SUPPORTING INFORMATION

Effect of Highly Fluorinated Amino Acids on Protein Stability at a Solvent-Exposed Position on an Internal Strand of Protein G B1 Domain

Hsien-Po Chiu, Bashkim Kokona, Robert Fairman, Richard P. Cheng*

Figures

Figure S1. Thermal denaturation of GB1-based proteins monitored by CD at 222 nm. S2

Figure S2. Chemical shift deviation ($\Delta\delta^1$H) of HCα from random coil values for the residues surrounding the solvent-exposed guest position 53 for proteins GB1-Ala, GB1-Atb, GB1-Qfl, and GB1-Hfl. S5

Figure S3. Schematic diagram depicting intra- and inter-strand NOE cross peaks involving backbone HN and HCα protons for residues near the guest position 53 of proteins GB1-Ala, GB1-Atb, GB1-Qfl, and GB1-Hfl. S6

Figure S4. Chemical shift deviation ($\Delta\delta^1$H) of HCα from random coil values for the helical region (Ala24-Asn37) for proteins GB1-Ala, GB1-Atb, GB1-Qfl, and GB1-Hfl. S7

Figure S5. The unfolding energy (ΔG_{unfold}) of the GB1-based proteins plotted against temperature near the T_m for each protein along with the corresponding curve fits. S8

Figure S6. The relative unfolding free energy at 60°C ($\Delta\Delta G_{unfold\ 60°C}$) for the GB1-based proteins plotted against side chain volume and the hydrophobic parameter logP. S10

Figure S7. Analytical ultracentrifuge data from sedimentation equilibrium experiments for the nine GB1-based proteins. S12

Tables

Table S1. ^{1}H Chemical Shift Assignments for 12 Strand Residues Near the Guest Site of GB1-Ala. S3

Table S2. ^{1}H Chemical Shift Assignments for 12 Strand Residues Near the Guest Site of GB1-Atb. S3

Table S3. ^{1}H Chemical Shift Assignments for 12 Strand Residues Near the Guest Site of GB1-Qfl. S4

Table S4. ^{1}H Chemical Shift Assignments for 12 Strand Residues Near the Guest Site of GB1-Hfl. S4

Table S5. ΔH_{Tm}, ΔS_{Tm}, T_m, and $\Delta G_{unfold\ 60°C}$ for Unfolding the GB1-Based Proteins. S9

Table S6. The Formula, Calculated *m*/*z*, Observed *m*/*z* by MALDI-TOF Mass Spectrometry, and Single Species Apparent Molecular Weigtht by Sedimentation Equilibrium Experiments for the GB1-Based Proteins. S11

Experimental Section S13

References S18

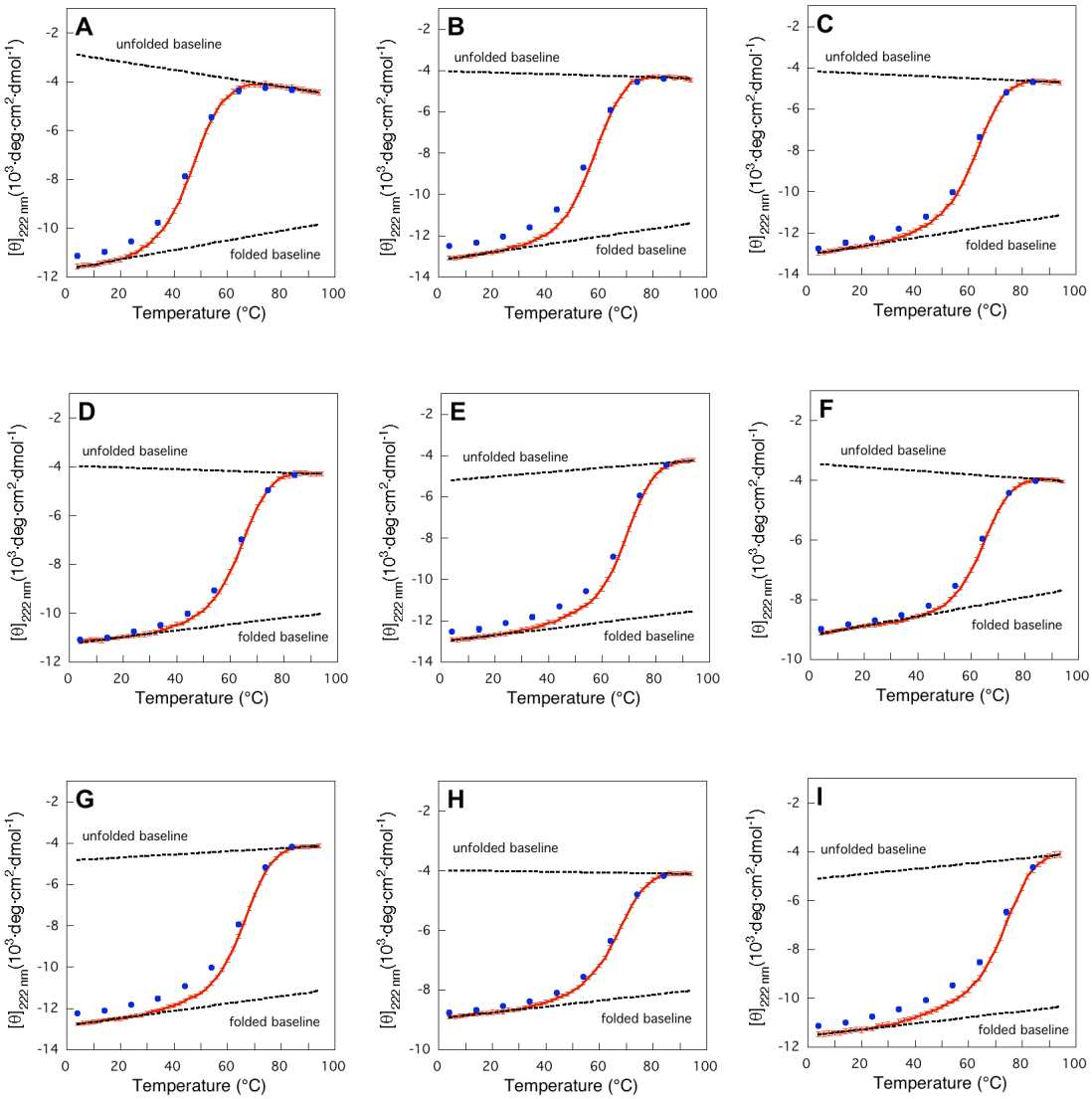

Figure S1. Thermal denaturation of GB1-based proteins monitored by CD at 222 nm. Since the folding and unfolding of the GB1 structure is highly cooperative,[1] the thermodynamic unfolding of the helix (monitored at 222 nm) has been used to represent the unfolding of the overall GB1 structure and thus β-sheet.[2] The red curves are the data collected upon raising the temperature to unfold the proteins, whereas the blue data points are data collected during the subsequent cooling to refold the proteins. The thermal unfolding and folding processes for all the proteins appear to be reversible, indicative of thermal equilibrium during the experiment with minimal protein decomposition to enable thermodynamic analysis. The pre-transition folded baseline and post-transition unfolded baseline were obtained by linear extrapolation following published procedures.[3-5] Panel A, GB1-Gly; panel B, GB1-Ala; panel C, GB1-Abu; panel D, GB1-Leu; panel E, GB1-Phe; panel F, GB1-Atb; panel G, GB1-Qfl; panel H, GB1-Hfl; panel I, GB1-Pff.

