Supporting Information: Flipping Out: Role of Arginine in Hydrophobic Interactions and Biological Formulation Design

Jonathan W. P. Zajac,[†] Praveen Muralikrishnan,[‡] Idris Tohidian,[¶] Xianci Zeng,[§] Caryn L. Heldt,[¶] Sarah L. Perry,[§] and Sapna Sarupria[*,†]

†*Department of Chemistry, University of Minnesota, Minneapolis, MN 55455, USA*

‡*Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN 55455, USA*

¶*Department of Chemical Engineering, Michigan Technological University, Houghton, MI 49931, USA*

§*Department of Chemical Engineering, University of Massachusetts Amherst, MA 01003, USA*

‖*Chemical Theory Center, University of Minnesota, Minneapolis, MN 55455, USA*

E-mail: sarupria@umn.edu

1 Simulation Details

A cubic box of length of 6.74 nm was constructed with a padding of 1.5 nm between the edge of the fully extended polymer and the nearest box edge. Chloride (Cl^-) counterions equal to the number of arginine molecules were added to achieve a net charge of zero. The TIP4P/2005[S1] model was used for water, and the CHARM22 force field was used for arginine and Cl^-.[S2] Lorentz-Berthelot mixing rules[S3,S4] were used to calculate non-bonded interac-

tions between different atom types, except polymer-water oxygen interactions (Table S2). Polymer-water oxygen interactions were adjusted iteratively until the folded and unfolded states of the polymer were approximately evenly distributed in straightforward MD simulations. The various Lennard-Jones parameters tested are presented in Table S2. Guided by radius of gyration (R_g) probability distributions, we selected parameters of model 2 for our study (Figure S2).

Table S1: Setup of simulated systems. Simulation time for REUS simulations are reported as $N \times M \times Q$, which represent the number of replicate simulations (N), the number of windows per replica (M), and the simulation length in each window (Q), respectively.

System	Simulation Time (ns)	Concentration (M)	N_{Exc}	N_{Wat}
Arginine	20	0.25	47	9653
Arginine	20	0.50	93	9111
Arginine	20	0.75	139	8582
Arginine	20	1.0	185	7933
Polymer	3 x 100 x 12	0.00	0	10599
Polymer + Arginine	3 x 12 x 100	0.25	47	10092
Polymer + Arginine	3 x 12 x 100	0.50	93	9511
Polymer + Arginine	3 x 12 x 250	1.0	185	8398
Polymer + Guanidinium	3 x 12 x 50	0.25	47	10364
Polymer + Guanidinium	3 x 12 x 50	0.50	93	10144
Polymer + Guanidinium	3 x 12 x 50	1.0	185	9702
Polymer + Glycine	3 x 12 x 50	0.25	47	10318
Polymer + Glycine	3 x 12 x 50	0.50	93	10022
Polymer + Glycine	3 x 12 x 50	1.0	185	9444

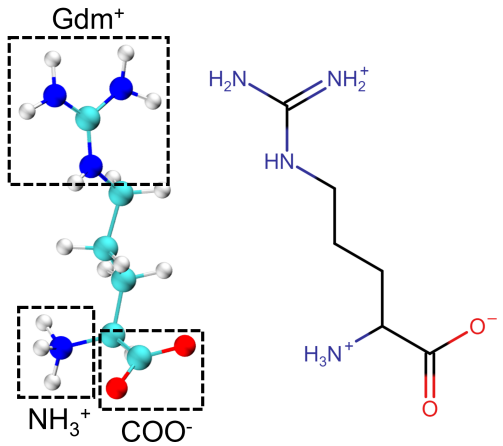

Figure S1: Representation of the structure of arginine. Boxes are drawn around the charged groups of arginine.

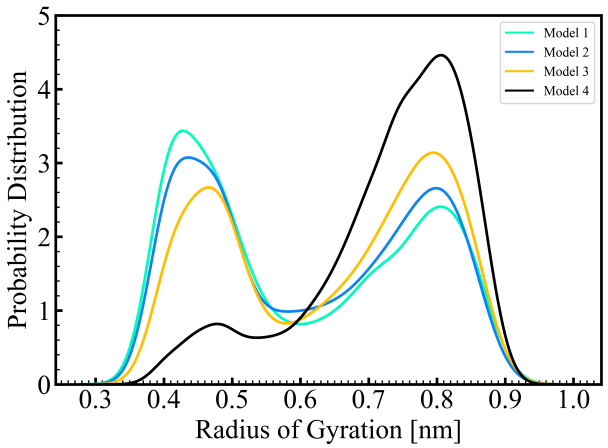

Figure S2: Probability distribution of radius of gyration obtained from 50 ns simulations of different polymer models in pure water. The models differ in their polymer-water interaction parameter, ϵ, having 85% (model 1), 88% (model 2), 92% (model 3), and 100% (model 4) of the value calculated from Lorentz-Berthelot mixing rules.

Table S2: Polymer interaction parameters used in the present study.

Interaction	Model	Sigma (nm)	Epsilon (kJ/mol)
Polymer-Polymer	All	0.373	0.586
Polymer-Water	Model 1	0.345	0.573
Polymer-Water	Model 2	0.345	0.593
Polymer-Water	Model 3	0.345	0.620
Polymer-Water	Model 4	0.345	0.674

REUS simulations were performed in 12 evenly-spaced windows along the R_g reaction coordinate, spanning 0.35 nm to 0.9 nm. Each window was biased according to a harmonic potential, with a force constant of 1000 kJ/mol/nm^2 for the window centered at 0.45 nm (window 3) and 5000 kJ/mol/nm^2 for all other windows. We observed inefficient sampling in window 3 region (Fig. S3). Subsequent simulations with varying force constants for window 3 regions revealed that a force constant of 1000 kJ/mol/nm^2 minimized differences between replicate runs in the regions close to the window centers.

The windows are first energy minimized using the steepest descent minimization with a tolerance of 10 kJ/mol/nm and step size of 0.01. For each window, 1 ns NVT equilibration is then performed using V-rescale thermostat (temperature coupling time constant, $\tau_T = 0.5$

