Universal Phase Identification of Block Copolymers from Physics-informed Machine Learning

Xinyi Fang,[†,‡,#] Elizabeth A. Murphy,[¶,§,‡,#] Phillip A. Kohl,[§,‡] Youli Li,[§,‡] Craig J. Hawker,[*,¶,§,‖,‡] Christopher M. Bates,[*,§,¶,‖,⊥,‡] and Mengyang Gu[*,†,‡]

†*Department of Statistics and Applied Probability, University of California, Santa Barbara, California 93106, United States*

‡*BioPACIFIC Materials Innovation Platform, University of California, Santa Barbara, California 93106, United States*

¶*Department of Chemistry and Biochemistry, University of California, Santa Barbara, California 93106, United States*

§*Materials Research Laboratory, University of California, Santa Barbara, California 93106, United States*

‖*Materials Department, University of California, Santa Barbara, California 93106, United States*

⊥*Department of Chemical Engineering, University of California, Santa Barbara, California 93106, United States*

#*These authors contributed equally*

E-mail: hawker@mrl.ucsb.edu; cbates@ucsb.edu; mengyang@pstat.ucsb.edu

Abstract

Block copolymers play a vital role in materials science due to their widely studied self-assembly behavior. Traditionally, exploring the phase space of block copolymer self-assembly and associated structure–property relationships involves iterative synthesis, characterization, and theory, which is labor-intensive both experimentally and computationally. Here, we introduce a versatile, high-throughput workflow towards materials discovery that integrates controlled polymerization and automated chromatographic separation with a novel physics-informed machine learning (ML) algorithm for the rapid analysis of small-angle X-ray scattering (SAXS) data. Leveraging the expansive and high-quality experimental datasets generated by 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 manual analysis, achieving out-of-sample predictive accuracy of around 95% for both novel and existing materials in the training dataset. 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 and achieve near-perfect predictions. In summary, 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.

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 (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 archi-

tecture.[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, the presence of coexisting phases, and extraneous atmospheric scattering.[9–13] At best, analyzing SAXS datasets 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—self-consistent field theory (SCFT).[14–19] Like with contemporary experiments, the extensive computational resources required to perform large numbers of simulations is often rate-limiting. In addressing these challenges, recent advances in machine learning have created new opportunities to automate structural analysis by detecting patterns in datasets with high dimensionality.[20–23] For example, Olsen and colleagues leveraged a random forest model to analyze the phase behavior of diblock copolymers using experimental data mined from the literature[24]. While 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 identities, 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 from small-angle X-ray scattering (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,[25,26] we incorporate experimental uncertainties across different wave vector magnitudes, q. Second, the use of physics-informed features transforms full-intensity curves into a small set of informative features, thereby enhancing the effectiveness and accuracy of our predictive models. This combination of advanced statistical modeling and informed feature reduction significantly enhances the robustness and reliability of our predictive models applied toward accelerated 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 different classification models, such as random forest,[27] bagging,[28] and gradient boosting[29] 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 block copolymers from a very small number of as-synthesized samples.[6,7,11,30–33] Together, the 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 dataset. This method can further identify and correct data that was initially mislabeled as a result of human error. Through uncertainty assessment, the accuracy of our method approaches nearly 100% after scrutinizing about 15% of the data with the highest uncertainty in predictive labels, marking a crucial advancement towards automated, high-throughput laboratories integrated with SAXS systems.[34,35] The data and code used in this paper are publicly available (`https://github.com/UncertaintyQuantification/automated_polymer_phase_identification`).