Table S1. ^{1}H Chemical Shift Sequence Specific Assignments for 12 Strand Residues Near the Guest Site of GB1-Ala.

Residue	HN	HC_{α}	HC_{β}	Remaining side chain proton signals
Thr51	7.366	5.390	3.756	0.993
Phe52	10.123	5.482	3.213; 3.116	7.100, 7.056, 6.751
Ala53	7.764	3.924	0.993	
Val54	9.129	5.278	1.917	1.591, 1.334
Lys4	9.141	5.114	4.861; 4.134	3.296, 3.085, 1.665, 1.429
Leu5	7.977	5.161	2.290	2.068, 1.929
Ala6	9.354	4.930	1.341	
Leu7	9.190	5.275	3.778	1.828, 1.229
Trp43	9.084	5.353	3.406; 2.718	10.569, 7.707, 7.559, 7.517, 7.329, 7.244
Ala44	7.189	3.117	0.573	
Tyr45	8.668	4.950	2.839; 2.504	7.098, 6.694
Asp46	7.687	4.565	2.643; 2.294	

Table S2. ^{1}H Chemical Shift Sequence Specific Assignments for 12 Strand Residues Near the Guest Site of GB1-Atb.

Residue	HN	HC_{α}	HC_{β}	Remaining side chain proton signals
Thr51	7.344	5.403	3.726	0.932
Phe52	10.030	5.403	3.186, 3.020	7.118, 7.048, 6.787
Atb53	9.082	5.353	3.381, 2.708	
Val54	8.781	5.022	2.858	2.612, 2.014
Lys4	9.155	5.239	2.851, 2.801	1.927, 1.586, 1.419, 1.323
Leu5	8.464	5.349	2.307	2.175, 2.061, 1.926
Ala6	9.349	5.501	2.589	
Leu7	7.984	5.169	2.293	2.068, 1.933
Trp43	9.223	4.895	3.253, 3.134	10.553, 7.752, 7.543, 7.407, 7.153, 7.054
Ala44	9.349	4.924	1.338	
Tyr45	8.760	4.934	2.813, 2.529	7.125, 6.329
Asp46	7.742	4.563	2.639, 2.281	

Table S3. ^{1}H Chemical Shift Sequence Specific Assignments for 12 Strand Residues Near the Guest Site of GB1-Qfl.

Residue	HN	HC_{α}	HC_{β}	Remaining side chain proton signals
Thr51	7.353	5.362	3.719	0.924
Phe52	9.965	5.298	3.155, 3.033	7.129, 7.047, 6.619
Qfl53	9.252	5.076	1.892	2.160, 6.069, 5.997, 5.939, 5.864
Val54	8.446	5.375	2.152	1.930, 1.866
Lys4	9.150	5.264	2.853, 2.801	1.948, 1.896, 1.580, 1.413, 1.325
Leu5	7.971	5.157		2.052, 1.925, 0.117
Ala6	9.339	4.924	1.349	
Leu7	9.065	5.347	3.389	2.715, 1.419
Trp43	9.196	5.019	3.284, 3.122	10.5, 7.636, 7.513, 7.471, 7.444, 7.210
Ala44	7.193	3.139	0.609	
Tyr45	8.643	4.980	2.801, 2.503	7.760, 7.134
Asp46	7.776	4.417	3.113, 2.934	

Table S4. ^{1}H Chemical Shift Sequence Specific Assignments for 12 Strand Residues Near the Guest Site of GB1-Hfl.

Residue	HN	HC_{α}	HC_{β}	Remaining side chain proton signals
Thr51	7.367	5.362	3.720	0.913
Phe52	10.006	5.308	3.158, 3.036	7.128, 7.051, 6.541
Hfl53	9.341	5.144	2.144, 2.077	2.821
Val54	8.774	4.993	2.883	1.587
Lys4	9.172	5.265	2.856, 2.803	1.952, 1.909, 1.596, 1.584, 1.418, 1.335
Leu5	7.988	5.170	2.275	2.064, 1.934, 0.863
Ala6	9.358	4.936	1.349	
Leu7	8.452	5.350	2.166	1.938, 1.866, 1.350
Trp43	9.227	4.988	3.282, 3.132	10.510, 7.578, 7.519, 7.432, 7.212, 7.171
Ala44	9.012	4.839	1.416	
Tyr45	9.081	5.349	3.393, 2.722	7.124, 6.613
Asp46	7.832	4.182	2.051, 1.395	

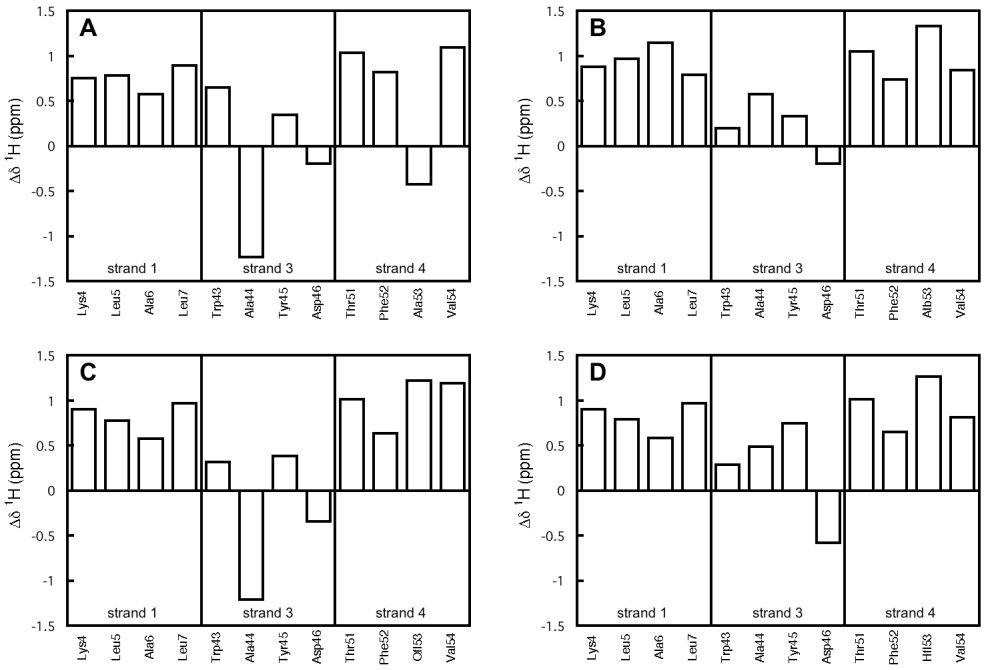

Figure S2. Chemical shift deviation ($\Delta\delta^1$H) of HCα from random coil values[6,7] for the residues surrounding the solvent-exposed guest position 53 for proteins GB1-Ala (panel A), GB1-Atb (panel B), GB1-Qfl (panel C), and GB1-Hfl (panel D). The guest position 53 is on internal strand 4. The cross strand position 6 is on internal strand 1 and position 44 is on edge strand 3 (Figure S1). Significant downfield chemical shift deviations (greater than 0.1 ppm) are consistent with sheet structures.[6,7] The chemical shift deviations for internal strands 1 and 4 are clearly consistent with stable β-sheet structure. There is variation in the chemical shift deviations for edge strand 3, suggesting a less stable structure as expected for edge strands due to more exposure to solvent compared to internal strands or perhaps different contribution of ring current effects from the side chain of Trp43,[8] which may have caused the difference in CD signal. It appears that GB1-Atb and GB1-Hfl have more β-strand structure for edge strand 3 compared to GB1-Ala and GB1-Qfl. The trend for edge strand 3 stability of these four proteins (GB1-Hfl ~ GB1-Atb > GB1-Qfl ~ GB1-Ala) is similar to the trend for the overall stability (GB1-Hfl > GB1-Atb ~ GB1-Qfl > GB1-Ala), but not identical. Nonetheless, sequential HCα(i)-HN(i+1) NOEs (indicative of stable sheet structure) have been observed edge strand 3 for all four mutant proteins (Figure S3).

