

AI-driven Automated Construction of Block Copolymer Phase Diagrams

Qing-Shu Dong*, Qing-Liang Song, Kun Tian, and Wei-Hua Li*

State Key Laboratory of Molecular Engineering of Polymers, Research Center of AI for Polymer Science, Key Laboratory of Computational Physical Sciences, Department of Macromolecular Science, Fudan University, Shanghai 200433, China

Abstract The self-assembly of block copolymers serves as an effective approach for fabricating various periodic ordered nanostructures. By employing self-consistent field theory (SCFT) to calculate the phase diagrams of block copolymers, one can accurately predict their self-assembly behaviors, thus providing guidance for the fabrication of various novel structures. However, SCFT is highly sensitive to initial conditions because it finds the free energy minima through an iterative process. Consequently, constructing phase diagrams using SCFT typically requires predefined candidate structures based on the experience of researchers. Such experience-dependent strategies often miss some structures and thus result in inaccurate phase diagrams. Recently, artificial intelligence (AI) techniques have demonstrated significant potential across diverse fields of science and technology. By leveraging AI methods, it is possible to reduce reliance on human experience, thereby constructing more robust and reliable phase diagrams. In this work, we demonstrate how to combine AI with SCFT to automatically search for self-assembled structures of block copolymers and construct phase diagrams. Our aim is to realize automatic construction of block copolymer phase diagrams while minimizing reliance on human prior knowledge.

Keywords Block copolymer; Self-consistent field theory; Phase diagram; Artificial intelligence; Bayesian optimization

Citation: Dong, Q. S.; Song, Q. L.; Tian, K.; Li, W. H. AI-driven automated construction of block copolymer phase diagrams. *Chinese J. Polym. Sci.* 2025, 43, 1730–1738.

INTRODUCTION

Block copolymers (BCPs) are a special class of polymers composed of two or more distinct monomers covalently bonded together. The mutual repulsion between different types of blocks, along with their covalent connectivity, endows BCPs with the ability to form a wide range of periodic ordered nanostructures. By rationally designing the molecular architecture of the BCPs, their self-assembly behavior can be precisely controlled, thereby yielding diverse nanostructures. For example, as the volume fraction of component A (f_A) increases from 0 to 0.5, AB diblock copolymers can self-assemble into various ordered structures, including the hexagonally close-packed (HCP) spherical phase, body-centered cubic (BCC) spherical phase, hexagonal cylindrical phase (C_6), double-gyroid phase (DG), and lamellar phase (Lam).^[1] For AB_n miktoarm copolymers, where n denotes the number of identical B blocks, the branched B blocks enhance the spontaneous curvature towards the A block. This enhanced spontaneous curvature significantly broadens the stability region of the spherical phases. More importantly, the Frank-Kasper σ and A15 phases can be stabilized in the enlarged

spherical phase region.^[2,3] It is essential to construct phase diagrams for different BCP systems to thoroughly investigate their self-assembly behavior and to guide the rational design of novel nanostructures. In experiments, phase diagrams can be obtained by synthesizing BCPs with varying parameters and characterizing their self-assembled structures using techniques such as small-angle X-ray Scattering (SAXS) and transmission electron microscope (TEM).^[4,5] However, experimentally constructing phase diagrams is highly labor-intensive and time-consuming. Even with the implementation of high-throughput synthesis methods,^[6,7] the process still requires considerable time and manpower.

The self-consistent field theory (SCFT) is a powerful theoretical tool for predicting the self-assembly behavior of BCPs.^[1,8–11] SCFT provides a standard framework for describing the chain architecture of BCPs and precisely calculating the free energy of various phases under different parameters, thus enabling the construction of accurate phase diagrams. SCFT results help researchers gain deep insights into the self-assembly mechanisms of block copolymers and even provide a useful guide for the design of novel nanostructured materials.^[12–19] Notably, many predictions of SCFT have been confirmed by experimental results,^[3,20–26] reinforcing its credibility and practical applicability. The general procedure for constructing phase diagrams using SCFT is as follows. First, define the chain architecture and interaction parameters of the

* Corresponding authors, E-mail: qsdong@fudan.edu.cn (Q.S.D.)
E-mail: weihuali@fudan.edu.cn (W.H.L.)

Special Topic: AI for Polymers

Received May 27, 2025; Accepted June 10, 2025; Published online September 2, 2025

target BCP system, and specify the parameter space to be explored. Next, a set of candidate self-assembled structures was selected. Then, the free energy of each candidate structure at every parameter point is computed using SCFT. Finally, the phase boundaries are determined by comparing the free energies of the different structures that comprise the phase diagram.

In the above workflow, selecting candidate self-assembled structures is a critical, yet highly experience-dependent step. The pseudo-spectral method is one of the most frequently used numerical schemes to solve SCFT equations.^[9,10] In the pseudo-spectral method for an AB system, self-assembled structures are described by the spatial density distributions of the components $\phi_A(\mathbf{r})$ and $\phi_B(\mathbf{r})$. The free energy of the system, F , is a functional of the density fields $\phi_A(\mathbf{r})$, $\phi_B(\mathbf{r})$ together with their corresponding conjugate fields $\omega_A(\mathbf{r})$, $\omega_B(\mathbf{r})$. The process of solving the SCFT equations begins with the initial conditions $\omega_{A,\text{init}}(\mathbf{r})$ and $\omega_{B,\text{init}}(\mathbf{r})$, and then proceeds iteratively to minimize the free energy and reach a self-consistent solution. According to the SCFT equations, specifying the initial conditions can be converted to assign an initial density distribution for one of the components, such as $\phi_{A,\text{init}}(\mathbf{r})$ ($\phi_{A,\text{init}}(\mathbf{r}) + \phi_{B,\text{init}}(\mathbf{r}) = 1$). In the numerical computations of SCFT, three-dimensional density distribution functions are typically represented on a grid of 32^3 – 128^3 points, meaning that approximately 10^4 – 10^6 parameters should be specified to determine the initial condition.

In the early stages, researchers commonly used random initialization^[27] in which the density at each grid point was assigned a random value as the initial condition. However, this method often fails to converge on well-ordered structures. This is because the free energy function is extremely high-dimensional and contains numerous local minima, many of which correspond to poorly ordered morphologies. In Fig. 1(a), we use a two-dimensional surface consisting of many local minima to illustrate the free energy landscape. The SCFT iteration process rarely converges to an ordered structure starting from a random point on the rugged surface (as indicated by the black arrow in Fig. 1a). To facilitate convergence to ordered structures, Li *et al.* proposed a specialized initialization scheme,^[14] in which the initial density distribution $\phi_{A,\text{init}}(\mathbf{r})$ is set based on known morphologies or artificially de-

signed structures. Special initialization can be regarded as placing the initial condition within the potential well of the local minimum of the given phase, enabling the iterative process to converge quickly to the minimum (as illustrated by the red arrow in Fig. 1a). Li *et al.* successfully obtained a variety of novel self-assembled structures using the special initialization scheme.^[14]

