TOPICS IN CURRENT CHEMISTRY 267

Volume Editor V. Wittmann

Glycopeptides and Glycoproteins

Synthesis, Structure, and Application

Springer

267
Topics in Current Chemistry

Topics in Current Chemistry
Recently Published and Forthcoming Volumes

Sulfur-Mediated Rearrangements II
Volume Editor: Schaumann, E.
Vol. 275, 2007

Sulfur-Mediated Rearrangements I
Volume Editor: Schaumann, E.
Vol. 274, 2007

Bioactive Conformation II
Volume Editor: Peters, T.
Vol. 273, 2007

Bioactive Conformation I
Volume Editor: Peters, T.
Vol. 272, 2007

Biomineralization II
Mineralization Using Synthetic Polymers and Templates
Volume Editor: Naka, K.
Vol. 271, 2007

Biomineralization I
Crystallization and Self-Organization Process
Volume Editor: Naka, K.
Vol. 270, 2007

Novel Optical Resolution Technologies
Volume Editors:
Sakai, K., Hirayama, N., Tamura, R.
Vol. 269, 2007

Atomistic Approaches in Modern Biology
From Quantum Chemistry
to Molecular Simulations
Volume Editor: Reiher, M.
Vol. 268, 2006

Glycopeptides and Glycoproteins
Synthesis, Structure, and Application
Volume Editor: Wittmann, V.
Vol. 267, 2006

Microwave Methods in Organic Synthesis
Volume Editors: Larhed, M., Olofsson, K.
Vol. 266, 2006

Supramolecular Chirality
Volume Editors: Crego-Calama, M.,
Reinhoudt, D. N.
Vol. 265, 2006

Radicals in Synthesis II
Complex Molecules
Volume Editor: Gansäuer, A.
Vol. 264, 2006

Radicals in Synthesis I
Methods and Mechanisms
Volume Editor: Gansäuer, A.
Vol. 263, 2006

Molecular Machines
Volume Editor: Kelly, T. R.
Vol. 262, 2006

Immobilisation of DNA on Chips II
Volume Editor: Wittmann, C.
Vol. 261, 2005

Immobilisation of DNA on Chips I
Volume Editor: Wittmann, C.
Vol. 260, 2005

Prebiotic Chemistry
From Simple Amphiphiles to Protocell Models
Volume Editor: Walde, P.
Vol. 259, 2005

Supramolecular Dye Chemistry
Volume Editor: Würthner, F.
Vol. 258, 2005

Molecular Wires
From Design to Properties
Volume Editor: De Cola, L.
Vol. 257, 2005

Glycopeptides and Glycoproteins

Synthesis, Structure, and Application

Volume Editor: Valentin Wittmann

With contributions by
S. J. Danishefsky · X. Geng · C. Haase · B. Meyer
H. Möller · S. Schoof · O. Seitz · D. Specker · R. D. Süssmuth
D. A. Thayer · J. D. Warren · V. Wittmann · F. Wolter · C.-H. Wong

Springer

The series *Topics in Current Chemistry* presents critical reviews of the present and future trends in modern chemical research. The scope of coverage includes all areas of chemical science including the interfaces with related disciplines such as biology, medicine and materials science. The goal of each thematic volume is to give the nonspecialist reader, whether at the university or in industry, a comprehensive overview of an area where new insights are emerging that are of interest to a larger scientific audience.
As a rule, contributions are specially commissioned. The editors and publishers will, however, always be pleased to receive suggestions and supplementary information. Papers are accepted for *Topics in Current Chemistry* in English.
In references *Topics in Current Chemistry* is abbreviated Top Curr Chem and is cited as a journal.
Visit the TCC content at springerlink.com

ISSN 0340-1022
ISBN 978-3-540-36760-4 Springer Berlin Heidelberg New York
DOI 10.1007/978-3-540-36761-6

Springer is a part of Springer Science+Business Media

springer.com

Cover design: WMXDesign GmbH, Heidelberg
Typesetting and Production: LE-TEX Jelonek, Schmidt & Vöckler GbR, Leipzig

Printed on acid-free paper 02/3100 YL – 5 4 3 2 1 0

Volume Editor

Prof. Dr. Valentin Wittmann

Fachbereich Chemie
Universität Konstanz
Fach M 709
78457 Konstanz, Germany
Valentin.Wittmann@uni-konstanz.de

Editorial Board

Topics in Current Chemistry Also Available Electronically

Preface

In the last 50 years molecular biology was dominated by the exploration of proteins and nucleic acids. Beside their role in energy metabolism, oligosaccharides, which represent the third class of biomacromolecules, have received less attention. Today it is well established that oligosaccharides are involved in many important biological regulation and recognition processes from protein folding to cell–cell communication. Glycosylation of proteins is the most complex form of co- and posttranslational modification. The determination of structure-function relationships, however, remains difficult due to the microheterogeneity of glycoproteins that exist in many different glycoforms. Thus chemical synthesis of glycoproteins and glycopeptides with defined glycan structures plays a pivotal role for the detailed determination of the role of protein glycosylation. This topic is covered by the first two chapters of this book dealing with the chemical and enzymatic synthesis of glycopeptides and glycoproteins. The third chapter describes the construction of glycopeptide and glycoprotein mimetics containing non-natural structural elements. These so-called neoglycopeptides and neoglycoproteins, respectively, can provide insight on the importance of distinct structural elements on biological activity and may have improved properties such as an increased stability. The application of synthetic glycopeptides, in many cases at the clinical level, as vaccines for both cancer and HIV is the subject of the fourth chapter. Glycopeptide antibiotics are glycosylated secondary metabolites of bacteria and fungi that are synthesized by non-ribosomal peptide synthetases. Some of them serve as antibiotics of last resort in the treatment of nosocomial infections with enterococci and methicillin-resistant *Staphylococcus aureus* (MRSA) strains. Their structure, biosynthesis, and mode of action are summarized in the fifth chapter. The last chapter covers current methods for the determination of high-resolution structures of glycopeptides and glycoproteins mainly based on NMR spectroscopy, X-ray crystallography, and molecular modeling.

It was my intention to provide a cross-section of modern methods in glycopeptide chemistry and I thank all the authors for their contributions. It is the interdisciplinary combination of these synthetic, biophysical, and genomic methods that will advance the field of glycobiology in future.

Konstanz, November 2006 Valentin Wittmann

Contents

Glycoscience – Chemistry and Chemical Biology I–III

Editors: Fraser-Reid B.O., Tatsuta K., Thiem J.
Section Editors: Coté G.L., Flitsch S., Ito Y., Kondo H., Nishimura S.-I., Yu B.
2nd edition, planned spring 2008
ISBN 978-3-540-36154-1

Glycostructures play a pivotal and crucial role in a myriad of organisms and important systems in biology, physiology, medicine, bioengineering and technology. Within recent years tools have been developed to promote the understanding of their highly complex functions and the associated chemistry and chemical biology, however, many facts still remain undiscovered. These days the area is experiencing a "quantum jump". Therefore the Editors in collaboration with six responsible Section Editors completely revised and updated the 1st edition to give up-to-date information on glycostructures, their chemistry and chemical biology, and present them in the form of a comprehensive and strictly systematic survey. The texts are furnished by more than 2500 figures, chemical structures and reaction schemes (including more than 12.000 individual chemical reactions), and more than 9.000 references.

In concept and content this Major Reference Work received an essential improvement and expansion. Whereas chapter 1, 2, 3, 7, 8, and 9 are updated, expanded and rearranged versions of the 1st edition, many new subchapters are included in the chapters on monosaccharides, oligosaccharides and polysaccharides. The importance of the saccharide metabolism as well as the involvement of complex glycostructures in biologically and medically relevant processes are considered by addition of two large totally newly conceived chapters. Further, another novel chapter featuring key technologies and tools for functional glycobiology is incorporated. Responsible for this major update are the Section Editors who solicited a large number of highly reputed new authors to cover the changes in the content and to write the major supplements.

1 **General Principles**
2 **General Synthetic Methods**
3 **Chemical Glycosylation Reactions**
4 **Monosaccharides**
5 **Oligosaccharides**
6 **Complex Polysaccharides**
7 **Glycolipids**
8 **Glycoproteins**
9 **Glycomimetics**
10 **Key Technologies and Tools for Functional Glycobiology**
11 **Biosynthesis and Degradation**
12 **Glycomedicine**

Top Curr Chem (2007) 267: 1–36
DOI 10.1007/128_043

Chemical Synthesis of Glycopeptides

Christian Haase · Oliver Seitz (✉)

Institut für Chemie, Humboldt-Universität zu Berlin,
Brook-Taylor-Str. 2, 12489 Berlin, Germany
christian.haase@chemie.hu-berlin.de, oliver.seitz@chemie.hu-berlin.de

Abstract Despite its omnipresence in nature, comparatively little is known about how glycosylation affects the function and activity of proteins. To understand the role of protein glycosylation at the molecular level, it is important to have access to homogeneous glycoproteins and glycopeptides. Chemical synthesis is particularly suited to provide uniform glycoprotein material, thus playing a pivotal role for the advancement of glycobiology. This chapter summarizes the recent progress achieved in glycopeptide synthesis including new glycosylation methods, enzymatic transformations, chemoselective ligation and recombinant expression.

Keywords Glycobiology · Glycopeptides · Chemoenzymatic · Solid-phase synthesis · Chemoselective ligation

Abbreviations

Ac	Acetyl
Acm	Acetamidomethyl
Aloc	Allyloxycarbonyl

Bn, Bzl	Benzyl
Boc	*tert*-Butyloxycarbonyl
Bu	Butyl
BSA	Bovine serum albumin
CLEAR	Cross-linked ethoxylate acrylate resin
CMP	Cytidine monophosphate
DCC	Dicyclohexylcarbodiimide
DCM	Dichloromethane, methylene chloride
DIEA, DIPEA	*N*,*N*-Diisopropylethylamine
DMAP	4-Dimethylaminopyridine
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
EDC, EDCI	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDT	Ethane-1,2-dithiol
Fmoc	9-Fluorenylmethoxycarbonyl
Gdm	Guanidinium
GDP	Guanosine diphosphate
HATU	2-(1H-7-Azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HBTU	2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
HMPA	4-Hydroxymethylphenoxyacetyl
HMPB	4-(4-Hydroxymethyl-3-methoxyphenoxy)butyryl
HOBt	*N*-Hydroxybenzotriazole
HYCRON	Hydroxycrotyloligoethylene glycol-*n*-alkanoyl
ivDde	(4,4-Dimethyl-2,6-dioxocyclohex-1-ylidene)-3-methylbutyl
MCA	Monochloroacetyl
MES-Na	Sodium mercaptoethanesulfonate
MOPS	3-(*N*-Morpholino)propanesulfonic acid
NIS	*N*-Iodosuccinimide
NMP	*N*-Methylpyrrolidone
P_i	Phosphate
Pbf	2,2,4,6,7-Pentamethyldihydrobenzofuran-5-sulfonyl
Phth	Phthaloyl
PTSA	*p*-Toluenesulfonic acid
PyBOP	Benzotriazole-1-yl-tris-pyrrolidino-phosphonium hexafluorophosphate
SASRIN	Super acid sensitve resin
Su	Succinimidyl
TBAF	Tetrabutylammonium fluoride
TBS	*tert*-Butyldimethylsilyl
TCEP	Tris(2-carboxyethyl)phosphine
Tf	Trifluoromethanesulfonyl
TFA	Trifluoroacetic acid
TFE	Trifluoro ethanol
TfOH	Trifluoromethanesulfonic acid
THF	Tetrahydrofuran
TIS	Triisopropylsilane
TMS	Trimethylsilyl
Tris	Tris(hydroxymethyl)aminomethane

Troc	2,2,2-Trichloroethoxycarbonyl
Trt	Trityl
Tsoc	Triisopropylsilyloxycarbonyl
Z	Benzyloxycarbonyl

1
Introduction

Protein glycosylation is an ubiquitous process that introduces tremendous structural diversity to proteins. Glycosylation is the most complex protein modification event and nature does not provide a coding template as it is available for nucleic acids and peptides. Rather, a plethora of enzymes of diverse activities confers the transfer of complex preformed carbohydrate structures or the attachment of bridgehead, extension and branching building blocks. Hence a given protein can exist in variant glycoforms, which complicates the isolation of well-defined glycoconjugates. Other problems arise from recombinant expression techniques. The host organism's glycosylation pattern might not match that of the original organism. Moreover, the artificial cultivation conditions might result in the display of aberrant glycosylation. Chemistry can solve this availability problem. Chemical and chemoenzymatic synthesis provides access to homogeneous glycopeptides that enable a quantitative assessment of the relationship between the structure and function of protein glycosylation. This chapter was written with the intention to give insight into the very recent developments in the area of glycopeptide synthesis. Readers who seek concise treatments of the field are guided to excellent reviews [1–10].

2
The Glycosidic Linkage

Carbohydrates are commonly linked to proteins either *N*-glycosidically via the amide side chain of asparagine or *O*-glycosidically via the hydroxyl groups of serine and threonine. Less common are *O*-glycosidic linkages to hydroxylysine [11], hydroxyproline [12] and tyrosine [13], and the *C*-glycosidic attachment to tryptophane [14].

The biosynthesis of *N*-linked glycopeptides starts at the rough endoplasmatic reticulum in which a common glycan unit is transferred from a dolichol phosphate to an Asn-Xxx-Ser/Thr sequon, Xxx being any amino acid except proline [15]. Various glycosidases shear the glycan to a common pentasaccharide core (Fig. 1) that is subject to further glycosylations.

O-Linked glycopeptides do not share a common biosynthetic pathway. Unlike their *N*-linked glycoconjugate counterparts biosynthesis is accom-

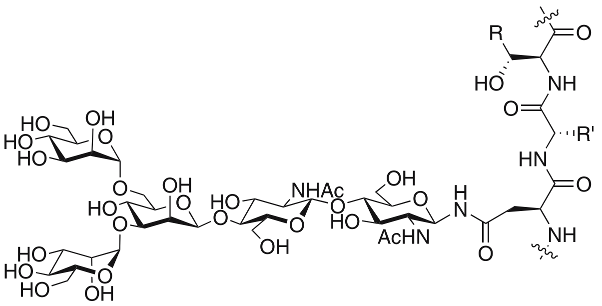
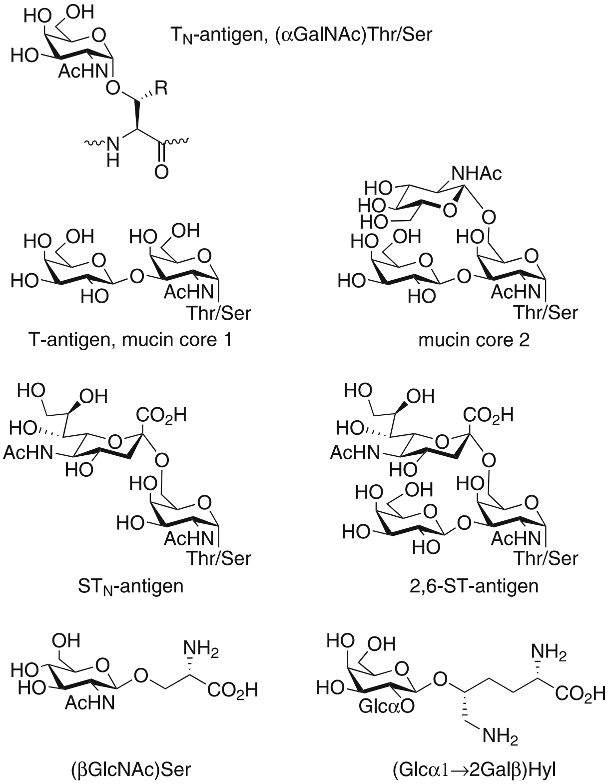


Fig. 1 Pentasaccharide core fragment and consensus sequence Asn – Xxx – Ser/Thr of *N*-glycosidically-linked oligosaccharides (R = H, CH_3)



T_N-antigen, (αGalNAc)Thr/Ser

T-antigen, mucin core 1

mucin core 2

ST_N-antigen

2,6-ST-antigen

(βGlcNAc)Ser

(Glcα1→2Galβ)Hyl

Fig. 2 Some representative examples of *O*-linked carbohydrates

plished after translation had occurred. *O*-Glycoproteins of extracellular domains frequently feature a 2-acetamido-2-deoxy-D-galactose (GalNAc) as bridgehead *α*-*O*-linked to serine or threonine. Hereupon a large quantity of glycosyltransferases attaches other monosaccharide building blocks thereby generating a bewildering number of carbohydrate structures. In cytosolic proteins, the serine and threonine hydroxyl group residues can carry *β*-*O*-linked 2-acetamido-2-deoxy-D-glucose (GlcNAc) which appears to serve regulatory purposes [16,17]. Galactose or the Glcα1 → 2Gal disaccharide structures can be found *β*-*O*-glycosidically attached to hydroxylysine or hydroxyproline on structural proteins such as collagen [18]. Some representative examples are given in Fig. 2. Many more types of glycosylation are known; for more detailed coverage the reader is addressed to the review literature [19–27].

3
Biological Role

N-Linked glycosylation plays an important role in protein folding and trafficking [28]. The lectins of the calnexin-calreticulin-chaperone system found in the endoplasmatic reticulum (ER) interact with the *N*-linked carbohydrate structures. Only properly folded glycopeptides are processed by glycosidases, which remove terminal glucose moieties from the glycan thereby liberating the functional protein. Improperly folded glycopeptides are retained in the ER by the antagonistic action of a glycosyltransferase that reattaches a glucose and in addition acts as a folding sensor [29–31]. This process allows for correct folding and disulfide bond isomerization. In an extracellular environment, *N*-linked carbohydrates act as receptor ligands and structural elements. For example, it was shown that the carbohydrate content of erythropoetine makes up 40% of its total weight and variations of *N*-linked carbohydrates alter the hormone's serum half-life [32]. NMR experiments with CD2 – a surface glycoprotein found on natural killer cells and T lymphocytes – revealed that the proper orientation of the protein-proximal GlcNAc-GlcNAc disaccharide on its surface is essential for the folding that enables CD58 binding [33]. In prokaryotic pathogens such as *Campylobacter jejuni N*-glycosylation is an important factor for adherence to and invasion into in vitro cultured eukaryotic cells [34–36].

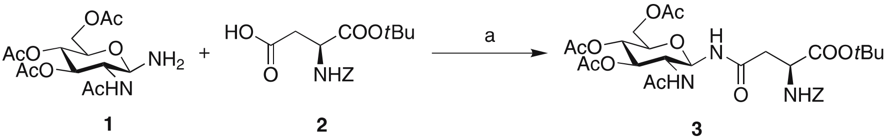
In eukaryotic systems, the most prevalent form of *O*-linked glycosylation is of the so-called mucin-type of glycosylation, in which carbohydrates are displayed on a GalNAc bridgehead *α*-*O*-linked to serine or threonine. Various epithelial cell lines express mucins on their surface. This outer mucin layer physically confines the cells and has a central role in maintaining homeostasis and, therefore, in promoting cell survival in variable conditions [37–43]. Aberrant glycosylation induced by changed expression of glycosyltransferases in the course of cancer development enables the malign degeneration to control its local microenvironment likewise [44–47]. The altered glycosylation pattern is also utilized to encapsulate from the immune system [47–55] and for tumor invasion and metastasis [56–66]. Nuclear proteins, transcription factors, and skeletal proteins have been found to contain a monosaccharide GlcNAc, which is *β*-*O*-linked to serine and threonine side chains [67,68]. The introduction of the *β*-*O*-GlcNAc moiety seems to be a dynamic regulatory modification involved in signal-transduction cascades [69,70].

4
Synthesis of Preformed Glycosyl Amino Acid Building Blocks

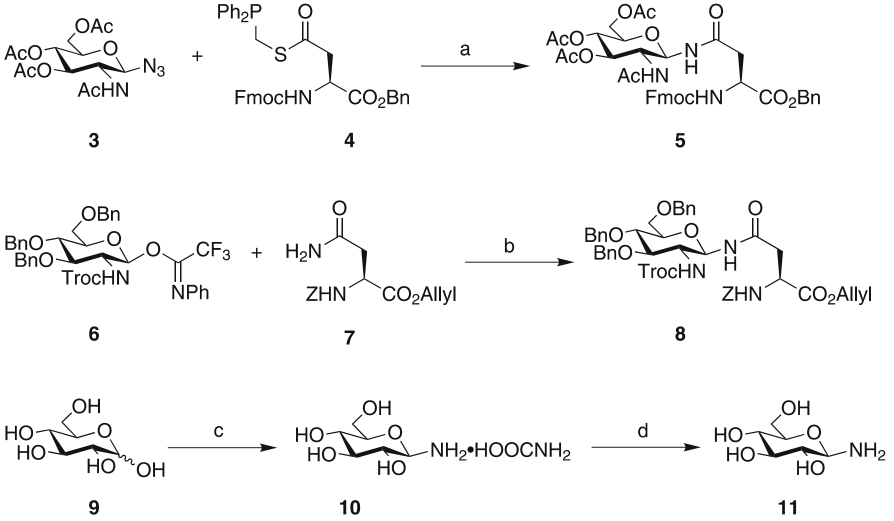
4.1
N-Glycosides

The synthetic route commonly used for the synthesis of *N*-glycosides exploits the formation of a peptidic bond between a glycosylamine **1** and an protected aspartic acid derivative **2** (Scheme 1 showing the principle) [71]. As glycosylamine precursors usually glycosylazides are employed, which can be obtained by treatment of glycosyl halides or 2-iodoglycosylsulfonamides with azide salts [72–83].

In a strategy reported by Kiessling et al., glycosylazides are used directly for a stereoselective *N*-glycosylation by Staudinger Ligation. Herein the glycosylazide **3** was reacted with phosphinothioester **4** furnishing glycosylasparagine **5** in a moderate yield (Scheme 2) [84]. Takahashi and coworker described a synthesis that avoids the need to prepare glycosylazides. The *β*-glycosyl trifluoroacetoimidate **6** was shown to react with asparagine **7** to deliver *β*-configurated *N*-glycosyl asparagine **8** [85]. Glycosylamines can be prepared directly from unprotected sugars by Kochetkov's procedure, which involves treatment with ammonium bicarbonate [86,87]. An improved procedure based upon ammonium carbamate was introduced by Likhosherstov [88–90]. The carbamic acid salts **10** are stable upon storage and are converted into the free amines **11** by base-treatment or by applying vacuum. Flitsch et al. have shown the acceleration of the Kochetkov procedure by irradiation with microwaves. The conversions of the sugars into the corresponding amines were finished within 90 min with yields ranging from 35 to 87% [91]. Complex *N*-linked glycosyl amino acids can be obtained by releasing *N*-linked carbohydrates from natural glycoconju-



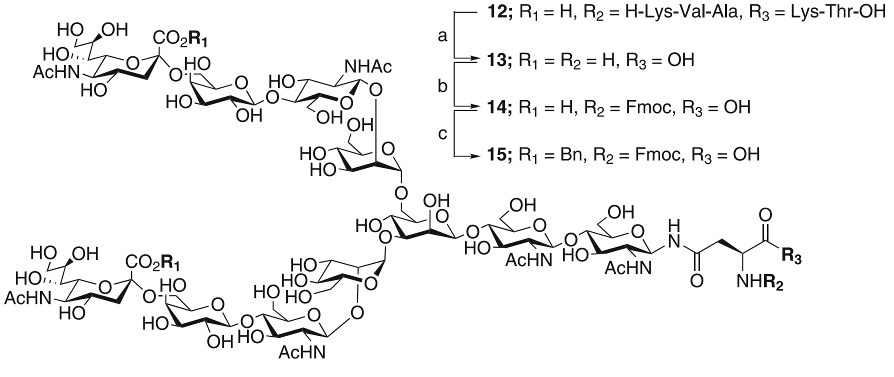
Scheme 1 (a) DCC, HOBt, THF, 79.5% [71]



Scheme 2 (a) DMF, 45%; (b) TMSOTf, $MeNO_2$, 0 °C to r.t., 68%; (c) NH_4COONH_2 in sat. NH_3 solution (d) base or high vacuum

gates [92]. In a recent example, chitin was enzymatically digested to furnish chitobiose, which was subsequently converted to the corresponding glycosyl amine [93].

Kajihara and co-workers isolated the sialylglycopeptide **12** in multigramm quantities from hen egg yolk (Scheme 3) [94]. Enzymatic digest of **12** with actinase E furnished glycosyl amino acid **13**. For application in solid phase peptide synthesis the amino group and the sialic acid carboxyl groups had to be protected. After *N*-Fmoc-protection the glycosyl amino acid **14** was treated with benzyl bromide and Cs_2CO_3 to block the sialic acid as ben-

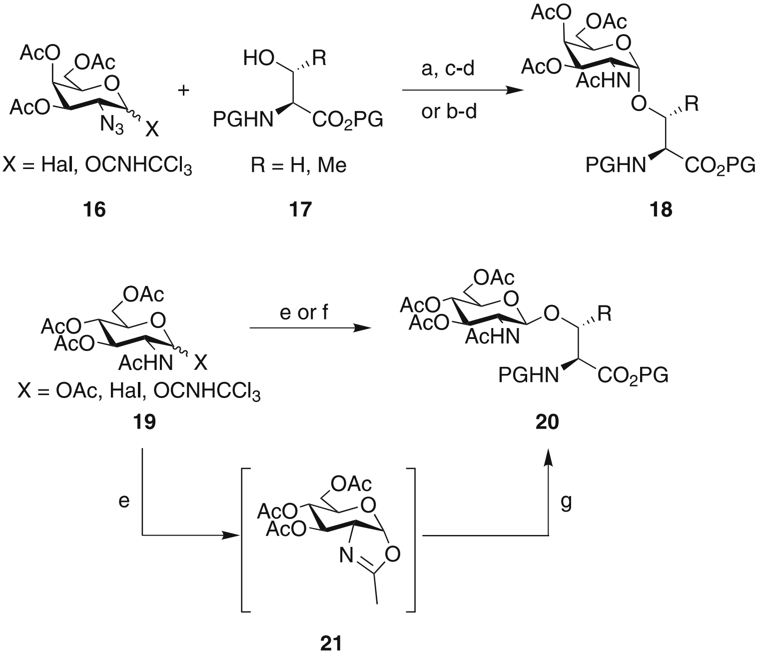


Scheme 3 (a) Actinase E, Tris-HCl buffer, NaN_3, pH 7.5, 86%; (b) Fmoc-OSu, $NaHCO_3$, acetone, H_2O, 68%; (c) Cs_2CO_3, H_2O, then BnBr, DMF, 85%

zyl ester. Interestingly, the aspartic acid carboxyl group was not affected by that treatment and remained available to activation for coupling in peptide synthesis. A marginal restriction arose during the application in SPPS. The coupling of the adjacent amino acid had to be stopped at 85% conversion, otherwise hydroxyl groups of the glycan moiety were acylated also [95]. Glycosyl amino acid **14** was used as versatile building block that allowed the remodeling of the carbohydrate structure by chemical and enzymatical means [96, 97].

4.2
O-Glycosides

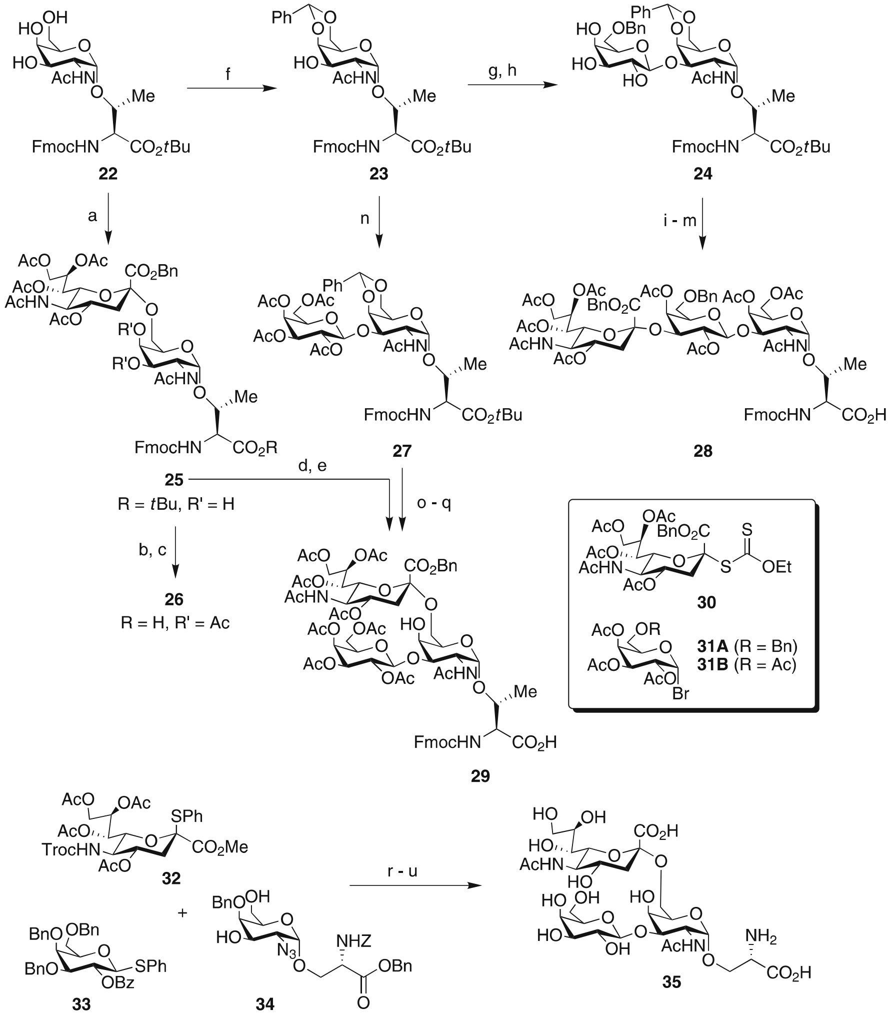
The preparation of the basic *O*-linked glycosyl amino acids (αGalNAc)Ser/Thr **18** is typically accomplished by treatment of glycosyl halides such as **16** with the free hydroxyl groups of otherwise protected amino acids **17** in the presence of silver salts (Scheme 4) [98–114]. High α-selectivities can be achieved when benzyl-protected glycosyl fluorides are employed as glycosyl acceptors [115]. The alternative reaction of acetimidates **16** with the protected amino acids **17** in the presence of Lewis acids also assures access to the



Scheme 4 (a) $Ag_2CO_3/AgClO_4$ (X = Hal); (b) TMSOTf (X = $OCNHCCl_3$); (c) $NaBH_4$, $NiCl_2$; (d) Ac_2O, pyridine; (e) BF_3OEt_2, **17** (X = OAc); (f) AgOTf, **17** (X = Hal, $OCNHCCl_3$); (g) $CuCl_2$

desired compounds [98, 106, 107, 109, 113]. The synthesis of α-configured glycosyl amino acids **18** relies on the 2-azido group as a non-participating precursor of the 2-acetamido moiety [116]. In either case, the 2-substituent will achieve its final state by reduction of the azido moiety. A conceptionally different approach developed by Schmidt et al. utilized a Michael-type reaction of a 2-nitroglycal with Boc/*t*Bu-protected serine or threonine [117–120]. Glycosyl donors such as **19** [121–131] are employed when aiming for (βGlcNAc)Ser/Thr building blocks **20**. The anchimeric assistance of the acetamido group (R = Ac) or other acylamido groups (R = Troc, Phth, Aloc) directs the glycosylation to the β-side. Rapid access to glycosyl amino acids is granted by the use of peracetylated sugars as glycosyl donors [121–125]. An accelerated procedure applying microwave irradiation has recently been reported by Seibel et al. [132]. In the synthesis of (βGlcNAc)Ser/Thr building blocks, the relatively unreactive oxazolines such as **21** are commonly found by-products. It has been shown that $CuCl_2$ can act as mild activator, enabling efficient coupling to Fmoc/All-protected serine [133].

For the synthesis of complex *O*-glycosides, a stepwise glycosylation strategy has emerged as reliable. Thus, the carbohydrate part is assembled by using a glycosyl amino acid as an acceptor, to which further glycan residues are appended. Versatile syntheses for the preparation of the more complex 2,6-sialyl-T, 2,3-sialyl-T and sialyl-T_N-antigens were developed by Kunz et al. The assembly of the sialyl-T_N building blocks **28** and **29** started from glycosyl threonine **22** (Scheme 5). In the reaction of **22** with the sialyl xanthate **30** only the primary hydroxyl function reacted [134]. Subsequent protecting group manipulations furnished the sialyl-T_N building block **26**. In galactose the equatorial 3-OH is more reactive than the axial 4-OH. Hence, glycosylation of **25** with glucosyl bromide **31B** delivered sialyl-T_N building block **29**. When necessary, the primary hydroxyl function of **22** was shielded from reacting by blocking it as benzylidene acetal as seen in **23**. Thereby it was possible to generate the T-antigen precursor **24**. The reaction of **24** with the sialyl xanthate **30**, in which the C3-OH-function of **24** was attacked selectively, afforded the (2,3)-sialyl-T-threonine **28** [135]. The sialyl amino acids **26**, **28** and **29** were used for solid-phase synthesis [134–139]. Takahashi and co-workers developed a one-pot two-step strategy for the synthesis of the 2,6- [140, 141] and 2,3-sialyl-T-antigens [141]. In the synthesis of the former, glycosyl serine **34** was glycosylated at the primary hydroxyl group by using β-thiophenyl sialoside **32**. Addition of thiogalactoside **33** to the reaction mixture furnished a fully protected 2,6-sialyl-T-antigen. An enzymatic approach towards the synthesis of 2,3 and 2,6-sialyl-T-antigens as well as other *O*-linked sialyl oligosaccharides was published by Paulson et al. in which sialylation was established by the use of two sialyltransferases, porcine Galβ(1–3)GalNAcα-2,3-sialyltransferase (pST3Gal I) and chicken GalNAcα2,6-sialyltransferase (chST6GalNAc I) [142]. A one-pot reaction sequence furnishing the sialyl-T threonine derivative was reported by Kunz et al. Herein Fmoc/*t*Bu-protected

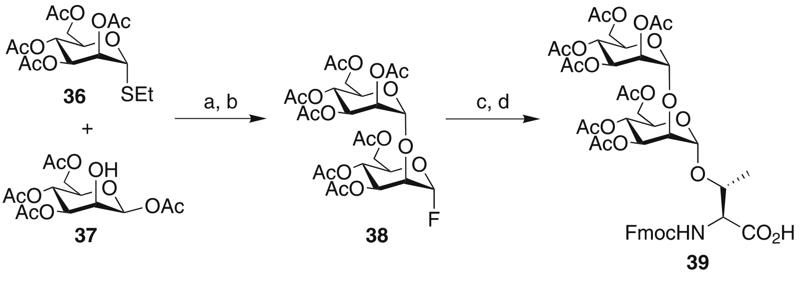


Scheme 5 (a) **30**, MeSBr, 3 Å MS, MeCN – CH_2Cl_2, – 65 °C, 59%; (b) Ac_2O, pyridine; (c) TFA, anisole, 85% (two steps); (d) **31B**, $Hg(CN)_2$, 4 Å MS, CH_3NO_2 – CH_2Cl_2; (e) TFA, anisole, 34% (two steps); (f) PTSA (cat.), $PhCH(OMe)_2$, DMF, 50 °C, vacuum, 75%; (g) **31A**, $Hg(CN)_2$, 4 Å MS, CH_3NO_2 – CH_2Cl_2; (h) NaOMe, MeOH, pH 8.5, 58% (two steps); (i) **30**, MeSBr, 3 Å MS, MeCN – CH_2Cl_2, – 65 °C; (j) aq. AcOH (80%); (k) Ac_2O, pyridine; (l) TFA, anisole, 41% (four steps); (m) **31B**, $Hg(CN)_2$, 4 Å MS, CH_3NO_2 – CH_2Cl_2, 86%; (n) aq. AcOH (80%), 87%; (o) **30**, MeSBr, 3 Å MS, MeCN – CH_2Cl_2, – 65 °C, 58% α and 8% β-anomer; (p) TFA, anisole, 98%; (q) **32** (1.2 equiv.), **34** (1.0 equiv.), NIS (1.44 equiv.), TfOH (0.1 equiv.), CH_3CN (5 mL mmol^{-1}), 3 Å MS (0.5 g mmol^{-1}); – 35 °C, then **33** (1.8 equiv.), NIS (2.70 equiv.), TfOH (0.2 equiv.), CH_2Cl_2 (45 ml mmol^{-1}), – 30 °C, 77%, $\alpha : \beta = 78 : 22$; (r) Zn dust, THF, HOAc, Ac_2O, 0 °C, 1 h; (s) $Pd(OH)_2$, THF, MeOH, HOAc, H_2O, 2 h; (t) 0.1 N NaOH, MeOH, then H_2O, 84% (three steps)

(GalNAcα)-threonine was subjected to a transgalactosylation employing lactose as donor and β-galactosidase as biocatalyst, and a subsequent sialyl transfer. The resulting *O*-unprotected glycoconjugate was then fully protected before acidolysis of the amino acids *t*Bu-ester set the stage for solid phase synthesis [143–145]. Albeit low yields the direct glycosylation of *O*-unprotected serine and threonine building blocks employing glycosidases has been reported [146–149].

The synthesis of α-mannosylated serine or threonine building blocks usually provides no specific challenges. Recently, Unverzagt et al. described the concise synthesis of dimannosyl amino acids (Scheme 6). Disaccharide **38** was prepared by allowing thioglycoside **36** to react with mannosyl acetate **37**. Under the conditions of thioglycoside activation, the anomeric acetate of **37** remained intact. Conversion to the glycosyl fluoride activated the disaccharide **38** for the Lewis acid mediated reaction with Fmoc/Bn-protected threonine. Subsequent hydrogenolysis afforded the glycosyl amino acid **39** as suitable building block for solid-phase synthesis [150].

The synthesis of glycosylated hydroxylysine derivatives is not an easy task due to the need for three amino acid protecting groups. In the group of Kihlberg it was shown that the synthesis of galactosylated hydroxylysine building blocks can be achieved by an elaborate protective group manipulation sequence [151]. Blocking the ε-nitrogen with the acid-stable Z-group allowed for the extension of the galactose moiety. Unfortunately, the β-selectivity of the galactosylation suffered from steric demands of the protective group when sugar moiety was introduced with the α-1,2-anhydrosugar-precursor [152]. This selectivity problem was circumvented by applying the Koenigs–Knorr glycosylation with anchimeric assisting sugar halogenides [153–155]. A different approach towards a galactosylated 5-hydroxylysine mimetic started from *N*-protected-6-substituted piperidin-2-ones [156]. It is noteworthy that the work of Kihlberg and co-workers substantially contributed to the understanding of the influence of glycosylated collagen in rheumatoid arthritis [157–161].

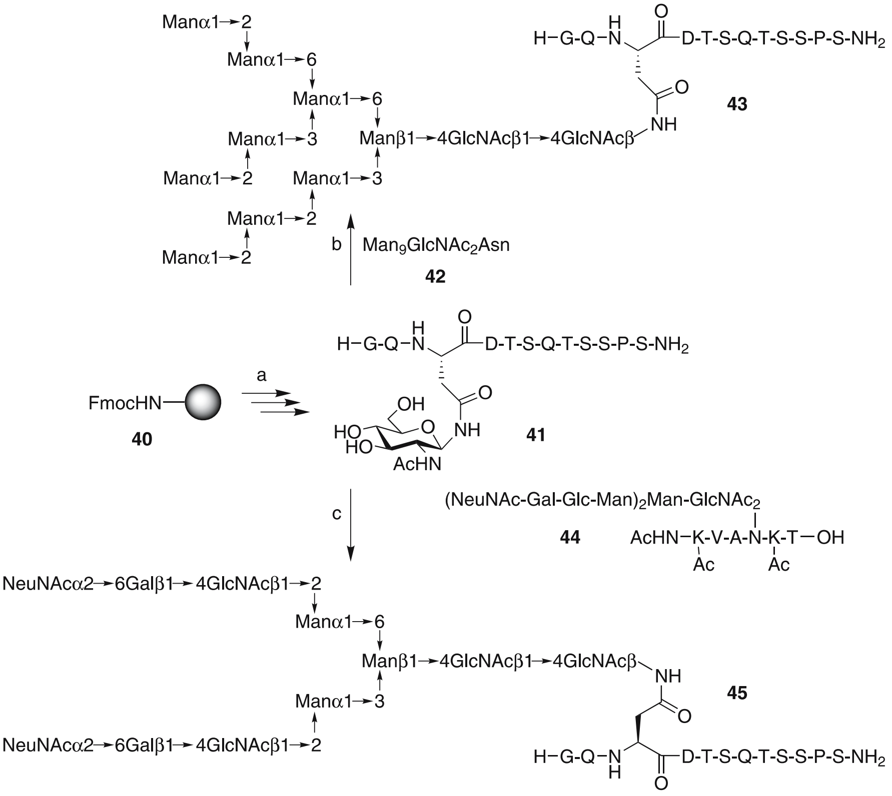


Scheme 6 (a) NIS, TfOH, 4 Å MS, 0 °C, 30 min, 67%; (b) HF-pyridine, CH_2Cl_2, 2 h, 70%; (c) Fmoc – Thr – OBn, $BF_3 \cdot OEt_2$, 4 Å MS, CH_2Cl_2, 2 h, 70%; (d) Pd/C 10%, H_2, MeOH, 2 h, 91%

5
Glycopeptide Synthesis in Solution

5.1
N-Glycopeptides

Enzymatic synthesis has been demonstrated to be a very powerful technique in constructing the carbohydrate part of glycopeptides. In analogy to biosynthetic pathways, glycosyltransferases can be used to step-by-step assemble carbohydrate structures by using sugar nucleotides as donors. This approach was predominantly utilized in the synthesis of *O*-linked glycopeptides like fragments of the P-selectin glycoprotein ligand-1 (PSGL-1) [162–165]. Certain glycosidases catalyze the release of *N*-glycans by cleaving the β-glycosidic



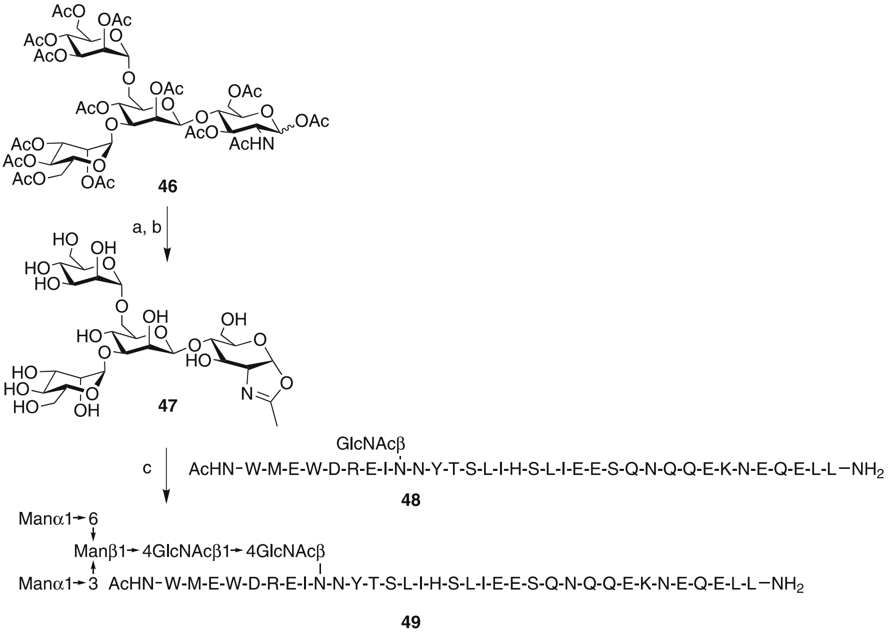
Scheme 7 (a) (i) 20% piperidine/DMF (ii) Fmoc-amino acid/glyco-amino acid, HATU, DIPEA (iii) repeat i–ii (iv) Cleavage/deprotection (95 : 5 TFA-H_2O or 90 : 5 : 3 : 2 TFA/thioanisole/anisole/EDT) (v) De-*O*-acetylation (5% hydrazine/H_2O) (v) RP-HPLC; (b) Endo-A, acetate buffer (pH 6), 25% acetone, $Man_9GlcNAc_2Asn$; (c) Endo-M, Ac-Lys(Ac)-Val-Ala-Asn[$(NeuNAc\text{-}Gal\text{-}GlcNAc\text{-}Man)_2Man\text{-}GlcNAc_2$]-Lys(Ac)-Thr-OH

linkage between two GlcNAc residues adjacent to asparagines in *N*-glycans. In a trans-glycosylation approach these enzymes can transfer saccharide moieties from glycosyl asparagine donors to *N*-linked GlcNAc-peptides.

Wang and co-workers applied the endo-*β*-*N*-acetylglucosaminidases from *Arthrobacter protophormiae* (Endo-A) and from *Mucor hiemalis* (Endo-M) for the synthesis of peptides displaying high-mannose type *N*-glycans in **43** and complex type *N*-glycans in **45**, respectively (Scheme 7) [166–168]. Glycopeptide products such as **43** and **45** are obtained in relatively modest yields and, though elegant, access to the complex *N*-glycopeptide donors such as **42** and **44** is required. Significant improvements of trans-glycosylation yields were brought about by Shoda's proposal of using 1,2-oxazolines as substrates [169].

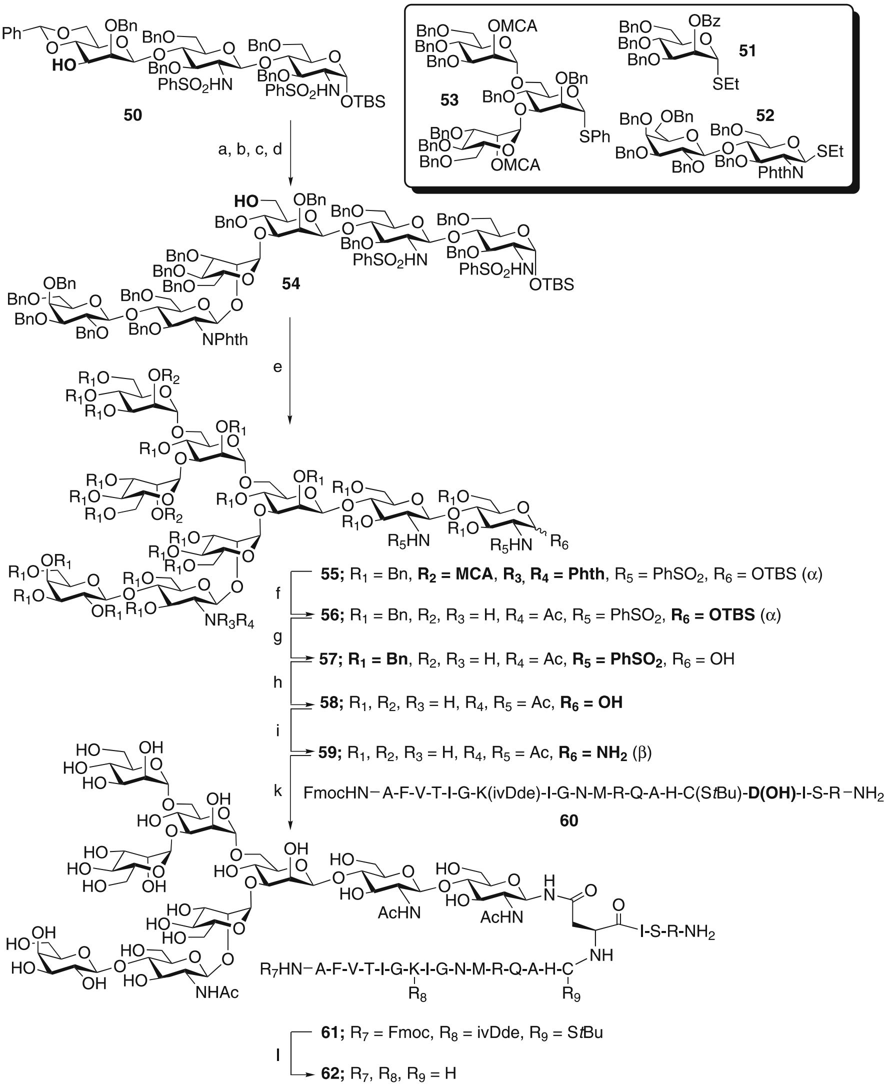
For the synthesis of the tetrasaccharide-peptide **49**, oxazoline **47** was prepared by allowing peracetylated tetrasaccharide **46** to react with a hard Lewis-acid (Scheme 8). Oxazoline **47** was offered to endo-A in a trans-glycosylation involving 34-mer GlcNAc-glycopeptide **48** as carbohydrate acceptor. A strong preference for oxazoline turnover rather than product hydrolysis was observed giving rise to spectacular 75% yield of product **49** [170].

A different approach towards enzyme-supported glycopeptide synthesis takes advantage of proteases that can be employed to catalyze the ligation



Scheme 8 (a) TMSBr, $BF_3 \cdot OEt_2$, collidine; (b) NaOMe, MeOH, 81%; (c) Endo-A, buffer (pH 6.5), 75%

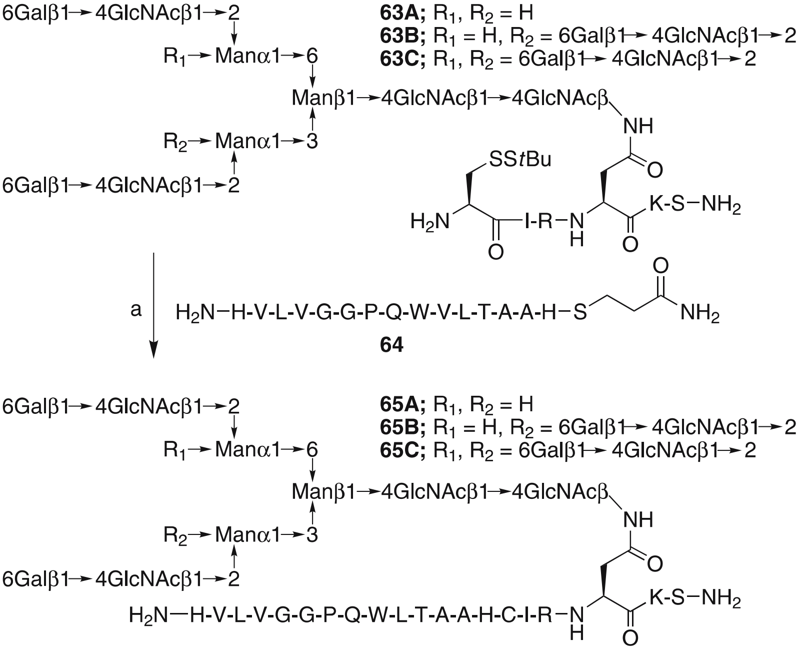
of peptide fragments. The group of Wong employed the protease subtilisin to achieve the condensation of glycopeptide fragments [171]. Recently, Davis et al. discovered that a combination of site-directed mutagenesis and chem-



Scheme 9 (a) **51**, $(BrC_6H_4)_3NSbCl_6$, MeCN, 80%; (b) NaOMe/MeOH, 89%; (c) **52**, $(BrC_6H_4)_3NSbCl_6$, MeCN, 60%; (d) $BH_3 \cdot THF$, Bu_2BOTf, CH_2Cl_2, 75%; (e) **53**, $(BrC_6H_4)_3NSbCl_6$, MeCN, 78%; (f) (i) $NH_2CH_2CH_2NH_2$, *n*-BuOH/toluene, 90 °C; (ii) Ac_2O, pyridine; (iii) NaOMe/MeOH, 69%; (g) TBAF, AcOH, 89%; (h) (i) Na/NH_3, – 78 °C; (ii) $NaHCO_3$, Ac_2O, 70%; (i) NH_4HCO_3/H_2O; (j) HATU, DIPEA, DMSO; (k) (i) NH_2NH_2, piperidine, DMF; (ii) MES-Na, DMF, DIPEA, H_2O, 19% last three steps

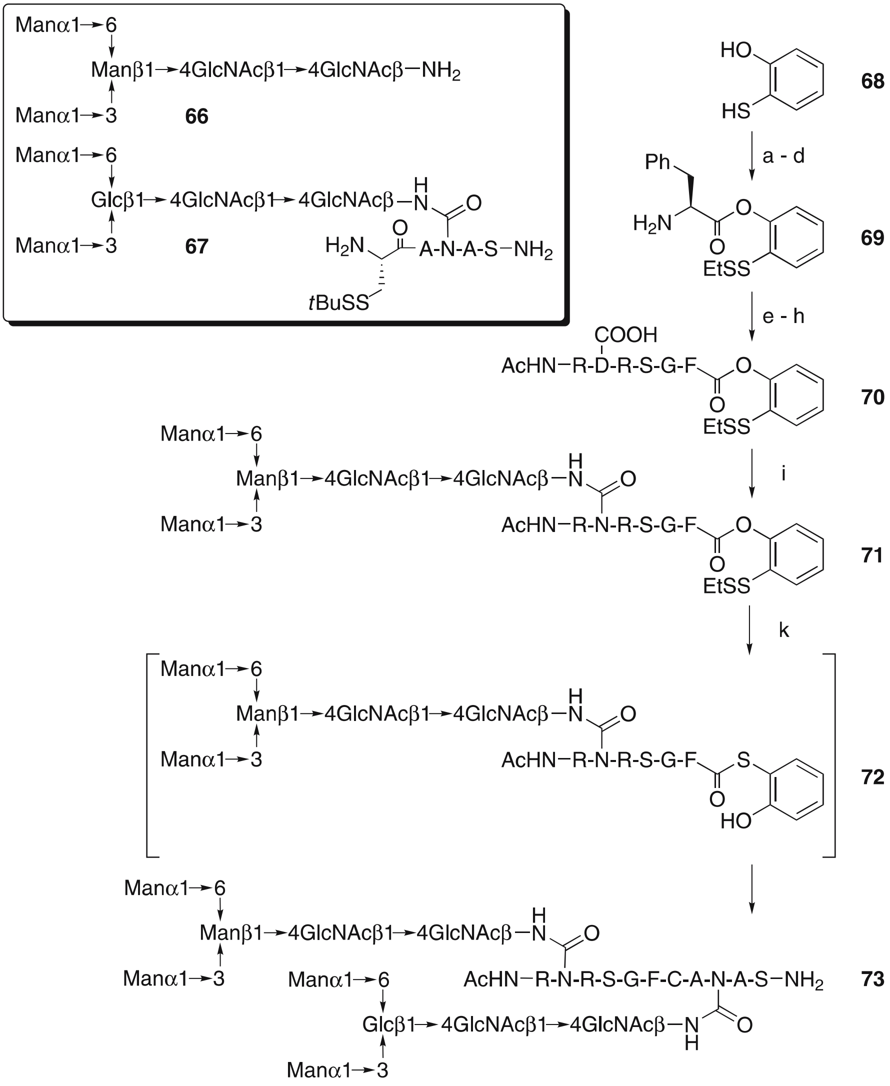
ical modification of subtilisin from *B. lentus* allows to improve the substrate tolerance and the coupling efficiency [172].

In the chemical solution-phase synthesis of *N*-glycopeptides it is practical to employ a block glycosylation approach in which the full-length carbohydrate is coupled to the aspartic acid side chain [173]. For example, in Danishefsky's synthesis of the HIV envelope glycoprotein gp120-fragment **62** (Scheme 9) the 3-OH of the mannose residue of precursor **50** [174] was subjected to an α-mannosylation with building block **51** followed by debenzoylation and extension with thiolactoside donor **52**. Regioselective opening of the benzylidene exposed the primary hydroxyl group of hexasaccharide **54**. The glycosylation with the trimannosyl building block **53** completed the assembly of the nonasaccharide **55**. In preparing the stage for the connection of the sugar unit with the peptide, the protecting groups were removed followed by conversion of **58** to the glycosylamine **59** by using Kochetkov's procedure. The aspartic acid side chain of *N*,*S*-protected peptide **60** was activated by HATU in DMSO to accomplish the coupling with glycosylamine **59**. Removal of the remaining protecting groups afforded glycopeptide **62** [175]. A similar strategy was employed in the synthesis of the high-mannose-type analogue of glycoconjugate **62**, which also spanned a fragment of HIV virus envelope-protein gp120. Both highly complex glycopeptides are putative HIV vaccines [176–178].



Scheme 10 (a) MES-Na, pH = 7.4, 15% (**65A**), 38% (**65B**) and 65% (**65C**), respectively

For further elaboration of the peptide part it is attractive to take advantage of convergent methods of peptide synthesis. Probably the most powerful method of achieving fragment condensations of deprotected peptides is the native chemical ligation, the reaction of a peptide thioester



Scheme 11 (a) I_2, MeOH, H_2O; (b) $BF_3 \cdot OEt_2$, EtSSEt, CH_2Cl_2, 99%, two steps; (c) Boc-Phe-OH, EDCI, DMAP, CH_2Cl_2/THF, 93%; (d) 4 N HCl/dioxane, 94%; (e) Fmoc-R(Pbf)D(*t*Bu)R(Pbf)S(*t*Bu)G-OH, HATU, DIEA, DMF, 61%; (f) piperidine, DMF; (g) Ac_2O, pyridine; (h) TFA/phenol/TIS/H_2O, 35 : 2 : 1 : 1, 60%; (i) **66**, HATU, DIEA, DMSO, 50%; (k) **67**, 0.2 M phosphate, 0.2 M NaCl, pH ~ 7.4, excess MES-Na, 75%

with a cysteinylpeptide [179]. The application of native chemical ligation in *O*-glycoprotein synthesis was pioneered by Bertozzi [180, 181].

In Scheme 10 examples of *N*-glycopeptide synthesis are depicted. For the synthesis of prostate-specific antigen fragments the *S*-protected cysteinylglycopeptides **63A**, **63B** and **63C** were prepared. The native chemical ligation was executed by adding sodium mercaptoethanesulfonate (MES-Na) at pH 7.4 in an aqueous buffer, which conferred the removal of the *St*Bu-protecting group and the coupling with peptide thioester **64** [182, 183].

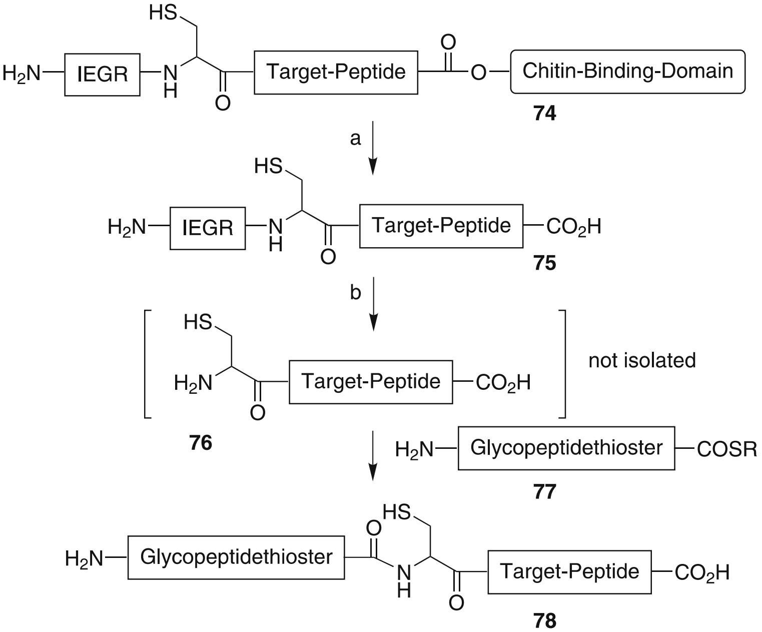
The lability of thioesters can set limitations to the synthesis of glycopeptides fragments to be employed as acyl donor. Danishefsky has introduced a latent thioester that becomes activated for ligation upon exposure to reducing conditions. Herein the phenylalanine mercaptophenyl ester **69** played a pivotal role (Scheme 11). Extension to the hexapeptide mercaptophenyl ester **70** and subsequent coupling of glycosylamine **66** afforded **71**. The disulfide in **71** was cleaved during native chemical ligation enabling the formation of thioester **72** and the subsequent reaction with cysteinylglycopeptide **67** [184].

Large protein structures are available by means of recombinant technologies. Wong and co-workers reported the recombinant synthesis of fragments of the bioactive "cysteinylproteins" IL-2 and C37-H6, a peptide derived from the HIV gp41 which were subsequently used in the native chemical ligation with a glycopeptidethioester [185]. An analogous methodology was applied by Imperiali et al. For a comparative folding study of glycosylated and unglycosylated bacterial immunity protein Im7 a recombinant cysteinylprotein and an artificial glycopeptidethioester were ligated [186]. Guo showed in the synthesis of a fucose-containing glycopeptide that comparably short segments can be ligated in organic solvents by using pentafluorophenylesters [187, 188]. The authors emphasized the convenience of work-up procedures due to the use of *O*-unprotected carbohydrates throughout the peptide assembly [189]. Fluorous phase-tags that facilitate removal of reagents by organic phase/fluorous phase extractions have been reported to support the synthesis of a GlcNAc-pentapeptide [190].

5.2
O-Glycopeptides

In principle, the synthesis of *O*-linked glycopeptides can be performed by attaching the glycan as block to the amino acid/peptide aglycon or by stepwise assembly of the peptide part by using preformed glycosyl amino acids. The modest yields of block glycosylations and their relatively poor α/β-selectivities call for the application of the latter approach. The step-by-step condensation of presynthesized building blocks in solution has found widespread application. Danishefsky reported the synthesis of various glycopeptides bearing complex *O*-linked carbohydrate structures [106, 109, 113, 191, 192]. The synthesis of a short trivalent T antigen glycopeptide was reported by Guo et al. [115]. The approach benefited from straightforward work-up steps due to the *O*-unprotected glycan moieties. The convergent synthesis of the peptide part of *O*-glycopeptides has been performed by fragment ligations of protected and unprotected peptide segments in organic and aqueous solvents, respectively.

Ligation methods that allow the use of unprotected (glyco)peptides provide the opportunity to employ segments accessed by both chemical and recombinant means. In an approach known as expressed protein ligation, Macmillan and Bertozzi coupled chemically synthesized multiply glycosylated peptide thioesters with recombinant peptide fragments. The peptides **74** were expressed in *E. coli* as conjugates with a *C*-terminal chitin-binding domain as a means of purification (Scheme 12). The removal of the chitin-binding domain proceeded through the intermediary formation of a thioester that was hydrolyzed to obtain peptide **75**. Subsequently peptide **75** was treated with factor Xa protease which cleaved behind *C*-terminal to the recognition sequence Ile – Glu – Gly – Arg. The exposed cysteinylpeptide **76** was submitted without purification to native chemical ligation with glycopeptidethioester **77** [193]. The synthetic strategy employing latent thioesters developed by Danishefsky and co-workers was also demonstrated for the syn-

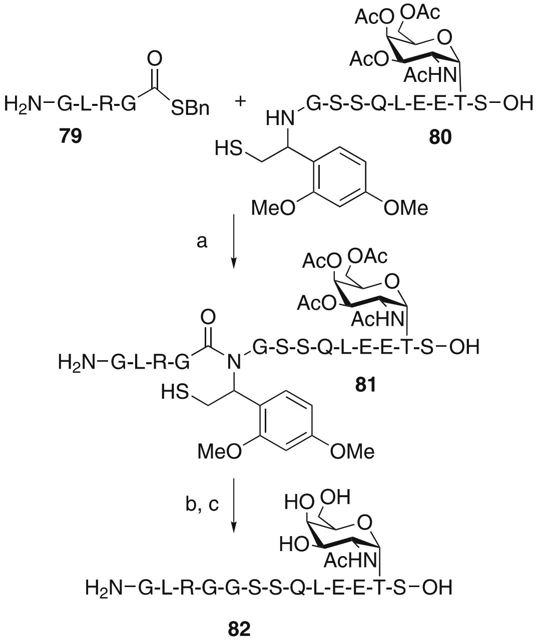


Scheme 12 (a) 50 mM DTT, 100 mM $NaHPO_3$, pH 8.0, 100 mM NaCl; (b) Factor Xa protease, 5 mM $CaCl_2$, 100 mM $NaHPO_3$, 100 mM NaCl, 2% MES-Na

thesis of *O*-linked glycopeptides [184]. A detailed description has been given in the preceding section (Scheme 11). To enable “cysteine-free” native chemical ligation reactions thiol-containing auxiliary groups have been attached to the peptide *N*-terminus.

For the synthesis of *O*-linked glycopeptides the cysteine-free ligation was scrutinized. Macmillan and Anderson explored the 1-(2,4-dimethoxyphenyl)-2-mercaptoethyl auxiliary in the synthesis of glycopeptide **82** (Scheme 13). The phenethyl thiol group in **80** conferred the coupling with peptide thioester **79**. The conversion of the *N*-benzylamine **80** to the *N*-benzylamide **81** allowed the acid-mediated cleavage of the auxiliary group to afford the unprotected glycopeptide **82** [194].

The application of glycosyltransferases for the stepwise assembly of oligosaccharides in *O*-linked glycopeptides has been demonstrated in a plethora of examples [121, 126, 162–165, 195, 196]. Recently, the synthesis of a MUC1 peptide bearing a Lewis X *O*-glycan was reported by Vliegenthart and co-workers. The authors showed a completely enzymatic synthesis in which the carbohydrate structure was elaborated starting from the unglycosylated peptide [197]. Sialylation of T$_N$ glycopeptides employing a mutated α2,6-sialyltransferase originating from *Photobacterium damsela* was described by Lin et al. To facilitate an activity screening, the authors labelled



Scheme 13 (a) 6 M Gdm-HCl, 200 mM sodium phosphate buffer, pH 8.0, 2% MES-Na, 20 mM TCEP; (b) 95% TFA; (c) 2% $N_2H_4 \cdot H_2O$, 10 mM sodium phosphate buffer

the *N*-terminus of the glycopeptide substrates with the dabsyl-dye, which was well tolerated by the enzyme [198].

6
Synthesis on Solid Support

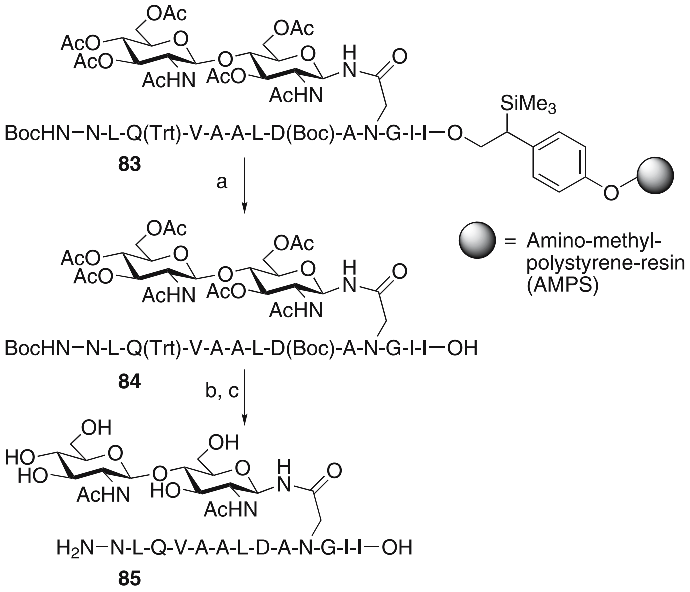
Solid-Phase synthesis allows the automation of the highly repetitive process of building block coupling – if desired in parallel and combinatorial formats. Further advantages of solid-phase synthesis are the speed and ease of synthesis and the possibility to drive reactions to completion by using a large excess of building blocks and reagents. To cope with the acid sensitivity of many carbohydrates, the solid-phase synthesis of glycopeptides is most commonly performed by coupling *N*-Fmoc-protected amino acid building blocks, which allow the application of relatively mild conditions for the cleavage of temporary and permanent protecting groups. With few alterations concerning the removal of carbohydrate protecting groups, the standard protocols of Fmoc-based solid-phases peptide synthesis can be applied.

6.1
N-Glycopeptides

In solid-phase synthesis, the choice of the linker that connects the growing glycopeptide with the solid support is of critical importance. Acid-labile linkers such as the Wang [199], the HMPA [200], the Rink [201] and or the PAL-linkers [202] have frequently been used to release unprotected glycopeptide acids and glycopeptide esters. The synthesis of protected glycopeptide fragments has been accomplished by using very acid-labile linkers such as SASRIN [203], HMPB [204], Sieber [205] or trityl linkers [206] or the Pd(0)-labile allylic HYCRON anchor [207].

Recently, Kunz and co-workers introduced the (2-Phenyl-2-trimethylsilyl)-ethyl-linker in **83** that allowed the fluoride ion-mediated release of protected *N*-glycopeptide **84** under practically neutral conditions (Scheme 14). The removal of the peptide side-chain protecting groups and the *O*-acyl groups was accomplished by acid and base treatment, respectively [134].

Guo and co-workers presented the synthesis of the impressive glycoconjugate **91** bearing an *N*-linked fucosylglycan moiety and in addition a glycosylphosphatidylinositol-anchor (Scheme 15). The solid-phase synthesis of peptide **88** proceeded on a trityl-resin, which allowed, after coupling of benzyl-protected glycosylasparagine **86** and peptide extension, the detachment of fully protected glycopeptide **90**. Subsequently, the peptide bond between **90** and the protected GPI-anchor-precursor **87** was established. The benzyl protecting groups were removed from the glycan moieties by hydrogenolysis. Although known for its acid lability, the unprotected fucose

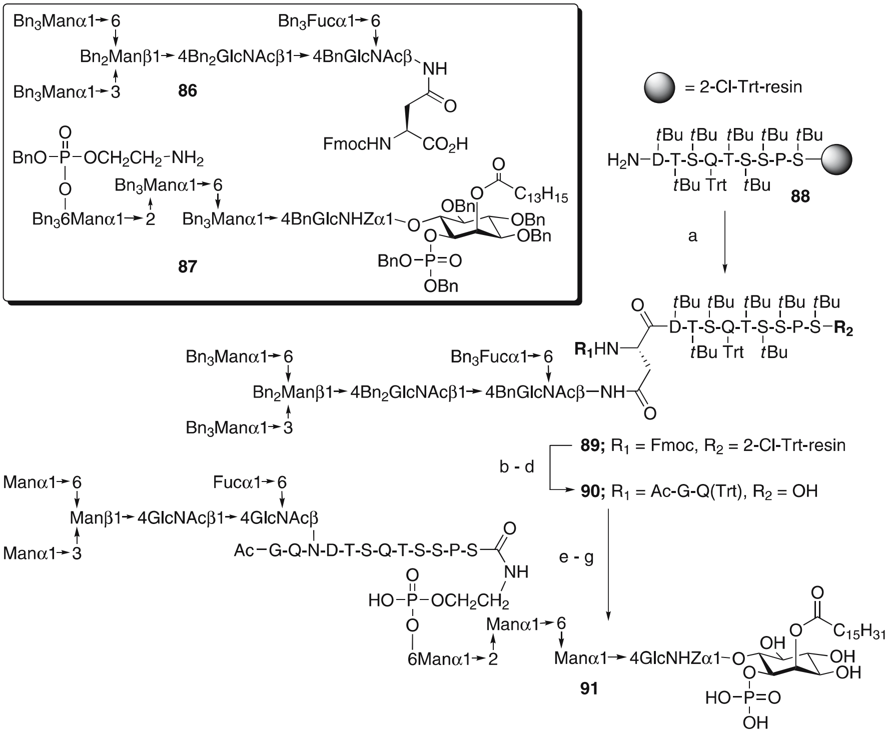


Scheme 14 (a) TBAF · $3H_2O$ (2 equiv), CH_2Cl_2, RT, 25 min; (b) TFA, Et_3SiH, H_2O (38 : 1 : 1), RT, 4 h, 97%; (c) NaOMe, MeOH, pH 10, RT, 7 h, 79%

moiety in **91** remained unaffected during the final acid treatment which removed the peptide side chain protecting groups [208].

Most commonly, solid-phase synthesis of glycopeptides is performed with the sugar hydroxyl groups protected as benzyl ethers or acetates. Kajihara and co-workers reported an example in which the hydroxyl groups of the complex-type sialyl glycosyl asparagine **15** remained unprotected [95].

Solid-phase synthesis provides a convenient means of synthesizing peptide segments to be used in further ligation reactions. Hojo et al. used the thioester ligation to gain access to a fragment of the extracellular matrix metalloproteinase inducer (Emmprin), a glycopeptide located on the surface of cancer cells [209]. However, the routine protocols of Fmoc-solid-phase synthesis cannot be applied due to the lability of thioesters against nucleophilic cleavage during the Fmoc-deprotection step. To prevent undesired losses, a mixture of 1-methylpyrrolidine, hexamethyleneimine and HOBt in *N*-methylpyrrolidinone/DMSO was employed [210]. The introduction of the asparagine **92** carrying the sterically demanding benzyl-protected pentasaccharide core was performed by using HATU as coupling reagent (Scheme 16). The resin-bound glycopeptide **98** was extended by coupling pentafluorophenyl-activated amino acids. Treatment of **99** with TFA and subsequently TFMSA cleaved the amino acid side-chain and the sugar protecting groups and detached the unprotected glycopeptide. After reprotecting the amino groups, glycopeptide **101** was obtained. The silver-mediated thioester

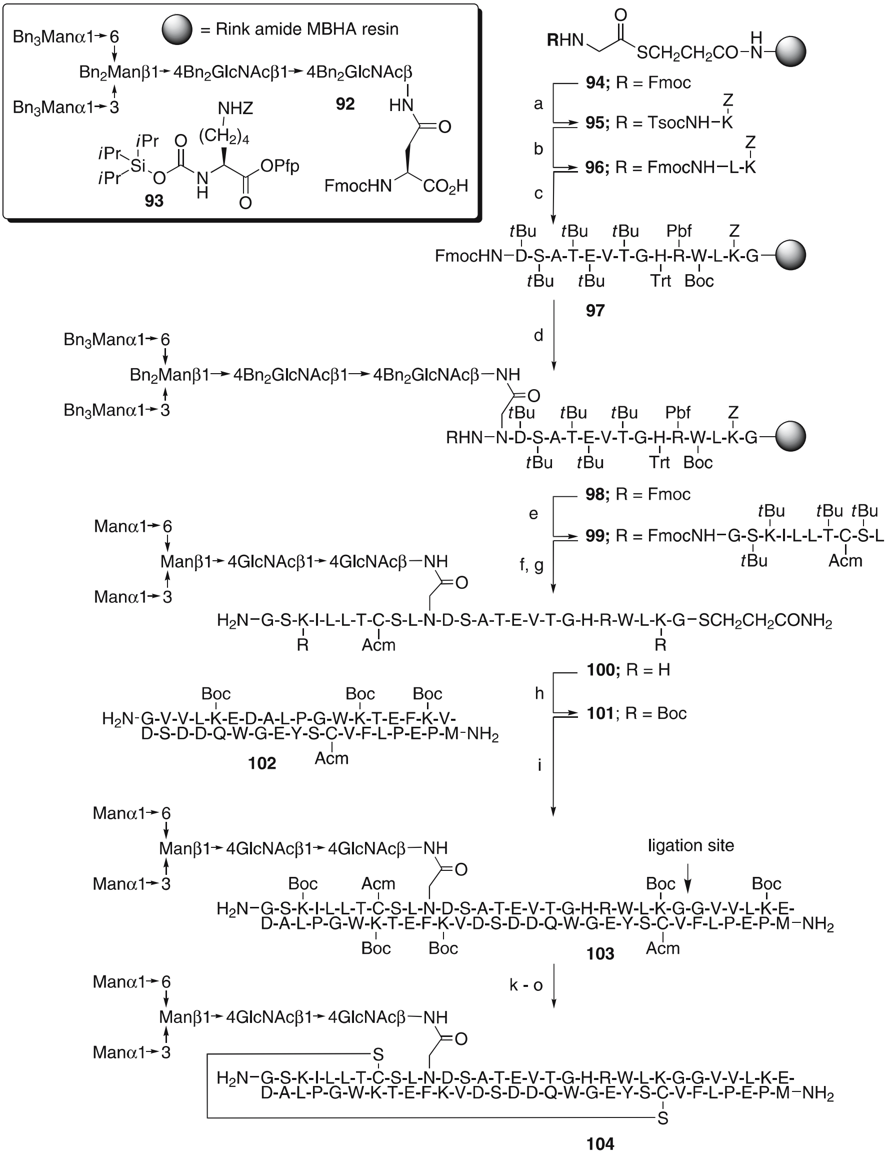


Scheme 15 (a) **86** (2 equiv.), HOBt, DCC, 0 °C to RT, 24 h; (b) (i) 20% piperidine in NMP, RT, 2 h; (ii) Fmoc-Gln(Trt)-OH, HOBt, DCC, NMP, RT, 2 h; (iii) 20% piperidine in NMP, RT, 2 h; (c) Ac-Gly-OH, HOBt, DCC, NMP, RT, 2 h; (d) HOAc, TFE, DCM (1 : 1 : 8), RT, 2 h, 70% four steps; (e) **87**, HOBt, EDC, DCM–NMP (2 : 1), RT, 1 d, 70%; f) 10% Pd/C, H_2, $CHCl_3$ – MeOH – H_2O (10 : 10 : 3), 2 d; g) TFA, TIS, DCM then HPLC, 85% (two steps)

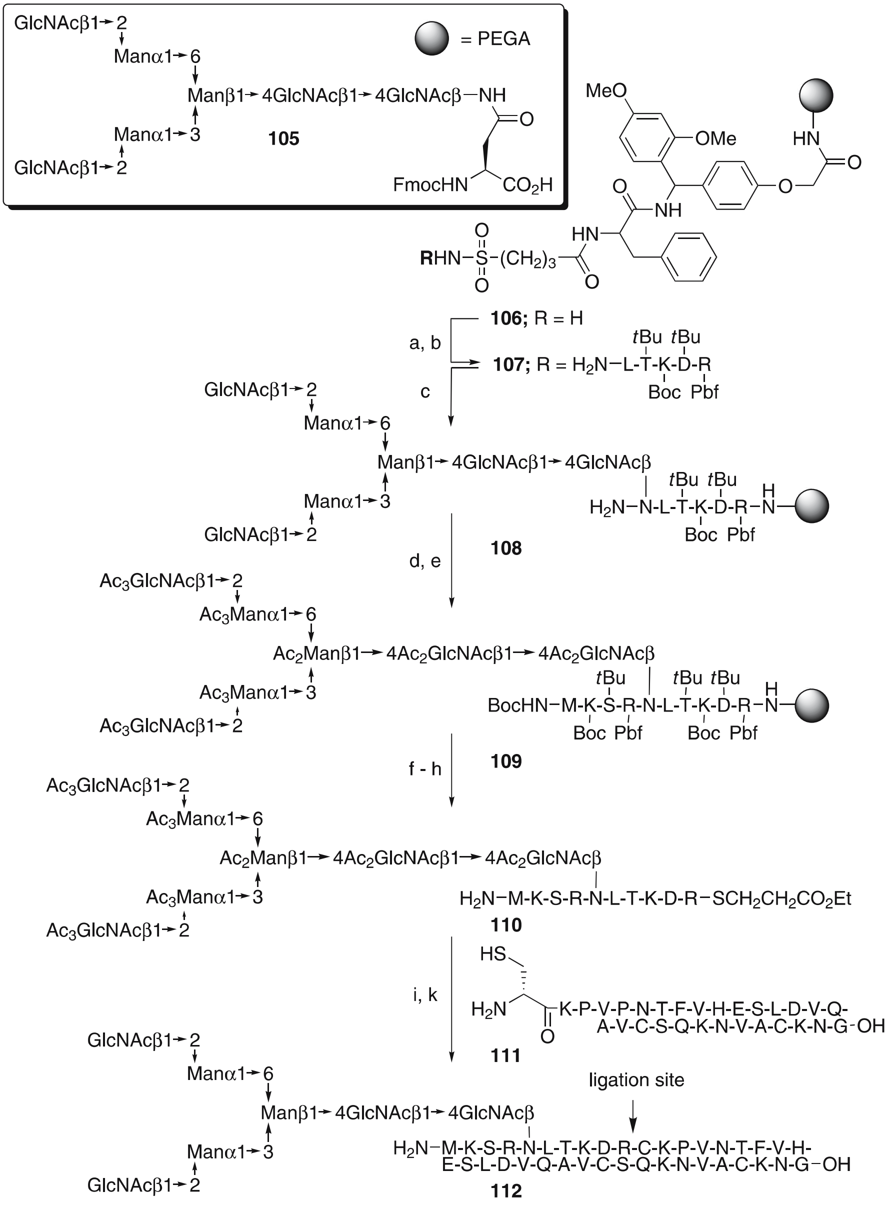
coupling of **101** with **102** furnished **103**. After the ligation the Boc- and the Acm-groups were removed. Air oxidation provided cyclic disulfide **104** [211].

In native chemical ligation reactions (see Schemes 10–12), the re-protection of amino groups of the thioester fragment can be omitted. To access glycopeptide thioesters such as the 38mer complex type *N*-glycopeptide **112** by Fmoc-based solid-phase synthesis. Unverzagt and others used the safety-catch sulfonamide linker [180, 181, 212]. An alternative silyl linker-based approach was reported by Nakahara [213].

The double-linker strategy shown in Scheme 17 facilitated the monitoring of the reaction steps. Interestingly, it was found that the *N*-acylsulfonamide structure **108** was subject to *N*-acetylation during capping. To avoid this side reaction and the resulting peptide losses capping was performed with acetic acid buffered Ac_2O/pyridine mixtures. The glycosyl asparagine **105** was introduced in the *O*-unprotected form. However, exposure to subsequent capping steps led to *O*-acetylation of the glycan. After the linear chain assembly had



Scheme 16 (a) (i) Fmoc deprotection (1-methylpyrrolidine, hexamethyleneimine, HOBt); (ii) **93**; (b) Fmoc-Leu-F, TBAF (cat.); (c) (i) Fmoc deprotection; (ii) Fmoc-AA-OH, HBTU, HOBt, DIEA; (iii) repeat i-ii; (d) (i) Fmoc deprotection; (ii) **92**, HATU, DIEA; (iii) Z(2-Cl)-OSu (capping); (e) (i) Fmoc deprotection; (ii) Fmoc-AA-OPfp; (iii) repeat i-ii; (f) TFA–thioanisole–H_2O–phenol–EDT (82.5 : 5 : 5 : 5 : 2.5); (g) TfOH–TFA–DMS–*m*-cresol (1 : 5 : 3 : 1), – 10 °C, 2 h, 1.8% from **94**; (h) Boc-OSu, 100%; (i) **102**, AgCl, HOOBt; (k) TFA, 5% EDT; (l) $AgNO_3$, DIEA/DMSO; (m) DTT, HCl; (n) RP-HPLC; (o) air oxidation, 32% from **100**



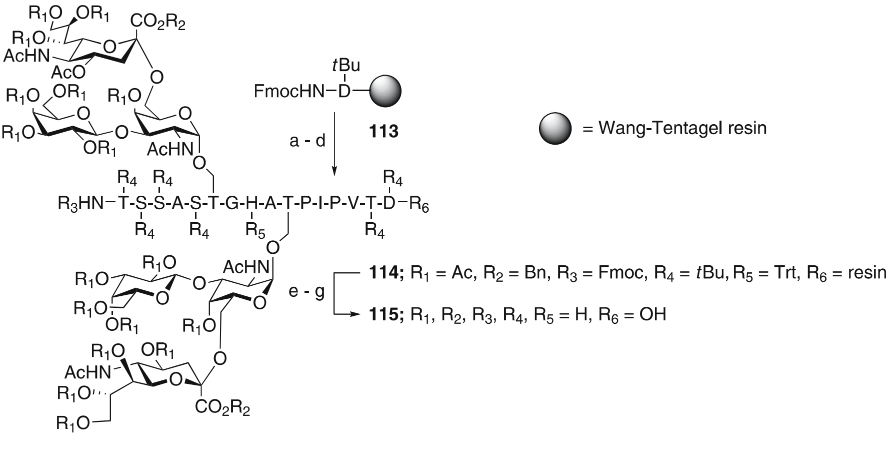
Scheme 17 (a) Fmoc-Arg(Pbf)-OH, DIC, 1-MeIm, CH_2Cl_2, DMF; (b) (i) 20% piperidine in DMF; (ii) Fmoc-AA-OH, HOBt, TBTU, DIPEA; (iii) repeat i–ii; (iv) 20% piperidine in DMF; (c) **105**, PyBOP, DIPEA, DMSO, NMP; (d) Ac_2O, pyridine, HOAc; (e) (i) Fmoc deprotection; (ii) Fmoc-AA-OH and Boc-Met-OH, respectively, HOBt, TBTU, DIPEA; (iii) repeat i–ii; (f) Me_3SiCHN_2, *n*-hexane, CH_2Cl_2; (g) ethyl 3-sulfanylpropionate, sodium thiophenolate, DMF; (h) TFA, H_2O, Et_3SiH, ethyl 3-sulfanylpropionate, rt., 2 h (46% from **105**); (i) **111** (1.3 equiv), 6 M guanidinium chloride, thiophenol (2%), phosphate buffer (pH 7.6); (k) hydrazine hydrate (10%), DTT (5%), 100% (two steps)

been completed, the safety-catch linker was activated for nucleophilic cleavage by treatment with $TMSCHN_2$. Thiolysis and deprotection of the amino acid side chains furnished **110**. The glycopeptide thioester was then subjected to a native chemical ligation with **111**. The deacetylation of the *N*-linked glycan moiety yielded the RNase-B-fragment **112** [214].

6.2
O-Glycopeptides

Glycan *O*-acetylation offers a convenient means to stabilize *O*-glycosidic bonds against electrophilic cleavage. Using glycosyl serine and threonine building blocks with *O*-acetyl protected glycans the Kunz group prepared a host of *O*-glycopeptides [135–139].

In the representative example shown in Scheme 18, a tandem repeat sequence of the epithelial mucin MUC4 was synthesized on Wang–Tentagel resin. The synthesis started from Fmoc-Asp(O*t*Bu) resin **113**. The chain assembly was performed by employing standard protocols of Fmoc solid-phase peptide synthesis. However, the excess of acylating agents was reduced from ten to two equivalents in the coupling of the precious glycosyl amino acid building blocks. After the TFA treatment, performed to simultaneously cleave the Wang-linker and the amino acid side chain protection groups, the glycan moiety was deblocked. The sialic acid benzyl ester groups were removed by hydrogenolysis followed by mild acetate hydrolysis at pH 9 [139]. Recently, Nishimura and co-workers demonstrated the beneficial effects of microwave irradiation in the solid-phase synthesis of MUC1-fragments [215]. Only 7 h



Scheme 18 (a) piperidine (20%), NMP; b) HBTU, HOBt, DIPEA, Fmoc-AA-OH (10 equiv.) or Fmoc-T(glycan)-OH (2 equiv.); c) cat. HOBt, Ac_2O, DIPEA, NMP; d) repeat a–c; e) TFA, iPr_3SiH, H_2O (17 : 1 : 1), 28% overall yield; f) cat. Pd/C, H_2, MeOH, 84%; g) NaOMe, MeOH, pH 9, 81%

were needed to complete the synthesis of a 20-residue glycopeptide featuring five *O*-linked mucin-type core-2 glycans.

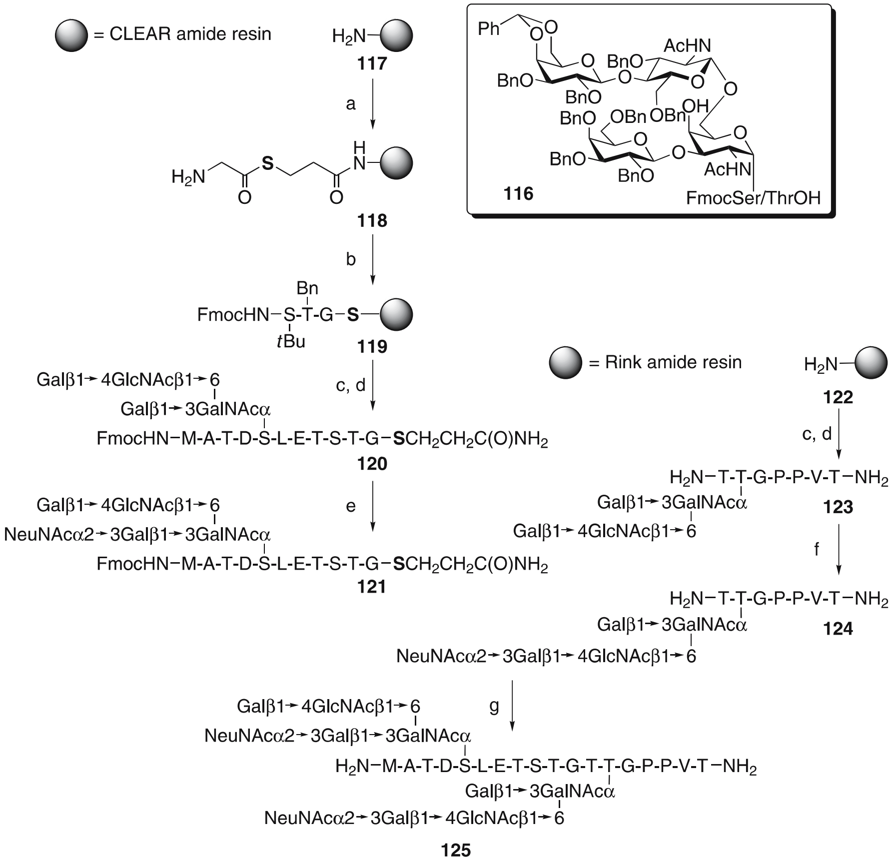
Benzyl-protected *O*-linked glycans were explored by Nakahara and coworkers [216–225]. In the chemoenzymatic synthesis of a segment of the human T-lymphocyte transmembrane glycoprotein leukosialin (CD 43) a core 2 *O*-hexasaccharide-containing decapeptide was prepared first by coupling a benzyl-protected glycosylthreonine. After global deprotection, a combination of two sialyltransferases, β-Gal-β1,3/4-GlcNAc-α2,3-sialyltransferase and β-Gal-β1,3-GlcNAc-α2,3-sialyltransferase was used to enzymatically add two sialic acid residues [226].

For the synthesis of a mimic for the heterogeneous surface of a core 2 sialoglycan-linked glycoprotein enzymatic carbohydrate synthesis and chemical fragment condensation were combined. First glycopeptide thioester **120** was prepared according to the modified Fmoc strategy described above (see Scheme 16) on a CLEAR amide resin (Scheme 19) [227]. After deprotection and release from the solid phase, the glycan moiety of **120** was extended enzymatically. The *C*-terminal glycopeptide fragment **124** was synthesized by a similar reaction sequence which involved classical Fmoc solid-phase synthesis to afford **123** and subsequent enzymatic sialylation. Then the thioester ligation method was applied to join **121**, activated by HOOBt/silver chloride, with **124** to afford **125** [228]. Sometimes the complete removal of multiple benzyl groups can be an arduous task. For this reason, it may be preferable to use *O*-unprotected glycans. Nakahara et al. demonstrated that the thioester ligation method can also be applied on the solid phase with resin-bound glycopeptides with *O*-unprotected carbohydrates [229]. Hojo demonstrated the synthesis of the EGF-like domain of blood coagulation factor IX relying on unprotected building blocks [230, 231].

In most chemoenzymatic approaches, solid-phase synthesis serves to provide soluble glycopeptide precursors for further enzymatic solution-phase elaboration of the carbohydrate or peptide part. This is mainly due to the intolerance of the biocatalyst towards many of the commonly used supports. Withers described a strategy in which so-called glycosynthases were applied in an on-resin glycosylation reaction. Glycosynthases are mutated glycosidases with the active-site carboxylate nucleophile replaced by a nonnucleophilic side chain. Thereby the ability to hydrolyze glycosidic bonds is lost, but the ability to catalyze glycosylations of sugar acceptors by using glycosyl fluoride donors is maintained.

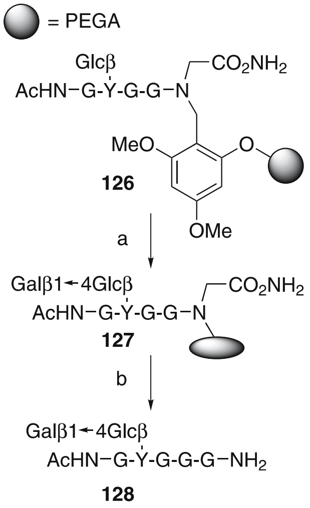
The mutant of *Agrobacterium* sp. β-glucosidase Abg E358G was shown to catalyze the reaction of α-D-galactosyl fluoride with resin-bound glycopeptide **126** (Scheme 20) [232]. Recently, the activity and substrate repertoire of the Abg E358G mutant's was expanded by directed evolution [233].

