# The Scaled Effective Solvent Method for Predicting the Equilibrium Ensemble of Structures with Analysis of Thermodynamic Properties of Amorphous Polyethylene Glycol-Water Mixtures

*Hyeyoung Shin[a], Tod A. Pascal[ab], William A. Goddard III[abc*], and Hyungjun Kim[a*]*

[a]Graduate School of EEWS, Korea Advanced Institute of Technology (KAIST), Daejeon, Korea

[b]Material and Process Simulation Center, California Institute of Technology, Pasadena, CA 91125

[c]World Class University Professor, KAIST, Daejeon, Korea

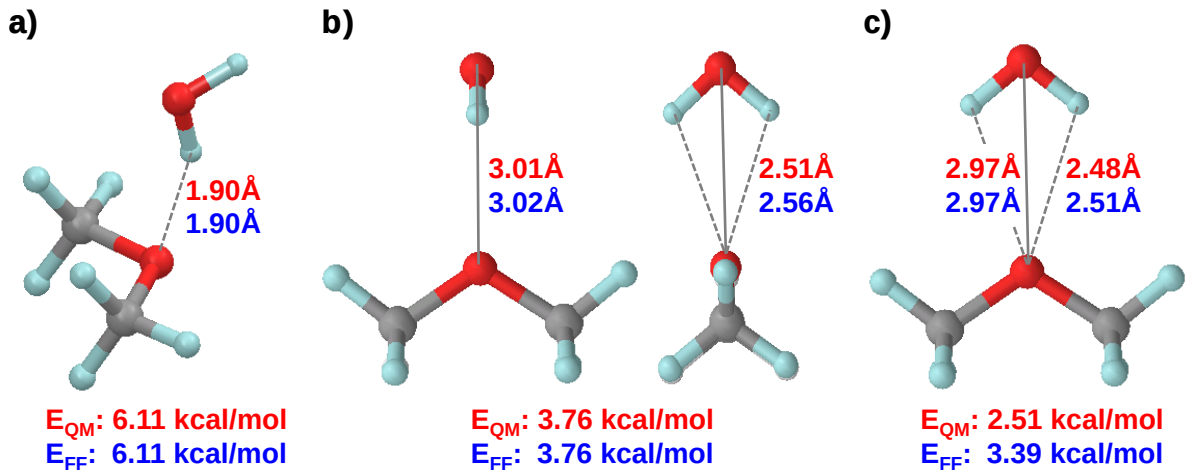



**Figure S1.** To describe the interaction between Polyethylene glycol (PEG) and water, we carried out quantum mechanical (QM) calculations using the M06-2X flavor of DFT with the cc-PVTZ++ basis set. We considered three types of optimized dimer structures between dimethyl ether (DME) and a water molecule from a) global minimum configuration, b) and c) structures optimized while constraining the point group symmetry of $C_{2v}$. QM calculation results and force field (FF) calculation results are described as red and blue fonts, respectively. Carbon atoms are gray-colored, oxygen atoms are red-colored, and hydrogen atoms are light blue-colored.

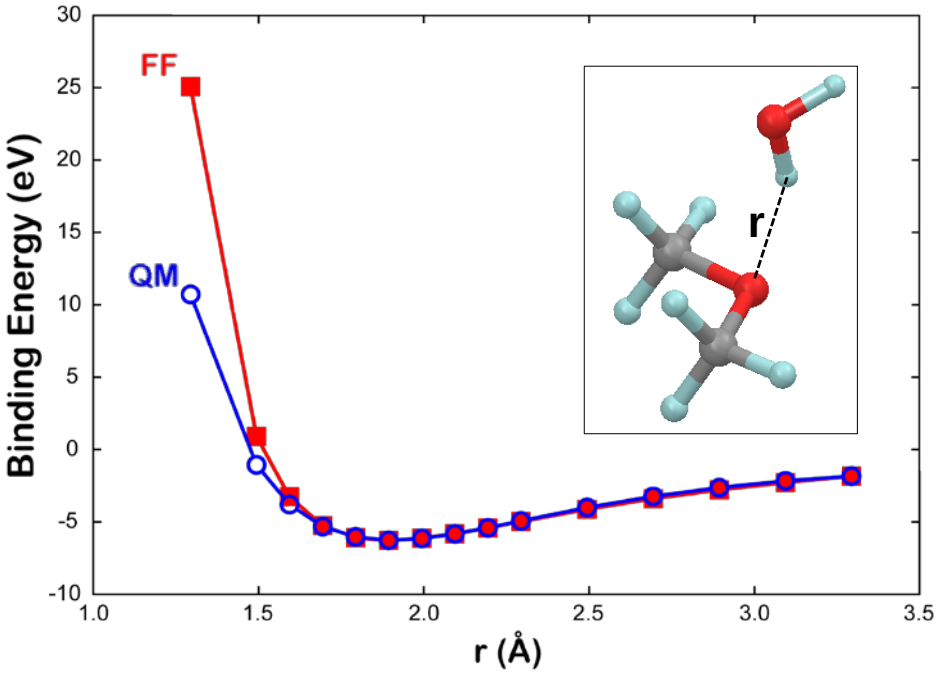



**Figure S2.** Dispersion curve obtained using the global minimum configuration (inset, Fig. S1a). Red solid line is calculated with the force field (FF) while the blue solid line is obtained from quantum mechanics (QM) calculation using M06-2X flavor of density functional theory (DFT) with the cc-PVTZ++ basis set. Carbon atoms are gray-colored, oxygen atoms are red-colored, and hydrogen atoms are light blue-colored.

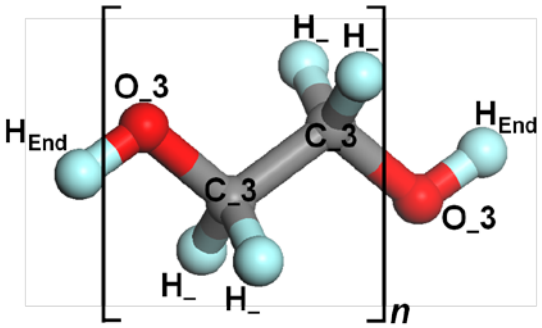
**Table S1.** In our simulations all force field parameters are DREIDING except that a) Van der Waal's (vdW) interactions between ether oxygen in a PEG backbone ($O_{PEG}$) and hydrogen in water ($H_{wat}$) are described with Morse potential [eq. (1)] and b) the explicit hydrogen bonding interactions are described by DREIDING hydrogen bond (H-bond) parameters above with the standard $(\cos\theta_{OHO})^4$ angular dependence and a cutoff at $\theta_{OHO}$ =90° [12-10 Lennard-Jones (LJ) potential, eq. (2)]. We optimized off-diagonal vdW parameters first to reproduce the geometry and binding energy of dimethyl ether (DME) and water interaction imposing $C_{2v}$ symmetry (Fig. S1b and S1c) where the explicit H-bond term has no contribution due to the angle cutoff. Then, we optimized the explicit H-bond term to reproduce the geometry and binding energy of the fully relaxed structure of DME and water (Fig. S1a).

| $O_{PEG}$-$H_{wat}$ | Van der Waals (Morse) | | | Hydrogen bond (12-10 LJ) | |
|---|---|---|---|---|---|
| | $D_0$ (kcal/mol) | $R_0$ (Å) | $\alpha$ | $D_0$ (kcal/mol) | $R_0$ (Å) |
| | 0.3100 | 2.867 | 6.300 | 1.967 | 2.850 |

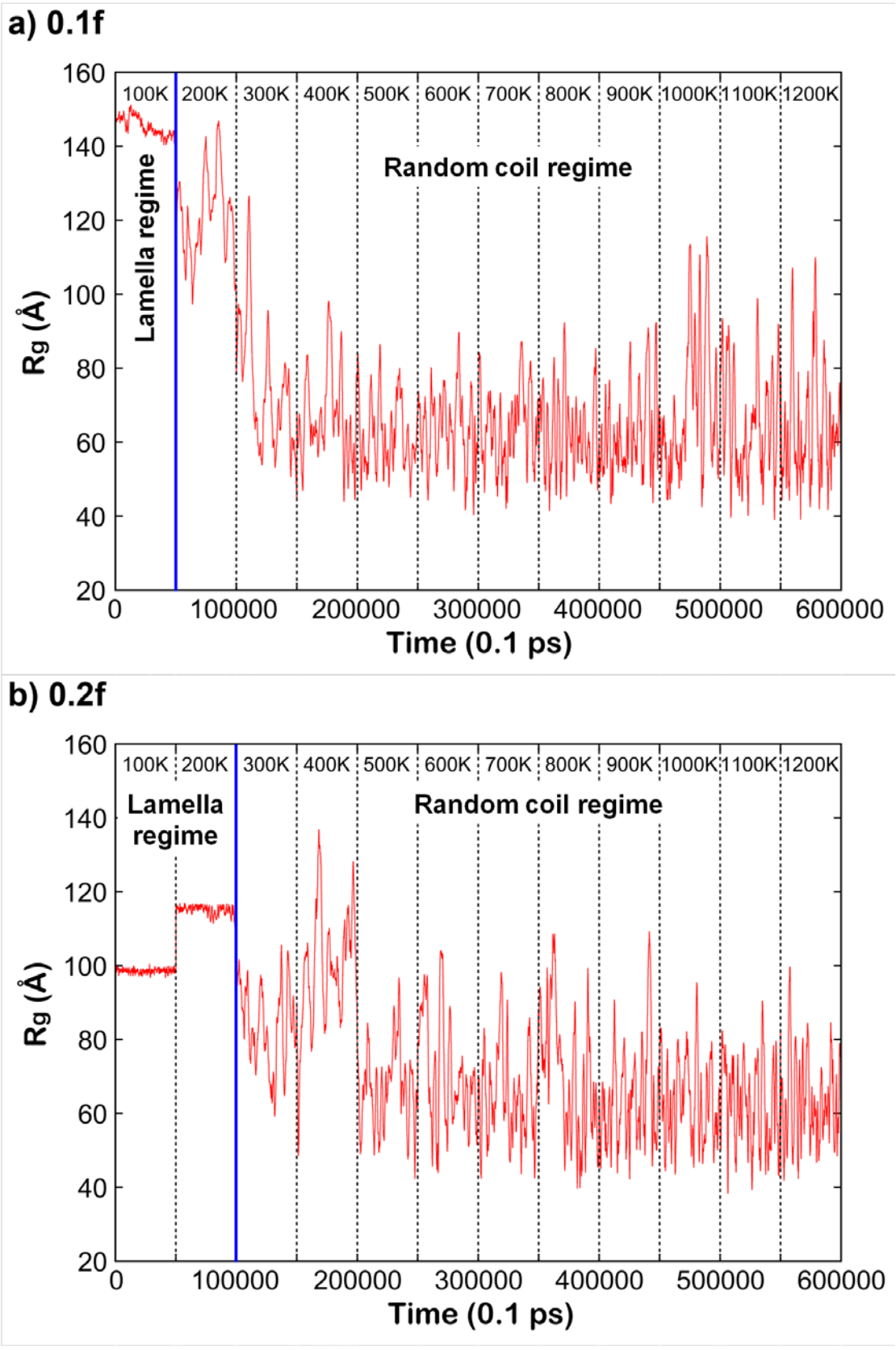
Morse potential: $U(R) = \mathrm{D}_0\left\{\left(\exp\left(-\left(\frac{\alpha}{2}\right)\left(\frac{R}{R0}-1\right)\right)\right)^2 - 2\exp\left(-\left(\frac{\alpha}{2}\right)\left(\frac{R}{R0}-1\right)\right)\right\}$ (1)

