

Automated Identification of Ordered Phases for Simulation Studies of Block Copolymers

Yu-Chen Zhang, Wei-Ling Huang, and Yi-Xin Liu*

State Key Laboratory of Molecule Engineering of Polymers, Department of Macromolecular Science, Fudan University, Shanghai 200433, China

 Electronic Supplementary Information

Abstract In unit cell simulations, identification of ordered phases in block copolymers (BCPs) is a tedious and time-consuming task, impeding the advancement of more streamlined and potentially automated research workflows. In this study, we propose a scattering-based automated identification strategy (SAIS) for characterization and identification of ordered phases of BCPs based on their computed scattering patterns. Our approach leverages the scattering theory of perfect crystals to efficiently compute the scattering patterns of periodic morphologies in a unit cell. In the first stage of the SAIS, phases are identified by comparing reflection conditions at a sequence of Miller indices. To confirm or refine the identification results of the first stage, the second stage of the SAIS introduces a tailored residual between the test phase and each of the known candidate phases. Furthermore, our strategy incorporates a variance-like criterion to distinguish background species, enabling its extension to multi-species BCP systems. It has been demonstrated that our strategy achieves exceptional accuracy and robustness while requiring minimal computational resources. Additionally, the approach allows for real-time expansion and improvement to the candidate phase library, facilitating the development of automated research workflows for designing specific ordered structures and discovering new ordered phases in BCPs.

Keywords Block copolymer; Phase identification; Scattering function

Citation: Zhang, Y. C.; Huang, W. L.; Liu, Y. X. Automated identification of ordered phases for simulation studies of block copolymers. *Chinese J. Polym. Sci.* 2024, 42, 683–692.

INTRODUCTION

Over the past few decades, block copolymers (BCPs) have gained significant attention and found potential applications in various fields, including materials science,^[1] chemistry,^[2] micro-nano electronics,^[3] and biomedical engineering.^[4,5] The remarkable potential of BCPs lies in their diverse, complex, and customizable ordered phases. Even two-species BCPs can spontaneously assemble into a wide range of phases,^[6] such as packing of cylinders, cubic lattices of spheres^[7–9] (face-centered cubic, A15 phase, etc.), triply periodic minimal surfaces^[10–12] (double gyroid, double diamond, etc.) and so on. Understanding and characterizing these ordered phases is essential for designing functional materials with desired properties.

Theoretical and experimental studies have led to significant advancements in understanding ordered phases of BCPs. When studying the equilibrium phase behavior and phase diagrams, one of the most successful theoretical methods is the unit cell self-consistent field theory (SCFT) calculation.^[13,14] The unit cell calculations ensure the converged structures from SCFT calculations being perfectly ordered. In recent decades, the application of unit cell SCFT calculations has fa-

cilitated the interpretation and prediction of diverse ordered phases in various BCP systems.^[6,15–17] New stable phases continue to emerge even in the simplest diblock copolymers, such as Frank-Kasper σ ,^[18] A15^[19,20] and so forth. Recent studies have further revealed how the compositional and topological manipulations provide opportunities for designing materials with complex morphologies.^[21–23]

During computation of a phase diagram, the thermodynamic equilibrium phase at a particular point in the parameter space is determined by calculating and comparing the free energies of multiple candidate phases.^[24] Due to the continuous discovery of stable phases, more and more candidate phases have to be computed. Typically, the construction of a single phase diagram requires thousand times of free energy calculations. Thus, it is of great interest to automate the phase diagram calculation as much as possible. However, the primary bottleneck hindering the automation of such process lies in the identification and confirmation of phases being computed due to the fact that SCFT calculations often converge to structures that are different from the one given initially. In addition, manual identification of ordered phases relies on visual inspection of the converged density distributions which is hard to distinguish subtle differences between perfect ordered structures and slightly distorted ordered structures. Therefore, it is highly desirable to develop a reliable identification method to automate the calculation of free energies,

* Corresponding author, E-mail: lyx@fudan.edu.cn

Received November 9, 2023; Accepted December 19, 2023; Published online January 26, 2024

thereby improve the efficiency of phase diagram calculations.

In recent years, efforts have been made to achieve automated phase identification of ordered structures obtained from SCFT calculations, thereby providing valuable feedback for simulation workflows. Reported methods include calculating the variance of density distributions between unknown structures and the target structure as well as utilizing neural networks for the classification and identification of density distributions formed by each polymer species.^[25,26] However, the former method suffers from the issue of a narrow identification range, while the latter approach requires modifying and retraining the neural network model after each change to the candidate phase library. This leads to time-consuming procedures, limited scalability, and high computational demands. Moreover, the complexity of neural networks increases when dealing with the variability in the degree of fuzziness of interfaces in the ordered phases due to changes of interaction parameters, often limiting their applicability to high χN scenarios.

In this study, we present a novel scattering-based automated identification strategy (SAIS) for the characterization and identification of ordered phases in a single unit cell under periodic boundary conditions based on their unique scattering patterns. The distinguishing feature of our approach lies in its capacity to handle a wide range of ordered phases within an extensive parameter space, while delivering rapid and precise identification results with minimal computational requirements. Moreover, our strategy not only characterizes the phases but also extracts essential information for recognition, enabling real-time expansion and improvement to the candidate phase library. It should be noted that our SAIS is specially designed for automating simulation studies where both density distributions within a single unit cell and associated lattice parameters are available beforehand. Thus, it is not applicable to identify phase structures based on experimental results such as scattering curves or images of crystalline structures where state-of-the-art methods already present.^[27]

METHODS

To demonstrate our strategy, we adopt unit cell SCFT calculations to generate the ordered phases for analysis. However, it should be noted that our strategy is not limited to the density distributions obtained from SCFT but applicable to any numerical methods that can produce density distribution of arbitrary ordered phases in a single unit cell with periodic boundary conditions. The d -dimensional unit cell is evenly discretized along each dimension into $N = \prod_i N_i$ grid points, where $N_i = |\mathbf{a}_i| / \Delta x$ for $i = 1, 2, \dots, d$ and Δx is the spatial resolution. Here, \mathbf{a}_i , represented by column vectors, are the basis vectors of the unit cell. The density distribution of each species is represented by an array \mathbf{D} . Therefore, the discretized unit cell represents a one-period sampling of the periodic ordered structures. While we take 3D unit cell as an example, the formulation is applicable to 1D and 2D unit cells as well.

