**Protein-Like Tertiary Folding Behavior from Heterogeneous Backbones**

Zachary E. Reinert, George A. Lengyel, and W. Seth Horne*

*Department of Chemistry, University of Pittsburgh, Pittsburgh, PA 15260.*

**SUPPORTING INFORMATION**

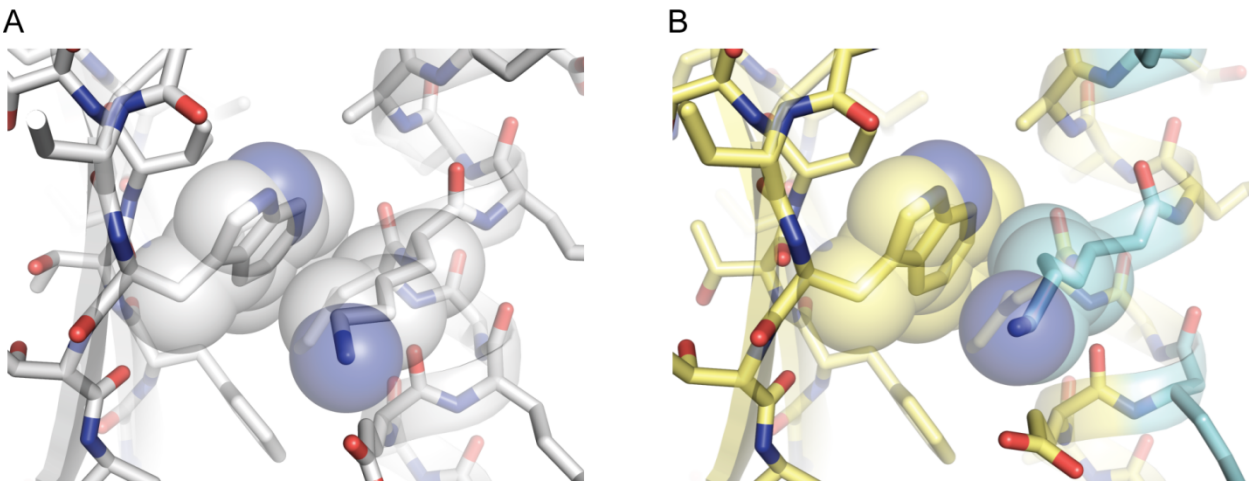


**Figure S1.** Packing interactions involving (A) $Lys_{31}$ in wild-type protein **1** and (B) $\beta^3$-$Lys_{31}$ in helix-modified variant **2**.

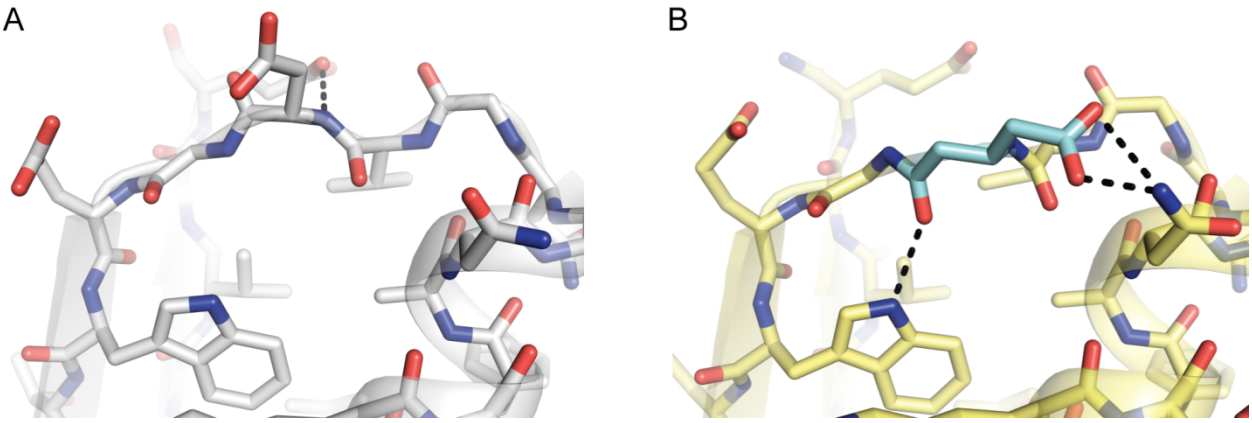


**Figure S2.** Polar contacts involving (A) $Asp_{40}$ in wild-type protein **1** and (B) $\beta^3$-$Asp_{40}$ in loop-modified variant **3**.

