# Supporting Information: "Cooperative Collapse" of the Denatured State Revealed through Clausius–Clapeyron Analysis of Protein Denaturation Phase Diagrams.

Alexander Tischer[1], Venkata R. Machha[1],
Jörg Rösgen[2*], and Matthew Auton[1*]

January 19, 2018

[1] Division of Hematology, Departments of Internal Medicine and Biochemistry and Molecular Biology, Mayo Clinic, Rochester, MN, USA

[2] Penn State University College of Medicine, Department Biochemistry and Molecular Biology, Hershey PA, 17033, USA

[*] To whom correspondence is addressed, auton.matthew@mayo.edu and jur19@psu.edu.

**Author Contributions:** AT expressed, purified and carboxyamidated the A3 domain. AT performed the spectroscopy, size-exclusion chromatography, trypsin proteolysis, mass-spectrometry and analyzed the primary data. VRM performed and analyzed the SPR experiments. JR and MA designed the phase diagram analysis. MA, JR and AT wrote the manuscript. MA and JR designed the research.

**Running Title:** Disulfide Bonds and Denatured State Collapse.

**Keywords:** Urea-temperature phase diagram, Circular Dichroism, Fluorescence, RCAM A3, von Willebrand factor

**Abbreviations Used:** vWF, von Willebrand Factor; CD, Circular Dichroism; FL, Fluorescence

# 1 Materials and Methods

## 1.1 Materials

Tris HCl, GdnHCl, Tween 20 and disodium hydrogen phosphate were from Fisher Scientific. EDTA and sodium acetate were purchased from ICN Biomedicals. Urea, Glycine, Iodoacetamide, DTT and NaCl were from Sigma Aldrich. All chemicals were of analytical grade or higher purity. Human placenta Collagen III was obtained from Sigma. For all spectroscopic measurements a buffer consisting of 10mM Disodium hydrogen phosphate, 10mM Glycine, 10mM sodium acetate (PGA) and 150mM NaCl, pH 8 was used.

## 1.2 Expression, purification and preparation of RCAM A3

The vWF A3 domain ($Ser_{1671}$-$Gly_{1874}$) was expressed in *E. coli* M15–cells as a fusion protein containing an N–terminal 6× Histidine–tag and refolded and purified from inclusion bodies by $Ni^{2+}$ affinity chromatography. The purity of the protein was confirmed using reducing SDS PAGE and analytical size–exclusion chromatography. RCAM A3 was prepared by an initial complete denaturation of A3 in 6M urea. After addition of 10mM EDTA and of 6mM DTT, the A3 was incubated for 3hrs. After the addition of 12mM Iodoacetamide, the sample was incubated for 1hr in the dark and the reaction was stopped by addition of 2–Mercaptoethanol followed by excessive dialysis against buffer overnight at 4°C.

## 1.3 Collagen binding via surface plasmon resonance

**Fig.1** compares the binding of A3 and RCAM A3 to collagen. SPR kinetic experiments were performed on a Biacore T–100 at 25°C. Collagen III was covalently coupled via primary amines to the active channel of a CM5 chip resulting in ∼3000RU (∼30 ng/mm$^2$ collagen). The reference channel lacking collagen was subtracted as background. The flow rate was 30$\mu$L/min and the association and dissociation phase both were 500s. Regeneration of the biochip was performed using 1mM EDTA, 2M NaCl, 100mM sodium citrate, pH 5 for 60s at 10$\mu$L/min, followed by injection of 100mM phosphoric acid for 30s at 30$\mu$L/min. The sensorgram response at the end of the association phase was used to determine the collagen binding affinity of A3 and RCAM A3 using the following **Eq.1**.

$$RU[c] = \frac{c(1/K_D)R_{max}}{c(1/K_D)+1} \tag{1}$$

## 1.4 Limited proteolysis Mass Spectrometry

Table 1 list the theoretical fragments of A3 trypsinolysis with the secondary structure regions of the fragments in the crystal structure of A3 (pdbID = 1ao3).

