ACS POLYMERS Au
AN OPEN ACCESS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/polymerau
Perspective

Navigating the Expansive Landscapes of Soft Materials: A User Guide for High-Throughput Workflows

Published as part of ACS Polymers Au virtual special issue "2023 Rising Stars".

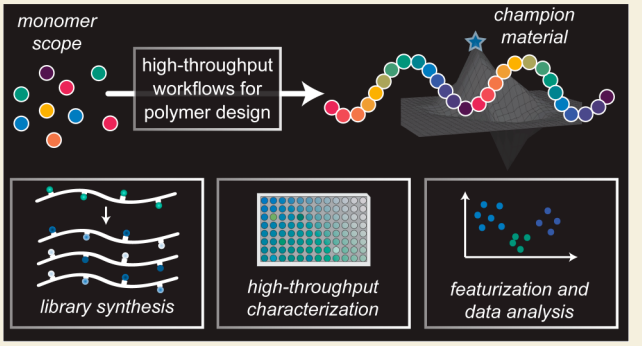
Erin C. Day,[†] Supraja S. Chittari,[†] Matthew P. Bogen, and Abigail S. Knight*

Cite This: *ACS Polym. Au* 2023, 3, 406–427

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Synthetic polymers are highly customizable with tailored structures and functionality, yet this versatility generates challenges in the design of advanced materials due to the size and complexity of the design space. Thus, exploration and optimization of polymer properties using combinatorial libraries has become increasingly common, which requires careful selection of synthetic strategies, characterization techniques, and rapid processing workflows to obtain fundamental principles from these large data sets. Herein, we provide guidelines for strategic design of macromolecule libraries and workflows to efficiently navigate these high-dimensional design spaces. We describe synthetic methods for multiple library sizes and structures as well as characterization methods to rapidly generate data sets, including tools that can be adapted from biological workflows. We further highlight relevant insights from statistics and machine learning to aid in data featurization, representation, and analysis. This Perspective acts as a "user guide" for researchers interested in leveraging high-throughput screening toward the design of multifunctional polymers and predictive modeling of structure–property relationships in soft materials.



KEYWORDS: *high-throughput screening, polymer libraries, soft materials design, machine learning, bioinspired*

1. INTRODUCTION

From water purification to medical devices, functional polymers provide solutions to global challenges owing to their diverse structures and chemistries that can engender environmental resilience and targeted functionality.[1] Synthetic macromolecules introduce unique chemical compositions, dispersities, and architectures beyond those of native biopolymers, thereby affording new avenues toward improved functions. However, identifying structure–property and resulting structure–function relationships in these materials remains challenging. Diverse efforts toward modern multifunctional materials, from antifungal activity[2] to drug delivery injectables,[3] have underscored the intricacies and complexities of macromolecular properties. Nonintuitive and emergent characteristics in these materials challenge rational design, limiting the use of existing design principles and stepwise iteration. Therefore, the development of high-throughput synthetic workflows can unravel these complex relationships and advance materials research.

While high-throughput screening has been prevalent in biological sciences for decades[4] (e.g., for drug candidates[5] and protein-based coatings[6]), its application to synthetic macromolecules is comparatively recent. Some biological strategies can be leveraged to expedite the development of synthetic polymers with desirable properties, including library design principles, characterization methods, and analysis platforms. However, challenges unique to the polymer community can prevent small-scale setups from readily translating to a high-throughput workflow. The past two decades have seen key advancements that address these bottlenecks. These developments span oxygen-tolerant polymerization,[7,8] automated data processing,[9] robotics,[10] and entry points to machine learning softwares.[11] The broadening accessibility of these tools has resulted in an influx of high-throughput research efforts as directly interrogating complex design questions comes within reach.

In this Perspective, we outline strategies to design high-throughput workflows for solution-phase macromolecules,

Received: September 15, 2023
Revised: November 2, 2023
Accepted: November 7, 2023
Published: December 5, 2023

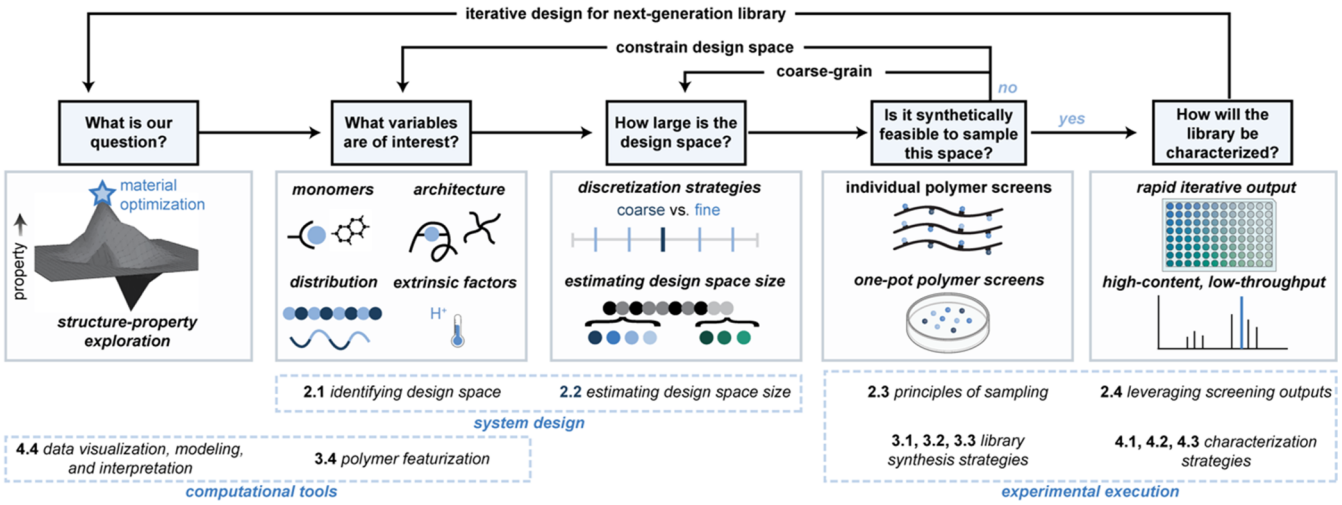


Figure 1. High-throughput screening workflow design. First, a scientific objective must be established to optimize a material or explore a structure−property relationship. Then, variables or features of interest must be chosen and discretized appropriately to result in a design space that can be feasibly sampled. A library can then be generated, screened, and the resulting characterization can be used in designing a new library and further material discovery. The relevant sections of this perspective are highlighted below the flowchart.

focusing on recent advances in library synthesis, characterization, and the use of statistics and machine learning techniques for data analysis and interpretation. Due to the diversity and depth of these fields, we seek to provide insight into how these components interface and augment one another. Further insight into high-throughput polymerizations[12] and screening of bulk polymeric materials,[13] as well as machine learning for data-driven polymer design[14−17] can be found in previous articles. Our recommendations focus on optimizing library design with respect to size (i.e., number of library members), ease of synthesis (i.e., time and purification steps required), and characterization efficiency (i.e., time per sample, simultaneous or automated measurement), as these selections are critical to maximizing experimental efficiency. Furthermore, we provide a summary of rapid and iterative characterization strategies capable of reporting on desired macromolecular properties. We conclude with an outlook on high-throughput materials development, including the need for shared databases toward designing and understanding structure−property relationships for next-generation materials to address ongoing societal challenges.

2. WORKFLOW DESIGN TO UNVEIL STRUCTURE−PROPERTY RELATIONSHIPS IN A HIGH-DIMENSIONAL DESIGN SPACE

A comprehensive understanding of macromolecular structure−property relationships offers two major advantages: first, insight into how microscopic descriptors and chemical moieties result in a macroscopic property and second, the ability to predict this property for materials in *de novo* design.[18] We herein focus on using rapid workflows to develop soft materials including dilute polymer systems, sequence-defined oligomers, and biomimetic materials, combining expertise from both synthetic and machine learning fields. We discuss approaches to the design and synthesis of libraries composed of 10^2−10^5 members, rapid characterization methods, and implementation of this information to guide further efforts.

High-throughput approaches are well-suited for systems with many tunable variables that show complex interactions with one another. The modularity inherent to many polymer systems due to structural features such as composition, sequence, architecture, and molecular weight results in a high dimensional feature space that is challenging to interrogate directly. A universal workflow for these studies can be deconstructed into a handful of steps (Figure 1). First, a scientific objective for the study must be established, often categorized as either optimization or exploration of a structure−property relationship (described further in Section 2.1). Next, features of interest must be selected; these can include variables such as material structure and extrinsic factors such as the reaction conditions and sample preparation methods (Section 2.2). Chosen features must then be appropriately bounded and discretized to estimate the size of the design space (Sections 2.2.1 and 2.2.2). With this estimate in hand, a method for library synthesis that can generate a representative fraction of the total space can be selected (Section 2.3). At this point, if it is unfeasible to sample the design space given the library size, the size of the design space can be constrained to generate a smaller study through adjusting the number of variables or further discretization. Once a design space is chosen, the library can be synthesized (Section 3), and a suite of characterization tools are available for screening (Section 4). Outputs of the characterization stage can be used to inform the design of future libraries, the generation of databases, and the synthesis of novel materials (Section 2.4).

2.1. Identifying the Desired Design Spaces: What Is Our Question?

Screening can be effective for systems with some established design principles, but complex relationships between features and target outcomes have yet to be uncovered. However, for novel systems where the role of features such as new monomers (e.g., catalytic,[19] structural[20]) or architectures (e.g., star,[21] cross-linked,[22] branched[23] polymers) are poorly defined, beginning with a small-scale library instead of a much larger screen is a valuable first step. These libraries can be rationally designed or guided by Design of Experiments to efficiently explore a feature space.[24,25] A systematic study presents an opportunity to troubleshoot the synthesis, characterization, and analysis protocols on a small-scale. Such

https://doi.org/10.1021/acspolymersau.3c00025
ACS Polym. Au 2023, 3, 406−427

prototyping will generate practical information about the system such as optimal synthesis, purification, and sample preparation methods in addition to limits of solubility. Further, preliminary functional and structural characteristics can be generated to inform rational material design or further screening. If a high-throughput approach is needed, a target objective of the screen must be chosen: (1) optimization, where a target property must be enhanced by tuning material structure or processing, or (2) exploration of a structure−property relationship, where a model can predict a property using descriptors of the system. While both these aims are related, optimization and exploration pose different challenges and benefit from different experimental and statistical tools.

The goal of optimization is to develop a high-performance material with desired property. These can encompass a specific function, such as the binding of target molecules (e.g., metals[26] or sugars[27]) or catalytic activity,[28] or structural properties such as compactness[29] or helicity.[30] Many structural features (e.g., polymer composition or architecture) and/or extrinsic descriptors (e.g., reaction conditions) can contribute to the target property and must be tuned to generate a material with optimal performance. Visualizing the structure−property relationship as a surface in a high-dimensional space would result in many "peaks" and "valleys" that correspond to high- or low-performance materials, respectively (Figure 2). The optimization approach searches for peaks within this surface, and positions of valleys are considered obstacles to avoid or overcome through strategic library design. Challenges arise as complex surfaces may often contain multiple peaks, and identification of the tallest peak (i.e., highest performing material) requires that the library describes a large amount of this surface.

In contrast, the objective of exploration is to map a structure−property relationship over the entire feature space, such that the properties of a material can be predicted for any arbitrary position in the space. Therefore, exploration demands knowledge of both the peaks and valleys within a given feature space, as well as the combinations of different features that result in changing peak or valley heights (Figure 2). Quantitative structure−activity relationship (QSAR) or structure−property relationship (QSPR) models are generated through such an exploratory search that predicts the property of a material using feature descriptors, such as material composition or architecture. Valleys are no longer inconveniences as they were in optimization; they are important sources of information in the development of an exhaustive model.

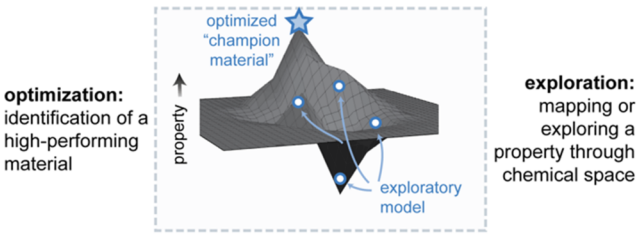


Figure 2. Objective of a high-throughput screen over a complex design surface (gray) falls into either optimization (left) or exploration (right) categories, where the former involves identification of a high-performing "champion material" (blue star), and the latter involves mapping a structure−property relationship over the entirety of the surface (blue dots).

Exploration and optimization present distinct hurdles. The challenge of optimization lies in reaching the global maximum (i.e., the highest-performance material) due to the presence of local maxima that are difficult to navigate away from or "activity cliffs" where similar materials have very different performances.[31] To decrease the experimental burden, a material that reaches a local maximum can instead be selected as a champion material based on relative improvement over another material or exceeding a desired threshold. For example, optimization workflows in drug discovery frequently involve setting threshold values for target objectives such as potency and cytotoxicity.[32] Statistical tools such as adaptive sampling are useful in navigating to materials that fulfill multiple objectives (also known as multiobjective optimization) while reducing experimental burden.

The challenge of exploration is the requirement for large data sets. Experimentally relevant questions tend to span extremely high-dimensional spaces that are difficult to sample effectively through library generation—often referred to as the "curse of dimensionality."[33,34] Insufficient data may not fit a regression model or result in poor predictiveness. Experimental tools in library generation (further discussed in Section 3) can assist in the rapid synthesis of a larger, more representative sample size to reach all corners of the feature space.

2.2. Feature Selection and Estimation of Design Space Size: What Are Variables of Interest and How Large Is the Design Space?

Identification of relevant descriptors for samples within a library is critical for extracting information from the greater surface (Figure 3a). Common intrinsic descriptors of a material include composition (e.g., hydrophobic,[35] functional,[36] charged,[37] and stimuli-responsive[38]), architecture (e.g., cross-linked,[22] branched,[23] and star[39]), sequence patterning,[40] and molecular weight. However, descriptors that are extrinsic to the material itself, such as sample preparation protocols[41−44] or substrate choice for a polymer catalyst,[45,46] can also be important variables to probe with a library.

2.2.1. Variable Discretization. Once individual features are selected, each feature can be subdivided into a set of intervals that span the desired range. Consider an example study with a library of random copolymers with the objective of determining polymer compositions that are capable of binding a target, such as in protein stabilization.[47,48] Monomers with different chemistries as well as different polymer architectures and molecular weights are all synthetically accessible. The incorporation of a selected monomer in a polymer is a continuous variable—a random copolymer can be synthesized with any arbitrary percentage of a monomer, so this variable must be both restricted and discretized. Bounds on this variable would be the minimum and maximum percentage of monomer allowed in the overall composition, which can be determined by factors like the polymer solubility in a solvent of choice (Figure 3b, left). Discretization of this variable is the selection of interval percentages that the monomer changes by, for example increasing within a selected range in steps of 10 versus 30 mol % incorporation (Figure 3b, right).

Variable discretization can depend on both limits of detection and practical constraints. For example, differences between a polymer with 20 mol % of a given monomer versus 25 mol % may be negligible in an assay output, or monomer

https://doi.org/10.1021/acspolymersau.3c00025
ACS Polym. Au 2023, 3, 406−427

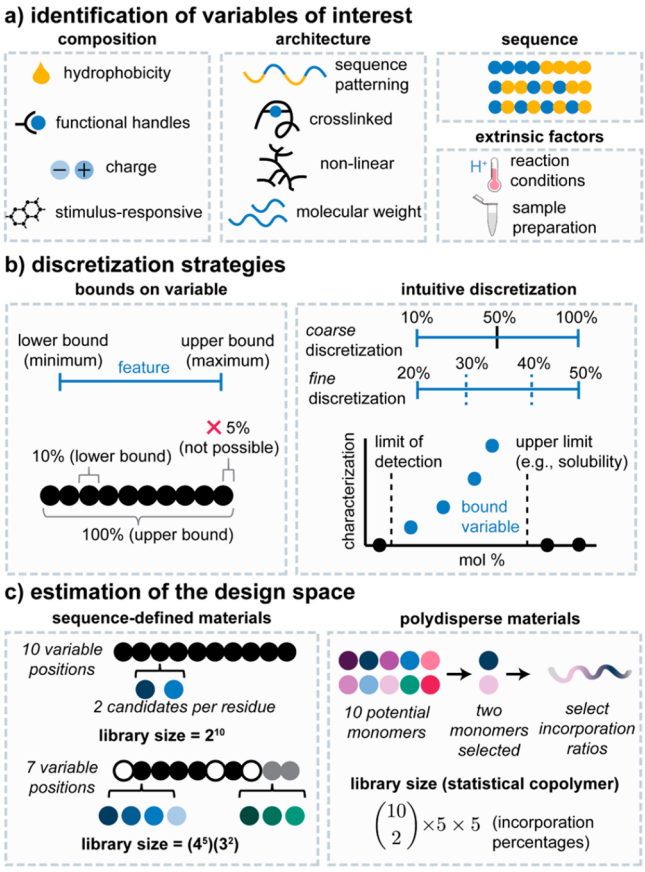


Figure 3. Featurization strategies and estimating library size. (a) Common variables for macromolecule libraries include composition, molecular weight, architecture, sequence patterning, and extrinsic factors. (b) Variables must be bounded (left) and discretized (right) based on physical limitations. (c) Estimation of the size of the design space for sequence-defined materials (left) or polydisperse materials (right).

incorporation quantification can be limited by resolution via NMR integration. Given inherent variability within experimental systems, the choice of discretization should yield materials with reproducible results by sample and measurement replicates. In both cases, discretization to smaller increments may not reveal additional insight. Further, discretization into finer segments increases the number of members within the design space, thereby increasing the library size required to sample it. Discretization to 1 mol % may be suitable in limited-scope libraries (e.g., subsequent iterations of a screening workflow), or if a 1 mol % change has a disproportionately large effect on a property. If choosing a small discretization is appropriate and necessary, the minimum and maximum bounds on the variable can be chosen strategically to minimize the total number of possible values this variable can take on. If the sensitivity of the target outcome to a given variable is not well-known, an initial sparse library can be generated that spans a large range of values, and subsequent libraries can focus further on any range of interest (Figure 3b). Thus, taking care to appropriately restrict and discretize each variable with chemical intuition in mind will improve the efficacy of a screen.

Variables may not always be continuous—they can take on discrete values as well. As a second example, a sequence-defined oligomer library is screened for a feature, such as antifungal properties.[49] In the first example, monomer incorporation was a continuous feature in disperse copolymers that was subdivided into discrete percentages. In a system where oligomers are composed of 10 discrete residues, generating an oligomer with only 5 mol % of one monomer is not synthetically possible, as the smallest discretization available is 10 mol %. However, the principles of bounding and discretizing variables still apply. To reduce the size of a library of 10-mer oligomers, the library can be synthesized with monomer incorporation discretized to 20 mol % increments instead (i.e., groups of two residues can vary).

2.2.2. Estimating Design Space Size. Once features are selected, combinatorics can be used to calculate the size of a design space, aiding in the estimation of sample representativeness: a comparison of the library size relative to that of the full design space. Increasing the complexity of a design space also increases the necessary sampling and experimental burden. Consider a statistical copolymer library, where combinations of ten functional monomers are investigated. Each polymer sample is designed to be comprised of two different functional monomers of the ten and a consistent filler monomer for the remaining composition (Figure 3c). If the mole percent incorporation of two of the monomers is fixed, the total number of polymers is described by the combination function 10 choose 2, or 45. If the two functional monomers can take on different percentages of the total polymer, we can discretize those percentages into an arbitrary number. For this example, say that there are five possible percentages for each monomer; the total number of polymers now is 45 × 5 × 5 = 1125. In a second case study with sequence-defined macromolecules synthesized by a modular synthetic strategy, the total design space is represented by y^x, where x is the number of variable positions and y is the possible residues at each position. If 10-residue oligomers are being synthesized with two possible residues at each position, the size of the design space would be 2^{10} or 1024 members. If 10-mer oligomers are being synthesized with seven variable positions, where five positions could be one of four residues, and the other two could be one of three residues, the total design space would be $(4^5)(3^2)$ or 9216 members (Figure 3c). A 100-member library would be more effective at sampling the first design space (approximately 10% sampling) than the second design space (approximately 1% sampling). By keeping library design principles in mind from objective selection through data interpretation, data collection becomes more efficient and effective.

