Journal of Polymer Science

WILEY

JOURNAL OF POLYMER SCIENCE

RESEARCH ARTICLE OPEN ACCESS

Universal Phase Identification of Block Copolymers From Physics-Informed Machine Learning

Xinyi Fang[1,2] | Elizabeth A. Murphy[2,3,4] | Phillip A. Kohl[2,4] | Youli Li[2,4] | Craig J. Hawker[2,3,4,5] | Christopher M. Bates[2,3,4,5,6] | Mengyang Gu[1,2]

[1]Department of Statistics and Applied Probability, University of California, Santa Barbara, California, USA | [2]BioPACIFIC Materials Innovation Platform, University of California, Santa Barbara, California, USA | [3]Department of Chemistry and Biochemistry, University of California, Santa Barbara, California, USA | [4]Materials Research Laboratory, University of California, Santa Barbara, California, USA | [5]Materials Department, University of California, Santa Barbara, California, USA | [6]Department of Chemical Engineering, University of California, Santa Barbara, California, USA

Correspondence: Craig J. Hawker (hawker@mrl.ucsb.edu) | Christopher M. Bates (cbates@ucsb.edu) | Mengyang Gu (mengyang@pstat.ucsb.edu)

Received: 15 November 2024 | **Revised:** 10 January 2025 | **Accepted:** 10 January 2025

Funding: The research reported here was primarily supported by the National Science Foundation Materials Research Science and Engineering Center (MRSEC) at UC Santa Barbara (NSF DMR-2308708, IRG-1) and the BioPACIFIC Materials Innovation Platform of the National Science Foundation under Award No. DMR-1933487 (C.M.B., C.J.H. equipment, and characterization). C.M.B. thanks DMR-1844987 for supporting the synthetic and chromatography work. The research reported here made use of shared facilities of the Materials Research Science and Engineering Center (MRSEC) at UC Santa Barbara: NSF DMR-2308708. The UC Santa Barbara MRSEC is a member of the Materials Research Facilities Network (www.mrfn.org). E.A.M. gratefully acknowledges the National Science Foundation Graduate Research Fellowship Program under Grant No. 2139319. M.G. acknowledges the support from the National Science Foundation under Award No. OAC-2411043. X.F. acknowledges the support of the UC Multicampus Research Programs and Initiatives program under Grant No. M23PL5990.

Keywords: block copolymers | feature selection | machine learning | physics-informed | self-assembly

ABSTRACT

Block copolymers play a vital role in materials science due to their diverse self-assembly behavior. Traditionally, exploring the block copolymer self-assembly and associated structure–property relationships involve iterative synthesis, characterization, and theory, which is labor-intensive both experimentally and computationally. Here, we introduce a versatile, high-throughput workflow toward materials discovery that integrates controlled polymerization and automated chromatographic separation with a novel physics-informed machine-learning algorithm for the rapid analysis of small-angle X-ray scattering data. Leveraging the expansive and high-quality experimental data sets generated by fractionating polymers using automated chromatography, this machine-learning method effectively reduces data dimensionality by extracting chemical-independent features from SAXS data. This new approach allows for the rapid and accurate prediction of morphologies without repetitive and time-consuming manual analysis, achieving out-of-sample predictive accuracy of around 95% for both novel and existing materials in the training data set. By focusing on a subset of samples with large predictive uncertainty, only a small fraction of the samples needs to be inspected to further improve accuracy. Collectively, the synergistic combination of controlled synthesis, automated chromatography, and data-driven analysis creates a powerful workflow that markedly expedites the discovery of structure–property relationships in advanced soft materials.

Xinyi Fang and Elizabeth A. Murphy have contributed equally to this work.

1 | Introduction

Block copolymers are an important class of materials known for undergoing self-assembly into well-defined nanostructures, underpinning their use in applications including drug delivery, high-performance materials, and advanced electronics [1, 2]. Self-assembly into a variety of nanostructures, including body-centered cubic spheres (BCC), hexagonally packed cylinders (HEX), double gyroid networks (GYR), and lamellae (LAM), can be precisely tuned by various parameters including block chemistry, volume fraction (f), molecular weight (M_n), and architecture [3]. More recently, Frank–Kasper phases, hexagonally close-packed spheres, hierarchical X-in-Y structures, and complex network phases have been discovered in block copolymers, highlighting the ever-expanding palette of potential morphologies and properties achievable in this class of soft materials [4–8].

Traditionally, studying the phase behavior of block copolymers is laborious, involving iterative synthesis across many different compositions and molecular weights, coupled with characterizing each distinct material using tools such as small-angle X-ray scattering (SAXS). The complexity of this approach is underscored by the need for rigorous peak indexing of each SAXS pattern to accurately determine a given structure, which is complicated by issues including poor long-range order, missing reflections (i.e., form-factor minima that suppress allowed reflections), limited peak resolution, sample purity, coexisting phases, and extraneous atmospheric scattering [9–13]. At best, analyzing SAXS data sets is time-consuming and slow; at worst, the aforementioned issues cause challenges that only an expert-level understanding of X-ray theory can help resolve. The considerable complexity, time, and effort required to accurately identify nanostructures formed by block copolymers—and other materials—is compounded by an expansive parameter space, highlighting the potential utility of a workflow that accelerates the study of new materials, ideally in a fashion that is accessible to researchers from many different backgrounds.