Our aim is to calculate the 1D scattering pattern of an ordered structure from its discrete density distribution in a single unit cell. According to the scattering theory,^[28] the scattering intensity of a perfect crystal is given by:

$$I(\mathbf{q}) = |F(\mathbf{q})|^2 \delta(\mathbf{q} - \mathbf{q}_{hkl}) \quad (1)$$

where \mathbf{q} is the scattering vector, $|F(\mathbf{q})|^2$ is the form factor of a unit cell, and δ denotes the Kronecker delta function. The above equation implies that the scattering intensity can have non-zero value only when the scattering vector satisfies the Laue condition, i.e., $\mathbf{q} = \mathbf{q}_{hkl} = 2\pi(h\mathbf{a}_1^* + k\mathbf{a}_2^* + l\mathbf{a}_3^*)$, where \mathbf{a}_i^* are basis vectors of the reciprocal lattice and h, k, l are integers known as Miller indices.^[29] Note that reciprocal basis vectors, \mathbf{a}_i^* , can be computed from \mathbf{a}_i .^[30]

The form factor of the unit cell evaluated only at \mathbf{q}_{hkl} can be expressed as:^[28,31]

$$F(\mathbf{q}_{hkl}) = \int_{\Omega} d\mathbf{r} \rho(\mathbf{r}) \exp(-i\mathbf{q}_{hkl} \cdot \mathbf{r}) \quad (2)$$

where $\rho(\mathbf{r})$ is the density distribution of scattering length within the unit cell, and Ω denotes the domain of the unit cell. In our case, $\rho(\mathbf{r})$ is discretized into $b\mathbf{D}$, where b is the scattering length of the polymer species. Consequently, the form factor can be rewritten as:^[31]

$$F(\mathbf{q}_{hkl}) = \frac{b}{N} \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} \sum_{n_3=0}^{N_3-1} \mathbf{D} \exp(-i\mathbf{q}_{hkl} \cdot \mathbf{r}_{n_1 n_2 n_3}) \quad (3)$$

where $\mathbf{r}_{n_1 n_2 n_3} = \frac{n_1}{N_1} \mathbf{a}_1 + \frac{n_2}{N_2} \mathbf{a}_2 + \frac{n_3}{N_3} \mathbf{a}_3$ with $n_i = 0, 1, \dots, N_i - 1$

are the position vectors of each grid point, (n_1, n_2, n_3) , in the unit cell.

By utilizing the relationship between the basis vectors in the real space and the reciprocal space,^[29,30] Eq. (3) can be further simplified into

$$F(\mathbf{q}_{hkl}) = \frac{b}{N} \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} \sum_{n_3=0}^{N_3-1} \mathbf{D} \exp \left[-i2\pi \left(\frac{n_1 h}{N_1} + \frac{n_2 k}{N_2} + \frac{n_3 l}{N_3} \right) \right] \quad (4)$$

It is important to note that the right-hand side is nothing but the 3D discrete Fourier transform (DFT) of the array \mathbf{D} .^[32] usually denoted as $\hat{\mathbf{D}}$. Therefore, $F(\mathbf{q}_{hkl}) = b\hat{\mathbf{D}}$. Finally, the scattering intensity can be calculated by the following equation:

$$I(\mathbf{q}_{hkl}) = b^2 |\hat{\mathbf{D}}|^2 \quad (5)$$

where $\hat{\mathbf{D}}$ can be efficiently computed via the fast Fourier transform (FFT) algorithm.

To obtain the 1D scattering pattern, we first compute the length of each scattering vector using $q = |\mathbf{q}_{hkl}| = 2\pi |\mathbf{M}[h \ k \ l]^T|$ with $\mathbf{M} = [\mathbf{a}_1^* \ \mathbf{a}_2^* \ \mathbf{a}_3^*]$ being the transformation matrix of the reciprocal lattice.^[14] We then group all Miller indices with identical q values. The scattering intensity of each group is the sum of intensities corresponding to all \mathbf{q}_{hkl} presented in the group as computed by Eq. (5). Reflections are those groups with intensities higher than a given threshold value, otherwise they are considered as extinctions. Finally, the 1D scattering pattern is represented by a full list of all possible Miller indices and the information whether they are reflections or extinctions, ordering by their q values. The power of our scattering theory-based approach lies in the fact both q values and their associated Miller indices are obtained simultaneously, in contrast to the SAXS experiments where process of annotating scattering peaks is much more challenging.

It should be noted that altering lattice constants can lead to a change in the ordering of q values computed for different Miller indices. Since Miller indices only depend on the symmetry and space group of specific structures, changes in

the proportional lattice parameters do not alter the characteristic Miller indices. Therefore, each structure possesses its unique "standard lattice constants" representing the ratio of lattice parameters. During the identification process, the standard lattice constants are employed to establish a fixed sequence of Miller indices. Examples of the standard lattice constants can be found in the section 1 of the electronic supplementary information (ESI).

Another point to note is that, to distinguish between different phases with the same symmetry in multi-species systems, we need to differentiate between "structural species" and "background species" to obtain the density distribution used for calculations. This will be detailed in the "Multi-species Block Copolymers" section later.

RESULTS AND DISCUSSION

In this study, we consider a wide range of candidate phases including all stable ordered phases in two-species BCPs known to date and some selected metastable phases, which are listed below:

1. One-dimensional (1D) ordered phases: asymmetrical lamellae (LAMa), symmetrical lamellae (LAMs).
2. Two-dimensional (2D) ordered phases: Square-packed cylinders (SPC, No.10), C^4C^4 (No.10),^[22] Hexagonally-packed cylinders (HEX, No.17), C^3C^6 (No.17).^[22]
3. Three-dimensional (3D) ordered phases: O^{70} (No.70),^[33,34] Perforated layer (PL (No.123) & PL (No.139)),^[35] σ phase (No.136),^[8,36] Hexagonal close-packed spheres (HCP, No.194),

