Macromolecular Rapid Communications
www.mrc-journal.de

A Simple Generic Model of Elastin–Like Polypeptides with Proline Isomerization

*Yani Zhao, Robinson Cortes–Huerto, and Debashish Mukherji**

A generic model of elastin–like polypeptides (ELP) is derived that includes proline isomerization (ProI). As a case study, conformational transition of a –[valine–proline–glycine–valine–glycine]– sequence is investigated in aqueous ethanol mixtures. While the non–bonded interactions are based on the Lennard–Jones (LJ) parameters, the effect of ProI is incorporated by tuning the intramolecular 3– and 4–body interactions known from the underlying all–atom simulations into the generic model. One of the key advantages of such a minimalistic model is that it readily decouples the effects of geometry and the monomer–solvent interactions due to the presence of ProI, thus gives a clearer microscopic picture that is otherwise rather nontrivial within the all–atom setups. These results are consistent with the available all–atom and experimental data. The model derived here may pave the way to investigate large scale self–assembly of ELPs or biomimetic polymers in general.

1. Introduction

Elastin–like polypeptides (EPL) are a modern class of synthetic biomimetic polymers where establishing a structure–property relation is fundamentally challenging.[1–4] Typically, ELPs have a standard sequence –[VPG**X**G]– with V, P, and G being valine, proline, and glycine amino acids, respectively. Here, **X** can be any amino acid but proline. ELPs are of particular importance because their underlying peptide–based nature makes them biocompatible and a suitable material class for the design of functional soft biomaterials.[5–10]

For the broad applications of ELPs, their solution processing in a variety of single and multiple component solvents are routinely employed.[1–4,11,12] In this context, ELPs in pure water show the typical lower critical solution (LCST) behavior with a transition temperature T_ℓ, i.e., an ELP remains expanded for a temperature $T < T_\ell$ and collapses when $T > T_\ell$.[1,13] More specifically, when $T \to T_\ell$ a certain number of solvent–polymer interactions break. These expelled solvent molecules gain translational entropy that is significantly larger than the polymer conformational entropy loss upon collapse, making LCST a solvent entropy driven process.[14] Note that in the microgel community, T_ℓ is referred to as the volume phase transition temperature T_{VPTT}.[15–17]

A particular chemical sequence usually have a well–defined T_ℓ. However, T_ℓ can be tuned by using complex copolymer sequences. In the case of standard commodity polymers, T_ℓ can change by changing the relative hydrophobic–hydrophilic sequences along a backbone.[18–22] Here, ELPs are a natural choice where their peptide–based sequence can be used to obtain a desired T_ℓ, which can be achieved by changing the nature of **X** along an ELP backbone. For example, $T_\ell \simeq 300 - 305$ K when **X** is valine[23] and $T_\ell \simeq 305 - 310$ K for glycine,[3,24] which is not surprising give that a glycine is more hydrophilic than valine.

Traditionally, phase behavior of ELP is mostly investigated in pure water.[1,3,4] A recent experimental study also revealed that the ELP sequences can exhibit the co–non–solvency (CNS) phenomenon in aqueous ethanol mixtures.[11] The term CNS is associated with an intriguing phenomenon of polymer collapse in a mixture of two (fairly) miscible and individual good solvents for the same polymer.[8,9,25] This counter–intuitive solvation phenomenon was initially reported for polystyrene in cyclohexane and DMF mixtures,[26] while it has gained popularity in the context of the "smart" polymers[15,17,27–30] and ELPs in binary mixtures.[11,12] More specifically, when ethanol molecules are added to an ELP–water solution, T_ℓ first decreases with increasing ethanol mole fraction x_e and then increases again at higher x_e. The direct consequence of such a LCST behavior with x_e is that– when x_e increases at $T \simeq 300$ K, an ELP first collapses around $x_e \simeq 0.05$ and a further increase of x_e above $\simeq 0.14$ again expands the same ELP. As a result an ELP goes from a coil–to–globule–to–coil transition with x_e, as known from the CNS phenomenon.

Y. Zhao
Bruker Daltonics GmbH & Co. KG
28359 Bremen, Germany
R. Cortes–Huerto
Max Planck Institute for Polymer Research
Ackermannweg 10, 55128 Mainz, Germany
D. Mukherji
Quantum Matter Institute
University of British Columbia
Vancouver V6T 1Z4, Canada
E-mail: debashish.mukherji@ubc.ca

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/marc.202400304

[The copyright line for this article was changed on August 5, 2024 after original online publication.]

© 2024 The Author(s). Macromolecular Rapid Communications published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

DOI: 10.1002/marc.202400304

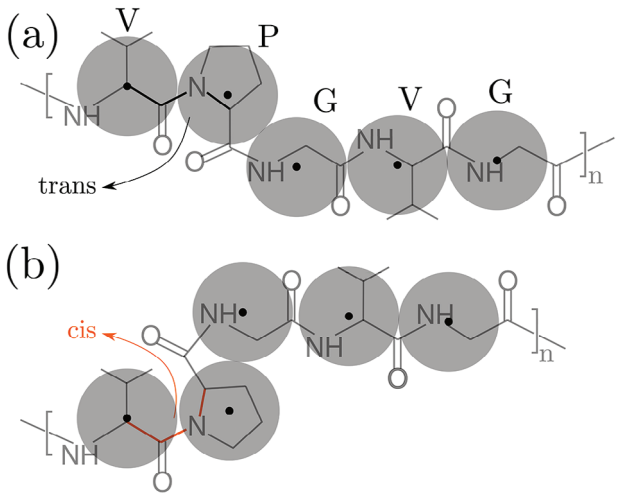

Figure 1. Schematic representation of a *trans* (part a) and a *cis* (part b) states of a –[VPGVG]– ELP sequence. Here, V, P, and G are valine, proline, and glycine amino acids, respectively. For this study, we have chosen the number of repeat unit $n = 10$. The grey circles represent a mapping scheme from the all–atom to the generic models, where one amino acid is represented one generic bead.

