Chem Soc Rev

RSC Publishing

REVIEW ARTICLE

View Article Online
View Journal | View Issue

Cite this: *Chem. Soc. Rev.*, 2013, **42**, 4408

Chemical assembly of *N*-glycoproteins: a refined toolbox to address a ubiquitous posttranslational modification†‡

Carlo Unverzagt*[a] and Yasuhiro Kajihara*[b]

Received 28th November 2012

DOI: 10.1039/c3cs35485g

www.rsc.org/csr

Incremental developments in the chemistry of peptides, proteins and carbohydrates have enabled researchers to assemble entire glycoproteins with high precision. Based on sophisticated ligation chemistries pure glycoproteins bearing a single glycosylation pattern have become available. The impact of N-glycosylation on the function of glycoproteins is generally recognized but not well understood. Based on the recent advances in the synthesis of glycoproteins by chemical methods researchers can finally start to elucidate the various roles of carbohydrates in complex biomolecules in detail.

1. Introduction

The many functions of proteins associated with their complex three-dimensional architectures are fascinating to scientists from many fields. Additional functionality of proteins is provided by their covalent modification with sugars, lipids, phosphates and other residues summarized as posttranslational modifications (PTMs). Phosphorylation is frequently involved in signal transduction, ubiquitination regulates proteolysis, and the attachment of lipids results in membrane anchoring. Depending on the type of PTM, the physical and chemical properties of a protein including folding, conformational properties and stability are altered resulting in a modified functionality of the protein.[1] Thus PTMs have received considerable attention. Glycosylation is a frequently encountered posttranslational modification based on a complex biosynthetic

[a] *Bioorganische Chemie, Gebäude NWI, Universität Bayreuth, 95440 Bayreuth, Germany. E-mail: carlo.unverzagt@uni-bayreuth.de; Fax: +49-921-555365; Tel: +49-921-552670*

[b] *Department of Chemistry, Graduate School of Science, Osaka University, 1-1, Machikaneyama, Toyonaka, 560-0043, Japan. E-mail: kajihara@chem.sci.osaka-u.ac.jp; Fax: +81-6-6850-5382; Tel: +81-6-6850-5380*

† Part of the carbohydrate chemistry themed issue.

‡ This work is dedicated to Prof. Werner Reutter on the occasion of his 75th birthday.

Carlo Unverzagt

Carlo Unverzagt studied chemistry in Mainz where he received his doctorate in 1988 with Prof. H. Kunz. After postdoctoral training at UCLA (Prof. J. Paulson) he moved to Technical University München in 1990 (Prof. H. Kessler) where he obtained his habilitation in 1997. Subsequently, he was appointed as a Professor for Organic Chemistry at the University of München and since 1998 he has held a chair in Bioorganic Chemistry at the University of Bayreuth. His research interests focus on the chemistry and biochemistry of carbohydrates, peptides, proteins and natural products.

Yasuhiro Kajihara

Yasuhiro Kajihara received his PhD from Tokyo Institute of Technology in 1993 under the supervision of Prof. H. Hashimoto. He spent two years at the Life Science Research Laboratory of Japan Tobacco Inc. as a post doctoral fellow. During this period he studied the synthesis of glycosyltransferase inhibitors and methods for sugar nucleotide synthesis. In 1995 he joined Yokohama City University as an Assistant Professor and was then promoted to Associate Professor in 2001 and Full Professor in 2007. At YC, he developed synthetic methods for oligosaccharides as well as for glycoproteins. In 2009 he moved to the Department of Chemistry at Osaka University.

machinery affecting protein function in many ways.[2] Nearly all the secretory proteins found on cell surfaces and in body fluids are glycoproteins. Thus more than 50% of all human proteins are expected to be glycosylated.[3] Glycoproteins play important roles in many biological events[4] such as cell–cell recognition, immune response and development. A general goal of PTM-related research is to understand how these modifications effectively regulate protein functions.

In order to reveal the functions of carbohydrates attached to proteins, biological experiments rely on glycoproteins prepared by recombinant expression. Due to their biosynthesis natural glycoproteins exhibit considerable heterogeneity in their oligosaccharide structures.[5] Unfortunately, the microheterogeneity of glycoproteins cannot be resolved routinely by chromatographic methods.[6] Thus the attempts to determine the biological role of a particular oligosaccharide in the protein of interest are frequently based on mixtures of glycoforms. In order to resolve these questions, the availability of homogeneous glycoproteins is essential. Until recently, glycoproteins could only be obtained by biological means[7] since chemical methods were not available for the preparation of these complex biomacromolecules.

This review describes the recent chemical approaches resulting in the preparation of homogeneous glycoproteins bearing complex asparagine linked (N-linked) oligosaccharides.

2. Biosynthesis of glycoproteins

The glycosylation of proteins is found in all classes of organisms and displays a variety of related carbohydrate structures.[2] All the enzymes involved in glycoprotein biosynthesis are potentially useful for recombinant production[7] as well as enzymatic synthesis of glycoproteins. In mammalian cells cytosolic glycoproteins bear only a short O-linked modification (O-GlcNAc) attached to Ser/Thr moieties.[8] The main source of glycoproteins is linked to the secretory pathway originating in the endoplasmic reticulum (ER). There a complex biosynthetic pathway occurs in a co- and/or posttranslational manner.

According to the mode of attachment the sugar part of secretory glycoproteins can be classified as an O-linked or N-linked type.[2] In the O-linked type, an *N*-acetyl-α-D-galactosamine (GalNAc) residue is connected to the OH of serine or threonine. The transfer of additional sugars in the Golgi-apparatus gives rise to eight basic core structures of moderate complexity. For the N-linked type, a unique 14-mer oligosaccharide is transferred *en bloc* to the amide nitrogen of an asparagine of the Asn-X-Ser/Thr recognition sequence. The 14-mer on the nascent glycoprotein is subjected to enzymatic degradation and remodelling in the ER and the Golgi apparatus. Subsequently, the final *N*-glycoprotein is translocated to the cell surface or secreted (Fig. 1).

The resulting *N*-glycans can be distinguished by their terminal carbohydrate composition into three types: complex, hybrid, and high-mannose structures (Fig. 2). The highest diversity is found in the complex structures where multiple

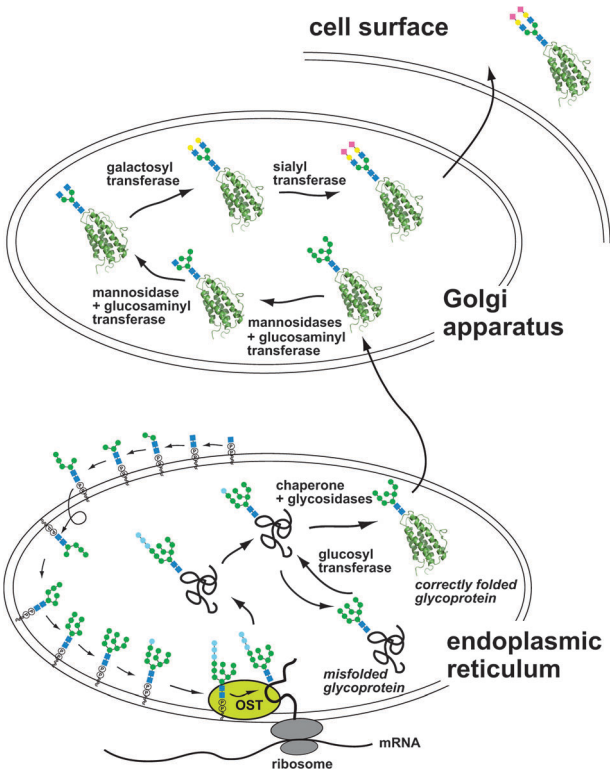

Fig. 1 Key steps in the biosynthesis of *N*-glycoproteins (OST = oligosaccharyltransferase).

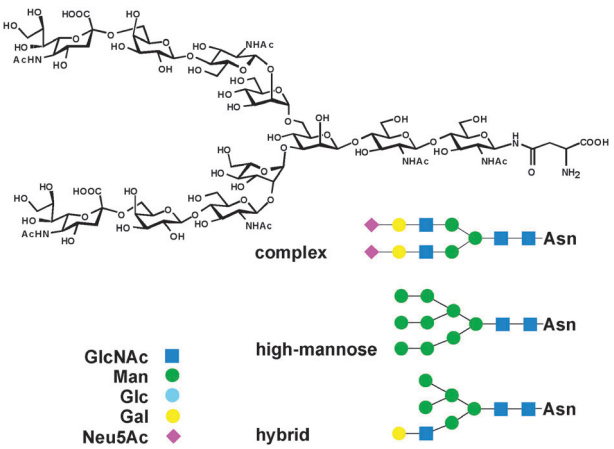

Fig. 2 Main types of *N*-glycans found on *N*-glycoproteins.

branching is accompanied by terminal capping with sialic acid containing motifs and distinct modifications of the core pentasaccharide. High-mannose *N*-glycans from higher eukaryotes are weakly processed structures bearing only terminal mannose residues derived from the nascent 14-mer. In hybrid *N*-glycans complex-type termini are present in the 1,3-branch combined with a high-mannose part in the 1,6-branch.

