed for the synthesis of **480**.

In their last effort, Vogel and co-workers described¹¹⁶ the synthesis of imino-*C*-disaccharides bearing D-gluco- and D-allopyranoside sugar moieties. To this aim, cross-aldol condensation of the readily available¹¹⁷ methyl 4,6-*O*-benzylidene-2-deoxy- α -D-*erythro*-hexopyranosid-3-ulose **485** (Scheme 82) with the unstable aldehyde **486** (KHMDS, then HMDA),



Scheme 81



Scheme 82

followed by reduction of the ketone gave the diol **487** as the major product in 55% yield. The absolute configuration of the hydroxyl-methylene linker could not be determined even if the Zimmerman-Traxler model suggests the (*S*)-configuration. The target disaccharide **488** was then obtained in 60% yield after hydrogenolysis using Adam's catalyst in acidic media.

Following another pathway, the diol **487** was converted, in the presence of DAST, into enones (*E*)-**489** and (*Z*)-**489** that could be isolated by flash chromatography on silica gel in 34 and 49% yields, respectively (Scheme 83). Reduction and deprotection steps on (*E*)-**489** afforded **490** featuring the *D*-gluco configuration (48%, 4 steps), whereas the same reaction sequence applied to (*Z*)-**489** afforded the *D*-allo configured **491** disaccharide in 47% overall yield.

In 1997, Goti, Brandi and their co-workers exploited¹¹⁸ the cycloaddition of cyclic nitrones to glycals to synthesize new imino-*C*-disaccharides. Indeed, they obtained 6 different cycloadducts (**494–499**, Fig. 27) in low to

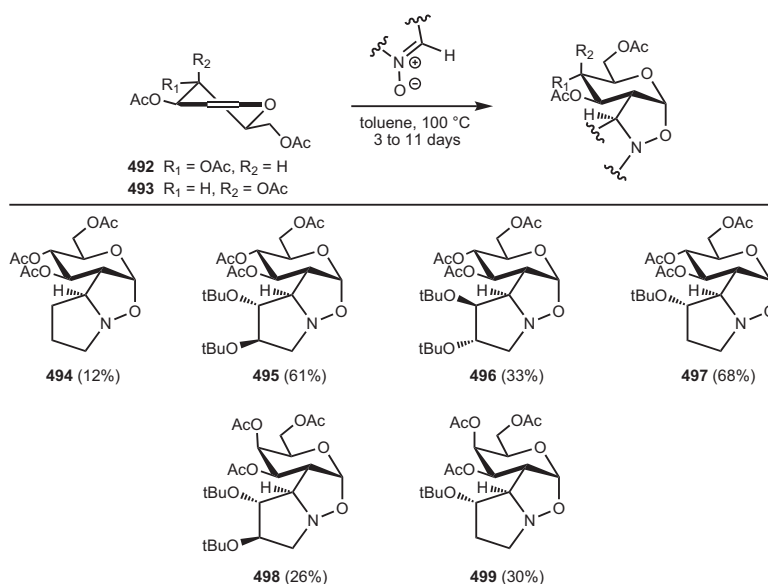
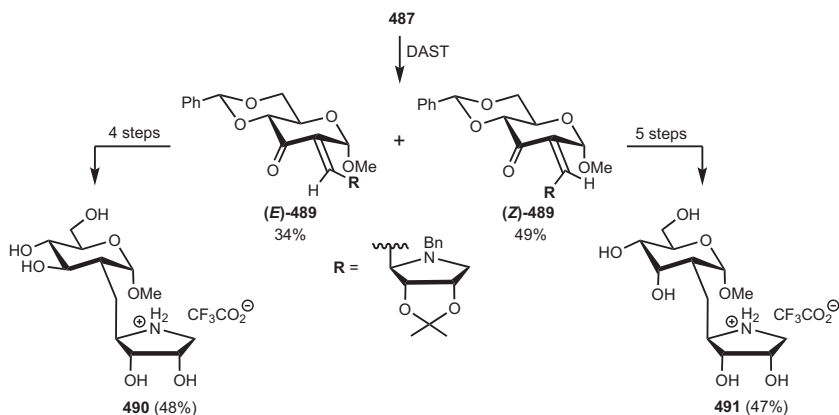
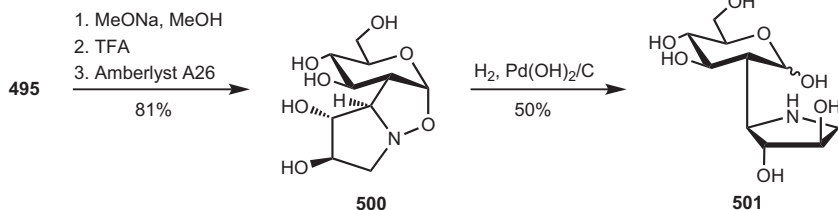


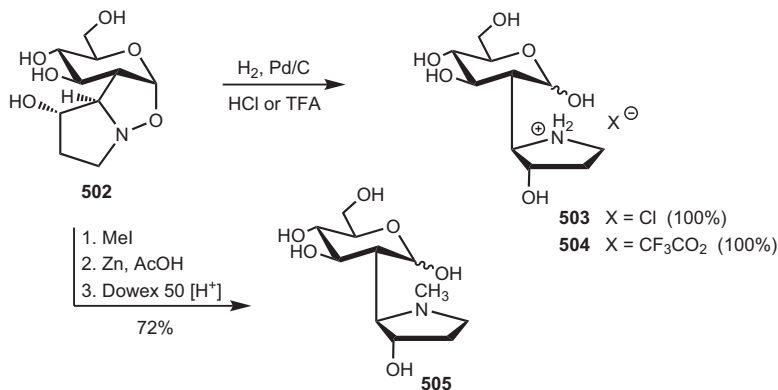
Fig. 27 Adducts obtained via cycloaddition of cyclic nitrones to glycals by Goti, Brandi and their co-workers.¹¹⁸

good yields (12–68%) starting from D-glucal **492** or D-galactal **493** and various enantiopure nitrones. These cycloaddition reactions were carried out in toluene at 100 °C for several days using an excess of glycals. These drastic conditions were necessary for complete conversion of the nitrones. Unfortunately, only the cycloadduct **495** was brought further in the synthesis (see Scheme 84).

The protecting groups of the isoxazolidine ring in **495** (Scheme 84) were removed to give the free-OH cycloadduct **500** (81%). Then, cleavage of the N–O bond through a reductive opening of **500** (H_2 , $\text{Pd}(\text{OH})_2/\text{C}$) afforded an anomeric mixture of the desired 2,4-linked imino-*C*-disaccharide **501** in 50% yield.



Scheme 84



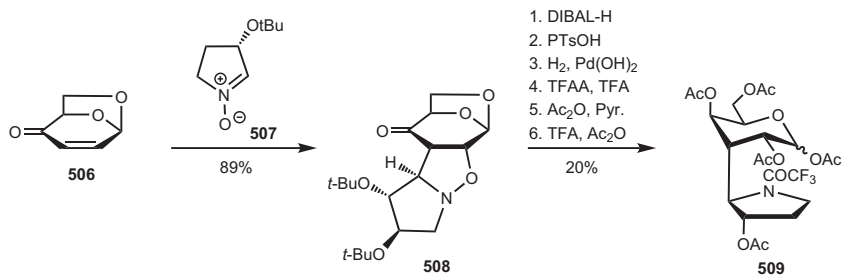
Scheme 85

One year later, starting from the deprotected isoxazolidine derived from **497**, the same research team was able to prepare¹¹⁹ two new imino-*C*-disaccharides (Scheme 85). Indeed, by catalytic hydrogenolysis of **502** in the presence of HCl or TFA, both salts **503** and **504** were obtained in quantitative yield. Moreover, adding a methylation step (MeI) before the cleavage of the N–O bond led to the *N*-methyl imino-*C*-disaccharide **505** (72%, 3 steps).

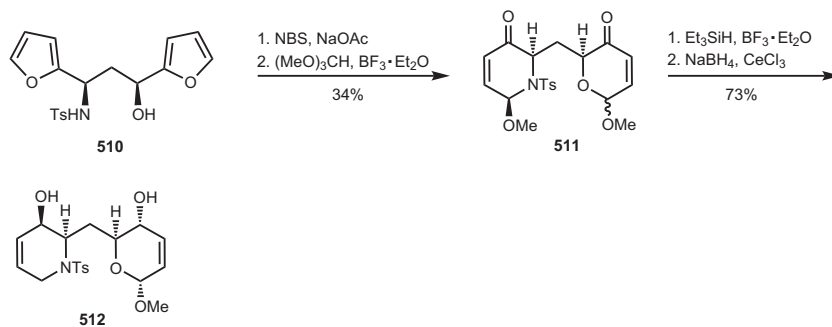
Finally, using isolevoglucosene **506** and the cyclic nitrone **507**, Cardona and co-workers were able to obtain¹²⁰ isoxazolidine **508** in 89% yield (Scheme 86). Reduction of ketone **508** with DIBAL-H at -78°C gave only the *endo* alcohol. Then, ring opening N–O bond reductive cleavage (H_2 , $\text{Pd}(\text{OH})_2$), followed by several deprotection/protection steps led to the protected imino-*C*-disaccharide **509** (20%, 6 steps).

In 2005, an alternative approach to imino-*C*-disaccharides was proposed¹²¹ by Nelson and co-workers using the Upjohn and Donohoe dihydroxylation methods as key steps. Their strategy consisted of the oxidative ring expansion of the di(2-furyl) amino alcohol derivative **510** followed by the elaboration of the newly formed bis-enone **511** to afford **512** in 25% overall yield (Scheme 87).

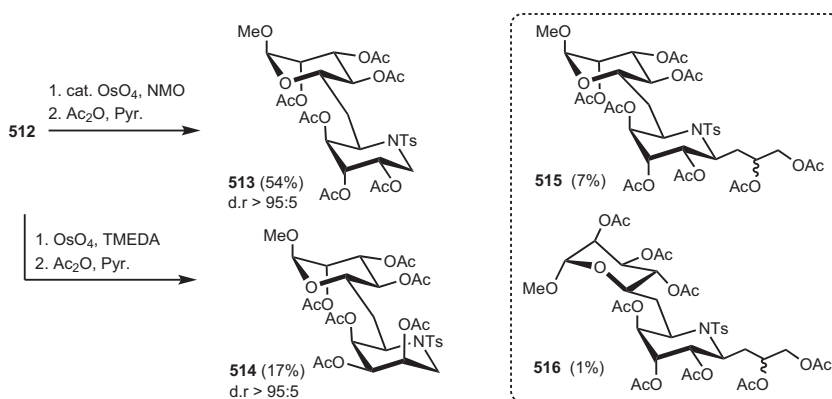
Both dihydroxylation protocols were then applied to **512** (Scheme 88) followed by a per-acetylation step leading to the imino-*C*-disaccharides **513** and **514** in 54 and 17% yield, respectively (d.r. > 95:5). Using the same strategy, the disaccharides **515** and **516** were also synthesized in 7 and 1% yield, respectively, over 6 steps.



Scheme 86



Scheme 87



Scheme 88

A special class of imino-*C*-disaccharides, *i.e.* homo- and heterodimers of nojirimycin linked through their (pseudo)anomeric positions, was reported¹²² by Cipolla, Cardona and their co-workers as new iminosugar-based trehalase inhibitors. In this work, both compounds were prepared starting from allyl α -*C*-nojirimycin¹²³ **517** (Scheme 89) exploiting the cross-metathesis reaction. Homodimer **518** was synthesized using Grubbs 2nd generation catalyst followed by the concomitant reduction of the double bond, hydrogenolysis of the benzyl ethers and the Cbz group (51%). On the other hand, the heterodimer **519** was obtained by the same reaction sequence in the presence of allyl α -*C*-glucopyranoside **463** (32%).

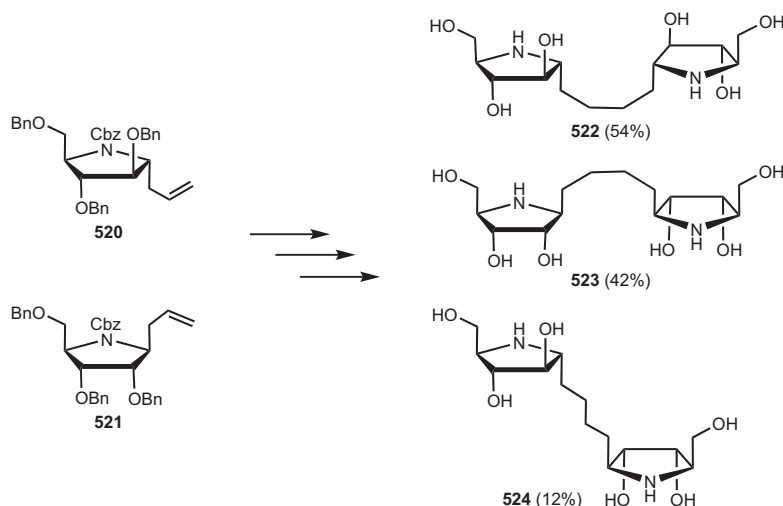
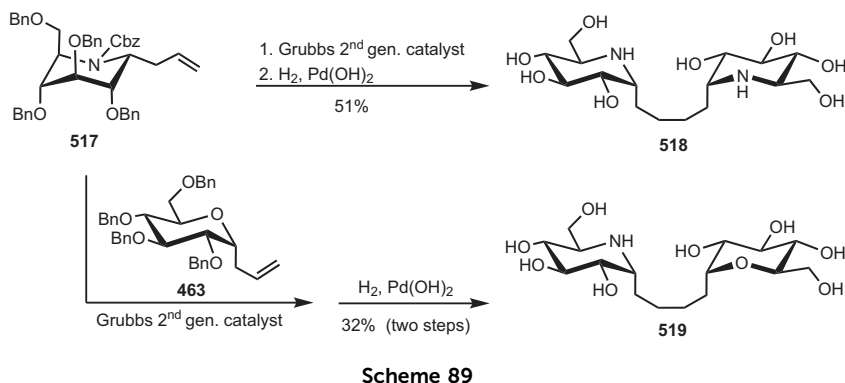


Fig. 28 Homodimers and heterodimers of iminosugars prepared by Cipolla, Cardona and their co-workers.¹²²

Moreover, combining allyl α -C-imino-arabinofuranoside **520** and allyl β -C-imino-ribofuranoside **521**, homodimers **522** and **523** were prepared in 54 and 42% yield, respectively, while heterodimer **524** was obtained in 12% overall yield (Fig. 28).

6 Glycosidase inhibition properties

Since 1985 and the first occurrence of an imino-disaccharide in the literature, no more than 235 compounds (di- and oligosaccharides) were synthesized. Unfortunately, only 69 of them were biologically evaluated. Indeed, often the newly prepared disaccharides were not deprotected thus preventing any evaluation of their activities against glycosidases or glycosyltransferases. Among these 70 compounds, a large majority are imino-*O*-disaccharides (47) and only a few of them are *S*- (2), *N*- (4) and *C*-disaccharides (6). In order to get a general view, the following Table 1 describes in a comprehensive manner the inhibition activities (K_i or IC_{50})

Table 1 Glycosidase and glycosyltransferase inhibition values (K_i or IC_{50} , μM) found for the known (ref.) imino-disaccharide and imino-oligosaccharide derivatives described in the previous sections (2 to 5) of the present chapter.

Enzyme	Entry	Compound	K_i (μM)	IC_{50} (μM)	Ref.	Section
α-Amylase						
<i>Human pancreatic</i>	1	3 ($n = 1$)		20 ^a	9	2
	2	3 ($n = 2$)		41 ^a	9	2
	3	3 ($n = 3$)		14 ^a	9	2
	4	3 ($n = 4$)		12 ^a	9	2
	5	3 ($n = 5$)		11 ^a	9	2
	6	3 ($n = 6$)		10 ^a	9	2
	7	3 ($n = 7$)		10 ^a	9	2
	8	31		500	13	2
	9	32		43	13	2
	10	33		30	13	2
	11	34		230	13	2
	12	43		3000	13	2
	13	44		34	13	2
	14	45		25	13	2
	15	46		70	13	2
	16	49	265		15	2
<i>Human salivary</i>	17	3 ($n = 1$)		21 ^a	9	2
	18	3 ($n = 2$)		40 ^a	9	2
	19	3 ($n = 3$)		14 ^a	9	2
	20	3 ($n = 4$)		12 ^a	9	2
	21	3 ($n = 5$)		9 ^a	9	2
	22	3 ($n = 6$)		9 ^a	9	2
	23	3 ($n = 7$)		10 ^a	9	2
	24	31		2000	13	2
	25	32		82	13	2
	26	33		49	13	2
	27	34		360	13	2
	28	43		2000	13	2
	29	44		46	13	2
	30	45		42	13	2
	31	46		70	13	2
<i>Bacterial liquefying</i>	32	3 ($n = 3$)		5 ^a	9	2
	33	3 ($n = 4$)		24 ^a	9	2
	34	3 ($n = 5$)		29 ^a	9	2
	35	3 ($n = 6$)		24 ^a	9	2
<i>Bacterial saccharifying</i>	36	3 ($n = 7$)		23 ^a	9	2
	37	3 ($n = 1$)		91 ^a	9	2
	38	3 ($n = 2$)		93 ^a	9	2
	39	3 ($n = 3$)		87 ^a	9	2
	40	3 ($n = 4$)		85 ^a	9	2
	41	3 ($n = 5$)		85 ^a	9	2
	42	3 ($n = 6$)		85 ^a	9	2
	43	3 ($n = 7$)		85 ^a	9	2
	44	4		44 ^a	9	2
Taka amylase A	45	3 ($n = 1$)		13 ^a	9	2
	46	3 ($n = 2$)		11 ^a	9	2
	47	3 ($n = 3$)		23 ^a	9	2
	48	3 ($n = 4$)		27 ^a	9	2
	49	3 ($n = 5$)		29 ^a	9	2
	50	3 ($n = 6$)		28 ^a	9	2
	51	3 ($n = 7$)		32 ^a	9	2

Table 1 (Continued)

Enzyme	Entry	Compound	K_i (μM)	IC_{50} (μM)	Ref.	Section
<i>Aspergillus niger</i> (acid stable)	52	3 ($n = 1$)		91 ^a	9	2
	53	3 ($n = 2$)		94 ^a	9	2
	54	3 ($n = 3$)		95 ^a	9	2
	55	3 ($n = 4$)		95 ^a	9	2
	56	3 ($n = 5$)		93 ^a	9	2
	57	3 ($n = 6$)		90 ^a	9	2
	58	3 ($n = 7$)		90 ^a	9	2
	59	4		41 ^a	9	2
	60	3 ($n = 1$)		12 ^a	9	2
<i>Aspergillus niger</i> (acid unstable)	61	3 ($n = 2$)		9 ^a	9	2
	62	3 ($n = 3$)		17 ^a	9	2
	63	3 ($n = 4$)		18 ^a	9	2
	64	3 ($n = 5$)		19 ^a	9	2
	65	3 ($n = 6$)		21 ^a	9	2
	66	3 ($n = 7$)		23 ^a	9	2
	67	3 ($n = 2$)		48 ^a	9	2
<i>P. Stutzeri</i>	68	3 ($n = 6$)		43 ^a	9	2
<i>Hog pancreatic</i>	69	3 ($n = 1$)		22 ^a	9	2
	70	3 ($n = 2$)		37 ^a	9	2
	71	3 ($n = 3$)		15 ^a	9	2
	72	3 ($n = 4$)		9 ^a	9	2
	73	3 ($n = 5$)		9 ^a	9	2
	74	3 ($n = 6$)		10 ^a	9	2
	75	3 ($n = 7$)		10 ^a	9	2
β-Amylase						
<i>Sweet potato</i>	76	3 ($n = 2$)		4 ^a	9	2
	77	4		70 ^a	9	2
<i>Soybean</i>	78	3 ($n = 2$)		19 ^a	9	2
	79	4		84 ^a	9	2
<i>Malt</i>	80	3 ($n = 2$)		6 ^a	9	2
	81	4		32 ^a	9	2
<i>Barley</i>	82	3 ($n = 2$)		6 ^a	9	2
	83	4		33 ^a	9	2
<i>B. polymyxa</i>	84	3 ($n = 2$)		23 ^a	9	2
	85	4		82 ^a	9	2
<i>B. megalerium</i>	86	3 ($n = 2$)		9 ^a	9	2
	87	4		66 ^a	9	2
<i>B. circulans</i>	88	3 ($n = 2$)		8 ^a	9	2
	89	4		75 ^a	9	2
Sucrase						
	90	4		22	8	2
	91	5a		23	8	2
	92	5b		64	8	2
	93	5c		22	8	2
	94	5d		49	8	2
	95	5e		20	8	2
	96	5f		24	8	2
	97	5g		23	8	2
	98	5h		72	8	2
	99	5i		16	8	2
	100	5j		14	8	2
	101	5k		32	8	2
	102	5l		10	8	2
	103	5m		16	8	2

Table 1 (Continued)

Enzyme	Entry	Compound	K_i (μM)	IC_{50} (μM)	Ref.	Section
<i>Intestinal rat</i>	104	59	2		20	2
Amyloglucosidase						
	105	296	0.063		72	4
	106	297	0.24		72	4
	107	298	94		72	4
	108	299	160		72	4
<i>Aspergillus niger</i>	109	328	21	41	3	5
	110	334	5	5	3	5
<i>Rhizopus mold</i>	111	328	10	17	3	5
	112	334	4	5	3	5
Maltase						
	113	4		610	8	2
	114	5a		170	8	2
	115	5b		950	8	2
	116	5c		72	8	2
	117	5d		900	8	2
	118	5e		69	8	2
	119	5f		830	8	2
	120	5g		850	8	2
	121	5h		450	8	2
	122	5i		1000	8	2
	123	5j		500	8	2
	124	5k		790	8	2
	125	5l		370	8	2
	126	5m		460	8	2
<i>Yeast</i>	127	251	5.5		65	2
	128	253	24		65	2
	129	262	30		65	3
	130	293	17		65	4
	131	295	108		65	4
Endomannosidase						
<i>Rat golgi</i>	132	135		1.7	40	2
	133	136		10	40	2
<i>Rat liver</i>	134	135	5.6		40	2
	135	136	25.1		40	2
<i>Hen egg white lysozyme</i>						
	136	51		374	18	2
	137	52		16.7	18	2
	138	53		0.57	18	2
Endocellulase E₁						
<i>T. Fusca</i>	139	71	47		23	2
	140	72	63		23	2
Heparanase						
<i>Colon 26N-17 cell</i>	141	120		56–63	35	2
α-Glucosidase						
	142	296	59		72	4
	143	297	70		72	4
	144	298	28		72	4

Table 1 (Continued)

Enzyme	Entry	Compound	K_i (μM)	IC_{50} (μM)	Ref.	Section
<i>Rice</i>	145	9		0.034	11	2
	146	11		230	11	2
	147	334	107	85	3	5
<i>Rat liver</i>	148	9		25	11	2
<i>Yeast</i>	149	313		175	79	4
	150	314		340	79	4
	151	315		940	79	4
	152	316		2400	79	4
	153	317		53	79	4
	154	317		1020	79	4
	155	334		500	3	5
β-Glucosidase						
	156	296	2.3		72	4
	157	297	0.38		72	4
	158	298	150		72	4
	159	299	510		72	4
<i>Bovine liver</i>	160	293	195		65	4
	161	295	190		65	4
<i>Almonds</i>	162	264	209		65	3
	163	272	100		68	4
	164	295	375		65	4
	165	281		60	70	4
	166	284		1000	70	4
<i>C-saccharolyticum</i>	167	281		3.6	70	4
	168	284		2	70	4
	169	285	3.3		70	4
β-Glucocerebrosidase						
<i>human</i>	170	218		56	56	2
Isomaltase						
	171	296	100		72	4
	172	297	19		72	4
	173	299	190		72	4
<i>Yeast</i>	174	251	55		65	2
	175	253	36		65	2
	176	262	52		65	3
	177	293	53		65	4
	178	295	83		65	4
	179	334		500	3	5
Trehalase						
<i>Pig kidney</i>	180	518	44	88	42	5
	181	293	549		65	4
Naringinase						
<i>Penicillium decumbens</i>	182	251	702		65	2
	183	253	160		65	2
	184	262	187		65	3
	185	293	742		65	4
	186	295	443		65	4

Table 1 (Continued)

Enzyme	Entry	Compound	K_i (μM)	IC_{50} (μM)	Ref.	Section
Pullulanase						
<i>Aerobacter aerogenes</i>	187	3 ($n = 1$)		47 ^a	9	2
	188	3 ($n = 2$)		62 ^a	9	2
	189	3 ($n = 3$)		86 ^a	9	2
	190	3 ($n = 4$)		92 ^a	9	2
	191	3 ($n = 5$)		87 ^a	9	2
	192	3 ($n = 6$)		90 ^a	9	2
	193	3 ($n = 7$)		94 ^a	9	2
α-Mannosidase						
<i>Jack bean</i>	194	296	1200		72	4
	195	299	400		72	4
	196	272	2.6		68	4
	197	365		43 ^b	9	5
<i>Almonds</i>	198	365		27 ^b	9	5
β-Mannosidase						
<i>Snail</i>	199	272	120		68	4
α-Galactosidase						
<i>Coffee beans</i>	200	335	0.092	0.77	3	5
<i>Aspergillus niger</i>	201	335	2	10	3	5
<i>E. coli</i>	202	328	630	228	3	5
	203	334	200	310	3	5
	204	335	0.22	0.205	3	5
β-Galactosidase						
<i>E. coli</i>	205	285	0.1		70	4
<i>Coffee beans</i>	206	285	250		70	4
β-N-Acetylhexosaminidase						
<i>Human placenta</i>	207	452		269	24	5
<i>Bovine kidney</i>	208	452		310	24	5
<i>HL-60</i>	209	452		416	24	5
<i>Jack bean</i>	210	452		917	24	5
α-N-Acetylhexosaminidase						
<i>Chicken liver</i>	211	452		567	24	5
α-N-Acetylgalactosaminidase						
	212	335	9.8	16	3	5
Xylanase						
<i>Cellulomonas fimi</i>	213	172	0.15		45	2
	214	174	0.37		45	2
	215	179	5.8		45	2
	216	180	110		45	2
	217	181	0.13		45	2
α-1,3-Fucosyltransferase						
	218	300	2700		73	4
	219	304	81		75	4
	220	305	233		75	4

^a % of inhibition at 0.25 mM.^b % inhibition at 1 mM.

of all tested compounds against a panel of 24 glycosidases (and even one glycosyltransferase) obtained from different species.

7 Conclusions

During more than three decades, many efforts have been devoted to the synthesis of imino-disaccharides and -oligosaccharides, in particular those featuring an oxygen atom or carbon chains as interglycosidic linkers. Unfortunately, many of these complex products have not been deprotected to afford sugar mimics suitable for the biological assays. Therefore, numerous syntheses outlined in the previous sections of the review can be considered as mere exercises of synthetic organic chemistry. Nevertheless, it appears from the survey of the literature that it can be interesting to explore new synthetic approaches towards the *S*- and *N*-linked iminodisaccharides, compounds that can be endowed with useful biological activity. It is worth noting that if pharmacological applications are envisaged, the new imino-disaccharides and -oligosaccharides should be obtained by means of metal-free reactions to avoid the contamination of the active principle by residual metal catalysts, *e.g.* copper or palladium. The use of these sugar mimics as actual drugs should also require the presence of enzymatically and hydrolytically stable interglycosidic bonds, as those found in the *S*- and *C*-glycoside derivatives, to allow the desired activity in the biological fluids.

Regretfully, due to the variety of the glycosidases (and their sources) employed in the inhibition assays, as well as the limited number of tested compounds, it is quite difficult to evaluate the structure-activity relationship of the imino-disaccharides and -oligosaccharides. This finding is in sharp contrast with the glycosidase inhibition studies performed with multivalent iminosugars,³ where a specific enzyme, the Jack bean α -mannosidase, has been largely adopted as model glycosidase.

References

- (a) B. G. Winchester, *Tetrahedron: Asymmetry*, 2009, **20**, 645–651; (b) P. Compain, V. Chagnault and O. R. Martin, *Tetrahedron: Asymmetry*, 2009, **20**, 672–711; (c) D. D'Alonzo, A. Guaragna and G. Palumbo, *Curr. Med. Chem.*, 2009, **16**, 473–505; (d) G. Horne, F. X. Wilson, J. Tinsley, D. H. Williams and R. Storer, *Drug Discovery Today*, 2011, **16**, 107–118; (e) P. Compain, *Synlett*, 2014, 1215–1240.
- IUPAC Nomenclature of Carbohydrates, Recommendations 1996, rule 2-Carb-34.1.
- (a) P. Compain and A. Bodlenner, *ChemBioChem*, 2014, **15**, 1239–1251; (b) S. G. Gouin, *Chem. Eur. J.*, 2014, **20**, 11616–11628; (c) R. Zelli, J.-F. Longevial, P. Dumy and A. Marra, *New J. Chem.*, 2015, **39**, 5050–5074; (d) N. Kanfar, E. Bartolami, R. Zelli, A. Marra, J.-Y. Winum, S. Ulrich and P. Dumy, *Org. Biomol. Chem.*, 2015, **13**, 9894–9906; (e) C. Matassini, C. Parmeggiani, F. Cardona and A. Goti, *Tetrahedron Lett.*, 2016, **57**, 5407–5415.
- (a) I. Robina and P. Vogel, *Synthesis*, 2005, 675–702; (b) “Imino-*C*-disaccharides and analogues: synthesis and biological activity” P. Vogel, S. Gerber-Lemaire and L. Juillerat-Jeanneret in *Iminosugars: From Synthesis*

- to *Therapeutic Applications*, ed. P. Compain and O. R. Martin, Wiley, New York, 2007, pp. 87–130.
- 5 P. Merino, I. Delso, E. Marca, T. Tejero and R. Matute, *Curr. Chem. Biol.*, 2009, **3**, 253–271.
 - 6 Y. Ezure, *Agric. Biol. Chem.*, 1985, **49**, 2159–2165.
 - 7 Y. Ezure, Y. Yoshikuni, N. Ojima and M. Sugiyama, *Acta Crystallogr.*, 1987, **C43**, 1809–1811.
 - 8 Y. Yoshikuni, Y. Ezure, T. Seto, K. Mori, M. Watanabe and H. Enomoto, *Chem. Pharm. Bull.*, 1989, **37**, 106–109.
 - 9 M. Arai, M. Sumida, K. Fukuhara, M. Kainisho and S. Murao, *Agric. Biol. Chem.*, 1986, **50**, 639–644.
 - 10 M. Kojima, T. Seto, Y. Kyotani, H. Ogawa, S. Kitazawa, K. Mori, S. Maruo, T. Ohgi and Y. Ezure, *Biosci., Biotechnol., Biochem.*, 1996, **60**, 694–696.
 - 11 N. Asano, K. Oseki, E. Kaneko and K. Matsui, *Carbohydr. Res.*, 1994, **258**, 255–266.
 - 12 N. S. Paek, D. J. Kang, H. S. Lee, J. J. Lee, Y. J. Choi, T. H. Kim and K. W. Kim, *Biosci., Biotechnol., Biochem.*, 1998, **62**, 588–589.
 - 13 R. Uchida, A. Nasu, S. Tokutake, K. Kasai, K. Tobe and N. Yamaji, *Chem. Pharm. Bull.*, 1999, **47**, 187–193.
 - 14 R. Bernotas, G. Papandreou, J. Urbach and B. Ganem, *Tetrahedron Lett.*, 1990, **31**, 3393–3396.
 - 15 S. Numao, I. Damager, C. Li, T. M. Wrodnigg, A. Begum, C. M. Overall, G. D. Brayer and S. G. Withers, *J. Biol. Chem.*, 2004, **279**, 48282–48291.
 - 16 I. Damager, S. Numao, H. Chen, G. D. Brayer and S. G. Withers, *Carbohydr. Res.*, 2004, **339**, 1727–1737.
 - 17 R. Hoos, A. B. Naughton, W. Thiel, A. Vasella, W. Weber, K. Rupitz and S. G. Withers, *Helv. Chim. Acta*, 1993, **76**, 2666–2686.
 - 18 M. Ogata, N. Umemoto, T. Ohnuma, T. Numata, A. Suzuki, T. Usui and T. Fukamizo, *J. Biol. Chem.*, 2013, **288**, 6072–6082.
 - 19 C. Gautheron-Le Narvor and C.-H. Wong, *J. Chem. Soc., Chem. Commun.*, 1991, 1130–1131.
 - 20 P. S. Liu, *J. Org. Chem.*, 1987, **52**, 4717–4721.
 - 21 P. B. Anzeveno, L. J. Creemer, J. K. Daniel, C.-H. R. King and P. S. Liu, *J. Org. Chem.*, 1989, **54**, 2539–2542.
 - 22 (a) S. Inouye, T. Tsuruoka, T. Ito and T. Niida, *Tetrahedron*, 1968, **23**, 2125–2144; (b) H. Iida, N. Yamazaki and C. Kibayashi, *J. Org. Chem.*, 1987, **52**, 3337–3342.
 - 23 L. J. Liotta, R. C. Bernotas, D. B. Wilson and B. Ganem, *J. Am. Chem. Soc.*, 1989, **111**, 783–785.
 - 24 R. C. Bernotas and B. Ganem, *Tetrahedron Lett.*, 1985, **26**, 1123–1126.
 - 25 (a) S. F. Moss and S. L. Vallance, *J. Chem. Soc. Perkin Trans.*, 1992, **1**, 1959–1967; (b) S. F. Moss and R. Southgate, *J. Chem. Soc. Perkin Trans.*, 1993, **1**, 1787–1794.
 - 26 M. Kiso, H. Katagiri, H. Furui and A. Hasegawa, *J. Carbohydr. Chem.*, 1992, **11**, 627–644.
 - 27 S. Takahashi, H. Terayama and H. Kuzuhara, *Tetrahedron*, 1996, **42**, 13315–13326.
 - 28 H. Terayama, S. Takahashi and H. Kuzuhara, *J. Carbohydr. Chem.*, 1993, **12**, 81–93.
 - 29 H. Furui, M. Kiso and A. Hasegawa, *Carbohydr. Res.*, 1992, **229**, C1–C4.
 - 30 (a) M. Kiso, H. Furui, K. Ando and A. Hasegawa, *J. Carbohydr. Chem.*, 1993, **12**, 673–677; (b) M. Kiso, H. Furui, K. Ando, H. Ishida and A. Hasegawa, *Bioorg. Med. Chem.*, 1994, **2**, 1295–1308.

- 31 A. Kondo, K. Ando, H. Ishida, I. Kato, A. Hasegawa and M. Kiso, *J. Carbohydr. Chem.*, 1994, **13**, 545–554.
- 32 M. Kiso, H. Furui, H. Ishida and A. Hasegawa, *J. Carbohydr. Chem.*, 1996, **15**, 1–14.
- 33 H. Furui, K. Ando-Furui, H. Inagaki, T. Ando, H. Ishida and M. Kiso, *J. Carbohydr. Chem.*, 2001, **20**, 789–812.
- 34 H. Ogawa, Y. Harada, Y. Kyotani, T. Ueda, S. Kitazawa, K. Kandori, T. Seto, K. Ishiyama, M. Kojima, T. Ohgi, Y. Ezure and M. Kise, *J. Carbohydr. Chem.*, 1998, **17**, 729–738.
- 35 (a) S. Takahashi and H. Kuzuhara, *Chem. Lett.*, 1994, 2119–2122;
(b) S. Takahashi, H. Kuzuhara and M. Nakajima, *Tetrahedron*, 2001, **57**, 6915–6926.
- 36 M. Kiso, H. Katagiri, H. Furui, K. Ando, H. Ishida and A. Hasegawa, *J. Carbohydr. Chem.*, 1994, **13**, 163–174.
- 37 M. Kiso, K. Ando, H. Furui, H. Ishida and A. Hasegawa, *J. Carbohydr. Chem.*, 1993, **12**, 985–990.
- 38 H. Ardron, T. D. Butters, F. M. Platt, M. R. Wormald, R. A. Dwek, G. W. Fleet and G. S. Jacob, *Tetrahedron: Asymmetry*, 1993, **4**, 2011–2024.
- 39 G. W. Fleet, N. G. Ramsden and D. R. Witty, *Tetrahedron*, 1989, **45**, 327–336.
- 40 S. Hiraizumi, U. Spohr and R. G. Spiro, *J. Biol. Chem.*, 1993, **268**, 9927–9935.
- 41 U. Spohr, M. Bach and R. G. Spiro, *Can. J. Chem.*, 1993, **71**, 1928–1942.
- 42 U. Spohr and M. Bach, *Can. J. Chem.*, 1993, **71**, 1943–1954.
- 43 S. Vonhoff, K. Piens, M. Pipelier, C. Braet, M. Claeysens and A. Vasella, *Helv. Chim. Acta*, 1999, **82**, 963–980.
- 44 S. Vonhoff and A. Vasella, *Synth. Commun.*, 1999, **29**, 551–560.
- 45 S. J. Williams, R. Hoos and S. G. Withers, *J. Am. Chem. Soc.*, 2000, **122**, 2223–2235.
- 46 S. J. Williams, V. Notenboom, J. Wicki, D. R. Rose and S. G. Withers, *J. Am. Chem. Soc.*, 2000, **122**, 4229–4230.
- 47 F. D'Andrea, G. Catelani, M. Mariani and B. Vecchi, *Tetrahedron Lett.*, 2001, **42**, 1139–1142.
- 48 P. L. Barili, G. Catelani, F. D'Andrea, F. De Rensis and P. Falcini, *Carbohydr. Res.*, 1997, **298**, 75–84.
- 49 G. Catelani, F. D'Andrea and L. Puccioni, *Carbohydr. Res.*, 2000, **324**, 204–209.
- 50 L. Guazzelli, G. Catelani, F. D'Andrea, T. Gagnani and A. Griselli, *Eur. J. Org. Chem.*, 2014, 6527–6537.
- 51 P. L. Barili, G. Berti, G. Catelani, F. D'Andrea and L. Puccioni, *J. Carbohydr. Chem.*, 2000, **19**, 79–81.
- 52 E. Attolino, G. Catelani, F. D'Andrea and M. Nicolardi, *J. Carbohydr. Chem.*, 2004, **23**, 179–190.
- 53 A. J. Steiner and A. E. Stütz, *Carbohydr. Res.*, 2004, **339**, 2615–2619.
- 54 J. Spreitz and A. E. Stütz, *Carbohydr. Res.*, 2004, **339**, 1823–1827.
- 55 R. Blattner, R. H. Furneaux and Z. Pakulski, *Carbohydr. Res.*, 2006, **341**, 2115–2125.
- 56 C. Boucheron, S. Toumieux, P. Compain, O. Martin, K. Ikeda and N. Asano, *Carbohydr. Res.*, 2007, **342**, 1960–1965.
- 57 Z. Csiki and P. Fügedi, *Tetrahedron*, 2010, **66**, 7821–7837.
- 58 Z. Csiki and P. Fügedi, *Tetrahedron Lett.*, 2010, **51**, 391–395.
- 59 D. Sawada, H. Takahashi, M. Shiro and S. Ikegami, *Tetrahedron Lett.*, 2005, **46**, 2399–2403.
- 60 T. Granier and A. Vasella, *Helv. Chim. Acta*, 1998, **81**, 865–880.
- 61 J. Fuentes, N. R. Al Bujuq, M. Angulo and C. Gasch, *Tetrahedron Lett.*, 2008, **49**, 910–913.

- 62 J. Fuentes Mota, D. Mostowicz, C. Ortiz, M. A. Pradera and I. Robina, *Carbohydr. Res.*, 1994, **257**, 305–316.
- 63 J. Fuentes, J. M. Illangua, F. J. Sayago, M. Angulo, C. Gasch and M. A. Pradera, *Tetrahedron: Asymmetry*, 2004, **15**, 3783–3789.
- 64 A. Marra, J. Esnault, A. Veyrières and P. Sinaÿ, *J. Am. Chem. Soc.*, 1992, **114**, 6354–6360.
- 65 E. M. Sanchez-Fernandez, R. Risquez-Cuadro, C. Ortiz Mellet, J. M. Garcia Fernandez, P. M. Nieto and J. Angulo, *Chem. – Eur. J.*, 2012, **18**, 8527–8539.
- 66 (a) K. Suzuki and H. Hashimoto, *Tetrahedron Lett.*, 1994, **35**, 4119–4122; (b) K. Suzuki and H. Hashimoto, *Carbohydr. Res.*, 2000, **323**, 14–27.
- 67 S. Knapp, Y. H. Choe and E. Reilly, *Tetrahedron Lett.*, 1993, **34**, 4443–4446.
- 68 Y. Blériot, T. Dintinger, N. Guillo and C. Tellier, *Tetrahedron Lett.*, 1995, **36**, 5175–5178.
- 69 K. Suzuki, T. Fujii, K. Sato and H. Hashimoto, *Tetrahedron Lett.*, 1996, **37**, 5921–5924.
- 70 S. Vonhoff, T. D. Heightman and A. Vasella, *Helv. Chim. Acta*, 1998, **81**, 1710–1725.
- 71 E. M. Sanchez-Fernandez, R. Risquez-Cuadro, M. Aguilar-Moncayo, M. I. Garcia-Moreno, C. Ortiz Mellet and J. M. Garcia Fernandez, *Org. Lett.*, 2009, **11**, 3306–3309.
- 72 W. Dong, T. Jespersen, M. Bols, T. Skrydstrup and M. R. Sierks, *Biochemistry*, 1996, **35**, 2788–2795.
- 73 L. Qiao, B. W. Murray, M. Shimazaki, J. Schultz and C.-H. Wong, *J. Am. Chem. Soc.*, 1996, **118**, 7653–7662.
- 74 R. Wischnat, R. Martin and C.-H. Wong, *J. Org. Chem.*, 1998, **63**, 8361–8365.
- 75 I. Jefferies and B. R. Bowen, *Bioorg. Med. Chem. Lett.*, 1997, **7**, 1171–1174.
- 76 L. Sun, P. Li, N. Amankulor, W. Tang, D. W. Landry and K. Zhao, *J. Org. Chem.*, 1998, **63**, 6472–6475.
- 77 B. Ruttens and J. Van der Eycken, *Tetrahedron Lett.*, 2002, **43**, 2215–2221.
- 78 V. Pistrà, A. Rescifina, F. Punzo, G. Greco, V. Barbera and A. Corsaro, *Eur. J. Org. Chem.*, 2011, 7278–7287.
- 79 L. O. B. Zamoner, V. Aragao-Leoneti, S. P. Mantoani, M. D. Rugen, S. A. Nepogodiev, R. A. Field and I. Carvalho, *Carbohydr. Res.*, 2016, **429**, 29–37.
- 80 C. R. Johnson, M. W. Miller, A. Golebiowski, H. Sundram and M. B. Ksebati, *Tetrahedron Lett.*, 1994, **48**, 8991–8994.
- 81 (a) J. Rouden and T. Hudlicky, *J. Chem. Soc., Perkin Trans.*, 1993, **1**, 1095–1097; (b) T. Hudlicky, J. Rouden, H. Luna and S. Allen, *J. Am. Chem. Soc.*, 1994, **116**, 5099–5107.
- 82 B. A. Johns, Y. T. Pan, A. D. Elbein and C. R. Johnson, *J. Am. Chem. Soc.*, 1997, **119**, 4856–4865.
- 83 M. A. Leeuwengurgh, S. Picasso, H. S. Overkleeft, G. A. van der Marel, P. Vogel and J. H. van Boom, *Eur. J. Org. Chem.*, 1999, 1185–1189.
- 84 C.-H. Wong, L. Provencher, J. A. Porco Jr., S.-H. Jung, Y.-F. Wang, L. Chen, R. Wang and D. H. Steensma, *J. Org. Chem.*, 1995, **60**, 1492–1501.
- 85 O. M. Saavedra and O. R. Martin, *J. Org. Chem.*, 1996, **61**, 6987–6993.
- 86 O. R. Martin, L. Liu and F. Yang, *Tetrahedron Lett.*, 1996, **37**, 1991–1994.
- 87 A. Baudat and P. Vogel, *Tetrahedron Lett.*, 1996, **37**, 483–484.
- 88 A. Baudat and P. Vogel, *J. Org. Chem.*, 1997, **62**, 6252–6260.
- 89 K. Kraehenbuehl, S. Picasso and P. Vogel, *Bioorg. Med. Chem. Lett.*, 1997, **7**, 893–896.
- 90 K. Kraehenbuehl, S. Picasso and P. Vogel, *Helv. Chim. Acta*, 1998, **81**, 1439–1451.
- 91 Y.-H. Zhu and P. Vogel, *J. Org. Chem.*, 1999, **64**, 666–669.
- 92 F. Cardona, I. Robina and P. Vogel, *J. Carbohydr. Chem.*, 2000, **19**, 555–571.

- 93 L. Campanini, A. Duréault and J.-C. Depezay, *Tetrahedron Lett.*, 1996, **29**, 5095–5098.
- 94 P. A. Risbood, T. S. Philips and L. Goodman, *Carbohydr. Res.*, 1981, **94**, 101–107.
- 95 R. L. Whistler and J. N. BeMiller, *Methods Carbohydr. Chem.*, 1972, **VI**, 220–222.
- 96 F. J. Duff, V. Vivien and R. H. Wightman, *Chem. Commun.*, 2000, 2127–2128.
- 97 N. G. Argyropoulos, E. Coutouli-Argyropoulos and P. Gkizis, *ARKIVOC*, 2008, **16**, 223–234.
- 98 V. Liautard, V. Desvergnés and O. R. Martin, *Tetrahedron: Asymmetry*, 2008, **19**, 1999–2002.
- 99 N. G. Argyropoulos and V. C. Sarli, *Tetrahedron Lett.*, 2004, **45**, 4237–4240.
- 100 E. Marca, I. Delso, T. Tejero and P. Merino, *Tetrahedron*, 2012, **68**, 6674–6687.
- 101 (a) A. Dondoni, P. P. Giovannini and A. Marra, *Tetrahedron Lett.*, 2000, **41**, 6195–6199; (b) A. Dondoni, P. P. Giovannini and D. Perrone, *J. Org. Chem.*, 2002, **67**, 7203–7214.
- 102 G. V. M. Sharma, N. Pendem, K. R. Reddy, P. R. Krishna, K. Narsimulu and A. C. Kunwar, *Tetrahedron Lett.*, 2004, **45**, 8807–8810.
- 103 G. V. M. Sharma, V. Goverdhan Reddy, A. Subhash Chander and K. Ravinder Reddy, *Tetrahedron: Asymmetry*, 2002, **13**, 21–24.
- 104 Y. Blériot, N. Auberger, Y. Jagadeesh, C. Gauthier, G. Prencipe, A. T. Tran, J. Marrot, J. Désiré, A. Yamamoto, A. Kato and M. Sollogoub, *Org. Lett.*, 2014, **16**, 5512–5515.
- 105 A. L. Haines, *Org. Biomol. Chem.*, 2004, **2**, 2352–2358.
- 106 L. Werner, L. Kniezo and H. Dvorakova, *Tetrahedron Lett.*, 2007, **48**, 609–612.
- 107 E. Marca, I. Delso, T. Tejero, J. T. Vazquez, R. L. Dorta and P. Merino, *Tetrahedron Lett.*, 2009, **50**, 7152–7155.
- 108 P. Zhang, C. Li, H. Yang, H. Wei, Z. Xia, D. Ma, H. Chen, K. Wang and X. Li, *Carbohydr. Res.*, 2014, **398**, 36–39.
- 109 P. Z. Zhang, X. L. Li, H. Chen, Y. N. Li and R. Wang, *Tetrahedron Lett.*, 2007, **48**, 7813–7816.
- 110 P. Zhang, C. Wei, A. Wang, W. Wang, M. Liu, Q. Yin, H. Chen, K. Wang, X. Li and J. Zhang, *Carbohydr. Res.*, 2012, **351**, 7–16.
- 111 E. Frérot, C. Marquis and P. Vogel, *Tetrahedron Lett.*, 1996, **37**, 2023–2026.
- 112 S. Allemann and P. Vogel, *Tetrahedron*, 1994, **50**, 2469–2478.
- 113 R. Mampuya Bimwala and P. Vogel, *J. Org. Chem.*, 1992, **57**, 2076–2083.
- 114 Y.-H. Zhu and P. Vogel, *Chem. Commun.*, 1999, 1873–1874.
- 115 I. Navarro and P. Vogel, *Helv. Chim. Acta*, 2002, **85**, 152–160.
- 116 E. R. Garcia, M. A. Brimble and P. Vogel, *Eur. J. Org. Chem.*, 2006, 3845–3855.
- 117 (a) A. Klemer and G. Rodemeyer, *Chem. Ber.*, 1974, **107**, 2612–2614; (b) D. Horton and W. Weckerle, *Carbohydr. Res.*, 1975, **44**, 227–240; (c) D. N. Jones, M. G. Taylor and W. W. Wood, *Synth. Commun.*, 1992, **22**, 1687–1690.
- 118 F. Cardona, S. Valenza, A. Goti and A. Brandi, *Tetrahedron Lett.*, 1997, **38**, 8097–8100.
- 119 F. Cardona, S. Valenza, S. Picasso, A. Goti and A. Brandi, *J. Org. Chem.*, 1998, **63**, 7311–7318.
- 120 F. Cardona, A. Goti and A. Brandi, *Org. Lett.*, 2003, **5**, 1475–1478.
- 121 A. Kennedy, A. Nelson and A. Perry, *Chem. Commun.*, 2005, 1646–1648.
- 122 D. Bini, M. Forcella, L. Cipolla, P. Fusi, C. Matassini and F. Cardona, *Eur. J. Org. Chem.*, 2011, 3995–4000.
- 123 L. Cipolla, M. Reis Fernandes, M. Gregori, C. Airoidi and F. Nicotra, *Carbohydr. Res.*, 2007, **342**, 1813–1830.

Bacterial polysaccharides as major surface antigens: interest in O-acetyl substitutions

Laurence A. Mulard^{a,b}

DOI: 10.1039/9781788010641-00071

Polysaccharides represent essential, although highly structurally diverse, components on microbial cell surfaces. They are the primary interface with the host and play critical roles in survival strategies. Acting as shields against environmental assaults, they are actively investigated as attractive vaccine components. Contributing to a tremendous structural diversity, a subtle but nonetheless essential microbial polysaccharide modification is O-acetylation. Focusing on bacterial capsular polysaccharides (CPS), this chapter provides some highlights on this widespread substitution. CPS O-acetylation is discussed first from a genetic and biochemical perspective, then in view of its implication in the host-pathogen crosstalk and ability to modulate CPS biological properties in a context-dependent manner. Lastly, the chapter addresses CPS O-acetylation in the context of antibacterial vaccine development.

1 Introduction

Polysaccharides represent prime components on bacterial cell surfaces. Produced by both pathogenic and non-pathogenic bacteria, they are involved in the bacterium cross talk with its environment, and often play critical roles in host-bacterium interactions. Occurring in the form of capsular polysaccharides (CPS), lipooligosaccharides (LOS), or lipopolysaccharides (LPS), they are important virulence factors contributing to, among other processes, surface charge, phase variation, resistance to serum-mediated killing, and more generally modulation of the host immune response.^{1–4} Whereas CPS may be present in both Gram-positive and Gram-negative bacteria, LPS is restricted to the outer membrane of the latter. LPS consists of three structural parts: the lipid A that serves as an anchor into the membrane, a core oligosaccharide (OS), and an O-specific polysaccharide (O-SP), which is the most surface-exposed and structurally diverse constituent.^{5,6} CPS and O-SP may be homopolymers, as exemplified by the high molecular weight negatively charged capsule shared by *Escherichia coli* K1, *Neisseria meningitidis* serogroup B (MenB), *Mannheimia haemolytica* and *Moraxella nonliquefaciens* (Fig. 1A),⁷ or by the neutral O-SP of *Vibrio cholerae* O1 serotype Inaba,⁸ respectively (Fig. 1B).

However, bacterial heteropolysaccharides, defined either by a linear or a branched repeating unit, are more widespread than homopolysaccharides. Their repeating units vary from di- to octasaccharides, with up to three side chains made of one to four residues. These polysaccharide fingerprints are built up from a wide repertoire of unique monosaccharide components, the number of which is unknown. Still, a

^aInstitut Pasteur, Unité de Chimie des Biomolécules, 28 rue du Dr Roux, 75 724, Paris Cedex 15, France. E-mail: laurence.mulard@pasteur.fr

^bCNRS UMR 3523, Institut Pasteur, 28 rue du Dr Roux, 75 724, Paris Cedex 15, France

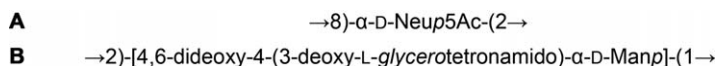


Fig. 1 (A) Basic repeating unit of the CPS common to *E. coli* K1, MenB, *M. haemolytica* and *M. nonliquefaciens*.⁷ In the case of *E. coli* K1, acetylation at O-7 and O-9 is observed.⁹ (B) Repeating unit of the O-SP from *V. cholerae* O1 Inaba.⁸

systematic database analysis of the bacterial glycome indicated that diversity at the monosaccharide level was more than ten-fold greater than that of the human glycome.¹⁰ Besides, the regiochemistry and α/β stereochemistry of the glycosidic bonds bring in increased complexity.^{10,11} Likewise, monosaccharide substitutions with non-sugar constituents, among which acetylation, methylation, and phosphorylation are frequently identified, thus contributing to additional diversity.^{10,11} The tremendous structural variety of CPS and O-SP gives rise to a high degree of antigenic heterogeneity whether between or within bacterial species; a property advantageously exploited for serotyping. For example, more than 80 CPS and 180 O-SP have been proposed for *E. coli*.¹² Of interest despite remaining difficulties in obtaining homogeneous harvested materials, progress in analytical methods has paved the way to a more clear-cut elucidation of complex carbohydrate structures with an enhanced interest for post-assembly modifications.^{13,14} Accordingly, repeating units from an increasing number of bacterial polysaccharides are being identified, or revised.^{12,15–18}

Herein, patterns of O-acetylation occurring on bacterial polysaccharides, whether CPSs or O-SPs, are exemplified. However, the main part of the chapter deals with O-acetylated CPSs. Thus, CPS O-acetylation is discussed first from a genetic perspective, then in view of its implication in the host-pathogen crosstalk. Increasing developments in the field have contributed to a renewal in CPS structural analysis, some of which will be highlighted. Owing to the key role of CPS in the field, the last part of the chapter addresses CPS O-acetylation in the context of antibacterial vaccine development.

2 PS O-acetylation: a widespread modification

2.1 Bacterial PS O-acetylation as a source of tremendous structural diversity

Whether stoichiometric or non-stoichiometric, O-acetylation of bacterial surface polysaccharides is frequent.¹⁰ Obviously, it has long been known as a common post glycosylation modification of major biological importance. It was essentially addressed in relation to sialic acid diversification, thereby underlining its implication in bacterial virulence and disease pathogenesis as well as its ability to alter the host innate and adaptive immunity, and ultimately contribute to bacterial escape.^{13,14,19} Many factors, such as the site of O-acetylation or the phase variation O-acetylation profile as in the polysialic acid K1 capsule of *E. coli*, contribute to this aptitude.^{9,20}

As evidenced, structural variation associated to bacterial polysaccharide non-stoichiometric O-acetylation is almost infinite.²⁴ For example,

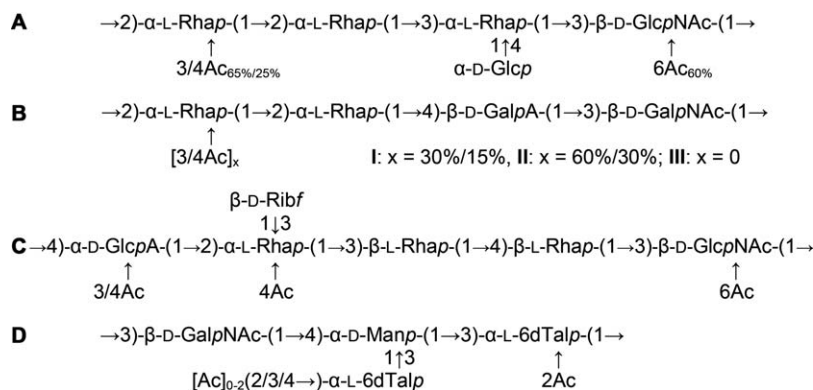


Fig. 2 Repeating units of the O-SPs from (A) *Shigella flexneri* serotype 2a,¹⁸ (B) *S. flexneri* serotype 6 (I), 6a (II),¹⁸ and *E. coli* serogroup O147 (III),²¹ (C) *Shigella boydii* type 11,²² (D) *Aeromonas hydrophila* serotype O:34.²³

data extracted from various *E. coli* K1 and Group B *Streptococcus* (GBS) isolates indicate that the degree of CPS O-acetylation may vary from 5% to 95%,²⁴ and 5% to 55%,²⁵ respectively. In contrast, the extent of O-acetylation lies in a narrower range in the case of the O-SP harvested from various strains of *S. flexneri* 2a (Fig. 2A).^{18,26} In some instances, as recently proposed for *S. flexneri* 6 (I) and 6a (II), the degree of O-acetylation differentiates subtypes (Fig. 2B).¹⁸ It is of note that in this last example, the corresponding non O-acetylated repeating unit defines the O-SP from the enterotoxigenic *E. coli* O147 (Fig. 2B, III).²¹

As an additional source of diversity, modifications may not be distributed evenly within the chain. For example, O-acetylation in the caryan moiety from *Pseudomonas* (*Burkholderia*) *caryophylli* LPS leads to a block pattern.²⁷ Moreover, as illustrated in the extreme with the O-SP from *S. boydii* type 11 (Fig. 2C),²² repeating units O-acetylated at multiple sites are frequently encountered in bacterial polysaccharides. As detected by NMR analysis in the case of *S. flexneri* 2a O-SP, the pentasaccharide repeating unit of which is O-acetylated in a non-stoichiometric manner at two residues (Fig. 2A), all possible combinations of O-acetylation may occur along the chain.²⁶ This is without counting the non-enzymatic migration of acetyl groups to vicinal,²⁸ and even non vicinal hydroxyl groups,²⁴ which also adds to structural diversity. Although less common, there is evidence for random multiple O-acetylation of a single residue within repeating units as on the branched 6-deoxy-L-talose residue from the O-SP of *A. hydrophila* O:34 (Fig. 2D).²³ This additional source of bacterial polysaccharide structural diversity cannot be left aside.

2.2 The genetic basis of bacterial CPS O-acetylation: highlights

The need for a better understanding of the biological significance of surface polysaccharide O-acetylation in bacteria has promoted major interest in the genetic basis for this important non-carbohydrate modification. Two principal families of proteins involved in the O-acetylation of exported polysaccharides have been identified: cytoplasmic proteins that use acetyl-CoA on the one hand and integral membrane proteins

on the other hand. Nevertheless, diversity is enormous. As seen with MynC, the O-3 and O-4 ManNAc transferase required for *N. meningitidis* serogroup A (MenA) CPS O-acetylation, several of the known O-acetyltransferases do not fall in these main categories.²⁹

2.2.1 O-Acetylation of fully assembled CPSs. Interestingly, in the search for an efficient *in vitro* production of MenA CPS, the molecular cloning, recombinant expression, functional characterization, and combination of the three key enzymes taking part in the biosynthesis of MenA CSP provided a clear demonstration that O-acetylation is a post-assembly CPS modification.³⁰ In particular, attempts at the *in vitro* synthesis of the MenA polysaccharide from the 3-O-acetyl-ManpNAc-UDP donor failed. The MenA poly-ManpNAc-1-phosphate transferase was undoubtedly shown to prefer non-O-acetylated over O-acetylated primers, while enzymatic O-acetylation of the resulting ManpNAc-1-phosphate polymer provided a polysaccharide identical to the natural MenA CPS.³⁰

One of the most studied bacterial O-acetyltransferases is NeuO, the prophage-encoded protein controlling the phase-variable CPS O-acetylation in *E. coli* K1.²⁰ Following extensive biochemical characterization,³⁴ the protein three-dimensional structure was solved, shedding light into the O-acetylation mechanism. NeuO, which uses acetyl-CoA as donor substrate, belongs to the left-handed β -helix (L β H) family of acetyltransferases.³⁵ *In vitro*, it catalyzes acetyl transfer to O-7 and O-9 of sialic acid within oligomers comprising at least 14 residues. This observation provides strong evidence for a co- or post-synthetic process *in vivo*.³⁵ NeuO is closely related to OatWY, the O-acetyltransferase shared by MenY and MenW.³⁶ The corresponding CPSs are heteropolymers, the repeating units of which have a [$\rightarrow$ 4]- α -D-Neup5Ac-(2 $\rightarrow$] residue in common (Fig. 3C and D). O-Acetylation at position 7 or 9 of the sialic acid residue was demonstrated in both cases.¹⁵ The structure of OatWY in complex with its donor substrate acetyl-CoA was solved. It paved the way to the first proposed mechanism for acetyl transfer to CPS.³⁷ Structural data revealed that the enzyme was also a member of the L β H family, and uncovered key features for the enzyme to accommodate large negatively

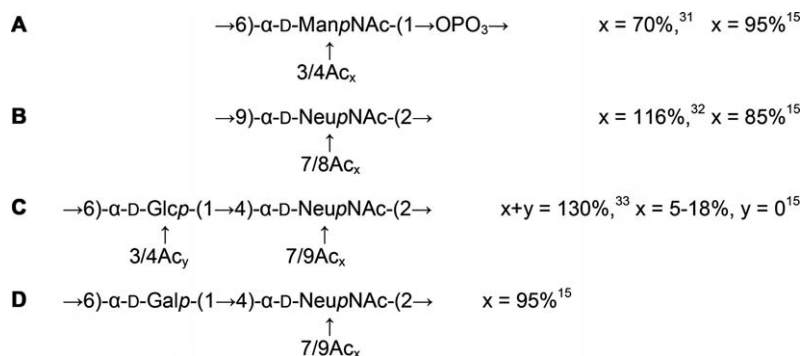


Fig. 3 Repeating units of the CPS from (A) MenA,^{15,31} (B) *N. meningitidis* serogroup C (MenC),^{15,32} (C) *N. meningitidis* serogroup Y (MenY),^{15,33} and (D) *N. meningitidis* serogroup W135 (MenW).¹⁵

charged acceptor substrates. They also provided insights on the origin of OatWY acceptor promiscuity with regards to MenY and MenW for an enzyme otherwise qualified of being highly specific.³⁷

In contrast, no homology was found for OatC, the CPS O-acetyltransferase of MenC.³⁶ Subsequent biochemical investigations demonstrated that the enzyme was highly specific for $[\rightarrow 9)\text{-}\alpha\text{-D-Neup5Ac-}(2\rightarrow)]$ oligomers and polymers, whereas neither Neup5Ac nor CMP-Neup5Ac were acceptor substrates. This indicated that *in vivo* CPS O-acetylation occurs at the polymer level. It supports a post-synthetic process for OatC, as shown for OatWY.³⁸ In spite of this similarity, OatC is strikingly distinct from OatWY. It was proposed that the former adopts an α/β -hydrolase fold structure and that CPS O-acetylation might occur selectively at O-8 of sialic acid by a ping-pong mechanism involving acetyl-CoA as donor substrate.³⁸

Phase-variable expression of the enzymes responsible for CPS O-acetylation was detected in MenY and MenW, albeit not in MenC.³⁶ In the latter case, slipped-strand mispairing was identified as the cause of phase variable CPS O-acetylation.³⁶ Instead, O-acetylation status in MenW and MenY CPSs was shown to correlate with clonal lineages.³⁶

2.2.2 O-Acetylation of CPS monosaccharide precursors. The O-acetylation of sialylated CPS is highly regulated. However, regulation processes differ among bacteria as exemplified in GBS, the only Gram positive bacterium reported to produce a sialic acid containing capsule. Clinical isolates of GBS elaborate nine different CPSs, responsible for their classification into types. Despite significant antigenic diversity, the $\alpha\text{-D-NeupNAc-}(2\rightarrow 3)\text{-}\beta\text{-D-Galp-}(1\rightarrow$ motif – whereby the sialyl residue is always present in the form of a side chain – is strictly conserved among all known GBS CPS repeating units (Fig. 4).³⁹ This shared element is thought to be central to the antiphagocytic properties of GBS CPSs and critical to GBS survival. The side chain sialyl residue common to all CPS multicomponent repeating units may be O-acetylated in varying degrees (Fig. 4).⁴⁰ In contrast to O-acetylation of the fully assembled polymer in meningococci, GBS employs only an intracellular O-acetylation mechanism. The extent of CPS O-acetylation, preferentially taking place at O-7 of the exocyclic sialyl chain followed by a unidirectional migration to O-9, is controlled at the monosaccharide stage prior to sialyl transfer to the polymer.⁴⁰ Fine tuning involves the GBS O-acetyltransferase activity of NeuD on the one hand, and the sialyl O-acetyltransferase activity of NeuA on the other hand. Both enzymes participate in a cyclic O-acetylation/de-O-acetylation process in addition to taking part in the biosynthesis of sialic acid and CMP-Neup5Ac, respectively.²⁵ Moreover, a single nucleotide polymorphism in *neuD* is associated with varied O-acetylation levels in GBS strains, and

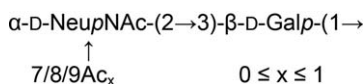


Fig. 4 Conserved motif among all known repeating units from GBS CPS types and sites of O-acetylation on the exocyclic side-chain of NeupNAc.^{39,40}

contributes to distinguish between high (45–55%) and low (2–15%) O-acetylated CPSs.⁴¹

Following up with former conclusions,²⁰ and despite low sequence identity with GBS NeuD, the O-acetyltransferase function of the homologous *E. coli* K1 NeuD was demonstrated in GBS.⁴¹ This novel finding suggested the occurrence of two separate pathways for O-acetylation of polymeric and monomeric sialic acids in *E. coli* K1.⁴² Furthermore, using a bioinformatic approach, *neuD* homologs that are physically associated with sialic acid biosynthetic gene clusters were found in the genomes of 18 bacterial species, some of which were not known to express sialic acid.⁴¹ Phylogenetic analysis revealed that members of the NeuD O-acetyltransferase family have a common evolutionary lineage, which is distinct from that of the *E. coli* K1 NeuO and from that of the meningococci OatWY, both known to act on polysialic acid.⁴¹

3 On the role of CPS O-acetylation on the host-pathogen crosstalk

3.1 *E. coli* K1: O-acetylation as a source of phase variation and unique properties

E. coli K1 is an intestinal commensal of mammals and birds, and possibly a source of disease for its host. Above all, it is a common cause of sepsis and meningitis in neonates. Early estimates indicated that K1-encapsulated *E. coli* accounts for approximately 80% of all *E. coli*-originating neonatal meningitis.⁴³ The bacterial CPS is a major virulence factor owing to its ability to inhibit phagocytosis and to resist antibody-independent serum bactericidal activity. As for MenB CPS, structural similarity of the *E. coli* K1 CPS (Fig. 1) with the α -(2→8)-polysialic acid moiety of mammalian neural cell adhesion molecules, contributes to bacterial neuroinvasiveness, particularly in embryos.²⁴ In the case of *E. coli* K1, phase variation to which acetylation at O-7/O-9 of the sialic acid residues contributes for a large part,⁹ modifies the CPS physicochemical properties, which in turn affect the bacterium interaction with the host and contribute to a rapid adaptation of the bacterium to environmental changes. For instance, this modification, which occurs at high frequency (1:50–1:20), was shown to enhance *E. coli* resistance to dessication, but to reduce its aptitude to biofilm formation, thereby suggesting a delicate balance between functions triggered by phase-variation.⁴⁴ Furthermore, O-acetylation-acquired CPS resistance to hydrolysis by neuraminidases is thought to favor *E. coli* K1 survival in the intestinal tract. On another aspect, CPS-reversible O-acetylation may hamper the binding of cationic antimicrobial peptides to the bacterial membrane by modulating its hydrophobicity, and therefore interfere with the host innate immunity.²⁴

Furthermore, CPS O-acetylation in *E. coli* K1 was demonstrated to alter antigenicity and to increase immunogenicity. Immunodominant epitopes are generated^{9,45} even though O-acetylated CPS are also recognized by antibodies induced following immunization with O-acetyl negative *E. coli* K1 variants.⁹ Each *E. coli* K1 strain predominantly expresses an O-acetylated or non-O-acetylated CPS form, associated to a high reversion

rate to the opposite phenotype. Thus, it is hypothesized that phase-variation through reversible O-acetylation may facilitate bacterial evasion from a specific immune response against one CPS structure,⁴⁶ and more generally contributes to avoid the host immune defenses. Interestingly, clinical data suggest that CPS O-acetylation correlates with higher bacterial virulence.⁴⁷

3.2 *Streptococcus agalactiae* group B (GBS): potential for O-acetyl-mediated survival

GBS is an opportunistic bacterium, which asymptotically colonizes the lower digestive and vaginal tract in up to one-third of healthy women. It is the leading agent of bacterial sepsis and meningitis in newborns⁴⁸ and also a cause of serious infections in the elderly and immune-compromised individuals.⁴⁹ The α -D-NeupNAc-(2→3)- β -D-Galp-(1→ motif present in all GBS CPSs is also frequently encountered in N- and O-glycans on the surface of mammalian cells. In this context, O-acetylation appears as a source of differentiation. Indeed, O-acetylation of the α -D-NeupNAc-(2→3)-linked residue has to our knowledge never been described in mammals. In contrast, a significant portion of several GBS type-specific CPSs exhibit some variable degree (5–55%) of O-acetyl substitution of the exocyclic side chain of the outer terminal sialyl residue of their repeating unit.⁴⁰

In GBS, the level of O-acetylation obeys conserved serotype-specific patterns.⁵⁰ Genetic and biochemical manipulation of sialyl O-acetylation revealed that it did not prejudice the bacterium hydrophilicity, neither did it significantly affect complement C3b binding to the GBS surface. However, as for *E. coli* K1, acetylation at O-7 was demonstrated to be efficient at reducing sialic acid susceptibility to enzymatic removal by a variety of microbial or host sialidases. In doing so, this subtle modification contributes at protecting GBS from losing a key virulence factor, and therefore benefits GBS survival in the gastrointestinal and vaginal tracts.⁵¹ Moreover, acetylation at O-7 was shown to block Siglec9 binding to GBS, while the effect is diminished following acetyl migration from O-7 to O-9 under physiological pH.⁵¹ Detailed cellular investigations revealed that CPS O-acetylation reduced sialyl-mediated GBS escape from isolated human neutrophils.⁵⁰ Likewise, it was demonstrated that CPS O-acetylation impaired GBS evasion of neutrophil killing mechanisms in the human bloodstream, therefore contributing to some extent to attenuate GBS virulence *in vivo*.⁵⁰

While CPS may confer a survival advantage for GBS, it is also a primary target of the humoral immune response mounted by the infected host. On that basis, CPS-based vaccines have been investigated for decades against GBS infection.⁵² The importance of CPS sialylation was demonstrated, in particular with regards to GBS type III.^{53,54} Moreover, diverging from previous beliefs, a recent in depth investigation of the molecular bases of GBS III CPS by a protective monoclonal antibody revealed a linear six-residue epitope highlighting the direct involvement of the branched sialic acid.⁵⁵ For long, the possible implication of sialyl O-acetylation in CPS immunogenicity was overlooked, owing to the

chemical lability of O-acetyl groups.⁴⁰ Actually, vaccine development has focused on de-O-acetylated GBS CPSs. The subsequent finding that all CPS-conjugate vaccine prototypes evoked functional antibodies independent of the extent of O-acetylation of CPS sialic acid residues was therefore of high relevance.⁵⁶ It supported the pursuit of investigational de-O-acetylated CPS-based vaccine formulations against major disease-causing types of GBS. Noticeably, a possible explanation for these *in vivo* observations emerged from the recently disclosed X-ray data of a synthetic oligosaccharide featuring two non-O-acetylated RUs of the GBS III CPS in complex with a protective antibody. It was observed that the antibody interacts with the O-7 of the NeupNAc through a water molecule, suggesting the possible hosting of an O-acetyl group in the corresponding space.⁵⁵

3.3 *Streptococcus pneumoniae* O-acetylation: source of diversity and/or immune escape

CPS shields pneumococci from the host phagocytes and is therefore recognized as a major bacterial virulence factor.⁵⁷ With over 90 known serotypes differentiated based on their capsule, *S. pneumoniae* is a striking example of bacterial CPS diversity.⁵⁸ The CPS synthesis locus has been sequenced for most serotypes.⁵⁹ It is estimated that some 14 different putative acetyltransferase genes are present in the loci of 47 pneumococcal CPSs.⁶⁰ Yet, the exact function of most identified activities was not precisely assigned. In recent years, novel issues emerging from epidemiological studies, such as changes in serotype prevalence⁵⁸ or potential for cross-reactivity with immune antisera, have promoted a renewed interest in the genetics and structure of several pneumococcal CPSs.⁶⁰ Additional subtypes were identified and previously determined structures from known CPSs were revised, while accounting for non-carbohydrate modifications such as O-acetylation.^{60–64}

3.3.1 *S. pneumoniae* type 9A versus *S. pneumoniae* type 9V. Contrasting with previous appreciations,⁶⁵ recent structural data indicate that *S. pneumoniae* type 9A CPS is highly O-acetylated and that it shares its O-acetylation sites with type 9V CPS,⁶⁶ the structure of which was confirmed (Fig. 5). It was proposed that the common O-acetylation at the α -D-GlcpA residue and to a smaller extent at the vicinal α -D-Glcp residue, was mediated by *wcjD*, which encodes a soluble O-acetyltransferase.⁶⁴ In contrast, 6-O-acetyl- β -D-ManpNAc is exclusively found in *S. pneumoniae* 9V CPS (Fig. 5), reflecting a change in antibody recognition. This type-specific stoichiometric O-acetylation was attributed to the putative *wcjE* gene product, in this case a membrane O-acetyltransferase. Given that the *wcjE* gene is conserved among 14 pneumococci serotypes, including type 9A, it is assumed that loss-of-function mutations to *wcjE* took place while type 9A arose within the host originally colonized or infected by *wcjE*-intact *S. pneumoniae* type 9V.⁶⁷

Considering that individuals immunized with type 9V CPS produce antibodies that may either not cross-react with type 9A CPS (10–20% of vaccinees),⁶⁸ or bind more strongly to type 9V CPS than to type 9A CPS,⁶⁴

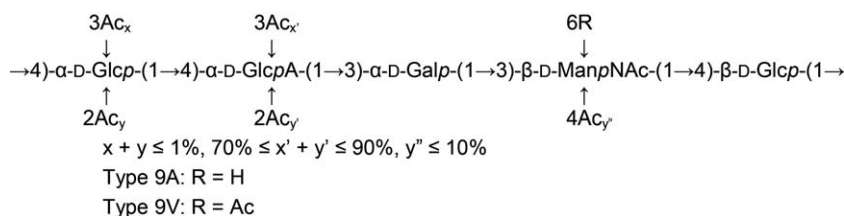


Fig. 5 Repeating units of the CPS from *S. pneumoniae* type 9A and 9V, showing specific patterns of O-acetylation.⁶⁴

the loss of ManpNAc 6-O-acetylation might offer a means to escape a host humoral immune response restricted to a *wcjE*-dependent epitope.⁶⁷ It was hypothesized that ManpNAc 6-O-acetylation such as in type 9V may facilitate host-to-host transmission, whereas its absence such as in type 9A may contribute to enhanced bacterial survival during invasive disease.⁶⁴

The differentiation of *S. pneumoniae* type 9A clinical isolates into two subtypes based on their level of expression of *wcjE*-associated epitopes brought in an additional level of complexity of the site-specific O-acetylation “on/off switch” process. It was argued that some 9A strains comprise partially functional *wcjE*, resulting in a diminished ManpNAc 6-O-acetylation of their CPS instead of its total abolition.⁶⁷ On the basis of similar, albeit more deeply investigated, findings for the types 11E and 11A from *S. pneumoniae*, it was proposed that types 9A and 9V correspond to two extremes of an antigenic spectrum with intermediate sero-variants.⁶⁹ The impact of the phenomenon on bacterial transmission, persistence and disease manifestation is not yet fully understood.⁶⁷

3.3.2 *S. pneumoniae* type 11E versus *S. pneumoniae* type 11A. Similarly, *S. pneumoniae* type 11E is thought to have materialized from *S. pneumoniae* type 11A in an independent *wcjE* evolution process.⁶⁰ As a result, the acetyl group at O-6 of the $\beta\text{-D-Galp}$ residue, known to be present in type 11A, is absent in type 11E (Fig. 6).⁶¹ In support to this finding, the $\beta\text{-D-Galp}$ residue in *S. pneumoniae* type 11F, which contain a putatively functional *wcjE* gene, is 6-O-acetylated.⁶¹ Thus, while type 11A CPS has four distinct O-acetylation sites, the CPS from *S. pneumoniae* 11E has only three.

Analysis of several *S. pneumoniae* type 11E clinical isolates revealed that each one of them exhibits a unique irreversible disrupting mutation to *wcjE*, suggesting that they are not transmitted among hosts.⁶⁰ It was proposed that every 11E strain emerged independently from a *S. pneumoniae* 11A progenitor by serotype conversion within the host, as a way to promote survival, therefore reflecting a unique model of microevolution.⁷⁰ In some instances, discrepancies arose between monoclonal antibody-based 11A/11E strain serotyping and *wcjE* sequencing analysis, suggesting that some strains share the properties of more than one serotype.⁷⁰ Subsequent structural and molecular analysis revealed that *S. pneumoniae* types 11E and 11A represent the two extremes of a population comprising variants, which differ by the activity of the *wcjE* gene, and consequently by the level of expression of a

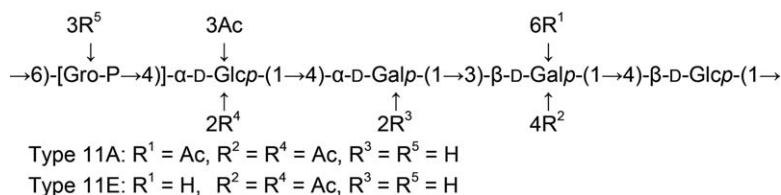


Fig. 6 Core of the CPS repeating units from *S. pneumoniae* group 11 members and sites of possible submolar O-acetylation (R^1 , R^2 , R^3 , R^4) or substitution (R^5). Type-specific patterns of O-acetylation are shown for type 11A and type 11E CPSs.⁶¹ In the case, of *S. pneumoniae* type 11B, 11C and 11F CPSs, an α -D-GlcpNAc residue replaces the α -D-Glcp residue. In the case of *S. pneumoniae* type 11B and 11F CPSs, a ribitol moiety replaces the Gro moiety.

6-O-acetyl- β -D-Galp residue in their CPS repeating unit.⁶⁹ The molecular mechanisms responsible for restraining the *wcjE* function are not yet elucidated.

In analogy with type 9A, *S. pneumoniae* type 11E is significantly more likely to occur among blood isolates – up to 50% of the strains originally typed as *S. pneumoniae* 11A by the Quellung reaction – than among strains isolated from asymptomatic nasopharyngeal colonization.^{70,71} Indeed, initial evidence suggest that the survival advantage correlated to a minor change in the chemical composition of the CPS could be specific to blood localization, either by enabling the bacterium to evade the *S. pneumoniae* type 11A specific humoral immune response mounted during asymptomatic colonization or by masking the targets for innate immune factors expressed during systemic infection.⁷⁰ More recently, it was found that Ficolin-2, a serum-associated pattern-recognition, which is involved at the early stage of the lectin complement pathway and direct opsonophagocytosis in humans, specifically binds type 11A CPS but not type 11E CPS.⁷² Concomitantly, Ficolin-2 was demonstrated to recognize most *wcjE*-encoding serotypes, despite differences in the corresponding patterns of CPS O-acetylation. In contrast, none of the corresponding *wcjE*-null isolates nor any strain producing *wcjE*-independent O-acetylated CPSs were recognized. Whereas Ficolin-2 was demonstrated to have a *wcjE*-dependent O-acetylation binding profile, *wcjE*-mediated O-acetylation is not a sufficient criteria *per se*.⁷² A model of immunity to invasive pneumococcal disease, whereby serum protection is mediated by Ficolin-2 recognition of specific O-acetyl-induced epitopes located on *S. pneumoniae* CPSs was proposed.⁷²

3.3.3 *S. pneumoniae* type 33A versus *S. pneumoniae* type 33F. Interestingly, *S. pneumoniae* types 33A and 33F may present a similar situation. They have almost identical CPS biosynthetic loci, including the membrane-bound O-acetyltransferase gene *wciG*. However, whereas the membrane-bound O-acetyltransferase gene *wcjE* is intact in the type 33A locus, it is disrupted in the case of type 33F.⁷⁴ Accordingly, the recent determination of the structure of the repeating unit from serotype 33A CPS revealed high sequence similarity with the repeating unit from type 33F CPS.⁷³ While the two CPSs have identical backbone

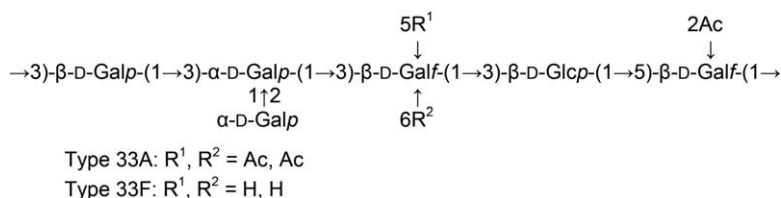


Fig. 7 Repeating units of the CPSs from *S. pneumoniae* serotypes 33A and 33F, showing the type-specific patterns of O-acetylation.^{62,73} In the case of type 33A CPS, a fourth O-acetylation site was proposed to be within the $\rightarrow 3\text{-}\beta\text{-D-Glcp}$ residue based on MS/MS data.⁷³

structure with at least one common O-acetylation site and possibly a second acetyl substitution located at an unidentified hydroxyl group within the $\rightarrow 3\text{-}\beta\text{-D-Glcp}$ residue, they differ in the 5/6-di-O-acetylation of the $\rightarrow 3\text{-}\beta\text{-D-Galf}$ residue (Fig. 7), resulting in antigenic differentiation between pneumococci 33A and 33F. More recent investigations revealed that loss of WcjE-mediated O-acetylation had little impact on cell wall adhesion or shielding. In particular, *S. pneumoniae* types 33A and 33F were shown to exhibit comparable nonspecific opsonophagocytic killing, biofilm production, and adhesion to nasopharyngeal cells, though type 33F survived short-term drying better than type 33A.⁷⁵ To our knowledge, the evolutionary relationship between the two serotypes is not yet identified.

Additional insight on the importance of O-acetyl substitutions and their occurrence within CPS repeating units emerged from the same detailed study, which also involved *wciG*-deficient variants of *S. pneumoniae* types 33A and 33F, created on purpose.⁷⁵ In contrast to WciE-mediated O-acetylation, WciG-mediated O-acetylation was found to have a major influence on the *S. pneumoniae* phenotype. Significant changes in the CPS biological properties strongly diminished its protective barrier function, resulting in a phenotype resembling that of nonencapsulated strains. The study demonstrated the importance of WciG-mediated, but not of WcjE-mediated, O-acetylation for producing protective capsules in *S. pneumoniae* type 33A.⁷⁵

4 O-Acetylated CPSs from pathogenic bacteria: implication in vaccine development

4.1 Bacterial surface CPSs as vaccine components

Protective immunity against bacterial infections may often involve an antibody response to surface polysaccharide antigens. On that basis, CPS vaccines were developed against diseases caused by *Haemophilus influenzae* type b, *N. meningitidis* (tetraivalent, Menomune[®], Mencevax[®]), *S. pneumoniae* (23-valent, Pneumovax[®]) and *Salmonella enterica typhi* (*S. Typhi*, monovalent, Typhim Vi[®]). However, these vaccines are poorly immunogenic in infants and in children younger than 18 months, limiting their usefulness. As a result, the polysaccharide vaccine licensed against *H. influenzae* b in 1985 was withdrawn from the market in the late

1980s. The need to improve protective immunity in populations at highest risk resulted in the development of polysaccharide conjugate vaccines. The strategy was most successful in the case of *H. influenzae* type b infections, whereby the covalent coupling of CPS to a carrier protein overcame the limitations encountered with the plain polysaccharide vaccines.⁷⁶ Subsequently, the attractive strategy was brilliantly extrapolated to several other diseases caused by encapsulated bacteria. Despite increased complexity, in part owing to the need for multivalency, several polysaccharide-protein conjugate vaccines were licensed over the past two decades. Major achievements occurred on the one hand in the field of *N. meningitidis* with the licensing of three quadrivalent vaccines, Menactra[®] (MenACWY-DT, Sanofi Pasteur), Menveo[®] (MenACWY-CRM₁₉₇, GSK Vaccines, formerly Novartis) and Nimenrix[®] (MenACWY-TT, Pfizer, formerly GSK Vaccines),⁷⁷ and on the other hand in the field of *S. pneumoniae* with the licensing of a 10-valent (Synflorix, GSK Vaccines) and 13-valent (Pneumovax13, Pfizer) vaccines, which followed that of Prevnar (Pfizer) – a 7-valent vaccine⁷⁸ – while a 15-valent candidate (PCV15-CRM₁₉₇, Merck Sharp & Dohme) is being investigated.

