Supporting Information for Lawrence, Phillips, and Liu

Materials and Methods

Construction of supercharged GFPs. Initial designs of supercharged proteins were created partially by manual inspection of the structure of GFP. Following the encouraging results with supercharged GFPs, the design algorithm was reformulated as a completely automated method and used to create the supercharged versions of GST and streptavidin.

We designed a GFP with net charge of +36 by mutating residues of the starting GFP that were likely to tolerate substitution to a positively charged amino acid (Lys or Arg). We chose residues seen to be highly solvent-exposed, based on manual inspection of the crystal structure.[1] Preference was given to negatively charged amino acids (Asp and Glu), not only because a basic side-chain is likely to have the same extreme hydrophilicity as an acidic one (thus decreasing the likelihood of a structural disruption), but also because these mutations achieved a double increment in charge. Additional salient surface-exposed positions to mutate were chosen on the basis of sequence variability observed at these positions among GFP homologues. In total, 29 residues were mutated to positively charged amino acids.

The gene encoding the GFP(+36) was constructed using overlapping synthetic oligonucleotides.[2] Genes encoding stGFP and GFP(+36), optimized for *E. coli* codon usage, were constructed using a series of 38 overlapping DNA oligonucleotides, each ~40 nt in length and annealing to each other over ~20-bp stretches. The oligonucleotides were mixed together in equimolar ratio, then 5'-phosphorylated using T4 polynucleotide kinase and ligated using T4 DNA ligase. Full-length genes were amplified by PCR using the outermost oligonucleotides as primers. The initial design for GFP(+36) was successful, yielding green fluorescent bacteria.

We also designed a supernegative GFP by identifying 20 highly solvent-exposed positions and mutating these to negatively charged amino acids (Asp and Glu), yielding an initial design with theoretical net charge of –39. However, expression of a gene for GFP(–39) yielded non-fluorescent bacteria, indicating that one or more of the mutations disrupted GFP expression or function. To avoid mutations responsible for the disruption of expression or fluorescence, we repeated the GFP(–39) gene construction process in the presence of 5 mol% stGFP oligonucleotides. Two of the resulting fluorescent colonies were characterized further: GFP(–30), which had 15 of the 20 planned mutations, and GFP(–25), which had 12.

As a test of how much further the net charge of GFP(+36) could be increased, we designed oligos for a potential GFP(+73) and mixed these together with the GFP(+36) oligos in various ratios, repeated the gene construction procedure and picked green colonies, thereby isolating GFP(+48), with a total of 36 mutations.

Table S2 is a list of the amino acid residues of stGFP, listed in order of solvent exposure, computed as Average # of Neighboring Atoms (within 10 Å) Per Sidechain Atom ("AvNAPSA"). Also shown are the sequences of GFP(–25), GFP(–30), GFP(+36), and GFP(+48). The table also indicates all the positions that were allowed to vary in the designs for GFP(–39) and GFP(+73).

Construction of supercharged GSTs and streptavidins. The sequences of the other proteins studied in this work are shown below. Solvent-exposed residues (shown in grey) are those having AvNAPSA < 150, computed from crystallographic data.[3, 4] Charged or highly polar solvent-exposed residues (DERKNQ) were mutated either to Asp or Glu, for negative-supercharging (red); or to Lys or Arg, for positive-supercharging (blue). The supercharging design process for streptavidin (SAV) and glutathione-*S*-transferase (GST) was fully automated: residues were first sorted by solvent exposure (see Tables S3 and S4), and then the most solvent-exposed charged or highly polar residues were mutated either to Lys for positive supercharging, or to Glu (unless the starting residue was Asn, in which case to Asp) for negative supercharging. Synthetic genes for GST and streptavidin variants were purchased from DNA 2.0.

```
                1
SAV(-40)        MGHHHHHHGGAEAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGDAESRYVLTGRYDSAPATDGSGTA
wtSAV           ---------AAEAGITGTWYNQLGSTFIVTAGADGALTGTYESAVGNAESRYVLTGRYDSAPATDGSGTA
SAV(-52)        MGHHHHHHGGAKAGITGTWYNQLGSTFIVTAGAKGALTGTYESAVGNAKSRYVLTGRYDSAPATKGSGTA

               71
SAV(-40)        LGWTVAWKNDYENAHSATTWSGQYVGGAEARINTQWLLTSGTTEADAWKSTLVGHDTFTKVEPSAAS
wtSAV           LGWTVAWKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAWKSTLVGHDTFTKVKPSAAS
SAV(-52)        LGWTVAWKNKYRNAHSATTWSGQYVGGAKARINTQWLLTSGTTKAKAWKSTLVGHDTFTKVKPSAAS


                1
GST(-40)        MGHHHHHHGGPPYTITYFPVRGRCEAMRMLLADQDQSWEEEVVTMETWPPLKPSCLFRQLPKFQDGDLTLYQSNA
wtGST           MGHHHHHHGGPPYTITYFPVRGRCEAMRMLLADQDQSWKEEVVTMETWPPLKPSCLFRQLPKFQDGDLTLYQSNA
GST(+50)        MGHHHHHHGGPPYTITYFPVRGRCEAMRMLLADQKQSWKEEVVTMKTWPPLKPSCLFRQLPKFQDGKLTLYQSNA

               75
GST(-40)        ILRHLGRSFGLYGEDEEEAALVDMVNDGVEDLRCKYATLIYTDYEAGKEEYVEELPEHLKPFETLLSENEGGEAF
wtGST           ILRHLGRSFGLYGKDQKEAALVDMVNDGVEDLRCKYATLIYTNYEAGKEKYVKELPEHLKPFETLLSQNQGGQAF
GST(+50)        ILRHLGRSFGLYGKKQKEAALVDMVNDGVEDLRCKYATLIYTKYKAGKKKYVKKLPKHLKPFETLLSKNKGGKAF

              151
GST(-40)        VVGSEISFADYNLLDLLRIHQVLNPSCLDAFPLLSAYVARLSARPEIEAFLASPEHVDRPINGNGKQ
```

```
wtGST         VVGSQISFADYNLLDLLRIHQVLNPSCLDAFPLLSAYVARLSARPKIKAFLASPEHVNRPINGNGKQ
GST(+50)      VVGSKISFADYNLLDLLRIHQVLNPSCLKAFPLLSAYVARLSARPKIKAFLASPEHVKRPINGNGKQ
```

Protein expression and purification. Genes were cloned into a pET expression vector (Novagen), and overexpressed in *E. coli* BL21(DE3)pLysS for 5–10 hours at 15 °C. Cells were harvested by centrifugation and resuspended in lysis buffer containing 10 mM Tris pH 7.5 and 1 M NaCl, then lysed by sonication. Proteins were purified by Ni-NTA agarose chromotography (Qiagen), buffer-exchanged into 100 mM NaCl, 50 mM potassium phosphate pH 7.5, and concentrated by ultrafiltration (Millipore). All GFP variants were purified under native conditions. Wild-type streptavidin was purchased from Promega. Supercharged streptavidin variants were purified under denaturing conditions and refolded as reported previously for wild-type streptavidin,[5] as was supercharged GST (Fig. S5). Wild-type GST was purified under either native or denaturing conditions, yielding protein of comparable activity.

Protein staining and UV-induced fluorescence (Fig. 2a). 0.2 μg of each GFP variant was analyzed by electrophoresis in a 10% denaturing polyacrylamide gel and stained with Coomassie brilliant blue dye. 0.2 μg of the same protein samples in 25 mM Tris pH 8.0 with 100 mM NaCl was placed in a 0.2 mL Eppendorf tube and photographed under UV light (360 nm).

Isoelectric focusing (Fig. S6). IEF gels with a focusing range of pI = 3–10 (Novex, Invitrogen) were run according to the manufacturer's instructions.

Electrostatic surface potential calculations (Fig. 1b). Models of GFP variants were based on the crystal structure of superfolder GFP.[1] Electrostatic potentials were calculated using APBS[6] and rendered with PyMol[7] using a scale of –25 kT/e (red) to +25kT/e (blue).

