Loop–Closure and Gaussian Models of Collective Structural Characteristics of Capped PEO Oligomers in Water

3 AUTHORS:

Mangesh Chaudhari
Sandia National Laboratories
19 PUBLICATIONS **17** CITATIONS

Lawrence R. Pratt
Tulane University
203 PUBLICATIONS **9,038** CITATIONS

Michael Paulaitis
Johns Hopkins Medicine
165 PUBLICATIONS **4,722** CITATIONS

Loop-Closure and Gaussian Models of Collective Structural Characteristics of Capped PEO Oligomers in Water

M. I. Chaudhari, L. R. Pratt[1] and M. E. Paulaitis[2]

[1] *Department of Chemical and Biomolecular Engineering, Tulane University, New Orleans, LA 70118*

[2] *Department of Chemical and Biomolecular Engineering, The Ohio State University, Columbus, OH 43210*

(Dated: 14 January 2015)

Parallel-tempering MD results for a $CH_3(CH_2\text{-}O\text{-}CH_2)_m CH_3$ chain in water are exploited as a data-base for analysis of collective structural characteristics of the PEO globule with a goal of defining models permitting statistical thermodynamic analysis of dispersants of Corexit type. The chain structure factor, relevant to neutron scattering from a deuterated chain in neutral water, is considered specifically. The traditional continuum-Gaussian structure factor is inconsistent with the simple $k \to \infty$ behavior, but we consider a discrete-Gaussian model that does achieve that consistency. Shifting-and-scaling the discrete-Gaussian model helps to identify the low-k to high-k transition near $k \approx 2\pi/0.6$ nm when an empirically matched number of Gaussian links is about one-third of the total number of effective-atom sites. This short distance-scale boundary of 0.6 nm is directly verified with the r-space distributions, and this distance is thus identified with a natural size for coarsened monomers. The probability distribution of $R_g{}^2$ is compared with the classic predictions for both Gaussian model and freely-jointed chains. $\langle R_g{}^2(j) \rangle$, the contribution of the j-th chain segment to $\langle R_g{}^2 \rangle$, depends on contour index about as expected for Gaussian chains despite significant quantitative discrepancies which express the swelling of these chains in water. Monomers central to the chain contour occupy the center of the chain globule. The density profiles of chain segments relative to their center of mass can show distinctive density structuring for smaller chains due close proximity of central elements to the globule center. But that density structuring washes-out for longer chains where many chain elements additively contribute to the density profiles. Gaussian chain models thus become more satisfactory for the density profiles for longer chains.

I. INTRODUCTION

Arguably the most important water-soluble synthetic polymers,[1,2] (-CH_2-O-CH_2-)$_n$ chain molecules are intrinsic to the dispersant materials applied to oil spills,[3] and can be soluble also in organic solvents. With -H (and thus methyl -CH_3) ends polyethylene oxide (PEO) is a common name, and we will use that name generically when the chains have arbitrary capping groups. With hydroxyl -OH terminations these chains are typically called polyethylene glycol (PEG), and we will use that name in discussing experiments that study that case specifically.

For dispersants used on oil spills,[3] and for other applications,[4] PEO chains are often decorated with junctions or tails or caps. Correlations associated with capping groups then focus molecular structural analyses, molecular specificity in understanding loop closure being an outstanding interest.[5] Those correlations can be the targets of neutron scattering experiments[6] with deuterated chains, under the important limitation of chemical feasibility of specific isotopic labeling for the caps. We have studied X(-CH_2-O-CH_2-)$_n$X with a variety of capping groups X for just those reasons.[7] Capping groups can non-trivially change solution properties, particularly for the short-chain oligomers, and those changes have been of specific interest.[8]

Here we analyze the X = CH_3 case. With this capping choice, neutron scattering experiments also characterize C-C contacts for C atoms closer along the chain contour than the end-caps, *i.e.,* short-circuited loops contribute to those neutron scattering results. We utilize simulation results theoretically to investigate anticipated neutron scattering studies that interrogate loop-closure. We consider a structure factor model that spans small-angle and diffraction regimes, which thus highlights the molecular-scale features that might be sought.

