Supporting Information

Lindman et al. 10.1073/pnas.1005689107

SI Materials and Methods

Cloning of NGFP-PGB41-56 and PGB1-40-CGFP. DNA fragments covering residues 1 to 40 and 41 to 56 of PGB1-QDD were amplified by PCR from a PetSac vector carrying a synthetic PGB1-QDD gene. The primers used (Table S1) were designed to introduce restriction sites for subsequent cloning into vectors pET11a-link-NGFP and pMRBAD-link-CGFP carrying the residues 1 to 157 (NGFP) and 158 to 238 (CGFP) of dissected GFP gene, respectively (1), kind gifts from the Professor Lynne Regan, Yale University. The PCR was carried out using the Expand High Fidelity DNA polymerase (Roche Diagnostics). The PCR products were purified after agarose gel electrophoresis using GFX PCR, DNA and gel band purification Kit (GE Healthcare) or by electroelution inside a small dialysis bag (MWCO 3,500 Da; Spectrum Laboratories). Plasmid pET11a-link-NGFP was amplified and doubly digested with BamHI [New England Biolabs (NEB)] and XhoI (NEB) overnight at 37 °C. Plasmid pMRBAD-link-CGFP was amplified and first digested for two hours by AatII (NEB) at 37 °C, followed by addition of NcoI (NEB) and double digestion continued overnight. The amplified and purified PCR product carrying PGB1-QDD residues 1 to 40 was digested with AatII and NcoI, and the PCR band with residues 41 to 56 by BamHI and XhoI. Digested vectors and PGB1-40 fragments were purified after gel electrophoresis using the GFX PCR, DNA and gel band purification kit; whereas doubly digested PGB41-56 fragment was purified by electroelution. The double digestion was repeated once more in order to increase the ratio of digested plasmids and PCR bands available for successful ligation.

The doubly digested PCR fragments were ligated into the appropriate vector using T4 DNA ligase (USB) during 16 h at 16 °C using different ratios of fragment over vector. 1 μL of each ligation product was mixed with 50 μL of *Escherichia coli* ER2566 calcium competent cells and tubes kept in ice bath for 30 min, incubated at 42 °C during 45 s and incubated again 10 min in ice before plating on LB/agar. The pET11a-NGFP-PGB1 ligation products were plated on LB agar containing 50 μg/mL ampicillin; while pMRBAD-CGFP were plated on LB agar containing 100 μg/mL kanamycin. Several colonies from the plates grown overnight at 37 °C were picked and amplified during at least 8 h at 37 °C in 1.5–2 mL of LB medium containing appropriate antibiotic. The bacteria were amplified in 15 mL falcon tubes laid down for maximal exposition surface to air during agitation at 130 rpm in an incubator. Amplified plasmids were recovered and purified using Illustra plasmidPrep Mini Spin Kit (GE Healthcare Life Sciences). To identify plasmids containing the correct gene, DNA was sequenced using BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems) and separated by capillary chromatography (purchased at BM Unit). The primer 5′ CACAACGTTCCCATCATGGCAGAC 3′ was used for sequences in the pET11A plasmid and 5′ CTCTCTACTGTTTCTCCATACCCG 3′ for the pMRBAD plasmid.

Library Cloning. A PGB1-40 focused library was constructed by overlap extension PCR of degenerate oligonucleotides to yield a degenerate DNA library that translates into a degenerate protein fragment library (Fig. 2*E*). The DNA library (120 base pairs plus cloning sites) is constant at most base positions but randomized over two or three alternatives at 11 base positions to yield the following amino acid alternatives: Y3 or F3; L7 or I7; V39 or I39; T16 or I16; T18, P18, I18 or L18; T25, A25, P25, K25, E25, or Q25; V29, F29, K29, E29, Y29, D29, I29, L29, or N29. The randomization at DNA level will also give a stop codon at position 29, but this alternative will be lost at colony level as the GFP part will not be expressed. The library DNA was first built by PCR amplification of two overlapping oligonucleotides followed by digestion by TseI (NEB) and ligation at 16 °C for 16 h. In the case of GCG (Ala) at position 25, there will be additional Tse1 cleavage site, which may results in a deletion, thus a small proportion of the library will have DAAAE instead of DAAAAE over positions 22 to 27. Deletions were found in one of top 25 clones after screening. However, insertion of an extra Ala was seen in four cases, probably due to mismatch in PCR using many identical codons in a row. Approximately 1 μL of the library DNA was amplified using the primers in Table S1, digested by AatII and NcoI, purified and ligated with cleaved (by AatII and NcoI) vector pMRBAD-link-CGFP carrying the gene for GFP residues 158 to 238 (CGFP, kanamycin resistant). After transformation into calcium competent *E. coli* ER2566 and growth on LB agar plates with 100 μg/mL kanamycin overnight at 37 °C, several colonies were picked, amplified in liquid culture over night and plasmids purified. The plasmids were sequenced in order to corroborate the correct cloning of the library and occurrence of the mutations.