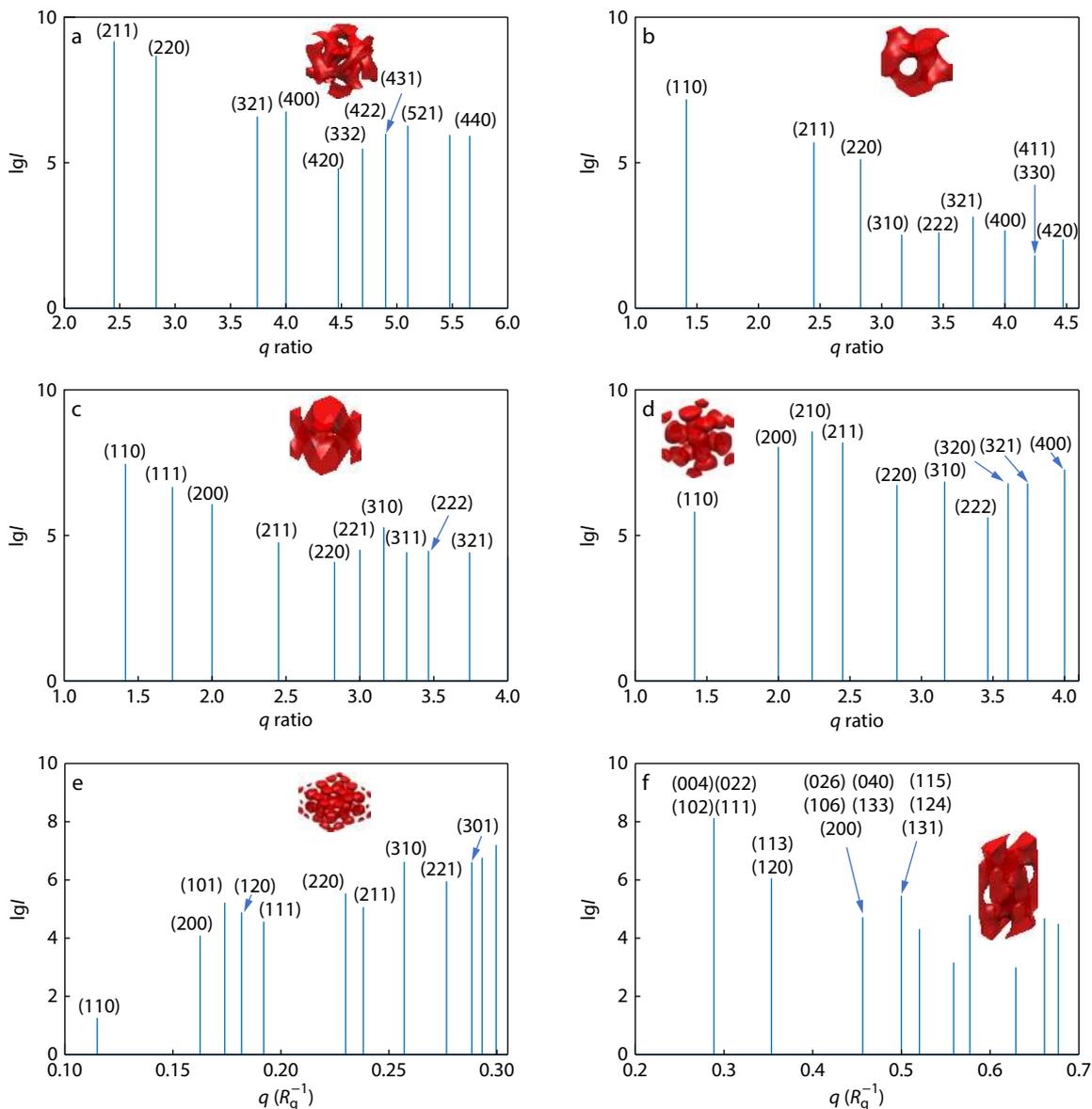


Fig. 1 Scattering patterns of six typical phases: (a) Double gyroid (DG), (b) Single gyroid (SG), (c) Double diamond (DD), (d) A15, (e) σ , (f) O^{70} . Red represents species A (species B forming the matrix is not shown). For the sake of clarity, only the first 10 peaks are shown in each subplot. In (a-d), the abscissa shows the relative q ratio instead of absolute q values for a clearer visualization and comparison, which is most appropriate for the cubic crystal system.

Single gyroid (SG, No.214),^[33,37] Simple cubic (SC, No.221), CsCl (No.221),^[22] A15 (No.223),^[9,20] Double diamond (DD, No.224),^[33,37] Face-centered cubic (FCC, No.225), NaCl (No.225),^[22] Body-centered cubic (BCC, No.229), Double gyroid phase (DG, No.230).

Fig. 1 shows the scattering patterns of six typical phases: DG,^[33] SG,^[37] DD,^[33] A15,^[20] σ ,^[8] and O^{70} .^[34] These results are consistent with previous experimental and theoretical observations. Plots of the remaining phases can be found in Figs. S1–S9 in ESI. As can be seen, each phase obeys its unique reflection conditions in its scattering pattern. To facilitate SAIS, we maintain a database of reflection conditions for each candidate phase. Hence, during SAIS we can systematically exclude candidate phases by comparing the corresponding Miller indices of each reflection and the remaining candidate phase, if there is any, should be assigned to the test phase. A detailed description of SAIS and related results are presented in following sections.

First Stage of SAIS

Initially, the sample can be classified into one of the seven crystal systems based on known lattice parameters: triclinic, monoclinic, orthorhombic, tetragonal, trigonal, hexagonal, and cubic,^[29] significantly narrowing down the list of candidate phases to be considered. Next, we compute the 1D scattering patterns of the test phase and all possible candidate phases according to Eq. (5). Then, a comparison of Miller indices between the test phase and each candidate phase is performed in an ascending order of q values. Whenever a mismatch is detected, meaning that one phase exhibits reflection while the other phase exhibits extinction at the same q or *vice versa*, the corresponding candidate phase is excluded from consideration. Meanwhile, the Miller index of the first mismatch is recorded for future reference.

In practice, due to the significant differences in reflection conditions among various space groups, distinguishing them only requires a subset of the reflection conditions. This can also help avoid accidental zeros (extinctions) that may arise at high q values^[38,39] (an example can be found in Fig. S10 in ESI). Therefore, it is important to choose an appropriate threshold for the number of q values to be compared. The threshold, N_t , can be adjusted based on the specific charac-

teristics of the candidate phases being analyzed, allowing users to fine-tune this criterion based on their specific requirements and knowledge of the system under investigation. For example, setting N_t to 6 is sufficient to distinguish all known stable ordered phases for two-species BCPs.

Fig. 2 provides an example of applying SAIS to cubic phases where the test phase is A15. The absence of (100) reflection excludes five candidate phases and the presence of (110) reflection excludes another two more. DD, SG and BCC are further excluded after the comparison of the (111), (200) and (210) reflections, respectively. Only A15 survives after the comparison of first five Miller indices. As the number of matches, which is at least 12 shown in Fig. 2, exceeds the threshold ($N_t = 6$), the test phase is identified as A15 as expected.

For non-cubic crystal systems, the matching process is expected to be much easier since there are fewer candidate phases to be compared. Fig. 3 presents an example of the identification of the HCP phase in the orthorhombic crystal system. Both HCP and HEX phases are computed in a conventional hexagonal unit cell. As can be seen, all possible candidate phases exhibit quite divergent scattering patterns, leading to quick exclusion of incorrect phases. Note that both LAM and HEX can orientate towards either basis vector of the unit cell. The zero in the Miller index suffixed to the phase indicates the orientation of a low-dimensional phase. For example, the direction of the cylinders in HEX(hk0) phase and the normal direction of the layers in LAM(00l) are both parallel to the vector (00l).

Second Stage of SAIS

Through the first stage, candidate phases are systematically eliminated by comparing every Miller index in the order of q values, narrowing down the range of search, and ensuring that erroneous reflections are absent in the test phase. However, when exploring completely new and unknown BCP systems, due to the complexity of possible ordered phases, N_t may become either too small or too large. If N_t is too small, multiple candidate phases may be left after the first stage of SAIS, making the situation ambiguous. On the other hand, a large N_t may mistakenly remove the correct candidate phase due to accidental extinctions in the high- q region.

