CrossMark

ORIGINAL PAPER

Construction, Purification, and Characterization of a Homodimeric Granulocyte Colony-Stimulating Factor

Gitana Mickiene[1,2] · Indre Dalgediene[1] · Zilvinas Dapkunas[1,2] · Gintautas Zvirblis[1,2] ·
Henrikas Pesliakas[2] · Algirdas Kaupinis[3] · Mindaugas Valius[3] · Edita Mistiniene[2] ·
Milda Pleckaityte[1]

Abstract Granulocyte colony-stimulating factor (G-CSF) has found widespread clinical application, and modified forms with improved biopharmaceutical properties have been marketed as well. PEGylation, the covalent modification of G-CSF with polyethylene glycol (PEG), has a beneficial effect on drug properties, but there are concerns connected to the immunogenicity of PEGylated compounds and bioaccumulation of the synthetic polymer. To overcome challenges connected with chemical modifications, we developed fusion proteins composed of two G-CSF molecules connected via different peptide linkers. Three different homodimeric G-CSF proteins were purified, and their in vitro and in vivo activities were determined. A G-CSF dimer, GCSF-Lα, was constructed using an alpha-helix-forming peptide linker, and it demonstrated an extended half-life in serum with a stronger neutrophil response as compared to the monomeric G-CSF protein. The GCSF-Lα protein, therefore, might be selected for further studies as a potential drug candidate.

Keywords Biopharmaceuticals · Fusion proteins · G-CSF · Linker · Protein purification · Half-life · Biological activity · Absolute neutrophil count

✉ Milda Pleckaityte
milda.pleckaityte@bti.vu.lt

1   Institute of Biotechnology, Vilnius University, Sauletekio al. 7, 10257 Vilnius, Lithuania

2   Profarma UAB, V.A. Graiciuno 6, 02241 Vilnius, Lithuania

3   Institute of Biochemistry, Vilnius University, Sauletekio al. 7, 10257 Vilnius, Lithuania

Introduction

Biopharmaceuticals have become key players in the drug market for the treatment of a variety of diseases. Diabetes, cancer, inflammatory diseases, and hemophilia have already been targeted by biological drugs [1]. Biopharmaceuticals, which are mainly based on polypeptides and proteins, face several drawbacks, such as physicochemical instability, susceptibility to proteolytic degradation, and short circulation times. Many technologies have been developed to enhance the biopharmaceutical properties of the administered drugs [2–5]. Amino acid replacement targets mainly proteolytic instability and immunogenicity. Half-life extension strategies include chemical modification by the attachment of polyethylene glycol (PEGylation) and carbohydrate polymers (HESylation, polysialylation), glycosylation, and chemical lipidation [3, 5]. Fusion of proteins to human serum albumin (e.g., albiglutide) results in prolonged circulation time due to a large size that prevents renal clearance [6]. A decreased susceptibility to intracellular degradation can be achieved by fusion with the Fc portion of IgG [7]. Novel fusion partners for therapeutic proteins include human transferrin and highly sialylated carboxy-terminal peptide CTP. G-CSF fusion with human transferrin was used to develop an orally administered myelopoietic agent [8]. CTP fusion to erythropoietin [9] and follicle-stimulating hormone [10] produced recombinant proteins with increased half-lives. Overall, the development of multimeric proteins has been gaining speed as they have proven to be more effective than a single protein, with easier and faster refolding, increased biological activity, and/or prolonged circulation time [2]. Linkers are an indispensable tool in recombinant fusion protein technology because their type, length, and flexibility can have a crucial effect on the physical and biological characteristics of fusion proteins.

Granulocyte colony-stimulating factor (G-CSF) has gained clinical use due to its ability to selectively stimulate proliferation and differentiation of progenitor cells and to activate maturation of neutrophils [11]. This protein is marketed as a recombinant glycosylated protein (lenograstim), *E. coli*-derived G-CSF (filgrastim), and also as a modified version, PEGylated G-CSF (pegfilgrastim). Glycosylation of the G-CSF molecule does not affect its circulation half-life, although it is more stable and active than the recombinant G-CSF protein marketed as filgrastim [12]. PEGylation of G-CSF results in a substantially reduced renal clearance that allows for less frequent dosing during administration [13]. The safety of PEGyated compounds is still under consideration. There is some concern that PEGylated proteins have a greater potential for accumulation in tissues and cells, where lysosomal enzymes efficiently process the protein but not the PEG moiety [14]. PEGylation reduces the immunogenicity of the proteins [15, 16]; however, the data on antibody formation against PEG-conjugated compounds have been reported in recent studies [17, 18]. The Fc fusion format was applied to generate a not-yet marketed G-CSF dimer, which exhibited biological activity in vivo [19].

To improve the therapeutic properties of G-CSF and to overcome problems connected with chemical modification, we developed fusion proteins composed of two G-CSF molecules connected via different peptide linkers. Three homodimeric G-CSF proteins were purified, characterized, and compared with monomeric G-CSF. GCSF-Lα protein, constructed using an alpha-helix-forming peptide linker, exhibited an extended circulation half-life and increased biological activity in vivo.

Materials and Methods

Generation of Three Variants of Dimeric G-CSF and Their Expression in *E. coli*

The DNA fragment coding for the mature isoform B with extra N-terminal Met of G-CSF protein (Genbank accession no. NM_172219) was amplified using a synthetic cDNA gene of human G-CSF by polymerase chain reaction (PCR) to introduce NdeI at the 5′-end and Kpn2I and BamHI sites at the 3′-end of the PCR fragment using 5′-CATATGACACCTTTAGGACCTGCT and 5′-GGATCC GCATCCGGACGGCTGCGCAAGGTGGCGTAG primers (all primers were purchased from Metabion, Germany). The obtained PCR product was verified by sequencing, then digested with Kpn2I and BamHI restriction enzymes (all enzymes were purchased from Thermo Fisher Scientific, Vilnius, Lithuania) and fused with the Kpn2I/BamHI digested DNA fragment containing the Lα

linker coding sequence (SGLEA(EAAAK)$_4$ALEA (EAAAK)$_4$ALEGS) [20]. The resulting construct included the first copy of the *gcsf* gene fused with the linker coding sequence. The second copy of the *gcsf* gene was amplified to introduce BamHI and HindIII sites at the 5′- and 3′-ends of the PCR fragment using primers 5′-GGATCCACACC TTTAGGACCT and 5′-AAGCTTATTACGGCTGCGCA AGGTGGCG, respectively. The construct obtained by fusion of the two copies of the G-CSF coding gene interspaced by the Lα linker was named GCSF-Lα. To introduce the L2 linker (coding for (Ser-Gly$_4$)$_2$-Ser) and the L7 linker (coding for (Ser-Gly$_4$)$_7$-Ser), the DNA coding for GCSF-Lα was further digested with Kpn2I and BamHI and fused with the DNA fragments coding for the respective linker sequences. The constructs obtained by fusion of the two *gcsf* gene copies interspaced by the L2 or L7 linker sequences were named GCSF-L2 and GCSF-L7, respectively.

The DNA constructs for GCSF-Lα, GCSF-L2, and GCSF-L7 were digested with NdeI and HindIII restriction endonucleases and cloned into expression vector pET21b(+) (Merck Millipore, Darmstadt, Germany). The resulting plasmids were transformed into *E. coli* BL21(DE3) (Merck Millipore). Protein synthesis was induced with 1 mM isopropyl-b-D-thiogalactopyranoside (Thermo Fisher Scientific). After induction, the cell pellet was disrupted by sonication and centrifuged. The supernatant (soluble fraction) and the cell pellet (insoluble fraction) were then analyzed by 12.5% polyacrylamide gel electrophoresis (SDS-PAGE) under reducing conditions.