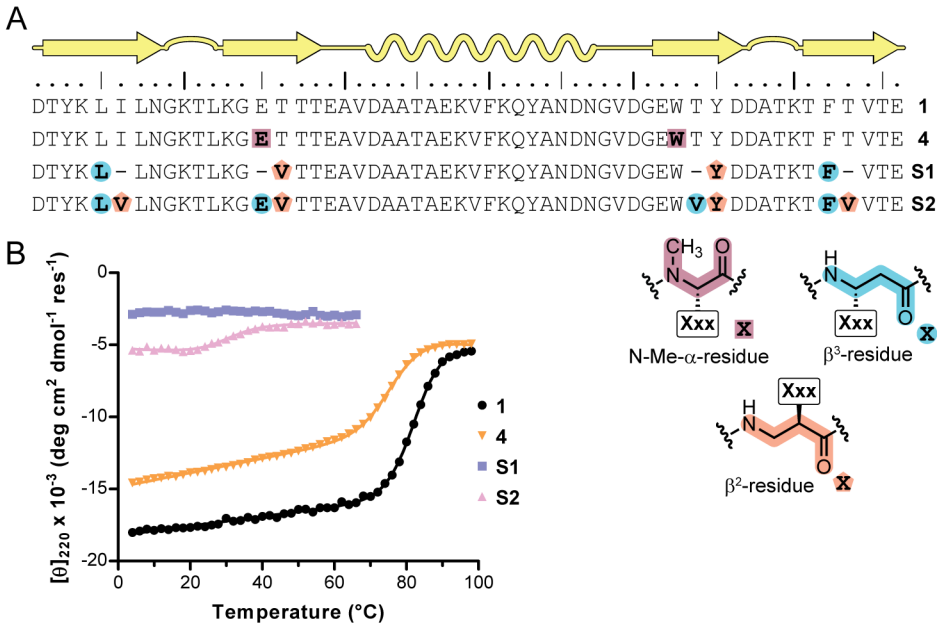


**Figure S3.** Chemical structures (A) and temperature-dependent CD data (B) for wild-type GB1 (**1**) and sheet-modified variants **4**, **S1**, and **S2**. The folds of proteins **S1** and **S2**, based on sequence-guided modification of the sheet with $\beta^3$-residues and $\beta^2$-residues, were drastically destabilized relative to wild-type GB1.

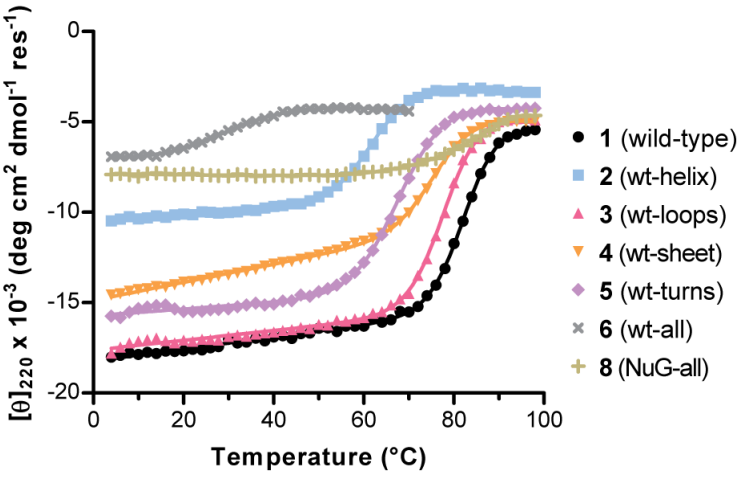


**Figure S4.** Raw molar ellipticity data from temperature-dependent CD measurements on proteins **1-6** and **8**. Values in Figure 2B, based on the data shown here, are normalized to percent unfolded for ease of comparison.