Cubic	(100)	(110)	(111)	(200)	(210)	(211)	(220)	(221)	(300)	(310)	(311)	(222)
Test phase												
A15	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗
BCC	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗
SG	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗
DD	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗
FCC	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗
DG	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗
SC	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗
SPC	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗
PL(No.123)	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗
LAMa	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗
LAMs	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗	⊗

Fig. 2 The matching of reflections and extinctions by the sequence of Miller indices during the first stage of SAIS. The test phase is A15. The left most column lists the test phase and all candidate cubic phases, while the top row displays Miller indices arranged in the ascending order of q value. The PL (No.123) and three low-dimensional phases (SPC, LAMa and LAMs) are also considered because they are also possible in the cubic lattice. Red and green cells indicate the presence of reflections of unidentified and candidate phases, respectively. Note that for cubic crystal systems, the Miller indices (hkl) are permutable.^[39]

Orthorhombic	(100)	(010)	(001)	(110)	(101)	(011)	(111)	(200)	(020)	(002)	(102)	(112)
Test phase												
HCP	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
HEX(<i>hk0</i>)	✓	✓	✓	✓	✓	✓	✗					
HEX(<i>h0l</i>)	✓	✓	✓	✗								
HEX(<i>0kl</i>)	✓	✓	✓	✗								
O ⁷⁰	✓	✓	✓	✗								
LAMa(<i>00l</i>)	✓	✓	✗									
LAMs(<i>00l</i>)	✓	✓	✗									
LAMa(<i>0kl</i>)	✓	✗										
LAMs(<i>0kl</i>)	✓	✗										
LAMa(<i>h00</i>)	✗											
LAMs(<i>h00</i>)	✗											

Fig. 3 Applying the first-stage of SAIS to the orthorhombic crystal system similar to Fig. 2. The test phase is HCP which is correctly identified.

To circumvent above difficulties, a second stage of SAIS is introduced to validate and refine the results obtained by the first stage. In this stage, a residual value (ε) between scattering patterns of the test phase and each candidate phase is computed according to

$$\varepsilon = \frac{1}{N_q} \sqrt{\sum_{i=1}^{N_q} |v_i^{-1} - u_i^{-1}|^2} \quad (6)$$

where, v and u represent the scattering patterns (the sequence of q) of the candidate phase and the test phase, respectively, and N_q denotes the number of reflections. If the total number of reflections, denoted as N_x , in a particular phase is lower than N_q , then N_x is used for calculation instead. To emphasize the importance of smaller q values, we use the reciprocal of q , which is equivalent to the distance between two neighboring lattice planes, to measure the distance of two vectors instead of conventional mean squared error.

In the second stage of SAIS, the candidate phase giving lowest ε is of most interest. If the first stage of SAIS assigns a candidate phase to the test phase, the second stage confirms such identification if the residual for that phase is indeed the lowest. Otherwise, the identification made by the first stage is in doubt, implying that a bad N_t is chosen. However, this scenario should rarely occur for a sufficiently large N_t . Indeed, we have never encountered such case throughout this study. In the case of multiple phases left after the first stage of SAIS, the one with lowest ε is chosen as the phase for the test phase which resolves the ambiguous issue.

For the case when no candidate phase is left after the first stage of SAIS, two scenarios are possible: the correct phase is accidentally removed or the test phase is actually a new phase which is not presented in the candidate library. The former scenario occurs when N_t is too large and accidental extinctions cause the failure of the first stage of SAIS. In this scenario, the q sequence is a subset of the perfect q sequence without accidental extinctions. The correct result can still be obtained through the second stage of SAIS because the ε of a phase with accidental extinctions is still significantly smaller than that of other phases. An example of accidental extinctions can be found in Fig. S10 (ESI). Further verification can be

done by checking whether the Miller index of the first mismatch lies in the high- q region. For the later scenario, SAIS can automatically incorporate such new phase into the candidate structure library to identify its later occurrences. And by utilizing explicit lattice parameters and reflection conditions, users can determine its corresponding space group by consulting reference tables.^[39]

To demonstrate the robustness of the second stage of SAIS, we perform tests by varying three typical physical parameters: the volume fraction of the species A (f_A), lattice parameters (a), and the Flory-Huggins interaction parameter (χN). Changes in these parameters lead to variations in the sizes of phase-separated domains, the number of sampling points, and the interface between A-rich and B-rich domains. An increase in the χN parameter results in a sharper interface, causing a transition from diffuse to well-defined boundaries between individual domains.^[6] These variations present notable challenges for image classification techniques that rely on neural networks, necessitating larger and more intricate models.^[40,41] In contrast, our scattering theory-based approach effectively addresses these challenges. A more detailed comparison will be given in a later section.

Fig. 4 presents the residuals as a function of various parameters used to generate 18 samples (12 for the f_A and χN , 6 for a), all of which are different from the standard phases stored in the candidate phase library. The phase region from which the parameters are selected is illustrated in left column of subplots, with the phase diagram reproduced from reference.^[17] Note that these ordered phases can be either stable or metastable depending on where they are located in the phase diagram. As can be seen, the lowest residuals correspond to BCC, DG and FCC in Figs. 4(a), 4(b) and 4(c), respectively, confirming the identification results of the first stage of SAIS. This indicates that our approach is capable of effectively handling a wide range of parameters to provide reliable identification results.

Multi-species Block Copolymers

To extend SAIS to multi-species BCPs, it is necessary to address the challenge of distinguishing distinct ordered structures formed by different species that share identical scattering pat-