Coexpression. The plasmids coding for NGFP-PGB41-56 and PGB1-40-CGFP (or NGFP-PGB41-56 and PGB1-40_{lib}-CGFP) were cotransformed into calcium competent *E. coli* ER2566 and spread on LB agar plates with 100 μg/mL ampicillin, 35 μg/mL kanamycin, 10 μM isopropyl-β-D-1-thiogalactosid (IPTG) and 0.2% (w/v) arabinose (inducing plates) or with 100 μg/mL ampicillin and 35 μg/mL kanamycin (noninducing plate). The plates were first incubated at 37 °C for 16 h and then taken out and left at RT for up to five days. The appearance of green fluorescence from the colonies was monitored using a transilluminator with light emission between 420 and 500 nm (Dark Reader DR45M nonUV blue light transilluminator; Clare Chemical Research, Inc.). Plates were imaged using a digital CCD camera Sony Cybershot DSC-W5. Coexpression of the constructs GFPN-EF1 and GFPC-EF2 (26) were used as positive controls for GFP assembly and fluorescence development.

Screening for High Affinity Variants and DNA Sequencing. The plasmids coding for NGFP-PGB41-56 and PGB1-40_{lib}-CGFP were cotransformed into *E. coli* ER2566 and spread on inducing plates. The parent constructs NGFP-PGB41-56 and PGB1-40-CGFP were cotransformed in parallel for comparison. Single colonies that were judged to have higher green fluorescence intensity than parent were picked and cultivated overnight in LB with 100 μg/mL ampicillin and 35 μg/mL kanamycin, followed by plasmid preparation using Illustra plasmidPrep Mini Spin Kit (GE Healthcare Life Sciences). Single clone plasmids were cotransformed with NGFP-PGB41-56 in *E. coli* ER2566 by thermal shock, spread on inducing plates, incubated during 16 h at 37 °C, followed by incubation at RT up to five days. The intensity of green fluorescence was ranked by human eye after comparison of inducing plates. All single clones were compared versus one another and each clone versus parent construct, using the transilluminator. The DNA for the library insert of single clones was sequenced using the primer 5′CTCTCTACTGTTTCTCCATACCCG3′ and BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems) followed by capillary chromatography (purchased at BM Unit).

Cloning, Expression, and Purification of Intact Protein of Selected Variants. Synthetic genes for intact PGB1 corresponding to the top

three variants with codons preferred by *E. coli* were cloned into the ampicillin resistant pET3a plasmid (Genscript). The plasmids were transformed into calcium competent *E. coli* BL21star(DE3) pLysS. Single colonies were used to inoculate overnight cultures of LB medium with 50 μg/ml ampicillin and 30 μg/ml chloramphenicol. The overnight cultures were diluted 1:100 in the day cultures of 500 mL each in 2.5 L baffled flasks. Protein production was induced by adding 0.4 mM IPTG at an OD600 of 0.6 to 0.8, and the culture was harvested by centrifugation 3 to 4 hours later. The cell pellet was resuspended in buffer A (10 mM Tris/HCl, 1 mM ethylenedinitrilotetraacetic acid (EDTA), pH 7.5) and sonicated for 5 min and centrifuged for 10 min at 15,000 g. The supernatant was poured into boiling buffer A at a 1:1 ratio, heated to 80 °C, and then directly cooled on ice. The solution was centrifuged at 15,000 g for 10 min, and the supernatant was pumped onto 100 ml DEAE cellulose (Whatman). The column was washed with buffer A and eluted using a 0–400 mM linear NaCl gradient in buffer A. Fractions containing PGB1 were pooled and lyophilized, dissolved in 10 mL Millipore water, and separated on a 3.4 × 180 cm Sephadex G50 superfine gel filtration column using 50 mM ammonium acetate, pH 6 as running buffer. Fractions containing PGB1 were pooled, lyophilized and desalted on a Sephadex G25 superfine gel filtration column in water. Desalted protein was lyophilized and stored at −20 °C. The proteins were judged homogeneous by SDS/PAGE, analytical gel filtration chromatography, and mass spectrometry.