The biosynthesis of *N*-glycoproteins has been studied extensively. A dolichol phosphate embedded in the lipid bilayer of the endoplasmic reticulum (ER) is enzymatically elongated with GlcNAc, Man and Glc residues. The 14-mer oligosaccharide of

the $Glc_3Man_9GlcNAc_2$ (G3M9) diphospho-lipid can be transferred to the nascent peptide chain entering the ER following an intriguing mechanism.[9] The glycosylated polypeptide is further elongated and subsequent folding yields a native glycoprotein. This folding process is either spontaneous or can be supervised by several enzymes and chaperones, which rigorously discriminate misfolded and correctly folded glycoproteins. The carbohydrate-guided refolding machinery is referred to as the ER glycoprotein quality control system.[10]

Properly folded glycoproteins are transported into the Golgi apparatus, whereas misfolded glycoproteins are translocated to the cytoplasm for degradation. In the Golgi apparatus, several glycosidases and glycosyltransferases convert the initial high-mannose type oligosaccharides into complex-type oligosaccharides or hybrid-type oligosaccharides. During this process, a considerable variety of the final oligosaccharides is generated. The resulting mixture of glycoforms makes it difficult to elucidate to what extent individual *N*-glycans are involved in trafficking, secretion and bioactivity of each glycoform. Structure–activity relationships have been established *e.g.* for sialylated *N*-glycans on circulatory glycoproteins[11,12] as well as for core-fucosylated *N*-glycans on immunoglobulins (ADCC).[13] An immunosuppressive component of IgG was found to be dependent on a unique sialylated *N*-glycan.[14] *N*-Glycans can directly affect the stability of the glycoprotein by intramolecular interaction.[15]

3. General synthetic considerations for *N*-glycoproteins

Many approaches for the synthesis of glycoproteins and their analogues[16–23] have been established, however, only in a few cases homogeneous glycoproteins were obtained. Retrosynthetic analysis gives a multitude of possible disconnections, but not all of them have a synthetic equivalent (Fig. 3). A possibility for attaching *N*-glycans to non-glycosylated proteins would be most desirable. Hitherto no chemical or enzymatic options have been found to be applicable for this task. A convenient alternative is a disconnection within the chitobiose part of the *N*-glycan to a simple GlcNAc glycoprotein and a truncated *N*-glycan. The endoglycosidases catalyzing this cleavage reaction are involved in the breakdown of glycoproteins. It is known, however, that these endoglycosidases can also be used for synthesis,[24] since they catalyze the reverse reaction at preparatively useful rates, whenever the transferred sugar is activated as an oxazoline.[25]

The highest degree of flexibility is currently available by assembling the glycoprotein from shorter peptides. The most powerful method is native chemical ligation,[26] which takes advantage of the high reactivity of N-terminal cysteines in unprotected peptides with peptide thioesters (Fig. 4). Native chemical ligation has enabled the synthesis of many full-length proteins including modified proteins.[27] The required peptide fragments (thioesters or Cys-peptides) can be obtained by chemical or enzymatic synthesis as well as by recombinant expression.[28]

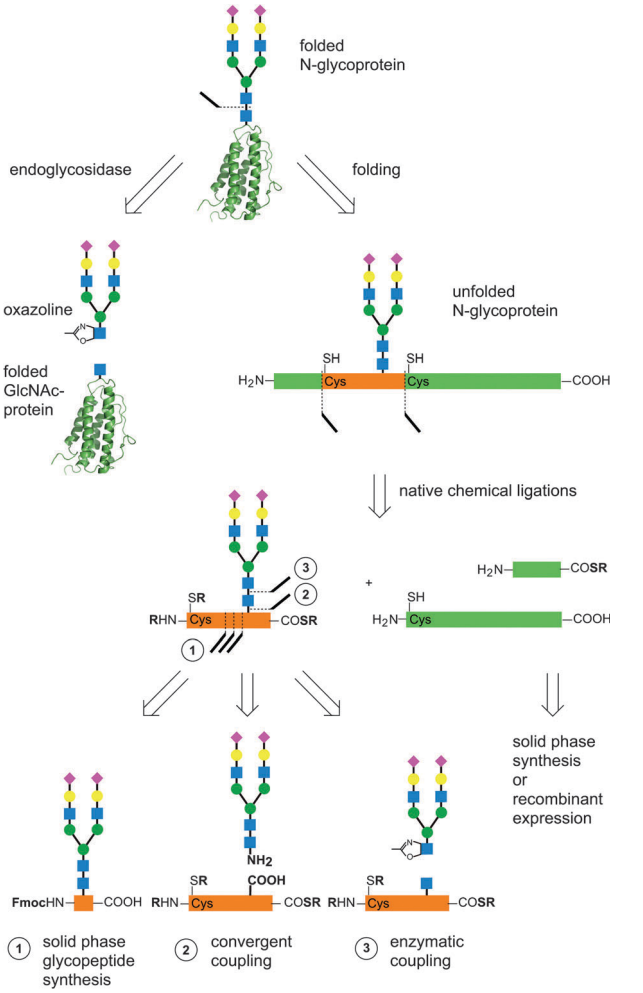

Fig. 3 Useful retrosynthetic pathways for the strategic disconnection of *N*-glycoproteins.

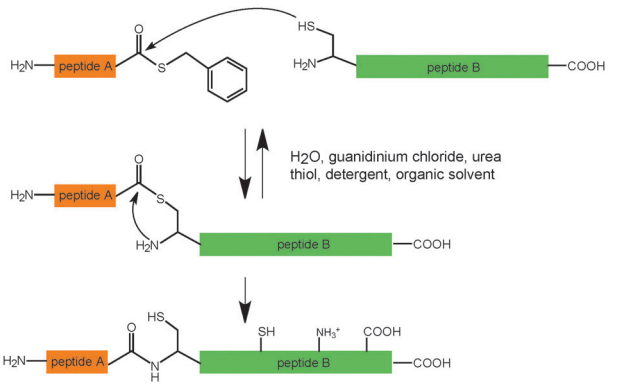

Fig. 4 Fragment coupling of unprotected peptides by native chemical ligation (NCL).

Based on this elaborated approach the incorporation of glycopeptides should lead to glycoproteins. Recombinant protein fragments for native chemical ligation are not limited by length restrictions, whereas peptides obtained by solid phase synthesis are limited to 40–50-mers for technical reasons.

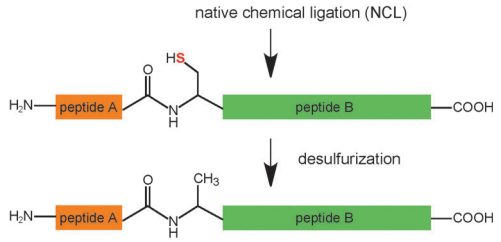

Fig. 5 Selective removal of thiol groups after native chemical ligation converts cysteine to alanine with high efficiency.

Usually synthetic glycopeptides bearing biorelevant *N*-glycans are quite difficult to obtain in a similar length. Therefore, multiple ligations may be necessary, depending on the position of the oligosaccharides and the length of the available fragments.

Generally, the low overall abundance of cysteine residues in proteins (1–2%) may require alternative or additional ligation sites. Two solutions of this problem have emerged: the use of temporary auxiliaries[29] and the selective desulfurization of thiol groups.[30–32] Whereas the cysteine-mimicking auxiliaries have shown limited applicability and may result in low efficiency ligations, the selective desulfurization of cysteine and a number of other thiol-containing amino acids has found wide acceptance among ligation chemists (Fig. 5).

Regardless of the availability of a fairly elaborated spectrum of useful ligation sites, the synthesis of the required glycopeptides[33,34] still remains challenging. This is mainly caused by the limited availability of *N*-glycans in larger amounts, followed by carbohydrate related side reactions during glycopeptide synthesis. Synthetic *N*-glycans of nearly any desired structure are available for specialized labs only and multistep syntheses curb the final amounts of these oligosaccharides. Sufficient amounts of biantennary *N*-glycan model compounds can be obtained by isolation of a sialoglycopeptide from egg yolk.[20] For the synthesis of *N*-glycopeptides two main strategies exist, namely the

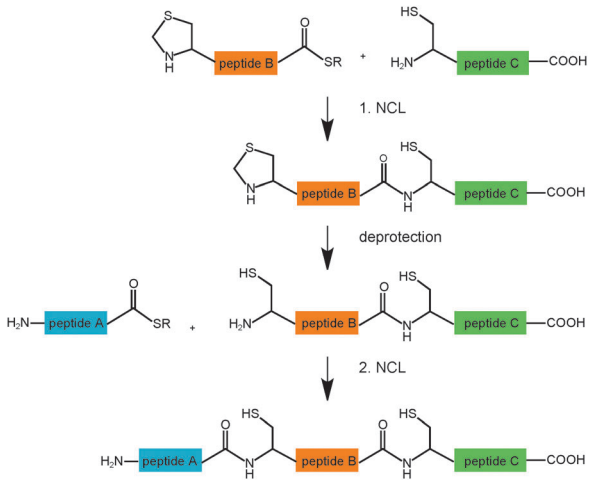

Fig. 6 Sequential native chemical ligations with N-terminal cysteine masked as a thiazolidine.

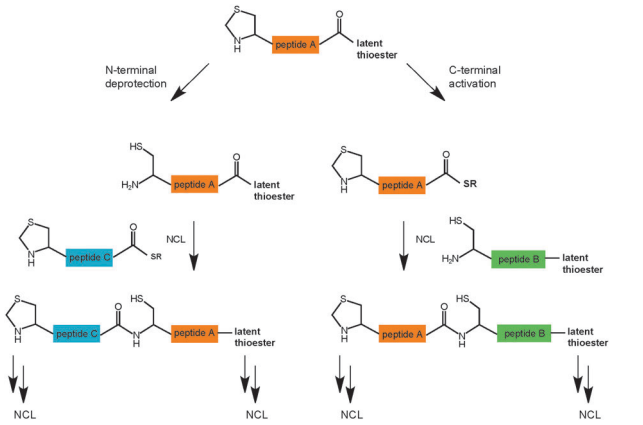

Fig. 7 Latent thioesters in combination with orthogonally protected N-terminal cysteines allow selective peptide elongations in N- and C-terminal directions.

cassette approach, wherein preformed glycosylamino acids are employed in stepwise elongations, or the convergent method,[35] where the peptide is synthesized first and subsequently the *N*-glycan is coupled to the peptide chain (Lansbury aspartylation).

Sequential ligations are mainly carried out using peptide thioesters, wherein a temporary protecting group masks an N-terminal cysteine residue.[36] After the first ligation the N-terminal cysteine is liberated and selectively ligated with another thioester fragment (Fig. 6). A less general alternative is provided by kinetically controlled ligations.[37] In this case a more reactive aromatic peptide thioester is ligated with a Cys-peptide bearing a less reactive aliphatic thioester, thereby allowing the extension of the peptide by ligation in the C-terminal direction. An improved modification of this concept takes advantage of latent thioesters[38] (thioester precursors), which permit a free choice of the direction of sequential ligations (Fig. 7).

