Control Over Conformational Landscapes of Polypeptoids by Monomer Sequence Patterning

Audra J. DeStefano[1†], Shawn D. Mengel[1†], Morgan W. Bates[2], Sally Jiao[1], M. Scott Shell[1], Songi Han[4*], Rachel A. Segalman[1,3,5]*

[1]Department of Chemical Engineering, University of California, Santa Barbara, California 93106 U.S.A.

[2]California NanoSystems Institute, University of California, Santa Barbara, California, 93106 U.S.A.

[3]Department of Chemistry and Biochemistry, University of California, Santa Barbara, California, 93106 U.S.A.

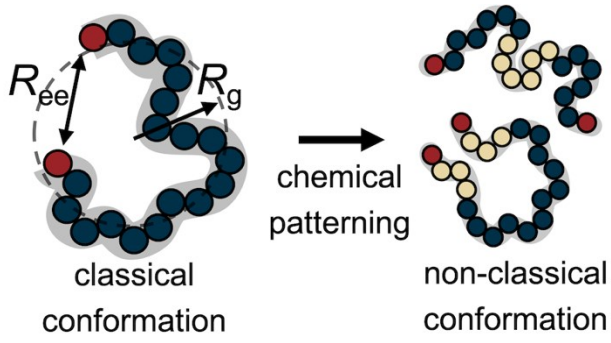
[4]Department of Chemistry, Northwestern University, Evanston, Illinois, 60208 U.S.A.

[5]Department of Materials, University of California, Santa Barbara, California, 93106 U.S.A.

[†]AJD and SM contributed equally to this work.

*Corresponding authors. Email: songi@chem.ucsb.edu, segalman@ucsb.edu

For Table of Contents use only



Keywords: sequence control, conformation, polypeptoids, double electron-electron resonance, hydrophobic

Abstract:

The ability to program chain conformation and structure through control over the monomer sequence of synthetic polymers has broad implications for next-generation material design. While related problems of protein-folding and *de novo* design have generated accurate predictions of 3D folded chain structures, generalization to synthetic polymers remains intractable due to the requirement of large structural databases and the intrinsically disordered nature of polymer building blocks. In this work, polypeptoids, a class of peptidomimetic synthetic polymers, are utilized to build a general workflow for the study of relationships between monomer sequence and dynamic 3D chain structure in solution. This work demonstrates how control over the monomer sequence can alter the conformational landscape of synthetic polymers to deviate dramatically from classical chain statistics. Specifically, the distribution of end-to-end distances, as measured by double electron-electron resonance spectroscopy in dilute solvent, is systematically skewed towards shorter distances with an increasing number of hydrophobes and further refined by hydrophobe arrangement in amphiphilic polypeptoid chains.

1. Introduction

Polymer properties relate to a hierarchy of chain structures that originate from the sequence of monomers and evolve due to the unique intermolecular interactions present in the given environment. The importance of sequence on chain conformation is best exemplified by biological macromolecules, such as proteins, where the precise arrangement of amino acids controls the formation of helical- and sheet-shaped segments, which guide the chain into a specific three-dimensional shape. The protein's folded shape then dictates its unique function. A large class of protein sequences give rise to partially or fully intrinsically disordered protein architectures that still exert important biological function, despite or because of this disorder, and whose ensemble conformational distributions are therefore critical to their properties and functions. Replication of specificity in function through precise sequence control in synthetic polymer systems can similarly afford opportunities to design highly tunable materials for optoelectronic devices, catalysis, lithography, stimuli-responsive materials, and drug delivery. [1, 2]

This grand challenge is closely related to the protein folding and *de novo* protein design problems, which rely heavily on AI-based algorithms for protein structure prediction. However, such methods lack fundamental insights into sequence/structure relationships and are ungeneralizable to synthetic backbone chemistries due to their reliance on statistical learning methods applied to large structural datasets of known protein structures in highly complex sequence space. [3-5] As a result, application of sequence control to synthetic polymers has been limited to block copolymers with relatively few blocks (typically $\leq$3) to restrict the design space to an experimentally manageable size, which has now been thoroughly explored. Increases in sequence complexity through post-functionalization of block copolymers with additional interaction types has demonstrated improved control over polymer structure and properties,

motivating study of more precise sequence effects[6] and an expanded design space. Future advancements in synthetic polymer sequence design hence require new approaches to investigate this space and understand the interplay between chemical information and the resulting polymeric structure. [7, 8]

Polypeptoids bridge proteins and traditional synthetic polymers, providing an excellent platform for studying sequence-structure relationships. Solid-phase polypeptoid synthesis enables precise control over monomer sequence akin to polypeptides, while accessing a wider library of sidechain functionalities through the addition of primary amines rather than amino acids. *N*-substitution of the side chains simplifies intramolecular interactions by eliminating backbone chirality and hydrogen bonding. [9, 10] Previous studies leveraged these advantages to demonstrate sequence effects on solvent-induced single chain conformational changes and block copolymer self-assembly. [11-16] It remains unclear, however, what design elements of the chain sequence stabilize polymer chain conformation. Inspired by the importance of hydrophobic/hydrophilic patterning in determining protein structure, [17, 18] this study demonstrates relationships that describe how amphiphilic polymer sequences affect chain shape in dilute solution.

Intuitive rules of solvophobic/solvophilic interactions can inform limited design of sequence-defined polymers, but do not provide concrete sequence design principles. Further, even simple design motifs (e.g., binary) cannot be exhaustively searched through experiment, as their sequence possibilities grow exponentially with chain length. The lengths considered here exceed 69 trillion sequence possibilities. To navigate sequence space and resulting chain structures, low-cost computational models capable of high-throughput screening have become an important tool, particularly for protein-inspired macromolecular design. [7, 19] Validated atomistic simulations provide accurate and detailed conformational ensembles. However, their computational expense

limits their utility in design workflows searching large sequence spaces, and they must be paired with coarser models or statistical learning methods. In particular, long conformational transition timescales prohibit the use of atomistic simulations for the 38-mer polypeptoids used in this study. [20] The sequences here are hence designed for compatibility with atomistic simulation methods and compared to a low-cost computational bead-spring model to validate the accuracy for future development of high-throughput sequence screening methods.

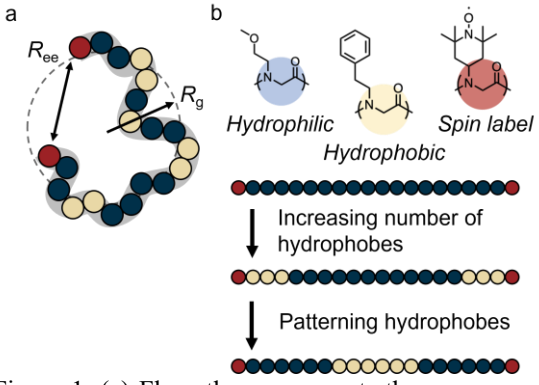


Figure 1. (a) Flory theory connects the average end-to-end distance, $\langle R_{ee}^2 \rangle^{\frac{1}{2}}$, to the radius of gyration, $\langle R_g^2 \rangle^{\frac{1}{2}}$, but few methods exist for experimentally measuring $\langle R_{ee}^2 \rangle^{\frac{1}{2}}$. (b) Incorporation of nitroxide labels at both ends of each polypeptoid uniquely enables characterization of the distribution of R_{ee} from an ensemble of polypeptoids via double electron-electron resonance (DEER). Polypeptoids with varying hydrophobe content and arrangement along the polymer chain (patterning) modulate polymer conformation.

