**Glycoproteins**

# Semisynthesis of Homogeneous, Active Granulocyte Colony-Stimulating Factor Glycoforms

*Lukas Kerul[+], Maximilian Schrems[+], Alanca Schmid, Rajeshwari Meli,*
*Christian F. W. Becker,\* and Claudia Bello\**

**Abstract:** Granulocyte colony stimulating factor (G-CSF) is a cytokine used to treat neutropenia. Different glycosylated and non-glycosylated variants of G-CSF for therapeutic application are currently generated by recombinant expression. Here, we describe our approaches to establish a first semisynthesis strategy to access the aglycone and *O*-glycoforms of G-CSF, thereby enabling the preparation of selectively and homogeneously post-translationally modified variants of this important cytokine. Eventually, we succeeded by combining selenocysteine ligation of a recombinantly produced N-terminal segment with a synthetic C-terminal part, transiently equipped with a side-chain-linked, photocleavable PEG moiety, at low concentration. The transient PEGylation enabled quantitative enzymatic elongation of the carbohydrate at Thr133. Overall, we were able to significantly reduce the problems related to the low solubility and the tendency to aggregate of the two protein segments, which allowed the preparation of four G-CSF variants that were successfully folded and demonstrated biological activity in cell proliferation assays.

## Introduction

Cytokines are small proteins or glycoproteins produced by specific cells of the immune system. They play a central role in cellular signaling as immune modulators and are involved in central biological events, such as hematopoiesis, inflammation and repair, survival and proliferation, differentiation commitment, maturation and functional activation of cells.[1–3] Among them, colony-stimulating factors (CSFs) are small glycoproteins essential for blood cell formation, immunocompetence and lymphocyte function.[4,5] The first four proteins identified in this group were given names indicating the major type of cells they stimulate: GM-CSF, stimulating granulocyte and macrophage colony formation, M-CSF, stimulating macrophage colony formation, G-CSF, granulocyte-CSF, and multi-CSF (now more commonly termed interleukin 3, IL3) stimulating a broad range of hematopoietic cell colony types. CSFs constitute an attractive class of biomolecules due to their therapeutic applications, ranging from inflammation, neurological diseases and autoimmunity.[6–8] In particular, different expressed variants of G-CSF are currently used for the treatment of chemotherapy-induced neutropenia, e.g. G-CSF aglycone from *E. coli* (filgrastim), its PEGylated version (pegfilgrastim) and a mixture of *O*-glycoforms of G-CSF isolated from CHO cells (lenograstim). The latter carry the following *O*-glycans: Neu5Ac-α(2-3)Gal-β(1-3)[Neu5Ac-α(2-6)]GalNAc and the truncated form, Neu5Ac-α(2-3)Gal-β(1-3)-GalNAc.[9] The physiological role of G-CSF *O*-glycosylation remains unclear although in vitro studies indicate a higher potency of glycosylated G-CSF (lenograstim), which might be due to an increased in vitro stability.[10–18] Reports based on clinical studies comparing activities of unmodified G-CSF (filgrastim) and *O*-glycosylated G-CSF (lenograstim) are contradictory with respect to beneficial effects of glycosylation.[19–22] The development of a robust method for the preparation of homogeneous glycoforms of G-CSF would help in shedding light on the role of individual glycosylation patterns and their impact on the physiological activity of this cytokine.

Whilst GM-CSF was chemically synthesized, including different homogeneous glycoforms,[23] the only reported total chemical synthesis of G-CSF aglycone[24] revealed the complexity of this task because of the aggregation tendency linked to limited solubility and structural complexity of the protein. To the best of our knowledge, the only reported example of homogeneous *O*-glycosylated G-CSF variant carrying an artificial mannose at Thr 133 was generated by

[*] Dr. L. Kerul,[+] M. Schrems,[+] A. Schmid, Dr. R. Meli,
Prof. Dr. C. F. W. Becker
Institute of Biological Chemistry, Faculty of Chemistry, University of Vienna
Währinger Str. 38, 1090 Vienna (Austria)
E-mail: christian.becker@univie.ac.at

Prof. Dr. C. Bello
Interdepartmental Research Unit of Peptide and Protein Chemistry and Biology, Department of Chemistry "Ugo Schiff", University of Florence
via della Lastruccia 13, 50019 Sesto Fiorentino (Florence) (Italy)
E-mail: claudia.bello@unifi.it

[+] These authors contributed equally to this work.

*Angew. Chem. Int. Ed.* **2022**, *61*, e202206116 (1 of 10)

© 2022 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

recombinant expression in yeast.[25] Native G-CSF contains 174 amino acids organized in four antiparallel alpha-helices (Scheme 1A); its tertiary structure is stabilized by two disulfide bridges (Cys36 to Cys42, &Cys-64 to Cys74) and the protein contains one additional free cysteine (Cys17)[26] as well as one O-glycosylation site at Thr133. The latter leads to G-CSF being present endogenously as a mixture of glycoforms that is almost indistinguishable from lenograstim by current analytical techniques.[9] To the best of our knowledge, no total chemical or semisynthesis of homogeneously glycosylated G-CSF variants has been reported so far.

Therefore, we set out to develop a semisynthetic route to G-CSF, including chemoenzymatic glycosylation of a transiently polyethylene-glycol modified (PEGylated) protein segment,[27] to demonstrate the power of combining these approaches. Together they should help to overcome solubility issues previously reported for G-CSF peptide segments[24] and avoid synthetic challenges within the N-terminal part of the protein.

Protein semisynthesis is particularly well-suited for accessing proteins carrying modifications only on a small to medium-sized segment, which is generated by chemical synthesis while the unmodified part is recombinantly expressed, thus simplifying the preparation of the whole protein.[28,29] Our strategy for the synthesis of homogeneous

G-CSF glycoforms involves the use of chemoenzymatic glycosylation for the preparation of homogeneous variants of the synthetic, O-glycosylated fragments starting from a synthetic peptide containing Thr-α-D-GalNAc (Tn antigen, Scheme 1B) in position 133. As previously demonstrated, tagging of peptides with monodisperse PEG via an (enzyme- or photo-)removable linker modulates the solubility properties of the peptides, enhancing solubility in aqueous solution and enabling full precipitation by addition of an appropriate organic solvent, thus greatly facilitating its enzymatic glycosylation and recovery of the glycopeptide product.[27,30,31] Furthermore, utilizing a 1-nitrophenyl-2-mercaptoethyl-based photocleavable ligation auxiliary linked to glycine (Aux-Gly) as PEG attachment point enabled the elongation of the glycosylated peptide by joining two or more (glyco)peptides via auxiliary-mediated native chemical ligation without the need for a cysteine residue.[31] Even though G-CSF contains five cysteine residues, they are inconveniently located within the sequence to chemically access Thr133.

