Sequence Engineering of Copolymers using Evolutionary Computing

Ashwin A Bale\textsuperscript{1} and Tarak K Patra\textsuperscript{2}\textsuperscript{*}

\textsuperscript{1}Department of Chemical Engineering, Birla Institute of Technology and Science Pilani-Hyderabad, Hyderabad, TS 500078, India

\textsuperscript{2}Department of Chemical Engineering, Center for Atomistic Modeling and Materials Design and Center for Carbon Capture Utilization and Storage, Indian Institute of Technology Madras, Chennai TN 600036, India

Abstract

The correlations between the sequence of monomers in a macromolecule and its three-dimensional (3D) structure is a grand challenge in polymer science and biology. The properties and functions of macromolecules – both synthetic polymers and biomolecules depend on their 3D shape that has appeared to be dictated by their monomer sequence. However, the progress towards understanding the sequence-structure-property correlations and their utilization in materials engineering are slow because it is almost impossible to characterize astronomically large number of possible sequences of a copolymer using traditional experimental and simulation methods. To address this problem, here, we combine evolutionary computing and coarse-grained molecular dynamics (CGMD) simulation and study the sequence-structure correlations of a model AB type copolymer system. The CGMD based evolutionary algorithm screens the sequence space of the copolymer efficiently and identifies wide range of single molecule structures including extremal radius of gyrations ($R_{gs}$). The data provide new insights on the sequence-$R_g$ correlations of the copolymer system and their impact on the structure and functionality of polymeric materials. The work highlights the opportunities of sequence specific control of macromolecular structure for designing materials with exceptional properties.

Key Words: Evolutionary Computing, Polymer Design, Polymer Sequence, Molecular Simulation

\textsuperscript{*}Corresponding Author, Email: tpatra@iitm.ac.in
INTRODUCTION

Copolymers are a special class of macromolecules that are made of multiple chemical moieties with a rich phase behaviour\textsuperscript{1-3} and tunability in their thermophysical properties.\textsuperscript{4-9} Recent progresses in synthetic chemistry such as solid supported reaction, chemo-selective reaction,\textsuperscript{10} orthogonal organic reaction,\textsuperscript{11,12} acid-catalysed cascade reaction\textsuperscript{13} and soluble phase synthetic strategy with molecular sieving\textsuperscript{14} have been enabling to exercise an unprecedented level of control over sequences of different chemical moieties in a copolymer molecule.\textsuperscript{15,16} These precision polymers provide enormous opportunities for molecular scale materials design with wide range of microstructure, functionality and activity being included in a polymeric material.\textsuperscript{17-25} In a single molecular level, a copolymer’s folding and three dimensional structure can be controlled to a great extent by adjusting its monomer sequences. Consequently, the bulk properties of polymeric materials such as elasticity, conductivity, rigidity, biodegradability, anti-bacterial activity and optoelectronic property can be precisely tuned by controlling the sequence of monomers in their constituting polymers. For example, copolymers with varying sequences are employed to improve the thermodynamic stability of polymer mixtures, and they, therefore, have wide applicability in emulsions and composite materials.\textsuperscript{5,26,27} Even monomer level sequence specific control can lead to compatibilizers that yield zero interfacial energy between two immiscible homopolymers.\textsuperscript{28,29} Similarly, the sequence of monomers in a charged copolymer is found to control its structural phase transitions,\textsuperscript{30,31} liquid-liquid phase transition and associated critical point.\textsuperscript{32} The morphology and interface of tapered block copolymers are found to be largely controlled by the details of the sequence of monomers in the middle block that is present in between two immiscible homogeneous blocks.\textsuperscript{33-35} These works clearly suggest that minor changes in the sequence lead to a large variation in the structure, dynamics and phase behaviour of macromolecules. Moreover, the structure and properties of a copolymer in water and other solutions are also strongly influenced by its monomer sequence. In biological context, intrinsically disordered proteins (IDP) and structured proteins are primarily differed by their amino acid sequences. Why does a particular sequence of monomer prefer crystalline structure while a slight variation in the sequence yields disordered structure is not well understood. Among non-crystalline structures, how does a specific sequence determine swelling or collapsing of a polymer chain is also an unsolved problem. In spite of clear connections between sequence and three dimensional structure of a macromolecule, understanding and establishing their relationship is a complex and largely unsolved problem. The primary bottleneck to address this problem is the vast combinatorial
sequence space of a macromolecule that precludes efficient screening. For instance, a linear copolymer chain with $n$ number of possible monomers and $m$ type of chemical moieties will have $m^n/2$ sequences that need to be explored, excluding the double counting due to the fact that a sequence and its reverse sequence represent one copolymer. Even for an AB type copolymer i.e. 2 types of chemical moieties with a chain length of 100, the total combinations possible are $2^{99}$ which is close to $10^{30}$. In practice, many copolymers possess more than two chemical moieties and several hundreds of monomer units. Given such an enormous sequence space that needs to be explored, it is highly desirable to minimize the number of property measurements – computer simulations or experiments needed to arrive at a sequence that corresponds to a desired target property and sample data across the entire space to establish sequence-structure correlations. Many of the previous studies have focused on diblock copolymers or copolymers with periodic blocks of monomers wherein a copolymer is usually characterized by its mean block length and mass fraction. The number of possible periodic sequences in a given copolymer of fixed chain length is very small and manageable using traditional experiment and simulation methods. However, the sequence space is mostly occupied by the non-periodic sequences that are astronomically large in number. Due to this vastness, non-periodic sequences are not well explored and whether mean block length can explain the variability of non-periodic copolymer’s properties is not known. Therefore, the progress towards developing a comprehensive understanding of sequence-structure-property correlations is limited by the astronomically large number of possible non-periodic sequences of a copolymer that are impossible to characterize using traditional approaches. Towards this end an improved understanding of sequence-structure relations that are valid for all possible sequences (periodic and non-periodic) are essential for their successful adaptation in materials engineering. This requires rapid characterization and uniform sampling of the sequence space of copolymers.

