Supplementary information

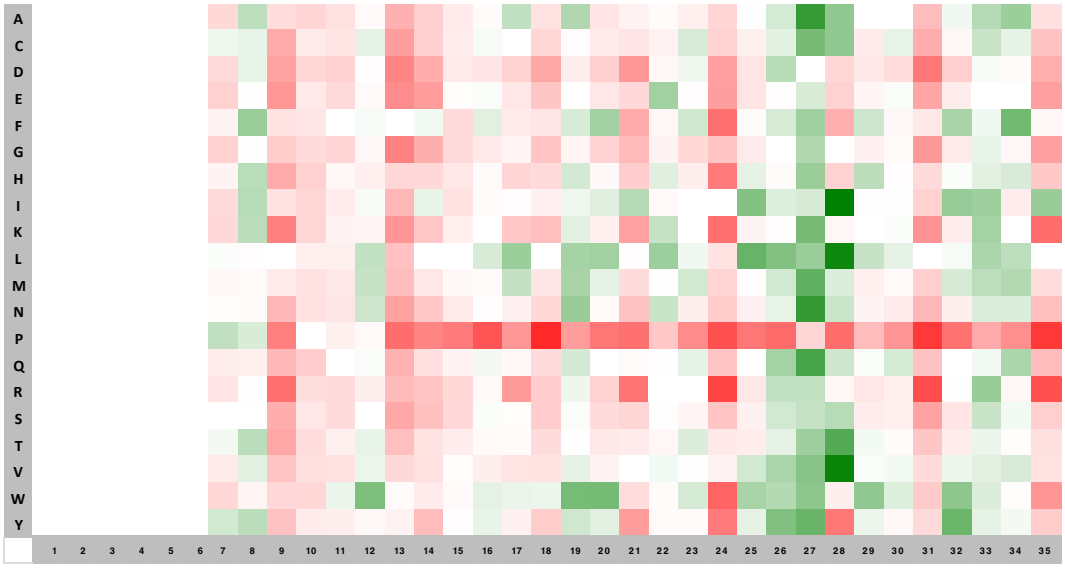
Protein engineering and HDX identifies structural regions of G-CSF critical to its stability and aggregation

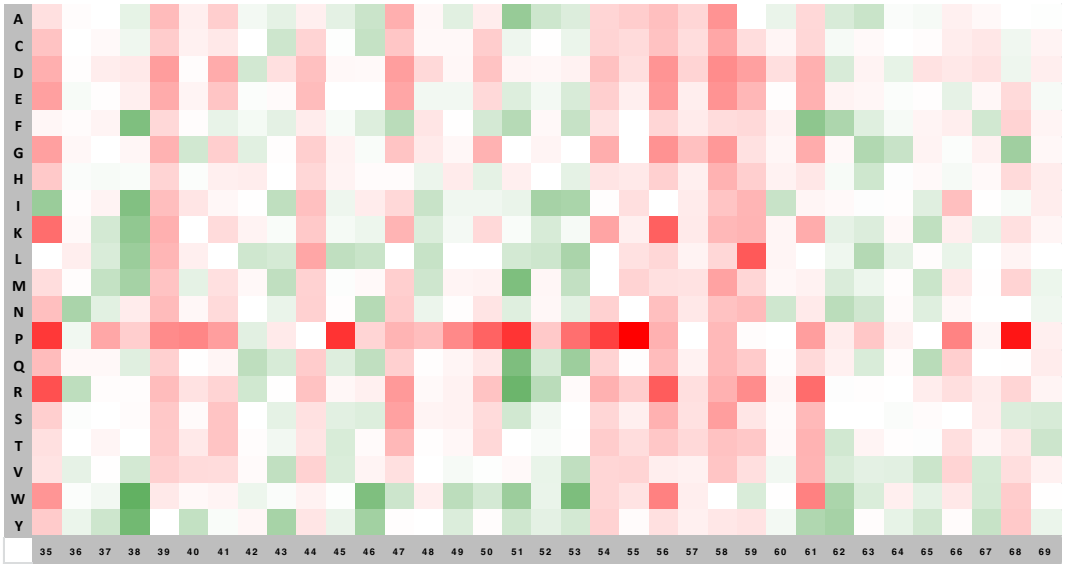
Victoria E. Wood[1], Kate Groves[3], Lok Man Wong[1], Luyan Kong[1], Christopher Bird[2], Meenu Wadhwa[2], Milena Quaglia[3], Paul Matejtschuk[2], Paul A. Dalby[1]

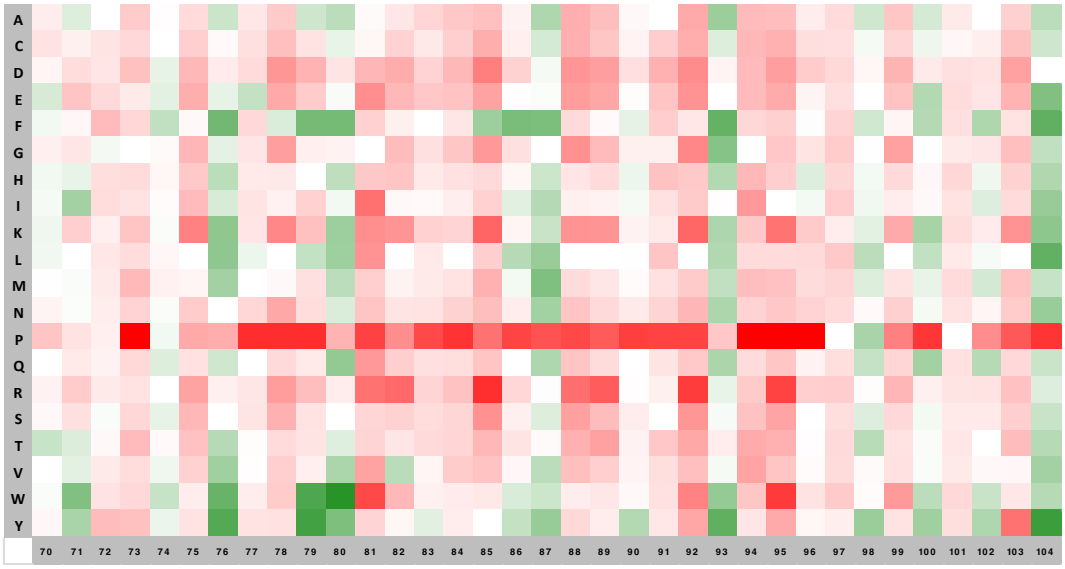
[1]Department of Biochemical Engineering, University College London, Gower Street, London, WC1E 6BT

[2]National Institute for Biological Standards and Control (NIBSC), Blanche Lane, South Mimms, Potters Bar, Hertfordshire, EN6 3QG

[3]National Measurement Laboratory at LGC Ltd, Queens Road, Teddington, TW11 0LY


A C D E F G H I K L M N P Q R S T V W Y
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35


A C D E F G H I K L M N P Q R S T V W Y
35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69


A C D E F G H I K L M N P Q R S T V W Y
70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104

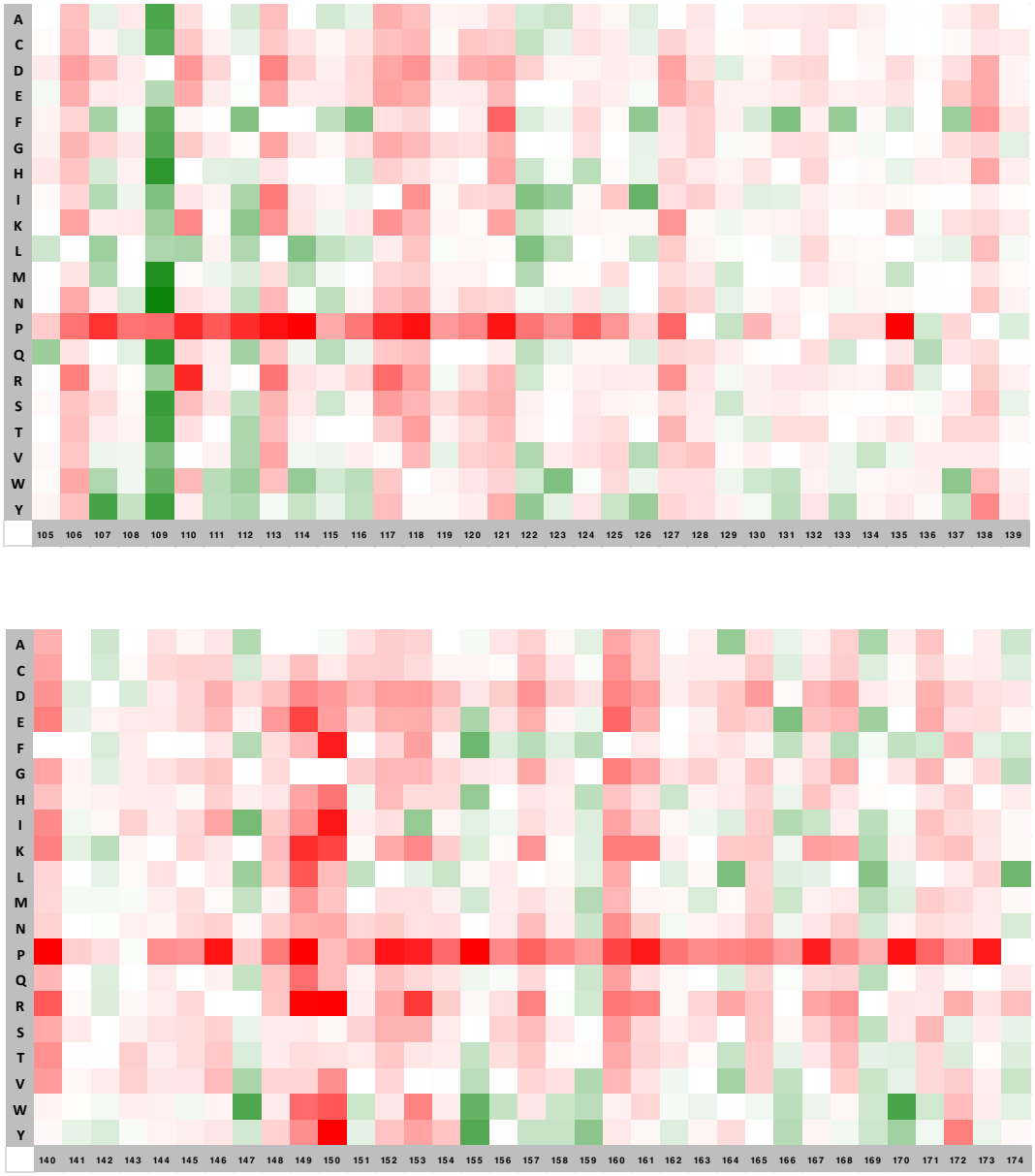


Figure S1. Heat map of ΔΔ*G* values for all point mutations of G-CSF as predicted by Rosetta_ddg_monomer. The x-axis represents the protein residue number and the y-axis represents the possible amino acid substitutions. The colour key represents the value of ΔΔ*G*, where green indicates negative values (stabilising), white indicates a value of 0 or no mutation, and red indicates a positive (destabilising) value.

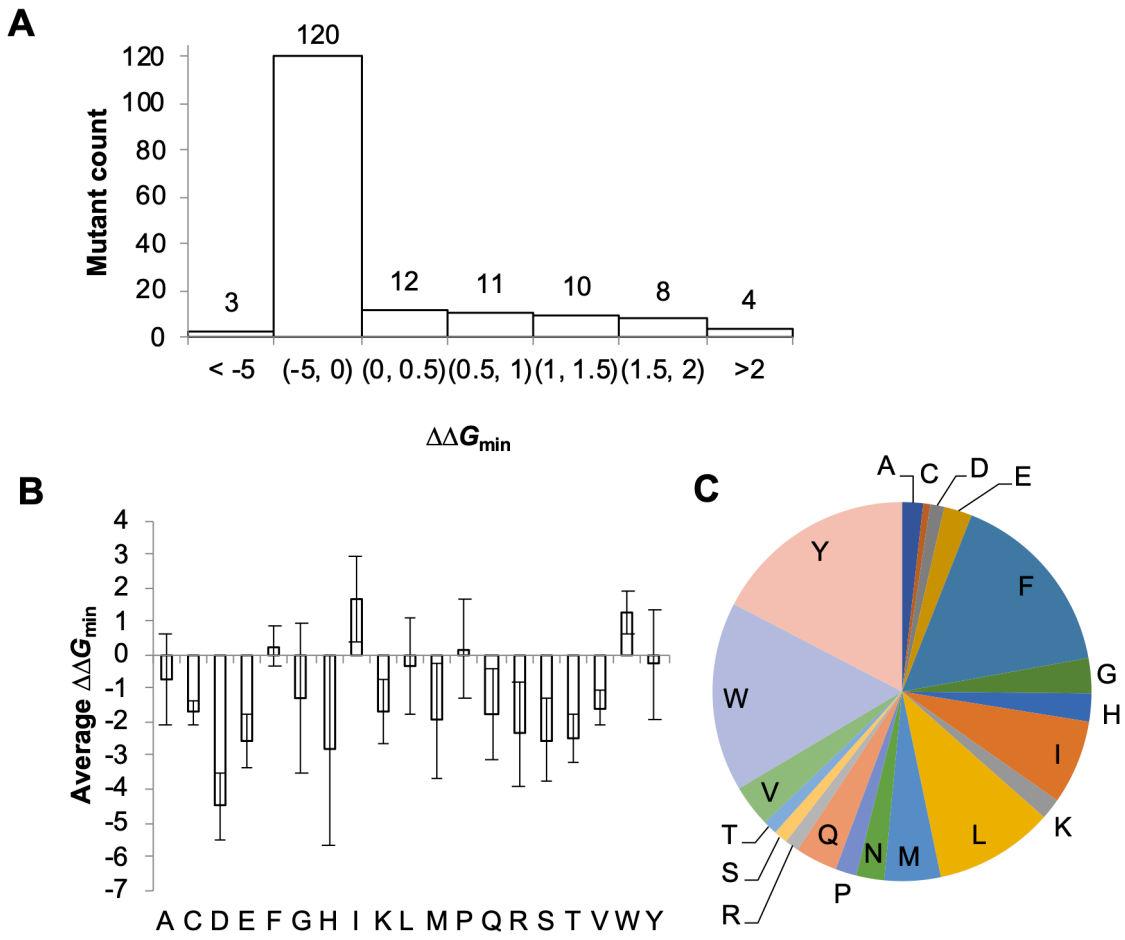


Figure S2. Lowest ΔΔ*G* variant analysis. $\Delta\Delta G_{min}$ values were obtained from the lowest ΔΔG mutant per residue of G-CSF as predicted by Rosetta_ddg_monomer. A) Histogram for the distribution of $\Delta\Delta G_{min}$ values obtained, B) Average $\Delta\Delta G_{min}$ obtained for each target residue type within G-CSF C) Selection frequency for amino acids having $\Delta\Delta G_{min}$.