Thermal denaturation and aggregation (Fig. 3a). Purified GFP variants were diluted to 2 mg/mL in 25 mM Tris pH 8.0, 100 mM NaCl, and 10 mM beta-mercaptoethanol (BME), then photographed under UV illumination ("native"). The samples were heated to 100 °C for 1 minute, then photographed again under UV illumination ("boiled"). Finally, the samples were cooled 2 h at room temperature and photographed again under UV illumination ("cooled").

In order to verify that the thermally induced aggregation of stGFP was not due to the particular buffer conditions, we repeated the experiment with stGFP, using a buffer that included 5 mM dithiothreitol (DTT), and 10% glycerol. Figure S2 shows that stGFP still aggregated and failed to recover fluorescence under these conditions, despite the inclusion of these bufferadditives, over a range of protein concentrations tested.

Chemically induced aggregation (Fig. 3b). 2,2,2-trifluoroethanol (TFE) was added to produce solutions with 1.5 mg/mL protein, 25 mM Tris pH 7.0, 10 mM BME, and 40% TFE. Aggregation at 25 °C was monitored by right-angle light scattering.

Size-exclusion chromatography (Figs. S8, S10, and Table S1). The multimeric state of SAV and GST variants was determined by analyzing 20–50 μg of protein on a Superdex 75 gel-filtration column. Buffer was 100 mM NaCl, 50 mM potassium phosphate pH 7.5. Molecular weights were determined by comparison with a set of monomeric protein standards of known molecular weights analyzed separately under identical conditions.

References Cited

1. Pedelacq, J.D., Cabantous, S., Tran, T., Terwilliger, T.C. & Waldo, G.S. Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol* **24**, 79-88 (2006).
2. Stemmer, W.P., Crameri, A., Ha, K.D., Brennan, T.M. & Heyneker, H.L. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164**, 49-53 (1995).
3. Dirr, H., Reinemer, P. & Huber, R. Refined crystal structure of porcine class Pi glutathione S-transferase (pGST P1-1) at 2.1 A resolution. *J Mol Biol* **243**, 72-92 (1994).
4. Weber, P.C., Ohlendorf, D.H., Wendoloski, J.J. & Salemme, F.R. Structural origins of high-affinity biotin binding to streptavidin. *Science* **243**, 85-88 (1989).
5. Thompson, L.D. & Weber, P.C. Construction and expression of a synthetic streptavidin-encoding gene in Escherichia coli. *Gene* **136**, 243-246 (1993).
6. Baker, N.A., Sept, D., Joseph, S., Holst, M.J. & McCammon, J.A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**, 10037-10041 (2001).
7. Delano, W.L., The PyMOL Molecular Graphics System, http://www.pymol.org 2002).
8. Ali-Osman, F., Akande, O., Antoun, G., Mao, J.X. & Buolamwini, J. Molecular cloning, characterization, and expression in Escherichia coli of full-length cDNAs of three human glutathione S-transferase Pi gene variants. Evidence for differential catalytic activity of the encoded proteins. *J Biol Chem* **272**, 10004-10012 (1997).
9. Kada, G., Kaiser, K., Falk, H. & Gruber, H.J. Rapid estimation of avidin and streptavidin by fluorescence quenching or fluorescence polarization. *Biochim Biophys Acta* **1427**, 44-48 (1999).

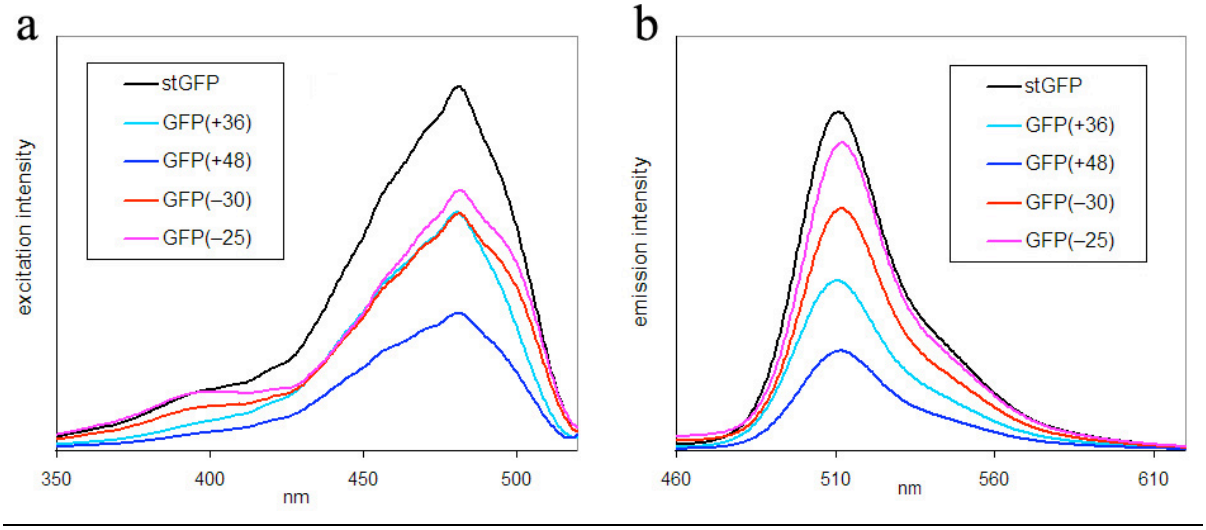

Figure S1. **(a)** Excitation and **(b)** emission spectra of GFP variants. Each sample contained an equal amount of protein as quantitated by chromophore absorbance at 490 nm.

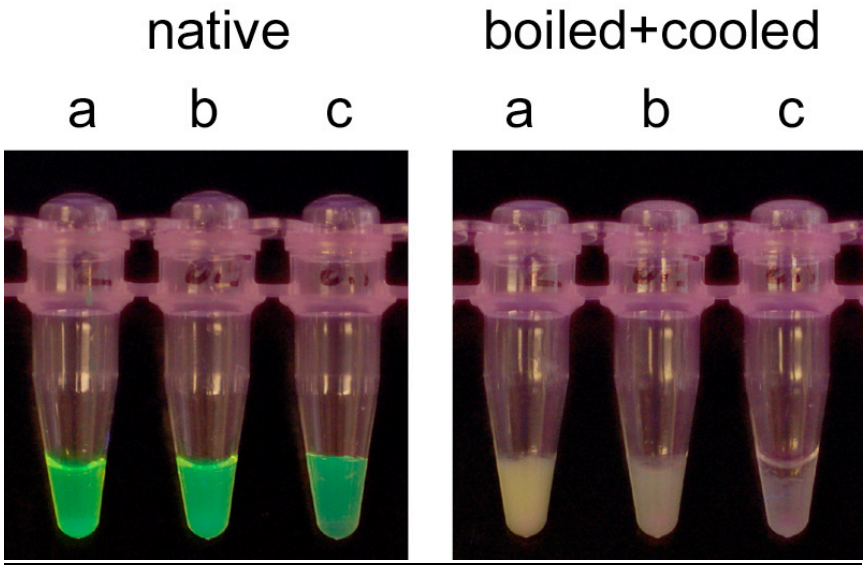

Figure S2. Thermally induced aggregation of stGFP under alternative conditions. The starting GFP (stGFP) was mixed to final concentrations of 2.0 (a), 0.5 (b), or 0.1 (c) mg/mL in a buffer containing 10 mM Tris pH 7.5, 5 mM NaCl, 5 mM dithiothreitol (DTT), and 10% glycerol. The left panel shows the samples before heating, under UV illumination. Samples were heated for 1 minute at 100°C, then cooled to room temperature for 2 hours and photographed again under UV illumination.