4. Synthesis of glycoprotein fragments

The availability of protein fragments for the native chemical ligation of proteins has improved substantially in the nearly two decades since the seminal publication[26] in 1994. At first only chemical synthesis could provide peptides with a thioester functionality. However, the use of inteins has revealed a viable path for recombinant thioesters of virtually any length.[28,39] The chemical synthesis of peptide thioesters follows two principles: either the thioester is generated on the resin or a peptide acid is cleaved off the resin and subsequently converted to a thioester. Thioesters are accessible *via* Fmoc-[40] and Boc-strategies.[41] The presence of a thioester functionality is known to facilitate epimerization of the C-terminal amino acid even under weakly basic ligation conditions.[42] This stimulated the recent development of several latent thioester concepts,[38] where thioester precursors remain stable during peptide synthesis and global deprotection. Under special conditions, also including ligation conditions, the thioester can be generated from the stable precursors.

Peptide thioesters, which are not accessible by solid phase methods (above 50–100 amino acids), can be obtained routinely

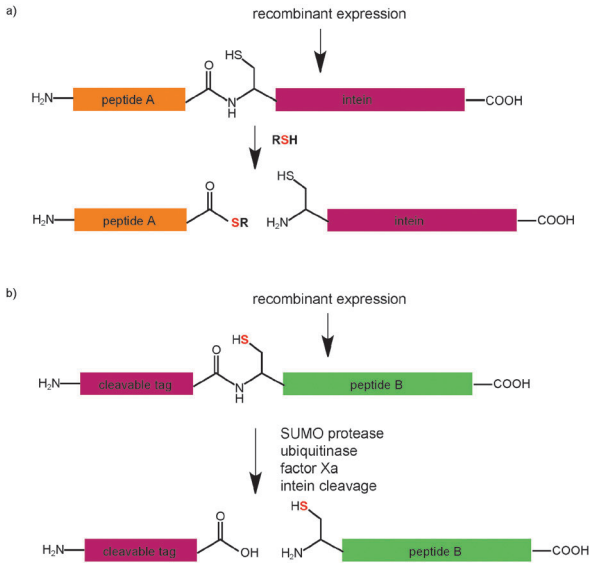

Fig. 8 Recombinant methods for the generation of (a) thioesters and (b) Cys-peptides.

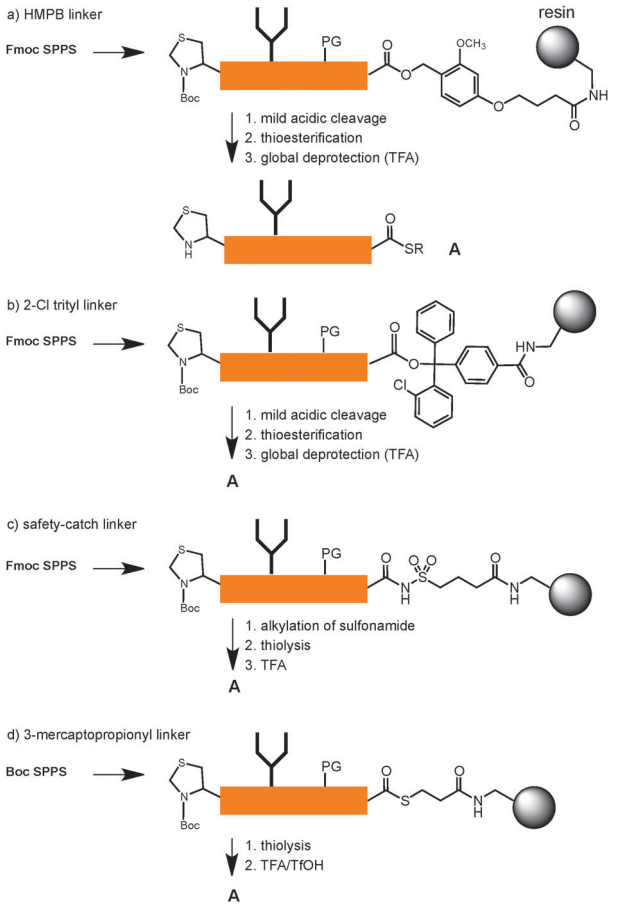

Fig. 9 Linker systems compatible with the synthesis of complex glycopeptide thioesters.

by recombinant expression of the peptide C-terminally fused to an intein (Fig. 8).[39] Inteins are self-cleaving proteases involving internal thioester intermediates. By intercepting these thioester intermediates with external thiols, the thioester is liberated from the intein and can be isolated. It is recommended to choose the intein according to the C-terminal amino acid, since most inteins show preferences for certain amino acids at the cleavage site. However, this method usually retains a methionine at the N-terminus of the peptide thioester.

In order to recombinantly generate peptides (including their thioesters) with an N-terminal cysteine or any other native N-terminal amino acid special approaches are needed. Established methods include the use of specific proteases *e.g.* SUMO-protease,[43] factor Xa,[44] TEV protease[45] or self-cleaving inteins[39] in some cases.

The linear chemical synthesis of glycopeptides and their thioesters is mainly performed on the solid phase. In this approach, an asparagine linked *N*-glycan is employed for peptide elongation. Due to the lability of oligosaccharides to strong acids this approach usually follows the Fmoc strategy. However, the acid lability of the *O*-glycosidic linkages of *N*-glycans can be overcome by esterification of free hydroxyl groups[46] and carboxylates,[47] enabling also Boc-based glycopeptide synthesis, which will be discussed later.

When the *N*-glycan-Asn building block is readily available the Fmoc-based cassette approach can be applied. After choosing an appropriate resin–linker combination the elongation is carried out. In our hands the use of more hydrophilic resins is crucial, when larger *N*-glycans of about 10 residues are to be incorporated into the peptide chain. Typically PEGA[48] or ChemMatrix resins are preferred over polystyrene-based supports. A multitude of robust linkers is compatible with glycopeptide synthesis leaving the choice to the chemist whether the thioester should be generated on the resin or after cleavage of the glycopeptide (Fig. 9).

Post-cleavage thioesterification of protected glycopeptide acids can be troublesome whenever the cleaved peptide is not well soluble or aggregates. Depending on the sequence the final glycopeptide may also be retained on the resin.[49] However, for shorter and well soluble glycopeptides the post-cleavage thioesterification appears to be more efficient.[50]

The elongation of the solid phase-bound peptide with the sugar–Asn conjugate is typically carried out using a low excess of the precious building block. When incorporating GlcNAc-Asn the sugar hydroxyls can be protected by acetylation. However, larger *N*-glycans with acetyl protection[51] were found to block extension of the peptide chain beyond 20 amino acids.[52] The example in Fig. 10 shows the use of temporary acetyl protection on a biantennary nonasaccharide *N*-glycan, which was compatible with the elongation to the 14mer glycopeptide thioester **2**. The acetates were removed on the resin by hydrazinolysis prior to establishing the thioester functionality.[52] As a consequence the hydroxyl groups of *N*-glycan-Asn building blocks (*e.g.* **1** and **3**) were kept unprotected after incorporation into the peptide, which improved the following peptide elongation. However, care must be taken in order to avoid esterification of the hydroxyl groups during chain extension and capping procedures. It was found that maintaining the concentration of activated amino acids below 40 mM may reduce esterification of

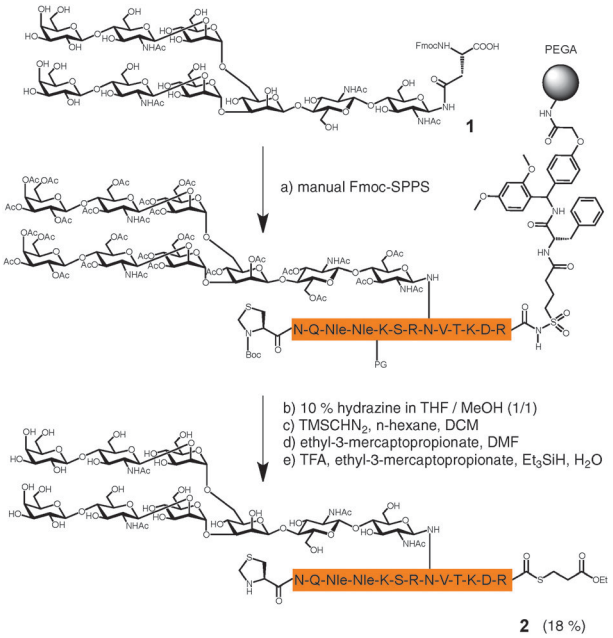

Fig. 10 Synthesis of the RNase 26–39 glycopeptide thioester **2** using temporary acetyl protection on the oligosaccharide.

the sugar to a very low level.[53] Alternatively, on-resin hydrazinolysis can remove undesired esterification, whenever the linker is not susceptible to these conditions.[52] Hydrazinolysis is not compatible with sialic acid moieties protected by ester groups and some types of linkers. Sialylated *N*-glycan-Asn derivatives can be incorporated into Fmoc-SPPS as a benzyl protected ester (**3**, Fig. 11).[47,54] The esters can be cleaved under basic conditions after completion of the peptide synthesis.