To tackle this problem, machine learning and advanced optimization method are recently used for designing new polymer with target structures, and characterizing sequence specific aggregates in solution phase. Also, new copolymer sequences in bulk phase have been identified using evolutionary algorithm that improve thermal conductive of polymers. Further, in our recent study, we have used coarse-grained molecular simulation based evolutionary search and Monte Carlo tree search to efficiently identify optimal sequences of polymer compatibilizers. These optimization methods, particularly evolutionary methods, are progressively being used as a versatile strategy for material property optimization and
They have been increasingly adopted in the field of polymers and other materials design problems as well. Motivated by these previous studies, here, we employ evolutionary computing to understand the sequence space, conformational space of an 1:1 AB type linear copolymer and design target structures. Designing sequences for complex structures of copolymers are always challenging. It requires a priori understanding of the interactions of the building blocks and sequence-morphology correlations. For example, Khokholov and co-workers introduce “instant image” strategy for protein-like homopolymer design. Within this strategy, a homopolymer with strong attractive interaction between all the monomers are simulated to form a globule structure. Then, monomers on the surface of the globule are identified by human intervention and hydrophilic interactions are selectively assigned to them. Unlike “instant image” approach, here our object is to develop a simple and autonomous method for target structure design. Moreover, we aim to assess the impact of sequence on a copolymer structure by generating homogeneous sequence-structure data across the entire sequence space. These two interconnected objectives are addressed by combining implicit solvent molecular dynamics (MD) simulation and evolutionary algorithm (EA). Our MD-EA determines the optimal sequences of a copolymer that give rise to its highest and lowest in equilibrium radius of gyration in an implicit solvent environment. The MD-EA workflow rapidly identifies target sequences of extremal gyration despite the search space extending from one million to $10^{30}$. The MD-EA algorithm screens around 3000 candidate sequences to identify an optimal candidate and has proved to be very efficient and scalable for sequence engineering. The data suggest that both the compact structure and loose structure of the copolymer correspond to non-periodic sequences for all the polymer chain lengths studied in this work. The low $R_g$ structures are similarly to globular protein-like aggregates that consist of a hydrophobic monomers in the core and polar groups on the surface. The current design scheme can identify protein-like copolymers autonomously. The data generated using this molecular simulation based evolutionary computing are analyzed to establish new sequence-structure correlations of copolymers. We further test the impact of sequence on the packing of a single macromolecule as well as bulk polymeric material. We demonstrate that the known range of material structures and assemblies can be expanded by monomer-to-monomer sequence control. The work outlines important design rules of sequence defined polymers, and has fundamental implications in designing exceptional polymeric materials.
MODEL AND METHODS

Model Copolymer. A generic coarse-grained model of a copolymer is considered to study the impact of sequence on its structure. In this model system, two adjacent coarse-grained monomers of the copolymer are connected by the Finitely Extensible Nonlinear Elastic (FENE) potential of the form \( E = -\frac{1}{2} K R_0^2 \ln \left( 1 - \left( \frac{r}{R_0} \right)^2 \right) \), where \( K \approx 30 \epsilon/\sigma^2 \) and \( R_0 = 1.5 \sigma \). The pair interaction between any two monomer is modelled by the Lennard-Jones (LJ) potential of the form \( V(r) = 4 \epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right] \). The \( \epsilon \) is the unit of pair interaction energy. The size of all the monomers is \( \sigma \). The LJ interaction is truncated and shifted to zero at a cut-off distance \( r_c = 2.5 \sigma \) to represent attractive interaction among the monomers. The AA interaction strength is \( \epsilon_{AA} = \epsilon \), while the BB and AB type interaction strengths are \( \epsilon_{BB} = \epsilon_{AB} = 0.2 \epsilon \). This ensures immiscibility between A and B type monomers. These set of model parameters are previous used to study copolymer systems and are very successful in understanding the generic properties of both synthetic polymers and biomolecules.

Molecular Dynamics Simulation. We conduct implicit solvent molecular dynamics simulations of a copolymer chain in a canonical ensemble. The initial configuration of a polymer chain is placed in a cubic simulation box of fixed dimension and box is periodic in all three directions. We use the Velocity Verlet algorithm with a timestep of 0.001 \( \tau \) to integrate the equation of motion. Here, \( \tau = \sigma \sqrt{m/\epsilon} \) is the unit of time, and \( m \) is the mass of a monomer, which is same for both the A and B type moieties. All the simulations are conducted at a reduced temperature \( T^* = T \epsilon/k_B = 1 \), which is maintained by the Langevin thermostat within the LAMMPS simulation environment. The simulations are conducted for four different polymer chains of length \( N= 20, 50, 100 \) and \( 150 \). All the simulations are equilibrated for \( 10^7 \) MD steps followed by a production run of \( 10^7 \) steps. The data during the production cycles are collected for computing the equilibrium properties of the polymer chains. We further conduct simulations of bulk system for few selected sequence specific copolymers in a isothermal isobaric (NPT) ensemble. For NPT simulations, the temperature and pressure are maintained by the Langevin thermostat and Nose-Hoover barostat, respectively. All the bulk simulation are conducted at a reduced temperature \( T^* = 1 \), and reduced pressure \( P^* = 0 \). For each of these bulk simulations, 50 polymer chains are placed in a periodic simulation box, and the system is equilibrated for \( 10^7 \) steps followed by an production run of \( 10^7 \) steps.

Molecular Simulation based Evolutionary Computing. The central component of the sequence space screening and molecular design framework in this work is the use of an
optimization algorithm for selection of candidate copolymer sequences en route to target properties. We adopt a metaheuristic approach viz. evolutionary computing for this purpose. The evolutionary computing is a class of algorithms that aim to mimic the process of natural selection to optimize a system’s properties. These algorithms are commonly known as evolutionary algorithms (EAs) that begin with a set of initial candidates and compute an objective function quantifying their properties relative to target values. Afterword, the algorithms iteratively select new candidates in an effort to converge to a desired value of the

Figure 1: Evolutionary computing for polymer design and sequence sampling. An AB type copolymer is mapped to a binary genome where A and B moieties are represented by 1 and 0, respectively. In each cycle, evolutionary operations – selection, crossover and mutation are applied to the binary sequence to create new copolymers. The fitness of copolymers i.e. the properties are computed using molecular dynamics simulations. Molecular simulations for all the candidates in a particular generation are conducted parallelly. For all the case studies, the evolutionary cycles are carried out for about 75 generations by which highest fitness show a plateau as a function of the generation number.
objective function. The evolutionary processes usually continue until a criterion for convergence is satisfied or they are terminated because of a maximal time constraint. In these algorithms, new candidates are iteratively selected via several evolutionary operators – selection, crossover, mutation and elitism. Here, we choose one such strategy that is suitable for the design of a binary copolymer with 1:1 composition of moieties and the objective function is its radius of gyration. The schematic structure of the molecular simulation based evolutionary algorithm that is employed for copolymer design and sequence space exploration is shown in Figure 1. It begins with a random population of 48 candidate copolymers. Each copolymer’s monomer sequence is mapped to a binary genomic representation, as the model copolymer consists of only two types of chemical moieties. In the binary genome representation, 0 indicates A-type moiety and 1 indicates B-type moiety. Since, we consider equal proportion of A and B in the polymer chain, the number of 0s and 1s in the binary genome are equal in number. The fitness of all the candidates, which are their radius of gyration, is determined simultaneously by conducting multiple MD simulations in parallel. The MD simulations within the EA are conducted for sufficiently long time, and we ensure that the value of Rg converges over the simulation time. Then the top 16 out of the available 48 candidates are selected based on their ranking and moved to the next generation. These 16 candidates are the ‘parents’ who are used to produce new candidates using the crossover and mutation operations in a particular generation. In each step, two parents are randomly selected for crossover operation. The EA employs two-points crossover to combine the selected parents. Afterward, a point mutation with a rate of 0.01 is applied to generate a new candidate. In the newly generated candidates, if the ratio of two genes deviates from 1:1, the EA applies additional point mutations on the gens such that the desired ratio is maintained. During the mutation, a randomly chosen gene changes its type from 0 to 1 or vice versa (bit flip) as shown in Figure 1. The gene selection for mutation are also random. This process continues until EA produce required number of new candidates in a generation. Therefore, every new generation would again consist of 48 new candidates: 16 of which are the parents retained from the previous generation and 32 new candidates generated using the evolutionary operators. The entire evolutionary cycle is iteratively continued for around 75 generations for all the cases. The total number of generation is decided based on the criteria that the optimal fitness does not improve over more than 10 generations.
RESULTS AND DISCUSSION

