Supporting Information

Backbone Circularization Coupled with Optimization of Connecting Segment in Effectively Improving the Stability of Granulocyte-Colony Stimulating Factor

Takamitsu Miyafusa†, Risa Shibuya‡, Wataru Nishima†, Rie Ohara‡, Chuya Yoshida†, and Shinya Honda†‡*

†Biomedical Research Institute, The National Institute of Advanced Industrial Science and Technology, Central 6, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566, Japan

‡Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8562, Japan

Table of Contents

Supplemental Figures

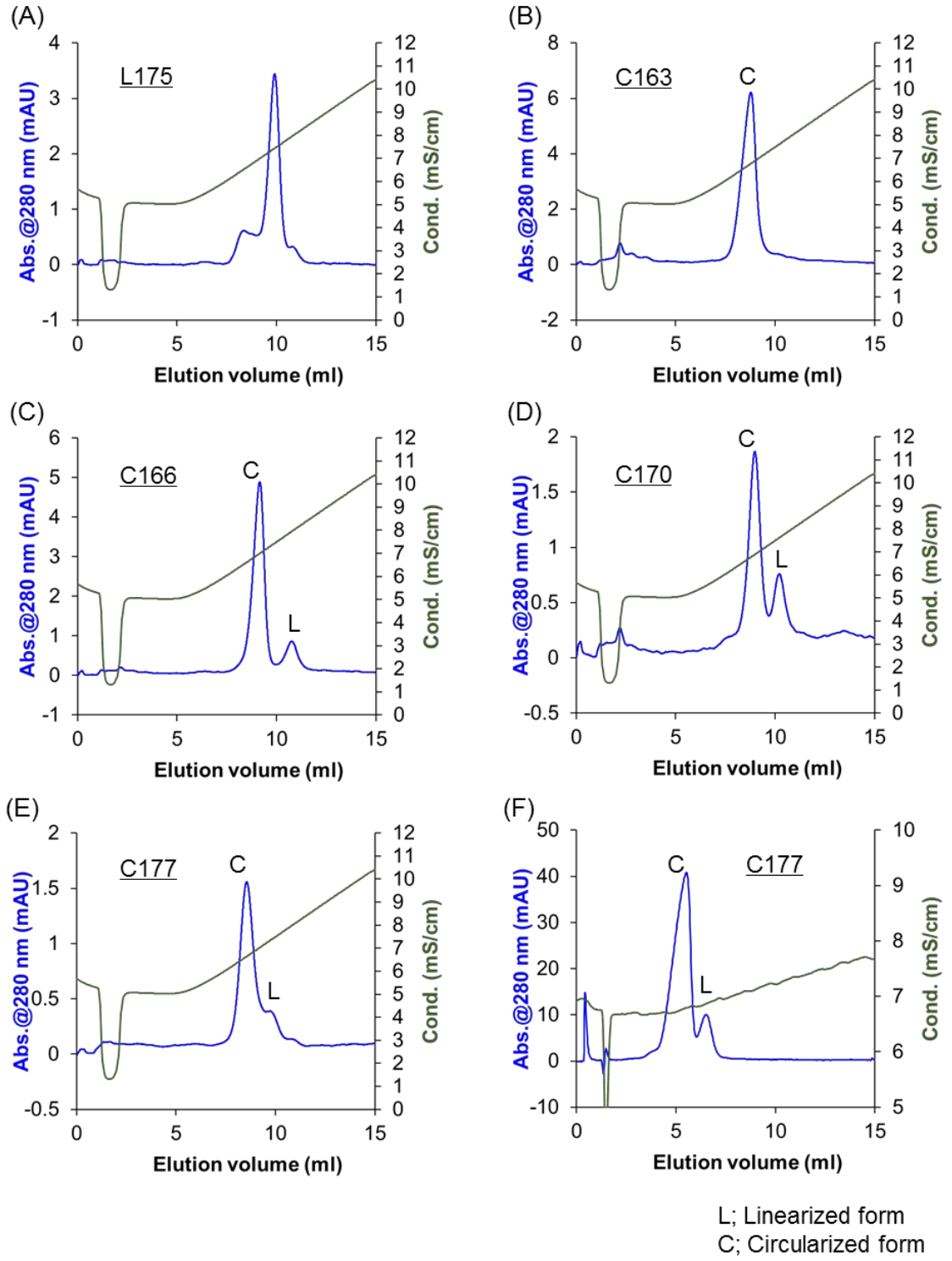