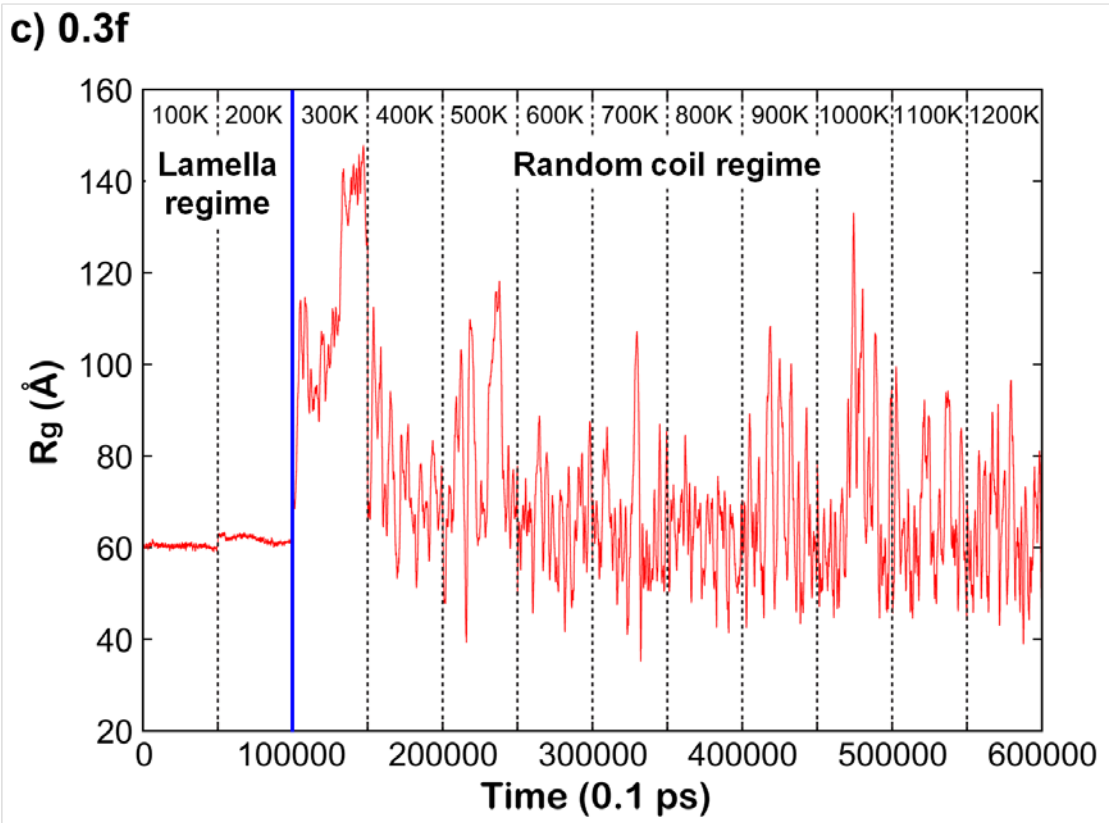
12-10 LJ potential: $U(R) = D_0\left\{5\left(\frac{R_0}{R}\right)^{12} - 6\left(\frac{R_0}{R}\right)^{10}\right\}\cos^4\theta$ (2)

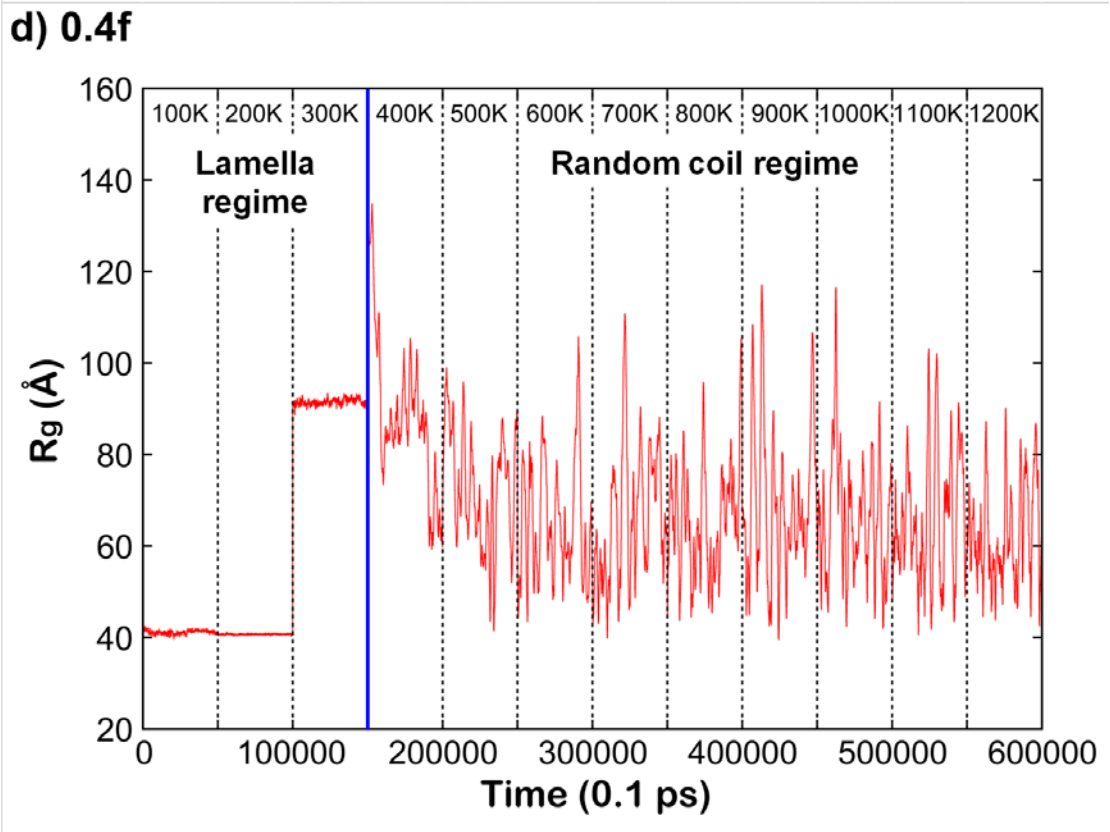
**Table S2.** The partial charge distributions (*q*) of a 1,2-Ethanediol which represents the PEG model with a repeating unit (*n*=1) and a PEG-20kDa molecule with 455 repeating units (*n*=455). $H_{End}$ is a hydrogen atom at both end sites of the PEG molecule. The partial charges for the 1,2-Ethanediol are electrostatic potential (ESP) charges derived from QM calculation with M06-2X/cc-PVTZ++ level and the partial charges of the PEG-20kDa are determined by modifying the partial charges of the 1,2-Ethanediol.