2.3. Principles of Sampling: Is It Feasible to Sample the Desired Space?

The sample representativeness, or how well a chemical library represents the larger high-dimensional space,[50] determines how well the library will achieve the intended goal of identifying global trends or reaching an optimum. The sample size and representativeness are directly related to the ability to validate a hypothesis or fit a model. In general, benchmarks on sampling sizes, such as the minimal percentage of a design space that must be sampled to fulfill an optimization or exploration objective, are typically not known at the start of library design and vary from system to system. For example, in the case of fitting data to a machine learning model, a heuristic sampling guideline (>5%) was suggested, but we emphasize that this value is intended only as a starting point and not a

https://doi.org/10.1021/acspolymersau.3c00025
ACS Polym. Au 2023, 3, 406−427

definitive threshold.[51] Often, the representativeness of the library can only be assessed in *post hoc* analysis.[50,52]

Insufficient library sizes can result in a design space being sampled ineffectively, making it challenging to draw conclusions (Figure 4a, left). In the design of a target binding polymer, if the design space is very large, low sampling may miss potential high-performers or be too small to elucidate important relationships. Further, if the target behavior is low frequency (i.e., only a small fraction of polymers in a design space are adequate binders), an exhaustively large library will need to be synthesized to discover them. This example is the imbalanced data problem, where certain features or classes (i.e., poor binders) are overrepresented and information from minority classes (i.e., good binders) is important but hard to access (Figure 4a, center). A final challenge in high-throughput workflows is that libraries may contain candidate molecules that are unusable. For example, a structure−property model may predict that highly hydrophobic polymers will show the highest propensity for protein-like self-assembly. However, these candidates may be insoluble in aqueous conditions and therefore unusable for the desired application (Figure 4a, right). When many features, both structural and extrinsic, are relevant, predicting which materials are unusable is challenging intuitively.

The most straightforward approach to each of the above three challenges—imbalanced data, insufficient library size, and unusable outputs—is increasing sampling. A larger library will encompass more of the design space and mitigate some of these issues. Traditional sampling approaches are "space-filling" in that they span the entire design space and include simple random, grid, Latin hypercube,[53] and sobol sampling.[54] However, these approaches may demand a library size that may be experimentally unfeasible to synthesize and characterize. In these cases, adaptive sampling, also called active sampling or response-adaptive designs, (Section 2.4) presents an alternative strategy. Adaptive sampling does not require searching the entire design space directly but instead samples with a "feedback loop" between experimental results and subsequent sampling. Important feature interactions are "learned" and candidate materials with high predicted performance are suggested. Virtual screening (Section 2.4) is an alternative method, where a large library is characterized through computationally inexpensive simulations to uncover target materials at low experimental burden. Additional strategies to treat imbalanced data sets include further statistical methods.[55]

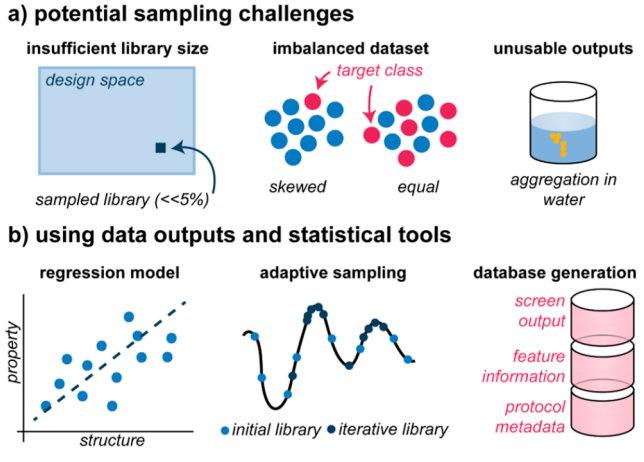


Figure 4. Potential sampling challenges and harnessing data outputs. (a) Potential sampling challenges arise with insufficient library size (left), overrepresentation of certain data classes in imbalanced data sets (center), and practical constraints on samples such as insolubility in water (right). (b) Outputs of screening workflows include development of structure−property regression models (left), iterative library design using adaptive sampling (center), and database generation of screening outputs, feature information, and protocol metadata (right).

2.4. Modeling and Leveraging Screening Outputs: How Can This Library Be Characterized?

Following library synthesis and characterization (further discussed in Sections 3 and 4), the subsequent data set can be fit to a model (Figure 4b, left). If a known physical model exists, these principles can be used to fit data directly, as in with scattering[56] or diffusion.[57] For a data-first approach, various statistical models can instead be easily implemented through premade software packages, such as Scikit-learn in Python.[11] Two primary types of models exist: regression and classification. Regression models are used when outputs can take on any value. However, we may determine a material is "good" or "poor" if the value falls above or below a selected threshold. Then, a classification model is more appropriate as output values fall into a set of predefined classes. Apart from these two model types, model learning can also be either supervised or unsupervised. Supervised learning involves first fitting a model on a set of training data with known outputs or classifications and then predicting on data it has not seen, out-of-sample data. In contrast, unsupervised learning instead finds groupings directly in a data set where classifications or outputs are unknown.

In the first step of model development, many different models are fit to the same data set to determine the best performance.[16,58] Model performance is traditionally quantified through prediction error such as root-mean-square error (RMSE), where 0 is theoretical perfect performance in a noise-free data set. Additional metrics such as "discovery" scores are being developed to evaluate a model's capability to propose new high-performing materials.[59] Further details on supervised and unsupervised learning models, model training, validation, evaluation, and interpretation can be found in a user-guide to machine learning for materials design by Gormley and co-workers.[16] Other helpful resources include the software QSARINS, which focuses on multiple linear regression modeling and includes tools for data preprocessing, validation, outlier detection, and visualization[60] and polyBERT, an end-to-end machine learning pipeline for polymer informatics and optimization.[61]

Different model types necessitate different data set sizes. Neural nets are powerful nonlinear models consisting of "neurons" organized and interconnected in complex, versatile architectures.[62] Therefore, these models are "data hungry" and a large data set (thousands to millions of data points) is usually demanded.[63] Other models, such as random forest approaches, where decision trees are trained on different subsets of data,[16,64] can readily accommodate smaller data sets on the order of hundreds of members more relevant to some library synthesis approaches. Even smaller data sets (<100 members) may only be fit using linear models. Further understanding of how small data sets can be fit to machine learning models is the focus of recent work.[65]

In some cases, directly interrogating a large and complex design space directly is impractical. Instead, several smaller

https://doi.org/10.1021/acspolymersau.3c00025
ACS Polym. Au 2023, 3, 406−427

libraries can be synthesized sequentially, with insights from each informing the next through active sampling (Figure 4c, center).[66,67] Practical considerations, such as solubility limits or interactions between different features, are typically not known at the outset. Active sampling can more rapidly accommodate these restrictions and avoid the synthesis of a large library where a significant fraction may be unusable. Machine learning and Bayesian approaches are popular active learning schemes that have been successfully applied to diverse chemical challenges including the synthesis of metallic alloys[68,69] and nanoparticles,[70] drug discovery,[71,72] catalyst development,[73] and the evaluation of properties of bulk polymers ranging from electronic bandgap to thermal transitions.[74] Examples of high-throughput synthesis coupled to iterative sampling include the design of protein-stabilizing random copolymers using automated polymer synthesis,[47,75] the identification of ^{19}F MRI contrast agents using continuous-flow chemistry,[51] and the development of polymeric injectables for drug delivery.[3]

Virtual screening also rapidly narrows a design space through computationally inexpensive simulations.[76−78] Genetic algorithms, a type of optimization algorithm inspired by mutations and natural selection, have harnessed virtual screening to identify novel materials such as photovoltaics[79] and dielectrics[80] by optimizing property criteria such as glass transition temperature or electronic bandgap.[81] Polymer chemistry has also begun to benefit from these iterative approaches with examples including polymer sequence design toward achieving compactness,[82,83] optimization of polymeric catalysts,[84] and multiobjective discovery and optimizations.[85]

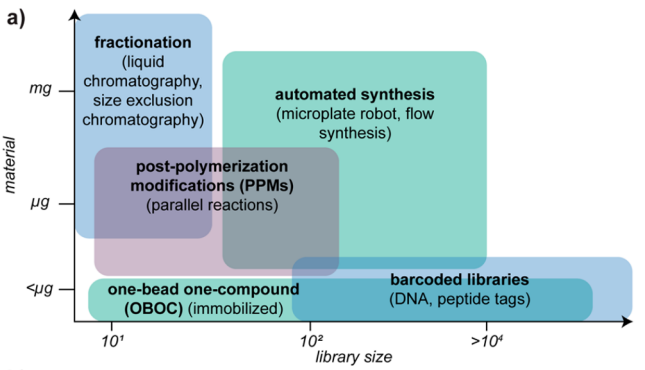
High-throughput screening results can also contribute to database generation (Figure 4c, right). Some existing polymer informatics databases[86,87] are PolyInfo,[88] Polymer Genome,[89] CRIPT,[90] Polymer Handbook,[91] and CHEMnetBASE-Polymers.[92] These databases contain characterizations of properties and relevant structural descriptors, such as monomer identity, molecular weight, and material classification. In addition to these empirical descriptors, databases can benefit from additional metadata, such as reaction conditions, material preparation methods, and calibration information, as small differences between measurements can be attributed to these metadata. As high-throughput measurements become increasingly accessible, the parallel growth of open-access databases will facilitate database benchmarking,[93] assessing how different databases perform on a similar model, and model benchmarking,[94] assessing how different models perform on the same data set. The FAIR guiding principles (findable, accessible, interoperable, and reusable) ensure that shared data are well-annotated, meet community guidelines, and are easily obtainable and verifiable, to readily support informatics.[95]

3. LIBRARY SYNTHESIS METHODS

For small sample sets (~10^1), each member of a library can be synthesized individually. Preliminary small libraries (5−25 samples) are useful to synthesize manually to troubleshoot synthetic challenges, such as different monomer reactivities, and characterization workflows. These libraries can also be constructed through a Design of Experiments workflow, where multiple parameters are varied simultaneously to rapidly uncover feature importance and interactions.[25] The efficiency of these approaches has been demonstrated in catalyst design,[96,97] where material optimization was possible with small sample sets and few iterations. Systematic studies of a selection of polymers can also be for a desired property with rapid screening, exemplified in structural characterization by Terashima and co-workers.[98−101] However, large libraries spanning a broad chemical space require different methods for efficient and high-fidelity syntheses. Three main strategies exist: (1) sequencable libraries for one-pot screening (e.g., one-bead one-compound and barcoding), (2) modification of a single synthesis (e.g., post-polymerization modification and fractionation), and (3) simultaneous, independent syntheses (e.g., parallel reactors and automation). For each of these strategies, we outline the time required, the typical size of the resulting library, and synthetic materials best suited.

3.1. Library Synthesis Methods That Enable One-Pot Screening

Libraries can be designed to enable a one-pot characterization method, such as dye-based visualization or isolation via pull-down assays, followed by sample identification. For one-bead one-compound (OBOC) screening, immobilized sequence-defined oligomers libraries (10^3−10^5) can be rapidly analyzed (Figure 5).[102] While primarily used for biopolymers such as peptides, this combinatorial library synthesis technique extends to various sequence-defined polymers and peptidomimetics, including peptoids, oligocarbamates, oligoureas, vinylogous sulfonyl peptides, peptidosulfonamides, azatides, and ketides.[103] Libraries are synthesized on a solid support, typically a cross-linked polymer resin (i.e., micron-sized beads), using the combinatorial split-and-pool synthesis method (Figure 6a).[104] Synthesis typically requires one to three hours per residue (e.g., approximately 40 h for a 20-mer library). However, OBOC systems become challenging to physically handle and manipulate with greater than ~10^6 library members. A library can be rapidly analyzed through a colorimetric or fluorometric output correlated to the property or function of interest.[105,106]



b)

	simultaneous screening	immobilized screening	single batch synthesis	parallel synthesis	property varied in library
OBOC libraries	✓	✓		✓	sequence, composition
barcoded libraries	✓				composition, unlimited
fractionation			✓		molecular weight, blocks
PPMs			✓		side-chain functionality
automated synthesis				✓	unlimited

Figure 5. Methods for efficient library generation. (a) Library types organized by the size of space (*x*-axis) and amount of material (*y*-axis) that can be screened. (b) High-throughput synthesis methods categorized by how the libraries can be screened, how the material is produced, and what macromolecular properties are varied.

https://doi.org/10.1021/acspolymersau.3c00025
ACS Polym. Au 2023, 3, 406−427

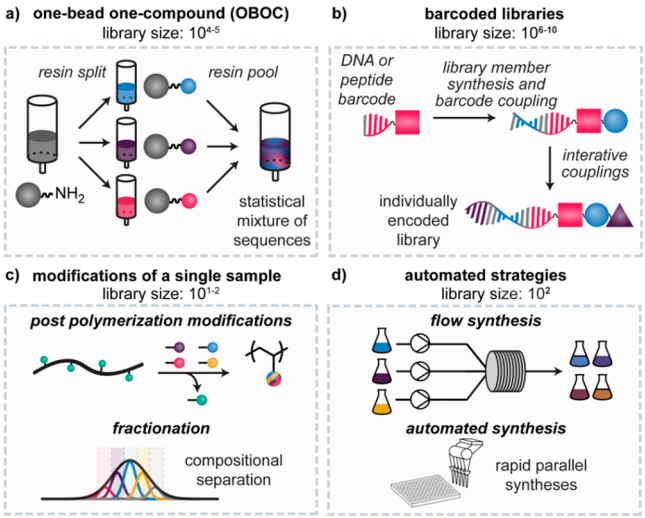


Figure 6. Methods for efficient library synthesis. Schematic of (a) split-and-pool synthesis of one-bead one-compound libraries, (b) large library synthesis with barcodes, (c) modification or separation of a single polymerization batch into a library, and (d) options for automated library generation.

Hits are identified following screening by isolating oligomers from individual beads and sequencing, typically using tandem mass spectrometry dissociation.[107−109]

Barcoded libraries also offer one-pot library screening for materials that are challenging to sequence directly. In barcoded libraries, each member has a sequencable tag covalently linked to the library member. Traditionally, DNA is used as a tag, but recent use of peptides expands the chemistries for potential library synthesis (Figure 6b).[110,111] Encoded libraries are analyzed simultaneously and selected commonly using affinity selection (e.g., binding to a target). In drug discovery, DNA encoded libraries can be greater than 10^{10} in size (Figure 5).[112] While DNA has been used to assemble polymers,[113,114] its use in encoding macromolecule libraries has been limited.[115] High-conversion synthetic steps are recommended for OBOC and barcoded libraries to eliminate the need for purification of library members.

3.2. Library Synthesis from a Single Polymerization

Post-polymerization modifications (PPMs) and fractionation allow for a single polymerization batch to yield a library while maintaining properties such as dispersity and chain-end fidelity. Subsequent screening must be done in parallel (e.g., in a well plate). Molecular weight, dispersity, and monomer patterning can be maintained by substitution or functionalization of modifiable handles on a single parent polymer to generate a library 10^{1-2} in size.[116−119] Independent syntheses may result in unwanted chemical diversity driven by reactivity ratios and small differences in monomer ratios. Additionally, PPMs can be efficiently realized at small scales in well plates, making library synthesis tractable without the need for automation or parallel processing. Champion materials can also be both generated and screened in one-pot using dynamic covalent chemistry, where functional handles exchange onto a polymer scaffold in the presence of a template material.[120,121]

Fractionation strategies are also able to separate disperse polymer batches to generate a library. Chromatographic techniques including thin layer chromatography (TLC), size-exclusion chromatography (SEC), reversed-phase liquid chromatography (LC), normal-phase chromatography, and ion exchange chromatography have been used in fractionation strategies (Figure 6c).[122−124] Although fractionation is faster than manual synthesis, only ∼10^1 can be generated with this scheme (Figure 5).

3.3. Parallel Material Synthesis

Polymer library members can also be synthesized independently using liquid-handling robots[125] or parallel reactors.[126,127] Recently, photoinduced electron/energy transfer reversible addition−fragmentation chain transfer polymerization (PET-RAFT) has enabled polymerizations in well plates using oxygen-tolerant conditions, which has enhanced the efficiency of existing parallel synthesis systems.[12,128] Purification of polymers can be done in 96-well filter plates[129] in addition to commercially available miniaturized dialysis products. Well plate-compatible library synthesis and characterization has also been demonstrated, including testing for antimicrobial activity (Figure 6d).[130] Library members (∼10^2) can be further functionalized post-polymerization prior to screening, with examples including coupling peptides to engender functionality[130] and polyethylene glycol to imbue brush-like structures.[131]

A series of polymerizations can also be run using liquid-handling robots and continuous flow reactors, available at select academic and national laboratories, such as BioPACIFIC MIP[123,132] and Argonne National Laboratory's Polybot, a self-driving laboratory for polymer development.[133] Flow synthesis has been used for anionic, cationic, radical, and ring-opening polymerizations,[134,135] in addition to sequence-defined oligomers.[136] Additionally, flow systems can synthesize diverse chemical structures such as gradient copolymers by tuning the monomer feed ratio through the reaction.[137] While flow setups are commercially available, and the challenges of building a flow setup for custom polymerizations can often outweigh the advantages.[134] While access to necessary instrumentation such as liquid-handlers or flow synthesis reactors is currently limited, the efficiency, fidelity, and ease of integration into machine learning workflows will likely continue to increase the popularity of these methods.[47,51]

3.4. Representing Polymers Using Molecular Descriptors

Section 2.2 describes intuitive methods to bound and discretize features that will parametrize a chemical design space. While a researcher has intuition about chemical information describing a library, such as monomer structure and polymer architecture, these nuances are not always included as inputs but may be critical to a predictive model. We must therefore consider methods for alternate chemical or molecular representation.

