Electronic Supplementary Material (ESI) for Chemical Science.

Folding Thermodynamics of Protein-Like Oligomers with Heterogeneous Backbones

Zachary E. Reinert and W. Seth Horne*

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

SUPPORTING INFORMATION

Materials and Methods

Protein Synthesis and Purification. Proteins were synthesized by automated methods on a PTI Tribute synthesizer using NovaPEG Rink Amide resin (70 μmol scale). Fmoc-protected α- and β^3-residues were purchased with suitable side-chain protecting groups, and Fmoc-(*S*,*S*)-β-ACPC was synthesised as previously described.[1] Coupling reactions were performed by combining 3.0 mL of 0.4 M *N*-methylmorpholine in DMF with 7 equivalents each of Fmoc-amino acid and HCTU. Following a 2 min preactivation, the solution was added to 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 Fmoc deprotection reaction, the resin was washed with 3 mL of dichloromethane followed by 3 mL of methanol. Resin was dried treated with a solution of 94% TFA, 1% triisopropylsilane, 2.5% water, and 2.5% ethanedithiol for 3 h to cleave protein from resin and remove side chain protecting groups. After filtration, the 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 for purification.

Each protein was purified by a two-stage protocol of HPLC followed by ion exchange. Preparative reverse-phase HPLC was carried out on a C18 column using gradients between 0.1% TFA in water and 0.1% TFA in acetonitrile. Anion-exchange chromatography was carried out on a monoQ 5/50GL column (GE Healthcare) using 0.02 M Tris buffer at pH 8 and eluting with increasing concentrations of KCl. Final protein samples were ≥ 95% pure by analytical reversed-phase HPLC (Figure S1). The identity of each protein was confirmed by MS analysis on a Voyager DE Pro MALDI-TOF instrument (Table S2).

Circular Dichroism Spectroscopy and Data Analysis. CD measurements were performed on an Olis DSM17 Circular Dichroism Spectrometer in 1 mm quartz cells. Samples consisted of 40 μM protein in 20 mM sodium phosphate buffer, pH 7 with varying concentrations of guanidinium hydrochloride. Thermal melts were monitored at 220 nm over the range of 4 °C to 98 °C with 2 °C increments, a dead band of 0.5 °C, and a 2 min equilibration time at each temperature. All measurements were baseline corrected for buffer. Raw CD data were fit using Mathematica 8 (Wolfram) and equations reported previously,[2] summarized briefly below. The protein folding free energy (ΔG) at a given temperature (T) and concentration of guanidinium ([Gdm]) is given by eq. 1:

$$\Delta G = \Delta H° - T\Delta S° + \Delta C_p * (T - T_0 + T * \ln(\tfrac{T_0}{T}) - m * [Gdm] \qquad (1)$$

where $\Delta H°$ and $\Delta S°$ are the folding enthalpy and entropy at a reference temperature $T°$, ΔC_p the change in heat capacity, and m the dependence of the folding free energy on [Gdm]. The observed ellipticity (θ_{obs}) at a particular T and [Gdm] is given by eq. 2:

$$\theta_{obs} = \frac{(\theta_n + \theta_u * \exp(\frac{-\Delta G°}{RT}))}{(1+\exp(\frac{-\Delta G°}{RT})} \qquad (2)$$

where θ_n and θ_u are the ellipticity of the folded and unfolded states. Based on literature precedent[2,3] and inspection of the raw data, θ_n and θ_u were assumed to vary linearly with T and [Gdm] according to eq. 3 and 4:

$$\theta_u = a + bT + c * [Gdm] \qquad (3)$$

$$\theta_n = d + eT + f * [Gdm] \qquad (4)$$

Some proteins (**2-4** and **6-7**) lacked well-defined fully folded baselines as a function of [Gdm]. The parameter *f* was constrained to zero for these fits; this approximation did not change the observed thermodynamic values by more than 10% when applied to proteins with better defined baselines.[3] The folding/unfolding transitions of GB1 and its analogues were assumed to follow a two-state model,[4] and it was assumed that ΔC_p and *m* do not vary over the range of experimental conditions. Results from the fits are summarized in Table S1.