In addition, because any periodic function can be expanded into a Fourier series, self-assembled structures can be represented in a reciprocal space composed of a set of basis functions. In this approach, specifying the density values at each grid point is equivalent to assigning coefficients to corresponding basis functions. Accordingly, the SCFT equations can be solved in Fourier space using the spectral method.^[8] Matsen *et al.* proposed that basis functions can be grouped according to the symmetry of the space group, thereby reducing the number of parameters required.^[8] However, whether using the pseudo-spectral method or the spectral method, researchers still need to empirically set the initial conditions to converge to a specific ordered structure.^[28] Due to this reliance on empirical knowledge, previous studies primarily derived candidate structures from experimental observations, known inorganic crystal structures, and intuitive assumptions of researchers. This reliance has hindered the comprehensive exploration of candidate structures, thereby compromising the reliability of theoretical phase diagrams. In fact, even for the simplest AB diblock copolymer, its complete phase diagram was finally obtained through long-term interplay between experimental observations and theoretical calculations.^[8,13,29–32]

In recent years, the rapid development of artificial intelligence (AI) technologies has profoundly transformed numerous scientific and technological domains, including the polymer science research paradigm.^[33–42] To reduce or eliminate reliance on the prior experience of researchers, heuristic algorithms, optimization techniques, and machine learning methods have been explored for the automatic discovery of self-assembled structures.^[43–47] Tsai and Fredrickson *et al.* were the pioneer to develop such intelligent algorithms. They employed a genetic algorithm (GA) using real-space density distribution functions as genes to search for globally stable phases.^[43] This approach successfully identified structures

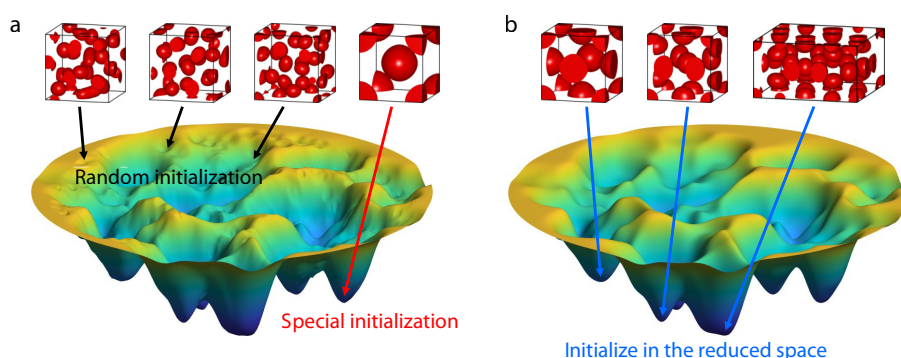


Fig. 1 Schematic illustration of search strategies for self-assembled structures. (a) Search on the original rugged free energy landscape, where the black arrow indicates the use of random initialization and the red arrow indicates the use of special initialization. (b) Search on a smoothed surface after applying the low-frequency modes approximation (blue arrow).

such as BCC, C_6 , and Lamellae. However, owing to the high dimensionality of the parameter space, it is inefficient for recovering more complex structures such as the double-gyroid phase (DG). To lower the dimensionality, Tsai and Fredrickson initialized structures using peaks distributed near a single spherical shell in reciprocal space, reducing the parameters to several peak angles and amplitudes.^[44] Building on this representation, they applied a particle swarm optimization (PSO) algorithm and successfully identified many common self-assembled structures, including DG. However, because initialization was restricted to peaks near a single spherical shell, important structural features may have been overlooked during the search process, limiting the ability of the algorithm to find complex morphologies.

Inspired by the two aforementioned approaches, as well as the spectral method of SCFT^[8] and the level set method,^[48] Dong *et al.* proposed initializing structures using multiple sets of basis functions in reciprocal space, with the basis functions grouped according to the space group symmetry. This strategy significantly reduces the dimensionality of the parameter space, while retaining the ability to represent essential structural information.^[46] Dong *et al.* combined this multimode basis function initialization scheme with Bayesian optimization and the pseudo-spectral method and successfully identified a variety of novel structures, including complex Frank-Kasper phases such as A15, Z, and σ . The initialization of structures using single or multiple basis functions can be viewed as a projection of the high-dimensional free-energy landscape onto a smoothed, lower-dimensional surface. As shown in Fig. 1(b), the smoothed free-energy landscape contains fewer local minima, which typically correspond to well-ordered structures. Consequently, the SCFT iteration process is more likely to converge to defect-free morphologies.

Unlike the optimization-based approaches described above, Chen and Dorfman developed a structure-generation method based on generative adversarial networks (GANs).^[45] The GAN framework involves the simultaneous training of two deep neural networks: a generator, which takes random noise as input to produce candidate samples, and a discriminator, which compares these samples against a real dataset to determine whether they are real or generated. After training, the generator can produce samples that closely resemble those of a real dataset. By integrating GAN with SCFT, Chen and Dorfman successfully generated a wide range of novel self-assembled structures, including many low-symmetry network morphologies. Although the types of structures currently generated by this approach are constrained by the training

dataset, we believe that as the dataset continues to grow, the GAN-SCFT method will transition from incremental improvements to qualitative breakthroughs, ultimately evolving into a powerful tool for discovering novel self-assembled structures.

In addition to developing methods for searching for self-assembled structures, many recent studies have focused on achieving the automatic construction of phase diagrams.^[49–52] For example, given a predefined set of candidate structures, approaches such as particle swarm optimization (PSO)^[53] and Bayesian optimization (BO)^[54] have been employed to automatically identify the stability region of a target phase. Zhao *et al.* employed an active learning approach to automatically select sampling points in the parameter space, thereby reducing the number of samples required for phase diagram construction.^[55] By combining these methods with automatic structure search schemes, it is possible to achieve fully automated construction of block copolymer phase diagrams.

In the following sections, we introduce the workflow for the automated construction of the phase diagrams, as illustrated in Fig. 2. First, the parameters of the block copolymer, space group, and number of basis functions were specified. Next, multiple points within the parameter space are selected for the BO-SCFT iterative searches, through which a large number of self-assembled structures can be obtained. At each phase point, several low-free-energy structures were selected as candidate phases. Subsequently, a 3D convolutional neural network is trained to classify these structures. Finally, the trained neural network was integrated with the SCFT to compute the phase diagram over the defined parameter space.