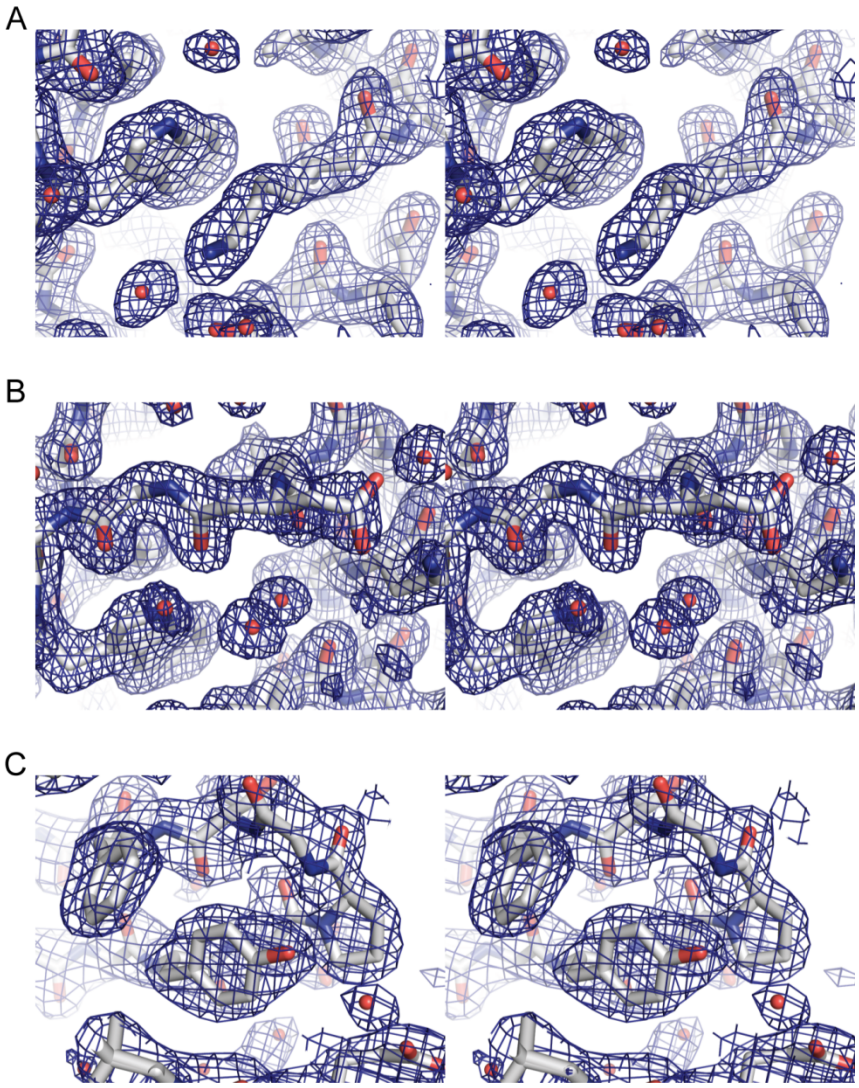


**Figure S5.** Stereo views of electron density ($\sigma_a$ weighted $2F_o-F_c$ maps) from the refined structures of protein **2** (A), protein **3** (B), and protein **5** (C).