Achieving chain shape control requires both the ability to pattern intra- and inter-molecular interactions along the polymer backbone and a means for observing conformational changes. While folded proteins exhibit precise conformations with well-defined atomic locations and intrachain contacts, polymer chain conformations are disordered and transient. Ideal polymer chains are classically described by end-to-end distance, $\boldsymbol{R_{ee}}$, and radius of gyration, $\boldsymbol{R_g}$, vectors (Figure 1a), both of which exhibit Gaussian distributions. [21] These distributions can be extended to real polymer chains through use of renormalization group theory to account for excluded volume effects. [22] A combination of X-ray scattering and magnetic resonance techniques are widely used to characterize the structure of both folded and, more recently, intrinsically disordered proteins. [23, 24] In contrast, characterization of polymer chain conformations is typically limited to small angle neutron or X-ray scattering, which yield an average radius of gyration, $\langle R_g^2 \rangle^{\frac{1}{2}}$. Few methods exist for experimentally

determining $\langle R_{ee}^2 \rangle^{\frac{1}{2}}$. [20, 25-28] Rather, the end-to-end distance is frequently inferred by a proportionality predicted by Flory theory ($\langle R_g^2 \rangle = \frac{1}{6}\langle R_{ee}^2 \rangle$) with minor corrections for excluded volume effects. [22, 29-32] These theories do accurately reflect the chain conformations of homopolymers and some random copolymers, but do not incorporate the chemical information required to predict sequence effects. Introduction of protein-like sequence control will, by design, induce modest to complete deviation from classical chain statistics by programming chain shapes with narrower, non-Gaussian distributions of $\boldsymbol{R}_{ee}$ and $\boldsymbol{R}_g$. Theoretical predictions of such deviations already exist for block copolymers; however, these theories are challenging to validate by scattering alone and do not extrapolate to sequences of arbitrary complexity. [33-37] As the sequence and conformational complexity of synthetic polymers approach that of proteins, development of chain statistical models for each new sequence motif becomes burdensome. New methods are required to quantify and predict these sequence effects in a way that is compatible with high throughput polymer design.

Direct measurement of $\langle R_{ee}^2 \rangle^{\frac{1}{2}}$ is hence highly desirable to quantify sequence effects on polymer conformations. Critically, the average value $\langle R_{ee}^2 \rangle^{\frac{1}{2}}$ offers limited information. Instead, measurement of the entire end-to-end distance distribution, $P(R_{ee})$, is needed to obtain more detailed insight into the variation of conformations that reflect on the degree and nature of disorder. The $P(R_{ee})$ originating from the full conformational ensemble is uniquely accessible by double electron-electron resonance (DEER) spectroscopy, a pulsed electron paramagnetic resonance technique that can measure distances between 2-8 nm, depending on the experimental conditions. DEER is most widely used for studying structural biology of biomolecular complexes, [26] with comparatively few demonstrations in polymer physics. [25, 38] DEER measures a time domain decay

of an ensemble of pairwise dipolar couplings between radical spin labels that is then transformed into an ensemble distribution of distances between the labels. [20, 26] In this case, these distances are the magnitude of the end-to-end vector, R_{ee}, without any angular information.

Here, control over the end-to-end distance distributions of amphiphilic polypeptoids by precise monomer sequencing is demonstrated via DEER spectroscopy and coarse-grained molecular simulations. Sequence control and determination of the P(R_{ee}) reveals how the number of hydrophobes and their arrangement affects the chains' conformational ensemble (Figure 1b), as quantified by deviations from classic chain statistics. Nitroxide radical-based spin labels at both ends of all sequences enable detection of the intra-polymer distances between chain ends by DEER. Together, the synthetic, characterization, and simulation approaches described here provide foundations for a cohesive workflow capable of identifying polymer sequences with user-defined properties.

2. Experimental and Theoretical Methods

2.1 Polypeptoid Synthesis

All solvents and reagents were purchased from commercial suppliers and were used without further purification. Polypeptoids were synthesized using an automated Prelude peptide synthesizer. 2-chlorotrityl chloride resin (1.02 mmol/g loading, 200 μmol scale, purchased from Aapptec) was swollen overnight in a 50/50 mixture of anhydrous dimethylformamide and dichloromethane. The following morning, the swollen resin was transferred to the prelude and drained. To activate the resin, 2 mL of 1.2 M bromoacetic acid in anhydrous dimethylformamide and 2 mL of 1.2 M diisopropylethylamine in anhydrous dichloromethane were added and mixed for two hours. The resin was then drained and rinsed with dimethylformamide. A series of amine displacement and bromoacetylation steps were performed to synthesize the desired sequences.

Methoxyethylamine and 4-amino TEMPO solutions were prepared at a concentration of 1 M in dimethylformamide. Phenethylamine was prepared at 1.5 M in dimethylformamide. Amine displacement steps were allowed to proceed for 1.5 hours each. Bromoacetylation steps were conducted in a solution of 0.8 M bromoacetic acid and 0.74 M diisopropylcarbodiimide for 20 minutes. The bromoacetylation step was performed twice for each monomer unit in order to maximize yields. The final sequences were acetylated for 30 minutes with a mixture of 0.4 M pyridine and 0.4 M acetic anhydride in dimethylformamide to cap the reactive chain ends.

Polypeptoids were cleaved in a 30% v/v mixture of hexafluoroisopropanol in dichloromethane for 1 hour. The cleaved product was recovered by filtration from the resin beads, followed by rinsing of the resin with dichloromethane. Solvent was then removed under vacuum before lyophilizing to yield a crude polypeptoid product. Polypeptoids were purified by preparative high performance liquid chromatography (HPLC) (Section S1). The mass and purity of the final polypeptoid products was confirmed by analytical HPLC mass spectrometry (Section S2).