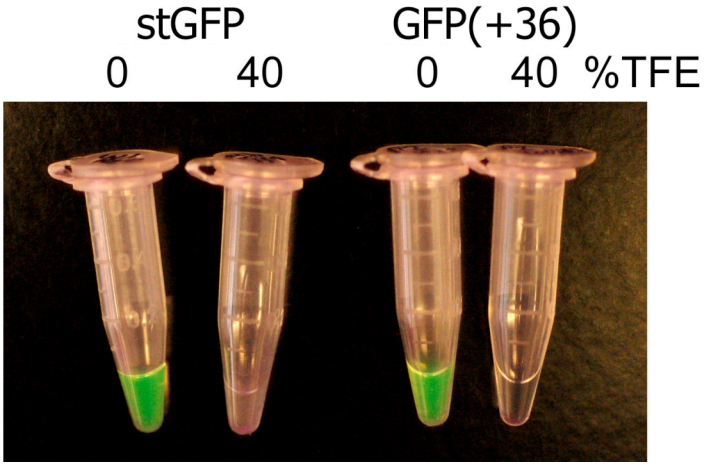

Figure S3. stGFP and GFP(+36) lose fluorescence when exposed to 40% 2,2,2-trifluoroethanol (TFE). Furthermore, the stGFP solution forms a white precipitate, while the GFP(+36) solution remains clear.

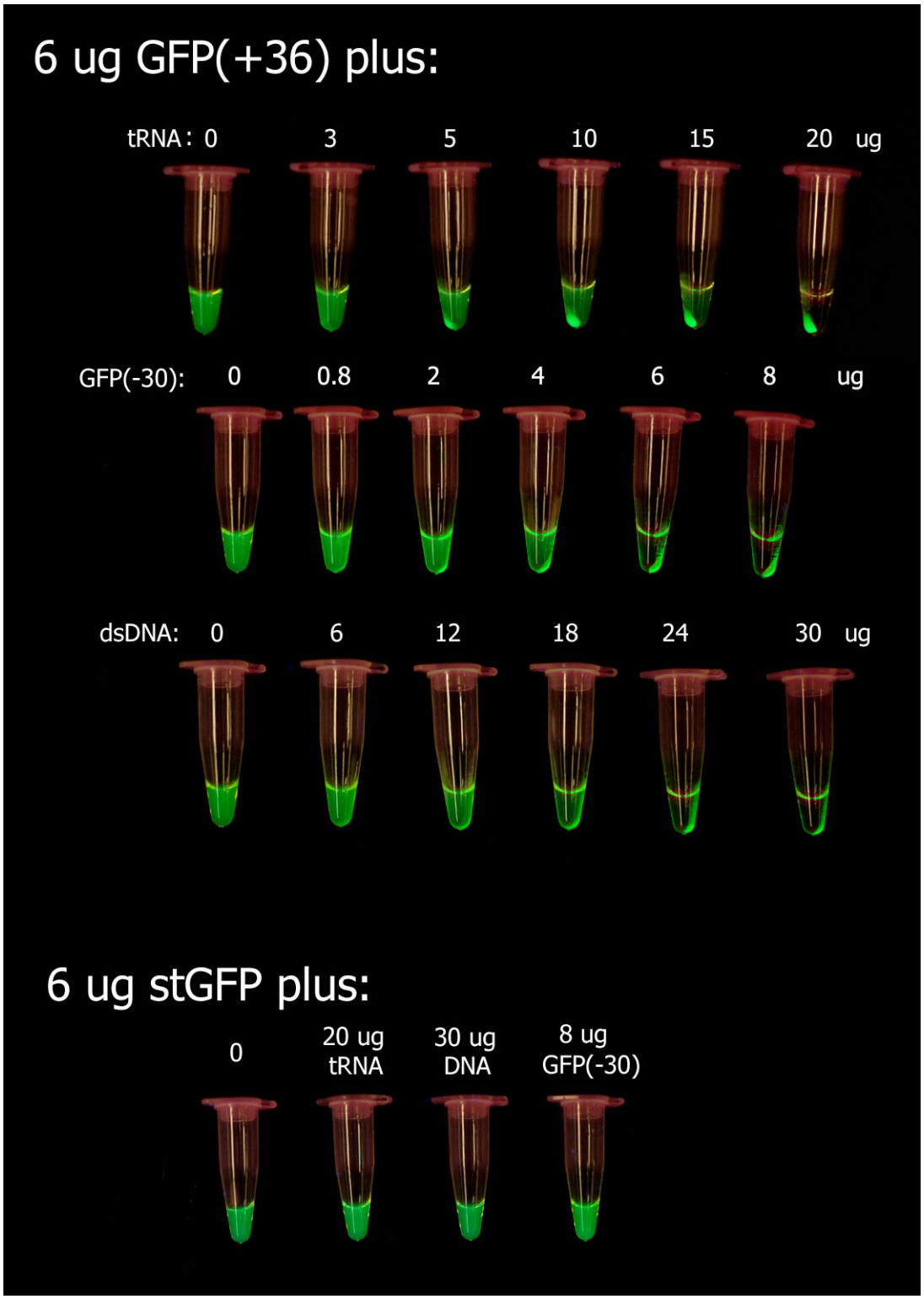

Figure S4. "Molecular velcro" effect demonstrated by titration of macromolecular polycounterions into solutions of supercharged GFPs.

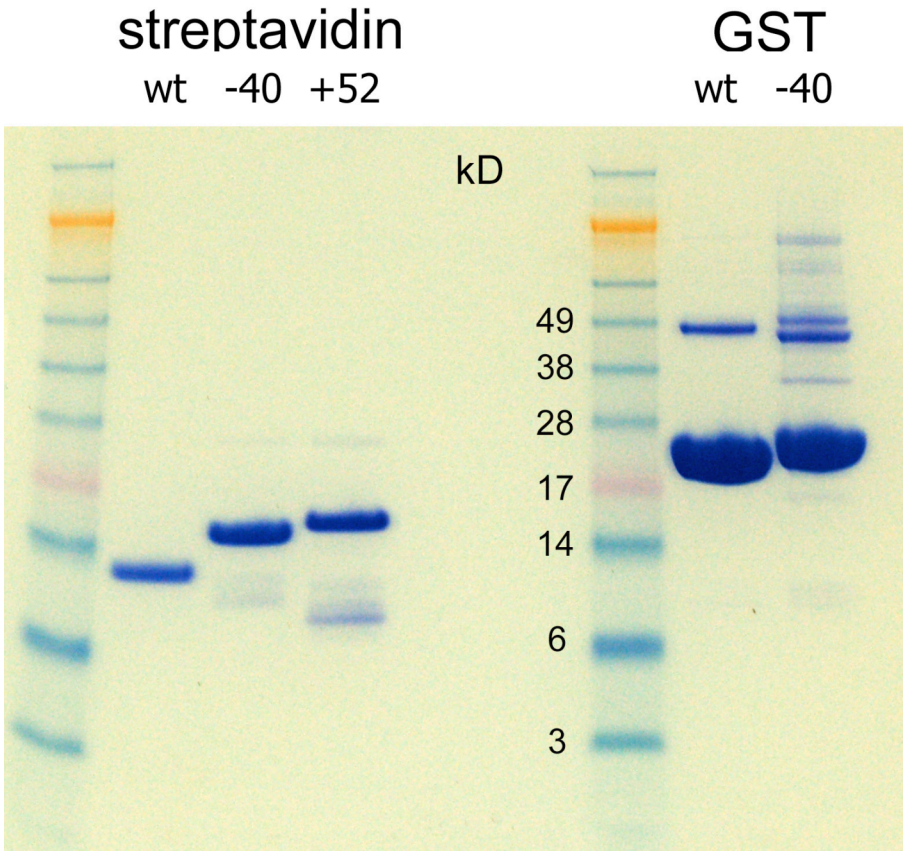

Figure S5. Purified streptavidin and GST variants examined by PAGE and Coomassie staining.