Top3 appeared to be sensitive to storage before purification and was purified in a more rapid manner using batch chromotography. The cell pellet was sonicated and directly mixed with DEAE cellulose (Whatman). The protein was washed with buffer A on a Büchner funnel and eluted using increasing NaCl concentration in steps of 50 mM from 200 to 500 mM. Fractions containing the protein of interest were directly applied to centrifugal devices with a 10 kDa cut-off. The flow through was concentrated and buffer exchanged on a 3.5 kDa cutoff centrifugal device. As for the other variants, the purity and size was judged by SDS/PAGE, and mass spectrometry.

MALDI-TOF Mass Spectrometry. MALDI-TOF mass spectrometry was performed using a 4700 proteomics analyzer (Applied Biosystems). The parent and mutant PGB1-QDDs were analyzed in comparison to the control proteins cytochrome c (12,352 Da), BPTI (6,511 Da), amyloid β protein (4,459 Da), and calbindin D_{9k} (8,661 Da). Desalted protein samples were diluted 1:10 or 1:100 in 50% acetonitrile, 0.3% TFA, 15 mg/ml sinapic acid and 0.5 μl dispensed onto a MALDI sample support on top of a dried layer of 0.5 μl of 95% acetonitrile, 0.03% TFA, 1.5 mg/ml sinapic acid, and allowed to air dry (2). All analyses were performed in positive linear mode collecting data from approximately 900 to 1,500 single laser shots (5 kV).

Fragment Expression and Purification. The PGB1-QDD 1-41 fragment (N41) with parent sequence and with Top1, Top2, and Top3 sequence were cloned as intein-fusions in the PTXb1 vector, expressed in *E. coli* and purified by affinity chromatography, thiol cleavage as described (3), followed by incubation in carbonate buffer, pH 9, to achieve the correct C-terminal carboxylate as verified by MALDI-TOF. PGB1-41-56 (C16) was purchased as a synthetic peptide with >98% purity from Innovagen (Lund).

Fragment Complementation. The affinity between complementary fragments (with parent or mutated sequence) was measured using far-UV CD spectroscopy. CD spectra were recorded between 190 and 250 nm in a 0.1 mm cuvette (scan rate 20 nm/min, response 8 s, data interval 0.5 nm, band width 1 nm, accumulations 2) using a JASCO J-815 spectropolarimeter with a PTC-423S Peltier element at 5 °C and at 37 °C. In each titration the concentration of PGB1-41-56 (C16) was held constant at 0.6 mM while the concentration of PGB1-QDD-1-41 (N41) with parent or mutated sequence was varied between 0 and 1.2 mM. Two or three titration experiments were conducted for each variant and for parent. The equilibrium dissociation constant, K_D, was estimated by fitting a 1:1 binding model to the CD signal (θ) as a function of total N41 concentration, C_N:

$$\theta = \theta_N * [N41] + \theta_C * [C16] + \theta_{NC} * [N41{\cdot}C16]$$

where θ_{N41}, θ_{C16} and θ_{NC} are the responses per mM of free N41, free C16 and the N41-C16 complex, respectively. All concentrations and K_D were expressed in mM. The free N41 concentration was solved from the 1:1 binding equation as

$$[N41] = 0.5 * (C_N - 0.6 - K_D) + \sqrt{(0.25 * (C_N - 0.6 - K_D)^2 + C_N * K_D)}.$$

The complex concentration was calculated as

$$[N41{\cdot}C16] = 0.6 * [N41]/([N41] + K_D),$$

and the free C16 concentration as

$$[C16] = 0.6 - [N41{\cdot}C16].$$

SI Results

SDS/PAGE. The purified proteins PGB1-QDD, Top1, Top2. and Top3 were analyzed by SDS/PAGE (18% gel) (Fig. S1). All four proteins migrate as one band. None of PGB1-QDD or variants migrates in SDS/PAGE according to expected molecular weights but similar to one another and somewhat longer than PGB1-QDD (Fig. S1). The low mobility is probably due to poor SDS binding. It is not uncommon for small protein to migrate in a very sequence-dependent manner on SDS/PAGE.

