Fully Synthetic Granulocyte Colony-Stimulating Factor Enabled by Isonitrile-Mediated Coupling of Large, Side-Chain-Unprotected Peptides

Andrew G. Roberts[†,‡,*,⊥], **Eric V. Johnston**[†,‡,#,⊥], **Jae-Hung Shieh**[§], **Joseph P. Sondey**[‡], **Ronald C. Hendrickson**[‡], **Malcolm A. S. Moore**[§], and **Samuel J. Danishefsky**[†,‡,§,‖,*]

[†]Laboratory for Bio-Organic Chemistry, Sloan Kettering Institute (SKI) for Cancer Research, New York, New York 10065, United States

[‡]Chemical Biology Program, Sloan Kettering Institute (SKI) for Cancer Research, New York, New York 10065, United States

[§]Cell Biology Program, Sloan Kettering Institute (SKI) for Cancer Research, New York, New York 10065, United States

[‖]Department of Chemistry, Columbia University, New York, New York 10027, United States

Abstract
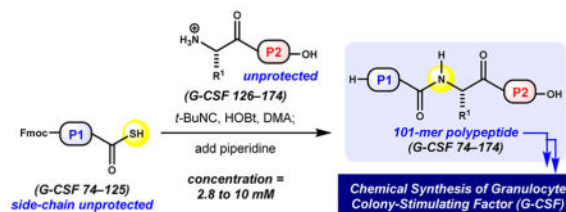
Human granulocyte colony-stimulating factor (G-CSF) is an endogenous glycoprotein involved in hematopoiesis. Natively glycosylated and nonglycosylated recombinant forms, lenograstim and filgrastim, respectively, are used clinically to manage neutropenia in patients undergoing chemotherapeutic treatment. Despite their comparable therapeutic potential, the purpose of O-linked glycosylation at Thr133 remains a subject of controversy. In light of this, we have developed a synthetic platform to prepare G-CSF aglycone with the goal of enabling access to

[*]Corresponding Authors: (A.G.R.) roberta1@mskcc.org. (S.J.D.) s-danishefsky@ski.mskcc.org.
[#]**Present Address**
(E.V.J.) Department of Organic Chemistry, Stockholm University, Arrhenius Laboratory, SE-106 91 Stockholm, Sweden.
[⊥]**Author Contributions**
A.G.R. and E.V.J. contributed equally to this work.

native and designed glycoforms with site-selectivity and glycan homogeneity. To address the synthesis of a relatively large, aggregation-prone sequence, we advanced an isonitrile-mediated ligation method. The chemoselective activation and coupling of C-terminal peptidyl Gly thioacids with the N-terminus of an unprotected peptide provide ligated peptides directly in a manner complementary to that with conventional native chemical ligation–desulfurization strategies. Herein, we describe the details and application of this method as it enabled the convergent total synthesis of G-CSF aglycone.

1. INTRODUCTION

Human granulocyte colony stimulating factor (G-CSF) **1** is the principal growth factor responsible for the regulation of neutrophil granulocyte proliferation and maturation.[1,2] In its major active form, G-CSF **1** is a 19.6 kDa glycoprotein containing 174 amino acid residues and belongs to a family of structurally similar cytokines (Figure 1A).[3,4] Its tertiary structure is composed of four antiparallel alpha-helices rigidified by two crucial intramolecular disulfide linkages, Cys36–Cys42 and Cys64–Cys74 (Figure 1B).[3] Natively glycosylated lenograstim **1** (Granocyte, 1–174) and aglycone filgrastim **2b** (r-metHuG-CSF, Neupogen, Met0 + 1–174) produced by recombinant expression[5] are used clinically to treat neutropenia in cancer patients undergoing chemotherapy.[2,6] Recent interest in the use of G-CSF for the treatment of neurological disease,[6c] among other therapeutic uses,[6] provides impetus for continued research. Our interest in G-CSF stems from its importance among related hematopoietic regulatory glycoproteins[4,7,8] as well as the curious nature of its glycosylation with regard to therapeutic function and structure.[9]

Both endogenous G-CSF **1** and lenograstim (**1**) exist as mixtures of O-linked glycans at Thr133 (Figure 1C).[10] These equipotent glycoforms have been purified to homogeneity, and their carbohydrate structures have been characterized as Neu5Ac-α(2–3)Gal-β(1–3)[Neu5Ac-α(2–6)]GalNAc and its sialic acid-truncated form, Neu5Ac-α(2–3)Gal-β(1–3)-GalNAc.[10b,c] The presence of O-glycosylation does not appear to directly influence biological function, yet it is deemed to be important for proteolytic stability and prevention of severe aggregate formation in solution.[10b] Native glycosylation may indirectly protect the Cys17[11] sulfhydryl moiety from degradative reactivity,[10b] although the carbohydrate is neither proximal to Cys17 nor immediately involved at the principal ligand-binding region of **1** bound to its glycoprotein receptor.[12]

Aggregation in nonglycosylated **2b**[13] can be reduced through covalent attachment of polyethylene glycol at N-terminal Met0, which also augments circulation time in the blood.[14] Several clinical studies express concern with regard to the biological stability of these glycosylated versus nonglycosylated therapeutics (compare **1** and **2b**).[15] Despite this dissimilarity, the role of G-CSF, its respective clinical forms, and emerging biosimilars in hematopoiesis is well-established.[16] Recent studies demonstrate that the incorporation of novel N-linked glycans at single-residue-mutated G-CSF variants provides novel glycosylated analogues.[17] Interestingly, only the Phe140Asn mutant shows notable activity, perhaps due to the proximity of its glycan relative to the native glycosylation site (Thr133).

In the broader context of glycoprotein synthesis and function, natively glycosylated G-CSF **1** and its aglycone **2a** (1–174) are viewed as a veritable targets for chemical synthesis. Additionally, designed N-linked glycoproteins, such as **3**, are intriguing[17] and potentially useful as therapeutics with improved stability.[9,18] Site-specific access to designed N-linked glycoforms at the native position, Thr133Asn, are unattainable by current expression techniques. For these reasons, we aimed to develop a platform for the chemical synthesis of aglycone **2a** using a route amenable to site-selective glycan incorporation.

Our initial strategy relied on conventional chemical protein synthesis methods,[19] including 9-fluorenylmethyloxycarbonyl (Fmoc)-based solid-phase peptide synthesis (SPPS),[20] native chemical ligation (NCL),[21] and desulfurization methodological advances,[22] as delineated in Figure 2. The later two-step sequence is well-recognized as a powerful strategy within the field.[23] Despite its broad utility, enabling access to numerous protein targets, our preliminary endeavors were thwarted due to a single recalcitrant dethiylation of Cys127 to provide native Ala127 (vide infra). Notably, target subsequence G-CSF 74–174 harbors no internal Cys residues available for its convergent assembly by NCL. This impediment serves to emphasize the need for the continued development of thiol-independent ligation techniques[24,25] that do not employ Cys or Cys surrogates,[21,23] both of which require postligation desulfurization[22,23] or ligation auxiliaries[26] that require removal (Figure 2). Therefore, we were inspired to develop a ligation method that would provide G-CSF 74–174 directly, without postligative manipulation of an aggregation-prone sequence.[13]

The structural complexity, size, and physical properties[13] of aglycone **2a** present a synthetic challenge addressed by the advance of a solution-phase ligation of large, side-chain-unprotected polypeptides (>20-mer residue sequences).[27] This effort builds upon previous research in our laboratory pertaining to the intriguing reactivity of isonitriles with carboxylic and thiocarboxylic acids alike.[28,29] The chemo-selective activation of C-terminal thiocarboxylic acids by an isonitrile at room temperature presumably generates a reactive thio-formimidate carboxylate mixed anhydride (thio-FCMA) or 1-hydroxybenzotriazole (HOBt) ester intermediate, permitting subsequent bimolecular ligation (Figure 2).[28,29] Our previous studies demonstrate the utility of such mild reactivity[29a] in the context of macrocyclic peptide synthesis,[29b–d] peptide and glycopeptide ligation,[29e] and solid-phase fragment coupling,[29f] as well as peptide aspartylation.[29g] The current process is effective even at the low molar concentrations required to solubilize long, hydrophobic peptides in side-chain-unprotected form. The first chemical total synthesis of G-CSF **2a** enabled by evaluation of this method in the context of large peptidyl substrates is described herein.