Purification of the G-CSF Dimers

Isolation of inclusion bodies (IBs) from collected *E. coli* cells was performed as described previously [21]. Washed pellets were solubilized in urea containing buffer (8 M urea, 1 mM EDTA, 50 mM Tris–HCl, pH 8.0). To reduce the disulfide bonds within the proteins, 1,4-dithiothreitol (DTT) was added to a final concentration of 0.5–5 mM. After centrifugation at 25,000 rpm for 25 min at 4 °C, the supernatants containing the solubilized IBs were collected. Refolding of recombinant proteins was initiated by rapid dilution of the denatured–reduced proteins into buffer (50 mM Tris–HCl, 1 mM EDTA, pH 8.0) that contained oxidized glutathione (GSSG) until concentrations of 3 M urea and 0.5 mg/mL protein were reached. The final molar ratio of GSSG to DTT was 1–5 for GCSF-L2 and 1–6 for GCSF-L7 and GCSF-Lα. Refolding of dimeric G-CSFs was carried out for 24 h at 4 °C with gentle stirring. Recombinant proteins were loaded onto a DEAE Sepharose Fast Flow column (GE Healthcare, Uppsala, Sweden) that was equilibrated with 10 mM Tris–HCl (pH 7.5) and contained 20 mM NaCl. Two-step elution was performed

at NaCl concentrations of 0.08 and 0.45 M. Fractions containing G-CSF dimers were collected and diluted with 20 mM sodium acetate to reach a pH of 5.4 and a conductivity of 3.2 mS/cm. The solutions containing GCSF-Lα and GCSF-L2 proteins were applied to SP Sepharose Fast Flow columns (GE Healthcare, Uppsala, Sweden), whereas GCSF-L7 was applied onto a CM Sepharose Fast Flow column (GE Healthcare). The target proteins were eluted with a NaCl gradient using 20 mM sodium acetate (pH 5.4) that contained 450 mM NaCl. Recombinant proteins were subsequently loaded onto a Sephadex G-25 Medium column that was equilibrated with 10 mM acetic acid/NaOH buffer, pH 4.0. Fractions containing the dimeric proteins were pooled and kept in a storage buffer (10 mM acetic acid/NaOH buffer, pH 4.0, containing 5% D-sorbitol, 0.0025% Tween 80) at +4 °C.

RP-HPLC Analysis

The dimeric G-CSF proteins that were extracted from IBs, their folding intermediates, and fractions derived from ion-exchange chromatography were analyzed by reverse-phase high-performance liquid chromatography (RP-HPLC) using a C_{18} reverse-phase column (Hi-Pore RP-318, 4.6 × 250 mm, Bio-Rad, Richmond, CA, USA). The column was equilibrated with mobile phase A (0.1% trifluoroacetic acid (TFA) in water), and separation of the proteins was performed via gradient and isocratic elution with mobile phase B (10% water, 89.9% acetonitrile, and 0.1% TFA) at a flow rate of 1 mL/min, as follows: (1) initial equilibration with 100% A, (2) 5-min gradient to 55% B, (3) 50-min gradient to 80% B, (4) 1-min gradient to 90% B, (5) 4-min isocratic elution at 90% B, (6) 2-min gradient to 0% B, and (7) 12-min isocratic elution at 0% B.

The purified dimeric proteins collected after Sephadex G-25 Medium were also analyzed with RP-HPLC using a C_{18} reverse-phase column (Zorbax SB-C18, 4.6 × 250 mm, Agilent Technologies, USA). Mobile phases A (0.1% TFA in water) and B (0.1% TFA in acetonitrile) were used for gradient and isocratic elution at a flow rate of 0.9 mL/min, as follows: (1) initial equilibration at 100% A, (2) 2-min gradient to 50% B, (3) 3-min gradient to 53% B, (4) 38-min gradient to 67% B, (5) 2-min gradient to 81% B, (6) 3-min isocratic elution at 81% B, (7) 2-min gradient to 0% B, and (8) 5-min isocratic elution at 0% B. For both elution conditions, the temperature of the columns was maintained at 30 °C and eluted proteins were detected at 215 nm.

Size-Exclusion HPLC

Twelve μg of purified protein was injected into a TSK-gel G3000 SWXL column (7.8 × 300 mm, Tosoh Bioscience, Tokyo, Japan) connected to an Alliance e2695 HPLC system (Waters, USA). Samples were eluted with an isocratic mobile phase of 0.1 M Na_2HPO_4, 0.1 M Na_2SO_4 (pH 7.2), and a flow rate 0.6 mL/min (22 °C). Absorbance was recorded at 215 nm. Low molecular weight standards were obtained from GE Healthcare.

Western Blotting

After sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), purified recombinant proteins and the G-CSF monomer (Filgrastim, Sicor Biotech, Teva) were electro-transferred onto polyvinyl difluoride membranes (Millipore, Bedford, MA, USA). The membranes were blocked with 2% milk powder in phosphate-buffered saline (PBS) for 2 h. The blocking solution was removed, and the blots were incubated for 1 h with monoclonal antibody (mAb) (clone no. 5D7) against human G-CSF (Abcam, Cambridge, UK) at 1:2000 dilution in PBS with 1% Tween 20 (PBST). The membranes were washed with PBST and incubated for 1 h with goat anti-mouse IgG conjugated to horseradish peroxidase diluted 1:4000 in PBST. The enzymatic reaction was developed using tetramethylbenzidine chromogenic substrate (Sigma-Aldrich, St. Louis, MO, USA).

RP-HPLC Peptide Mapping

Peptide mapping was done according to a slightly modified method described previously [22]. Purified GCSF-Lα and G-CSF monomer were dissolved in 50 mM Na_2HPO_4 (pH 8.0) buffer at a concentration of 25 μg/mL. Endoproteinase Glu-C (Sigma-Aldrich) was added to the protein solution at a concentration of 2.5 μg/mL, and digestion was performed for 18 h at 37 °C. Separation of peptides was carried out using a Hi-Pore RP-318 column (Bio-Rad, Hercules, CA, USA) with an acetonitrile gradient.

Fluorescence Spectroscopy

The fluorescence emission spectra of GCSF-Lα and G-CSF were recorded at 290–420 nm with an excitation at 280 nm on a FluoroMax-4 spectrofluorometer (Horiba Jobin Yvon Inc., USA) using a 1-cm path-length cuvette. The protein samples were diluted with 10 mM sodium citrate (pH 3.2 or 7.5) containing 100 mM NaCl to a concentration of 0.14–0.16 mg/mL. Fluorescence signals were recorded after mixing and standing the solutions for 5 min at room temperature.

Protease Digestion, Amino Acid Composition, and Sequence Analysis of Peptides

Purified G-CSF and GCSF-Lα proteins were dissolved in 25 mM ammonium bicarbonate buffer to a concentration

of 0.01–4 µg. TCEP (tris(2-carboxyethyl)phosphine) or DTT was added to the protein solution at a concentration of 5 mM and the reaction mixture was incubated for 5 min at 100 °C and cooled to the room temperature. The reduced protein was reacted with 10 mM iodoacetamide, and the S-alkylation was allowed to proceed for 20 min at room temperature in the dark. An aliquot of trypsin (Sigma-Aldrich) was added to the protein solution at a concentration of 3 ng/mL, and digestion was performed for 3 h at 37 °C. Liquid chromatography-mass spectrometry (LC-MS) of the protein digests was performed with a nanoAcquity UPLC system (Waters Corporation, Manchester, UK) as described earlier [23]. Sequence analysis of peptides was carried out according to a procedure reported previously [23].

