Angewandte International Edition Chemie

Glycoproteins

DOI: 10.1002/anie.201207906

Total Synthesis of Erythropoietin through the Development and Exploitation of Enabling Synthetic Technologies

*Richard J. Payne**

chemical ligation · glycopeptides · glycoproteins · glycosylation · total synthesis

Glycosylation is the most common co- or posttranslational modification of proteins, with 50–70% of all human proteins estimated to bear covalently linked glycans.[1] Glycosylation of proteins is implicated in a diverse number of processes including fertilization, immune surveillance, hormone activity, neuronal development, and protein folding.[2] Despite the unquestionable importance of glycosylation, progress towards understanding the role of this modification on both the structure and function of proteins has been hampered by difficulties in accessing them in pure form. Specifically, these issues arise from the fact that, unlike protein synthesis, the glycosylation process is not templated. Rather, it is controlled by the activities of glycosyltransferases that give rise to mixtures of glycoforms. As these glycoforms differ only in the nature of covalently appended glycan(s), they are often impossible to separate using chromatographic methods. Recombinant human erythropoietin (rhEPO) a multi-billion dollar drug for the treatment of anemia is a 166 amino acid protein bearing four glycosylation sites, three highly variable N-linked (at Asn-24, Asn-38, and Asn-83) and one more conserved O-linked (at Ser-126). As a consequence of its production in mammalian cells, rhEPO is sold as a complex mixture of glycoforms, and the glycosylation state of EPO that exhibits optimal activity remains unknown.

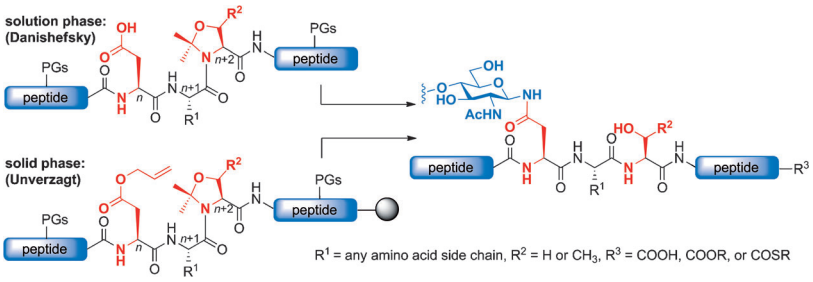
Wild-type EPO, in which all four glycosylation sites are occupied with glycans, is the proverbial "holy grail" of synthetic targets and, as such, has been the subject of intense research efforts. There have been a number of elegant syntheses of non-wild-type analogues of EPO[3] and, recently, the Kajihara group disclosed the synthesis of full-length EPO bearing one complex sialylated biantennary glycan, albeit with the introduction of two amino acid mutations (Glu21Ala and Gln78Ala) to aid in the synthesis.[4] Despite these synthetic feats, access to a pure glycoform of EPO bearing all four glycans has remained elusive.

In practice, accessing a molecule with the inherent complexity of EPO could be accomplished by a large team of synthetic chemists using currently available technologies; however, the philosophy must be to develop and employ enabling technologies that can be used simply and efficiently to produce any glycoprotein in pure form for biological study. Synthetic methods such as solid-phase peptide synthesis (SPPS) and native chemical ligation fit such criteria and have been used extensively for the chemical synthesis of hundreds of proteins. The lack of robust and operationally simple methods for installing glycans into peptides is a major impediment for the rapid assembly of glycoproteins such that progress in this area has lagged behind that of unglycosylated proteins.

Recently the Danishefsky[5] and Unverzagt[6] groups independently reported a new synthetic strategy for the simple and efficient preparation of N-linked glycopeptide fragments. In a breakthrough study, the Danishefsky group subsequently used this strategy en route to the first total synthesis of homogeneous synthetic EPO bearing four glycans.[7] This Highlight aims to outline the new methodology as well as this landmark achievement in glycoprotein total synthesis.

