Total Synthesis of Glycosylated Proteins

Alberto Fernández-Tejada,

Laboratory for Bioorganic Chemistry, Molecular Pharmacology and Chemistry Program, 1275 York Avenue, New York, NY 10065, USA. Chemical and Physical Biology, CIB-CSIC, Ramiro de Maeztu 9, Madrid 28040, Spain

John Brailsford,

Laboratory for Bioorganic Chemistry, Molecular Pharmacology and Chemistry Program, 1275 York Avenue, New York, NY 10065, USA

Qiang Zhang,

Laboratory for Bioorganic Chemistry, Molecular Pharmacology and Chemistry Program, 1275 York Avenue, New York, NY 10065, USA

Jae-Hung Shieh,

Cell Biology Program, Sloan Kettering Institute for Cancer Research, 1275 York Avenue, New York, NY 10065, USA

Malcolm A.S. Moore, and

Cell Biology Program, Sloan Kettering Institute for Cancer Research, 1275 York Avenue, New York, NY 10065, USA

Samuel J. Danishefsky

Laboratory for Bioorganic Chemistry, Molecular Pharmacology and Chemistry Program, 1275 York Avenue, New York, NY 10065, USA

Alberto Fernández-Tejada: aftejada@cib.csic.es; Samuel J. Danishefsky: s-danishefsky@ski.mskcc.org

Abstract

Glycoproteins are an important class of naturally occurring biomolecules which play a pivotal role in many biological processes. They are biosynthesized as complex mixtures of glycoforms through post-translational protein glycosylation. This fact, together with the challenges associated with producing them in homogeneous form, has hampered detailed structure-function studies of glycoproteins as well as their full exploitation as potential therapeutic agents. By contrast, chemical synthesis offers the unique opportunity to gain access to homogeneous glycoprotein samples for rigorous biological evaluation. Herein, we review recent methods for the assembly of complex glycopeptides and glycoproteins and present several examples from our laboratory towards the total chemical synthesis of clinically relevant glycosylated proteins that have enabled synthetic access to full-length homogeneous glycoproteins.

Correspondence to: Alberto Fernández-Tejada, aftejada@cib.csic.es; Samuel J. Danishefsky, s-danishefsky@ski.mskcc.org.

1 Introduction

The study of protein post-translational modification (PTM) [1] is an active area in protein research. Protein glycosylation is a major form of PTM which expands the functional diversity and structural heterogeneity of proteins. More than 50% of human proteins are reported to be glycosylated [2], a modification known to play an important role in a wide range of biological events, at both cellular (cell–cell interaction, signaling, immune response) and protein levels (molecular recognition, protein folding) [3–5]. Aberrant glycosylation of proteins is also implicated in a number of illnesses, including autoimmune diseases, infectious diseases, and cancer [6, 7]. As a result, glycans and glycoproteins are increasingly being investigated in clinical settings as potential targets for the development of therapeutics, diagnostics, and vaccines [8, 9]. For all these reasons, a great deal of effort has been directed toward elucidating the effects of protein glycosylation on the function, conformation, stability, and bioactivity of glycoproteins [10–13].

Structurally, all native glycoproteins can be classified into two major types: *N*-linked and *O*-linked glycoproteins. In *N*-linked glycoproteins, the glycan (*N*-acetylglucosamine, GlcNAc) is attached through a β-glycosidic linkage to the amide side chain of an asparagine within an Asn-Xaa-Ser/Thr consensus sequence. In common *O*-glycoproteins, the sugar (mainly *N*-acetyl-galactosamine, GalNAc and GlcNAc) is α- or β-linked to the hydroxyl group of serine or threonine residues, respectively (Fig. 1). Other forms of *O*-glycosylation are also observed with several different glycans such as mannose, xylose, glucose, or fucose [14].

Unlike protein synthesis, protein glycosylation is not regulated genetically, and the biosynthesis of typical *N*- and *O*-glycoproteins often comprises multiple steps dictated by the relative activities of a large number of enzymes. As a result, glycoproteins are naturally expressed as complex mixtures of glycoforms possessing the same peptide backbone but differing in both the nature and site of glycosylation [15]. This structural diversity lays the molecular basis for a wide range of biological functions. However, the structural heterogeneity of glycosylation and the difficulty associated with obtaining homogeneous glycoproteins from biological sources has limited progress towards correlating glycoprotein structure with function [16]. The urgent need for chemically pure, synthetic glycoproteins for detailed structure–activity relationship studies and therapeutic applications has been the main driving force toward the development of several strategies for the preparation of these structures [17]. In particular, chemical synthesis has emerged as a practical way to gain access to well-defined glycoproteins, as is demonstrated through several relevant examples described herein.

2 Chemical Synthesis of Glycopeptides

The chemical synthesis of glycoproteins requires the ability to prepare glycosylated peptide fragments efficiently [18, 19]. Peptide synthesis efforts typically rely on solid-phase peptide

synthesis (SPPS) [20], a process which allows for the efficient assembly of peptide fragments comprising up to 50 amino acid residues. Of the two conventional SPPS methodologies (Boc method and Fmoc method), Fmoc-based SPPS is usually favored in the synthesis of glycopeptides because the glycosidic linkages are relatively labile in the presence of strong acid.

For the synthesis of a glycopeptide, bearing oligosaccharide side chains, there exist two general strategies: the cassette-based approach and the convergent method. In the more linear cassette approach, the glycan is attached to an amino acid residue and the resulting glycosyl amino acid is then used in SPPS to assemble the complete glycopeptide fragment (Scheme 1). In principle, the pre-formed glycosyl amino acid "cassette" can consist of either the fully elaborated oligosaccharide, or a monosaccharide residue, which, following SPPS, may be elongated further by enzymatic or chemical glycosylation to yield the glycopeptide containing the full glycan. Several research groups have used this stepwise, "cassette-based" approach to generate successfully *N*-linked [21, 22] and especially *O*-linked [23, 24] glycopeptide fragments related to natural glycoproteins. A particularly attractive application of this strategy is being implemented in a number of laboratories for the synthesis of mucin-type *O*-glycopeptides containing tumor carbohydrate antigens (Tn, TF, STn) as promising targets for anticancer vaccines [25–27].

In contrast, the convergent approach to glycopeptide synthesis relies on the direct conjugation of an oligosaccharide domain with the full-length peptide to generate the corresponding glycopeptide. Whereas the cassette strategy has been widely used for the synthesis of *O*-glycopeptides, the convergent method has been mainly applied for the preparation of *N*-glycopeptide fragments. In this approach, the partially protected peptide is first synthesized by SPPS. Next, the *N*-glycan, bearing an anomeric amine, is directly introduced through an amide linkage to the free aspartic acid (Asp) side chain of the peptide under Lansbury aspartylation conditions [28]. The Asp residue to be coupled requires an additional orthogonal protecting group which can be selectively removed prior to Lansbury aspartylation. The carboxylic acid is activated using coupling agents (i.e., HATU, HBTU, or PyAOP), and the anomeric amine of the sugar reacts with the activated ester to form the amide bond. When the peptidyl Asp side chain is activated, care should be taken to minimize competitive intramolecular aspartimide formation [29]. To address this problem, we and Unverzagt independently developed a simple solution, involving emplacement of a temporary pseudoproline motif, derived from Ser/Thr, at the $n + 2$ position (relative to Asp) of the Asn-Xaa-Ser/Thr consensus sequence (Scheme 2) [30, 31]. Adoption of this pseudoproline strategy has enabled the efficient and convergent syntheses of extended glycopeptide fragments, thereby providing enabling progress in the chemical synthesis of homogeneous complex *N*-glycopeptides [32].

3 Chemical Synthesis of Glycoproteins

3.1 Advances in Chemical Ligation

Extension of the methodologies described above to the assembly of full-length glycoproteins is hindered by a number of factors, including low coupling efficiency, poor solubility of protected polypeptides, acid-lability of the oligosaccharide chains, increased epimerization,

and formation of side-products resulting in low yields and purities of the final target molecules. For these reasons, several chemoselective ligation methods have been developed which permit the assembly of complex glycoproteins from shorter glycosylated peptides [33].

Most notable among these is the powerful native chemical ligation (NCL) method, first described by Kent et al. in 1994 [34]. Over the past 20 years, NCL has proved to be a remarkably robust methodology for the efficient, epimerization-free assembly of proteins. As shown in Scheme 3, the process accomplishes chemoselective ligation, in aqueous solution, of two fully deprotected peptides, one of which presents a C-terminal thioester and the other an N-terminal cysteine residue. The reaction involves transthioesterification of the C-terminal thioester with the N-terminal cysteine to generate a transient thioester intermediate, which then undergoes an irreversible $S \rightarrow N$ acyl transfer to form the native peptide bond. Interestingly, large peptide fragments may be obtained by recombinant expression and further used as ligation partners through a method termed expressed protein ligation (EPL) [35, 36].

To date, NCL-based strategies have been successfully applied to the chemical synthesis of over 200 proteins [37]. This technology has also been adopted for the total synthesis of a number of biologically relevant glycoproteins [38]. The first significant examples were reported by Bertozzi and coworkers, who synthesized several mucin-type glycoproteins bearing GalNAc residues at the relevant O-glycosylation sites [39, 40]. One such target was diptericin, an 82-residue antimicrobial glycoprotein prepared via NCL between two fragments, a 58-mer Cys-glycopeptide and a 24-mer glycopeptide thioester (Scheme 4). Notably, an alkane sulfonamide "safety-catch" linker was used for the Fmoc-based SPPS synthesis of the glycopeptide thioester [39]. Unverzagt et al. also used the "safety-catch" linker strategy followed by NCL to synthesize the first N-linked glycopeptide fragment carrying a complex-type oligosaccharide [41]. The first total chemical synthesis of a full-length glycoprotein bearing a complex type N-glycan was reported by Kajihara and coworkers, who prepared a single glycoform of the 76-amino acid chemokine monocyte chemotactic protein-3 (MCP-3) via two consecutive NCLs of three (glyco)peptide fragments, followed by protein folding (not shown) [42].

Despite these impressive examples, efforts to extend directly the NCL method toward the synthesis of complex glycopeptides have been complicated by the difficulties associated with synthesizing the requisite pre-formed glycopeptide thioester fragments, particularly by Fmoc-based SPPS [43]. To address this issue, our group has been engaged in the development of a number of synthetic strategies that may have broad utility for the preparation of homogeneously glycosylated peptides and proteins [44, 45]. In pursuing modified NCL strategies, we and others have sought to replace the labile C-terminal thioester with more stable latent thioester or alternative acyl donor groups.

In a key advance, we synthesized a C-terminal glycopeptide phenolic ester equipped with an o-disulfide moiety which, upon reductive treatment with subsequent unmasking of the free thiol, can rearrange to generate the corresponding aryl thioester via a reversible $O \rightarrow S$ acyl transfer [46]. The utility of this latent acyl donor in glycopeptide ligation has been widely

demonstrated. In the example shown in Scheme 5, two glycopeptide fragments – one bearing the newly crafted *o*-disulfide phenolic ester at the C-terminus and the other a *t*Bu-thio-protected cysteine at the N-terminus – were subjected to reductive cleavage by the addition of sodium 2-mercaptoethanesulfonate (MESNa). The resulting glycopeptide intermediates underwent ligation in situ to furnish the doubly glycosylated peptide.