Crystallography. Crystals of proteins **4**, **5**, and **8** were grown by hanging drop vapor diffusion. The drop, 1 μL of protein (~15 mg/mL in water) mixed with 1 μL of well buffer, was allowed to equilibrate over 1 mL of well buffer at room temperature. The well buffers were 0.2 M calcium chloride, 0.1 M sodium acetate pH 4.6, 30% v/v isopropanol for protein **4**; 0.1 M sodium citrate pH 5.6, 20% v/v isopropanol, 20% w/v PEG 4000 for protein **5**; and 0.2 M magnesium chloride, 0.1 M HEPES pH 7.5, 30% v/v isopropanol for protein **8**. Single crystals were flash frozen in liquid nitrogen after cryoprotection in well buffer diluted to 30% (**4**), 5% (**5**), or 10% (**8**) v/v glycerol. Diffraction data were collected using CuK_α radiation on a Rigaku/MSC diffractometer operated at 100 K. Raw frames were indexed, integrated, and scaled with d*TREK. Structure solution and refinement were carried out using the Phenix[5] software suite. The structures were solved by molecular replacement using published coordinates of wild-type GB1 (PDB 2QMT).[6] Statistics for data collection and structure refinement are given in the Supporting Information (Table S3). Coordinates and structure factors were deposited in the PDB under accession codes 4OZA (**4**), 4OZB (**5**), and 4OZC (**8**).

References

1 P. R. LePlae, N. Umezawa, H.-S. Lee and S. H. Gellman, *J. Org. Chem.*, 2001, **66**, 5629-5632.

2 B. Kuhlman and D. P. Raleigh, *Protein Sci.*, 1998, **7**, 2405-2412.

3 B. C. Buer, B. J. Levin and E. N. G. Marsh, *J. Am. Chem. Soc.*, 2012, **134**, 13027-13034.

4 P. Alexander, S. Fahnestock, T. Lee, J. Orban and P. Bryan, *Biochemistry*, 1992, **31**, 3597-3603.

5 P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart, *Acta Crystallogr., Sect D: Biol. Crystallogr.*, 2010, **66**, 213-221.

6 H. L. Frericks Schmidt, L. J. Sperling, Y. G. Gao, B. J. Wylie, J. M. Boettcher, S. R. Wilson and C. M. Rienstra, *J. Phys. Chem. B*, 2007, **111**, 14362-14369.

