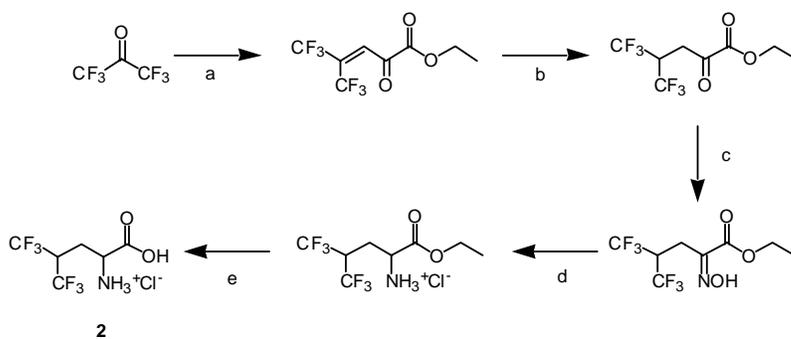


Supporting Information for Tang and Tirrell, *Biosynthesis of a Highly Stable Coiled-coil Protein Containing Hexafluoroleucine in an Engineered Bacterial Host*

Hexafluoroleucine synthesis

The synthesis (Scheme S1) is modified from that reported by Zhang et al.¹ An alternative procedure was reported by Xing et al.²

Scheme S1



(a) $\text{PPh}_3\text{C}_5\text{H}_6\text{O}_3$, $(\text{C}_2\text{H}_5)_2\text{O}$, -78°C . (b) H_2 (30 psi), PtO_2 , RT. (c) NH_2OH , $\text{C}_2\text{H}_3\text{O}_2\text{Na}$, 50% EtOH, 50°C . (d) TiCl_3 , L-tartaric acid, NaBH_4 , pH 7.0, RT. (e) 6 N HCl, 120°C , o/n.

Procedures for steps a, b and e are adopted from Ref. 1. The procedure for α -oxime ester reduction (step d) is adopted from Hoffman et al.³ The overall yield of synthesis is 12%. Hexafluoroleucine hydrochloride (**2**) is used directly in the *in vivo* incorporation studies.

¹H-NMR of **2** (CD_3OD): 4.1 (dd, 1H), 3.9 (m, 1H), 2.4 (m, 1H), 2.25 (dt, 1H). ¹⁹F of **2** (10% D_2O , referenced to CF_3COOH): 7.96 (q, $J=10$), 7.72 (q, $J=10$).

¹ Zhang, C.; Ludin, C.; Eberle, M. K.; Stoeckli-Evans, H.; Keese, R. *Helv. Chim. Acta.* **1998**, *81*, 174-181.

² Xing, X. C.; Fichera, A.; Kumar, K. *Org. Lett.* **2001**, *3*, 1285-1286.

³ Hoffman, C.; Tanke, R. S.; Miller, M. J. *J. Org. Chem.* **1989**, *54*, 3750-3751.

Figure S1 is the ^{19}F NMR spectrum of **2** (10% D_2O) referenced to trifluoroacetic acid. Signals that correspond to those of trifluoroleucine were observed (2.32 (d, $J=10$), 2.18 (d, $J=10$)). Trifluoroleucine is determined to be present at a level of $< 0.2\%$ by integration.