2. RESULTS AND DISCUSSION

2.1. Preliminary Approach and Efforts toward the Chemical Synthesis of G-CSF Aglycone (2a)

At the onset of this research, we devised a chemical synthesis of G-CSF aglycone **2a** utilizing state-of-the-art NCL-metal-free desulfur-ization (MFD) techniques (Figure 2).[21,22c] In this vein, we envisioned reliable means dependent on "now conventional" methods[19,23] to provide expedient access to glycosylated **1** and designed N-linked glycoforms such as **3** (Figure 1). We hypothesize that N-linked glycosylation[9,17,18] may impart improved relative

stability and solubility as well as decreased potential for aggregation during chemical assembly.[10b,13]

We proposed access to the 174-residue protein **2a** by its retrosynthetic dissection into the following four functionalized polypeptides: G-CSF 1–35 (**4**), G-CSF 36–73 (**5**), G-CSF 74–126 (**7**), and G-CSF 127–174 (**8**) (Scheme 1). These disconnections were made with consideration of manageable sequence length (averaging 44 residues per sequence), elements of anticipated secondary structure,[3] and appropriate Cys sites for convergent assembly by NCL (Cys36, Cys74, Cys127). This dictates preparation of G-CSF 1–73 by native ligation of **4** with **5**[21] and access to G-CSF 74–174 from respective partners **7** and **8** via a two-step Ala-ligation.[22,23] The platform hinges on a penultimate union of G-CSF 1–73 (**6**) with G-CSF 74–174 (**9b**) by NCL at a kinetically favorable Gly73–Cys74 junction. Peptide thioesters **4**, **5**, and **7** were prepared using microwave-mediated Fmoc-based SPPS[20,30] with selected use of pseudoproline dipeptides[31] and late-stage installation of thioester-containing residues by a modified Sakakibara elongation.[32] With access to C-terminal thioesters **4**, **5**, and **7** as well as polypeptide **8**, we began preliminary studies to assemble G-CSF aglycone **2a**.[33]

Our research commenced uneventfully to access G-CSF 1–73 (**6**) via kinetic chemical ligation (KCL) of Leu35 phenyl thioester **4** with N-terminal Cys36 **5** (Scheme 1).[34] The strategic use of *tert*-butyl thio-protection enables access to Gly73 ethyl thioester **6** with four of the five native sulfhydryl moieties unveiled at Cys17, -36, -42, and -64.[8] With G-CSF 1–73 (**6**) prepared, we turned our attention to the synthesis of its requisite ligation partner, G-CSF 74–174 (**9b**). The NCL of Met126 ethyl thioester **7** and N-terminal Cys127 **8** proceeded with apparent full conversion under standard conditions. However, isolated yields of **9a** were consistently poor, and ligations conducted with excess nucleophile **8** (2 equiv) did not improve product recovery.[35] These observations led us to believe that poor isolated yield reflects the unfavorable properties of thioester **7**, which seem to translate into product **9a**.[36] Subsequent attempts to dethiylate **9a** (Cys127 → Ala127) in the presence of radical initiator, VA-044, tris(2-carboxyethyl)phosphine, and 2-methylpropane-2-thiol under standard conditions failed to deliver target **9b** (Scheme 1, X = H). Prolonged reaction time, elevated temperature, and alternative solvents, such as *N,N*-dimethylformamide (DMF) or dimethyl sulfoxide, provided an inseparable mixture of **9a** and **9b** with poor conversion. Partially successful reactions generated small quantities of **9b** in the presence of presumed irreversible adduct **9c** as a major species.[37,38] It became apparent that low conversion was likely reflective of the poor solubility properties of **9a**. To the best of our knowledge, sequence length is not a known limitation, as our laboratory[8,39] and others[40] have reported successful desulfurization of comparably sized peptides and glycopeptides.

2.2. Evaluation of the Sequence of G-CSF 74–174 and Development of an Isonitrile-Mediated Coupling Method for Its Preparation

The structure of rh-metG-CSF **2b** in solution positions the corresponding native Ala127 at the beginning of the flexible CD loop region (126–144) and in close proximity to helices B and C (Figure 1A).[3d,41] Although direct comparison of Cys127 in G-CSF 74–174 (**9a**) through analogy with structural data from folded G-CSF 0–174 (**2b**) is tenuous, the insight

offers a structural hindrance explanation as to why efficient dethiylation may be precluded. Alternative Ala ligation sites could have been examined (i.e., Phe113–Ala114, Pro128–Ala129); however, there is currently no a priori method for determining the efficiency of dethiylation. To this end, the development of Cys surrogate residues for NCL with more facile reduction potential,[42] as well as homocysteine[43a] or selenocysteine variants,[43b] represents attractive alternative solutions. Recent disclosures that simplify tandem NCL–desulfurization processes are equally intriguing.[44] Possible alternatives aside, our first approach was unable to deliver isolable quantities of G-CSF 74–174 (**9b**). We desired a ligation method that permits coupling at non-Cys sites, requires minimal post SPPS substrate modification, can be conducted in the presence of unprotected side-chain functionality and requires minimal postligation manipulation.

Despite the growing body of literature focused on the development of chemoselective methods for amide bond-forming reactions,[25] particularly those concerned with cysteine-or thiol-independent ligation of unprotected peptides,[24] few studies address the limitations related to sequence length and the poor solubility of large hydrophobic peptides in bimolecular reactions. Notable exceptions include silver-promoted thio-acid[46a] and thioester[46b,c] ligations, reverse NCL,[46d] Ser/Thr ligations,[46e,f] and ketoacid-hydroxylamine ligations.[46g–k] To evaluate the suitability of alternative methods, we took a closer look at the sequence of G-CSF 74–174, which comprises three of four alpha helical regions in folded G-CSF (Figure 1A).[3] Sequence analysis indicates the target 101-mer is considerably hydrophobic, harboring 12 Leu residues within G-CSF 74–125, as well as both Ala- and Leu-rich stretches within G-CSF 126–174. Our initial studies demonstrate that the exemplary sequence is challenging to solubilize, purify, and isolate in unprotected form by conventional techniques. Predictions based upon the coil conformation parameter, P_c, suggest that these sequences would be poorly soluble in fully protected form, rendering efficient solution-phase coupling difficult.[45] Our successes with the chemoselective activation of thiocarbox-ylic acids prompted us to pursue this modality for large peptide ligations.[28,29]

In recent years, a variety of other methods demonstrate the potential of thiocarboxylic and peptidyl C-terminal thiocarbox-ylic acids alike to serve as effective acyl donors.[47–49] The activation of thiocarboxylic acids under various oxidative conditions is well-studied.[47,48] Additionally, thiocarboxylic acid substrates provide access to various amide types upon engagement with acyl azides,[49a] azides,[49b–d] isocyanates,[49e] 2-pyridyl thioethers,[49f,g] thiols,[49h] and aryl sulfonamides.[49i–k] The utility of peptidyl aziridines in context with activated thioacids has also been realized.[50] Recent methods report mild conditions for acyl transfer through the unique activation of thioacids[51a] and thioesters[51b] with *N,O*-bis(trimethylsilyl)-acetamide. Lastly, thioacid activation with N-terminal dithio-carbamates[52a] or N-terminal thioamides[52b] highlights the utility of presumed 1,3-diacyl intermediates. Despite the variety of conditions available to chemoselectively engage thiocarboxylic acids, few studies exploit their potential in the context of large peptide substrates.[27] Furthermore, low to moderate isolated yields during such bimolecular activation–ligation sequences are common in native amide-bond–forming reactions for examples that are larger than tripeptides. This limitation may be due, in part, to the

propensity of peptide C-terminal thiocarboxylic acids to undergo epimerization during activation as well as the inherently poor solubility of large sequences in compatible solvents.[29] To examine the potential of the later prospect with confidence, we opted to study *tert*-butyl isonitrile-mediated activations and ligations at Gly residues of large peptidyl thioacids to access to G-CSF 74–174.