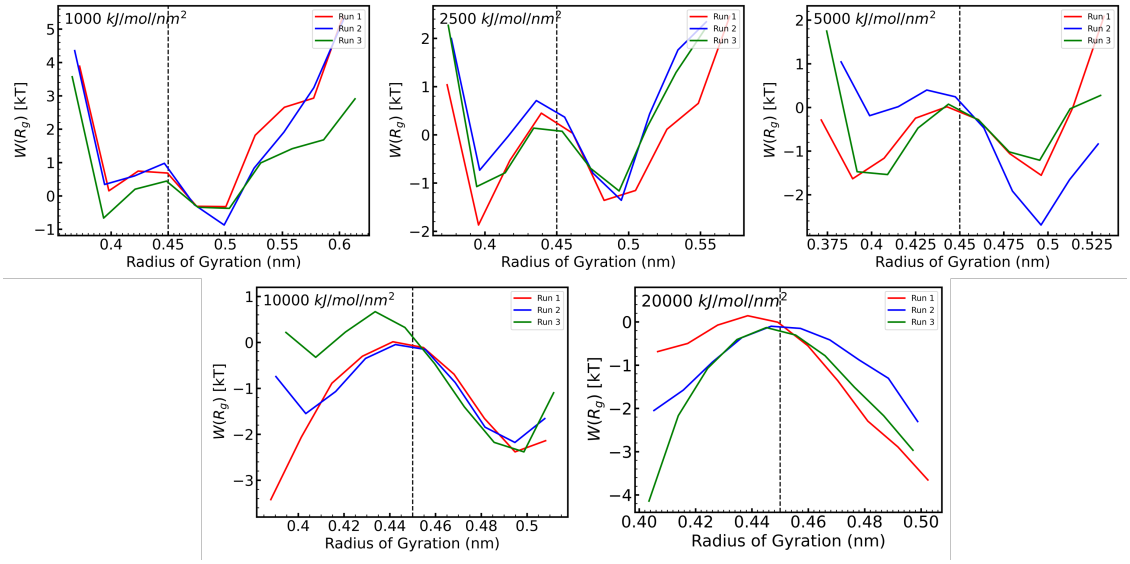

Figure S3: Sampling of 3 replicate runs in the window 3 (reference radius of gyration = 0.45 nm) region for polymer in 0.75M arginine solution with different force constants ranging from 1000 - 20000 $kJ/mol/nm^2$. For regions close to the reference, the uncertainty between runs is lesser for the lower force constant values. Force constant 1000 $kJ/mol/nm^2$ was chosen for window 3 based on these observations

ps),[S5] followed by 1 ns NPT equilibration using the V-rescale thermostat (τ_T = 0.5 ps)[S5] and Berendsen barostat (τ_P = 0.5 ps)[S6] to bring the system to a temperature of 300 K and pressure of 1 atm. NPT production run for 100 ns is simulated for each window using Nosé-Hoover temperature coupling (τ_T = 5 ps)[S7] and Parrinello-Rahman pressure coupling (τ_P = 25 ps).[S8] A Hamiltonian exchange move is attempted every 200 timesteps, with a 2 fs time step. The Particle Mesh Ewald (PME) algorithm was used for electrostatic interactions with a cut-off of 1 nm. A reciprocal grid of 42 x 42 x 42 cells was used with 4th order B-spline interpolation. A single cut-off of 1 nm was used for van der Waals interactions. The neighbor search was performed every 10 steps.

To further investigate the hypothesis that attractive polymer-arginine interactions are driven by the guanidinium sidechain while indirect effects are driven by the backbone, additional independent REUS simulations including either guanidinium or glycine as the additive were carried out. Guanidinium parameters were based on the CHARMM22 parameters of arginine. This was achieved by truncating an arginine molecule up to the first guanidinium

nitrogen, protonating this atom, and imposing a symmetric charge distribution according to the existing parameters. Glycine parameters were taken directly from the CHARMM22 force field. Systems in the same concentration range as arginine were generated to study sidechain and backbone contributions to hydrophobic polymer collapse.

2 PMF Convergence Checks

PMF convergence for hydrophobic polymer folding/unfolding was assessed by comparing the Kolmogorov-Smirnov statistic between either time-lagged PMFs or between replicate REUS simulations (Fig. S4). Simulations were stopped once the deviation between replicate PMFs was observed to fluctuate around an average value less than 0.2, resulting in total simulation times per replica of between 100-250 ns. Across 3 replicate simulations, each with 12 windows, an aggregate simulation time of 3.6-9.0 μs was carried out for each system.

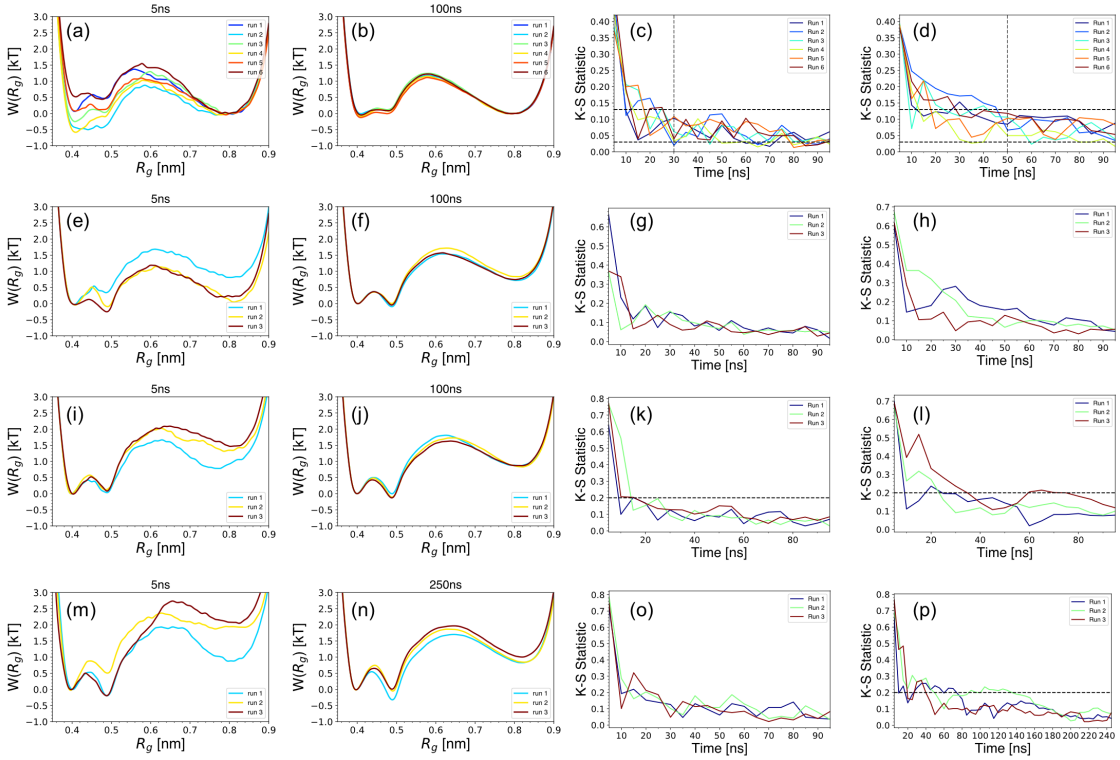

Figure S4: PMF convergence checks for all REUS simulations. Hydrophobic polymer in (a-d) water, (e-h) 0.25 M arginine, (i-l) 0.5 M arginine, and (m-p) 1.0 M arginine. In the first column, PMFs obtained after 5 ns per replica are shown. In the second column, PMFs obtained after 100 or 250 ns are shown, highlighting converged replicate simulations. In column 3, Kolmogorov-Smirnov (K-S) statistics are plotted as a comparison for PMFs obtained in 5 ns blocks (i vs $i+1$). In column 4, K-S statistics are plotted as a comparison between PMFs at time t vs the PMF obtained at either 100 ns (d,h,l) or 250 ns (p).