Gel Filtration. The purified proteins PGB1-QDD, Top1, Top2, and Top3 were analyzed by analytical size exclusion chromatography on a Superdex 75 column. The elution volume is highly sequence-dependent for small proteins; therefore literature data on the elution volumes of monomeric and dimeric PGB1 were used as controls (4). The elution volume and profile for PGB1-QDD and mutants are in agreement with monomer (Fig. S2) at all concentrations tested.

Mass Spectrometry. The purified proteins PGB1-QDD, Top1, Top2, and Top3 were analyzed by MALDI-TOF mass spectrometry, in comparison with four control proteins BPTI, calbindin D_{9k}, Aβ(M1-40), and cytochrome *c*. The instrument was focused at 6,000 and therefore BPTI was used as an internal standard (6,511.1 Da). A singly charged species with the expected monomer Mw was found to be strongly dominating for PGB1-QDD and the three variants (Fig. S3). The observed and theoretical masses are within the error of the measurement for PGB1-QDD (measured 6,224, theoretical 6,225), Top1 (measured 6,256.3, theoretical 6,256.8), Top2 (measured 6,241.7, theoretical 6,242.8), Top3 (measured 6,223.8, theoretical 6,223.8), as well as the control proteins Aβ(M1-40) (measured 4,461, theoretical 4,459), calbindin D_{9k} (measured 8,665, theoretical 8,661), cytochrome *c* (measured 12,355, theoretical 12,352). Cytochrome *c* was used to verify that a species with the Mw of a PGB1 dimer would be detected. The data support that the PBG1 samples are monomeric.

Affinity Between Fragments. Examples of data from the affinity measurements are shown in Fig. S5. For the parent construct we obtain a K_D of 0.14–0.02 mM, while for the mutants we obtain

K_D values of 0.03–0.01 mM for Top1, ≤0.02 mM for Top2 and 0.02–0.01 mM for Top3. Our results clearly discriminates all variants as having higher affinity than the parent construct. However, the data does not allow for ranking of the mutants whose values overlap within error. The affinity for the variants has increased ca. 10-fold compared to parent. Better estimates of K_D for the mutants would require mutants and parent to be studied in different concentration ranges, but this would deteriorate comparison between parent and mutants (5).

SI Discussion.

A general method for protein stabilization based on the current results is outlined as follows.

1. **Fragment design.** The protein in need for stabilization is dissected into two fragments, A and B. It is required that the two fragments recombine into a native arrangement with at least moderate affinity ($KD \leq 1$ mM). The ideal cut site is in a flexible region in a loop on the protein surface. If a fragment complementation system is already described for the protein that may be an ideal choice. Otherwise one may try different cut sites in parallel before embarking on library production.
2. **Fragment cloning and coexpression.** The fragments are cloned into the split GFP system (1) and emergence of some level of green fluorescence verified for coexpression of the wild-type construct under inducing compared to noninducing conditions.
3. **Library generation.** The fragments are subjected to library generation at DNA level using multiple designed substitutions, error-prone PCR and/or gene shuffling. By error-prone PCR, single base substitutions will dominate at DNA level limiting the possible substitutions of each amino acid to 5 to 9 alternatives. To reach high levels of stabilization one may have to do several rounds of error-prone PCR or a combination of designed substitutions and error-prone PCR. The library DNA is cloned into the split GFP system.
4. **In vivo screening.** Library A is cotransformed with wild-type fragment B, library B with wild-type fragment A, and also library A with library B to allow for wider possibilities of synergistic mutations. A very large number of transformants is desirable and the brightest green colonies are picked for single clone plasmid preparations. These single clone plasmids are cotransformed with their cognate partner and the intensity of green fluorescence ranked compared to other clones and to wild type.
5. **Protein production and stability measurement.** The sequences of the top fluorescent clones are used to produce new intact proteins and their stability toward denaturation measured in comparison to wild type to confirm the degree of stabilization. The method can be reiterated for further stability increase using the original libraries or new libraries based on the findings of the first round. The method can be combined with functional assays to retain only functional variants.