In Vitro Bioassay

In vitro biological activity of the G-CSF dimers was determined with a cell proliferation assay with the G-CSF-dependent cell line M-NFS-60 [24, 25]. The G-CSF monomer was used as a reference standard. Before the assay, M-NFS-60 cells were centrifuged and resuspended at a concentration of 5.0×10^7 cells/mL in the test medium (RPMI 1640 supplemented with 10% fetal bovine serum, antibiotic gentamicin sulfate, and 0.05 mM 2-mercaptoethanol). Fifty µL of the test medium was aliquoted into each well of a 96-well tissue culture plate. The purified dimeric proteins and G-CSF were serially diluted in the test medium. A volume of 50 µL of the diluted protein was added to the wells to give concentrations of 0.004–7.8 pg/mL. Each protein was tested in triplicate. Fifty µL of the cell suspension was added to each well. The plates were incubated at 37 °C in a 5% CO_2 atmosphere. After 48 h of incubation, 20 µL of tetrazolium salt solution (MTS, 5 g/L) (Promega, Madison, WI, USA) was added to each well, and incubation was continued for 4 h under the same conditions. The absorbance of formazan derived from MTS cleavage by cellular mitochondrial dehydrogenases [26] was measured using a multi-well scanning spectrofluorometer (FluoroMax-4, Horiba Scientific, USA) at 490 nm. The biological activity of each protein was calculated from proliferation curves using OriginLab Origin and Microsoft Excel.

The proliferation curves were constructed by plotting the $\log_2$ dilution of the dimeric protein or standard G-CSF against the absorbance value at 490 nm. To calculate specific biological activity of the recombinant proteins in vitro, the following equations were used:

$$A_T = \frac{2^{D_T} \cdot S_T}{2^{D_S} \cdot S_S} \cdot A_S$$

$$SA = \frac{A_T}{c_T}$$

where A_T and A_S are the activity (IU/mL) of the tested dimeric protein (GCSF-Lα, GCSF-L2, or GCSF-L7) and the standard G-CSF monomer, respectively; S_T and S_S are the initial dilutions of the dimeric protein and G-CSF, respectively; D_T is the $\log_2$ dilution of the dimeric protein at 50% of the maximum absorbance; D_S is the $\log_2$ dilution of the G-CSF monomer at 50% of the maximum absorbance (these values are obtained from the proliferation curves); c_T is the concentration (mg/mL) of the dimeric protein; and SA is the specific activity (IU/mg) of the dimeric protein.

Bioavailability and Biological Activity In Vivo

Assays for bioavailability of the purified GCSF-Lα protein and the G-CSF monomer were performed using healthy Wistar rats (8–12 weeks old). The study with laboratory animals was approved by the State Food and Veterinary Service of the Republic of Lithuania (approval no. 0182). Groups of six rats received a single subcutaneous injection of G-CSF and a control group received sodium acetate buffer. Protein was injected in a volume of 1 mL/kg at a dose of 150 µg of protein/kg. Blood samples were drawn from the rats at selected time points (0, 3, 6, 12, 18, and 24 h after injection). Protein concentrations in the blood serum were determined by a sandwich enzyme-linked immunosorbent assay (ELISA) using G-CSF ELISA Development Kit according to manufacturer's instructions (PeproTech, Rocky Hill, NJ, USA). Concentration values of the G-CSF in serum at selected time intervals after protein injection were fitted with an exponential function using GRAFIT software, which obtained a coefficient kappa for every protein function. The circulation half-life of the proteins was determined using an equation $t\frac{1}{2} = \ln(2)/\text{kappa}$.
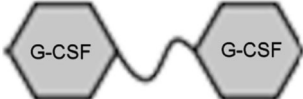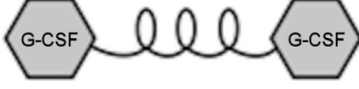
In vivo absolute neutrophil count (ANC) in rats was determined to assess the biological activity of the G-CSF proteins. Groups of six rats received a single subcutaneous injection of control buffer, GCSF-Lα, or monomeric G-CSF. Protein injections were administered with a dose of 200 µg/kg and a volume of 1 mL/kg. Blood samples were drawn from the rats at selected time points (0, 24, 48, and 72 h after injection). The ANC was determined using a microcell counter (hematology analyzer Exigo EOS, Boule Medical AB, Spånga, Sweden).

Results

Preparation of the G-CSF Dimers

Dimeric G-CSFs connected by a covalent fusion of two G-CSF molecules with various linker peptides (Table 1,

Table 1 Linkers connecting two G-CSF molecules

Name	Amino acid sequence	Schematic representation
L2	(SG_4)–(SG_4)–S	
L7	(SG_4)–(SG_4)–(SG_4)–(SG_4)–(SG_4)–(SG_4)–(SG_4)–S	
Lα	SGLEA–$(EAAAK)_4$–ALEA–$(EAAAK)_4$–ALEGS	

[27]) were constructed. Three homodimeric G-CSF proteins GCSF-L2, GCSF-L7, and GCSF-Lα were expressed in *E. coli* cells. SDS-PAGE analysis of soluble and insoluble fractions of *E. coli* lysate revealed that recombinant dimeric G-CSFs were expressed as insoluble proteins that accumulated in IBs (data not shown). The main focus of this work was on the selection of appropriate refolding conditions for recombinant proteins. RP-HPLC was the main analytical tool to monitor the transition of the reduced protein form into the oxidized state (peak 1, Supplementary Fig. 1B) and to determine the purity level of the proteins throughout purification (Table 2). RP-HPLC analysis showed that the highest purity was reached for GCSF-Lα (61%) after refolding (Table 2). The refolding purities of GCSF-L2 and GCSF-L7 were significantly lower (43.2 and 51.2%, respectively) because of the accumulation of protein forms with longer retention times (peak 2, Supplementary Fig. 1B). The GCSF-Lα protein (37 kDa) was detected as a single band on the SDS-PAGE gel, while some heterogeneity was visible in the GCSF-L2 and GCSF-L7 preparations under non-reducing conditions. The proteins migrated as a single band after reduction, indicating the presence of heterogenic disulfide-bonded species in the non-reduced preparations (Supplementary Fig. 1C).

After refolding, GCSF-L2, GCSF-L7, and GCSF-Lα were subjected to further purification using a combination of ion-exchange chromatography columns and were then applied to a Sephadex G-25 Medium column for desalting. An efficient purification of GCSF-Lα was achieved with the first chromatography step (DEAE Sepharose FF column) using a linear salt gradient between 0.02 and 0.08 M NaCl (Table 2). With the same conditions, a considerable amount of the longer, RP-HPLC-retained protein forms were eluted along with the oxidized protein forms of GCSF-L2 and GCSF-L7. To reduce accompanying impurities, all G-CSF derivatives were further loaded onto a cation-exchange chromatography column. A strong cation-exchange medium, SP Sepharose Fast Flow, was used for GCSF-L2 and GCSF-Lα, while a weak cation-exchange matrix (CM Sepharose Fast Flow) was applied to the purification of GCSF-L7. Application of CM sepharose fast flow facilitated better separation of the refolded form of GCF-L7 from impurities (data not shown). Fractions of the purified proteins were applied to Sephadex G-25 Medium for desalting and buffer exchange.