While polymer scientists have developed widely understood and accepted notations for representing chemical structures of polymers, it is challenging to mathematically describe these schematics for use in a statistical model. Unique descriptions that preserve the geometric symmetries are required to capture the full chemical complexity of a structure. Small molecule organic chemistry and drug discovery have developed methods to describe chemical structures for feature inputs that may be leveraged to describe polymer properties.[138] Molecular descriptors pertinent to polymeric materials include string notation, graph representation, and learned descriptors. String representations, such as Simplified Molecular Input Line Entry System (SMILES)[139] and International Chemical Identifier (InChI),[140] describe atoms and their connectivity within an organic molecule (Figure 7a). As these representations are not

https://doi.org/10.1021/acspolymersau.3c00025
ACS Polym. Au 2023, 3, 406−427

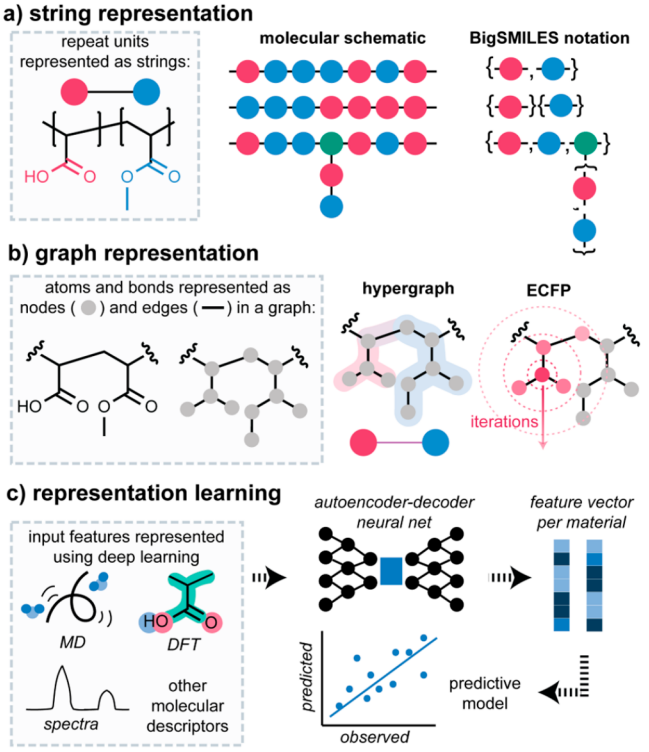


Figure 7. Overview of molecular descriptors. (a) String representation for polymer materials can use existing SMILES-type notation (left). BigSMILES notation also supports architecture representation unique to polydisperse materials (right, adapted from ref 141. Copyright 2019 American Chemical Society). (b) Graph representation uses nodes and edges to represent atoms and bonds in a molecule (left). Hypergraph (e.g., PolyGrammar) and ECFP are graph-driven representation techniques that preserve information on the connectivity of atoms in a polymer (right). (c) Representation learning is a powerful tool that can take a diverse set of inputs, including MD simulation trajectory data, electronic structure calculations from DFT, spectroscopic inputs, and other types of molecular descriptors (left). These inputs can be converted to a feature vector using deep learning (autoencoder-decoder neural net), and the feature vectors can be used to fit a predictive model (right).

unique (i.e., many strings can be written for a single structure), canonicalization is critical to ensure the same representation per structure. However, string notation is ill-suited to capture polydispersity, sequence distribution, and complex topological features specific to synthetic polymers. BigSMILES has been developed to describe polydisperse, statistical materials, such as synthetic polymers by accommodating both different monomer patterning and nonlinear architectures.[141] As it builds on widely available SMILES notation, BigSMILES presents an easily accessible alternative notation that captures structural nuances specific to the polymer community.

In addition to string notation, molecular graphs, where atoms and bonds can be represented as nodes and edges in graphs, have been successful (Figure 7b).[142−144] Extended-connectivity fingerprints (ECFPs), also known as "Morgan fingerprints",[145] are generated through a circular approach, where the "extended connectivity" of atoms is described with increasingly large radii centered around non-hydrogen atoms.[146] PolyGrammar is a polymer specific graph-type approach that is designed to support architecture and monomer chemistries.[147] Other types of molecular representation include chemical table representations, such as MDL molfiles.[148] While these techniques may be less intuitive than their string counterparts, they can represent complex polymer topologies with a high level of specificity.

Feature engineering can generate learned descriptors, which are nonintuitive but high-performing molecular representations subsequently fit to a machine learning model (Figure 7c). Representation learning algorithms automatically determine the most significant features of large data sets.[149] Neural nets, a common representation learning tool, recast features into a representation in "machine language" enabling a single graph structure to represent descriptors of interest in diverse contexts such as small organic molecules,[150] chemical reactions,[151] and crystal structures.[152] Representation learning has broad applicability, for example natural language processing algorithms have succeeded in representing complex problems in organic reactions,[153] sustainable chemistry,[154] genomics,[155] protein properties,[156] and drug discovery.[157] Improving featurization and representation for stochastic materials is critical to adapting language learning models in polymer chemistry, such that models can make accurate predictions about emergent properties. Further discussion on interpreting new representations developed by deep learning is discussed in Section 4.4.

Polymer chemists have begun to harness the potential of representation learning using chemical descriptors to fit predictive neural networks.[158−160] New graph-based representations are also being developed for polymer materials. These graphs capture the composition and architecture of polydisperse polymers, and have improved prediction of polymer properties compared to molecular descriptors alone.[161,162] For example, they have been successful in predicting catalysis conditions for ring-opening polymerizations across multiple different data sets.[84] TransPolymer, developed from the Transformer-based language processing algorithm, has been successfully trained on diverse polymer data sets for different bulk properties.[163] Additionally, graph representation has been successful in theoretical studies, such as predicting the radius of gyration (R_g) of coarse-grained model polymers with defined sequence and composition.[83]

Sequence-defined monodisperse materials can leverage techniques used for biopolymers, which benefit from relatively limited composition, well-defined sequence, and training data sets such as the Protein Data Bank. Amino acid structures can be described traditionally by SMILES or other string representations[164] or through representation learning approaches, as with the successful neural net AlphaFold.[165,166] Molecular descriptors for peptides span diverse properties such as sequence, conformation, charge, and hydrophobicity.[167−169] However, descriptors are less well developed for noncanonical amino acids and peptidomimetics (e.g., peptoids and β-peptides), and structure prediction for these materials therefore is still simulation- and experiment-driven.[170] The applicability of molecular descriptor approaches, like SMILES, in peptidomimetic systems holds promise for identifying structure−property relationships using statistical tools.

Beyond sequence descriptors, molecular dynamics (MD) and electronic structure calculations have been important inputs to representation algorithms. These calculations provide insight into microscopic dynamics and energetics otherwise opaque in experiments or chemical structure descriptors. Peptide structure−property regression frequently uses descriptor types (e.g., electrostatics and hydrophobicity) for amino acids.[171] Electronic structure descriptors from density func-

https://doi.org/10.1021/acspolymersau.3c00025
ACS Polym. Au 2023, 3, 406−427

tional theory calculations have also been successful in models of small molecule catalysis[172] and optimizing band gaps for conducting polymers.[173] Trajectory data from MD calculations have been used to analyze phase transitions[174] in materials including thermoresponsive polymers.[175,176] Additionally, MD simulation data has even reconstructed underlying pathways of protein folding producing interpretable models consisting only of a handful of states.[177,178] MD simulations therefore play important roles as features and descriptors in library design.

New open-source packages including PaDEL,[179] Mordred,[180] RDKit,[145] and ChemDes[181] have been developed to rapidly calculate different molecular descriptors and fingerprints, benefiting diverse fields. We envision that these software packages will facilitate the use of more complex descriptors in future work with polymer libraries. Model performance has been shown to depend significantly on the type of molecular descriptor,[182] influencing predictions of polymer glass transition temperature[94] and structure−activity modeling for peptides.[183] With broader accessibility of more complex molecular descriptors through these software, higher accuracy predictive models are anticipated in the future.

4. HIGH-THROUGHPUT CHARACTERIZATION AND ANALYSIS OF POLYMER LIBRARIES

Efficient and effective characterization of libraries is critical to differentiate and to optimize material performance. Factors such as the library size, amount of required material, and the specific property being tested all can affect the speed and precision of material characterization (Figure 8).[10,184] Traditional polymer characterization techniques are material- and time-intensive, acting as a bottleneck in workflows that must support large numbers of samples. High-throughput screening techniques can process upward of 10^6 samples with the development of rapid, precise measurement approaches, 384−6144 well plates, droplet microfluidics, and automated sample handling. Colorimetric and fluorometric screening are commonly used due to their convenience and simplicity, but a wider array of techniques including chromatography and scattering approaches exist. We explore how high-throughput techniques from biochemical fields[5,184−187] can be adapted for rapid and comprehensive analysis of synthetic macromolecules.

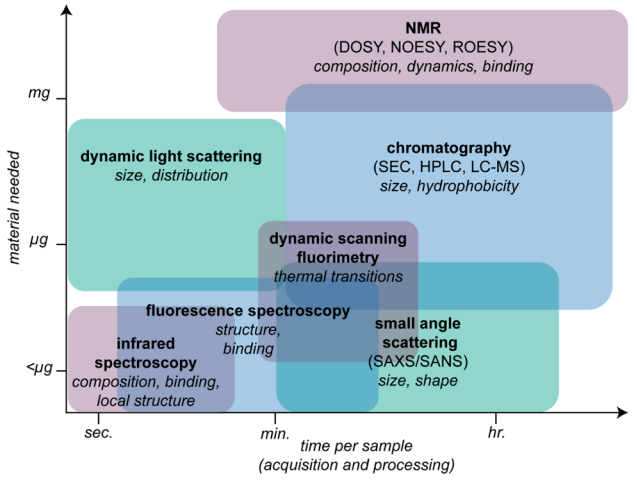


Figure 8. Characterization methods organized by data acquisition and processing times (*x*-axis) and material required (*y*-axis). Italicized are the properties that are evaluated by each technique.

4.1. Planning a Characterization Workflow

Well plate assays and instrumentation offer significant advantages for rapid screening due to miniaturization and parallel processing of samples. This allows analysis of small-scale material quantities, analogous to the capabilities of OBOC and barcoded libraries. Rapid and parallel sample processing are particularly valuable for imbalanced data sets, where the number of promising candidates is outnumbered by those with low or mediocre performance. Each characterization method has unique material limitations and time constraints that must be considered when designing a workflow (Figure 8). For these workflows, the objective is to quickly sift through the library to identify relatively high-performing candidates to be studied with more exhaustive or information-rich characterization. Well plates facilitate analysis with various analytical techniques such as spectrophotometry, fluorimetry, and light scattering, enabling the collection of multiple measurements within a single experiment. Moreover, consolidating sample preparation to a single standard container streamlines workflows, even without automated systems. This reduces the potential for error in arduous and repetitive tasks performed numerous times, such as switching containers or mixing reagents. Well plates are most commonly employed with aqueous systems or high-boiling solvents (e.g., DMSO) due to their compatibility with the plastics they are commonly made of (e.g., polystyrene, polypropylene, and cyclic olefin polymers);[188] however, glass-coated well plates are available for systems that require more specific chemical or temperature needs.

For sequence-defined polymers and other polymers that can be synthesized on a solid support, immobilized screening methods may be considered. These methods can offer improved material stability and tolerance against temperature and organic solvents, as well as a physical means of material separation, recovery, and reuse. The solid support can be cross-linked resins (i.e., beads) or surfaces (i.e., microarrays[189−191]). Immobilized assays are particularly effective in affinity screening, allowing testing against multiple targets and visualization of affinity using dyes or labels.[103,105,106,192] However, false positives may occur due to autofluorescence or nonspecific interactions between targets or ligands and the solid support. Effective strategies to mitigate these biases include lowering the loading densities, screening replicate libraries, and cross-validating hit materials with solution-phase screening.[102,107,193−196] Consequently, immobilized screening assays have proven useful in drug discovery and therapeutic development, with emerging potential for incorporation into automated workflows.[195−197]

4.2. Rapid Characterization Techniques from Both Synthetic Polymer Chemistry and Biological Polymers

A variety of polymer characterization techniques have been successful in high-throughput workflows. In the following sections, we discuss methods to incorporate traditional polymer characterization tools (e.g., chromatography and scattering) as well as to adapt techniques developed in adjacent fields that find applicability in polymer chemistry (e.g., colorimetric assays).

4.2.1. Incorporating Common Polymer Characterization Tools into High-Throughput Workflows. Some traditional characterization tools for polymeric materials that report on polymer properties, such as chromatography, scattering, and thermal analysis, have been incorporated into

https://doi.org/10.1021/acspolymersau.3c00025
ACS Polym. Au 2023, 3, 406−427

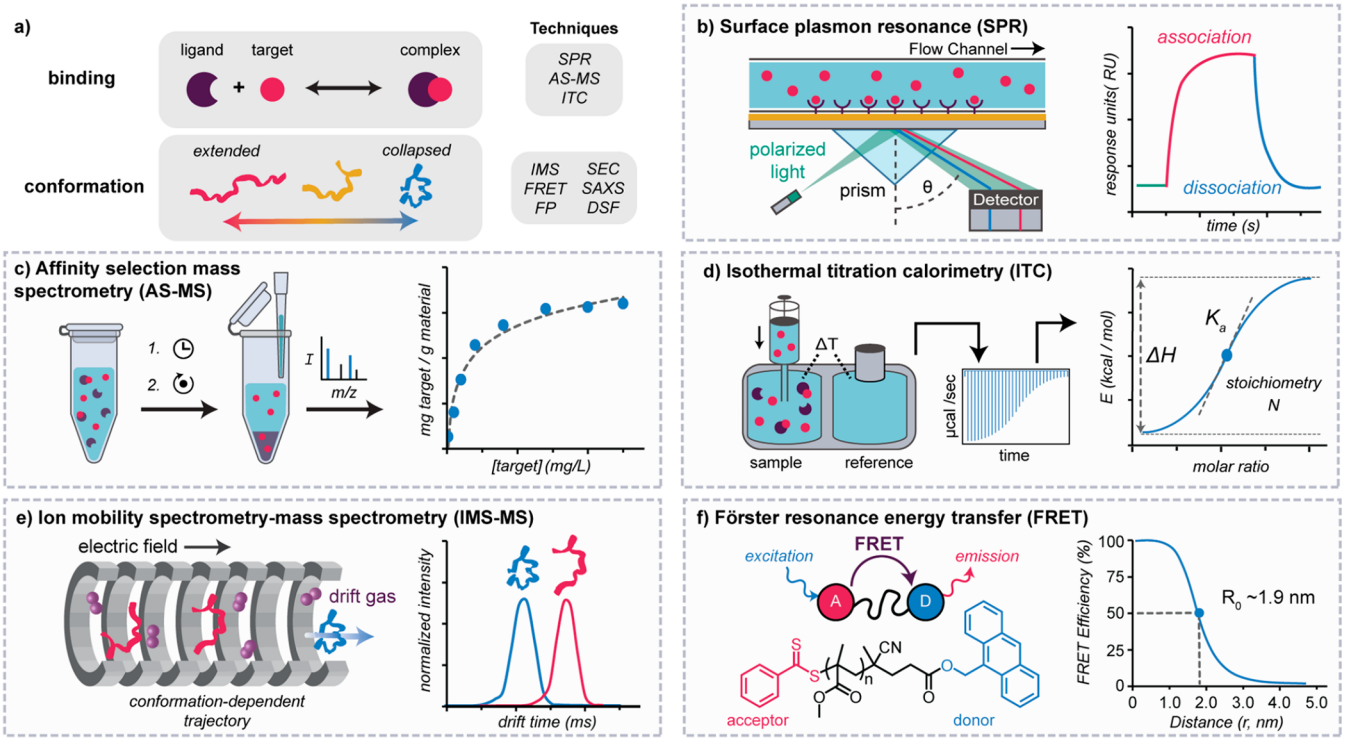


Figure 9. Biological techniques adaptable to synthetic macromolecule characterization. (a) Overview of the structural and functional characterization techniques described herein. (b) Surface plasmon resonance (SPR): the target flows through a channel over a ligand immobilized on a metal sensor surface. Changes in the refractive index upon binding provide real-time information regarding kinetics and affinities between binding partners. (c) Affinity selection mass spectrometry (AS-MS): unbound target molecules or ions are isolated physically or through dialysis techniques from a macromolecular binding partner, then quantified to determine the bound fraction across a range of target concentrations. (d) Isothermal titration calorimetry (ITC): one binding partner, typically the target, is titrated gradually into a dilute solution of the other, while the resulting heat change is measured against a reference cell. Peak integration of each binding event and subsequent curve fitting yields ΔH, binding stoichiometry (N), and association constant (K_a). (e) Ion mobility spectrometry-mass spectrometry (IMS-MS): ions are separated based on size, charge state, and collisional cross section, resolving differences in polymer architecture or conformation. (f) Förster resonance energy transfer (FRET): the transfer of energy from excitation of a donor group to an acceptor moiety in close proximity enables the measurement of through-space interactions, and this can probe properties such as conformation or end-to-end distance (Adapted from ref 234. Copyright 2023 American Chemical Society).

high-throughput workflows through autosamplers and well-plate formats. While these approaches have increased the speed and efficiency of data acquisition, it should be noted that user input is often required for common analysis techniques and can limit throughput. Chromatographic techniques such as size exclusion chromatography (SEC) and liquid chromatography (LC) can provide diverse information, including polymer conformation[20,98,198] and function.[198] Common size descriptors (M_n, M_w, R_g, and $Đ$) can be acquired through SEC, and coupling these techniques with light scattering detectors (SEC-MALS) or mass spectrometry (LC-MS) can expand the information obtained per sample run.[131,199,200]

Scattering methods also provide information-rich data and can be high throughput. Dynamic light scattering (DLS) can be performed in 384-well plate formats[201] to yield the size distribution of particles or polymers in solution or measure viscosity.[202,203] Additionally, techniques such as wide-angle X-ray scattering (WAXS) and small-angle X-ray or neutron scattering (SAXS, SANS) are used to characterize shape and structure on the length scale of a few angstroms up to hundreds of nanometers.[204−207] This instrumentation is more commonly found at national laboratories (e.g., Oak Ridge, Brookhaven, Lawrence Berkeley, and Argonne National Laboratories),[131,206] but availability may fluctuate for user submissions. Additionally, the quality of light scattering data is generally sensitive to sample preparation, and practical guidelines to strengthen SAXS data acquisition and processing are available.[208] Further, SAXS data collection can be coupled with rapid flow synthesis.[209]

Thermal analysis can also be performed by small scale, rapid analysis. Microdifferential scanning calorimetry (micro-DSC) yields the melting point, revealing conformational changes in proteins and soft matter, such as the denaturing of collagen.[210] While DSC is limited in throughput, differential scanning fluorimetry (DSF) can rapidly detect thermal transitions by tracking changes in solvatochromic fluorescent dyes with affinity for hydrophobic regions of a protein that become exposed upon unfolding.[211] DSF can also probe binding interactions between proteins and small molecules or polymer networks[5,212,213] with low sample burden (10 pmol) by measuring temperature-dependent fluorescence changes with simultaneous real time-PCR instrumentation.[214] However, DSF depends strongly on protein-reporter interactions and not every reporter will capture relevant thermal shifts.[215]

4.2.2. Adapting Biological Characterization Techniques for Synthetic Polymers. Polymer chemistry has also found success in adapting tools from biological workflows (Figure 9a). These tools are typically capable of characterizing conformation and binding profiles for biopolymers, including proteins, which are laborious to synthesize and can be unstable

https://doi.org/10.1021/acspolymersau.3c00025
ACS Polym. Au 2023, 3, 406−427

in ambient conditions. These techniques therefore produce information-rich outputs using small amounts of materials, and they can typically be performed in well plates.