2.2 **Measurement of Polypeptoid Conformational Landscapes by DEER**

End-to-end distance distributions were measured for each polypeptoid sequence using DEER (Figure 2), as described previously. [20] Briefly, samples containing approximately 100 μM of polypeptoid were dissolved in 50/50 v/v D_2O/d_8-tetrahydrofuran to ensure full solubility across a range of hydrophobic contents. Solutions were cryo-protected with 30% deuterated glycerol, by volume. Then, approximately 40 μL of solution was loaded into a 3 mm OD, 2 mm ID quartz tube. Samples were flash frozen in liquid nitrogen immediately prior to performing DEER. A Bruker/ColdEdge FlexLine Cryostat (Model ER 4118HV-CF100) maintained the sample temperature at 60 K. All DEER spectra were obtained using a Bruker QT-II resonator with a pulsed

Q-band Bruker E580 Elexsys spectrometer with an Applied Systems Engineering, Model 177 Ka 300 W TWT amplifier. The following four pulse DEER sequence was applied to all samples: $\pi_{obs}/2$- τ_1-π_{obs}-(t-π_{pump})- (τ_2-t)-π_{obs}- τ_2-echo. Nutation experiments were used to determine optimal observer pulses (approximately 10 ns for 90 pulses and 20 ns for 180 pulses). The linear chirp π_{pump} frequency width was set at 80 MHz and its duration at 100 ns, while π_{obs} was 90 MHz above the center of the pump frequency range. τ_1 was set at 126 ns and τ_2 was set at 6 µs. All DEER experiments were signal averaged over at least 10 averages. Distance distributions were extracted from time-domain DEER signals using "model-free" Tikhonov regularization, as implemented by the LongDistances software. [39] Raw data, background-corrected data, and fits were plotted in Section S3.

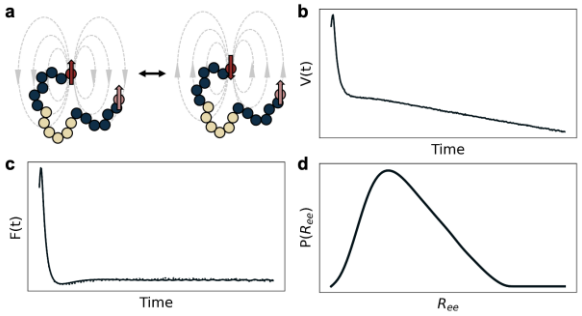


Figure 2. Measuring end-to-end distance distributions via DEER. (a) A four-pulse sequence measures changes in the electron spin echo due to manipulation of the local microwave field between two spins on a macromolecule. (b) Raw DEER data (V(t)) contains the amplitude of the electron spin echo as a function of the pump pulse delay. (c) V(t) is background corrected to remove effects from intermolecular interaction and isolate intramolecular interactions. (d) Tikhonov regularization is typically used to extract a distribution of distances, in this case ($P(R_{ee})$), from the background corrected time-domain data (F(t)).

2.3 Bead Model Simulations of Polypeptoid Sequences

The hydrophobic-hydrophilic polymer model from Panagiotopoulos and coworkers[40] was simulated using the OpenMM simulation engine. The model comprises either hydrophobic or hydrophilic beads, bonded with a Finite Extensible Nonlinear Elastic potential. Spin label residues were not included in the simulations; the simulated chains are thus 18-mers and 36-mers rather than the 20-mers and 38-mers used experimentally. Hydrophobic beads interact with each other

through the Lennard Jones potential, while all interactions involving hydrophilic beads (hydrophilic-hydrophilic, hydrophilic-hydrophobic) are purely repulsive. The only parameters in the model are an energy well depth, ε, the mass of a bead, m, and a lengthscale, σ. Langevin dynamics were simulated with an integration timestep of 0.005τ, where $\tau = \sqrt{m\sigma^2/\varepsilon}$, and a friction coefficient of $0.2m/\tau$. The simulations were run at a temperature of $T = 3\varepsilon/4k_B$, which gives stronger correlations between the experimental and computationally measured mean end-to-end distances compared to higher (T=ε/k_B) and lower ($\varepsilon/2k_B$) temperatures (Fig. S40). Each polymer was simulated at infinite dilution in a box of length 24σ without periodic boundary conditions using the reaction field method to cut off interactions at distances specified by Panagiotopoulos and coworkers.[40] Each simulation was run for $2.5 \times 10^6\tau$, saving configurations every 50τ. The first 5000τ were discarded and the remaining snapshots were used to compute the end-to-end distance distributions. For each snapshot, the end-to-end distance was computed as the distance between the first and last beads. The distributions shown in Figure S11 have a bin width of $2\sigma/3$. To compute the distribution uncertainty, a basic bootstrap was utilized, resampling the dataset 10,000 times, with a sample size corresponding to the number of uncorrelated samples computed from $N_{\mathrm{prod}}/2t_{\mathrm{c}}$ where N_{prod} is the length of the production run ($2.495 \times 10^6\tau$) and t_{c} is the autocorrelation time of the end-to-end distance of the fully hydrophilic sequence for each sequence length ($t_{\mathrm{c}} = 152.5\tau$ for the 18mers and $t_{\mathrm{c}} = 928\tau$ for the 36mers). The uncertainties represent a 95% confidence interval around the plotted median distribution.

2. Results and Discussion

In this work, we present a combined computational and experimental platform for controlling polymer conformation via monomer sequence. Polypeptoids with precisely defined monomer

sequences are synthesized with systematically changing hydrophobic compositions and patterning as a preliminary screen of the sequence design space. Experimental measurements of end-to-end distance distributions indicate the role of sequence parameters on chain conformations and offer new tools for quantifying chain shape. A computationally inexpensive bead-spring model qualitatively reproduces the trends in the polymer R_{ee} obtained by DEER spectroscopy, suggesting the utility of coarse-grained molecular dynamics simulations in predicting material properties and motivating future development of high-throughput sequence screening algorithms.

This study considers conformations of polypeptoid sequences with increasing hydrophobicity and varying hydrophobe arrangement of two lengths (Figure 3). In particular, the effect of hydrophobic content is determined using sequences with increasing number of hydrophobic monomers at both ends ("0H" to "4H"). Because "3H" experimentally shows the maximum deviation from the fully hydrophilic sequence ("0H"), additional sequences of equivalent composition but varied placement of the hydrophobic monomers along the polymer backbone are considered to determine the extent to which sequence patterning can tune chain shape. All of the peptoid chain conformations are measured in the same solvent mixture of deuterated water/tetrahydrofuran/glycerol. The water/tetrahydrofuran ratio is specifically chosen to maximize hydrophobic interactions (maximum water content) while maintaining solubility of the chains.