Fortunately, the G-CSF 74–174 sequence is particularly well-suited for evaluation of this method.[53] Although lacking internal Cys residues, the native sequence is replete with well-spaced Gly residues, which would serve as ideal acyl donors, and free of Lys residues, which would require orthogonal side-chain protection. By analogy to the synthesis of Met126 ethyl thioester **7**, we prepared Gly125 thioacid **12** by activation of Leu124 carboxylic acid **11** in the presence of excess Gly125 thioester **11** followed by complete side-chain deprotection and purification (Scheme 2).[48k,54]

We evaluated several parameters for the bimolecular activation–ligation of Gly125 thioacid **12** in the presence of nucleophilic partner Met126 **13** (1.5 equiv). Reaction progress was monitored at various time points.[33] A convenient procedure was developed wherein the reaction was quenched with piperidine followed by precipitation.[55] In a two-step, one-pot process, isonitrile-mediated ligation afforded 101-mer, G-CSF 74–174 (**14**) in 10–18% isolated yield under optimal conditions. For this ligation, *N,N*-dimethylacetamide (DMA) provides superior conversion relative to that with DMF and *N*-methylpyrrolidone (NMP). Relative to observed side products, Gly125 carboxylic acid **15** and its respective piperidide **16**, reaction conversion is estimated to be between 30 and 50%.[33] The relative quantity of **16** decreased during reaction progression. This suggests that activated acyl intermediate(s) derived from **12** are stable to coupling conditions and diverge to ligation product **14** or hydrolysis product **15**. The absence of remaining thioacid **12** suggests that activation occurs quantitatively, yet bimolecular ligation competes with an undetermined pathway, resulting in formal hydrolysis. Under all productive conditions examined, the reaction rate slows after 48 h such that prolonged experiments do not result in a significant increase in yield. Experiments run open to air in the absence of *tert*-butyl isonitrile resulted in poor conversion (entries 6 and 7, <10% conversion).[29e] We attribute overall attenuated isolated yields to the aggregation-prone properties and poor solubility of target 101-mer **14**, as well as difficult chromatographic separation from carboxylic acid **15**. Notwithstanding poor recovery, this method demonstrates an effective means to access an aggregation-prone sequence relative to conventional strategies (vide supra) and represents the largest isonitrile-mediated ligation to date.

To further evaluate this method, we examined couplings at Gly149–Gly150 and Gly135–Ala136 junctions in the context of G-CSF 126–174 (Scheme 3). Target sequences G-CSF 126–174 (**13**) and glycosylated derivative **22** were chosen to provide insight as to how novel N-linked glycoforms, such as **3**, may be accessed. The ligation of Gly149 thioacid **17** with nucleophilic partner Gly150 **18** provides 49-mer **13** in 21–33% isolated yield. The median junction point Gly149–Gly150 provides convergent access to **13** from two soluble, equal-sized sequences and completes a formal four-step synthesis of target 101-mer **14**. The ligation of glycan containing Gly135 thioacid **19a** with nucleophilic partner Ala136 **20** provides the analogous 49-mer glycopeptide **21** in 37–65% isolated yield.[33] Interestingly,

the shortened glycosylated thioacid sequence **19a** confers improved physical properties and permits efficient ligation conducted at higher concentration (10 mM) without compromised solubility.

With few examples, it is difficult to establish general conditions with regard to effective concentration and solvent. It is apparent that solubility is a limiting, sequence-dependent factor. As a guideline, ideal conditions should be identified by screening compatible solvents at concentrations near saturation of the ligation partners. In the context of the relevant G-CSF sequences described, NMP and DMA (>2 mM) are optimal for such large peptide sequences. Bimolecular acyl transfer to the incoming N-terminal peptide resulting in product formation competes with C-terminal thioacid substrate hydrolysis. Efforts to understand and minimize factors that contribute to hydrolysis are ongoing. The extent of hydrolysis may vary by sequence and is perhaps dependent on the propensity of an activated substrate to form transient oxazolone species.[56] This study has demonstrated successful ligations between Gly–Met, Gly–Gly, and Gly–Ala junctions.

2.3. Fully Synthetic G-CSF 1–174 Aglycone (2a): Ultimate Ligation, Folding, Characterization, and Evaluation of Biological Competency

With access to G-CSF 1–73 (**6**) and G-CSF 74–174 (**14**), we performed the ultimate NCL to access the complete linear sequence of G-CSF **2a** (Scheme 4). Full-length G-CSF 1–174 (**22**) formed in high conversion in the presence of excess Gly ethyl thioester **6** (1.5 equiv) after 20 h. This Gly73–Cys74-ligated material **22** was routinely purified to homogeneity by reverse-phase HPLC. Folding of **22** under oxidative conditions previously established for recombinant **2b**[57] proceeded uneventfully to provide synthetic G-CSF **2a** in 28% isolated yield from **22**. Streamlined access to **2a** was achieved wherein crude **22** is folded under identical conditions immediately following its precipitation from ligation buffer (6% isolated yield, two steps).[58] As expected, synthetic **2a** and recombinant **2b**, differing only by the addition of N-terminal Met0 residue, have similar retention when analyzed under various chromatographic methods.[33] Moreover, high-resolution mass spectrometry data supports the partial oxidation of G-CSF 1–174 (**22**) to provide G-CSF 1–174 (**2a**) with two disulfide linkages and an intact sulfhydryl moiety (Figure 3B). Additional characterization data of synthetic G-CSF **2a** in comparison with commercial recombinant G-CSF **2b**, including analysis by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (Figure 3A), circular dichroism spectroscopy, and proteolytic mapping, support[59] its identity.[33]

2.4. Biological Competency of Synthetic G-CSF 1–174 Aglycone (2a)

A proliferation assay of murine myeloblastic NFS-60 cells and a colony-forming assay of human primary cord blood CD34+ cells were used to validate the biological activity of synthetic G-CSF 1–174 aglycone (**2a**) compared with control, recombinant G-CSF 0–174 (**2b**, Neupogen, Amgen Inc.).[7b,33,60] Preliminary evaluation in NFS-60 cells demonstrates that synthetic G-CSF **2a** does induce cellular proliferation (Figure 4), albeit with reduced competency relative to recombinant G-CSF **2b** (**2a**, EC_{50} of ~0.2 ng/mL vs **2b**, 2 ng/mL). Synthetic **2a** also stimulates the proliferation of normal human CD34+ cells to form clusters and colonies alone, as well as synergistically with recombinant human klotho protein (KL) (**2a**, EC_{50} of ~1.25 ng/mL vs **2b**, 0.08 ng/mL).[33] Despite the reduced activity, these results

are encouraging in light of the known propensity for nonglycosylated G-CSF **2b** to aggregate.[13] Such physical property challenges are reported during denaturation, refolding, and isolation of **2b** from recombinant preparations,[57] as well as during formulation and storage in nonstabilized media.[61]

Interestingly, recent reports by Broxmeyer and co-workers provide an explanation for diminished cytokine potency attributed to enzymatic regulation by dipeptidylpeptidase (DPP4).[62] The study suggests the additional Met0 residue at the N-terminus of recombinant G-CSF **2b** serves as a fortuitous and distinguishing modification relative to native G-CSF **1**.[62a] The analogous experiments with synthetic G-CSF **2a**, lacking this beneficial residue (Met0), relative to recombinant G-CSF**2b** are of considerable interest. Efforts to better emulate reported isolation and formulation techniques to afford a more potent synthetic G-CSF **2a** as well as glycosylated variants thereof are ongoing.

3. CONCLUSIONS

In response to a challenging dethiylation, we revised our initial approach to access synthetic G-CSF aglycone **2a**. The strategy hinges on the development and application of an isonitrile-mediated ligation of large, side-chain-unprotected peptides in solution. This advance enables a highly convergent assembly of **2a** and provides insight as to how future G-CSF glycoform variants may be constructed. Building in the C- to N-terminal direction, the method is complementary to NCL for the chemical synthesis of large hydrophobic proteins and is a useful alternative in instances where tandem NCL–desulfurization strategies are troublesome. Moreover, the application is distinct from most thiol-independent ligation methods in that one of the two coupling partners is an unmodified peptide available from routine SPPS. Efforts to showcase the generality of this ligation method in series as well as in convergent fashion for the synthesis of large peptides and glycoproteins are ongoing.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. For seminal research contributions, see: Welte K, Platzer EL, Lu H, Gabrilove J, Levi E, Mertelsmann R, Moore MA. Proc Natl Acad Sci U S A. 1985; 82:1526–1530. [PubMed: 3871951] Souza LM, Boone TC, Gabrilove J, Lai PH, Zsebo KM, Murdock DC, Chazin VR, Bruszewski J,

Lu H, Chen KK, Barendt J, Platzer E, Moore MAS, Mertelsmann R, Welte K. Science. 1986; 232:61–65. [PubMed: 2420009] Platzer E, Oez S, Welte K, Sendler A, Gabrilove JL, Mertelsmann R, Moore MA, Kalden JR. Immunobiology. 1986; 172:185–193. [PubMed: 3492429] Nomura H, Imazeki I, Oheda M, Kubota N, Tamura M, Ono M, Ueyama Y, Asano S. EMBO J. 1986; 5:871–876. [PubMed: 3487449] Cohen AM, Zsebo KM, Inoue H, Hines D, Boone TC, Chazin R, Tsai L, Ritch T, Souza LM. Proc Natl Acad Sci U S A. 1987; 84:2484–2488. [PubMed: 3550811] Bussolino F, Wang JM, Defilippi P, Turrini F, Sanavio F, Edgell CS, Aglietta M, Arese P, Mantovani A. Nature. 1989; 337:471–473. [PubMed: 2464767]