The binding profile of a material can be quantified through multiple techniques. Surface plasmon resonance (SPR) determines binding stoichiometry and kinetics with high sensitivity by measuring changes in the refractive index of a thin film sensor surface with an immobilized ligand or analyte.[216] Continuous flow of the complementary partner through a single channel can screen upward of 10^2 samples per day and has been used to study polymer−polymer interactions (Figure 9b).[5,217] Affinity selection mass spectrometry (AS-MS) characterizes protein−ligand interactions by quantifying either the bound or unbound ligands with mass spectrometry, often coupled with reversed phase or size exclusion chromatography (Figure 9c).[218,219] This method has also been used to study adsorption of metals[105,220] and per- and polyfluorinated alkyl substances (PFAS)[221] to polymer resins, as well as binding of polymers to other larger biomolecules.[222] Additionally, this method is amenable to both immobilized or solution-phase applications by using physical filtration or dialysis to isolate the bound material, respectively. In small systematic studies ($\sim 10^1$), isothermal titration calorimetry (ITC) has be used to probe binding thermodynamics for proteins,[223,224] peptides,[225,226] and soft materials[227] by measuring solution heat changes as a target and ligand are slowly combined (Figure 9d). Although ITC is limited by longer experiment times, autosampling capabilities improve viability for characterization of hit materials identified in larger screens.[35,227] Finally, metallochromic dyes have been further used to visually screen the binding efficacy of materials, such as with immobilized peptoid ligands for hexavalent chromium and cadmium ions.[105,228]

Apart from binding a target material, several techniques exist to quantify the intrinsic conformation or size of a polymer, including ion mobility spectrometry (IMS), fluorescence techniques, and colorimetric dyes. IMS can be coupled with mass spectrometry to differentiate molecules by size based on drift time (Figure 9e). IMS has been used to study protein disorder,[229,230] sequence-defined polymers,[231] and polymer architecture.[232] Förster resonance energy transfer (FRET) instead quantifies through-space proximity between donor and acceptor fluorescent moieties to report on conformational changes in polymer brush networks[233] and in single chain polymers (Figure 9f).[234] Fluorescence anisotropy, also called fluorescence polarization (FP), measures the tumbling rate of a fluorophore through solution, which is correlated with apparent size.[5] This technique also requires a fluorescent probe, and has been used to determine the fluidity or rigidity of lipid and polymer membranes[235] and yield time-resolved conformation and dynamics analysis of polymers in solution.[236−238]

Dyes not coupled to the polymer can span solvatochromic, colorimetric, or fluorogenic types, and are capable of characterizing diverse properties including conformational changes,[98,214] aggregation,[239,240] cell viability,[241,242] metal ion uptake,[105,243−245] and catalytic activity.[45,46] Characterization using dyes is both accessible and straightforward, involving minimal sample preparation and rapid throughput in well plates. For example, Nile red is a fluorescent dye used to visualize hydrophobic surfaces of proteins,[246] polymers,[247] and lipid droplets.[248] Other probes, such as pyrene and Reichardt's dye, exhibit solvatochromic properties, with absorption or emission spectra dependent on the polarity of the solvent or local environment.[249] These probes are useful for surveying both single-chain conformation[98,250] and multichain assemblies.[251−253] Fluorescent dyes, including pentameric formylthiophene acetic acid and Thioflavin T, can monitor fibrilization and β-sheet assemblies,[254] and nucleic acid assemblies can be similarly visualized.[255,256] Dye assays also benefit from careful optimization and validation of assay conditions, including quantification of nonspecific interactions, background signal, and photobleaching, in addition to characterization with existing materials as internal calibrants.

4.3. Balancing Throughput and Data-Rich Characterization

A key consideration in study design is whether a high-throughput screen or a high-content measurement is most appropriate. For smaller systematic studies ($\sim 10^1$), implementation of multiple experimental techniques is feasible to capture structural and dynamic behavior. When using multiple characterization methods, a practical guideline is to select techniques for which the total estimated analysis time does not significantly exceed the time it takes to prepare the library or its subsequent expansion(s). Ultimately, the choice of characterization techniques should balance throughput and obtaining meaningful structure−property information with consideration of sampling and time constraints.

To examine the interplay between characterization throughput and information quality, we can compare how different techniques vary in utility based on library size. For example, an ITC can provide a detailed profile of binding interactions between a ligand and target (i.e., ΔG, ΔS, ΔH, K_d, and stoichiometry), but the lengthy sample run times, replicates, and controls required for analysis generally limit its application to small sample sets on the order of 10^1. In a large library of target-binding polymers, where it may not be practical to characterize every library member, alternative techniques such as UV−vis or fluorescence may be used to compare binding via displacement of a competing dye or fluorophore.[257−259] This is an example of proxy characterization by monitoring binding indirectly, via displacement of the dye, that enables higher-throughput screening. These alternatives allow for rapid comparisons of relative affinity across the sample set, ultimately aiding in the identification of standout materials for further characterization.

When extrinsic factors significantly influence the properties of macromolecules, multiple replicates or conditions may be required. Samples in a library may be assessed with a set baseline condition, but it may also be insightful to monitor behavior across different concentrations, pHs, or ionic strengths. For instance, in recent studies on the design of polymeric catalysts, the structure and activity profiles (i.e., turnover rates and final yields) are sensitive to external reaction conditions such as solvent, temperature, and substrate identity.[45,46,260]

Instrument parameters can be strategically tuned to improve the throughput of many characterization methods. In spectroscopic techniques, parameters such as the spectral width, resolution, and number of scans can be optimized to reduce acquisition times without compromising the ability to identify key features or make preliminary assessments. Similar enhancements can be achieved by changing factors such as the heat rate or temperature range in thermal analysis techniques (e.g., DSC and DSF), or the flow rates and column parameters in chromatographic separation techniques (e.g., HPLC, SEC,

https://doi.org/10.1021/acspolymersau.3c00025
ACS Polym. Au 2023, 3, 406−427

and LC-MS).[199,261,262] For techniques that are not amenable to well-plate formats, semiautomated workflows can be achieved by manually preparing samples for autosamplers.[126,137] Additionally, complementary or tandem approaches can also be explored. For example, diffuse reflectance infrared Fourier transform (DRIFT) spectroscopy is an alternative to NMR to rapidly determine polymer composition through automated analysis of dry samples,[127] and size-exclusion chromatography has been coupled with an orthogonal DLS for fast, high-information screening of aggregation behavior.[199]

4.4. Data Visualization, Modeling, and Interpretation

High-throughput screens frequently produce multidimensional data outputs, including multiple feature descriptors and associated measurements. While some approaches such as representation learning are capable of handling large amounts of data in the original feature space, even if it high-dimensional, direct visualization or interpretation of the data set typically poses a challenge (Figure 10a).

Dimensionality reduction is the compression of a large set of features while maximizing the amount of information preserved from the original set (Figure 10b). Because information loss is inherent to dimension reduction, which can be beneficial or detrimental, determining the success of the approach *a priori* is not possible. Principal component analysis (PCA) is a widely used unsupervised technique that generates

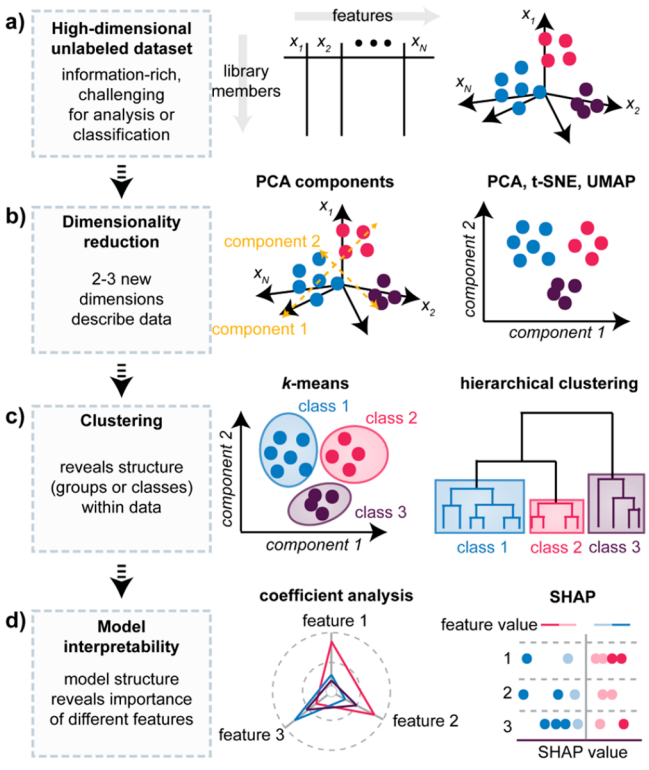


Figure 10. Data visualization and regression workflow. (a) A high-dimensional data set presents challenges for both analysis and visualization. (b) Dimension reduction techniques such as principal component analysis (PCA) can represent the data on new reduced axes. (c) Low-dimensional data can then be clustered into different classes using strategies such as k-means clustering or hierarchical clustering. (d) Models can be fit and interpreted in a variety of ways, including regression analysis and calculating SHAP values, to determine the importance of various features.

new features that are linear combinations of the original features, ranked in order of the variance captured.[263] Typically, the top one to three principal components are selected to use in further analysis or visualization. PCA can also represent experimental outputs such as mass spectra,[264,265] FT-IR spectra,[266] fluorescence data,[120,267] or molecular dynamics trajectories.[268] PCA has found applications in the development of gene delivery agents,[269] membranes in separations,[270] and protein stabilizers.[47] Other types of dimensionality reduction tools include t-distributed stochastic neighbor embedding (t-SNE), designed to visualize high-dimensional data using a mapping onto two- or three-dimensions,[271] and uniform manifold approximation and projection (UMAP), a topology driven technique.[272] For example, the TransPolymer model uses t-SNE to visualize millions of unlabeled training data points and data from specific property databases.[163] Many other types of dimension reduction techniques are detailed thoroughly by Banerjee and co-workers.[273]

Clustering is another unsupervised technique that can identify groupings within unlabeled data sets referred to as classes or clusters, and these algorithms are useful when groupings are not intuitive or measurable (Figure 10c). While dimensionality reduction decreases the size of the feature space, clustering algorithms calculate distances between data points to place them into clusters. However, distance metrics tend to be less effective in a high-dimensional space, so dimensionality reduction prior to clustering can be useful. This assumes that the reduced dimensions adequately represent the original data set, and in the case of very large data sets, distance metrics developed to be robust to high dimensions can be used for clustering directly.[274] One common clustering method is *k*-means, which divides data points into nonoverlapping clusters by minimizing the distance between each point to its assigned cluster center.[275,276] This technique does not describe how pairs of data points are related. Alternatively, hierarchical clustering iteratively calculates distances between data to create clusters between the most similar points, giving not only cluster assignments but a branched representation of similarity between data points via dendrogram plots.[275,276] A challenge in cluster analysis is that the true number of clusters contained within a data set is often unknown, so different scores are used to assess cluster validity.[277] For example, a molecular photocatalyst library has been visualized through a UMAP representation and then clustered using *k*-means to reveal distinct classes of chemically similar catalysts further compared for activity.[278] These techniques are not limited to very large data sets: molecular descriptors for 39 thermoplastic stabilizers were also successfully reduced with PCA and classified with *k*-means clustering.[279]

Both clustering and dimensionality reduction serve purposes beyond visualization; they are applicable in conjunction with machine learning models such as neural nets. Fitting a model to a lower-dimensional space reduces the computational cost of model training and may improve the model performance, but the technique and model selection depend on the data set. For example, a combination of the discussed techniques was applied in a data-driven study of polymeric drug delivery injectables[3] and to predict accurate annotation of protein sequences.[280]

Methods to interpret and glean chemical insight from optimized models are available. Model interpretability can be quantified using the "predictive, descriptive, and relevant" framework discussed by Yu and co-workers, covering model

https://doi.org/10.1021/acspolymersau.3c00025
ACS Polym. Au 2023, 3, 406−427

development stages such as model selection, training, testing, and analysis.[281] Current strategies primarily focus on *post hoc* interpretability.[282] For instance, linear regression models are easily interpretable, as regression coefficients are correlated to the impact of different features on the output. This was demonstrated in a study of Pd-driven catalysis.[283] Decision trees in random forest or gradient boosting algorithms can reveal which features play inhibitory or cooperative roles in target outcomes, such as deconvoluting ligand-target interactions for peptides capable of inactivating biological targets.[284]

For more complex models like neural nets, different analysis methods are necessary. Shapley additive explanations (SHAP) values, which quantify the importance of different features, can be used to rank the relative importance of features within predictions.[285] These features include different polymer structure descriptors in diverse problems, including protein stabilization,[47] gas separation membranes,[286] and electronics.[287,288] Salience or attention maps can also extract specific information on how neural nets interact differently with various features from particular inputs.[282] Such interpretation of neural nets requires caution. Sometimes these outputs corroborate chemical or physical intuition, like reaction classification based on functional groups,[289] but they may also lead to unexplainable decisions, referred to as shortcut learning.[290] For example, the K_{DEEP} algorithm, used to predict binding affinity of protein−ligand complexes, sometimes missed important interactions (e.g., key functional moieties) or assigned importance to erroneous interactions, which was analyzed using PlayMolecule Glimpse.[291] As the versatility of deep learning grows, we must continue to carefully interrogate its representation learning and be mindful of its interpretability.

5. CONCLUSIONS AND OUTLOOK

As the prevalence and accessibility of high-throughput methods continues to expand in polymer chemistry, strategies to efficiently design and perform these experiments and extract information will be critical. Herein, we have offered different considerations for the design, synthesis, and characterization of polymer libraries to enhance structure−property understanding and accelerate material design. A holistic approach to high-throughput screening that incorporates both experimental and statistical tools, many of which can be adapted from strategies implemented with biomacromolecules, will expedite the *de novo* design of synthetic materials. While automated liquid handlers have improved parallel syntheses, there are many ways to systematically screen a macromolecular space. From strategic initial sampling to synthetic techniques, we have provided guidance for accessible methods to study a large chemical space.

Improving the quality of data sets generated by high-throughput techniques is critical to successful predictive modeling and statistics. The rapid development of high-performing predictive tools in structural biology, such as AlphaFold, is facilitated by the extensive and standardized Protein Data Bank data set. While initial successes in predictive models for polymer science have been achieved, data sets are specific to each study and scattered across research groups. Consolidating this data into a larger, open-access database will facilitate the emergence of more powerful predictive software. Key to this advance is improving standardization of measurement. The FAIR guiding principle—findable, accessible, interoperable, reusable—are excellent considerations for data management. Considering the sensitivity of polymer characterization to experimental conditions, incorporating metadata such as sample preparation and protocols into data sets will be useful in data comparisons. Further, calibration data points, such as material standards, are also useful benchmarks and points of comparison. These open science approaches, where high-quality data is freely shared, will accelerate and improve the success of material development.[292,293]

The increasing abundance of machine learning algorithms indicates a growing interest in using statistical tools in chemical workflows. The design, assessment, and reporting of these tools is essential to facilitate further improvement and accessibility. Open-access tools should be comprehensible to researchers from a variety of backgrounds and experience levels with statistical tools and coding languages. Regular updates and annotations on open-access platforms like GitHub, along with detailed instructions on download, installation, and usage will enhance their accessibility and reach to allow more researchers to analyze greater amounts of data. Providing an accompanying set of sample data in the desired format is recommended to reduce the learning curve and provide a template for new data inputs.

In summary, the rapid growth of high-throughput characterization in polymer science has enabled new data-driven approaches. We have highlighted advances in the development, synthesis, and screening of polymer libraries and delineated practical strategies for harnessing statistical tools in data representation and interpretation. We are optimiztic about the future of high-throughput platforms in *de novo* design of materials, pushing the boundaries of foundational science and addressing global challenges.

AUTHOR INFORMATION

Corresponding Author

Abigail S. Knight − *Department of Chemistry, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, United States;* orcid.org/0000-0003-1524-473X; Email: aknight@unc.edu

Authors

Erin C. Day − *Department of Chemistry, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, United States;* orcid.org/0000-0001-9278-9560

Supraja S. Chittari − *Department of Chemistry, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, United States;* orcid.org/0000-0002-5010-5201

Matthew P. Bogen − *Department of Chemistry, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acspolymersau.3c00025

Author Contributions

†E.C.D. and S.S.C. contributed equally to this work. CRediT: **Erin C. Day** conceptualization, visualization, writing-original draft, writing-review & editing; **Supraja S Chittari** conceptualization, visualization, writing-original draft, writing-review & editing; **Matthew P Bogen** conceptualization, visualization, writing-original draft, writing-review & editing; **Abigail S**

https://doi.org/10.1021/acspolymersau.3c00025
ACS Polym. Au 2023, 3, 406−427

Knight conceptualization, funding acquisition, supervision, visualization, writing-review & editing.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported in part by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Award Number DE-SC0021295 (focus on synthetic copolymers) and in part by the National Science Foundation under Grant DMR-2045021 (focus on bioinspired techniques). E.C.D. acknowledges support from the National Defense Science & Engineering Graduate (NDSEG) Fellowship Program. S.S.C. acknowledges the National Science Foundation Graduate Research Fellowship (DGE-2040435). We thank Dr. Rishi Kulkarni and Johann Rapp for valuable feedback and discussions.

■ ABBREVIATIONS

AS-MS, affinity selection mass spectrometry; $Đ$, dispersity; DFT, density-functional theory; DLS, dynamic light scattering; DNA, deoxyribonucleic acid; DOSY, diffusion-ordered spectroscopy; DRIFT, diffuse reflectance infrared Fourier transform spectroscopy; DSC, differential scanning calorimetry; DSF, differential scanning fluorimetry; ECFP, extended-connectivity fingerprints; FRET, Förster resonance energy transfer; FT-IR, Fourier transform infrared spectroscopy; ΔG, Gibbs free energy change; ΔH, change in enthalpy; HPLC, high-performance liquid chromatography; IMS, ion mobility spectrometry; ITC, isothermal titration calorimetry; K_a, association constant; K_d, dissociation constant; LC-MS, liquid chromatography−mass spectrometry; MALS, multiangle light scattering; MD, molecular dynamics; M_n, number-average molecular weight; M_w, weight-average molecular weight; N, binding stoichiometry; NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser effect spectroscopy; OBOC, one-bead one-compound; PCA, principle component analysis; PCR, polymerase chain reaction; PET-RAFT, photoinduced electron/energy transfer reversible addition−fragmentation chain transfer polymerization; PFAS, per/polyfluorinated alkyl substances; PPM, post-polymerization modification; QSAR, quantitative structure−activity relationship; QSPR, quantitative structure−property relationship; R_g, radius of gyration; RMSE, root-mean-square error; ROESY, rotating frame Overhauser enhancement spectroscopy; ΔS, change in entropy; SAXS, small-angle X-ray scattering; SEC, size exclusion chromatography; SHAP, Shapley additive explanations; SMILES, simplified molecular-input-line-entry system; SPR, surface plasmon resonance; t-SNE, t-distributed stochastic neighbor embedding; TLC, thin layer chromatography; UMAP, uniform manifold approximation and projection; UV−vis, ultraviolet−visible (light); WAXS, wide-angle X-ray scattering

■ REFERENCES

(1) Namazi, H. Polymers in Our Daily Life. *Bioimpacts* **2017**, *7* (2), 73−74.

(2) Schaefer, S.; Pham, T. T. P.; Brunke, S.; Hube, B.; Jung, K.; Lenardon, M. D.; Boyer, C. Rational Design of an Antifungal Polyacrylamide Library with Reduced Host-Cell Toxicity. *ACS Appl. Mater. Interfaces* **2021**, *13* (23), 27430−27444.

(3) Bannigan, P.; Bao, Z.; Hickman, R. J.; Aldeghi, M.; Häse, F.; Aspuru-Guzik, A.; Allen, C. Machine Learning Models to Accelerate the Design of Polymeric Long-Acting Injectables. *Nat. Commun.* **2023**, *14* (1), 35.

(4) Pereira, D. A.; Williams, J. A. Origin and Evolution of High Throughput Screening: Origin and Circumscribed History of HTS. *Br. J. Pharmacol.* **2007**, *152* (1), 53−61.

(5) Blay, V.; Tolani, B.; Ho, S. P.; Arkin, M. R. High-Throughput Screening: Today's Biochemical and Cell-Based Approaches. *Drug Discovery Today* **2020**, *25* (10), 1807−1821.

(6) Van Gaal, R. C.; Vrehen, A. F.; Van Sprang, J. F.; Fransen, P.-P. K. H.; Van Turnhout, M. C.; Dankers, P. Y. W. Biomaterial Screening of Protein Coatings and Peptide Additives: Towards a Simple Synthetic Mimic of a Complex Natural Coating for a Bio-Artificial Kidney. *Biomater. Sci.* **2021**, *9* (6), 2209−2220.

(7) Xu, J.; Jung, K.; Atme, A.; Shanmugam, S.; Boyer, C. A Robust and Versatile Photoinduced Living Polymerization of Conjugated and Unconjugated Monomers and Its Oxygen Tolerance. *J. Am. Chem. Soc.* **2014**, *136* (14), 5508−5519.

(8) Zugates, G. T.; Anderson, D. G.; Langer, R. High-Throughput Methods for Screening Polymeric Transfection Reagents. *Cold Spring Harb Protoc* **2013**, *2013* (11), pdb.prot078634.