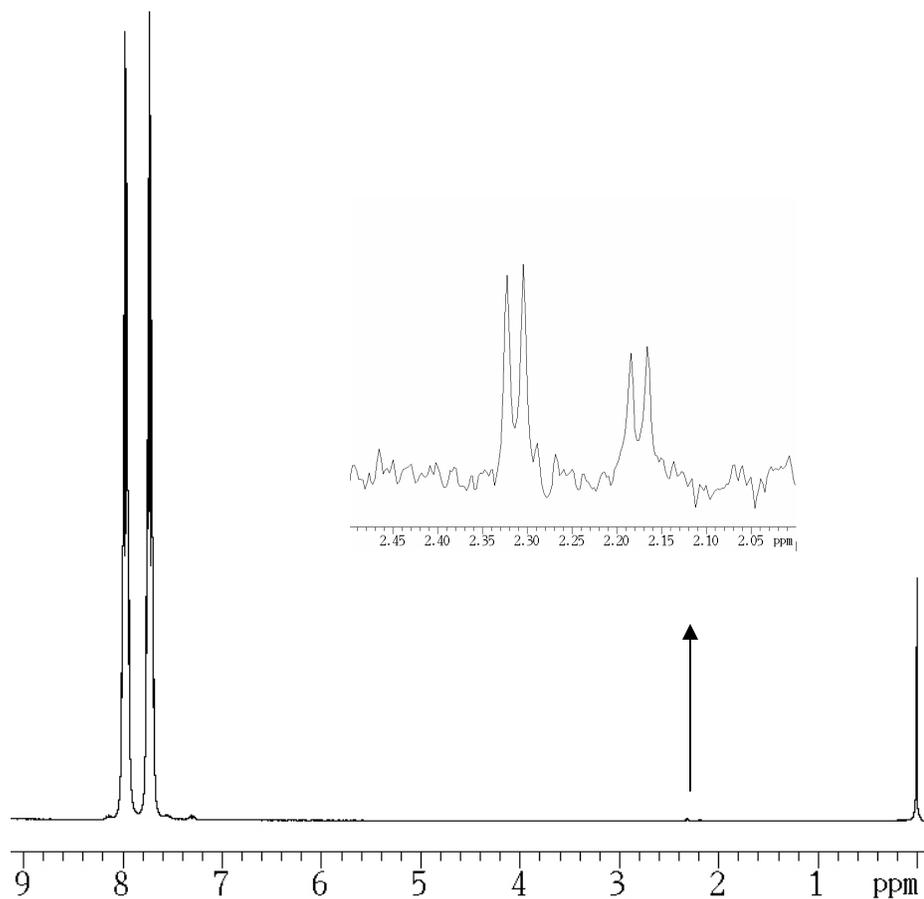


Figure S1: ^{19}F NMR spectrum of **2** showing traces of trifluoroleucine contamination.

Cloning of LeuS and overexpression

PCR amplification of leuS: *E. coli* genomic DNA was prepared from an overnight culture using DNEasy kit from Qiagen (Chatsworth, CA). Primer sequences are as follows:

LeuS53: ACGTTCTTTGTTGCTAGCTTTGCTAATACG

LeuS35: TATCACGCAGATGCTAGCCACACCCGGCC

The *NheI* restriction sites are underlined. PCR conditions are: 50 ng/100 μ l of template DNA, 250 ng/100 μ l of each primer. 55°C annealing temperature (2 minutes) and 72°C extension temperature (4 minutes). The resulting 2850 bp target DNA also includes the *leuS* promoter.

Construction of pA1EL: The amplified DNA is digested using *NheI* and ligated into the *NheI* site of expression plasmid pQE41 to yield the expression vector pA1EL. The orientation was checked by *Sall* digestion. Overexpression of LeuRS is checked by SDS-PAGE of overnight cultures of strains with pA1EL. The accumulation of LeuRS (86kDa) is visible in strains bearing the overexpression vector (Figure S2).