Boc-based glycopeptide synthesis can be carried out using *N*-glycan-Asn derivatives with protected sialic acid.[55] When stronger

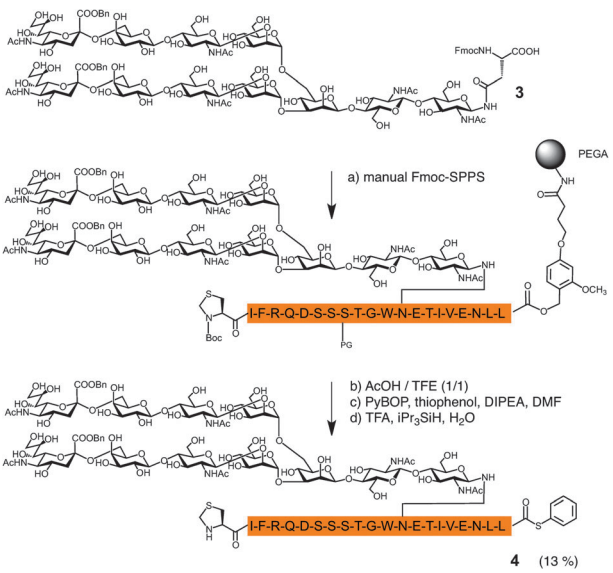

Fig. 11 Synthesis of IFN beta glycopeptide thioester **4** bearing benzyl protected sialic acids.

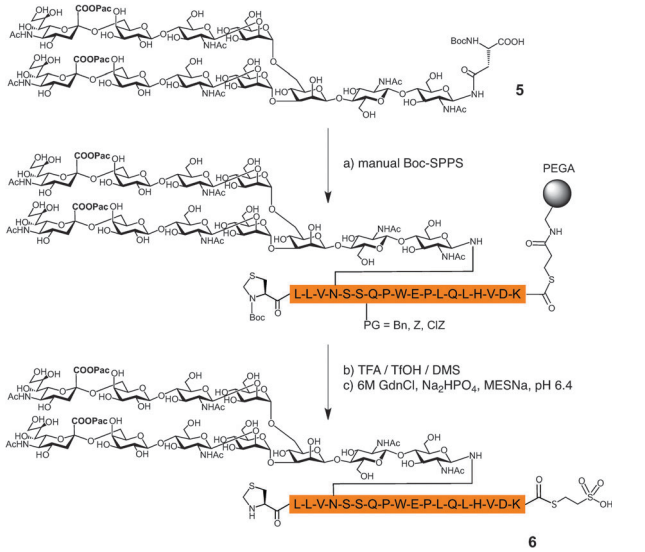

Fig. 12 Synthesis of a glycopeptide thioester by use of Boc SPPS with sialic acids protected by phenacyl esters (Pac).

acids are needed for side chain deprotection a more stable protecting group for the sialic acid carboxylate is required.

This requirement was met with phenacyl esters.[56] In Fig. 12 the synthesis of glycosylated EPO 79–97 thioester **6** is shown. Strong acid was applied to the resin-bound glycopeptide thioester followed by transthioesterification using thiol (MESNa) in 6 M GdnCl. In general the incorporation of sialic acid into glycopeptides adds more complexity and requires a case-by-case evaluation. A less demanding approach started with the synthesis of a glycopeptide bearing a synthetic *N*-glycan followed by enzymatic galactosylation and sialylation at a later stage.[57]

By following the stepwise approach (elongation by single amino acids) glycopeptides and the corresponding thioesters of about 20 amino acids bearing full length *N*-glycans were obtained. This barrier was overcome by convergent elongation of the peptide chain on the resin with longer peptide fragments. Thus side reactions on the sugar hydroxyls could be reduced and reverted by on-resin hydrazinolysis. However, C-terminal epimerization during fragment condensation poses a more serious problem. If C-terminal glycines or prolines are not available, a racemization-free glycopeptide elongation can be carried out using fragments with a C-terminal pseudoproline. This method led to the 1–39 RNase thioester **7** (Fig. 13).[58]

Compared to peptides the sequential elongation of glycopeptides is more demanding and can lead to unsatisfactory results in particular when difficult or longer sequences are targeted. Under these circumstances an alternative approach can be helpful, wherein the carbohydrate is attached to the peptide chain as the last step. This convergent approach is known as the Lansbury aspartylation[35] and can be carried out with peptides obtained by automatic peptide synthesis. The Lansbury aspartylation was developed as a solution phase method, but can also be carried out on the solid phase. Complex glycopeptide thioesters were synthesized by this

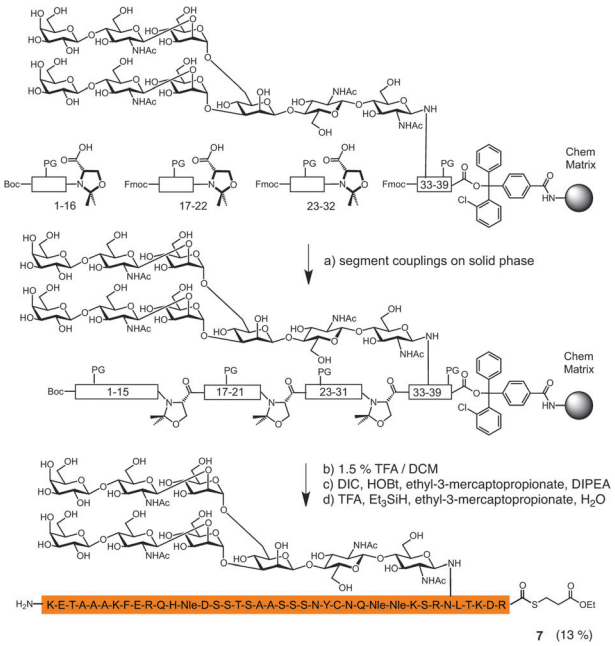

Fig. 13 Racemization-free elongation of glycopeptides by fragments containing C-terminal pseudoproline.

method mainly by the Danishefsky group.[19] Despite many advantages the aspartylation approach also suffers from a few drawbacks. The aspartate to be coupled with a glycosyl amine requires an additional orthogonal protecting group and is prone to aspartimide formation throughout the synthesis. Only recently this drawback could be reduced efficiently by use of pseudoprolines at the Ser/Thr residues of the consensus Asp-X-Ser/Thr sequence. Using this approach in solution[59] or on the solid phase[49] long glycopeptide thioesters of high complexity have become available (Fig. 14). This approach should further pave the way for the extension of the chemical synthesis of glycoproteins, as shown in a recent example of a synthetic EPO where all four glycosylation sites were occupied.[60]

At this point the availability of *N*-glycan-Asn derivatives as well as *N*-glycans needs to be addressed. For biantennary *N*-glycans a convenient source for a short sialoglycopeptide (**SGP**) from egg yolk[61] was exploited. Kajihara *et al.* showed that it is possible to degrade the disialylated undecasaccharide to numerous valuable Fmoc-Asn building blocks (*e.g.* **1**),[62] which allows us to obtain the desired compounds routinely in amounts of several hundred milligrams (Fig. 15).

Chemical synthesis of *N*-glycans is carried out at a high level in several specialized laboratories and mainly follows a convergent strategy based on derivatives of the core trisaccharide. The core trisaccharide bears a β-mannoside as the most challenging linkage. Selective elongation of this core scaffold allows the attachment of antennae as well as typical core modifications, *e.g.* core fucose or bisecting GlcNAc. Fig. 16 shows a synthetic dodecasaccharide (**15**) representing the highest density of branches in complex-type *N*-glycans.[63] By use of less branched building blocks virtually all the desired cores of complex-type *N*-glycans can be accessed.

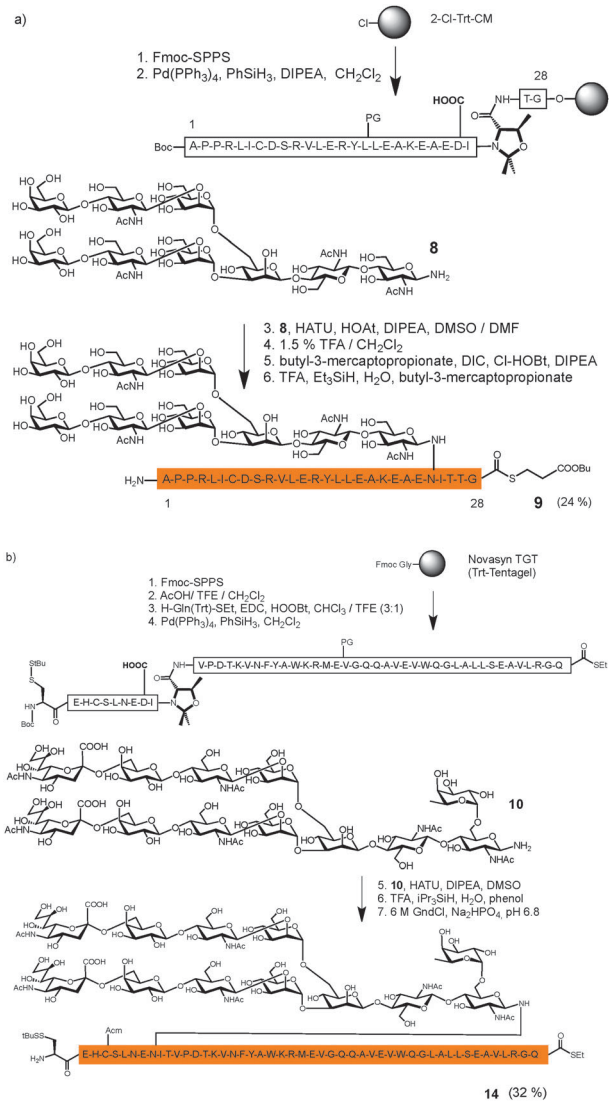

Fig. 14 Convergent synthesis of glycopeptide thioesters assisted by consensus Ser/Thr pseudoproline: (a) on the solid phase; (b) in solution.

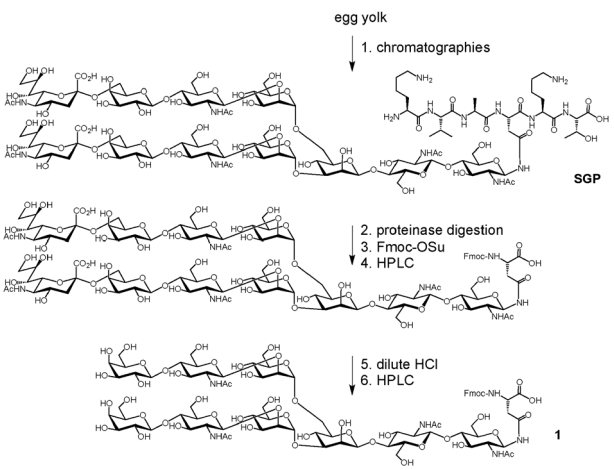

Fig. 15 Isolation of biantennary *N*-glycan-Asn building blocks from egg yolk.