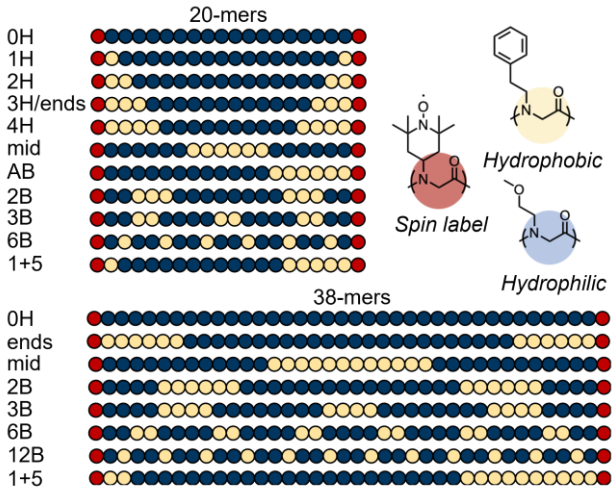


Figure 3. Complete list of polypeptoid sequences tested in this study. Each sequence is comprised of precisely patterned hydrophilic (blue beads) and hydrophobic (yellow beads) monomers. TEMPO spin probes at the chain ends (red beads) enable measurement of P(R_{ee}) by DEER spectroscopy. These sequences probe the effects of increasing hydrophobic content and patterning.

This thereby maximizes the effect of hydrophobic sequence patterning on the conformations of these compositionally equivalent sequences.

The patterned set of sequences are designed using intuition from block copolymers and hydrophobic/hydrophilic patterning common to protein folding. Hydrophobic interactions between the aromatic side chains and the surrounding solvent are expected to drive the hydrophobes together to minimize solvent-hydrophobe contacts. [41] Sequences "ends," "AB," and "mid" therefore test the extremes of this effect. Sequences "2B" to "12B" vary the hydrophobic block size, which has previously been shown to control the density of hydrophobic/hydrophilic copolymer globules. [42-44]

While longer polypeptoids are expected to be more capable of achieving protein-like structural control, even state-of-the-art atomistic simulations possess chain length limitations of around 20 residues. [20] The development of coarse-grained methods capable of accurately simulating longer polymer sequences, therefore, will be necessary to reach the full potential of polypeptoids as biomimetic materials. Toward this goal, this work considers polypeptoids of two chain lengths, 20- and 38-mers. The longer 38-mers possess larger sequence effects and R_{ee} values centered in the 2–8 nm range, making them better suited for DEER spectroscopy. Although pushing the lower boundaries of the DEER technique, the 20-mers offer a bridge to connect experimental measurements with atomistic simulations for the future development of coarse-grained models. Together, these materials demonstrate that sequence can be used to tune chain shape and motivate the use of high throughput computational models to guide non-intuitive sequence discovery.

Distributions of R_{ee} ($P(R_{ee})$) obtained from DEER show that increasing the polypeptoid hydrophobicity at the chain ends drives them together, shifting the ensemble of chain conformations toward more compact structures (Figure 4a). This effect is most drastically seen in

the root-mean-squared end-to-end distances of each sequence, with an increasing number of hydrophobes resulting in shorter average R_{ee} values (Figure 4b). Despite equivalent contour lengths across all sequences, the more hydrophobic polypeptoids exhibit a clear trend toward more compact chain conformations, likely because increasing the hydrophobicity of the polymer changes the effective solvent quality experienced by the chains. As the chain becomes more hydrophobic, it is more poorly solvated, resulting in contraction of the chains toward the globule limit.

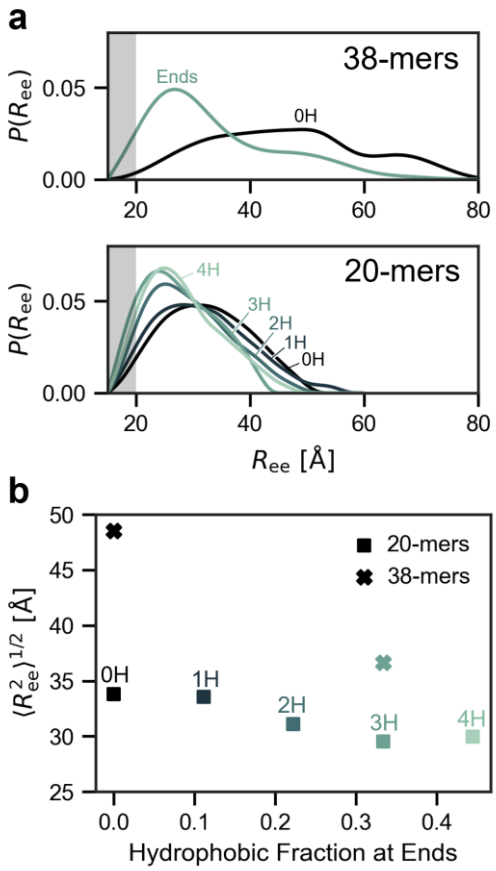


Figure 4. Increasing the number of hydrophobes on both ends leads to more compact conformations. (a) P(R_{ee}) narrow and shift to shorter distances with increasing hydrophobicity for both 20-mer and 38-mer sequences. (b) The rotationally averaged end-to-end distance, $\langle R_{ee}^2 \rangle^{\frac{1}{2}}$, central to Flory theory reflects the trend of chain collapse with increased hydrophobic content.

In addition to average chain composition, the arrangement of a fixed number of hydrophobes also influences the P(R_{ee}), as shown by the compositionally equivalent series of sequences (Fig. 5a). Placing the hydrophobic residues at the center of the chain ("Mid" sequence, Fig. 5b) extends the $\langle R_{ee}^2 \rangle^{\frac{1}{2}}$, while placement at the chain ends ("Ends") contracts $\langle R_{ee}^2 \rangle^{\frac{1}{2}}$ in comparison to the more evenly distributed "6B" sequence. This indicates that chain contraction due to solvent interaction is influenced by engineering of monomer sequence in the polypeptoids to achieve specific chain conformations, rather

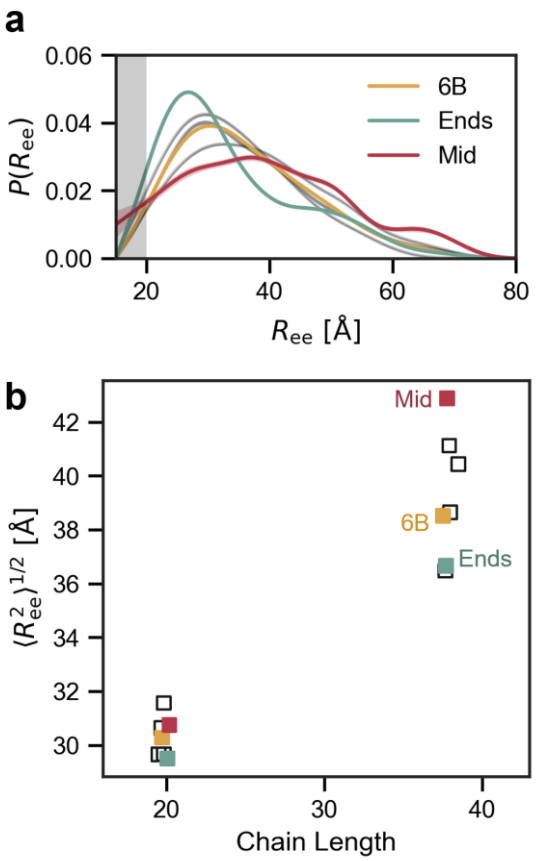


Figure 5. Patterning a set number of hydrophobes fine-tunes polymer conformation. (a) The longest polymers considered in this study (38-mer) best exemplify sequence-dependent conformational changes. For ease of viewing, several sequences are highlighted. These sequences show that placing the hydrophobes on both ends ("Ends") leads to more compact conformations than more evenly distributed hydrophobes ("6B"). Placing all hydrophobes at the middle of the sequence ("Mid") leads to relatively extended conformations compared to the other patterns considered. (b) Trends in $\langle R_{ee}^2 \rangle^{\frac{1}{2}}$ remain consistent for both chain lengths, suggesting that monomer sequence provides a handle to tune polymer conformation.