Herein we describe the application of this strategy towards G-CSF, the various challenges met along the way and the development of a semisynthesis approach enabling the preparation of homogeneous O-glycosylated G-CSF variants.
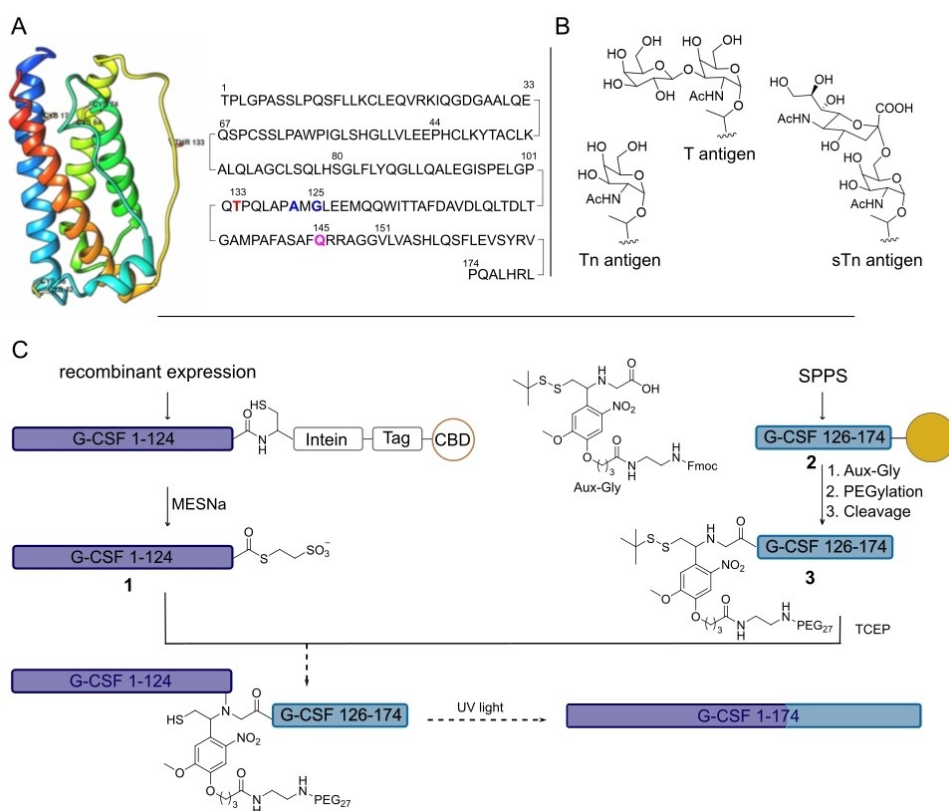


**Scheme 1.** A) Three-dimensional structure and sequence of granulocyte colony-stimulating factor (G-CSF). Crystal structure: PDB entry: 2D9Q. B) O-Glycans used in this study. C) First semisynthesis strategy for full-length G-CSF, the dashed arrows indicate that the products have not been obtained (CBD = Chitin Binding Domain). Highlighted amino acids in the G-CSF sequence indicate ligation sites used (blue), the glycosylation site (red) and the site of attachment of the photocleavable linker **14** (magenta).

## Results and Discussion

### Auxiliary-Mediated G-CSF Semisynthesis

The envisioned strategy involved an auxiliary-mediated native chemical ligation (NCL) as key step for the ligation of the recombinant and the synthetic fragments. The PEGylated Aux-Gly building block (Scheme 1C) should facilitate the chemoenzymatic glycosylation and purification of the C-terminal segment, by improving its solubility, and allow the NCL reaction at a non-Cysteine site as well as subsequent mild removal of the solubility tag. This strategy requires a glycine at the ligation site and we chose the Leu124-Gly125 junction. We thus recombinantly expressed the N-terminal segment comprising aa 1–124, and synthesized the C-terminal segment with aa 125–174 including the Aux-Gly building block[31] (Scheme 1C).

For the preparation of G-CSF1–124 α-thioester **1**, a G-CSF1–124-*Mxe*Intein-His-tag fusion construct was generated. Highest expression yields were obtained with the *E. coli* Rosetta 2(DE3) strain 4 h after induction at 37 °C. G-CSF1–124-*Mxe*Intein-His-tag was found in inclusion bodies and had to be re-solubilized, which lowered the yield. The construct was isolated via Ni-affinity chromatography and subsequently cleaved with 250 mM sodium mercaptoethane-sulfonate (MESNa) for 24 h to obtain G-CSF1–124 α-thioester **1**. The product was purified via RP-HPLC and characterized by LC-MS (Figure S3). Purified **1** was obtained with modest yields of $\approx 1$ mg per liter of culture.

The C-terminal fragment G-CSF126–174 **2** (Scheme 1C) was prepared via Fmoc-based solid phase peptide synthesis (SPPS) on a Pro TentaGel® R Trt resin, with low loading (see ESI). The Aux-Gly building block was installed via *O*-(7-azabenzotriazol-1-yl)-*N,N,N′,N′*-tetramethyluronium-hexafluorophosphate (HATU)-mediated coupling, followed by PEGylation with monodisperse Fmoc-NH-PEG$_{27}$-COOH and Fmoc removal to obtain PEG-Aux-G-CSF125–174 **3**. Cleavage of **3** from the resin was successfully performed using Reagent H without ammonium iodide (82.5% trifluoroacetic acid (TFA), 5% phenol, 5% thioanisole, 2.5% ethane-1,2-dithiol (EDT), 3% ddH$_2$O, 2% dimethyl sulfide (DMS)) as cleavage cocktail. In the presence of ammonium iodide we observed (most likely radical induced) cleavage of the 1-nitrophenyl-2-mercaptoethyl group. Consequently, ammonium iodide was omitted from the cleavage solution. The desired product **3** was purified by RP-HPLC and obtained in 14% yield (based on synthesis scale) with minor impurities corresponding to methionine oxidation ($\Delta m = +16$ Da) and TFA counter ion ($\Delta m = +114$ Da, Figure S13).

With both segments in hand, the first ligation was started by briefly incubating **3** in 0.2 M sodium phosphate buffer (NaPi), 35 mM TCEP at pH = 7.5 to remove the *tert*-butylsulfanyl group from the auxiliary. Subsequently, 1.5 eq. of **1** dissolved in 6 M guanidinium chloride (Gdn-HCl) buffer, (final concentrations: 2 mM **3**, 3 mM **1**, 4 M Gdn HCl, 0.2 M NaPi, 100 mM MESNA, 10 mM tris-(2-carboxyethyl) phosphine (TCEP), pH 7 at 30 °C). Here, a first challenge was immediately evident: the low solubility of thioester **1** and the long reaction time, due to the slow

ligation rates as a consequence of the 2-mercaptoethyl auxiliary[32] and the Leu-thioester,[33] led to aggregation and precipitation of the majority of **1**. Only traces of oxidized ligation product were detected. Unfortunately, we were unable to identify reaction conditions that would give acceptable yields even after screening a large set of conditions such as concentration of reactants and of chaotropes, thiol additives, reducing agents, buffer systems, pH and temperature (see Table S1).

### Selenocysteine-Mediated G-CSF Semisynthesis

Therefore, we had to consider alternative routes to obtain G-CSF by accelerating the ligation reaction, thus limiting the risk of aggregation by speeding up the reaction and avoiding the Leu-Aux-Gly ligation site (Scheme 2). Due to the inconveniently placed cysteine residues and the need to maintain them in the final product we opted for a selenocysteine (Sec)-mediated ligation, as Sec can be selectively deselenized to give Ala in a process that leaves all native cysteine residues untouched.[34–38] Consequently, we moved the ligation site two residues towards the C-terminus to generate a Met-Ala junction (Scheme 1A). For the preparation of the elongated N-terminal segment, the available G-CSF1–124 gene was elongated by two amino acids.