Figure S1. Chromatograms of the purification using MonoQ column. Related to Figure 2. (A)–(E) Main products and non-circularized by-products are separated by anion exchange chromatography. (F) The gradient condition was modified for the separation of C177. The peaks corresponding to the linearized and circularized forms are indicated by L and C, respectively.

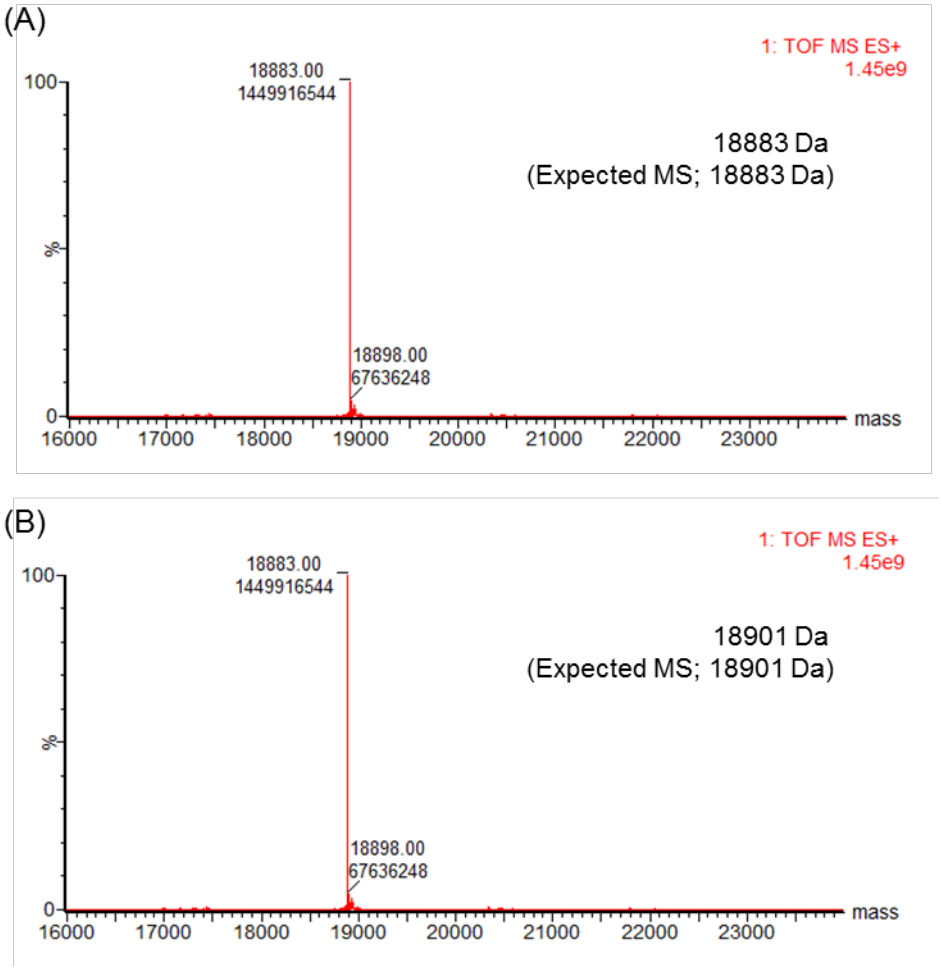

Figure S2. Deconvoluted mass spectrum of C177. Related to Figure 2. (A) Main product of C177. (B) By-product of C177. All spectra were obtained using the MaxEnt 1 software. Deconvolution was performed at a resolution of 1 Da. The average molecular weights of the circularized and hydrolyzed C177 are 18882.7 and 18900.7 Da, respectively.

