

# Data-Driven Design of Protein-Like Single-Chain Polymer Nanoparticles

*Rahul Upadhyaya,<sup>†</sup> Matthew J. Tamasi,<sup>†</sup> Elena Di Mare, N. Sanjeeva Murthy, Adam J. Gormley\**

Department of Biomedical Engineering, Rutgers, The State University of New Jersey,  
Piscataway, NJ 08854, USA

**KEYWORDS:** small-angle X-ray scattering, SAXS, dynamic light scattering, DLS, machine learning, automation, high throughput

## **ABSTRACT**

The functional structure of proteins is heavily influenced by their folding behavior. AlphaFold, a powerful artificial intelligence (AI) program trained on information from the Protein Data Bank (PDB), was developed to predict the 3D structure of proteins from its amino acid sequence. Inspired by this, we aim to elucidate structural features of synthetic single-chain polymer nanoparticles (SCNPs) based on compositional information (monomers, chain length, molecular weight, charge, and valency) by machine learning (ML). Specifically, we demonstrate the effectiveness of ML to improve the efficiency of SCNP design and uncover important polymer design attributes to mimic protein-like structural features. To start, we randomly screened over

1000 synthesized SCNPs through a combination of high-throughput dynamic light scattering (DLS) and small-angle X-ray scattering (SAXS) and compared these results to simulated protein data from the PDB. Then, utilizing evidential neural networks (ENets), we predicted, synthesized, and characterized 30 novel compact SCNPs. Incredibly, this data-driven approach yielded 58% of the predicted SCNPs with Porod exponent  $\geq 3.5$  as opposed to 5% of SCNPs from the random screen. Using Shapely additive explanation (SHAP) values, we further uncovered interesting contributions of monomer content on Porod exponent and radius of gyration. From this work, we have shown that an ML-guided approach proves effective for the challenging, unintuitive problem of nanoparticle design.

## INTRODUCTION

Proteins are remarkably complex macromolecules with specific functionality. The ultimate principle in protein science is that amino acid sequence impacts the structure which in turn enables function.<sup>1-5</sup> Protein structure and folding requires a combination of hydrophobic interactions, electrostatics, and hydrogen bonding.<sup>2, 4, 6, 7</sup> A key requirement is dehydration whereby hydrophobic amino acids form a hydrophobic core surrounded by solvent-accessible amino acids. Additional intra- and intermolecular interactions permit even greater structural control.<sup>6-8</sup> Further, concentration effects exist since below a characteristic concentration, the entropic cost of chain collapse and chain-chain repulsion is greater than the benefit of avoiding contact with water.<sup>3, 6</sup> This extensive theoretical knowledge about protein folding along with significant information in the Protein Data Bank (PDB) paved the path for AlphaFold, an artificial intelligence (AI) platform that provides protein structural predictions given only the corresponding amino acid sequence.<sup>9</sup>

While this landmark advancement has significantly closed the gap in our ability to predict how protein sequence organization impacts its ultimate morphology, these advancements beg the question – is it possible to bridge the gap between design features and structural attributes for synthetic macromolecules which lack the precise organization of globular proteins?

A class of synthetic macromolecules that mimic proteins are referred to as single-chain polymer nanoparticles (SCNPs) which can collapse into higher ordered structures.<sup>10-18</sup> These nanoparticles have potential as enzyme<sup>19, 20</sup> or protein mimetics,<sup>16, 21, 22</sup> sensing nanomaterials,<sup>23</sup> and drug delivery systems.<sup>24-26</sup> Of particular interest is the use of SCNPs as multivalent therapeutics because of their biophysical tunability and therefore ligand presentation.<sup>27, 28</sup> Multivalent strategies provide avidity or accumulated binding strength from presentation of multiple ligands and can be explained mathematically using statistical thermodynamics. Greater valency theoretically increases permutation entropy, or number of possible ligand-receptor interaction pairs, and thus results in a more negative or favorable binding free energy. This effect tends to be more pronounced for weaker binding ligands (*e.g.* peptides) which have a faster off-rate and allow for more potential interactions.<sup>29</sup> Providing some optimal amount of polymer scaffold flexibility can impart sufficient conformational freedom for ligands to position towards receptors and be in position to interact and bind. Recently, Chapman and co-workers utilized a similar strategy with end-functionalized star polymer-peptide conjugates to mimic the chemotherapeutic protein TNF-related apoptosis-inducing ligand (TRAIL).<sup>22</sup> Multivalency has also been implemented to improve binding affinity and biological activity of erythropoietin (EPO) mimetic peptides, illustrating the appeal of this approach.<sup>30-34</sup>