The discussions presented so far deal with most common ELP sequences. However, another interesting feature of the ELP sequences is their ability to change conformations via proline isomerization (ProI). For example, ProI can either be in a *trans* or in a *cis* state, see the representative schematics in **Figure 1**. The free–energy difference between these two states is about $2k_BT$, while their barrier is over $30k_BT$, suggesting that the transition between states is a slow kinetic process.[31,32] Furthermore, such a *cis*–to–*trans* transition is switchable via light responsiveness. In this context, light responsive polymers are of great technological relevance[33–35] and thus the ELP sequences with ProI provide a natural choice. A direct consequence of this conformational behavior is their possible use in obtaining the tunable large aggregation and scale self–assembly.[36]

So far, the majority of studies have dealt with the generic ELP behavior, while only a limited number of studies have investigated the influence of *cis*–to–*trans* switching on ELP conformations.[31,32,37] Here, one of the important questions is– what causes such a conformational switching via ProI? More specifically, if an ELP changes conformation by changing ProI, it is predominantly because of the modified solvation structure around an ELP. Such a modification in the solvation shell can either be due to the change in proline interactions with the solvent molecules or because of the local geometric arrangement of a short segment involving a proline along an ELP backbone. In particular, it is of great importance to exactly understand the extent by which these two (possibly competing, yet related) effects influence the solvation of ELPs. This is a grand challenge because the solvation properties are dictated by large conformational/compositional fluctuations,[38] especially within the standard all–atom simulations where complex competing interactions complicate the scenario even more and thus hinders a clear distinction of different microscopic effects. Given the above discussion, a simple generic model of ELP with ProI would be more suitable to capture the specific solvent interactions and the local conformational effects, which to the best of our knowledge is lacking.

Motivated by the above discussion, we derive a generic model of ELP in multi–component solvent mixtures that incorporates ProI. This derived model is used to study the conformational behavior of ELPs in binary solvents. While the results are compared with the available experimental and all–atom simulation data, a key added advantage of our model is that it readily provides separate microscopic effects dictating the ELP solvation in single solvent or in binary mixtures.

The remainder of this draft is organized as follows: in Section 2, we sketch the key ingredients of the basic model of ELP in binary solution. In Section 3 A, we start by discussing the derived ProI–based generic model and use this model to then discuss our results in Section 3 B. A discussion based on the local geometry and intra–molecular interactions in present in Section 3 C. Finally, the conclusions are drawn in Section 4.

2. Model and Method

For this study, we employ a generic molecular dynamics using the LAMMPS package,[39] where specific inter– and intra–molecular interactions are incorporated that are motivated by the underlying all–atom model of ELP. Here, the interactions are categorized in– the polymer model, the monomer–solvent and the monomer–cosolvent interactions, and the intra–molecular (angular and dihedral) interactions. In the following, we sketch the specific details of all these interactions.

2.1. The Polymer Model

The bare ELP backbone is modelled using the well–known bead–spring model of polymers.[40] Within this model, the bonded monomers interact by a combination of repulsive 6–12 Lennard–Jones (LJ) potential with a cutoff $r_c = 2^{1/6}\sigma$ and a finitely extensible non–linear elastic (FENE) potential. As illustrated by the grey circles in Figure 1, this work maps one amino acid onto one generic monomer. Therefore, a –[VPGVG]– segment consists of five generic monomers, while the number of repeat units of such a segment forming an ELP chain is chosen as $n = 10$. This results in a representative generic ELP chain with a total length $N_\ell = 5n = 50$. This specific N_ℓ is chosen because it shows a reasonable coil–to–globule transition, while equilibration of the chain conformation and the solvation structure around a chain remains within the reasonable range of the computational time. We also note in passing that the synthesis of ELP sequences with $n > 10 - 20$ is a grand challenge. Here, most experimental studies usually deal with small sequences and study their aggregation under their finite concentration.[11,36] In this context, we believe that our choice of N_ℓ is within the reasonable synthesis range.

The non–bonded monomers interact via an attractive LJ potential with $r_c = 2.5\sigma$. For the simplicity of modelling, we have taken monomer–monomer interactions between different amino acids to be identical, i.e., $\epsilon_{ij} = 0.5\epsilon$ and $\sigma_{ij} = 1.0\sigma$ with i & j = {V, P, G}. This is a safe choice given that our study aims to investigate an ELP under good solvent condition in pure solvent, i.e., the typical case of $T < T_\ell$. The results are presented in the units

 © 2024 The Author(s). Macromolecular Rapid Communications published by Wiley-VCH GmbH

15213927, 2024, 18, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/marc.202400304 by MPI 355 Polymer Research, Wiley Online Library on [02/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

of LJ energy ϵ, LJ length σ, LJ time $\tau = \sigma\sqrt{m/\epsilon}$, and mass of a monomer m. The representative real units for the above mentioned mapping scheme gives $\sigma \simeq 0.5$ nm, $\epsilon/k_B \simeq 300$ K, pressure $p = 32\epsilon/\sigma^2 \simeq 1$ atm, $m \simeq 62$ gmol^{-1}, and thus lead to an unit of time $\tau = \sigma\sqrt{m/\epsilon} \simeq 2.5$ ps.

A single chain is solvated in a cubic box of $N = 5 \times 10^4$ solvent particles, i.e., the total number of solvent and cosolvent molecules, but at different cosolvent mole fraction $0.0 \leqslant x_c \leqslant 1.0$. Here, $x_c = 0.0$ represents the pure solvent phase and $x_c = 1.0$ corresponds to the pure cosolvent phase.

Given that the monomer sizes are taken as $\sigma_{ij} = 1\sigma$, the sizes of the solvent (s) and the cosolvent (c) molecules are chosen as $\sigma_s = 0.5\sigma$ and $\sigma_c = 0.9\sigma$, respectively. These relative sizes are taken from the all–atom data of aqueous ethanol mixtures,[17] where the typically diameter of a water molecules is about half the size of a monomer and an ethanol is about 1.8 times the water molecules.[12,17] Furthermore, ϵ_{ij} between the (co–)solvent components is tuned to reproduce the total (co–)solvent number density ρ with x_c at a constant pressure, see refs. [12, 41] for more details.

2.2. (Co–)Solvent–Monomer Interactions

To model the monomer–solvent and monomer–cosolvent interactions, we use our earlier developed model that mimics correct solvation behavior of a generic ELP chain in solvent–cosolvent mixtures[12] as known from the experimental data.[11] Here, the specific LJ interaction parameters are parameterized to fit the all–atom data.[12] While these parameters are listed in the Table S1 (Supporting Information), we only sketch its key ingredients herein.

2.2.1. Chain in Pure Solvent

We start with a good solvent case of ELP in pure solvent. This requires accurate modeling of the relative monomer–monomer, monomer–solvent, and solvent–solvent interactions. Here, one of the key aspects is that how the solvent interaction with a particular monomer (amino acid) type influences the solvent structure of the other neighboring monomers (with different amino acid types). Based on this consideration, the relative monomer–solvent interactions for different monomer species are parameterized so that it mimics the correct relative solvent coordination around a particular monomer type (water coordination around an amino acid) at $T = 300$ K.