The structure factor model that we analyze assumes ideal Gaussian chains, and is particularly simple. Nevertheless, it is more complicated than the most traditional continuum-Gaussian model, and the distinction is essential for the success of the model in these comparisons.

We then further test other aspects of Gaussian chain models against simulation results. Since Gaussian chain models are particularly simple, these alternative aspects are typically collective characteristics. Though limited in molecular detail, collective characteristics have countervailing advantages of wide utility. Initial examples include the probability distribution, $P(R_g{}^2)$, of the square radius of gyration, and then the $\langle R_g{}^2(j) \rangle$ for the j-th chain segment, contributing to the decomposition

$$\langle R_g{}^2 \rangle = \frac{1}{n+1} \sum_{j=0}^{n} \langle R_g{}^2(j) \rangle \ . \qquad (1.1)$$

Building from $\langle R_g{}^2(j) \rangle$ results, we consider density profiles of chain globules described by a Gaussian chain model. Our goal is to establish simple models that permit statistical thermodynamic evaluation of the surface

tensions of aqueous electrolyte solutions with hydrocarbon liquids[9–12] when dispersant materials are deployed. As an example of a specific characteristic that should be helpful, we note that the dielectric constant of aqueous PEO solutions depends linearly on the water volume fraction.[7,13]

These systems are notorious for physical complexity despite their chemical simplicity.[14] But a broad physical description of these solutions is that water is a good solvent for PEO chains, which are swollen by the solvent. Our results for the osmotic second virial coefficients for $CH_3(CH_2\text{-}O\text{-}CH_2)_{11}CH_3$, $B_2 > 0$ obtained elsewhere from multi-chain solution simulations,[7] indicates repulsive inter-segment interactions at ambient (T, p) conditions. For an experimental perspective on PEG osmotic pressures, see Cohen, et al.[15,16] Consistent with repulsive inter-segment interactions, we find $\langle R_g{}^2 \rangle \propto m^{1.3}$. Of course, that exponent was not the goal of the calculations implemented, which are detailed below.

Nevertheless, the solution environment can sensitively affects PEG conformations.[1,2] PEG molecules are helical in n-propanoic, isobutyric, and isopentanoic acid solutions coexisting with liquid water,[2] with helix formation requiring a trace of water.[2] In contrast, these chains are generically coiled in aqueous solutions and also in acetic acid, isobutanol, and n-butanol. Conformational sensitivity is associated with size fractionation of PEGs between water and isobutyric acid.[1,2]

The versatility of PEG polymers solutions makes them a challenge for molecular thermodynamics. Flory-Huggins interaction parameters, experimentally evaluated, show substantial but different *composition* dependences for PEG in water and methanol,[17,18] but in ethanol only minor dependence on composition.[19]

The results below extend the aqueous solution calculations discussed in a preliminary report that compared n-hexane solvent with water.[20] Previous simulation calculations[13,21–25] evaluate different aspects of these solutions, and give a helpful baseline on which the present modeling builds.

II. METHODS

The simulation calculations below treated a single $CH_3(CH_2\text{-}O\text{-}CH_2)_mCH_3$ molecule in water by molecular dynamics, using parallel tempering[26] to achieve enhanced sampling of chain conformations. We evaluated system sizes of $N_{\text{water}} = 1000$ ($m = 11$), and $N_{\text{water}} = 2000$ ($m = 21, 31$). The chain molecules were represented by optimized potentials for liquid simulations (OPLS-AA),[27] and the SPC/E model was adopted for water[28] implemented with the GROMACS 4.5.3 molecular dynamics simulation package.[29] Long-range electrostatic interactions were treated in standard periodic boundary conditions using particle mesh Ewald with a cutoff of 0.9 nm. The Nosé-Hoover thermostat maintained the temperature and hydrogen atom bond-lengths were constrained