An interesting approach of achieving enzymatic on-resin glycosylations was presented by Nishimura et al. In their synthesis of a sialyl Lewis X bearing dodecapeptide, photocleavable ketone **129** was appended to resin-bound

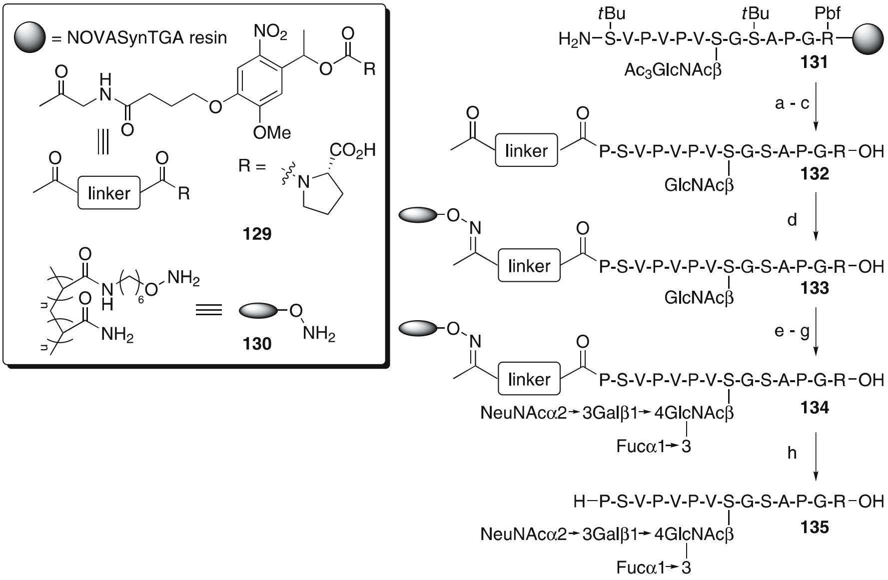


Scheme 19 (a) (i) Fmoc – Gly – $SCH_2CH_2CO_2H$, DCC, HOBt; (ii) 1-methylpyrrolidine, hexamethyleneimine, HOBt; (b) (i) Tsoc-T(Bn)-OPfp; (ii) Fmoc-S(*t*Bu)-F, Bu_4NF; (c) (i) 1-methylpyrrolidine, hexamethyleneimine, HOBt; (ii) Fmoc-AA-OH, DCC, HOBt or **116**, HATU, DIEA; (iii) repeat i–ii; d) (i) TFA–thioanisole–H_2O–phenol–EDT (82.5 : 5 : 5 : 5 : 2.5); (ii) DMS, *m*-cresol, EDT, TFA, then TfOH; (e) β-Gal-β1,3-GalNAc-α2,3-Sialyltransferase, CMP-sialic acid, BSA, cacodylate buffer (pH 6.0); (f) β-Gal-β1,3/4-GalNAc-α2,3-Sialyltransferase, CMP-sialic acid, BSA, MOPS buffer (pH 7.4); (g) **121**, AgCl, HOOBt, DMSO, 16% overall yield

glycopeptide **131** (Scheme 21). The glycopeptide was deblocked, detached and then attached to alkoxyamino-functionalized polyacrylamide **130** by means of oxim formation. Remarkably, chain-terminated capped products remained in solution since they lacked the terminal ketone group. As a result, only the desired full-length glycopeptide **133** was bound to the polymer. Analogous to a strategy introduced by Schuster, Seitz and Wong, a set of three different glycosyltransferase-mediated reactions provided the SLe^X-tetrasaccharide-containing peptide **134** [121, 234]. UV-light irradiation released the target



Scheme 20 (a) Abg E358G (1.0 mg/ml), α-D-galactosyl fluoride (20 mM), NaP$_i$ (100 mM, pH 9.0); (b) TFA, H_2O (19 : 1), 33% overall yield



Scheme 21 (a) (i) **129**, HBTU, HOBt, DIPEA, NMP/DMF (ii) Ac_2O, HOBt, DIPEA, DMF; (b) TFA, H_2O (9 : 1); (c) NaOH, MeOH; (d) **130**, acetate-buffer (pH 5.0–5.5); (e) UDP-Gal (5 mM), β-1,4-galactosyltransferase (100 mU mL^{-1}), $MnCl_2$, 0.1% BSA, HEPES buffer (50 mM, pH 7.0), 100%; (f) CMP-NeuNAc (5 mM), α-2,3-sialyltransferase (37 mU mL^{-1}), 0.1% BSA, HEPES buffer (50 mM, pH 7.0), 100%; (g) GDP-Fuc (5–7 mM), α-1,3-fucosyltransferase (20–30 mU mL^{-1}), $MnCl_2$ (10 mM), 0.1% BSA, HEPES buffer (50 mM, pH 7.0); h) UV irradiation (365 nm), 12% overall yield

compound **135** from the polymer in 12% overall yield [235]. To simplify the attachment onto the water-soluble polyacrylamide polymer, a more practical technique was developed. A *B. licheniformis* glutamic acid-specific protease cleavable linker was assembled of a Glu, a Phe and a 5-oxohexanoic acid building block replacing the UV-labile nitro-veratroyl brigde in **132**. This procedure gained fast access to a set of diverse and complex glycopeptides using parallel and combinatorial synthesis, respectively [236]. The combinatorial strategy reported by Nakahara and Narimatsu accelerates the elaboration of glycoconjugates libraries. Glycopeptides were constructed by an iterative repetition of adding glycosyltransferases and their corresponding glycosyl donors and the inactivation of the enzymes [237].

7
Glycopeptide Synthesis and the Expanded Genetic Code

Recently, the in vivo suppressor *t*RNA technology [238] has been exploited as a new technique for glycoprotein synthesis. To incorporate glycosylated amino acids by using an expanded genetic code an orthogonal *t*RNA synthetase-*t*RNA pair from *M. jannaschii* has been evolved and inserted into the *E. coli* genome. The *t*RNA synthetase is able to accept and charge the unnatural amino acid onto the Amber-suppressing $t\mathrm{RNA}_{\mathrm{CUA}}$ [239–242]. In addition, a permissible TAG-Amber stop codon introduced into the protein of interest served the site-directed incorporation of the desired carbohydrate building block. Due to better membrane permeability, peracetylated glycosyl amino acids were offered for in vivo protein synthesis. The acetyl protection was sheared off by unspecific intracellular esterases and the modified translation machinery of *E. coli* expressed myoglobin-variants bearing α-*O*-linked GalNAc [243] and β-*O*-linked GlcNAc residues [244].

8
Summary and Outlook

The methodology of glycopeptide synthesis has improved to a state at which the preparation of glycopeptides with a degree of complexity that meets many of the needs of biological and medicinal research is a doable task. For the synthesis of complex *N*-glycopeptides, the carbohydrate part can be coupled chemically or enzymatically as one block to a peptide or a glycopeptide acceptor. Though elegant and rapid, the preparation of complex *N*-glycopeptides is limited by the availability of the complex *N*-glycan part which has to be provided from natural sources or through challenging carbohydrate synthesis. However, the recent progress that has been brought about by minimal protection and sequential one-pot glycosylations as well as the improved availability of glycosyltransferases and glycosidesynthases has facilitated the preparation of complex glycan structures. The synthesis of *O*-glycopeptides is most successfully performed by assembling preformed glycosyl amino acid building blocks. With glycosyl amino acid building blocks in hand, solid-phase synthesis allows rapid and convenient access to various glycopeptides.

The development of convergent peptide ligation techniques was a significant milestone, which provided access to complex glycoproteins. At present, it appears that the "Native Chemical Ligation" and "Expressed Protein Ligation" are the most powerful methods for such a fragment coupling of two unprotected peptide segments prepared by chemical and/or enzymatical means. The presented examples illustrate that smart combinations of chemical and enzymatic synthesis with solid-phase techniques and convergent ligation strategies allow a significant shortening of the often time-consuming synthesis of complex glycopeptides, which are essential for the advancement of molecular glycobiology. However, the synthesis of very complex targets is still a challenge and it usually is the speed of glycan synthesis that is rate limiting. Here, solid-phase carbohydrate synthesis promises a solution to one of the most difficult tasks of glycoscience, that is to match the different time scales of chemical and biological research.

References

1. Seitz O (2000) ChemBioChem 1:215
2. Seitz O, Heinemann I, Mattes A, Waldmann H (2001) Tetrahedron 57:2247
3. Brocke C, Kunz H (2002) Bioorg Med Chem 10:3085
4. Grogan MJ, Pratt MR, Marcaurelle LA, Bertozzi CR (2002) Annu Rev Biochem 71:593
5. Hanson S, Best M, Bryan MC, Wong CH (2004) Trends Biochem Sci 29:656
6. Holeman A, Seeberger PH (2004) Curr Opin Biotechnol 15:615
7. Hang HC, Bertozzi CR (2005) Bioorg Med Chem 13:5021
8. Kajihara Y, Yamamoto N, Miyazaki T, Sato H (2005) Curr Med Chem 12:527
9. Pratt MR, Bertozzi CR (2005) Chem Soc Rev 34:58
10. Wong C-H (2005) J Org Chem 70:4219
11. Butler WT, Cunningh LW (1966) J Biol Chem 241:3882
12. Lamport DTA (1967) Nature 216:1322
13. Rodriguez IR, Whelan WJ (1985) Biochem Biophys Res Commun 132:829
14. Hofsteenge J, Müller DR, Debeer T, Löffler A, Richter WJ, Vliegenthart JFG (1994) Biochemistry 33:13524
15. Hart GW, Brew K, Grant GA, Bradshaw RA, Lennarz WJ (1979) J Biol Chem 254:9747
16. Zachara NE, Hart GW (2002) Chem Rev 102:431
17. Wells L, Vosseller K, Hart GW (2001) Science 291:2376
18. Spiro RG (1967) J Biol Chem 242:4813
19. Hart GW (1997) Annu Rev Biochem 66:315
20. Vliegenthart JFG, Casset F (1998) Curr Opin Struct Biol 8:565
21. Endo T (1999) Biochim Biophys Acta, Gen Subj 1473:237
22. Haltiwanger RS (2002) Curr Opin Struct Biol 12:593

23. Haines N, Irvine KD (2003) Nat Rev Mol Cell Biol 4:786
24. Shao L, Haltiwanger RS (2003) Cell Mol Life Sci 60:241
25. Upreti RK, Kumar M, Shankar V (2003) Proteomics 3:363
26. Lomako J, Lomako WM, Whelan WJ (2004) Biochim Biophys Acta, Gen Subj 1673:45
27. Szymanski CM, Wren BW (2005) Nat Rev Microbiol 3:225
28. Helenius A, Aebi M (2001) Science 291:2364
29. Cannon KS, Helenius A (1999) J Biol Chem 274:7537
30. Huppa JB, Ploegh HL (1998) Cell 92:145
31. Hammond C, Braakman I, Helenius A (1994) Proc Natl Acad Sci USA 91:913
32. Fukuda MN, Sasaki H, Lopez L, Fukuda M (1989) Blood 73:84
33. Wyss DF, Choi JS, Li J, Knoppers MH, Willis KJ, Arulanandam ARN, Smolyar A, Reinherz EL, Wagner G (1995) Science 269:1273
34. Szymanski CM, Burr DH, Guerry P (2002) Infect Immun 70:2242
35. Larsen JC, Szymanski C, Guerry P (2004) J Bacteriol 186:6508
36. Arora SK, Neely AN, Blair B, Lory S, Ramphal R (2005) Infect Immun 73:4395
37. Tanaka S, Meiselman HHJ, Engel E, Guth PH, Furukawa O, Wenby RB, Lee J, Kaunitz JD (2002) Dig Dis Sci 47:967
38. Fujita T, Ohara S, Sugaya T, Saigenji K, Hotta K (2000) Comp Biochem Physiol B Biochem Mol Biol 126:353
39. Cao XX, Bansil R, Bhaskar KR, Turner BS, LaMont JT, Niu N, Afdhal NH (1999) Biophys J 76:1250
40. Tanaka S, Podolsky DK, Engel E, Guth PH, Kaunitz JD (1997) Am J Physiol, Gastroint Liver Physiol 276:G1473
41. Nordman H, Davies JR, Lindell G, Carlstedt I (1995) Biochem Soc Trans 23:S533
42. Ho SB, Roberton AM, Shekels LL, Lyftogt CT, Niehans GA, Toribara NW (1995) Gastroenterology 109:735
43. Bhaskar KR, Garik P, Turner BS, Bradley JD, Bansil R, Stanley HE, Lamont JT (1992) Nature 360:458
44. Jepson S, Komatsu M, Haq B, Arango ME, Huang DM, Carraway CAC, Carraway KL (2002) Oncogene 21:7524
45. McDermott KM, Crocker PR, Harris A, Burdick MD, Hinoda Y, Hayashi T, Imai K, Hollingsworth MA (2001) Int J Cancer 94:783
46. Hayashi T, Takahashi T, Motoya S, Ishida T, Itoh F, Adachi M, Hinoda Y, Imai K (2001) Digestion 63:87
47. Brinkman-Van der Linden ECM, Varki A (2000) J Biol Chem 275:8625
48. Borsig L, Wong R, Feramisco J, Nadeau DR, Varki NM, Varki A (2001) Proc Natl Acad Sci USA 98:3352
49. Nath D, Hartnell A, Happerfield L, Miles DW, Burchell J, Taylor-Papadimitriou J, Crocker PR (1999) Immunology 98:213
50. Kim YJ, Borsig L, Han HL, Varki NM, Varki A (1999) Am J Pathol 155:461
51. Chan AK, Lockhart DC, von Bernstorff W, Spanjaard RA, Joo HG, Eberlein TJ, Goedegebuure PS (1999) Int J Cancer 82:721
52. Komatsu M, Yee L, Carraway KL (1999) Cancer Res 59:2229
53. Agrawal B, Gendler SJ, Longenecker BM (1998) Mol Med Today 4:397
54. Agrawal B, Krantz MJ, Reddish MA, Longenecker BM (1998) Nat Med 4:43
55. Van de Wiel-van Kemenade E, Ligtenberg MJL, Deboer AJ, Buijs F, Vos HL, Melief CJM, Hilkens J, Figdor CG (1993) J Immunol 151:767
56. Lüttges J, Feyerabend B, Buchelt T, Pacena M, Kloppel G (2002) Am J Surg Pathol 26:466
57. Rahn JJ, Dabbagh L, Pasdar M, Hugh JC (2001) Cancer 91:1973

58. Kashiwagi H, Kijima H, Dowaki S, Ohtani Y, Tobita K, Tsukui M, Tanaka Y, Matsubayasi H, Tsuchida T, Yamazaki H, Nakamura M, Ueyama Y, Tanaka M, Tajima T, Makuuchi H (2000) Int J Oncol 16:455
59. Nitta T, Sugihara K, Tsuyama S, Murata F (2000) Cancer 88:245
60. Hiraga Y, Tanaka S, Haruma K, Yoshihara M, Sumii K, Kajiyama G, Shimamoto F, Kohno N (1998) Oncology 55:307
61. Kondo K, Kohno N, Yokoyama A, Hiwada K (1998) Cancer Res 58:2014
62. Spicer AP, Rowse GJ, Lidner TK, Gendler SJ (1995) J Biol Chem 270:30093
63. Nakamori S, Ota DM, Cleary KR, Shirotani K, Irimura T (1994) Gastroenterology 106:353
64. Ligtenberg MJL, Buijs F, Vos HL, Hilkens J (1992) Cancer Res 52:2318
65. Wesseling J, Van der Valk SW, Hilkens J (1996) Mol Biol Cell 7:565
66. Wesseling J, Van der Valk SW, Vos HL, Sonnenberg A, Hilkens J (1995) J Cell Biol 129:255
67. Guinez C, Morelle W, Michalski JC, Lefebvre T (2005) Int J Biochem Cell Biol 37:765
68. Whelan SA, Hart GW (2003) Circ Res 93:1047
69. Khidekel N, Hsieh-Wilson LC (2004) Org Biomol Chem 2:1
70. Wells L, Whalen SA, Hart GW (2003) Biochem Biophys Res Commun 302:435
71. Clark RS, Banerjee S, Coward JK (1990) J Org Chem 55:6275
72. Bolton CH, Jeanloz RW (1963) J Org Chem 28:3228
73. Szilagyi L, Gyorgydeak Z (1985) Carbohydr Res 143:21
74. Augé C, Gautheron C, Pora H (1989) Carbohydr Res 193:288
75. Sabesan S, Neira S (1992) Carbohydr Res 223:169
76. Tropper FD, Andersson FO, Braun S, Roy R (1992) Synthesis, p 618
77. Unverzagt C, Kunz H (1992) J Prakt Chem 334:570
78. Li C, Arasappan A, Fuchs PL (1993) Tetrahedron Lett 34:3535
79. Li C, Shih TL, Jeong JU, Arasappan A, Fuchs PL (1994) Tetrahedron Lett 35:2645
80. Matsubara K, Mukaiyama T (1994) Chem Lett:247
81. Soli ED, DeShong P (1999) J Org Chem 64:9724
82. McDonald FE, Danishefsky SJ (1992) J Org Chem 57:7001
83. Danishefsky SJ, Roberge JY (1995) Pure Appl Chem 67:1647
84. He Y, Hinklin RJ, Chang JY, Kiessling LL (2004) Org Lett 6:4479
85. Tanaka H, Iwata Y, Takahashi D, Adachi M, Takahashi T (2005) J Am Chem Soc 127:1630
86. Cohen-Anisfeld ST, Lansbury PT (1993) J Am Chem Soc 115:10531
87. Likhosherstov LM, Novikova OS, Derevitskaja VA, Kochetkov NK (1986) Carbohydr Res 146:C1
88. Likhosherstov LM, Novikova OS, Zheltova AO, Shibaev VN (2004) Russ Chem Bull 53:709
89. Likhosherstov LM, Novikova OS, Shibaev VN (2003) Dokl Chem 389:73
90. Likhosherstov LM, Novikova OS, Shibaev VN (2002) Dokl Chem 383:89
91. Bejugam M, Flitsch SL (2004) Org Lett 6:4001
92. Meinjohanns E, Meldal M, Paulsen H, Dwek RA, Bock K (1998) J Chem Soc, Perkin Trans 1:549
93. Hackenberger CPR, O'Reilly MK, Imperiali B (2005) J Org Chem 70:3574
94. Seko A, Koketsu M, Nishizono M, Enoki Y, Ibrahim HR, Juneja LR, Kim M, Yamamoto T (1997) Biochim Biophys Acta, Gen Subj 1335:23
95. Yamamoto N, Ohmori Y, Sakakibara T, Sasaki K, Juneja LR, Kajihara Y (2003) Angew Chem, Int Ed 42:2537

96. Kajihara Y, Suzuki Y, Yamamoto N, Sasaki K, Sakakibara T, Juneja LR (2004) Chem-Eur J 10:971
97. Fukae K, Yamamoto N, Hatakeyama Y, Kajihara Y (2004) Glycoconjugate J 21:243
98. Iijima H, Ogawa T (1989) Carbohydr Res 186:107
99. Paulsen H, Adermann K (1989) Liebigs Ann Chem:751
100. Bielfeldt T, Peters S, Meldal M, Bock K, Paulsen H (1992) Angew Chem, Int Ed Engl 31:857
101. Paulsen H, Peters S, Bielfeldt T, Meldal M, Bock K (1995) Carbohydr Res 268:17
102. Rio-Anneheim S, Paulsen H, Meldal M, Bock K (1995) J Chem Soc, Perkin Trans 1:1071
103. Vuljanic T, Bergquist KE, Clausen H, Roy S, Kihlberg J (1996) Tetrahedron 52:7983
104. Leuck M, Kunz H (1997) J Prakt Chem 339:322
105. Mathieux N, Paulsen H, Meldal M, Bock K (1997) J Chem Soc, Perkin Trans 1:2359
106. Sames D, Chen X-T, Danishefsky SJ (1997) Nature 389:587
107. Chen X-T, Danishefsky SJ, Sames D (1998) J Am Chem Soc 120:7760
108. Satyanarayana J, Gururaja TL, Naganagowda GA, Ramasubbu N, Levine MJ (1998) J Pept Res 52:165
109. Schwarz JB, Kuduk SD, Chen X-T, Sames D, Glunz PW, Danishefsky SJ (1999) J Am Chem Soc 121:2662
110. Danishefsky SJ, Allen JR (2000) Angew Chem, Int Ed Engl 39:836
111. Kunz H, Birnbach S (1986) Angew Chem, Int Ed Engl 25:360
112. Paulsen H, Bielfeldt T, Peters S, Meldal M, Bock K (1994) Liebigs Ann Chem:369
113. Kuduk SD, Schwarz JB, Chen XT, Glunz PW, Sames D, Ragupathi G, Livingston PO, Danishefsky SJ (1998) J Am Chem Soc 120:12474
114. Liebe B, Kunz H (1997) Angew Chem, Int Ed Engl 36:618
115. Shao N, Guo Z (2005) Org Lett 7:3589
116. Lemieux RU, Ratcliffe RM (1979) Can J Chem 57:1244
117. Winterfeld GA, Ito Y, Ogawa T, Schmidt RR (1999) Eur J Org Chem:1167
118. Winterfeld GA, Schmidt RR (2001) Angew Chem, Int Ed 40:2654
119. Winterfeld GA, Khodair AI, Schmidt RR (2003) Eur J Org Chem:1009
120. Geiger J, Barroca N, Schmidt RR (2004) Synlett, p 836
121. Seitz O, Wong CH (1997) J Am Chem Soc 119:8766
122. Pohl T, Waldmann H (1997) J Am Chem Soc 119:6702
123. Salvador LA, Elofsson M, Kihlberg J (1995) Tetrahedron 51:5643
124. Arsequell G, Krippner L, Dwek RA, Wong SYC (1994) J Chem Soc, Chem Commun, p 2383
125. Elofsson M, Walse B, Kihlberg J (1991) Tetrahedron Lett 32:7613
126. Schultz M, Kunz H (1992) Tetrahedron Lett 33:5319
127. Vargas-Berenguel A, Meldal M, Paulsen H, Bock K (1994) J Chem Soc, Perkin Trans 1:2615
128. Meinjohanns E, Meldal M, Bock K (1995) Tetrahedron Lett 36:9205
129. Meinjohanns E, Vargas-Berenguel A, Meldal M, Paulsen H, Bock K (1995) J Chem Soc, Perkin Trans 1:2165
130. Jensen KJ, Hansen PR, Venugopal D, Barany G (1996) J Am Chem Soc 118:3148
131. Saha UK, Schmidt RR (1997) J Chem Soc, Perkin Trans 1:1855
132. Seibel R, Hillringhaus L, Moraru R (2005) Carbohydr Res 340:507
133. Wittmann V, Lennartz D (2002) Eur J Org Chem:1363
134. Wagner M, Dziadek S, Kunz H (2003) Chem-Eur J 9:6018
135. Dziadek S, Brocke C, Kunz H (2004) Chem-Eur J 10:4150

136. Reipen T, Kunz H (2003) Synthesis, p 2487
137. Brocke C, Kunz H (2003) Synlett, p 2052
138. Dziadek S, Kunz H (2003) Synlett, p 1623
139. Brocke C, Kunz H (2004) Synthesis, p 525
140. Adachi M, Tanaka H, Takahashi T (2004) Synlett, p 609
141. Tanaka H, Adachi M, Takahashi T (2005) Chem-Eur J 11:849
142. Blixt O, Allin K, Pereira L, Datta A, Paulson JC (2002) J Am Chem Soc 124:5739
143. Dudziak G, Bézay N, Schwientek T, Clausen H, Kunz H, Liese A (2000) Tetrahedron 56:5865
144. Bézay N, Dudziak G, Liese A, Kunz H (2001) Angew Chem, Int Ed 40:2292
145. Kunz H (2003) J Pept Sci 9:563
146. Cantacuzene D, Attal S, Bay S (1991) Bioorg Med Chem Lett 1:197
147. Cantacuzene D, Attal S (1991) Carbohydr Res 211:327
148. Nilsson KGI, Scigelova M (1994) Biotechnol Lett 16:671
149. Nilsson KGI, Ljunger G, Melin PM (1997) Biotechnol Lett 19:889
150. Varon D, Lioy E, Patarroyo ME, Schratt X, Unverzagt C (2002) Aust J Chem 55:161
151. Broddefalk J, Backlund J, Almqvist F, Johansson M, Holmdahl R, Kihlberg J (1998) J Am Chem Soc 120:7676
152. Broddefalk J, Forsgren M, Sethson I, Kihlberg J (1999) J Org Chem 64:8948
153. Holm B, Broddefalk J, Flodell S, Wellner E, Kihlberg J (2000) Tetrahedron 56:1579
154. Syed BM, Gustafsson T, Kihlberg J (2004) Tetrahedron 60:5571
155. Cudic M, Lauer-Fields JL, Fields GB (2005) J Pept Res 65:272
156. Marin J, Violette A, Briand JP, Guichard G (2004) Eur J Org Chem:3027
157. Bäcklund J, Treschow A, Bockermann R, Holm B, Holm L, Issazadeh-Navikas S, Kihlberg J, Holmdahl R (2002) Eur J Immunol 32:3776
158. Holm B, Bäcklund J, Recio MAF, Holmdahl R, Kihlberg J (2002) ChemBioChem 3:1209
159. Holm B, Baquer SM, Holm L, Holmdahl R, Kihlberg J (2003) Bioorg Med Chem 11:3981
160. Dzhambazov B, Holmdahl M, Yamada H, Lu SM, Vestberg M, Holm B, Johnell O, Kihlberg J, Holmdahl R (2005) Eur J Immunol 35:357
161. Holm L, Kjellen P, Holmdahl R, Kihlberg J (2005) Bioorg Med Chem 13:473
162. Leppänen A, White SP, Helin J, McEver RP, Cummings RD (2000) J Biol Chem 275:39569
163. Koeller KM, Smith MEB, Wong CH (2000) Bioorg Med Chem 8:1017
164. Koeller KM, Smith MEB, Huang RF, Wong CH (2000) J Am Chem Soc 122:4241
165. Koeller KM, Smith MEB, Wong CH (2000) J Am Chem Soc 122:742
166. Wang LX, Song HJ, Liu SW, Lu H, Jiang SB, Ni JH, Li HG (2005) ChemBioChem 6:1068
167. Singh S, Ni JH, Wang LX (2003) Bioorg Med Chem Lett 13:327
168. Li HG, Singh S, Zeng Y, Song HJ, Wang LX (2005) Bioorg Med Chem Lett 15:895
169. Fujita M, Shoda S, Haneda K, Inazu T, Takegawa K, Yamamoto K (2001) Biochim Biophys Acta, Gen Subj 1528:9
170. Li B, Zeng Y, Hauser S, Song H, Wang L-X (2005) J Am Chem Soc 127:9692
171. Witte K, Seitz O, Wong CH (1998) J Am Chem Soc 120:1979
172. Doores KJ, Davis BG (2005) Chem Commun, p 168
173. Wang ZG, Zhang XF, Visser M, Live D, Zatorski A, Iserloh U, Lloyd KO, Danishefsky SJ (2001) Angew Chem, Int Ed 40:1728
174. Dudkin VY, Miller JS, Danishefsky SJ (2003) Tetrahedron Lett 44:1791

175. Mandal M, Dudkin VY, Geng XD, Danishefsky S (2004) Angew Chem, Int Ed 43:2557
176. Geng XD, Dudkin VY, Mandal M, Danishefsky SJ (2004) Angew Chem, Int Ed 43:2562
177. Dudkin VY, Orlova M, Geng XD, Mandal M, Olson WC, Danishefsky SJ (2004) J Am Chem Soc 126:9560
178. Warren JD, Geng X, Danishefsky SJ (2006) Top Curr Chem, in press
179. Dawson PE, Muir TW, Clarklewis I, Kent SBH (1994) Science 266:776
180. Shin Y, Winans KA, Backes BJ, Kent SBH, Ellman JA, Bertozzi CR (1999) J Am Chem Soc 121:11684
181. Marcaurelle LA, Mizoue LS, Wilken J, Oldham L, Kent SBH, Handel TM, Bertozzi CR (2001) Chem-Eur J 7:1129
182. Dudkin VY, Miller JS, Danishefsky SJ (2004) J Am Chem Soc 126:736
183. Miller JS, Dudkin VY, Lyon GJ, Muir TW, Danishefsky SJ (2003) Angew Chem, Int Ed 42:431
184. Warren JD, Miller JS, Keding SJ, Danishefsky SJ (2004) J Am Chem Soc 126:6576
185. Tolbert TJ, Franke D, Wong CH (2005) Bioorg Med Chem 13:909
186. Hackenberger CPR, Friel CT, Radford SE, Imperiali B (2005) J Am Chem Soc 127:12882
187. Shao N, Xue J, Guo ZW (2003) J Org Chem 68:9003
188. Xue J, Guo ZW (2003) J Org Chem 68:2713
189. Wen S, Guo Z (2001) Org Lett 3:3773
190. Mizuno M, Goto K, Miura T (2005) Chem Lett 34:426
191. Glunz PW, Hintermann S, Williams LJ, Schwarz JB, Kuduk SD, Kudryashov V, Lloyd KO, Danishefsky SJ (2000) J Am Chem Soc 122:7273
192. Glunz PW, Hintermann S, Schwarz JB, Kuduk SD, Chen XT, Williams LJ, Sames D, Danishefsky SJ, Kudryashov V, Lloyd KO (1999) J Am Chem Soc 121:10636
193. Macmillan D, Bertozzi CR (2004) Angew Chem, Int Ed 43:1355
194. Macmillan D, Anderson DW (2004) Org Lett 6:4659
195. George SK, Schwientek T, Holm B, Reis CA, Clausen H, Kihlberg J (2001) J Am Chem Soc 123:11117
196. Izumi M, Shen G-J, Wacowich-Sgarbi S, Nakatani T, Plettenburg O, Wong C-H (2001) J Am Chem Soc 123:10909
197. Gutiérrez Gallego R, Kamerling JP, Vliegenthart JFG, Dudziak G, Kragl U, Wandrey C (2003) Biochimie 85:275
198. Teo CF, Hwang TS, Chen PH, Hung CH, Gao HS, Chang LS, Lin CH (2005) Adv Synth Catal 347:967
199. Wang SS (1973) J Am Chem Soc 95:1328
200. Sheppard RC, Williams BJ (1982) Int J Pept Protein Res 20:451
201. Rink H (1987) Tetrahedron Lett 28:3787
202. Albericio F, Kneibcordonier N, Biancalana S, Gera L, Masada RI, Hudson D, Barany G (1990) J Org Chem 55:3730
203. Mergler M, Tanner R, Gosteli J, Grogg P (1988) Tetrahedron Lett 29:4005
204. Floersheimer A, Riniker B (1991) Pept. 1990, Proc Eur Pept Symp, 21st:131
205. Sieber P (1987) Tetrahedron Lett 28:2107
206. Barlos K, Gatos D, Kallitsis J, Papaphotiu G, Sotiriu P, Yao WQ, Schäfer W (1989) Tetrahedron Lett 30:3943
207. Seitz O, Kunz H (1995) Angew Chem, Int Ed Engl 34:803
208. Shao N, Xue B, Guo ZW (2004) Angew Chem, Int Ed 43:1569
209. Guo HM, Zucker S, Gordon MK, Toole BP, Biswas C (1997) J Biol Chem 272:24

210. Li XQ, Kawakami T, Aimoto S (1998) Tetrahedron Lett 39:8669
211. Hojo H, Haginoya E, Matsumoto Y, Nakahara Y, Nabeshima K, Toole BP, Watanabe Y (2003) Tetrahedron Lett 44:2961
212. Backes BJ, Virgilio AA, Ellman JA (1996) J Am Chem Soc 118:3055
213. Ishii A, Hojo H, Nakahara Y, Ito Y, Nakaharai Y (2002) Biosci Biotechnol Biochem 66:225
214. Mezzato S, Schaffrath M, Unverzagt C (2005) Angew Chem, Int Ed 44:1650
215. Matsushita T, Hinou H, Kurogochi M, Shimizu H, Nishimura SI (2005) Org Lett 7:877
216. Nakahara Y, Ogawa T (1996) Carbohydr Res 292:71
217. Guo ZW, Nakahara Y, Ogawa T (1997) Carbohydr Res 303:373
218. Guo ZW, Nakahara Y, Ogawa T (1997) Bioorg Med Chem 5:1917
219. Nakahara Y, Ito Y, Ogawa T (1997) Tetrahedron Lett 38:7211
220. Nakahara Y, Ito Y, Ogawa T (1998) Carbohydr Res 309:287
221. Ando S, Nakahara Y, Ito Y, Ogawa T (2000) Carbohydr Res 329:773
222. Hojo H, Watabe J, Nakahara Y, Ito Y, Nabeshima K, Toole BP (2001) Tetrahedron Lett 42:3001
223. Maemura M, Ohgaki A, Nakahara Y, Hojo H (2005) Biosci Biotechnol Biochem 69:1575
224. Takano Y, Habiro M, Someya M, Hojo H, Nakahara Y (2002) Tetrahedron Lett 43:8395
225. Guo ZW, Nakahara Y, Ogawa T (1997) Angew Chem, Int Ed Engl 36:1464
226. Takano Y, Kojima N, Nakahara Y, Hojo H (2003) Tetrahedron 59:8415
227. Kempe M, Barany G (1996) J Am Chem Soc 118:7083
228. Takano Y, Hojo H, Kojima N, Nakahara Y (2004) Org Lett 6:3135
229. Ichiyanagi T, Takatani M, Sakamoto K, Nakahara Y, Ito Y, Hojo H (2002) Tetrahedron Lett 43:3297
230. Takemura T, Hojo H, Nakahara Y, Ishimizu T, Hase S (2004) Org Biomol Chem 2:133
231. Kitamura M, Hojo H, Nakahara Y, Ishimizu T, Hase S (2004) Glycoconjugate J 21:197
232. Tolborg JF, Petersen L, Jensen KJ, Mayer C, Jakeman DL, Warren RAJ, Withers SG (2002) J Org Chem 67:4143
233. Kim YW, Lee SS, Warren RAJ, Withers SG (2004) J Biol Chem 279:42787
234. Schuster M, Wang P, Paulson JC, Wong CH (1994) J Am Chem Soc 116:1135
235. Fumoto M, Hinou H, Matsushita T, Kurogochi M, Ohta T, Ito T, Yamada K, Takimoto A, Kondo H, Inazu T, Nishimura SI (2005) Angew Chem, Int Ed 44:2534
236. Fumoto M, Hinou H, Ohta T, Ito T, Yamada K, Takimoto A, Kondo H, Shimizu H, Inazu T, Nakahara Y, Nishimura SI (2005) J Am Chem Soc 127:11804
237. Ito H, Kameyana A, Sato T, Kiyohara K, Nakahara Y, Narimatsu H (2005) Angew Chem, Int Ed 44:4547
238. Wang L, Schultz PG (2005) Angew Chem, Int Ed 44:34
239. Wang L, Schultz PG (2001) Chem Biol 8:883
240. Wang L, Magliery TJ, Liu DR, Schultz PG (2000) J Am Chem Soc 122:5010
241. Wang L, Brock A, Herberich B, Schultz PG (2001) Science 292:498
242. Santoro SW, Wang L, Herberich B, King DS, Schultz PG (2002) Nat Biotechnol 20:1044
243. Xu R, Hanson SR, Zhang ZW, Yang YY, Schultz PG, Wong CH (2004) J Am Chem Soc 126:15654
244. Zhang ZW, Gildersleeve J, Yang YY, Xu R, Loo JA, Uryu S, Wong CH, Schultz PG (2004) Science 303:371

Top Curr Chem (2007) 267: 37–63
DOI 10.1007/128_032

Enzymatic Synthesis of Glycopeptides and Glycoproteins

Desiree A. Thayer · Chi-Huey Wong (✉)

Department of Chemistry and The Skaggs Institute for Chemical Biology,
The Scripps Research Institute, 10550 N. Torrey Pines Rd., La Jolla, California 92037,
USA
wong@scripps.edu

Abstract Glycosylation is recognized as important in biology, though the discovery of previously unknown functions associated with carbohydrates is ongoing. As the most complex and diverse post-translational modification, determining the structure-function relationships for glycosylation remains difficult due to the heterogeneous expression of glycoproteins as mixtures of glycoforms. Thus the synthesis of glycoproteins and glycopeptides with defined glycan structures is essential for detailed determination of the role of glycosylation. This review describes the use of enzymes for the synthesis of glycopeptides and glycoproteins, including their use as catalysts for glycosidic linkage formation and elaboration, peptide bond formation, and modification.

Keywords Glycopeptide · Glycoprotein · Enzymatic synthesis

1
Introduction

The role of carbohydrates in biological systems is an active area of research. Even though carbohydrates are established as important in biology, elucidation of the biological functions of carbohydrates remains a major challenge. Glycosylation is the most complex and diverse post-translational modification of proteins, with diversity both chemically and biologically. The chemical diversity of glycosylation originates from the ability to combine monosaccharides into glycans of varying length, sequence, and anomeric linkage (α or β) with a variety of glycosylation positions, branch points, and modifications (e.g., sulfate, phosphate, acetyl, methyl). On the other hand, the biological diversity of glycoproteins derives from their post-translational nature, whereby the glycan structure depends on the cell or tissue where the glycoprotein is produced.

Glycosylation is of utmost importance for the structure of glycoproteins and their cellular functions. Associated with a number of biological processes, including protein folding and secretion, cell targeting and adhesion, cell differentiation, signal transduction, and immune response [1–4], it is estimated that more than half of human proteins are glycosylated. Alterations of glycosylation are established to occur in diseases such as cancer (including glycan changes marking malignancy, tumor cell invasiveness, and metastasis), rheumatoid arthritis, congenital disorders of glycosylation, lysosomal storage diseases, and some immunological diseases [5, 6]. Moreover several therapeutic human glycoproteins require the presence of appropriate glycans for proper folding and to ensure stability in human serum.

Natural glycoproteins are often expressed as complex mixtures of glycoforms, in which the protein backbone is common though the structure and/or position of glycans are different. Such heterogeneity is biologically relevant, as it can vary not only according to the cell type but also the physiological state of the tissue or cells where it is produced. The heterogeneity complicates the isolation and purification of individual homogeneous forms for structural and functional studies [7, 8]. Consequently several methods have been developed for the synthesis of glycoproteins and glycopeptides with defined glycans.

This review article focuses on the use of enzymes for the synthesis of glycopeptides and glycoproteins. As detailed herein, enzymes have proven to be valuable catalysts for several aspects of glycopeptide and glycoprotein synthesis, including glycosidic linkage formation and elaboration, peptide bond formation, protecting group removal, and modification (e.g., sulfation, acylation, and cyclization). Chemical methods and the use of neoglycopeptides are also important for the study of glycopeptides and glycoproteins, though these will not be covered here as they are the topics of other articles in this volume.

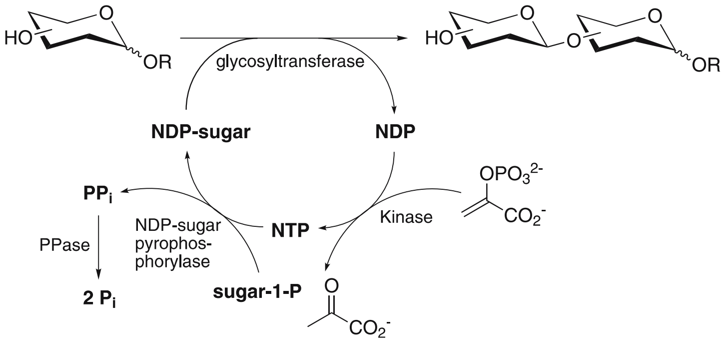
2
Enzymatic Glycosidic Linkage Formation in Glycopeptide Synthesis

Various types of sugar-modifying enzymes have been used in the synthesis of glycopeptides. Glycosyltransferases, exoglycosidases, and endoglycosidases have proven to be valuable catalysts for the formation and elaboration of glycans on glycopeptides.

2.1
Glycosyltransferases

The biosynthesis of oligosaccharides can be divided into Leloir and non-Leloir pathway enzymes. In mammalian systems, Leloir pathway enzymes account for the synthesis of most cell-surface glycoforms. These enzymes catalyze the transfer of a given carbohydrate from the corresponding sugar nucleotide donor substrate to a specific hydroxyl group of the acceptor molecule. Though a large number of glycosyltransferases of this type exist, it is amazing that these enzymes utilize only nine general sugar nucleotide substrates as glycosyl donors. Glucosyl-, galactosyl-, and xylosyltransferases consume substrates activated as uridine diphosphate sugars (α-UDP-Glc, α-UDP-GlcNAc, α-UDP-GlcUA, α-UDP-Gal, α-UDP-GalNAc, α-UDP-xylose). Fucosyl- and mannosyl-transferases employ guanosine diphosphate sugar donors (β-GDP-Fuc, α-GDP-Man), while sialyltransferases accept substrates with cytidine monophosphate as the anomeric leaving group (β-CMP-NeuAc). Non-Leloir glycosyltransferases, which consume sugar phosphates as glycosyl donors, have not been applied widely in the synthesis of glycopeptides and will not be discussed further.

Leloir glycosyltransferases are ideal for the synthesis of glycopeptides due to their high regio- and stereospecific glycosidic linkage formation and high



Scheme 1 Multi-enzyme sugar nucleotide regeneration protocols are used for glycosyltransferase reactions to minimize product inhibition by NDPs and reduce the expense of NDP-sugars by recycling NDPs to NDP-sugars

yield. Nevertheless, glycosyltransferases suffer from two shortcomings. First, the nucleoside diphosphates (NDPs) produced are often inhibitors of the glycosyltransferase. In addition, the expense of sugar nucleotide (NDP-sugar) can become a concern for large-scale synthesis. Fortunately, both of these problems can be solved by the use of multi-enzyme sugar nucleotide regeneration (Scheme 1), wherein the product inhibition is avoided and expense reduced by recycling catalytic amounts of the NDPs to NDP-sugars [9, 10].

Another obstacle is the availability of a wide diversity of glycosyltransferases. Though enzymes for all desired glycosidic linkages have not been isolated, several glycosyltransferases are available commercially and many others can be expressed recombinantly or isolated from tissue sources.

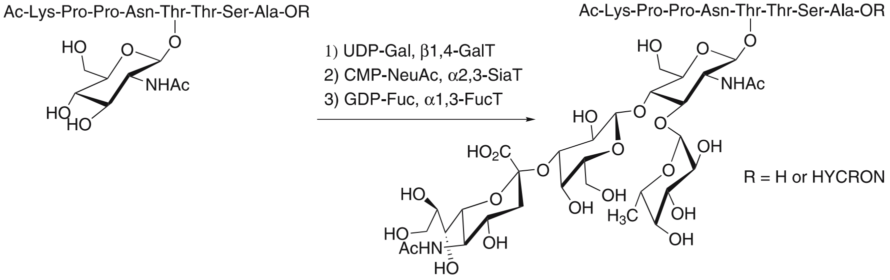
2.2

Use of Glycosyltransferases for the Synthesis of Glycopeptides with Complex *O*-glycans

Leloir glycosyltransferases are a favorable tool for the synthesis of *O*-linked glycopeptides, as the enzymatic transfer of individual monosaccharides to a preformed simple *O*-linked glycan of a synthetic glycopeptide is more convenient than difficult chemical glycosylations involving issues of protecting-group manipulation and stereoselectivity. Such elaboration of glycans was used in the synthesis of glycopeptides containing an *O*-linked sialyl-Lewis-X (SLe^X) tetrasaccharide, a glycopeptide fragment of P-selectin glycoprotein ligand-1 (PSGL-1), and tumor associated Tn (GalNAcαThr/Ser)- and T (Galα1-3GalNAcαThr/Ser)-antigens and their sialylated forms.

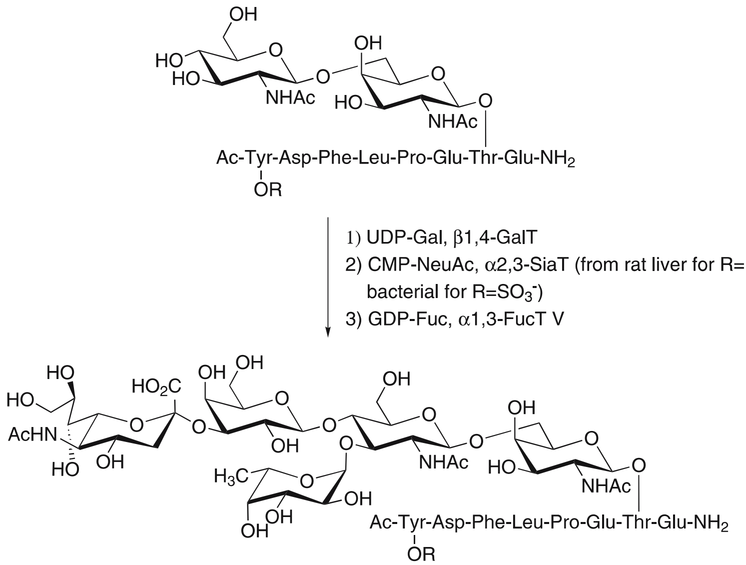
Various glycoforms of SLe^x, SLe^a, and their sulfated derivatives are moieties on adhesion molecules that are recognized by selectins. Interactions between selectins and their ligands leads to a "rolling" of the leukocyte on the endothelial cell surface as an inflammatory response to an injured site. In an effort to construct specific glycoforms of SLe^x of interest to the study of cell-cell adhesion, an *O*-linked glycopeptide containing β-linked GlcNAc was elaborated sequentially with β1,4-galactosyl-, α2,3-sialyl-, and α1,3-fucosyltransferases to yield a SLe^x tetrasaccharide-containing glycopeptide (Scheme 2) [11]. The enzymatic glycosylation of the *O*-GlcNAc-octapeptide substrates was performed both in solution and on solid support, highlighting the flexibility of glycosyltransferases in chemoenzymatic methodologies for glycopeptide synthesis. More recently, a similar approach has been reported for the chemoenzymatic synthesis of glycopeptides with a spacer between the SLe^x or SLe^x mimetic glycan and peptide [12, 13].

Another example for the synthesis of *O*-linked glycopeptides is the chemoenzymatic synthesis of glycopeptide fragments of PSGL-1. PSGL-1 is a membrane-bound counter-receptor for P-selectin that is expressed on leukocytes and involved in leukocyte initiation in the inflammatory response. Soluble glycopeptide fragments of PSGL-1 have been sought as P-selectin



Scheme 2 Sequential elaboration of a *β*-GlcNAc glycan on an *O*-linked glycopeptide by glycosyltransferases to a sialyl-Lewis-X tetrasaccharide

inhibitors, wherein glycosyltransferases were employed for extension of the glycan from a monosaccharide up to a hexasaccharide. Cummings and colleagues reported the chemoenzymatic synthesis of several PSGL-1 glycopeptide fragments by sequential enzymatic glycosylation [14–16]. These various PSGL-1 fragments provided the basis for revealing that both sulfate and sugar residues are structural requirements for high affinity binding to P-selectin. Wong and coworkers synthesized an associated fragment of PSGL-1, wherein a synthetic glycopeptide containing a disaccharide was expanded to a pentasaccharide by glycosyltransferases (Scheme 3) [17, 18].



Scheme 3 Enzymatic synthesis of a fragment of P-selectin glycoprotein ligand-1. The commercially available *α*-2,3-sialyltransferase from rat liver transferred *N*-acetylneuraminic acid only to the unsulfated glycopeptide. Recombinant *α*-2,3-sialyltransferase from *Neisseria meningitidis*, however, was used successfully with the sulfated glycopeptide

Interestingly, the commonly used α2,3-sialyltransferase from rat liver did not accept the sulfated glycopeptide [17]. However, recombinant expression of a bacterial α2,3-sialyltransferase from *Neisseria meningitidis* was efficient in transfer of sialic acid to the glycosulfopeptide [18]. Additional studies have revealed an α2,3-sialyltransferase from *N. gonorrheae* that shows application for a variety of substrates, including oligosaccharides, glycopeptides, and their sulfated derivatives and glycolipids [19].

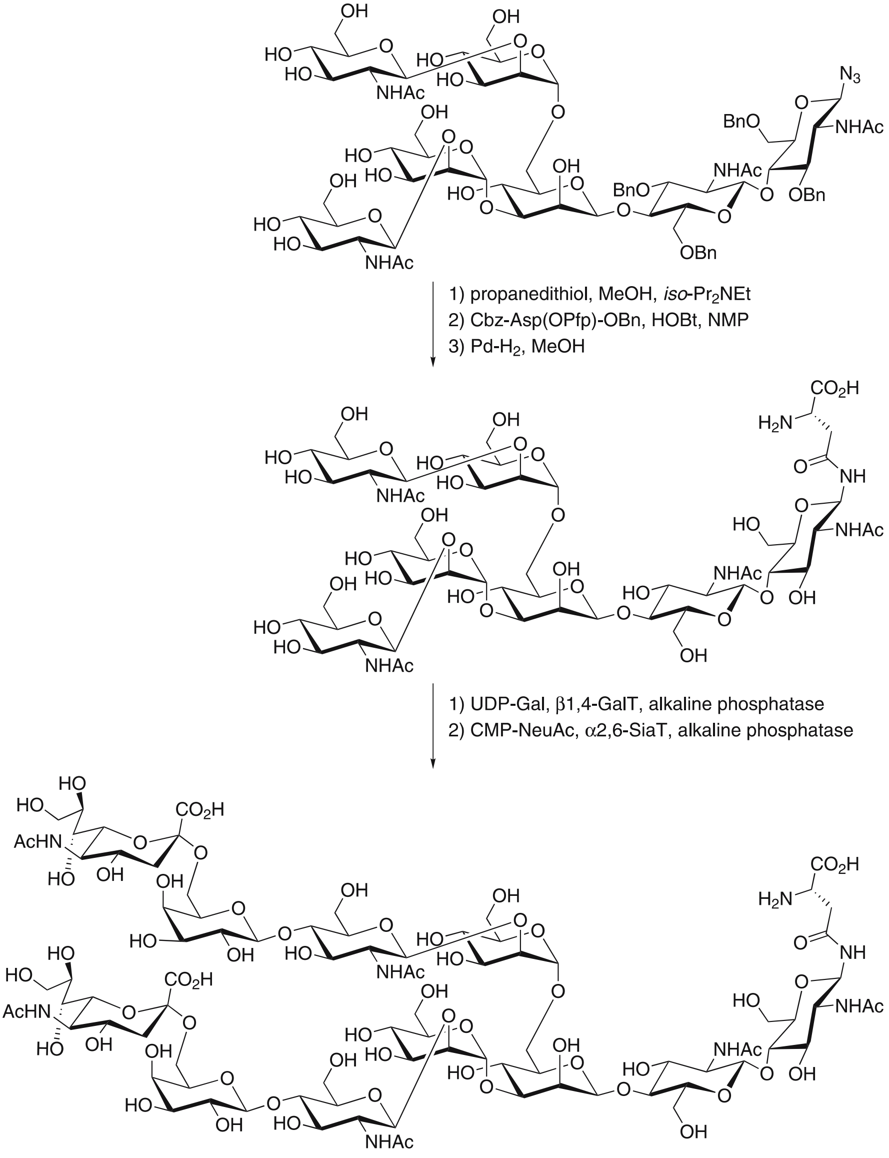
Additionally, glycosyltransferases have been used in the synthesis of glycopeptides containing mucins. The most common *O*-linked carbohydrates in eukaryotes are the mucins, which contain a core structure represented by the Tn- and T-antigens. Sialylated Tn- and T-antigens are expressed in low levels in many normal tissues but can become expressed abundantly in several types of human malignancies. In addition, mucins are involved in inflammation and cellular recognition. Due to such important biological connections, efficient methods for the preparation of Tn- and T-antigens and their sialylated forms have been realized. Chemoenzymatic syntheses of the α2-3sialylated T-antigen, including the glycosyl amino acid Neu5Acα2-3Galβ1-3GalNAcα-threonine and a neoglycopeptide, using the sialyltransferase ST3Gal I have been reported [20, 21]. Gambert and Thiem also used ST3Gal I in the synthesis of a sialylated T-antigen derivative using a one-pot multienzyme system with cofactor regeneration [22]. Kihlberg and coworkers have employed a murine ST6GalNAc I expressed recombinantly in insect cells for the synthesis of sialyl-Tn glycopeptide and a synthetic analog containing a spacer between the peptide and glycan [23, 24]. More recently, Paulson and associates have developed an approach amenable to the preparative scale synthesis of all four possible sialyl-Tn- and sialyl-T-antigens using chicken GalNAcα2,6-sialyltransferase and porcine Galβ(1-3)GalNAcα-2,3-sialyltransferase [25].

2.3

Use of Glycosyltransferases for the Synthesis of Glycopeptides with Complex *N*-glycans

The synthesis of *N*-linked glycopeptides is also enhanced by chemoenzymatic approaches incorporating the use of glycosyltransferases, with advantages analogous to those realized with *O*-linked glycopeptides.

Similar to *O*-glycopeptides, the use of sialyltransferases is favorable as chemical introduction of sialic acid and deprotection of sialylated glycans remains challenging. Earlier work highlighted an efficient method for the chemoenzymatic synthesis of *N*-linked sialyllactosamine glycopentapeptides using galactosyl- and sialyltransferases in a two-step, one-pot reaction [26]. Slight modification of this procedure was used for the enzymatic transfer of galactose and sialic acid onto glycopeptides with multiple acceptor sites, and these multivalent sialoglycopeptides were tested as inhibitors of influenza virus [27]. Additional employment of this method by Unverzagt demonstrated a chemoenzymatic synthesis of a diantennary complex-type *N*-linked glycosyl asparagine by means of sequential, one-pot glycan elaboration with *β*1,4-galactosyl- and *α*2,6-sialyltransferases [28, 29] (Scheme 4). In this case



Scheme 4 Chemoenzymatic synthesis of an *N*-linked complex-type diantennary glycosyl asparagine. The heptasaccharide core was elaborated to a complex-type undecasaccharide by sequential one-pot enzymatic glycosylation

the choice of protecting groups and the order of their removal was critical to obtaining the unprotected glycosyl asparagine for enzymatic glycosylation. This chemoenzymatic approach provides a complex glycosyl amino acid which can then be incorporated into glycopeptides by chemical synthetic strategies.

Also of importance are previous studies that highlight the wide-ranging applicability of glycosyltransferases for the synthesis of *N*-linked glycopeptides in other conditions. For example, glycosyltransferases were employed to make *N*-linked glycopeptides on solid support [30] and using multienzyme cofactor regeneration schemes in solution [31].

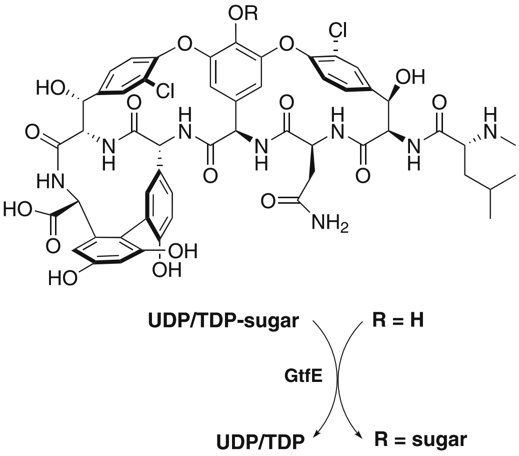
Another enzyme that has been used for the synthesis of *N*-linked glycopeptides is oligosaccharyl transferase, which uses a dolichol-linked tetradecasaccharide (Dol-PP-$\mathrm{GlcNAc_2Man_9Glc_3}$) donor to transfer the saccharide to an asparagine side chain of a protein inside the lumen of the endoplasmic reticulum. The protein or peptide acceptor contains the sequence Asn-Xaa-Thr/Ser, where Xaa is any amino acid except proline. Using the yeast oligosaccharyl transferase, the Imperiali laboratory studied the substrate specificity of the enzyme for glycosylation of a tetrapeptide [32]. The minimum requirement for the glycosyl donor is a GlcNAc unit (Dol-PP-GlcNAc), though transfer of this monosaccharide is not very efficient. Addition of another sugar improves the efficiency, with Dol-PP-$\mathrm{GlcNAc_2}$ a better substrate than Dol-PP-GlcNAcGlc. Such experiments have been valuable for studying the specificity and providing insight into the mechanism of oligosaccharyl transferase, yet further work is necessary to improve the yield for this enzyme to have synthetic utility.

2.4

Use of Antibiotic Glycosyltransferases for the Sythesis of Glycopeptide Antibiotics

Antibiotic glycosyltransferases, like traditional Leloir glycosyltransferases, utilize sugar nucleotide donor substrates, while the nucleophilic acceptor atom is present on the co-substrate aglycon or partially glycosylated antibiotic. Usually the substrates are TDP-hexoses, though some UDP-hexoses show activity in vitro, activated at C-1, including aminodeoxy-, deoxy-, and deoxymethyl-modified hexoses.

Glycopeptide antibiotics are biosynthesized by nonribosomal peptide synthetases and modified by oxidative crosslinking enzymes and antibiotic glycosyltransferases. Such late-stage glycosylation of glycopeptide antibiotics is significant for current methods toward engineering analogs with altered glycans via chemoenzymatic synthesis. Since the sugar moieties on glycopeptide antibiotics are known to be important for their biological activity (e.g., vancomycin aglycon does not have in vivo antimicrobial properties), the use of glycosyltransferases from the biosynthetic machinery of glycopeptide anti-



Scheme 5 Vancomycin derivatives with altered glycans made by using biosynthetic antibiotic glycosyltransferases

biotics provides an ideal avenue for introduction of structural diversity [33, 34]. Such modification of glycopeptide structures entails the coupling of chemical or enzymatic synthesis of TDP- or UDP-hexose substrates to subsequent enzymatic glycosylation of the aglycon by the appropriate antibiotic glycosyltransferase.

Glycopeptide analogs of vancomycin and teicoplanin have been produced via modification of the glycans through the action of antibiotic glycosyltransferases [35–37]. Vancomycin contains an L-vancosaminyl-1,2-D-glucosyl disaccharide attached to the 4-hydroxyphenylglycine residue and is formed by the action of two antibiotic glycosyltransferases, the D-glucosyltranferase GtfE and the L-vancosaminyltransferase GtfD. GftE is known to accept both vancomycin and teicoplanin aglycons as substrates [35, 36, 38], along with a variety of UDP- and TDP-sugar donors [36, 37]. Several vancomycin analogs have been synthesized using GtfE and UDP- or TDP-aminoglucoses, providing a chemoselective site for modification of the amine for further diversification (Scheme 5). An additional vancomycin analog incorporating a 6-azido-6-deoxyglucosyl moiety was made enzymatically with GtfE and varied chemically by 1,3-dipolar cycloaddition reaction with several acetylenes [37].

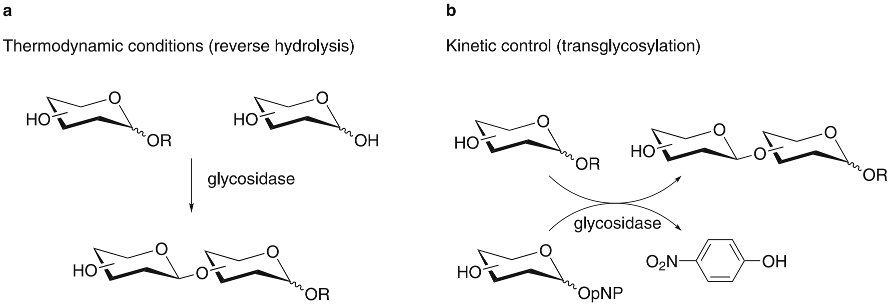
2.5
Exo-Glycosidases

The role of exo-glycosidases in vivo is glycan processing during glycoprotein synthesis via cleavage of glycosidic linkages. However, under appropriate controlled conditions, glycosidases can be used for glycosidic bond formation instead of cleavage (Scheme 6). Use of these enzymes for the synthesis of oligosaccharides has been employed extensively [39–42]. Compared to glycosyltransferases, glycosidases are privileged as accepting inexpensive glycosyl donors, more stable, and readily available. On the other hand, they suffer from weak regiospecificity and oftentimes low yield from competitive hydrolysis.

The use of glycosidases for glycosidic bond formation can be done under thermodynamic conditions (reverse hydrolysis), though this procedure oftentimes results in low yield of product. An alternative approach, under kinetic conditions (transglycosylation), results frequently in improved yields. The transglycosylation protocol usually utilizes an activated glycosyl donor, organic co-solvent, and/or excess donor. Even under such conditions, the yield can still be lower than those achieved by glycosyltransferase reactions. However, the availability of a glycosidase when the equivalent glycosyltransferase is not accessible makes it a valuable alternative route.

Glycosidases have also been applied as catalysts for glycopeptide synthesis. For example, recombinant β-galactosidase from *Bacillus circulans* was employed in the chemoenzymatic synthesis of a tetrapeptide containing a sialyl T-antigen [43], following an earlier study using this *B. circulans* β-galactosidase for the synthesis of Galβ(1–3)GalNAcα-serine [44]. More recently an enzymatic synthesis of a sialyl-Lewis-X *O*-glycan on a tumor-associated MUC1a$'$ peptide was reported [45], wherein β-galactosidase and a series of glycosyltransferases were used sequentially to assemble the desired glycan.

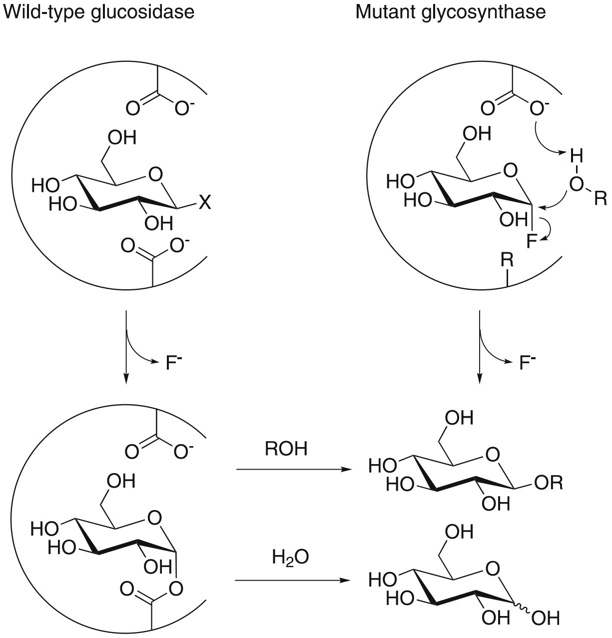
Exoglycosidases are also useful for glycan remodeling, wherein part of the glycan is removed by glycosidases followed by re-elaboration of the glycan either enzymatically or chemically. For instance, jack bean β-galactosidase was used in glycan remodeling to remove the terminal galactosyl residues on an isolated glycopeptide [46]. This modified glycopeptide was then elongated by glycosyltransferases to the glycan found on glycodelin A, a glycoprotein having immunosuppressive properties and an inhibitor of sperm–egg binding.



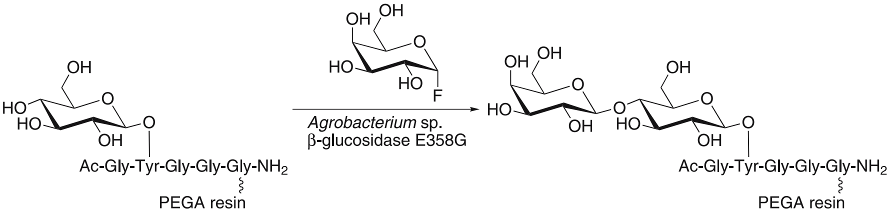
Scheme 6 Thermodynamically- (**a**) and kinetically-controlled (**b**) procedures for using glycosidases for glycosidic linkage formation

An alternative use of glycosidases involves the use of a specifically mutated glycosidase, known as a glycosynthase. A glycosynthase contains a mutation in the active site such that the active-site catalytic nucleophile is replaced with a non-nucleophilic amino acid side chain, thereby eliminating the hydrolysis activity (Scheme 7). However, the glycosynthase preserves its ability to glycosylate a sugar acceptor with an appropriate activated glycosyl donor. Typically glycosyl fluorides act as activated glycosyl donors for glycosynthases since they are synthesized readily and bear a small leaving group. Glycosynthases have been used widely in oligosaccharide synthesis [47, 48], and some have been applied to glycopeptides. For instance, Withers and coworkers used a glycosynthase mutant of *Agrobacterium* sp. β-glucosidase to transfer a galactose moiety from a galactosyl fluoride to a resin-bound glycopentapeptide (Scheme 8) [49]. Key considerations for this study included the use of PEGA resin for its compatibility with aqueous buffers and the choice of a backbone amide linker for attachment of the glycopeptide to the resin.

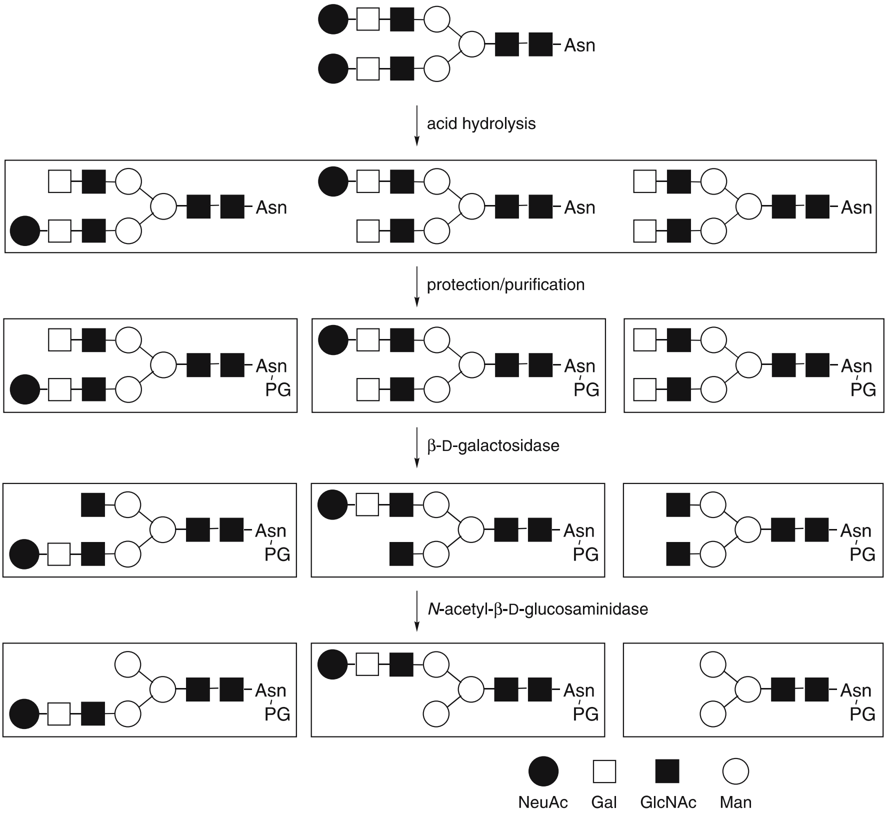
Lately exo-glycosidases have been used to make diverse asparagine-containing oligosaccharides (Asn-oligosaccharides) and a glycopeptide containing two of the oligosaccharides (Scheme 9) [50]. An asparagine-linked biantennary complex-type sialylundecasaccharide was isolated from egg yolk and processed by acid hydrolysis to yield a mixture of Asn-oligosaccharides.



Scheme 7 Glycosynthases, mutant glycosidases lacking an active-site catalytic nucleophile, have an altered mechanism of glycosyl bond formation without hydrolysis



Scheme 8 Glycosynthase of *Agrobacterium* sp. β-glucosidase for the galactosylation of a resin-bound glycopeptide



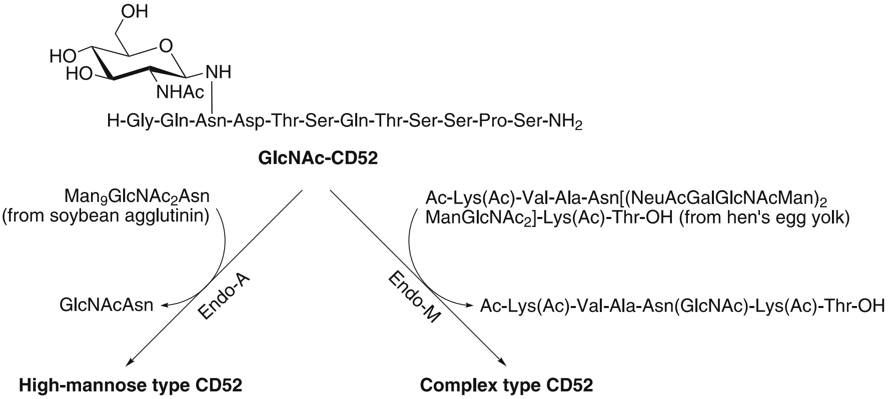
Scheme 9 Several diverse asparagine-containing oligosaccharides were made by processing an asparagine-linked biantennary complex-type sialylundecasaccharide from egg yolk. The isolated complex-type Asn-oligosaccharide was cleaved by acid hydrolysis. The resulting mixture was protected on the amine of the asparagine, and the components purified by HPLC separation. Subsequent processing by exoglycosidases yielded various Asn-oligosaccharides

The mixture was protected with Fmoc on the amine of the asparagine to permit purification of the Asn-oligosaccharides by HPLC. These Asn-oligosaccharides were then processed further with *β*-galactosidase and *N*-acetyl-*β*-glucosaminidase to produce a diverse collection of *N*-linked glycans. Two of these Asn-oligosaccharides, a sialyl-undeca- and a nonasaccharide, were used subsequently in solid-phase glycopeptide synthesis.

2.6
Endo-Glycosidases

Internal glycosidic linkages in an oligosaccharide or glycoconjugate are cleaved by endo-glycosidases. Under kinetic conditions, the transglycosylase activity of endo-glycosidases is favored over cleavage. The endo-*β*-*N*-acetylglucosaminidases (including Endo-A, Endo-CE, Endo-D, Endo-F, Endo-H, and Endo-M) cleave the *N*,*N*′-diacetylchitobiosyl glycosidic linkage at the reducing terminus of the *N*-linked glycan, leaving one *N*-acetylglucosamine unit on the peptide or protein. In general, *N*-linked glycans are classified as high-mannose, hybrid, or complex, with poly(LacNAc) oligosaccharides as a possible additional category. The substrate specificity of endo-glycosidases varies from activity for a strict type of *N*-glycan to broad inclusion of a variety of *N*-glycan types. For transglycosylase activity, these endo-glycosidases require an asparagine-linked glycan donor. The transglycosylase activity of endo-glycosidases has proven to be very useful as a single transformation that yields complex *N*-linked glycopeptides and glycoproteins with defined structure.

The endo-*β*-*N*-acetylglucosaminidase from *Arthrobacter protophormiae* (Endo-A) releases high-mannose *N*-glycans from glycopeptides or glycoproteins and also shows significant transglycosylase activity. However, the transglycosylase activity forms principally in the initial stage of the reaction, allowing the product to be degraded by competing cleavage activity. To yield the most product required quick termination of the reaction, which was achieved by an immobilized Endo-A fusion protein linked to glutathione *S*-transferase (GST) [51]. This immobilized GST-Endo-A approach has been used in the chemoenzymatic synthesis of glycoforms of HIV-1 gp120 glycopeptides containing penta-, hexa-, and nonamannosyl moieties [52]. Endo-A was also used for the single-step attachment of high-mannose type oligosaccharide to CD52 (Scheme 10) [53]. CD52 is a GPI-anchored cell surface glycoprotein expressed on most human lymphocytes and sperm cells. The protein core is the same in both types of cells, while the *N*-glycans are unlike in composition. Such differences likely contribute to the differences in function as lymphocyte CD52 is involved in human immune response, whereas sperm CD52 is implicated in sperm-egg interaction. Thus generating a homogenous glycoprotein of CD52 is valuable for structural and functional studies of CD52 antigens.



Scheme 10 Endo-A and Endo-M catalyze the transfer of high-mannose and complex-type oligosaccharides, respectively, to CD52 glycoprotein

In addition, a similar approach was used for the synthesis of a 34 amino acid glycopeptide fragment of HIV-1 gp41 with high-mannose type glycan [54]. Structural and functional studies of these gp41 fragments revealed that glycosylation affects both the ability to form α-helical bundles with an N-terminal peptide fragment of gp41 and the anti-HIV activity. More recently Wang and coworkers used oligosaccharide oxazolines as donors for constructing *N*-glycopeptide fragments of HIV-1 gp41 and gp120 with Endo-A [55]. Di- and tetrasaccharide oxazolines were transferred efficiently (yields of 73–82%) to GlcNAc-peptide acceptors with stereo- and regiospecificity.

Endo-M, an endo-β-*N*-acetylglucosaminidase from *Mucor hiemalis*, exhibits transglycosylase activity for a wide range of substrates, including high-mannose, hybrid-type, and complex-type oligosaccharides. Such broad substrate specificity has been exploited for the syntheses of several glycopeptides [56]. Endo-M was used to transfer oligosaccharides to yield glycopeptides of human chorionic gonadotropin peptide [57], Peptide T [58], calcitonen [59, 60], the neuropeptide Substance P [61], and *Saccharomyces cerevisiae* α-mating factor [62], allowing structural and functional studies of the homogenous glycopeptides. The chemoenzymatic synthesis of a complex-type CD52 glycoprotein was also performed using Endo-M (Scheme 10) [53].

Though the use of endoglycosidases is currently a favorable approach for the synthesis of homogeneous *N*-linked glycopeptides, this method often has the drawback of relatively low yields. This problem has been addressed recently to some extent for Endo-M, wherein the use of media containing organic solvents results in transglycosylase reactivity of Endo-M with improved yield [63].

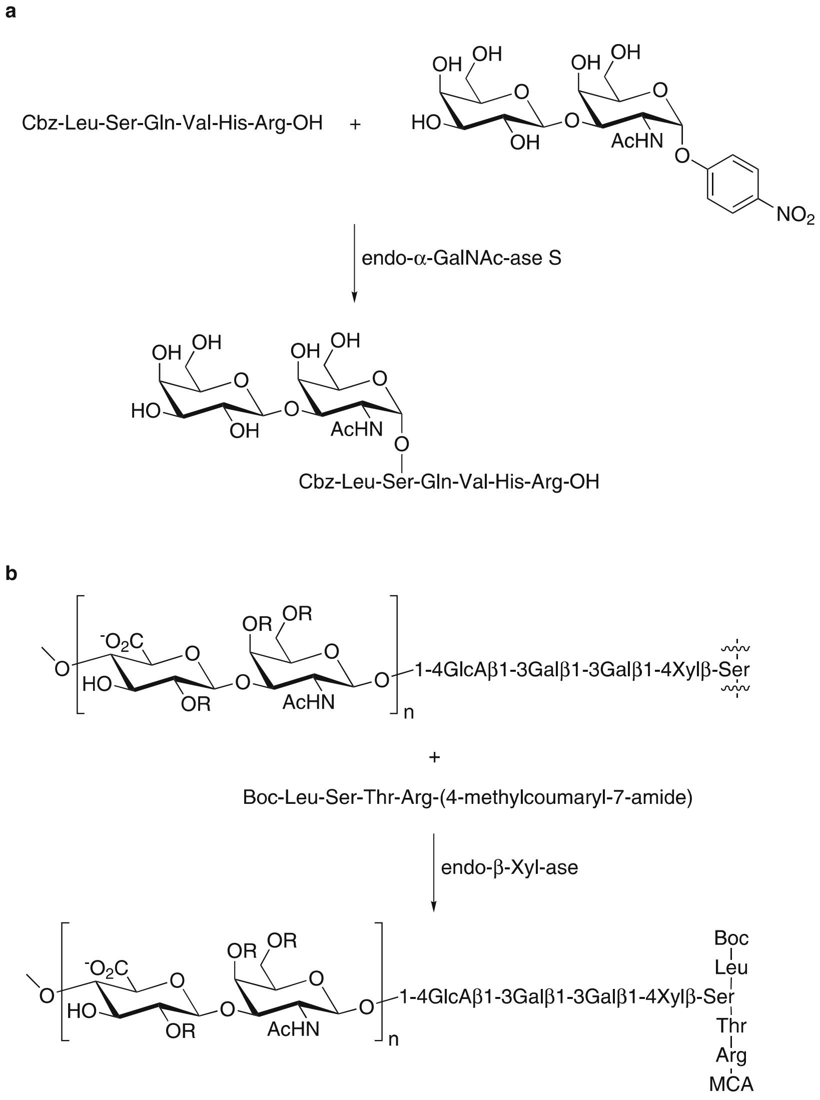
Endo-CE is an endo-β-*N*-acetylglucosaminidase from the nematode *Caenorhabditis elegans*. This enzyme has been expressed recombinantly and shows activity for high-mannose type glycans, though biantennary complex type oligosaccharides are poor substrates and triantennary complex glycans are not hydrolyzed [64]. In addition, Endo-CE exhibits transglycosylase activity, though it has not yet been applied to glycopeptide synthesis.

Endo-D and Endo-F are endo-β-*N*-acetylglucosaminidases from *Steptococcus pneumoniae* and *Flavobacterium meningosepticum*, respectively. Endo-D acts on the core structure of complex type *N*-linked glycopeptides and glycoproteins [65, 66], while Endo-F hydrolyzes core high-mannose and biantennary complex *N*-linked oligosaccharides [67]. The transglycosylase activity, and therefore synthetic utility, of Endo-D and Endo-F have not been reported.

While endo-glycosidases have been utilized extensively to synthesize *N*-linked glycopeptides, of late they have been applied to *O*-linked glycopeptides. Endo-α-*N*-acetyl-D-galactosaminidase (endo-GalNAc-ase) hydrolyzes the *O*-glycosidic linkage between Galβ1-3GalNAcα and serine or threonine in mucin-type glycopeptides and glycoproteins. Endo-GalNAc-ases from *S. pneumoniae* [68] and *Bacillus* sp. [69] display transglycosylase activity toward sugar acceptors. However, only the Endo-GalNAc-ase isolated from the culture medium of *Streptomyces* sp. OH-11242 [70] has been employed for the chemoenzymatic synthesis of a glycopeptide, a mucin-type hexapeptide (Scheme 11a) [71]. Endo-β-xylosidase from the molluscus *Patinopecten* cleaves the Xyl-β1-*O*-Ser linkage between a glycosaminoglycan chain and the core protein. Isolated peptidochondroitin sulfate (molecular mass 30 kD), dermatan sulfate (32 kD), and heparan sulfate (19 kD) were transferred by the endo-β-xylosidase to a fluorescently labeled tetrapeptide (Scheme 11b) [72], thus illustrating the potential of this endo-β-xylosidase for the synthesis of proteoglycans.

Endo-glycosidases are also useful for glycan remodeling of glycoproteins. In glycan remodeling, a glycoprotein is isolated with a heterogenous glycan, the glycan is trimmed back by glycosidases, and a homogeneous glycan is built, usually enzymatically by glycosyltranferases. Such an approach is limited to primary glycosylation sites predetermined by the cell type used for the protein expression. However, this approach is appropriate in many cases, as exemplified by the use of this method for the preparation of homogenous glycoforms of ribonuclease (RNase) [73] (Scheme 12). The high-mannose type glycan on commercially available RNase B was cleaved by the endo-β-*N*-acetylglucosaminidase from *Streptomyces plicatus* (Endo-H) to yield GlcNAc-RNase. Elaboration of the Endo-H-treated GlcNAc-RNase with glycosyltransferases produced a homogeneous glycoform of RNase containing the branched sLe^x glycan.

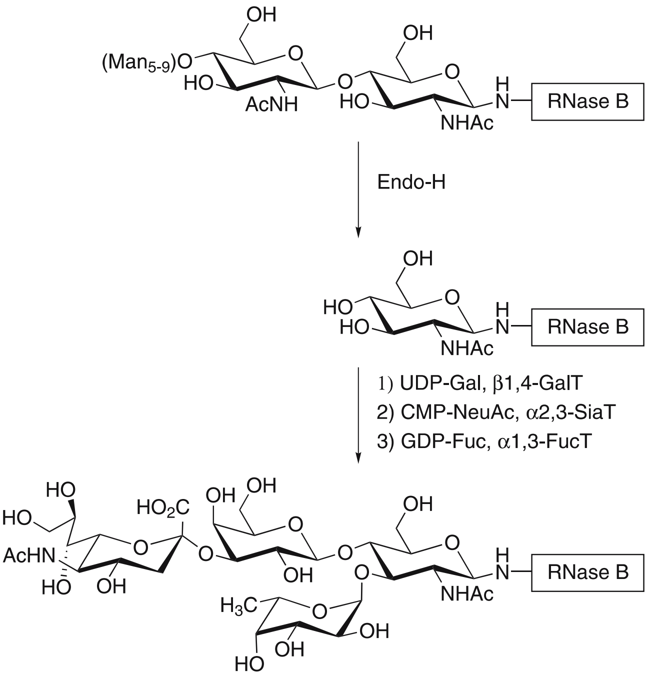
Recently enzymatic digestion of chitin with chitinase followed by chemical amination was used for the preparation of 1-β-aminochitobiose [74]. This β-glycosylamine was coupled to an Fmoc-protected aspartic acid allyl ester then protected appropriately for Fmoc solid phase glycopeptide synthe-



Scheme 11 Enzymatic *O*-linked glycosylation of peptides by endoglycosidases. **a** A mucin-type glycopeptide was synthesized using endo-α-GalNAc-ase from *Streptomyces* sp. **b** Endo-β-xylosidase transferred glycosaminoglycans to a fluorescently labeled peptide

sis. The *N*-linked chitobiose building block was then incorporated into a 39 amino acid glycopeptide, thereby demonstrating the synthetic utility of this approach along with its potential to make large peptide fragments for the semi-synthesis of glycoproteins.

Glycosynthases have also been made from mutated endo-glucanases and cellulase. Though these endo-glycosynthases have been used as a method



Scheme 12 RNase B containing the branched sLe^x glycan was synthesized by glycan remodeling, using Endo-H to remove the high-mannose type glycan followed by elongation with a series of glycosyltransferases

for oligosaccharide synthesis [75–79], their synthetic utility for glycopeptide synthesis has not been reported.

3
Non-glycosidic Bond Forming Enzymes for Glycopeptide Synthesis

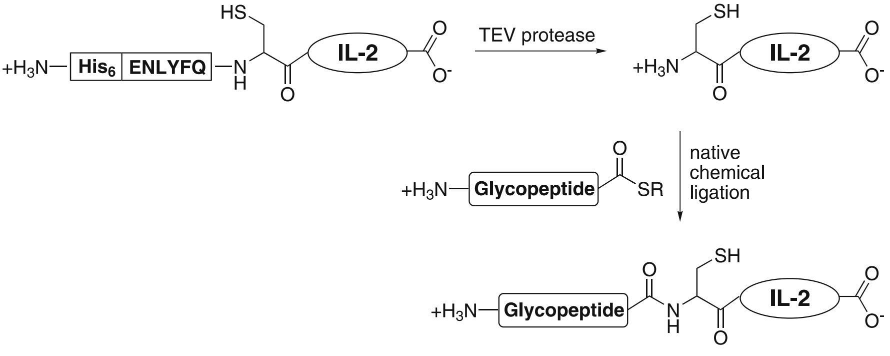
Enzymes involved in catalysis other than glycosidic bond formation have also been important for enzymatic glycopeptide and glycoprotein synthesis. Such enzymes include proteases for peptide bond formation and protecting group removal, lipases for removal of protecting groups, and sulfotransferases, acyltransferases, and thioesterases for glycopeptide modification.

3.1
Proteases

Though the function of proteases is the hydrolysis of peptide bonds in vivo, they can be utilized for amide or ester bond formation under appropriate conditions in vitro. Thus, proteases have great synthetic utility [80], including several functions in the synthesis of glycopeptides. First, they have been used for the formation of peptide bonds. Subtilisin and engineered variants thereof have been utilized widely to ligate synthetic peptide fragments to form larger peptides [81–83]. In order to achieve favorable yields, kinetically controlled conditions using an active site Ser to Cys subtilisin mutant to favor aminolysis to hydrolysis and acyl donors activated as esters (especially benzyl types of esters) are preferred. As expected, such kinetic conditions involve competition between hydrolysis and aminolysis. Nevertheless, hydrolysis can be minimized by reducing the amount of water and increasing the cosolvent concentration, the rate of aminolysis can be enhanced by increasing the concentration of acyl acceptor peptide, and subtilisin mutants have been optimized for peptide ligation or stability in organic solvents [82, 84–86]. In addition, self-condensation reactions are avoided by protecting the N-terminal amine of the acyl donor and the C-terminus of the acyl acceptor. Engineered variants of subtilisin have been made with improved thermostability and broadened substrate specificity, thereby providing access to peptide bond formation with several glycopeptide substrates, including *O*- and *N*-linked glycopeptides containing unprotected or acylated glycans [31, 73, 87]. Based on the substrate specificity of subtilisin BPN′ [87, 88], glycopeptides with sugar residues typically at least one residue away from the site of amide bond formation in the active site can be accommodated. The advantages of the use of subtilisin variants over chemical condensation techniques are the lack of racemization and opportunity to form a cysteine-free connection, as is necessary for native chemical ligation [89].

Proteases have also been used to generate proteins with N-terminal cysteines for native chemical ligation, which has been used extensively for the synthesis of lengthy peptides and proteins [90]. The protease factor Xa was used to produce N-terminal cysteines [91, 92], though the imperfect sequence specificity of factor Xa can result in cleavage after positively charged residues [93, 94]. In a recent effort to find a more specific cysteine protease, the tobacco etch virus NIa (TEV) protease has been demonstrated as an efficient and selective means for generating proteins with N-terminal cysteines from recombinantly expressed N-terminal fusion proteins [95]. This TEV protease approach was applied to produce a peptide derived from HIV gp41 and a truncated form of human interleukin-2 protein with N-terminal cysteines [96, 97]. Native chemical ligation of these peptides to a glycopeptide thioester resulted in a glycopeptide related to HIV entry and full-length glycosylated interleukin-2 (Scheme 13).

In addition, proteases have been used for selective ester hydrolysis of protecting groups on glycosyl amino acids [98], which can then be incorporated into glycopeptides. For example, papain was used for the hydrolysis of some methyl [99–101], 2-(2-methoxyethoxy)ethyl (MEE) [102], and *iso*-butyl [103] esters from protected amino acid glycosides. In other studies thermitase was used to remove a *tert*-butyl ester from a glycosylated dipeptide [104, 105] and cleaved methyl and *p*-nitrobenzyl esters from peptide substrates [106]. Subtil-



Scheme 13 Synthesis of IL-2 glycoprotein by native chemical ligation of a glycopeptide thioester to a truncated IL-2 with N-terminal cysteine. The truncated IL-2 was expressed recombinantly as a fusion protein and processed to the N-terminal cysteine by TEV protease

isin was also used to deprotect the C-terminus of glycodipeptides with methyl esters [107], albeit papain could not be used in this case due to its cleavage of the dipeptide. This illustrates the drawback of using proteases for the removal of protecting groups from peptides and glycopeptides. Additionally, enzymatic cleavage of glycopeptides from the resin after solid-phase synthesis has also been demonstrated with α-chymotrypsin [30].

3.2
Lipases

Lipases have been particularly valuable for the selective formation or hydrolysis of esters on a variety of chemical structures [98, 108–110], including the hydrolysis of protected glycopeptides. The use of lipases is free of protease activity and avoids chemical deprotection problems, as they react under essentially neutral pH conditions which are compatible with acid- or base-sensitive glycosidic linkages. Another key feature of lipases is their lack of specificity for L-amino acids; thus, they can be used with peptides containing D-amino acids [111]. In addition, a variety of lipases are available commercially and rather inexpensive.

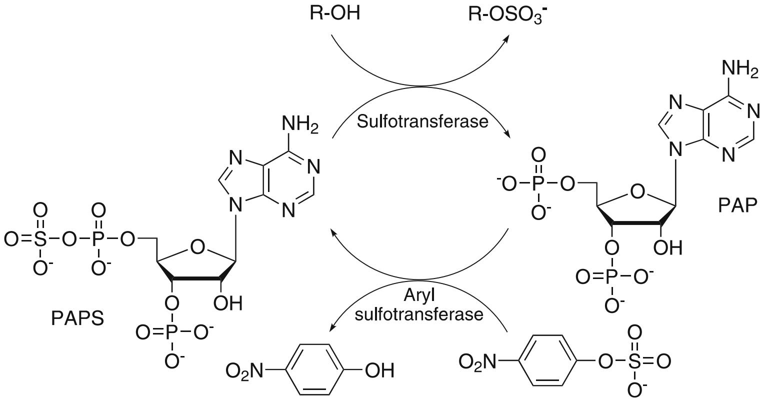
A number of lipase-labile ester protecting groups have been applied successfully for the deprotection of the C-terminus during glycopeptide synthesis [112, 113], including 2-bromoethyl [114], *n*-heptyl [114–117], MEE [102, 118], and (polyethylene)glycol [118]. Protection of the C-terminus as a choline ester has also been cleaved with butyrylcholine esterase for the synthesis of glycophosphopeptides [119, 120]. In addition, *p*-(phenylacetoxy)benzyloxycarbonyl has been used to protect the N-terminal amine, where removal of the phenylacetate is achieved by penicillin G acylase and the remaining phenolate undergoes spontaneous fragmentation to yield the free amine [119, 121].

3.3
Sulfotransferases

Sulfation is an important post-translational modification [122, 123] that can occur on glycoproteins or glycopeptides on either the hydroxyl group of a glycan or the side chain of an amino acid. Sulfotransferases, which install sulfate esters, have proven useful in the synthesis of glycoconjugates, with some application to the synthesis of glycosulfopeptides. For instance, a glycopeptide fragment of PSGL-1 was sulfated enzymatically at three tyrosine residues with a cloned tyrosylprotein sulfotransferase [14].

More recently, a simple one-pot enzymatic regeneration of 3′-phosphoadenosine-5′-phosphosulfate (PAPS) has been developed (Scheme 14) [124–126]. Similar to in situ regeneration for other transferase enzymes, this strategy is valuable for the synthetic use of sulfotransferases, since PAPS is an expensive sulfate donor, and PAP can cause feedback inhibition. This approach has been applied to the synthesis of sulfated carbohydrates and should also be valuable for glycosulfopeptide synthesis.

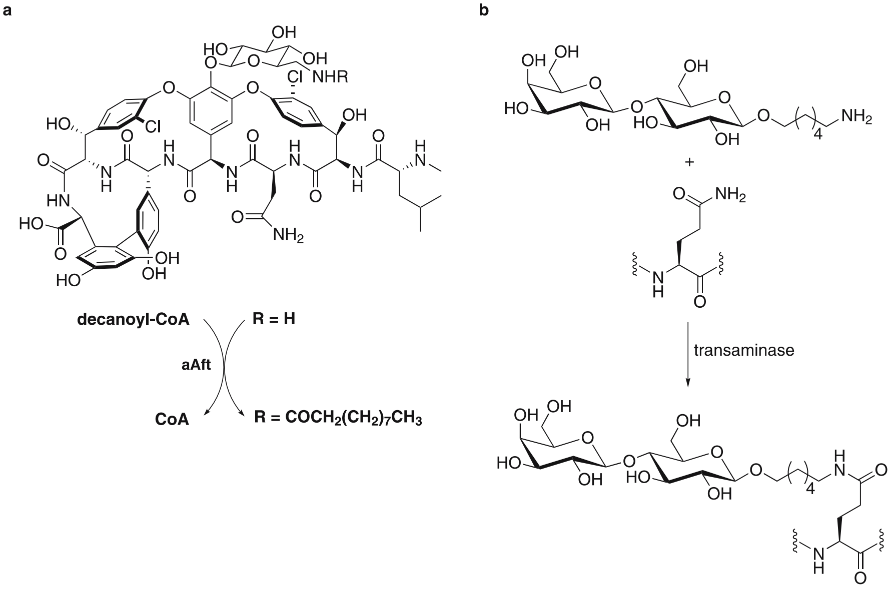
Other sulfotransferases that catalyze the sulfation of carbohydrates or peptides have been discovered [122, 123], though they have not been applied extensively to glycosulfopeptide synthesis.