RESULTS AND DISCUSSION

Multimode Initialization Scheme

First, we introduce a multimode basis-function initialization method. For any 3D periodic function $\rho(x, y, z)$, it can be expanded as a Fourier series in the form:

$$\rho(x, y, z) = \frac{1}{V} \sum_{h,k,l=-\infty}^{\infty} F_{hkl} \exp[-i2\pi(hx + ky + lz)] = \frac{1}{V} \sum_{h,k,l=-\infty}^{\infty} |F_{hkl}| \exp[-i2\pi(hx + ky + lz) + i\theta_{hkl}] \quad (1)$$

where i is the imaginary unit defined as the square root of -1 , V is the volume in real space; (x, y, z) are the real-space coordinates, equivalent to $\mathbf{r}$ as used earlier; (h, k, l) are the coordinates in the Fourier (reciprocal) space; F_{hkl} denotes the Fourier amplitude (coefficient); and θ_{hkl} represents the corresponding phase angle. In the pseudo-spectral method of SCFT, periodic func-

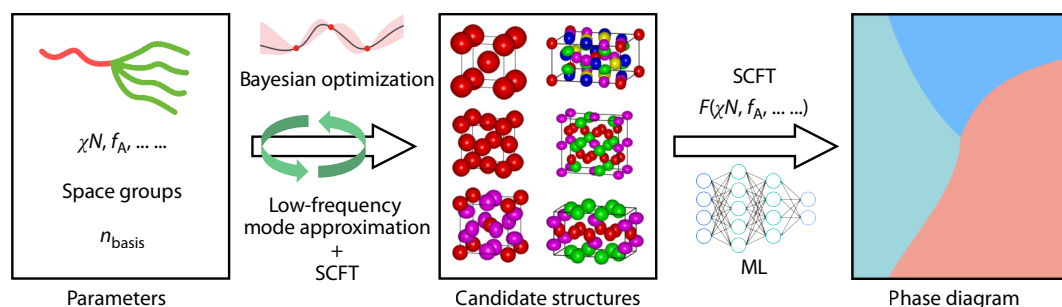


Fig. 2 Workflow for the automated construction of block copolymer phase diagrams.

tions, such as density and field variables, are represented on discrete grid points. Therefore, in Eq. (1), the Fourier series expansion can be truncated to a finite sum. This expansion does not result in any loss of information nor does it reduce the number of parameters involved.

In Fourier space, low-frequency components correspond to the primary features of the structure, whereas high-frequency components capture fine details. To initialize the structures, high-frequency components are not essential because SCFT is relatively insensitive to them during the early stages of iteration. Compared with low-frequency components, high-frequency details are more easily recovered automatically during the iterative process. Therefore, it is sufficient to retain only the first few terms in the expansion, which correspond to low-frequency modes. Moreover, the basis functions can be grouped according to the symmetry of the space group, thereby further reducing the number of parameters.

$$\rho(x, y, z) = C_1 + C_2 \times \sum_{j=1}^{n_{\text{basis}}} \sum_{hkl \in G_j} |F_{hkl}| \exp[-i2\pi(hx + ky + lz) + i\theta_{hkl}] \quad (2)$$

here, C_1 and C_2 are two automatically computed normalization parameters used to constrain the values of $\rho(x, y, z)$ within reasonable ranges, such as $[0, 1]$. G_j denotes the set of Miller indices associated with the j^{th} group of symmetry-adapted basis functions. When n_{basis} groups of basis functions are selected, the parameter space consists of n_{basis} non-negative amplitude coefficients $|F_{hkl}|$ with values in the range of $[0, 1]$, n_{basis} phase parameters θ_{hkl} with values in the range of $[0, 2\pi]$, and $1 - 6$ unit cell parameters ($a, b, c, \alpha, \beta, \gamma$) that depend on the chosen space group.

By exploiting symmetry, we only need to select basis functions with nonzero coefficients (*i.e.*, those that are not systematically extinct) according to the reflection conditions of the corresponding space group. For example, for the space group $Im\bar{3}m$ (No. 229), the reflection conditions are as follows: for permutable h, k, l ,

- $hkl : h + k + l = 2n$
- $0kl : k + l = 2n$
- $hhl : l = 2n$
- $h00 : h = 2n$

Therefore, the selected symmetry-adapted basis functions are $\{110\}$, $\{200\}$, $\{211\}$, $\{220\}$, and $\{310\}$. The reflection conditions for different space groups can be found in the International Tables for Crystallography: A.^[56]

For centrosymmetric space groups, the phase angle θ_{hkl} can only assume two discrete values: 0 or π , which corresponds to $\exp[i\theta_{hkl}] = \pm 1$. Therefore, the amplitude $|F_{hkl}|$ and phase angle θ_{hkl} can be combined into a single parameter, f_{hkl} with values ranging from -1 to 1 . Thus, a periodic function can be represented using a small number of parameters. For example, in the case of the centrosymmetric space group $Im\bar{3}m$, which belongs to the cubic crystal system, only $5 + 1 = 6$ parameters are required to represent a periodic function if five groups of basis functions are selected.

For non-centrosymmetric space groups, θ_{hkl} can, in principle, take any value in the range $[0, 2\pi]$. However, this does not necessarily mean that we need to specify n_{basis} independent amplitudes and n_{basis} independent phase angle parameters because the allowed phase angles are still constrained by

the symmetry of the space group for certain basis functions. For example, in the case of non-centrosymmetric space group $I4_132$ (No. 214), the reflection conditions are as follows: for permutable h, k, l ,

- $hkl : h + k + l = 2n$
- $0kl : k + l = 2n$
- $hhl : l = 2n$
- $h00 : h = 4n$

The first six reflected basis functions are $\{110\}$, $\{211\}$, $\{220\}$, $\{310\}$, $\{222\}$, and $\{321\}$, with phase angles $\theta_{110} = \pi/2$ or $3\pi/2$, $\theta_{211} = 0$ or π , $\theta_{220} = 0$ or π , $\theta_{310} = \pi/2$ or $3\pi/2$, $\theta_{222} = \pi/2$ or $3\pi/2$, $\theta_{321} = 0 - 2\pi$. For the five basis functions $\{110\}$, $\{211\}$, $\{220\}$, $\{310\}$, and $\{222\}$, the two possible values of θ_{hkl} differed by π , corresponding to a sign reversal in $\exp[i\theta_{hkl}]$. Therefore, θ_{hkl} and $|F_{hkl}|$ can still be combined into a single parameter f_{hkl} , with values ranging from -1 to 1 . Only θ_{321} must be independently specified. Consequently, when selecting the first six basis functions for the $I4_132$ space group, only $6 + 1 + 1 = 8$ parameters are required to represent a periodic function. Because $\exp[i(\theta_{hkl} + \pi)] = -\exp[i\theta_{hkl}]$, we can restrict the range of θ_{321} to $[0, \pi]$, which allows the corresponding parameter f_{321} to be defined within the interval $[-1, 1]$.

In summary, for all space groups, all f_{hkl} values can be defined within the range of $[-1, 1]$, and the values of all θ_{hkl} can be restricted to the interval $[0, \pi]$. This unified setting of the parameter bounds facilitates the implementation of subsequent optimization algorithms. According to the method described above, a parameter point in the variable space can be denoted by $\mathbf{X} = \{\{f_{hkl}\}, \{\theta_{hkl}\}, a, b, c, \alpha, \beta, \gamma\}$. When n_{basis} groups of basis functions are selected, the dimensionality of the parameter space lies between $n_{\text{basis}} + 1$ and $2n_{\text{basis}} + 6$. In this way, the original parameter space with a dimensionality of $10^4 - 10^6$ is reduced to a much more manageable space of the order of 10^1 .