2. For selected reviews, see: Roberts AW, Nicola NA. Granulocyte Colony-Stimulating Factor. Colony-Stimulating Factors: Molecular & Cellular Biology (2). Garland JM, Quesenberry PJ, Hilton DJ. Marcel DekkerNew York1997:203–226.Metcalf D. Nature. 1989; 339:27–30. [PubMed: 2469962]

3. For structural characterization of G-CSF, see: Hill CP, Osslund TD, Eisenberg D. Proc Natl Acad Sci U S A. 1993; 90:5167–5171. [PubMed: 7685117] Lovejoy B, Cascio D, Eisenberg D. J Mol Biol. 1993; 234:640–653. [PubMed: 7504736] Werner JM, Breeze AL, Kara B, Rosenbrock G, Boyd J, Soffe N, Campbell ID. Biochemistry. 1994; 33:7184–7192. [PubMed: 7516182] Zink T, Ross A, Lüers K, Cieslar C, Rudolph R, Holak TA. Biochemistry. 1994; 33:8453–8463. [PubMed: 7518249]

4. (a) Mott HR, Campbell ID. Curr Opin Struct Biol. 1995; 5:114–121. [PubMed: 7773741] (b) Maurer MH, Schäbitz WR, Schneider A. Curr Med Chem. 2008; 15:1407–1411. [PubMed: 18537618]

5. (a) Kubota N, Orita T, Hattori K, Oh-eda M, Ochi N, Yamazaki T. J Biochem. 1990; 107:486–492. [PubMed: 1692828] (b) Nagata S, Tsuchiya M, Asano S, Kaziro Y, Yamazaki T, Yamamoto O, Hirata Y, Kubota N, Oheda M, Nomura H, Ono M. Nature. 1986; 319:415–418. [PubMed: 3484805]

6. (a) Dale DC. Drugs. 2002; 62:1–15. [PubMed: 12479591] (b) Schneider A, Kuhn HG, Schäbitz WR. Cell Cycle. 2005; 4:1753–1757. [PubMed: 16258290] (c) Hamilton JA. Nat Rev Immunol. 2008; 8:533–544. [PubMed: 18551128] (d) Adusumilli A, Rao K, RK, Krothapalli SR. Asian J Biomed Pharm Sci. 2012; 2:1–10.

7. Wilson RM, Dong S, Wang P, Danishefsky SJ. Angew Chem, Int Ed. 2013; 52:7646–7665.Wang P, Dong S, Shieh JH, Peguero E, Hendrickson R, Moore MAS, Danishefsky SJ. Science. 2013; 342:1357–1360. [PubMed: 24337294] See also: Fernández-Tejada A, Brailsford J, Zhang Q, Shieh JH, Moore MAS, Danishefsky SJ. Top Curr Chem. 2014; 362:1–26. [PubMed: 25805144]

8. Zhang Q, Johnston EV, Shieh JH, Moore MAS, Danishefsky SJ. Proc Natl Acad Sci U S A. 2014; 111:2885–2890. [PubMed: 24516138]

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

10. (a) Oheda M, Hase S, Ono M, Ikenaka T. J Biochem. 1988; 103:544–546. [PubMed: 2455710] (b) Oh-eda M, Hasegawa M, Hattori K, Kubinowa J, Kojima T, Orita T, Tomonou K, Yamazaki T, Ochi N. J Biol Chem. 1990; 265:11432–11435. [PubMed: 1694845] (c) Rotondaro L, De Paolis E, Ferrero D, D'Alatri L, Raucci G, Fabbri A, Gerwig GJ, Kamerling JP, Mariani MF, Mele A, De Santis R. Mol Biotechnol. 1999; 11:117–128. [PubMed: 10464766]

11. (a) Ishikawa M, Iijima H, Satake-Ishikawa R, Tsumura H, Iwamatsu A, Kadoya T, Shimada Y, Fukamachi H, Kobayashi K, Matsuki S, Asano K. Cell Struct Funct. 1992; 17:61–65. [PubMed: 1375132] (b) Arakawa T, Prestrelski SJ, Narhi LO, Boone TC, Kenney WC. J Protein Chem. 1993; 12:525–531. [PubMed: 7511386]

12. Aritomi M, Kunishima N, Okamoto T, Kuroki R, Ota Y, Morikawa K. Nature. 1999; 401:713–717. [PubMed: 10537111] (b) Tamada T, Honjo E, Maeda Y, Okamoto T, Ishibashi M, Tokunaga M, Kuroki R. Proc Natl Acad Sci U S A. 2006; 103:3135–3140. [PubMed: 16492764] (c) Wells JA, de Vos AM. Annu Rev Biochem. 1996; 65:609–634. [PubMed: 8811191]

13. (a) Chi EY, Krishnan S, Kendrick BS, Chang BS, Carpenter JF, Randolph TW. Protein Sci. 2002; 12:903–913. [PubMed: 12717013] (b) Krishnan S, Chi EY, Webb JN, Chang BS, Shan D, Goldenberg M, Manning MC, Randolph TW, Carpenter JF. Biochemistry. 2002; 41:6422–6431. [PubMed: 12009905] (c) Raso SW, Abel J, Barnes JM, Maloney KM, Pipes G, Treuheit MJ, King J, Brems DN. Protein Sci. 2005; 14:2246–2257. [PubMed: 16131655] (d) Ribarska JT, Jolevska ST, Panovska A, Dimitrovska A. Acta Pharm. 2008; 58:199–206. [PubMed: 18515229]

14. (a) Satake-Ishikawa R, Ishikawa M, Okada Y, Kakitani M, Kawagishi M, Matsuki S, Asano K. Cell Struct Funct. 1992; 17:157–160. [PubMed: 1379521] (b) Kinstler OB, Brems DN, Lauren SL, Paige AG, Hamburger JB, Treuheit MJ. Pharm Res. 1996; 13:996–1002. [PubMed: 8842035] (c) Rajan RS, Li T, Aras M, Sloey C, Sutherland W, Arai H, Briddell R, Kinstler O, Lueras MK, Zhang Y, Yeghnazar H, Treuheit M, Brems DN. Protein Sci. 2006; 15:1063–1075. [PubMed: 16597829]

15. For discussions, see: Höglund M. Med Oncol Tumor Pharmacother. 1998; 15:229–233.Chamorey AL, Magné N, Pivot G. Eur Cytokine Netw. 2002; 13:154–160. [PubMed: 12101071] For selected examples, see: Bönig H, Silbermann S, Weller S, Kirschke R, Körholz D, Janssen G, Göbel U, Nürenberger W. Bone Marrow Transplant. 2001; 28:259–264. [PubMed: 11535993] Carter CRD, Keeble JR, Thorpe R. Biologicals. 2004; 32:37–47. [PubMed: 15026024] Carter CRD, Whitmore KM, Thorpe R. J Leukocyte Biol. 2004; 75:515–522. [PubMed: 14657210] Mattii L, Azzará A, Fazzi R, Carulli G, Chimenti M, Cecconi N, Galimberti S, Petrini M. Leuk Res. 2005; 29:1285–1292. [PubMed: 15916805] Ribeiro D, Veldwijk MR, Benner A, Laufs S, Wenz F, Ho AD, Fruehauf S. Transfusion. 2007; 47:969–980. [PubMed: 17524085]

16. (a) Welte K, Gabrilove J, Bronchud MH, Platzer E, Mortstyn G. Blood. 1996; 88:1907–1929. [PubMed: 8822908] (b) Welte, K. Discovery of G-CSF and Early Clinical Studies. In: Molineux, G.; Foote, M.; Arvedson, T., editors. Twenty Years of G-CSF, Milestones in Drug Therapy. Springer Basel AG; Basel, Switzerland: 2012. p. 15-24.(c) Welte K. Expert Opin Biol Ther. 2014; 14:983–993. [PubMed: 24707817]

17. Chung HY, Ko EM, Kim SW, Byun SJ, Chung HK, Kwon M, Lee HC, Yang BC, Han DW, Park JK, Hong SG, Chang WK, Kim KW. BMB Reports. 2012; 45:742–747. [PubMed: 23261062]