by the LINCS algorithm. After energy minimization, density equilibration with $(T = 300K, p = 1 \text{ atm})$ MD calculations established the constant volumes for each parallel tempering simulation. The parallel tempering spanned the 256-550K temperature range with 32 replicas (for $m = 11$, and 21 cases) and 40 replicas (for $m = 31$). Parallel tempering swaps were attempted at a rate of 100/ns, and the temperature grid resulted in a success rates of 15-25%. Production calculations for each replica set were extended to 10 ns.
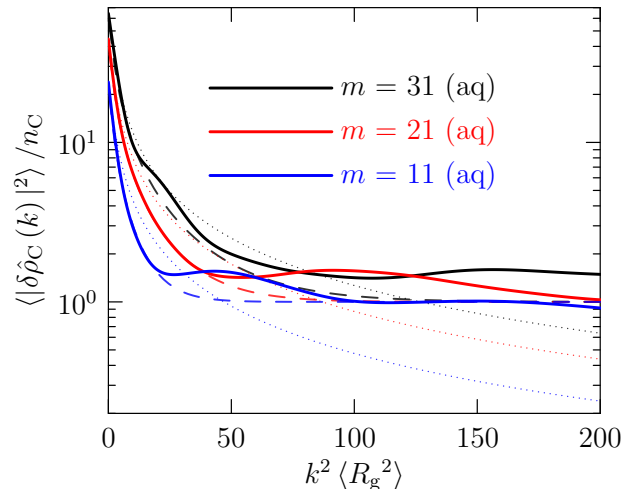

FIG. 1. The traditional continuum-Gaussian structure factor (Eq. (3.5), dotted curves) contrasted with a discrete-Gaussian model, shifted and scaled, which is consistent with the correct $k \to \infty$ limit (Eq. (3.7), dashed curves). The direct numerical simulation results are the solid curves. For the model Eq. (3.7), the number of Gaussian links, n, was adjusted for agreement throughout a low-k regime. In all cases here n was about a third of the number of heavy atoms of $CH_3(CH_2\text{-}O\text{-}CH_2)_mCH_3$, *i.e.*, in coarse-grained models of these oligomers, coarsened monomers can represent about three effective-atoms for this characteristic. The local maxima for $k > 0$ here correspond to $k \gtrsim 2\pi/0.6$nm.

III. RESULTS AND DISCUSSION

The structure factor

$$\langle |\delta\hat{\rho}_C(k)|^2 \rangle / n_C = 1 + \int \left(\frac{\sin kr}{kr} \right) \langle \rho_C(r)|0\rangle \, d\mathbf{r} , \quad (3.2)$$

addresses CC loop-closure contacts comprehensively, in contrast to chain-end closure exclusively which would be targeted by labelled ends.[6] Here n_C is the number of C-atoms in the solution, and $\langle \rho_C(r)|0\rangle$ is the density, conditional on placement of a C-atom at the origin, of other C-atoms at radius r. Since our calculations here treat only one chain molecule, $\langle \rho_C(r)|0\rangle$ is the density of other, *intra*molecular C-atoms, and is normalized to

one less than the number of C-atoms in a solute chain; in our case

$$\langle |\delta\hat{\rho}_{\mathrm{C}}(0)|^2 \rangle / n_{\mathrm{C}} = 2(m+1) \ . \qquad (3.3)$$

Inverse to Eq. (3.2) is

$$\left(\frac{1}{2\pi}\right)^3 \int \left(\frac{\sin kr}{kr}\right) \left(\langle |\delta\hat{\rho}_{\mathrm{C}}(k)|^2 \rangle / n_{\mathrm{C}}\right) \mathrm{d}\mathbf{k} =$$
$$\delta(\mathbf{r}) + \langle \rho_{\mathrm{C}}(r) |0 \rangle \ . \quad (3.4)$$