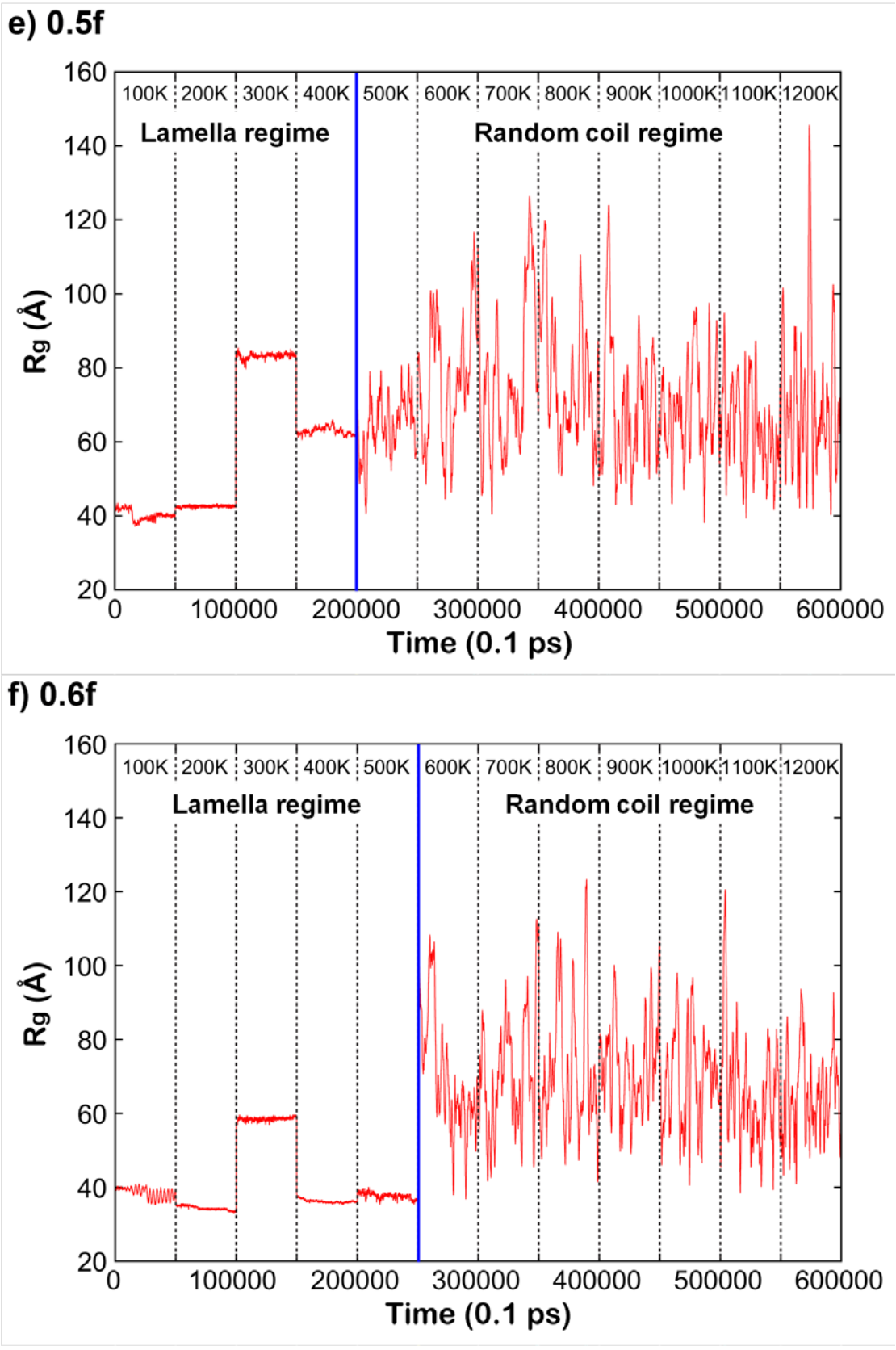


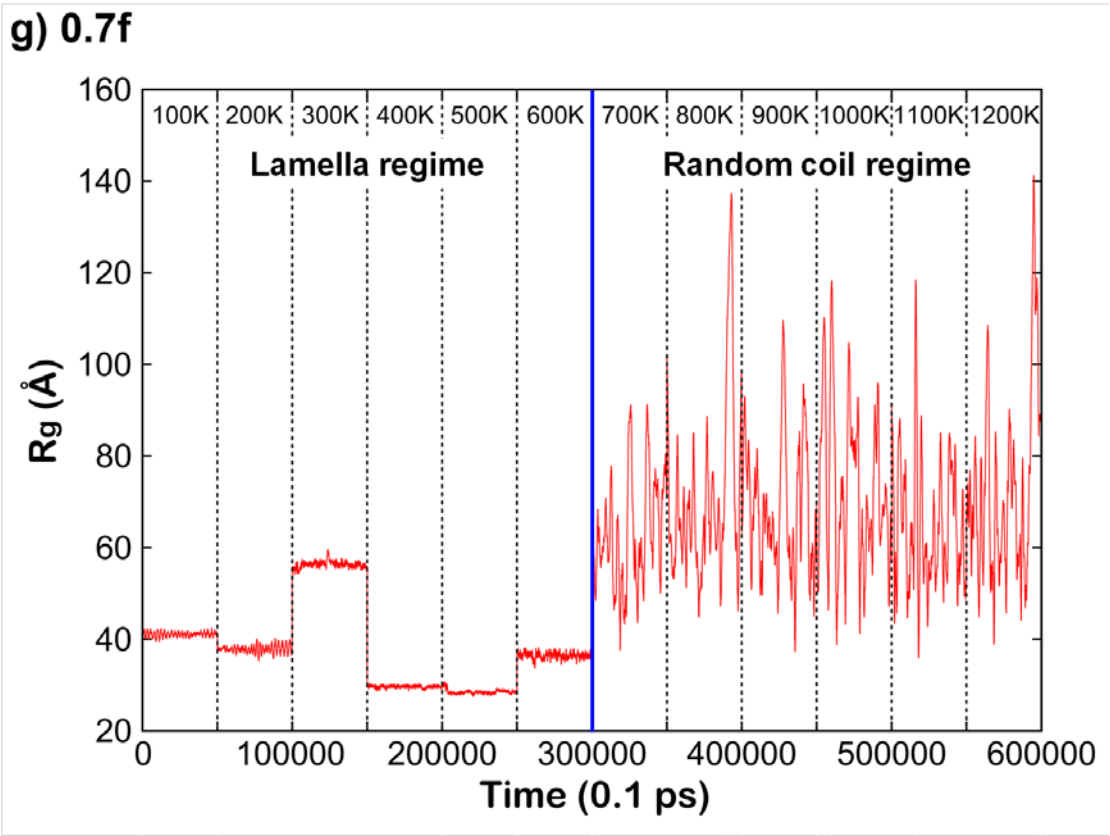

| | 1,2-Ethanediol (*n*=1) | PEG-20kDa (*n*=455) |
|---|---|---|
| **H_** | 0.00069 | 0.03699 |
| **O_3** | -0.68317 | -0.72384 |
| **C_3** | 0.28028 | 0.28765 |
| **$H_{End}$** | 0.40151 | 0.49387 |

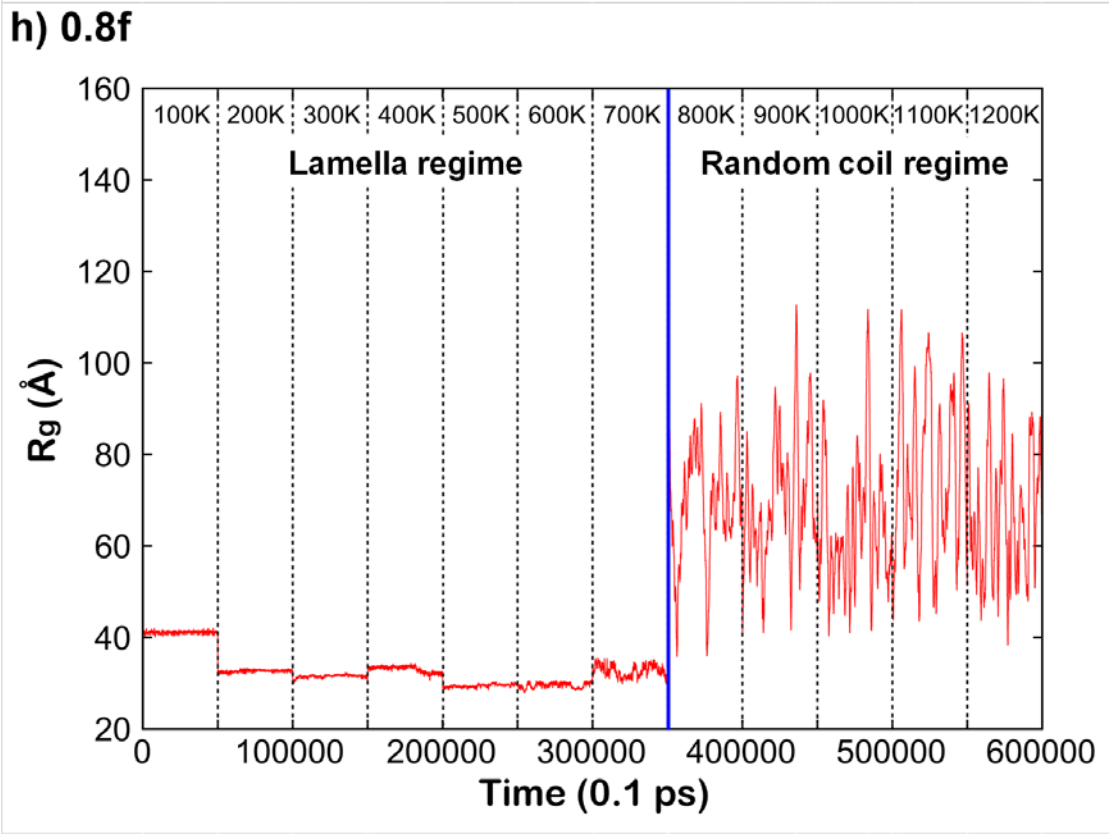

a) 0.1f
100K
200K
300K
400K
500K
600K
700K
800K
900K
1000K
1100K
1200K
Lamella regime
Random coil regime
Rg (Å)
Time (0.1 ps)
b) 0.2f
Lamella regime
Random coil regime
Rg (Å)
Time (0.1 ps)

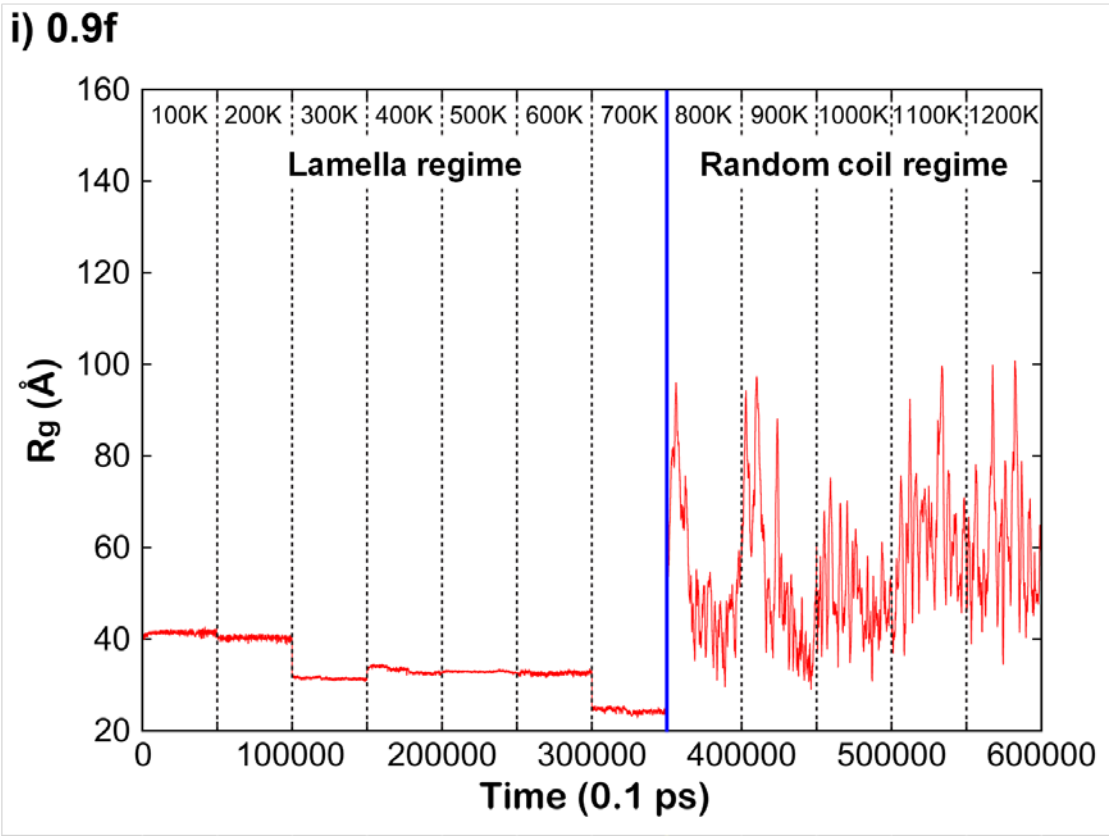

c) 0.3f
100K
200K
300K
400K
500K
600K
700K
800K
900K
1000K
1100K
1200K
Lamella regime
Random coil regime
Rg (Å)
Time (0.1 ps)
0
100000
200000
300000
400000
500000
600000
20
40
60
80
100
120
140
160


d) 0.4f
100K
200K
300K
400K
500K
600K
700K
800K
900K
1000K
1100K
1200K
Lamella regime
Random coil regime
Rg (Å)
Time (0.1 ps)
0
100000
200000
300000
400000
500000
600000
20
40
60
80
100
120
140
160


e) 0.5f
100K
200K
300K
400K
500K
600K
700K
800K
900K
1000K
1100K
1200K
Lamella regime
Random coil regime
Rg (Å)
Time (0.1 ps)
f) 0.6f
100K
200K
300K
400K
500K
600K
700K
800K
900K
1000K
1100K
1200K
Lamella regime
Random coil regime
Rg (Å)
Time (0.1 ps)


g) 0.7f
100K
200K
300K
400K
500K
600K
700K
800K
900K
1000K
1100K
1200K
Lamella regime
Random coil regime
Rg (Å)
Time (0.1 ps)
0
100000
200000
300000
400000
500000
600000
20
40
60
80
100
120
140
160


h) 0.8f
100K
200K
300K
400K
500K
600K
700K
800K
900K
1000K
1100K
1200K
Lamella regime
Random coil regime
Rg (Å)
Time (0.1 ps)
0
100000
200000
300000
400000
500000
600000
20
40
60
80
100
120
140
160


i) 0.9f
100K
200K
300K
400K
500K
600K
700K
800K
900K
1000K
1100K
1200K
Lamella regime
Random coil regime
Rg (Å)
Time (0.1 ps)
160
140
120
100
80
60
40
20
0
100000
200000
300000
400000
500000
600000