## 1.5 Global 2-dimensional Analysis of CD unfolding transitions.

Table 2 lists the resulting $N \rightleftharpoons D_1$ and $N \rightleftharpoons D_2$ fit parameters used in the quantitative analysis of the unfolding transitions shown in **Fig.2**. Three base–planes were fit to the CD data defining the native state, $N$ (b=1), the compact thermally–denatured state, $D_2$ (b=2), in low urea concentrations, and the expanded urea–denatured state, $D_1$ (b=3), as a function of temperature at high urea concentration. These base–planes for each state were defined as 2–dimensional Taylor expansions of $\Delta c$ and $\Delta T$.

$$S_b[T,c] = \sum_{j=0}^{2}\sum_{k=0}^{1}\left(\frac{\partial^{j+k}S_i[T,c]}{\partial T^j \partial c^k}\right)\frac{\Delta T^j \Delta c^k}{j!k!} \tag{2}$$

Once determined, the baseplanes were kept constant and the transitions were fit to a thermodynamic equation of state whose partition function, $Q = 1 + K_1 + K_2$, accounts for reversible unfolding from $N \rightleftharpoons D_1$ (i=1) and from $N \rightleftharpoons D_2$ (i=2) as a function of $c$ and $T$.

$$\ln K_i[T,c] = \sum_{j=0}^{3}\sum_{k=0}^{1}\left(\frac{\partial^{j+k}\ln K_i[T,c]}{\partial \beta^j \partial c^k}\right)\frac{\Delta \beta^j \Delta c^k}{ji!k!} \tag{3}$$

The overall equilibrium constant for unfolding, $K = K_1 + K_2$, and free energy $\Delta G = -RT \ln K$. The free energy of unfolding to each denatured state, $\Delta G_i = -RT \ln K_i$. Enthalpy was defined as the first derivative of $\ln K$ with respect to $\Delta\beta$.

$$\Delta H_i[T,c] = -\sum_{j=0}^{3}\sum_{k=0}^{1}\left(\frac{\partial^{j+k+1}\ln K_i[T,c]}{\partial\beta^{j+1}\partial c^k}\right)\frac{\Delta\beta^{j-1}\Delta c^k}{(j-1)!k!} \tag{4}$$

The overall unfolding enthalpy is defined by the individual unfolding equilibrium constants ($K_1$ and $K_2$) and individual enthalpies ($\Delta H_1$ and $\Delta H_2$).

$$\Delta H[T,c] = -\left(\frac{\partial \ln K[T,c]}{\partial\beta}\right) = \frac{K_1\Delta H_1 + K_2\Delta H_2}{K_1 + K_2} \tag{5}$$

Likewise, the $m$–value was defined as the first derivative of $\ln K$ with respect to $\Delta c$.

$$m_{K_i}[T,c] = \left(\frac{-m_i[T,c]}{RT}\right) = \sum_{j=0}^{3}\sum_{k=0}^{1}\left(\frac{\partial^{j+k+1}\ln K_i[T,c]}{\partial\beta^{j}\partial c^{k+1}}\right)\frac{\Delta\beta^{j}\Delta c^{k-1}}{j!(k-1)!} \tag{6}$$

The overall unfolding $m$–value is defined by the individual unfolding equilibrium constants ($K_1$ and $K_2$) and individual $m$–values ($m_{K_1}$ and $m_{K_2}$).

$$m_K[T,c] = \left(\frac{-m[T,c]}{RT}\right) = \left(\frac{\partial \ln K[T,c]}{\partial c}\right) = \frac{K_1 m_{K_1} + K_2 m_{K_2}}{K_1 + K_2} \tag{7}$$

The CD data as a function of $T$ and $c$ were fit to the following equation ...

$$S[T,c] = S_N P_N + S_{D_1} P_{D_1} + S_{D_2} P_{D_2} \tag{8}$$

where $P_N = 1/Q$, $P_{D_1} = K_1/Q$, and $P_{D_2} = K_2/Q$.