3 PPV System Validation

To ensure our surface representation of the PPV capsid accurately reflects the dynamics of a fully assembled PPV capsid, we carried out an additional production run of a fully assembled PPV capsid system (PDB: 1K3V). In the NVT ensemble at 300 K, even after 300 ns of simulation time (18 μs aggregated across monomers) and over 500,000 CPU hours, structural equilibration of the entire capsid was not reached (Fig. S5c). Therefore, we utilized a locally-stable equilibrium state of the trajectory from 50-100 ns for further analysis. This enabled a reasonable comparison of our PPV surface model and a fully assembled capsid under the same parameters and simulation conditions.

To quantify the global and local dynamics present in these systems, we computed dynamic cross-correlation (d_{corr}) matrices of either intra-monomer residue-residue fluctuations (Fig. S7) or monomer-monomer fluctuations (Fig. S8). Ultimately, we find that, during the unrestrained simulation, our surface model accurately reflects intra-monomer residue-residue dynamics observed in a fully assembled PPV capsid model. With respect to monomer-monomer correlations, however, we observe that the surface representation does not fully capture correlated motions of adjacent monomers.

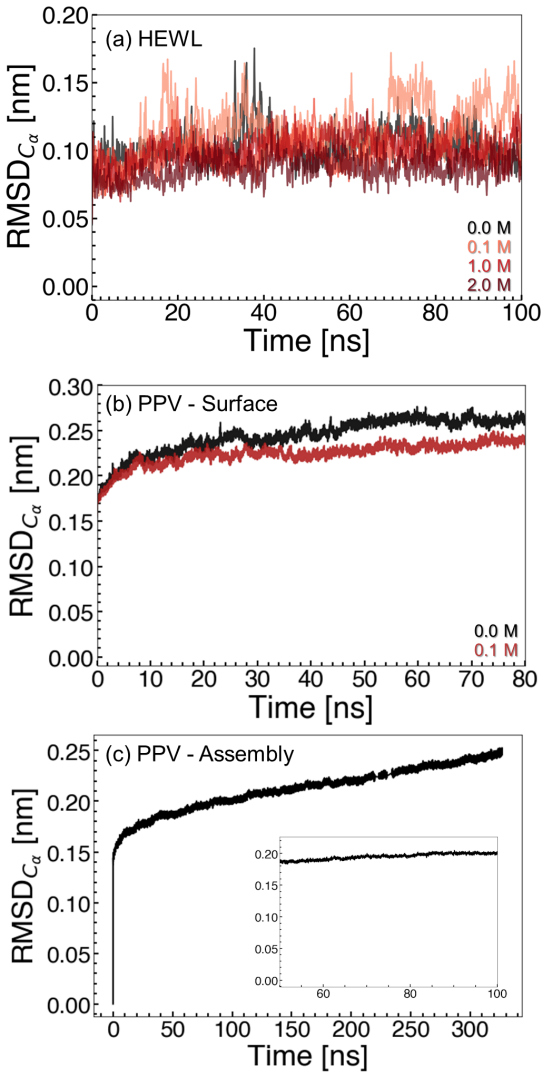

Figure S5: RMSD$_{C_\alpha}$ for larger biomolecular models, (a) HEWL (b) PPV surface model, and (c) PPV fully-assembled capsid (inset, c) local steady state. Systems with arginine as an excipient are colored in red, while the macromolecule in water is colored in black.

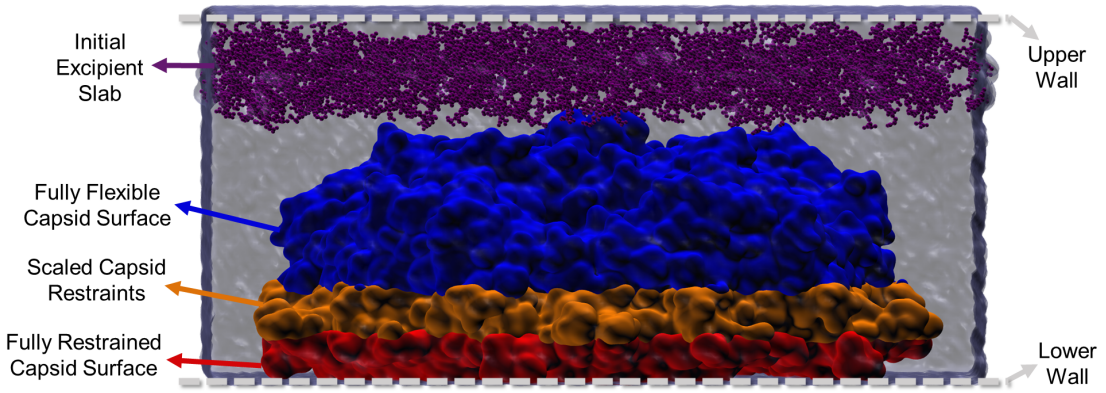

Figure S6: Schematic detailing the PPV 15-mer surface simulation.