18. (a) Imperiali B, O'Connor SE. Curr Opin Chem Biol. 1999; 3:643–649. [PubMed: 10600722] (b) Solá RJ, Griebenow K. BioDrugs. 2010; 24:9–21. [PubMed: 20055529]

19. For selected reviews pertaining to chemical protein synthesis, see: Nilsson BL, Soellner MB, Raines RT. Annu Rev Biophys Biomol Struct. 2005; 34:91–118. [PubMed: 15869385] Kimmerlin T, Seebach D. J Pept Res. 2005; 65:229–260. [PubMed: 15705167] Kent SB. Chem Soc Rev. 2009; 38:338–351. [PubMed: 19169452] Kan C, Danishefsky SJ. Tetrahedron. 2009; 65:9047–9065. [PubMed: 20161377] Unverzagt C, Kajihara Y. Chem Soc Rev. 2013; 42:4408–4420. [PubMed: 23403448]

20. (a) Merrifield RB. J Am Chem Soc. 1963; 85:2149–2154.(b) Fields GB, Noble RL. Int J Pept Protein Res. 1990; 35:161–214. [PubMed: 2191922]

21. Dawson PE, Muir TW, Clark-Lewis I, Kent SB. Science. 1994; 266:776–779. [PubMed: 7973629]

22. (a) Yan LZ, Dawson PE. J Am Chem Soc. 2001; 123:526–533. [PubMed: 11456564] (b) Pentelute BL, Kent SB. Org Lett. 2007; 9:687–690. [PubMed: 17286375] (c) Wan Q, Danishefsky SJ. Angew Chem, Int Ed. 2007; 46:9248–9252.

23. (a) Rohde H, Seitz O. Biopolymers. 2010; 94:551–559. [PubMed: 20593472] (b) Dawson PE. Isr J Chem. 2011; 51:862–867.(c) Malins LR, Payne RJ. Aust J Chem. 2015; 68:521–537.

24. Chow HY, Li X. Tetrahedron Lett. 2015; 56:3715–3720.

25. Pattabiraman VR, Bode JW. Nature. 2011; 480:471–479. [PubMed: 22193101]

26. (a) Coltart DM. Tetrahedron. 2000; 56:3449–3491.(b) Offer J. Biopolymers. 2010; 94:530–541. [PubMed: 20593473]

27. In this context, large is defined as peptide sequences of more than 20 residues in length and ligation target sequences that are conventionally inaccessible by routine solid-phase peptide synthesis. For discussion, evaluation, and limitations of chemoselective ligation reactions, see: Saito F, Noda H, Bode JW. ACS Chem Biol. 2015; 10:1026–1033. [PubMed: 25572124]

28. Li X, Danishefsky SJ. J Am Chem Soc. 2008; 130:5446–5448. [PubMed: 18370392] For a review, see: Wilson RM, Stockdill JL, Wu X, Li X, Vadola PA, Park PK, Wang P, Danishefsky SJ. Angew Chem, Int Ed. 2012; 51:2834–2848. and references therein.

29. (a) Rao Y, Li X, Danishefsky SJ. J Am Chem Soc. 2009; 131:12924–12926. [PubMed: 19697932] (b) Wu X, Stockdill JL, Wang P, Danishefsky SJ. J Am Chem Soc. 2010; 132:4098–4100. [PubMed: 20199071] (c) Wu X, Stockdill JL, Park PK, Danishefsky SJ. J Am Chem Soc. 2012; 134:2378–2384. [PubMed: 22280518] (d) Wang T, Danishefsky SJ. J Am Chem Soc. 2012; 134:13244–13247. [PubMed: 22830449] (e) Wang P, Danishefsky SJ. J Am Chem Soc. 2010; 132:17045–17051. [PubMed: 21049949] (f) Wang T, Danishefsky SJ. Proc Natl Acad Sci U S A.

2013; 110:11708–11713. [PubMed: 23821753] (g) Wang P, Li X, Zhu J, Chen J, Yuan Y, Wu X, Danishefsky SJ. J Am Chem Soc. 2011; 133:1597–1602. [PubMed: 21207981]

30. Palasek SA, Cox ZJ, Collins JM. J Pept Sci. 2007; 13:143–148. and references therein. [PubMed: 17121420]

31. Wöhr T, Wahl F, Nefzi A, Rohwedder B, Sato T, Sun X, Mutter M. J Am Chem Soc. 1996; 118:9218–9227.

32. (a) Sakakibara S. Biopolymers. 1995; 37:17–28. [PubMed: 7880964] (b) Stuhr-Hansen N, Wilbek TS, Strømgaard K. Eur J Org Chem. 2013; 24:5290–5294.

33. See the Supporting Information.

34. Bang D, Pentelute BL, Kent SB. Angew Chem, Int Ed. 2006; 45:3985–3988.

35. In some instances, we observed precipitation of substrate, presumed to be thioester **7**, from ligation media. Reaction monitoring was performed by UPLC-MS analysis following reaction precipitation with water, centrifugation, and reconstitution of the pellet in neat trifluoroacetic acid (see the Supporting Information).

36. Retrospectively, the challenging purification of thioester G-CSF 74–126 **7** served as a harbinger for the difficulty experienced in attempts to prepare G-CSF 74–174 **9a**. Low mass-balance recovery of side products **15** and **16** corroborates such loss due to purification. Relative to G-CSF sequence 1–73 (**6**), purified G-CSF 74–174 (**9a**) was generally insoluble.

37. For a similar report of VA-044 derived adduct formation, see: Cergol KM, Thompson RE, Malins LR, Turner P, Payne RJ. Org Lett. 2014; 16:290–293. [PubMed: 24294973] See the Supporting Information..

38. More forcing conditions at higher temperatures with compatible radical initiators (VA-050, AIBN) were equally unsuccessful. Geiermann A-S, Micura R. ChemBioChem. 2012; 13:1742–1745. [PubMed: 22786696]

39. The presence of multiple Met residues (Met121, -126, -137) precluded attempts to dethiylate G-CSF 74–174 with Raney-Nickel. For the metal-free dethiylation of comparably sized polypeptides, see: Brailsford JA, Danishefsky SJ. Proc Natl Acad Sci U S A. 2012; 109:7196–7201. [PubMed: 22499784] Li J, Dong S, Townsend SD, Dean T, Gardella TJ, Danishefsky SJ. Angew Chem, Int Ed. 2012; 51:12263–12267.Creech GS, Paresi C, Li YM, Danishefsky SJ. Proc Natl Acad Sci U S A. 2014; 111:2891–2896. [PubMed: 24516155]

40. For impressive chemical protein synthesis examples wherein large proteins (>300 residues) are desulfurized, see: Kumar KSA, Bavikar SN, Spasser L, Moyal T, Ohayon S, Brik A. Angew Chem, Int Ed. 2011; 50:6137–6141.Haj-Yahya M, Fauvet B, Herman-Bachinsky Y, Hejjaoui M, Bavikar SN, Karthikeyan SV, Ciechanover A, Lashuel HA, Brik A. Proc Natl Acad Sci U S A. 2013; 110:17726–17731. [PubMed: 24043770]

41. The solution structure of G-CSF was analyzed and images were rendered with the PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC, using PDB entry 1GNC. See ref 3d.

42. For -mercapto Glu and β-mercapto Asp residues used in chemical protein synthesis, see: Cergol KM, Thompson RE, Malins LR, Turner P, Payne RJ. Org Lett. 2014; 16:290–293. [PubMed: 24294973] Thompson RE, Chan B, Radom L, Jolliffe KA, Payne RJ. Angew Chem, Int Ed. 2013; 52:9723–9727.Guan X, Drake MR, Tan Z. Org Lett. 2013; 15:6128–6131. [PubMed: 24266801]

43. (a) Tam JP, Yu Q. Biopolymers. 1998; 46:319–327. [PubMed: 9754028] (b) Quaderer R, Sewing A, Hilvert D. Helv Chim Acta. 2001; 84:1197–1206.

44. (a) Thompson RE, Liu X, Alonso-Garcia N, Pereira PJB, Jolliffe KA, Payne RJ. J Am Chem Soc. 2014; 136:8161–8164. [PubMed: 24873761] (b) Reimann O, Smet-Nocca C, Hackenberger CPR. Angew Chem, Int Ed. 2015; 54:306–310.