(9) Dahlhauser, S. D.; Escamilla, P. R.; VandeWalle, A. N.; York, J. T.; Rapagnani, R. M.; Shei, J. S.; Glass, S. A.; Coronado, J. N.; Moor, S. R.; Saunders, D. P.; Anslyn, E. V. Sequencing of Sequence-Defined Oligourethanes via Controlled Self-Immolation. *J. Am. Chem. Soc.* **2020**, *142* (6), 2744−2749.

(10) Michael, S.; Auld, D.; Klumpp, C.; Jadhav, A.; Zheng, W.; Thorne, N.; Austin, C. P.; Inglese, J.; Simeonov, A. A Robotic Platform for Quantitative High-Throughput Screening. *ASSAY and Drug Development Technologies* **2008**, *6* (5), 637−657.

(11) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D. Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825−2830.

(12) Oliver, S.; Zhao, L.; Gormley, A. J.; Chapman, R.; Boyer, C. Living in the Fast Lane—High Throughput Controlled/Living Radical Polymerization. *Macromolecules* **2019**, *52* (1), 3−23.

(13) Baudis, S.; Behl, M. High-Throughput and Combinatorial Approaches for the Development of Multifunctional Polymers. *Macromol. Rapid Commun.* **2022**, *43* (12), 2100400.

(14) Martin, T. B.; Audus, D. J. Emerging Trends in Machine Learning: A Polymer Perspective. *ACS Polym. Au* **2023**, *3*, 239.

(15) Patel, R. A.; Webb, M. A. Data-Driven Design of Polymer-Based Biomaterials: High-Throughput Simulation, Experimentation, and Machine Learning. *ACS Appl. Bio Mater.* **2023**, DOI: 10.1021/acsabm.2c00962.

(16) Meyer, T. A.; Ramirez, C.; Tamasi, M. J.; Gormley, A. J. A User's Guide to Machine Learning for Polymeric Biomaterials. *ACS Polym. Au* **2023**, *3*, 141.

(17) Patra, T. K. Data-Driven Methods for Accelerating Polymer Design. *ACS Polym. Au* **2022**, *2* (1), 8−26.

(18) Le, T.; Epa, V. C.; Burden, F. R.; Winkler, D. A. Quantitative Structure-Property Relationship Modeling of Diverse Materials Properties. *Chem. Rev.* **2012**, *112* (5), 2889−2919.

(19) Xiong, T. M.; Garcia, E. S.; Chen, J.; Zhu, L.; Alzona, A. J.; Zimmerman, S. C. Enzyme-like Catalysis by Single Chain Nanoparticles That Use Transition Metal Cofactors. *Chem. Commun.* **2022**, *58* (7), 985−988.

(20) Warren, J. L.; Dykeman-Bermingham, P. A.; Knight, A. S. Controlling Amphiphilic Polymer Folding beyond the Primary Structure with Protein-Mimetic Di(Phenylalanine). *J. Am. Chem. Soc.* **2021**, *143* (33), 13228−13234.

(21) Gao, H.; Matyjaszewski, K. Synthesis of Star Polymers by A New "Core-First" Method: Sequential Polymerization of Cross-Linker and Monomer. *Macromolecules* **2008**, *41* (4), 1118−1125.

(22) Mavila, S.; Eivgi, O.; Berkovich, I.; Lemcoff, N. G. Intramolecular Cross-Linking Methodologies for the Synthesis of Polymer Nanoparticles. *Chem. Rev.* **2016**, *116* (3), 878−961.

(23) Voit, B. I.; Lederer, A. Hyperbranched and Highly Branched Polymer Architectures—Synthetic Strategies and Major Characterization Aspects. *Chem. Rev.* **2009**, *109* (11), 5924−5973.
(24) Meier, M. A. R.; Schubert, U. S. Combinatorial Polymer Research and High-Throughput Experimentation: Powerful Tools for the Discovery and Evaluation of New Materials. *J. Mater. Chem.* **2004**, *14* (22), 3289−3299.
(25) Weissman, S. A.; Anderson, N. G. Design of Experiments (DoE) and Process Optimization. A Review of Recent Publications. *Org. Process Res. Dev.* **2015**, *19* (11), 1605−1633.
(26) Archer, W. R.; Fiorito, A.; Heinz-Kunert, S. L.; MacNicol, P. L.; Winn, S. A.; Schulz, M. D. Synthesis and Rare-Earth-Element Chelation Properties of Linear Poly(Ethylenimine Methylenephosphonate). *Macromolecules* **2020**, *53* (6), 2061−2068.
(27) Arnaud, J.; Audfray, A.; Imberty, A. Binding Sugars: From Natural Lectins to Synthetic Receptors and Engineered Neolectins. *Chem. Soc. Rev.* **2013**, *42* (11), 4798−4813.
(28) Rothfuss, H.; Knöfel, N. D.; Roesky, P. W.; Barner-Kowollik, C. Single-Chain Nanoparticles as Catalytic Nanoreactors. *J. Am. Chem. Soc.* **2018**, *140* (18), 5875−5881.
(29) Hilburg, S. L.; Ruan, Z.; Xu, T.; Alexander-Katz, A. Behavior of Protein-Inspired Synthetic Random Heteropolymers. *Macromolecules* **2020**, *53* (21), 9187−9199.
(30) van Leeuwen, T.; Heideman, G. H.; Zhao, D.; Wezenberg, S. J.; Feringa, B. L. In Situ Control of Polymer Helicity with a Non-Covalently Bound Photoresponsive Molecular Motor Dopant. *Chem. Commun.* **2017**, *53* (48), 6393−6396.
(31) Wawer, M.; Peltason, L.; Weskamp, N.; Teckentrup, A.; Bajorath, J. Structure-Activity Relationship Anatomy by Network-like Similarity Graphs and Local Structure-Activity Relationship Indices. *J. Med. Chem.* **2008**, *51* (19), 6075−6084.
(32) Maynard, A. T.; Roberts, C. D. Quantifying, Visualizing, and Monitoring Lead Optimization. *J. Med. Chem.* **2016**, *59* (9), 4189−4201.
(33) Gormley, A. J.; Webb, M. A. Machine Learning in Combinatorial Polymer Chemistry. *Nat. Rev. Mater.* **2021**, *6* (8), 642−644.
(34) Tuckerman, M. E. The Curse of Dimensionality Loses Its Power. *Nat. Comput. Sci.* **2022**, *2* (1), 6−7.
(35) Archer, W. R.; Gallagher, C. M. B.; Vaissier Welborn, V.; Schulz, M. D. Exploring the Role of Polymer Hydrophobicity in Polymer-Metal Binding Thermodynamics. *Phys. Chem. Chem. Phys.* **2022**, *24* (6), 3579−3585.
(36) Karim, K.; Breton, F.; Rouillon, R.; Piletska, E. V.; Guerreiro, A.; Chianella, I.; Piletsky, S. A. How to Find Effective Functional Monomers for Effective Molecularly Imprinted Polymers? *Adv. Drug Delivery Rev.* **2005**, *57* (12), 1795−1808.
(37) Blackman, L. D.; Gunatillake, P. A.; Cass, P.; Locock, K. E. S. An Introduction to Zwitterionic Polymer Behavior and Applications in Solution and at Surfaces. *Chem. Soc. Rev.* **2019**, *48* (3), 757−770.
(38) Roy, D.; Brooks, W. L. A.; Sumerlin, B. S. New Directions in Thermoresponsive Polymers. *Chem. Soc. Rev.* **2013**, *42*, 7214.
(39) Ren, J. M.; McKenzie, T. G.; Fu, Q.; Wong, E. H. H.; Xu, J.; An, Z.; Shanmugam, S.; Davis, T. P.; Boyer, C.; Qiao, G. G. Star Polymers. *Chem. Rev.* **2016**, *116* (12), 6743−6836.
(40) Das, R. K.; Ruff, K. M.; Pappu, R. V. Relating Sequence Encoded Information to Form and Function of Intrinsically Disordered Proteins. *Curr. Opin. Struct. Biol.* **2015**, *32*, 102−112.
(41) Korevaar, P. A.; Newcomb, C. J.; Meijer, E. W.; Stupp, S. I. Pathway Selection in Peptide Amphiphile Assembly. *J. Am. Chem. Soc.* **2014**, *136* (24), 8540−8543.
(42) Ghosh, A.; Buettner, C. J.; Manos, A. A.; Wallace, A. J.; Tweedle, M. F.; Goldberger, J. E. Probing Peptide Amphiphile Self-Assembly in Blood Serum. *Biomacromolecules* **2014**, *15* (12), 4488−4494.
(43) Wehner, M.; Röhr, M. I. S.; Bühler, M.; Stepanenko, V.; Wagner, W.; Würthner, F. Supramolecular Polymorphism in One-Dimensional Self-Assembly by Kinetic Pathway Control. *J. Am. Chem. Soc.* **2019**, *141* (14), 6092−6107.
(44) Ghosh, G.; Barman, R.; Mukherjee, A.; Ghosh, U.; Ghosh, S.; Fernández, G. Control over Multiple Nano- and Secondary Structures in Peptide Self-Assembly. *Angew. Chem., Int. Ed.* **2022**, *61* (5), No. e202113403.
(45) Chen, J.; Wang, J.; Bai, Y.; Li, K.; Garcia, E. S.; Ferguson, A. L.; Zimmerman, S. C. Enzyme-like Click Catalysis by a Copper-Containing Single-Chain Nanoparticle. *J. Am. Chem. Soc.* **2018**, *140* (42), 13695−13702.
(46) Sanders, M. A.; Chittari, S. S.; Sherman, N.; Foley, J. R.; Knight, A. S. Versatile Triphenylphosphine-Containing Polymeric Catalysts and Elucidation of Structure-Function Relationships. *J. Am. Chem. Soc.* **2023**, *145* (17), 9686−9692.
(47) Tamasi, M. J.; Patel, R. A.; Borca, C. H.; Kosuri, S.; Mugnier, H.; Upadhya, R.; Murthy, N. S.; Webb, M. A.; Gormley, A. J. Machine Learning on a Robotic Platform for the Design of Polymer-Protein Hybrids. *Adv. Mater.* **2022**, *34*, 2201809.
(48) Hoshino, Y.; Kodama, T.; Okahata, Y.; Shea, K. J. Peptide Imprinted Polymer Nanoparticles: A Plastic Antibody. *J. Am. Chem. Soc.* **2008**, *130* (46), 15242−15243.
(49) Green, R. M.; Bicker, K. L. Discovery and Characterization of a Rapidly Fungicidal and Minimally Toxic Peptoid against *Cryptococcus Neoformans*. *ACS Med. Chem. Lett.* **2021**, *12* (9), 1470−1477.
(50) O'Hagan, S.; Kell, D. B. Analysing and Navigating Natural Products Space for Generating Small, Diverse, But Representative Chemical Libraries. *Biotechnology Journal* **2018**, *13* (1), 1700503.
(51) Reis, M.; Gusev, F.; Taylor, N. G.; Chung, S. H.; Verber, M. D.; Lee, Y. Z.; Isayev, O.; Leibfarth, F. A. Machine-Learning-Guided Discovery of ^{19}F MRI Agents Enabled by Automated Copolymer Synthesis. *J. Am. Chem. Soc.* **2021**, *143* (42), 17677−17689.
(52) Vabalas, A.; Gowen, E.; Poliakoff, E.; Casson, A. J. Machine Learning Algorithm Validation with a Limited Sample Size. *PLoS One* **2019**, *14* (11), No. e0224365.
(53) Viana, F. A. C. A Tutorial on Latin Hypercube Design of Experiments. *Quality and Reliability Engineering International* **2016**, *32* (5), 1975−1985.
(54) Renardy, M.; Joslyn, L. R.; Millar, J. A.; Kirschner, D. E. To Sobol or Not to Sobol? The Effects of Sampling Schemes in Systems Biology Applications. *Mathematical Biosciences* **2021**, *337*, 108593.
(55) Banerjee, P.; Dehnbostel, F. O.; Preissner, R. Prediction Is a Balancing Act: Importance of Sampling Methods to Balance Sensitivity and Specificity of Predictive Models Based on Imbalanced Chemical Data Sets. *Frontiers in Chemistry* **2018**, *6*, 362.
(56) Pedersen, J. S. Analysis of Small-Angle Scattering Data from Colloids and Polymer Solutions: Modeling and Least-Squares Fitting. *Adv. Colloid Interface Sci.* **1997**, *70*, 171−210.
(57) Masaro, L.; Zhu, X. X. Physical Models of Diffusion for Polymer Solutions, Gels and Solids. *Prog. Polym. Sci.* **1999**, *24* (5), 731−775.
(58) Jiang, D.; Wu, Z.; Hsieh, C.-Y.; Chen, G.; Liao, B.; Wang, Z.; Shen, C.; Cao, D.; Wu, J.; Hou, T. Could Graph Neural Networks Learn Better Molecular Representation for Drug Discovery? A Comparison Study of Descriptor-Based and Graph-Based Models. *Journal of Cheminformatics* **2021**, *13* (1), 12.
(59) Borg, C. K. H.; Muckley, E. S.; Nyby, C.; Saal, J. E.; Ward, L.; Mehta, A.; Meredig, B. Quantifying the Performance of Machine Learning Models in Materials Discovery. *Digital Discovery* **2023**, *2*, 327.
(60) Gramatica, P.; Chirico, N.; Papa, E.; Cassani, S.; Kovarich, S. QSARINS: A New Software for the Development, Analysis, and Validation of QSAR MLR Models. *J. Comput. Chem.* **2013**, *34* (24), 2121−2132.
(61) Kuenneth, C.; Ramprasad, R. polyBERT: A Chemical Language Model to Enable Fully Machine-Driven Ultrafast Polymer Informatics. *Nat. Commun.* **2023**, *14* (1), 4099.
(62) Goh, G. B.; Hodas, N. O.; Vishnu, A. Deep Learning for Computational Chemistry. *J. Comput. Chem.* **2017**, *38* (16), 1291−1307.

https://doi.org/10.1021/acspolymersau.3c00025
ACS Polym. Au 2023, 3, 406−427

(63) Gao, H.; Struble, T. J.; Coley, C. W.; Wang, Y.; Green, W. H.; Jensen, K. F. Using Machine Learning To Predict Suitable Conditions for Organic Reactions. *ACS Cent. Sci.* **2018**, *4* (11), 1465−1476.
(64) Xu, P.; Ji, X.; Li, M.; Lu, W. Small Data Machine Learning in Materials Science. *npj Comput. Mater.* **2023**, *9* (1), 1−15.
(65) Zhang, Y.; Ling, C. A Strategy to Apply Machine Learning to Small Datasets in Materials Science. *npj Comput. Mater.* **2018**, *4* (1), 1−8.
(66) Lookman, T.; Balachandran, P. V.; Xue, D.; Yuan, R. Active Learning in Materials Science with Emphasis on Adaptive Sampling Using Uncertainties for Targeted Design. *npj Comput. Mater.* **2019**, *5* (1), 1−17.
(67) Deringer, V. L.; Bartók, A. P.; Bernstein, N.; Wilkins, D. M.; Ceriotti, M.; Csányi, G. Gaussian Process Regression for Materials and Molecules. *Chem. Rev.* **2021**, *121* (16), 10073−10141.
(68) Kusne, A. G.; Yu, H.; Wu, C.; Zhang, H.; Hattrick-Simpers, J.; DeCost, B.; Sarker, S.; Oses, C.; Toher, C.; Curtarolo, S.; Davydov, A. V.; Agarwal, R.; Bendersky, L. A.; Li, M.; Mehta, A.; Takeuchi, I. On-the-Fly Closed-Loop Materials Discovery via Bayesian Active Learning. *Nat. Commun.* **2020**, *11* (1), 5966.
(69) Khatamsaz, D.; Vela, B.; Singh, P.; Johnson, D. D.; Allaire, D.; Arróyave, R. Bayesian Optimization with Active Learning of Design Constraints Using an Entropy-Based Approach. *npj Comput. Mater.* **2023**, *9* (1), 1−14.
(70) Tao, H.; Wu, T.; Aldeghi, M.; Wu, T. C.; Aspuru-Guzik, A.; Kumacheva, E. Nanoparticle Synthesis Assisted by Machine Learning. *Nat. Rev. Mater.* **2021**, *6* (8), 701−716.
(71) Mehta, S.; Laghuvarapu, S.; Pathak, Y.; Sethi, A.; Alvala, M.; Priyakumar, U. D. MEMES: Machine Learning Framework for Enhanced MolEcular Screening. *Chemical Science* **2021**, *12* (35), 11710−11721.
(72) Avila, C.; Cassani, C.; Kogej, T.; Mazuela, J.; Sarda, S.; Clayton, A. D.; Kossenjans, M.; Green, C. P.; Bourne, R. A. Automated Stopped-Flow Library Synthesis for Rapid Optimisation and Machine Learning Directed Experimentation. *Chemical Science* **2022**, *13* (41), 12087−12099.
(73) Wang, X.; Huang, Y.; Xie, X.; Liu, Y.; Huo, Z.; Lin, M.; Xin, H.; Tong, R. Bayesian-Optimization-Assisted Discovery of Stereoselective Aluminum Complexes for Ring-Opening Polymerization of Racemic Lactide. *Nat. Commun.* **2023**, *14* (1), 3647.
(74) Doan Tran, H.; Kim, C.; Chen, L.; Chandrasekaran, A.; Batra, R.; Venkatram, S.; Kamal, D.; Lightstone, J. P.; Gurnani, R.; Shetty, P.; Ramprasad, M.; Laws, J.; Shelton, M.; Ramprasad, R. Machine-Learning Predictions of Polymer Properties with Polymer Genome. *J. Appl. Phys.* **2020**, *128* (17), 171104.
(75) Kosuri, S.; Borca, C. H.; Mugnier, H.; Tamasi, M.; Patel, R. A.; Perez, I.; Kumar, S.; Finkel, Z.; Schloss, R.; Cai, L.; Yarmush, M. L.; Webb, M. A.; Gormley, A. J. Machine-Assisted Discovery of Chondroitinase ABC Complexes toward Sustained Neural Regeneration. *Adv. Healthcare Mater.* **2022**, *11* (10), 2102101.
(76) Sattari, K.; Xie, Y.; Lin, J. Data-Driven Algorithms for Inverse Design of Polymers. *Soft Matter* **2021**, *17* (33), 7607−7622.
(77) Pyzer-Knapp, E. O.; Suh, C.; Gómez-Bombarelli, R.; Aguilera-Iparraguirre, J.; Aspuru-Guzik, A. What Is High-Throughput Virtual Screening? A Perspective from Organic Materials Discovery. *Annu. Rev. Mater. Res.* **2015**, *45* (1), 195−216.
(78) Himanen, L.; Geurts, A.; Foster, A. S.; Rinke, P. Data-Driven Materials Science: Status, Challenges, and Perspectives. *Advanced Science* **2019**, *6* (21), 1900808.
(79) Kanal, I. Y.; Owens, S. G.; Bechtel, J. S.; Hutchison, G. R. Efficient Computational Screening of Organic Polymer Photovoltaics. *J. Phys. Chem. Lett.* **2013**, *4* (10), 1613−1623.
(80) Mannodi-Kanakkithodi, A.; Pilania, G.; Huan, T. D.; Lookman, T.; Ramprasad, R. Machine Learning Strategy for Accelerated Design of Polymer Dielectrics. *Sci. Rep* **2016**, *6* (1), 20952.
(81) Kim, C.; Batra, R.; Chen, L.; Tran, H.; Ramprasad, R. Polymer Design Using Genetic Algorithm and Machine Learning. *Comput. Mater. Sci.* **2021**, *186*, 110067.
(82) Ramesh, P. S; Patra, T. K Polymer Sequence Design via Molecular Simulation-Based Active Learning. *Soft Matter* **2023**, *19* (2), 282−294.
(83) Webb, M. A.; Jackson, N. E.; Gil, P. S.; de Pablo, J. J. Targeted Sequence Design within the Coarse-Grained Polymer Genome. *Science Advances* **2020**, *6* (43), No. eabc6216.
(84) Park, N. H. Artificial Intelligence Driven Design of Catalysts and Materials for Ring Opening Polymerization Using a Domain-Specific Language. *Nat. Commun.* **2023**, 3686.
(85) Jablonka, K. M.; Jothiappan, G. M.; Wang, S.; Smit, B.; Yoo, B. Bias Free Multiobjective Active Learning for Materials Design and Discovery. *Nat. Commun.* **2021**, *12* (1), 2312.
(86) Sha, W.; Li, Y.; Tang, S.; Tian, J.; Zhao, Y.; Guo, Y.; Zhang, W.; Zhang, X.; Lu, S.; Cao, Y.-C.; Cheng, S. Machine Learning in Polymer Informatics. *InfoMat* **2021**, *3* (4), 353−361.
(87) Xu, P.; Chen, H.; Li, M.; Lu, W. New Opportunity: Machine Learning for Polymer Materials Design and Discovery. *Advanced Theory and Simulations* **2022**, *5* (5), 2100565.
(88) Otsuka, S.; Kuwajima, I.; Hosoya, J.; Xu, Y.; Yamazaki, M. PoLyInfo: Polymer Database for Polymeric Materials Design. In *2011 International Conference on Emerging Intelligent Data and Web Technologies*; 2011; pp 22−29.
(89) Kim, C.; Chandrasekaran, A.; Huan, T. D.; Das, D.; Ramprasad, R. Polymer Genome: A Data-Powered Polymer Informatics Platform for Property Predictions. *J. Phys. Chem. C* **2018**, *122* (31), 17575−17585.
(90) Walsh, D. J.; Zou, W.; Schneider, L.; Mello, R.; Deagen, M. E.; Mysona, J.; Lin, T.-S.; de Pablo, J. J.; Jensen, K. F.; Audus, D. J.; Olsen, B. D. Community Resource for Innovation in Polymer Technology (CRIPT): A Scalable Polymer Material Data Structure. *ACS Cent. Sci.* **2023**, *9* (3), 330−338.
(91) *Polymer Handbook*, 4th ed.; Brandrup, J., Ed.; Wiley: New York Weinheim, 1999.
(92) Ellis, B.; Smith, R. *Polymers: A Property Database*, Second ed.; CRC Press, 2008.
(93) Ma, R.; Luo, T. PI1M: A Benchmark Database for Polymer Informatics. *J. Chem. Inf. Model.* **2020**, *60* (10), 4684−4690.
(94) Tao, L.; Varshney, V.; Li, Y. Benchmarking Machine Learning Models for Polymer Informatics: An Example of Glass Transition Temperature. *J. Chem. Inf. Model.* **2021**, *61* (11), 5395−5413.
(95) Wilkinson, M. D.; Dumontier, M.; Aalbersberg, Ij. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L. B.; Bourne, P. E.; Bouwman, J.; Brookes, A. J.; Clark, T.; Crosas, M.; Dillo, I.; Dumon, O.; Edmunds, S.; Evelo, C. T.; Finkers, R.; Gonzalez-Beltran, A.; Gray, A. J. G.; Groth, P.; Goble, C.; Grethe, J. S.; Heringa, J.; 't Hoen, P. A. C.; Hooft, R.; Kuhn, T.; Kok, R.; Kok, J.; Lusher, S. J.; Martone, M. E.; Mons, A.; Packer, A. L.; Persson, B.; Rocca-Serra, P.; Roos, M.; van Schaik, R.; Sansone, S.-A.; Schultes, E.; Sengstag, T.; Slater, T.; Strawn, G.; Swertz, M. A.; Thompson, M.; van der Lei, J.; van Mulligen, E.; Velterop, J.; Waagmeester, A.; Wittenburg, P.; Wolstencroft, K.; Zhao, J.; Mons, B. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci. Data* **2016**, *3* (1), 160018.
(96) Bobers, J.; Hahn, L. K.; Averbeck, T.; Brunschweiger, A.; Kockmann, N. Reaction Optimization of a Suzuki-Miyaura Cross-Coupling Using Design of Experiments. *Chemie Ingenieur Technik* **2022**, *94* (5), 780−785.
(97) Bowden, G. D.; Pichler, B. J.; Maurer, A. A Design of Experiments (DoE) Approach Accelerates the Optimization of Copper-Mediated 18F-Fluorination Reactions of Arylstannanes. *Sci. Rep* **2019**, *9* (1), 11370.
(98) Terashima, T.; Sugita, T.; Fukae, K.; Sawamoto, M. Synthesis and Single-Chain Folding of Amphiphilic Random Copolymers in Water. *Macromolecules* **2014**, *47* (2), 589−600.
(99) Matsumoto, K.; Terashima, T.; Sugita, T.; Takenaka, M.; Sawamoto, M. Amphiphilic Random Copolymers with Hydrophobic/Hydrogen-Bonding Urea Pendants: Self-Folding Polymers in Aqueous and Organic Media. *Macromolecules* **2016**, *49* (20), 7917−7927.