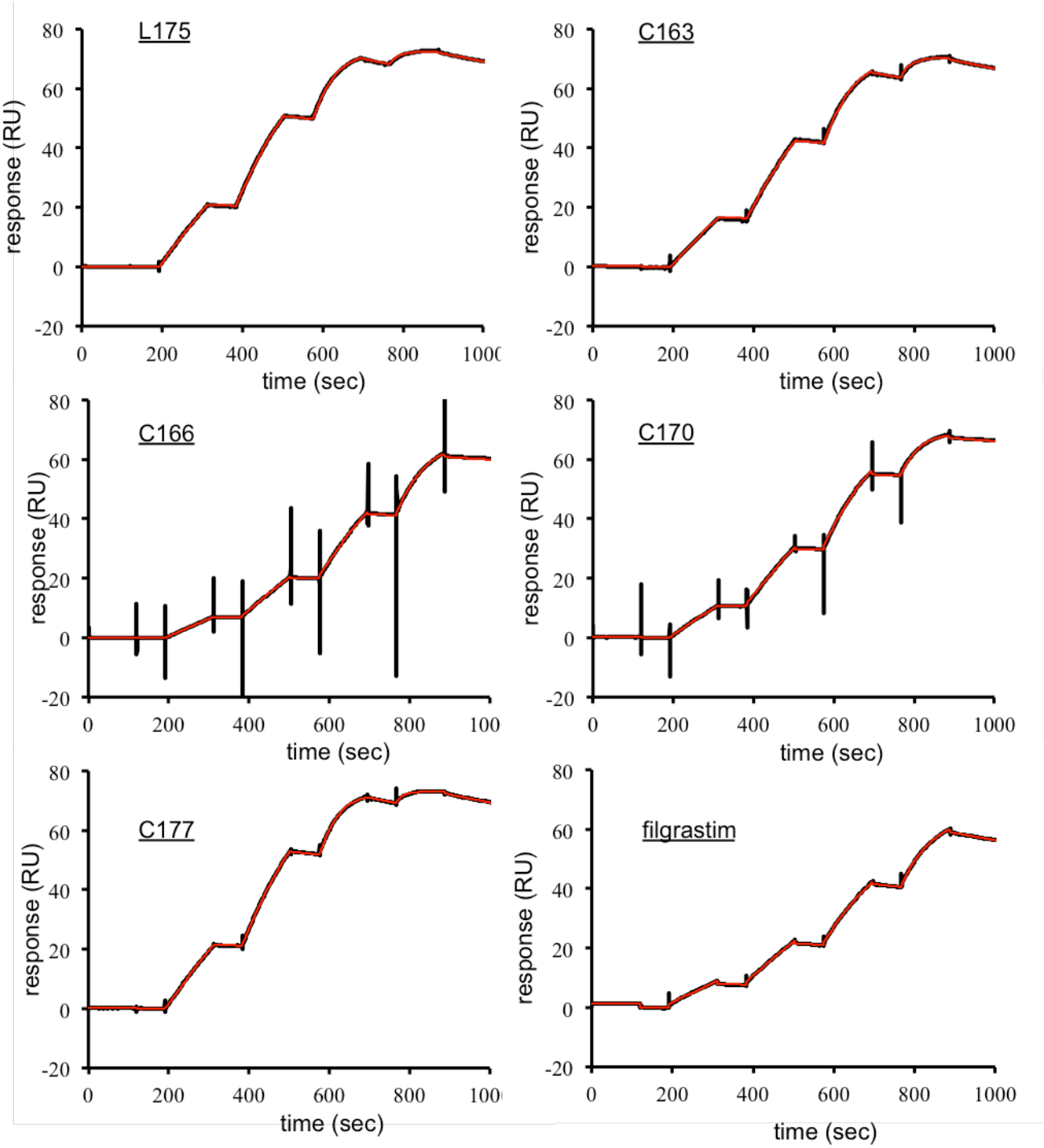

Figure S3 SPR sensorgrams to measure the affinity of circularized G-CSF to G-CSF receptor. Protein A was immobilized on the sensor chip, and an Fc-fused G-CSF receptor was captured. The G-CSF variants were injected in a single cycle.