Traditionally, N-linked glycopeptides are prepared using two methods: 1) linear incorporation of preformed glycosylamino acid "cassettes" into the growing resin-bound peptide or 2) convergent solution- or solid-phase aspartylation of glycosylamines by specific aspartic acid (Asp) residues within a peptide sequence.[1] The major disadvantages of the "cassette" strategy are the increased expenditure of the precious glycosylated building blocks over iterative steps and the reduced efficiency of coupling reactions because of the steric bulk of larger glycans. The major pitfall of the aspartylation method is the extensive formation of aspartimide by-products during activation of the Asp side chain which is often more rapid than formation of the desired amide linkage, especially for larger glycans. The new strategy reported by the groups of Danishefsky and Unverzagt is elegant in its simplicity. The authors demonstrate that the use of a *gem*-dimethyloxazolidine ring (otherwise known as a pseudoproline (ΨPro) residue) as a conformational constraint of X-serine(Ser)- and X-threonine(Thr)-containing dipeptide units found within the Asn-X-Ser/Thr consensus sequence (present in all N-linked glycoproteins) efficiently suppresses the formation of aspartimides upon reaction of an

[*] Dr. R. J. Payne
School of Chemistry, The University of Sydney
NSW 2006 (Australia)
E-mail: richard.payne@sydney.edu.au
Homepage: http://sydney.edu.au/science/chemistry/~payne

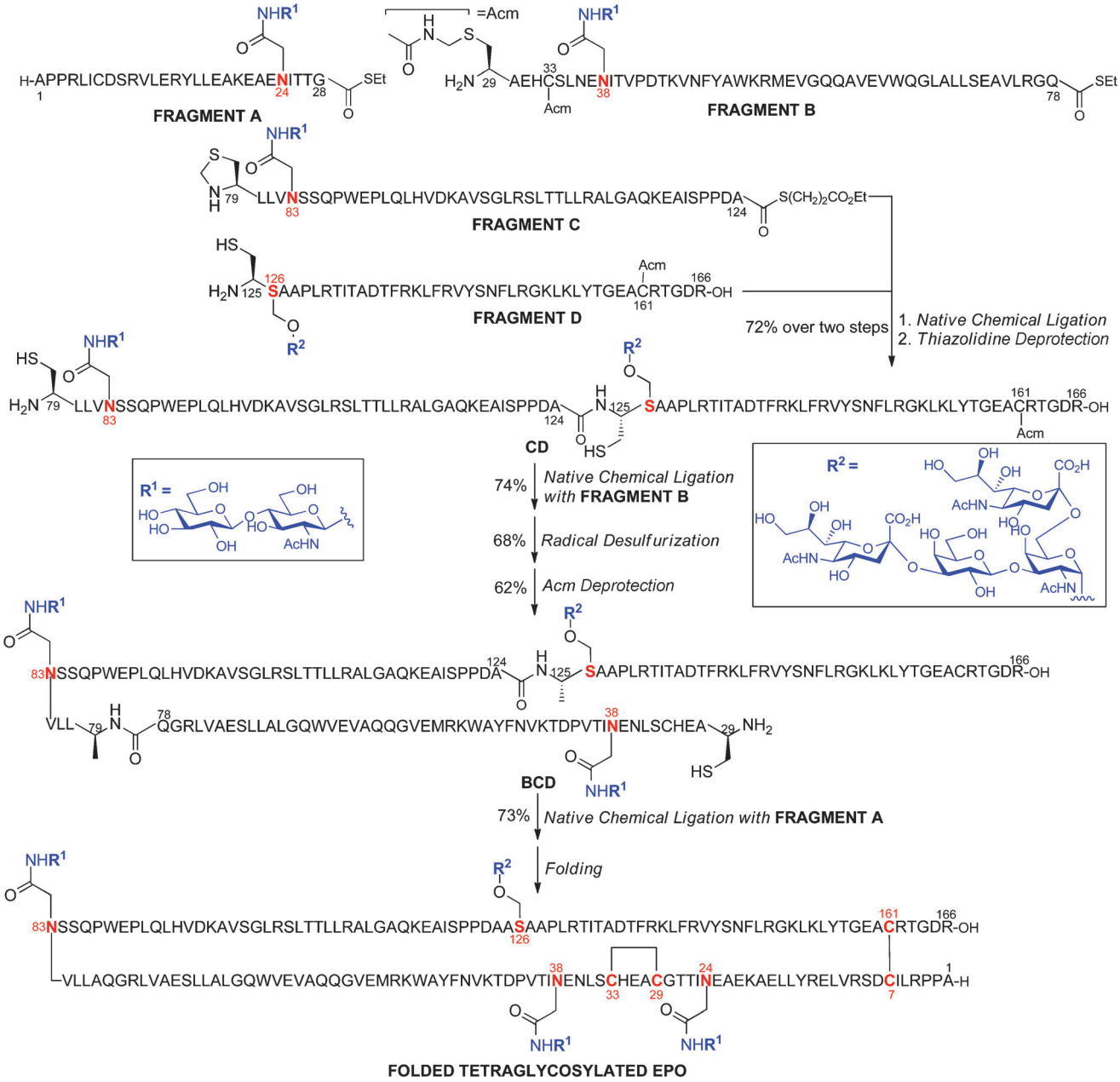


Scheme 1. Convergent synthesis of N-linked glycopeptides via ψPro-derived peptides. PGs = protecting groups.