To effectively mimic protein therapeutics, there is a requirement to further optimize SCNP design features that affect downstream signaling such as compactness, flexibility, and valency.<sup>35-</sup>

<sup>37</sup> While complexity in the form of orthogonal chemistry may be required,<sup>14, 38</sup> a polymer coil-to-globule transition is largely driven by hydrophobic collapse to form a core that is shielded from aqueous solvent. Hydrogen bonding,  $\pi$ - $\pi$  stacking, and electrostatics can both stabilize SCNPs and enhance solubility.<sup>6</sup> There is an obvious relationship between polymer chain flexibility and compactness: flexible polymers undergo intra-chain cross-linking and collapse and interacting segments tend to increase in stiffness with reduced conformational degrees of freedom.<sup>29</sup>

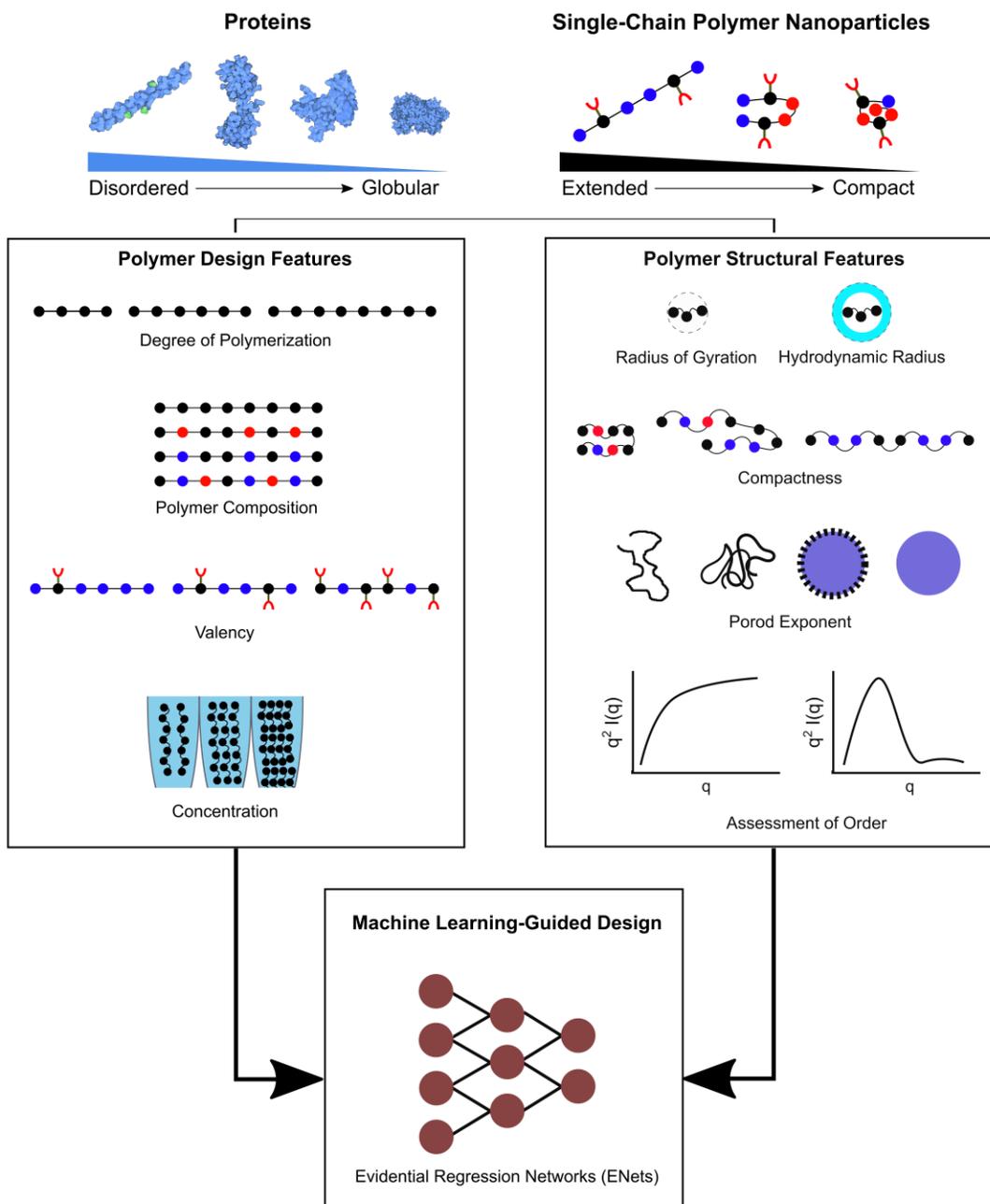
Despite our fundamental understanding of polymer physics, there is no established protocol to design polymers with a desired conformation because of the inability to experimentally probe the entire chemical design space. Traditional polymer chemistry techniques such as reversible addition-fragmentation chain transfer (RAFT) polymerization and atom transfer radical polymerization (ATRP) must be conducted in inert conditions.<sup>39, 40</sup> Introduced within the last decade were air-tolerant photoinduced electron/energy transfer-RAFT (PET-RAFT),<sup>41-45</sup> enzyme-assisted ATRP,<sup>46</sup> enzyme-assisted RAFT (Enz-RAFT),<sup>47, 48</sup> and ring-opening polymerization (ROP) through water<sup>49</sup> which have enabled polymerizations to be carried out directly in microplate format. To further the ability to complete combinatorial chemistry, our group previously demonstrated the potential of a plate-based gel filtration purification,<sup>50</sup> high-throughput structural characterization of SCNPs,<sup>51</sup> and the adaptation of liquid handling robotics to complete polymer synthesis and post-polymerization functionalization.<sup>52, 53</sup> Others such as Gibson and co-workers have also utilized liquid handling robotics for preparing and testing polymer libraries.<sup>54-56</sup>