The resulting G-CSF1–126 gene was cloned into a pTXB1 vector to generate a G-CSF1–126-*Mxe*Intein-His-tag fusion construct. Expression and purification were performed as described for segment **1** and the corresponding thioester **4** was obtained with a yield of 2 mg L⁻¹ of culture, after cleavage with MESNa and purification by RP-HPLC (Figure 1A).

Viability of this new synthesis route was assessed with unmodified and monoglycosylated C-terminal Sec-G-CSF128–174 peptides, as these segments do not require enzymatic glycosylation. Both C-terminal peptides Sec-G-CSF128–174 **5** and Sec-G-CSF128–174-Tn **6** were synthesized by SPPS (Scheme 2). Fmoc-Thr(GalNAc-3Ac)-OH was used as building block in order to introduce the glycosylation at position 133 of **6** and Fmoc-Sec(Mob)-OH for the addition of Sec at the N-terminus of both peptides. Crude peptides were obtained in good purity after cleavage from the solid support and deprotection. After Sec trans-protection[39] using 50 eq. of 2,2′-dithio-bis-(5-nitropyridin) (DTNP) in TFA for 30 min, dithiothreitol (DTT) treatment, and purification via RP-HPLC, fully deprotected dimers of Sec-G-CSF128–174 **5′** and Sec-G-CSF128–174-(Tn) **6′** (Scheme 2A & Figure 1B,C) were obtained in 11% and 8% yield, respectively (based on synthesis scale).

Ligation conditions were explored with segments **4** and **5′** and best results were obtained using equimolar amounts of both segments at 1.5 mM concentration in ligation buffer (6 M Gdn-HCl, 100 mM NaPi, 50 mM TCEP, pH 7), with addition of 4-mercaptophenylacetic acid (MPAA, 250 mM) to speed up the reaction, and of sodium ascorbate (100 mM) as radical quencher to avoid the premature radical deselenization of Sec-G-CSF128–174 **5′** by TCEP.[38] On the contrary
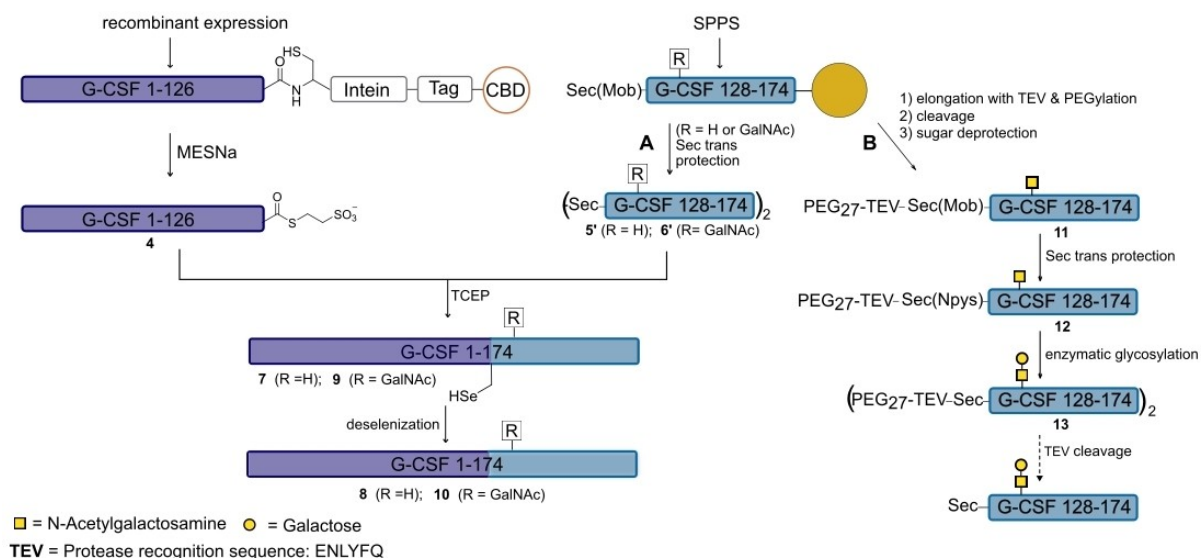
**Scheme 2.** Alternative approaches for the preparation of full-length, homogeneously glycosylated G-CSF. A) Sec-mediated ligation of chemically synthesized peptide **5'** (dimer of Sec-G-CSF128–174) or glycosylated peptide **6'** (dimer of Sec-G-CSF128–174-(Tn)), with thioester **4**. B) Preparation of the C-terminal fragment carrying a T-antigen at position 133 and the PEG chain attached to the N-terminus elongated by a TEV protease recognition sequence. The dashed arrows indicate that the products have not been obtained
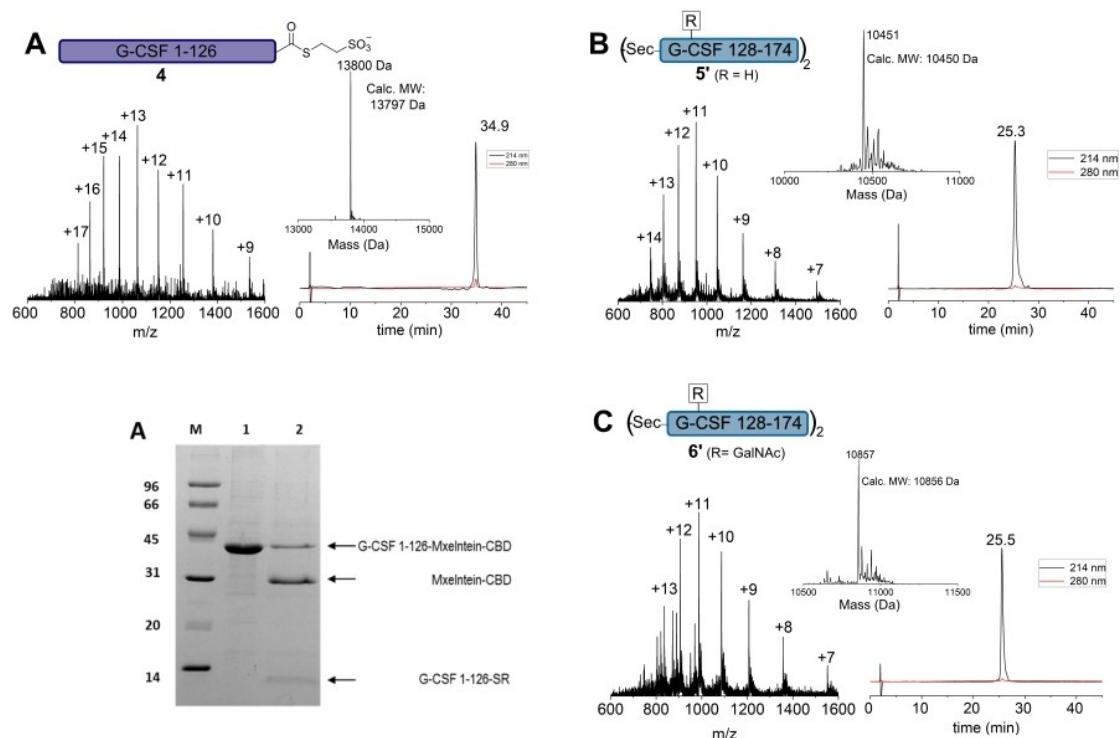


**Figure 1.** A) RP-HPLC chromatogram and MS analysis of purified thioester **4**, and SDS-PAGE of thiol induced cleavage of G-CSF1–126-*Mxe*Intein-CBD (cleavage efficiency: 78%; lane M: low molecular weight (LMW) marker, lane 1: 0 h reaction time, lane 2: 24 h reaction time). B) RP-HPLC chromatograms and MS analyses of deprotected dimers of Sec-G-CSF128–174 **5'** and C) Sec-G-CSF128–174-(Tn) **6'**. G-CSF1–126-SR = **4**; SR = C-terminal α-thioester.

to what we observed in the case of the auxiliary-mediated ligation, in which only traces of product were observed after prolonged time, the Sec-mediated ligation proceeded at acceptable speed thereby also limiting the precipitation of protein segments (Figure S20). Thioester hydrolysis was also observed, which initiated the addition of more **4** over time. However, this did not improve the overall yield, since G-CSF-1–126 (hydrolyzed thioester) partially coeluted with the ligation product complicating purification. After **4** was almost completely consumed (22 h), the reaction was quenched by addition of TCEP and an excess of 6 M Gdn-HCl buffer at pH 4.7. Ligation product **7** was recovered in 44 % yield after purification via RP-HPLC (Figure S20).

