Supplementary Information for Thermal compaction of disordered, hydrophobic elastin like polypeptides: a temperature-dependent, sequence-specific coarse-grained simulation model

Upayan Baul,*,† Michael Bley,*,† and Joachim Dzubiella*,†,‡

†*Applied Theoretical Physics – Computational Physics, Physikalisches Institut, Albert-Ludwigs-Universität Freiburg, Hermann-Herder Strasse 3, D-79104 Freiburg, Germany*

‡*Cluster of Excellence livMatS @ FIT - Freiburg Center for Interactive Materials and Bioinspired Technologies, Albert-Ludwigs-Universität Freiburg, Georges-Köhler-Allee 105, D-79110 Freiburg, Germany*

E-mail: upayan.baul@physik.uni-freiburg.de; michael.bley@physik.uni-freiburg.de; joachim.dzubiella@physik.uni-freiburg.de

Coarse grained beads in the SOP-IDP model

Table S1: Radii and charges of beads in the two bead per residue coarse-grained representation.[1,2] [a] Used as modifiable parameter in supplementary information section *Context-specific alterations within the scope of the model.*

bead type	description	vdW radius (Å)	charge (e)
C_α	backbone bead	1.90	0.0
Gly	single-bead Glycine	2.25	0.0
Val	Valine side-chain	2.93	0.0
Pro	Proline side-chain	2.78	0.0
Leu	Leucine side-chain	3.09	0.0
Ile	Isoleucine side-chain	3.09	0.0
Met	Methionine side-chain	3.09	0.0
Ala	Alanine side-chain	2.52[a]	0.0
Phe	Phenylalanine side-chain	3.18	0.0
Tyr	Tyrosine side-chain	3.23	0.0
Trp	Tryptophan side-chain	3.39	0.0
His	Histidine side-chain	3.04	0.0 / 1.0
Arg	Arginine side-chain	3.28	1.0
Lys	Lysine side-chain	3.18	1.0
Asp	Aspartic acid side-chain	2.79	-1.0
Glu	Glutamic acid side-chain	2.96	-1.0
Ser	Serine side-chain	2.59	0.0
Thr	Threonine side-chain	2.81	0.0
Asn	Asparagine side-chain	2.84	0.0
Gln	Glutamine side-chain	3.01	0.0
Cys	Cysteine side-chain	2.74	0.0

Numerical values of parameters in the energy function

Table S2: Numerical values of parameters in the energy function (equation 1 in main manuscript). [a] Related to ϵ_{ss} through $\epsilon_{SS} = \epsilon^0_{SS}\,|\epsilon_i - 0.7|$, where ϵ_i is obtained from the knowledge-based Betancourt-Thirumalai statistical potential.[4] [b] See Table S3 for numerical values as a function of temperature.

parameter	description	value
k	energy scale of FENE bond	20.0 kcal mol^{-1}Å^{-2}
R_0	tolerance in FENE bond fluctuations	2.0 Å
ϵ_{loc}	energy scale of local r^{-6} repulsion	1.0 kcal mol^{-1}
ϵ_{BB}	LJ interaction well depth	0.1186 kcal mol^{-1}
ϵ_{BS}	LJ interaction well depth	0.2372 kcal mol^{-1}
ϵ^0_{ss}	LJ interaction well depth [a]	0.1779 kcal mol^{-1}
ε	dielectric constant	temperature dependent [b], from reference[3]
κ	inverse Debye screening length	temperature and salt concentration dependent [b]

Table S3: Numerical values of dielectric constant (ε) and Debye screening length (κ^{-1}) at select simulation temperatures (T). Values for ε are obtained from published reference.[3] For monovalent salt solutions, $\kappa^{-1} = \left(\frac{\varepsilon\varepsilon_0 k_B T}{2\times 10^3 e^2 N_A C}\right)^{1/2}$, where ε_0 is the permittivity of free space, e is electronic charge, N_A is Avogadro number, and C is molar concentration of salt. Specific values for simulations at given temperatures were calculated using an online tool[5] for a simulated salt concentration of 0.15 M.

T / K	ε	κ^{-1} / Å
288	81.95	7.89
293	80.10	7.87
298	78.30	7.84
303	76.55	7.82
308	74.83	7.79
313	73.15	7.77
318	71.51	7.74
323	69.91	7.72
328	68.35	7.69
333	66.82	7.66
338	65.32	7.63
343	63.86	7.60
348	62.43	7.57
353	61.03	7.54

Derivation of functional form of $\epsilon(T)$ using square-well potential

Let us consider a square-well pair potential defined by

$$\begin{aligned} V(r) &= \infty \;\; ; r < \sigma \\ &\quad -\epsilon \;\; ; \sigma \leq r \leq (\sigma + l\sigma) \\ &\quad 0 \quad\; ; r > (\sigma + l\sigma). \end{aligned} \tag{S1}$$

The second virial coefficient is then defined as

$$\begin{aligned} B_2 &= -2\pi \int_0^\infty \left[e^{-\beta V(r)} - 1\right] r^2 dr \\ &= -2\pi \left[-\int_0^\sigma r^2 dr + \int_\sigma^{\sigma + l\sigma} \left(e^{\beta\epsilon} - 1\right) r^2 dr \right] \\ &= B_2^0 \sigma^3 - f(l)\sigma^3 \left(e^{\beta\epsilon} - 1\right), \end{aligned} \tag{S2}$$

whence

$$\frac{B_2}{\sigma^3} = B_2^0 - f(l)\left(e^{\beta\epsilon} - 1\right). \tag{S3}$$

Comparing with equation (2) in the main manuscript, and incorporating constant coefficients as well as $f(l)$ into prefactors c and d, we find the form

$$e^{\beta\epsilon} = 1 + cT + dT^2$$

or,

$$\epsilon = k_B T \ln\left[1 + cT + dT^2\right], \tag{S4}$$

which is the functional form for $\epsilon(T)$ expressed in equation (4) of the main manuscript.