1. Wilson CG, Magliery TJ, Regan L (2004) Detecting protein–protein interactions with GFP fragment reassembly. *Nat Methods* 1:255–262.
2. Önnerfjord P, et al. (1999) Homogeneous sample preparation for automated high throughput analysis with matrix assisted laser desorption/ionisation time of flight mass spectrometry. *Rapid Commun Mass Sp* 13:315–322.
3. Bauer MC, Xue WF, Linse S (2009) Protein GB1 folding and assembly from structural elements. *Int J Mol Sci* 10:1552–1566.
4. Jee J, Byeon IJ, Louis JM, Gronenborn AM. The point mutation A34F causes dimerization of GB1. *Proteins* 71:1420–1431.
5. Linse S, Jönsson B, Chazin WJ (1995) The effect of protein concentration on ion binding. *Proc Natl Acad Sci USA* 92:4748–4752.

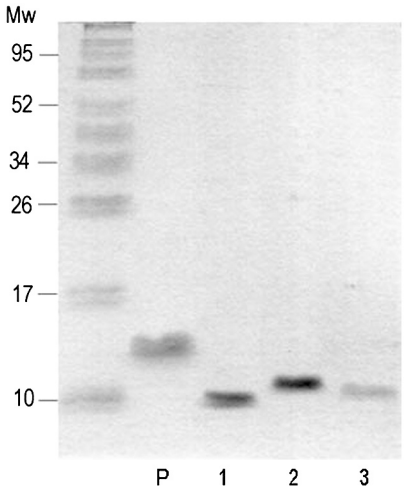

Fig. S1. SDS/PAGE (18%) of PQB1-QDD (lane P), Top1 (lane 1), Top2 (lane 2), and Top3 (lane 3).

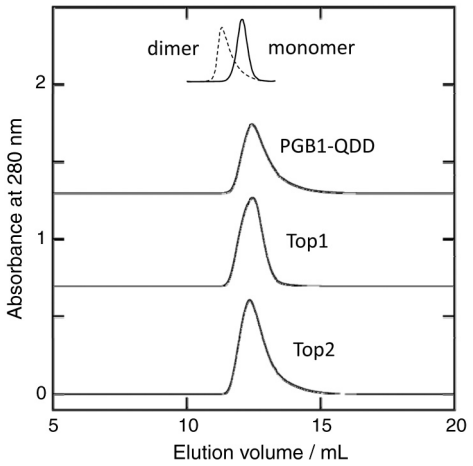

Fig. S2. Gel filtration on a superdex 75 column, of PGB1-QDD, Top1 and Top2 in 10 mM Na-phosphate buffer, 0.15 M NaF, pH 7.2. The reference spectra at the top for monomeric wild-type PGB1 (solid line) and dimeric A34F mutant (dashed line) are from ref. 3.