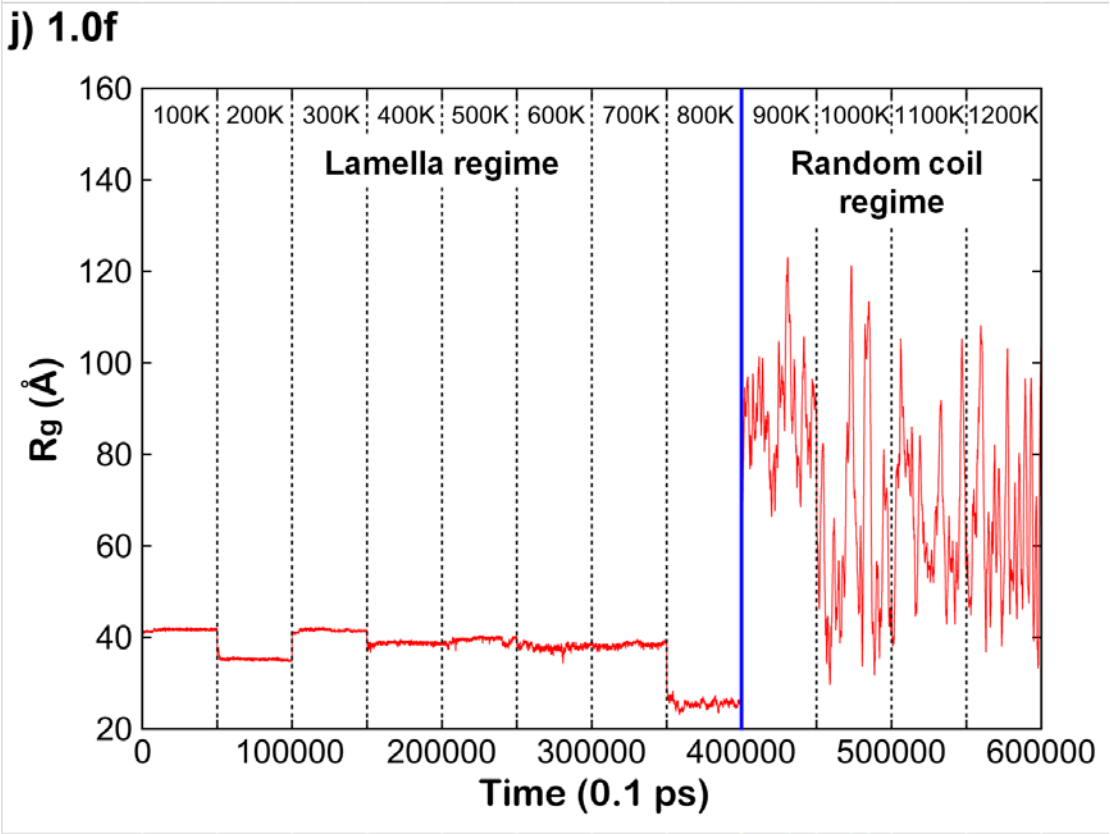

j) 1.0f
100K
200K
300K
400K
500K
600K
700K
800K
900K
1000K
1100K
1200K
Lamella regime
Random coil regime
Rg (Å)
Time (0.1 ps)
160
140
120
100
80
60
40
20
0
100000
200000
300000
400000
500000
600000

**Figure S3.** $R_g$ variation with time during individual last 5ns NVT dynamics using scaling factor, $f$; a) $f$=0.1, b) $f$=0.2, c) $f$=0.3, d) $f$=0.4, e) $f$=0.5, f) $f$=0.6, g) $f$=0.7, h) $f$=0.8, i) $f$=0.9, and j) $f$=1.0 for each temperature ranging from 100 K to 1200 K. On the left side of blue solid line, $R_g$ values change little with very small fluctuation while on the right side of the blue solid line the $R_g$ fluctuates dramatically with time. This distinction arises because below the $\theta$ temperature the strong intra-chain interactions dominate leading lamellar structures while above the θ temperature, thermal effects dominate leading to random coil (R-coil) structures. Therefore, we distinguish these two regimes as lamella regime and R-coil regime separated by the $\theta$ temperature.

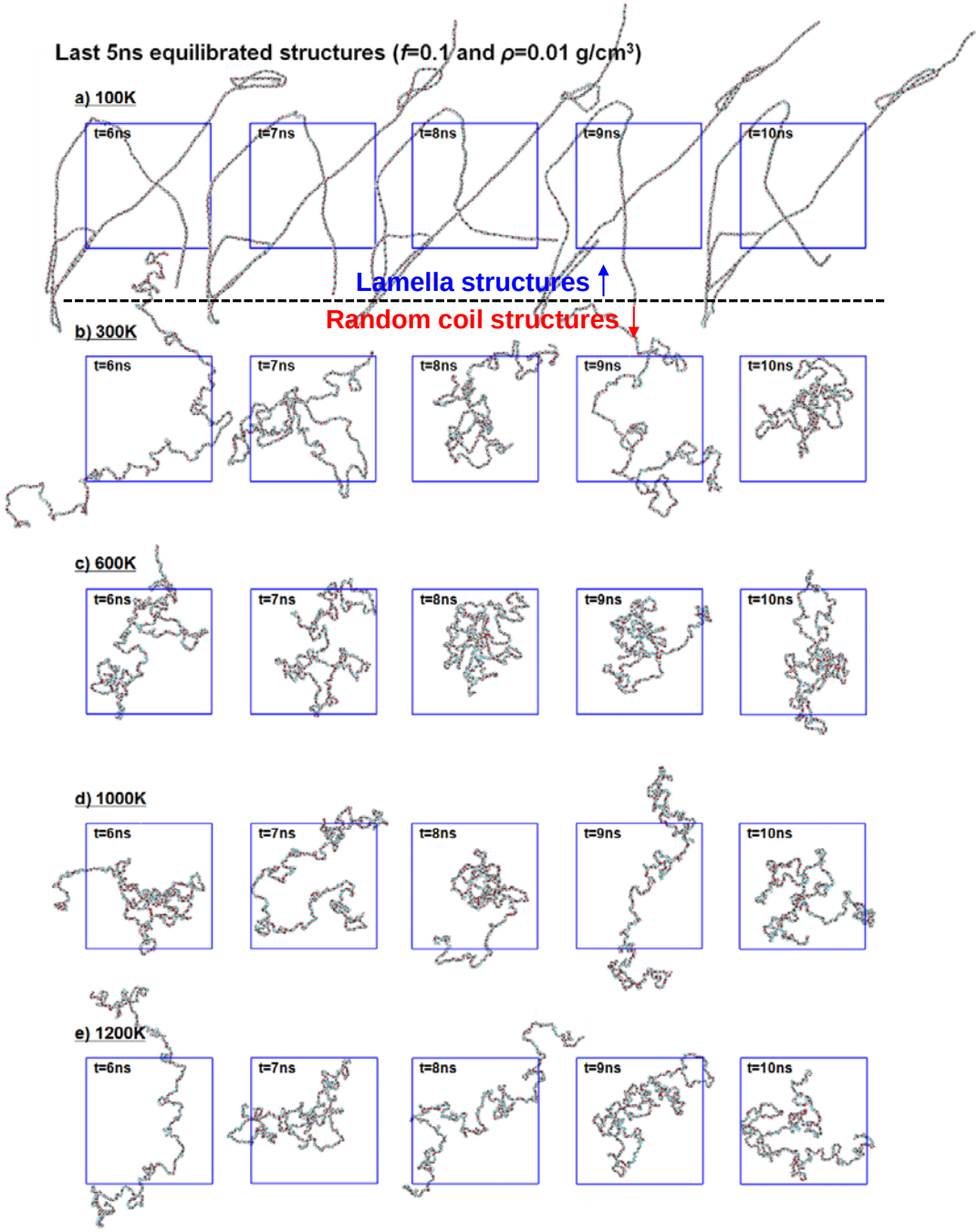



**Figure S4.** The $f$=0.1 scaled effective solvent (SES) case. Structures of a PEG-20kDa chain with $\rho$=0.01 g/cm$^3$ sampled over the last 5ns NVT dynamics a) at $T$=100 K, b) at $T$=300 K, c) at $T$=600 K, d) at $T$=1000 K, and e) at $T$=1200 K. Over the temperature range from 200 K to 1200 K, the last 5ns of equilibrated PEG-20kDa conformations all show random coil like behavior in which the backbone of the polymer moves randomly in three-dimensional space, while at 100 K and below the PEG structures maintain a lamellar shape. White, red and green atoms are hydrogen, oxygen and carbon atoms respectively. The periodic box is shown as a blue solid line.