The average $\Delta\Delta G_{min}$ for each amino acid type within the residues of G-CSF (Figure S2B) show that residues most likely to be stabilised (negative average $\Delta\Delta G_{min}$) included cysteine, aspartic acid, glutamic acid, lysine, methionine, glutamine, arginine, serine, threonine and valine. Residues most likely to be destabilized (positive average $\Delta\Delta G_{min}$) included isoleucine and tryptophan. The frequency of amino acids introduced to give the lowest $\Delta\Delta G$ at each residue (Figure S2C) showed that mutations towards tyrosine, tryptophan, phenylalanine, leucine and isoleucine were most likely to lead to stabilisation. Interestingly, these matched the target amino-acid types most likely to be destabilised or unchanged by mutations in Fig S2B.

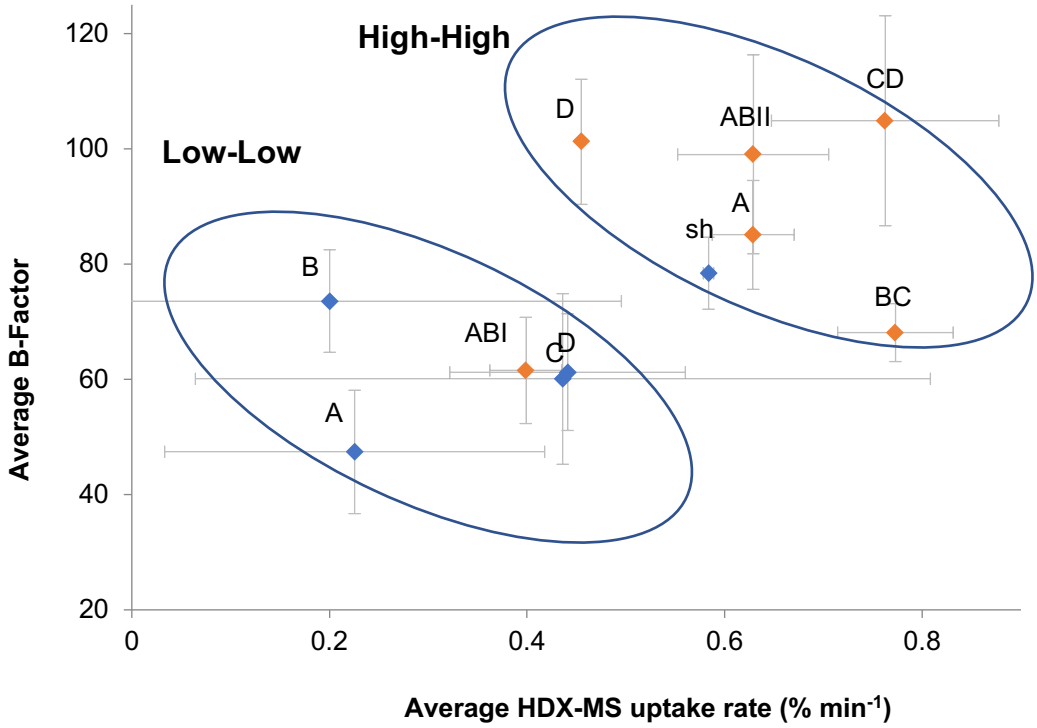


Figure S3. Comparison of HDX-MS average uptake rates (% min^{-1}) and B-factors from PDBID: 2D9Q, each averaged for different structural regions of G-CSF. The α-helices are coloured in blue and loops in orange. Ellipses group data with high values on both axes (High-High) and low values on both axes (Low-Low). HDX exchange data was obtained previously at pH 4.25 (Wood et al 2020).

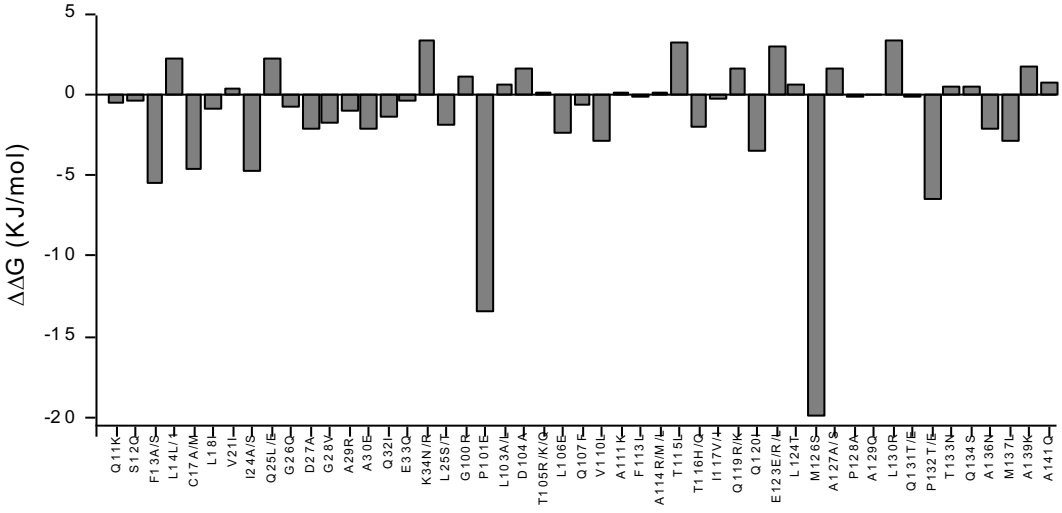


Figure S4. RosettaDesign-Flexibility predicted mutations and ΔΔ*G* values. The lowest energy yielding mutation for each G-CSF target residue is shown along the x-axis and the average of 10 ΔΔ*G* values plotted along the y-axis, as predicted by RosettaDesign with the G-CSF 3D structure PDB:2D9Q (Tamada et al. 2006).

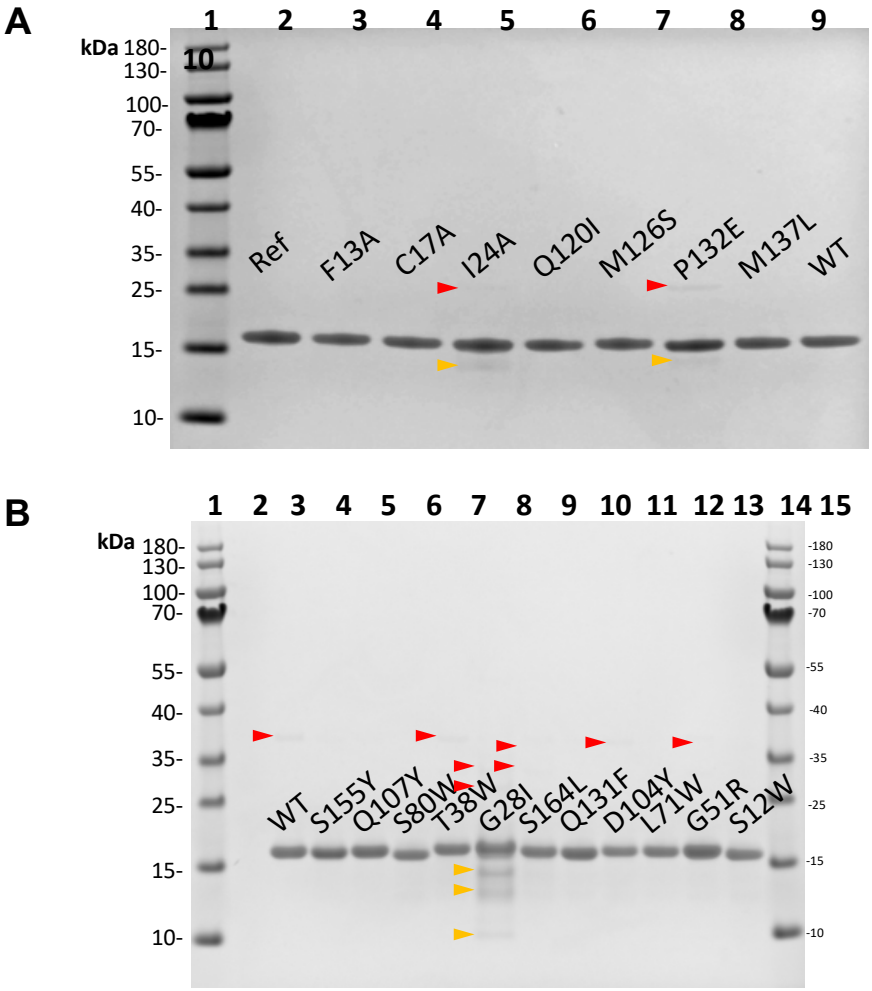
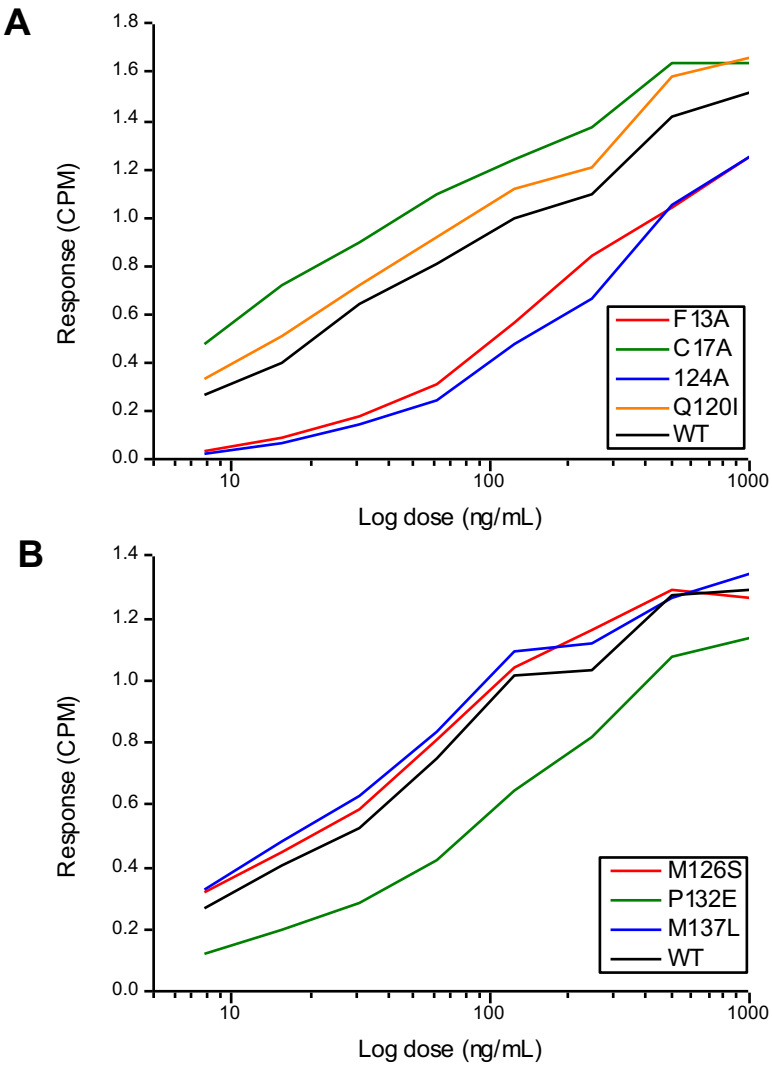


Figure S5. Non-reduced SDS-PAGE analysis of G-CSF WT and variant purified samples. A) RosettaDesign-Flexibility variants, B) Rosetta_ddg_monomer-Unguided variants. Ref is a NIBSC reference standard. Higher molecular weight impurities are highlighted with a red arrow and lower molecular weight impurities are highlighted with an orange arrow.