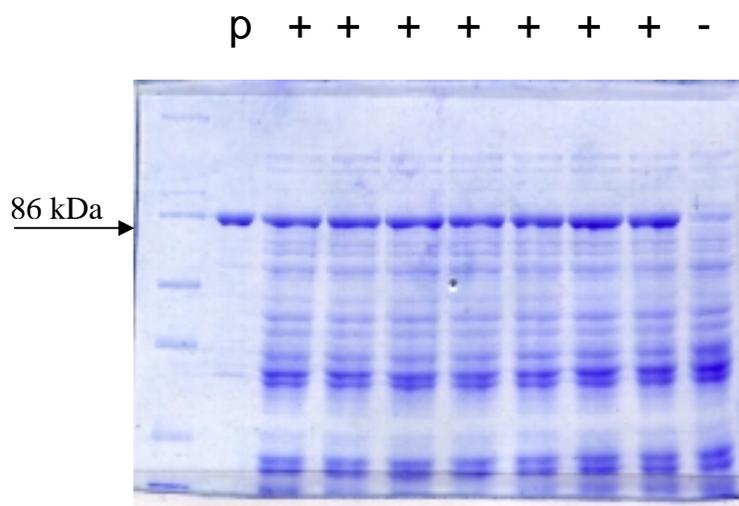


Figure S2. Protein products from strains transformed with pA1EL (+) and with pQE41 (-). Purified protein (p) is shown as a reference.

Protein expression and mass analysis

Protein expression procedures: LAM1000 strains were transformed with pA1EL and pREP4 to yield the expression strain LAM1000/pA1EL/pREP4. M9 medium (10 ml) supplemented with all 20 amino acids (40 mg/L), 1 mM MgSO₄, 1 mM CaCl₂, 0.4 wt% glucose, 1 mg/L thiamine and the antibiotics ampicillin (200 mg/L) and kanamycin (25 mg/L) was inoculated with 100 μ l of fresh overnight culture (M9) of

the expression strain. After the culture had grown to OD₆₀₀ of 1.0, the cells were pelleted and washed with cold 0.9% NaCl three times. The cell pellet was then resuspended in 10 mL M9 medium (as described above but without leucine) supplemented with **2** (320 mg/L). IPTG (1 mM) was added after a certain time period (see below) to induce protein expression. Cells were collected after 3 hr by centrifugation (15000 g, 2 min, 4°C). The pellets were resuspended in 600 µL of Buffer A (8 M urea, 0.1 M NaH₂PO₄, 0.01 M Tris, pH 8.0) and stored at -80°C overnight. The whole cell lysate was then subjected to purification by Ni-NTA spin columns (Qiagen) according to the manufacturer's instructions.

MALDI analysis of intact protein: The eluent from the spin column was dialyzed against water extensively and lyophilized to a fluffy powder. The powder was sent directly for MALDI matrix and amino acid analyses.

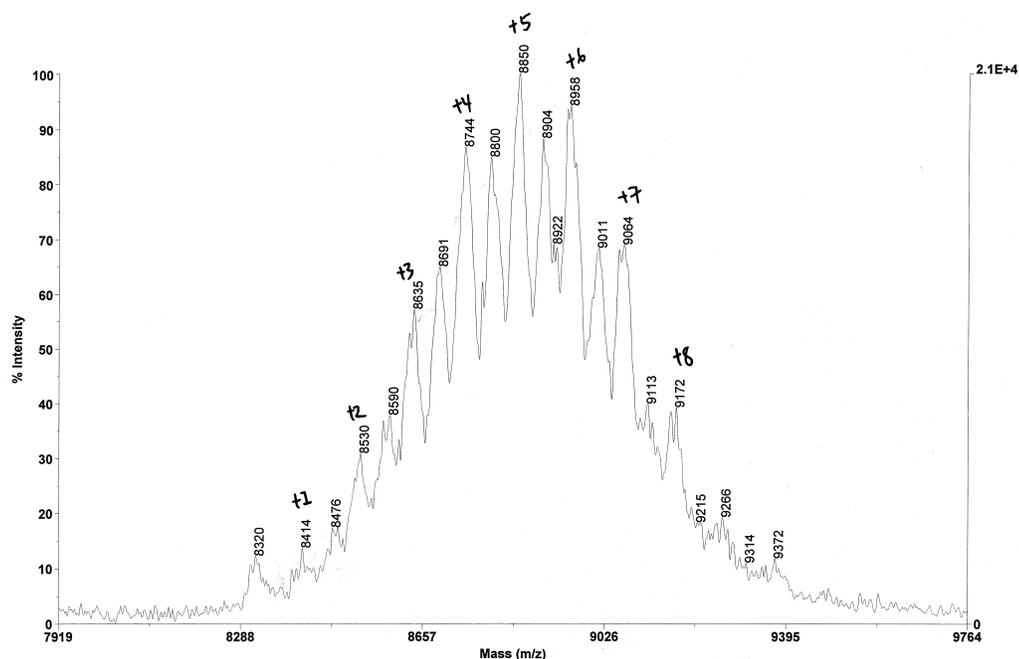
MALDI of tryptic fragments: The eluent from spin column purification was in Buffer A, pH 4.5. 450 µL of 50 mM (NH₄)₂CO₃ was added to 50 µL of eluent to bring the pH to optimal trypsin working pH (8.0). 5 µL of trypsin stock solution (20 µg/200 µL) was added and the sample was incubated at room temperature overnight. The reaction was quenched by addition of 2 µL of TFA. The reaction mixture was subjected to C18 ZipTip (Millipore) purification and eluted with 3 µL of 0.1 TFA, 50% CH₃CN. 1 µL of eluent was used for the MALDI analysis (α-cyano-β-hydroxycinnamic acid, 10 ml/mL in 1:1 H₂O/CH₃CN).