Evolutionary Screening of Sequence Space. We conduct two evolutionary searches for each chain length – one for identifying optimal sequence that forms lowest radius of gyration and another one for identifying optimal sequence that forms highest radius of gyration. This strategy ensures that we sample data over the entire range of structures possible as a function of monomer sequence. Figure 2 shows the performance of the evolution algorithm for polymers of chain lengths N=20, 50, 100 and 150. For each case, we plot the objective function viz. the radius of gyration of the fittest candidate as a function of the number of candidates screened during the evolution process. For both minimization and maximization runs, there is a rapid progress at the beginning wherein the algorithm identifies better candidates in every generation. Subsequently, the search process reaches a saturation and the search is terminated when there is no improvement over more than 10 generations. The MD snapshots of the minimum and maximum radius of gyrations for a representative case of N=150 are shown in Figure 2c and

![Figure 2: The evolution of the materials design algorithm. The radius of gyration of the fittest candidate in a given generation is plotted as a function of the total number of candidate screened for minimization run and maximization run in (a) and (b), respectively. The 3D structure of polymer conformation for N=150 is shown for the two optimal sequences that correspond to lowest and highest radius of gyration in (c) and (d), respectively. The red and blue beads are A and B type moieties, respectively.](image-url)
2d, respectively. A list of all the optimized sequences, which are identified by our MD simulation based evolutionary algorithm (MD-EA) can be found in the supporting information (SI) along with their MD snapshots of equilibrium structures. For both the minimum radius of gyration and maximum radius of gyration, the sequence of A and B type monomers in the copolymer appears to be disordered or non-periodic. For a given polymer chain, a total of ~5000 sequences are evaluated during the two evolutionary computing runs. A cumulative total of 20000 sequences are studied using MD simulations within our evolutionary computing framework spanning polymer chain length from N=20 to N=150. We separately run two MD simulations with pristine diblock copolymer and alternating block copolymer for all the four chain lengths, outside the evolutionary computing workflow. These two periodic sequences – diblock and alternating block copolymers are the two extreme cases of monomer distribution in a copolymer chain, and serve as a baseline to compare the properties of all other sequences. For the alternating block copolymer, the mean block length is 1σ and for the diblock copolymer the mean block length is Nσ/2 for a polymer of chain length N. The mean block length of all other possible sequences falls within this two extreme values. We now compare the radius of gyration of the optimal candidates along with the diblock and alternating block copolymer chains in Figure 3 for all the chain lengths considered in this study. The data suggest that the range of R_g possible in a copolymer increases with the polymer chain length. For very short chain like N=20, the impact of sequence on the radius of gyration of the chain is minimal. However, for N=150, The lowest and highest R_g are 3.4σ and 6.2σ, respectively. This is over more that 80% variation in radius of gyration as a function of sequence. It also clearly indicates that the range of R_g will further expand rapidly for higher chain lengths. The R_g of the EA maximized sequence is very close to that of a diblock polymer for all the cases as shown in Figure 3. For example, the EA maximized sequence for N=150 is 71A-1B-1A-35B-1A-19B-1A-19B-1A-1B, while that of a diblock copolymer is 75A-75B. The only difference between the EA maximized sequence and diblock

![Figure 3: Range of Single molecule structure. Radius of gyration of sequence specific copolymers as function of chain length is show for four different cases – diblock copolymer, alternating block copolymer and two evolutionary algorithm identified optimal sequences.](image-url)
sequence is that four A moieties are sparsely distributed in the B dominated half of the chain. Therefore, we infer that a copolymer, whose monomer distribution is close to that of an AB type diblock polymer, will have very high radius of gyration than any other distribution of moieties. On the other hand, visual analysis of the arrangement of A and B type moieties in the EA minimized sequence reveals that the polymer consists of multiple A segments and B segments. The lengths of these segments are not equal and there is no specific order of A and B segments in the polymer topology. For example, the sequence of As and Bs in a EA minimized copolymer of chain length $N=150$ is 2B-2A-5B-11A-2B-1A-6B-2A-4B-3A-1B-1A-1B-1A-5B-3A-7B-6A-5B-1A-2B-4A-4B-1A-4B-1A-1B-3A-2B-4A-1B-3A-7B-4A-2B-1A-1B-1A-6B-10A-4B-1A-1B-1A-2B-1B-8A-1B. Evidently, the polymer is made of segments ranging from block length 1$\sigma$ to 11$\sigma$. The combination of short segments – as short as 1$\sigma$ and long segments – as long as 11$\sigma$ without any specific order leads to the highest compact structure with lowest $R_g$.

Interestingly, the $R_g$ of the EA minimized sequence differs significantly from that of an alternating block copolymer, which is $(AB)^{75}$ for $N=150$. For example, for $N=150$, the $R_g$ of the EA minimized sequence is 3.4$\sigma$, while the $R_g$ for the alternating block copolymer is 4.5$\sigma$. This is a reduction of more than 20% in $R_g$ value from that of an alternating block copolymer. Visual inspection of these EA-minimized structures reveals that these are protein-like aggregates. The core of these aggregates are formed by A type moieties, and the surface of these aggregates are formed predominantly by B type moieties.