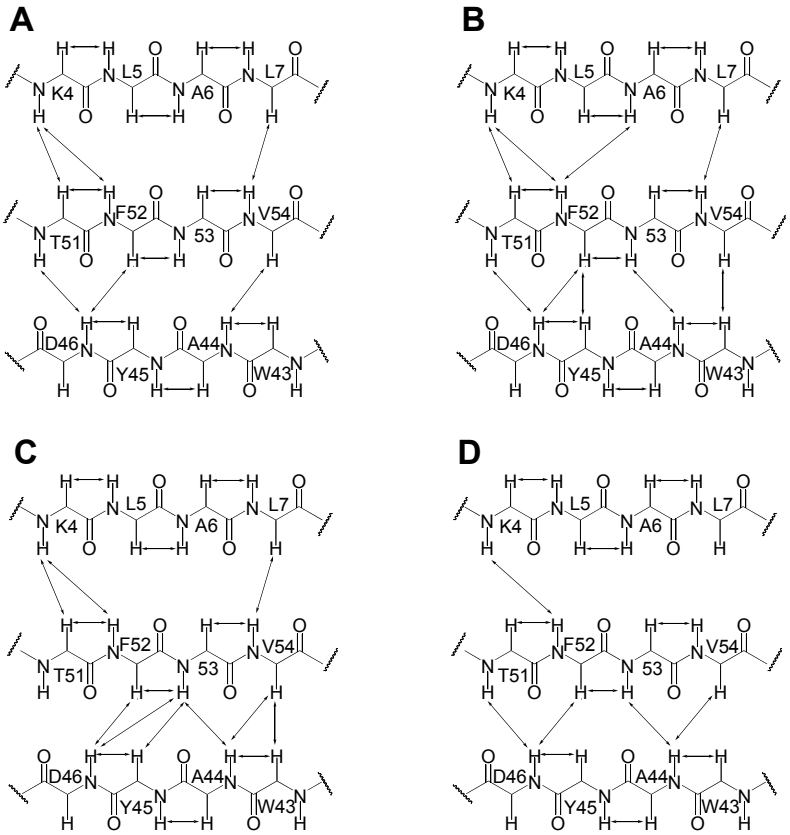

Figure S3. Schematic diagram depicting intra- and inter-strand NOE cross peaks involving backbone HN and HCα protons for residues near the guest position 53 of proteins GB1-Ala (panel A), GB1-Atb (panel B), GB1-Qfl (panel C), and GB1-Hfl (panel D). Three strands for each protein are shown: internal strand 1 (residues K4-L7), edge strand 3 (residues D46-W43), and internal strand 4 (residues T51-V54). Sequential HCα(*i*)-HN(*i*+*1*) connectivities indicative of β-strand formation[9] were observed for the depicted residues for all four proteins. Interstrand connectivies consistent with wild type sheet structure[1] were also observed for all four proteins.

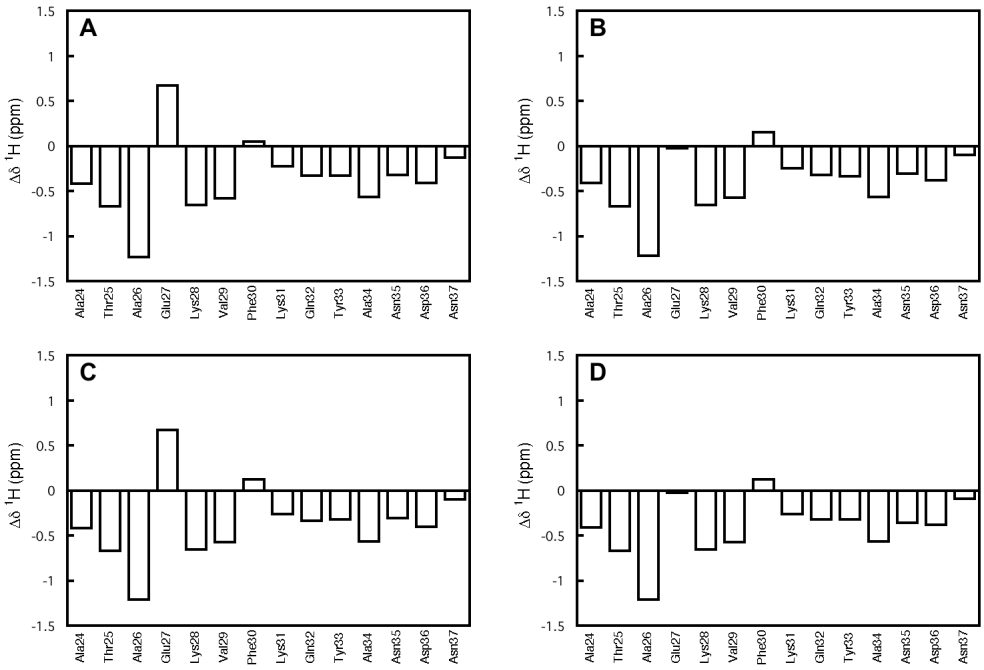

Figure S4. Chemical shift deviation ($\Delta\delta^1$H) of HCα from random coil values[6,7] for the helical region (Ala24-Asn37) for proteins GB1-Ala (panel A), GB1-Atb (panel B), GB1-Qfl (panel C), and GB1-Hfl (panel D). Significant upfield chemical shift deviations (greater than 0.1 ppm) are consistent with helical structures.[6,7] Several NOE patterns are indicative of helix formation, including HN(*i*)-HN(*i+1*) NOEs and HCα(*i*)-HCβ(*i+3*) NOEs.[9] All four proteins show 13 sequential HN(*i*)-HN(*i+1*) NOEs in this helical region. GB1-Ala and GB1-Qfl both show 8 HCα(*i*)-HCβ(*i+3*) NOEs, whereas GB1-Atb and GB1-Hfl both show 7 HCα(*i*)-HCβ(*i+3*) NOEs along with one ambiguous NOE due to spectral overlap.

