Published in final edited form as:

Science. 2013 December 13; 342(6164): 1357–1360. doi:10.1126/science.1245095.

Erythropoietin Derived by Chemical Synthesis

Ping Wang^{#1}, Suwei Dong^{#1}, Jae-Hung Shieh², Elizabeth Peguero², Ronald Hendrickson³, Malcolm A. S. Moore², and Samuel J. Danishefsky^{1,4,†}

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

²Cell Biology Program, Sloan-Kettering Institute for Cancer Research, 1275 York Avenue, New York, NY 10065, USA

³Department of Pharmacology and Chemistry, Sloan-Kettering Institute for Cancer Research, 1275 York Avenue, New York, NY 10065, USA.

⁴Department of Chemistry, Columbia University, Havemeyer Hall, 3000 Broadway, New York, NY 10027, USA

[#] These authors contributed equally to this work.

Abstract

Erythropoietin is a signaling glycoprotein that controls the fundamental process of erythropoiesis, orchestrating the production and maintenance of red blood cells. As administered clinically, erythropoietin has a polypeptide backbone with complex dishomogeneity in its carbohydrate domains. Here we describe the total synthesis of homogeneous erythropoietin with consensus carbohydrate domains incorporated at all of the native glycosylation sites. The oligosaccharide sectors were built by total synthesis and attached stereospecifically to peptidyl fragments of the wild-type primary sequence, themselves obtained by solid-phase peptide synthesis. The glycopeptidyl constructs were joined by chemical ligation, followed by metal-free dethylation, and subsequently folded. This homogeneous erythropoietin glycosylated at the three wild-type aspartates with N-linked high-mannose sialic acid-containing oligosaccharides and O-linked glycophorin exhibits Procrit-level in vivo activity in mice.

Erythropoietin (EPO) is functionally a hormone that plays a key role in erythropoiesis, which is central to the orchestration and production of erythrocytes (1, 2). EPO is a glycoprotein with both a highly conserved polypeptide domain and four tightly conserved sites of glycosylation. However, the actual carbohydrate ensembles, particularly at the N linkages, are strikingly variable (3). Before our earlier work (4, 5), there were no reports of a naturally isolated or synthetically derived structurally defined, homogeneously glycosylated, EPO sample of wild-type primary structure—bearing glycosyl domains at the conserved sites.

[†]Corresponding author. s-danishefsky@ski.mskcc.org.

Supplementary Materials

www.sciencemag.org/content/342/6164/1357/suppl/DC1 Materials and Methods Figs. S1 to S27 References (40-42)

The defining goal that we undertook in 2002 was to prepare, by total synthesis, a “wild-type” EPO polypeptide homogeneously glycosylated at the three conserved N-linked sites, as well as at the single O-linked center, by oligosaccharides of biolevel complexity (Fig. 1) (6). All of the carbohydrate domains would also be prepared through total synthesis (7-10). Given the wide-spread appearance of glycophorin (2, structure **S1**) in O linkages in complex EPO-type glycoproteins, we chose to install this motif at Ser¹²⁶ in our envisioned synthetic EPO (11). As for the three asparagine-linked domains, we designed what was perceived to be a consensus sequence of realistic EPO-level complexity (12-14). The N-linked glycan domain would display a characteristic chitobiose disaccharide at its reducing end, which would bear a signature branching β -linkage to L-fucose and a core branched trimannosyl ensemble. The latter would in turn be linked at C2 and C2' of its “wing” mannose residues to two identical trisaccharide sialyllactose domains. The sense of sialylation (α -2,3) of the lactosamine would correspond to the wild-type recombinant motif (15, 16). This thinking led us to propose structure **S2** as the consensus dodecasaccharide (fig. S1).