https://doi.org/10.1021/acspolymersau.3c00025
ACS Polym. Au 2023, 3, 406−427

(100) Kono, H.; Hibino, M.; Ida, D.; Ouchi, M.; Terashima, T. Self-Assembly of Amphiphilic Alternating Copolymers by Chain Folding in Water: From Uniform Composition and Sequence to Monodisperse Micelles. *Macromolecules* **2023**, *56*, 6086.
(101) Shibata, M.; Matsumoto, M.; Hirai, Y.; Takenaka, M.; Sawamoto, M.; Terashima, T. Intramolecular Folding or Intermolecular Self-Assembly of Amphiphilic Random Copolymers: On-Demand Control by Pendant Design. *Macromolecules* **2018**, *51* (10), 3738−3745.
(102) Pei, D.; Appiah Kubi, G. Developments with Bead-Based Screening for Novel Drug Discovery. *Expert Opinion on Drug Discovery* **2019**, *14* (11), 1097−1102.
(103) Lam, K. S.; Lebl, M.; Krchňák, V. The "One-Bead-One-Compound" Combinatorial Library Method. *Chem. Rev.* **1997**, *97* (2), 411−448.
(104) Gao, Y.; Amar, S.; Pahwa, S.; Fields, G.; Kodadek, T. Rapid Lead Discovery Through Iterative Screening of One Bead One Compound Libraries. *ACS Comb. Sci.* **2015**, *17* (1), 49−59.
(105) Knight, A. S.; Zhou, E. Y.; Pelton, J. G.; Francis, M. B. Selective Chromium(VI) Ligands Identified Using Combinatorial Peptoid Libraries. *J. Am. Chem. Soc.* **2013**, *135* (46), 17488−17493.
(106) Knight, A. S.; Kulkarni, R. U.; Zhou, E. Y.; Franke, J. M.; Miller, E. W.; Francis, M. B. A Modular Platform to Develop Peptoid-Based Selective Fluorescent Metal Sensors. *Chem. Commun.* **2017**, *53* (24), 3477−3480.
(107) Paulick, M. G.; Hart, K. M.; Brinner, K. M.; Tjandra, M.; Charych, D. H.; Zuckermann, R. N. Cleavable Hydrophilic Linker for One-Bead-One-Compound Sequencing of Oligomer Libraries by Tandem Mass Spectrometry. *J. Comb. Chem.* **2006**, *8* (3), 417−426.
(108) Ren, J.; Tian, Y.; Hossain, E.; Ho, J. S.; Mann, Y. S.; Zhang, Y.; Browne, M. D.; Connolly, M. D.; Zuckermann, R. N. Mass Spectrometry Studies of the Fragmentation Patterns and Mechanisms of Protonated Peptoids. *Biopolymers* **2020**, e23358.
(109) Witus, L. S.; Moore, T.; Thuronyi, B. W.; Esser-Kahn, A. P.; Scheck, R. A.; Iavarone, A. T.; Francis, M. B. Identification of Highly Reactive Sequences for PLP-Mediated Bioconjugation Using a Combinatorial Peptide Library. *J. Am. Chem. Soc.* **2010**, *132* (47), 16812−16817.
(110) Zimmermann, G.; Neri, D. DNA-Encoded Chemical Libraries: Foundations and Applications in Lead Discovery. *Drug Discovery Today* **2016**, *21* (11), 1828−1834.
(111) Rössler, S. L.; Grob, N. M.; Buchwald, S. L.; Pentelute, B. L. Abiotic Peptides as Carriers of Information for the Encoding of Small-Molecule Library Synthesis. *Science* **2023**, *379* (6635), 939−945.
(112) Gironda-Martínez, A.; Donckele, E. J.; Samain, F.; Neri, D. DNA-Encoded Chemical Libraries: A Comprehensive Review with Succesful Stories and Future Challenges. *ACS Pharmacol. Transl. Sci.* **2021**, *4* (4), 1265−1279.
(113) Lu, X.; Watts, E.; Jia, F.; Tan, X.; Zhang, K. Polycondensation of Polymer Brushes via DNA Hybridization. *J. Am. Chem. Soc.* **2014**, *136* (29), 10214−10217.
(114) Mondal, T.; Nerantzaki, M.; Flesch, K.; Loth, C.; Maaloum, M.; Cong, Y.; Sheiko, S. S.; Lutz, J.-F. Large Sequence-Defined Supramolecules Obtained by the DNA-Guided Assembly of Biohybrid Poly(Phosphodiester)s. *Macromolecules* **2021**, *54*, 3423.
(115) Kim, J.; Vaughan, H. J.; Zamboni, C. G.; Sunshine, J. C.; Green, J. J. High-Throughput Evaluation of Polymeric Nanoparticles for Tissue-Targeted Gene Expression Using Barcoded Plasmid DNA. *J. Controlled Release* **2021**, *337*, 105−116.
(116) Gibson, M. I.; Frohlich, E.; Klok, H.-A. Postpolymerization Modification of Poly(Pentafluorophenyl Methacrylate): Synthesis of a Diverse Water-Soluble Polymer Library. *J. Polym. Sci., Part A: Polym. Chem.* **2009**, *47*, 4332.
(117) Le Neindre, M.; Nicolaÿ, R. One-pot Deprotection and Functionalization of Polythiol Copolymers via Six Different Thiol-X Reactions. *Polymer international* **2014**, *63* (5), 887−893.
(118) Nicolaÿ, R. Synthesis of Well-Defined Polythiol Copolymers by RAFT Polymerization. *Macromolecules* **2012**, *45* (2), 821−827.
(119) Bachler, P. R.; Forry, K. E.; Sparks, C. A.; Schulz, M. D.; Wagener, K. B.; Sumerlin, B. S. Modular Segmented Hyperbranched Copolymers. *Polym. Chem.* **2016**, *7* (25), 4155−4159.
(120) Harrison, E. E.; Waters, M. L. Detection and Differentiation of Per- and Polyfluoroalkyl Substances (PFAS) in Water Using a Fluorescent Imprint-and-Report Sensor Array. *Chem. Sci.* **2023**, 928.
(121) Mahon, C. S.; Jackson, A. W.; Murray, B. S.; Fulton, D. A. Investigating Templating within Polymer-Scaffolded Dynamic Combinatorial Libraries. *Polym. Chem.* **2013**, *4* (2), 368−377.
(122) Cragg, L. H.; Hammerschlag, H. Hammerschlag, Hanna. The Fractionation of High-Polymeric Substances. *Chem. Rev.* **1946**, *39* (1), 79−135.
(123) 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 Morphological Analysis of ABC Triblock Terpolymer Libraries. *Macromolecules* **2022**, *55* (19), 8875−8882.
(124) Lawrence, J.; Lee, S. H.; Abdilla, A.; Nothling, M. D.; Ren, J. M.; Knight, A. S.; Fleischmann, C.; Li, Y.; Abrams, A. S.; Schmidt, B. V. K. J.; Hawker, M. C.; Connal, L. A.; McGrath, A. J.; Clark, P. G.; Gutekunst, W. R.; Hawker, C. J. A Versatile and Scalable Strategy to Discrete Oligomers. *J. Am. Chem. Soc.* **2016**, *138* (19), 6306−6310.
(125) Bosman, A. W.; Heumann, A.; Klaerner, G.; Benoit, D.; Fréchet, J. M. J.; Hawker, C. J. High-Throughput Synthesis of Nanoscale Materials: Structural Optimization of Functionalized One-Step Star Polymers. *J. Am. Chem. Soc.* **2001**, *123* (26), 6461−6462.
(126) Hoogenboom, R.; Wiesbrock, F.; Leenen, M. A. M.; Meier, M. A. R.; Schubert, U. S. Accelerating the Living Polymerization of 2-Nonyl-2-Oxazoline by Implementing a Microwave Synthesizer into a High-Throughput Experimentation Workflow. *J. Comb. Chem.* **2005**, *7* (1), 10−13.
(127) Baudis, S.; Lendlein, A.; Behl, M. High-Throughput Synthesis as a Technology Platform for Copolymer Libraries: High-Throughput Synthesis as a Technology Platform. *Macromol. Symp.* **2014**, *345* (1), 105−111.
(128) Phommalysack-Lovan, J.; Chu, Y.; Boyer, C.; Xu, J. PET-RAFT Polymerisation: Towards Green and Precision Polymer Manufacturing. *Chem. Commun.* **2018**, *54* (50), 6591−6606.
(129) Upadhya, R.; Kanagala, M. J.; Gormley, A. J. Purifying Low-Volume Combinatorial Polymer Libraries with Gel Filtration Columns. *Macromol. Rapid Commun.* **2019**, *40* (24), 1900528.
(130) Li, Z.; Han, Z.; Stenzel, M. H.; Chapman, R. A High Throughput Approach for Designing Polymers That Mimic the TRAIL Protein. *Nano Lett.* **2022**, *22* (7), 2660−2666.
(131) Upadhya, R.; Murthy, N. S.; Hoop, C. L.; Kosuri, S.; Nanda, V.; Kohn, J.; Baum, J.; Gormley, A. J. PET-RAFT and SAXS: High Throughput Tools to Study Compactness and Flexibility of Single-Chain Polymer Nanoparticles. *Macromolecules* **2019**, *52* (21), 8295−8304.
(132) Luo, Y.; Gu, M.; Edwards, C. E. R.; Valentine, M. T.; Helgeson, M. E. High-Throughput Microscopy to Determine Morphology, Microrheology, and Phase Boundaries Applied to Phase Separating Coacervates. *Soft Matter* **2022**, *18* (15), 3063−3075.
(133) Vriza, A.; Chan, H.; Xu, J. Self-Driving Laboratory for Polymer Electronics. *Chem. Mater.* **2023**, *35* (8), 3046−3056.
(134) Reis, M. H.; Leibfarth, F. A.; Pitet, L. M. Polymerizations in Continuous Flow: Recent Advances in the Synthesis of Diverse Polymeric Materials. *ACS Macro Lett.* **2020**, *9* (1), 123−133.
(135) Taylor, N. G.; Reis, M. H.; Varner, T. P.; Rapp, J. L.; Sarabia, A.; Leibfarth, F. A. A Dual Initiator Approach for Oxygen Tolerant RAFT Polymerization. *Polym. Chem.* **2022**, *13* (33), 4798−4808.
(136) Wicker, A. C.; Leibfarth, F. A.; Jamison, T. F. Flow-IEG Enables Programmable Thermodynamic Properties in Sequence-Defined Unimolecular Macromolecules. *Polym. Chem.* **2017**, *8* (37), 5786−5794.
(137) Reis, M. H.; Davidson, C. L. G.; Leibfarth, F. A. Continuous-Flow Chemistry for the Determination of Comonomer Reactivity Ratios. *Polym. Chem.* **2018**, *9* (13), 1728−1734.

