Published in final edited form as:

Biopolymers. 2010 ; 94(4): 373–384. doi:10.1002/bip.21374.

Toward Fully Synthetic, Homogeneous Glycoproteins: Advances in Chemical Ligation

Yu Yuan¹, Jin Chen², Qian Wan¹, Rebecca M. Wilson¹, and Samuel J. Danishefsky^{1,3}

¹Laboratory for Bioorganic Chemistry, Sloan-Kettering Institute for Cancer Research, 1275 York Avenue, New York, NY 10065

²Department of Chemistry, Michigan Technological University, 1400 Townsend Drive, Houghton, MI 49931

³Department of Chemistry, Columbia University, Havemeyer Hall, 3000 Broadway, New York, NY 10027

Abstract

Traditionally, in the pharma sciences, there has been an unstated but operative bifurcation into small molecules and biologics. Small molecules were seen to be, at the discovery level, in the province of chemistry, based on targets provided through biology. By contrast, “biologics” were seen to arise solely from the province of biology exploiting its accessible replicative mechanisms. Our laboratory has been dedicated to the proposition that explosive advances in chemical synthesis have been such as to render so called “biologics” as being accessible to chemical synthesis. In this paper, we focus particularly on the area of glycopeptides. Chemical synthesis, in principle, offers an advantage, in that it can lead to homogeneous glycopeptides characterized by a single glycoform of the glycosidic domain mounted at a particular amino acid in the polypeptide domain. In support of this defining goal, a variety of new methods have been developed. The key problem addressed is that of ligation. In this paper, we review how insights available from mechanistic organic chemistry have been used to create an imposing framework for the synthesis of structures which would, in an earlier day, have been seen to be strictly in the realm of chemically inaccessible “biologics”.

Introduction

The study of protein structure, activity, and interactions has emerged as an important focus of modern chemistry, as well as the biological and pharmaceutical sciences. Given the structural complexity of these biomacromolecules, it perhaps comes as no surprise that researchers often tend to focus their investigations on individual polypeptide regions of the protein, in order to gain insight into the fundamentals of protein behavior. The difficulty associated with isolating chemically homogeneous peptidic and glycopeptidic fragments provides strong impetus for the development of efficient and broadly useful methods for the total synthesis of these complex and valuable targets. Major technological and methodological advances have been made in this area in recent years. Solid phase peptide synthesis (SPPS)¹ allows for the efficient assembly of polypeptides composed of up to 50 amino acid residues. With the development of cysteine-based native chemical ligation (NCL) by Kent and co-workers,² it became possible to prepare even longer peptide fragments, provided that a cysteine residue was situated at an appropriate location along the

We dedicate this paper to the remarkable accomplishments of Professor Steven Kent – particularly, his discovery of NCL, which changed the course of peptide ligations.

target peptide sequence. In practice, however, this requirement is often problematic, due to the fact that cysteine residues occur with relatively low frequency in naturally occurring protein sequences. Furthermore, the Kent NCL protocol, which requires pre-formation of a thioacid acyl donor, is not generally applicable to the synthesis and ligation of glycopeptide targets.

A major focus of our laboratory is centered on the development of improved methods for the chemical synthesis of homogeneous, structurally defined biologic agents. Along these lines, we have developed a number of strategies, which we think are broadly useful for peptide and glycopeptide assembly. This review will focus on our efforts in three areas: (1) development of stable acyl donors to enable glycopeptide–glycopeptide ligation; (2) development of auxiliary-based protocols to allow for NCL with non-cysteine N-terminal acyl acceptors; and (3) development of a mild ligation/desulfurization sequence, which has allowed us to expand NCL capabilities to a range of other naturally occurring amino acids.

(1) Development of Acyl Donor for Glycopeptide Synthesis

An overarching objective of our laboratory is to develop the means by which to synthesize, *de novo*, complex, multiply glycosylated proteins. Toward this end, we envision that maximum convergence could be achieved through the preparation of individual glycopeptide domains of comparable size, which would be merged through sequential ligation reactions (Scheme 1).