As an example, we consider the space group $Im\bar{3}m$ to demonstrate how to generate a periodic function using Eq. (2). For simplicity, we selected the first two basis functions, $\{110\}$ and $\{200\}$. The general Wyckoff position coordinates of the $Im\bar{3}m$ space group are (x, y, z) , $(\bar{x}, \bar{y}, z)$, $(\bar{x}, y, \bar{z})$, $(x, \bar{y}, \bar{z})$, ..., as detailed in the International Tables for Crystallography: A.^[56] By substituting these coordinates into Eq. (2), we obtain the following simplified expression.

$$\rho(x, y, z) = C_1 + C_2 \times \{12 \times f_{110} \times 32 [\cos(2\pi x) \cos(2\pi y) + \cos(2\pi x) \cos(2\pi z) + \cos(2\pi y) \cos(2\pi z)] + 6 \times f_{200} \times 32 [\cos(4\pi x) + \cos(4\pi y) + \cos(4\pi z)]\} \quad (3)$$

here, 12 and 6 are multiplicity factors corresponding to the $\{110\}$ and $\{200\}$ families of crystal planes, respectively. By assigning different values to f_{110} and f_{200} , one can generate a variety of periodic functions. In addition to substituting the general Wyckoff position coordinates into Eq. (2), the corresponding expressions can also be found in the International Tables for Crystallography: B.^[57]

Fig. 3 illustrates the periodic functions generated from Eq. (3), in which f_{110} and f_{200} are assigned values of $-1, 0$, and 1 , respectively, and the isosurface level is set to 0.5 . As illustrated, using only the $\{110\}$ and $\{200\}$ basis functions can generate many distinct periodic morphologies including BCC,

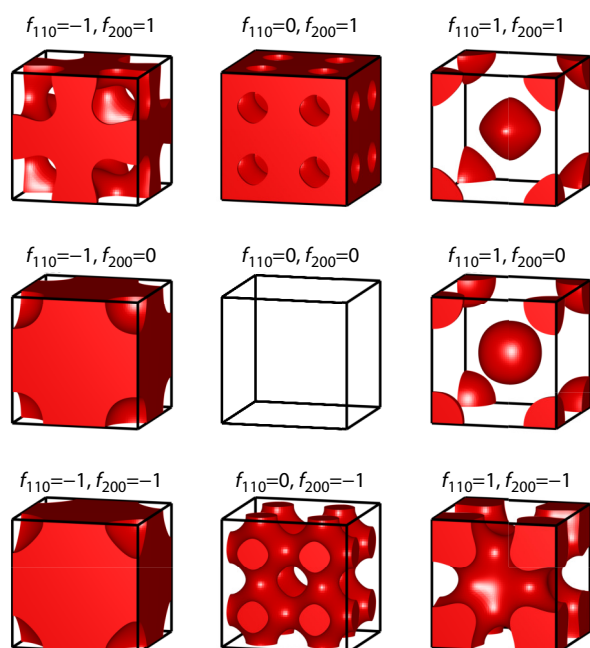


Fig. 3 Examples of periodic functions generated using the $\{110\}$ and $\{200\}$ basis functions in the $Im\bar{3}m$ space group, where the values of f_{110} and f_{200} are set to -1 , 0 , and 1 , respectively.

I-WP, and single primitive phases. Notably, simultaneously reversing the signs of both f_{110} and f_{200} leads to the formation of their corresponding inverse structures. This result is in line with our expectations because the initial structure generation does not rely on any prior assumptions regarding the target morphology.

For AB systems, once $\rho(x, y, z)$ is obtained, it can be directly used as the initial density distribution of component A, $\phi_{A,init}(r)$, while the initial density distribution of component B can be obtained via $\phi_{B,init}(r) = 1 - \phi_{A,init}(r)$. The initial condition for the SCFT iteration is then given by:

$$\omega_{A,init}(x, y, z) = \chi N \phi_{B,init}(x, y, z) \quad (4)$$

$$\omega_{B,init}(x, y, z) = \chi N \phi_{A,init}(x, y, z) \quad (5)$$

here, χN denotes the product of the Flory-Huggins interaction parameter χ and the number N of segments. It is worth noting that in standard SCFT equations, a Lagrange multiplier is introduced to enforce the incompressibility condition. However, in the present initialization procedure, the Lagrange multiplier is not required, because an incompressibility condition has already been explicitly imposed.

Searching for Candidate Structures

With the established structural initialization method, the next step is to determine the values of f_{hkl} and θ_{hkl} that enable the SCFT iteration to converge to an ordered structure or, ideally, the one with the lowest free energy. Fig. 4 presents the free energy landscape computed using SCFT, starting from structures initialized with $\{110\}$ and $\{200\}$ of the $Im\bar{3}m$ space group. Parameters used in the SCFT calculations are $\chi N = 40$ and $f_A = 0.3$ for an AB_4 block copolymer, with the initial unit cell sizes as $a = b = c = 5.0R_g$. Each color in the figure denotes a unique free-energy level, indicating a specific self-assembled structure.

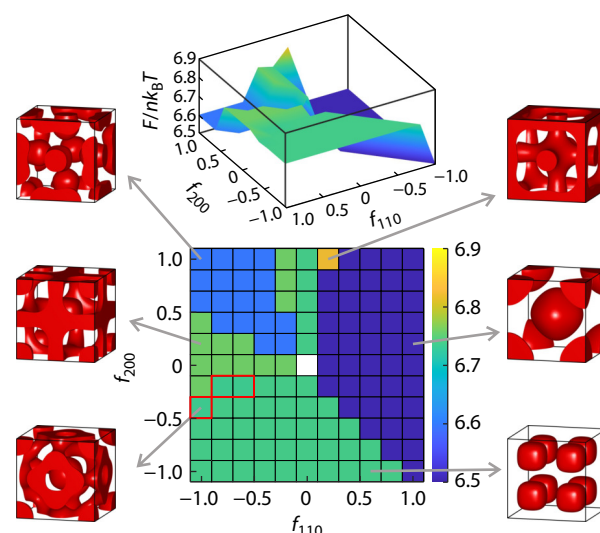


Fig. 4 Free energy landscape corresponding to the $Im\bar{3}m$ space group constructed using the 110 and 200 basis functions. The model BCP is AB_4 with $\chi N = 40$ and $f_A = 0.3$. The initial unit cell parameters are fixed as $a = b = c = 5.0R_g$. This plot can be viewed as a projected free energy surface under the low-mode approximation, corresponding to Fig. 1(b).

Starting from the initialized structures shown in Fig. 3, the SCFT iterations converged to the final morphologies depicted in Fig. 4. In certain cases, such as the BCC structure at $f_{110} = 1$ and $f_{200} = 0$, the final state matches the initial configuration. However, in several other cases, the resulting structure differs from the initial structure. Notably, the free energy minimum spans a region rather than a single point, suggesting vanishing gradients and highlighting the difficulty of identifying the global minimum.