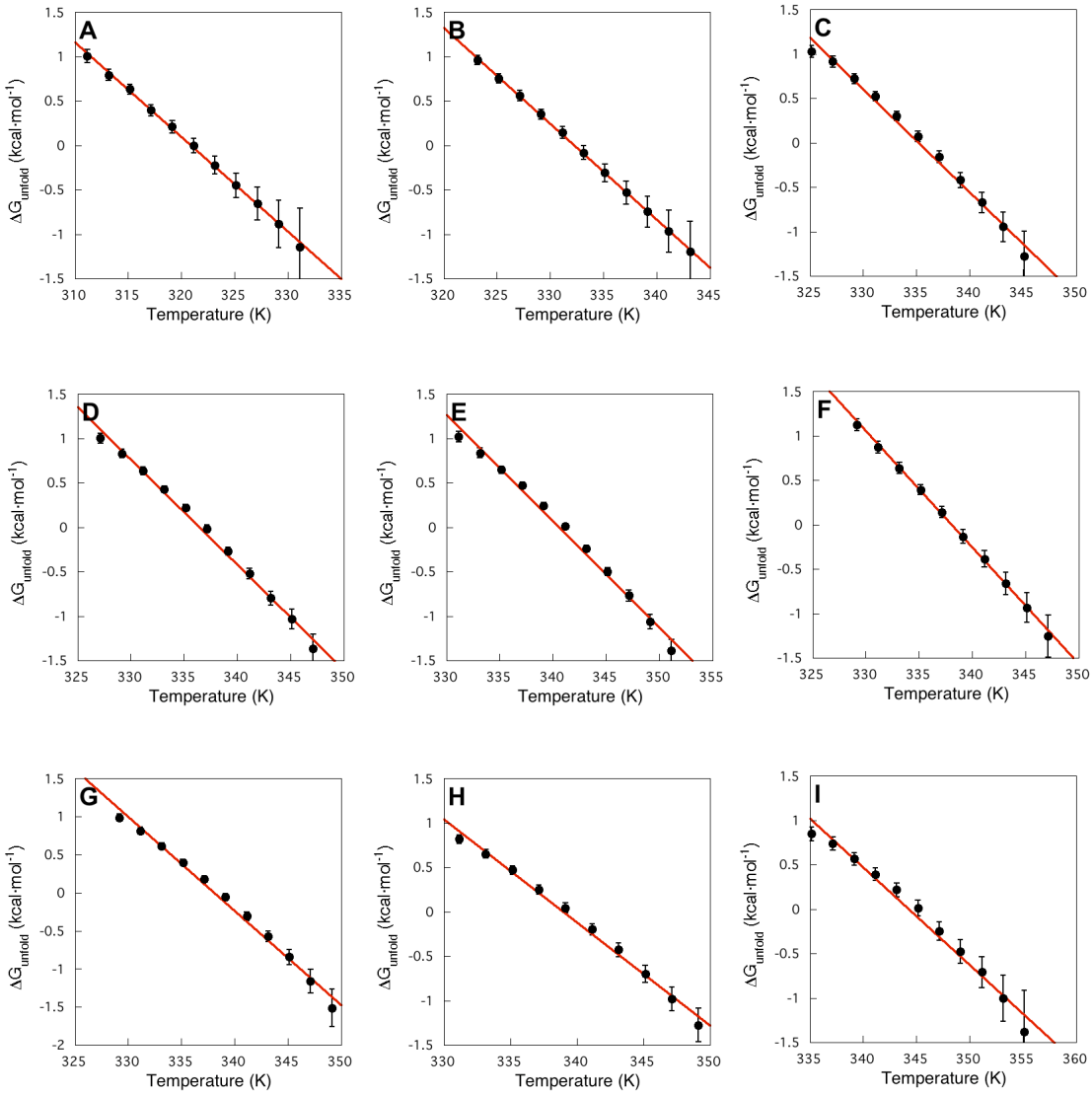

Figure S5. The unfolding energy (ΔG_{unfold}) of the GB1-based proteins plotted against temperature near the T_m for each protein along with the corresponding linear fits. The data points are plotted as black filled circles with error bars. The van't Hoff analysis by fitting to the equation $\Delta G_{unfold}=\Delta H_{Tm}-T\cdot\Delta S_{Tm}$ are shown as red curves. The resulting thermodynamic parameters ΔS_{Tm}, ΔH_{Tm}, and T_m are listed in Table S5. Panel A, GB1-Gly; panel B, GB1-Ala; panel C, GB1-Abu; panel D, GB1-Leu; panel E, GB1-Phe; panel F, GB1-Atb; panel G, GB1-Qfl; panel H, GB1-Hfl; panel I, GB1-Pff.

Table S5. ΔH_{Tm}, ΔS_{Tm}, T_m, and $\Delta G_{unfold\ 60°C}$ for Unfolding the GB1-Based Proteins.

Protein	ΔH_{Tm}[a] (kcal·mol^{-1})	ΔS_{Tm}[a] (cal·mol^{-1}·K^{-1})	T_m[b] (°C)	$\Delta G_{unfold\ 60°C}$[c] (kcal·mol^{-1})
GB1-Gly	34.2 ± 0.5	107 ± 2	47.7 ± 0.2	-1.29 ± 0.33
GB1-Ala	36.5 ± 0.4	110 ± 1	59.1 ± 0.3	-0.086 ± 0.132
GB1-Abu	40.2 ± 0.6	120 ± 2	62.1 ± 0.6	0.301 ± 0.139
GB1-Leu	39.0 ± 1.0	116 ± 3	63.3 ± 0.7	0.427 ± 0.083
GB1-Phe	42.3 ± 0.7	124 ± 2	67.5 ± 0.6	0.988 ± 0.151
GB1-Atb	43.9 ± 0.4	130 ± 1	64.9 ± 0.9	0.651 ± 0.078
GB1-Qfl	40.1 ± 0.9	118 ± 3	64.9 ± 1.1	0.636 ± 0.062
GB1-Hfl	39.3 ± 0.8	116 ± 2	65.9 ± 0.9	0.720 ± 0.075
GB1-Pff	38.7 ± 0.8	112 ± 3	71.1 ± 0.8	1.32 ± 0.24

[a]Obtained from van't Hoff analysis of the thermal denaturation data near $\Delta G_{unfold}=0$ in Figure S5 by linearly fitting to $\Delta G=\Delta H_{Tm}-T\cdot\Delta S_{Tm}$.

[b]Derived from $T_m=(\Delta H_{Tm}/\Delta S_{Tm})$; T_m is the temperature at which $\Delta G_{unfold}=0$.

[c]Derived from $\Delta G_{unfold\ 60°C}= \Delta H_{Tm}-333.15\cdot\Delta S_{Tm}$; 60°C = 333.15K.

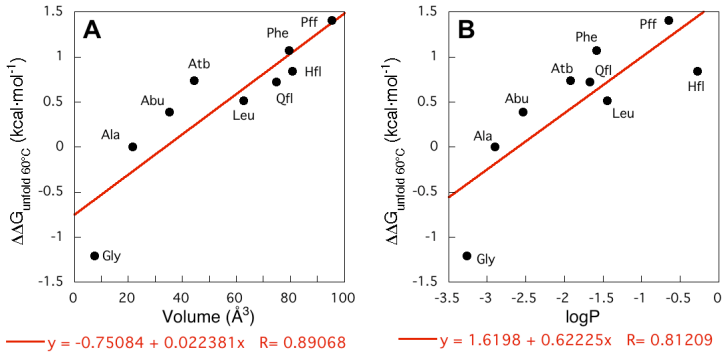

Figure S6. The relative unfolding free energy at 60°C ($\Delta\Delta G_{unfold\ 60°C}$) for the GB1-based proteins plotted against side chain volume (panel A) and the hydrophobic parameter logP (panel B). The side chain volumes were measured based on the energy minimized structures using the CHARMM forcefield[10] in Discovery Studio 2.1 (Accelrys, San Diego, CA). The parameter logP was derived experimentally using thin layer chromatography following published procedures.[11] The linear fit for side chain volume is only marginally better compared to the hydrophobic parameter logP, suggesting that size,[12,13] hydrophobicity,[14] or both[15,16] may be important factors for the observed stabilization.