Critical to the success of this approach would be the judicious choice of acyl donor (group X, compounds **2** and **3**). First, this functionality should be able to withstand strongly acidic conditions, since deprotection of the polypeptide side chains is likely to require exposure to 95% trifluoroacetic acid. Furthermore, the acyl donor (X) should be inert to the conditions required for glycan attachment, in order to ensure that the carbohydrates are conjugated at the desired position. Finally, upon activation, the resulting acyl donor should be capable of participating in the requisite aminolysis reaction. Upon considering these prerequisites, we sought to examine the viability of a masked *ortho*-thiophenolic acyl donor, of the type **8** (Scheme 2).³

We projected that a peptide (**8**) equipped with a latent C-terminal *ortho*-thiophenol functionality should be stable to standard aspartylation conditions (typically accomplished through amidation, **8**→**9**). Under reducing conditions, disulfide cleavage was expected to trigger acyl migration between the O and S atoms (**10**↔**11**). NMR studies have indicated that this equilibrium is pH-dependent: at lower pH (pH<6) compound **11** is predominant in the mixture; yet once the pH value reaches 7 or higher, **10** and **11** are in fast dynamic exchange.⁴ It was anticipated that either of these two species might be sufficiently activated to undergo ligation with a second glycopeptide, furnishing the key amide bond in product **12**.

Indeed, the logic underlying this proposal was borne out experimentally.³ In a representative example, pentasaccharide **13** was coupled with peptide **15**, pre-equipped with the C-terminal *ortho*-thiophenolic moiety, to provide glycopeptide **17** in 50% yield (Scheme 3). The latter readily underwent NCL with glycopeptide **18** to furnish **19** in 75% yield. This striking example serves to highlight the power of our novel glycopeptide ligation methodology as a means by which to gain rapid access to glycopeptides possessing unnatural carbohydrate domains, which are difficult to obtain through other methods.

Aside from its high reactivity in glycopeptide and peptide ligations, we recognized that the intrinsic electrophilicity of the aryl thioester might offer the possibility of achieving direct condensations with general free amine acceptors. Inspired by Aimoto's AgCl system,^{5,6} we

found that, following cleavage of the disulfide, the resulting thioester reacted readily to give ligated peptide in the presence of HOObt, DIEA and AgCl (Scheme 4). Furthermore, we observed that, in the absence of reducing agent, the same reaction gave even cleaner products, providing access to linear or cyclic peptides in good to excellent yields, with various amine acceptors, such as Leu, Asp and Gln.⁷ Unexpectedly, even under metal-free conditions, the masked phenolic ester can be transferred to a reactive acyl donor, to provide ligation adducts in good yields (Scheme 4). This finding is particularly important, since it indicates a suitable kinetic window within which it should be possible to differentiate between aryl thioesters and alkyl thioesters; thereby providing the theoretical basis for the development of a *reiterative* ligation strategy.

Experimentally, our reiterative ligation approach has been demonstrated with the synthesis of glycopeptide **25** (Scheme 5). Thus, a TCEP-mediated condensation reaction served to selectively activate the aryl thioester while leaving the alkyl thioester intact, providing glycopeptide **23** in excellent yield. The subsequent union of **23** and **24** furnished the desired product **25**, demonstrating the high promise of this reiterative ligation strategy.

Although powerful for a wide range of substrates, the acyl transfer ability provided by the *ortho*-thiophenolic ester does not meet the steric demands of some hindered amino acid coupling partners. In light of Bodanszky's research,⁸ which uses AA nitrophenol ester to elongate peptides, we sought to explore higher energy peptide acyl donors with oxo-ester functionality.⁹ In practice, these oxo-acyl donors turned out to be very reactive in NCL reactions, even for the most hindered C-terminal amino acids. For example, when C-terminal Thr, Val, Ile and D-*allo*-Ile were equipped with nitrophenolic ester, the ligations were typically completed within 6–7 h, with yields ranging from 60–70%; C-terminal Pro substrates required longer reaction times, furnishing products with slightly lower yields (~50%). The superior kinetic profile of the oxo-ester in NCL is illustrated in Scheme 6. Thus, ligation between peptide **26a**, bearing a C-terminal Ile thioester, and **27** under standard NCL conditions took 72 h to generate product **28** in 58% yield; while the corresponding *p*-nitrophenol ester afforded a 70% yield of the ligated product within only 7 h.