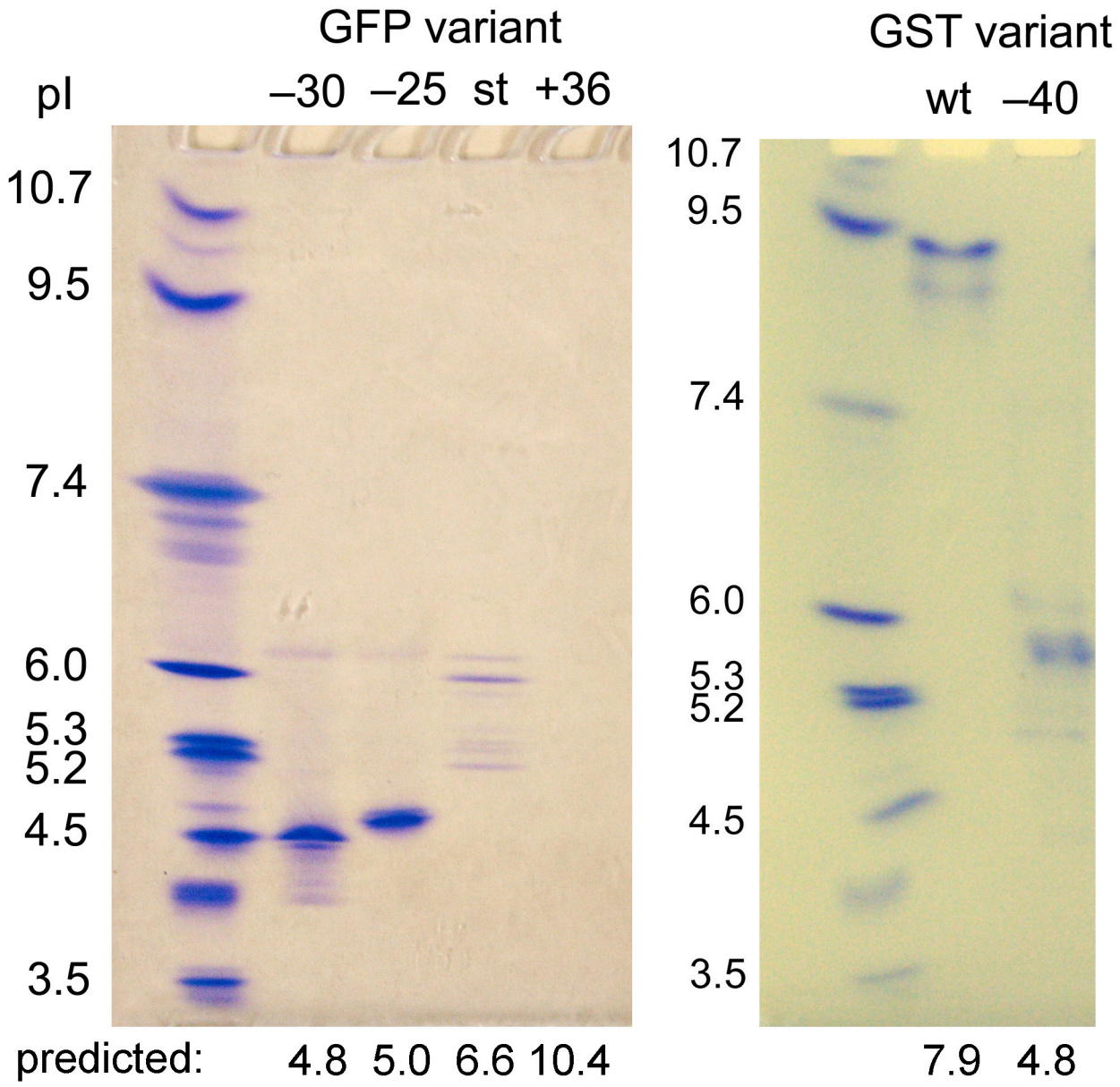

Figure S6. Isoelectric focusing (IEF) gels measuring the apparent isoelectric point (pI) of GFP and GST variants. Predicted pIs are shown at bottom.

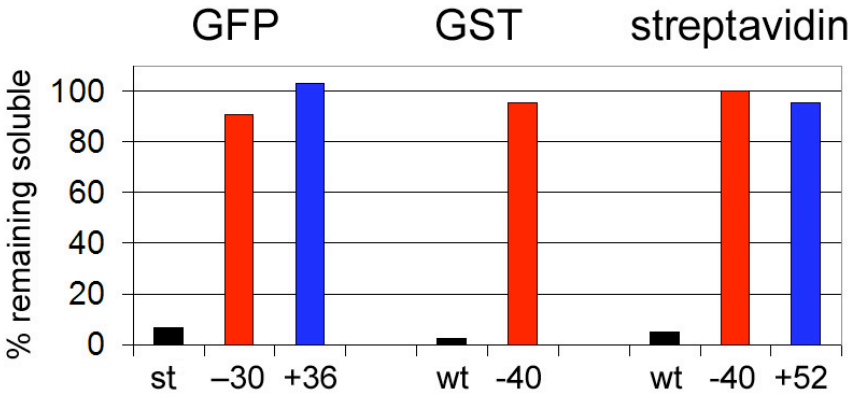

Figure S7. Resistance of supercharged proteins to thermally induced aggregation. Proteins were subjected to the boil-and-cool procedure demonstrated in Fig. 3a, by heating for 1 minute to 100°C and cooling for 2 hours at room temperature. Amount of protein remaining soluble was quantitated by brief centrifugation (1 min at 14,000 *G*) and remeasurement of the absorbance at 280 nm.

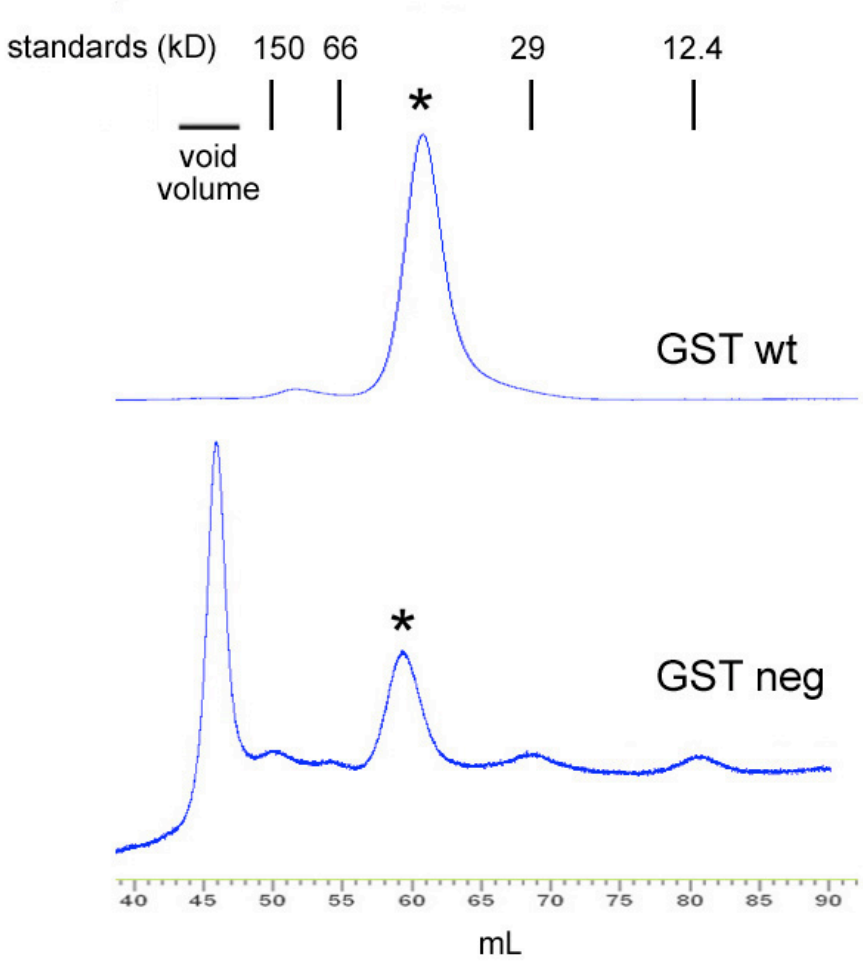

Figure S8. Analytical size-exclusion chromatography of GST variants. 50 μg of protein was injected onto a Superdex 75 gel-filtration column. Buffer was 100 mM NaCl, 50 mM potassium phosphate pH 7.5. Molecular weights were determined by comparison with a set of monomeric protein standards of known molecular weights analyzed separately under identical conditions.