Table S6. The Formula, Calculated *m*/*z*, Observed *m*/*z* by MALDI-TOF Mass Spectrometry, and Single Species Apparent Molecular Weight by Sedimentation Equilibrium Experiments for the GB1-Based Proteins.

Protein	Formula	Calc $[MH]^+$	Observed[a] $[MH]^+$	Observed[b] MW
GB1-Gly	$C_{269}H_{415}N_{69}O_{92}S$	6119.645	6119.029	5800 ± 180
GB1-Ala	$C_{270}H_{417}N_{69}O_{92}S$	6133.672	6133.826	6130 ± 530
GB1-Abu	$C_{271}H_{419}N_{69}O_{92}S$	6147.699	6147.584	6157 ± 280
GB1-Leu	$C_{273}H_{423}N_{69}O_{92}S$	6175.752	6175.854	5600 ± 200
GB1-Phe	$C_{276}H_{421}N_{69}O_{92}S$	6209.768	6210.423	6000 ± 200
GB1-Atb	$C_{271}H_{416}F_3N_{69}O_{92}S$	6201.670	6201.349	4700 ± 200
GB1-Qfl	$C_{273}H_{419}F_4N_{69}O_{92}S$	6247.714	6247.206	6900 ± 200
GB1-Hfl	$C_{273}H_{417}F_6N_{69}O_{92}S$	6280.970	6281.112	5780 ± 200
GB1-Pff	$C_{276}H_{416}F_5N_{69}O_{92}S$	6299.720	6299.330	6100 ± 170

[a]Obtained by MALDI-TOF mass spectrometry.
[b]Obtained by sedimentation equilibrium experiments using analytical ultracentrifugation.

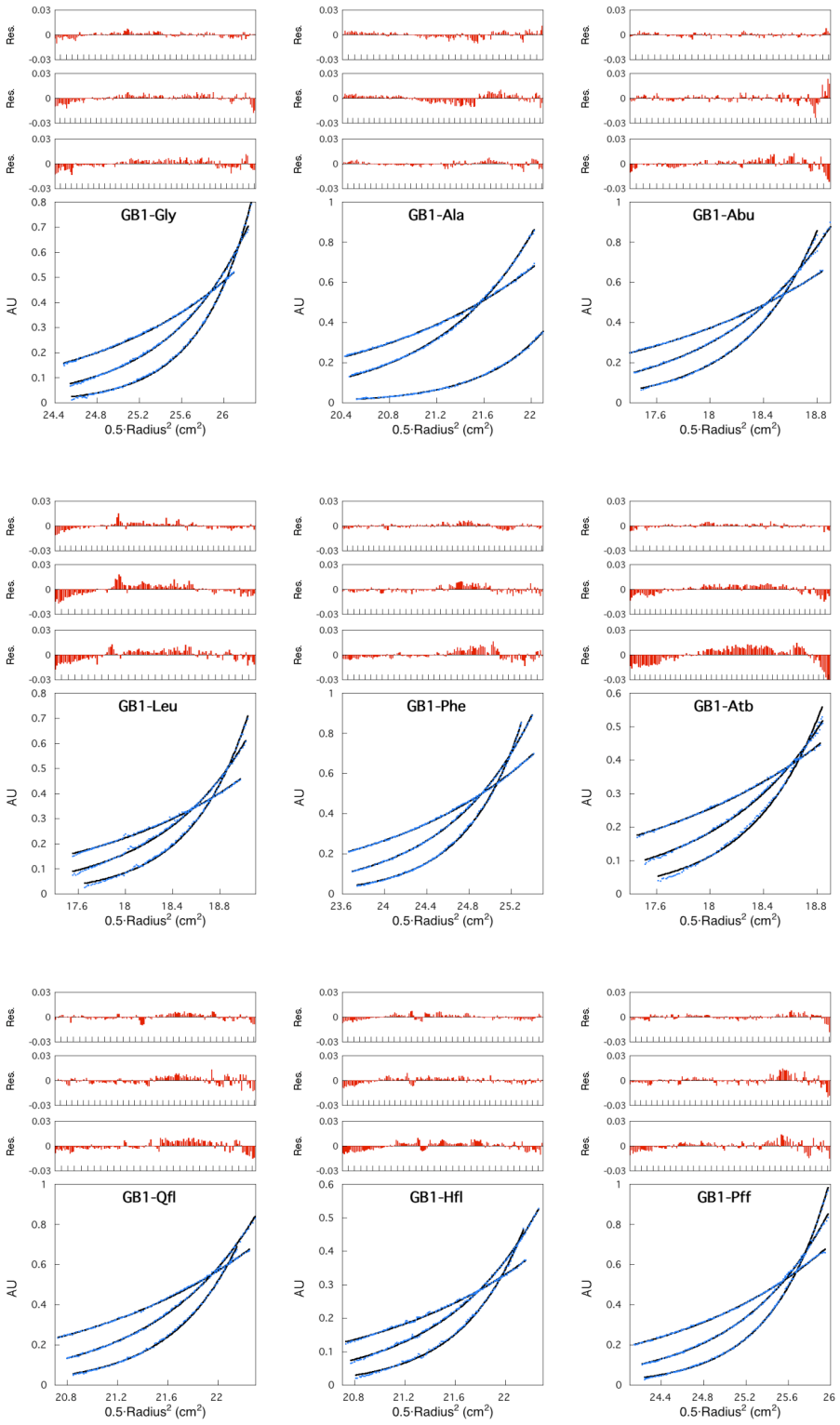

Figure S7. Analytical ultracentrifuge data from sedimentation equilibrium experiments for the nine GB1-based proteins (blue points). The curve fits using the monomer molecular weight based on amino acid composition of the peptides are shown in black. The residuals for the fits at the different rotation speeds (30k, 40k, and 50k rpm) are shown in red.