A
Response (CPM)
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
10
100
1000
Log dose (ng/mL)
F13A
C17A
I24A
Q120I
WT
B
Response (CPM)
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
10
100
1000
Log dose (ng/mL)
M126S
P132E
M137L
WT

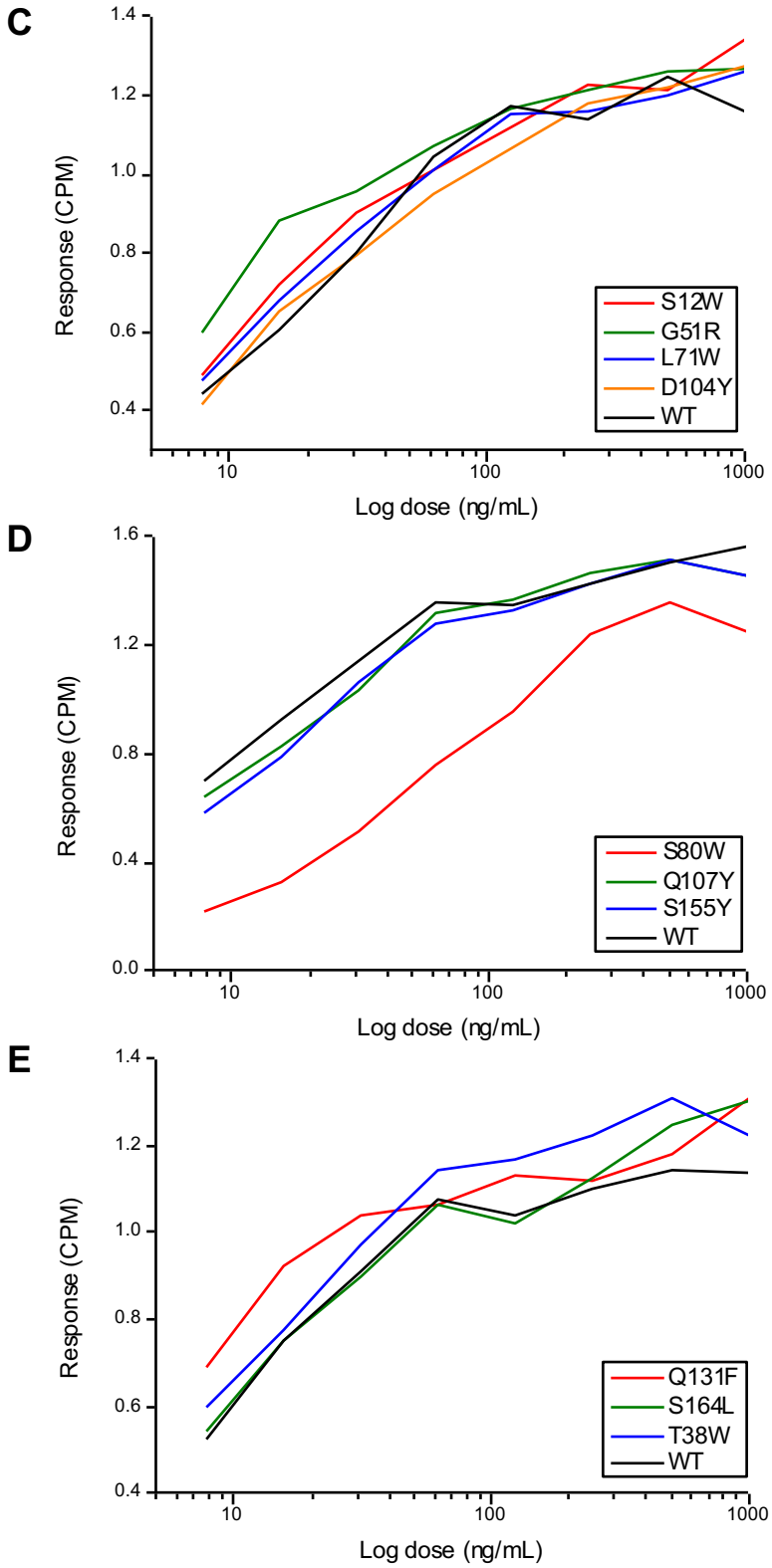
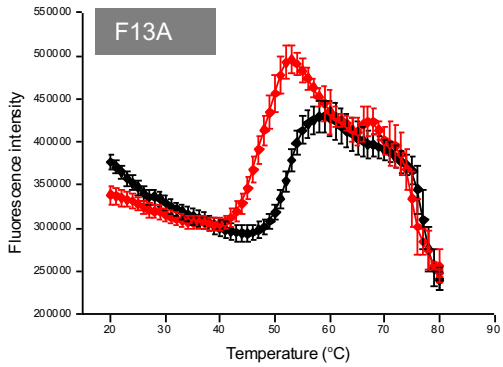
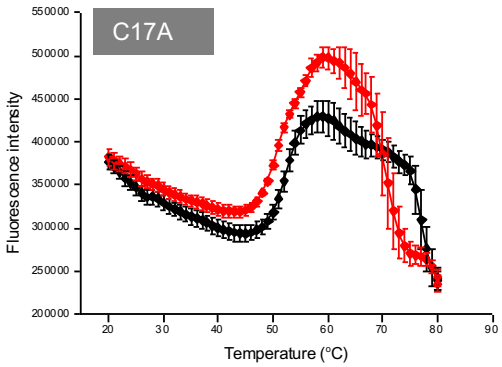
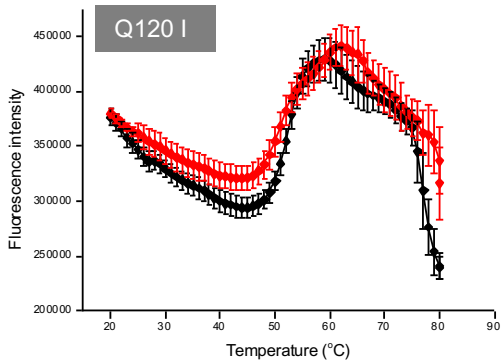
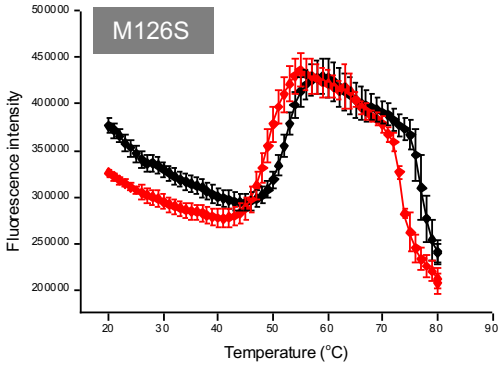


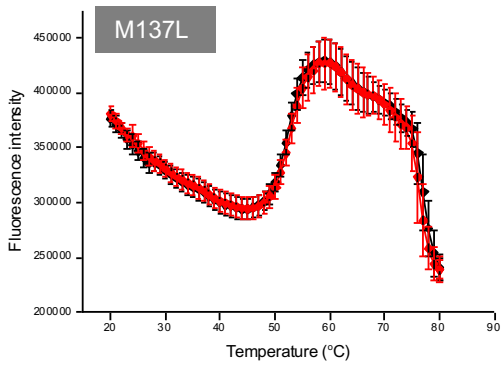
Figure S6. Comparison of G-CSF mutant bioactivity relative to WT using the GNFS-60 cell proliferation bioassay. A-E represent different microtitre plates, and the variants alongside a WT standard sample measured within each, are displayed in the individual figure legends.

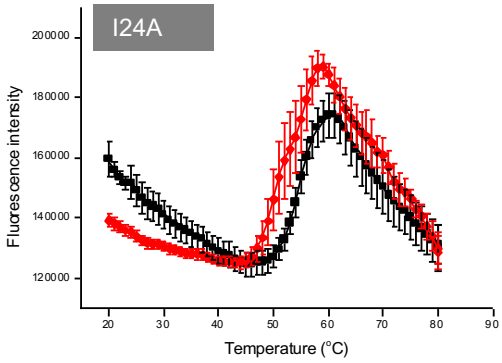

F13A
Fluorescence intensity
Temperature (°C)

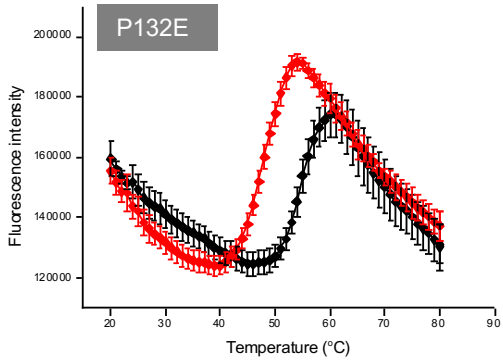

C17A
Fluorescence intensity
Temperature (°C)

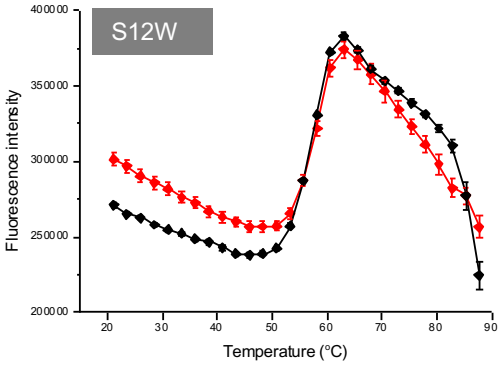

Q120 I
Fluorescence intensity
Temperature (°C)

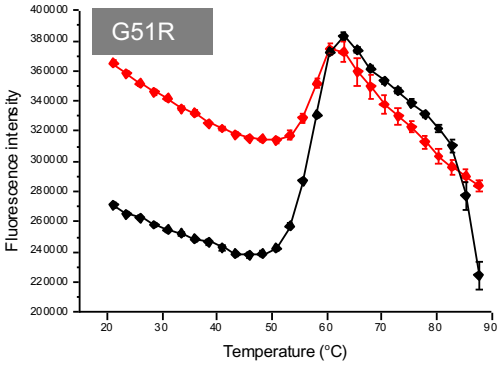

M126S
Fluorescence intensity
Temperature (°C)

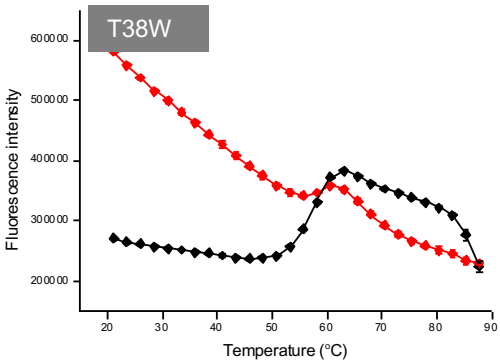

M137L
Fluorescence intensity
Temperature (°C)

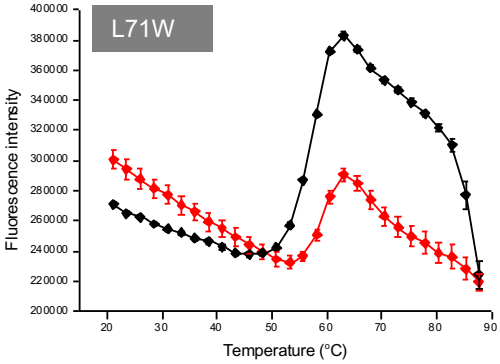

I24A
Fluorescence intensity
Temperature (°C)

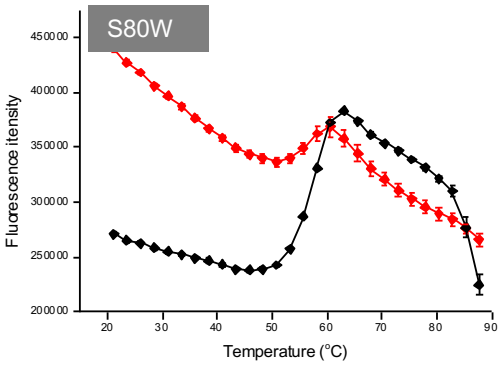

P132E
Fluorescence intensity
Temperature (°C)

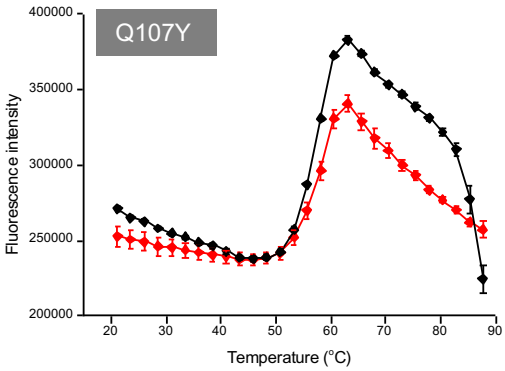

S12W
Fluorescence intensity
Temperature (°C)


G51R
Fluorescence intensity
Temperature (°C)


T38W
Fluorescence intensity
Temperature (°C)


L71W
Fluorescence intensity
Temperature (°C)


S80W
Fluorescence itensity
Temperature (°C)


Q107Y
Fluorescence intensity
Temperature (°C)