This kinetic advantage of oxo-ester donors over aryl thioesters have already allowed us to achieve important advances in glycopeptide synthesis. Thus, traditional peptide elongation is performed from C to N terminus. The discoveries in our laboratory indicate that, under appropriate conditions, it is possible to tune the ligation rates for different acyl donors; therefore, these combined technologies serve to provide a means by which to accomplish peptide ligation from N to C terminus. A persuasive example of such a strategy was recently demonstrated in our synthesis of the erythropoietin (EPO) Cys²⁹–Gly⁷⁷ fragment.¹⁰ It is worth noting that the diverse substitution potential on the aromatic region of the oxo-ester allows us to regulate reactivity to match particular substrate features. This flexibility turned out to be critical for the attachment of complex carbohydrate **29** to EPO fragment Cys²⁹–Pro⁴², which contains a C-terminal oxo-ester. When R₂ is a nitro group, product **33** is not observed; instead, either hydrolysis or direct aminolysis (not rigorously determined) overwhelmed the desired pathway. However, once the nitro group was replaced with a cyano functionality, the desired glycopeptide **34** was formed (28% yield). We were pleased to further observe that the cyanoester function also tolerates free secondary amine, and **34** was readily converted to glycopeptide **35**. Finally, direct condensation with peptide **36** furnished the key intermediate **37** in 47% yield. In this case, a modified *ortho*-thioester was successfully employed to suppress the hydrolysis of glycine phenol ester.

With a range of valuable acyl donor groups in hand, our focus now shifted to the development of a range of non-cysteine based acyl *acceptor* substrate types. Each of the

non-cysteine based ligation strategies described below relies on the use of a temporary thiol functionality, which serves the role typically played by the cysteine thiol group in initiating ligation. In these strategies, the transient thiol may be presented either as an N-terminal auxiliary (Section 2) or as an unnatural (or natural) amino acid, which itself serves as a surrogate for a non-thiol containing natural amino acid (Section 3). In the latter approach, the sulfur moiety is removed following ligation through a mild desulfurization methodology developed in our laboratory.

(2) Development of Auxiliary-Based Acyl Acceptors for Glycopeptide Synthesis

In our effort to expand the capabilities of NCL to accommodate acyl acceptors presenting non-cysteine residues at the N-termini, we first sought to examine an auxiliary-based approach, wherein a temporary N-terminal functionality would serve to mimic the function of the cysteine thiol. According to the auxiliary-based ligation strategy outlined in Scheme 8, a cleavable thiol-containing group would be introduced onto the N-terminus of the acceptor glycopeptide fragment. We envisioned that, under reaction conditions, the thiol functionality would undergo acyl exchange with the *in situ* generated thioester donor, to produce an intermediate in which the two glycopeptide fragments are covalently connected through a mercaptobenzyl linker. Due to the proximity of the carbonyl and amino groups, it was expected that a 1,5 acyl transfer reaction will be triggered, resulting in formation of the thermodynamically favored amide bond. Upon removal of the auxiliary, the native peptide backbone would be in hand.¹¹

This strategy was, in fact, reduced to practice. In a representative example, shown in Scheme 8, the two disulfide bonds of glycopeptides **38** and **39** were concomitantly reduced, and the resulting active species joined, presumably via the proposed mechanism, to afford glycopeptide **40**, bearing two distinct glycan domains. Finally, the auxiliary was removed through sequential methylation and TFA treatment, which serves to protonate the amide bond of the auxiliary, leading to cleavage of the electron-rich aromatic moiety. Notably, no carbohydrate decomposition was observed under these auxiliary removal conditions.

Interestingly, we have found that the initial organizing linkage between the two peptide domains does not necessarily need to be established through NCL. As outlined in Scheme 9, we recently conceived of a cysteine-free ligation protocol, wherein the two glycopeptide fragments to be joined are positioned in a *meta* arrangement on a benzylic scaffold, with a protected thiol residing between them. Thiol deprotection sets into motion an O→S acyl transfer, and the resulting thioester is then situated to undergo intramolecular S→N acyl transfer with the amine of the second glycopeptide.⁴ This migration is driven by the thermodynamic stability of the amide bond. This method has been successfully applied to the synthesis of the complex glycopeptide, **44** (Scheme 9). Efforts are currently underway to optimize the post-ligational removal of the auxiliary component.