Characterization of the Purified Dimeric G-CSF Proteins

The purified G-CSF dimers were characterized by a subset of analytical methods that included HPLC (RP and size exclusion), SDS-PAGE, Western blotting, MS, peptide

Table 2 Yield and purity of the G-CSFs dimers after three processing steps

Protein	Theoretical mol. weight (kDa)	Refolding step		I chromatography step (anion exchange)		II chromatography step (cation exchange)	
		Yield[a] (%)	Purity (%)	Yield[a] (%)	Purity (%)	Yield[a] (%)	Purity (%)
GCSF-L2	38.2	89.3 ± 2.2	43.2 ± 3.8	4.6 ± 1.6	55.4 ± 4.2	1.2 ± 0.5	79.1 ± 2.1
GCSF-L7	39.7	90.3 ± 2.7	51.2 ± 3.5	7.6 ± 2.1	68.1 ± 3.8	2.2 ± 1.1	82.4 ± 1.5
GCSF-Lα	42.5	91.0 ± 1.5	61.0 ± 2.7	17.0 ± 2.1	90.2 ± 3.2	14.1 ± 2.2	95.1 ± 1.0

Each value represents the mean of three independent analyses ± standard deviation (SD)

Purity of the protein was determined by RP-HPLC

[a] Protein concentrations were determined by the Bradford method [28] using bovine serum albumin as a standard
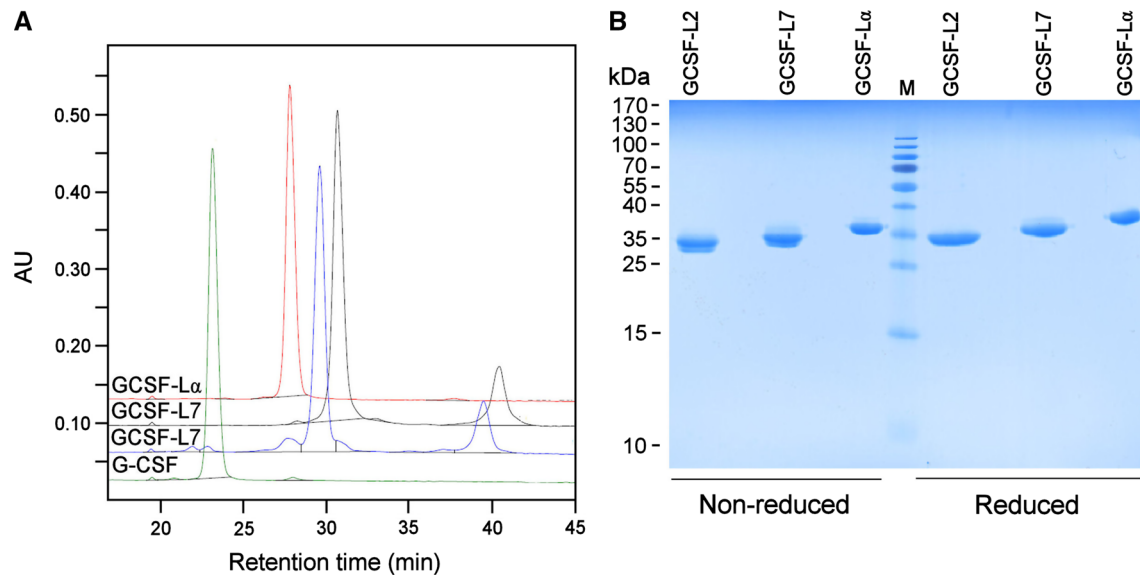
Fig. 1 a The RP-HPLC analysis of the purified dimeric G-CSF proteins collected after Sephadex G-25 Medium and the G-CSF monomer used as a control. Fifteen μg of each protein was loaded onto Zorbax 300SB-C18 column (Agilent Technologies) and separated via gradient and isocratic elution with mobile phase A (0.1% TFA in water) and mobile phase B (0.1% TFA in acetonitrile) at 30 °C. Absorbance at 215 nm is reported as AU. **b** SDS-PAGE of the purified dimeric proteins under non-reducing and reducing conditions. *M* molecular weight marker (Thermo Fisher Scientific)

mapping, and fluorescence spectroscopy. *E. coli*-derived G-CSF monomer was used as a reference.

The RP-HPLC analysis revealed different hydrophobicities and degrees of purity for the G-CSFs (Fig. 1a).

The shortest retention time was observed for G-CSF, whereas it increased for the dimeric proteins in the order of GCSF-Lα < GCSF-L7 < GCSF-L2. The purities of GCSF-Lα, GCSF-L7, and GCSF-L2 determined by RP-HPLC were 95, 82, and 79%, respectively (Table 2). The purified G-CSF dimers migrated on SDS-PAGE depending on their molecular weight; however, the corresponding bands were observed at slightly lower positions on the gel than their calculated molecular weights (38.2 kDa for GCSF-L2, 39.7 kDa for GCSF-L7, and 42.5 kDa for GCSF-Lα) (Fig. 1b). GCSF-Lα produced a single major band on the gel, while the bands corresponding to GCSF-L2 and GCSF-L7 had an extra, thinner band, which was present throughout the purification process. This extra band was not found under reducing conditions, suggesting that it might correspond to the later eluted protein form detected by RP-HPLC with a retention time between 37 and 42 min (Fig. 1a).

The mobility of the dimers and the G-CSF monomer on a calibrated SE-HPLC column was estimated (Fig. 2). The proteins were separated according to their relative molecular weights (55 kDa for GCSF-Lα, 49 kDa for GCSF-L7, and 42 kDa for GCSF-L2), which were higher than those calculated from the corresponding amino acid sequences (Table 2). Such discrepancies resulted from the fact that the SEC column is able to separate proteins by their shape
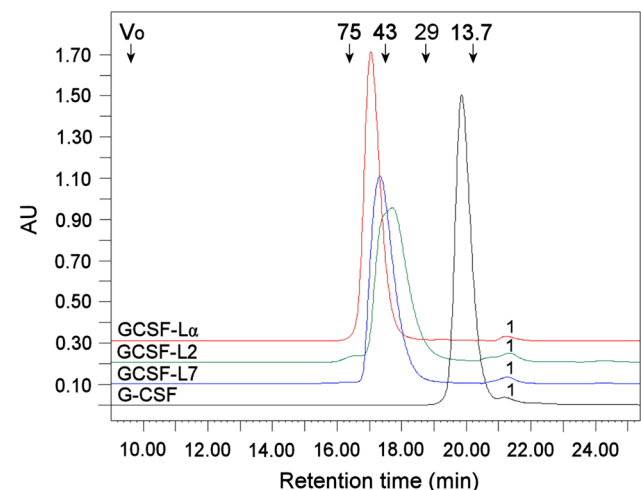

Fig. 2 Size-exclusion HPLC analysis of the purified dimeric G-CSF proteins collected after Sephadex G-25 medium and the G-CSF monomer used as a control. Twelve μg of each protein was injected into a TSK-gel G3000 SWXL (7.8/300) column and eluted at 0.6 mL/min in 0.1 M Na$_2$HPO$_4$ (pH 7.2) buffer, containing 0.1 M Na$_2$SO$_4$. *Arrows* indicate the retention time of molecular weight standards: conalbumin (75 kDa), ovalbumin (43 kDa), carbonic anhydrase (29 kDa), and ribonuclease A (13.7 kDa). Blue dextran was used to determine the void volume (Vo) of the column. Absorbance at 215 nm is reported as AU. Peak 1 obtained at retention time of 21.25 min represents excipients from the storage buffer