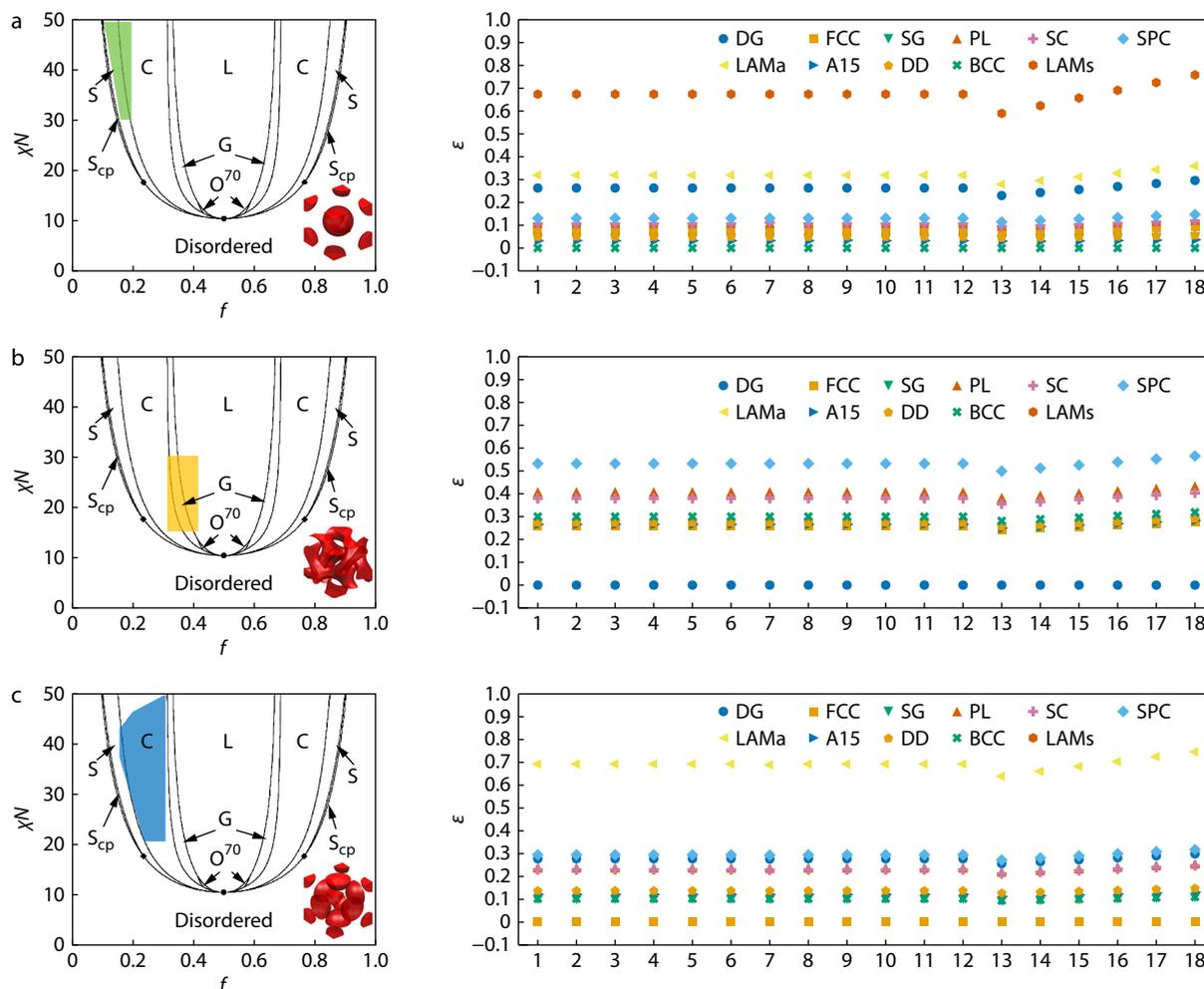


Fig. 4 Residuals between the test phase and each candidate phase in the library as a function of various parameters. (a) BCC, (b) DG, (c) FCC. Left column of each subplot shows the phase region from which we select parameters. Samples 1–12 demonstrate modifications of the f_A and χN parameters, while samples 13–18 showcase modifications of the lattice parameter. The lowest residuals in each subplot correspond to BCC, DG and FCC, respectively.

terns in the sense of ignoring differences in peak intensities. For instance, in the case of ABC-type BCPs, phases such as a simple cubic phase formed by species A (SC_A), a simple cubic phase formed by species C (SC_C), and a CsCl phase formed by species A and C (CsCl) are indistinguishable using SAIS alone. Similar situations can also arise in the comparison among FCC_A , FCC_C , & NaCl, and in the comparison between SPC & C^4C^4 . To properly identify these phases, it is crucial to find a reliable way to differentiate the "background" species from the species forming minor domains (referred to as "structural" species). In the above example, the background species are quite different: species B and C for the SC_A phase, species A and B for the SC_C phase, and species B for the CsCl phase. Thus, the accurate determination of the background species is essential for effectively discerning these distinct structures.

Here, we propose a variance-like quantity to capture the background species. For each species, $X \in \{A, B, C\}$, the variance can be computed using the equation

$$v_X = \sum_{i \in I^*} \left(\frac{D_{X,i}}{\bar{D}_X} - 1 \right)^2 \quad (7)$$

where I^* denotes a collection of grid points whose density exceed the mean value, *i.e.*, $D_{X,i} > \bar{D}_X$ with $\bar{D}_X = \frac{1}{N} \sum_{i=1}^N D_{X,i}$. As shown in Fig. 5, the variance of the structural species is significantly higher than those of the background species. The corresponding scattering patterns of SC_A , SC_C , CsCl, FCC_A , FCC_C , NaCl, SPC and C^4C^4 formed by ABC-type BCPs can be found in Figs. S11–S18 in ESI. In practice, we can set a proper threshold value for v_X , below which the species is considered as the background.

The variance-like criterion is particularly important for automated identification of a multi-species system. In the case of AB-type BCPs, this information serves as an additional parameter that aids in distinguishing phases like BCC_A and BCC_B . After determining the background species, scattering lengths, b_X , $X \in \{A, B, C, \dots\}$, are assigned to each individual species. In general, the scattering length of the background species can be set to zero, and the scattering lengths of individual structural species can be adjusted according to specific requirements.

Finally, the individual density distributions are combined

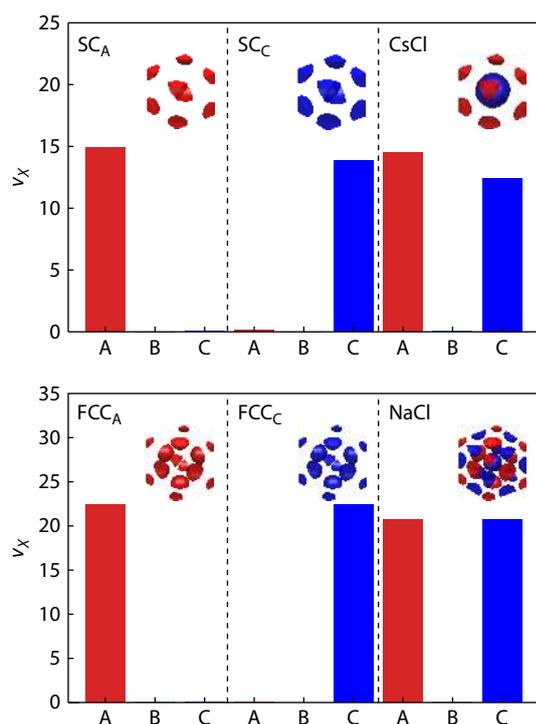


Fig. 5 The variance (v_x) of the species A, B, C in SC_A , SC_C , CsCl, FCC_A , FCC_C and NaCl. Red represents species A while blue represents species C (species B forming the matrix is not shown). The variance of the background species is significantly lower than that of the structural species, typically being less than 1/10, which is almost invisible in the plots.

into a single density distribution, $\mathbf{D} = \sum_x b_x \mathbf{D}_x$, which is then used as input for SAIS. This treatment effectively captures the inherent structural characteristics of multi-species systems and enables the identification of complex phases.