Scheme 14 Enzymatic regeneration of PAPS from PAP for the sulfation of biologically-relevant molecules, including glycopeptides and oligosaccharides

3.4
Acyltransferases

Recently, acyltransferases have been used for the modification of lipoglycopeptide antibiotics [127]. Acyltransferases were cloned from the biosynthetic genes encoding teicoplanin and A-40,926 and used for the acylation of vancomycin (Scheme 15a), teicoplanin, and A-40,926 with various acyl-



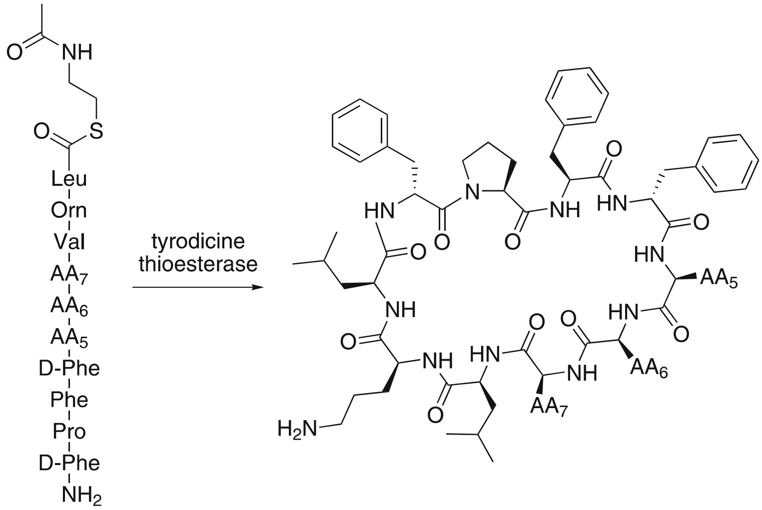
Scheme 15 Acyltransferases have been used (**a**) to modify glycopeptide antibiotics and (**b**) to generate neoglycopeptides

CoAs. Though these acyltransferases have been used in analytical reactions to demonstrate their ability to make novel lipoglycopeptides, it remains to be determined if they will have efficient synthetic utility.

Another area where acyltransferases have been useful is the synthesis of modified glycopeptides with non-native sugar-peptide linkages, often termed neoglycopeptides [128, 129]. Glutaminyl-peptide γ-glutamyl transferase (transglutaminase) catalyzes the acylation of primary amines with the γ-carboxamide group of glutamine or glutamine-containing peptides. Though the substrate specificity of transglutaminase allows only the preparation of neoglycopeptides (Scheme 15b), its use provided the framework for studying the effect of carbohydrates on cyclic peptides targeting hemagglutinin [130] and prolonging the glucose-lowering effect of insulin with incorporation of sialyloligosaccharides [131, 132].

3.5
Thioesterases

A thioesterase has been employed of late in the macrolactamization of cyclic head-to-tail glycopeptides [133]. The isolated thioesterase domain of the tyrocidine nonribosomal peptide synthetase was shown to be efficient in the macrolactamization of several glycopeptide analogs of tyrocidine (Scheme 16). Though the thioesterase is known to be intolerant of variations



Scheme 16 Monoglycosylated peptides made efficiently by cyclization with tyrocidine thioesterase. A variety of glycosyl amino acids were substituted at positions 5 through 7 and cyclized enzymatically, including AA_5 = Ser(β-Gal), Ser(Ac_4-β-gal); AA_6 = Ser(β-Gal), Ser(Ac_4-β-gal); AA_7 = Ser(β-Gal), Ser(Ac_4-β-gal), Ser(β-Glc), Ser(β-GlcNAc), Ser(β-$GlcNH_2$), Tyr(β-Gal), Tyr(β-Glc)

from the tyrocidine sequence near the site of macrolactamization, a diversity of head-to-tail cyclic glycopeptides with internal sequence variation can be achieved with distinct regio- and stereochemistry. The isolated tyrocidine thioesterase domain was also used in related work for the synthesis of cyclic neoglycopeptides [134].

4
In Vivo Synthesis of Glycoproteins

In vivo methods for the synthesis of glycoproteins have been established. Such methods, consequently, are the equivalent of in vivo enzymatic methods, whereby enzymes within the cell are used for synthetic purposes. Two types of in vivo glycoprotein synthesis have been explored, including co-translational incorporation of an unnatural amino acid by in vivo suppressor tRNA technology and post-translational modification of glycoproteins via pathway re-engineering in yeast.

In vivo suppressor tRNA technology allows the site-specific incorporation of non-natural amino acids into proteins [135]. This process involves the evolution of an orthogonal tRNA synthetase and tRNA pair that selectively charges and inserts a selected unnatural amino acid into the protein in the place of an amber codon TAG. Using this method, engineered *E. coli* were used to produce neoglycoproteins and glycoproteins. Neoglycoproteins were generated by in vivo insertion of *p*-acetylphenylalanine into recombinant protein followed by chemical modification with aminooxy saccharides [136]. In addition, homogeneous *O*-linked glycoproteins were produced by direct incorporation of the glycosyl amino acids *N*-acetylglucosamine-β-serine [137] and *N*-acetylgalactosamine-α-threonine [138]. These glycoproteins containing core monosaccharides glycans were then elaborated by glycosyltransferases in vitro. Though this method is distinguished for its ability to introduce glycans site-specifically into proteins, further development of this method is needed to achieve favorable production levels.

Another method for in vivo synthesis of glycopeptides is pathway re-engineering in yeast. *N*-linked glycosylation occurs in yeast, though conservation with human glycosylation only exists in the early steps of *N*-glycan assembly and processing. Differences in later steps result in high-mannose type glycans in yeast, which are not suitable for therapeutic human glycoproteins. Thus, efforts have been made to humanize *N*-glycosylation pathways in yeast for the production of therapeutic glycoproteins [139]. Such re-engineering requires altering endogenous *N*-glycosylation reactions and genetic incorporation of human *N*-glycosylation pathways into yeast and other fungi. Although the ability to humanize glycosylation pathways in yeast has progressed recently, future research is necessary for efficient yeast strains to be developed for homogeneous human glycoprotein synthesis.

5
Conclusion

The daunting complexity of glycans continues to challenge their synthetic progress and, hence, study of their biological functions. Though advances in enzymatic strategies for the synthesis of glycopeptides and glycoproteins are important complements to existing chemical approaches, continued research in these areas is required to provide access to all desired constructs. A variety of enzymes are instrumental in the synthesis of glycopeptides and glycoproteins: glycosyltransferases have facilitated the synthesis of difficult glycosyl linkages, glycosidases have been used to transfer complex glycans, proteases have catalyzed peptide bond formations, lipases and proteases have removed protecting groups, and sulfo- and acyltransferases and thioesterases have been employed to modify glycopeptides.

Acknowledgements We are grateful to the National Institutes for Health for research support. D.A.T. is supported by the Norton B. Gilula Graduate Student fellowship.

References

1. Dwek RA (1996) Chem Rev 96:683
2. Sears P, Wong CH (1998) Cell Mol Life Sci 54:223
3. Lis H, Sharon N (1993) Eur J Biochem 218:1
4. Varki A (1993) Glycobiol 3:97
5. Durand G, Seta N (2000) Clin Chem 46:795
6. Dube DH, Bertozzi CR (2005) Nat Rev Drug Discov 4:477
7. Rudd PM, Dwek RA (1997) Crit Rev Biochem Mol Biol 32:1
8. Prescher JA, Bertozzi CR (2005) Nat Chem Biol 1:13
9. Wong CH, Haynie SL, Whitesides GM (1982) J Org Chem 47:5416
10. Herrmann GF, Ichikawa Y, Wandrey C, Gaeta FCA, Paulson JC, Wong CH (1993) Tetrahedron Lett 34:3091
11. Seitz O, Wong CH (1997) J Am Chem Soc 119:8766
12. Sallas F, Nishimura SI (2000) J Chem Soc Perkin Trans 1:2091
13. Matsuda M, Nishimura SI, Nakajima F, Nishimura T (2001) J Med Chem 44:715
14. Leppänen A, Mehta P, Ouyang YB, Ju TZ, Helin J, Moore KL, van Die I, Canfield WM, McEver RP, Cummings RD (1999) J Biol Chem 274:24838
15. Leppänen A, White SP, Helin J, McEver RP, Cummings RD (2000) J Biol Chem 275:39569
16. Leppänen A, White SP, Helin J, McEver RP, Cummings RD (2000) Glycobiol 10:1106
17. Koeller KM, Smith MEB, Wong CH (2000) J Am Chem Soc 122:742
18. Koeller KM, Smith MEB, Huang RF, Wong CH (2000) J Am Chem Soc 122:4241
19. Izumi M, Shen GJ, Wacowich-Sgarbi S, Nakatani T, Plettenburg O, Wong CH (2001) J Am Chem Soc 123:10909
20. Dudziak G, Bezay N, Schwientek T, Clausen H, Kunz H, Liese A (2000) Tetrahedron 56:5865
21. Zeng XX, Nakaaki Y, Murata T, Usui T (2000) Arch Biochem Biophys 383:28
22. Gambert U, Thiem J (1999) Eur J Org Chem 1999:107
23. George SK, Holm B, Reis CA, Schwientek T, Clausen H, Kihlberg J (2001) J Chem Soc Perkin Trans 1:880
24. George SK, Schwientek T, Holm B, Reis CA, Clausen H, Kihlberg J (2001) J Am Chem Soc 123:11117
25. Blixt O, Allin K, Pereira L, Datta A, Paulson JC (2002) J Am Chem Soc 124:5739
26. Unverzagt C, Kunz H, Paulson JC (1990) J Am Chem Soc 112:9308
27. Unverzagt C, Kelm S, Paulson JC (1994) Carbohydr Res 251:285
28. Unverzagt C (1996) Angew Chem Int Ed Engl 35:2350
29. Unverzagt C (1997) Carbohydr Res 305:423
30. Schuster M, Wang P, Paulson JC, Wong CH (1994) J Am Chem Soc 116:1135
31. Wong CH, Schuster M, Wang P, Sears P (1993) J Am Chem Soc 115:5893
32. Tai VWF, Imperiali B (2001) J Org Chem 66:6217
33. Walsh CT, Losey HC, Meyers CLF (2003) Biochem Soc Trans 31:487
34. Walsh C, Meyers CLF, Losey HC (2003) J Med Chem 46:3425
35. Solenberg PJ, Matsushima P, Stack DR, Wilkie SC, Thompson RC, Baltz RH (1997) Chem Biol 4:195
36. Losey HC, Jiang JQ, Biggins JB, Oberthur M, Ye XY, Dong SD, Kahne D, Thorson JS, Walsh CT (2002) Chem Biol 9:1305
37. Fu X, Albermann C, Jiang JQ, Liao JC, Zhang CS, Thorson JS (2003) Nat Biotechnol 21:1467

38. Losey HC, Peczuh MW, Chen Z, Eggert US, Dong SD, Pelczer I, Kahne D, Walsh CT (2001) Biochemistry (Moscow) 40:4745
39. Watt GM, Lowden PAS, Flitsch SL (1997) Curr Opin Struct Biol 7:652
40. Crout DHG, Vic G (1998) Curr Opin Chem Biol 2:98
41. Palcic MM (1999) Curr Opin Biotechnol 10:616
42. Koeller KM, Wong CH (2000) Chem Rev 100:4465
43. Ajisaka K, Miyasato M (2000) Biosci Biotechnol Biochem 64:1743
44. Suzuki K, Fujimoto H, Ito Y, Sasaki T, Ajisaka K (1997) Tetrahedron Lett 38:1211
45. Gallego RG, Dudziak G, Kragl U, Wandrey C, Kamerling JP, Vliegenthart JFG (2003) Biochimie 85:275
46. Hokke CH, van den Eijnden DH (1997) Carbohydr Res 305:463
47. Jahn M, Withers SG (2003) Biocatal Biotransform 21:159
48. Perugino G, Trincone A, Rossi M, Moracci M (2004) Trends Biotechnol 22:31
49. Tolborg JF, Petersen L, Jensen KJ, Mayer C, Jakeman DL, Warren RAJ, Withers SG (2002) J Org Chem 67:4143
50. Kajihara Y, Suzuki Y, Yamamoto N, Sasaki K, Sakakibara T, Juneja LR (2004) Chem Eur J 10:971
51. Fujita K, Tanaka N, Sano M, Kato I, Asada Y, Takegawa K (2000) Biochem Biophys Res Commun 267:134
52. Singh S, Ni JH, Wang LX (2003) Bioorg Med Chem Lett 13:327
53. Li HG, Singh S, Zeng Y, Song HJ, Wang LX (2005) Bioorg Med Chem Lett 15:895
54. Wang LX, Song H, Liu S, Lu H, Jiang S, Ni J, Li H (2005) Chem Bio Chem 6:1068
55. Li B, Zeng Y, Hauser S, Song H, Wang LX (2005) J Am Chem Soc 127:9692
56. Yamamoto K (2001) J Biosci Bioeng 92:493
57. Haneda K, Inazu T, Yamamoto K, Kumagai H, Nakahara Y, Kobata A (1996) Carbohydr Res 292:61
58. Yamamoto K, Fujimori K, Haneda K, Mizuno M, Inazu T, Kumagai H (1997) Carbohydr Res 305:415
59. Haneda K, Inazu T, Mizuno M, Iguchi R, Yamamoto K, Kumagai H, Aimoto S, Suzuki H, Noda T (1998) Bioorg Med Chem Lett 8:1303
60. Mizuno M, Haneda K, Iguchi R, Muramoto I, Kawakami T, Aimoto S, Yamamoto K, Inazu T (1999) J Am Chem Soc 121:284
61. Haneda K, Inazu T, Mizuno M, Iguchi R, Tanabe H, Fujimori K, Yamamoto K, Kumagai H, Tsumori K, Munekata E (2001) Biochim Biophys Acta 1526:242
62. Saskiawan I, Mizuno M, Inazu T, Haneda K, Harashima S, Kumagai H, Yamamoto K (2002) Arch Biochem Biophys 406:127
63. Akaike E, Tsutsumida M, Osumi K, Fujita M, Yamanoi T, Yamamoto K, Fujita K (2004) Carbohydr Res 339:719
64. Kato T, Fujita K, Takeuchi M, Kobayashi K, Natsuka S, Ikura K, Kumagai H, Yamamoto Y (2002) Glycobiology 12:581
65. Muramatsu T (2000) Glycoconj J 17:577
66. Yamamoto S, Muramatsu H, Muramatsu T (2005) Glycoconj J 22:35
67. Trimble RB, Atkinson PH, Tarentino AL, Plummer TH, Maley F, Tomer KB (1986) J Biol Chem 261:2000
68. Bardales RM, Bhavanandan VP (1989) J Biol Chem 264:19893
69. Ashida H, Yamamoto K, Murata T, Usui T, Kumagai H (2000) Arch Biochem Biophys 373:394
70. Ishiikarakasa I, Iwase H, Hotta K, Tanaka Y, Omura S (1992) Biochem J 288:475
71. Ajisaka K, Miyasato M, Ishii-Karakasa I (2001) Biosci Biotechnol Biochem 65:1240

72. Ishido K, Takagaki K, Iwafune M, Yoshihara S, Sasaki M, Endo M (2002) J Biol Chem 277:11889
73. Witte K, Sears P, Martin R, Wong CH (1997) J Am Chem Soc 119:2114
74. Hackenberger CPR, O'Reilly MK, Imperiali B (2005) J Org Chem 70:3574
75. Malet C, Planas A (1998) FEBS Lett 440:208
76. Faijes M, Fairweather JK, Driguez H, Planas A (2001) Chem Eur J 7:4651
77. Faijes M, Perez X, Perez O, Planas A (2003) Biochemistry (Moscow) 42:13304
78. Faijes M, Ima T, Bulone V, Planas A (2004) Biochem J 380:635
79. van Lieshout J, Faijes M, Nieto J, van der Oost J, Planas A (2004) Archaea 1:285
80. Bordusa F (2002) Chem Rev 102:4817
81. Jackson DY, Burnier J, Quan C, Stanley M, Tom J, Wells JA (1994) Science 266:243
82. Sears P, Schuster M, Wang P, Witte K, Wong CH (1994) J Am Chem Soc 116:6521
83. Moree WJ, Sears P, Kawashiro K, Witte K, Wong CH (1997) J Am Chem Soc 119:3942
84. Nakatsuka T, Sasaki T, Kaiser ET (1987) J Am Chem Soc 109:3808
85. Abrahmsen L, Tom J, Burnier J, Butcher KA, Kossiakoff A, Wells JA (1991) Biochemistry (Moscow) 30:4151
86. Zhong ZY, Wong CH (1991) Biomed Biochim Acta 50:S9
87. Witte K, Seitz O, Wong CH (1998) J Am Chem Soc 120:1979
88. Tolbert TJ, Wong CH (2004) Methods Mol Biol 283:267
89. Dawson PE, Muir TW, Clarklewis I, Kent SBH (1994) Science 266:776
90. Dawson PE, Kent SBH (2000) Ann Rev Biochem 69:923
91. Erlanson DA, Chytil M, Verdine GL (1996) Chem Biol 3:981
92. Muir TW (2001) Synlett:733
93. Harris JL, Backes BJ, Leonetti F, Mahrus S, Ellman JA, Craik CS (2000) Proc Natl Acad Sci USA 97:7754
94. Matthews DJ, Wells JA (1993) Science 260:1113
95. Tolbert TJ, Wong CH (2002) Angew Chem Int Ed 41:2171
96. Tolbert TJ, Wong CH (2004) Methods Mol Biol 283:255
97. Tolbert TJ, Franke D, Wong CH (2005) Bioorg Med Chem 13:909
98. Kadereit D, Waldmann H (2001) Chem Rev 101:3367
99. Ishii H, Funabashi K, Mimura Y, Inoue Y (1990) Bull Chem Soc Jpn 63:3042
100. Cantacuzene D, Attal S, Bay S (1991) Bioorg Med Chem Lett 1:197
101. Cantacuzene D, Attal S, Bay S (1991) Biomed Biochim Acta 50:S231
102. Eberling J, Braun P, Kowalczyk D, Schultz M, Kunz H (1996) J Org Chem 61:2638
103. Tokuda Y, Takahashi Y, Matoishi K, Ito Y, Sugai T (2002) Synlett:57
104. Hermann P (1991) Biomed Biochim Acta 50:S19
105. Schultz M, Hermann P, Kunz H (1992) Synlett:37
106. Reissmann S, Greiner G (1992) Int J Pept Protein Res 40:110
107. Attal S, Bay S, Cantacuzene D (1992) Tetrahedron 48:9251
108. Reetz MT (2002) Curr Opin Chem Biol 6:145
109. Koeller KM, Wong CH (2001) Nature 409:232
110. Pathak T, Waldmann H (1998) Curr Opin Chem Biol 2:112
111. Margolin AL, Klibanov AM (1987) J Am Chem Soc 109:3802
112. Gewehr M, Kunz H (1997) Synthesis:1499
113. Eisele F, Owen DJ, Waldmann H (1999) Bioorg Med Chem 7:193
114. Waldmann H, Braun P, Kunz H (1991) Biomed Biochim Acta 50:S243
115. Braun P, Waldmann H, Kunz H (1992) Synlett:39
116. Braun P, Waldmann H, Kunz H (1993) Bioorg Med Chem 1:197
117. Braun P, Davies GM, Price MR, Williams PM, Tendler SJB, Kunz H (1998) Bioorg Med Chem 6:1531

118. Kunz H, Kowalczyk D, Braun P, Braum G (1994) Angew Chem Int Ed Engl 33:336
119. Sander J, Waldmann H (1999) Angew Chem Int Ed 38:1250
120. Sander J, Waldmann H (2000) Chem Eur J 6:1564
121. Pohl T, Waldmann H (1997) J Am Chem Soc 119:6702
122. Bowman KG, Bertozzi CR (1999) Chem Biol 6:R9
123. Chapman E, Best MD, Hanson SR, Wong CH (2004) Angew Chem Int Ed 43:3526
124. Burkart MD, Izumi M, Wong CH (1999) Angew Chem Int Ed 38:2747
125. Burkart MD, Wong CH (1999) Anal Biochem 274:131
126. Burkart MD, Izumi M, Chapman E, Lin CH, Wong CH (2000) J Org Chem 65:5565
127. Kruger RG, Lu W, Oberthur M, Tao JH, Kahne D, Walsh CT (2005) Chem Biol 12:131
128. Ramos D, Rollin P, Klaffke W (2000) Angew Chem Int Ed 39:396
129. Ramos D, Rollin P, Klaffke W (2001) J Org Chem 66:2948
130. Ohta T, Miura N, Fujitani N, Nakajima F, Niikura K, Sadamoto R, Guo CT, Suzuki T, Suzuki Y, Monde K, Nishimura SI (2003) Angew Chem Int Ed 42:5186
131. Sato M, Sadamoto R, Niikura K, Monde K, Kondo H, Nishimura SI (2004) Angew Chem Int Ed 43:1516
132. Sato M, Furuike T, Sadamoto R, Fujitani N, Nakahara T, Niikura K, Monde K, Kondo H, Nishimura SI (2004) J Am Chem Soc 126:14013
133. Lin H, Thayer DA, Wong CH, Walsh CT (2004) Chem Biol 11:1635
134. Lin H, Walsh CT (2004) J Am Chem Soc 126:13998
135. Wang L, Schultz PG (2004) Angew Chem Int Ed Engl 44:34
136. Liu H, Wang L, Brock A, Wong CH, Schultz PG (2003) J Am Chem Soc 125:1702
137. Zhang Z, Gildersleeve J, Yang YY, Xu R, Loo JA, Uryu S, Wong CH, Schultz PG (2004) Science 303:371
138. Xu R, Hanson SR, Zhang Z, Yang YY, Schultz PG, Wong CH (2004) J Am Chem Soc 126:15654
139. Wildt S, Gerngross TU (2005) Nat Rev Microbiol 3:119

Top Curr Chem (2007) 267: 65–107
DOI 10.1007/128_2006_104

Synthesis and Application of Glycopeptide and Glycoprotein Mimetics

Daniel Specker · Valentin Wittmann (✉)

Fachbereich Chemie, Universität Konstanz, 78457 Konstanz, Germany
mail@valentin-wittmann.de

Abstract Glycosylation of proteins is the most complex form of posttranslational modification. Glycan chains of glycoproteins are involved in numerous biological recognition events, such as protein folding, cell–cell communication and adhesion, cell growth and differentiation, as well as bacterial and viral infection. As an alternative to glycoprotein synthesis, approaches to the construction of glycopeptide and glycoprotein mimetics have been developed. These so-called neoglycopeptides and neoglycoproteins, respectively, provide insight on the importance of distinct structural elements on biological activity. Furthermore, they may have improved properties, such as increased stability in the case of *S*- and *C*-glycopeptides. This article summarizes recent progress in the field of synthetic glycosylated amino acids, peptides, and proteins with non-natural structural elements with a focus on the carbohydrate–peptide linkage. Linear and convergent approaches are discussed. Application of chemoselective ligation reactions provides access to pure glycoforms of glycoprotein mimetics.

Keywords Glycopeptides · Glycoproteins · Glycobiology · Neoglycoconjugates · Bioconjugation · Chemoselective ligation

Abbreviations

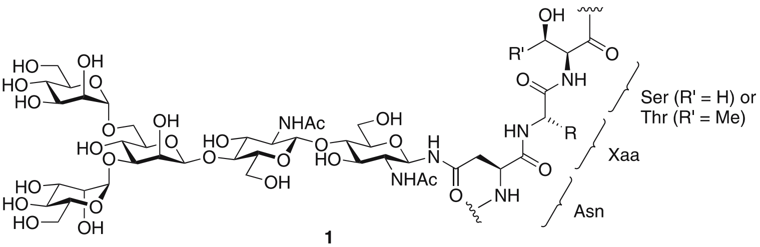
Ac	Acetyl
All	Allyl
Aloc	Allyloxycarbonyl
Azy	Aziridine-2-carboxylic acid
Bn	Benzyl
Boc	*tert*-Butoxycarbonyl
Bu	Butyl
t-Bu	*tert*-Butyl
Bz	Benzoyl
CM	Cross-metathesis
COD	Cycloocta-1,5-diene
dba	Dibenzylidene acetone
DBU	1,8-Diazabicyclo[5.4.0] undec-7-ene
DCL	Dynamic combinatorial library
Ddv	1-(4,4-dimethyl-2,6-dioxocyclohexylidene)isovaleryl
DIC	*N*,*N*′-Diisopropylcarbodiimide
DIEA	*N*,*N*′-Diisopropylethylamine
DMAP	4-(Dimethylamino)pyridine
DMF	*N*,*N*-dimethylformamide
dppf	1,1′-Bis(diphenylphosphanyl)ferrocene
DTT	Dithiothreitol
EDT	Ethane-1,2-dithiol
EPO	Erythropoietin
Et	Ethyl
FITC	Fluorescein isothiocyanate
Fmoc	9-Fluorenylmethoxycarbonyl
HATU	2-(1*H*-7-Azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HBTU	2-(1*H*-Benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HOBt	1-Hydroxybenzotriazole
KHMDS	Potassium hexamethyl disilazide
ManLev	*N*-Levulinoylmannosamine
MBHA	4-Methylbenzhydrylamine
Me	Methyl
NaHMDS	Sodium hexamethyl disilazide
NEM	*N*-Ethylmorpholine
NMM	*N*-Methylmorpholine
Pfp	Pentafluorophenyl
Ph	Phenyl
Pht	Phthaloyl
Piv	Pivaloyl
PTM	Posttranslational modification
py	Pyridine
PyBOP	Benzotriazol-1-yloxytripyrrolidinophosphonium hexafluorophosphate
rt	Room temperature
SBL	Subtilisin *Bacillus lentus*

SPPS	Solid-phase peptide synthesis
Su	Succinimidyl
TBAHS	Tetra-*n*-butylammonium hydrogen sulfate
TBS	*tert*-Butyldimethylsilyl
TFA	Trifluoroacetic acid
Tfa	Trifluoroacetyl
THF	Tetrahydrofuran
TIPS	Triisopropylsilyl
TMS	Trimethylsilyl
Ts	Tosyl
Z	Benzyloxycarbonyl

1
Introduction

The most surprising result of the human genome project probably was the fact that our genome consists of less than 30 000 genes. This low number underlines the importance of posttranslational modifications (PTMs) in order to reach the complexity of higher organisms. Among the PTM events that are known today, glycosylation is the most complex form and requires about 1% of mammalian genes [1]. The functions of the glycan chains of glycoproteins span the complete spectrum from being essentially unimportant to being crucial for the survival of an organism [1–10]. Glycan chains of glycoproteins modulate physicochemical properties of proteins, such as solubility, viscosity, charge, conformation, and dynamics and thereby can bestow stability and resistance to proteolytic degradation and can lead to improved protein efficacy. On the other hand, the glycans provide unique epitopes for molecular recognition that are involved in cell–cell communication, cell growth and differentiation, cancer metastasis, bacterial and viral infection, and they direct protein folding.

Oligosaccharides are attached to proteins mainly via an *N*-glycosidic bond to asparagine (Fig. 1) or an *O*-glycosidic bond to hydroxylated amino acids,



1

Fig. 1 Common structure of all *N*-glycoproteins: core pentasaccharide attached to the consensus sequence Asn-Xaa-Ser/Thr

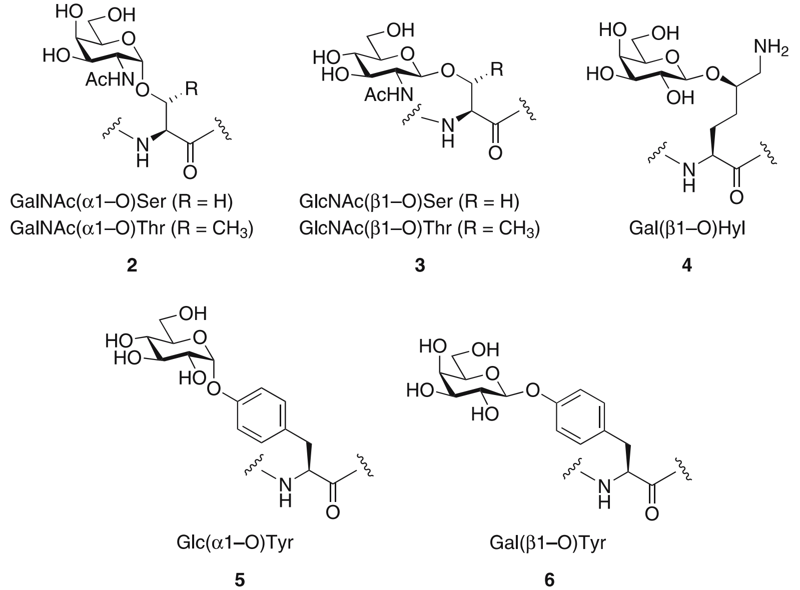


Fig. 2 Selected carbohydrate–peptide linkages found in *O*-glycoproteins

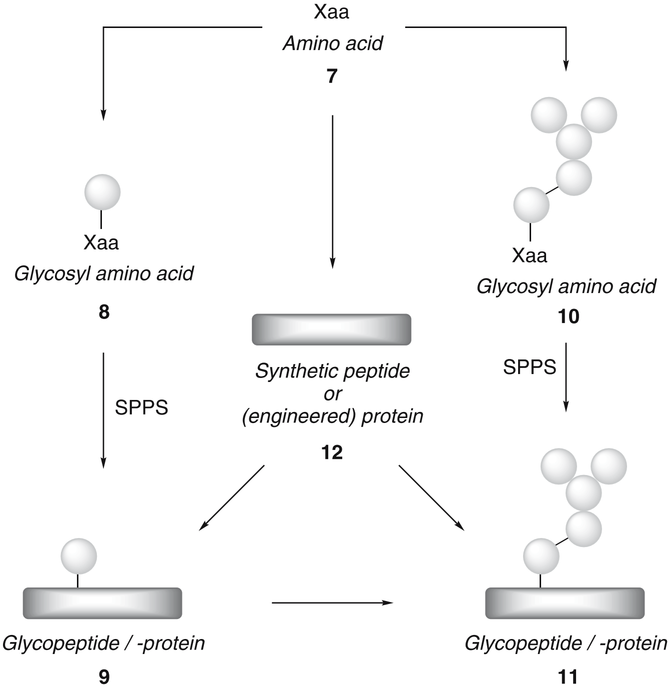
such as serine and threonine (Fig. 2) [11–15]. Whereas all *N*-glycoproteins share the same core structure **1**, which is derived from a common biosynthetic oligosaccharide precursor, a variety of carbohydrate–peptide linkages are found in *O*-glycoproteins. In the mucins, the GalNAc(*α*1–O)Ser/Thr linkage **2** (also called Tn antigen) is found. Another widely occurring *O*-glycosidic bond is the GlcNAc(*β*1–O)Ser/Thr linkage **3**. Less common *O*-glycosidic linkages include those to hydroxylysine (**4**) and tyrosine (**5**, **6**).

Since the glycan chains are secondary gene products that are not synthesized in a template-controlled fashion, they usually occur as complex mixtures of glycoforms, making their isolation and analysis a demanding task. Chemists, therefore, have developed synthetic strategies for the preparation of pure glycoforms of proteins required for the evaluation of structure–function relationships [16–26]. The total synthesis of glycoproteins, however, is still an enormous challenge and remains the domain of a few specialized laboratories around the world.

The synthesis of glycopeptide and glycoprotein mimetics, also referred to as neoglycoconjugates [27], represents an alternative strategy to produce single glycoforms for biological studies. The replacement of a natural structure with a non-natural one allows study of the influence of distinct structural elements on biological activity, but has many practical applications as well. Use of chemoselective ligation reactions makes glycoconjugate synthesis accessible to a broader community. Furthermore, *S*- and *C*-glycosidic bonds for example are more stable than the corresponding *O*-glycosides (both chemically and with respect to enzymatic degradation), which leads to an increased half life of a glycoconjugate within a biological system. Thus, beneficial properties of glycans may be permanently linked to a protein, which is an important aspect for pharmaceutical applications. This article summarizes recent progress in the field of synthetic glycosylated amino acids, peptides, and proteins with non-natural structural elements, with a focus on the carbohydrate–peptide linkage. For a more complete coverage of this wide area, the reader is referred to other excellent reviews [23, 28–38].

2
Strategies for Assembly of Glycopeptide and Glycoprotein Mimetics

Scheme 1 summarizes the main strategies that have been explored in the past for the preparation of glycopeptide and glycoprotein mimetics. Two major approaches can be distinguished: the linear and the convergent assembly. In the linear assembly, carbohydrates are coupled to amino acids **7** to give modified amino acids carrying either mono- (**8**) or oligosaccharides (**10**). These are used as building blocks in solution or (more commonly) solid-phase pep-



Scheme 1 Linear and convergent strategies that have been employed for the synthesis of glycopeptide and glycoprotein mimetics

tide synthesis (SPPS) to provide glycoconjugates **9** and **11**, respectively. Linear approaches are covered in Sect. 3 of the article. In convergent approaches, carbohydrates are coupled to presynthesized peptides or to proteins (**12**) to give **9** and **11**. Depending on the type of chemistry used for carbohydrate attachment, there might be a need for protecting groups at the peptide and carbohydrate, an approach which is confined to synthetic peptides. Conjugation of carbohydrates to full-length proteins is possible via chemoselective ligation to amino acids with a unique reactivity, for example cysteine residues. Combined with site-directed mutagenesis, this approach allows for control of both site of attachment and type of saccharide. Alternatively, genetically engineered proteins containing non-proteinogenic amino acids with a unique chemical reactivity can be employed.

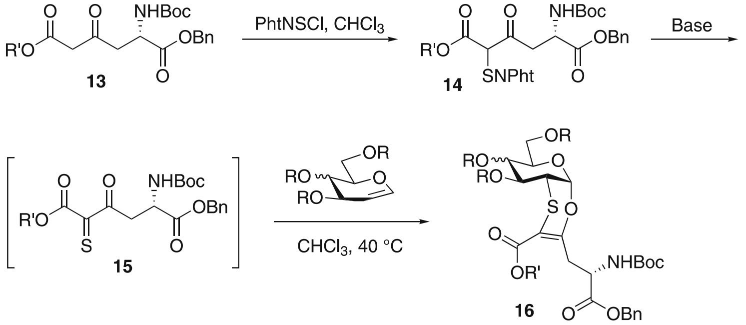
Mixed linear and convergent assembly strategies have also been described. Glycopeptide mimetics **9** or **11** synthesized by the linear route have been extended in the glycan part by convergent attachment of an oligosaccharide through chemoselective ligation. Convergent and mixed approaches for assembly of glycopeptide and glycoprotein mimetics are dealt with in Sect. 4. Not shown in Scheme 1 are the assembly of peptide and neoglycopeptide fragments by native chemical ligation [39] and the incorporation of non-proteinogenic glycosylated amino acids into proteins by in vivo suppressor tRNA technology [40–42].

3
Synthesis of Glycopeptide Mimetics by Linear Assembly of Glycosylated Amino Acids with Unnatural Linkages

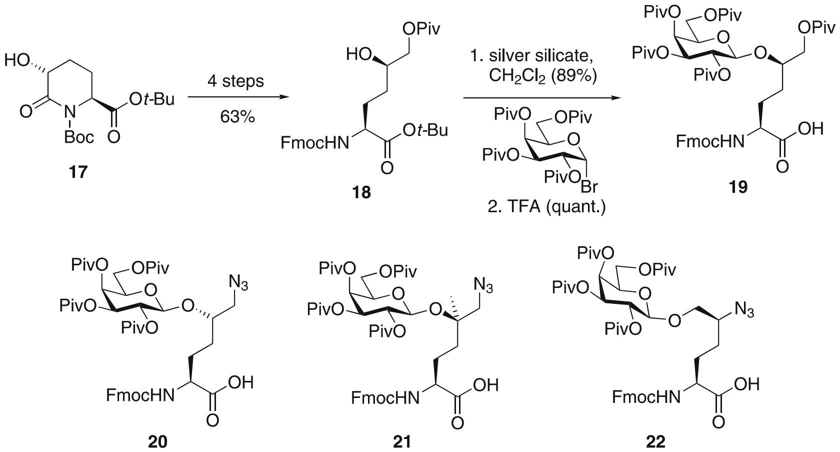
3.1
O-Glycosides

Several groups have described building blocks for the preparation of glycopeptide mimetics containing *O*-glycosidic bonds. Based on chemistry developed earlier [43], Nativi et al. prepared α-O-linked glycohomoglutamates **16** as replacement for naturally occurring α-*O*-glycopeptides found in mucins (Scheme 2) [44]. β-Ketoesters **13** obtained from aspartic acid and Meldrum's acid were reacted with phthalimidosulfenyl chloride. Base treatment of **14** gave reactive intermediate **15**, which underwent a hetero Diels–Alder reaction with glycals to give α-*O*-glycohomoglutamates **16** as diasteriomerically pure isomers. The same approach has been used by Franck et al. for the preparation of *N*-glycosides [45].

Preparation of the four galactosylated 5-hydroxylysine mimetic building blocks **19–22** was accomplished by Guichard and coworkers (Scheme 3) [46]. The synthesis of the 5-hydroxy amino acid derivative **18** started from hydroxylated piperidinone **17**. Subsequent galactosylation gave **19**. The use of



Scheme 2 Preparation of α-*O*-glycohomoglutamates **16** through hetero Diels–Alder reaction with glycals



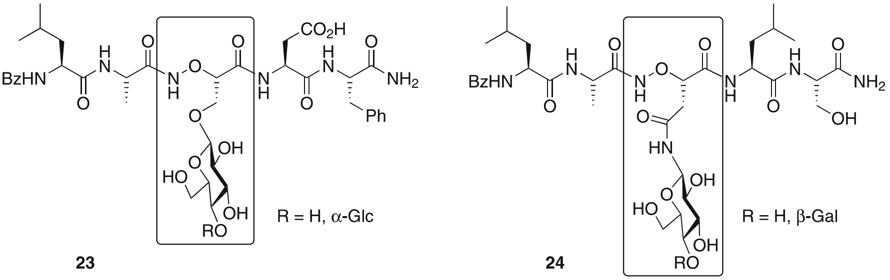
Scheme 3 Synthesis of building blocks for the preparation of collagen mimetics

pivaloyl-protected galactosyl bromide prevented undesired orthoester formation during the glycosylation step. Building blocks **20–22** were obtained analogously. The galactosylated hydroxylysine mimetics were incorporated into bovine type II collagen (bCII) immunodominant glycopeptides in order to study the fine specificity of bCII-reactive T cells involved in the initiation and/or regulation of collagen-induced arthritis, a mouse model for rheumatoid arthritis [47].

Danishefsky et al. applied β-*O*-allyl glycosides of mono- and oligosaccharides, including the breast and prostate cancer antigen Globo-H, for a cross-metathesis reaction with Fmoc-L-allylglycine benzyl ester, followed by reduction of the resulting olefins via catalytic hydrogenation, with the concomitant release of the free acid [48]. The obtained non-natural glycosyl amino acids can readily be applied in solid-phase synthesis of potential glycopeptide-based antitumor vaccines.

Peptoids (N-substituted oligoglycines) are peptide mimetics that are characterized by an improved proteolytic stability and high conformational flexibility [49, 50]. Roy and coworkers were the first to synthesize glycosylated derivatives of peptoids, which they named glycopeptoids. Several O-linked [51, 52] but also N-linked [53, 54] and C-linked [52] variants were prepared and used as building blocks for the construction of HIV-1 protease inhibitors [55] and Tn-antigen clusters [56].

Glycopeptide mimetics with a modified peptide backbone in order to reduce biodegradability have been reported by Shin et al. [57]. They synthesized the pentamers **23** and **24** containing O- and N-glycosylated α-aminooxy acids by fragment coupling on solid support (Scheme 4).

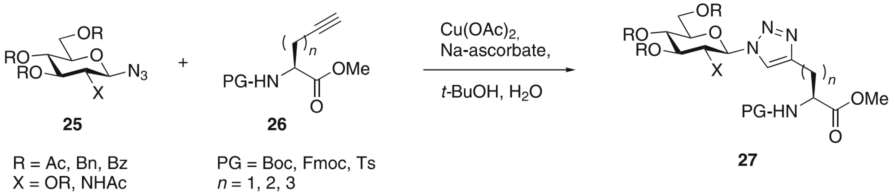


Scheme 4 Glycopeptide mimetics with a peptide backbone containing O- and N-glycosylated α-aminooxy acids

3.2
N-Glycosides

In recent years, the copper(I)-catalyzed [3 + 2] cycloaddition of azides and alkynes [58, 59] (often referred to as "click" reaction) has found wide application in the field of bioconjugation [60], including the preparation of glycopeptide mimetics. Depending on whether the azide or the alkyne component is connected to the carbohydrate, *N*-glycosides and *C*-glycosides, respectively, are obtainable. Rutjes and coworkers reacted various glycosyl azides **25** with a series of acetylenic amino acids **26** to obtain triazole-linked glycosyl amino acids **27** and also some peptides (Scheme 5) [61]. With a few exceptions, the yields were generally good to excellent. The triazole linkage appeared to be stable under several acidic and basic conditions (e.g., 1 M HCl in MeOH, 6 h, reflux or 1.25 M aqueous NaOH, 24 h, reflux). Using α- and β-glycosyl acetylenes, several C-glycosidically linked triazole-containing glycosyl amino acids were synthesized.

The azide-alkyne cycloaddition was also employed for the preparation of carbohydrate-modified 2(1*H*)-pyrazinones as precursors of glycopep-

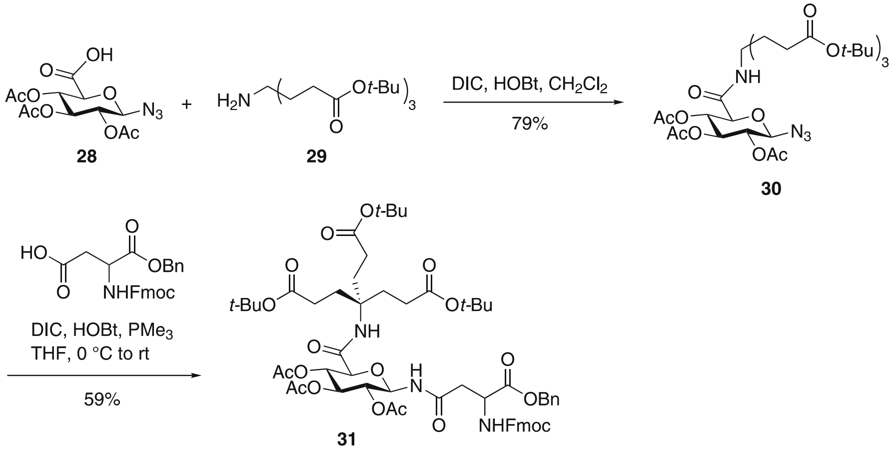


Scheme 5 Application of the azide-alkyne cycloaddition for the preparation of triazole-linked glycosyl amino acids

tidomimetics [62] and also for the preparation of glycopeptide and glycoprotein mimetics by convergent strategies (cf. Sect. 4.4).

A different approach towards non-natural *N*-glycosyl amino acids was reported by Klaffke and coworkers [63, 64]. They used disaccharides *N*-glycosidically linked to spacers bearing a second amino function at the other end. This primary amine was linked to the γ-carboxamide function of glutamine within the dipeptide Z-Gln-Gly-OH in a reaction catalyzed by the enzyme transglutaminase. The needed spacer-modified *N*-glycosides were accessible by reaction of the disaccharide (e.g., maltose) with allylamine followed by acetylation and photochemical thioetherification with cysteamine hydrochloride.

Starting from azido-functionalized glucuronic acid **28**, Lindhorst and coworkers prepared glycosyl amino acids with a dendron moiety attached to the sugar giving access to branched glycopeptide mimetics of variable sugar composition (Scheme 6) [65]. Serving as the branching element, amino triester **29** was introduced under standard peptide coupling conditions. Resulting **30**



Scheme 6 Application of a modified Staudinger reaction for the synthesis of dendron-functionalized *N*-glycoside **31**

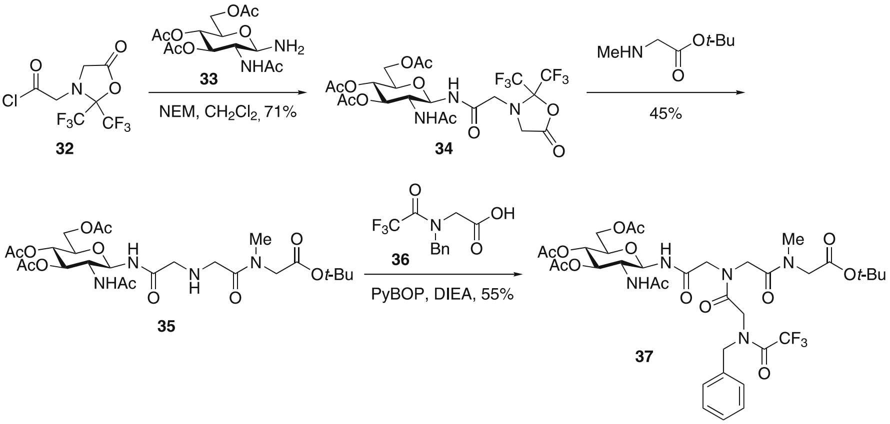
was converted into the dendron-functionalized *N*-glycoside **31** via a modified Staudinger reaction.

Van Ameijde and Liskamp coupled N-linked glyco amino acid building blocks, in which glycosyl amines are connected to the carboxy group of an amino acid via an amide bond, to trivalent cyclotriveratrylene scaffolds with varying spacers [66]. The obtained glycoclusters can be used to study multivalent interactions. Ichikawa et al. explored the use of Steyermark's glucopyranosyl oxazolidinone for anchoring glucosyl moieties onto amines and thiols in aqueous media [67]. The method was applied for the synthesis of a urea-tethered glucosyl lysine conjugate and a corresponding cysteine conjugate.

Thiem and coworkers prepared a N^4-(2-acetamido-2-deoxy-β-D-glucopyranosyl)-L-asparagine analogue in which the side-chain carboxy group of aspartic acid is linked to the 2-amino function of 1,5-anhydro-2-amino-2-deoxyglucitol [68]. The building block was incorporated into a mimic of the V3 loop of the HIV-1 gp 120 envelope glycoprotein by solid-phase peptide synthesis (SPPS).

Inspired by the work of Roy et al. [53, 54], already mentioned in Sect. 3.1, Burger and coworkers reported a preparatively simple synthesis of N-linked glycopeptoids starting from iminodiacetic acid using hexafluoroacetone as protecting and activating reagent (Scheme 7) [69]. Dielectrophile **32** reacts with weak nucleophiles such as glycosylamine **33** at the more reactive acid chloride to give the N-glycosylated activated glycine derivative **34**. Treatment with sarcosine-*tert*-butylester provided dipeptoide **35** under concurrent deprotection of the α-amino group. N-Terminal elongation with **36** gave tripeptoide **37**.

Vankar and coworkers recently introduced the novel reagent system trimethylsilylnitrate/trimethylsilyl azide for the synthesis of 2-deoxyglycosyl



Scheme 7 Synthesis of N-linked glycopeptoide **37**

azides from glycals [70]. After Staudinger reduction, the obtained glycosylamines were applied in the synthesis of 2-deoxy-*β*-*N*-glycopeptides.

3.3
S-Glycosides

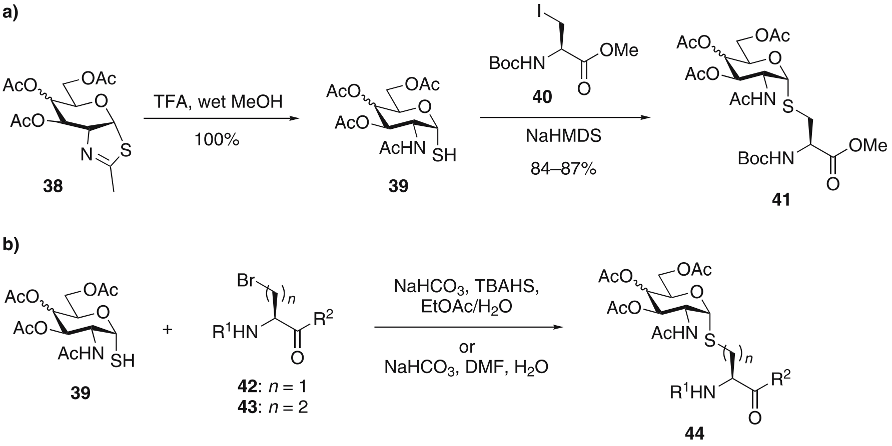
Replacement of the anomeric oxygen of *O*-glycosides by sulfur to furnish *S*-glycosides leads to an increased stability towards acid/base and enzymatic hydrolysis. The preparation of S-linked glycopeptides/proteins has been covered by several excellent reviews [23, 32, 38, 71], most notably a very recent one by Pachamuthu and Schmidt [28]. The present article, therefore, will focus on selected examples.

In the past, two main strategies have been employed for the synthesis of S-linked glycosyl amino acids. One uses an anomeric thiolate nucleophile, which is generally configurationally stable, in S_N2 reaction with a halogen-functionalized amino acid, often a 3-bromoalanine derivative. The major side reactions are disulfide formation via air oxidation and *β*-elimination to the corresponding dehydroalanine followed by Michael addition, which results in an epimeric mixture at the *α*-carbon of the amino acid product. Alternatively, a thiol-containing amino acid, e.g., cysteine or homocysteine, is glycosylated by reaction with a glycosyl donor. The reaction can be carried out under basic conditions employing a halogenose leading to an S_N2 displacement. This is particularly convenient for equatorial *S*-glycosides because the halogen atom is generally found in the axial position. Access to axial thioglycosides is given by acid-catalyzed S-glycosylation following an S_N1-type mechanism.

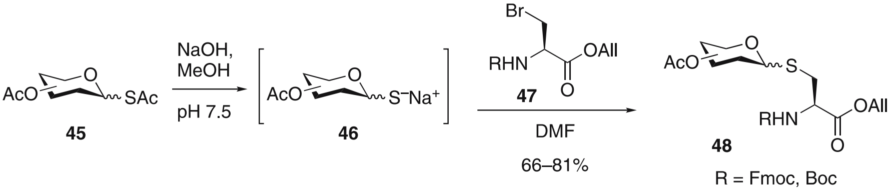
Knapp and Myers used thiazolines **38**, derived from peracetylated glucosamine and galactosamine, respectively, by treatment with Lawesson's reagent, for the formation of *α*-GlcNAc and *α*-GalNAc thiols **39** (Scheme 8a) [72, 73]. Alkylation with 3-iodoalanine derivative **40** in the presence of base provided the S-glycosylated amino acids **41** in high yields without apparent epimerization at the *α*-carbon of the amino acid.

A similar approach was applied by Zhu and Schmidt for the synthesis of S-glycosylated amino acids and di- and tripeptides **44** (Scheme 8b) [74, 75]. They reacted *α*-GlcNAc and *α*-GalNAc thiols **39** as well as *β*-GlcNAc and *β*-lactosyl thiols with 3-bromoalanine (**42**) and 4-bromohomoalanine derivatives (**43**) under phase transfer conditions or in a mixture of DMF and water at pH 8.5. Due to the mild conditions, epimerization was not observed and even Fmoc-protected peptides were applicable.

Wong et al. modified the procedure and developed a two-step one-pot reaction with in situ generation of sugar thiolates **46** (Scheme 9) [233]. Glycosyl thioacetates **45** were first S-deacetylated with NaOH in methanol at pH 7.5 followed by reaction with Boc- and Fmoc-protected 3-bromoalanines **47** to get a great variety of S-linked glycosyl amino acids **48.** Two of them were incorporated into a linear and a cyclic analogue of the cyclic cationic decapeptide



Scheme 8 Approaches to S-linked glycosyl amino acids using anomeric thiolate nucleophiles

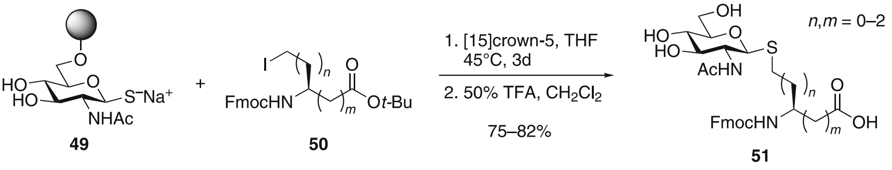


Scheme 9 A two-step one-pot reaction for the preparation of S-linked glycosyl amino acids **48** generating sugar thiolates **46** in situ

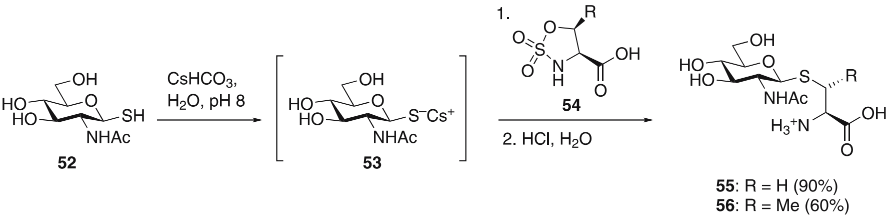
antibiotic tyrocidine A, which is produced in *Bacillus brevis*. The cyclic analogue was shown to have a better antibiotic profile than tyrocidine A.

An efficient solid phase strategy for the synthesis of *β*-*S*-glycosyl amino acids **51** was described by Jobron and Hummel (Scheme 10) [76]. Based on an earlier report [77], they used unprotected sugar thiols immobilized via the 6 position to generate sodium thiolates **49** for nucleophilic attachment to different iodine-substituted Fmoc-protected amino acids **50**.

For the stereoselective synthesis of S-linked glycosyl amino acids, e.g., **55** and **56**, Cohen and Halcomb prepared serine- and *allo*-threonine-derived



Scheme 10 Solid-phase synthesis of *β*-*S*-glycosyl amino acids **51**



Scheme 11 Application of cyclic sulfamidates **54** (R = H, Me) for the synthesis of S-linked glycosyl amino acids

cyclic sulfamidates **54** (R = H, Me) (Scheme 11) [78, 79]. In their approach, the sulfamidates reacted with in situ-formed *β*-cesium thiolates, such as **53**, to give the unprotected thioglycosides after hydrolysis of the *N*-sulfates. The use of cesium as the counter ion afforded higher yields due to an increased solubility of the thiolate, allowing the application of higher concentrations. The serine-derived sulfamidate was also incorporated into a peptide on a solid support. Addition of a 1-thio sugar afforded a solid-phase bound *S*-glycodipeptide, which could be extended to a tripeptide using standard peptide synthesis procedures.

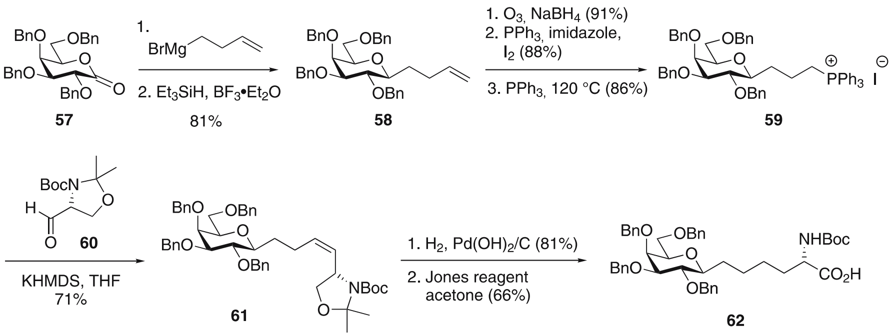
Direct base-catalyzed S-glycosylation of cysteine- and homocysteine-containing peptides using protected glycosyl bromides was reported by Schmidt and coworkers [80]. They reacted cysteine-centered tripeptides, which were accessible via chemical ligation, with peracetylated glucosyl and lactosyl bromides under two-phase conditions and observed the formation of S-linked glycopeptides in very good yields. The strategy was also applied for the synthesis of an S-linked glycopeptide analogue derived from human Tamm–Horsfall glycoprotein in the same laboratory [81].

Two types of novel *S*-neoglycopeptides were published by Schmidt et al. [82]. Glycosylthiomethyl amides were obtained by reaction of glycosylthiomethyl azides with amino acids and peptides containing aspartate and glutamate thio acids, whereas reaction of glycosylthiomethyl bromides with cysteine and homocysteine derivatives afforded the corresponding *S*-(glycosylthiomethyl) amino acids and peptides.

3.4
C-Glycosides

The existence of *C*-glycosidic linkages in naturally occurring glycoproteins was first discovered in 1994 in human RNase, in which a mannose residue is connected to C2 of the indole ring of a tryptophan [83]. A few years later, Manabe and Ito established the synthesis of a corresponding building block, which opened the way for the incorporation of ^{13}C or ^{15}N labels [84]. However, even before the discovery of the *C*-mannosyl-tryptophan linkage, carbon-linked glycosyl amino acids have attracted the interest of synthetic chemists because of the stability of this linkage. *C*-Glycosyl glycines for example were reported as early as in the late 1940s [85]. Yet, it took until 1992 to incorporate *C*-glycosyl amino acids into peptides. Kessler et al. synthesized *C*-glycosyl alanine derivatives via free radical addition to dehydroalanine derivatives and dehydroalanine-containing peptides [86]. Bertozzi et al. reported on a *C*-glycosyl analogue of Gal(β1–O)Ser (i.e., a *C*-glycosyl homoalanine) and its use for SPPS of an α-helical *C*-glycopeptide [87]. Since then, C-linked glycopeptides have been extensively studied by many groups as an approach to obtain stable glycopeptide mimetics and several review articles cover this subject [32, 35, 38]. Selected recent publications will be presented.

Nucleophilic attack at the anomeric carbon is a common pathway towards *C*-glycosides [88, 89]. Following this approach, Kihlberg et al. synthesized the C-linked isostere **62** of β-D-galactosyl hydroxynorvaline (Scheme 12) [90]. The two key steps in this synthesis were the addition of a homoallylic Grignard reagent to per-O-benzylated galactopyranolactone **57** with subsequent reduction of the resulting hemiacetal with triethylsilane and a Wittig reaction of phosphonium salt **59** employing Garner's aldehyde **60**. Building block **62** was then incorporated into a fragment from type II collagen by solid-phase synthesis. Helper T cell hybridomas obtained in a mouse model for rheumatoid arthritis responded to the C-linked glycopeptide when presented by class II MHC molecules although at a higher concentration compared to the natural O-glycosylated hydroxynorvaline.

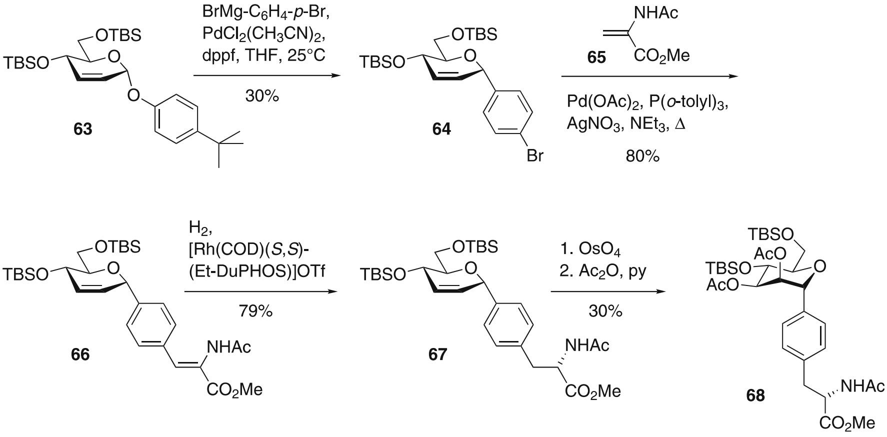


Scheme 12 Synthesis of the C-linked isostere **62** of β-D-galactosyl hydroxynorvaline from galactopyranolactone **57**

On the basis of these results, the same group succeeded in 2003 in the stereoselective synthesis of a *C*-glycoside analogue of β-D-galactosyl threonine in 14 steps [91]. Recently they reported the synthesis of a *C*-glycoside analogue of β-D-galactosyl hydroxylysine in 17 steps [92]. An Evans allylation established the configuration at the γ-carbon of hydroxylysine and an asymmetric hydrogenation using Burk's catalyst was applied for the α-amino acid moiety.

Schweizer and coworkers demonstrated that ulosonic acids, which are accessible from the corresponding sugar lactones by reaction with *tert*-butylacetate, can be converted into unnatural *C*-glycosyl amino acids via a one-pot intramolecular Ritter reaction [93]. A diaminobutyric acid-derived building block was used as serine replacement in the solid-phase synthesis of a modified opioid peptide.

The stereoselective synthesis of *C*-glycosyl analogues of phenylalanine was reported by Sinou and coworkers (Scheme 13) [94]. Condensation of hex-2-enopyranoside **63** with an aryl Grignard reagent [95] afforded unsaturated α-*C*-aryl glycoside **64**, which was subjected to a Heck coupling with acrylate **65**. The obtained enamide ester **66** was hydrogenated with a chiral DuPHOS-Rh^+ catalyst and subsequently dihydroxylated to give *C*-mannopyranosyl phenylalanine **68**.

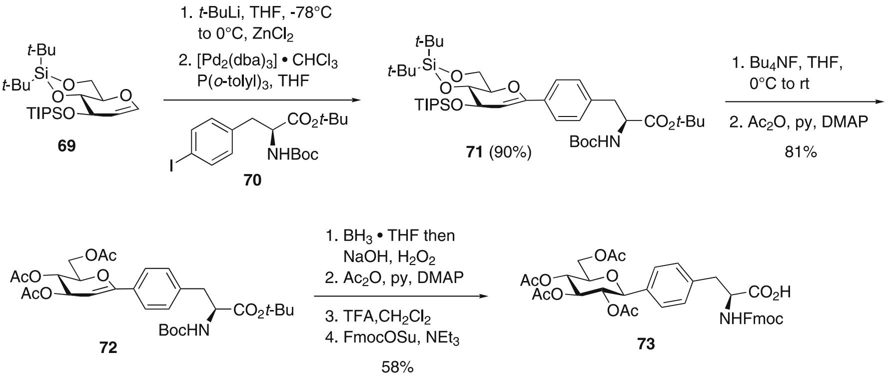


Scheme 13 Synthetic route to *C*-mannopyranosyl phenylalanine **68**

For the gram-scale preparation of the *C*-glucosyl phenylalanine derivative **73**, Augé and coworkers used the Negishi cross-coupling as key reaction (Scheme 14) [96]. High-yield coupling of *p*-iodo-L-phenylalanine **70** to silylated glycal **69** gave **71**. Further protecting group manipulations and stereoselective hydroboration of the double bond yielded C-glycosylated amino acid **73**, ready for use in SPPS.

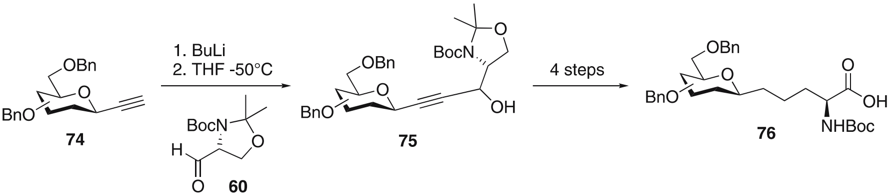
Gallagher et al. employed organozinc reagents carrying an aryl iodide function for the reaction with glycals [97]. Cross-coupling of the aryl iodide to an alanyl zinc reagent by Pd(0) catalysis led to *C*-glycosyl analogues of glycosyl tyrosines. A C-mannosylated building block was incorporated into a cyclic hexapeptide by solid-phase synthesis.

Through cross-coupling of C-linked glycosyl acetylenes with racemic N^{α}-acetyl-3- and -4-iodophenylalanine methyl esters by Pd(0) catalysis Meldal et al. received diastereomeric pairs of *C*-glycosylacetylene phenylalanine



Scheme 14 Application of the Negishi cross-coupling as key step for the synthesis of C-glycosylated phenylalanine **73**

building blocks [98]. Separation of the diastereoisomers was achieved by selective enzymatic hydrolysis of the N^{α}-acetyl group of the L-amino acid by acylase I. After Fmoc protection, the building blocks were used in the solid-phase synthesis of neoglycopeptide analogues of the mouse hemoglobin-derived decapeptide Hb(67–76). Dondoni et al. used this kind of sugar acetylene for the syntheses of ethylene isosteres **76** of *N*-glycosyl asparagines by coupling of metalated sugar acetylenes **74** with the Garner aldehyde **60** (Scheme 15) [99].



Scheme 15 Synthesis of ethylene isosteres **76** of *N*-glycosyl asparagines from metalated sugar acetylenes **74**

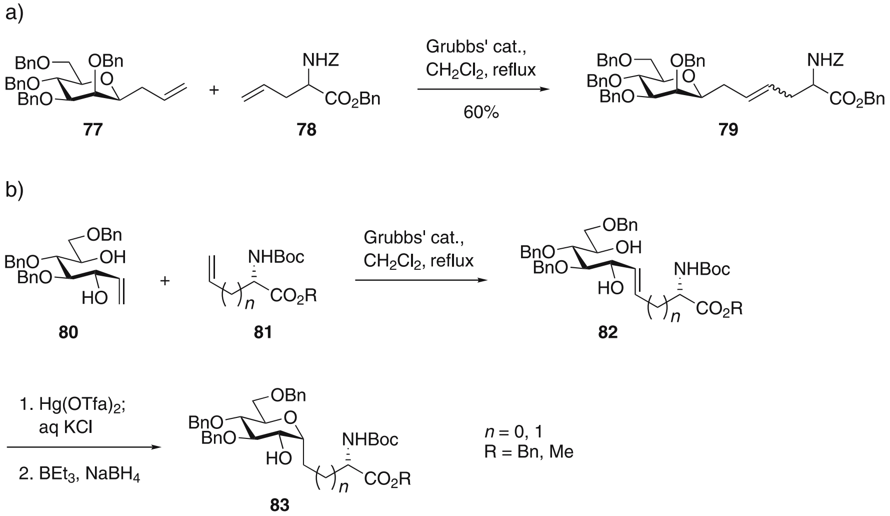
For the synthesis of *C*-glycosyl alanine derivatives, Wong and coworkers used an asymmetric Strecker reaction starting from *C*-glycosyl acetaldehyde derivatives [100]. The stereoselectivity of this reaction was strongly solvent-dependent and was reversed by changing from THF to CH_2Cl_2. Starting with the same type of C-glycosidically linked aldehyde, Arya and coworkers succeeded in the synthesis of carbon-linked analogues of α- and β-galactosyl-serine glycoconjugates using asymmetric enolate methodology [101]. Methylene-elongated serine and threonine *C*-glycosyl amino acid building blocks were prepared by Thiem et al. employing regio- and stereoselective aziridine ring opening methodology [102]. The needed *C*-glycosyl aldehyde intermediate was assembled by reaction of peracetylated *β*-D-galactopyranose with propargyl trimethyl silane in the presence of a Lewis acid, similar to a procedure developed by Bednarski et al. [103]. Wang and Linhardt reported the synthesis of serine-based neuraminic acid *C*-glycosides by using a neuraminic acid-based glycosyl donor in a samarium diiodide-mediated C-glycosidation reaction [104].

Despite early reports on the use of free radical reactions for the synthesis of *C*-glycosyl amino acids and peptides [86, 105], this approach is still rarely used in comparison to the above-mentioned nucleophilic approaches. Linker et al. reported a short three-step synthesis of 2-C-branched glyco-amino acids by addition of nitroacetate to various glycals in the presence of ceric(IV) ammonium nitrate and subsequent reduction of the obtained bicyclic isoxazoline *N*-oxides [106]. Caddick and coworkers used tetrafluorophenol-linked acrylate on solid phase for the synthesis of C1- and C6-linked glycopeptide mimetics [107]. The work represents the first *C*-glycoside synthesis on a solid phase using a free-radical methodology.

Radical allylation of acetobromomannose followed by ozonolysis gave a *C*-glycosyl aldehyde which was further elaborated into a *C*-mannosyl amino acid by Toone and coworkers [108]. This building block was incorporated into C-linked glycopeptides that acted as mono- and bivalent ligands for the legume lectin concanavalin A. Binding studies on resin beads and in solution were carried out to investigate the effect of carbohydrate presentation on the carbohydrate–lectin interaction. C-Linked glycopeptides have been used before by the group and evaluated as ligands for the Shiga-like toxin 1 binding subunit [109].

Dondoni et al. used *C*-glycosyl nitrile oxides and acetylenes for 1,3-dipolar cycloadditions to an alkyne or an azide, respectively, to obtain rigidified heterocycle-tethered *C*-glycosyl amino acids [110]. The isoxazole- or triazole-tethered *C*-glucosyl and *C*-galactosyl alanines are promising building blocks for new glycopeptide mimetics. At the same time, Rutjes and coworkers also reported the synthesis of triazole-linked *C*-glycosyl amino acids employing 1,3-dipolar cycloadditions (cf. Sect. 3.2) [61].

The cross-metathesis (CM) reaction, being an important method for the preparation of diverse neoglycoconjugates [111], was used by Roy and coworkers for the conversion of *β*-*C*-allyl mannoside **77** and (*R*/*S*)-allyl glycine **78** into C-mannosylated amino acid **79** using Grubbs' catalyst (Scheme 16a) [112]. Earlier, the group had already applied CM in the synthesis of C-glycosylated peptoids [52] which have also been reported by Kessler and coworkers [113]. In 2002 McGarvey et al. used this methodology for the reaction of *α*-*C*-allyl glycosides from *N*-acetylglucosamine and lactose with protected allyl glycine derivatives, leading to the corresponding unsaturated *C*-glycosyl amino acid building blocks [114]. In a convergent approach, the reaction was also applied for the modification of an allyl-



Scheme 16 Applications of cross-metathesis (CM) for the preparation of C-glycosylated amino acids **79** (**a**) and **83** (**b**)

glycine-containing tripeptide. Liu and Ben applied CM for the preparation of a series of C-linked antifreeze glycoprotein analogues [115]. Another type of C-linked antifreeze glycoprotein analogue was reported earlier by the same group [116, 117]. Nolen et al. used open-chain sugar precursor **80** for the preparation of C-glucosylated amino acids **83** with a linking carbon chain between the *C*-glycoside and the amino acid shorter than three carbons (Scheme 16b) [118]. Their CM/cyclization strategy provides a stereoselective entry to *C*-glucosyl alanines **83** ($n = 0$) and *C*-glucosyl analogues of glucosyl serines **83** ($n = 1$), without the seriously diminished yields previously observed using allyl or vinyl *C*-glycosides.

Michael and coworkers employed aminomethyl-*C*-glycosides in a photochemical approach towards unnatural *C*-glycosyl asparagine derivatives [119, 120]. The products could be obtained by phototransamidation of the aminomethyl-*C*-glycosides with photoreactive asparagine derivatives under neutral conditions. Conformationally constrained *C*-glycosyl α- and β-amino acids via Diels–Alder reaction have been reported by Jayakanthan and Vankar [121].

4

Convergent Approaches to Glycopeptide and Glycoprotein Mimetics

Compared to linear approaches, the convergent synthesis of glycopeptide and glycoprotein mimetics offers greater flexibility with respect to the sugars attached to a peptide. Thus, the preparation of several well-defined glycoforms of the same peptide/protein becomes possible. The approach is also of interest for combinatorial glycopeptide synthesis. If chemoselective ligation reactions are employed for the sugar attachment, it is possible to modify unprotected peptides and even whole proteins.

4.1
Formation of C = N Bonds

In 1996 Mutter et al. reported oxime formation as a means to chemoselectively couple reducing oligosaccharides to aminooxy-functionalized peptides [122]. Since then, oxime formation and related C = N bond formations have been widely used to link carbohydrates and peptides with either the carbohydrate or the peptide acting as the carbonyl component. The possibility of using water as solvent and the high chemoselectivity of the reaction, making the use of protecting groups superfluous, are essential advantages of this methodology, which has also been applied in analytical chemistry [123]. However, one has to keep in mind that the reaction is potentially reversible.

Mimetics of *O*-glycopeptides with native sugar–peptide linkages were obtained by Bertozzi and coworkers [124]. A synthetic *O*-glycopeptide with a single *N*-acetylgalactosamine residue was oxidized by galactose oxidase and the obtained GalNAc C-6 aldehyde was reacted with several aminooxy-substituted carbohydrates. Later Bertozzi et al. applied aminooxylated carbohydrates for a chemoselective ligation with (2*S*)-aminolevulinic acid-containing peptides prepared by SPPS to obtain glycopeptide mimetic **84** (Fig. 3) [125]. Similarly, hydrazide- and thiosemicarbazide-modified sugars were used to prepare the corresponding hydrazones and thiosemicarbazones, respectively. Wong and Schultz and their coworkers incorporated *p*-acetyl-L-phenylalanine into the Z domain of staphylococcal protein A by a genetic approach in response to the amber nonsense codon [126] according to a method established earlier [127, 128]. Aminooxy-functionalized mono- and oligosaccharides were ligated and in part further elongated by glycosyltransferase-catalyzed glycosylations. The relatively low mass of the obtained glycoprotein mimetic **85** (approx. 8 kD) facilitated high-accuracy

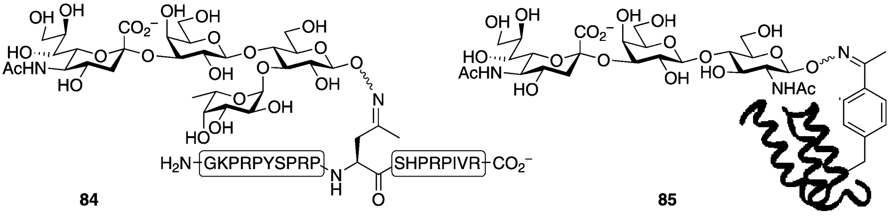
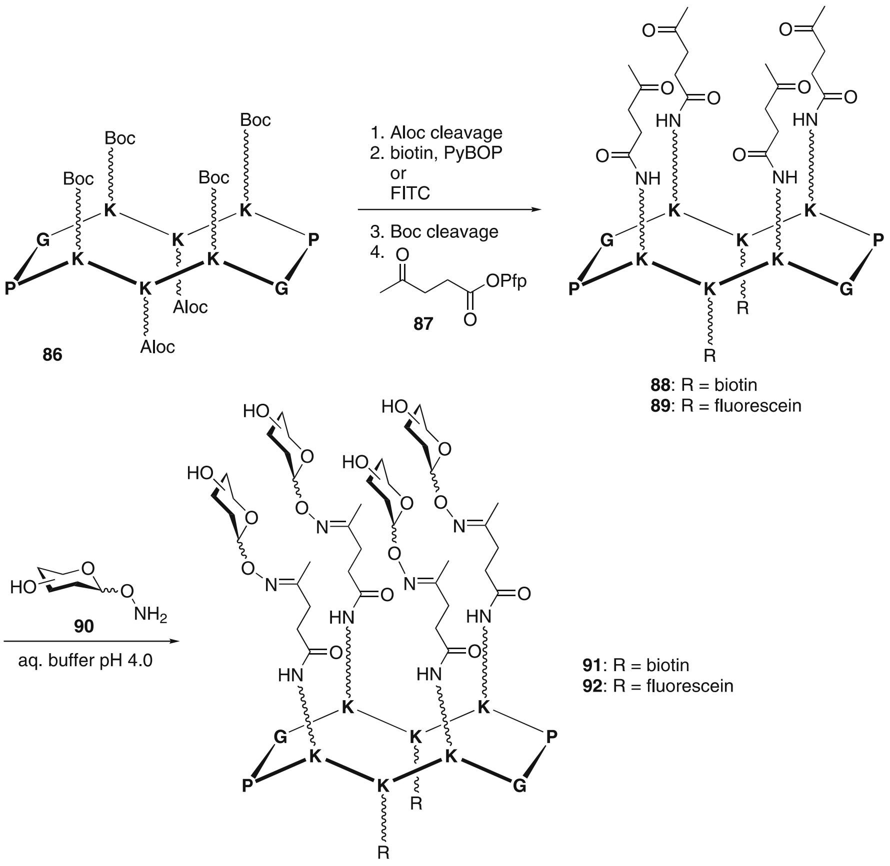


Fig. 3 Glycopeptide and glycoprotein mimetics obtained by oxime ligation

mass spectrometry and thus a good control of the selective attachment of the unprotected aminooxy-modified saccharides.

In a series of papers, the Dumy group explored cyclic peptides with regioselectively addressable side-chain amino groups for the tetravalent presentation of mono- and disaccharides [129–132]. Recently, they reported the synthesis of a cyclic peptide template with two separated addressable domains (Scheme 17) [131]. Four Boc-protected lysine residues were used for the tetravalent attachment of carbohydrates. Two lysines orthogonally protected by Aloc groups served for attachment of molecular probes. Thus, **86** was first treated with a Pd(0) catalyst and phenylsilane to remove the Aloc groups, followed by attachment of a biotin and a fluorescein residue, respectively. TFA treatment and coupling of pentafluorophenyl (Pfp) ester **87** under basic conditions gave tetra keto-substituted peptides **88** and **89**, re-

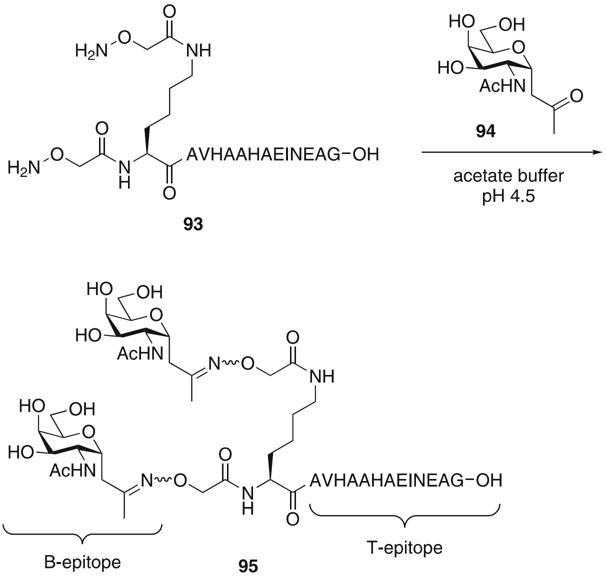


Scheme 17 Application of the cyclic peptide template **86** with two separated addressable domains for the preparation of labeled neoglycopeptides **91** and **92**

spectively. Oxime formation was achieved with several α- and β-configurated aminooxylated monosaccharides **90** under aqueous conditions. The obtained cyclic neoglycopeptides **91** and **92** were obtained in yields of 40–60% for the oxime formation step after RP-HPLC purification. Similar compounds have been used in lectin-binding studies [129, 132].

A chemoselective approach toward the sensitive detection of *O*-GlcNAc posttranslational modifications was developed by Hsieh–Wilson and coworkers [133]. Employing a modified UDP-Gal analogue with an acetylmethyl group replacing the 2-OH group and a genetically engineered galactosyltransferase it was possible to label *O*-GlcNAc-modified proteins with a ketone tag. Oxime ligation with an aminooxy-substituted biotin derivative permitted the identification of several low-abundance *O*-GlcNAc-modified proteins from brain [134]. Ketone isosteres of GlcNAc and GalNAc have also been incorporated into the glycan chains of cell-surface glycoproteins by metabolic oligosaccharide engineering and identified by hydrazone ligation with biotin hydrazide [135].

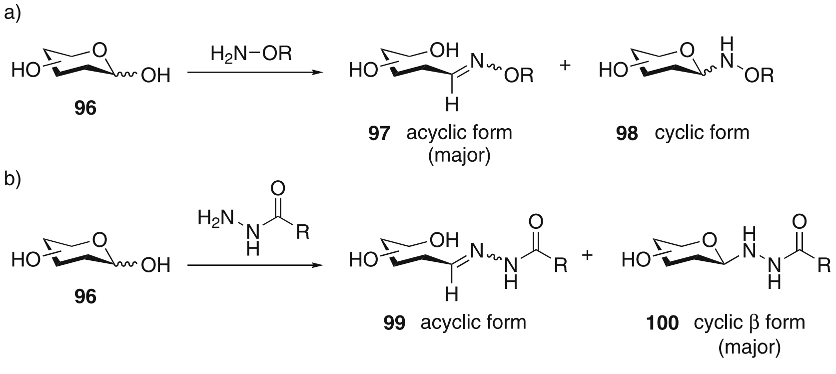
Nicotra and coworkers applied oxime ligation to the covalent attachment of the *C*-glycoside analogue **94** of the Tn epitope to T cell epitope peptides, such as **93** modified with two aminooxy acetic acid residues (Scheme 18) [136–139]. The obtained neoglycopeptide **95**, bearing two B-epitopes and one T-epitope, could be presented to and recognized by the



Scheme 18 Ligation of *C*-glycosyl ketone **94** to aminooxy-substituted T cell epitope peptide **93**

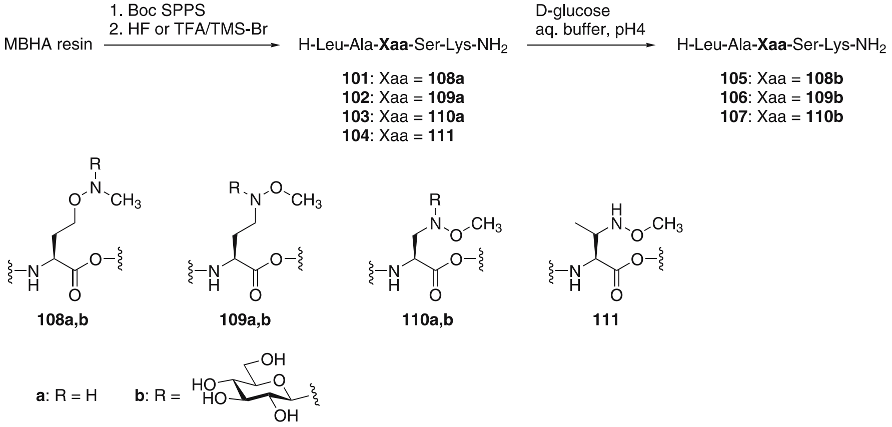
T cell receptor with higher efficiency than related neoglycopeptides having only one B-epitope.

A number of papers describe the use of aminooxy-modified peptides for the preparation of glycopeptide mimetics by chemoselective ligation to unmodified reducing mono- and oligosaccharides. The advantage of this approach is that oligosaccharides isolated from natural sources (e.g., from chemical or enzymatic degradation of glycoconjugates) can directly be used for conjugation to a peptide. Imperiali et al. used this approach to obtain several glycopeptide mimetics containing single GlcNAc residues as inhibitors of oligosaccharyl transferase [140, 141]. Besides aminooxy-modified peptides, peptides with aspartic acid hydrazide were also employed. A problem associated with oxime and acyl hydrazone formation is the fact that the ligation products can exist in two forms, cyclic and acyclic, and each of them as several diastereoisomers (Scheme 19). It has been reported that the acyclic form dominates in the case of ligations with alkoxyamines (Scheme 19a) [142, 143] whereas the cyclic form is preferred with hydrazides (Scheme 19b) [143–147]. The product ratio, however, is dependent on the carbohydrate. In order to fix the sugar in a defined acyclic form, subsequent reduction with, e.g., sodium cyanoborohydride to give the corresponding hydroxylamine and hydrazide, respectively, may be carried out. This process, however, generates new nucleophilic centers (which might have additional utility [148]) and is no longer a sheer ligation reaction.



Scheme 19 Oxime (**a**) and acyl hydrazone (**b**) formation leads to mixtures of acyclic and cyclic ligation products

As an alternative strategy, the use of *N*,*O*-disubstituted hydroxylamines has been described [149]. In contrast to oxime formation, the cyclic nature of the saccharide reducing unit is completely preserved. The reaction also proceeds with high stereoselectivity to give under thermodynamic control the *β* or *α* pyranose forms depending on the sugar. Carrasco et al. followed this approach and prepared differently protected amino acids with a methyl aminooxy group and a methoxyamino group, respectively, in their side chain [150–152]. SPPS following the Boc strategy gave unprotected peptides **101–104** that were chemoselectively glycosylated to generate neoglycopeptides **105–107** (Scheme 20). The ligation yields were in the range of 75% and more. Glycosylation with the sterically hindered threonine derivate **104**, however, failed. The same strategy was successfully applied by Rocchi et al. for the synthesis of opioid deltorphin I analogues [153].

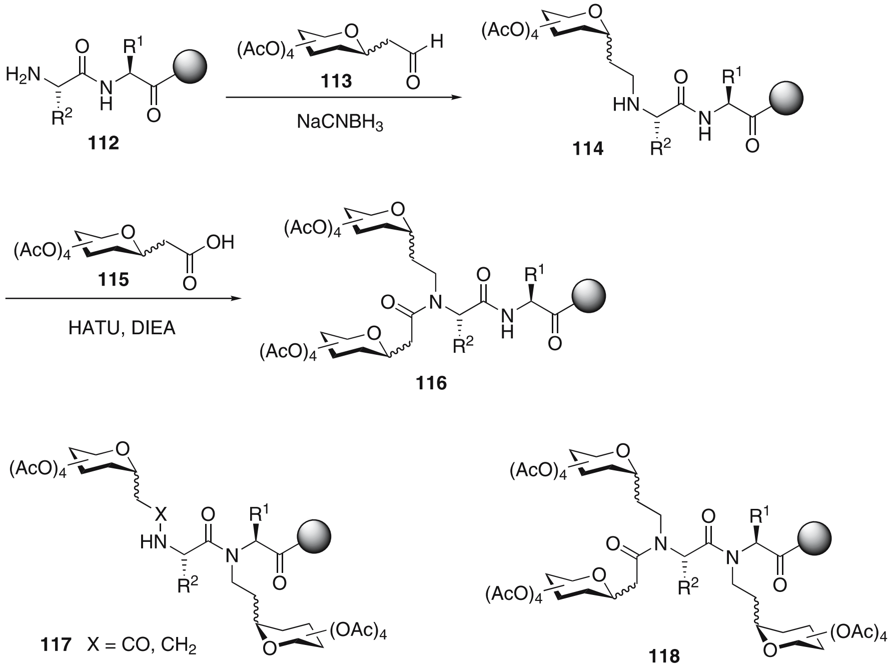


Scheme 20 Peptides containing *N*,*O*-disubstituted hydroxylamines and their use in chemoselective ligation reactions with unprotected glucose

Aminooxy- and methylaminooxy-substituted homoalanines were incorporated into T4 phage lysozyme and streptavidin by Nishimura and coworkers [154] employing the frameshift suppression strategy on the basis of the four-base codon/anticodon system [155]. Chemoselective coupling between the methylaminooxy-containing mutant streptavidin and unmodified mono-, di-, and trisaccharides allowed for the combinatorial synthesis of several glycoprotein mimetics.

4.2
Formation of C – N Bonds

Arya and coworkers developed automated, multistep, solid-phase strategies for the parallel synthesis of libraries of carbon-linked glycopeptide mimetics, such as **116–118** (Scheme 21) [156–158]. Starting from solid-phase bound peptide **112**, a *C*-glycosyl aldehyde **113** was attached by reductive amination followed by amide bond formation using *C*-glycosyl carboxylic acid derivatives **115** to result in neoglycopeptides **116**. Using this approach, a variety of α- or β-carbon linked carbohydrates can be incorporated either at the N-terminus or at internal nitrogens of short peptides in a highly flex-

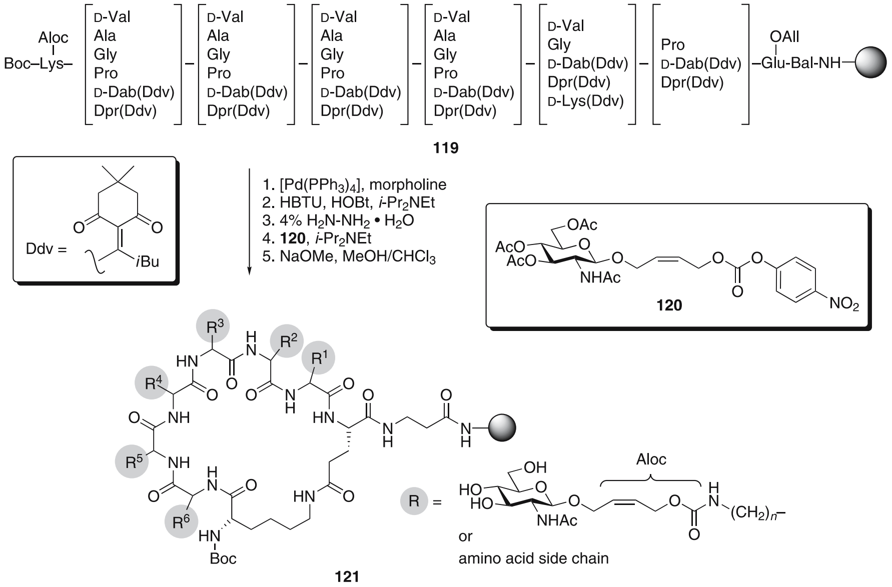


Scheme 21 Strategy for the combinatorial synthesis of carbon-linked glycopeptide mimetics **116**. Structures **117** and **118** may be obtained analogously

ible manner. The glycopeptide mimetics were used as probes for studying carbohydrate–protein interactions as well as protein folding and trafficking.

Wang and coworkers established a general method for rapid generation and screening of a combinatorial *C*-glycopeptide library in the one-bead one-compound format [159]. An α-*C*-galactosyl carboxylic acid derivative **115** was activated as a pentafluorophenyl ester and coupled to the N-terminus of the peptides under standard solid-phase conditions. Partial termination during peptide synthesis (ladder synthesis [160, 161]) allowed for mass spectrometric identification of lead structures bound to single beads. Several glycopeptide sequences mimicking the Gal(α-1,3)Gal epitope in binding to human natural anti-Gal antibodies were identified.

Wittmann and coworkers reported the convergent solid-phase synthesis of multivalent cyclic neoglycopeptides to study the influence of carbohydrate presentation on the binding affinity to lectins (Scheme 22) [162, 163]. Starting from library **119** of linear nonapeptides, side-chain cyclization was achieved by Pd(0)-catalyzed removal of the Aloc and allyl groups followed by HBTU/HOBt activation. Hydrazine treatment liberated Ddv-protected (Ddv = 1-(4,4-dimethyl-2,6-dioxocyclohexylidene)isovaleryl) side-chain amino groups of the cyclic peptides. For the attachment of carbohydrates a new linker based on the Aloc protecting group has been developed.



Scheme 22 Preparation of library **121** of 19 440 conformationally restricted glycoclusters for the spatial screening of multivalent lectin ligands

Thus, treatment with activated carbonate **120** led to urethane formation in virtually quantitative yields. The obtained combinatorial one-bead one-compound library **121** of 19440 conformationally restricted, spatially diverse glycoclusters has been screened for binding to wheat germ agglutinin (WGA) by a solid-phase enzyme-immuno lectin-binding assay. Several tetra-, penta-, and hexavalent high-affinity ligands were identified by single-bead Edman degradation after palladium-catalyzed detachment of the sugars according to a strategy described earlier [164]. It was concluded that the spatial presentation of the sugars on the cyclopeptide scaffolds is responsible for the high affinity to WGA.

Cyclic peptides as scaffolds for the attachment of carbohydrates have also been reported by Nishimura et al. [165]. Aminohexyl-spacer modified disaccharides were coupled to glutamine side chains by a reaction catalyzed by transglutaminase according to a method described by Klaffke et al. [63] and further elongated by an α-2,3-(*N*)-sialyltransferase-catalyzed glycosylation. The obtained trivalent glycopeptide mimetics displayed different binding affinity to influenza hemagglutinin depending on the peptide conformation. For the best ligand, a tridentate glycopeptide–protein complex was suggested, based on computer calculations.