Note also that a –[VPGVG]– sequence has $T_\ell \simeq 305$ K and thus remains in a coil state at $T = 300$ K. Furthermore, while our model mimics the good solvent condition of ELPs in pure solvent, it is not tuned to reproduce the coil–to–globule transition upon increasing T. This is particularly because when $T < T_\ell$, a coil configuration in pure solvent is stabilized by the solvent (or water) caging around a chain (or ELP). When T increases above T_ℓ, a certain number of solvent–monomer contacts breaks and thus also breaks the caging around a chain. Such a caging is due to complex multi–body balancing forces and the dipole orientations of solvent molecules that dictates the solvent packing within the first solvation shell.[25] Incorporating such multi–body effects within a two–body interaction model is a non–trivial task. Moreover, T dependence can be indirectly incorporate via an effective Flory–Huggins (FH) $\chi(T)$ parameter.[23] A recent study has also incorporated the dipole orientation within a FH–based lattice model via a Potts–like spin description.[42]

2.2.2. Chain in Binary Mixtures

We use an ELP model in binary solvent that is inspired by the earlier proposed generic microscopic picture of CNS by one of us with senior collaborators.[30,38] This picture suggests that the preferential monomer–cosolvent interaction drives the CNS transition. More specifically, when a small amount of cosolvent molecules are added in a good solvent polymer–solvent mixture, they preferentially bind to more than one monomer topologically far from one another along a chain contour, i.e., to reduce their binding free–energy. This process initiates the coil–to–globule transition in a polymer for lower x_c. When x_c increases above a critical value, such that there are sufficiently large number of cosolvent molecules that can decorate the full polymer chain, the same chain again goes from a re–entrant globule–to–coil transition.

A direct consequence of such a CNS picture is that the solvent quality always remains good and that the cosolvent–driven polymer collapse happens under the good solvent condition. Something that speaks in this favor is that the solvation free–energy of a polymer ΔG_m monotonically decreases with increasing x_c, i.e., the solvent quality becomes better and better with increasing x_c.[38] This observed solvation thermodynamics is completely decoupled from the conformation of a chain with changing x_c, as otherwise expected from the FH–like mean–field description.[30,43] ΔG_m between the pure solvent and the pure cosolvent phases vary between 3–5k_BT for different amino acids.[12,44]

The relative solvent–monomer and cosolvent–monomer ϵ_{ij} are tuned that reproduces the shift in ΔG_m with x_c, known from the all–atom simulations, is reproduced within the generic model.[12] The complete list of these parameters are highlighted in the Table S1 (Supporting Information).

2.3. Simulation Details

2.3.1. Generic Simulations

The equations of motion are integrated using the velocity Verlet algorithm with a time step $\Delta t = 2 \times 10^{-4}\tau$ and The temperature is set to $T = 1\epsilon/k_B$ using a Langevin thermostat with a damping coefficient of $\gamma = 1\tau^{-1}$. Individual systems are first equilibrated in a constant pressure p ensemble for 5×10^6 steps, with $p = 32\epsilon/\sigma^2$ that gives $\rho \simeq 5.5\sigma^{-3}$ for the pure solvent. The production runs are performed under the canonical ensemble for 10^8 MD steps with $\Delta t = 2 \times 10^{-4}\tau$. This choice is over two orders of magnitude longer than the typical relaxation time for $N_\ell = 50$. During the production runs observables, such as the single chain gyration radius R_g, the single chain structure factor $S(k)$, and the coordination numbers n_i, are calculated. Individual data sets are average of twenty independent runs and the error bars are the standard deviation.

© 2024 The Author(s). Macromolecular Rapid Communications published by Wiley-VCH GmbH

15213927, 2024, 18, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/marc.202400304 by MPI 355 Polymer Research, Wiley Online Library on [02/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

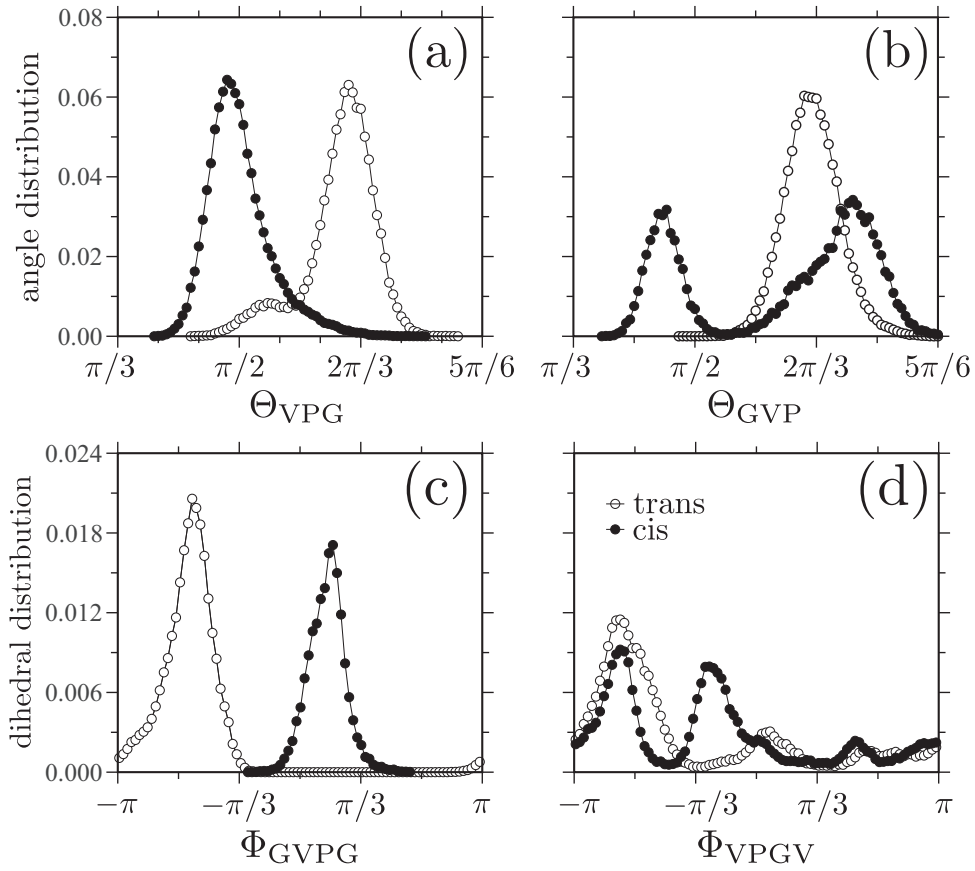

Figure 2. Distributions of angles (parts a–b) and dihedrals (parts c–d) showing two each most distinct effects with proline isomerizations between a *trans* and a *cis* configurations. The complete sets of distributions are shown in the Figures S1 and S2 (Supporting Information).

