From EPOthilone to EPO: A challenge for natural product synthesis

Zachary Z. Brown and Tom W. Muir[1]
Frick Chemistry Laboratory, Princeton University, Princeton, NJ 08544

Activate. Couple. Deprotect. Repeat. So goes the cycle of peptide synthesis, whereby amino acids are iteratively connected to produce oligopeptides. However, only peptides that are fewer than approximately 50 residues can typically be accessed through solid-phase peptide synthesis. Thus, a defining moment in the field of peptide synthesis came about with the advent of native chemical ligation (NCL), a mild and selective reaction that allows two unprotected polypeptides to be joined together via a peptide bond (1). This convergent synthesis approach significantly expanded the accessible chemical space of peptide chemistry, and advances have continued to push the field forward to access even larger protein constructs. In PNAS, Brailsford and Danishefsky (2) use a number of these methodological developments and detail their synthetic adventures to produce a nonglycosylated form of erythropoietin (EPO).

The protein of interest is EPO, a 166-aa glycoprotein that serves as the main hormonal regulator of red blood cell production. EPO contains three N-linked glycans (attached to asparagine residues 24, 38, and 83) and a single O-linked glycan at serine 126. The hormone is therapeutically administered to treat anemia that can result from kidney failure, cancer, and chemotherapy. Commercial overexpression of EPO results in a heterogeneous mixture of glycoforms (various sugar appendages that are not necessarily the canonical modifications), and the individual therapeutic contributions of each form are not known. Indeed, one study showed that enzymatic removal of N-linked glycans diminished in vivo activity but not in vitro activity (3). Other studies have found that increasing the sialic acid content through cell engineering techniques leads to advantageous in vivo therapeutic properties (4). Access to homogeneous samples of glycoproteins would lead to a better understanding of the individual roles of the glycans in tuning the folding and function of EPO, perhaps opening the way to more potent analogs. In principle, protein synthesis through NCL is perfectly suited to this problem. However, the synthetic challenges associated with the synthesis of a complex glycoprotein such as EPO are manifold, even with the use of state-of-the-art NCL methodology. Nonetheless, Brailsford and Danishefsky (2) embarked on this ambitious synthetic expedition several years ago, and, in this installment of the journey, report on the successful assembly of the full-length unmodified protein (i.e., without the appended glycosyl groups). With this, they have passed an important milestone.