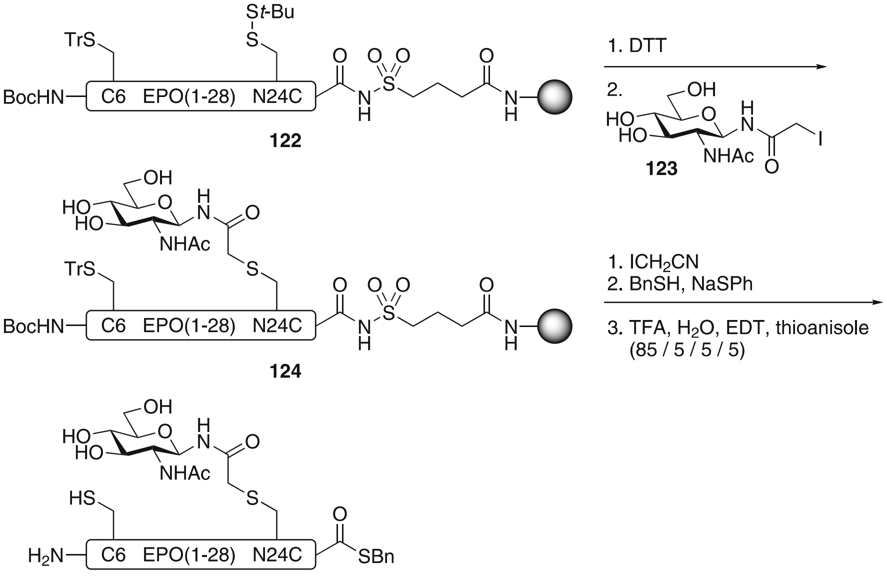
Finn and coworkers described the use of whole icosahedral cowpea mosaic virus particles as multivalent carbohydrate display platforms [166]. When treated with α-D-mannopyranosyl-phenylisothiocyanate, up to 120 mannose residues were attached to lysine residues on the virus surface by thiourea formation. Aggregation of labeled virus with the plant lectin concanavalin A was studied by UV/VIS spectroscopy, transmission electron microscopy, and fluorescence resonance energy transfer. The nature of this multivalent interaction was shown to depend upon the density and position of functionalization of the virus. Modification of adenovirus (AV) lysine residues by 2-imino-2-methoxyethyl-1-thioglycosides was used by Davis and coworkers to modulate the ability of AV to function as gene delivery vector [167]. It was possible to "switch off" AV transfection by the normal receptor pathway and "switch on" a specific sugar-mediated pathway. The same imidate chemistry which had originally been developed by Lee and coworkers for the preparation of neoglycoproteins [168] was applied by the Davis group to generate glycosylated forms of rhamnosidase for use in a bipartite drug-delivery system called lectin-directed enzyme-activated prodrug therapy (LEAPT) [169].

4.3
Sulfur as Nucleophile

4.3.1
S-Alkylation

The first report on the use of the reaction of glycosyl iodoacetamides with cysteine residues for the synthesis of N-linked glycoprotein mimetics came from the Flitsch group [170] and was later followed by a report on the modification of unprotected synthetic peptides containing a single cysteine residue in solution [171]. Dahmén and Kihlberg et al. coupled 2-bromoethyl β-glycosides of galabiose [Gal(α1–4)Gal], globotriose [Gal(α1–4)Gal(β1–4)Glc], and 3′-sialyllactose [Neu5Ac(α2–3)Gal(β1–4)Glc] to cysteine and homocysteine residues in T cell-stimulating peptides by cesium carbonate promoted alkylation in yields of 52–95% [172]. Neither β-elimination nor epimerization of the neoglycopeptide stereocenters occurred. More recently, the iodoacetamide approach was applied by Flitsch and coworkers for the semi-synthesis of homogeneous glycoforms of human erythropoietin (EPO) [173]. Several mutants in which N-glycosylation site asparagines had been changed to cysteines were reacted with GlcNAc iodoacetamides in yields of about 50%. This was possible even in the presence of additional cysteine residues involved in disulfide bridges. Purification was readily carried out by lectin affinity chromatography.

Glycosyl iodoacetamides may also be applied to solid-phase glycopeptide synthesis. This allows for selective modification of certain cysteine residues in peptides with several cysteines employing orthogonal protection. Macmillan and Flitsch and coworkers applied the 4-sulfamylbutyryl-AM (safety-catch) resin [174] to obtain neoglycopeptide thioesters derived from the first 28 amino acids of the glycoprotein hormone EPO (Scheme 23) [175]. Peptide **122**

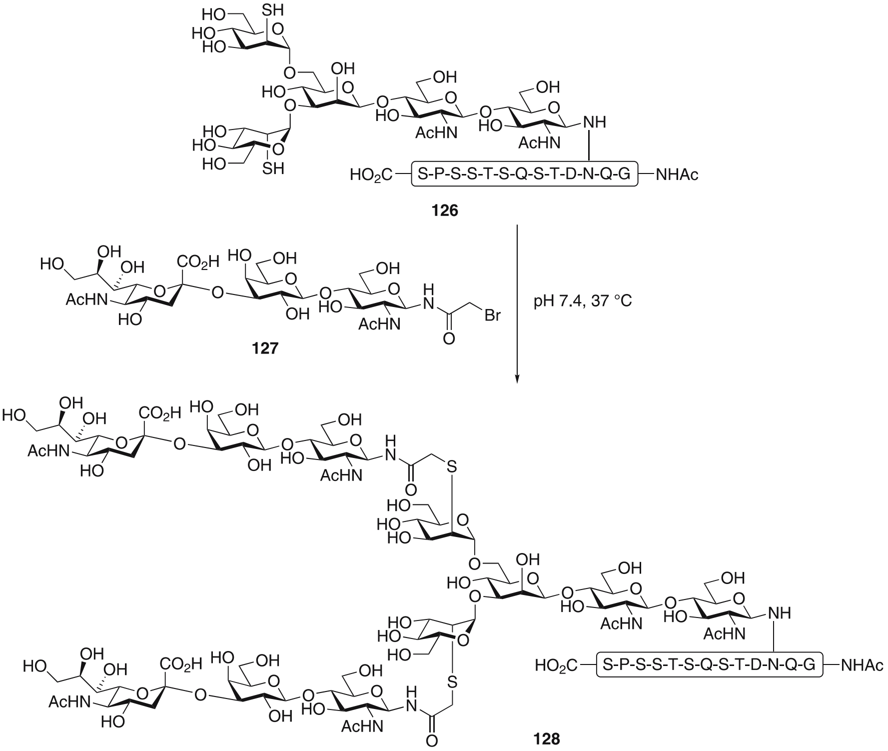


125

Scheme 23 Application of glycosyl iodoacetamide **123** to the solid phase synthesis of glycopeptide mimetic **125**

was prepared by Fmoc SPPS and selectively deprotected by reduction with dithiothreitol (DTT). The free thiol was then alkylated with iodoacetamide **123** to give glycopeptide mimetic **124**. The glycopeptide was released from the resin with benzyl mercaptan in the presence of catalytic NaSPh after activation with iodoacetonitrile. Deprotection with concentrated TFA gave glycopeptide thioester **125**. Later it was shown by Macmillan and Arham [176] that glycopeptide thioesters obtained by this approach are suitable for the synthesis of glycoproteins by native chemical ligation [177]. In the same year, a complex-type disialyl-undecasaccharide was appended to a cysteine residue of an undecapeptide using a bromoacetamidyl oligosaccharide [178]. It was shown that the obtained glycopeptide analogue was resistant to enzymatic digestion by peptide:N-glycosidase F (PNGase-F).

Bertozzi and coworkers adapted Flitsch's haloacetamide method to convergently ligate bromoacetamides of Gal and GlcNAc to a C-3 thiol of a GalNAc(*α*1-O)Thr-containing peptide with natural carbohydrate–peptide linkage [179]. The obtained thioether-linked structures can be regarded as mimetics of "core 1" and "core 3" structures of mucin-type glycoproteins. Later the work was extended to the synthesis of thioether-linked analogues of the mucin-related 2,3-sialyl-TF and MECA-79 antigens [180] as well as to the synthesis of N-linked glycopeptide mimetics that replace the glycosidic linkages extending from the core pentasaccharide with thioethers

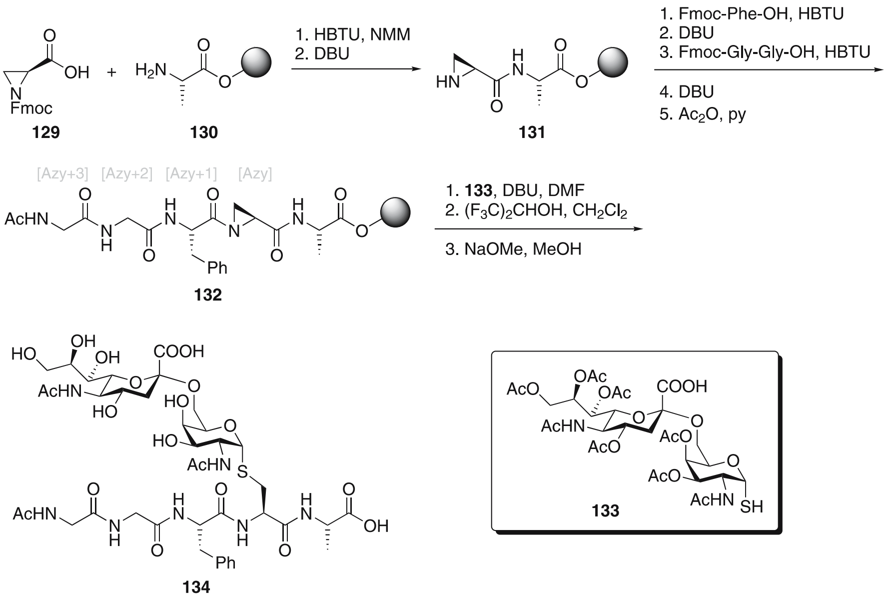


Scheme 24 Application of the haloacetamide ligation for the convergent assembly of neoglycopeptide **128** containing an undecasaccharide mimetic

(Scheme 24) [181]. A protected pentasaccharide-Asn analogue containing two thiol residues was incorporated into CD52 (a glycopeptide expressed on human lymphocytes and sperm cells) by Fmoc SPPS to give, after deprotection, **126**. Undecasaccharide mimetic **128** was then readily generated by alkylation of **126** with bromoacetamido trisaccharide **127**.

The application of 1-thiosugars for the preparation of S-glycosylated amino acid building blocks has already been discussed in Sect. 3.3. The approach, however, can also be carried out in a convergent manner. To this end, Zhu and Schmidt used O-acetyl protected 1-thio derivatives of GalNAc, GlcNAc, and lactose and attached them to β-bromoalanine- or γ-bromohomoalanine-containing di- to hexapeptides to give S-linked neoglycopeptides [75]. The reaction selectively gave α- and β-*S*-glycosides and even worked well with O-unprotected 1-thiosugars. Glycosylthiomethyl bromides and azides have also been applied in a convergent approach [82].

Aziridine-2-carboxylic acid (Azy)-containing peptides represent alternative electrophilic peptides for conjugation with thiols. Van der Donk and Gin and coworker prepared aziridine-containing peptide **132** starting from



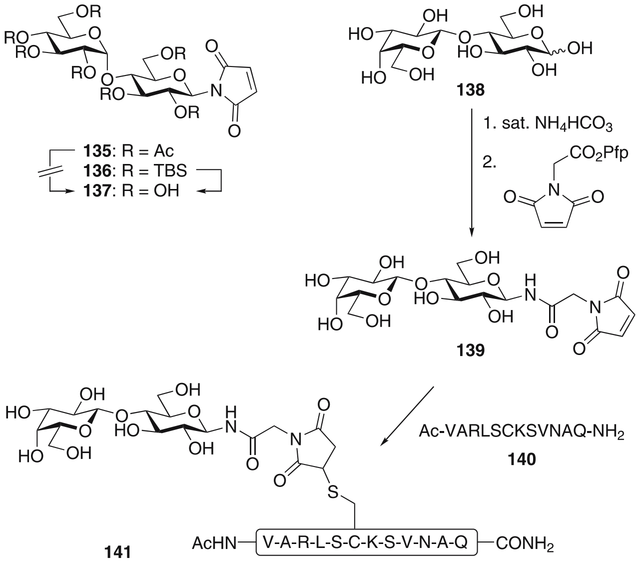
Scheme 25 Conjugation of 1-thio-disaccharide **133** to solid phase-bound Azy-containing peptide **132** leads to *S*-glycopeptide **134**

Fmoc-Azy-OH **129** and solid phase bound alanine **130** (Scheme 25) [182]. Attempts to extend **131** by sequential coupling of single Fmoc-amino acids were unsuccessful, presumably due to intramolecular aziridine N-deacylation by the liberated amine upon Fmoc removal at the [Azy + 2] position (numbering according to the direction of the synthesis). This side reaction could be avoided by employing a dipeptide building block for coupling of the residues in the positions [Azy + 2] and [Azy + 3] leading to **132**. Addition of 1-thio-disaccharide **133** and deprotection gave *S*-glycopeptide **134** containing the thio analogue of the sialyl Tn (STn) antigen. Similarly, other 1-thiosugars were conjugated. In all cases, base-promoted aziridine ring-opening occurred with high regioselectivity delivering only small amounts of the corresponding β^2-amino acid derivative.

4.3.2
Conjugate Addition to Michael Acceptors

The conjugate addition of thiols to maleimido moieties [183] has been widely used in bioconjugate chemistry [184]. Hansen and coworkers were the first to apply this reaction to the preparation of neoglycopeptides [185]. A reducing disaccharide was reacted with *S*-trityl-2-mercaptoethylamine and stabilized in the cyclic form by N-acetylation. After detritylation, the thiol-functionalized carbohydrate was coupled to maleimido-functionalized peptides to result in neoglycopeptides with a flexible linker between carbohydrate and peptide moieties.

Shin et al. synthesized maleimido-modified sugars for the chemoselective reaction with cysteine-containing peptides and proteins (Scheme 26) [186, 187]. Since deacetylation of **135** and similar sugars under mild basic conditions was not possible neither before nor after thiol ligation due to imide opening, TBS-protected sugars, such as **136**, were employed giving access to **137**. Alternatively, unprotected maleimido-modified sugars were obtained by a two-step procedure in 54–95% as shown for the conversion of **138** into **139**. These sugars were successfully attached to glutathione (γ-Glu-Cys-Gly), the synthetic Fas peptide **140**, and to bovine serum albumin possessing a single reduced cysteine at position 58. Later, Shin et al. used thiol-maleimide ligation also for the fabrication of carbohydrate microarrays [188, 189].



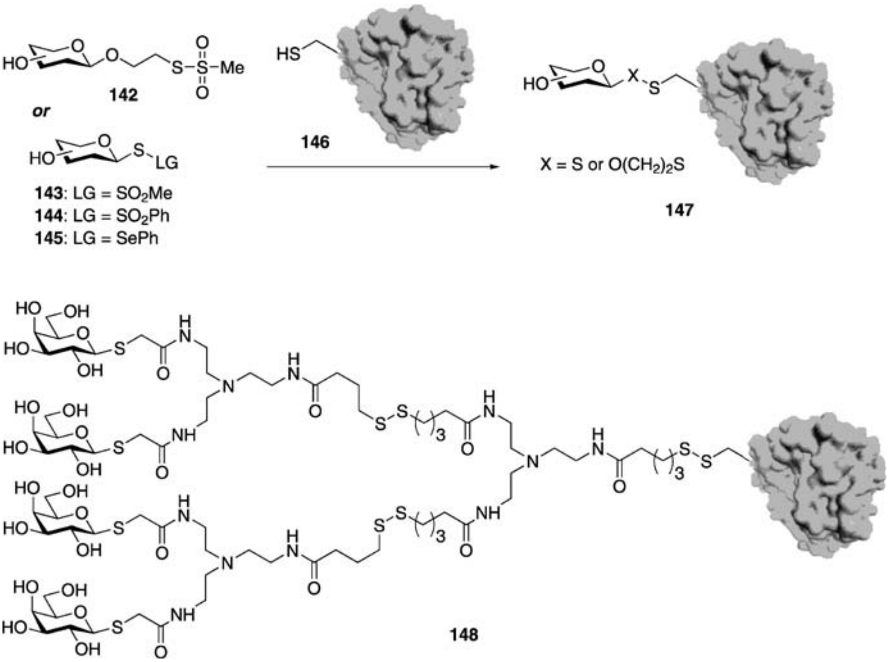
Scheme 26 Application of maleimido-modified sugars for the chemoselective ligation with cysteine-containing peptides

Van der Donk and coworkers used dehydroalanine-containing peptides for the chemoselective conjugate addition of glycosyl thiolates [190, 191]. As the dehydroalanine precursor, *Se*-phenylselenocysteine was introduced via SPPS. Oxidative elimination of this precursor was effected by sodium periodate addition and provided the dehydropeptide. Subsequent ligation with different α-glycosyl thiolates resulted in thio isosteres of the tumor-associated carbohydrate antigens Tn, T, STn, and 2,6-ST in high yields. The process proceeded with complete retention of the α-anomeric configuration. No stereoselectivity, however, was observed at the newly formed cysteine stereocenter. The incorporation of selenocysteine into peptides is not only of interest as a precursor for chemoselective modification of peptides but also for the preparation of larger peptides and proteins by selenocysteine-mediated native and expressed protein ligation [192].

4.3.3
Disulfide Bond Formation

The modification of cysteine residues within proteins combined with site-directed mutagenesis offers a strategy for site-selective incorporation of glycans into proteins. Beside the alkylation reactions discussed in Sects. 4.3.1 and 4.3.2, disulfide bond formation has been explored by several groups towards this goal. Boons et al. used 5-nitropyridine-2-sulfenyl activated thioglycosides for the preparation of disulfide-linked BSA–GlcNAc conjugates [193]. More recently, they reported the direct coupling of 1-thiosugars (thioaldoses) with the thiol group of free cysteine-containing peptides and proteins due to difficulties encountered in synthesizing nitropyridinesulfenyl thioglycosides of oligosaccharides [194]. A number of thioaldoses derived from several oligosaccharides including the core pentasaccharide of *N*-glycoproteins were synthesized. Incubation of 1–2 equivalents of 1-thio-chitobiose with cysteine-containing glutathione and BSA resulted in the site-specific formation of the disulfide-linked neoglycoconjugates together with the homodisulfides of glutathione and BSA in statistical amounts. The same technology was applied to the "glycosylation" of an aglycosylated human IgG1-Fc antibody fragment, engineered to have Cys-297 in place of glycan-linked Asn [195]. Mass spectrometry analysis verified site-selective and complete saccharide incorporation without affecting other Fc protein disulfides. The neoglycoproteins displayed enhanced inhibition of superoxide production by γ-interferon-stimulated U937 cells relative to the aglycosylated N297C protein, where increased glycan size correlated positively with increased inhibition.

Davis et al. reported a whole series of carbohydrate derivatives suitable for chemoselective, cysteine-specific synthesis of neoglycoproteins (Scheme 27) [196–202]. Methanthiosulfonates bound to carbohydrates either via a linker (**142**) or directly (**143**) (glyco-MTS derivatives) react specifically with thiols of cysteine-containing proteins **146** and allow the specific synthesis of disulfide-linked glycoproteins **147** in high yields [196–198]. This was demonstrated by site-selective conjugation of glyco-MTS derivatives to cysteine mutants of the serine protease subtilisin *Bacillus lentus* (SBL), which does not naturally contain cysteine. A library of 48 glycosylated forms of SBL as single, pure species has been prepared and detailed glycan structure–hydrolytic activity relationships have been determined. The efficiency of the approach of disulfide bond formation using glyco-MTS derivatives was later improved by using glycosyl phenylthiosulfonates (glyco-PTS) **144** [199]. In most cases these reagents gave superior yields of neoglycoproteins **147**.



Scheme 27 Glyco-MTS (**142**, **143**), glyco-PTS (**144**), and glycosyl selenenylsulfide (**145**) derivatives developed for the synthesis of neoglycoproteins **147** and glycodendriproteins, such as **148**

The glyco-MTS strategy has also been used for the attachment of different bi- and tetravalent glycodendrimers to the Ser156Cys mutant of SBL to give so-called glycodendriproteins, such as **148** [200, 201]. These galactose-presenting synthetic glycoprotein mimetics turned out to be nanomolar inhibitors of the co-aggregation of the pathogen *Actinomyces naeslundii* with co-pathogen *Streptococcus oralis*. It was shown that both the multivalent presentation of galactose and the protein degrading activity of this SBL neoglycoconjugate are essential for optimal inhibition.

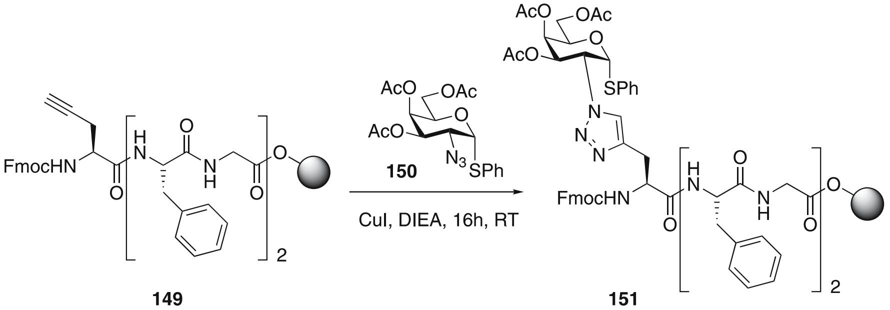
Recently, Davis and coworkers described selenenylsulfide-mediated glycoconjugation as a new strategy for the synthesis of disulfide-linked glycoproteins **147** (X = S) [202]. Two alternative routes were studied. Either a 1-thio mono- or oligosaccharide was converted into the glycosyl selenenylsulfide **145** and then reacted with a cysteine-containing protein **146** or the protein was first converted into the corresponding (phenylselenenyl)sulfide by treatment with phenylselenenyl bromide and then reacted with 1-thio mono- or oligosaccharides. With this approach it was possible to prepare neoglycoproteins from several model proteins modified with mono- and oligosaccharides of up to seven saccharide units.

The formation of disulfides under conditions allowing thiol/disulfide exchange can be used for the preparation of dynamic combinatorial libraries (DCLs) [203–206]. First applications of DCLs in carbohydrate chemistry were reported by Sasaki et al. [207] and Lehn et al. [208, 209]. Sando and Aoyama and coworker prepared DCLs of glycopeptide mimetics based on disulfide-linked sugar–peptide coupling [210]. A mixture of 1-thiogalactose and two peptides containing two cysteine residues each afforded in the presence of DTT a disulfide library of at least ten sugar–peptide conjugates and cyclic peptides. Preliminary surface plasmon resonance experiments showed that the library contained active divalent galactoside species capable of cross-linking soluble peanut lectin (PNL) to PNL immobilized on the sensor-chip surface.

A similar approach was applied by Davis et al. [211]. They used their selenenylsulfide-mediated glycoconjugation approach for the preparation of several disulfide-linked glycopeptide mimetics containing glucose and GlcNAc and observed the composition of the library under conditions allowing thiol/disulfide exchange. Upon addition of wheat germ agglutinin, a GlcNAc-binding plant lectin, an enrichment of GlcNAc-containing glycopeptides was observed.

4.4
Azide-Alkyne [3 + 2] Cycloaddition

The copper(I)-catalyzed azide-alkyne cycloaddition [58, 59] represents a mild method for the regioselective preparation of 1,4-disubstituted 1,2,3-triazoles even in a biological environment and is now widely used in bioconjugate chemistry [60]. In their seminal paper, Meldal et al. applied this reaction to the synthesis of unnaturally linked glycopeptides [58]. 2-Azido-thiogalactoside **150** was reacted in the presence of two equivalents of copper(I) iodide with solid-phase bound peptides substituted with a terminal alkyne group either at the peptide backbone or at the amino-acid side chain, as shown for propargylglycine derivative **149** (Scheme 28). The reactions gen-

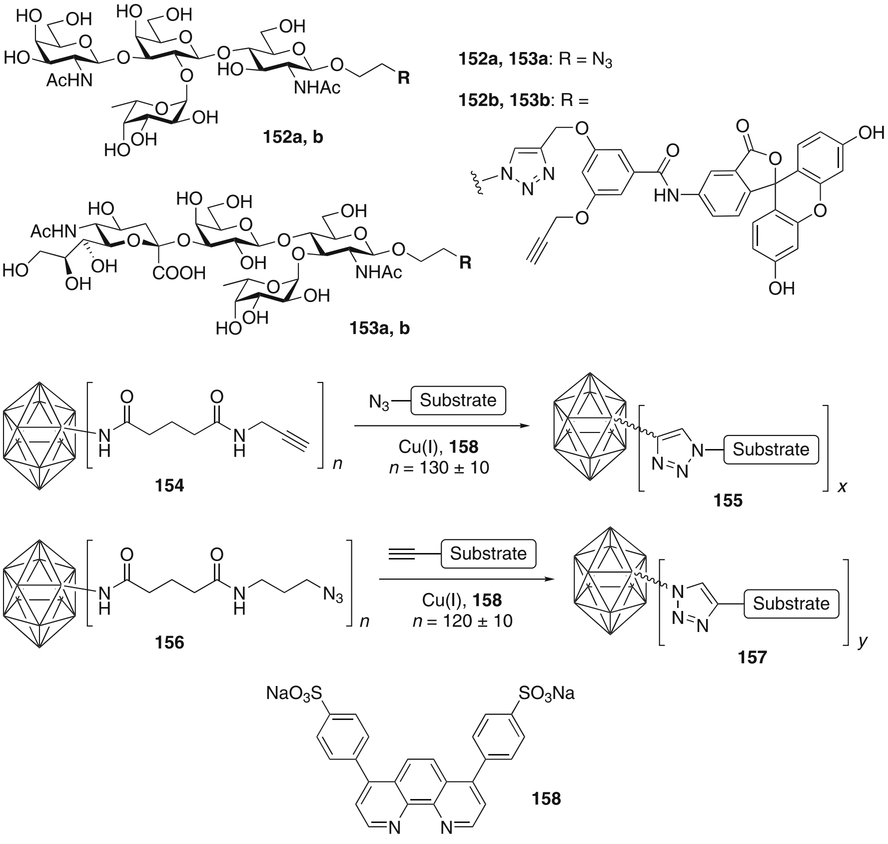


Scheme 28 Copper(I)-catalyzed azide-alkyne cycloaddition for the solid-phase synthesis of triazole-linked glycopeptide mimetics

erally worked with > 95% conversion and > 95% purity (78% purity in the case of **151**).

Lin and Walsh applied the copper(I)-catalyzed azide-alkyne cycloaddition for the attachment of 21 different azido-functionalized monosaccharides to 13 derivatives of the cyclic decapeptide tyrocidine with one to three propargylglycine residues incorporated at positions 3–8 [212]. Head-to-tail cyclization of the peptides was accomplished using a thioesterase domain from tyrocidine synthetase. Antibacterial and hemolytic assays showed that the two best glycopeptide mimetics had a sixfold better therapeutic index than the natural tyrocidine.

It has been shown that alkyne- and azido-modified amino acids, such as 2-amino-5-hexynoic acid [213], azidohomoalanine [214, 215], and with less efficiency also azidoalanine, azidonorvaline, and azidonorleucine [216], act as methionine surrogates that are activated by the methionyl-tRNA syn-



Scheme 29 Application of the copper(I)-catalyzed azide-alkyne cycloaddition for the preparation of glycoviruses

thetase of *E. coli* and replace methionine in proteins expressed in methionine-depleted bacterial cultures. This, together with other methods for the incorporation of non-canonical amino acids into proteins [41, 217], offers the possibility to use azide-alkyne cycloaddition (and also Staudinger ligation [218–220]) not only for protein labeling within cells or on cell surfaces [215, 216] but also for the preparation of glycoprotein mimetics.

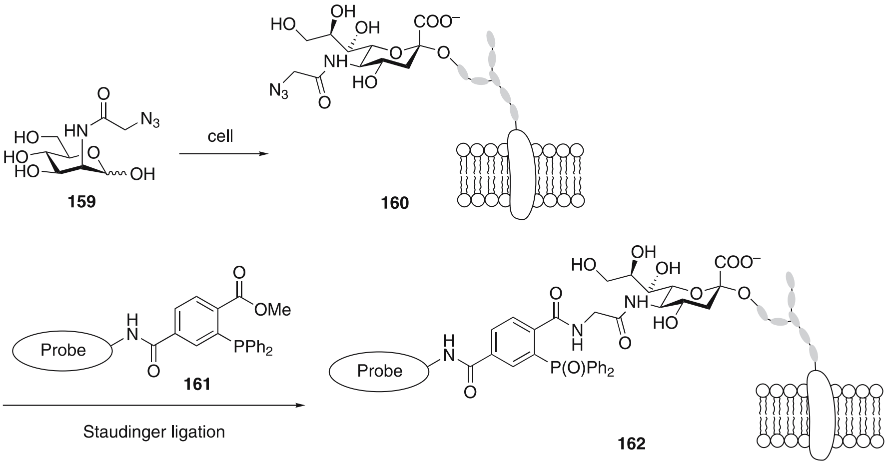
Based on work carried out earlier by the same group [221], Finn and coworkers succeeded in the attachment of more than 100 copies of the two complex tetrasaccharides **152** and **153** to the cowpea mosaic virus (CPMV) (Scheme 29) [222]. To allow for ready quantification of the conjugation, the sugar azides **152a** and **153a** were first converted into fluorescein-labeled compounds **152b** and **153b** by azide-alkyne cycloaddition to the corresponding fluorescein dialkyne. In a second step, dye alkyne derivatives **152b** and **153b** were grafted to CPMV azide **156** to give **157** with loadings of 115 and 105 per virion, respectively. Likewise, attachment of sugar azides **152a** and **153a** to alkyne-modified CPMV **154** under similar conditions gave glycoviruses **155**, assuming the same reactivity as with the fluorescein-tagged sugars. The unfortunate tendency of copper ions to accelerate the hydrolytic cleavage of peptide bonds was suppressed by the use of an excess of the sulfonated copper ligand bathophenanthroline **158** to restrict accessibility of the metal center.

4.5
Metabolic Oligosaccharide Engineering and Subsequent Bioorthogonal Ligation

Metabolic oligosaccharide engineering offers the possibility of using the cell's biosynthetic machinery to introduce carbohydrates with unnatural structural elements into the glycan chains of glycoproteins in cell culture and in animals. Reutter and coworkers have demonstrated that mannosamine derivatives, in which the *N*-acetyl group of ManNAc was substituted with for example *N*-propanoyl, *N*-butanoyl, or *N*-pentanoyl groups, can be taken up by cells and are converted into the corresponding sialosides. These are incorporated into sialoglycoconjugates, replacing in a cell-type specific manner 10–85% of normal sialic acids [223–225]. Application of this approach to different biological systems revealed important and unexpected functions of the *N*-acyl side chain of sialic acids. Sialic acids modified in such a way led, for example, to markedly altered binding and/or infection of different primate polyoma viruses that depend on cell surface sialic acids for entry. Both inhibition (over 95%) and enhancement (up to sevenfold) of virus binding and infection were observed, depending on the modification [224].

The promiscuity of the sialic acid biosynthetic pathway has been exploited by Bertozzi et al. to introduce ketone [226] and azide groups [218] into cell surface sialoglycoconjugates. Due to their unique chemical reactivity, these groups can be used for attachment of chemical probes through bioorthogonal coupling reactions. The versatility of this technology was demonstrated by an example of selective drug delivery [226]. Several human cell lines were treated with *N*-levulinoylmannosamine (ManLev). Ketone-containing cell surface glycans were conjugated to biotinamidocaproyl hydrazide through the formation of an acyl hydrazone to provide a tag for subsequent detection with fluorescein-labeled avidin or for delivery of a ricin A chain–avidin conjugate. Ricin toxin A chain is a potent inhibitor of protein synthesis. At certain concentrations, the ricin A chain–avidin conjugate showed selective toxicity for cells treated with ManLev.

Later, Bertozzi et al. developed an improved bioorthogonal ligation reaction based on a modified Staudinger reaction [218, 219]. This so-called Staudinger ligation allows a selective reaction between azides and modified triphenylphosphines containing an intramolecular electrophilic trap leading to formation of an amide bond (Scheme 30). The azide and phosphine are both abiotic and essentially unreactive toward cellular biomolecules. Since *N*-azidoacetylmannosamine (ManNAz) **159** is metabolized by cells and incorporated into surface glycoconjugates **160**, the Staudinger ligation can be used with labeled triarylphosphines **161** to produce stable cell-surface adducts **162**. Recently, Bertozzi et al. could show that the Staudinger ligation can also be executed in living animals [227]. The process of cell surface engineering offers a variety of applications including the generation of artificial cellular receptors for drug targeting, viral gene delivery, and diagnostics, the improvement of glycoprotein properties, and the attachment of labels to study the behavior of glycoproteins in vivo [225, 228, 229]. The metabolic labeling approach can also been used for proteomic analysis of mucin-type O-linked glycoproteins [230] and *O*-GlcNAc-modified proteins [231, 232].



Scheme 30 Cell surface engineering by the Staudinger ligation

5
Conclusion

The biological importance of glycans and the need for pure glycoforms of glycoproteins for biological studies and pharmaceutical applications have been the major driving forces toward the development of strategies for the preparation of these structures. The synthesis of glycopeptide and glycoprotein mimetics allows for structure–activity relationship studies but has many practical applications as well. Replacement of labile *N*- and *O*-glycosidic bonds by more stable *C*- and *S*-glycosides generates compounds with longer half-lives and simplifies their preparation by linear assembly strategies due to the increased stability of the corresponding building blocks. Convergent approaches using chemoselective ligation reactions provide access to homogeneous glycoprotein mimetics that are likely to impact our understanding of how specific glycoforms mediate physiological processes. Despite the progress made in the field, many challenges remain, e.g., the development of methods for the controlled introduction of multiple (different) glycans into proteins.

References

1. Lowe JB, Marth JD (2003) Annu Rev Biochem 72:643
2. Dube DH, Bertozzi CR (2005) Nat Rev Drug Discov 4:477
3. Shriver Z, Raguram S, Sasisekharan R (2004) Nat Rev Drug Discov 3:863
4. Ratner DM, Adams EW, Disney MD, Seeberger PH (2004) Chem Bio Chem 5:1375
5. Gabius H-J, Siebert H-C, André S, Jiménez-Barbero J, Rüdiger H (2004) Chem Bio Chem 5:740
6. Rudd PM, Elliott T, Cresswell P, Wilson IA, Dwek RA (2001) Science 291:2370
7. Helenius A, Aebi M (2001) Science 291:2364
8. Wittmann V (2001) Glycoproteins: properties. In: Fraser-Reid B, Tatsuta K, Thiem J (eds) Glycoscience: chemistry and chemical biology, vol III. Springer, Berlin Heidelberg New York, p 2289
9. Dwek RA (1996) Chem Rev 96:683
10. Varki A, Cummings R, Esko J, Freeze H, Hart G, Marth J (eds) (1999) Essentials of glycobiology. Cold Spring Harbor Laboratory Press, Cold Spring Harbor
11. Hang HC, Bertozzi CR (2005) Bioorg Med Chem 13:5021
12. Muramatsu T (2001) Glycoconjugate J 17:577
13. Wittmann V (2001) Glycoproteins: occurrence and significance. In: Fraser-Reid B, Tatsuta K, Thiem J (eds) Glycoscience: chemistry and chemical biology, vol III. Springer, Berlin Heidelberg New York, p 2253
14. Van den Steen P, Rudd PM, Dwek RA, Opdenakker G (1998) Crit Rev Biochem Mol Biol 33:151
15. Rudd PM, Dwek RA (1997) Crit Rev Biochem Mol Biol 32:1
16. Haase C, Seitz O (2007) In: Wittmann V (ed) Glycopeptides and glycoproteins. Topics in current chemistry, vol 276. Springer, Berlin Heidelberg New York (in press)

17. Thayer DA, Wong C-H (2007) In: Wittmann V (ed) Glycopeptides and glycoproteins. Topics in current chemistry, vol 276. Springer, Berlin Heidelberg New York (in press)
18. Warren JD, Geng X, Danishefsky SJ (2007) In: Wittmann V (ed) Glycopeptides and glycoproteins. Topics in current chemistry, vol 276. Springer, Berlin Heidelberg New York (in press)
19. Liu L, Bennett CS, Wong C-H (2006) Chem Commun, pp 21–33
20. Wong CH (2005) J Org Chem 70:4219
21. Pratt MR, Bertozzi CR (2005) Chem Soc Rev 34:58
22. Hölemann A, Seeberger PH (2004) Curr Opin Biotechnol 15:615
23. Davis BG (2002) Chem Rev 102:579
24. Grogan MJ, Pratt MR, Marcaurelle LA, Bertozzi CR (2002) Annu Rev Biochem 71:593
25. Brocke C, Kunz H (2002) Bioorg Med Chem 10:3085
26. Herzner H, Reipen T, Schultz M, Kunz H (2000) Chem Rev 100:4495
27. Lee YC, Lee RT (eds) (1994) Neoglycoconjugates: preparation and applications. Academic, San Diego
28. Pachamuthu K, Schmidt RR (2006) Chem Rev 106:160
29. Langenhan JM, Thorson JS (2005) Curr Org Synth 2:59
30. Davis BG (2004) Science 303:480
31. Peri F, Nicotra F (2004) Chem Commun, pp 623–627
32. Schweizer F (2002) Angew Chem Int Ed 41:230
33. Randell KD, Barkley A, Arya P (2002) Comb Chem High Throughput Screen 5:179
34. Hang HC, Bertozzi CR (2001) Acc Chem Res 34:727
35. Dondoni A, Marra A (2000) Chem Rev 100:4395
36. Barchi JJ Jr (2000) Curr Pharm Des 6:485
37. Marcaurelle LA, Bertozzi CR (1999) Chem Eur J 5:1384
38. Taylor CM (1998) Tetrahedron 54:11317
39. Tolbert TJ, Wong C-H (2000) J Am Chem Soc 122:5421
40. Arslan T, Mamaev SV, Mamaeva NV, Hecht SM (1997) J Am Chem Soc 119:10877
41. Wang L, Schultz PG (2005) Angew Chem Int Ed 44:34
42. Röhrig CH, Retz OA, Hareng L, Hartung T, Schmidt RR (2005) Chem Bio Chem 6:1805
43. Dios A, Franck RW, Geer A, Marzabadi C, Tamarez M, Capozzi G, Menichetti S, Nativi C (1996) Angew Chem Int Ed Engl 35:777
44. Venturi F, Venturi C, Liguori F, Cacciarini M, Montalbano M, Nativi C (2004) J Org Chem 69:6153
45. Bartolozzi A, Li B, Franck RW (2003) Bioorg Med Chem 11:3021
46. Marin J, Violette A, Briand J-P, Guichard G (2004) Eur J Org Chem:3027
47. Marin J, Blaton M-A, Briand J-P, Chiocchia G, Fournier C, Guichard G (2005) Chem Bio Chem 6:1796
48. Biswas K, Coltart DM, Danishefsky SJ (2002) Tetrahedron Lett 43:6107
49. Simon RJ, Kania RS, Zuckermann RN, Huebner VD, Jewell DA, Banville S, Ng S, Wang L, Rosenberg S, et al. (1992) Proc Natl Acad Sci USA 89:9367
50. Zuckermann RN, Kerr JM, Kent SBH, Moos WH (1992) J Am Chem Soc 114:10646
51. Kim JM, Roy R (1997) Carbohydr Res 298:173
52. Hu Y-J, Roy R (1999) Tetrahedron Lett 40:3305
53. Saha UK, Roy R (1995) Tetrahedron Lett 36:3635
54. Roy R, Saha UK (1996) Chem Commun, pp 201–202
55. Saha UK, Roy R (1997) Tetrahedron Lett 38:7697
56. Kim JM, Roy R (1997) Tetrahedron Lett 38:3487

57. Lee M-r, Lee J, Shin I (2002) Synlett, pp 1463–1466
58. Tornøe CW, Christensen C, Meldal M (2002) J Org Chem 67:3057
59. Rostovtsev VV, Green LG, Fokin VV, Sharpless KB (2002) Angew Chem Int Ed 41:2596
60. Breinbauer R, Koehn M (2003) Chem Bio Chem 4:1147
61. Kuijpers BHM, Groothuys S, Keereweer AR, Quaedflieg PJLM, Blaauw RH, van Delft FL, Rutjes FPJT (2004) Org Lett 6:3123
62. Ermolat'ev D, Dehaen W, Van der Eycken E (2004) QSAR Comb Sci 23:915
63. Ramos D, Rollin P, Klaffke W (2000) Angew Chem Int Ed 39:396
64. Ramos D, Rollin P, Klaffke W (2001) J Org Chem 66:2948
65. Roeckendorf N, Lindhorst TK (2004) J Org Chem 69:4441
66. van Ameijde J, Liskamp RMJ (2003) Org Biomol Chem 1:2661
67. Ichikawa Y, Matsukawa Y, Isobe M (2006) J Am Chem Soc 128:3934
68. Schäfer A, Klich G, Schreiber M, Paulsen H, Thiem J (1998) Carbohydr Res 313:107
69. Burger K, Böttcher C, Hennig L, Essawy SA (2004) Monatsh Chem 135:865
70. Reddy BG, Madhusudanan KP, Vankar YD (2004) J Org Chem 69:2630
71. Seeberger PH, Haase W-C (2000) Chem Rev 100:4349
72. Knapp S, Myers DS (2001) J Org Chem 66:3636
73. Knapp S, Myers DS (2002) J Org Chem 67:2995
74. Zhu X, Schmidt RR (2003) Tetrahedron Lett 44:6063
75. Zhu X, Schmidt RR (2004) Chem Eur J 10:875
76. Jobron L, Hummel G (2000) Org Lett 2:2265
77. Hummel G, Hindsgaul O (1999) Angew Chem Int Ed 38:1782
78. Cohen SB, Halcomb RL (2001) Org Lett 3:405
79. Cohen SB, Halcomb RL (2002) J Am Chem Soc 124:2534
80. Zhu X, Pachamuthu K, Schmidt RR (2003) J Org Chem 68:5641
81. Zhu X, Haag T, Schmidt RR (2004) Org Biomol Chem 2:31
82. Zhu X, Pachamuthu K, Schmidt RR (2004) Org Lett 6:1083
83. Hofsteenge J, Mueller DR, de Beer T, Loeffler A, Richter WJ, Vliegenthart JFG (1994) Biochemistry 33:13524
84. Manabe S, Ito Y (1999) J Am Chem Soc 121:9754
85. Kostír JV, Queisnerová M (1949) Chem Listy 43:277
86. Kessler H, Wittmann V, Köck M, Kottenhahn M (1992) Angew Chem Int Ed Engl 31:902
87. Bertozzi CR, Hoeprich PD, Bednarski MD (1992) J Org Chem 57:6092
88. Postema MHD (1995) C-Glycoside synthesis. CRC, Boca Raton, FL
89. Levy DE, Tang C (1995) The chemistry of C-glycosides. Elsevier, New York
90. Wellner E, Gustafsson T, Backlund J, Holmdahl R, Kihlberg J (2000) Chem Bio Chem 1:272
91. Gustafsson T, Saxin M, Kihlberg J (2003) J Org Chem 68:2506
92. Gustafsson T, Hedenstroem M, Kihlberg J (2006) J Org Chem 71:1911
93. Penner M, Taylor D, Desautels D, Marat K, Schweizer F (2005) Synlett, pp 212–216
94. Xu X, Fakha G, Sinou D (2002) Tetrahedron 58:7539
95. Wernicke A, Sinou D (2001) J Carbohydr Chem 20:181
96. Ousmer M, Boucard V, Lubin-Germain N, Uziel J, Augé J (2006) Eur J Org Chem, pp 1216–1221
97. Pearce AJ, Ramaya S, Thorn SN, Bloomberg GB, Walter DS, Gallagher T (1999) J Org Chem 64:5453
98. Lowary T, Meldal M, Helmboldt A, Vasella A, Bock K (1998) J Org Chem 63:9657

99. Dondoni A, Mariotti G, Marra A (2002) J Org Chem 67:4475
100. Vincent SP, Schleyer A, Wong CH (2000) J Org Chem 65:4440
101. Ben RN, Orellana A, Arya P (1998) J Org Chem 63:4817
102. Kröger L, Henkensmeier D, Schäfer A, Thiem J (2004) Bioorg Med Chem Lett 14:73
103. Kobertz WR, Bertozzi CR, Bednarski MD (1992) Tetrahedron Lett 33:737
104. Wang Q, Linhardt RJ (2003) J Org Chem 68:2668
105. Michael K, Wittmann V, König W, Sandow J, Kessler H (1996) Int J Pept Protein Res 48:59
106. Sommermann T, Kim BG, Peters K, Peters EM, Linker T (2004) Chem Commun, pp 2624–2625
107. Caddick S, Hamza D, Wadman SN, Wilden JD (2002) Org Lett 4:1775
108. Debenham SD, Snyder PW, Toone EJ (2003) J Org Chem 68:5805
109. Lundquist JJ, Debenham SD, Toone EJ (2000) J Org Chem 65:8245
110. Dondoni A, Giovannini PP, Massi A (2004) Org Lett 6:2929
111. Leeuwenburgh MA, van der Marel GA, Overkleeft HS (2003) Curr Opin Chem Biol 7:757
112. Dominique R, Liu B, Das SK, Roy R (2000) Synthesis, pp 862–868
113. Dechantsreiter MA, Burkhart F, Kessler H (1998) Tetrahedron Lett 39:253
114. McGarvey GJ, Benedum TE, Schmidtmann FW (2002) Org Lett 4:3591
115. Liu S, Ben RN (2005) Org Lett 7:2385
116. Ben RN, Eniade AA, Hauer L (1999) Org Lett 1:1759
117. Eniade A, Murphy AV, Landreau G, Ben RN (2001) Bioconjugate Chem 12:817
118. Nolen EG, Kurish AJ, Potter JM, Donahue LA, Orlando MD (2005) Org Lett 7:3383
119. Vizvardi K, Kreutz C, Davis AS, Lee VP, Philmus BJ, Simo O, Michael K (2003) Chem Lett 32:348
120. Simo O, Lee VP, Davis AS, Kreutz C, Gross PH, Jones PR, Michael K (2005) Carbohydr Res 340:557
121. Jayakanthan K, Vankar YD (2005) Org Lett 7:5441
122. Cervigni SE, Dumy P, Mutter M (1996) Angew Chem Int Ed Engl 35:1230
123. Zhao Y, Kent SBH, Chait BT (1997) Proc Natl Acad Sci USA 94:1629
124. Rodriguez EC, Winans KA, King DS, Bertozzi CR (1997) J Am Chem Soc 119:9905
125. Rodriguez EC, Marcaurelle LA, Bertozzi CR (1998) J Org Chem 63:7134
126. Liu H, Wang L, Brock A, Wong C-H, Schultz PG (2003) J Am Chem Soc 125:1702
127. Wang L, Brock A, Herberich B, Schultz PG (2001) Science 292:498
128. Wang L, Zhang Z, Brock A, Schultz PG (2003) Proc Natl Acad Sci USA 100:56
129. Renaudet O, Dumy P (2003) Org Lett 5:243
130. Renaudet O, Dumy P (2004) Tetrahedron Lett 45:65
131. Renaudet O, Dumy P (2005) Bioorg Med Chem Lett 15:3619
132. Singh Y, Renaudet O, Defrancq E, Dumy P (2005) Org Lett 7:1359
133. Khidekel N, Arndt S, Lamarre-Vincent N, Lippert A, Poulin-Kerstien KG, Ramakrishnan B, Qasba PK, Hsieh-Wilson LC (2003) J Am Chem Soc 125:16162
134. Khidekel N, Ficarro SB, Peters EC, Hsieh-Wilson LC (2004) Proc Natl Acad Sci USA 101:13132
135. Hang HC, Bertozzi CR (2001) J Am Chem Soc 123:1242
136. Peri F, Cipolla L, La Ferla B, Dumy P, Nicotra F (1999) Glycoconjugate J 16:399
137. Cipolla L, Rescigno M, Leone A, Peri F, La Ferla B, Nicotra F (2002) Bioorg Med Chem 10:1639
138. Peri F, Cipolla L, Rescigno M, La Ferla B, Nicotra F (2001) Bioconjugate Chem 12:325
139. Peri F, Cipolla L, La Ferla B, Nicotra F (2003) C R Chim 6:635

140. Peluso S, Imperiali B (2001) Tetrahedron Lett 42:2085
141. Peluso S, Ufret ML, O'Reilly MK, Imperiali B (2002) Chem Biol 9:1323
142. Hatanaka Y, Kempin U, Jong-Jip P (2000) J Org Chem 65:5639
143. Lee M-r, Shin I (2005) Org Lett 7:4269
144. Shinohara Y, Sota H, Gotoh M, Hasebe M, Tosu M, Nakao J, Hasegawa Y, Shiga M (1996) Anal Chem 68:2573
145. Ojala CR, Ostman JM, Ojala WH (2002) Carbohydr Res 337:21
146. Ernholt BV, Thomsen IB, Lohse A, Plesner IW, Jensen KB, Hazell RG, Liang X, Jakobsen A, Bols M (2000) Chem Eur J 6:278
147. Bendiak B (1997) Carbohydr Res 304:85
148. Lohse A, Martins R, Jørgensen MR, Hindsgaul O (2006) Angew Chem Int Ed 45:4167
149. Peri F, Dumy P, Mutter M (1998) Tetrahedron 54:12269
150. Carrasco MR, Nguyen MJ, Burnell DR, MacLaren MD, Hengel SM (2002) Tetrahedron Lett 43:5727
151. Carrasco MR, Brown RT, Serafimova IM, Silva O (2003) J Org Chem 68:195
152. Carrasco MR, Brown RT (2003) J Org Chem 68:8853
153. Filira F, Biondi B, Biondi L, Giannini E, Gobbo M, Negri L, Rocchi R (2003) Org Biomol Chem 1:3059
154. Matsubara N, Oiwa K, Hohsaka T, Sadamoto R, Niikura K, Fukuhara N, Takimoto A, Kondo H, Nishimura S-I (2005) Chem Eur J 11:6974
155. Hohsaka T, Sisido M (2002) Curr Opin Chem Biol 6:809
156. Kutterer KMK, Barnes ML, Arya P (1999) J Comb Chem 1:28
157. Arya P, Kutterer KMK, Barkley A (2000) J Comb Chem 2:120
158. Arya P, Barkley A, Randell KD (2002) J Comb Chem 4:193
159. Xian M, Fatima Z, Zhang W, Fang J, Li H, Pei D, Loo J, Stevenson T, Wang PG (2004) J Comb Chem 6:126
160. Youngquist RS, Fuentes GR, Lacey MP, Keough T (1995) J Am Chem Soc 117:3900
161. St. Hilaire PM, Lowary TL, Meldal M, Bock K (1998) J Am Chem Soc 120:13312
162. Wittmann V, Seeberger S (2000) Angew Chem Int Ed 39:4348
163. Wittmann V, Seeberger S (2004) Angew Chem Int Ed 43:900
164. Wittmann V, Seeberger S, Schägger H (2003) Tetrahedron Lett 44:9243
165. Ohta T, Miura N, Funitani N, Nakajima F, Niikura K, Sadamoto R, Guo C-T, Suzuki T, Suzuki Y, Monde K, Nishimura S-I (2003) Angew Chem Int Ed 42:5186
166. Raja KS, Wang Q, Finn MG (2003) Chem Bio Chem 4:1348
167. Pearce OMT, Fisher KD, Humphries J, Seymour LW, Smith A, Davis BG (2005) Angew Chem Int Ed 44:1057
168. Lee YC, Stowell CP, Krantz MJ (1976) Biochemistry 15:3956
169. Robinson MA, Charlton ST, Garnier P, Wang X-t, Davis SS, Perkins AC, Frier M, Duncan R, Savage TJ, Wyatt DA, Watson SA, Davis BG (2004) Proc Natl Acad Sci USA 101:14527
170. Davis NJ, Flitsch SL (1991) Tetrahedron Lett 32:6793
171. Wong SYC, Guile GR, Dwek RA, Arsequell G (1994) Biochem J 300:843
172. Bengtsson M, Broddefalk J, Dahmén J, Henriksson K, Kihlberg J, Lonn H, Srinivasa BR, Stenvall K (1998) Glycoconjugate J 15:223
173. Macmillan D, Bill RM, Sage KA, Fern D, Flitsch SL (2001) Chem Biol 8:133
174. Backes BJ, Ellman JA (1999) J Org Chem 64:2322
175. Macmillan D, Daines AM, Bayrhuber M, Flitsch SL (2002) Org Lett 4:1467
176. Macmillan D, Arham L (2004) J Am Chem Soc 126:9530
177. Dawson PE, Muir TW, Clark-Lewis I, Kent SBH (1994) Science 266:776

178. Yamamoto N, Sakakibara T, Kajihara Y (2004) Tetrahedron Lett 45:3287
179. Marcaurelle LA, Bertozzi CR (2001) J Am Chem Soc 123:1587
180. Marcaurelle LA, Pratt MR, Bertozzi CR (2003) Chem Bio Chem 4:224
181. Pratt MR, Bertozzi CR (2003) J Am Chem Soc 125:6149
182. Galonic DP, van der Donk WA, Gin DY (2004) J Am Chem Soc 126:12712
183. Kitagawa T, Aikawa T (1976) J Biochem (Tokyo) 79:233
184. Hermanson GT (1996) Bioconjugate techniques. Academic, San Diego, p 148
185. Hansen PR, Olsen CE, Holm A (1998) Bioconjugate Chem 9:126
186. Shin I, Jung H-J, Cho J (2000) Bull Korean Chem Soc 21:845
187. Shin I, Jung H-J, Lee M-r (2001) Tetrahedron Lett 42:1325
188. Park S, Shin I (2002) Angew Chem Int Ed 41:3180
189. Park S, Lee M-r, Pyo S-J, Shin I (2004) J Am Chem Soc 126:4812
190. Zhu Y, van der Donk WA (2001) Org Lett 3:1189
191. Galonic DP, van der Donk WA, Gin DY (2003) Chem Eur J 9:5997
192. Gieselman MD, Zhu YT, Zhou H, Galonic D, van der Donk WA (2002) Chem Bio Chem 3:709
193. Macindoe WM, van Oijen AH, Boons G-J (1998) Chem Commun, pp 847–848
194. Watt GM, Boons G-J (2004) Carbohydr Res 339:181
195. Watt GM, Lund J, Levens M, Kolli VSK, Jefferis R, Boons G-J (2003) Chem Biol 10: 807
196. Davis BG, Maughan MAT, Green MP, Ullman A, Jones JB (2000) Tetrahedron: Asymmetry 11:245
197. Davis BG, Lloyd RC, Jones JB (2000) Bioorg Med Chem 8:1527
198. Lloyd RC, Davis BG, Jones JB (2000) Bioorg Med Chem 8:1537
199. Gamblin DP, Garnier P, Ward SJ, Oldham NJ, Fairbanks AJ, Davis BG (2003) Org Biomol Chem 1:3642
200. Davis BG (2001) Chem Commun, pp 351–352
201. Rendle PM, Seger A, Rodrigues J, Oldham NJ, Bott RR, Jones JB, Cowan MM, Davis BG (2004) J Am Chem Soc 126:4750
202. Gamblin DP, Garnier P, van Kasteren S, Oldham NJ, Fairbanks AJ, Davis BG (2004) Angew Chem Int Ed 43:828
203. Ramström O, Lehn J-M (2002) Nat Rev Drug Discov 1:26
204. Rowan SJ, Cantrill SJ, Cousins GRL, Sanders JKM, Stoddart JF (2002) Angew Chem Int Ed 41:898
205. Otto S, Furlan RLE, Sanders JKM (2002) Drug Discov Today 7:117
206. Lehn J-M, Eliseev AV (2001) Science 291:2331
207. Sakai S, Shigemasa Y, Sasaki T (1997) Tetrahedron Lett 38:8145
208. Ramström O, Lehn J-M (2000) Chem Bio Chem 1:41
209. Ramström O, Lohmann S, Bunyapaiboonsri T, Lehn J-M (2004) Chem Eur J 10:1711
210. Sando S, Narita A, Aoyama Y (2004) Bioorg Med Chem Lett 14:2835
211. Hotchkiss T, Kramer HB, Doores KJ, Gamblin DP, Oldham NJ, Davis BG (2005) Chem Commun, pp 4264–4266
212. Lin H, Walsh CT (2004) J Am Chem Soc 126:13998
213. van Hest JCM, Kiick KL, Tirrell DA (2000) J Am Chem Soc 122:1282
214. Kiick KL, Saxon E, Tirrell DA, Bertozzi CR (2002) Proc Natl Acad Sci USA 99:19
215. Link AJ, Tirrell DA (2003) J Am Chem Soc 125:11164
216. Link AJ, Vink MKS, Tirrell DA (2004) J Am Chem Soc 126:10598
217. Link AJ, Mock ML, Tirrell DA (2003) Curr Opin Chem Biol 14:603
218. Saxon E, Bertozzi CR (2000) Science 287:2007

219. Lin FL, Hoyt HM, Van Halbeek H, Bergman RG, Bertozzi CR (2005) J Am Chem Soc 127:2686
220. Köhn M, Breinbauer R (2004) Angew Chem Int Ed 43:3106
221. Wang Q, Chan TR, Hilgraf R, Fokin VV, Sharpless KB, Finn MG (2003) J Am Chem Soc 125:3192
222. Gupta SS, Kuzelka J, Singh P, Lewis WG, Manchester M, Finn MG (2005) Bioconjugate Chem 16:1572
223. Kayser H, Zeitler R, Kannicht C, Grunow D, Nuck R, Reutter W (1992) J Biol Chem 267:16934
224. Keppler OT, Stehling P, Herrmann M, Kayser H, Grunow D, Reutter W, Pawlita M (1995) J Biol Chem 270:1308
225. Keppler OT, Horstkorte R, Pawlita M, Schmidt C, Reutter W (2001) Glycobiology 11:11R
226. Mahal LK, Yarema KJ, Bertozzi CR (1997) Science 276:1125
227. Prescher JA, Dube DH, Bertozzi CR (2004) Nature 430:873
228. Prescher JA, Bertozzi CR (2005) Nat Chem Biol 1:13
229. Dube DH, Bertozzi CR (2003) Curr Opin Chem Biol 7:616
230. Hang HC, Yu C, Kato DL, Bertozzi CR (2003) Proc Natl Acad Sci USA 100:14846
231. Vocadlo DJ, Hang HC, Kim E-J, Hanover JA, Bertozzi CR (2003) Proc Natl Acad Sci USA 100:9116
232. Laughlin ST, Bertozzi CR (2004) Glycobiology 14:1192
233. Thayer DA, Yu HN, Galan MC, Wong C-H (2005) Angew Chem Int Ed 44:4596

Top Curr Chem (2007) 267: 109–141
DOI 10.1007/128_031

Synthetic Glycopeptide-Based Vaccines

J. David Warren[1] · Xudong Geng[3] · Samuel J. Danishefsky[1,2] (✉)

[1]Laboratory for Bioorganic Chemistry, Sloan-Kettering Institute for Cancer Research,
1275 York Avenue, New York, NY 10021, USA
s-danishefsky@ski.mskcc.org

[2]Department of Chemistry, Columbia University, Havemeyer Hall, 3000 Broadway,
New York, NY 10027, USA
s-danishefsky@ski.mskcc.org

[3]Global Discovery Chemistry, Novartis Institute for Biomedical Research, Inc.,
100 Technology Square, Cambridge, MA 02139, USA

Abstract This review provides an overview of our explorations into oligosaccharide and glycoconjugate construction for the creation and evaluation of glycopeptide-based vaccines. The basis for these investigations is the known tendency of both cancer cells and viruses to express selective carbohydrate motifs in the form of glycoproteins or glycolipids. Utilization of these carbohydrates in a glycopeptide-based vaccine could potentially trigger immune recognition, generating a protective response against the disease. However, obtaining large quantities of such compounds from natural sources is extremely difficult. Over the past two decades, our lab has been engaged in the total synthesis of complex oligosaccharides and glycoconjugates. With this knowledge and experience, we have begun to evaluate, in many cases at the clinical level, whether the human immune system is capable of mounting a response against such fully synthetic carbohydrate antigens in a focused and useful way. Toward this goal, we have merged the powers of both chemistry and immunology to provide insight into this problem. The synthesis and evaluation of potential vaccines for both cancer and HIV will be described.

Keywords Cancer vaccine · Drug discovery · Glycoconjugates · HIV vaccine · Synthetic carbohydrate antigens

Abbreviations

Ac	acetyl
Bn	benzyl
Boc	*tert*-butoxycarbonyl
Bz	benzoyl
DMDO	3,3-dimethyldioxirane
DTBP	di-*tert*-butylpyridine
ELISA	enzyme-linked immunosorbent assay
Fmoc	fluoren-9-ylmethylcarbonyl
ivDde	4,4-dimethyl-2,6-dioxocyclohex-1-ylidine-3-methylbutyl
KLH	keyhole limpet hemocyanin
MBS	*m*-maleimidobenzoyl-N-hydroxysuccinimide ester
MCA	monochloroacetyl
MesNa	sodium 2-mercaptoethanesulfonate
MMCCH	4-(4-*N*-maleimidomethyl)cyclohexane-1-carboxylic acid hydrazide
NCL	native chemical ligation
Pam_3Cys	tripalmitoyl-*S*-glycerylcysteinylserine
Phth	phthaloyl
py	pyridine
SCLC	small cell lung cancer
SPR	surface plasmon resonance
TBAF	tetrabutylammonium fluoride
TBS	*tert*-butyldimethylsilyl
TFA	trifluoroacetic acid
TIPS	triisopropylsilyl

1
Introduction

The first successful human vaccination was performed by Edward Jenner in 1796. Having observed that a number of milkmaids who had previously contracted cowpox were resistant to smallpox, a similar virus, Jenner attempted to protect a teenaged boy, James Phipps, from the disfiguring and possibly deadly effects of smallpox. In his experiment, he "vaccinated" (from the Latin term *vacca*, for cow) Phipps with pustular material from the hand of Sarah Nelmes. She had contracted the infection from a cow named Blossom, whose hide hangs today in London, at St. George's Hospital. After six weeks, Jenner challenged Phipps with virulent smallpox and fortunately he did not contract the disease. Almost 75 years later, Louis Pasteur first used the terms "immune" and "immunity" in the scientific sense; however, he acknowledged Jenner's pioneering research by retaining the word "vaccination" to describe his own accomplishments in the prevention of rabies and anthrax. Since then, great success has been realized in the development of vaccines to manage and reverse infections caused by bacteria, viruses and parasites, notably those for diphtheria (von Behring) and polio (Salk) among many others [1–3].

It has been the ultimate goal of many researchers to develop a vaccine that would rely on stimulating the immune system to treat (or even prevent) diseases such as cancer and HIV/AIDS [4–9]. Traditional immunization against bacterial and viral infection include killed, live-attenuated, and toxoid vaccines. The advent of molecular biology, the development of analytical methods, and improvements in synthetic methodology have allowed for the production of vaccines based on purified subunits and genetically altered attenuated strains of the infective agents; however, one could argue that a future vaccine for either cancer or HIV would necessitate the development of a fully synthetic immunogen. Such a vaccine would contain a well-defined composition allowing for highly reproducible biological properties.

An important advance in vaccine development has been the use of conjugate vaccines. Our laboratory has been focused on the generation of vaccines for cancer and HIV utilizing fully synthetic glycopeptide-based antigens [10–12]. The attachment of a glycopeptide-based antigen to an immunogenic carrier would, in theory, better arm the immune system to generate a protective response against the disease. Experience suggests that only by conjugating an otherwise non-immunogenic glycopeptide antigen to an immunogenic carrier can it be presented to the immune system in a truly effective fashion. The immunogenic carriers often used are proteins or short peptide sequences, although others are currently being explored.

A number of factors can contribute to the effectiveness of a glycopeptide-based conjugate vaccine, including proper selection of the glycopeptide antigen construct, the nature of the immunogenic carrier, the ratio of the antigen to the carrier, and the nature of the linkage between the two, not to mention the homogeneity of the entire construct. There are a number of different methods for generating the glycoconjugates; however, they will not be covered in this review. Instead we will focus on our efforts toward the synthesis of the glycopeptide portion of the vaccine.

Many of the glycopeptide-based antigens we have considered for inclusion in either a cancer or HIV vaccine are not widely available from natural sources. For those that are, the measures necessary for their isolation, purification and identification are tedious, low-yielding, and in the end, impractical. It is here that synthetic organic chemistry plays a critical role. Synthesis allows for the rapid assembly of complex oligosaccharide domains through the use of glycosylation methodology developed in the last half century [13–20]. By utilizing synthesis, one can arrive at a product of unquestioned structural integrity and of purity appropriate for the clinic. Furthermore, synthesis allows for alteration of the carbohydrate domain during optimization of the vaccine construct, an option not easily achievable if the antigen must come from a natural source.

2
Cancer Vaccines

During embryogenesis, changes in cellular glycosylation patterns are characteristic of activation and embryonic development [21]. It is not surprising, then, that malignant transformation and tumor progression are also accompanied by glycosylation changes. Certain carbohydrates of cancer cells frequently display anomalous glycosylation patterns, and the specific type of antigen displayed is often associated with a particular cancer. The degree of the tumor's antigen expression is often diagnostic of the disease's progression, and can therefore foreshadow the prognosis for treatment. The notion that tumor-associated antigens could be used as vaccines has been considered and debated for some time, and is emerging as a promising therapy for inducing the immune system to mount a tumor-specific immune response. However, many tumor-bearing hosts do not recognize malignant cells, indicating that the expressed tumor antigens may be insufficiently immunogenic.

The primary responsibility of an antibody generated by an immune response is to eliminate circulating pathogens from the blood stream. Antibodies generated from a carbohydrate-based vaccine could then identify cancerous cells and initiate a series of cascades, hopefully leading to their selective elimination [22]. Since the target of this type of therapy includes micrometastases and circulating tumor cells, it would be adjuvant in nature and would function to prevent tumor metastasis or reoccurrence after primary therapy (chemotherapy, radiation, surgery, etc.) has relieved tumor burden.

A number of issues must be addressed when considering the use of carbohydrate antigens as targets for active immunotherapy. Classified as T-cell-independent antigens, carbohydrates typically evade an adaptive immune response (i.e., the response characteristic of an antigenic protein) [23]. Studies of carbohydrate-based vaccines in mice have indicated the production of IgM antibodies as a predominate response (e.g., [24]). Despite repeated vaccination, isotype switching to IgG antibodies is not common, although some IgG production has been observed (e.g., [25]). The use of conjugate vaccines has been viewed as critical in overcoming the lack of a T-cell response. An appropriately conjugated vaccine can induce a T-cell-dependent response, leading to the production of protective IgG antibodies and immunologic memory specific for the carbohydrate antigen [26].

Another factor to be considered en route to a carbohydrate-based cancer vaccine is that isolation of carbohydrate antigens from natural sources is difficult at best, and often results in minute quantities of material obtained. The immense impediments associated with their purification from natural sources renders them virtually impossible to obtain as homogeneous starting materials for a clinical program. Accordingly, the burden (or opportunity) falls to the organic chemist to solve the purity and availability problems if the program is ever to advance to a clinical setting. Furthermore, chemistry must also play a vital role in the conjugation phase. For a synthetically derived construct to be useable, it must be obtained in a readily conjugatable form.

Our laboratory has been engaged in the synthesis of complex oligosaccharides and glycoconjugates for a number of years [10, 16, 27]. The following sections will detail our synthetic efforts into the realm of complex oligosaccharide-based cancer vaccines. Particular emphasis will be placed on our more recent use of glycopeptide-based vaccine constructs that mimic the cell surface of tumors.

2.1
Vaccine Synthesis Strategy

When contemplating the synthesis of complex glycoconjugate-containing substructures, one must consider a route that is not only efficient but also allows for the ability to establish a linker domain through a spacer unit in order to generate a functional immunogen. From the standpoint of synthetic economy, the use of protecting group manipulations should be minimized. Toward this goal, we perceived a great advantage in using the glycal assembly method to rapidly construct the oligosaccharides of carbohydrate-based tumor antigens. The logic and details of the power of glycal assembly have been thoroughly reviewed and will not be covered further, except by application [16]. Once the synthesis of the carbohydrate is complete, attachment of an appropriate immunostimulant would follow. While the optimal spacer-linker combinations have yet to be thoroughly established, the overall goal would be the development of a system that would allow for molecular recognition of the synthetic antigen by the immune system to take place in the presence of the conjugated biocarrier [28–30].

Thus, the overall preclinical phases of a successful cancer vaccine program include:

1. Identification of tumor-selective, or tumor-specific antigens.
2. Rigorous structural proof of the antigen.
3. Chemical synthesis of the antigen.
4. Introduction of an appropriate spacer unit that would allow for attachment to the protein domain while maintaining the immunological integrity of the antigen.

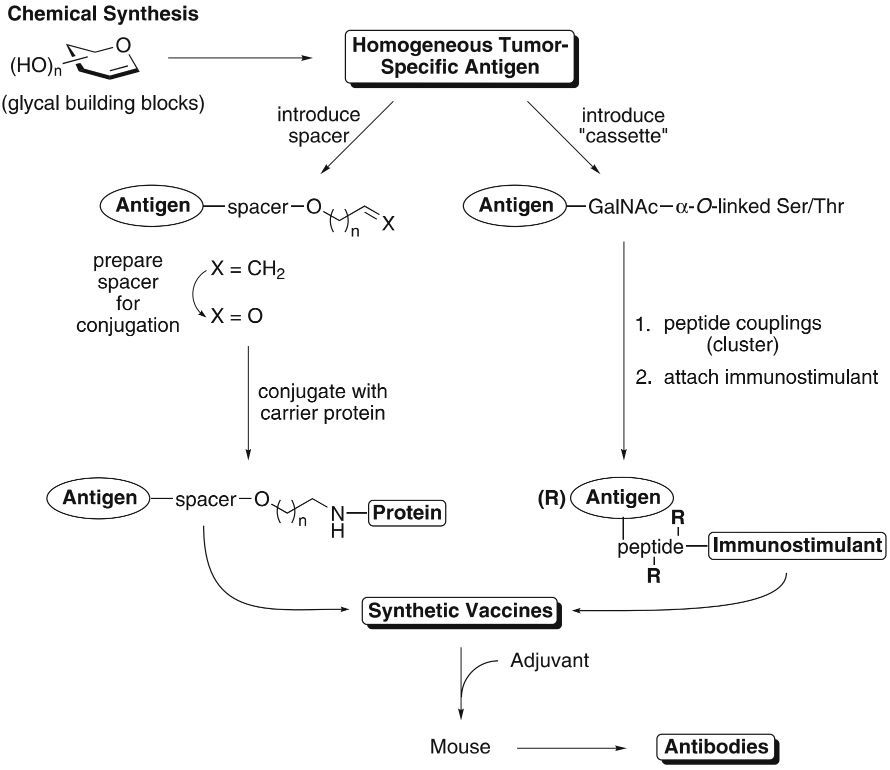


Fig. 1 General approach to synthetic carbohydrate vaccines

5. Covalent attachment of an immunogenic carrier protein or other immunostimulant to allow for a more potent vaccine.
6. Immunological studies using murine hosts to evaluate the vaccine's efficacy.

These goals having been accomplished; the vaccine could then emerge as a candidate for advancement into human clinical trials. The interface of these goals with the glycal assembly method is depicted in Fig. 1.

2.2
Monomeric Vaccines

After the first reports of tumor associated carbohydrate antigens, the development of cancer vaccines which exploit such moieties has been the focus of intense study. In the past two decades our lab has concentrated on many aspects of vaccine construction, including: choice of antigen [10], methodological development of antigen synthesis [16], conjugation to an appropriate

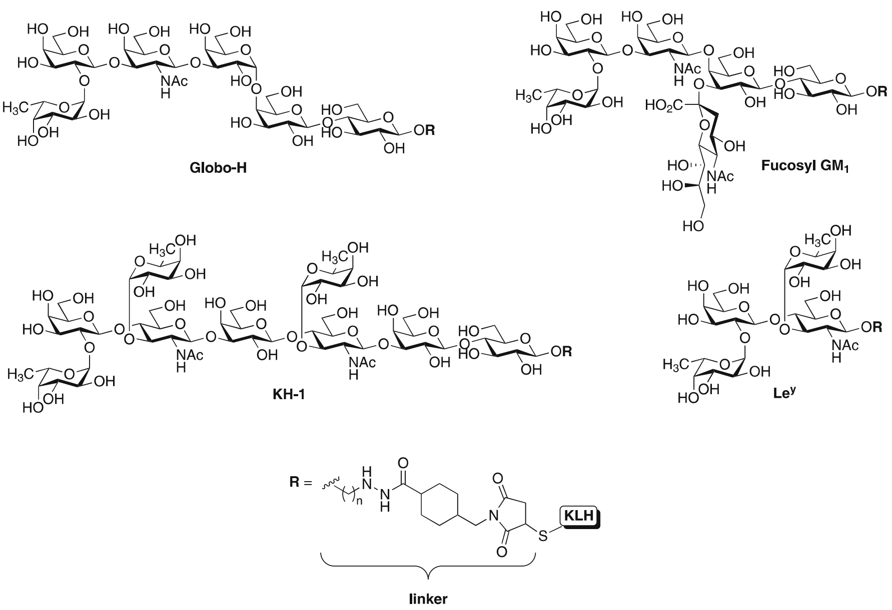


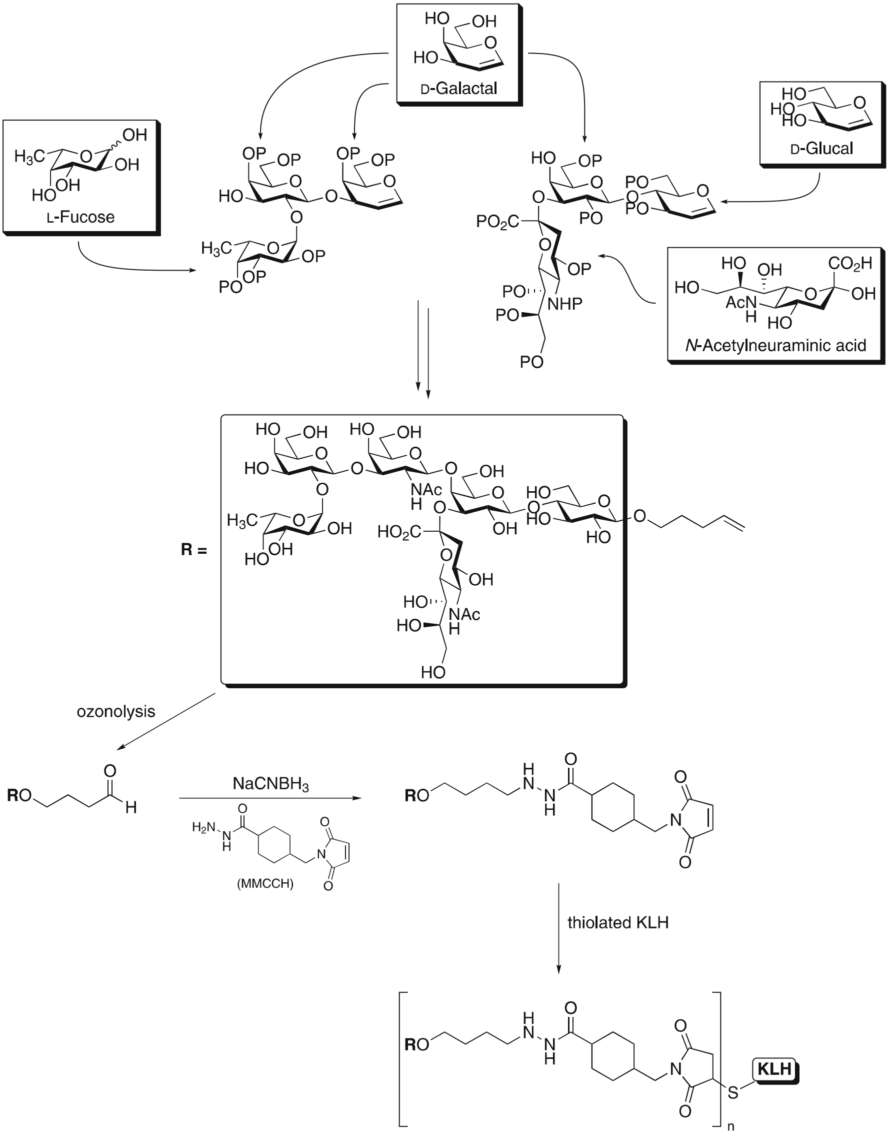
Fig. 2 Monomeric vaccine constructs discussed in this account

immunogenic carrier [27], and evaluation of various immunological adjuvants for coadministration [31]. Our first entries into the field included vaccines that contained a single carbohydrate conjugated to a carrier protein (i.e., monomeric) (Fig. 2). These carbohydrates were synthesized as allyl or pentenyl glycosides such that ozonolysis of the terminal olefin would generate an aldehyde that could be used as a chemical implement for the attachment of a spacer unit and then conjugated to an immunogenic protein.

2.2.1
Fucosyl GM_1

The glycolipid fucosyl GM_1 has been identified as a highly specific marker associated with small cell lung cancer (SCLC) cells [32–35]. It is the major ganglioside component contained in human SCLC tissue, expressed more frequently and more abundantly than either GM_2 or GD_3 [36]. Immunohistochemistry studies using the F12 monoclonal antibody (mAb) have indicated that the distribution of fucosyl GM_1 in normal tissue is highly restricted and virtually nonexistent in any human cancer cell line other than SCLC [37], making it an attractive vaccine target. Early clinical studies using bovine-derived fucosyl GM_1 demonstrated that a potent immunological response could be obtained, with induction of both IgM and IgG antibodies, upon conjugation to KLH [38] and coadministration with the adjuvant QS-21 [39]. In this case, conjugation occurred after ozonolysis of the ceramide olefin followed by reductive amination of the resulting aldehyde to the lysine – NH_2 groups on KLH.

Given the prospective importance of fucosyl-GM_1 in a carbohydrate-based attack on SCLC, we began a synthetic program that would allow access to the hexasaccharide epitope. A total chemical synthesis would not only authenticate the structural assignments in the carbohydrate but



Scheme 1 Synthesis of fucosyl-GM_1-KLH conjugate

would also confirm that the specificity of the F12 mAb is directed at the carbohydrate sector. The construct was synthesized as a pentenyl glycoside using the glycal assembly approach to oligosaccharide synthesis (Scheme 1) [40]. Elaboration of the pentenyl group to an aldehyde followed by reductive coupling to the $-NH_2$ group of the heterobifunctional cross-linker 4-(4-*N*-maleimidomethyl)cyclohexane-1-carboxylic acid hydrazide (MMCCH) allowed for direct coupling to thiolated KLH.

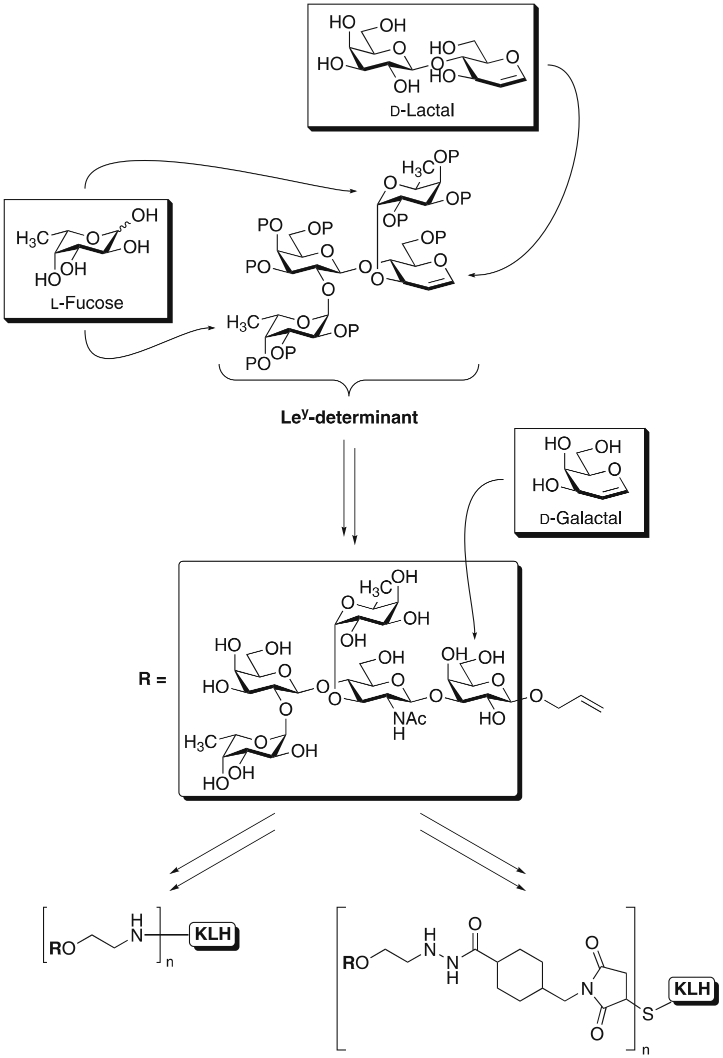
With a chemical synthesis realized, focus on the fucosyl GM_1 antigen shifted to issues of immunology and vaccinology. Early Phase I clinical trials have demonstrated that vaccination with the synthetic glycoconjugate induces an IgM antibody response against fucosyl GM_1 and tumor cells expressing fucosyl GM_1 comparable with the response induced by the bovine-derived counterpart [41].

2.2.2
Lewis ^{y}

Lewis ^{y} (Le ^{y}) is a carbohydrate specificity belonging to the A, B, H, and Lewis family of blood group determinates. Members of this group serve to control cellular growth and differentiation, mediate cell–cell adhesion, and instigate an immune response in inflammatory processes. Le ^{y} has been identified as one of the cell surface carbohydrates that is over-expressed in a majority of carcinomas, including ovary, breast, pancreas, colon, prostate, and non-small cell lung cancers [42]. In ovarian cancers, Le ^{y} is preferentially expressed in both serous and endometrioid carcinomas, but is poorly expressed in mucinous tumors [43]. It is also expressed on some normal cells, including epithelial cells and their esophageal secretions, stomach, proximal small intestine, and some acinar cells of the pancreas [42, 44].

Our first Le ^{y}-containing vaccine incorporated the Le ^{y} pentasaccharide as an allyl glycoside [45, 46]. The synthetic route took full advantage of the *N*-acetyllactosamine backbone by starting with the readily available D-lactal (Scheme 2). Again, utilization of the powerful glycal assembly method allowed for the generation of significant amounts of the allyl glycoside.

For immunological studies, the allyl glycoside was ozonolyzed to the corresponding aldehyde and conjugated directly to KLH. Alternatively, it was attached to the MMCCH linker and then coupled to thiolated KLH. Initial studies using murine models indicated both the Le ^{y}-KLH and Le ^{y}-MMCCH-KLH conjugates were able to induce the formation of both IgG and IgM antibodies capable of reacting with epitopes carried on Le ^{y} glycolipids and mucins and with tumor cells expressing Le ^{y}, when administered with the immunological adjuvant QS21 [47]. In a Phase I clinical trial in ovarian cancer patients, the Le ^{y}-KLH conjugate together with QS21 was able to induce an antibody response in 67% of the patients [48]. A majority of the responses were of the IgM class, although 17% did exhibit a clear IgG response.

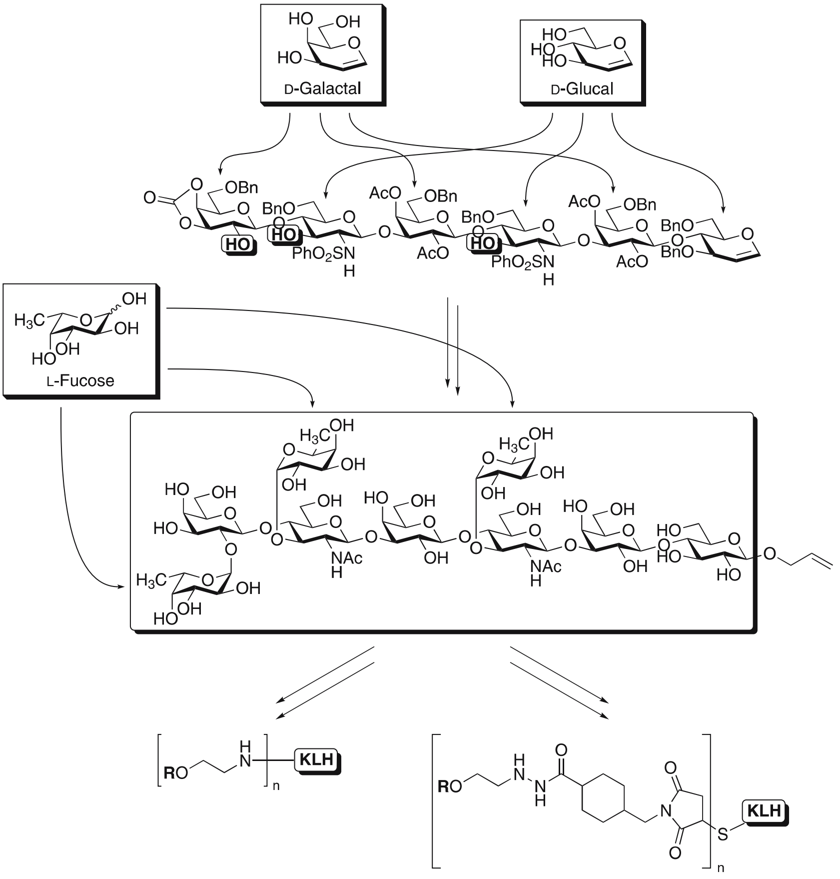


Scheme 2 Synthesis of the Le^y-KLH glycoconjugate

2.2.3
KH-1

The glycolipid KH-1 is perhaps the most daunting carbohydrate-based tumor antigen thus far characterized [49], and synthesized [50–52]. The KH-1 antigen was identified in 1986 using the monoclonal antibody AH6, and its overexpression has been observed in several human adenocarcinomas [42, 53, 54]. Structural studies of the isolate indicated the presence of epitopes corresponding to both the Le^y tetrasaccharide and the Le^x trisaccharide. Monoclonal antibodies raised against the KH-1 antigen were found to have specific binding, leading to the postulation that it is a highly specific marker for malignancy and pre-malignancies involving colonic adenocarcinoma. This, coupled with the fact that the presence of the KH-1 antigen has not been detected in normal colonic extracts, makes it an attractive target for active immunotherapy.

In selecting a route to the KH-1 antigen, we came to favor a plan that would build the linear hexasaccharide portion utilizing principles that emerged from the logic of glycal assembly (Scheme 3) [50, 51]. The hexasaccharide would be so differentiated in terms of its protecting group patterns as to allow for the simultaneous unveiling of the three hydroxyl groups which would serve as subsequent fucosylation acceptors. The featured asset of the synthesis would then be the three concurrent fucosylations to arrive at the fully intact antigen. In addition, the route allows for ready access to the allyl gly-



Scheme 3 Synthesis of the KH-1-KLH glycoconjugate

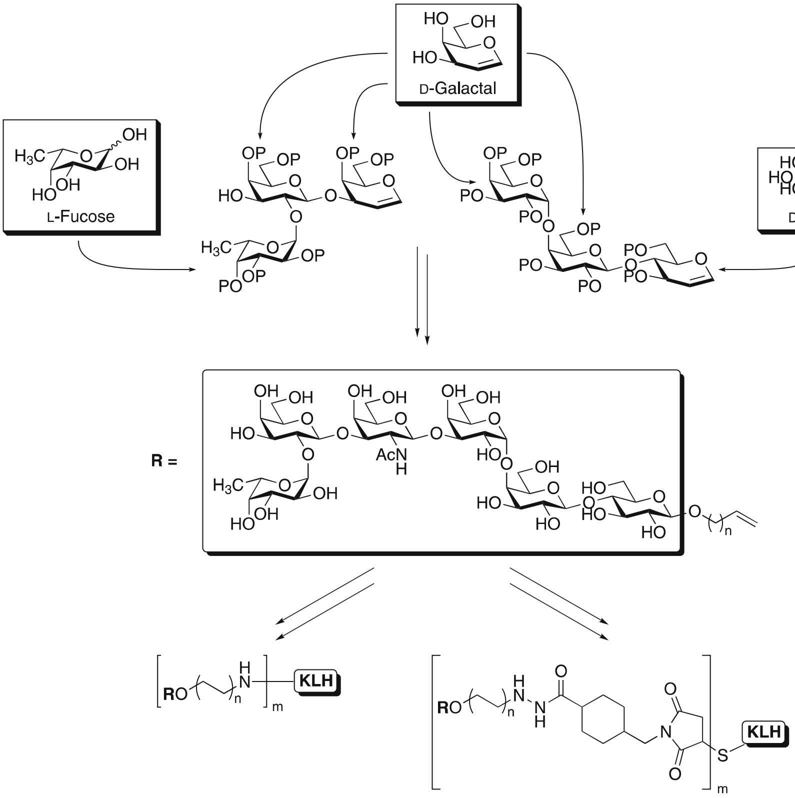
coside, which would, in the usual manner, serve as a handle for either direct conjugation to KLH or attachment of the MMCCH linker.

Immunological studies of both constructs in murine subjects were performed [55]. Upon vaccination with the KH-1-MMCCH-KLH construct, high titers of both IgM and IgG antibodies were obtained, while vaccination with the directly linked construct generated only IgM antibodies. Interestingly, antibodies elicited by both constructs were able to recognize not only the KH-1 antigen, but also the Le^y antigen. This result is not entirely surprising since the two antigens share the same four saccharides at their non-reducing end, and is indeed consistent with the well-established paradigm that anti-carbohydrate antibodies react primarily with the terminal residues of oligosaccharides [56]. Based on these studies, a Phase I clinical trial is being prepared with patients having epithelial cancers, using the KH-1-MMCCH-KLH vaccine construct along with the QS21 adjuvant.

2.2.4
Globo-H

Our most clinically advanced monomeric vaccine, Globo-H, has been actively studied as a vaccine for breast, prostate, and ovarian cancers. The Globo-H hexasaccharide was first isolated in sub-milligram quantities from human MCF-7 breast cancer cells using the monoclonal antibody MBr1 [57–59]. It has been shown to reside at the cell surface as a glycolipid and also possibly as a glycoprotein. Further immunohistological studies have determined that Globo-H is also expressed on a number of other carcinomas, including colon, lung, ovary, prostate, and small cell lung cancers [60]. There is evidence, though, that there are carbohydrates, assumed to be Globo-H, residing on the cell surface of normal breast, pancreas, small bowel and prostate tissues. However, the antigen is predominantly localized where access to the immune system is restricted [61]. Because of this, we felt that only a detailed investigation would reveal the usefulness of Globo-H as an antigen in mouse immunizations and whether it would be of value as a cancer treatment in an adjuvant setting.

In studying the structure of Globo-H with respect to a synthetic plan, we were sensitive to an important requirement. Not only would the program have to produce adequate quantities for immunocharacterization, conjugation, and mouse vaccinations, but the synthesis would be required to generate much larger amounts if the intent were to move into the clinic after positive serological studies. We came to favor a route in which we could further demonstrate the power of the glycal assembly method in the context of a convergent synthesis (Scheme 4) [62–64]. In the end, Globo-H was produced as either an allyl- or pentenyl glycoside in sufficient quantities as to provide material for in-depth studies in both pre-clinical and clinical settings. Stud-



Scheme 4 Synthesis of the Globo-H-KLH glycoconjugate

ies in mice using the Globo-H-KLH construct showed high titer IgM and IgG responses against the Globo-H antigen [65]. Furthermore, the antibodies reacted with Globo-H positive MCF-7 cells and not with Globo-H negative B78.2 cells, and were able to effectively induce complement-mediated cytotoxicity (48% lysis).

A full Phase I trial has been completed in patients with progressive and recurring prostate cancers [66–68]. All immunized patients exhibited good IgM responses against Globo-H, with the antibodies able to recognize Globo-H-expressing cell lines. In some cases, they were able to induce complement-mediated lysis. On the basis of these encouraging results, a Phase I trial was initiated in patients with metastatic breast cancer without evidence of disease or with stable disease on hormone therapy [69]. Again, vaccination was able to stimulate the production of IgM antibodies in a majority of patients. Significant binding of the antibodies to MCF-7 was observed in 60% of the patients (a tripling of the values between pre- and post therapy samples).

2.3
Clustered Vaccines

As a second-generation synthetic investigation in our program, we became interested in the development of methodology aimed at the preparation of vaccine constructs that would more closely mimic the cell surface of tumor cells. Initially we focused on the mucin related O-linked glycopeptides. Mucins comprise a family of large glycoproteins expressed on the surface of epithelial tissues, and carry large glycodomains in clustered modes [70]. The amino acid sequence of mucins possesses a relatively high percentage of serine and threonine residues, often arranged in continuous arrays ranging in number from two to five. Despite the diversity found in mucin glycostructures, the appearance of an *N*-acetylgalactosamine moiety (GalNAc) α-linked to the serine/threonine residue appears to be highly conserved. A relevant example of this is found in the glycophorin family of carbohydrate antigens (Fig. 3). The Tn, sialyl-Tn (STn), and Thomsen-Friedenreich disaccharide (TF) antigens are quite common in malignant carcinomas [71]. More than 80% of breast, prostate, and ovarian carcinomas express STn, whereas the expression levels of STn in normal tissues is much reduced and appears to be restricted to only a few epithelial tissues at secretory borders [44]. Other more complex members of the glycophorin family (2,3-STF, 2,6-STF, and glycophorin) have not been extensively studied, but represent an interesting group of carbohydrate antigens.