Aggregation of the ligation product and co-elution of the hydrolyzed thioester during RP-HPLC purification limited the amount of recovered product **7**. Due to its poor solubility full-length G-CSF could only be dissolved in reducing Gdn-HCl buffer (containing 25 mM DTT), in which it was deselenized using conditions reported by Metanis *et al.*[40] Based on the limited solubility of G-CSF1–174(Ala127Sec), concentrations of **7** had to be 10 times lower than originally reported for this procedure.[38,40]

The reaction was finished after 3 h (Figure S20) and product **8** was directly used in folding experiments without further purification (see below).

Similar conditions were used for ligating **4** and **6′** (Scheme 2) to give G-CSF with GalNAc in position 133. This reaction was finished after 12 h (Figure S21) and the product was recovered in 48 % yield after purification via RP-HPLC. The obtained ligation product **9** was deselenized giving **10** (Figure S21) and folded as described for **8**. Being able to obtain folded full-length G-CSF via thioester-Sec ligation allowed us to further develop this strategy to temporarily improve solubility via a solubility tag that would allow additional chemoenzymatic glycosylation steps.

This was necessary as enzymatic glycosylation with the non-PEGylated **6** to elongate the α-D-GalNAc on Thr 133 to β-D-Gal-1,3-α-D-GalNAc (T antigen, Scheme 1B) gave only incomplete reactions and precipitation of the starting material. In a first approach, we reverted to a strategy previously used for the chemoenzymatic glycosylation of mucin peptides.[30] The PEG polymer was added to the N-terminus of Sec-G-CSF128–174-Tn on resin including a TEV protease recognition sequence. Coupling of PEG and of additional amino acids was achieved under base-free conditions with Oxyma Pure and DIC at rt to avoid racemization and deselenization.[41] After cleavage from the resin, deprotection of the acetyl groups on the sugar with sodium methanolate to give product **11** (Scheme 2B), and trans-protection of Sec with DTNP, the obtained **12** (Figure S14) was purified via RP-HPLC and recovered in 8 % yield (based on synthesis scale). To ensure full solubility under conditions suitable for enzymatic glycosylation, the peptide was dissolved in 8 M urea and subsequently diluted with glycosylation buffer (75 mM TRIS (pH = 7.5), 0.06 % Triton X-100, 10 mM MnCl$_2$, 2 mM UDP-Gal, 0.01 mg mL$^{-1}$ human C1GalT1) to a final concentration of 2 M urea. Quantitative conversion of PEG-TEV-Sec-G-CSF128–174-Tn **12** to PEG-TEV-Sec-G-CSF128–174-T **13** was achieved after 2 h (Fig-

ure S15). The product was recovered by precipitation with ethanol and diethyl ether at low temperature after centrifugation.[27] Unfortunately, we could not identify conditions to remove the N-terminal PEG chain by TEV-protease digestion from PEG-TEV-Sec-G-CSF128–174 carrying a β-D-Gal-1,3-α-D-GalNAc disaccharide at Thr133 (Scheme 2B), even though we started out with conditions previously described to cleave a TEV site before Sec.[42] A variety of conditions were tested (different temperatures, ranging from 4 °C to 34 °C, various amount of reducing agent (0 –10 mM DTT) and addition of urea (2 M) to avoid precipitation) but none of the tested conditions led to detectable amounts of the desired product. This behavior of the PEGylated G-CSF peptide **13** could be explained by limited accessibility of the cleavage site to the enzyme because of the tendency of the peptide to aggregate or due to formation of the diselenide.

### Side-Chain Solubilization Tag

In our different synthesis attempts of G-CSF we verified on the one hand that Sec-mediated ligation is effective in generating the full-length protein and on the other that the presence of the PEG polymer is essential for enzymatic glycosylation. In order to take advantage of both strategies, we designed a photocleavable linker that can carry the PEG moiety without interfering with ligations, is attached to an amino acid side chain and that can be easily removed after assembly of G-CSF. Removable backbone or side-chain modifications based on oligoethylene glycol[43] or on poly-Lys/Arg tags.[44–49] have been previously used to increase the solubility of peptides. Photocleavable linker **14** was synthesized in 7 steps with an overall yield of 42 % on a multigram scale (Scheme 3; for details on the synthesis see Supporting Information Scheme S1). Sec-G-CSF128–174-Tn was prepared by SPPS including a Glu(OAll) building block at position 145 for installing **14**.

Linker **14** was coupled to the side chain of Glu145 after allyl group removal. Deprotection of the linker on resin allowed efficient coupling of PEG$_{27}$ (Scheme 3). The Mob-protected Sec-G-CSF128–145(Aux′-PEG$_{27}$)-174-Tn **15** was removed from the solid support and recovered in excellent purity in 11 % yield (based on synthesis scale) after removal of the acetyl protecting groups on the sugar and purification via RP-HPLC (Figure 2).