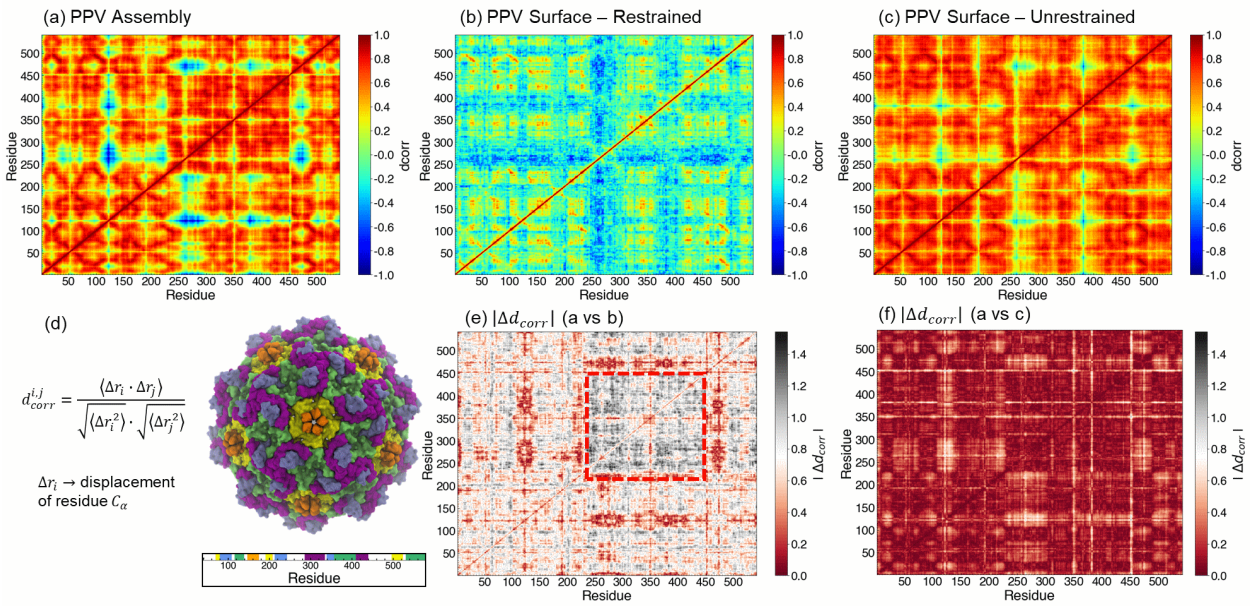

Figure S7: Dynamic cross-correlation of capsid protein residues for (a) PPV assembly, (b) PPV restrained surface, and (c) PPV unrestrained surface systems. (d) Representation of the distinct structural units present in the capsid systems, and approximate residue compositions. (e) Absolute difference in d_{corr} between plots (a) and (b), reflecting similarities of residue-residue dynamics in the fully assembled capsid and restrained surface model. (f) Absolute difference in d_{corr} between plots (a) and (c), reflecting similarities of residue-residue dynamics of the fully assembled capsid and unrestrained surface model. In (e) and (f), red regions highlight a high similarity between models. In (e), the dashed box denotes a region of high dissimilarity between the fully assembled capsid and restrained surface model.

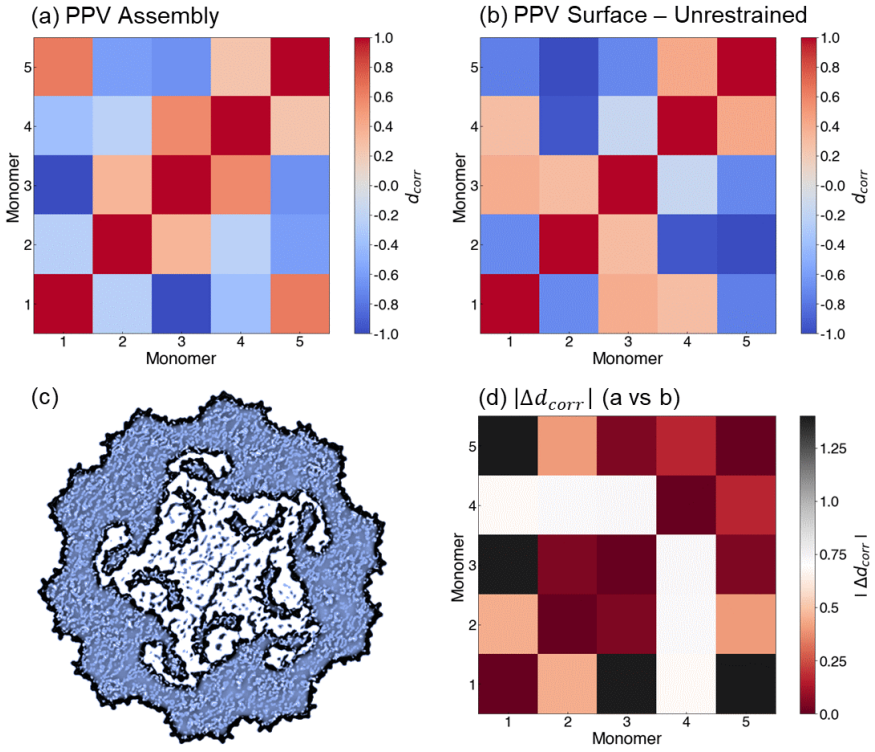

Figure S8: Dynamic cross-correlation of capsid protein centers-of-mass for (a) PPV assembly and the (b) PPV unrestrained surface model. (c) Representation of the pentameric unit used in generating each plot. (d) Absolute difference in d_{corr} between plots (a) and (b), reflecting similarities of the monomer-monomer dynamics in the fully assembled capsid and unrestrained surface model.

4 Error Calculations

The errors for PMF were calculated through the propagation of uncertainty using 3 replicate simulations (N = 3). The derivation of uncertainty in the free energy of unfolding is shown below. σ represents the standard deviation, exp represents the exponential term, ln represents the logarithmic term and int represents the integral.

$$\Delta G_{\text{unfold}} = k_B T \ln \frac{\int_{R_g^{\text{cut}}}^{R_g^{\max}} \exp\left(-\frac{W(R_g)}{k_B T}\right) dR_g}{\int_{R_g^{\min}}^{R_g^{\text{cut}}} \exp\left(-\frac{W(R_g)}{k_B T}\right) dR_g} \tag{S1}$$

The integral is approximated as a sum and divided into discrete bins in the Rg coordinate. The Rg space (from 0.3 to 0.9 nm) is divided into 600 bins, giving a $\Delta R_g = 0.001$ nm.

$$\sigma_{W(R_g)} = \sqrt{\frac{\sum \left(W\left(R_g\right)_i - \mu_{W(R_g)}\right)^2}{N}} \tag{S2}$$

$$\sigma_{\exp} = \left|\exp\left(-\frac{W\left(R_g\right)}{k_B T}\right)\right| * \left|\frac{1}{k_B T} * \sigma_{W(R_g)}\right| \tag{S3}$$

$$\sigma_{\text{int}} = \Delta R_g * \sqrt{\sum \sigma_{\exp}^2} \tag{S4}$$

$$\sigma_{ln} = \frac{\sigma_{\text{int}}}{\text{int}} \tag{S5}$$

$$\sigma_{\Delta G} = k_B T * \sqrt{(\sigma_{ln})_{num}^2 + (\sigma_{ln})_{den}^2} \tag{S6}$$

The errors in PMF decomposition were calculated using error propagation rules. An example of error calculation for ΔE_{unfold} is shown below:

$$\Delta E_{unfold} = \langle E \rangle_u - \langle E \rangle_f \tag{S7}$$

$$\langle E\rangle_f = \frac{\sum_{r_{\min}}^{r_{\text{cut}}} E\left(R_g\right) P\left(R_g\right)}{\sum_{r_{\min}}^{r_{\text{cut}}} P\left(R_g\right)}, \quad \langle E\rangle_u = \frac{\sum_{r_{\text{cut}}}^{r_{\max}} E\left(R_g\right) P\left(R_g\right)}{\sum_{r_{\text{cut}}}^{r_{\max}} P\left(R_g\right)} \tag{S8}$$

$$\sigma_{E(R_g)} = \sqrt{\frac{\sum \left(E\left(R_g\right)_i - \mu_{E(R_g)}\right)^2}{N}} \tag{S9}$$