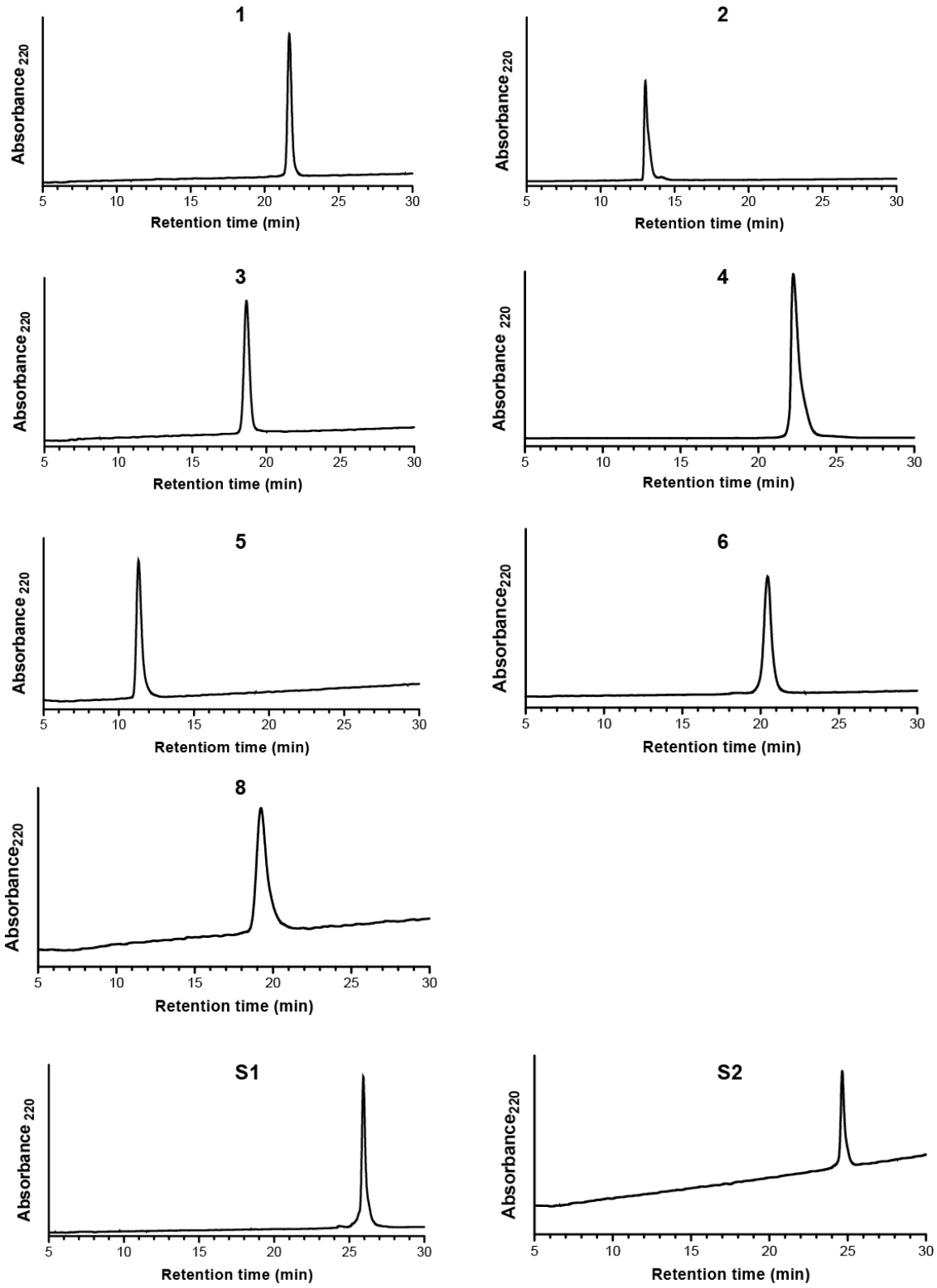


**Figure S6.** Analytical HPLC chromatograms of purified proteins **1-6**, **8**, **S1**, and **S2**.

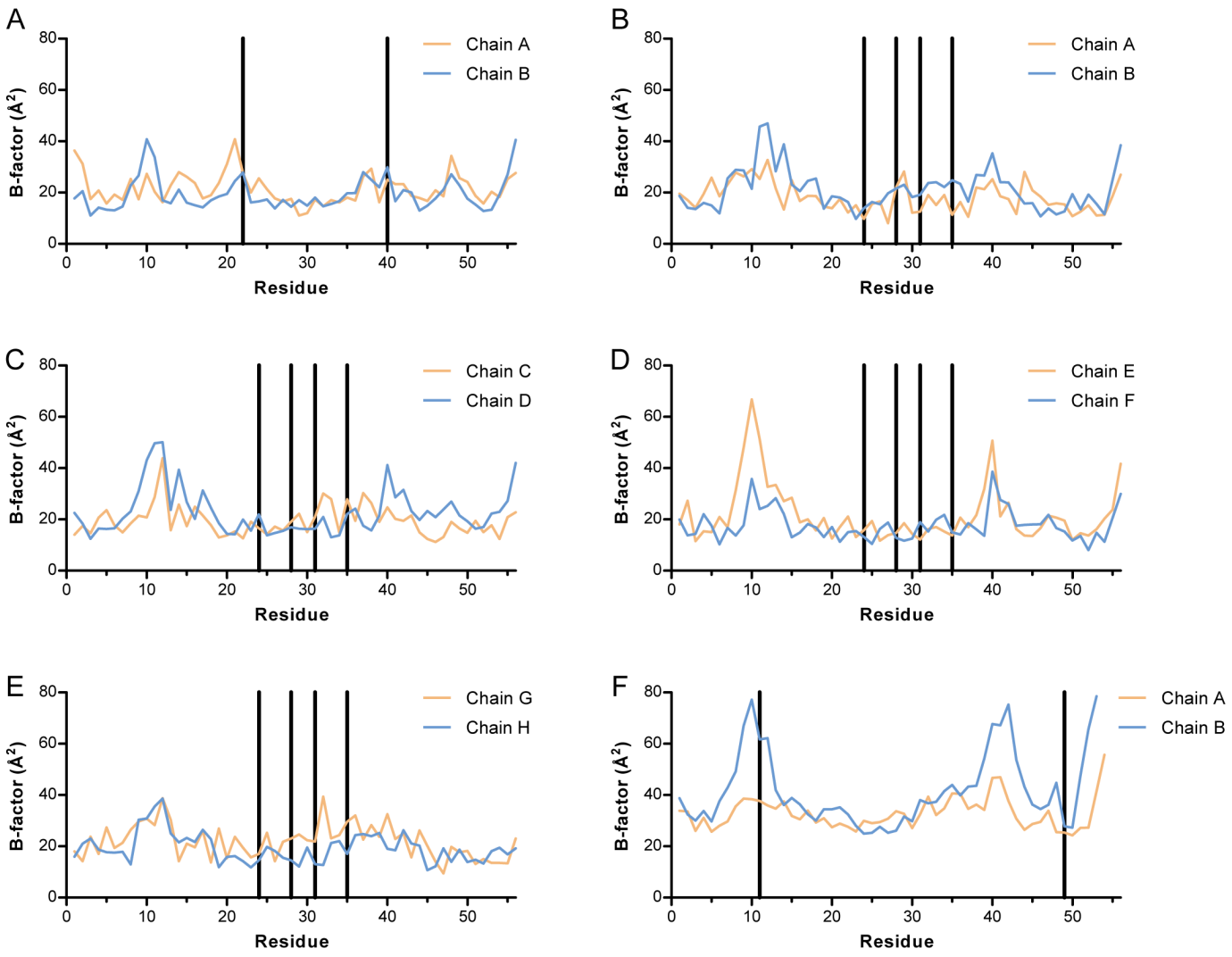


**Figure S7.** Plots of per-residue B-factors for each crystallographically independent chain in in the crystal structures of **2** (A), **3** (B-E), and **5** (F). Positions of backbone alteration are marked with vertical lines.