The first demonstration that O-acetylation can exert a profound effect on a polysaccharide antigenic and immunogenic properties goes back to the work by O.T. Avery and W.F. Goebel in the early 1930s.⁷⁹ The authors showed that in its O-acetylated form, the CPS from *S. pneumoniae* type 1 possesses all the immunological characteristics of its de-O-acetylated counterpart, while exhibiting additional distinctive properties. In particular, only in its O-acetylated form was the CPS able to absorb all type 1-specific antibodies from an anti-type 1 serum. In addition, immunizing mice with minute amounts of the O-acetylated CPS induced active immunity.⁷⁹ Although not all O-acetylation patterns may be relevant for biological activity, the role of CPS O-acetylation has evolved into a major concern in several instances, especially with regards to immune escape and vaccine development. The following illustrates this increasing interest by highlighting selected examples.

4.2 From CPS vaccines to CPS-conjugate vaccines: *S. pneumoniae*

Infection by *S. pneumoniae* is a major cause of morbidity and mortality especially in young children and in the elderly. Besides causing respiratory tract infections such as acute otitis media and community-acquired pneumonia, *S. pneumoniae* may disseminate and infection may evolve into a systemic disease, including among others meningitis and bacteremia. In children less than five years old, pneumococcal infections account for around 11% deaths worldwide.⁸⁰

Whereas Pneumovax[®] evoked broad serotype immunity in responders, the introduction of Prevnar[®] reduced coverage to the seven most prevalent serotypes in the vaccinated population. As a result of widespread vaccination, carriage of vaccine-targeted serotypes, whether asymptomatic or causing invasive pneumococcal disease (IPD), was drastically reduced. As a drawback to success, changes in serotype prevalence have emerged or increased substantially. While pneumococcal capsule switching has been a regular occurrence over more than half a century,⁵⁸ serotype

replacement is essentially responsible for the observed evolution.⁸¹ Nevertheless, a major discrepancy was noted in the magnitude of replacement between non vaccine type carriage and disease, with replacement being complete in carriage but not in disease.⁸¹ Moreover, as vaccines with more valences are being developed,⁸² complexity increases. In this context, the importance of labile substitutions – such as acetyl groups – may turn critical owing to possible loss during CPS sizing and chemical conjugation to a carrier. For that reason, the fine contribution of O-acetylation as part of immunodominant and possibly cross-reactive antigenic determinants is being increasingly investigated. The influence of O-acetylation on the ability of CPS-based vaccines to induce a functional antibody response is also questioned since the non-interference of O-acetyl groups would simplify vaccine development.

4.2.1 *S. pneumoniae* types 9V and 18C: non-essential O-acetylation.

As discussed above, immunization with type 9V conjugates evoked antibodies specific for the 6-O-acetyl- β -D-ManpNAc moiety on the one hand and for the CPS backbone – itself O-acetylated at multiple sites in a non-stoichiometric manner (Fig. 5) – on the other hand. Nevertheless, opsonophagocytic activity was observed in antisera mostly directed at de-O-acetylated 9V CPS. Therefore, CPS O-acetylation does not appear to be essential to induce an anti-9V functional antibody response.⁶⁸

Similarly to *S. pneumoniae* type 9V, type 18C is included in all commercially available pneumococcal conjugate vaccines. The homologous CPS is stoichiometrically acetylated at O-6 of its α -D-Glcp side chain residue (Fig. 8).⁶³ Antigenicity analysis with rabbit and human sera demonstrated that the acetate was not important for antibody recognition. Moreover, it was found that conjugates issued from the de-O-acetylated type 18C CPS elicited antibodies in rabbits, which were specific for native type 18C CPS and functional. In confirming that the acetyl group was not part of any crucial protective epitope, this finding suggested a potential for inducing cross-protection among members of the pneumococci group 18, while simplifying process development.⁶³

4.2.2 *S. pneumoniae* group 15: is O-acetylation essential for immunogenicity?

Group 15 pneumococcus is subdivided into four types (A, B, C and F). Interest in type 15B and type 15C emerged owing to the finding that these two pneumococci were repeatedly simultaneously recovered from exudates in the course of otitis media.⁸⁷ A reversible switching of the two serotypes was demonstrated *in vitro*.⁸⁷ The revised pentasaccharide structure of the core repeating units from *S. pneumoniae*

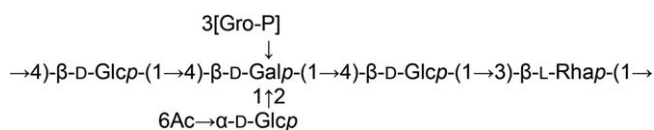


Fig. 8 Repeating units of the CPSs from *S. pneumoniae* serotypes 18C, showing stoichiometric O-acetylation at the side chain residue (Gro-P = Glycerol phosphate).⁶³

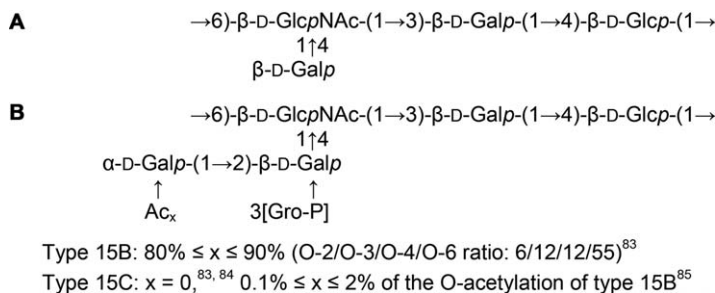


Fig. 9 Repeating units of the CPSs from *S. pneumoniae* serotypes (A) 14,⁸⁶ and (B) 15B and 15C, showing the common core tetrasaccharide, and the sites and amounts (x) of O-acetylation differentiating type 15B from type 15C.^{83–85}

types 15B and 15C CPSs⁸³ encompasses the branched tetrasaccharide repeating unit of the CPS from *S. pneumoniae* type 14.⁸⁶ The two pentasaccharides defining type 15B and 15C CPSs are closely related to the type 14 CPS repeat and only differ in O-acetylation (Fig. 9).⁸³ Both CPS loci contain an allele of the *wcjE*-like gene, *wciZ*, which is functional in type 15B but was originally established as nonfunctional in 15C, suggesting similarity with the type 9A/9V and 11A/11E systems.⁸⁸ The occurrence of intermediate serovariants as in systems 11A/11E and 9A/9V was hypothesized.⁶⁹

In this context, the effect of CPS O-acetylation on functional antibody activity and potential cross-reactivity was questioned. A measurable antigenic difference between CPSs from type 15B and type 15C was shed to light by use of sera from individuals vaccinated with Pneumovax[®]. It was demonstrated that the O-acetylation of type 15B CPS was part of the primary functional epitope for this polysaccharide. In particular, the observed correlation between loss of functional antibody activity and type 15B CPS de-O-acetylation was substantiated by the absence of functional antibodies cross-reacting with type 15C in post-vaccination sera.⁸⁹ In contrast, a recent investigation of the specificity of recognition of sera from a larger number of Pneumovax[®] immunized individuals, concluded on a rather slight difference of the opsonophagocytic activity of the induced sera in favor of type 15B in comparison to type 15C.⁸⁵ Interestingly, by use of antibodies specifically targeting WciZ-mediated O-acetylation the same study revealed that the CPS from *S. pneumoniae* 15B and 15C were O-acetylated at the same location, albeit to a much lesser extent for the later (Fig. 9B). Moreover, the extent of O-acetylation was shown to vary among serotype 15C strains and to be controlled by the number of TA repeats within the *wciZ* gene. Whereas the type 15B *wciZ* gene coding for a functional *wciZ* acetyl transferase contains eight TA repeats, the 15C *wciZ* gene is truncated, encompassing between six, seven or nine TA repeats. It was found that *wciZ* coded by (TA)₇ or (TA)₉ *wciZ* retaining partial activity, whereas type 15C strains with (TA)₆ *wciZ* had barely detectable CPS O-acetylation levels.⁸⁵ In support to the assumption that the influence of O-acetylation on CPS biological properties is polysaccharide-dependent,⁷⁵ it was suggested that CPS O-acetylation

had little effect on the biological properties of CPS in the case of *S. pneumoniae* types 15B and 15C and allowed for only limited evasion of Pneumovax[®]-elicited anti-serotype 15B antibodies.⁸⁵

Whereas Pneumovax[®] comprises type 15B, no member of pneumococci group 15 was included in Prevnar. However, issues of possible serotype replacement with non-vaccine serotypes have emerged after Prevnar introduction, and concern arose with serotypes 15A, 15B and 15C.⁹⁰ All three serotypes were isolated more frequently in children than in adults, but they diverged in terms of invasiveness. While types 15A and 15C were predominantly associated with noninvasive infections, type 15B was linked to both invasive and noninvasive infections. This properties somewhat resemble the situation described in the case of type 11A and type 11E pneumococci, respectively.

Although none of type 15 CPSs was included in Synflorix, Prevnar13 or in the 15-valent pneumococcal vaccine under development, the increasing isolation of strains from group 15 pneumococci questions the addition of representatives of this group as part of future developments. The most recent report suggests that inclusion of serotype 15B in a *S. pneumoniae* polysaccharide conjugate vaccine would induce antibodies recognizing both the acetylation pattern as well as the CPS core structure therefore preventing both types 15B and 15C extension. Alternatively, type 15C might represent a better option to elicit cross-reactive antibodies to both serotypes.⁸⁵

4.2.3 *S. pneumoniae* group 33: O-acetylation as a source of cross-protection. Interest in serogroup 33 pneumococci has risen in recent years. Serological assays have identified five serotypes organized into subgroups, namely 33A/33F (see 3.3.3), 33B/33D and 33C. Serotype 33F was included in Pneumovax[®]. While it is not part of Prevnar nor of Prevnar 13, type 33F was added in a 15-valent prototype pneumococcal conjugate vaccine under development.⁸² Recently, the elucidation of the structure of the repeating units of all five type-specific CPSs was completed,^{73,91} opening the way to a better understanding of the cross-reactivity observed with group 33 typing sera.

Except that of type 33C, all CPSs from members of group 33 pneumococci have the $\rightarrow 3\text{-}\beta\text{-D-Glcp-(1}\rightarrow 5\text{)-[2Ac]-}\beta\text{-D-Galf-(1}\rightarrow \text{disaccharide}$ in common (Fig. 7 and Fig. 10).⁹¹ O-Acetylation at O-2 of the Galf residue is mediated by WciG, an acetyltransferase common to types 33A/33F and types 33B/33D. In contrast, type 33C uses WcyO, an acetyltransferase specific for O-6 of the same Galf residue. The $\rightarrow 3\text{-}\beta\text{-D-Glcp-(1}\rightarrow 5\text{)-[2Ac]-}\beta\text{-D-Galf-(1}\rightarrow \text{disaccharide}$ was identified as an immunodominant epitope, which is not shared by type 33C CPS, suggesting a major role of the 2-O-acetyl group located on the $\rightarrow 5\text{-}\beta\text{-D-Galf}$ residue in antibody recognition.⁹¹ Moreover, as discussed above, the $\rightarrow 3\text{-}[5,6\text{-diAc}]\text{-}\beta\text{-D-Galf-(1}\rightarrow 3\text{)-}\beta\text{-D-Glcp-(1}\rightarrow \text{disaccharide}$, which was identified in the repeating unit from type 33A CPS, enabled to distinguish the former from type 33F CPS.⁷³ The $[5,6\text{-diAc}]\text{-}\beta\text{-D-Galf}$ moiety was not found in any of the other type 33 CPSs. In contrast, the $\rightarrow 3\text{-}[5,6\text{-diAc}]\text{-}\beta\text{-D-Galf-(1}\rightarrow 3\text{)-}\beta\text{-D-Glcp-(1}\rightarrow \text{disaccharide}$ is also present in the CPSs from types 35A

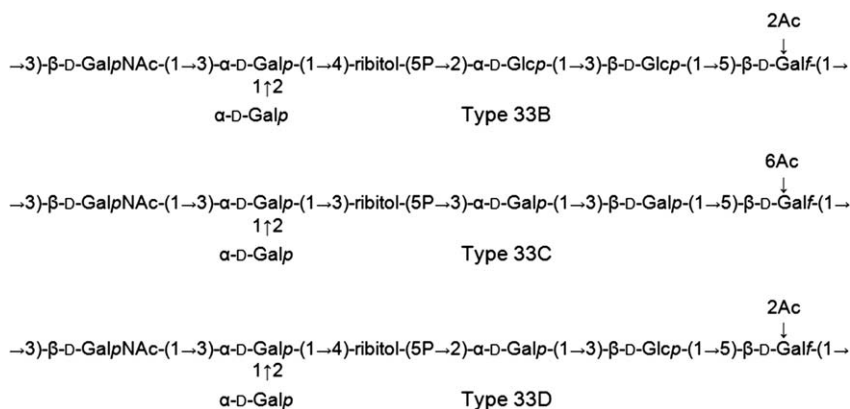


Fig. 10 Repeating units of the CPSs from *S. pneumoniae* serotypes 33B, 33C and 33D, showing the type-specific patterns of O-acetylation.^{62,91}

and 20.^{92,93} This disaccharide was proposed to be the antigenic determinant common to the three pneumococci, again suggesting a key contribution of CPS O-acetylation to pneumococci serotyping.⁷³

4.2.4 *S. pneumoniae* group 35: WciG-mediated O-acetylation as an immunodominant epitope. The historically rare *S. pneumoniae* type 35B is not included in any of the available pneumococcal vaccines. Following vaccine licensure, changes in serotype prevalence featured an increase in type 35B occurrence among *S. pneumoniae* isolates. This observation has encouraged the detailed serological analysis of clinical isolates determined to pertain to type 35B according to a genetic basis. While this study confirmed the previously established structure for the homologous CPS (Fig. 11A),⁹⁴ it also led to the identification of a novel *S. pneumoniae* serotype, named 35D by the authors.⁹⁶ *S. pneumoniae* type 35D is genetically very similar to serotype 35B, nevertheless the structure, and immunoreactivity of its CPS differ from those of the parent CPS. In particular, the isolate failed to bind to group 35 antiserum and factor serum 35a, while retaining the other 35B-factor sera specificity. It was found that *S. pneumoniae* type 35D has two inactivating mutations in the *wciG* gene coding for WciG, a predicted integral membrane O-acetyltransferase. In support to this observation, NMR analysis revealed that the CPS from *S. pneumoniae* type 35D is chemically identical to that of type 35B, except that it is not acetylated (Fig. 11A). On that basis, it was proposed that factor serum 35a targets the 2-O-acetyl group located on the $\rightarrow 6\text{-}\beta\text{-D-Galf}$ residue and conserved in all members of serogroup 35 other than the newly established serotype 35D. Accordingly, the WciG-mediated O-acetylation appears as being associated to an immunodominant epitope for members of *S. pneumoniae* group 35.⁹⁶

In contrast to non-natural types 33A and 33F *wciG*-deficient variants described above (see § 3.3.3), *S. pneumoniae* 35D is the first described serotype naturally arising from *wciG* inactivation. Its identification and that of additional clinical isolates exhibiting similar genetic and serologic profiles⁹⁷ provide additional support to the importance of genetic

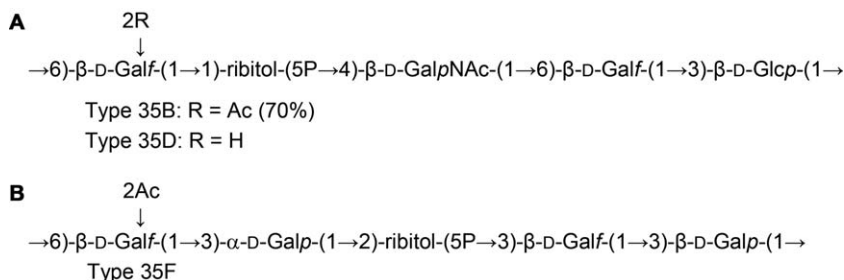


Fig. 11 Repeating units of the CPSs from *S. pneumoniae* serotypes (A) 35B and 35D and (B) 35F, exemplifying serological differences or similarities related to O-acetylation.^{94–96}

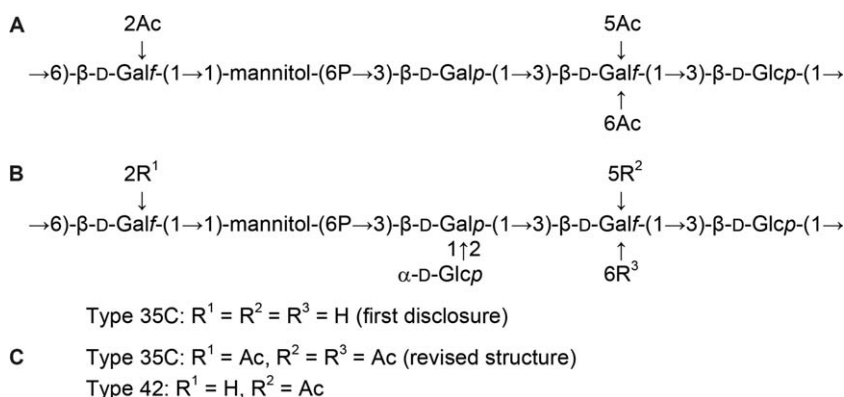


Fig. 12 Repeating units of the CPSs from *S. pneumoniae* serotypes (A) 35A,⁹² (B) 35C⁹⁵ and (C) 35C and 42,⁹⁸ exemplifying newly disclosed serological differences related to WciG-mediated O-acetylation.

inactivation of membrane-bound O-acetyltransferases as a common mechanism for *S. pneumoniae* to diversify its CPSs.^{60,96} Taking into account the fact that several *S. pneumoniae* serotypes encode *wciG*,⁵⁹ this phenomenon may deserve further exploration especially in the context of epidemiological prevalence following routine vaccination.

S. pneumoniae serotype pairing through functional/non-functional acetyltransferases has become increasingly relevant. It was demonstrated to be occasionally associated to mistyping,⁹⁸ especially since some isolates may express reduced amount of O-acetylation, a *per se* labile and possibly variable CPS substitution. In this regard, the *S. pneumoniae* 42 mistyping as *S. pneumoniae* 35C is an exquisite example.⁹⁹ The origin of the distinction between these two closely related serotypes was recently shed to light. Although, the repeating unit of *S. pneumoniae* 35C was originally reported as being a branched non-O-acetylated hexasaccharide (Fig. 12B), a more recent investigation revealed O-acetylation at three sites of this hexasaccharide (Fig. 12C).⁹⁸ The updated repeating unit for *S. pneumoniae* 35C CPS has the WcjE-controlled O-acetylation at OH-5 and OH-6 of the →3)-D-Galp residue in common with the repeating units from the CPS of serotypes 35A (Fig 12A) and 42 (Fig. 12C). Moreover, it was found to differ from the later by a WciG-mediated 2-O-acetylation at

In light of the above examples and others,¹⁰⁰ the way *S. pneumoniae* modulates its CPS O-acetylation resulting in heterogeneity within established serotypes should be looked at on a case to case basis. This aptitude of pneumococci to modify their molecular properties by means of a controlled chemical variation may raise an increased interest when considering epidemiological survey and the development of a broader serotype coverage CPS-based pneumococcal vaccine, by use of a minimal number of valences. Alternatively, a clear understanding of *S. pneumoniae* capsular microevolution will be important to perceive vaccine long term potency and potential for vaccine escape in a context of widespread vaccination.

With an estimated 21 million episodes and 200 000 deaths in 2000, typhoid fever remains a serious public health problem in developing countries.¹⁰¹ This systemic disease is caused by the bacterium *S. Typhi*, a highly adapted human-specific pathogen.¹⁰² *S. Typhi* produces a CPS known as the Vi antigen, which plays a crucial role in its virulence. In contrast to many CPSs, which are characteristic of a unique bacterium, the Vi polysaccharide is shared by *S. Paratyphi C*, *S. Dublin* and *Citrobacter freundii*. In particular and of importance for vaccine development, the Vi polysaccharide from *C. freundii*, a BSL-1 pathogen, is structurally similar and immunologically indistinguishable from the Vi polysaccharide from *S. Typhi*, a BSL-3 organism.¹⁰³ The Vi antigen is a linear homopolymer of α -(1 $\rightarrow$ 4)-linked *N*-acetyl-D-galactosaminurate partially acetylated at O-3 (Fig. 13A).¹⁰⁴

A $\rightarrow 4$)- α -D-GalpANAc-(1 $\rightarrow$
1 $\uparrow$ 3
Ac_{60-90%}

B $\rightarrow 4$)- α -D-GalpA-(1 $\rightarrow$

88 | Carbohydr. Chem., 2018, 43, 71–103

Vi vaccine should contain $0.085 (\pm 25\%) \mu\text{mol}$ O-acetyl group based on the Hestrin colorimetric assay.¹⁰⁸

The contribution of the 3-O-acetyl group to a Vi immunodominant epitope was revealed in the 1980s by use of sera,¹⁰⁹ or monoclonal antibodies.¹¹⁰ Structural analysis involving chemical modifications of the Vi antigen supported these findings. Partial base- or acid-mediated de-O-acetylation of the native Vi CPS from 65% to 45% could increase immunogenicity slightly. In contrast, immunogenicity of the Vi antigen was eliminated in the O-deacetylated Vi polysaccharide, paralleling loss of antigenicity.¹¹¹ Instead, full Vi acetylation at O-3 conferred rigidity to the CPS, possibly diminishing important intermolecular interaction.¹¹² A CPK space-filling model was proposed, showing that the bulky hydrophobic O-acetyl and N-acetyl groups dominate the molecular surface of Vi and could shield the carboxyl groups from interaction with other molecules.¹¹¹ This observation is consistent with the relative unimportance of the Vi carboxyl groups in determining its immunological properties, a rather unusual phenomenon.¹¹³ Moreover, an independent investigation also shed light on the influence of the 3-O-acetylation of the Vi repeating unit on Vi interaction with mononuclear phagocytes and lymphocytes and ability to modulate MHC class II expression.¹¹⁴

4.3.2 Analytical methods for O-acetyl detection and quantification.

The CPS O-acetylation pattern often varies for different cultures of a single bacteria strain. As a result of the established importance of the acid- and base-labile O-acetyl moieties in determining the immunologic properties of the Vi antigen, and of other CPSs of interest for vaccine development, a key test in the control of the bulk material aims at an estimation of the O-acetylation pattern. Therefore, potent analytical methods enabling the detection, content measurement, and localization of acetyl groups on purified CPSs were developed as a complement to existing colorimetric assays. In that regard, NMR spectroscopy, which is sensitive to subtle structural differences, provides a fingerprint typical of each individual CPS. It may advantageously substitute for a diversity of wet chemical approaches and has wide applications in the field of polysaccharide vaccines.¹¹⁵ In the case of the Vi antigen, a validated method, which could replace the Hestrin test,¹¹⁶ was established whereby comparison of the integrals of the resolved N-acetyl and acetate anion resonances in the spectrum of the freshly de-O-acetylated CPS provides an estimate of the 3-O-acetyl content in the purified Vi polysaccharide.¹¹⁷ Subsequently, an enzyme immunoassay (EIA) involving a Vi-specific serum produced by immunizing rabbits with a Vi conjugate was also developed. A non-linear correlation between antibody detection and the degree of 3-O-acetylation was observed, showing the non-inferiority of preparations featuring O-acetylation levels of 50% or more and native Vi CPS. In contrast, lowering further the CPS 3-O-acetyl content resulted in partial loss of detection in EIA. A threshold reactivity of 90 EU was proposed to ensure that the Vi samples meet the European Pharmacopeia specifications.¹⁰⁸ Otherwise, an anion-exchange HPLC method has been developed for quantification of

O-acetyl groups in CPSs after their hydrolysis into anions. The high-performance anion-exchange chromatography with conductivity detection (HPAEC-CD) used to determine CPS O-acetyl content is 10–20-fold more sensitive than the Hestrin test and requires less material for analysis than NMR spectroscopy. Similarly to the latter technology, HPAEC-CD had been used to quantify monosaccharide components in CPSs submitted to quality control in the context of PS or PS-conjugate vaccine development.¹¹⁸

4.3.3 Strategies towards *S. Typhi* conjugate vaccines. Besides the available Vi vaccines, the development of Vi conjugate vaccines is actively pursued.¹¹² The source and manufacturing of Vi are among the important issues under investigation. Vi CPS purified from the non-pathogenic bacteria *C. freundii* appears as a promising alternative to fermentation of the dangerous pathogen *S. Typhi*.¹¹⁹ This finding led to develop a high yielding robust and scalable process for Vi purification that gives high quality material retaining high O-acetylation levels.¹²⁰ Interestingly, taking into consideration the essentiality of the Vi O- and N-acetyl groups for both antigenicity and immunogenicity while attempting at solving technical challenges in industrial manufacturing of Vi conjugates, pectin was investigated as a replacement raw material for Vi. Pectin purified from fruits or plants is an α -(1→4)-linked D-galacturonate polymer.¹²¹ Following acetylation to a 70% extent at both O-2 and O-3, the modified pectin was shown to be antigenically indistinguishable from Vi antigen. Conjugation to a carrier protein resulted in an immunogenic preparation, which could induce an anti-Vi booster response in mice and guinea pigs, albeit at somewhat lower levels than Vi conjugates.¹²² An O-acetylated pectin conjugate was found stable over 2.5 years, as well as safe and immunogenic in adult volunteers.¹²³ Although more thorough studies are needed, this original strategy, which explores antigen sources other than pathogens, opens new avenues towards a polysaccharide-based *S. Typhi* vaccine.¹²³

4.4 O-Acetylation contribution to CPS immunogenicity is group dependent: *N. meningitidis*

N. meningitidis is an important cause of meningitis in human. This family of Gram negative bacteria is organized into 12 groups based on the expression of chemically and serologically different CPSs. Invasive infections are most commonly associated to *N. meningitidis* strains expressing CPSs featuring group A, B, C, Y, W135 and in recent years X. With the exception of MenB CPS, the use of which is believed to be at risk owing to structural similarities with human glycans, and of that from the emerging MenX, CPSs from *N. meningitidis* causing disease are major vaccine components. Besides a tetravalent combination of plain polysaccharides, CPS conjugate vaccines exist as mono- (MenA, MenC) and tetravalent combinations owing to geographical prevalence. A bivalent MenA-MenC formulation was not developed further as interest in the tetravalent combinations rose.¹²⁴ Interestingly, the four most important CPSs are diversely O-acetylated. O-acetylation is highest for MenC (88%),

MenA (75–90%), and MenY (79%) and variable for MenW (8–85%).¹²⁵ In a study demonstrating that for serogroups C, Y and W135, CPS O-acetylation did not impair recognition of meningococci by human dendritic cells, it was concluded that the modulatory effects of CPS O-acetylation on immune recognition appeared to be restricted to the humoral response.¹²⁶ As detailed below, O-acetylation is important for immunogenicity following vaccination, albeit to an extent that differs from one serogroup to another.

4.4.1 *N. meningitidis* serogroup C. The non-stoichiometric O-acetylation of MenC CPS at O-7 and O-8 of its 9)- α -D-NeupNAc-(2 $\rightarrow$ repeating unit (Fig. 3B) was described early on,³² and subsequently analyzed in more detail in the context of vaccine production.¹⁵ O-acetylation was shown to be random along the chain while its status remained constant over time.¹⁵ Moreover, the NMR study revealed an evolution of the O-acetylation pattern with time, starting from an O-8:O-7 acetylation ratio of 5:1 in a freshly dissolved sample to an equilibrium value of 1:3 after a few days at 22 °C.

Not all MenC circulating strains are O-acetylated. Epidemiological data from the US and later from UK indicated that 15% of strains isolated from patients were de-O-acetylated,^{127,128} and that the relative proportion of fatal cases was independent of the O-acetylation status. Limited data comparing vaccine and carrier sera demonstrated that the latter have a higher propensity for O-acetyl negative CPS specificity than the former,¹²⁹ suggesting that O-acetyl negative strains were more prevalent in carriers.¹²⁸

Early development of meningococci CPS vaccines led to the licensing of formulations encompassing an O-acetylated MenC CPS component. Yet, the corresponding de-O-acetylated CPS was also shown to be highly immunogenic in humans, including in children.^{130,131} Subsequent immunogenicity studies on bivalent (groups A and C) and tetravalent (groups A, C, Y and W) *N. meningitidis* CPS vaccines containing O-acetyl negative or O-acetyl positive MenC CPS also supported these findings, demonstrating that O-acetylation of the MenC CPS did not significantly influence the bactericidal anti-O-acetyl positive MenC CPS antibody titers induced by the vaccines.^{132,133} Nonetheless, a tendency for a better immunogenicity of the de-O-acetylated CPS vaccine was underlined.¹³⁰ On this basis, both O-acetyl negative and O-acetyl positive MenC CPS-conjugate vaccines were investigated for use in infants and young children. Preliminary evaluation of an O-acetyl negative MenC CPS-conjugate vaccine in adults demonstrated that the vaccine could induce bactericidal antibodies against an O-acetylated MenC strain after a single dose.¹³⁴ Such a conjugate was well tolerated and highly immunogenic, inducing immune memory against MenC strains independently of their O-acetylation status, in infants receiving three doses on a 2-, 3- and 4-month schedule.¹³⁵ Interestingly, although the conjugate induced high bactericidal IgG titers against both O-acetyl negative and O-acetyl positive MenC CPSs, a higher amount of IgGs was directed at the former. Recognition of both common backbone epitopes and unique epitopes exposed in the absence of O-acetylation was suggested.¹³⁵ Investigating

epitope-specificity and bactericidal efficacy of sera induced by a series of MenC CPS-conjugates differing in the CPS O-acetylation status, the influence of the distribution of epitope-binding on functional activity was discussed in relation to structural and conformational analysis, also taking into account the high propensity for acetyl migration from O-8 to O-7 of the α -(2 $\rightarrow$ 9)-linked sialic acid residue.¹³⁶ NMR data suggested that the O-acetyl negative and 7-O-acetylated CPSs adopt similar conformations around their glycosidic linkages. In contrast, the CPS acetylated at O-8, which is likely to feature the properties of the polysaccharide exposed at the bacterial surface, adopted a significantly different conformation.¹³⁶ Moreover, the higher potency of the O-acetyl negative CPS relative to the 7-O- and 8-O-acetylated CPSs at inhibiting the serum bactericidal activity of vaccine-induced antibodies was demonstrated, and was found even more pronounced in the latter case. In addition, the serum bactericidal activity was shown to be correlated with the anti-O-acetyl negative CPS IgG antibody titer. It was hypothesized that the protective epitopes on MenC CPS are present in the backbone polysaccharide and that acetylation at O-8 contributes to bacterial escape from immune surveillance by generating less immunogenic epitopes or masking functional epitopes.¹³⁶ The phenomenon was confirmed later on.¹³⁷ Whether simply related to O-acetylation-mediated steric hindrance or to O-acetylation-induced conformational change, the increase in inhibition of serum bactericidal activity of various vaccine-induced sera against O-acetyl positive strains appeared to grossly reflect the CPS O-acetylation status. It was suggested that O-acetylation may have evolved to facilitate survival in the host of commensal bacteria that only occasionally cause disease.¹³⁷

Three monovalent MenC conjugate vaccines were introduced in the years 1999–2000, primarily into the UK. One of them comprises a de-O-acetylated CPS component (NeisVac-C, Pfizer, formerly Baxter), whereas the other two feature an O-acetylated CPS constituent (Meningitec, Nuron Biotech formerly Pfizer, and Menjugate, GSK Vaccines formerly Novartis).⁷⁷ Although the vaccines differ in their CPS O-acetylation pattern, they are immunogenic in infants and elicit a boostable serum meningococcal bactericidal antibody response against MenC strains expressing O-acetylated or de-O-acetylated CPS. Moreover, all three vaccines induced immunological memory after a single dose in UK toddlers, justifying the use of a single dose in catch-up immunization programs.¹³⁸ The O-acetylation status of disease-causing MenC isolates in the UK did not appear to have been influenced by vaccine implementation, at least on a short term basis, especially when taking into account the natural fluctuation in CPS O-acetylation and the diminished number of MenC isolates post vaccination.¹³⁹

In contrast to the monovalent formulations, all *N. meningitidis* licensed tetravalent conjugate vaccines and combination vaccines feature an O-acetyl positive CPS for the MenC component.⁷⁷ Novel formulations are being developed. Undoubtedly, a number of factors can have a significant impact on the profile and immunogenicity of the vaccines.⁷⁷ Particular attention is paid to batch to batch consistency and CPS structural integrity post conjugation, including in terms of O-acetylation content.

NMR assays are part of the arsenal of analytical methods in use for stability studies.¹⁴⁰ They were validated as potent tools to ensure proper control of the later parameter and respect of the pharmacopeial specifications, which require O-acetylation of more than half CPS repeats in the case of MenC CPS.^{115,141} In particular, ¹H NMR spectroscopy, which provides a fingerprint for each novel vaccine lot, is part of the established physicochemical technologies employed by vaccine manufacturers.¹⁴² As for *S. Typhi*, HPAEC-CD was proposed as an alternative sensitive method for the routine quantification of total O-acetyl groups in meningococcal CPSs. However, the method does not distinguish between O-acetylation sites if more than one positions is O-acetylated as in meningococcal CPSs.¹¹⁸ It is noteworthy that a candidate international standard for MenC CPS was proposed. It is 95% O-acetylated, which implies that a correction is required for measuring the MenC content of samples featuring a different O-acetylation level.¹⁴³

4.4.2 *N. meningitidis* serogroup Y. As shown in Fig. 3C, the core → 6)-α-D-Glcp-(1→4)-α-D-NeupNAc-(2→ repeating unit of the MenY CPS was originally thought to be O-acetylated at each one of the two residues.³³ Subsequent detailed NMR analysis of the purified CPS revealed acetylation at O-9 and to a lesser extent at O-7 of the sialic acid residue.¹⁵ Acetylation is thought to be present at O-7 of the CPS sialic acid residue on the bacterial cell surface and to eventually migrate to O-9 upon CPS purification and storage.⁷⁷

At a time when the overall incidence of *N. meningitidis* cases remained stable, a dramatic increase in MenY incidence from 0% in 1989 to 32.5% in 1995 was noticed in the USA.¹⁴⁴ In a survey of meningococci isolated in the UK in the years 1996, 2000 and 2001, 79% of MenY isolates were found to express O-acetylated CPSs, a proportion that remained stable over the period of the study.¹⁴⁴ The O-acetylation status of the MenY component of the plain *N. meningitidis* CPS tetravalent vaccines was not reported. In contrast, paralleling the development of tetravalent conjugate vaccines, a growing interest for MenY CPS O-acetylation and its impact on functional properties has emerged.¹⁴⁴ Likewise, the pharmacopeial specifications requires that at least 14.3% repeats be O-acetylated in MenY CPS used in vaccine production,¹⁴¹ a level that seems to be attainable.¹⁴²

As for MenC, the influence of O-acetylation on CPS protective immunogenicity was investigated by use of a series of MenY CPS featuring various degrees of O-acetylation to serve in serological assays and conformational analysis.¹³⁷ On the basis of ¹H NMR data, it was concluded that the 9-O-acetylated CPS and the O-acetyl negative CPS adopted similar conformations around their glycosidic linkages, whether considering the glucose residue or the sialic acid. Likewise, the α-D-Glcp-(1→4)-α-D-NeupNAc bond did not seem to be affected by the 7-O-acetyl groups likely present in the MenY CPS surrounding the bacterium. In contrast, the 9-O-acetylated CPS displayed significant conformational differences around the α-D-NeupNAc-(2→6)-α-D-Glcp bond.¹³⁷ Recently, conformational analysis of the de-O-acetylated MenY CPS by molecular dynamics simulations of a three-repeat oligosaccharide in aqueous

solution revealed that the polysaccharide has a single dominant conformation.¹⁴⁵

Conjugates encompassing an O-acetyl negative MenY CPS were found more immunogenic in mice than O-acetyl positive CPS conjugates, when tested against a de-O-acetylated CPS containing antigen. Although less pronounced than in the case of MenC, this propensity was confirmed by use of a competitive inhibition bactericidal assay against an O-acetyl positive MenY strain. The O-acetyl negative MenY CPS was a more potent inhibitor, whether considering O-acetyl negative or positive CPS conjugate vaccine-induced sera. As with MenC, the study demonstrated that backbone epitopes are the primary targets for bactericidal antibodies. It was suggested that by misdirecting the immune response, O-acetylation may contribute to an escape mechanism.¹³⁷

4.4.3 *N. meningitidis* serogroup W135. Interestingly, the core repeating unit of the MenW CPS is a disaccharide closely resembling the repeating unit from MenY CPS. Defined by the common $\rightarrow 6)-\alpha\text{-D-hexp}-(1 \rightarrow 4)-\alpha\text{-D-NeupNAc}-(2 \rightarrow$ sequence, they differ from one another by the nature of the hexose component. Instead of a Glcp residue in the case of MenW, the MenY CPS comprises an $\alpha\text{-D-Galp}$ residue $1 \rightarrow 4$ -linked to the sialic acid moiety.¹⁵ For that reason, the recently reported molecular dynamics simulations of the conformation adopted in aqueous solution by a three-repeat de-O-acetylated MenY oligosaccharide was extended to the corresponding MenW oligosaccharide. The latter was found to exhibit a family of conformations including that primarily adopted by the MenY segment.¹⁴⁵ This important finding supported the previously observed cross-protection induced by the monovalent CPS vaccines,¹⁴⁶ and that suggested in the context of a recent clinical trial involving a Hib/MenC/MenY conjugate vaccine, whereby the vaccinees showed significantly higher MenW-specific seroprotection than the control group.¹⁴⁷ Moreover, by demonstrating the partial overlap between the two CPS dynamics, the molecular simulation study also provided some insights on the superior capacity of the MenY CPS to induce cross-protection with MenW.¹⁴⁵

The similarity between the MenY and MenW CPS extends to their O-acetylation pattern. As for the MenY CPS, acetylation was reported at O-7 and O-9 of the sialic acid in the MenW CPS. Likewise, migration of the acetyl group from the former position to the latter dominates to reach an equilibrium value of 1 : 2, which is probably sensitive to pH.¹⁵ However, not all MenW strains display an O-acetylated CPS. For example, isolates responsible for the Hajj outbreak in 2000 were not O-acetylated. Moreover, despite an increase in MenW O-acetylated strains in 2001, a survey at that period in the UK indicated that only 8% of the MenW isolates expressed an O-acetylated CPS with a similar distribution in carrier and case isolates.¹⁴⁴

Outbreaks of MenW in West Africa in 2001 and cases identified at various sites in pilgrims to the Hajj supported interest in the development of conjugate vaccines against this serogroup.¹⁴⁸ The impact of CPS O-acetylation status on serological measurements of anti-MenW IgG

antibodies in adults immunized with a tetravalent meningococcal CPS vaccine containing MenY and MenW O-acetyl positive CPS was investigated. Overall, there was no difference in functional activity, as measured by serum bactericidal assay against O-acetyl positive and negative MenW strains. Nevertheless, for some sera, the agreement in anti-O-acetyl positive *versus* O-acetyl negative MenW IgG assignment was serum-specific and did not reflect the functional activity *in vitro*.¹⁴⁹ An original study in mice demonstrated that MenW CPS conjugates displaying low O-acetylation levels were more immunogenic than the plain CPS and that hydrazine-mediated reduction of the O-acetyl groups in the MenW CPS had no adverse effect on immunogenicity nor on the CPS structure as detected by NMR.¹⁵⁰ Consistent with these observations, a subsequent study aimed at finding appropriate materials and methods to produce MenW conjugates revealed that MenW O-acetyl groups are not immunologically critical.¹⁴⁸ In particular, although the O-acetyl positive CPS conjugate induced some O-acetyl specific antibodies, these did not appear to contribute to the bactericidal effect. Besides, better bactericidal titers against both O-acetyl positive and negative MenW strains were detected following immunization with conjugates encompassing an O-acetyl negative or a de-O-acetylated CPS component.¹⁴⁸ The study also shed light on the influence of the degree of O-acetylation of the sialic residue in MenW CPS on periodate oxidation and in turn on the composition of the resulting conjugates. Acetylation at O-7 or O-9 can modify the availability of the more accessible exocyclic vicinal hydroxyl groups on the sialic acid, and therefore affect the involvement of galactose residues in periodate oxidation-mediated coupling of MenW CPS to proteins.¹⁴⁸ The authors concluded that the O-acetyl negative CPS may be a good starting material for preparing MenW conjugate vaccines using periodate oxidation.

Despite the low abundance of O-acetylated strains and the demonstration that O-acetylation does not contribute to an important epitope in raising functional antibodies,¹⁴⁸ all licensed MenW CPS and conjugate vaccines contain an O-acetylated CPS component. Moreover, as for MenY, the pharmacopeial specifications for MenW CPS used in vaccine production requires that more than 14.3% repeats be O-acetylated.¹⁴¹ Available data suggest that a much higher O-acetyl content can be achieved as analyzed by ¹H NMR.¹⁴²

4.4.4 *N. meningitidis* serogroup A. The MenA CPS consists of N-acetyl-D-mannosamine residues α -(1 $\rightarrow$ 6)-linked through phosphodiester bridges (Fig. 3A). Non-stoichiometric acetylation at O-3 and to a lesser extent at O-4 was reported to occur in high, albeit varying, amounts.¹⁵ A study aimed at defeating the biosynthetic origin of MenA CPS O-acetylation reported that, as revealed by ¹H NMR, the purified CPS had 60–70% ManpNAc residues that contained acetyl groups at O-3, with some species acetylated at O-4, and at both O-3 and adjacent O-4.²⁹ These findings were subsequently supported by applying whole cell high-resolution magic angle spinning NMR (HRMAS NMR) spectroscopy for the *in vivo* determination of the precise structure of MenA

CPS expressed on the bacterial surface. In both studies, the level of acetylation at O-4 was estimated to be half of that at O-3.¹⁵¹

In contrast to other meningococcal CPS, O-acetylation has a critical influence on MenA CPS immunogenicity, as demonstrated in the case of post-immunization human sera and mouse immunization.¹⁵² Thus, inhibition ELISA assay revealed that the vast majority of anti-MenA CPS antibodies induced in human receiving a CPS vaccine appear to bind epitopes involving O-acetyl groups. Moreover, comparing the immune response induced by CPS and CPS conjugates in mice revealed that de-O-acetylation of MenA CPS resulted in a marked loss of immunogenicity regardless of protein conjugation. Most importantly, the ability to induce functional bactericidal antibodies was drastically reduced in the case of the de-O-acetylated formulations.¹⁵² Nevertheless, mice immunized with a de-O-acetylated CPS conjugate did develop some functional antibodies, suggesting that epitopes other than those involving O-acetyl groups may also contribute to the development of a protective response. Obviously, the antigenic importance of MenA CPS O-acetylation resembles that observed in the case of *S. Typhi*,¹¹¹ a phenomenon that could be interpreted in terms of O-acetylation sites on the CPS backbones.¹⁵² Similarly, O-acetylation was identified as one of the critical parameters to maintain during vaccine production process. In that regard, the conjugation chemistry may have an impact when dealing with the development of CPS conjugate vaccines.

In addition to the development of meningococcal tetravalent vaccines featuring a MenA component, the incidence of MenA in the “meningitis belt” justified the development, licensure and large-scale distribution initiated in 2010 of *MenAfriVac*[™], a monovalent MenA conjugate vaccine (MenA-TT, Serum Institute of India).^{153,154} Early attempts at preparing such conjugates involved diverse methods, some of which conducted at high pH, which questioned compatibility with a sensitive CPS. Alternatively, following improvement limited periodate oxidation at the vicinal diol of the non-O-acetylated ManpNAc residues and subsequent sodium borohydride-mediated reductive amination was demonstrated viable in the case of MenA. The method was adopted starting from a MenA CPS O-acetylated to a 77–85% extent to produce high-molecular weight cross-linked lattice structures.¹⁵⁵

Otherwise in an independent assay, the O-acetylation level reached 90% in average in CPS samples provided by vaccine manufacturers, as measured by ¹H NMR.¹⁴¹ In a study aimed at demonstrating batch to batch profile consistency in a tetravalent MenACWY conjugate preparation, the MenA CPS O-acetyl content ranged between 76.6% and 84.6%, far above the pharmacopeial specifications (>61.5%),¹⁴¹ indicating that the single-site conjugation process in use retained these labile CPS decorations.¹⁴² In order to assess further the importance of CPS O-acetyl content in MenA containing vaccine formulations, an independent phase III study was conducted in healthy adults to compare the immunogenicity of two lots of tetravalent MenACWY conjugates differing in the percentage – 68% or 92% – of O-acetylation of the MenA CPS component. The tetravalent formulation with the lower level of O-acetylation was non-inferior to the

vaccine lot with the higher level of O-acetylation, and the authors concluded that in the studied range, the level of MenA CPS O-acetylation did not affect the immunogenicity of the vaccine.

5 Conclusion

Bacterial CPSs are the primary interface with the host. They act as important chemical and physical shields against environmental assaults and are significant virulence factors for pathogens. Decades ago, they were identified as attractive vaccine targets and have since then been the subject of major interest leading to remarkable breakthroughs in the prevention of bacterial diseases. However, a growing body of literature demonstrates a large propensity of bacterial cell-surface carbohydrates for modifications, which can have profound effects on host-microbe interactions. For instance, the possible occurrence of stoichiometric and/or variable degree of O-acetylation spread over multiple sites along the polysaccharide chain was established long ago. These substitutions involve a diversity of complex biosynthetic pathways evolved by the bacteria. They can simply act as alternative substituent but are also known to alter CPS chemical and physical properties, such as molecular conformation and hydrophobicity. Moreover, they can modulate CPS biological properties in a context-dependent manner, as discussed for some pneumococcal serogroups, and play a key role in the bacteria propensity for immune escape.

As highlighted in this chapter, O-acetylation can interfere with naturally exposed epitopes or generate diverging epitopes, therefore affecting the antigenicity and immunogenicity of CPSs of relevance for vaccine design. The phenomenon extends far beyond bacterial CPSs exemplified in this chapter. The CPS of other bacteria of relevance for human health, such as *Staphylococcus aureus* serotype 5 and *Burkholderia pseudomallei*, or those displayed at the surface of pathogenic fungi, for example *Cryptococcus neoformans*, are also affected. Likewise, O-acetylation of the exopolysaccharide produced by mucoid *Pseudomonas aeruginosa* in the lung of cystic fibrosis patients was shown to alter the biofilm, facilitating cell-adherence to lung epithelium, microcolony formation, and resistance to host defenses. Last but not least, as illustrated in the introduction, numerous LPSs of concern in vaccine production also display subtle antigenic diversity related to O-acetylation, mimicking in that regard the increasing complexity of pneumococcus CPSs. This is without accounting O-acetylation of plant cell wall polysaccharides or that of mammalian glycans.

References

- 1 E. R. Moxon and J. S. Kroll, *Curr. Top. Microbiol.*, 1990, **150**, 65–85.
- 2 I. Lerouge and J. Vanderleyden, *FEMS Microbiol. Rev.*, 2002, **26**, 17–47.
- 3 L. E. Comstock and D. L. Kasper, *Cell*, 2006, **126**, 847–850.
- 4 S. K. Mazmanian and D. L. Kasper, *Nat. Rev. Immunol.*, 2006, **6**, 849–858.
- 5 M. Caroff and D. Karibian, *Carbohydr. Res.*, 2003, **338**, 2431–2447.

- 6 Y. A. Knirel, in *Bacterial Lipopolysaccharides*, ed. Y. A. Knirel and M. A. Valvano, Springer-Verlag, Wien, 2011, ch. 3, pp. 41–115.
- 7 L. M. Willis and C. Whitfield, *Carbohydr. Res.*, 2013, **378**, 35–44.
- 8 S. N. Chatterjee and K. Chaudhuri, *Biochim. Biophys. Acta*, 2003, **1639**, 65–79.
- 9 F. Orskov, I. Orskov, A. Sutton, R. Schneerson, W. Lin, W. Egan, G. E. Hoff and J. B. Robbins, *J. Exp. Med.*, 1979, **149**, 669–685.
- 10 S. Herget, P. V. Toukach, R. Ranzinger, W. E. Hull, Y. A. Knirel and C. W. von der Lieth, *BMC Struct. Biol.*, 2008, **8**, 20.
- 11 A. Adibekian, P. Stallforth, M. L. Hecht, D. B. Werz, P. Gagneux and P. H. Seeberger, *Chem. Sci.*, 2011, **2**, 337–344.
- 12 R. Stenutz, A. Weintraub and G. Widmalm, *FEMS Microbiol. Rev.*, 2006, **30**, 382–403.
- 13 H. Yu and X. Chen, *Org. Biomol. Chem.*, 2007, **5**, 865–872.
- 14 S. M. Muthana, C. T. Campbell and J. C. Gildersleeve, *ACS Chem Biol.*, 2012, **7**, 31–43.
- 15 X. Lemercinier and C. Jones, *Carbohydr. Res.*, 1996, **296**, 83–96.
- 16 E. L. Nazarenko, R. J. Crawford and E. P. Ivanova, *Mar. Drugs*, 2011, **9**, 1914–1954.
- 17 Y. A. Knirel, A. V. Perepelov, A. N. Kondakova, S. y. N. Senchenkova, Z. Sidoreczyk, A. Rozalski and W. Kaca, *Innate Immun.*, 2011, **17**, 70–96.
- 18 A. V. Perepelov, M. E. Shekht, B. Liu, S. D. Shevelev, V. A. Ledov, S. N. Senchenkova, V. L. L'Vov, A. S. Shashkov, L. Feng, P. G. Aparin, L. Wang and Y. A. Knirel, *FEMS Immunol. Med. Microbiol.*, 2012, **66**, 201–210.
- 19 R. Schauer, G. V. Srinivasan, D. Wipfler, B. Kniep and R. Schwarz-Albiez, in *The Molecular Immunology of Complex Carbohydrates-3*, ed. A. M. Wu, Springer Science + Business Media, LLC, New York, 2011, ch. 28, pp. 525–548.
- 20 E. L. Deszo, S. M. Steenbergen, D. I. Freedberg and E. R. Vimr, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 5564–5569.
- 21 K. Hygge Blakeman, A. Weintraub and G. Widmalm, *Eur. J. Biochem.*, 1998, **251**, 534–537.
- 22 V. L. L'vov, A. P. Iakovlev, A. S. Shashkov and B. A. Dmitriev, *Bioorg. Khim.*, 1991, **17**, 111–120.
- 23 Y. A. Knirel, A. S. Shashkov, S. N. Senchenkova, S. Merino and J. M. Tomas, *Carbohydr. Res.*, 2002, **337**, 1381–1386.
- 24 M. R. King, S. M. Steenbergen and E. R. Vimr, *Trends Microbiol.*, 2007, **15**, 196–202.
- 25 A. L. Lewis, H. Cao, S. K. Patel, S. Diaz, W. Ryan, A. F. Carlin, V. Thon, W. G. Lewis, A. Varki, X. Chen and V. Nizet, *J. Biol. Chem.*, 2007, **282**, 27562–27571.
- 26 J. Kubler-Kielb, E. Vinogradov, C. Chu and R. Schneerson, *Carbohydr. Res.*, 2007, **342**, 643–647.
- 27 A. Molinaro, C. De Castro, B. O. Petersen, J. O. Duus, M. Parrilli and O. Holst, *Angew. Chem. Int. Ed. Engl.*, 2000, **39**, 156–160.
- 28 A. V. Perepelov, V. L. L'Vov, B. Liu, S. y. N. Senchenkova, M. E. Shekht, A. S. Shashkov, L. Feng, P. G. Aparin, L. Wang and Y. A. Knirel, *Carbohydr. Res.*, 2009, **344**, 687–692.
- 29 S. K. Gudlavalleti, A. K. Datta, Y. L. Tzeng, C. Noble, R. W. Carlson and D. S. Stephens, *J. Biol. Chem.*, 2004, **279**, 42765–42773.
- 30 T. Fiebig, F. Freiburger, V. Pinto, M. R. Romano, A. Black, C. Litschko, A. Bethe, D. Yashunsky, R. Adamo, A. Nikolaev, F. Berti and R. Gerardy-Schahn, *J. Biol. Chem.*, 2014, **289**, 19395–19407.

- 31 D. R. Bundle, I. C. Smith and H. J. Jennings, *J. Biol. Chem.*, 1974, **249**, 2275–2281.
- 32 A. K. Bhattacharjee, H. J. Jennings, C. P. Kenny, A. Martin and I. C. Smith, *J. Biol. Chem.*, 1975, **250**, 1926–1932.
- 33 A. K. Bhattacharjee, H. J. Jennings, C. P. Kenny, A. Martin and I. C. Smith, *Can. J. Biochem.*, 1976, **54**, 1–8.
- 34 A. K. Bergfeld, H. Claus, U. Vogel and M. Muhlenhoff, *J. Biol. Chem.*, 2007, **282**, 22217–22227.
- 35 E. C. Schulz, A. K. Bergfeld, R. Ficner and M. Muhlenhoff, *PLoS One*, 2011, **6**, e17403.
- 36 H. Claus, R. Borrow, M. Achtman, G. Morelli, C. Kantelberg, E. Longworth, M. Frosch and U. Vogel, *Mol. Microbiol.*, 2004, **51**, 227–239.
- 37 H. J. Lee, B. Rakic, M. Gilbert, W. W. Wakarchuk, S. G. Withers and N. C. J. Strynadka, *J. Biol. Chem.*, 2009, **284**, 24501–24511.
- 38 A. K. Bergfeld, H. Claus, N. K. Lorenzen, F. Spielmann, U. Vogel and M. Muehlenhoff, *J. Biol. Chem.*, 2009, **284**, 6–16.
- 39 M. J. Cieslewicz, D. Chaffin, G. Glusman, D. Kasper, A. Madan, S. Rodrigues, J. Fahey, M. R. Wessels and C. E. Rubens, *Infect. Immun.*, 2005, **73**, 3096–3103.
- 40 A. L. Lewis, V. Nizet and A. Varki, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 11123–11128.
- 41 A. L. Lewis, M. E. Hensler, A. Varki and V. Nizet, *J. Biol. Chem.*, 2006, **281**, 11186–11192.
- 42 S. M. Steenbergen, Y.-C. Lee, W. F. Vann, J. Vionnet, L. F. Wright and E. R. Vimr, *J. Bacteriol.*, 2006, **188**, 6195–6206.
- 43 J. B. Robbins, G. H. McCracken, E. C. Gotschlich, F. Ørskov, I. Ørskov and L. A. Hanson, *N. Engl. J. Med.*, 1974, **290**, 1216–1220.
- 44 I. L. Mordhorst, H. Claus, C. Ewers, M. Lappann, C. Schoen, J. Elias, J. Batzilla, U. Dobrindt, L. H. Wieler, A. K. Bergfeld, M. Mühlenhoff and U. Vogel, *Environ. Microbiol.*, 2009, **11**, 3154–3165.
- 45 R. Torensma, A. Vanwijk, M. J. C. Visser, A. Bouter, M. Rozenbergarska and J. Verhoef, *J. Clin. Microbiol.*, 1991, **29**, 1356–1358.
- 46 J. Colino and I. Outschoorn, *Microb Pathogenesis*, 1999, **27**, 187–196.
- 47 H. Frasa, J. Procee, R. Torensma, A. Verbruggen, A. Algra, M. Rozenbergarska, K. Kraaijeveld and J. Verhoef, *J. Clin. Microbiol.*, 1993, **31**, 3174–3178.
- 48 Centres for Disease Control and Prevention, *MMWR Morb. Mortal. Wkly. Rep.*, 2009, **59**, 109–112.
- 49 M. S. Edwards and C. J. Baker, *Clin. Infect. Dis.*, 2005, **41**, 839–847.
- 50 S. Weiman, S. Uchiyama, F.-Y. C. Lin, D. Chaffin, A. Varki, V. Nizet and A. L. Lewis, *Biochem. J.*, 2010, **428**, 163–168.
- 51 S. Weiman, S. Dahesh, A. F. Carlin, A. Varki, V. Nizet and A. L. Lewis, *Glycobiology*, 2009, **19**, 1204–1213.
- 52 C. M. Healy and C. J. Baker, in *Vaccines*, ed. S. A. Plotkin, W. A. Orenstein and P. A. Offit, Elsevier Saunders, 6th edn., 2013, ch. 59, pp. 1176–1181.
- 53 H. J. Jennings, C. Lugowski and D. L. Kasper, *Biochemistry*, 1981, **20**, 4511–4518.
- 54 D. Safari, H. A. T. Dekker, G. T. Rijkers, A. van der Ende, J. P. Kamerling and H. Snippe, *Glycoconjugate J.*, 2011, **28**, 557–562.
- 55 F. Carboni, R. Adamo, M. Fabbrini, R. De Ricco, V. Cattaneo, B. Brogioni, D. Veggi, V. Pinto, I. Passalacqua, D. Oldrini, R. Rappuoli, E. Malito, I. y. R. Margarit and F. Berti, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, 5017–5022.

- 56 P. S. Pannaraj, M. S. Edwards, K. T. Ewing, A. L. Lewis, M. A. Rench and C. J. Baker, *Vaccine*, 2009, **27**, 4452–4456.
- 57 C. Hyams, E. Camberlein, J. M. Cohen, K. Bax and J. S. Brown, *Infect. Immun.*, 2010, **78**, 704–715.
- 58 K. L. Wyres, L. M. Lambertsen, N. J. Croucher, L. McGee, A. von Gottberg, J. Linares, M. R. Jacobs, K. G. Kristinsson, B. W. Beall, K. P. Klugman, J. Parkhill, R. Hakenbeck, S. D. Bentley and A. B. Brueggemann, *J. Infect. Dis.*, 2013, **207**, 439–449.
- 59 S. D. Bentley, D. M. Aanensen, A. Mavroidi, D. Saunders, E. Rabinowitsch, M. Collins, K. Donohoe, D. Harris, L. Murphy, M. A. Quail, G. Samuel, I. C. Skovsted, M. S. Kaltoft, B. Barrell, P. R. Reeves, J. Parkhill and B. G. Spratt, *PLoS Genet.*, 2006, **2**, 262–269.
- 60 J. J. Calix and M. H. Nahm, *J. Infect. Dis.*, 2010, **202**, 29–38.
- 61 J. J. Calix, M. H. Nahm and E. R. Zartler, *J. Bacteriol.*, 2011, **193**, 5271–5278.
- 62 X. Lemercinier and C. Jones, *Carbohydr. Res.*, 2006, **341**, 68–74.
- 63 J. Chang, Y. Serrano, R. Garrido, L. M. Rodriguez, J. Pedrosa, F. Cardoso, Y. Valdes, D. Garcia, V. Fernandez-Santana and V. Verez-Bencomo, *Vaccine*, 2012, **30**, 7090–7096.
- 64 J. J. Calix, J. S. Saad, A. M. Brady and M. H. Nahm, *J. Biol. Chem.*, 2012, **287**, 13996–14003.
- 65 T. J. Rutherford, C. Jones, D. B. Davies and A. Clare Elliott, *Carbohydr. Res.*, 1994, **265**, 97–111.
- 66 T. J. Rutherford, C. Jones, D. B. Davies and A. C. Elliott, *Carbohydr. Res.*, 1991, **218**, 175–184.
- 67 J. J. Calix, M. B. Oliver, L. K. Sherwood, B. Beall, S. K. Hollingshead and M. H. Nahm, *J. Infect. Dis.*, 2011, **204**, 1585–1595.
- 68 T. B. McNeely, J. M. Staub, C. M. Rusk, M. J. Blum and J. J. Donnelly, *Infect. Immun.*, 1998, **66**, 3705–3710.
- 69 J. J. Calix, A. M. Brady, V. Y. Du, J. S. Saad and M. H. Nahm, *J. Clin. Microbiol.*, 2014, **52**, 758–765.
- 70 J. J. Calix, R. Dagan, S. I. Pelton, N. Porat and M. H. Nahm, *Clin. Infect. Dis.*, 2012, **54**, 794–799.
- 71 R. Camilli, B. L. Spencer, M. Moschioni, V. Pinto, F. Berti, M. H. Nahm and A. Pantosti, *PLoS One*, 2014, **9**, e100722.
- 72 A. M. Brady, J. J. Calix, J. Yu, K. A. Geno, G. R. Cutter and M. H. Nahm, *J. Infect. Dis.*, 2014, **210**, 1155–1165.
- 73 F. L. Lin, E. Vinogradov, C. Deng, S. Zeller, B. A. Green, K. U. Jansen and V. Pavliak, *Carbohydr. Res.*, 2013, **380**, 101–107.
- 74 A. Mavroidi, D. M. Aanensen, D. Godoy, I. C. Skovsted, M. S. Kaltoft, P. R. Reeves, S. D. Bentley and B. G. Spratt, *J. Bacteriol.*, 2007, **189**, 7841–7855.
- 75 B. L. Spencer, J. S. Saad, A. T. Shenoy, C. J. Orihuela and M. H. Nahm, *Infect. Immun.*, 2017, **85**, e00132–0017, DOI: 10.1128/IAI.00132-17.
- 76 H. Peltola, *Clin. Microbiol. Rev.*, 2000, **13**, 302–317.
- 77 M. Bröker, F. Berti and P. Costantino, *Human Vaccines Immunother.*, 2016, **12**, 1808–1824.
- 78 J. T. Poolman, C. C. Peeters and G. P. van den Dobbelsteen, *Expert Rev. Vaccines*, 2013, **12**, 1379–1394.
- 79 O. T. Avery and W. F. Goebel, *J. Exp. Med.*, 1933, **58**, 731–755.
- 80 K. L. O'Brien, L. J. Wolfson, J. P. Watt, E. Henkle, M. Deloria-Knoll, N. McCall, E. Lee, K. Mulholland, O. S. Levine and T. Cherian, *Lancet*, 2009, **374**, 893–902.
- 81 D. M. Weinberger, R. Malley and M. Lipsitch, *Lancet*, 2011, **378**, 1962–1973.

- 82 J. M. Skinner, L. Indrawati, J. Cannon, J. Blue, M. Winters, J. MacNair, N. Pujar, W. Manger, Y. H. Zhang, J. Antonello, J. Shiver, M. Caulfield and J. H. Heinrichs, *Vaccine*, 2011, **29**, 8870–8876.
- 83 C. Jones and X. Lemercinier, *Carbohydr. Res.*, 2005, **340**, 403–409.
- 84 P. E. Jansson, B. Lindberg, U. Lindquist and J. Ljungberg, *Carbohydr. Res.*, 1987, **162**, 111–116.
- 85 B. L. Spencer, A. T. Shenoy, C. J. Orihuela and M. H. Nahm, *Clin. Vaccine Immunol.*, 2017, **24**, e00099–17, DOI: 10.1128/CVI.00099-17.
- 86 B. Lindberg, J. Lonngren and D. A. Powell, *Carbohydr. Res.*, 1977, **58**, 177–186.
- 87 P. S. Venkateswaran, N. Stanton and R. Austrian, *J. Infect. Dis.*, 1983, **147**, 1041–1054.
- 88 S. van Selm, L. M. van Cann, M. A. B. Kolkman, B. A. M. van der Zeijst and J. P. M. van Putten, *Infect Immun.*, 2003, **71**, 6192–6198.
- 89 G. Rajam, G. M. Carlone and S. Romero-Steiner, *Clin. Vaccine Immunol.*, 2007, **14**, 1223–1227.
- 90 M. R. Jacobs, C. E. Good, S. Bajaksouzian and A. R. Windau, *Clin. Infect. Dis.*, 2008, **47**, 1388–1395.
- 91 F. L. Lin, E. Vinogradov, C. Deng, S. Zeller, L. Phelan, B. A. Green, K. U. Jansen and V. Pavliak, *Carbohydr. Res.*, 2014, **383**, 97–104.
- 92 L. M. Beynon, J. C. Richards and M. B. Perry, *Eur. J. Biochem.*, 1997, **250**, 163–167.
- 93 J. J. Calix, R. J. Porambo, A. M. Brady, T. R. Larson, J. Yother, C. Abeygunwardana and M. H. Nahm, *J. Biol. Chem.*, 2012, **287**, 27885–27894.
- 94 L. M. Beynon, J. C. Richards, M. B. Perry and P. J. Kniskern, *Can. J. Chem.*, 1995, **73**, 41–48.
- 95 C. A. Bush, J. O. Cisar and J. Yang, *J. Bacteriol.*, 2015, **197**, 2762–2769.
- 96 K. A. Geno, J. S. Saad and M. H. Nahm, *J. Clin. Microbiol.*, 2017, **55**, 1416–1425.
- 97 M. Staples, R. M. A. Graham, V. Hicks, J. Strachan, A. Gonçalves da Silva, J. Peverall, V. Wicks and A. V. Jennison, *Clin. Microbiol. Infect.*, 2017, **23**, 476–479.
- 98 K. A. Geno, C. A. Bush, M. Wang, C. Jin, M. H. Nahm and J. Yang, *J. Clin. Microbiol.*, 2017, **55**, 2775–2784.
- 99 H. B. Konradsen, *Vaccine*, 2005, **23**, 1368–1373.
- 100 K. A. Geno, G. L. Gilbert, J. Y. Song, I. C. Skovsted, K. P. Klugman, C. Jones, H. B. Konradsen and M. H. Nahm, *Clin. Microbiol. Rev.*, 2015, **28**, 871–899.
- 101 M. I. Khan, R. L. Ochiai and J. D. Clemens, *Expert. Rev. Vaccines*, 2010, **9**, 485–496.
- 102 C. M. Parry, T. T. Hien, G. Dougan, N. J. White and J. J. Farrar, *N. Engl. J. Med.*, 2002, **347**, 1770–1782.
- 103 X. Hu, Z. Chen, K. Xiong, J. Wang, X. Rao and Y. Cong, *Crit. Rev. Microbiol.*, 2017, **43**, 440–452.
- 104 L. Kenne and A. A. Lindberg, in *Bacterial Polysaccharides*, ed. G. O. Aspinall, Academic Press, Inc., New York, 1981, vol. 2, p. 315.
- 105 J. D. Robbins and J. B. Robbins, *J. Infect. Dis.*, 1984, **150**, 436–449.
- 106 *WHO Technical Report Series*, 1994, **840**, 14–29.
- 107 *European Pharmacopeia. Edition 4.2*, 2003, 2842–2843.
- 108 S. Rijpkema, Z. Durrani, X. Lemercinier and C. Jones, *Biologicals*, 2004, **32**, 11–16.
- 109 B. Szewczyk and A. Taylor, *Infect. Immun.*, 1980, **29**, 539–544.

- 110 A. Qadri, S. Ghosh and G. P. Talwar, *J. Immunoassay*, 1990, **11**, 235–250.
- 111 S. C. Szu, X. R. Li, A. L. Stone and J. B. Robbins, *Infect. Immun.*, 1991, **59**, 4555–4561.
- 112 S. C. Szu, *Expert. Rev. Vaccines.*, 2013, **12**, 1273–1286.
- 113 S. Bystrycky and S. C. Szu, *Biophys. Chem.*, 1994, **51**, 1–7.
- 114 A. Qadri, *Immunology*, 1997, **92**, 146–152.
- 115 C. Jones, *J. Pharm. Biomed. Anal.*, 2005, **38**, 840–850.
- 116 S. Hestrin, *J. Biol. Chem.*, 1949, **180**, 249–261.
- 117 X. Lemercinier, I. Martinez-Cabrera and C. Jones, *Biologicals*, 2000, **28**, 17–24.
- 118 G. Kao and C. M. Tsai, *Vaccine*, 2004, **22**, 335–344.
- 119 S. Rondini, F. Micoli, L. Lanzilao, C. Hale, A. J. Saul and L. B. Martin, *Clin. Vaccine Immunol.*, 2011, **18**, 460–468.
- 120 F. Micoli, S. Rondini, I. Pisoni, C. Giannelli, V. Di Cioccio, P. Costantino, A. Saul and L. B. Martin, *Vaccine*, 2012, **30**, 853–861.
- 121 W. Pilnik and A. G. J. Voragen, in *The Biochemistry of Fruits and Their Products*, ed. A. C. Hulme, Academic Press, London, 1970, vol. 1, pp. 53–87.
- 122 Z. Kossaczka, S. Bystrycky, D. A. Bryla, J. Shiloach, J. B. Robbins and S. C. Szu, *Infect. Immun.*, 1997, **65**, 2088–2093.
- 123 S. C. Szu, K. F. Y. Lin, S. Hunt, C. Chu and N. D. Thinh, *Vaccine*, 2014, **32**, 2618–2622.
- 124 D. Pace, *Expert. Opin. Biol. Ther.*, 2013, **13**, 11–33.
- 125 Y. L. Tzeng, J. Thomas and D. S. Stephens, *Crit. Rev. Microbiol.*, 2016, **42**, 759–772.
- 126 A. Villwock, C. Schmitt, M. Frosch and O. Kurzai, *Intl. J. Med. Microbiol.*, 2008, **298**, 591–597.
- 127 M. A. Apicella and H. A. Feldman, *Proc. Soc. Exp. Biol. Med.*, 1976, **152**, 289–291.
- 128 R. Borrow, E. Longworth, S. J. Gray and E. B. Kaczmarek, *FEMS Immunol. Med. Microbiol.*, 2000, **28**, 189–191.
- 129 G. Arakere and C. E. Frasch, *Infect. Immun.*, 1991, **59**, 4349–4356.
- 130 M. P. Glode, E. B. Lewin, A. Sutton, C. T. Le, E. C. Gotschlich and J. B. Robbins, *J. Infect. Dis.*, 1979, **139**, 52–59.
- 131 M. C. Steinhoff, E. B. Lewin, E. C. Gotschlich and J. B. Robbins, *Infect. Immun.*, 1981, **34**, 144–146.
- 132 I. Vodopija, Z. Baklaic, P. Hauser, P. Roelants, F. E. André and A. Safary, *Infect Immun.*, 1983, **42**, 599–604.
- 133 H. Peltola, A. Safary, H. Kayhty, V. Karanko and F. E. Andre, *Pediatrics*, 1985, **76**, 91–96.
- 134 P. Richmond, D. Goldblatt, P. C. Fusco, J. D. Fusco, I. Heron, S. Clark, R. Borrow and F. Michon, *Vaccine*, 1999, **18**, 641–646.
- 135 P. Richmond, R. Borrow, J. Findlow, S. Martin, C. Thornton, K. Cartwright and E. Miller, *Infect. Immun.*, 2001, **69**, 2378–2382.
- 136 F. Michon, C. H. Huang, E. K. Farley, L. Hronowski, J. Di and P. C. Fusco, *Dev. Biol.*, 2000, **103**, 151–160.
- 137 P. C. Fusco, E. K. Farley, C.-H. Huang, S. Moore and F. Michon, *Clin. Vaccine Immunol.*, 2007, **14**, 577–584.
- 138 P. Richmond, R. Borrow, D. Goldblatt, J. Findlow, S. Martin, R. Morris, K. Cartwright and E. Miller, *J. Infect. Dis.*, 2001, **183**, 160–163.
- 139 P. Balmer, R. Borrow and E. Miller, *J. Med. Microbiol.*, 2002, **51**, 717–722.
- 140 M. M. Ho, X. Lemercinier, B. Bolgiano, D. Crane and M. J. Corbel, *Biotechnol. Appl. Bioc.*, 2001, **33**, 91–98.
- 141 C. Jones and X. Lemercinier, *J. Pharm. Biomed. Anal.*, 2002, **30**, 1233–1247.

- 142 A. Bardotti, G. Averani, F. Berti, S. Berti, V. Carinci, S. D'Ascenzi, B. Fabbri, S. Giannini, A. Giannozzi, C. Magagnoli, D. Proietti, F. Norelli, R. Rappuoli, S. Ricci and P. Costantino, *Vaccine*, 2008, **26**, 2284–2296.
- 143 C. Vipond, C. J. Swann, T. W. Dougall, P. Rigsby, F. Gao, N. J. Beresford and B. Bolgiano, *Biologicals*, 2017, **47**, 33–45.
- 144 E. Longworth, P. Fernsten, T. L. Mininni, U. Vogel, H. Claus, S. Gray, E. Kaczmarek and R. Borrow, *FEMS Immunol. Med. Microbiol.*, 2002, **32**, 119–123.
- 145 M. M. Kuttel, Z. Timol and N. Ravenscroft, *Carbohydr. Res.*, 2017, **446–447**, 40–47.
- 146 J. M. Griffiss, B. L. Brandt, P. L. Altieri, G. B. Pier and S. L. Berman, *Infect. Immun.*, 1981, **34**, 725–732.
- 147 M. Leonardi, T. Latiolais, K. Sarpong, M. Simon, J. Twiggs, P. Lei, S. Rinderknecht, M. Blatter, V. Bianco, Y. Baine, L. R. Friedland and J. M. Miller, *Vaccine*, 2015, **33**, 933–941.
- 148 S. K. Gudlavalleti, C. H. Lee, S. E. Norris, M. Paul-Satyaseela, W. F. Vann and C. E. Frasch, *Vaccine*, 2007, **25**, 7972–7980.
- 149 P. C. Giardina, E. Longworth, R. E. Evans-Johnson, M. L. Bessette, H. Zhang, R. Borrow, D. Madore and P. Fernsten, *Clin. Diag. Lab. Immunol.*, 2005, **12**, 586–592.
- 150 Z. G. Jin, G. A. Bohach, J. Shiloach, S. E. Norris, D. I. Freedberg, C. Deobald, B. Coxon, J. B. Robbins and R. Schneerson, *Infect. Immun.*, 2005, **73**, 7887–7893.
- 151 S. K. Gudlavalleti, C. M. Szymanski, H. C. Jarrell and D. S. Stephens, *Carbohydr. Res.*, 2006, **341**, 557–562.
- 152 D. S. Berry, F. Lynn, C. H. Lee, C. E. Frasch and M. C. Bash, *Infect. Immun.*, 2002, **70**, 3707–3713.
- 153 C. E. Frasch, M. P. Preziosi and F. M. LaForce, *Human Vaccines Immunother.*, 2012, **8**, 715–724.
- 154 C. E. Frasch, S. V. Kapre, C. H. Lee and J. M. Preaud, *Clin. Infect. Dis.*, 2015, **61**, S404–S409.
- 155 C. H. Lee, W. C. Kuo, S. Beri, S. Kapre, J. S. Joshi, N. Bouveret, F. M. LaForce and C. E. Frasch, *Vaccine*, 2009, **27**, 726–732.

Regioselective glycosylation: What's new?

Laurent Legentil* and Vincent Ferrières*

DOI: 10.1039/9781788010641-00104

The aim of this chapter is to show how increased understanding of the mechanisms related to both carbohydrate-processing enzymes and intrinsic properties of mono-glycosyl residue have significantly impacted our way of thinking about the glycosylation reaction. A focus is made on the importance of low energy interactions. Some recent examples related to biocatalytic approaches and chemical synthesis are given.

1 Introduction

Biological processes are the results of billions years of evolution through different geological eras. Their most striking characteristic remains the use of very simple architectures, like DNA or protein in order to perform very efficiently complex cascade of events. Proteins in particular are the perfect example where form and function are intimately linked to provide the proper results. The understanding of such ballet gives new insights for scientists and allowed to mine new ideas based on the diversion of such natural processes. The imitation of Nature and the inspiration by Nature is called biomimicry and has influenced greatly many chemists over the history of sciences.^{1–3}

Enzymatic glycosylations present all required assets for biomimicry.⁴ The first one is molecular recognition. Active sub-sites of enzymes are able to select both donor and acceptor thanks to specific local hydrogen bond networks and low energy interactions. The second asset relies on the activation of the glycosyl donor which also results in anomeric stereochemical control. Thirdly, the hydroxyl group, which will be further involved in the new glycosidic linkage, is specifically activated thanks to controlled conformation of the acceptor entity within the active site and hydrogen bond with the amino acid residues bearing a negative charge. This finally results in regioselective couplings. Nevertheless working with enzymes shows some drawbacks like the lack of flexibility and stability. Attempt to mimic such selectivity remains the grail for glycochemists. Indeed taking advantage of this biomimicry may present many benefits: decreasing the number of synthetic and purification steps, saving time, optimizing financial resources including wastes management, and, very importantly, democratizing glycochemistry. This is why different strategies have been devised in order to differentiate the different hydroxyl groups on the acceptor or to modulate the reactivity of the donors.

Based on these two parallel approaches, this review will first focus on recent developments based on intakes of biocatalysis. The second chapter will be dedicated to bioinspired chemical glycosylations.

Ecole Nationale Supérieure de Chimie de Rennes, CNRS, UMR 6226, 11 Allée de Beaulieu, CS 50837, 35708, Rennes Cedex 7, France.

E-mail: laurent.legentil@ensc-rennes.fr; vincent.ferrieres@ensc-rennes.fr

2 Tackling the regioselectivity problem using biocatalysts

Regioselectivity in glycosylation remains one of the key problems to tackle when dealing with the elaboration of glycosides. As exemplified in the second part of this chapter, elaborate strategies have been conceived to minimize the number of protection/deprotection steps using transient intermediates or activating protecting groups. Nevertheless, Nature manages to provide complex oligosaccharides with defined structures very efficiently using a powerful arsenal of enzymes involved in the biosynthesis, metabolisms and catabolism of glycans and glycoconjugates. The development of proteomics and genomics as well as the democratization of the molecular biology techniques have enabled the discovery and characterisation of these carbohydrate-processing enzymes and have opened a new era in the synthesis of glycans.

2.1 Carbohydrate-processing enzymes as biocatalysts

Many biological events are controlled through a subtle interaction of a specific saccharidic pattern with another partner (glycans, lectins, antibodies...). Such patterns were obtained *in vivo* thanks to highly specific enzymes, namely the glycosyltransferases (GT). Their role is crucial since they transfer a monosaccharidyl entity onto an acceptor from the corresponding nucleotide-sugars or the sugar 1-phosphate. As organisms are in constant transformation, the glycosylation profile must also be adaptable and flexible. Another category of enzyme, the glycosyl hydrolases (GH) or glycosidases, which are used for the remodelling and the recycling of glycans, is making this possible. As the natural glycans are multifunctional, the glycosidases are more promiscuous and accept a large panel of oligosaccharides or polysaccharides as substrate.

Even if their biological roles are different, the mechanism of action of GT and GH present numerous common points.⁵ Both glycosylation (GT) and hydrolysis of the glycosidic bond (GH) occur within a cavity at the surface of the protein, constituted of a catalytic site and a binding site. It involves mostly two catalytic amino acids that work in tandem (Fig. 1). Depending on the stereoisomerism of the reaction, two mechanisms are possible.⁶ In the case of a mechanism with retention of configuration (meaning that the final stereoisomerism of the glycosidic bond is identical to the initial one), the acid/base residue activates the aglycon in order to weaken the phosphoester or the glycosidic bond. It is followed by the attack of the nucleophilic residue to give the glycosyl-enzyme intermediate and the departure of the aglycon. A second displacement takes place in presence of an acceptor (GT) or water (GH). Inverting carbohydrate-processing enzymes work in a different fashion. Direct attack of the anomeric position by the acceptor or by water leads to an inversion of configuration. No glycosyl-enzyme intermediate is formed and both amino acids are involved through acid/base catalysis.

Glycosyltransferases and glycosyl hydrolases are the product of million years of evolution and they accordingly possess the specificity and the efficiency that most synthetic catalysts lack. They process carbohydrates under mild conditions without needing the presence of transient

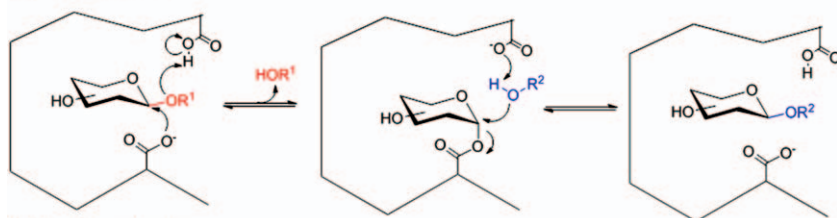
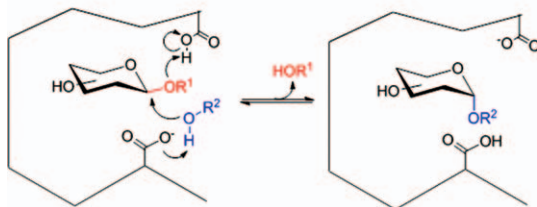
(a) Retaining mechanism**(b) Inverting mechanism**

Fig. 1 Catalytic mechanism for retaining (a) and inverting (b) glycosyltransferases (R^1 = nucleotide, R^2 = saccharide) and glycosidases (R^1 = saccharide, R^2 = H).

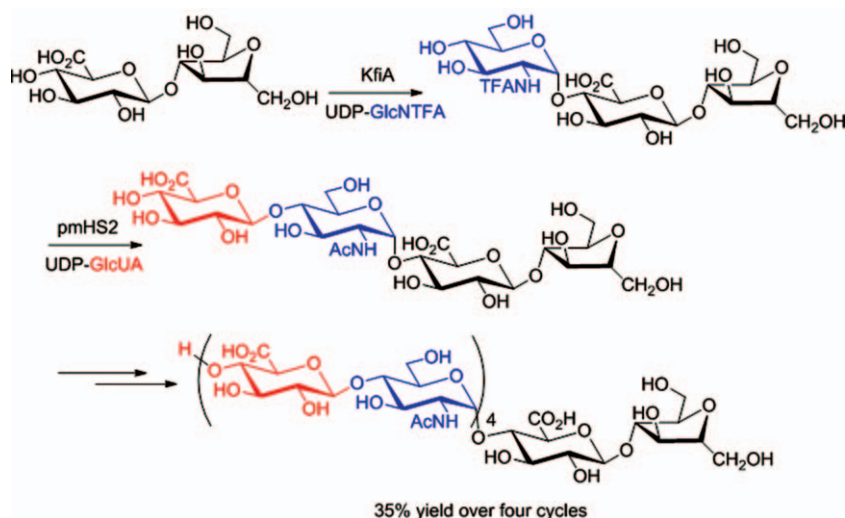
protecting groups or toxic chemicals. Such quality lies in the existence of a perfectly defined binding site that constitutes a perfect fit for a perfect match with the substrate.⁷ The tight hydrogen bond network that surrounds the substrates allows the actors of the reaction to position ideally. It results in only one product. Consequently, both GH and GT play the roles of catalyst as well as transient protecting group. Their unique properties have attracted much effort to bring them into a chemistry laboratory in order to perform regioselective and stereoselective glycosylations. In the followings examples, such processes are described.

Glycosyltransferases can be classified as Leloir or non-Leloir depending on the nature of the donor. Most known GTs use a nucleotide-sugar as the source of carbohydrate and are named Leloir. Others that use glycosides phosphate or disaccharides like sucrose (sucrose synthase) are categorized as non-Leloir. They catalyzed the glycosylation reaction with an exceptional regio- and stereoselectivity. This is linked with the strong interaction of the binding site with both the donor and the acceptor. The selectivity is so high that GTs only use a specific pair of sugar and nucleotide. The acceptor approaches the glycosyl donor in a defined orientation that allows only one hydroxyl group to react, namely the one nearest to the catalytic acid/base amino acid. The others are buried into the protein and are involved in H-bonds with amino acids of the binding site. Subtle modifications of the microenvironnement during the reaction also drive the reaction towards one regioisomer.