The simple result

$$\langle |\delta\hat{\rho}_{\mathrm{C}}(k)|^2 \rangle / n_{\mathrm{C}} = (2m+1)$$
$$\times \left[\exp\left(-k^2 \langle R_g{}^2 \rangle\right) - 1 + k^2 \langle R_g{}^2 \rangle \right]$$
$$\times \frac{2}{(k^2 \langle R_g{}^2 \rangle)^2} \ . \quad (3.5)$$

is the continuum-Gaussian model that we consider.[30] This satisfies the anticipated low-k behavior, but not the $k \to \infty$ value associated with the $\delta(\mathbf{r})$ function of Eq. (3.4). The result for a discrete-Gaussian chain is

$$\left[\langle |\delta\hat{\rho}_{\mathrm{C}}(k)|^2 \rangle / n_{\mathrm{C}} \right]_{\mathrm{DG}} =$$
$$\left[\mathrm{e}^{2\zeta/n}(n+1) - 2\left(\mathrm{e}^{\zeta/n} - \mathrm{e}^{-\zeta}\right) - (n+1) \right]$$
$$/ \left[(n+1)\left(\mathrm{e}^{\zeta/n} - 1\right)^2 \right] \ , \quad (3.6)$$

where $\zeta = k^2 \langle R_g{}^2 \rangle$, and n is the number of *Gaussian links*. This has the expected $\zeta \to 0$ limit, namely, $(n+1)$ (Eq. (3.3)). Then $n \to \infty$, with ζ fixed, leads to Eq. (3.5) and clarifies "continuum" here.

We shift and scale the discrete-Gaussian model,

$$\langle |\delta\hat{\rho}_{\mathrm{C}}(k)|^2 \rangle / n_{\mathrm{C}} - 1 =$$
$$\left(\frac{2m+1}{n}\right) \left\{ \left[\langle |\delta\hat{\rho}_{\mathrm{C}}(k)|^2 \rangle / n_{\mathrm{C}} \right]_{\mathrm{DG}} - 1 \right\} \ , \quad (3.7)$$

to compare with the simulation data (FIG. 1). The discrete-Gaussian model matches the data through the low-k regime more effectively than does the continuum model Eq. (3.5). Beyond that low-k regime, the data deviate from the discrete-Gaussian model positively through a local maximum indicating a short length scale $2\pi/k_{\max} \lesssim 0.6$ nm.

We can directly turn to $\langle \rho_{\mathrm{C}}(r)|0 \rangle$ for confirmation of this inference (FIG. 2). Indeed, the short-distance regime begins with $r < 0.6$ nm, consistent with identification of high-angles for $k > 2\pi/0.6$ nm (FIG. 1).

In addition to the assumption of ideal behavior for the n coarsened monomers, the model tested above obviously utilizes an empirical $\langle R_g{}^2 \rangle$. We next consider the distribution of $R_g{}^2$ for these chains, compared to Gaussian model results (FIG. 3). The distinctions deriving from molecular-scale resolution, cutoffs at minimum and maximum lengths, are prominent.
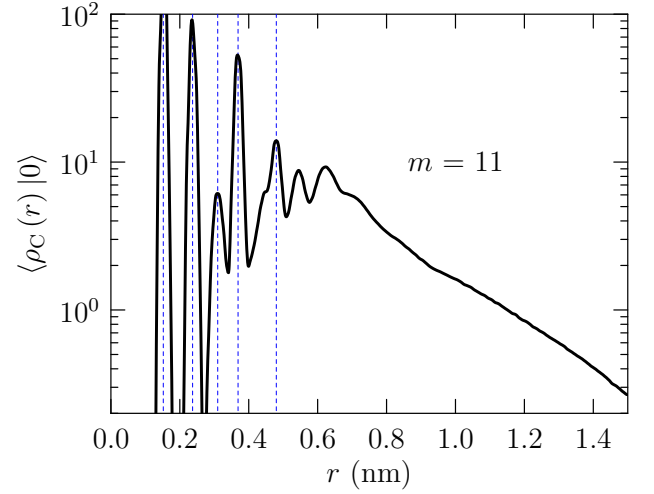