45. Although accessible, the condensation of fully protected peptides, C-terminal Gly125 acid (G-CSF 74–125) with G-CSF 126–174, was not pursued due to the anticipated difficulty associated with their purification. See: Narita M, Ishikawa K, Chen J-Y, Kim Y. Int J Pept Protein Res. 1984; 24:580–587. [PubMed: 6530333]

46. (a) Blake J, Li CH. Proc Natl Acad Sci U S A. 1981; 78:4055–4058. [PubMed: 6270658] (b) Hojo H, Matsumoto Y, Nakahara Y, Ito E, Suzuki Y, Suzuki M, Suzuki A, Nakahara Y. J Am Chem Soc. 2005; 127:13720–13725. [PubMed: 16190738] (c) Asahina Y, Kamitori S, Takao T, Nishi N, Hojo H. Angew Chem. 2013; 125:9915–9919.(d) Tam JP, Lu YA, Liu CF, Shao J. Proc Natl Acad

Sci U S A. 1995; 92:12485–12489. [PubMed: 8618926] (e) Zhang Y, Xu Ci Lam HY, Li X. Proc Natl Acad Sci U S A. 2013; 110:6657–6662. [PubMed: 23569249] (f) Wong CTT, Li T, Lam HY, Zhang Y, Li X. Front Chem. 2014; 2:28–34. [PubMed: 24904921] (g) Wu J, Ruiz-Rodríguez J, Comstock JM, Dong JZ, Bode JW. Chem Sci. 2011; 2:1976–1979.(h) Ogunkoya AO, Pattabiraman VR, Bode JW. Angew Chem, Int Ed. 2012; 51:9693–9697.(i) Wucherpfennig TG, Pattabiraman VR, Limberg FRP, Ruiz-Rodríguez J, Bode JW. Angew Chem, Int Ed. 2014; 53:12248–12252.(j) Rohrbacher F, Wucherpfennig TG, Bode JW. Top Curr Chem. 2014; 363:1–32. [PubMed: 25761549] (k) Pusterla I, Bode JW. Nat Chem. 2015; 7:668–672. [PubMed: 26201744]

47. Liu R, Orgel LE. Nature. 1997; 389:52–54. and references therein. [PubMed: 9288964]

48. (a) Pan J, Devarie-Baez NO, Xian M. Org Lett. 2011; 13:1092–1094. [PubMed: 21271725] (b) Joseph R, Dyer FB, Garner P. Org Lett. 2012; 15:732–735. [PubMed: 23346888] (c) Mali SM, Bhaisare RD, Gopi HN. J Org Chem. 2013; 78:5550–5555. [PubMed: 23672478] (d) Mail SM, Gopi HN. J Org Chem. 2014; 79:2377–2383. [PubMed: 24548175]

49. (a) Mhidia R, Boll E, Fécourt F, Ermolenko M, Ollivier N, Sasaki K, Crich D, Delpech B, Melnyk O. Bioorg Med Chem. 2013; 21:3479–3485. [PubMed: 23602526] (b) Shangguan N, Katukojvala S, Greenberg R, Williams LJ. J Am Chem Soc. 2003; 125:7754–7755. [PubMed: 12822965] (c) Kolakowski RV, Shangguan N, Sauers RR, Williams LJ. J Am Chem Soc. 2006; 128:5695–5702. [PubMed: 16637636] (d) Raz R, Rademann J. Org Lett. 2012; 14:5038–5041. [PubMed: 22991969] (e) Mühlberg M, Siebertz KD, Schlegel B, Schmieder P, Hackenberger CPR. Chem Commun. 2014; 50:4603–4606.(f) Crich D, Sasaki K. Org Lett. 2009; 11:3514–3517. [PubMed: 19719195] (g) Liu CF, Rao C, Tam JP. Tetrahedron Lett. 1996; 37:933–936.(h) Zhang X, Li F, Liu CF. Chem Commun. 2011; 47:1746–1748.(i) Shigenaga A, Sumikawa Y, Tsuda S, Sato K, Otaka A. Tetrahedron. 2010; 66:3290–3296.(j) Crich D, Sana K, Guo S. Org Lett. 2007; 9:4423–4426. [PubMed: 17900128] (k) Crich D, Sharma I. Angew Chem, Int Ed. 2009; 48:7591–7594.(l) Sasaki K, Crich D. Org Lett. 2010; 12:3254–3257. [PubMed: 20568826] (m) Karmakar P, Talan RS, Sucheck SJ. Org Lett. 2011; 13:5298–5301. [PubMed: 21916403]

50. (a) Assem N, Natarajan A, Yudin AK. J Am Chem Soc. 2010; 132:10986–10987. [PubMed: 20666404] (b) Dyer FB, Park CM, Joseph R, Garner P. J Am Chem Soc. 2011; 133:20033–20035. [PubMed: 22112239] (c) Murray C, Dyer FB, Garner P. Tetrahedron Lett. 2015; 56:3636–3638.

51. (a) Wu W, Zhang Z, Liebeskind LS. J Am Chem Soc. 2011; 133:14256–14259. [PubMed: 21846113] (b) Chen H, He M, Wang Y, Zhai L, Cui Y, Li Y, Li Y, Zhou H, Hong X, Deng Z. Green Chem. 2011; 13:2723–2726.

52. (a) Chen W, Shao J, Hu M, Yu W, Giulianotti MA, Houghten RA, Yu Y. Chem Sci. 2013; 4:970–976.(b) Pourvali A, Cochrane JR, Hutton CA. Chem Commun. 2014; 50:15963–15966.

53. Within the G-CSF 74–174 sequence, Gly residues are abundant and represent potential sites for isonitrile-mediated ligation (Gly81, -87, -93, -125, -135, -149, -150).

54. Munson MC, García-Escheverría C, Albericio F, Barany G. J Org Chem. 1992; 57:3013–3018.

55. This affects quantitative removal of the Fmoc blocking group at Cys74 in all products derived from **12**. Additionally, this analysis provides insight into the ratio of product **14** formed relative to hydrolysis product, Gly125 carboxylic acid **15**, and piperidide **16**. Piperidide **16** is presumed to arise from the quenching of activated acyl intermediate(s).

56. Oxazolone formation in C-terminal activated peptides: DeTar DF, Silverstein R, Rogers FF. J Am Chem Soc. 1966; 88:1024–1030.

57. (a) Lu HS, Clogston CL, Narhi LO, Merewether LA, Pearl WR, Boone TC. J Biol Chem. 1992; 267:8770–8777. [PubMed: 1374379] (b) Brems DN. Protein Sci. 2002; 11:2504–2511. [PubMed: 12237471] (c) Tiwari K, Shebannavar S, Kattavarapu K, Pokalwar S, Mishra MK, Chauhan UK. Indian J Biochem Biophys. 2012; 49:285–288. [PubMed: 23077791] (d) Kim CK, Lee CH, Lee S-B, Oh J-W. PLoS One. 2013; 8:e80109. [PubMed: 24224041]

58. Attenuated yields may reflect attrition due to aggregation during isolation and purification by semipreparative HPLC.

59. Mo J, Tymiak AA, Chen G. Rapid Commun Mass Spectrom. 2013; 27:940–946. [PubMed: 23592195]

60. Shirafuji N, Asano S, Matsuda S, Watari K, Takaku F, Nagata S. Exp Hematol. 1989; 17:116–119. [PubMed: 2463930]

61. Of several potential factors, the known propensity of recombinant G-CSF **2b** to aggregate and precipitate from solution at higher pH (>5), as well as in the absence of stabilizing cofactors during HPLC purification, isolation, and prolonged storage, may explain the reduced bioactivity exhibited by synthetic G-CSF **2a**. Data from high-resolution mass spectrometry and proteolytic mapping support the primary sequence and disulfide connectivity in synthetic G-CSF **2a**. However, we cannot exclude the presence of misfolded intermediates that may also contribute to lower activity. Also, see ref 62.