than averaged mixing effects. This result is consistent with previously observed sequence-dependent variations in polypeptoid hydration water diffusivity.[45] The $\langle R_{ee}^2 \rangle^{\frac{1}{2}}$ values of all the patterned sequences are smaller than that of the hydrophilic chain ("0H"), suggesting that some chain contraction still occurs, potentially due to global effects of decreased effective solvent quality with increasing peptoid hydrophobicity. However, placing the hydrophobic residues in the center or ends of the chain restricts globular collapse to those portions of the chain, while the rest fluctuates in a more extended state.

The 4H 20-mer sequence is an interesting outlier to the trend of hydrophobicity in Figure 4, possessing a larger $\langle R_{ee}^2 \rangle^{\frac{1}{2}}$ than the less hydrophobic 3H sequence. This is in part an artifact of the DEER cutoff which has an increasingly significant impact at these short distances. This increase in $\langle R_{ee}^2 \rangle^{\frac{1}{2}}$ may also be a result of the hydrophobic interactions being spread over a larger section of the polymer chain. For the 1H and 2H sequences, the hydrophobes are concentrated only at the chain ends, drawing the ends directly toward one another with increasing force due to the increased number of hydrophobic interactions. However, in the 4H sequence, 44%

of the chain is hydrophobic. As a result, there is a larger set of favorable pairwise interactions between hydrophobes that do not necessarily require close proximity of the chain ends. The additional conformations associated with these interactions hence lead to an increase in $\langle R_{ee}^2 \rangle^{\frac{1}{2}}$.

Sequence-dependent variations in R_{ee} are generally consistent between the two chain lengths. As shown in Figure 5b, at both lengths, the "mid" sequences possess greater $\langle R_{ee}^2 \rangle^{\frac{1}{2}}$ than the "ends" sequences, with the blocky "6B" in between. However, differences in P(R_{ee}) between the sequences are more pronounced among the 38-mer polypeptoids compared to the 20-mers. This is likely due to the reduced conformational freedom of the shorter chains, which reduces the range of possible R_{ee} values, as well as limitations in the measurable length scale of the DEER technique. Based on the typical $\langle R_{ee}^2 \rangle^{\frac{1}{2}}$ values measured by DEER (approximately 3 nm), roughly 15% of the P(R_{ee}) distribution for the 20-mer sequences should lie outside of the range accessible to DEER, reducing the technique's ability to resolve differences in $\langle R_{ee}^2 \rangle^{\frac{1}{2}}$. Nonetheless, the general agreement of the shorter sequences with the 38-mers suggests that they can still provide basic information relating sequence to chain shape for future development of coarse-grained simulation models. Further comparison between the full set of compositionally equivalent sequences can be found in the Supporting Information.

While atomistic simulations provide high resolution, they are prohibitively expensive for screening large design spaces for new, non-intuitive sequence designs. Bead-spring models by contrast are computationally inexpensive, even for long chains or multi-chain systems, but lack chemical information, making them difficult to relate to real materials. Coarse-grained models

offer a compromise between cost and atomistic resolution, yet it is not currently understood what level of resolution is required to accurately model monomer sequence effects. To better understand this trade-off, a bead-spring model capturing the main interaction forces (hydrophobic attractions) is used to simulate the 20-mer and 38-mer sequences.

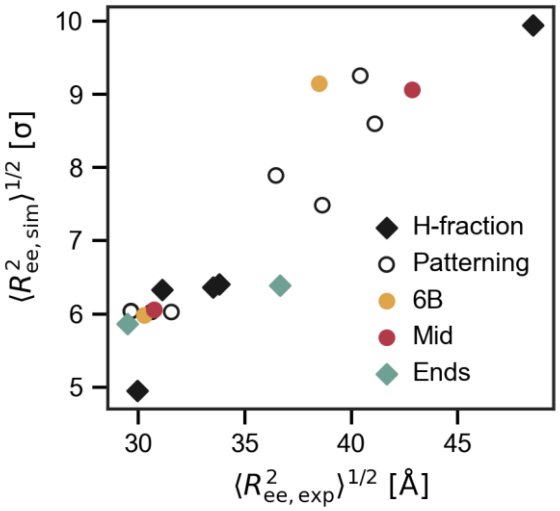


Figure 6. Low-cost bead model predicts conformational changes. Trends in $\langle R^2_{ee}\rangle^{\frac{1}{2}}$ generally agree with experiment, partially replicating experimental results for the patterned sequences. Because the model represents each monomer as a bead with size σ, computed $\langle R^2_{ee}\rangle^{\frac{1}{2}}$ are plotted in units of σ, rather than Å.

Simulations from the bead-spring model reflect trends in $\langle R^2_{ee}\rangle^{\frac{1}{2}}$ due to changes in chain length, hydrophobic fraction, and sequence patterning effects. As seen in Figure 6, there exists a positive correlation ($r = 0.929$) between the experimentally measured $\langle R^2_{ee}\rangle^{\frac{1}{2}}$ and those calculated from the simulation. The 38-mers all possess larger average values than the 20-mers. The "0H" to "4H" series (represented as diamond symbols) and the 38-mer patterned sequences (circle symbols) are also correlated. However, experiments show small differences in $\langle R^2_{ee}\rangle^{\frac{1}{2}}$ among 20-mer patterned sequences that are not predicted by simulation. The 2 nm lower cut-off intrinsic to the DEER technique likely accounts for some of the deviation between the model and experiment for the shorter patterned polypeptoids, while extending sequences to 38 monomers largely overcomes this limitation. From the computational side, variations in chain stiffness not captured in the model may also contribute to differences in $\langle R^2_{ee}\rangle^{\frac{1}{2}}$. While unable to resolve the most nuanced sequence effects, this simplistic model demonstrates a baseline of conformational resolution that predicts compositional and sequence effects well. We expect that incorporation of

some atomistic detail through coarse-grained MD simulations will further improve predictions to enable nuanced, high throughput screening of vast sequence spaces.