The over-expression of STn in various carcinomas has been correlated to an aggressive phenotype and worsened prognosis [72]. Murine immunization with STn has been shown to elicit anti-STn antibodies capable of protection from subsequent tumor challenge with syngeneic cancer cell lines expressing STn [73]. This finding is consistent with an expanding body of evidence

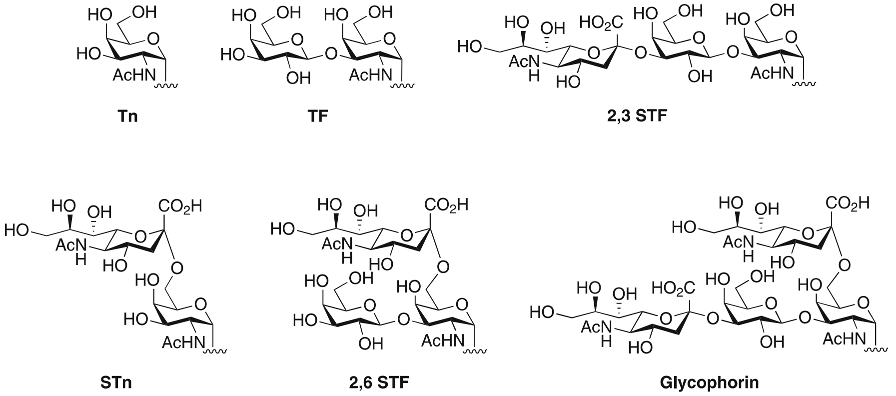
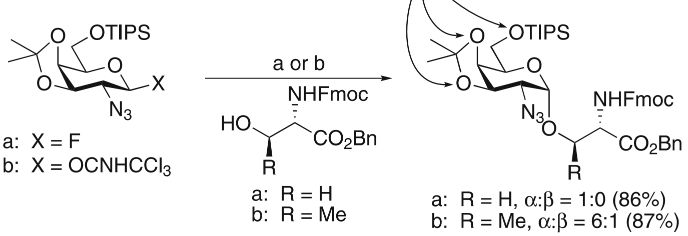


Fig. 3 The glycophorin family of α-O-linked antigens

suggesting that antibodies against defined tumor antigens are able to provide some protection against circulating tumor cells and micrometastases [22, 74]. Studies have shown that monoclonal antibodies against STn are able to recognize not only the STn monomer, but also STn clusters (STn(c)), indicating that it is identified in at least two distinct configurations at the tumor cell surface [75]. The STn(c) consisted of a linear tripeptide comprised of serine or threonine residues whose side chain hydroxy groups contained the carbohydrate antigen. Since the immune system tends to recognize clustered motifs of carbohydrate antigens [75, 76], we developed a clear interest in the ability to synthesize vaccine constructs that could take advantage of this phenomenon.

2.3.1
The Cassette Approach

At the outset of our investigations into the synthesis of clustered glycopeptide-based vaccine constructs, we identified a significant problem; the synthesis of carbohydrate domains O-linked to either serine or threonine with strong stereochemical control in the formation of α-glycosidic linkages. The glycopeptide assembly method, we came to favor is that of the "cassette" modality rather than a maximally convergent approach. In this approach, the basic building block, a GalNAc synthon, is stereospecifically α-linked to the side-chain oxygen of either serine, threonine or more recently, hydroxynorleucine (*vide infra*), with a differentiable acceptor site on the GalNAc (Scheme 5). This construct then serves as a general insert (cassette) that can be subsequently joined to an appropriately activated carbohydrate [77–81]. In this way, we obviate the need for direct coupling of the fully elaborated, complex glycosyl donor to the side chain of the desired amino acid. The clear advantage of this method is that the difficult O-linkage step is accomplished on a simple monosaccharide very early in the synthesis. The α-O-linked amino acid is, consequently, already in place as the oligosaccharide is being built.



Scheme 5 Synthesis of the cassette: (a) X = $OCNHCCl_3$, R = H, TMSOTf, THF, – 78 °C; (b) X = F, R = Me, Cp_2ZrCl_2, AgOTf, CH_2Cl_2

2.3.2
Monoantigenic Clustered Vaccines

Using the cassette methodology, the synthesis of a Tn clustered epitope was completed (Fig. 4) [82]. The Tn(c) was attached to a heterobifunctional linker, *m*-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS), and then conjugated to thiolated KLH, providing the desired construct. Preclinical ELISA studies demonstrated the ability of the Tn(c), when coadministered with QS21, to generate both IgM and IgG antibodies in mice after three immunizations [82, 83]. The cell surface activities of the antibodies were measured using Tn(c) positive LS-C and Tn(c) negative LS-B colon cancer cells and sera from the vaccinated mice showed clear IgM and IgG reactivity by flow cytometry and complement-dependent cytotoxicity assays. A Phase I human trial in prostate cancer patients has just been completed using the Tn(c) vaccine [84]. All patients immunized with Tn(c)-MBS-KLH developed good

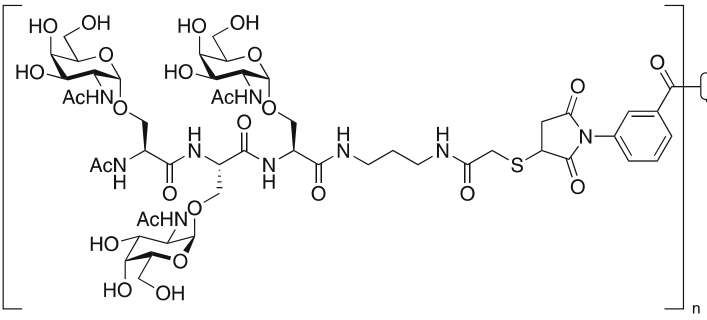


Fig. 4 Structure of the clustered Tn vaccine

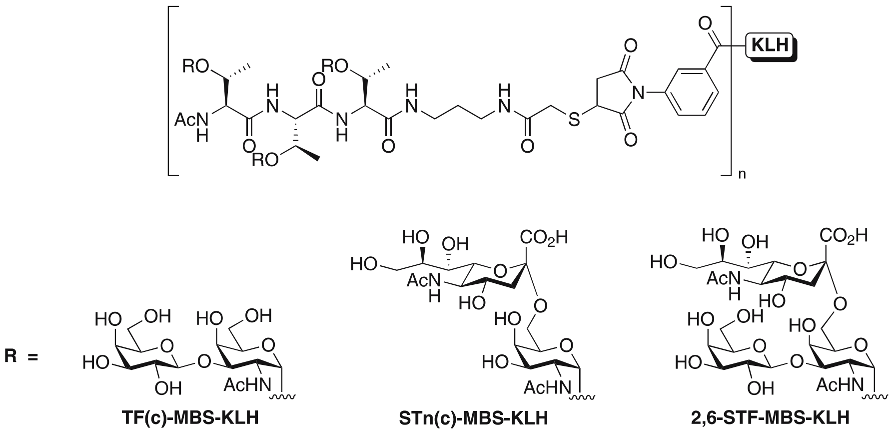


Fig. 5 Structures of other clustered glycophorin vaccines

antibody responses against Tn(c). In addition, the levels of prostate specific antigen (PSA) observed in the treated patients either stabilized or declined; the clinical impact and relevance of this remains to be validated.

Other glycophorin clusters have been prepared using the cassette method, including TF [82], STn [85], and 2,6-STF [86] (Fig. 5), as well as a Le^y cluster [87]. Conjugation and immunological evaluation of the TF(c)-MBS-KLH and STn(c)-MBH-KLH vaccines have produced similar results. Interestingly, when mice were vaccinated with the Le^y(c)-peptide-MBS-KLH conjugate, both IgM and IgG antibodies were produced, in contrast to the non-clustered vaccine (*vide supra*). However, the specificity of the response was limited to the immunizing epitope [88].

2.3.3
Polyantigenic Clustered Vaccines

All of the cancer vaccines described above were designed to target one antigen per vaccine. It has now been well established that cancer cells can display several different carbohydrate antigens at their cell surfaces [37, 44]. It is, in fact, entirely possible that during each stage of cellular development differential levels of antigenic expression may occur. Accordingly, a monoantigenic approach may not be sufficient for targeting a population of transformed cells. By contrast, a polyantigenic approach incorporating several type-specific carbohydrate antigens could well provide a heightened and more varied immune response, increasing the efficiency of binding to targeted cells.

There have been two distinctly different approaches taken in the search for multiantigenic vaccines. One approach relies on the construction of a polyantigenic vaccine from the admixture of several existing monoantigenic conjugate vaccines, in either clustered or non-clustered form [31, 89]. This idea has been used extensively in fields outside of cancer (e.g., bacterial vaccines) [90, 91]. Preclinical trials using polymolecular glycoconjugates have demonstrated the viability of this method. Using the conjugate vaccines GD3-KLH, Le^y-KLH, and the two peptidic antigens MUC1-KLH and MUC2-KLH, along with QS21, high titer IgM and IgG antibodies were produced that reacted specifically with the representative antigens in tumor cell lines expressing them [89]. This result was obtained regardless of administration method or injection site, demonstrating that the immunogenicity of the individual components was not diminished in the polyantigenic vaccine. On the basis of these results, a Phase II clinical trial featuring three to seven individual antigen-KLH conjugates (all of which have been proven safe and effective in Phase I trials) is being initiated in breast, ovarian, and prostate cancer patients.

Although this approach may seem attractive at the outset, a number of concerns must be addressed. First, an increase in the level of carrier protein used in the vaccine may have potential consequences [92, 93]. Additionally, in order to gain regulatory approval, one would have to demonstrate the validity of each individual component in the polyantigenic "mixture". There are further disadvantages from a synthetic perspective. In addition to its synthesis, each individual carbohydrate antigen would have to undergo a potentially low-yielding conjugation to the carrier protein. Perhaps a better alternative, and one we strongly prefer, would be a multiantigenic conjugate vaccine. In this case, the construct would consist of multiple carbohydrate antigens displayed on a single molecular entity that could then undergo a single conjugation step. Although the presumably low-yielding conjugation step is still required on the total construct, it would only need to be performed once for the entire vaccine. Moreover, such a consolidation would reduce the amount of carrier protein needed, minimizing the potential for carrier protein-induced immune suppression [94].

Early investigations into the requirements for optimal antigenicity led to the assembly of two glycopeptides, each containing three different carbohydrate antigens (Fig. 6). One construct contained the Tn, TF, and Le^y antigens linked to a peptide backbone through serine hydroxy groups, mimicking natural mucin-like architecture [95]. In this case the serine-GalNAc cassette was incorporated into each glycosylamino acid. The other construct consisted of the Tn, Le^y and Globo-H antigens attached to the peptide backbone through the hydroxy group of the non-natural amino acid hydroxynorleucine, or tris homoserine [96]. Preliminary studies of the two constructs indicate the hydroxynorleucine-based conjugate to be considerably more antigenic than the mucin-derived conjugate, and so this was selected for additional investigation in murine hosts. Further evaluation confirmed that both IgM and IgG antibodies were generated and that the antibod-

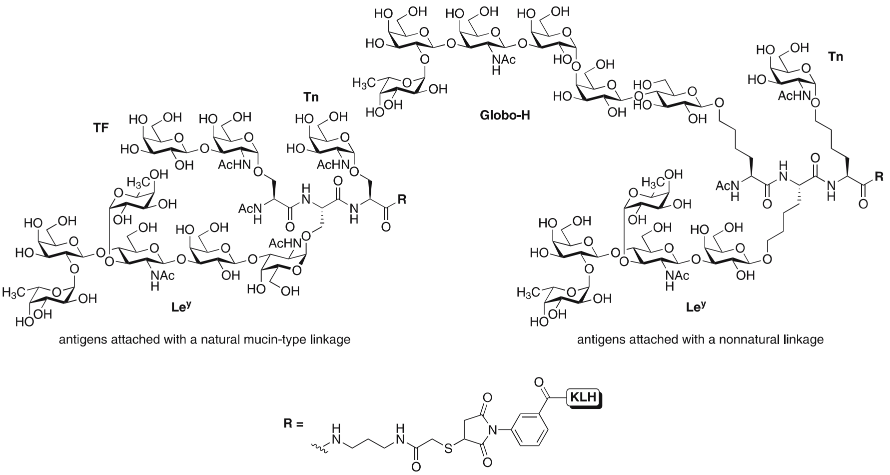


Fig. 6 Structures of two multiantigenic unimolecular vaccines

ies react selectively with cancer cells expressing the Tn, Le^y, and Globo-H antigens [97].

The results of these experiments suggest that the stimulation of a multifaceted immune response may be possible using a polyantigenic clustered vaccine. The development of a polyantigenic vaccine targeting prostate cancer provides a unique opportunity to explore the efficacy of such an approach as we have synthetic access to several of the carbohydrate antigens associated with it. Prostatic cancer tissue has been found to display the blood group related antigens Tn, STn, TF and Le^y, along with the gangliosides GM2 and Globo-H in abundance on the cell surface [37, 44]. Many of these antigens have been included in monoantigenic vaccines and have demonstrated success in early clinical trials (*vide supra*). Given this success, we have targeted a multiantigenic conjugate vaccine containing five different carbohydrate antigens (Globo-H, STn, Tn, TF, and Le^y) that could potentially be used to elicit an immune response directed specifically to prostate cancer. In addition to using KLH, we hoped to assess the efficacy of the macrolipid *N*-α-palmitoyl-*S*-[2,3-bis(palmitoyloxy)-(2*RS*)-propyl]-L-cysteine (Pam3Cys) [98] as an immuno carrier. Although KLH has been extensively studied, and provides a favorable immune response [99], the Pam3Cys lipid, a potent B cell stimulant, has only seen minimal preclinical and clinical evaluation [84, 88].

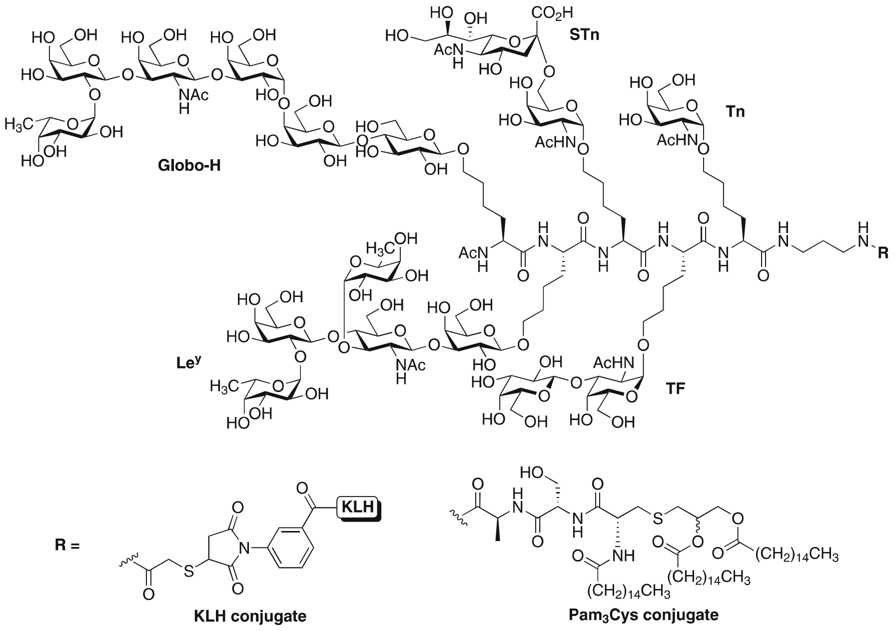


Fig. 7 Pentaantigenic vaccine designed for use in prostate cancer

The synthesis of the pentaantigenic vaccine was accomplished starting from a pool of glycosylamino acids in which the carbohydrate domain was completely protected and the amino terminus was blocked with the fluorenylmethyl carbamate (Fmoc) group (Fig. 7) [96, 100–102]. Utilizing solution-phase Fmoc-based peptide chemistry, we assembled the requisite glycopeptide in such a way as to allow for conjugation to either KLH or the Pam_3Cys lipid [103]. Evaluation of the two glycoconjugates is currently underway and will be reported in due course.

3
HIV/AIDS Vaccines

N-linked glycoproteins represent a class of compounds with significant biological importance. For instance, gp120 [104] is the N-linked envelope glycoprotein on the surface of the HIV virus. The extensive glycosylation of this protein leads to the formation of a "glycan shield" which is believed to play an integral role in the protection of the peptidic backbone against an immune response [105]. Recently however, it has been demonstrated that the monoclonal antibody 2G12, a general neutralizing anti-HIV antibody [106], is able to recognize specific structural motifs on gp120, i.e., a N-linked high-mannose or hybrid-type glycan attached to the Asn 332, 339 and 392 residues of gp120 [107, 108]. These findings pave the way, in principle, for the design of a new generation of HIV vaccines which use the virus' own defense mechanism against it, targeting the 2G12 binding motif of the highly glycosylated gp120.

The goal of our gp120 program is the construction of gp120-based glycopeptides that contain the 2G12 binding domain through total chemical synthesis [11, 12, 109]. Figure 8 illustrates our general synthetic strategy for the assembly of such elaborate targets. Through a series of glycosylation

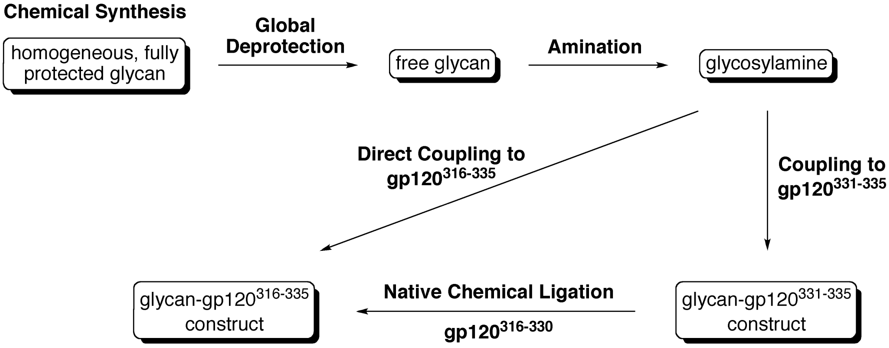
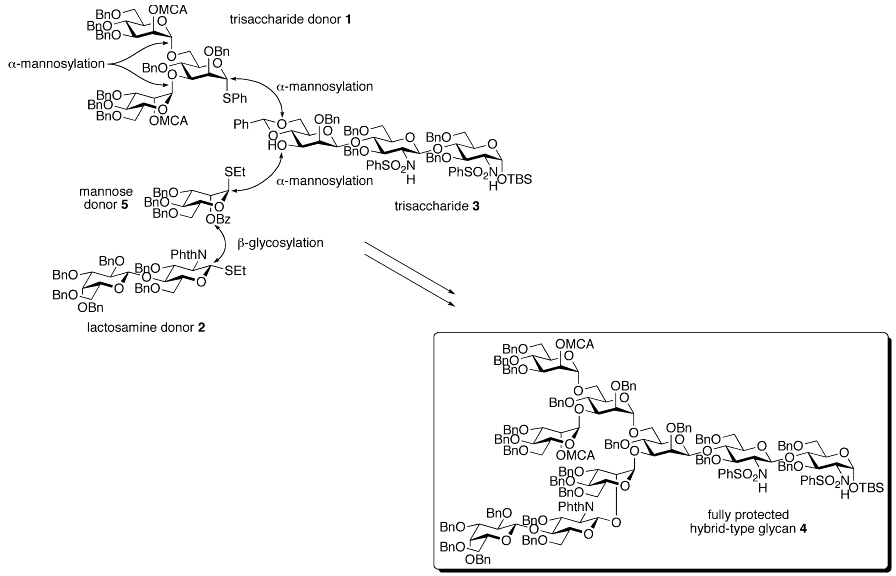


Fig. 8 Fully synthetic approach to glycan-gp120 peptide constructs

sequences, fully protected glycans could be constructed. Conversion of the anomeric hydroxy group to an amine is then accomplished following a global deprotection step, giving a fully deprotected glycosylamine. The carbohydrate would then be attached to Asn332 in a properly assembled gp120$^{331-335}$ peptide. In this setting we have chosen a linkage point of the carbohydrate proximal to a cysteine (Cys331), which could be used to attach a larger peptide chain *via* native chemical ligation (NCL) [110], giving the glycan-gp120$^{316-335}$ conjugate. This methodology has been used successfully in our previous reported synthesis of glycopeptides related to the prostate specific antigen [111, 112]. Alternatively, the glycosylamine could be directly attached to the full gp120$^{316-335}$ segment. The remainder of this review will focus on our synthetic efforts into the synthesis of both hybrid-type and high-mannose type gp120 fragments.

3.1
Hybrid-Type gp120 Fragments

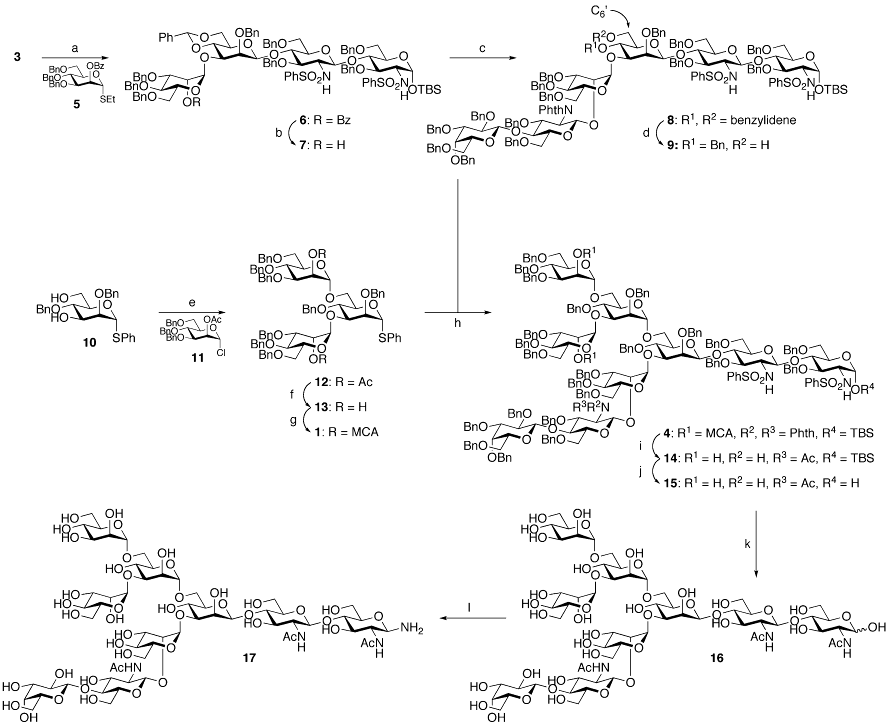
We first investigated the synthesis of a hybrid-type glycan, and began with the core β-mannose/chitobiose trisaccharide **3**, which already has the C3′and C6′ branching points distinguished (Scheme 6) [11, 109, 113]. The lower branching point (C3′) would be utilized first, attaching the mannose-lactosamine side chain ("D1 arm") through consecutive α- then β-glycosylations. The upper branching point (C6′) would then be introduced by way of an



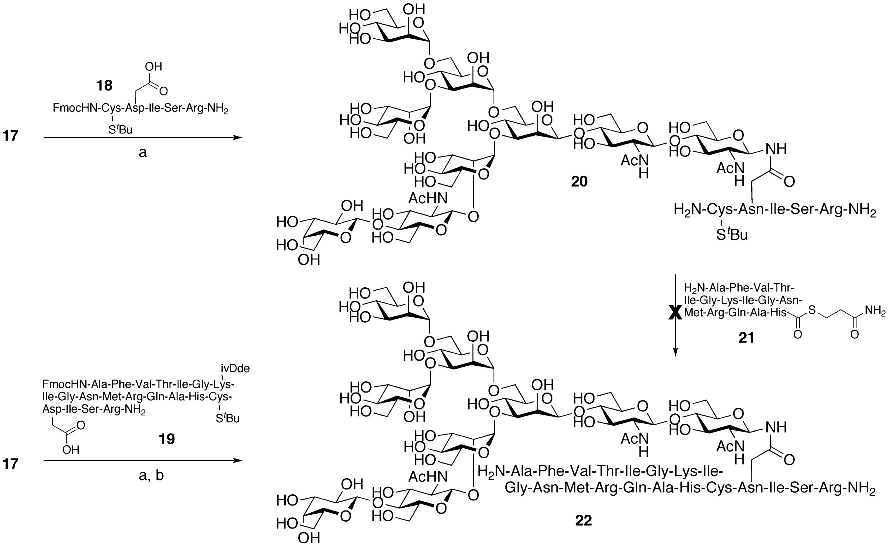
Scheme 6 Synthetic strategy for the preparation of a hybrid-type glycan

α-glycosylation using trisaccharide **1** as the donor. **1**, in turn, can be readily prepared by a bis α-mannosylation.

The synthesis commenced with trisaccharide **3** (Scheme 7). α-Mannosylation of **3** using donor **5** under Sinaÿ radical activation conditions [114] followed by debenzoylation provided alcohol **7**. Hexasaccharide **8** was then obtained upon coupling **7** with the lactosamine thio donor **2**. The "virtual" branching point on **8**, i.e., C'_6, was liberated by regioselective opening of the 4,6-benzylidene acetal, and then glycosylated with donor **1** to yield the fully protected hybrid-type nonasaccharide **4**. We now note the presence of a monochloroacetyl (MCA) protecting group on the axial hydroxyls of donor **1**. In earlier studies we had found that the use of the benzoyl protecting group at this position led to complications during final deprotection, i.e., hydrolysis was slow and led to concurrent removal



Scheme 7 Synthesis of hybrid glycan **17**: (a) **5**, $(BrC_6H_4)_3NSbCl_6$, MeCN, 80%; (b) NaOMe/MeOH, 89%; (c) **2**, $(BrC_6H_4)_3NSbCl_6$, MeCN, 60%; (d) $BH_3{\cdot}THF$, Bu_2BOTf, CH_2Cl_2, 75%; (e) **11**, AgOTf, DTBP, CH_2Cl_2; (f) NaOMe/MeOH, 65% (two steps); (g) $(ClCH_2CO)_2O$, py, CH_2Cl_2, 79%; (h) **1**, $(BrC_6H_4)_3NSbCl_6$, MeCN, 78%; (i) 1. $NH_2CH_2CH_2NH_2$, *n*-BuOH/toluene, 90 °C; 2. Ac_2O, py; 3. NaOMe/MeOH, 69%; (j) TBAF/AcOH, 89%; (k) 1. Na/NH_3, – 78 °C, 70%; 2. $NaHCO_3$, Ac_2O, 70%; (l) NH_4HCO_3/H_2O



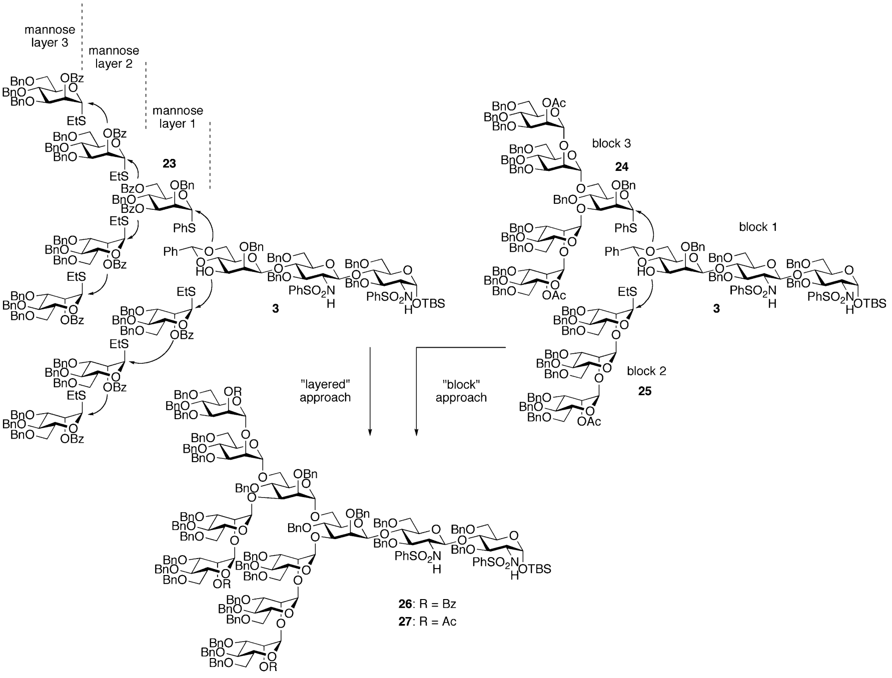
Scheme 8 Synthesis of the hybrid-glycan glycopeptide **22**. (a) 1. **18** or **19**, HATU, DIEA, DMSO; 2. NH_2NH_2 piperidine, DMF; (b) MES-Na, DMF, DIEA, H_2O, 95%

of the phthalimide. By employing MCA protection, selective cleavage was possible. In the event, the MCA protecting groups were quantitatively removed by the action of thiourea, followed by conversion of the phthalimide into the required acetamide function and liberation of the anomeric hydroxy site. Compound **15** was then subjected to global deprotection with sodium in liquid ammonia followed by acetylation of the free amines with acetic anhydride in saturated $NaHCO_3$ to afford glycan **16** bearing an anomeric hydroxy group. Amination of the anomeric position under Kochetkov conditions [115] led to **17**, which was used for peptide coupling without further protection [11].

Coupling of glycosylamine **17** with excess pentapeptide **18**, representing gp120$^{331-335}$, under previously developed conditions [111, 112, 116], followed by removal of the *N*-terminal Fmoc group led to glycopeptide **20**, our precursor for NCL (Scheme 8). Unfortunately, **20** was not a competent participant in NCL with peptide **21**, representing gp120$^{316-330}$. This failure could be attributed to steric hindrance resulting from the bulky nonasaccharide attached to asparagine directly adjacent to the ligation site. These findings illustrated potential limitations of NCL in a highly elaborate setting. In the end, the hybrid-type gp120$^{316-335}$ glycopeptide fragment could be obtained using a direct coupling method with peptide **19**, followed by final deprotection.

3.2
High-Mannose-Type gp120 Fragments

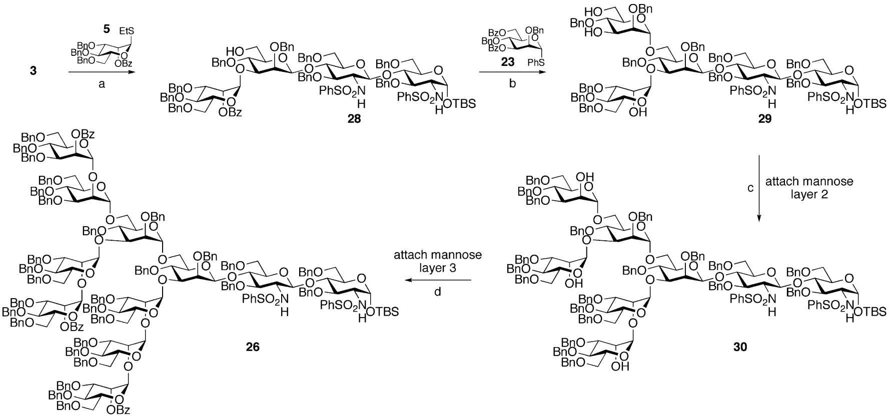
In parallel with the synthetic development of hybrid-type fragments, our efforts were also concentrated on the high-mannose-type constructs [12]. As shown in Scheme 9, we investigated two distinct approaches: the "layer" approach, featuring two consecutive tri-mannosylations, and the "block" approach, featuring a 3 + 3 glycosylation followed by a 6 + 5 glycosylation.



Scheme 9 Synthetic strategy for the preparation of a high-mannose-type glycan

3.2.1
The "Layer" Approach

The "layer" approach was investigated first, and began with the same core *β*-mannose/chitobiose trisaccharide **3** used during the synthesis of the hybrid-type glycan. Again *α*-mannosylation at C_3' followed by regioselective reduction of the benzylidene ring gave tetrasaccharide **28** (Scheme 10). A second *α*-mannosylation at the now revealed C_6' hydroxy group followed by benzoyl saponification exposed the three required acceptor sites of **29**. The next "layer" was added by glycosylation under Sinaÿ conditions [114] delivering octasaccharide **30** in respectable yield (55%). The process of



Scheme 10 The "layer" approach to assemble high-mannose-type glycan **26**. (a) 1. **5**, $(BrC_6H_4)_3NSbCl_6$, MeCN, 78%; 2. $BH_3 \cdot THF$, Bu_2BOTf, THF 0 °C, 90%; (b) 1. **23**, $(BrC_6H_4)_3NSbCl_6$, MeCN, 74%; 2. NaOMe/MeOH, 91%; (c) 1. **5**, $(BrC_6H_4)_3NSbCl_6$, MeCN, 55%; 2. NaOMe/MeOH, 84%; (d) **5**, $(BrC_6H_4)_3NSbCl_6$, MeCN, 51%

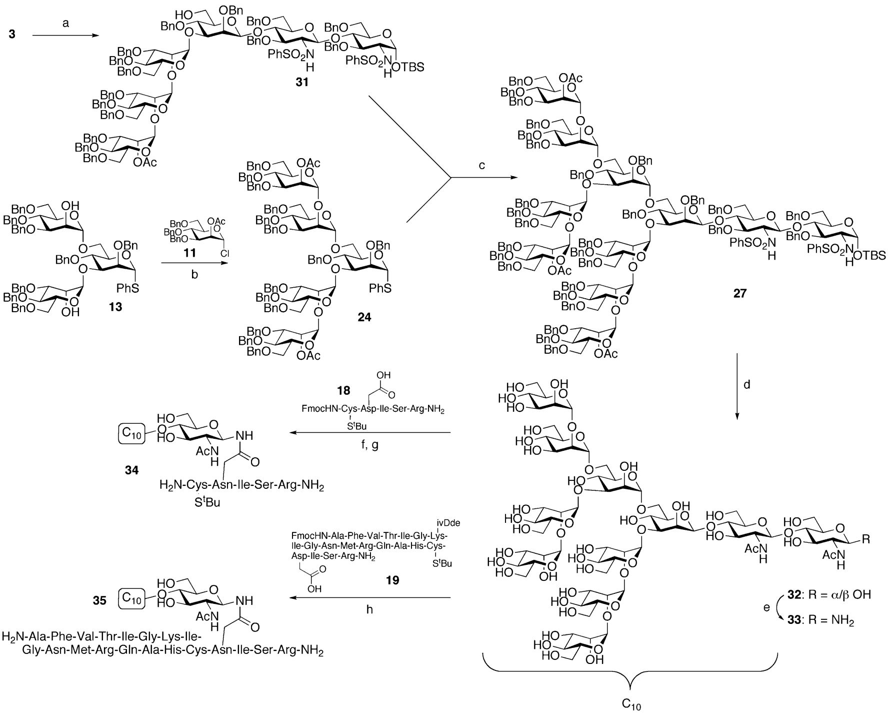
saponification/glycosylation was then repeated to furnish the desired fully protected undecasaccharide **26**.

3.2.2
The "Block" Approach

While the "layer" approach proved to be highly successful, we felt that **26** could be fashioned in an even more efficient manner. For this purpose, we explored a still more convergent "block" approach (Scheme 11) [12]. Hexasaccharide **31** was constructed from "blocks" **3** and **25** by a MeOTf-mediated glycosylation, followed by reductive opening of the benzylidene acetal. The third "block", pentasaccharide **24**, was assembled efficiently through two consecutive dimannosylation reactions beginning with phenylthiol mannoside **13** and chloro donor **11**. Pentasaccharide **24** was then subjected to a 6 + 5 glycosylation with acceptor **31**, delivering the intact fully protected high-mannose undecasaccharide **27** in 63% yield (85% based on recovered acceptor).

With the protected undecasaccharide in hand, we proceeded to the next phase, global deprotection. Similar to the strategy used in the hybrid-type glycan (*vide supra*), **27** was subjected to deacetylation, desilylation, and dissolving metal reduction to yield the free glycan, which was advanced to glycosylamine **33** by Kochetkov amination.

Following similar thinking, glycosylamine **33** was coupled to pentapeptide **18**, representing gp120$^{331-335}$ followed by Fmoc removal. Similar to the hybrid-type case, construct **34** was not a substrate for NCL. However, we were pleased to find that the direct coupling strategy employed earlier had also



Scheme 11 The "block" approach to assemble high-mannose-type glycopeptide **35**. (a) 1. **25**, MeOTf, DTBP, CH_2Cl_2, – 40 °C to r.t., 70%; 2. BH_3·THF, Bu_2BOTf, THF, 0 °C, 86%; (b) **11**, AgOTf, DTBP, CH_2Cl_2, – 10 °C to r.t. 87%; (c) $(BrC_6H_4)_3NSbCl_6$, CH_3CN, 63% (85% based on recovered **31**); (d) 1. NaOMe/MeOH, 96%; 2. TBAF, AcOH, THF, 0 °C, 98%; 3. Na/NH_3, – 78 °C; 4. Ac_2O, $NaHCO_3$, 87% (two steps); (e) NH_4HCO_3/H_2O; (f) 1. **18**, HATU, DIEA, DMSO; (g) piperidine, DMF 24% from **33**; (h) 1. **19**, HATU, DIEA, DMSO; 2. N_2H_4, piperidine, DMF, 16% from **33**

worked in this case, providing the high-mannose-type gp120$^{316-335}$ glycopeptide fragment **35**, after final side chain deprotection.

3.3
2G12 Binding Studies

With both hybrid-type and high-mannose type glycopeptide fragments in hand, we set out to evaluate their ability to bind the monoclonal antibody 2G12. While not necessarily establishing construct immunogenicity, such binding assays could provide valuable information regarding the structural requirements of a potential HIV vaccine.

The binding assays were conducted on a Biacore 3000 system, utilizing surface plasmon resonance (SPR) technology [117, 118]. The amine coupling method was used to immobilize 2G12 on a CM5 sensor chip, providing the active binding surface. In addition, a human IgG1 isotype-control antibody was also immobilized, allowing for a reference surface. A single injection of substrate is exposed sequentially to the reference surface and then the active surface, allowing for qualitative evaluation of the data. For comparative purposes, recombinant HIV-$1_{\text{JR-FL}}$ gp120 [119] was also tested.

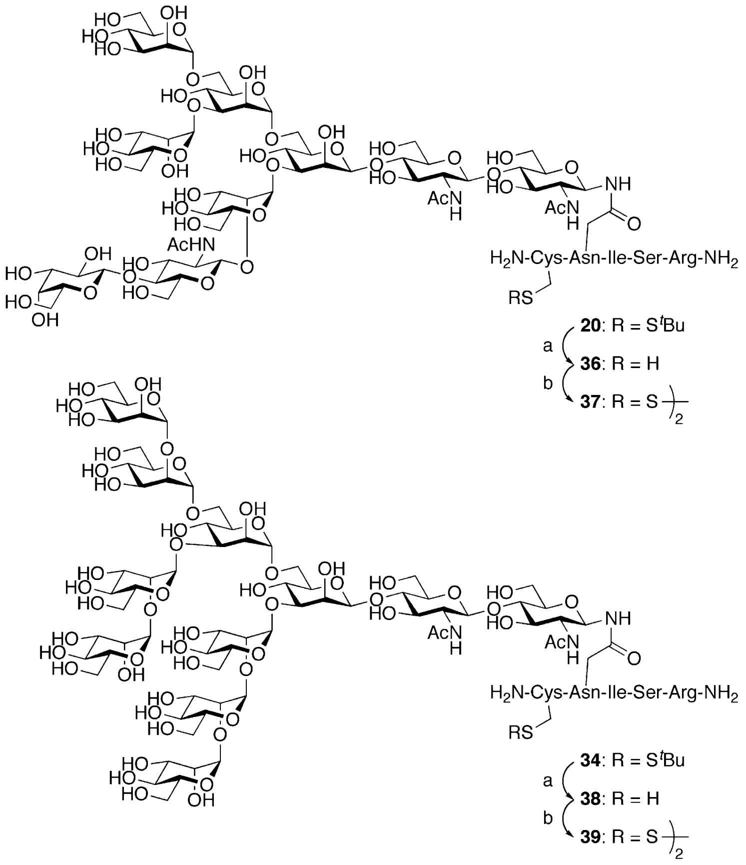
As shown in Table 1, neither the hybrid-type glycan **16**, the hybrid-type gp$120^{331-335}$ glycopeptide fragment **20**, nor the high-mannose-type glycan **32** displayed binding ability to 2G12 [109]. While the high-mannose glycan itself does not bind, the high-mannose-type gp$120^{331-335}$ glycopeptide fragment **38** displayed significant binding. Interestingly, when the sulfhydryl side chain of Cys^{331} was protected (**34**), binding was diminished. This was particularly puzzling, as it has been shown that the peptide motifs are not directly recognized by 2G12 [106].

As we examined the testing stock solution of **38** by liquid chromatography/mass spectroscopy (LC-MS), we discovered that the solution consisted of a mixture of the monomeric and oxidized disulfide forms, with a majority existing as the dimer (Scheme 12). Moreover, upon reduction of the disulfide, 2G12 binding was dramatically diminished. In control experiments, we confirmed that the 2G12 surface was still able to bind unreduced compound and that there were no detectable effects on the ability of 2G12 to bind gp120, suggesting that the dimeric form of the glycopeptide was responsible for the observed binding. This hypothesis was validated upon testing **39**, homogeneously formed by DMSO oxidation of **38**, as it exhibited the highest binding to 2G12. Of note is that dimeric hybrid-type gp$120^{331-335}$ construct **37** still did not demonstrate significant binding ability.

Table 1 Qualitative assessment of 2G12 binding to free glycan and glycan-gp120 glycopeptide[a]

Carbohydrate type	Compound (conc)	Cys SH state	Binding, RU
Hybrid-type	**16** (40 μM)	N/A	< 1
Hybrid-type	**20** (20 μM)	protected	< 1
Hybrid-type	**36** (20 μM)	free SH	< 1
Hybrid-type	**37** (20 μM)	dimer	< 1
High-mannose-type	**32** (40 μM)	N/A	< 1
High-mannose-type	**34** (20 μM)	protected	5
High-mannose-type	**38** (10 μM)	free SH*	75
High-mannose-type	**38** (10 μM) + DTT	free SH	9.5
High-mannose-type	**39** (10 μM)	dimer	78

[a] A single injection of substrate was exposed sequentially to the reference surface and then the active surface. Binding experiments were performed at 25 °C in HBS-P buffer (10 mM HEPES, pH 7.4, 150 mM NaCl, 0.005% Surfactant P20). The sensor surface was regenerated with a short pulse of 3.5 M MgCl_2. Recombinant HIV-$1_{\text{JR-FL}}$ gp120 was tested for comparison.
* In this case, it was later found that the thiol actually existed as an oxidized disulfide.



Scheme 12 Synthesis of the hybrid and high-mannose dimers. (a) MesNa, DMF, DIEA, H_2O; (b) DMSO, H_2O, air

While earlier findings revealed that both hybrid-type and high-mannose-type glycans may be part of the binding epitope on gp120, our experimental results suggest that the high-mannose-type is more likely to be recognized by 2G12. Our findings also support the notion that structural motifs featuring multiple glycans with proper spatial orientation might be crucial for 2G12 binding, which is consistent with the crystal structure results reported by Wilson [106].

4
Concluding Remarks

The generation of significant amounts of homogeneous carbohydrate antigens is no longer a stumbling block for the development of glycopeptide-based vaccines. It is clear that total synthesis has met the challenge of complex glycoconjugate construction and will continue to do so, regardless of the targeted structure. As such, the ability of researchers to probe the intricate structural requirements for an effective vaccine will only continue to grow. Although these studies are in their infancy and there is still much to learn, the necessary complex materials required for further preclinical and clinical studies are becoming increasingly available. It is hoped that through creative collaboration, advances in complex target oriented synthesis and immunology will allow for the development of novel vaccines to fight the diseases of tomorrow.

Acknowledgements The research reviewed in this manuscript was possible only through the dedication, enthusiasm, and creativity of those co-workers whose names are acknowledged on the publications cited from our laboratory. It is to these people, and to other members of the laboratory (who were constantly available for the extensive discussions that helped fashion this research) to whom the program is lastingly indebted. Financial support was largely provided by the National Institutes of Health, grants CA-103823 and CA-28824 (to S.J.D). We especially thank Mr. William H. Goodwin and Mrs. Alice Goodwin and the Commonwealth Cancer Foundation for Research for additional financial support. J.D.W. gratefully acknowledges the US Army Breast Cancer Research Foundation (DAMD17-03-1-0443) for a Postdoctoral Research Fellowship.

References

1. Haurowitz F (1979) Trends Biochem Sci 4:N268
2. Silverstein AM (1989) In: A History of Immunology, chap. 1 and 12. Academic Press, San Diego
3. Rosen FS, Mackay I (2001) New Eng J Med 345:1343
4. Livingston PO (1992) Curr Opin Immunol 4:624
5. Lanzavecchia A (1993) Science 260:937
6. Pardoll DM (1993) Curr Opin Immunol 5:719
7. Dranoff G, Jaffee E, Lazenby A, Golumbek P, Levitsky H, Brose K, Jackson V, Hamada H, Pardoll D, Mulligan RC (1993) Proc Natl Acad Sci USA 90:3539
8. Tao MH, Levy R (1993) Nature 362:755
9. Boon T (1993) Int J Cancer 54:177
10. Danishefsky SJ, Allen JR (2000) Angew Chem Int Ed Engl 39:836
11. Mandal M, Dudkin VY, Geng X, Danishefsky SJ (2004) Angew Chem Int Ed Engl 43:2557
12. Geng X, Dudkin VY, Mandal M, Danishefsky SJ (2004) Angew Chem Int Ed Engl 43:2562
13. Paulsen H (1982) Angew Chem Int Ed Engl 21:155
14. Schmidt RR (1991) In: Trost BM, Flemming I (eds) Comprehensive organic synthesis: selectivity, strategy and efficiency in modern organic chemistry, vol 6. Pergamon Press, Inc. p 33
15. Toshima K, Tatsuta K (1993) Chem Rev 93:1503
16. Danishefsky SJ, Bilodeau MT (1996) Angew Chem Int Ed Engl 35:1380
17. Boons GJ (1996) Contemp Org Synth 3:173
18. Boons GJ (1996) Tetrahedron 52:1095

19. Davis BG (2000) J Chem Soc, Perkin Trans 1:2137
20. Seeberger PH (2002) J Carbohydr Chem 21:613
21. Varki A (1999) In: Varki A, Cummings R, Esco J, Freeze H, Hart G, Marth J (eds) Essentials of Glycobiology. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, p. 537
22. Livingston PO, Zhang SL, Lloyd KO (1997) Cancer Immunol Immunother 45:1
23. Cobb Ba, Kasper DL (2005) Eur J Immunol 35:352
24. Beuvery EC, Vanrossum F, Nagel J (1982) Infect Immun 37:15
25. Barrett DJ, Ammann AJ, Stenmark S, Wara DW (1980) Infect Immun 27:411
26. Kasper DL, Paoletti LC, Wessels MR, Guttormsen HK, Carey VJ, Jennings HJ, Baker CJ (1996) J Clin Invest 98:2308
27. Keding SJ, Danishefsky SJ (2003) In: Wong CH (ed) Carbohydrate-Based Drug Discovery, vol 1. Wiley, Weinheim, Germany, p 381
28. Stroud MR, Levery SB, Nudelman ED, Salyan ME, Towell JA, Roberts CE, Watanabe M, Hakomori S (1991) J Biol Chem 266:8439
29. Stroud MR, Levery SB, Martensson S, Salyan ME, Clausen H, Hakomori S (1994) Biochemistry 33:10672
30. Yuen CT, Bezouska K, Obrien J, Stoll M, Lemoine R, Lubineau A, Kiso M, Hasegawa A, Bockovich NJ, Nicolaou KC, Feizi T (1994) J Biol Chem 269:1595
31. Ragupathi G, Koide F, Sathyan N, Kagan E, Spassova M, Bornmann W, Gregor P, Reis CA, Clausen H, Danishefsky SJ, Livingston PO (2003) Cancer Immunol Immunother 52:608
32. Nilsson O, Mansson J-E, Brezicka T, Holmgren J, Lindholm L, Sorensen S, Yngvason F, Szennerholm L (1984) Glycoconjugate J 1:43
33. Nilsson O, Brezicka FT, Holmgren J, Sorenson S, Svennerholm L, Yngvason F, Lindholm L (1986) Cancer Res 46:1403
34. Brezicka FT, Olling S, Nilsson O, Bergh J, Holmgren J, Sorenson S, Yngvason F, Lindholm L (1989) Cancer Res 49:1300
35. Vangsted AJ, Clausen H, Kjeldsen TB, White T, Sweeney B, Hakomori S, Drivsholm L, Zeuthen J (1991) Cancer Res 51:2879
36. Brezicka T, Bergman B, Olling S, Fredman P (2000) Lung Cancer 28:29
37. Zhang SL, CordonCardo C, Zhang HS, Reuter VE, Adluri S, Hamilton WB, Lloyd KO, Livingston PO (1997) Int J Cancer 73:42
38. Markl J, Lieb B, Gebauer W, Altenhein B, Meissner U, Harris JR (2001) J Cancer Res Clin Oncol 127 Suppl 2:R3
39. Dickler MN, Ragupathi G, Liu NX, Musselli C, Martino DJ, Miller VA, Kris MG, Brezicka FT, Livingston PO, Grant SC (1999) Clin Cancer Res 5:2773
40. Allen JR, Danishefsky SJ (1999) J Am Chem Soc 121:10875
41. Krug LM, Ragupathi G, Hood C, Kris MG, Miller VA, Allen JR, Keding SJ, Danishefsky SJ, Gomez J, Tyson L, Pizzo B, Baez V, Livingston PO (2004) Clin Cancer Res 10:6094
42. Hellström I, Garrigues HJ, Garrigues U, Hellström KE (1990) Cancer Res 50:2183
43. Federici MF, Kudryashov V, Saigo PE, Finstad CL, Lloyd KO (1999) Int J Cancer 81:193
44. Zhang SL, Zhang HS, CordonCardo C, Reuter VE, Singhal AK, Lloyd KO, Livingston PO (1997) Int J Cancer 73:50
45. Behar V, Danishefsky SJ (1994) Angew Chem Int Ed Engl 33:1468
46. Danishefsky SJ, Behar V, Randolph JT, Lloyd KO (1995) J Am Chem Soc 117:5701
47. Kudryashov V, Kim HM, Ragupathi G, Danishefsky SJ, Livingston PO, Lloyd KO (1998) Cancer Immunol Immunother 45:281

48. Sabbatini P, Kudryashov V, Ragupathi G, Danishefsky SJ, Livingston PO, Bornmann W, Spassova M, Zatorski A, Spriggs D, Aghajanian C, Soignet S, Peyton M, O'Flaherty C, Curtin J, Lloyd KO (2000) Int J Cancer 87:79
49. Nudelman E, Levery SB, Kaizu T, Hakomori S (1986) J Biol Chem 261:11247
50. Deshpande PP, Danishefsky SJ (1997) Nature 387:164
51. Deshpande PP, Kim HM, Zatorski A, Park TK, Ragupathi G, Livingston PO, Live D, Danishefsky SJ (1998) J Am Chem Soc 120:1600
52. Spassova MK, Bornmann WG, Ragupathi G, Sukenick G, Livingston PO, Danishefsky SJ (2005) J Org Chem 70:3383
53. Kaizu T, Levery SB, Nudelman E, Stenkamp RE, Hakomori S (1986) J Biol Chem 261:11254
54. Kim YS, Yuan M, Itzkowitz SH, Sun QB, Kaizu T, Palekar A, Trump BF, Hakomori S (1986) Cancer Res 46:5985
55. Ragupathi G, Deshpande PP, Coltart DM, Kim HM, Williams LJ, Danishefsky SJ, Livingston PO (2002) Int J Cancer 99:207
56. Kabat EA (1976) In: Structural Concepts in Immunology and Immunochemistry. Holt, Rinehart and Winston, New York, p 119
57. Kannagi R, Levery SB, Ishigami F, Hakomori S, Shevinsky LH, Knowles BB, Solter D (1983) J Biol Chem 258:8934
58. Bremer EG, Levery SB, Sonnino S, Ghidoni R, Canevari S, Kannagi R, Hakomori S (1984) J Biol Chem 259:14773
59. Menard S, Tagliabue E, Canevari S, Fossati G, Colnaghi MI (1983) Cancer Res 43:1295
60. Livingston PO (1995) Semin Cancer Biol 6:357
61. Zhang SL, Zhang HS, Reuter VE, Slovin SF, Scher HI, Livingston PO (1998) Clin Cancer Res 4:295
62. Bilodeau MT, Park TK, Hu SH, Randolph JT, Danishefsky SJ, Livingston PO, Zhang SL (1995) J Am Chem Soc 117:7840
63. Park TK, Kim IJ, Hu SH, Bilodeau MT, Randolph JT, Kwon O, Danishefsky SJ (1996) J Am Chem Soc 118:11488
64. Allen JR, Allen JG, Zhang XF, Williams LJ, Zatorski A, Ragupathi G, Livingston PO, Danishefsky SJ (2000) Chem Eur J 6:1366
65. Ragupathi G, Park TK, Zhang SL, Kim IJ, Graber L, Adluri S, Lloyd KO, Danishefsky SJ, Livingston PO (1997) Angew Chem Int Ed Engl 36:125
66. Ragupathi G, Slovin SF, Adluri S, Sames D, Kim IJ, Kim HM, Spassova M, Bornmann WG, Lloyd KO, Scher HI, Livingston PO, Danishefsky SJ (1999) Angew Chem Int Ed Engl 38:563
67. Slovin SF, Ragupathi G, Adluri S, Ungers G, Terry K, Kim S, Spassova M, Bornmann WG, Fazzari M, Dantis L, Olkiewicz K, Lloyd KO, Livingston PO, Danishefsky SJ, Scher HI (1999) Proc Natl Acad Sci USA 96:5710
68. Wang ZG, Williams LJ, Zhang XF, Zatorski A, Kudryashov V, Ragupathi G, Spassova M, Bornmann W, Slovin SF, Scher HI, Livingston PO, Lloyd KO, Danishefsky SJ (2000) Proc Natl Acad Sci USA 97:2719
69. Gilewski T, Ragupathi G, Bhuta S, Williams LJ, Musselli C, Zhang XF, Bornmann WG, Spassova M, Bencsath KP, Panageas KS, Chin J, Hudis CA, Norton L, Houghton AN, Livingston PO, Danishefsky SJ (2001) Proc Natl Acad Sci USA 98:3270
70. Carlstedt I, Davies JR (1997) Biochem Soc Trans 25:214
71. Springer GF (1984) Science 224:1198
72. Werther JL, Riveramacmurray S, Bruckner H, Tatematsu M, Itzkowitz SH (1994) Brit J Cancer 69:613
73. Fung PY, Madej M, Koganty RR, Longenecker BM (1990) Cancer Res 50:4308

74. Livingston PO, Ragupathi G (1997) Cancer Immunol Immunother 45:10
75. Zhang SL, Walberg LA, Ogata S, Itzkowitz SH, Koganty RR, Reddish M, Gandhi SS, Longenecker BM, Lloyd KO, Livingston PO (1995) Cancer Res 55:3364
76. Nakada H, Inoue M, Numata Y, Tanaka N, Funakoshi I, Fukui S, Mellors A, Yamashina I (1993) Proc Natl Acad Sci USA 90:2495
77. Nakahara Y, Iijima H, Ogawa T (1994) Tetrahedron Lett 35:3321
78. Liebe B, Kunz H (1994) Tetrahedron Lett 35:8777
79. Meinjohanns E, Meldal M, Schleyer A, Paulsen H, Bock K (1996) J Chem Soc, Perkin Trans 1:985
80. Mathieux N, Paulsen H, Meldal M, Bock K (1997) J Chem Soc, Perkin Trans 1:2359
81. Chen XT, Sames D, Danishefsky SJ (1998) J Am Chem Soc 120:7760
82. Kuduk SD, Schwarz JB, Chen XT, Glunz PW, Sames D, Ragupathi G, Livingston PO, Danishefsky SJ (1998) J Am Chem Soc 120:12474
83. Kagan E, Ragupathi G, Yi SS, Reis CA, Gildersleeve J, Kahne D, Clausen H, Danishefsky SJ, Livingston PO (2005) Cancer Immunol Immunother 54:424
84. Slovin SF, Ragupathi G, Musselli C, Olkiewicz K, Verbel D, Kuduk SD, Schwarz JB, Sames D, Danishefsky S, Livingston PO, Scher HI (2003) J Clin Oncol 21:4292
85. Schwarz JB, Kuduk SD, Chen XT, Sames D, Glunz PW, Danishefsky SJ (1999) J Am Chem Soc 121:2662
86. Sames D, Chen XT, Danishefsky SJ (1997) Nature 389:587
87. Glunz PW, Hintermann S, Williams LJ, Schwarz JB, Kuduk SD, Kudryashov V, Lloyd KO, Danishefsky SJ (2000) J Am Chem Soc 122:7273
88. Kudryashov V, Glunz PW, Williams LJ, Hintermann S, Danishefsky SJ, Lloyd KO (2001) Proc Natl Acad Sci USA 98:3264
89. Ragupathi G, Cappello S, Yi SS, Canter D, Spassova M, Bornmann WG, Danishefsky SJ, Livingston PO (2002) Vaccine 20:1030
90. Lee CJ, Lee LH, Koizumi K (2002) Infect Med 19:127
91. Lee CJ, Lee LH, Koizumi K (2002) Infect Med 19:179
92. Dagan R, Eskola J, Leclerc C, Leroy O (1998) Infect Immun 66:2093
93. Fattom A, Cho YH, Chu C, Fuller S, Fries L, Naso R (1999) Vaccine 17:126
94. Herzenberg LA, Tokuhisa T, Herzenberg LA (1980) Nature 285:664
95. Williams LJ, Harris CR, Glunz PW, Danishefsky SJ (2000) Tetrahedron Lett 41:9505
96. Allen JR, Harris CR, Danishefsky SJ (2001) J Am Chem Soc 123:1890
97. Ragupathi G, Coltart DM, Williams LJ, Koide F, Kagan E, Allen J, Harris C, Glunz PW, Livingston PO, Danishefsky SJ (2002) Proc Natl Acad Sci USA 99:13699
98. Bessler WG, Cox M, Lex A, Suhr B, Wiesmuller KH, Jung G (1985) J Immunol 135:1900
99. Musselli C, Livingston PO, Ragupathi G (2001) J Cancer Res Clin Oncol 127:R20
100. Biswas K, Coltart DM, Danishefsky SJ (2002) Tetrahedron Lett 43:6107
101. Keding SJ, Atsushi E, Biswas K, Zatorski A, Coltart DM, Danishefsky SJ (2003) Tetrahedron Lett 44:3413
102. Keding SJ, Endo A, Danishefsky SJ (2003) Tetrahedron 59:7023
103. Keding SJ, Danishefsky SJ (2004) Proc Natl Acad Sci USA 101:11937
104. Burton DR (1997) Proc Natl Acad Sci USA 94:10018
105. Wei XP, Decker JM, Wang SY, Hui HX, Kappes JC, Wu XY, Salazar-Gonzalez JF, Salazar MG, Kilby JM, Saag MS, Komarova NL, Nowak MA, Hahn BH, Kwong PD, Shaw GM (2003) Nature 422:307
106. Calarese DA, Scanlan CN, Zwick MB, Deechongkit S, Mimura Y, Kunert R, Zhu P, Wormald MR, Stanfield RL, Roux KH, Kelly JW, Rudd PM, Dwek RA, Katinger H, Burton DR, Wilson IA (2003) Science 300:2065

107. Scanlan CN, Pantophlet R, Wormald MR, Saphire EO, Stanfield R, Wilson IA, Katinger H, Dwek RA, Rudd PM, Burton DR (2002) J Virol 76:7306
108. Sanders RW, Venturi M, Schiffner L, Kalyanaraman R, Katinger H, Lloyd KO, Kwong PD, Moore JP (2002) J Virol 76:7293
109. Dudkin VY, Orlova M, Geng X, Mandal M, Olson WC, Danishefsky SJ (2004) J Am Chem Soc 126:9560
110. Dawson PE, Muir TW, Clark-Lewis I, Kent SBH (1994) Science 266:776
111. Miller JS, Dudkin VY, Lyon GJ, Muir TW, Danishefsky SJ (2003) Angew Chem Int Ed Engl 42:431
112. Dudkin VY, Miller JS, Danishefsky SJ (2004) J Am Chem Soc 126:736
113. Dudkin VY, Miller JS, Danishefsky SJ (2003) Tetrahedron Lett 44:1791
114. Zhang YM, Mallet JM, Sinaÿ P (1992) Carbohydr Res 236:73
115. Likhosherstov LM, Novikova OS, Derevitskaja VA, Kochetkov NK (1986) Carbohydr Res 146:C1
116. Cohen-Anisfeld ST, Lansbury PT (1993) J Am Chem Soc 115:10531
117. Rich RL, Myszka DG (2000) Curr Opin Biotech 11:54
118. Rich RL, Day YSN, Morton TA, Myszka DG (2001) Anal Biochem 296:197
119. Trkola A, Dragic T, Arthos J, Binley JM, Olson WC, Allaway GP, ChengMayer C, Robinson J, Maddon PJ, Moore JP (1996) Nature 384:184

Top Curr Chem (2007) 267: 143–185
DOI 10.1007/128_041

Synopsis of Structural, Biosynthetic, and Chemical Aspects of Glycopeptide Antibiotics

Falko Wolter · Sebastian Schoof · Roderich D. Süssmuth (✉)

Institut für Chemie/FG Organische Chemie, Technische Universität Berlin,
Straße des 17. Juni 124, 10623 Berlin, Germany
suessmuth@chem.tu-berlin.de

Abstract Glycopeptide antibiotics represent a very important group of natural products with regard to medicinal application as antibacterials and for cancer treatment. A characteristic for assignment as a glycopeptide antibiotic, next to antibiotic potential, is the modification of the peptide aglycon with various types of carbohydrates. According to this definition, vancomycin (including five subtypes), ramoplanin, bleomycin, mannopeptimycins and salmochelin belong to the family of glycopeptide antibiotics. In this article, an overview is given on the structural aspects, the biosynthesis and the mode of action of glycopeptide antibiotics. Structure–activity relationships of important functional groups contributing to enhanced antibiotic properties are highlighted. Finally, similarities with regard to structural features, the biosyntheses and the modes of actions are discussed. According to current knowledge, glycopeptide antibiotics mostly act as inhibitors of bacterial cell wall biosynthesis. It is to be expected that the actual list of glycopeptides will be continuously extended due to discoveries from ongoing screening programs.

Keywords Bleomycin · Mannopeptimycin · Ramoplanin · Salmochelin · Vancomycin

Abbreviations

ACP	Acyl carrier protein
BLM	Bleomycin
Chp	3-Chloro-4-hydroxyphenylglycine
Dab	2,4-Diaminobutyric acid
Dhb	2,3-Dihydroxybenzoic acid
Dpg	3,5-Dihydroxyphenylglycine
Dpr	2,3-Diaminopropionic acid
FTICR	Fourier transform ion cyclotron resonance
Hpg	4-Hydroxyphenylglycine
Hty	β-Hydroxytyrosine
Mephe	β-Methylphenylalanine
MIC	Minimum inhibitory concentration
MPM	Mannopeptimycin
MRSA	Methicillin-resistant *Staphylococcus aureus*
NRPS	Non-ribosomal peptide synthetase
orf	Open reading frame
PBA	Pyrimidoblamic acid
PBP	Penicillin binding protein
PKS	Polyketide synthase
SAR	Structure–activity relationship
VRE	Vancomycin-resistant enterococci
VRSA	Vancomycin-resistant *Staphylococcus aureus*

1
Introduction

According to the IUPAC convention, a glycopeptide is defined as a compound consisting of a carbohydrate linked to an oligopeptide composed of L- and/or D-amino acids [1]. However, for glycopeptide antibiotics this definition has to be sharpened towards glycosylated secondary metabolites of bacteria and fungi, synthesized by non-ribosomal peptide synthetases (NRPS), with antibiotic properties in the most common sense. Unfortunately, the term glycopeptide antibiotic often solely refers to vancomycin and its analogues and thus excludes other compounds that also belong to the family of glycopeptide antibiotics.

Until now, there has existed a relatively limited but important number of structurally diverse glycopeptide antibiotic classes. Probably the biggest class of glycopeptide antibiotics is represented by the vancomycin group. Among these, the most prominent representative vancomycin (type I) is listed together with avoparcin (type II), ristocetin (type III), teicoplanin (type IV) and complestatin (type V). Except for the complestatin-type, which shows antiviral properties, types I–IV are all active against Gram-positive bacteria. The group of vancomycin-related glycopeptides is followed by bleomycins (BLMs), which have been known for decades in cancer therapy. More recently, the ensemble of glycopeptides has been extended by the highly potent antibacterial ramoplanin. The mannopeptimycins represent the smallest and least-known group of glycopeptides. Although they were discovered in the 1950s, they have only recently been re-examined for their antibacterial activities. An outsider in this group is the recently discovered salmochelin. It is produced by *E. coli* and *Salmonella* strains and most likely represents an important pathogenic factor in salmonellosis. Although, at first glance, structurally the salmochelins are not considered to be peptides rather than cyclic esters they are closely related to glycopeptide antibiotics with regard to their biosynthetic origin. After structural features, the nature and mechanisms of biosynthetic assembly, which is non-ribosomal peptide synthesis, unifies all the glycopeptides mentioned above.

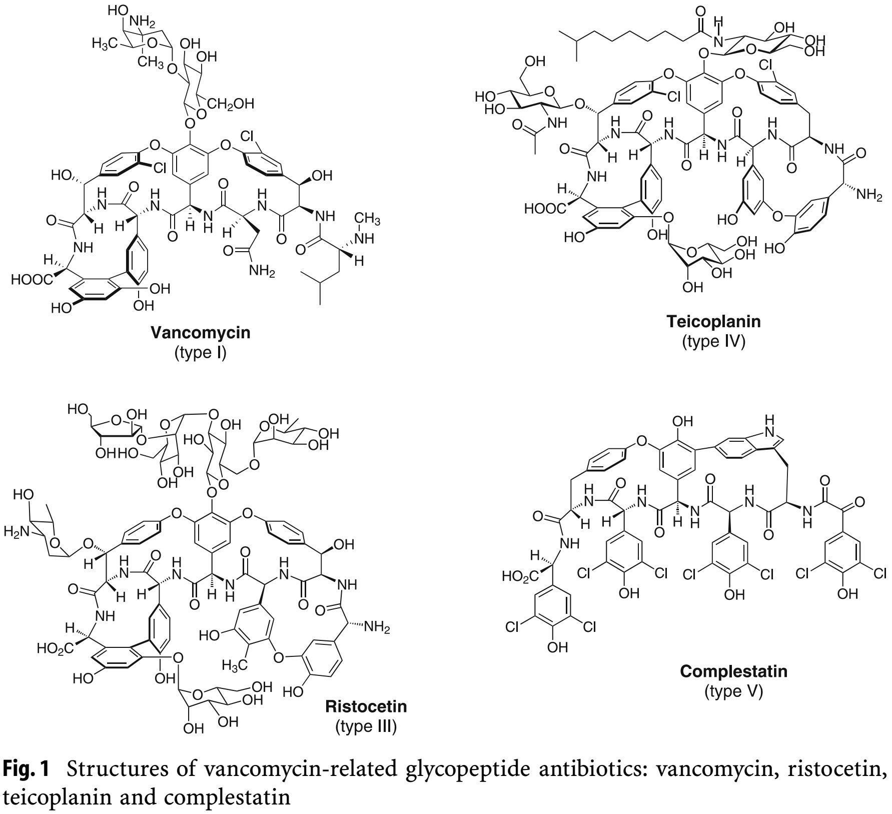
Non-ribosomal peptide synthetases (NRPS) are widespread in the assembly of peptidic secondary metabolites [2]. They share a common architecture, with each module being responsible for the selection, activation and condensation of one specific amino acid. A representative module of a biosynthetic gene cluster contains the characteristic C-A-T domain arrangement (with the exception of the first module, which bears only A-T domains). The A-domain specifically recognizes an amino acid and converts it to aminoacyl-AMP with consumption of ATP. The phosphopantetheine moiety of the T-domain bearing a free thiol group then forms a covalent thioester with the activated amino acid. Finally, the *C*-domain catalyses the formation of the peptide bonds between the amino acids. Thus, the peptide chain grows from the *N*-terminus to the *C*-terminus to yield a peptide. Additionally, epimerase domains (E), methylation domains (M), cyclization (Cy) and oxidation (Ox) domains provide additional structural diversity to NRPS biosynthetic assembly.

In this article, selected aspects of the investigation of glycopeptide antibiotic biosyntheses, the modes of action, and the development of novel derivatives are presented. The synopsis of glycopeptides highlights similarities and differences with regard to the features of glycopeptide antibiotics.

2
Vancomycin and Related Glycopeptide Antibiotics

2.1
Structure

The discovery of vancomycin (Fig. 1) dates back to the mid-1950s when it was isolated from culture broth of *Streptomyces orientalis* (reclassified as *Amycolatopsis orientalis*) [3]. Vancomycin proved an effective antibacterial agent and soon thereafter it was used clinically for the treatment of infection by Gram-positive bacteria. However, due to vancomycin's toxic side effects, other antibiotics like methicillin first gained increasing popularity in medicinal application. Later it was discovered that these side effects arose from impurities in the preparation. Since 1986 highly purified vancomycin has been available and nowadays vancomycin, together with teicoplanin (Europe) (Fig. 1), serve as antibiotics of last resort in the treatment of severe noso-



Vancomycin
(type I)

Teicoplanin
(type IV)

Ristocetin
(type III)

Complestatin
(type V)

Fig. 1 Structures of vancomycin-related glycopeptide antibiotics: vancomycin, ristocetin, teicoplanin and complestatin comial infections with enterococci and methicillin-resistant *Staphylococcus aureus* (MRSA) strains.

It took several decades to resolve the structure of vancomycin by means of X-ray crystallography and nuclear magnetic resonance (NMR) studies [4–6]. Total synthesis, a prime example of the art of peptide synthesis, was achieved some 20 years later by the groups of Nicolaou [7–9], Evans [10] and Boger [11]. Comprehensive reviews on the strategies for the total synthesis of vancomycin have been published recently [12, 13]. Meanwhile, total synthesis of teicoplanin has also been achieved by the Boger group [14].

According to a commonly accepted classification, the family of glycopeptide antibiotics is subdivided into five different types (type I–V), four of which have antimicrobial activities (type I–IV). Apart from the rather diverse composition of the sugar moieties and attachment sites of chlorine they all share a common heptapeptide backbone with three characteristic ring systems in the aromatic side chains: ^{1}Xxx-D-^{2}Tyr/Hty-^{3}Yyy-D-^{4}Hpg-D-^{5}Hpg-L-^{6}Hty-L-^{7}Dpg, with Xxx and Yyy as randomized amino acids among different subtypes. While type I glycopeptide antibiotics (e.g. vancomycin, balhiymcin) have aliphatic amino acids in these positions, type II glycopeptides (e.g. actinoidin) possess aromatic residues. Furthermore, in type III glycopeptides (e.g. ristocetin) these aromatic residues form an additional ether-bridged ring system. Teicoplanin is an example of type IV glycopeptides that share a general architecture with type III, however, in addition they possess a fatty acid attached to the sugar moiety. Therefore teicoplanin-type glycopeptides are sometimes referred to as lipoglycopeptides. Type V glycopeptides (e.g. complestatin) have only two ring systems and hold ^{2}Trp in the backbone structure. Interestingly, it was shown that complestatin displays antiviral properties [15]. All other structural variations of glycopeptide antibiotics occur almost exclusively with regard to halogenation, glycosylation and lipidation. These can vary broadly and contribute to the heterogeneity of the compound class.

2.2
Mechanism of Action

Vancomycin glycopeptides interfere with the late stages of peptidoglycan synthesis of the bacterial cell wall. This is also the case for the well-known β-lactam antibiotics (e.g. penicillin). Because of a relatively high molecular mass ($\sim$ 1400 Da) vancomycin and its analogues are not able to cross the inner membrane of the bacterial cell wall. The inhibition of peptidoglycan synthesis is based on binding of vancomycin to the D-Ala-D-Ala terminus of peptidoglycan precursors of the bacterial cell wall [16]. Further insights into the mechanism of action came along with the elucidation of vancomycin's structure [4–6]. It was shown by NMR experiments that vancomycin forms five hydrogen bonds with the Lys-D-Ala-D-Ala tripeptide moiety [17] (Fig. 2).

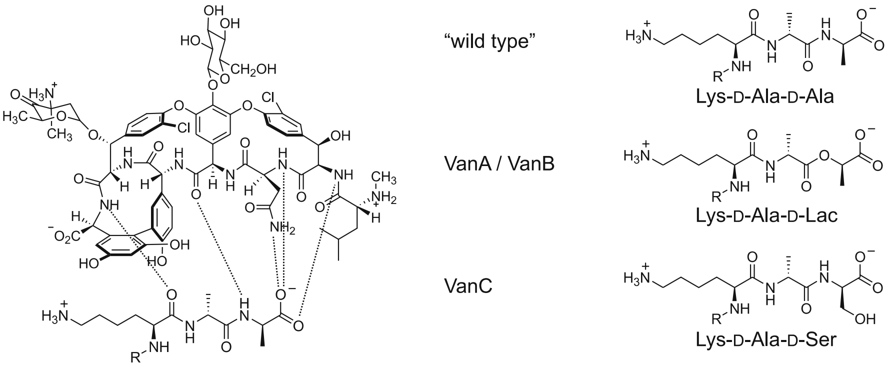


Fig. 2 Interaction of vancomycin with its target Lys-D-Ala-D-Ala peptide via five hydrogen bonds. VanA/B- and VanC-resistance types represent structural variations in the peptide sequence of the bacterial cell wall biosynthesis, thus leading to reduced binding

Vancomycin is thought to bind to lipid II monomers or to the termini of glycan strains, thus shielding the substrates from transglycosylation and transpeptidation steps. Because of an inhibited cross-linking of the peptidoglycan, the bacterial cells become susceptible to lysis.

Interestingly, one characteristic of vancomycin-type glycopeptides is their ability to dimerize. Aggregation of vancomycin in aqueous solution was first reported by Nieto and Perkins in 1971 [18]. Later, NMR and X-ray experiments resolved the nature of dimerization for various glycopeptide antibiotics [19–21]. It was shown that in the dimer hydrogen bond interactions were formed between the back sides of the two antibiotic molecules (Fig. 3). The affinity of the resulting homodimer for cell wall fragments increases by a factor of up to ten compared to the monomolecular glycopeptide antibiotic [22]. In turn, cell wall fragments were reported to enhance dimerization by factors of two to 100 [22]. These findings led to the development of a model of cooperative binding, which was further supported by titration NMR experiments [23, 24]. Based on these in vitro data one can assume that the process of dimerization plays an important role in the physiological mode of action of a number of vancomycin-type antibiotics. In contrast, the lipoglycopeptide teicoplanin (type IV) is suggested to insert into membranes (i.e. the cytoplasmic membrane) via its fatty acid moiety. Both effects, dimerization and membrane insertion, have recently been discussed in detail [25].

In contrast to many other antibiotics, vancomycin had been used for nearly 30 years without the observation of noteworthy bacterial resistance. However, in 1988, resistance to vancomycin was reported for strains of enterococci [26] and almost 10 years later resistance was observed in staphylococcal isolates [27]. Three basic types of enterococcal vancomycin resistance have been identified and are understood on a molecular basis [28].

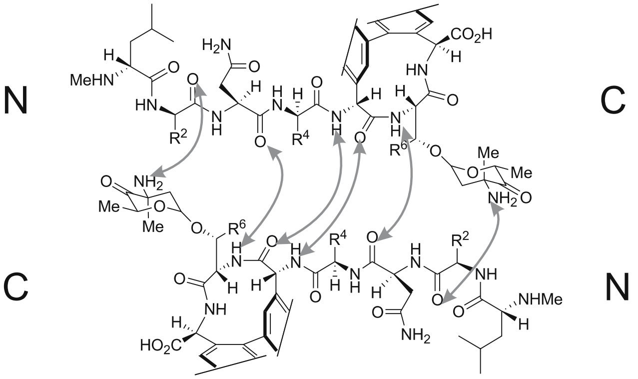


Fig. 3 Back-to-back homodimerization of glycopeptide antibiotics (e.g. balhimycin) via hydrogen bonds

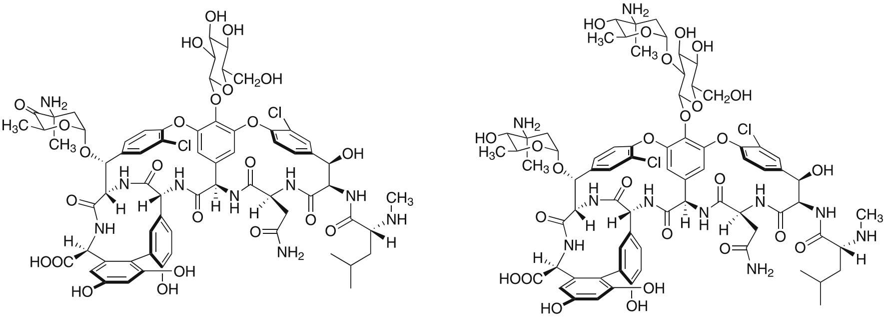
These are the shifting of bacterial cell wall biosynthesis from D-Ala-D-Ala to D-Ala-D-Lac (VanA/VanB) or D-Ala-D-Ser (VanC) peptides in bacterial cell wall biosynthesis (Fig. 2). The resulting significant decrease in affinity of vancomycin leads basically to a reduced susceptibility of resistance of the target strain. Staphylococcal resistance is presumably based on a thickened cell wall and an increased synthesis of D-Ala-D-Ala peptides scavenging glycopeptides [29]. This resistance problem induced vigorous screening efforts to find novel alternative antibiotics, but also for the semi-synthetic generation of vancomycin antibiotics based on SAR studies (see below).

2.3
Biosynthesis

Biosyntheses of the glycopeptide antibiotics balhimycin and chloroeremomycin (Fig. 4) have been studied in detail and are considered to be representative for vancomycin biosynthesis [30, 31]. Compared to vancomycin, the structures differ only in the carbohydrate moieties attached to the aglycon. Over the past years considerable progress has been made in understanding the biosynthesis of this type of glycopeptide.

Balhimycin (Fig. 4) biosynthesis has mainly been investigated with gene inactivation mutagenesis [32]. The gene cluster of the balhimycin producer *Amycolatopsis balhimycina* was published by Pelzer et al. in 1999 [33] shortly after the sequence of the chloroeremomycin (Fig. 4) biosynthesis gene cluster [34]. The gene clusters of teicoplanin [35, 36] and complestatin [37] have been sequenced recently.

Based on sequence homologies, 31 of the 33 open reading frames (*orfs*) of the balhimycin gene cluster were attributed to known enzyme functions (Fig. 5). The biosynthesis of vancomycin glycopeptides is divided into three



Balhimycin **Chloroeremomycin**

Fig. 4 Structures of vancomycin-type glycopeptide antibiotics balhimycin and chloroeremomycin

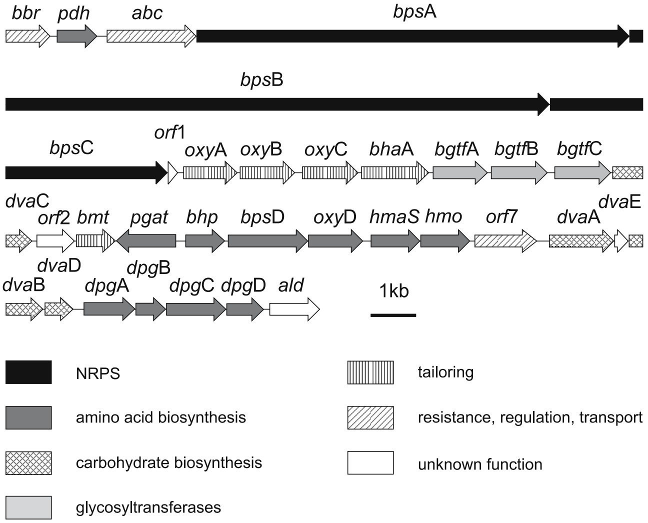


Fig. 5 Balhimycin biosynthesis gene cluster of the *A. balhimycina* and assignment of proposed functions [33]

steps: (1) synthesis of non-proteinogenic amino acids and carbohydrates; (2) non-ribosomal peptide synthesis; (3) post-NRPS tailoring.

Although not explicitly shown, it is reasonable to assume that the amino acid building blocks Leu and Asn are derived from primary metabolism. The biosynthetic genes of non-proteinogenic amino acids 4-hydroxyphenylglycine (Hpg), 3,5-dihydroxyphenylglycine (Dpg) and β-hydroxytyrosine (Hty) are all located in the biosynthesis gene cluster (Fig. 5). The biosynthesis of Hpg has been characterized by protein overexpression and involves the precursor metabolites prephenate and tyrosine [38–40]. Biosynthesis of Dpg involves a polyketide synthase (DpgA), which assembles a linear polyketide from four malonyl-CoA units that is subsequently transformed into 3,5-dihydroxyphenylacetate supported by DpgB and DpgD [41]. DpgC then converts the phenylacetate into phenylglyoxylate [42], which is transformed into Dpg by the transaminase Pgat [41]. Interestingly, polyketide synthase DpgA shows strong similarities to plant chalcon synthases (PKS III).

The amino acid Hty is thought to be synthesized from tyrosine covalently bound to the NRPS BpsD by β-hydroxylation with P450-type monooxygenase OxyD. After β-hydroxylation Hty is cleaved from BpsD by the hydrolase Bhp [43–45]. The chlorination time-point of the putative halogenase BhaA is still elusive, but is likely to occur after Hty biosynthesis [43, 44, 46]. Vancosamines are the characteristic carbohydrates of vancomycin antibiotics and their biosynthetic assembly has been elucidated recently by protein overexpression and substrate conversion experiments [47].

The balhimycin biosynthesis gene cluster contains three genes coding for three NRPS (BpsA/B/C) with a 3 : 3 : 1 modular structure. Apart from the usual C-A-T-motif for recognition, activation and condensation there exist epimerase domains (E) in modules 2, 4 and 5, which are responsible for transforming (S,R)-Hty into (R,R)-Hty and (S)-Hpg into (R)-Hpg. The final NRPS-products are hexa- and heptapeptide precursors, respectively.

One prominent hallmark of the group of glycopeptide antibiotics is the tricyclic nature of the heptapeptide that defines the overall rigid structure, which is the base for the antibiotic activity of vancomycin-type glycopeptides. The side-chain bridging reactions are therefore crucial steps in the biosynthetic assembly. They are carried out by three P450-type monooxygenases OxyA/B/C [31, 33]. The sequence of side-chain cyclization was elaborated by directed gene inactivation of *oxyA/B/C* and analysis of metabolites accumulated in culture filtrates of oxygenase mutants [48, 49]: OxyB introduces the aa2-4 diarylether cross-link, then OxyA catalyses the second ring formation (aa4-6) and finally OxyC forms the aa5-7 biaryl cross-link (Fig. 6). Two out of three P450-monooxygenases involved in vancomycin biosynthesis have been crystallized [50, 51]. Recently, the biosynthetic picture has been refined with the hypothesis that side-chain cyclization by P450-monooxygenases steps would take place on NRPS-bound peptide [46]. Biosynthesis mutants inactivated in heptapeptide formation and Dpg-delivery were thoroughly investigated for truncated metabolites accumulated in culture filtrates. The complete series of di- to pentapeptides as well as variously cyclized hexapeptides were detected. This suggestion was supported by in vitro experiments with overexpressed OxyB [52]. In addition, the Robinson group could show that synthetic linear hexapeptides were cyclized to monocyclic hexapeptides only if they were covalently bound as phosphopantetheine derivatives. Likewise, the conversion of linear heptapeptides was also achieved, raising some doubts on the exact point of time during NRPS assembly. Al-

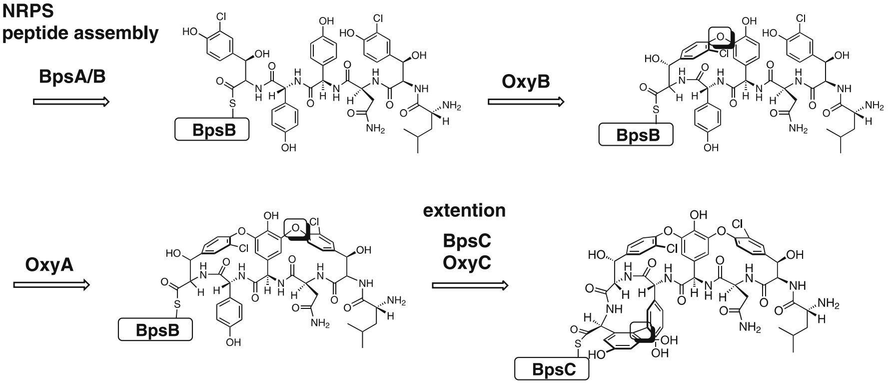


Fig. 6 Model of oxidative ring closure reactions in vancomycin-type glycopeptide antibiotic biosynthesis [46]

though it has not been shown, it is fair to assume that all three biosynthetic steps catalysed by OxyA/B/C take place on NRPS-bound peptide intermediates (Fig. 6) [46]. As a consequence, aglycon assembly by NRPS in interaction with oxygenases would represent a strongly controlled and complex biosynthetic process. Furthermore, one could speculate that this peptide processing increases efficiency of the biosynthesis and the prevention of putatively toxic cross-linking reactions of e.g. tyrosine residues of proteins.

The final stages of vancomycin-type biosynthesis are glycosylation of the aglycon with glucose and vancosamine. The vancosamin sugar is characteristic for the whole vancomycin family of glycopeptide antibiotics and also occurs in its *epi-* and *oxo*-forms (chloroeremomycin and balhimycin). For balhimycin, the attachment of carbohydrates to the aglycon is performed by glycosyltransferases BgtfA/B/C [33]. Glycosylations of chloroeremomycin and vancomycin show corresponding analogies. Glycosyltransferases from vancomycin and chloroeremomycin biosynthesis pathways have been crystallized [53–55] and the in vitro reconstitution with overexpressed glycosyltransferases of chloroeremomycin was performed [56]. This is the base for biocombinatorial attempts to generate novel glycopeptide antibiotics [57–60]. Finally, methylation at the amino terminus of D-Leu is performed by methyltransferase Bmt [61].

These findings represent the main features of vancomycin-type glycopeptides and the results may be extrapolated to the biosyntheses of other related glycopeptides. For the lipoglycopeptide teicoplanin, with four rings, the attachment of fatty acids to the amino sugar side chain has already been described [35]. Similarly, the sequence of the biosynthetic gene cluster of complestatin [37] gives insights into the possible course of the biosynthesis of this antibiotic.

2.4
Semi-synthetic Derivatives of Vancomycin and Related Glycopeptide Antibiotics

Since the discovery of vancomycin, a considerable number of semi-synthetic glycopeptide derivatives have been generated (Fig. 7). The synthetic aspects

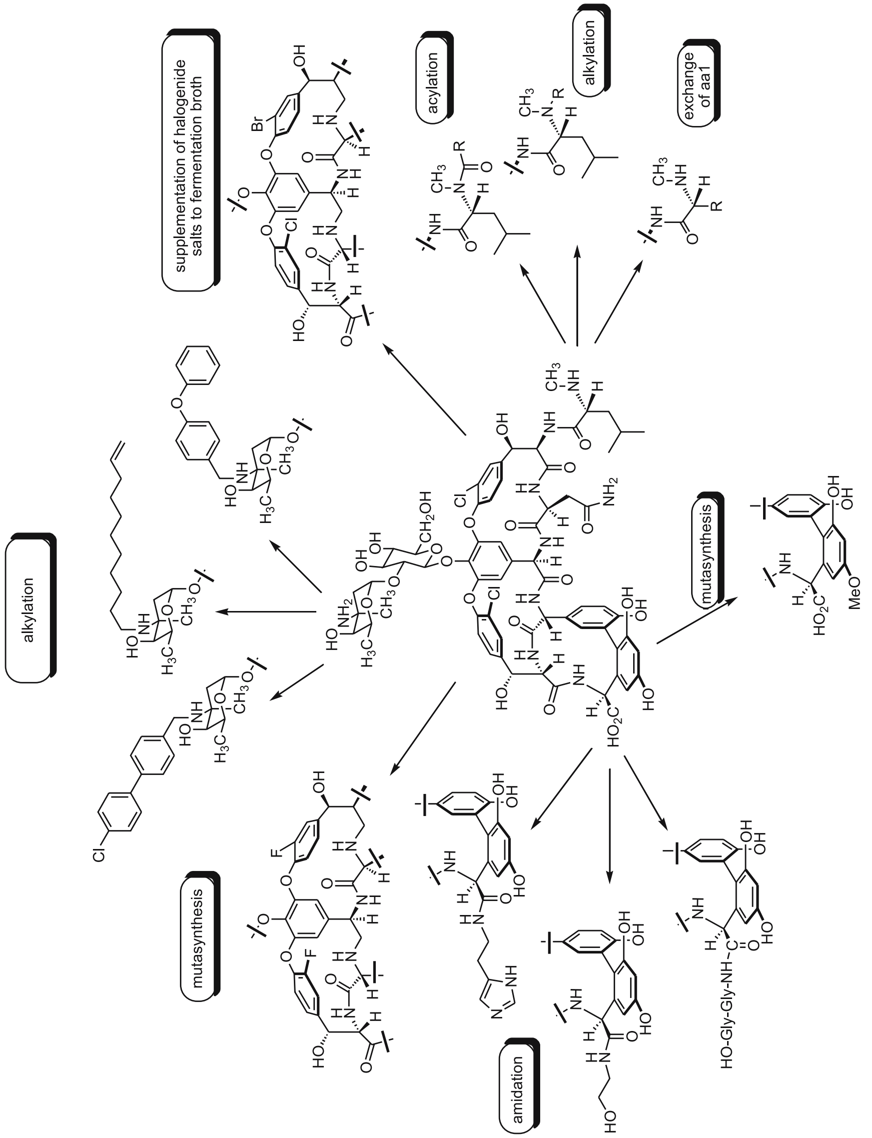


Fig. 7 Selection of representative examples of semi-synthetic and mutasynthetic vancomycin-type derivatizations

of these derivatives have been highlighted in a number of reviews [12, 29, 62, 63] and herein only a selection of a few derivatives is given. Initial SAR investigations probed the role of halogenation and glycosylation, mostly from natural isomers. Chemical derivatization reactions were directed to easily modifiable groups, i.e. amino groups and carboxy groups. The most important studies were performed with acylation or alkylation of the *N*-terminus, later selectivity was improved for modification of the amino sugar. In summary, these studies showed that modifications of the *N*-terminus basically abolish antibiotic activity.

Other examples comprise modifications of the *C*-terminus as amides [64], hydrazides [65], carboxamides [66], urea derivatives [65] and the attachment of peptides [67]. Also *C*-terminal esters with hydrophobic alcohols were synthesized [68, 69]. Dimer formation of glycopeptides by hydrogen bonding inspired the chemical synthesis of covalently linked dimers [70, 71] and trimers [72]. Modifications at the aromatic moieties of the peptide backbone have been introduced by using biosynthesis mutants or change of media composition. Thus, glycopeptides with fluorine [73], bromine [74] and methoxy derivatives were obtained [75]. The lack of aromatic side chain cross-links is accompanied by complete loss of antibiotic activity. This has been extensively proven by linear, mono- and bicyclic biosynthesis intermediates [48, 49, 76] that showed no antibiotic activity. Furthermore, with the exception of a Mannich reaction at ^{7}Dpg [77], the aromatic ring systems show an almost complete chemical inaccessibility of the aromatic residues. Most chemical modifications have no or too little effect in enhancing antibiotic activity. The exceptions are *N*-alkyl derivatives of the amino sugar (oritavancin, dalbavancin, telavancin) and amide derivatives of the *C*-terminus (dalbavancin). Other contributions are based on a fully synthetic rather than a semi-synthetic approach in reducing molecular complexity, e.g. structurally simplified monocyclic analogues [78]. However, apart from promising preliminary data on antibiotic activity, additional data have not been reported using this approach.