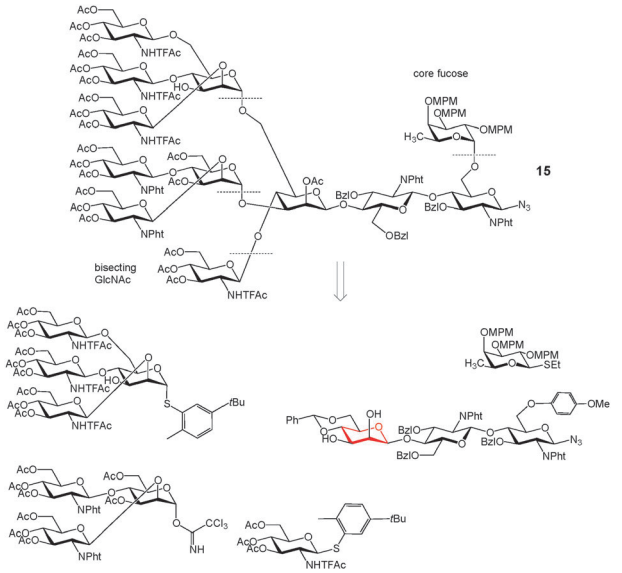

Fig. 16 Key building blocks for modular synthesis of multiantennary *N*-glycans bearing bisecting GlcNAc and core fucose.

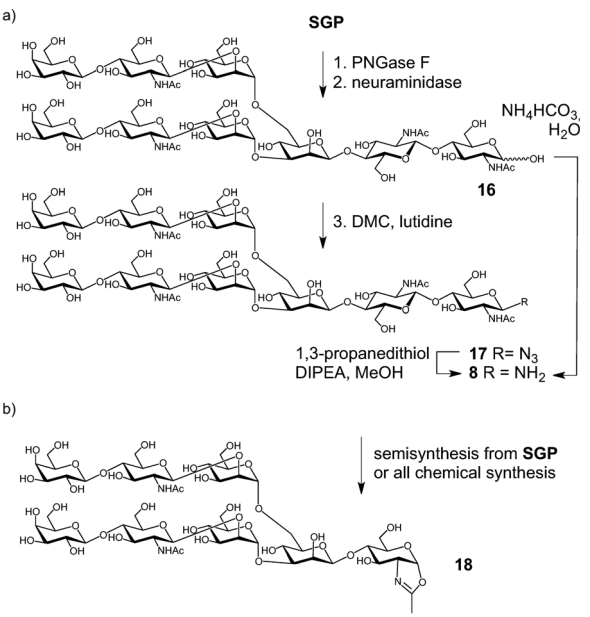

Fig. 17 (a) Semisynthetic approaches to *N*-glycan azide **17** and the corresponding amine **8**; (b) complex-type oxazoline **18**.

Even *N*-glycans with completely sialylated termini can be established chemically, however, enzymatic methods can provide a more rapid and easily variable access to biorelevant structures.[57] Depending on the synthetic strategy an anomeric azide may be incorporated early in the synthesis[63] or attached as a final step (Fig. 17a).[64] The glycosyl amine needed for coupling to the peptide can be obtained by reduction of an azide or *via* Kochetkov amination.[65] Care should be taken since a side reaction of the amination can cause loss of the reducing end GlcNAc moiety.[66] Oxazolines of truncated *N*-glycans required for transglycosylation were initially synthesized only chemically but recently convenient methods emerged for the semisynthetic conversion of truncated *N*-glycans to oxazolines even in aqueous solvents (Fig. 17b).[67]

Ligation methods

Native chemical ligation of peptides can be carried out in the presence of a multitude of buffers, solvents and additives ranging from plain water to 6 M guanidinium chloride or 8 M urea containing organic solvents, detergents, thiols and reducing agents (Fig. 4). The first step (transthioesterification) is reversible whereas the second step (S → N acyl transfer) traps the desired native amide product. The free side chains of the peptide usually do not interfere and thioester formation on interior thiols is reversible and non-productive. Once the fragments needed for glycoprotein assembly are available, their ligation can be investigated. It is recommended to test the efficiency of the ligation sites early on using small model peptides.[68] The typically employed alkyl thioesters are only moderately activated, but after addition of 50–200 mM MPAA (or other thiophenols) transthioesterification to a reactive phenolic thioester occurs.[69]

In case the glycoprotein contains cysteines at appropriate positions, the length of the fragments can be adjusted according to the synthetic availability and the minimal number of ligations. For proteins without cysteines at strategically useful positions (about 40 amino acids maximum for fragments) other ligation sites are required. One option is the use of temporary auxiliaries.[29] However, the use of thiolated amino acids followed by desulfurization bears more advantages and has recently become the method of choice for ligations without a native Cys. Desulfurization can be carried out efficiently and selectively by a radical process under homogeneous conditions[31] and was shown for a third of the natural amino acids (Ala, Val, Leu, Pro, Thr, Phe, Lys, Gln) (Fig. 18). Radical desulfurization is compatible with thioether functionalities (methionine, Acm protected cysteine and thiazolidine). A special ligation method

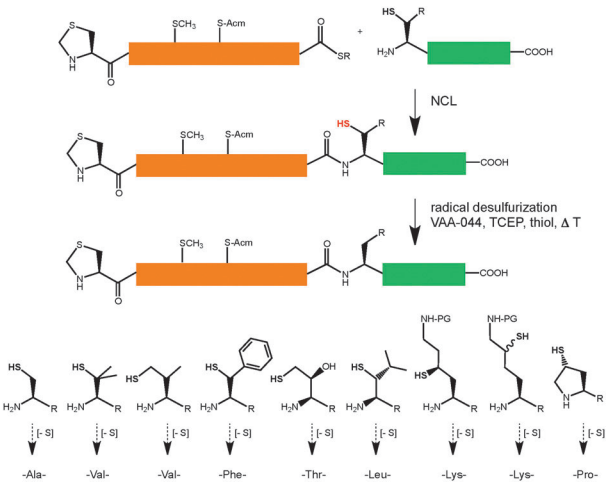

Fig. 18 Various thiolated amino acids are compatible with NCL and can be converted to sulfur-free amino acids within the ligation product.

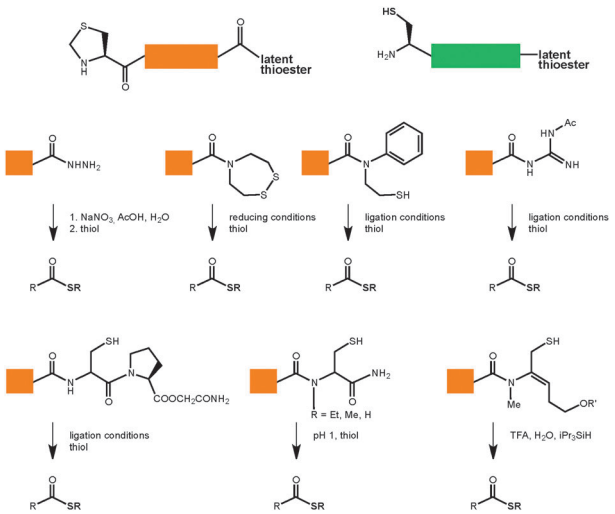

Fig. 19 Examples for latent thioesters and their activation conditions.

requiring no thiol groups is the direct aminolysis of peptide thioesters using partially protected peptides.[70]

Once the ligation sites are defined the ligations need to be arranged in a sequence depending on the features of the fragments. Sequential ligations (more than two fragments) are usually carried out from the C-terminus to the N-terminus. The cysteines of the inner fragments carry a temporary protection *e.g.* as thiazolidine,[36] which needs to be removed prior to the following ligation. This directional limitation can be overcome by applying kinetically controlled ligations.[37] Herein a more reactive (aromatic) thioester is C-terminally elongated by a less reactive thioester, which needs to remain stable under the ligation conditions.

A safer way to carry out multiple C-terminal extensions is to employ latent thioesters (Fig. 19). Several variants of latent thioesters were developed recently,[42,71–75] which led to a generally increased flexibility in the sequence of ligations. By means of latent thioesters the instability and base sensitivity of the thioester functionality can be circumvented elegantly. This robustness gives new options during assembly of the fragments and for establishing a desired order of the ligations. Whenever the size of the fragments varies considerably purification issues may arise, which can be resolved by alternative ligation sequences. Additionally, selective desulfurization steps can be integrated with greater ease.

5. Synthesis of *N*-glycoproteins

The following examples are organized in a timely sequence, which underlines the rapid incorporation of synthetic advances and the increasing complexity of the target glycoproteins. A key example reported by the Wang group, in which chemoenzymatic *N*-glycoprotein remodeling was carried out efficiently on RNase B, has led to an RNase B glycoform with a natural *N*-glycan structure. Isolated RNase B is a mixture of at least five natural glycoforms ($Man_{5-9}GlcNAc_2$), which can be degraded enzymatically to a single GlcNAc-RNase by use of Endo-A.