FIG. 2. Density, conditional on placement of a C-atom at the origin, of other C-atoms at radius r, see text. Consistent with FIG. 1, the short-distance regime begins with $r < 0.6$ nm (high-angles for $k > 2\pi/0.6$ nm). The dashed vertical lines indicated distances of specific interest. The $r \approx 0.38$ nm peak represents direct, non-bonded CC contacts that are of interest to investigations of hydrophobic interactions.[6]
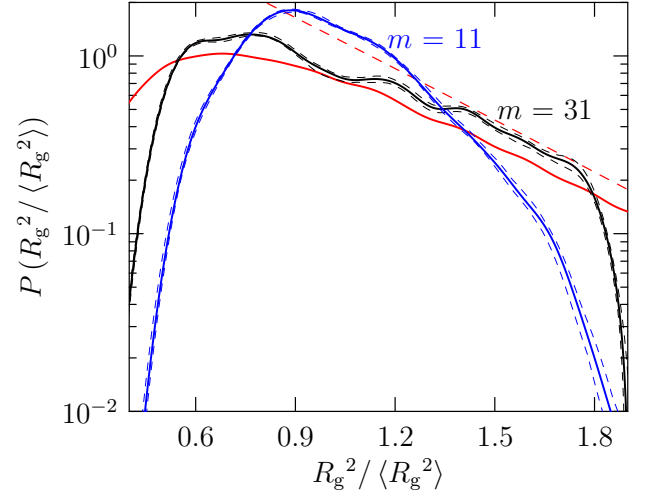

FIG. 3. The solid red curve is the result for the $m = 31$ case of a freely jointed chain, obtained numerically by a straightforward Monte Carlo calculation; the dashed red curve is the asymptotic $R_{\mathrm{g}}{}^2 / \langle R_{\mathrm{g}}{}^2 \rangle \sim \infty$ result for an ideal Gaussian model,[31] close to a simple Gaussian function. The dashed lines bracket the 95% confidence intervals approximated by a bootstrap method.

The comparison (FIG. 4) of the observed dependence of $\langle R_g{}^2(j) \rangle$ on contour index j with that for ideal Gaussian models shows encouraging agreement. On the other-hand, the discrepancies of ideal Gaussian behavior from the observed results are much larger than the difference of the ideal Gaussian model from the results for a freely
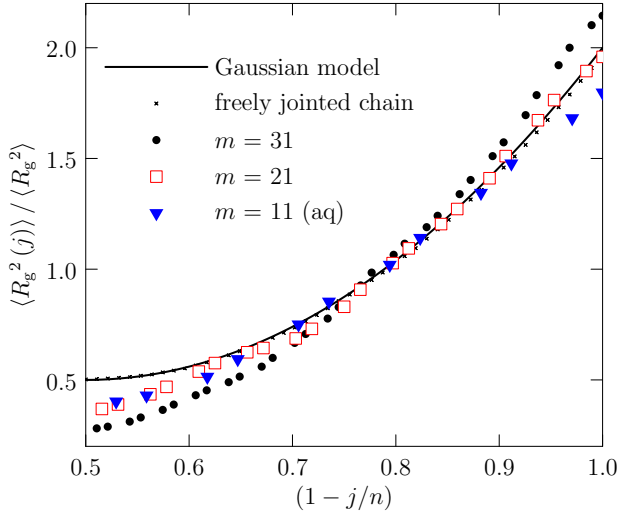
FIG. 4. The number of bonds along the heavy atom chain contour is indexed by n, *i.e.* $n = 34$ which has 35 heavy atoms for the $m = 11$ chain. The chain molecule length index m is defined on the basis of the molecular formula $CH_3(CH_2\text{-}O\text{-}CH_2)_mCH_3$. The solid curve is the function $2\left[1 - 3\left(j/n\right)\left(1 - j/n\right)\right]$ appropriate for an ideal Gaussian chain.[32] The crosses are the results for the $m = 31$ case of a freely jointed chain, obtained numerically by a straightforward Monte Carlo calculation. The right-most triangle corresponds to a methyl C atom, here index $j=34$; $j=0$ is chemically equivalent. The third triangle from the right boundary corresponds to $j=31$, chemically equivalent to $j = 3$.