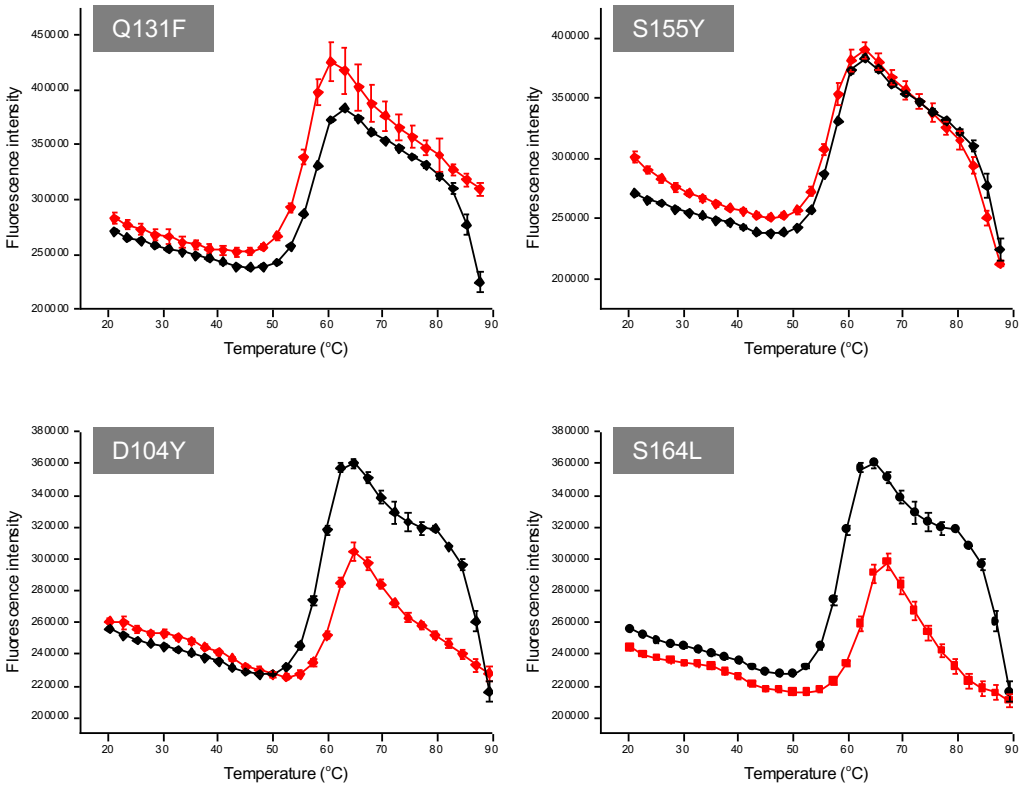
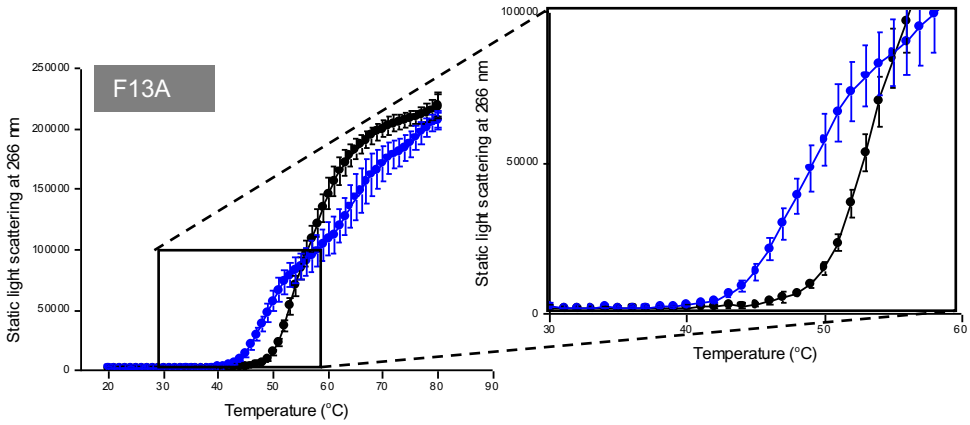
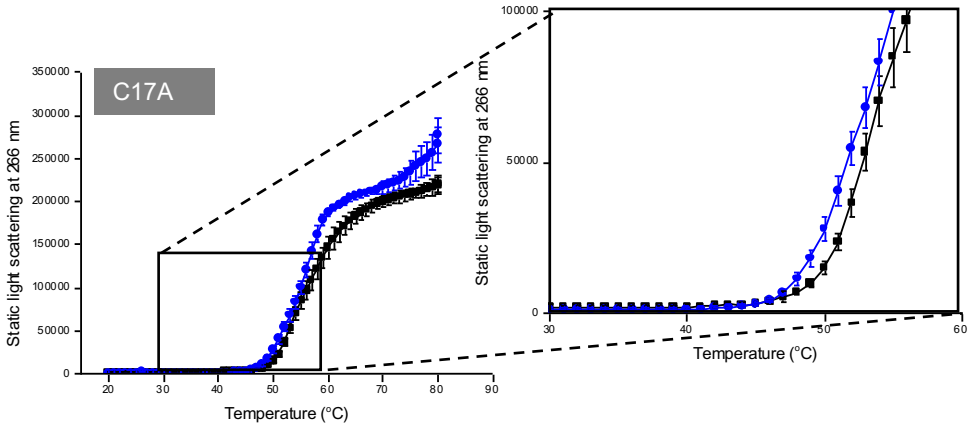
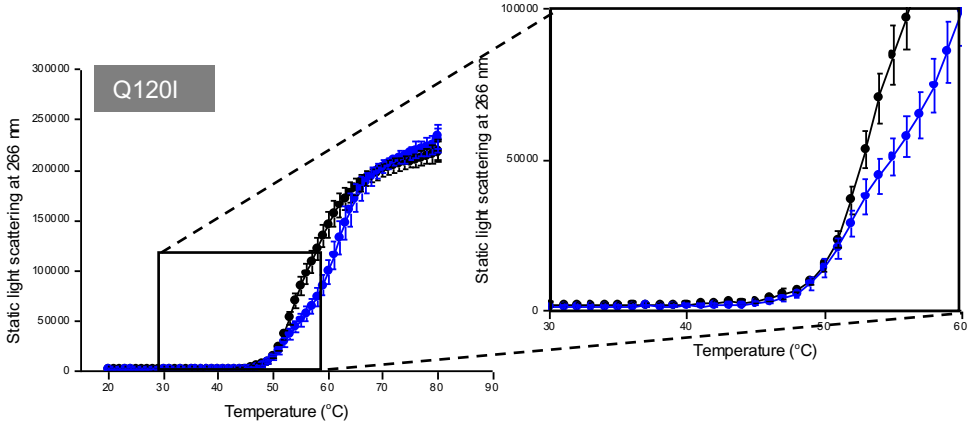
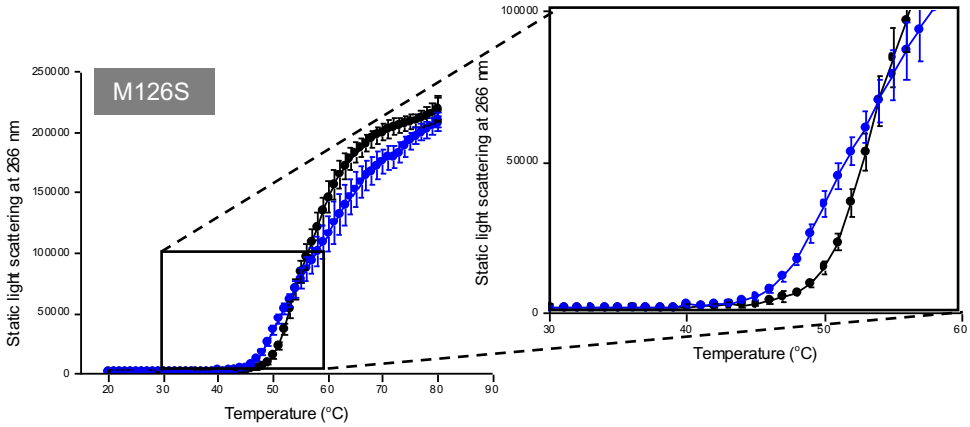


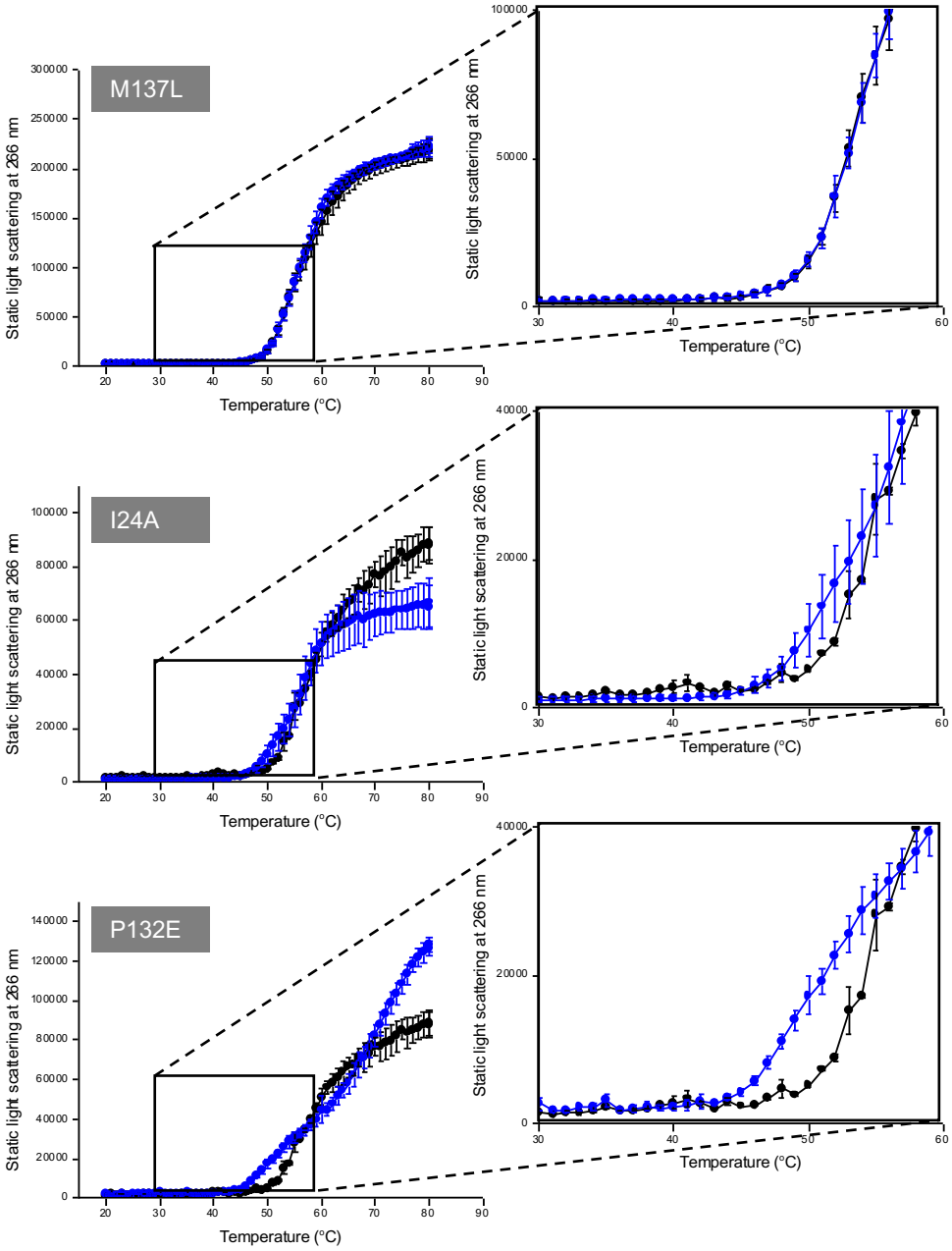
Figure S7. Thermal denaturation profiles measured using intrinsic fluorescence for all G-CSF variants, compared directly to that of WT G-CSF. Variants analysed in each figure are named in the grey boxes. Fluorescence measurements were obtained every 1 or 2 °C during a thermal ramp from 20 to 80 or 90°C at a step rate of 1 °C/min. Variants were formulated at 1 mg/mL in 10 mM sodium acetate pH 4.25. Each formulation was made with five replicates for variants (red), and ten replicates for WT (black), fitted independently to two-state transitions (Eq 1 in manuscript), and their data averaged.