Although this method proved highly useful for a wide range of substrates, we next sought to increase the acyl transfer capability of the *o*-thiophenolic ester in order to improve ligation efficiency with peptides bearing hindered C-terminal amino acids. Along these lines, we found that, when activated *p*-nitrophenyl or *p*-cyanophenyl oxo-ester variants were employed as thioester surrogates [47], efficient ligation was attainable at sterically demanding C-terminal sites. Indeed, this strategy was ultimately adopted for the synthesis of a challenging glycopeptide fragment en route to erythropoietin (EPO) [48].

Another important limitation for the general applicability of the NCL is the requirement that an N-terminal cysteine be present on the target peptide or glycopeptide fragment. Given the relative scarcity of cysteine residues in naturally occurring proteins and glycoproteins, it is often difficult to locate a feasible ligation site containing this amino acid, which makes direct ligation at cysteine itself of somewhat limited practical value. As a consequence, a number of alternative strategies have emerged for the cysteine-free, ligation-based assembly of glycopeptides and glycoproteins [49]. Early studies led to the development of thiol-based auxiliaries, wherein a mercapto group present in a provisional N-terminal moiety serves to mimic the function of the cysteine thiol. A pioneering example by Macmillan and Anderson demonstrated the utility of this ligation approach at an accessible glycine–glycine junction in the synthesis of an *O*-glycopeptide fragment bearing a single GalNAc residue [50]. We have reported several examples of the use of such auxiliary-based ligations for the construction of complex glycopeptides bearing multiple *O*- and *N*-glycans [51]. In a representative case, a glycopeptide acceptor containing a cleavable thiol-containing trimethoxybenzyl auxiliary on the N-terminus was ligated with a masked glycopeptide thioester at a glycine –glutamine junction to afford the desired glycopeptide bearing two *N*-linked glycans (Scheme 6).

As shown, under the reducing ligation conditions, $O \rightarrow S$ acyl transfer occurred at the *o*-disulfide phenolic ester as previously described (see above), to provide the thioester, which underwent transthioesterification with the thiol-containing glycopeptide (upon in situ reduction of the auxiliary disulfide bond); the transient intermediate underwent an $S \rightarrow N$ acyl transfer to generate the thermodynamically favored amide bond of the doubly glycosylated peptide adduct. The auxiliary was subsequently removed through sequential methylation of the free thiol to prevent the reverse acid-mediated $N \rightarrow S$ acyl shift, followed by TFA treatment.

While useful as a tool for peptide or glycopeptide ligation, the general application of these cysteine-free ligation strategies presents some drawbacks. First, steric hindrance at the ligation site renders the key $S \rightarrow N$ acyl transfer step difficult and, as a result, this method can only be effectively applied to junctions where at least one of the amino acids is glycine or alanine. Another potential problem lies in the harsh acidic conditions required for cleavage of the thiol-auxiliary, which could compromise the labile glycosidic linkages

within the glycopeptide. An alternative approach to the development of non-cysteine-based ligation strategies is to utilize a thiol-containing amino acid surrogate located at the N-terminus of the peptide fragment. In this strategy, the thiol moiety acts as a temporary "handle" to promote ligation with the corresponding peptide coupling partner. After the ligation reaction, the thiol functionality is removed through desulfurization to afford the natural amino acid at the site of ligation (Scheme 7).

This idea was first introduced by Yan and Dawson, who took advantage of cysteine-based NCL and then converted the N-terminal cysteine residue into alanine by reduction of the side chain with Raney nickel or Pd/Al$_2$O$_3$ [52]. In essence, this method provided a means to use cysteine as an alanine surrogate, which is a much more abundant amino acid in natural protein sequences. Subsequent work from Pentelute and Kent showed that unprotected cysteine residues could be desulfurized in the presence of methionine and Acm-protected cysteine residues with Raney nickel, although a large excess of the metal was required [53]. In 1998, González and Valencia had introduced a photochemical method for the efficient removal of the thiol group of cysteine without metal reagents, using triethyl phosphite as the nucleophile and triethylborane in acetonitrile as a radical initiator [54]. Driven by our interest in the total synthesis of complex glycoproteins, which often contain extensive and sensitive functional groups, we developed a mild, metal-free protocol for the selective reduction of cysteine to alanine compatible with a wide range of functionalities, including carbohydrate, thiazolidine, biotin, and thioester moieties [55]. This novel desulfurization strategy is aqueous-compatible and employs TCEP as a reducing agent, in conjunction with the water-soluble radical initiator VA-044 and *tert*-butylthiol as a thiol additive. This mild, radical-based protocol represented a significant improvement over previously reported metal-based NCL desulfurization methods and has been successfully applied in the synthesis of a number of complex proteins and glycoproteins [56–59]. Having identified a highly versatile cysteine reduction method, this ligation/ desulfurization approach opened the door to a new set of thiol-based ligation strategies with a range of different amino acid residues. Thus, through recourse to desulfurization methods, the reach of the NCL logic was extended, in our laboratory and others, to include alanine [52, 55], valine [60, 61], threonine [62], leucine [63, 64], lysine [65, 66], proline [67, 68], phenylalanine [69], methionine [70, 71], serine [72], arginine [73], aspartate [74], tryptophan [75], and glutamate [76] as viable ligation sites. Overall, these non-cysteine-based thiol ligation strategies, which are referred to as extended NCL reactions, followed by post-ligation desulfurization, have greatly facilitated the total chemical synthesis of complex, biologically relevant proteins and glycoproteins.

3.2 Total Synthesis of Homogeneous *N*-Glycosylated Proteins

Our laboratory has a long-standing interest in the total synthesis of naturally occurring, complex *N*-linked glycoproteins of important therapeutic value. This pursuit arises from the motivation to understand the structural and functional effects of protein glycosylation at the molecular level and, hopefully, to bring these complex biologics into the realm of structures that can be accessed by chemical means for structure-activity relationship studies. We have primarily focused our synthetic efforts on the human glycoprotein hormones (hGPH), granulocyte–macrophage colony-stimulating factor (GM-CSF) and EPO.

3.2.1 Human Glycoprotein Hormones—We are interested in the synthesis of a family of glycoproteins involved in a variety of important biological processes – the human glycoprotein hormones (hGPH). The members of this family are follicle stimulating hormone (FSH), luteinizing hormone (LH), chorionic gonadotropin (CG), and thyroid stimulating hormone (TSH). These glycoproteins are composed of two non-covalently associated subunits, a common α-subunit (α-hGPH) and a unique hormone-specific β-subunit (β-FSH, β-LH, β-CG, and β-TSH), each of which contains diverse carbohydrates at defined glycosylation sites. As the α-subunit is shared amongst these glycoproteins, we initially focused on the synthesis of the α-subunit (α-GPH) as the first step to access the human glycoprotein hormones.

Human α-Glycoprotein Hormone (α-hGPH): In our early studies, a simplified version of this domain incorporating chitobiose as a model glycan was synthesized as a proof of principle. The validated route was then applied to prepare the α-hGPH bearing the dibranched, sialylated and fucosylated *N*-linked dodecasaccharide. This complex-type *N*-glycan is a consensus sequence which has also been found to exist on the β-subunits of the human glycoproteins and incorporates the key features associated with these hormones. The challenging synthesis of this sugar, which involves approximately 70 chemical steps [77], utilizes glycal chemistry to build the core trisaccharide and highlights the use of the Sinaÿ radical glycosylation [78] for the simultaneous coupling of both biantennary side-chains of the dodecasaccharide. A seven-step deprotection procedure followed by Kochetkov amination [79] afforded the fully deprotected *N*-glycan bearing an anomeric amine (not shown).

With the synthetic dodecasaccharide amine in hand, we then built up the corresponding glycopeptide fragments using the Lansbury aspartylation reaction [28] with peptide thioesters α-hGPH[31–58] and α-hGPH[59–81], themselves prepared by Fmoc-based SPPS followed by single amino acid attachment under Sakakibara's epimerization-free conditions [80] (Scheme 8). After removal of all acid-labile protecting groups with a TFA-based cocktail, the fully deprotected (except Acm-containing cysteines) α-hGPH[82–92] and glycosylated α-hGPH [59–81] and α-hGPH[31–58] were then combined by means of a sequential two-step protocol involving ligation and subsequent cysteine unmasking. Lastly, the peptide thioester α-hGPH[1–30] (prepared by SPPS and direct thioesterification of the C-terminal glycine residue) was coupled with α-hGPH[31–92] via NCL to provide the Acm-protected primary sequence of α-hGPH featuring two defined, complex *N*-linked glycans [81].

Human β-Follicle Stimulating Hormone (β-hFSH): As our first complete target (with both α and β subunits synthesized), we selected hFSH, a member of the family of the gonadotropic hormones which play a key role in the regulation and maintenance of important reproductive processes. Naturally synthesized and secreted in the anterior pituitary gland, it has been used in clinical settings to treat anovulatory disorders associated with infertility [82]. Currently, hFSH is obtained from recombinant techniques as a complex mixture of glycoforms. This variability may well be the cause of the formidable liabilities and side effects associated with its use [83]. Faced with the unavailability of homogeneous

forms of hFSH, chemical synthesis seems to be the most straightforward approach to overcome this limitation.

The β-subunit of hFSH (β-FSH) was synthesized as described in Scheme 9 [84]. The individual peptide and glycopeptide segments were prepared by SPPS (with subsequent Sakakibara elongation or C-terminal glycine thioesterification to obtain the corresponding thioesters) and Lansbury aspartylation. Thus, fragment β-hFSH[66–111], peptide thioester β-hFSH[28–65], and glycosylated β-hFSH[20–27] were combined by NCL followed by cleavage of the N-terminal thiazolidine (Thz) blocking group in a stepwise manner. The last glycopeptide segment β-hFSH [1–19] was then joined under NCL conditions to form the glycosylated β-subunit of hFSH with two *N*-linked dodecasaccharides at the native glycosylation sites and Acm-protecting groups at the cysteine residues.