Thinking about solutions to such experimental bottlenecks can find inspiration from recent advances in computation and theory. In many ways, similar issues are encountered with the *de-facto* tool for simulating block copolymer phase behavior, such as self-consistent field theory (SCFT) and coarse-grained model [14–20]. Yet SCFT and other simulation approaches rely on idealized parameters and assumptions that may not capture the distinct phase behaviors of block copolymers with different monomers. In addressing these challenges, recent advances in machine learning have created new opportunities to automate structural analysis by detecting patterns in data sets with high dimensionality [21–24]. For example, Jayaraman and coworkers developed a high-throughput machine learning enhanced computational method which analyzes SAXS patterns to reconstruct real-space 3D structures of materials [25–27]. Pozzo and colleagues introduced an algorithm which analyzes the shape of SAXS profiles to automatically generate phase maps [28]. Furthermore, Olsen and colleagues leveraged a random forest model to analyze the phase behavior of diblock copolymers using experimental data mined from literature [29]. While Olsen's method is effective in predicting phases of existing polymers in the database, this method, which we refer to as a chemistry-dependent random forest (CD-RF), relies on chemical-specific features, such as monomer identity, volume fraction, molecular weight, and temperature, which limits the predictive accuracy of phases formed by new monomers that are not in the database.

Here, we build on these advances and report a machine-learning method that rapidly, automatically, and accurately determines the morphology of block copolymers SAXS data to enable real-time data analysis. Unlike conventional machine-learning methods for analyzing SAXS data that process the full intensity curve without explicit noise filtering, our approach is informed by X-ray scattering theory and automatically extracts pivotal physics-informed morphological features from the reduced 1D intensity scattering pattern to construct a universal classification model and enable real-time data analysis for researchers. Statistically, our method introduces two major innovations. First, by modeling each intensity curve with a Gaussian process [30, 31], we incorporate experimental uncertainties across different wave vector magnitudes, q. Second, by integrating the physical principles of X-ray diffraction, we enhanced the effectiveness and accuracy of our predictive models. X-ray diffraction measures the power spectrum of the Fourier transform of the sample's electron density. Thus, details of a sample's phase and nanostructure manifest most prominently as diffraction peaks in the X-ray data profile; where the q-location of the strongest peak, peak width, and peak locations of all peaks with respect to the primary peak are the most important features with direct physical significance. With this knowledge, full-intensity curves were instead transformed into a small set of informative features, dramatically reducing the required size of the training set.

The combination of advanced statistical modeling and informed feature reduction significantly enhances the robustness and reliability of our predictive model for accelerating materials discovery. Our machine-learning process is structured into three stages: filtering the noise of intensity curves with a Gaussian process, detecting peaks and extracting crucial curve features, and applying a classification model, such as random forest [32], bagging [33], and gradient boosting [34] for material phase prediction. Importantly, we couple this new physics-informed machine-learning algorithm with a powerful experimental technique recently developed by our groups—automated chromatography—that yields a large set of well-defined and purified block copolymers from a very small number of as-synthesized samples [6, 7, 11, 35–38]. A notable advantage of this separation process over traditional iterative synthesis is its ability to remove homopolymer impurities and reduce the dispersity of each fraction compared to the as-synthesized parent block copolymer, enhancing the purity and reproducibility of fractionated samples' morphologies. Together, this combined workflow minimizes the synthetic burden of creating comprehensive and systematic sets of training data that were used to evaluate the accuracy and predictive capabilities of the machine-learning framework. Notably, this method achieves approximately 95% predictive accuracy across a variety of block copolymer chemistries, molecular weights, domain spacings, and nanostructures, including both new and existing materials in the training data set, which is much higher than other alternative approaches. Furthermore, this method can further identify and correct data that was initially mislabeled as a result of human error

based on quantified uncertainty of the prediction. After scrutinizing about 15% of the data with the highest uncertainty in predictive labels, the accuracy of our method approaches nearly 100%, marking a crucial advancement toward automated, high-throughput laboratories integrated with SAXS systems [39, 40]. The data and code used in this paper are publicly available (https://github.com/UncertaintyQuantification/automated_polymer_phase_identification).