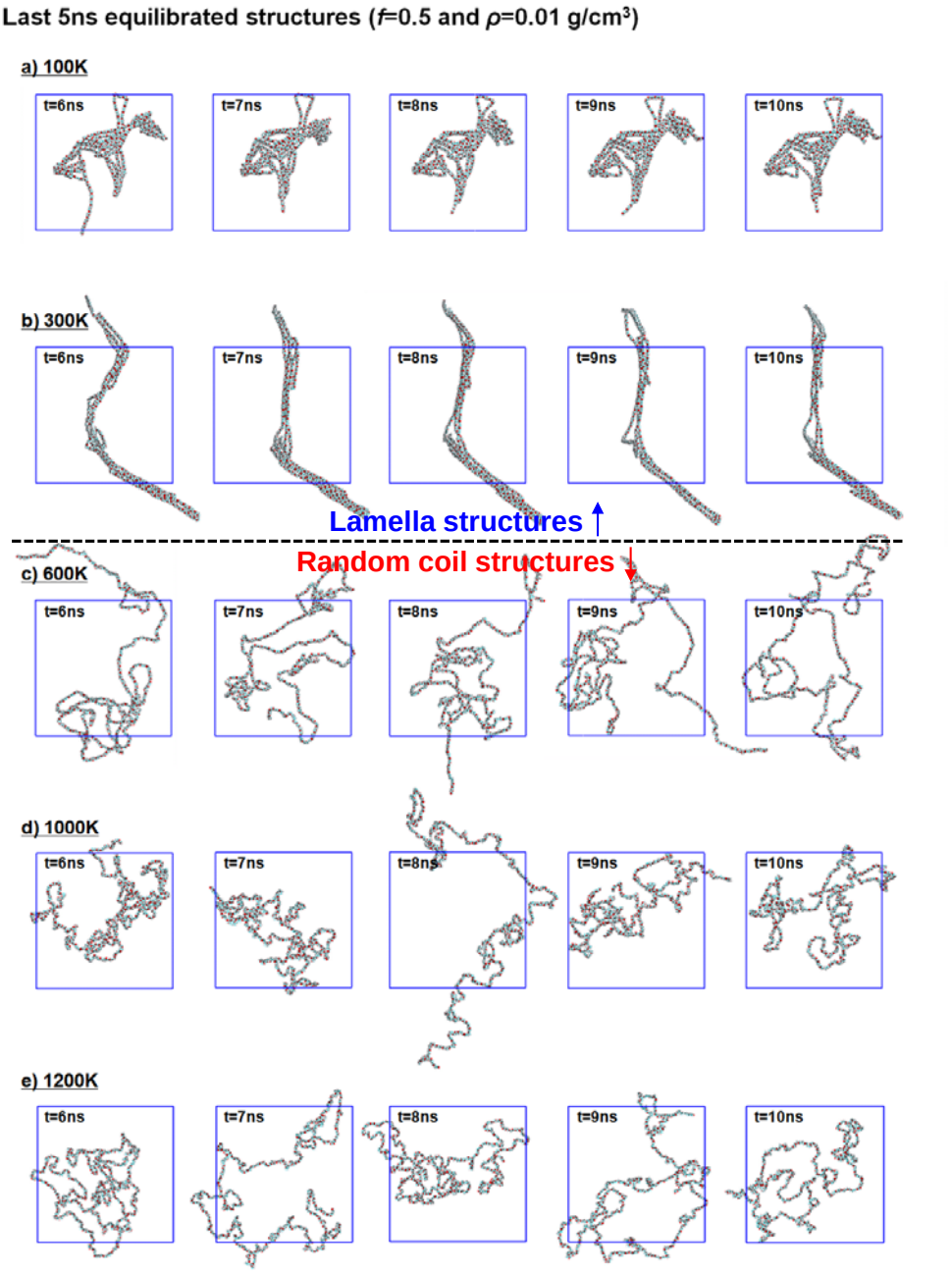



**Figure S5.** The $f$=0.5 scaled effective solvent (SES) case. Structures of a PEG-20kDa chain with $\rho$=0.01 g/cm$^3$ sampled over the last 5ns NVT dynamics a) at $T$=100 K, b) at $T$=300 K, c) at $T$=600 K, d) at $T$=1000 K, and e) at $T$=1200 K. Over the temperature range from 500 K to 1200 K, the last 5ns of equilibrated PEG-20kDa conformations all show random coil like behavior in which the backbone of the polymer moves randomly in three-dimensional space, while at 400 K and below the PEG structures maintain a lamellar shape. White, red and green atoms are hydrogen, oxygen and carbon atoms respectively. The periodic box is shown as a blue solid line.

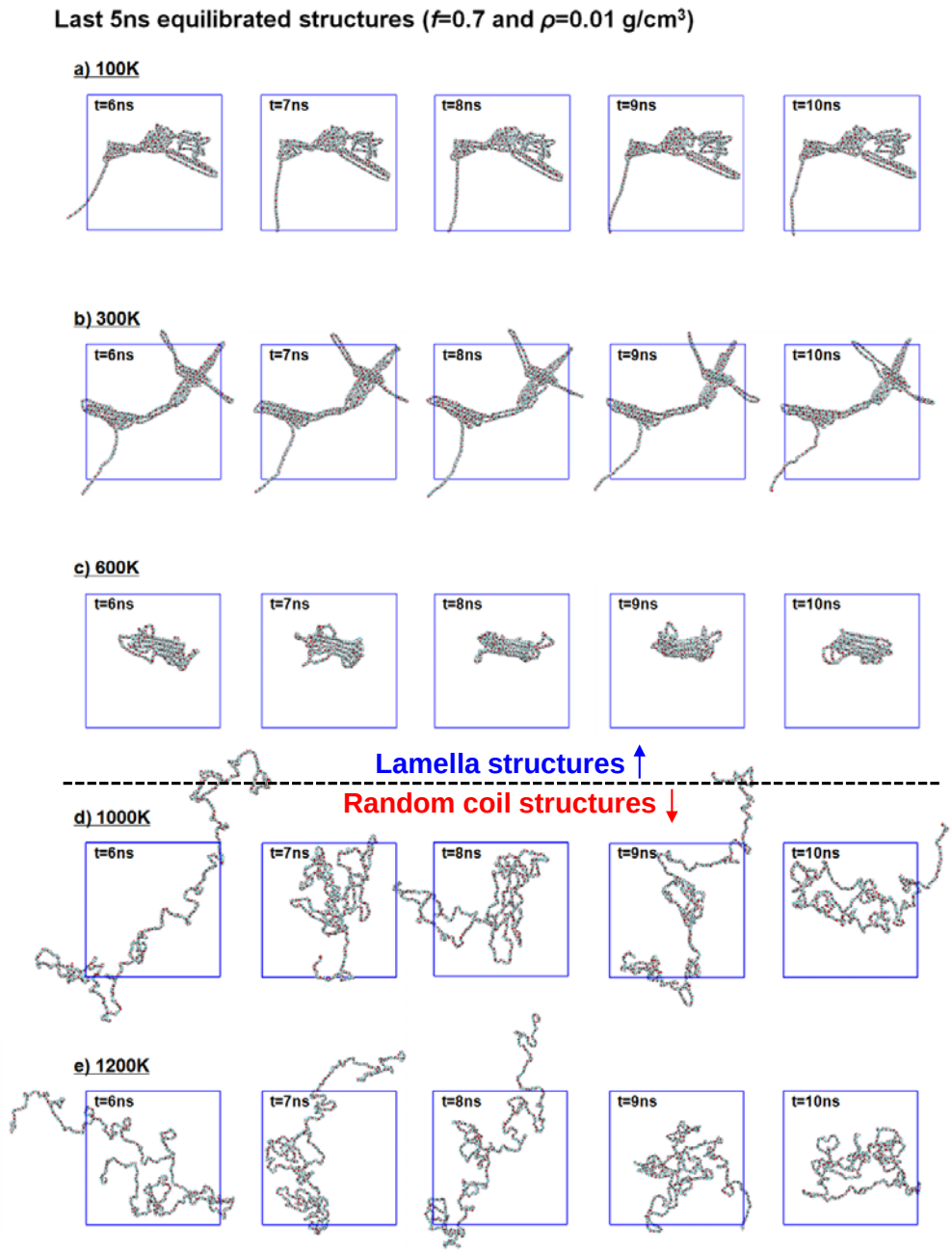
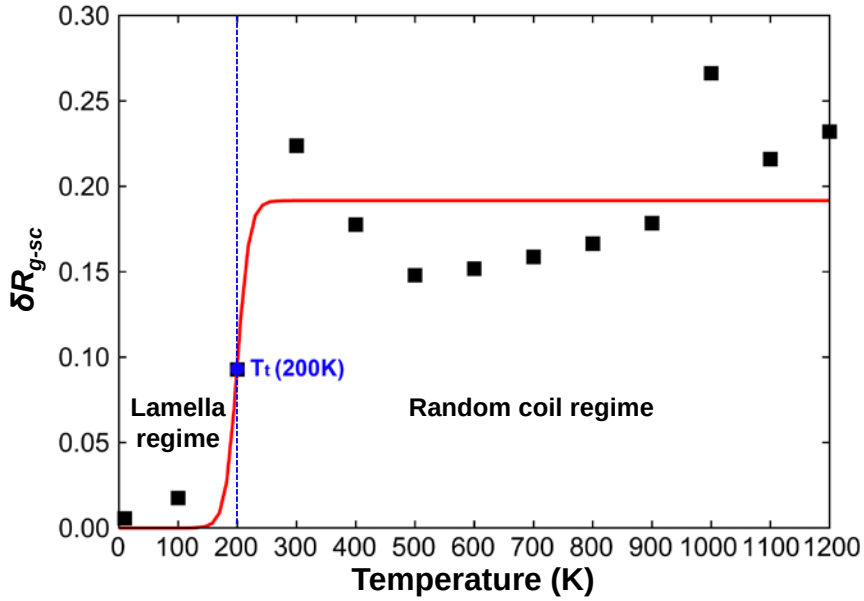
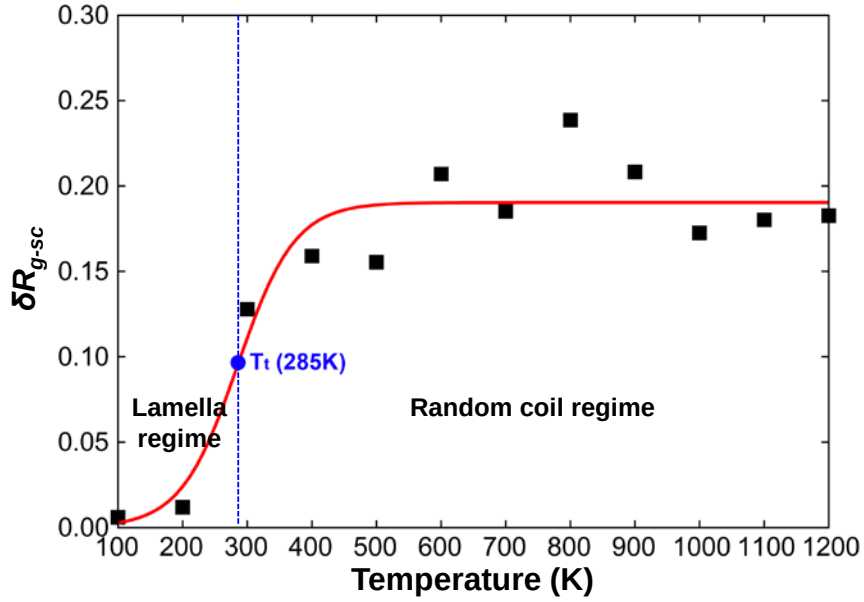



**Figure S6.** The $f$=0.7 scaled effective solvent (SES) case. Structures of a PEG-20kDa chain with $\rho$=0.01 g/cm$^3$ sampled over the last 5ns NVT dynamics a) at $T$=100 K, b) at $T$=300 K, c) at $T$=600 K, d) at $T$=1000 K, and e) at $T$=1200 K. Over the temperature range from 700 K to 1200 K, the last 5ns of equilibrated PEG-20kDa conformations all show random coil like behavior in which the backbone of the polymer moves randomly in three-dimensional space, while at 600 K and below the PEG structures maintain a lamellar shape. White, red and green atoms are hydrogen, oxygen and carbon atoms respectively. The periodic box is shown as a blue solid line.