**Sequence-Structure Correlations.** We now study the sequence-structure mapping of all the candidates. To understand this mapping of sequence space to the structure space, we calculate the fraction of sequences among the EA screened candidates, for a given chain length, that have very close $R_g$ values. Figure 4 represents the histogram of $R_g$ for two representative cases -

![Figure 4: Degeneracy of copolymer structures. Histogram of the radius of gyration that are screened during the evolutionary search for (a) $N=100$ and (b) $N=150$. The total number of candidate structures used in histogram calculations are 5000 for both (a) and (b).](image-url)
$N=100$ and $N=150$. It indicates a broad distribution of $R_g$ possible in the model system as a function of sequence. It also clearly suggests many-to-one mapping between the sequence space and structure space wherein multiple candidates in the sequence space are connected to one point in the structure space. Now, the question is why so many different sequences forming similar structures? Is there any structural descriptor/feature that are common among these sequences that lead to same $R_g$? To further understand this similarity in the 3D structures of a copolymer chain as a function of its monomer sequence, we compute the mean block length of all the sequences. The mean block length of a copolymer with $n$ blocks can be written as $L_b = \sum_i l_i / n$, where $l_i$ is the length of the $i^{th}$ block. There have been many recent efforts to correlate copolymer properties such as thermal conductivity and surface tension with its mean block length.\textsuperscript{28,29,38} Here, we aim to understand the correlation between mean block length and $R_g$ among all the sequences that are identified by the EA runs. We plot the radius of gyration of all the sequences screened during the evolutionary search as a function of their mean block length in Figure 5 for the same two representative cases $N=100$ and $N=150$. Although the exact sequences of A and B of all the candidate copolymers are different, but a subset of them have same mean block length. In fact, many of the chains, which show same $R_g$, have same mean block length $L_b$, in spite of different order of As and Bs in their architecture. However, Figure 5 also shows many exceptions, especially in the longer mean block length region. In the long mean block length region, many of the candidate sequences lead to different $R_g$s in spite of having same mean block length. Overall, the mean block length of a sequence does not show

![Figure 5: $R_g$-mean block length correlation. Radius of gyration of all the sequences screened during evolutionary computing are plotted as a function of their mean block length for chain lengths $N=100$ and $N=150$ in (a) and (b), respectively. The data points are connected by solid red line for periodic sequences as a guide to eyes. The bottom x-axis represents mean block lengths for non-periodic sequences while the top x-axis corresponds to the mean block length of the periodic sequences. The dotted perpendicular line separate two regions – The left side of the dotted line shows a modest $R_g$-$L_b$ correlation, while the right side of the line indicates a large variation of $R_g$ as a function of $L_b$.](image-url)
any strong correlation with the radius of gyration for the entire range of $R_g$ and $L_b$ that are captured in this study. However, there is a narrow range of mean block length for which $R_g$ broadly decreases as $L_b$ increases. On the other hand, copolymers with higher mean block length form very wide range of structures – both high and low $R_g$s. The perpendicular dotted line in Figure 5 qualitatively shows the cross-over of these two regions. The left side of the dotted line, we observe a qualitative inverse correlation between $L_b$ and $R_g$. On the right side of the dotted line, such correlation can’t be drawn as the variation of $R_g$ is very large. We also collect a subset of sequences that are periodic for a given chain length and added them in Figure 5. As show by circular legends and connected by solid line in Figure 5, there is a modest correlation between their radius of gyration and mean block length. Considering only the periodic sequences possible for a given chain length, we observe that the radius of gyration decreases as a function of mean block length initially, and then it increases for the further increment in mean block length. The EA-minimization run yields large number of sequences that form compact structures with lower $R_g$ than that an alternating block copolymers. These sequences are protein like sequences. The degree of “blocky-ness” in these MD-EA identified “protein-like’ sequences are higher that a random copolymer or an alternating block copolymer. Unlike any random sequences, these protein-likes sequences exhibit long range correlations which can be described by Levy-flight statistics. 

Figure 6: Variance of polymer structure. The radius of gyration of a polymer chain is plotted against its variance of block length for a fixed mean block length in (a), (b) and (c) for $L_b = 5.5\sigma$, $L_b = 12.5\sigma$ and $L_b = 8.8\sigma$,respectively. In equilibrium MD snapshots are shown in (d) and (c) for $S_2 = 76.35\sigma^2$ and $S_2 = 93.24\sigma^2$, respectively for a fixed mean block length of 5.5$\sigma$. The red and blue particles are A and B type moieties, respectively.
Now, we intend to understand the variability of the radius of gyration among the polymer chains with fixed mean block length. We compute the variance in their block lengths to understand this Rg variability. The variance of block length of a polymer chain is defined as $S^2 = \sum (l_i - L_b)^2 / (n - 1)$. Here, $l_i$ is the length of $i^{th}$ block of a polymer chain with $n$ number of blocks. We select four polymer chains that have same mean block lengths however their radius of gyrations vary significantly, for the case of $N=100$. We aim to explain the variability of the radius of gyration as a consequence of the variances of block length in the polymer chain. Figure 6a and b presents two such cases where the radius of gyration is plotted against $S^2$ for $L_b=5.5\sigma$ and $12.5\sigma$. For $L_b=5.5\sigma$ and $12.5\sigma$, there is clearly two regimes – once corresponding to low $R_g$ and the other corresponds to high $R_g$ as shown in Figure 6a and 6b, respectively. Among the four cases, one is in the lower $R_g$ regime while the remaining three are in the higher $R_g$ regime. The lower $R_g$ is seen when the block length variance of the polymer chain is low. On the other hand when the block length variance is high, we observer higher radius of gyration. Further, we choose a case where the mean block length is same and the block length variances are reasonably close. As shown in Figure 6c, the $R_g$ values of all the four candidates are similar. Thus, the combination of mean block length and block length variance provide a plausible explanation for two questions – 1. It demystifies the variability of $R_g$ among the candidates, particularly in the long $L_b$ region. These are the sequences that have same $L_b$ but significantly different $S^2$. 2. It explains why many different sequences of a copolymer lead to a single $R_g$, particularly in the short $L_b$ region. These are the sequences that have same $L_b$ as well as same $S^2$ in spite of varying the order of A and B moieties in the polymer architecture.