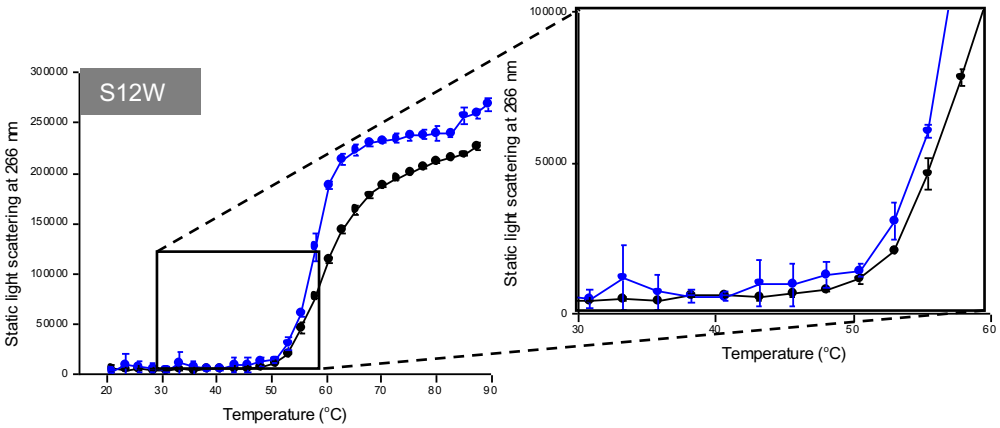

F13A
Static light scattering at 266 nm
Temperature (°C)
Static light scattering at 266 nm
Temperature (°C)

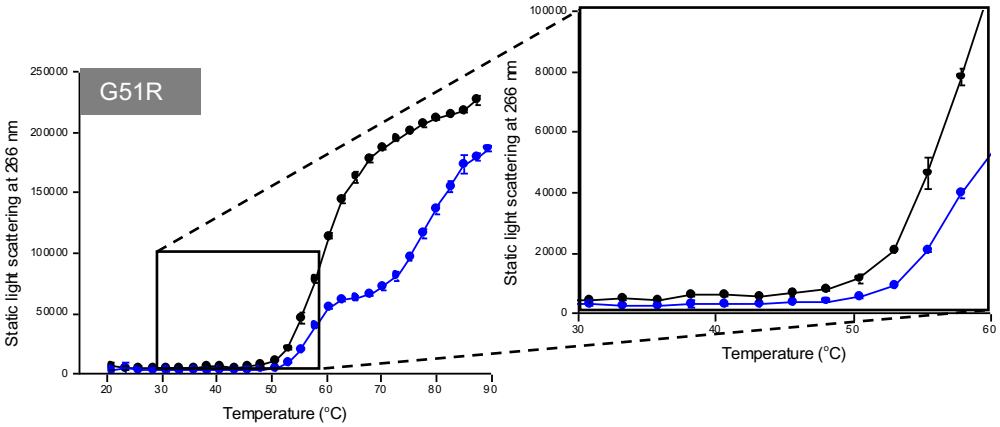

C17A
Static light scattering at 266 nm
Temperature (°C)
Static light scattering at 266 nm
Temperature (°C)

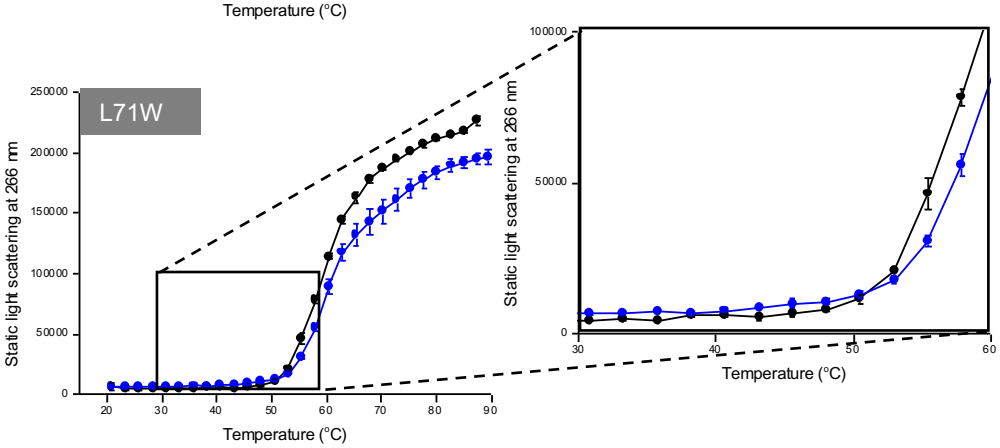

Q120I
Static light scattering at 266 nm
Temperature (°C)
Static light scattering at 266 nm
Temperature (°C)

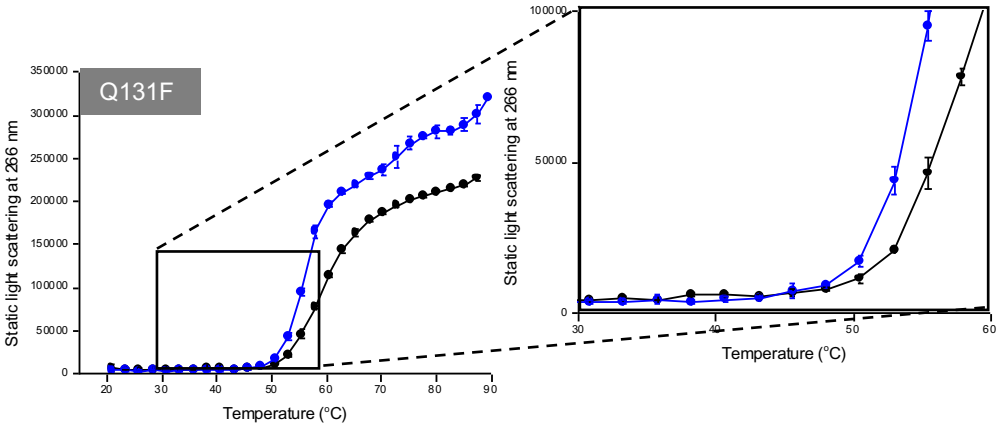

M126S
Static light scattering at 266 nm
Temperature (°C)
Static light scattering at 266 nm
Temperature (°C)

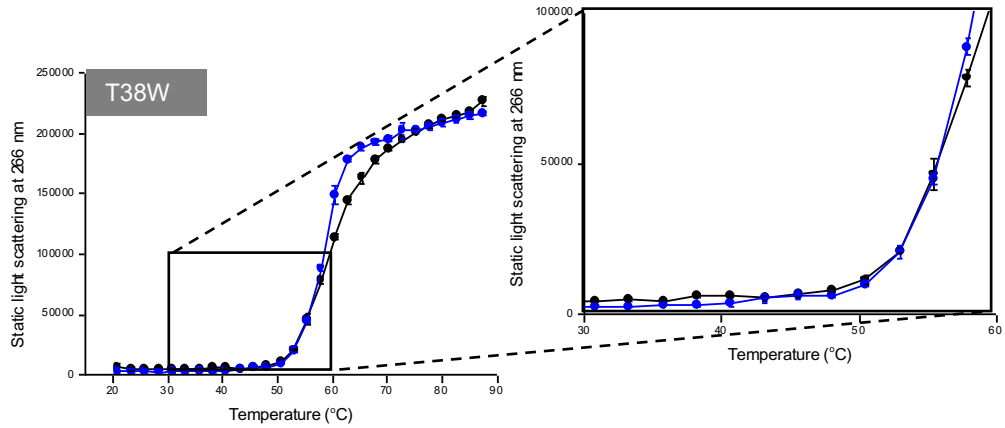

M137L
I24A
P132E
Static light scattering at 266 nm
Temperature (°C)

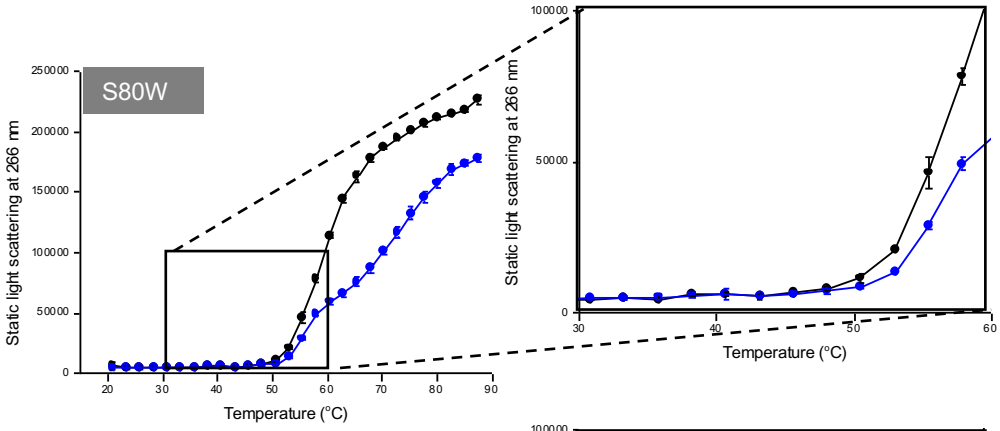

S12W
Static light scattering at 266 nm
Temperature (°C)
Static light scattering at 266 nm
Temperature (°C)

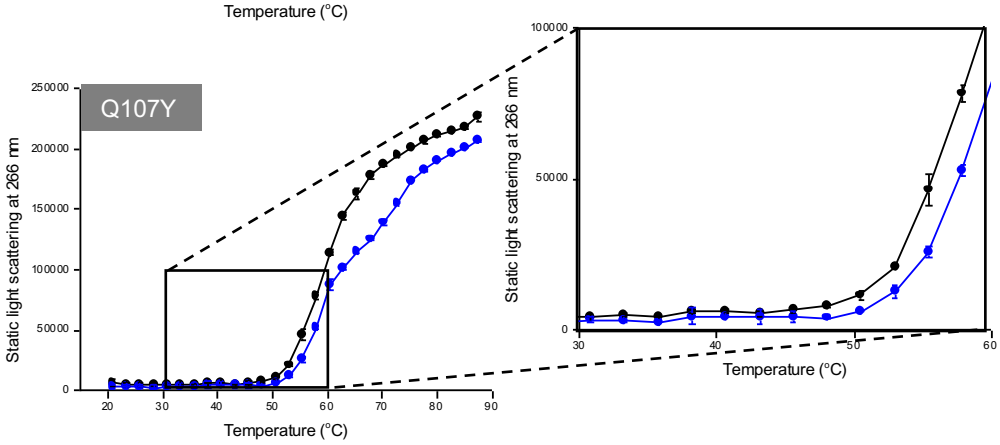

G51R
Static light scattering at 266 nm
Temperature (°C)
Static light scattering at 266 nm
Temperature (°C)

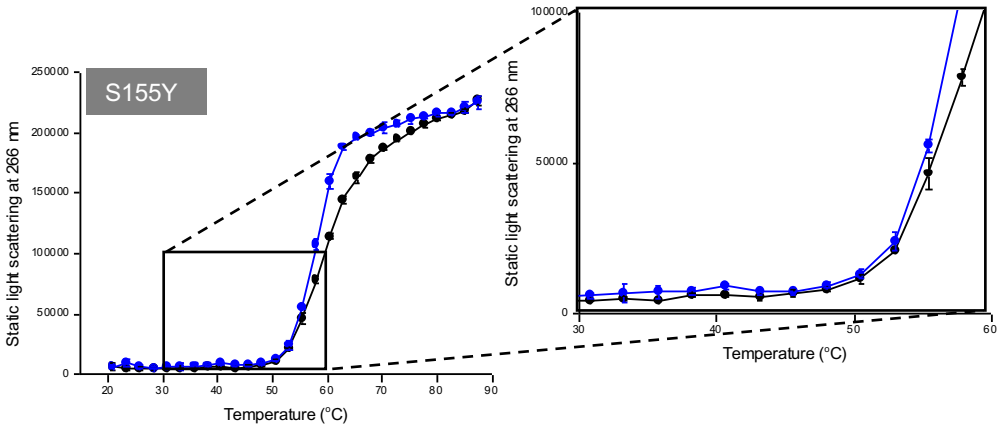

L71W
Static light scattering at 266 nm
Temperature (°C)
Static light scattering at 266 nm
Temperature (°C)


Q131F
Static light scattering at 266 nm
Temperature (°C)
Static light scattering at 266 nm
Temperature (°C)


T38W
Static light scattering at 266 nm
Temperature (°C)
Static light scattering at 266 nm
Temperature (°C)


S80W
Static light scattering at 266 nm
Temperature (°C)
Static light scattering at 266 nm
Temperature (°C)


Q107Y
Static light scattering at 266 nm
Temperature (°C)
Static light scattering at 266 nm
Temperature (°C)


S155Y
Static light scattering at 266 nm
Temperature (°C)
Static light scattering at 266 nm
Temperature (°C)

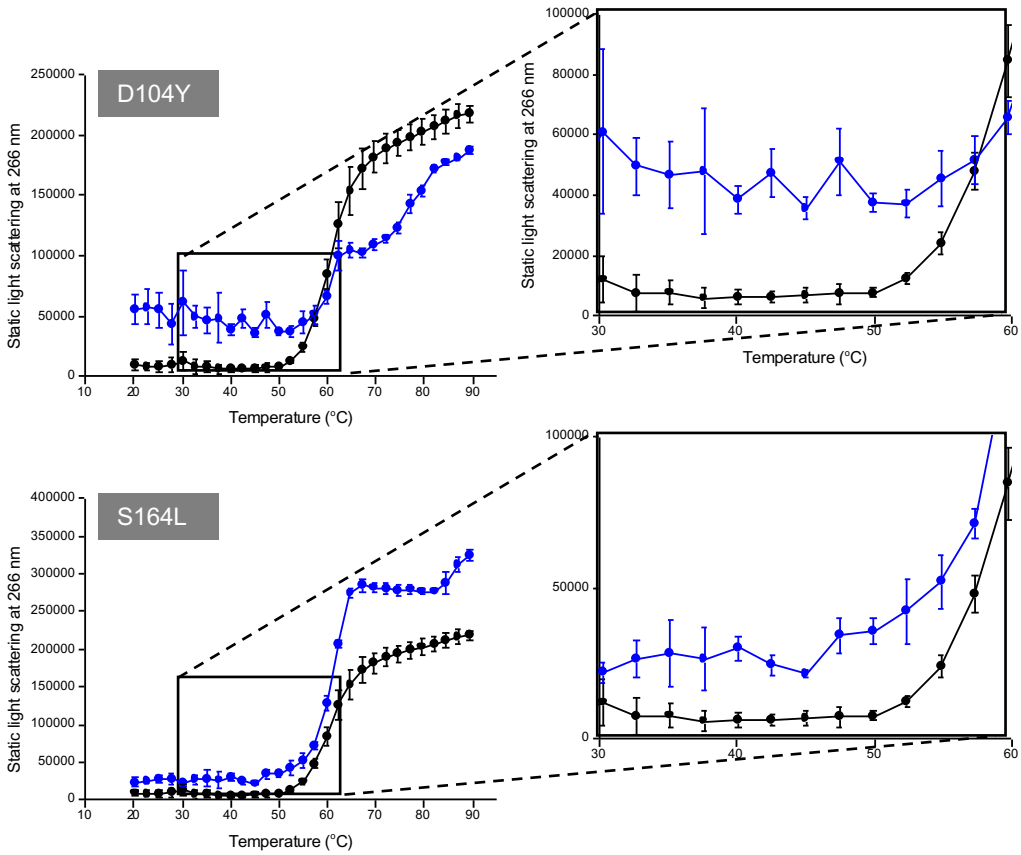


Figure S8. Thermal denaturation profiles measured using static light scattering for all G-CSF variants, compared directly to that of WT G-CSF. Variants analysed in each figure are named in the grey boxes. Static slight scattering measurements at 266 nm were obtained simultaneously with fluorescence measurements, every 1 or 2 °C during a thermal ramp from 20 to 80 or 90°C at a step rate of 1 °C/min. Variants were formulated at 1 mg/mL in 10 mM sodium acetate pH 4.25. Each formulation was made with five replicates for variants (blue), and ten replicates for WT (black), fitted independently and their resulting T_{agg} data averaged.