However, utilizing only high-throughput approaches to synthesize and characterize SCNPs runs the risk of inefficiency, false positives and negatives, and difficulty with interpreting large datasets.<sup>57</sup> In previous work, we conducted high-throughput screening of a large combinatorial library (> 450 copolymers) via dynamic light scattering (DLS) and small-angle X-ray scattering

(SAXS) to identify compact nanostructures.<sup>51</sup> Due to polymer hydrophobicity and having a relatively small dataset, we were unable to define a model to quantitatively inform further design, thus serving as a motivation for this study. For such complex and non-intuitive material design problems,<sup>58-60</sup> machine learning (ML) is emerging as an exciting approach for informing design. ML has seen success in small-molecule organic synthesis<sup>61-63</sup> and is being introduced for synthetic polymers<sup>64, 65</sup> to accelerate discovery of new materials with desired structural or functional properties.<sup>63, 64, 66</sup> In particular, ML can be exploited to classify large datasets, uncover hidden patterns, and reduce the number of variables for a multivariate dataset.<sup>67</sup> By accomplishing these tasks, a data-driven approach can make predictions about material features of interest and further our understanding of how these features can be designed in future experiments.<sup>68-70</sup>

In this study, we combined high-throughput polymer synthesis and characterization with ML to aid the design of novel SCNPs that are compact and exhibit similar flexibility to ordered proteins. We hypothesized that an ML-guided approach can be leveraged in an SCNPs synthetic workflow to identify difficult-to-design compact and rigid nanoparticles. First, we synthesized over 1000 amphiphilic random heteropolymer backbones containing varied composition (neutral, cationic, and hydrophobic monomers), degree of polymerization (DP) (200, 300, and 400), functionalization (2 kDa PEG, TRAIL mimetic peptide, or EPO mimetic peptide), and valency (2, 5, and 10 mol%). Synthesized copolymers were characterized by SAXS and DLS to provide compactness and flexibility information that was then compared against a parallel dataset of over 300 protein structures extracted from the PDB with analogous theoretical compactness and flexibility. The dataset of over 1000 synthesized copolymers were leveraged to train evidential neural networks (ENets) to provide novel predictions for compact SCNPs.<sup>71, 72</sup> This approach successfully identified novel SCNPs structures that demonstrated remarkable compactness. Further, we

performed post hoc analysis of our data to understand how copolymer features influence SCNP compactness and flexibility (**Figure 1**).