Enzymatic glycosylation was carried out using conditions previously described.[27] Fortunately, the new side-chain-linked solubility tag considerably increased the solubility of **15** under glycosylation conditions and complete conversion was achieved after 4 h, with the product remaining fully soluble. In order to generate sufficient amount of peptide, we reacted **15** at higher concentration (1 mM) and with less activated sugar (UDP-Gal at 1 mM). After 6 h starting material **15** was completely consumed and the solution containing Mob-protected Sec-G-SCF128–145(Aux′-PEG$_{27}$)-174-T Mob-**16** was lyophilized and treated with DTNP in TFA for Sec trans-protection to give Sec(Npys)-G-CSF128–145(Aux′-PEG$_{27}$)-174-T Npys-**16**. Subsequently, Npys-**16**
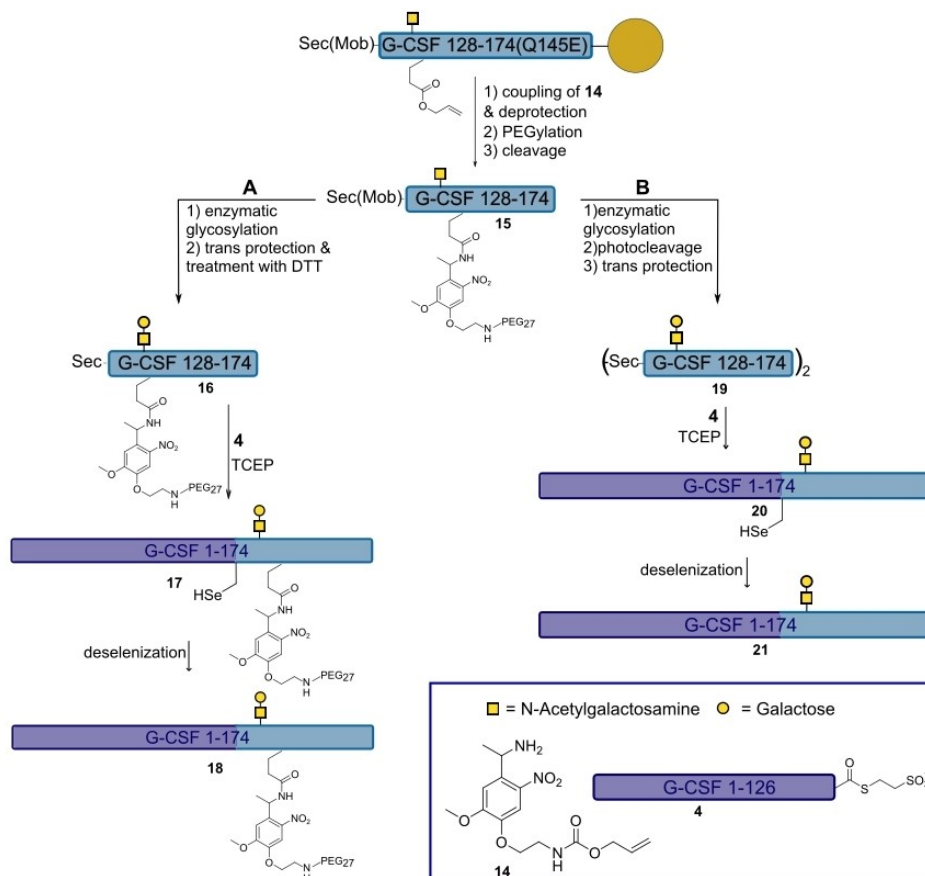
**Scheme 3.** Preparation of glycosylated variants of the C-terminal fragment carrying a T-antigen at position 133 and a photoremovable side-chain linker as solubility tag. A) Variant **16** is ligated to thioester **4**. The side-chain-attached PEG could not be removed from full-length deselenized product **18**. B) Variant **19**, obtained from **15** by photocleavage and transprotection, is ligated to thioester **4** giving full-length G-CSF modified with a β-D-Gal-1,3-α-D-GalNAc at Thr133 **21** after deselenization of the ligated product.
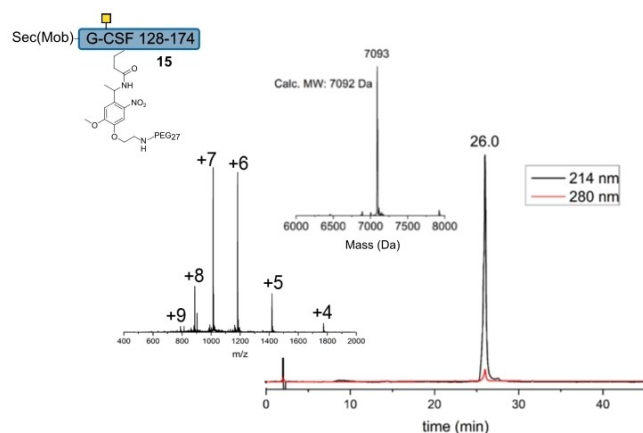


**Figure 2.** RP-HPLC chromatogram and ESI-MS of purified Sec-G-CSF128–145(Aux′-PEG$_{27}$)-174-(Tn) **15**, expected mass: 7092 Da, observed mass: 7093 Da, error of the instrument 5 Da.

was precipitated with diethyl ether, centrifuged and the pellet resuspended in a 6 M Gdn-HCl solution (pH=4.7) containing 10 mM DTT. Unprotected Sec-G-CSF128–

145(Aux′-PEG$_{27}$)-174-(T) **16** (Scheme 3A) was obtained with a yield of 19 % (based on synthesis scale) after RP-HPLC purification (Figure S16) and used in the ligation reaction with G-CSF1–126 thioester **4**.

Due to the improved solubility of **16** compared to **6**, a higher concentration (2 mM) could be reached, thus the ligation was conducted using the thioester **4** as limiting reagent in order to reduce the problem of partial coelution of the ligation product with hydrolyzed **4**. After 20 h the reaction was quenched by addition of TCEP and acidic Gdn-HCl solution. The crude reaction mixture was purified via RP-HPLC giving full length G-CSF1–126-Sec-128–145(Aux′-PEG$_{27}$)174-(T) **17** in 55 % yield. Deselenization of **17** gave **18** in excellent conversion (Figure S22). Unfortunately, the final step of UV-light-induced removal of the Aux′-PEG$_{27}$ tag was not as straightforward as anticipated and a variety of conditions had to be screened to give satisfactory yields. We added TCEP and sodium ascorbate to the reaction mix, the first to keep the five cysteines in reduced form, the second to protect the thiol groups from radical desulfurization, but none of these precautions could prevent considerable formation of undesired side products, mainly desulfurized starting material. The same complex

*Angew. Chem. Int. Ed.* **2022**, *61*, e202206116 (6 of 10)

© 2022 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

mixture was obtained when **17** was used in the photo-cleavage reaction before deselenization.

To circumvent this problem, we then tried to remove the PEGylated Aux′ before ligation (Scheme 3B), thus on crude Mob-Sec-G-CSF128–145(Aux′-PEG$_{27}$)-174-(T) Mob-**16** that does not contain any cysteine residues. Photocleavage in water led to precipitation of the product that then could not be dissolved anymore. The reaction was thus conducted in 6 M Gdn HCl solution and was finished in 2.5 h. Product **19** could be obtained in 6 % yield (based on synthesis scale) after purification via RP-HPLC. After DTNP trans-protection the product was precipitated with diethyl ether, centrifuged and lyophilized, and then used in the ligation reaction with thioester **4** (Figure S23). After 20 h the ligation was finished and pure G-CSF1–174(Ala127Sec)-(T) **20** was recovered in 18 % yield after purification via RP-HPLC. Losses of product are probably due to its tendency to aggregate on the HPLC column and to the problematic partial coelution of hydrolyzed thioester that reduced the collection window for the ligation product. For the subsequent deselenization reaction, **20** was dissolved in 6 M Gdn-HCl, 25 mM DTT (49 eq.), 200 mM NaP$_i$ at pH 7.2. After 10 minutes a solution of 6 M Gdn-HCl, 200 mM TCEP (420 eq.), 200 mM NaP$_i$ at pH 5 was added. The starting material was completely deselenized in 30 min and G-CSF1–174-(T) **21** was obtained in moderate purity (Figure S23). Impurities are mostly based on a mixture of desulfurization and oxidation reactions.

Folding of all three semisynthetic G-CSF variants **8**, **10** and **21** was achieved following published protocols by dialysis of a 1 mg ml$^{-1}$ solution of the polypeptide in 6 M Gdn-HCl (pH 4.7) against a buffer containing sarkosyl and catalytic amounts of CuSO$_4$ (2 % Sarkosyl, 50 mM TRIS, 40 μM CuSO$_4$, pH 8.0), in order to remove Gdn-HCl and to form the correct disulfide bridges.[24] This process was followed by RP-HPLC (Figure S25). Retention times of the folded protein variants shifted towards lower values because of the decreased exposure of hydrophobic residues after folding. This was observed for all G-CSF variants. Retention time of the folded aglycone (21.7 min, Figure S25A) is almost identical to the commercially available and chemically identical G-CSF aglycone NEUPOGEN® (21.8 min, Figure S25A).