Figure 7: Radial distribution of monomers in a single molecule. The radial distribution function of A and B type moieties around an A type moiety is shown for Rg-minimized, alternating block, diblock and Rg-maximized sequences in (a), (b), (c) and (d), respectively. The corresponding in equilibrium molecular snapshots are shown in (e), (f), (g) and (h). Here, red and blue bead represent A and B type moieties, respectively. The red and blue particles are A and B type moieties respectively.
Sequence Defined Packing of Single Molecule: In this section, we focus on local structure of a copolymer and study the impact of sequence on the single molecular packing for a representative case of $N=100$. We calculate radial distribution function which is defined as the local density of a specific moiety around a given moiety as a function of separation distance between them. Figure 7a, b, c and d represent radial distribution function for $R_g$-minimized sequence, alternating block copolymer, diblock copolymer and $R_g$-maximized sequence, respectively. The MD snapshots of the corresponding four molecular conformations are shown in Figure 7e, f, g and h for EA-minimized sequence, alternating block copolymer, diblock copolymer and EA-maximized sequence, respectively. For each of these cases, we compute two radial distribution functions – one for the A-A pair and the second one is for the A-B pair of moieties. For the EA minimized sequence, the local density of A type moiety around an A type moiety $g(r_{AA})$ is higher than the local density of B type moiety around an A type moiety, $g(r_{AB})$, as shown by the first peak heights of the two functions in Figure 7a. However, for an alternating block copolymer the first peak of $g(r_{AB})$ is significantly higher than that of the $g(r_{AA})$. As the sequence of the alternating block is (AB)$_{50}$, which is made of AB dimers only, the distribution of B around A is higher than that of A around an A. On the other hand, the EA minimized sequence is 3B-6A-3B-4A-8A-2A-1B-3A-1B-4A-5B-1A-4B-8A-1B-1A-3B-1A-7B-10A-6B-1A-1B-4B-4A-1B-1A-2B-3A, which contains AB dimers as well as large number of pure A and B strings with varying size. A balanced combination of $g(r_{AB})$ and $g(r_{AA})$ lead to a very compact structure with lowest $R_g$ in the EA minimized sequence. For the diblock copolymer, the AA pair distribution function shows highest first peak intensity and AB pair distribution function exhibits lowest pick intensity, as shown in Figure 7c. This leads to a phase separation of the two blocks as visually seen in Figure 7g. The distribution function for EA-maximized sequence is very similar to that of the diblock case with slightly higher density for the AB pair. This is due to the fact that EA maximized sequence (47A-25B-1A-20B-1A-4B-1A-1B) has 3 A-type moieties that are sparsely placed in the B dominant half of the chain. This is

![Figure 8: Pair interaction energy of a copolymer. The per particle Lennard-Jones interaction energy of four systems – alternating block copolymer, diblock copolymer, EA maximized sequence and EA minimized sequence during production runs are shown as a function of MD time.](image)
also in line with the fact that the $R_g$ of a diblock copolymer is very close to that of the EA maximized sequence. A strong $g(r_{AA})$ and a very weak $g(r_{AB})$ lead to extended structure with highest $R_g$ in the EA maximized sequence. The packing of the polymer can be further understood by analyzing the energy of the systems. Figure 8 shows the in equilibrium Lennard-Jones energy of the system for the four cases. As the diblock forms a complete phase separation of A and B type moieties, it has maximum favourable interaction (A-A interactions), the system has lowest pair energy as shown in Figure 8. For the alternating block copolymer, A-B pair density is strongest and it accounts for maximum unfavourable interactions (A-B interactions). Therefore, the energy of the alternating block copolymer system is highest. For the two EA optimized sequences, the pair energy is in between that of a diblock and an alternating block copolymer. For the EA maximized sequence, the presence of three BAB trimers in the B dominant region of the chain leads to its slightly higher energy in comparison to that of the diblock copolymer. For the case of EA minimized sequence, the number of ABA trimers are higher than that of a diblock but lower than that of the alternating block copolymer. This leads to the total pair energy higher than that of the diblock and lower than that of the alternating block copolymers.

**Sequence Specific Packing of Copolymers in a Material.** We further study the impact of sequence in a bulk polymeric material via isothermal isobaric molecular dynamics simulations. Specifically, packing of the EA-optimized sequences along with alternating and diblock copolymers in a bulk system are tested for a representative case of polymer chain length.

![Figure 9: Copolymer packing in a bulk system. The radial distribution function for alternating block copolymer, EA minimized sequence, diblock copolymer and EA maximized sequence are shown in (a), (b), (c) and (d), respectively. For all the cases two radial distribution functions are shown – one for AA pair and the second one for AB pair, as mentioned in the legends. The in equilibrium molecular conformations are shown in (e), (f), (g) and (h) for alternating block copolymer, EA minimized sequence, diblock copolymer, EA maximized sequence, respectively. The read and blue colours represent A and B type moieties, respectively. The B type moieties are not shown in (g) and (f) for better visual inspection of the structures.](image-url)
We note that these sequences are optimized for solution. It is quite possible that the optimal sequences for bulk system are different. Here, our intention is to test whether it is possible to develop sequence specific copolymer that fall outside the known range of density of a bulk copolymer system defined by the two extremes – diblock copolymer and alternating block copolymer. Evidently, our simulation shows that the equilibrium number densities are $0.48\sigma^{-3}$, $0.52\sigma^{-3}$, $0.34\sigma^{-3}$ and $0.33\sigma^{-3}$ for alternating block copolymer, EA-minimized sequence, diblock copolymer, EA-maximized sequence, respectively. First of all, this suggests more than 50% change in the density of the system by changing the copolymer sequence from the EA-maximized sequence to the EA-minimized sequence. Secondly, the density of our EA-minimized sequence is greater than that of an alternating block copolymer. In principle, addressing the question of highest possible density in a copolymer system requires conducting optimization calculation with bulk density as the fitness function. This is not within the scope of the current study. Nonetheless, our data clearly indicates that the range of possible density in a bulk copolymer system can be expended by monomer level sequence control. We further, analyze the arrangement of A and B type moieties for these four systems. Figure 9 show the radial distribution function for all the cases along with MD snaps shots. In these bulk systems, the distribution of B type moieties around an A type moiety is similar to that observed for single molecule packing. In the alternating block copolymer system, the height of the first peak of $g(r_{AB})$ is significantly greater than that of the $g(r_{AA})$ as the system consists of AB dimers, which is very clear from Figure 9a. On the other hand, the EA-minimized sequence, the $g(r_{AA})$ peak is higher than that of the $g(r_{AB})$, as shown in Figure 9b. There is a significant aggregation of A type moieties and B type moieties separately in the system. These local domains of A moieties and B moieties form a percolating gyroid-like network structure as shown in Figure 9f. It is also significantly different form that of the alternating block copolymer where A and B type moieties are uniformly distributed in the system as shown in Figure 9e. Moving from EA-minimized sequence to a diblock copolymer, we see a complete phase separation. In the diblock system, $g(r_{AA})$ shows highest first peak, and the $g(r_{AB})$ peak height is minimal as shown in Figure 9c. This indicates a phase separation between A and B type moieties. The $g(r_{AA})$ peak height slightly reduces when the diblock polymer is replaced by the EA-maximized sequence as depicted in Figure 9d. This reduction leads to a sparse dispersion of A type moieties. The bulk density of a diblock copolymer system is close to that of the EA maximized sequence. But their microstructures have noticeable differences. As shown in Figure 9g, the diblock copolymer system forms a perfect cylindrical micelle structure wherein...
the A type moieties form a cylinder within the simulation box, and the remaining volume fraction is occupied by B type moieties. However, for the EA-maximized sequence, in addition to a cylindrical structure of A type moieties, few A moieties are sparsely distributed in the B rich domain of the simulation box as shown in Figure 9h. Therefore, this analysis suggests a wide range of new structures and functions along with well know spherical micelles, cylindrical micelles, lamellar and gyroids structures\textsuperscript{3,59–61} can be leveraged by engineering monomer level sequence of a copolymer.