**Figure 1. Schematic of ML-guided approach for designing SCNPs.** Proteins undergo a complex folding process and possess a high level of structural order. Using synthetic materials, a ML approach is utilized to model important design (DP, polymer composition, valency, and concentration) and structural features ( $R_h$ ,  $R_g$ , compactness as  $R_g/R_h$ , and Porod exponent) to provide predictions for synthesizing novel copolymers with desired properties.

## RESULTS AND DISCUSSION

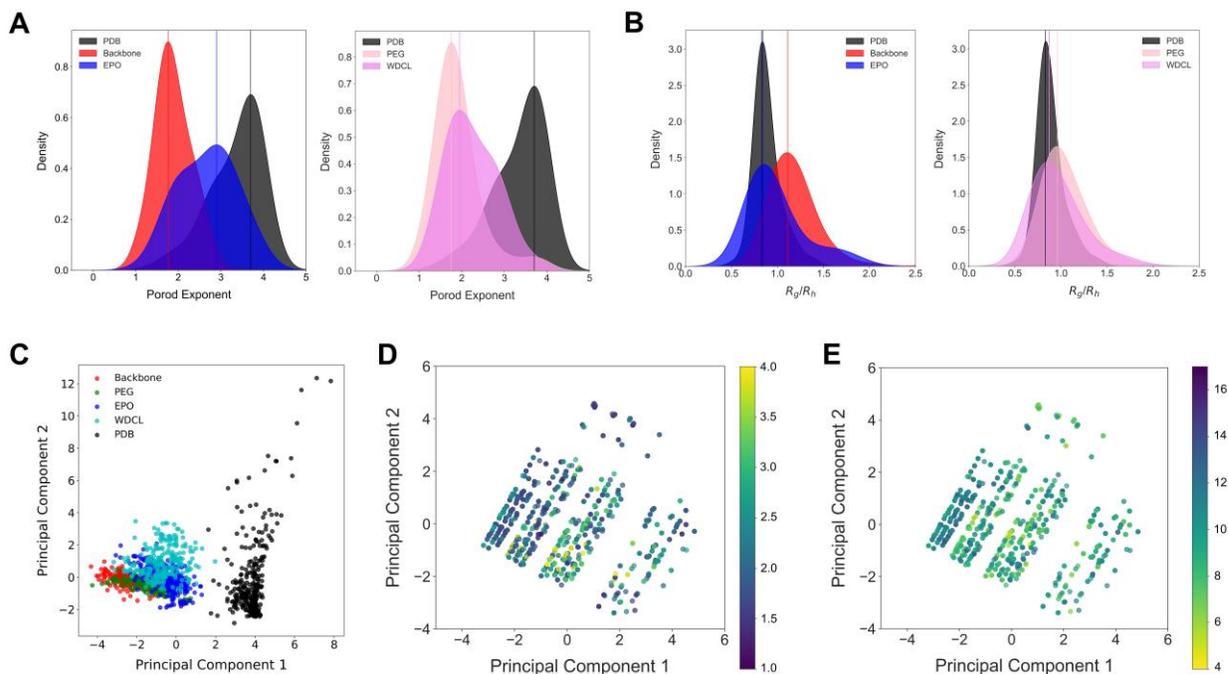
### *Random Screen of Copolymer Backbones and Polymer Conjugates*