Synthetic strategies that rely on glycosyltransferase to catalyze glycosylation are still hampered by the low availability of the catalyst, the cost of the nucleotide-diphospho-sugars, and the necessity to recycle or destroy the resulting nucleotide. The first transferases isolated were from plants then from mammals.⁸ They are often membrane bound and their large scale production is somehow tricky. The discovery of

glycosyltransferases from bacterial origin has removed this obstacle and different strategies involving bacterial GTs have been developed in the last decades.⁹ These strategies were particularly useful to obtain glycosidic linkages often difficult to synthesize by chemical ways. For example, α -sialyltransferases from bacterial origin were used in order to add sialic acid on linear and branch oligosaccharides.¹⁰ Also the α -(1 $\rightarrow$ 2)-fucosyltransferase from *Helicobacter pylori* catalyzed the selective transfer of fucose from guanosine-5'-diphospho- β -L-fucose to various oligosaccharides acceptors to provide valuable mimics of Lewis^y oligosaccharide.¹¹ The resulting oligomers played a crucial role in many biological processes like cell adhesion or bacterial invasion. As GTs are highly specific, they can perfectly be used in tandem reactions. For example, an oligomer of heparin made of eight units was built from a tagged disaccharide by the combined action of *N*-acetylglucosaminyl transferase from *E. coli* K5 and heparosan synthase-2 from *Pasteurella multocida* (Scheme 1).¹² The substrate of the first enzyme was an UDP-GlcNTFA where the glucosamine was protected as a trifluoroacetamide. Surprisingly, such unnatural substrate was an excellent donor for the transferase. There is therefore some flexibility in the substrate accommodation by GTs. Such alternative substrate recognition was also described with β -(1 $\rightarrow$ 4)-galactosyltransferase from *Helicobacter pylori*, but this time on the acceptor side. Indeed the enzyme successfully performed thioglycosylation between UDP-Gal and *p*-nitrophenyl (*p*NP) 4-S- β -GlcNAc.¹³

Finally Chen and coworkers have developed a clever system to overcome the low availability and stability of the nucleotide-sugar.¹⁴ The one-pot multienzyme system groups a glycosyltransferase and a sugar nucleotide generating enzyme in one block. The corresponding blocks work independently starting from free sugar as donor precursor and are dedicated to the specific transfer of one glycosyl entity to the growing



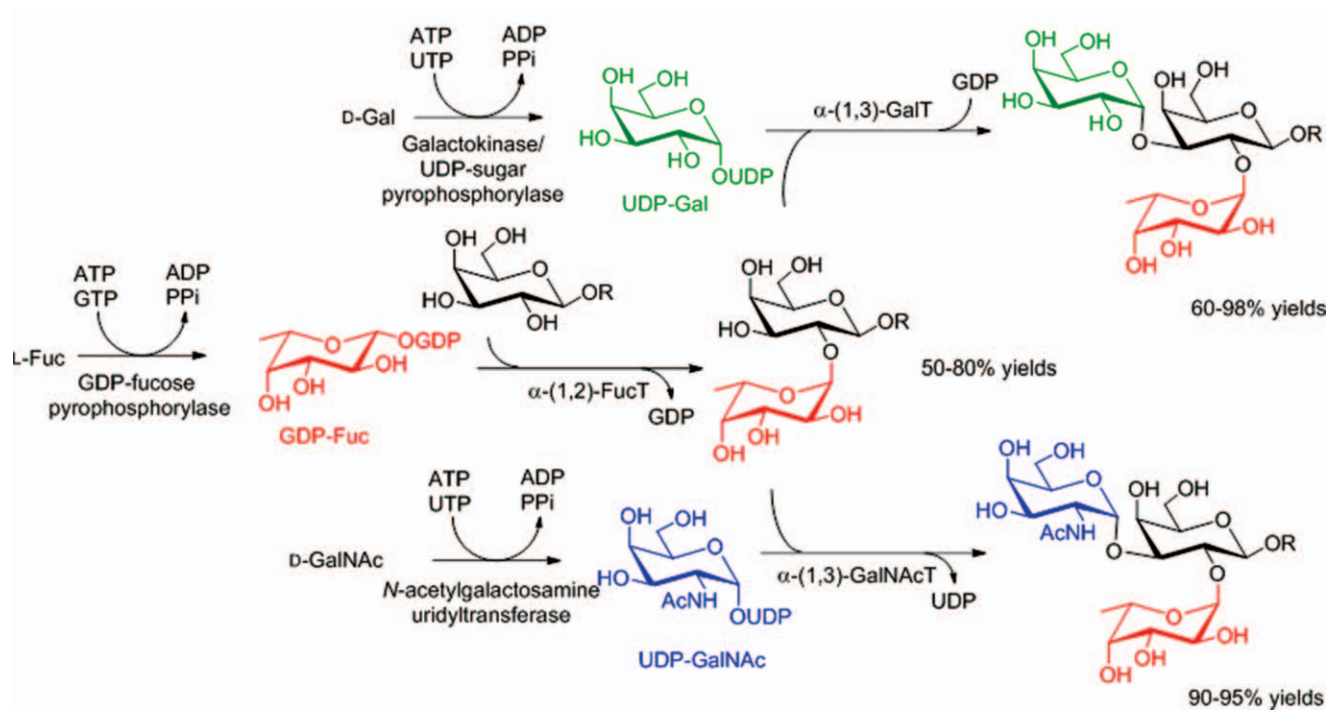
Scheme 1 Synthesis of heparin oligomer by the tandem use of *N*-acetylglucosaminyl transferase (KfiA) and heparosan synthase-2 (pmHS2).

chain. This strategy has been successfully applied for the enzymatic synthesis of all 15 naturally occurring ABH antigens¹⁵ (Scheme 2) and Lewis^x oligosaccharides.¹⁶ Alternative generation or regeneration of UDP-sugar can also be implemented by the combination of a GT with a sucrose synthase.¹⁷ This enzyme, classified as a glycosyltransferase, catalyzes the reversible formation of NDP-glucose from sucrose and a nucleoside diphosphate. It can also be used for the regioselective synthesis of disaccharides.

Even if most obstacles are now overcome when dealing with GTs as catalyst for glycosylation, some drawbacks remain, mostly related to the lack of flexibility and stability of the protein. This is why numerous research groups prefer to work with glycosylhydrolases that are much more robust and easy to produce. The choice of GHs to perform glycosylation might appear inappropriate as such enzymes rather catalyze the reverse reaction meaning the cleavage of the glycosidic bond. Nevertheless, if the reaction is performed in the presence of an excess of acceptor, a competition with water occurs and some disaccharides can be formed. This process is named transglycosylation.

Two approaches were investigated for the development of transglycosylation (Fig. 2). On one side, the thermodynamic strategy relies on the saturation of the active site by the acceptor and is generally associated with the removal of water. In this approach, the released aglycon (water or saccharide) can compete with the acceptor for the glycosylation. On the other side, the kinetic approach uses activated sugar like *p*-nitrophenyl glycosides, glycosyl fluorides or glycosyl azides¹⁸ as donors. When the glycosyl-enzyme intermediate is formed, the leaving group is released. As it possesses very poor nucleophilicity, the reverse reaction is impossible and the accumulation of glycosyl-enzyme occurs. Attack of the acceptor releases the resulting disaccharide. This strategy however faces the competitive addition of the acceptor or water. In both cases, the product formed is also a possible substrate of the reaction and can be hydrolyzed. Consequently, the yields of transglycosylation are often low. Only few GHs showed natural transglycosylase activity. One can mention, for example, cyclodextrin glycosyltransferases that are starch degrading enzymes,^{19,20} trypanosomal *trans*-sialidase with high activity for the transfer of sialic acid²¹ or xyloglucan endotransglycosylases that play key role in plant cell wall growth.²²

Like GTs, the specificity of substrate and the regioselectivity of the hydrolysis reaction catalyzed by the native GH often dictate the selectivity of the transfer in transglycosylation reaction. GHs show a rather tight binding in the negative subsite, the site occupied in transglycosylation by the donor, but a rather loose H-bond network in the positive sub-site, so with the potential acceptor. This is an advantage when one wants to glycosylate a large panel of acceptors. However, this is detrimental to the regioselectivity of the reaction. For example, the α -L-arabinofuranosidase 51 from *Clostridium thermocellum* is a thermostable enzyme able to hydrolyse α -(1 $\rightarrow$ 3), α -(1 $\rightarrow$ 2) and α -(1 $\rightarrow$ 5)-linked arabinoside of arabinoxylan with the same efficiency. When used in transglycosylation condition with *p*NP α -L-arabinofuranoside (Araf) as both donor and acceptor,



Scheme 2 Modular assembly of ABH antigens using one-pot multienzyme system.

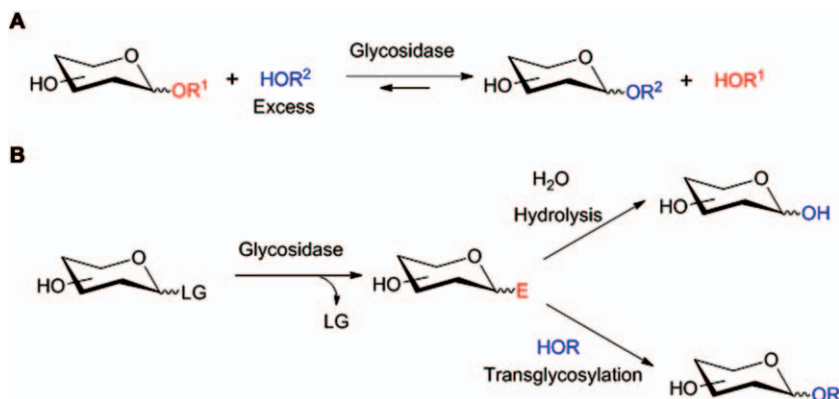
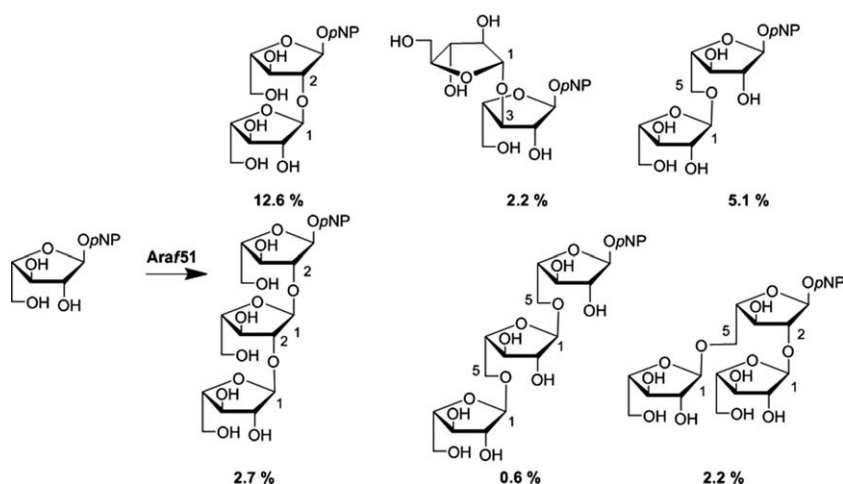


Fig. 2 Transglycosylation approaches. A. Thermodynamic control ($R^2 = \text{H}$ or saccharide). B. Kinetic control (LG = pNP, F or N_3).



Scheme 3 CtAraf51 in organic synthesis: synthesis of oligoarabinofuranosides.

different regioisomers were isolated with the pNP $\alpha\text{-Araf}^2\text{-(1}\rightarrow\text{2)-}\alpha\text{-Araf}$ as the major one (Scheme 3). The $\alpha\text{-(1}\rightarrow\text{2)}$ linkage rapidly hydrolyzed to provide the $\alpha\text{-(1}\rightarrow\text{3)}$ and the $\alpha\text{-(1}\rightarrow\text{5)}$ isomers that are thermodynamically favored.²³ Such kinetic observations emphasize the fact that, often, information on the hydrolytic substrate specificity of glycosidases can be very valuable in order to predict the regioselectivity of transglycosylation reactions.²⁴

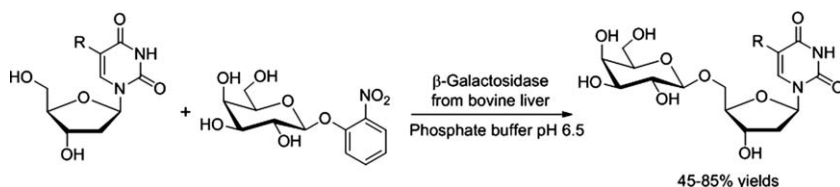
An effort was also made recently to identify the factors of such broad regioselectivity using a thermodynamic and a kinetic approach.²⁵ The authors postulate that one determinant factor of regioselectivity is the thermodynamic stability of the glycosidic bond that is broken or formed.²⁶ They determined the rate of hydrolysis of pNP galactobiosides in presence of the α -galactosidase from *T. maritima*. The calculated free energy for formation and hydrolysis of individual glycosidic bond highlighted a higher stability for the $\alpha\text{-(1}\rightarrow\text{6)-}$ linkage over the $\alpha\text{-(1}\rightarrow\text{3)-}$ and the $\alpha\text{-(1}\rightarrow\text{2)-}$ ones. The $\alpha\text{-(1}\rightarrow\text{4)-}$ linkage was completely disfavored as

confirmed by docking calculation. Indeed, the docked α -Galp-(1 $\rightarrow$ 4)-Galp showed a geometric clash between the reducing galactose and the +1 binding site. These results were in accordance with the natural distribution of galactobiosides.

From these studies, complete regioselectivity will be favored in transglycosylation when using glycosidases with strict regioselectivity. Such enzymes exist when looking at the Carbohydrate Active Enzyme (CAZY) database²⁷ and some of them have been extensively exploited for transglycosylation purpose as exemplified in recent reviews.^{28–31} Recent examples include the 5'-regioselective galactosylation of deoxynucleoside catalyzed by a commercially available β -galactosidase from bovine liver (Scheme 4).³²

A recent breakthrough in the development of transglycosylation is associated with glycoprotein remodeling. Glycosylation is a common post-translational modification of eukaryotic protein. It often leads to a mixture of glycoforms. Glycobiologists are often keen on dealing with homogeneous glycoproteins for analysis purpose but also to determine the influence of the glycosylation profile on the protein activity. The endo- β -N-acetylglucosaminidases (ENGases) are glycosylhydrolases that cleave selectively the β -GlcNAc-(1 $\rightarrow$ 4)-GlcNAc linkage found in most N-glycoprotein. Interestingly, their mechanism of action differs from the classic double displacement established by Koshland. They rely on the neighboring participation of the acetamide at position 2 to promote the bond cleavage (Fig. 3A).³³ Rather than a glycosyl-enzyme, an oxazoline constitutes the reactional intermediate. In 2001, the group of Shoda managed to perform a transglycosylation reaction using such oxazoline as donor and an oligosaccharide bearing a GlcNAc at the non-reducing end.³⁴ The reaction was highly stereo- and regioselective for the formation of β -(1 $\rightarrow$ 4)-linkages. From there, several microbial endohexosaminidases were explored in order to increase the library of recognized substrates.³⁵ A recent example includes the endoglycosidase from *Flavobacterium meningosepticum* (Endo-F) that was able to glycosylate the O-6 fucosylated GlcNAc (Fig. 3B).³⁶ This powerful methodology was successfully applied for the remodeling of antibodies, synthetic peptides and so on.³⁷

The last available biocatalyst family to perform regioselective glycosylation is the glycoside phosphorylase (GPs). They have been categorized either as GTs or as GHs depending on their sequence. Their mechanism of action however is mostly related to the one of glycosidases. Glycoside phosphorylases catalyze the formation of sugar 1-phosphate from a



Scheme 4 Regioselective glycosylation of nucleoside catalyzed by a β -galactosidase from bovine liver.

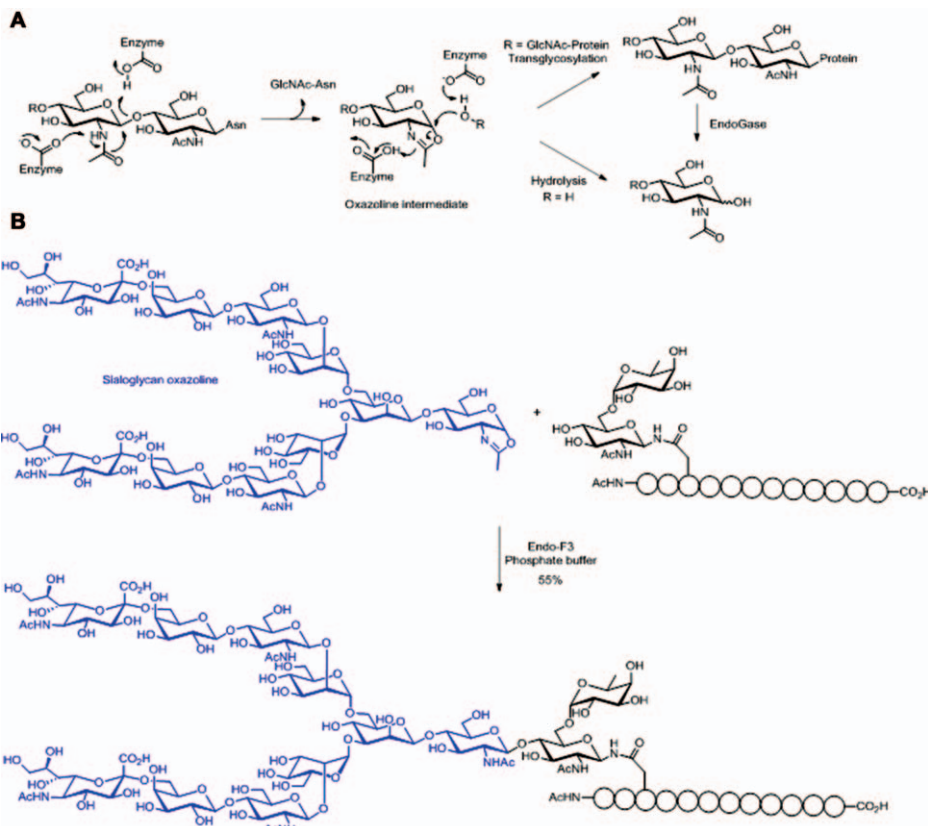


Fig. 3 (A) Mechanism of hydrolysis and transglycosylation catalyzed by endoglucosaminidases. (B) Chemoenzymatic synthesis of sialylated fucosylated CD52 antigen.

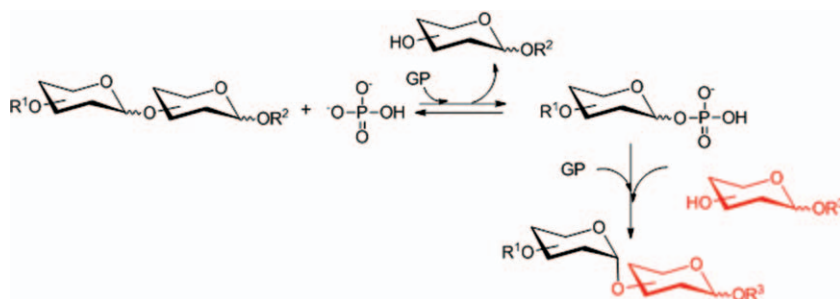


Fig. 4 Reaction catalyzed by glycoside phosphorylases.

disaccharide. They can work according to retaining or inverting mechanisms where inorganic phosphate plays the role of water (Fig. 4).³⁸ Contrary to GTs and GHs, the reversible reaction of disaccharide formation happens naturally because of the high energy of the sugar-phosphate bond. The GPs like the GTs have strong substrate specificity due to the tight binding in the catalytic pocket of both the donor and the acceptor. They also proceed with high regiospecificity. Modification of

the acceptor is allowed only if such modification occurs at a position that is far away from the catalytic site or if new interaction with the protein occurs upon binding to the active site. GPs have the advantage to use cheap disaccharides like sucrose or maltose as ligands and efficiently provide the corresponding phosphate. If a proper acceptor is introduced in the reaction media, transfer of the sugar to this acceptor gives a new disaccharide with added value. Even if such process is at the moment less popular than the utilization of GTs or GHs, its applications are in constant growth. The only drawback is the low yield of transglycosylation. At best 35% conversion into the new disaccharide can be obtained (except for sucrose phosphorylase where conversion yield reach 70%) because of the lack of versatility of such processing enzyme for other sugars than glucose at the reducing end.³⁹ Nevertheless, sucrose phosphorylase was successfully employed for the synthesis of α -glucosides,⁴⁰ and cellobiose phosphorylase gave excellent catalytic properties for the transfer of glucose to various monosaccharides.⁴¹

Finally, efforts were also made in order to provide an efficient and selective synthesis of prebiotics like the FructoOligoSaccharides using either β -fructofuranosidases or fructosyltransferase. For example, levansucrase from *Bacillus subtilis* selectively transferred fructose to the OH-6 of trehalose.⁴² The final action of trehalase gave blastose, a disaccharide with bifido-stimulating effect, with high yield. Also β -fructofuranosidases from *Xanthophyllomyces dendrorhous* showed a unique activity on disaccharide acceptors like maltose to give selectively neo-erlose.⁴³

2.2 Tailoring of the biocatalysts: a driving force for a better regioselectivity?

The development of *in silico* modeling coupled with protein engineering has recently opened a new frontier for the improvement of biocatalyzed glycosylation reaction. Nevertheless, such techniques mainly focus on the improvement of the specificity and the performance of the enzyme rather than the regioselectivity.⁴⁴ A striking example is the engineering of glucansucrase, a specific enzyme that catalyzes transglycosylation to various substrates from sucrose.⁴⁵ Owing to the high plasticity of the +1 subsite of this enzyme, different acceptors could be glycosylated. Random mutagenesis as well as rational design were applied to increase both the versatility and the efficiency of the enzyme.^{46,47} No less than seven mutations in the first shell of coordination were, for example, introduced on an amylsucrase to perform the regioselective glycosylation of rhamnose, a non-natural substrate.

The major contribution for the improvement of transglycosylation strategies through protein engineering was reported by Withers and colleagues in 1998.⁴⁸ They performed a single mutation of the nucleophilic residue of a retaining glycosidase in order to switch off its hydrolytic properties (Fig. 5). As no glycosyl-enzyme is formed with this now called glycosynthase, it is compulsory to use an activated glycosyl fluoride with the opposite anomeric configuration as donor. After attack of the

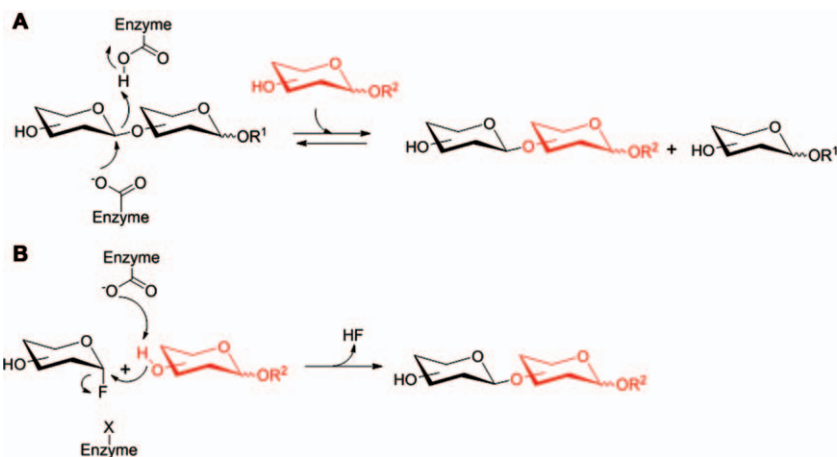


Fig. 5 Comparative mechanism of transglycosylation catalyzed by retaining glycosidase (A) and glycosynthase (B). X represents a non-nucleophilic amino acid.

acceptor, the resulting disaccharide is not hydrolyzed by the enzyme and therefore accumulates in the reaction media. Yields of transglycosylation increase greatly using this technology. Glycosyl azides were also described as potential donors for the reaction.^{49,50} This methodology was subsequently extended to inverting glycosidase, endoglycosidase^{37,51} as well as glycoside phosphorylase.⁴¹ In the case of inverting glycosidases, mutation of the nucleophilic amino-acid does not prevent the hydrolysis of the product of the reaction. The glycosynthase activity was however possible thanks to the mutation of the amino acid involved in hydrogen bonding with the water molecule involved in the reaction.^{52,53} Since this discovery, many disaccharides were obtained using glycosynthase methodology.^{44,54}

The contribution of glycosynthase to regioselective glycosylation still remains limited. Indeed, only the catalytic activity of the enzyme is changed and not the binding properties. Nevertheless, it is important to remember that glycosidases, even if they often lack regioselectivity, cleaved preferentially one bond before the others. This is reflected in the reverse reaction by the preferential formation of the kinetically favored disaccharides. These products are then recycled by the glycosidase to give the thermodynamically favored disaccharides. As glycosynthases do not show any hydrolytic activity, the kinetic product accumulates and no recycling occurs. For example, the GH35 β -galactosidase from *Bacillus circulans* was first used in transglycosylation of various acceptors by pNP-Gal donor.⁵⁵ A fluctuating regioselectivity was obtained according to the nature of the acceptor. In particular β -(1 $\rightarrow$ 3)-bond with GlcNAc and GalNAc could be obtained but the resulting disaccharide was rapidly hydrolyzed. Recently, Li and Kim reported the selective formation of galactosyl- β -(1 $\rightarrow$ 3)-linkages catalyzed by galactosynthase derived from such β -galactosidase with yield up to 98%.⁵⁶ They hypothesized that the β -(1 $\rightarrow$ 3)-bond was formed first and as no hydrolysis could happen, no other isomer could be synthesized.

Mutation of key amino acids of the binding site could also favor the regiospecificity of glycosidase or glycosynthase. Regioselectivity in reaction catalyzed by retaining GTs or GHs is governed by two factors, the distance from the acid/base catalytic residue and the nucleophilic hydroxyl group on the acceptor, and the distance between the same alcohol and the anomeric carbon of the glycosyl-enzyme intermediate.⁵⁷ Identification of crucial amino acids in the binding site that control the positioning of the acceptor and therefore the resulting distances allowed modulating the regioselectivity of the reaction by subsequent mutations. For example, three mutations of an α -L-fucosidases from *T. maritima* were necessary in order to increase the regioselectivity for pNP α -L-Fuc-(1 $\rightarrow$ 2)- β -D-Gal from 7% to 62%.⁵⁸ Another example was reported by Blanchard *et al.* who performed some mutations on the binding site of the glycosynthase from the endocellulase Cel7B. The resulting mutants exhibited a drastic change of regiospecificity.⁵⁹ While the single and triple mutant allowed the selective formation of β -(1 $\rightarrow$ 3)-linked tetrasaccharide, the double mutant favored the β -(1 $\rightarrow$ 4)-bond between two disaccharides. The reshaping of the +1 sub-site appeared to be the driving force for the regioselectivity through the reorientation of the acceptor. Finally, the regiospecificity of GH-65 maltose phosphorylase was also modulated thanks to the mutation of a sequence of amino acid on the loop next to the catalytic site. Sequence alignment with trehalose phosphorylase and kojibiose phosphorylase showed that this loop is particularly crucial for the substrate recognition. Swapping such sequence by the one found, for example, in trehalose phosphorylase modifies the regioselectivity of the reverse phosphorolysis from α -(1 $\rightarrow$ 4) to α -(1 $\rightarrow$ 1).⁶⁰

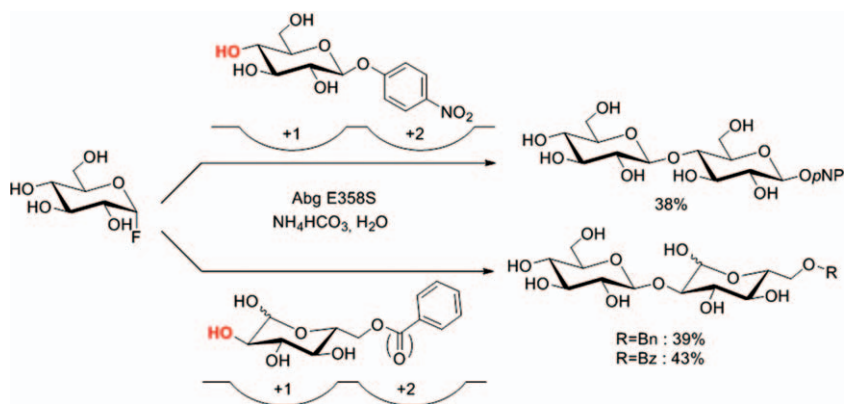
2.3 Modification of the experimental parameters to modulate the regioselectivity

As exemplified above, regioselectivity in the transfer of one glycoside to the other is often a result of the choice of the enzyme. Mutagenesis can modulate the selectivity but most research focused rather on the improvement of the versatility of the enzyme. Nevertheless, other factors like the nature of the acceptor, the temperature or the solvent can also positively or negatively influence the outcome of the reaction. This is particularly true when dealing with glycosidases and glycosynthases.

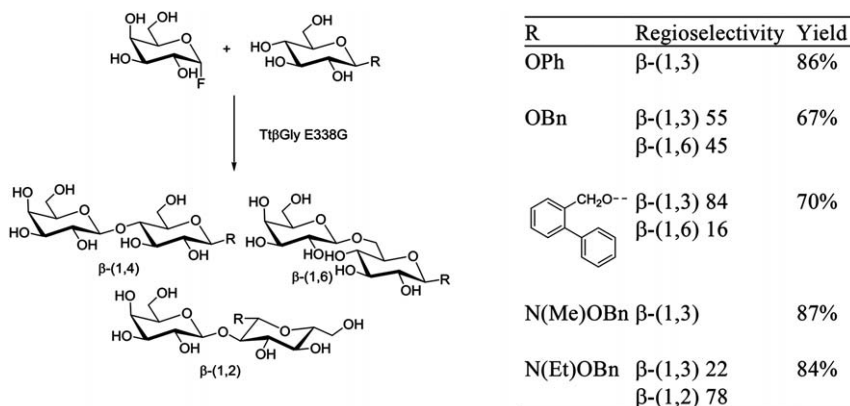
The nature of the anomeric configuration as well as the substitution on the acceptor is crucial for the binding to the positive sub-sites. In particular, it can generate different conformation of the acceptor in the binding site and the hydroxyl group nearest to the catalytic amino acid can vary. For example the presence or absence of an *N*-acetyl group can influence the regioselectivity of β -galactosidase⁶¹ or sialidases.⁶² Stick *et al.* showed that the introduction of a benzyl or a benzoyl group on position *O*-6 of the glucopyranose acceptor swapped the regioselectivity induced by a β -glucosidase mutated as GS from β -(1 $\rightarrow$ 4) to β -(1 $\rightarrow$ 2).⁶³ This is the consequence of an alternating binding mode of the acceptor where the aromatic ring was accommodated in the +2 sub-site thus

changing the orientation of the acceptor (Scheme 5). Trincone and coworkers also reported the influence of the aglycon on the anomeric position. Indeed the change from 2-*p*NP- β -D-Glc to 4-*p*NP- β -D-Glc as acceptor of the hyperthermophilic glycosynthase from *S. solfataricus* changed the regioselectivity from β -(1 $\rightarrow$ 3) to β -(1 $\rightarrow$ 6).⁶⁴ The strict regiospecificity of GS from *Streptomyces* for β -(1 $\rightarrow$ 3) linkages was as well greatly altered when using aryl disaccharides instead of monosaccharides.⁶⁵ In this example, the presence of a glycosyl unit in subsite +2 switched the regioselectivity from β -(1 $\rightarrow$ 3) to β -(1 $\rightarrow$ 4) because of the possibility of an alternative mode of binding.

More recently Tellier's group showed that substrate engineering can give as good results as protein engineering to enhance the regioselectivity of glycosynthases. They used a mutated β -glucosidase from *Thermus thermophilus* as glycosynthase. Transglycosylation reaction performed with phenyl glycoside as acceptor lead to disaccharides with β -(1 $\rightarrow$ 3) selectivity (Scheme 6).⁶⁶ However, the introduction of a more flexible benzyl group instead of the phenyl group decreased the regioselectivity.



Scheme 5 Modulation of the regioselectivity of glucosylation catalyzed by *Agrobacterium* sp. β -glucosidase according to the nature of the O-6 substituent.



Scheme 6 Reaction between glucosyl fluoride and various glucoside acceptors catalyzed by E338G mutant glycosidase from *Thermus thermophilus*.

A new aglycon, the 2-biphenylmethane, was then chosen from *in silico* calculations. According to those, such aryl group fitted perfectly the binding site of the enzyme thanks to a stabilizing stacking with a tryptophan. Gratifyingly, transglycosylation results with the corresponding acceptor showed a restored regioselectivity.⁶⁷ Another aryl moiety, the *N*-alkyl-*O*-benzylhydroxyamine was then introduced on the glucoside acceptor. This moiety has the advantage to be easily cleavable. Interestingly, a methyl substitution on the alkoxyamino group did not change the regioselectivity of the reaction while bulkier ethyl substituent favoured the transglycosylation on *O*-2.⁶⁸

Other parameters were screened to evaluate their influence on the regioselectivity of the glycosylation catalyzed by glycoside-processing enzyme. The ratio acceptor/donor can be a factor, but with unpredictable results. An excess of acceptor can sometimes improve the regioselectivity of the transglycosylation or on the contrary, generate a drop of selectivity.^{69,70} An increase of temperature can also be detrimental to the yield and the regioselectivity.⁷¹ For example, Elling *et al.* have implemented a transgalactosylation with a thermostable β -glycosidase under micro-wave irradiation.⁷² The reaction proceeded far below its temperature optimum in order to avoid any thermal effect. Thanks to the use of the micro-wave, the time of the reaction was shortened and the amount of hydrolysis was low. It favored the transglycosylation towards the kinetic product bearing a β -(1 $\rightarrow$ 4)-linkage.

One of the main parameters that have been proven to influence regioselectivity is the presence of organic solvents. The addition of solvent was initially devised in order to reduce the amount of hydrolysis in transglycosylation conditions. Nevertheless, the action of β -galactosidases for example in presence of co-solvent can be significantly improved in term of regioselectivity.^{73–75} Recent developments with bio-solvents were reported by Hernáiz and coworkers.^{76–78} The regioselectivity of transglycosylation catalyzed by biolacta β -galactase was switched from β -(1 $\rightarrow$ 4) to β -(1 $\rightarrow$ 6) when a high concentration of glycerol based solvents was added to the reaction media. This modulation was explained by measuring the distance between the anomeric carbon *C*-1' of the glycosyl-enzyme intermediate and the *O*-4 or the *O*-6 of the GlcNAc acceptor. Differences were found in presence or absence of glycerol in the calculation cells. In pure water, the acceptor binds the enzyme in a deep mode because of the hydrophobic clash with the molecules of water that surround the binding site. In this case, the distance *C*-1'/*O*-4 is smaller than the *C*-1'/*O*-6 one. In presence of glycerol, interaction of the co-solvent occurred with both the enzyme and the acceptor that adopted here a shallow mode of interaction. The *O*-4 was involved in an H-bond with the bio-solvent and the *C*-1'/*O*-6 distance is the smallest.

Finally, a recent report from Castillo and co-workers highlighted the influence of the acceptor anomerism on the regiospecificity of cyclodextrin glycosyltransferase.⁷⁹ Using *p*NP β -Glucp, a mixture of α -(1 $\rightarrow$ 4)-, α -(1 $\rightarrow$ 3)- and α -(1 $\rightarrow$ 6)-linkage was obtained. However, in the presence of the α -anomer, a complete regioselectivity for α -(1 $\rightarrow$ 4)-linkage was found. Similar results were described when switching from the pyranose to the

furanose form. While a mixture of maltoside and isomaltoside was obtained using *p*-NP Glucp as acceptor, only one regioisomer, namely the *p*-NP α -D-Glcp-(1 $\rightarrow$ 3)- α -L-Araf was isolated using *p*-NP arabinofuranoside as acceptor.⁸⁰

3 Bioinspired chemical glycosylation

3.1 Prerequisites

Since the synthesis of a disaccharide involves a glycosyl donor and a glycosyl acceptor, which are both chiral reagents, it should be kept in mind that succeeding in glycosylation reactions also depends on double diastereodifferentiation, *i.e.* match-mismatch effect or reciprocal donor-acceptor reactivity, and not only on electrophilicity of the donor and nucleophilicity of the acceptor.⁸¹

As a result, desired selectivity, related to protection steps and glycosylation reactions, requires a basic knowledge of the main parameters that could finely affect reactivity of hydroxyl functions and induce differences. The first approaches aimed to single out one specific hydroxyl by a set of orthogonal protecting groups. Although the resulting strategies are time-consuming, they have proved to be highly efficient, so efficient that they are still used for synthesizing unambiguously structurally well-defined oligosaccharides.^{82–84} Amongst other parameters is the nature of the hydroxyl functions: acetalic, primary, secondary, axial or equatorial positions are different enough that more or less subtle differences may be expected. Consequently, patterns of reactivity also vary according to the nature of the explored carbohydrate.

In order to estimate the nucleophilicity of hydroxyl functions in various hexopyranosides, Bols and coworkers have synthesized monoamino-deoxy-glycosides and measured their pK_a .⁸⁵ They hypothesized that increasing pK_a values should correspond to higher electron density on the oxygen atom. Based on this assumption, some important trends have been drawn: (i) the primary hydroxyl is always the most nucleophilic, and also benefits from the least steric hindrance, (ii) antiperiplanar relationships are very important. For instance, in *gluco* and *manno* series (Fig. 6), the 4-amino group and the endocyclic oxygen O-5 are indeed fixed antiperiplanar, that increases electron-withdrawing ability of O-1 atom, and so mirrors lowering nucleophilicity of the 4-OH. This antiperiplanar effect also explains why the pK_a of an amine increases when an axial substituent is present.

As a magnificent application of the control of the hydroxyl reactivity was the regioselective one-pot protection of monosaccharides performed

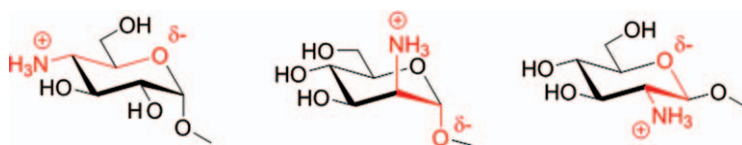


Fig. 6 Antiperiplanar impacts charge density on ammonium groups.

by Hung and collaborators.⁸⁶ They could obtain fully protected building blocks accordingly avoiding in-between purification steps. Thanks to the orthogonal properties of the protecting groups, further intermediates with one free hydroxyl for regioselective glycosylation could be synthesized (Scheme 7).

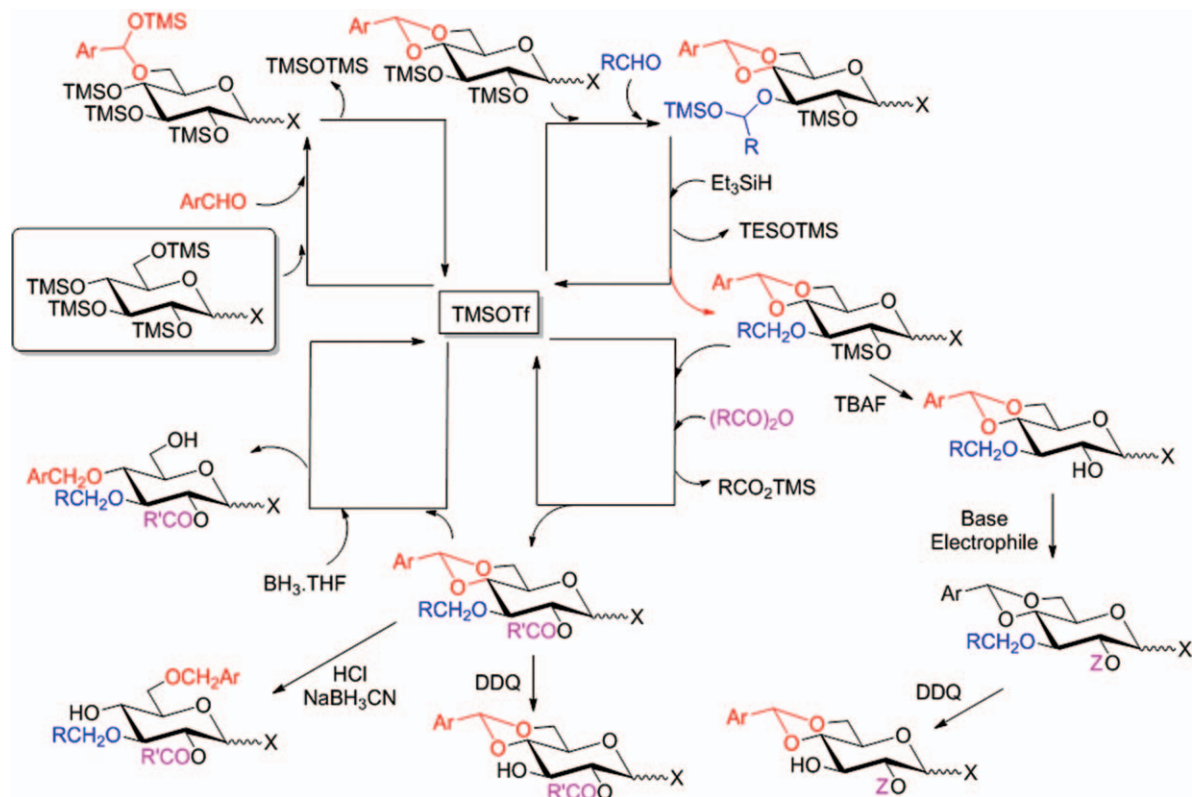
3.2 Hydrogen bonding impact and regioselective hydroxyl protections

It has been known for a long time that intramolecular hydrogen bonding may result in a network which could affect reactivity of free hydroxyls.^{87,88} This has been recently reviewed by Moitessier⁸⁹ and by Minnaard.⁹⁰ It is clear that intramolecular hydrogen bond networks depend on the carbohydrate series, solvent, temperature, anomeric centre configuration or the nature of protecting groups already present on the monosaccharide. It was also shown that the amounts of electrophilic reagents may influence the final regioselectivity. The selectivity could also be impacted by the nature of the aglycon, *i.e.* *O*-glycoside *vs.* *S*-glycoside when protection was performed in the presence of freshly prepared silver(I) oxide.⁹¹

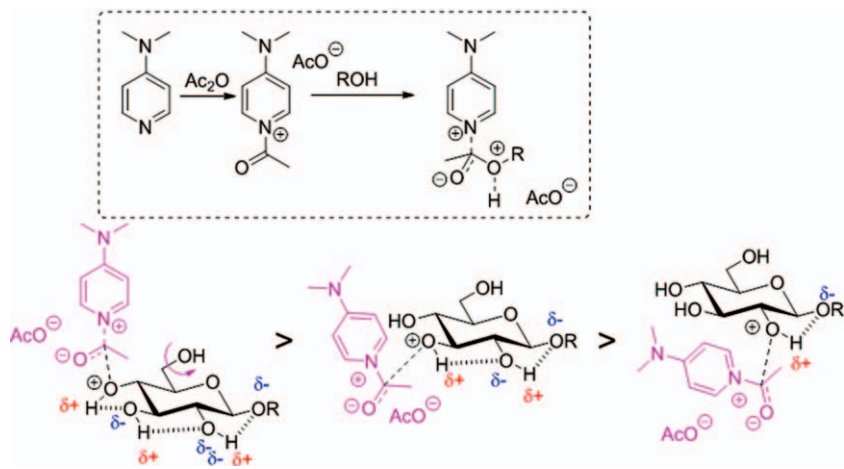
As another example, Yoshida and co-workers have performed a rigorous acetylation of alkyl α - or β -gluco-, galacto- and mannopyranosides in chloroform, using acetic anhydride as an acylating agent and DMAP as a catalyst.⁹² Amongst main results, they have demonstrated that secondary hydroxyls are more reactive than the primary function. This was explained by the fact that the rate determining step is the addition to the carbonyl function from the acyl-pyridinium intermediate and that the positive charge is delocalized through hydrogen-bonding networks (Scheme 8). The primary OH was not considered in these models on the assumption that the C6–O6 bond rotates more freely.

It was very recently proposed to substitute DMAP by (*R*)- or (*S*)-benzotetramisoles (BTM) that is able to stabilize the actual acylation species thanks to coordination of the sulfur atom and the carbonyl function (Fig. 7).⁹³ The authors were able to model and predict the regioselectivity of the target acylation on a set of mono- and disaccharides, and based this orientation on cation-*n* interactions with 1,2-*trans*-diequatorial diols.

It was also established that hydrogen bond network could be judiciously impacted thanks to acetate anions.^{94,95} The resulting dual hydrogen bonding approach allowed acetylation of the primary and the 3-hydroxyl function in *gluco* and *galacto* series. However, the axial hydroxyl groups still remained less reactive than the corresponding equatorial ones. Very recently, Schmidt and co-workers demonstrated that dual hydrogen bond could be induced by cyanide anions. Due to its basicity, this anion was suspected to activate preferentially the more acidic axial hydroxyl, provided that the primary group was previously masked and that the reaction was carried out at very low temperature. As a result, using benzoyl cyanide and DMAP, the authors were able to acylate position *O*-4 of galactosides, fucosides, and lactosides, position *O*-2 of mannosides as well as the axial hydroxyl of myoinositols (Scheme 9).⁹⁶



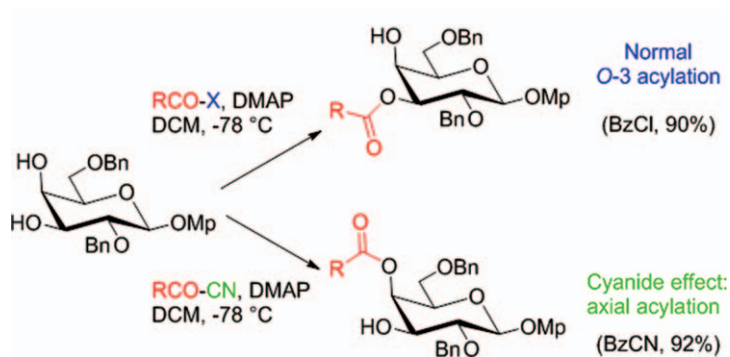
Scheme 7 One-pot regioselective protections.



Scheme 8 Proposed intermediates for DMAP-catalyzed acetylation of pyranosides.



Fig. 7 Benzotetramisoles to perform regioselective acylations.

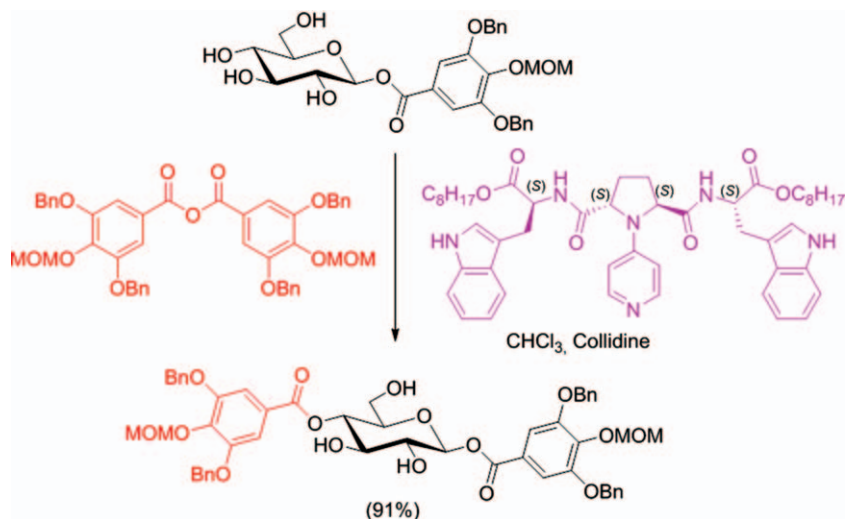


Scheme 9 Dual hydrogen bond induced by cyanide anion for the protection of axial hydroxyls.

Organocatalysis recently emerges to efficiently ensure regioselective protection of carbohydrate derivatives. This approach was pioneered by Miller and Griswold who have screened peptide libraries to acetylate monosaccharides.⁹⁷ Acetylation, using acetic anhydride as acylating agent and 6-*O*-silylated glucosamine or octyl β -D-glucopyranoside as the substrate, was significantly perturbed by the presence of tetra- to octa-peptide. The authors also demonstrated that the main monoacetylated derivatives were obtained under kinetic selection. Inspired by this

concept, Kawabata and co-workers were further able to produce very efficient lipopeptide catalysts and applied them to esterify the 4-OH of gallate β -D-glucose.⁹⁸ The resulting diester, a useful intermediate in the synthesis of ellagitannins, could be obtained with a 91% selectivity (Scheme 10).

Among famous methods proposed to single out two vicinal hydroxyl groups is also the use of organotin derivatives. Despite their recognized toxicity, which inescapably limits subsequent developments,^{89,90,99} new results are still published regularly with such catalysts. In 2013, Muramatsu and Takemoto established that the length of the alkyl chain in the dichlorodialkyl tin catalysts was responsible for the regioselectivity of thiocarbonylation, acylation, and sulfonylation (Fig. 8).¹⁰⁰ In fact, Me_2SnCl_2 was more efficient for the protection at the primary position of



Scheme 10 Peptide-based acylation of glucose derivative for the synthesis of ellagitannins.

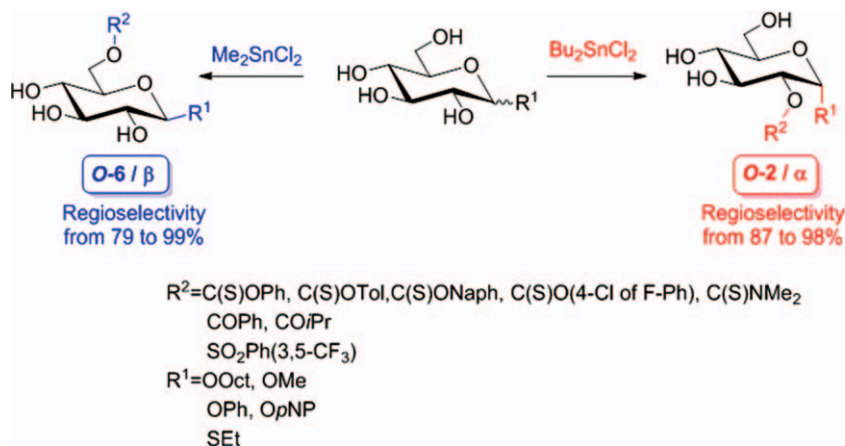
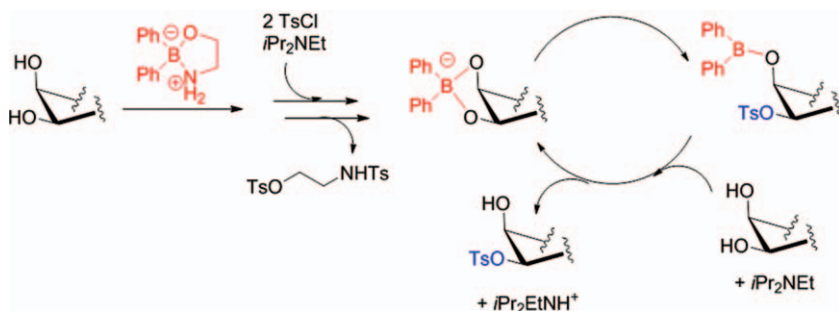


Fig. 8 Effect of organotin catalysts on regioselective protections.



Fig. 9 How valence of boron influences reactivity of hydroxyl groups.



Scheme 11 Regioselective monoprotection of monosaccharides thanks to borinic catalysts.

β -glucopyranosides, while Bu_2SnCl_2 was found to be more appropriate for the protection of the OH-2 in α -glucosides, even if the OH-6 remained free from starting material.

Parallel to these developments, the introduction of organoboron compounds in glycochemistry opened without contest new great opportunities. Very interestingly, tricoordinate boron alkoxides result in attenuated nucleophilicity of hydroxyls involved in the boronate, thus giving OH-protection strategies.^{101–105} On the other hand, the tetra-coordinate boron alkoxides display increased reactivity towards electrophiles (Fig. 9).^{106–108} The development of this approach allows regioselective monotosylation, acylation, sulfonylation and alkylation of monosaccharides bearing up to three unprotected hydroxyl functions.¹⁰⁸ The reaction requires a polyol which is activated by a borinic ester used as a pre-catalyst. The resulting true catalytic entity further reacts with the electrophilic agent and releases a monoester boron intermediate. The catalytic ring is finally completed by the transfer of the diarylboryl function to the polyol to form the borinic specie (Scheme 11).¹⁰⁸

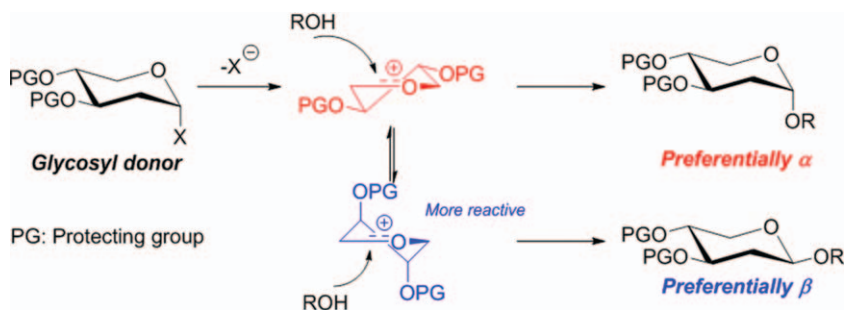
3.3 Regioselective glycosylation of partially protected acceptors

After considering selective protections, the more arduous problem of regioselective glycosylation was also studied. This chapter will first deal with the selective glycosylation depending on the glycosyl donor reactivity. Thanks to hydroxypiperidines as models to mimic glycoside hydrolysis intermediates, Bols and co-workers could explain increased reactivity of glycosyl donors with axial hydroxyl substituents.¹⁰⁹ As a result, since axial hydroxyl substituents are less electron-withdrawing,⁸⁵ a donor with more axial groups is more reactive. This is why galactosyl donors present higher relative reactivity values than glucosyl donors.

Moreover, an oxycarbenium intermediate able to adopt a conformation with axial substituents is much more reactive. This leads to the concept of super-armed donors (Scheme 12).

Reactivity is not only affected by conformational effects but also by electronic ones. To summarize numerous observations, a donor bearing electron-withdrawing protecting groups (ester for instance) are classified among disarmed donors and those with ether-type groups are among armed donors. However, the substitution of the 2-*O*-acyl group by a benzyl one in disarmed donors cancels the anchimeric assistance that helps the releasing of the leaving group, thus giving superdisarmed donors. On the contrary, substituting the 2-*O*-benzyl protection by a participating group makes easier departure of the anomeric activating group, thus giving a superarmed donor (Fig. 10).¹¹⁰

As a result, a judicious choice of donors allowed one-pot synthesis of a trisaccharide, thanks to the successive activations of the superarmed then the armed donors, but not activation of the disarmed donor, which was therefore used as an acceptor (Scheme 13).¹¹¹



Scheme 12 Impact of axial/equatorial orientation on the reactivity of the glycosyl donors.

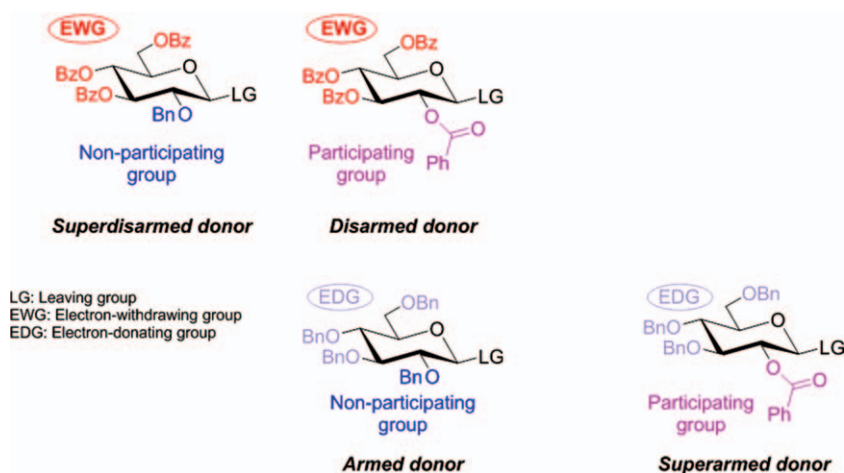
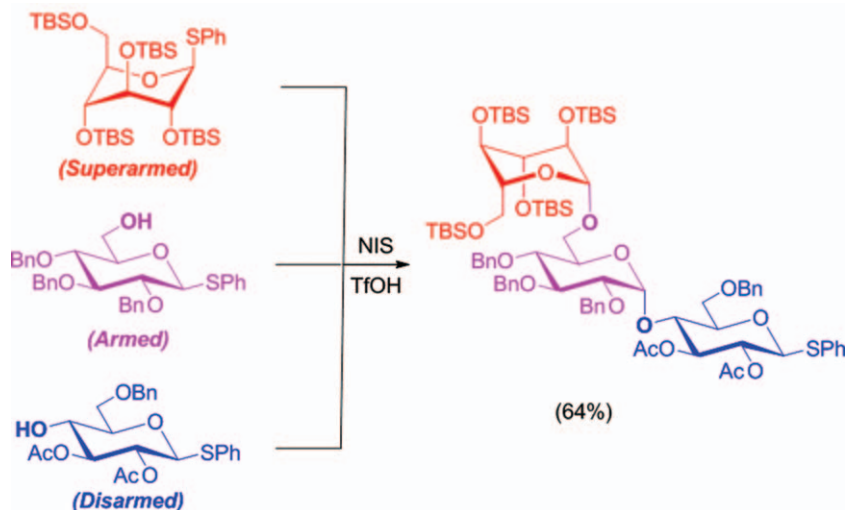
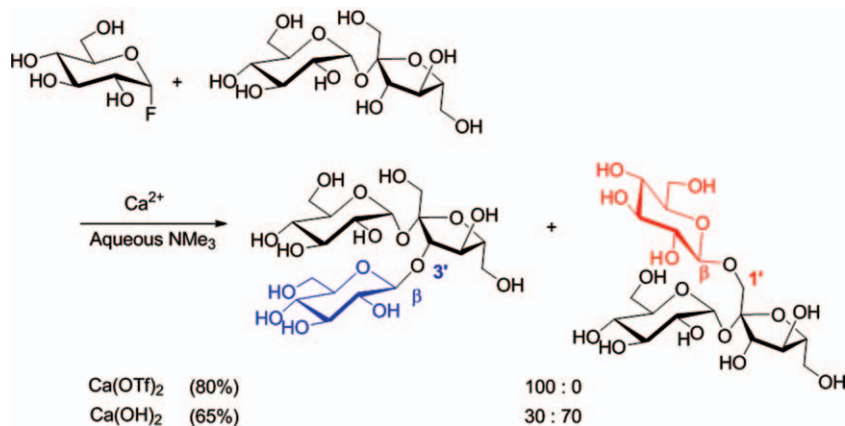


Fig. 10 From superdisarmed to superarmed donors.



Scheme 13 Selective glycosylation based on reactivity differences of donors.

This concept is of great importance to explain, or to approach rationalization of the reactivity of so many diversified glycosyl donors. In that case, the regioselectivity of the coupling is generally ensured by the standard protecting group strategy. Another challenge relies however on the use of reactants bearing less protecting groups, and so presenting several free hydroxyl functions, keeping in mind that the Grail quest for glycochemists still remains biomimicry. To approach this, first acid catalyzed glycosylation of unprotected oses afforded very complex mixtures of oligo- and polysaccharides, accompanied by colorization and many by-products. Moreover, the newly formed glycosidic linkages are sensitive under the reactions conditions. Alternatively, Suzuki *et al.* proposed to apply thermal condensation of totally unprotected glycosyl fluoride as donor to yield highly branched polysaccharides, which means that criterion of selectivity was not addressed under these conditions.¹¹² Few years later, Thiem and collaborators performed self-condensation of glycosyl fluorides but under anionic activation by lithium hydride. It first gave an alcoholate, which subsequently substituted fluoride from the donor. As a result, the authors indeed obtained homopolysaccharides characterized by the presence of 10–25 saccharidic units. Degrees of branching were sugar-dependant.¹¹³ In order to synthesize structurally well-defined disaccharides, further improvements of the base-promoted glycosylation favoured partially protected acceptors and fully protected glycosyl chlorides.¹¹⁴ Interestingly, high 1,2-*trans*-diastereoselectivities were observed. Very recently, Miller and collaborators were able to glucosylate saccharose in water, using glucosyl fluoride as a donor, in the presence of trimethyl amine and calcium cation. Under these conditions, only the fructofuranosyl residue was functionalized to afford trisaccharides (Scheme 14).¹¹⁵ More precisely, best results for 3'-glycosylation were obtained using 6 equivalents of both glucopyranosyl fluoride and calcium triflate in a

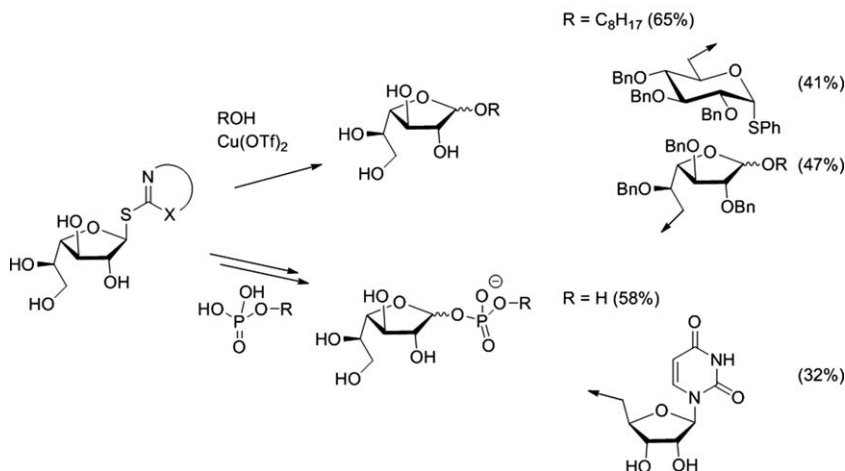


Scheme 14 Glucosylation of saccharose in aqueous solutions.

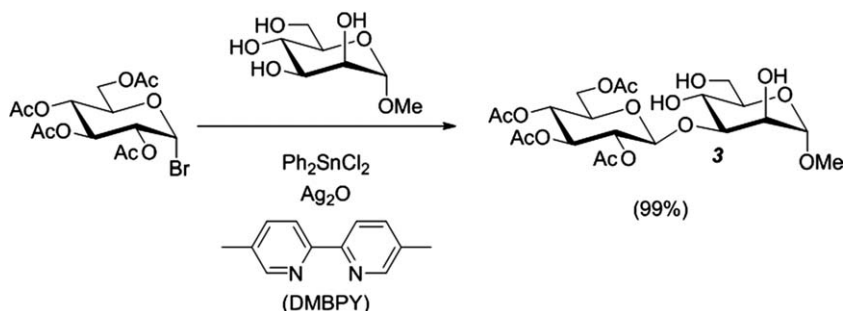
0.3 M aqueous solution of sucrose in the presence of 45% trimethylamine solution. Under these conditions, the trisaccharide was isolated in an excellent 80% yield, limiting formation of by-products such as glucose, 1,6-anhydroglucose and fructose (that resulted from glucose isomerization). Very interestingly, and under the same conditions, the reaction was efficiently extended to sucrose-like substrates as acceptors (xylosucrose, raffinose, stachyose, lactosyl fructofuranoside, erlose, galactose), and to a disaccharidyl donor (fluoromaltose). In a second phase, regioselectivity of the coupling of fluoroglucose to sucrose was performed thanks to substitution of calcium triflate by calcium hydroxide and by adding sodium chloride. Once again, the authors underlined the critical role of hydrogen-bonding network to explain their results.

Nevertheless, a very large proportion of glycochemists continue to favour the standard acid-promoted glycosylation. As already described, applying (super)arm/(super)disarm concept allowed desired glycosidic coupling without self-condensation. As an instance, a branched heptasaccharide was obtained according to one-pot glycosylation using seven independent building blocks.¹¹⁶ More rarely, fully unprotected donors were used once again without formation of oligosaccharides. Some examples were published from thioimidoyl donors for the synthesis of galactofuranose-containing disaccharides¹¹⁷ or nucleotide-furanoses.^{118,119} Interestingly, no activation of the nucleotide part was required and both natural and non-natural nucleotide-furanoses were obtained in only a few minutes (Scheme 15).

Most of the time, fully protected donors reacted with partially protected acceptors. In those cases, regioselectivity is allowed by intrinsic properties (see for examples Ref. 116, 120–122) of each free hydroxyl groups or by extrinsic orientations. Once again, organotin catalysts were efficient in controlled Koenigs-Knorr glycosylation.^{123–125} The catalyst diphenyldichlorotin was able to coordinate *cis*-diol, *i.e.* OH-2 and OH-3 in mannopyranosides (Scheme 16). This coordination increased the acidity of the hydroxyl groups, so that the weak and hindered aromatic base 5,5'-dimethyl-2,2'-bipyridyl (DMBPY) could deprotonate selectively OH-3.



Scheme 15 Unprotected thioimidoyl furanosyl donors for the synthesis of disaccharides and nucleotide sugars.

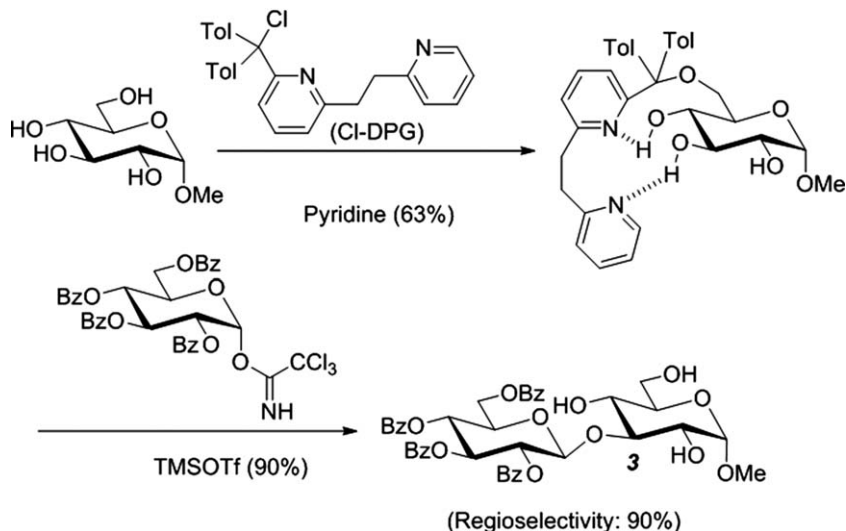


Scheme 16 Regioselective glycosylation of methyl α -D-mannopyranoside catalyzed by Ph_2SnCl_2 in the presence of Ag_2O and DMBPY.

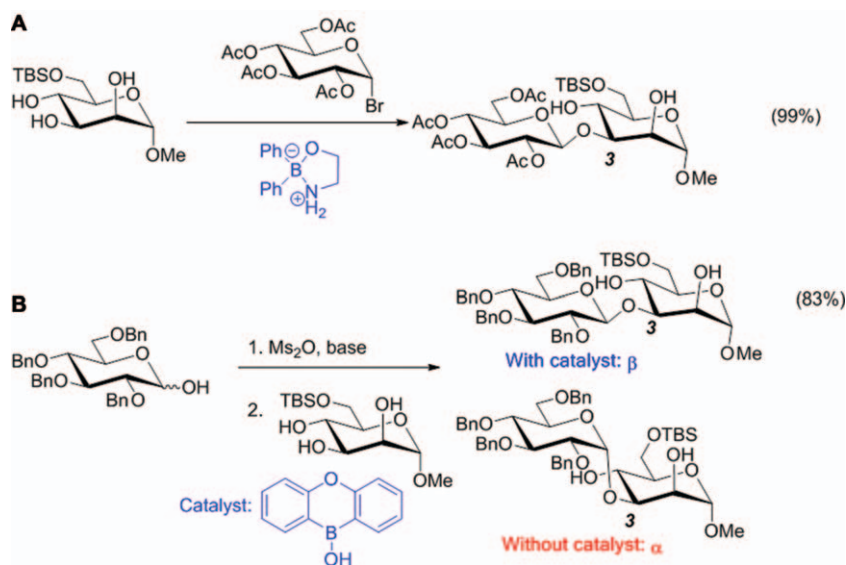
Very interestingly, this method did not require protection of the primary position.¹²³

Among other highlights is the introduction by Chapleur and Moitessier of directing-protecting groups.^{89,126} After connection of such a group to the primary position of a glucopyranoside, the presence of two pyridyl groups triggered intramolecular hydrogen bonds, which orientate glycosylation towards the 3-position, provided that a disarmed trichloroacetimidate was used (Scheme 17).

Finally, borinic acid-catalyzed glycosylation was able to induce regioselectivity of glycosidic coupling, thanks to the activation of the acceptor which however needs to possess a *cis*-diols. The desired disaccharides were obtained through abstraction of the bromide atom by halophilic Lewis acid (Scheme 18A).¹⁰⁷ More recently, this approach was extended to glycosyl methanesulfonates as donors (Scheme 18B).¹²⁷ Interestingly, the diastereoselectivity of the resulting glycosidic bond depended on the presence (β -orientation) or the absence (α -orientation) of the boron catalyst.

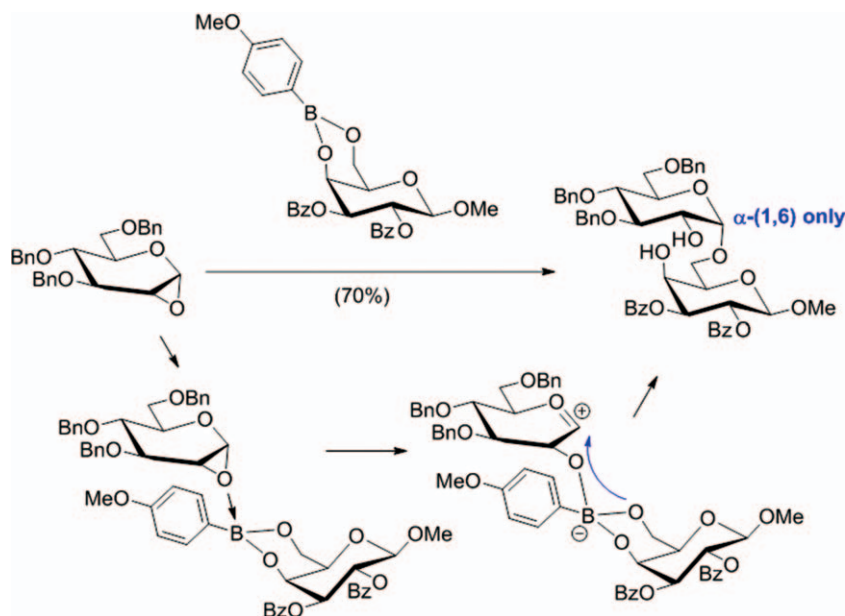


Scheme 17 The concept of directing-protecting groups.



Scheme 18 Some examples of regioselectivities under borinic acid-catalyzed glycosylation.

In 2015, Toshima and co-workers cleverly combined two great principles of glycosylation reactions: regioselective orientation thanks to boronic ester intermediates and intramolecular aglycon delivery thanks to activation of 1,2-epoxides as donors activated by the previously formed boronic species. As expected, the resulted disaccharides were characterized by new 1,2-*cis* glycosidic linkages in *gluco* and *galacto* series (Scheme 19).¹²⁸



Scheme 19 Synthesis of 1,2-*cis*-glycosides through activation of 1,2-epoxides by boronic esters.

4 Conclusion

Regioselective glycosylation relies on the subtle reactivity of the donor as well as on the subtle differentiation of the various hydroxyl groups on the acceptor. Nature managed to tackle the second problem by positioning the acceptor in an environment that favours one regioisomer and not the other. Tacking advantage of such specificity, chemists use now glycosyl-transferases or glycosylhydrolases in order to perform *O*-glycosylation. In addition, the emergence of molecular modelling and mutagenesis has boosted the versatility of such process and a large array of oligosaccharides is now available thanks to biocatalyzed glycosylation. Nevertheless, the production of carbohydrate-processing enzymes in a chemistry laboratory can still be hampered by the cost and the lack of availability of the native biocatalyst itself. This is why chemists, using the concept of hydroxyl group differentiation, have conceived and develop some purely chemical strategies to perform regioselective protections or glycosylations. For that, calculation of the nucleophilicity of each hydroxyl group was performed to gain insight into the reactivity of most carbohydrates. Different tools are now available, like the use of transient protecting groups or, on the contrary, of inducers of regioselectivity that are able to activate one hydroxyl group and not the others. Most strategies involve organic solvent and efficient glycosylation in water remains to be optimized so as to approximate the enzymatic synthesis conditions. To conclude, biomimicry is not only to get inspiration from Nature, but also to outweigh Nature. The explosion of data in genomics and metabolomics has still to be followed by time-consuming annotation. This gene

mining would certainly lead to the discovery of new enzymes, and would therefore increase the library of available biocatalysts. The emergence of a simpler model of protein, like a foldamer for example,¹²⁹ could also trigger a new era in regioselective glycosylation. For the moment, in the geometry-function duet, only the geometry was mimicked. Efforts remain to be done in order to mimic the function of a protein.

References

- 1 R. W. Hoffmann, in *Bioinspiration and Biomimicry in Chemistry*, John Wiley & Sons, Inc., 2012, pp. 419–453.
- 2 E. Lurie-Luke, *Biotechnol. Adv.*, 2014, **32**, 1494–1505.
- 3 B. Bhushan, *Philos. Trans. R. Soc., A*, 2009, **367**, 1445–1486.
- 4 R. M. Schmaltz, S. R. Hanson and C.-H. Wong, *Chem. Rev.*, 2011, **111**, 4259–4307.
- 5 A. Ardevol and C. Rovira, *J. Am. Chem. Soc.*, 2015, **137**, 7528–7547.
- 6 D. E. Koshland, *Biol. Rev.*, 1953, **28**, 416–436.
- 7 K. Hiromi, *Biochem. Biophys. Res. Commun.*, 1970, **40**, 1–6.
- 8 M. M. Palcic, *Curr. Opin. Biotechnol.*, 1999, **10**, 616–624.
- 9 D.-M. Liang, J.-H. Liu, H. Wu, B.-B. Wang, H.-J. Zhu and J.-J. Qiao, *Chem. Soc. Rev.*, 2015, **44**, 8350–8374.
- 10 C.-C. Yu and S. G. Withers, *Adv. Synth. Catal.*, 2015, **357**, 1633–1654.
- 11 D. B. Stein, Y.-N. Lin and C.-H. Lin, *Adv. Synth. Catal.*, 2008, **350**, 2313–2321.
- 12 R. Liu, Y. Xu, M. Chen, M. Weiwer, X. Zhou, A. S. Bridges, P. L. DeAngelis, Q. Zhang, R. J. Linhardt and J. Liu, *J. Biol. Chem.*, 2010, **285**, 34240–34249.
- 13 D.-J. Namdjou, H.-M. Chen, E. Vinogradov, D. Brochu, S. G. Withers and W. W. Wakarchuk, *ChemBioChem*, 2008, **9**, 1632–1640.
- 14 H. Yu and X. Chen, *Org. Biomol. Chem.*, 2016, **14**, 2809–2818.
- 15 J. Ye, X.-W. Liu, P. Peng, W. Yi, X. Chen, F. Wang and H. Cao, *ACS Catalysis*, 2016, **6**, 8140–8144.
- 16 C. Chen, Y. Zhang, M. Xue, X.-w. Liu, Y. Li, X. Chen, P. G. Wang, F. Wang and H. Cao, *Chem. Commun.*, 2015, **51**, 7689–7692.
- 17 K. Schmoelzer, A. Gutmann, M. Diricks, T. Desmet and B. Nidetzky, *Biotechnol. Adv.*, 2016, **34**, 88–111.
- 18 P. Bojarová, L. Petrásková, E. E. Ferrandi, D. Monti, H. Pelantová, M. Kuzma, P. Simerská and V. Křen, *Adv. Synth. Catal.*, 2007, **349**, 1514–1520.
- 19 B. Henrissat and G. Davies, *Curr. Opin. Struct. Biol.*, 1997, **7**, 637–644.
- 20 R. Han, J. Li, H.-d. Shin, R. R. Chen, G. Du, L. Liu and J. Chen, *Biotechnol. Adv.*, 2014, **32**, 415–428.
- 21 S. Meinke and J. Thiem, in *SialoGlyco Chemistry and Biology II: Tools and Techniques to Identify and Capture Sialoglycans*, ed. R. Gerardy-Schahn, P. Delannoy and M. von Itzstein, Springer International Publishing, Cham, 2015, pp. 231–250.
- 22 C. P. Darley, A. M. Forrester and S. J. McQueen-Mason, *Plant Mol. Biol.*, 2001, **47**, 179–195.
- 23 I. Chlubnova, D. Filipp, V. Spiwok, H. Dvorakova, R. Daniellou, C. Nugier-Chauvin, B. Kralova and V. Ferrieres, *Org. Biomol. Chem.*, 2010, **8**, 2092–2102.
- 24 Š. Malá, H. Dvořáková, R. Hrabal and B. Králová, *Carbohydr. Res.*, 1999, **322**, 209–218.

- 25 A. S. Borisova, D. R. Ivanen, K. S. Bobrov, E. V. Eneyskaya, G. N. Rychkov, M. Sandgren, A. A. Kulminskaya, M. L. Sinnott and K. A. Shabalin, *Carbohydr. Res.*, 2015, **401**, 115–121.
- 26 Y. B. Tewari and R. N. Goldberg, *J. Biol. Chem.*, 1989, **264**, 3966–3971.
- 27 V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho and B. Henrissat, *Nucl. Acids Res.*, 2014, **42**, D490–D495.
- 28 A. Planas, M. Faijes and V. Codera, in *Carbohydrate Chemistry: State of the Art and Challenges for Drug Development*, Imperial College Press, 2016, pp. 215–245.
- 29 P. M. Danby and S. G. Withers, *ACS Chem. Biol.*, 2016, **11**, 1784–1794.
- 30 A. Trincone, *Biomolecules*, 2013, **3**, 334–350.
- 31 B. Zeuner, C. Jers, J. D. Mikkelsen and A. S. Meyer, *J. Agric. Food Chem.*, 2014, **62**, 9615–9631.
- 32 M. Ye, L.-Q. Yan, N. Li and M.-H. Zong, *J. Mol. Catal. B: Enzym.*, 2012, **79**, 35–40.
- 33 A. J. Fairbanks, *Pure Appl. Chem.*, 2013, **85**, 1847–1863.
- 34 M. Fujita, S.-i. Shoda, K. Haneda, T. Inazu, K. Takegawa and K. Yamamoto, *Biochim. Biophys. Acta*, 2001, **1528**, 9–14.
- 35 R. M. Schmaltz, S. R. Hanson and C.-H. Wong, *Chem. Rev.*, 2011, **111**, 4259–4307.
- 36 W. Huang, J. Li and L.-X. Wang, *ChemBioChem*, 2011, **12**, 932–941.
- 37 K. Yamamoto, *Biotechnol. Lett.*, 2013, **35**, 1733–1743.
- 38 V. Puchart, *Biotechnol. Adv.*, 2015, **33**, 261–276.
- 39 E. C. O'Neill and R. A. Field, *Carbohydr. Res.*, 2015, **403**, 23–37.
- 40 C. Goedl, T. Sawangwan, P. Wildberger and B. Nidetzky, *Biocatal. Biotransform.*, 2010, **28**, 10–21.
- 41 M. R. M. de Groeve, T. Desmet and W. Soetaert, *J. Biotechnol.*, 2011, **156**, 253–260.
- 42 A. Miranda-Molina, E. Castillo and A. Lopez Munguia, *Food Chem.*, 2017, **227**, 202–210.
- 43 M. Gimeno-Perez, P. Santos-Moriano, L. Fernandez-Arrojo, A. Poveda, J. Jimenez-Barbero, A. O. Ballesteros, M. Fernandez-Lobato and F. J. Plou, *Process Biochem.*, 2014, **49**, 423–429.
- 44 D. Kwan and S. Withers, *J. Carbohydr. Chem.*, 2011, **30**, 181–205.
- 45 D. Daude, I. Andre, P. Monsan and M. Remaud-Simeon, *Carbohydr. Chem.*, 2014, **40**, 624–645.
- 46 A. Verges, E. Cambon, S. Barbe, S. Salamone, Y. Le Guen, C. Moulis, L. A. Mulard, M. Remaud-Siméon and I. André, *ACS Catalysis*, 2015, **5**, 1186–1198.
- 47 S. Salamone, C. Guerreiro, E. Cambon, I. Andre, M. Remaud-Simeon and L. A. Mulard, *Chem. Commun.*, 2015, **51**, 2581–2584.
- 48 L. F. Mackenzie, Q. Wang, R. A. J. Warren and S. G. Withers, *J. Am. Chem. Soc.*, 1998, **120**, 5583–5584.
- 49 B. Cobucci-Ponzano, C. Zorzetti, A. Strazzulli, S. Carillo, E. Bedini, M. M. Corsaro, D. A. Comfort, R. M. Kelly, M. Rossi and M. Moracci, *Glycobiology*, 2011, **21**, 448–456.
- 50 B. Cobucci-Ponzano, C. Zorzetti, A. Strazzulli, E. Bedini, M. M. Corsaro, G. Sulzenbacher, M. Rossi and M. Moracci, *Biocatal. Biotransform.*, 2012, **30**, 288–295.
- 51 M. Umekawa, W. Huang, B. Li, K. Fujita, H. Ashida, L.-X. Wang and K. Yamamoto, *J. Biol. Chem.*, 2008, **283**, 4469–4479.
- 52 M. Kitaoka, Y. Honda, S. Fushinobu, H. Masafumi, K. Takane and Y. Kenji, *Trends Glycosci. Glycotechnol.*, 2009, **21**, 23–39.

- 53 Y. Honda, S. Fushinobu, M. Hidaka, T. Wakagi and H. Shoum, *J. Appl. Glycosci.*, 2009, **56**, 119–125.
- 54 B. Cobucci-Ponzano, A. Strazzulli, M. Rossi and M. Moracci, *Adv. Synth. Catal.*, 2011, **353**, 2284–2300.
- 55 X. Zeng, R. Yoshino, T. Murata, K. Ajisaka and T. Usui, *Carbohydr. Res.*, 2000, **325**, 120–131.
- 56 C. Li and Y.-W. Kim, *ChemBioChem*, 2014, **15**, 522–526.
- 57 N. F. Brás, P. A. Fernandes and M. J. Ramos, *Theor. Chem. Acc.*, 2009, **122**, 283.
- 58 G. Osanjo, M. Dion, J. Drone, C. Solleux, V. Tran, C. Rabiller and C. Tellier, *Biochemistry*, 2007, **46**, 1022–1033.
- 59 S. Blanchard, S. Armand, P. Couthino, S. Patkar, J. Vind, E. Samain, H. Driguez and S. Cottaz, *Carbohydr. Res.*, 2007, **342**, 710–716.
- 60 H. Nakai, B. O. Petersen, Y. Westphal, A. Dilokpimol, M. Abou Hachem, J. Ø. Duus, H. A. Schols and B. Svensson, *Prot. Eng., Des. Sel.*, 2010, **23**, 781–787.
- 61 N. Bridiau and T. Maugard, *Biotechnol. Prog.*, 2011, **27**, 386–394.
- 62 D. Schmidt, B. Sauerbrei and J. Thiem, *J. Org. Chem.*, 2000, **65**, 8518–8526.
- 63 R. V. Stick, K. A. Stubbs and A. G. Watts, *Aust. J. Chem.*, 2004, **57**, 779–786.
- 64 A. Trincone, A. Giordano, G. Perugino, M. Rossi and M. Moracci, *Chem-BioChem*, 2005, **6**, 1431–1437.
- 65 M. Faijes, M. Saura-Valls, X. Pérez, M. Conti and A. Planas, *Carbohydr. Res.*, 2006, **341**, 2055–2065.
- 66 H.-Y. Feng, J. Drone, L. Hoffmann, V. Tran, C. Tellier, C. Rabiller and M. Dion, *J. Biol. Chem.*, 2005, **280**, 37088–37097.
- 67 Z. Marton, V. Tran, C. Tellier, M. Dion, J. Drone and C. Rabiller, *Carbohydr. Res.*, 2008, **343**, 2939–2946.
- 68 D. Teze, M. Dion, F. Daligault, V. Tran, C. André-Miral and C. Tellier, *Bioorg. Med. Chem. Lett.*, 2013, **23**, 448–451.
- 69 J. Thiem and B. Sauerbrei, *Angew. Int. Ed. Eng.*, 1991, **30**, 1503–1505.
- 70 W. Li, Y. Sun, H. Ye and X. Zeng, *Eur. Food Res. Technol.*, 2010, **231**, 55–63.
- 71 A. Vetere and S. Paoletti, *Biochem. Biophys. Res. Commun.*, 1996, **219**, 6–13.
- 72 M. Henze, D. Merker and L. Elling, *Int. J. Mol. Sci.*, 2016, **17**, 210.
- 73 M. Sandoval, A. Cortes, C. Civera, J. Trevino, E. Ferreras, M. Vaultier, J. Berenguer, P. Lozano and M. J. Hernaiz, *RSC Adv.*, 2012, **2**, 6306–6314.
- 74 C. Bayón, Á. Cortés, J. Berenguer and M. J. Hernáiz, *Tetrahedron*, 2013, **69**, 4973–4978.
- 75 N. Bridiau, N. Issaoui and T. Maugard, *Biotechnol. Prog.*, 2010, **26**, 1278–1289.
- 76 M. Pérez-Sánchez, M. Sandoval and M. J. Hernáiz, *Tetrahedron*, 2012, **68**, 2141–2145.
- 77 M. Pérez-Sánchez, Á. Cortés Cabrera, H. García-Martín, J. V. Sinisterra, J. I. García and M. J. Hernáiz, *Tetrahedron*, 2011, **67**, 7708–7712.
- 78 M. Perez-Sanchez, M. Sandoval, A. Cortes-Cabrera, H. Garcia-Marin, J. V. Sinisterra, J. I. Garcia and M. J. Hernaiz, *Green Chem.*, 2011, **13**, 2810–2817.
- 79 S. Strompen, A. Miranda-Molina, A. López-Munguía, E. Castillo and G. Saab-Rincón, *Carbohydr. Res.*, 2015, **404**, 46–54.
- 80 A. Pennec, L. Legentil, L. Herrera-Estrella, V. Ferrieres, A.-L. Chauvin and C. Nugier-Chauvin, *Green Chem.*, 2014, **16**, 3803–3809.
- 81 L. Bohe and D. Crich, *Trends Glycosci. Glycotechnol.*, 2010, **22**, 1–15.
- 82 S. S. Nigudkar and A. V. Demchenko, *Chem. Sci.*, 2015, **6**, 2687–2704.