2 | Results and Discussion

2.1 | Generating Block Copolymer Libraries

To evaluate the efficacy of our new machine-learning algorithm as applied to analyzing small-angle X-ray scattering data, we leveraged a large experimental data set (364 SAXS patterns) spanning four different block copolymer chemistries with systematically varying molecular weights and volume fractions as derived from automated chromatography. Note that we recently reported the phase behavior of these materials, which was manually determined by painstaking analysis of each individual SAXS pattern; these experimental phase portraits are reproduced in Figure 1 [11]. Each material is a diblock copolymer with a poly(dodecyl acrylate) block connected to one of four semi-fluorinated acrylates; we denote the volume fraction of the semi-fluorinated block in each case by f_F. Four classes of AB diblock copolymers with increasing degrees of fluorination were synthesized via sequential photo-initiated atom transfer radical polymerization: poly(dodecyl acrylate)-*b*-poly(2-fluoroethyl acrylate) (D-1F), poly(dodecyl acrylate)-*b*-poly(2,2,3,3,3-pentafluoropropyl acrylate) (D-5F), poly(dodecyl acrylate)-*b*-poly(1H,1H,2H,2H-nonafluorohexyl acrylate) (D-9F), poly(dodecyl acrylate)-*b*-poly(2,2,3,3,4,4,5,5,6,6,7,7-dodecafluoroheptyl acrylate) (D-12F). Readers interested in the experimental details of the synthesis and chromatographic separation are referred to our previous publication and the Supporting Information [11].

2.2 | Filtering and Feature Extraction

Automated chromatography yielded an extensive library of materials with high-quality self-assembly across a wide range of volume fractions ($f_F = 0.02 - 0.80$). Representative scattering patterns from six distinct morphologies (disordered (DIS), BCC, σ, HEX, GYR, and LAM) are plotted in Figure 2, where the blue curve represents the scattering intensities after noise filtration via a twice differentiable Gaussian process with a Matérn kernel [41]. We have innovated an approach that integrates experimental uncertainties in filtering and denoising with minimal computational demand, which is crucial for processing experimental data. In X-ray diffraction, certain peaks may be absent or suppressed due to factors like poor resolution or limited long-range order. For example, the σ pattern typically requires synchrotron-level capabilities to resolve its intricate diffraction pattern often comprised of circa 48 distinct peaks, whereas using traditional bench-top SAXS instruments would impart significant peak broadening despite analyzing the same nanostructure [4].

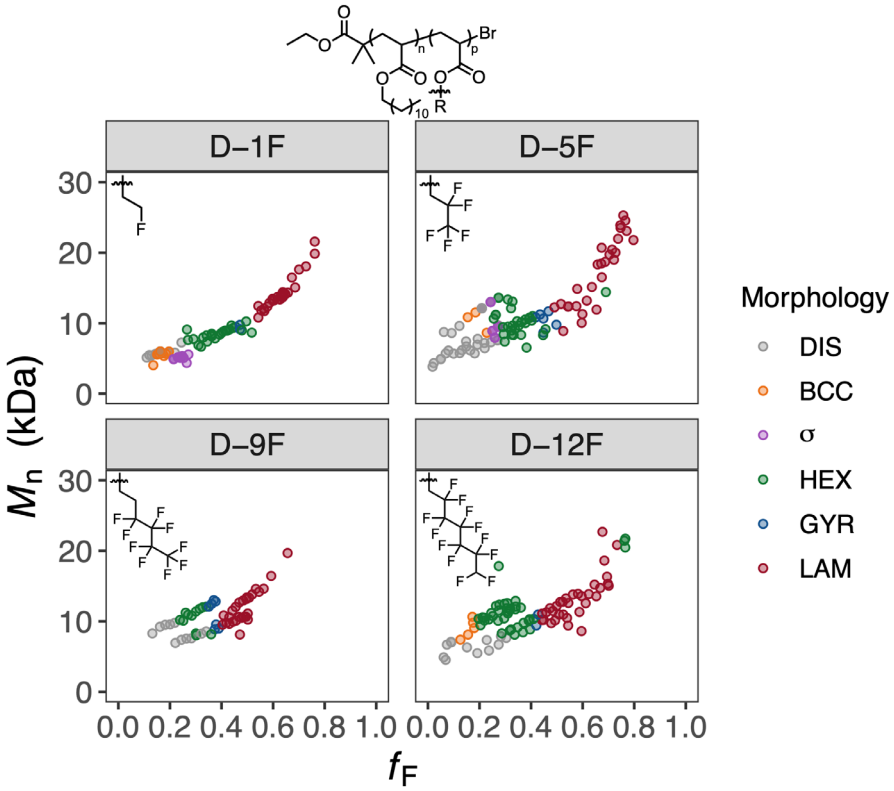

FIGURE 1 | A library of 364 well-ordered diblock copolymers derived from the synthesis and separation of only 16 parent copolymer samples was used to evaluate the machine-learning algorithms developed herein. These data sets were recently reported in *Physical Review Materials* and are reproduced here for clarity [11]. Overall block copolymer molecular weight is denoted by M_n. Volume fraction of the semi-fluorinated block is denoted by f_F. Color indicates the morphology as determined by manual analysis of SAXS data.

Excessive smoothing can obscure these critical small or overlapping peaks, whereas insufficient smoothing might misinterpret noise as peaks. The smoothing level in Gaussian process regression is controlled by the range and nugget parameter of the covariance [42], which are determined using the maximum likelihood estimate from all DIS samples in our data set. Comprehensive details on the formulae for smoothing and peak detection are provided in the Supporting Information.