We began by investigating our initial library of copolymer backbones, copolymer-PEG conjugates, and copolymer-peptide conjugates that we will refer to as a random screen. To explore a broad compositional space, we employed *N,N*-dimethylacrylamide (DMA) as a neutral monomer, [2-(methacryloyloxy)ethyl]trimethylammonium chloride (TMAEMA) as a cationic monomer, and methyl methacrylate (MMA) and butyl methacrylate (BMA) as hydrophobic monomers. All representative monomer, copolymer, and copolymer conjugate structures are provided (**Figures S1-S4**). Additionally, we varied DP (200, 300, and 400) and valency of SPAAC-functionalized PEG or peptide (2, 5, and 10 mol%). For polymer-peptide conjugates, we focused on two peptides that have been previously validated and characterized for binding with an objective of enabling nanoparticle design.<sup>22, 30-34, 73</sup> All characterization data of PDB simulated proteins (simulation details provided in Materials and Methods), copolymer backbones, PEG-SCNPs, EPO-SCNPs, and WDCL-SCNPs are provided in **Tables S3-S7**.

After synthesis and characterization of copolymer backbones and copolymers conjugated with PEG, EPO, and WDCL, various approaches were taken to visualize attributes of the large dataset. Compactness ( $R_g/R_h$ ) and flexibility (Porod exponent) are two parameters of interest in the realm of SCNP design. **Figure 2** illustrates kernel density estimation (KDE) and principal component analysis (PCA) that were completed for the dataset to observe differences in flexibility between subgroups. Compared to copolymer backbones, the conjugation of the EPO peptide on average seemed to shift behavior towards that of proteins. These EPO-SCNPs tended towards increased Porod exponent compared to that of copolymer backbones, showing greater similarity to the behavior of the PDB proteins. The other subgroups of synthesized SCNPs (PEG and WDCL)

displayed more of a disparity with PDB proteins compared to EPO-SCNPs (**Figure 2A**).  $R_g/R_h$  seems to have greater complexity as the two groups of polymer-peptide conjugates, the EPO-SCNPs and WDCL-SCNPs, have greater overlap with compactness of PDB proteins than copolymer backbones and PEG-SCNPs (**Figure 2B**). This is also reflected in the shift in median  $R_g/R_h$  and Porod exponent, respectively, between the groups: Backbone (1.13, 1.8), PEG-SCNPs (0.98, 1.8), WDCL SCNPs (0.91, 2.1), EPO-SCNPs (0.88, 2.8), and PDB proteins (0.85, 3.6). Individual physical characterization data including values obtained from SAXS and DLS along with feature information (molecular weight, charge, and hydrophobicity) are also provided (**Tables S3-S7**).

While KDE plots in **Figure 2A-C** quantitatively show the influence of composition on structure, 2D PCA in **Figure 2D-F** seems to reflect that there are three distinct groups within the dataset (**Figure 2C**) when we featurize the data using all known values ( $R_g$ ,  $R_h$ ,  $R_g/R_h$ , molecular weight, total charge, total positive charge, total negative charge,  $\log P$ , and HLB) (representation of explained variance is provided in **Figure S12**). We also observe that when the data is featurized by only polymer design features (DP, mol% DMA, mol% TMAEMA, mol% MMA, mol% BMA, mol% NHS methacrylic acid, mol% DBCO, mol% PEG, mol% EPO, mol% WDCL, and molecular weight) and colored by Porod exponent and  $R_g$  (**Figure 2D,E**), there does not seem to be a clear gradient or dependency. These illustrations reveal that SCNP structural design contains non-linear complexities that cannot be explained by simple feature correlations alone. Additional two-dimensional KDE plots are provided for only chemical descriptors and experimentally derived features (**Figures S13 and S14**).