To undertake this program, four synthetic methodology entries were crucial to our endeavors, including native chemical ligation (NCL), metal-free desulfurization (MFD), o-mercaptoaryl ester rearrangement (OMER)-mediated ligation, and one-flask aspartylation (fig. S2). Developed by Kent and co-workers (17), NCL serves as a key technology in complex peptide synthesis. However, we could take advantage of only one NCL to affix an N-terminal cysteine to a suitable C-terminal thioester donor. Our contribution for extending the range of NCL logic to enable ligations at ultimately noncysteine N termini was that of MFD (18), which has allowed for the implementation of an earlier recognition by Dawson that MFD of a cysteine can lead to an alanine (19). Despite a dearth of useful cysteine ligation sites, the primary structure of target EPO glycoform **1** offers a variety of alanine centers, thereby inviting a range of retrosynthetic options for reaching our goal. OMER-mediated ligation served as a device for in situ generation of the activated thioesters (20, 21) and proved to be particularly useful in the synthesis of fragment **S7** (fig. S4) (22). Finally, and critical to our mission, was the capacity to realize maximally convergent amidation of highly complex oligosaccharide glycosylamines with suitably differentiated aspartates, even in substantially sized polypeptides (23, 24).

For programmatic expediency, we first field-tested our methodological capabilities and the implementability of our retrosynthetic logic in the context of a simpler target, where the three N linkages serve to join chitobiose disaccharides to the conserved asparagines of the wild-type 166-oligomer (4, 5). Naively, we thought at the time that the pathway to **S3**, if followed faithfully (with the highly challenging provision that the N-linked oligosaccharides corresponding to **S2** would be in place), would soon lead us to our long-term, noncompromisable target, **1**. We were able to synthesize the N-linked fragments **S4**, **S5**, and **S6** (fig. S4A) containing the fully synthetic dodecamer polysaccharide moieties.

With the required fragments in hand, we proceeded toward **1** (Fig. 1). The coupling of glycopeptides **S6** and **S7** under NCL conditions followed by deprotection of the thiazolidine (Thz) (25) afforded a prospective EPO (79-166) domain (**S8**) in good yield (fig. S4B). Unfortunately, the coupling of **S8** to prospective EPO fragment **S5**, intended to produce the prospective EPO (29-166) domain (**S9**), was not successful (at best, 10%). After numerous

investigations, it was surmised that the breakdown in the attempted merger of **S8** and **S5** arose from the large consensus carbohydrate domain at position 83 in fragment **S8**.

Accordingly, we had to reconfigure the erstwhile fragments **S5** and **S6**, generating in their wake more manageable NCL-worthy prealanine coupling domains for filling in the space between amino acids 29 to 124. We reorganized the domain associated with pre-EPO (29-124) into three fragments, shown in Fig. 2 as building blocks **2**, **3**, and **4** (26). Thus, following the one-flask convergent aspartylation protocol, glycopeptide fragments EPO (29-59) (**2**) and EPO (60-97) (**3**) were prepared in good overall yields. In this manner, the previously troublesome Asn⁸³-glycosylated fragment was reengineered so that the consensus carbohydrate domain was placed at a distance from the terminus involved in the required ligations.

Having accomplished the syntheses of pre-EPO (29-124) fragments **2** to **4**, we proceeded toward their assembly. We began by extending glycopeptide **S7** to create the EPO (98-166) domain, which upon Thz deprotection afforded N-terminal ligation partner **5** (Fig. 3). Subsequent ligation between **5** and **3** (bearing a C-terminal lysine thioester donor at position 97) proceeded smoothly. This coupling was followed again by Thz deprotection to provide **6**, encompassing amino acids 60-166 of our projected EPO glycoprotein. Glycopeptide **6** showed much-improved reactivity in the ligation with EPO (29-59) domain (**2**) and afforded the corresponding ligation product **7** in excellent conversion (85%). These observations indicate that the large carbohydrate domain may block ligation because of sheer steric bulk. Alternatively, it may alter the conformation in the N-terminal region of **S8** in such a fashion as to hide its crucial Cys⁷⁹ ligation machinery. These results further underscore the synthetic challenges associated with the preparation of glycopeptides incorporating multiple complex-type oligosaccharides and suggest opportunities in retrosynthetic analysis for building multifaceted biologics in the laboratory.