Asp side chain with a glycosylamine (Scheme 1). The authors show the utility of this method through the preparation of a wide range of glycopeptides and glycopeptide thioesters (including several fragments of EPO) both in solution and on the solid phase. It should be noted that the generation of glycopeptide thioesters by this strategy is particularly powerful given the use of these building blocks as acyl donors in ligation chemistry. The Danishefsky group generated glycopeptides and glycopeptide thioesters in solution through a coupling reaction between free Asp side chains on otherwise completely side-chain-protected peptides bearing ΨPro-derived serine or threonine residues at the $n+2$ site.[5] Acidic deprotection then provided the desired N-glycosylated targets. Using the allyl ester protecting group on the side chain of Asp and a ΨPro-derived serine or threonine residue at the $n+2$ site, Unverzagt and co-workers were able to selectively unmask the side chain of Asp followed by on-resin aspartylation of a glycosylamine.[6] Glycosylamines ranging in size from monosaccharides up to dodecasaccharides were introduced by both groups, and although yields for the larger oligosaccharides dropped to 20–30%, little to no aspartimides were formed in all cases. Presumably, conversion of the normally *trans*-X-Ser/Thr bond to the *cisoid* geometry results in an unfavorable local conformation which lowers the propensity for nucleophilic attack of the backbone amide onto the activated side chain of the Asp. One of the enormous advantages of this methodology, in addition to the suppression of aspartimides, is that preformed Ser- and Thr-derived ΨPro dipeptides are commercially available (used as traceless aggregation suppressants) and, as such, peptide precursors can be simply prepared by standard automated SPPS techniques. In addition, glycosylamines can be prepared directly from the corresponding reducing sugars, many of which are commercially available, using well-established chemistry. Given the operationally simple nature of this method it should become widely adopted by researchers.

Danishefsky and co-workers highlighted the utility of the methodology in the synthesis of three N-linked glycopeptide fragments of EPO which were then assembled into wild-type EPO bearing four glycan modifications.[7] The disconnection approach adopted involved the sole use of the venerable native chemical ligation reaction.[8] The group rationalized that the native Cys residues in EPO (at positions 7, 29, 33, and 161) were not ideally located to enable assembly of the linear structure and therefore opted to introduce unnatural Cys residues into the sequence. The group took advantage of the native chemical ligation–desulfurization manifold reported by Yan and Dawson in which Cys residues can be desulfurized to afford alanine (Ala) residues following the ligation event.[9] The presence of 18 Ala residues in EPO provided ample disconnection sites. The authors chose to disconnect the linear sequence between Gly-28 and Cys-29, Gln-78 and Ala-79, and Ala-124 and Ala-125, leading to four target fragments **A**–**D** which in turn were assembled in the C- to N-terminal direction through three sequential native chemical ligation events (Scheme 2). The native Ala-79 was replaced by a thiazolidine residue in fragment **C** while Ala-125 was replaced with a Cys residue in fragment **D**. Cys-33 and Cys-161 were protected with acetamidomethyl (Acm) groups to enable selective desulfurization of the unnatural Cys residues using the metal free dethiylation (MFD) methodology developed by Wan and Danishefsky.[10] Fragments **A**–**C**, each possessing an N-linked glycosylation site, were prepared from ΨPro-derived peptides using the convergent solution-phase aspartylation methodology. The authors incorporated the simplified chitobiose disaccharide in place of the complex glycan that would normally be present in endogeneous EPO or rhEPO. Using a stepwise, linear assembly approach, fragments **C** and **D** were first ligated under standard native chemical ligation conditions followed by demasking of the thiazolidine unit to generate glycopeptide **CD**. Subsequent ligation of **CD** with the glycopeptide thioester **B**, followed by radical desulfurization of the free Cys residues (at positions 79 and 125) and Acm deprotection furnished **BCD**. Finally, ligation of **BCD** with glycopeptide thioester **A** provided the complete unfolded linear sequence of EPO (**ABCD**) bearing three N-linked chitobiose units and one O-linked tetrasaccharide moiety. An elegant extension of this work was the subsequent preparation of EPO through a kinetically controlled one-pot ligation approach, this time in the N- to C-terminal direction. In this instance, modified variants of glycopeptide thioester fragments **A** and **B** were prepared and ligated to leave a C-terminal ethyl thioester moiety at Gln-78 which could be reacted directly with fragment **CD** to generate full-length EPO rapidly and in high yield. Desulfurization and Acm deprotection then provided unfolded EPO, and the full-length glycoprotein was successfully folded. This synthetic tetraglycosylated EPO was found to elicit erythrocyte production in vitro and its activity was superior to that of a synthetic EPO aglycone (prepared by the linear approach). It is not yet clear how the activity of this single EPO glycoform compares to that of rhEPO which contains more complex N-linked glycans; however, this will no doubt be the subject of future investigations now that the challenge of obtaining a single glycosylated form has been overcome.