Experimental Section

General Methods and Reagents

All of the chemical reagents except those indicated otherwise were purchased from Aldrich. Organic and high performance liquid chromatography (HPLC) solvents were from EMD Science. *N*-9-Fluorenylmethoxycarbonyl (Fmoc)-amino acids, 1-hydroxybenzotrazole (HOBt), *O*-1*H*-benzotriazol-1-yl-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) were from NovaBiochem, Fmoc-L-pentafluorophenylalanine (Fmoc-Pff-OH) was from Synthetech, and Fmoc-PAL-PEG-PS resin was from Applied Biosystems. Reagents and solvents were used without further purification. (*S*)-Fmoc-5,5,5,5',5',5'-hexafluoroleucine (Fmoc-Hfl-OH) and (*S*)-Fmoc-5,5,5',5'-tetrafluoroleucine (Fmoc-Qfl-OH) were synthesized following our chemoenzymatic route.[17,18] (*S*)-Fmoc-2-amino-4,4,4-trifluorobutyric acid (Fmoc-Atb-OH) was obtained by enzymatic resolution following published procedures.[19,20] Analytical reverse phase (RP)-HPLC was performed on an Agilent 1100 series chromatography system using a Vydac C_{18} column (4.6 mm diameter, 250 mm length). Preparative RP-HPLC was performed on a Waters Breeze chromatography system using a Vydac RP C_{18} column (22 mm diameter, 250 mm length). The association state of peptides was determined by sedimentation equilibrium. Mass spectrometry of the proteins was performed on a matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) spectrometer (Bruker Daltonics Bioflex IV) using α-cyano-4-hydroxycinnamic acid as the matrix. Determination of peptide concentration was performed on a UV-Vis spectrophotometer (Agilent 8453). Circular dichroism (CD) spectra were collected on a Jasco J715 spectrometer.

Protein Synthesis by Solid Phase Peptide Synthesis Based on Fmoc Chemistry

Fmoc-PAL-PEG-PS (50 μmol) was swollen in *N,N*-dimethylformamide (DMF, 5 mL) for 30 minutes before the first coupling. The resin was then washed with DMF (5 mL, 5x1.5 min). This was followed by Fmoc deprotection with 20% piperidine/DMF (5 mL, 3x8 min). The resin was subsequently washed with DMF (5 mL, 5x1.5 min). A mixture of 5 equivalents of the appropriately protected Fmoc amino acid, HOBt, and HBTU was dissolved in DMF (1 mL). Diisopropylethylamine (DIEA, 20 equivalents) was then added to the solution. The solution was then mixed thoroughly and applied to the resin. The vial that contained the solution was rinsed with DMF (2x1 mL), and added to the reaction. The first amino acid was coupled for 8 hours. The rest of the residues that were attached to the resin were coupled for 4.75 hrs. Proteins GB1-Ala, GB1-Leu, GB1-Phe, GB1-Gly, and GB1-Abu were synthesized using an automated peptide synthesizer (Advanced ChemTech Apex 396). For proteins of GB1-Hfl, GB1-Qfl, GB1-Pff, and GB1-Atb, the first 7 residues (50-56) were coupled manually, and the rest of the protein was completed using the peptide synthesizer. To remove the Fmoc group of residue 51 on GB1-Hfl,

2% DBU/DMF (5 mL, 3x8 min) was used. After coupling the last residue for all proteins, the resin was washed with DMF (5 mL, 5x1.5 min) followed by Fmoc deprotection using 20% piperidine/DMF (5 mL, 3x8 min). Since the amino group of N-terminal residue Met appears to form a salt bridge to the carboxylic acid side chain of Glu19 in the crystal structure,[1] the N-terminus was left uncapped for all proteins. The resin was subsequently washed with DMF (5 mL, 5x1.5 min) and CH_2Cl_2 (5 mL, 5x0.5 min) and lyophilized overnight.

The protein was deprotected and cleaved off the resin by treating the resin with 90:5:5 trifluoroacetic acid (TFA)/triisopropylsilane/ethanedithiol (10 mL) for 2 hrs. The reaction was then filtered through glass wool and the resin was washed with TFA (3x3 mL). The combined filtrate was then evaporated by a gentle stream of N_2. The resulting oil was washed with hexanes, dissolved in water and lyophilized. The proteins (1 mg mL^{-1} aqueous solution) were analyzed using analytical RP-HPLC on a 25 cm C_{18} column (dia 4.6 mm) using 1 mL/min flow rate, linear 1%/min gradient from 100% A to 0% A (solvent A: 99.9% water, 0.1% TFA; solvent B: 90% acetonitrile, 10% water, 0.1% TFA.). Appropriate linear solvent A/ solvent B gradients were used for purification on a RP-HPLC preparative C_{18} column. The identity of the proteins was confirmed by MALDI-TOF (Table S5). The monomeric association state of proteins was confirmed using sedimentation equilibrium by analytical ultracentrifugation.

GB1-Ala

(MetGlnTyrLysLeuAlaLeuAsnGlyLysThrLeuLysGlyGluThrThrThrGluAlaValAspAlaAlaThr-AlaGluLysValPheLysGlnTyrAlaAsnAspAsnGlyValGluGlyGluTrpAlaTyrAspAspAlaThrLysThr-PheAlaValThrGlu-NH_2). The protein was synthesized using 274.3 mg (0.046 mmol) of Fmoc-PAL-PEG-PS resin. The synthesis gave 475.3 mg of resin (49.9% yield). The cleavage yielded 121.7 mg of crude protein (77.8% yield, 11.7% purity). The peptide was purified by preparative RP-HPLC using a C_{18} column to 98.8% purity. Retention time on analytical RP-HPLC was 39.4 min. The identity of the protein was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{270}H_{417}N_{69}O_{92}S$ $[MH]^+$: 6133.672; observed: 6133.826.

GB1-Gly

(MetGlnTyrLysLeuAlaLeuAsnGlyLysThrLeuLysGlyGluThrThrThrGluAlaValAspAlaAlaThr-AlaGluLysValPheLysGlnTyrAlaAsnAspAsnGlyValGluGlyGluTrpAlaTyrAspAspAlaThrLysThr-PheGlyValThrGlu-NH_2). The protein was synthesized using 281.8 mg (0.05 mmol) of Fmoc-PAL-PEG-PS resin. The synthesis gave 436.8 mg of resin (38.0% yield). The cleavage yielded 115.0 mg of crude protein (99.0% yield, 8.5% purity). The peptide was purified by preparative RP-HPLC using a C_{18} column to 96.4% purity. Retention time on analytical RP-HPLC was 39.3 min. The identity of the protein was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{269}H_{415}N_{69}O_{92}S$ $[MH]^+$: 6119.645; observed: 6119.029.

GB1-Leu

(MetGlnTyrLysLeuAlaLeuAsnGlyLysThrLeuLysGlyGluThrThrThrGluAlaValAspAlaAlaThr-AlaGluLysValPheLysGlnTyrAlaAsnAspAsnGlyValGluGlyGluTrpAlaTyrAspAspAlaThrLysThr-PheLeuValThrGlu-NH_2). The protein was synthesized using 279.8 mg (0.05 mmol) of Fmoc-PAL-PEG-PS resin. The synthesis gave 495.3 mg of resin (50.6% yield). The cleavage yielded 156.2 mg of crude protein (97.8% yield, 6.8% purity). The peptide was purified by preparative RP-HPLC using a C_{18} column to 96.3% purity. Retention time on analytical RP-HPLC was 40.4 min. The identity of the protein was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{273}H_{423}N_{69}O_{92}S$ $[MH]^+$: 6175.752; observed: 6175.854.

GB1-Phe

(MetGlnTyrLysLeuAlaLeuAsnGlyLysThrLeuLysGlyGluThrThrThrGluAlaValAspAlaAlaThr-AlaGluLysValPheLysGlnTyrAlaAsnAspAsnGlyValGluGlyGluTrpAlaTyrAspAspAlaThrLysThr-PhePheValThrGlu-NH_2). The protein was synthesized using 295.0 mg (0.05 mmol) of Fmoc-PAL-PEG-PS resin. The synthesis gave 515.6 mg of resin (49.8% yield). The cleavage yielded 119.4 mg of crude protein (69.5% yield, 10.5% purity). The peptide was purified by preparative RP-HPLC using a C_{18} column to 99.8% purity. Retention time on analytical RP-HPLC was 40.5 min. The identity of the protein was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{276}H_{421}N_{69}O_{92}S$ $[MH]^+$: 6209.768; observed: 6210.423.