- 83 L. K. Mydock and A. V. Demchenko, *Org. Biomol. Chem.*, 2010, **8**, 497–510.
- 84 X. Zhu and R. R. Schmidt, *Angew. Chem., Int. Ed.*, 2009, **48**, 1900–1934.
- 85 C. M. Pedersen, J. Olsen, A. B. Brka and M. Bols, *Chem. – Eur. J.*, 2011, **17**, 7080–7086.
- 86 C.-C. Wang, J.-C. Lee, S.-Y. Luo, S. S. Kulkarni, Y.-W. Huang, C.-C. Lee, K.-L. Chang and S.-C. Hung, *Nature*, 2007, **446**, 896–899.
- 87 J. M. Williams and A. C. Richardson, *Tetrahedron*, 1967, **23**, 1369–1378.
- 88 P. R. Muddasau, E. Bozo, B. Bernet and A. Vasella, *Helv. Chim. Acta*, 1994, **77**, 257–290.
- 89 J. Lawandi, S. Rocheleau and N. Moitessier, *Tetrahedron*, 2016, **72**, 6283–9319.
- 90 M. Jäger and A. J. Minnaard, *Chem. Commun.*, 2016, **52**, 656–664.
- 91 J. S. Haisheng Wang, Li-He Zhang and Xin-Shan Ye, *J. Org. Chem.*, 2004, **69**, 5774–5777.
- 92 T. Kurahashi, T. Mizutani and J.-i. Yoshida, *J. Chem. Soc., Perkin Trans. 1*, 1999, 465–473.
- 93 G. Xiao, G. A. Cintron-Rosado, D. A. Glazier, B.-m. Xi, C. Liu, P. Liu and W. Tang, *J. Am. Chem. Soc.*, 2017, **139**, 4346–4349.
- 94 B. Ren, M. Rahm, X. Zhang, Y. Zhou and H. Dong, *J. Org. Chem.*, 2014, **79**, 8134–8142.
- 95 X. Zhang, B. Ren, J. Ge, Z. Pei and H. Dong, *Tetrahedron*, 2016, **72**, 1005–1010.
- 96 P. Peng, M. Linseis, R. F. Winter and R. R. Schmidt, *J. Am. Chem. Soc.*, 2016, **138**, 6002–6009.
- 97 K. S. Griswold and S. J. Miller, *Tetrahedron*, 2003, **59**, 8869–8875.
- 98 H. Takeuchi, K. Mishiro, Y. Ueda, Y. Fujimori, T. Furuta and T. Kawabata, *Angew. Chem., Int. Ed.*, 2015, **54**, 6177–6180.
- 99 H. Dong, Y. Zhou, X. Pan, F. Cui, W. Liu, J. Liu and O. Ramström, *J. Org. Chem.*, 2012, **77**, 1457–1467.
- 100 W. Muramatsu and Y. Takemoto, *J. Org. Chem.*, 2013, **78**, 2336–2345.
- 101 E. Dimitrijević and M. S. Taylor, *ACS Catal.*, 2013, **3**, 945–9962.
- 102 E. Kaji, D. Yamamoto, Y. Shirai, K. Ishige, Y. Arai, T. Shirahata, K. Makino and T. Nishino, *Eur. J. Org. Chem.*, 2014, 3536–3539.
- 103 M. S. Taylor, *Acc. Chem. Res.*, 2014, **48**, 295–305.
- 104 S. Rocheleau, J. Pottel, I. Huskić and N. Moitessier, *Eur. J. Org. Chem.*, 2017, 646–656.
- 105 R. S. Mancini, J. B. Lee and M. S. Taylor, *Org. Biomol. Chem.*, 2017, **15**, 132–143.
- 106 L. Chan and M. S. Taylor, *Org. Lett.*, 2011, **13**, 3090–3093.
- 107 D. Lee and M. S. Taylor, *J. Am. Chem. Soc.*, 2011, **133**, 3724–3727.
- 108 D. Lee, C. L. Williamson, L. Chan and M. S. Taylor, *J. Am. Chem. Soc.*, 2012, **134**, 8260–8267.
- 109 M. Heuckendorff, C. M. Pedersen and M. Bols, *Chem. – Eur. J.*, 2010, **16**, 13982–13994.
- 110 H. D. Premathilake and A. V. Demchenko, *Top. Curr. Chem.*, 2011, **301**, 189–221.
- 111 H. H. Jensen, C. M. Pedersen and M. Bols, *Chem. – Eur. J.*, 2007, **13**, 7576–7582.
- 112 A. Kanazawa, S. Namiki and M. Suzuki, *J. Polym. Sci., Part A: Polym. Chem.*, 2007, **45**, 3851–3860.
- 113 A. Steinmann, J. Thimm, M. Matwiejuk and J. Thiem, *Macromolecules*, 2010, **43**, 3606–3612.
- 114 S. Boettcher, M. Matwiejuk and J. Thiem, *Beilstein J. Org. Chem.*, 2012, **8**, 413–420.

- 115 G. Pelletier, A. Zwicker, C. L. Allen, A. Schepartz and S. J. Miller, *J. Am. Chem. Soc.*, 2016, **138**, 3175–3182.
- 116 H. Tanaka, M. Adachi, H. Tsukamoto, T. Ikeda, H. Yamada and T. Takahashi, *Org. Lett.*, 2002, **4**, 4213–4216.
- 117 R. Euzen, J.-P. Guegan, V. Ferrieres and D. Plusquellec, *J. Org. Chem.*, 2007, **72**, 5743–5747.
- 118 P. Peltier, R. Daniellou, C. Nugier-Chauvin and V. Ferrieres, *Org. Lett.*, 2007, **9**, 5227–5230.
- 119 P. Peltier, J.-P. Guegan, R. Daniellou, C. Nugier-Chauvin and V. Ferrieres, *Eur. J. Org. Chem.*, 2008, 5988–5994.
- 120 D. V. Yashunsky, Y. E. Tsvetkov, A. A. Grachev, A. O. Chizhov and N. E. Nifantiev, *Carbohydr. Res.*, 2016, **419**, 8–17.
- 121 H. W. a. J. Ning, *J. Org. Chem.*, 2003, **68**, 2521–2524.
- 122 R. Geurtsen and G.-J. Boons, *Eur. J. Org. Chem.*, 2002, 1473–1477.
- 123 W. Muramatsu and H. Yoshimatsu, *Adv. Synth. Catal.*, 2013, **355**, 2518–2524.
- 124 D. A. Niedbal and R. Madsen, *Tetrahedron*, 2016, **72**, 415–419.
- 125 E. Kaji, K. Shibayala and K. In, *Tetrahedron Lett.*, 2003, **44**, 4881–4885.
- 126 N. Moitessier and J.-Y. Chapleur, *Tetrahedron Lett.*, 2003, **44**, 1731–1735.
- 127 K. A. D'Angelo and M. S. Taylor, *J. Am. Chem. Soc.*, 2016, **138**, 11058–11066.
- 128 A. Nakagawa, M. Tanaka, S. Hanamura, D. Takahashi and K. Toshima, *Angew. Chem., Int. Ed.*, 2015, **54**, 10935–10939.
- 129 A. M. Czyzewski and A. E. Barron, *AIChE J.*, 2008, **54**, 2–8.

Glycosyltransferase inhibitors: a promising strategy to pave a path from laboratory to therapy

Paula Alexandra Videira,^{*a} Filipa Marcelo^a and Ravneet Kaur Grewal^b

DOI: 10.1039/9781788010641-00135

Abnormal glycosylation is a common feature in disease that typically results from misregulation in expression and/or activity of glycosyltransferases (GTs) and glycosidases. Interfering with these enzymes, by developing a prospect of targeted inhibitors, has opened newer avenues to meet the challenge of abnormal glycosylation. Progress in GT inhibitors has been relatively slower in comparison to glycosidases. In case of GTs, their polyspecific nature, structural homology, overlapping specificities, multi-substrate catalytic mechanism and relatively less available 3D-structural data stance a big challenge to explore the whole potential of GT target inhibitors in comparison to glycosidases for therapeutic purposes. In this review, we focus on GT specific inhibitors, which are organised as conventional donor analogues (*viz.* donor, acceptor and transition state mimetic), glycomimetic, alternative inhibitor chemotype and metabolic chain terminator. Conventional donor analogues are however limited by the poor membrane permeability and chemical instability. Thus, in the last two decades, alternative inhibitor chemotypes caught attention as a promising lead, which are not structurally derived from GT substrates and possess drug like properties offering an alternate non-substrate like inhibitor based strategy for drug development. Although successful cellular GT-targeted molecules are yet to be achieved, recent designing of metabolic inhibitors *i.e.*, dead end substrates are emerging as an impetus to explore the potential of GT families as therapeutic targets.

Glycans (carbohydrates, sugars and glycoconjugates) are relevant biomolecules that mediate numerous biological activities *viz.* inter- and intra-cellular communication, immune response, microbial adhesion among other feature.¹ Glycosylation has the potential to generate a wide range of glycan assemblies composing the most complex language of life: the glycome. From interaction of distinct glycan receptors, the glycome is translated into the most varied cellular effects in health and disease. Glycan-structure is influenced by pathological and clinical conditions such as microbial infection, inflammation, malignant transformation, alcoholism, diabetes and neuronal disorders.^{2–6} In this perspective, the biosynthesis of glycans is thus a critical step and a potential therapeutic target.

1 Glycan biosynthesis and glycan processing enzymes

Glycan biosynthesis is not template driven, as is observed with proteins and nucleic acids. However, the glycan structures are encoded indirectly

^aUCIBIO, Departamento Ciências da Vida, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516, Caparica, Portugal.

E-mail: p.videira@fct.unl.pt

^bDepartment of Biotechnology, Sri Guru Granth Sahib World University (SGGSWU), Fatehgarh Sahib, Punjab - 140406, India

by the location and coordinated activities of specific glycan processing enzymes *i.e.*, glycosyltransferases (GT) and glycosidases and availability of the substrates.⁷

GTs are multisubstrate enzymes that transfer sugar residues from activated donor substrate (usually sugar nucleotides) to an acceptor substrate such as protein, lipid or glycan.¹ The glycosylation involves either inversion or retention of the anomeric configuration, depending on the nature of the GTs (Fig. 1).^{8,9} GTs are classified into three structural super

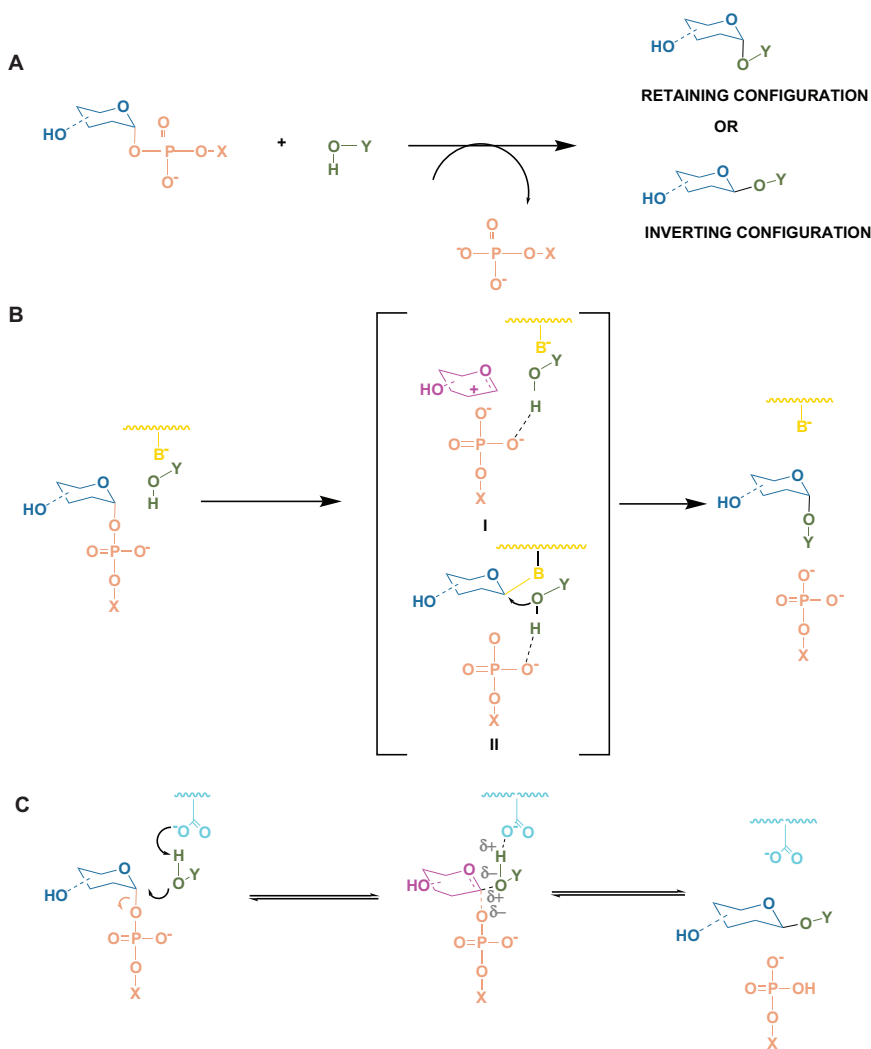


Fig. 1 A: GTs catalyse the glycosylation involves either retention or inversion of the anomeric configuration. B: Retaining GT are classified as two types depending on the presence or absence of a potential nucleophile in the catalytic site *i.e.*, glycosylation involves either of two models as a variation of a common mechanism in a two-step reaction with the formation of an oxocarbenium ion-like transition state which could be stabilized *via* the formation of an oxocarbenium ion or a covalent glycosyl-enzyme depending on the structure of the catalytic site. C: Inverting GT involves a single-displacement mechanism with an oxocarbenium ion-like transition state.

families: GTA (mostly metal ion dependent), GTB (metal ion independent), and GTC folds.¹⁰ They are organised into 94 structural families at the carbohydrate active enzymes (CAZy) database (available at <http://www.cazy.org/>), on the basis of similarity in amino acid sequence.¹¹ However, GTs have not been efficiently organized based on their activities due to: (a) Most of these enzymes are membrane bound and are difficult to obtain in sufficient amount or properly folded to perform kinetic studies, (b) Deduction of GT function is cumbersome as both nucleotide donor and acceptor substrates need to be identified, (c) Methods to address GT activity are complex and require experienced and well equipped teams. Another family of glycan processing enzymes are glycosidases (glycoside hydrolases), which are classified based on the sugar unit they hydrolyse and orientation of the glycosidic bond they cleave.⁷

The dominant mean to decipher the role of glycans in physiological and pathological conditions, involves the modulation of the enzyme activities responsible for the formation or degradation of the glycan, in cellular and animal models and correlating the changes with physiological expression. The common strategies involve genetic or chemical methods: (a) Genetic approaches overexpress, knock down or knock out the genes coding the glycan processing enzymes, and (b) Chemical approaches use small molecule inhibitors to modulate glycan processing enzymes. These approaches complement each other and both efficiently intercept biosynthetic pathways, in a cellular context. Chemically synthesized molecules are already being used to disrupt pathological carbohydrate-dependent biological processes and are emerging as important therapeutic agents due to their easiness in production and scale up process.

2 Small molecule inhibitors to modulate glycan processing enzymes

When considering the generation, and use of chemical inhibitors, one must understand the desirable features of inhibitors and most importantly the issues of general *versus* class specific promiscuity, which collectively constitute the possible off-target effects. Parameters like ligand efficiency or ligand-lipophilicity efficiency provide useful measures for evaluating general promiscuity. Electrophilic groups should also generally be avoided.¹² However, class-specific promiscuity is a concern in glycobiology because inhibitors can affect a class of functionally related glycan processing enzymes with structurally similar active site. This turns quite challenging designing inhibitors which can selectively target an enzyme and not the entire class. The main aspects to be considered to achieve inhibitors with potential clinical relevance are the characteristics of the inhibitor, which make it an appropriate candidate in biological studies, and the manufacturability and efficiency (Table 1).⁷ Impressive advances have been made in the case of glycosidases inhibitors and their relevance attested by the number of drugs approved in clinic with notable examples: Tamiflu, Miglustat, Vogalibose, Acarbose, Miglitol, Castanospermine, Swainsonine, Kiflunensine.^{13–15} Several reviews covering the progress

Table 1 What to consider when determining drug potency?

Requirements of chemical glycan processing inhibitor
Higher potency, when benchmarked against known inhibitors
Reduced cLog <i>P</i> values
<i>In vitro</i> inhibitor potency and half-maximum effective concentration (EC ₅₀)
Evaluation of inhibitor's stability in biological fluids for validating its application
<i>in vivo</i> and in cellular models
Assay of inhibitor in cells to identify probes
<i>In vitro</i> screening of inhibitors and histology of tissues to evaluate off target effects
Minimise the off-target effects by using lowest concentration of inhibitor
<i>In vivo</i> analysis of dose-dependent effects on glycosylation

in the field of glycosidases inhibitor are extensively reported in the literature,^{13–20} Herein, we have paid attention to glycosyltransferase inhibitors attempting to describe the most recent advances in the field with special emphasis for the challenges and perspectives for the future. While reviews on GTs inhibitors are available in the literature^{21–26} with discussion mostly focused on the synthesis of chemical tools specific for a particular class of GT from a chemist's perspective, we provide herein an overview of different glycosyltransferase inhibitors classified on the basis of different strategies employed till date for designing the GT specific chemical tools; in particular GTs involved in pathological conditions *e.g.*, cancer, inflammation, infection and metabolic disorders to develop an understanding among readers about the road map between glycochemistry and glycomedicine to pave a path from laboratory to therapy.

3 Glycosyltransferase inhibitors

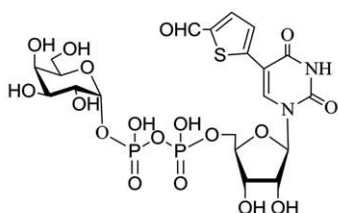
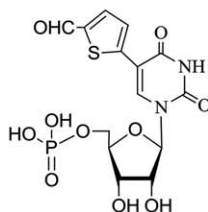
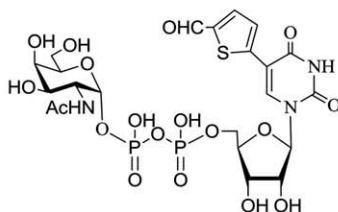
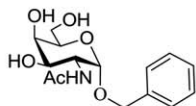
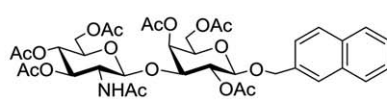
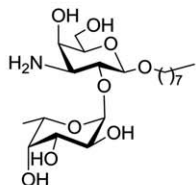
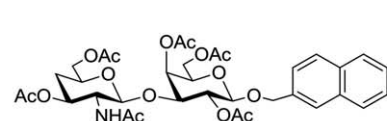
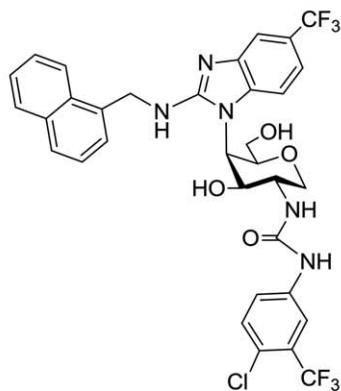
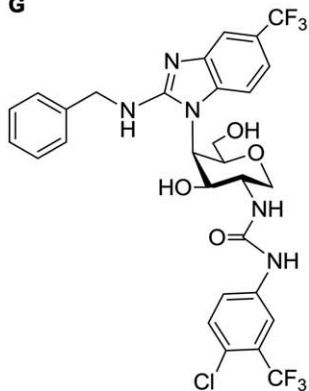
Although GTs have similar physiological significance to other enzyme families like proteases and kinases, the potential of GTs for chemical biology and drug discovery is relatively unexplored.⁷ This is due, at least in part, to relative lack of GT specific inhibitors for structural, mechanistic and cellular studies.²⁷ Furthermore, progress in inhibitors of GTs has been slower than glycosidases and it may be attributed to comparatively less information available on 3D-structure of GTs and on the polyspecific nature of GT families.⁷ The real challenge in creating these chemical tools is due to many factors: (a) Structural homology and overlapping specificities of the GTs, which makes the impact of GT inhibitor on glycan level less obvious. As an example, the large family of human polypeptide *N*-acetylgalactosamine(GalNAc) transferase(ppGalNAcTs) catalyzes the same enzymatic step, having distinct, but overlapping kinetic properties and substrate specificities that allows to determine which and where proteins are *O*-glycosylated; (b) Heterogeneity of glycans, which exist not only at the point of linkage, but also in the structure of attached glycan; (c) Although most GTs belong to either GTA or GTB fold, many are highly dynamic proteins which follow complex, multi-substrate reaction mechanism involving several conformational changes. This unusual plasticity has complicated the rationale of *de novo* design of inhibitors; (d)

Complexity of GT catalytic mechanism due to participation of distinct components in transition state (*i.e.*, nucleotide-sugar donor, acceptor substrate, and in certain cases metal co-factor, and (e) Glycans are primarily secreted and are expressed at cell surface and mediate intercellular interactions, and are evident as phenotypes only in context to whole organism.^{7,28–31} GTs catalyze the transfer of the sugar unit from a nucleotide-sugar substrate donor to the substrate acceptor. In this sense, most of the approaches to design GT inhibitors are classified into two major categories: (a) Nucleotide-sugar donor based inhibitors, and (b) carbohydrate scaffolds to mimic acceptor substrates.

In the first approach, the base of the nucleotide diphosphate (*i.e.*, NDP) is decorated with other substituent groups.^{32–37} In this sense the nucleotide scaffold will warrant the binding while the substituent groups in the base will contribute to improve specificity to GTs. Most of the inhibitors based on this approach contain the phosphate groups instead of only the nucleoside portion in order to increase the affinity towards GTs.^{35,38,39} However, NDP based inhibitors have poor membrane permeability with poor *in vivo* pharmacokinetic properties due to highly hydrophilic and negative charge nature.³⁷ Second approach involves exploring carbohydrates as scaffolds to design mimics of the sugar acceptor substrates to interfere in GT activity.^{31,37,40} Alternative approach to identify GT inhibitors is to screen a purified enzyme against large combinatorial libraries to generate leads that can be optimised to probe glycosylation pathways.^{37,41,42} Existing inhibitors for GTs can be classified into five classes: substrate analogues which include donor and acceptor mimics, transition state analogs where the donor and the acceptor are covalently attached, glycomimetics, alternative inhibitor chemotypes and metabolic chain terminators.

4 Substrate analogs

Donor mimics are commonly developed by modifying the sugar moiety of donor substrate. Pesnot *et al.*⁴³ reported a new synthesis strategy by altering the base moiety instead of sugar in donor mimics with 5-substituted UDP-Gal derivative for galactosyltransferases (GalTs). It is demonstrated that 5-formylthien-2-yl group locks Gal-T in an unproductive conformation that inhibit the enzyme activity (Fig. 2A). Further studies showed that this strategy can be extended to other GTs *e.g.*, a bifunctional α -(1,4)-glycosyltransferase [AA(Gly)B; an enzyme involved in the biosynthesis of blood group A and B in humans]. The modification of uridine of UDP-Gal with 5-formylthien-2-yl yields a poor substrate, but a potent nanomolar inhibitor (Fig. 2A) for AA(Gly)B, emphasizing that 5-formylthien-2-yl group interfere with the folding of an internal loop and C-terminus, which are essential for catalysis.²⁷ Interestingly, this study proposed for the first time the existence of pseudo-closed conformation for AA(Gly)B, different from the known closed conformation typically established in presence of the donor and acceptor thus exemplifying the considerable conformation plasticity of GTs. Another notable example is the base-modified UDP-sugars *i.e.*, 5-(5-formylthien-2-yl UDP)

A**B****C****D****E****F****G**

(Fig. 2A) and its close derivatives (Fig. 2B) which reduce the basal levels of cell surface P-selectin glycoprotein ligand-1 (PSGL-1).⁴⁴ PSGL-1 is a cell surface glycoprotein decorated with sialyl Lewis X (SLe^x), which is implicated in the recruitment of inflammatory cells. P-selectin/PSGL-1 interaction is a putative therapeutic target in asthma and chronic obstructive pulmonary disease. Such discrimination potential of base-modified UDP-sugars for PSGL-1 in pro-inflammatory and physiological conditions may open new avenues for targeting inflammatory settings without compromising host defence status.⁴⁴

Donor mimics by altering the base moiety may yield inhibitors with affinity in μM range for GTs^{45–47} however, since these inhibitors use a common donor it may lead to non-selective inhibition of GTs. In this sense, acceptor mimics can be more selective GT inhibitors,^{48,49} noteworthy lower affinity ($K_m = \text{mM}$).

Acceptor mimics can be classified as: (a) Acceptor substrate decoys, and (b) Competitive acceptor analogs. Substrate decoys are small-molecule analogs of endogenous acceptor substrates that outcompete the natural acceptor, thus leaving it unmodified when used at high concentrations. This has been demonstrated by Kuan SF *et al.*⁵⁰ with benzyl 2-acetamido-2-deoxy- α -D-galactopyranoside (Fig. 2C) to block mucin-type glycosylation, and it has since been widely explored to disrupt the biosynthesis of various glycan structures. Brown *et al.*⁵¹ employed per-*O*-acetylated GlcNAc- β -1,3-Gal- β -*O*-naphthalenemethanol (Fig. 2D) as a substrate decoy to inhibit formation of SLe^x in cancer cells and *in vivo*. SLe^x is an important selectin ligand, crucial for leukocyte migration and playing an important role in metastization. The drawback of this approach is the accumulation of un-natural free glycans in cells, which can be toxic. This strategy could be improved by modifying the substrate decoy to yield a mimic that can serve as a competitive inhibitor instead of an acceptor substrate decoy^{52,53} As an example, 3-amino-3-deoxy-[Fuc α (1-2)]Gal β -*O*(CH₂)₇CH₃] (Fig. 2E) is a potent inhibitor of blood group A α -1,3-*N*-acetyl-galactosaminyltransferase *in vitro*. Interestingly, this acceptor mimic blocks enzymatic activity against an exogenously introduced cell-permeable substrate but exhibits restricted activity on cell-surface glycosylation.⁵² Indeed the amino charge only present in the inhibitor decrease the diffusion into the cells hampering it to reach the Golgi apparatus in high concentrations. Another example includes 4-deoxy analog of per-*O*-acetylated GlcNAc- β -1,3-Gal- β -*O*-naphthalenemethanoldisaccharide (Fig. 2F) which blocks SLe^x expression in a cancer cell line and reduces tumor metastasis *in vivo*.⁵¹ Per-acetylation of the

Fig. 2 A: Structure of Substrate analogs. The figure represents two types of substrate analogs: donor mimics (A, B) and acceptor mimics (C–G). A: 5-(5-formylthien-2-yl UDP-Gal) blocks Gal-T and AA(Gly)B. B: 5-(5-formylthien-2-yl UDP-Gal) based derivatives reduce the basal levels of PSGL-1. C: Benzyl 2-acetamido-2-deoxy- α -D-galactopyranoside blocks mucin-type glycosylation. D: per-*O*-acetylated GlcNAc- β -1,3-Gal- β -*O*-naphthalenemethanol inhibits SLe^x expression. E: 3-amino-3-deoxy-[Fuc α (1-2)]Gal β -*O*(CH₂)₇CH₃] inhibits blood group A α -1,3-*N*-acetyl-galactosaminyltransferase. F: per-*O*-acetylated GlcNAc- β -1,3-Gal- β -*O*-naphthalenemethanoldisaccharide targets SLe^x. G: Monosaccharide scaffold based GT inhibitor with antibacterial activity.

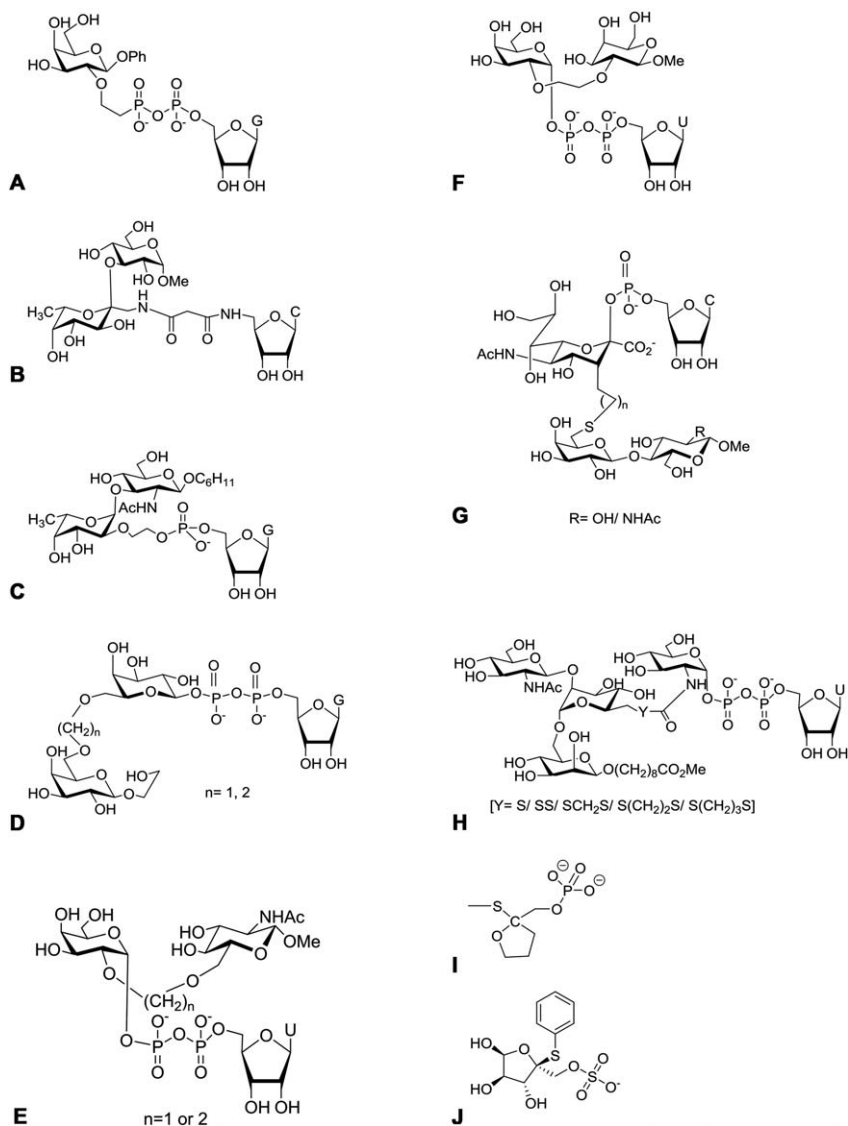


Fig. 3 Structure of Transition State analogs. TS analogs are described as inhibitors (or potential inhibitors) of Fuc-T (A–D); GalT (E and F); ST (G); GnT (H). A: Bisubstrate analog for α -1,2-Fuc-T with the acceptor sugar and the nucleotide which mimics the putative transition state of the enzymatic reaction and inhibits the enzyme competitively. B and C: Trisubstrate analog for α -1,3-Fuc-T with malondiamide (B) or ethylene links (C) between guanosine and fucose residues, respectively, instead of a diphosphate linker. D: Trisubstrate analog transform into GDP analogs for inhibitory activity against α -1,3-Fuc-T. E: Trisubstrate analog for β -1,4-GalT based on its S_N2 like transition state with methylene or ethylene tether. F: Trisubstrate analog for retaining enzyme α -1,3-GalT. G: Trisubstrate methylene linked analog for α -2,3-ST and α -2,6-ST. H: Transition state analog with longer linker strongly inhibit GnT-V.

disaccharide allows passive diffusion through the cell membrane. Acetyl groups are then removed by cytoplasmic or membrane-associated carboxyesterases. In this way, the inhibitors gain access to GTs in the Golgi

complex in much higher concentrations. Recently, Zuegg *et al.*⁴⁰ constructed compound libraries based on the acceptor carbohydrate scaffold of bacterial GTs. The developed GT inhibitors (Fig. 2G) block the GT responsible for the assemble of the sugar units in the peptidoglycan (PG) and show meonomycin-like antibacterial activity *in vitro* however much better pharmacokinetic properties and efficacy in *in vivo* models of infection. The new compounds present a potent antibacterial activity against wide range of gram positive bacteria resistant to common antibiotics.⁵⁴ This study highlights a new mode of action that can be used to develop antibiotics with low resistance frequencies.

5 Transition state analogs

Donor mimics have high affinity (micromolar range) for the enzyme but poor selectivity while acceptor mimics based on carbohydrate scaffolds are less potent (milimolar range) but much more selective inhibitors. Therefore, to get more potent and selective GT inhibitors, bisubstrate analogs in which donor and acceptor substrates are covalently attached to each other were developed⁴⁷ (Fig. 3). Noteworthy, for the rational design of an efficient transition state mimetic the knowledge of the relative orientation for concomitant binding to their respective binding sites on GTs, is of paramount importance.^{47,55,56} Regrettably, the lack of molecular details concerning the ternary complex GTs/donor-substrate/acceptor-substrate has been hampered the advances of a structure-based rational development of GTs inhibitors.

The first TS analog for GT was designed to inhibit α -1,2-fucosyltransferases (Fuc-T), enzymes responsible for biosynthesis of fucosylated oligosaccharides, such as ABH blood group antigen, Le^x, Le^y and SLe^x. Palcic *et al.*⁵⁷ designed a bisubstrate analog for α -1,2-Fuc-T without donor sugar moiety (Fig. 3A), which mimics the putative transition state of the enzymatic reaction and inhibits the enzyme competitively. Heskamp *et al.*^{58,59} reported TS analog (Fig. 3B and C) for α -1,3-Fuc-T, with malondiamide or ethylene links between guanosine and fucose residues, respectively, instead of a diphosphate linker, to overcome poor membrane permeability barrier. However, their inhibitory activities are not reported. Izumi *et al.*⁶⁰ developed two TS analogs for α -1,3-Fuc-Ts, that were examined against α -1,3-Fuc-T V and α -1,3-Fuc-TVI (Fig. 3D). Intriguingly, these transition state analogs act as potent inhibitors for α -1,3-Fuc-T V, but are substrate for α -1,3-Fuc-T VI.

Galactosyltransferase (Gal-T) is an important target for the inhibitor development due to its involvement in the synthesis of cellular antigen *e.g.*, blood group B antigen, lactosamide, LacNAc and Gal α (1 $\rightarrow$ 3)Gal β (1 $\rightarrow$ 4)GlcNAc sequence and is responsible for hyper acute organ rejection in the xenotransplantation of organs from pigs to humans.⁶¹ Hashimoto *et al.*^{62,63} designed TS analog (Fig. 3E) for the inverting GT β -1,4-Gal-T based on its S_N2 like transition state with methylene or ethylene tether. Both compounds inhibit the acceptor and donor UDP-Gal substrate of β -1,4-Gal-T. However, remarkable differences were deduced with respect to the mode of action. Interestingly, mode of

inhibition is determined competitive for substrates (*i.e.*, donor and acceptor) in the presence of methylene linker, while in the case of ethylene tethered the inhibition showed to be non-competitive for both donor and acceptor substrates. The design of a TS analog for an inverting enzyme is relatively simple as incoming nucleophile and leaving nucleotide can be connected in the proper configuration. However, design of a TS analog for retaining enzyme is somewhat challenging because donor and product have the same anomeric configuration. Endo *et al.*⁶⁴ developed first TS analog (Fig. 3F) for retaining enzyme α -1,3-GalT, that inhibits enzyme competitively and is the only retaining glycosyltransferase to which a TS analog inhibitor was developed until date.

Sialyltransferases (STs) are an important class of GTs, synthesize sialic acid containing epitopes, and regulate physiological functions *viz.* inflammation and cell adhesion.¹ The potent inhibitors of STs may be useful as anti-metastatic, immunosuppressive or anti-inflammatory agents. Hinou *et al.*^{39,65} reported a TS analog, where the donor and acceptor moieties are connected by sulfide bonds and distant by an alkyl linker (Fig. 3G), and those inhibitory activities were investigated with α -2,3-ST and α -2,6-ST. In general, analogs with LacNAc moiety are better inhibitors than analogs containing lactose, which is consistent with the substrate preference of the enzymes. Methylene-linked analog act as potent inhibitor for both STs as K_i varies with difference in linker length of the analogs, suggesting that linker length is an important determinant of inhibition. Jung and colleagues⁶⁶ designed TS mimetics, most of which have CMP-substituted- α -hydroxy phosphonate attached to dehydroneuraminic acid, with lactose or galactose derivatives as acceptor moiety. These inhibitors are effective against α -2,3- and α -2,6-STs, however lack enzyme selectivity. Hanashima *et al.*^{67,68} developed TS analog (Fig. 3H), consists of UDP-GalNAc and an acceptor trisaccharide, GlcNAc β (1 $\rightarrow$ 2)Man α (1 $\rightarrow$ 6)Man β with varying linker length and observed that inhibitors with longer linker strongly inhibit GnT-V(N-acetylglucosaminyltransferase produces GlcNAc β (1 $\rightarrow$ 6)Man branched glycans), which correlates with the malignant transformation and metastatic potential of tumor cells and is putative target for inhibition in cancer therapy.

Burkart *et al.*⁶⁹ reported that fluorinated monosaccharides can be employed as transition state inhibitors of sialyl- and fucosyl-transferases. However, their utility as global inhibitors of ST and Fuc-T is limited: (a) Core fucosylation of N-linked glycans by α -1,6-Fuc-T in HL-60 cells is reduced but not abolished, and (b) some sialylation of O-glycans still exists despite obliterated sialylation of N-linked glycans; emphasizes that α -1,6-FucT and at least one ST has less sensitivity to these inhibitors.⁷⁰

In an approach to explore the catalytic mechanism of GTs using density functional theory (DFT) *ab initio* method, tetrahydro-2-(methylthio)furan-2-yl]methyl phosphate dianion⁷¹ (Fig. 3I) and (2S,3R,4R,5S)-3,4,5-trihydroxy-2-(phenylsulfanyl)tetrahydroxy-2 (phenylsulfanyl)tetrahydrofuran-2-yl]methyl sulphate anion⁷² (Fig. 3J), emerged as scaffolds that mimics the TS of the reactions catalyzed by inverting GTs. This presents a promising strategy to design putative transition state analogs for GTs and can be explored further for retaining GTs as is predicted by Raab *et al.*⁷¹

6 Glycomimetics

Glycomimetics is another generation of GT inhibitors that emerged based on carbohydrate mimetics, which include compounds that are structurally altered analogs of carbohydrates; designed to simulate the shape and most functionalities of natural substrate. The progress in glycomimetics, prompted primarily in the field of glycosidase inhibitors, which has led to the development of a wide variety of novel structures *viz.* iminosugars, carbasugars and C-glycosides.³⁵ Then they emerged as putative substrate analogs for GTs. The iminosugars, (*e.g.*, pyrrolidinols, piperidinols and morpholine), which are small organic molecules which mimic monosaccharides but contain a nitrogen atom in place of the endocyclic oxygen, have attracted attention and are already in clinical evaluation.

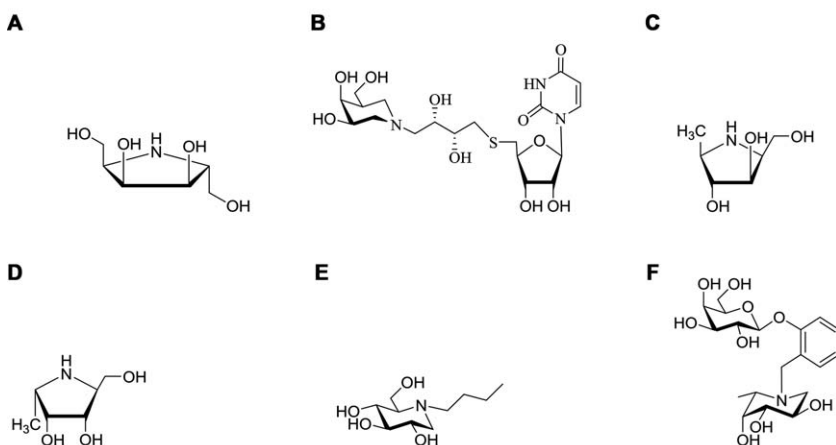
Saotome *et al.*⁷³ developed an iminosugar (Fig. 4A) for β 1,4-GalT as potent as UDP and acting as uncompetitive inhibitor vs UDP-Gal. Kim and colleagues⁷⁴ reported 1-N-iminosugar based UDP-galactose analog (Fig. 4B), which is potent and selective for α 1,3-GalT but not for β 1,4-GalT. Two pyrrolidines (Fig. 4C and D) exhibit a strong synergetic inhibition of α 1,3-FucT in the presence of GDP, and it is attributed to the formation of complex between GDP and iminosugars which mimics the transition state of fucosyltransferase reaction.⁷⁵

Miglustat is a prominent example of clinically employed iminosugar (Fig. 4E). Miglustat mimics the ceramide, acceptor substrate of glucosylceramide synthase (GCS) and are used in the treatment of lysosomal storage disorders Gaucher and Niemann-Pick type C disease.⁷⁶ Increasing the alkyl chain length and therefore inherent hydrophobicity of N-alkyl chain increases potency, membrane adsorption, with further persistence in tissue and improved brain penetration.⁷⁷ Among the piperidinol-based iminosugars, TS analogs are the best inhibitors. An example is a compound which employs two glycomimics: an iminosugar for mimicking the fucose moiety and phenol as GlcNAc-mimic, and is more potent than GlcNAc itself⁷⁸ (Fig. 4F).

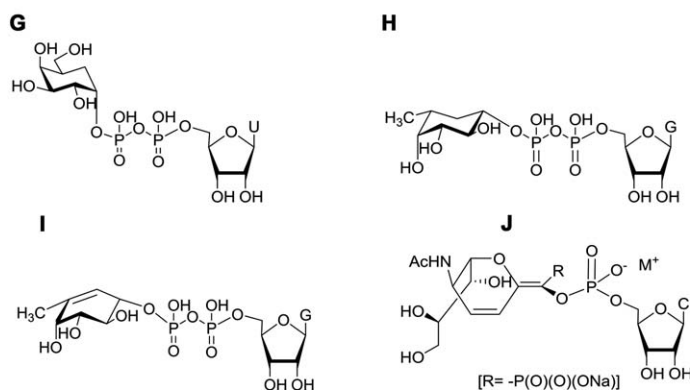
Carbasugars also attracted attention of chemists because of their chemical stability as well as their biological properties specifically as antibiotics and glycosidase inhibitors. The carbasugar analogs of UDP-Gal (Fig. 4G) and GDP-Fuc (Fig. 4H and I) are competitive inhibitors of β 1,4-GalT and α 1,3-4Fuc-T respectively.^{79,80} Schaub and coworkers⁸¹ developed most potent analog (Fig. 4J), with 1000 folds high affinity for ST than natural substrate and this remarkable TS analog can be explored for delineating biological role of sialic acid containing glycoconjugates.

C-glycosides began to be used for design of stable analogs of sugar nucleotides which can fit into enzyme active site without undergoing enzymatic cleavage. Another interest in C-glycosides is the possibility of altering aglycon moiety to prepare TS analogs or pyrophosphate mimics.³⁵ Schmidt and colleagues⁸² designed a mimic of TS analog (Fig. 4K) by combining a glycal structure with non-cleavable C–C bond to the pseudo anomeric centre, that demonstrated to be an effective inhibitor for β -galactosyltransferase. A potent inhibitor of α 1,3-Fuc-T (Fig. 4L) was

Immunosugars



Carbasugars



C-glycosides



Fig. 4 Structure of glycomimetics used to inhibit GTs. The figure shows the three main types of glycomimetics: iminosugar (A–F); carbasugar (G–J); C-glycoside (K and L). A: iminosugar for β 1,4-GalT inhibition. B: 1-N-iminosugar based UDP-galactose analog for α 1,3-GalT. Pyrrolidines C and D form a complex between GDP and iminosugars, mimicking the transition state of fucosyltransferase reaction and inhibit α 1,3-Fuc-T. E: TS analogs which employ an iminosugar for mimicking the fucose moiety and phenol as GlcNAc-mimic. F mimics acceptor substrate and inhibits GCS. Carbasugars G, H and I inhibit GTs competitively *i.e.*, G targets β 1,4-GalT; H and I inhibit α 1,3-4Fuc-T. J: Transition state analog for ST. K: C-glycoside based analog mimicking the hypothetical transition state for β -galactosyltransferase. L: potent inhibitor for both donor (UDP-Fuc) and acceptor (LacNAc) substrates of α 1,3-Fuc-T.

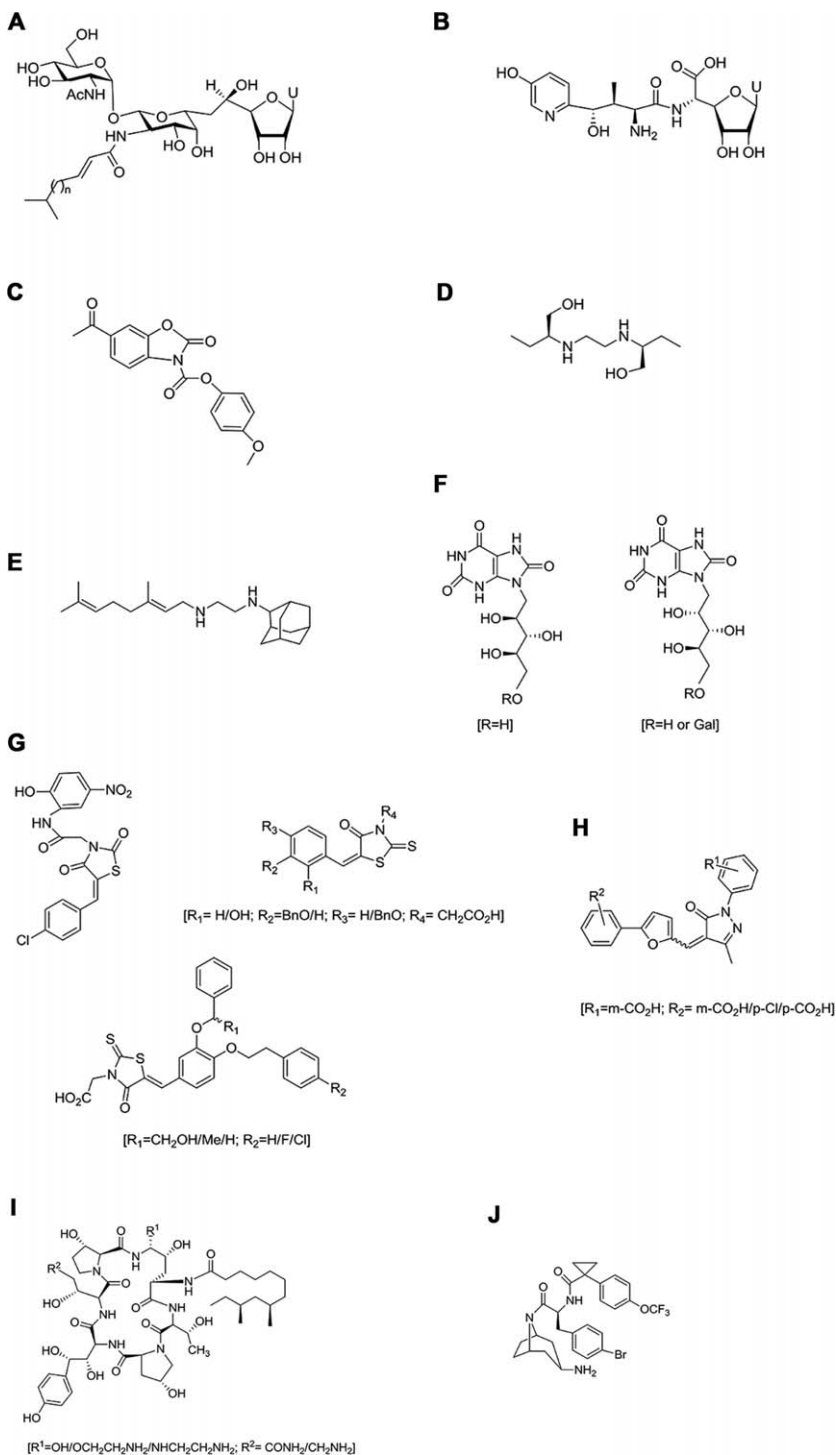
developed by Pasquarell and coworkers⁸³ displaying mixed type inhibition for both donor GDP-Fuc and acceptor substrate LacNAc and represents first member of promising class of simple and selective inhibitors.

7 Alternate chemotype analogs

Most existing GT inhibitors are either donor or acceptor analogs with limited potential for further development due to intrinsic unfavourable properties such as lack of cell penetration and limited chemical stability. Furthermore, the substrate based inhibitors are only accessible *via* multistep synthesis, which compromises their practical utility. These limitations of substrate analogs created interest among chemists to identify alternate inhibitor chemotypes, which are not structurally derived from GT substrates and may possess drug like properties, to explore the potential of this important enzyme family as therapeutic targets. In the past, GT inhibitor discovery suffered due to lack of 3D structural data and relative lack of operational assays for glycosyltransferases. However, the recent advances in available X-ray structures, sensitive high throughput screening (HTS) assays and techniques in medicinal chemistry contributed significantly to the growing number of new inhibitor chemotypes.

Tunicamycin (Fig. 5A), a mixture of homologous nucleoside antibiotics, was the first chemical tool for inhibition of N-glycosylation.⁸⁴ Tunicamycin interacts with wide range of molecular targets including GlcNAc-1-phosphotransferase and the glycosyltransferase oligosaccharide transferase (OST) in eukaryotes. Tunicamycin also exhibits effective inhibition in non-eukaryotes, *e.g.*, it inhibits the incorporation of glycans into hepatitis C virus glycoproteins E1 and E2 by intervening in their N-glycosylation at an early stage, without any effect on core protein expression.⁸⁵ McCarthy *et al.*⁸⁶ reported nikkomycin as a potent inhibitor of fungal GlcNAc-transferase chitin synthase, which mimics the natural donor UDP-GlcNAc (Fig. 5B). Either nikkomycin as well as tunicamycin and other uracil based inhibitors are screened against the uridine diphospho-N-acetylglucosamine: polypeptide β -N-acetylglucosaminyltransferase (OGT). OGT is responsible of O-GlcNAc glycosylation, which is abundant in cytoplasmic and nuclear proteins, and it is linked to insulin resistance, diabetic complications, cancers and neurodegenerative diseases.²⁶ Walker group made significant contributions to potent and selective OGT inhibitors.^{87–89} Intriguingly, one of OGT inhibitors is found to reduce the expression of FoxM1 in breast cancer cells and to inhibit its growth and invasion in cellular assays,^{87,88} emphasizing the potential of small molecule GT inhibitors as chemical tools in biological studies. Jiang *et al.*⁸⁹ identified another OGT inhibitor (Fig. 5C), which mimics the diphosphate moiety of endogenous ligand *via* carbamate formation and leads to complete enzyme inhibition by cross linking lysine and cysteine residues in the catalytic site.

Ethambutol, an established treatment for tuberculosis, is an inhibitor for arabinosyltransferase, an enzyme responsible for polymerisation of arabinose into arabinan of arabinogalactan in mycobacterial cell wall⁹⁰ (Fig. 5D). A combinatorial approach to find ethambutol analogs led to discovery of diamine SQ109 (Fig. 5E), exhibiting effective selectivity and efficacy in mouse models of tuberculosis and is advanced to early stage phase II trials^{91,92} Intriguingly, primary target of SQ109 appears to be



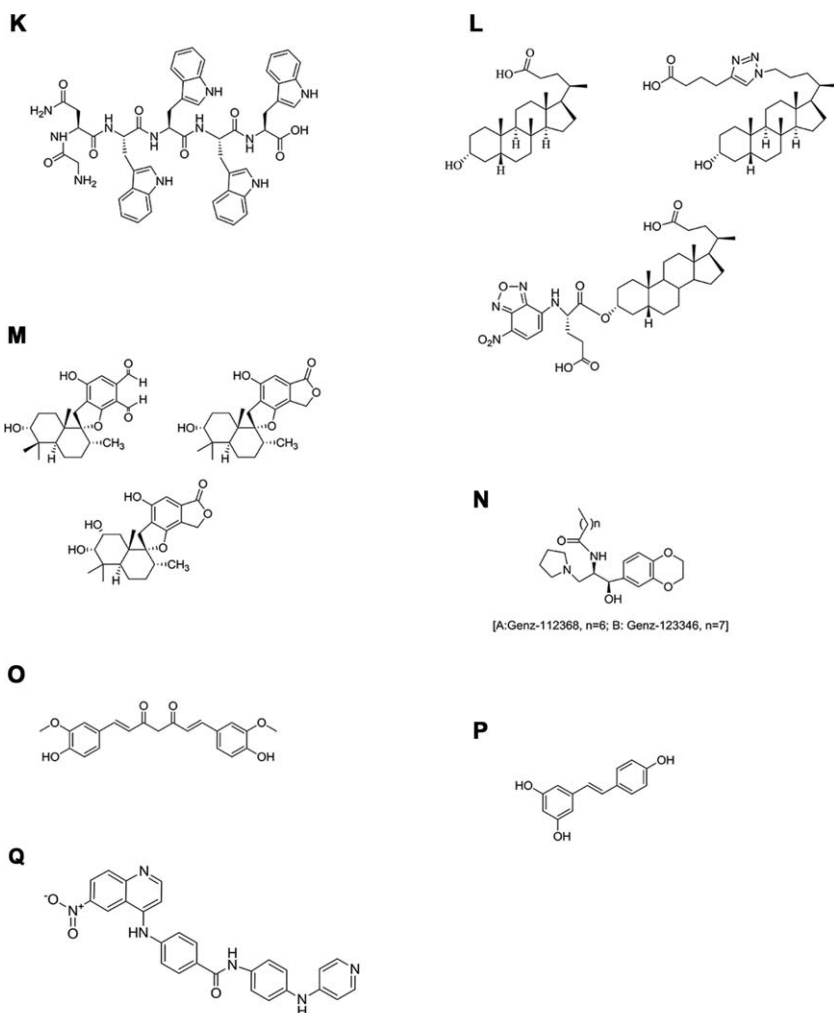


Fig. 5 Structure of alternate chemotype analogs. **A:** Tunicamycin interfere with the activities of GlcNAc-1-phosphotransferase and the glycosyltransferase oligosaccharide transferase (OST) in eukaryotes. **B:** Nikkomycin is a potent inhibitor of fungal GlcNAc-transferase chitin synthase. **C:** OGT inhibitor inhibits the enzyme by cross linking lysine and cysteine residues in the catalytic site. **D:** Ethambutal acts as an inhibitor for arabinosyltransferase. **E:** SQ109 (Ethambutal analog) targets a membrane transporter of an important lipid involved in cell wall biosynthesis instead of arabinosyltransferase. **F:** Uric acid derivatives inhibit GTB. **G:** Thiazolidinones are the potent inhibitors of bacterial GlcNAc-transferase MurG, parasitic Dol-P-mannose synthase (DPMS) and fungal Dol-P-Man:Proteinmannosyltransferase (PMT1). **H:** Pyrazol-3-one is a potent inhibitor of WaaC and MurG. **I:** Pneumocandin B₀ exhibits inhibitory activity against *C. albicans* β -(1,3)-D-glucan synthase. **J:** EXEL-0346 targets GCS. **K:** NH₂-GNWWWW inhibits N-glycan specific ST3Gal I and ST6Gal I. **L:** Lithocholic acid derivatives effectively inhibit STs. **M:** Stachybotrydial and its derivatives inhibit Fuc-Ts and STs. **N:** Eligustat is a potent inhibitor of GCS. **O:** Curcumin and **P:** Resveratrol appeared as putative inhibitors of Gal α 2,6-ST1, ST3GalSI3 and ST8. **Q:** T3-inh-1 targets GalNAc-T3.

membrane transporter of an important lipid involved in cell wall biosynthesis instead of for arabinosyltransferase as for its parent ethambutol.²⁶ Schaefer *et al.*⁹³ demonstrated that uric acid derivatives (Fig. 5F)

are the inhibitors of blood group B galactosyltransferase (GTB), and docking studies suggest that pentitol-linked uric acid mimics uracil, ribofuranoside and pyrophosphate group of UDP-Gal, the natural donor substrate for GTB. Thiazolidinones (Fig. 5G) are the heterocyclic compounds which have been identified as inhibitors for diverse GTs *e.g.*, thiazolidinones are the potent inhibitors of bacterial GlcNAc-transferase MurG, parasitic Dol-P-mannose synthase (DPMS) and fungal Dol-P-Man:Proteinmannosyltransferase (PMT1), respectively. Structurally similar to thiazolidinone, an aryl pyrazolone (Fig. 5H) is identified as a potent inhibitor of WaaC and MurG.^{94,95} Recently, structurally related aryl pyrazolones with inhibitory activities against LgtC are also identified.²⁶ Since LgtC is the key enzyme responsible for the expression of the digalactose motif in bacterial lipo-oligosaccharide coat, and also increases the serum resistance; the design of LgtC targeted inhibitors are of considerable interest in pharmacology, as emerging anti-virulence agents.⁹⁶

Another class of non-substrate chemotypes that have emerged as potential tool in the treatment of fungal infections are the echinocadins, a family of macrocyclic lipopeptides. Echinocadin, *e.g.*, pneumocandin B₀ (Fig. 5I) is demonstrated to exhibit inhibitory activity against *C. albicans* β -(1,3)-D-glucan synthase *in vitro*. L-733560, is the structurally modified analog of pneumocandin B₀, with 70-fold more potent inhibition against glucan synthase than the parent compound. Furthermore, the modification around macrocyclic scaffold in these inhibitors enhances their water solubility and makes them potent antifungal agents clinically.²⁶ A novel L-amino acid structure derived scaffold EXEL-0346 (Fig. 5J) is identified as a promising inhibitor of GCS by Richards *et al.*⁹⁷ Peptide based inhibitors are also explored for STs and hexapeptide NH₂-GNWWWW (Fig. 5K) is found to exhibit best inhibition profile for N-glycan specific ST3Gal I and ST6Gal I *in vitro*, emphasizing that this hexapeptide may have potential for development for a broad range of STs, regardless of their linkage specificity. Lithocholic acid derivatives (Fig. 5L) are identified which possess effective inhibition against STs, which are expressed at high level in metastatic cells.⁹⁸ Stachybotrydial and related compounds (Fig. 5M) containing decalin ring system are demonstrated active against GTs involved in terminal glycosylation of cell surface glycans including Fuc-Ts and STs.⁹⁹ Eligustat (Genz-112638, ceramide analog) is currently in clinical development for the treatment of type 1 Gaucher disease (Fig. 5N). A close homologue of eliglustat is Genz-123346, with longer acyl chain and slightly better inhibitory activity. Genz-123346 blocks the accumulation of glucosylceramide in cells with advantage to be oral available that make it suitable to be used successfully in a range of animal models, including type 2 diabetes, polycystic kidney disease and asthma. This GCS inhibitor is reported to protect rats against the cytotoxic effects of shiga toxin.²⁶ Using computational modelling, Grewal *et al.*¹⁰⁰ discerned the inhibitory activity of phytochemicals *i.e.*, curcumin (Fig. 5O) and resveratrol (Fig. 5P) against sialyltransferases *i.e.*, Gal α 2,6-ST1, ST3GalSI3 and ST8 involved in metastasis. Recently, Song and Linstedt¹⁰¹ identified quinoline based inhibitor (*i.e.*, T3-inh-1;

Fig. 5Q) which targets GalNAc-T3 mediated O-glycosylation and blocks cancer cell invasiveness. Furthermore, this inhibitor reduces FGF23 levels *in vitro* and *in vivo*, which could be a putative lead for the treatment of chronic kidney disease.¹⁰¹

At present, non-substrate like inhibitors have been reported for only a relatively small number of GTs, including GCS (lysosomal storage diseases), Fuc-Ts and STs (cancer), OGT (cancer and diabetes), as well as several bacterial enzymes *e.g.*, MraY, WaaC, MurG and LgtC. To realise the potential of alternate inhibitor chemotype as putative lead compounds for drug discovery and tools in glycobiology, further key steps should be addressed (a) establishing target selectivity, (b) cellular assays, and (c) mechanism of interaction of non-substrate inhibitor with target enzyme. Further progress with these issues addressed can change the long-followed perception of GTs as difficult targets for drug discovery.²⁶

8 Metabolic chain terminator

Recent advances in glycobiology verified that conservative modifications of sugar residues could be tolerated by GTs within cells.⁷ Based on this observation, another strategy for designing GT inhibitors has been recently developed which employs metabolic chain terminators, called dead-end substrates. These are unnatural donor or acceptor substrates that are incorporated into cells, leading to the biosynthesis of a corresponding unnatural nucleotide sugar. The modification can disrupt recognition by downstream enzymes or by receptor which governs the glycan function, thus terminating the metabolic chain. Pioneer study by Kayser *et al.*¹⁰² demonstrated that mannosamine inhibits glycosyl phosphatidylinositol anchor synthesis in mammalian cells. Gloster and colleagues¹⁰³ have reported a new rationally designed OGT inhibitor based on metabolic chain terminator strategy. O-GlcNAc is a nuclear and cytoplasmic modification of many eukaryotic proteins, which is generally installed and removed several times during lifetime of a given protein, and is regulated by combined activities of OGT and glycoside hydrolases. Inhibitors of OGT are of interest to delineate the role held by O-GlcNAc in physiology. 5-thio analog of GlcNAc and Ac-5S-GlcNAc are salvaged by cells and are processed by hexosamine biosynthetic pathway (HBP) into UDP-5S-GlcNAc, which act as inhibitors of OGT and reduce the level of cellular O-GlcNAc.⁹⁶ Consistent with this study, Zandberg *et al.*¹⁰⁴ demonstrated that peracetylated 5S-Fuc is taken up by HpG2 cells, and converts into GDP-5S-Fuc; which inhibits $\alpha 1,3/\alpha 1,4$ fucosyltransferases, limits SLe^x expression and selectin mediated cell adhesion. Although thio analogs are not evaluated *in vivo*, yet these findings imply the potential of thio analogs as putative tools to modulate glycosylation in physiology.

Glycoform aberration *e.g.*, enhanced expression of highly branched N-glycans is the characteristic of cancer cell proliferation and metastasis, which emphasize the need to explore the potential candidates for intervening such biosynthetic pathways for cancer therapy. There is plethora of reports available in literature suggesting the use of peracetylated 4F-GlcNAc for the reduction of selectin ligands *e.g.*, SLe^x,

both *in vitro* and *in vivo*.^{105–111} Fluorosugar analogs of UDP-GlcNAc and UDP-Gal are demonstrated to inhibit N-acetylglucosamine- and galactosyl-transferases.^{105,106} Dimitroff *et al.*¹⁰⁷ demonstrated that per-O-acetylated 2-acetamido-2, 4-dideoxy-4-fluoro-glucosamine blocks the formation of SLe^x to prevent leukocyte adhesion in CLA-T cells. Later, Dimitroff group reported an *in vivo* reduction of E- and P-selectin ligands by 4FGlcNAc and impairment of leukocyte trafficking to inflamed skin.¹⁰⁸ Barthel *et al.*¹¹⁰ proved that Gal-1 and E selectin ligands reduction in KG1a cells by 4FGlcNAc, is not due to chain termination by direct incorporation of 4FGlcNAc.¹⁰⁸ However, it is attributed to intervention by acetylated fluorosugar (*i.e.* 4FGlcNAc) in reducing UDP-GlcNAc pool required for synthesis of LacNAc and SLe^x moieties on N- and O-glycans. This study was further corroborated by Nishimura *et al.*¹¹¹ that per-O-acetylated GlcNAc with a fluorine atom at C4 position is the precursor of metabolic inhibitor of cancer cell proliferation, either through changes in the availability of UDP-GlcNAc or through an inhibitory effect of UDP-4FGlcNAc against different GlcNAc transferases. It was reported that peracetylated derivatives of fluorinated fucose (2F-Fuc) and sialic acid (3F-Neu5Ac) are membrane permeable and could be transformed into GDP-2F-Fuc and CMP-3F-Neu5Ac respectively.⁷⁰ As their cellular level increases, these derivatives interfere with natural biosynthetic pathway and act as competitive inhibitors of FucTs and STs and substantially reduce the expression of fucosylated and sialylated glycan epitopes *in vitro* and in cells (Fig. 6). Latter Macauley *et al.*¹¹² demonstrated that *in vivo* administration of 3F- Neu5Ac results in global blockade of sialylation in mice and impact is most evident on metabolically active tissues with high turnover of membrane proteins *viz.* liver and kidney. Okeley *et al.*¹¹³ reported that 2F-Fuc is orally bioavailable and blocks endogenous fucosylation without any apparent toxicity in high dose administered mice. Unlike 3F-Neu5Ac, inhibition of fucosylation is reversed on discontinuing 2F-Fuc treatment.^{70,112,113} Furthermore, the demonstration that oral 2F-Fuc inhibits vaso-occlusion, interactions of leukocytes and sickle red blood cells with endothelium, NF- κ B activation and adhesion molecule expression in transgenic sickle mice, suggests its propitious benefits in the treatment of sickle cell disease and other inflammatory conditions.¹¹⁴ The therapeutic potential of P-3F- Neu5Ac is explored by Bull *et al.*,¹¹⁵ which demonstrated that treatment of murine melanoma cells with this global sialyltransferase inhibitor readily deplete α 2,3/ α 2,6 linked sialoglycans; strongly impair their binding to extracellular matrix (ECM) components and migratory capacity *in vitro* without any cellular toxicity, and delay tumor growth *in vivo*. Recently, the same group reported that intravenous injection of P-3F- Neu5Ac- PLG-nanoparticles prevents metastasis formation in murine lung metastasis model.¹¹⁶ This combined glycobio-nano-technology based strategy to block sialoglycan expression to interfere with sialic acid- dependent processes in metastasis can be extended to other types of cancers, infection and inflammation. Van Wijk *et al.*¹¹⁷ identified peracetylated 6-F-GalNAc for its capacity to intervene with glycosaminoglycan (GAG) chain elongation or sulfation, and inhibit growth factor signalling and reduced *in vivo*

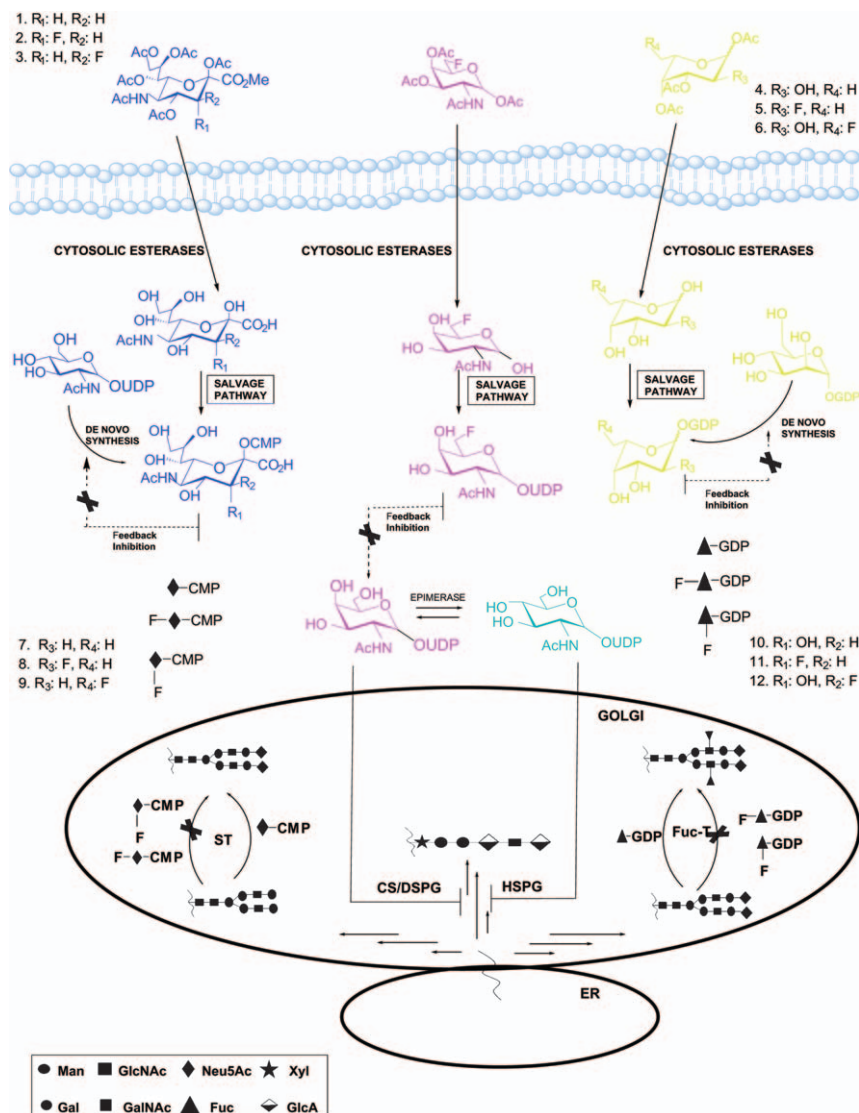


Fig. 6 Fluorinated monosaccharide [sialic acid, N-acetylgalactosamine (GalNAc) and fucose] analogs act as metabolic GT inhibitors. Peracetylated analogs of sialic acid [1–3] or GalNAc [6F-GalNAc (Ac)] or fucose [4–6] enter cell by passive diffusion and are converted into nucleotide sugars viz., sialic acid [7–9], GalNAc [UDP-6F-GalNAc] and fucose [10–12] through salvage pathways by cytosolic esterases. Fluorinated analogs [8–9], UDP-6F-GalNAc and [11–12] act as inhibitors of ST or GalNAcT or Fuc-T in Golgi apparatus, and are accumulated in the cell and interrupt *de novo* synthesis of GDP-fucose or UDP-GalNAc or CMP-Neu5Ac in cytosol via feedback loop. As UDP-GalNAc and UDP-GlcNAc are in equilibrium, a decrease in the level of UDP-GalNAc may lead to reduced level of UDP-GlcNAc, and may be manifested as reduction in biosynthesis of glycosaminoglycan (GAGs) or other glycans e.g., chondroitin sulfate (CS), dermatan sulfate proteoglycan (DSPG), and heparan sulfate proteoglycan (HSPG).

angiogenesis. GAGs (*i.e.*, chondroitin/dermatan/heparin sulphate) are polysaccharides present in ECM and virtually on every mammalian cell surface and any alteration in their quantity or structure is associated with

diseases including cancer. They demonstrated that 6-F- GalNAc (Ac₃) is taken up by SKOV3 cells and is transformed into UDP-6-F-GalNAc by GalNAc salvage pathway, depleting natural UDP-GalNAc pool, results in the inhibition of GAGs biosynthesis (Fig. 6).¹¹⁷ These findings may motivate on exploring acetylated fluorosugars as potential inhibitors for other families of glycosyltransferases to elucidate their role in mammalian physiology.

9 Conclusion

GTs play key roles in many biological processes of human health and disease such as inflammation, infection, cancer and metabolic disorders; and thus, are potential therapeutic targets. Despite systematic efforts in this area, few high-selective and potent inhibitors have been developed and progress in GT inhibitors has been slower than glycosidases. General and class promiscuity is also a big concern with most of GT inhibitors. Conventional donor analogs including transition state mimics have been proposed as GT inhibitors, but are limited by poor membrane permeability. Library screening may pave a new way to identify leads to generate non-substrate inhibitor chemotypes. Recently, metabolic inhibitors of GTs emerged as a new promising inhibitor based strategy for drug development. Future studies including structural data, optimization of existing compounds, improvement of rationale approaches and *in vivo* testing of these inhibitors could unravel the physiological role of various glycoconjugates and mechanism of GTs under pathological conditions, which will be a lead for designing GT inhibitor-based drugs for therapeutic purpose.

References

- 1 A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, *In Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press, NY, 2nd, 2009.
- 2 S. Hakomori, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 10231.
- 3 R. Grewal and A. Mahmood, *Mol. Cell. Biochem.*, 2009, **331**(1–2), 19.
- 4 R. Grewal and A. Mahmood, *Ann. Gastroenterol.*, 2009, **2**(3), 178.
- 5 R. Grewal and A. Mahmood, *Alcohol*, 2010, **44**, 515.
- 6 M. B. Lazarus, Y. Nam, J. Jiang, P. Sliz and S. Walker, *Nature*, 2011, **27**(469), 564.
- 7 T. M. Gloster and D. J. Vocadlo, *Nat. Chem. Biol.*, 2012, **8**, 683.
- 8 V. Rojas-Cervellera, A. Ardevol, M. Boero, A. Planas and C. Rovira, *Chem. - Eur. J.*, 2013, **19**, 14018.
- 9 D. Jove, D. Albesa, M. Giganti, P. Jackson, M. Alzari and M. E. Guerin, *Glycobiology*, 2014, **24**(2), 108.
- 10 T. M. Gloster, *Curr. Opin. Struct. Biol.*, 2014, **28**, 131.
- 11 V. Lombard, H. Golaconda-Ramulu, E. Drula, P. M. Coutinho and B. Henrissat, *Nucleic Acids Res.*, 2014, **42**, D490.
- 12 G. M. Rishton, *Drug Discovery Today*, 2003, **8**, 86.
- 13 T. Kajimoto and M. Node, *Curr. Top. Med. Chem.*, 2009, **9**(1), 13–33.
- 14 D. Giuseppe and P. Maffioli, *Arch. Med. Sci.*, 2012, **8**(5), 899.
- 15 T. M. Gloster and G. J. Davies, *Org. Biomol. Chem.*, 2010, **8**(2), 305.
- 16 B. P. Rempel and S. G. Withers, *Glycobiology*, 2008, **18**(8), 570.

- 17 P. Compain and A. Bodlenner, *ChemBioChem.*, 2014, **15**(9), 1239.
- 18 S. G. Gouin, *Chemistry*, 2014, **20**(37), 11616.
- 19 R. Zelli, J. F. Longevial, P. Dumy and A. Marra, *New J. Chem.*, 2015, **39**, 5050.
- 20 N. Kanfar, E. Bartolami, R. Zelli, A. Marra, J. Y. Winum, S. Ulrich and P. Dummy, *Org. Biomol. Chem.*, 2015, **13**, 9894.
- 21 S. H. Khan, K. S. Matta, In *Glycoconjugates: Composition: Structure, and Function*, ed. H. J. Allen and E. C. Kisailus, CRC Press, NY, 1992, *Recent advances in the development of potential inhibitors of glycosyltransferases*, 361.
- 22 K. H. Zung, R. R. Schimdt, In *Carbohydrate based Drug Discovery*, ed. C. H. Wong, Wiley, NY, 2nd, 2003, *Glycosyltransferase inhibitors*, 609.
- 23 J. D. Esko, C. Bertozzi, In *Essentials of Glycobiology*, ed. A. Varki, R. D. Cummings, J. D. Esko, *et al.*, Cold Spring Harbour Laboratory Press, NY, *Chemical tools for inhibiting glycosylation*, 2009, 50.
- 24 T. Kajimoto, M. Node, 2009, **19**, 3179.
- 25 P. Merino, I. Delso, T. Tejero, M. Ghirardello and V. Juste-Navarro, *Asian J. Org. Chem.*, 2016, **5**(12), 1413.
- 26 L. Tedaldi and G. K. Wagner, *MedChemComm*, 2014, **5**, 1106.
- 27 R. Jorgenson, T. Pesnot, J. H. Lee, M. M. Palcic and G. K. Wagner, *J. Biol. Chem.*, 2013, **288**, 26201.
- 28 N. J. Agard and C. R. Bertozzi, *Acc. Chem. Res.*, 2002, **42**(6), 788.
- 29 D. J. Gill, H. Clausen and F. Bard, *Trends Cell Biol.*, 2011, **21**(3), 149.
- 30 A. A. Yakhnina and Z. Gitai, *J. Bacteriol.*, 2013, **195**(19), 4527.
- 31 K. W. Moremen, M. Tiemeyer and A. V. Nairn, *Nat. Rev. Mol. Cell Biol.*, 2014, **13**(7), 448.
- 32 R. Wang, D. H. Steensma, Y. Takaoka, J. W. Yun, T. Kajimoto and C. H. Wong, *Bioorg. Med. Chem.*, 1997, **5**(4), 661.
- 33 X. Qian, M. M. Palcic, In *carbohydrates in chemistry and biology: A comprehensive handbook*, ed. B. Ernst, G. W. Hart and P. Sana, Wiley-VCH (NY), 2000, 293.
- 34 C. Saotome, C. H. Wong and O. Kanie, *Chem. Biol.*, 2001, **8**, 1061.
- 35 P. Compain and O. R. Martin, *Bioorg. Med. Chem.*, 2001, **9**, 3077.
- 36 P. Compain and O. R. Martin, *Curr. Top. Med. Chem.*, 2003, **3**, 541.
- 37 H. U. Yanan, J. S. Helm, L. Chem, C. Ginsberg, B. Gross, B. Kraybill, K. Tiyanont, X. Fang, T. Wu and S. Walker, *Chem. Biol.*, 2004, **5**(11), 703.
- 38 L. V. Lee, M. L. Mitchell, S. J. Huang and V. V. Fokin, *J. Am. Chem. Soc.*, 2003, **125**(32), 9588.
- 39 H. Hinou, X. L. Sun and Y. Ito, *J. Org. Chem.*, 2003, **68**, 5602.
- 40 J. Zuegg, C. Muldoon, G. Adamson, D. McKeveney, G. Le Thanh and R. Premraj, *et al.*, *Nat. Comm.*, 2015, **6**, 7719.
- 41 K. A. Winans and C. R. Bertozzi, *Chem. Biol.*, 2002, **9**, 113.
- 42 M. C. Schuster, D. A. Mann, T. J. Buchholz, K. M. Johnson, W. D. Thomas and L. L. Kiessling, *Org. Lett.*, 2003, **5**, 1407.
- 43 T. Pesnot, R. Jorgensen, M. M. Palcic and G. K. Wagner, *Nat. Chem. Biol.*, 2010, **6**, 321.
- 44 V. Kanabar, J. Jiang, X. Nie, I. Panina, K. Descroix, F. Man, S. C. Pitchford, C. P. Page and G. K. Wagner, *Glycobiology*, 2016, **26**(10), 1059.
- 45 R. Schworer and R. R. Schmidt, *J. Am. Chem. Soc.*, 2002, **124**, 1632.
- 46 L. V. Lee, M. L. Mitchell, S. J. Huang and V. V. Fokin, *J. Am. Chem. Soc.*, 2003, **125**(32), 9588.
- 47 M. Izumi, H. Yuasa and H. Hashimoto, *Curr. Top. Med. Chem.*, 2009, **9**, 87.
- 48 K. Okazaki, S. Nishigaki, F. Ishizuka, Y. Kajihara and S. Ogawa, *Org. Biomol. Chem.*, 2003, **1**, 2229.
- 49 M. C. Galan, A. P. Venot, R. S. Phillips and G. J. Boons, *Org. Biomol. Chem.*, 2004, **2**, 1376.

- 50 S. F. Kuan, J. C. Byrd, C. Basbaum and Y. S. Kim, *J. Biol. Chem.*, 1989, **264**, 19271.
- 51 J. R. Brown, M. M. Li, R. Fuster, N. Varki, C. A. Glass and J. D. Esko, *Clin. Cancer Res.*, 2003, **12**, 2894.
- 52 S. Laferte, N. W. Chan, K. Sujino, T. L. Lowary and M. M. Palcic, *Eur. J. Biochem.*, 2000, **267**, 4840.
- 53 J. R. Brown, F. Yang, A. Sinha, B. Ramakrishnan, Y. Tor, P. K. Qasba and J. D. Esko, *J. Biol. Chem.*, 2009, **284**, 4952.
- 54 P. Butaye, L. A. Devriese and F. Haesebrouck, *Antimicrob. Agents Chemother.*, 2001, **45**(5), 1374.
- 55 U. M. Unligil and J. M. Rini, *Curr. Opin. Struct. Biol.*, 2000, **10**(5), 510.
- 56 C. Breton, L. Snajdrova, C. Jeanneau, J. Koca and A. Imberty, *Glycobiology*, 2006, **16**, 29R.
- 57 M. M. Palcic, L. D. Heerze, O. P. Srivastava and O. Hindsgaul, *J. Biol. Chem.*, 1989, **264**, 17174.
- 58 B. M. Heskamp, G. H. Veeneman, G. A. van der Marel, C. A. A. van Boeckel and J. H. van Boom, *Tetrahedron*, 1995, **51**, 8397.
- 59 B. M. Heskamp, G. A. van der Marel and J. H. van Boom, *J. Carbohydr. Chem.*, 1995, **14**, 1265.
- 60 M. Izumi, S. Kaneko, H. Yuasa and H. Hashimoto, *Org. Biomol. Chem.*, 2006, **4**, 681.
- 61 U. Galili, *Biochimie*, 2001, **83**, 557.
- 62 H. Hashimoto, T. Endo, M. Natori, Y. Kajihara, H. Yuasa, *2nd Euroconference on carbohydrate mimics*, 1996, C11.
- 63 H. Hashimoto, T. Endo and Y. Kajihara, *J. Org. Chem.*, 1997, **62**(7), 1914.
- 64 T. Endo, H. Hasegawa, Y. Kadota, H. Yuasa, H. Hashimoto, K. Furukawa, *et al. In abstracts of 21st Japanese Carbohydrate Symposium*, Nagoya, Japan, 2000, A1–A02.
- 65 H. Hinou, X. L. Sun and Y. Ito, *Tetrahedron Lett.*, 2002, **43**, 9147.
- 66 K. H. Jung, R. Schworer and R. R. Schmidt, *Trends Glycosci. Glycotechnol.*, 2003, **15**, 275.
- 67 S. Hanashima, K. I. Inamori, S. Manabe, N. Taniguchi and Y. Ito, *Angew. Chem. Int. Ed.*, 2004, **43**, 5674.
- 68 S. Hanashima, K. I. Inamori, S. Manabe, N. Taniguchi and Y. Ito, *Chem. – Eur. J.*, 2006, **12**, 3449.
- 69 M. D. Burkart, S. P. Vincent, A. Duffels, B. W. Murray, S. V. Ley and C. H. Wong, *Bioorg. Med. Chem.*, 2000, **8**, 1937–1946.
- 70 C. D. Rillahan, A. Antonopoulos, C. T. Lefort, R. Sonon, P. Azadi and K. Ley, *et al., Nat. Chem. Biol.*, 2012, **8**(7), 661.
- 71 M. Raab, S. Kozmon and I. Tvaroska, *Carbohydr. Res.*, 2005, **340**, 1051.
- 72 L. Sihelnikova, S. Kozmon and I. Tvaroska, *Collect. Czech. Chem. Commun.*, 2008, **73**(5), 591.
- 73 C. Saotome, Y. Kanie, O. Kanie and C. H. Wong, *Bioorg. Med. Chem.*, 2000, **8**(9), 2249.
- 74 Y. J. Kim, M. Ichikawa and Y. J. Ichikawa, *J. Am. Chem. Soc.*, 1999, **121**, 5829.
- 75 L. Qiao, B. W. Murray, M. Shimazaki, J. Schultz and C. H. J. Wong, *J. Am. Chem. Soc.*, 1996, **118**, 7653.
- 76 T. D. Butters, L. A. G. M. van den Broek, G. W. J. Fleet, T. M. Krulle, M. R. Wormald and R. A. Dwek, *et al., Tetrahedron: Asymmetry*, 2000, **11**, 113.
- 77 H. R. Mellor, F. M. Platt, R. A. Dwek and T. D. Butters, *Biochem. J.*, 2003, **374**, 307.
- 78 I. Jefferies and R. R. Rowen, *Bioorg. Med. Chem. Lett.*, 1997, **7**, 1171.