Figure 3a shows a t-distributed stochastic neighbor embedding (t-SNE) plot of logarithmic SAXS intensities curves [43], where each point represents an individual curve, color-coded by morphology. Utilizing the entire 1D intensity scattering curves leads to poor separation in t-SNE visualization, likely due to subtle differences in overall scattering patterns across various morphologies as well as baseline noise interference. To address these challenges, we employ X-ray scattering theory to construct physics-informed morphological features (PIMF), which are automatically extracted and input into the machine-learning model rather than using the full scattering curve. Specifically, the analysis of the location of the first three peaks proved highly diagnostic of distinct morphologies. Figure 3b,c, respectively, show the first primary scattering peak location against the ratio between the first and second peaks and the ratio between the second and third peaks. These plots illustrate that physics-informed features, such as ratios between peak locations, facilitate a more effective structural determination than traditional methods which analyze the entire scattering curve. These refined features—the location of the first primary scattering peak and pairwise ratios of the first three peak locations—hold particular significance for researchers seeking to analyze novel materials that may lack extensive long-range order. We limit our analysis to the first three peaks here, as using fewer yields insufficient information and extending beyond often proves impractical, given that over 30% of samples lack a fourth peak, leading to less accurate classification.

In addition to peak locations, the width and sharpness of the peaks are also critical for identifying morphologies in block copolymers. Typically, disordered samples exhibit a single broad and low-intensity peak, whereas ordered block copolymers often display multiple well-resolved and sharp reflections. However, there are instances, such as in HEX, BCC, and σ phases, where a single peak might emerge due to overlapping peaks or being close to the order–disorder boundary. To address these challenges, we include in our model the width of the first primary scattering peak and its sharpness, measured by the second derivative of the peak.

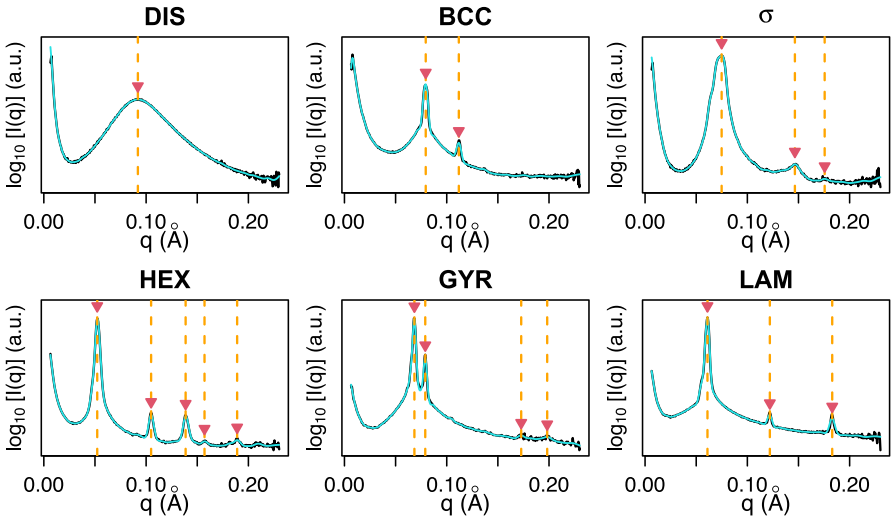

FIGURE 2 | Panels display a representative example of each morphology. Black and blue curves represent the original log intensity and the smoothed intensity curves, respectively. Detected peak locations are highlighted by red triangles and marked with orange dashed lines.

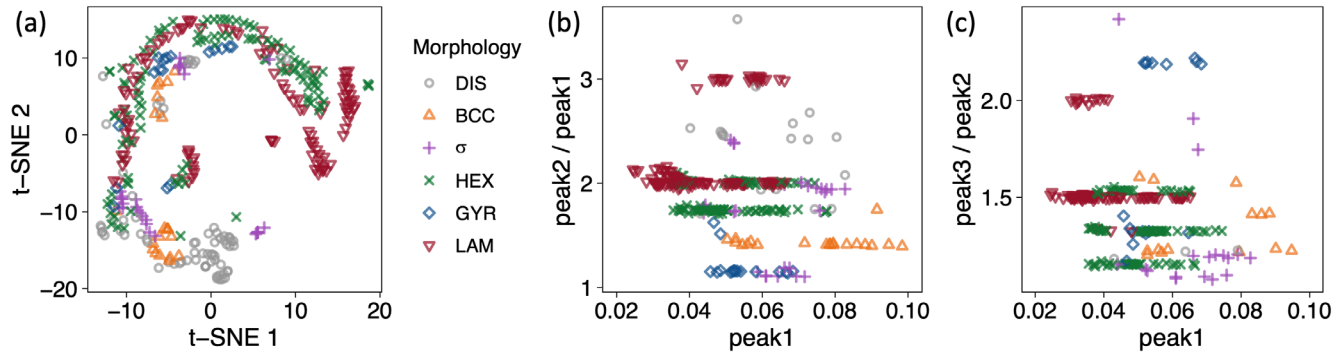

FIGURE 3 | (a) T-distributed stochastic neighbor embedding (t-SNE) for visualization of original log intensity. (b) First peak location versus the ratio between second and first peak. (c) First peak location vs. the ratio between third and second peak.