Measurement of the entire P(R_{ee}) by DEER, rather than an average value, provides additional insights into the conformational ensembles as a function of polypeptoid sequence through analysis of higher order statistical moments of the distributions (e.g., width and skew). Due to the DEER cutoff, higher order moments are only calculated for the 38-mer polypeptoids, where an estimated 92% of P(R_{ee}) is estimated to reside within the measurable range. [46] Underlying variations in $\langle R_{ee}^2 \rangle^{\frac{1}{2}}$, the variance and skewness of the 38-mer P(R_{ee}) are sequence-dependent (Figure 7). Here, the variance (distribution width) and skewness (distribution symmetry) are calculated as the second central and third standardized moments, respectively, from P(R_{ee}) distributions above the 2 nm DEER cutoff. The 38-mers are compared to calculated moments of a theoretical P(R_{ee}) for the excluded volume model.[22] This theoretical P(R_{ee}) is defined to have an average end-to-end distance matching that of the experimental 0H 38-mer and is similarly truncated at 2 nm to mimic the effect of the DEER cutoff on the distribution. The variance is normalized by the first moment (distribution mean) squared to provide a size-independent metric of the distribution width. A positive skewness indicates a higher density of shorter conformations with a tail at higher R_{ee} values.

Differences in variance between the sequences are small, consistent with the design that all of the polypeptoid ensembles studied here have significant disorder and hence access a similarly wide range of conformations. The measured variances are smaller than that computed from the excluded volume chain due to artificial reduction of the distribution width imposed by the DEER cutoff. The presence and sequence of hydrophobic moieties cause the ensemble to skew substantially from excluded volume behavior previously observed for hydrophilic polypeptoids. [20] In particular, the

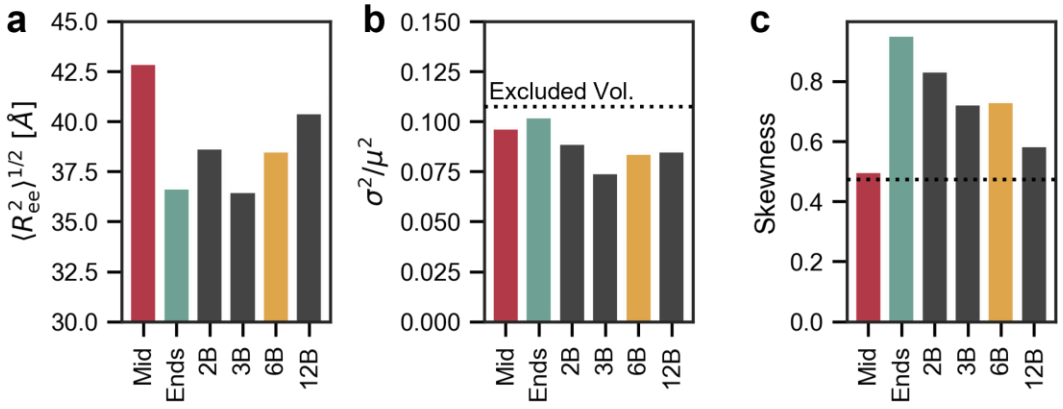


Figure 7. Measurement of full P(R_{ee}) enables analysis of higher order moments. (a) Polymer conformations are often characterized in terms of average quantities such as R_g and $\langle R_{ee}^2 \rangle^{\frac{1}{2}}$. For the 38-mer sequences considered here, $\langle R_{ee}^2 \rangle^{\frac{1}{2}}$ is shortest when hydrophobes are placed at both ends ("Ends"), longest when all hydrophobes are placed at the center of the polymer chain ("Mid"), in between the two extremes when the hydrophobes are distributed into blocks ("2B" to "12B"). (b) Distribution variances show small sequence effects, and all are lower than that predicted for excluded volume chains, due in part to the lower limit of the DEER technique falling at 2nm. (c) Sequence more strongly impacts skewness with decreasing the block size of evenly distributed blocks ("2B" to "12B") generally leading to less positive skewness.

"ends" sequence shows the greatest skew because the hydrophobic chain ends give rise to a greater proportion of short R_{ee}. Shifting the position of the hydrophobic blocks closer to the center decreases skewness ("2B" and "mid"). Increasing the number of hydrophobic blocks also brings the skewness closer to that of ideal excluded volume polymers. This behavior is seen both in the experimental DEER data as well as the simulated distributions (Figure S42). As the number of hydrophobic blocks increases, the block size decreases, yielding a more homogenous composition among each blob along the backbone. The decreasing skewness toward the excluded volume model limit hence provides experimental evidence of an upper limit on the level of sequence complexity that can tune chain shape of disordered polymers, where sequence features smaller than the polymer blob size become averaged out by neighboring residues. The blob size of disordered polypeptides has been estimated to be 5-7 residues, which reasonably matches the length scale over which the patterned polypeptoid sequences become more ideal. [47] This study observes conformational differences between polypeptoid sequences of equivalent overall

composition, indicating an expected, but previously difficult to observe, sequence-induced deviation from Flory theory. [29]

Conclusions

This study demonstrates that control over polypeptoid conformational ensembles can be achieved through precise monomer sequence patterning. Measurement of P(R_{ee}) and $\langle R_g^2 \rangle^{\frac{1}{2}}$ reveals that both chain composition and exact hydrophobe placement create deviations from classical polymer theory. Initial results agree qualitatively with theories and experimental observations of BAB triblock copolymers[35] and hydrophobic self-associating polymers[48] in selective solvents but demonstrate new capabilities for characterizing more complicated sequence-defined polymers. Past studies have been limited by indirect measures of chain conformation (e.g., viscometry) as well as poor sequence precision and high dispersity. The workflow presented here that relies on the direct measurement of the full R_{ee} distribution by DEER overcomes these issues while generalizing to arbitrary sequence complexity. The low-cost computational model presented here produces trends that generally match experimental results, predicting both the compositional and patterning effects on chain conformations. This demonstrates the potential of high-throughput models to screen and predict sequence-encoded properties and motivates further development of coarse-grained models to improve resolution in nuanced sequence effects. Together, this work pushes beyond analysis of mean-field polymer-solvent interactions and random walk chain statistics to provide a powerful approach for elucidating sequence-structure relationships of synthetic polymers. Such capabilities are necessary for achieving chain shape design rivaling the structural complexity of proteins.

Supporting Information

Reverse-phase liquid chromatographs, time domain DEER data and fits, bead-spring model end-to-end distance distributions, additional discussion of distribution moment analysis and comparison between chain lengths.

Acknowledgements

The polymer synthesis and characterization were supported by the National Science Foundation under Grant No. 2203179 (SDM, AJD, RAS) leveraging facilities and expertise from the BioPACIFIC Materials Innovation Platform of the National Science Foundation under Award No. DMR-1933487 (MWB). Development of the DEER technique and computational model was supported by the Center for Materials for Water and Energy Systems (M-WET), an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Award #DE-SC0019272. SDM and SJ acknowledge support from the National Science Foundation Graduate Research Fellowship (DGE 2139319, DGE 1650114). AJD acknowledges support from the Department of Defense through the National Defense Science & Engineering Graduate (NDSEG) Fellowship Program. S.H. thanks the National Institute of General Medicine grant (R35GM136411) for support. We also thank Dr. Xiangxi (Zoey) Meng for optimizing the polypeptoid submonomer method for 2-CTC resin.