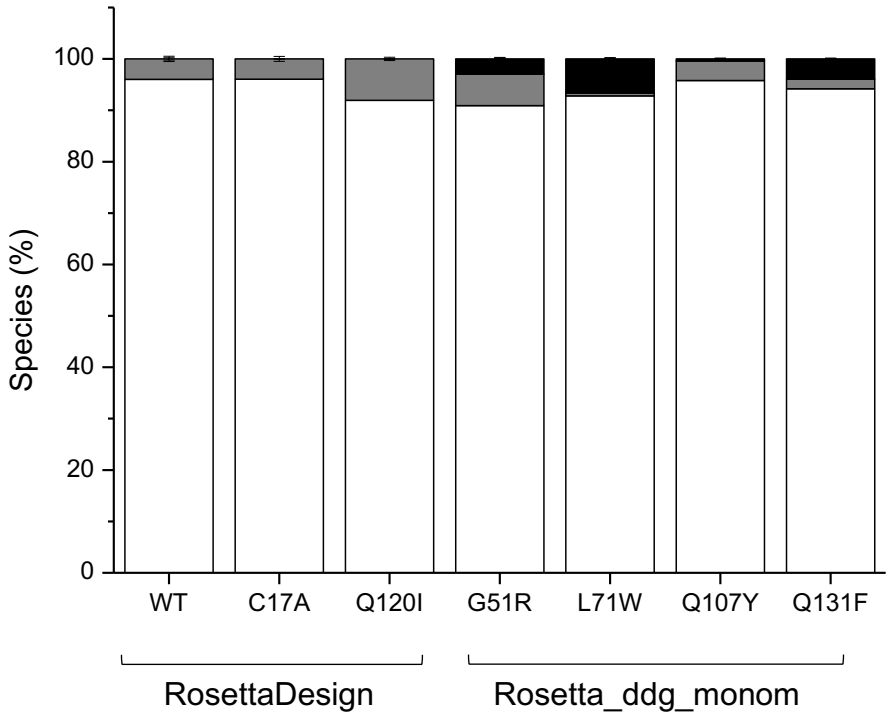


Figure S9. Initial sample content of WT GCSF and variant samples by SEC-HPLC. Samples were formulated at 0.2 mg/mL in 10 mM sodium acetate pH 4.25, frozen at -70 °C followed by defrosting for 20 mins prior to analysis. SEC-HPLC was performed using a TSK3000 SEC column with 0.1 M phosphate buffer pH 2.5 over 40 mins. Resolved species were aggregate (grey), dimer (black) and monomer (white).

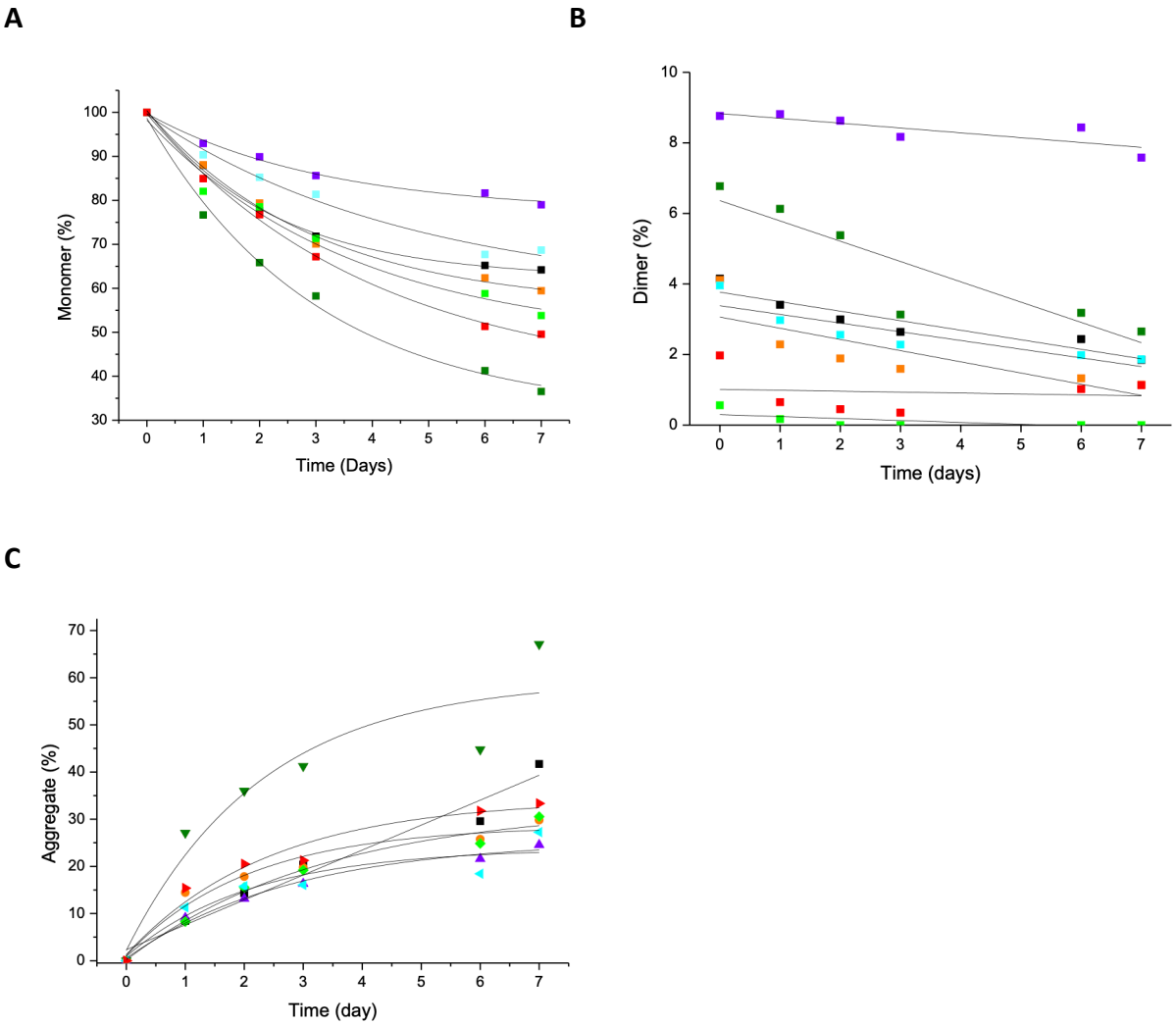
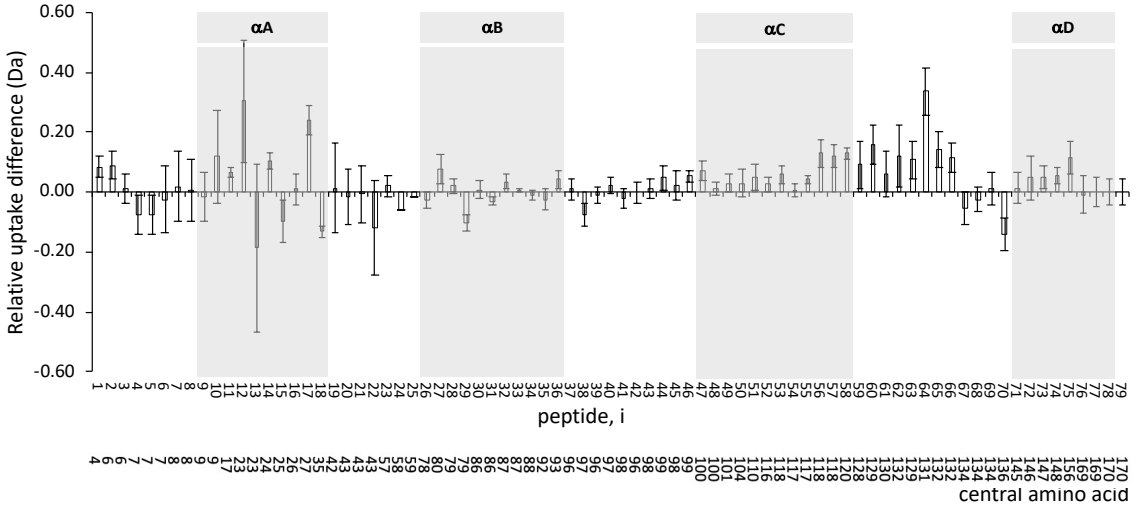
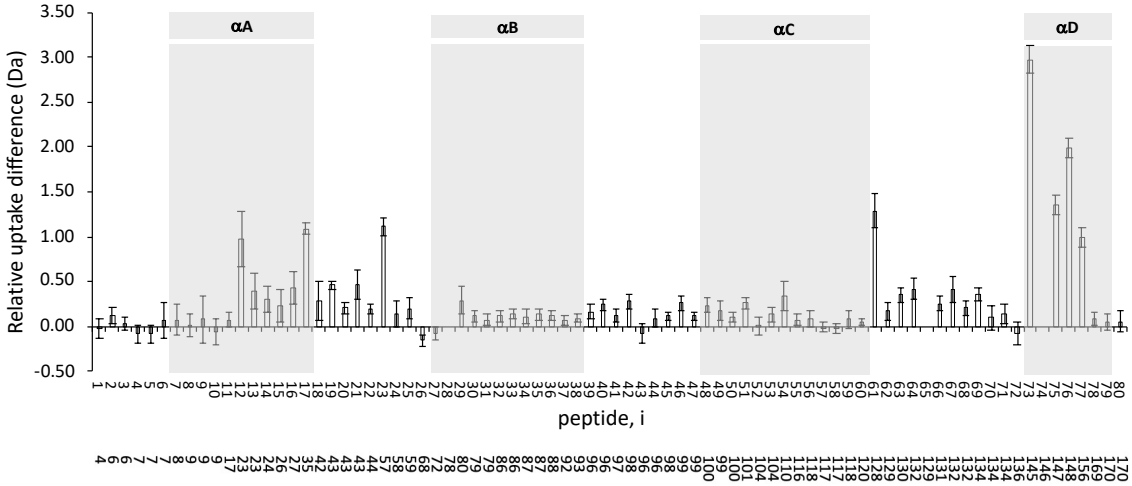


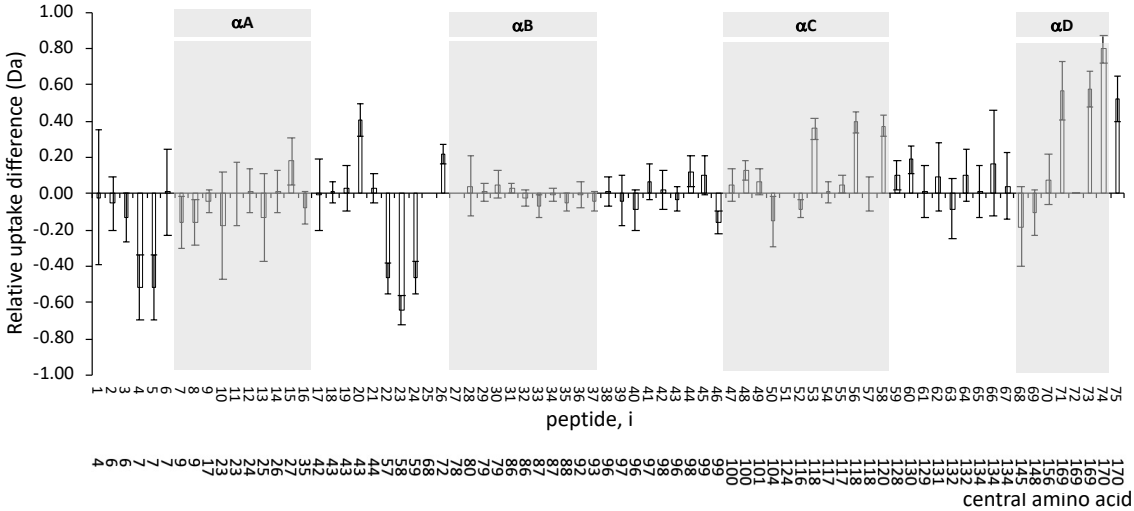
Figure S10. Degradation kinetics for G-CSF and variants at 45 °C incubation. Time-dependence of A) Monomer loss; B) Dimer loss; C) Aggregate formation, each as % of original starting material. Curves were fitted using the equations y = y0-Aexp(R0*x) or y = mx + c. Symbols are coloured by variant as (Black) WT; (Orange) C17A; (Purple) Q120I; (Dark green) G51R; (Light green) L71W; Cyan (Q107Y); (Red) Q131F.