Results and discussion

Generating block copolymer libraries

To evaluate the efficacy of our new machine-learning algorithm as applied to analyzing small-angle X-ray scattering (SAXS) data, we leveraged a large experimental dataset (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),

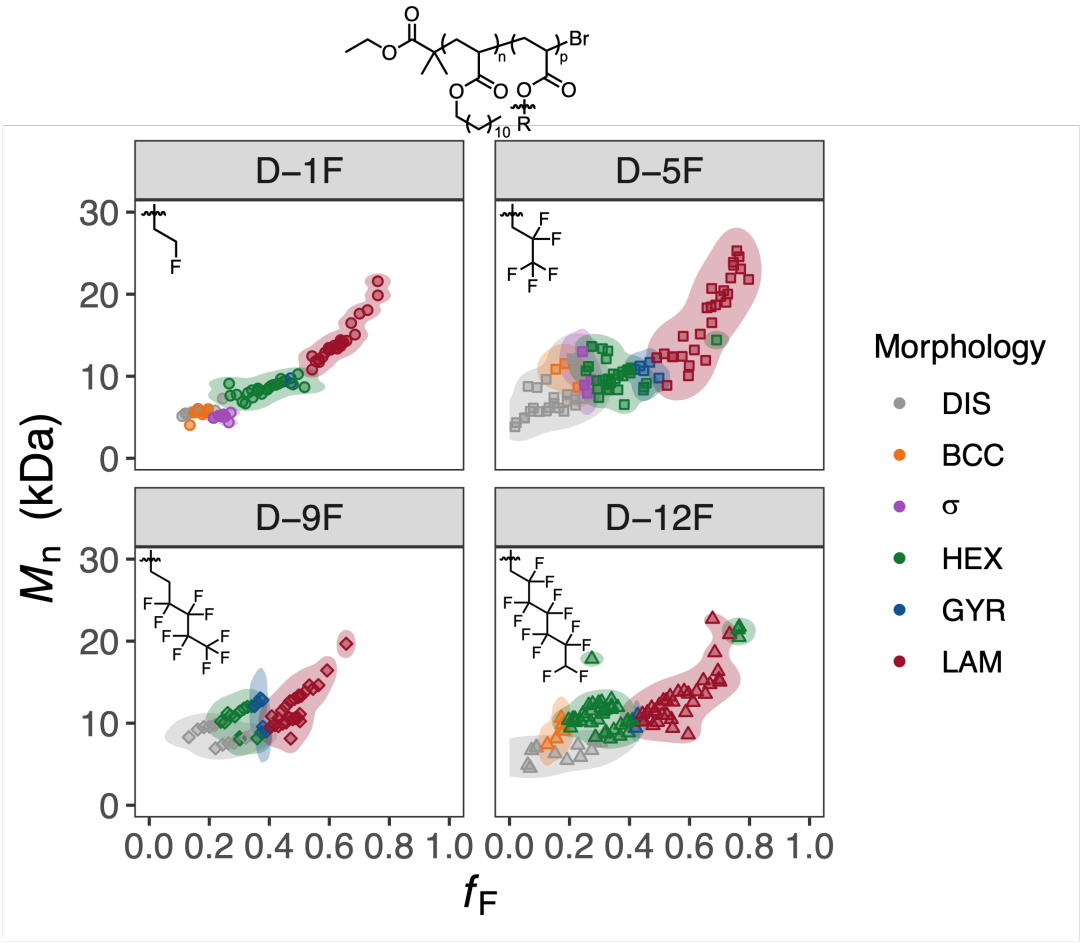

Figure 1: A library of 364 well-ordered diblock copolymers derived from the synthesis and separation of only 16 parent copolymer samples was used to evaluate the machine-learning algorithms developed herein. These datasets were recently reported in *Physical Review Materials* and are reproduced here for clarity.[11] In this figure color indicates the morphology as determined by manual analysis of SAXS data.

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]