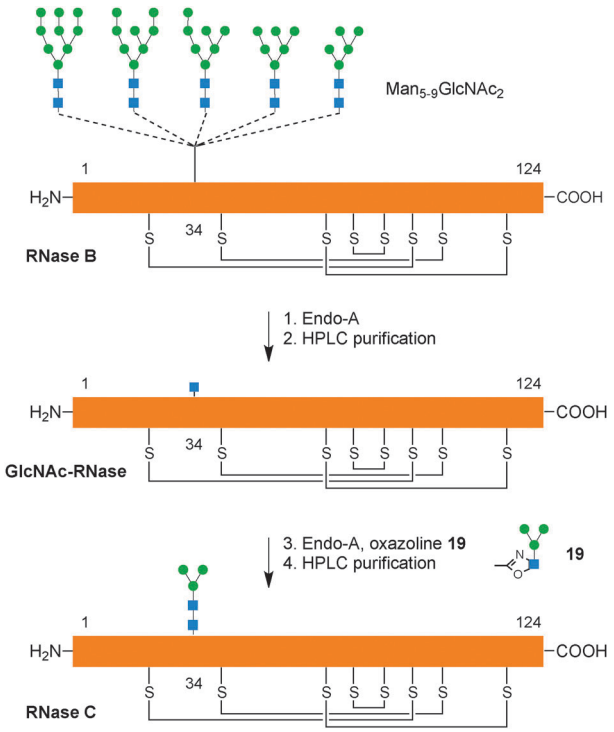

Fig. 20 Synthesis of a homogeneous glycoform of RNase B by enzymatic remodeling using the endoglycosidase Endo-A.

The truncated glycoprotein (GlcNAc-RNase) can be purified by HPLC and extended enzymatically, *e.g.* with the tetrasaccharide oxazoline **19**, to yield a homogeneous glycoform of RNase B bearing the core pentasaccharide (Fig. 20).[76] This approach was extended and further elaborated also by Fairbanks *et al.*[77] employing Endo-glycosynthases capable of transferring more complex oxazolines without hydrolyzing the final glycoprotein products.[22]

The first synthesis of homogenous glycoprotein bearing a fully synthetic peptide backbone and a semisynthetic sialylated *N*-glycan[55] was carried out in a joint project of the Kajihara and the Dawson group. The cytokine MCP-3 has 76 amino acids, two disulfides and an N-terminal N-glycosylation site, which allowed assembly by two sequential ligations using a short glycopeptide thioester and two longer peptide segments (Fig. 21). The sialylated glycopeptide thioester was accessible by Fmoc methodology, but additionally a novel Boc based approach with no protection on the side chains could be established. The minimal protection approach effectively circumvented the use of HF, which is not compatible with the oligosaccharide. Ligation of the three segments proceeded smoothly, and after oxidative refolding the benzyl protection of the terminal sialic acid residues was removed as the final step. The native state of the synthetic glycoprotein cytokine MCP-3 was confirmed by CD-spectroscopy, ELISA as well as disulfide mapping.

One year later the first semisynthesis of a glycoprotein with enzymatic activity was completed by the Unverzagt group.[52,78] RNase C is a glycoform of bovine RNase bearing a

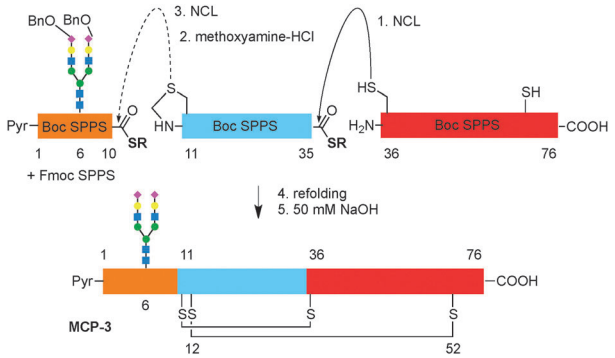

Fig. 21 Synthesis of sialylated MCP-3 by sequential NCL.

complex type *N*-glycan. The glycoprotein encompasses 124 amino acids and eight cysteines engaged in four disulfide bridges. A three segment strategy was chosen since the 1–39 glycopeptide thioester[58] was not accessible by conventional methods at the time. Another key fragment was the 40–124 peptide, which was obtained recombinantly using intein cleavage and *in situ* protection as mixed disulfides (Fig. 22). This type of protection facilitated the isolation as well as the following expressed protein ligation. The full length peptide was refolded from the second ligation mixture, purified and characterized by CD-spectroscopy and enzymatic activity.

Among many examples from the Danishefsky lab the syntheses of the alpha and beta subunit of the follicle stimulating hormone FSH bearing two sialylated and core fucosylated *N*-glycans are particularly noteworthy (Fig. 23).[79,80] The alpha subunit (92 aa) serves as a universal subunit for the four human glycoprotein hormones and displays two *N*-glycans and ten cysteine moieties. Fmoc synthesis of the required glycopeptide thioesters was unexpectedly demanding and led to an optimized ligation scheme. The convergent synthesis of the glycopeptide fragments was based on acid labile protection of the aspartate side chain and optional allyl type side chain protection for other acidic or basic side chains. After optimizing the scheme for chitobiose the valuable synthetic dodecasaccharide *N*-glycan (71 steps) was incorporated substoichometrically in good yields. Further difficulties arose in the deallylation step and a guanidinium adduct formation on a histidine. However, the three sequential ligations proceeded without affecting the two sensitive *N*-glycans. In a similar fashion the β-subunit

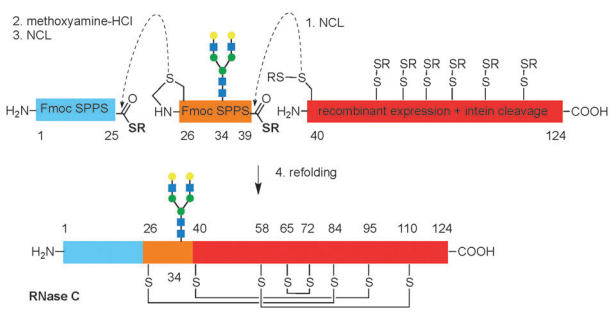

Fig. 22 Synthesis of RNase C by sequential NCL.

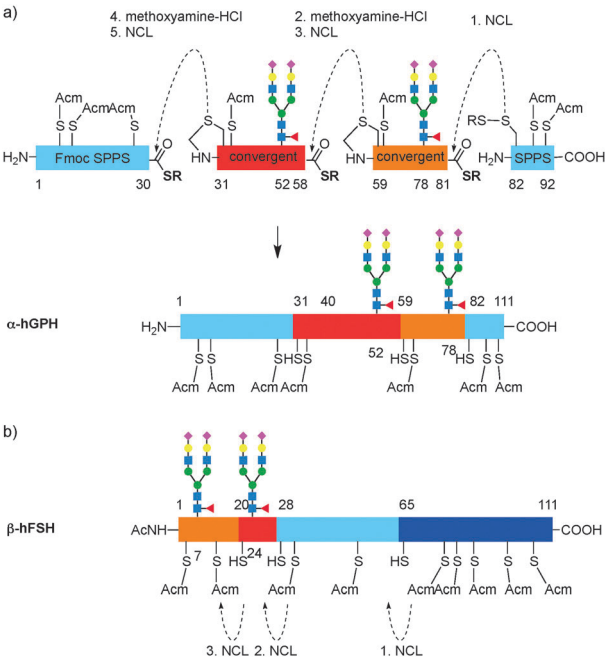

Fig. 23 Synthesis of the alpha and beta subunit of human FSH.

could be assembled successfully. Both subunits were kept with an Acm protection on most of the cysteines and should be amenable to refolding of the entire hormone after deprotection.

Synthetic glycopeptide chemistry also permits the generation of designed model glycoproteins not available in biological systems. In this example from the Kajihara lab an IL-8, which is not glycosylated in nature, was equipped with a $Man_9GlcNAc_2$ *N*-glycan (Fig. 24). This glycan can be glucosylated by enzymes of the glycoprotein quality control in the ER and was intended to serve as a probe for misfolded protein folds.[81] The oligomannosidic *N*-glycan-Asn building block was isolated from egg yolk as a side product and incorporated by Fmoc chemistry close to the N-terminus of the 39mer fragment. Ligation gave the full length IL-8, which was subjected to oxidative refolding conditions with and without disulfide shuffle. Refolding in the presence of a shuffle gave a high yield of glycoprotein with

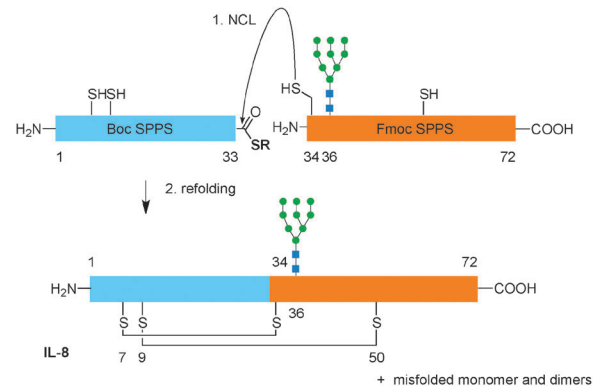

Fig. 24 Synthesis of an *N*-glycosylated IL-8 model glycoprotein used in folding studies.

the native fold, whereas in the absence of a shuffle three additional misfolded entities could be detected and separated by HPLC. The foldamers were characterized by disulfide mapping, CD spectroscopy, ANS binding and a functional assay using the ER glucosyltransferase UGGT.

The glycosylated interferon-β-1a is a powerful cytokine, which is approved as a drug against multiple sclerosis. IFN-β-1a comprises 166 amino acids, one N-glycosylation site in the central part and three cysteines at strategically less suitable positions. Thus a three segment sequential ligation strategy was devised based on artificially introduced cysteines for convenient ligation and subsequent desulfurization to the native alanine residues (Fig. 25).[54] The sialic acid moieties of the central glycopeptide thioester were protected as benzyl esters. Each ligation step was followed by a mild desulfurization and finally yielded the full length glycoprotein chain. After removal of the Acm groups and the benzylesters the glycoprotein was folded to the native state. The synthetic IFN-β-1a was characterized by disulfide mapping, CD spectroscopy and comparative *in vitro* and *in vivo* bioactivity studies revealing striking similarity to commercial products.