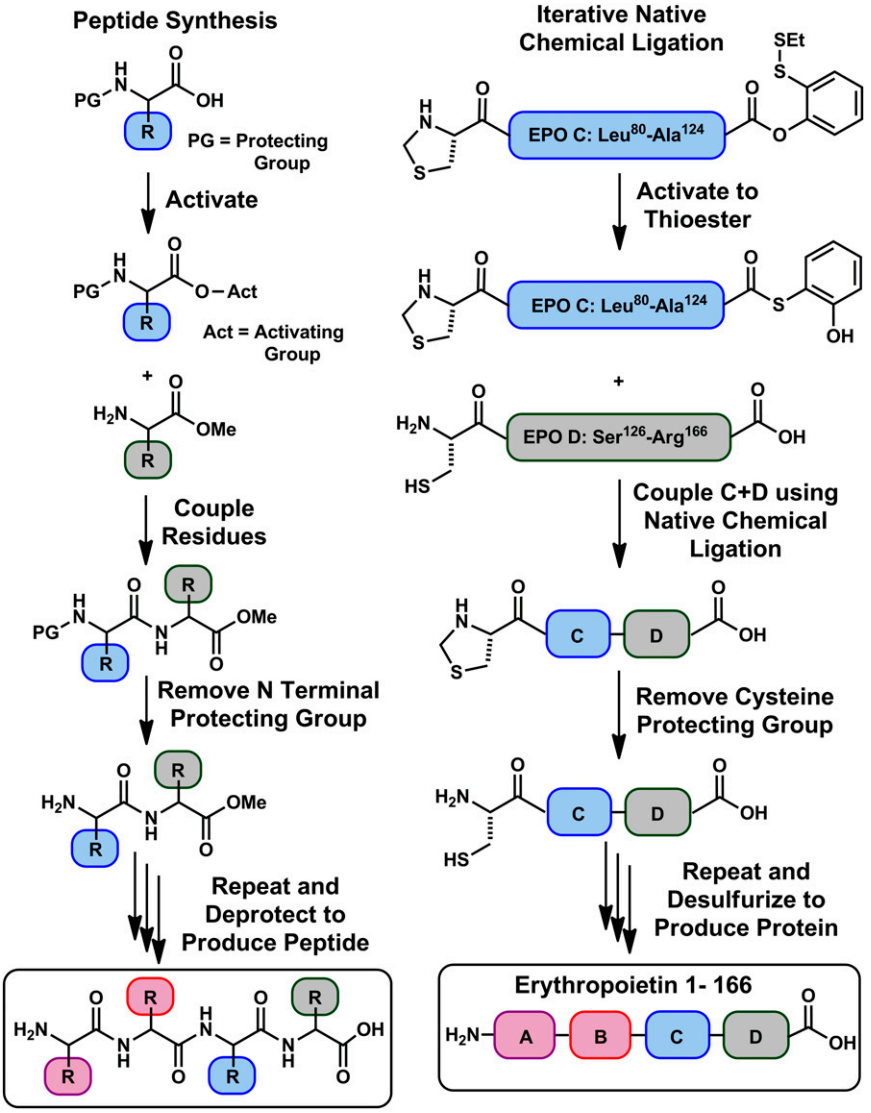

Fig. 1. The logical flow of peptide synthesis (*Left*) compared with Iterative NCL for the synthesis of EPO (*Right*). The sequential steps of activation of the *C* terminus of one building block, coupling of the two fragments, and removing the next protecting group are repeated to assemble peptides (*Left*) or proteins (*Right*).

Brailsford and Danishefsky (2) exploit a number of synthetic advances toward the creation of synthetic proteins, highlighting how far the field has come since the introduction of NCL some 18 years ago (5). The methodology they use demonstrates how a toolbox of useful reactions for the generation, manipulation, and elaboration of large peptides has been developed. These include the use of semipermanent protecting groups and active esters, tactics that are routinely used

Author contributions: Z.Z.B. and T.W.M. wrote the paper.
The authors declare no conflict of interest.

See companion article on page 7196.

[1]To whom correspondence should be addressed. E-mail: muir@princeton.edu.

www.pnas.org/cgi/doi/10.1073/pnas.1205667109

in the assembly of peptides from amino acid monomers, but that have now been adapted to the assembly of proteins from peptide building blocks (Fig. 1). Indeed, the methodology has been developed to the point that consecutive ligations can be performed, leading to assembly of much larger protein constructs. Extending the analogy to peptide synthesis, iterative NCL reactions have even been performed on a solid support (6). Arguably one of the most impactful developments in NCL methodology in the past decade was the advent of the desulfurization technique (7). In this strategy, a thiol containing NCL auxiliary mediates the desired transthioesterification reaction and subsequent acyl transfer to yield the native peptide bond. Having performed this vital role, the sulfur atom is then selectively removed by using mild conditions that only erase primary thiol groups. This concept freed NCL from the requirement of cysteine at the ligation junction. EPO is an attractive target for this methodology because, like many proteins, the endogenous cysteines are not in an optimal location for the necessary ligation reactions (i.e., in the interior of the protein and evenly distributed). One common implementation of desulfurization is to substitute a native alanine for a cysteine, proceed with the NCL, and then desulfurize the cysteine back to alanine. Notable developments in desulfurization include a radical-based procedure introduced by Wan and Danishefsky (8), rather than the initial hydrogenation chemistry (7), as a more convenient implementation. The tremendous success of the "alanine equals cryptic cysteine" logic in NCL has fueled the development of other desulfurization schemes, which now extend to other amino acids such as phenylalanine, leucine, valine, and lysine (9). The upshot of all this is that it is now much easier to find a suitable ligation spot in a protein target.

The synthetic route developed by Brailsford and Danishefsky (2) shows how the desulfurization methodology can assist in creating completely native backbone structures. It is worth noting, however, that other analogs of EPO have been previously constructed through NCL-based strategies. Kochendoerfer et al. (10) used a strategy whereby endogenous glutamic acid residues were replaced with cysteines, which were then used as ligation junctions (positions 89 and 117). Following ligation, alkylation with bromoacetic acid yielded structural analogues of the native glutamate side chains, which differed only slightly by having an additional sulfur inserted between the two methylene carbons of the side chains. This chemistry was invaluable to access the backbone of EPO, and then to further derivatize the protein with discreet polyethylene-based oligomers that replaced the glycosyl groups. More recently, Hirano et al. (11) prepared a glycosylated EPO analogue by using a semisynthetic approach that combined recombinant and synthesis polypeptide fragments (conventionally termed "expressed protein ligation"). By using this methodology, they were able to synthesize an EPO construct with homogenous yet complex sialyloligosaccharides that were derived from biological origin. With these nonnative saccharide groups, they were also able to fold their EPO construct into the correct 3D structure and show that it had biological activity similar to recombinant EPO. As evidenced by this semisynthetic work, as well as several other recent glycoprotein syntheses (12, 13), the field of glycoprotein synthesis is certainly gaining momentum.