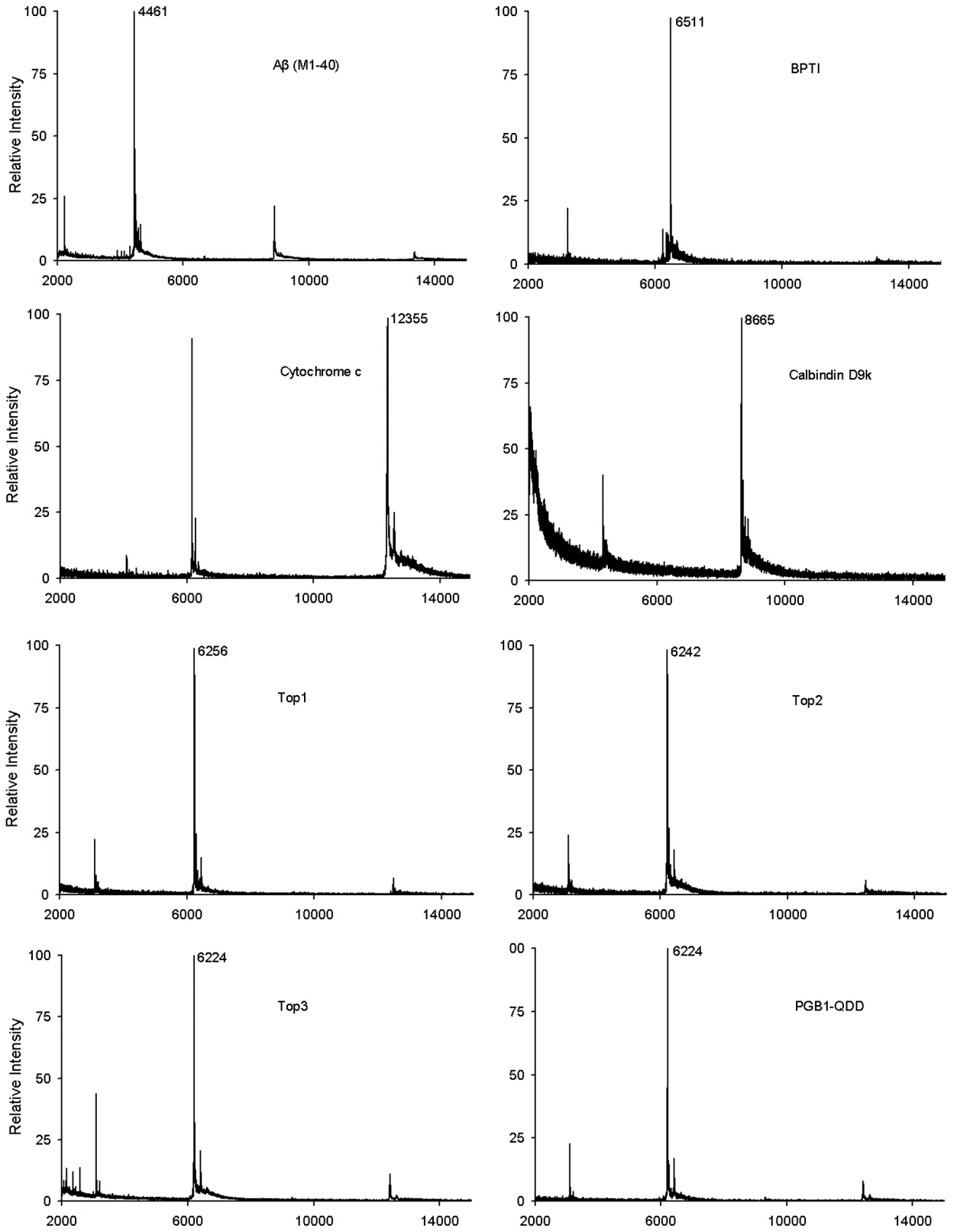

Fig. S3. MALDI-TOF of control proteins and PGB1-QDD and variants.

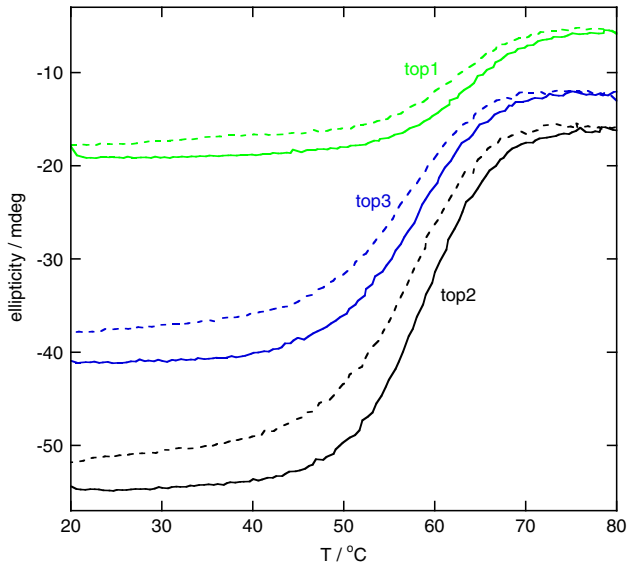

Fig. S4. Thermal scans in forward (solid line) and reverse (dashed line) mode as monitored by CD signal at 218 nm for top1 (green), top 2 (black), and top3 (blue) at three different concentrations.