The retention time of the folded glycosylated variants is decreased compared to NEUPOGEN® because of the hydrophilic sugar moieties present in the homogeneously glycosylated G-CSF variants. After RP-HPLC purification, the folded G-CSF variants were concentrated, buffer was exchanged to 10 mM sodium acetate buffer and concentration of the variants was measured at 280 nm. The folded **8**, **10** and **21**, were successfully obtained in 11 %, 5 % and 6 % yield, respectively.

The yields were lowered by the poor solubility of the folded products and by losses during HPLC purification, but they remain in good agreement with literature reports for folding experiments.[24] The samples were then further analyzed via CD, HPLC, ESI-MS and SDS-PAGE (Figure 3).

The secondary structure of the variants does not seem to be affected by the glycosylation as all three variants show similar CD spectra and are also almost identical to NEUPOGEN® with absorption minima at 222 nm and 208 nm and a maximum at 193 nm, characteristic for α-helical proteins.[50]

### Expressed Protein Selenoester Ligation

To further facilitate access to G-CSF variants with variable carbohydrates modifications, we explored an alternative route based on the recently introduced expressed protein selenoester ligation (EPSL).[51] This method combines expressed protein ligation with diselenide-selenoester ligation, thus enabling fast ligations of recombinant and synthetic segments at low concentration, helping to overcome limited solubility of the starting materials. We employed EPSL (Scheme 4) for the preparation of another sialylated G-CSF variant carrying a Neu5Ac-α(2–6)-GalNAc glycan at Thr 133 (Scheme 1B, Figure S24). G-CSF(1–126)-selenoester **23** was prepared from the G-CSF1–126-*Mxe*Intein-His-tag fusion construct by addition of a hydrazine solution (5 % hydrazine, 100 mM DTT) at room temperature for 24 h to produce the precursor G-CSF1–126 hydrazide **22**, which was obtained with a yield of 3.2 mg per liter of culture after HPLC purification (Figure S4). G-CSF1–126 selenoester **23** was prepared by treatment of **22** with acetylacetone in the presence of diphenyl diselenide (DPDS, 50 mM), in 6 M Gdn·HCl, 200 mM HEPES buffer containing TCEP at pH 1.5. After 2.5 h the solution was extracted with hexane and pure **23** was recovered after RP-HPLC of the aqueous layer and subsequent lyophilization (Figure S5). The C-terminal fragment Sec-G-CSF128–145(-Aux′-PEG$_{27}$)-174-(sTn) **24** was obtained through highly efficient enzymatic glycosylation of **15** using the enzyme ST6GalNAc1 to transfer sialic acid to position 6 of the GalNAc in 24 h at 37 °C (Figure S17). After lyophilization, crude **24** was directly submitted to cleavage of the PEGylated linker via irradiation at 365 nm (Figure S19). After 2 h, Mob-protected Sec-G-CSF128–174-(sTn) **25** was purified via RP-HPLC and lyophilized. Subsequently, the peptide was treated with DPDS in TFA, in the presence of thioanisole and DMS, for Sec trans-protection. Deprotected Sec-G-CSF128–174-(sTn) **26** was directly used in the EPSL reaction with selenoester **23**. EPSL of G-CSF segments was two-times faster than NCL and selenocysteine-mediated ligation reaction, and was shown to be very efficient even at much lower concentration than generally needed for NCL,[51] clearly demonstrating the opportunities within EPSL for challenging ligation reactions. The ligation solution was extracted with diethyl ether and diluted with degassed reducing buffer (0.25 M TCEP, 25 mM DTT in 6 M Gdn·HCl, pH 7.7) for direct deselenization of the ligation product within 7 h. Despite the poor solubility of the protein and its tendency to aggregate on the HPLC column, leading to reduced recovery, sialylated full length G-CSF-(sTn) **27** was obtained in acceptable purity (Figure S24). However, the mass spectrum of the final product showed signals corresponding to G-CSF-(Tn) **10**
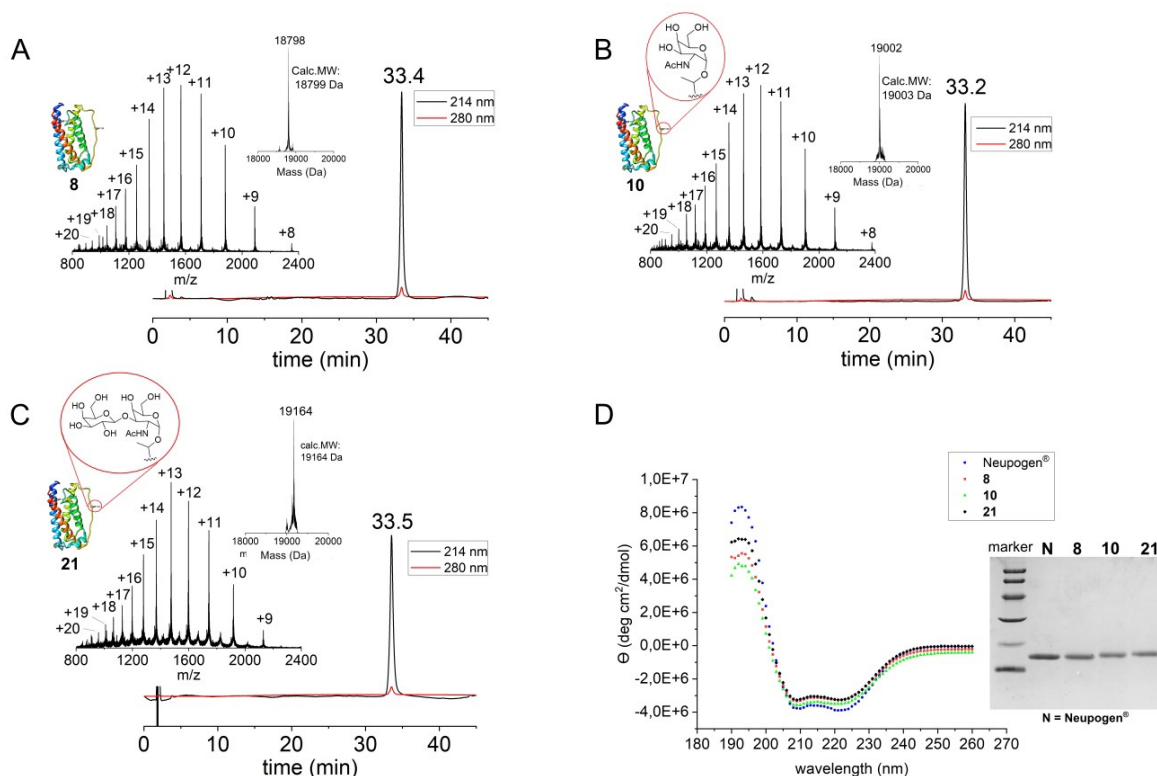
**Figure 3.** HPLC and ESI-MS of full-length G-CSF aglycon **8** (A) and homogeneously glycosylated G-CSF-(Tn) **10** (B) and G-CSF-(T) **21** (C). CD spectra and SDS-PAGE of the three variants (D); legend for the SDS-PAGE in (D): marker = low molecular weight marker, N = Neupogen, **8**, **10**, **21** = G-CSF variant **8**, **10** and **21**, respectively.