(3) Development of Non-Cysteine Based Acyl Acceptor Ligation Strategies

While useful as a tool for peptide or glycopeptide ligation, the general application of the above described mercaptobenzyl-based auxiliary strategies is limited by the harsh conditions required for auxiliary removal. An alternative approach to the development of non-cysteine based ligation strategies would make use of a "traceless auxiliary", namely, a thiol-containing amino acid surrogate located at the N-terminus of the acyl acceptor. According to this strategy, the thiol moiety would promote ligation with the requisite acyl donor. Following ligation, the thiol functionality would be removed through desulfurization, to afford the natural amino acid at the site of ligation. With this goal in mind, we recently disclosed the development of a mild and selective thiol reduction protocol. Toward this end, we were drawn to the earlier work of Hoffmann and coworkers, which described a

desulfurization reaction between mercaptan and trialkylphosphite.¹² This reaction was later extended to trialkylphosphines and a radical mechanism was proposed.^{13–17} We were pleased to observe that tris(2-carboxyethyl) phosphine (TCEP) could be used as a phosphine source in our system. TCEP is considered an excellent phosphine source for these purposes, due to its water solubility and compatibility with a range of glycopeptide functionalities. Thus, in the presence of TCEP, *t*BuSH, and water-soluble radical initiator VA-044, we observed nearly quantitative conversion of the cysteine residue into alanine in model substrates.¹⁸ In a key demonstration, our reduction protocol was found to efficiently accommodate the variety of important and potentially labile functionalities found in substrate **45**, including Thz, methionine, Cys(Acm), thioester, and carbohydrate (Scheme 10).¹⁸ This novel reduction protocol represents a significant advance over previously reported metal-based NCL desulfurization methods, which suffer from general incompatibility with thiol containing groups.

This ligation/radical desulfurization sequence was recently applied to the synthesis of a complex target glycopeptide in our laboratory. Thus, as shown in Scheme 11, the target EPO glycopeptide fragment Ala¹-Gly²⁸, incorporating a complex N-linked dodecasaccharide, presented a challenge to standard ligation approaches, due to the absence of a suitably located cysteine residue for NCL, or an appropriate glycine or proline residue for direct condensation. Instead, we decided to target Ala²² as the disconnection site, and sought to employ our ligation/desulfurization strategy to accomplish formal alanine ligation at this site. As shown, peptide **47** and glycopeptide **48**, incorporating a cysteine residue in place of alanine at the N-terminus, were subjected to our standard ligation conditions. The reaction was accompanied by instantaneous thiolactone formation; however, this issue was readily addressed through the addition of the thiolactone opening agent, thiopropionic acid. Upon exposure of **51** to our mild desulfurization conditions, the target glycopeptide **52**, incorporating an alanine residue at the site of ligation, was in hand.¹⁹

The ability to perform glycopeptide couplings at a range of different amino acid residues would, in principle, provide greatly enhanced flexibility in terms of synthetic planning. The development of non-cysteine based ligation capacities is particularly important in the context of our longterm objectives toward synthesizing complex glycoproteins. As part of our effort to extend the above described two-step ligation/reduction strategy to different amino acids, we have demonstrated the ability to achieve formal ligation at valine,²⁰ which is a much more naturally abundant amino acid than is cysteine. There are two possible sites at which to install a temporary thiol group on the valine residue: at the β position or γ position. We envisioned that γ -thiol valine would serve as a more effective valine surrogate. According to our strategy, a synthetic γ -thiol valine would be appended to the N-terminus of peptide **2** (Scheme 12). This peptide was anticipated to undergo ligation with peptide **1**, presenting either ester or thioester acyl donor functionality at the C-terminus. Desulfurization of the ligated intermediate would provide the desired product, incorporating a valine residue at the ligation site.