What we perceived as the final step in the synthesis would be the only NCL that would actually result in a surviving cysteine residue. For this to be possible, it was first necessary to achieve four concurrent MFD transformations to reveal the required alanines (from their erstwhile cysteine precursors) before exposing the prospective surviving cysteines at positions 29, 33, and 161 in the large 29-166 subunit. This task was accomplished as shown in Fig. 3 (**8**). After smooth cleavage of the Ac_m protecting groups at cysteines 29, 33, and 161 using silver acetate (AgOAc) (**9**) (27), the stage for the final ligation was in place. Fortunately, the projected NCL was smoothly accomplished (~80 to 85% conversion), thereby providing **10**, the tetrahydro (unfolded) precursor of **1** (28, 29).

Compound **10** exhibited enhanced stability relative to its glycotruncated counterpart **11** (the unfolded tetrahydro version of **S3**; Fig. 3, inset). For instance, NCL-generated **10** survived lyophilization of solvent, whereas under comparable treatment, at least in our hands, **11** suffered considerable aggregation during the lyophilization and thus required resolubilization before folding. However, even in the case of **10**, storage caused attrition of the material. Hence we proceeded rapidly to its twofold oxidation/folding in a dialysis cassette using cysteine/cystine as the redox shuffling agent (30). After gradual dilution of the folding buffer (25), **10** was successfully converted to the long-pursued **1**. That success

had been achieved was first indicated by the circular dichroism (CD) criterion (Fig. 4A) (31, 32) and was further supported by the remarkable erythropoietic activity of **1** as judged by in vitro and in vivo assays. Formulated recombinant EPO (Procrit) was used as the standard in these experiments. This recombinant version of EPO has only 165 amino acids as compared with our synthesized human EPO, which has an additional C-terminal Arg known to be cleaved by the Golgi complex in recombinant EPO produced in mammalian cells (33).

As seen in Fig. 4C, the activities of **1** and Procrit track remarkably closely in their effects on the proliferation of TF-1 cells. The in vivo comparison of the effect on **1** and Procrit on the peripheral blood reticulocyte numbers in mice (Fig. 4D) was equally heartening. The initial burst of activity of fully synthetic **1** was ~70% that of Procrit when clinically relevant doses were used (supplementary materials). Further confidence that we indeed generated homogeneous polyglyco-polypeptides arose from comparison of the electrophoretic mobility of commercial rhEPO- α (expressed in CHO cells), compound **1**, compound **S3**, and compound **9** under reducing conditions [SDS-polyacrylamide gel electrophoresis (SDS-PAGE); compare bandwidths in Fig. 4B].

At least for now, we propose the descriptor EPO(S)1 as a name for this homogeneous, wild-type, biocompetent EPO (**1**). We emphasize, however, that the particular rendering of folded EPO **1**, shown in Fig. 4, is provided for convenience solely by analogy with a reported structure of a tri-Lys24,38,83-mutated EPO aglycone bound to the EPO receptor (34).

The ability to reach a molecule of the complexity of **1** by entirely chemical means provides convincing testimony about the growing power of organic synthesis. As a result of synergistic contributions from many laboratories, the aspirations of synthesis may now include, with some degree of realism, structures hitherto referred to as “biologics”—a term used to suggest accessibility only by biological means (isolation from plants, fungi, soil samples, corals, or microorganisms, or by recombinant expression). Formidable as these methods are for the discovery, development, and manufacturing of biologics, one can foresee increasing needs and opportunities for chemical synthesis to provide the first samples of homogeneous biologics. As to production, the experiments described above must be seen as very early days. In no way do we mean to suggest that chemical synthesis is competitive with biologic-type methods to produce EPO, if product homogeneity is not of concern. However, one can well imagine that with continuing improvements in the practicality of solid-phase peptide synthesis (35, 36), new ligation methods, the operational handling of synthetic polypeptides, and enzymatic means of oligosaccharide assembly, synthesis is likely to be of growing value, even in terrains currently described as strictly biologic (39).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors are grateful for support from the National Institutes of Health (grant HL025848), for spectroscopic assistance from G. Sukenick and H. Fang of the Sloan-Kettering Institute's nuclear magnetic resonance core

facility, and to L. Wilson and L. Ambrosini Vadola for preparation of the manuscript. We also thank M. Luo and H. Guo for their assistance on gel electrophoresis and Y. Xia for helpful discussions regarding mass spectroscopy.