- 79 S. Cai, M. R. Stroud, S. Hakomori and T. Toyokuni, *J. Org. Chem.*, 1992, **57**, 6693.
- 80 T. Miura, T. Kajimoto, M. Mimbo, K. Yamagishi, J. C. Inokuchi and C. H. Wong, *Bioorg. Med. Chem.*, 1998, **6**, 1481.
- 81 C. Schaub, B. Muller and R. Schmidt, *Eur. J. Org. Chem.*, 2000, **9**, 1745.
- 82 R. R. Schmidt and K. Frische, *Bioorg. Med. Chem. Lett.*, 1993, **3**, 1747.
- 83 C. Pasquarell, S. Picasso, R. Demange, M. Malissard, E. G. Berger and P. J. Vogel, *J. Org. Chem.*, 2000, **65**, 4251.
- 84 J. S. Tkacz and J. O. Lampen, *Biochem. Biophys. Res. Commun.*, 1975, **65**, 248.
- 85 B. Imperiali, S. E. O. Connor, T. Hentrickson and C. Kellenberger, *Pure Appl. Chem.*, 1999, **71**, 777.
- 86 P. J. McCarthy, P. F. Troke and K. Gull, *J. Gen. Microbiol.*, 1985, **131**(4), 775.
- 87 B. J. Gross, B. Kraybill and S. Walker, *J. Am. Chem. Soc.*, 2005, **127**, 14588.
- 88 B. J. Gross, J. G. Swoboda and S. Walker, *J. Am. Chem. Soc.*, 2008, **130**, 440.
- 89 J. Jiang, M. B. Lazarus, L. Pasquina, P. Sliz and S. Walker, *Nat. Chem. Biol.*, 2012, **8**, 72.
- 90 A. E. Belanger, G. S. Besra, M. E. Ford, K. Mikusova, T. Belisle, P. J. Brennan, J. M. Inamine, *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 11919.
- 91 C. E. Barry and J. S. Blanchard, *Curr. Opin. Chem. Biol.*, 2010, **14**(4), 456.
- 92 K. A. Sacksteder, M. Protopopova, C. E. Barry, K. Andries and C. A. Nacy, *Future Microbiol.*, 2012, **7**(7), 823.
- 93 K. Schaefer, N. Sindhuwinata, T. Hackl, M. P. Kotzler, F. C. Niemeyer, M. M. Palcic, T. Peters and B. Meyer, *J. Med. Chem.*, 2013, **56**(5), 2150.
- 94 P. R. Reeves, M. Hobbs, M. A. Valvano, M. Skurunil, V. Gerusz and G. Le Fralliec, *et al.*, *Trends Microbiol.*, 1996, **4**, 495.
- 95 Y. Hu, J. S. Helm, L. Chen, C. Ginsberg, B. Gross and B. Kraybill, *et al.*, *Chem. Biol.*, 2004, **11**, 703.
- 96 K. Persson, H. D. Ly, M. Dieckelmann, W. W. Wakarchuk, S. G. Withers and N. C. J. Strynadka, *Nat. Struct. Mol. Biol.*, 2001, **8**, 166.
- 97 S. Richards, C. J. Larson, E. S. Koltun, A. Hanel and V. Chan, *et al.*, *J. Med. Chem.*, 2012, **55**, 4322.
- 98 C. H. Chiang, C. H. Wang, H. C. Chang, S. V. More, W. S. Li and W. C. Hung, *J. Cell. Physiol.*, 2010, **223**, 492.
- 99 T. W. Lin, W. W. Chang, C. C. Chen and Y. C. Tsai, *Biochem. Biophys. Res. Commun.*, 2005, **331**, 953.
- 100 R. K. Grewal, A. Pathania and M. Chawla, *Int. J. Pharma Biol. Sci.*, 2014, **5**(3), B691.
- 101 L. Song and A. D. Lindstedt, *eLife*, 2017, DOI: 10.7554/eLife.24051.
- 102 H. Kayser, R. Zeitler, C. Kannicht, D. Grunow, R. Nuck and W. Reutter, *J. Biol. Chem.*, 1992, **267**(24), 16934.
- 103 T. M. Gloster, W. F. Zandberg, J. E. Heinonen, D. L. Shen, L. Deng and D. J. Vocadlo, *Nat. Chem. Biol.*, 2011, **7**(3), 174.
- 104 W. F. Zandberg, J. Kumarasamy, B. M. Pinto and D. J. Vocadlo, *J. Biol. Chem.*, 2012, **287**, 40021.
- 105 T. Hayashi, B. W. Murray, R. Wang and C. H. Wong, *Bioorg. Med. Chem.*, 1997, **5**, 497.
- 106 M. D. Burkart, S. P. Vincent, A. Duffels, B. W. Murray, S. V. Ley and C. H. Wong, *Bioorg. Med. Chem.*, 2000, **8**, 1937.
- 107 C. J. Dimitroff, R. J. Bernacki and R. Sackstein, *Immunobiology*, 2003, **101**, 602.
- 108 C. J. Dimitroff, T. S. Kupper and R. Sackstein, *J. Clin. Invest.*, 2003, **112**, 1008.
- 109 M. E. Gainers, L. Descheny, S. R. Barthel, L. Liu, M. A. Wurbel and C. J. Dimitroff, *J. Immunol.*, 2007, **179**, 8509.

- 110 S. R. Barthel, A. Antonopoulos, F. Cedeno-Laurent, L. Schaffer, G. Hernandez, S. A. Patil, S. J. North, A. Dell, K. L. Matta, S. Neelamegham, S. M. Haslam and C. J. Dimitroff, *J. Biol. Chem.*, 2011, **286**(24), 21717.
- 111 S. I. Nishimura, M. Hato, S. Hyugaji, F. Feng and M. Amano, *Angew. Chem., Int. Edn. Engl.*, 2012, **51**(14), 3386.
- 112 M. S. Macauley, B. M. Arlian, C. D. Rillahan, P. C. Pang, N. Bortell and M. C. Marcondes, *et al.*, *J. Biol. Chem.*, 2014, **289**(51), 35149.
- 113 N. M. Okeley, S. C. Alley, M. E. Anderson, T. E. Boursalian and P. J. Burke, *et al.*, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 5404.
- 114 J. D. Belcher, C. Chen, J. Nguyen, F. Abdulla, P. Nguyen and M. Nguyen, *et al.*, *Plos One*, 2015, **10**(2), e0117772, DOI: 10.1371/journal.pone.0117772.
- 115 C. Bull, T. J. Boltje, M. Wassink, A. M. de Graaf, F. L. van Delft and M. H. den Brok, *et al.*, *Mol. Cancer Ther.*, 2013, **12**(10), 1935.
- 116 C. Bull, T. J. Boltje, E. A. van Dinther, T. Peters, A. M. de Graaf and J. H. Leusen, *ACS Nano*, 2015, **9**(1), 733.
- 117 X. M. van Wijk, R. Lawrence, V. L. Thijssen, S. A. van den Broek, R. Troost, M. van Scherpenzeel, M. Naidu, A. Oosterhof, A. W. Griffioen, D. J. Lefeber, F. L. van Delft and T. H. van Kuppevelt, *FASEB J.*, 2015, **29**(7), 2993.

Targeting protein-carbohydrate interactions in plant cell-wall biodegradation: the power of carbohydrate microarrays

Diana O. Ribeiro, Benedita A. Pinheiro, Ana Luísa Carvalho and Angelina S. Palma*

DOI: 10.1039/9781788010641-00159

The plant cell-wall is constituted by structurally diverse polysaccharides. The biodegradation of these is a crucial process for life sustainability. Cellulolytic microorganisms are highly efficient in this process by assembling modular architectures of carbohydrate-active enzymes with appended non-catalytic carbohydrate-binding modules (CBMs). Carbohydrate microarrays offer high-throughput and sensitive tools for uncovering carbohydrate-binding specificities of CBMs, which is pivotal to understand the function of these modules in polysaccharide biodegradation mechanisms. Features of this technology will be here briefly reviewed with highlights of microarray approaches to study plant-carbohydrates and CBM-carbohydrate interactions, along with an overview of plant polysaccharides and microorganisms strategies for their recognition.

1 Structural diversity of plant cell-wall polysaccharides

The plant cell-wall is an intricate structure composed in its majority by complex polysaccharides and a smaller number of structural proteins. The composition in polysaccharides is highly variable, depending if the plant cell-wall is expanding (primary cell-wall) or if its role is to give additional structural support to the cell (secondary plant cell-wall). It also differs between species, with distinct chemical compositions among the cell-walls of grass and flowering plant species.^{1,2} The wider molecular and functional diversity of the polysaccharides is mainly observed in the primary cell-wall with their configurations changing throughout the plant cell development, expansion and division.³

In plant cell-walls, microfibrils of the major polysaccharide cellulose form a network embedded in a matrix of various complex polysaccharides, such as hemicelluloses, β -glucans and pectins. The hemicelluloses are interconnected with cellulose reinforcing the strength and resilience of the network, while the hydrated gels composed of pectin that intercalate this network, determine the porosity and thickness of the cell-wall (Fig. 1). The entire structure is maintained by non-covalent interactions, both spontaneous physico-chemical interactions and enzymatic-crosslinking, that exist between these polysaccharides.^{2,4}

The plant cell-wall polysaccharides are structurally diverse (Fig. 2). Cellulose is composed of aligned linear homopolymers of $(1 \rightarrow 4)$ - β -D-linked

UCIBIO-REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal.
E-mail: angelina.palma@fct.unl.pt

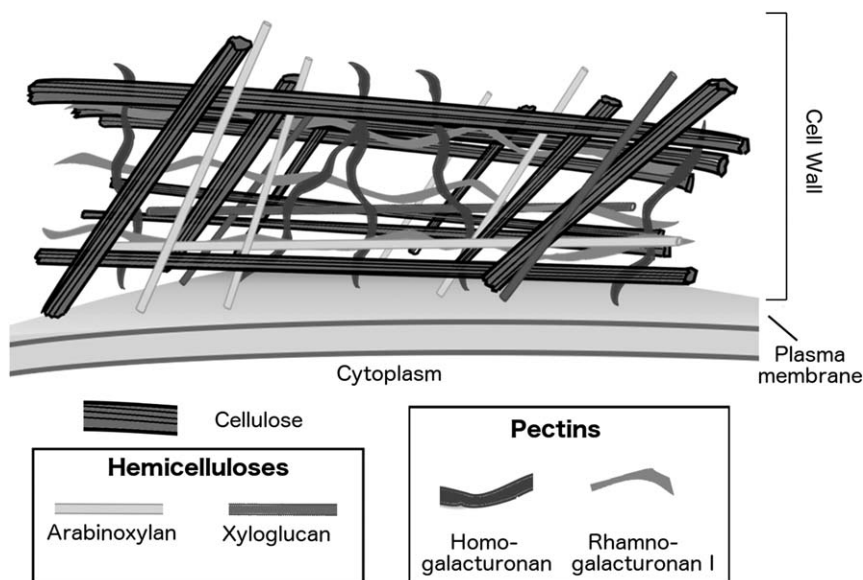


Fig. 1 Illustrative representation of the diversity of major polysaccharides in the plant cell wall. In combination with hemicelluloses, cellulose microfibrils form a network interspersed by pectin polysaccharides. The main hemicellulose polysaccharides found in plant cell walls are xyloglucan (dicot species) and arabinoxylan (grasses).² The major pectin polysaccharides are rhamnogalacturonan-I and homogalacturonan.²

glucosyl residues, organized in sheets packed in a “parallel-up” fashion forming a structure that is in its majority crystalline.⁵ Hemicelluloses maintain the (1→4)- β -D-linked backbone but are structurally more complex and, in addition to glucose, are composed of other residues such as mannose and xylose in linear or branched sequences.

Examples of hemicelluloses include xyloglucans, xylans, mannans, glucomannans and (1→3)- β -D-(1→4)- β -D-glucans ((1→4)- β -D-linked glucans with interspersed single (1→3)- β -D-linkages). All hemicelluloses have significant structural similarity as their backbone residues share the same equatorial configuration at C1 and C4 positions. While xyloglucans are widespread in land plants, the others are more species specific: xylans (dicots and commelinid monocots), mannans (charophytes) and (1→3)- β -D-(1→4)- β -D-glucans (grasses).⁶ Callose is a (1→3)- β -D-glucan that is widespread and occurs in specialized walls or wall-associated structures, specifically at stages of differentiation.⁷ Pectins are a structurally diverse group of polysaccharides constituted by galacturonic acid (GalpA) in their backbone sequences. The predominant pectins in the primary plant cell-wall are: homogalacturonan, a polysaccharide with an unsubstituted backbone of (1→4)- α -D-linked GalpA residues and rhamnogalacturonan-I, a polysaccharide with a backbone of the repeating disaccharide [2)- α -L-Rhap-(1→4)- α -D-GalpA-(1→)]_n, substituted at the rhamnose residue with different structural domains, such as galactans, arabinans or arabinogalactans (Fig. 2). Other pectins such as xylogalacturonan, which has a homogalacturonan backbone substituted by xylose, and rhamnogalacturonan-II, a highly-ramified polysaccharide

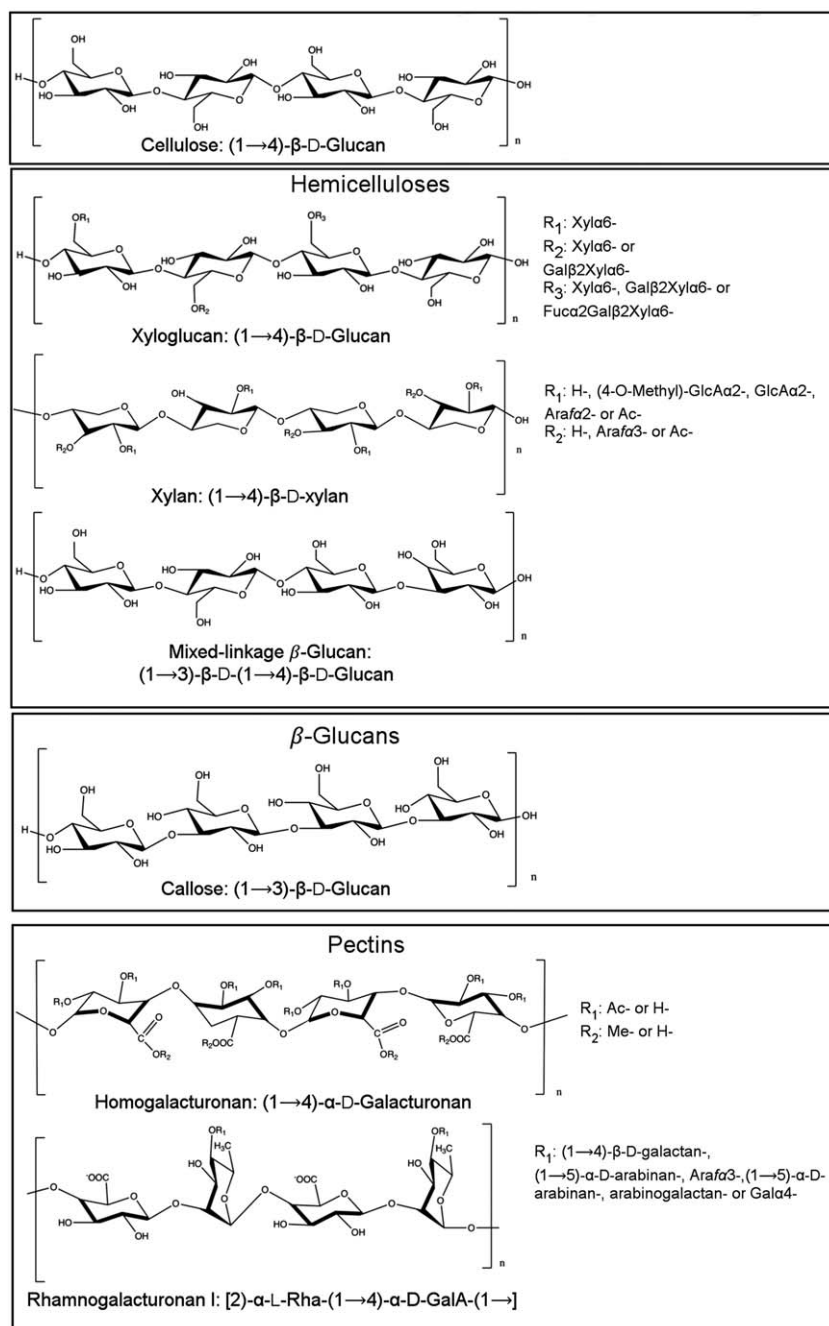


Fig. 2 Examples of structures of polysaccharides found in the plant cell wall. The Haworth conformational structure of the main chain representative tetrasaccharide sequences is represented; –R, possible ramification position; the different possible sequences of the ramifications to the main backbone chain are depicted. In the mixed linkage β-glucans, the (1→3)-β-D-linkages separate segments of 2 up to 14 glucose residues linked with (1→4)-β-D-linkages.⁷

with a homogalacturonan backbone comprising 7 to 9 (1 → 4)- α -D-linked GalpA residues are present in smaller amounts. Rhamnogalacturonan-II is the most structurally complex cell-wall polysaccharide as it has 12 different monosaccharide residues interconnected by more than 20 glycosidic types of linkages.⁸

As consequence of this structural diversity, cellulolytic microorganisms that degrade the plant cell-wall, have developed a consortium of Carbohydrate Active Enzymes (CAZymes) appended to Carbohydrate Binding Modules (CBMs) for which the diversity of specificities matches the variety of polysaccharides.

2 Cellulolytic microorganisms express proteomes highly efficient in plant cell-wall biodegradation

The machinery for degrading plant cell-wall polysaccharides differs from anaerobic to aerobic microorganisms, however the modular organization of the polysaccharide-degrading enzymes is maintained in both.

Due to energetic constraints and competition between the species found in anaerobic environments, most anaerobic cellulolytic microorganisms have arranged an efficient but rather elaborate system, where the produced and secreted CAZymes, such as endoglucanases, exoglucanases and β -glucosidases, are assembled in supramolecular complexes (molecular weight >3 MDa), termed the 'cellulosome' (see references^{9–11} for a comprehensive review on cellulosomes) (Fig. 3A). Cellulosomes show different levels of complexity and are mainly localized in the cell surface and glycocalyx matrix.¹⁰ These are known as integrating systems and are composed by one or more scaffoldins, where CAZymes are integrated and brought to the vicinity of the substrate. The scaffoldin is a structural subunit composed of several cohesin modules that bind their binding partners, the dockerin modules, present in CAZymes and other relevant proteins. Most of the scaffoldins have in their modular structure a CBM from family 3a, that specifically binds to recalcitrant cellulosic substrates.¹¹ These multi-enzyme complexes can be attached to the bacteria surface and display very complex and dynamic assemblies. One example is the cellulosome from *Clostridium thermocellum*, which has a primary scaffoldin comprising nine highly conserved type I cohesins, which allow the incorporation of different CAZymes and associated CBMs, through their type I dockerins (Fig. 3A). To attach the scaffoldin subunit to the surface of the bacterial cell, membrane-associated proteins are bound to a type II cohesin.¹⁰ Bacteria of the genera *Acetivibrio*, *Clostridium*, *Ruminococcus*, *Thermotoga*^{12,13} and fungi of the genera *Neocallimastix*, *Piromyces* and *Orpinomyces*¹⁰ are examples of anaerobic cellulolytic microorganisms. The ecosystems where anaerobes are found to degrade plant polysaccharides to soluble sugars are as diverse as soils, sediments or water bodies. Recently, much attention has been given to the polysaccharide-degrading systems from anaerobic organisms that reside in the

digestive tracts of invertebrates and vertebrates.^{14,15} These offer novel systems for studying carbohydrate-recognition and are important for communication with the host, such as the human host, promoting health and nutritional benefits.

Aerobic microorganisms secrete large quantities of CAZymes to the environment, organized in much simpler non-integrating systems. Bacteria from genera *Bacillus*, *Micromonospora*, *Cellvibrio* and *Pseudomonas*¹⁶ and fungi from genera *Aspergillus*¹⁷ are examples of aerobic cellulolytic microorganisms. The enzymatic activities of the CAZymes are still complementary, showing strong synergy in the degradation of plant cell-walls.

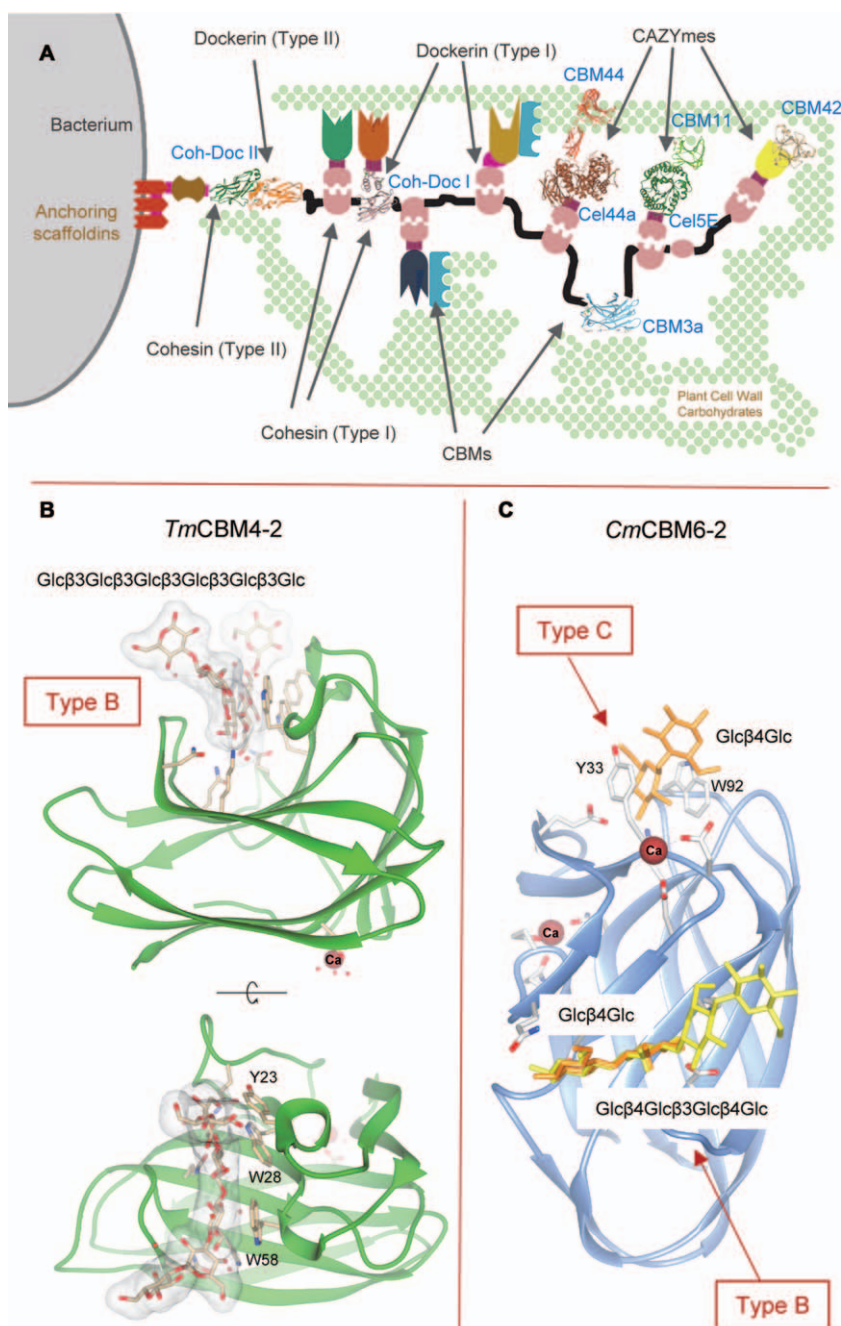
It is worth emphasizing in this section the recently identified lytic polysaccharide monooxygenases (LPMO) as key players in the first steps of plant biomass degradation.¹⁸ LPMOs are copper-dependent enzymes that cleave crystalline substrates by oxidizing the glycosidic bonds. By introducing chain breaks in insoluble polysaccharides, such as cellulose and chitin, and also in some hemicelluloses, LPMOs have the ability to enhance the activity of the glycoside hydrolases.

CAZymes and CBMs are classified into sequence-based families in the CAZy database (www.cazy.org).¹⁹ This database, with regular updates, is dedicated to classifying and analyse genomic, structural and biochemical information concerning CAZymes and associated CBMs involved in the synthesis, modification and breakdown of oligo- and polysaccharides.¹⁹ Currently (as of June 2017), there are numerous different CAZymes and CBM families identified: 145 families of glycoside hydrolases, 103 families of glycosyl transferases, 26 families of polysaccharide lyases, 16 families of carbohydrate esterases, 13 families of auxiliary activities (including the 4 LPMO families), and 81 families of CBMs.

2.1 Carbohydrate-binding modules: the non-catalytic domains associated to Carbohydrate-Active enZymes

CBMs are defined as non-catalytic protein domains, with sequences ranging from 30 to 200 amino acids,^{20,21} which are classified and divided into families based on sequence similarity.¹⁹ These modules were initially defined as cellulose-binding domains, as the first examples of CBMs mainly bound to crystalline cellulose.³¹ However, these modules show a highly diverse range of ligand specificities, between different families and even within the same family. Several characterized CBMs recognize non-crystalline cellulose, chitin, xylan, mannan, galactan, soluble α - and β -glucans and insoluble storage polysaccharides, such as starch and glycogen.¹⁹ CBMs are most commonly found associated to polysaccharide-degrading enzymes, where the structural proximity promotes the CBM function in enhancing the enzyme activity. Four functional major roles are recognized for CBMs: proximity effect, targeting function, disruptive function and cell attachment.^{20,21} However, these modules can also be found in isolated or tandem forms, and may exert other functions, such as the case of polysaccharide sensing domains¹¹ and LysM signalling domains.³²

Given its wide variety of ligand specificities, CBMs are an excellent model to study protein-carbohydrate recognition mechanisms. Additionally, some properties make these modules interesting candidates for various applications in biotechnology, such as enhancers in biomass degradation,³³ as affinity support in purification techniques and as biosensors.³⁴



2.2 The functional CBM types

Based on the conservation of protein fold, CBMs are divided into 7 different families. Most CBMs identified to date are classified in the β -sandwich family of protein folds. To provide additional functional relevance to the CBM classification, these modules have been grouped into three types: A, B, and C, according to the mode of interaction with the carbohydrates and the architecture of the binding site.^{20,21}

CBMs from type A have a planar hydrophobic surface decorated by aromatic residues that interact with flat crystalline polysaccharides, such as chitin or cellulose. This type of interaction is observed in the crystal structure of a CBM63-containing *Bacillus subtilis* expansin in complex with (1 $\rightarrow$ 4)- β -D-linked cellohexaose.³⁵ Some type A CBMs have been reported to bind not only to crystalline cellulose but also to soluble polysaccharides, such as CBMs from family 2 and 3, which can also bind xyloglucans,³⁶ and CBM64 from *Spirochaeta thermophila* that binds a variety of hemicelluloses.³⁷

CBMs from type B or endo-type are classified as CBMs that bind to internal oligosaccharide sequences. These CBMs exhibit a cleft or groove, that accommodates oligosaccharide chains with four or more residues, and show higher binding affinities with the increase of the oligosaccharide chain length. One example is the highly thermostable family 4 CBM from *Thermotoga maritima* that binds (1 $\rightarrow$ 3)- β -D-glucans and (1 $\rightarrow$ 3)- β -D-(1 $\rightarrow$ 4)- β -D-glucans^{28,38} (Fig. 3B and Fig. 5). This CBM comprises the C-terminus of the putative laminarinase Lam16A and the recognition of (1 $\rightarrow$ 3)- β -D-linked laminarioligosaccharides involves three tryptophan residues (W28, W58, and W99) and one tyrosine residue (Y23). One other example is the family 11 CBM from *C. thermocellum*, which recognizes mixed-linked (1 $\rightarrow$ 3)- β -D-(1 $\rightarrow$ 4)- β -D-glucans.^{25,38} This CBM is associated to the enzyme Lic26A-Cel5E, an enzyme that contains GH5 and GH26 catalytic domains that display (1 $\rightarrow$ 4)- β -D-glucan and (1 $\rightarrow$ 3)- β -D-(1 $\rightarrow$ 4)- β -D-glucan endoglucanase activity, respectively.²⁵

CBMs from type C, or exo-type, are classified as CBMs that recognize the non-reducing end of an oligosaccharide sequence, binding in an optimal way to mono-, di- or trisaccharides, due to steric restriction in the binding

Fig. 3 The anaerobic bacterial cellulosome and the carbohydrate-recognition by CBMs. (A) Schematic representation of a cellulosomal assembly from *C. thermocellum*, depicting known 3D structures of selected modules determined by X-ray Crystallography. Depicted 3D structures are: Type I Cohesin-Dockerin (Coh-Doc I) complex (PDB ID: 1OHZ²²), type II Cohesin-Dockerin (Coh-Doc II) complex (PDB ID: 2B59²³), family 3a CBM (CBM3a; PDB ID: 4B9F¹¹), enzyme Cel5E endoglucanase (Cel5E; PDB ID: 4U3A²⁴), family 11 CBM (CBM11; PDB ID: 1V0A²⁵), enzyme Cel44A (Cel44A; PDB ID: 2EEX²⁶) and family 42 CBM (CBM42; PDB ID: 3KMV²⁷), all from *C. thermocellum*. (B) Type B protein-carbohydrate interactions illustrated by 2 views of the 3D structure of family 4 CBM from *Thermotoga maritima* in complex with (1 $\rightarrow$ 3)- β -D-linked laminarihexaose (PDB ID: 1GUI²⁸). This type of CBMs displays a cleft arrangement in which the binding site accommodates glycan chains with four or more monosaccharide units. (C) Family 6 CBM from *Cellvibrio mixtus* exhibits a type B cleft capable of recognizing (1 $\rightarrow$ 3)- β -D-(1 $\rightarrow$ 4)- β -D-glucans (PDB ID: 1UZ0) and a type C cleft that interacts with xylooligosaccharides (PDB ID: 1UYX).²⁹ Representations (not to scale) of individual 3D structures were done with program Chimera³⁰ using the PDB atomic coordinates.

site. Unlike the type B CBMs, type C do not contain the extended grooves in the binding-sites. Examples of this CBM type are the family 6 CBM from *Bacillus halodurans* in complex with laminarihexaose,³⁹ the family 42 CBM from *C. thermocellum*, a β -trefoil lectin that binds the arabinose side-chains of complex hemicelluloses,²⁷ and family 13 CBM from *Streptomyces lividans*, another β -trefoil binding xylose or xylo-oligosaccharides.⁴⁰

Remarkably, the family 6 CBM from *Cellvibrio mixtus* exhibits a type B cleft capable of recognizing (1 $\rightarrow$ 4)- β -D-linked, (1 $\rightarrow$ 3)- β -D-linked, and (1 $\rightarrow$ 3)- β -D-(1 $\rightarrow$ 4)- β -D-linked glucose oligosaccharides and a type C cleft that interacts with terminal residues of (1 $\rightarrow$ 4)- β -D-linked and (1 $\rightarrow$ 3)- β -D-linked glucose oligosaccharides^{29,41} (Fig. 3C).

The number of newly identified CBM sequences with putative carbohydrate binding is growing fast due to the exponential increase of sequence information derived from microbial genomics, metagenomics and transcriptomics data. Many of these proteins await elucidation and assignment of a carbohydrate-binding function. The development of carbohydrate microarrays in the recent decades,^{42–48} came to revolutionize the study of carbohydrate-protein interactions, satisfying the high demand for high-throughput methods to systematically array carbohydrate libraries and identify the specificity and biological role of carbohydrate-binding proteins. The important aspects of this technology will be highlighted in the sections below, with insights into approaches to study CBM-plant carbohydrate interactions.

3 Carbohydrate microarrays

The main advantage of the microarray technology is that a wide diversity of carbohydrate probes can be immobilized on a microarray surface and simultaneously assessed for binding events, using only minute amounts of samples. This miniaturization feature of the microarrays takes the most out of precious materials, both carbohydrates and protein analytes, while generating a large amount of information on a variety of carbohydrate-recognition systems.^{49,50} In addition, the multivalent display of arrayed carbohydrates enables the microarray to mimic to some extent the display at the cell surfaces, which is ideal for detecting the usually very low affinities of carbohydrate-protein interactions.

Carbohydrate microarray methods are generally of two categories: polysaccharide or glycoprotein microarrays and oligosaccharide microarrays.⁴⁹ The carbohydrate samples can either be isolated from natural sources or be chemically or chemo-enzymatically synthesized. On the one hand, polysaccharide or glycoprotein microarrays can comprise the full diversity of a particular glycome and may avoid the loss of any labile or conformational determinants during the release of the oligosaccharides from their molecules of origin.^{43,44,46,49} On the other hand, oligosaccharide microarrays are powerful tools to assess binding-specificity in carbohydrate-recognition events and to identify the binding epitopes.^{46,49} Polysaccharides and glycoproteins can be readily and randomly immobilized on solid matrices based on hydrophobic physical adsorption or charge-based interaction.^{43,44,49} The immobilization of oligosaccharides

is more challenging, given their low mass and hydrophilic nature, and chemical derivatization at the reducing monosaccharide is usually required prior to immobilization to introduce suitable functional groups.^{49–51} Being in equilibrium between the hemiacetal closed-ring and the aldehyde open-chain form, the reducing monosaccharide can serve as an electrophilic group for a chemoselective reaction with numerous nucleophilic amine-, hydrazide-, or oxyamine-containing reagents.^{49–51} For oligosaccharides obtained through chemical synthesis, the functionality is generally carried out by placing a linker at the reducing terminal monosaccharide residue in a form suitable for flexible modifications.^{49–51} These different strategies allow oligosaccharides to be immobilized on a compatible microarray surface. Polymer-based surfaces such as nitrocellulose or plastic are usually an attractive solid surface for non-covalent immobilization,^{42–44,47} whereas gold or functionalised glass are used to covalently attach carbohydrate probes.^{45,52}

3.1 Carbohydrate microarray platforms

To date, several carbohydrate microarray platforms have been developed that: (1) use alternative chemical strategies to overcome the limitation of direct immobilization of oligosaccharides onto solid matrices; (2) differ on the type of carbohydrates and how they are displayed on the array surface; and (3) are based on covalent or non-covalent immobilization to different surfaces. These are reviewed in detail in recent references^{48,50,51,53,54} and some are highlighted in the special issue of *Current Opinion in Chemical Biology* dedicated to arrays.⁵⁵ Here, we do not intend to extensively review these, but we will highlight high-throughput platforms that contain a high diversity of oligosaccharide probes and that use different strategies for their immobilization and presentation.

Feizi and colleagues have developed a microarray system based on the neoglycolipid (NGL) technology,⁵⁶ in which the oligosaccharides are linked to a lipid.^{42,46,47,53,57} The generated NGL probes have amphipathic properties, which enables efficient display onto nitrocellulose-coated glass slides using a liposome formulation in the presence of carrier lipids⁵⁸ (Fig. 4A). Natural or chemically synthesized reducing oligosaccharides are conjugated through microscale reductive amination to the aminolipid 1,2-dihexadecyl-*sn*-glycero-3-phosphoethanolamine (DHPE, DH-NGLs) (Fig. 4C). This procedure yields the ring-opening of the monosaccharide at the reducing end.⁴⁹ To overcome this limitation, NGLs with ring-closed monosaccharide cores have been introduced by Liu and colleagues.⁵⁹ These are prepared by conjugating reducing oligosaccharides to an aminoxy-functionalized DHPE by microscale oxime ligation (without reduction) (AOPE, AO-NGLs) (Fig. 4C). This procedure enables the efficient presentation of short oligosaccharides for direct binding assays.⁵⁹ The non-covalent immobilization of NGLs in a lipid environment onto a nitrocellulose surface introduces an element of mobility. This mode of presentation simulates to some extent the cell surface display of glycans and may be advantageous for detection of binding for particular recognition systems.⁵³ The NGL-based microarray

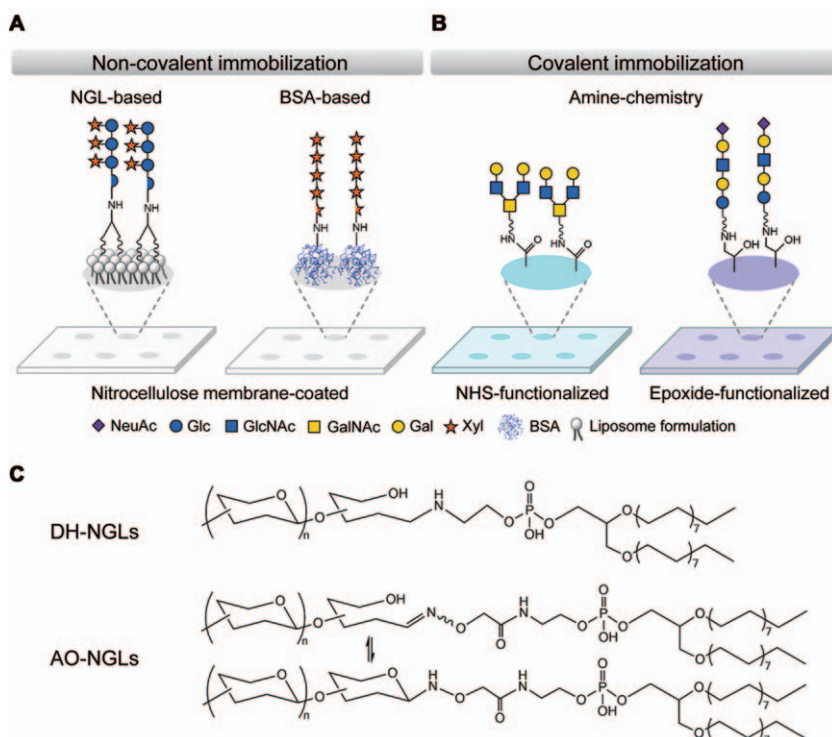


Fig. 4 Graphic representation of examples of immobilization strategies used to generate carbohydrate microarrays. (A) Non-covalent microarrays: immobilization onto nitrocellulose-coated glass slides of reducing oligosaccharides derivatized by reductive amination to an aminolipid, to prepare neoglycolipids (NGLs),⁵³ or to BSA, to prepare neoglycoproteins.⁶¹ (B) Covalent microarrays: immobilization of synthetic oligosaccharides derivatized at the reducing end to an amino-terminating linker onto *N*-hydroxysuccinimide (NHS)-functionalized glass slides⁴⁵ or onto epoxide-functionalized glass slides.⁶² (C) NGL probes prepared from reducing oligosaccharides by reductive amination (DHPE, DH-NGLs)⁶³ and by oxime ligation (AOPE, AO-NGLs);⁵⁹ the derivatization of the oligosaccharide by oxime ligation produces an equilibrium between the open- and closed-ring form of the reducing monosaccharide.⁵⁹ Examples of carbohydrate structures in the different libraries are shown using the symbol nomenclature according to Varki *et al.*, 2015.⁶⁴

system currently contains a repertoire of around 800 sequence-defined probes, with a high content of natural oligosaccharide sequences, including NGLs derived from various oligosaccharides of mammalian sources, from polysaccharides of bacterial, fungal, and plant origins, and natural and synthetic glycolipids.⁵³

The microarray platform of the Consortium for Functional Glycomics (CFG) developed by the early work of Blixt and colleagues^{45,60} is also based upon amine chemistry, whereby oligosaccharides linked at the reducing end with an amine-terminating linker are covalently immobilized onto *N*-hydroxysuccinimide (NHS) ester-derivatized glass slides (Fig. 4B). Recent microarray versions are composed of around 600 mammalian-type probes (mammalian printed array version 5.3). Other strategies that also use an amino-linker involve immobilization of the amine-terminated oligosaccharides onto epoxide-derivatized slides (Fig. 4B). Examples are by Cummings and colleagues that used this

method for immobilization of naturally-derived oligosaccharide libraries⁶⁵ and by Varki and colleagues who developed a structurally diverse microarray of sialylated oligosaccharides.⁶² Gildersleeve and colleagues demonstrated that oligosaccharides conjugated to BSA or HSA (displayed as neoglycoproteins) may also be efficiently immobilized using amine chemistry onto epoxide functionalized glass slides for binding studies.⁶⁶ Other groups, have developed covalent microarrays based on different chemistries, such as the early work by Shin and colleagues using the thiol chemistry, whereby maleimide-functionalized oligosaccharides are immobilized onto thiol-derivatized slides.⁶⁷ In these covalent oligosaccharide microarray platforms, the nature and length of the linkers between the oligosaccharide and the array surface are important for accessibility of the oligosaccharide to the protein and detection of specific binding.

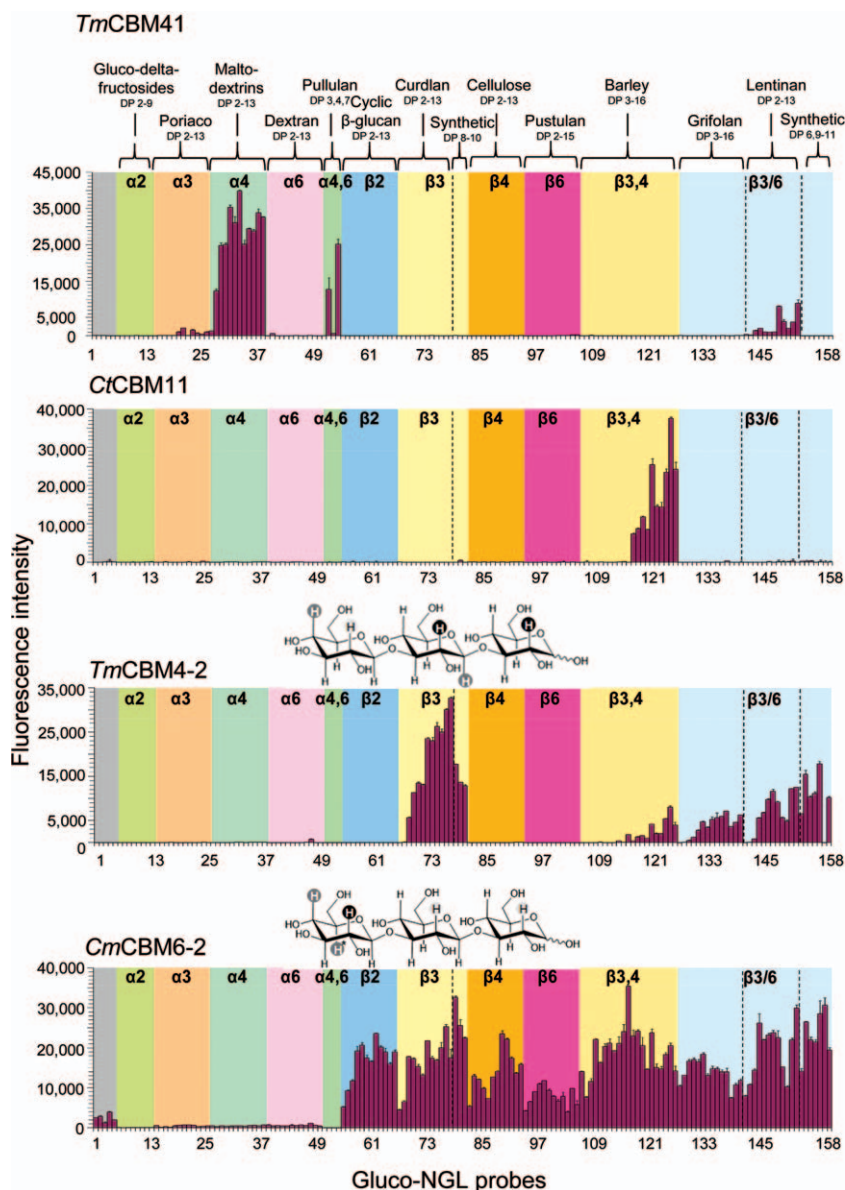
The NGL-based microarray facility and that of the Consortium of Functional Glycomics (accessed through the links <https://www.imperial.ac.uk/glycosciences/> and <http://www.functionalglycomics.org/>, respectively) are the two largest platforms assembled to date that are open to the broad scientific community for microarray screening analyses of carbohydrate-binding proteins in different biological contexts.

Although the number of sequence-defined probes in carbohydrate microarrays has been expanding, the increase in diversity to date has been mainly on mammalian-type sequences. Some groups, however, have focused on development of microarrays from plant-derived carbohydrates.^{38,61,68–73} Sequence-defined oligosaccharides can be derived from natural polysaccharides and the development of methods for fine-tuned depolymerisation, purification, high-sensitive sequencing and structural characterisation of the oligosaccharide fragments is crucial.^{38,61} Methods for chemical^{74–76} or chemo-enzymatic synthesis⁷⁷ of structural elements from complex plant cell-wall polysaccharides, or for genetically engineering bacterial strains with specific CAZymes gene deletions to produce oligosaccharides in the presence of a target substrate,⁷⁸ offer alternative and complementary approaches to achieve the much needed structural diversity. In the sections below, selected examples of microarray approaches to study plant-carbohydrate recognition by CBMs will be highlighted.

3.2 Microarrays focused on plant carbohydrates for recognition studies

Early work by Willats and colleagues, reported on a carbohydrate-based approach for high-throughput plant polysaccharide cell-wall profiling.⁷⁰ This Comprehensive Microarray Polymer Profiling (CoMPP) is based on the extraction of *Arabidopsis thaliana* and *Physcomitrella patens* polysaccharides and printing of the polysaccharide-rich fractions onto nitrocellulose-based arrays. These are then probed with CBMs and monoclonal antibodies of known specificities for plant cell-wall polysaccharides. The CoMPP strategy enables the plant cell-wall composition to be assessed in a semi-quantitative high-throughput way by revealing the relative abundance of polysaccharide epitopes.⁷⁰ More recently, using

the same principle of arraying chemically extracted polysaccharides and the information of monoclonal antibodies, Waldron and colleagues have analysed quantitatively the abundance of different non-cellulosic polysaccharides in 331 genetically different *Brassica napus* cultivars.⁷¹ These studies are providing insights to plant cell-wall biosynthesis and restructuring.⁷¹ This high-throughput screening of polysaccharide structures requires the use of proteins for which carbohydrate-specificity is known. Thus, development of sequence-defined plant-based carbohydrate microarrays is highly important to provide these protein tools.



Willats and colleagues have developed a suitable platform for high-throughput analysis of the specificities of CBMs and monoclonal anti-carbohydrate antibodies.⁶¹ This microarray is composed of linear and branched oligosaccharides, either isolated from polysaccharides, such as glucans, xylans, mannans, galactans, xyloglucans or arabinans, using enzymatic or chemical hydrolysis or generated by chemical synthesis.⁶¹ The oligosaccharides are coupled to BSA by reductive amination, producing a ring-opened monosaccharide at the reducing end of the oligosaccharide (Fig. 4A).⁶¹ These neoglycoproteins are arrayed non-covalently together with plant polysaccharides onto a solid matrix, such as nitrocellulose-coated glass slides. The developed microarrays were recently used to characterize unknown CBMs of *Ruminococcus flavefaciens* cellulosome, revealing six previously unidentified CBM families targeting β -glucans, β -mannans and pectic homogalacturonan.⁷⁹ This study was important to gain knowledge on the complexity of the *R. flavefaciens* cellulosome and its extended repertoire of CBMs for efficient plant cell-wall degradation in absence of CBMs that target cellulose. More recently, the sequence-diversity in these microarrays was expanded to contain linear, branched and phosphorylated (1 $\rightarrow$ 4)- α -D-linked glucose maltooligosaccharides.⁸⁰ These were applied to characterize the starch binding domain CBM20 of *Aspergillus niger* as tool for high-throughput screening of starch structures during development and germination.⁸⁰

3.3 Targeting glucan recognition using glycome-based microarrays

A key feature of the NGL technology is its interface with mass spectrometry and high-performance TLC (HPTLC) or HPLC.^{49,53} This enables bioactive oligosaccharides released from a glycome source to be resolved from heterogeneous mixtures, characterized and purified, allowing the discovery and characterization of novel ligands of biological relevance.^{42,46,53,58} Based on their previous work, which used this ‘designer’ approach from ligand-bearing glucans to assign the oligosaccharide ligands for the immune receptor Dectin-1⁴⁷ and anti-fungal therapeutic antibodies,⁴⁹ Palma and colleagues have developed a sequence-defined ‘glucome’ microarray as a screening tool for glucan-binding proteins (Fig. 5).³⁸

Fig. 5 Neoglycolipid (NGL)-based designer microarray comprised of sequence-defined gluco-oligosaccharide NGL probes (glucome microarray) as a tool for glucan recognition studies. Differing specificities and chain length requirements obtained in the microarray analysis of glucan-binding CBMs from different CAZy families: family 41 and family 4 CBMs from the marine hyperthermophile *T. maritima* (TmCBM41 and TmCBM4-2, respectively); family 11 CBM from *C. thermocellum* (CtCBM11); and family 6 from the aerobic soil bacterium *C. mixtus* (CmCBM6-2); the inset shows the binding epitopes as determined by STD NMR of (1 $\rightarrow$ 3)- β -D-linked glucose trisaccharide in the presence of TmCBM4-2 and CmCBM6-2 (dark, medium and light gray circles indicate strong, medium, and weak STD effects, respectively). Depiction of the sequence diversity in the microarray is on the top of the panel. DP: degree of polymerization. This research was originally published in Molecular and Cellular Proteomics. A. S. Palma, Y. Liu, H. Zhang, Y. Zhang, B. V. McCleary, G. Yu, Q. Huang, L. S. Guidolin, A. E. Ciocchini, A. Torosantucci, D. Wang, A. L. Carvalho, C. M. G. A. Fontes, B. Mulloy, R. a Childs, T. Feizi and W. Chai. Unravelling glucan recognition systems by glycome microarrays using the designer approach and mass spectrometry. Molecular and Cellular Proteomics. 2015; 14:974–988. © the American Society for Biochemistry and Molecular Biology.

The glucome microarray comprises 153 gluco-oligosaccharide probes, with diverse sequences and chain lengths representing major sequences in glucans (Fig. 5), including those present in plant cell-wall (Fig. 2). The oligosaccharides are prepared by depolymerization of glucans and multiple chromatographic methods at microscale, or are synthesized chemically. The linkage and sequence determination of the linear and branched oligosaccharides are determined at high-sensitivity by development of a negative-ion ESI-CID-MS/MS method for gluco-oligosaccharides.³⁸ The oligosaccharides are converted to NGL probes by oxime ligation to an aminoxy-functionalized lipid (AO-NGLs)⁵⁹ (Fig. 4C). The optimization of this method enabled long chains of gluco-oligosaccharides, otherwise difficult to derivatize, to be displayed in the microarrays for interaction studies.^{59,81}

Combining the microarray analysis with MS sequencing enabled the high sensitivity detection and unambiguous assignment of specificity of glucan recognition. Examples of the binding patterns obtained for bacterial CBMs are in Fig. 5, highlighting different specificities for glucan sequences and chain length requirements. Importantly, the chain length requirements observed in the microarray analysis of these CBMs were correlated with the different modes of (1 → 3)-β-D-glucan interaction in solution using STD-NMR (insets in Fig. 5) and with their assigned functional types (Fig. 3B and C) as follows: type B *TmCBM4-2*,²⁸ for which the binding site is a cleft that accommodates 4 residues, and interacts most strongly with internal residues of (1 → 3)-β-D-linked gluco-oligosaccharides (Fig. 3B and inset of Fig. 5), requires a degree of polymerization (DP) of 4 for detectable binding; *CmCBM6-2*,^{29,41} for which one of two binding sites has a type C arrangement (Fig. 3C) and interacts most strongly with the non-reducing end of the (1 → 3)-β-D-linked glucose trisaccharide (inset of Fig. 5), binds to shorter sequences in the microarray (DP-2 or DP-3).

The purity of glucan polysaccharides of different structural types is of particular importance for assignment of specificity. The partial fragmentation of the polysaccharides, the sequencing of the oligosaccharides by the negative-ion ESI-CID-MS/MS method and their interrogation on the microarrays, not only provide detailed information on linkage, sequence and chain length requirements of glucan-recognizing proteins, but are also a sensitive means of revealing unsuspected sequences in the polysaccharides.³⁸

4 Conclusions

Carbohydrate microarrays have become essential in carbohydrate-recognition research in the post-genomic era. The technology is ideal for screening proteomes for carbohydrate-binding activities and assigning natural ligands for proteins, as well as their biological roles. The selected examples highlight the importance of different approaches in developing microarray platforms to present and study the structural diversity of plant polysaccharides and their recognition by microbial proteins. The continuous expansion of sequence-defined carbohydrate probe libraries in numbers and diversity, assembled from multiple sources and by

different methods, is critical to increase the scope of these microarray approaches. These establish the microarray technology as an effective tool to target carbohydrate-binding function in plant polysaccharide biodegradation, and to functionally classify newly identified CBMs or CBMs assigned to known families in the CAZy database. The vast information that is generated can then be exploited in biotechnology, nutrition and plant biology.

Acknowledgements

For grant support, the authors acknowledge the Portuguese Science and Technology Foundation (MCTES) through grants PTDC/QUI-QUI/112537/2009, PTDC/BBB-BEP/0869/2014; SFRH/BD/100569/2014; Unidade de Ciências Biomoleculares Aplicadas-UCIBIO UID/Multi/04378/2013; POCI-01-0145-FEDER-007728.

References

- 1 N. C. Carpita and D. M. Gibeaut, *Plant J.*, 1993, **3**, 1–30.
- 2 D. J. Cosgrove, *Nat. Rev. Mol. Cell Biol.*, 2005, **6**, 850–861.
- 3 J. P. Knox, *Curr. Opin. Plant Biol.*, 2008, **11**, 308–313.
- 4 R. A. Burton, M. J. Gidley and G. B. Fincher, *Nat. Chem. Biol.*, 2010, **6**, 724–732.
- 5 S. Pérez and D. Samain, in *Advances in Carbohydrate Chemistry and Biochemistry*, Elsevier Inc., 2010, vol. 64, pp. 25–116.
- 6 H. V. Scheller and P. Ulvskov, *Annu. Rev. Plant Biol.*, 2010, **61**, 263–289.
- 7 B. A. Stone, in *Chemistry, Biochemistry, and Biology of 1-3 Beta Glucans and Related Polysaccharides*, Elsevier Inc, 2009, pp. 5–46.
- 8 D. Mohnen, *Curr. Opin. Plant Biol.*, 2008, **11**, 266–277.
- 9 L. Artzi, E. A. Bayer and S. Morais, *Nat. Rev. Microbiol.*, 2016, **15**, 83–95.
- 10 C. M. G. A. Fontes and H. J. Gilbert, *Annu. Rev. Biochem.*, 2010, **79**, 655–681.
- 11 O. Yaniv, E. Morag, I. Borovok, E. A. Bayer, R. Lamed, F. Frolov and L. J. W. Shimon, *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.*, 2013, **69**, 733–737.
- 12 P. L. Bergquist, M. D. Gibbs, D. D. Morris, V. S. J. Te'o, D. J. Saul and H. W. Morgan, *FEMS Microbiol. Ecol.*, 1999, **28**, 99–110.
- 13 J. L. A. Brás, B. A. Pinheiro, K. Cameron, F. Cuskin, A. Viegas, S. Najmudin, P. Bule, V. M. R. Pires, M. J. Romão, E. A. Bayer, H. L. Spencer, S. Smith, H. J. Gilbert, V. D. Alves, A. L. Carvalho and C. M. G. A. Fontes, *Sci. Rep.*, 2016, **6**, 38292.
- 14 H. J. Flint, E. A. Bayer, M. T. Rincon, R. Lamed and B. A. White, *Nat. Rev. Microbiol.*, 2008, **6**, 121–131.
- 15 G. R. Hemsworth, G. Dejean, G. J. Davies and H. Brumer, *Biochem. Soc. Trans.*, 2016, **44**, 94–108.
- 16 L. Lynd, *Microbiol. Mol. Biol. Rev.*, 2002, **66**, 506–577.
- 17 R. P. de Vries and J. Visser, *Microbiol. Mol. Biol. Rev.*, 2001, **65**, 497–522.
- 18 G. R. Hemsworth, G. Déjean, G. J. Davies and H. Brumer, *Biochem. Soc. Trans.*, 2016, **44**, 94–108.
- 19 V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho and B. Henrissat, *Nucleic Acids Res.*, 2014, **42**, 490–495.
- 20 A. B. Boraston, D. N. Bolam, H. J. Gilbert and G. J. Davies, *Biochem. J.*, 2004, **382**, 769–781.

- 21 H. J. Gilbert, J. P. Knox and A. B. Boraston, *Curr. Opin. Struct. Biol.*, 2013, **23**, 669–677.
- 22 A. L. Carvalho, F. M. V. Dias, J. a M. Prates, T. Nagy, H. J. Gilbert, G. J. Davies, L. M. A. Ferreira, M. J. Romao and C. M. G. A. Fontes, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 13809–13814.
- 23 J. J. Adams, G. Pal, Z. Jia and S. P. Smith, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 305–310.
- 24 S.-F. Yuan, T.-H. Wu, H.-L. Lee, H.-Y. Hsieh, W.-L. Lin, B. Yang, C.-K. Chang, Q. Li, J. Gao, C.-H. Huang, M.-C. Ho, R.-T. Guo and P.-H. Liang, *J. Biol. Chem.*, 2015, **290**, 5739–5748.
- 25 A. L. Carvalho, A. Goyal, J. A. M. Prates, D. N. Bolam, H. J. Gilbert, V. M. R. Pires, L. M. A. Ferreira, A. Planas, M. J. Romão and C. M. G. A. Fontes, *J. Biol. Chem.*, 2004, **279**, 34785–34793.
- 26 Y. Kitago, S. Karita, N. Watanabe, M. Kamiya, T. Aizawa, K. Sakka and I. Tanaka, *J. Biol. Chem.*, 2007, **282**, 35703–35711.
- 27 T. Ribeiro, T. Santos-Silva, V. D. Alves, F. M. V. Dias, A. S. Luís, J. A. M. Prates, L. M. A. Ferreira, M. J. Romão and C. M. G. A. Fontes, *Biochim. Biophys. Acta, Proteins Proteomics*, 2010, **1804**, 2054–2062.
- 28 A. B. Boraston, D. Nurizzo, V. Notenboom, V. Ducros, D. R. Rose, D. G. Kilburn and G. J. Davies, *J. Mol. Biol.*, 2002, **319**, 1143–1156.
- 29 V. M. R. Pires, J. L. Henshaw, J. A. M. Prates, D. N. Bolam, L. M. A. Ferreira, C. M. G. A. Fontes, B. Henrissat, A. Planas, H. J. Gilbert and M. Czjzek, *J. Biol. Chem.*, 2004, **279**, 21560–21568.
- 30 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605–1612.
- 31 N. R. Gilkes, R. A. Warren, R. C. Miller and D. G. Kilburn, *J. Biol. Chem.*, 1988, **263**, 10401–10407.
- 32 G. Buist, A. Steen, J. Kok and O. P. Kuipers, 2008, **68**, 838–847.
- 33 V. Reyes-Ortiz, R. A. Heins, G. Cheng, E. Y. Kim, B. C. Vernon, R. B. Elandt, P. D. Adams, K. L. Sale, M. Z. Hadi, B. A. Simmons, M. S. Kent and D. Tullman-Ercek, *Biotechnol. Biofuels*, 2013, **6**, 93.
- 34 O. Shoseyov, Z. Shani and I. Levy, *Microbiol. Mol. Biol. Rev.*, 2006, **70**, 283–295.
- 35 N. Georgelis, N. H. Yennawar and D. J. Cosgrove, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 14830–14835.
- 36 M. C. Hernandez-Gomez, M. G. Rydahl, A. Rogowski, C. Morland, A. Cartmell, L. Crouch, A. Labourel, C. M. G. A. Fontes, W. G. T. Willats, H. J. Gilbert and J. P. Knox, *FEBS Lett.*, 2015, **589**, 2297–2303.
- 37 V. M. R. Pires, P. M. M. Pereira, J. L. A. Brás, M. Correia, V. Cardoso, P. Bule, V. D. Alves, S. Najmudin, I. Venditto, L. M. A. Ferreira, M. J. Romão, A. L. Carvalho, C. M. G. A. Fontes and D. M. Prazeres, *J. Biol. Chem.*, 2017, **292**, 4847–4860.
- 38 A. S. Palma, Y. Liu, H. Zhang, Y. Zhang, B. V. McCleary, G. Yu, Q. Huang, L. S. Guidolin, A. E. Ciocchini, A. Torosantucci, D. Wang, A. L. Carvalho, C. M. G. A. Fontes, B. Mulloy, R. A. Childs, T. Feizi and W. Chai, *Mol. Cell. Proteomics*, 2015, **14**, 974–988.
- 39 A. L. Van Bueren, C. Morland, H. J. Gilbert and A. B. Boraston, *J. Biol. Chem.*, 2005, **280**, 530–537.
- 40 V. Notenboom, A. B. Boraston, S. J. Williams, D. G. Kilburn and D. R. Rose, *Biochemistry*, 2002, **41**, 4246–4254.
- 41 J. L. Henshaw, D. N. Bolam, V. M. R. Pires, M. Czjzek, B. Henrissat, L. M. A. Ferreira, C. M. G. A. Fontes and H. J. Gilbert, *J. Biol. Chem.*, 2004, **279**, 21552–21559.

- 42 S. Fukui, T. Feizi, C. Galustian, A. M. Lawson and W. Chai, *Nat. Biotechnol.*, 2002, **20**, 1011–1017.
- 43 D. Wang, S. Liu, B. J. Trummer, C. Deng and A. Wang, *Nat. Biotechnol.*, 2002, **20**, 275–281.
- 44 W. G. T. Willats, S. E. Rasmussen, T. Kristensen, J. D. Mikkelsen and J. P. Knox, *Proteomics*, 2002, **2**, 1666–1671.
- 45 O. Blixt, S. Head, T. Mondala, C. Scanlan, M. E. Huflejt, R. Alvarez, M. C. Bryan, F. Fazio, D. Calarese, J. Stevens, N. Razi, D. J. Stevens, J. J. Skehel, I. van Die, D. R. Burton, I. A. Wilson, R. Cummings, N. Bovin, C.-H. Wong and J. C. Paulson, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 17033–17038.
- 46 T. Feizi and W. Chai, *Nat. Rev. Mol. Cell Biol.*, 2004, **5**, 582–588.
- 47 A. S. Palma, T. Feizi, Y. Zhang, M. S. Stoll, A. M. Lawson, E. Díaz-Rodríguez, M. A. Campanero-Rhodes, J. Costa, S. Gordon, G. D. Brown and W. Chai, *J. Biol. Chem.*, 2006, **281**, 5771–5779.
- 48 J. Y. Hyun, J. Pai and I. Shin, *Acc. Chem. Res.*, 2017, **50**, 1069–1078.
- 49 Y. Liu, A. S. Palma and T. Feizi, *Biol. Chem.*, 2009, **390**, 647–656.
- 50 C. D. Rillahan and J. C. Paulson, *Annu. Rev. Biochem.*, 2011, **80**, 797–823.
- 51 S. Park, J. C. Gildersleeve, O. Blixt and I. Shin, *Chem. Soc. Rev.*, 2013, **42**, 4310–4326.
- 52 B. Donczko, J. Kerekyarto, Z. Szurmai and A. Guttman, *Analyst*, 2014, **139**, 2650.
- 53 A. S. Palma, T. Feizi, R. A. Childs, W. Chai and Y. Liu, *Curr. Opin. Chem. Biol.*, 2014, **18**, 87–94.
- 54 X. Song, J. Heimbürg-Molinario, R. D. Cummings and D. F. Smith, *Curr. Opin. Chem. Biol.*, 2014, **0**, 70–77.
- 55 S. Matson Robert and D. F. Smith, Eds., *Arrays*, Current Opinion in Chemical Biology, 2014, vol. 18.
- 56 P. W. Tang, H. C. Gool, M. Hardy, Y. C. Lee and T. Felzi, *Biochem. Biophys. Res. Commun.*, 1985, **132**, 474–480.
- 57 T. Feizi, *Ann. N. Y. Acad. Sci.*, 2013, **1292**, 33–44.
- 58 Y. Liu, R. A. Childs, A. S. Palma, M. A. Campanero-Rhodes, M. S. Stoll, W. Chai and T. Feizi, in *Carbohydrate Microarrays: Methods and Protocols*, *Methods in Molecular Biology*, ed. Y. Chevolut, Springer Science+Business Media, 2012, vol. 808, pp. 117–136.
- 59 Y. Liu, T. Feizi, M. A. Campanero-Rhodes, R. A. Childs, Y. Zhang, B. Mulloy, P. G. Evans, H. M. I. Osborn, D. Otto, P. R. Crocker and W. Chai, *Chem. Biol.*, 2007, **14**, 847–859.
- 60 D. F. Smith, X. Song and R. D. Cummings, in *Methods in Enzymology: Functional Glycomics*, Elsevier Inc., 1st edn., 2010, vol. 480, pp. 417–444.
- 61 H. L. Pedersen, J. U. Fangel, B. McCleary, C. Ruzanski, M. G. Rydahl, M. C. Ralet, V. Farkas, L. Von Schantz, S. E. Marcus, M. C. F. Andersen, R. Field, M. Ohlin, J. P. Knox, M. H. Clausen and W. G. T. Willats, *J. Biol. Chem.*, 2012, **287**, 39429–39438.
- 62 V. Padler-Karavani, X. Song, H. Yu, N. Hurtado-Ziola, S. Huang, S. Muthana, H. A. Chokhawala, J. Cheng, A. Verhagen, M. A. Langereis, R. Kleene, M. Schachner, R. J. De Groot, Y. Lasanajak, H. Matsuda, R. Schwab, X. Chen, D. F. Smith, R. D. Cummings and A. Varki, *J. Biol. Chem.*, 2012, **287**, 22593–22608.
- 63 T. Feizi, F. Fazio, W. Chai and C.-H. Wong, *Curr. Opin. Struct. Biol.*, 2003, **13**, 637–645.
- 64 A. Varki, R. D. Cummings, M. Aebe, N. H. Packer, P. H. Seeberger, J. D. Esko, P. Stanley, G. Hart, A. Darvill, T. Kinoshita, J. J. Prestegard, R. L. Schnaar,

- H. H. Freeze, J. D. Marth, C. R. Bertozzi, M. E. Etzler, M. Frank, J. F. G. Vliegthart, T. Lütke, S. Perez, E. Bolton, P. Rudd, J. Paulson, M. Kanehisa, P. Toukach, K. F. Aoki-Kinoshita, A. Dell, H. Narimatsu, W. York, N. Taniguchi and S. Kornfeld, *Glycobiology*, 2015, **25**, 1323–1324.
- 65 X. Song, Y. Lasanajak, B. Xia, D. F. Smith and R. D. Cummings, *ACS Chem. Biol.*, 2009, **4**, 741–750.
- 66 J. C. Manimala, T. A. Roach, Z. Li and J. C. Gildersleeve, *Angew. Chemie - Int. Ed.*, 2006, **45**, 3607–3610.
- 67 S. Park, M. R. Lee, S. J. Pyo and I. Shin, *J. Am. Chem. Soc.*, 2004, **126**, 4812–4819.
- 68 M. Shipp, R. Nadella, H. Gao, V. Farkas, H. Sigrist and A. Faik, *Glycoconjugate J.*, 2008, **25**, 49–58.
- 69 O. Kosík, R. P. Auburn, S. Russell, E. Stratilová, S. Garajová, M. Hrmová and V. Farkaš, *Glycoconjugate J.*, 2010, **27**, 79–87.
- 70 I. Møller, I. Sørensen, A. J. Bernal, C. Blaukopf, K. Lee, J. Øbro, F. Pettolino, A. Roberts, J. D. Mikkelsen, J. P. Knox, A. Bacic and W. G. T. Willats, *Plant J.*, 2007, **50**, 1118–1128.
- 71 I. P. Wood, B. M. Pearson, E. Garcia-Gutierrez, L. Havlickova, Z. He, A. L. Harper, I. Bancroft and K. W. Waldron, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, 201619033.
- 72 J. M. van Munster, B. Thomas, M. Riese, A. L. Davis, C. J. Gray, D. B. Archer and S. L. Flitsch, *Sci. Rep.*, 2017, **7**, 43117.
- 73 S. Vidal-Melgosa, H. L. Pedersen, J. Schückel, G. Arnal, C. Dumon, D. B. Amby, R. N. Monrad, B. Westereng and W. G. T. Willats, *J. Biol. Chem.*, 2015, **290**, 9020–9036.
- 74 S. A. Nepogodiev, M. Fais, D. L. Hughes and R. A. Field, *Org. Biomol. Chem.*, 2011, **9**, 6670.
- 75 S. Kaothip and G.-J. Boons, *Org. Biomol. Chem.*, 2013, **11**, 5136–5146.
- 76 A. N. Zakharova, R. Madsen and M. H. Clausen, *Org. Lett.*, 2013, **15**, 1826–1829.
- 77 M. Lopez, S. Fort, H. Bizot, A. Buléon and H. Driguez, *Carbohydr. Polym.*, 2012, **88**, 185–193.
- 78 D. Ndeh, A. Rogowski, A. Cartmell, A. S. Luis, A. Baslé, J. Gray, I. Venditto, J. Briggs, X. Zhang, A. Labourel, N. Terrapon, F. Buffetto, S. Nepogodiev, Y. Xiao, R. A. Field, Y. Zhu, M. A. O'Neill, B. R. Urbanowicz, W. S. York, G. J. Davies, D. W. Abbott, M.-C. Ralet, E. C. Martens, B. Henrissat and H. J. Gilbert, *Nature*, 2017, **544**, 65–77.
- 79 I. Venditto, A. S. Luis, M. Rydahl, J. Schückel, V. O. Fernandes, S. Vidal-Melgosa, P. Bule, A. Goyal, V. M. R. Pires, C. G. Dourado, L. M. A. Ferreira, P. M. Coutinho, B. Henrissat, J. P. Knox, A. Baslé, S. Najmudin, H. J. Gilbert, W. G. T. Willats and C. M. G. A. Fontes, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 7136–7141.
- 80 V. Tanackovic, M. G. Rydahl, H. L. Pedersen, M. S. Motawia, S. S. Shaik, M. D. Mikkelsen, S. L. Kronic, J. U. Fangel, W. G. Willats and A. Blennow, *Sci. Rep.*, 2016, **6**, 30551.
- 81 H. Zhang, A. S. Palma, Y. Zhang, R. A. Childs, Y. Liu, D. A. Mitchell, L. S. Guidolin, W. Weigel, B. Mulloy, A. E. Ciocchini, T. Feizi and W. Chai, *Glycobiology*, 2016, **26**, 1086–1096.

Low melting carbohydrate mixtures and aqueous carbohydrates – an effective green medium for organic synthesis

Palanisamy Ravichandiran and Yanlong Gu*

DOI: 10.1039/9781788010641-00177

Low melting mixtures comprising of carbohydrates/urea or inorganic salts as new unconventional renewable solvents for organic transformations. These are stable, environmentally benign and easily biodegradable. Moreover, the properties include nontoxic, highly polar and contains low vapour pressure. In this review, we summarize the recent developments and their usefulness in organic synthesis, yet this is the extensive fast rising field and the most important examples in the past two decades have been taken into the account for the discussion.

1 Introduction

When the solvent is safe in an operation for both human beings and the environment and generated from the renewable feedstock obviously comes under the classification of green solvent.^{1,2} Ionic liquids (ILs) are considered to be a greener solvent due to its great physical and chemical properties, nevertheless, their impact towards the environment still questioned.³ Besides, ILs cannot be categorized as either green or toxic but their impact on the humans and environment are mainly concerned with the kind of anions and cations used to generate the ionic liquids and these factors are rigorously discussed in recent literature.^{4,5}

Organic transformation in water also an attractive and fast-growing field since last three decades, nevertheless, most of the organic compounds do not display sufficient solubility in the pure water and many of the starting materials decompose in the water at elevated temperature. During the work-up procedure for the product isolation at end of the reaction, it may also lead to the formation of the by-product in most of the cases, and consequently it is notable energy consuming process at the bulk level, as well. The effect of environmental problems causes by classical organic solvents in a bulk scale at the industry level, if they classified under volatile organic compounds (VOC) which covers chlorinated hydrocarbons derived from ethane, propane and methane.⁶ In order to address the downsides of the above stated conventional solvents, still, there is a way to find the benefit from the solvent effects, “green solvents” which are discovering their own path into the laboratories and chemical industries.

Instead, the utilization of renewable feedstock's into the valuable chemicals gains major importance in the 21st century. This utility of sustainable raw materials leads to the diminishing of fossil fuel reserves

Key Laboratory of Material Chemistry for Energy Conversion and Storage, Ministry of Education, Hubei Key Laboratory of Material Chemistry and Service Failure, School of Chemistry and Chemical Engineering, Huazhong University of Science and Technology, Wuhan 430074, China. E-mail: klgyl@hust.edu.cn

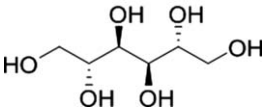
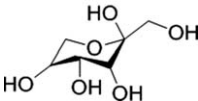
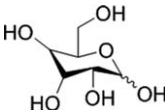
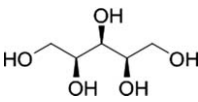
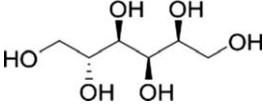
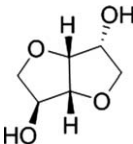
and it is an auspicious substitute for the sustainable supply of valuable building blocks and chemicals to the chemical industry.^{7,8} In the biomass, carbohydrate form occupies more than 75% weight and this can be utilized directly or with the modification of hydrolysis of poly- and oligosaccharides to monosaccharides like D-xylose and D-glucose.⁹ If the most oxidized sites in D-xylose and D-glucose substituted, the anomeric centre of the molecule pave the way to the contact of the significant classes of the glycosides and O-,¹⁰ S-,¹¹ C-,^{12,13} and N-glycosides¹⁴ are the best illustrations in this group of C-1 substituted monosaccharides.

Besides, the amalgamation of the carbohydrates with an aqueous media tailoring the properties of the solvent such as polarity, boiling point and the chirality, what the substrates need to be converted into the end product. Of late, low melting mixtures with the combination of the carbohydrates, urea, and inorganic salts may enlarge the field of organic-aqueous reactions.¹⁵ This stable melt considered as environmentally friendly due to the reason of easily biodegradable, relatively nontoxic and non-volatile compounds readily existing from natural resources. Meanwhile, the high polarity of these melt favours the reactions involving a polar transition state. The low melt organic ingredients consist of very less toxicity, for example, NaCl LD₅₀ orally in rats 3.8 g kg⁻¹ in addition, no toxicity was observed from lactose, sorbitol, N,N'-dimethylurea or urea. Also, an easy work-up without the use of any classical organic solvents turn into possible.¹⁶ With the simple production of these organic melt is an advantage to establish the biomass industries consists of limited industrial infrastructure.

Grafting of low organic melt generally considered by (2-hydroxyethyl)trimethylammonium chloride named choline chloride (ChCl) belonging to the vitamin B family, used as a green synthon in deep eutectic solvents in a number of organic transformations, though this quaternary ammonium salt choline plays a significant role of a multitude of metabolic processes, and assists as a dietary addition of animal forages. By a simple gas phase reaction among trimethylamine, ethylene oxide, and HCl, it has been commercially produced.¹⁷ Though, choline chloride can be used as potential green synthon in deep eutectic solvent nevertheless the production of ethyleneoxide is not so green and the product it-self is toxic. Immediate local irritation of the eyes, skin, and upper respiratory tract by ethylene oxide gas may produce. At high concentrations, it may cause an accretion of fluid in the lungs. Breath of ethylene oxide can produce CNS depression, and in risky cases, respiratory distress and coma. Based on the report in 2003 by Abbott *et al.*¹⁸ the foundation for solvents from the renewable resources established. According to that, low melting mixtures of urea/carbohydrate and ChCl are liquid at room temperature named as “deep eutectic solvents” (DESS). This is defined as a simple halide salt can produce liquids with the combination of hydrogen bond donor (HBD) and the physical/chemical properties are similar to ILs nevertheless, most of the time ILs melting/freezing points are unpredictable (See Table 1).

Our research group is being contributed on precise reviews in green or sustainable and unconventional bio-based solvents for organic

Table 1 Melting, freeing points and the molar ratio of sugars with ChCl.^a

Entry	Compound	Structure	ChCl:HBD ratio		<i>T</i> (°C)	Ref.
1	D-Mannitol		1:1	<i>T_f</i>	108	19
2	D-Fructose		1:2	<i>T_f</i>	5	19
3	D-Glucose		1:2	<i>T_f</i>	14	19
4	D-Xylitol		1:1	<i>T_m</i>	Liquid at rt	20
5	D-Sorbitol		1:1	<i>T_m</i>	Liquid at rt	20
6	D-Isosorbide		1:2	<i>T_m</i>	Liquid at rt	20

^a ChCl:HBD ratio in (mol: mol).

transformations.^{21,22} Nevertheless, to our curiosity, the concise reviews on the utility of carbohydrates as solvents are very limited at the current scenario and the use of sugar/carbohydrate-based low organic melt and aqueous carbohydrates as solvents for organic reactions continues particularly to increase. The present review summarizes thus the recent developments on organic reactions in low melting/aqueous sugar mixtures between 1996 to till date.

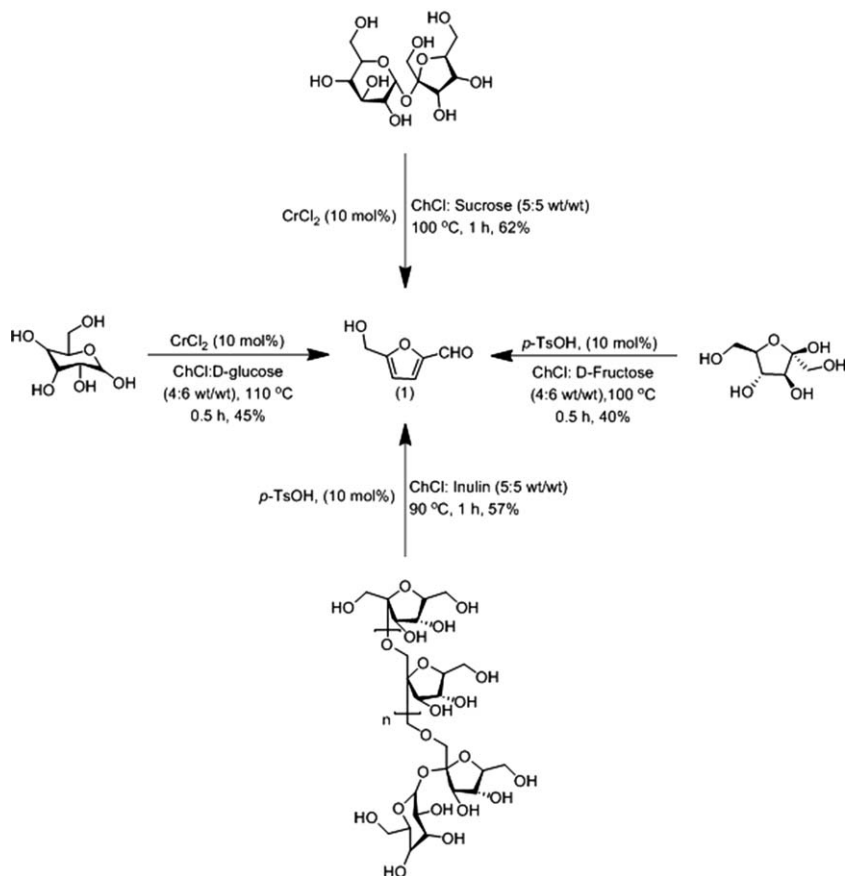
2 Organic Transformations in carbohydrate-based low melt

In the utilization of biomass into the useful chemicals, in recent time, bulk carbohydrates used as a promoting green medium and importantly, König *et al.* have developed a low melting carbohydrates mixtures as a solvent for many organic reactions such as Diels–Alder, Stille, Heck reaction, and cycloaddition reaction. The melting point of this low melting carbohydrate mixtures fall between 65 °C to 85 °C, which is the opted solvent medium with the combination of D-fructose, D-sorbitol, D-glucose and D-maltose and inorganic salts includes NH₄Cl and CaCl₂, urea and *N,N'*-dimethylurea (DMU). The polarity of these melt mimic like water and DMSO significantly, the physicochemical behaviour of these melt can be tailored based on the reaction ambience.

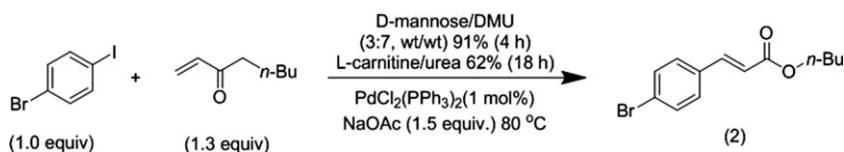
5-(Hydroxymethyl)furfural (HMF, **1**) is one of the most important building block in a number of organic synthesis significantly by rehydration it produces levulinic and formic acid. By the condensation, it produces soluble polymers and insoluble humins. Of late, HMF (**1**) was synthesized and described by Ilgen *et al.*²³ whereas, concentrated CHCl and 50 wt% of carbohydrate (2.9 to 3.1 mol L⁻¹) have been employed to convert sugar contents into HMF with the aid of catalysts (CrCl₂ or *p*-TsOH). Monosaccharides such as D-fructose and D-glucose were also converted into the corresponding HMF with a moderate yield, adequately, disaccharide includes sucrose and polyfructan inulin was also be successfully demonstrated (Scheme 1).

Similarly, in recent time, Heck reaction (product **2**), Sonogashira cross-coupling, Cu-catalysed 1,3-dipolar cycloaddition and Diels–Alder reaction have been described by König *et al.*²⁴ and in their report, a new L-carnitine/urea melt developed. The synthesized melt utilized in the reactions and the obtained results were compared with the sugar and sugar alcohol melts. The melting point and polarity of the L-carnitine/urea melt were evaluated by differential scanning calorimetry (DSC) and solvatochromatic measurements, respectively. However, the lower thermal stability of L-carnitine/urea melt, its applicability to the high-temperature organic reactions are limited (Scheme 2). Likewise, the same research group, described the Rh-catalysed hydrogenation and Pd-catalysed Suzuki reaction in 2006,²⁵ and this time the mixture of sugars, sugar alcohols or citric acid with urea and inorganic salts were used to accelerate the reactions affords the final product (**3**) under the clean condition. The polarity of these melt has been identified and it behaves like *N,N*-dimethylformamide and water closely and the obtained results suggest that non-toxic sugar–urea–salt melt are the feasible sustainable medium for organic reactions (Scheme 3).

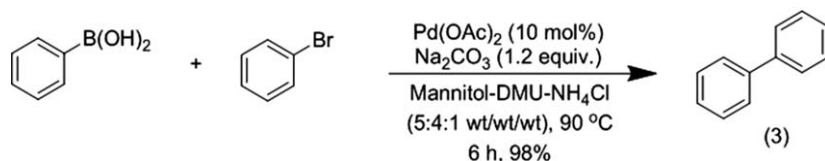
Equally, again Diels–Alder reaction in carbohydrate melt was developed and reported by Imperato *et al.*¹⁵ and the reaction between cyclopenta-1,3-diene and *n*-butyl acrylate led to the formation of Diels–Alder products (**4c–4f**) with an *endo/exo* ratio of 3.0:1, which was measured by gas chromatography. D-Fructose/DMU (7:3 wt/wt) exhibit



Scheme 1 Synthesis of HMF from mono- and disaccharides.

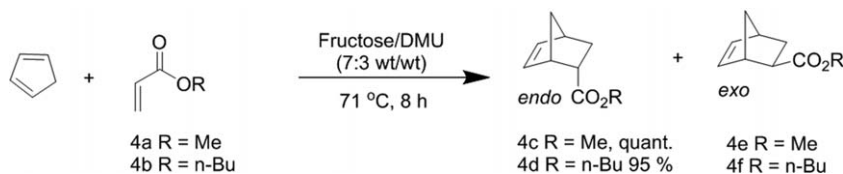


Scheme 2 Sugar melt promoted Heck coupling between *n*-butyl acrylate and 4-bromoiodobenzene.

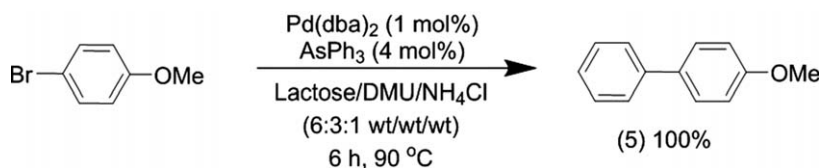


Scheme 3 Pd-catalysed Suzuki coupling in sugar-urea-salt melt.

better yield of 95% without the aid of any external catalyst, in addition glucose/urea/CaCl₂ (5:4:1 wt/wt/wt) also showed promising product yield of 93% with an *ee* of 2.6:1 (Scheme 4).



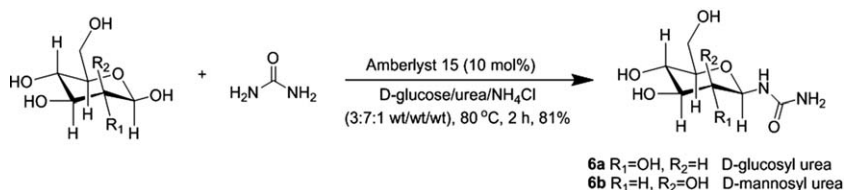
Scheme 4 Catalyst-free Diels–Alder reaction in fructose/DMU melt.



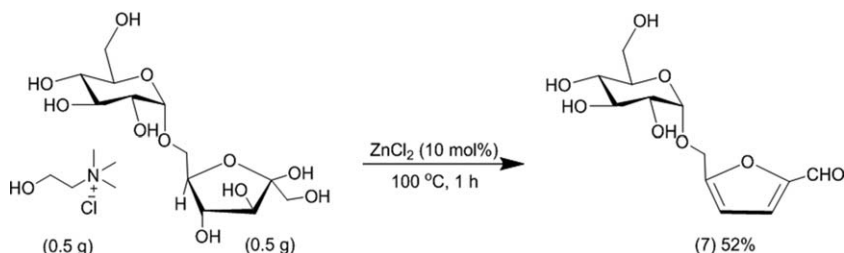
Scheme 5 Stille coupling performed in sugar-urea-salt melt.

Stille reaction involve the transfer of simple alkyl group needs some special condition of solvents such as hexamethylphosphoramide or organotin reagents (stannatranes, monoorganotin halides). Recently, the sustainable green solvent medium, low-melting mixture of sugar, urea and inorganic salts promote the palladium-catalysed alkyl transfer of tetraalkyltin reagents have been extensively studied.²⁶ In credit, the high polarity and nucleophilic character of the organic low melt promoted the biaryl synthesis using electron-poor and electron-rich aryl bromides with the 100% of conversion and also in excellent yield (5). Stille alkylation/coupling of arylbromides and tributylphenylstannane, Stille coupling of 4-bromoanisole and phenyltributylstannane also were successfully documented recently (Scheme 5).