(hydrodynamic radius), as well as their molecular weight [29]. The G-CSF monomer, which has the smallest hydrodynamic radius, was eluted with a relative molecular weight of 17 kDa, which is lower than what was established for filgrastim (18.8 kDa), indicating that G-CSF is a
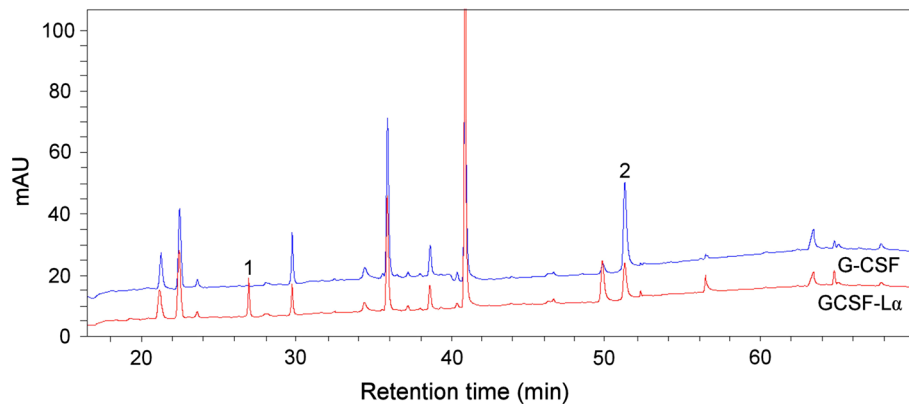
Fig. 3 The RP-HPLC peptide mapping of GCSF-Lα (*red line*) and the G-CSF monomer (*blue line*). Eight μg of each protein was used for analysis. The separation of peptides was performed on a Hi-Pore RP-318 column via gradient elution of two mobile phases: 94.5% water, 5% acetonitrile, and 0.05% TFA (A) and 5% water, 94.5% acetonitrile, and 0.05% TFA (B) at a constant flow rate of 1 mL/min at 50 °C, as follows: (1) initial equilibration at 3% B, (2) 12-min gradient to 6% B, (3) 34-min gradient to 34% B, (4) 29-min gradient to 90% B, (5) 8-min isocratic elution at 90% B, (6) 2-min gradient to 3% B, and (7) 12-min isocratic elution at 3% B. Absorbance was recorded at 215 nm (Color figure online)
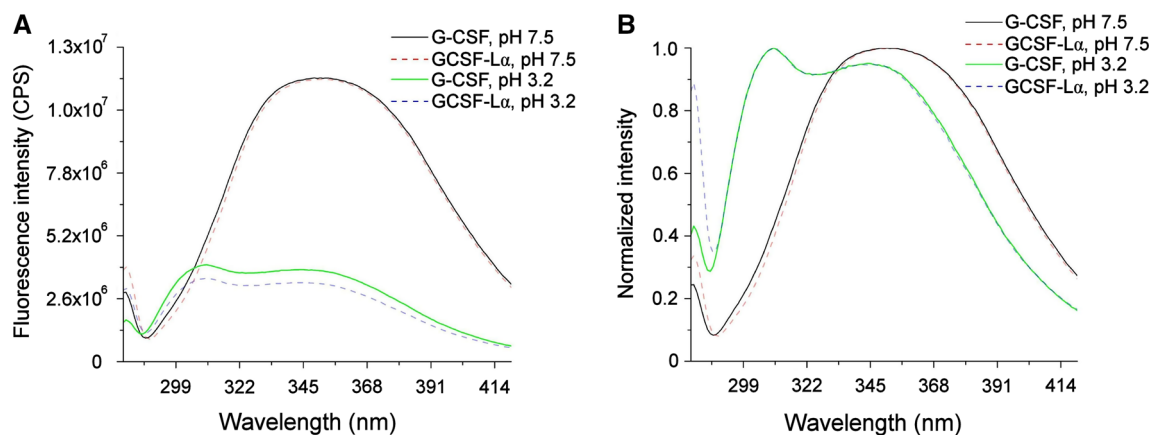

Fig. 4 Raw (**a**) and normalized (**b**) fluorescence emission spectra of the purified GCSF-Lα protein and G-CSF monomer at pH 3.2 and 7.5. The spectra were normalized at their maximum

compact molecule. The SE-HPLC profile of GCSF-L2 showed the presence of a small amount of impurities (approximately 6% of the higher molecular weight forms), whereas GCSF-Lα, GCSF-L7, and G-CSF protein preparations reached approximately 99% purity.

The molecular weights of GCSF-Lα, GCSF-L7, and GCSF-L2 (42,513.95, 39,742.90, and 38,166.67 Da, respectively) determined by electrospray ionization mass spectrometry (ESI-MS) were in good agreement with the ones calculated from their amino acid sequences (within a range of 2 Da of the theoretical values). The G-CSF dimers were detected by Western blot analysis using a mAb specific to the G-CSF monomer (Supplementary Fig. 2). The major bands corresponding to GCSF-L7 and GCSF-L2 on the Western blot had an extra lower molecular weight band that is suggested to be degradation products of the dimeric proteins.

The RP-HPLC peptide mapping of GCSF-Lα was used to confirm the identity of this protein sequence and its structural alterations. GCSF-L2 and GCSF-L7, due to insufficient purities (Table 2), were not subjected to this analysis. Peptide mapping of the G-CSF monomer and purified GCSF-Lα was performed by digestion with endoproteinase Glu-C, which cleaves G-CSF at the carboxyl side of glutamyl and/or aspartyl peptide bonds [30]. The obtained peptide mixture was analyzed by RP-HPLC (Fig. 3). The peptide maps of GCSF-Lα and G-CSF showed similar profiles; however, several discrepancies were found. The chromatogram of G-CSF showed peptide 2 at 51.8 min (Fig. 3), which might indicate a partially digested larger fragment from the peptide mixture. The additional peak (peptide 1, Fig. 3), which eluted from the GCSF-Lα peptide mixture at 26.5 min, may belong to the linker sequence joining the two G-CSF molecules into a whole polypeptide chain. To resolve the peptide mapping differences observed by RP-HPLC (Fig. 3), mass spectrometry (MS) of purified GCSF-Lα and peptide liquid