References and Notes

1. Elliott, S.; Foote, MA.; Molineux, G. Erythropoietins, Erythropoietic Factors, and Erythropoiesis. ed. 2. Birkhäuser; Basel, Switzerland: 2009.
2. Sytkowski, AJ. Erythropoietin. Wiley-VCH; Weinheim, Germany: 2004.
3. Rush RS, et al. Anal. Chem. 1995; 67:1442–1452. [PubMed: 7741215]
4. Wang P, et al. Angew. Chem. Int. Ed. 2012; 51:11576–11584.
5. Wilson RM, Dong S, Wang P, Danishefsky SJ. Angew. Chem. Int. Chem. 2013; 52:7646–7665.
6. During the course of these investigations, ongoing studies from other laboratories provided striking results in reaching congeners of various sorts with promising erythropoietic activity. This work is summarized in (5) and includes references (31, 37, 38).
7. Wu B, et al. Tetrahedron Lett. 2006; 47:5577–5579.
8. Wu B, et al. Tetrahedron Lett. 2006; 47:8009–8011.
9. Danishefsky SJ, Bilodeau MT. Angew. Chem. Int. Ed. Engl. 1996; 35:1380–1419.
10. Danishefsky SJ, Allen JR. Angew. Chem. Int. Ed. 2000; 39:836–863.
11. Schwarz JB, et al. J. Am. Chem. Soc. 1999; 121:2662–2673.
12. There are over 50 glycoforms of recombinant EPO containing combinations of bi-, tri-, and tetraantennary carbohydrate domains, which have not been separated or fully characterized [see (13)].
13. Sasaki H, Bothner B, Dell A, Fukuda M. J. Biol. Chem. 1987; 262:12059–12076. [PubMed: 3624248]
14. Takeuchi M, Kobata A. Glycobiology. 1991; 1:337–346. [PubMed: 1820196]
15. Hamilton SR, et al. Science. 2006; 313:1441–1443. [PubMed: 16960007]
16. Nett JH, et al. J. Biotechnol. 2012; 157:198–206. [PubMed: 22100268]
17. Dawson PE, Muir TW, Clark-Lewis I, Kent SB. Science. 1994; 266:776–779. [PubMed: 7973629]
18. Wan Q, Danishefsky SJ. Angew. Chem. Int. Ed. 2007; 46:9248–9252.
19. Yan LZ, Dawson PE. J. Am. Chem. Soc. 2001; 123:526–533. [PubMed: 11456564]
20. Warren JD, Miller JS, Keding SJ, Danishefsky SJ. J. Am. Chem. Soc. 2004; 126:6576–6578. [PubMed: 15161285]
21. Chen G, et al. J. Am. Chem. Soc. 2006; 128:7460–7462. [PubMed: 16756298]
22. Dong S, Shang S, Tan Z, Danishefsky SJ. Isr. J. Chem. 2011; 51:968–976. [PubMed: 23585694]
23. Wang P, Aussedat B, Vohra Y, Danishefsky SJ. Angew. Chem. Int. Ed. 2012; 51:11571–11575.
24. Ullmann V, et al. Angew. Chem. Int. Ed. 2012; 51:11566–11570.
25. Bang D, Kent SB. Angew. Chem. Int. Ed. 2004; 43:2534–2538.
26. Sakakibara S. Biopolymers. 1995; 37:17–28. [PubMed: 7880964]
27. Liu S, Pentelute BL, Kent SB. Angew. Chem. Int. Ed. 2012; 51:993–999.
28. The mass spectrum of **10** (figs. S24 and S25) is supportive of the assignment. However, as is the case with previous attempts at obtaining high-quality readouts from extensively glycosylated polysialic acid-containing domains, this attempt has not been fully successful.
29. Llop E, Gutiérrez-Gallego R, Segura J, Mallorquí J, Pascual JA. Anal. Biochem. 2008; 383:243–254. [PubMed: 18804089]
30. Saxena VP, Wetlaufer DB. Biochemistry. 1970; 9:5015–5023. [PubMed: 4991411]
31. Kochendoerfer GG, et al. Science. 2003; 299:884–887. [PubMed: 12574628]
32. Narhi LO, et al. Protein Eng. 2001; 14:135–140. [PubMed: 11297671]
33. Recny MA, Scoble HA, Kim Y. J. Biol. Chem. 1987; 262:17156–17163. [PubMed: 3680293]
34. Syed RS, et al. Nature. 1998; 395:511–516. [PubMed: 9774108]
35. Merrifield RB. J. Am. Chem. Soc. 1963; 85:2149–2154.