62. Broxmeyer HE, Hoggatt J, O'Leary HA, Mantel C, Chitteti BR, Cooper S, Messina-Graham S, Hangoc G, Farag S, Rohrabaugh SL, Ou X, Speth J, Pelus LM, Srour EF, Campbell TB. Nat Med. 2012; 18:1786–1796. [PubMed: 23160239] For further discussion, see: Ou X, O'Leary HA, Broxmeyer HE. Blood. 2013; 122:161–169. [PubMed: 23637126]
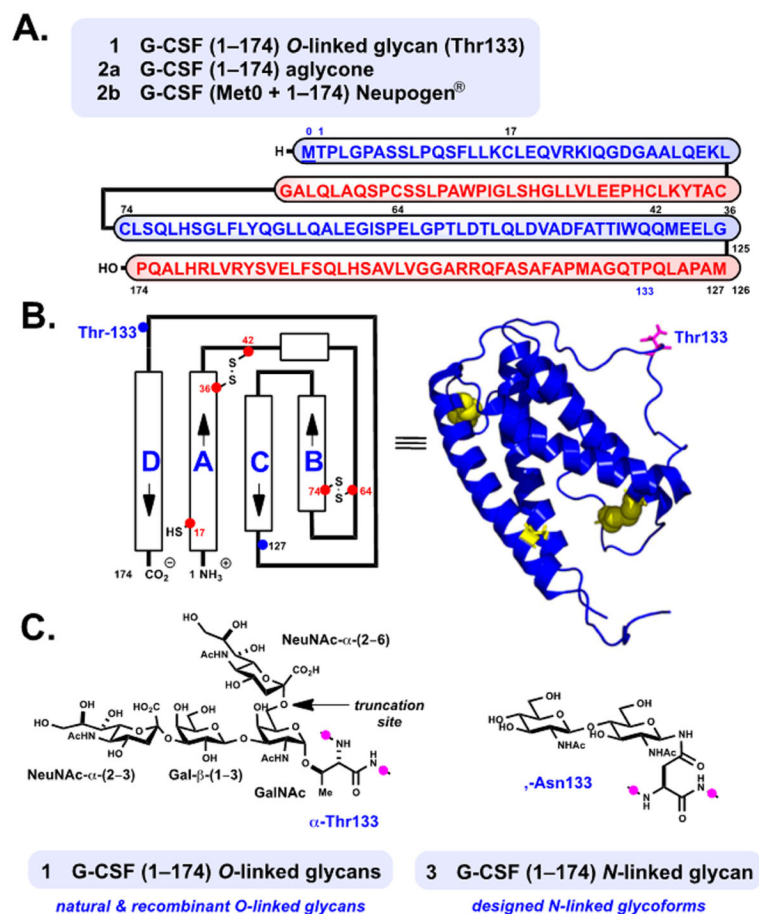
Figure 1.
Granulocyte colony-stimulating factor (G-CSF). Structure and therapeutic forms of **1**, **2**, and designed **3**. (A) Linear representation of G-CSF (Met0 + 1–174) with disulfide linkages removed for clarity (Cys36–Cys42, Cys64–Cys74). (B) Representation of protein structure and helix directionality.[3a] (C) Structure of the native O-linked carbohydrates at Thr133 (**1**) and targeted N-linked glycoform at Asn133 (**3**).
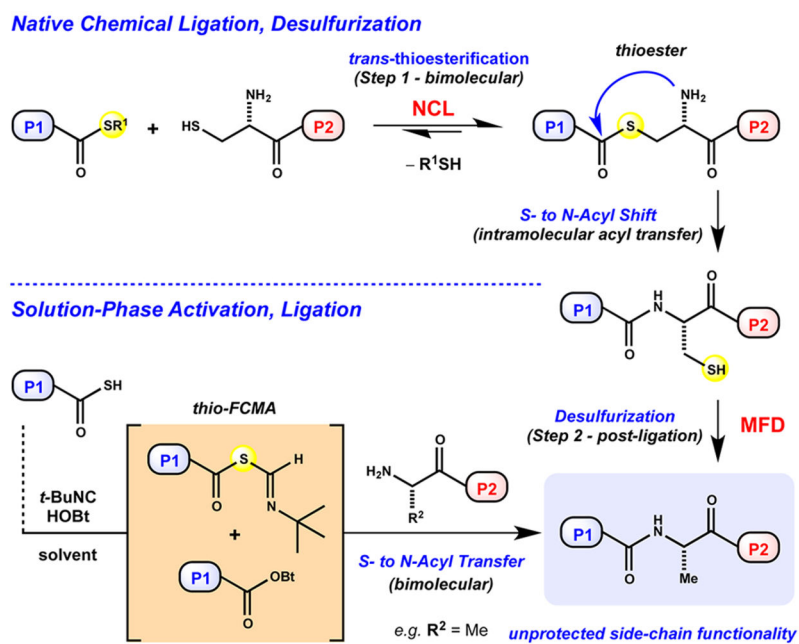
Figure 2.
Chemical protein synthesis methods.
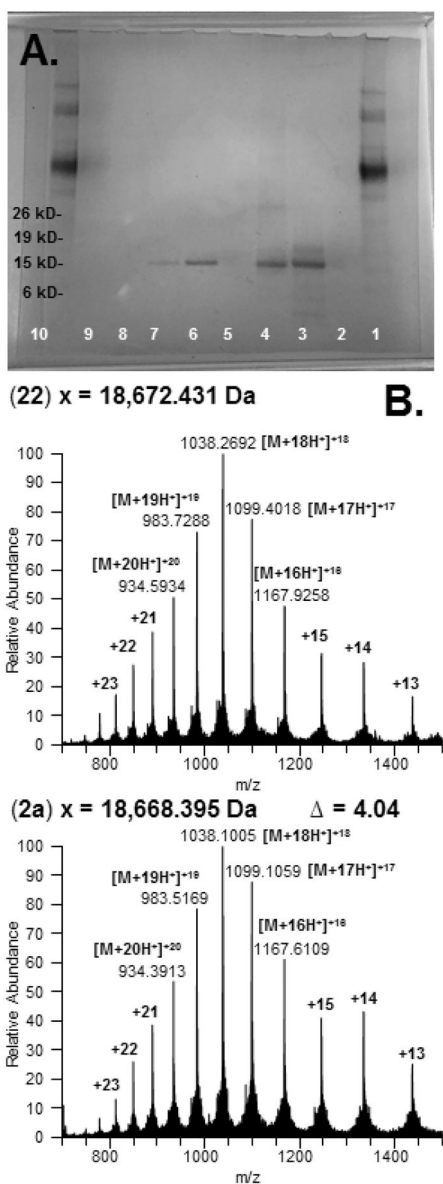
Figure 3.
Characterization of synthetic G-CSF 1–174 (**22**) and G-CSF (**2a**). (A) Silver-stained SDS-PAGE, NuPage 4–12% Bis-Tris gel: lanes 1 and 10, benchmark protein ladder (MW marker); lanes 2, 5, and 9, blank; lane 3, synthetic G-CSF 1–174 **22**; lane 4, synthetic G-CSF **2a**; lanes 6–8, recombinant G-CSF **2b** gradient. (B) High-resolution mass spectrometry: top, synthetic G-CSF 1–174 (**22**) and bottom, synthetic G-CSF 1–174 (**2a**).
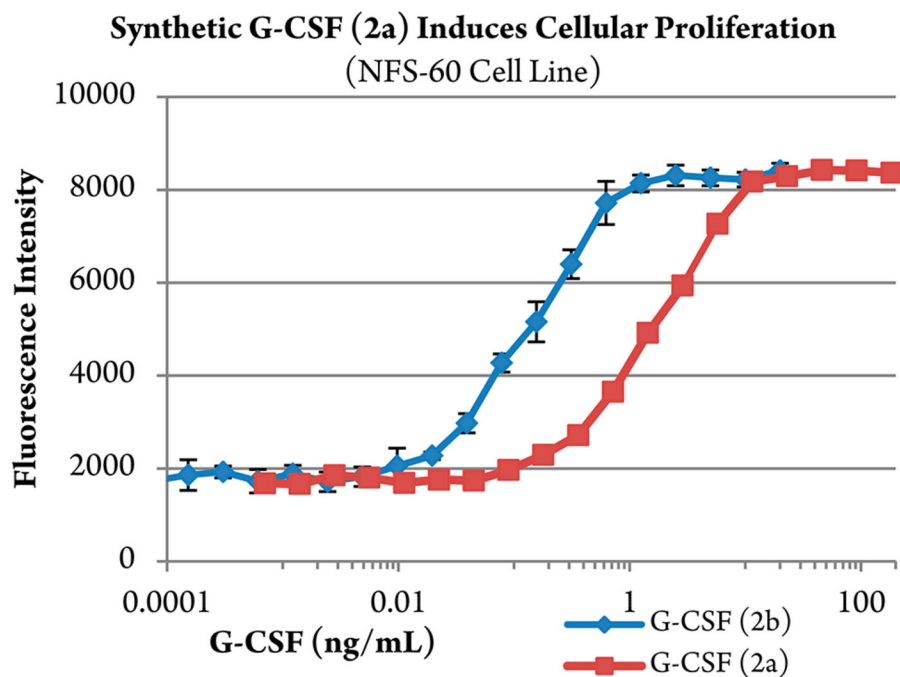
Figure 4.
Murine myeloblastic NFS-60 cellular proliferation assay. In vitro assay measuring the effect of synthetic G-CSF **2a** and recombinant G-CSF **2b** on the proliferation of NFS-60 cells. Results are expressed as average relative fluorescent intensity ± SD, run in triplicate.
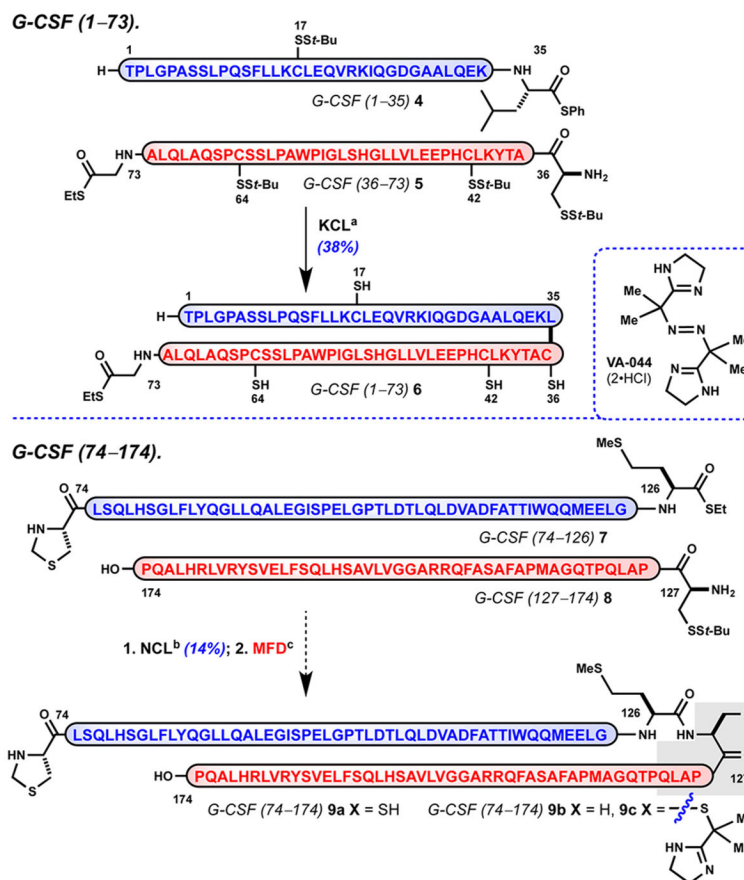
Scheme 1. Preliminary Approach to G-CSF 1–174 (2a) by Conventional Means[a]