Our approach effectively reduced the original intensity curves to six salient features: the pairwise ratio of the first three peak locations, and the location, width, and sharpness of the first primary scattering peak. These physics-informed features substantially simplify the high-dimensional scattering intensity patterns to enable facile structure determination. Importantly, these features are independent of the specific chemical composition of the material, underscoring the versatility of this approach as a universal algorithm for phase identification in a wide range of block copolymers. Including more morphologically relevant features can improve predictive accuracy in similar tasks, whereas adding a large number of less informative features may also degrade the predictive accuracy of the model. Significantly, this method's ability to extract essential morphological information from scattering patterns enables its broad applicability across a diverse chemical landscape extending to various block chemistries, molecular weights, disparities, number of blocks, and polymer architectures. Notably, well-ordered materials of a specific morphology exhibit identical series of scattering reflections, regardless of polymer identity. The q values of these reflections may shift based on polymer molecular weight, but the overall pattern remains consistent. By removing the impact of specific chemical characteristics, our model offers a powerful tool for analyzing diverse block copolymer systems efficiently.

2.3 | Phase Identification of New Block Copolymer Chemistries

To demonstrate the effectiveness of our method with PIMF in rapidly analyzing novel block copolymers, we first trained our model using manually identified SAXS patterns from three block copolymer chemistries to predict the morphologies of a fourth material group. Due to the increased conformational asymmetry of D-1F and D-5F block copolymers, these materials exhibit a window of σ stability that vanishes in more conformationally symmetric D-9F and D-12F materials [11]. Holding out these two groups as the training data could lead to inaccurate predictions due to the limited number of σ samples. Thus, we focused on two testing scenarios: using D-1F, D-5F, and D-12F to predict the morphology of D-9F, and using the other three block copolymer libraries to predict D-12F.

We present the results of out-of-sample prediction by the physics-informed morphological features in the random forest (PIMF-RF) model [32] and compare the accuracy with the CD-RF model that relies on volume fraction, total molar mass, temperature, and monomer identity of the diblock copolymers [29] and the Curve-RF model that directly uses the original SAXS curve. We also employ bagging and the gradient boosting model with our feature set, which provides similar results detailed in the Supporting Information.

Table 1 presents the classification performance of different approaches when predicting the phase behavior of group D-9F. Notably, our approach only has two misclassifications out of 57 samples, significantly outperforming the 22 and 38 misclassifications observed in the CD-RF and Curve-RF methods, respectively. Compared with using the entire SAXS curve as the input in the Curve-RF method, the six physically formed features in PIMF-RF are informative in reducing the high dimensionality of the SAXS curve, and they lead to much higher accuracy in predictions. Furthermore, the CD-RF approach does not rely on the use of SAXS data, and it is shown to be accurate for predicting phases with monomers that appear in the training data set [29]. The results indicate that the automated analysis of SAXS data in PIMF-RF effectively enhances the accuracy of predicting the phase of the novel monomers not in the training data set.

TABLE 1 | Prediction results of random forest models on predicting D-9F block copolymer morphologies using PIMF-RF, CD-RF, and Curve-RF.

	Pred	DIS	BCC	σ	HEX	GYR	LAM	# Misclassified
PIMF-RF								
True	DIS	10		1				1
	HEX				12			0
	GYR					7		0
	LAM				1		26	1
CD-RF								
True	DIS	4	2		5			7
	HEX				12			0
	GYR				7			7
	LAM				4	4	19	8
Curve-RF								
True	DIS	10	1					1
	HEX	8		2	2			10
	GYR			2	3		2	7
	LAM	11			9		7	20

The RF model classifies phases based on the predicted probability for each phase, where the phase with the highest probability is selected as the predicted outcome. The low maximum predicted probability indicates high uncertainty of the method, which can be used to control the predictive error [44]. Figure 4a represents the maximum predicted probabilities for all test samples, overlaid on violin plots that illustrate the distribution of these probabilities [45]. The plot reveals that the two misclassified samples in group D-9F have probabilities below 0.55, indicating low prediction confidence. Conversely, correctly classified samples generally display maximum probabilities exceeding 0.7, demonstrating a higher confidence level. This pattern suggests that inspecting a small subset of test samples with low maximum predicted probabilities could enhance the accuracy close to 100%, which will substantially reduce the manual review workload compared to examining all samples.