Human β-Luteinizing Hormone (β-hLH) and β-Chorionic Gonadotropin Hormone (β-hCG): Human luteinizing hormone (hLH) and chorionic gonadotropin hormone (hCG) belong to the family of the gonadotropins and are implicated in the stimulation of the gonadal and endocrine functions, as well as in pregnancy. hLH is released by the pituitary gland and hCG is produced in the human placenta. They both have similar physiological functions, and are used clinically as fertility drugs in reproductive medicine [85]. The β-subunit of hCG has been found to be overexpressed in several types of cancers and encompasses an epitope for hCG-based monoclonal antibodies. These features have made hCG a promising target for important biomedical and therapeutic applications [86, 87]. hLH and hCG share a high degree of similarity in their peptide sequence although hCG incorporates an extended C-terminal domain with four *O*-glycosylation sites. Thus, β-hLH contains 121 residues and only one *N*-glycosylation site (Asn30), whereas β-hCG comprises 145 amino acids and incorporates both *N*-linked (Asn13 and Asn30) and *O*-linked sugars (Ser121, Ser127, Ser132, and Ser138), making it the longest and most complex human glycoprotein hormone. Several studies investigating the importance of the oligosaccharide chains on the biological activity of these hormones suggest a more relevant role of the underlying carbohydrate residues vs the peripheral sugar units [88]. We sought to gain access to homogeneous glycoforms of β-hLH and β-hCG by chemical means for further exploration of the influence of the glycans on hormone action and function. Initially, we accomplished the synthesis of the simpler β-subunit of hLH bearing chitobiose as a model glycan to explore the feasibility of the synthetic route. Upon successful completion of the target molecule by sequential NCL, we then applied this strategy to assemble the more complex hCG β-subunit bearing not only two *N*-glycans but also four *O*-linked sugars [89].

The synthesis included two key features to access both terminal domains of the molecule. A successful double Lansbury aspartylation enabled the simultaneous, convergent installation of two chitobiose units onto the N-terminus of the peptide backbone to provide β-hCG[1–37]. The C-terminal fragment bearing closely spaced *O*-glycosylation sites, β-hCG[110–145], was prepared following a more linear "cassette" approach, whereby the four *O*-linked glycosylamino acids were sequentially incorporated in a single, solid-supported synthesis. The assembly of the Acm-protected, full-length β-hCG containing two *N*-linked (chitobiose) and four *O*-linked glycans (GalNAc) was performed in a modular and convergent fashion by ligation of the individual (glyco)peptide fragments (i.e., *O*-glycosylated β-hCG[110–145],

β-hCG[72–109], β-hCG[38–71], and *N*-glycosylated β-hCG[1–37]) from the C- to the N-terminus, using a multiple NCL-based strategy (Scheme 10). Importantly, this synthesis of the β-subunit of hCG represents the longest human glycoprotein hormone (hGPH) accessed as a single glycoform by strictly chemical means.

Having accomplished the total synthesis of both α- and β-subunits of these glycosylated proteins, we are now investigating the late-stage challenges towards preparing homogeneous, folded heterodimeric hFSH, hLH, and hCG in our laboratory.

3.2.2 Granulocyte–Macrophage Colony-Stimulating Factor (GM-CSF)—Our continuing interest in complex, therapeutic-level glycoproteins led us to pursue the synthesis of homogeneous glycoforms of GM-CSF with the aim of studying the effect of glycosylation on the biological activity of this target. GM-CSF is a glycoprotein cytokine which promotes cell growth and proliferation and functions as an immune modulator. It is used to stimulate the immune system after bone-marrow transplant and chemotherapy and is being investigated as a vaccine immunoadjuvant [90]. The GM-CSF structure comprises 127 amino acids with both *N*- and *O*-glycosylation sites. Our aim was to establish a synthetic strategy to enable access to a number of homogeneous glycosylated variants of GM-CSF. We designed a route with maximal convergence whereby the full-length molecule would be assembled by merging three (glyco)peptide fragments using exclusively alanine ligations.

In the synthesis of a bis-*N*-glycosylated version containing the chitobiose disaccharide, glycopeptide GM-CSF[34–80] and peptide GM-CSF[81–127] were first connected via NCL to give, following Thz cleavage, GM-CSF[34–127] (Scheme 11) [91]. This fragment was then combined with GM-CSF[1–33], containing a chitobiose moiety, to provide complex glycopeptide intermediate GM-CSF[1–127]. The late-stage transformations towards the doubly-glycosylated GM-CSF congener involved reduction of the two free thiol groups via metal-free desulfurization (MFD) to reveal the native alanine residues, followed by deprotection of the four Acm-bearing cysteine residues. A monoglycosylated variant bearing a single chitobiose (at Asn37) was prepared via a similar synthetic route. Final folding of both fully synthetic analogues under glutathione redox buffering conditions provided the structurally defined GM-CSF glycoforms bearing either one (Asn37) or two (Asn27 and Asn37) chitobiose units at asparagine wild-type sites. Evaluation of the in vitro and in vivo biological activity of these synthetic GM-CSF samples, including a non-glycosylated congener, showed that they all function in a similar manner to commercially available GM-CSF, which is obtained recombinantly as a mixture of glycoforms.

3.2.3 Erythropoietin (EPO)—Erythropoietin (EPO) is a glycoprotein hormone primarily implicated in the regulation of red blood cell production in the bone marrow, a process known as erythropoiesis [92–94]. Recombinantly derived EPO has been used for the treatment of anemia associated with renal failure and cancer chemotherapy [95]. However, production of recombinant EPO results in a complex mixture of glycoforms, where the primary protein sequence and the glycosylation sites are highly conserved but the carbohydrate domains present significant heterogeneity. Despite studies suggesting that specific sugar modifications affect the stability and erythropoietic activity of EPO [96, 97], the unavailability of single glycoforms has complicated elucidation of the relationship

between EPO glycosylation and its biological activity. Chemical synthesis uniquely holds the potential to provide access to homogeneous EPO for systematic evaluation of the effect of the glycan structure on biological activity.

EPO is composed of a 166-amino acid backbone incorporating three *N*-linked glycans (Asn24, Asn38, and Asn83) and an *O*-linked glycan (Ser126). In our first total synthesis of homogeneously glycosylated EPO, we chose to include three simpler chitobiose units and the naturally-abundant glycophorin tetrasaccharide at the four native glycosylation sites [98]. From a retrosynthetic perspective, we envisioned gaining access to the full-length glycoprotein, EPO[1–166], by assembling four glycopeptide domains through iterative NCL-based ligations (Scheme 12).

The *O*-linked glycophorin was introduced into the peptide chain as a glycosylated serine cassette [99] via addition of a single amino acid at each end of the glycophorin-linked Ser residue followed by NCL of the resulting glycosylated tripeptide with the adjacent peptide segment to provide EPO[125–166] (Scheme 13).

The chitobiose-containing *N*-glycopeptide constructs were prepared by our one-flask Lansbury aspartylation/deprotection protocol [30] from the corresponding peptide thioester fragments, themselves obtained by Fmoc-based SPPS followed by single amino acid attachment. On the basis of our previous experience toward a non-glycosylated version of EPO [59], we envisioned a synthetic strategy for merging the four component glycopeptides via a series of iterative alanine ligations, followed by late-stage desulfurization to convert the non-native cysteines into the requisite alanine residues. This first-generation synthesis featured a kinetically controlled NCL (KCL)-based approach [100] to enable the one-flask coupling of three fully elaborated EPO fragments in a highly convergent fashion (Scheme 14). In this KCL route, the aryl thioester of EPO[1–29] undergoes transthioesterification with the N-terminal cysteine of EPO[30–78] at a faster rate than the alkyl thioester, resulting in selective ligation between Cys-29 and Cys-30. Thus, EPO[1–29] and EPO[30–78] were first merged by NCL; upon completion of this ligation, EPO[79–166], itself prepared by NCL between EPO[79–124] and EPO[125–166], was added to assemble the protein sequence. The synthesis was completed by MFD, followed by removal of the Acm groups to afford the primary structure of homogeneously glycosylated EPO.

An alternative, more linear strategy was also developed, whereby the full protein chain was assembled via NCL as the final step of the synthetic sequence. The main advantage of this route is that it avoids the need for post-ligation transformations on the full-length sequence prior to protein folding. The approach employs peptides which slightly differ from the fragments used in the convergent route in that the ligation site at the N-terminus is shifted by one residue. Thus, glycopeptide fragments EPO[125–166], EPO[79–124], and EPO[29–78] were combined by successive NCLs to afford the EPO[29–166] domain (Scheme 15). At this point, desulfurization of the non-native cysteines that participated in the ligations, and subsequent Acm deprotection on the remaining cysteine residues set the stage for the merger of the final peptide fragment, EPO[1–28], by NCL. The fully synthetic, homogeneously glycosylated EPO[1–166] thus obtained was then subjected to disulfide oxidation and

protein folding to provide a simplified variant of folded EPO which, despite having truncated *N*-linked glycans, demonstrated erythropoietic activity.

Notwithstanding the major breakthrough of generating a fully synthetic version of EPO as a single glycoform, our ultimate goal was to achieve the total synthesis of "wild-type" EPO with consensus carbohydrate structures of realistic, biological-level complexity at all four native glycosylation sites. Therefore, we set out on this effort with the objective of incorporating the elaborate biantennary dodecasaccharide at the three *N*-glycosylation sites and glycophorin at Ser126 [101]. We initially envisioned adapting our previously described linear strategy, developed in the context of the chitobiose glycoform, to this more complex target. However, the greater steric bulk of the full *N*-glycans rendered ligation of EPO[29–78] and EPO[79–166] fruitless. Accordingly, the synthetic route was modified and a different ligation site, more distant from the problematic, large dodecasaccharide, was selected. In the newly designed approach, the domain related to pre-EPO[29–124] was reconfigured into three novel fragments, *N*-glycosylated EPO[29–59] and EPO[60–97], and peptide segment EPO[98–124], which were sequentially assembled, starting from the *O*-glycosylated, C-terminal domain EPO[125–166] via a series of smooth alanine ligations followed by thiazolidine unmasking (Scheme 16). The resulting fragment EPO[29–166] was then subjected to fourfold MFD to reveal the required alanines. After cleavage of the Acm protecting groups, cysteine-based NCL between EPO[29–166] and EPO[1–28] afforded the complete primary sequence of EPO containing wild-type complex glycan structures at each of the four glycosylation sites. This unfolded material gave rise to a mass spectrum of marginal quality (see Supporting Information in Wang et al. [101]). Not surprisingly, the material assigned as the EPO primary sequence did not even exhibit in vitro activity.

Upon exposure of the EPO primary sequence to folding conditions, a product was obtained which was reported [101] to exhibit strong erythropoietic activity and, thus, presumed to correspond to a biologically active version of compound 1 (Scheme 12) (although our synthetic EPO showed suitable activity in an in vivo assay, literature precedent suggests that heterogeneous mixtures of tetra-antennary structures have approximately 7× the in vivo activity of analogous bi-antennary EPO glycoforms [102–104]). More recently, an approximately 1-year-old sample of this material, bearing identical mobility properties to those reported from the original sample (see Wang et al. [101]), was found to retain approximately 15% of the in vitro activity of Procrit. It has been found that thawing and re-freezing of the sample tend to erode its erythropoietic activity. In retrospect, however, the inability to obtain a supportive mass spectrum of this material raises concerns with respect to the nature of the folded product. Related sialic acid-containing recombinantly derived EPO mixtures of unspecified structure gave rise to useful mass spectra in the context of glycoform profiling studies ([105, 106] and references therein). The corresponding author of reference [101] assumes full responsibility for not taking cognizance of these precedents. Further clarification of the matter of folding of the presumed precursor would require access to freshly prepared material, which can only be accomplished by repetition of the total synthesis. This has recently been undertaken in order to settle these questions.