https://doi.org/10.1021/acspolymersau.3c00025
ACS Polym. Au 2023, 3, 406−427

(138) Wigh, D. S.; Goodman, J. M.; Lapkin, A. A. A Review of Molecular Representation in the Age of Machine Learning. *WIREs Computational Molecular Science* **2022**, *12* (5), No. e1603.
(139) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28* (1), 31−36.
(140) Goodman, J. M.; Pletnev, I.; Thiessen, P.; Bolton, E.; Heller, S. R. InChI Version 1.06: Now More than 99.99% Reliable. *J. Cheminform* **2021**, *13* (1), 40.
(141) Lin, T.-S.; Coley, C. W.; Mochigase, H.; Beech, H. K.; Wang, W.; Wang, Z.; Woods, E.; Craig, S. L.; Johnson, J. A.; Kalow, J. A.; Jensen, K. F.; Olsen, B. D. BigSMILES: A Structurally-Based Line Notation for Describing Macromolecules. *ACS Cent. Sci.* **2019**, *5* (9), 1523−1531.
(142) Coley, C. W.; Jin, W.; Rogers, L.; Jamison, T. F.; Jaakkola, T. S.; Green, W. H.; Barzilay, R.; Jensen, K. F. A Graph-Convolutional Neural Network Model for the Prediction of Chemical Reactivity. *Chemical Science* **2019**, *10* (2), 370−377.
(143) Qin, S.; Jin, T.; Van Lehn, R. C.; Zavala, V. M. Predicting Critical Micelle Concentrations for Surfactants Using Graph Convolutional Neural Networks. *J. Phys. Chem. B* **2021**, *125* (37), 10610−10620.
(144) Shi, J.; Rebello, N. J.; Walsh, D.; Zou, W.; Deagen, M. E.; Leao, B. S.; Audus, D. J.; Olsen, B. D. Quantifying Pairwise Similarity for Complex Polymers. *Macromolecules* **2023**, *56*, 7344.
(145) Landrum, G. A. *RDKit: Open-Source Cheminformatics*, 2006. https://www.rdkit.org (accessed 2023-11-02).
(146) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50* (5), 742−754.
(147) Guo, M.; Shou, W.; Makatura, L.; Erps, T.; Foshey, M.; Matusik, W. Polygrammar: Grammar for Digital Polymer Representation and Generation. *Advanced Science* **2022**, *9* (23), 2101864.
(148) Dalby, A.; Nourse, J. G.; Hounshell, W. D.; Gushurst, A. K. I.; Grier, D. L.; Leland, B. A.; Laufer, J. Description of Several Chemical Structure File Formats Used by Computer Programs Developed at Molecular Design Limited. *J. Chem. Inf. Comput. Sci.* **1992**, *32* (3), 244−255.
(149) Li, M. M.; Huang, K.; Zitnik, M. Graph Representation Learning in Biomedicine and Healthcare. *Nat. Biomed. Eng.* **2022**, *6* (12), 1353−1369.
(150) Atz, K.; Grisoni, F.; Schneider, G. Geometric Deep Learning on Molecular Representations. *Nat. Mach Intell* **2021**, *3* (12), 1023−1032.
(151) Heid, E.; Green, W. H. Machine Learning of Reaction Properties via Learned Representations of the Condensed Graph of Reaction. *J. Chem. Inf. Model.* **2022**, *62* (9), 2101−2110.
(152) Na, G. S.; Jang, S.; Lee, Y.-L.; Chang, H. Tuplewise Material Representation Based Machine Learning for Accurate Band Gap Prediction. *J. Phys. Chem. A* **2020**, *124* (50), 10616−10623.
(153) Ismail, I.; Chantreau Majerus, R.; Habershon, S. Graph-Driven Reaction Discovery: Progress, Challenges, and Future Opportunities. *J. Phys. Chem. A* **2022**, *126* (40), 7051−7069.
(154) Leonard, K. C.; Hasan, F.; Sneddon, H. F.; You, F. Can Artificial Intelligence and Machine Learning Be Used to Accelerate Sustainable Chemistry and Engineering? *ACS Sustainable Chem. Eng.* **2021**, *9* (18), 6126−6129.
(155) Prihoda, D.; Maritz, J. M.; Klempir, O.; Dzamba, D.; Woelk, C. H.; Hazuda, D. J.; Bitton, D. A.; Hannigan, G. D. The Application Potential of Machine Learning and Genomics for Understanding Natural Product Diversity, Chemistry, and Therapeutic Translatability. *Natural Product Reports* **2021**, *38* (6), 1100−1108.
(156) Wang, Z.; Combs, S. A.; Brand, R.; Calvo, M. R.; Xu, P.; Price, G.; Golovach, N.; Salawu, E. O.; Wise, C. J.; Ponnapalli, S. P.; Clark, P. M. LM-GVP: An Extensible Sequence and Structure Informed Deep Learning Framework for Protein Property Prediction. *Sci. Rep* **2022**, *12* (1), 6832.
(157) Öztürk, H.; Özgür, A.; Schwaller, P.; Laino, T.; Ozkirimli, E. Exploring Chemical Space Using Natural Language Processing Methodologies for Drug Discovery. *Drug Discovery Today* **2020**, *25* (4), 689−705.
(158) Hirohara, M.; Saito, Y.; Koda, Y.; Sato, K.; Sakakibara, Y. Convolutional Neural Network Based on SMILES Representation of Compounds for Detecting Chemical Motif. *BMC Bioinformatics* **2018**, *19* (19), 526.
(159) Mohapatra, S.; An, J.; Gómez-Bombarelli, R. Chemistry-Informed Macromolecule Graph Representation for Similarity Computation, Unsupervised and Supervised Learning. *Mach. Learn.: Sci. Technol.* **2022**, *3* (1), 015028.
(160) Ethier, J. G.; Casukhela, R. K.; Latimer, J. J.; Jacobsen, M. D.; Rasin, B.; Gupta, M. K.; Baldwin, L. A.; Vaia, R. A. Predicting Phase Behavior of Linear Polymers in Solution Using Machine Learning. *Macromolecules* **2022**, *55* (7), 2691−2702.
(161) Aldeghi, M. W.; Coley, C. A Graph Representation of Molecular Ensembles for Polymer Property Prediction. *Chemical Science* **2022**, *13* (35), 10486−10498.
(162) Antoniuk, E. R.; Li, P.; Kailkhura, B.; Hiszpanski, A. M. Representing Polymers as Periodic Graphs with Learned Descriptors for Accurate Polymer Property Predictions. *J. Chem. Inf. Model.* **2022**, *62* (22), 5435−5445.
(163) Xu, C.; Wang, Y.; Barati Farimani, A. TransPolymer: A Transformer-Based Language Model for Polymer Property Predictions. *npj Comput. Mater.* **2023**, *9* (1), 1−14.
(164) Minkiewicz, P.; Iwaniak, A.; Darewicz, M. Annotation of Peptide Structures Using SMILES and Other Chemical Codes-Practical Solutions. *Molecules* **2017**, *22* (12), 2075.
(165) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583−589.
(166) Vani, B. P.; Aranganathan, A.; Wang, D.; Tiwary, P. AlphaFold2-RAVE: From Sequence to Boltzmann Ranking. *J. Chem. Theory Comput.* **2023**, *19*, 4351.
(167) Kingsbury, J. S.; Saini, A.; Auclair, S. M.; Fu, L.; Lantz, M. M.; Halloran, K. T.; Calero-Rubio, C.; Schwenger, W.; Airiau, C. Y.; Zhang, J.; Gokarn, Y. R. A Single Molecular Descriptor to Predict Solution Behavior of Therapeutic Antibodies. *Science Advances* **2020**, *6* (32), No. eabb0372.
(168) Mei, H.; Liao, Z. H.; Zhou, Y.; Li, S. Z. A New Set of Amino Acid Descriptors and Its Application in Peptide QSARs. *Peptide Science* **2005**, *80* (6), 775−786.
(169) Chen, Z.; Zhao, P.; Li, F.; Leier, A.; Marquez-Lago, T. T.; Wang, Y.; Webb, G. I.; Smith, A. I.; Daly, R. J.; Chou, K.-C.; Song, J. iFeature: A Python Package and Web Server for Features Extraction and Selection from Protein and Peptide Sequences. *Bioinformatics* **2018**, *34* (14), 2499−2502.
(170) Butterfoss, G. L.; Yoo, B.; Jaworski, J. N.; Chorny, I.; Dill, K. A.; Zuckermann, R. N.; Bonneau, R.; Kirshenbaum, K.; Voelz, V. A. De Novo Structure Prediction and Experimental Characterization of Folded Peptoid Oligomers. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (36), 14320−14325.
(171) Li, Y.-W.; Li, B. Characterization of Structure-Antioxidant Activity Relationship of Peptides in Free Radical Systems Using QSAR Models: Key Sequence Positions and Their Amino Acid Properties. *J. Theor. Biol.* **2013**, *318*, 29−43.
(172) Tang, T.; Hazra, A.; Min, D. S.; Williams, W. L.; Jones, E.; Doyle, A. G.; Sigman, M. S. Interrogating the Mechanistic Features of Ni(I)-Mediated Aryl Iodide Oxidative Addition Using Electroanalytical and Statistical Modeling Techniques. *J. Am. Chem. Soc.* **2023**, DOI: 10.1021/jacs.3c01726.
(173) Xu, P.; Lu, T.; Ju, L.; Tian, L.; Li, M.; Lu, W. Machine Learning Aided Design of Polymer with Targeted Band Gap Based on DFT Computation. *J. Phys. Chem. B* **2021**, *125* (2), 601−611.

(174) Xu, X.; Wei, Q.; Li, H.; Wang, Y.; Chen, Y.; Jiang, Y. Recognition of Polymer Configurations by Unsupervised Learning. *Phys. Rev. E* **2019**, *99* (4), 043307.
(175) Bhattacharya, D.; Patra, T. K. dPOLY: Deep Learning of Polymer Phases and Phase Transition. *Macromolecules* **2021**, *54* (7), 3065−3074.
(176) Bejagam, K. K.; An, Y.; Singh, S.; Deshmukh, S. A. Machine-Learning Enabled New Insights into the Coil-to-Globule Transition of Thermosensitive Polymers Using a Coarse-Grained Model. *J. Phys. Chem. Lett.* **2018**, *9* (22), 6480−6488.
(177) Noé, F.; Schütte, C.; Vanden-Eijnden, E.; Reich, L.; Weikl, T. R. Constructing the Equilibrium Ensemble of Folding Pathways from Short Off-Equilibrium Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (45), 19011−19016.
(178) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov Models of Molecular Kinetics: Generation and Validation. *J. Chem. Phys.* **2011**, *134* (17), 174105.
(179) Yap, C. W. PaDEL-Descriptor: An Open Source Software to Calculate Molecular Descriptors and Fingerprints. *J. Comput. Chem.* **2011**, *32* (7), 1466−1474.
(180) Moriwaki, H.; Tian, Y.-S.; Kawashita, N.; Takagi, T. Mordred: A Molecular Descriptor Calculator. *Journal of Cheminformatics* **2018**, *10* (1), 4.
(181) Dong, J.; Cao, D.-S.; Miao, H.-Y.; Liu, S.; Deng, B.-C.; Yun, Y.-H.; Wang, N.-N.; Lu, A.-P.; Zeng, W.-B.; Chen, A. F. ChemDes: An Integrated Web-Based Platform for Molecular Descriptor and Fingerprint Computation. *Journal of Cheminformatics* **2015**, *7* (1), 60.
(182) Liu, T.; Johnson, K. R.; Jansone-Popova, S.; Jiang, D. Advancing Rare-Earth Separation by Machine Learning. *JACS Au* **2022**, *2* (6), 1428−1434.
(183) Zhou, P.; Liu, Q.; Wu, T.; Miao, Q.; Shang, S.; Wang, H.; Chen, Z.; Wang, S.; Wang, H. Systematic Comparison and Comprehensive Evaluation of 80 Amino Acid Descriptors in Peptide QSAR Modeling. *J. Chem. Inf. Model.* **2021**, *61* (4), 1718−1731.
(184) Payne, E. M.; Holland-Moritz, D. A.; Sun, S.; Kennedy, R. T. High-Throughput Screening by Droplet Microfluidics: Perspective into Key Challenges and Future Prospects. *Lab Chip* **2020**, *20* (13), 2247−2262.
(185) Xiao, H.; Bao, Z.; Zhao, H. High Throughput Screening and Selection Methods for Directed Enzyme Evolution. *Ind. Eng. Chem. Res.* **2015**, *54* (16), 4011−4020.
(186) Turner, N. J. Directed Evolution Drives the next Generation of Biocatalysts. *Nat. Chem. Biol.* **2009**, *5* (8), 567−573.
(187) Markel, U.; Essani, K. D.; Besirlioglu, V.; Schiffels, J.; Streit, W. R.; Schwaneberg, U. Advances in Ultrahigh-Throughput Screening for Directed Enzyme Evolution. *Chem. Soc. Rev.* **2020**, *49* (1), 233−262.
(188) Auld, D. S.; Klumpp-Thomas, C.; Michael, S.; Sittampalam, G. S.; Trask, O. J.; Wildey, J. Microplate Selection and Recommended Practices in High-Throughput Screening and Quantitative Biology. In *Assay Guidance Manual*; Eli Lilly & Company and the National Center for Advancing Translational Sciences: Bethesda, MD, 2020.
(189) Govindarajan, R.; Duraiyan, J.; Kaliyappan, K.; Palanisamy, M. Microarray and Its Applications. *J. Pharm. Bioall Sci.* **2012**, *4* (6), 310.
(190) Horton, M.; Su, G.; Yi, L.; Wang, Z.; Xu, Y.; Pagadala, V.; Zhang, F.; Zaharoff, D. A.; Pearce, K.; Linhardt, R. J.; Liu, J. Construction of Heparan Sulfate Microarray for Investigating the Binding of Specific Saccharide Sequences to Proteins. *Glycobiology* **2021**, *31* (3), 188−199.
(191) Sun, Y. S.; Landry, J. P.; Fei, Y. Y.; Zhu, X. D.; Luo, J. T.; Wang, X. B.; Lam, K. S. Macromolecular Scaffolds for Immobilizing Small Molecule Microarrays in Label-Free Detection of Protein-Ligand Interactions on Solid Support. *Anal. Chem.* **2009**, *81* (13), 5373−5380.
(192) Lam, K. S.; Lebl, M. Selectide Technology: Bead-Binding Screening. *Methods* **1994**, *6* (4), 372−380.
(193) Meldal, M. Pega: A Flow Stable Polyethylene Glycol Dimethyl Acrylamide Copolymer for Solid Phase Synthesis. *Tetrahedron Lett.* **1992**, *33* (21), 3077−3080.
(194) Murrell, E.; Luyt, L. G. Incorporation of Fluorine into an OBOC Peptide Library by Copper-Free Click Chemistry toward the Discovery of PET Imaging Agents. *ACS Comb. Sci.* **2020**, *22* (3), 109−113.
(195) Komnatnyy, V. V.; Nielsen, T. E.; Qvortrup, K. Bead-Based Screening in Chemical Biology and Drug Discovery. *Chem. Commun.* **2018**, *54* (50), 6759−6771.
(196) Cho, C.-F.; Behnam Azad, B.; Luyt, L. G.; Lewis, J. D. High-Throughput Screening of One-Bead-One-Compound Peptide Libraries Using Intact Cells. *ACS Comb. Sci.* **2013**, *15* (8), 393−400.
(197) Cha, J.; Lim, J.; Zheng, Y.; Tan, S.; Ang, Y. L.; Oon, J.; Ang, M. W.; Ling, J.; Bode, M.; Lee, S. S. Process Automation toward Ultra-High-Throughput Screening of Combinatorial One-Bead-One-Compound (OBOC) Peptide Libraries. *SLAS Technology* **2012**, *17* (3), 186−200.
(198) Maag, P. H.; Feist, F.; Frisch, H.; Roesky, P. W.; Barner-Kowollik, C. Fluorescent and Catalytically Active Single Chain Nanoparticles. *Macromolecules* **2022**, *55*, 9918.
(199) Bhirde, A.; Chikkaveeraiah, B. V.; Venna, R.; Carley, R.; Brorson, K.; Agarabi, C. High Performance Size Exclusion Chromatography and High-Throughput Dynamic Light Scattering as Orthogonal Methods to Screen for Aggregation and Stability of Monoclonal Antibody Drug Products. *J. Pharm. Sci.* **2020**, *109* (11), 3330−3339.
(200) Maag, P. H.; Feist, F.; Frisch, H.; Roesky, P. W.; Barner-Kowollik, C. Fluorescent and Catalytically Active Single Chain Nanoparticles. *Macromolecules* **2022**, *55* (22), 9918−9924.
(201) Dauer, K.; Pfeiffer-Marek, S.; Kamm, W.; Wagner, K. G. Microwell Plate-Based Dynamic Light Scattering as a High-Throughput Characterization Tool in Biopharmaceutical Development. *Pharmaceutics* **2021**, *13* (2), 172.
(202) He, F.; Becker, G. W.; Litowski, J. R.; Narhi, L. O.; Brems, D. N.; Razinkov, V. I. High-Throughput Dynamic Light Scattering Method for Measuring Viscosity of Concentrated Protein Solutions. *Anal. Biochem.* **2010**, *399* (1), 141−143.
(203) Palmieri, V.; Lucchetti, D.; Gatto, I.; Maiorana, A.; Marcantoni, M.; Maulucci, G.; Papi, M.; Pola, R.; De Spirito, M.; Sgambato, A. Dynamic Light Scattering for the Characterization and Counting of Extracellular Vesicles: A Powerful Noninvasive Tool. *J. Nanopart. Res.* **2014**, *16* (9), 2583.
(204) Norman, A. I.; Cabral, J. T.; Ho, D. L.; Amis, E. J.; Karim, A. Scattering Methods Applied to High Throughput Materials Science. *Polymeric Materials: Science & Engineering* **2004**, *90*, 339.
(205) Wei, Y.; Hore, M. J. A. Characterizing Polymer Structure with Small- Angle Neutron Scattering : A Tutorial. *J. Appl. Phys.* **2021**, *129*, 171101.
(206) González-Burgos, M.; Asenjo-Sanz, I.; Pomposo, J. A.; Radulescu, A.; Ivanova, O.; Pasini, S.; Arbe, A.; Colmenero, J. Structure and Dynamics of Irreversible Single-Chain Nanoparticles in Dilute Solution. A Neutron Scattering Investigation. *Macromolecules* **2020**, *53* (18), 8068−8082.
(207) Hura, G. L.; Menon, A. L.; Hammel, M.; Rambo, R. P.; Poole Ii, F. L.; Tsutakawa, S. E.; Jenney, F. E., Jr; Classen, S.; Frankel, K. A.; Hopkins, R. C.; Yang, S.; Scott, J. W.; Dillard, B. D.; Adams, M. W. W.; Tainer, J. A. Robust, High-Throughput Solution Structural Analyses by Small Angle X-Ray Scattering (SAXS). *Nat. Methods* **2009**, *6* (8), 606−612.
(208) Dyer, K. N.; Hammel, M.; Rambo, R. P.; Tsutakawa, S. E.; Rodic, I.; Classen, S.; Tainer, J. A.; Hura, G. L. High-Throughput SAXS for the Characterization of Biomolecules in Solution: A Practical Approach. In *Structural Genomics*; Chen, Y. W., Ed.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2014; Vol. *1091*, pp 245−258.
(209) Guild, J. D.; Knox, S. T.; Burholt, S. B.; Hilton, E.M.; Terrill, N. J.; Schroeder, S. L. M.; Warren, N. J. Continuous-Flow Laboratory SAXS for In Situ Determination of the Impact of Hydrophilic Block