Estimation of model parameters

As described in eq 4 of the main manuscript, and in eqs S1 - S4 above, our final form for the temperature-dependent well depth between pairwise hydrophobic beads i and j follows

$$\epsilon^{ij}(T) = k_B T \ln\left[1 + c(T - T_0^{ij}) + d(T - T_0^{ij})^2\right]. \tag{S5}$$

Even though we envisage that c and d should be universal, a brute-force search for optimal parameters c, d, and all T_0^{ij} is out of our computational reach. Instead, we adopt an efficient, albeit somewhat phenomenological methodology to estimate the values of the model parameters which is described below.

As discussed in the main manuscript, the dimensionless second virial coefficient $B_2(T)/\sigma^3$ of hydrophobic pair potentials $U(r, T)$ between two monomers is negative and a decreasing function of temperature.[6,7] For a thermosensitive polymer, it was observed that B_2 is simply linear in T over a sufficiently large temperature interval (ca. 30 K) covering the transition temperature (T_c).[8] It is thus expected that the simple linear form

$$B_2(T)/\sigma^3 = a_0 - a_1 T\,, \tag{S6}$$

and the temperature dependence of $\epsilon(T)$ derived based on it, should be able to reproduce LCST transitions in hydrophobic polymers. Under this simplified scenario eq S4 simplifies to $\epsilon = k_B T \ln[1 + c_1 T]$, and we can write eq S5 as

$$\epsilon^{ij}(T) \approx k_B T \ln\left[1 + c_1(T - T_0^{ij})\right]. \tag{S7}$$

Here, we have added a subscript to c (c_1) to distinguish its value from the parameter value corresponding to a second order expansion. While the simplified functional form is unable to reproduce the reduction of strength of hydrophobic attraction at $T \gg T_c$ through construction, we indeed observe that c_1 can be parameterized to reproduce the LCST transition

in $(\mathrm{VPGVG})_n$ ELPs.

Moghaddam and coworkers reported the potential of mean force (PMF) between a methane molecule and a methane dimer from explicit water simulations as a function of temperature.[9] We digitally extracted the global (contact) minima of the PMF profiles (see Figure 3 in reference[9]) and used the temperature-dependent values as estimates for $\epsilon(T)$. Following this, we fit the functional form of eq S7 to obtain an initial *order-of-magnitude* estimation for c_1. The fit, which resulted in $c_1 \approx 0.04$ K^{-1}, is shown in Figure S1.

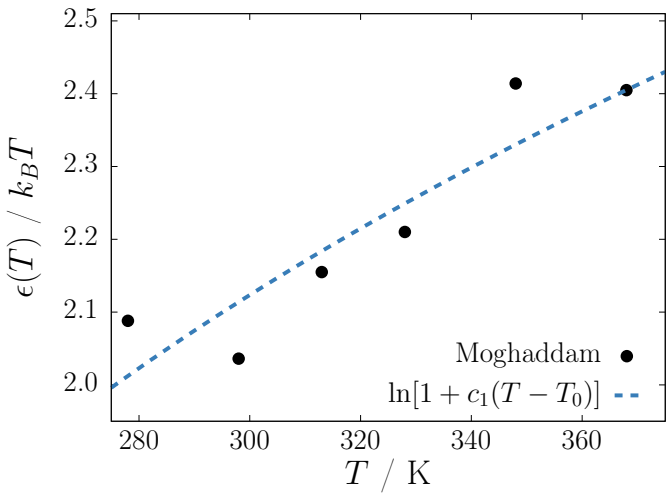

Figure S1: Order-of-magnitude estimation of c_1 in eq S7. The black dots represent global minima of the PMF profiles between a methane molecule and a methane dimer reported by Moghaddam and coworkers,[9] digitally extracted by us using online tool.[10] From the fitted line (blue), $c_1 \approx 0.04$ K^{-1} with $T_0 \approx 116$ K.

Following this initial estimate, we varied c_1 in the range $0.01\ \mathrm{K}^{-1} \leq c_1 \leq 0.04\ \mathrm{K}^{-1}$ and simulated the two ELP sequences $(\mathrm{VPGVG})_{60}$ and $(\mathrm{VPGVG})_{90}$ over a temperature range encompassing their respective experimental T_c values. To obtain T_0^{VV}, T_0^{VP} and T_0^{PP}, the reference temperatures for the temperature-dependent pairwise interactions, we used eq 5 in the main manuscript (with c replaced by c_1). We observed that for $c_1 = 0.025$ K^{-1}, the experimental T_c for both sequences can be reproduced to within (1 to 2) K. The results for selected c_1 values are shown in Figure S2.

It is important to note that even though the LCST can be captured by the functional form in eq S7, arising from a linear form of $B_2(T)/\sigma^3$ (eq S6), it suffers from the limitation that it

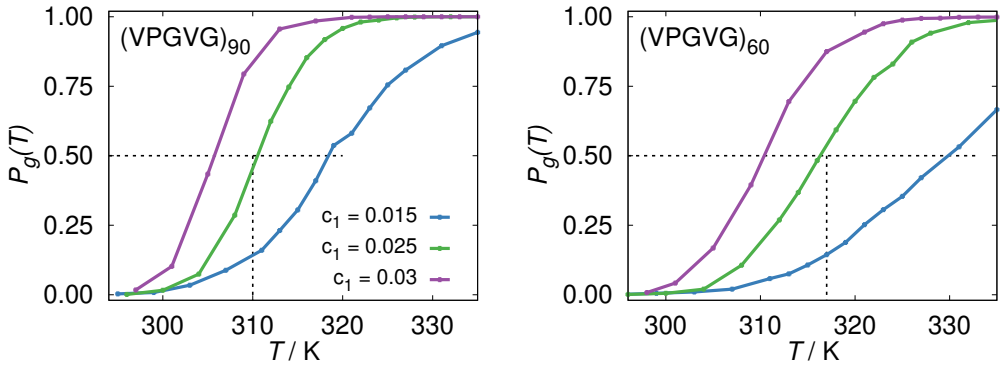

Figure S2: Fraction of collapsed globule-like conformations ($P_g(T)$, eq 8 in main manuscript) as a function of temperature for $(VPGVG)_{90}$ (left) and $(VPGVG)_{60}$ (right) for varying c_1 (eq S7, units of K^{-1}). The dotted horizontal lines show $P_g(T) = 0.5$ and the dotted vertical lines show the experimental T_c values (see main manuscript).[11] The results for for $c_1 = 0.025$ K^{-1} show that LCST transitions in $(VPGVG)_n$ ELPs can be reproduced by a functional form such as eq S7, which is a simplified version of the functional form of $\epsilon(T)$ used by us (eq S5, eq 4 in main manuscript).

can not capture the decreases in strength of hydrophobic interactions at higher temperatures. Accordingly, we next parameterized the full functional form of $\epsilon(T)$ (eq S5, eq 4 of main manuscript) which considers a second order (quadratic in T) form of $B_2(T)/\sigma^3$. For this, we adopted a phenomenological approach, instead of a parameter search through a large number of simulations.