(Figure S24), indicating fragmentation during analysis or during ligation and handling. The latter is less likely as the sialylated peptide remained intact under similar conditions.

Folding of sialylated G-CSF **27** was achieved under the same conditions used for aglycone **8** and glycoforms **10** and **21** in 9% yield (Figure S25). Considering the yields obtained for the folding reactions of **8**, **10** and **21**, we conclude that the use of the EPSL reaction is compatible with generating glycoproteins and constitutes a considerable improvement with respect to reaction time and conversion. The faster ligation reaction reduces precipitation of badly soluble protein segments, as well as degradation of sensitive functional groups, thus enhancing recovery of the product.

Semisynthetic G-CSFs **8**, **10** and **21** were further tested for bioactivity in proliferation assays with murine myeloblastic NFS-60 cells, in direct comparison to the commercially available G-CSF aglycone Neupogen[®][52] (Figure 4). All three (glyco) variants show significant biological activity in the proliferation assays even though the two glycovariants are less active in this assay. As previous studies have shown that glycosylation typically increases stability in cell proliferation assays, this lower activity could be explained by either concentration effects due to the low amounts of protein used or, more interestingly, due the nature and positioning of the carbohydrates, which have not yet been studied as homogeneous species. The latter will be explored next, also with the now available sialylated G-CSF variants.
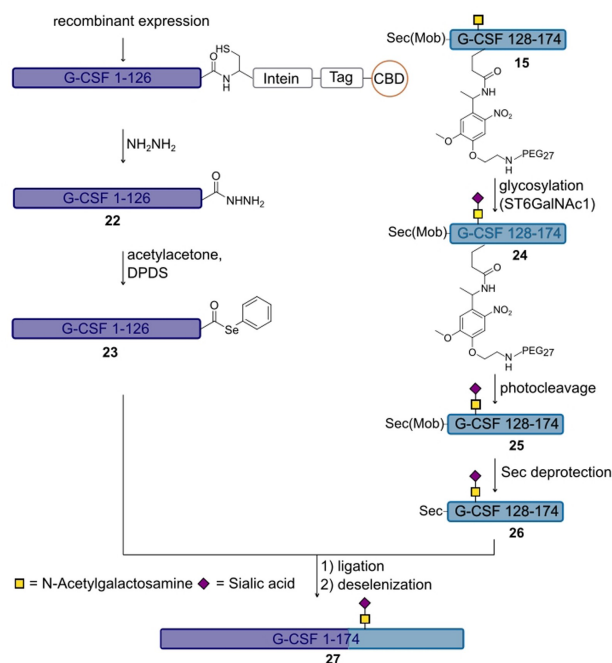


**Scheme 4.** Semisynthesis of sialylated G-CSF variant **27** via expressed protein selenoester ligation.
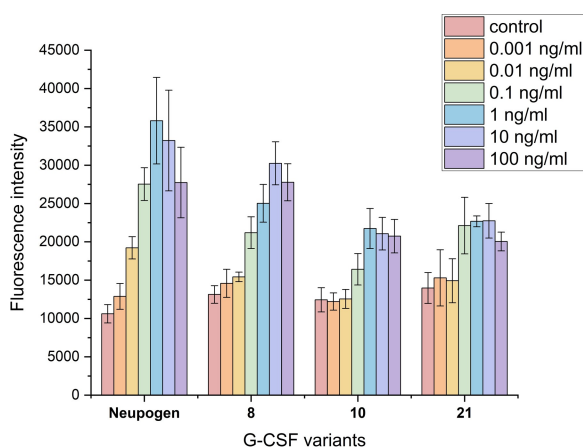
**Figure 4.** In vitro cellular proliferation assay of murine myeloblastic NFS-60 cells with semisynthetic G-CSF variants and Neupogen® as positive control. For Neupogen®, compound **8** and compound **10**, 7 independent measurements were carried out in triplicate for each concentration tested; for compound **21**, 3 independent measurements were carried out in triplicate for each concentration tested. Reported data are means ± SD of the mean.

## Conclusion

Access to homogeneous post-translationally modified hydrophobic proteins remains a significant challenge and requires a complex set of peptide and protein chemistry tools. In approaching the semisynthesis of different glycosylated variants of G-CSF we relied on the ability of a monodisperse PEG chain to transiently modulate the solubility of the segment obtained by SPPS, thus supporting its quantitative enzymatic glycosylation. Our semisynthesis strategy had to be adjusted along the way to overcome challenges due to poor solubility and the tendency to aggregation of the two segments to be ligated as well as of the full-length product. Overall, we could obtain three fully functional G-CSF variants that showed full biological activity. To the best of our knowledge this is the first semisynthetic strategy to access G-CSF and the first time that homogeneous, glycosylated variants of G-CSF have been obtained besides an artificially *O*-mannosylated G-CSF from yeast.[25]

Beyond the successful semisynthesis of G-CSF, our findings indicate the need for more general semisynthesis approaches that work at low concentrations and under conditions, which can keep hydrophobic/aggregation-prone peptides and proteins in solution without the need for complex modification reactions. To this end, we established access to G-CSF by using the recently introduced EPSL approach,[51] thereby demonstrating the high potential of this method for generating challenging protein targets with complex modifications.

## Acknowledgements

## Conflict of Interest

The authors declare no conflict of interest.

## Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

[1] J.-X. Lin, W. J. Leonard, *Annu. Rev. Immunol.* **2019**, *37*, 295–324.

[2] J. J. O'Shea, M. Gadina, R. M. Siegel, in *Clinical Immunology*, 5th ed. (Eds.: R. R. Rich, T. A. Fleisher, W. T. Shearer, H. W. Schroeder, A. J. Frew, C. M. Weyand), Elsevier, London, **2019**, pp. 127–155.e1.

[3] M. M. Khan, in *Immunopharmacology* (Ed.: M. M. Khan), Springer International Publishing, Cham, **2016**, pp. 57–92.

[4] E. Trus, S. Basta, K. Gee, *Cytokine* **2020**, *127*, 154939.

[5] J. M. Garland, *Colony-Stimulating Factors: Molecular & Cellular Biology, Second Edition*, CRC Press, Boca Raton, **2020**.

[6] A. Singhal, M. Subramanian, *Cytokine* **2019**, *122*, 154190.

[7] P. Mehta, J. C. Porter, J. J. Manson, J. D. Isaacs, P. J. M. Openshaw, I. B. McInnes, C. Summers, R. C. Chambers, *Lancet Respir. Med.* **2020**, *8*, 822–830.

[8] M. Kostic, N. Zivkovic, A. Cvetanovic, I. Stojanovic, *J. Neuroimmunol.* **2018**, *323*, 1–9.

[9] M. Oheda, S. Hase, M. Ono, T. Ikenaka, *J. Biochem.* **1988**, *103*, 544–546.

[10] C. Nissen, *Eur. J. Cancer* **1994**, *30A*, S12–S14.

[11] M. Oh-eda, M. Hasegawa, K. Hattori, H. Kuboniwa, T. Kojima, T. Orita, K. Tomonou, T. Yamazaki, N. Ochi, *J. Biol. Chem.* **1990**, *265*, 11432–11435.

[12] M. Ono, *Eur. J. Cancer* **1994**, *30A*, S7–S11.