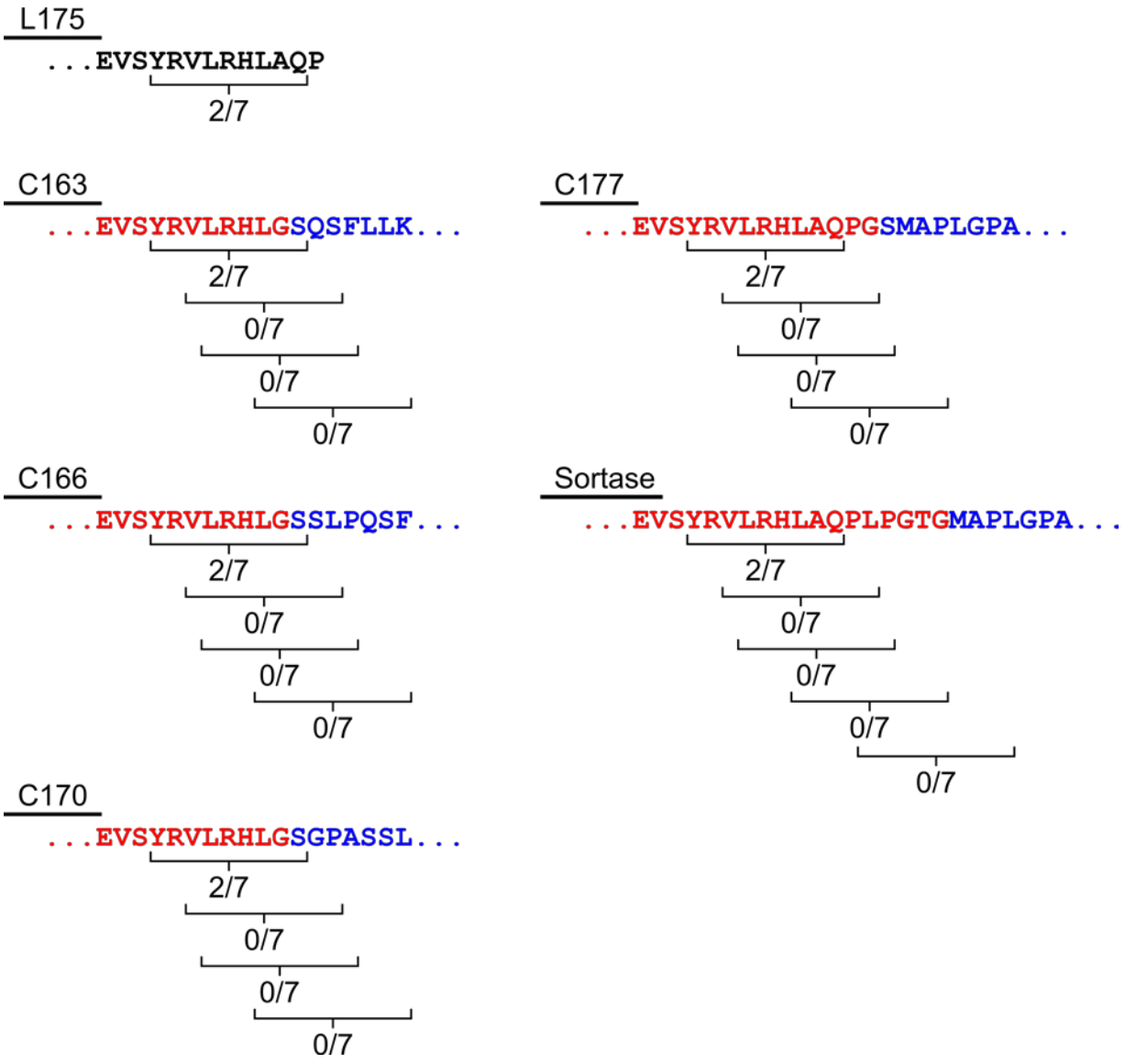

Figure S4. Prediction of immunogenicity. The T-cell epitope predictor we used scores the immunogenicity of each 9mer peptide. The sequence realized by circularization includes a total of four potentially immunogenic peptides. However, none of them raised the possibility of being recognized by the seven representative MHC II alleles.