In this study, we employed Bayesian optimization (BO)^[58] to explore the parameter space Ω , which consists of all parameter points $\mathbf{X}$. BO is a global optimization strategy aimed at locating the global minimum of a target function using a limited number of samples. Here, the target function $F(\mathbf{X})$ represents the free energy obtained after SCFT convergence from an initial configuration at $\mathbf{X}$. BO approximates $F(\mathbf{X})$ with a surrogate model and suggests the next sampling point $\mathbf{X}_{next}$ by balancing the exploration and exploitation through an acquisition function. Repeating this process leads to the identification of the optimal parameter $\mathbf{X}^*$ that minimizes free energy. Importantly, a number of local minima were also obtained through this process, each associated with a metastable ordered structure. Gaussian Process Regression (GPR) was employed as the surrogate model.^[58,59] The conventional Lower Confidence Bound (LCB) acquisition function is modified by introducing a distance-based penalty term, which encourages the BO algorithm to sample points farther from currently known data.^[46] This modification effectively imposes tabu behavior, enhancing the exploratory capability of BO. The modified LCB acquisition function is given by

$$\alpha_{LCB}(\mathbf{X}) = -\mu(\mathbf{X}) + w \times \sigma(\mathbf{X}) + w_d \times \left[1 - \tanh\left(\frac{d - \sigma_d}{\lambda_d}\right) \right] / 2 \quad (6)$$

where, $\mu(\mathbf{X})$ and $\sigma(\mathbf{X})$ denote the mean and standard deviation

tion, respectively, predicted by the GPR model; w is the weight parameter that balances exploration and exploitation; d represents the minimum distance from the current point $\mathbf{X}$ to all previously sampled points; w_d , σ_d and λ_d are the hyperparameters that control the distance penalty. Here, we set $w = 10$, $w_d = -10$, $\sigma_d = 0.1$, and $\lambda_d = 0.1$. It should be noted that various alternatives to BO exist for exploring the parameter space, including GA,^[43] PSO,^[44] and grid search. Nevertheless, this study did not aim to evaluate or compare the performance.

Combining the multimode basis function initialization method with BO and the pseudo-spectral method of SCFT enables the automated discovery of self-assembled morphologies. First, the block copolymer parameters, target space group, and number of basis functions, n_{basis} are set. GPR was then used to model the free energy landscape, providing uncertainty estimates based on prior evaluations. The acquisition function suggests the next sampling point $\mathbf{X}_{\text{next}}$ by balancing exploration and exploitation. Then, the initial fields $\omega_{A,\text{init}}$ and $\omega_{B,\text{init}}$ are generated using the basis functions corresponding to $\mathbf{X}_{\text{next}}$, and the SCFT is applied to compute the free energy $F(\mathbf{X})$. If the SCFT converges, the result is added to the dataset; otherwise, a penalized high value is assigned to $F(\mathbf{X})$. This loop continues until the maximum number of iterations is reached, yielding a collection of self-assembled structures and their free energies. A detailed implementation can be found in our open-source GitHub repository: <https://github.com/qsdong/BOSearchStructures>.

We demonstrate the automated search for self-assembled structures and phase diagram construction using the AB₄ miktoarm copolymer as an example. The considered phase space spans $\chi N = 20 - 40$ and $f_A = 0.2 - 0.4$. According to previous studies, this region includes both spherical and cylindrical structures. Multiple phase points within this region were sampled to ensure the discovery of diverse morphological features. Specifically, we uniformly selected three values along each axis of the phase diagram, namely $\chi N = 20, 30, 40$ and $f_A = 0.2, 0.3, 0.4$, resulting in nine parameter points. Among them, (20,0.2), (20,0.3), and (30,0.2) correspond to disordered phases and were excluded from the search.

As a demonstration, we selected the following three-dimensional space groups: $la\bar{3}d$, $Im\bar{3}m$, $Fd\bar{3}m$, $Fm\bar{3}m$, $Pn\bar{3}m$, $Pm\bar{3}n$, $Pm\bar{3}m$, $I4_132$, $P23$, $P6_3/mmc$, $P6/mmm$, $P4_2/mnm$, and the following two-dimensional plane groups: $p6mm$, $p6$, $p31m$, $p3m1$, $p3$, $p4gm$, $p4mm$, $p4$. It is important to note that for space groups $P6_3/mmc$, $P6/mmm$, and $P4_2/mnm$, the c/a ratio is not fixed, and as the value of c/a varies, the ordering of the basis functions changes accordingly. To account for this, we selected basis functions with three representative values: $c/a = 0.5, 1.0$, and 2.0 . In our previous work, we found that 24 basis functions were required to successfully recover the σ phase. Therefore, in this study, we fixed the number of basis functions at $n_{\text{basis}} = 24$, which is sufficient to capture highly complex morphologies. For each space group, 128 iterations of the BO-SCFT loop were performed.

Fig. 5 presents a representative search result for the $Im\bar{3}m$ space group with $\chi N = 40$ and $f_A = 0.2$. In Fig. 5, the horizontal axis denotes the number of BO-SCFT iterations, while the vertical axis shows the corresponding free energy. The BCC structure, which is known to be the most stable phase for this space group, was identified within only 20 iterations, demonstrating the high efficiency of the algorithm. Additionally, FCC has been observed multiple times, suggesting that certain symmetry elements of the FCC structure are also included in the selected basis functions. Numerous unknown complex spherical structures have been discovered. Although the BO-SCFT algorithm successfully identified the BCC structure within the first 20 iterations, it continued to explore the parameter points associated with higher free energies in the subsequent steps. This behavior stems from the design of our algorithm, which aims to discover as many self-assembled structures as possible rather than focusing solely on the one with the lowest free energy. To encourage such an exploration, we set the weight parameter to $w = 10$ and introduced a distance penalty term, as described in Eq. (6).

Once the BO-SCFT searches over all space groups are completed for each phase point, the resulting structures are sorted according to the free energy at each point. Structures with periods exceeding $15R_g$ or unit cell length ratios greater than 3 were excluded, as such configurations are rarely observed in

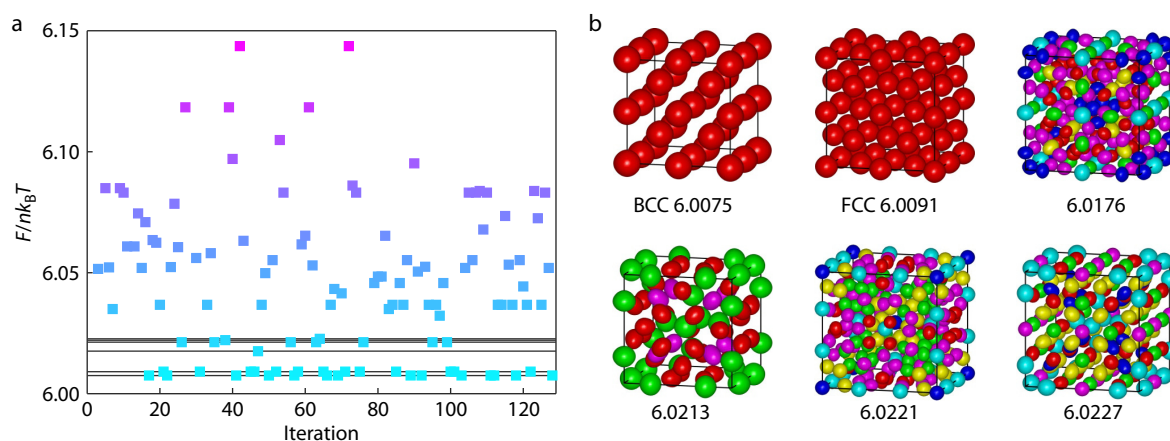


Fig. 5 Typical BO-SCFT search for $Im\bar{3}m$ space group with $\chi N = 40$ and $f_A = 0.2$. (a) Evolution of free energy during the searching process; (b) Representative structures obtained during the search process, with the corresponding free energy values shown below each structure. Non-equivalent domains are shown with different colors.