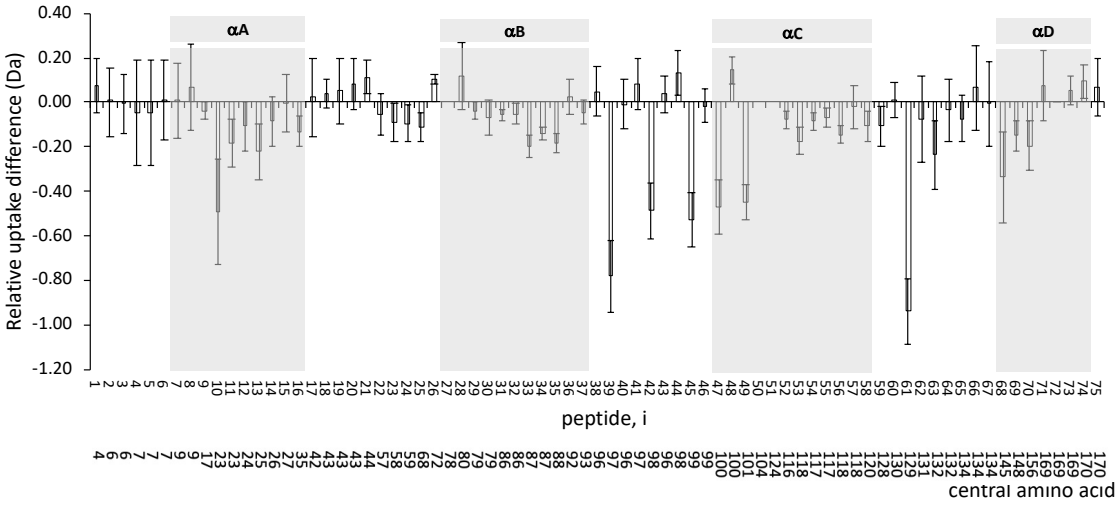
C17A

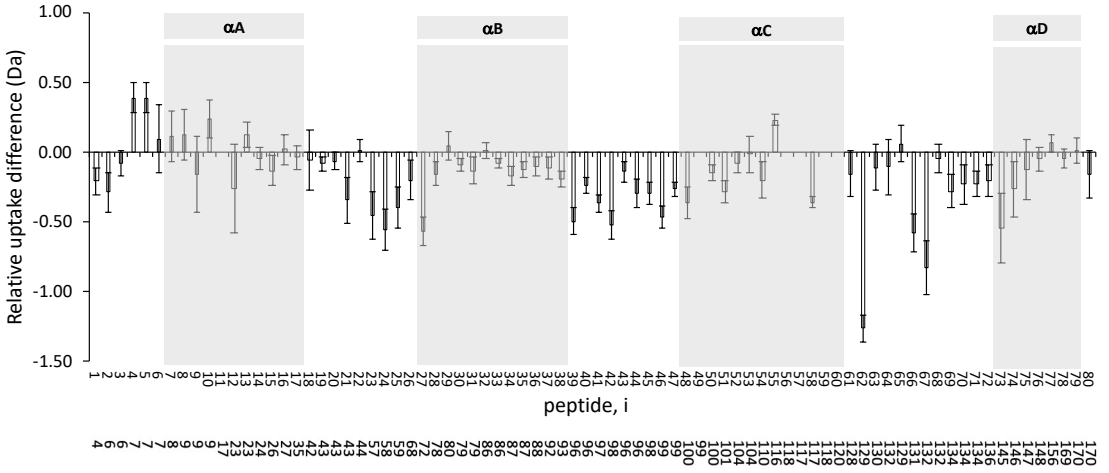


G51R



L71W





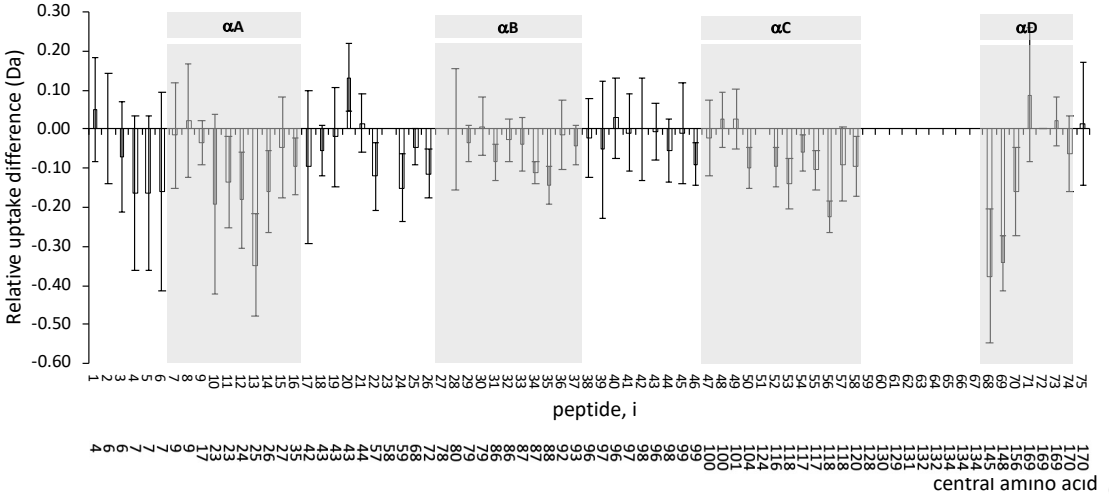
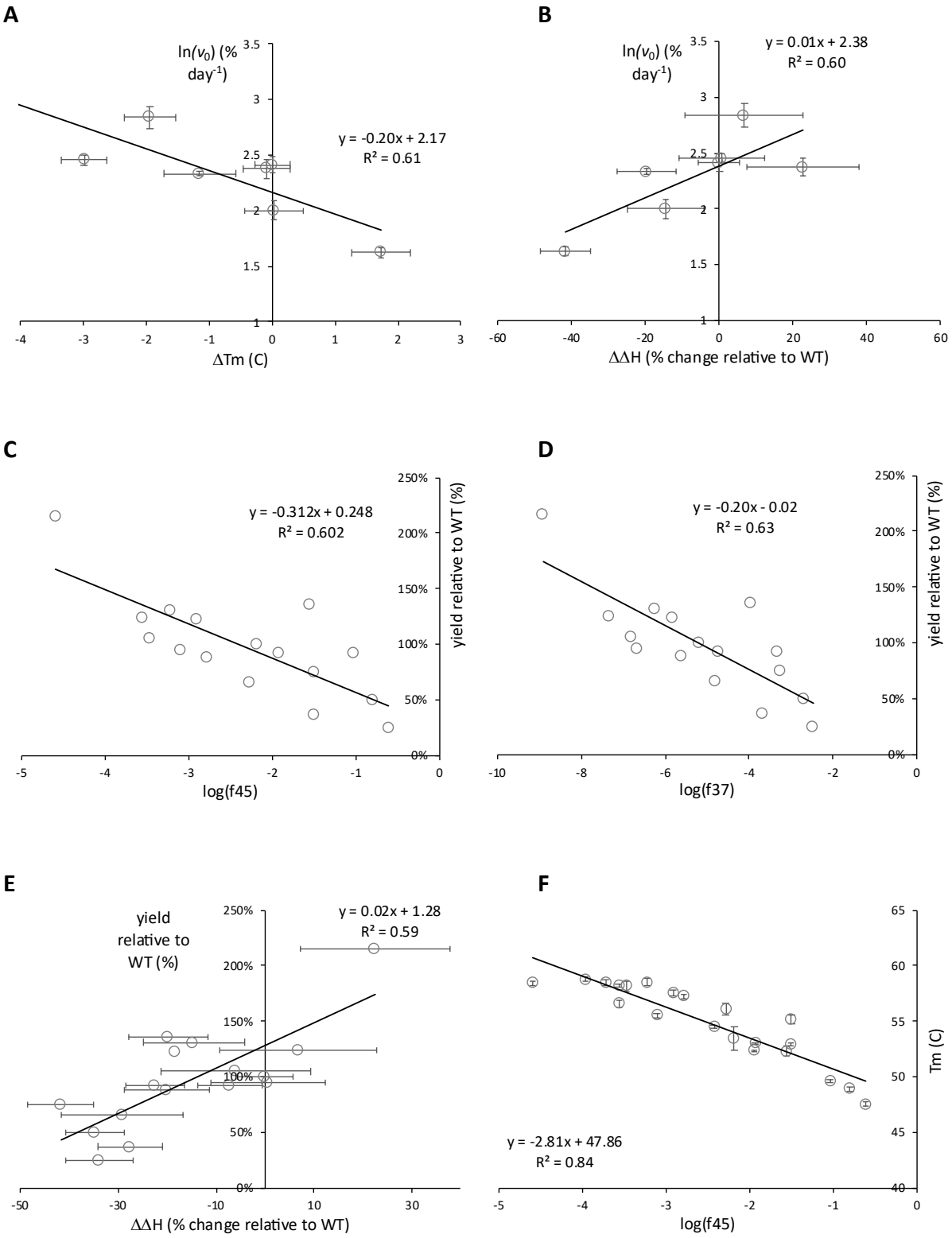


Figure S11. Peptide differential uptake plots for each G-CSF variant relative to WT. Individual HDX-MS time point relative uptake values were added together and differential calculated (variant uptake minus wild-type uptake). The *y*-axis denotes the relative uptake calculated from the change in mass between the undeuterated and deuterated values. The *x*-axis denotes the ordinal peptide number, a sequential arrangement of 80 peptides of GCSF by the midpoints in the sequence. The different helical regions of G-CSF are coloured in the background. The non-coloured regions represent the

connecting loop regions. The supplementary information (Table S2) contains the G-CSF peptide sequence, residue numbers and locations.


A
ln(v_0) (% day^{-1})
y = -0.20x + 2.17
R² = 0.61
ΔTm (C)
B
ln(v_0) (% day^{-1})
y = 0.01x + 2.38
R² = 0.60
ΔΔH (% change relative to WT)
C
y = -0.312x + 0.248
R² = 0.602
yield relative to WT (%)
log(f45)
D
y = -0.20x - 0.02
R² = 0.63
yield relative to WT (%)
log(f37)
E
yield relative to WT (%)
y = 0.02x + 1.28
R² = 0.59
ΔΔH (% change relative to WT)
F
y = -2.81x + 47.86
R² = 0.84
Tm (C)
log(f45)

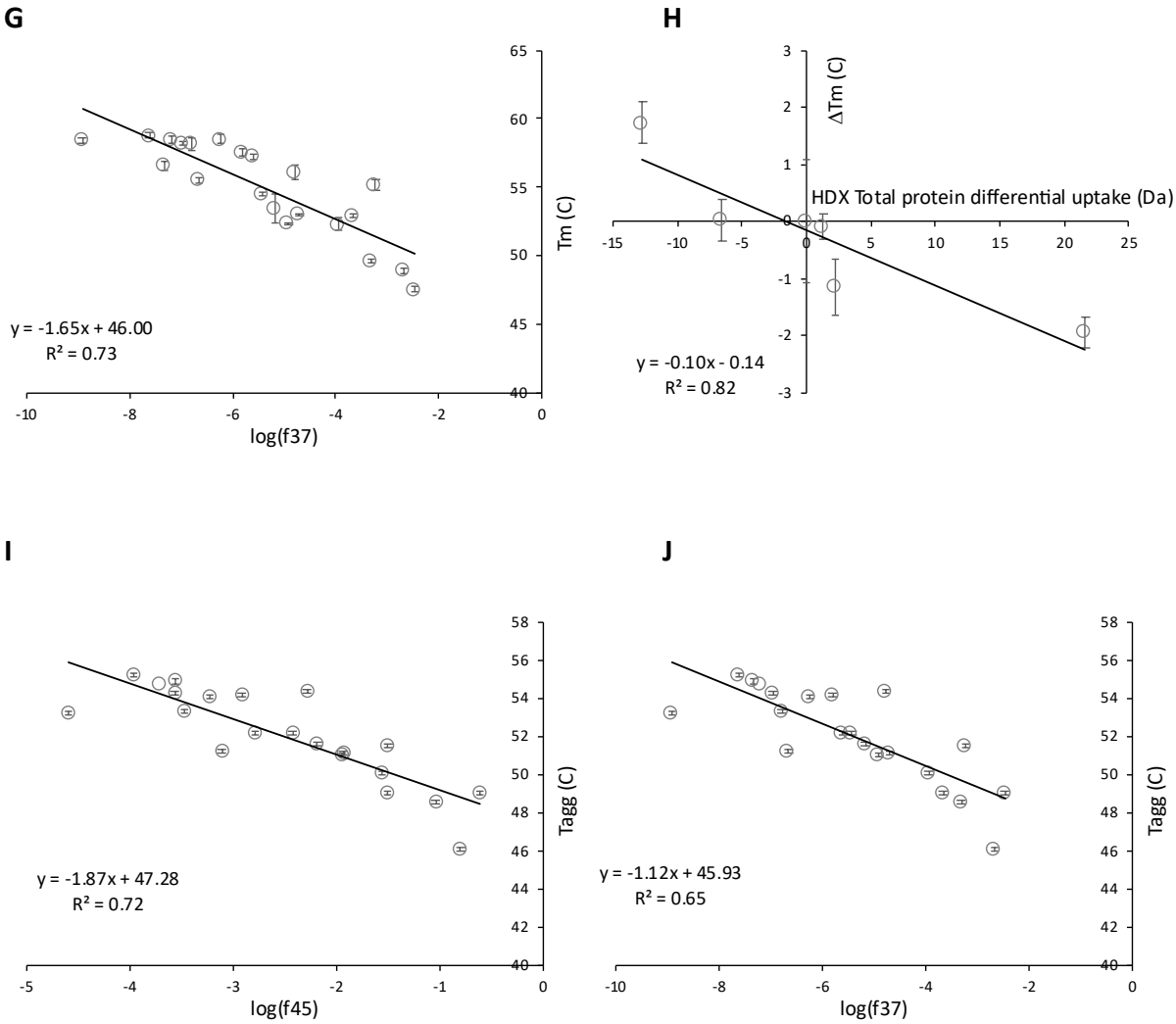


Figure S12. Significant correlations between various variant properties.
A) ln(v_0) vs T_m; B) ln(v_0) vs $\Delta\Delta H$; C) Expression yield vs log(f45); D) Expression yield vs log(f37); E) Expression yield vs $\Delta\Delta H$; F) T_m vs log(f45); G) T_m vs log(f37); H) T_m vs HDX total protein differential uptake; I) T_{agg} vs log(f45); J) T_{agg} vs log(f37)

Table S1. Peptide map summary for WT and variant samples

Sample	No. of peptides	Redundancy	Coverage (%)
WT	105	6.96	97.7
C17A	94	6.06	98.9
G51R	95	6.42	97.7
L71W	103	7.02	97.7
Q107Y	119	8.11	97.7
Q120I	90	5.74	97.7
Q131F	106	6.47	97.7

Table S2. G-CSF peptide sequences identified and used in HDX-MS measurements shown in Figure 3, for variant G-CSF in in 10mM sodium acetate pH 4.25.