Supplementary Tables

Table S1 Number of loop templates extracted from the PDB. Related to Figure 1

Loop length	redundant	non-redundant
1	0 (38,920)	0 (10,787)
2	***92*** (36,994)	***10*** (11,365)
3	20 (42,263)	2 (13,266)
4	16 (32,137)	3 (9,203)
5	***19*** (26,551)	***11***(7,975)
6	11 (20,885)	4 (5,712)
7	6 (13,851)	3 (4,171)
8	3 (12,683)	0 (3,547)
9	***128*** (12,529)	***6*** (2,963)
10	4 (8,862)	2 (2,199)
11	0 (6,203)	0 (1,524)
12	0 (4,012)	0 (1,106)
13	0 (3,914)	0 (1,079)
14	1 (2,913)	1 (758)
≥15	1 (11,571)	1 (2,890)
total	301 (274,288)	43 (78,545)

The number of loop templates, having RMSD ≤ 1.0 Å at α-helix patch, are listed in the table. The numbers in parentheses indicate the number of loops without RMSD restriction (but having a resolution ≤ 3.0 Å and α-helices at both ends). Redundant and non-redundant (≤ 90% similarities) datasets were examined. We used the redundant dataset for designing the cyclic GCSF linker. Loop length: the number of residues for the template loops; redundant/non-redundant: obtained from redundant/non-redundant datasets.

Table S2 Top 10 optimal sequences Related to Figure 1

	sequence	Rosetta score	template
1	GS	-21.88	1HVD(A), 73-74
2	GC	-19.84	1AEI(F), 302-303
3	LS	-18.33	1HAK(A), 304-305
4	FS	-18	1N44(A), 302-303
5	ES	-17.56	1A8B(A), 71-72
6	HS	-17.52	1N44(A), 302-303
7	DS	-17.41	2IE7(A), 302-303
8	NS	-17.3	1ANN(A), 302-303
9	AS	-17.2	1HVD(A), 73-74
10	MS	-17.12	1HAK(A), 304-305

Supplemental Methods

Construction of circularizing vector

The DNA that encodes a methionyl human G-CSF with two-point mutations (T2A and C17S) and DNA that encodes a split DnaE intein from *Nostoc puctiforme* were purchased from Takara. For the expression vector of the G-CSF (L175), the methionyl human G-CSF with two-point mutations (T2A and C17S) was amplified and cloned into vector pET16b (Novagen) using the NcoI and BamHI restriction sites. For the expression vector of the circularized G-CSF variants, the split DnaE intein was amplified and cloned into vector pET16b using the NcoI and XhoI restriction sites. DNA-encoded C163, C166, C170, and C177 were inserted using the NheI and NdeI restriction sites. All nucleic-acid sequences are listed below.

1. a split DnaE intein from *Nostoc puctiforme*

ATGGTGAAAATAGCCACACGCAAATATCTGGGCAAACAGAACGTGTATGATATTGGCGTGGAACGCGATCAT
AACTTTGCGCTGAAAAACGGCTTCATAGCTAGCAATTGTTTTAACAAAAGCCATTTTGCGGAATATTGTTTAT
CATATGAAACGGAAATATTGACCGTGGAATATGGCCTGCTGCCGATTGGCAAAATTGTGGAAAAACGCATTG
AATGCACCGTGTATAGCGTGGATAACAACGGCAACATTTATACCCAGCCGGTGGCGCAGTGGCATGATCGCG
GCGAACAGGAAGTGTTTGAATATTGCCTGGAAGATGGCAGCCTGATTCGCGCGACCAAAGATCATAAATTTA
TGACCGTGGATGGCCAGATGCTGCCGATTGATGAAATTTTTGAACGCGAACTGGATCTGATGCGGGTTGATA
ATTTGCCGAATAGCGGCAGCGGCCATCACCATCACCATCACTAA