**Table S1.** X-ray diffraction data collection and refinement statistics for proteins **2**, **3**, and **5**.

| | Protein **2** (GB1-helix) | Protein **3** (GB1-loops) | Protein **5** (GB1-turns) |
|---|---|---|---|
| **PDB ID** | 4KGR | 4KGS | 4KGT |
| **Data Collection** | | | |
| Unit cell dimensions (Å, °) | $a$ = 52.2, $b$ = 81.2 $c$ = 52.1 $\alpha$ = 90, $\beta$ = 90, $\gamma$ = 90 | $a$ = 80.7, $b$ = 35.7, $c$ = 46.5 $\alpha$ = 90, $\beta$ = 120.4, $\gamma$ = 90 | $a$ = $b$ = 83.8, $c$ = 97.5 $\alpha$ = $\beta$ = $\gamma$ = 90 |
| Space group | $P2_1$ | C2 | $I4_122$ |
| Resolution (Å) | 32.03–2.00 (2.07–2.00) | 28.27–1.95 (2.02–1.95) | 23.28–2.00 (2.07–2.00) |
| Total observations | 97,753 | 51,377 | 166,708 |
| Unique observations | 27,508 | 8,453 | 12,054 |
| Redundancy | 3.55 (2.68) | 6.08 (3.64) | 13.83 (13.81) |
| Completeness (%) | 93.7 (82.7) | 99.6 (96.3) | 100 (100) |
| $I/\sigma$ | 23.6 (5.1) | 16.2 (4.1) | 16.7 (4.5) |
| $R_{merge}$ (%) | 4.4 (20.3) | 7.8 (22.5) | 7.6 (39.6) |
| **Refinement** | | | |
| Resolution (Å) | 32.03–2.00 | 28.27–1.95 | 23.28–2.00 |
| $R$ (%) | 16.2 | 17.6 | 19.8 |
| $R_{free}$ (%) | 19.2 | 21.7 | 21.0 |
| Avg. B factor (Å$^2$) | 23.6 | 24.9 | 39.4 |
| RMSD | | | |
| Bonds (Å) | 0.011 | 0.008 | 0.007 |
| Angles (°) | 1.54 | 1.12 | 1.05 |

**Table S2.** MALDI-TOF MS data for synthetic proteins.

| # | $[M+H]^+$ *m/z* (avg.) | |
|---|---|---|
| | **Calculated** | **Observed** |
| **1** | 6179.6 | 6178.6 |
| **2** | 6235.8 | 6232.8 |
| **3** | 6207.7 | 6204.4 |
| **4** | 6207.7 | 6207.5 |
| **5** | 5831.2 | 5828.5 |
| **6** | 5943.5 | 5944.1 |
| **8** | 5995.6 | 5997.8 |
| **S1** | 5789.3 | 5791.5 |
| **S2** | 6271.9 | 6274.6 |

## Experimental Methods

**Protein Synthesis and Purification.** Proteins were synthesized by automated methods on a PTI Tribute synthesizer using NovaPEG Rink Amide resin (70 μmol scale). Coupling reactions were performed by combining 3.0 mL of 0.4 M *N*-methylmorpholine in DMF with 7 equivalents relative to resin of the Fmoc-amino acid and HCTU. All unnatural amino acids were purchased in suitably protected form, except for Fmoc-$\beta^2$-Val-OH and Fmoc-$\beta^2$-Tyr(*t*Bu)-OH, which were synthesized using published routes.[1,2] Following a two minute preactivation, the activated amino acid was added to the resin and vortexed for 45 min. Deprotection reactions were carried out twice with 3.0 mL of a 20% v/v solution of 4-methylpiperidine in DMF for 4 min. The resin was washed three times with 3 mL of DMF for 40 sec between each cycle. After the final deprotection step, the resin was washed with 3 mL of dichloromethane followed by 3 mL of methanol. Resin was dried and subjected to cleavage by treatment with a solution of 94% TFA, 1% triisopropylsilane, 2.5% water, and 2.5% ethanedithiol. Crude protein was precipitated by addition of cold diethyl ether. The solid was pelleted by centrifugation and dissolved in 6 M guanidinium chloride, 25 mM sodium phosphate, pH 6. This solution was subjected to purification by preparative C18 reverse-phase HPLC using gradients between 0.1% TFA in water and 0.1% TFA in acetonitrile. The identity of each protein was confirmed by MS analysis on a Voyager DE Pro MALDI-TOF instrument (Table S2). Each protein was subjected to a second purification by anion-exchange chromatography on a monoQ 5/50GL column (GE Healthcare) using 0.02 M Tris buffer at pH 8 and eluting with increasing concentrations of KCl. In the case of protein **8**, peak doubling was observed in the crude HPLC, which was attributed to partial N→O acyl transfer from backbone N to side-chain O at N-Me-$Thr_{15}$ under the strongly acidic conditions of resin cleavage.[3] Material containing mass corresponding to desired product was dissolved in 0.02 M phosphate buffer at pH 8 for 24 h (leading to rapid O→N acyl migration to the native amide),[4] subjected to a second round of HPLC purification and then ion exchange as detailed above. Final protein samples were ≥ 95% pure by analytical reverse-phase HPLC (Figure S6).