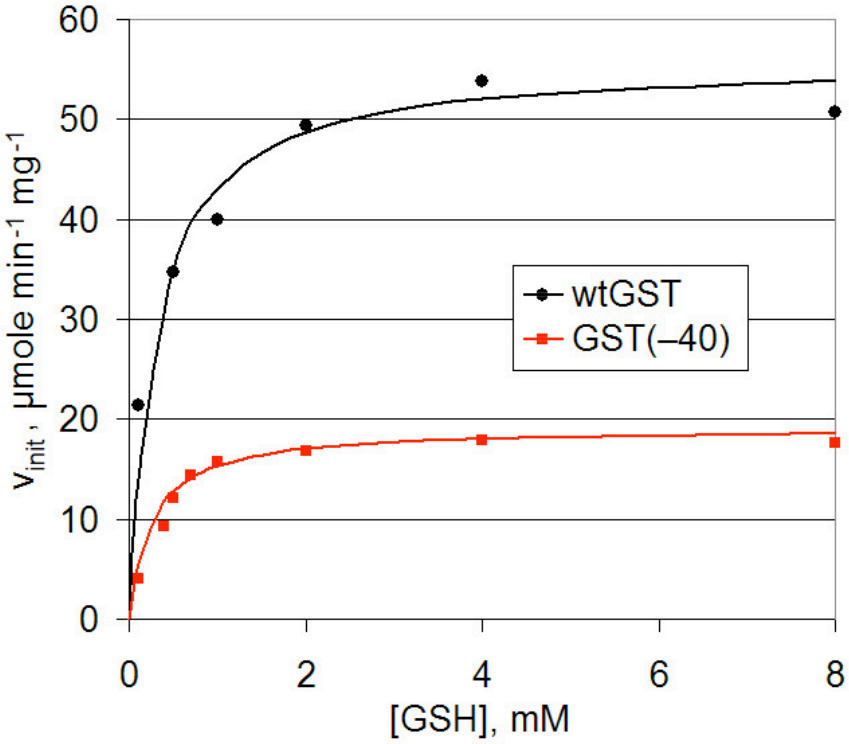

Figure S9. Determination of kinetic parameters for glutathione S-transferase (GST) variants.[8] Reactions contained 0.1 mM to 8 mM glutathione (GSH), as shown, plus 1 mM chlorodinitrobenzene (CDNB), 100 mM potassium phosphate buffer pH 6.5, and 0.01 mg/mL GST variant. Timepoints were removed and quenched by the addition of 0.2 M acetic acid, and reaction progress was measured by following product absorbance at 340 nm. Initial rates (v_{init}) were determined by measuring the maximum slope of the reaction curve.

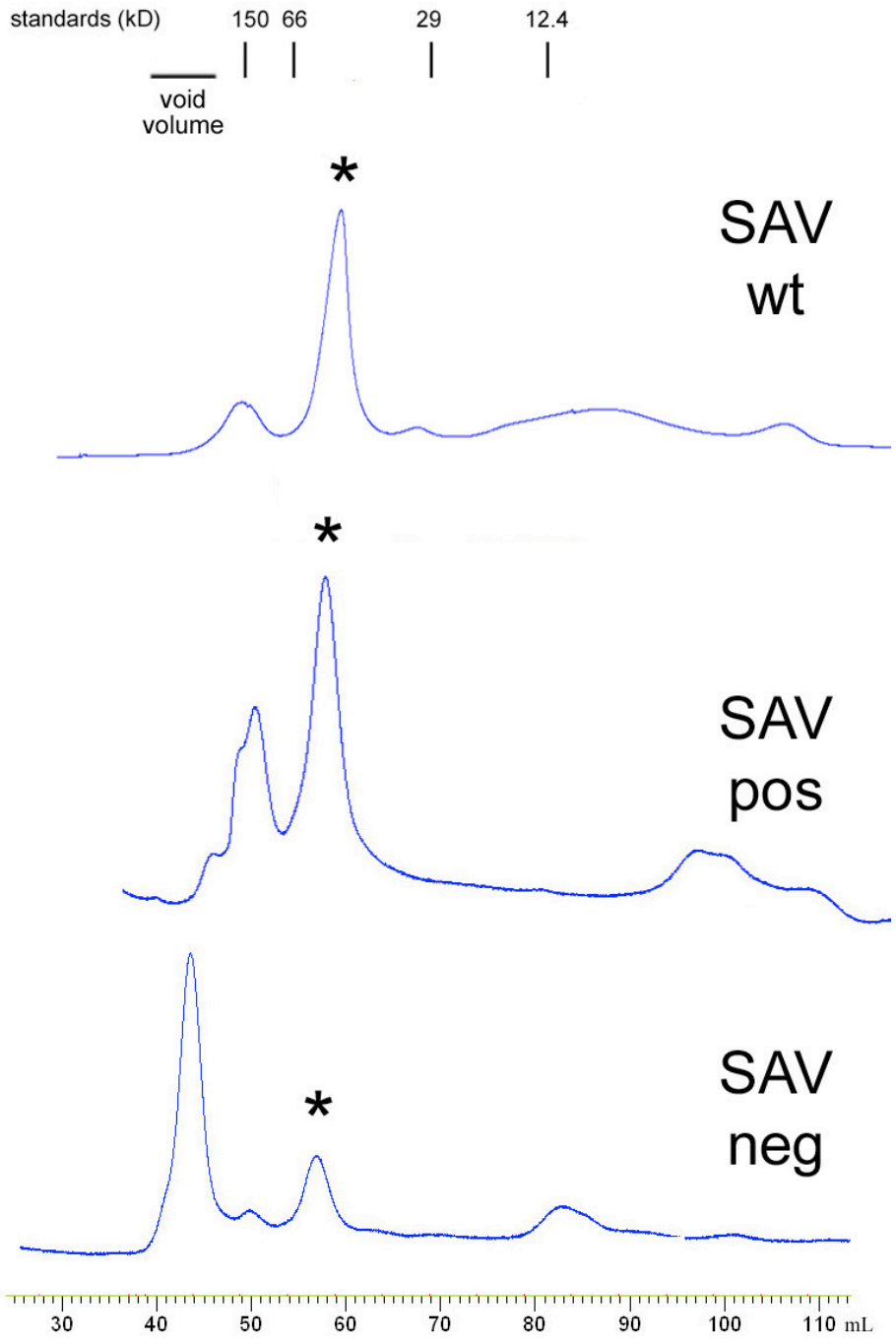

Figure S10. Analytical size-exclusion chromatography of streptavidin variants. 20 μg of protein was injected onto a Superdex 75 gel-filtration column. Conditions were as in Fig. S8.