jointed chain. Because Eq. (1.1) is a sum of positive contributions, the ratio plotted in FIG. 4 is normalized, and therefore the behavior seen in FIG. 4 expresses the swelling of these chains in water. Thus it is clear that the quantitative discrepancies are significant, though these characteristics offer minimal expression of molecular detail.

Identifying chemically distinct C atoms permits a layered display of density profiles (FIG. 5). C atoms near the center of the chain are likely in the interior of the chain droplet, and the end-atoms are more likely on the outside. For the $m = 11$ chains some molecular-scale structure is evident, but that is less prominent (FIG. 5) for $m = 21$ chains.

The natural comparative model for the density profiles,[32]

$$\rho_C\left(r\right) = \sum_{j=\text{C atoms}} \frac{e^{-3r^2/2\left\langle R_g{}^2(j)\right\rangle}}{\sqrt[3]{2\pi \left\langle R_g{}^2(j)\right\rangle /3}} , \qquad (3.8)$$

is obtained by superposing of the Gaussian distributions associated with the observed $\left\langle R_g{}^2(j)\right\rangle$. The density contributions of C atoms interior the chains are more structured than the overall density profile.
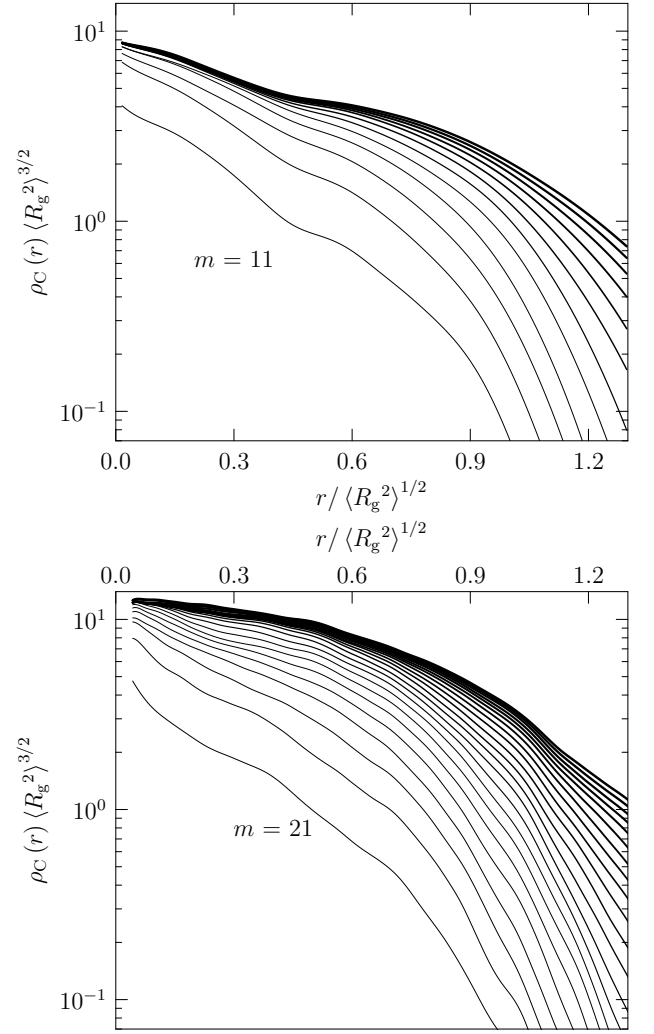

FIG. 5. C atom contributions to the chain density profiles from their center of mass, cumulatively indexed by the contour distance from the chain center. The lowest curve corresponds to the mid-pair C atoms, the next higher curve includes two additional C atoms, and the upper-most curve includes all C atoms.