**Circular Dichroism Spectroscopy.** CD measurements were performed on an Olis DSM17 Circular Dichroism Spectrometer in 2 mm quartz cells. Samples consisted of 40 μM protein in 20 mM sodium phosphate buffer, pH 7.0. Scans were carried out at 25 °C over the range of 200-260 nm with 1 nm increments and a 2 nm bandwidth. Scan data were smoothed by the Savitzky-Golay method. Melts were monitored at 220 nm over the range of 4 ºC to 98 ºC with 2 ºC increments, a dead band of 0.5 ºC, and a 2 min equilibration time at each temperature. All measurements were baseline corrected for blank buffer. Temperature-dependent CD data were fit to a two-state unfolding model to obtain melting temperature ($T_m$). The change in free energy of folding for each mutant relative to wild-type ($\Delta\Delta G_{fold}$) was estimated from the change in $T_m$ ($\Delta T_m$),[5] using the enthalpy of folding determined for GB1 by differential scanning calorimetry.[6]

**Protein Crystallization, X-ray Diffraction Data Collection, and Structure Determination.** Crystals of proteins **2**, **3**, and **5** were grown by hanging drop vapour diffusion from 17 mg/mL stock solutions of protein in water and well buffers as follows: 0.15 M sodium acetate pH 4.6, 20% w/v PEG 4000 for protein **2**; 0.1 M sodium acetate pH 4.6, 16% w/v PEG 3350 for protein **3**; and 0.1 M sodium acetate pH 4.6, 8% w/v PEG 4000 for protein **5**. A single crystal of each protein was flash frozen in liquid nitrogen after cryoprotection in well buffer supplemented with 30% v/v glycerol. Diffraction data were collected using CuKα radiation on a Rigaku/MSC diffractometer operated at 100 K. Raw diffraction data were indexed, integrated, and scaled with

d*TREK. Structures of **2**, **3**, and **5** were solved by molecular replacement using a published structure of wild-type GB1 (PDB: 2QMT)[7] to generate the search model. Structure solution and refinement were carried out using the CCP4[8] and Phenix[9] software suites. In the case of protein **2** (GB1-helix), the diffraction pattern indexed as tetragonal, but subsequent analysis and structure determination revealed the actual lattice to be $P2_1$ with near perfect pseudomerohedral twinning (twin operator: L, –K, H). Refinement for **2** was performed using the twin refinement algorithm implemented in Phenix. Coordinates and structure factors for **2**, **3**, and **5** have been deposited in the Protein Data Bank under accession codes 4KGR, 4KGS, and 4KGT.

**References**


1. Lengyel, G. A.; Frank, R. C.; Horne, W. S., *J. Am. Chem. Soc.* **2011**, *133*, 4246-4249.
2. Lengyel, G. A.; Horne, W. S., *J. Am. Chem. Soc.* **2012**, *134*, 15906-15913.
3. Oliyai, R.; Stella, V. J., *Pharm. Res.* **1992**, *9*, 617-22.
4. Mutter, M.; Chandravarkar, A.; Boyat, C.; Lopez, J.; Dos Santos, S.; Mandal, B.; Mimna, R.; Murat, K.; Patiny, L.; Saucède, L.; Tuchscherer, G., *Angew. Chem. Int. Ed.* **2004**, *43*, 4172-4178.
5. Becktel, W. J.; Schellman, J. A., *Biopolymers* **1987**, *26*, 1859-1877.
6. Alexander, P.; Fahnestock, S.; Lee, T.; Orban, J.; Bryan, P., *Biochemistry* **1992**, *31*, 3597-3603.
7. Frericks Schmidt, H. L.; Sperling, L. J.; Gao, Y. G.; Wylie, B. J.; Boettcher, J. M.; Wilson, S. R.; Rienstra, C. M., *J. Phys. Chem. B* **2007**, *111*, 14362-14369.
8. *Acta Crystallogr., Sect D: Biol. Crystallogr.* **1994**, *50*, 760-3.
9. Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., *Acta Crystallogr., Sect D: Biol. Crystallogr.* **2010**, *66*, 213-221.