Given the simultaneous reproduction of the experimental T_c for $(VPGVG)_{90}$ (310 K) and $(VPGVG)_{60}$ (317 K), we consider that $\epsilon(T)$ between hydrophobic residues in $(VPGVG)_n$, specifically Val-Val, Val-Pro and Pro-Pro, is reasonably approximated at temperatures $T =$ 310 K and $T = 317$ K by eq S7, with $c_1 = 0.025$ K^{-1}. These values in units of kcal/mol are $\epsilon^{VV} = 0.376$, $\epsilon^{VP} = 0.277$, and $\epsilon^{PP} = 0.276$ at $T = 310$ K, and $\epsilon^{VV} = 0.442$, $\epsilon^{VP} = 0.350$, and $\epsilon^{PP} = 0.349$ at $T = 310$ K respectively. Subsequently, we enumerate $\epsilon^{VV}(T)$, $\epsilon^{VP}(T)$ and $\epsilon^{PP}(T)$ using the full functional form given by eq S5 for different choices of c and d. To always have real values of $\epsilon^{ij}(T)$, we make an additional approximation, and drop the quadratic term in evaluating T_0^{ij} (eq 5 in main manuscript). In the process, we filter out combinations of c and d which satisfy the conditions stated below.

- $|c| \gg |d|$, $c > 0$ and $d < 0$ to account for increase in hydrophobicity with T, followed

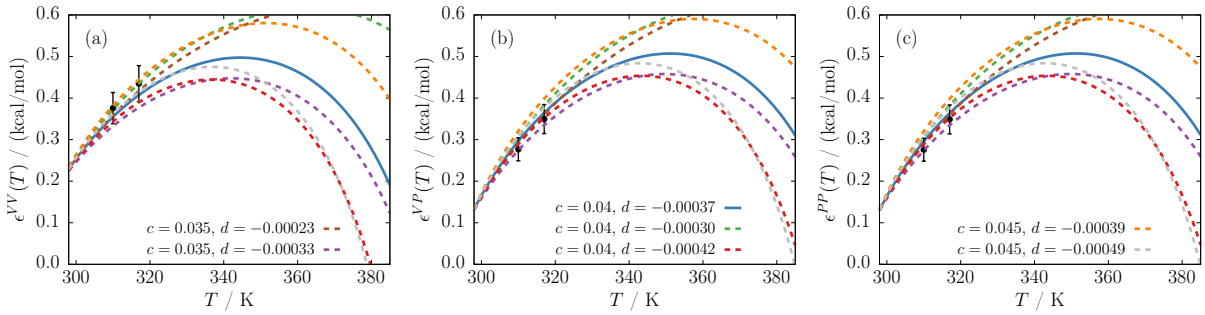

Figure S3: Rational estimation of possible ranges in c and d using conditions described in the text (eq S5). Units for c and d are K^{-1} and K^{-2} respectively. Black circles with erorbars show $\pm 10\%$ spread in ϵ^{VV}, ϵ^{VP}, and ϵ^{PP} values at $T = 310$ K and $T = 317$ K respectively, obtained using eq S7 with $c_1 = 0.025$ K^{-1}. The final parameter set is shown using solid blue lines. The dashed lines show several extreme combinations of c and d, at and beyond which at least one of the conditions listed in the text are not satisfied.

by decrease. In fact, as the quadratic term $d(T - T_0^{ij})^2$ compensates for the increase in $\epsilon(T)$ arising from $c(T - T_0^{ij})$, we only have to consider values of c that are greater than c_1 ($= 0.025$ K^{-1}). We varied c in the range 0.025 $K^{-1} \leq c \leq 0.045$ K^{-1} in intervals of 0.005 K^{-1}.

- At $T = 310$ K and $T = 317$ K, ϵ^{VV}, ϵ^{VP}, and ϵ^{PP} values are found to be within $\pm 10\%$ of the values quoted above.

- The maxima of ϵ^{VV}, ϵ^{VP}, and ϵ^{PP} are found to be within the wide temperature range of 320 K and 380 K. This condition is justified by observations from Marx *et al.*[12] who reported *unfolding regimes* in ELPs for $T \geq 333$ K, and is consistent with theory of hydrophobic solvation.[13]

As shown in Figure S3, these conditions help to greatly limit the possible combinations of c and d. The figure highlights, shown using dashed lines, the allowed ranges in d for three values of c. The chosen values of c lie across the final parameterized value of $c = 0.04$ K^{-1}. Additionally, it can be clearly seen that the quadratic term $d(T - T_0^{ij})^2$ with the condition $c \gg |d|$ has only marginal influence on $\epsilon(T)$ at $T = 298$ K. This justifies the approximate equation used to evaluate T_0^{ij} (eq 5 of main manuscript), where this quadratic term is neglected at $T = 298$ K. Following the identification of the limiting parameter ranges,

simulations were performed for $(VPGVG)_{90}$ and $(VPGVG)_{60}$, and the final parameter set $c =$ 0.04 K^{-1} and $d =$ -0.00037 K^{-2} was chosen via the simultaneous reproduction of experimental T_c values in a manner similar to described above in Figure S2. The good agreement in T_c within 1 to 2 K for the final parameter set has been highlighted in the main manuscript in Figure 4.