2. a methionyl human G-CSF with two-point mutations (T2A, C17S)

ATGGCACCATTAGGTCCAGCGAGCAGCCTGCCGCAGAGCTTTCTGCTGAAAAGCCTGGAACAGGTGCGTAAA
ATTCAGGGTGATGGTGCGGCGCTGCAAGAAAAACTGTGCGCGACCTATAAACTGTGCCATCCGGAAGAGCTG
GTGCTGCTGGGCCATAGCCTGGGTATTCCGTGGGCACCGCTGTCTAGCTGTCCGAGCCAGGCGCTGCAACTGG
CCGGTTGTCTGAGCCAGCTGCATAGCGGCCTGTTTCTGTATCAGGGCCTGCTGCAAGCGCTGGAAGGCATTAG
CCCGGAGCTGGGCCCGACTCTGGATACCCTGCAACTGGATGTGGCGGATTTTGCGACCACCATTTGGCAGCA
GATGGAAGAGCTGGGCATGGCACCGGCGCTGCAACCGACCCAGGGTGCCATGCCGGCGTTTGCGAGCGCGTT
TCAGCGTCGTGCGGGCGGTGTTCTGGTGGCGAGCCATCTGCAATCTTTTCTGGAAGTGAGCTATCGTGTGCTG
CGTCATCTGGCCCAGCCG

3. forward primer for L175

TATACCATGGCACCATTAGGTCCAGCGAGCAGCCTGCCGCAGAGCTTT

4. back primer for L175

ATGGATCCTTACGGCTGGGCCAGAT

5. forward primer for C177

AAAGCTAGCAATAGCATGGCACCATTAGGTCCA

6. back primer for C177

AAAAATTCATATGATAAACAGCCCGGCTGGGCCAGATG

7. forward primer for C170

AAAGCTAGCAATAGCGGTCCAGCGAGCAGC

8. back primer for C170

AAAAATTCATATGATAAACAGCCCAGATGACGCAGCACACG

9. forward primer for C166

AAAGCTAGCAATAGCAGCCTGCCGCAG

10. back primer for C166

AAAAATTCATATGATAAACAGCCCAGATGACGCAGCACACGATAG

11. forward primer for C163

AAAAAAGCTAGCAATAGCCAGAGCTTTCTGCTGAAA

12. back primer for C163

AAAAATTCATATGATAAACAGCCCAGATGACGCAGCAC

Design of terminal regions

The circularized G-CSF was rationally designed based on three-dimensional structures of a human G-CSF. Because the positions of terminal helices A and D maintain rigid structures among the several X-ray crystal structures of the G-CSF, we aimed to use the fragment of the loop to link the N- and C-terminals of the G-CSF together. A G-CSF scaffold structure was chosen from a complex structure with human GCSF receptor (PDB: 2D9Q; chain A). The length and amino-acid sequence of the connecting segment were determined by the frequency of appearance in the PDB and free-energy-based (Rosetta score)[1] optimization, respectively.

Backbone and side-chain optimizations were conducted for all possible sequences on the two-residue connector, as explained next. For a given amino-acid sequence of the loop, the side-chain structures in each template loop were initially modeled ahead of the optimization. Then, potential energy minimization was performed for 100 steps to eliminate unfavorable distortion in the backbone structure in the patched regions. After the backbone optimization, rotamer optimization was performed to refine the side-chain structures. We used the FASTER algorithm[2] and Rosetta score function[1] for the initial construction of the side-chain and rotamer optimizations implemented in the CHOMP program[3]. The molecular dynamic package AMBER[4] and AMBER force field[5] were also used for the energy minimization.