**Figure 2. KDE and PCA of copolymers, copolymer conjugates, and proteins.** KDE and PCA were used to determine differences between the subgroups: proteins simulated from the PDB, copolymer backbones (Backbone), PEG-SCNPs (PEG), EPO-SCNPs (EPO), and WDCL-SCNPs (WDCL). (A) The probability densities associated with Porod exponents of PDB proteins, copolymer backbones, and EPO-SCNPs along with PDB proteins, PEG-SCNPs, and WDCL-SCNPs are shown. (B) The probability densities associated with  $R_g/R_h$  of PDB proteins, copolymer backbones, and EPO-SCNPs along with PDB proteins, PEG-SCNPs, and WDCL-SCNPs are provided. (C) PCA was completed for all groups using all features ( $R_g$ ,  $R_h$ , Porod exponent, Porod volume,  $R_g/R_h$ , molecular weight, total charge, total positive charge, total negative charge,  $\log P$ , and HLB) with explained variance of 0.608. PCA was also done using only polymer design features (DP, % DMA, % TMAEMA, % MMA, % BMA, % NHS methacrylic acid, % DBCO, % PEG, % EPO, % WDCL, and molecular weight), colored by (D) Porod exponent and (E)  $R_g$ .

Comparisons of the random experimental screen of copolymer backbones and copolymer conjugates with the PDB protein dataset is important, however, some limitations should be noted. While our PDB dataset would ideally contain a large percentage of both disordered and globular proteins, intrinsically disordered proteins (IDPs) that lack secondary or tertiary structure only comprise about 30% of the eukaryotic proteome.<sup>74</sup> Also a large fraction of the structural information in the PDB are for globular proteins with almost 90% of structures from crystallography while the other 10% are from solution NMR and electron microscopy.<sup>75</sup> Although we attempted to diversify our PDB dataset, it still included approximately 64% of structures from X-ray crystallography. CRY SOL is a powerful program in the ATSAS package that enables computation of theoretical SAXS curves given a PDB structure file as an input.<sup>76, 77</sup> While other groups have also implemented this program to obtain information about IDPs,<sup>74, 78-80</sup> CRY SOL is limited by size or  $D_{\max}$  due to its use of spherical harmonics which can make the program challenging for generating SAXS profiles of extended proteins.<sup>80</sup> Especially for IDPs, the maximum number of harmonics (50) should be selected, but there remains the possibility to provide experimental scattering data to CRY SOL to improve fitting.

Additionally, KDE allows us to visualize the chemical space of our dataset in terms of total charge, HLB, and  $\log P$  compared to the simulated PDB proteins (**Figures S18-S20**). While copolymer backbones are somewhat similar in total charge to a large portion of the PDB protein population, there are differences between the two subgroups in HLB and  $\log P$ . EPO-SCNPs and WDCL-SCNPs appear to shift copolymer HLB closer to that of proteins, and these plots represent that a combination of descriptors are likely required for the complex task of SCNP design. For analysis and interpretation of SAXS and DLS data, we focused on two parameters of interest in the realm of therapeutic receptor binding: compactness ( $R_g/R_h$ ) and flexibility (Porod exponent).

From the KDE plots, there seems to be three distinct groups based on distributions of PDB proteins (narrow compact range with Porod exponent  $> 3.0$ ), copolymer backbones and PEG-SCNPs (with extended conformation and Porod exponent  $\leq 2.0$ ), and EPO-SCNPs and WDCL-SCNPs (favor compactness with a broad range of flexibility and intermediate Porod exponent from 2.0-3.0). The data also indicates that peptidic side chains seem to shift copolymer behavior closer to that of proteins in terms of both  $R_g/R_h$  and Porod exponent (**Figure 2A,B**).

We supplemented the PCA analysis displayed in **Figure 2** by separating each subgroup and labeling by features of interest (Porod exponent or  $R_g$ ). In contrast to PDB proteins which displayed greater Porod exponent throughout, the copolymer backbones and conjugates contained distinct pockets of lower flexibility regions indicative of higher order that even differed between each synthetic subgroup (**Figure S15**). This implies that it would be nearly impossible to logically formulate a *de novo* framework for designing similarly structured compact copolymers. Conversely,  $R_g$  seems to have relatively more consistency between these regions of compact structures (**Figure S16**). Simply focusing the PCA on SCNPs allowed us to expand our feature set to include more polymer specific characteristics, resulting in greater spread within the EPO and WDCL groups and demonstrating the utility of including relevant polymer descriptors in data visualization and interpretation (**Figure S17**). While PCA and KDE are powerful tools for dimensionality reduction or drawing conclusions about the overall dataset, they cannot provide us feature importance or predictive capability which is why we must turn to other methods.<sup>81</sup>