Pair specific T_0 values

Table S4: Temperature-dependent, hydrophobic amino-acid pair-specific LJ interaction well depths ($\epsilon(T)$, *eg.* Figure S4) are calculated using listed T_0 values and equation 4 in the main manuscript. The T_0 values are calculated using equation 5 in the main manuscript with $c = 0.04$ K^{-1}. All T_0 values are in Kelvin (K) units. While the parameter d does not explicitly contribute to the evaluation of T_0 values, it is important to note that at a given value for c, d can not be varied arbitrarily (see Figure S3) without affecting the performance of the model. The T_0 values, through this indirect influence, also are associated with $d = -0.00037$ K^{-2}. [a]Only for de-protonated ($q = 0$) state.

	Ala	Val	Leu	Ile	Pro	Phe	Met	Trp	Tyr	His[a]
Ala	290.1	288.3	288.4	288.6	292.7	288.8	289.9	288.1	290.6	294.0
Val	-	284.5	283.6	285.0	291.3	285.1	287.3	285.6	289.4	293.7
Leu	-	-	283.4	283.7	291.3	283.8	285.0	284.7	287.7	293.0
Ile	-	-	-	285.9	292.5	285.3	285.9	285.3	288.8	293.8
Pro	-	-	-	-	291.4	290.2	290.5	284.4	288.1	291.6
Phe	-	-	-	-	-	283.3	282.5	283.8	287.1	290.2
Met	-	-	-	-	-	-	286.3	281.9	286.9	290.4
Trp	-	-	-	-	-	-	-	284.3	286.4	287.4
Tyr	-	-	-	-	-	-	-	-	289.4	290.0
His[a]	-	-	-	-	-	-	-	-	-	288.8

Hydrodynamic radius calculation using the Nygaard relation

Nygaard and co-workers recently proposed a relationship among the hydrodynamic radius R_h^{Nyg}, radius of gyration R_g, and sequence length N of disordered polypeptide sequences.[14]

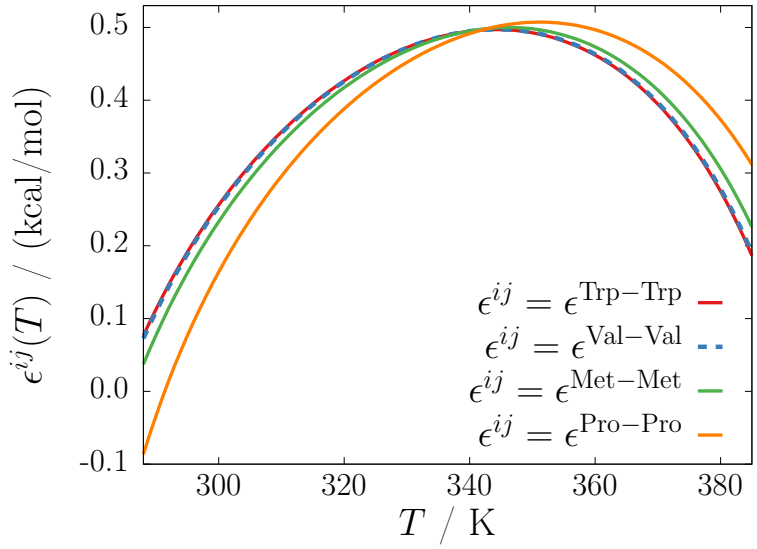

Figure S4: Plot showing a few selected pair-specific LJ interaction well depths $\epsilon(T)$ as a function of temperature. $\epsilon_{ij}(T)$ between residues i and j are calculated using equation 4 in the main manuscript with $c = 0.04$ K^{-1}, $d = -0.00037$ K^{-2}, and T_0^{ij} from Table S4. Note that $\epsilon(T)$ is not symmetric about the peak. The effective LJ interactions between beads depend also on their vdW radii given in Table S1.

The Nygaard relation is given by

$$\frac{R_g}{R_h^{\mathrm{Nyg}}}(N, R_g) = \frac{\alpha_1 \left(R_g - \alpha_2 N^{0.33}\right)}{N^{0.60} - N^{0.33}} + \alpha_3. \tag{S8}$$

Here we use the super-script notation R_h^{Nyg} to distinguish from R_h calculated using the Kirkwood double-sum formula described in eq 6 of the main manuscript. The optimized parameter values were reported to be $\alpha_1 = 0.216$ $\mathring{A}^{-1}$, $\alpha_2 = 4.06$ $\mathring{A}$ and $\alpha_3 = 0.821$ for IDPs.[14] We used eq S8 to calculate R_h^{Nyg} from the simulated R_g values (eq 7 in main manuscript) and the results are reported below for $(\mathrm{VPGVG})_n$ ELP sequences. In Table S6, we also report the temperature-dependent estimates for R_h^{Nyg} for IDP p53-IDR.

As can be seen from Table S5, under good-solvent conditions of $T = 293$ K, neither R_h, nor R_h^{Nyg} consistently outperforms the other in reproducing the experimental hydrodynamic radii R_h^{exp}. For shorter chain lengths R_h^{Nyg} is closer to R_h^{exp}, but for longer sequences the Kirkwood relation appears to be more accurate. In absence of experimental values at $T = 340$ K, the relative performance can not be assessed. At both temperatures, R_h and

Table S5: Comparison of hydrodynamic radii calculated using the Kirkwood double-sum formula (R_h), reported in Table 1 of the main manuscript, and the Nygaard relation (R_h^{Nyg}) for $(\mathrm{VPGVG})_n$ ELPs. Experimentally reported hydrodynamic radii (R_h^{exp}) at 293 K [15] are also included for reference.

ELP	R_h^{exp} (293 K)	R_h (293 K)	R_h^{Nyg} (293 K)	R_h (340 K)	R_h^{Nyg} (340 K)
$(\mathrm{VPGVG})_{150}$	–	7.02	8.89	3.46	4.15
$(\mathrm{VPGVG})_{120}$	6.0	6.12	7.68	3.20	3.91
$(\mathrm{VPGVG})_{90}$	–	5.22	6.47	3.00	3.62
$(\mathrm{VPGVG})_{60}$	4.2	4.24	5.14	2.69	3.28
$(\mathrm{VPGVG})_{40}$	3.7	3.49	4.07	2.43	2.98
$(\mathrm{VPGVG})_{30}$	3.4	3.10	3.53	2.28	2.79
$(\mathrm{VPGVG})_{20}$	2.7	2.56	2.86	2.04	2.48

R_h^{Nyg} show consistent qualitative trends with chain length, as envisaged by Nygaard and co-workers.[14] It should also be noted that the performance of R_h^{Nyg} depends also on the accuracy of R_g calculated from simulations.