Expression and purification

Expression, refolding, and purification were partially performed in reference to previous studies[6,7]. All the G-CSF variants (L175, C177, C170, C166, and C163) were expressed in *E. coli* BL21 (DE3) strains. The cells were incubated at 37 °C in Lysogeny broth for 4 h. To induce protein expression, 200-μM Isopropyl β-D-1-thiogalactopyranoside was added to the culture. The cells were incubated for 4 h after induction and thereafter harvested by centrifugation.

The harvested cells were resuspended in a lyses buffer containing 50-mM Tris-HCl with pH of 8.0, 1% (w/v) sodium deoxycholate (DOC), 1.2-kU/ml lysozyme, and 25-U/mL benzonase® nuclease and incubated for 15 min at room temperature. After centrifugation, the pellets were washed (resuspension and centrifugation) in three steps using buffers containing 50-mM Tris-HCl, 5-mM ethylenediaminetetraacetic acid (EDTA), and three different wash reagents [(1) 2% (w/v) Tween20, (2) 1% (w/v) DOC, and (3) 1-M NaCl]. The washed inclusion bodies were solubilized in a buffer containing 50-mM Tris-HCl with pH of 8.0, 5-mM EDTA, and 6-M guanidine hydrochloride. The supernatants were diluted into 20 volumes of refolding buffer containing 50-mM Tris-HCl, 2-mM EDTA 400-mM Arg-HCl, 1-mM reduced glutathione, and 0.1-mM oxidized glutathione. After equilibration in a buffer containing 20-mM Tris-HCl with pH of 8.0, purification was performed.

L175 was purified in two steps. First, we employed anion-exchange chromatography in a HiTrap Q HP column (GE Healthcare). Second, the fractions containing G-CSF were subjected to gel-filtration chromatography in a Superdex 75 10/300 GL column (GE Healthcare) equilibrated with 20-mM Tris-HCl with pH of 8.0 and 150-mM NaCl. The circularized variants (C177, C170, C166, and C163) were subjected to one more purification step. Subsequently, the fractions of the G-CSF were dialyzed into a buffer containing 20-mM Tris-HCl with pH of 8.0 and were purified in a MonoQ 5/50 column (GE Healthcare).

The fractions containing G-CSF were concentrated on a 10-kDa Centriprep filtration unit (Merck Millipore). The protein concentration was determined by ultraviolet absorbance at 280 nm using the calculated molar extinction coefficient of the G-CSF.

SDS-PAGE

SDS-PAGE was performed to confirm the purity and to quantify the proteolytic stability assay. Some 10- to 100-ng of G-CSF samples were loaded in each well of 4%–20% Mini-PROTEAN® TGX™ Gel (Bio-Rad) and stained using Oriole™ fluorescent gel stain (Bio-Rad). All procedures followed the protocol described by the supplier.

Mass Spectrometry (MS)

Liquid chromatography–MS experiments were performed using a hybrid electrospray quadrupole time-of-flight mass spectrometer (Synapt G2 HDMS, Waters). Samples of G-CSF variants were dissolved in 100-mM ammonium carbonate and 1-mM dithiothreitol with pH of 8.0 at a concentration of 100 μg/mL. In all measurements, 10 μL of samples

was injected and desalted online using a MassPREP micro desalting cartridge with a dimension of 2.1 × 5 mm, 20 μm, and 1000 Å (Waters, Manchester). Calibration was performed using singly charged ions produced by a 2-μg/mL solution of cesium iodide in 2-propanol/water (1/1). Data analysis was performed using MassLynx 4.1 (Waters, Manchester, U. K.).