2.3.2. All–Atom Simulations

We have also performed a set of all–atom simulations to obtain the angular $\mathcal{D}\left(\Theta_{ijk}\right)$ and the dihedral $\mathcal{D}\left(\Phi_{ijkl}\right)$ distributions to distinguish between a *trans* and a *cis* configurations. For this purpose, an all–atom ELP chain with $n = 30$ is simulated in vacuum using the GROMACS molecular dynamics package.[45] The amino acids were modelled using the CHARMM36m force field.[46] The velocity rescaling thermostat[47] is used to impose $T = 300$ K with a coupling constant $\Delta t_{\mathrm{T}} = 1$ ps. The bond vibrations are constrained using the LINCS algorithm.[48] The electrostatic interactions are treated using the particle mesh Ewald (PME) algorithm.[49] The cutoff of the electrostatic and the van der Waals interactions were set to 1.4 nm. The equations of motion were integrated using the leap-frog integrator with a time step of $\Delta t = 2$ fs. The angular $\mathcal{D}\left(\Theta_{ijk}\right)$ and the dihedral $\mathcal{D}\left(\Phi_{ijkl}\right)$ distributions are calculated for from the all–atom trajectories for 1 μs.

3. Results and Discussion

3.1. Proline Isomerization

We begin by discussing the generic model of ELP with ProI using the mapping scheme shown in Figure 1. One of the key differences between a *trans* and a *cis* configurations is the relative geometric arrangement of proline with its neighboring amino acids, especially those within a three or a four amino acid segment along an ELP chain. The all–atom data from for $\mathcal{D}\left(\Theta_{ijk}\right)$ and $\mathcal{D}\left(\Phi_{ijkl}\right)$ for two most dominant cases are shown in **Figure 2**. The full set of distributions are shown in the Figures S1a–e and S2a–e (Supporting Information). The angular $u(\Theta_{ijk})$ and dihedral $u(\Phi_{ijkl})$ interactions are derived by using the Boltzmann inversion of $\mathcal{D}\left(\Theta_{ijk}\right)$,[50]

$$u(\Theta_{ijk}) = -k_{\mathrm{B}}T \ln\left[\mathcal{D}\left(\Theta_{ijk}\right)\right] + C_{\Theta} \tag{1}$$

and $\mathcal{D}\left(\Phi_{ijkl}\right)$,

$$u(\Phi_{ijkl}) = -k_{\mathrm{B}}T \ln\left[\mathcal{D}\left(\Phi_{ijkl}\right)\right] + C_{\Phi} \tag{2}$$

respectively. The resultant data are shown in the Figures S1f–j and S2f–j (Supporting Information).

ProI is incorporated within the generic model via the tabulated angular (3–body between $i - j - k$ monomers) and dihedral (4–body between $i - j - k - l$ monomers) interactions. Here, $u(\Theta_{ijk})$ and $u(\Phi_{ijkl})$, shown in the Figures S1–S2 (Supporting Information), are used between different generic monomer combinations. Using this model we first investigate ELP conformation in pure solvent (see Section 2.2.1) with both *trans* and *cis* states. In

 © 2024 The Author(s). Macromolecular Rapid Communications published by Wiley-VCH GmbH

15213927, 2024, 18, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/marc.202400304 by MPI 355 Polymer Research, Wiley Online Library on [02/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

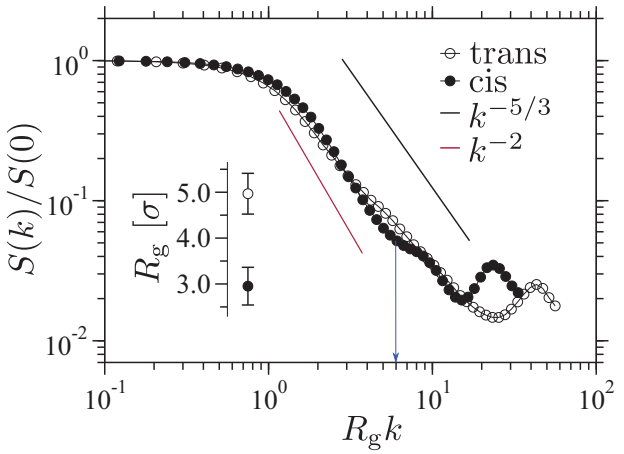

Figure 3. Single chain structure factor $S(k)$ comparing an all *trans* and an all *cis* configurations. $S(k)$ is normalized with the chain length $S(0) = N_\ell = 50$. The vertical down arrow shows an effective blob–size. The corresponding gyration radii R_g are shown in the inset. Error bars are the standard deviation calculated from twenty independent runs.

Figure 3, we show $S(k)$ for two isomerizations. It can be appreciated that a *trans* chain shows a clear good solvent scaling (see the ○ data set), i.e., $k^{-5/3}$.[14] However, a *cis* chain (see the • data set) shows a weak signature of cross–over around $R_g k \simeq 6.0$ (or $k \simeq 2.0\sigma^{-1}$). More specifically, while a *cis* chain remains in a random walk configuration (given by k^{-2}), it consists of small geometric segments of typical size $\ell_{blob} \simeq 3\sigma$, i.e., typically consisting of three monomers, with a central proline, arranged in $\pi/2$ angular orientation, see Figure 2a. This specific local geometry is a key for the solvent packing, as will be discussed in Section 3.3.

The $S(k)$ behavior is also supported by R_g shown in the inset of Figure 3. It can be seen that $R_g \simeq 3.0\sigma$ for a *cis* chain. This value is somewhat larger than the typical R_g of a collapsed chain with $N_\ell = 50$, i.e., $R_g \simeq 2.0 - 2.5\sigma$.[12]

We note in passing that this structural behavior is mostly consistent with the all–atom data in ref. [37]. However, a slight difference between the all–atom data in ref. [37] and our present generic data is because the all–atom simulation used $N_\ell = 150$, where some degree of secondary structure was also observed. For the shorter chains, as in this study, ELPs always remain in their disordered configurations akin of intrinsically disordered proteins.