Temperature-dependent size estimates for IDP p53-IDR

Table S6: Experimental hydrodynamic radius (R_h^{exp}) for IDP p53-IDR as reported in DLS experiments by Whitten *et al.*,[16] together with size estimates from our simulations at the corresponding simulation temperatures. Here, R_g stands for radius of gyration from simulations calculated using eq 7 in the main manuscript. Expressions R_h and R_h^{Nyg} stand for hydrodynamic radii calculated using the Kirkwood double-sum formula (eq 6 in main manuscript) and the Nygaard relation (eq S8). The R_g values quoted in the table were used to calculate R_h^{Nyg}. At a first glance the R_h^{Nyg} appear to be closer to R_h^{exp}. However, between the temperatures of 288 K and 338 K, R_h^{exp} decreases by $\sim 17.5\%$. The corresponding compaction is $\sim 14.6\%$ for R_h but only $\sim 8.5\%$ for R_h^{Nyg}. As such, we concluded that change in R_h provides a closer approximation for the compaction of p53-IDR with temperature. For visual comparisons and simulated values at additional temperatures please refer to Figure 9(c) in the main manuscript. Detailed commentary on the agreements and deviations between experiment and simulations can also be found therein. Experimental data were extracted digitally using online tool *WebPlotDigitizer*.[10]

T / K	R_h^{exp} / nm	R_h / nm	R_g / nm	R_h^{Nyg} / nm
288	3.42 (0.33)	2.47 (0.20)	3.01 (0.52)	2.83
298	3.28 (0.25)	2.37 (0.20)	2.88 (0.52)	2.78
308	3.28 (0.23)	2.29 (0.21)	2.73 (0.51)	2.71
318	3.10 (0.22)	2.22 (0.20)	2.59 (0.51)	2.65
328	2.95 (0.42)	2.15 (0.21)	2.50 (0.50)	2.60
338	2.82 (0.39)	2.11 (0.21)	2.47 (0.51)	2.59
348	2.65 (0.31)	2.14 (0.21)	2.50 (0.49)	2.60

Distribution of R_g is at high temperatures

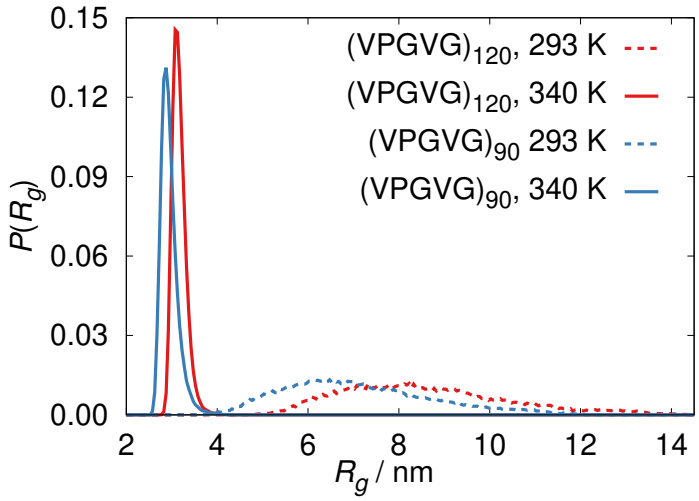

Figure S5: Probability distributions of radius of gyration (R_g) at coil-like (293 K, dashed lines) and globule-like (340 K, solid lines) phases for $(VPGVG)_{120}$ (red) and $(VPGVG)_{90}$ (blue). As clearly seen, broad distributions at 293 K shift towards sharply peaked distributions at 340 K. However, as discussed in the main manuscript using Figure 7, the polymer conformations maintain disorder at 340 K. 14.6 8.5

Disordered ensemble and uncorrelated conformations

In Figure 7 of the main manuscript, the globule-like phase of ELP $(VPGVG)_{90}$ was shown to have similar scaled end-to-end distance distributions as theoretical polymer models. This indicates that the underlying ensembles of ELPs are disordered and dynamic, even at high temperatures above T_c. In Figure S6 we provide supporting evidence for the assertion by representing individual simulated conformations of $(VPGVG)_{90}$ in the scaled R_g - R_{ee} plane.

In Figure S7 we plot the time auto-correlation function of instantaneous the end-to-end vector $\overrightarrow{R_{ee}}$ for $(VPGVG)_{90}$ in the globule-like phase at 340 K. The auto-correlation function is given by

$$C(t) = \left\langle \frac{\overrightarrow{R_{ee}}(t) \cdot \overrightarrow{R_{ee}}(0)}{|\overrightarrow{R_{ee}}(t)||\overrightarrow{R_{ee}}(0)|} \right\rangle \tag{S9}$$

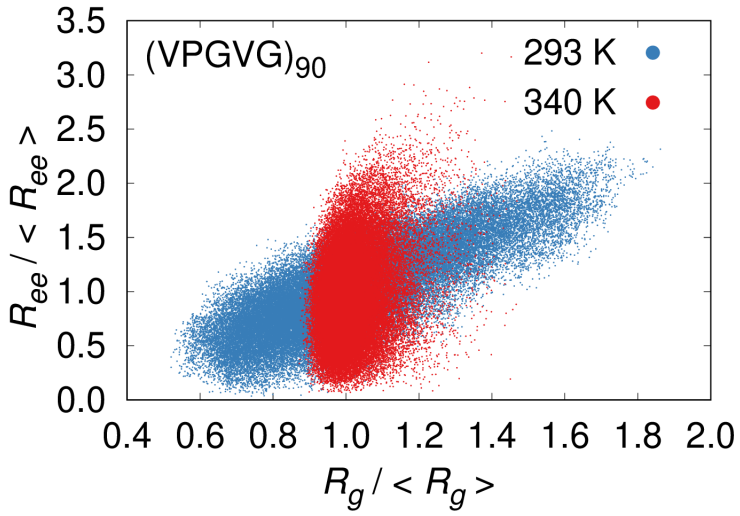

Figure S6: Scatter plot representation of individual simulated conformations of $(VPGVG)_{90}$ in the scaled R_g - R_{ee} plane. Each point in the plot represents a simulated conformation. The scaling of both R_g and R_{ee} are with respect to mean values obtained from simulations, and are shown using angular brackets in the figure axis labels. Even in the globule-like phase (340 K, red dots) where fluctuation of R_g is constrained to $\sim \pm 15\%$ of the mean value, one can observe a large ($\sim \pm 100\%$) spread in R_{ee}. This highlights the disordered and dynamic nature of the globule-like ensemble at 340 K, and is consistent with Figure 7 in the main manuscript. The disordered and dynamic nature for the low temperature coil-like phase (293 K, blue dots) is trivially ascertained following same rationale.