IV. CONCLUSIONS

Chain structure factors, relevant to neutron scattering from a chain in neutral water, are compared in detail to a traditional continuum-Gaussian model result. The most serious limitation of the traditional continuum-Gaussian structure factor is the failure to match the trivially known $k \to \infty$ limiting value. A discrete-Gaussian model that is consistent with the correct $k \to \infty$ value is considered. Shifting-and-scaling the discrete-Gaussian model helps to identify the low-k to high-k transition near $k \approx 2\pi/0.6$ nm when an empirically matched number of Gaussian links is about one-third of the total number of effective-atom sites. The shifted-and-scaled discrete-Gaussian model better identifies the transition from low-
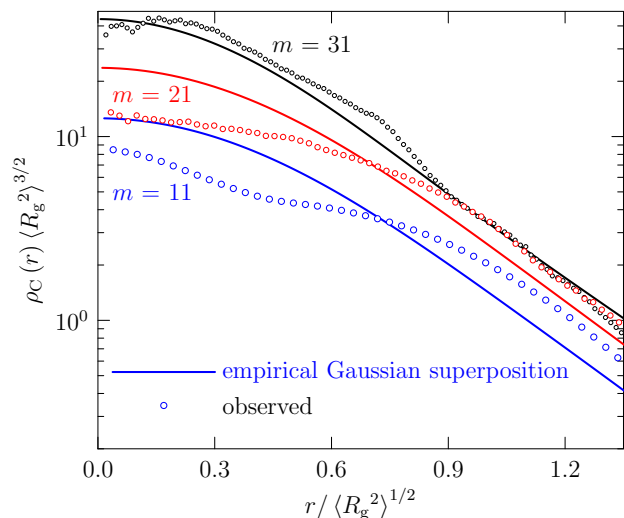
FIG. 6. Density profiles for C atoms, relative to their center of mass. The Gaussian superposition models (solid curves) follow from the discussion of Yamakawa,[32] using the empirical $\langle R_g{}^2(j) \rangle$ shown in FIG. 3.

k to high-k behavior near $k \approx 2\pi/0.6$nm, which thus provides a natural spatial size for the coarsened monomers. This short distance-scale boundary of 0.6 nm is directly verified with the r-space distributions.

Further testing of Gaussian chain models for these systems shows that $\langle R_g{}^2(j) \rangle$, the contribution of the j-th chain segment to $\langle R_g{}^2 \rangle$, depends on contour index about as expected for Gaussian chains despite quantitative discrepancies. The quantitative comparison expresses the swelling of these chains in water. Monomers central to the chain contour are usually central to the chain globule. The density profiles of chain molecule segments relative to their center of mass can show distinctive density structuring for smaller chains due close proximity of central elements to the globule center. That density structuring washes-out for longer chains, and due to the coarsened length-scale $\langle R_g{}^2 \rangle^{1/2}$, many chain elements then contributing additively to the density profiles. Gaussian chain models thus become more satisfactory for the density profiles for longer chains.