CD

The CD spectra and CD melting curves were recorded by a J-805 spectropolarimeter (JASCO). The proteins were dissolved at 5 μm in 10-mm HEPES-NaOH with pH of 7.4 and 150-mM NaCl. The spectra were obtained at 10 and 90 °C and presented in units of molecular ellipticity per mol of residue. The melting curves were obtained by monitoring the ellipticity at 222 nm while increasing the temperature from 10 to 90 °C at a heating rate of 1.0 °C/min. The thermodynamic parameters of the proteins for equilibrium unfolding were obtained through fitting calculation using a two-state transition model[8]. Numerical-fitting calculations were carried out using the IGOR software (Wavemetrics, Inc., Lake Oswego, OR).

SPR

SPR assay was performed using Biacore T200 (GE Healthcare). Because the human-cell-expressed G-CSF receptor Fc Chimera (SYMANSYS) was denatured when covalently immobilized on sensor chip CM5 (GE Healthcare), it was captured on sensor-chip-immobilized protein A via Fc-protein A interaction. Protein A (Nacalai, Japan) was first covalently immobilized on sensor chip CM5 (GE Healthcare) via amine coupling. The G-CSFR-Fc in the running buffer [10-mM HEPES-NaOH, 150-mm NaCl, 0.05% (v/v) surfactant Tween 20, and pH of 7.4] was then injected and captured by protein A. Kinetic characterization of the G-CSF was performed using the single-cycle kinetic method. Sensorgrams were obtained by injecting G-CSF with increasing concentrations from 1.25 to 20 nM at a flow rate of 30 μL/min. After each measurement, the sensor chip was regenerated by removing the G-CSFR-Fc using a solution containing 10-mM sodium acetate with pH of 5.0. All measurements were performed in duplicate.

Digestion by elastase

The G-CSF variants at a concentration of 10 μg/mL were mixed with 1 μg/mL elastase (Wako) in a buffer containing 100-mM HEPES-NaOH with pH of 7.4. The solutions were incubated at 37 °C for 15, 30, 45, 60, 90, and 120 min (n = 3 for all samples). The elastase reaction was stopped by cooling in an ice bath and subsequently adding phenylmethanesulfonyl fluoride at a final concentration of 1 mM. Thereafter, the degraded samples were diluted 50-fold and measured using AlphaLISA technique, EnVision-Alpha Reader (PerkinElmer), and Human G-CSF Kit (PerkinElmer) according to the manufacturer instruction. The concentration was calculated using the calibration curves shown by the non-degraded G-CSF variants at a concentration of 0.5–500 ng/mL.

Supplemental References

(1) Rohl, C. A., Strauss, C. E. M., Misura, K. M. S., and Baker, D. (2004) Protein structure prediction using Rosetta. *Methods Enzymol. 383*, 66–93.
(2) Allen, B. D., and Mayo, S. L. (2006) Dramatic performance enhancements for the FASTER optimization algorithm. *J. Comput. Chem. 27*, 1071–1075.
(3) Loksha, I. V, Maiolo, J. R., Hong, C. W., Ng, A., and Snow, C. D. (2009) SHARPEN-systematic hierarchical algorithms for rotamers and proteins on an extended network. *J. Comput. Chem. 30*, 999–1005.
(4) Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem. 26*, 1668–1688.
(5) Ponder, J. W., and Case, D. A. (2003) Force fields for protein simulations. *Adv. Protein Chem. 66*, 27–85.
(6) Arakawa, T., Prestrelski, S. J., Narhi, L. O., Boone, T. C., and Kenney, W. C. (1993) Cysteine 17 of recombinant human granulocyte-colony stimulating factor is partially solvent-exposed. *J. Protein Chem. 12*, 525–531.
(7) Luo, P., Hayes, R. J., Chan, C., Stark, D. M., Hwang, M. Y., Jacinto, J. M., Juvvadi, P., Chung, H. S., Kundu, A., Ary, M. L., and Dahiyat, B. I. (2002) Development of a cytokine analog with enhanced stability using computational ultrahigh throughput screening. *Protein Sci. 11*, 1218–1226.
(8) Honda, S., Kobayashi, N., and Munekata, E. (2000) Thermodynamics of a beta-hairpin structure: evidence for cooperative formation of folding nucleus. *J. Mol. Biol. 295*, 269–278.