where $|\overrightarrow{R_{ee}}|$ represents the magnitude of $\overrightarrow{R_{ee}}$, and angular brackets represent ensemble average. The dashed line in black in Figure S7 shows an exponential fit of $C(t)$, given by $C(t) = C_0 \exp(-t/\tau)$. Owing to the inherent noise in the data the fit is only approximate, with estimates of $C_0 \sim 0.5$ and $\tau \sim 20$ ns. Correlation time of 20 ns is small compared to the duration of the production simulations (3000 ns, see main manuscript).

FASTA sequences for IDPs

Simulated sequences for the simulated IDPs hTau40 and p53-IDR are provided below. The sequences are identical to the ones used in previous published work.[1]

FASTA sequence for hTau40 :

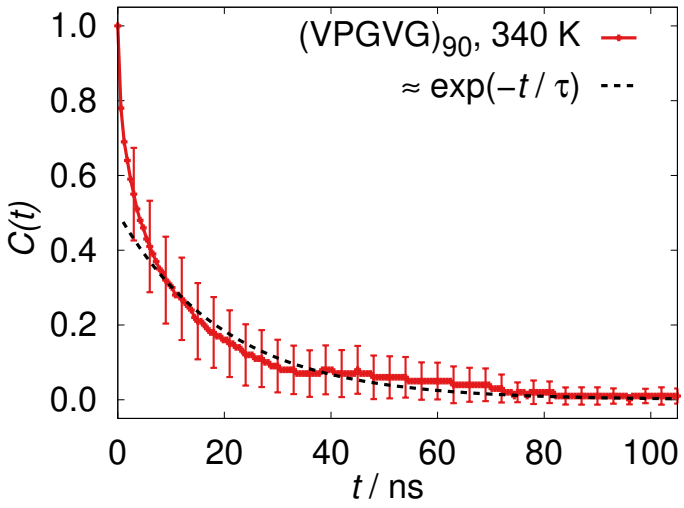

Figure S7: Decay of time auto-correlation (equation S9) of the unit end-to-end vector in ELP $(VPGVG)_{90}$ at 340 K. The dashed line in black shows an approximate fit using $C(t) = C_0 \exp(-t/\tau)$ where $C_0 \sim 0.5$ and $\tau \sim 20$ ns. The duration of production simulations (3000 ns) is much larger than τ, which establishes that the generated ensemble does not sample only kinetically trapped conformations.

MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKESPLQTPTEDGSEEPGSETSD

AKSTPTAEDVTAPLVDEGAPGKQAAAQPHTEIPEGTTAEEAGIGDTPSLEDEAAGHVTQARMVSK

SKDGTGSDDKKAKGADGKTKIATPRGAAPPGQKGQANATRIPAKTPPAPKTPPSSGEPPKSGDRS

GYSSPGSPGTPGSRSRTPSLPTPPTREPKKVAVVRTPPKSPSSAKSRLQTAPVPMPDLKNVKSKI

GSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSL

GNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGA

EIVYKSPVVSGDTSPRHLSNVSSTGSIDMVDSPQLATLADEVSASLAKQGL

FASTA sequence for p53-IDR :

MEEPQSDPSVEPPLSQETFSDLWKLLPENNVLSPLPSQAMDDLMLSPDDIEQWFTEDPGPDEAPR

MPEAAPPVAPAPAAPTPAAPAPAPSWPL

Context-specific alterations within the scope of the model

In spite of the limitations of our simulation model to capture some of the sequence (guest residue) specific trends in T_c of ELPs, the model performs appreciably well for systems with manifold hydrophobic interactions, namely IDPs p53-IDR and hTau40. In retrospect, this is unsurprising. Indeed, the ranking of amino acid hydrophobicities, or the *hydrophobicity scale*, is non-unique and depends on the context-dependent definition of hydrophobicity.[17–22] As such, it is not unexpected that the underlying Betancourt-Thirumalai statistical potential,[4] which has no inputs from T_c of ELPs but rather from observed spatial proximities of amino acids in folded protein structures, will result in deviations of temperature-dependent relative hydrophobicities of amino acids. For example, if relative contact propensities are treated as a measure of hydrophobicity, temperature-dependent relative hydrophobicities of amino acids reported recently by Abeln *et.al*[23] do not conform fully to the observations with ELPs.[24]

These observations simply highlight the complexities associated with identifying a 'universal' scheme for weighting the pairwise interactions in a CG computational model. For specific applications, however, modifications to the observed hydrophobicities in our model may be desirable. To conclude our results, in the following we show that the simulation model is amenable to context-specific modifications, without violating the axioms set forth in the methods section. We note in passing, that the *relative* hydrophobicities of amino acids (assigned through equation 5 in main manuscript) has little, if any connection with our envisaged temperature dependence of strength of hydrophobic interactions (equation 4 therein).