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

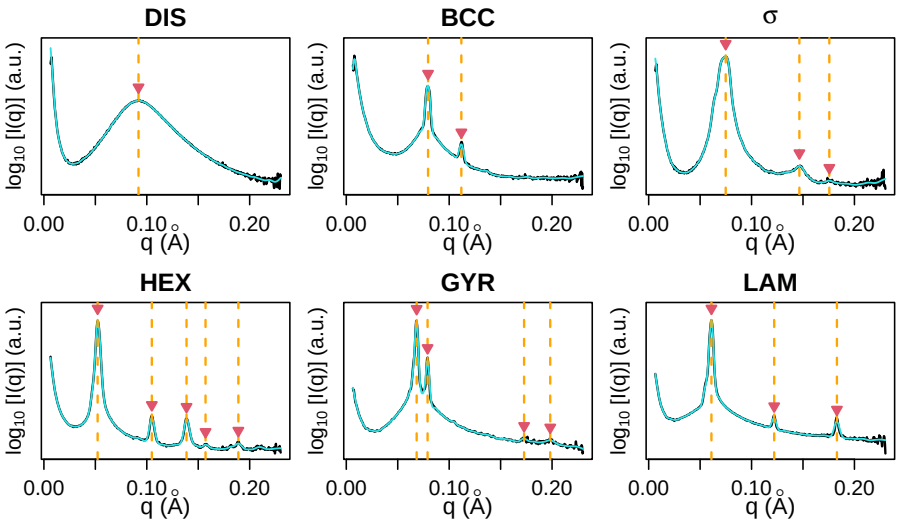

Figure 2: Panels display a representative example of each morphology. Black and blue curves represent the original log intensity and the smoothed intensity curves, respectively. Detected peak locations are highlighted by red triangles and marked with orange dashed lines.

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.[36] We have innovated an approach that integrates experimental uncertainties into the Kalman filter for 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 exhibit significant peak broadening despite analyzing the same nanostructure.[4] 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,[37] which are determined using the maximum likelihood estimate from all DIS samples in our dataset. Comprehensive details on the formulae for smoothing and peak detection are provided in SI.

In Figure 3, the left panel shows a t-distributed stochastic neighbor embedding (t-SNE) plot of logarithmic SAXS intensities curves,[38] 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. The middle and right panels of Figure 3 respectively show the first primary scattering peak location against the distance between the first and second peaks, and the second peak location against the distance between the second and third peaks. These plots illustrate that physics-informed features, such as ratios between peak locations, facilitate a more effective morphology interpretation than a method directly utilizing the entire scattering curve without accounting for physical information, as different morphologies can be better separated and identified by physics-informed features. 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.

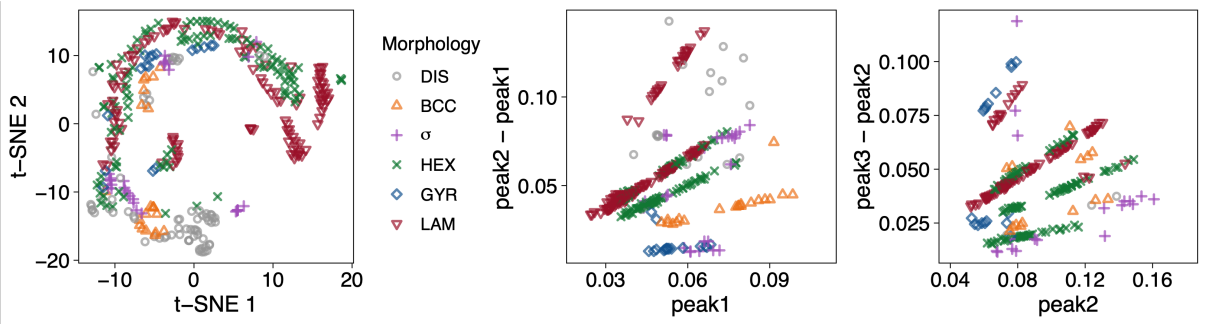

Figure 3: Left panel: T-distributed stochastic neighbor embedding (t-SNE) for visualization of original log intensity. Middle panel: First peak location vs. distance between second and first peak. Right panel: Second peak location vs. distance between third and second peak.

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 the width of the first primary scattering peak and its sharpness, measured by the second derivative of the peak, into the model.

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.

Phase identification of new block copolymer chemistries

To demonstrate the versatility 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 D-9F and D-12F.[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[27] 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[24]. We also employ bagging and the gradient boosting model with our feature set, which provides similar results detailed in the SI. The upper portions of Table 1 present classification accuracy comparing the performance of different approaches, both applied to predict the phase behavior of group D-9F. Notably, our approach only has 2 misclassifications out of 57 samples, significantly outperforming the 22 misclassifications observed in the CD-RF method.