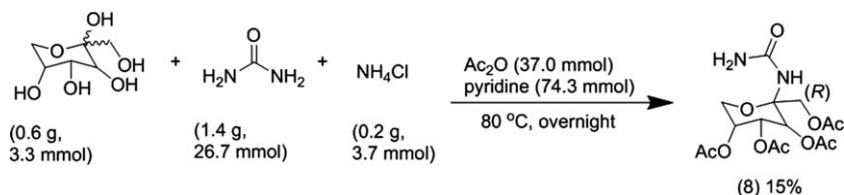
Unprotected hexoses converted into sugar-ureides (**6a** and **6b**) in sugar melt such as D-glucose/urea/NH₄Cl (3 : 7 : 1 wt/wt/wt) under the influence of acid catalyst Amberlyst-15 has been described by Ruß *et al.*²⁷ In a one-pot synthesis, β-D-glucosyl- and β-D-mannosyl urea were obtained in a moderate to good yield (81%). The conversion of D-galactose, N-acetyl-D-glucosamine, L-rhamnose and 2-deoxy-D-glucose into the corresponding glycosyl ureas were greatly demonstrated whereas, N,N'-ethylene urea, N,N'-allylurea and ethyl carbamate were employed as a urea source. However, N-octylurea did not produce any remarkable product (Scheme 6). Interestingly, the same researchers,²⁸ utilized isomaltulose-choline chloride melt as a starting material in 1:1 ratio (0.5 g) with the concentration of 1.7 mol L⁻¹ to prepare the important building block of 5-(α-D-glucosyloxymethyl)furfural (GMF) (7) in good yield (52%). ZnCl₂ (10 mol%) used as an acid catalyst to promote the conversion by the dehydration reaction with the shorter reaction time of 1 h (Scheme 7). Similar protocol has been adopted to synthesize sorbosyl urea and HMF from sorbose-urea and L-sorbose-choline chloride melt.²⁹ Under the acidic conditions at overnight which affords the desired N-(1,3,4,5-tetra-O-acetyl-L-sorbopyranosyl)urea (8) with the 15% of yield, similarly HMF (9) also been prepared with 11–24% of product yield (Schemes 8 and 9).



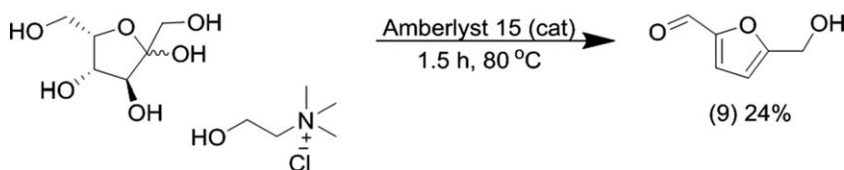
Scheme 6 D-Glucose/urea/ NH_4Cl triggered formation of β -form condensation product 6a/b.



Scheme 7 Conversion of isomaltulose to 5-(α -D-glucosyloxymethyl)furfural catalysed by ZnCl_2 .

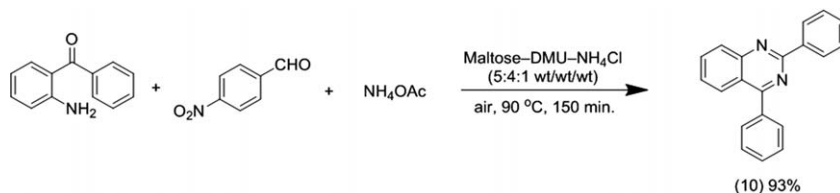


Scheme 8 Synthesis of *N*-(1,3,4,5-tetra-*O*-acetyl-L-sorbosepyranosyl)urea from sorbose-urea melt.

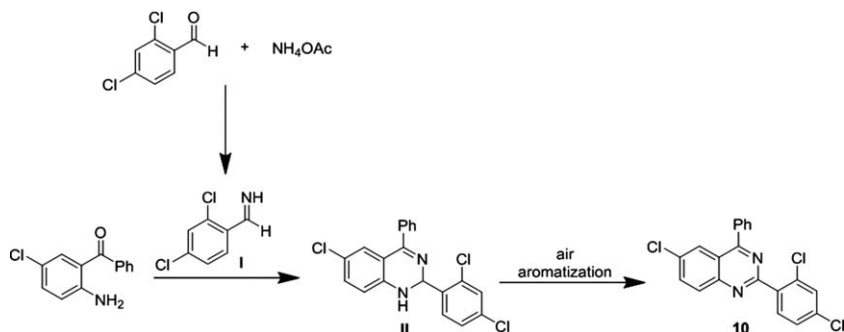


Scheme 9 Synthesis of HMF from L-sorbose in choline chloride melt.

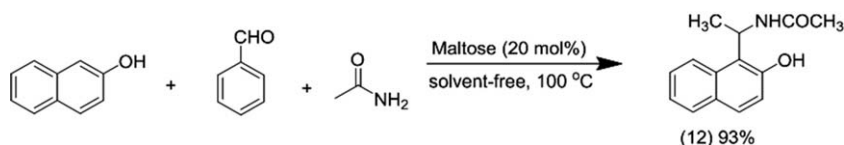
Quinazolines proved to be an important nitrogen-containing heterocycle usually found in a number of pharmaceutical molecules, natural products and functional materials. This type of privileged molecules developed under the green medium of a mixture of D-maltose-DMU- NH_4Cl (5:4:1 wt/wt/wt).³⁰ Efficient one-pot, three-component reaction between 2-aminoaryl ketones, aldehyde, and ammonium acetate affords the respective quinazolines (**10**) in higher yields of 93%. The mechanism of this three-component reaction involves, the reaction between 2,4-dichlorobenzaldehyde and ammonium acetate gives the intermediate **I** and it further condensed with (2-amino-5-chlorophenyl)(phenyl)methanone and intermediate **II** has formed. Under the aerobic oxidation, the



Scheme 10 Synthesis of quinazolines in low melt sugar.



Scheme 11 Mechanistic pathway of synthesis of quinazolines under the influence of maltose-DMU- NH_4Cl melt.



Scheme 12 Maltose-catalysed synthesis of 1-amidoalkyl-2-naphthol.

intermediate **II** pave the way to form quinazolines **10** whereas, the trace amount of desired product obtained under nitrogen ambience. The results suggest, oxygen plays a significant role in the increment of the reaction yield further (Schemes 10 and 11).

Disaccharide such as maltose not only used as a solvent in the form of the organic low melt, which also been utilized as a catalyst to promote the one-pot, three-component reaction. The synthesis of 1-amidoalkyl-2-naphthols (**12**) from 2-naphthol, amides or urea and aromatic aldehydes has been developed by Adrom *et al.*³¹ and 20 mol% of the maltose at 100°C promoted the reaction with yields up to 93%. The other sugar catalysts like D-lactose, D-sucrose and D-xylose also been employed in this reaction but they showed a lower product yield (48%, 65%, 25% respectively). The mechanism of this reaction involves 2-naphthol reacts with aldehydes in the occurrence of maltose to produced *ortho*-quinone methides (*o*-QMs). By the conjugate addition of amides with *o*-QMs led to the formation of 1-amidoalkyl-2-naphthol derivatives (Scheme 12).

3 Organic Transformations in an aqueous Sugar

The organic reactions in water exhibit many benefits moreover, the solubility of the substrates is the major issue to be solved. The number of monosaccharides are highly hydrophilic and readily dissolve in water, also glyco-organic substrates incorporate a carbohydrate moiety upsurges the water solubility, tuning the polarity and inducing chirality.^{32,33} Anyhow, no surety of expected advancement in the reaction rate or selectivity when the use of this aqueous sugars have been proved. Nevertheless, in 1996 Plusquellec *et al.* established a new approach to accessing an aqueous glycosidic media as a sustainable and effective medium for the regio- and stereoselective reductions of carbonyl compounds.³⁴ Afterwards, a number of methodologies have been developed and reported on aqueous carbohydrates in recent time. At the beginning, the regioselective reductions of α,β -unsaturated ketones have been studied and documented by Plusquellec,³⁴ in 1996 and the conversion of corresponding allylic alcohols was obtained in moderate to good yield. One molar solution of glycosidic surfactants/amphiphilic carbohydrates such as non-ionic glycosidic surfactants **13**, **14** and aqueous solutions of amphiphilic carbohydrates **15**–**17** (see Fig. 1) were used as a sustainable green medium and sodium borohydrate employed as a reducing agent. Importantly, under the similar ambience, cyclohexanones and cyclohexenones also were reduced to its corresponding allylic alcohols (**18**), it contains an equatorial alcohol function (Scheme 13).

Later on, in 2001,³⁵ the similar protocol has been adopted in the epoxidation of allylic alcohols and the different cyclic and acyclic allylic alcohols were successfully converted into the respective chemo-, regio- and/or stereoselective epoxidation products (**19**) with the good yields range between 28%–92%. Amphiphilic carbohydrates such as sucrose (**15**), L-arabinose (**16**), and methyl or ethyl β -D-fructopyranoside (**17a,b**)

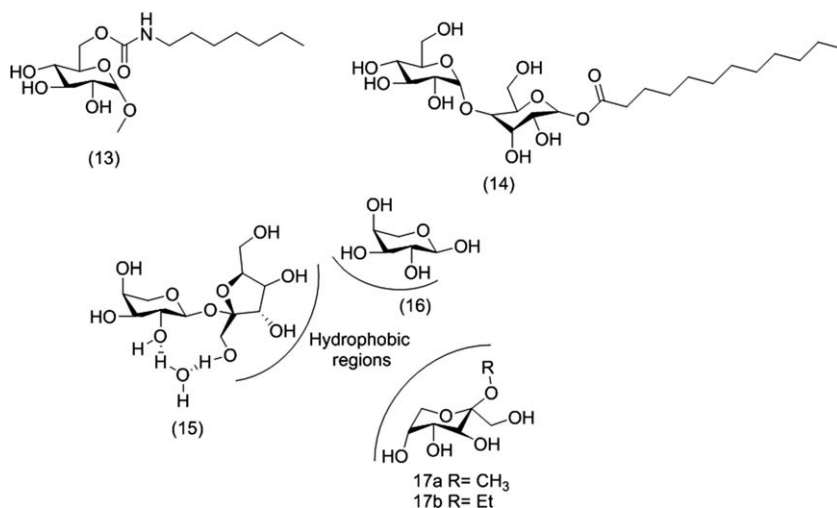
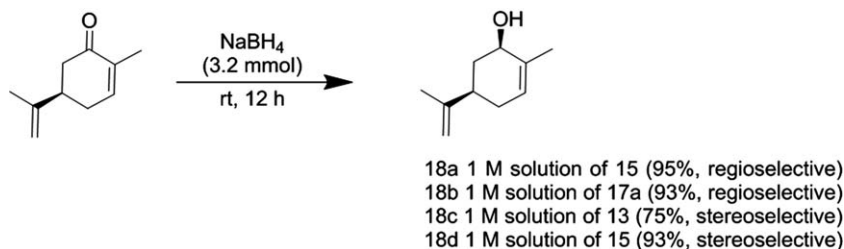
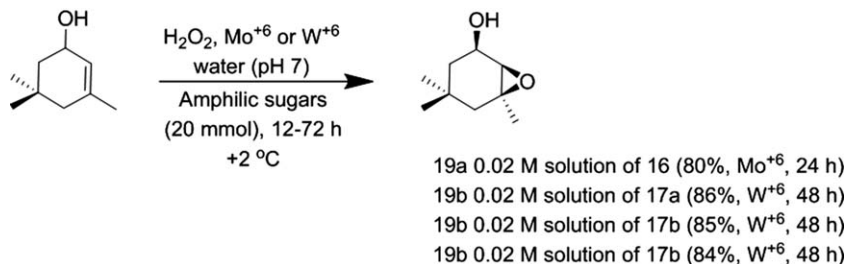


Fig. 1 Structure/hydrophobic regions of non-ionic glycosidic surfactants and amphiphilic carbohydrates **13**–**17**.



Scheme 13 Regio/stereoselective conversion of carbonyl compounds in aqueous sugars.



Scheme 14 Chemo-, regio- and/or stereoselective epoxidation of allylic alcohols.

with the pH 7 were used to promote the reaction under the influence of H₂O₂-Mo⁺⁶ or W⁺⁶ metal catalysts (Scheme 14).

Similar protocol has also been described by Plusquellec *et al.*³⁶ in 2010 but this time the indium-catalysed allylation reaction was achieved in moderate to good yields. During the reaction with substituted cyclohexanones to consistent alcohols, facial amphiphilic character of the sugars was observed with the pyranoside ring which facilitates the solubility of the carbohydrate in an aqueous region and water (see Fig. 2). 1 M of an aqueous solution of fructopyranosides (**20** and **21**) utilized in the promotion of this allylation of cyclohexanones to alcohols (**22-ax/eq**) whereas, stereochemical outcome of the reaction also been screened by the same solvent medium (Scheme 15).

In 2012, same researchers,³⁷ developed direct aldol reaction in amphiphilic carbohydrates directed by organocatalysts and aldol condensed products (**23**) were achieved in a good yield with high diastereoselectivity. The example for the effect of this green sustainable medium in this reaction reflects, reduced the quantity of the catalyst needed only 2% of organocatalyst to take forward the reaction. The obtained results in aqueous carbohydrates were compared with the reactions carried out in water and it proved that an aqueous carbohydrates are efficient green medium to promote the reaction with higher yield in a shorter time (Scheme 16).

Starch is the carbohydrate derived inexpensive and completely ecological natural biopolymer. It has been largely employed as a surfactant in many of the organic synthesis.³⁸ When the combination of starch and acids includes amino acid, fatty acid, sulfuric acid, this system promote the reaction very efficiently and this phenomenon is well-documented in

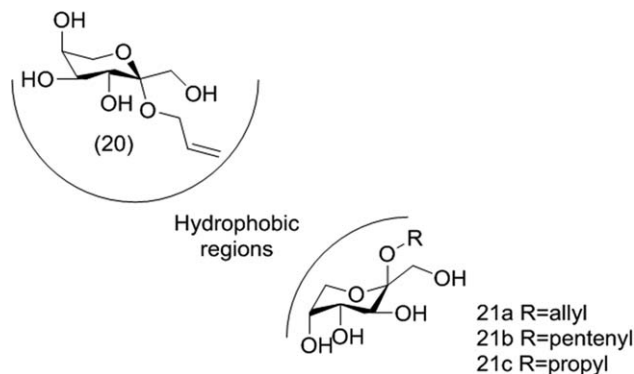


Fig. 2 Structure and hydrophobic regions of facial amphiphilic fructopyranosides.

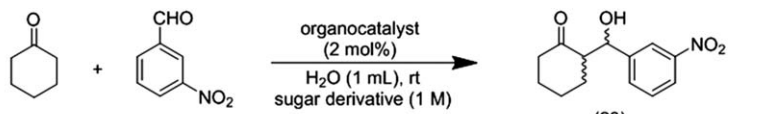


R	Solvent	time (h)	yield (%)	22-ax/22-eq
4-Me	Water	2.5	88	3: 1
4-Me	1 M aqueous solution of 21a	2.0	75	1.56: 1
4-Me	1 M aqueous solution of 21c	24.0	75	1.77: 1
4-Me	1 M aqueous solution of 21b	1.5	75	2.33: 1
3-Me	Water	6.0	91	5.25: 1
3-Me	1 M aqueous solution of 21a	6.0	91	6.70: 1
2-OMe	Water	8.0	93	24: 1
2-OMe	1 M aqueous solution of 21a	8.0	90	19: 1
2-SMe	Water	7.0	79	6.14: 1
2-Sme	1 M aqueous solution of 21a	7.0	60	4: 1

Scheme 15 Allylation of cyclohexanones with allyl bromide in facial amphiphilic fructopyranosides catalysed by In(0).

the modern years.^{39,40} Recently, Mannich reaction under the catalyst of starch sulfonic acid (SSA) in water has been described by Wu *et al.*⁴¹ and in this reaction aldehydes, amines, and ketones were used as benchmark partners, affords the respective β -amino ketones (24) at higher yields with excellent stereoselectivity (*anti*-selectivity). At room temperature for 8–12 h, the reaction produces the desired product and the catalyst SSA used in the minimum quantity to run the reaction (0.08 g). In this reaction, there is no hydrolysis of the heterogeneous catalyst, it is starch. (Scheme 17).

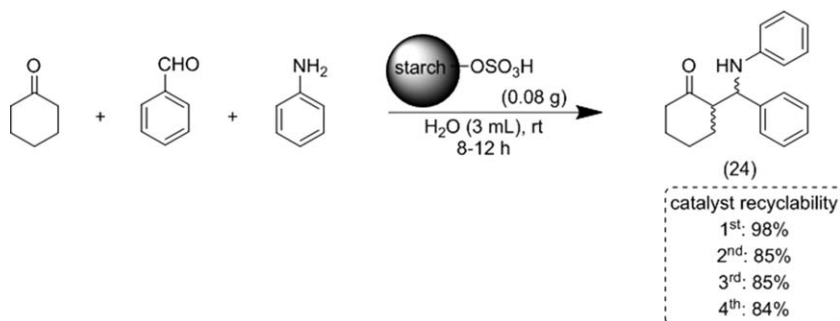
Blackmond *et al.*⁴² have developed an enantioselective synthesis of amino acid precursors (25) catalysed by chiral pentose sugars and the magnitude of the chiral induction observed by the indirect cooperation between sugar hydroxyl groups. The opposite chiral preferences were obtained from D-ribose and D-lyxose whereas, the theoretical calculations disclose the pseudoenantiomeric behaviour of transition state structures from D-ribose and D-lyxose. Interestingly, mixtures of natural D-sugars (0.50 M) give enantioenriched natural L-amino acid precursors under the identical conditions. The obtained results are evidence for the interaction of sole chirality of the two most vital classes of biological molecules and



entry	organocatalyst	additive	time (h)	Yield (%) ^a	<i>anti/syn</i>	<i>ee</i> ^b (%)
1		-	96	No reaction	-	-
2		-	36	90	1.1: 1	<5
3		17b	36	87	1.4: 1	<5
4		-	1.5	85	1: 1.3	<5
5		17b	1.0	86	1.9: 1	<5
6		-	72	77	3: 1	<5
7		17b	72	65	2.6: 1	<5
8		-	2.5	88	3: 1	<5
9		17b	1.5	73	9: 1	<5
10		-	4	90	3: 1	<5
11		17b	3	91	13.3: 1	<5
12		15	2.5	88	7.3: 1	<5

^a Combined yield of isolated diastereoisomers. ^b Determined through ¹H NMR experiments in the presence of Europium.

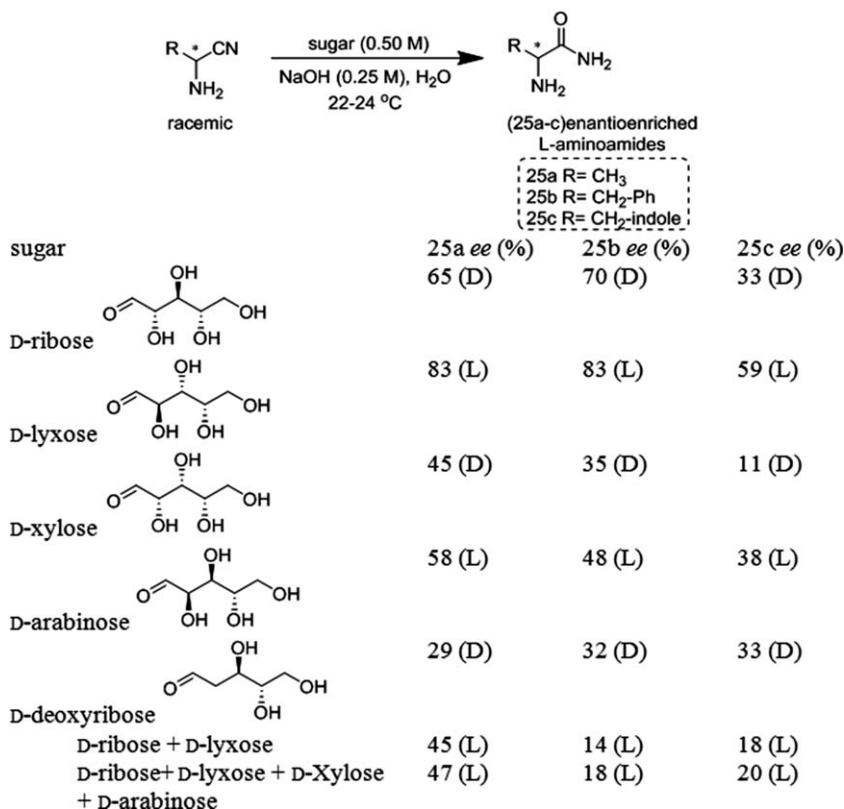
Scheme 16 Direct aldol reaction of cyclohexanone catalysed by various organocatalysts.



Scheme 17 Mannich reaction between aldehydes, amines, and ketones in SSA/water system.

sugars, are capable of making DNA, RNA and the amino acids lead to the formation of proteins (Scheme 18).

In a modern area, some of the bio-based materials, such as carbohydrates in the form of monosaccharides were employed in producing low melting mixtures and an aqueous solution of carbohydrates that can be consumed in some organic conversions.⁴³ In green chemistry, meglumine used as a promoting medium with water in the construction of many of the industrially important heterocycles. Meglumine is a hexosamine converting from sorbitol with molecular formula C₇H₁₇NO₅. Owing to its extraordinary low toxic properties, it has been employed in a number of pharmaceutical formulations as an excipient, and also a mixture of meglumine-iodinated compounds such as diatrizoate



Scheme 18 Enantioselective synthesis of enantioenriched amino acids from racemic aminonitriles under the influence of chiral sugars.

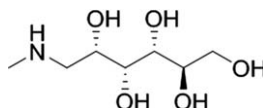


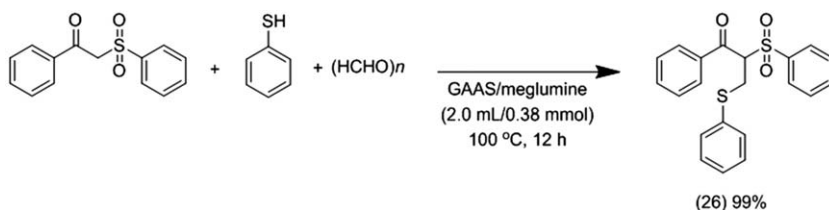
Fig. 3 Structure of meglumine constituted with amino group, primary and secondary hydroxyl groups.

meglumine and iodipamide meglumine used as optical brighteners in contrast media. Meglumine encompasses with an amino group, primary and secondary hydroxyl groups that can be facilitate to activate the nucleophilic and electrophilic substrates in the reaction vessel by means of hydrogen bonding and contribution of the lone pair of electrons correspondingly (see Fig. 3). To the credit of its extraordinary properties, such as biodegradability, physiological inertness, low toxicity, inexpensive price, non-corrosion nature and reusability, meglumine has enthused swelling of research curiosity as an imperative and operative catalyst/solvent in organic synthesis.^{44–49}

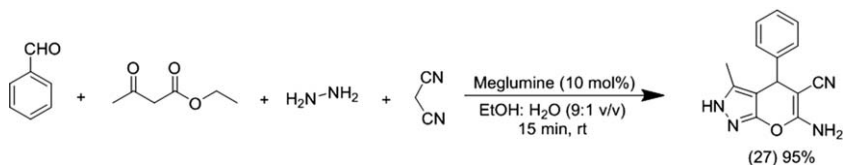
Gu *et al.*⁵⁰ developed a mixture of two bio-based chemicals as a solvent system for the hydroxymethylation of β -ketosulfones with formaldehyde which affords hydroxymethylation product (26). In this report, meglumine and gluconic acid aqueous solution (GAAS, 50 wt%) (0.38 mmol of

meglumine in 2.0 mL of GAAS) were employed to be a task-specific bio-based solvent. The one-pot, stepwise, three- and four-component reactions also were performed with obtained hydroxymethylation product (26) and the nucleophiles includes β -ketosulfones, α -bromo ketone, sodium benzenesulfinate, thiophenol and formaldehyde also employed in this green synthesis. The GAAS and meglumine system is highly hydrophilic and the solvent has been recovered and reused without the significant loss of activity at the minimum of four spells in the three-component reaction of paraformaldehyde, β -ketosulfone and α -methylstyrene (Scheme 19).

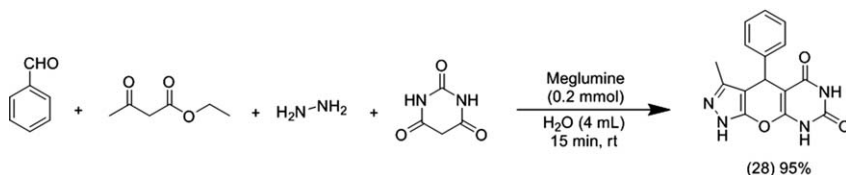
Meglumine with water/ethanol mixture has also been used as a promoting medium (solvent as well as a catalyst) to synthesize pyrazolopyranopyrimidines, pyranopyrazoles, functionalized 2-amino-4*H*-pyrans, pyrazoles bearing a coumarin unit and hydrazones lately, Zhan-Hui Zhang has contributed in this area for more than five research articles. Of late, pyranopyrazole derivatives have been synthesized and described *via* one-pot, four-component reaction between hydrazine hydrate, carbonyl compound or isatin, malononitrile, and β -keto ester in EtOH-H₂O (9 : 1 v/v) and the meglumine utilized as a catalyst in 10 mol% at room temperature.⁵¹ The mechanism of this reaction involves, Knoevenagel condensation of benzaldehyde and malononitrile lead to the formation of intermediate and this intermediate condensed with ethyl acetoacetate and hydrazine, which would be transformed to its consistent enolate form in the presence of catalyst. Michael-type addition, intramolecular cyclization and finally, tautomerization gave the path to the formation of the desired product pyranopyrazoles (27) in moderate to good yield (Scheme 20). A year later,⁵² the similar starting materials except of malononitrile instead barbituric acid employed and the previous synthetic protocol has been employed to synthesize pyrazolopyranopyrimidines in meglumine-water mixture (0.2 mmol/4 mL) at room temperature for 15 minutes. In this synthesis, just only a low quantity of the catalyst (0.2 mmol) was used to achieve the desired product (28) in



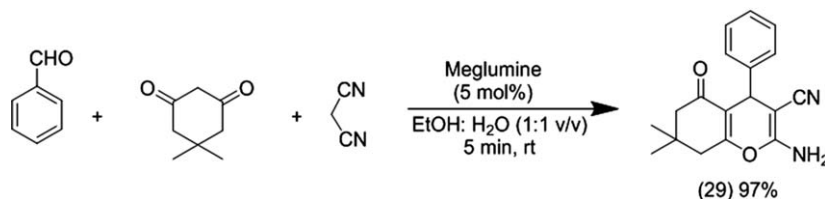
Scheme 19 One-pot reaction of β -ketosulfone, paraformaldehyde and thiophenol or thiol in GAAS/meglumine.



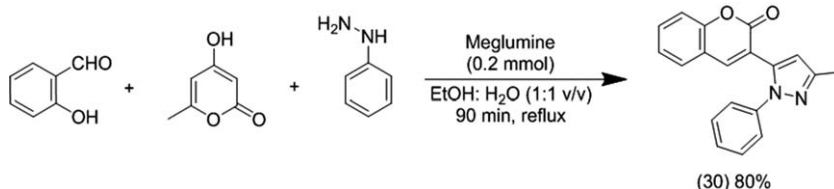
Scheme 20 Synthesis of pyranopyrazole derivatives in meglumine.



Scheme 21 Meglumine-catalysed four-component reaction of ethyl acetoacetate, benzaldehyde, hydrazine hydrate, barbituric acid.



Scheme 22 Synthesis of 2-amino-4H-pyrans in EtOH/water system.



Scheme 23 One-pot synthesis of pyrazolylcoumarins catalysed by meglumine.

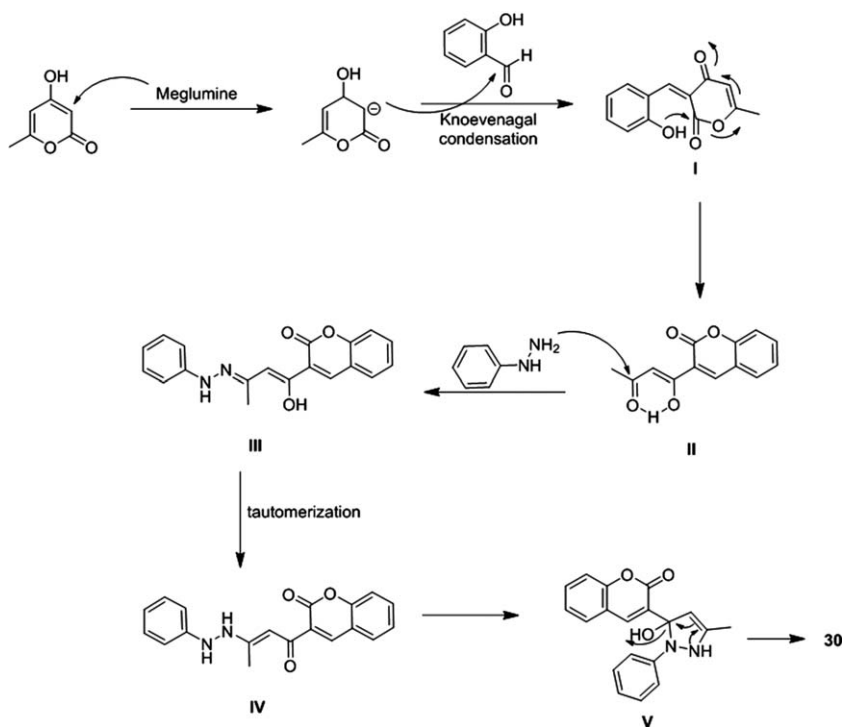
95% yield. The other organic base catalysts were also tested in the reaction system and all failed to promote the reaction efficiently (Scheme 21).

Also, an efficient one-pot, three-component synthesis of diversely functionalized 2-amino-4H-pyrans (29) has been reported by Zhang *et al.*⁵³ Benzaldehyde (1 mmol), 5,5-dimethylcyclohexane-1,3-dione (1 mmol), malononitrile (1 mmol) were exploited as a benchmark partner in EtOH/water medium (1:1 v/v). A wide range of substrates scope has been established by this methodology which includes aromatic/heteroaromatic aldehydes, acenaphthenequinone/isatin derivatives involved the condensation of enolizable C–H activated compounds and alkylmalonates to produce the anticipated products in excellent yields (97% at maximum). The advantage of this methodology includes shorter reaction time, higher product yield, biodegradable and inexpensive solvent/catalyst medium, mild reaction conditions and simple work-up procedure (Scheme 22).

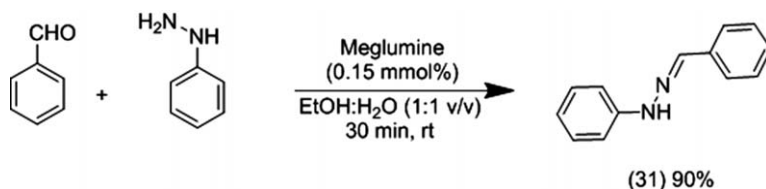
Recently, pyrazoles bearing a coumarin unit has been developed and reported by the same group⁵⁴ adopted one-pot, three-component reaction between salicylaldehyde, 4-hydroxy-6-methyl-2H-pyran-2-one, and hydrazine in water-ethanol solvent medium (1:1 v/v) under the influence of 0.2 mmol of meglumine as a catalyst at the reflux condition, which affords the desired pyrazolylcoumarins (30, 80%) in moderate to good yields (Scheme 23). The mechanism of this reaction involves Knoevenagel condensation between salicylaldehyde and 4-hydroxy-6-methyl-2H-pyran-2-one catalysed

by meglumine lead to the formation of intermediate **I**. The intramolecular cyclization of intermediate **I** by the reaction of enolate oxygen to carbonyl group *via* nucleophilic addition to paying for the intermediate **II**. Phenylhydrazine further reacted with intermediate **II** followed by the tautomerization to produce the intermediate **IV**. Finally, meglumine promoted intramolecular cyclization of intermediate **IV** afforded the desired pyrazolylcoumarins by eradicating one molecule of water (Scheme 24).

Very recently, a simple, environmentally benevolent procedure for the synthesis of hydrazones (**31**) from hydrazides and carbonyl compounds has been described by Zhang *et al.*⁵⁵ and the inexpensive catalyst meglumine (0.15 mmol) in an aqueous-ethanol (1:1 v/v) system at room temperature were adopted in this. The salient features of this method proved that, metal-free synthesis, short reaction time, operational simplicity, mild reaction conditions, high product yields, applicability towards gram-scale synthesis (Scheme 25).



Scheme 24 Mechanistic pathway for synthesis of pyrazolylcoumarins catalysed by meglumine.



Scheme 25 Synthesis of hydrazones catalysed by meglumine

4 Conclusions and Outlook

In summary, the present contribution aims to project on the beginning and development of the use of bio-based material, such as carbohydrate melt and aqueous carbohydrates, as a solvent, hydrogen donor and a catalyst for a number of organic transformations. Though the up growing research in this ground is still limited, the reports available from the last seven years becomes outward that this waste generated by the bio-based industries and other sources can be a consistent substitute to swap the classical and hazardous organic solvents.

The existing accessibility of carbohydrate on a bulk scale with cheaper price, collectively with its essential properties, *i.e.* high polarity, low vapour pressure, no flammability, biodegradability, no toxicity and obtaining from renewable sources makes it a perfect candidate to grow greener synthetic developments. This carbohydrate solvents meaningfully subsidise to the future bio-based economy and the best examples of the advancements in the sustainable technologies at biomass processing. However, with a respect of all the positive reflections of carbohydrate solvents, some problems also be associated with it such as its high polarity and high viscous nature, besides, the active metal coordinating hydroxyl groups also must be taken into the account for its utilization in a particular case. Even with, the solvent properties of this bio-based solvent can be tailored based on the requirement thus, facilitate to overcome some of the above-said disadvantages.

Last but not least, while the environmental impact and low toxicity of carbohydrates as a solvent are frequently presumed, additionally, toxicological studies are essential if these type of solvents are successfully to be used in larger scale at industries.

References

- 1 D. Reinhardt, F. Ilgen, D. Kralisch, B. König and G. Kreisel, *Green Chem.*, 2008, **10**, 1170–1181.
- 2 L. Moity, M. Durand, A. Benazzouz, C. Pierlot, V. Molinier and J.-M. Aubry, *Green Chem.*, 2012, **14**, 1132–1145.
- 3 (a) M. Petkovic, K. R. Seddon, L. P. N. Rebelo and C. Silva Pereira, *Chem. Soc. Rev.*, 2011, **40**, 1383–1403; (b) A.-O. Diallo, G. Fayet, C. Len and G. Marlair, *Ind. Eng. Chem. Res.*, 2012, **51**, 3149–3156; (c) A.-O. Diallo, C. Len, A. B. Morgan and G. Marlair, *Sep. Purif. Technol.*, 2012, **97**, 228–234; (d) A.-O. Diallo, A. B. Morgan, C. Len and G. Marlair, *Energy Environ. Sci.*, 2013, **6**, 699–710; (e) L. Chancelier, A.-O. Diallo, C. C. Santini, G. Marlair, T. Gutel, S. Mailley and C. Len, *Phys. Chem. Chem. Phys.*, 2014, **16**, 1967–1976; (f) A. B. Nilles, A.-O. Diallo, G. Marlair, P. Pandard, L. Chabot, A. Geffard, C. Len, J.-M. Porcher and W. Sanchez, *J. Hazard. Mater.*, 2015, **283**, 202–210.
- 4 S. Stolte, J. Arning and J. Thöming, *Chem. Ing. Tech.*, 2011, **83**, 1454–1467.
- 5 M. Petkovic, J. L. Ferguson, H. Q. N. Gunaratne, R. Ferreira, M. C. Leitão, K. R. Seddon, L. P. N. Rebelo and C. S. Pereira, *Green Chem.*, 2010, **12**, 643–649.
- 6 K. Tanaka and F. Toda, *Chem. Rev.*, 2000, **100**, 1025–1074.
- 7 Y. Roman-Leshkov, C. J. Barrett, Z. Y. Liu and J. A. Dumesic, *Nature*, 2007, **447**, 982–985.

- 8 Y. Roman-Leshkov, J. N. Chheda and J. A. Dumesic, *Science*, 2006, **312**, 1933–1937.
- 9 D. Peters, *Chem. Ing. Tech.*, 2006, **78**, 229–238.
- 10 R. R. Schmidt and W. Kinzy, in *Advances in Carbohydrate Chemistry and Biochemistry*, ed. H. Derek, Academic Press, 1994, vol. 50, pp. 21–123.
- 11 J. D. Wander and D. Horton, in *Advances in Carbohydrate Chemistry and Biochemistry*, ed. R. S. Tipson, Academic Press, 1976, vol. 32, pp. 15–123.
- 12 M. D. Lewis, J. K. Cha and Y. Kishi, *J. Am. Chem. Soc.*, 1982, **104**, 4976–4978.
- 13 I. Paterson and L. E. Keown, *Tetrahedron Lett.*, 1997, **38**, 5727–5730.
- 14 W. Pigman, D. Horton and A. Herp, *The Carbohydrates: Chemistry and Biochemistry*, Academic Press, New York, 2nd edn, 1970.
- 15 G. Imperato, E. Eibler, J. Niedermeier and B. König, *Chem. Commun.*, 2005, 1170–1172.
- 16 The Merck Index: An Encyclopedia of Chemicals, Drugs, and Biologicals, 11th edn, Merck, 1989.
- 17 C. Atwater, *Ullmann's Encyclopedia of Industrial Chemistry*, Wiley-VCH Verlag GmbH & Co. KGaA, 2000.
- 18 A. P. Abbott, G. Capper, D. L. Davies, R. K. Rasheed and V. Tambyrajah, *Chem. Commun.*, 2003, 70–71.
- 19 A. P. Abbott, G. Capper, D. L. Davies, R. K. Rasheed, V. Tambyrajah, WO 02/26701 (A2), 2002 (Scionix Limited).
- 20 Z. Maugeri and P. D. de Maria, *RSC Adv.*, 2012, **2**, 421–425.
- 21 Y. Gu and F. Jérôme, *Chem. Soc. Rev.*, 2013, **42**, 9550–9570.
- 22 Y. Gu, *Green Chem.*, 2012, **14**, 2091–2128.
- 23 F. Ilgen, D. Ott, D. Kralisch, C. Reil, A. Palmberger and B. König, *Green Chem.*, 2009, **11**, 1948–1954.
- 24 F. Ilgen and B. König, *Green Chem.*, 2009, **11**, 848–854.
- 25 G. Imperato, S. Hoger, D. Lenoir and B. König, *Green Chem.*, 2006, **8**, 1051–1055.
- 26 G. Imperato, R. Vasold and B. König, *Adv. Synth. Catal.*, 2006, **348**, 2243–2247.
- 27 C. Ruß, F. Ilgen, C. Reil, C. Luff, A. H. Begli and B. König, *Green Chem.*, 2011, **13**, 156–161.
- 28 C. Ruß, F. Ilgen, C. Reil, C. Luff, A. H. Begli and B. König, *Synth. Commun.*, 2012, **42**, 3112–3116.
- 29 C. Ruß, A. H. Begli and B. König, *Synth. Commun.*, 2013, **43**, 2452–2456.
- 30 Z.-H. Zhang, X.-N. Zhang, L.-P. Mo, Y.-X. Li and F.-P. Ma, *Green Chem.*, 2012, **14**, 1502–1506.
- 31 B. Adrom, N. Hazeri, M. T. Maghsoodlou and M. Mollamohammadi, *Res Chem Intermed.*, 2015, **41**, 4741–4747.
- 32 C. J. Li, *Chem. Rev.*, 1993, **93**, 2023–2035.
- 33 A. Lubineau, J. Auge, H. Bienayme and Y. Queneau, *Synthesis*, 1994, 741–760.
- 34 C. Denis, B. Laignel, D. Plusquellec, J.-Y. Le Marouille and A. Botrel, *Tetrahedron Lett.*, 1996, **37**, 53–56.
- 35 C. Denis, K. Misbahi, A. Kerbal, V. Ferrières and D. Plusquellec, *Chem. Commun.*, 2001, 2460–2461.
- 36 A. Bellomo, R. Daniellou and D. Plusquellec, *Tetrahedron Lett.*, 2010, **51**, 4934–4936.
- 37 A. Bellomo, R. Daniellou and D. Plusquellec, *Green Chem.*, 2012, **14**, 281–284.
- 38 J. E. Anderson, D. Casarini, A. I. Ijeh, L. Lunazzi and D. A. Tocher, *J. Am. Chem. Soc.*, 1995, **117**, 3054–3056.
- 39 (a) J. Merta, M. Torkkeli, T. Ikonen, R. Serimaa and P. Stenius, *Macromolecules*, 2001, **34**, 2937–2946; (b) H. Feitkenhauer, D. Fischer and D. Fähr, *Biotechnol. Prog.*, 2003, **19**, 874–879; (c) J. Merta, V. M. Garamus, A. I. Kuklin,

- R. Willumeit and P. Stenius, *Langmuir*, 2000, **16**, 10061–10068; (d) A. Ito, M. Hattori, T. Yoshida, A. Watanabe, R. Sato and K. Takahashi, *J. Agric. Food Chem.*, 2006, **54**, 10191–10196.
- 40 G. Zhang, M. D. Maladen and B. R. Hamaker, *J. Agric. Food Chem.*, 2003, **51**, 2801–2805.
- 41 Y. Wan, W. Lin, X.-M. Chen, Y. Shen, H.-Q. Xin, H.-H. Xu, R. Yuan, L.-L. Pang, R. Ma, C.-H. Yue, W. Yin, R.-C. Bo and H. Wu, *Lett. Org. Chem.*, 2009, **6**, 456–461.
- 42 A. J. Wagner, D. Yu. Zubarev, A. Aspuru-Guzik and D. G. Blackmond, *ACS Cent. Sci.*, 2017, **3**, 322–328.
- 43 A. P. Abbott, R. C. Harris, K. S. Ryder, C. D'Agostino, L. F. Gladden and M. D. Mantle, *Green Chem.*, 2011, **13**, 82–90.
- 44 M. Li, W. Kong, L.-R. Wen and F.-H. Liu, *Tetrahedron*, 2012, **68**, 4838–4845.
- 45 H. Gao, J. Sun and C.-G. Yan, *Tetrahedron*, 2013, **69**, 589–594.
- 46 Y. Li, Q.-Y. Li, H.-W. Xu, W. Fan, B. Jiang, S.-L. Wang and S.-J. Tu, *Tetrahedron*, 2013, **69**, 2941–2946.
- 47 L. L. Bian, J. C. Xu, L. Q. Xie, J. Chen, H. M. Deng, M. Shao, T. L. Ding, H. Zhang and W. G. Cao, *Tetrahedron*, 2013, **69**, 6121–6128.
- 48 Z.-J. Quan, Q. Xu, Z. Zhang, Y.-X. Da and X.-C. Wang, *Tetrahedron*, 2013, **69**, 881–887.
- 49 S. F. Chen, L. Li, H. Y. Zhao and B. G. Li, *Tetrahedron*, 2013, **69**, 6223–6229.
- 50 J. Yang, H. Li, M. Li, J. Peng and Y. Gu, *Adv. Synth. Catal.*, 2012, **354**, 688–700.
- 51 R.-Y. Guo, Z.-M. An, L.-P. Mo, S.-T. Yang, H.-X. Liu, S.-X. Wang and Z.-H. Zhang, *Tetrahedron*, 2013, **69**, 9931–9938.
- 52 X.-T. Li, A.-D. Zhao, L.-P. Mo and Z.-H. Zhang, *RSC Adv.*, 2014, **4**, 51580–51588.
- 53 R.-Y. Guo, Z.-M. An, L.-P. Mo, R.-Z. Wang, H.-X. Liu, S.-X. Wang and Z.-H. Zhang, *ACS Comb. Sci.*, 2013, **15**, 557–563.
- 54 X.-T. Li, Y.-H. Liu, X. Liu and Z.-H. Zhang, *RSC Adv.*, 2015, **5**, 25625–25633.
- 55 M. Zhang, Z.-R. Shang, X.-T. Li, J.-N. Zhang, Y. Wang, K. Li, Y.-Y. Li and Z.-H. Zhang, *Synth. Commun.*, 2017, **47**, 178–187.

Surfactants based on green/blue sugars: towards new functionalities in formulations

Louise Renault,^{a,b} Freddy Pessel^{a,b} and Thierry Benvegnu^{*a,b}

DOI: 10.1039/9781788010641-00196

Sugar-based surfactants are surface-active agents that are gaining more and more attention due to advantages regarding to performance, biodegradability, low toxicity and environmental compatibility. They have been commercially available for several decades and they represent attractive substitutes for petroleum-derived surfactants in many industrial applications such as food, detergency and cosmetics. Indeed, these poly(oligo)ethylene glycol-free products exhibit a wide range of physicochemical properties that can compete with or differ from those of traditional ethoxylates. With the aim to enlarge their application fields and to facilitate their penetration in markets dominated by non-renewable surfactants, research work is still necessary to propose 'greener' or 'bluer' molecules utilizing carbohydrate-based raw materials of terrestrial or marine origin, and that display innovative physicochemical behaviours. The purpose of this chapter is to present the most recent findings concerning the chemical syntheses of new carbohydrate-containing surfactants and their potentiality in formulation, with a special focus on the development of ionic versions of sugar surfactants, the use of original carbohydrate source, the incorporation of new linkers between hydrophilic and hydrophobic domains, and the one-pot transformation of natural polysaccharides as eco-friendly synthetic procedures.

1 Introduction

Surfactants can be defined as amphiphilic molecules consisting of hydrophilic (water-soluble) and hydrophobic (water-insoluble) portions that reduce surface tension between two liquids or a liquid and a solid.¹ They are key ingredients in consumer and industrial cleaning compounds such as detergents, cleansing or wetting agents, emulsifiers, foaming or defoaming agents, viscosity builders, degreasers or dispersants. In 2016, the surfactant market size was estimated to be USD 30.64 Billion with a production exceeding 16 million tons worldwide. The majority of surfactants are produced from petrochemical raw materials, which is a non-renewable source. These petroleum-based products are facing a significant increase in the stringent regulations worldwide, due to their poor biodegradability and questionable toxicity profile. Therefore, alternatives made entirely from renewable resources – named Green or Blue surfactants depending on the terrestrial or marine origin of the vegetable raw materials – are being developed for use as substitutes in several applications.² In particular, there has been a push in recent years toward surfactants composed of a sugar for the hydrophilic portion and a

^aEcole Nationale Supérieure de Chimie de Rennes, CNRS, UMR 6226, 11 Allée de Beaulieu, CS 50837, 35708 Rennes Cedex 7, France.

E-mail: Thierry.Benvegnu@ensc-rennes.fr

^bUniversité Bretagne Loire, France

variety of natural fats and oils for the hydrophobic portion.³ Indeed, these sugar-based surfactants display remarkable beneficial physico-chemical properties, reduced sensitivity to many factors, *e.g.*, temperature (in contrast to the non-ionic poly(ethylene oxide) surfactants), surfactant concentration and salinity, and significant advantages regarding biodegradability,⁴ low toxicity, mildness to human skin, and environmental compatibility.⁵ They can be produced from monomeric (such as glucose, sorbitol), dimeric (such as sucrose, lactose) or polymeric (such as starch, cellulose, pectin, chitin, and other natural polysaccharides) sugar raw materials⁶ as well as from agricultural or food processing by-products and waste (carbohydrate-rich waste of peanut oil cake⁷ or orange peelings⁸ for biosurfactant production, pentoses from wheat straw or wheat bran⁹).

Today the most important carbohydrate-based surfactants are alkyl polyglycosides (APG), sorbitan/sucrose esters, and alkyl glucamides (Fig. 1) differing through the nature of the carbohydrate moiety and the linkage between the sugar and the hydrophobic domain (glycosides, esters and amides).¹⁰ Nonionic APG – the largest group of sugar surfactants manufactured worldwide – are produced applying the green chemistry principles¹¹ by the Fisher synthesis between fatty alcohols derived from coconut or palm and glucose from corn starch or other sources of renewable carbohydrates. Due to their good compatibility with the skin and eyes, their strong synergetic effects with anionic surfactants, and their foam-stabilizing properties, APG have found applications in manual dishwashing detergents and cosmetics. Sorbitan esters and sucrose esters are another important class of non-ionic surfactants obtained by either esterification of sorbitol with fatty acids under acidic conditions or

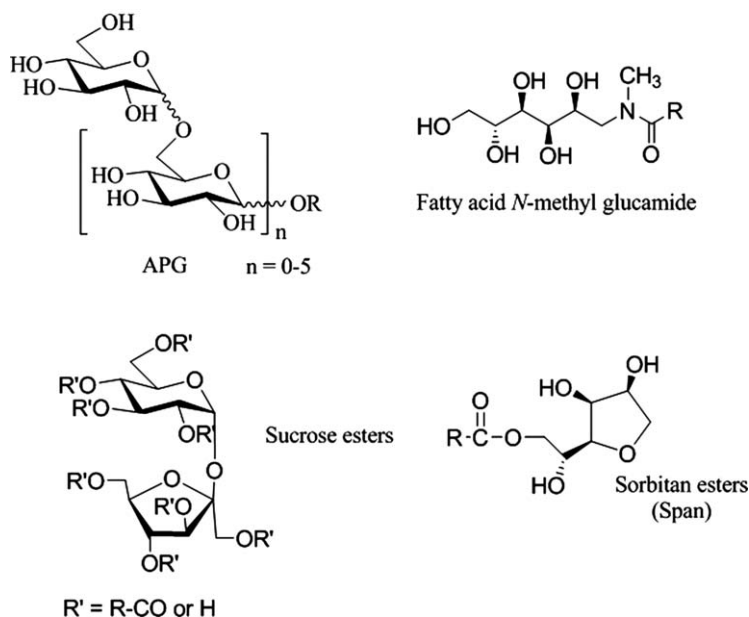


Fig. 1 Structure of traditional sugar-based surfactants.

transesterification of fatty acid methyl esters or triglycerides and sucrose. They are mainly used as emulsifiers and solubilizers in food, cosmetics and pharmaceuticals where dermatological and toxicological criteria are stringent. Fatty acid glucamides are characterized by the presence of a linear head group derived from D-glucitol coupled *via* an amide linkage to an alkyl chain.^{2,5} They are manufactured through the reductive amination of D-glucose with methylamine followed by a base-catalysed acylation reaction with fatty acid methyl ester. Like APG, fatty acid glucamides exhibit high synergism with anionic surfactants with a low irritation potential but they have a strong affinity with calcium ions and are less soluble. An advantageous property of glucamides and APG over sugar esters is their weak sensitivity to hydrolysis under alkaline conditions.

Despite the clear advantages of sugar-based surfactants with regard to biodegradability and ecological behaviour, the growth of these bio-based products still suffers from several limitations compared to petroleum-derived surfactants. First, sugars do not permit the same variety of applications provided by the ethylene oxide source in terms of capacity to modulate the hydrophilicity of non-ionic surfactants to fulfil the intended function. Secondly, their non-ionic character cannot afford to cover the application fields of anionic surfactants (linear alkylbenzene sulfonates (LAS), alpha olefin sulfonates (AOS) and alcohol ether sulfates (AES used in laundry and dishwashing applications) and cationic surfactants (amines, alkylimidazolines, alkoxyated amines, quaternized ammonium compounds (Quats) used as fabric softeners, conditioning agents, disinfectants). Finally, the cost of the carbohydrate raw materials and the relative complexity of the process technology involved in their manufacture, make their price substantially higher than for other surfactants. As a consequence, a need is evident to broaden the spectrum of properties and applications for sugar surfactants so as to facilitate their penetration in markets dominated by non-renewable surfactants. Recent research work has focused on the design, the synthesis and the characterization of novel carbohydrate surfactants with the aim to enlarge their technical viabilities in different formulation applications and to enhance cost-effectiveness using cheaper sugar-based raw materials and/or more direct synthetic routes. In particular, four strategic approaches were investigated to reach these objectives: (1) developing novel ionic versions of sugar surfactants to get a level of performance close to, or higher than, those of anionic and cationic surfactants; (2) producing surfactants based on new carbohydrate motifs derived from terrestrial or marine vegetable source to exhibit original functionality and/or improved surface activities; (3) investigating and rationalizing the effect of original linkers between sugar units and alkyl chains on the surfactant properties and phase behaviour in water; and (4) exploring new environmentally benign synthetic routes based upon the direct transformation of abundant polysaccharides by using one-pot procedures (Fig. 2).

This chapter aims to cover recent developments of sugar based-surfactants (patents excluded), with a special focus on the four items previously mentioned dealing with the introduction of ionic functional

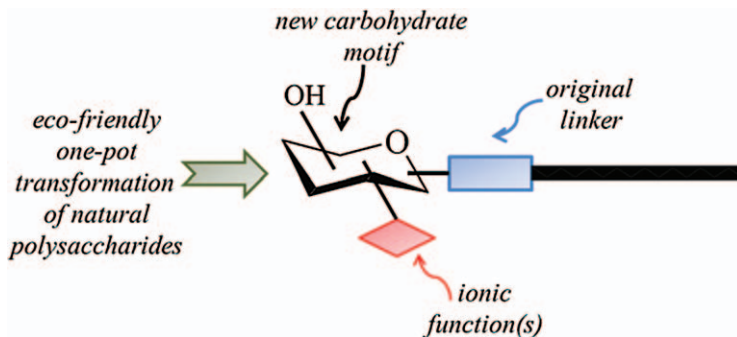


Fig. 2 Schematic representation of innovative sugar-based surfactants reported in recent literature.

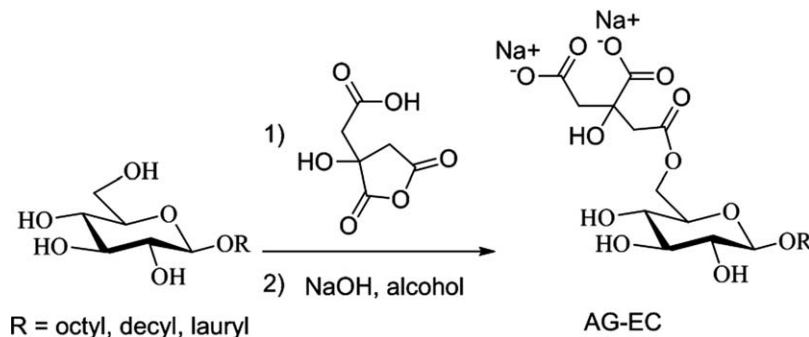
groups, the valorisation of innovative sugar source, the incorporation of new spacers between hydrophilic and hydrophobic domains, and the one-pot conversion of biomass-derived polysaccharides into non-ionic surfactant compositions. The scope of this chapter is limited to production of surfactants using chemical syntheses and will not include enzymatic and microbial processes.

2 Sugar-based ionic surfactants

The derivatization of APG surfactants through the incorporation of ionic groups or the synthesis of additional sugar derivatives containing one or more anionic or cationic moieties has been investigated in the last decades with the goal to modify the properties of nonionic surfactants. In particular, anionic versions of alkyl polyglucosides based on carboxylates, citrates, sulfosuccinates and tartrates have been manufactured for applications in personal care.^{12–14} A variety of other negatively or positively charged compounds were synthesized and evaluated for their potential use in separation, pharmaceutical, and medical sciences. However as outlined in the literature,⁵ only a few products are established in the market and data on the behaviour of these ionic carbohydrate surfactants in aqueous solutions are still insufficient. This section reviews recent bibliography on the syntheses, properties, and potential applications of anionic and cationic surfactants possessing D-glucose or D-mannuronic acid head groups.

2.1 Sugar-based anionic surfactants

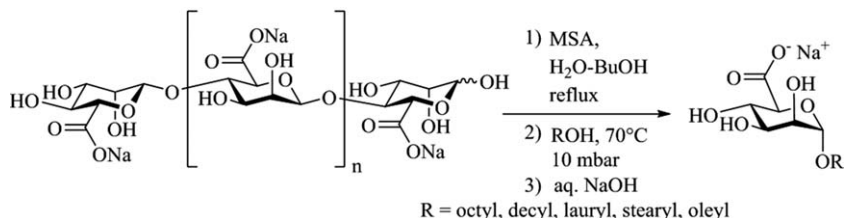
The development of new anionic carbohydrate surfactants was mainly envisaged through the direct introduction of negatively charged citrate, phosphate or sulfonate groups onto alkyl glycosides. Thus, a series of disodium alkyl monoglucoside citric monoesters (AG-EC) were recently synthesized using alkyl (lauryl, decyl or octyl) monoglucosides purified from APG and citric acid anhydride (Scheme 1).¹⁵ These esterification reaction conditions proceeded efficiently (95% yield) to furnish after an additional aqueous sodium hydroxide treatment, the anionic sugar monoester derivatives AG98-EC, AG10-EC and AG12-EC. The surface



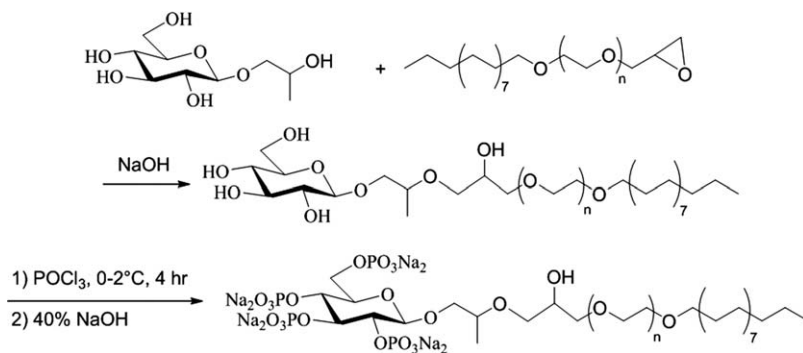
Scheme 1 Synthesis of AG-EC.

properties, foaming ability and wetting ability were investigated. These AG-EC amphiphilic molecules reduced the surface tension of distilled water to a range of $30\text{--}34\text{ mN m}^{-1}$, which indicated that they could adsorb at the air-water interface strongly. The lowest surface tension values were obtained with 0.23 mmol L^{-1} AG12-EC, 1.04 mmol L^{-1} AG10-EC and 5.10 mmol L^{-1} AG08-EC are 31.94 , 33.60 and 29.72 mN m^{-1} that compares favorably to values found with nonionic APG. Due to improved water solubility resulting from the introduction of the COONa head group, monoesters in the decyl and lauryl series exhibited good foaming performance, especially in hard water. The presence of Ca^{2+} in the hard water drove the carboxylated molecules to pack tighter which decreased the cross-sectional area of surfactant monomers and the electrostatic repulsion between charged head groups and improved the strength of the surface film. On the contrary, AG08-EC possessing a shorter hydrocarbon chain showed poor foam ability. Because of their attractive properties in surface activity and non-irritating effects on skin, the decyl and lauryl AG-EC surfactants may be useful in cosmetics and personal care applications.

A greener alternative to the chemical introduction of carboxylate into sugar residues, (precluding the use of toxic chemicals), lies in using natural uronate derivatives possessing carboxylate functionality as starting materials for the preparation of negatively charged surfactants. In particular, homopolymeric oligomannuronates prepared from alginate by acid hydrolysis,¹⁶ were efficiently converted into monomeric anionic mannuronate surfactants through a one-pot process without requiring any isolation and purification step of all intermediates formed during the reactions (Scheme 2): (1) methane sulfonic acid-catalyzed hydrolysis and butanolysis of oligomers; (2) *in situ* transesterification and transglycosidation reactions with fatty alcohols and (3) saponification reaction with aqueous NaOH solution. These mannuronate products exhibit attractive surface activity and foaming properties.¹⁰ Anionic (COONa form) and neutral (COOH form) surfactants with $\text{C}_{12}\text{--C}_{14}$ fatty chains reduce the surface tension to value ($29\text{--}30\text{ mN m}^{-1}$) comparable to those obtained with commercial non-ionic surfactants (Polyethylene glycol, APG). In all cases, carboxylate derivatives exhibited higher CMC values than their neutral acidic counterparts ($0.28\text{--}1.62\text{ mmol L}^{-1}$ for anionic compounds and $0.13\text{--}0.16\text{ mmol L}^{-1}$ for neutral counterparts). Mannuronic acid



Scheme 2 Synthesis of mannuronate surfactants from oligomannuronates (obtained by depolymerization of alginates).



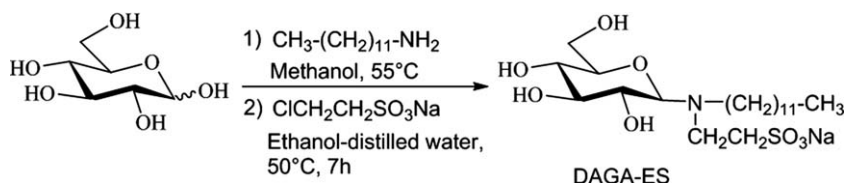
Scheme 3 Synthesis of sugar-based anionic surfactants with phosphate groups.

derivatives are high-foaming surfactants which can be compared with the ether sulfate derivative, the Sodium Dodecyl Sulfate (SDS). Besides, the mannuronate compounds are characterized by a high biodegradability profile and the absence of aquatic toxicity according to the European regulations. These original 'Blue/Green' surfactants derived from both marine and terrestrial raw materials are being manufactured by SurfactGreen, a start-up specialized in the production and formulation of 100% bio-based surfactants.

Another recent family of anionic glucosidic surfactants results from reactions of propylene glycol dextrin with oxiran-containing polyoxyethylene stearyl ethers and phosphorus oxychloride. These polyanionic compounds contain four sodium phosphate groups on the glucose moiety and a polyoxyethylene chain connected to a saturated C₁₈ alkyl chain (Scheme 3).¹⁷ By changing the length of the polyoxyethylene chain, it was possible to modulate the properties of these surfactants. Short oxyethylene chains provided sugar based anionic products with low areas occupied by the surfactant molecules at the air-water interface, and great effectiveness at minimizing surface-tension, due to a lower hydrophilic character. Increasing the length of the oxyethylene chain resulted in increased foam production and foam stability. This behaviour was due to the ability of the micelles to break up to give a larger amount of free sugar based anionic surfactants in solution. Concomitant higher interactions between the oxyethylene chains of the surfactants occurred and resulted in more stable foam. Excellent emulsifying properties were also observed arising from the combination of the adsorption ability of the dextrin

(which, due to the hydrophobicity of the stearyl moiety was able to rapidly adsorb and rearrange at the Oil/Water (O/W) interface, resulting in the formation of a coherent protective molecular film) and the high hydrophilicity of the polar domain (dextrose unit, polyoxyethylene units, and phosphate groups) that formed a strongly solvated layer near the O/W interface, conferring steric stabilization to the emulsion oil droplets. Such behaviours suggest the applicability of these novel anionic sugar surfactants as emulsifiers, especially in paints and water phase corrosion inhibitors in oil fields.¹⁸

The access to anionic sugar surfactants was alternatively envisaged through the introduction of the negatively charged function into the linker between the carbohydrate head group and the alkyl chain. Thus, an eco-friendly sugar-based anionic-nonionic surfactant (DAGA-ES) was prepared through a two step-process involving the synthesis of lauryl glucosylamine and its subsequent reaction with 2-chloride ethyl sulfonic acid sodium, employing ethanol, methanol and water as solvents (Scheme 4).¹⁹ This surface active molecule was designed for its potential application in the enhanced oil recovery (EOR) industry in order to displace oil by reducing oil/water interfacial tension, forming an emulsion, and achieving wetting inversion on the surface of rock.²⁰ The presence of the anionic sulfonate group is supposed to possess thermal stability whereas the carbohydrate residue is expected to display salt tolerance, which constitutes two requirements for application in EOR processes. Surface activity measurements of the solution containing DAGA-ES in the absence or presence of electrolyte gave CMC and γ_{CMC} values in 3 mol L⁻¹ NaCl solution of 1 mmol L⁻¹ and 22.3 mN m⁻¹, which were lower than that in distilled water. The excellent water solubility was confirmed by the value of the hydrophilic-lipophilic balance, HLB (11.72). Other fundamental surfactant parameters such as maximum surface excess (Γ_{max}) (also named adsorption quantity) and minimum surface area (A_{min}) were determined and showed that more surfactant molecules adsorbed on the air/water interface and arranged closely when electrolyte was added. In addition, the oil-water interface tension was reduced significantly when DAGA-ES interacted with NaCl to exhibit the minimum value of 0.05 mN m⁻¹. Furthermore, with the surfactant concentration increasing, the hydrophilicity for both the hydrophile surface and lipophilic surface (glass sheets were soaked in distilled water and silicone oil, respectively, for about one week) was gradually increased. Wettability reversal was achieved just on the lipophilic surface. Finally, the oil displacement experiment showed that the oil recovery increased by 16.29% using



Scheme 4 Synthesis of sugar-based anionic-nonionic surfactant DAGA-ES.

DAGE-ES *versus* 12.82% using AES. These results indicated that this original surfactant displayed more attractive properties in EOR process than the conventional anionic-nonionic surfactant AES.

2.2 Sugar-based cationic surfactants

Cationic surfactants are amphiphilic molecules that possess a positively charged head group such as amines, alkylimidazolines, alkoxyated amines protonated in acidic media, and quaternized ammonium compounds (called Quats).²¹ They are good emulsifying agents, and do not form insoluble scums with positively-charged hard-water ions. These surfactants have also been found to be good bactericides and some find use as topical antiseptics. Their germicidal properties make them especially useful in bathroom and hand sanitizers. Furthermore, cationic surfactants are attracted to negatively-charged sites that occur naturally on most fabrics. They can bind to these sites and provide the fabric with a soft, luxurious feel. For this reason, they are often employed as fabric softeners. Other industrial applicability of traditional cationic surfactants deals with uses as antistatic agents, conditioning agents in cosmetics, asphalt additives, corrosion inhibitors and organophilic clays. Nevertheless, most of these products have been found to have an acute toxicity and they are generally poorly biodegradable. Besides, there are some other problems for these cationic surfactants which limit their applications, including skin irritation and low compatibility with anionic surfactants forming insoluble aggregates in aqueous media. One possible strategy to propose more efficient, greener and safer molecules is to insert sugar moieties into the hydrophilic cationic head group. In this context, cationic versions of glucose and other sugar-based surfactants were investigated and some of them are commercially available as for cationic alkyl polyglucosides.²² These products are expected to provide substantivity to skin and hair, much more than their nonionic APG precursors. The sugar moiety decreases the irritation substantially over traditional Quats. However, the chemical syntheses to produce them are not generally eco-friendly (use of hazardous reagents and solvents, production of pollutants) which minimizes the benefit of these sugar-based cationic surfactants for sustainable development. Besides, cationic units introduced onto the carbohydrate head groups are not derived from vegetable resources but are of petrochemical origin.

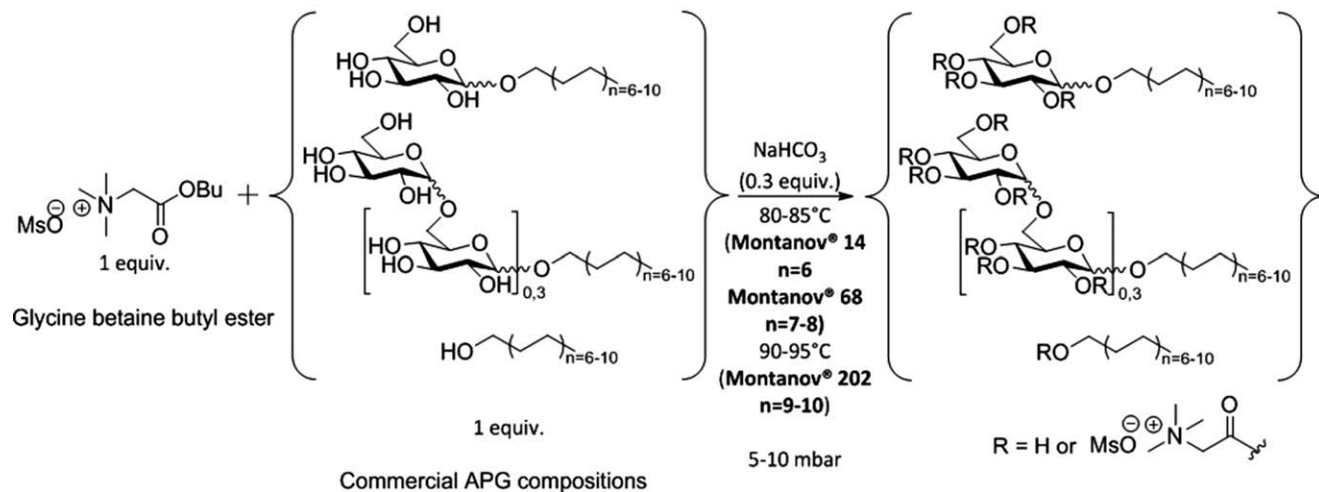
Very recently,²³ a versatile and efficient method for the cationization of fatty alcohols and nonionic sugar-based surfactants was developed through NaHCO_3 -catalyzed transesterification reactions using a 'green' glycine betaine butyl ester reagent to provide 100% bio-based cationic sugar surfactants. Glycine betaine (GB) is an abundant natural molecule composed of quaternary trimethylalkylammonium moiety and carboxylate functionality, and it is present in many fruits, cereals and beet. It was easily transformed into the cationizing agent through an esterification reaction with butanol in the presence of methane sulfonic acid (MSA) as a biodegradable catalyst. The environmentally benign solvent free cationization reaction was successfully applied to mixtures of fatty

alcohols and glucose or xylose-based surfactant, leading to high levels of conversion of these nonionic substrates into cationized forms. In particular, APG corresponding to commercial surfactant compositions – combinations of fatty alcohols and alkyl glucosides as a mixture of α/β isomeric compounds with one to four glucose units – sold under the name Montanov[®] 14, 68 and 202 by the company Seppic, were used as starting sugar surfactants. It is noteworthy that the grafting of the glycine betaine residue into alkyl glycosides proceeded with high efficiencies even in the presence of a large excess of fatty alcohols. The cationization reaction was found to operate preferentially at both positions 3 and 9 of glucose units. These novel cationic GB-containing APG versions were produced at a kg scale and exhibited promising emulsifying properties (Scheme 5).

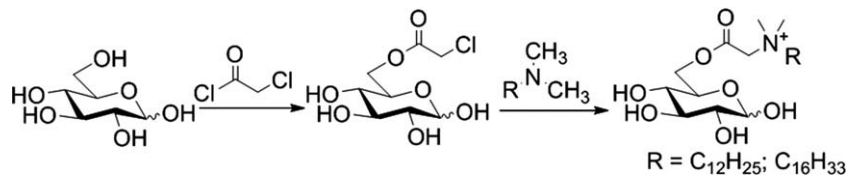
A couple of new glucose-based cationic surfactants were also synthesized *via* a two step-process involving an esterification at position 6 of the glucose unit with 2-chloroacetyl chloride followed a quaternization reaction of a tertiary amine bearing a dodecyl or hexadecyl chain in addition to two methyl groups (Scheme 6).²⁴ These products exhibited interesting surface active properties with CMC and γ_{CMC} values of 0.955 and 0.794 mmol L⁻¹ and 24.5 and 22.9 mNm⁻¹ for the C₁₂ and C₁₆ derivatives, respectively. The foamability of these synthetic surfactants is quite low, especially for the hexadecyl compound. Antimicrobial activity was investigated towards three bacterial strains (Gram-positive and Gram-negative) in comparison to that of cetyl trimethylammonium bromide (CTAB). The results showed that both of the cationic sugar surfactants displayed greater bactericidal activity to that of the control on *E. coli* and *S. aureus*. However, they were less active on *B. subtilis* than CTAB. Interestingly, these sugar-based ester quaternary ammonium salts displayed an excellent compatibility with anionic surfactant SDS over a large range of concentration and a synergistic effect was evidenced. It was assumed that the introduction of the sugar head group to the traditional quaternary ammonium compound increased the steric hindrance of the cationic ion and weakened the binding tightness of the cationic and anionic polar heads.

3 Surfactants based on novel carbohydrate motifs

The need still exists to identify new sources of saccharide materials (especially industrial by-products and waste, marine raw materials which are non- or poorly exploited in food industry, new chemical structures) for designing and characterizing sugar surfactants based on new motifs and with economic and technical viabilities in different applications. Recent research results reported in this section, demonstrated that uronic acids, isosorbide, and trehalose could represent attractive starting materials for the synthesis of sugar surfactants having a range of beneficial physical and performance properties, including high levels of surface and interfacial activity, effective emulsification properties, solubilizing and stabilizing abilities.



Scheme 5 Conditions used for the cationization reaction of commercial APG compositions (Montanov® series).

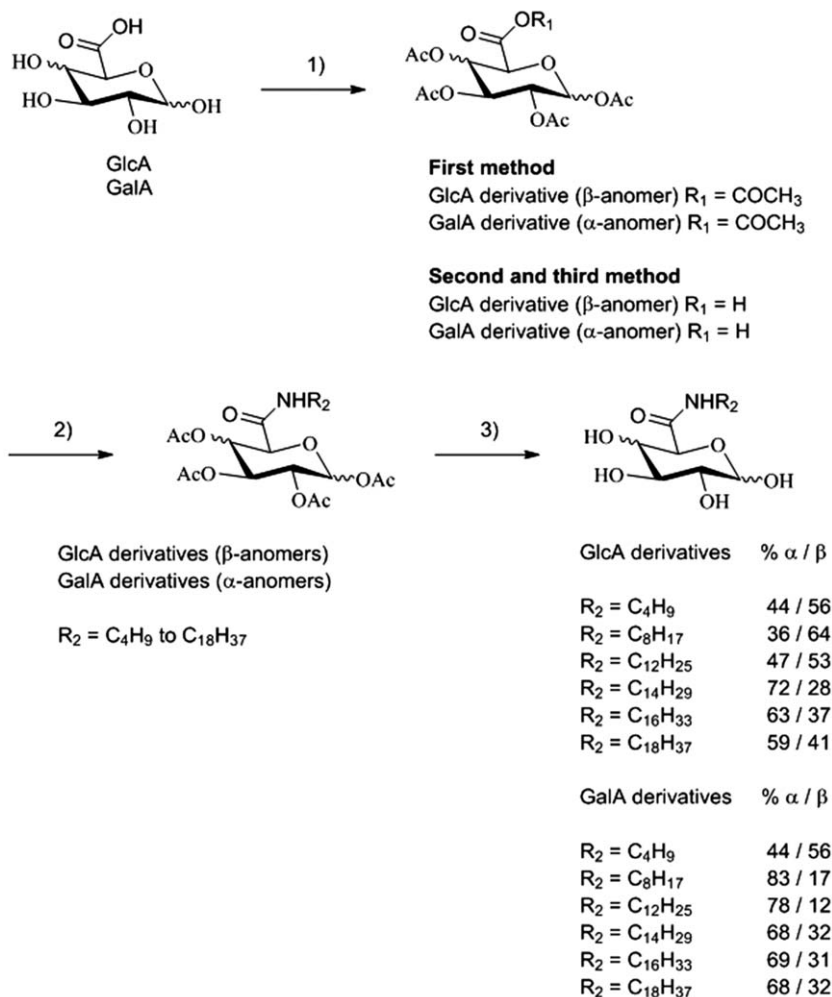


Scheme 6 Synthetic route of glucose-based cationic surfactants.

3.1 Uronic amide derivatives

Uronic acids are monosaccharides possessing a carboxyl group instead of a primary alcohol unit.²⁵ They are constituents of many important biopolymers of plant and animal origin. D-glucuronic acid is a component of hemicelluloses, gums, hyaluronic acid, and heparin; D-galacturonic acid, which is the monomeric unit of pectic substances, is a constituent of certain bacterial polysaccharides. In animals, D-glucuronic acid (found in the blood and urine) removes toxic substances by forming glycosides. These sugars contain appropriate functional sites (hydroxyl and carboxylic groups) for the introduction of lipophilic chains which can promote the development of new added value bio-based surfactants. Through several research works during last decades, monocatenary and bicatenary alkyl uronates possessing either a free anomeric position or an alkyl chain at C-1 position were evaluated as nonionic surfactants. In particular, the effect of structural modifications (stereochemistry, alkyl chain length and number) on surface activities and micelle formation was reported.^{26,27}

As an amide function linking the hydrophilic sugar head group to the lipophilic tail is more resistant towards hydrolysis compared to ester function in both neutral and alkaline media, uronic amide surfactants were recently produced from uronic acids derived from widely available raw materials. Three short synthetic routes toward long-chain *N*-alkylamides of glucuronic and galacturonic acids were proposed *via* the increase of carboxyl group reactivity using an anhydride as synthetic intermediate, adding carbodiimide in the medium, or employing an acid chloride as activated compound (Scheme 7).²⁸ Next nucleophilic attack of primary linear alkylamines followed by removal of acetate protecting groups under basic conditions by reaction with sodium carbonate in methanol afforded a mixture of amide-type α and β glycopyranosides. The third synthetic pathway based on the formation of an acyl chloride led to the best overall yields (46–58%). Surface-active properties (CMC, γ_{CMC} , Γ_{max} , A_{min}) of homologous series of glucuronic and galacturonic acids *N*-alkylamides from C₈ to C₁₈ were also assessed. In general, the two series of uronic amide-based surfactants exhibited γ_{CMC} and CMC values in the same order of magnitude, even if a slight trend of higher surface activities was observed for galacturonic *N*-alkylamides in comparison with the corresponding glucuronic *N*-alkylamides (for example, γ_{CMC} values of 24.4 mN m⁻¹ and 20.5 mN m⁻¹ and CMC values of 3.3 ± 1.1 mmol L⁻¹ and 4.0 ± 1.0 mmol L⁻¹ for C₈Glc-N and C₈Gal-N products, respectively). Furthermore the maximum surface excess and the minimum molecular areas were also on the same range for both series of surfactants. More interestingly, these compounds were found to be more efficient surface-active agents compared to the corresponding alkyl ester analogues having the same alkyl chain length or to standard nonionic surfactants such as alkylphenol ethylene oxide condensate or alkyl(poly)glycosides. Additionally, the results showed that an amide-type linkage between the head group and the hydrocarbon tail instead of an ester bond led to superior surface-active properties. Indeed, they are supposed to form a more condensed film at the air-water interface

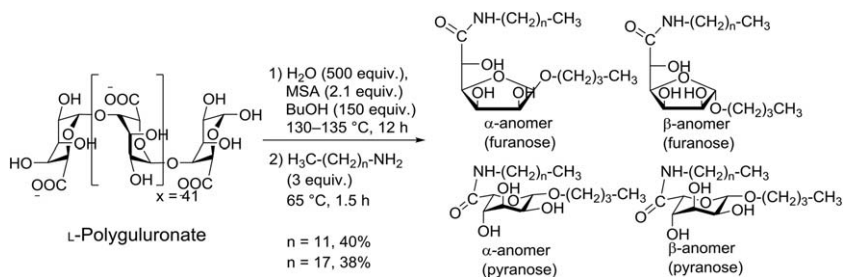


Scheme 7 Synthesis of various *N*-alkylamides from D-glucuronic and D-galacturonic acids. Reagents: a First method: (1) Ac_2O , I_2 , 0°C ; (2) RNH_2 , dry CH_2Cl_2 ; (3) Na_2CO_3 , MeOH; b Second method: (1) Ac_2O , H_2SO_4 , 60°C then H_2O ; (2) DCC, dry CH_3CN then RNH_2 ; (3) Na_2CO_3 , MeOH; c Third method: (1) Ac_2O , H_2SO_4 , 60°C then H_2O ; (2) $(\text{COCl})_2$, DMF, dry CH_2Cl_2 then RNH_2 ; (3) Na_2CO_3 , MeOH.

possibly because of a smaller bulk size of the amide bond compared to that of the ester one. As pointed out by the authors, due to their total biodegradability, lack of taste, non-skin irritation, their being odorless and non-toxic, these uronic amide derivatives could find several applications in the cosmetic, food or pharmaceutical sectors.

Despite the clear interest of most amide-type surfactants reported in the literature over the past ten years, alternative solutions were still required both to develop 'greener' synthetic routes for these nonionic sugar-based products and to propose novel saccharide materials extracted from original natural sources. In this context, in order to complement the access to mannuronamide surfactants produced from

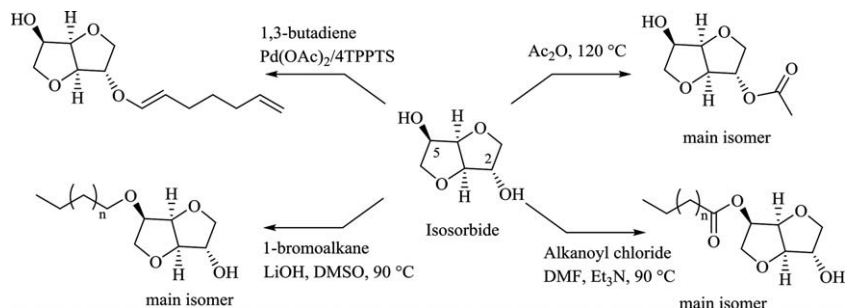
D-mannuronic acid oligomers,²⁹ innovative alkyl-L-gulonamide-based surfactant compositions were recently developed from oligogulonates obtained as for oligomannuronates, by an acidic depolymerization of alginates.³⁰ The approach followed in this work aimed at using eco-friendly conditions which satisfied the principles of blue chemistry (corresponding to the concept of green chemistry), carrying out all the reactions involved in the synthesis in a one-pot process, and providing surfactant compositions with emulsifying properties superior to the conventional alkyl(poly)glycosides. Direct acid hydrolysis of the oligosaccharides coupled with Fischer glycosylation and esterification reactions of L-guluronic acid with butanol proceeded efficiently in the presence of readily biodegradable methane sulfonic acid (MSA) to furnish the corresponding *n*-butyl uronate intermediates. Subsequent transformation of these mono-gulonate derivatives into long-chain alkyl L-gulonamide surfactant mixtures were carried out without any isolation and purification step of the intermediates through a one-pot solvent-free aminolysis reaction with C₁₂ or C₁₈ fatty amines (Scheme 8). The surfactant compositions are α,β anomeric mixtures of furanose and pyranose forms bearing a short butyl chain at the anomeric position and a longer alkyl (C₁₂ and C₁₈) chain amide-linked to the sugar head. The β -L-gulofuranosiduronamide was found to be the major component in the isomeric mixture whereas the β -L-gulopyranosiduronamide was present in the lowest quantity. The emulsifying properties of these L-gulonamide-based surfactant compositions both in water-in-oil (W/O) and oil-in-water (O/W) systems were investigated in comparison with those of a commercial alkylpolyglycoside emulsifier, named Montanov 82[®] manufactured by the company Seppic. High stabilities of O/W emulsions were engendered by the L-gulonamide derivatives at a concentration of 0.5% in weight, since the emulsions remained unchanged after at least one month at 20 °C. On the contrary a faster demixing of the emulsions was observed when the commercial emulsifier Montanov 82[®] (<24 h) was used. Furthermore, the uronamide surfactant composition displayed excellent W/O emulsion stabilizing abilities unlike Montanov 82[®] which led to more modest efficiencies as emulsion stabilizers. Additionally, attractive antibacterial effects of these sugar-based surfactants against Gram positive bacteria were observed.



Scheme 8 Conditions used for the transformation of L-polyguluronate into L-gulonamide surfactants.

3.2 Isosorbide derivatives

Another bio-based material produced through a double dehydration of sorbitol (an important product of the starch industry), is isosorbide (1,4:3,6-dianhydro-D-sorbitol).³¹ It is a V-shaped bicyclic intramolecular diether with two secondary hydroxyl groups in the 2- and 5-positions (Scheme 9). It has been reported that 5-hydroxyl group in *endo* configuration forms intramolecular hydrogen bonds, leading to higher reactivity because of the more nucleophilic character of the oxygen at the 5-position. Using different reaction conditions, the less sterically hindered OH-2 in *exo* configuration may be more reactive; clearly the two secondary hydroxyl groups are not of equivalent reactivity. Isosorbide is being developed industrially as a platform chemical – it is now available at the scale of several thousand tons a year – for applications as monomers and building blocks for new polymers and functional materials, new organic solvents, for medical and pharmaceutical applications, and even as fuels or fuel additives. In addition, this non-toxic diol produced from bio-based feedstocks was used as a synthon for the preparation of surfactants, with the aim to propose greener alternatives to the ethoxylates. In particular, isosorbide-based hydrotropes, also known as ‘solvo-surfactants’, capable of enhancing the aqueous solubility of hydrophobic compounds were investigated through the synthesis of ether and ester derivatives. In industrial applications, hydrotropes are of great interest, in particular for the formulation of cleaning products, in drug solubilization, and in the paint and coating industries. Monoethers of isosorbide possessing a short alkyl chain (butyl, pentyl, hexyl and octadienyl) were produced under classical Williamson conditions³² or by the Pd-catalyzed telomerisation of butadiene³³ (Scheme 9) leading to either the 5-*O*-alkyl monoether (Williamson reaction) or the 2-*O*-octadienyl monoether (telomerisation reaction in water) as the main products. Isosorbide monoesters (acetate to octanoate) were obtained by esterification with acetic anhydride in the absence of base and solvent or with acyl chlorides in dimethylformamide, in the presence of trimethylamine.³⁴ Isosorbide monoacetates were isolated in 40% yield and with selectivity in favor of the 2-*O*-isomer. On the contrary for the longer monoesters, the 5-*O*-acyl isomer was formed in 45% yield, with a

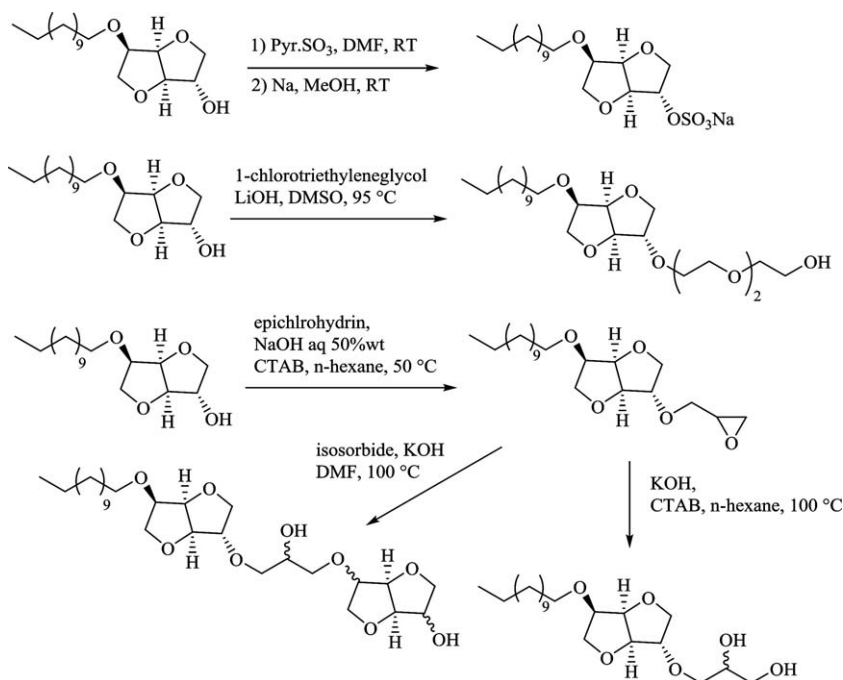


Scheme 9 Conditions used for the transformation of isosorbide into monoether and monoester hydrotropes.

moderate selectivity due to the higher acidity of the 5-OH proton which exhibits a hydrogen bond with the oxygen atom of the adjacent cycle.