An example for glycoprotein synthesis, wherein radical desulfurization was employed extensively, is the synthesis of an erythropoietin (EPO) with a sialylated *N*-glycan at Asn 83 (Fig. 26).[56] Due to its unique biological activity EPO is one of the best studied therapeutic glycoproteins. The four cysteines of the 166 aa EPO are located close to the termini warranting alternative ligation sites. Thus the four native cysteines were kept protected by Acm throughout, and five alanines were selected as ligation sites for the six fragments. For the central glycopeptide thioester novel conditions based on Boc chemistry were established. In order to incorporate a sialylated *N*-glycan, the carboxylic acid of the sialoside must be esterified. Since the established benzyl ester was not stable during strong acid treatment, the more robust phenacyl ester was employed. This protection allowed the generation of the glycopeptide thioester using a thiol linker. After global side chain deprotection the Pac protected sialoglycopeptide thioester was released from the resin by thiolysis (Fig. 11). Four fragments were ligated sequentially from the C-terminus. Prior to the third ligation the Pac group (in conjunction with Trp-formyl) was conveniently cleaved by piperidine. The glycopeptide 50–166 was subjected

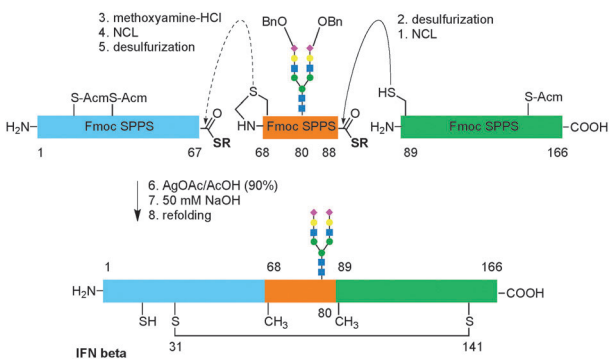

Fig. 25 Synthesis of sialylated human interferon beta (IFN beta).

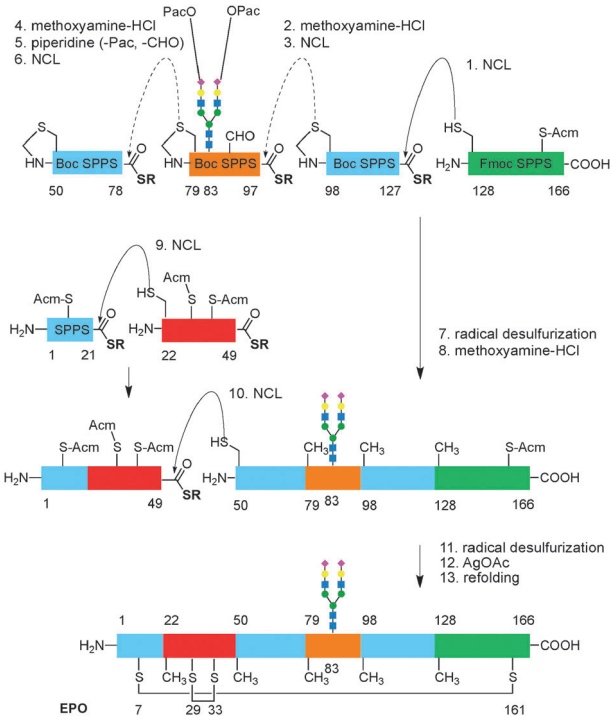

Fig. 26 Multi-segment ligations at non-native cysteines giving a mutant EPO with a sialylated *N*-glycan at Asn 83 after desulfurization.

to desulfurization of the three free thiol groups. A final ligation gave the full-length polypeptide, which was dethiolated prior to Acm deprotection of the four native cysteines. Oxidative refolding gave an EPO with three mutations and a sialylated *N*-glycan at Asn 83. CD spectroscopy suggests correct folding. The compound is designed to probe if *in vivo* bioactivity of EPO can be attained by a single *N*-glycan.

The synthesis of a member of very hydrophobic family of glycoprotein was reported by Hojo *et al.*[82] (Fig. 27). Despite a length of 80 amino acids only two of the six cysteines of saposin C are useful for ligations. Unfortunately the N-terminal 1–34 thioester was poorly soluble and led to serious difficulties in synthesis and ligation. Both difficulties could be improved by

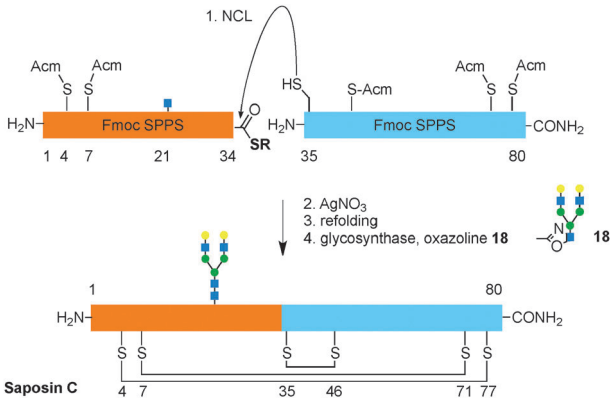

Fig. 27 Synthesis of saposin C using an *O*-acylisopeptide thioester and enzymatic transglycosylation by an Endo-M glycosynthase.

incorporating an *O*-acylisopeptide bond, which facilitated the synthesis of the GlcNAc containing segment 1–34. The thioester was generated from a latent *N*-ethyl-Cys precursor. Both segments were colyophilized prior to ligation, thus ensuring proper solubility in the denaturant. After removal of the Acm groups the GlcNAc saposin C was refolded in the presence of DMSO and purified. Enzymatic transfer of a complex-type *N*-glycan was accomplished using the fully synthetic octasaccharide oxazoline **18** and a glycosynthase derived from Endo-M. The reaction required DMSO as a cosolvent. The biological activity of the differently glycosylated saposin C forms was quite similar. However, only in the case of the glycoform with the biantennary nonasaccharide the solubility properties of the lipophilic glycoprotein were markedly improved.

6. Conclusions

The long-standing demand for homogeneous glycoproteins can finally be met by chemical synthesis. Significant advances in the synthesis of glycopeptides bearing full length *N*-glycans were fruitfully combined with peptide ligation methods. This has led to a rapidly increasing number of fully elaborated *N*-glycoproteins of impressive size and diversity. The implementation of advanced ligation methods will simplify the approaches further and allow taking on more complex targets including the generation of libraries of glycoforms. With this outlook the rapid transfer from basic glycobiology to industrial applications is starting to become reality.

Notes and references

1 S. van Kasteren, *Biochem. Soc. Trans.*, 2012, **40**, 929–944.
2 A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart and M. E. Etzler, *Essentials of Glycobiology*, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY, 2009.
3 R. Apweiler, H. Hermjakob and N. Sharon, *Biochim. Biophys. Acta*, 1999, **1473**, 4–8.
4 J. N. Arnold, M. R. Wormald, R. B. Sim, P. M. Rudd and R. A. Dwek, *Annu. Rev. Immunol.*, 2007, **25**, 21–50.
5 S. Thobhani, C. T. Yuen, M. J. Bailey and C. Jones, *Glycobiology*, 2009, **19**, 201–211.
6 P. M. Rudd, H. C. Joao, E. Coghill, P. Fiten, M. R. Saunders, G. Opdenakker and R. A. Dwek, *Biochemistry*, 1994, **33**, 17–22.
7 S. R. Hamilton, R. C. Davidson, N. Sethuraman, J. H. Nett, Y. Jiang, S. Rios, P. Bobrowicz, T. A. Stadheim, H. Li, B. K. Choi, D. Hopkins, H. Wischnewski, J. Roser, T. Mitchell, R. R. Strawbridge, J. Hoopes, S. Wildt and T. U. Gerngross, *Science*, 2006, **313**, 1441–1443.
8 L. Wells, K. Vosseller and G. W. Hart, *Science*, 2001, **291**, 2376–2378.
9 C. Lizak, S. Gerber, S. Numao, M. Aebi and K. P. Locher, *Nature*, 2011, **474**, 350–355.
10 A. Helenius and M. Aebi, *Annu. Rev. Biochem.*, 2004, **73**, 1019–1049.
11 R. L. Hudgin, W. E. Pricer, Jr., G. Ashwell, R. J. Stockert and A. G. Morell, *J. Biol. Chem.*, 1974, **249**, 5536–5543.
12 M. Takeuchi, N. Inoue, T. W. Strickland, M. Kubota, M. Wada, R. Shimizu, S. Hoshi, H. Kozutsumi, S. Takasaki and A. Kobata, *Proc. Natl. Acad. Sci. U. S. A.*, 1989, **86**, 7819–7822.
13 T. Shinkawa, K. Nakamura, N. Yamane, E. Shoji-Hosaka, Y. Kanda, M. Sakurada, K. Uchida, H. Anazawa, M. Satoh, M. Yamasaki, N. Hanai and K. Shitara, *J. Biol. Chem.*, 2003, **278**, 3466–3473.
14 R. M. Anthony, F. Nimmerjahn, D. J. Ashline, V. N. Reinhold, J. C. Paulson and J. V. Ravetch, *Science*, 2008, **320**, 373–376.
15 E. K. Culyba, J. L. Price, S. R. Hanson, A. Dhar, C. H. Wong, M. Gruebele, E. T. Powers and J. W. Kelly, *Science*, 2011, **331**, 571–575.
16 D. P. Gamblin, E. M. Scanlan and B. G. Davis, *Chem. Rev.*, 2009, **109**, 131–163.
17 R. J. Payne and C. H. Wong, *Chem. Commun.*, 2010, **46**, 21–43.
18 L. Liu, C. S. Bennett and C. H. Wong, *Chem. Commun.*, 2006, 21–33.
19 Y. Yuan, J. Chen, Q. Wan, R. M. Wilson and S. J. Danishefsky, *Biopolymers*, 2010, **94**, 373–384.
20 Y. Kajihara, N. Yamamoto, R. Okamoto, K. Hirano and T. Murase, *Chem. Rec.*, 2010, **10**, 80–100.
21 C. P. Hackenberger and D. Schwarzer, *Angew. Chem., Int. Ed.*, 2008, **47**, 10030–10074.
22 L. X. Wang and J. V. Lomino, *ACS Chem. Biol.*, 2012, **7**, 110–122.
23 H. Hojo and Y. Nakahara, *Biopolymers*, 2007, **88**, 308–324.
24 K. Takegawa, M. Tabuchi, S. Yamaguchi, A. Kondo, I. Kato and S. Iwahara, *J. Biol. Chem.*, 1995, **270**, 3094–3099.
25 M. Fujita, S. Shoda, K. Haneda, T. Inazu, K. Takegawa and K. Yamamoto, *Biochim. Biophys. Acta*, 2001, **1528**, 9–14.
26 P. E. Dawson, T. W. Muir, I. Clark-Lewis and S. B. Kent, *Science*, 1994, **266**, 776–779.
27 S. Kent, *J. Pept. Sci.*, 2003, **9**, 574–593.
28 T. W. Muir, *Annu. Rev. Biochem.*, 2003, **72**, 249–289.
29 J. Offer, *Biopolymers*, 2010, **94**, 530–541.
30 L. Z. Yan and P. E. Dawson, *J. Am. Chem. Soc.*, 2001, **123**, 526–533.
31 Q. Wan and S. J. Danishefsky, *Angew. Chem., Int. Ed.*, 2007, **46**, 9248–9252.
32 B. L. Pentelute and S. B. Kent, *Org. Lett.*, 2007, **9**, 687–690.
33 H. Herzner, T. Reipen, M. Schultz and H. Kunz, *Chem. Rev.*, 2000, **100**, 4495–4538.
34 T. Buskas, S. Ingale and G. J. Boons, *Glycobiology*, 2006, **16**, 113R–136R.
35 S. T. Cohen-Anisfeld and P. T. Lansbury, *J. Am. Chem. Soc.*, 1993, **115**, 10531.
36 D. Bang and S. B. Kent, *Angew. Chem., Int. Ed.*, 2004, **43**, 2534–2538.
37 D. Bang, B. L. Pentelute and S. B. Kent, *Angew. Chem., Int. Ed.*, 2006, **45**, 3985–3988.
38 D. MacMillan, A. Adams and B. Premdjee, *Isr. J. Chem.*, 2011, **51**, 885–899.