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.[39] The left panel of Figure 4 represents the maximum predicted probabilities for all test samples, overlaid on violin plots that illustrate the distribution of these probabilities.[40] The plot reveals that the 2 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.

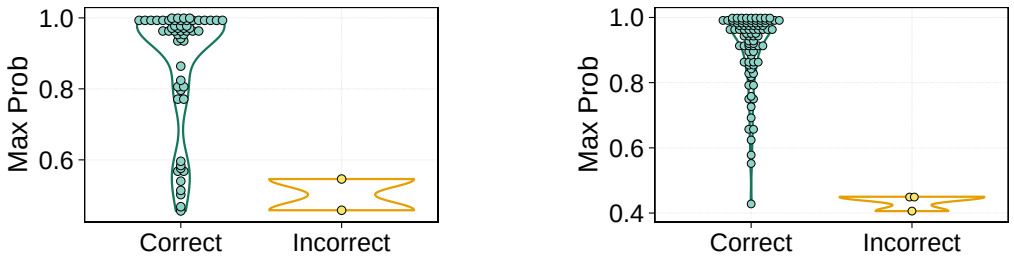

Figure 4: Maximum predicted probability for correctly and incorrectly predicted morphologies for the D-9F block copolymer library (left) and the D-12F block copolymer library (right) using PIMF-RF.

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 the lower two

Table 1: Prediction results of random forest models on predicting D-9F and D-12F block copolymer morphologies using PIMF-RF versus CD-RF. The table is segmented into four sections: the performance of (i) PIMF-RF on D-9F, (ii) CD-RF on D-9F, (iii) PIMF-RF on D-12F, and (iv) CD-RF on D-12F.

PIMF-RF		pred					
		DIS	σ	HEX	GYR	LAM	# Misclassified
true	DIS	10	1				1
	HEX			12			0
	GYR				7		0
	LAM			1		26	1

CD-RF		pred					
		DIS	BCC	HEX	GYR	LAM	# Misclassified
true	DIS	4	2	5			7
	HEX			12			0
	GYR			7			7
	LAM			4	4	19	8

PIMF-RF		pred					
		DIS	BCC	HEX	GYR	LAM	# Misclassified
true	DIS	12					0
	BCC		5				0
	HEX		3	36			3
	GYR				2		0
	LAM					33	0

CD-RF		pred					
		DIS	BCC	HEX	GYR	LAM	# Misclassified
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

sections of Table 1. Of the 91 D-12F block copolymers, only three are inaccurately predicted, an improvement over the CD-RF approach, which misclassified 21 morphologies[24]. The right panel of Figure 4 shows that the maximum predicted probabilities of 3 misclassified samples in the D-12F library are all below 0.5, reinforcing the earlier observation that reviewing samples with low predicted probabilities can avoid prediction errors from the machine learning 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 dataset, highlighting the power of this synergistic approach combining machine learning with automated chromatographic separation towards accelerated materials discovery.

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 5-fold cross-validation strategy. The dataset 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 SI. The left panel of Figure 5 displays the maximum predicted probabilities for all morphologies, with mislabeled materials highlighted in red. We retrained our model of the revised dataset 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[24] using the same training data (detailed results are shown in the SI).

The right panel of Figure 5 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.

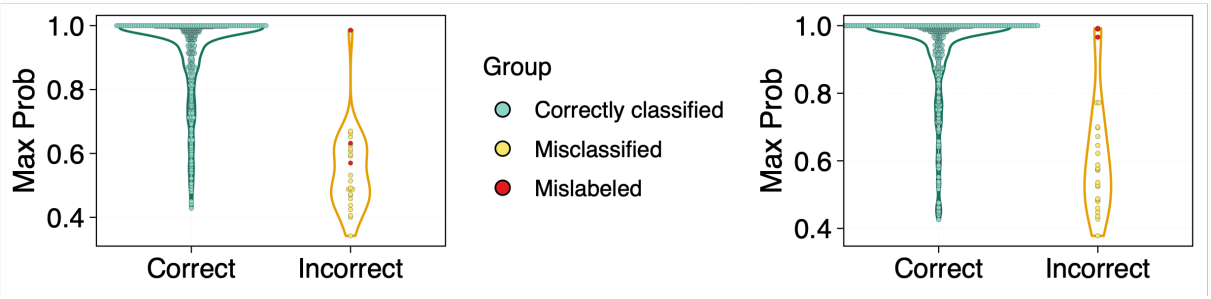

Figure 5: Maximum predicted probability for correctly and incorrectly predicted samples of all diblock copolymers using the physics-informed features for the original samples (left panel) and revised samples after correcting three mislabeled detected by our method (right panel).