References

(1) Polymeropoulos, G.; Zapsas, G.; Ntetsikas, K.; Bilalis, P.; Gnanou, Y.; Hadjichristidis, N. 50th Anniversary Perspective: Polymers with Complex Architectures. *Macromolecules* **2017**, *50* (4), 1253-1290. DOI: 10.1021/acs.macromol.6b02569.
(2) Bates, F. S.; Hillmyer, M. A.; Lodge, T. P.; Bates, C. M.; Delaney, K. T.; Fredrickson, G. H. Multiblock Polymers: Panacea or Pandora's Box? *Science* **2012**, *336* (6080), 434-440. DOI: 10.1126/science.1215368.
(3) Korendovych, I. V.; DeGrado, W. F. De novo protein design, a retrospective. *Q Rev Biophys* **2020**, *53*. DOI: 10.1017/S0033583519000131.
(4) Pan, X. J.; Kortemme, T. Recent advances in de novo protein design: Principles, methods, and applications. *J Biol Chem* **2021**, *296*. DOI: 10.1016/j.jbc.2021.100558.
(5) Kuhlman, B.; Bradley, P. Advances in protein structure prediction and design. *Nat Rev Mol Cell Bio* **2019**, *20* (11), 681-697. DOI: 10.1038/s41580-019-0163-x.
(6) Schadler, V.; Wiesner, U. Salt-controlled lamellar spacing in ionically end-capped symmetric diblock copolymers. *Macromolecules* **1997**, *30* (21), 6698-6701. DOI: DOI 10.1021/ma9707852.
(7) Perry, S. L.; Sing, C. E. 100th Anniversary of Macromolecular Science Viewpoint: Opportunities in the Physics of Sequence-Defined Polymers. *ACS Macro Lett* **2020**, *9* (2), 216-225. DOI: 10.1021/acsmacrolett.0c00002
(8) Bates, F. S.; Brant, P.; Coates, G. W.; Lipson, J.; Osuji, C.; Pablo, J. d.; Rowan, S.; Segalman, R.; Winey, K. I. *Frontiers in Polymer Science and Engineering*; NSF Workshop, University of Minnesota Twin Cities, 2016.
(9) Zuckermann, R. N.; Kerr, J. M.; Kent, S. B. H.; Moos, W. H. Efficient Method for the Preparation of Peptoids [Oligo(N-Substituted Glycines)] by Submonomer Solid-Phase Synthesis. *J Am Chem Soc* **1992**, *114* (26), 10646-10647. DOI: 10.1021/ja00052a076.
(10) Rosales, A. M.; Segalman, R. A.; Zuckermann, R. N. Polypeptoids: a model system to study the effect of monomer sequence on polymer properties and self-assembly. *Soft Matter* **2013**, *9* (48), 11713-11713.
(11) Murnen, H. K.; Rosales, A. M.; Dobrynin, A. V.; Zuckermann, R. N.; Segalman, R. A. Persistence length of polyelectrolytes with precisely located charges. *Soft Matter* **2013**, *9* (1), 90-98. DOI: 10.1039/c2sm26849c.
(12) Yu, B.; Li, R.; Segalman, R. A. Tuning the Double Gyroid Phase Window in Block Copolymers via Polymer Chain Conformation Near the Interface. *Macromolecules* **2021**, *54* (12), 5388-5396. DOI: 10.1021/acs.macromol.1c00048.
(13) Sternhagen, G. L.; Gupta, S.; Zhang, Y. H.; John, V.; Schneider, G. J.; Zhang, D. H. Solution Self-Assemblies of Sequence-Defined Ionic Peptoid Block Copolymers. *J Am Chem Soc* **2018**, *140* (11), 4100-4109. DOI: 10.1021/jacs.8b00461.
(14) Kang, L.; Chao, A.; Zhang, M.; Yu, T.; Wang, J.; Wang, Q.; Yu, H.; Jiang, N.; Zhang, D. Modulating the Molecular Geometry and Solution Self-Assembly of Amphiphilic Polypeptoid Block Copolymers by Side Chain Branching Pattern. *J Am Chem Soc* **2021**, *143* (15), 5890-5902. DOI: 10.1021/jacs.1c01088.
(15) Jiang, N. S.; Yu, T. Y.; Darvish, O. A.; Qian, S.; Tsengam, I. K. M.; John, V.; Zhang, D. H. Crystallization-Driven Self-Assembly of Coil-Comb-Shaped Polypeptoid Block Copolymers: Solution Morphology and Self-Assembly Pathways. *Macromolecules* **2019**, *52* (22), 8867-8877. DOI: 10.1021/acs.macromol.9b01546.