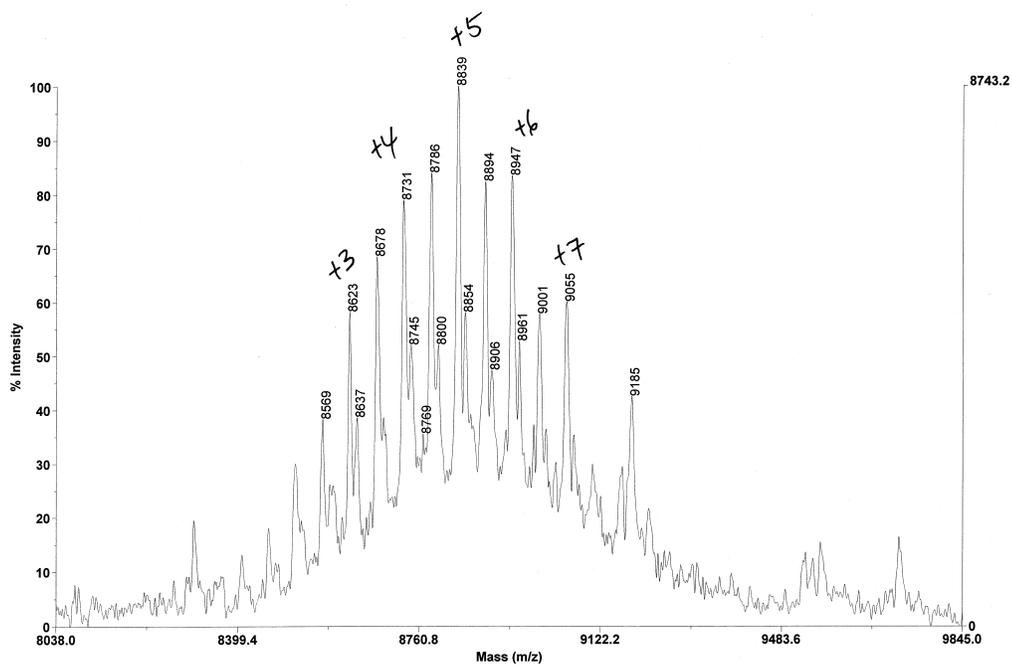
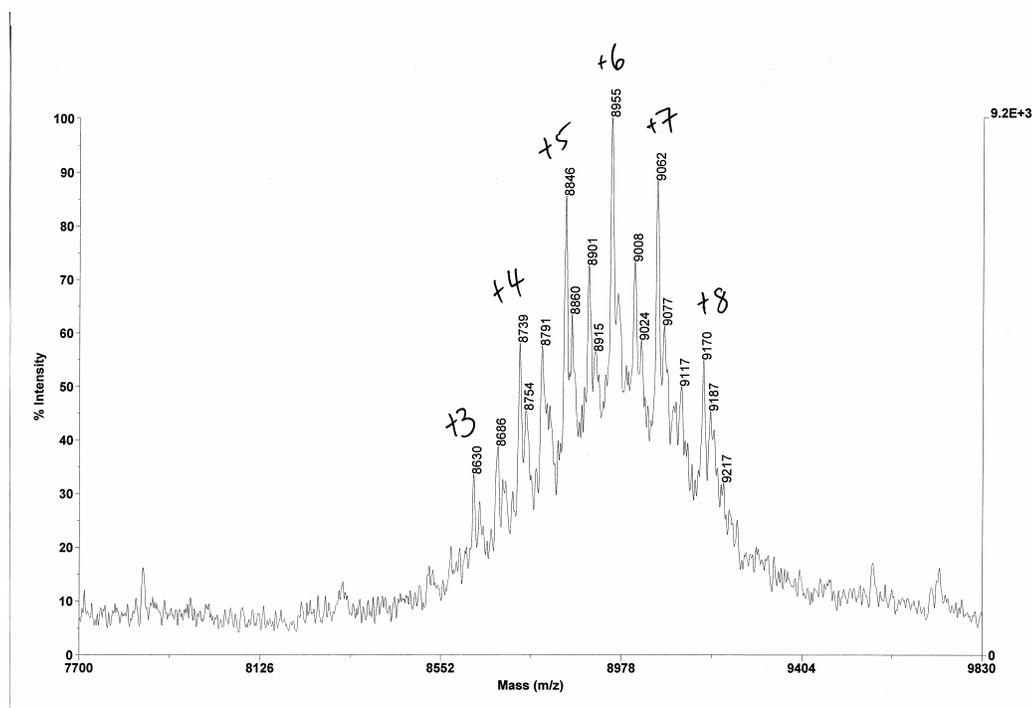
Minimizing trifluoroleucine contamination of target protein