## 1.6 Size-Exclusion Chromatography

Given molecular weight standards of a variety of proteins, we calibrated the retention times at 20°C for elution in terms of the reciprocal retention coefficient ($1/K_d$) with respect to their known Stokes radii in **Fig.3**. Based on this calibration, the Stokes radius (SR) was calculated from the reciprocal retention coefficient $1/K_d$ of A3 and RCAM A3 at each urea concentration according to **Eq.9**.

$$1/K_d = 1.247(\pm 0.153) + 0.077(\pm 0.00432) * SR \tag{9}$$

## 1.7 Assessment of the thermal reversibility of unfolding.

The reversibility of thermal unfolding of A3 and RCAM A3 was assessed by following intrinsic protein fluorescence as a function of thermal scan rate (**Fig.4**) and by repeated 2°C/min thermal scans on the the same sample (**Fig.5**). Intrinsic reversibility was confirmed for both proteins via scan rate independent transition midpoints ($T_m$) and enthalpies ($\Delta H$). The average $T_m$ of A3 was found to be 67.5 ± 1.1 °C with an average $\Delta H$ of 363 ± 6 kJ/mol. RCAM A3 was less stable to thermal unfolding with a reduced average $T_m$ of 59.8 ± 0.8 °C and a slightly reduced $\Delta H$ of 318 ± 8 kJ/mol. Experimental verification of thermal reversibility was assessed by repeated fluorescence scans on the same sample to successively increasing temperatures (**Fig.5A**). Protein concentrations were 0.1$\mu$M and the % fluorescence recovery after scanning to the indicated temperatures was determined by measuring the fluorescence at 10°C following a 30min equilibration. **Fig.5B** demonstrates that the % fluorescence recovery of the native state was ≥50% even after scanning well beyond the transition temperature although both proteins were susceptible to thermal aggregation at high temperature.

Table 1: Theoretical fragments of the trypsinolysis of the VWF A3 domain sorted by residue sequence.

| Fragment | Residues | VWF Residues | $2^o$ Structure | Sequence |
|---|---|---|---|---|
| T1 | 1 - 2 | | | (-) MR (G) |
| T3 | 54 - 57 | $Ser_{1712}$ - $Lys_{1715}$ | $\alpha 1$ | (K) SFAK (A) |
| T4 | 58 - 62 | $Ala_{1716}$ - $Lys_{1720}$ | $\alpha 1$ | (K) AFISK (A) |
| T5 | 63 - 68 | $Ala_{1721}$ - $Arg_{1726}$ | $\alpha 1$ - loop | (K) ANIGPR (L) |
| T6 | 69 - 93 | $Leu_{1727}$ - $Lys_{1751}$ | $\beta 2$ - hairpin - $\beta 3$ - loop - $\alpha 2$ | (R) LTQVSVLQYGSITT IDVPWNVVPEK (A) |
| T7 | 94 - 105 | $Ala_{1752}$ - $Arg_{1763}$ | $\alpha 2$ - loop | (K) AHLLSLVDVMQR (E) |
| T8 | 106 - 121 | $Glu_{1764}$ - $Arg_{1779}$ | loop - $\alpha 3$ | (R) EGGPSQIGDALGFAVR (Y) |
| T9 | 122 - 136 | $Tyr_{1780}$ - $Lys_{1794}$ | $\alpha 3$ - loop - $\beta 4$ | (R) YLTSEMHGARPGASK (A) |
| T10 | 137 - 158 | $Ala_{1795}$ - $Arg_{1816}$ | $\beta 4$ - loop - $\alpha 4$ | (K) AVVILVTDVSVDSVDAAADAAR (S) |
| T11 | 159 - 161 | $Ser_{1817}$ - $Arg_{1819}$ | $\alpha 4$ - loop - $\beta 5$ | (R) SNR (V) |
| T12 | 162 - 172 | $Val_{1820}$ - $Arg_{1830}$ | $\beta 5$ | (R) VTVFPIGIGDR (Y) |
| T13 | 173 - 179 | $Tyr_{1342}$ - $Arg_{1837}$ | $\alpha 5$ | (R) YDAAQLR (I) |
| T14 | 180 - 192 | $Ile_{1838}$ - $Lys_{1850}$ | $\alpha 5$ - loop - $\beta 6$ | (R) ILAGPAGDSNVVK (L) |
| T15 | 193 - 195 | $Leu_{1851}$ - $Arg_{1853}$ | $\beta 6$ - loop - $\alpha 6$ | (K) LQR (I) |
| T16 | 196 - 212 | $Ile_{1854}$ - $Lys_{1870}$ | $\alpha 6$ | (R) IEDLPTMVTLGNSFLHK (L) |
| T18 | 218 - 219 | $Leu_{1876}$ - $Asn_{1877}$ | C-terminus | (K) LN (-) |
| Disulfide T2 | 3 - 53 | $Ser_{1671}$ - $Lys_{1711}$ | N-terminus, $\beta 1$, $\alpha 1$ | (R)GSHHHHHHGSSGEGLQIPTLSPAP DCSQPLDVILLLDGSSSFPASYFDEMK(S) |
| T17 | 213 - 217 | $Leu_{1871}$ - $Lys_{1875}$ | $\beta 1$ | (K)LCSGK(L) |