3.2. Effect of Cosolvent and Proline Isomerization

We will now use our ELP model with ProI to investigate their conformations in solvent–cosolvent mixtures, using the non–bonded parameters listed in the Table S1 (Supporting Information). Here, $r_c = 2.5\sigma_{ij}$ between different particle pairs with $\sigma_{ij} = (\sigma_i + \sigma_j)/2$. We have investigated four different chain configurations, namely, an all *trans*, an all *cis*, an alternating copolymer of *trans* and *cis* segments, and a diblock of half *trans* and half *cis* segments. In **Figure 4**, we show the representative R_g as a function of x_c. For comparison, we have also included previously published all–atom data of ELP in aqueous ethanol mixtures.[51] As expected, the generic feature of the coil–to–globule–to–coil

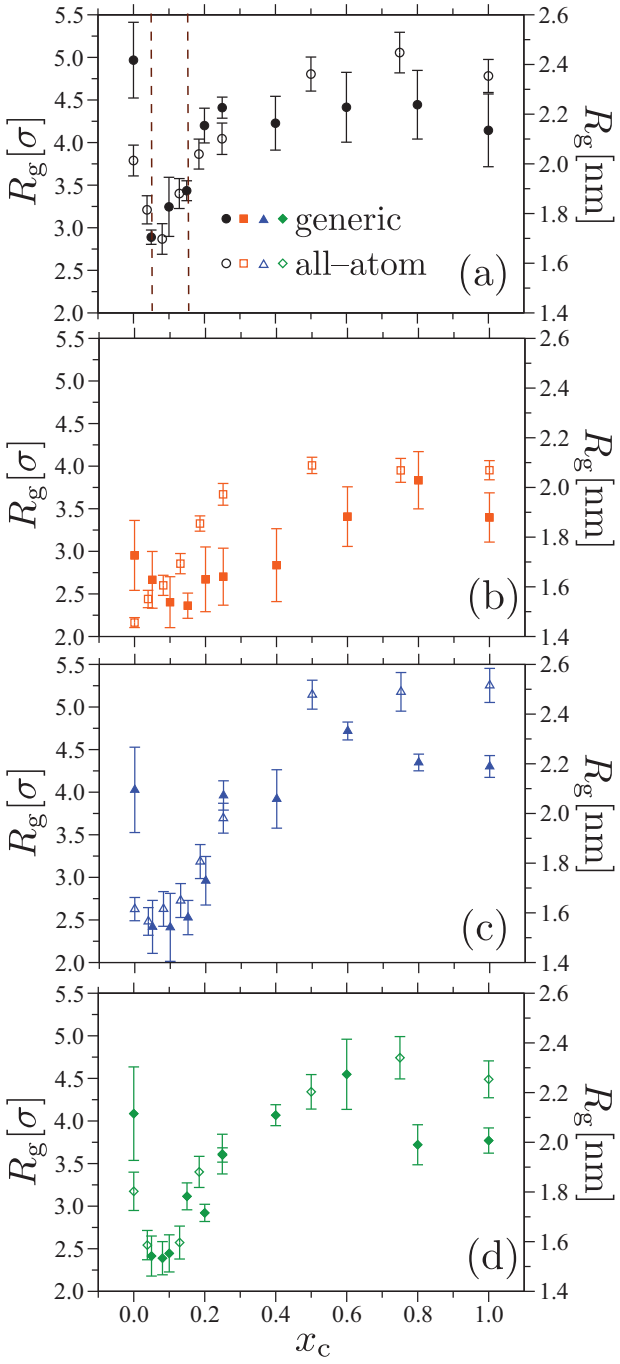

Figure 4. Radius of gyration R_g as a function of cosolvent mole fractions x_c. The data is shown for an all *trans* (part a), an all *cis* (part b), an alternating heteropolymer consisting of alternating *trans* and *cis* residues (part c), and a diblock consisting of half *trans* and half *cis* (part d) configurations. For comparison purpose, we have also included previously published all–atom data.[51] The two vertical dashed lines in part (a) highlight the window of collapse observed in experiment for $T = 300$ K.[11] Note the generic data is represented by left y −axis and the all–atom data is shown by the right y −axis.

© 2024 The Author(s). Macromolecular Rapid Communications published by Wiley-VCH GmbH

15213927, 2024, 18, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/marc.202400304 by MPI 355 Polymer Research, Wiley Online Library on [02/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

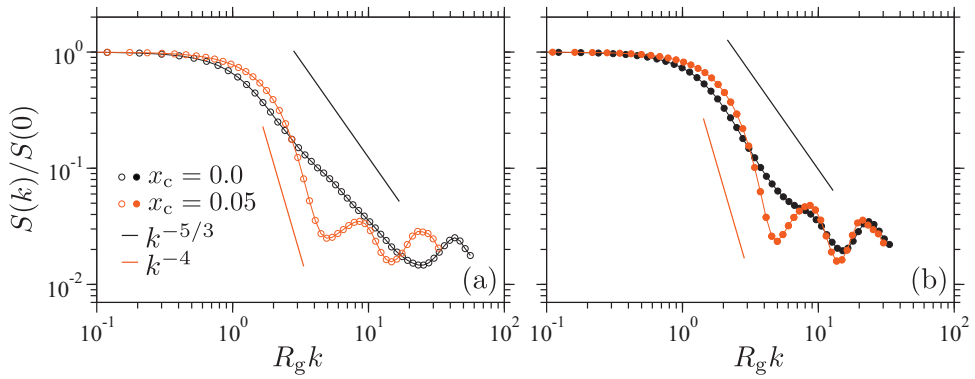

Figure 5. Single chain structure factor $S(k)$ comparing an all *trans* (part a)) and an all *cis* (part b) configurations. $S(k)$ is normalized with the chain length $S(0) = N_\ell$. The data is shown for two different cosolvent concentrations x_c, i.e., an expanded configuration at $x_c = 0.0$ and collapse under the influence of cosolvent $x_c = 0.05$.

transition (akin of CNS) is quite evident in Figure 4. It can also be seen that our simple generic model reasonably captures the window of ELP collapse observed in the all–atom simulations,[37] especially for an all *trans* and a diblock configurations, see Figure 4a,d. Furthermore, the window of collapse in an all *trans* case is also consistent with the experimental data,[11] shown by the vertical lines in Figure 4a.