CONCLUSIONS

Characterizing the impact of sequence on the 3D shape of a macromolecule is the key to understand the functions of many biological systems including protein and DNA. Moreover, an improved understanding of the impact of sequence on a molecular topology will enable a more effective control of structure-property relations in synthetic polymeric materials and open a viable pathway to tailor materials properties that are essential for miniaturized, more efficient and durable next generation devices. The central barrier in designing such materials is that the exploration of sequence space of a copolymer is challenging due to the astronomically large number of possible combinations. Therefore, efficiently generating and analyzing the big data of copolymer sequences and structures have been one of the major challenge in polymer informatics. Recent progress in machine learning and advanced optimization methods along with coarse-grained MD simulations can play a vital role for this purpose. Herein, we use one such approach viz., MD simulation based evolutionary algorithm to design sequence specific polymer and study sequence-structure relationship of a model AB type linear copolymer with equal composition. Within this framework, starting from a random copolymer, we repetitively employ evolutionary operations – selection, mutation and crossovers to identify new sequences and conduct CGMD simulations to measure their equilibrium structures. During this CGMD simulation based evolutionary computing, we determine in-equilibrium structures of more than 20000 sequences of the model copolymer, spanning across different chain lengths. The EA search spans across the two extremal points of highest and lowest radius of gyrations for all the chain lengths. We find that the optimal sequences for the lowest Rg and highest $R_g$ are non-periodic and non-intuitive for all the chain lengths. However, we note that the sequence of our EA-identified copolymer with highest Rg is very close to that of a diblock copolymer, where half of the chain is predominantly consists of A type moieties, and the remaining half of the chain is predominantly consists of B type moieties. The EA-identified copolymers with lowest
Rg yield protein-like aggregates. Identifying protein-like synthetic copolymer is always a challenging design task. Here, starting from a random sequence of AB type copolymer with equal composition and setting Rg as the fitness of an evolutionary search, we are able to track specific sequences that fold like a globular protein. We also demonstrate that the structures of a copolymer with non-periodic sequences deviate significantly from that of periodic sequences. The mean block length and the block length variance of a copolymer together provide a plausible explanation for the variation in Rg as a function of sequence. As the current study is restricted to AB type copolymer with equal proportion of A and B type moieties, future studies are required towards understanding a universal sequence-structure correlations of copolymers with varying ratios of moieties as well as the presence of side chains and other possible molecular architectures. The work illustrates the non-periodic sequences are very important and monomer-to-monomer sequence control of a random copolymer can expand the range of 3D structures possible in a copolymer system. We also test the packing of EA-identified solution phase optimal sequences in a bulk environment. It suggests that packing density of a bulk system can be enhanced by monomer level sequence specific control. It would be of further interest to optimize the density of the bulk system as a function of sequence. Over all, current study demonstrates the effectiveness of evolutionary computing for sequence engineering of macromolecules for target structures and functionalities, and reports new insights on the sequence-Rg correlation of copolymers.

SUPPORTING INFORMATION

The MD-EA identified optimal sequences of all four polymer chains and the single molecule structures of the optimal sequences.

NOTES

The authors declare no competing financial interest.

ACKNOWLEDGMENT

The work is made possible by financial support from SERB, DST, Gov of India through a start-up research grant (SRG/2020/001045) and National Supercomputing Mission’s research grant (DST/NSM/R&D_HPC_Applications/2021/40). TKP acknowledges ICSR, IIT Madras for initiation, seed and exploratory research grants. This research used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357. We also used computational facility of the Center for
Nanoscience Materials. Use of the Center for Nanoscale Materials, an Office of Science user facility, was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357. We acknowledge the use of the computing resources at HPCE, IIT Madras.

REFERENCE

(1) Bates, F. S. Polymer-Polymer Phase Behavior. Science 1991, 251 (4996), 898–905. https://doi.org/10.1126/science.251.4996.898.

(2) Chremos, A.; Nikoubashman, A.; Panagiotopoulos, A. Z. Flory-Huggins Parameter $\chi$ from Binary Mixtures of Lennard-Jones Particles to Block Copolymer Melts. J. Chem. Phys. 2014, 140 (5), 054909. https://doi.org/10.1063/1.4863331.

(3) Bates, F. S.; Fredrickson, G. H. Block Copolymer Thermodynamics: Theory and Experiment. Annu. Rev. Phys. Chem. 1990, 41 (1), 525–557. https://doi.org/10.1146/annurev.pc.41.100190.002521.

(4) Hamley, I. W. Ordering in Thin Films of Block Copolymers: Fundamentals to Potential Applications. Prog. Polym. Sci. 2009, 34 (11), 1161–1210. https://doi.org/10.1016/j.progpolymsci.2009.06.003.

(5) Cigana, P.; Favis, B. D.; Jerome, R. Diblock Copolymers as Emulsifying Agents in Polymer Blends: Influence of Molecular Weight, Architecture, and Chemical Composition. J. Polym. Sci. Part B Polym. Phys. 1996, 34 (9), 1691–1700. https://doi.org/10.1002/(SICI)1099-0488(19960715)34:9<1691::AID-POLB18>3.0.CO;2-2.

(6) Meijer, H. E. H.; Lemstra, P. J.; Elemans, P. H. M. Structured Polymer Blends. Makromol. Chem. Macromol. Symp. 1988, 16 (1), 113–135. https://doi.org/10.1002/masy.19880160109.

(7) Sundararaj, U.; Macosko, C. W. Drop Breakup and Coalescence in Polymer Blends: The Effects of Concentration and Compatibilization. Macromolecules 1995, 28 (8), 2647–2657. https://doi.org/10.1021/ma00112a009.

(8) Khokhlov, A. R.; Khalatur, P. G. Conformation-Dependent Sequence Design (Engineering) of AB Copolymers. Phys. Rev. Lett. 1999, 82 (17), 3456–3459. https://doi.org/10.1103/PhysRevLett.82.3456.

(9) Sing, C. E.; Zwanikken, J. W.; Olvera de la Cruz, M. Electrostatic Control of Block Copolymer Morphology. Nat. Mater. 2014, 13 (7), 694–698. https://doi.org/10.1038/nmat4001.

(10) Trinh, T. T.; Oswald, L.; Chan-Seng, D.; Lutz, J.-F. Synthesis of Molecurly Encoded Oligomers Using a Chemoselective “AB + CD” Iterative Approach. Macromol. Rapid Commun. 2014, 35 (2), 141–145. https://doi.org/10.1002/marc.201300774.