As already noted, the Danishefsky group (14) have developed a vibrant program toward the synthesis of complex glycoproteins. This effort has yielded several notable developments, as so often happens during the course of natural product synthesis; a classical example of this would be in the total synthesis of the eponymous epothilones of the title, by this same group (15). One of these contributions, used in the current study by Brailsford and Danishefsky (2), was that of the latent thioester, whereby unmasking of an *ortho*-thiophenolic ester derivative rearranges to produce the thioester that is a required component of NCL. To accommodate the active ester and to chemically distinguish between the various cysteines (i.e., ones used for ligation that will be desulfurized and endogenous residues that need to be retained), different thiol protecting groups were used. The naturally occurring cysteines were masked with acetamidomethyl protecting groups, which will be stable to all the conditions used throughout the assembly (including desulfurization). In contrast, the N-terminal cysteines of the peptide building blocks, which are the ligation handles, are temporarily protected as thiazolidine rings (the 5-membered *N*,*S* heterocycle) and then selectively exposed when needed. These advances highlight how prominent the role of cysteine is in modern protein chemistry, with the selective unmasking of the thiol group at various stages made possible by the advent of different protecting groups.

After successful assembly of the complete backbone through iterative NCL, certainly an achievement in its own right, Brailsford and Danishefsky (2) encountered severe solubility problems with this advanced intermediate. Although the polypeptide could be desulfurized, the protecting groups on the endogenous cysteines could not be removed because of the unfavorable physical properties of the molecule. This finding hints at how intricately the glycosyl groups might be intertwined with the native structure and behavior of the protein, such that, in the absence of the glycans, the protein is difficult to handle. If correct, this explanation would be another example of how glycosyl modifications can assist in protein folding and stabilize certain protein structures. Thus, it will be fascinating to see if subsequent synthetic endeavors from this group, which include various glycan frameworks, produce a more soluble protein.

In sum, the report by Brailsford and Danishefsky (2) is a tour de force of modern protein synthesis and is an impressive example as to how far the methodology for the ligation of peptide segments has evolved. Further advancements of NCL methodology will continue to push the field toward the synthesis of larger and more complex protein constructs. This will allow unprecedented investigation as to how modifications such as glycosylation are able to modulate protein function.

1. Dawson PE, Muir TW, Clark-Lewis I, Kent SBH (1994) Synthesis of proteins by native chemical ligation. *Science* 266:776–779.
2. Brailsford JA, Danishefsky SJ (2012) Probing the stability of nonglycosylated wild-type erythropoietin protein via reiterative alanine ligations. *Proc Natl Acad Sci USA* 109:7196–7201.
3. Higuchi M, et al. (1992) Role of sugar chains in the expression of the biological activity of human erythropoietin. *J Biol Chem* 267:7703–7709.
4. Egrie JC, Browne JK (2001) Development and characterization of novel erythropoiesis stimulating protein (NESP). *Br J Cancer* 84(suppl 1):3–10.
5. Kent SBH (2009) Total chemical synthesis of proteins. *Chem Soc Rev* 38:338–351.
6. Cotton GJ, Muir TW (2000) Generation of a dual-labeled fluorescence biosensor for Crk-II phosphorylation using solid-phase expressed protein ligation. *Chem Biol* 7:253–261.
7. Yan LZ, Dawson PE (2001) Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurization. *J Am Chem Soc* 123:526–533.
8. Wan Q, Danishefsky SJ (2007) Free-radical-based, specific desulfurization of cysteine: A powerful advance in the synthesis of polypeptides and glycopolypeptides. *Angew Chem Int Ed Engl* 46:9248–9252.
9. Dawson PE (2011) Native chemical ligation combined with desulfurization and deselenization: A general strategy for chemical protein synthesis. *Isr J Chem* 51: 862–867.
10. Kochendoerfer GG, et al. (2003) Design and chemical synthesis of a homogeneous polymer-modified erythropoiesis protein. *Science* 299:884–887.
11. Hirano K, Macmillan D, Tezuka K, Tsuji T, Kajihara Y (2009) Design and synthesis of a homogeneous erythropoietin analogue with two human complex-type sialyloligosaccharides: Combined use of chemical and bacterial protein expression methods. *Angew Chem Int Ed Engl* 48:9557–9560.
12. Sakamoto I, et al. (2012) Chemical synthesis of homogeneous human glycosyl-interferon-β that exhibits potent antitumor activity in vivo. *J Am Chem Soc* 134:5428–5431.
13. Piontek C, et al. (2009) Semisynthesis of a homogenous glycoprotein enzyme: Ribonuclease C: Part 2. *Angew Chem Int Ed* 48:1941–1945.
14. Kan C, Danishefsky SJ (2009) Recent departures in the synthesis of peptides and glycopeptides. *Tetrahedron* 65:9047–9065.
15. Balog A, et al. (1996) Total synthesis of (-)-epothilon A. *Angew Chem Int Ed Engl* 35:2801–2803.