The data in Figure 4 also reveal that the initial drop in R_g between $x_c \simeq 0.0 -$ to $- 0.1$ is most prominent in the generic model, see the solid symbols in Figure 4. On the contrary, the all–atom configurations in pure solvent (i.e., $x_c = 0.0$) are relatively more compact. There are two possible reasons for such a compact structure in the all–atom simulations: a) an expanded configuration of in pure solvent is stabilized by the caging of solvent molecules around the first solvation shell of a chain. Building caging is a rather non–trivial task within the standard all–atom setups, where a delicate competition between the microscopic interactions and the fluctuations play a key role. On the contrary, a generic model helps in fast equilibration of the solvent structure and chain conformation. b) The all–atom reference simulations were performed for a reasonably long chain of $N_\ell = 150$ that usually also have a tendency to form a certain degree of secondary structures, leading to somewhat shrinkage of an ELP chain.[37] Here, we note in passing that a stable secondary structure in mixed solvents is attained not only by the specific local interactions, but also due to the dipole orientations between the residue and solvent molecules.[52] The later effect is beyond the scope of our present study, where we are interested in the sorter sequences that can be classified as the intrinsically disordered poly–peptides.

Comparing the data in Figure 4a–b, we observe that R_g of a maximally collapsed ELP in an all *cis* case is about 25% smaller than an all *trans* configurations. This is not surprising because of the rather compact local configuration of a *cis* isomer, see Figure 1. These two configurations show the same collapsed structures, akin of the sphere scattering scaling q^{-4}, see **Figure 5**. This further consolidates the fact that the cosolvent molecules have a greater effect on an ELP (or "smart" polymers in general) than ProI in the complex mixtures.

The effect of ProI has a rather subtle effect on the copolymer configurations in Figure 4c–d. While their collapsed $R_g(x_c \leqslant 0.4)$, an alternating copolymer is slightly more expanded than a diblock copolymer. The later is because the *cis* half of a diblock is certainly more compact, as also seen in Figure 4b.

3.3. Local Geometry and Interactions

So far we have discussed the conformation of a generic chain in single and binary mixtures. However, a key advantage of such a minimalistic generic model is that the microscopic details are easily accessible, which are otherwise rather difficult with the standard all–atom setups because of their complex nature of the interactions. Therefore, in this section we will now discuss a distinction between the local geometric alterations via ProI and the monomer–(co–)solvent interactions in dictating an ELP conformation. For this purpose, we start by calculating the number of solvent n_s and cosolvent n_c molecules within the first solvation shell of a generic monomer representing proline, i.e., the number of solvent or cosolvent molecules directly in contact with it within a distance of $r \leqslant 1.0\sigma$. In **Figure 6**, we show the normalized n_i as a function of x_c. It can be appreciated that there is an excess of the cosolvent molecules (ethanol) around a generic proline monomer, see the hump in Figure 6b. This excess of the cosolvent molecules replace the solvent molecules within the first solvation shell and thus leads to solvent depletion, see the dip in Figure 6a. This behavior is consistent with the picture that the preferential cosolvent–monomer binding drives the CNS transition in "smart" polymers.[30,38]

Figure 6b also reveal that ProI does not have a noticeable effect on n_c, while n_s for a *trans* chain is about 10–15% larger than a *cis* chain, see Figure 6a. Here, it is important to mention that the interactions of generic proline monomer with solvent and cosolvent molecules are the same in the generic model. The observed noticeable difference in n_s is primarily due to the local geometrical arrangement due to ProI.

A closer look at the data in Figure 2a reveal that the V–P–G angle in *trans* state is about $2\pi/3$ that give enough geometrical (packing) space to accommodate a cosolvent (ethanol) and a solvent (water) molecules to in the interior region, see a schematic representation in **Figure 7a**. On the contrary, the same angle is about $\pi/2$ for a *cis* case, hence restricts the available space to fit

 © 2024 The Author(s). Macromolecular Rapid Communications published by Wiley-VCH GmbH

15213927, 2024, 18, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/marc.202400304 by MPI 355 Polymer Research, Wiley Online Library on [02/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

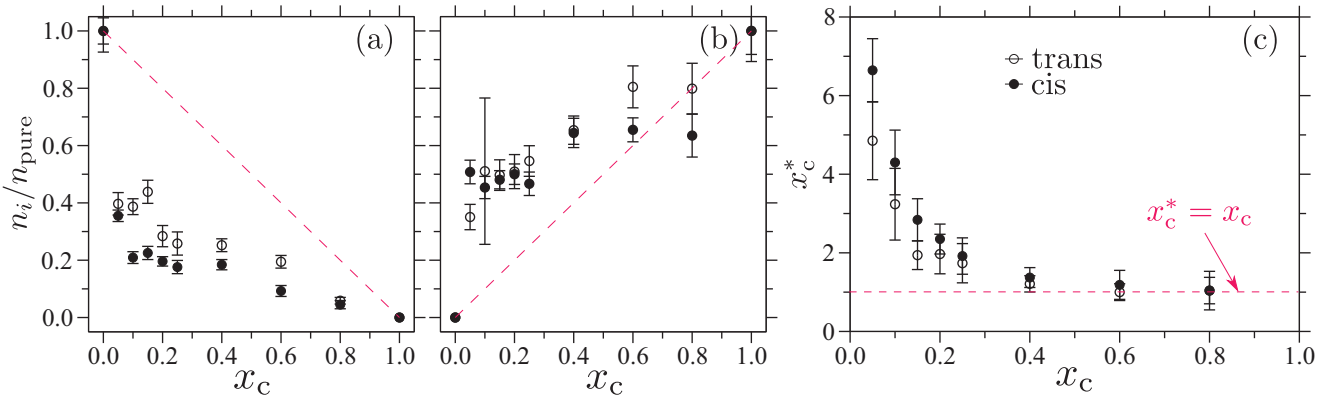

Figure 6. Parts(a–b) show the normalized coordination number n_i as a function of cosolvent (or ethanol in all–atom) mole fractions x_c. Part (a) shows the solvent coordination n_s, while the cosolvent coordination n_c is shown in part (b). Here, n_i is the number of solvent or cosolvent molecules directly in contact with a proline monomer within a distance $r \leqslant 1\sigma$. n_i is normalized with their respective coordination numbers in the pure phases, i.e., with $n_s(x_c = 0.0)$ in part (a) and $n_c(x_c = 1.0)$ in part (b). The data is shown for both proline isomerizations. The dashed lines are linear extrapolation between the pure solvent ($x_c = 0.0$) and the pure cosolvent ($x_c = 1.0$) phases. When n_i is larger than this linear extrapolated value, there is an excess. Otherwise there is a solvent depletion. Part (c) shows the excess coordination of the cosolvent molecules x_c^* as a function of x_c. x_c^* is calculated using Equation (3).

both molecules. As a result only a cosolvent molecule may fit because of their preferential binding, see Figure 7b. This simple discussion above further consolidate the fact that the generic model can clearly help distinguishing different factors responsible for the complex solvation properties.