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 large datasets for training, our novel ML approach rapidly analyzed 1D scattering patterns and achieved a remarkable 95% accuracy in identifying the nanoscale morphologies of diblock copolymers. By integrating this automated synthesis method with advanced statistical models and machine learning techniques, we established a robust, high-throughput framework to enhance the predictive accuracy and efficiency of analyzing polymer phase behavior. This advance paves the way for accelerated materials discovery and a deeper comprehension of structure-property relationships.[41] 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.

Author Contributions

X.F. and E.A.M. contributed equally. The manuscript was written by X.F., E.A.M., C.J.H., C.M.B., and M.G.. Experiments were designed by X.F., E.A.M., P.A.K., Y.L., C.J.H., C.M.B., and M.G. and performed by X.F., E.A.M., and P.A.K.. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

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. X.F. acknowledges the support of the UC Multicampus Research Programs and Initiatives (MRPI) program under Grant No.

M23PL5990.

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`)

Supporting Information Available

Methodology of block copolymer synthesis, automated chromatography, data collection, peak detection, feature extraction, and model development; numerical results of bagging and boosting models; analysis of misclassification instances.

References

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

(2) Liggins, R.; Burt, H. Polyether–Polyester Diblock Copolymers for the Preparation of Paclitaxel Loaded Polymeric Micelle Formulations. *Advanced Drug Delivery Reviews* **2002**, *54*, 191–202.

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

(4) 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.

(5) Bates, M. W.; Lequieu, J.; Barbon, S. M.; Lewis III, R. M.; Delaney, K. T.; Anastasaki, A.; Hawker, C. J.; Fredrickson, G. H.; Bates, C. M. Stability of the A15 phase

in Diblock Copolymer Melts. *Proceedings of the National Academy of Sciences* **2019**, *116*, 13194–13199.

(6) Zhang, C.; Vigil, D. L.; Sun, D.; Bates, M. W.; Loman, T.; Murphy, E. A.; Barbon, S. M.; Song, J.-A.; Yu, B.; Fredrickson, G. H.; Whittaker, A. K.; Hawker, C. J.; Bates, C. M. Emergence of Hexagonally Close-Packed Spheres in Linear Block Copolymer Melts. *Journal of the American Chemical Society* **2021**, *143*, 14106–14114.

(7) Cui, S.; Murphy, E. A.; Zhang, W.; Zografos, A.; Shen, L.; Bates, F. S.; Lodge, T. P. Cylinders-in-Undulating-Lamellae Morphology from ABC Bottlebrush Block Terpolymers. *Journal of the American Chemical Society* **2024**, *146*, 6796–6805.

(8) Lee, H.; Kwon, S.; Min, J.; Jin, S.-M.; Hwang, J. H.; Lee, E.; Lee, W. B.; Park, M. J. Thermodynamically Stable Plumber's Nightmare Structures in Block Copolymers. *Science* **2024**, *383*, 70–76.

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

(10) Gupta, R.; Misra, M.; Zhang, W.; Mukhtyar, A.; Gido, S. P.; Ribbe, A.; Escobedo, F. A.; Coughlin, E. B. Topological Frustration as a New Parameter to Tune Morphology Revealed through Exploring the Continuum between ABC 3-Arm Star and Linear Triblock Polymers. *Macromolecules* **2021**, *54*, 4401–4411.

(11) 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. *Physical Review Materials* **2024**, *8*, 015602.