GB1-Abu

(MetGlnTyrLysLeuAlaLeuAsnGlyLysThrLeuLysGlyGluThrThrThrGluAlaValAspAlaAlaThr-AlaGluLysValPheLysGlnTyrAlaAsnAspAsnGlyValGluGlyGluTrpAlaTyrAspAspAlaThrLysThr-PheAbuValThrGlu-NH_2). The protein was synthesized using 294.6 mg (0.05 mmol) of Fmoc-PAL-PEG-PS resin. The synthesis gave 472.5 mg of resin (40.6% yield). The cleavage yielded 139.5 mg of crude protein (99.4% yield, 11.2% purity). The peptide was purified by preparative RP-HPLC using a C_{18} column to 99.9% purity. Retention time on analytical RP-HPLC was 39.4 min. The identity of the protein was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{271}H_{419}N_{69}O_{92}S$ $[MH]^+$: 6147.699; observed: 6147.584.

GB1-Atb

(MetGlnTyrLysLeuAlaLeuAsnGlyLysThrLeuLysGlyGluThrThrThrGluAlaValAspAlaAlaThr-AlaGluLysValPheLysGlnTyrAlaAsnAspAsnGlyValGluGlyGluTrpAlaTyrAspAspAlaThrLysThr-PheAtbValThrGlu-NH_2). The protein was synthesized using 544.6 mg (0.1 mmol) of Fmoc-PAL-PEG-PS resin. The synthesis gave 777.9 mg of resin (25.8% yield). The cleavage yielded 280.5 mg of crude protein (>99% yield, 6.6% purity). The peptide was purified by preparative RP-HPLC using a C_{18} column to 99.4% purity. Retention time on analytical RP-HPLC was 39.3 min.The identity of the protein was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{271}H_{416}F_3N_{69}O_{92}S$ $[MH]^+$: 6201.670; observed: 6201.349.

GB1-Hfl

(MetGlnTyrLysLeuAlaLeuAsnGlyLysThrLeuLysGlyGluThrThrThrGluAlaValAspAlaAlaThr-AlaGluLysValPheLysGlnTyrAlaAsnAspAsnGlyValGluGlyGluTrpAlaTyrAspAspAlaThrLysThr-PheHflValThrGlu-NH_2). The protein was synthesized using 544.6 mg (0.1 mmol) of Fmoc-PAL-PEG-PS resin. The synthesis gave 1067.2 mg of resin (57.2% yield). The cleavage yielded 380.0 mg of crude protein (95.4% yield, 7.1% purity). The peptide was purified by preparative RP-HPLC using a C_{18} column to 99.3% purity. Retention time on analytical RP-HPLC was 40.4 min. The identity of the protein was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{273}H_{417}F_6N_{69}O_{92}S$ $[MH]^+$: 6280.970; observed: 6281.112.

GB1-Pff

(MetGlnTyrLysLeuAlaLeuAsnGlyLysThrLeuLysGlyGluThrThrThrGluAlaValAspAlaAlaThr-AlaGluLysValPheLysGlnTyrAlaAsnAspAsnGlyValGluGlyGluTrpAlaTyrAspAspAlaThrLysThr-PhePffValThrGlu-NH_2). The protein was synthesized using 294.9 mg (0.05 mmol) of Fmoc-PAL-PEG-PS resin. The synthesis gave 477.2 mg of resin (40.8% yield). The cleavage yielded 138.3 mg of crude protein (97.1% yield, 12.4% purity). The peptide was purified by preparative RP-HPLC using a C_{18} column to 97.6% purity. Retention time on analytical RP-HPLC was 41.3 min. The identity of the protein was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{276}H_{416}F_5N_{69}O_{92}S$ $[MH]^+$: 6299.720; observed: 6299.330.

GB1-Qfl

(MetGlnTyrLysLeuAlaLeuAsnGlyLysThrLeuLysGlyGluThrThrThrGluAlaValAspAlaAlaThr-AlaGluLysValPheLysGlnTyrAlaAsnAspAsnGlyValGluGlyGluTrpAlaTyrAspAspAlaThrLysThr-PheQflValThrGlu-NH_2). The protein was synthesized using 293.3 mg (0.050 mmol) of Fmoc-PAL-PEG-PS resin. The synthesis gave 484.0 mg of resin (43.0% yield). The cleavage yielded 128.3 mg of crude protein (86.1% yield, 10.6% purity). The peptide was purified by preparative RP-HPLC using a C_{18} column to 98.0% purity. Retention time on analytical RP-HPLC was 40.5 min. The identity of the protein was confirmed by MALDI-TOF mass spectrometry. Calculated for $C_{273}H_{419}F_4N_{69}O_{92}S$ $[MH]^+$: 6247.714; observed: 6247.206.

Sedimentation Equilibrium Experiments by Analytical Ultracentrifugation

Proteins GB1-Ala, GB1-Abu, GB1-Atb, GB1-Qfl, GB1-Phe, and GB1-Pff were dissolved in 120 μL buffer of 10 mM MOPS ((3-*N*-morpholino)- propanesulfonic acid), 150 mM NaCl, pH 7.0 to give an absorbance between 0.2 and 0.4 at 275 nm. Samples were spun at 30, 40, and 50k rpm at 20°C with 8 hours of equilibration time for each spin speed prior to data collection. Proteins GB1-Gly, GB1-Leu, and GB1-Hfl were dissolved in 120 μL buffer of 10 mM MOPS, 50 mM NaCl, pH 7.0 to give an absorbance between 0.2 and 0.4 at 275 nm. Samples were spun at 30, 40, and 50k rpm at 4°C with 8 hours of equilibration time for each spin speed prior to data

collection. The data were truncated using the program WinReedit (version 0.999, 1998) and analyzed using WinNonLin (version 1.035, 1997).[21] For the analysis, the solution density and the peptide partial specific volume were calculated by using Sednterp as described by Laue *et al.*[22] The partial specific volumes for modified amino acids was calculated by using a procedure described by Durschschlag *et al*.[23] Single species analysis for all proteins suggested a molecular weight consistent with a monomer (Table S6). Theoretical curves generated using monomer molecular weights based on composition fit the experimental data reasonably well (Figure S7).