AB₄ block copolymers. The 10 structures with the lowest free energy were then selected at each phase point, yielding a total of 60 candidates. Because the searches were conducted at different values of χN and f_A , some of the selected candidate structures emerged as geometrically identical, differing only in their unit cell dimensions. These duplicates were removed and only one representative was preserved. Additionally, some structures exhibit hybrid morphologies in which the A component forms two distinct domains simultaneously, for example, a combination of spheres and networks. It is important to note that AB₄ contains only a single A block, which typically does not favor the formation of stable hybrid structures. For simplicity, such hybrid structures were excluded. Note that this step is just to reduce the computational time, but is not required. In practical applications, these hybrid morphologies should be considered among the candidate structures, as they may be stabilized in systems such as ABAB multiblock copolymers.^[16,19]

Following the above criteria, we retained 16 candidate structures, as shown in Fig. 6. These include well-known morphologies such as BCC, FCC, HCP, and C₆, as well as several Frank-Kasper phases, including A15, σ , C14, C15, and Z. It is worth noting that commonly observed structures such as DG and Lam were not included among the final candidates, as their stability regions lie outside the selected phase region. Importantly, the selected structures successfully covered all known stable phases within the chosen phase region. While most of these phases have been well documented in previous studies, it is important to emphasize that their selection was entirely automated based on our predefined rules, without relying on any prior structural knowledge. This result highlights the effectiveness of the proposed method.

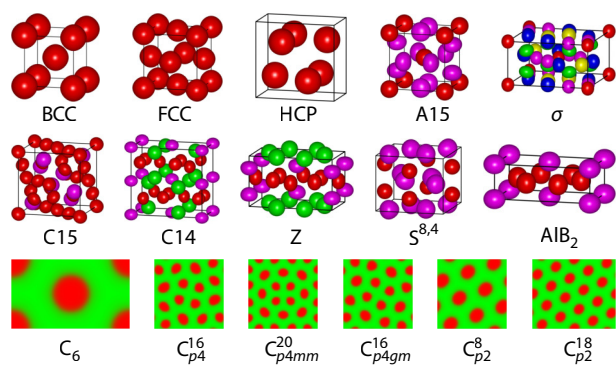


Fig. 6 Automatically selected candidate structures.

Constructing the Phase Diagram

After obtaining the candidate structures, we developed a machine-learning model for structure recognition. Consistent with our previous study,^[54] we first applied FFT low-pass filtering to preprocess the data in order to unify the number of grid points for all structures. The preprocessed data were then used to train a three-dimensional convolutional neural network (FFT-3DCNN), which enabled the automatic classification of SCFT-converged structures. A detailed implementation can be found in our open-source GitHub repository: <https://github.com/qsdong/FFT3DCNN>.

By integrating the pseudo-spectral SCFT and FFT-3DCNN,

we automatically constructed the phase diagram of the AB₄, as shown in Fig. 7. We adopt a grid scanning approach, with χN ranging from 20 to 40 with a step size of 4, and f_A from 0.2 to 0.4 with a step size of 0.01. At each parameter point, the free energy of each candidate structure was computed using SCFT. The FFT-3DCNN classifier was applied to verify whether each structure converged correctly. The structure with the lowest free energy among the valid results is considered to be the stable phase and is plotted as a point in the phase diagram. In Fig. 7, the colored dots denote automatically computed points, and the gray lines represent manually drawn phase boundaries. Different colors indicate different self-assembled phases. As shown in the figure, the automated results were consistent with the manually drawn phase diagram, including the stability regions of the disordered, HCP, σ , A15, and C₆ phases.

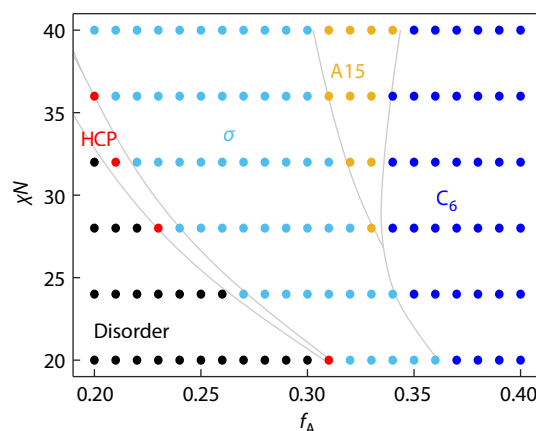


Fig. 7 Automatically constructed phase diagram, with colored points denoting automatically computed results and gray lines representing manually calculated phase boundaries.

Notably, in the work by Xie *et al.*,^[3] the focus was primarily on the stability of the Frank-Kasper phases. Because FCC and HCP are both close-packed structures with small free energy differences, only the FCC phase was considered in the phase diagram. In contrast, the candidate structures in the current study were selected automatically based on predefined criteria. As a result, both the FCC and HCP phases were included in the automatically selected candidates. Notably, a stable region corresponding to the HCP phase was identified in our automatically constructed phase diagram, which was consistent with the manually drawn phase boundaries. This demonstrates that our method effectively reduced the subjectivity involved in the manual selection of candidate structures.

CONCLUSIONS

In this study, we demonstrated the automatic construction of the phase diagram of block copolymers with minimal reliance on human prior knowledge. First, we introduce a multimode basis function initialization method that significantly reduces the dimensionality of the parameter space. We then used Bayesian optimization (BO) to explore the basis function coefficient space and integrated it with self-consistent field theory (SCFT) to automatically discover novel self-assembled morphologies. Subse-

quently, we constructed an FFT low-pass filtering-assisted 3D convolutional neural network (FFT-3DCNN) to classify the structures.

Taking AB₄ miktoarm copolymer as an example, we illustrate the complete workflow for automatic phase diagram construction. First, multiple parameter points are uniformly selected within the phase diagram to search for self-assembled structures. The resulting structures were sorted by the free energy, and the top 10 lowest-free-energy structures at each point were retained as candidates. These automatically selected candidates were then used to train FFT-3DCNN to classify the structures. Finally, the trained FFT-3DCNN was integrated with SCFT to compute the phase diagram over the defined parameter space. The resulting scatter plot of the stable phases closely matched the manually constructed phase diagram, thus validating the accuracy of our automated approach. This framework significantly reduces the dependence on human experience while greatly improving both the efficiency and reliability of the phase diagram construction.

Conflict of Interests

The authors declare no interest conflict.

Data Availability Statement

The related data for this paper are available from the corresponding author upon reasonable request.

ACKNOWLEDGMENTS

This work was financially supported by the National Natural Science Foundation of China (Nos. 52394272, 22333002, 22203018 and 22303017).