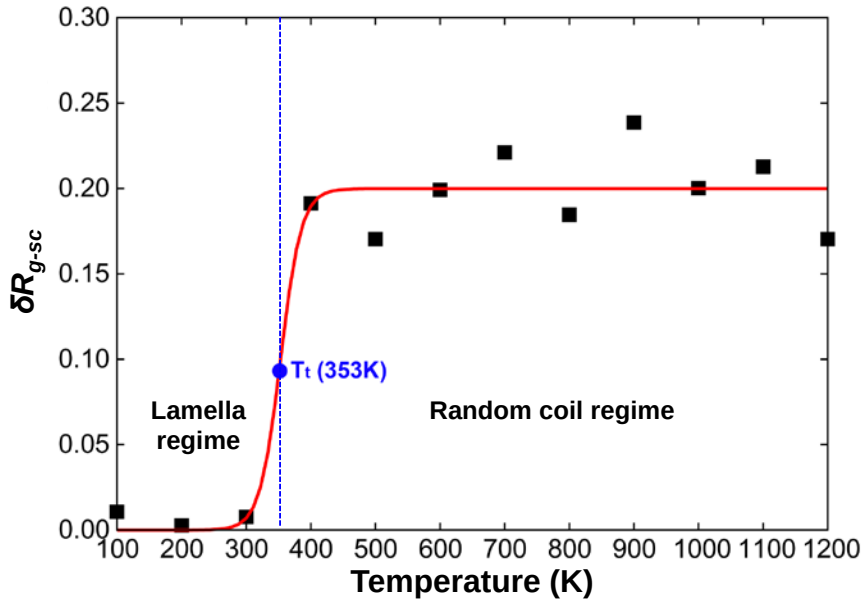
a) *f*=0.1


0.30
0.25
0.20
0.15
0.10
0.05
0.00
δR_g-sc
Tt (200K)
Lamella regime
Random coil regime
0 100 200 300 400 500 600 700 800 900 1000 1100 1200
Temperature (K)

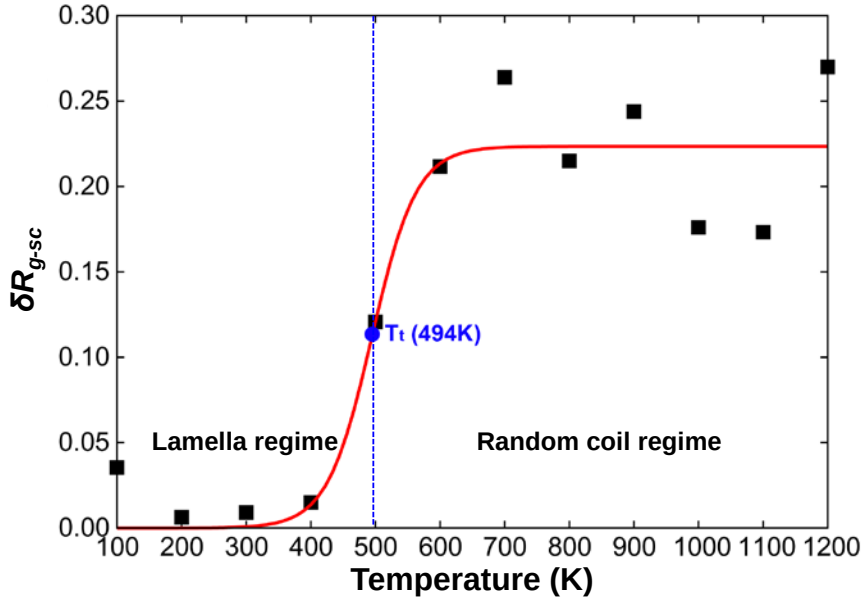
b) *f*=0.2


0.30
0.25
0.20
0.15
0.10
0.05
0.00
δR_g-sc
Tt (285K)
Lamella regime
Random coil regime
100 200 300 400 500 600 700 800 900 1000 1100 1200
Temperature (K)

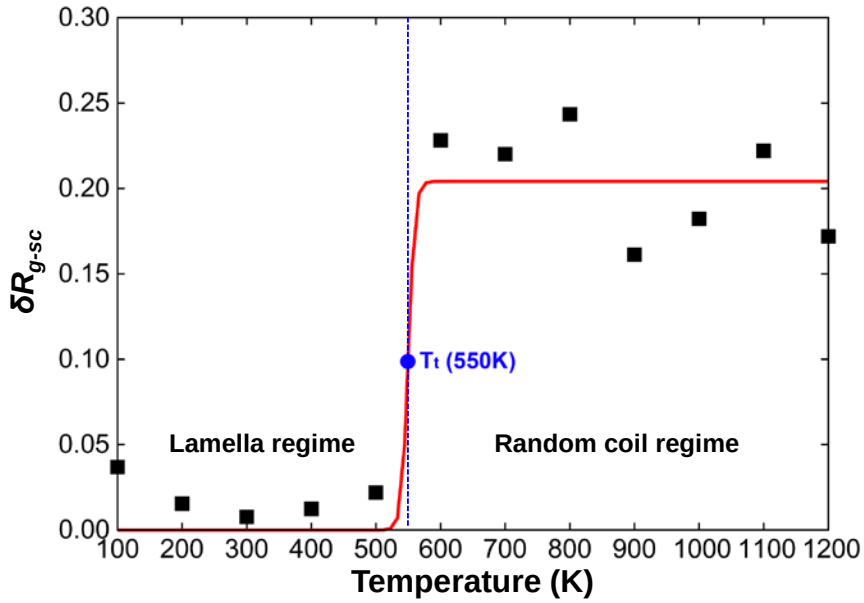
c) $f$=0.4


0.30
0.25
0.20
0.15
0.10
0.05
0.00
δR_g-sc
Tt (353K)
Lamella regime
Random coil regime
100 200 300 400 500 600 700 800 900 1000 1100 1200
Temperature (K)

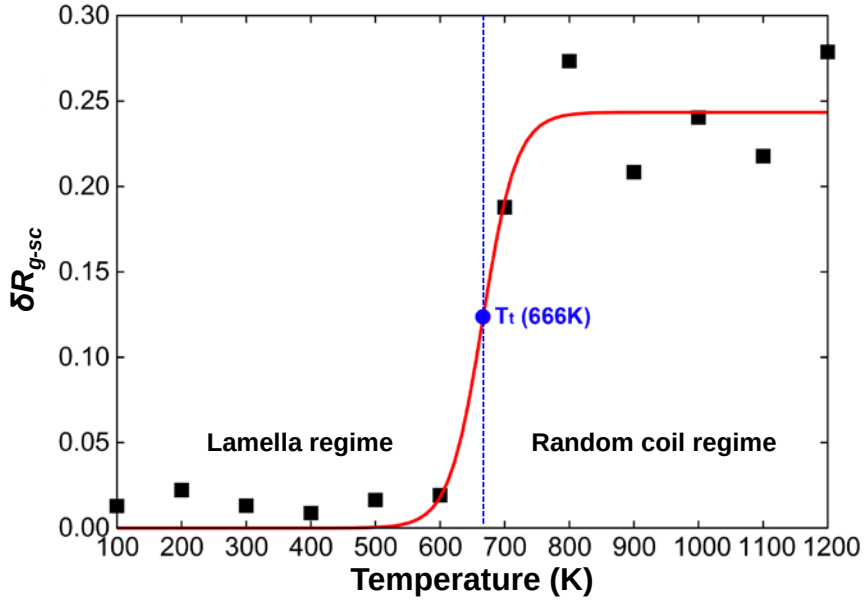
d) $f$=0.5


0.30
0.25
0.20
0.15
0.10
0.05
0.00
δR_g-sc
Tt (494K)
Lamella regime
Random coil regime
100 200 300 400 500 600 700 800 900 1000 1100 1200
Temperature (K)

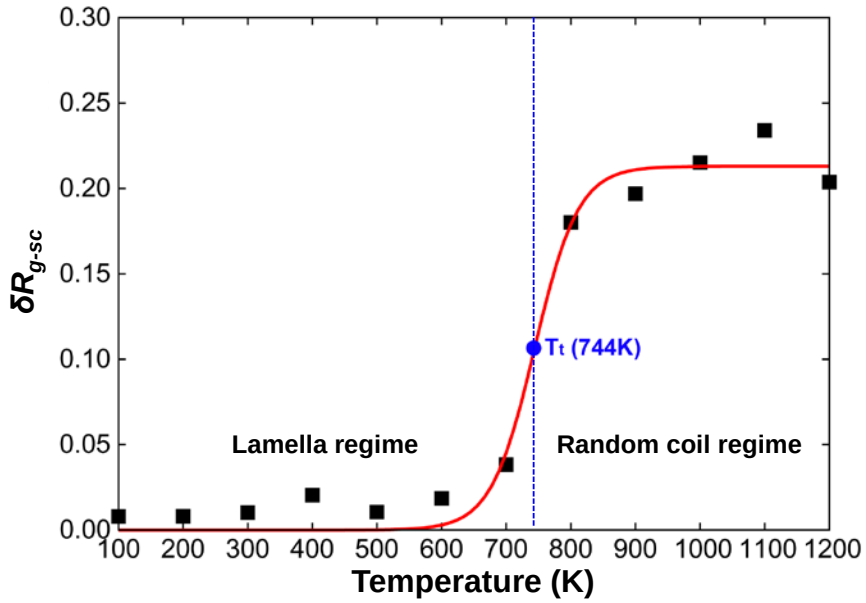
e) *f*=0.6


0.30
0.25
0.20
0.15
0.10
0.05
0.00
δR_g-sc
Tt (550K)
Lamella regime
Random coil regime
100 200 300 400 500 600 700 800 900 1000 1100 1200
Temperature (K)

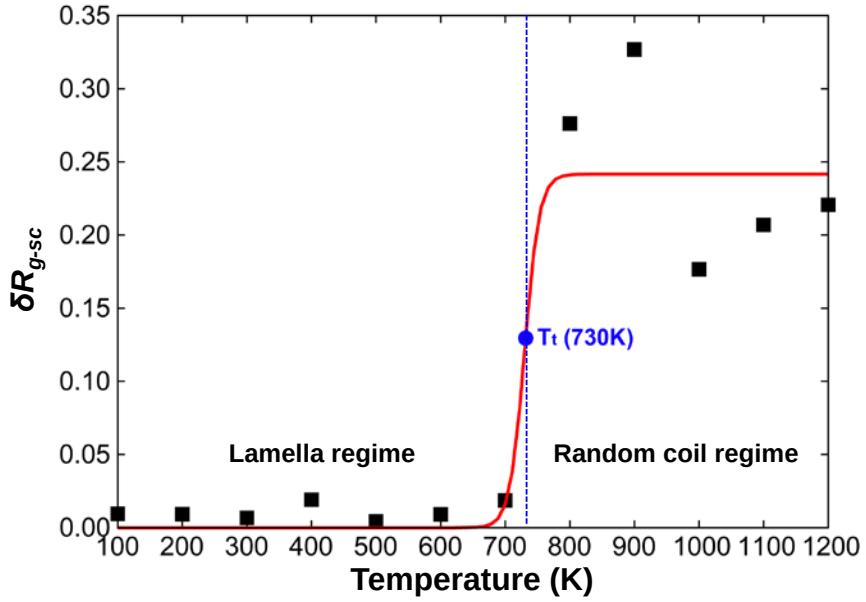
f) *f*=0.7


0.30
0.25
0.20
0.15
0.10
0.05
0.00
δR_g-sc
Tt (666K)
Lamella regime
Random coil regime
100 200 300 400 500 600 700 800 900 1000 1100 1200
Temperature (K)

g) *f*=0.8


0.30
0.25
0.20
0.15
0.10
0.05
0.00
δR_g-sc
Tt (744K)
Lamella regime
Random coil regime
100 200 300 400 500 600 700 800 900 1000 1100 1200
Temperature (K)

h) *f*=0.9


0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
δR_g-sc
Tt (730K)
Lamella regime
Random coil regime
100 200 300 400 500 600 700 800 900 1000 1100 1200
Temperature (K)