Subsequently, we applied the same PIMF-RF approach to predict the morphologies of the D-12F block copolymer library with the classification results detailed in Table 2. Of the 91 D-12F block copolymers, only three are inaccurately predicted, an improvement over the CD-RF [29] and Curve-RF approaches, which misclassify 21 and 71 morphologies, respectively. Figure 4b shows that the maximum predicted probabilities of three misclassified samples in the D-12F library are all below 0.5, reinforcing the earlier observation that reviewing samples with low predicted probabilities can reduce prediction errors of our approach. Significantly, our PIMF-RF model accurately identifies the morphology of 142 out of 147 block copolymers (96.6% accuracy) consisting of novel monomers not present in the training data set, highlighting the power of this synergistic approach combining machine learning with automated chromatographic separation toward accelerated materials discovery.

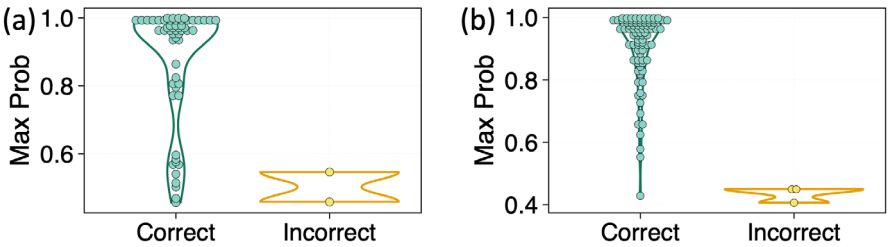

FIGURE 4 | Maximum predicted probability for correctly and incorrectly predicted morphologies for (a) the D-9F block copolymer library; (b) the D-12F block copolymer library using PIMF-RF.

TABLE 2 | Results of random forest models on predicting D-12F block copolymer morphologies using PIMF-RF, CD-RF, and Curve-RF.

	Pred	DIS	BCC	σ	HEX	GYR	LAM	# Misclassified
PIMF-RF								
True	DIS	12						0
	BCC		5					0
	HEX		3		36			3
	GYR					2		0
	LAM						33	0
CD-RF								
True	DIS	8	1		3			4
	BCC	4	1					4
	HEX	4	1		29	1	4	10
	GYR				1		1	2
	LAM					1	32	1
Curve-RF								
True	DIS	9			2		1	3
	BCC				3	2		5
	HEX	8		1	6	5	19	33
	GYR	1		1				2
	LAM	6			22		5	28

FIGURE 5 | Maximum predicted probability for correctly and incorrectly predicted samples of all diblock copolymers using the physics-informed features for (a) the original samples; (b) revised samples after correcting three mislabeled detected by our method.

2.4 | Phase Identification of Mixed Block Copolymer Chemistries

To further evaluate the robustness of our high-throughput analysis method under diverse training conditions, we combined all four classes of diblock copolymers and implemented a five-fold cross-validation strategy. The data set was randomly divided into five folds, with each iteration using four-folds for training and one-fold for testing. Employing the PIMF-RF approach results in 25 misclassified morphologies out of 364. Upon careful review, we found three block copolymers initially mislabeled during manual morphology assignment were identified and correctly predicted by our approach. These three SAXS patterns are shown in the Supporting Information. Figure 5a displays the maximum predicted probabilities for all morphologies, with mislabeled materials highlighted in red. We retrained our model of the revised data set after correcting these labels, and the number of misclassified morphologies using our PIMF-RF method reduced to 20, which is much smaller than the 62 misclassifications observed with the CD-RF approach [29] and 85 misclassified samples in Curve-RF approach using the same training data. Detailed results are provided in the Supporting Information.

Figure 5b presents the violin plot of the maximum predicted probabilities of all held-out test samples by the PIMF-RF method. Analysis shows that when using a threshold of 0.6 for the predicted probabilities, we need to examine only 32 SAXS patterns (9% of the data) to achieve a predictive accuracy of 98.4%. Increasing this threshold to 0.8 requires inspecting 69 samples (19% of the data), which results in remarkably accurate predictions with 100% accuracy. This pattern suggests that strategically examining a small subset of SAXS patterns with lower predicted probabilities enables the ML approach to attain high predictive accuracy.

3 | Conclusions

In summary, we have developed a high-throughput, material-independent workflow for characterizing the morphology of unidentified block copolymers using a physics-informed machine-learning algorithm that rapidly analyzes SAXS data. Leveraging automated chromatography and controlled polymerization to generate high-quality and expansive data sets for training, our novel machine-learning approach rapidly extracted 1D scattering patterns and achieved a remarkable 95% accuracy in identifying the nanoscale morphologies of diblock copolymers for new monomers not in the training data set. By integrating this automated synthesis method with advanced feature extraction and uncertainty analysis by machine-learning techniques, we established a robust, high-throughput framework to enhance the predictive accuracy and efficiency of analyzing polymer phase behavior. This advancement paves the way for accelerated materials discovery and a deeper comprehension of structure–property relationships in soft materials. This research inspires several directions, including predicting a new phase rarely appearing in the training data and inverse design to achieve a given polymer phase by optimizing chemical identity, composition, and external conditions. This approach is not limited to diblock copolymers but also holds the potential for extension to higher-order multiblock copolymers and non-linear architectures, indicating its broad applicability in materials science. We envision that the synergy between laboratory automation and machine learning will further catalyze the development of laboratories of the future, ushering in new possibilities for research and innovation in polymer science.