$$\sigma_{\langle E\rangle} = \frac{\sqrt{\sum_{r_{\min}}^{r_{\text{cut}}} \sigma_{E(R_g)}^2 P\left(R_g\right)^2}}{\sum_{r_{\min}}^{r_{\text{cut}}} P\left(R_g\right)} \tag{S10}$$

$$\sigma_{\Delta E} = \sqrt{\sigma_{int,f}^2 + \sigma_{int,u}^2} \tag{S11}$$

5 Clustering Analysis

Clustering was achieved via the leaf algorithm of HDBSCAN.[S9] The minimum cluster size parameter was set to 100, while the minimum samples parameter was set to 50. Clustering was carried out on the principal moments of the gyration tensor of the hydrophobic polymer. Data were obtained from the final 100 ns in each window (3.6 μs total), saving coordinates every 100 ps. Data points not belonging to clusters were removed, for clarity. Clusters identified in principal moment space were projected onto end-to-end vs radius of gyration space. Representative snapshots are shown in Fig S9 to illustrate the configurations obtained in each cluster. Clusters at $R_g = 0.4$ and $R_g = 0.5$ are separated by a free energy barrier in the calculated PMFs.

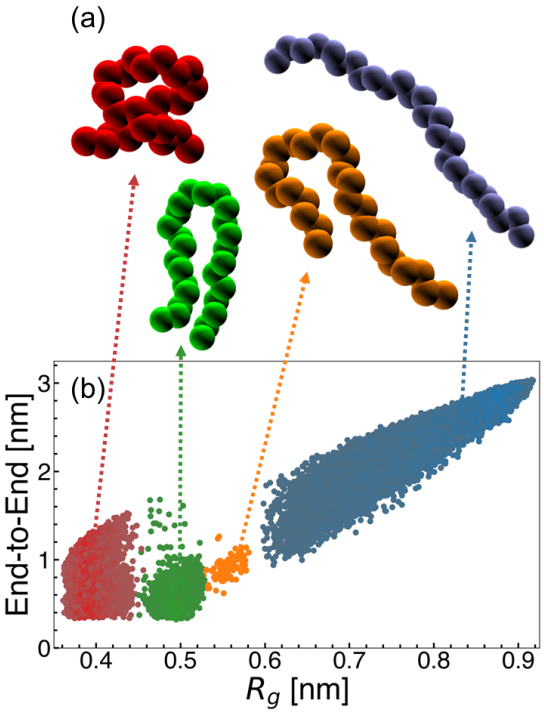

Figure S9: (a) Representative configurations from HDBSCAN clustering in 0.25 M arginine solution. (b) Polymer configurations projected onto end-to-end distance and radius of gyration space.

6 Additional Solvent Characterizations

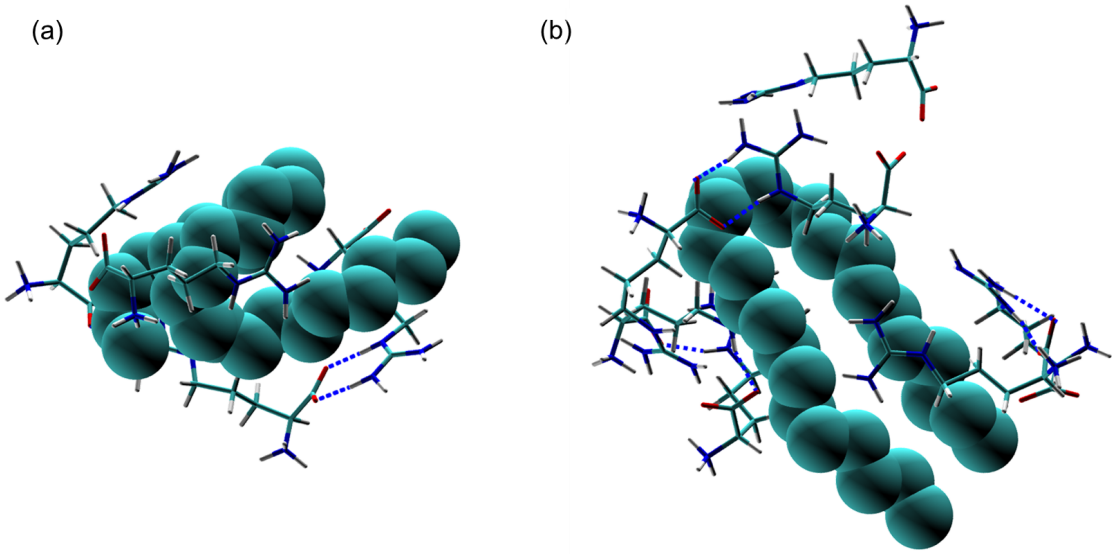

Figure S10: Representative snapshots of arginine encapsulating structures observed at the hydrophobic polymer surface. Snapshots extracted from the hydrophobic polymer in (a) unfolded and (b) folded REUS windows.

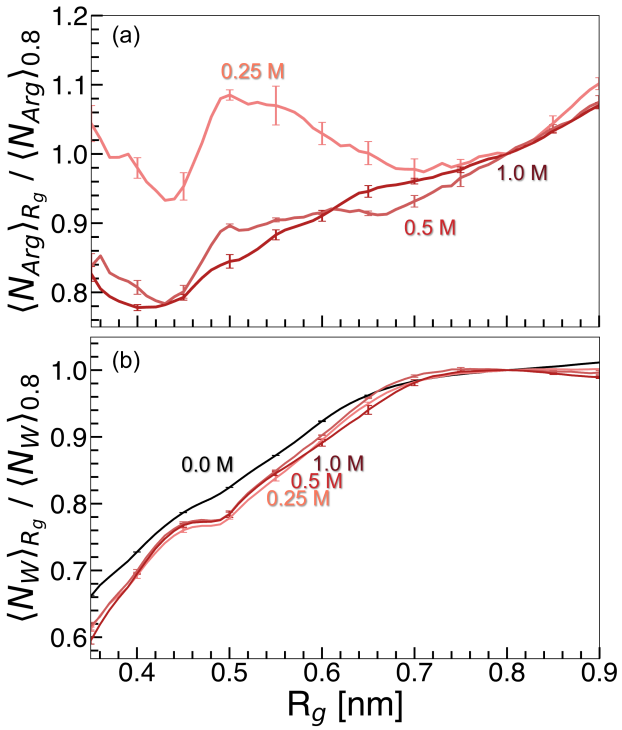

Figure S11: Quantification of arginine molecules and water molecules in the local domain of the hydrophobic polymer (within 0.5 nm). (a) Average number of arginine molecules in a given R_g window. (b) Average number of water molecules in a given R_g window. Values are normalized by the average value obtained at $R_g = 0$ nm. Means are estimated as the average value in a given bin for three replicate REUS simulations. Concentration is denoted by increased shading (light to dark).