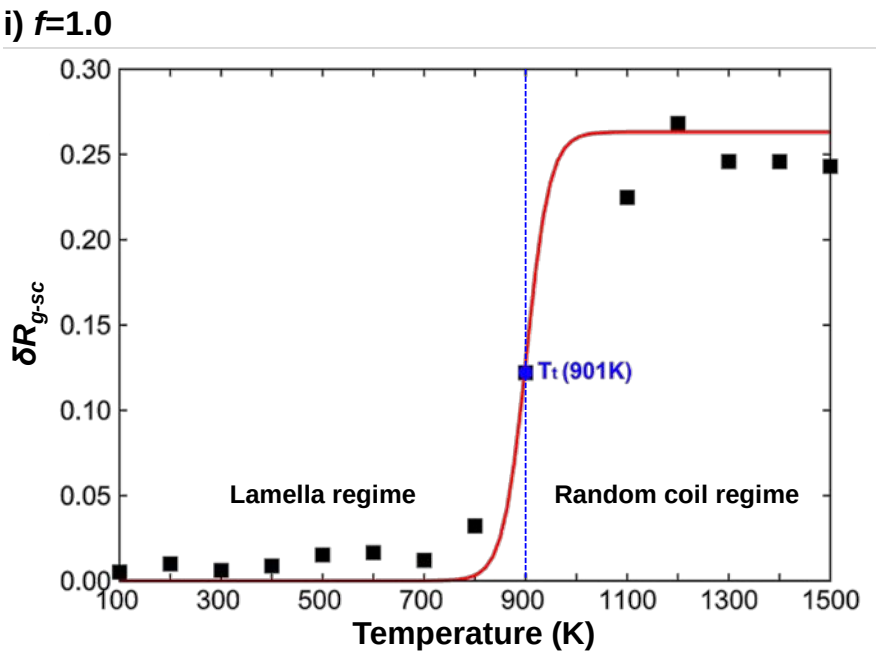



**Figure S7.** The scaled standard deviation of $R_g$ , $\delta R_{g\text{-}sc}$ = (standard deviation of $R_g$)/(average $R_g$), over the last 5ns NVT dynamics for each temperature ranging from 100 K to 1200 K (100 K to 1500 K for the case of $f$=1.0) using a large simulation cell with $\rho$=0.01 g/cm$^3$ a) $f$=0.1, b) $f$=0.2, c) $f$=0.4, d) $f$=0.5, e) $f$=0.6, f) $f$=0.7, g) $f$=0.8, h) $f$=0.9, and i) $f$=1.0 For all values of $f$, we find a similar character that low temperature leads to a globular or lamella structures with small $\delta R_{g\text{-}sc}$ (< ~0.04) and high temperature leads to a random coil (R-coil) structures with $\delta R_{g\text{-}sc}$ > ~0.09. At the highest temperature, the value of $\delta R_{g\text{-}sc}$ is similar to the value expected for an infinite Gaussian coil, 0.26. To determine a precise $\theta$ temperature ($T_t$) to serve as the boundary between these two regimes, we used the hyperbolic tangent function, $\delta R_{g\text{-}sc}(T) = a[\tanh\{(T - T_t,)/b\} + 1]$ , where $T_t$ is indicated by blue circle point.

**Table S3.** At the highest temperatures (1200K), the scaled standard deviation of $R_g$, $\delta R_{g\text{-}sc}$ = (standard deviation of $R_g$)/(average $R_g$), calculated from the hyperbolic tangent function which was derived to determine the $\theta$ temperature as a function of the scaling factor, $f$.

| $f$ | **0.1** | **0.2** | **0.3** | **0.4** | **0.5** | **0.6** | **0.7** | **0.8** | **0.9** | **1** |
|---|---|---|---|---|---|---|---|---|---|---|
| $\delta R_{g\text{-}sc}$ | 0.192 | 0.190 | 0.198 | 0.200 | 0.223 | 0.204 | 0.243 | 0.213 | 0.242 | 0.263 |

**Table S4.** The scaled standard deviation of $R_g$, $\delta R_{g\text{-}sc}$ = (standard deviation of $R_g$)/(average $R_g$), for the ideal Gaussian chains with different number of monomers or polymer chains. We carried out stochastic simulation to generate number of polymers, $N_{pol}$, with the number of segments, $n$. The fluctuation of $R_g$ is then numerically calculated, which is in turn determined as ~0.26 at the limit of $N_{pol}$ and $n$ are both infinity (regardless of the segmentation length). This compares well with the values we calculate above the transition temperature of $\delta R_{g\text{-}sc}$=0.19–0.26 (see Table S3), validating that the polymer behaves like a random coil above the transition.

| $N_{pol}$ \ $n$ | 10 | 50 | 100 | 200 | 500 | 1000 |
|---|---|---|---|---|---|---|
| **100** | 0.231 | 0.263 | 0.254 | 0.257 | 0.255 | 0.253 |
| **1000** | 0.230 | 0.246 | 0.259 | 0.257 | 0.249 | 0.259 |
| **10000** | 0.227 | 0.255 | 0.262 | 0.259 | 0.255 | 0.261 |
| **100000** | 0.227 | 0.255 | 0.258 | 0.259 | 0.258 | 0.258 |
| **1000000** | 0.227 | 0.254 | 0.257 | 0.258 | 0.259 | 0.259 |

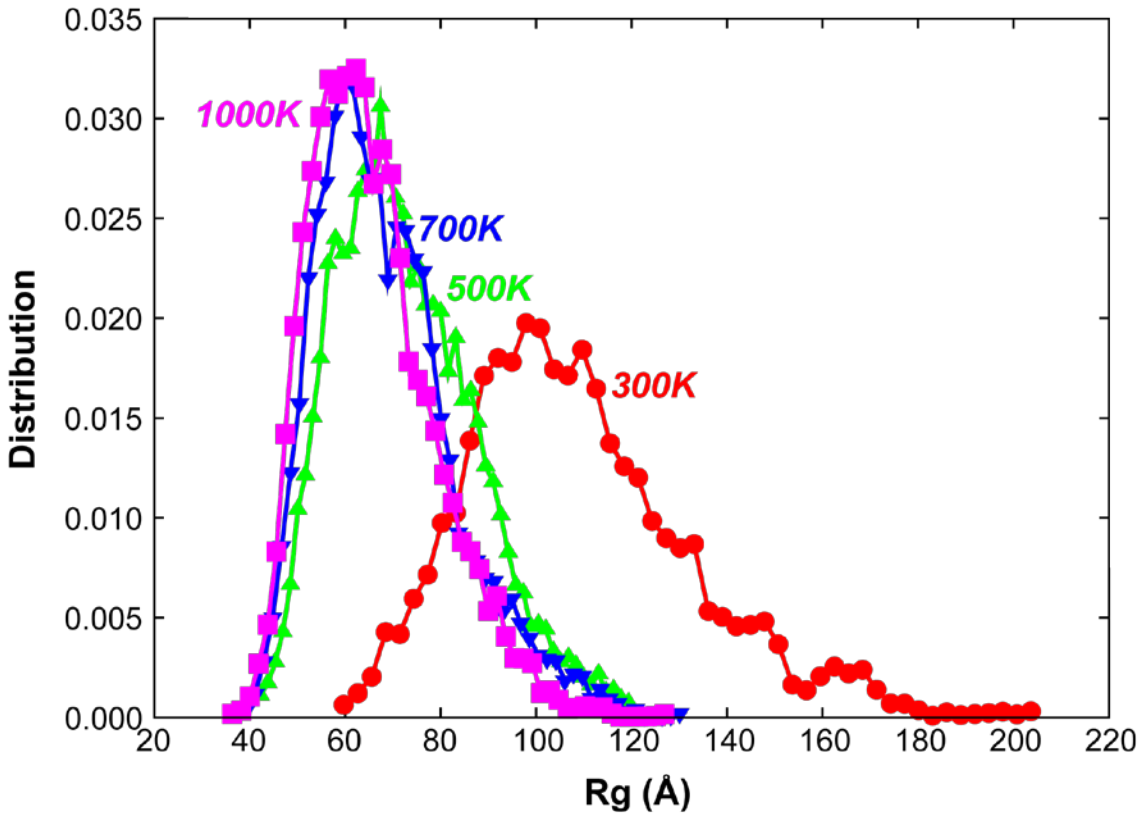



**Figure S8.** $R_g$ distribution of random coils sampled with scaling factor of 0.3 as a function of temperature. The case of 300 K is red-colored, the case of 500 K is green-colored, the case of 700 K is blue colored, and the case of 1000 K is magenta-colored. For each temperature, $R_g$ distribution of random coils is obtained from MD simulation for 110 ns.

In addition to the globule region we find two range of temperature.

1) Intermediate temperature range where there is nearly a balance between thermal effects and the scaled chain-chain interaction, but not sufficiently high to freely overcome the torsional barriers. This leads the sampling of random-coiled structures but biased to sample the expanded forms due to the stiffness of the polymer backbone, see for example the $T$=300 K case for $f$=0.3 (red solid line)

2) High temperature range where $kT$ is high enough to freely overcome the torsional barriers, leading to bias toward more packed random-coiled conformations. See the $T$=500 K and above for the case with $f$=0.3

From the 2PT free energy analysis in the last part of our manuscript, sampling this full temperature range is important since all sampled conformations from all different temperatures have comparable free energy values after including solvation effect explicitly in full solvent 300 K dynamics.