Modifications through well depth ($\epsilon(T)$) : In absence of charged residues, the only explicit temperature-dependent parameters in our simulation model are the pair-specific $\epsilon(T)$ between hydrophobic residues. As discussed in the methods section of the main manuscript, $\epsilon(T)$ between a pair of residues i and j is determined by a single parameter T_0^{ij},- which we assigned using the SOP-IDP model (equation 5 in main manuscript). As such T_0^{ij} is an obvious choice for affecting minor modifications to relative hydrophobicities. For example,

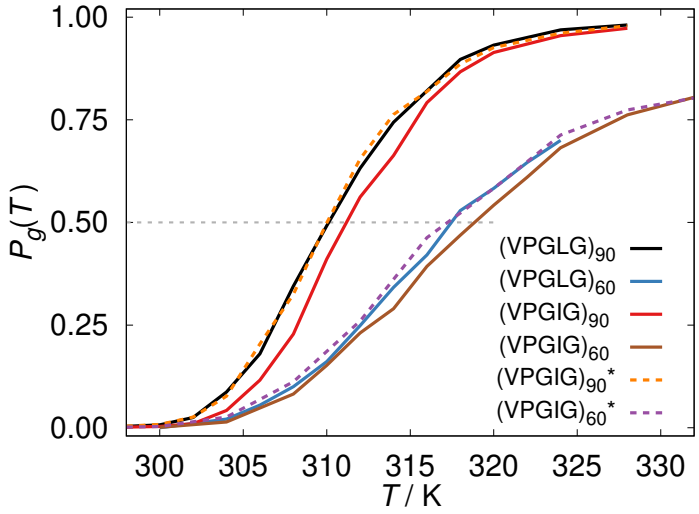

Figure S8: Fraction of collapsed globule-like conformations ($P_g(T)$, see equation 8 in main manuscript) as a function of T. Solid lines are obtained with parameter values consistently used in the manuscript. Results with modified parameters (see text) are shown with dashed lines and are labeled with asterisks in legends. Solid lines : (VPGLG)$_{90}$, (VPGLG)$_{60}$, (VPGIG)$_{90}$ and (VPGIG)$_{60}$ are shown in black, blue, red, and brown respectively. Curves for (VPGIG)$_{90}$ and (VPGIG)$_{60}$ with LJ parameters modified through T_0^{ij} are shown with dashed lines,– orange and purple respectively.

in experiments the T_c-s of ELPs (VPGIG)$_n$ and (VPGLG)$_n$ have been reported to be nearly identical.[24] In computational models that treat L and I residues as indistinguishable, *eg.* the HPS model,[25] this would be trivially reproduced. In the SOP-IDP model, residues L and I are distinguishable through interaction parameters (see Table S4), and we observed the T_c for (VPGIG)$_n$ to be higher than the respective T_c for (VPGLG)$_n$ in our simulations (Figure S8). However, making temperature-dependent interactions of I identical to those of L (*i.e.*, $T_0^{II*} = T_0^{LL}$, $T_0^{IV*} = T_0^{LV}$, and $T_0^{IP*} = T_0^{LP}$), the LCST transitions for (VPGIG)$_n$ can be made virtually indistinguishable from (VPGLG)$_n$, as shown with dashed curves in Figure S8.

As a note of caution, this approach is not without pitfalls, and changing T_0^{ij} arbitrarily strongly impacts the balance of inter-residue interactions set forth by the SOP-IDP model[1] at ambient temperatures ($T \sim 298$ K) through the underlying Betancourt-Thirumalai potential map[4] (minor differences are already affected through dropping the quadratic term

of equation 4 in equation 5, both in main manuscript). With compostionally diverse sequences such as generic IDPs, the influences of such reparameterizations are likely to be unpredictable.

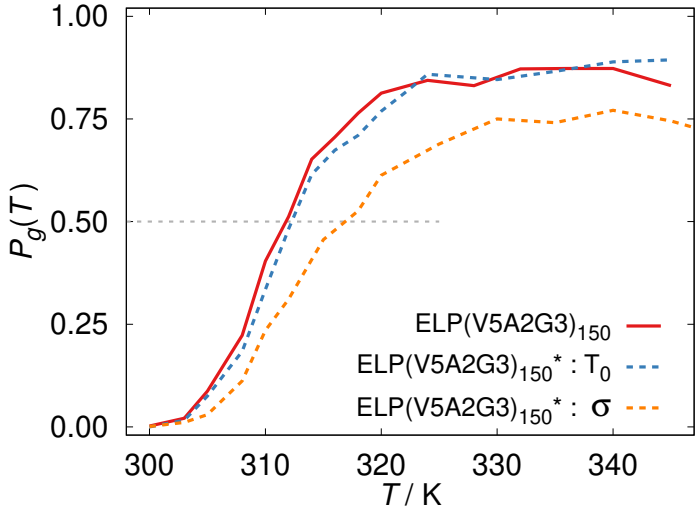

Figure S9: Fraction of collapsed globule-like conformations ($P_g(T)$, see equation 8 in main manuscript) as a function of T. Solid red line is for ELP(V5A2G3)$_{150}$ with parameter values consistently used in the manuscript. Results with modified parameters are shown with dashed lines and are labeled with asterisks in legends. In blue : LJ parameters modified through T_0^{ij}, and in orange : LJ parameters modified through σ. See text for description of ELP(V5A2G3)$_n$ notation.

Modifications through vdW radius: A second approach is to modify the bead dimensions through pairwise σ_i in equation 1 of main manuscript. This does not directly violate the Betancourt-Thirumalai potential map, but still influences the LJ interactions and hence $B_2(T)$ (equation 2 in main manuscript). Also, noting that the σ_i for a bead pair is defined as the sum of their vdW radii, this leads to only one free parameter per residue,– namely the vdW radius of the corresponding side chain bead. As an example of such modifications, we consider the LCST transitions in a *weakly hydrophobic* ELP sequence with non-homogeneous distribution and composition of pentameric repeats, denoted by ELP(V5A2G3)$_{150}$ in the following and in Figure S9. The sequence has a total of 150 pentameric units, among which (VPGVG), (VPGAG) and (VPGGG) are found in ratio 5:2:3. The distributions of the

pentamers along the sequence was chosen at random.

The experimental T_c for ELP(V5A2G3)$_{150}$, $\sim$ 323 K at lowest reported concentration of 1 μM,[11] deviates strongly from our estimate of $\sim$ 312 K under the parameterization dictated by equation 5 of main manuscript (red curve, Figure S9. Furthermore, relatively major modifications to the T_0^{ij} values involving alanine also does not lead to any marked improvement in the estimate for T_c. This is shown using the blue dashed curve in Figure S9, which uses T_0^{AV*} = 293 K, T_0^{AP*} = 295 K, T_0^{AA*} = 298 K (modified from T_0^{AV} = 288.3 K, T_0^{AP} = 292.7 K, T_0^{AA} = 290.1 K),– a substantial modification when compared with default T_0^{ij} values reported in Table S4. A more significant improvement ($T_c \sim$ 317 K) is observed when instead of T_0^{ij} values, the vdW radius of alanine side chain bead is changed from 2.52 Å(Table S1) to 1.90 Å. This is shown using the orange dashed line in Figure S9.