Imperfect Ordered Structures

In unit cell SCFT calculations, structures with "crystallographic defects", such as point defects, line defects, and planar defects, are impossible to appear in theory.^[14] Due to the periodic boundary conditions employed in unit cell calculations, any resulted structure is inherently periodic, including the "defects" that cause deviations from a perfect structure. Since these periodic results do not contain conventional crystallographic defects that typically occur at random positions, they can still be considered as "perfect crystals". Therefore, in this context, "defect-free" refers to the absence of conventional crystallographic defects that typically occur randomly and disrupt the overall structure. And the term "defect" refers to structural deviations that occur periodically but do not adhere to the symmetry requirements specified by a space group. In the subsequent sections, we refer to this type of structures as distorted structures.

Two types of imperfect structures are commonly encountered: defect-free but non-symmetrical structures and distorted structures. The former refers to the case where domains of interest, *i.e.*, motifs, do not position at crystal lattice point as one usually does in crystallography. Non-symmetrical structures often emerge due to the periodic boundary condition imposed in numerical simulations which allows translating the ordered structure along basis vectors by any amount.

Such non-symmetrical density distribution within a unit cell may introduce extra difficulty into identification methods relying on visualization or image recognition techniques. In contrast, our scattering-based approach is naturally immune to translation since the Fourier transform of periodic functions is not affected by translation.

To validate this assertion, we perform random translations of the density distributions of FCC and A15 phases and utilize SAIS to identify them. As shown in Fig. S19 in ESI, the scattering patterns of translated structures undergo negligible changes, resulting in the accurate identification of these translated structures.

In addition, SAIS are capable of handling noise-mediated structures and distorted structures. Although the density distributions calculated by the unit cell SCFT is free of any noise, such situations may arise in molecular dynamic simulations due to thermal fluctuations.^[42] To demonstrate the ability of SAIS in identifying noise-mediate structures, we randomly selected 10% of the data points from the density distribution and multiply them with a Gaussian noise $N(\mu=1.0, \sigma=0.1)$. A slightly larger threshold of the scattering intensity is set to filter out extinctions. As demonstrated in Fig. S20 (in ESI), scattering patterns for various noise-mediated structures are consistent with those of noise-free structures. Therefore, they can be correctly identified by SAIS.

Distorted structures are difficult to distinguish visually and often exhibit excess reflections which do not belong to the original space group. This situation frequently occurs in the vicinity of phase boundaries. The free energy of distorted structures is often higher than that of normal structures, thereby introducing further uncertainty into calculation of phase diagrams. Taking the identification of distorted BCC phases as an example, due to the difference between sizes of the spheres on the corner and that on the center of the unit cell, the rule of translation symmetry is broken and these ordered structures, in a strict manner, are no longer belong to BCC. These distorted structures and their scattering patterns are shown in Figs. 6(b)–6(d). SAIS identifies them as new phases due to the mismatch of the reflection conditions between these structures and all candidate phases. Similar situations may also arise when varying parameter χN for FCC phases (Figs. 6e and 6f). These kinds of distorted structures are difficult to distinguish *via* identification methods based on neural network models that are built on top of image recognition since these neural network models typically are not aware of the crystallographic nature of density distributions within a unit cell. The early detection of defects in the target structure can effectively reduce the time requirement for the determination of phase boundaries.

Comparison with Existing Methods

Existing methods for automatic phase recognition include density-variance-based calculation^[25] and machine learning methods relying on image recognition neural networks.^[26] The former method has been proven to be efficient in measuring the distance between certain predicted and target structure, but it suffers from poor robustness and works only in a very limited range of parameters. For instance, the changes in f , χN and translation of the structure easily leads to failures of this method, as demonstrated in Fig. S21 (in ESI).

As mentioned earlier, the change in the interface between

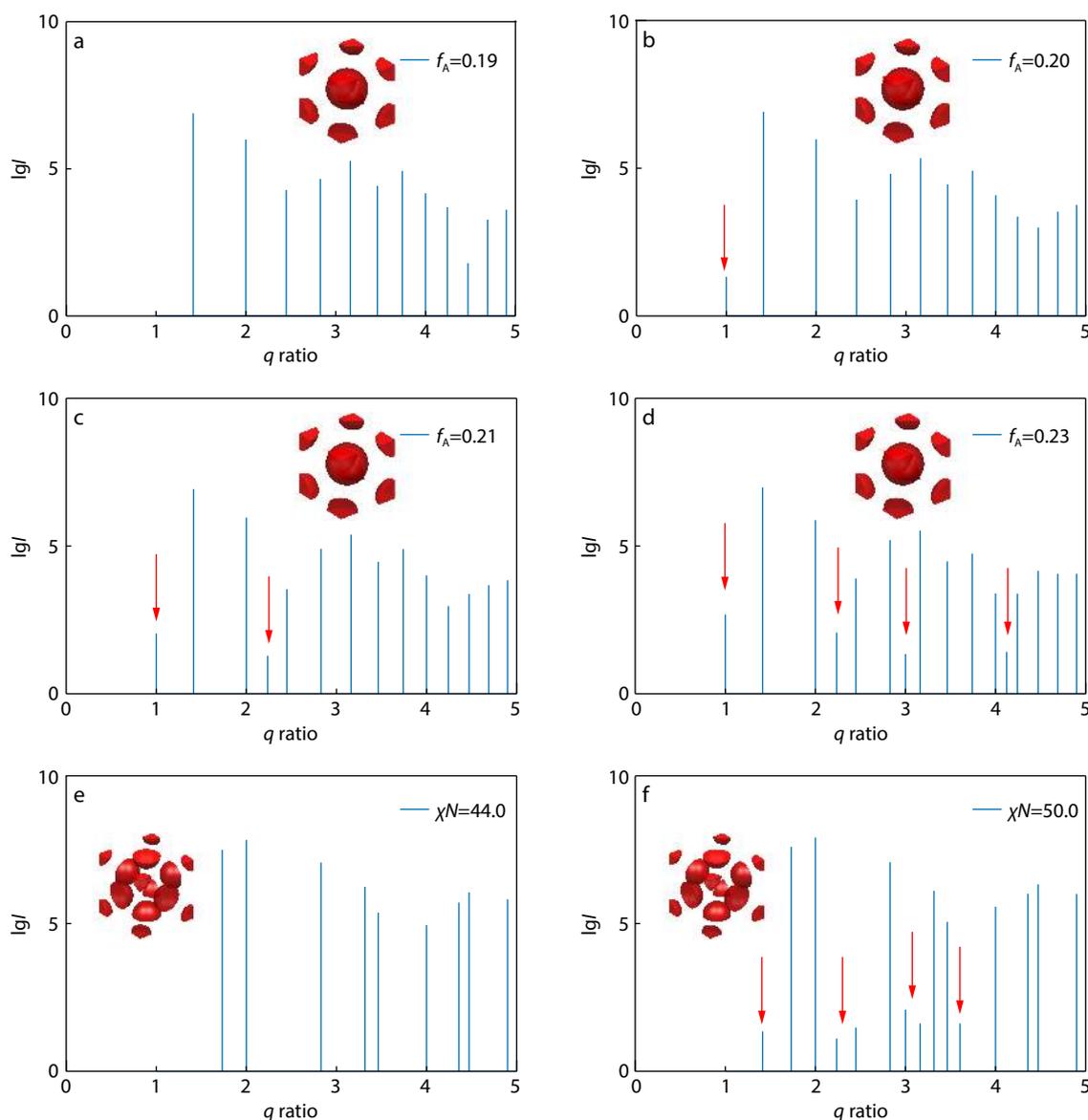


Fig. 6 Scattering patterns of standard BCC phase (a), distorted BCC phases (b–d), standard FCC phase (e), distorted FCC phase (f). (a) $f_A = 0.19$. (b) $f_A = 0.20$. (c) $f_A = 0.21$. (d) $f_A = 0.23$. (e) $\chi N = 44.0$. (f) $\chi N = 50.0$. The red arrows indicate excess reflections that should have been extinctions.