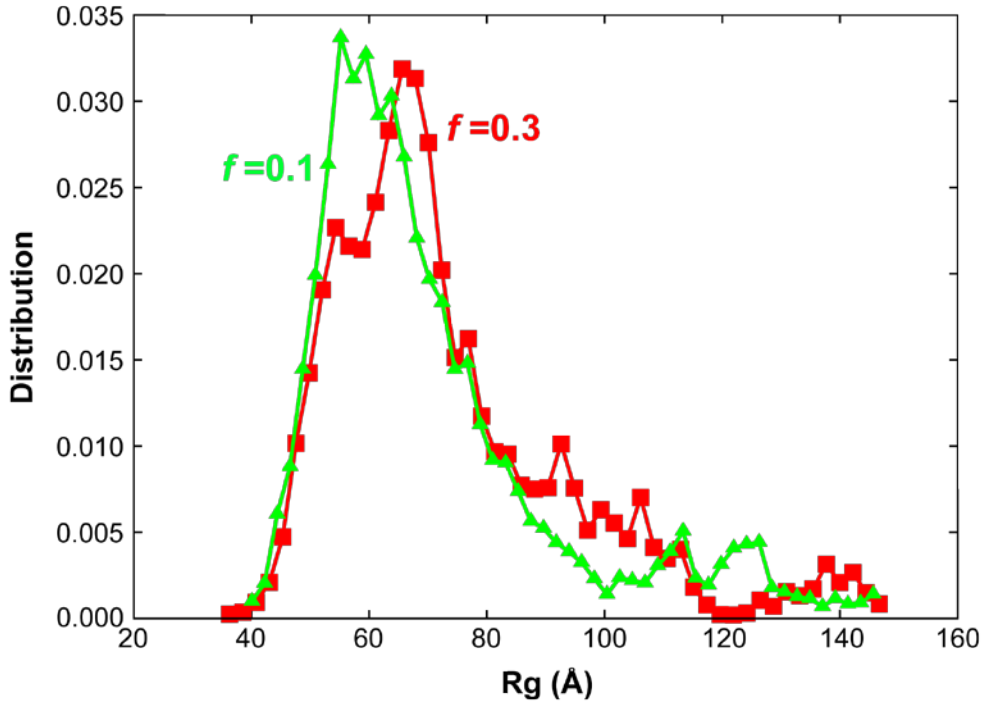



**Figure S9.** $R_g$ distributions obtained by using $f$=0.3 (red) or 0.1 (green) and varying temperature in large simulation box (140×140×140 Å$^3$) with $\rho$=0.01 g/cm$^3$. For all cases, we include all temperatures above the $\theta$ temperature (that is all temperatures for the R-coil structures from 100 K up to ~1,200 K).

This figure shows that the case of $f$=0.1 but varying the temperature up to ~1,200 K leads to a similar distribution to that of the case with $f$=0.3 with a similar $\Delta R_g$ ~ 110 Å although it leads to a most probable value of $R_g$, 55.17 Å while $f$=0.3 leads to 65.58 Å. This infers that using $f$=0.1 is also an option for SES dynamics.

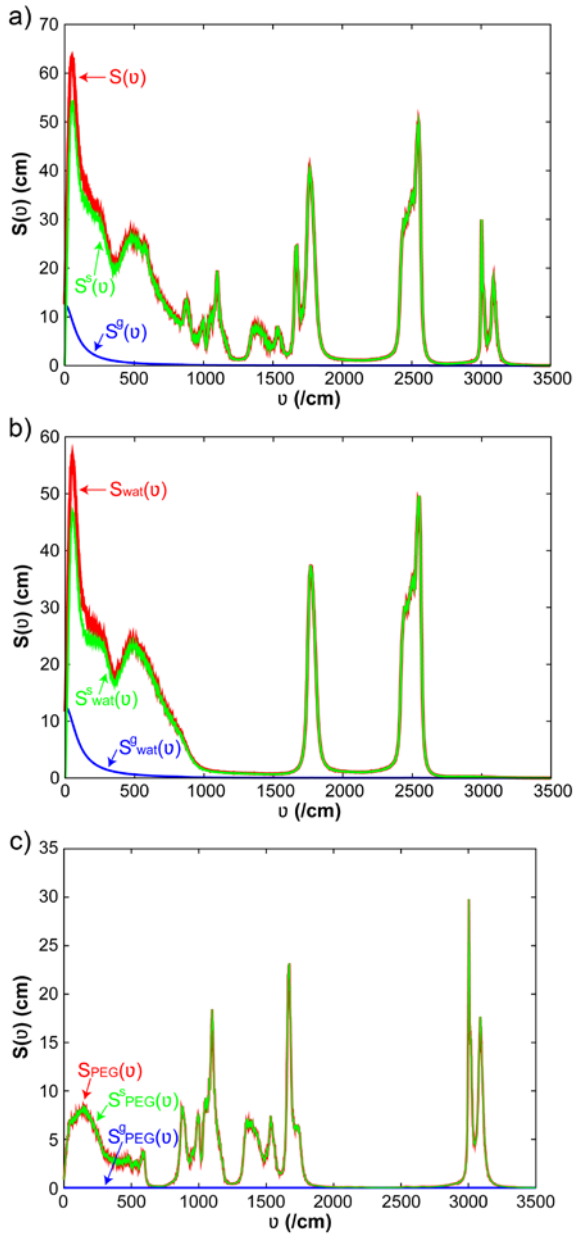



**Figure S10.** Starting with the PEG chain having $R_g$=34.86 Å, density of states (DoS) from the Fourier transform of the velocity autocorrelation function by the 2PT analysis at 300 K for the 25 wt% PEG aqueous solution

a) Total DoS $S(\upsilon)$ and the gas-like $S^g(\upsilon)$ and solid-like $S^s(\upsilon)$ components.

b) DoS of water $S_{wat}(\upsilon)$ and gas-like $S^g_{wat}(\upsilon)$ and solid-like $S^s_{wat}(\upsilon)$ components.

c) DoS of PEG $S_{PEG}(\upsilon)$ and gas-like $S^g_{PEG}(\upsilon)$ and solid-like $S^s_{PEG}(\upsilon)$ components. Most diffusional contribution is from the water dynamics.

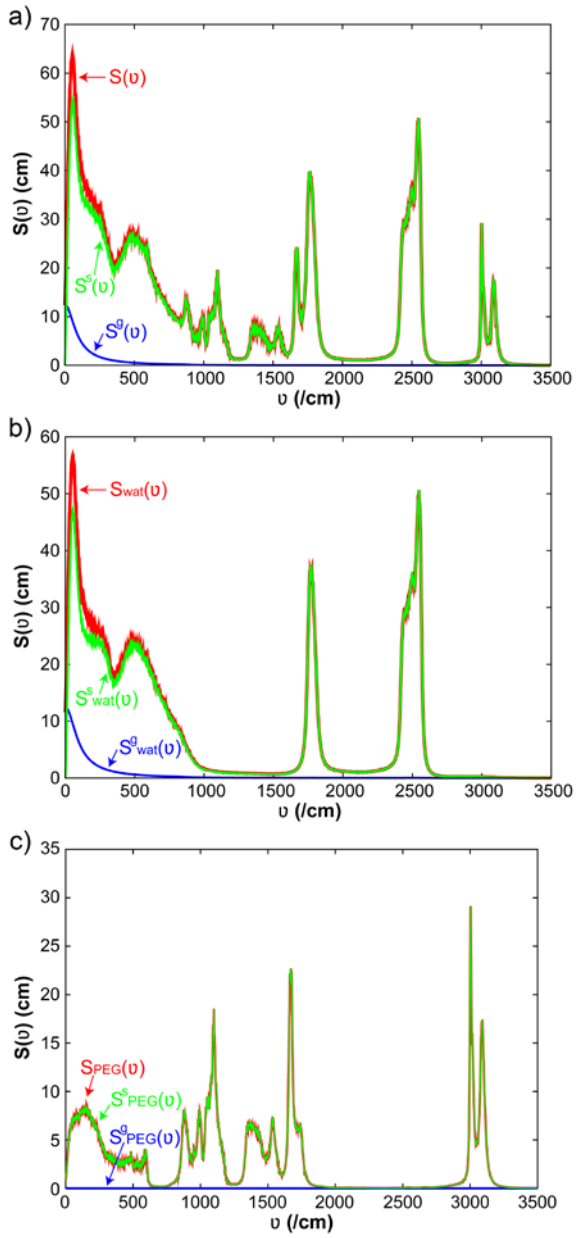



**Figure S11.** Starting with the PEG chain having $R_g$=55.67 Å, density of states (DoS) from the Fourier transform of the velocity autocorrelation function by the 2PT analysis at 300 K for the 25 wt% PEG aqueous solution

a) Total DoS $S(\upsilon)$ and the gas-like $S^g(\upsilon)$ and solid-like $S^s(\upsilon)$ components.

b) DoS of water $S_{wat}(\upsilon)$ and gas-like $S^g_{wat}(\upsilon)$ and solid-like $S^s_{wat}(\upsilon)$ components.

c) DoS of PEG $S_{PEG}(\upsilon)$ and gas-like $S^g_{PEG}(\upsilon)$ and solid-like $S^s_{PEG}(\upsilon)$ components. Most diffusional contribution is from the water dynamics.

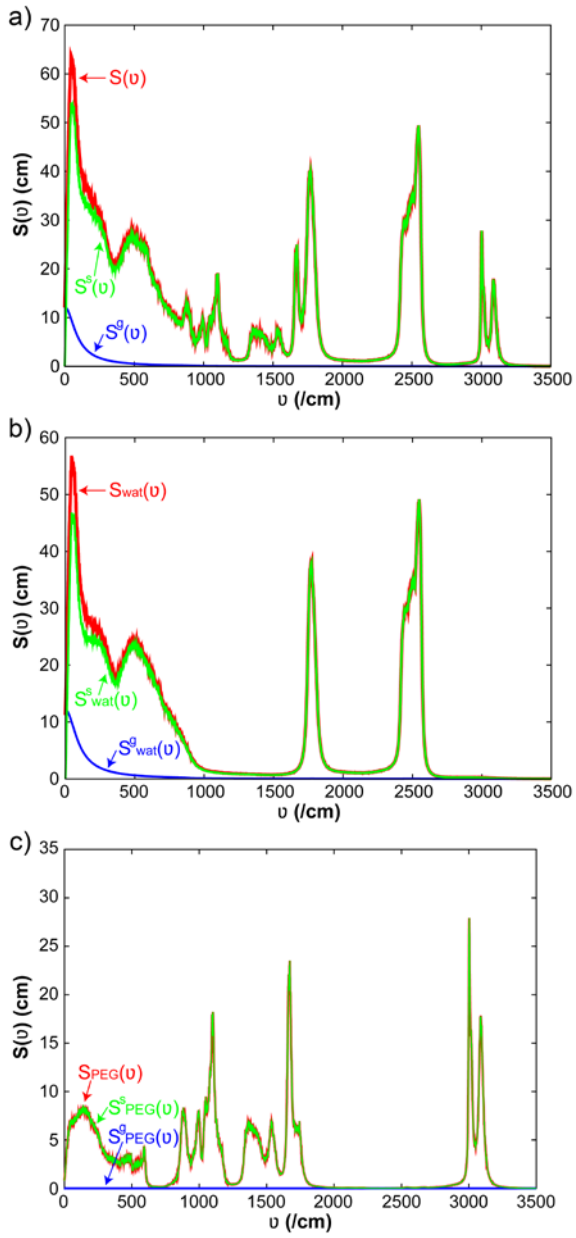



**Figure S12.** Starting with the PEG chain having $R_g$=78.77 Å, density of states (DoS) from the Fourier transform of the velocity autocorrelation function by the 2PT analysis at 300 K for the 25 wt% PEG aqueous solution

a) Total DoS $S(\upsilon)$ and the gas-like $S^g(\upsilon)$ and solid-like $S^s(\upsilon)$ components.

b) DoS of water $S_{wat}(\upsilon)$ and gas-like $S^g_{wat}(\upsilon)$ and solid-like $S^s_{wat}(\upsilon)$ components.

c) DoS of PEG $S_{PEG}(\upsilon)$ and gas-like $S^g_{PEG}(\upsilon)$ and solid-like $S^s_{PEG}(\upsilon)$ components. Most diffusional contribution is from the water dynamics.