Fig. 5 In vitro proliferation assay profiles for the purified GCSF-L2 (**a**), GCSF-L7 (**b**), and GCSF-Lα (**c**). The G-CSF monomer (control) included in each assay. The curves were obtained at twofold doubling (log 2) serial dilutions of the tested proteins. *Error bars* represent standard deviation (SD) of the absorbance means (490 nm) obtained in triplicate
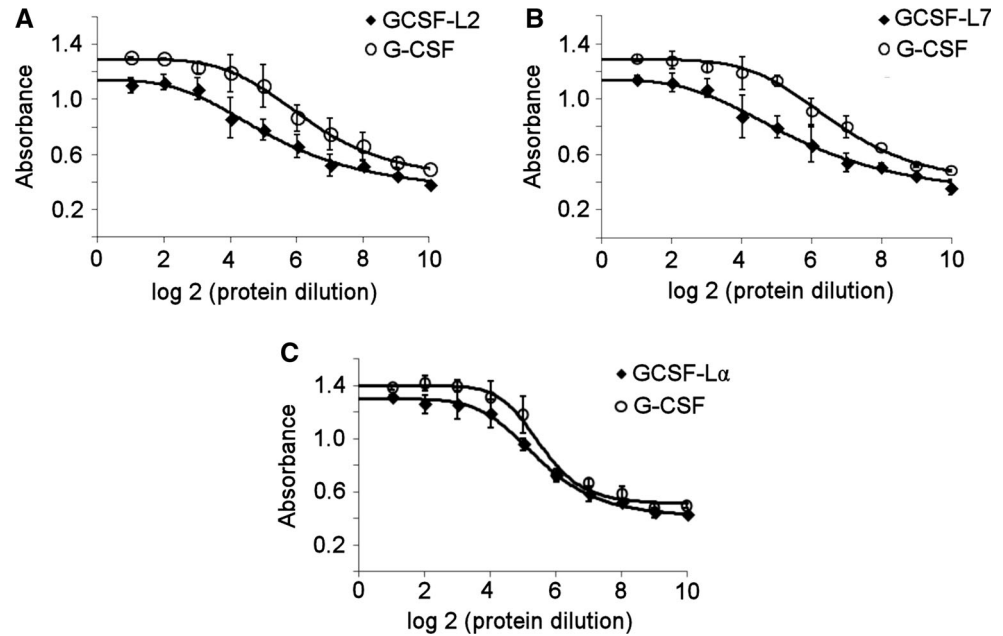
chromatography-mass spectrometry (LC-MS) were performed. The molecular weight of GCSF-Lα was confirmed by MS (Supplementary Fig. 3). The tryptic peptides identified by LC-MS matched with the linker sequence of GCSF-Lα (Table S1), resolving the discrepancies reflected in Fig. 3.

To verify that native disulfide bonds were formed, a sample of GCSF-Lα was subjected to fluorescence analysis. It was demonstrated that under acidic conditions (pH 3.2) correctly folded active G-CSF exhibits a characteristic Tyr and Trp fluorescence spectrum, which was not observed for any disulfide-reduced intermediates [31]. At neutral pH, the fluorescence emission spectra of GCSF-Lα and G-CSF were characterized by the tryptophan fluorescence with a maximum at 350 nm. At pH 3.2, the Trp fluorescence decreased and a shoulder with a maximum at 310 nm, attributable to Tyr fluorescence, was present (Fig. 4). Such changes are characteristic feature of the native G-CSF molecule, since the disulfide-reduced intermediates as well as the disulfide-unpaired analogs show a decrease in the intensity of the Trp fluorescence, but no change in the Tyr fluorescence [31]. The obtained data demonstrated (Fig. 4) the formation of native disulfide bonds in the dimeric GCSF-Lα protein.

Biological Activity In Vitro and In Vivo

In vitro activity of the G-CSF dimers was determined using a cell proliferation assay with the G-CSF dependent cell line M-NFS-60 [25]. The G-CSF monomer was used as a reference. Results were obtained by absorbance readings of the colored formazan product accessed by the cleavage of tetrazolium salt (MTS) by viable cells [26]. The proteins induced proliferation of the cells in a dose-dependent manner (Fig. 5). Calculated in vitro activity of GCSF-Lα reached 48% of that of the G-CSF monomer, while GCSF-L2 and GCSF-L7 demonstrated relative activities of 22%.

A bioavailability comparison between the G-CSF monomer and GCSF-Lα was made using groups of six rats. GCSF-L2 and GCSF-L7 were not included in the assay because of unsatisfactory yield and purity (Table 2). Each rat in the group received a single subcutaneous injection of G-CSF or GCSF-Lα at a dose of 150 μg/kg. Protein in the blood serum at selected time intervals after subcutaneous injection was assessed using a human G-CSF ELISA kit. A curve of the protein concentration in the blood serum samples versus time was generated (Fig. 6a). Calculated protein circulation half-life ($t_{1/2}$) for G-CSF and GCSF-Lα was 1.2 and 8.7 h, respectively. The data indicate that the clearance of GCSF-Lα was reduced by more than sevenfold compared to that of the G-CSF monomer.

The biological activity of G-CSF in vivo comprises its ability to stimulate neutrophil release from the bone marrow by inducing a transient increase in circulating neutrophils. An increase in ANC in the peripheral blood was detected using three groups of rats. After subcutaneous administration of GCSF-Lα, G-CSF, and control buffer, ANCs were determined at selected time intervals after injection (Fig. 6b). The ANC peaked at 24 h post-injection of G-CSF and GCSF-Lα and decreased at 48 h post-injection. In response to injection of GCSF-Lα, rats exhibited a 1.8-fold increase in circulating neutrophils 24 h post-injection.
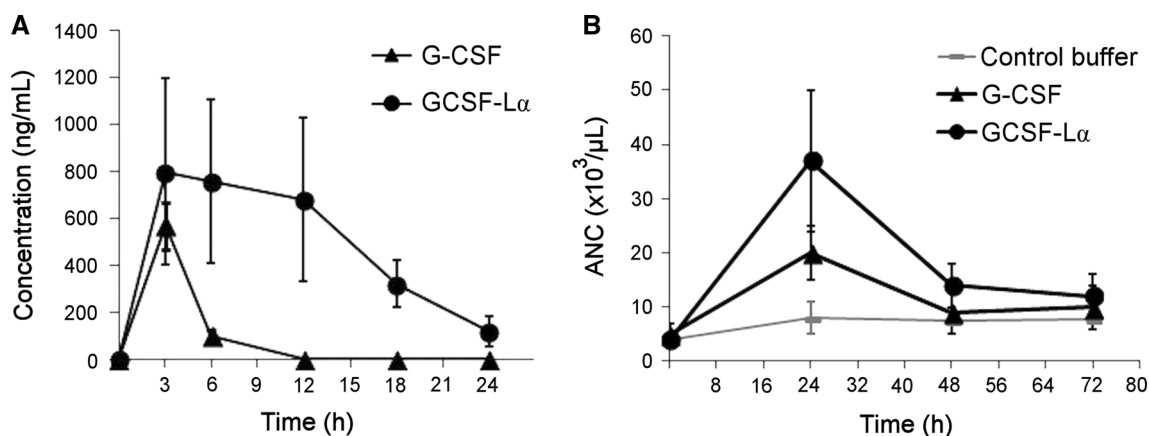
Fig. 6 a Pharmacokinetics profile of subcutaneously administered GCSF-Lα and the G-CSF monomer. *Error bars* represent standard deviation (SD) of G-CSF serum concentration obtained from four rats per group and GCSF-Lα serum concentration obtained from five rats per group. **b** The ANC count versus time profiles of subcutaneously

administered GCSF-Lα, G-CSF, and control buffer. *Error bars* represent standard deviation (SD) of the ANC obtained from blood of five rats per group after injection of G-CSF, blood of six rats per group was used after injection of GCSF-Lα, and three rats were in the control group

Discussion

In this work, we describe the construction, purification, and characterization of three novel human G-CSF dimers, which were developed by genetic fusion of two G-CSF molecules via three specific linkers (Table 1). The length and amino acid composition of a linker sequence had a profound effect on physical characteristics of G-CSF dimeric proteins.