(11) Porel, M.; Alabi, C. A. Sequence-Defined Polymers via Orthogonal Allyl Acrylamide Building Blocks. J. Am. Chem. Soc. 2014, 136 (38), 13162–13165. https://doi.org/10.1021/ja507262t.

(12) Olivos, H. J.; Alluri, P. G.; Reddy, M. M.; Salony, D.; Kodadek, T. Microwave-Assisted Solid-Phase Synthesis of Peptoids. Org. Lett. 2002, 4 (23), 4057–4059. https://doi.org/10.1021/ol0267578.

(13) Porel, M.; Thornlow, D. N.; Phan, N. N.; Alabi, C. A. Sequence-Defined Bioactive Macrocycles via an Acid-Catalysed Cascade Reaction. Nat. Chem. 2016, 8 (6), 590–596. https://doi.org/10.1038/nchem.2508.
Dong, R.; Liu, R.; Gaffney, P. R. J.; Schaepttoens, M.; Marchetti, P.; Williams, C. M.; Chen, R.; Livingston, A. G. Sequence-Defined Multifunctional Polyethers via Liquid-Phase Synthesis with Molecular Sieving. Nat. Chem. 2019, 11 (2), 136–145. https://doi.org/10.1038/s41557-018-0169-6.

Lutz, J.-F.; Ouchi, M.; Liu, D. R.; Sawamoto, M. Sequence-Controlled Polymers. Science 2013, 341 (6146), 1238149. https://doi.org/10.1126/science.1238149.

Nanjan, P.; Porel, M. Sequence-Defined Non-Natural Polymers: Synthesis and Applications. Polym. Chem. 2019, 10 (40), 5406–5424. https://doi.org/10.1039/C9PY00886A.

Lutz, J.-F. Defining the Field of Sequence-Controlled Polymers. Macromol. Rapid Commun. 2017, 38 (24), 1700582. https://doi.org/10.1002/marc.201700582.

Neve, J. D.; Haven, J. J.; Maes, L.; Junkers, T. Sequence-Definition from Controlled Polymerization: The next Generation of Materials. Polym. Chem. 2018, 9 (38), 4692–4705. https://doi.org/10.1039/C8PY00190G.

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. https://doi.org/10.1021/acsmacrolett.0c00002.

Lutz, J.-F. Coding Macromolecules: Inputting Information in Polymers Using Monomer-Based Alphabets. Macromolecules 2015. https://doi.org/10.1021/acs.macromol.5b00890.

Yang, C.; Wu, K. B.; Deng, Y.; Yuan, J.; Niu, J. Geared Toward Applications: A Perspective on Functional Sequence-Controlled Polymers. ACS Macro Lett. 2021, 10 (2), 243–257. https://doi.org/10.1021/acsmacrolett.0c00855.

Peng, C.; Joy, A. Alternating and Random-Sequence Polyesters with Distinct Physical Properties. Polym. Chem. 2017, 8 (15), 2397–2404. https://doi.org/10.1039/C7PY00165G.

Stubbs, C.; Lipecki, J.; Gibson, M. I. Regioregular Alternating Polyampholytes Have Enhanced Biomimetic Ice Recrystallization Activity Compared to Random Copolymers and the Role of Side Chain versus Main Chain Hydrophobicity. Biomacromolecules 2017, 18 (1), 295–302. https://doi.org/10.1021/acs.biomac.6b01691.

Knight, A. S.; Zhou, E. Y.; Francis, M. B.; Zuckermann, R. N. Sequence Programmable Peptoid Polymers for Diverse Materials Applications. Adv. Mater. 2015, 27 (38), 5665–5691. https://doi.org/10.1002/adma.201500275.

Chang, L.-W.; Lytle, T. K.; Radhakrishna, M.; Madinya, J. J.; Vélez, J.; Sing, C. E.; Perry, S. L. Sequence and Entropy-Based Control of Complex Coacervates. Nat. Commun. 2017, 8 (1), 1273. https://doi.org/10.1038/s41467-017-01249-1.

Lyatskaya, Y.; Gersappe, D.; Gross, N. A.; Balazs, A. C. Designing Compatibilizers To Reduce Interfacial Tension in Polymer Blends. J. Phys. Chem. 1996, 100 (5), 1449–1458. https://doi.org/10.1021/jp952422e.

Macosko, C. W.; Guégan, P.; Khandpur, A. K.; Nakayama, A.; Marechal, P.; Inoue, T. Compatibilizers for Melt Blending: Premade Block Copolymers. Macromolecules 1996, 29 (17), 5590–5598. https://doi.org/10.1021/ma9602482.

Meenakshisundaram, V.; Hung, J.-H.; Patra, T. K.; Simmons, D. S. Designing Sequence-Specific Copolymer Compatibilizers Using a Molecular-Dynamics-Simulation-Based Genetic Algorithm. Macromolecules 2017, 50 (3), 1155–1166. https://doi.org/10.1021/acs.macromol.6b01747.

K. Patra, T.; D. Loeffler, T.; S. Sankaranarayanan, S. K. R. Accelerating Copolymer Inverse Design Using Monte Carlo Tree Search. Nanoscale 2020. https://doi.org/10.1039/D0NR06091G.
(30) Sing, C. E. Micro- to Macro-Phase Separation Transition in Sequence-Defined Coacervates. *J. Chem. Phys.* **2020**, *152* (2), 024902. https://doi.org/10.1063/1.5140756.

(31) Madinya, J. J.; Chang, L.-W.; Perry, S. L.; Sing, C. E. Sequence-Dependent Self-Coacervation in High Charge-Density Polyampholytes. *Mol. Syst. Des. Eng.* **2020**, *5* (3), 632–644. https://doi.org/10.1039/C9ME00074G.

(32) 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), 075101. https://doi.org/10.1063/1.5141095.

(33) Luo, M.; Brown, J. R.; Remy, R. A.; Scott, D. M.; Mackay, M. E.; Hall, L. M.; Epps, T. H. Determination of Interfacial Mixing in Tapered Block Polymer Thin Films: Experimental and Theoretical Investigations. *Macromolecules* **2016**, *49* (14), 5213–5222. https://doi.org/10.1021/acs.macromol.6b00946.

(34) Seo, Y.; Brown, J. R.; Hall, L. M. Effect of Tapering on Morphology and Interfacial Behavior of Diblock Copolymers from Molecular Dynamics Simulations. *Macromolecules* **2015**, *48* (14), 4974–4982. https://doi.org/10.1021/ma502309h.

(35) Patterson, A. L.; Yu, B.; Danielsen, S. P. O.; Davidson, E. C.; Fredrickson, G. H.; Segalman, R. A. Monomer Sequence Effects on Interfacial Width and Mixing in Self-Assembled Diblock Copolymers. *Macromolecules* **2020**, *53* (9), 3262–3272. https://doi.org/10.1021/acs.macromol.9b02426.