4 Conclusion

The total synthesis of glycosylated proteins has long posed a daunting challenge to practitioners of organic chemistry. Recent key developments in this field, including NCL, non-cysteine-based ligations, MFD, and novel glycosylation methodologies, have led to the realization of the long-standing goal of gaining synthetic access to homogeneous peptides, proteins, and glycoproteins. Several truncated-carbohydrate versions of important biologically active glycoproteins have been synthesized in homogeneous form by a combination of these methods. These include GM-CSF, TSH, and EPO. An application to EPO bearing sialic acid-containing biantennary 12-mer *N*-glycosides has also been reported. The nature of the previously described biologically active version of this compound warrants further investigation because we did not obtain a useful mass spectrum to support its structure.

In parallel to efforts to achieve the total synthesis of glycoproteins by strictly chemical means, a number of powerful chemical and biological technologies have emerged for making tailored, full-size glycoproteins bearing glycan-defined domains for the study of specific biological questions. These highly enabling strategies, which have been extensively reviewed elsewhere, include the chemoselective, site-specific glycosylation of recombinant proteins via the so-called "tag and modify" approach [107], in vitro chemoenzymatic glycosylation remodeling of heterogeneous glycoproteins [108], direct enzymatic glycosylation of proteins [109], specific glyco engineering of host biosynthetic pathways [108], and in vivo suppressor tRNA technology [110, 111].

All these approaches complement each other and it is envisaged that, either alone or in combination, they will continue to enable the generation of glycosylated proteins with predesigned modifications and increasing complexity for the elucidation of the molecular basis of protein glycosylation. This realization will certainly contribute to the development of novel and improved glycoprotein-derived therapeutics in the next few years.

Acknowledgments

References

1. Walsh G, Jefferis R. Nat Biotechnol. 2006; 24:1241. [PubMed: 17033665]

2. Apweiler R, Hermjakob J, Sharon N. BBA-Gen Subj. 1999; 1473:4.

3. Dwek RA. Chem Rev. 1996; 96:683. [PubMed: 11848770]

4. Varki A. Glycobiology. 1993; 3:97. [PubMed: 8490246]

5. Ohtsubo K, Marth JD. Cell. 2006; 126:855. [PubMed: 16959566]

6. Miller LH, Good MF, Milon G. Science. 1994; 264:1878. [PubMed: 8009217]

7. Dube DH, Bertozzi CR. Nat Rev Drug Discov. 2005; 4:477. [PubMed: 15931257]

8. Hudak JE, Bertozzi CR. Chem Biol. 2014; 21:16. [PubMed: 24269151]

9. Dalziel M, Crispin M, Scanlan CN, Zitzmann N, Dwek RA. Science. 2014; 343:1235681. [PubMed: 24385630]

10. Bosques CJ, Tschampel SM, Woods RJ, Imperiali B. J Am Chem Soc. 2004; 126:8421. [PubMed: 15237998]

11. Imperiali B, O'Connor SE. Curr Opin Chem Biol. 1999; 3:643. [PubMed: 10600722]

12. Wormald MR, Petrescu AJ, Pao YL, Glithero A, Elliott T, Dwek RA. Chem Rev. 2002; 102:371. [PubMed: 11841247]

13. Li H, d'Anjou M. Curr Opin Biotechnol. 2009; 20:678. [PubMed: 19892545]

14. Spiro RG. Glycobiology. 2002; 12:43R.

15. Rudd PM, Dwek RA. Crit Rev Biochem Mol Biol. 1997; 32:1. [PubMed: 9063619]

16. Rich JR, Wither SG. Nat Chem Biol. 2009; 5:206. [PubMed: 19295526]

17. Bertozzi CR, Kiessling LL. Science. 2001; 291:2357. [PubMed: 11269316]

18. Hojo H, Nakahara Y. Biopolymers. 2007; 88:308. [PubMed: 17278124]

19. Buskas T, Ingale S, Boons GJ. Glycobiology. 2006; 16:113R.

20. Merrifield RB. J Am Chem Soc. 1963; 85:2149.

21. Yamamoto N, Ohmori Y, Sakakibara T, Sasaki K, Juneja LR, Kajihara Y. Angew Chem Int Ed. 2003; 42:2537.

22. Yamamoto N, Takayanagi A, Yoshino A, Sakakibara T, Kajihara Y. Chem Eur J. 2007; 13:613. [PubMed: 16977655]

23. Mathieux N, Paulsen H, Meldal M, Bock K. J Chem Soc Perkin Trans. 1997; 1:2359.

24. Herzner H, Reipen T, Schultz M, Kunz H. Chem Rev. 2000; 100:4495. [PubMed: 11749356]

25. Gaidzik N, Westerlind U, Kunz H. Chem Soc Rev. 2013; 42:4431.

26. Wilson RM, Danishefsky SJ. J Am Chem Soc. 2013; 135:14462. [PubMed: 23944352]

27. Lakshminarayanan V, Thompson P, Wolfert MA, Buskas T, Bradley JM, Pathangey LB, Madsen CS, Cohen PA, Gendler SJ, Boons GJ. Proc Natl Acad Sci U S A. 2012; 109:261. [PubMed: 22171012]

28. Cohen-Anisfeld ST, Lansbury PT. J Am Chem Soc. 1993; 115:10531.

29. Bodanszky M, Natarajan S. J Org Chem. 1975; 40:2495. [PubMed: 1165511]

30. Wang P, Aussedat B, Vohra Y, Danishefsky SJ. Angew Chem Int Ed. 2012; 51:11571.

31. Ullmann V, Rädisch M, Boos I, Freund J, Pöhner C, Schwarzinger S, Unverzagt C. Angew Chem Int Ed. 2012; 51:11566.

32. Aussedat B, Vohra Y, Park PK, Fernández-Tejada A, Alam SM, Dennison SM, Jaeger FH, Anasti K, Stewart S, Blinn JH, Liao HX, Sodroski JG, Haynes BF, Danishefsky SJ. J Am Chem Soc. 2013; 135:13113. [PubMed: 23915436]

33. Payne RJ, Wong CH. Chem Comm. 2010; 46:21. [PubMed: 20024291]

34. Dawson P, Muir T, Clark-Lewis I, Kent SBH. Science. 1994; 266:776. [PubMed: 7973629]

35. Muir TW. Annu Rev Biochem. 2003; 72:249. [PubMed: 12626339]

36. Macmillan D, Bertozzi CR. Angew Chem Int Ed. 2004; 43:1355.

37. Kent SBH. Chem Soc Rev. 2009; 38:338. [PubMed: 19169452]

38. Unverzagt C, Kajihara Y. Chem Soc Rev. 2013; 42:4408. [PubMed: 23403448]

39. Shin Y, Winans KA, Backes BJ, Kent SBH, Ellman JA, Bertozzi CR. J Am Chem Soc. 1999; 121:11684.

40. Marcaurelle LA, Mizoue LS, Wilken J, Oldham L, Kent SB, Handel TM, Bertozzi CR. Chem Eur J. 2001; 7:1129. [PubMed: 11303872]

41. Mezzato S, Schaffrath M, Unverzagt C. Angew Chem Int Ed. 2005; 44:1650.

42. Yamamoto N, Tanabe Y, Okamoto R, Dawson PE, Kajihara Y. J Am Chem Soc. 2008; 130:501. [PubMed: 18085777]

43. Mende F, Seitz O. Angew Chem Int Ed. 2011; 50:1232.

44. Kan C, Danishefsky SJ. Tetrahedron. 2009; 65:9047. [PubMed: 20161377]

45. Yuan Y, Chen J, Wan Q, Wilson RM, Danishefsky SJ. Biopolymers. 2010; 94:373. [PubMed: 20593459]

46. Warren JD, Miller JS, Keding SJ, Danishefsky SJ. J Am Chem Soc. 2004; 126:6576. [PubMed: 15161285]

47. Wan Q, Chen J, Yuan Y, Danishefsky SJ. J Am Chem Soc. 2008; 130:15814. [PubMed: 18855357]

48. Yuan Y, Chen J, Wan Q, Tan Z, Chen G, Kan C, Danishefsky SJ. J Am Chem Soc. 2009; 131:5432. [PubMed: 20560636]

49. Haase C, Seitz O. Angew Chem Int Ed. 2008; 47:1553.

50. Macmillan D, Anderson DW. Org Lett. 2004; 6:4659. [PubMed: 15575654]

51. Wu B, Chen J, Warren JD, Chen G, Hua Z, Danishefsky SJ. Angew Chem Int Ed. 2006; 45:4116.

52. Yan LZ, Dawson PE. J Am Chem Soc. 2001; 123:526. [PubMed: 11456564]

53. Pentelute BL, Kent SBH. Org Lett. 2007; 9:687. [PubMed: 17286375]

54. González A, Valencia G. Tetrahedron Asymmetry. 1998; 9:2761.

55. Wan Q, Danishefsky SJ. Angew Chem Int Ed. 2007; 46:9248.

56. Shang S, Tan Z, Danishefsky SJ. Proc Natl Acad Sci U S A. 2011; 108:5986. [PubMed: 21444787]

57. Dong S, Shang S, Li J, Tan Z, Dean T, Maeda A, Gardella TJ, Danishefsky SJ. J Am Chem Soc. 2012; 134:15122. [PubMed: 22891619]

58. Li J, Dong S, Townsend SD, Dean T, Gardella TJ, Danishefsky SJ. Angew Chem Int Ed. 2012; 51:12263.

59. Brailsford JA, Danishefsky SJ. Proc Natl Acad Sci U S A. 2012; 109:7196. [PubMed: 22499784]

60. Chen J, Wan Q, Yuan Y, Zhu J, Danishefsky SJ. Angew Chem Int Ed. 2008; 47:8521.

61. Haase C, Rohde H, Seitz O. Angew Chem Int Ed. 2008; 47:6807.

62. Chen J, Wang P, Zhu J, Wan Q, Danishefsky SJ. Tetrahedron. 2010; 66:2277. [PubMed: 20798898]

63. Harpaz Z, Siman P, Kumar KSA, Brik A. ChemBioChem. 2010; 11:1232. [PubMed: 20437446]

64. Tan Z, Shang S, Danishefsky SJ. Angew Chem Int Ed. 2010; 49:9500.

65. Yang R, Pasunooti KK, Li F, Liu X-W, Liu C-F. J Am Chem Soc. 2009; 131:13592. [PubMed: 19728708]

66. Ajish Kumar KS, Haj-Yahya M, Olschewski D, Lashuel HA, Brik A. Angew Chem Int Ed. 2009; 48:8090.

67. Shang S, Tan Z, Dong S, Danishefsky SJ. J Am Chem Soc. 2011; 133:10784. [PubMed: 21696220]

68. Townsend SD, Tan Z, Dong S, Shang S, Brailsford JA, Danishefsky SJ. J Am Chem Soc. 2012; 134:3912. [PubMed: 22332757]