REFERENCES

- Matsen, M. W. Effect of architecture on the phase behavior of AB-type block copolymer melts. *Macromolecules* **2012**, *45*, 2161–2165.
- Grason, G. M.; Kamien, R. D. Interfaces in diblocks: A study of miktoarm star copolymers. *Macromolecules* **2004**, *37*, 7371–7380.
- Xie, N.; Li, W. H.; Qiu, F.; Shi, A. C. σ Phase formed in conformationally asymmetric AB-type block copolymers. *ACS Macro Lett.* **2014**, *3*, 906–910.
- Barbon, S. M.; Song, J.-A.; Chen, D.; Zhang, C.; Lequeieu, J.; Delaney, K. T.; Anastasaki, A.; Rolland, M.; Fredrickson, G. H.; Bates, M. W.; Hawker, C. J.; Bates, C. M. Architecture effects in complex spherical assemblies of (AB)_n-type block copolymers. *ACS Macro Lett.* **2020**, *9*, 1745–1752.
- Liu, X. Y.; Yan, X. Y.; Liu, Y.; Qu, H.; Wang, Y.; Wang, J.; Guo, Q. Y.; Lei, H.; Li, X. H.; Bian, F.; Cao, X. Y.; Zhang, R.; Wang, Y.; Huang, M.; Lin, Z.; Meijer, E. W.; Aida, T.; Kong, X.; Cheng, S. Z. D. Self-assembled soft alloy with Frank–Kasper phases beyond metals. *Nat. Mater.* **2024**, *23*, 570–576.
- Zhang, C.; Bates, M. W.; Geng, Z. S.; Levi, A. E.; Vigil, D.; Barbon, S. M.; Loman, T.; Delaney, K. T.; Fredrickson, G. H.; Bates, C. M.; Whittaker, A. K.; Hawker, C. J. Rapid generation of block copolymer libraries using automated chromatographic separation. *J. Am. Chem. Soc.* **2020**, *142*, 9843–9849.
- Murphy, E. A.; Skala, S. J.; Kottage, D.; Kohl, P. A.; Li, Y.; Zhang, C.; Hawker, C. J.; Bates, C. M. Accelerated discovery and mapping of block copolymer phase diagrams. *Phys. Rev. Mater.* **2024**, *8*, 15602.
- Matsen, M. W.; Schick, M. Stable and unstable phases of a diblock copolymer melt. *Phys. Rev. Lett.* **1994**, *72*, 2660–2663.
- Tzeremes, G.; Rasmussen, K. Ø.; Lookman, T.; Saxena, A. Efficient computation of the structural phase behavior of block copolymers. *Phys. Rev. E* **2002**, *65*, 41806.
- Arora, A.; Qin, J.; Morse, D. C.; Delaney, K. T.; Fredrickson, G. H.; Bates, F. S.; Dorfman, K. D. Broadly accessible self-consistent field theory for block polymer materials discovery. *Macromolecules* **2016**, *49*, 4675–4690.
- Qiang, Y. C.; Li, W. H. Accelerated pseudo-spectral method of self-consistent field theory via crystallographic fast fourier transform. *Macromolecules* **2020**, *53*, 9943–9952.
- Grason, G. M.; DiDonna, B. A.; Kamien, R. D. Geometric theory of diblock copolymer phases. *Phys. Rev. Lett.* **2003**, *91*, 58304.
- Tyler, C. A.; Morse, D. C. Orthorhombic Fddd network in triblock and diblock copolymer melts. *Phys. Rev. Lett.* **2005**, *94*, 208302.
- Xie, N.; Liu, M. J.; Deng, H. L.; Li, W. H.; Qiu, F.; Shi, A. C. Macromolecular metallurgy of binary mesocrystals via designed multiblock terpolymers. *J. Am. Chem. Soc.* **2014**, *136*, 2974–2977.
- Liu, M. J.; Qiang, Y. C.; Li, W. H.; Qiu, F.; Shi, A. C. Stabilizing the Frank-Kasper phases via binary blends of AB diblock copolymers. *ACS Macro Lett.* **2016**, *5*, 1167–1171.
- Zhao, B.; Liu, M. J.; Wang, C.; Chen, Y. C.; Xu, Y. C. Enlarged stable phase region of hybrid lamella-sphere phase enabled by A₁B₁A₂(B₂)_m branched tetrablock copolymer. *Chinese J. Polym. Sci.* **2021**, *39*, 925–933.
- Park, S. J.; Bates, F. S.; Dorfman, K. D. Single gyroid in H-shaped block copolymers. *Phys. Rev. Mater.* **2023**, *7*, 105601.
- Dong, Q. S.; Li, L. Y.; Xu, Z. W.; Li, W. H. Useful strategy for assuming complex ordered phases possibly formed by ABC-type block copolymers. *Phys. Rev. Mater.* **2024**, *8*, 25601.
- Zhao, B.; Wang, C.; Li, W. H.; Qiang, Y. C. Laves phases emerge in neat AB-type block copolymer as hybrid spherical phases. *ACS Macro Lett.* **2025**, 721–726.
- Lee, S.; Bluemle, M. J.; Bates, F. S. Discovery of a Frank-Kasper sigma phase in sphere-forming block copolymer melts. *Science* **2010**, *330*, 349–353.
- Schulze, M. W.; Lewis, R. M.; Lettow, J. H.; Hickey, R. J.; Gillard, T. M.; Hillmyer, M. A.; Bates, F. S. Conformational asymmetry and quasicrystal approximants in linear diblock copolymers. *Phys. Rev. Lett.* **2017**, *118*, 207801.
- Duan, C.; Zhao, M. T.; Qiang, Y. C.; Chen, L.; Li, W. H.; Qiu, F.; Shi, A.-C. Stability of two-dimensional dodecagonal quasicrystalline phase of block copolymers. *Macromolecules* **2018**, *51*, 7713–7721.
- Miyamori, Y.; Suzuki, J.; Takano, A.; Matsushita, Y. Periodic and aperiodic tiling patterns from a tetrablock terpolymer system of the A₁BA₂C type. *ACS Macro Lett.* **2020**, *9*, 32–37.
- Matsushita, Y.; Ohta, M.; Suzuki, N.; Suzuki, M.; Kitahara, A.; Takano, A. Nearly ideal random tiling with dodecagonal symmetry from pentablock quarterpolymers of the AB₁CB₂D type. *ACS Nano* **2024**, *18*, 17135–17144.
- Li, C. C.; Dong, Q. S.; Li, W. H. Largely tunable asymmetry of phase diagrams of A(AB)_n miktoarm star copolymer. *Macromolecules* **2020**, *53*, 10907–10917.
- Seo, Y.; Woo, D.; Li, L. Y.; Li, W. H.; Kim, J. K. Phase behavior of PS-(PS-*b*-P2VP)₃ miktoarm star copolymer. *Macromolecules* **2021**, *54*, 7822–7829.
- Drolet, F.; Fredrickson, G. H. Combinatorial screening of complex block copolymer assembly with self-consistent field theory. *Phys.*