References

(1) Baul, U.; Chakraborty, D.; Mugnai, M. L.; Straub, J. E.; Thirumalai, D. Sequence effects on size, shape, and structural heterogeneity in intrinsically disordered proteins. *J. Phys. Chem. B* **2019**, *123*, 3462–3474.

(2) Chakraborty, D.; Straub, J. E.; Thirumalai, D. Differences in the free energies between the excited states of Aβ40 and Aβ42 monomers encode their distinct aggregation propensities. *bioRxiv* **2020**, `https://doi.org/10.1101/2020.02.09.940676`.

(3) Malmberg, C. G.; Maryott, A. A. Dielectric constant of water from 0°C to 100°C. *J. Res. Natl. Bur. Stand.* **1956**, *56*, 2641.

(4) Betancourt, M. R.; Thirumalai, D. Pair potentials for protein folding: choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. *Protein Sci.* **1999**, *8*, 361–369.

(5) Kocherbitov, V. Calculation of debye length. 2019 (accessed 01 Apr 2019); `http://www.surfchem.info/calculate/Debye/`.

(6) Lüdemann, S.; Abseher, R.; Schreiber, H.; Steinhauser, O. The temperature-dependence of hydrophobic association in water. Pair versus bulk hydrophobic interactions. *J. Am. Chem. Soc.* **1997**, *119*, 4206–4213.

(7) Chaudhari, M. I.; Holleran, S. A.; Ashbaugh, H. S.; Pratt, L. R. Molecular-scale hydrophobic interactions between hard-sphere reference solutes are attractive and endothermic. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 20557–20562.

(8) Reinhardt, M.; Dzubiella, J.; Trapp, M.; Gutfreund, P.; Kreuzer, M.; Gröschel, A. H.; Müller, A. H. E.; Ballauff, M.; Steitz, R. Fine-tuning the structure of stimuli-responsive polymer films by hydrostatic pressure and temperature. *Macromolecules* **2013**, *46*, 6541–6547.

(9) Moghaddam, M. S.; Shimizu, S.; Chan, H. S. Temperature dependence of three-body hydrophobic interactions : potential of mean force, enthalpy, entropy, heat capacity, and nonadditivity. *J. Am. Chem. Soc.* **2005**, *127*, 303–316.

(10) Rohtagi, A. WebPlotDigitizer Version 4.2. 2019 (accessed 01 May 2019); `https://automeris.io/WebPlotDigitizer`.

(11) Meyer, D. E.; Chilkoti, A. Quantification of the effects of chain length and concentration on the thermal behavior of elastin-like polypeptides. *Biomacromolecules* **2004**, *5*, 846–851.

(12) Rousseau, R.; Schreiner, E.; Kohlmeyer, A.; Marx, D. Temperature-dependent conformational transitions and hydrogen-bond dynamics of the elastin-like octapeptide GVG(VPGVG): a molecular-dynamics study. *Biophys. J.* **2004**, *86*, 1393 – 1407.

(13) Huang, D. M.; Chandler, D. Temperature and length scale dependence of hydrophobic effects and their possible implications for protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 8324–8327.

(14) Nygaard, M.; Kragelund, B. B.; Papaleo, E.; Lindorff-Larsen, K. An efficient method for estimating the hydrodynamic radius of disordered protein conformations. *Biophys. J.* **2017**, *113*, 550 – 557.

(15) Fluegel, S.; Fischer, K.; McDaniel, J. R.; Chilkoti, A.; Schmidt, M. Chain stiffness of elastin-like polypeptides. *Biomacromolecules* **2010**, *11*, 3216–3218.

(16) Langridge, T. D.; Tarver, M. J.; Whitten, S. T. Temperature effects on the hydrodynamic radius of the intrinsically disordered N-terminal region of the p53 protein. *Proteins: Struct., Funct., Bioinf.* **2014**, *82*, 668–678.

(17) Nozaki, Y.; Tanford, C. The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions establishment of a hydrophobicity scale. *J. Biol. Chem.* **1970**, *246*, 2211–2217.

(18) Kyte, J.; Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105 – 132.

(19) Eisenberg, D. Three-dimensional structure of membrane and surface proteins. *Annu. Rev. Biochem.* **1984**, *53*, 595–623.

(20) Moret, M. A.; Zebende, G. F. Amino acid hydrophobicity and accessible surface area. *Phys. Rev. E* **2007**, *75*, 011920.

(21) Bandyopadhyay, D.; Mehler, E. L. Quantitative expression of protein heterogeneity: response of amino acid side chains to their local environment. *Proteins: Struct., Funct., Bioinf.* **2008**, *72*, 646–659.

(22) Zhu, C.; Gao, Y.; Li, H.; Meng, S.; Li, L.; Francisco, J. S.; Zeng, X. C. Characterizing hydrophobicity of amino acid side chains in a protein environment via measuring contact angle of a water nanodroplet on planar peptide network. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 12946–12951.

(23) van Dijk, E.; Hoogeveen, A.; Abeln, S. The hydrophobic temperature dependence of amino acids directly calculated from protein structures. *PLoS Comput. Biol.* **2015**, *11*, 1–17.

(24) Urry, D. W.; Gowda, D. C.; Parker, T. M.; Luan, C.-H.; Reid, M. C.; Harris, C. M.; Pattanaik, A.; Harris, R. D. Hydrophobicity scale for proteins based on inverse temperature transitions. *Biopolymers* **1992**, *32*, 1243–1250.

(25) Dignon, G. L.; Zheng, W.; Kim, Y. C.; Best, R. B.; Mittal, J. Sequence determinants of protein phase behavior from a coarse-grained model. *PLoS Comput. Biol.* **2018**, *14*, e1005941.