Water miscibility studies showed that the shorter monoester compounds (acetates, propanoates and butanoates) were completely miscible with water whereas the pentanoate to octanoate derivatives exhibited an extended miscibility gap. The ether analogues in *endo* configuration are miscible with water in the whole concentration and temperature ranges up to a butyl chain, while the pentyl derivative has a cloud point around 40 °C. These results clearly indicate that in the case of these short-chain amphiphiles derived from isosorbide, ethers are more hydrophilic than esters. Whatever the bond (ether or ester), the compounds in *exo* configuration are less hydrophilic than the *endo*-isomers. In terms of hydrotropic properties, the longest water-soluble compounds (butanoate esters in *endo* or *exo* configuration and pentyl monoether *endo*-isomer) exhibited the best solubilizing capacities and they compare favorably to the reference hydrotrope ethylene glycol butyl ether C₄E₁. Furthermore it was found that for the ester derivatives, the *exo* series was more efficient than the *endo* series.

Derivatizations of isosorbide-based monoethers were also envisaged to provide new anionic or non-ionic surfactants. Due to the weak hydrophilicity of isosorbide residue, the introduction of negatively charged sulfate group or non-ionic glycerol, triethylene glycol and glyceryl-isosorbide moieties was performed into this bicyclic structure so as to balance the lipophilic dodecyl alkyl chain grafted at either 5- or 2-position (Scheme 10). Thus, 2-*O*-dodecyl isosorbide sulfate and 5-*O*-dodecyl isosorbide sulfate³⁵ were prepared as bio-sourced alternatives to the lauryl ether sulfate, LES that is based on petroleum-derived ethylene oxides. The sulfation of 2-*O*- and 5-*O*-monododecyl isosorbide was carried out with pyridine-SO₃ complex in dimethylformamide at room temperature. The pyridinium salt of the 5-*O*-monododecyl isosorbide sulfate was isolated as a pure form after flash chromatography and the pyridinium counterion was exchanged with sodium through the treatment with a methanolic solution in which metallic sodium was previously introduced. As for non-ionic isosorbide derivatives, the grafting of the triethylene glycol hydrophilic unit was achieved using standard Williamsom conditions whereas glycerol and glycerylisosorbide moieties were introduced through the synthesis of a glycidyl intermediate followed by its ring opening by hydroxide ions or isosorbide (Scheme 10).³⁶ With regard to physicochemical properties, the two 5-*O* and 2-*O*-dodecyl isosorbide sulfate isomers showed significantly different aqueous properties (Krafft temperature, critical micellar concentration), probably due to the fairly different conformations adopted and the rigidity of isosorbide. The 5-*O*-isomer is characterized by the same Krafft temperature as SDS but its CMC is four times lower, which makes it a more efficient surfactant. Due to a more linear conformation, the 2-*O*-isomer packs more efficiently in the solid state, and consequently has a Krafft temperature higher than the 5-*O*-isomer. The foaming properties of 5-*O*-dodecyl isosorbide sulfate were found to be comparable to the ones of SDS with a three times lower concentration, with slightly lower foam stability.



Scheme 10 Synthetic pathways to the sulfated and non-ionic triethyleneglycol-, glycerol- and glycerolisorbide-containing dodecylisorbide surfactants.

Physicochemical properties of isorbide-based non-ionic surfactants were further evaluated in comparison to the ones of polyethylene glycol dodecyl ether. The surfactant functionalized by a triethylene glycol displayed very similar aqueous behaviour to the one of pentaethylene glycol dodecyl ether ($C_{12}E_5$), which indicates that the isorbide fragment is rather equivalent to a diethylene glycol (E_2) when inserted between the apolar and polar fragments of the surfactant. Moreover, it does not much affect the packing parameters since the same succession of liquid crystal phases (Hexagonal H_1 – Bicontinuous cubic V_1 – Lamellar L_α) was obtained in water when increasing concentration. Only the concentration and temperature limits are slightly shifted. With surfactants containing a glycerol or a glycerolisorbide residue, very poorly water-solubility was observed, and only cubic and lamellar liquid crystal phases in water are formed resulting from a higher packing parameter.

Very recently,³⁷ an emulsifier based on diglycidyl ether of isorbide (Fig. 3) was synthesized *via* a condensation of isosorbide and epichlorohydrin yielding a mass prepolymer and subsequent condensation to dodecylamine. This new bio-sourced surfactant was characterized by a molar mass of 1955 g mol^{-1} and a thermal stability up to 300°C that makes it suitable for applications at high temperatures. Moreover, it exhibited a very low CMC value of $1.7 \times 10^{-6} \text{ mol L}^{-1}$ compared to other products and a strong O/W emulsion stabilizing effect was observed at lower concentrations (0.1%) than usual emulsifiers. The isoelectric point was at the alkaline pH 12.3, so that emulsion droplets were positively

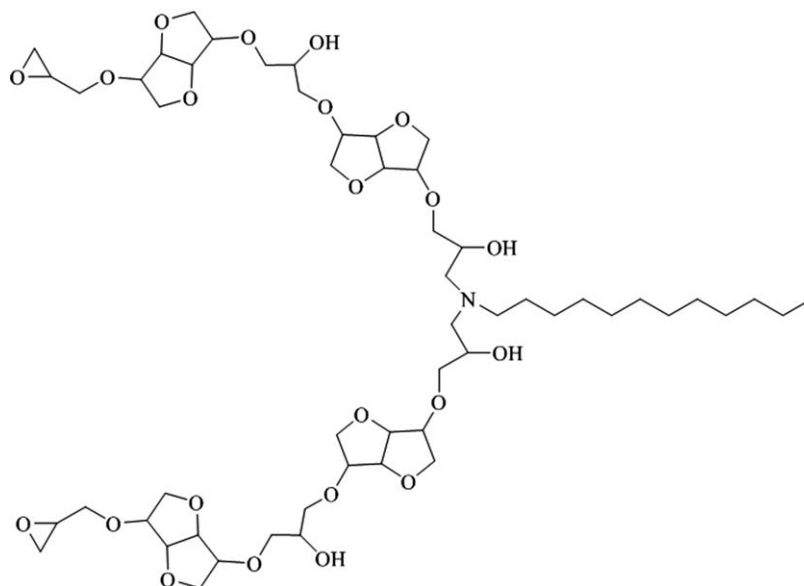


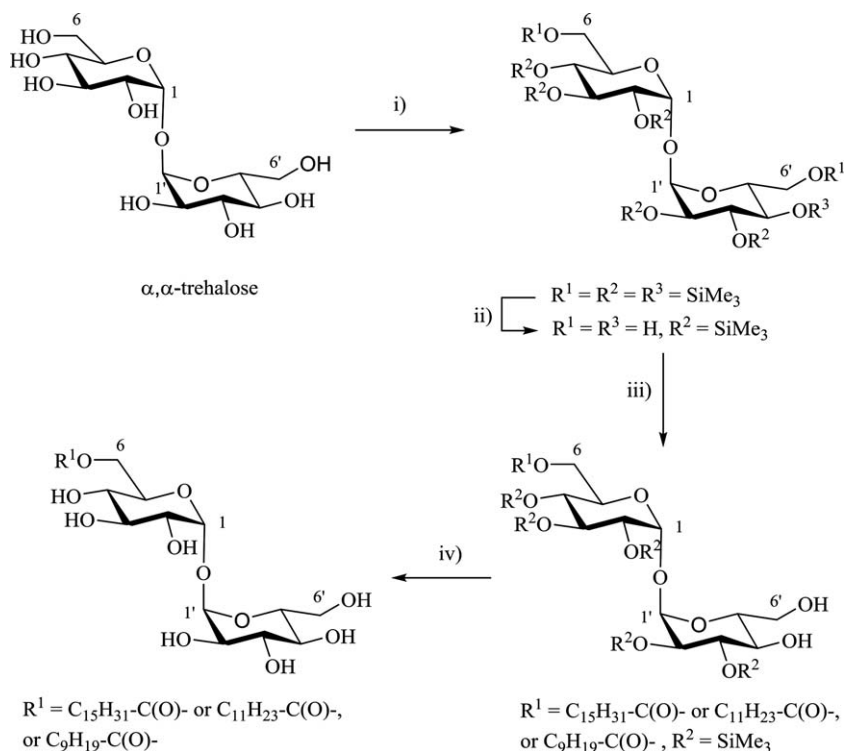
Fig. 3 Structure of an emulsifier based on oligomers of isosorbide.

charged in the pH range of practical applications. Thereby, these cationic surfactants could be used in cosmetic and textile applications as fabric softeners or hair conditioners.

3.3 Trehalose fatty acid esters

Trehalose is a disaccharide composed of two glucose molecules linked by an α - α -1,1'-O-glycosidic bond.³⁸ Since the reducing end of a glucosyl residue is connected with the other, it does not possess reducing properties. Trehalose occurs widely in nature (in mushrooms, yeasts, fungi, and insects) and it is well-known to act as a protectant against various environmental stresses, like desiccation, heat, freezing, or osmotic shock. Due to its clam shell-like shape and its unique hydration structure, this natural carbohydrate is able to interact with both hydrophilic and hydrophobic molecules. Trehalose structure ensures its stability under low pH values even at elevated temperatures and unlike other disaccharides, trehalose will not readily hydrolyze. Because of nonreducing sugar, this saccharide does not show Maillard reaction with amino compounds such as amino acids or proteins. Until 1990s, trehalose has been produced in relatively small amounts and with a high cost, by extraction from yeast and vegetable sources. Since then, its price has dramatically decreased (from 700 \$ per Kg in 1990 to 3 \$ per Kg in 2015) after the development of its enzymatic manufacturing from poly- and oligosaccharides making possible its use in pharmaceutical (excipient), medicine (agent for preservation of organs and tissues for transplantation, cryopreservation of stem cells), cosmetic (moisturizing ingredient) and food (sweetener and bulking agent) sectors.³⁷

In view of the unique properties of this atypical disaccharide in terms of chemical and thermal stability and hydration structure, Schiefelbusch *et al.*³⁹ recently proposed its chemical transformation into non-ionic surfactants as bio-sourced alternatives of polysorbates used in pharmaceutical protein formulations. Indeed, these polyethylene glycol-derived surfactants are efficient protein stabilizers against various kinds of surface-correlated stress, *e.g.* adsorption at container surfaces, denaturation at the air-water interface, or the ice-water interface. Nevertheless, they cannot ensure long-term stability of different proteins due to polyethylene glycol-related induction of oxidation processes under static storage conditions. In this context, the preparation of well-defined PEG-free chemical structures instead of a heterogeneous mixture of ethoxylated sorbitan derivatives esterified with lauric acid (Polysorbate 20) or oleic acid (Polysorbate 80), could provide higher performing protein stabilizers for application in biopharmaceutical formulations. Thus, three different sugar-based surfactants, 6-*O*-monocaprinoyl- α,α -trehalose, 6-*O*-monolauroyl- α,α -trehalose and 6-*O*-monopalmitoyl- α,α -trehalose were synthesized in a chemical process based on four successive steps (Scheme 11): (i) complete silylation using trimethylsilyl chloride and hexamethyldisilazane in anhydrous pyridine; (ii) selective desilylation of both primary hydroxyl using a methanolic potassium



Scheme 11 Reaction scheme of the preparation of 6-*O*-monocaprinoyl- α,α -trehalose, 6-*O*-monolauroyl- α,α -trehalose and 6-*O*-monopalmitoyl- α,α -trehalose: (i) TMSCl, HMDS, anh. Pyridine; (ii) K_2CO_3 , MeOH, 0 °C; (iii) 1,3-DCC, 4-DMAP, fatty acid, dry CH_2Cl_2 ; (iv) TFA, THF, H_2O .

carbonate solution; (iii) predominant monoacylation through the use of fatty acids in the presence of 1,3-DCC (dicyclohexylcarbodiimide) and a stoichiometric amount of 4-DMAP (dimethylaminopyridine) and (iv) final deprotection of the silylated hydroxyl groups with trifluoroacetic acid in aqueous tetrahydrofuran (Scheme 11).

The long-chain 6-*O*-monopalmitoyl- α,α -trehalose was insoluble in water at room temperature and therefore appeared to be less suitable for protein formulation. On the contrary, the two other surfactants with shorter C_{10} and C_{12} alkyl chains exhibited CMC and surface tension values (C_{10} compound: $CMC = 1.92 \text{ mg mL}^{-1}$, $\gamma_{CMC} = 29 \text{ mN m}^{-1}$; C_{12} compound: $CMC = 0.33 \text{ mg mL}^{-1}$, $\gamma_{CMC} = 39 \text{ mN m}^{-1}$) that were comparable to those of Polysorbate 20 ($CMC = 0.15 \text{ mg mL}^{-1}$, $\gamma_{CMC} = 38 \text{ mN m}^{-1}$) and Polysorbate 80 ($CMC = 0.014 \text{ mg mL}^{-1}$, $\gamma_{CMC} = 45 \text{ mN m}^{-1}$). Both these trehalose-based molecules showed slightly higher hemolytic activity compared to polysorbates. In addition, they formed transparent and colorless aqueous solutions and did not display phase transitions or separation at pharmaceutically applied temperatures. Finally, similar to polysorbates, they were able to stabilize human growth hormone, hGH, against aggregation upon shaking as well as adsorption. Consequently, these new sugar-based surfactants offer a promising alternative and have potential for application in protein formulations.

4 Sugar-based surfactants including original linkers between polar and lipophilic domains

The idea of amphiphilic linkers was recently introduced in order to increase the surfactant-oil interaction. Surfactants can be found in various forms (monocatenar, bicatenar, bolaamphiphile, gemini, *etc.*). Commonly, they are synthesized by linking hydrophilic (sugar unit) and lipophilic (hydrocarbon chain) moieties *via* a glycosidic linkage but it may be of a different nature like esters, amides and amines. Each of these linkers presents their own advantages and disadvantages; glycosides and esters hydrolyze at low pH whereas amines and amides are stable under these conditions. However, amides have constraining limitations because of their strong hydrogen bonding, which potentially leads to high Krafft temperatures. Here are some examples of these chemical products (Fig. 4).

In this section, we will discuss about novel and original linkers and focus particularly on the synthesis of these chemical structures and their influence on the interfacial properties. In this part, we have chosen to differentiate the molecules according to their chemical functions.

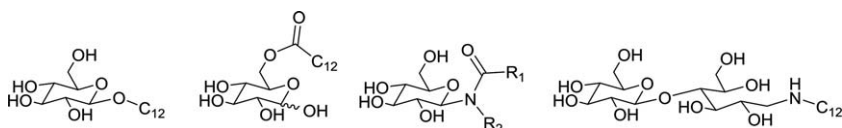


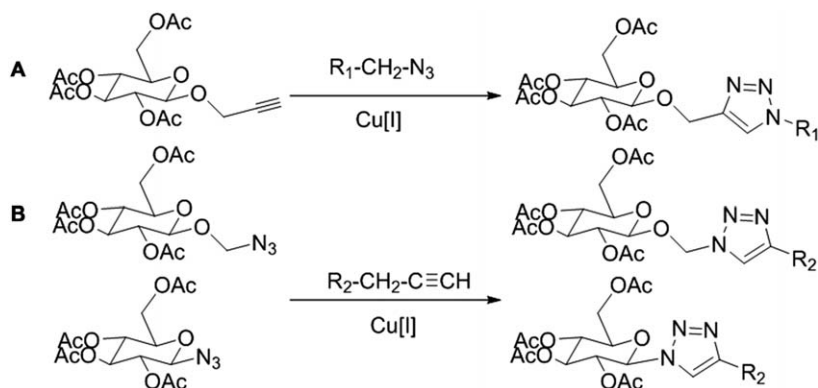
Fig. 4 Examples of sugar surfactants possessing glycosidic, ester, amide or amine linkers.

4.1 Synthesis and characterization

4.1.1 Triazole-based linkers. The triazole-based linkers are by far the best documented. The general synthetic scheme is a multi-step process starting from activation and functionalization to coupling reaction with fatty derivatives by using click chemistry methodologies. In the 1,3-dipolar [3 + 2] cycloaddition, also known as Huisgen cycloaddition, a 1,3-dipole – such as azides, nitriloxides or diazoalkanes – reacts with a dipolarophile – like alkenes, alkynes or carbonyl – *via* a concerted mechanism to form a five-membered heterocycle. This reaction gives poor regioselectivity which leads to a mixture of 1,4 and 1,5-substituted triazoles. A modification of this reaction was initially described by Sharpless *et al.*⁴⁰ and Meldal *et al.*⁴¹ to specifically provide a 1,4-disubstituted-1,2,3-triazole regioisomer using Cu(I), better termed the Copper(I)-catalyzed Azide–Alkyne Cycloaddition (CuAAC). The active Cu(I) catalyst can be generated *in situ* from Cu(I) salts (iodide, bromide, chloride), coordination complexes such as [Cu(CH₃CN)₄]PF₆ and [Cu(CH₃CN)₄]OTf, or Cu(II) salts (sulfate, acetate) using a reducing agent like sodium ascorbate. This latter quickly became the method of choice and water appeared to be an ideal solvent capable of supporting Cu(I) acetylides in their reactive state. Hence, contrary to Cu(I), any oxidative reaction could take place. 1,2,3-triazoles could be synthesized using microwave irradiation and ultrasound.^{42–44} The CuAAC is simple to perform because of its reaction conditions (no protecting step, no sensitivity to the presence of oxygen or water) and its high yields. Moreover, it requires a stoichiometric amount of starting material and does not furnish by-products. For example, non-ionic 1,2,3-triazole surfactants have been obtained from D-glucose and D-galactose.⁴⁵

There is a large panel of 1,2,3-triazole surfactants in the literature, which involved several positions of the sugar: either anomeric position or any other position which needs a prior protection step. Two different routes were proposed to provide the desired products. The linkage between the sugar and the hydrophilic tail can be formed by either a glycosyl azide or an alkynyl glycoside to be processed into triazole by CuAAC. Initially, this reaction involved an alkynyl glycoside which reacted with an alkyl azide (Reaction A, Scheme 12).^{46,47} Recently, a complementary alternative has been developed using an azidoalkyl glycoside and an acetylenic group included in a hydrophobic chain (Reaction B, Scheme 12).^{48,49} Thus, the strategy for synthesizing such products relies on the accessibility of the raw materials.

Introduction of the triazole linker into the anomeric position. Functionalization of the anomeric position was achieved essentially *via* the formation of a glycosidic linkage, according to the two approaches described in Scheme 12. Han-Su *et al.*⁵⁰ have recently investigated the effect of the type of the triazole linkage and the number of linking methylene groups between the sugar and the triazole, especially on the surface properties of the alkyl triazole glycosides (ATG). Their approach consisted in coupling reactions that involved two types of sugar derivatives. Firstly, the most common one was based upon the reaction of an



Scheme 12 Two synthetic approaches for the synthesis of 1,2,3-triazole derivatives.

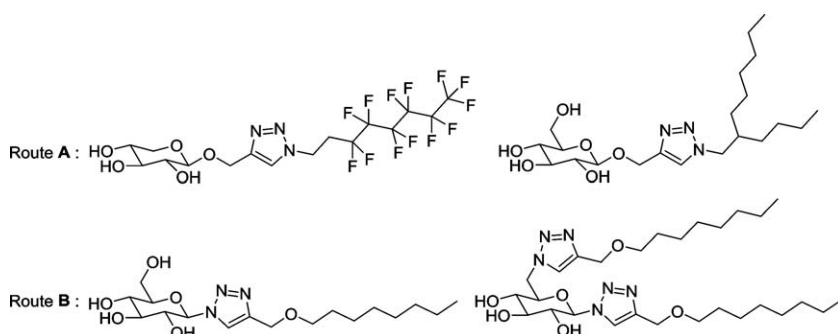
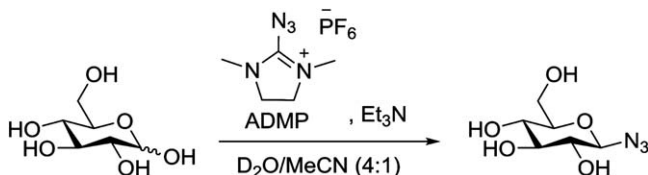


Fig. 5 Structures of triazole-based surfactants according to route A,^{46,47} or route B.⁵³

alkynyl glycoside with an alkyl azide^{48,51} and for the second one, the reaction involved an azidoalkyl glycoside and a propargylated fatty alcohol.^{52–54} Furthermore, sugars bearing an azido group are very attractive because of the stability of the azide functionality under various reaction conditions.

In the case of glycoside alkylation, the strategy was based on the formation of 1,2,3-triazole linkers at the anomeric position of sugars starting from an alkynyl glycoside previously prepared from a per-*O*-acetylated-glycopyranose. Sani *et al.*⁴⁶ realized the homogeneous glycosylation of propargyl alcohol in the presence of Amberlite IR120, followed by CuAAC with an alkyl azide, Cu(OAc)₂ and sodium ascorbate in 69–73% yields. The same procedure was followed by Oldham *et al.*^{47,51} with BF₃·OEt₂ as the promoter for the glycosylation reaction and CuSO₄·5H₂O and sodium ascorbate for the CuAAC in 64–80% yields. The resulting surfactants (Fig. 5) differed in terms of hydrophobic chain (fluorinated chain⁴⁷ or double hydrophobic chain⁴⁶) and polar head group (glucose⁴⁶ and xylose⁴⁷).

Regarding glycoside azides, different approaches for their preparation were reported in the literature even if they were commercially available. The oldest method⁵³ involved a first step of chlorination with SOCl₂ and



Scheme 13 Direct synthesis of glycosyl azides using ADMP reagent.

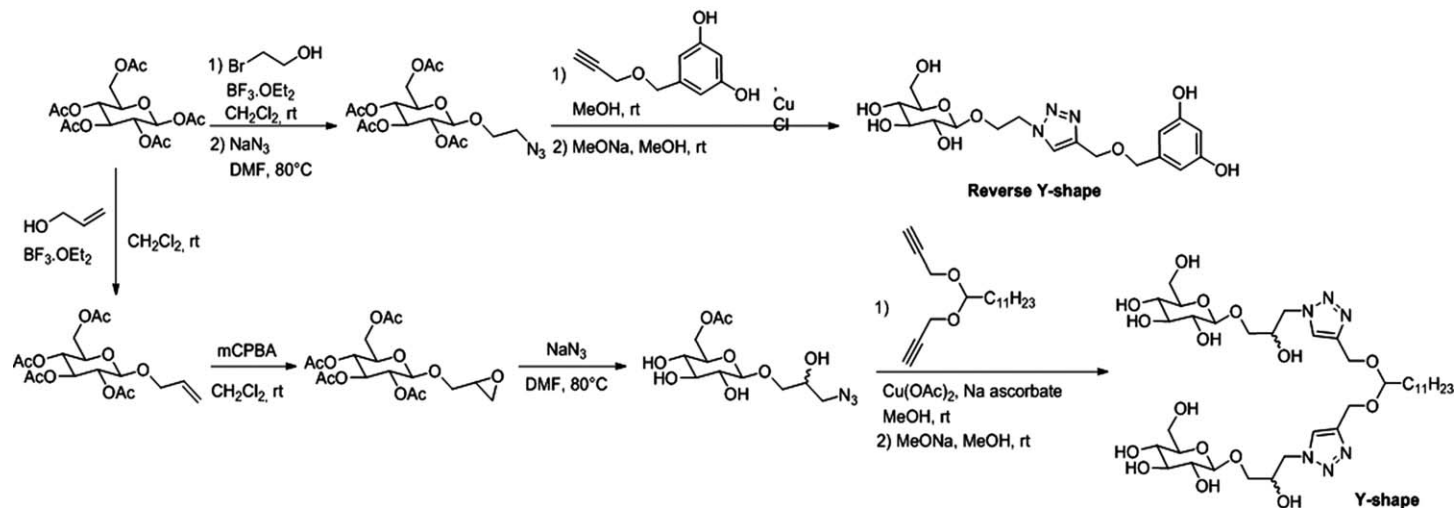
BiCl_3 , generated *in situ* in CH_2Cl_2 at room temperature, in 82–97% yields.⁵⁵ The subsequent substitution reaction with NaN_3 in aqueous acetone was carried out in 67–78% yields from 2,3,4,6-*O*-tetraacetylated-(gluco, galacto, manno, rhamno, arabino, xylo, lacto, malto)pyranose.⁵⁶ Finally, the click reactions were performed with *n*-octyl or *n*-dodecyl propargyl ether in the presence of copper(II) acetate and sodium ascorbate (Fig. 5).

A recent one-pot method was developed by Lim *et al.*⁵⁴ for the direct and completely stereoselective synthesis of β -glycosyl azides from the reducing sugars. They used 2-azido-1,3-dimethylimidazolinium hexafluorophosphate (ADMP), as both activating agent for the anomeric hydroxyl group and source of azide, in the presence of trimethylamine in $\text{D}_2\text{O}/\text{MeCN}$ (4:1) (Scheme 13). Then, the glycosyl triazoles were synthesized in the presence of propargyl alcohol, copper(II) sulfate pentahydrate and L-ascorbic acid in yields ranging from 73% to 88%.

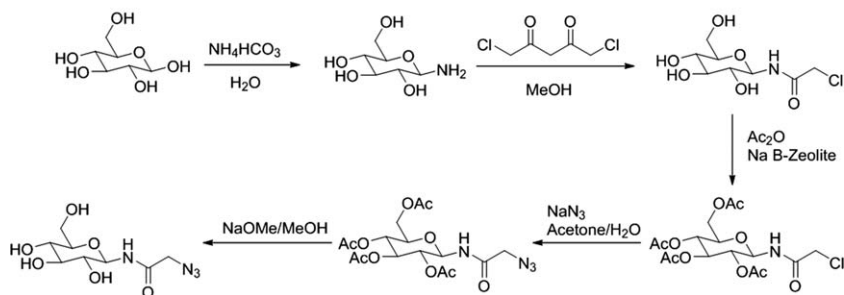
Starting from commercially available glycosyl azides, Schuster *et al.*⁵² developed branched fluorinated amphiphiles in which the hydrophobic moieties are fluorophilic residues (*e.g.* perfluoroalkyl chain, R^{F}). These fluorinated blocks were alkyne-functionalized by Michael-addition reaction with propyne-thiol in yields exceeding 95%. In the final step, the hydrophilic-fluorophilic amphiphiles were prepared by CuAAC using tetrakis(CH_3CN)copper(I) hexafluorophosphate and 2,6-lutidine in yields of 83% for $\text{R}^{\text{F}} = 2$ and 43% for $\text{R}^{\text{F}} = 3$ under identical reaction conditions. Such a difference can be explained by the lower accessibility of the alkyne group due to a more pronounced steric hindrance.

Alternatives for the introduction of azide function involved a nucleophilic substitution reaction at the anomeric position bearing an acetate group with either 2-bromoethanol⁴⁹ or an allylic alcohol^{48,49} in the presence of a Lewis acid, $\text{BF}_3 \cdot \text{OEt}_2$. In the first case, the glycosyl azide was formed by additional nucleophilic substitution with NaN_3 whereas the second method required more steps which are epoxidation and ring opening with NaN_3 , leading to two types of substrates. Depending on the hydrophobic residues and the stoichiometric amount of each product, the authors have succeeded in synthesizing a large panel of derivatives: (1) Y-shaped surfactants⁴⁸ (two head groups and one hydrocarbon chain) after CuAAC with copper(II) acetate or sulfate pentahydrate in 47–63% yields; (2) Reverse Y-shaped surfactants⁴⁹ (two hydrophobic tails and one carbohydrate head group) after CuAAC with CuCl with yields of 70% (Scheme 14).

The last strategy consisted in creating an amide function at the anomeric position prior to the introduction of the azide functionality.



Scheme 14 Synthetic routes toward Y-shaped and (reverse)-Y-shaped surfactants.

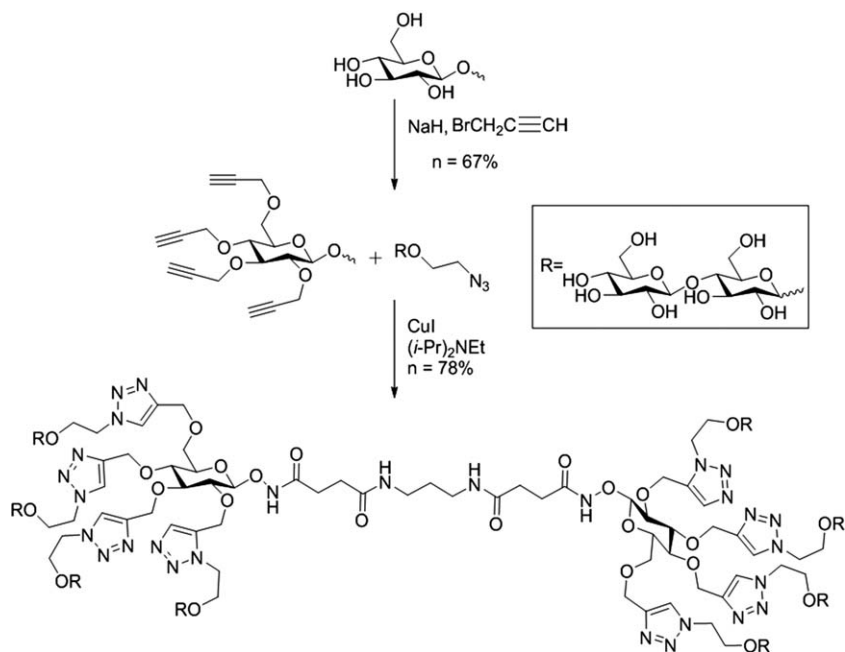


Scheme 15 Insertion of an azide function at the anomeric position via an amide linkage.

This approach was developed by Loganathan *et al.*^{53,57} via a multi-step process described in Scheme 15. The first step of this synthetic pathway involved the addition of ammonia to yield a glycosylamine. Then, they performed the *N*-chloroacetylation with chloroacetic anhydride in dry methanol and acetylation of the secondary hydroxyl groups with acetic anhydride. Afterwards, the nucleophilic substitution with NaN_3 in aqueous acetone proceeded with yields of 81–95% to end up de-*O*-acetylating.⁵⁷ This multi-step synthesis is rather tedious but it furnishes azidoacetamido sugars which can lead to interesting triazolyl derivatives.⁵³

Introduction of the triazole linker into other positions. In order to synthesize new sugar-based surfactants, some researchers have investigated functionalization or substitution of the other hydroxyl groups of the carbohydrate using various reactions such as etherification and esterification. In all cases, the anomeric position was alkylated (usually –OMe), because of its higher reactivity. For this purpose, a glycosidic bond with different alkyl chain lengths was formed.

The etherification strategy relied on the functionalization of one or more hydroxyl groups of the sugar with propargyl bromide in the presence of an effective base. Then, the reaction followed the general synthetic pathway described in Scheme 12 (Reaction A). Several authors used this approach to prepare a large variety of surfactant structures. For example, Sebyakin *et al.*⁵⁸ developed an octavalent bolaamphiphile (*i.e.* an amphiphilic molecule that has hydrophilic groups at both ends of a sufficiently long hydrophobic hydrocarbon chain) starting from penta-*O*-acetyl- α -D-glucose (Scheme 16). The linker which created the bolaamphiphile structure was introduced through the opening of a dioxypyrrolidine ring with propane-1,3-diamine. Then, multiple acetylenic groups were introduced by the reaction of the hydroxyl groups with propargyl bromide in the presence of sodium hydride (20 equiv.) in 67% yield. Triazole function was built by conjugation of acetylenic functions with 2-azidoethyl β -lactoside as the hydrophilic moiety, in the presence of a catalytic amount of copper(I) iodide and ethyl(diisopropyl)amine. This strategy was based on a simple nucleophilic substitution with a bromide derivative followed by a 1,3-dipolar cycloaddition in 52% overall yield.



Scheme 16 General synthesis of a multi-OCH₂-triazole-bolaamphiphile.

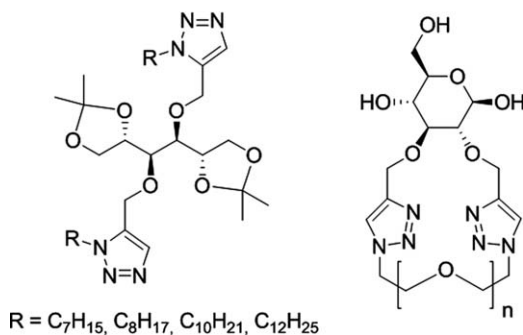


Fig. 6 Structure of ether-linked bis-1,2,3-triazoles.

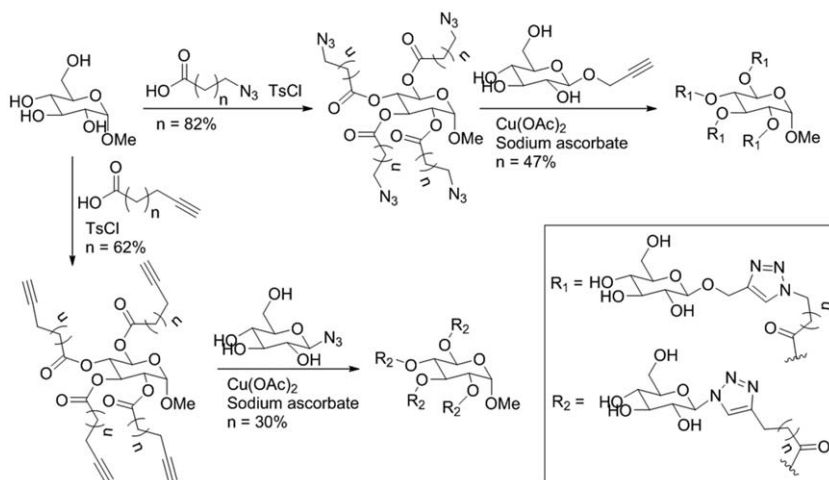
Synthesis of 1,2,3-triazole sugar-based surfactants were also investigated at privileged positions *via* a prior protection step of hydroxyl groups, the triazole linkage being compatible with a wide range of them, such as acetals and benzyl ethers. Some researchers explored this strategy by using copper(II) derivatives. Mohammed *et al.*⁵⁹ prepared gemini surfactants (*i.e.* two single-chain surfactants chemically linked by a spacer molecule), starting from D-mannitol, *via* ether-linked bis-1,2,3-triazoles (Fig. 6). For the etherification step, they investigated the reaction of propargyl bromide in the presence of various basic reagents (NaH, NaOH, KOH) and they found that the best conditions involved the use of NaOH in DMF from $-20\text{ }^{\circ}\text{C}$ to room temperature. Indeed, these conditions allowed a very good yield of 81%. For the second step of

cycloaddition, another copper derivative, copper(II) sulfate pentahydrate was used in the presence of sodium ascorbate with the aim of generating Cu(I) acetylides *in situ*. In these conditions, very good yields (80–89%) were obtained, which was much more better than results obtained with direct Cu(I) derivatives.

Salman *et al.*⁶⁰ synthesized various glycolipid-crown-ether analogues (Fig. 6) through the bis-propargylation of lauryl glucoside with propargyl bromide under phase transfer using tetrabutylammonium bromide, Bu₄NBr.⁶¹ Indeed, these compounds are soluble both in aqueous and organic solutions. The deprotonated sugar, in association with the phase-transfer catalyst, is entirely soluble in the organic phase, thus enabling the nucleophilic substitution to work more effectively. In these conditions, the products were isolated in 90% yield. The second step was performed with copper(II) acetate in the presence of sodium ascorbate which afforded the different macrocycles in yields exceeding 80%.

A tetra-substituted “star-like” compound was synthesized by Neto *et al.*⁶² (Scheme 17). The esterification reaction was performed between the hydroxyl groups of the sugar and long-chain functionalized carboxylic acid, substituted accordingly by either a terminal azide or acetylenic group to be coupled, in the presence of tosyl chloride with yields of 82% and 62% respectively. These results were quite satisfactory given the lack of basis to neutralize the hydrogen chloride released during the reaction. Thus, the cycloaddition reactions with the corresponding acetylenic- or azide-functionalized unprotected glucose in the presence of copper(II) acetate and sodium ascorbate provided the two “star-like” compounds with respective yields of 30% and 47%. These lower yields of cycloaddition can be explained by the high level of steric hindrance. Nevertheless, the two different synthetic routes led to two original surfactants in moderate yields (38% and 20%).

Salman *et al.*⁶³ proposed the synthesis of a sugar-based surfactant *via* the reaction B described in Scheme 12. The primary alcohol at the



Scheme 17 Synthetic route toward the “star-like” compounds.

6'-position underwent chloride substitution, with *N*-chlorosuccinimide in the presence of triphenylphosphine⁶⁴ to facilitate the reaction of sodium azide.⁶⁵ Then, they performed the CuAAC between this azide and fatty acid propargyl ester ranging C₁₂ and C₁₈⁶⁶ in the presence of copper(II) acetate and sodium ascorbate in 74–78% yields.

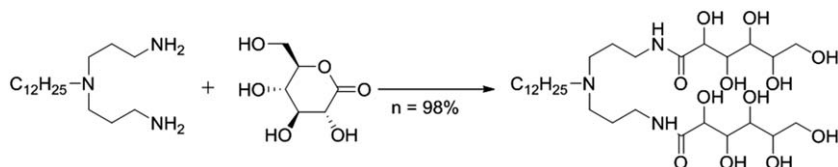
4.1.2 Other linkers

Introduction of linkers into open-chain sugars. While sugars exist largely in cyclic form *in vivo*, surfactants of alkylated, open-chain sugars are widely studied. Indeed, the flexibility of the sugar surfactant is also an important factor and was most commonly achieved by reducing the sugar group to obtain alditols. One of the best known methods involved an amide bond at the anomeric/C-6 position of the sugar. This strategy has been significantly expanded to more original linkages to furnish a wide range of amphiphilic compounds.

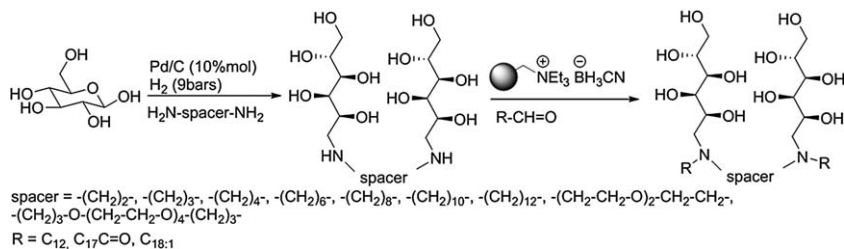
The synthesis of alkylaminoamide sugar surfactants was described by Oskarsson *et al.*⁶⁷ by using a long-chain alkylamine possessing a couple of aminopropyl chains attached to the alkylamine nitrogen. The general method consisted in the simple stirring of *N,N*-di(3-aminopropyl)dodecylamine with *D*-gluconic acid δ -lactone leading to the open-chain product with a 98% yield after a final precipitation step (Scheme 18).

With the idea of developing gemini surfactants, novel nonionic amphiphiles characterized by α,ω -diamino-alkyl or α,ω -diamino-ethylene oxide spacer have been synthesized by Wagenaar *et al.*⁶⁸ (Scheme 19). In a first step, a reductive amination was carried out in the presence of a catalytic amount of Pd/C (10 %mol) under hydrogen pressure between the H₂N-spacer-NH₂ and the sugar. A large range of α,ω -diaminoalkyl spacers (varying from ethyl to dodecyl chain) and α,ω -diamino-ethylene-oxide, $-(CH_2-CH_2-O)_2-CH_2-CH_2-$ or $-(CH_2)_3-O-(CH_2-CH_2-O)_4-(CH_2)_3-$ spacers was used with various sugars or aldehydes (glucitol, mannitol, galactitol, lactitol, talitol). Secondly, reductive alkylation or acylation of the bipolar intermediates with various aldehydes/aclys (C₁₂, C_{18:1}, C₁₇C=O) was performed by using NaBH₃CN or a polymer-bound cyanoborohydride to afford the corresponding gemini products in 43–89% yields.

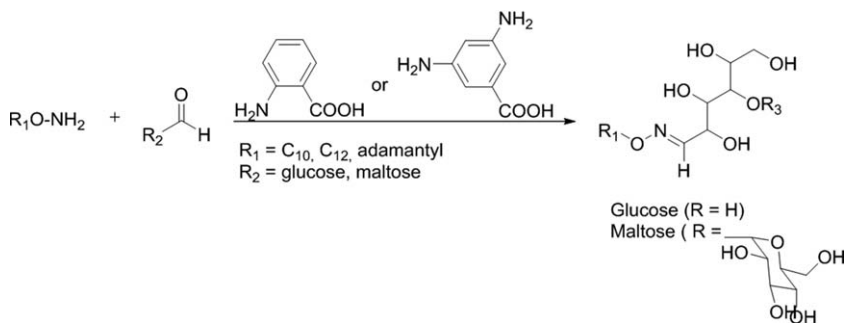
Since the traditional linkages (acetals, esters and amides) are rather labile, further limiting their applications, another approach tried to prevent this weakness to extend the use of sugar surfactants. A novel ether oxime linkage was described by Ewan *et al.*⁶⁹ to be used in neutral applications such as emulsifiers and stabilizers. The synthetic route involved a chemoselective condensation (without requiring protection/deprotection steps) of the aldehyde/ketone with hydrophobic



Scheme 18 Synthesis of an alkylaminoamide sugar surfactant (C₁₂-DGA).



Scheme 19 Two-step synthesis of gemini surfactants.



Scheme 20 Synthesis of the sugar oxime ether surfactants.

alkoxyamine under mild conditions to afford the sugar oxime ether surfactants in quantitative yields (Scheme 20). The different alkoxyamines were prepared through the reaction of alkyl bromides⁷⁰ or alcohol derivatives⁷¹ with *N*-hydroxyphtalimide followed by a hydrazinolysis. For the condensation reaction, various acid catalysts have been tested; hence, anthranilic acid or 3,5-diaminobenzoic acid was selected to furnish the ether oxime surfactant in good to excellent yields (51–98%).

Finally, Kaplânek *et al.*⁷² used as starting materials open form derivatives, 1-deoxy-1-methylamino-D-glucitol or 1-amino-1-deoxy-D-glucitol. Various polyfluoroalkylated glucamine compounds were prepared *via* the reaction of glucitol substrates with 2-[(perfluoroalkyl)methyl]oxiranes in refluxing ethanol through the opening of the epoxide ring (85–98% yields).

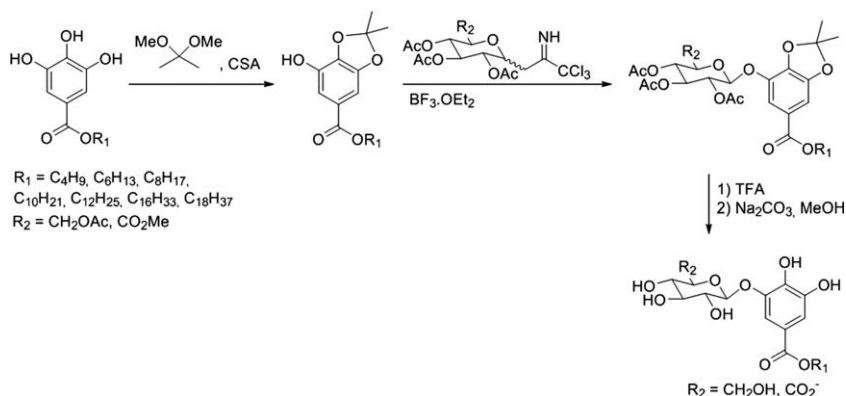
Introduction of novel linkers into cyclic sugars. In recent years, considerable progress has been made in the synthesis of glycosides. Considering surfactant molecules, a wide variety of linkage connecting carbohydrates and hydrophobic tails is observed, including different heteroatoms (*e.g.* O, N, S) and functional groups. The most prevalent types comprise *O*-alkyl glycosides and glycosyl amides. A number of approaches – including esterification, glycosylation with different nucleophiles or click chemistry – were considered for the synthesis of *O*-glycosides employing protected or unprotected sugars as starting materials. The click chemistry was performed with *O*-glycosyl trichloroacetimidates as a new type of glycosyl donors. They are easily prepared, sufficiently stable and can be activated for the glycosylation

reactions with catalytic amount of Lewis acids such as TMSOTf, $\text{BF}_3 \cdot \text{OEt}_2$, $\text{Sn}(\text{OTf})_2$, AgOTf and $\text{ZnCl}_2 \cdot \text{Et}_2\text{O}$.

According to this method, gallate units were used as linkers for the synthesis of new surfactants with antioxidant properties. A series of glycosyl alkyl gallates that contain glucose and glucuronic acid attached to the alkyl gallate scaffold were designed by Maldonado *et al.*⁷³ The synthesis started with the acetal protection of the different alkyl gallates with different chain lengths from C_4 to C_{18} . This step is necessary to maintain a di-*ortho* phenolic moiety which would be responsible for the antioxidant activity. Next, glycosylation reactions were carried out using classical Schmidt's protocol, with $\text{BF}_3 \cdot \text{OEt}_2$ as the promoter, either with the glucosyl or the glucuronosyl trichloroacetimidates donors. The final products were obtained after two steps of deprotection. Thus, treatment with trifluoroacetic acid to remove the isopropylidene group followed by hydrolysis with Na_2CO_3 afforded the final products with good to excellent yields (Scheme 21).

Starting from the same glycosyl trichloroacetimidates, Liu *et al.*⁷⁴ prepared novel gemini alkyl glucosides using ethylene glycol diglycidyl ether as the spacer because it is hydrophilic, nontoxic, and has readily available polymeric forms of variable length (Fig. 7). The linker was positioned between the two alkyl chains in order to achieve a tighter connection owing to elimination of the steric interaction between bulky sugar moieties which exists if two sugar moieties are linked together.

As alternative to *O*-glycosides, *C*-glycosides have attracted much attention, considering that many of them demonstrated interesting properties. The increasing interest for these molecules is due to the fact that the conformational differences of *C*-glycosides compared to *O*- or *N*-glycosides are minimal. Furthermore, the base-labile ester or acid-labile *O*-glycosidic linkages are unstable whereas *C*-glycosides are resistant toward enzymatic and acid hydrolysis since the anomeric center has been transformed from acetal to ether. Recent advances in carbohydrate C–C bond at the anomeric position have been useful for the synthesis of new classes of surfactants.⁷⁵ For this purpose, two series of these new surfactants have been designed – linear and cyclic *C*-glycosides – from a



Scheme 21 General synthesis for the sugar-based alkyl gallates.

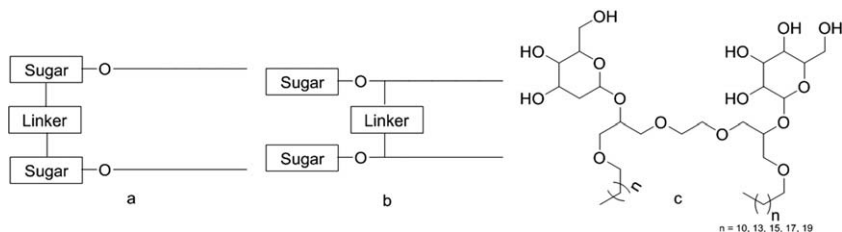
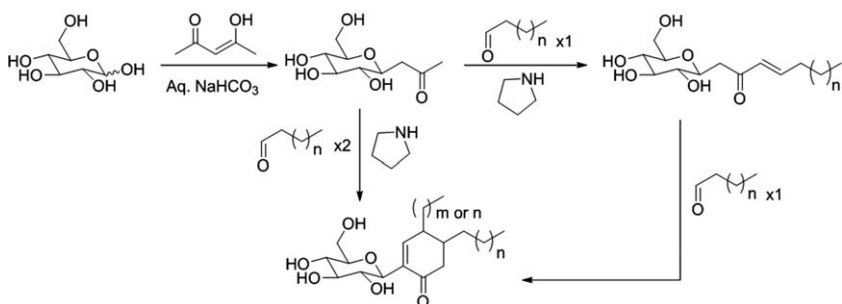


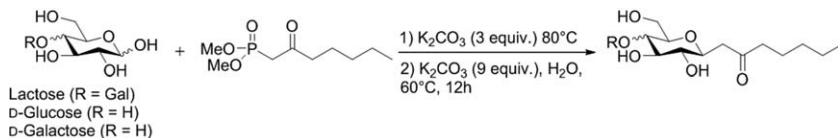
Fig. 7 General structures of gemini alkyl glucosides with the linker connected to two sugars (a) or two alkyl chains (b). In (c), example of new gemini alkyl glucosides.



Scheme 22 General synthetic approach to C-glycoside surfactants.

nonulose intermediate. This latter was first developed by Lubineau and coworkers⁷⁶ from D-Glucose and D-Mannose using the Knoevenagel condensation with pentane-2,4-dione in the presence of aqueous NaHCO_3 or $\text{Yb}(\text{OTf})_3$. The synthesis of C-glycosides through an enamine-catalyzed aldol condensation with an alkyl aldehyde in the presence of pyrrolidine was described by Foley *et al.*⁷⁷ By varying the number of equivalents of the aldehyde, they were able to produce linear (1.0 equiv.) or cyclic (≥ 2.0 equiv.) C-glycosides with good to excellent yields (Scheme 22). Furthermore, the linear enone can be modified by photochemical isomerization that could provide compounds with different absorbance properties.

Another expeditious synthetic route described by Ranoux *et al.*^{78,79} was based upon an one-step Horner–Wadsworth–Emmons (HWE) reaction from unprotected mono- (D-glucose, D-galactose and D-xylose) and disaccharide (lactose) using β -ketophosphonates and potassium carbonate as the base (Scheme 23). The olefination of the masked aldehyde led to an electrophilic double bond that underwent a Michael addition reaction to cyclic C-pyranosides or C-furanosides. HWE reaction in aqueous media was successfully applied to short chain length pure alkyl-oxo-phosphonates and it provided very good selectivity in favor of the β -pyranoside forms. Solvent-free HWE reaction followed by a simple aqueous basic treatment yielded β -C-glycosidic ketones with various alkyl chain lengths. In particular, C-glycosides possessing a short (C_7) alkyl chain were obtained in moderate to good yields with high β -selectivity. These sugar-based surfactants were found to exhibit attractive solubilizing properties comparatively to O-glycoside references.



Scheme 23 One-step Horner–Wadsworth–Emmons reaction.

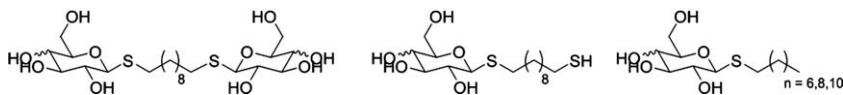


Fig. 8 Examples of thioglycosides from InBr_3 -promoted glycosylations.

There are few examples of *S*-glycosides as surfactants in the literature. Thio-glycosides (Fig. 8) were synthesized with a molar excess of a strong Lewis acid, $\text{BF}_3 \cdot \text{OEt}_2$. A better method to get alkyl- β -*S*-glycopyranoside has been recently reported by Szabo *et al.*⁸⁰ starting from peracetyl- β -D-glycopyranose and various thiol derivatives, in the presence of a catalytic amount of the Lewis acid, InBr_3 (1 mol%). As previously reported, InBr_3 proved to be excellent glycosylation promoters.⁸¹ Bolaamphiphiles and mono-glycosides were prepared by adjusting the stoichiometry of the glycosyl donor and the acceptor to favour either di- or mono-glycosylation of the α/ω dithiol (69–88% yields).

4.2 Physico-chemical properties

In this section, we will discuss how the surface active properties in terms of critical micelle concentration (CMC), equilibrium surface tension γ_{CMC} or whether the area/molecule depend on the structure of the sugar-based surfactants and particularly the nature of the linkers. It is known that the CMCs of sugar-based surfactants are dependent on the alkyl chain length⁸² *i.e.*, when the alkyl chain length is increased by two methylene groups, it decreases with a factor 10.

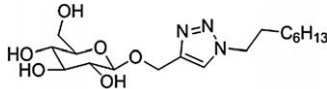
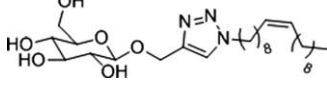
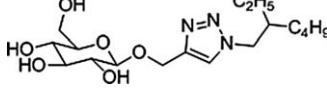
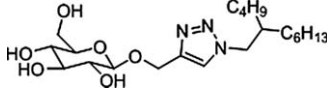
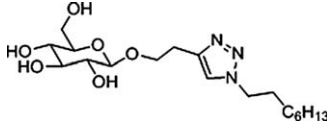
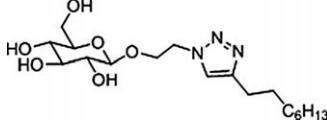
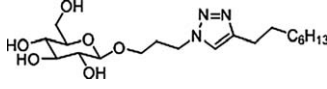
Table 1 gives an overall view of the difference between alkyl glucosides and alkyl triazole glucosides from C_8 to C_{16} . As we can see, ATG have the same dependence on the alkyl chain length but there is 10% decrease of the CMCs with the introduction of four methylene groups. Furthermore, the introduction of the triazole functionality enabled to significantly decrease the CMC for each chain length, thus these products exhibited interesting surface active properties. The most significant difference appears for C_8 with a CMC of 20 mmol L^{-1} for the *n*-octyl- β -glucoside and 1.3 mmol L^{-1} for the 4-(β -glucopyranosyloxymethyl)-1-octyl-1,2,3-triazole. This reduced solubility for APG is probably due to the triazole moiety which adds an additional hydrophobic element to the glycolipid. Inversely, the surface tension of ATG solutions above the CMC does not follow the same trend *i.e.*, γ_{CMC} values remained constants at 29 mN m^{-1} regardless of the chain length.

The effects of the distance between triazole moiety and sugar, the orientation of the triazole connection on the sugar and the nature of the hydrophobic tails have been studied and the results are listed in Table 2.

Table 1 Comparison of CMC and surface tension for ATGs and corresponding AG.^{74,83}

C_n	AG		ATG		$T_{k,ATG}$ ($^{\circ}\text{C}$)
	CMC (mmol L^{-1})	γ_{CMC} (mN m^{-1})	CMC (mmol L^{-1})	γ_{CMC} (mN m^{-1})	
8	20	30.2	1.3	29	<10
10	2	28.5	0.77	28	<10
12	0.19	39.5	0.11	29	<10
14	0.58	28.82	0.065	29	42
16	5.32	29.94	0.0045	29	53

Table 2 Effects of the distance and type of the hydrophobic moieties.^{46,50}

Entry	ATG	CMC $\times 10^{-3}$ (mol L^{-1})	γ_{CMC} (mN m^{-1})	Surface Area A ($\text{\AA}^2$)	T_k ($^{\circ}\text{C}$)
1		1.3	29	—	<10
2		0.25	29	—	<10
3		8.0	29	—	<10
4		0.58	29	—	<10
5		2.8	28.5	65	<10
6		3.6	28.5	41	<10
77		1.2	28.7	40	<10

Triazole surfactants with branched tails have also been examined and exhibited higher CMC values than the corresponding straight analogues *e.g.*, 4-(β -glucopyranosyloxymethyl)-1-(2-ethyl-hexyl)-1,2,3-triazole (entry 3 = 8.0 mmol L^{-1}) compared to ATG-C₈ (Table 1, CMC = 1.3 mmol L^{-1}) and 4-(β -glucopyranosyloxymethyl)-1-(2-butyl-octyl)-1,2,3-triazole (entry 4 = 0.58 mmol L^{-1}) compared to ATG-C₁₂ (Table 1, CMC = 0.11 mmol L^{-1}). Indeed, the CMC increases with the branching and represents the potential for increased solubility in water of the hydrophobic tails. Inversely, the

insaturated surfactant (entry 2) exhibited lower CMC while keeping the same surface tension at 29 mN m^{-1} . This behaviour is consistent with the values previously reported for AGs which also show a chain independent surface tension in the range above C_{12} . Nearly all ATG surfactants showed Krafft temperature below 10°C except for linear long chain above C_{12} . The Krafft point reflected the high melting point for the surfactant and the results were in accordance with those obtained for AGs *i.e.*, much higher chain length presented a constraint towards the solubilisation temperature.

Regarding entries 1,5–7 in Table 2, all compounds exhibited very similar surfactant properties based on surface tension values. However, results showed that a longer distance between the sugar and the triazole led to a better water solubility (CMC values of entries 6 and 7) when the triazole was connected to the sugar *via* the nitrogen atom. The surface areas for the last two surfactants were quite similar, thus indicating that the triazole has no major effect neither on the interfacial orientation or on the surface activity. Furthermore, the higher surface area for the ATG, entry 5, is probably caused by minor impurities, which could not be detected in the NMR spectrum.

Additional triazole-based surfactants were widely characterized in terms of CMC, surface tension, Krafft temperature and surface area (Fig. 9). Compounds A–E characterized by the presence of one or two (2-hydroxy-3- β -D-glucopyranosyloxy) scaffolds displayed very similar surface active properties with Krafft point below 10°C while no clouding was observed upon heating. This renders the surfactants as promising candidates for emulsifier applications. Surface tension values ($30\text{--}50 \text{ mN m}^{-1}$) and surface area per surfactant ($51\text{--}62 \text{ \AA}^2$) exceed typical values around 30 mN m^{-1} and $40 \text{ \AA}^2$ for simple ATG which is consistent with the previously reported range of values for double headed surfactants. It reflected an enhanced tendency of the side-by-side alignment for a curved assembly. All surfactants exhibited similar CMCs, reflecting the size of the hydrocarbon chain. The lowest CMCs are provided by compounds D, E and G, involving longer alkyl chain into the linking unit, like isopropyl or diethyl-1,3-dioxane linkers. The inclusion of a “phenol” group – which is equivalent to about three methylene group in terms of hydrophobicity – can explain this interesting CMC value, surprisingly without changing the Krafft temperature. This effect was also observed for compound F which provided attractive surface active properties. This hypothesis is in part confirmed by the physico-chemical properties of compound C ($\text{CMC} = 3.7 \text{ mmol L}^{-1}$) which presents a short alkyl chain into the linking unit.

The comparison between compounds A, B and D that contain a two-oxygen or an amine linker, shows that the CMC is not much influenced by the nature of the heteroatom but primarily by the number of carbon above the link between the two polar head groups. However, compounds A and C with amine-based linkers led to poor emulsification efficiency, likely due to repulsive ionic interactions.

Fluorous surfactants are generally more effective than hydrocarbon surfactants of comparable properties. This can be explained by the

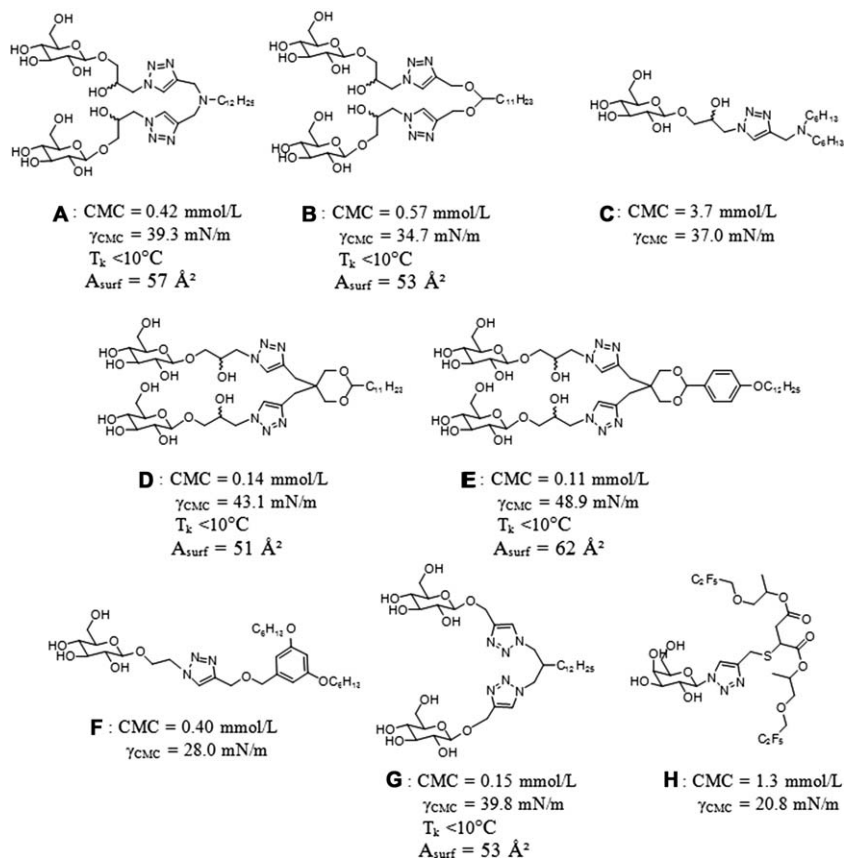


Fig. 9 Structure and physico-chemical properties of triazole-based surfactants.

greater free energy difference of transferring a $-\text{CF}_2-$ group from the bulk state to a micellar environment relative to that of a $-\text{CH}_2-$ group. Hence, compound **H** presented interesting surface tension, sufficient for application in processes such as wetting of paints or water repellent coatings.

The physico-chemical properties described in Table 3 concern surfactants with no triazole-based linker. They have been synthesized *via* common ways to link hydrophilic and hydrophobic moieties like ether and amide bond or newly developed, *C*-glycosides. Entries 1–2 show two open form surfactants which involve ether oxime or amide bond. The pH-dependence of γ -amine derivatives has an effect on the solubility and affects physico-chemical properties, CMC and surface tension values. Indeed, the surface tension of 36.2 mN m^{-1} was obtained at pH 7 and is most likely due to a favorable mixture of positively charged and non-charged in the interfacial film. This difference is also expected considering surfactant hydrophilicity and bulkiness of the polar head group.

We can observe that for each case of the *O*-glycoside series, an increase in the length of the alkyl chain led to a decrease of the CMC, as occurred for conventional surfactants. Nevertheless, for alkyl chains longer than C_{12} , the surfactant performance decreased until not behaving as

Table 3 Physico-chemical properties of open or *O*-/*C*-glycosides.

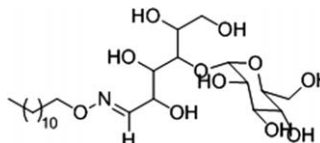
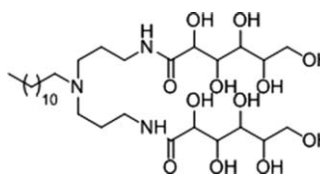
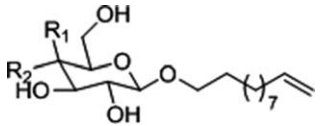
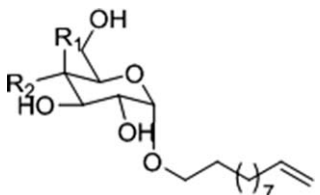
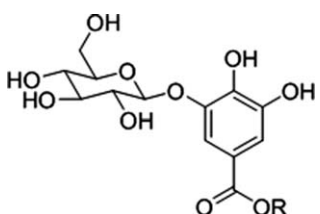
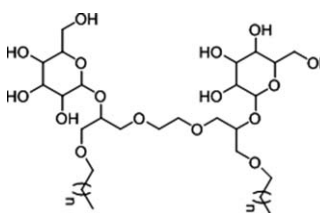
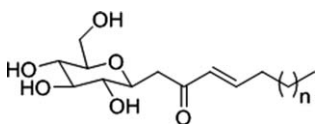
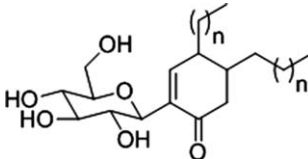
Entry	Surfactants	CMC×10 ³ (mol L ⁻¹)	γ _{CMC} (mN m ⁻¹)	Surface Area A (Å ²)
Open forms				
1 ⁶⁹		0.12	20.0	—
2 ⁶⁷		0.63	36.2	42.3
<i>O</i> -glycosides				
3 ⁸⁵		<i>R</i> ₁ = OH : 1.70	29.0	44.3
		<i>R</i> ₂ = OH : 2.05	29.6	47.7
4 ⁸⁵		<i>R</i> ₁ = OH : 3.80	26.8	43.7
		<i>R</i> ₂ = OH : 0.60	25.1	52.9
5 ⁷³		<i>R</i> = C ₄ H ₉ : 2.40	37.5	27.4
		<i>R</i> = C ₆ H ₁₃ : 1.30	33.5	54.9
		<i>R</i> = C ₈ H ₁₇ : 0.50	31.0	42.3
		<i>R</i> = C ₁₀ H ₂₁ : 0.032	39.5	41.6
		<i>R</i> = C ₁₂ H ₂₅ : 0.010	44.3	42.9
6 ⁷⁴		<i>n</i> = 8 : 0.11	30.45	45.5
		<i>n</i> = 10 : 0.029	36.09	25.6
		<i>n</i> = 12 : 0.089	30.31	24.0
		<i>n</i> = 14 : 0.22	35.24	22.2
		<i>n</i> = 16 : 0.67	31.85	21.8
<i>C</i> -glycosides				
7 ⁷⁷		<i>n</i> = 3 : 3.9	24.4	—
		<i>n</i> = 9 : 0.1	28.0	—

Table 3 (Continued)

Entry	Surfactants	CMC $\times 10^3$ (mol L ⁻¹)	γ_{CMC} (mN m ⁻¹)	Surface Area A (Å ²)
8 ⁷⁷		$n = 1 : 55.0$ $n = 3 : 0.17$	30 22	— —

surfactants because of an excess of hydrophobicity (entries 5–6). The various glucosyl alkyl gallates (entry 5) displayed interesting surfactant properties when the alkyl chain reached octyl carbon length. This behaviour must be due to the incorporation of a glucosyl group that made the polar head group of the alkyl gallates more hydrophilic and consequently larger alkyl chains are useful to fit the proper range of HLB. In addition, a larger alkyl chain can speed up the process of auto-aggregation into micelles avoiding the contact of the hydrophobic chain with the water molecules. For glucosyl alkyl gallate series, the best surfactant effectiveness (CMC = 0.5 mmol L⁻¹, γ_{CMC} = 31.0 mN m⁻¹ and A = 42.3 Å²) was obtained with the octyl chain. Concerning the gemini surfactants (entry 6), the linkage originality results from the fact that the linkage position was attached to the alkyl chains rather than the sugar moieties. The authors outlined that CMC values are roughly 6–31 times lower than monomeric counterparts and 118 times lower than those of reported gemini alkyl glucosides in which the linkage was placed at the sugar moieties.⁸⁴ These gemini surfactants had also a great capacity to reduce surface tension as indicated by the surface tension at CMC, γ_{CMC} = 30–36 mN m⁻¹ and the area/molecule that is 1.1–2.0 times larger than that of the monomeric surfactants. The data of γ_{CMC} and A suggested a greater hydrophobic packing effect and thus a highly organized alignment of the gemini molecules at the air/water interface.

Considering linear *O*-glycosides (entries 3–4), the effect of the sugar unit, the anomeric configuration and the presence of an insaturation has been studied. For β -anomers, CMC values are quite similar (2.05 mmol L⁻¹ for glucoside and 1.70 mmol L⁻¹ for galactoside) whereas α -anomers showed a significant difference as the function of the sugar moiety (0.6 mmol L⁻¹ for glucose unit and 3.80 mmol L⁻¹ for galactose unit) but the water solubility of β -anomers is slightly lower. However, surfactant efficiency, and in particular the γ_{CMC} value, was not expected to depend on the hydrophilic part, because no drastic changes were observed for glucose and galactose (γ_{CMC} = 25–29 mN m⁻¹), but depends on the chemical nature and length of the hydrophobic part. Indeed, the addition of an insaturation increased the CMC which can be explained by its hydrophilic behaviour compared to its saturated analogues (Table 1) and also by the rigid geometry of the double bond that can disturb the tail organization inside the micelles.

Finally, the *C*-glycoside surfactants displayed promising surface active properties including drastic surface tension reduction and flexible foaming properties. These results showed that *C*-glycosides (entries 7–8) behave as their *O*-glycoside analogues (Table 1) in terms of CMC and interfacial tension reduction which are directly related to the hydrophobic-lipophilic balance (HLB). Indeed the CMC increased with the HLB, which is commonly attributed to a polydispersity of micelles. Furthermore, the difference between a linear and a cyclic linker is notable with moderate alkyl chain length considering the CMC values (3.9 mmol L^{-1} for linear derivative and 55 mmol L^{-1} for cyclic derivative).

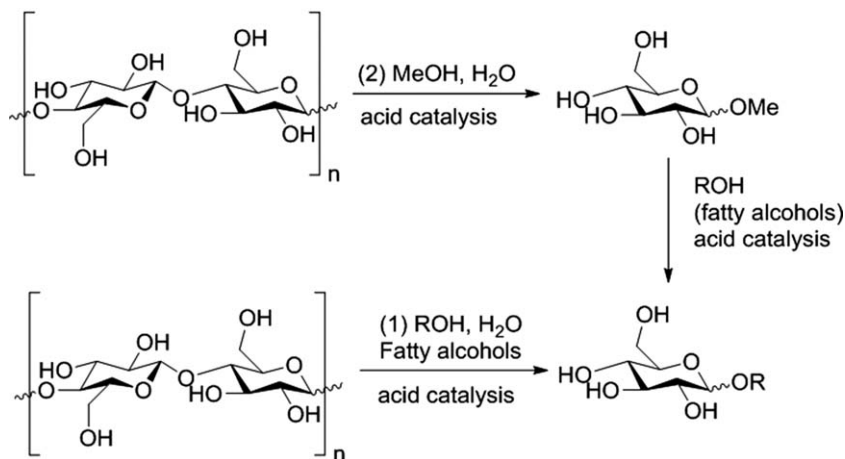
5 Surfactant compositions resulting from the direct transformation of polysaccharides

Over the past ten years, a challenging approach for the synthesis of sugar-based surfactants investigated the direct transformation of natural abundant polysaccharides into glycosyl surfactants through the one-pot hydrolysis of these biopolymers into monomers and their subsequent functionalization with lipophilic alkyl chains. The main goals of these attractive synthetic routes were to enhance cost-effectiveness of the process and to provide original surfactant compositions. Generally, cellulose and starch were used as raw materials for the direct production of alkyl glycoside surfactants but additional polysaccharides derived from lignocellulosic or algal biomass were recently explored as sources of original sugar motifs. This section will describe the reaction conditions used for the conversion of these polysaccharides into non-ionic surfactants, with a particular attention on the choice of solvents, catalysts and temperature.

5.1 Alkyl glucoside surfactants from cellulose

Cellulose, which represents an important source of biofuels and chemicals, is a crystalline polymer of *D*-glucopyranoside units linked together by β -1,4 glycosidic bonds. The different units are linked by hydrogen bonds and Van der Waals interactions which confer to cellulose a high stability, making the hydrolysis process difficult. The production of alkyl glucoside surfactants from cellulose has been previously studied and processes to provide monomeric glucose by acid or enzymatic hydrolysis have been largely reported.

A simple method to produce alkyl glycosides is the Fischer glycosylation reaction that involves the acid-catalyzed condensation (acetalization) between a carbohydrate residue and an aliphatic alcohol. There are currently two acetalization methods that can be used from cellulose (Scheme 24): (1) the direct acetalization consists of condensing cellulose and fatty alcohols; (2) the transacetalization requires the preliminary synthesis of a methylglucoside which is then reacted with an aliphatic alcohol. These reactions require the presence of different homogeneous or heterogeneous acids such as mineral and organic acids, heteropolyacids, zeolites, amorphous silica-alumina, metal oxides or sulfonated



Scheme 24 Two different routes to obtain alkyl glucosides from cellulose.

carbons. To favour the cellulose transformation into glucose, it is possible to dissolve cellulose in ionic liquids.

In a pioneering work, Villandier *et al.*^{86,87} developed a one-pot synthesis of long-chain alkyl glucosides in good yields by reacting cellulose and fatty alcohols with different acid catalysts in ionic liquid media. The two strategic routes shown in Scheme 24 were studied *via* the use of Amberlyst-15, a sulfonic resin, or H₃PW₁₂O₄₀, a heteropolyacid, as acid catalysts and 1-butyl-3-methylimidazolium chloride, BMIMCl, as the ionic liquid. In the case of direct glycosylation, the reaction was carried out using Amberlyst 15Dry (A15) as the catalyst and in the presence of a small amount of water to fulfil the requirements of cellulose hydrolysis and to limit the formation of HMF. By coupling properly the rate of cellulose hydrolysis and the rate of glycosidation of the monosaccharides formed with C₄ to C₁₂ alcohols, it was possible to obtain 82% mass yield of alkyl- α,β -glucoside plus alkyl- α,β -xyloside. Furthermore, it was found that the hydrolysis of cellulose continued during Fisher glycosidation and the selectivity decreased with alcohol chain length.

The transacetalization process started by the direct transformation of cellulose into methyl- α,β -glucopyranosides in the presence of H₃PW₁₂O₄₀ at 468 K. After 30 min, the conversion of the cellulose was 87% and two methyl glucosides (furanose and pyranose isomers) were formed in a total yield of 65%. Then, transacetalization was carried out in the presence of 1-octanol or 1-decanol and A15 at 378 K furnishing the corresponding two alkyl- α,β -glucoside isomers. These results are explained by the kinetic formation of the furanoside compound followed by the isomerization to the pyranoside product, which is thermodynamically more stable.

More recently, Climent *et al.*⁸⁸ examined the use of various sulfonated carbon, C-SO₃H, in comparison with A-15 and silica composite catalyst, Nafion SAC-13 and the possibility of a recirculation of unconverted cellulose. The different carbon materials bearing -SO₃H groups can play an important role during the successive steps and they were first evaluated

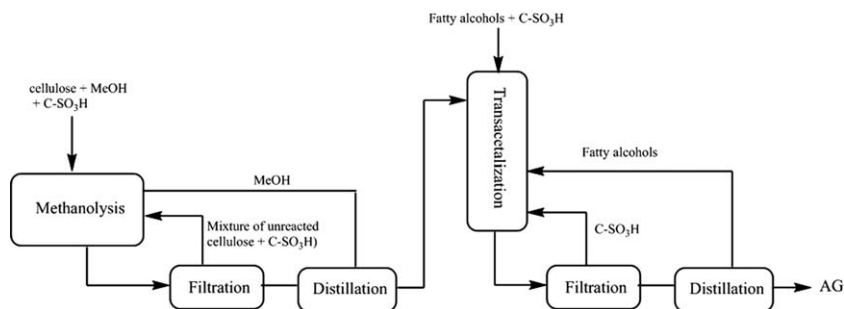


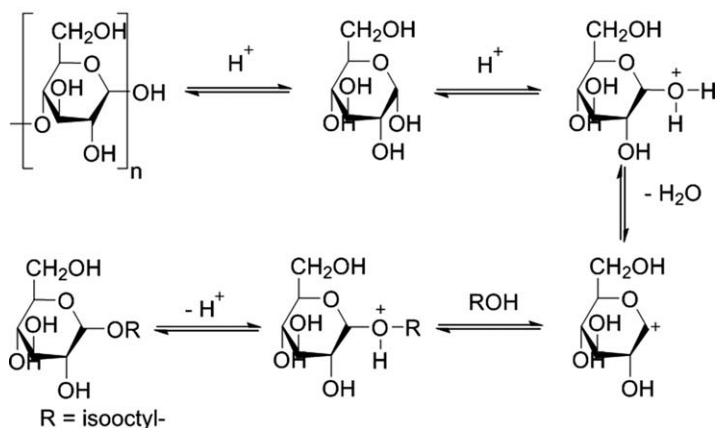
Fig. 10 Diagram of the overall process of conversion of cellulose into alkyl glucoside surfactants.

in the transacetalization step. Three of them led to promising results, as they permitted a similar conversion than A-15. The latter gave a 94% conversion of methyl- α -glucopyranoside and a 85% yield of α,β -octyl-glucopyranoside. The carbon materials bearing sulfonic groups gave also good yields (79–85%) with a lower conversion rate (84–92%). The influence of the sulfur content (1.83 wt%, 2.30 wt% and 3.55 wt%) was investigated in the methanolysis reaction. The results showed that the cellulose conversion increased with increasing the sulfur content but led also to the formation of methyl levulinate byproduct. In order to maximize the yield of depolymerized cellulose, the catalyst with 2.30 wt% was selected. With this C-SO₃H catalyst, they also studied the effect of the temperature (160, 180, 200 and 220 °C). The results showed a decrease of selectivity with increasing reaction temperature and a lower yield in low temperatures. Finally, the ideal reaction temperature has been set to 200 °C.

The original reaction process described in the publication is the recirculation of cellulose to improve the selectivity to octyl glucosides (Fig. 10). This was performed by converting the cellulose under the same conditions for 3 h and then filtering the unreacted cellulose and catalyst. Fresh methanol was added to the latter mixture and further reacted for 3 h. This process was repeated until complete conversion of cellulose was achieved. Then, all the filtrates were combined and the transacetalization was performed as previously described. This process led to a significant increase in the conversion of cellulose (97%), and in the yield of the surfactants, which includes alkyl glucosides and alkyl cellobiosides (73%). In addition, a considerable decrease in the formation by-products was observed. Finally, the determination of the surfactant properties was investigated and was a slightly better than the commercial alkyl β -glucopyranosides *e.g.*, α,β -octylGP + α,β -octyloligosaccharides (CMC = 5.70 g L⁻¹, γ_{CMC} = 31.68 mN m⁻¹) compared to commercial β -octylGP (CMC = 6.07 g L⁻¹ and γ_{CMC} = 33.71 mN m⁻¹) and α,β -decylGP + α,β -decyloligosaccharides (CMC = 0.76 g L⁻¹, γ_{CMC} = 25.09 mN m⁻¹) compared to commercial β -octylGP (CMC = 0.82 g L⁻¹ and γ_{CMC} = 27.34 mN m⁻¹).

5.2 Alkyl polyglycosides surfactants from starch

As previously mentioned, the degradation of the β -glycosidic linkage in the cellulose chain is very difficult. Therefore, an advantageous



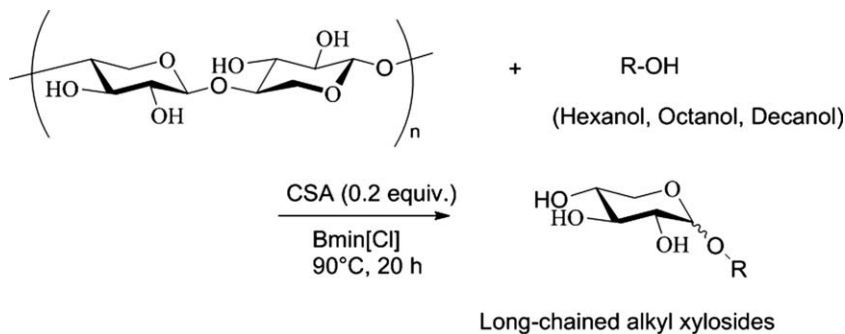
Scheme 25 Reaction mechanism of the alcoholysis of starch with sub-critical isooctyl alcohol.

alternative is the direct use of starch to produce alkyl polyglycosides through a two-step transglycosylation Fischer method.⁸⁹ Some problems such as high production cost and poor quality of product have hindered their industrial application. In this context, Zou *et al.*⁹⁰ reported a one-step direct glycosylation between starch and long chain alcohols. To alleviate the problem of compatibility between the two reagents, they used the sub-critical isooctyl alcohol for the alcoholysis of starch without the need for any catalyst. The sub-critical alcoholysis was performed in the presence of methanol and hydrogen peroxide (oxidative decoloration) under the appropriate pressure at 260 °C with different weight ratios of isooctyl alcohol/starch (5, 10 and 15 wt%). In these conditions, starch was quickly depolymerized and transformed into isooctyl glucoside in 109, 132 and 122 % yields respectively, *versus* a theoretical yield (defined as the final mass of isooctyl glycoside/initial mass of starch) of 183%. The mechanism for the alcoholysis reaction is shown in Scheme 25.

It was found that with the increase in the weight ratio of isooctyl alcohol/starch, the yield of isooctyl glucoside had a transition and it attained the maximum of 132% when the weight ratio of isooctyl alcohol/starch was 10. The increase of isooctyl alcohol sped up the alcoholysis of starch but in the presence of an excess of isooctyl alcohol the formation of a disaccharide by-product could be envisaged. The CMC and HLB of the produced isooctyl glucoside were 2.5 g L⁻¹ and 16, respectively.

5.3 Alkyl xyloside surfactants from xylan

Xylan, a hemicellulose largely found in nature, is considered as the second most abundant polysaccharide after cellulose. It occurs in close association with cellulose and lignin and contributes to the rigidity of plant cell walls in lignified tissues. The most potential sources of xylan include many agricultural crops such as straw, sorghum, sugar cane, corn stalks and cobs, and hulls from starch production, as well as forest and pulping waste products from hardwoods and softwoods.⁹¹ Xylan is a heteropolysaccharide whose main chain 1,4- β -D-xylopyranose is



Scheme 26 Transglycosylations of xylan with fatty alcohols using CSA in Bmin[Cl].

substituted with α -L-arabinofuranose, 4-O-methyl- α -D-glucuronic acid, ferulic and *p*-coumaric acids. Recently, Sekine *et al.*⁹² proposed its direct conversion to long-chained alkyl xylosides through a transglycosylation process in an ionic liquid, with the alcoholysis of xylan by a fatty alcohol (hexanol, octanol, or decanol) rather than a two-step reaction involving hydrolysis by water followed by Fisher glycosylation of xylose by the alcohol as in cellulose transformation (Scheme 26). An optimization of the reaction conditions was achieved regarding the ionic liquid, amount of catalyst, reaction time, reaction temperature, and amount of alcohol. The results obtained showed that not only the catalyst, but also the counter anion part of the ionic liquid was very important for promoting the transglycosylation process. Furthermore a sufficient amount of catalyst was required to produce alkyl α,β -xylosides but a higher quantity caused significant decomposition of both xylan and alkyl α,β -xylosides. Finally it was found that optimal transformation of xylan (5 wt% to ionic liquid) with fatty alcohol (30 equiv. to one xylosidic unit in xylan) in 1-*n*-butyl-3-methylimidazolium chloride (Bmim[Cl]), using 10-camphorsulfonic (CSA, 0.2 equiv. to one xylosidic unit in xylan) at 90 °C for 20 h gave the best result, providing alkyl α,β -xylosides in modest to good yields (with octanol: 69 wt% yield).

5.4 Alkyl L-guluronamide and D-mannuronamide surfactant compositions from alginate

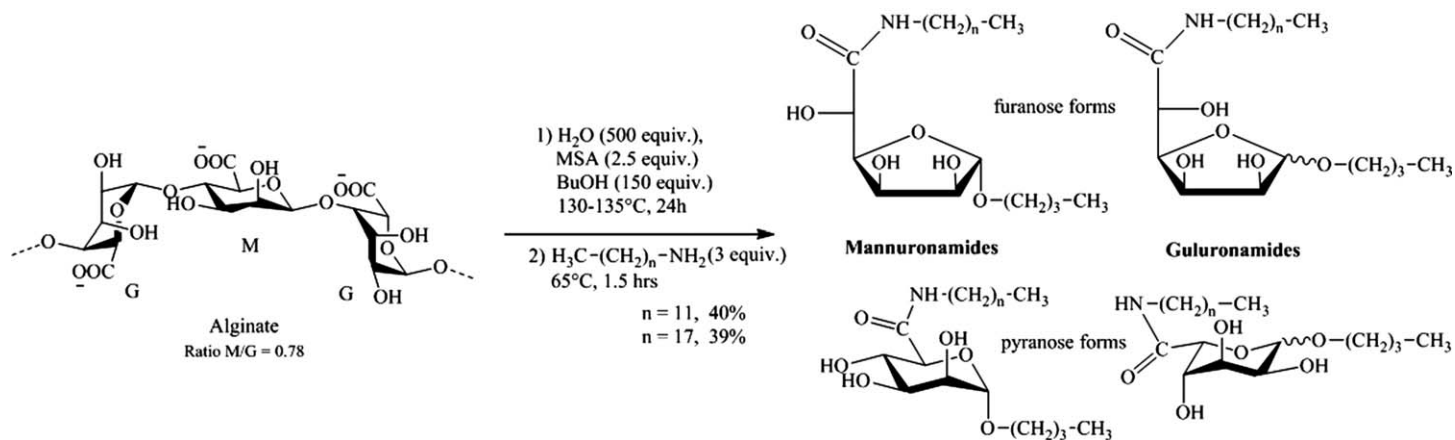
Alginate is a naturally occurring anionic polymer typically obtained from brown seaweed, and has been extensively investigated and used for many biomedical applications, due to its biocompatibility, low toxicity, relatively low cost, and mild gelation by addition of divalent cations such as Ca^{2+} .⁹³ It is a linear unbranched copolymers of (1,4)-linked β -D-mannuronate (M) and (1,4)-linked α -L-guluronate (G) residues with widely varying composition and sequence. Within the heteropolysaccharide chain, the G and M residues can be found both distributed in homopolymer blocks (*e.g.* MMMM or GGGG) comprising between 20 and 30 repeating units, and in longer statistical or strictly alternating segments (MGMG).