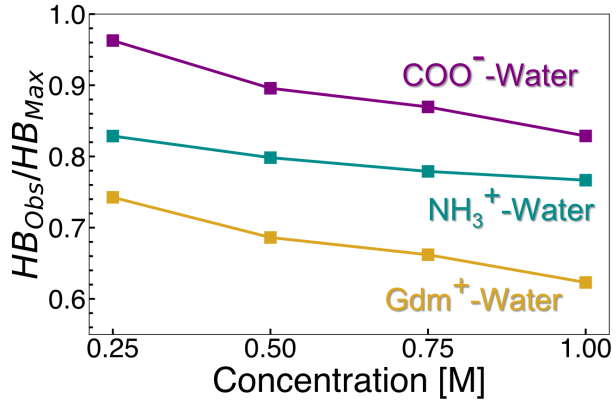

Figure S12: Fraction of observed hydrogen bonds (HB_{Obs}) relative to the maximum number of hydrogen bonds (HB_{Max}) per interaction group.

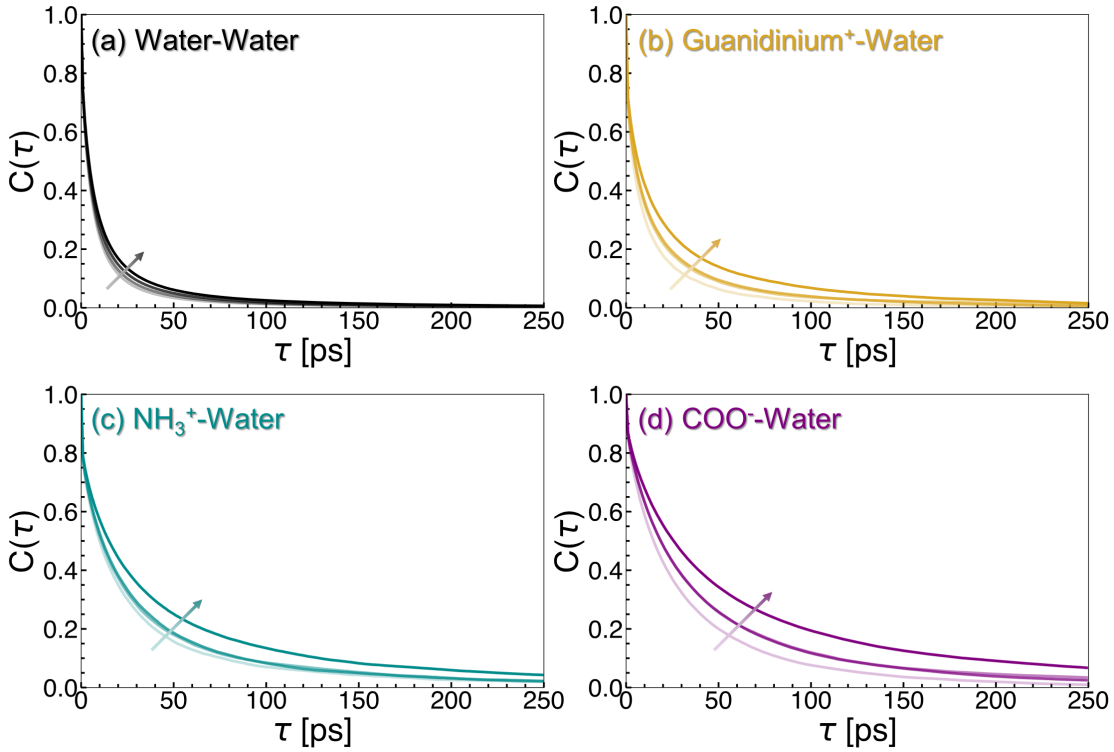

Figure S13: Hydrogen bond existence correlation functions for (a) water-water, (b) guanidinium$^+$-water, (c) NH_3^+-water, and (d) COO^--water. Each plot is shown as a function of concentration, with increased shading (light to dark) denoting increasing arginine concentration.

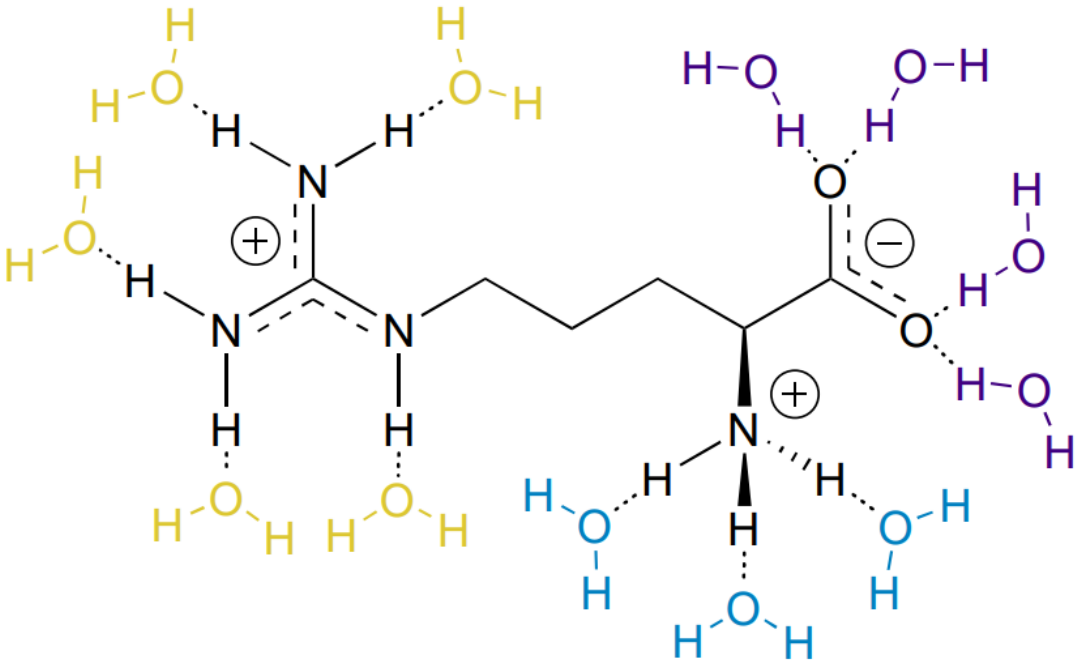

Figure S14: Illustration of arginine-water hydrogen bond interactions. Water molecules interacting with the Gdm^+ sidechain are highlighted in yellow, while those interacting with NH_3^+ and COO^- are shaded in blue and purple, respectively.

7 Preferential Interaction Coefficients

In the main text, we denote water, polymer, and additive as W, P, and A, respectively. Here, we follow traditional notation found in literature, denoting water, polymer, and additive as 1, 2, and 3, respectively. At higher concentrations, no preference for folded versus unfolded conformations was observed. Cl^- was found to preferentially deplete from the local domain of the polymer at both high and low concentrations (Fig. S15), as expected. For a binary electrolyte such as ArgCl, the net preferential interaction coefficient is obtained as[S10]

$$\Gamma_{23} = 0.5(\Gamma_{23}^{-} + \Gamma_{23}^{+} - |Z|) \tag{S12}$$

where $\Gamma_{23,-}$ denotes the preferential interaction coefficient for the anion, $\Gamma_{23,+}$ for the cation, and Z is the charge of the solute (for the polymer, $Z = 0$).