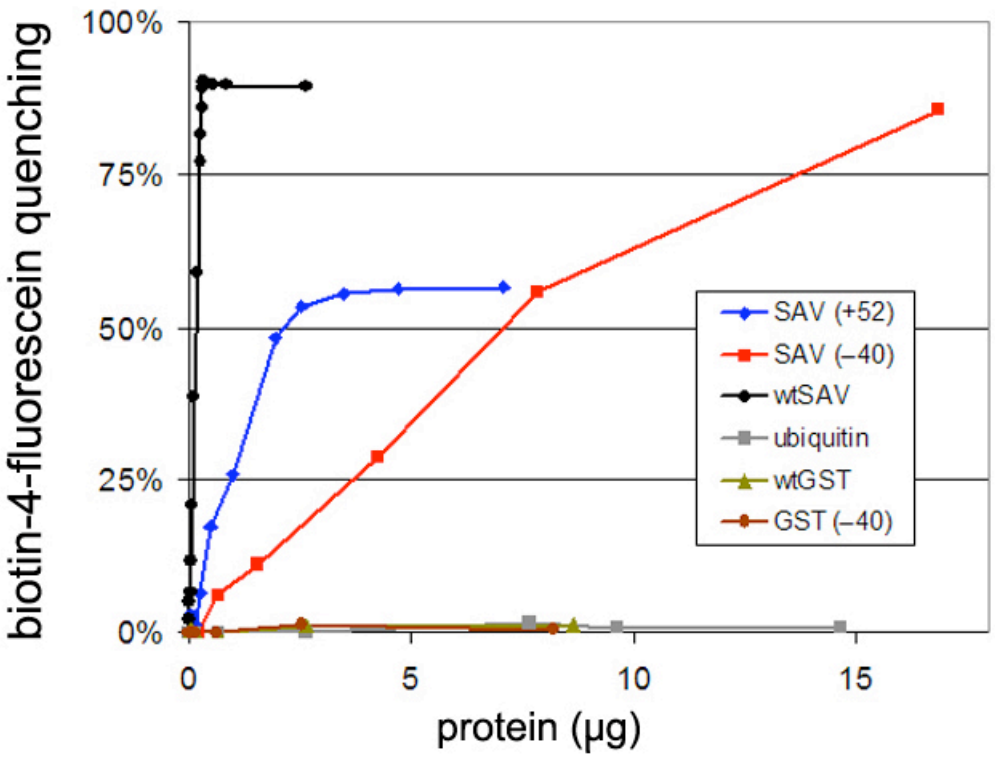

Figure S11. Biotin-binding activity of streptavidin variants, measured as described previously[9] by monitoring binding-dependent of biotin-4-fluorescein (Invitrogen). Protein samples were titrated into 0.3 µM biotin-4-fluorescein (B4F), 100 mM NaCl, 1 mM EDTA, 0.1 mg/mL bovine serum albumin (BSA), 50 mM potassium phosphate pH 7.5. Quenching of fluorescence at 526 nm was measured on a Perkin-Elmer LS50B luminescence spectrometer with excitation at 470 nm. Measurements were normalized to control titrations that contained a 600-fold excess of non-fluorescent biotin. The three proteins in the bottom of the legend are included as negative controls.

Table S1. Calculated and experimentally determined protein properties.

name	MW (kD)	length (aa)	n_{pos}	n_{neg}	$n_{charged}$	Q_{net}	pI	ΔG (kcal/mol)[a]	native MW (kD)[b]	% soluble after boiling[c]
GFP (–30)	27.8	248	19	49	68	–30	4.8	10.2	n.d.	98
GFP (–25)	27.8	248	21	46	67	–25	5.0	n.d.	n.d.	n.d.
stGFP	27.8	248	27	34	61	–7	6.6	11.2	n.d.	4
GFP (+36)	28.5	248	56	20	76	+36	10.4	8.8	n.d.	97
GFP (+48)	28.6	248	63	15	78	+48	10.8	7.1	n.d.	n.d.
SAV (–40)	14.3	137	5	15	20	–10	5.1	n.d.	55 ± 5 (tetramer)	99
wtSAV	13.3	128	8	9	17	–1	6.5	n.d.	50 ± 5 (tetramer)	7
SAV (+52)	14.5	137	16	3	19	+13	10.3	n.d.	55 ± 5 (tetramer)	97
GST (–40)	24.7	217	17	37	54	–20	4.8	n.d.	50 ± 5 (dimer)	96
wtGST	24.6	217	24	23	47	+1	7.9	n.d.	50 ± 5 (dimer)	3
GST (+50)[d]	24.7	217	39	14	53	+25	10.0	n.d.	n.d.	n.d.

n_{pos}, number of positively charged amino acids (per monomer)
n_{neg}, number of negatively charged amino acids
$n_{charged}$, total number of charged amino acids
Q_{net}, theroretical net charge at neutral pH
pI, calculated isoelectric point
n.d., not determined

[a]measured by guanidinium denaturation (Fig. 2c).
[b]measured by size-exclusion chromatography.
[c]percent protein remaining in supernatant after 5 min at 100 ºC, cooling to 25 °C, and brief centrifugation.
[d]protein failed to express in *E. coli*.

Table S2. Sequences of GFP variants. Residues are ordered by solvent exposure (AvNAPSA), from most exposed to least exposed. C-terminal residues disordered in the crystal structure are indicated as "(unstr"). Residues mutated in characterized variants are colored in red (negatively supercharged) and blue (positively supercharged). Residues varied in attempted target constructs are colored in light red and light blue.

AvNAPSA	aa #	–39	–30	–25	st	+36	+48	+73
(unstr)	238	K	K	K	K	K	K	K
(unstr)	237	Y	Y	Y	Y	Y	Y	Y
(unstr)	236	L	L	L	L	R	R	R
(unstr)	235	E	E	E	E	E	E	E
(unstr)	234	D	D	D	D	D	K	K
(unstr)	233	M	M	M	M	R	R	R
43	212	N	N	N	N	K	K	K
44	232	G	G	G	G	G	G	G
50	132	E	E	E	E	E	E	R
55	230	D	D	D	T	K	K	K
56	157	Q	Q	Q	Q	R	R	R
56	211	P	P	P	P	P	P	P
59	214	D	D	D	K	K	K	K
59	231	H	H	H	H	H	H	H
60	193	V	V	V	V	V	V	V
61	117	D	D	D	D	R	R	R
64	156	K	K	K	K	K	K	K
65	1	A	A	A	A	A	R	R
66	133	D	D	D	D	K	K	K
67	116	G	G	G	G	G	G	G
68	6	E	E	E	E	R	R	R
68	26	E	E	E	K	K	K	K
69	102	D	D	D	D	K	K	K
70	194	L	L	L	L	L	L	L
71	174	G	G	G	G	G	G	G
74	76	D	D	D	D	K	K	K
75	190	D	D	D	D	R	R	R
78	173	D	D	D	D	D	D	R
83	9	D	D	T	T	R	R	R
83	39	E	E	E	N	R	R	R
83	129	D	D	D	D	D	D	R
84	52	E	E	E	K	K	K	K
85	101	E	K	K	K	K	K	K
87	10	G	G	G	G	G	G	G
87	90	E	E	E	E	K	K	K
88	24	G	G	G	G	G	G	G
91	172	E	E	E	E	K	K	K
92	51	G	G	G	G	G	G	G
92	198	D	D	D	N	N	K	K
93	175	S	S	S	S	S	S	S
97	189	G	G	G	G	G	G	G
97	210	D	D	D	D	D	D	D

97	215	R	R	R	R	R	R	R
98	34	E	E	E	E	K	K	K
98	197	D	D	D	D	R	R	R
102	109	D	R	R	R	R	R	R
103	11	E	V	V	V	K	K	K
103	126	D	K	K	K	K	K	K
104	107	K	K	K	K	K	K	K
105	25	H	H	H	H	H	H	H
106	32	E	E	E	E	K	K	K
106	77	H	H	H	H	H	H	R
106	158	E	E	E	K	K	K	K
106	166	E	E	E	K	K	K	K
106	206	A	A	A	A	K	K	K
107	21	D	D	D	D	D	D	R
107	142	E	E	E	E	R	R	R
108	73	D	D	R	R	R	R	R
109	41	E	E	E	K	K	K	K
109	50	T	T	T	T	T	T	T
109	80	Q	Q	Q	Q	R	R	R
109	115	E	E	E	E	E	K	R
109	124	E	E	E	E	K	K	K
109	131	K	K	K	K	K	K	K
110	38	T	T	T	T	T	T	K
110	128	I	I	I	I	R	R	R
110	147	S	S	S	S	S	S	R
110	182	Y	Y	Y	Y	Y	Y	R
110	208	S	S	S	S	S	S	S
110	213	E	E	E	E	E	E	E
111	140	K	K	K	K	K	K	K
112	221	L	L	L	L	L	L	K
113	30	R	R	R	R	R	R	R
114	138	G	G	G	G	G	G	G
114	155	D	D	D	D	D	D	D
114	159	N	N	N	N	N	N	N
114	176	V	V	V	V	V	V	V
115	13	P	P	P	P	P	P	P
116	151	Y	Y	Y	Y	Y	Y	Y
116	204	E	E	E	Q	R	R	R
117	3	K	K	K	K	K	K	K
117	144	N	N	N	N	N	N	R
118	153	T	T	T	T	T	T	R
118	162	K	K	K	K	K	K	K
118	178	L	L	L	L	L	L	L
119	89	P	P	P	P	P	P	P
119	139	H	H	H	H	H	H	R
119	168	R	R	R	R	R	R	R
119	223	F	F	F	F	F	F	R
122	19	D	D	D	D	K	K	K
122	146	N	N	N	N	N	N	N