In summary, access to N-linked glycopeptides should be greatly accelerated through the convergent aspartylation strategy developed by the Danishefsky and Unverzagt groups, as demonstrated in the total chemical synthesis of EPO. This methodology should find wide application as an enabling tool for the synthetic generation of a whole range of glycoproteins for detailed biological studies. This includes access to the much anticipated wild-type EPO bearing three complex



Scheme 2. Total synthesis of EPO bearing four glycans using native chemical ligation.

biantennary or triantennary N-linked glycans which will no doubt serve as the next milestone in glycoprotein synthesis.

Received: October 1, 2012
Published online: November 23, 2012

[1] a) D. P. Gamblin, E. M. Scanlan, B. G. Davis, *Chem. Rev.* **2009**, *109*, 131 – 163; b) R. J. Payne, C. H. Wong, *Chem. Commun.* **2010**, *46*, 21 – 43.
[2] A. Varki, *Glycobiology* **1993**, *3*, 97 – 130.
[3] a) G. G. Kochendoerfer et al., *Science* **2003**, *299*, 884 – 887; b) S. H. Liu, B. L. Pentelute, S. B. H. Kent, *Angew. Chem.* **2012**, *124*, 1017 – 1023; *Angew. Chem. Int. Ed.* **2012**, *51*, 993 – 999; c) K. Hirano, D. Macmillan, K. Tezuka, T. Tsuji, Y. Kajihara, *Angew. Chem.* **2009**, *121*, 9721 – 9724; *Angew. Chem. Int. Ed.* **2009**, *48*, 9557 – 9560.
[4] M. Murakami, R. Okamoto, M. Izumi, Y. Kajihara, *Angew. Chem.* **2012**, *124*, 3627 – 3632; *Angew. Chem. Int. Ed.* **2012**, *51*, 3567 – 3572.
[5] P. Wang, B. Aussedat, Y. Vohra, S. J. Danishefsky, *Angew. Chem.* **2012**, *124*, 11739 – 11743; *Angew. Chem. Int. Ed.* **2012**, *51*, 11571 – 11575.
[6] V. Ullmann, M. Rädisch, I. Boos, J. Freund, C. Pöhner, S. Schwarzinger, C. Unverzagt, *Angew. Chem.* **2012**, *124*, 11734 – 11738; *Angew. Chem. Int. Ed.* **2012**, *51*, 11566 – 11570.
[7] P. Wang, S. Dong, J. A. Brailsford, K. Iyer, S. D. Townsend, Q. Zhang, R. C. Hendrickson, J. Shieh, M. A. S. Moore, S. J. Danishefsky, *Angew. Chem.* **2012**, *124*, 11744 – 11752; *Angew. Chem. Int. Ed.* **2012**, *51*, 11576 – 11584.
[8] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. H. Kent, *Science* **1994**, *266*, 776 – 779.
[9] L. Z. Yan, P. E. Dawson, *J. Am. Chem. Soc.* **2001**, *123*, 526 – 533.
[10] Q. Wan, S. J. Danishefsky, *Angew. Chem.* **2007**, *119*, 9408 – 9412; *Angew. Chem. Int. Ed.* **2007**, *46*, 9248 – 9252.