Active, recombinant human G-CSF contains a free cysteine residue at position 17 and two intramolecular disulfide bonds, Cys36-Cys42 and Cys64-Cys74 [32]. To recover the biological activity of dimeric G-CSFs accumulated as insoluble proteins, the main focus was laid on the selection of an appropriate refolding procedure for them. The successful refolding of IBs solubilized in urea in the presence of sufficient amount of DTT was achieved by a subsequent slow dilution procedure, which gave purities of the refolded proteins of higher than 40%, as determined by RP-HPLC (Table 2). Two consecutive chromatographic steps were selected for the purification of GCSF-L2, GCSF-L7, and GCSF-Lα. The purity of GCSF-Lα determined by RP-HPLC was approximately 95%. Complete separation of the oxidized and longer-retained forms of the GCSF-L2 and GCSF-L7 proteins was practically infeasible by conventional ion-exchange chromatography methods. The final purity of GCSF-L2 and GCSF-L7 was less than 83%, as determined by RP-HPLC, and the final yield was 6–11 times lower than that of GCSF-Lα.

The longer RP-HPLC-retained form in the GCSF-L2 and GCSF-L7 preparations was detected as an extra band in SDS-PAGE, which was present throughout the purification process. This extra band disappeared in SDS-PAGE under reducing conditions, indicating that this

accompanying impurity originated from the formation of extra disulfide bonds within these proteins. We presume that this structure was formed via –S–S– bonds of unpaired cysteine residue that originated from the G-CSF monomer units. This assumption may explain why such protein forms were not completely separated during chromatography steps and confirms the importance of the linker composition and length. The L2 and L7 linkers increased spatial separation between the domains, allowing them to interact with one another [33, 34], while the Lα linker may rigidly ensure separation and elongation of the two G-CSF monomers to a distance favorable for their independent functioning. GCSF-Lα has a significant advantage over the other two dimers, as it is easier to purify and to obtain the large quantities of high purity protein.

The interaction between the G-CSF molecule and its G-CSFR receptor resulted in a stoichiometric ratio of 2:2 [35]. Using spectroscopic total internal reflection ellipsometry (TIRE), it was demonstrated that the G-CSF molecules included into GCSF-Lα interact with the receptor [36]. The obtained data presume that the length of the Lα linker allows spatial separation of the G-CSF molecules in a way that their binding sites become accessible for interaction with their receptors. GCSF-Lα therefore may activate the receptor with higher efficiency supporting the 1:2 stoichiometry between the dimeric G-CSF protein and its G-CSFR receptor. GCSF-L2 and possibly GCSF-L7 having a shorter linker sequence act as the G-CSF monomer (1:1 stoichiometry between GCSF-L2 and G-CSFR). The L2 and L7 linkers, which consisted of stretches of Gly and Ser residues, implemented greater degrees of freedom in the overall conformation, whereas the Lα linker had a more rigid, helical-alpha, spiral-like structure [33, 34, 37].

The ability to stimulate cell proliferation in vitro was reduced for all dimeric G-CSF molecules compared to the G-CSF monomer. Retention of about 50% of in vitro biological activity by GCSF-Lα might indicate that spatial interference has an impact on G-CSF interaction with the receptor as the N- and C-termini of G-CSFs in the dimeric molecule are joined with the linker sequence. In vitro biological activity of GCSF-Lα was comparable to that of PEGylated G-CSF (Pegfilgrastim), whereby PEG is attached to amine groups found at the N-terminus of G-CSF, which retained from 45 to 70% activity relative to the G-CSF monomer [38, 39]. The fusion of two G-CSF molecules into a homodimer obviously has an advantage over longer acting forms of G-CSF, as the generation of dimeric proteins was obtained by more simple technology than PEGylation. Low biological activity of GCSF-L2 and GCSF-L7 could be a consequence of the steric hindrance caused by improper linker length and the presence of accompanying impurities.

The biological activity assays in healthy rats demonstrated that the GCSF-Lα molecule has advantages over the G-CSF monomer. GCSF-Lα showed a circulation half-life of 8.7 h, more than sevenfold longer than that of G-CSF. In another study, the dimer (F_{dim}) generated by joining two G-CSF molecules via disulfide bridges between non-paired cysteines had the same circulation half-life as the G-CSF monomer [38]. We assume that F_{dim} is a more compact molecule than GCSF-Lα; therefore, the prolonged action of GCSF-Lα is primarily associated with the hydrodynamic radius of the molecule. The in vivo response, which comprised the ability of G-CSF to stimulate neutrophil release, was more pronounced for GCSF-Lα compared to the monomeric protein. After 24 h of a single subcutaneous injection of GCSF-Lα, rats exhibited a 1.8-fold increase in circulating neutrophils, albeit with a larger margin of error.

Conclusions

In this paper, we described the construction and characterization of three G-CSF dimeric proteins generated using different linker peptides. GCSF-Lα had the best performance in terms of purity and in vitro activity. However, the primary idea on the ability of GCSF-Lα to activate the receptor with the higher efficiency was not confirmed as the dimer acted as the G-CSF monomer in vitro. To explain these findings better, an understanding of the molecular mechanism of receptor activation by the G-CSF dimer is needed. Biological activity studies with healthy rats demonstrated that GCSF-Lα had a longer half-life in the blood serum and produced a stronger neutrophil response compared to monomeric G-CSF. The GCSF-Lα protein might be selected for further studies as a potential drug candidate. The Lα-type linker can be applied to the development of structurally similar receptor/ligand systems.

Compliance with Ethical Standards

Conflict of interest The authors declare that there are no conflicts of interest.

Ethics Approval All procedures performed in studies involving animals were in accordance with the ethical standards of the institution or practice at which the studies were conducted. The study with laboratory animals was approved by the State Food and Veterinary Service of the Republic of Lithuania (approval no. 0182).

References

1. Walsh, G. (2014). Biopharmaceutical benchmarks 2014. *Nature Biotechnology, 32,* 992–1000.
2. Strohl, W. R. (2014). Fusion proteins for half-life extension of biologics as a strategy to make biobetters. *BioDrugs, 29,* 215–239.
3. Pasut, G. (2014). Pegylation of biological molecules and potential benefits: pharmacological properties of certolizumab pegol. *BioDrugs, 28*(Suppl 1), S15–S23.
4. Metzner, H. J., Pipe, S. W., Weimer, T., & Schulte, S. (2013). Extending the pharmacokinetic half-life of coagulation factors by fusion to recombinant albumin. *Thrombosis and Haemostasis, 110,* 931–939.
5. Van Witteloostuijn, S. B., Pedersen, S. L., & Jensen, K. J. (2016). Half-life extension of biopharmaceuticals using chemical methods: Alternatives to PEGylation. *ChemMedChem, 11,* 2474–2495.
6. Rogers, B., Daoyuan, D., Li, Z., & Li, Z. (2015). Recombinant human serum albumin fusion proteins and novel applications in drug delivery and therapy. *Current Pharmaceutical Design, 21,* 1899–1907.
7. Czajkowsky, D. M., Hu, J., Shao, Z., & Pleass, R. J. (2012). Fc-fusion proteins: New developments and future perspectives. *EMBO Molecular Medicine, 4,* 1015–1028.
8. Bai, Y., Ann, D. K., & Shen, W. C. (2005). Recombinant granulocyte colony-stimulating factor-transferrin fusion protein as an oral myelopoietic agent. *Proceedings of the National Academy of Sciences, 102,* 7292–7296.
9. Fares, F., Ganem, S., Hajouj, T., & Agai, E. (2007). Development of a long-acting erythropoietin by fusing the carboxyl-terminal peptide of human chorionic gonadotropin beta-subunit to the coding sequence of human erythropoietin. *Endocrinology, 148,* 5081–5087.
10. Duijkers, I. J., Klipping, C., Boerrigter, P. J., Machielsen, C. S., De Bie, J. J., & Voortman, G. (2002). Single dose pharmacokinetics and effects on follicular growth and serum hormones of a long-acting recombinant FSH preparation (FSH-CTP) in healthy pituitary-suppressed females. *Human Reproduction, 17,* 1987–1993.
11. Caselli, D., Cesaro, S., & Aricò, M. (2016). Biosimilars in the management of neutropenia: Focus on filgrastim. *Biologics, 10,* 17–22.