122	180	D	D	D	D	D	K	R
122	200	Y	Y	Y	Y	Y	Y	Y
124	28	S	S	S	S	S	S	R
124	49	T	T	T	T	T	T	T
124	97	T	T	T	T	T	T	T
125	36	D	D	D	D	D	D	R
125	118	T	T	T	T	T	T	K
125	209	K	K	K	K	K	K	K
126	43	T	T	T	T	T	T	R
126	99	S	S	S	S	S	S	S
126	188	I	I	I	I	I	I	I
126	202	S	S	S	S	S	S	R
126	228	G	G	G	G	G	G	G
127	95	E	E	E	E	E	E	E
127	111	E	E	E	E	E	E	R
128	170	N	N	N	N	N	N	N
129	7	L	L	L	L	L	L	L
130	134	G	G	G	G	G	G	G
130	149	D	D	D	N	K	K	K
130	164	E	E	E	N	K	K	K
130	184	Q	Q	Q	Q	Q	Q	Q
131	17	E	E	E	E	E	K	K
132	122	D	R	R	R	R	R	R
133	171	V	V	V	V	V	V	V
133	227	A	A	A	A	A	A	A
135	154	A	A	A	A	A	A	A
136	23	N	N	N	N	N	N	N
138	4	G	G	G	G	G	G	G
138	79	D	D	K	K	K	K	K
138	93	V	V	V	V	V	V	V
140	143	Y	Y	Y	Y	Y	Y	Y
141	113	K	K	K	K	K	K	K
141	195	L	L	L	L	L	L	L
141	225	T	T	T	T	T	T	T
142	2	S	S	S	S	S	S	S
142	5	E	E	E	E	E	K	K
142	15	L	L	L	L	L	L	L
142	105	T	T	T	T	K	K	K
142	216	D	D	D	D	D	D	D
142	229	I	I	I	I	I	I	I
143	186	T	T	T	T	T	T	T
144	33	G	G	G	G	G	G	G
144	219	V	V	V	V	V	V	R
146	45	K	K	K	K	K	K	K
150	217	H	H	H	H	H	H	H
151	47	I	I	I	I	I	I	I
151	75	P	P	P	P	P	P	P
151	104	G	G	G	G	G	G	G
153	191	G	G	G	G	G	G	G

154	81	H	H	H	H	H	H	H
156	48	C	C	C	C	C	C	C
157	120	V	V	V	V	V	V	V
157	137	L	L	L	L	L	L	L
158	187	P	P	P	P	P	P	P
162	103	D	D	D	D	D	D	D
163	54	P	P	P	P	P	P	P
164	40	G	G	G	G	G	G	G
164	192	P	P	P	P	P	P	P
165	20	G	G	G	G	G	G	G
165	160	G	G	G	G	G	G	G
166	12	V	V	V	V	V	V	V
166	114	F	F	F	F	F	F	F
171	135	N	N	N	N	N	N	N
172	31	G	G	G	G	G	G	G
172	207	L	L	L	L	L	L	L
176	78	M	M	M	M	M	M	M
178	148	H	H	H	H	H	H	H
179	56	P	P	P	P	P	P	P
179	177	Q	Q	Q	Q	Q	Q	Q
180	205	S	S	S	S	S	S	S
181	86	S	S	S	S	S	S	S
181	91	G	G	G	G	G	G	G
181	127	G	G	G	G	G	G	G
183	35	G	G	G	G	G	G	G
185	53	L	L	L	L	L	L	L
187	27	F	F	F	F	F	F	F
187	58	P	P	P	P	P	P	P
189	196	P	P	P	P	P	P	P
190	130	F	F	F	F	F	F	F
191	22	V	V	V	V	V	V	V
192	44	L	L	L	L	L	L	L
193	37	A	A	A	A	A	A	A
195	203	T	T	T	T	T	T	T
198	14	I	I	I	I	I	I	I
198	163	A	A	A	A	A	A	A
200	123	I	I	I	I	I	I	I
201	57	W	W	W	W	W	W	W
201	179	A	A	A	A	A	A	A
202	18	L	L	L	L	L	L	L
202	55	V	V	V	V	V	V	V
202	110	A	A	A	A	A	A	A
202	125	L	L	L	L	L	L	L
202	141	L	L	L	L	L	L	L
202	199	H	H	H	H	H	H	H
202	226	A	A	A	A	A	A	A
203	87	A	A	A	A	A	A	A
203	108	T	T	T	T	T	T	T
203	220	L	L	L	L	L	L	L

205	8	F	F	F	F	F	F	F
205	165	F	F	F	F	F	F	F
206	145	F	F	F	F	F	F	F
206	224	V	V	V	V	V	V	V
208	61	V	V	V	V	V	V	V
208	88	M	M	M	M	M	M	M
208	218	M	M	M	M	M	M	M
209	16	V	V	V	V	V	V	V
210	42	L	L	L	L	L	L	L
210	60	L	L	L	L	L	L	L
210	66	Y	Y	Y	Y	Y	Y	Y
210	98	I	I	I	I	I	I	I
210	100	F	F	F	F	F	F	F
211	29	V	V	V	V	V	V	V
211	67	G	G	G	G	G	G	G
211	185	N	N	N	N	N	N	N
212	65	T	T	T	T	T	T	T
212	183	Q	Q	Q	Q	Q	Q	Q
213	82	D	D	D	D	D	D	D
214	64	L	L	L	L	L	L	L
214	74	Y	Y	Y	Y	Y	Y	Y
214	85	K	K	K	K	K	K	K
214	96	R	R	R	R	R	R	R
214	152	I	I	I	I	I	I	I
214	222	E	E	E	E	E	E	E
215	46	F	F	F	F	F	F	F
215	59	T	T	T	T	T	T	T
215	119	L	L	L	L	L	L	L
215	121	N	N	N	N	N	N	N
216	62	T	T	T	T	T	T	T
216	71	F	F	F	F	F	F	F
216	72	S	S	S	S	S	S	S
217	83	F	F	F	F	F	F	F
217	167	I	I	I	I	I	I	I
217	181	H	H	H	H	H	H	H
218	106	Y	Y	Y	Y	Y	Y	Y
218	136	I	I	I	I	I	I	I
218	150	V	V	V	V	V	V	V
218	161	I	I	I	I	I	I	I
219	94	Q	Q	Q	Q	Q	Q	Q
219	201	L	L	L	L	L	L	L
220	63	T	T	T	T	T	T	T
221	84	F	F	F	F	F	F	F
221	112	V	V	V	V	V	V	V
221	169	H	H	H	H	H	H	H
222	69	Q	Q	Q	Q	Q	Q	Q
222	92	Y	Y	Y	Y	Y	Y	Y
223	68	V	V	V	V	V	V	V
224	70	C	C	C	C	C	C	C

Table S3. Sequences of GST variants. Residues are ordered as in Table S2. Residues mutated in GST(–40) are colored red. Residues chosen for mutation in the GST(+50) variant, which did not express in *E. coli*, are shown in light blue.