(16) Hudson, B. C.; Battigelli, A.; Connolly, M. D.; Edison, J.; Spencer, R. K.; Whitelam, S.; Zuckermann, R. N.; Paravastu, A. K. Evidence for cis Amide Bonds in Peptoid Nanosheets. *Journal of Physical Chemistry Letters* **2018**, *9* (10), 2574-2578. DOI: 10.1021/acs.jpclett.8b01040.
(17) Cheng, Y. K.; Rossky, P. J. Surface topography dependence of biomolecular hydrophobic hydration. *Nature* **1998**, *392* (6677), 696-699. DOI: 10.1038/33653.
(18) Kauzmann, W. Some Factors in the Interpretation of Protein Denaturation. *Adv Protein Chem* **1959**, *14*, 1-63. DOI: 10.1016/S0065-3233(08)60608-7.
(19) DeStefano, A. J.; Segalman, R. A.; Davidson, E. C. Where Biology and Traditional Polymers Meet: The Potential of Associating Sequence-Defined Polymers for Materials Science. *Jacs Au* **2021**, *1* (10), 1556-1571. DOI: 10.1021/jacsau.1c00297.
(20) Jiao, S.; DeStefano, A.; Monroe, J. I.; Barry, M.; Sherck, N.; Casey, T.; Segalman, R. A.; Han, S.; Shell, M. S. Quantifying Polypeptoid Conformational Landscapes through Integrated Experiment and Simulation. *Macromolecules* **2021**, *54* (11), 5011-5021. DOI: 10.1021/acs.macromol.1c00550.
(21) Hiemenz, P. C.; Lodge, T. P. *Polymer Chemistry*; CRC Press, 2007.
(22) Schäfer, L. Dilute Limit: Details on the Internal Structure of Isolated Coils. In *Excluded Volume Effects in Polymer Solutions*, Springer, 1999; p 311.
(23) Bernado, P.; Svergun, D. I. Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Mol Biosyst* **2012**, *8* (1), 151-167. DOI: 10.1039/c1mb05275f From NLM Medline.
(24) Kosol, S.; Contreras-Martos, S.; Cedeno, C.; Tompa, P. Structural characterization of intrinsically disordered proteins by NMR spectroscopy. *Molecules* **2013**, *18* (9), 10802-10828. DOI: 10.3390/molecules180910802.
(25) Sherck, N.; Webber, T.; Brown, D. R.; Keller, T.; Barry, M.; DeStefano, A.; Jiao, S.; Segalman, R. A.; Fredrickson, G. H.; Shell, M. S.; et al. End-to-End Distance Probability Distributions of Dilute Poly(ethylene oxide) in Aqueous Solution. *J Am Chem Soc* **2020**, *142* (46), 19631-19641. DOI: 10.1021/jacs.0c08709.
(26) Jeschke, G. DEER Distance Measurements on Proteins. *Annu Rev Phys Chem* **2012**, *63*, 419-446. DOI: 10.1146/annurev-physchem-032511-143716.
(27) Moeglich, A.; Joder, K.; Kiefhaber, T. End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation (vol 103, pg 12394, 2006). *P Natl Acad Sci USA* **2008**, *105* (18), 6787-6787. DOI: 10.1073/pnas.0803144105.
(28) Merchant, K. A.; Best, R. B.; Louis, J. M.; Gopich, I. V.; Eaton, W. A. Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *P Natl Acad Sci USA* **2007**, *104* (5), 1528-1533. DOI: 10.1073/pnas.0607097104.
(29) Flory, P. J. *Principles of Polymer Chemistry*; Cornell University Press, 1953.
(30) Ewart, R. H.; Roe, C. P.; Debye, P.; Mccartney, J. R. The Determination of Polymeric Molecular Weights by Light Scattering in Solvent-Precipitant Systems. *J Chem Phys* **1946**, *14* (11), 687-695. DOI: 10.1063/1.1724085.
(31) Li, B.; Madras, N.; Sokal, A. D. Critical Exponents, Hyperscaling, and Universal Amplitude Ratios for 2-Dimensional and 3-Dimensional Self-Avoiding Walks. *J Stat Phys* **1995**, *80* (3-4), 661-754. DOI: 10.1007/Bf02178552.
(32) Duplantier, B. Geometrical properties of a Kuhnian polymer chain. *Journal de Physique* **1986**, *47* (10), 1633-1656. DOI: 10.1051/jphys:0198600470100163300.

(33) Tanaka, T.; Kotaka, T.; Ban, K.; Hattori, M.; Inagaki, H. Conformation of Block Copolymers in Dilute-Solution - Molecular Dimension-Block Architecture Relationships. *Macromolecules* **1977**, *10* (5), 960-967. DOI: 10.1021/ma60059a014.
(34) Tanaka, T.; Omoto, M.; Inagaki, H. Conformation of Block Copolymers in Dilute-Solution .3. Determination of the Center-to-Center Distance between the 2 Blocks by Light-Scattering. *Macromolecules* **1979**, *12* (1), 146-152. DOI: 10.1021/ma60067a030.
(35) Kotaka, T.; Tanaka, T.; Inagaki, H. Thermodynamic and Conformational Properties of Styrene-Methyl Methacrylate Block Copolymers in Dilute-Solution .4. Behavior of Diblock and Triblock Copolymers in Selective Solvents. *Polym J* **1972**, *3* (3), 327-&. DOI: 10.1295/polymj.3.327.
(36) Tanaka, T.; Kotaka, T.; Inagaki, H. Conformation of Block Copolymers in Dilute-Solution - Monte-Carlo Calculations and Light-Scattering-Studies on Di-Block Copolymer Systems. *Macromolecules* **1976**, *9* (4), 561-568. DOI: 10.1021/ma60052a006.
(37) Dondos, A.; Rempp, P.; Benoit, H. Investigations on Conformations of Block and of Random Copolymers in Dilute Solution. *Makromolekul Chem* **1969**, *130* (Dec), 233-&.
(38) Pfannebecker, V.; Klos, H.; Hubrich, M.; Volkmer, T.; Heuer, A.; Wiesner, U.; Spiess, H. W. Determination of end-to-end distances in oligomers by pulsed EPR. *J Phys Chem-Us* **1996**, *100* (32), 13428-13432. DOI: 10.1021/jp960895v.
(39) *LongDistances*; 2022. https://sites.google.com/site/altenbach/labview-programs/epr-programs/long-distances (accessed May 2022).
(40) Statt, A.; Casademunt, H.; Brangwynne, C. P.; Panagiotopoulos, A. Z. Model for disordered proteins with strongly sequence-dependent liquid phase behavior. *J Chem Phys* **2020**, *152* (7). DOI: 10.1063/1.5141095.
(41) Lau, K. F.; Dill, K. A. Theory for Protein Mutability and Biogenesis. *P Natl Acad Sci USA* **1990**, *87* (2), 638-642. DOI: 10.1073/pnas.87.2.638.
(42) Khokhlov, A. R.; Khalatur, P. G. Protein-like copolymers: Computer simulation. *Physica A* **1998**, *249* (1-4), 253-261. DOI: 10.1016/S0378-4371(97)00473-1.
(43) Khokhlov, A. R.; Khalatur, P. G. Conformation-dependent sequence design (engineering) of AB copolymers. *Phys Rev Lett* **1999**, *82* (17), 3456-3459. DOI: 10.1103/PhysRevLett.82.3456.
(44) Murnen, H. K.; Khokhlov, A. R.; Khalatur, P. G.; Segalman, R. A.; Zuckermann, R. N. Impact of Hydrophobic Sequence Patterning on the Coil-to-Globule Transition of Protein-like Polymers. *Macromolecules* **2012**, *45* (12), 5229-5236. DOI: 10.1021/ma300707t.
(45) Jiao, S.; Mirabal, D. M. R.; DeStefano, A. J.; Segalman, R. A.; Han, S.; Shell, M. S. Sequence Modulates Polypeptoid Hydration Water Structure and Dynamics. *Biomacromolecules* **2022**, *23* (4), 1745-1756. DOI: 10.1021/acs.biomac.1c016137.
(46) Schäfer, L. *Excluded Volume Effects in Polymer Solutions*; Springer Berlin, Heidelberg, 1999. DOI: 10.1007/978-3-642-60093-7.
(47) Pappu, R. V.; Wang, X.; Vitalis, A.; Crick, S. L. A polymer physics perspective on driving forces and mechanisms for protein aggregation. *Arch Biochem Biophys* **2008**, *469* (1), 132-141. DOI: 10.1016/j.abb.2007.08.033.
(48) Bock, J.; Siano, D. B.; Valint, P. L.; Pace, S. J. Structure and Properties of Hydrophobically Associating Polymers. *Adv Chem Ser* **1989**, (223), 411-424.