69. Crich D, Banerjee A. J Am Chem Soc. 2007; 129:10064. [PubMed: 17658806]

70. Tam JP, Yu Q. Biopolymers. 1998; 46:319. [PubMed: 9754028]

71. Saporito A, Marasco D, Chambery A, Botti P, Monti SM, Pedone C, Ruvo M. Biopolymers. 2006; 83:508. [PubMed: 16888773]

72. Okamoto R, Kajihara Y. Angew Chem Int Ed. 2008; 47:5402.

73. Malins LR, Cergol KM, Payne RJ. ChemBioChem. 2013; 14:559. [PubMed: 23426906]

74. Thompson RE, Chan B, Radom L, Jolliffe KA, Payne RJ. Angew Chem Int Ed. 2013; 52:9723.

75. Malins LR, Cergol KM, Payne RJ. Chem Sci. 2014; 5:260.

76. Cergol KM, Thompson RE, Malins LR, Turner P, Payne RJ. Org Lett. 2014; 16:290. [PubMed: 24294973]

77. Nagorny P, Fasching B, Li X, Chen G, Aussedat B, Danishefsky SJ. J Am Chem Soc. 2009; 131:5792. [PubMed: 19341309]

78. Zhang YM, Mallet JM, Sinaÿ P. Carbohydr Res. 1992; 236:73. [PubMed: 1291063]

79. Likhosherstov LM, Novikova OS, Derevitskaja VA, Kochetkov NK. Carbohydr Res. 1986; 146:C1.

80. Sakakibara S. Biopolymers. 1995; 37:17. [PubMed: 7880964]

81. Aussedat B, Fasching B, Johnston E, Sane N, Nagorny P, Danishefsky SJ. J Am Chem Soc. 2012; 134:3532. [PubMed: 22280541]

82. Howles CM. Hum Reprod Update. 1996; 2:172. [PubMed: 9079412]

83. Pang SC. Womens Health. 2005; 1:87.

84. Nagorny P, Sane N, Fasching B, Aussedat B, Danishefsky SJ. Angew Chem Int Ed. 2012; 51:975.

85. Stenman UH, Tiitinen A, Alfthan H, Valmu L. Hum Reprod Update. 2006; 12:769. [PubMed: 16877746]

86. Talwar GP, Vyas HK, Purswani S, Gupta JC. J Reprod Immunol. 2009; 83:158. [PubMed: 19854518]

87. Iversen PL, Mourich DV, Moulton HM. Curr Opin Mol Ther. 2003; 5:156. [PubMed: 12772505]

88. Valmu L, Alfthan H, Hotakainen K, Birken S, Stenman UH. Glycobiology. 2006; 16:1207. [PubMed: 16880503]

89. Fernández-Tejada A, Vadola PA, Danishefsky SJ. J Am Chem Soc. 2014; 136:8450. [PubMed: 24806200]

90. Shi Y, Liu CH, Roberts AI, Das J, Xu G, Ren G, Zhang Y, Zhang L, Yuan ZR, Tan HS, Das G, Devadas S. Cell Res. 2006; 16:126. [PubMed: 16474424]

91. Zhang Q, Johnston EV, Shieh J-H, Moore MAS, Danishefsky SJ. Proc Natl Acad Sci U S A. 2014; 111:2885. [PubMed: 24516138]

92. Krantz SB. Blood. 1991; 77:419. [PubMed: 1991159]

93. Lacombe C, Mayeux P. Haematologica. 1998; 83:724. [PubMed: 9793257]

94. Koury MJ. Exp Hematol. 2005; 33:1263. [PubMed: 16263408]

95. Tonelli M, Hemmelgarn B, Reiman T, Manns B, Reaume MN, Lloyd A, Wiebe N, Klarenbach S. CMAJ. 2009; 180:E62. [PubMed: 19407261]

96. Mattio M, Ceaglio N, Oggero M, Perotti N, Amadeo I, Orozco G, Forno G, Kratje R, Etcheverrigaray M. Biotechnol Prog. 2011; 27:1018. [PubMed: 21608141]

97. Higuchi M, Oh-eda M, Kuboniwa H, Tomonoh K, Shimonaka Y, Ochi N. J Biol Chem. 1992; 267:7703. [PubMed: 1560005]

98. Wang P, Dong S, Brailsford JA, Iyer K, Townsend S, Zhang Q, Hendrickson RC, Shieh JH, Moore MAS, Danishefsky SJ. Angew Chem Int Ed. 2012; 51:11576.

99. Schwarz JB, Kuduk SD, Chen X-T, Sames D, Glunz PW, Danishefsky SJ. J Am Chem Soc. 1999; 121:2662.

100. Bang D, Pentelute BL, Kent SBH. Angew Chem Int Ed. 2006; 45:3985.

101. Wang P, Dong S, Shieh J-H, Peguero E, Hendrickson R, Moore MAS, Danishefsky SJ. Science. 2013; 342:1357. [PubMed: 24337294]

102. Takeuchi M, Inoue N, Strickland TW, Kubota M, Wada M, Shimizu R, Hoshi S, Kozutsumi H, Takasaki S, Kobata A. Proc Natl Acad Sci U S A. 1989; 86:7819. [PubMed: 2813359]

103. Kochendoerfer GG, Chen S-Y, Mao F, Cressman S, Traviglia S, Shao H, Hunter CL, Low DW, Cagle EN, Carnevali M, Gueriguian V, Keogh PJ, Porter H, Stratton SM, Wiedeke MC, Wilken J, Tang J, Levy JJ, Miranda LP, Crnogoran MM, Kalbag S, Potti P, Schindler-Horvat J, Savatski L, Adamson JW, Kung A, Kent SBH, Bradburne JA. Science. 2003; 299:884. [PubMed: 12574628]

104. Hamilton SR, Davidson RC, Sethuraman N, Nett JH, Jiang Y, Rios S, Bobrowicz P, Stadheim TA, Li H, Choi B-K, Hopkins D, Wischnewski H, Roser J, Mitchell T, Strawbridge RR, Hoopes J, Wildt S, Gerngross TU. Science. 2006; 313:1441. [PubMed: 16960007]

105. Sanz-Nebot V, Benavente F, Vallverdú A, Guzman NA, Barbosa J. Anal Chem. 2003; 75:5220. [PubMed: 14708798]

106. Harazono A, Hashii N, Kuribayashi R, Nakazawa S, Kawasaki N. J Pharm Biomed Anal. 2013; 87:65.

107. Chalker JM, Bernardes GJ, Davis BG. Acc Chem Res. 2011; 44:730. [PubMed: 21563755]

108. Wang LX, Lomino JV. ACS Chem Biol. 2012; 7:110. [PubMed: 22141574]

109. Wang LX, Amin MN. Chem Biol. 2014; 21:51. [PubMed: 24439206]

110. Wang L, Schultz PG. Angew Chem Int Ed. 2005; 44:34.

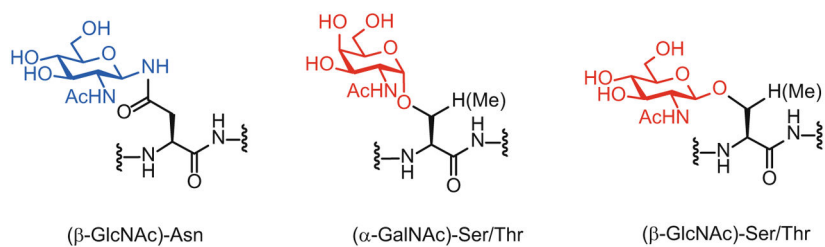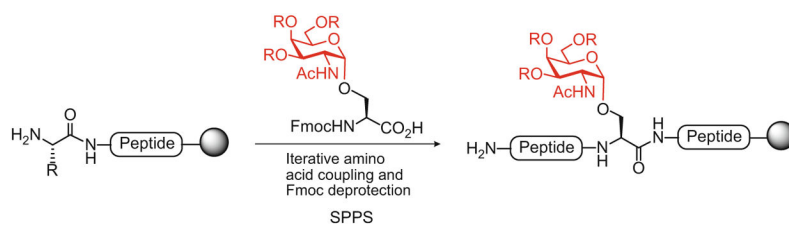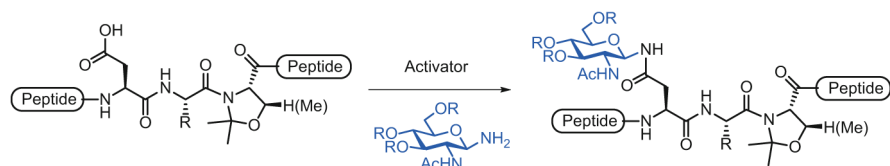111. Liu L, Bennett CS, Wong CH. Chem Comm. 2006; 42:21.