- Rev. Lett.* **1999**, 83, 4317–4320.
- 28 Jiang, K.; Huang, Y. Q.; Zhang, P. W. Spectral method for exploring patterns of diblock copolymers. *J. Comput. Phys.* **2010**, 229, 7796–7805.
 - 29 Leibler, L. Theory of microphase separation in block copolymers. *Macromolecules* **1980**, 13, 1602–1617.
 - 30 Vavasour, J. D.; Whitmore, M. D. Self-consistent mean field-theory of the microphases of diblock copolymers. *Macromolecules* **1992**, 25, 5477–5486.
 - 31 Hajduk, D. A.; Harper, P. E.; Gruner, S. M.; Honeker, C. C.; Kim, G.; Thomas, E. L.; Fetters, L. J. The gyroid: a new equilibrium morphology in weakly segregated diblock copolymers. *Macromolecules* **1994**, 27, 4063–4075.
 - 32 Bailey, T. S.; Hardy, C. M.; Epps, T. H.; Bates, F. S. A noncubic triply periodic network morphology in poly(isoprene-*b*-styrene-*b*-ethylene oxide) triblock copolymers. *Macromolecules* **2002**, 35, 7007–7017.
 - 33 Kim, C.; Chandrasekaran, A.; Huan, T. D.; Das, D.; Ramprasad, R. Polymer genome: a data-powered polymer informatics platform for property predictions. *J. Phys. Chem. C* **2018**, 122, 17575–17585.
 - 34 Wei, D.; Zhou, T. J.; Huang, Y. Q.; Jiang, K. A multi-category inverse design neural network and its application to diblock copolymers. *Mathematics* **2022**, 10, 4451.
 - 35 Liu, Z. H.; Liu, Y.-X.; Yang, Y. L.; Li, J. F. Template design for complex block copolymer patterns using a machine learning method. *ACS Appl. Mater. Interfaces* **2023**, 15, 31049–31056.
 - 36 Rebello, N. J.; Arora, A.; Mochigase, H.; Lin, T.-S.; Shi, J.; Audus, D. J.; Muckley, E. S.; Osmani, A.; Olsen, B. D. The block copolymer phase behavior database. *J. Chem. Inf. Model.* **2024**, 64, 6464–6476.
 - 37 Tran, H.; Gurnani, R.; Kim, C.; Pilania, G.; Kwon, H.-K.; Lively, R. P.; Ramprasad, R. Design of functional and sustainable polymers assisted by artificial intelligence. *Nat. Rev. Mater.* **2024**, 9, 866–886.
 - 38 Zhang, K. H.; Gong, X. R.; Jiang, Y. Machine learning in soft matter: from simulations to experiments. *Adv. Funct. Mater.* **2024**, 34, 2315177.
 - 39 Gao, L.; Lin, J. P.; Wang, L. Q.; Du, L. Machine learning-assisted design of advanced polymeric materials. *Acc. Mater. Res* **2024**, 5, 571–584.
 - 40 Qiu, H. K.; Sun, Z. Y. On-demand reverse design of polymers with PolyTAO. *Npj Comput. Mater* **2024**, 10, 273.
 - 41 Qiu, H. K.; Liu, L. Y.; Qiu, X. P.; Dai, X. M.; Ji, X. L.; Sun, Z. Y. PolyNC: a natural and chemical language model for the prediction of unified polymer properties. *Chem. Sci.* **2024**, 15, 534–544.
 - 42 Hu, W. L.; Jing, E. Z.; Qiu, H. K.; Sun, Z. Y. Discovering polyimides and their composites with targeted mechanical properties through explainable machine learning. *J. Mater. Inform.* **2025**, 5, 1.
 - 43 Tsai, C. L.; Delaney, K. T.; Fredrickson, G. H. Genetic algorithm for discovery of globally stable phases in block copolymers. *Macromolecules* **2016**, 49, 6558–6567.
 - 44 Tsai, C. R. L.; Fredrickson, G. H. Using particle swarm optimization and self-consistent field theory to discover globally stable morphologies of block copolymers. *Macromolecules* **2022**, 55, 5249–5262.
 - 45 Chen, P. Y.; Dorfman, K. D. Gaming self-consistent field theory: generative block polymer phase discovery. *Proc. Natl. Acad. Sci. U. S. A.* **2023**, 120, e2308698120.
 - 46 Dong, Q. S.; Xu, Z. W.; Song, Q. L.; Qiang, Y. C.; Cao, Y.; Li, W. H. Automated search strategy for novel ordered structures of block copolymers. *ACS Macro Lett.* **2024**, 13, 987–993.
 - 47 Dorfman, K. D. Computational phase discovery in block polymers. *ACS Macro Lett.* **2024**, 13, 1612–1619.
 - 48 Wohlgemuth, M.; Yufa, N.; Hoffman, J.; Thomas, E. L. Triply periodic bicontinuous cubic microdomain morphologies by symmetries. *Macromolecules* **2001**, 34, 6083–6089.
 - 49 Dai, C. Y.; Glotzer, S. C. Efficient phase diagram sampling by active learning. *J. Phys. Chem. B* **2020**, 124, 1275–1284.
 - 50 Aoyagi, T. Deep learning model for predicting phase diagrams of block copolymers. *Comput. Mater. Sci.* **2021**, 188, 110224.
 - 51 Zhang, Y. C.; Huang, W. L.; Liu, Y. X. Automated identification of ordered phases for simulation studies of block copolymers. *Chin. J. Polym. Sci.* **2024**, 42, 683–692.
 - 52 Zhang, Y. C.; Huang, W. L.; Liu, Y. X. Automated chain architecture screening for discovery of block copolymer assembly with graph enhanced self-consistent field theory. *Commun. Mater.* **2024**, 5, 266.
 - 53 Khadilkar, M. R.; Paradiso, S.; Delaney, K. T.; Fredrickson, G. H. Inverse design of bulk morphologies in multiblock polymers using particle swarm optimization. *Macromolecules* **2017**, 50, 6702–6709.
 - 54 Dong, Q. S.; Gong, X. R.; Yuan, K. R.; Jiang, Y.; Zhang, L. S.; Li, W. H. Inverse design of complex block copolymers for exotic self-assembled structures based on Bayesian optimization. *ACS Macro Lett.* **2023**, 12, 401–407.
 - 55 Zhao, S. C.; Cai, T. Y.; Zhang, L. S.; Li, W. H.; Lin, J. P. Autonomous construction of phase diagrams of block copolymers by theory-assisted active machine learning. *ACS Macro Lett.* **2021**, 10, 598–602.
 - 56 Hahn, T. *International Tables for Crystallography, Volume A: Space-Group Symmetry*, 5th ed.; Springer, **2005**, p. 92–715.
 - 57 Shmueli, U. *International Tables for Crystallography, Volume B: Reciprocal Space*, 2nd ed.; Springer, **2001**, p. 120–149.
 - 58 Garnett, R. *Bayesian Optimization*; Cambridge University Press: Cambridge, **2023**, p. 15.
 - 59 Williams, C. K.; Rasmussen, C. E. *Gaussian Processes for Machine Learning*; MIT press Cambridge, MA: London, **2006**, p. 1.