Currently, three semi-synthetic glycopeptide derivatives, oritavancin (LY 333328), dalbavancin and telavancin, are in clinical-phase trials [79] (Fig. 8). With regard to the chemical structure, oritavancin and telavancin are vancomycin derivatives, whereas dalbavancin is a teicoplanin analogue. Remarkably, oritavancin and telavancin are active against vancomycin-resistant *S. aureus* (VRSA) and vancomycin-resistant enterococci (VRE). In contrast, dalbavancin does not target VRE of the VanA-type, although it has enhanced activity against MRSA. Speculations on the mode of action of oritavancin originally proposed a membrane insertion mechanism. However, recent investigations by Kahne and coworkers on oritavancin suggest an additional glycosyltransferase inhibition by the chlorobiphenyl moiety [80–82].

In conclusion, from SAR studies on glycopeptide antibiotics three semi-synthetic derivatives evolved that have reached clinical trials, particularly for infections with vancomycin-resistant bacteria. The tuning of the van-

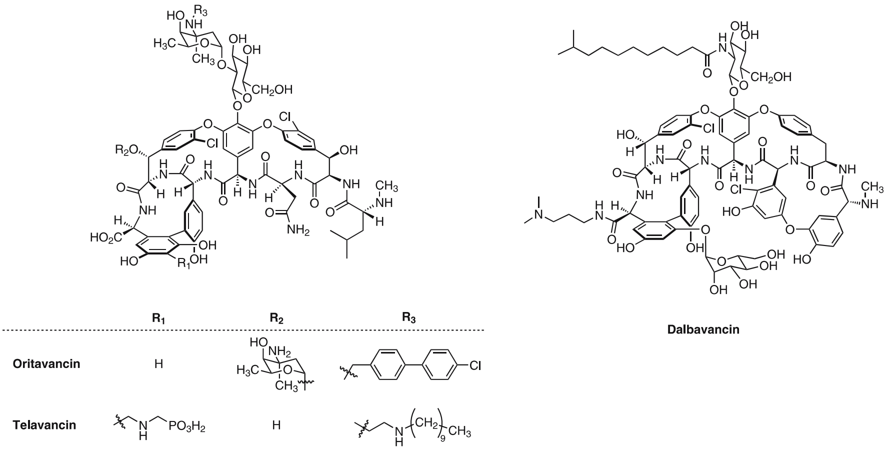


Fig. 8 Structures of glycopeptides oritavancin, telavancin, and dalbavancin in clinical phases

comycin core structure represents a remarkable example of a natural compound, chemically transformed into a drug, which is active against previously resistant bacteria.

3
Ramoplanin

3.1
Structure

The growing resistance of Gram-positive staphylococci and enterococci against vancomycin promoted the search for alternative antibiotics. The glycopeptide ramoplanin is considered an alternative for the treatment of infections by multiresistant Gram-positive bacteria. The growing interest over the past years has fuelled investigations on the biosynthesis, mode of action, and total synthesis by several research groups, which have been summarized in several review articles [83, 84].

Ramoplanin was discovered in 1984 in the course of an industrial drug discovery program directed against cell wall biosynthesis inhibitors [85]. Ramoplanin factors A1, A2 and A3 (Fig. 9) were isolated at Biosearch Italia from culture broths of *Actinoplanes* strain ATCC 33076, which originated from a soil sample collected at Vaghalbod Village, Indore, India.

Ramoplanin is active against a broad spectrum of Gram-positive bacteria, including VRE and MRSA. Low MIC (minimum inhibitory concentration) values of about 1–2 μg/mL were reported for most of the microorganisms

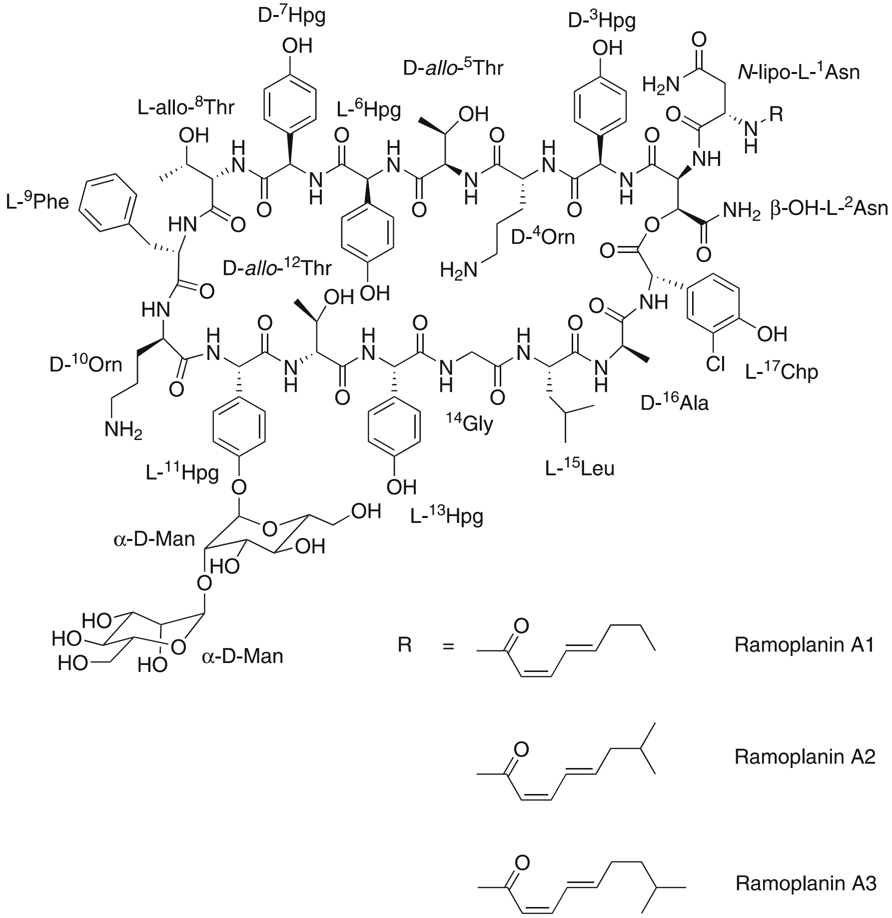


Fig. 9 Structures of ramoplanin factors A1–A3

tested [86]. Like vancomycin, ramoplanin shows no activity against Gram-negative bacteria. The reason for this may be that it does not penetrate the outer membrane characteristic for Gram-negatives. Ramoplanin has been found more potent compared to vancomycin in different antibacterial assays [84, 87].

Ramoplanin factor A2 is currently undergoing clinical-phase trials for the treatment of Gram-positive bacterial infections with pathogens that are resistant to vancomycin, macrolides or *β*-lactams [79]. Ramoplanin retains its effectiveness despite efforts to elicit resistance in the laboratory [88, 89]. Remarkably, studies with patients have shown that oral applicability of ramoplanin might be attained [79], which is unusual for peptides as well as in comparison to intravenous application of vancomycin.

Ramoplanin is a lipoglycodepsipeptide antibiotic [90, 91] and shares some structural moieties with teicoplanin as well as with mannopeptimycins. It consists of a 49-membered macrocyclic depsipeptide ring with 17 amino acids. All three ramoplanins A1–A3 share the same peptide skeleton.

Like many natural products from the NRPS biosynthetic pathway, ramoplanin amino acids have D- and L-configurations and consist of a considerable number of non-proteinogenic amino acids. These include ornithine (Orn), 4-hydroxyphenylglycine (Hpg), 3-chloro-4-hydroxyphenylglycine (Chp), D/L-*allo*-Thr and β-hydroxyasparagine (β-OH-Asn). A remarkable structural feature of ramoplanin is the high number of aromatic amino acids (5 × Hpg, 1 × Chp, 1 × Phe) and β-hydroxy amino acids (3 × Thr, 1 × β-OH-Asn). The *N*- and *C*-termini of the aglycon are linked by a lactone bridge between the carboxyl of residue ^{17}Chp and the hydroxyl of β-OH-^{2}Asn. Amino acid ^{1}Asn stands exocyclic of the macrolactone ring and represents the *N*-terminus, which is subject to modifications with unsaturated fatty acids. All fatty acids are $\alpha,\beta,\gamma,\delta$-unsaturated. An α-1,2-dimannosyl disaccharide is attached to the phenol of ^{11}Hpg. Comparison of the antimicrobial activity of ramoplanins A1–A3 shows only marginal differences. Obviously, ramoplanin has strong structural similarities to ramoplanose, a trimannosylated ramoplanin [92] and to enduracidin [93], which in fact was discovered much earlier. Enduracidin is also a 17mer peptide forming a 16-amino acid depsipeptide macrocycle with an exocyclic ^{1}Asn. Compared to ramoplanin, the aromatic amino acids (5 × Hpg, 1 × Dpg) and β-hydroxy amino acids (3 × Thr, 1× Ser) of enduracidin share identical positions in the aglycon. The characteristic feature of enduracidin is the amino acid D/L-enduracidine (End) that contains a guanidine side chain. This amino acid has also been described from mannopeptimycins (see Sect. 4).

The 3D-structure of ramoplanose determined from multidimensional NMR experiments revealed a double-stranded antiparallel β-sheet with seven intramolecular hydrogen bonds [92]. Apart from some more advanced data sets and evaluation conditions, these findings were basically confirmed for ramoplanin A2 with a β-sheet and six hydrogen bonds with turns in the regions of amino acids 8/9 and 14–16, respectively [94]. Remarkably, the solution structure of ramoplanin in MeOH rendered a dimer at the interface of amino acid positions 10–14 [95]. This finding has some analogy to the mode of action of vancomycin-type glycopeptides and might provide some explanation on the mode of action (see below).

3.2
Biosynthesis

The non-proteinogenic nature of most amino acids assembled in the ramoplanin aglycon indicates an NRPS origin. Cloning and sequencing studies revealed a biosynthetic gene cluster of 88.5 kilobase (kb) identified from

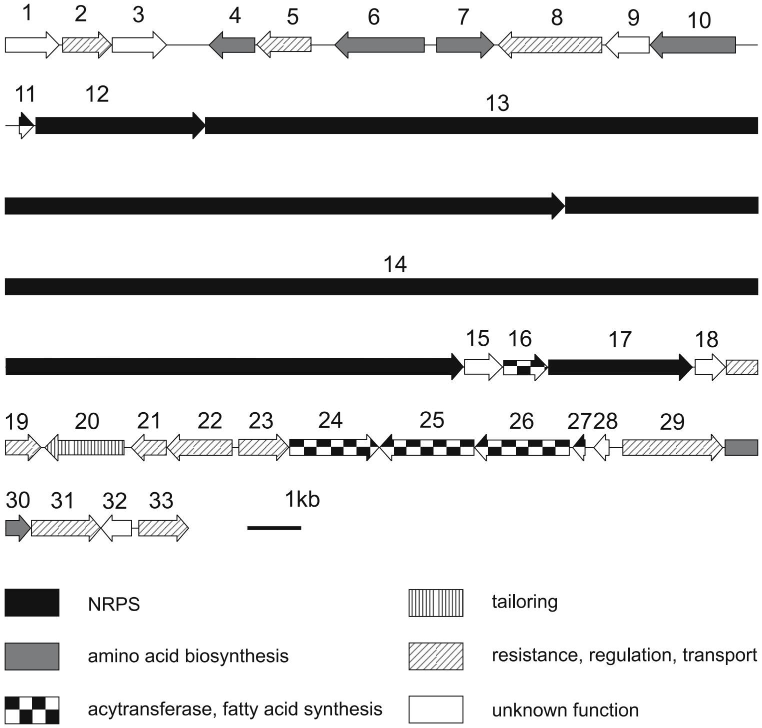


Fig. 10 Ramoplanin biosynthesis gene cluster and assignment of proposed functions [84]

Actinoplanes ATCC 33076 (Fig. 10) [96]. Sequencing and homology searches identified 33 open reading frames (*orfs*) and putative functions could be attributed to 29 *orfs* [97]. As for the biosyntheses of other NRPS-peptides, the biosynthesis may be grouped into building block synthesis, peptide assembly and the final tailoring steps. Among the genes for the synthesis of non-proteinogenic amino acids, *orfs4/6/7/30* resemble enzymes of Hpg biosynthesis from vancomycin-type glycopeptides [30]. Interestingly, *orf28* codes for a putative chorismate mutase that may provide prephenate from chorismate due to the raised need of prephenate as a precursor for Hpg biosynthesis.

Several *orfs* are attributed to the synthesis of the acyl chain attached to the *N*-terminus of the peptide aglycon. The fatty acid is likely to be synthesized in analogy to fatty acid biosynthesis: An acyl-CoA ligase (*orf26*), two acyl carrier proteins (ACP, *orf11/27*), two FAD-dependent dehydrogenases (*orfs24/25*) and a NAD-dependent reductase (*orf16*) have been identified. However, no ketosynthase that is required for chain elongation could be assigned. This function may be provided by primary metabolism. The direct precursor for the attachment of the acyl chain to the peptide is thought to be an ACP-bound (*orf11/27*) fatty acid. The transfer of the acyl chain to the Orf12-bound ^{1}Asn-NRPS assembly line is the subject of current investigations.

Four genes, *orf12–14* and *orf17*, code for NRPSs. Orf13 and Orf14 are large multimodular peptide synthetases for the attachment of D-^{3}Hpg to ^{9}Phe and D-^{10}Orn to ^{17}Hpg to the growing peptide chain, respectively. The *C*-terminal end of Orf14 bears a thioesterase (Te)-domain for macrolactonization of the peptide aglycon. Despite the occurrence of D-amino acids, no epimerization domains have been found on ramoplanin NRPSs. It is currently speculated that these functions are provided by enzymes working *in trans*. Remarkably, in the sequence of Orf13 no A-domain was identified for ^{8}Thr. It has been suggested that adenylating function is performed by NRPS of Orf17. Orf12 is putatively responsible for adenylation of both ^{1}Asn and β-OH-^{2}Asn, since these functions could not be assigned otherwise. Orf15 is an enzyme with homologies to a type II thioesterase. Tailoring reactions are currently assigned to include glycosylation at the phenol of ^{11}Hpg, and chlorination of ^{17}Hpg by a halogenase (*orf20*) to yield ^{17}Chp, respectively. However, no *orf* could be attributed to glycosyl transferase activity. Enzymes from the primary metabolism may provide this function. Two *orfs* could be identified that probably encode a non-heme Fe-hydroxylase (*orf10*) responsible for β-hydroxylation of ^{2}Asn. Results obtained for vancomycin-type glycopeptide balhimycin suggest halogenation and β-hydroxylation to occur on a NRPS-bound form, putatively before or during peptide assembly [43–45]. Finally, several *orfs* are most likely involved in transport (*orf2/8/23/31*) and transcriptional regulation/resistance (*orf5/21/22/33*). However, the mechanism of self-resistance is yet unknown. The function of several *orfs* still remains to be clarified in order to complete the picture of ramoplanin biosynthesis.

3.3
Mechanism of Action

From the initial antibiotic screen it was known that ramoplanin interferes with bacterial cell wall synthesis. Subsequent studies by Somner et al. considered the membrane-associated glycosyltransferase MurG of the peptidoglycan biosynthesis pathway as the cellular target for ramoplanin [98, 99]. They assumed ramoplanin to act before transpeptidation by blocking lipid II formation (MurG inhibition). No interference with lipid I biosynthesis was found. Several years later, Brötz et al. [100, 101] observed that ramoplanin (and also nisin/epidermin lanthionin peptides) co-migrated with lipid II in thin-layer chromatography. Subsequent assays by Lo et al. proposed the inhibition of the transglycosylation step as the primary target of ramoplanin [102]. Inhibition kinetics of ramoplanin revealed that ramoplanin binds lipid II with a dissociation constant of 50 nM in a 2 : 1 (ramoplanin : lipid II) stoichiometry [103]. A completely different kinetic behaviour was observed for inhibition of MurG, which supports the abovementioned argument of transglycoslyation reaction as the ramoplanin target [103]. With regard to an interaction with MurG catalysis it was shown that ramoplanin directly inhibits MurG rather than by binding to the lipid I-substrate [104]. Further arguments which likely suggest transglycosylation as the target of ramoplanin are, that MurG is located at the cytoplasmic side of the inner membrane and that it is therefore not accessible to ramoplanin. Furthermore, the concentrations necessary for MurG inhibition appear too high. Interestingly, fibril formation of lipid II analogues was observed upon titration with ramoplanin. However, it is not clear yet, whether this observation is of physiological relevance [102, 105]. To date, no resistance against ramoplanin has been observed, which rises early hopes for a future long-term use as an antibiotic. However, as mentioned above the nature of self-resistance of the producing organism is also elusive.

3.4
Structure–Activity Relationship Studies

Structure–activity relationship studies were performed in order to assign important molecular features modulating antibiotic activity (Fig. 11). Hydrolytic cleavage of the macrolactone ring to the linear 17mer peptide fully deleted antibiotic activity [87]. In contrast, the catalytic reduction of unsaturated fatty acids [106, 107], lipid exchange [108] or even deacylation [109] showed only modulating influence on antibiotic activity. Deglycosylated ramoplanins turned out to be as antibiotically active as ramoplanin [110, 111]. Further simple modifications were the acylation, the reductive amination and the guanidylation of Orn residues. Antibacterial activity decreased

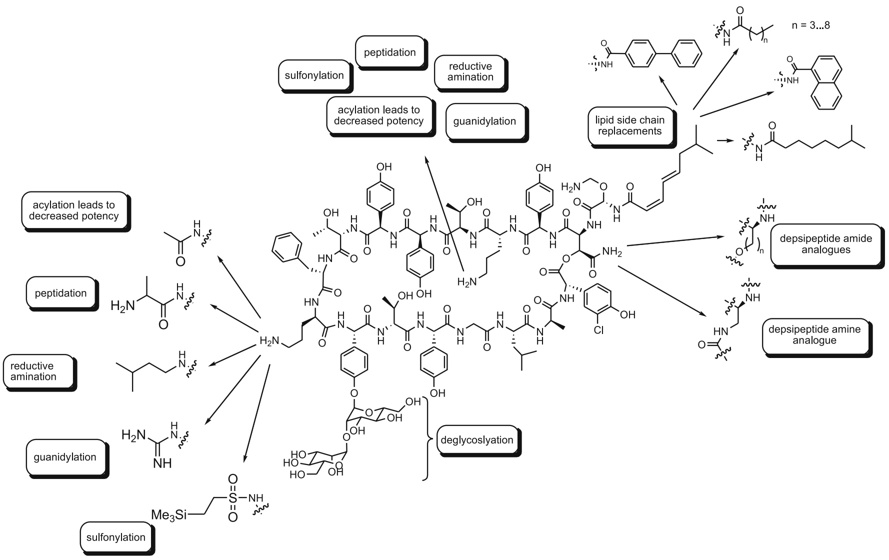


Fig. 11 Examples of structural variations of ramoplanin

significantly with modification of Orn residues, particularly upon loss of the positive charge [87]. The contribution of each single Orn residue to antibiotic activity is still under debate [83].

Total synthesis of ramoplanin A2 aglycon has recently been achieved by the Boger group [112], providing further important means for detailed SAR studies. Furthermore, two amide analogues have been synthesized where β-OH-^{2}Asn has been replaced by diaminopropionic acid (Dpr) and diaminobutyric acid (Dab), respectively [109]. The Dpr derivative proved to be even more active than ramoplanin itself. As a consequence, the Dpr-macrolactam provides a more chemically stable alternative to natural ramoplanin. In contrast, the ring extension with Dab-analogue by one methylene group led to a loss of antibiotic activity proving the importance of exactly maintaining the ring size of the 49-membered ring.

In conclusion, total synthesis set the stage for comprehensive but laborious investigations of structure–activity relationships. With the synthetic tools at hand it would be interesting to study, for example, the importance of Hpg residues by exchange with phenylglycine or even other hydrophobic (non-aromatic amino acids). Furthermore, the mechanistic effects of those derivatives will aid understanding of lipid II binding.

4
Mannopeptimycins

4.1
Structure and Biosynthesis

In the late 1950s researchers at Wyeth isolated a mixture of antibiotic compounds produced by the strain *Streptomyces hygroscopicus* LL-AC98. This compound mixture was named the AC98 complex and was shown to be effective against Gram-positive bacteria [113]. Not until four decades later did Wyeth Research re-examine the AC98 complex. In these studies, AC98 showed considerable in vitro activity against methicillin-resistant staphylococci (MRS) and vancomycin-resistant enterococci (VRE) by inhibition of cell wall synthesis.

The AC98 complex was dereplicated into five compounds [114], and their structures (Fig. 12) elucidated by extensive NMR analysis, high-resolution Fourier transform ion cyclotron resonance (FTICR) mass spectrometry and chemical degradation.

Structure elucidation revealed these compounds as glycopeptide derivatives. The cyclic hexapeptide is composed of the six amino acids: L-serine (Ser), glycine (Gly), β-methylphenylalanine (Mephe), D-tyrosine (Tyr), and two stereoisomers of the amino acid, β-hydroxyenduricidin (Aiha-A and -B). The characteristic feature of this amino acid is a cyclic guanidine moiety. The sequence and absolute stereochemistry of the amino acids constituting

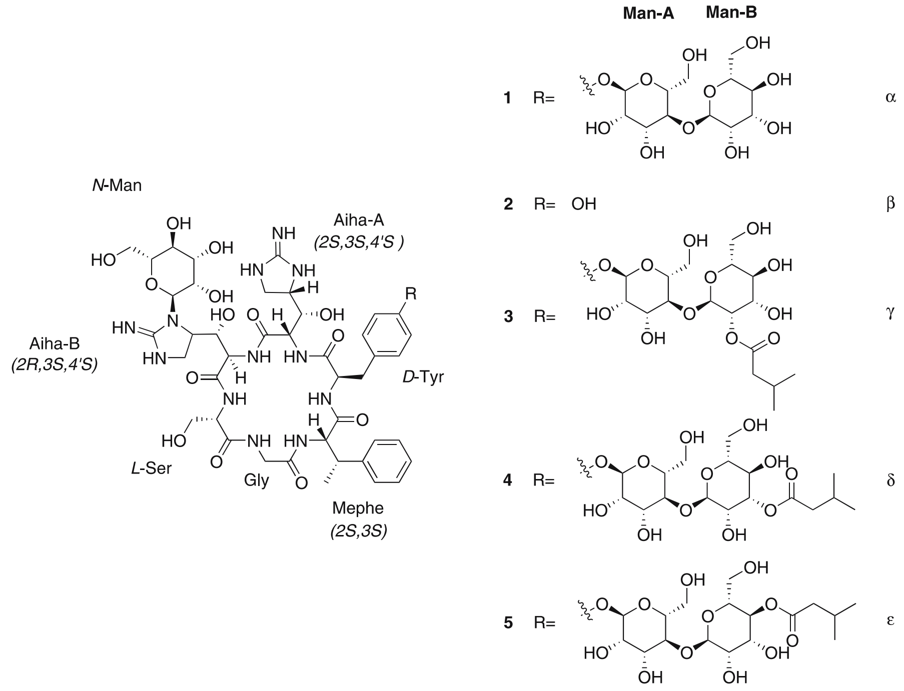


Fig. 12 Structures of mannopeptimycins α–ε

the hexapeptide aglycon was determined as [114]: *c*-[L-Ser-Gly-(2*S*3*S*)MePhe-D-Tyr-(2*S*3*S*4′*S*)AihaA-(2*R*3*S*4′*S*)AihaB], which was primarily determined by NMR spectroscopy. Because of the extensive mannosylation the five compounds were named mannopeptimycins α through ε.

The mannopeptimycins (MPM) differ from each other in their glycosylation patterns. All mannopeptimycins bear mannopyranose bound *N*-glycosidic to Aiha-B. In addition, compounds MPM-α and MPM-γ-ε possess a mannosyl disaccharide moiety (Man-A and Man-B) bound to D-tyrosine. This ((1 → 4)-α-Man)-α-Man disaccharide is attached via an *O*-glycosidic bond to the peptide aglycon. The presence and the attachment site of an isovaleryl group in the terminal mannose (Man-B) at C-3/C-4 and at C-5 for esters MPM-γ-ε are crucial for antibacterial potency. These compounds show a significantly higher antibacterial activity than the parent compound mannopeptimycin α.

With regard to the biosynthesis and the biosynthetic gene cluster no data are available yet. However, according to the chemical structure and similarities to previously mentioned glycopeptide antibiotics it is reasonable to assume a non-ribosomal peptide assembly. Mannosylation may be due to a specific mannosyltransferase or alternatively based on primary metabolism, as supposed for ramoplanin.

4.2
Mechanism of Action

The recent re-examination of the antibacterial potential of mannopeptimycins prompted further investigations, particularly on the mode of action [114, 115]. Initial studies indicated that the mannopeptimycins are inhibitors of bacterial cell wall biosynthesis. This conjecture was strengthened by the structure of the mannopeptimycins. For molecules with a molecular mass > 1000 Da it is fair to assume the bacterial cell wall is the distinct target of mannopeptimycins.

The biggest antibiotic potential was assigned to mannopeptimycins bearing isovaleryl esters, with an exclusive but strong activity against Gram-positive bacteria. Mannopeptimycin δ inhibits the incorporation of radiolabelled *N*-acetylglucosamine into peptidoglycan [116]. Furthermore, in *S. epidermidis* cultures a significant accumulation of UDP-MurNAc-Ala-Glu-Lys-D-Ala-D-Ala (UDP-MurNAc-pentapeptide) was observed in the presence of mannopeptimycin δ [116]. As a consequence, early steps of peptidoglycan biosynthesis (MurA through MurF) were not affected by mannopeptimycins, and subsequent lipid I and II formation was also not inhibited [116]. The formation of lipid II in the presence of mannopeptimycin was proved with whole-cell membrane preparations of *S. aureus* with UDP-MurNAc-pentapeptide and radiolabelled UDP-[^{14}C]GlcNAc. Similarly, lack of competitive binding with vancomycin disproved mannopeptimycins as binders of D-Ala-D-Ala-peptides. These results suggested an interaction of mannopeptimycins with later stages of the cell wall biosynthesis, transglycosylation or on transpeptidation.

In a gel-shift assay, tritiated mannopeptimycin α was used in order to prove binding of mannopeptimycins to lipid II [116]. In a modified version of this assay, components of an in vitro peptidoglycan biosynthesis were UV-cross-linked to tritiated mannopeptimycin. Furthermore, in competition assays with various mannopeptimycin derivatives a correlation between the potential for lipid II interaction and the antibacterial activity was shown. In competitive PBP-binding tests with benzylpenicillin and mannopeptimycin δ the inhibition of transpeptidation reaction was assayed. The putative targets, staphylococcal or *E. coli* PBPs, still showed activity. As a consequence, lipid II is currently considered to be the primary target of mannopeptimycins inhibiting transglycosylation reaction.

4.3
Semi-synthetic Derivatives

A number of semi-synthetic derivatives were synthesized from natural mannopeptimycins (Fig. 13). Degradation experiments revealed that the hydrophobic isovaleryl group attached to the mannosyl side chains of com-

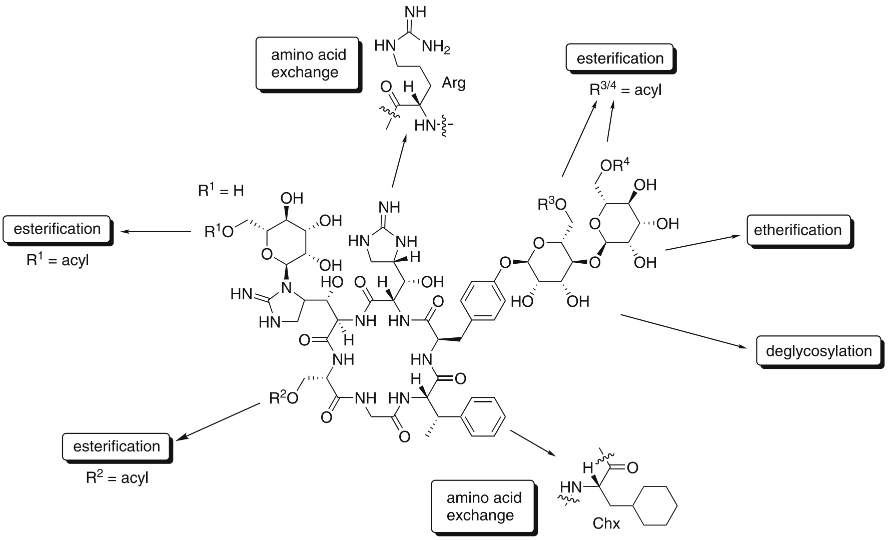


Fig. 13 Semi-synthetic modification of mannopeptimycin α

pounds γ, δ and ε is crucial for antibacterial activity. Enzymatic methods for smooth preparation of deglycosylated mannopeptimycins were developed [117] and mannopeptimycin β was prepared from the α-compound by exposure to jack-bean α-mannosidase specific for terminal cleavage of unsubstituted α-mannopyranosides. In addition, with almond α-mannosidase the *O*-monosaccharide of mannopeptimycin α was obtained. Both products may be useful starting materials for the preparation of new mannopeptimycin derivatives.

Thus far, the only modification known for the peptide backbone was the replacement of β-methylphenylalanine with cyclohexylalanine (Chx) by precursor-directed biosynthesis [119]. Other modifications are focussed at the mannosyl residues and the fatty acid (isovaleryl) side chain of mannopeptimycin α.

Initial mannopeptimycin modifications were performed by randomized acylations with acyl chlorides or anhydrides [118]. By this method monoesters were obtained modified at positions C-6 of *N*-Man, Man-A and Man-B and at the hydroxy side chain of L-Ser. Modification at C-6 of *N*-glycosidic bound mannose or at the terminal hydroxy group of L-Ser suppressed antibacterial activity. In contrast, esterification of the disaccharide moiety, particularly at C-6 of Man-B led to improved antibacterial activities. Likewise, two carbonate derivatives at C-6 of Man-B, showed comparable antibiotic activities. Based on the results of the SAR studies, further hydrophobic modifications were prepared as cyclic acetals and ketals of the 4,6-positions of the terminal mannose Man-B [119].

With improved derivatization methods, the positions at C-4 and C-6 of Man-B could be selectively addressed [120]. The MIC values of these novel hydrophobic derivatives tested against a variety of Gram-positive bacteria proved all of these derivatives more potent than the natural parent compound. From a reaction sequence, four ether modifications at C-4 of the terminal mannose (Man-B) were obtained as by-products [121, 122]. All isomers were tested for their in vitro antibacterial activities for quantitative MIC against a spectrum of Gram-positive bacteria. The antibacterial activity is dependant on the ether attachment site: 4-*O*-ether > 6-*O*-ether > 3-*O*-ether > 2-*O*-ether. Although ether derivatives are metabolically more stable than esters and acetals, their preparation is more complex and appears less suited for semi-synthetic large-scale modification.

Solid phase peptide synthesis of mannopeptimycin analogues was established by Wang et al. [123]. Arg was used as substitute for non-proteinogenic amino acid Aiha. Dimannosyl-D-tyrosine was prepared as a building block prior to coupling on solid support. Cleavage of the protected linear hexapeptide from the resin was followed by intramolecular cyclization at high dilution. However, this fully synthetic analogue turned out to show only poor antibacterial activity against the bacterial strains tested [123]. The evident lack in antibacterial activity shows that the cyclic guanidine residues of Aiha-A and Aiha-B play a crucial role in maintaining antibacterial activity.

5
Bleomycins

5.1
Structure

The bleomycins (BLMs) are a family of structurally related glycopeptides produced by *Streptomyces verticillus*. They were isolated in the late 1960s from a soil sample collected at a coal mine in the south of Japan [124, 125]. Unlike the above-mentioned glycopeptides, bleomycin has strong cytotoxic properties against eukaryontic cells and is widely applied as a drug in combination chemotherapy against cancer. It is supposed that the biological activity of bleomycin is based on a sequence-selective, metal-dependent oxidative cleavage of DNA and RNA in the presence of oxygen. The chemistry [126], biosynthesis [127, 128] and mode of action [129, 130], has been the subject of a number of recent reviews.

The structure of bleomycin A_2 (Fig. 14) was established by the group of Umezawa by means of chemical degradation, X-ray crystallography and NMR studies [131–133]. In 1978 the proposed structure had to be slightly corrected [134, 135]. Finally, the structural formula was confirmed by total synthesis in 1982 in the laboratories of Hecht [136] and Umezawa [137]. Im-

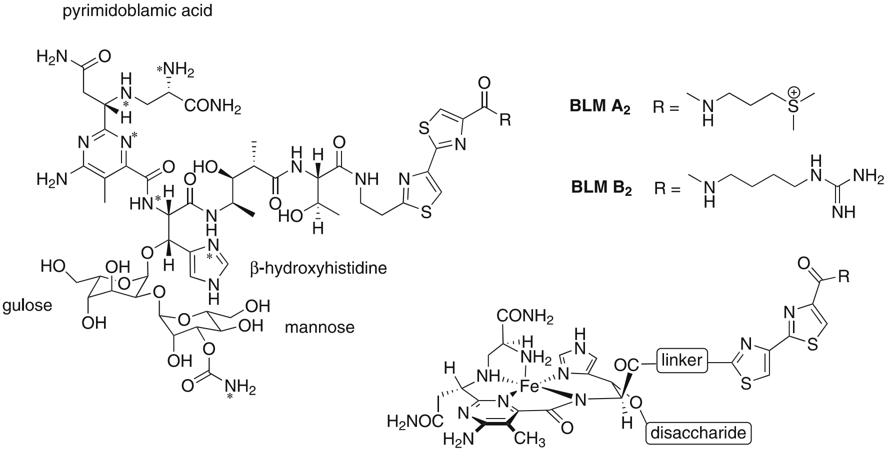


Fig. 14 Structures of bleomycin A2 and B2 and a model of a putative bleomycin–Fe^{2+} complex. *Asterisks* mark the putative coordination sites for metal binding

provements in synthetic methodology in both laboratories led to easier and more efficient strategies [138, 139] culminating in a total synthesis with enhanced stereochemical control by Boger et al. in 1994 [140].

A view on the bleomycin A2 structure reveals that BLMs are peptide-related with pyrimidine, bisthiazole and a disaccharide unit as dominating structural elements (Fig. 14). A total of 19 asymmetric centres have been found for bleomycin A2. The structure of bleomycin is commonly dissected into four functional domains. The *N*-terminus, which is deduced according to the logic of biosynthetic assembly, is represented by pyrimidoblamic acid (PBA) along with (2*S*, 3*S*)-*β*-hydroxy histidine. It is thought that these molecular entities provide the coordination sites responsible for metal binding, thus enabling activation of molecular oxygen. This part of the molecule is directly responsible for DNA/RNA cleavage. A linker domain represented by (2*S*, 3*S*, 4*R*)-4-amino-3-hydroxy-2-methylpentanoic acid and threonine connects the *N*- and the *C*-terminal halves of the molecule. In addition, the linker is thought to modulate the efficiency of DNA cleavage by having an influence on conformation. The *C*-terminal part of bleomycin is formed by a bithiazole moiety with a cationic tail. The cationic tail consists of a guanidine group or of a sulfonium cation. This structural part provides most of the DNA affinity. The disaccharide moiety consists of gulosyl ((1→ 2)-*α*-Gul)-*α*-Man attached via an *α*-glycosidic bond to the hydroxy group of *β*-OH-His. It is hypothesized that this carbohydrate residue mediates cellular uptake by serving as the cellular recognition site.

Compounds that are structurally related to bleomycins are the phleomycins (PLMs) [141–143] and the tallysomycins (TLMs) [144, 145]. Compared to bleomycin, phleomycins have thiazinyl-thiazol in lieu of the bithiazol. The most important differences of tallysomycins are a structural variation in the linker domain and an additional glycosylation site (talose residue) in close proximity to the thiazolyl-thiazole.

5.2
Biosynthesis

Compared to the glycopeptide antibiotics discussed above, identification of the biosynthetic origin of all structural features of BLMs is not easily assigned at first glance. Initial biosynthetic investigations of BLMs were based on feeding isotope-labelled compounds, which indicated a hybrid NRPS-PKS system. The *N*-terminal part was assigned to Dpr, pyrimidoblamic acid and β-OH-His and has an NRPS origin. Similarly, the *C*-terminus, composed of Thr and the bithiazolyl unit, has a NRPS biosynthetic origin. In contrast, the linker domain with (2*S*,3*S*,4*R*)-4-amino-3-hydroxy-2-methylpentanoic acid situated between both termini provides evidence for polyketide biosynthesis.

The *blm* gene cluster has been recently cloned [146–148], which provided deeper insight into bleomycin assembly. A total of 77.5 kb sequences are located directly upstream of *blmA* and *blmB*, two resistance genes which have been cloned earlier [149]. The gene cluster contains *orfs* for ten NRPS genes, a PKS gene encoding one PKS module, and five sugar biosynthesis genes (Fig. 15). Further genes code for other biosynthesis proteins and regulatory

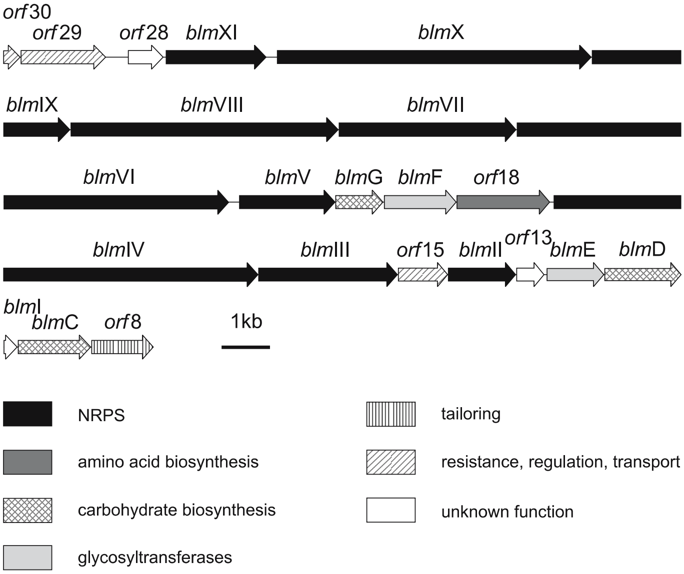


Fig. 15 Sketch of the bleomycin biosynthesis gene cluster and assignment of proposed functions

proteins. The function of resistance protein BlmA is *N*-acetylation of Dpr and the function of BlmB is drug sequestration.

From the chemical structure of bleomycin combined with the predicted functions of NRPS and PKS domains, a biosynthetic assembly line was deduced. The core of this assembly line is represented by modular enzymes in the order: BlmVI-BlmV-BlmX-BlmIX-BlmVIII-BlmVII-BlmIV-BlmIII (Fig. 16). The biosynthetic alignment of BlmIX-BlmVIII-BlmVII constitutes an interesting hybrid NRPS/PKS/NRPS system combining features of hybrid NRPS/PKS systems as well as hybrid PKS/NRPS systems [128]. The deduced sequence according to currently proposed substrates would be: starter-^{1}Ser-^{2}Asn-^{3}Asn-^{4}His-^{5}Ala-^{6}ketidyl-^{7}Thr-$^{8}\beta$-Ala-^{9}Cys-^{10}Cys, not taking in account tailoring reactions taking place on NRPS and PKS modules.

Pyrimidoblamic acid is biosynthesized by BlmVI/BlmV/BlmX and the bithiazole moiety is synthesized by NRPS BlmIV and BlmIII. The NRPS BlmIV and BlmIII contain cyclization (Cy) and oxidation (Ox) domains, respectively (Fig. 16). From in vitro experiments it was concluded that cyclization is performed *in cis* and *in trans* by the Cy-domain of BlmIV. In contrast, oxidation is performed *in cis* and *in trans* by BlmIII and adenylation of Cys is solely performed with the A-domain of BlmIV, since the A-domain of BlmIII is non-functional. Finally, the aglycon is released from the megasynthetase by aminolysis to form a *C*-terminally modified derivative.

The biosynthesis of bleomycin is completed by the attachment of the carbohydrates (L-gulose and 3-*O*-carbamoyl-D-mannose) to the aglycon. The sequence of assembling the modified sugars, and of their attachment to the aglycon, is still unknown. According to sequenced sugar biosynthesis genes (*blmC*, *blmD*, *blmE*, *blmF*, *blmG*) and to their deduced functions, as yet only a proposal for the pathway has been made starting with D-mannose. The above-mentioned findings roughly reflect the current status of knowledge on the biosynthesis of bleomycins. The sequence of the gene cluster represents considerable progress in understanding of the biosynthesis, however, some work has still to be performed on the level of protein characterization.

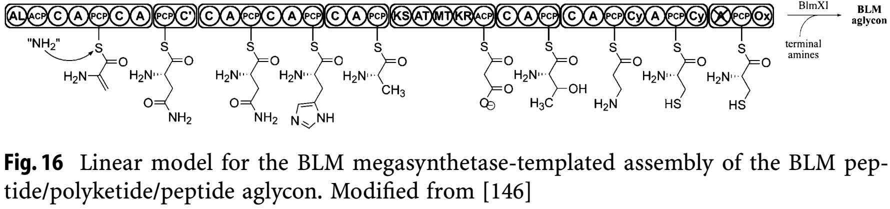


Fig. 16 Linear model for the BLM megasynthetase-templated assembly of the BLM peptide/polyketide/peptide aglycon. Modified from [146]

5.3
Mechanism of Nucleic Acid Cleavage and Structure–Activity Relationships

It is thought that the biological activity of bleomycin is based on a sequence-selective, metal-dependant oxidative cleavage of DNA (ds, ss), RNA (*t*RNA, *m*RNA, *r*RNA) and DNA-RNA hybrids in the presence of oxygen [129, 130]. Several pathways are discussed for this catalytic process. In both cases the presence of a redox-active metal ion, preferably Fe^{2+}, is a prerequisite for DNA attack [130]. Octahedrally coordinated Fe^{2+} and molecular oxygen serve as cofactors in bleomycin activation and represent essential characteristics for the biological activity of bleomycin. The oxidative cleavage of DNA by bleomycin involves the activation of the Fe(II)–bleomycin complex with oxygen (Fig. 17). Subsequently, this complex binds to DNA, probably intercalating by its dithiazol moiety to the minor grove. The cationic tail provides additional ionic interactions with the negatively charged DNA backbone. Then DNA degradation steps occur with H-atom abstraction from C-4′ of ribose. Analysis of degradation products revealed a selectivity to $^{5'}GC^{3'}$ and $^{5'}GT^{3'}$ sites. According to the O_2-dependant mechanism, an aldehyde base is formed from DNA cleavage. Bleomycin is formally oxidized under secession of a hydroxyl ion. The intermediate compound is $[BLMFe=O]^{V}$, which can react with DNA under formation of a DNA-C4′ radical and $[BLMFe=O]^{IV}$. In the presence of molecular oxygen, the DNA radical is readily oxidized to a DNA peroxyl radical species. Finally, under reconstitution of $[BLMFe=O]^{V}$ a DNA peroxide is generated, which degrades under DNA cleavage and base propenal formation.

Alternatively, the base is liberated in an O_2-independent pathway: a DNA-C4′ cation is formed from a DNA radical under reduction of $[BLMFe=O]^{IV}$ to $BLMFe^{III}$ and formation of a hydroxyl anion. The DNA-carbocation reacts with water and releases free nucleic base and renders the DNA strand alkali-labile.

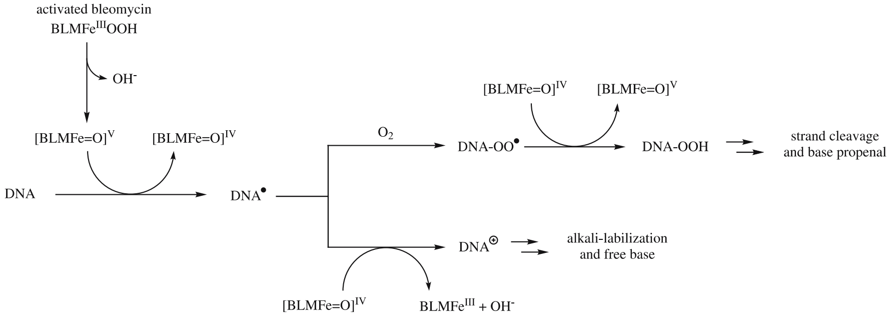


Fig. 17 Models for reaction pathways for DNA cleavage under the influence of an activated bleomycin species

The improvements in synthetic methodology of bleomycin analogues by the groups of Hecht [138, 139, 150–152] and Boger [126, 140] were the base for extensive and systematic investigations of these derivatives in structure–activity relationship studies. Besides spectroscopic analyses, the investigation of synthetic analogues shaped a deeper understanding of the mechanism of action and biological activity of bleomycins. With regard to the mechanism of action four bleomycin subunits are commonly defined. This division is based on the individual contributions of structural moieties to the biological effectiveness:

5.3.1
The Metal-Binding Domain

The metal-binding domain is represented by the *N*-terminal part with PBA and β-OH-His and determines the sequence selectivity of DNA cleavage [126], which was elucidated by alteration of *C*-terminal substituents as well as structural alterations of the metal-binding domain. The octahedral coordination of the transition metal, preferably Fe^{2+}, is performed by the *N*-terminal part of bleomycin. Interestingly, in contrast to Fe^{3+}-specific siderophores, the coordinating atoms are predominantly nitrogens: The equatorial ligands are the secondary amine of diaminopropionic acid (Dpr), the amide nitrogen of β-OH-His, the N5 nitrogen of pyrimidine, and the N1 nitrogen of imidazole (Fig. 14). From X-ray diffraction studies, the primary amine of Dpr was identified as an axial ligand [153]. Participation of nitrogen [154] and oxygen [155] from the carbamoyl group of the carbohydrate moiety serving as axial ligands is discussed.

5.3.2
The Linker Domain

The linker domain is supposed to provide a conformation that facilitates DNA-cleavage. It connects the pyrimidoblamic acid moiety with the *C*-terminal part of bleomycin. Initial studies showed that substituent variations and their absolute stereochemistry in this compartment have critical influence on the efficiency of DNA cleavage [156, 157]. Extensive variation studies in linker length and linker substitutions [126] have contributed to an understanding of the influence of the linker domain on catalytic efficiency, specificity and strand selection.

5.3.3
The *C*-Terminal DNA-Binding Domain

The bithiazole moiety and the *C*-terminal tail are involved in DNA binding. The removal of this *C*-terminal substituent or substituents that lack the positive charge leads to a significant decrease in the efficiency of DNA cleavage [158–161]. The uncharged derivatives of bleomycin A_2, a sulfoxide and the methyl thioether have up to tenfold less efficiency in DNA cleavage.

5.3.4
The Carbohydrate Domain

The influence of the carbohydrate domain on the activity of BLMs is not yet completely understood. The dependence of the efficiency of DNA cleavage on the carbohydrate moiety were studied with synthetic compounds [126]. The terminal sugar of the disaccharide moiety in bleomycin A_2, 2-*O*-(3-*O*-carbamoyl)-α-D-manno-pyranoside has only a negligible influence on the selectivity of DNA cleavage. In contrast, the lack of the gulose residue seemed to decrease efficiency and the ratio of ds-DNA and ss-DNA cleavage.

In summary, the mode of action of bleomycin with regard to the selectivity, specificity and efficiency of the cleavage reactions is rather complex. In the past, bleomycin has served as an example for a number of bleomycin models [162] but, to our knowledge, application of these models or of semisynthetic derivatives for clinical use has never been achieved.

6
Salmochelin

Iron is an essential element for the metabolism of all organisms and serves as the catalytic centre of a considerable number of enzymes, e.g. P450 monooxygenases. However, the availability of iron in nature is relatively low. In mammals, all iron is tightly bound to enzymes or to transport and storage proteins (e.g. transferrins), which lower iron concentration to an estimated 10^{-24} M. Siderophores are iron chelators that are produced and excreted by many bacteria under low iron media conditions. Siderophores have extremely high complex-forming constants in the range of 10^{23}–10^{52} M^{-1}. Upon complexation of Fe^{3+} in the extracellular space, siderophores are taken up by the bacterial cell and iron is liberated by hydrolysis of the complex [163, 164]. Therefore, it is possible to assign siderophores the role of pathogenic factors in bacterial or fungal infections. The two main classes of siderophores are of the catecholate-type and of the hydroxamate-type.

Enterobactin (Fig. 18), produced by *E. coli* and *Salmonella* strains [165, 166], is a characteristic catecholate-type siderophore and is probably the most prominent representative among all siderophores known to date. The enterobactin structure is not a peptide, rather a cyclic trilactone. Three L-Ser are linked in the hydroxymethyl side chains with the carboxyl group of a subsequent L-Ser residue. The three remaining *N*-termini are amidated with 2,3-dihydroxybenzoic acid (Dhb) [163]. Salmochelin (diglycosyl enterobactin, DGE) (Fig. 18) is a siderophore of the enterobactin-type, which was recently isolated from *Salmonella enterica* and *E. coli* strains [167, 168]. Salmochelins were named after the *Salmonella* strains they were first isolated from, in analogy to enterobactin, which was isolated from *Enterobacter* strains.

The recent structure elucidation of salmochelin by means of 2D-NMR experiments revealed a twofold *C*-glycosylated enterobactin [168]. Furthermore, with salmochelins S2, S1 and SX a number of putative degradation products of the parent compound were characterized. In all cases, 1-desoxyglucose is attached via a *β*-*C*-glycosidic bond to the 5-position of Dhb.

As for the above-mentioned glycopeptide antibiotics, enterobactin is synthesized by NRPS (Fig. 19). Acquirement of the *iroA* gene cluster transforms a bacterial enterobactin producer into a salmochelin-producing strain. The *ent/fep* locus codes for the assembly of enterobactin, which is under the control of iron-dependant repressor Fur [169]. The *iroA* locus [167] is responsible for the modification of enterobactin to yield salmochelin and also codes for some proteins involved in the secretion, uptake and degradation of salmochelin (Fig. 19).

To date, the enterobactin biosynthesis is probably the most thoroughly investigated biosynthetic pathway among all siderophores [164, 169]. 2,3-Dihydroxybenzoic acid is synthesized in a three-step synthesis by EntC/B/A from chorismate. EntB is a two-domain protein and also acts as a Dhb-carrier protein, whereas EntF is a peptidyl-carrier protein with C-A-T-Te domain organization. In a first step, both proteins are phosphopantetheinylated by EntD. Then EntB is charged with Dhb by AMP-ligase EntE. In parallel, the T-domain of EntF is charged with L-Ser by its own AMP-ligase (represented by the A-domain of EntF). In a subsequent step, the *C*-domain of EntF forms an amide between Dhb and the *N*-terminus of L-Ser, which liberates EntB from its substrate. The resulting Dhb-L-Ser is stored as an ester on the Te-domain of EntF. This procedure is repeated another two times and finally the

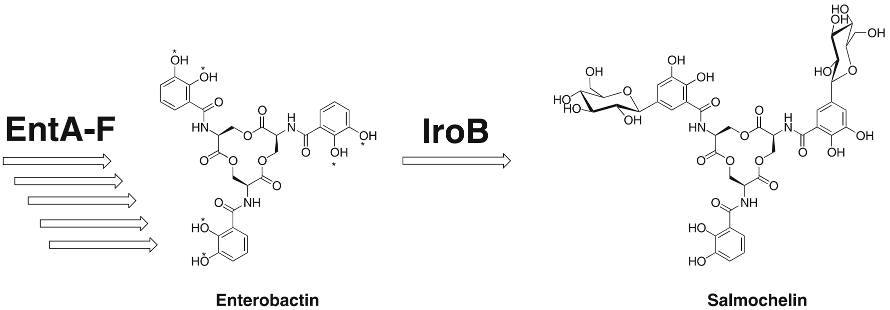


Fig. 18 Structures of enterobactin and salmochelin, and the biosynthetic proteins involved in biosynthesis. *Asterisks* mark the coordination sites for Fe^{3+}

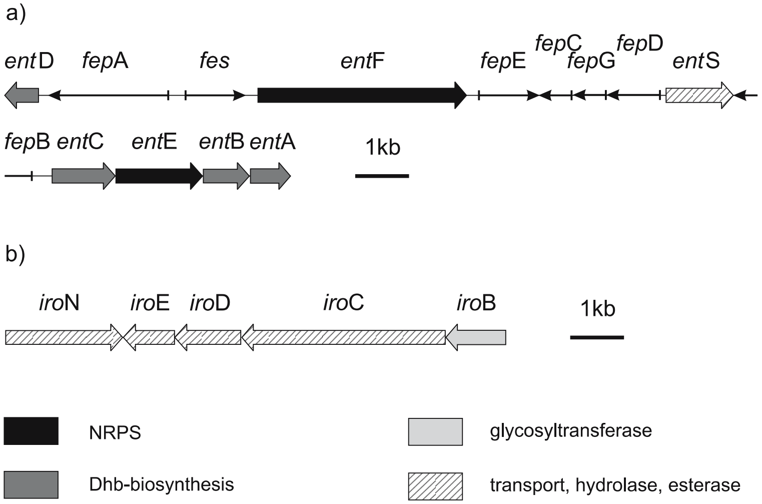


Fig. 19 Biosynthesis gene clusters of enterobactin and salmochelin. **a** Organization of the *ent/fep*-gene cluster of Gram-negative *E. coli* **b** Organization of the *iroA*-gene cluster of Gram-negative *Salmonella enterica*

resulting linear ester is cyclized head-to-tail under cleavage from EntF. Subsequent steps comprise the export and import of the charged Fe^{3+}–enterobactin complex. EntS, also known as YdbA, is the exporting protein of enterobactin into the periplasmic space. Further details on the export mechanism into the medium are unknown. FepA is the Fe^{3+}–enterobactin receptor in the outer membrane of *E. coli*. FebB binds the Fe^{3+}–enterobactin complex in the periplasm. The complex is then transported through the cytoplasmic membrane by FebD/G and ABC-transporter FebC into the cytoplasm. Fes is the hydrolase for cleavage of the Fe^{3+}–enterobactin complex. Regulation of the transcription of *ent/fep* genes is iron-dependant. In this context, Fe^{2+}-binding Fur-proteins (ferric uptake iron regulators) play a central role for suppression of gene expression under high iron conditions.

The *iroA* locus (Fig. 19) consists of five genes *iroN* and *iroBCDE*, which have originally been described for a number of *Salmonella enterica* strains [170, 171]. According to sequence comparison, the following functions have been assigned to the gene products of *iroA* [167, 168, 172]: IroN – outer membrane siderophore receptor, IroD – esterase, IroE – hydrolase, IroC – export protein, IroB – glycosyltransferase. Deductions from sequence data were supported by gene inactivations of *iroA* genes [167]. With *iroN* mutants it was shown that IroN is the main salmochelin receptor and *iroB*-inactivated mutants only produced enterobactin, assigning the function of IroB as a glycosyltransferase [168]. Complementation with *iroA* genes again partially or completely restored biosynthesis and function of salmochelins.

The function of glycosyltransferase IroB has recently been confirmed by overexpression in *E. coli* [173]. In vitro studies howed that most likely enterobactin serves as the natural substrate, with UDP-glucose as the glycosyl donor. Control experiments with IroB and the enterobactin monomer Dhb-serine did not yield any substrate conversion. Interestingly, the authors also observed monoglycosyl enterobactin (MGE) and triglycosyl enterobactin (TGE) in minor amounts. In contrast, in culture media of *Salmonella* strains only the diglycosyl form of salmochelin has been described [168], and this finding may be explained by rapid export of salmochelin after diglycosylation.

The function and activity of IroD, IroE and Fes has been recently investigated in detail within the context of enterobactin and salmochelin degradation [174, 175]. Extensive kinetic data (k_{cat}, K_D) of overexpressed proteins have been generated by assaying enterobactin and its variously *C*-glycosylated analogues MGE, DGE, TGE as well as their Fe^{3+} forms [174]. Fes was assigned the role of hydrolysing Fe^{3+}–enterobactin because of its lowered enzymatic efficiency for *apo*-enterobactin. Kinetic data on glycosylated forms of enterobactin are negligable. IroD preferably catalyses the hydrolysis of all Fe^{3+} forms of glycosylated and non-glycosylated enterobactins, particularly of diglycolylsalmochelin. Furthermore, in contrast to IroE, IroD hydrolyses the linear trimer forms down to dimer and monomer units. Regioselectivity has been determined. IroE shows a significantly greater catalytic efficiency for the *apo* forms. In conclusion, IroD is the Fes analogue for the salmochelin–iron uptake pathway. Similarly, Hantke et al. [175] suggest a periplasmic function of IroE and a cytoplasmic function of IroD. In contrast, however, IroC would act as a receptor in the periplasmic membrane for linearized salmochelins rather than act as a salmochelin export protein. In conclusion, there still exist some ambiguities on the biosynthetic picture which need to be resolved. With respect to the pathogenic relevance, it might be that salmochelins respectively the presence of *iroA* genes increase the virulence of strains [170, 176, 177] compared to enterobactins. The elucidation of the role of salmochelin is the subject of current investigations.

7
Conclusions

The previous chapters have presented selected aspects of the biosynthesis, the mode of action, and the structure–activity relationships of the most representative glycopeptide antibiotics. The subsequent comparison provides insights into basic principles and highlights the interesting parallels and differences of this compound class.

Among all glycopeptides, cyclic forms dominate with macrolactone, macrolactam or biaryl/diarylether linkages. This fact underscores the significance of conformational rigidity for the exertion of antibiotic activity, e.g. as shown for ramoplanin [87] and biosynthetic intermediates of vancomycin-type glycopeptides [48, 49, 76]. The structure of bleomycin is linear, however, only in its metal-free form. In the presence of Fe^{2+}, for example, the *N*-terminal part of the molecule is winded around the metal centre to yield a conformationally fixed metal complex.

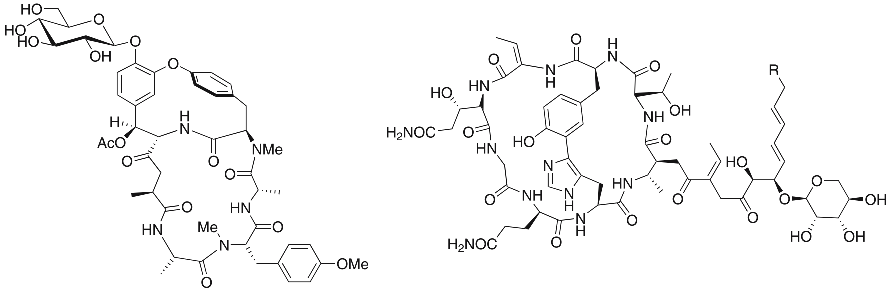
A deeper consideration of the glycopeptide antibiotic structures shows an unusually high content of certain amino acids, which can be termed "lead amino acids". In general, aromatic amino acid residues are clearly overrepresented. Among these, 4-hydroxyphenylglycine (Hpg) is an essential structural element of vancomycin-related antibiotics as well as of ramoplanin. Other aromatic amino acids are Phe, Tyr/β-Hty, Dpg, His, β-MePhe, as well as pyrimidine and thiazylthiazole derivatives. Salmochelin consists of aromatic Dhb as a major constituent. Similarly, the presence of β-hydroxy amino acids is a characteristic of all glycopeptide structures. This is exemplified by proteinogenic Ser in mannopeptimycins and salmochelins, and of threonines in ramoplanin and bleomycin. But, diverse non-proteinogenic β-hydroxy amino acids are also represented. A curiosity among β-hydroxy amino acids is β-hydroxyenduracididin (Aiha), from mannopeptimycins, and ramoplanin-related enduracidin [93] with a guanidino side chain reminiscent of Arg. The β-hydroxy group of β-OH-Asn forms the macrolactone with the *C*-terminus of the ramoplanin peptide. In salmochelins the β-hydroxy group of Ser forms the triseryl macrolactone. Among aromatic β-hydroxy amino acids, β-Hty is characteristic of vancomycin-related glycopeptide antibiotics and β-OH-His of bleomycin. For the latter two antibiotics the β-hydroxy functions serve as attachment sites for carbohydrates. Moreover, the free hydroxy groups might contribute to hydrogen bonding and/or enhanced water solubility.

In contrast to the amino acids as constituents of glycopeptide antibiotics, the assignment of characteristic "lead carbohydrates" appears to be more difficult. From a comprehensive overview of the currently known vancomycin-related glycopeptides (type I–IV) [7–9], the amino sugars vancosamine and eremosamine dominate. Furthermore, rhamnose, glucose and mannose essentially contribute to the glycosylation of this glycopeptide family, but an enormous variety of other carbohydrates have also been described. In other glycopeptide antibiotics mannose is mostly the dominating carbohydrate constituent. With regard to the glycosylation sites, it is interesting to note that three types are represented within the group of glycopeptide antibiotics: *O*-glycosylation clearly dominates (aromatic versus aliphatic hydroxyl) over *N*-glycosylation (mannopeptimycin) and *C*-glycosylation (salmochelin). However, in some cases glycosylation seems to be derived from the secondary metabolism, e.g. in the case of ramoplanin no glycosyltransferase was identified in the gene cluster.

An important question in this context arises concerning the function of carbohydrates for biological activity. It has been shown that glycosylation has a strong influence on the antibiotic activity of vancomycin-related glycopeptides, i.e. dimerization as the base for an allosteric binding of the D-Ala-D-Ala-peptide motif of the bacterial cell wall [25]. In the case of teicoplanin and mannopeptimycins the sugars constitute attachment sites for fatty acids, which strongly influence antibiotic activity. The glycosylation of salmochelin is the base for specific recognition and uptake by the receptor IroN [167]. In contrast, deglycosyl ramoplanin shows the same antibiotic activity as ramoplanin, thus disproving a significant contribution of carbohydrates to antibiotic effects [107, 108]. In any case, glycosylations increase the water solubility of glycopeptides, although likely signalling functions (serving to cell-cell communication) have not clearly been shown. Further work has to be performed in order to shed more light on the role of glycosylation.

With the exception of salmochelin, produced by Gram-negative *Salmonella* and *E. coli* strains, glycopeptide antibiotics have been found to be mainly produced by Gram-positive actinomycete strains. Only recently has hassallidin A [178] been isolated and characterized from cyanobacteria. Other glycopeptides (e.g. the antitumour glycopeptide bouvardin) have been isolated from the plant *Bouvardia ternifolia* (Fig. 20) [179], and antifungal aciculitins (Fig. 20) [180] and theopalauamide [181] have been described from sponges. The latter findings indicate a broader distribution among different organisms and it is likely that novel glycopeptide antibiotics will be discovered in the future.

So far as is known, the common biosynthetic origin of the glycopeptide antibiotics presented here is the non-ribosomal peptide synthesis with activation (A), thiolation (T), condensation (C) and thioesterase (Te) domains [2]. However, some deviations among biosynthetic clusters are apparent since (e.g. in the case of ramoplanin) E-domains for the epimerization of (*S*)- into their (*R*)-peptidyl counterparts are missing. The bleomycin biosynthesis shows the biggest variety of NRPS domains occurring in genes coding for NRPS. Cyclization (Cy) and oxidation (Ox) domains provide biosynthetic



Bouvardin RA-XVI

Aciculitin A R = C_5H_{11}

Fig. 20 Structures of glycopeptides bouvardin and acicultin

Table 1 Comparison of structural and biological features of the glycopeptides vancomycin, ramoplanin, mannopeptimycin, bleomycin and salmochelin

	Vancomycin	Ramoplanin	Mannopeptimycin	Bleomycin	Salmochelin
Producing organism	*Amycolatopsis orientalis*	*Actinoplanes* strains	*Streptomyces hygroscopius*	*Streptomyces verticillus*	Uropathogenic *E.coli* strains, *Salmonella* strains
	Gram-positive	Gram-positive	Gram-positive	Gram-positive	Gram-negative
Backbone	Aromatic crosslinked heptapeptide	Macrocyclic depsipeptide with lactone linkage	Cyclic hexapeptide	Linear peptide-ketide backbone	Amino acid trilactone
Macrocycles	Biaryl linkage biarylether linkage	Lactone linkage	Lactam linkage	"Metallocycle"	Cyclic trilactone
Type of carbohydrate	*O*-Glycosidic	*O*-Glycosidic	*O*-Glycosidic and *N*-glycosidic	*O*-Glycosidic	*C*-Glycosidic
Number of carbohydrate residues	2	3	3	2	2
Molecular mass	~ 1450	~ 2550	~ 1600	~ 1500	~ 1000
Biological activity	Antibacterial	Antibacterial	Antibacterial	Antitumour (antibacterial) (antifungal)	(Antibacterial)
Molecular target	Peptidoglycan biosynthesis, D-Ala-D-Ala	Peptidoglycan biosynthesis, lipid I, lipid II	Peptidoglycan biosynthesis, lipid II	DNA cleavage	Fe^{3+}
Structurally related compounds	Balhimycin, avoparcin, ristocetin teicoplanin	Enduracidin		Phleomycin tallysomycin	Enterobactin

means for the assembly of thiazylthiazole. Furthermore, the biosynthetic assembly of bleomycin peptides is based on a mixed NRPS/PKS/NRPS system, unique among the glycopeptide antibiotics thus far described. In contrast, side-chain cross-linking of aryls in vancomycin-type glycopeptides are performed by P450-type monooxygenases acting *in trans* [46, 52].

The biosynthesis of enterobactin as the precursor for salmochelin has been shown to occur in an iterative way [2, 169]. Three single structural Dhb-Ser units are iteratively linked and finally macrolactonized by a Te domain. A macrolactonization has to be assumed also for ramoplanin. With regard to biocombinatorial approaches, vancomycin-type glycopeptide models are most advanced with mutasynthesis [73, 75] and the characterization of glycosyltransferases and attempts to generate novel glycopeptides [57].

The identified targets of glycopeptide antibiotics are subject to three fields of action. The action as peptidoglycan biosynthesis inhibitors clearly dominates the targeting of Gram-positive bacteria, e.g. staphylococci and enterococci. These findings are in accordance with the molecular architecture of glycopeptide antibiotics which, due to their polarity and high molecular mass, are unable to cross the outer membrane of Gram-negative bacteria. Furthermore, glycopeptides are considered still too big in size to cross the plasma membrane of Gram-positive organisms in order to enter the cytoplasm. With regard to the current knowledge on cell wall biosynthesis, it is mainly the final stages of peptidoglycan formation (i.e. transglycosylation and transpeptidation reactions) that are inhibited by vancomycin, teicoplanin, ramoplanin and mannopeptimycins. However, the exact targets and their mechanistic inhibition on a molecular base are subtle and are, at least in some cases, still under debate. Muraymycins are antibiotics with structural relationships and related targets to glycopeptide antibiotics [182]. Structurally they can be considered as nucleoside-lipopeptide or nucleoside-lipoglycopeptide antibiotics. Remarkably, the above-mentioned murayamycins [182] and mureidomycins have been reported as inhibitors of translocase MraY involved in late-stage peptidoglycan biosynthesis. Semi-synthesis and SAR studies were performed in order to evaluate their antibiotic potential [183].

Bleomycin and its mode of action is an outsider among other glycopeptide antibiotics. The tripartite molecule unifies a catalytic oxidative and a DNA-intercalating function. As a consequence, bleomycin also has some antibacterial (Gram-negative and Gram-positive) and antifungal activity [184, 185]. However, despite its clinical use for cancer therapy, to our knowledge the mode of crossing the cell membrane is elusive. It is presumed that the carbohydrate moieties might aid this process.

The siderophore salmochelin has a transport function for Fe^{3+}. Its duty is to secure the iron-need under low iron conditions from the extracellular space. Like many transport molecules it possesses a specific receptor in Gram-negative bacteria. The deletion of this transporter inhibits salmochelin uptake and compromises the viability of thus-mutated strains [167].

Of the antibacterial glycopeptides, only vancomycin (USA/Europe) and teicoplanin (Europe) are used as anti-infective drugs in humans. Ramoplanin is currently under investigation for clinical applications and it might well be that it reaches the market in the near future. The re-examination of mannopeptimycins, which were put on hold in drug development for several years, implies a strong consideration for their pharmaceutical use as an alternative to vancomycin, or for the treatment of vancomycin-resistant infections. Bleomycin has been used in cancer therapy for decades. However, due to its strong toxicity and side effects at the moment there is no increase in significance expected.

The urgent need for new antibacterials in clinical applications is due to the increase of nosocomial (hospital-acquired) infections by resistant pathogens, notably methicillin-resistant *S. aureus* (MRSA), vancomycin-resistant *E. faecium* (VRE) and vancomycin-resistant *S. aureus* (VRSA). One approach is based on the exploitation of glycopeptide structures as lead structures in order to develop antibacterials with stronger action over an extended spectrum of antibacterial activity, e.g. VRSA. Of all the glycopeptide antibiotics, these investigations and developments are most advanced for vancomycin-type and teicoplanin-type antibiotics, combined with recent progress in the understanding of the mode of action [80, 81]. Structure–activity relationship studies have led to oritavancin, telavancin and dalbavancin, which are in the late phases of clinical evaluation and which should reach clinical application.

In summary, glycopeptide antibiotics represent an interesting and highly important natural product family. With regard to structure elucidation and organic synthesis, glycopeptides have posed significant challenges to researchers in the past. Currently, the understanding of their biosynthesis sets the stage for future biocombinatorial applications for the production of novel glycopeptide antibiotics. From the viewpoint of a bioorganic chemist, their structures and their diverse modes of action reveal a great aesthetics and beauty.

Acknowledgements This work was supported by a grant of the European Union (COMBIG-TOP, LSHG-CT-2003-503491), the Deutsche Forschungsgemeinschaft (DFG, SU 239/3-3) and by an Emmy-Noether-Fellowship for young investigators of the DFG (SU 239/2-2).

References

1. Sharon N (1986) Eur J Biochem 159:1
2. Sieber SA, Marahiel MA (2005) Chem Rev 105:715
3. McCormick MH, Stark WM, Pittenger GE, Pittenger RC, McGuire JM (1955-1956) Antibiot Annu 606
4. Sheldrick GM, Jones PG, Kennard O, Williams DH, Smith GA (1978) Nature 271:223
5. Williams DH, Kalman JR (1977) J Am Chem Soc 99:2768
6. Harris CM, Harris TM (1982) J Am Chem Soc 104:4293

7. Nicolaou KC, Natarajan S, Li H, Jain NF, Hughes R, Solomon ME, Ramanjulu JM, Boddy CNC, Takayanagi M (1998) Angew Chem 110:2872, Angew Chem Int Ed 37:2708
8. Nicolaou KC, Jain NF, Natarajan S, Hughes R, Solomon ME, Li H, Ramanjulu JM, Takayanagi M, Koumbis AE, Bando T (1998) Angew Chem 110:2879, Angew Chem Int Ed 37:2714
9. Nicolaou KC, Takayanagi M, Jain NF, Natarajan S, Koumbis AE, Bando, Ramanjulu JM (1998) Angew Chem 110:2881, Angew Chem Int Ed 37:2717
10. Evans DA, Wood MR, Trotter BW, Richardson TI, Barrow JC, Katz JL (1998) Angew Chem 110:2864, Angew Chem Int Ed 37:2700
11. Boger DL, Miyazaki S, Kim SH, Wu JH, Loiseleur O, Castle SL (1999) J Am Chem Soc 121:3226
12. Nicolaou KC, Boddy CNC, Bräse S, Winssinger N (1999) Angew Chem 111:2230, Angew Chem Int Ed 38:2096
13. Nicolaou KC, Snyder SA (2003) Classics in total synthesis II. Wiley, Weinheim
14. Boger DL, Kim SH, Mori Y, Weng JH, Rogel O, Castle SL, McAtee JJ (2001) J Am Chem Soc 123:1862
15. Matsuzaki K, Ikeda H, Ogino T, Matsumoto A, Woodruff HB, Tanaka H, Omura S (1994) J Antibiot 47:1173
16. Perkins HR (1969) Biochem J 111:195
17. Williams DH, Williamson MP, Butcher DW, Hammond SJ (1983) J Am Chem Soc 105:1332
18. Nieto A, Perkins HR (1971) Biochem J 123:773
19. Waltho JP, Williams DH (1989) J Am Chem Soc 111:2475
20. Batta G, Sztaricskai F, Kover KE, Rudel C, Berdnikova TFJ (1991) J Antibiot 44:1208
21. Loll PJ, Bevivino AE, Korty BD, Axelsen PH (1997) J Am Chem Soc 119:1516
22. Mackay JP, Gerhard U, Beauregard DA, Westwell MS, Searle MS, Williams DH (1994) J Am Chem Soc 116:4581
23. Williams DH, Maguire AJ, Tsuzuki W, Westwell MS (1998) Science 280:711
24. Shiozawa H, Chia BCS, Davies NL, Zerella R, Williams DH (2001) J Am Chem Soc 124:3914
25. Williams DH, Bardsley B (1999) Angew Chem 111:1264, Angew Chem Int Ed 38:1172
26. Uttley AH, Collins CH, Naidoo J, George RC (1988) Lancet 1:57
27. Hiramatsu K, Hanaki H, Ino T, Yabuta K, Oguri T, Tenover FC (1997) J Antimicrob Chemother 40:135
28. Cetinkaya Y, Falk P, Mayhall CG (2000) Clin Microbiol Rev 13:686
29. Allen NE, Nicas TI (2003) FEMS Microbiol Rev 26:511
30. Hubbard BK, Walsh CT (2003) Angew Chem 115:752, Angew Chem Int Ed 42:730
31. Süssmuth RD, Wohlleben W (2004) Appl Microb Biotech 63:344
32. Pelzer S, Reichert W, Huppert M, Heckmann D, Wohlleben W (1997) J Biotech 56:115
33. Pelzer S, Süssmuth RD, Heckmann D, Recktenwald J, Huber P, Jung G, Wohlleben W (1999) Antimicrob Agents Chemother 43:1565
34. van Wageningen AMA, Kirkpatrick PN, Williams DH, Harris BR, Kershaw JK, Lennard NJ, Jones M, Jones SJM, Solenberg PJ (1998) Chem Biol 5:155
35. Li TL, Huang F, Haydock SF, Mironenko T, Leadlay PF, Spencer JB (2004) Chem Biol 11:107
36. Sosio M, Kloosterman H, Bianchi A, de Vreugd P, Dijkhuizen L, Donadio S (2004) Microbiology 150:95
37. Chiu HT, Hubbard BK, Shah AN, Eide J, Fredenburg RA, Walsh CT, Khosla C (2001) Proc Natl Acad Sci USA 98:8548

38. Choroba OW, Williams DH, Spencer JB (2000) J Am Chem Soc 122:5389
39. Hubbard BK, Thomas MG, Walsh CT (2000) Chem Biol 42:1
40. Li TL, Choroba OW, Hong H, Williams DH, Spencer JB (2001) J Chem Soc Chem Commun 20:2156
41. Pfeifer V, Nicholson GJ, Ries J, Recktenwald J, Schefer AB, Shawky RM, Schröder J, Wohlleben W, Pelzer S (2001) J Biol Chem 276:38370
42. Chen H, Tseng CC, Hubbard BK, Walsh CT (2001) Proc Natl Acad Sci USA 98:14901
43. Puk O, Bischoff D, Kittel C, Pelzer S, Weist S, Stegmann E, Süssmuth RD, Wohlleben W (2004) J Bacteriol 186:6093
44. Puk O, Huber P, Bischoff D, Recktenwald J, Jung G, Süssmuth RD (2002) Chem Biol 9:225
45. Chen H, Walsh CT (2001) Chem Biol 74:1
46. Bischoff D, Bister B, Bertazzo M, Pfeifer V, Stegmann E, Nicholson GJ, Keller S, Pelzer S, Wohlleben W, Süssmuth RD (2005) ChemBioChem 6:267
47. Chen H, Thomas MG, Hubbard BK, Losey HC, Walsh CT, Burkart MD (2000) Proc Natl Acad Sci USA 96:11942
48. Bischoff D, Pelzer S, Höltzel A, Nicholson GJ, Stockert S, Wohlleben W, Jung G, Süssmuth RD (2001) Angew Chem 113:1736, Angew Chem Int Ed 40:1693
49. Bischoff D, Pelzer S, Bister B, Nicholson GJ, Stockert S, Schirle M, Wohlleben W, Jung G, Süssmuth RD (2001) Angew Chem 113:4824, Angew Chem Int Ed 40:4688
50. Pylypenko O, Vitali F, Zerbe K, Robinson JA, Schlichting I (2003) J Biol Chem 278:46727
51. Zerbe K, Pylypenko O, Vitali F, Zhang W, Rouset S, Heck M, Vrijbloed JW, Bischoff D, Bister B, Süssmuth RD, Pelzer S, Wohlleben W, Robinson JA, Schlichting I (2002) J Biol Chem 277:47476
52. Zerbe K, Woithe K, Li DB, Vitali F, Bigler L, Robinson JA (2004) Angew Chem 116:6877, Angew Chem Int Ed 43:6709
53. Mulichak AM, LW, Losey HC, Walsh CT, Garavito RM (2004) Biochemistry 43:5170
54. Mulichak AM, Losey HC, Lu W, Wawrzak Z, Walsh CT, Garavito RM (2003) Proc Natl Acad Sci USA 100:9238
55. Mulichak AM, Losey HC, Walsh CT, Garavito RM (2001) Structure 9:547
56. Lu W, Oberthür M, Leimkuhler C, Tao J, Kahne D, Walsh CT (2004) Proc Natl Acad Sci USA 101:4390
57. Solenberg PJ, Matsushima P, Stack DR, Wilkie SC, Thompson RC, Baltz RH (1997) Chem Biol 4:195
58. Oberthür M, Leimkuhler C, Kruger RG, Lu W, Walsh CT, Kahne D (2005) J Am Chem Soc 127:10747
59. Losey HC, Jiang J, Biggins JB, Oberthür M, Ye XY, Dong SD, Kahne D, Thorson JS, Walsh CT (2002) Chem Biol 9:1305
60. Losey HC, Peczuh MW, Chen Z, Eggert US, Dong SD, Pelczer I, Kahne D, Walsh CT (2001) Biochemistry 40:4745
61. O'Brien DP, Kirkpatrick PN, O'Brien SW, Staroske T, Spencer JB, Williams DH (2000) J Chem Soc Chem Commun 1:103
62. Feliu L, Planas M (2005) Int J Pept Res Therap 11:53
63. Preobrashenskaya MN, Olsufyeva EN (2004) Expert Opin Ther Patents 14:141
64. Shi Z, Griffin JH (1993) J Am Chem Soc 115:6482
65. Pavlov AY, Berdnikova TF, Olsufyeva EN, Miroshnikova OV, Filiposyanz ST, Preobrazhenskaya MN, Scottani C, Colombo L, Goldstein BP (1996) J Antibiot 49:149
66. Miroshnikova OV, Printsevskaya SS, Olsufyeva EN, Pavlov AY, Nilius A, Hensey-Rudloff D, Preobrazhenskaya MN (2000) J Antibiot 53:286

67. Sundram UN, Griffin JH (1995) J Org Chem 60:1102
68. Kenny MT, Brackman MA, Dulworth JK (1995) Antimicrob Agents Chemother 39:1589
69. Goldstein BP, Candiani G, Arain TM (1995) Antimicrob Agents Chemother 39:1580
70. Nicolaou KC, Hughes R, Cho SY, Winssinger N, Smethurst C, Labischinski H, Endermann R (2000) Angew Chem 112:3981, Angew Chem Int Ed 39:3823
71. Williams DH, Staroske T (1998) Tetrahadron Lett 39:4917
72. Rao J, Lahiri J, Isaacs L, Weis RM, Whitesides GM (1998) Science 280:708
73. Weist S, Bister B, Puk O, Bischoff D, Nicholson G, Stockert S, Wohlleben W, Jung G, Süssmuth RD (2002) Angew Chem 114:3531; Angew Chem Int Ed 40:3383
74. Bister B, Bischoff D, Nicholson GJ, Stockert S, Wink J, Brunati C, Donadio S, Pelzer S, Wohlleben W, Süssmuth RD (2003) Chem Bio Chem 4:658
75. Weist S, Kittel C, Bischoff D, Bister B, Pfeifer V, Nicholson GJ, Wohlleben W, Süssmuth RD (2004) J Am Chem Soc 126:5942
76. Süssmuth RD, Pelzer S, Nicholson GJ, Walk T, Wohlleben W, Jung G (1999) Angew Chem 111:2096, Angew Chem Int Ed 38:1976
77. Pavlov AY, Lazhko EI, Preobrazhenskaya MN (1996) J Antibiot 50:509
78. Xu R, Greiveldinger G, Marenus LE, Cooper A, Ellman JA (1999) J Am Chem Soc 121:4898
79. Alekshun MN (2005) Exp Opin Investig Drugs 14:117
80. Ge Min, Chen Z, Onishi HR, Kohler J, Silver LL, Kerns R, Fukuzawa S, Thompson C, Kahne D (1999) Science 284:507
81. Sun B, Chen Z, Eggert US, Shaw SJ, LaTour JV, Kahne D (2001) J Am Chem Soc 123:12722
82. Chen L, Walker D, Sun B, Hu Y, Walker S, Kahne D (2003) Proc Natl Acad Sci USA 100:5658
83. Walker S, Chen L, Hu Y, Rew Y, Shin D, Boger DL (2005) Chem Rev 105:449
84. McCafferty DG, Cudic P, Frankel BA, Barkallah S, Kruger RG, Li W (2002) Biopolymers 66:261
85. Cavalleri B, Pagani H, Volpe G, Selva E, Parenti F (1984) J Antibiot 37:309
86. Espersen F (1999) Curr Opin Anti-Infect Invest Drugs 1:78
87. Cudic P, Behenna DC, Kranz JK, Kruger RG, Wand AJ, Veklich YI, Weisel JW, McCafferty DG (2002) Chem Biol 9:897
88. Mobarakai N, Quale JM, Landman D (1994) Antimicrob Agents Chemother 38:385
89. Ristow TA, Noskin GA, Warren JR, Peterson LR (1995) Microb Drug Resist 1:335
90. Ciabatti R, Kettenring JK, Winters G, Tuan G, Zerilli L, Cavalleri B (1989) J Antibiot 42:254
91. Kettenring JK, Ciabatti R, Winters G, Tamborini G, Cavalleri B (1989) J Antibiot 42:268
92. Skelton NJ, Harding MM, Mortishire-Smith RJ, Rahman SK, Williams DH, Rance MJ, Ruddock JC (1991) J Am Chem Soc 113:7522
93. Higashide E, Hatano K, Shibata M, Nakazawa K (1968) J Antibiot 21:126
94. Kurz M, Guba W (1996) Biochemistry 35:12570
95. Lo M-C, Helm JS, Sarngadharan G, Pelczer I, Walker S (2001) J Am Chem Soc 123:8640
96. Farnet CM, Zazopoulos E, Ecopia SA (2002) WO Patent 0231155
97. McCafferty DG, Cudic P, Frankel BA, Barkallah S, Kruger RG, Li W (2002) Biopolymers 66:261
98. Reynolds PE, Somner EA (1990) Drugs Exp Clin Res 16:385
99. Somner EA, Reynolds PE (1990) Antimicrob Agents Chemother 34:413

100. Brötz H, Bierbaum G, Leopold K, Reynolds PE, Sahl HG (1998) Antimicrob Agents Chemother 42:154
101. Brötz H, Josten M, Wiedemann I, Schneider U, Götz F, Bierbaum G, Sahl HG (1998) Mol Microbiol 30:317
102. Lo M-C, Men H, Branstrom A, Helm J, Yao N, Goldman R, Walker S (2000) J Am Chem Soc 122:3540
103. Hu YN, Helm JS, Chen L, Ye XY, Walker SJ (2003) J Am Chem Soc 125:8736
104. Helm JS, Chen L, Walker S (2002) J Am Chem Soc 124:13970
105. Lo M-C (2000) PhD Thesis, Princeton University
106. Ciabatti R, Cavalleri B (1989) EP Patent 321696
107. Ciabatti R, Cavalleri B (1992) US Patent 5108988
108. Ciabatti R, Maffioli S, Checchia A, Romano G, Candiani G, Panzone G (2003) WO Patent 2003076460
109. Rew Y, Shin D, Hwang I, Boger DL (2004) J Am Chem Soc 126:1041
110. Ciabatti R, Cavalleri B (1989) EP Patent 337203
111. Ciabatti R, Cavalleri B (1996) US Patent 5491128
112. Jiang W, Wanner J, Lee RJ, Bonnaud P-Y, Boger DL (2003) J Am Chem Soc 125:1877
113. De Voe SE, Kunstman MP (1970) US Patent 3495004
114. He H, Williamson RT, Shen B, Graziani EI, Yang HY, Sakya SM, Petersen PJ, Carter GT (2002) J Am Chem Soc 124:9729
115. Singh MP, Petersen PJ, Weiss WJ, Janso JE, Luckman SW, Lenoy EB, Bradford PA, Testa RT, Greenstein M (2003) Antimicrob Agents Chemother 47:62
116. Ruzin A, Singh G, Severin A, Yang Y, Dushin RG, Sutherland AG, Minnick A, Greenstein M, May MK, Shlaes DM, Bradford PA (2004) Antimicrob Agents Chemother 48:728
117. Sutherland AG, Bailey AR, Cai P, Dushin RG, Liu M, Lotvin J, Ruppen ME (2004) Tetrahedron Asymm 15:2821
118. He H, Shen B, Petersen PJ, Weiss WJ, Yang HY, Wang TZ, Dushin RG, Koehn FE, Carter GT (2004) Bioorg Med Chem Lett 14:279
119. Dushin RG, Wang TZ, Sum PE, He H, Sutherland AG, Ashcroft JS, Graziani EI, Koehn FE, Bradford PA, Petersen PJ, Wheless KL, How D, Torres N, Lenoy EB, Weiss WJ, Lang SA, Projan SJ, Shlaes DM, Mansour TS (2004) J Med Chem 47:3487
120. He H, Wang TZ, Dushin RG, Feng X, Shen B, Ashcroft JS, Koehn FE, Carter GT (2004) Tetrahedron Lett 45:5889
121. Sum PE, How D, Torres N, Petersen PJ, Lenoy EB, Weiss WJ, Mansour TS (2003) Bioorg Med Chem Lett 13:1151
122. Sum PE, How D, Torres N, Petersen PJ, Ashcroft J, Graziani EI, Koehn FE, Mansour TS (2003) Bioorg Med Chem Lett 13:2805
123. Wang TZ, Wheless KL, Sutherland AG, Dushin RG (2004) Heterocycles 62:131
124. Umezawa H, Maeda K, Takeuchi T, Okami Y (1966) J Antibiot Ser A 19:200
125. Umezawa H, Suhara Y, Takita T, Maeda K (1966) J Antibiot Ser A 19:210
126. Boger DL, Cai H (1999) Angew Chem 111:470, Angew Chem Int Ed 38:448
127. Galm U, Hager MH, van Lanen SG, Ju J, Thorson JS, Shen B (2005) Chem Rev 105:739
128. Shen B, Du L, Sanchez C, Edwards DJ, Chen M, Murrell JM (2002) J Nat Prod 65:422
129. Hecht SM (2000) J Nat Prod 63:158
130. Burger RM (1998) Chem Rev 98:1153
131. Umezawa H (1971) Pure Appl Chem 28:665
132. Takita T, Muraoka Y, Yoshika T, Fujii A, Maeda K, Umezawa H (1972) J Antibiot 25:755
133. Muraoka Y, Fujii A, Yoshioka T, Takita T, Umezawa H (1977) J Antibiot 30:178

134. Takita T, Muraoka Y, Nakatani T, Fujii A, Umezawa Y, Naganawa H, Umezawa H (1978) J Antibiot 31:801
135. Fukuoka T, Muraoka Y, Fujii A, Naganawa H, Takita T, Umezawa H (1980) J Antibiot 33:114
136. Aoyagi Y, Katano K, Suguna H, Primeau J, Chang LH, Hecht SM (1982) J Am Chem Soc 104:5537
137. Takita T, Umezawa Y, Saito S, Morishima H, Naganawa H, Umezawa H (1982) Tetrahedron Lett 23:521
138. Saito S, Umezawa Y, Yoshioka T, Takita T, Umezawa H, Muraoka Y (1983) J Antibiot 36:92
139. Katano K, An H, Aoyagi Y, Overhand M, Sucheck SJ, Stevens WC Jr, Hess CD, Zhou X, Hecht SM (1998) J Am Chem Soc 120:11285
140. Boger DL, Honda T (1994) J Am Chem Soc 116:5647
141. Maeda K, Kosaka H, Yagishita K, Umezawa H (1956) J Antibiot Ser A 9:82
142. Takita T, Maeda K, Umezawa H (1959) J Antibiot Ser A 12:111
143. Takita T (1959) J Antibiot Ser A 12:285
144. Kawaguchi H, Tsukiura H, Tomita K, Konishi M, Saito KI, Kobaru S, Numata KI, Fujisawa KI, Miyaki T, Hatori M, Koshiyama H (1977) J Antibiot 30:779
145. Konishi M, Saito KI, Numata KI, Tsuno T, Asama K, Tsukiura H, Naito T, Kawaguchi H (1977) J Antibiot 30:789
146. Du L, Sánchez C, Chen M, Edwards DJ, Shen B (2000) Chem Biol 7:623
147. Du L, Shen B (1999) Chem Biol 6:507
148. Du L, Chen M, Sánchez C, Shen B (2000) FEMS Microbiol Lett 189:171
149. Sugiyama M, Davies J (1994) Gene 151:11
150. Hecht SM (2002) Eur J Cancer 38:S13
151. Rishel MJ, Hecht SM (2001) Org Lett 3:2867
152. Leitheiser CJ, Smith KL, Rishel MJ, Hashimoto S, Konishi K, Thomas CJ, Li C, McCormick MM, Hecht SM (2003) J Am Chem Soc 125:8218
153. Iitaka Y, Nakamura H, Nakatani T, Muraoka Y, Fujii A, Takita T, Umezawa H (1978) J Antibiot 31:1070
154. Lehmann TE (2002) J Biol Inorg Chem 7:305
155. Smolentsev G, Soldatov AV, Wasinger EC, Solomon EI (2004) Inorg Chem 43:1825
156. Otsuka M, Masuda T, Haupt A, Ohno M, Shiraki T, Sugiura Y, Maeda K (1990) J Am Chem Soc 112:838
157. Owa T, Haupt A, Otsuka M, Kobayashi S, Tomioka N, Itai A, Ohno M, Shiraki T, Uesugi M, Sugiura Y, Maeda K (1992) Tetrahedron 48:1193
158. Hecht SM (1995) Cancer chemotherapeutic agents. Am Chem Soc, Washington DC
159. Berry DE, Chang LH, Hecht SM (1985) Biochem 24:3207
160. Shipley JB, Hecht SM (1988) Chem Res Toxicol 1:25
161. Carter BC, Murty VS, Reddy KS, Wang SN, Hecht SM (1990) J Biol Chem 265:4193
162. Nivorozhkin AL, Girerd JJ (1996) Angew Chem 108:665; Angew Chem Int Ed 35:609
163. Raymond KN, Dertz EA, Kim SS (2003) Proc Natl Acad Sci USA 100:3584
164. Walsh CT, Liu J, Rusnak F, Sakaitani M (1990) Chem Rev 90:1105
165. Pollack JR, Neilands JB (1970) Biochem Biophys Res Commun 38:989
166. O'Brien IG, Gibson F (1970) Biochim Biophys Acta 215:393
167. Hantke K, Nicholson G, Rabsch W, Winkelmann G (2003) Proc Natl Acad Sci USA 100:3677
168. Bister B, Bischoff D, Nicholson GJ, Valdebenito M, Schneider K, Winkelmann G, Hantke K, Süssmuth RD (2004) Biometals 17:471
169. Crosa JH, Walsh CT (2002) Microbiol Mol Biol Rev 66:223

170. Bäumler AJ, Tsolis RM, van der Velden AW, Stojiljkovic I, Anic S, Heffron F (1996) Gene 183:207
171. Bäumler AJ, Norris TL, Lasco T, Voigt W, Reissbrodt R, Rabsch W, Heffron F (1998) J Bacteriol 180:1446
172. Rabsch W, Voigt W, Reissbrodt R, Tsolis RM, Baumler AJ (1999) J Bacteriol 181:3610
173. Fischbach MA, Lin H, Liu DR, Walsh CT (2005) Proc Natl Acad Sci USA 102:571
174. Lin H, Fischbach MA, Liu DR, Walsh CT (2005) J Am Chem Soc 127:11075
175. Zhu M, Valdebenito M, Winkelmann G, Hantke K (2005) Microbiology 151:2363
176. Bjarnason J, Southward CM, Surette MG (2003) J Bacteriol 185:4973
177. Konopka K, Neilands JB (1984) Biochemistry 23:2122
178. Neuhof T, Schmieder P, Preussel K, Dieckmann R, Pham H, Bartl F, von Doehren H (2005) J Nat Prod 68:695
179. Jolad SD, Hoffman JJ, Torrance SJ, Wiedhopf RM, Cole JR, Arora SK, Bates RB, Gargiulo RL, Kriek GR (1977) J Am Chem Soc 99:8040
180. Bewley CA, He H, Williams DH, Faulkner DJ (1996) J Am Chem Soc 118:4314
181. Bewley CA, Holland ND, Faulkner DJ (1996) Experiencia 52:716
182. McDonald LA, Barbieri LR, Carter GT, Lenoy E, Lotvin J, Petersen PJ, Siegel MM, Singh G, Williamson RT (2002) J Am Chem Soc 124:10260
183. Lin Y, Li Z, Francisco GD, McDonald LA, Davis RA, Singh G, Yang Y, Mansour TS (2002) Bioorg Med Chem Lett 12:2341
184. Umezawa H, Kenji M, Tomio T, Yoshiro O (1996) J Antibiot Ser A 19:200
185. Ichikawa T, Matsuda A, Miyamoto K, Tsubosaki M, Kaihara T, Sakamoto K, Umezawa H (1967) J Antibiot Ser A 20:149

Top Curr Chem (2007) 267: 187–251
DOI 10.1007/128_2006_078

Conformation of Glycopeptides and Glycoproteins

Bernd Meyer[1] · Heiko Möller[2] (✉)

[1]Institut für Organische Chemie, Universität Hamburg, Martin-Luther-King-Platz 6,
20146 Hamburg, Germany

[2]Fachbereich Chemie, Universität Konstanz, Universitätsstraße 10, 78457 Konstanz,
Germany
Heiko.Moeller@uni-konstanz.de

Abstract Glycosylation is the most frequent post-translational modification found in proteins. Glycoproteins are involved in a highly diverse spectrum of biological functions, ranging from protein folding and molecular and cellular recognition to attack of pathogens and regulation of biological half-life. Consequently, understanding the structural role of glycosylation provides key insight into biological processes as well as ideas for medicinal applications. Here, we review the current methodology to arrive at high-resolution structures of glycopeptides and glycoproteins. To achieve this, significant obstacles have to be overcome like preparation of sufficient amounts of sufficiently pure and homogeneous material, analyzing highly ambiguous spectroscopic data, and dealing with molecules that show local or global disorder. We will provide an overview of our knowledge of glycoprotein and glycopeptide structure, and illustrate with selected examples the performance and limitations of current methodology.