Fig. 1.
Common (β-GlcNAc)-*N*-linked asparagine and (α-GalNAc)/(β-GlcNAc)-*O*-linked serine/threonine in glycosylated proteins
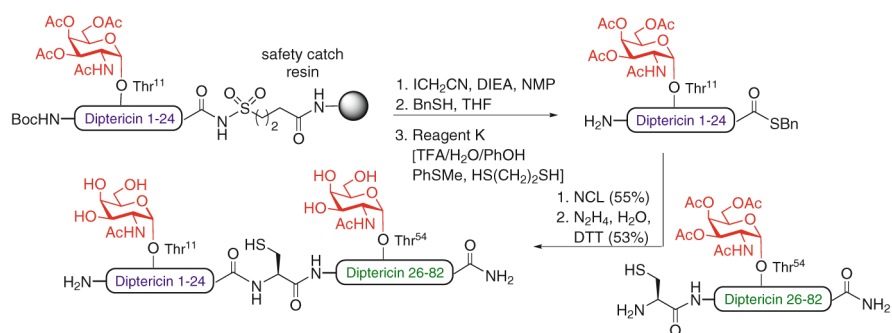
Scheme 1.
Linear "cassette"-based approach for the synthesis of *O*-linked glycopeptides
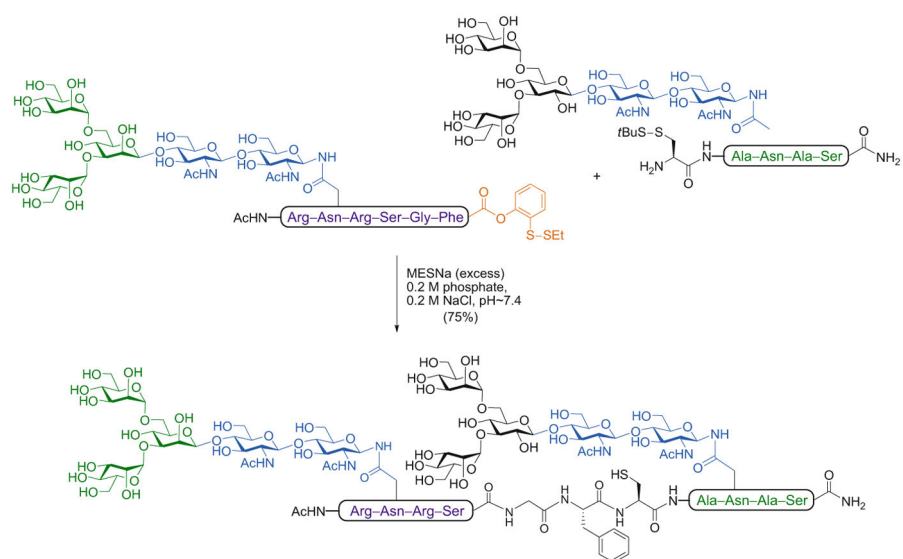
Scheme 2.
Lansbury aspartylation reaction facilitated by a pseudoproline motif for the convergent synthesis of *N*-linked glycopeptides
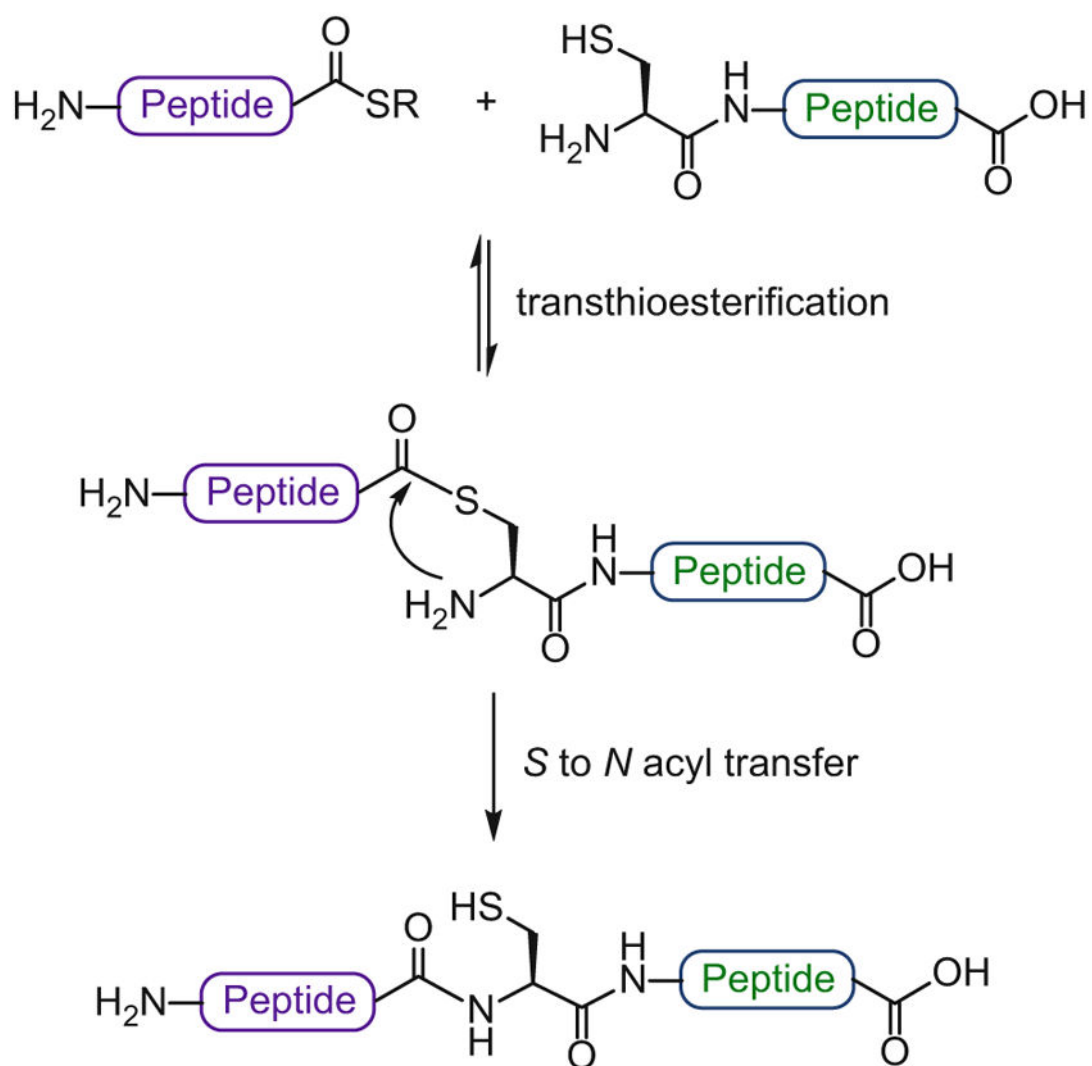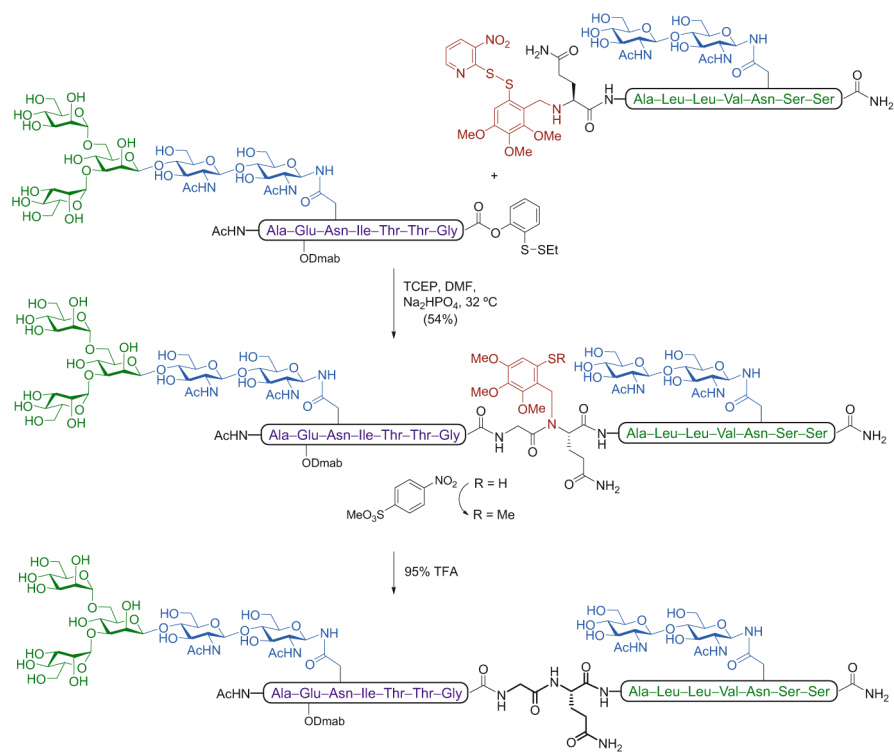
Scheme 3.
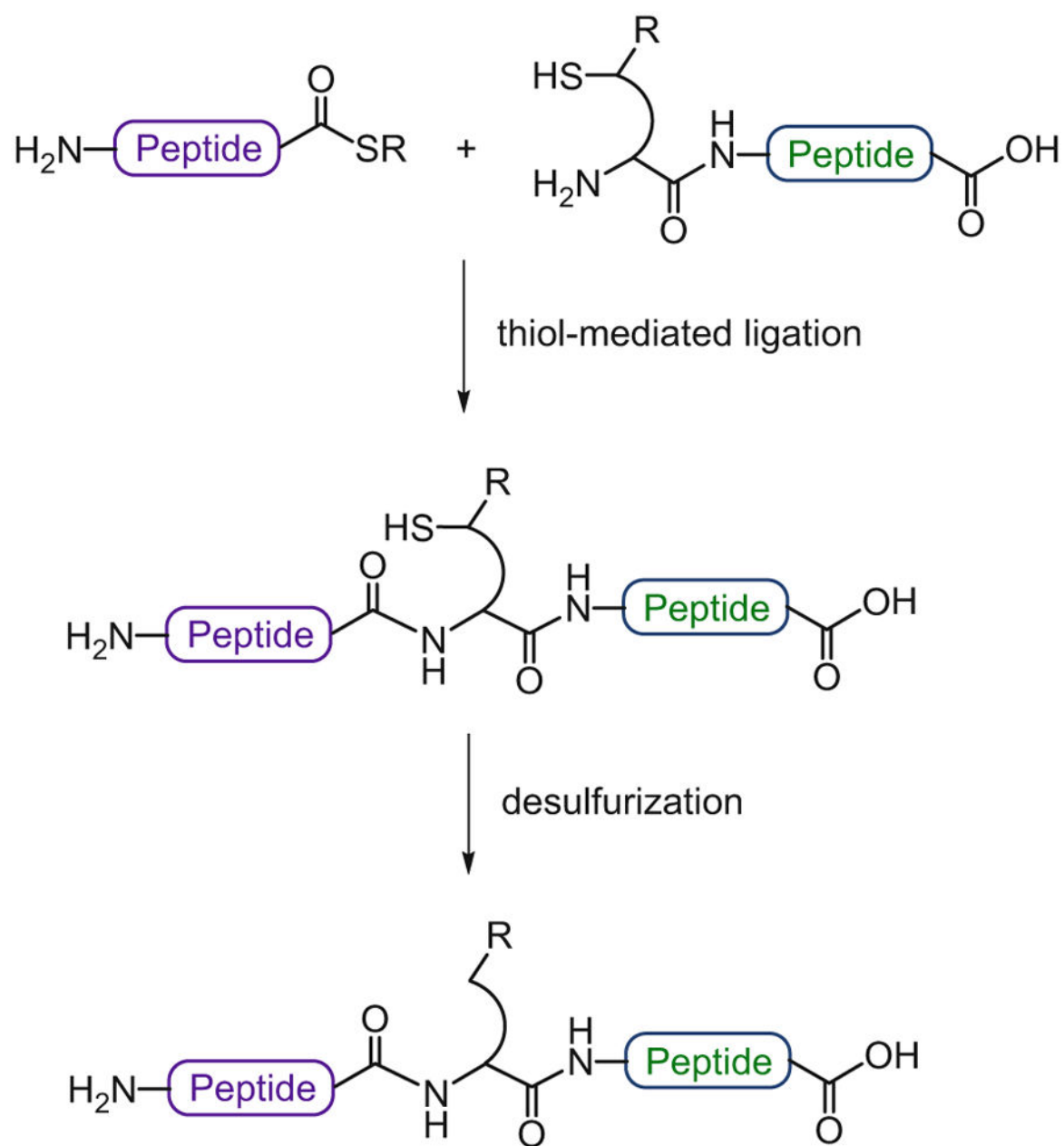Native chemical ligation (NCL)

Scheme 4.
Total synthesis of antimicrobial *O*-glycoprotein diptericin by NCL

Scheme 5.
NCL-based assembly of a complex model *N*-glycopeptide using an *o*-disulfide phenolic ester as latent thioester

Scheme 6.
Synthesis of a complex model *N*-glycopeptide using an auxiliary-based, cysteine-free ligation strategy