Table 2: Thermodynamic parameters defining the urea-temperature phase diagrams.

| $i=1$ $N \overset{K_1}{\rightleftharpoons} D_1$ | | Thermodynamic | $i=2$ $N \overset{K_2}{\rightleftharpoons} D_2$ | | Parameter |
|---|---|---|---|---|---|
| A3 | RCAM A3 | Parameters | A3 | RCAM A3 | Units |
| – | – | Observed $T_m$ | 342.5 | 334.4 | $K$ |
| | | Temperature Dependence of $\ln K_i$ | | | |
| $-12.21 \pm 0.03$ | $-7.02 \pm 0.01$ | $\ln K_i$ | 0 [a] | 0 [a] | |
| $-159.9 \pm 55$ | $-129.4 \pm 69$ | $\partial \ln K_i / \partial \beta$ | $-400.7 \pm 6.7$ | $-349.8 \pm 13$ | kJ/mol |
| $-3622 \pm 1827$ | $-297 \pm 2489$ | $\partial^2 \ln K_i / \partial \beta^2$ | $3776 \pm 1047$ | $3358 \pm 2103$ | $(\text{kJ/mol})^2$ |
| $231188 \pm 28691$ | $120795 \pm 41409$ | $\partial^3 \ln K_i / \partial \beta^3$ | 0 | 0 | $(\text{kJ/mol})^3$ |
| | | Temperature Dependence of $m_{K_i}$ | | | |
| $2.63 \pm 0.28$ | $2.24 \pm 0.18$ | $\partial \ln K_i / \partial c$ | $0.779 \pm 0.006$ | $1.007 \pm 0.11$ | kJ/mol $\text{M}^{-1}$ |
| $27.9 \pm 8.8$ [b] | $18.4 \pm 6.9$ | $\partial^2 \ln K_i / \partial \beta \partial c$ | $27.9 \pm 8.8$ [b] | $25.9 \pm 8.2$ | $(\text{kJ/mol})^2$ $\text{M}^{-1}$ |
| $-237.3 \pm 129.9$ [b] | $-146.6 \pm 120.4$ | $\partial^3 \ln K_i / \partial \beta^2 \partial c$ | $-237.3 \pm 129.9$ [b] | $-237.3 \pm 129.9$ [b] | $(\text{kJ/mol})^3$ $\text{M}^{-1}$ |

All parameters are referenced to the indicated $T_m$ in the absence of urea. $\beta = 1/RT$ (mol/kJ).
[a] $\ln K_2$ at the $T_m$ is 0 by definition.
[b] Temperature dependence of the pre- and post- sigmoid transition of $m_{K_i}$ constrained to be equal for simplicity in fitting.