Even though trifluoroleucine is present at less than 0.2% in the sample of **2** used to supplement the medium, its faster rate of activation by LeuRS allows it to be incorporated into proteins readily. The result is enrichment of trifluoroleucine in the target protein. Under normal expression conditions where the cells were incubated for 10 minutes after medium shift prior to induction of A1 expression by IPTG, the amount of trifluoroleucine in the protein was greater than 15%. We postulated that extending the incubation time could allow the limited amount of trifluoroleucine to be utilized in the synthesis of other cellular proteins. Subsequently at the time of target protein expression, the amount of

trifluoroleucine is diminished. We performed experiments with incubation times extended to 30, 60, 90 and 120 minutes. We did not extend the time beyond 2 hours because cell death was observed (shown by a decrease of OD_{600}). The proteins produced under these conditions were analyzed by MALDI-MS. The results are shown in Figure S3. As we anticipated, the amount of trifluoroleucine present significantly decreased as we increased the incubation time. When the cells were incubated in M9 supplemented with **2** for 2 hours prior to induction, the amount of trifluoroleucine was reduced to below 3% as compared to 15% for 10 minute incubation, and the amount of **2** in A1 was elevated to 80% as compared to 50%. The mass spectra and thermal denaturation data reported in the main text are for the protein expressed after 2 hours of incubation, followed by 3 hours protein expression.

S3-A (10 min)