A) Peptide information for G51R, Q120I and associated WT G-CSF used to calculate peptide differential uptake shown in Figure 3.

Peptide number	Peptide residue range	Sequence	Region
1	1-10	MTPLGPASSL	LoopA
2	1-13	MTPLGPASSLPQS	
3	1-14	MTPLGPASSLPQSF	
4	1-15	MTPLGPASSLPQSFL	
5	1-15	MTPLGPASSLPQSFL	
6	3-14	PLGPASSLPQSF	
7	3-15	PLGPASSLPQSFL	
8	5-15	GPASSLPQSFL	LoopA/αA
9	6-14	PASSLPQSF	
10	6-15	PASSLPQSFL	
11	15-21	LLKCLEQ	αA
12	16-32	LKCLEQVRKIQGDGAAL	
13	17-32	KCLEQVRKIQGDGAAL	
14	19-32	LEQVRKIQGDGAAL	
15	22-32	VRKIQGDGAAL	
16	22-34	VRKIQGDGAALQE	
17	33-39	QEKLCAT	αA/LoopAB
18	40-47	YKLCHPEE	LoopAB
19	40-48	YKLCHPEEL	
20	41-47	KLCHPEE	
21	41-48	KLCHPEEL	
22	42-48	LCHPEEL	
23	51-65	LGHSLGIPWAPLSSC	
24	51-68	LGHSLGIPWAPLSSCPSQ	
25	51-69	LGHSLGIPWAPLSSCPSQA	
26	66-72	PSQALQL	
27	70-76	LQLAGCL	LoopAB/αB
28	76-83	LSQLHSGL	αB
29	76-86	LSQLHSGLFLY	
30	77-83	SQLHSGL	
31	77-84	SQLHSGLF	
32	84-90	FLYQGLL	
33	84-91	FLYQGLLQ	
34	84-92	FLYQGLLQA	
35	85-92	LYQGLLQA	
36	86-92	YQGLLQA	
37	90-97	LQALEGIS	αB/LoopBC
38	91-97	QALEGIS	
39	91-104	QALEGISPELGPTL	LoopBC
40	93-101	LEGISPELG	
41	93-104	LEGISPELGPTL	

42	93-106	LEGISPELGPTLDT	
43	94-100	EGISPEL	
44	94-101	EGISPELG	
45	94-104	EGISPELGPTL	
46	94-106	EGISPELGPTLDT	LoopBC/αC
47	96-104	ISPELGPTL	
48	96-106	ISPELGPTLDT	
49	97-104	SPELGPTL	
50	98-104	PELGPTL	
51	98-106	PELGPTLDT	
52	100-110	LGPTLDTLQLD	αC
53	100-111	LGPTLDTLQLDV	
54	107-116	LQLDVADFAT	
55	114-120	FATTIWQ	
56	114-124	FATTIWQQMEE	
57	115-121	ATTIWQQ	
58	115-122	ATTIWQQM	
59	115-124	ATTIWQQMEE	
60	118-124	IWQQMEE	
61	121-137	QMEELGMAPALQPTQGA	LoopCD
62	123-138	EELGMAPALQPTQGAM	
63	125-138	LGMAPALQPTQGAM	
64	125-141	LGMAPALQPTQGAMPAF	
65	126-135	GMAPALQPTQ	
66	126-138	GMAPALQPTQGAM	
67	126-141	GMAPALQPTQGAMPAF	
68	129-138	PALQPTQGAM	
69	129-141	PALQPTQGAMPAF	
70	130-140	ALQPTQGAMPA	
71	132-138	QPTQGAM	
72	133-141	PTQGAMPAF	
73	139-153	PAFASAFQRRAGGVL	αD
74	141-153	FASAFQRRAGGVL	
75	143-153	SAFQRRAGGVL	
76	146-153	QRRAGGVL	
77	154-161	VASHLQSF	
78	167-174	RVLRHLAQ	αD/LoopD
79	167-175	RVLRHLAQP	
80	168-175	VLRHLAQP	

B) Peptide information for C17A and associated WT G-CSF used to calculate peptide differential uptake shown in Figure 3.

Peptide number	Peptide residue range	Sequence	Region
1	1-10	MTPLGPASSL	LoopA
2	1-13	MTPLGPASSLPQS	
3	1-14	MTPLGPASSLPQSF	

4	1-15	MTPLGPASSLPQSFL	
5	1-15	MTPLGPASSLPQSFL	
6	3-14	PLGPASSLPQSF	
7	5-15	GPASSLPQSFL	LoopA/αA
8	6-15	PASSLPQSFL	
9	15-21	LLKCLEQ	αA
10	16-32	LKCLEQVRKIQGDGAAL	
11	17-32	KCLEQVRKIQGDGAAL	
12	19-32	LEQVRKIQGDGAAL	
13	20-32	EQVRKIQGDGAAL	
14	22-32	VRKIQGDGAAL	
15	22-34	VRKIQGDGAALQE	
16	33-39	QEKLCAT	αA/LoopAB
17	40-47	YKLCHPEE	LoopAB
18	40-48	YKLCHPEEL	
19	41-47	KLCHPEE	
20	41-48	KLCHPEEL	
21	42-48	LCHPEEL	
22	51-65	LGHSLGIPWAPLSSC	
23	51-68	LGHSLGIPWAPLSSCPSQ	
24	51-69	LGHSLGIPWAPLSSCPSQA	
25	66-72	PSQALQL	
26	70-76	LQLAGCL	LoopAB/αB
27	76-83	LSQLHSGL	αB
28	76-86	LSQLHSGLFLY	
29	77-83	SQLHSGL	
30	77-84	SQLHSGLF	
31	84-90	FLYQGLL	
32	84-91	FLYQGLLQ	
33	84-92	FLYQGLLQA	
34	85-92	LYQGLLQA	
35	86-92	YQGLLQA	
36	90-97	LQALEGIS	αB/LoopBC
37	91-97	QALEGIS	
38	91-104	QALEGISPELGPTL	LoopBC
39	91-106	QALEGISPELGPTLDT	
40	93-101	LEGISPELG	
41	93-104	LEGISPELGPTL	
42	93-106	LEGISPELGPTLDT	
43	94-101	EGISPELG	
44	94-104	EGISPELGPTL	
45	94-106	EGISPELGPTLDT	LoopBC/αC
46	96-104	ISPELGPTL	
47	96-106	ISPELGPTLDT	
48	98-104	PELGPTL	
49	98-106	PELGPTLDT	

50	100-110	LGPTLDTLQLD	αC
51	109-141	LDVADFATTIWQQMEELGMAPALQPTQGAMPAF	
52	114-120	FATTIWQ	
53	114-124	FATTIWQQMEE	
54	115-121	ATTIWQQ	
55	115-122	ATTIWQQM	
56	115-124	ATTIWQQMEE	
57	116-122	TTIWQQM	
58	118-124	IWQQMEE	
59	121-137	QMEELGMAPALQPTQGA	LoopCD
60	125-138	LGMAPALQPTQGAM	
61	126-135	GMAPALQPTQ	
62	126-138	GMAPALQPTQGAM	
63	126-141	GMAPALQPTQGAMPAF	
64	129-138	PALQPTQGAM	
65	129-141	PALQPTQGAMPAF	
66	130-140	ALQPTQGAMPA	
67	132-138	QPTQGAM	
68	139-153	PAFASAFQRRAGGVL	αD
69	146-153	QRRAGGVL	
70	154-161	VASHLQSF	
71	166-175	YRVLRHLAQP	αD/LoopD
72	167-174	RVLRHLAQ	
73	167-175	RVLRHLAQP	
74	168-175	VLRHLAQP	

C) Peptide information for L71W, Q107Y and Q131F and associated WT G-CSF used to calculate peptide differential uptake shown in Figure 3.

Peptide number	Peptide residue range	Sequence	Region
1	1-10	MTPLGPASSL	LoopA
2	1-13	MTPLGPASSLPQS	
3	1-14	MTPLGPASSLPQSF	
4	1-15	MTPLGPASSLPQSFL	
5	1-15	MTPLGPASSLPQSFL	
6	3-14	PLGPASSLPQSF	
7	5-15	GPASSLPQSFL	LoopA/αA
8	6-15	PASSLPQSFL	
9	15-21	LLKCLEQ	αA
10	16-32	LKCLEQVRKIQGDGAAL	
11	17-32	KCLEQVRKIQGDGAAL	
12	19-32	LEQVRKIQGDGAAL	
13	20-32	EQVRKIQGDGAAL	
14	22-32	VRKIQGDGAAL	
15	22-34	VRKIQGDGAALQE	
16	33-39	QEKLCAT	αA/LoopAB

17	40-47	YKLCHPEE	LoopAB
18	40-48	YKLCHPEEL	
19	41-47	KLCHPEE	
20	41-48	KLCHPEEL	
21	42-48	LCHPEEL	
22	51-65	LGHSLGIPWAPLSSC	
23	51-68	LGHSLGIPWAPLSSCPSQ	
24	51-69	LGHSLGIPWAPLSSCPSQA	
25	66-72	PSQALQL	
26	70-76	LQLAGCL	LoopAB/αB
27	76-86	LSQLHSGLFLY	αB
28	77-83	SQLHSGL	
29	77-84	SQLHSGLF	
30	84-90	FLYQGLL	
31	84-91	FLYQGLLQ	
32	84-92	FLYQGLLQA	
33	85-92	LYQGLLQA	
34	86-92	YQGLLQA	
35	90-97	LQALEGIS	αB/LoopBC
36	91-97	QALEGIS	
37	91-104	QALEGISPELGPTL	LoopBC
38	91-106	QALEGISPELGPTLDT	
39	93-101	LEGISPELG	
40	93-104	LEGISPELGPTL	
41	93-106	LEGISPELGPTLDT	
42	94-101	EGISPELG	
43	94-104	EGISPELGPTL	
44	94-106	EGISPELGPTLDT	LoopBC/αC
45	96-104	ISPELGPTL	
46	96-106	ISPELGPTLDT	
47	98-104	PELGPTL	
48	98-106	PELGPTLDT	
49	100-110	LGPTLDTLQLD	αC
50	114-120	FATTIWQ	
51	114-124	FATTIWQQMEE	
52	115-121	ATTIWQQ	
53	115-122	ATTIWQQM	
54	115-124	ATTIWQQMEE	
55	116-122	TTIWQQM	
56	118-124	IWQQMEE	
57	121-137	QMEELGMAPALQPTQGA	LoopCD
58	125-138	LGMAPALQPTQGAM	
59	126-135	GMAPALQPTQ	
60	126-138	GMAPALQPTQGAM	
61	126-141	GMAPALQPTQGAMPAF	
62	129-138	PALQPTQGAM	

63	129-141	PALQPTQGAMPAF	
64	130-140	ALQPTQGAMPA	
65	132-138	QPTQGAM	
66	139-153	PAFASAFQRRAGGVL	αD
67	146-153	QRRAGGVL	
68	154-161	VASHLQSF	
69	166-175	YRVLRHLAQP	αD/LoopD
70	167-174	RVLRHLAQ	
71	167-175	RVLRHLAQP	
72	168-175	VLRHLAQP	

References

Wood VE, Groves K, Cryar A, Quaglia M, Matejtschuk P, Dalby PA (2020) HDX and in-silico docking reveal that excipients stabilise G-CSF via a combination of preferential exclusion and specific hotspot interactions. *Molecular Pharmaceutics* 17 (12) 4637-4651.

Tamada, T.; Honjo, E.; Maeda, Y.; Okamoto, T.; Ishibashi, M.; Tokunaga, M.; Kuroki, R. Homodimeric cross-over structure of the human granulocyte colony-stimulating factor (GCSF) receptor signaling complex. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 3135-3140.