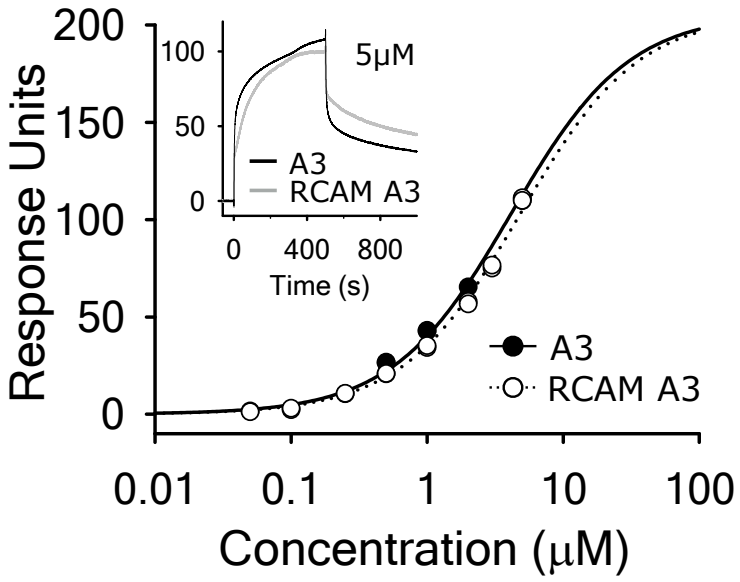


Figure 1: Surface plasmon resonance of the binding of A3 and RCAM A3 to collagen III. $K_D$ for A3 $= 4.1 \pm 0.4\ \mu\text{M}^{-1}$ and $K_D$ for RCAM A3 $= 4.8 \pm 0.4\ \mu\text{M}^{-1}$. $R_{max} = 206 \pm 10$ response units. *Inset:* Representative sensorgrams at $5\mu$M concentration of A3 and RCAM A3.

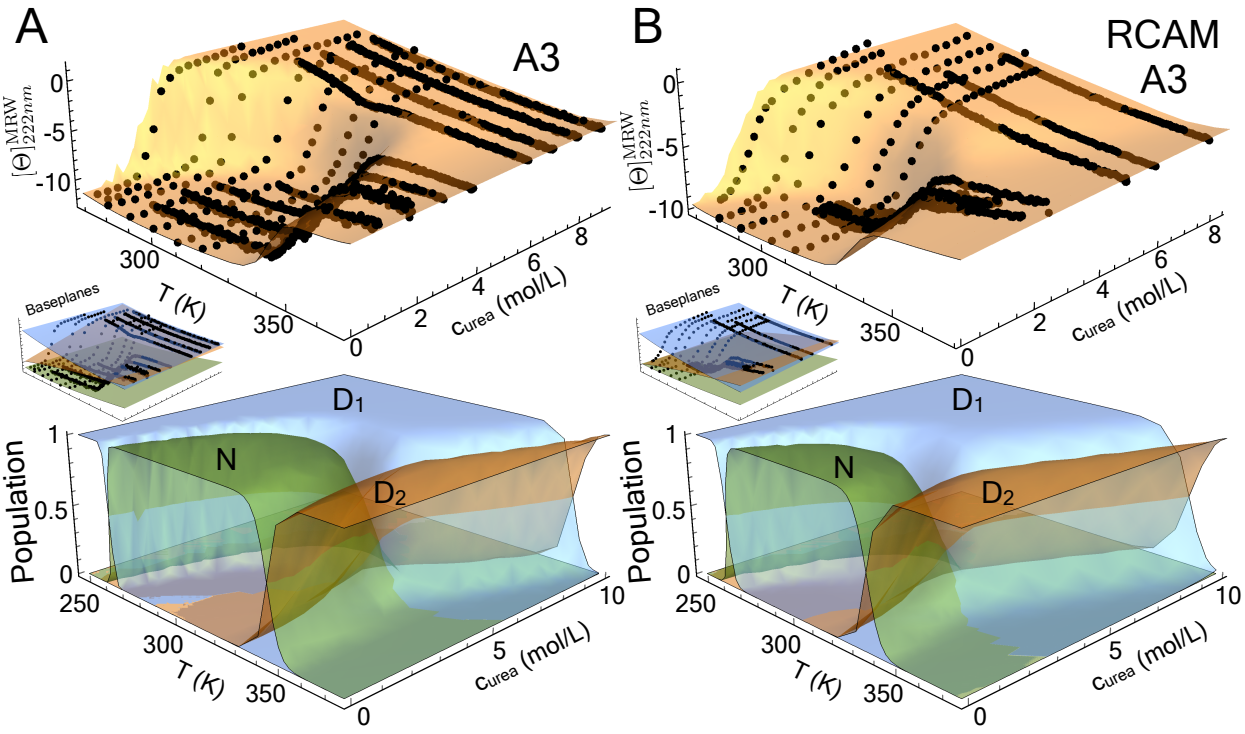


Figure 2: Global 2–dimensional analysis of the CD unfolding data. *A)* A3 domain. *B)* RCAM A3 domain. *Top* – Combined urea and thermal unfolding data with the resulting fit described by **Eq.8**. *Bottom* – Population of native (N, green) and denatured states, ($D_1$, blue; $D_2$, orange). *Insets:* Base planes defining the CD as a function of temperature and urea concentration for native (green), expanded urea–denatured (blue), and compact thermally–denatured (orange) states.