The net preferential interaction coefficient of the binary electrolyte ArgCl is reported in Fig. S15. The observed increase in Γ_{23}^{ArgCl} with increasing concentration is in contrast to experimental evidence suggesting arginine tends to preferentially interact with proteins at low concentrations and becomes excluded with increasing concentration.[S11–S14] Our findings suggest that this concentration-dependent behavior of arginine is likely not mediated by the presence of hydrophobic interaction sites.

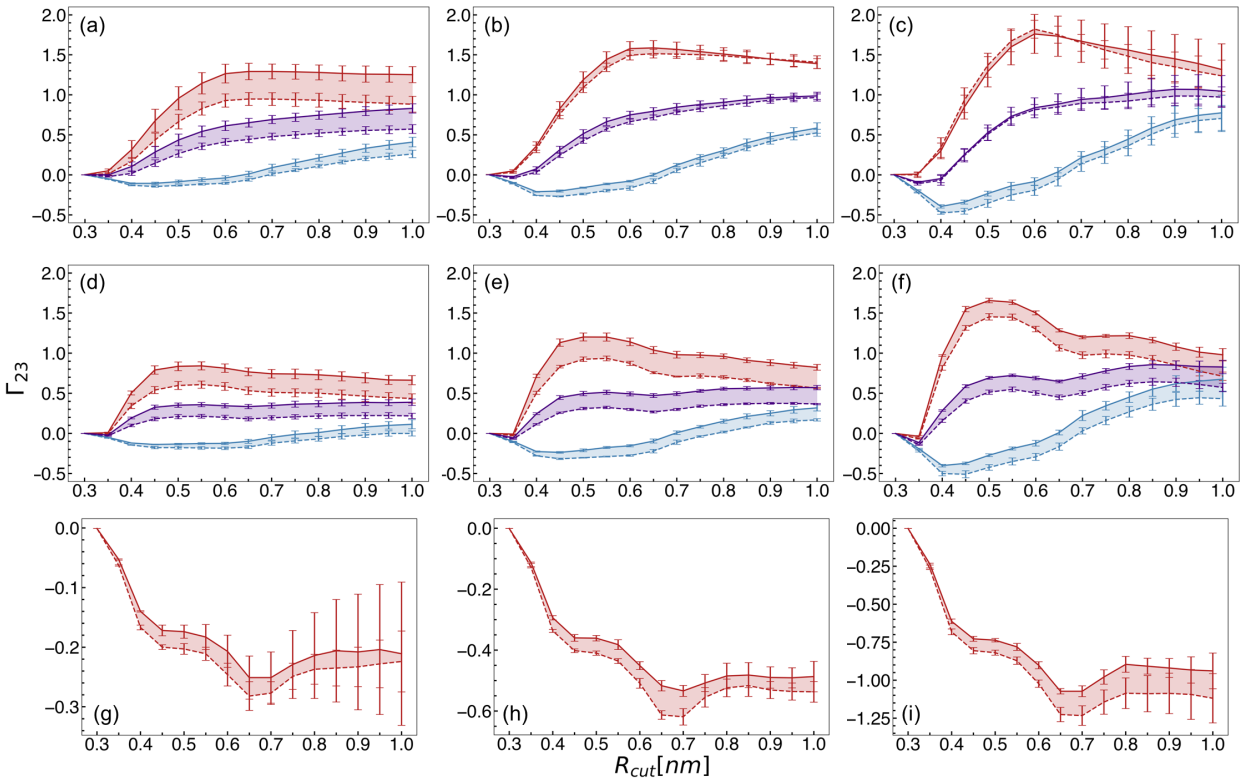

Figure S15: Preferential interaction coefficients for (a-c) arginine, (d-f) guanidinium, (g-i) glycine solutions. The additive is colored in red, counterion (if present) is colored in blue, and the net preferential interaction coefficient is colored in purple. Dashed lines indicate values for the unfolded state, while solid lines denote the folded state. Increasing arginine concentration is denoted by increased shading (light to dark). Mean values are reported from three replicate REUS simulations. Error bars were estimated as standard deviations from three replicate simulations.

8 Experimental Details

8.1 Temperature Stability of PPV

Materials

Eagle's minimum essential media (EMEM), sodium bicarbonate (7.5% solution), penicillin/streptomycin (pen/strep, 10,000 U/ml), fetal bovine serum (FBS, qualified, USDA-approved regions), phosphate-buffered saline (1 X PBS, pH 7.2), and trypsin/EDTA (0.25%) used for cell culture were purchased from Gibco™ (Grand Island, NY). MTT (2-(3,5-diphenyltetrazol-2-ium-2-yl)- 4,5-dimethyl-1,3-thiazole; bromide, 98%) and sodium dodecyl sulfate (SDS, BioReagent, ≥98.5%) were purchased from Fisher Scientific (Waltham, MA) for virus titration. Arginine monohydrochloride (reagent grade, ≥98% (HPLC)) was purchased from Millipore Sigma (Burlington, MA) as the stabilizing excipient. Sodium phosphate monobasic monohydrate (reagent ACS grade) was purchased from Millipore. Sodium phosphate dibasic heptahydrate (ACS reagent, ≥98.0%) was purchased from Sigma Aldrich (St. Luis, MO).

Methods

Cell line and virus

Porcine kidney cells (PK-13) were purchased from the American Type Culture Collection (ATCC®) (cat# CRL-6489™) and cultured in EMEM supplemented with 10 v/v% FBS and 1 v/v% pen/strep. The cells were incubated at 37 °C, 5% CO2, and 100% relative humidity. Porcine parvovirus (PPV) strain NADL-2 was a generous gift from Dr. Ruben Carbonell at North Carolina State University (Raleigh, NC). PPV strain NADL-2 was propagated in PK-13 cells using a previously established method.[S15] After three freeze-thaw cycles, the cell lysate was clarified by centrifugation at 5,000 rpm at 4 °C for 15 minutes in an ST16R centrifuge with a TX-400 swing-bucket rotor (Thermo Scientific (Waltham, Ma)). The PPV-containing supernatant was stored at -80 °C prior to use.

Virus quantification

The titer of PPV was found by the MTT colorimetric cell viability assay.[S16] PK-13 cells were seeded at a density of 8×10^4 cells/mL in 96-well plates and incubated overnight. The next day, the cells were infected with a 1:5 serial dilution of samples. After six days, 5 mg/mL of MTT in 1X pH 7.2 PBS was added to each well. Four hours later, 10 w/v% SDS with 0.01 M hydrochloric acid (HCl) was added to each well and the absorbance at 550 nm was measured with a Synergy^{TM} Mx microplate reader from BioTek (Winoski, VT) the next day. The 50% viral infectious dose was determined in units of MTT_{50}/mL.