36. Wang T, Danishefsky SJ. *Proc. Natl. Acad. Sci. U.S.A.* 2013; 110:11708–11713. [PubMed: 23821753]
37. Hirano K, Macmillan D, Tezuka K, Tsuji T, Kajihara Y. *Angew. Chem. Int. Ed. Engl.* 2009; 48:9557–9560. [PubMed: 19798711]
38. Murakami M, Okamoto R, Izumi M, Kajihara Y. *Angew. Chem. Int. Ed. Engl.* 2012; 51:3567–3572. [PubMed: 22307754]
39. Information on materials and methods, including spectroscopic data of new compounds, is available on *Science* Online.

Fig. 1. Schematic representation of the target homogenous EPO glycoform 1

Fig. 2. Synthesis of revised glycopeptide fragments

Reaction conditions for Sakakibara elongation(26): Peptidylacid, amino acidthioester, EDC free base, HOObt, CHCl_3 /TFE, room temperature (rt). Reaction conditions for aspartylation: (i) Peptidyl thioester, **S2**, HATU, DIEA, DMSO, rt. (ii) TFA/TIS/ H_2O /phenol, rt. EDC = N-(3-Dimethylaminopropyl)-N'-ethylcarbodiimide; HOObt = 3-Hydroxy-1,2,3-benzotriazin-4(3H)-one; HATU = 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate; TFE = trifluoroethanol; DIEA = *N,N*-Diisopropylethylamine; DMSO = dimethyl sulfoxide; TFA = trifluoroacetic acid; TIS = triisopropylsilane.

Fig. 3. Synthesis of homogeneous EPO glycoform 1

SPPS, solid-phase peptide synthesis. Reaction conditions for NCL: 6 M GND•HCl, 300 mM Na₂HPO₄, 20 mM TCEP•HCl, 200 mM MPAA, pH 7.2, rt. Reaction conditions for Thz opening: MeONH₂•HCl, rt. Reaction conditions for MFD: 5.7 M GND•HCl, 200 mM Na₂HPO₄, 300 mM TCEP (Bond-Breaker), *t*-BuSH, VA-044, pH 6.8, 37°C. Reaction conditions for AcM removal: AgOAc, AcOH/H₂O, rt. Reaction conditions for folding: cysteine/cystine (8:1). GND = guanidine; TCEP = tris(2-carboxyethyl)phosphine; MPAA = 4-Mercaptophenylacetic acid; Thz = thiazolidine; VA-044 = 2,2'-Azobis[2-(2-imidazolin-2-yl) propane]dihydrochloride.

Fig. 4. Functional characterization of 1

(A) CD spectrum of EPO(S)1 (1). (B) SDS-PAGE conducted on two separate gels (1, lanes a and b; 2, lanes c, d, and e) both using rEPO α (EMD) as control (lanes a and e). Lane b: EPO(S)1 (1); Lane c: EPO glycoform S3; Lane d: EPO (29-166) (9). (C) In vitro assay of the effect of Procrit and EPO(S)1 (1) on the proliferation of TF-1 cells. The results are expressed as average relative fluorescent intensity $\pm$ SD, run in triplicate. Relative fluorescent intensity = fluorescent intensity of EPO-treated TF-1 cells/maximal fluorescent intensity of 10 ng/ml Procrit-EPO-treated TF-1 cells. (D) Effect of Procrit and EPO(S)1 (1) on peripheral blood reticulocyte numbers in vivo. C57 mice (three mice per group) were subcutaneously injected with 100 μ l of saline containing 1 mg/ml bovine serum albumin and various doses of EPO daily for 3 days. After 24 hours since the last injection, peripheral blood from each mouse was harvested and stained with reticulocyte stain reagent, and reticulocytes per 200 red blood cells were counted under a microscope with a 100 $\times$ objective lens. The results are expressed as the average $\pm$ SD, $n = 3$ mice.