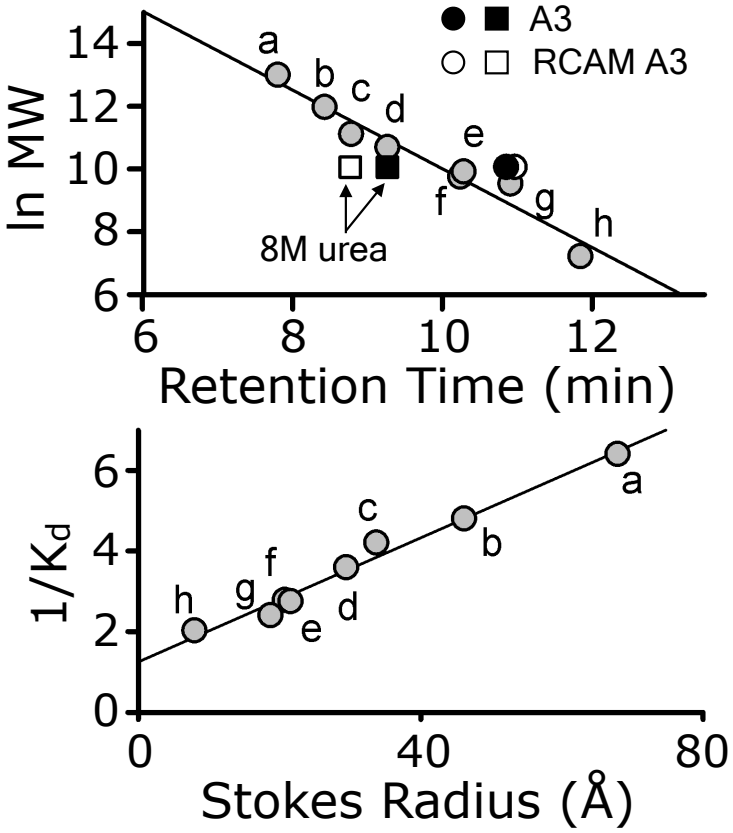


Figure 3: Molecular weight calibration curve for standard proteins (top) and $1/K_d$ as a function of Stokes radius (bottom). Proteins used for the calibration curves were: Ferritin (a), Gamma Globulin (b), Bovine Serum Albumin (c), Ovalbumin (d), Soybean Trypsin Inhibitor (e), Myoglobin (f), Ribonuclease (g), and B12 (h). *Top panel:* Retention times of A3 and RCAM A3 in presence and absence of 8M urea.