Circular Dichroism Spectroscopy

CD data was collected using a 1 mm pathlength cell. The concentration of peptide stock solution were determined by the tyrosine and tryptophan absorbance in 6 M guanidinium chloride (ε_{276}=9770, ε_{278}=9750, ε_{280}=9540, ε_{282}=9295).[24,25] Thermal denaturation of the proteins was performed at protein concentrations 0.5 mg/mL in 50 mM sodium acetate buffer (pH 5.2)[8,26] from 4 °C to 94 °C (98 °C for GB1-Pff) and monitored by the ellipticity of the proteins at 222 nm^2 on a Jasco J715 spectrometer. Each reported CD value was the mean of at least 60 determinations. Two-state transition model was used to analyze the thermal denaturation curves. Linear extrapolation of the pre-transition folded and post-transition unfolded baselines[3-5] was performed using the curve-fitting program Kaleidagraph 3.52 (Synergy Software CA). The fraction native (F_n) at each temperature was calculated using the equation $F_n = (\theta_{obs} - \theta_u)/(\theta_f - \theta_u)$, where θ_{obs} is the observed ellipticity at 222 nm, and θ_f and θ_u are the folded and unfolded ellipticities derived from the extrapolated baselines. The fraction unfolded (F_u) is given by F_u=1-F_n. The unfolding equilibrium constant (K_u) for various temperatures was calculated by $K_u=F_u/F_n$. The unfolding free energy (ΔG_{unfold}) for temperatures near T_m (at which 50% of the protein is unfolded, i.e. ΔG_{unfold}=0) was calculated by $\Delta G_{unfold} = -RT\cdot\ln(K_u)$. ΔG_{unfold} was plotted against temperature, and van't Hoff analysis was performed on data near T_m by linearly fitting to $\Delta G_{unfold}=\Delta H_{Tm}-T\cdot\Delta S_{Tm}$ (Figure S5), to determine the T_m, and van't Hoff enthalpy (ΔH_{Tm}) and entropy (ΔS_{Tm}) (Table S5). The values of ΔH_{Tm} and ΔS_{Tm} were used to calculate ΔG_{unfold} for each protein at 60 °C (333.15K). The protein mutant with Ala at position 53 was used as the reference state,[8,26] and $\Delta\Delta G_{unfold}$ for each mutant was calculated relative to $\Delta G_{GB1\text{-}Ala}$ using the equation $\Delta\Delta G_{Xaa} = \Delta G_{GB1\text{-}Xaa} - \Delta G_{GB1\text{-}Ala}$.

NMR Spectroscopy

Protein samples for NMR studies were prepared using 2 mg of protein dissolved in a buffer containing 90% 50 mM d_4-acetate, pH 5.2, 10% D_2O. ^{1}H 2D homonuclear spectra TOCSY[27,28] and NOESY[29-31] were acquired on a Varian Inova 750 MHz spectrometer. All spectra were recorded in a phase-sensitive manner using TPPI methods.[32] In all spectra, 16-32

scans for each of the 512 t_1 values were collected with 4K points. Water suppression was achieved by presaturation during 1.0-1.3 s relaxation delay. For the NOESY spectra, low-power presaturation was also used during the 100 ms mixing time for additional water suppression. Data were processed and analyzed using NMR PIPE and CARA (Computer Aided Resonance Assignment) software, respectively. In the acquisition dimension, low–frequency deconvolution was applied to remove the residual water signal.[33] Data were multiplied by a 60^o-shifted squared sine bell or 90^o-shifted sine bell apodization function prior to Fourier transformation. A polynomial baseline correction was applied to the rows of the transformed matrices. In the t_1 dimension, the data were zero filled to 1K points, and a 60^o-shifted squared sine bell or 90^o-shifted sine bell apodization function was applied before Fourier transformation.

References:

(1) Gronenborn, A. M.; Filpula, D. R.; Essig, N. Z.; Achari, A.; Whitlow, M.; Wingfield, P. T.; Clore, G. M. *Science* **1991**, *253*, 657.

(2) Merkel, J. S.; Sturtevant, J. M.; Regan, L. *Structure* **1999**, *7*, 1333.

(3) Pace, C. N.; Shirley, B. A.; Thomson, J. A. In *Protein structure: A practical approach*; Creighton, T. E., Ed.; IRL Press: Oxford, 1990, p 311.

(4) Allen, D. L.; Peilak, G. J. *Protein Sci.* **1998**, *7*, 1262.

(5) John, D. M.; Weeks, K. M. *Protein Science* **2000**, *9*, 1416.

(6) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *J. Mol. Biol.* **1991**, *222*, 311.

(7) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *Biochemistry* **1992**, *31*, 1647.

(8) Minor, D. L., Jr.; Kim, P. S. *Nature* **1994**, *367*, 660.

(9) Wuthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, NY, 1986.

(10) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comp. Chem.* **1983**, *4*, 187.

(11) Pliska, V.; Schmidt, M.; Fauchere, J. L. *J. Chromatogr.* **1981**, *216*, 79.

(12) Street, A. G.; Mayo, S. L. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 9074.

(13) Bai, Y.; Englander, S. W. *Proteins Struct. Func. Genet.* **1994**, *18*, 262.

(14) Yang, A.-S.; Honig, B. *J. Mol. Biol.* **1995**, *252*, 366.

(15) Otzen, D.; Fersht, A. *Biochemistry* **1995**, *34*, 5718.

(16) Koehl, P.; Levitt, M. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 12524.

(17) Chiu, H.-P.; Suzuki, Y.; Gullickson, D.; Ahmad, R.; Kokona, B.; Fairman, R.; Cheng, R. P. *J. Am. Chem. Soc.* **2006**, *128*, 15556.

(18) Chiu, H.-P.; Cheng, R. P. *Org. Lett.* **2007**, *9*, 5517.

(19) Black, S.; Wright, N. G. *J. Biol. Chem.* **1955**, *213*, 39.

(20) Tsushima, T.; Kawada, K.; Ishihara, S.; Uchida, N.; Shiratori, O.; Higaki, J.; Hirata, M.

Tetrahedron **1988**, *44*, 5375.
(21) Johnson, M. L.; Correia, J. J.; Yphantis, D. A.; Halvorson, H. R. *Biophys. J.* **1981**, *36*, 575.
(22) Laue, T. M.; Shah, B.; Ridgeway, T. M.; Pelletier, S. L. In *Analytical Ultracentrifugation in Biochemistry and Biopolymer Science*; Harding, S. E., Rowe, A. J., Horton, J. C., Eds.; Royal Society of Chemistry: Cambridge, 1992, p 90.
(23) Durchschlag, H.; Zipper, P. *Prog. Colloid Polym. Sci.* **1994**, *94*, 20.
(24) Edelhoch, H. *Biochemistry* **1967**, *6*, 1948.
(25) Pace, C. N.; Vajdos, F.; Fee, L.; Grimsley, G.; Gray, T. *Protein Sci.* **1995**, *4*, 2411.
(26) Smith, C. K.; Withka, J. M.; Regan, L. *Biochemistry* **1994**, *33*, 5510.
(27) Braunschweiler, L.; Ernst, R. R. *J. Magn. Reson.* **1983**, *53*, 521.
(28) Bax, A.; Davis, D. G. *J. Magn. Reson.* **1985**, *65*, 355.
(29) Jeener, J.; Meier, B. H.; Bachmann, P.; Ernst, R. R. *J. Chem. Phys.* **1979**, *71*, 4546.
(30) Kumar, A.; Ernst, R. R.; Wuthrich, K. *Biochem. Biophys. Res. Commun.* **1980**, *95*, 1.
(31) States, D. J.; Haberkorn, R. A.; Ruben, D. J. *J. Magn. Reson.* **1982**, *48*.
(32) Marion, D.; Wuthrich, K. *Biochem. Biophys. Res. Commun.* **1983**, *113*, 967.
(33) Marion, D.; Ikura, M.; Bax, A. *J. Magn. Reson.* **1989**, *84*, 425.