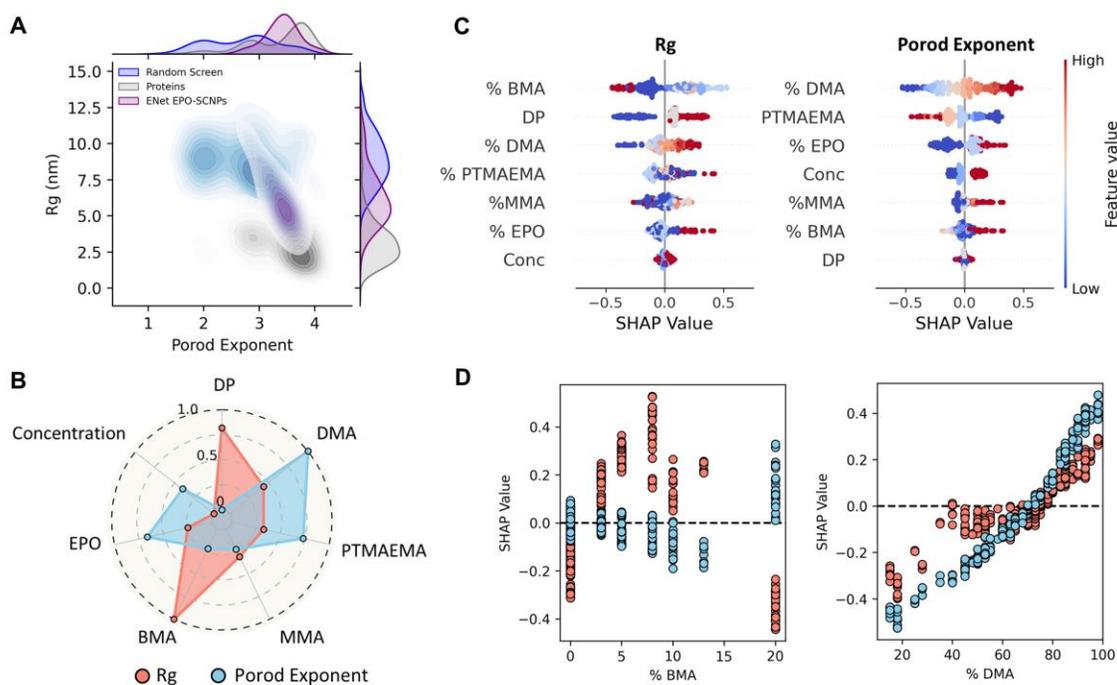
### ***ML Investigation of SCNPs***

While many copolymer conjugates were investigated in our initial study, only a small fraction of SCNPs demonstrated significant compactness and flexibility. This is not surprising as designing

self-folding behavior in synthetic macromolecules remains challenging. To facilitate further control over these higher order behaviors, we utilized ML techniques to identify samples that demonstrate compactness and protein-like flexibility. To do so, we began by training multiple ENets to make predictions on SCNP parameters of interest ( $R_g$ ,  $R_h$ ,  $R_g/R_h$ , Porod exponent) directly from chemical representations of polymer conjugate chemistry and chosen molecular descriptors (see Materials and Methods). Predictions of  $R_h$  and  $R_g/R_h$  demonstrated low accuracy on held-out validation data during training (**Figure S5**) and prediction accuracy of  $R_g$  and Porod exponent was significantly greater, leading us to utilize these models in our future design. The tuning of the ENet regularization parameter is illustrated in **Figure S6**. To generate compactness and flexibility predictions of copolymer conjugates similar to our synthesized samples, respective ENet models were used to predict expected mean and variance of  $R_g$  and Porod exponent for  $\sim 700,000$  copolymer conjugate chemistries *in silico*. Comparable to our observations of compactness and flexibility in our measured data of SCNPs, predictions of PEG, WDCL, and EPO suggested EPO-SCNPs generally exhibited high Porod exponent and PEG-SCNPs were anticipated to demonstrate the lowest  $R_g$  (**Figure S7**).