S3-B (30 min)**S3-C (60 min)**

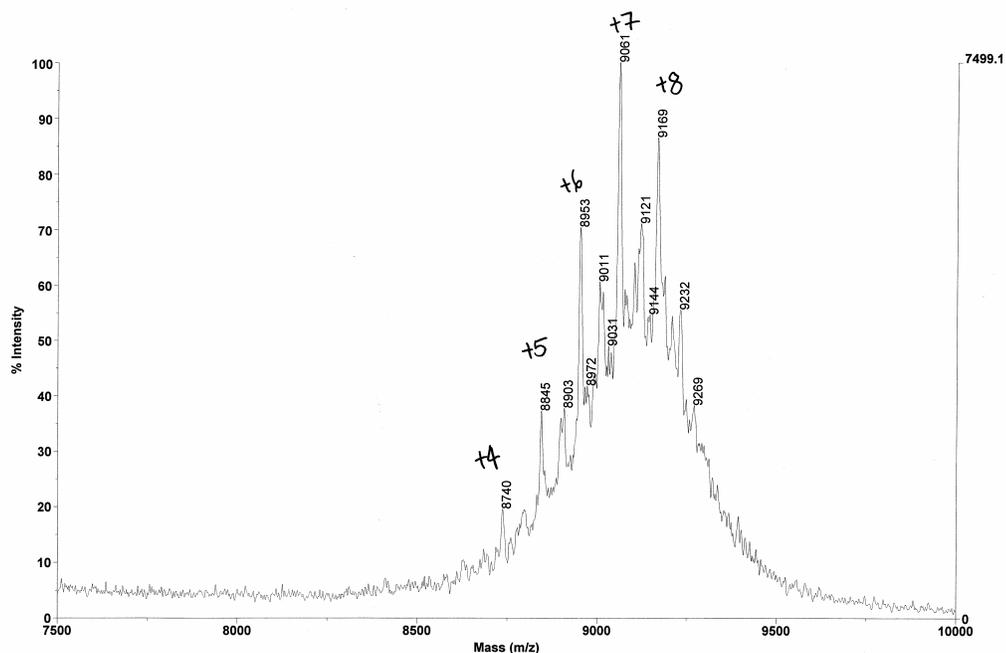
S3-D (120 min)

Figure S3. Comparison of overall protein composition as a function of increasing incubation time after medium shift and prior to protein expression. The incubation times for the proteins shown are A: 10 min; B: 30 min; C: 60 min; D: 120 min (same as that shown in the main text). The amount of trifluoroleucine (as can be determined from the half integer peaks in the MS) decreased as the incubation time was extended, while the mass of protein increased as a result of depletion of leucine and trifluoroleucine.

Amino acid analysis of HA1

Amino Acid	# of Res.	Expected		Observed	
		Mole Fractions	pMoles	Mole Fractions	% Err
ASP	9	12.2	4828	11.53	5.22
GLU	13	17.6	7599	18.15	3.29
SER	6	8.1	3492	8.34	2.84
GLY	5	6.8	3147	7.51	11.21
HIS	6	8.1	3352	8.00	1.28
ARG	6	8.1	3360	8.02	1.05
ALA	7	9.5	3743	8.94	5.52
VAL	3	4.1	1585	3.78	6.65
MET	2	2.7	1218	2.91	7.60
ILE	3	4.1	1535	3.67	9.59
LEU	8	10.8	1036	2.47	77.12
LYS	4	5.4	2049	4.89	9.49
HFL			3354	8.01	
Total	74	100	41747	100	

The combined pmoles of LEU and HFL total 4390. The combined mole fractions of LEU+HFL amount to 10.48% compared to the expected value of 10.8%. We attribute the difference of 0.32% to presence of TFL. Calculations based on mole fraction of LEU, HFL and TFL lead to the following distribution of amino acids at the leucine positions:

LEU: $2.47 / 10.8 = 22.9 \%$

HFL: $8.01 / 10.8 = 74.2 \%$

TFL: $0.32 / 10.8 = 2.9\%$

TFL / HFL: 0.04 (Figure reported in main text)

Secondary Structure of HA1

The secondary structure of HA1 was compared to that of A1 in PBS, pH 7.4, using circular dichroism. The data collection was performed at 0°C. The protein concentration was determined to be 10 μM . Figure S4 shows that fluorination does not change the secondary structure of the protein. HA1 remains > 90% helical as judged by the ellipticity at 222 nm. (Chen, Y. H.; Yang, J. T.; Chau, K. H. *Biochemistry* **1974**, *13*, 3350.)