To further validate this geometry versus interaction argument, we have also calculated the excess mole fraction of the cosolvent molecules using,

$$x_c^* = \frac{1}{x_c}\left(\frac{n_c}{n_c + n_s}\right) \tag{3}$$

In Figure 6c, x_c^* is shown. Consistent with the discussions above–(a) an excess is observed within the region where an ELP collapses and (b) x_c^* is more pronounced for the *cis* configuration because of the lower n_s. These observations are not surprising given that–

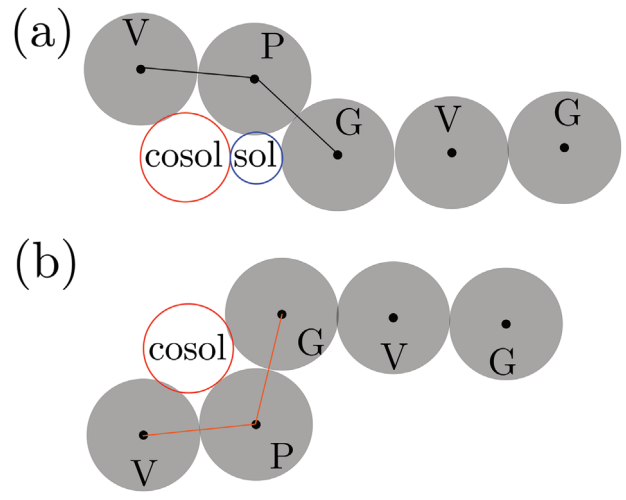

Figure 7. Schematics showing the solvent (sol) and cosolvent (cosol) packing in the interior of a –[VPG]– segment with *trans* (part a) and *cis* (part b) configurations.

when x_c is small, cosolvents aggregate near an ELP with an excess coordination within the first solvation shell.[17,30] When an ELP again goes from a globule–to–coil transition for $x_c \geqslant 0.4$, it gets completely decorated by the cosolvent molecules and thus gives $x_c^* \simeq x_c$. It is important to highlight that such a picture is not only applicable for the ELP solvation, rather is also observed within the context of "smart" polymer solvation in binary solvents.[17,30]

4. Conclusion

We have derived a simple generic model of elastin–like polypeptides (ELP) that incorporates proline isomerization (ProI). For this purpose, we have incorporated the angular and dihedral interactions known from the underlying all–atom systems into the generic model. Our generic model reasonably captures the conformational behavior of ELPs in a solvent and in the solvent–cosolvent mixtures consistent with the previous all–atom[51] and experimental[11] data. Going beyond the standard all–atom simulations, our simple generic model clearly reveals the microscopic distinction between the importance of local geometry that modifies the solvation structure around a segment involving proline and the monomer–level interactions.

The generic model proposed here in one hand can be used to study the single chain conformation in single and binary solvents. On the other hand, however, the application of such a minimalistic model might be to study the large scale self–assembly of ELPs and intrinsically disordered poly–peptide sequences in multi–component mixtures. In this context, we wish to highlight that the all–atom setups suffer from two main disadvantages– 1) Attaining equilibrium large scale assemblies of ELPs is severely hindered by the slow kinetics of the process. Complexities get even more elevated because of the delicate interplay between the system size effects, spacial heterogeneities, and unreasonable compositional fluctuations within the mid–sized all–atom simulation setups. 2) When an assembly is driven by the preferential monomer–cosolvent interactions, an excess of cosolvent around an assembly leads to depletion elsewhere within the simulation box.[38] This disturbs solvent equilibrium and thus leads

 © 2024 The Author(s). Macromolecular Rapid Communications published by Wiley-VCH GmbH

15213927, 2024, 18, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/marc.202400304 by MPI 355 Polymer Research, Wiley Online Library on [02/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

to wrong solvation effects. The later effect can be circumvented by using a semi–grand–canonical approach[38] where solvent or co-solvents are replaced within a simulation domain based on their chemical potential at a given x_c. Such a semi–grand–canonical move is usually impossible in a high density all–atom system, but typically is much easier in a coarse–grained system.[38] Therefore, the far reaching implication of our work may pave the way for operation understanding of complex assemblies and thus may lead to functional design of bio–compatible advanced materials.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

The content presented in this work has greatly benefited from a long standing continual collaboration with Kurt Kremer, whom the authors take this opportunity to gratefully acknowledge. Simulations in this work were performed at the Advanced Research Computing Sockeye allocation to D.M. and at the MPI–P computational facility of R.C.-H. For D.M. this research was undertaken thanks, in part, to the Canada First Research Excellence Fund (CFREF), Quantum Materials and Future Technologies Program.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

elastin-like polypeptide, generic model, molecular dynamics, polymer solvation, proline isomerization, structure-property relationship