(12) Garcia, R. V.; Murphy, E. A.; Sinha, N. J.; Okayama, Y.; Urueña, J. M.; Helgeson, M. E.; Bates, C. M.; Hawker, C. J.; Murphy, R. D.; Read de Alaniz, J. Tailoring Writability and Performance of Star Block Copolypeptides Hydrogels through Side-Chain Design. *Small* **2023**, *19*, 2302794.

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

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

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

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

(17) Shefelbine, T.; Vigild, M. E.; Matsen, M.; Hajduk, D.; Hillmyer, M.; Cussler, E.; Bates, F. Core-Shell Gyroid Morphology in a Poly(isoprene-*block*-styrene-*block*-dimethylsiloxane) Triblock Copolymer. *Journal of the American Chemical Society* **1999**, *121*, 8457–8465.

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

(19) Vorselaars, B.; Kim, J. U.; Chantawansri, T. L.; Fredrickson, G. H.; Matsen, M. W. Self-Consistent Field Theory for Diblock Copolymers Grafted to a Sphere. *Soft Matter* **2011**, *7*, 5128–5137.

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

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

(22) Deagen, M. E.; Dalle-Cort, B.; Rebello, N. J.; Lin, T.-S.; Walsh, D. J.; Olsen, B. D.

Machine Translation between BigSMILES Line Notation and Chemical Structure Diagrams. *Macromolecules* **2023**, *57*, 42–53.

(23) Mysona, J. A.; Nealey, P. F.; de Pablo, J. J. Machine Learning Models and Dimensionality Reduction for Prediction of Polymer Properties. *Macromolecules* **2024**,

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

(25) Rasmussen, C. E. *Gaussian processes for machine learning*; MIT Press, 2006.

(26) Gu, M.; Wang, X.; Berger, J. O. Robust Gaussian stochastic process emulation. *Annals of Statistics* **2018**, *46*, 3038–3066.

(27) Breiman, L. Random forests. *Machine learning* **2001**, *45*, 5–32.

(28) Breiman, L. Bagging predictors. *Machine learning* **1996**, *24*, 123–140.

(29) Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine. *Annals of statistics* **2001**, 1189–1232.

(30) Murphy, E. A.; Zhang, C.; Bates, C. M.; Hawker, C. J. Chromatographic Separation: A Versatile Strategy to Prepare Discrete and Well-Defined Polymer Libraries. *Accounts of Chemical Research* **2024**, *57*, 6306–6310.

(31) Zhang, C.; Bates, M. W.; Geng, Z.; 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. *Journal of the American Chemical Society* **2020**, *142*, 9843–9849.

(32) Murphy, E. A.; Chen, Y.-Q.; Albanese, K.; Blankenship, J. R.; Abdilla, A.; Bates, M. W.; Zhang, C.; Bates, C. M.; Hawker, C. J. Efficient Creation and Mor-

phological Analysis of ABC Triblock Terpolymer Libraries. *Macromolecules* **2022**, *55*, 8875–8882.

(33) Oschmann, B.; Lawrence, J.; Schulze, M. W.; Ren, J. M.; Anastasaki, A.; Luo, Y.; Nothling, M. D.; Pester, C. W.; Delaney, K. T.; Connal, L. A.; McGrath, A. J.; Clark, P. G.; Bates, C. M.; Hawker, C. J. Effects of Tailored Dispersity on the Self-Assembly of Dimethylsiloxane–Methyl Methacrylate Block Co-Oligomers. *ACS Macro Letters* **2017**, *6*, 668–673.

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

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

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

(37) Gu, M.; Palomo, J.; Berger, J. O. RobustGaSP: Robust Gaussian Stochastic Process Emulation in R. *The R Journal* **2019**, *11*, 112–136.

(38) Van der Maaten, L.; Hinton, G. Visualizing Data Using t-SNE. *Journal of Machine Learning Research* **2008**, *9*, 2579–2605.

(39) Fang, X.; Gu, M.; Wu, J. Reliable Emulation of Complex Functionals by Active Learning with Error Control. *The Journal of Chemical Physics* **2022**, *157*, 214109.

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

(41) Xu, Z.; Dong, Q.; Li, W. Architectural Design of Block Copolymers. *Macromolecules* **2024**,

TOC Graphic