Data Availability Statement

All curated data are available in the manuscript or the Supporting Information. The data and code that support the findings of this study are openly available in GitHub at: (https://github.com/UncertaintyQuantification/automated_polymer_phase_identification).

References

1. R. A. Segalman, "Patterning With Block Copolymer Thin Films," *Materials Science and Engineering: R: Reports* 48 (2005): 191–226.

2. R. Liggins and H. Burt, "Polyether–Polyester Diblock Copolymers for the Preparation of Paclitaxel Loaded Polymeric Micelle Formulations," *Advanced Drug Delivery Reviews* 54 (2002): 191.

3. C. M. Bates and F. S. Bates, "50th Anniversary Perspective: Block Polymers Pure Potential," *Macromolecules* 50 (2017): 3–22.

4. S. Lee, M. J. Bluemle, and F. S. Bates, "Discovery of a Frank-Kasper σ Phase in Sphere-Forming Block Copolymer Melts," *Science* 330 (2010): 349–353.

5. M. W. Bates, J. Lequieu, S. M. Barbon, et al., "Stability of the A15 Phase in Diblock Copolymer Melts," *Proceedings of the National Academy of Sciences* 116 (2019): 13194–13199.

6. C. Zhang, D. L. Vigil, D. Sun, et al., "Emergence of Hexagonally Close-Packed Spheres in Linear Block Copolymer Melts," *Journal of the American Chemical Society* 143 (2021): 14106–14114.

7. S. Cui, E. A. Murphy, W. Zhang, et al., "Cylinders-in-Undulating-Lamellae Morphology From ABC Bottlebrush Block Terpolymers," *Journal of the American Chemical Society* 146 (2024): 6796–6805.

8. H. Lee, S. Kwon, J. Min, et al., "Thermodynamically Stable Plumber's Nightmare Structures in Block Copolymers," *Science* 383 (2024): 70–76.

9. C. Sinturel, F. S. Bates, and M. A. Hillmyer, "High χ-Low N Block Polymers: How Far Can We Go?," *ACS Macro Letters* 4 (2015): 1044–1050.

10. R. Gupta, M. Misra, W. Zhang, et al., "Topological Frustration as a New Parameter to Tune Morphology Revealed Through Exploring the Continuum Between ABC 3-Arm Star and Linear Triblock Polymers," *Macromolecules* 54 (2021): 4401–4411.

11. E. A. Murphy, S. J. Skala, D. Kottage, et al., "Accelerated Discovery and Mapping of Block Copolymer Phase Diagrams," *Physical Review Materials* 8 (2024): 015602.

12. R. V. Garcia, E. A. Murphy, N. J. Sinha, et al., "Tailoring Writability and Performance of Star Block Copolypeptides Hydrogels Through Side-Chain Design," *Small* 19 (2023): 2302794.

13. S. Bae and K. G. Yager, "Chain Redistribution Stabilizes Coexistence Phases in Block Copolymer Blends," *ACS Nano* 16 (2022): 17107–17115.

14. E. Helfand, "Block Copolymer Theory. III. Statistical Mechanics of the Microdomain Structure," *Macromolecules* 8 (1975): 552–556.

15. L. Leibler, "Theory of Microphase Separation in Block Copolymers," *Macromolecules* 13 (1980): 1602–1617.

16. M. W. Matsen and M. Schick, "Stable and Unstable Phases of a Diblock Copolymer Melt," *Physical Review Letters* 72 (1994): 2660–2663.

17. T. Shefelbine, M. E. Vigild, M. Matsen, et al., "Core−Shell Gyroid Morphology in a Poly(Isoprene-Block-Styrene-Block-Dimethylsiloxane) Triblock Copolymer," *Journal of the American Chemical Society* 121, no. 37 (1999): 8457–8465.

18. C. A. Tyler and D. C. Morse, "Orthorhombic Fddd Network in Triblock and Diblock Copolymer Melts," *Physical Review Letters* 94 (2005): 208302.

19. B. Vorselaars, J. U. Kim, T. L. Chantawansri, G. H. Fredrickson, and M. W. Matsen, "Self-Consistent Field Theory for Diblock Copolymers Grafted to a Sphere," *Soft Matter* 7 (2011): 5128.

20. G. Khaira, M. Doxastakis, A. Bowen, et al., "Derivation of Multiple Covarying Material and Process Parameters Using Physics-Based Modeling of X-Ray Data," *Macromolecules* 50 (2017): 7783–7793.

21. T. Aoyagi, "Deep Learning Model for Predicting Phase Diagrams of Block Copolymers," *Computational Materials Science* 188 (2021): 110224.

22. S. Zhao, T. Cai, L. Zhang, W. Li, and J. Lin, "Autonomous Construction of Phase Diagrams of Block Copolymers by Theory-Assisted Active Machine Learning," *ACS Macro Letters* 10 (2021): 598–602.