Received: May 3, 2024
Revised: May 23, 2024
Published online: June 12, 2024

[1] D. E. Meyer, A. Chilkoti, *Nat. Biotechnol.* **1999**, *17*, 1112.
[2] D. E. Meyer, A. Chilkoti, *Biomacromolecules* **2002**, *3*, 357.
[3] D. T. McPherson, J. Xu, D. W. Urry, *Protein Expression Purif.* **1996**, *7*, 51.
[4] B. Zhao, N. K. Li, Y. G. Yingling, C. K. Hall, *Biomacromolecules* **2016**, *17*, 111.
[5] Y. Zhou, K. Jiang, Q. Song, S. Liu, *Langmuir* **2007**, *23*, 13076.
[6] E. G. Kelley, T. P. Smart, A. J. Jackson, M. O. Sullivan, T. H. Epps, *Soft Matter* **2011**, *7*, 7094.
[7] M. L. Adams, A. Lavasanifar, G. S. Kwon, *J. Pharm. Sci.* **2003**, *92*, 1343.
[8] A. Halperin, M. Kröger, F. M. Winnik, *Angew. Chem., Int. Ed.* **2015**, *54*, 15342.
[9] D. Mukherji, C. M. Marques, K. Kremer, *Annu. Rev. Condens. Matter Phys.* **2020**, *11*, 271.
[10] M. Müller, *Prog. Polym. Sci.* **2020**, *101*, 101198.
[11] C. E. Mills, E. Ding, B. D. Olsen, *Biomacromolecules* **2019**, *20*, 2167.
[12] Y. Zhao, M. K. Singh, K. Kremer, R. Cortes-Huerto, D. Mukherji, *Macromolecules* **2020**, *53*, 2101.
[13] Y. Zhao, K. Kremer, *J. Phys. Chem. B* **2021**, *125*, 9751.
[14] M. Doi, S. F. Edwards, *The Theory of Polymer Dynamics*, Oxford Science Publications, UK **1986**.
[15] C. Scherzinger, P. Lindner, M. Keerl, W. Richtering, *Macromolecules* **2010**, *43*, 6829.
[16] S. Micciulla, J. Michalowsky, M. A. Schroer, C. Holm, R. von Klitzing, J. Smiatek, *Phys. Chem. Chem. Phys.* **2016**, *18*, 5324.
[17] S. Backes, P. Krause, W. Tabaka, M. U. Witt, D. Mukherji, K. Kremer, R. von Klitzing, *ACS Macro Lett.* **2017**, *6*, 1042.
[18] A. S. Hoffman, P. S. Stayton, V. Bulmus, G. Chen, J. Chen, C. Cheung, A. Chilkoti, Z. Ding, L. Dong, R. Fong, C. A. Lackey, C. J. Long, M. Miura, J. E. Morris, N. Murthy, Y. Nabeshima, T. G. Park, O. W. Press, T. Shimoboji, S. Shoemaker, H. J. Yang, N. Monji, R. C. Nowinski, C. A. Cole, J. H. Priest, J. M. Harris, K. Nakamae, T. Nishino, T. Miyata, *J. Biomed. Mater. Res.* **2000**, *52*, 577.
[19] Z. Shen, K. Terao, Y. Maki, T. Dobashi, G. Ma, T. Yamamoto, *Colloid Polym. Sci.* **2006**, *284*, 1001.
[20] S. Samanta, D. R. Bogdanowicz, H. H. Lu, J. T. Koberstein, *Macromolecules* **2016**, *49*, 1858.
[21] T. E. de Oliveira, D. Mukherji, K. Kremer, P. A. Netz, *J. Chem. Phys.* **2017**, *146*, 034904.
[22] C. C. De Silva, P. Leophairatana, T. Ohkuma, J. T. Koberstein, K. Kremer, D. Mukherji, *J. Chem. Phys.* **2017**, *147*, 064904.
[23] A. Prhashanna, P. A. Taylor, J. Qin, K. L. Kiick, A. Jayaraman, *Biomacromolecules* **2019**, *20*, 1178.
[24] D. T. McPherson, J. Xu, D. W. Urry, *Protein Expr. Purif.* **1996**, *7*, 51.
[25] D. Mukherji, K. Kremer, *Polymers* **2023**, *15*, 3229.
[26] B. A. Wolf, G. Blaum, *J. Polym. Sci., Polym. Phys. Ed.* **1975**, *13*, 1115.
[27] F. M. Winnik, H. Ringsdorf, J. Venzmer, *Macromolecules* **1990**, *23*, 2415.
[28] H. G. Schild, M. Muthukumar, D. A. Tirrell, *Macromolecules* **1991**, *24*, 948.
[29] G. Zhang, C. Wu, *Phys. Rev. Lett.* **2001**, *86*, 822.
[30] D. Mukherji, C. M. Marques, K. Kremer, *Nat. Commun.* **2014**, *5*, 4882.
[31] D. Kern, M. Schutkowski, T. Drakenberg, *J. Am. Chem. Soc.* **1997**, *119*, 8403.
[32] A. Valiaev, D. W. Lim, T. G. Oas, A. Chilkoti, S. Zauscher, *J. Am. Chem. Soc.* **2007**, *129*, 6491.
[33] F. D. Jochum, P. Theato, *Macromolecules* **2009**, *42*, 5941.
[34] K. Y. Lam, C. S. Lee, M. R. Pichika, S. F. Cheng, R. Y. Hang Tan, *RSC Adv.* **2022**, *12*, 15261.
[35] J. Volarić, J. Buter, A. M. Schulte, K.-O. van den Berg, E. Santamaría-Aranda, W. Szymanski, B. L. Feringa, *J. Org. Chem.* **2022**, *87*, 14319.
[36] V. Walter, T. Schmatko, P. Muller, A. P. Schroder, S. R. MacEwan, A. Chilkoti, C. M. Marques, *Biophys. J.* **2024**, *123*, 901.
[37] Y. Zhao, K. Kremer, *J. Phys. Chem. B* **2021**, *125*, 9751.
[38] D. Mukherji, K. Kremer, *Macromolecules* **2013**, *46*, 9158.
[39] S. Plimpton, *J. Comput. Phys.* **1995**, *117*, 1.
[40] K. Kremer, G. S. Grest, *J. Chem. Phys.* **1990**, *92*, 5057.
[41] D. Mukherji, C. M. Marques, K. Kremer, *J. Phys.: Condens. Matter* **2017**, *30*, 024002.
[42] S. Dhamankar, M. A. Webb, Asymmetry in polymer-solvent interactions yields complex thermoresponsive behavior, *arXiv* **2024**, p. 2312.00940.
[43] D. Mukherji, C. M. Marques, T. Stuehn, K. Kremer, *J. Chem. Phys.* **2015**, *142*, 114903.
[44] The data in ref. [12] was reported for the central monomer of a tri–proline. Therefore, as shown earlier in ref. [38], going from a trimer to a polymer, the solvation volume around a monomer gets modified by a factor of about 2.

 © 2024 The Author(s). Macromolecular Rapid Communications published by Wiley-VCH GmbH

15213927, 2024, 18, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/marc.202400304 by MPI 355 Polymer Research, Wiley Online Library on [02/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[45] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, **2015**, *1*, 19.
[46] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmülle, A. D. MacKerell Jr, **2017**, *14*, 71.
[47] G. Bussi, D. Donadio, M. Parrinello, *J. Chem. Phys.* **2007**, *126*, 014101.
[48] B. Hess, H. Bekker, H. J. Berendsen, J. G. Fraaije, *J. Comp. Chem.* **1997**, *18*, 1463.
[49] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.* **1995**, *103*, 8577.
[50] W. Tschöp, K. Kremer, J. Batoulis, T. Bürger, O. Hahn, *Acta Polymerica* **1998**, *49*, 61.
[51] Y. Zhao, K. Kremer, *Macromol. Rapid Commun.* **2022**, *43*, 2100907.
[52] L. A. Baptista, Y. Zhao, K. Kremer, D. Mukherji, R. Cortes-Huerto, *ACS Macro Lett.* **2023**, *12*, 841.

 © 2024 The Author(s). Macromolecular Rapid Communications published by Wiley-VCH GmbH

15213927, 2024, 18, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/marc.202400304 by MPI 355 Polymer Research, Wiley Online Library on [02/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License