(36) Webb, M. A.; Jackson, N. E.; Gil, P. S.; Pablo, J. J. de. Targeted Sequence Design within the Coarse-Grained Polymer Genome. *Sci. Adv.* **2020**, *6* (43), eabc6216. https://doi.org/10.1126/sciadv.abc6216.

(37) Statt, A.; Kleeblatt, D. C.; Reinhart, W. F. Unsupervised Learning of Sequence-Specific Aggregation Behavior for a Model Copolymer. *ArXiv210511528 Cond-Mat* **2021**.

(38) Zhou, T.; Wu, Z.; Chilukoti, H. K.; Müller-Plathe, F. Sequence-Engineering Polyethylene–Polypropylene Copolymers with High Thermal Conductivity Using a Molecular-Dynamics-Based Genetic Algorithm. *J. Chem. Theory Comput.* **2021**. https://doi.org/10.1021/acs.jctc.1c00134.

(39) Chakraborti, N. Genetic Algorithms in Materials Design and Processing. *Int. Mater. Rev.* **2004**, *49* (3–4), 246–260. https://doi.org/10.1179/095066004225021909.

(40) Arora, V.; Bakhshi, A. K. Molecular Designing of Novel Ternary Copolymers of Donor-Acceptor Polymers Using Genetic Algorithm. *Chem. Phys.* **2010**, *373* (3), 307–312. https://doi.org/10.1016/j.chemphys.2010.06.013.

(41) Mitra, K. Genetic Algorithms in Polymeric Material Production, Design, Processing and Other Applications: A Review. *Int. Mater. Rev.* **2008**, *53* (5), 275–297. https://doi.org/10.1179/174328008X348174.

(42) Jaeger, H. M.; de Pablo, J. J. Perspective: Evolutionary Design of Granular Media and Block Copolymer Patterns. *APL Mater.* **2016**, *4* (5), 053209. https://doi.org/10.1063/1.4948270.

(43) Kasat, R. B.; Ray, A. K.; Gupta, S. K. Applications of Genetic Algorithm in Polymer Science and Engineering. *Mater. Manuf. Process.* **2003**, *18* (3), 523–532. https://doi.org/10.1081/AMP-120022026.

(44) Khaira, G. S.; Qin, J.; Garner, G. P.; Xiong, S.; Wan, L.; Ruiz, R.; Jaeger, H. M.; Nealey, P. F.; de Pablo, J. J. Evolutionary Optimization of Directed Self-Assembly of Triblock Copolymers on Chemically Patterned Substrates. *ACS Macro Lett.* **2014**, *3* (8), 747–752. https://doi.org/10.1021/mz5002349.

(45) Kansters, R. P. F.; Donald, K. J. Cluster: Searching for Unique Low Energy Minima of Structures Using a Novel Implementation of a Genetic Algorithm. *J. Chem. Theory Comput.* **2014**, *10* (12), 5729–5737. https://doi.org/10.1021/ct500744k.
(46) Deaven, D. M.; Ho, K. M. Molecular Geometry Optimization with a Genetic Algorithm. *Phys. Rev. Lett.* 1995, 75 (2), 288–291.

(47) Fornleitner, J.; Verso, F. L.; Kahl, G.; Likos, C. N. Genetic Algorithms Predict Formation of Exotic Ordered Configurations for Two-Component Dipolar Monolayers. *Soft Matter* 2008, 4 (3), 480–484. https://doi.org/10.1039/B717205B.

(48) Chua, A. L.-S.; Benedek, N. A.; Chen, L.; Finnis, M. W.; Sutton, A. P. A Genetic Algorithm for Predicting the Structures of Interfaces in Multicomponent Systems. *Nat. Mater.* 2010, 9 (5), 418–422. https://doi.org/10.1038/nmat2712.

(49) Khokhlov, A. R.; Khalatur, P. G. Protein-like Copolymers: Computer Simulation. *Phys. Stat. Mech. Its Appl.* 1998, 249 (1), 253–261. https://doi.org/10.1016/S0378-4371(97)00473-1.

(50) Chertovich, A. V.; Govorun, E. N.; Ivanov, V. A.; Khalatur, P. G.; Khokhlov, A. R. Conformation-Dependent Sequence Design: Evolutionary Approach. *Eur. Phys. J. E* 2004, 13 (1), 15–25. https://doi.org/10.1140/epje/e2004-00036-1.

(51) Wang, W.; Zhao, P.; Yang, X.; Lu, Z.-Y. Coil-to-Globule Transitions of Homopolymers and Multiblock Copolymers. *J. Chem. Phys.* 2014, 141 (24), 244907. https://doi.org/10.1063/1.4904888.

(52) Bachmann, M.; Arkin, H.; Janke, W. Multicanonical Study of Coarse-Grained off-Lattice Models for Folding Heteropolymers. *Phys. Rev. E* 2005, 71 (3), 031906. https://doi.org/10.1103/PhysRevE.71.031906.

(53) Grest, G. S.; Lacasse, M.; Kremer, K.; Gupta, A. M. Efficient Continuum Model for Simulating Polymer Blends and Copolymers. *J. Chem. Phys.* 1996, 105 (23), 10583–10594. https://doi.org/10.1063/1.472978.

(54) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* 1995, 117 (1), 1–19. https://doi.org/10.1006/jcph.1995.1039.

(55) De Jong, K. A. Evolutionary Computation: A Unified Approach; MIT Press: Cambridge, MA, USA, 2016.

(56) Vikhar, P. A. Evolutionary Algorithms: A Critical Review and Its Future Prospects. In 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC); 2016: pp 261–265. https://doi.org/10.1109/ICGTSPICC.2016.7955308.

(57) Goldberg, D. E. *Genetic Algorithms in Search, Optimization and Machine Learning*, 1st ed.; Addison-Wesley Longman Publishing Co., Inc.: Boston, MA, USA, 1989.

(58) Govorun, E. N.; Ivanov, V. A.; Khokhlov, A. R.; Khalatur, P. G.; Borovinsky, A. L.; Grosberg, A. Yu. Primary Sequences of Proteinlike Copolymers: Levy-Flight-Type Long-Range Correlations. *Phys. Rev. E* 2001, 64 (4), 040903. https://doi.org/10.1103/PhysRevE.64.040903.

(59) Park, M. J.; Balsara, N. P. Phase Behavior of Symmetric Sulfonated Block Copolymers. *Macromolecules* 2008, 41 (10), 3678–3687. https://doi.org/10.1021/ma702733f.

(60) Cochran, E. W.; Garcia-Cervera, C. J.; Fredrickson, G. H. Stability of the Gyroid Phase in Diblock Copolymers at Strong Segregation. *Macromolecules* 2006, 39 (7), 2449–2451. https://doi.org/10.1021/ma0527707.

(61) 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. https://doi.org/10.1126/science.1215368.