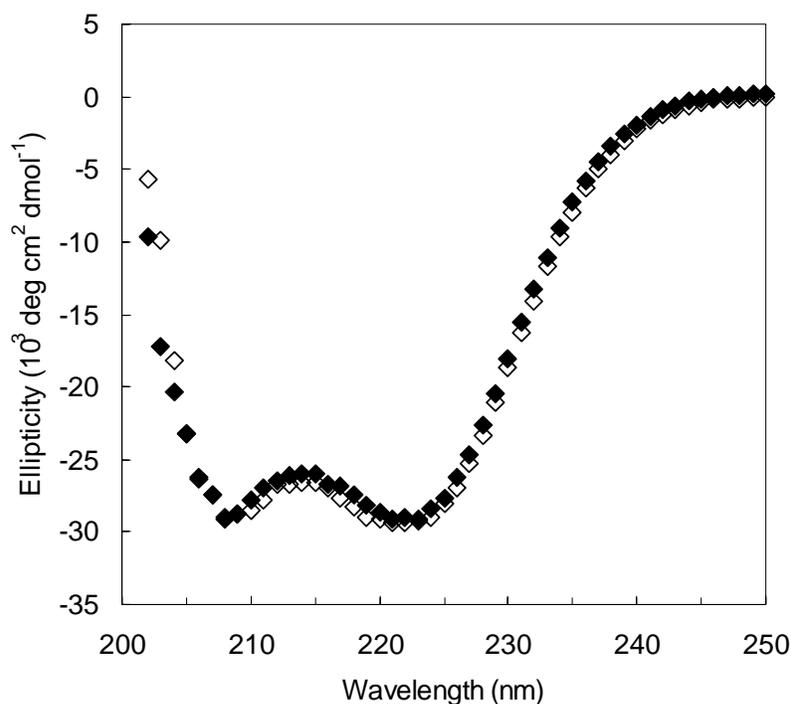


Figure S4. Secondary structure comparison between A1 (filled symbols), HA1 (open symbols). The concentrations of both proteins were at 10 μM . The measurements were performed at 0°C in PBS.

Ultracentrifugation Studies

Sedimentation equilibrium analysis was performed using a Beckman XLI analytical ultracentrifuge, recording interference data and radial absorbance at 236 and 280 nm at the same time. Initial peptide concentrations ranged between 100 and 300 μM ; buffer was 0.01M sodium phosphate, pH 7.4, containing 0.1 M NaCl. The samples were centrifuged at

35000, 40000, 45000 rpm, until equilibrium was reached. The equilibrium constant for dimer-tetramer transition is determined to be $188 \pm 4 \mu\text{M}$, and is used to determine the species plot shown in Figure S5.

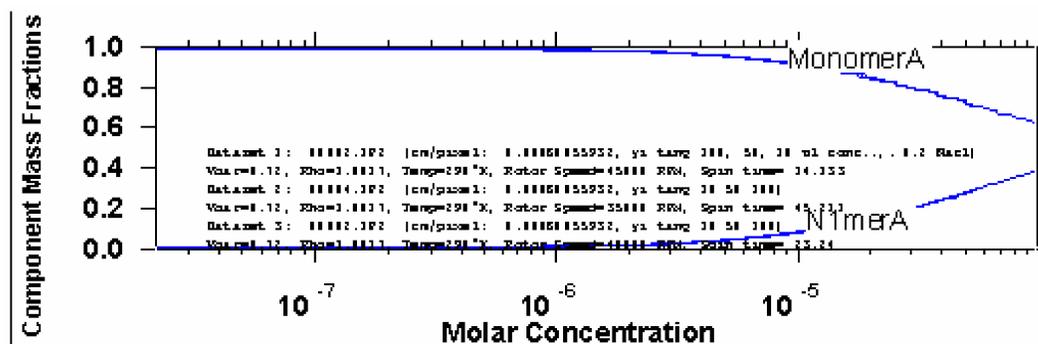


Figure S5. Species fit of HA1. *MonomerA* represents the dimeric species; *N1merA* represents tetramers.