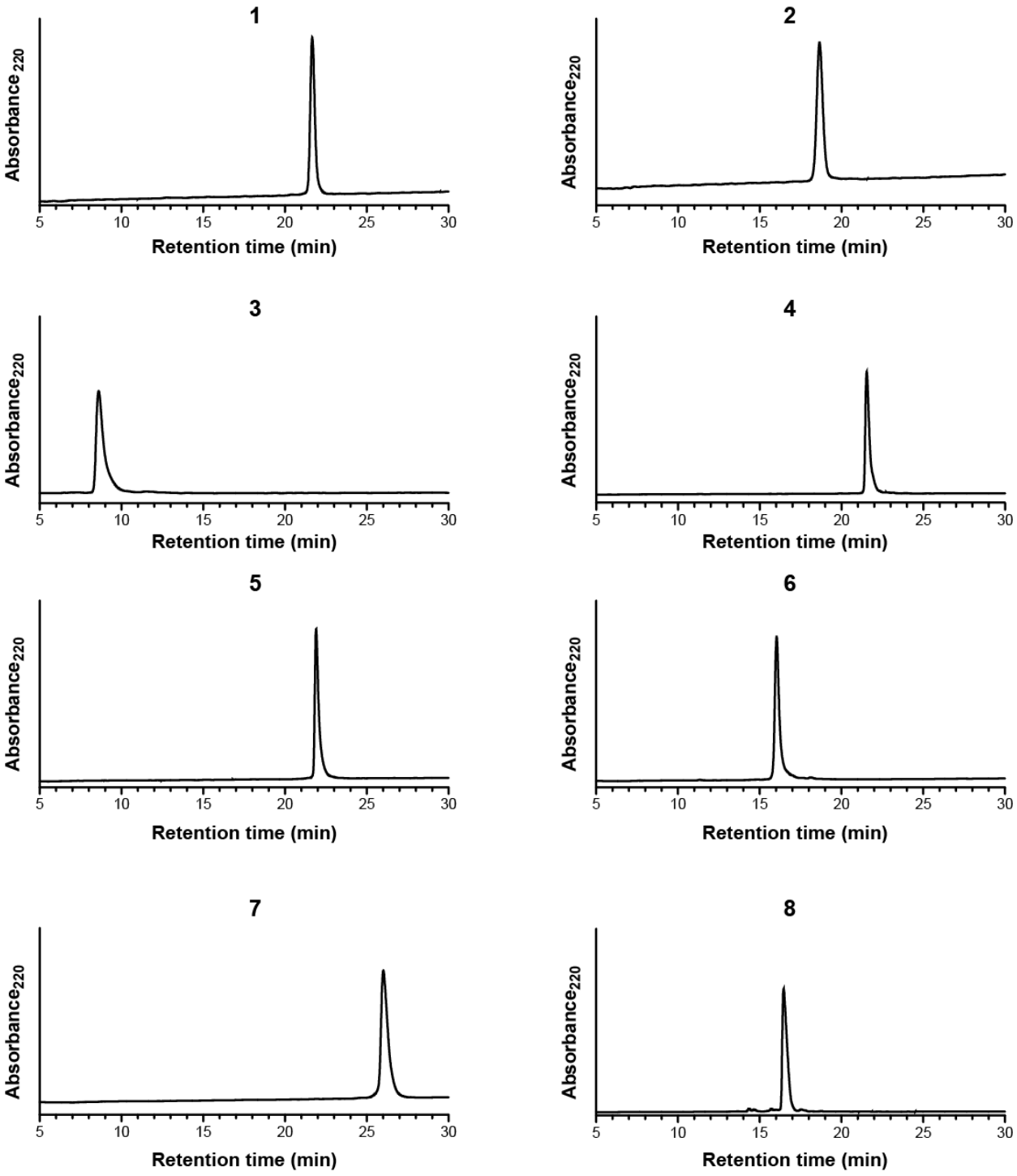

Figure S1. Analytical HPLC chromatograms of purified proteins **1-8**.