12. Bönig, H., Silberman, S., Weller, S., Kirschke, R., Körholz, D., Janssen, G., et al. (2001). Glycosylated vs non-glycosylated granulocyte colony-stimulating factor (G-CSF)—Results of a prospective randomized monocenter study. *Bone Marrow Transplantation, 28,* 259–264.

13. Molineux, G. (2004). The design and development of pegfilgrastim (PEG-rmetHuG-CSF, Neulasta). *Current Pharmaceutical Design, 10,* 1235–1244.

14. Baumann, A., Tuerck, D., Prabhu, S., Dickmann, L., & Sims, J. (2014). Pharmacokinetics, metabolism and distribution of PEGs and PEGylated proteins: quo vadis? *Drug Discovery Today, 19,* 1623–1631.

15. Tsutsumi, Y., Onda, M., Nagata, S., Lee, B., Kreitman, R. J., & Pastan, I. (2000). Site-specific chemical modification with polyethylene glycol of recombinant immunotoxin anti-Tac(Fv)-PE38 (LMB-2) improves antitumor activity and reduces animal toxicity and immunogenicity. *Proceedings of the National Academy of Sciences, 97*(85), 48–8553.

16. Zeidan, A., Wang, E. S., & Wetzler, M. (2009). Pagasparaginase: Where do we stand? *Expert Opinion on Biological Therapy, 9,* 111–119.

17. Garay, R. P., El-Gewely, R., Armstrong, J. K., Garratty, G., & Richette, P. (2012). Antibodies against polyethylene glycol in healthy subjects and in patients treated with PEG-conjugated agents. *Expert Opinion on Drug Delivery, 9,* 1319–1323.

18. Hershfield, M. S., Ganson, N. J., Kelly, S. J., Scarlett, E. L., Jaggers, D. A., & Sundy, J. S. (2014). Induced and pre-existing anti-polyethylene glycol antibody in a trial of every 3-week dosing of pegloticase for refractory gout, including in organ transplant recipients. *Arthritis Research & Therapy, 16*(2), R63.

19. Yan, X., Huang, Z., Yang, H., Sun, N. C., & Huang S. (2016). Recombinant human G-CSF dimer and use thereof for the treatment of neurological diseases. U.S. Patent 9,273,108, March 1, 2016.

20. Arai, R., Ueda, H., Kitayama, A., Kamiya, N., & Nagamune, T. (2001). Design of the linkers which effectively separate domains of a bifunctional fusion protein. *Protein Engineering, 14,* 529–532.

21. Zaveckas, M., Luksa, V., Zvirblis, G., Chmieliauskaite, V., Bumelis, V., & Pesliakas, H. (2003). Mutation of surface-exposed histidine residues of recombinant human granulocyte-colony stimulating factor (Cys17Ser) impacts on interaction with chelated metal ions and refolding in aqueous two-phase systems. *Journal of Chromatography B, 786,* 17–32.

22. Council of Europe. (2009). European pharmacopoeia. In *Supplement 6.3. Filgrastim concentrated solution* (pp. 4142–4144). Strasbourg: Council of Europe.

23. Ger, M., Kaupinis, A., Nemeikaite-Ceniene, A., Sarlauskas, J., Cicenas, J., Cenas, N., et al. (2016). Quantitative proteomic analysis of anticancer drug RH1 resistance in liver carcinoma. *Biochimica et Biophysica Acta, 1864,* 219–232.

24. Weinstein, Y., Ihle, J. N., Lavu, S., & Reddy, E. P. (1986). Truncation of the c-myb gene by retroviral integration in an interleukin 3-dependent myeloid leukemia cell line. *Proceedings of the National Academy of Sciences, 83,* 5010–5014.

25. Matsuda, S., Shirafuji, N., & Asano, S. (1989). Human granulocyte colony-stimulating factor specifically binds to murine myeloblastics NFS-60 cells and activates their guanosine triphosphate binding proteins/adenylate cyclase system. *Blood, 74,* 343–348.

26. Monsmann, T. (1983). Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. *Journal of Immunology Methods, 65,* 55–63.

27. Mistiniene, E., Pesliakas, J. H., Pleckaityte, M., & Zvirblis, G. (2016). Derivatives of recombinant proteins, homo-multimers of granulocyte colony-stimulating factor and method of preparation thereof. U.S. Patent 9,243,049, January 26, 2016.

28. Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry, 72,* 248–254.

29. Hagel, L. (2003). Gel-filtration chromatography. In *Current protocols in molecular biology* (Vol. 44, pp. 10.9.1–10.9.2). New York: Wiley.

30. Drapeau, G. R. (1976). Protease from *Staphylococcus aureus*. In L. Lorand (Ed.), *Methods in enzymology* (Vol. 45B, pp. 469–475). New York: Academic Press.

31. Lu, H. S., Clogston, C. L., Narhi, L. O., Merewether, L. A., Pearl, W. R., & Boone, T. C. (1992). Folding and oxidation of recombinant human granulocyte colony stimulating factor produced in *Escherichia coli*. *Journal of Biological Chemistry, 267,* 8770–8777.

32. Lu, H. S., Boone, T. C., Souza, L. M., & Lai, P. H. (1989). Disulfide and secondary structures of recombinant human granulocyte colony stimulating factor. *Archives of Biochemistry and Biophysics, 268,* 81–92.

33. Chen, X., Zaro, J. L., & Shen, W. C. (2013). Fusion protein linkers: Property, design and functionality. *Advanced Drug Delivery Reviews, 65,* 1357–1369.

34. Reddy Chichili, V. P., Kumar, V., & Sivaraman, J. (2013). Linkers in the structural biology of protein–protein interactions. *Protein Science, 22,* 153–167.

35. Tamada, T., Honjo, E., Maeda, Y., Okamoto, T., Ishibashi, M., Tokunaga, M., et al. (2006). Homodimeric cross-over structure of the human granulocyte colony-stimulating factor (GCSF) receptor signaling complex. *Proceedings of the National Academy of Sciences, 103,* 3135–3140.

36. Balevicius, Z., Baleviciute, I., Tumenas, S., Tamosaitis, L., Stirke, A., Makaraviciute, A., et al. (2014). In situ study of ligand–receptor interaction by total internal reflection ellipsometry. *Thin Solid Films, 571,* 744–748.

37. Yamasaki, M., Konishi, N., Yamaguchi, K., Itoh, S., & Yokoo, Y. (1998). Purification and characterization of recombinant human granulocyte colony-stimulating factor (rhG-CSF) derivatives: KW-2228 and other derivatives. *Bioscience, Biotechnology, and Biochemistry, 62,* 1528–1534.

38. Fidler, K., Jevsevar, S., Milunovic, T., Skrajnar, S., Premzl, A., Kunstelj, M., et al. (2011). The characterization and potential use of G-CSF dimers and their pegylated conjugates. *Acta Chimica Slovenica, 58,* 1–8.

39. Molineux, G. (2009). Pegfilgrastim—Designing an improved form of rmetHu-GCSF. In F. M. Veronese (Ed.), *Pegylated protein drugs: Basic science and clinical applications* (pp. 169–185). Basel: Birkhauser.