23. M. E. Deagen, B. Dalle-Cort, N. J. Rebello, T.-S. Lin, D. J. Walsh, and B. D. Olsen, "Machine Translation Between BigSMILES Line Notation and Chemical Structure Diagrams," *Macromolecules* 57, no. 1 (2023): 42–53.

24. J. A. Mysona, P. F. Nealey, and J. J. de Pablo, "Machine Learning Models and Dimensionality Reduction for Prediction of Polymer Properties," *Macromolecules* 57 (2024): 1988–1997.

25. M. G. Wessels and A. Jayaraman, "Computational Reverse-Engineering Analysis of Scattering Experiments (CREASE) on Amphiphilic Block Polymer Solutions: Cylindrical and Fibrillar Assembly," *Macromolecules* 54 (2021): 783–796.

26. C. M. Heil, A. Patil, A. Dhinojwala, and A. Jayaraman, "Computational Reverse-Engineering Analysis for Scattering Experiments (CREASE) With Machine Learning Enhancement to Determine Structure of Nanoparticle Mixtures and Solutions," *ACS Central Science* 8 (2022): 996–1007.

27. C. M. Heil, Y. Ma, B. Bharti, and A. Jayaraman, "Computational Reverse-Engineering Analysis for Scattering Experiments for Form Factor and Structure Factor Determination ("p (q) and s (q) Crease")," *JACS Au* 3 (2023): 889–904.

28. K. Vaddi, K. Li, and L. D. Pozzo, "Metric Geometry Tools for Automatic Structure Phase Map Generation," *Digital Discovery* 2 (2023): 1471–1483.

29. A. Arora, T.-S. Lin, N. J. Rebello, S. H. Av-Ron, H. Mochigase, and B. D. Olsen, "Random Forest Predictor for Diblock Copolymer Phase Behavior," *ACS Macro Letters* 10 (2021): 1339–1345.

30. C. K. Williams, and C. E. Rasmussen, *Gaussian Processes for Machine Learning* (Cambridge, MA: MIT Press, 2006).

31. M. Gu, X. Wang, and J. O. Berger, "Robust Gaussian Stochastic Process Emulation," *Annals of Statistics* 46 (2018): 3038–3066.

32. L. Breiman, "Random Forests," *Machine Learning* 45 (2001): 5–32.

33. L. Breiman, "Bagging Predictors," *Machine Learning* 24 (1996): 123–140.

34. J. H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," *Annals of Statistics* 29 (2001): 1189–1536.

35. E. A. Murphy, C. Zhang, C. M. Bates, and C. J. Hawker, "Chromatographic Separation: A Versatile Strategy to Prepare Discrete and Well-Defined Polymer Libraries," *Accounts of Chemical Research* 57 (2024): 6306.

36. C. Zhang, M. W. Bates, Z. Geng, et al., "Rapid Generation of Block Copolymer Libraries Using Automated Chromatographic Separation," *Journal of the American Chemical Society* 142 (2020): 9843.

37. E. A. Murphy, Y.-Q. Chen, K. Albanese, et al., "Efficient Creation and Morphological Analysis of ABC Triblock Terpolymer Libraries," *Macromolecules* 55 (2022): 8875–8882.

38. B. Oschmann, J. Lawrence, M. W. Schulze, et al., "Effects of Tailored Dispersity on the Self-Assembly of Dimethylsiloxane–Methyl Methacrylate Block Co-Oligomers," *ACS Macro Letters* 6 (2017): 668–673.

39. N. J. Szymanski, C. J. Bartel, Y. Zeng, M. Diallo, H. Kim, and G. Ceder, "Adaptively Driven X-Ray Diffraction Guided by Machine Learning for Autonomous Phase Identification," *npj Computational Materials* 9 (2023): 31.

40. J.-W. Lee, W. B. Park, J. H. Lee, S. P. Singh, and K.-S. Sohn, "A Deep-Learning Technique for Phase Identification in Multiphase Inorganic Compounds Using Synthetic XRD Powder Patterns," *Nature Communications* 11 (2020): 86.

41. M. Gu, X. Liu, X. Fang, and S. Tang, "Scalable Marginalization of Correlated Latent Variables With Applications to Learning Particle Interaction Kernels," *New England Journal of Statistics in Data Science* 1 (2023): 172–186.

42. M. Gu, J. Palomo, and J. O. Berger, "RobustGaSP: Robust Gaussian Stochastic Process Emulation in R," *R Journal* 11, no. 1 (2019): 112.

43. L. Van der Maaten and G. Hinton, "Visualizing Data Using t-SNE," *Journal of Machine Learning Research* 9 (2008): 2579.

44. X. Fang, M. Gu, and J. Wu, "Reliable Emulation of Complex Functionals by Active Learning With Error Control," *Journal of Chemical Physics* 157 (2022): 214109.

45. J. L. Hintze and R. D. Nelson, "Violin Plots: A Box Plot-Density Trace Synergism," *American Statistician* 52 (1998): 181–184.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.