As shown below, this strategy was effectively reduced to practice. The reaction sequence was found to be quite robust, accommodating a variety of C-terminal residues, including Gln, Phe, Thr, and Pro. In the case of more sterically demanding ligation sites, such as Pro-Val, use of the *p*-nitrophenol ester at the C-terminus of the acyl donor served to greatly enhance ligation efficiency. A representative example is shown in Scheme 13. Thus, glycopeptide **53**, incorporating a C-terminal *ortho*-thiophenyl ester, successfully participated in the ligation reaction with peptide **54**, equipped with the N-terminal valine surrogate, to produce, following desulfurization, the target glycopeptide **56**.²⁰ On the basis of these encouraging findings, we are looking to extend this general idea to the execution of formal

ligation at other amino acids, such as threonine. Our recent advances in the development of a threonine ligation strategy will be reported shortly.

Conclusions

We have described herein the development of recent peptide ligation methodologies which have emerged from our group in the context of our program directed toward the chemical synthesis of complex glycopeptide and glycoprotein target systems. These methods not only enable us to construct biologics in a convergent and efficient manner; but also change chemists' vision in terms of building block designing and synthetic strategy planning. In closing, we note that these methodological advances will enable us to progress toward our goal of building glycopeptides and glycoproteins which possess extraordinary biological activity. These methods are expected to be of broad value to the synthetic community at large.

References

1. Merrifield RB. J. Am. Chem. Soc. 1963; 85:2149.
2. Dawson PE, Muir TW, Clark-Lewis I, Kent SBH. Science. 1994; 266:776. [PubMed: 7973629]
3. Warren JD, Miller JS, Keding SJ, Danishefsky SJ. J. Am. Chem. Soc. 2004; 126:6576. [PubMed: 15161285]
4. Chen G, Warren JD, Chen J, Wu B, Wan Q, Danishefsky SJ. J. Am. Chem. Soc. 2006; 128:7460. [PubMed: 16756298]
5. Blake J. Int. J. Pept. Protein Res. 1981; 17:273. [PubMed: 7228502]
6. Aimoto S, Mizoguchi N, Hojo H, Yoshimura S. Bull. Chem. Soc. Jpn. 1989; 62:524.
7. Chen G, Wan Q, Tan Z, Kan C, Hua Z, Ranganathan K, Danishefsky SJ. Angew. Chem. Int. Ed. 2007; 46:7383.
8. Bodanszky M. Nature. 1955; 175:685. [PubMed: 14370194]
9. Wan Q, Chen J, Yuan Y, Danishefsky SJ. J. Am. Chem. Soc. 2008; 130:15814. [PubMed: 18855357]
10. Yuan Y, Chen J, Wan Q, Tan Z, Chen G, Kan C, Danishefsky SJ. J. Am. Chem. Soc. 2009; 131:5432. [PubMed: 20560636]
11. Wu B, Chen J, Warren JD, Chen G, Hua Z, Danishefsky SJ. Angew. Chem. Int. Ed. 2006; 45:4116.
12. Hoffmann FW, Ess RJ, Simmons TC, Hanzel RS. J. Am. Chem. Soc. 1956; 78:6414.
13. Walling C, Rabinowitz R. J. Am. Chem. Soc. 1957; 79:5326.
14. Walling C, Basedow OH, Savas ES. J. Am. Chem. Soc. 1960; 82:2181.
15. Gonzalez A, Valencia G. Tetrahedron: Asymmetry. 1998; 9:2761.
16. Cuesta J, Arsequell G, Valencia G, Gonzalez A. Tetrahedron: Asymmetry. 1999; 10:2643.
17. Arsequell G, Gonzalez A, Valencia G. Tetrahedron Lett. 2001; 42:2685.
18. Wan Q, Danishefsky SJ. Angew. Chem. Int. Ed. 2007; 46:9248.
19. Kan C, Trzupek JD, Wu B, Wan Q, Chen G, Tan Z, Yuan Y, Danishefsky SJ. J. Am. Chem. Soc. 2009; 131:5438. [PubMed: 19334679]
20. Chen J, Wan Q, Yuan Y, Zhu JL, Danishefsky SJ. Angew. Chem. Int. Ed. 2008; 47:8521.

Scheme 1.

Scheme 2.

Scheme 3.

Scheme 4.

Scheme 5.

Scheme 6.

Scheme 7.

Scheme 8.

Scheme 9.

NIH-PA Author Manuscript

Scheme 11.

Scheme 12.