Scheme 7.
Non-cysteine ligation followed by desulfurization
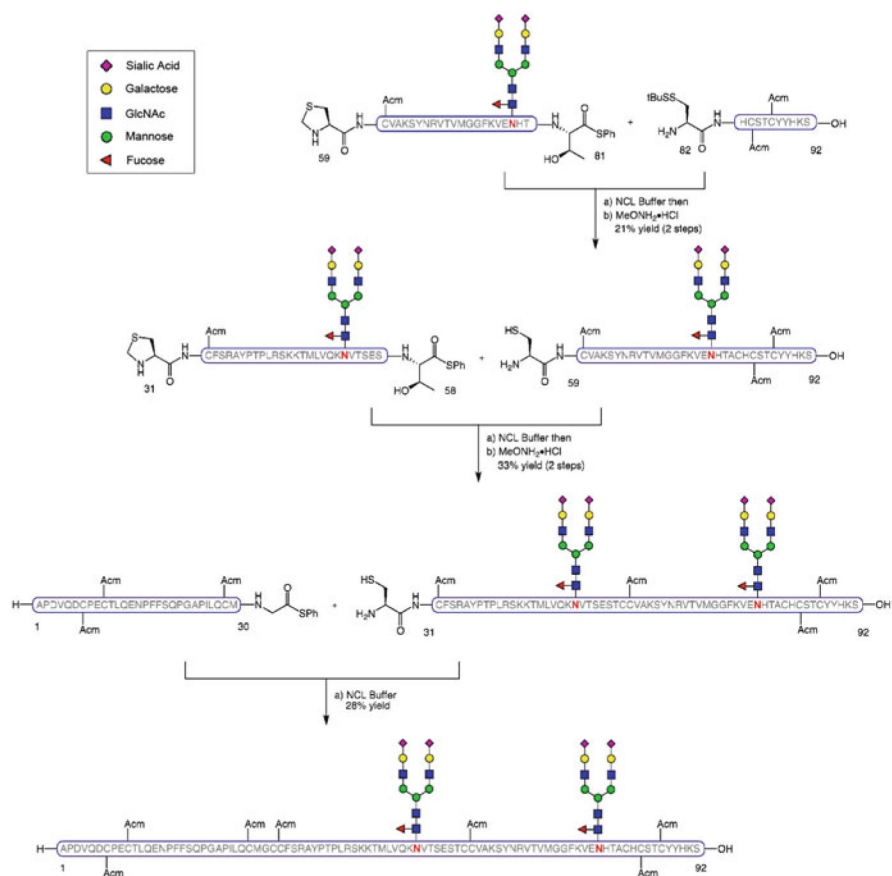
Scheme 8.
Synthesis of the α-subunit of human glycoprotein hormones, α-hGPH[1–92]
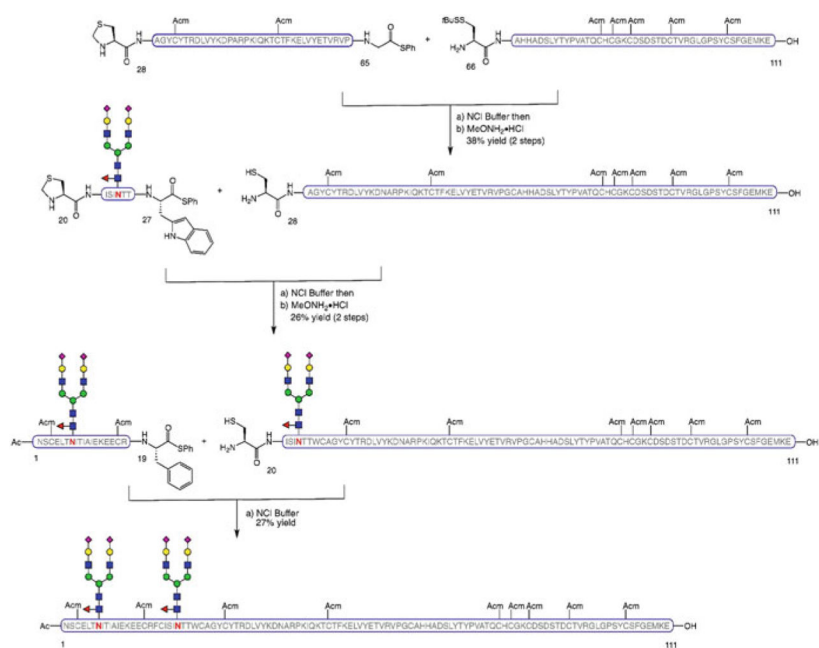
Scheme 9.
Synthesis of the β-subunit of human follicle stimulating hormone, β-hFSH[1–111]
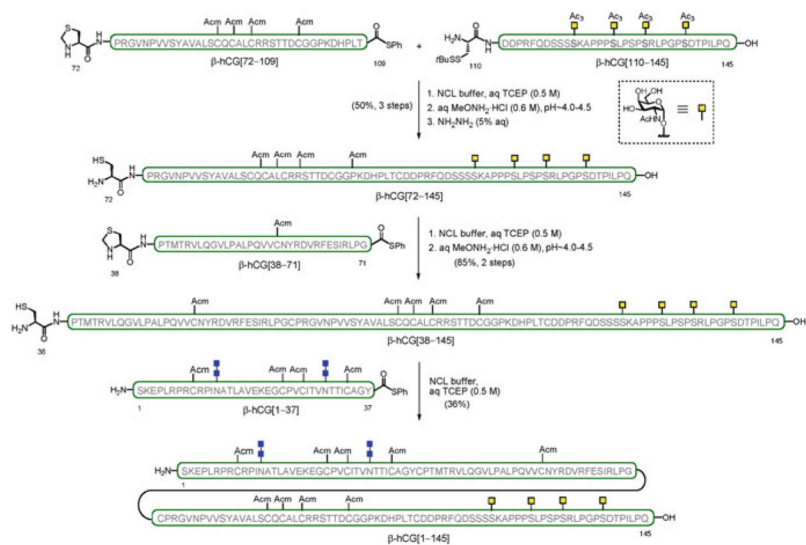
Scheme 10.
Synthesis of β-subunit of human chorionic gonadotropin hormone, β-hCG[1–145]
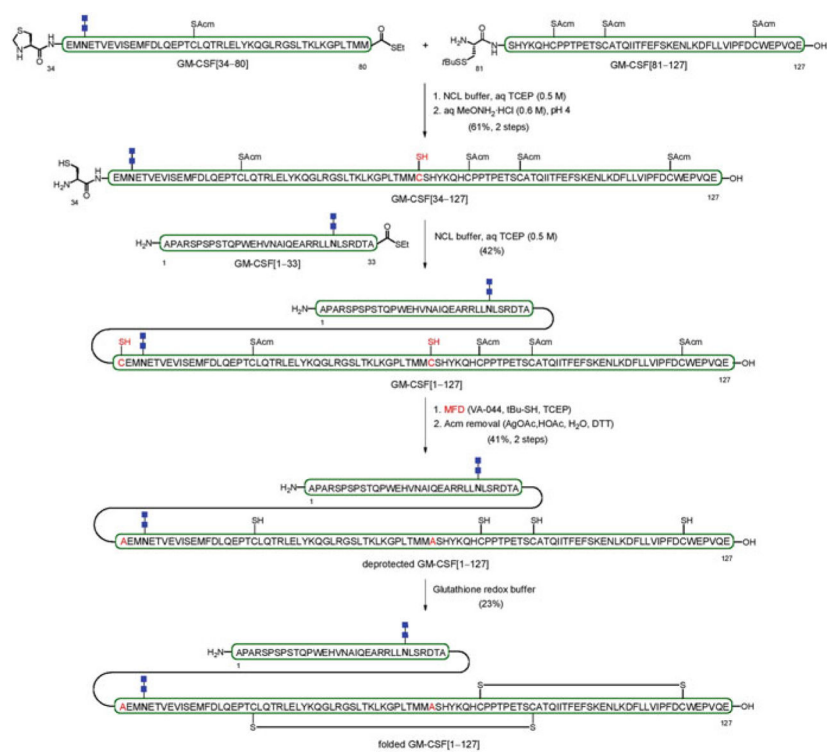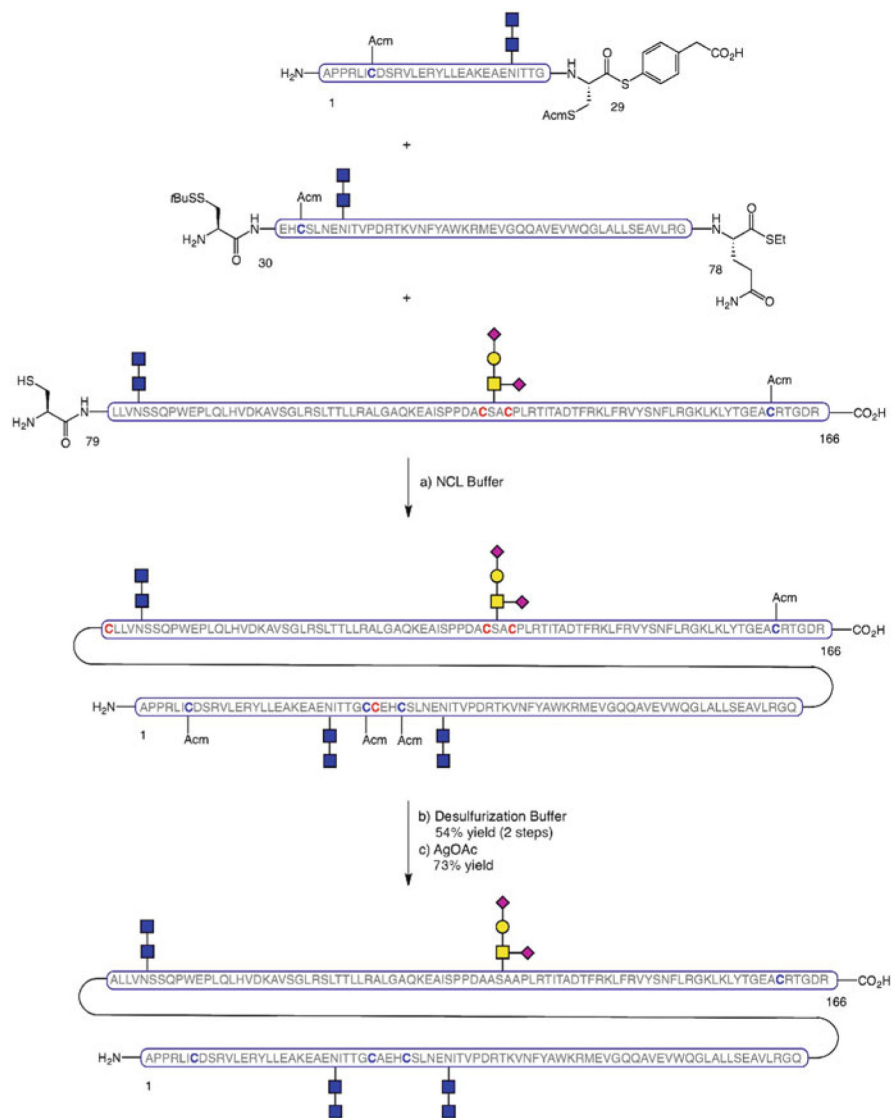
Scheme 11.