[13] P. Pedrazzoli, N. Gibelli, L. Pavesi, P. Preti, M. Piolini, F. Bertolini, G. Robustelli della Cuna, *Anticancer Res.* **1996**, *16*, 1781–1785.

[14] A. R. Mire-Sluis, R. G. Das, R. Thorpe, *J. Immunol. Methods* **1995**, *179*, 117–126.

[15] S. Querol, J. A. Cancelas, L. Amat, G. Capmany, J. Garcia, *Haematologica* **1999**, *84*, 493–498.

[16] C. R. D. Carter, K. M. Whitmore, R. Thorpe, *J. Leukocyte Biol.* **2004**, *75*, 515–522.

[17] L. Mattii, A. Azzarà, R. Fazzi, G. Carulli, M. Chimenti, N. Cecconi, S. Galimberti, M. Petrini, *Leuk. Res.* **2005**, *29*, 1285–1292.

[18] F. Martin-Christin, *Anti-Cancer Drugs* **2001**, *12*, 185–191.

[19] R. Farhan, E. Urbanowska, H. Zborowska, M. Król, M. Król, T. Torosian, I. Piotrowska, K. Bogusz, K. Skwierawska, W. Wiktor-Jędrzejczak, E. Snarski, *Ann. Hematol.* **2017**, *96*, 1735–1739.

[20] S. Sivgin, E. Karakus, M. Keklik, G. Zararsiz, M. Solmaz, L. Kaynar, B. Eser, M. Cetin, A. Unal, *Transfusion Apheresis Sci.* **2016**, *54*, 410–415.

[21] G. M. Keating, *Drugs* **2011**, *71*, 679–707.

[22] F. Marchesi, M. Vacca, D. Giannarelli, F. Ipsevich, A. Pandolfi, S. Gumenyuk, D. Renzi, F. Palombi, F. Pisani, A. Romano, *Transfusion* **2018**, *58*, 1143–1148.

[23] Q. Zhang, E. V. Johnston, J.-H. Shieh, M. A. S. Moore, S. J. Danishefsky, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 2885–2890.

[24] A. G. Roberts, E. V. Johnston, J. H. Shieh, J. P. Sondey, R. C. Hendrickson, M. A. Moore, S. J. Danishefsky, *J. Am. Chem. Soc.* **2015**, *137*, 13167–13175.

[25] B. Gong, I. Burnina, H. Lynaugh, H. Li, *J. Chromatogr. B* **2014**, *945–946*, 135–140. Note: the glycosylation site here is indicated as Thr 134 instead of Thr 133. The different numbers result from counting the N-terminal Met.

[26] N. Kubota, T. Orita, K. Hattori, M. Oh-eda, N. Ochi, T. Yamazaki, *J. Biochem.* **1990**, *107*, 486–492.

[27] C. Bello, S. Wang, L. Meng, K. W. Moremen, C. F. Becker, *Angew. Chem. Int. Ed.* **2015**, *54*, 7711–7715; *Angew. Chem.* **2015**, *127*, 7823–7828.

[28] A. C. Conibear, E. E. Watson, R. J. Payne, C. F. W. Becker, *Chem. Soc. Rev.* **2018**, *47*, 9046–9068.

[29] R. E. Thompson, T. W. Muir, *Chem. Rev.* **2020**, *120*, 3051–3126.

[30] C. Bello, K. Farbiarz, J. F. Moller, C. F. W. Becker, T. Schwientek, *Chem. Sci.* **2014**, *5*, 1634–1641.

[31] C. Bello, C. F. W. Becker, *Bioorg. Med. Chem.* **2017**, *25*, 5016–5021.

[32] C. Marinzi, S. J. Bark, J. Offer, P. E. Dawson, *Bioorg. Med. Chem.* **2001**, *9*, 2323–2328.

[33] T. M. Hackeng, J. H. Griffin, P. E. Dawson, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 10068–10073.

[34] R. J. Hondal, B. L. Nilsson, R. T. Raines, *J. Am. Chem. Soc.* **2001**, *123*, 5140–5141.

[35] R. Quaderer, A. Sewing, D. Hilvert, *Helv. Chim. Acta* **2001**, *84*, 1197–1206.

[36] M. D. Gieselman, L. Xie, W. A. van Der Donk, *Org. Lett.* **2001**, *3*, 1331–1334.

[37] L. R. Malins, N. J. Mitchell, S. McGowan, R. J. Payne, *Angew. Chem. Int. Ed.* **2015**, *54*, 12716–12721; *Angew. Chem.* **2015**, *127*, 12907–12912.

[38] S. Dery, P. S. Reddy, L. Dery, R. Mousa, R. N. Dardashti, N. Metanis, *Chem. Sci.* **2015**, *6*, 6207–6212.

[39] A. L. Schroll, R. J. Hondal, S. Flemer Jr, *J. Pept. Sci.* **2012**, *18*, 155–162.

[40] N. Metanis, E. Keinan, P. E. Dawson, *Angew. Chem. Int. Ed.* **2010**, *49*, 7049–7053; *Angew. Chem.* **2010**, *122*, 7203–7207.

[41] D. Besse, L. Moroder, *J. Pept. Sci.* **1997**, *3*, 442–453.

[42] J. Liu, Q. Chen, S. Rozovsky, *J. Am. Chem. Soc.* **2017**, *139*, 3430–3437.

[43] J. A. Karas, A. Noor, C. Schieber, T. U. Connell, F. Separovic, P. S. Donnelly, *Chem. Commun.* **2017**, *53*, 6903–6905.

[44] J.-S. Zheng, Y. He, C. Zuo, X.-Y. Cai, S. Tang, Z. A. Wang, L.-H. Zhang, C.-L. Tian, L. Liu, *J. Am. Chem. Soc.* **2016**, *138*, 3553–3561.

[45] M. T. Jacobsen, M. E. Petersen, X. Ye, M. Galibert, G. H. Lorimer, V. Aucagne, M. S. Kay, *J. Am. Chem. Soc.* **2016**, *138*, 11775–11782.

[46] T. Yoshiya, S. Tsuda, S. Masuda, *ChemBioChem* **2019**, *20*, 1906–1913.

[47] K. Sato, S. Tanaka, J. Wang, K. Ishikawa, S. Tsuda, T. Narumi, T. Yoshiya, N. Mase, *Org. Lett.* **2021**, *23*, 1653–1658.

[48] A. Schmid, C. F. Becker in *Total Chemical Synthesis of Proteins* (Eds.: A. Brik, P. Dawson, L. Liu), Wiley-VCH, Weinheim, **2021**, pp. 437–462.

[49] S. Bondalapati, E. Eid, S. M. Mali, C. Wolberger, A. Brik, *Chem. Sci.* **2017**, *8*, 4027–4034.

[50] G. Holzwarth, P. Doty, *J. Am. Chem. Soc.* **1965**, *87*, 218–228.

[51] S. S. Kulkarni, E. E. Watson, J. W. C. Maxwell, G. Niederacher, J. Johansen-Leete, S. Huhmann, S. Mukherjee, A. R. Norman, J. Kriegesmann, C. F. W. Becker, R. J. Payne, *Angew. Chem. Int. Ed.* **2022**, *61*, e202200163; *Angew. Chem.* **2022**, *134*, e202200163.

[52] N. Shirafuji, S. Asano, S. Matsuda, K. Watari, F. Takaku, S. Nagata, *Exp. Hematol.* **1989**, *17*, 116–119.