Keywords Glycopeptide · Glycoprotein · NMR spectroscopy · X-ray crystallography · Molecular modeling

Abbreviations

AFGPs	antifreeze glycoproteins
BEVS	baculovirus expression vectors
BMRB	BioMagResBank
CCR	cross-correlated relaxation
CCSD	Complex Carbohydrate Structure Database
CD	circular dichroism
CFG	Consortium for Functional Glycomics
EPL	expressed protein ligation
GalT	galactosyltransferase
GnT	glucosaminyltransferase
hCG	human chorionic gonadotropin
HMBC	heteronuclear multiple-bond correlation
IgSF	immune globulin superfamily
Kdo	3-deoxy-D-manno-octulosonic acid
KEGG	Kyoto Encyclopedia of Genes and Genomes
Le^y-Antigen	Lewis^y-antigen
MAS	magic angle spinning
MD	molecular dynamics
NCL	native chemical ligation
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser enhancement
PDB	Protein Data Bank
RDCs	residual dipolar couplings
SPPS	solid-phase peptide synthesis
SPR	surface plasmon resonance
ST	sialyltransferase
STF-Antigen	sialyl-Thomsen-Friedenreich-antigen
TF-Antigen	Thomsen-Friedenreich-antigen, Galβ1-3GalNAcα1-O-Ser/Thr
Tn-Antigen	Thomsen-antigen, GalNAcα1-O-Ser/Thr

1
Introduction

In the era of post genomic and post proteomic research it has been widely accepted that post-translational modifications of proteins are key to understanding their function and regulation. Glycosylation is probably the most complex subset of post-translational modifications of proteins that is in fact very difficult to analyze. The challenges are illustrated by the fact that, as of October 2005, only about 3% of the 33 500 structures deposited in the Protein Data Bank (PDB [1]) contain glycoprotein chains.

On the other hand, the relevance of research on glycoproteins becomes clear when regarding that an estimate of more than 50% of the eukaryotic proteins are glycosylated, which in turn demands considerable physiological effort and a whole range of enzymes to work together.

This estimate was derived from the SWISS-PROT database that, at that time, contained 75 000 protein sequences two thirds of which have the Asn-X-Ser/Thr motif and are thus potentially *N*-glycosylated. From a subset of 749 well-characterized glycoproteins an occupancy of 2/3 of these *N*-glycosylation sequons was determined. The majority of *N*-glycoproteins carry *O*-glycans, as well. The proportion of solely *O*-glycosylated proteins was determined at

Table 1 Potential and real glycosylation sites in the 749 well-characterized glycoproteins listed in the SWISS-PROT database by the end of 1998. (Reproduced with permission from [2])

	Glycoproteins with at least one biochemically characterized ("real") glycosylation site		Glycoproteins with at least one real *N*-glycosylation site and at least one real *O*-glycosylation site		Glycoproteins with at least one real *N*-glycosylation site and no real *O*-glycosylation site		Glycoproteins with at least one real *O*-glycosylation site and no real *N*-glycosylation site	
	Sites	Entries	Sites	Entries	Sites	Entries	Sites	Entries
Potential *N*-glycosylation sites (sequons)	2066	697	289	80	1679	582	98	35
Real glycosylation sites	1965	749	556	80	1041	582	368	87
Real *N*-glycosylation sites	1279	662	238	80	1041	582	0	0
Real *O*-glycosylation sites	686	167	318	80	0	0	368	87

10% (Table 1) [2]. Together with less frequently observed post-translational modifications by carbohydrates like *C*-mannosylation or ribosylation, glycoproteins are found almost anywhere in the cell and are involved in a highly diverse spectrum of biological functions, ranging from protein folding, molecular and cellular recognition, and regulation of biological half-life [3].

Glycoproteins are, among others, preferred targets of pathogens, e.g. parasites or viruses. Furthermore, aberrantly glycosylated proteins play key roles in many degenerative diseases like cancer and autoimmune disease. Consequently, these properties make glycoproteins highly attractive targets for the development of new therapies and diagnostic methods. From the rapid increase of protein drug prescriptions in recent years we may expect a major industrial effort in glycoprotein production as well as to obtain a greater understanding of their functions [4].

This work reviews selected examples of our current knowledge of 3D structure and structural dynamics of selected glycopeptides and glycoproteins and gives an overview of the methodologies to arrive at high resolution structures. This review is not intended to be comprehensive and the reader is referred to other excellent reviews for further information [5–12]. A prerequisite of a structure elucidation is the availability of pure material. It is beyond the scope of this review to cover all synthetic and biochemical approaches towards glycopeptides and glycoproteins, which are described in other parts of this volume. Here, we will focus on high resolution 3D structures and their implications obtained from NMR or X-ray methods.

Aspects of glycopeptide and glycoprotein structure determination have very often been split into structure analysis of the glycan part on the one hand and of the peptide/protein part on the other, respectively. Excellent reviews cover these subjects including elucidation of primary structure, conformational analysis and molecular modeling [5, 9–11, 13–17]. We will focus on methods and developments that have been applied to intact glycoproteins or glycopeptides.

Even today, we should be cautious when trying to generalize what we know from glycoprotein structures solved so far. Our view is biased because the majority of 3D structures could be solved only after extensive manipulation of oligosaccharides before X-ray crystallography, expression in non-glycosylating organisms for NMR spectroscopy or from synthesis of truncated glycopeptide structures. Many proteins that essentially require glycosylation for structural stability or proper function have probably not been analyzed yet including the majority of membrane glycoproteins.

There are only a few very-well studied glycoproteins for which we have in-depth information on conformation of both the protein and the glycan fraction, on molecular flexibility and dynamics, and on the influence of glycosylation on biological function. One example is the human T-cell surface protein CD2 that crucially requires its *N*-glycan for binding to its receptor CD58. Another example is the glycoprotein hormone human chorionic gonadotropin (hCG) which is biologically inactive when deglycosylated. Both cases will be presented in greater detail.

2
Preparation of Glycopeptides and Glycoproteins for Conformational Analysis

Conformational analysis of glycoproteins relies on the availability of sufficiently pure protein in sufficient amount. Depending on the type of analysis this ranges from microgram quantities of heterogeneous material, for example for CD experiments to multiple milligrams of very pure and homogeneous preparations for NMR studies or crystallography. Current methodology to obtain homogeneous glycopeptides and glycoproteins has been reviewed by Grogan et al. [18], and is covered in more detail in other parts of this volume.

2.1
Natural Sources

Apart from abundant glycoproteins like soy bean agglutinin, etc., usually only small quantities of glycoproteins can be extracted from natural sources, which can even make primary structure determination challenging. In addition, isotope labeling for NMR experiments is generally not easily possible at reasonable costs if proteins have to be extracted at basal expression levels.

An example for a glycoprotein that can be purified in sufficient amounts from natural sources is the complement regulatory protein CD59. It was extracted from human urine and its structure was determined already in 1994 by Neuhaus and coworkers by 2D NMR spectroscopy without ^{15}N or ^{13}C labeling [19]. Complete resonance assignments were obtained for the polypeptide consisting of 77 residues and an extensive set of nuclear Overhauser effects (NOEs) allowed the protein structure to be determined at high resolution. Signal overlap prevented an unambiguous NMR assignment of the carbohydrate and GPI anchor part and of their NOEs except for the GlcNAc residue which is directly connected to Asn18. This glycan residue is the only one that shows NOEs to the polypeptide chain. The structure calculation led to two families of conformers consistent with these protein carbohydrate NOEs exhibiting disorder in the χ_1 dihedral angle of Asn18 (for a definition of dihedral angles, see Fig. 1). In both conformers, the proximal GlcNAc residue has a similar orientation with respect to the protein covering hydrophobic side chains on the surface of the protein component. Sharp lines and the lack of protein carbohydrate NOEs for the remaining residues suggest that they are highly flexible and point into solution. Glycosylation of CD59 seems to be important for organization of this protein in the membrane and for prevention of aggregation (Fig. 2).

Natural sources of glycoproteins have proven very useful for the isolation of oligosaccharides when only the glycan part is to be studied as there are

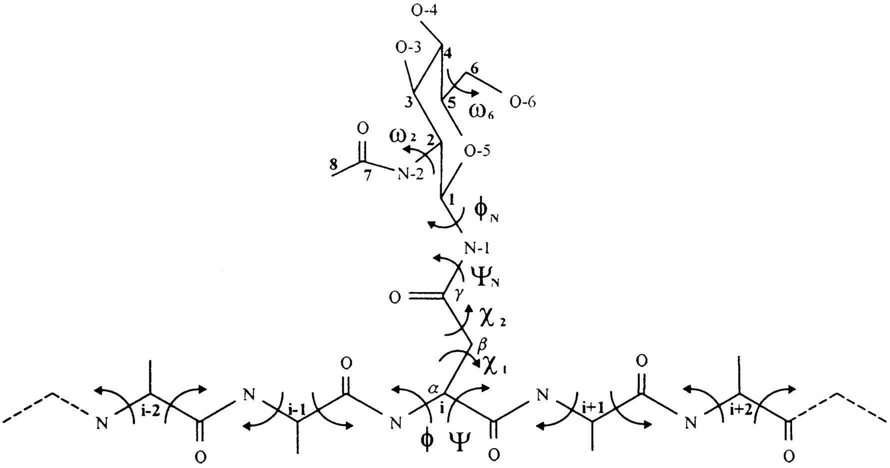


Fig. 1 Schematic representation of the *N*-linked GlcNAc together with the Asn-Xxx-Ser/Thr-containing peptide fragment. The GlcNAc and Asn non-hydrogen atoms, together with torsion angles of interest, are labeled. (Adapted and reproduced with permission from [20])

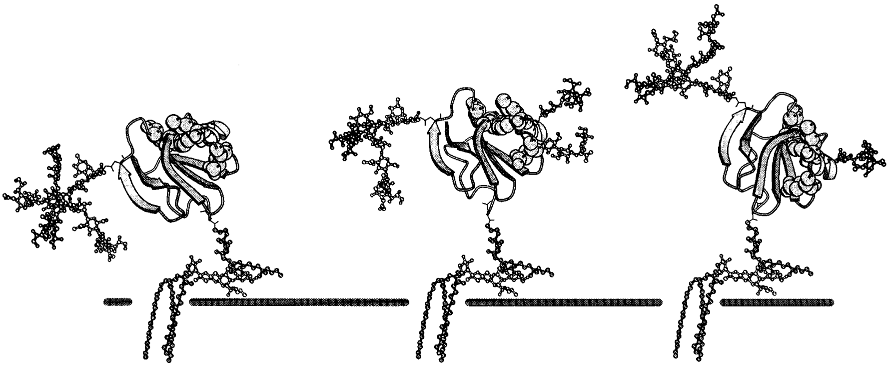


Fig. 2 A schematic figure showing the effect of CD59 glycosylation on the flexibility of the protein relative to the GPI anchor [19, 21]. Three different glycoforms are shown. Both the *O*- and *N*-linked oligosaccharides (size range 2–6 nm) attached to CD59 (diameter of approximately 3 nm) restrict the conformational space available to the protein and limit its interaction with the lipid bilayer. The sugars may therefore orient the active site of CD59 toward the C5b-9 complex, which is also inserted into the cell membrane. In addition, the heterogeneity of the sugars suggests that the glycans influence the geometry of the packing, and it is unlikely that CD59 molecules will form a regular array on the cell surface. The active site residues are highlighted in CPK. (Reproduced with permission from [21])

well-established sources for complex-type, high-mannose, and hybrid-type oligosaccharides, which can be released by trifluoromethanesulfonic acid, hydrazinolysis or enzymatic treatment and subsequent purification.

2.2
Expression Systems

Structural genomics initiatives have boosted the development of expression systems. After focusing initially on small bacterial and archeal genomes the field has subsequently moved to eukaryotic organisms which are more complex and more interesting due to the occurrence of a whole range of post-translational modifications. Separate reviews cover the latest developments in this field and we will only highlight select aspects relevant to studying glycoproteins [22–24].

One common approach to simplify the issue is based on bacterial expression systems to produce a non-glycosylated variant of the protein of interest. This approach however often fails, as glycans are known to be often indispensable for proper folding or to avoid protein aggregation. Expression of glycoprotein sequences in yeast can give very good yields of protein. However, if one is interested in intact, natively glycosylated proteins, yeast expression results in a yeast-specific high mannose glycosylation that is very different from the glycosylation pattern in mammalian cells. Expression in the yeast *P. pastoris* is more suitable than in *S. cerevisiae*, because *P. pastoris* produces glycoproteins with shorter glycans attached to the same Asn-X-Ser/Thr site as in mammalian cells, and it also does not produce structures with terminal α1,3-linked mannose residues which are not found in mammalian cells. Table 2 gives an overview of the glycosylation properties of a variety of expression systems.

Baculovirus-infected insect cells are well-studied expression systems to yield glycoproteins with a simple oligosaccharide moiety called paucimannose (Fig. 3) [24]. Recently, modified baculoviral expression systems yielded glycoprotein structures carrying humanized oligosaccharides. This was achieved by transfecting the insect cell lines with Glucosaminyltransferase II (GnTII), Galactosyltransferase (β1,4-GalT) and Sialyltransferase (α2,6-ST) to obtain biantennary complex-type structures terminated in 2,6-linked sialic acids. Chinese hamster ovary (CHO) cells have been engineered to express human-like glycoproteins by knocking out α1,3-GalT and introducing α2,6-ST.

As far as NMR experiments on isotopically labeled material are to be carried out, all of the above mentioned expression systems have to be adapted to minimal media. Homans and collaborators were the first to succeed in producing a ^{13}C-,^{15}N-labeled glycoprotein, namely the α-subunit of hCG, by growing CHO cells in a medium made from algal hydrolysates and supplemented with isotopically labeled amino acids (cf. Sect. 5.1.3) [26]. Even earlier Fesik et al. obtained isotopically labeled urokinase-type plasminogen acti-

Table 2 The major glycosylation properties for different cell expression systems. (Reproduced with permission from [24])

Organism	Cell type	Type of glycosylation [a]	Length of oligosaccharide	Composition of oligosaccharide	Glycosidic linkage
Bacterium	*E. coli*	–	–	–	–
Yeast	*S. cerevisiae*	*O*-linked	L	Man, Gal. GlcNAc, GalNAc	β1,3
Yeast	*S. cerevisiae*	*N*-linked	L	Man, GlcNAc	α1,6; α1,2; α1,3
	Pichia pastoris	*N*-linked	S	Man, GlcNAc	α1,6; α1,2
Plant	Tobacco BY2	*N*-linked	S	Fuc, GlcNAc	α1,3
Insect	Baculovirus-infected insect cells sf9	*N*-linked	S	Gal, Man, Fuc, GlcNAc	α1,6; α1,3
	Baculovirus-infected insect cells sf21	*N*-linked	S	Gal, Man, Fuc, GlcNAc	α1,6
	E. acrea Ea-4	*N*-linked	S	Man, Fuc, GlcNAc	α1,6
Frog	Xenopus oocyte	*N*-linked	S	Man, GlcNAc, GalNAc, sialic acid	α2,3
Hamster	CHO *O*-linked	*N*-linked,	S	Gal, Man, Fuc, GlcNAc/GalNAc/NeuAc	α1,6; α2,6
	BHK *O*-linked	*N*-linked,	S	Gal, Man, Fuc, GlcNAc/GalNAc/NeuAc	α1,6; α1,3; α2,3
Mouse	C127	*O*-linked	S	Man, Fuc, Gal, GalNAc, sialic acid, NeuGc	α1,6; α1,3; α2,3; α2,6
	Hybridoma	*O*-linked	S	Man, Fuc, Gal, sialic acid, NeuGc	α1,6; α1,3; α2,6
	J558L	*O*-linked	ND	Man, Fuc, Gal, sialic acid, NeuGc	α1,6; α1,3; α2,3; α2,6
	NSO myeloma	*O*-linked, *N*-linked	S	Man, Fuc, Gal, sialic acid, GlcNAc, NeuGc	α1,6; α1,3; α2,3; α2,6
	Transgenic	*N*-linked	S	Man, Fuc, Gal, GlcNAc	α1,6
Human	Namalwa	*N*-linked, *O*-linked	S	Man, Fuc, sialic acid, NeuGc, GlcNAc	α1,6; α2,3; α2,6
	Pituitary	*N*-linked, *O*-linked	S	Man, Fuc, GalNAc, sialic acid	α1,6; α2,3; α2,6
	Heterohybridoma	*N*-linked	S	Man, Fuc, sialic acid, GlcNAc, NeuGc	α1,6; α2,3; α2,6

[a] Although the characteristics of *O*-linked and *N*-linked glycosylations are not completely studied in some literature, it may possess both the profiles of glycosylation in the same expression system.

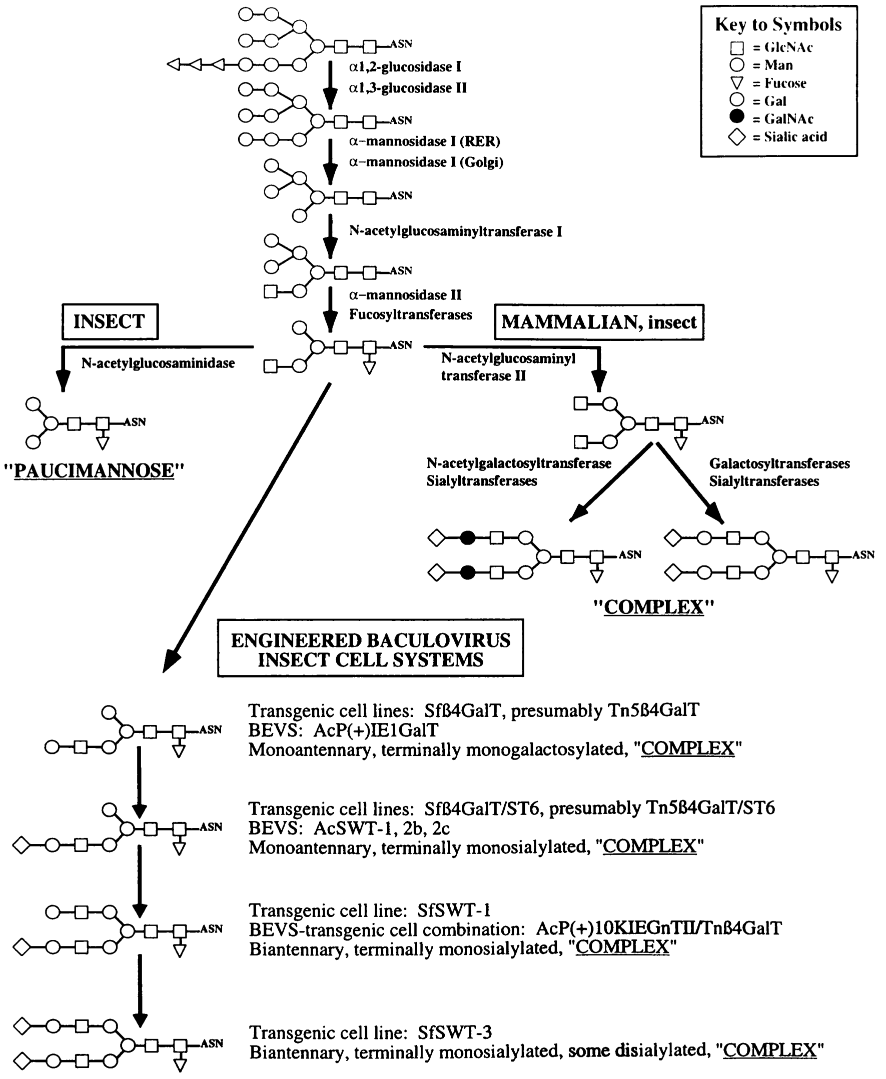


Fig. 3 Protein *N*-glycosylation pathways in insect and mammalian cells. Monosaccharides are indicated by their standard symbolic representations, as defined in the key. The insect and mammalian *N*-glycan processing pathways share a common intermediate, as shown. The major products derived from this intermediate are paucimannose and complex *N*-glycans in insect and mammalian cells, respectively. It is generally recognized that insect cells have only a limited capacity, at best, to produce complex *N*-glycans. However, this model accommodates the possibility that some insect cells can produce complex *N*-glycans under certain circumstances. Complex *N*-glycans are extremely diverse and only representative examples are shown in the figure. The structures of the *N*-glycans produced by transgenic lepidopteran insect cell lines, modified baculovirus expression vectors (BEVS), and BEVS-transgenic insect cell combinations are shown as well. (Reproduced with permission from [25])

vator from mammalian cell lines, although they purified an unglycosylated portion of this glycoprotein [27]. The work on the adhesion domain of human CD2 in the laboratory of Wagner started out with unlabeled protein preparations. In later stages of the work, refined structures, especially concerning the glycosylation site, were obtained from amino acid-specifically or ubiquitously labeled CD2 expressed in CHO cells (cf. Sect. 5.1.2) [28].

Eukaryotic expression systems experience an increased application for isotopic labeling purposes [22, 29, 30]. One recent example is the work of Kessler et al. who succeeded in producing functionally folded, glycosylated GM2 activator protein suitable for NMR purposes in *P. pastoris* following a modified protocol reported by Laroche et al. [31, 32]. Sometimes, extensive protein engineering is mandatory in order to improve the quality of NMR spectra [33].

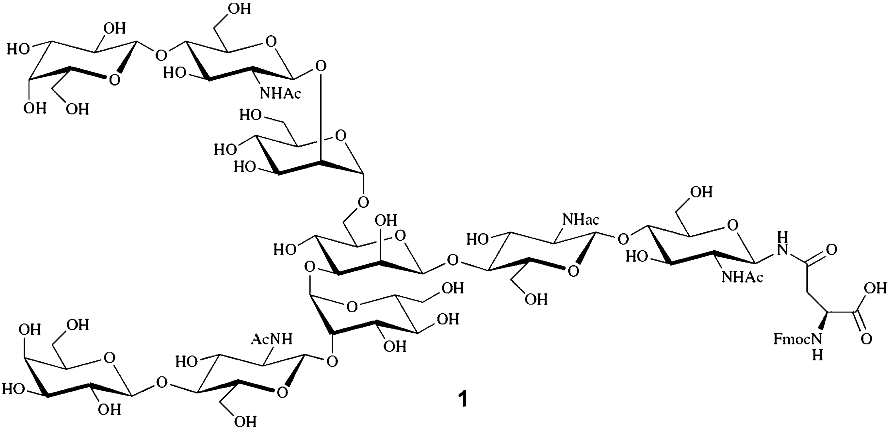
The glycosyltransferases and the glycosylhydrolases of the cell lines chosen for expression of the glycoproteins normally produce a non homogeneous pattern of glycosylation at each glycosylation site (microheterogeneity). It has been shown that introduction and/or overexpression of additional enzymes of the glycosylation machinery can lead to glycoproteins with a more homogeneous structure [34]. Alternatively, oligosaccharides may have to be trimmed down to a common core by glycosidase treatment prior to structure determination [18]. A way of obtaining homogeneously glycosylated proteins was found by Schultz, Wong and coworkers who exploited an enhanced genetic code to incorporate glycosyl amino acids during protein translation in *E. coli* [35, 36].

Recently introduced cell-free expression systems are now well established in structural genomics efforts due to advantages in cloning, purification and because expression in these systems can be automatized. In addition, proteins specifically labeled at certain amino acids can be synthesized in a straightforward manner by providing suitably labeled monomers. As there is no amino acid metabolism present no scrambling of isotopic labels occurs. It will be very difficult, though, to transfer the eukaryotic glycosylation machinery into such a system.

2.3
Chemical Synthesis

The chemical synthesis of glycopeptides and smaller glycoproteins can provide very pure molecules [8]. It is, however, very costly to produce isotopically labeled compounds. Also, there is a limitation in terms of size of these substances. Large proteins in their glycosylated form can still not be produced easily by chemical synthesis. Also, the introduction of arbitrary complex glycoside structures into these molecules is not easy. However, the glycosyl residue interacts mostly via the first and next few sugars with the protein. Therefore, truncated forms can be obtained by chemical synthesis as model compounds with maximum flexibility in choosing the site and type of glycosylation. Moreover, only by chemical means can we introduce a whole range of modifications that do not occur naturally which enables us to investigate the influence on glycopeptide conformation of specific structural elements. Both for *O*- and *N*-glycosylated structures synthetic protocols have been established, by which it is possible to generate singly and multiply glycosylated glycopeptides. Both types of glycosylated peptides are usually being synthesized by introducing a glycosyl amino acid building block into the growing peptide chain in a conventional peptide synthesis. In the case of the *N*-type glycosides, either carbohydrates derived from natural sources are being used to obtain larger oligosaccharide structures or the carbohydrate portion is being chemically assembled, as well. Subsequently, the oligosaccharides are aminated at the reducing end to produce a glycosylamine structure. These are linked to the carboxyl group of the side chain of aspartic acid in an amidation reaction to produce a glycosyl amino acid building block (Fig. 4). The carbohydrate structure varies in size between just the *N*-acetylglucosamine and a decasaccharide [37]. This glycosyl amino acid building block is then introduced into the growing peptide.

O-type glycosylation can be obtained in a similar way by chemical synthesis of the *O*-linked glycosyl linkage to serine or threonine where all types of *O*-linked core structures have been synthetically produced. Further modification of these structures can be obtained by chemo-enzymatic synthesis. These building blocks can then be introduced into the growing peptide chain to produce *O*-linked glycopeptides with one or several attachment sites of the carbohydrate to the peptide backbone. Chemo-enzymatic modifications can also occur after the complete synthesis of the glycopeptides [38]. For a detailed discussion of chemical and chemo-enzymatic synthesis of glycopeptides we refer to other overview articles that have recently been published, in part in this volume [8, 39–43].



1

Fig. 4 Structure of a biantennary *N*-type building block **1** ready for use in solid-phase peptide synthesis (SPPS)

2.4
Semisynthesis: Native Chemical Ligation and Expressed Protein Ligation

The synthesis of proteins has experienced a major advancement by the invention of two ligation methods, which allow for fusing one part of the target protein that has been obtained by expression techniques with another part that originates, for example, from solid-phase peptide synthesis. Both methods rely on transesterification of thioester groups followed by a sulfur-to-nitrogen rearrangement that finally yields native backbone polypeptides. In native chemical ligation (NCL) a chemically synthesized peptide with a C-terminal thioester group is coupled to a second peptide or protein which has an N-terminal cysteine residue [44]. The other method, expressed protein ligation (EPL), utilizes the peculiar behavior of protein domains called inteins which catalyze a protein splicing reaction leading to excision of the intein part and joining the polypeptides (exteins) that surround the intein at its N- and *C*-terminus. With this reaction one can introduce a C-terminal thioester functionality into an expressed protein which can then be linked via NCL to a peptide containing an N-terminal cysteine [45].

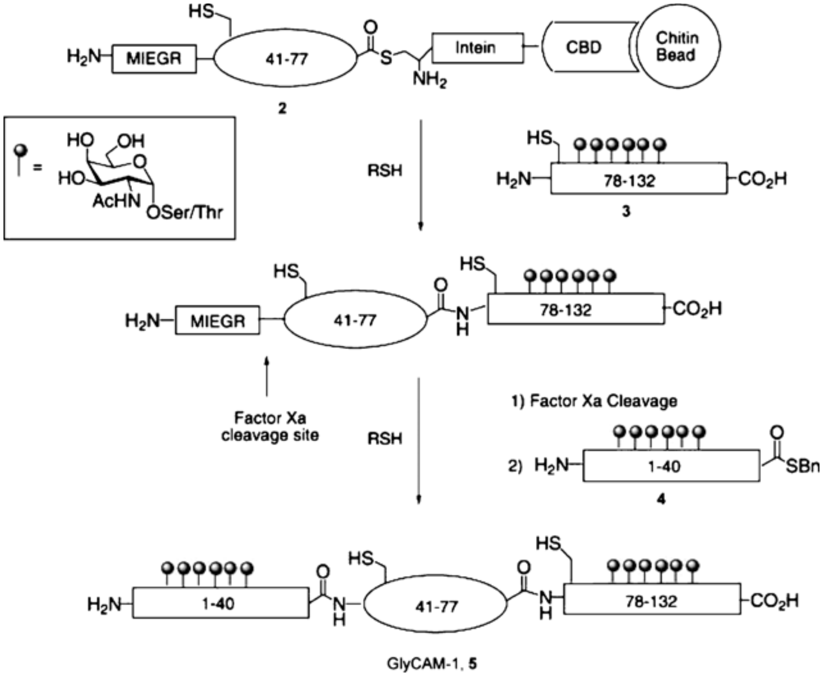


Fig. 5 Synthesis of glycosylated GlyCAM-1 (**5**) by expressed protein ligation (EPL) to synthetic mucin domains **4** and **5**. Factor Xa proteolysis allows late-stage exposure of an *N*-terminal cysteine for the timed NCL to **4** (CBD, chitin-binding domain) [46]. (Reproduced with permission from [18])

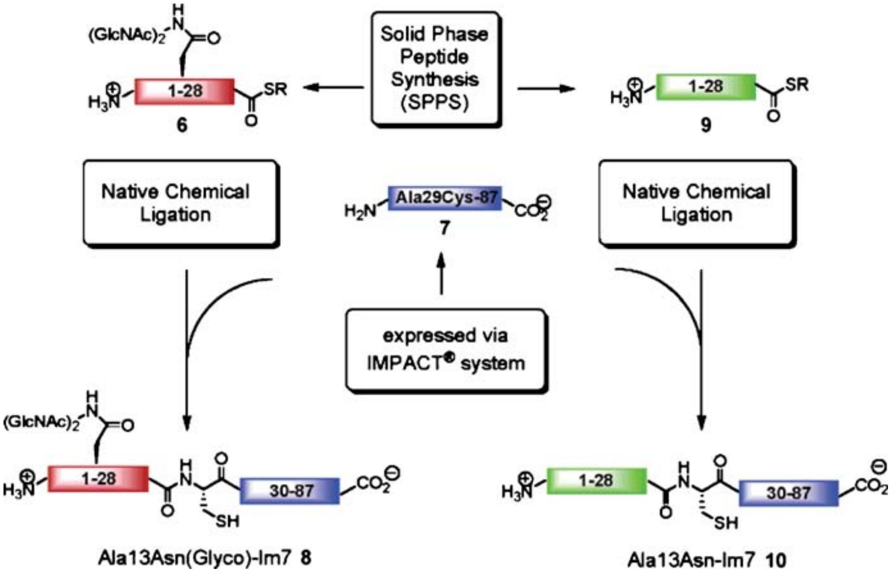


Fig. 6 Semisynthetic strategy for glycosylated and unglycosylated Im7 analogues **8** and **10**. (Reproduced with permission from [48])

These thioester ligation reactions are compatible with peptide or protein fragments carrying unprotected carbohydrates and are thus an attractive route to glycoproteins that are too large for SPPS or that cannot be expressed in sufficient amount or homogeneity. The potential of the EPL method for glycoprotein synthesis has been demonstrated by Bertozzi and coworkers who synthesized different glycoforms of the murine L-selectin ligand GlyCAM-1 (Fig. 5) [46]. Native chemical ligation of SPPS-derived peptides has been employed in Bertozzi's group to yield the *O*-glycosylated T-cell chemokine lymphotactin and by Hojo et al. for synthesizing the *N*-chitobiosylated Ig-1 domain of Emmprin, a cell surface protein of tumor cells [47].

Hackenberger et al. constructed, following the NCL strategy, *N*-glycosylated and unglycosylated variants of the four helix immunity protein Im7 in order to investigate the influence of the *N*-chitobiosyl residue upon folding of this well characterized model system (Fig. 6) [48].

3

Determination of Composition and Primary Structure

A prerequisite of 3D structure determination to high resolution is the precise knowledge of the primary structure and composition of the glycoprotein of interest. Apart from classical biochemical approaches like gene sequencing and Edman degradation that yield the primary structure of the polypeptide and potential glycosylation sites, special techniques have to be employed in order to accurately define (i) the presence of carbohydrate modifications at specific positions in the amino acid sequence, (ii) the type and the size of glycan, (iii) the carbohydrate branching pattern, i.e., the positions of glycosidic linkages, (iv) the anomeric configuration of glycosidic linkages, and (v) covalent modifications like sulfation or acetylation. In this regard, adept application of specific enzymatic cleavage reactions followed by chromatographic separation and identification of the purified carbohydrate substructures by lectins can yield the desired information [49].

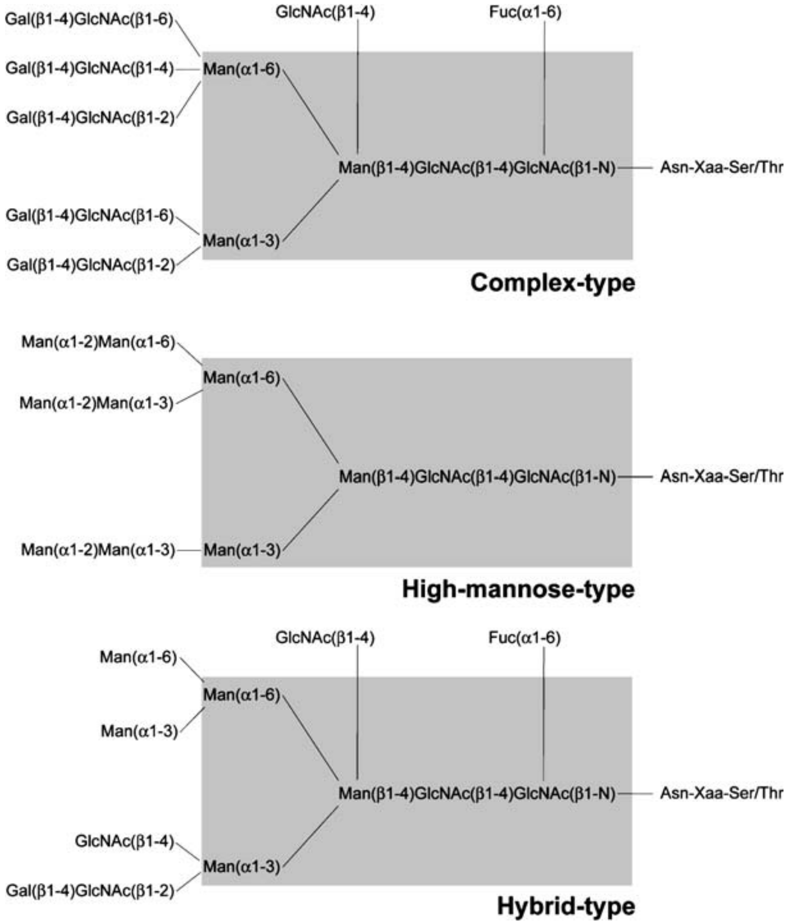


Fig. 7 Structures of the complex-, high-mannose- and hybrid-type sugar chains which constitute the three most common subgroups of *N*-linked glycans. The structure within the gray box represents the pentasaccharide core common to all *N*-glycans. All sugars except Fuc are D-sugars

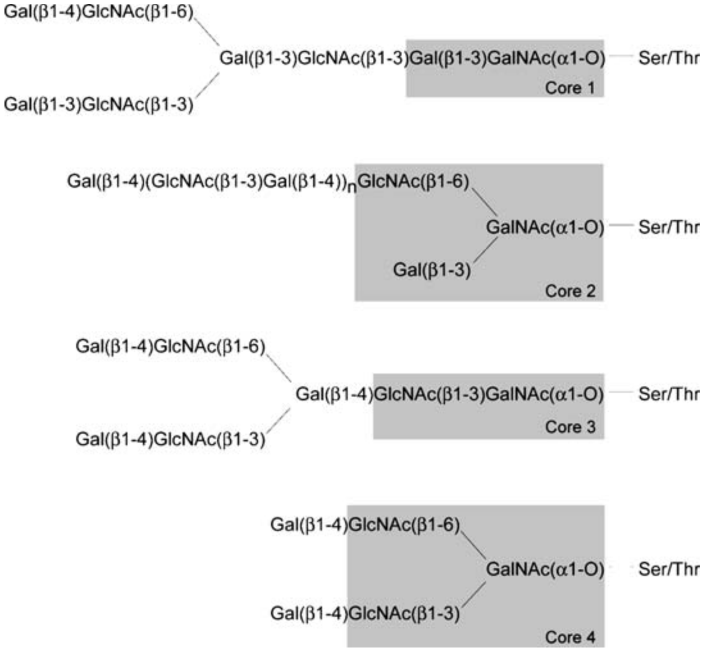


Fig. 8 Examples for *O*-glycans found in mucin-type glycoproteins. The oligosaccharides shown here are based on four O-glycan core structures which are indicated by *gray boxes*

In recent years, ultrahigh-sensitivity techniques like combinations of HPLC or capillary electrophoresis with tandem mass spectrometry have become increasingly important, especially for analyzing sub-picomol amounts of mixtures of glycoforms [50–52]. Because of the advent of high-sensitivity NMR probes with cryo-technology and extremely small sample volumes for HPLC-NMR as well as for MAS applications, NMR spectroscopy is of major importance not only for 3D structure determination but for elucidating glycoconjugate primary structures as well [9, 53, 54]. Examples of glycan structures found in *N*- and *O*-glycoproteins are given in Figs. 7 and 8.

4
Determination of 3D Structure

The structure determination of proteins by X-ray crystallography or NMR spectroscopy is well established. However, extension of this methodology to glycopeptides and glycoproteins bears major challenges due to several reasons [5, 10–12]. In addition to problems in obtaining pure homogeneous glycopeptides or glycoproteins from chemical or biochemical synthesis, these compounds are inherently difficult to crystallize. There are probably several reasons for this behavior. (a) No larger oligosaccharide structure has been crystallized even as pure isolated carbohydrate only, which is probably due to an unfavorable packing in the crystal due to structural irregularities; (b) the large hydration shell of the oligosaccharides make direct contacts of the molecules in a crystal less favorable; and (c) the flexibility and dynamics of the oligosaccharides attached to proteins often produce no defined 3D arrangements of the individual residues, which results in a crystal disorder effect from the carbohydrates. Therefore, even when crystals are obtained, very often part or all of the glycan is invisible in the electron density map due to local disorder. NMR on the other hand does not require single crystals. However, the population of multiple conformational states leads to time-averaged NMR parameters that cannot directly be converted into a 3D structure. Rather, NMR experiments have to be combined with extensive computational studies to predict possible conformations and to extract their relative populations.

The conformational analysis of glycopeptides and glycoproteins has very often been separated into the peptide and the glycan part ignoring the potential interdependence of both. Only recently have experimental tools become available that allow structure determinations of natively glycosylated proteins and there are a couple of new techniques that still await being employed for conformational analysis of glycoproteins. Structure determinations and analyses of molecular flexibility and dynamics of entire glycoproteins are still rare. Methodical advances in this regard shall be the focus of this review.

4.1
Conformational Analysis by NMR Spectroscopy

Being a technique for structure determination in solution capable of handling molecular disorder and dynamic equilibria, NMR spectroscopy should be the method of choice for the investigation of glycoproteins. Whereas for carbohydrates a wealth of knowledge has been achieved on conformational analysis, and excellent reviews cover this subject [5, 10–12], reports on glycopeptide and glycoprotein structure determination by NMR are still of limited number.

4.1.1
Chemical Shift Assignment

Conformational analysis by NMR requires chemical shift assignments for as many resonances of the glycoconjugate of interest as possible. In this regard, the polypeptide part poses fewer challenges than the carbohydrate moiety due to generally better dispersion of its resonances, much lower variation of the connectivity between monomers—there are only peptide bonds between

Table 3 Selected sugars and their characteristic NMR features. All the sugars listed are D unless otherwise noted. (Adapted and reproduced with permission from [9])

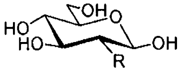
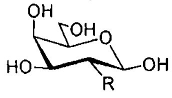
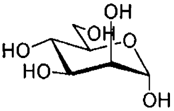
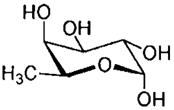
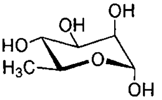
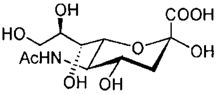
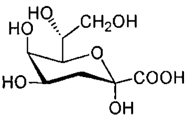
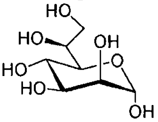
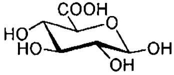
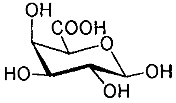
β-Glc and β-GlcNAc  R = OH, NHAc	$^3J_{HH}$ (H^1 – H^2, H^2 – H^3, H^3 – H^4, H^4 – H^5, [Hz]) αGlc: 3.6, 9.5, 9.5, 9.5 [60] βGlc: 7.8, 9.5, 9.5, 9.5 [60] The coupling pattern is similar for GlcNAc. GlcNAc, $\delta(CH_3) \sim 2$ ppm Upfield shift of δ(C2): αGlcNAc: δ(C2) $\sim$ 55.4 ppm βGlcNAc: δ(C2) $\sim$ 58 ppm [60, 61]
β-Gal and β-GalNAc  R = OH, NHAc	$^3J_{HH}$ (H^1 – H^2, H^2 – H^3, H^3 – H^4, H^4 – H^5, [Hz]) αGal: 3.8, 10, 3.8, 1 [60] βGal: 8, 10, 3.8, 1 [60] The coupling pattern is similar for GalNAc. GalNAc, $\delta(CH_3) \sim 2$ ppm Upfield shift of δ(C2): αGalNAc: δ(C2) $\sim$ 51.4 ppm βGalNAc: δ(C2) $\sim$ 54.9 ppm [60, 61]
α-Man 	$^3J_{HH}$ (H^1 – H^2, H^2 – H^3, H^3 – H^4, H^4 – H^5, [Hz]) αGal: 1.8, 3.6, 10.0, 9.8 [60] βGal: 1.5, 3.8, 10.0, 9.8 [60]
α-L-Fuc 	δ(C6) $\sim$ 16.3 ppm [61] δ(H6) $\sim$ 1.1 ppm [60] The coupling pattern is similar to Gal. $^3J_{H5H6}$ = 6.3 Hz [60]
α-L-Rha 	δ(C6) $\sim$ 18.0 ppm [61] δ(H6) $\sim$ 1.2 ppm [60] The coupling pattern is similar to Man. $^3J_{H5H6}$ = 6.2 Hz [60]
α-NeuAc 	δ(H3ax) $\sim$ 1.9 ppm, δ(H3eq) $\sim$ 2.3 ppm [62] Upfield shift of δ(C3) αNeuAc: $^3J_{C1H3ax} \sim 6$ Hz βNeuAc: $^3J_{C1H3ax} < 1$ Hz [63, 64]
α-Kdo[b] 	Upfield shift of δ(C3) $\sim$ 34.8 ppm [61] Coupling pattern similar to NeuAc acid for anomeric configuration
α-Heptose 	The coupling pattern is similar to Man. For identification assignment of C7 is required, where both a L and D form can be found.

Table 3 (continued)

β-Glucuronic acid  	The coupling pattern is similar to Glc. δ(C6) is pH dependent [61]: αGlcA: pH = 7.8: δ(C6) ~ 176.9 ppm pH = 1.8: δ(C6) ~ 172.9 ppm βGlcA: pH = 7.8: δ(C6) ~ 177.6 ppm pH = 1.8: δ(C6) ~ 173.8 ppm
β-Galacturonic acid  	The coupling pattern is similar to Gal δ(C6) is pH dependent [61]: αGalA: pH = 6: δ(C6) ~ 172.6 ppm βGalA: pH = 6: δ(C6) ~ 173.5 ppm

α-amino and α-carboxyl groups and most of them are *trans*—and, in most cases, due to a lower degree of flexibility. Classical strategies of peptide and protein NMR spectroscopy usually lead to sufficiently complete resonance assignment [15, 55, 56].

Chemical shift assignment of the glycan part turns out to be much more difficult due to severe overlap of resonances in the region between 3 and 4 ppm. Provided that glycoconjugates isolated from natural sources or expressed in cellular systems are being investigated, the glycan part raises additional questions in that the composition, sequence and connectivity have to be characterized prior to conformational analysis. Strategies in this area have been summarized in detail by the groups of Bock and Vliegenthart [9, 57]. Table 3 gives an overview on characteristic NMR parameters of monosaccharide building blocks found in glycoproteins. Databases that compile and provide statistical analysis of chemical shifts greatly facilitate the assignment task. The corresponding database to the BioMagResBank (BMRB [58]) of the protein field is Sugabase for the carbohydrate part. After termination of funding its content is now accessible, for example via the Glycosciences.de internet portal (cf. Sect. 4.5). Recently, a new method for metabolic incorporation of ^{13}C- and ^{2}H-isotopic labels into glycoproteins has been described that on the one hand facilitates heteronuclear NMR experiments and, in addition, provides clues for the assignment of specific building blocks from the ratio of H/D incorporation [59].

4.1.2
Nuclear Overhauser Enhancements

Despite the advent of NMR methods that extend the repertoire of structural restraints, like residual dipolar couplings (RDCs), cross-correlated relaxation

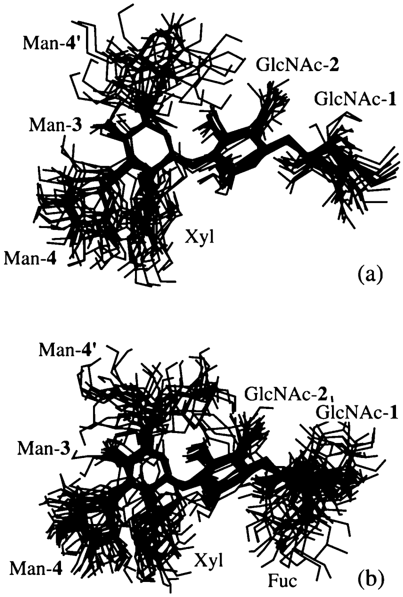
(CCR), and paramagnetic pseudocontact shifts, most 3D structure determinations published to date are still based on the analysis of nuclear Overhauser enhancements (NOEs) providing short-range distance information between pairs of nuclei, and on the measurement of scalar couplings that can be translated into dihedral angles by application of a Karplus relationship.

NOE-based structure determination is most successful in closely packed regions of a protein where many redundant distance restraints can be extracted. However, structures calculated from NOEs alone suffer from two major complications. Extended structures or protruding protein loops give rise to a relatively sparse number of NOEs that lead to low structural definition in this area. In addition, peptide chains with little contact to the protein core tend to adopt multiple conformations. The NOEs originating from these conformers are being time-averaged with the inverse sixth power of the internuclear distance as a weighting factor, i.e. short distances contribute much more strongly to the intensity of an NOE cross peak than long distances of the same pair of nuclei present in other conformers.

Both of these complications have a large impact on structure calculations of the oligosaccharide part of glycopeptides and glycoproteins. If the glycan protrudes into solution, there are usually very few long-range NOEs to the peptide part or between monosaccharide units. The conformation would have to be calculated mainly from NOEs bridging glycosidic linkages. This approach inherently suffers from the effect that small inaccuracies during distance determination accumulate to large structural uncertainties when looking from the core to the periphery of the glycan. The flexible nature of oligosaccharides gives rise to non-linear averaging of NOEs that complicate distance calculations even more.

Apart from evaluating other structural parameters with different averaging behavior or distance dependence, one way to address these problems is employing more sophisticated molecular modeling that involves back-calculation of NOE and ROE intensities from molecular dynamics (MD) trajectories as has been recently demonstrated by Lommerse et al. for two core *N*-glycans of snail α_D-hemocyanin (Fig. 9 and Table 4) [65].

A variety of NMR methods has been developed in order to provide an optimum of conformational information when oligosaccharides are involved. Spectral overlap is addressed by heteronuclear experiments that have recently become more feasible due to the development of cryogenic NMR probes with drastically improved sensitivity. Homonuclear 3D spectra have been shown to provide additional information not accessible by 2D spectra [66, 67]. The number of NOEs originating from the carbohydrate part of a glycoconjugate can be significantly increased, if the exchange rates of hydroxyl protons are being decreased by application of solvent mixtures and/or low temperature measurements [68].



11

12

Fig. 9 Conformations appearing during the MD simulations of two *N*-glycan core structures in α_D-hemocyanin of the snail Helix pomatia superimposed on the Man-**3**-(*β*1-4)GlcNAc-**2**-(*β*1-4)GlcNAc-**1** element. **a** Xylosylated oligosaccharide Man(*α*1-6)[Man(1*α*-3)][Xyl(*β*1-2)]Man(*β*1-4)GlcNAc(*β*1-4)GlcNAc(*β*1-OMe) (**11**) and **b** xylosylated (*α*1-6)-fucosylated oligosaccharide Man(*α*1-6)[Man(*α*1-3)][Xyl(*β*1-2)]Man(*β*1-4)GlcNAc(*β*1-4)[Fuc(*α*1-6)]GlcNAc(*β*1-OMe) (**12**). (Reproduced with permission from [65])

Table 4 Stability of local minima at the glycosidic linkages, probed by short MD simulations for **11** and **12** from Fig. 9. Criterion for a stable local minimum is a difference between starting and final conformation (after 50 ps simulation for **11** and 40 ps simulation for **12**) of less than 30° for both ϕ and ψ. Given are the starting ϕ,ψ values. s = stable, ns = not stable. (Reproduced with permission from [65])

Glycosidic linkage	Compound **11**			Compound **12**		
	ϕ	ψ		ϕ	ψ	
Xyl(*β*1-2)-Man-**3**	– 150	– 150	s	– 150	– 150	ns
	40	– 105	s	40	– 105	ns
	– 30	– 70	ns	– 80	– 140	ns
Man-**4**-(*α*1-3)Man-**3**	170	– 70	ns	170	– 70	s
	60	70	ns	60	170	s
	90	90	s	90	90	s
				140	180	ns
Man-**4'**-(*α*1–6)Man-**3**	120	90	ns	120	90	s
(ω = – 60°)	120	90	ns	120	90	s
	150	– 110	ns	150	– 110	ns
	140	165	s	140	– 165	ns

Table 4 (continued)

Glycosidic linkage	Compound **11**			Compound **12**		
	ϕ	ψ		ϕ	ψ	
	40	100	ns	160	100	ns
				100	150	ns
				40	180	ns
Man-**4′**-(α1–6)Man-**3** (ω = 60°)	140	100	s	140	100	ns
	160	– 140	ns	160	– 140	ns
	70	90	ns	100	140	s
	150	170	ns	180	120	ns
				60	40	ns
				180	40	ns
				80	– 100	s
Man-**3**-(β1–4)GlcNAc-**2**	– 90	– 70	s	– 120	70	s
	– 130	70	ns	– 130	– 50	s
	– 60	– 30	s			
	– 90	60	ns			
GlcNAc-**2**-(β1–4)GlcNAc-1	– 60	– 30	ns	– 60	– 30	s
	– 90	80	ns	– 90	80	s
	20	120	s	20	120	ns
				– 90	– 70	s
				80	60	s
Fuc(α1–6)GlcNAc-**1** (ω = – 60°)				– 100	60	s
				– 160	120	ns
				– 20	– 120	ns
				– 60	120	s
Fuc(α1–6)GlcNAc-**1** (ω = 60°)				– 100	80	s
				– 140	– 140	s
				– 160	140	ns
				– 160	40	ns
				– 50	– 70	s
				– 60	– 160	ns
				– 80	100	ns

4.1.3
Scalar Couplings

NOE-based structure calculations are generally supported by dihedral angle constraints that can be derived from ^3J scalar coupling constants. This is well established in protein structure determination. Whereas $^3J_{HH}$-couplings have been used already for a long time to determine the ring stereochemistry of monosaccharide building blocks and to derive the ω torsion angle of the hydroxymethyl group of pyranoses, gaining information on the ϕ and ψ angle of the glycosidic linkage requires more experimental effort as heteronuclear couplings have to be measured. This is traditionally done by HMBC-type experiments at natural abundance that exploit the sensitivity increase of inverse detection [69, 70]. ^{13}C-labeled compounds open the route to a variety of experiments that yield long-range H – C and even C – C couplings as proposed by Homans, Bush and Serianni [67, 71, 72]. Xu et al. demonstrate that $^{3}J_{CH}$-couplings can be reliably measured by quantitative J-correlation methods [73]. Deriving a Karplus relationship for transglycosidic angles is complicated by the high diversity of possible linkages and stereochemistries of the building blocks depending on the relative position of oxygens attached to the four atom fragment that defines the torsion angle [72]. Karplus parameters for interglycosidic J-couplings have been reported for H – C and C – C scalar couplings [67, 72, 74, 75].

In the case of molecules with flexible portions, like protein loops or oligosaccharides, structure determination based on scalar couplings has the major advantage of a much simpler averaging behavior compared to the NOE. The J-coupling values are simple linear averages over the ensemble of conformers. Martin-Pastor and Bush have shown that, for flexible oligosaccharides that populate distinct minima separated by substantial energy barriers, modeling based on scalar coupling data is superior to NOE-based modeling [71]. For a heptasaccharide of *Streptococcus mitis* it was possible to narrow the search space for allowed torsion angles for each individual glycosidic linkage. It turned out that the flexibility of this heptasaccharide is mainly localized to certain linkages and that only very few conformers are required to reproduce experimental coupling and NOE data. Although these methods have proved to be very powerful for determination of carbohydrate structure and dynamics, an application to intact glycopeptides or glycoproteins is still lacking.

4.1.4
Residual Dipolar Couplings

Whereas dipolar couplings are completely canceled out under conditions of isotropic tumbling, partial alignment of molecules with respect to the magnetic field of the NMR spectrometer gives rise to observable line splittings. The magnitude and sign of these residual dipolar couplings (RDCs) depend on the extent of molecular alignment and on the angle between the magnetic field and the vector connecting the coupled nuclei (Fig. 10) [76–78]. In the protein field, residual dipolar couplings are today widely used as additional restraints during structure calculation, for structure validation and refinement, and recently, as probes of dynamic processes [79, 80].

The major advantage of RDC-derived restraints compared to NOEs and scalar couplings is that RDCs provide long-range information from which the relative orientation of remote substructures can be concluded. The problem of small errors in short-range distances or torsion angles that accumulate to

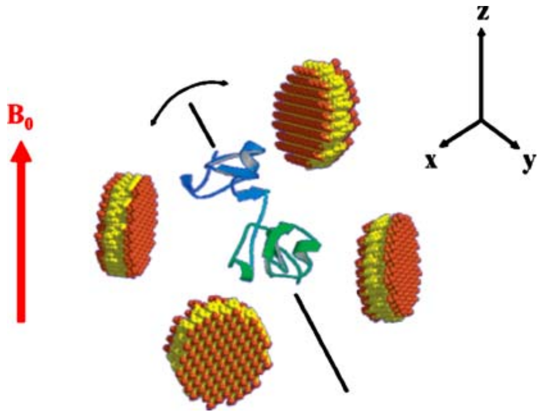


Fig. 10 Partial alignment of biomolecules: A two-domain construct from barley lectin oriented by collisions with bicelles. The protein's long axis tends to align parallel to the bicelle surfaces. (Adapted and reproduced with permission from [78])

large deviations in the entire structure is absent when looking at RDCs. This predestines RDCs as parameters for structure calculations of glycan chains of glycopeptides and glycoproteins that are otherwise poorly defined due to lack of long-range NOE contacts.

Applications to glycosylated peptides or proteins have not been reported, yet. The major reason for this might be the lack of reliable methods for introducing isotopic labels into glycoconjugates and complications due to motional averaging. The former limitation can be circumvented in part by using highest sensitivity NMR spectrometers equipped with cryogenic probes and/or operating at high field. In addition, significant advances have been made regarding expression systems for glycoproteins (cf. Sect. 2.2).

Motional averaging makes the analysis of RDC data more challenging compared to the situation encountered with more or less rigid, globular protein domains. On the other hand, RDCs are sensitive probes of dynamic processes and, complementary to relaxation rate and NOE techniques, cover a timescale that is not accessible by other NMR parameters (Fig. 11) [80–83].

Recently, characterization of conformation and dynamics of oligosaccharides by RDCs has been reported [84–86]. Azurmendi et al. could show from measurements of one-bond C – H dipolar couplings combined with Monte-Carlo simulations that two blood group epitopes, namely Fuc(α1-2)[GalNAc(α1-3)]Gal(α1-OH) and Fuc(α1-2)[Gal(α1-3)]Gal(α1-OH), adopt essentially one single conformation in solution in contrast to theoretical models that suggest the existence of two conformers [84].

Duus and coworkers analyzed the dynamic conformational equilibrium of the trisaccharide Man(α1-3)[Man(α1-6)]Man(α1-OMe) (**13**) which represents the core region of *N*-glycans. Hetero- and homonuclear RDCs served as restraints for long (50 ns) molecular dynamics simulations in explicit water.

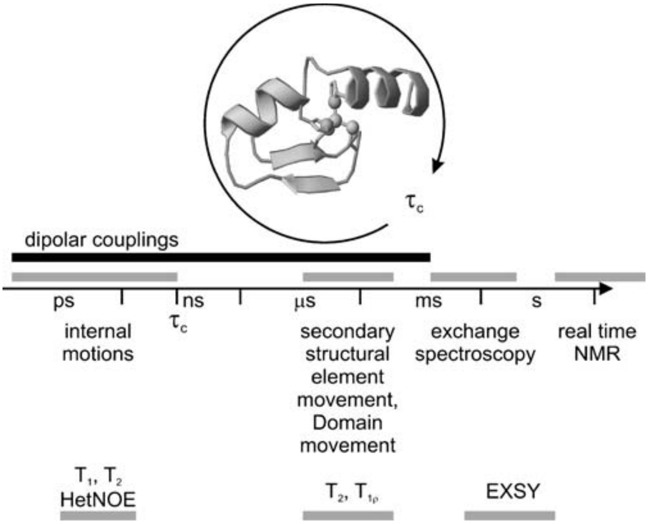


Fig. 11 Time scales that can be detected by various methods in NMR: The fast motions ps to ns are measurable by relaxation of, e.g., ^{15}N, the motion between approximately 50 µs and 10 ms by $T1\rho$ measurements, the slower time scales by exchange spectroscopy and even slower ones by real time NMR. Dipolar couplings cover all time scales from ps to ms and therefore can report also about the "dark" region of time scales. (Adapted and reproduced with permission from [80])

The results indicate almost equal populations of the *gg* and *gt* rotamers of the α1-6-linkage, whereas the *tg* rotamer was predicted to be unpopulated in aqueous solution (Fig. 12) [85].

13 **13**

Fig. 12 The two dynamic conformations of the trisaccharide Man(α1-3)[Man(α1-6)] Man(α1-OMe) (**13**). *On the left* is the *gg* conformation of the (α1-6) linkage and *on the right* the *gt*. In this diagram only the ring carbons, oxygens and the glycosidic oxygens have been drawn. They have been overlayed by best-fitting to the central sugar ring. (Reproduced with permission from [85])

4.1.5
Cross-Correlated Relaxation

Another NMR parameter that allows for measuring angles between bond vectors is the relaxation rate that originates from cross-correlation between relaxation mechanisms like dipole–dipole interaction or chemical shift anisotropy. This cross-correlated relaxation (CCR) rate provides direct information on the projection angle between bond vectors, provided that double or zero quantum coherence can be exited between the nuclei of interest as shown in Fig. 13 for two amide bond vectors of the protein rhodniin [87, 88]. In addition CCR rates contain information on local anisotropic dynamics [89, 90].

CCR methodology is still in rapid development for proteins and nucleic acids and has only recently been applied to the carbohydrate field for studying internal dynamics of monosaccharides [91], and to support primary structure elucidation of bacterial polysaccharides [92]. It remains to be seen whether CCR measurements will extend the repertoire for 3D structure determination of glycoconjugates, as well.

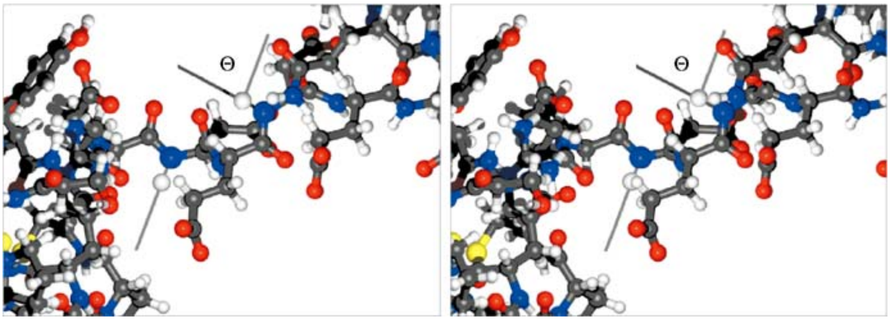


Fig. 13 Stereoview showing a close-up of the structure of the protein rhodniin. The *two gray lines* indicate the bond vectors of two spin pairs. The angle between the two interatomic vectors is indicated by θ. Atoms are depicted by the following colors: hydrogen, *white*; carbon, *black*; sulfur, *yellow*; oxygen, *red*; and nitrogen, *blue*. (Reproduced with permission from [87])

4.2
X-ray Crystallography

As of October 2005, the Protein Data Bank (PDB [1]) listed 1094 3D structures containing glycoprotein chains, 1073 of which were solved by X-ray crystallography. Many of the molecules studied had to have their glycans truncated or removed in order to be crystallized. This might bias us to see mainly glycoprotein structures with glycan residues that are, at least in the mature form of the protein, not of crucial importance for structure or function.

The disorder problem associated with flexible glycan chains of glycoproteins can at least in part be overcome when the oligosaccharide conformation and/or orientation is being stabilized by intra- or intermolecular contacts. This leads to another family of X-ray structures that display oligosaccharides at a comparatively high degree of resolution caused by lectin-like protein carbohydrate interactions that stabilize the glycan in the crystal. One should on the other hand be cautious when interpreting this kind of study as the carbohydrate binding auxiliary might force the glycoprotein or its oligosaccharides to be trapped in one favorable conformation or even adopt non-native conformations. The dynamic behavior of the glycoprotein is most certainly altered compared to the free molecule. Structure determination of highly flexible carbohydrates by X-ray crystallography is facing inherent problems such that dynamic features of these molecules are much better addressed by solution techniques.

Nevertheless, methodological advances have been reported that allow for determining medium-resolution structures from partly disordered glycoprotein crystals as exemplified by the work of Chen et al. who were able to obtain the structure of a fully glycosylated SIV gp120 envelope glycoprotein in an unliganded conformation by X-ray crystallography at 4.0 Å resolution [93, 94]. By optimizing beam geometry and the strategy for phase improvement including multi-crystal averaging, a 3D structure was obtained that showed at least four ordered sugars on many of the 13 oligosaccharides (Sect. 5.1.1).

4.3
Molecular Modeling

Both experimental methods for structure determination at atomic resolution, X-ray crystallography and NMR spectroscopy, face significant complications when applied to highly flexible molecules coexisting in numerous conformations as is mostly the case for the oligosaccharide portion of glycoproteins and is of importance for the peptide part of glycopeptides, as well.

This makes molecular modeling an invaluable tool for searching conformational space and for calculating structures which are consistent with experimental data. Apart from very small glycopeptides, most molecules of biological interest still exceed the capabilities of quantum mechanical approaches in terms of the number of atoms and degrees of freedom. Therefore, force field calculations dominate the field. A whole variety of potential energy functions has been developed and extended in order to accommodate the polypeptide as well as the glycan part, including AMBER [95–99], CHARMM [100–102], GROMOS [103, 104], OPLS [105, 106], and TRIPOS [107, 108].

All implementations have to face as a major challenge the high degree of hydrophilicity and thus extensive solvation of the carbohydrate part, electronic arrangements that largely vary with conformation, leading to the anomeric, *exo*-anomeric and gauche effects, and a correct representation of the forces that act between the glycan and the protein part, including electrostatic and van der Waals interactions as well as entropic effect due to solvation/desolvation processes. A comparison of the performance of 20 different force fields including their carbohydrate-specific parameters has been carried out by Pérez et al. [109]. Even though computational methods for structure calculation are highly sophisticated, today, many features of physical reality are still being neglected, ranging from induced polarization effects via protonation equilibria to a thorough treatment of solvation [10, 11].

Apart from considerations relating to the force field, an important aspect is the software implementation that has to be capable of handling all available experimental restraints in a flexible way. Regarding NMR spectroscopic data of glycopeptides and glycoproteins, special emphasis pertains to treating highly ambiguous distance restraints due to overlap of carbohydrate resonances, and to incorporating the growing number of restraints derived from tensorial interactions. In addition, the high degree of conformational flexibility of these molecules demands incorporating restraints in an ensemble and/or time-averaged form [110]. Flexible options for this have been implemented, for example into AMBER [95], X-PLOR/CNS [111, 112], and GROMOS [103].

4.4
Statistical Analysis of Available 3D Structures

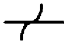
The entirety of available crystallographic data on oligosaccharides, glycoproteins, and glycan-binding proteins was analyzed by Petrescu et al. in 1999 [113]. From 639 glycosidic linkage structures distinct conformers could be identified. The O5-C1-O-C(x)′ torsion angles for all these distinct conformers appears to be determined chiefly by the *exo*-anomeric effect. The favored conformations identified in this study should prove useful when modeling "average" glycan structures and should, in analogy to the Ramachandran plot for proteins, also allow the easy identification of distorted glycosidic linkages.

More recently, Petrescu et al. focused on the peptide glycosidic linkage, and the peptide primary, secondary, and tertiary structures around *N*-glycosylation sites [114]. At that time, less than 3% of the proteins in the PDB were glycosylated. Regarding that over 50% of eukaryotic genes code for glycoproteins this again illustrates the immense technical difficulties associated with crystallization of glycoproteins [2]. Out of 2592 glycosylation sequons 1683 (65%) were occupied by *N*-glycans. An increased occurrence of aromatic residues and a deficit of acidic residues upstream from occupied glycosylation sites was found. Glycosylated sequons showed strong preference of threonine with respect to serine at position +2 whereas proline at position

+ 1 seems to completely prevent glycosylation. Other than that, no pattern of remote amino acid distribution was found.

Glycosylation can occur on all types of secondary structure, with a bias towards turns and bends. It is noteworthy that glycosylation sites occur at or very close to protein regions where there is a change in secondary structure. The authors hypothesize that this reflects a specific role of glycans in organizing the folding process. As might be expected, many of the occupied *N*-glycosylation sites are found at convex surfaces (33%). On the other hand, a surprisingly large number of glycosylated asparagine residues have a low accessibility with 10% being in deep recesses and 20% with the attached glycan filling a cleft both primarily formed by hydrophobic, especially aromatic, amino acid side chains (Table 5). These findings provide evidence for the im-

Table 5 Analysis of surface geometry and relative accessibility of the Asn residue for all occupied glycosylation sites in a non-redundant data set of 386 *N*-glycosylation sites. The surface geometry was qualitatively classified in two orthogonal directions by inspection and the following attributes were assigned for each direction: convex (*x*), concave (*v*), inflection (*i*), and flat (*f*). Very frequently, *N*-glycosylation was found at the edge of a groove. These sites were separately labeled as *e* (edge). The relative accessibility of the Asn residue to a probe of radius 3 Å was defined as the ratio between the side chain accessibility of the Asn residue and that of an Asn residue located in a fully extended Gly-Asn-Gly tripeptide. Asn SC (–G), Asn side chain relative accessibility in the absence of the glycan; Asn SC (+G), Asn side chain relative accessibility in the presence of the glycan; Asn Bkb (+G), Asn backbone relative accessibility in the presence of the glycan. (Adapted and reproduced with permission from [114])

Sites (%)	Asn SC (–G)	Asn SC (+G)	Asn Bkb (+G)	Surface type	Local geometry of the protein protein surface at the glycosylation site
33.2	57.6 ± 19.8	34.4 ± 27.0	28.2 ± 30.8	xx	
18.6	29.5 ± 13.9	13.9 ± 21.9	12.9 ± 23.2	e	
11.0	36.9 ± 11.6	12.6 ± 15.1	6.8 ± 16.3	xf	
6.8	42.4 ± 11.9	16.8 ± 19.7	6.7 ± 11.8	xi	
6.5	23.4 ± 10.0	6.3 ± 11.5	3.8 ± 10.6	fi	
6.0	7.6 ± 5.2	0.0 ± 0.0	0.2 ± 0.6	vv	
6.0	25.1 ± 6.4	3.6 ± 5.6	1.9 ± 5.7	ff	
4.5	21.5 ± 11.1	3.8 ± 7.2	7.6 ± 14.4	vx	
3.4	14.9 ± 9.1	1.4 ± 3.6	1.1 ± 3.9	vi	
3.1	32.5 ± 9.9	10.0 ± 9.5	6.0 ± 11.0	ii	
0.8	29.1 ± 2.7	10.4 ± 10.6	0.0 ± 0.1	fv	

portance of hydrophobic protein/glycan interactions. In addition, the authors propose that, by occupying grooves or holes, these glycosylation sites become a direct marker of proper folding as enzymes of the quality control machinery, like UDP-glucose glycoprotein:glucosyltransferase (UGGT), would not process these sites during quality control which would in turn prevent protein degradation.

N-Glycosylation is correlated with altered asparagine side chain torsion angle distribution and reduced flexibility [Figs. 14 and 15 (and Fig. 1 for angle definitions)]. Unmodified Asn residues show a preference for the g^- conformer compared to the *t* conformer, largely irrespective of whether Asn is located inside the protein, at a solvent-accessible surface or in a glycosylation

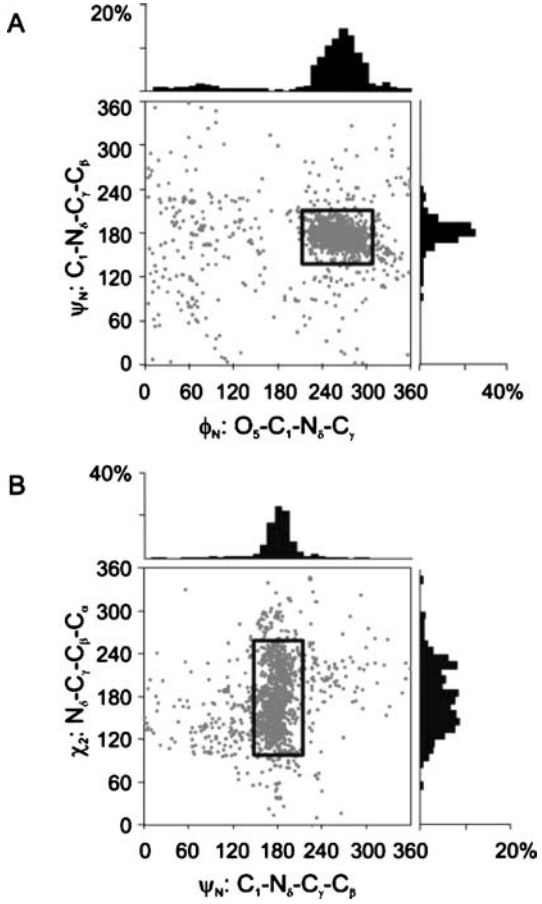


Fig. 14 Torsion angle and histogram plots for the Asn-GlcNAc linkage. Structures associated with a distinct conformer are shown by the *boxes*. **A** Plot of ϕN versus ψN. **B** Plot of ψN versus χ2. (Reproduced with permission from [114])

sequon. In contrast, glycosylated Asn residues populate the *t* conformer in preference to the g^- conformer which corresponds to an extended side chain conformation increasing the distance between the first GlcNAc and the peptide back bone. A further important outcome of the statistical analysis is that, in the glycosylated case, the $\chi 2$-angle displays a much narrower distribution compared to unmodified Asn residues while in both cases the $\chi 2$-dihedral centers at 180° in agreement with previous work [20]. The reduced standard deviation of $\chi 2$ in the glycopeptide case is in agreement with a whole body

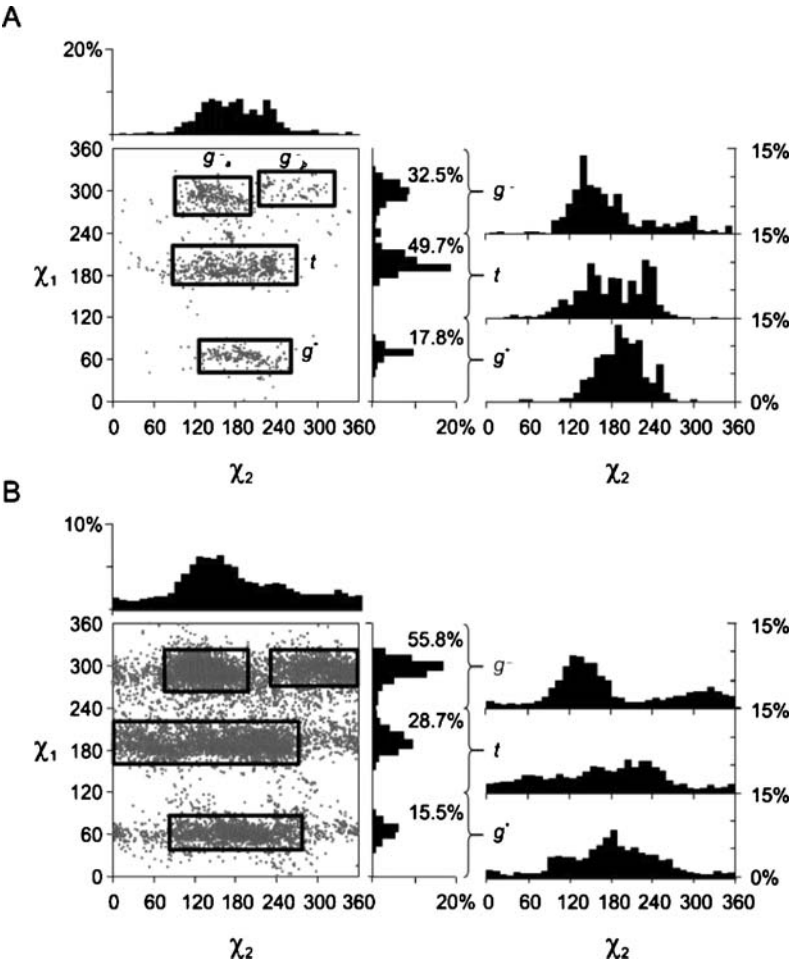


Fig. 15 Torsion angle and histogram plots of $\chi 1$ versus $\chi 2$ for Asn residue side chain subsets. The $\chi 2$ histogram plot is shown for each distinct $\chi 1$ conformer, as well as for the total population. **A** Asn residues in occupied glycosylation sequons. **B** Asn residues not in glycosylation sequons and with a solvent accessibility greater than zero. (Reproduced with permission from [114])

of literature saying that *N*-glycosylation leads to decreased flexibility of the peptide chain.

Ninety percent of *N*-glycosylation sites found in the PDB have the same conformation of the glycosidic linkage with a ψN-angle of 180° as expected for a *trans* amide bond and a ϕN-angle centered around 240°. This is in agreement with the previous statistical analysis of PDB structures by Imberty and Pérez and with structural information obtained by NMR spectroscopy [20, 115–117]. Interestingly, the broader data basis surveyed in the latest statistical analysis by Petrescu et al. allowed the identification of a second conformational family populated by roughly 12% of structures. Here, ϕN averages at 75° significantly deviating from the value of 40° predicted by force field calculations [20].

In a previous study, the peptide backbone conformation and the amino acid distribution at position + 1 of the glycosylation sequon was analyzed by Veluraja and coworkers on a set of 696 Asn – X – Ser/Thr motifs found in structures of the PDB [118]. Here, as well as by analyzing the amino acid sequences of the SWISS-PROT database it turned out that Gly is found with increased probability in the glycosylation sequon. In addition, the authors report elevated occurrence of Asn and Phe at this position although this could not be verified by the more recent study of Petrescu et al. [114].

It is noteworthy that the backbone angles of Gly residues found at glycosylation sites cluster around $\phi G = +60°$ to $+110°$ and $\phi G = -30°$ to $+30°$, a conformation that is disallowed for non-glycine residues (Fig. 16). This backbone conformation allows the formation of an H-bond between the side chain of Asn and the hydroxy amino acid that has been proposed earlier to be essential for the function of the oligosaccharyltransferase [119, 120].

The whole set of confirmed *N*-glycosylating sequences could be classified into nine backbone conformations of the Asn-X-Ser/Thr motif with angular deviations smaller than 30°. In the majority of these conformations the authors were able to demonstrate the possibility of direct or water-mediated hydrogen bonds [118]. The significant number of structures analyzed by Christlet et al. that do not possess this kind of H-bonding and the fact that glycosylation occurs during synthesis of the polypeptide chain that is presumed to be unfolded at this stage, suggests on the other hand that conformations found in mature proteins need not reflect circumstances present at the time of glycosyltransfer [20].

4.5
Databases

Databases that compile and order chemical, structural and biological information and make it searchable are largely responsible for the increasing pace in the life sciences and for major breakthroughs seen, for example in genetics and structural biology of proteins.

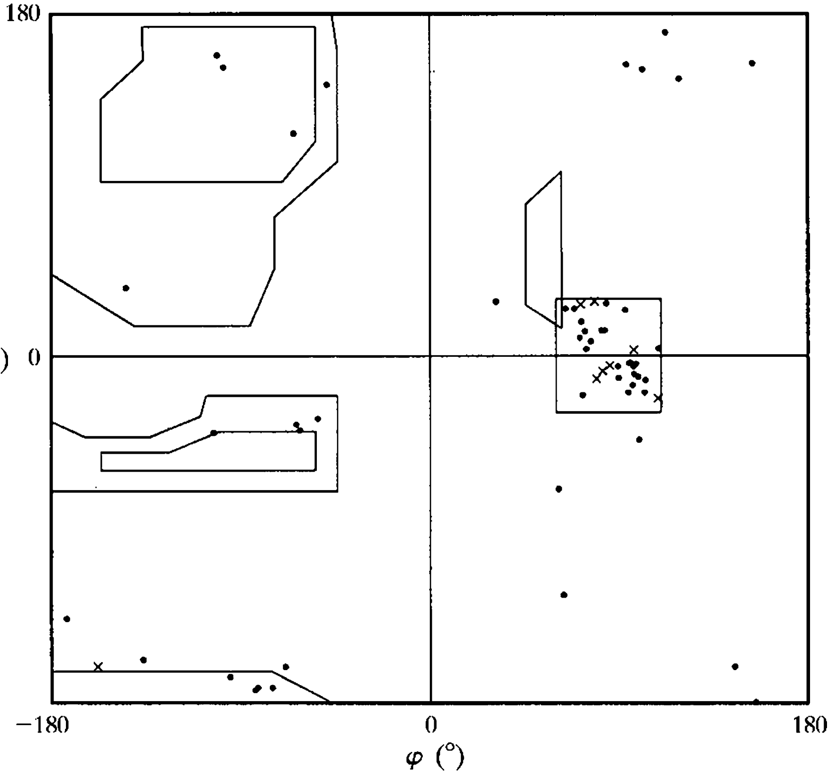


Fig. 16 Ramachandran plot showing the clustered region for Gly in the glycosylated consensus sequences Asn-Gly-Ser/Thr. The clustered region is marked as a *rectangular box*. *Dots and crosses* represent consensus sequences from a set of 488 non-homologous proteins and confirmed glycosylated sequences, respectively. (Reproduced with permission from [118])

Protein Data Bank (PDB) (http://www.rcsb.org/)

The most important source for 3D structural information on biomolecules is the PDB containing 36710 entries (May 2006) [1, 121]. Structures of proteins, glycoproteins as well as nucleic acids and biomolecular complexes determined by X-ray, NMR or theoretical methods are compiled together with information on sequence, experimental conditions, and methodical details. From 2006 on, the PDB has fully switched to a new web interface with improved search and browse functionality. However, it is still not straightforward to identify all PDB entries of glycoproteins that actually contain structural information on the carbohydrate part. In order to do so, Petrescu et al. applied their own search algorithm to a downloaded version of the PDB (cf. Sect. 4.4) [113, 114].

Biological Magnetic Resonance Data Bank (BMRB) (http://www.bmrb.wisc.edu/)
The pedant to the PDB in terms of NMR spectroscopic data is the BMRB maintained at the University of Wisconsin, Madison [58]. At the time of writing, it contained ca. 7000 entries on chemical shift assignment, and, in many cases, restraint lists and even NMR time domain data. However, chemical shift data on glycopeptides and glycoproteins are scarce. Most of the examples reviewed here are not, or only very incompletely, referenced in the BMRB.

4.5.1
Glycoconjugate-Specific Databases

The need for cross-linking available information on glycoconjugates has triggered the founding of several glyco-specific databases and internet resources. Currently, three major initiatives are, in part jointly, building-up comprehensive knowledge bases for glycobiology, namely Glycosciences.de, the Consortium for Functional Glycomics, and the Kyoto Encyclopedia of Genes and Genomes.

Glycosciences.de (http://www.glycosciences.de/)
A broad range of information and a versatile collection of tools is available at the Glycosciences.de web site located at the DKFZ, Heidelberg, Germany. The SWEET-DB [122] associates primary structures of carbohydrates with 3D structures from the PDB, NMR data from Sugabase, and literature references from NCBI Pubmed and Carbbank. In addition, theoretically calculated mass spectra and 3D-coordinates can be retrieved for a given carbohydrate (sub)structure. Web-based tools allow for searching the PDB for carbohydrate structures, predicting protein *N*-glycosylation, building of 3D models and calculating of mass spectra of arbitrary glycan structures. Glycosciences.de has joined with the Consortium for Functional Glycomics and is part of the EuroCarbDB initiative which aims at creating a comprehensive database of glycoconjugate-specific information (http://www.eurocarbdb.org/).

Consortium for Functional Glycomics (CFG) (http://www.functionalglycomics.org/)
The US pendant to Glycosciences.de is the CFG [123]. Its database integrates glycan structures from Carbbank, Glycominds Ltd., as well as structure synthesized or identified by members of the CFG. In addition, the CFG curates databases on glycosyl transferases, glycan binding proteins, and various microarray and profiling experiments.

Kyoto Encyclopedia of Genes and Genomes (KEGG) Glycan Database (http://www.genome.jp/kegg/glycan/)
The KEGG Glycan database, part of the Kyoto Encyclopedia of Genes and Genomes, currently contains 11 118 entries on glycans covering primary structure, pathways and related proteins [124].

Glyco3D (http://www.cermav.cnrs.fr/lectines/)
At the CERMAV division of the CNRS, 3D structural information on mono-, oligo-, and polysaccharides, lectins, glycosyltransferases, and glycosaminoglycan binding proteins has been organized in the form of a web site that provides links to the corresponding PDB entry.

O-Glycbase 6.0 (http://www.cbs.dtu.dk/databases/OGLYCBASE/)
This database, recently updated to version 6.0, lists collected information on primary structure, site and type of glycosylation for 242 *O*- and *C*-glycoproteins with experimentally confirmed glycosylation [125]. In addition, content has been linked to various other databases like Medline, PDB, and GlycosuiteDB. Associated with this O-Glycbase is the netOglyc server that predicts *O*-glycosylation sites in proteins based on an artificial neural network [126] (http://www.cbs.dtu.dk/services/NetOGlyc/).

GlycoSuite (http://www.glycosuite.com)
Release 8.0 of the commercial GlycosuiteDB lists 9436 curated entries which are searchable for information related to structure, disease, mass spectrometry, etc. [127].

Sugabase (http://www.boc.chem.uu.nl/sugabase/sugabase.html)
SUGABASE is a carbohydrate-NMR database that combines CarbBank Complex Carbohydrate Structure Data (CCSD) with proton and carbon chemical shift values. Funding has been discontinued and this database is no longer being updated. Its information content has been integrated into other databases, for example SWEET-DB.

Complex Carbohydrate Structure Data (CCSD) and CarbBank (http://www.boc.chem.uu.nl/sugabase/carbbank.html)
The CCSD, formerly hosted at the Complex Carbohydrate Research Center, Georgia, USA, collected primary structures of carbohydrates and associated literature references. Funding has been discontinued [128]. The content of the CCSD has been integrated into other databases, for example SWEET-DB, CFG, and KEGG Glycan.

5
3D Structures of Glycopeptides and Glycoproteins and Implications for Biological Function

In the following we summarize recent results on glycoprotein structure that illustrate the diversity of glycan function and that highlight how fruitful structural research on glycoproteins and glycopeptides has become. We have compiled two sections, one on *N*-glycosylated and one on *O*-glycosylated peptides and proteins. Instead of presenting every structural work of the last five years, we will concentrate on a few examples that an extensive set of experimental data has been collected on, and we have to apologize to all researchers whose work we decided not to present here.

5.1
Effect of *N*-Glycosylation on 3D Structure

5.1.1
HIV Envelope Glycoprotein GP120

One of the great success stories of crystallography of glycoproteins centers around structure and interaction of the HIV envelope glycoprotein GP120. GP120 mediates adhesion of the viral particle to the host cell and, together with the transmembrane glycoprotein GP41, triggers viral entry. GP120, with a molecular mass of 120 kDa half of which originates from carbohydrates, associates with GP41 to form a heterodimer. Three of these heterodimers assemble to the so-called viral spike which protrudes from the viral envelope, a membrane that originally stems from the virus-producing host cell. GP120 is the only viral protein accessible to the immune system. Tragically, the virus has developed a whole variety of mechanisms to escape host defense. First, GP120 is heavily *N*-glycosylated. Glycans, synthesized by the glycosylation machinery of the host cell are attached at 27 sequence positions. These oligosaccharides shield a significant portion of the virus from recognition by immune cells and antibodies. Second, GP120 contains five hypervariable peptide loops which are immunogenic, however, mutations in these areas quickly make antibodies useless, obviously without impairing viral infectivity. Third, viral entry is essentially organized as a two-step process, starting with a primary interaction of GP120 with the immune globulin superfamily (IgSF) protein CD4 of the host cell, which leads to a conformational change of GP120 and results in presentation of the co-receptor binding site. The subsequent interaction with a co-receptor belonging to the family of chemokine receptors finally triggers a reorganization of the viral spike, insertion of a fusion peptide of GP41 into the host cell membrane and membrane fusion. This two-step entry mechanism protects vital areas of GP120 inside the protein until the primary interaction with CD4 has taken place.

The high content in carbohydrates and the presence of several flexible peptide loops constituted a significant hindrance for obtaining crystals. A major breakthrough was achieved by the groups of Hendrickson and Sodroski in 1998, when they succeeded in crystallizing a GP120-core in complex with two domains of CD4 and the F_{ab} fragment of an anti-GP120 antibody [129, 130]. This GP120-core construct had all of its carbohydrates cleaved off beyond the innermost GlcNAc residue and contained short Gly-Ala-Gly substitutions at the sequence positions of the V1/V2 and V3 hypervariable loops. Despite these radical modifications, GP120-core retains its affinity for CD4 and, upon interaction with CD4, presents an epitope which is recognized by the antibody 17b, the epitope of which overlaps with the chemokine receptor binding site.

This crystallographic study has put our knowledge about antigenicity, receptor requirements and sequence variability- and glycosylation-induced immune escape on a structural basis. GP120 has an inner and outer domain both of which are linked by a region called the bridging sheet, providing a basis for receptor-induced conformational changes. Contacts with the CD4 receptor are mainly made by main chain groups explaining why many mutations do not affect CD4 binding. The proposed co-receptor binding site is composed of conserved residues, yet, it is only exposed close to the host cell membrane after contact with CD4. A highly immunogenic surface, called a non-neutralizing face, is only accessible when monomeric GP120 is shed into solution, and consequently does not elicit antibodies with the capacity to bind to the fully assembled viral spike. Another surface, called the silent face, is not recognized by antibodies because of complete masking by complex carbohydrates.

On the basis of the described structure Kwong et al. followed-up by modeling the arrangement of GP120 in the oligomeric state of the viral spike [131]. Following criteria like carbohydrate exposure, occlusion of conserved residues and steric considerations with regard to CD4 and antibody binding, a model was developed that allows binding of multiple CD4 entities, contacting the oligomer from the periphery whereas the co-receptor binding site is exposed in the vicinity to the host cell membrane close to the threefold symmetry axis of the viral spike (Fig. 17). It appears that co-receptor binding is much more dependent on the charge of the third variable (V3) domain than on its precise structure providing yet another explanation for immune escape.