[1] Alessi, M. L.; Norman, A. I.; Knowlton, S. E.; Ho, D. L.; Greer, S. C. *Macromolecules* **2005**, *38*, 9333–9340.
[2] Norman, A. I.; Fei, Y.; Ho, D. L.; Greer, S. C. *Macromolecules* **2007**, *40*, 2559–2567.
[3] *Understanding Oil Spill Dispersants: Efficacy and Effects*; National Academies Press, Washington DC, 2005.
[4] Lin, Z.; Rubtsov, I. V. *Proc. Nat. Acad. Sci. USA* **2012**, *109*, 1413–1418.
[5] Weikl, T. R. *Arch. Biochem. Biophys.* **2008**, *469*, 67–75.
[6] Chaudhari, M. I.; Pratt, L. R.; Paulaitis, M. E. *J. Chem. Phys.* **2010**, *133*, 231102.
[7] Chaudhari, M. I. Molecular Simulations to Study Thermodynamics of Polyethylene Oxide Solutions. Ph.D. thesis, Department of Chemical & Biomolecular Engineering, Tulane University, 2013.
[8] Dormidontova, E. E. *Macromolecules* **2004**, *37*, 7747–7761.
[9] Nichols, A. L.; Pratt, L. R. *Faraday Symp. Chem. Soc.* **1982**, *17*, 129–140.
[10] Wilson, M. A.; Nichols III, A. L.; Pratt, L. R. *J. Chem. Phys.* **1984**, *81*, 579–580.
[11] Nichols III, A. L.; Pratt, L. R. *J. Chem Phys.* **1984**, *80*, 6225–6233.
[12] Pratt, L. R. *J. Phys. Chem.* **1992**, *96*, 25–33.
[13] Borodin, O.; Bedrov, D.; Smith, G. D. *Macromolecules* **2002**, *35*, 2410–2412.
[14] Israelachvili, J. *Proc. Nat.l Acad. Sci. USA* **1997**, *94*, 8378–8379.
[15] Cohen, J. A.; Podgornik, R.; Hansen, P. L.; Parsegian, V. A. *J. Phys. Chem. B* **2009**, *113*, 3709–3714.
[16] Cohen, J. A.; Podgornik, R.; Parsegian, V. A. *Biophys. J.* **2012**, *102*, 400A–400A.
[17] Bae, Y. C.; Shim, D. S., J. J .and Soane; Prausnitz, J. M. *J. Appl. Poly. Sci.* **1993**, *47*, 1193–1206.
[18] Zafarani-Moattar, M. T.; Tohidifar, N. *J. Chem. Eng. Data* **2006**, *51*, 1769–1774.
[19] Zafarani-Moattar, M. T.; Tohidifar, N. *J. Chem. Eng. Data* **2008**, *53*, 785–793.
[20] Chaudhari, M. I.; Pratt, L. R. In *OIL SPILL REMEDIATION: COLLOID CHEMISTRY-BASED PRINCIPLES AND SOLUTIONS*; Somasundaran, P., Farinato, R., Patra, P., Papadopoulos, K. D., Eds.; John Wiley and Sons, Inc., 2012; See also: arXiv:1208.0349v2.
[21] Borodin, O.; Bedrov, D.; Smith, G. D. *Macromolecules* **2001**, *34*, 5687–5693.
[22] Lee, H.; Venable, R. M.; MacKerell Jr, A. D.; Pastor, R. W. *Biophys. J.* **2008**, *95*, 1590–1599.
[23] Choi, E.; Mondal, J.; Yethiraj, A. *J. Phys. Chem. B* **2013**, 131218200604003.
[24] Mondal, J.; Choi, E.; Yethiraj, A. *Macromolecules* **2014**, *47*, 438–446.
[25] Starovoytov, O. N.; Borodin, O.; Bedrov, D.; Smith, G. D. *J. Chem. Theory Comput.* **2011**, *7*, 1902–1915.
[26] Earl, D. J.; Deem, M. W. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3910–3916.
[27] Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
[28] Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269–6271.
[29] van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comp. Chem.* **2005**, *26*, 1701–1718.
[30] Berne, B. J.; Pecora, R. *Dynamic Light Scattering*; John Wiley & Sons: New York, 1976.
[31] Yamakawa, H. *Modern Theory of Polymer Solutions*; Harper & Row: New York, 1971; Eq. (8.50).
[32] Yamakawa, H. *Modern Theory of Polymer Solutions*; Harper & Row: New York, 1971; Sec. 7a.