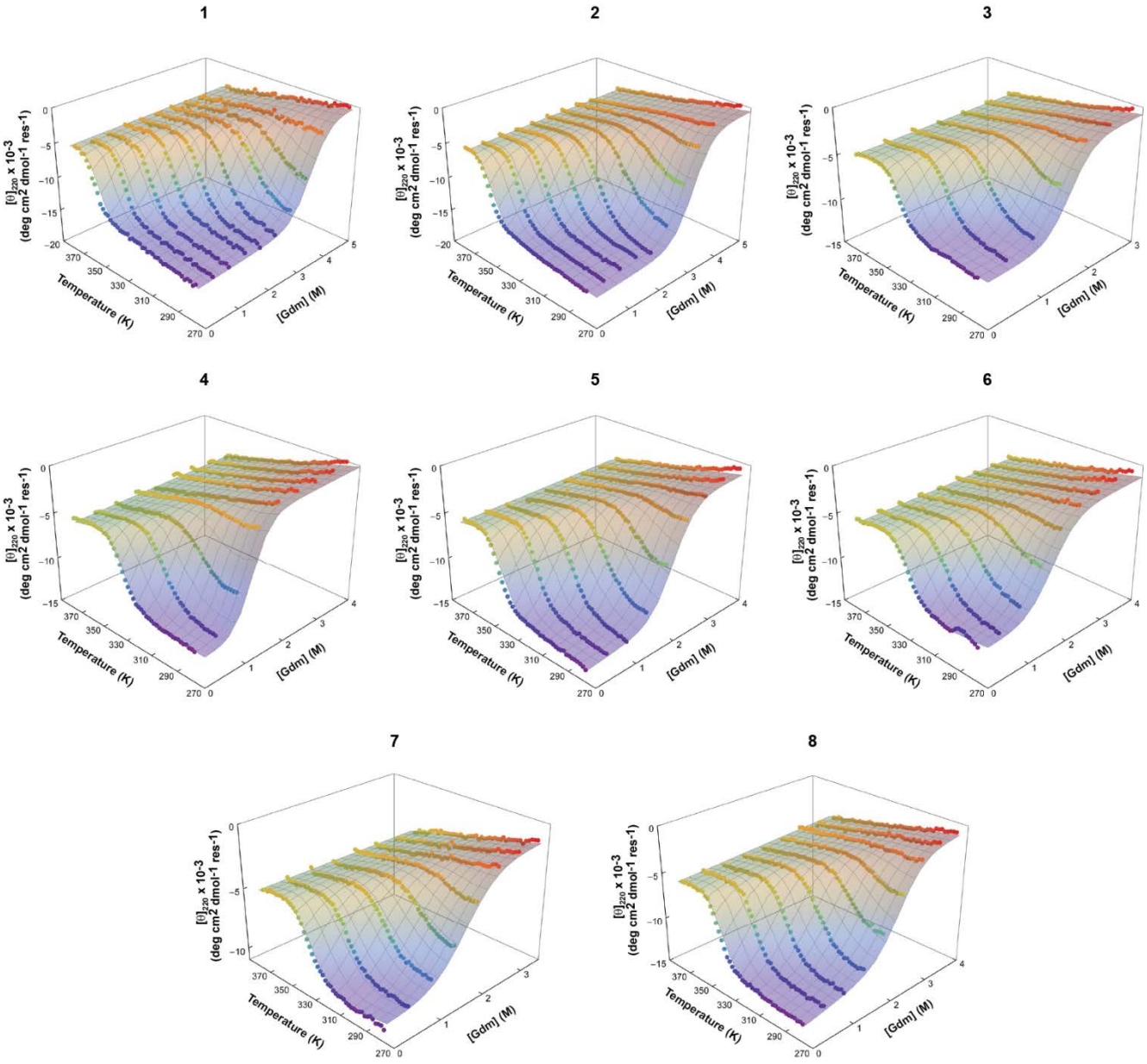

Figure S2. CD signature of proteins **1**-**8** at 220 nm as a function of temperature and chemical denaturant. Raw data (points) are fit (surface) to extract thermodynamic parameters for the folding equilibrium.

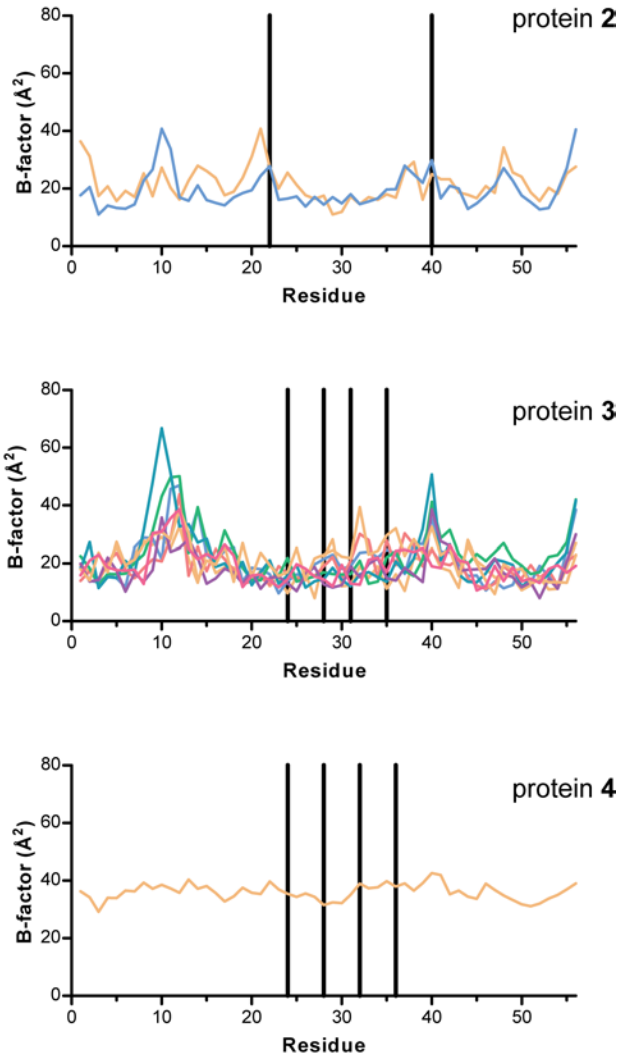

Figure S3. B-factor analysis for the crystal structures of proteins **2-4**. Each line color is a different chain in the asymmetric unit and positions of β^3-residues are marked as vertical lines.