Liquid viral sample preparation

The excipient solutions were made by dissolving different concentrations of arginine monohydrochloride in phosphate buffer containing 1.54 mM sodium phosphate monobasic monohydrate and 2.71 mM sodium phosphate dibasic. The virus samples were made by adding 10 v/v% viral stock solutions to the excipient solution.

Thermostability studies

Liquid samples were prepared in triplicates and were put either in a heat block at 60 °C[S17] or in a fridge at 4 °C as the control samples. 72 hours later, the titer of virus in each sample was determined using the MTT assay.

8.2 Temperature Stability of HEWL

Materials

Hen egg white lysozyme (HEWL $\geq$ 95%) was purchased from Hampton Research. 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) (HEPES $\geq$ 99%), hydrochloric acid (HCl, ACS grade), and sodium hydroxide (NaOH, ACS grade) were purchased from Fisher Scientific. L-arginine hydrochloride ($\geq$ 99%) was purchased from Sigma Aldrich. The Protein Thermal Shift^{TM} Dye Kits) were purchased from Applied Biosystems.

Methods

Stock Solution Preparation

Stock solutions of 1 M NaOH and 1 M HCl were prepared gravimetrically in DI water.

A stock solution of 10 mM HEPES was prepared gravimetrically in DI water and adjusted to pH = 7.00 ± 0.03 with HCl and NaOH, as needed (Thermo Scientific ROSS Sure-Flow Combination pH). Stock solutions of 1.24 mM HEWL (18 mg/mL) and 1.64 M arginine were prepared in 10 mM HEPES. A stock solution of 50X Sypro Orange was prepared using dye and buffer provided in the Protein Thermal Shift kit.

Test Sample Preparation

Samples were prepared as in Table S3 by mixing 10 mM HEPES, arginine, HEWL, and Sypro Orange to a microcentrifuge tube. Samples were mixed after the addition of HEWL and Sypro Orange via vortexing.

Table S3: Sample preparation for measuring hydrophobic exposure temperature of HEWL in ArgHCl solutions.

50x Sypro Orange (μL)	1.24 mM HEWL (μL)	1.64 M ArgHCl (μL)	10 mM HEPES (μL)
8	6	0	66
8	6	11	55
8	6	22	44
8	6	33	33
8	6	44	22
8	6	55	11
8	6	60	6
8	6	63	3
8	6	66	0

Thermal Shift Characterization

Differential scanning fluorimetry (DSF) can be used to determine the hydrophobic exposure temperature (T_{HE}) in a high-throughput manner.[S18,S19] The Sypro Orange dye shows enhanced fluorescence upon binding to hydrophobic regions of a protein, allowing for detection of protein unfolding events as a function of temperature. Previous reports have shown a strong correlation between T_{HE} and the actual thermodynamic melting temperature of a protein.[S18,S19] T_{HE} is typicallly defined as the temperature where the DSF melting curve reaches a minimum of the first derivative, marked as $-\frac{dF}{dT}$. We will use this technique to determine T_{HE} for HEWL in the presence and absence of added arginine.

Experimentally, three 25 μL replicate aliquots of each test sample were prepared as in

Table S3 and pipetted into a 96-well PCR plate (Thermo Scientific). The experiment was then run using a CFS Connect RT-PCR instrument (Bio-Rad). Fluorescence intensity was collected over the range of 10°C to 95°C in 1°C increments. The total time for the experiment was approximately 103 minutes.

References

(S1) Abascal, J. L. F.; Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123*, 234505.

(S2) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S. et al. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.

(S3) Lorentz, H. A. Ueber die Anwendung des Satzes vom virial in der Kinetischen Theorie der Gase. *Annalen der Physik* **1881**, *248*, 127–136.

(S4) Berthelot, D. Sur le mélange des gaz. *Compt. Rendus* **1898**, *126*.

(S5) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(S6) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(S7) Evans, D. J.; Holian, B. L. The Nose–Hoover thermostat. *J. Chem. Phys.* **1985**, *83*, 4069–4074.

(S8) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(S9) McInnes, L.; Healy, J.; Astels, S. hdbscan: Hierarchical density based clustering. *JOSS* **2017**, *2*, 205.

(S10) Record, M.; Anderson, C. Interpretation of preferential interaction coefficients of nonelectrolytes and of electrolyte ions in terms of a two-domain model. *Biophys. J.* **1995**, *68*, 786–794.

(S11) Shukla, D.; Trout, B. L. Interaction of Arginine with Proteins and the Mechanism by Which It Inhibits Aggregation. *J. Phys. Chem. B* **2010**, *114*, 13426–13438.

(S12) Schneider, C. P.; Trout, B. L. Investigation of CosoluteProtein Preferential Interaction Coefficients: New Insight into the Mechanism by Which Arginine Inhibits Aggregation. *J. Phys. Chem. B* **2009**, *113*, 2050–2058.

(S13) Schneider, C. P.; Shukla, D.; Trout, B. L. Arginine and the Hofmeister Series: The Role of Ion–Ion Interactions in Protein Aggregation Suppression. *J. Phys. Chem. B* **2011**, *115*, 7447–7458.

(S14) Vagenende, V.; Han, A. X.; Mueller, M.; Trout, B. L. Protein-Associated Cation Clusters in Aqueous Arginine Solutions and Their Effects on Protein Stability and Size. *ACS Chem. Biol.* **2013**, *8*, 416–422.

(S15) Heldt, C. L.; Hernandez, R.; Mudiganti, U.; Gurgel, P. V.; Brown, D. T.; Carbonell, R. G. A colorimetric assay for viral agents that produce cytopathic effects. *Journal of Virological Methods* **2006**, *135*, 56–65.

(S16) Joshi, P. U.; Decker, C.; Zeng, X.; Sathyavageeswaran, A.; Perry, S. L.; Heldt, C. L. Design Rules for the Sequestration of Viruses into Polypeptide Complex Coacervates. *Biomacromolecules* **2024**, *25*, 741–753.

(S17) Mi, X.; Blocher McTigue, W. C.; Joshi, P. U.; Bunker, M. K.; Heldt, C. L.; Perry, S. L. Thermostabilization of viruses *via* complex coacervation. *Biomater. Sci.* **2020**, *8*, 7082–7092.

(S18) Seabrook, S. A.; Newman, J. High-Throughput Thermal Scanning for Protein Stability: Making a Good Technique More Robust. *ACS Comb. Sci.* **2013**, *15*, 387–392.

(S19) Shi, S.; Semple, A.; Cheung, J.; Shameem, M. DSF Method Optimization and Its

Application in Predicting Protein Thermal Aggregation Kinetics. *Journal of Pharmaceutical Sciences* **2013**, *102*, 2471–2483.