[a]Reagents and conditions: (a) 1.5 equiv of **4** and 1.0 equiv of **5**, 6 M Gnd·HCl, 300 mM Na₂HPO₄, 20 mM TCEP, pH 7.2, isolated **6** (38%); (b) 1 equiv of **7** and 1.1 equiv of **8**, 6 M Gnd·HCl, 300 mM Na₂HPO₄, 20 mM TCEP, 200 mM MPAA, pH 7.6, isolated **9a** (14%); (c) **9a**, VA-044, *t*-BuSH, TCEP, solvent (see the Supporting Information); TCEP = tris(2-carboxyethyl)phosphine, MPAA = 4-mercaptophenylacetic acid, VA-044 = 2,2′-azobis[2-(2-imidazolin-2-yl)propane] dihydrochloride.

Scheme 2. Solution-Phase Ligation of Side-Chain-Unprotected Peptides: Optimized Synthesis of Aggregation-Prone 101-mer Polypeptide G-CSF 74–174 (14)[a]

[a]Additional reaction parameters: (a) Relative ratio of products **14/15**/ **16** determined by integration of UV trace at 280 nm (area under the curve); (b) isolated yield of peptide **14** following HPLC purification and lyophilization; (c)triethylamine (10 equiv) as an additive. EDC = 1-ethyl-3-(3-(dimethylamino)propyl)carbodiimide, HOOBt = hy-droxy-3,4-dihydro-4-oxo-1,2,3-benzotriazine, t-BuNC = tert-butyl isonitrile.
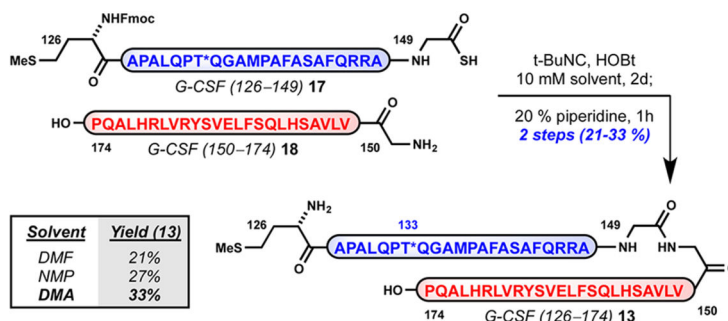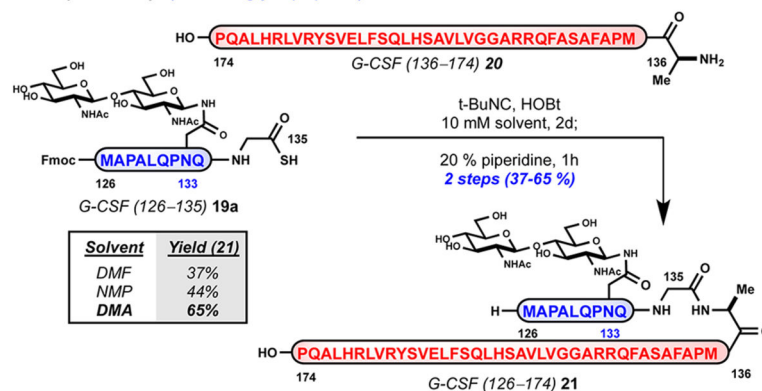
Scheme 3.
Solution-Phase Ligation of Side-Chain-Unprotected Peptides: Preparation of G-CSF 125–174 (13) and Glycopeptide (21)
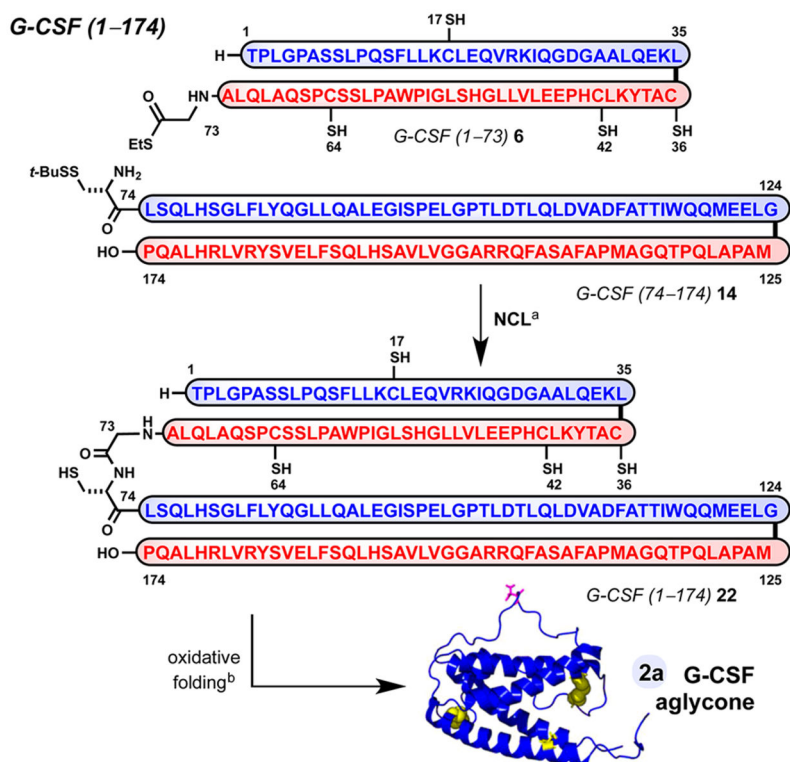
Scheme 4. Synthesis of G-CSF 1–174 (2a): Union of 6 and 14 by NCL, Oxidative Folding of G-CSF 22, and Isolation of Synthetic G-CSF 2a[a]

[a]Reagents and conditions: (a) 1.5 equiv of **6** and 1 equiv of **14**, 6 MGnd·HCl, 300 mM Na$_2$HPO$_4$, 40 mM TCEP, 200 mM MPAA, pH 6.8, isolated **22** (16–19%); (b) **22**, sarkosyl, H$_2$O; CuSO$_4$, H$_2$O, t = 20 h (see the Supporting Information); TCEP = tris(2-carboxyethyl)-phosphine, MPAA = 4-mercaptophenylacetic acid.