Recently, Huang et al. succeeded in crystallizing the GP120-core including the V3-loop which constitutes an immunodominant epitope and is of central importance for co-receptor binding and co-receptor usage, i.e. for using either CXCR4 or CCR5 as a secondary receptor [132]. With this structure, the whole co-receptor binding site of GP120 could be defined. From this, a consistent model of co-receptor binding could be proposed characterized by interactions of the positively charged base region of V3 with the

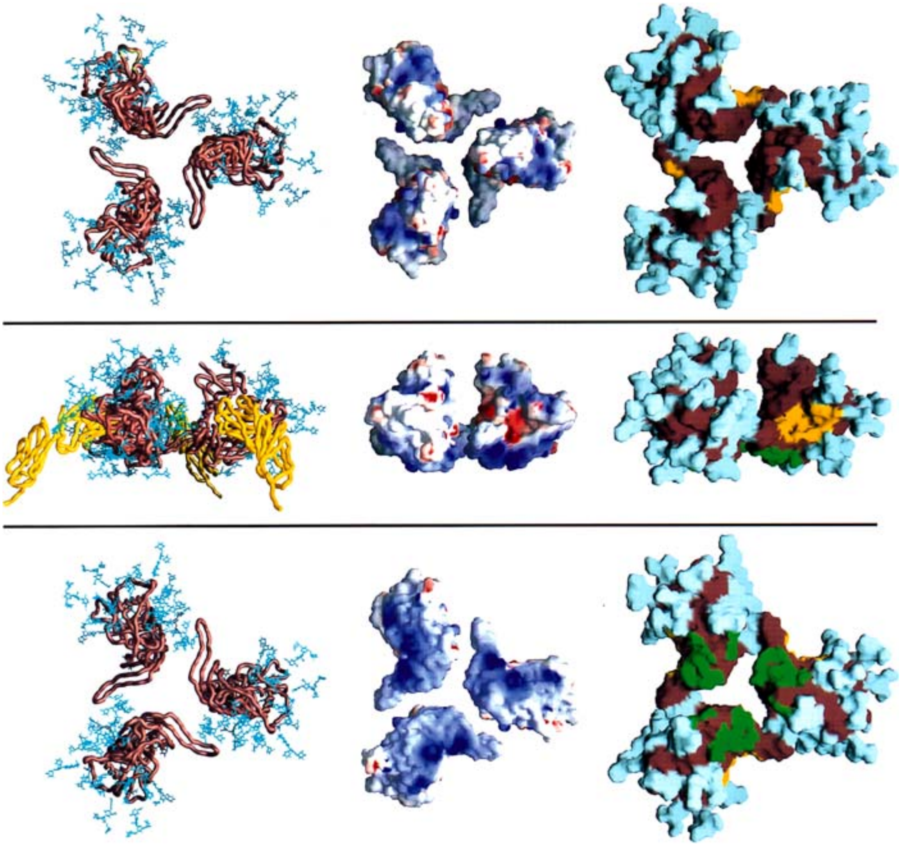


Fig. 17 Trimeric model of gp120. Three orientations of the model are shown. The images at the top depict the view from the orientation of the viral membrane. The *middle images* depict the view from the side, in between the viral and target cell membranes. The *images at the bottom* depict the view from the target cell membrane. The *left-most three images* are Cα worm representations of core gp120 (*copper brown*) and the two membrane-distal domains of CD4 (*yellow*). Also shown are the gp120 carbohydrate cores (*blue*), the $(Man)_3$-$(GlcNAc)_2$ cores shared by both high-mannose and complex *N*-linked glycan moieties. The carbohydrate shown here represents approximately half the carbohydrate on gp120, with the rest extending further from the gp120 surface. The *middle images* show the electrostatic surface of gp120 for the core. The electrostatic potential is depicted at the solvent-accessible surface, which is colored according to the local electrostatic potential, ranging from *dark blue* (most positive) to *red* (negative). The *right-most images* show the gp120 core with carbohydrate, with the solvent-accessible surface colored *cyan* for carbohydrate, *yellow* for the surface of gp120 less than 3 Å from CD4, *green* for the surface of gp120 less than 3 Å from the 17b antibody, and *copper brown* for the remaining surface of core gp120. The degree to which carbohydrate covers all of the solvent-accessible trimer surface is remarkable. Other than a small region at the viral proximal portion of the oligomer (where the missing N and C termini most likely reside), the only carbohydrate-free surfaces large enough to serve as an antibody epitope correspond to regions of receptor binding. (Reproduced with permission from [131])

sulfated N-terminus of the co-receptor and by binding of the conserved tip of V3 to the second extracellular loop of the seven-helix transmembrane co-receptor. It remains to be seen whether this model of receptor interaction will be confirmed by a high-resolution structure of the whole assembly present just before viral entry.

Another "snapshot" of HIV entry has recently been contributed by Chen et al. who determined the crystal structure of an unliganded SIV GP120-core, i.e. in the state prior to CD4 contact, something, that has been tried for more than 20 years [93, 94]. Interestingly, in this case, the typical deglycosylation procedure prior to crystallization attempts prevented rather than facilitated crystal formation. Thus, the structure of the fully glycosylated GP120-core, expressed in insect cells, was determined. The oligosaccharide part shows an astonishing degree of order resulting in electron density to be found for all 13 oligosaccharides, identification of seven α1,6-linked fucose residues, and eight glycans with order at least out to the mannose branch (Fig. 18). These well-ordered sugars are most likely a consequence of sugar-protein contacts as well as tight hydrogen bonding within sugar clusters. The structure of the unliganded GP120-core illustrates once more how oligosaccharides shield the protein surface from recognition by the immune system. In addition, this pre-fusion state shows significantly different organization compared to the CD4-bound structure, providing detailed insight into conformational changes that accompany receptor binding. Structural elements necessary for co-receptor binding are spatially separated in the unliganded form and thus not detectable by the immune system. On the other hand, the dramatic structural changes upon CD4 contact require GP120 to have several cavities and pockets, and Chen et al. have identified one of those as the binding site for a known entry inhibitor (BMS-378806 [133]).

In conclusion, crystallographic research on HIV GP120 has provided invaluable insight into viral entry, immune escape and, not to the least, experimental aspects of crystallography of glycoproteins. We can hope that these results trigger the development of new vaccines and pharmaceuticals that will be urgently needed in the fight against HIV and AIDS.

5.1.2
Human T-Cell Specific Surface Glycoprotein CD2

One of the relatively rare examples of glycoproteins that we have in-depth information on regarding structure, dynamics, interaction and biological function is the adhesion domain of the T-cell specific surface glycoprotein CD2 which is a key component of immune stimulation. It mediates adherence of T-cells to antigen-presenting cells by binding to the counter receptor protein CD58. Both proteins require *N*-glycosylation for biological activity although it turns out that the mechanism of stabilization is completely different in each case.

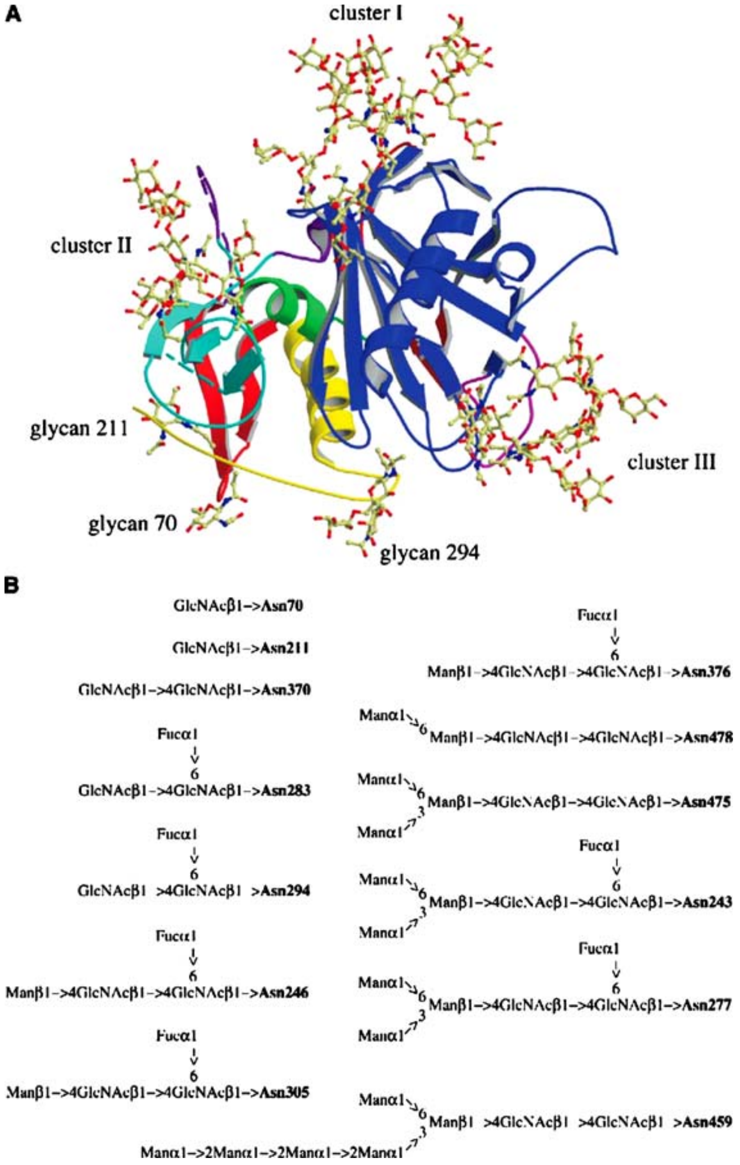
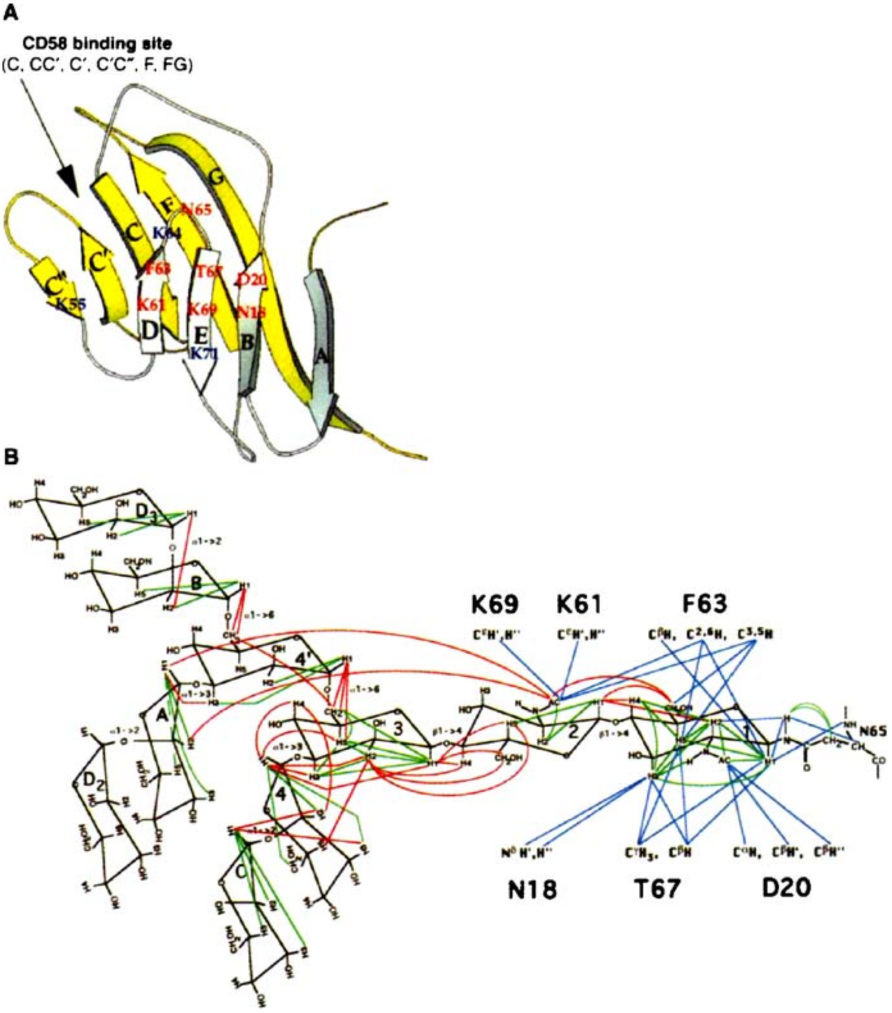


Fig. 18 Crystal structure of unliganded SIV GP120-core. The density for all 13 *N*-linked oligosaccharides was found in the final electron density map. **A** A view of SIV gp120 core structure in a ribbon diagram showing that the long *N*-linked glycans (stick model) form apparent sugar clusters on the surface of the molecule. Cluster I includes glycan 277, 305, and 459; Cluster II, glycan 243 and 246; Cluster III, glycan 376, 475, and 478. **B** The structures are shown with ordered residues as well as linkages; Asn in *bold* shows the glycosylated asparagine. Glycan 459 was built as a high-mannose type, but the identity of the last three ordered residues of the longer branch, which make contacts with a symmetry-related molecule, is questionable at current resolution. (Reproduced with permission from [93])

Already in 1993, a solution structure of the glycosylated N-terminal domain of CD2, expressed in CHO cells was reported, based on NMR spectra acquired at natural isotopic abundance, and from a protein sample specifically labeled with ^{15}N-lysine. Because of severe overlap of the sugar resonances the precise conformation and orientation of the single $(\mathrm{Man})_{5-8}(\mathrm{GlcNAc})_2$ glycan which is attached to Asn65 located at the tip of the DE loop could not be determined [134]. This was achieved two years later after successful expression of uniformly ^{15}N-labeled glycoprotein and with the aid of a high-field (750 MHz) NMR spectrometer [28]. The N-terminal domain of CD2 adopts the characteristic overall fold of a V-set IgSF domain although lacking half of strand A. The *N*-glycan is oriented toward the *β*-sheet containing



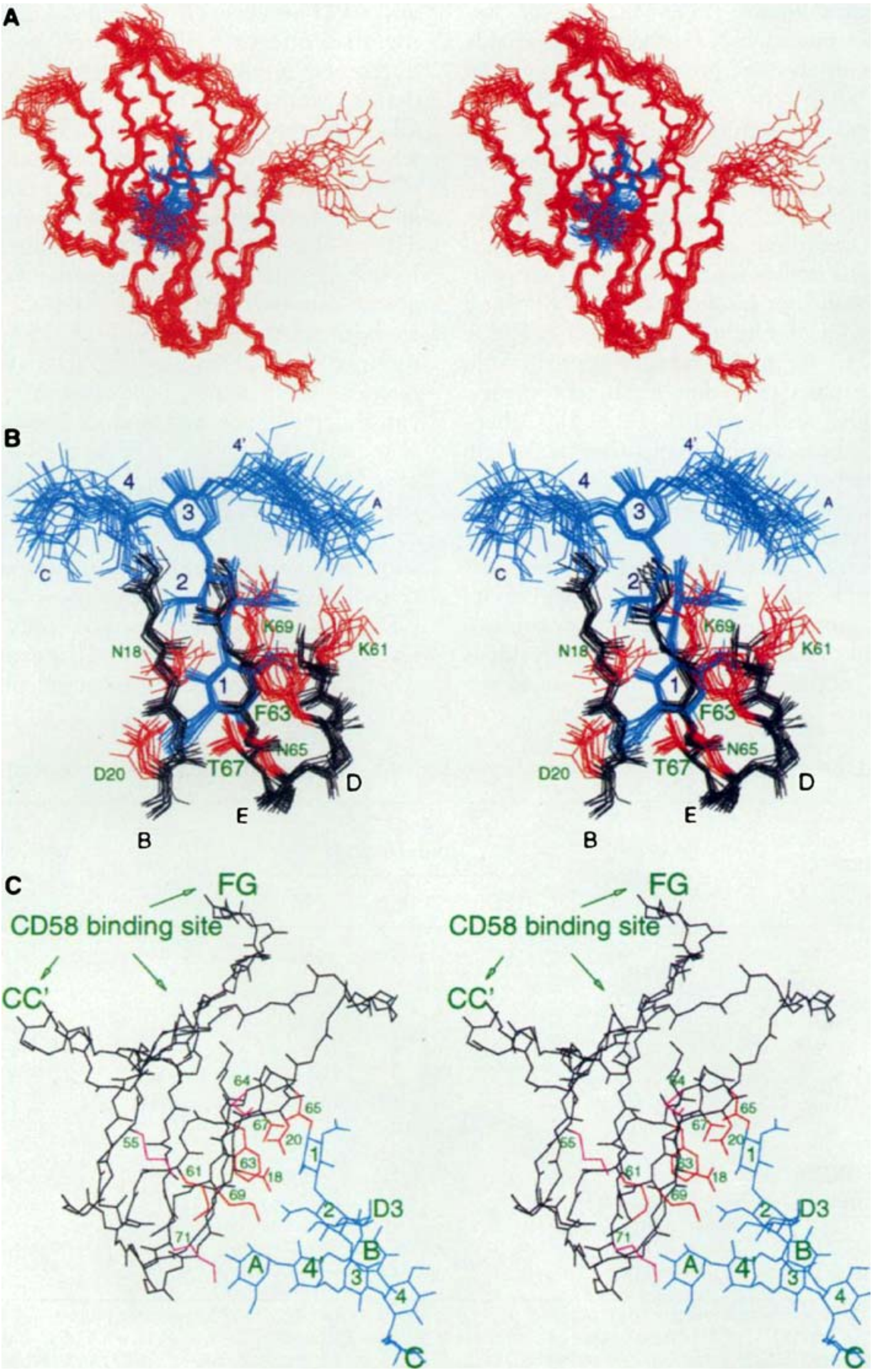
the strands D, E, and B and thus pointing away from the CD58 binding site (Figs. 19 and 20). There is no evidence for a direct interaction of CD2's glycan with CD58 or with residues of CD2 that have contact to the counter receptor. Rather, the carbohydrate stabilizes the conformation of the whole protein domain and prevents unfolding and aggregation. It is striking that already a single GlcNAc residue that remains attached to Asn65 after treatment with endoglycosidase H accounts for the major part of this effect [28].

The decisive difference between human CD2 and its homologs from other species that do not require glycosylation for activity is the presence of a lysine residue at position 61 whereas all other species have glutamic acid at this position. Lys61 leads to an unfavorable clustering of positive charges that obviously requires counterbalancing by hydrogen bonds, van der Waals interaction and entropic contribution of the *N*-glycan [28].

This and a follow-up study also provided a very detailed picture of the dynamics of CD2 and its high-mannose type oligosaccharide [135]. ^{13}C line width illustrate a significant reduction in the mobility of sugar residues close to the Asn65-GlcNAc linkage. In terms of flexibility GlcNAc-1 is as rigid as amino acid residues of the protein core. Sugar residues at the end of the *N*-glycan exhibit a mobility similar to that measured for amino acid side chain carbons and α-carbons of terminal residues. The assessment of mobility is also reflected in the structural ensemble showing a well-ordered trisaccharide core (Man)-(GlcNAc)$_2$ contacting the DEB-β-sheet.

^{15}N-relaxation data confirm this picture showing the glycosylated face of CD2 to be as rigid as the protein core. The termini and several loops show, as expected, fast motions on the ps-ns time scale, but surprisingly, in

◀ **Fig. 19** **A** Ribbon diagram of a single representative solution structure of human CD2 (residues 1–105), which is composed of two β sheets containing three (strands D, E, and B) and five (strands C″, C′, C, F, and G) antiparallel β strands. An additional β strand, A, makes parallel NOE contacts to strand G. The overall fold is characteristic of a V-set lgSF domain, but lacks the first half of strand A. A single high-mannose *N*-glycan is attached to the protein at Asn65 at the tip of the DE loop and is oriented toward the β sheet containing the strands D, E, and B. Polypeptide residues that contact the *N*-glycan are *highlighted in red*. The CD58 binding site, which includes residues in the C, C′, and F strands as well as in the CC″, C′C″, and FG′ loops, is located on the opposite face of the glycoprotein. Lys55, Lys64, and Lys71, are *highlighted in purple*, and together with Lys69 form a clustering of positively charged residues around Lys61. The view is with the DEB face in front and the C″C′CFG face in the back. **B** Summary of the NOE data obtained for the high-mannose *N*-glycan of CD2, NOEs are indicated only if they were unambiguously assigned. More NOEs to carbohydrate protons are expected, but they were either not resolved in the NOESY spectra or could not be assigned because many of the H^2, H^3, H^4, H^5, and H^6 sugar resonances exhibit similar chemical shifts. Carbohydrate intraresidue and inter-residue NOEs are shown by *green and red lines*, respectively, and carbohydrate-protein contacts are indicated by *blue lines*. Carbohydrate residues are labeled as in [139], and amino acid residues are abbreviated with the single-letter code and residue numbers. (Adapted and reproduced with permission from [28])


A
B
4'
4
3
A
C
2
K69
N18
K61
1
F63
N65
D20
T67
D
E
B
C
FG
CD58 binding site
CC'
64
65
67
20
55
61
63
18
69
1
2
D3
B
71
A
4'
3
4
C

◀ **Fig. 20** **A** Stereo diagram of backbone atoms of the polypeptide (*red*) and heavy atoms of the trisaccharide core Man-**3**-(β1-4)GlcNAc-**2**-(β1-4)GlcNAc-**1**-(β1-N-Asn) (*blue*) for 16 NMR structures of human CD2 superimposed on the NMR mean coordinates. The orientation is the same as in Fig. 19A. **B** DEB sheet and part of the high-mannose *N*-glycan of 16 model structures based on the superposition of the backbone heavy atoms of all residues (16 to 20, 60 to 71) in the DEB sheet and heavy atoms of the trisaccharide core Man-**3**-(β1-4)GlcNAc-**2**-(β1-4)GlcNAc-**1**-(β1-N-Asn). The backbone, the side chains for the contact residues, and the high-mannose *N*-glycan are shown in *black*, *red*, and *blue*, respectively. Glycoprotein structures were calculated for the $(\mathrm{Man})_7$-$(\mathrm{GlcNAc})_2$ glycoform, but for clarity the more disordered arm Man-**B**-Man-**D3** is not shown. The sugar residues are annotated as in Fig. 19B. **C** Representative glycosylated model structure of CD2 showing the polypeptide backbone in *black*, the side chains of the polypeptide residues that contact the glycan in *red*, the side chains of the polypeptide residues that together with Lys69 form a clustering of positively charged residues around Lys61 in *purple*, and the high-mannose *N*-glycan $(\mathrm{Man})_7$-$(\mathrm{GlcNAc})_2$ in *blue*. This carbohydrate is located on the opposite site of the CD58 binding site, and a conformation in which its [**4**′-A] arm is oriented toward the trisaccharide core Man-**3**-(β1-4)GlcNAc-**2**-(β1-4)GlcNAc-**1**-(β1-N-Asn) is significantly populated. (Reproduced with permission from [28])

the CD58 binding region these motions are largely absent and slow ms-μs fluctuations are found [135]. Later on, these relaxation data could be interpreted in terms of a model of concerted motion resembling conformational changes during counter receptor binding, albeit not taking slow motions into account [136].

As is the case for CD2, its receptor on antigen presenting cells, CD58, crucially requires glycosylation for bioactivity. Mature CD58 consists of 40–70% carbohydrate whereas the pure polypeptide part has a molecular weight of 11 KDa only. For a long time structural studies were unsuccessful as no way of preparing CD58 suitable for crystallization or NMR could be found. In contrast to the situation with CD2, the glycans of CD58 seem to stabilize the protein structure not by interacting with charged surfaces but by shielding hydrophobic patches which, in the absence of carbohydrate, lead to aggregation and insolubility. A homology model of CD58, based on different 3D structures of CD2 allowed for identifying hydrophobic residues of CD58 that are not involved in CD2 binding but are located on different faces of the receptor, and are, in its native form, presumably shielded by carbohydrates. These surface-exposed residues were mutated into polar amino acids finally yielding a glycan free, yet fully functional adhesion domain suitable for NMR purposes [137]. These mutational and structural studies on each binding partner finally opened the route for engineering glycan-free protein domains that could be crystallized in complex with each other [138]. Overall, the shielding of clustered positive charges as found for the adhesion domain of CD2 seems to be a special case. Statistical analysis of available 3D structures indicates that carbohydrates predominantly cover hydrophobic regions [114].

5.1.3
Human Chorionic Gonadotropin

Human chorionic gonadotropin (hCG) is a glycoprotein hormone involved in the maintenance of the *corpus luteum* during early pregnancy. It consists of two subunits, α and β, with molecular weights of 17 and 22 kDa, respectively, carrying in total four *N*-glycans (Glycan structures are shown in Fig. 21). It is remarkable that hCG is still a subject of intense structural studies regarding that X-ray structures of the $\alpha\beta$-heterodimeric protein were solved more than 10 years ago [140, 141]. This is mainly due to the observation that hCG preparations used for crystallization bind to the receptor but do not trigger signaling which is probably due to the presence of truncated *N*-glycans, and possibly further alterations, as a result of treatment with HF.

While a solution structure of a functional heterodimer has not been reported, yet, a wealth of knowledge has been gained for the 92 residue α-subunit in its native and various deglycosylated forms. αhCG carries *N*-glycans at two sites, Asn52 and Asn78, both of which are important for biological activity. The glycans at these sites display very different behavior in terms of flexibility and carbohydrate protein contacts, as became clear by work done in the groups of Vliegenthart, Kamerling and Homans.

Homans and coworkers pioneered NMR studies with ^{13}C- and ^{15}N-isotopically labeled αhCG produced in CHO cells [26, 142, 143]. One key feature of this expression method is that hCG carrying complex oligosaccharides is obtained. In order to further approach the glycosylation state found in human hCG, sialyl moieties were enzymatically attached to the CHO-cell hCG. The study focused on conformational analysis of the oligosaccharides in the free α subunit and in complex with the βhCG, and revealed that neither the glycan at Asn52 nor the one attached to Asn78 is in contact with the β subunit of hCG. The authors report that both glycans when attached to the protein behave very similar to the glycans studied free in solution. For the oligosaccharide at Asn52, no NOE connectivities to the protein were detected. MD simulations further support that this oligosaccharide is highly flexible and not significantly constrained by the protein.

On the basis of chemical shift assignments achieved by deBeer et al. at natural abundance the solution structure of the isolated α subunit was recently determined by Erbel et al. [112, 144]. In this case hCG was purified from urine and its glycans were trimmed down to the GlcNAc residue directly attached to asparagine by treatment with endo-β-*N*-acetylglucosaminidase-B. Free αhCG is a glycoprotein with an unusual fold containing two twisted β-hairpins and a cysteine knot motif displaying structural disorder in a loop region from residue 33 to 57 and at the *N*- and C-termini. Differences with respect to the crystal structure regarding the arrangement of the β-hairpins and the flexible loop region are caused by contacts with the β-subunit in the heterodimeric protein.

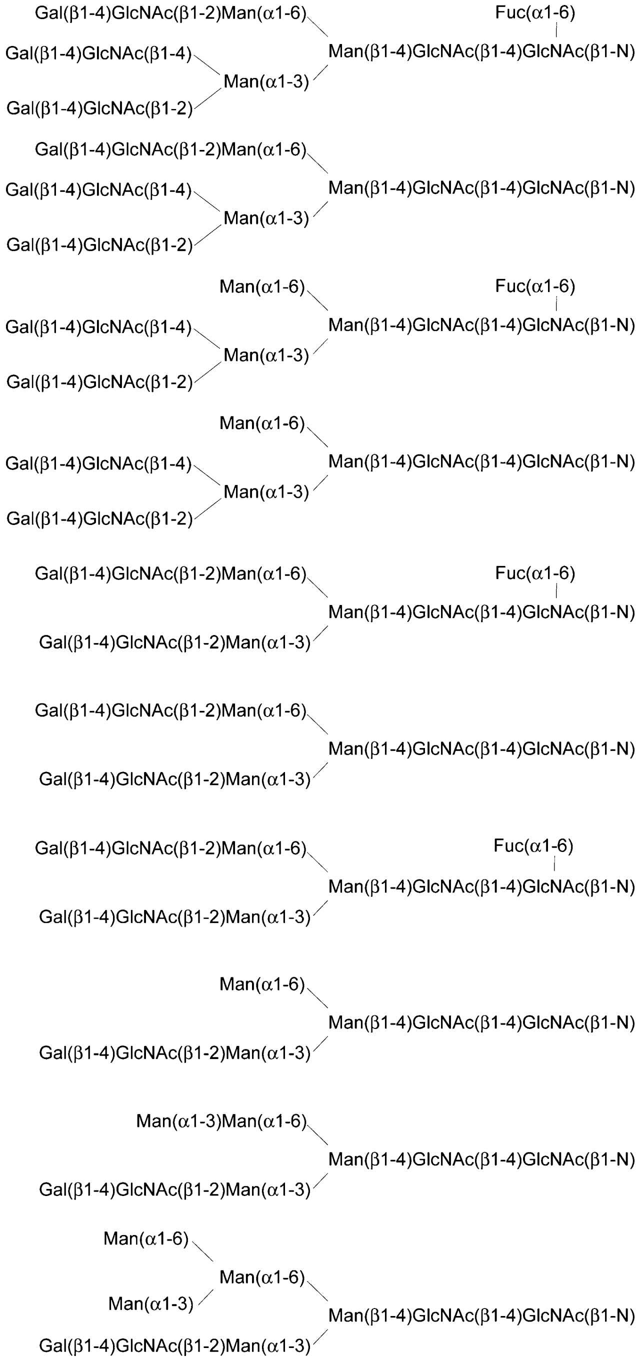


Fig. 21 Structures of desialylated *N*-linked sugar chains isolated from various hCG preparations (Adapted and reproduced with permission from [148])

Numerous NOE connectivities between GlcNAc-1 at Asn78 and hydrophobic amino acid side chains indicate that this residue is tightly packed onto the surface of the protein shielding this hydrophobic face from the solvent (Figs. 22 to 24). From that, the crucial role of this glycosylation site for protein stability becomes evident. In contrast, the glycan at Asn52 is located in a disordered region of free αhCG. These findings disprove the suggestion that, based on a single set of glycan signals measured by NMR at ^{13}C- and ^{15}N-labeled αhCG sample, both glycans should have a similar dynamic behavior, like a free glycan [143].

It has previously been reported that removal of carbohydrate residues other than GlcNAc-1 does not exert a major effect on the conformation of the protein part [144]. However, van Zuylen et al. demonstrated that glycosylation beyond GlcNAc-1 at Asn78 has a small but significant effect on the thermal stability of αhCG [145]. A further investigation of αhCG *N*-glycosylated at Asn78, only, revealed additional NOE connectivities between GlcNAc-2 and the protein core. From the values of ^{13}C-relaxation parameters T_2 and $T_{1\rho}$ of the constituting monosaccharide residues, it was concluded that the inner three residues have a severely restricted mobility (Fig. 22). The Man-4 and Man-4′ residues of the biantennary oligosaccharide exhibit a similar relaxation behavior, suggesting that the Man-4′ branch occurs in a single conformation of the C5–C6 linkage of Man-3 instead of in rapidly interconverting conformations that are known to exist for this linkage for the free oligosaccharide [146].

A structure calculation of this αhCG variant, carried out by Erbel et al., revealed restricted conformational freedom of the glycan at Asn78 as well as

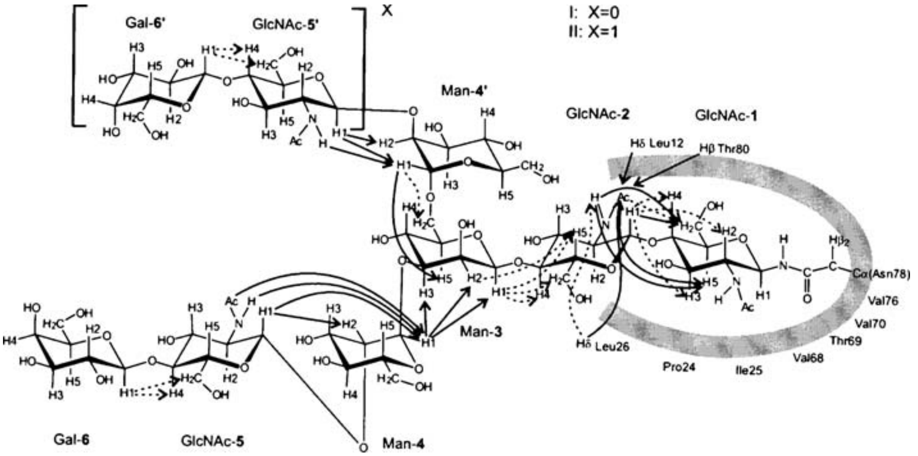


Fig. 22 NOE network of the glycan at Asn78 of αhCG. NOEs from GlcNAc-1 to the protein are identical with those published previously [144] and not indicated in this figure. Intraresidual NOEs and NOEs from GlcNAc-2 to protein are indicated with *arrows*. In case of overlap, *arrows* are *dashed*. (Reproduced with permission from [146])

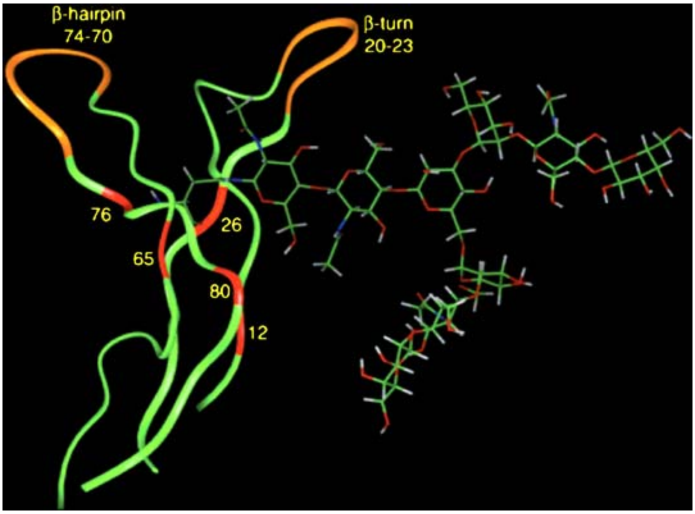


Fig. 23 Representation of the residues of αhCG influenced by the presence of the glycan at Asn78. The residues indicated in *red* show slightly different NOE intensities due to specific interactions with the glycan. The residues located in the heterogeneous β-turn and hairpin loop are drawn in *orange*. The residues that are unaffected by the presence of the glycan are represented in *green*. (Reproduced with permission from [112])

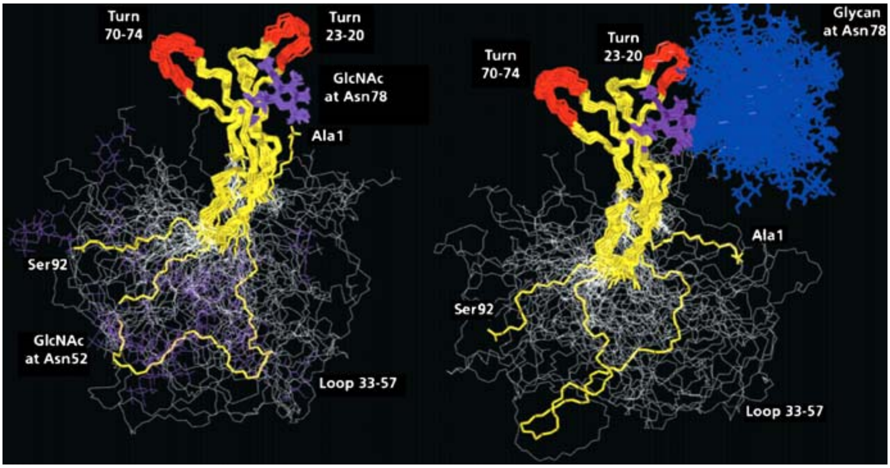


Fig. 24 Representation of the backbone conformation of the 26 conformers of αhCG[GlcNAc52,78] (*left*) and αhCG[glycan78] (*right*). The superposition was carried out including the backbone atoms of the residues 10–28, 59–70, and 75–84. The residues outside this region indicated in *white* represent the flexible and undefined part of the α-subunit, except one conformer which is represented in a *yellow ribbon*. The β-sheet is drawn in *yellow*. The tight turn and the hairpin loop are colored in *red*. GlcNAc-**1** is presented in *purple* and the residues beyond GlcNAc-**1** in *blue*. (Reproduced with permission from [112])

increased order within the β-hairpin loops (Figs. 23 and 24) [112]. Interestingly, the glycan at Asn78 leads to a similar stabilization of protein structure as the one of the adhesion domain of human CD2 [28]. However, the underlying molecular mechanism by which this effect is achieved is completely different in both cases. In the case of αhCG the glycan shields a hydrophobic surface of the protein from the solvent whereas in CD2 the glycan counterbalances five clustered positively charged lysines.

To further elucidate the role of the glycan at Asn52 of αhCG, Vliegenthart and coworkers compared native αhCG with a variant that had its Asn52 glycan enzymatically removed without affecting the glycan at Asn78 or disturbing the protein folding. The β-subunit was recombined in solution with either the alpha-subunit or the alpha-subunit enzymatically deglycosylated at αAsn52. The effect of the glycan at Asn52 on refolding and recombination of the α/β-hCG heterodimer was studied by SPR, CD and NMR spectroscopy and susceptibility toward proteolysis. The results give additional evidence that the glycan at Asn52 neither contributes significantly to the stability of the dimer nor is it responsible for large, overall, conformational differences with respect to the unglycosylated protein. However, minor changes regarding conformation and NOE pattern have been detected by CD and NMR spectroscopy, respectively. It is proposed that for bioactivity the bulky and extended glycan at αAsn-52 is necessary for inducing and stabilizing a conformational change in hCG upon binding to the receptor, resulting in activation of the signal-transduction pathway [147]. Whereas the role of the glycan at Asn78 has become relatively clear, the mechanism by which glycosylation at Asn52 renders αhCG biologically active is still a matter of debate. There is evidence for a variety of models including interaction of the glycans with secondary receptors, prevention of downregulation of the hCG receptor or induction of subtle structural changes. To answer this question, a structure determination of a functional heterodimer including complex, sialylated oligosaccharides might be the key.

5.1.4
Effect of *N*-Glycosylation on the Structure of Peptides

Imperiali and coworkers carried out extensive research on the influence of natural and non-natural glycosylation on the conformation of *N*-linked glycopeptides. In the case of peptides derived from influenza virus hemagglutinin, *N*-glycosylation of the sequence Ac-Orn-Ile-Thr-Pro-Asn-Gly-Thr-Trp-Ala-NH_2 with chitobiose (**14**) causes a conformational switching from an extended Asx-turn found in the peptide to a more closely packed type I β-turn of the glycopeptide. Astonishingly, this structural alteration is not accompanied by the formation of a strong NOE network between sugar and peptide but seems to be caused mainly by steric crowding or solvent-mediated effects [149]. The work on hemagglutinin glycopeptides was extended to the investigation of non-natural structures containing saccharide moieties other than chitobiose or a single GlcNAc residue. Peptides *β*-*N*-glycosylated with cellobiose (Glc(*β*1-4)Glc (**15**)), Glc(*β*1-4)GlcNAc (**16**), and GlcNAc(*β*1-4)Glc (**18**) were synthesized and characterized by NMR (Figs. 25 and 26) [150]. A dramatic effect on conformation was only found when GlcNAc was directly attached to asparagine leading to a preference for a *β*-turn conformation. A weak effect was observed when the *N*-acetyl group was present on the sugar residue distal from asparagine. The cellobiose carrying glycopeptide **15** that lacks *N*-acetyl groups was most flexible among the compounds studied and remains in an extended structure similar to the one found for the unglycosylated peptide. These findings were further strengthened by molecular

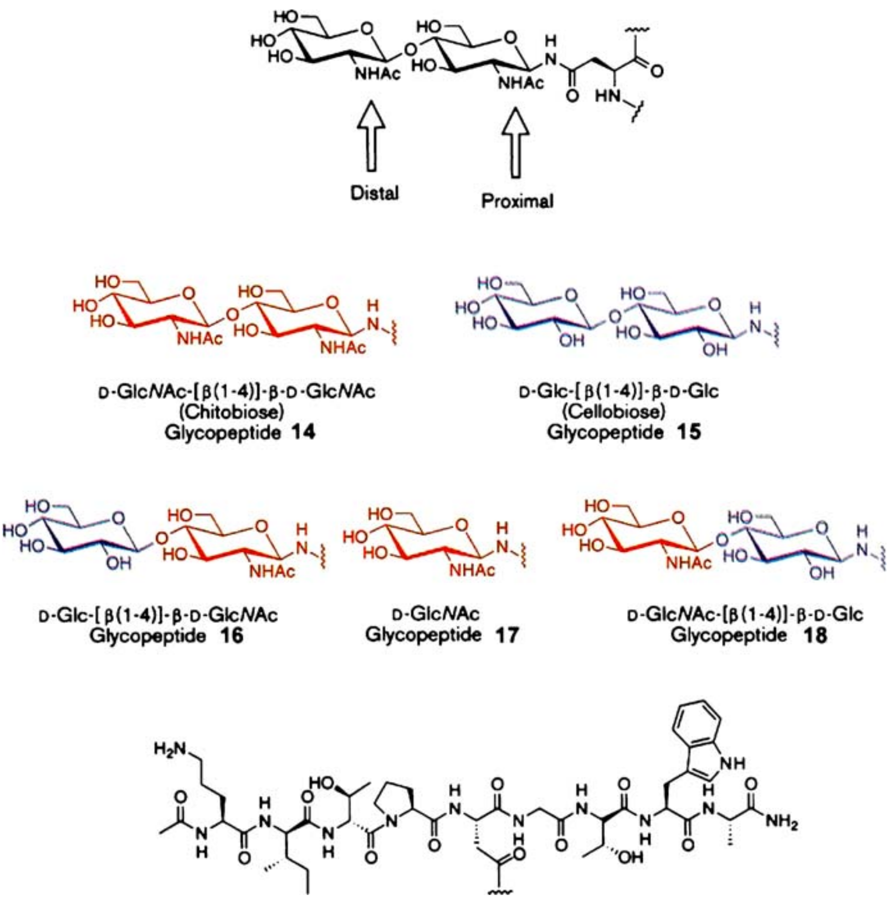
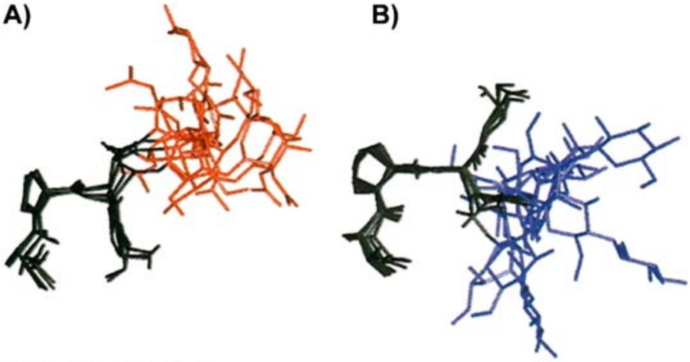
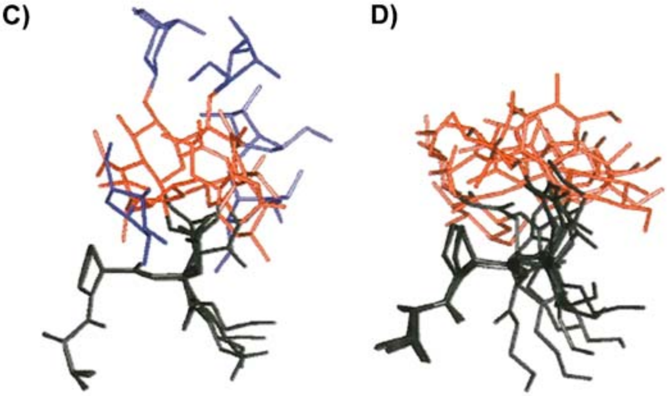


Fig. 25 Chemical structures of all glycopeptides examined by O'Conner et al. In each figure, *N*-acetylglucosamine is *red*, and glucose is *blue*. (Reproduced with permission from [150])



D-Glc*N*Ac-[β(1-4)]-β-D-Glc*N*Ac
(Chitobiose)
Glycopeptide **14**

D-Glc-[β(1-4)]-β-D-Glc
(Cellobiose)
Glycopeptide **15**



D-Glc-[β(1-4)]-β-D-Glc*N*Ac
Glycopeptide **16**

D-Glc*N*Ac
Glycopeptide **17**

D-Glc*N*Ac-[β(1-4)]-β-D-Glc
Glycopeptide **18**

◀ **Fig. 26** Minimized structures of the glycopeptides **14–18** shown in Fig. 25 resulting from a simulated annealing protocol that incorporated NOE-derived distance and $^3J_{HNH\alpha}$-derived dihedral restraints. Root mean square deviation (rmsd) values were calculated from all atoms. In all cases, the large rmsd value for asparagine reflects the highly disordered carbohydrate that is covalently linked to this residue. Residues Thr3-Gly6 are shown. **A** Five of ten NMR-derived structures of glycopeptide **14** that converge to a type I β-turn. Thr3, 0.76 ± 0.07 Å; Pro4, 1 ± 0.09 Å; Asn5, 6.5 ± 2.6 Å; Gly6, 2.3 ± 0.7 Å. **B** Seven of ten NMR-derived structures of glycopeptide **15**, illustrating the more extended backbone. Thr3, 1.2 ± 0.4 Å; Pro4, 0.9 ± 0.2 Å; Asn5, 7 ± 1.6 Å; Gly6, 1.8 ± 0.5 Å. **C** Five of ten NMR-derived structures of glycopeptide **16** that converge to a type I β-turn. Thr3, 0.82 ± 0.06 Å; Pro4, 0.88 ± 0.3 Å; Asn5, 7.9 ± 3.1 Å; Gly6, 1.2 ± 0.3 Å. **D** Seven of ten NMR-derived structures of glycopeptide **17** that converge to a disordered β-turn. Thr3, 0.86 ± 0.05 Å; Pro4, 0.76 ± 0.3 Å; Asn5, 4.1 ± 1.5 Å; Gly6, 2.4 ± 0.2 Å. **E** Six of ten NMR-derived structures of glycopeptide **18** that converge to a disordered β-turn. Thr3, 2.9 ± 1.6 Å; Pro4, 1.2 ± 0.1 Å; Asn5, 12.5 ± 3.7 Å; Gly6, 4.2 ± 0.7 Å. (Adapted and reproduced with permission from [150])

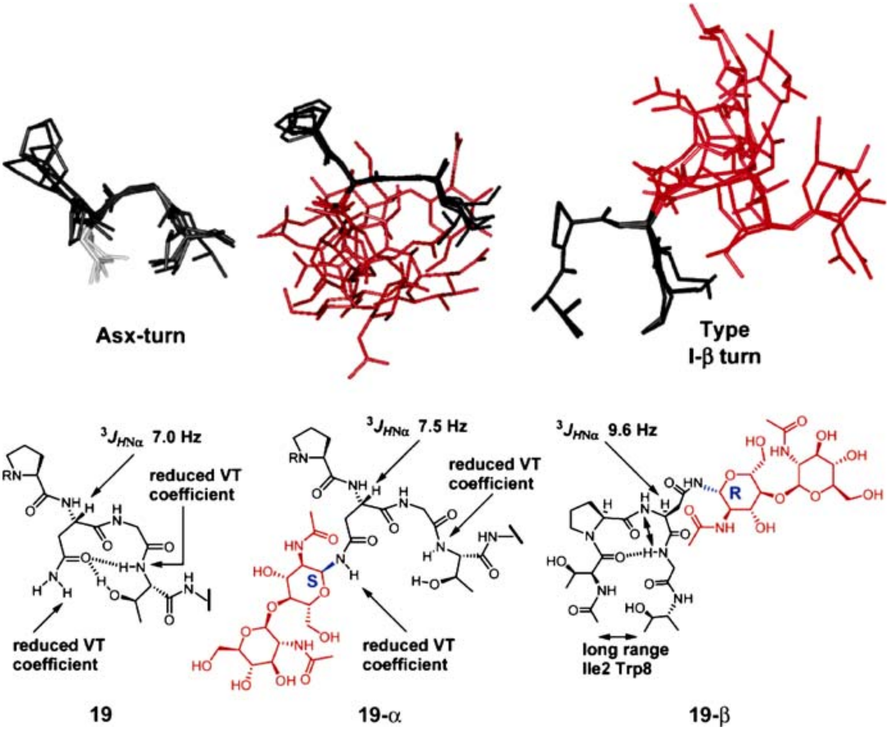


Fig. 27 Comparison of the solution-state structure between unglycosylated (**19**), α-chitobiosylated (**19**-α), and β-chitobiosylated (**19**-β) peptide Ac-Orn-Ile-Thr-Pro-Asn-Gly-Thr-Trp-Ala-NH_2 derived from hemagglutinin. **19** and **19**-α are shown from Pro4 to Thr7. **19**-β is shown from Thr3 to Thr7. (Reproduced with permission from [151])

modeling studies indicating a rigidifying effect of *N*-acetyl groups on the glycosidic linkage. The authors conclude that it is likely that the *N*-acetyl groups of the carbohydrates have a critical role in promoting the more compact β-turn conformation through steric interactions with the peptide.

Recently, a conformational analysis of a glycopeptide with α-*N*-linked chitobiose (**19-α**) was reported. Thereby, the impact of this non-natural glycosidic linkage could be compared to that of the commonly present β-linkage. It was shown that only the β-linkage is capable of inducing a type I β-turn whereas the α-glycosylated peptide adopted a conformation similar to the unmodified reference peptide displaying a more open Asx-turn-like structure in the vicinity of the glycosylation site (Fig. 27) [151]. In this regard it is noteworthy that statistical analysis of the stereochemistry of *N*-glycosylation sites carried out by Imberty and Pérez and by Petrescu et al. revealed that only about a quarter of glycosylated asparagines are located at a β-turn [20, 114]. This could mean, that the induction of β-turn structure by *N*-glycosylation is of transient character and of major importance only during the folding process.

5.2
Effect of *O*-Glycosylation on 3D Structure

5.2.1
Mucin-Type Glycoproteins

Danishefsky and coworkers carried out extensive research on the influence of *O*-glycosylation on mucin-type peptides. Glycopeptides derived from the CD43 glycoprotein were prepared by total synthesis with a range of glycans attached to adjacent serine and/or threonine residues. Not only a α-GalNAc monosaccharide (α-Tn-Antigen) but also di-(α-TF and β-TF), tri-(α-STF) and hexasaccharides (α-Le^y) were attached to peptide scaffolds (Fig. 28). The 3D structure of the triglycosylated peptide Ac-Ser(-O-α-STF)-Thr(-O-α-STF)-Thr(-O-α-STF)-Ala-Val (**24**) was investigated in detail by NMR (Fig. 29) [152, 153]. The peptide part turns out to be very similar to that of a hexaglycosylated peptide fragment of glycophorin A investigated by Schuster et al. [154]. There, the clustered glycosylation of adjacent amino acid residues in the glycopeptide His-Thr(O-α-D-GalNAc)-Ser(O-α-D-GalNAc)-Thr(O-α-D-GalNAc)-Ser(O-α-D-GalNAc)-Ser(O-α-D-GalNAc)-Ser(O-α-D-GalNAc)-Val-Thr-Lys (**26**) leads to an extended, "wave-type" structure with the saccharides projecting away from the peptide into alternating directions (Fig. 30). In terms of position and conformation of the glycan residues, there are obvious differences between the glycopeptide derived from glycophorin A and the one related to the mucin CD43, illustrated, for example, by deviating $\chi 1$-angles. In the case of the glycophorin A-derived glycopeptide **26**, strong NOEs between the GalNAc-NH and the

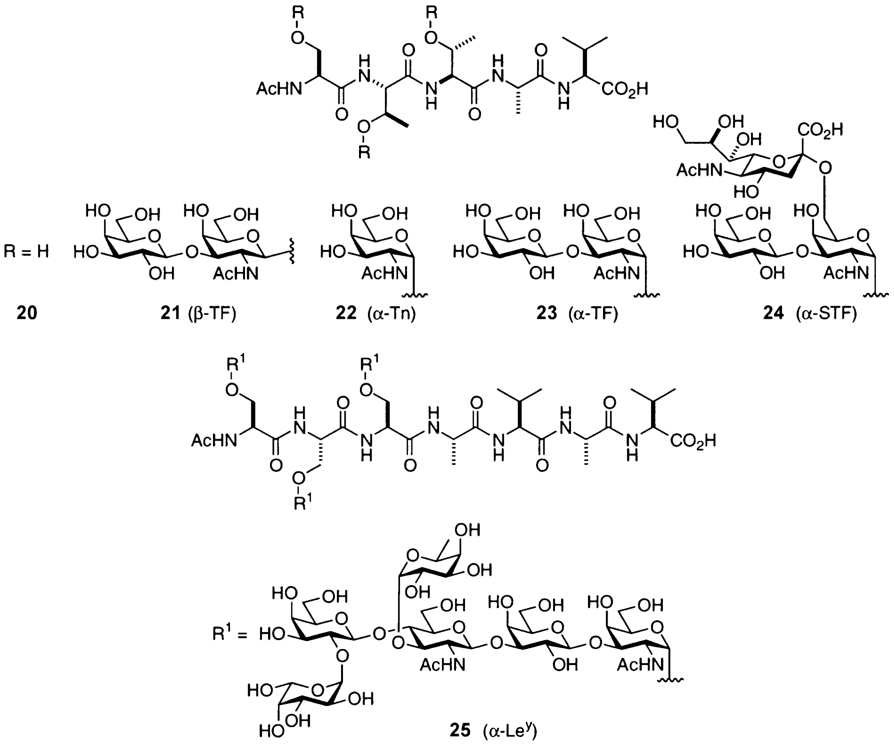


Fig. 28 Synthetic Muc-1 peptides with different *O*-Glycans attached to Ser and Thr residues. (Reproduced with permission from [152])

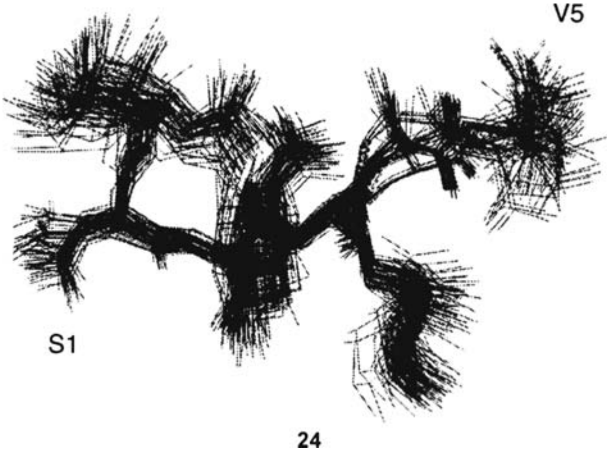


Fig. 29 Superposition of the peptide backbone and first sugar residue of the 59 best calculated structures of **24** from the restrained dynamics calculations (The chemical structure is shown in Fig. 28). (Reproduced with permission from [152])

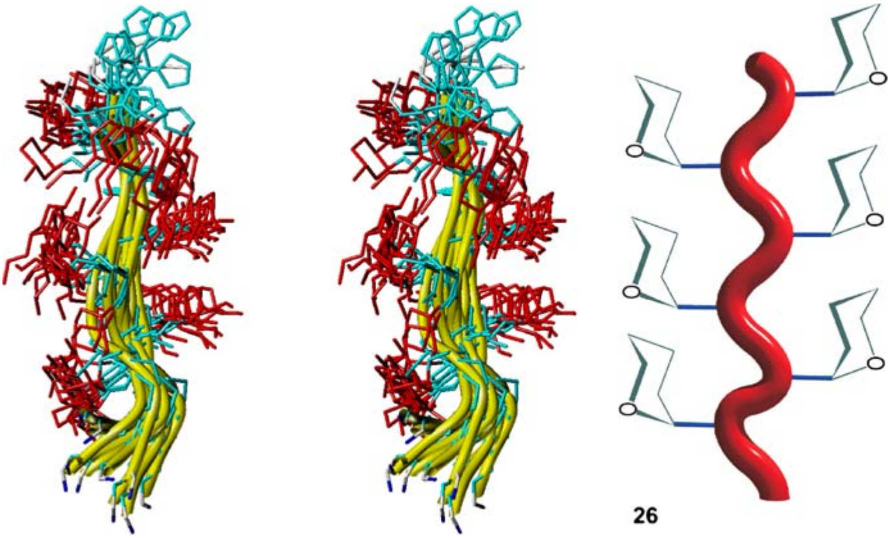

26

Fig. 30 Stereo plot of a superposition of 10 structures, obtained from a 1000 ps constrained MD simulation of the glycopeptide His-Thr(O-α-D-GalNAc)-Ser(O-α-D-GalNAc)-Thr(O-α-D-GalNAc)-Ser(O-α-D-GalNAc)-Ser(O-α-D-GalNAc)-Ser(O-α-D-Gal NAc)-Val-Thr-Lys (**26**) derived from glycophorin A. Each structure is a snapshot every 100 ps. The sugar residues are colored *red*, the peptide *cyan*. The peptide backbone conformation is highlighted by a *yellow tube*. On the *right* a schematic drawing of the "wave-type" structure with the sugars projecting toward the N-terminus is shown. (Adapted and reproduced with permission from [154])

peptide NH of the same amino acid were observed, in combination with other constraints supporting an orientation of the sugar towards the N-terminus. In contrast, in the CD43 glycopeptide **24**, experimental data are indicative for hydrogen bonding of the sugar NH to both the glycosidic oxygen and to the carbonyl oxygen of its own amino acid. In addition, this conformation allows for hydrophobic interactions between *N*-acetyl methyl groups and side chain methyl groups C-terminal from the site of glycosylation to stabilize the arrangement of the peptide backbone and of the sugars (Fig. 31). Structural rigidity is also supported by NMR-relaxation parameters.

The recent study on CD43 glycopeptides provides insight into the influence of inner and outer glycan residues and allows a comparison between native α- and the abnormal β-linkage. It was found that the innermost GalNAc residue, when attached α-anomerically, is key to the structural organization of these glycopeptides. A distinct NOE fingerprint was found for all glycopeptides with the native *O*-α-linkage, surprisingly independent from further glycosylation extending from the first GalNAc moiety. Strong induction of regular structure was observed for *O*-α-linked glycosylation only. In contrast,

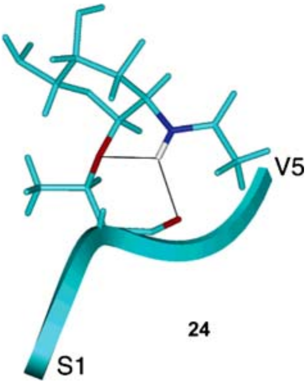


Fig. 31 Illustration of the hydrogen-bonding-type interactions of *N*-acetyl amide proton of the T3 GalNAc, shown by *black lines*. These are typical for all the glycosylated residues of **24** (cf. also Figs. 28 and 29). The relevant oxygens are *highlighted in red* and the nitrogen in *blue*. The backbone trace of the rest of the peptide portion is shown as a *ribbon*. (Reproduced with permission from [152])

the non-native *O*-β-linkage (compound **21**) did not seem to affect peptide conformation considerably leading to structures resembling the parent, unglycosylated peptides with a high degree of flexibility. This contrasts findings of Wong and coworkers who report significant conformational changes of a mucin peptide derived from MAdCAM-1 induced by glycosylation with β-O-SLe^x indicating a turn-like conformation in the vicinity of the glycosylated threonine [155, 156].

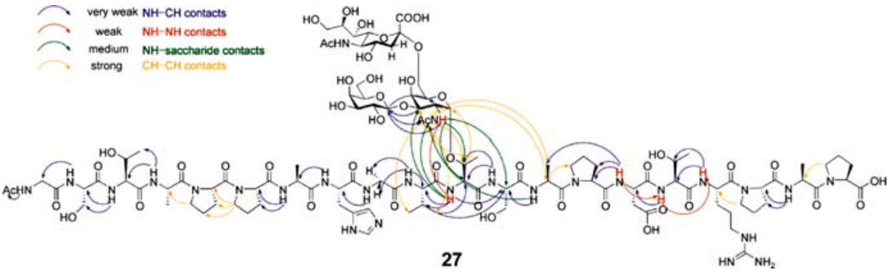
Levine and coworkers investigated the effect of *O*-glycosylation on the structure of human salivary mucin (MUC-7) glycopeptides. GalNAc-monosaccharide and Gal(β1-3)GalNAc-disaccharides were attached to adjacent serine or threonine residues of the two peptide sequences PAPPSSSAPPE and APPETTAAPPT derived from the tandem repeat region of MUC-7 [157]. The combined information from CD spectra, NMR chemical shifts, NOEs, and temperature dependence of amide proton chemical shifts, is indicative for an extended poly-L-proline type II or random conformation of the peptide backbone. Strong interaction with the peptide backbone is only reported for glycans attached to threonine. There, similar to the results of Coltart et al. [152], the NOE pattern supports the formation of an H-bond between the *N*-acetyl NH and the C-terminal carbonyl group of the threonine residue that carries the glycan. In the case of glycosylated serine, a reduced number of sugar-peptide NOEs is observed and there is less influence on chemical shifts of α-CH-protons of neighboring amino acids. Glycans attached to serine residues are thus likely to possess more rotational freedom and to have less influence on peptide conformation.

Kirnarsky et al. report on the conformation of the MUC-1 15-mer peptide PPAHGVTSAPDTRPA and the influence of glycosylation at Thr7 [158].

At lowered temperature (5–10 °C) and slightly acidic pH of 4.5, a weak but distinct effect on conformational propensities could be detected by cluster analysis of backbone conformation. Under all conditions studied, a strong NOE between the GalNAc-NH and the NH of the glycosylated Thr could be detected. In addition, weak dipolar interaction was present between the GalNAc methyl group and the methyl group of Ala9 indicative for a similar conformation as reported for the CD43-related glycopeptide [152].

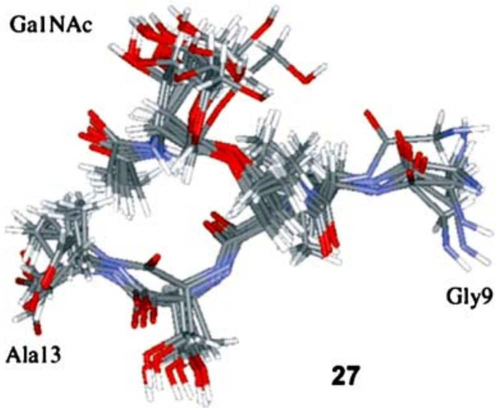
Very recently, Kunz and coworkers presented a conformational analysis of a MUC-1 glycopeptide (**27**) encompassing the whole tandem repeat of 20 amino acids with (*α*2,6)-sialyl-T-glycosylation at the same threonine residue as in the study of Kirnarsky et al. (Fig. 32) [158, 159]. In this case it was possible to carry out NMR experiments very close to physiological conditions at pH 6.5 and ambient temperature. Cluster analysis of the conformational ensemble yielded a rod-like structure with the GalNAc methyl group oriented towards the C-terminus, similar to results by Coltart et al. (Fig. 33) [152]. Grinstead et al. report on the influence of glycosylation on a MUC-1 21-mer peptide single and double Tn-glycosylated at Thr3 and Ser4, respectively [160]. Both glycosylation sites display specific NOE connectivities with the *N*-acetyl group of GalNAc at Thr3 oriented towards the N-terminus and the one at Ser4 contacting C-terminal residues.

Kirnarsky et al. investigated the effect of MUC-1 glycosylation at different sites of the tandem repeat, namely at Thr5, Thr10, and Thr17 of the sequence AHGVTSAPDTRPAPGSTAPPA [161]. The NMR-based structure determinations indicate that glycosylation at different sites leads to distinct alterations of the backbone conformation around the glycosylation site. In the case of glycosylation at Thr5 of the VTSA motif, only minor changes of the peptide backbone were observed. In contrast, attaching *O*-*α*-D-GalNAc to Thr17 of the STAP motif promotes a significant change of the peptide backbone around Thr17. The authors propose that these conformational requirements could be the cause for the finding that GalNAc-transferases show distinct preferences for each glycosylation site [161]. A higher conformational flexi-



27

Fig. 32 ROE contacts, determining the structure of a Sialyl-Tn-glycosylated MUC-1 peptide **27** encompassing the whole tandem repeat sequence. (Reproduced with permission from [159])



27

Fig. 33 The GVTSA sequence of glycopeptide **27** (chemical structure shown in Fig. 32) exhibits a rod-like secondary structure. Of the trisaccharide, only GalNAc is shown. (Reproduced with permission from [159])

bility of MUC-1 glycopeptides was reported by von Mensdorff-Pouilly et al. when MUC-1 sequence variants were investigated that have an ESR instead of a DTR motif. This could be a reason for differential antibody-mediated immune response observed in humans which is in turn related to the survival prognosis in breast cancer patients [162].

Interestingly, a type I β-turn-like structure was found for the immunodominant PDTRP region in the study of Dziadek et al. when Thr11 of the adjacent GVTSAP motif is glycosylated [159]. This epitope is often said to adopt a "knob-like" structure [163]. In contrast, the MUC-1 21-mer peptide studied by Kirnarsky et al. was found to have an S-shaped structure of the peptide backbone formed by consecutive inverse γ-turn, irrespective of glycosylation at Thr7 [158]. Also in the case of the diglycosylated MUC-1 21mer studied by Campbell and coworkers, a long range effect on the conformation of the immunodominant PDTRP epitope was not observed, however, their data supports a type I β-turn for both the peptide and the glycopeptide [160]. It is noteworthy that the breast cancer-specific antibody SM3 recognizes an extended structure in the case of an unglycosylated peptide as well as when binding to a glycopeptide bearing GalNAc at the central threonine of the PDTRP motif [164, 165].

In conclusion, it appears that the subtle interplay between peptide sequence, site of glycosylation, type of attached carbohydrate, and the clustering of *O*-glycans leads to induction of a variety of conformational motifs with a varying degree of stabilization. The unifying principle could be that all these conformations lead to a maximum exposure of presented carbohydrates facilitating interactions with receptors, especially in a multivalent fashion.

5.2.2
Antifreeze Glycoproteins

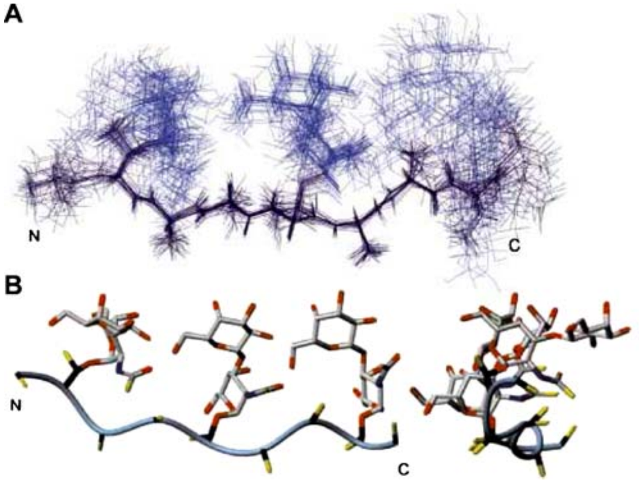
Antifreeze glycoproteins (AFGPs) have attracted attention due to their peculiar biophysical behavior as well as because of a variety of potential medical and commercial applications [166, 167]. AFGPs which can be isolated from fish blood plasma are complex mixtures comprising of repetitive Ala-Thr-Ala tripeptide units carrying (Gal(β1-3)GalNAc(α1-O) moieties at each threonine with some content in proline. Molecular weights range from 2.6 to 33 kDa. All AFGPs are said to inhibit freezing by binding to embryonic ice crystals, however, the precise mechanism of action is still unknown.

Recently, Tachibana et al. presented an extensive study on synthetic AFGPs composed of one to seven Ala-Thr-Ala repeating units [168]. Tripeptide oligomers with two or more units showed thermal hysteresis, a measure for the magnitude of the antifreeze effect, and were capable of changing the morphology of ice crystals into hexagonal bipyramids. Strong evidence indicates that the conformation of the peptide backbone plays a central role for the function of AFGPs as cryoprotectants. First, attaching GalNAc as a β-anomer leads to a complete loss of thermal hysteresis, and second, a glycopeptide made of Ala-Ser-Ala units exhibits no antifreeze activity. Both variations, and several others tested, either lead to loss of secondary structure or to a change from a polyproline type II helix in the active compounds to a largely α-helical structure in the case of the Ala-Ser-Ala repeating unit as determined by CD spectroscopy.

Despite the absence of long- and medium-range NOE contacts, a surprisingly well-defined structure could be calculated for a triple repeat of the Ala-(Gal(β1-3)GalNAc(α1-O)Thr-Ala unit (**28**) (Fig. 34). From a left-handed helix similar to the polyproline II helix all three disaccharides project to one side of the molecule forming a highly hydrophilic surface whereas the methyl groups of alanine and of the *N*-acetyl groups cluster on another face of the molecule. The authors propose that the amphipathic structure of this AGFP is crucial to its antifreeze activity. Interestingly, also non-glycosylated antifreeze proteins were found to possess amphipathic structure [169].

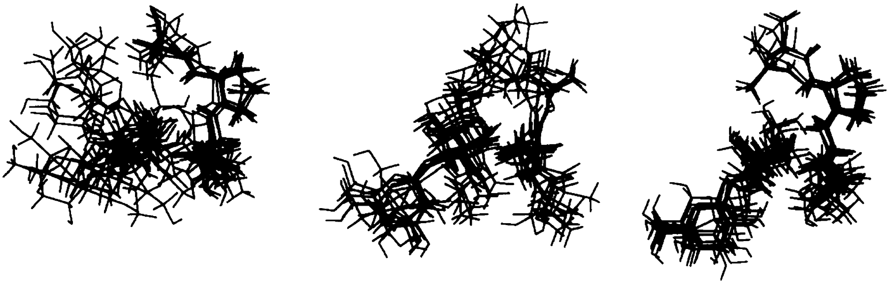
Even though the study on synthetic AFGPs by Nishimura and coworkers presents a consistent picture the situation for naturally occurring AFGPs seems to be more complicated [167]. Structural studies on these molecules, isolated from fish, have a long history yielding conflicting interpretations [170–175].

In the latest NMR study on the tetra-[Gal(β1-3)GalNAc(α1-O)]-glycosylated 14-mer peptide AATAATPATAATPA (**29**), known as AFGP fraction 8 or AFGP-8, from arctic cod, Lane et al. find relatively well-determined ATPA segments resembling polyproline helix conformation according to NOEs and scalar couplings (Fig. 35). ^{13}C relaxation data indicate motional anisotropy,



28

Fig. 34 **A** Superposition of the 25 lowest energy structures of the synthetic antifreeze glycopeptide [Ala-(Gal(β1-3)GalNAc(α1-O))Thr-Ala]$_3$ (**28**) calculated from NMR-based constraints. The peptide backbone is *navy*, and the carbohydrate moieties are *royal blue*. **B** Structure closest to the average of the 25 best calculated models for **28**: *yellow*, methyl carbon; *white*, carbon in carbohydrate; *gray*, carbon in peptide side chains except for methyl carbon; *blue*, nitrogen; *red*, oxygen; the N- and C-terminal ends are identified. (Reproduced with permission from [168])



29 (A5-A8) **29** (A8-A11) **29** (A11-A14)

Fig. 35 Preferred conformations of the AFGP-8 peptide (**29**). Segments 2 (A5-A8), 3 (A8-A11), and 4 (A11-A14) are shown *left to right* as superpositions of eight structures. (Reproduced with permission from [175])

however, long-range order in the form of a single preferred conformer could not be detected [175].

Nguyen et al. followed-up with 20 ns molecular dynamics simulations starting from 10 different conformers of the NMR ensemble obtained by Lane et al. [176]. From the simulation it was concluded that the AFGP8 backbone is structurally segmented into four semi-rigid segments while its disaccharides adopt well-defined conformations with respect to the peptide backbone.

H-bonding between *N*-acetyl NH groups and the peptide backbone stabilize the structure. The polyproline II helix is the overall conformation, yet AFGP8 can reach many distinct and energetically equal structures. The authors propose that these interconverting conformations represent a thermal reservoir that inhibits ice crystal growth by local energy transfer.

6
Concluding Remarks

By combining NMR spectroscopy, X-ray crystallography and molecular modeling, detailed insight into the influence of *O*- and *N*-glycosylation on peptide and protein conformation and dynamics has been obtained. From this, the structural role of glycosylation for biological function could be identified in several cases. Synthetic glycopeptides have proven to be invaluable model systems for studying the impact of specific structural elements on conformation and molecular flexibility, and for identifying minimal glycopeptide epitopes suitable for medicinal applications. Future research will show whether our current knowledge already represents the general principles of glycoprotein structure and function. To answer this question, it will be of central importance to develop and optimize experimental and theoretical methods for handling molecules that are highly hydrated and possess, at least in part, a high degree of conformational flexibility. This will profit immensely from our capability to prepare more and more complex, yet defined glycoconjugates, either by chemical or by molecular biology techniques.

References

1. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) Nucleic Acids Res 28:235
2. Apweiler R, Hermjakob H, Sharon N (1999) Biochim Biophys Acta 1473:4
3. Dwek RA (1998) Dev Biol Stand 96:43
4. Borman S (2005) Chem Eng News 83:41–50
5. Bush CA, Martin-Pastor M, Imberty A (1999) Annu Rev Biophys Biomol Struct 28:269
6. Wyss DF, Wagner G (1996) Curr Opin Biotechnol 7:409
7. Imperiali B, O'Connor SE (1999) Curr Opin Chem Biol 3:643
8. Haase C, Seitz O (2007) Top Curr Chem (in this volume)
9. Duus J, Gotfredsen CH, Bock K (2000) Chem Rev 100:4589
10. Woods RJ (1998) Glycoconj J 15:209
11. Imberty A, Pérez S (2000) Chem Rev 100:4567
12. Wormald MR, Petrescu AJ, Pao YL, Glithero A, Elliott T, Dwek RA (2002) Chem Rev 102:371
13. Liu HL, Hsu JP (2005) Proteomics 5:2056
14. Heinemann U, Illing G, Oschkinat H (2001) Curr Opin Biotechnol 12:348

15. Sattler M, Schleucher J, Griesinger C (1999) Prog NMR Spectrosc 34:93
16. Dauter Z (2006) Acta Crystallogr D Biol Crystallogr 62:1
17. Williams SP, Kuyper LF, Pearce KH (2005) Curr Opin Chem Biol 9:371
18. Grogan MJ, Pratt MR, Marcaurelle LA, Bertozzi CR (2002) Annu Rev Biochem 71:593
19. Fletcher CM, Harrison RA, Lachmann PJ, Neuhaus D (1994) Structure 2:185
20. Imberty A, Pérez S (1995) Protein Eng 8:699
21. Rudd PM, Morgan BP, Wormald MR, Harvey DJ, van den Berg CW, Davis SJ, Ferguson MA, Dwek RA (1997) J Biol Chem 272:7229
22. Yokoyama S (2003) Curr Opin Chem Biol 7:39
23. Pickford AR, O'Leary JM (2004) Methods Mol Biol 278:17
24. Li P, Gao XG, Arellano RO, Renugopalakrishnan V (2001) Protein Expr Purif 22:369
25. Jarvis DL (2003) Virology 310:1
26. Lustbader JW, Birken S, Pollak S, Pound A, Chait BT, Mirza UA, Ramnarain S, Canfield RE, Brown JM (1996) J Biomol NMR 7:295
27. Hansen AP, Petros AM, Mazar AP, Pederson TM, Rueter A, Fesik SW (1992) Biochemistry 31:12713
28. Wyss DF, Choi JS, Li J, Knoppers MH, Willis KJ, Arulanandam AR, Smolyar A, Reinherz EL, Wagner G (1995) Science 269:1273
29. Wood MJ, Komives EA (1999) J Biomol NMR 13:149
30. Cubeddu L, Moss CX, Swarbrick JD, Gooley AA, Williams KL, Curmi PM, Slade MB, Mabbutt BC (2000) Protein Expr Purif 19:335
31. Wendeler M, Hoernschemeyer J, John M, Werth N, Schoeniger M, Lemm T, Hartmann R, Kessler H, Sandhoff K (2004) Protein Expr Purif 34:147
32. Laroche Y, Storme V, De Meutter J, Messens J, Lauwereys M (1994) Biotechnology (NY) 12:1119
33. McAlister MS, Davis B, Pfuhl M, Driscoll PC (1998) Protein Eng 11:847
34. Weikert S, Papac D, Briggs J, Cowfer D, Tom S, Gawlitzek M, Lofgren J, Mehta S, Chisholm V, Modi N, Eppler S, Carroll K, Chamow S, Peers D, Berman P, Krummen L (1999) Nat Biotechnol 17:1116
35. Xu R, Hanson SR, Zhang Z, Yang YY, Schultz PG, Wong CH (2004) J Am Chem Soc 126:15654
36. Zhang Z, Gildersleeve J, Yang YY, Xu R, Loo JA, Uryu S, Wong CH, Schultz PG (2004) Science 303:371
37. Meinjohanns E, Meldal M, Paulsen H, Dwek RA, Bock K (1998) J Chem Soc Perkin Trans 1:549
38. George SK, Schwientek T, Holm B, Reis CA, Clausen H, Kihlberg J (2001) J Am Chem Soc 123:11117
39. Thayer DA, Wong CH (2007) Top Curr Chem (in this volume)
40. Warren DJ, Geng X, Danishefsky SJ (2007) Top Curr Chem (in this volume)
41. Pratt MR, Bertozzi CR (2005) Chem Soc Rev 34:58
42. Holemann A, Seeberger PH (2004) Curr Opin Biotechnol 15:615
43. Herzner H, Reipen T, Schultz M, Kunz H (2000) Chem Rev 100:4495
44. Dawson PE, Muir TW, Clark-Lewis I, Kent SB (1994) Science 266:776
45. Evans TC Jr, Benner J, Xu MQ (1998) Protein Sci 7:2256
46. Macmillan D, Bertozzi CR (2004) Angew Chem Int Ed Engl 43:1355
47. Hojo H, Watabe J, Nakahara Y, Nakahara Y, Ito Y, Nabeshima K, Toole BP (2001) Tetrahedron Lett 42:3001
48. Hackenberger CP, Friel CT, Radford SE, Imperiali B (2005) J Am Chem Soc 127:12882
49. Smith KD, Davies MJ, Hounsell EF (1994) Methods Mol Biol 32:143
50. Dell A, Morris HR (2001) Science 291:2351

51. Peter-Katalinic J (2005) Methods Enzymol 405:139
52. Rudd PM, Dwek RA (1997) Curr Opin Biotechnol 8:488
53. Gallego RG, Blanco JL, Thijssen-van Zuylen CW, Gotfredsen CH, Voshol H, Duus JO, Schachner M, Vliegenthart JF (2001) J Biol Chem 276:30834
54. Halkes KM, Gotfredsen CH, Grotli M, Miranda LP, Duus JO, Meldal M (2001) Chemistry 7:3584
55. Wüthrich K (1986) NMR of Proteins and Nucleic Acids. Wiley, New York
56. Kessler H, Seip S (1994) NMR of Peptides. In: Croasmun WR, Carlson RMK (eds) Two-Dimensional NMR Spectroscopy. VCH, New York, p 619
57. Leeflang BR, Faber EJ, Erbel P, Vliegenthart JF (2000) J Biotechnol 77:115
58. Seavey BR, Farr EA, Westler WM, Markley JL (1991) J Biomol NMR 1:217
59. Yamaguchi Y, Takizawa T, Kato K, Arata Y, Shimada I (2000) J Biomol NMR 18:357
60. Bock K, Thoegersen H (1982) Annu Rep NMR Spectrosc 13:1
61. Bock K, Pedersen C (1983) Adv Carbohydr Chem Biochem 41:27
62. Prytulla S, Lambert J, Lauterwein J, Klessinger M, Thiem J (1990) Magn Reson Chem 28:888
63. Prytulla S, Lauterwein J, Klessinger M, Thiem J (1991) Carbohydr Res 215:345
64. Haverkamp J, Spoormaker T, Dorland L, Vliegenthart JFG, Schauer R (1979) J Am Chem Soc 101:4851
65. Lommerse JP, van Rooijen JJ, Kroon-Batenburg LM, Kamerling JP, Vliegenthart JF (2002) Carbohydr Res 337:2279
66. Homans SW (1992) Glycobiology 2:153
67. Milton MJ, Harris R, Probert MA, Field RA, Homans SW (1998) Glycobiology 8:147
68. Harris R, Rutherford TJ, Milton MJ, Homans SW (1997) J Biomol NMR 9:547
69. Poppe L, van Halbeek H (1991) J Magn Reson 92:636–41
70. Poppe L, van Halbeek H (1991) J Magn Reson 93:214–17
71. Martin-Pastor M, Bush CA (1999) Biochemistry 38:8045
72. Bose B, Zhao S, Stenutz R, Cloran F, Bondo PB, Bondo G, Hertz B, Carmichael I, Serianni AS (1998) J Am Chem Soc 120:11158
73. Xu Q, Gitti R, Bush CA (1996) Glycobiology 6:281
74. Mulloy B, Frenkiel TA, Davies DB (1988) Carbohydr Res 184:39
75. Tvaroska I, Hricovini H, Perakova E (1989) Carbohydr Res 189:359
76. Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH (1995) Proc Natl Acad Sci USA 92:9279
77. Tjandra N, Bax A (1997) Science 278:1111
78. Prestegard JH, Bougault CM, Kishore AI (2004) Chem Rev 104:3519
79. de Alba E, Tjandra N (2002) Prog Nucl Magn Reson 40:175
80. Griesinger C, Peti W, Meiler J, Bruschweiler R (2004) Methods Mol Biol 278:107
81. Deschamps M, Campbell ID, Boyd J (2005) J Magn Reson 172:118
82. Meiler J, Prompers JJ, Peti W, Griesinger C, Bruschweiler R (2001) J Am Chem Soc 123:6098
83. Peti W, Meiler J, Bruschweiler R, Griesinger C (2002) J Am Chem Soc 124:5822
84. Azurmendi HF, Bush CA (2002) Carbohydr Res 337:905
85. Almond A, Duus JO (2001) J Biomol NMR 20:351
86. Almond A, Petersen BO, Duus JO (2004) Biochemistry 43:5853
87. Reif B, Hennig M, Griesinger C (1997) Science 276:1230
88. Reif B, Diener A, Hennig M, Maurer M, Griesinger C (2000) J Magn Reson 143:45
89. Scheurer C, Skrynnikov NR, Lienin SF, Straus SK, Bruschweiler R, Ernst RR (1999) J Am Chem Soc 121:4242
90. Markwick PR, Sprangers R, Sattler M (2005) Angew Chem Int Ed Engl 44:3232

91. Kover KE, Batta G, Kowalewski J, Ghalebani L, Kruk D (2004) J Magn Reson 167:273
92. Larsson EA, Urbina F, Yang Z, Weintraub A, Widmalm G (2004) Carbohydr Res 339:1491
93. Chen B, Vogan EM, Gong H, Skehel JJ, Wiley DC, Harrison SC (2005) Structure (Camb) 13:197
94. Chen B, Vogan EM, Gong H, Skehel JJ, Wiley DC, Harrison SC (2005) Nature 433:834
95. Case DA, Cheatham TE III, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) J Comput Chem 26:1668
96. Homans SW (1990) Biochemistry 29:9110
97. Woods RJ, Dwek RA, Edge CJ, Fraser-Reid B (1995) J Phys Chem 99:3832
98. Glennon TM, Merz KMJ (1997) Theochem 395–396:157
99. Momany FA, Willett JL (2000) Carbohydr Res 326:194
100. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) J Comp Chem 4:187
101. Ha SN, Giammona A, Field M, Brady JW (1988) Carbohydr Res 180:207
102. Reiling S, Schlenkrich M, Brickmann J (1996) J Comp Chem 17:450
103. Christen M, Hunenberger PH, Bakowies D, Baron R, Burgi R, Geerke DP, Heinz TN, Kastenholz MA, Krautler V, Oostenbrink C, Peter C, Trzesniak D, van Gunsteren WF (2005) J Comput Chem 26:1719
104. Ott K-H, Meyer B (1996) J Comp Chem 17:1068
105. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) J Am Chem Soc 118:11225
106. Damm W, Frontera A, Tirado-Rives J, Jorgensen WL (1997) J Comp Chem 18:1955
107. Clark M, Cramer RDI, Van Opdenbosch N (1989) J Comp Chem 10:982
108. Imberty A, Hardman KD, Carver JP, Pérez S (1991) Glycobiology 1:631
109. Pérez S, Imberty A, Engelsen SB, Gruza J, Mazeau K, Jimenez-Barbero J, Poveda A, Espinosa J-F, Van Eyck BP, Johnson G, French AD, Kouwijzer MLCE, Grootenuis PDJ, Bernardi A, Raimondi L, Senderowitz H, Durier V, Vergoten G, Rasmussen K (1998) Carbohydr Res 314:141
110. van Gunsteren WF, Brunne RM, Gros P, van Schaik RC, Schiffer CA, Torda AE (1994) Methods Enzymol 239:619
111. Weis WI, Brunger AT, Skehel JJ, Wiley DC (1990) J Mol Biol 212:737
112. Erbel PJ, Karimi-Nejad Y, van Kuik JA, Boelens R, Kamerling JP, Vliegenthart JF (2000) Biochemistry 39:6012
113. Petrescu AJ, Petrescu SM, Dwek RA, Wormald MR (1999) Glycobiology 9:343
114. Petrescu AJ, Milac AL, Petrescu SM, Dwek RA, Wormald MR (2004) Glycobiology 14:103
115. Wormald MR, Wooten EW, Bazzo R, Edge CJ, Feinstein A, Rademacher TW, Dwek RA (1991) Eur J Biochem 198:131
116. Davis JT, Hirani S, Bartlett C, Reid BR (1994) J Biol Chem 269:3331
117. Kessler H, Matter H, Gemmecker G, Kling A, Kottenhahn M (1991) J Am Chem Soc 113:7550
118. Christlet TH, Biswas M, Veluraja K (1999) Acta Crystallogr D Biol Crystallogr 55(Pt8):1414
119. Bause E, Legler G (1981) Biochem J 195:639
120. Marshall RD (1972) Annu Rev Biochem 41:673
121. Berman H, Henrick K, Nakamura H (2003) Nat Struct Biol 10:980
122. Loss A, Bunsmann P, Bohne A, Loss A, Schwarzer E, Lang E, von der Lieth CW (2002) Nucleic Acids Res 30:405
123. Raman R, Venkataraman M, Ramakrishnan S, Lang W, Raguram S, Sasisekharan R (2006) Glycobiology 16:82R

124. Hashimoto K, Goto S, Kawano S, Aoki-Kinoshita KF, Ueda N, Hamajima M, Kawasaki T, Kanehisa M (2006) Glycobiology 16:63R
125. Gupta R, Birch H, Rapacki K, Brunak S, Hansen JE (1999) Nucleic Acids Res 27:370
126. Julenius K, Molgaard A, Gupta R, Brunak S (2005) Glycobiology 15:153
127. Cooper CA, Joshi HJ, Harrison MJ, Wilkins MR, Packer NH (2003) Nucleic Acids Res 31:511
128. Doubet S, Bock K, Smith D, Darvill A, Albersheim P (1989) Trends Biochem Sci 14:475
129. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA (1998) Nature 393:648
130. Wyatt R, Kwong PD, Desjardins E, Sweet RW, Robinson J, Hendrickson WA, Sodroski JG (1998) Nature 393:705
131. Kwong PD, Wyatt R, Sattentau QJ, Sodroski J, Hendrickson WA (2000) J Virol 74:1961
132. Huang CC, Tang M, Zhang MY, Majeed S, Montabana E, Stanfield RL, Dimitrov DS, Korber B, Sodroski J, Wilson IA, Wyatt R, Kwong PD (2005) Science 310:1025
133. Lin PF, Blair W, Wang T, Spicer T, Guo Q, Zhou N, Gong YF, Wang HG, Rose R, Yamanaka G, Robinson B, Li CB, Fridell R, Deminie C, Demers G, Yang Z, Zadjura L, Meanwell N, Colonno R (2003) Proc Natl Acad Sci USA 100:11013
134. Withka JM, Wyss DF, Wagner G, Arulanandam AR, Reinherz EL, Recny MA (1993) Structure 1:69
135. Wyss DF, Dayie KT, Wagner G (1997) Protein Sci 6:534
136. Kitao A, Wagner G (2000) Proc Natl Acad Sci USA 97:2064
137. Sun ZY, Dotsch V, Kim M, Li J, Reinherz EL, Wagner G (1999) EMBO J 18:2941
138. Wang JH, Smolyar A, Tan K, Liu JH, Kim M, Sun ZY, Wagner G, Reinherz EL (1999) Cell 97:791
139. Vliegenthart JFG, Dorland L, Van Halbeek H (1983) Adv Carbohydr Chem Biochem 41:209
140. Wu H, Lustbader JW, Liu Y, Canfield RE, Hendrickson WA (1994) Structure 2:545
141. Lapthorn AJ, Harris DC, Littlejohn A, Lustbader JW, Canfield RE, Machin KJ, Morgan FJ, Isaacs NW (1994) Nature 369:455
142. Lustbader JW, Pollak S, Lobel L, Trakht I, Homans S, Brown JM, Canfield RE (1996) Mol Cell Endocrinol 125:21
143. Weller CT, Lustbader J, Seshadri K, Brown JM, Chadwick CA, Kolthoff CE, Ramnarain S, Pollak S, Canfield R, Homans SW (1996) Biochemistry 35:8815
144. De Beer T, Van Zuylen CW, Leeflang BR, Hard K, Boelens R, Kaptein R, Kamerling JP, Vliegenthart JF (1996) Eur J Biochem 241:229
145. van Zuylen CW, Kamerling JP, Vliegenthart JF (1997) Biochem Biophys Res Commun 232:117
146. van Zuylen CW, de Beer T, Leeflang BR, Boelens R, Kaptein R, Kamerling JP, Vliegenthart JF (1998) Biochemistry 37:1933
147. Erbel PJ, Haseley SR, Kamerling JP, Vliegenthart JF (2002) Biochem J 364:485
148. Kobata A, Takeuchi M (1999) Biochim Biophys Acta 1455:315
149. O'Connor SE, Imperiali B (1997) J Am Chem Soc 119:2295
150. O'Conner SE, Imperiali B (1998) Chem Biol 5:427
151. Bosques CJ, Tschampel SM, Woods RJ, Imperiali B (2004) J Am Chem Soc 126:8421
152. Coltart DM, Royyuru AK, Williams LJ, Glunz PW, Sames D, Kuduk SD, Schwarz JB, Chen XT, Danishefsky SJ, Live DH (2002) J Am Chem Soc 124:9833
153. Live DH, Williams LJ, Kuduk SD, Schwarz JB, Glunz PW, Chen XT, Sames D, Kumar RA, Danishefsky SJ (1999) Proc Natl Acad Sci USA 96:3489

154. Schuster O, Klich G, Sinnwell V, Kranz H, Paulsen H, Meyer B (1999) J Biomol NMR 14:33
155. Seitz O, Wong CH (1997) J Am Chem Soc 119:8766
156. Wu W, Pasternack L, Huang D-H, Koeller KM, Lin C-C, Seitz O, Wong C-H (1999) J Am Chem Soc 121:2409
157. Naganagowda GA, Gururaja TL, Satyanarayana J, Levine MJ (1999) J Pept Res 54:290
158. Kirnarsky L, Prakash O, Vogen SM, Nomoto M, Hollingsworth MA, Sherman S (2000) Biochemistry 39:12076
159. Dziadek S, Griesinger C, Kunz H, Reinscheid UM (2006) Chem Eur J 12:4981
160. Grinstead JS, Koganty RR, Krantz MJ, Longenecker BM, Campbell AP (2002) Biochemistry 41:9946
161. Kinarsky L, Suryanarayanan G, Prakash O, Paulsen H, Clausen H, Hanisch FG, Hollingsworth MA, Sherman S (2003) Glycobiology 13:929
162. von Mensdorff-Pouilly S, Kinarsky L, Engelmann K, Baldus SE, Verheijen RH, Hollingsworth MA, Pisarev V, Sherman S, Hanisch FG (2005) Glycobiology 15:735
163. Fontenot JD, Mariappan SV, Catasti P, Domenech N, Finn OJ, Gupta G (1995) J Biomol Struct Dyn 13:245
164. Dokurno P, Bates PA, Band HA, Stewart LM, Lally JM, Burchell JM, Taylor-Papadimitriou J, Snary D, Sternberg MJ, Freemont PS (1998) J Mol Biol 284:713
165. Möller H, Serttas N, Paulsen H, Burchell JM, Taylor-Papadimitriou J, Meyer B (2002) Eur J Biochem 269:1444
166. Bouvet V, Ben RN (2003) Cell Biochem Biophys 39:133
167. Harding MM, Anderberg PI, Haymet AD (2003) Eur J Biochem 270:1381
168. Tachibana Y, Fletcher GL, Fujitani N, Tsuda S, Monde K, Nishimura S (2004) Angew Chem Int Ed Engl 43:856
169. Sicheri F, Yang DS (1995) Nature 375:427
170. Bush CA, Feeney RE, Osuga DT, Ralapati S, Yeh Y (1981) Int J Pept Protein Res 17:125
171. Rao BN, Bush CA (1987) Biopolymers 26:1227
172. Filira F, Biondi L, Scolaro B, Foffani MT, Mammi S, Peggion E, Rocchi R (1990) Int J Biol Macromol 12:41
173. Dill K, Huang L, Bearden DW, Feeney RE (1992) J Carbohydr Chem 11:499
174. Mimura Y, Yamamoto Y, Inoue Y, Chujo R (1992) Int J Biol Macromol 14:242
175. Lane AN, Hays LM, Feeney RE, Crowe LM, Crowe JH (1998) Protein Sci 7:1555
176. Nguyen DH, Colvin ME, Yeh Y, Feeney RE, Fink WH (2002) Biophys J 82:2892

Author Index Volumes 251–272

Author Index Vols. 26–50 see Vol. 50
Author Index Vols. 51–100 see Vol. 100
Author Index Vols. 101–150 see Vol. 150
Author Index Vols. 151–200 see Vol. 200
Author Index Vols. 201–250 see Vol. 250

The volume numbers are printed in italics

Subject Index