AvNAPSA	aa #	–40	wt	+50
46	36	E	E	K
55	57	D	D	K
67	114	E	E	K
68	118	E	K	K
69	25	D	D	K
69	194	P	P	P
70	198	D	N	K
71	169	D	D	K
72	111	A	A	A
74	115	E	K	K
74	166	S	S	S
75	35	M	M	M
75	79	E	K	K
77	80	D	D	K
77	185	P	P	P
79	29	E	K	K
79	165	P	P	P
80	27	S	S	S
80	107	T	T	T
80	110	E	E	K
81	133	E	Q	K
81	188	E	K	K
82	40	P	P	P
83	34	T	T	T
83	138	E	Q	K
84	136	G	G	G
85	144	S	S	S
88	43	P	P	P
88	192	A	A	A
88	205	G	G	G
89	39	P	P	P
90	135	E	Q	K
90	172	P	P	P
90	179	A	A	A
92	56	G	G	G
92	183	A	A	A
92	197	V	V	V
94	170	A	A	A
95	119	E	E	K
95	145	E	Q	K
96	122	E	E	K
96	189	A	A	A
97	108	D	N	K
98	44	S	S	S

100	82	**E**	K	K
102	37	T	T	T
105	103	T	T	T
105	186	**E**	K	K
108	81	**E**	Q	Q
109	31	E	E	E
109	195	E	E	E
110	132	S	S	S
111	75	G	G	G
112	32	V	V	V
112	176	A	A	A
112	193	S	S	S
113	125	K	K	K
114	206	K	K	K
115	204	N	N	N
116	112	G	G	G
118	200	P	P	P
120	42	K	K	K
120	59	T	T	T
120	175	S	S	S
120	207	Q	Q	Q
122	9	P	P	P
122	162	V	V	V
124	26	Q	Q	Q
125	121	P	P	P
126	139	A	A	A
128	100	K	K	K
129	83	E	E	E
129	113	K	K	K
129	129	T	T	T
130	99	C	C	C
130	134	N	N	N
130	180	R	R	R
130	203	G	G	G
131	164	N	N	N
132	54	Q	Q	Q
133	2	P	P	P
134	1	P	P	P
135	4	T	T	T
136	143	G	G	G
137	73	S	S	S
138	106	Y	Y	Y
138	109	Y	Y	Y
140	48	R	R	R
140	117	V	V	V
140	173	L	L	L
144	58	L	L	L
145	184	R	R	R
146	74	F	F	F

146	137	G	G	G
146	182	S	S	S
147	24	Q	Q	Q
147	161	Q	Q	Q
148	49	Q	Q	Q
150	41	L	L	L
150	55	D	D	D
150	171	F	F	F
151	102	A	A	A
152	52	K	K	K
152	140	F	F	F
153	3	Y	Y	Y
153	128	E	E	E
154	123	H	H	H
156	86	L	L	L
158	158	R	R	R
160	10	V	V	V
160	38	W	W	W
160	96	D	D	D
161	178	V	V	V
162	33	V	V	V
162	191	L	L	L
164	30	E	E	E
164	78	G	G	G
165	46	L	L	L
166	11	R	R	R
167	23	D	D	D
167	196	H	H	H
168	28	W	W	W
168	45	C	C	C
168	126	P	P	P
169	6	T	T	T
170	71	G	G	G
170	130	L	L	L
170	146	I	I	I
170	201	I	I	I
171	61	Y	Y	Y
172	167	C	C	C
173	76	L	L	L
174	13	R	R	R
174	84	A	A	A
174	95	E	E	E
175	8	F	F	F
176	199	R	R	R
177	72	R	R	R
178	22	A	A	A
178	202	N	N	N
179	105	I	I	I
180	163	L	L	L

181	104	L	L	L
186	63	S	S	S
186	168	L	L	L
187	120	L	L	L
188	181	L	L	L
188	190	F	F	F
190	89	M	M	M
190	97	L	L	L
190	116	Y	Y	Y
190	187	I	I	I
191	69	H	H	H
191	142	V	V	V
192	93	G	G	G
194	85	A	A	A
194	159	I	I	I
196	7	Y	Y	Y
196	12	G	G	G
196	124	L	L	L
197	47	F	F	F
197	150	D	D	D
198	18	R	R	R
198	21	L	L	L
199	154	L	L	L
199	155	D	D	D
199	156	L	L	L
199	174	L	L	L
200	60	L	L	L
200	77	Y	Y	Y
201	15	E	E	E
201	62	Q	Q	Q
201	98	R	R	R
202	94	V	V	V
202	101	Y	Y	Y
202	131	L	L	L
203	50	L	L	L
203	87	V	V	V
204	16	A	A	A
204	90	V	V	V
204	92	D	D	D
205	5	I	I	I
205	64	N	N	N
205	151	Y	Y	Y
207	51	P	P	P
207	127	F	F	F
207	160	H	H	H
208	65	A	A	A
208	157	L	L	L
209	147	S	S	S
209	177	Y	Y	Y

210	66	I	I	I
211	14	C	C	C
211	91	N	N	N
212	70	L	L	L
213	153	L	L	L
216	19	M	M	M
216	53	F	F	F
216	141	V	V	V
216	152	N	N	N
217	20	L	L	L
218	67	L	L	L
219	17	M	M	M
220	88	D	D	D
221	149	A	A	A
223	68	R	R	R
223	148	F	F	F

Table S4. Sequences of streptavidin variants. Residues are ordered as in Table S2. C-terminal residues disordered in crystal structure are indicated as "unstr". Residues mutated in SAV(–40) are shown in red; those mutated in SAV(+52) are shown in blue.

AvNAPSA	aa #	–40	wt	+52
(unstr)	123	E	K	K
(unstr)	124	P	P	P
(unstr)	125	S	S	S
(unstr)	126	A	A	A
(unstr)	127	A	A	A
(unstr)	128	S	S	S
32	89	A	A	A
46	55	T	T	T
55	90	E	E	K
58	4	A	A	A
58	71	D	N	K
59	88	G	G	G
63	24	A	A	A
69	56	D	D	K
69	106	A	A	A
72	72	Y	Y	Y
79	3	E	E	K
80	105	E	E	K
81	57	G	G	G
81	122	V	V	V
82	25	D	D	K
86	104	T	T	T
90	42	E	R	R
90	107	D	N	K
95	40	E	E	K
97	52	A	A	A
98	54	A	A	A
102	41	S	S	S
103	73	E	R	R
104	21	T	T	T
104	38	D	N	N
112	11	Y	Y	Y
113	87	G	G	G
114	5	G	G	G
114	29	T	T	T
114	110	K	K	K
115	15	G	G	G
115	23	G	G	G
115	92	R	R	R
116	9	T	T	T
117	7	T	T	T
118	27	A	A	A
119	8	G	G	G
122	33	E	E	E

124	35	A	A	A
126	19	I	I	I
126	31	T	T	T
126	37	G	G	G
126	58	S	S	S
127	118	T	T	T
129	91	A	A	A
129	120	T	T	T
131	44	V	V	V
133	51	S	S	S
135	13	Q	Q	Q
136	53	P	P	P
136	121	K	K	K
137	76	H	H	H
138	70	N	N	N
139	86	V	V	V
140	74	N	N	N
141	17	T	T	T
141	36	V	V	V
141	108	A	A	A
142	103	T	T	T
143	26	G	G	G
148	30	G	G	G
148	69	K	K	K
149	14	L	L	L
150	2	A	A	A
151	94	N	N	N
153	109	W	W	W
154	101	S	S	S
157	75	A	A	A
158	116	H	H	H
159	85	Y	Y	Y
165	22	A	A	A
167	50	D	D	D
173	113	L	L	L
174	48	R	R	R
175	102	G	G	G
176	46	T	T	T
176	59	G	G	G
178	16	S	S	S
180	84	Q	Q	Q
182	39	A	A	A
182	49	Y	Y	Y
183	34	S	S	S
183	60	T	T	T
184	77	S	S	S
185	111	S	S	S
190	10	W	W	W
191	43	Y	Y	Y

192	12	N	N	N
193	112	T	T	T
194	45	L	L	L
196	20	V	V	V
197	115	G	G	G
198	6	I	I	I
198	100	T	T	T
200	96	Q	Q	Q
206	99	L	L	L
208	28	L	L	L
209	95	T	T	T
209	117	D	D	D
210	47	G	G	G
210	97	W	W	W
210	114	V	V	V
212	78	A	A	A
213	61	A	A	A
215	83	G	G	G
219	67	A	A	A
220	98	L	L	L
222	93	I	I	I
223	79	T	T	T
224	62	L	L	L
224	68	W	W	W
225	66	V	V	V
225	119	F	F	F
230	81	W	W	W
230	82	S	S	S
231	18	F	F	F
232	32	Y	Y	Y
233	65	T	T	T
234	64	W	W	W
235	63	G	G	G
240	80	T	T	T