individual domains poses serious challenges for machine learning methods relying on image recognition neural networks. For example, according to Dong *et al.*,^[26] most errors occur in the distinction between the C^3C^6 phase formed by species A and C, and the HEX_C phase formed by species C alone. Furthermore, as the volume fraction of species C decreases, the probability of misidentification increases. Within a specific parameter space ($f_A = 0.10, f_C \in [0.20, 0.30]$), the machine learning method frequently ignores the presence of the A domain due to its small volume fraction, resulting in the misclassification of C^3C^6 as HEX_C .

Using SAIS, the given situation is analogous to identification of distinct phases with identical scattering patterns as previously mentioned. The scattering patterns can be found in the Supporting Information (Fig. S22 in ESI). As illustrated in Fig. 7, it can be seen that the variance of the structural

species remains significantly larger than the variance of the background species. After accurate identification of both structural and background species, SAIS effectively circumvents potential misjudgments and successfully identifies these two phases.

CONCLUSIONS

In conclusion, our proposed two-stage strategy, SAIS, for performing automated phase identification of SCFT results utilizing scattering patterns of various ordered phases, offers significant advantages over existing solutions. By leveraging the scattering theory and utilizing efficient FFT techniques, our approach achieves reduced computational requirements while providing accurate and scalable phase recognition.

The first stage of SAIS relies on comparing reflection condi-

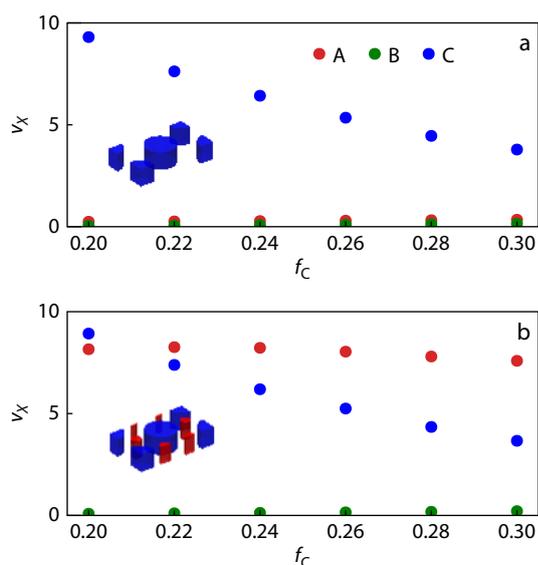


Fig. 7 The variances of the species A, B, C in HEX_C (a) and C^3C^6 (b). The variance of the structural species (A and C) remains significantly larger than that of the background species.

tions at given Miller indices, enabling efficient and accurate identification of known phases. The second stage of SAIS introduces a special tailored residual between the test phase and each phase in the candidate phase library, which serves as a tool to aid the identification process. Once the scattering pattern of a new phase is computed, they can be conveniently incorporated into the candidate phase library for future reference. This allows for continuous expansion and improvement of the candidate phase library by incorporating new phases. Additionally, by employing a variance-like criterion to distinguish background species, SAIS can be extended to multi-species BCP systems, further expanding its range of applications.

Our lightweight approach can be easily integrated into existing research workflows that involve unit cell calculations, serving as an automated phase identification module with negligible computational overhead. Through precise automated identification, the free energies of all candidate phases at each point in the parameter space can be computed automatically and reliably, consisting of an essential procedure for the automated construction of phase diagrams. The flexibility of online updates and adjustments to the candidate phase library enables efficient and automated exploration of a wide parameter space, facilitating the discovery of new phases. With the assistance of the SAIS, programs have the capacity to initiate the construction of the candidate phase library from scratch, continuously identifying and expanding this library to capture every phase that may emerge. We anticipate that our automated identification strategy will become a fundamental tool for the simulation studies of ordered phases of BCPs.

Conflict of Interests

The authors declare no interest conflict.

Electronic Supplementary Information

Electronic supplementary information (ESI) is available free of charge in the online version of this article at <http://doi.org/10.1007/s10118-024-3084-x>.

Data Availability Statement

The supplementary information and relevant data associated with this article can be found in the provided database from Data Repository of China Association for Science and Technology (DOI: 10.1007/s10118-024-3084-x). These files have also been uploaded on GitHub for reference, accessible at the following URL: <https://github.com/DShKM118/Supporting-Information-for-SAIS>. Should you have any inquiries, please feel free to contact the author via email at 22210440033@fudan.m.edu.cn, and I will be pleased to address any questions you may have.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grants No. 21873021).

REFERENCES

- Bates, F. S.; Fredrickson, G. H. Block copolymers—designer soft materials. *Phys. Today* **1999**, *52*, 32–38.
- Cummins, C.; Lundy, R.; Walsh, J. J.; Ponsinet, V.; Fleury, G.; Morris, M. A. Enabling future nanomanufacturing through block copolymer self-assembly: a review. *Nano Today* **2020**, *35*, 100936.
- Pester, C. W.; Liedel, C.; Ruppel, M.; Böker, A. Block copolymers in electric fields. *Prog. Polym. Sci.* **2017**, *64*, 182–214.
- Shao, G.; Liu, Y.; Cao, R.; Han, G.; Yuan, B.; Zhang, W. Thermo-responsive block copolymers: assembly and application. *Polym. Chem.* **2023**, *14*, 1863–1880.
- Fan, X.; Zhao, Y.; Xu, W.; Li, L. Linear-dendritic block copolymer for drug and gene delivery. *Mate. Sci. Eng. C* **2016**, *62*, 943–959.
- Matsen, M. W.; Schick, M. Stable and unstable phases of a diblock copolymer melt. *Phys. Rev. Lett.* **1994**, *72*, 2660–2663.
- Dorfman, K. D. Frank-Kasper phases in block polymers. *Macromolecules* **2021**, *54*, 10251–10270.
- Chen, M.; Huang, Y.; Chen, C.; Chen, H. Accessing the Frank-Kasper σ phase of block copolymer with small conformational asymmetry via selective solvent solubilization in the micellar corona. *Macromolecules* **2022**, *55*, 10812–10820.
- Li, W.; Liu, Y. Simplicity in mean-field phase behavior of two-component miktoarm star copolymers. *J. of Chem. Phys.* **2021**, *154*, 014903–014903.
- Han, L.; Che, S. An overview of materials with triply periodic minimal surfaces and related geometry: from biological structures to self-assembled systems. *Adv. Mater.* **2018**, *30*, 1705708.
- Ha, S.; La, Y.; Kim, K. T. Polymer cubosomes: infinite cubic mazes and possibilities. *Acc. of Chem. Res.* **2020**, *53*, 620–631.
- Feng, X.; Burke, C. J.; Zhuo, M.; Guo, H.; Yang, K.; Reddy, A.; Prasad, I.; Ho, R.; Avgeropoulos, A.; Grason, G. M.; Thomas, E. L. Seeing mesoatomic distortions in softmatter crystals of a double-gyroid block copolymer. *Nature* **2019**, *575*, 175–179.
- Hamley, I. W. in *Developments in block copolymer science and technology*. Wiley, **2004**.
- Fredrickson, G. H. in *The Equilibrium Theory of Inhomogeneous Polymers*. Oxford University Press, **2005**.