Sari-Chmaysem *et al.*⁹⁴ described the conversion of alginates into novel L-guluronamide and D-mannuronamide surfactant compositions.

Direct acid hydrolysis of the algal polysaccharides coupled with Fischer glycosylation and esterification reaction of D-mannuronic acid and L-guluronic acid with butanol proceeded efficiently at 130–135 °C in the presence of the readily biodegradable methane sulfonic acid (MSA) promoter to provide the corresponding *n*-butyl uronate intermediates. Subsequent transformation of these mono-mannuronate and guluronate derivatives into long-chain alkyl L-guluronamide and D-mannuronamide surfactant mixtures has been carried out without any isolation and purification step of the intermediates through a one-pot solvent-free aminolysis reaction with dodecyl or octadecyl amines at 65 °C. The compositions isolated in 40% overall yields are composed of six isomeric forms in the L-guluronic acid and D-mannuronic acid series (Scheme 27): two α,β -furanose forms and two α,β -pyranose forms for *N*-(*n*-alkyl)-*n*-butyl L-guluronamides; one α -furanose form and its α -pyranose isomer for *N*-(*n*-alkyl)-*n*-butyl D-mannuronamides. The 1,2-*cis*- β -D-mannofuranosiduronamide and 1,2-*cis*- β -D-mannopyranosiduronamide were not formed since the formation of 1,2-*cis*- β -mannosides is unfavorable thermodynamically as well as kinetically owing to steric hindrance of axial 2-hydroxy group. In addition, a process based on the neutralization of MSA prior to aminolysis reaction and the use of reduced amount of fatty amine was developed to manufacture surfactants directly from alginates without requiring the standard steps of solid-liquid separations while removing simply inorganic salts and other water-soluble products. These uronamide surfactant compositions were found to exhibit promising emulsifying properties compared to other sugar-based surfactant references. Interestingly, the composition resulting from the reaction of dodecyl amine without any final purification step engendered very high stabilities of O/W emulsions since the emulsions remained unchanged after several months at 20 °C. The study of their application in cosmetic formulations is under investigation.

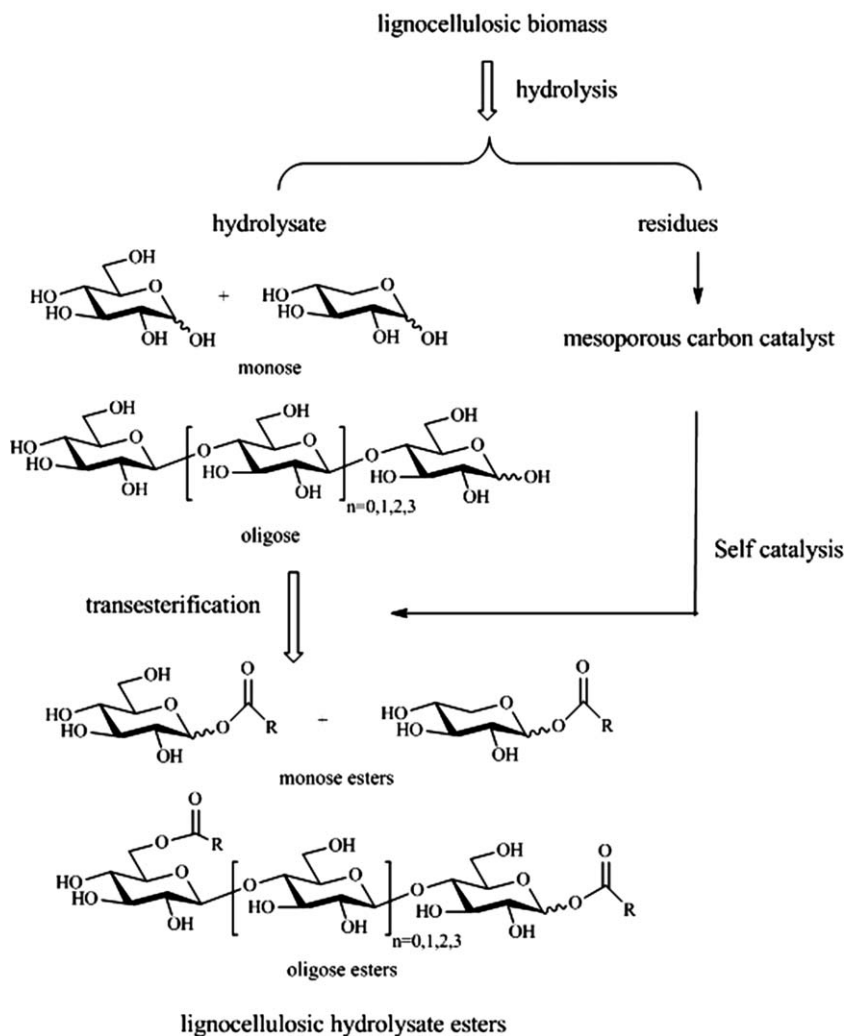
5.5 Monose and oligose esters from lignocellulose

To meet demands for environmental benign and food safety of surfactant industry, Wang *et al.*⁹⁵ proposed to transform lignocellulose derived from shrub willow into monose and oligose esters. For that purpose, a green process was developed based on the production of lignocellulosic hydrolysate used as substrate for a transesterification reaction with fatty acid esters (Scheme 28). Firstly, shrub willow was ground into powders that were then hydrolyzed in water under microwave irradiation for one hour. The hydrolysis step provided a clarified hydrolysate composed of monose (26.8 wt%) and oligose (72.2 wt%) in addition to a brown residue that contained lignin (68.5 wt%), cellulose (23.3 wt%) and others (8.2 wt%). The residue part was converted into a mesoporous carbon catalyst (K₂O/C) *via* impregnated pre-activating and high temperature activating process. Due to the immiscible phase between the non-purified hydrolysate and fatty acid esters (ethyl laurate, ethyl myristate, ethyl palmitate and ethyl stearate), ultrasonic assistant was introduced to promote the heterogeneous transesterification. With the mixture of fatty acid ester



Scheme 27 Conditions used for the transformation of alginate into L-gulonamide and D-mannuronamide surfactants.

(99.5 wt%) and hydrolysate solution (25.0 wt%), a light yellow emulsion was formed by magnetic stirring under ultrasonic assistant. After slow heating to 60 °C, most of water in the emulsion was removed under vacuum condition, followed by the addition of K_2O/C catalyst. The reaction was then conducted in the emulsion at a temperature higher than 100 °C until no more ethanol could be removed under reduced pressure. These reaction conditions led to the formation of lignocellulosic monose and oligose esters mainly under a monoester form (72.9% for monose esters and 83.9% for oligose esters) characterized by the presence of the lipophilic fatty acid ester at the anomeric position. A certain amount of di- and polyester was also detected in the mixtures, which could be attributed to the active primary hydroxyl group (C6-OH) of hexose unit.



Scheme 28 Green process for the preparation of novel glycosyl surfactant from lignocellulosic biomass.

Significantly, these lignocellulosic hydrolysate esters exhibited good surfactant capacity as glycosyl surfactants with the surface tension of 33.0 mN m^{-1} , CMC value of 0.514 g L^{-1} , emulsifying value of 11.63 min and HLB number of 13.9 which is similar to that of sucrose laurate esters and AlkylPolyGlycosides¹²¹⁴ and more better than Span-80, used as sugar surfactant references. In view of the promising economic and environmental performance, the authors concluded that these lignocellulosic-derived surfactant compositions represent remarkable novel emulsifiers for applied in commercial O/W glycosyl surfactant.

6 Conclusion

Sugar-based surfactants are becoming of great importance because they have a range of beneficial physical and performance properties, including high surface and interfacial activity, very rapid biodegradability, low human and animal toxicity, effective emulsifying properties, and surface interactions. The enormous variability of the glyco-repertoire available for developing innovative glycosyl surfactants may open new opportunities in the surfactant market, as alternatives to petroleum-based products. Recent research works have studied the interest in designing and characterizing novel sugar surfactants either functionalized with ionic groups or possessing original linkers between the hydrophilic and hydrophobic domains, or even based on new carbohydrate motifs. Another promising approach consisted in proposing more direct transformation routes of polysaccharide sources into surfactants without requiring the isolation and the purification of reaction intermediates. The objective was to find conditions that prevent the prior step of polysaccharide depolymerization and therefore their direct use in a one-pot process. Further investigations could be considered in the future, in particular the opportunity to propose eco-friendly conditions which satisfy in a more effective way the principles of green/blue chemistry, for example organic solvent-free reactions, without production of waste and using biodegradable reagents. Another challenging feature may explore more intensely the utilization of harvest saccharides from biomass processing (and related secondary streams) or from agricultural/food by-products, with the aim to adjust the discovery of new sugar surfactant to economic constraints encountered in the surfactant market.

Abbreviations

ADMP	2-azido-1,3-dimethylimidazolium hexafluorophosphate
AES	alcohol ether sulfates
AG	alkyl glycosides
AG-EC	alkyl monoglucoside citric monoesters
AOS	alpha olefin sulfonates
APG	alkyl polyglycosides
ATG	alkyl triazole glycosides
CMC	critical micelle concentration
CTAB	cetyl trimethylammonium bromide

CuAAC	copper(I)-catalyzed azide-alkyne cycloaddition
EOR	enhanced oil recovery
GB	glycine betaine
HLB	hydrophilic-lipophilic balance
HWE	Horner–Wadsworth–Emmons
LAS	linear alkylbenzene sulfonates
MSA	methane sulfonic acid
O/W	oil/water
SDS	sodium dodecyl sulfate

References

- 1 M. J. Rosen, *Surfactants and Interfacial Phenomena*, John Wiley & Sons, Inc., Hoboken, New Jersey, 2004.
- 2 *Surfactants from Renewable Resources*, ed. M. Kjellin and I. Johansson, John Wiley & Sons, Inc., Chichester, UK, 2010.
- 3 M. E. Abdel-Raoul, A. M. Abdul-Raheim and A. A. Abdel-Azim, *J. Surfactants Deterg.*, 2011, **14**, 113.
- 4 I. J. A. Baker, B. Matthews, H. Soares, I. Krodkiewska, D. N. Furlong, F. Grieser and C. J. Drummond, *J. Surfactants Deterg.*, 2000, **1**, 1.
- 5 *Sugar-Based Surfactants: Fundamentals and Applications*, ed. C. C. Ruiz, CRC Press, Taylor & Francis Group, Boca Raton, 2008.
- 6 P. S. Piispanen, *Synthesis and characterization of surfactants based on natural products*, Kungl Tekniska Högskolan, Stockholm, 2002.
- 7 S. George and K. Jayachandran, *Appl. Biochem. Biotechnol.*, 2009, **158**, 694.
- 8 R. Thavasi, S. Jayalakshmi, T. Balasubramanian and I. M. Banat, *World J. Microbiol. Biotechnol.*, 2008, **24**, 917.
- 9 S. Marinkovic, J. Le Bras, V. Nardello-Rataj, M. Agach and B. Estrine, *Int. J. Mol. Sci.*, 2012, **13**, 348.
- 10 T. Benvegnu, D. Plusquellec and L. Lemiègre, *Surfactants from Renewable Sources: Synthesis and Applications*, in *Monomers, Polymers and Composites from Renewable Resources*, ed. M. N. Belgacem and A. Gandini, Elsevier Limited, Amsterdam, 2008.
- 11 P. T. Anastas and J. C. Warner, *Green Chemistry: Theory and Practice*, Oxford University Press, New York, 1998.
- 12 P. Bernardi, D. Fornara, L. Paglino and T. Verzotti, Eucarol, APG series: a new generation of surfactants derived from glucose, in *The 1st Concise Surfactants Directory*, 1996, p. 15.
- 13 Anonymous, *Chem. Market Rep.*, 1996, **17**, 23.
- 14 A. Behler, H. Hensen and W. Seipel, *Proceedings of the 6th World Surfactant Congress, CESIO*, 2004, Berlin.
- 15 Z. Jin, J. Zhang, X. Yang and Y. Zhou, *J. Surfactants Deterg.*, 2016, **19**, 885.
- 16 T. Benvegnu and J. F. Sassi, *Oligomannuronates from Seaweeds as Renewable Sources for the Development of Green Surfactants*, in *Carbohydrates in Sustainable Development I: Renewable Products for Chemistry and Biotechnology, Book Series: Topics in Current Chemistry*, 2010, vol. 294, pp. 143–164.
- 17 L. H. Lin, Y. C. Lai, K. M. Chen and H. M. Chang, *Colloids Surf., A*, 2015, **485**, 118.
- 18 M. A. Aym and A. Al-L Hamad, *J. Surfactants Deterg.*, 2014, **17**, 1043.
- 19 T. H. Zhao, J. Y. Gu, W. F. Pu, Z. M. Dong and R. Liu, *RSC Adv.*, 2016, **6**, 70165.
- 20 X. F. Zheng and Q. Lian, *J. Dispersion Sci. Technol.*, 2015, **36**, 245.

- 21 *Cationic Surfactants: Analytical and Biological Evaluation*, ed. J. Cross and E. J. Singer, Surfactant Science Series, Marcel Dekker, 1994, vol. 53.
- 22 *Alkyl Polyglycosides: Technology, Properties and Applications*, ed. K. Hill, W. von Rybinski and G. Stoll, 1996, VCH, Weinheim.
- 23 D. Pérusse, J. P. Guégan, H. Rolland, J. Guilbot and T. Benvegnu, *Green Chem.*, 2016, **18**, 1664.
- 24 C. Gan, H. Wang, Z. Zhao and B. Yin, *J. Surfactants Deter.*, 2014, **17**, 1069.
- 25 V. A. Timoshchuk, *Russ. Chem. Rev.*, 1995, **64**, 675.
- 26 H. Lu, A. Drelich, M. Omri, I. Pezron, A. Wadouchi and G. Pourceau, *Molecules*, 2016, **21**, 1301.
- 27 C. Blecker, S. Piccicuto, G. Lognay, C. Deroanne, M. Marlier and M. Paquot, *J. Colloid Interface Sci.*, 2002, **247**, 424.
- 28 P. Laurent, H. Razafindralambo, B. Wathelet, C. Blecker, J. P. Wathelet and M. Paquot, *J. Surfactants Deterg.*, 2011, **14**, 51.
- 29 T. Benvegnu and J. F. Sassi, Carbohydrates in sustainable development, in *I. Renewable Resources for Chemistry and Biotechnology*, ed. A. P. Rauter, P. Vogel, Y. Queneau, 2010, vol. 294, pp. 143–164.
- 30 N. Sari-Chmayssem, F. Pessel, J. P. Guégan, S. Taha, H. Mawlawi and T. Benvegnu, *Green Chem.*, 2016, **18**, 6573.
- 31 M. Rose and R. Palkovits, *ChemSusChem*, 2012, **5**, 167.
- 32 Y. Zhu, M. Durand, V. Molinier and J. M. Aubry, *Green Chem.*, 2008, **10**, 532.
- 33 J. Lai, S. Bigot, M. Sauthier, V. Molinier, I. Suisse, Y. Castanet, J. M. Aubry and A. Mortreux, *ChemSusChem*, 2011, **4**, 1104.
- 34 A. Lavergne, L. Moity, V. Molinier and J. M. Aubry, *RSC Adv.*, 2013, **3**, 5997.
- 35 A. Lavergne, Y. Zhu, A. Pizzino, V. Molinier and J. M. Aubry, *J. Colloid Interface Sci.*, 2011, **360**, 645.
- 36 A. Lavergne, Y. Zhu, V. Molinier and J. M. Aubry, *Colloids Surf., A*, 2012, **404**, 56.
- 37 M. Ben Aldelkader, N. Azizi, M. Chemli, Y. Chevalier, O. Boyron and M. Majdoub, *Colloids Sur., A*, 2016, **492**, 1.
- 38 M. Burek, S. Waśkiewicz, I. Wandzik and K. Kamińska, *Chemik*, 2015, **8**, 469.
- 39 L. Schiefelbein, M. Keller, F. Weissmann, M. Lubert, F. Bracher and W. Frieß, *Eur. J. Pharm. Biopharm.*, 2010, **76**, 342.
- 40 V. V. Rostovtsev, L. G. Green, V. V. Fokin and K. B. Sharpless, *Angew. Chem., Int. Ed.*, 2002, **41**, 2596.
- 41 C. W. Tornøe, C. Christensen and M. Meldal, *J. Org. Chem.*, 2002, **67**, 3057.
- 42 J. Joosten, N. Tholen, F. El Maate, A. Brouwer, G. Esse, D. Rijkers, R. Liskamp and R. Pieters, *Eur. J. Org. Chem.*, 2005, 3182.
- 43 P. Appukkuttan, W. Dehaen, V. Fokin and E. Van Der Eycken, *Organic Lett.*, 2004, **6**, 4223.
- 44 G. Cravotto, V. Fokin, D. Garella, A. Binello, L. Boffa and A. Barge, *J. Comb. Chem.*, 2010, **12**, 13.
- 45 A. I. Mohammed, *Asian J. Chem.*, 2012, **24**, 5585.
- 46 F. Aliasghari Sani, T. Heidelberg, R. Hashim and Farhanullah, *Colloids Surf., B*, 2012, **97**, 196.
- 47 E. D. Oldham, S. Seelam, C. Lema, R. J. Aguilera, J. Fiegel, S. E. Rankin, B. L. Knutson and H.-J. Lehmler, *Carbohydr. Res.*, 2013, **379**, 68.
- 48 T. H. Ali, R. S. Duali Hussen and T. Heidelberg, *Colloids Surf., B*, 2014, **123**, 981.
- 49 T. H. Ali, H. A. Tajuddin, R. S. Duali Hussen and T. Heidelberg, *J. Surfactants Deterg.*, 2015, **18**, 881.
- 50 N. Han-Su, T. Heidelberg and A. A. Salman, *J. Dispersion Sci. Technol.*, 2016, **37**, 105.
- 51 E. D. Oldham, L. M. Nunes and A. Varela-Ramirez, *Chem. Central J.*, 2015, **9**, 3.

- 52 T. Schuster, S. Schellenberger, R. Friedrich, M. Klapper and K. Müllen, *J. Fluorine Chem.*, 2013, **154**, 30.
- 53 K. J. V. Paul and D. Loganathan, *Tetrahedron Lett.*, 2008, **49**, 6356.
- 54 D. Lim, M. A. Brimble, R. Kowalczyk, A. J. A. Watson and A. J. Fairbanks, *Angew. Chem., Int. Ed.*, 2014, **53**, 11907.
- 55 R. Ghosh, A. Chakraborty and S. Maiti, *Tetrahedron Lett.*, 2004, **45**, 9631.
- 56 F. M. Ibatullin and K. A. Shabalin, *Synth. Commun.*, 2000, **30**, 2819.
- 57 U. Aich and D. Loganathan, *J. Carbohydr. Chem.*, 2005, **24**, 1.
- 58 I. S. Shchelik and Y. L. Sebyakin, *Russ. J. Org. Chem.*, 2015, **51**, 1717.
- 59 A. I. Mohammed, Z. H. Abboud and A. H. O. Alghanimi, *Tetrahedron Lett.*, 2012, **53**, 5081.
- 60 A. A. Salman, M. Tabandeh, T. Heidelberg and R. S. Duali Hussen, *Carbohydr. Res.*, 2015, **406**, 41.
- 61 N. Karst and J.-C. Jacquinet, *J. Chem. Soc., Perkin Trans. 1*, 2000, **1**, 2709.
- 62 V. Neto, R. Granet, G. Mackenzie and P. Krausz, *J. Carbohydr. Chem.*, 2008, **24**, 231.
- 63 S. M. Salman and W. A. R. Abdul Razak, *Diyala J. Pure Sci.*, 2015, **11**, 13.
- 64 S. Hanessian, D. Ducharme, R. Masse and M. L. Campmau, *Carbohydr. Res.*, 1978, **63**, 265.
- 65 S. M. Salman, T. Heidelberg and H. A. B. Tajuddin, *Carbohydr. Res.*, 2013, **375**, 55.
- 66 R. Ramapanicker, R. Gupta, R. Megha and S. Chandrasekaran, *Int. J. Peptides*, 2011, **1**.
- 67 H. Oskarsson, M. Frankenberg, A. Annerling and K. Holmberg, *J. Surfactants Deterg.*, 2007, **10**, 41.
- 68 A. Wagenaar and J. B. F. N. Engberts, *Tetrahedron Lett.*, 2007, **63**, 10622.
- 69 H. S. Ewan, C. S. Muli, S. Touba, A. T. Bellinghiere, A. M. Veitschegger, T. B. Smith, W. L. Pistel, W. T. Jewell, R. K. Rowe, J. P. Hagen and H. Palandoken, *Tetrahedron Lett.*, 2014, **55**, 4962.
- 70 M. Capela, N. J. Mosey, L. Xing, R. Wang and A. Petitjean, *Chem. – Eur. J.*, 2011, **17**, 4598.
- 71 H. Palandoken, C. M. Bocian, M. R. McCombs and M. H. Nantz, *Tetrahedron Lett.*, 2005, **46**, 6667.
- 72 R. Kaplanek, R. Polak, O. Paleta, K. Kefurt, J. Moravcova, I. Krenova and M. Kodicek, *Carbohydr. Res.*, 2010, **345**, 1008.
- 73 O. S. Maldonado, R. Lucas, F. Comelles, M. J. Gonzales, J. L. Parra, I. Medina and J. C. Morales, *Tetrahedron*, 2011, **67**, 7268.
- 74 S. Liu, R. Sang, S. Hong, Y. Cai and H. Wang, *Langmuir*, 2013, **29**, 8511.
- 75 K. Lalitha, K. Muthusamy, Y. S. Prasad, P. K. Vemula and S. Nagarajan, *Carbohydr. Res.*, 2015, **402**, 158.
- 76 F. Rodrigues, Y. Canac and A. Lubineau, *Chem. Commun.*, 2000, 2049.
- 77 P. M. Foley, A. Phimpachanh, E. S. Beach, J. B. Zimmerman and P. T. Anastas, *Green Chem.*, 2011, **13**, 321.
- 78 A. Ranoux, L. Lemiègre, M. Benoit and T. Benvegno, *Eur. J. Org. Chem.*, 2010, 1314.
- 79 A. Ranoux, L. Lemiègre and T. Benvegno, *Sci. Chin. Chem.*, 2010, **53**, 1957.
- 80 L. Z. Szabo, D. J. Hanrahan, E. M. Jones, E. Martin and J. E. Pemberton, *Carbohydr. Res.*, 2016, **422**, 1.
- 81 M. R. Lefever, L. J. Szabo, B. Anglin, M. Ferracane, J. Hogan, L. Cooney and R. Polt, *Carbohydr. Res.*, 2012, **351**, 121.
- 82 D. Geetha and R. Tyagi, *Tenside, Surfactants, Deterg.*, 2012, **49**, 5.
- 83 P. M. Claesson and U. R. M. Kjellin, *Encycl. Surf. Colloid Sci.*, 2015, 6919.
- 84 M. J. L. Castro, J. Kovensky and A. F. Cirelli, *Langmuir*, 2002, **18**, 2477.

- 85 V. Neto, R. Granet and P. Krausz, *Tetrahedron*, 2010, **66**, 4633.
- 86 N. Villandier and A. Corma, *Chem. Commun.*, 2010, **46**, 4408.
- 87 N. Villander and A. Corma, *ChemSusChem*, 2011, **4**, 508.
- 88 M. J. Climent, A. Corma, S. Iborra and S. Martinez-Silvestre, *ChemSusChem*, 2016, **9**, 3492.
- 89 Q. N. Wang, F. L. Zhang, L. Li and T. Shen, *Fine Chem.*, 2007, **24**, 544.
- 90 M. Zou, J. Chen, Y. Wang, M. Li, C. Zhang and X. Yang, *J. Surfactants Deterg.*, 2016, **19**, 879.
- 91 E. E. Oliveira, A. E. Silva, J. Nagashima, M. C. S. Gomes, L. M. Aguiar, I. B. Marcelino, I. B. Araujo, M. P. Bayer, N. M. P. S. Ricardo, A. G. Oliveira and E. S. T. Agito, *Bioresour. Technol.*, 2010, **101**, 5402.
- 92 M. Sekine, T. Kimura, Y. Katayama, D. Takahashi and K. Toshima, *RSC Adv.*, 2013, **3**, 19756.
- 93 T. Benvegnu and J. F. Sassi, *Carbohydrates in Sustainable Development I. Renewable Resources for Chemistry and Biotechnology*, ed. A. P. Rauter, P. Vogel and Y. Queneau, 2010, **294**, 143.
- 94 N. Sari-Chmaysem, F. Pessel, J. P. Guegan, S. Taha, H. Mawlawi and T. Benvegnu, *Green Chem.*, 2016, **18**, 6585.
- 95 K. Wang, J. Jiang and X. Liang, *Ind. Crops Prod.*, 2016, **92**, 127.

Low molecular weight carbohydrate-based hydrogelators

Cédric Peyrot,^{a,b,c} Pierre Lafite,^a Loïc Lemiègre^{b,c} and Richard Daniellou^{*a}

DOI: 10.1039/9781788010641-00245

This review will focus on the ability of synthetic small (low-molecular weight) carbohydrates to form hydrogels and describe their key parameters.

1 Introduction

With polymeric substances either from natural, synthetic sources or even hybrids, the formation of gel in water (hydrogels) is well described in literature and believed to occur through chemical and/or physical cross-linking of polymeric chains.¹ A highly entangled three-dimensional network is therefore obtained, trapping water molecules presumably through a combination of surface forces and hydrogen bonding interactions. With such properties, hydrogels are becoming nowadays key-players for hi-tech applications in the biomedical, pharmaceutical, biotechnology, bioseparation, biosensor, agriculture, oil recovery and cosmetics fields.² Unfortunately, the development of polymeric hydrogels is hampered by the poor biodegradability of the available polymers and this has led to the search for low molecular weight gelators (LMWG).^{3,4} Accomplishments in such area were recently reported with amino acids, oligopeptide- or carbohydrate-based hydrogelators for examples. The latter based on sugar are of tremendous interest for the scientific community. Indeed, carbohydrates have been mainly studied for a long time for their chemistry and their role in the metabolism. They were only thought to constitute a source of energy or to play structural roles. Their biological functions have recently emerged and they are now considered to be central in most biological events, such as in cell-cell or cell-pathogen interactions.⁵ Bound to proteins, carbohydrates play a crucial role in expression and correct protein folding, thermal and proteolytic stability.⁶ In addition, they can also represent strong enzyme inhibitors and even powerful drugs.⁷ Finally their inherent biocompatibility make them promising molecules for material applications. Herein, this review will focus on the ability of synthetic small (low-molecular weight) carbohydrates to form hydrogels and describe their key parameters.

^aICOA UMR CNRS 7311, University of Orléans, rue de Chartres, BP6759, 45067 Orléans cedex 2, France. E-mail: Richard.daniellou@univ-orleans.fr

^bEcole Nationale Supérieure de Chimie de Rennes, CNRS, UMR 6226, 11 Allée de Beaulieu, CS 50837, 35708 Rennes Cedex 7, France

^cUniversité de Bretagne Loire, France

2 Monosaccharides

2.1 Protected carbohydrates

Two main types of fully or partially protected glycosyl conjugates induced successful gelation properties of water-based solutions. All structures are gathered in Fig. 1. The first series involves 4,6-benzylidene functionalisation. Kowalczyk *et al.* and Gonwald *et al.* prepared the methyl-4,6-*O*-(*p*-nitrobenzylidene)- α -D-glucopyranoside and mannopyranoside from corresponding methyl- α -D-glucopyranoside and *p*-nitro-phenylbenzaldehyde in the presence of sulfuric acid in dimethyl sulfoxide (10 days, r.t., 44%) (Fig. 1A).^{8,9} This compound gave opaque hydrogels at 1.5–5.0 wt% in water. Scanning Electron Microscopy¹⁰ analysis permitted to bring out fibrillar structures and junction points characteristic of a hydrogel network. The authors claimed that the nitro function offered a possible gelification of both polar and non-polar solvents.

A very similar compound was synthesized by Chen *et al.*, replacing the nitro group by an aldehyde function (Fig. 1A).¹¹ Desired product was obtained by the condensation reaction between methyl- α -D-glucopyranoside and terephthalaldehyde in presence of sulfuric acid in 4 days at room temperature. After column chromatography, the corresponding

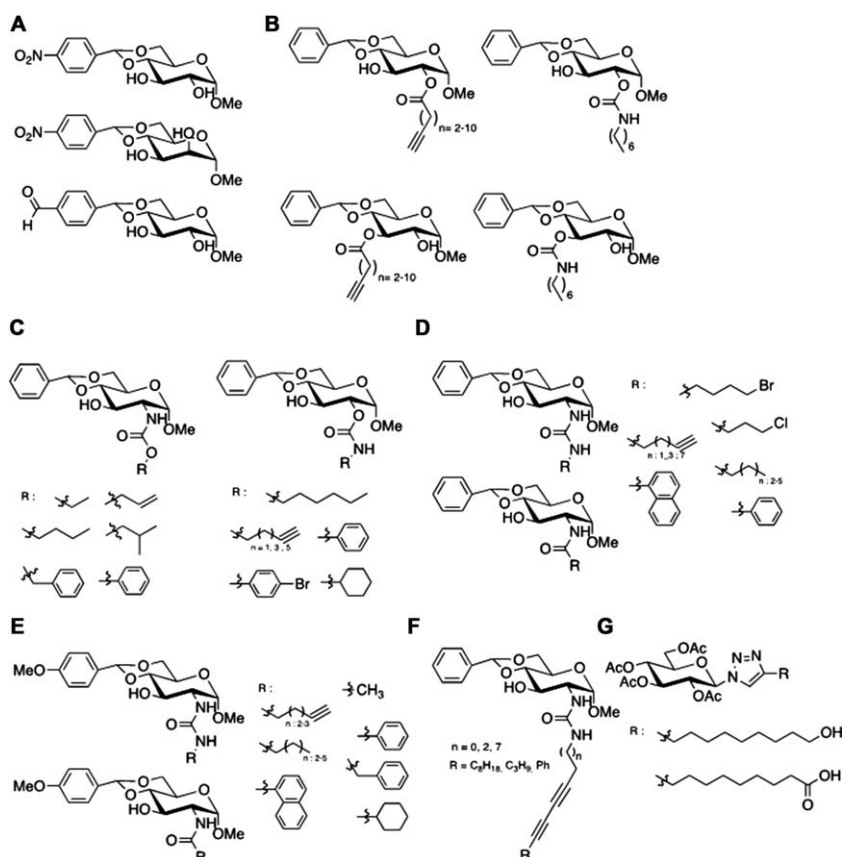


Fig. 1 Protected glycolipid hydrogelators.

benzylidene was obtained in 40% yield. This compound led to a white hydrogel and a critical gelation concentration of 0.8 wt% in water. The SEM images showed long fibres entangled each other to create a three-dimensional network. XRD studies provided a diffraction band corresponding to a d-spacing of 1.23 nm which corresponded to the length of the hydrogelator. Rheological experiments (G' , G'' measurements) revealed elastic more than viscous properties. The aldehyde group provided a cysteine responsive hydrogel by formation of a thiazolidine heterocycle with this specific amino acid. Thanks to the benzylidene acetal, the hydrogels were also sensitive to acidic conditions and started to disrupt at pH less than 1.2 (at 37 °C).

Starting from the same methyl- α -D-glucopyranoside, Wang *et al.* synthesized methyl-4,6-O-benzylidene- α -D-glucopyranoside followed by the esterification of the hydroxyls in position 2 or 3 with several acid chlorides.¹² Their synthetic strategy led to a large diversity of structures exhibiting interesting gelation abilities. Shorter alkyl chains bearing a terminal alkyne function ($n = 5-8$) gave hydrogelation properties at concentration ranging from 0.4 to 1.0 wt% in water (Fig. 1B). Interestingly, almost no differences were observed between the two regioisomers (esterification on position 2 or 3) meaning that the position of the alkyl chain had not a dramatic influence on the hydrogelation properties. SEM images showed characteristic networks of such hydrogels. In addition to ester linkage, the same research group synthesised carbamate analogues from glucosamine or glucose derivatives equipped with different type of fatty chains (linear, cyclic, bearing phenyl, acetylene or ethylene functions) in good yields (68–91%) following standard procedures (Fig. 1B–C). These compounds behaved as hydrogelators both in water and in EtOH/H₂O 1:2 mixture at concentrations ranging from 0.2 to 2 wt%.¹³ Corresponding hydrogels were characterised by SEM analysis. Glucoside derivatives led to uniform fibrillar structures with lengths > 100 μ m and diameters < 0.5 μ m. Fibres appeared in different type of objects. Depending on the molecule used, smooth sheets or ribbons were mainly found. Amino-glucoside carbamate derivatives self-assembled as large tubular structures composed by bundle of small fibres or cylindrical or rectangular planar sheets in water (lengths of 100 μ m, diameters > 0.1 μ m).

Goyal *et al.* described the synthesis and evaluation of two other series of hydrogelators also derived from benzylidene glycosides (Fig. 1D).¹⁴ Methyl- α -D-N-acetyl glucosamine was converted into the corresponding 4,6-benzylidene using benzylidene dimethyl acetal under acidic media then the acetyl group was removed and the released free amino group reacted finally with acid chlorides or isocyanates. Amide and urea derivatives were thus obtained in 45% yields (5 steps). The hydrogels, formed at concentration of 0.2 wt% in water, were characterised by optical microscopy and the SEM images were recorded on the corresponding xerogel, exhibiting long and uniform fibres. Tubular structures were also found only for the urea derivatives with diameters less than 1 μ m. This work demonstrated that the N–H present in the urea or amide functions is essential and provides a stabilizing hydrogen bond within the hydrogel network.

Based on the previous work Goyal *et al.* evaluated similar compounds but equipped with an additional methoxy group on the benzylidene acetal (Fig. 1E).¹⁵ The synthetic strategy was exactly the same leading to urea and amide compounds with 70% yield. Both ureas and amides provided gelation ability in H₂O/EtOH or H₂O/DMSO mixture (2:1; 0.13–1.0 wt%). Amide and urea gel morphologies were studied by optical microscopy demonstrating the presence of a characteristic uniform three-dimensional network. Fibre lengths were estimated to 40 to 100 μ m. The authors demonstrated that the extra *p*-methoxy group increased the solubility of the gelator and permitted to reduce the minimum gelation concentrations. Furthermore, the *p*-MeO derivative was found to be more sensitive to acidic pH. Indeed, the *p*-MeO-benzylidene ureas were less stable at pH 1–2 compared to the simpler benzylidene derivatives.

Following a similar synthetic approach Wang *et al.* introduced a polymerisable diacetylenic moiety on the fatty chain linked to a urea function (Fig. 1F).¹⁶ Due to solubility issues in water, all gels were obtained in EtOH/H₂O mixture (2:1). Upon UV irradiation, the polymerisation occurred within the gel network resulting to gel coloration. However, the morphologies of the gels before and after polymerisation were very similar as demonstrated by SEM images. However, this light-induced polymerisation increased significantly the melting temperatures of the gels by 10 to 20 °C. Rheological studies were performed on these gels and all storage modulus G' (400–2500 Pa) were higher than the loss modulus G'' (100–900 Pa) for all shearing frequencies. Thus, it confirmed the visco-elastic property of these types of gels. The gel ($n = 7$, $R = C_3H_9$; Fig. 1F) with the highest storage modulus is characterised by a position of the diacetylenic moiety at the end of the hydrophobic chain.

All together the gelation studies performed on the benzylidene series seem to demonstrate that the benzylidene ring is essential for gelation abilities. However, this first requirement can be modulated by addition of functions inducing hydrogen-bonding network such as amides or ureas. Among them, ureas which are known to afford strong hydrogen bond networks, induced the formation of stronger hydrogels.

The second series involving protected sugars was based on peracetylated glucosides bearing a triazole function at the anomeric position linked on the other side to a fatty chain.¹⁷ The first step of the synthetic strategy was the peracylation of D-glucose in the presence of acetic anhydride. HBr in acetic acid permitted to selectively activate the anomeric position, which can undergo an attack of sodium azide to obtain the glycopyranosyl azide (81% yield, 3 steps) (Fig. 1G). Then the fatty chain was introduced thanks to a CuAAC reaction with several alkynes. Among the library of glycolipids those bearing an extra hydroxyl or carboxyl groups gave hydrogelation abilities in DMSO/H₂O 1:1 or 1:2 at concentrations of 0.6–0.2 %wt. The soft elastic gels were characterized by storage modulus (G' : 2000–11000 Pa) higher than loss modulus (G'' : 400–3000 Pa). Optical microscopy displayed typical uniform fibrillar network.

2.2 Unprotected carbohydrates

Unprotected glycolipids offer other types of self-assembly based on the amphiphilic properties of such molecules. Usually, the hydrophilic/hydrophobic balance should be well adjusted to reach hydrogelation abilities. Indeed, the hydrophilic part provides solubility and the hydrophobic part gives self-assembling properties. Thus dissolution/aggregation balance is an important parameter to be considered.¹⁸ Within this series, *O*-glycosides were the most studied (Fig. 2) in addition to few other structures (Fig. 3).

Starting from peracetylated β -D-glucopyranose and cardanol derivatives John et al.¹⁹ synthesized three phenyl- β -D-glucopyranoside derivatives bearing different alkyl chains (Fig. 2A). Among them the saturated fatty chain (C15) and the unsaturated counterpart (C15:1) gave hydrogels in various solvent/H₂O 1:1 mixtures (MeOH, EtOH, Acetone, DMF, THF, DMSO). SEM studies revealed the presence of characteristic network self-assemblies. Transmission Electron Microscopy (TEM) analysis confirmed the formation of such network and permitted to observe twisted ribbons in the case of the saturated fatty chain. X-ray diffraction provided d spacing of 3.14 nm (C15) and 3.90 nm (C15:1) corresponding to a lamellar packing of the glycolipids. The double bond influenced this lamellar packing as demonstrated by a longer d spacing value. Indeed, it reduced the interdigitation of the alkyl chains.

Replacing the alkyl chain by a phenyl ring increased significantly the ability to form hydrogels in pure water (Fig. 2B).^{20,21} The corresponding gels were characterised by a three-dimensional network displaying long planar ribbons that changed to twisted chiral ribbons upon the addition of 1,4-dioxane (up to 60% in water). This tendency to switch to a chiral

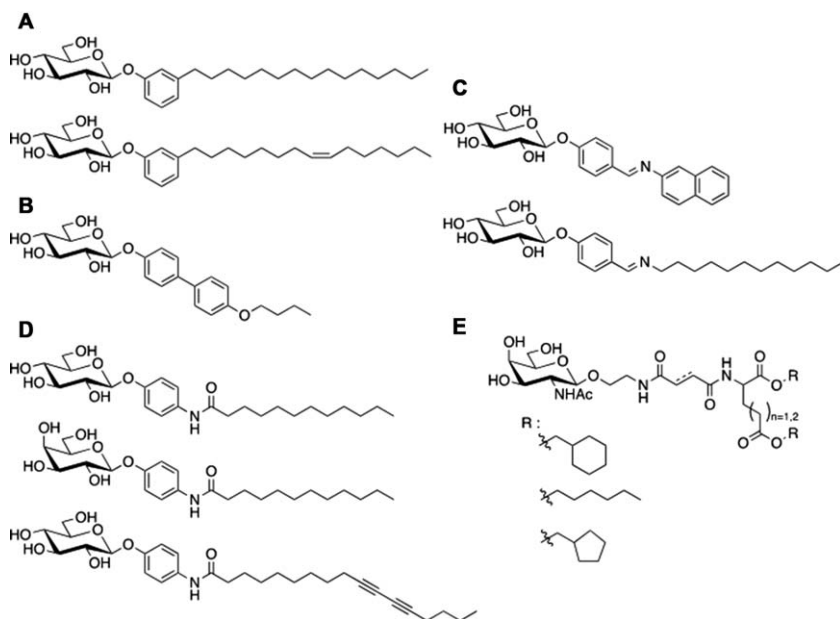


Fig. 2 Unprotected *O*-Glycolipid hydrogelators.

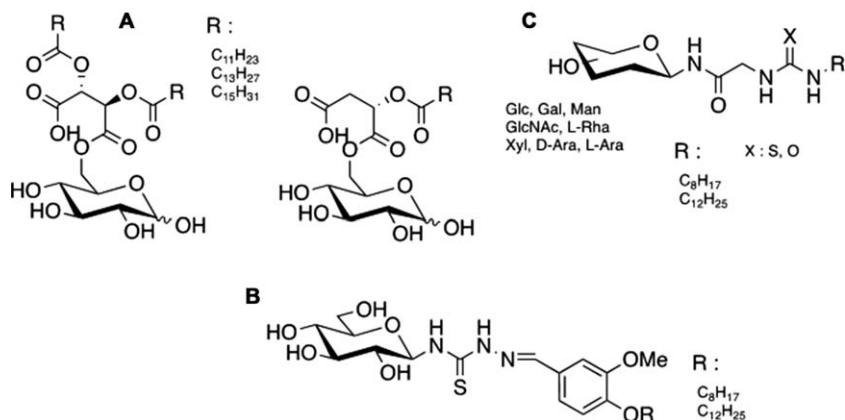


Fig. 3 Other unprotected hydrogelators.

self-assembly was also demonstrated by an increase of the signal in Circular Dichroism (CD) analysis. Wide-Angle X-ray Scattering (WAXS) was performed on xerogel and revealed sharp diffraction peaks, indicating the existence of crystallinity. The ratio of the calculated d-spacing corresponded to 1:1/2:1/3:1/4 which clearly implied a lamellar structure organisation. The higher diffraction peak corresponded to 2.4 nm, which could be related to the thickness of an interdigitated bilayer structure. This compound was obtained from peracetylated glucopyranoside which reacted with (4-bromophenoxyl)trimethylsilane to obtain the 4-bromophenyl β -D-tetraacetylglucoside. The later underwent a Suzuki cross-coupling reaction with 4-butoxyphenylboronic acid followed by a deprotection step to give the final 4(4-butoxyphenyl)phenyl- β -D-glucoside (43% yield, 3 steps).

Bao *et al.* developed phenyl-O-glucoside derivatives bearing an aldehyde function which easily reacted with alkyl or aromatic amines to form imine linkages.^{22,23} Peracetylated α -D-glucopyranosyl bromide reacted with *p*-hydroxybenzaldehyde which was put in reaction with dodecyl- or naphthyl-amine. After deprotection the two glycolipids were obtained in 51–53% yield respectively (Fig. 2C).

Both glycolipids permitted the gelation of $H_2O/EtOH$ 20 : 1 mixtures at concentrations around 2 w/v%. π -Stacking interactions and hydrogen bonding network between sugar polar heads were characterised by UV and IR analysis. XRD was performed on xerogels and showed the following d-spacing: 4.08, 2.02, 1.34 nm characteristic of a bilayer structure with interdigitated lipid chains. TEM and SEM images of the hydrogel obtained from the glycolipids bearing a linear alkyl chain showed the presence of tubular structures with tens of μm length, diameters of 75 nm and wall thickness of 20 nm. From these data, Bao and co-workers proposed a packing of 5 lipid bilayers which then form the nanotube. For the naphthyl derivatives TEM images demonstrated the presence of fibres with diameters between 30 to 80 nm.

Still within the phenyl glycoside series, Jung *et al.*²⁴ introduced an amide function, which added the possibility of a new hydrogen bond

between each amide groups (Fig. 2D). The synthetic strategy involved the reduction of the nitrophenyl- β -D-glucopyranoside (H_2 , Pd/C) and the subsequent functionalization of the so-formed amino group with different acyl chlorides in the presence of trimethylamine. After deprotection, the glycolipids were obtained in 34–55% yield (2 steps). In presence of trace amount of EtOH or MeOH (1.0 wt% in H_2O) these molecules led to hydrogels at low concentrations (0.1 wt%). SEM and TEM images demonstrated the formation of a three-dimensional network, characteristic fibres having 20–500 nm diameters and twisted ribbons. Hydrogen bonding network and π -stacking interactions were identified by NMR analysis. The amide hydrogen bond was also confirmed by FT-IR analysis with bands at 1645 and 1514 cm^{-1} . X-ray diffraction performed on the xerogel showed periodical diffraction peaks (2.90, 1.46, 0.97 nm), which were characteristic of a lamellar organization. The higher pic at 2.90 nm was compatible with the formation of interdigitated bilayered structures. The galactoside counterpart was also synthesised by the same strategy and this compound gave very similar results in terms of hydrogelation properties.²⁵ Hydrogels were obtained at concentrations starting from 0.2 wt% in water (In the presence of trace amount of MeOH or EtOH). SEM images of the hydrogels demonstrated the presence of linear fibres with diameters between 140 and 200 nm and lengths of several μm . TEM analysis showed twisted helical ribbons with 315 nm pitches and 85 nm wide. ^1H -NMR at variable temperatures highlighted the presence of self-assembled molecules interacting each other by π -stacking and intermolecular hydrogen bonds. The amide IR absorption band shifted to lower values (1639–1645 cm^{-1}), which confirmed that the carbonyl function was involved in a hydrogen bond. X-ray diffraction gave d-spacing corresponding to the ratio 1 : 1/2 : 1/3 characteristic of a lamellar structure with the same highest d-spacing value (2.90 nm). Almost no differences were observed between the gluco- and the galactoside derivatives. The same research group also evaluated the effect of longer chains (C17) and the presence of polymerisable diacetylene units (Fig. 2D).²⁶ These glycolipids were obtained following the same strategy. Hydrogelation abilities were obtained only for the unsaturated compound at 0.05 wt% in water. After irradiation at 540 nm, a red coloration of the hydrogel appeared which was related to the effective polymerisation of the diacetylene units. The authors performed comparative studies of the untouched and polymerised hydrogels. Circular dichroism permitted to conclude on a higher degree of organisation before irradiation than within the polymerised material. This hypothesis was supported by AFM images. Indeed, before irradiation helical ribbons were observed with diameter between 20 and 150 nm and length of several hundred μm . However, after irradiation AFM images showed disordered fibres devoid of helical ribbons. ^1H -NMR analysis demonstrated here again the importance of π -stacking interactions as a driving force for the formation of hydrogels. X-ray diagram before polymerisation revealed a single peak at 4.05 nm, which permitted to propose a molecular model based on an interdigitated bilayer packing. After polymerisation, the d-spacing slightly shifted to 4.20 nm which would correspond to a disordered bilayer packing.

Even if most of sugar based hydrogelators bear aromatic rings, few successful examples without aromatic unit are described in the literature. Among them Kiyonaka *et al.*²⁷ developed the synthesis of glutamate or aspartate derivatives, which linked a sugar on one side and two alkyl chains on the other side (Fig. 2E). The synthesis was carried out on solid phase from 2-azidoethyl β -*N*-acetyl-galactosamine, which permitted to reach good overall yields (20 to 85% depending on the alkyl chain). Hexyl, methyl cyclohexyl, methyl cyclopentyl chain gave very low hydrogelation concentrations (less than 0.25 wt%). Structural analysis was performed by TEM and showed fibrous network with diameters between 40 to 150 nm. IR spectra showed a peak corresponding to the amide carbonyl stretching at 1622 cm^{-1} . This value, shifted in the lower energy side, clearly demonstrated the presence of well-developed hydrogen bonding network.

The same research group introduced a double bond on the succinic moiety, this *E/Z* UV-sensitive double bond led to photo responsive hydrogels.^{28,29} Indeed, the *trans* configuration promoted hydrogelation whereas the *cis* isomer staid soluble in water. Comparative rheology, NMR and TEM/SEM analysis supported this photo gel-sol transition.

Unprotected sugars have also been involved in hydrogelator structures through other types of link instead of an *O*-glycosidic bond. In 2012, Nandi *et al.* described a new category of glucose based amphiphiles functionalised on the position 6 of the sugar letting unreacted the hydroxyl at the anomeric position (Fig. 3A).³⁰ Tartaric and malic anhydrides were acylated with different fatty acid chlorides. Then, these anhydrides reacted with free glucose on the most reactive hydroxyl (primary position) leading to the corresponding glycolipids in 20–30% yields. The authors obtained hydrogelation abilities for both tartaric (C11, C13, C15) and malic (C13, C15) derivatives. The malic-based glycolipids gave transparent hydrogels characterised by low Tgel (20–38 °C) whereas tartaric counterparts gave opaque hydrogels with higher Tgel (50–67 °C).

Some hydrogelator structures involved *N*-glycosidic bond between the sugar and the fatty chain parts (Fig. 3B). Bao *et al.*³¹ accessed to the per-*O*-acetyl- β -D-glucopyranosyl isothiocyanate from the bromine counterpart. Then reaction with hydrazine followed by condensation on benzaldehyde derivatives permitted to introduce an aromatic ring bearing different fatty chains. The final deprotection led to new glycolipids in 40–50% yield. These structures gave hydrogels at low concentrations ranging from 0.1 to 0.6 wt%. Fibrous networks were observed by TEM and presented micrometer lengths and diameters of 30 to 100 nm. IR spectra confirmed that the gelation was due to the presence of hydrogen bonding networks with the apparition of characteristic hydrogen bonded OH bands at 3303 and 3215 cm^{-1} . X-ray diffraction was carried out on the wet hydrogels, it permitted to give a model of molecular organisation within the fibres. Authors concluded in the presence of a bilayer structure with interdigitated fatty chains leading to strong hydrophobic interactions.

More recently, Mathiselvam and al.³² designed similar glycolipid structures but including a glycine spacer. Thiourea and urea were both envisaged and the sugar moiety was derived from D-glucose, D-galactose,

D-mannose, D-glucosamine, D-rhamnose, D-xylose, L- and D-arabinose (Fig. 3C). The synthesis scheme was based on the following steps; (1) Peracetylation of the sugar followed by the bromination of the anomeric position; (2) Azidolysis to obtain the per *O*-acetylated *N*-azido glycoside; (3) Reduction of the azide group and reaction with chloroacetyl anhydride; (4) nucleophilic substitution of the chlorine atom with sodium azide; (5) reduction of the azide group and introduction of the fatty chain by reaction with octyl or dodecyl isothiocyanate or isocyanate. All compounds were obtained in high overall yields (>70%). Most of these glycolipids gave hydrogelation properties at 0.5 wt% in water and the fibrillar network was characterised by SEM analyses. The urea series gave better results and higher stability of the hydrogels. Differences in the sugar nature were clearly demonstrated in the morphology of the fibres in SEM images. The influence of the polar head was also observed on thermal transition endotherms (DSC), for instance the glucoside derivative gave a lower sol–gel transition temperature (64 °C) compared to the galactoside derivative (70 °C). Thanks to SAXS patterns of corresponding xerogels, the authors proposed a model of organisation. Most of the proposed models were based on tail to tail packing with or without tilt depending on the nature of the sugar.

2.3 Open form

When designing hydrogelators, sugars are usually used in their cyclic form but they are also interesting in their opened form as it leads to other types of hydrogen bonding networks. Three different cases are possible to connect such opened sugars to any hydrophobic parts depending on the oxidation state of the anomeric position: (1) The sugar can react on the masked aldehyde function; (2) the sugar can be reduced; (3) the sugar can be oxidized to the corresponding reactive lactone.

Rajamalli *et al.*³³ described an easy access to dendron-based hydrogelators equipped with an opened glucoside unit (Fig. 4A). The poly(aryl ether) dendron was functionalised with a hydrazylamide which was nucleophilic enough to form an imine bond with the masked aldehyde of D-glucose (85% yield). This molecule required the presence of DMSO in water to provide hydrogels (DMSO/H₂O: 1:9 at 0.1 wt%). Hydrogen bonding networks involving the amide (1646 cm⁻¹) and the hydroxyl functions (3435 cm⁻¹) were highlighted by IR analysis. TEM images of the corresponding xerogels showed 3D network composed of entangled fibres presenting diameters around 100–150 nm and micrometer length. A lamellar arrangement was identified by XRD with a d-spacing of 3.14 nm. A dramatic morphological change was observed by addition of KOH on the hydrogel (until pH = 10), turning the lamellar organisation to spherical aggregates. The authors also described the possibility to incorporate graphene oxide into the hydrogel. This new material was found to have higher mechanical strength increasing the storage modulus (*G'*) from 34747 to 190071 Pa.

Li *et al.*³⁴ developed a hydrogelator based on D-sorbitol, which reacted with aldehydes to form the corresponding cyclic acetals (Fig. 4B). 3,4-dichlorobenzylidene acetal was synthesised following this strategy

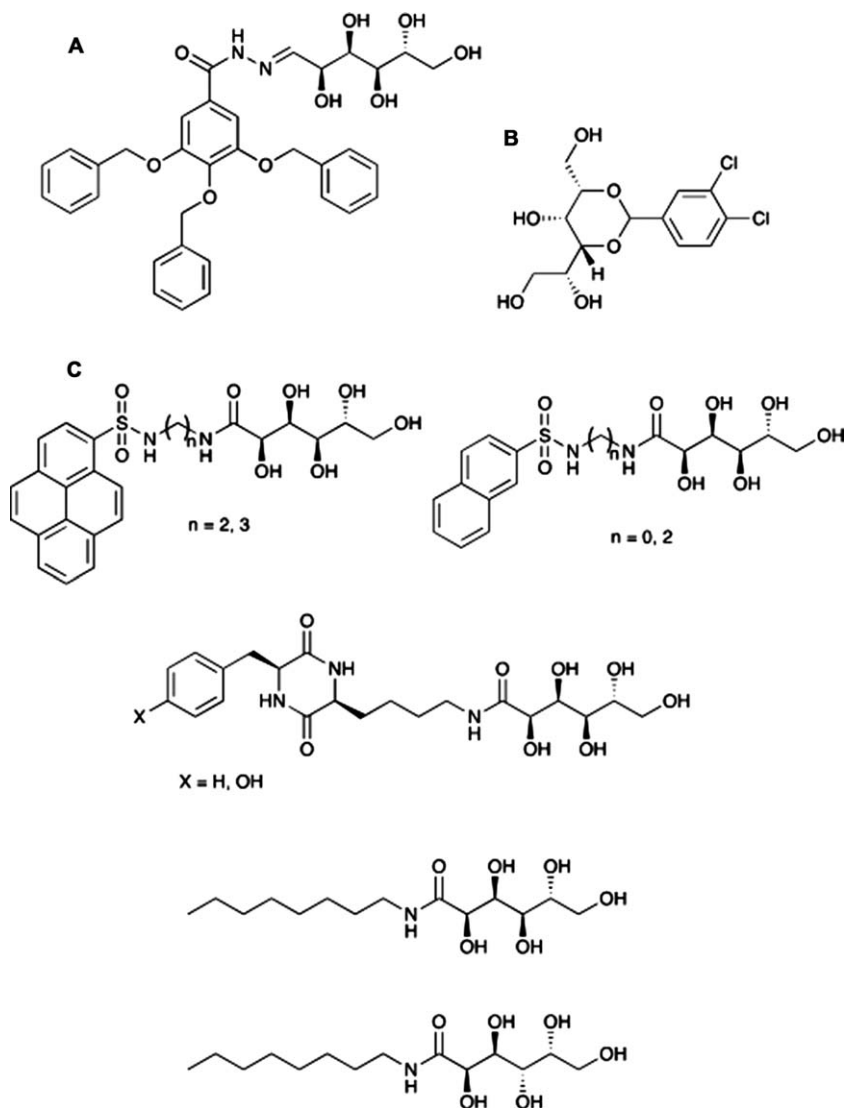


Fig. 4 Unprotected hydrogelators with sugar in an open form.

permitting a straight forward access to the hydrogelator structure. Hydrogelation appeared with different aqueous solutions (pure H_2O , 2% NaOH, 2% NaCl) at 1.0 wt%. The authors demonstrated that the presence of NaCl strengthened the hydrogel network. The hydrogels were also characterised by SEM analysis, which supported clearly this dramatic salt effect on the morphology of the fibrillar network. Combination of WAXS analysis and DFT calculations permitted to propose a model of organisation in which the molecules are interdigitated thanks to a strong π -stacking interaction between two aromatic rings.

The last example involved the reaction of D-gluconolactone with alkyl amine leading to the corresponding amides. Therefore, the main differences depend on the structure of the alkyl amine used (Fig. 4C).

Yan *et al.*³⁵ proposed to introduce pyrenyl or naphthyl sulphonamide groups as hydrophobic moieties. The length of the alkyl link between the sugar and the sulphonamide had a strong effect on the hydrogelation properties. Indeed, shorter linkages gave better hydrogelation abilities. Opaque hydrogels were obtained at concentrations less than 2.0 wt% in water. These hydrogels were characterised by SEM images demonstrating that the pyrenyl derivatives self-organised as helical supramolecular structures. XRD analysis permitted to propose totally different models of molecular self-organisation for the pyrenyl and the naphthyl derivatives. The pyrenyl one adopted a bilayered structure with an interdigitation of the sugar moiety. However, the naphthyl counterpart self-organized as a monolayer. Xie *et al.*³⁶ proposed on their side to introduce a cyclic dipeptide derived from lysine and phenylalanine or tyrosine aminoacids. After the synthesis of the dipeptide structures, the pendent amino function of lysine reacted with D-gluconolactone to offer the final hydrogelator structures. Both compounds derived from phenylalanine or from tyrosine gave hydrogelation properties at 2.0–3.0 wt% in water. Interestingly, the formation of the hydrogels required a stirring activation that accelerated the formation of aggregates from a supersaturated solution, which then evolved rapidly to the hydrogel state. TEM analysis of the xerogels demonstrated the formation of a three-dimensional network characterised by long fibres.

Capicciotti *et al.*³⁷ worked on the development of hydrogelators within the context of inhibition of ice recrystallization (Cryoprotectant) (Fig. 4C). Octylamine reacted with D-gluconolactone and D-galactonolactone leading to the corresponding *N*-octyl D-gluconamide and *N*-octyl D-galactonamide. Both compounds presented hydrogelation abilities at concentration of 1.0–0.5 wt% in water or in PBS. Interestingly, the glucoside derivative gave hydrogels with better ice recrystallization inhibition activity than the galactose derivative.

3 Disaccharides

The introduction of a disaccharide unit instead of a monosaccharide still provides molecular structures which can act as hydrogelators. However, it increases the hydrophilic part of the corresponding amphiphiles, which needs to be balanced by longer hydrophobic chains for instance.

3.1 Unprotected carbohydrates

Clemente *et al.*^{38–40} developed a series of disaccharide-based amphiphiles using a triazole spacer between an alkyl chain and a maltose, a lactose or a cellobiose unit (Fig. 5A). The synthetic scheme was based on a CuAAC reaction between the per-acetylated diglycosyl azide and a propargyl amide bearing the alkyl chain (C16). After deprotection the glycolipids were obtained in moderate overall yields (40–60%). These molecules offered hydrogelation abilities at concentration of 0.5–1.0 wt% in water. Interestingly, the presence of the triazole moiety is essential for achieving hydrogelation and π -stacking interaction in addition to other weak interactions were identified by variable temperature ¹H-NMR

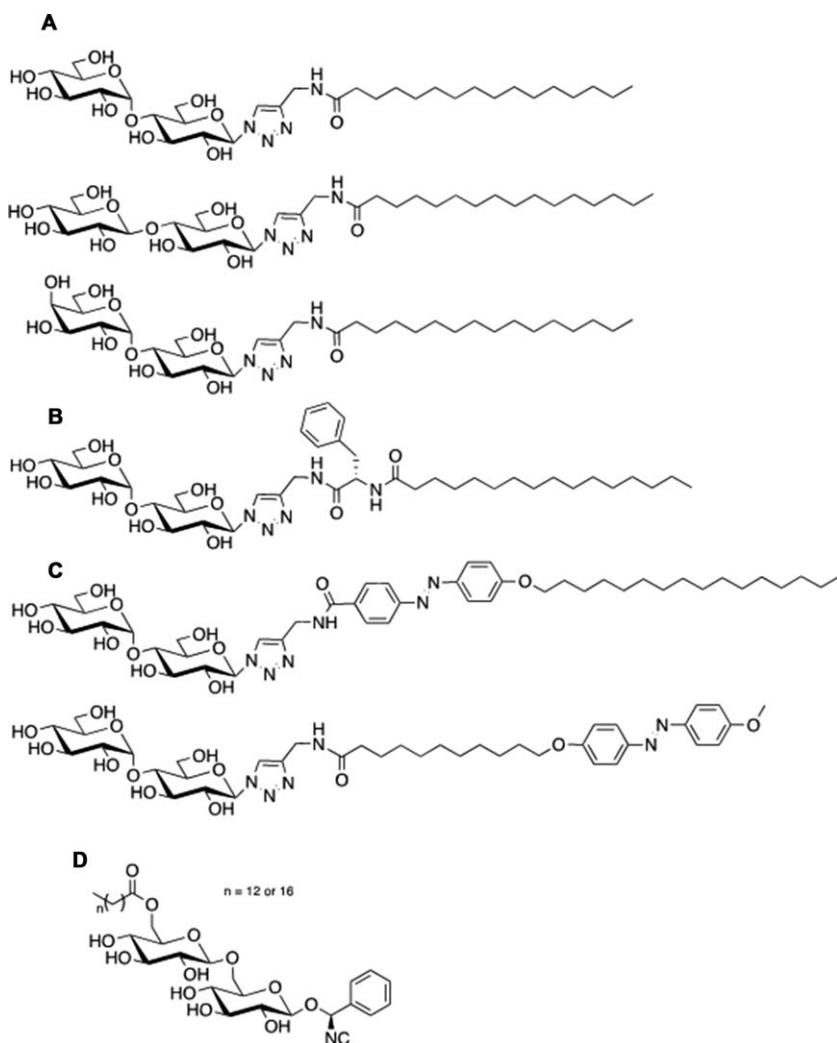


Fig. 5 Unprotected hydrogelators derived from disaccharides.

spectroscopy. SEM and TEM images of the corresponding xerogels revealed characteristic fibrillar network with an important proportion of twisted ribbons responsible for a supramolecular chirality. CD spectra were also recorded for which opposite sign were observed comparing lactosyl and cellobiosyl derivatives. It demonstrated the dramatic importance of the nature of the disaccharide on the supramolecular chirality transfer. A slight stereochemical difference in the structure of the sugar moiety (axial or equatorial orientation of a hydroxyl group) have a profound impact on the conformation of the molecules within the fibrillar network and finally on the supramolecular self-organisation.

The introduction of an extra aromatic ring thanks to the addition of a phenylalanine amino acid within the structure of the glycolipid also gave hydrogelation ability at concentration of 1.0 wt% (Fig. 5B). However, the

authors described an inhomogeneous gel compared to the compounds devoid of this phenyl ring.

Following the same synthetic scheme, Clemente *et al.*³⁸ could introduce an azobenzene group working as a light sensitive unit (Fig. 5C). The presence of the azobenzene on the structure of the hydrogelators did not disturb the self-organisation leading to hydrogels at concentration of 5 wt% in water or 1.5 wt% in a DMSO/water: 1 : 1 mixture. As expected, the azobenzene isomerisation upon irradiation at 365 nm modified the organisation of the molecules within the hydrogels. However, the effect was stronger when the azobenzene unit was placed in the middle of the structure between the sugar and the fatty chains parts than positioned at the end of the alkyl chain.

Amygdalin is a natural diglycoside found in plants. The structure of amygdalin is a cyanobenzyl 1->6-diglucoside which was selectively esterified on the 6' position with various fatty acid vinyl esters following an enzymatic step (novozyme 435) (Fig. 5D).⁴¹ When the fatty acids were derived from myristic or stearic acid, the corresponding esters gave hydrogelation properties in water at a concentration of 2.0 wt%. The hydrogels were characterised by SEM, the images showed a clear tridimensional network composed of helical ribbons of 50 nm wide and several micrometres length. X-ray diffraction permitted to determine a d-spacing of 4.0 nm which permitted to propose a model in which the fatty chains of two molecules are interdigitated within a bilayered organisation.

3.2 Open form

Bhattacharya *et al.*⁴² converted easily two disaccharides lactose and maltose into four glycolipids from *N*-hexadecyl glycosylamine (Fig. 6A). The first series was obtained by reduction with sodium borohydride leading to an open form of lactoside or maltoside derivatives. The second series was obtained by introduction of a second fatty chain through the reaction of the nitrogen atom with hexadecanoyl chloride. The best hydrogelation properties were obtained with the first series (one fatty chain) but only in the presence of a cosolvent, which could be an alcohol, DMF or DMSO for instance.

After oxidation of disaccharides such as lactose, maltose and cellobiose into the corresponding lactone, Ogawa *et al.*⁴³ envisaged to open the lactone ring with the nitrogen atom of a glycine derivative equipped with an azobenzene unit (Fig. 6B). This two-step synthesis yielded the amphiphiles in 70 to 80%. Hydrogels were formed at a concentration of 0.3 wt% in water and were visualised by low vacuum SEM. The corresponding images showed fibrillar structures. The driving force of gel formation was found to be related to π -stacking interactions between azobenzene rings. CD-spectra measurements also demonstrated the presence of chiral aggregates in the gel state for lactose and maltose derivatives but with opposite signs. However, cellobiose counterpart did not provide such chiral self-organisation. Interestingly, the presence of the azobenzene unit provided an impressive photoresponsive effect upon

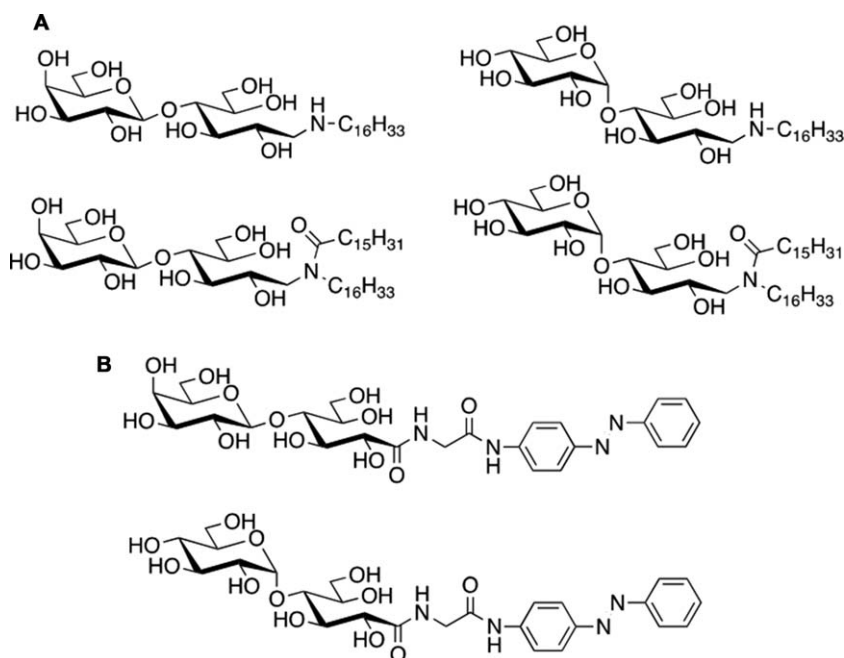


Fig. 6 Unprotected hydrogelators derived from open disaccharides.

irradiation at 365 nm. Indeed, gel-sol transitions were observed upon irradiation and gel state was recovered after leaving the sample at r. t. for 12h.

4 Bipolar hydrogelators

Bipolar lipids or bolamphiphiles are characterised by a unique hydrophobic domain bearing two hydrophilic units at both ends.⁴⁴ Therefore this situation corresponds to the creation a covalent linkage between the alkyl chains of two monopolar lipids. Few examples of glycolipid-based bolaamphiphiles have been described as good hydrogelators. In addition to their work on monopolar glycolipids, Jung *et al.*⁴⁵ also investigated the case of corresponding bipolar counterparts (Fig. 7A). A very similar synthetic strategy was applied, leading to the glucoside and galactoside derivatives in moderate yields (30%, 2 steps). Transparent hydrogels were obtained in pure water at very low concentrations (0.05 wt%). The comparison of CD spectra of the hydrogels obtained from the monopolar and the bipolar glycolipids were very informative. Indeed, the higher intensity recorded for the bipolar lipids indicated the presence of a highly ordered chiral structure for this later. The XRD analysis ($d = 3.58$ nm) confirmed the stretched conformation of the bipolar lipid within a monolayered self-organisation. The same research groups described also the synthesis and the evaluation of unsymmetrical bipolar glycolipids (Fig. 7B).⁴⁶ The synthesis followed again the same strategy and provided bipolar lipids bearing one sugar unit on one end (β -glucoside) and a carboxylic acid, an

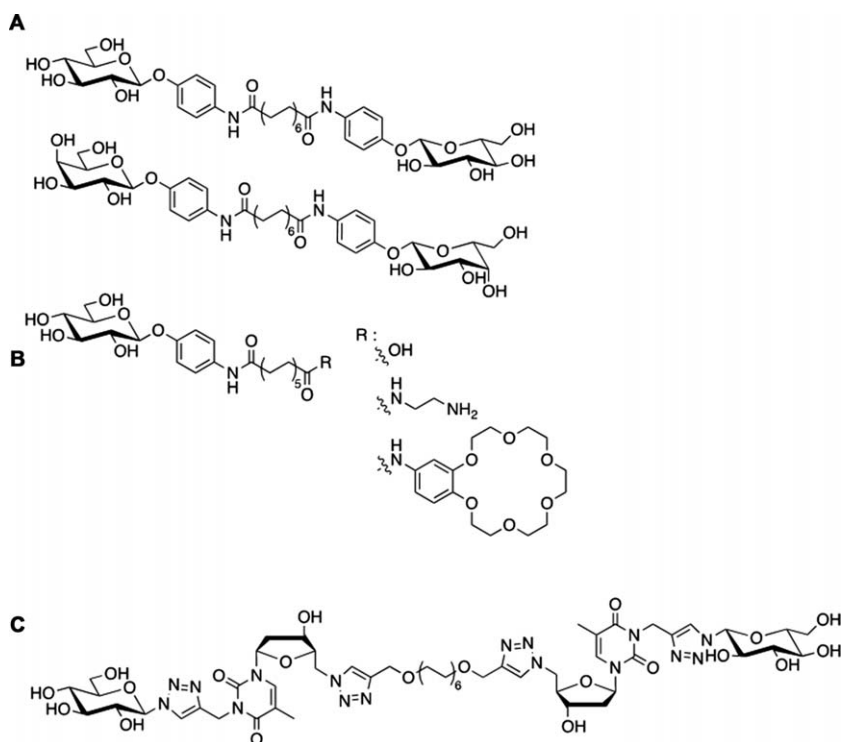


Fig. 7 Bolaamphiphile hydrogelators.

amine or a crown ether moiety on the other side. Hydrogels were obtained at 5.0 wt% concentrations. The glycolipids bearing a carboxylic acid or an amine function led to planar fibrillar structures visible by TEM. However, the crown ether equipped glycolipid led to helical fibres and strong CD spectra intensities. Interestingly the crown ether could accept several cationic guests such as ammonium, silver or potassium salts which induced dramatic structure stabilization or modification. This type of hydrogels was applied to the sol–gel polymerisation of TEOS (tetraethyl orthosilicate). Latxague *et al.*⁴⁷ developed the synthesis of another bipolar glycolipid through a double CuAAC reaction. Their strategy led to the successive bisfunctionalisation of dodecanediol with two nucleosidic units followed by two glucoside residues. The corresponding symmetrical glucosyl-nucleoside bipolar lipid led to a particularly effective hydrogelation of water at 1% w/v. TEM images showed a fibrillar network characterised by a 6–9 nm width. The storage modulus G' was found to be much higher (30225 Pa) than for the monopolar counterpart (1749 Pa). The authors explained that the stability of this hydrogel was mainly due to the several possible π -stacking interactions between the triazole moieties. The thixotropicity of this hydrogel was demonstrated by rheological experiments which permitted to define this material as a self-healing hydrogel. Finally, this hydrogel was applied to the culture of isolated stem cells and this soft material can be envisaged in regenerative medicine and tissue engineering applications.

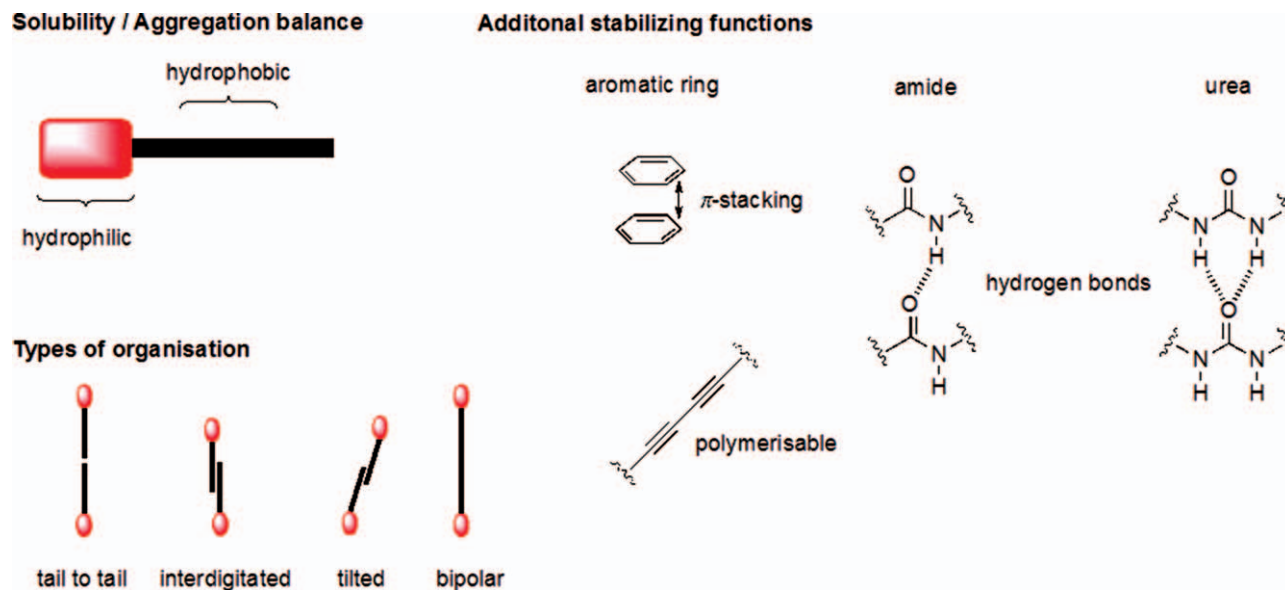


Fig. 8 Structure requirements and types of organization.

5 Rationalisation and structure/property relationships

The main requirement for a sugar-based hydrogelator is first a right solubility/aggregation balance. Therefore, amphiphilic molecules are usually good hydrogelator candidates (Fig. 8). Furthermore, the stability of the hydrogel can be enhanced by adding several other functions to the original molecular structure. Indeed, aromatic rings offering π -stacking interaction or amide/urea function providing strong hydrogen bonding network are often used in the design of new hydrogelators. Within this context, sugar derivatives usually provide a sufficient water solubility in addition to the possible formation of intermolecular hydrogen bonds. The diversity of sugar structures permits to fine tune the hydrophilicity, galactosides are less hydrophilic than glucosides for instance. Also, the sugar structure is responsible for the stereochemical control of the self-assembly leading usually to chiral supramolecular organisations such as helical or twisted structures.

The hydrophobic part which usually includes linear alkyl chains permits the self-assembly of the molecules within aqueous solutions. For monopolar glycolipids, tail to tail arrangement but mainly interdigitation of the alkyl chains are involved. This interdigitation can be tuned by the introduction of unsaturated fatty chains (especially *cis* double bonds). Thanks to their specific structures, the bipolar lipids behave as mono-layered structures with length corresponding more or less to the length of a unique stretched molecule. In addition, UV-polymerisable functions such as diacetylenic moiety provide a chemically stabilisation of the so-formed hydrogels.

6 Conclusion

Carbohydrate-based polymers and small synthetic molecules constitute a class of molecules that can efficiently provide hydrogels under specific conditions. Polysaccharides have been historically used for such application and they have the main advantage of coming from inexpensive renewable resources. In parallel, there is a need for carbohydrate-based low-molecular weight hydrogelators with highly well defined structures. Even if some requirements have been identified, the precise rules that govern their gelifying properties are not totally understood. In parallel, the development of enzymatic means of synthesis of such molecules will give the opportunity to the glycoscience community to increase the molecular diversity and to obtain more biocompatible and biodegradable carbohydrate hydrogelators. Given the role of sugars in so many crucial biological aspects, such research will pave the way for the development of smart biomaterials, combining multiple properties going from drug delivery to enzyme inhibition and lectin recognition for examples.

References

- 1 S. Bhattacharya, U. Maitra, S. Mukhopadhyay and A. Srivastava, in *Molecular Gels: Materials with Self-Assembled Fibrillar Networks*, ed. R. G. Weiss and

- P. Terech, Springer, Netherlands, Dordrecht, 2006, DOI: 10.1007/1-4020-3689-2_18, p. 613.
- 2 F. Ullah, M. B. H. Othman, F. Javed, Z. Ahmad and H. M. Akil, *Mater. Sci. Eng. C*, 2015, **57**, 414.
- 3 R. G. Weiss, *J. Am. Chem. Soc.*, 2014, **136**, 7519.
- 4 X. Du, J. Zhou, J. Shi and B. Xu, *Chem. Rev.*, 2015, **115**, 13165.
- 5 A. Varki, *Glycobiology*, 1993, **3**, 97.
- 6 P. Lafite and R. Daniellou, *Nat. Prod. Rep.*, 2012, **29**, 729.
- 7 B. Ernst and J. L. Magnani, *Nat. Rev. Drug. Discovery*, 2009, **8**, 661.
- 8 J. Kowalczyk, S. Jarosz and J. Tritt-Goc, *Tetrahedron*, 2009, **65**, 9801–9806.
- 9 O. Gronwald and S. Shinkai, *J. Chem. Soc., Perkin Trans. 2*, 2001, 1933–1937, DOI: 10.1039/B104366H.
- 10 I. Van Nieuwenhove, A. Salamon, K. Peters, G.-J. Graulus, J. C. Martins, D. Frankel, K. Kersemans, F. De Vos, S. Van Vlierberghe and P. Dubrue, *Carbohydr. Polym.*, 2016, **152**, 129.
- 11 Q. Chen, Y. Lv, D. Zhang, G. Zhang, C. Liu and D. Zhu, *Langmuir*, 2010, **26**, 3165–3168.
- 12 G. Wang, S. Cheuk, K. Williams, V. Sharma, L. Dakessian and Z. Thorton, *Carbohydr. Res.*, 2006, **341**, 705–716.
- 13 G. Wang, S. Cheuk, H. Yang, N. Goyal, P. V. N. Reddy and B. Hopkinson, *Langmuir*, 2009, **25**, 8696–8705.
- 14 N. Goyal, S. Cheuk and G. Wang, *Tetrahedron*, 2010, **66**, 5962–5971.
- 15 N. Goyal, H. P. R. Mangunuru, B. Parikh, S. Shrestha and G. Wang, *Beilstein J. Org. Chem.*, 2014, **10**(3111–3121), 3111.
- 16 G. Wang, N. Goyal, H. P. R. Mangunuru, H. Yang, S. Cheuk and P. V. N. Reddy, *J. Org. Chem.*, 2015, **80**, 733–743.
- 17 H. P. R. Mangunuru, J. R. Yerabolu, D. Liu and G. Wang, *Tetrahedron Lett.*, 2015, **56**, 82–85.
- 18 N. Yan, Z. Xu, K. K. Diehn, S. R. Raghavan, Y. Fang and R. G. Weiss, *J. Am. Chem. Soc.*, 2013, **135**, 8989.
- 19 G. John, J. H. Jung, M. Masuda and T. Shimizu, *Langmuir*, 2004, **20**, 2060–2065.
- 20 J. Cui, J. Zheng, W. Qiao and X. Wan, *J. Colloid Interface Sci.*, 2008, **326**, 267–274.
- 21 J. Cui, Z. Shen and X. Wan, *Langmuir*, 2010, **26**, 97–103.
- 22 C. Bao, R. Lu, M. Jin, P. Xue, C. Tan, Y. Zhao and G. Liu, *Carbohydr. Res.*, 2004, **339**, 1311–1316.
- 23 C. Bao, R. Lu, P. Xue, M. Jin, C. Tan, G. Liu and Y. Zhao, *J. Nanosci. Nanotechnol.*, 2006, **6**, 807–812.
- 24 J. H. Jung, G. John, M. Masuda, K. Yoshida, S. Shinkai and T. Shimizu, *Langmuir*, 2001, **17**, 7229–7232.
- 25 J. H. Jung, S. Shinkai and T. Shimizu, *Chem. – Eur. J.*, 2002, **8**, 2684–2690.
- 26 J. H. Jung, J. A. Rim, W. S. Han, S. J. Lee, Y. J. Lee, E. J. Cho, J. S. Kim, Q. Ji and T. Shimizu, *Org. Biomol. Chem.*, 2006, **4**, 2033–2038.
- 27 S. Kiyonaka, S. Shinkai and I. Hamachi, *Chemistry*, 2003, **9**, 976–983.
- 28 S. Matsumoto, S. Yamaguchi, S. Ueno, H. Komatsu, M. Ikeda, K. Ishizuka, Y. Iko, K. V. Tabata, H. Aoki, S. Ito, H. Noji and I. Hamachi, *Chem. – Eur. J.*, 2008, **14**, 3977–3986.
- 29 S. Matsumoto, S. Yamaguchi, A. Wada, T. Matsui, M. Ikeda and I. Hamachi, *Chem. Commun.*, 2008, 1545–1547, DOI: 10.1039/b719004b.
- 30 S. Nandi, H.-J. Altenbach, B. Jakob, K. Lange, R. Ihizane, M. P. Schneider, U. Guen and A. Mayer, *Org. Lett.*, 2012, **14**, 3826–3829.

- 31 C. Bao, R. Lu, M. Jin, P. Xue, C. Tan, Y. Zhao and G. Liu, *J. Nanosci. Nanotechnol.*, 2004, **4**, 1045–1051.
- 32 M. Mathiselvam, D. Loganathan and B. Varghese, *RSC Adv.*, 2013, **3**, 14528–14542.
- 33 P. Rajamalli, P. S. Sheet and E. Prasad, *Chem. Commun.*, 2013, **49**, 6758–6760.
- 34 J. Li, K. Fan, L. Niu, Y. Li and J. Song, *J. Phys. Chem. B*, 2013, **117**, 5989–5995.
- 35 N. Yan, Z. Xu, K. K. Diehn, S. R. Raghavan, Y. Fang and R. G. Weiss, *Langmuir*, 2013, **29**, 793–805.
- 36 Z. Xie, A. Zhang, L. Ye, X. Wang and Z.-G. Feng, *J. Mater. Chem.*, 2009, **19**, 6100–6102.
- 37 C. J. Capicciotti, M. Leclere, F. A. Perras, D. L. Bryce, H. Paulin, J. Harden, Y. Liu and R. N. Ben, *Chem. Sci.*, 2012, **3**, 1408–1416.
- 38 M. J. Clemente, R. M. Tejedor, P. Romero, J. Fitremann and L. Oriol, *RSC Adv.*, 2012, **2**, 11419–11431.
- 39 M. J. Clemente, J. Fitremann, M. Mauzac, J. L. Serrano and L. Oriol, *Langmuir*, 2011, **27**, 15236–15247.
- 40 M. J. Clemente, P. Romero, J. L. Serrano, J. Fitremann and L. Oriol, *Chem. Mater.*, 2012, **24**, 3847–3858.
- 41 P. K. Vemula, J. Li and G. John, *J. Am. Chem. Soc.*, 2006, **128**, 8932–8938.
- 42 S. Bhattacharya and S. N. G. Acharya, *Chem. Mater.*, 1999, **11**, 3504–3511.
- 43 Y. Ogawa, C. Yoshiyama and T. Kitaoka, *Langmuir*, 2012, **28**, 4404–4412.
- 44 J.-H. Fuhrhop and T. Wang, *Chem. Rev.*, 2004, **104**, 2901.
- 45 J. H. Jung, S. Shinkai and T. Shimizu, *Chem. – Eur. J.*, 2002, **8**, 2684.
- 46 H. Takamura, Y. Kadonaga, Y. Yamano, C. Han, I. Kadota and D. Uemura, *Tetrahedron*, 2009, **65**, 7449.
- 47 L. Latxague, M. A. Ramin, A. Appavoo, P. Berto, M. Maisani, C. Ehret, O. Chassande and P. Barthélémy, *Angew. Chem., Int. Ed.*, 2015, **54**, 4517.