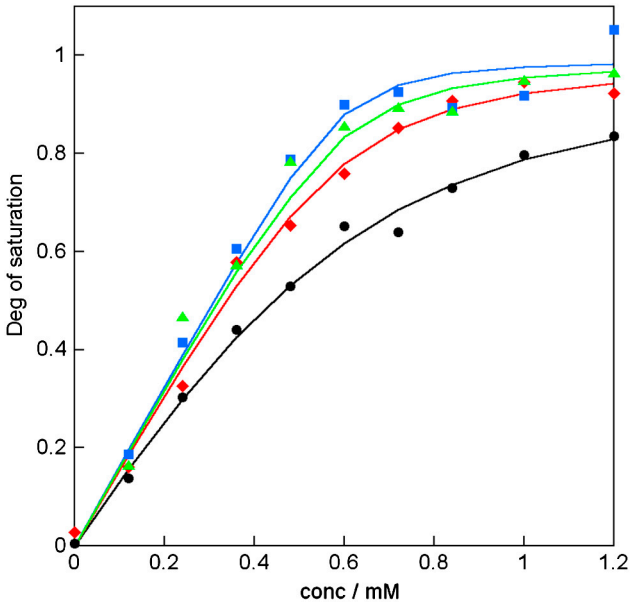

Fig. S5. Fragment complementation. Normalized CD titration data presented as the degree of saturation of C16 as a function of total concentration of N41 with parent (black) or Top1 (red) or Top2 (blue) or Top3 (green) sequence. The total C16 concentration held constant at 0.6 mM. The solid lines are fitted curves according to a 1:1 binding model. The curves were fitted to the raw data and then normalized for presentation.

Table S1. PGB1-QDD 1–40 sequences of 25 top fluorescent colonies

		Sequence
		10 20 30 40
Rank Order	Colony	....|....|....|....|..x..|....|....|....|
–	parent	MQYKLILDGKTLKGETTTEAVD-AATAEKVFKQYANDDGVD
1	L37*	......I.....-..A...F.....I.
2	L76	......I.....-..A...F.....
3	L84	......I.....I....-..A..T.....T....I.
4	L87	..F.....P....-..P.....I.
5	L16	.....P....-..A.....
6	L104	..F.....I.....-..A...I.....
7	L52	......I.....I.....-..A...F.....I.
8	L30	..F.....I.P....-..E...F.....
9	L42	......I.....P....A..A...F.....I.
10	L6	..F.....I....-..A...F.....
11	L8	.....I.I....-..E...F.....I.
12	L45	.....L....-..P...F.....
13	L65	......I.....-..K...E.....I.
14	L7	..F...I.....I.L..D.-..P...F.....I.
15	L96	......I.....I....-..Q...F.....
16	L17	.....P....A..P.....
17	L34	..F...I.....I.P....-.A...F.....I.
18	L22	......I.....I.P....-..E...D.....I.
19	L29	..F.....I.L....-..E...Y.....
20	L9	.....I....-..P...L.....I.
21	L112	..F...I.....I.....A..A...F.....I.
22	L36	..F...I.....I.P....-..P...Y.....I.
23	L106	.....I.I....-..P...I.....
24	L26	..F.....I.P....A.....N.....
25	L85	......I.....I....-..P...F.....
	HS1†	..F...I.....I.I....-..E...L.....I.
	HS2‡	..F...I.....I.I....-..E...F.....I.
	HS3‡	..F...I.....I.I....-..E...Y.....I.

*Colonies underlined developed the highest fluorescence after four days of incubation, others after three days.

†HS1: the most highly stable sequences found with computational calculations by Malakauskas and Mayo (1).

‡HS2 and 3: the most highly stable sequences found by phage display by Wunderlinch et al. (2). They only tested the positions 16, 18, 25, and 29, and the mutations Y3F, L7I, and V39I were introduced deliberately.

1 Malakauskas SM, Mayo SL (1998) Design, structure and stability of a hyperthermophilic protein variant. *Nat Struct Mol Biol* 5:470–475.

2 Wunderlich M, Martin A, Staab CA, Schmid FX (2005) Evolutionary protein stabilization in comparison with computational design. *J Mol Biol* 351:1160–1168.

Table S2. Primers used to clone GFPN-PGB41-56 and PGB1-40-GFPC

Name	Sequence	Restriction site
NGFP_PGB41-56 start:	5′-AATAATCTCGAGCGGTGAATGGACCTACGACGAC-3′	XhoI
NGFP_PGB41-56 stop:	5′-AATAATGGATCCTTATTCGGTAACGGTGAAGGTTTTGG-3′	BamHI
CGFP_PGB1-40 start:	5′ AATAATCCATGGCTATGCAGTACAAACTGATCCTGGA-3′	NcoI
CGFP_PGB1-40 stop:	5′ AATAAGACGTCCCGTCAACACCGTCGTCGTTAGC-3′	AatII

Underlined bases are the annealing sequence; italics are the restriction sites, and gray is the sequence lost after enzyme digestion.