- 15 Matsen, M. W. The standard Gaussian model for block copolymer melts. *J. Phys. Condensed Matter* **2002**, *14*, R21.
- 16 Epps, T. H.; Cochran, E. W.; Hardy, C. M.; Bailey, T. S.; Waletzko, R. S.; Bates, F. S. Network phases in ABC triblock copolymers. *Macromolecules* **2004**, *37*, 7085–7088.
- 17 Matsen, M. W. Effect of Architecture on the phase behavior of AB-type block copolymer melts. *Macromolecules* **2012**, *45*, 2161–2165.
- 18 Lee, S.; Bluemle, M. J.; Bates, F. S. Discovery of a Frank-Kasper σ phase in sphere-forming block copolymer melts. *Science* **2010**, *330*, 349–353.
- 19 Jung, H. Y.; Park, M. J. Thermodynamics and phase behavior of acid-tethered block copolymers with ionic liquids. *Soft Matter* **2016**, *13*, 250–257.
- 20 Schulze, M. W.; Lequieu, J.; Barbon, S. M.; Lewis, R. W.; Delaney, K. T.; Anastasaki, A.; Hawker, C. J.; Fredrickson, G. H.; Bates, C. M. Stability of the A15 phase in diblock copolymer melts. *Proc. Nat. Acad. Sci. U.S.A.* **2019**, *116*, 13194–13199.
- 21 Xie, Q.; Qiang, Y.; Li, W. Regulate the stability of gyroids of ABC-type multiblock copolymers by controlling the packing frustration. *ACS Macro Lett.* **2020**, *9*, 278–283.
- 22 Dong, Q.; Li, W. Effect of molecular asymmetry on the formation of asymmetric nanostructures in ABC-type block copolymers. *Macromolecules* **2021**, *54*, 203–213.
- 23 Xie, Q.; Qiang, Y.; Li, W. Single gyroid self-assembled by linear BABAB pentablock copolymer. *ACS Macro Lett.* **2022**, *11*, 205–209.
- 24 Arora, A.; Morse, D. C.; Bates, F. S.; Dorfman, K. D. Accelerating self-consistent field theory of block polymers in a variable unit cell. *J. Chem. Phys.* **2017**, *146*, 244902.
- 25 Paradiso, S. P.; Delaney, K. T.; Fredrickson, G. H. Swarm intelligence platform for multiblock polymer inverse formulation design. *ACS Macro Lett.* **2016**, *5*, 972–976.
- 26 Dong, Q.; Gong, X.; Yuan, K.; Jiang, Y.; Zhang, L.; Li, W. Inverse design of complex block copolymers for exotic self-assembled structures based on bayesian optimization. *ACS Macro Lett.* **2023**, *12*, 401–407.
- 27 Cullity, B. D.; Stock, S. R.; Education, I. *Elements of X-ray diffraction*. Pearson India Education Services, **2015**.
- 28 Li, T.; Senesi, A. J.; Lee, B. Small angle X-ray scattering for nanoparticle research. *Chem. Rev.* **2016**, *116*, 11128–11180.
- 29 Giacovazzo, C.; Al, E. *Fundamentals of crystallography*. Oxford University Press, **2009**.
- 30 Kittel, C. *Introduction to solid state physics*. Wiley, 2005.
- 31 Yager, K. G.; Zhang, Y.; Lu, F.; Gang, O. Periodic lattices of arbitrary nano-objects: modeling and applications for self-assembled systems. *J. Appl. Crystallography* **2013**, *47*, 118–129.
- 32 Duhamel, P.; Vetterli, M. Fast fourier transforms: a tutorial review and a state of the art. *Signal Processing* **1990**, *19*, 259–299.
- 33 Meuler, A. J.; Hillmyer, M. A.; Bates, F. S. Ordered network mesostructures in block polymer materials. *Macromolecules* **2009**, *42*, 7221–7250.
- 34 Takenaka, M.; Wakada, T.; Akasaka, S.; Nishitsuji, S.; Saijo, K.; Shimizu, H.; Kim, M. I.; Hasegawa, H. Orthorhombic Fddd network in diblock copolymer melts. *Macromolecules* **2007**, *40*, 4399–4402.
- 35 Tenneti, K. K.; Chen, X.; Li, C. Y.; Tu, Y.; Wan, X.; Zhou, Q.; Sics, I.; Hsiao, B. S. Perforated layer structures in liquid crystalline rod-coil block copolymers. *J. Am. Chem. Soc.* **2005**, *127*, 15481–15490.
- 36 Xie, N.; Li, W.; Qiu, F.; Shi, A. σ Phase formed in conformationally asymmetric AB-type block copolymers. *ACS Macro Lett.* **2014**, *3*, 906–910.
- 37 Xiang, L.; Li, Q.; Li, C.; Yang, Q.; Xu, F.; Mai, Y. Block copolymer selfassembly directed synthesis of porous materials with ordered bicontinuous structures and their potential applications. *Adv. Mater.* **2023**, *35*, 2207684.
- 38 Hammond, C. *The basics of crystallography and diffraction*. Oxford University Press, **2015**.
- 39 Aroyo, M. I.; Kristallographie, I. U. F. *International tables for crystallography. Volume A Space-group symmetry*; Chichester, West Sussex Wiley, **2016**.
- 40 Aoyagi, T. Deep learning model for predicting phase diagrams of block copolymers. *Comput. Mater. Sci.* **2021**, *188*, 110224.
- 41 Aoyagi, T. High-throughput prediction of stress-strain curves of thermoplastic elastomer model block copolymers by combining hierarchical simulation and deep learning. *MRS Adv.* **2021**, *6*, 32–36.
- 42 Dünweg, B.; Kremer, K. Molecular dynamics simulation of a polymer chain in solution. *J. Chem. Phys.* **1993**, *99*, 6983–6997.