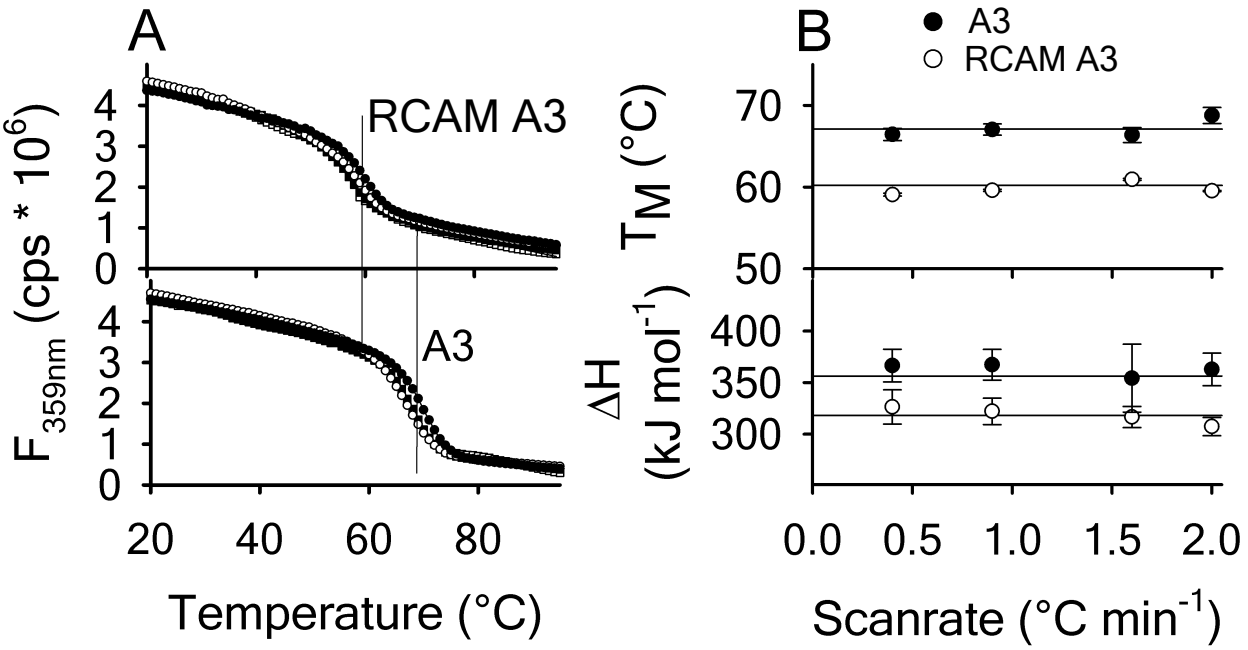


Figure 4: Thermal unfolding scan rate dependence of A3 and RCAM A3. *A)* Thermal unfolding of 1$\mu$M RCAM A3 (top) and A3 (bottom) measured with fluorescence ($\lambda_{Ex}$ = 280nm, $\lambda_{Em}$ = 359nm) at scan rates of 2 (●), 1.6 (○), 0.9 (■), and 0.4°C/min (□). *B)* Constant $T_m$ and $\Delta H$ as a function of the instrumental thermal scan rate indicates reversible thermal unfolding. Lines represent average values from all scan rates.

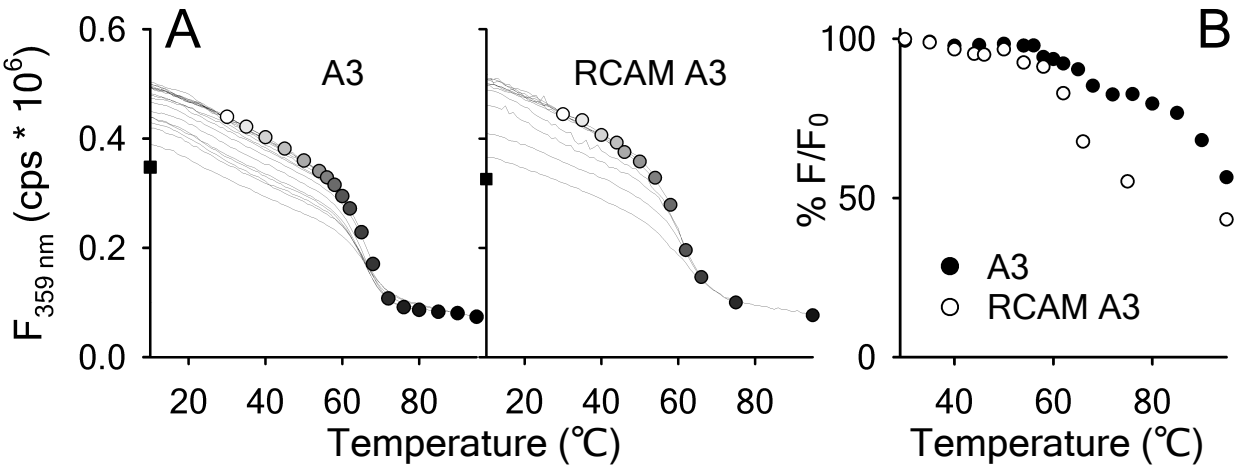


Figure 5: Experimental thermal unfolding reversibility of A3 and RCAM A3. *A)* Test for reversibility of thermal unfolding of $0.1\mu M$ A3 and RCAM A3 using fluorescence at a scan rate of 2.0°C/min. Repeated thermal scans were carried out on the same sample from 10°C to the indicated temperature followed by a drop back and an equilibration phase of 30 min at 10°C. *B)* Percent reversibility at 10°C after scanning to the indicated temperature. Loss of fluorescence is due to protein aggregation at high temperature.