Table S1. Thermodynamic parameters for the unfolding transition of wild-type GB1 (**1**) and backbone-modified analogues **2-8**.[a]

Protein	ΔH° (kcal mol^{-1})	TΔS° (kcal mol^{-1})	ΔG° (kcal mol^{-1})	ΔC_p (kcal mol^{-1} K^{-1})	*m* (kcal mol^{-1} M^{-1})	T_m (°C)[b]
1	21.9 ± 0.6	16.1 ± 0.3	5.8 ± 0.6	0.60 ± 0.02	1.80 ± 0.04	82.1
2	19.4 ± 0.4	14.6 ± 0.3	4.8 ± 0.5	0.59 ± 0.02	1.82 ± 0.04	77.6
3	18.3 ± 0.4	15.2 ± 0.3	3.1 ± 0.5	0.53 ± 0.02	2.48 ± 0.05	61.5
4	15.4 ± 0.4	12.8 ± 0.3	2.6 ± 0.5	0.63 ± 0.02	2.58 ± 0.05	56.5
5	18.4 ± 0.5	14.9 ± 0.3	3.5 ± 0.6	0.48 ± 0.02	2.00 ± 0.05	70.8
6	16.3 ± 0.3	14.3 ± 0.3	2.0 ± 0.4	0.49 ± 0.02	1.76 ± 0.03	57.1
7	13.8 ± 0.3	11.9 ± 0.3	1.9 ± 0.4	0.39 ± 0.02	1.49 ± 0.03	60.5
8	18.2 ± 0.5	14.9 ± 0.3	3.3 ± 0.6	0.52 ± 0.02	1.71 ± 0.04	67.9

[a] Thermodynamic values are at 298 K and reported errors are from parameter uncertainties in the fit. [b] Midpoint of the CD thermal unfolding transition in the absence of chemical denaturant.

Table S2. MALDI-TOF MS data for synthetic proteins **1**-**8**.

#	$[M+H]^+$ *m/z* (average)	
	Calculated	**Observed**
1	6179.6	6178.2
2	6207.7	6204.6
3	6235.8	6232.1
4	6235.8	6237.3
5	6244.8	6241.0
6	6173.6	6169.2
7	6182.7	6179.6
8	6252.9	6247.7

Table S3. X-ray diffraction data collection and refinement statistics for proteins **4**, **5**, and **8**.

	Protein **4**	Protein **5**	Protein **8**
PDB ID	4OZA	4OZB	4OZC
Data Collection			
Unit cell dimensions (Å, °)	$a = b = 65.9$, $c = 21.9$ $\alpha = \beta = \gamma = 90$	$a = 92.5$, $b = 22.6$, $c = 64.5$ $\alpha = \gamma = 90$, $\beta = 120.9$	$a = b = 79.3$, $c = 22.5$ $\alpha = \beta = \gamma = 90$
Space group	$I4_1$	C2	$I4_1$
Resolution (Å)	23.31–2.20 (2.28–2.20)	22.75–1.80 (1.86–1.80)	28.03–2.30 (2.38–2.30)
Total observations	13,203	46,166	17,236
Unique observations	2,418	10,465	3,245
Redundancy	5.46 (5.42)	4.4 (3.2)	5.3 (5.5)
Completeness (%)	96.6 (98.8)	94.9 (89.1)	98.8 (100)
I/σ	11.0 (3.2)	18.0 (3.3)	11.9 (2.2)
R_{merge} (%)	10.2 (26.4)	6.9 (20.9)	8.8 (31.8)
Refinement			
Resolution (Å)	23.31–2.20	25.00–1.80	28.03–2.30
R (%)	23.3	19.2	21.9
R_{free} (%)	26.0	21.0	25.3
Avg. B factor ($Å^2$)	38.1	22.8	44.8
RMSD			
Bonds (Å)	0.007	0.011	0.004
Angles (°)	1.14	1.4	1.05