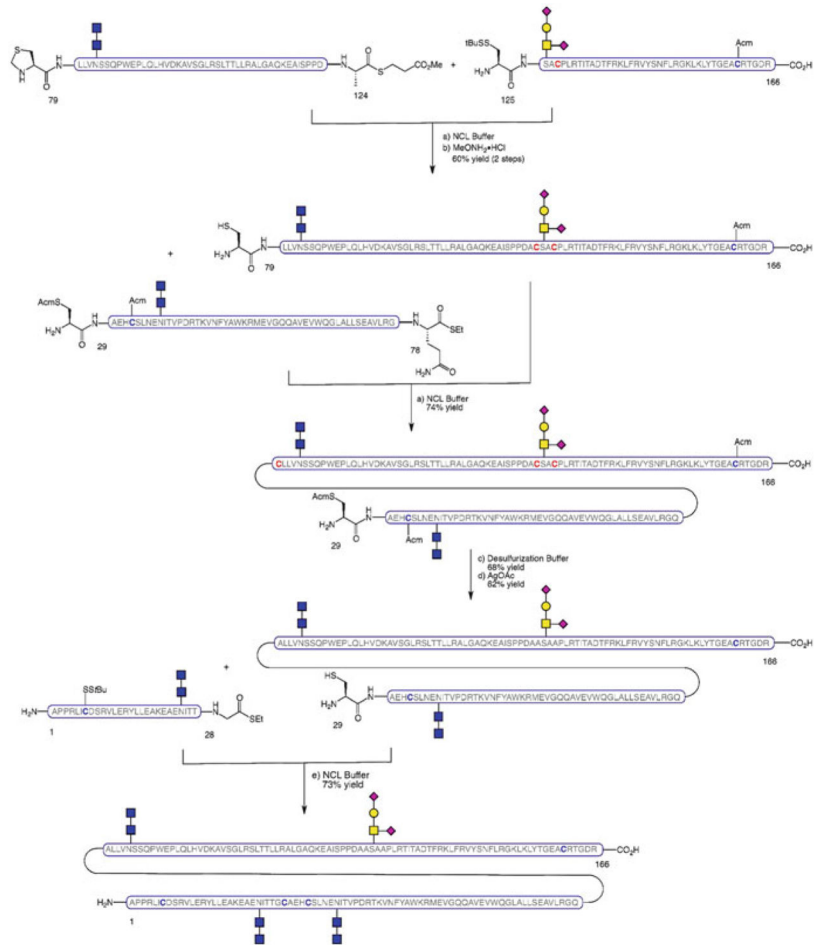Synthesis of the bis-glycosylated GM-CSF glycoform
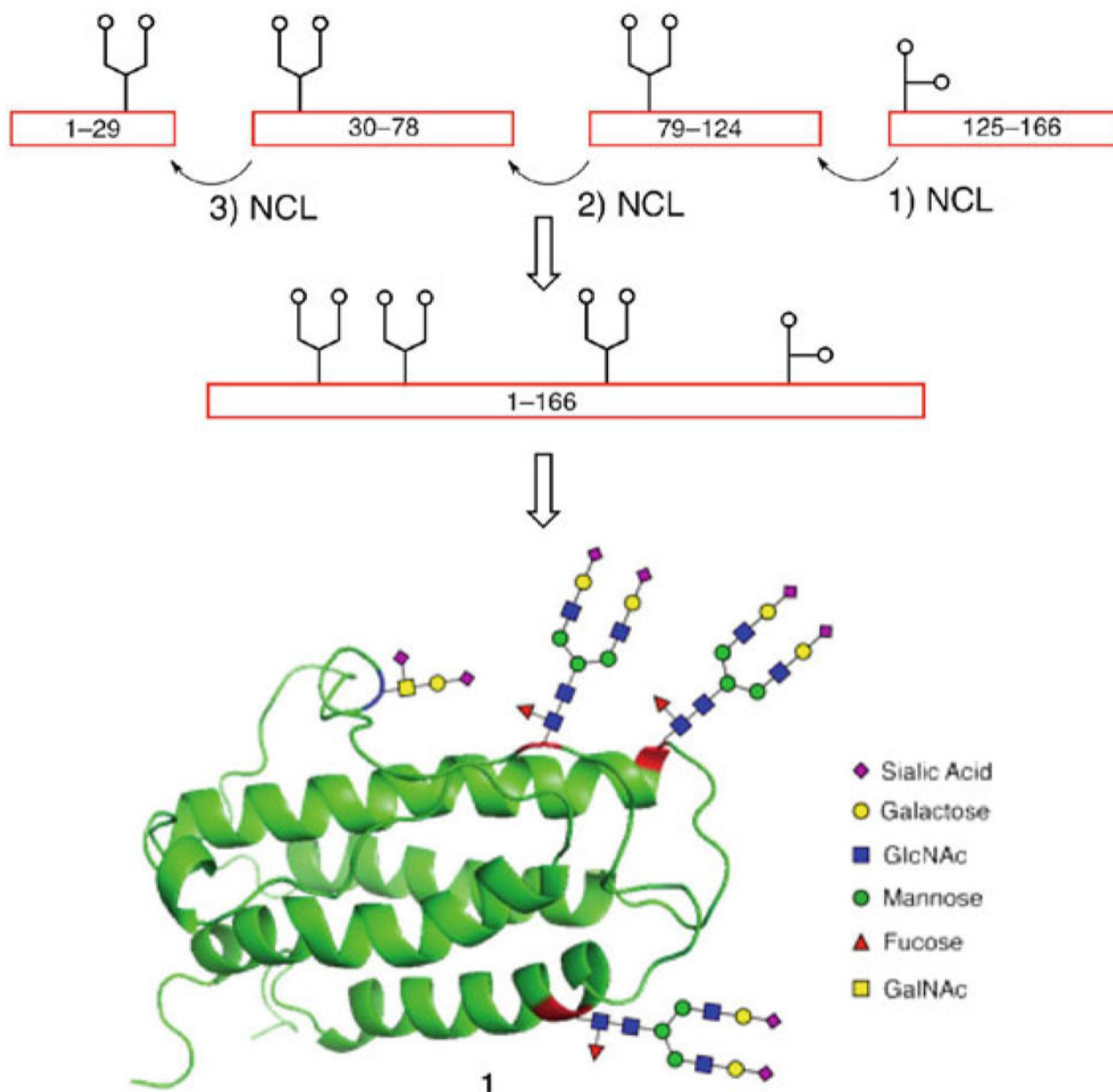
Scheme 12.
Proposed synthetic strategy for convergent synthesis of erythropoietin, EPO[1–166] and ribbon diagram of the fully glycosylated protein
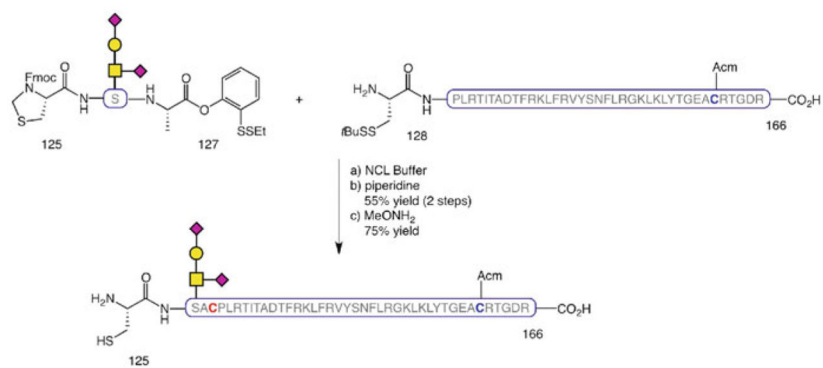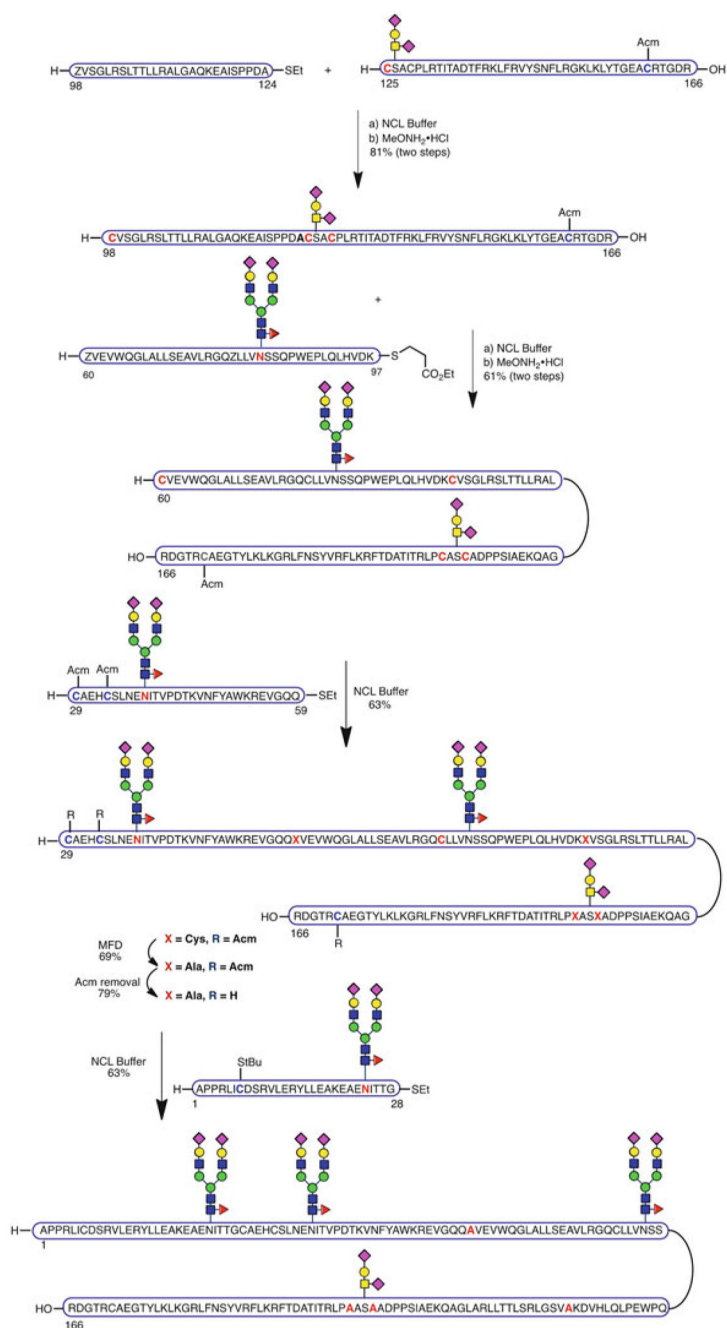
Scheme 13.
Synthesis of EPO[125–166]

Scheme 14.
Initial convergent synthesis of EPO

Scheme 15.
Linear synthesis of erythropoietin, EPO[1–166]

Scheme 16.
Synthesis of fully glycosylated EPO[1–166] bearing wild-type glycans at each of the four glycosylation sites