39 C. J. Noren, J. Wang and F. B. Perler, *Angew. Chem., Int. Ed.*, 2000, **39**, 450–466.
40 F. Mende and O. Seitz, *Angew. Chem., Int. Ed.*, 2011, **50**, 1232–1240.
41 H. Hojo and S. Aimoto, *Bull. Chem. Soc. Jpn.*, 1991, **64**, 111–117.
42 T. Kawakami and S. Aimoto, *Tetrahedron*, 2009, **65**, 3871–3877.
43 C. D. Lee, H. C. Sun, S. M. Hu, C. F. Chiu, A. Homhuan, S. M. Liang, C. H. Leng and T. F. Wang, *Protein Sci.*, 2008, **17**, 1241–1248.
44 D. A. Erlanson, M. Chytil and G. L. Verdine, *Chem. Biol.*, 1996, **3**, 981–991.
45 T. J. Tolbert and C. H. Wong, *Angew. Chem., Int. Ed.*, 2002, **41**, 2171–2174.
46 H. Kunz and C. Unverzagt, *Angew. Chem., Int. Ed. Engl.*, 1988, **27**, 1697–1699.
47 N. Yamamoto, Y. Ohmori, T. Sakakibara, K. Sasaki, L. R. Juneja and Y. Kajihara, *Angew. Chem., Int. Ed.*, 2003, **42**, 2537–2540.
48 E. Meinjohanns, M. Meldal, H. Paulsen, R. A. Dwek and K. Bock, *J. Chem. Soc., Perkin Trans. 1*, 1998, 549.
49 V. Ullmann, M. Rädisch, I. Boos, J. Freund, C. Pöhner, S. Schwarzinger and C. Unverzagt, *Angew. Chem., Int. Ed.*, 2012, **51**, 11566–11570.
50 Y. Kajihara, A. Yoshihara, K. Hirano and N. Yamamoto, *Carbohydr. Res.*, 2006, **341**, 1333–1340.
51 S. Mezzato, M. Schaffrath and C. Unverzagt, *Angew. Chem., Int. Ed.*, 2005, **44**, 1650–1654.
52 C. Piontek, D. Varón Silva, C. Heinlein, C. Pöhner, S. Mezzato, P. Ring, A. Martin, F. X. Schmid and C. Unverzagt, *Angew. Chem., Int. Ed.*, 2009, **48**, 1941–1945.
53 N. Yamamoto, A. Takayanagi, A. Yoshino, T. Sakakibara and Y. Kajihara, *Chem.–Eur. J.*, 2007, **13**, 613–625.
54 I. Sakamoto, K. Tezuka, K. Fukae, K. Ishii, K. Taduru, M. Maeda, M. Ouchi, K. Yoshida, Y. Nambu, J. Igarashi, N. Hayashi, T. Tsuji and Y. Kajihara, *J. Am. Chem. Soc.*, 2012, **134**, 5428–5431.
55 N. Yamamoto, Y. Tanabe, R. Okamoto, P. E. Dawson and Y. Kajihara, *J. Am. Chem. Soc.*, 2008, **130**, 501–510.
56 M. Murakami, R. Okamoto, M. Izumi and Y. Kajihara, *Angew. Chem., Int. Ed.*, 2012, **51**, 3567–3572.
57 S. Mezzato and C. Unverzagt, *Carbohydr. Res.*, 2010, **345**, 1306–1315.
58 C. Heinlein, D. Varon Silva, A. Tröster, J. Schmidt, A. Gross and C. Unverzagt, *Angew. Chem., Int. Ed.*, 2011, **50**, 6406–6410.
59 P. Wang, B. Aussedat, Y. Vohra and S. J. Danishefsky, *Angew. Chem., Int. Ed.*, 2012, **51**, 11571–11575.
60 P. Wang, S. Dong, J. A. Brailsford, K. Iyer, S. D. Townsend, Q. Zhang, R. C. Hendrickson, J. Shieh, M. A. Moore and S. J. Danishefsky, *Angew. Chem., Int. Ed.*, 2012, **51**, 11576–11584.
61 A. Seko, M. Koketsu, M. Nishizono, Y. Enoki, H. R. Ibrahim, L. R. Juneja, M. Kim and T. Yamamoto, *Biochim. Biophys. Acta*, 1997, **1335**, 23–32.
62 Y. Kajihara, Y. Suzuki, N. Yamamoto, K. Sasaki, T. Sakakibara and L. R. Juneja, *Chem.–Eur. J.*, 2004, **10**, 971–985.
63 S. Eller, R. Schuberth, G. Gundel, J. Seifert and C. Unverzagt, *Angew. Chem., Int. Ed.*, 2007, **46**, 4173–4175.
64 T. Tanaka, H. Nagai, M. Noguchi, A. Kobayashi and S. Shoda, *Chem. Commun.*, 2009, 3378–3379.
65 L. M. Likhosherstov, O. S. Novikova, V. A. Derevitskaya and N. K. Kochetkov, *Carbohydr. Res.*, 1986, **146**, C1–C5.
66 T. Murase and Y. Kajihara, *Carbohydr. Res.*, 2010, **345**, 1702–1707.
67 M. Noguchi, T. Tanaka, H. Gyakushi, A. Kobayashi and S. Shoda, *J. Org. Chem.*, 2009, **74**, 2210–2212.
68 T. M. Hackeng, J. H. Griffin and P. E. Dawson, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 10068–10073.
69 E. C. Johnson and S. B. Kent, *J. Am. Chem. Soc.*, 2006, **128**, 6640–6646.
70 R. J. Payne, S. Ficht, W. A. Greenberg and C. H. Wong, *Angew. Chem., Int. Ed.*, 2008, **47**, 4411–4415.
71 G. M. Fang, J. X. Wang and L. Liu, *Angew. Chem., Int. Ed.*, 2012, **51**, 10347–10350.
72 A. Otaka, K. Sato, H. Ding and A. Shigenaga, *Chem. Rec.*, 2012, **12**, 479–490.
73 N. Ollivier, J. Vicogne, A. Vallin, H. Drobecq, R. Desmet, O. El Mahdi, B. Leclercq, G. Goormachtigh, V. Fafeur and O. Melnyk, *Angew. Chem., Int. Ed.*, 2012, **51**, 209–213.
74 R. Okamoto, K. Morooka and Y. Kajihara, *Angew. Chem., Int. Ed.*, 2012, **51**, 191–196.
75 C. Ozawa, H. Katayama, H. Hojo and Y. Nakahara, *Org. Lett.*, 2008, **10**, 3531–3533.
76 B. Li, H. Song, S. Hauser and L.-X. Wang, *Org. Lett.*, 2006, **8**, 3081–3084.
77 C. D. Heidecke, Z. Ling, N. C. Bruce, J. W. Moir, T. B. Parsons and A. J. Fairbanks, *ChemBioChem*, 2008, **9**, 2045–2051.
78 C. Piontek, P. Ring, O. Harjes, C. Heinlein, S. Mezzato, N. Lombana, C. Pöhner, M. Püttner, D. Varón Silva, A. Martin, F. X. Schmid and C. Unverzagt, *Angew. Chem., Int. Ed.*, 2009, **48**, 1936–1940.
79 P. Nagorny, N. Sane, B. Fasching, B. Aussedat and S. J. Danishefsky, *Angew. Chem., Int. Ed.*, 2012, **51**, 975–979.
80 B. Aussedat, B. Fasching, E. Johnston, N. Sane, P. Nagorny and S. J. Danishefsky, *J. Am. Chem. Soc.*, 2012, **134**, 3532–3541.
81 M. Izumi, Y. Makimura, S. Dedola, A. Seko, A. Kanamori, M. Sakono, Y. Ito and Y. Kajihara, *J. Am. Chem. Soc.*, 2012, **134**, 7238–7241.
82 H. Hojo, H. Tanaka, M. Hagiwara, Y. Asahina, A. Ueki, H. Katayama, Y. Nakahara, A. Yoneshige, J. Matsuda and Y. Ito, *J. Org. Chem.*, 2012, **77**, 9437–9446.