Length on Spherical Nano-Object Formation during Polymerization-Induced Self-Assembly. *Macromolecules* **2023**, *56* (16), 6426−6435.
(210) Badea, E.; Della Gatta, G.; Usacheva, T. Effects of Temperature and Relative Humidity on Fibrillar Collagen in Parchment: A Micro Differential Scanning Calorimetry (Micro DSC) Study. *Polym. Degrad. Stab.* **2012**, *97* (3), 346−353.
(211) Vollrath, F.; Hawkins, N.; Porter, D.; Holland, C.; Boulet-Audet, M. Differential Scanning Fluorimetry Provides High Throughput Data on Silk Protein Transitions. *Sci. Rep* **2014**, *4* (1), 5625.
(212) Ciulli, A. Biophysical Screening for the Discovery of Small-Molecule Ligands. In *Protein-Ligand Interactions*; Williams, M. A., Daviter, T., Eds.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2013; Vol. *1008*, pp 357−388.
(213) Ashley, J.; Shukor, Y.; Tothill, I. E. The Use of Differential Scanning Fluorimetry in the Rational Design of Plastic Antibodies for Protein Targets. *Analyst* **2016**, *141* (23), 6463−6470.
(214) Wu, T.; Yu, J. C.; Suresh, A.; Gale-Day, Z. J.; Alteen, M. G.; Woo, A. S.; Millbern, Z.; Johnson, O. T.; Carroll, E. C.; Partch, C. L.; Fourches, D.; Vinueza, N. R.; Vocadlo, D. J.; Gestwicki, J. E.Conformationally Responsive Dyes Enable Protein-Adaptive Differential Scanning Fluorimetry. *bioRxiv*, January 28, 2023, ver. 1.
(215) Wu, T.; Yu, J.; Gale-Day, Z.; Woo, A.; Suresh, A.; Hornsby, M.; Gestwicki, J. E. Three Essential Resources to Improve Differential Scanning Fluorimetry (DSF) Experiments. *bioRiv*, March 25, 2020, ver. 1.
(216) Stahelin, R. V. Surface Plasmon Resonance: A Useful Technique for Cell Biologists to Characterize Biomolecular Interactions. *MBoC* **2013**, *24* (7), 883−886.
(217) Vollmer, N.; Trombini, F.; Hely, M.; Bellon, S.; Mercier, K.; Cazeneuve, C. Methodology to Study Polymers Interaction by Surface Plasmon Resonance Imaging. *MethodsX* **2015**, *2*, 14−18.
(218) Bergsdorf, C.; Ottl, J. Affinity-Based Screening Techniques: Their Impact and Benefit to Increase the Number of High Quality Leads. *Expert Opinion on Drug Discovery* **2010**, *5* (11), 1095−1107.
(219) Annis, D. A.; Nickbarg, E.; Yang, X.; Ziebell, M. R.; Whitehurst, C. E. Affinity Selection-Mass Spectrometry Screening Techniques for Small Molecule Drug Discovery. *Curr. Opin. Chem. Biol.* **2007**, *11* (5), 518−526.
(220) Zhang, Y.; Yan, J.; Xu, J.; Tian, C.; Matyjaszewski, K.; Tilton, R. D.; Lowry, G. V. Phosphate Polymer Nanogel for Selective and Efficient Rare Earth Element Recovery. *Environ. Sci. Technol.* **2021**, *55* (18), 12549−12560.
(221) Kumarasamy, E.; Manning, I. M.; Collins, L. B.; Coronell, O.; Leibfarth, F. A. Ionic Fluorogels for Remediation of Per- and Polyfluorinated Alkyl Substances from Water. *ACS Central Science* **2020**, *6*, 487.
(222) Pomplun, S.; Jbara, M.; Quartararo, A. J.; Zhang, G.; Brown, J. S.; Lee, Y.-C.; Ye, X.; Hanna, S.; Pentelute, B. L. *De Novo* Discovery of High-Affinity Peptide Binders for the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* **2021**, *7* (1), 156−163.
(223) Freyer, M. W.; Lewis, E. A. Isothermal Titration Calorimetry: Experimental Design, Data Analysis, and Probing Macromolecule/Ligand Binding and Kinetic Interactions. *Methods in Cell Biology* **2008**, *84* (07), 79−113.
(224) Park, J.; Cleary, M. B.; Li, D.; Mattocks, J. A.; Xu, J.; Wang, H.; Mukhopadhyay, S.; Gale, E. M.; Cotruvo, J. A. A Genetically Encoded Fluorescent Sensor for Manganese(II), Engineered from Lanmodulin. *Proc. Natl. Acad. Sci. U.S.A.* **2022**, *119* (51), No. e2212723119.
(225) Ramirez, J.; Nominé, Y. High-Quality Data of Protein/Peptide Interaction by Isothermal Titration Calorimetry. In *Microcalorimetry of Biological Molecules*; Ennifar, E., Ed.; Methods in Molecular Biology; Springer: New York, 2019; Vol. *1964*, pp 99−117.
(226) Gutenthaler, S. M.; Tsushima, S.; Steudtner, R.; Gailer, M.; Hoffmann-Röder, A.; Drobot, B.; Daumann, L. J. Lanmodulin Peptides - Unravelling the Binding of the EF-Hand Loop Sequences Stripped from the Structural Corset. *Inorg. Chem. Front.* **2022**, *9* (16), 4009−4021.
(227) Archer, W. R.; Schulz, M. D. Isothermal Titration Calorimetry: Practical Approaches and Current Applications in Soft Matter. *Soft Matter* **2020**, *16* (38), 8760−8774.
(228) Knight, A. S.; Zhou, E. Y.; Francis, M. B. Development of Peptoid-Based Ligands for the Removal of Cadmium from Biological Media. *Chem. Sci.* **2015**, *6* (7), 4042−4048.
(229) Robb, C. G.; Dao, T. P.; Ujma, J.; Castañeda, C. A.; Beveridge, R. Ion Mobility Mass Spectrometry Unveils Global Protein Conformations in Response to Conditions That Promote and Reverse Liquid-Liquid Phase Separation. *J. Am. Chem. Soc.* **2023**, *145*, 12541.
(230) Stuchfield, D.; Barran, P. Unique Insights to Intrinsically Disordered Proteins Provided by Ion Mobility Mass Spectrometry. *Curr. Opin. Chem. Biol.* **2018**, *42*, 177−185.
(231) Weber, P.; Hoyas, S.; Halin, E.; Coulembier, O.; De Winter, J.; Cornil, J.; Gerbaux, P. On the Conformation of Anionic Peptoids in the Gas Phase. *Biomacromolecules* **2022**, *23* (3), 1138−1147.
(232) Foley, C. D.; Zhang, B.; Alb, A. M.; Trimpin, S.; Grayson, S. M. Use of Ion Mobility Spectrometry-Mass Spectrometry to Elucidate Architectural Dispersity within Star Polymers. *ACS Macro Lett.* **2015**, *4* (7), 778−782.
(233) Besford, Q. A.; Yong, H.; Merlitz, H.; Christofferson, A. J.; Sommer, J.; Uhlmann, P.; Fery, A. FRET-Integrated Polymer Brushes for Spatially Resolved Sensing of Changes in Polymer Conformation. *Angew. Chem., Int. Ed.* **2021**, *60* (30), 16600−16606.
(234) Wang, Y.; Fortenberry, A. W.; Zhang, W.; Simon, Y. C.; Qiang, Z. Direct Measurement of Polymer-Chain-End-to-End Distances by Using RAFT Chain Transfer Agent as the FRET Acceptor. *J. Phys. Chem. B* **2023**, *127* (13), 3100−3108.
(235) Ohno, H.; Shimidzu, N.; Tsuchida, E.; Sasakawa, S.; Honda, K. Fluorescence Polarization Study on the Increase of Membrane Fluidity of Human Erythrocyte Ghosts Induced by Synthetic Water-Soluble Polymers. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1981**, *649* (2), 221−228.
(236) Viovy, J. L.; Monnerie, L.; Brochon, J. C. Fluorescence Polarization Decay Study of Polymer Dynamics: A Critical Discussion of Models Using Synchrotron Data. *Macromolecules* **1983**, *16* (12), 1845−1852.
(237) Chee, C. K.; Rimmer, S.; Soutar, I.; Swanson, L. Time-Resolved Fluorescence Anisotropy Studies of the Interaction of N-Isopropyl Acrylamide Based Polymers with Sodium Dodecyl Sulphate. *Soft Matter* **2011**, *7* (10), 4705.
(238) Thomsson, D.; Lin, H.; Scheblykin, I. G. Correlation Analysis of Fluorescence Intensity and Fluorescence Anisotropy Fluctuations in Single-Molecule Spectroscopy of Conjugated Polymers. *ChemPhysChem* **2010**, *11* (4), 897−904.
(239) Xie, S.; Wong, A. Y. H.; Chen, S.; Tang, B. Z. Fluorogenic Detection and Characterization of Proteins by Aggregation-Induced Emission Methods. *Chem.—Eur. J.* **2019**, *25* (23), 5824−5847.
(240) Hawe, A.; Sutter, M.; Jiskoot, W. Extrinsic Fluorescent Dyes as Tools for Protein Characterization. *Pharm. Res.* **2008**, *25* (7), 1487−1499.
(241) Kamiloglu, S.; Sari, G.; Ozdal, T.; Capanoglu, E. Guidelines for Cell Viability Assays. *Food Frontiers* **2020**, *1* (3), 332−349.
(242) Li, Y.; Beija, M.; Laurent, S.; Elst, L. V.; Muller, R. N.; Duong, H. T. T.; Lowe, A. B.; Davis, T. P.; Boyer, C. Macromolecular Ligands for Gadolinium MRI Contrast Agents. *Macromolecules* **2012**, *45* (10), 4196−4204.
(243) Hogendoorn, C.; Roszczenko-Jasińska, P.; Martinez-Gomez, N. C.; de Graaff, J.; Grassl, P.; Pol, A.; Op den Camp, H. J. M.; Daumann, L. J. Facile Arsenazo III-Based Assay for Monitoring Rare Earth Element Depletion from Cultivation Media for Methanotrophic and Methylotrophic Bacteria. *Appl. Environ. Microbiol.* **2018**, *84* (8), No. e02887-17.
(244) Lee, B.-C.; Chu, T. K.; Dill, K. A.; Zuckermann, R. N. Biomimetic Nanostructures: Creating a High-Affinity Zinc-Binding Site in a Folded Nonbiological Polymer. *J. Am. Chem. Soc.* **2008**, *130* (27), 8847−8855.
(245) Parker, B. F.; Knight, A. S.; Vukovic, S.; Arnold, J.; Francis, M. B. A Peptoid-Based Combinatorial and Computational Approach to

https://doi.org/10.1021/acspolymersau.3c00025
ACS Polym. Au 2023, 3, 406−427

Developing Ligands for Uranyl Sequestration from Seawater. *Ind. Eng. Chem. Res.* **2016**, *55* (15), 4187−4194.
(246) Sackett, D. L.; Wolff, J. Nile Red as a Polarity-Sensitive Fluorescent Probe of Hydrophobic Protein Surfaces. *Anal. Biochem.* **1987**, *167* (2), 228−234.
(247) Dutta, A. K.; Kamada, K.; Ohta, K. Spectroscopic Studies of Nile Red in Organic Solvents and Polymers. *J. Photochem. Photobiol., A* **1996**, *93* (1), 57−64.
(248) Greenspan, P.; Mayer, E. P.; Fowler, S. D. Nile Red: A Selective Fluorescent Stain for Intracellular Lipid Droplets. *J. Cell Biol.* **1985**, *100* (3), 965−973.
(249) Fletcher, K. A.; Storey, I. A.; Hendricks, A. E.; Pandey, S.; Pandey, S. Behavior of the Solvatochromic Probes Reichardt' s Dye, Pyrene, Dansylamide, Nile Red and 1-Pyrenecarbaldehyde within the Room-Temperature Ionic Liquid bmimPF6. *Green Chem.* **2001**, *3* (5), 210−215.
(250) Murnen, H. K.; Khokhlov, A. R.; Khalatur, P. G.; Segalman, R. A.; Zuckermann, R. N. Impact of Hydrophobic Sequence Patterning on the Coil-to-Globule Transition of Protein-like Polymers. *Macromolecules* **2012**, *45* (12), 5229−5236.
(251) Li, Q.; Zhang, H.; Lou, K.; Yang, Y.; Ji, X.; Zhu, J.; Sessler, J. L. Visualizing Molecular Weights Differences in Supramolecular Polymers. *Proc. Natl. Acad. Sci. U. S. A.* **2022**, *119* (9), 1−9.
(252) Varadaraj, R.; Bock, J.; Valint, P.; Brons, N. Micropolarity and Water Penetration in Micellar Aggregates of Linear and Branched Hydrocarbon Surfactants. *Langmuir* **1990**, *6* (8), 1376−1378.
(253) Fuller, A. A.; Jimenez, C. J.; Martinetto, E. K.; Moreno, J. L.; Calkins, A. L.; Dowell, K. M.; Huber, J.; McComas, K. N.; Ortega, A. Sequence Changes Modulate Peptoid Self-Association in Water. *Frontiers in Chemistry* **2020**, *8* (April), 1−13.
(254) Taylor, C. G.; Meisl, G.; Horrocks, M. H.; Zetterberg, H.; Knowles, T. P. J.; Klenerman, D. Extrinsic Amyloid-Binding Dyes for Detection of Individual Protein Aggregates in Solution. *Anal. Chem.* **2018**, *90* (17), 10385−10393.
(255) Bucevičius, J.; Lukinavičius, G.; Gerasimaitė, R. The Use of Hoechst Dyes for DNA Staining and Beyond. *Chemosensors* **2018**, *6* (2), 18.
(256) Haines, A. M.; Tobe, S. S.; Kobus, H. J.; Linacre, A. Properties of Nucleic Acid Staining Dyes Used in Gel Electrophoresis: Nucleic Acids. *ELECTROPHORESIS* **2015**, *36* (6), 941−944.
(257) Wilkinson-White, L. E.; Easterbrook-Smith, S. B. A Dye-Binding Assay for Measurement of the Binding of Cu(II) to Proteins. *Journal of Inorganic Biochemistry* **2008**, *102* (10), 1831−1838.
(258) Severin, K. Pattern-Based Sensing with Simple Metal-Dye Complexes. *Curr. Opin. Chem. Biol.* **2010**, *14* (6), 737−742.
(259) Sedgwick, A. C.; Brewster, J. T.; Wu, T.; Feng, X.; Bull, S. D.; Qian, X.; Sessler, J. L.; James, T. D.; Anslyn, E. V.; Sun, X. Indicator Displacement Assays (IDAs): The Past, Present and Future. *Chem. Soc. Rev.* **2021**, *50* (1), 9−38.
(260) Artar, M.; Souren, E. R. J.; Terashima, T.; Meijer, E. W.; Palmans, A. R. A. Single Chain Polymeric Nanoparticles as Selective Hydrophobic Reaction Spaces in Water. *ACS Macro Lett.* **2015**, *4* (10), 1099−1103.
(261) Pasch, H.; Kilz, P. Fast Liquid Chromatography for High-Throughput Screening of Polymers. *Macromol. Rapid Commun.* **2003**, *24* (1), 104−108.
(262) Lubbad, S.; Buchmeiser, M. R. Monolithic High-Performance SEC Supports Prepared by ROMP for High-Throughput Screening of Polymers. *Macromol. Rapid Commun.* **2002**, *23* (10−11), 617.
(263) Ivosev, G.; Burton, L.; Bonner, R. Dimensionality Reduction and Visualization in Principal Component Analysis. *Anal. Chem.* **2008**, *80* (13), 4933−4944.
(264) Michel, R.; Pasche, S.; Textor, M.; Castner, D. G. Influence of PEG Architecture on Protein Adsorption and Conformation. *Langmuir* **2005**, *21* (26), 12327−12332.
(265) Tran, N. K.; Howard, T.; Walsh, R.; Pepper, J.; Loegering, J.; Phinney, B.; Salemi, M. R.; Rashidi, H. H. Novel Application of Automated Machine Learning with MALDI-TOF-MS for Rapid High-Throughput Screening of COVID-19: A Proof of Concept. *Sci. Rep* **2021**, *11* (1), 8219.
(266) Miranda, T. M. R.; Gonçalves, A. R.; Amorim, M. T. P. Ultraviolet-Induced Crosslinking of Poly(Vinyl Alcohol) Evaluated by Principal Component Analysis of FTIR Spectra. *Polym. Int.* **2001**, *50* (10), 1068−1072.
(267) Huang, L.; Yu, K.; Zhou, W.; Teng, Q.; Wang, Z.; Dai, Z. Quantitative Principal Component Analysis of Multiple Metal Ions with Lanthanide Coordination Polymer Networks. *Sens. Actuators, B* **2021**, *346*, 130469.
(268) Papaleo, E.; Mereghetti, P.; Fantucci, P.; Grandori, R.; De Gioia, L. Free-Energy Landscape, Principal Component Analysis, and Structural Clustering to Identify Representative Conformations from Molecular Dynamics Simulations: The Myoglobin Case. *Journal of Molecular Graphics and Modelling* **2009**, *27* (8), 889−899.
(269) Bishop, C. J.; Abubaker-Sharif, B.; Guiriba, T.; Tzeng, S. Y.; Green, J. J. Gene Delivery Polymer Structure-Function Relationships Elucidated via Principal Component Analysis. *Chem. Commun.* **2015**, *51* (60), 12134−12137.
(270) Wang, M.; Xu, Q.; Tang, H.; Jiang, J. Machine Learning-Enabled Prediction and High-Throughput Screening of Polymer Membranes for Pervaporation Separation. *ACS Appl. Mater. Interfaces* **2022**, *14* (6), 8427−8436.
(271) Karlov, D. S.; Sosnin, S.; Tetko, I. V.; Fedorov, M. V. Chemical Space Exploration Guided by Deep Neural Networks. *RSC Adv.* **2019**, *9* (9), 5151−5157.
(272) Cihan Sorkun, M.; Mullaj, D.; Koelman, J. M. V. A.; Er, S. ChemPlot, a Python Library for Chemical Space Visualization**. *Chemistry-Methods* **2022**, *2* (7), No. e202200005.
(273) Ray, P.; Reddy, S. S.; Banerjee, T. Various Dimension Reduction Techniques for High Dimensional Data Analysis: A Review. *Artif Intell Rev.* **2021**, *54* (5), 3473−3515.
(274) Aggarwal, C. C.; Hinneburg, A.; Keim, D. A. On the Surprising Behavior of Distance Metrics in High Dimensional Space. In *Database Theory — ICDT 2001*; Van den Bussche, J., Vianu, V., Eds.; Lecture Notes in Computer Science; Springer: Berlin, Heidelberg, 2001; pp 420−434.
(275) Lukasiak, B. M.; Faria, R.; Zomer, S.; Brereton, R. G.; Duncan, J. C. Pattern Recognition for the Analysis of Polymeric Materials. *Analyst* **2006**, *131* (1), 73−80.
(276) Frades, I.; Matthiesen, R. Overview on Techniques in Cluster Analysis. In *Bioinformatics Methods in Clinical Research*; Matthiesen, R., Ed.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2010; pp 81−107.
(277) Akhanli, S. E.; Hennig, C. Comparing Clusterings and Numbers of Clusters by Aggregation of Calibrated Clustering Validity Indexes. *Stat Comput* **2020**, *30* (5), 1523−1544.
(278) Li, X.; Maffettone, P. M.; Che, Y.; Liu, T.; Chen, L.; Cooper, A. I. Combining Machine Learning and High-Throughput Experimentation to Discover Photocatalytically Active Organic Molecules. *Chemical Science* **2021**, *12* (32), 10742−10754.
(279) Liu, A. L.; Venkatesh, R.; McBride, M.; Reichmanis, E.; Meredith, J. C.; Grover, M. A. Small Data Machine Learning: Classification and Prediction of Poly(Ethylene Terephthalate) Stabilizers Using Molecular Descriptors. *ACS Appl. Polym. Mater.* **2020**, *2* (12), 5592−5601.
(280) Bileschi, M. L.; Belanger, D.; Bryant, D. H.; Sanderson, T.; Carter, B.; Sculley, D.; Bateman, A.; DePristo, M. A.; Colwell, L. J. Using Deep Learning to Annotate the Protein Universe. *Nat. Biotechnol.* **2022**, *40* (6), 932−937.
(281) Murdoch, W. J.; Singh, C.; Kumbier, K.; Abbasi-Asl, R.; Yu, B. Definitions, Methods, and Applications in Interpretable Machine Learning. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (44), 22071−22080.
(282) Oviedo, F.; Ferres, J. L.; Buonassisi, T.; Butler, K. T. Interpretable and Explainable Machine Learning for Materials Science and Chemistry. *Acc. Mater. Res.* **2022**, *3* (6), 597−607.
(283) Liu, D.; Xu, Z.; Lu, X.; Yu, H.; Fu, Y. Linear Regression Model for Predicting Allyl Alcohol C-O Bond Activity under Palladium Catalysis. *ACS Catal.* **2022**, *12* (22), 13921−13929.

https://doi.org/10.1021/acspolymersau.3c00025
ACS Polym. Au 2023, 3, 406−427

(284) Polash, A. H.; Nakano, T.; Takeda, S.; Brown, J. B. Applicability Domain of Active Learning in Chemical Probe Identification: Convergence in Learning from Non-Specific Compounds and Decision Rule Clarification. *Molecules* **2019**, *24* (15), 2716.
(285) Wojtuch, A.; Jankowski, R.; Podlewska, S. How Can SHAP Values Help to Shape Metabolic Stability of Chemical Compounds? *J. Cheminform* **2021**, *13* (1), 74.
(286) Yang, J.; Tao, L.; He, J.; McCutcheon, J. R.; Li, Y. Machine Learning Enables Interpretable Discovery of Innovative Polymers for Gas Separation Membranes. *Science Advances* **2022**, *8* (29), No. eabn9545.
(287) Liang, J.; Xu, S.; Hu, L.; Zhao, Y.; Zhu, X. Machine-Learning-Assisted Low Dielectric Constant Polymer Discovery. *Materials Chemistry Frontiers* **2021**, *5* (10), 3823−3829.
(288) Sahu, H.; Li, H.; Chen, L.; Rajan, A. C.; Kim, C.; Stingelin, N.; Ramprasad, R. An Informatics Approach for Designing Conducting Polymers. *ACS Appl. Mater. Interfaces* **2021**, *13* (45), 53314−53322.
(289) Schwaller, P.; Probst, D.; Vaucher, A. C.; Nair, V. H.; Kreutter, D.; Laino, T.; Reymond, J.-L. Mapping the Space of Chemical Reactions Using Attention-Based Neural Networks. *Nat. Mach Intell* **2021**, *3* (2), 144−152.
(290) DeGrave, A. J.; Janizek, J. D.; Lee, S.-I. AI for Radiographic COVID-19 Detection Selects Shortcuts over Signal. *Nat. Mach Intell* **2021**, *3* (7), 610−619.
(291) Varela-Rial, A.; Maryanow, I.; Majewski, M.; Doerr, S.; Schapin, N.; Jiménez-Luna, J.; De Fabritiis, G. PlayMolecule Glimpse: Understanding Protein-Ligand Property Predictions with Interpretable Neural Networks. *J. Chem. Inf. Model.* **2022**, *62* (2), 225−231.
(292) Woelfle, M.; Olliaro, P.; Todd, M. H. Open Science Is a Research Accelerator. *Nature Chem.* **2011**, *3* (10), 745−748.
(293) McKiernan, E. C.; Bourne, P. E.; Brown, C. T.; Buck, S.; Kenall, A.; Lin, J.; McDougall, D.; Nosek, B. A.; Ram, K.; Soderberg, C. K.; Spies, J. R.; Thaney, K.; Updegrove, A.; Woo, K. H.; Yarkoni, T. How Open Science Helps Researchers Succeed. *eLife* **2016**, *5*, No. e16800.

https://doi.org/10.1021/acspolymersau.3c00025
ACS Polym. Au 2023, 3, 406−427