As we were particularly interested in designing compact samples with protein-like features, 30 EPO-SCNPs were proposed for synthesis and characterization with low predicted  $R_g$  and high predicted Porod exponent (see Materials and Methods). Since many of these samples demonstrated limited solubility, likely due to large hydrophobic monomer fractions (**Table S8**), DLS and SAXS characterization were completed on 12 of 30 samples. Remarkably, **Figure 3A** demonstrates that ML-informed design appropriately identified numerous EPO-SCNPs exhibiting more similar physical measurements to the PDB sampled proteins than prior EPO-SCNPs measured during our random screen. While only 5% of SCNPs in the random screen demonstrated Porod exponent  $\geq$

3.5, ~58% of EPO-SCNPs proposed by the ENets demonstrated Porod exponents above this threshold. Further, these differences are readily observed in the distinction in median  $R_g$  and Porod exponent between initial EPO-SCNPs (8.4 nm and 2.7) and ENet proposed EPO-SCNPs (5.7 nm and 3.5). As  $R_g$  and Porod exponent shifted towards more protein-like values,  $R_h$  remained large for EPO-SCNPs (**Figure S9**). The  $R_g/R_h$  of an ideal sphere is 0.775, however, some of the characterized SCNPs exhibited  $R_g/R_h < 0.775$ . This phenomenon was even more prevalent in the ML predicted EPO-SCNPs. A few potential reasons for this include adoption of a core-shell structure or molten globule with tight packing of the hydrophobic alkyl side chain surrounded by a more dispersed shell.<sup>82, 83</sup>



**Figure 3. ML guided design of compact SCNPs.** (A) Joint distribution plot comparing  $R_g$  and Porod exponent of random screen EPO-SCNPs, ML predicted EPO-SCNPs, and PDB proteins. (B) Normalized mean absolute Shapely additive explanation (SHAP) values calculated for Porod exponent and  $R_g$ . (C) Summary of SHAP values calculated for all SCNP features for Porod exponent and  $R_g$ , respectively. (D) Porod exponent and  $R_g$  SHAP values for BMA and DMA monomer content.

Because ENet models were generally successful in predicting more compact EPO-SCNPs, we sought to understand the underlying chemical designs that led to compactness within our study. For this, we computed Shapely additive explanations (SHAP) values to quantify the impact of model features (monomers, peptide, DP) on prediction of both  $R_g$  and Porod exponent.<sup>84, 85</sup> In these interpretations, positive SHAP values indicate a positive contribution to predicted  $R_g$  or Porod exponent, and negative values indicate a negative contribution. Further, mean absolute SHAP values for individual features are used as a metric for overall feature importance. **Figure 3B** demonstrates that different EPO-SCNP features play primary roles in the predictions of  $R_g$  or Porod exponent. For example, while BMA is measured to be the most impactful parameter for predicting  $R_g$ , DMA demonstrates the largest absolute SHAP value for predictions of Porod exponent. To further investigate these differences, we compare all SHAP values for predictions of  $R_g$  and Porod exponent (**Figure 3C**). As our objective in this study was to identify highly compact SCNPs, we anticipated that EPO-SCNPs hydrophobicity would be important to induce single chain folding through induction of a hydrophobic core. Interestingly, BMA, a highly hydrophobic monomer, not only demonstrates the largest feature importance for  $R_g$  but also demonstrates a large negative SHAP for  $R_g$  at high levels of incorporation, in line with our design objective of minimizing  $R_g$  (**Figure 3D**). Additionally, despite not being as impactful to the ENet model for Porod exponent, BMA demonstrates a positive SHAP contribution to Porod exponent at high monomer incorporation. This suggests that high incorporation of BMA is ultimately synergistic to both design objectives. In contrast to this synergistic behavior for BMA, DMA, the chemical feature with highest absolute SHAP value for predicted Porod exponent, demonstrated competing behavior in predicting high Porod exponent and small  $R_g$ . Highly correlated SHAP values for DMA between Porod exponent and  $R_g$  demonstrate that high DMA incorporation leads to positive SHAP

values for Porod exponent (in line with our objective), yet high DMA also leads to positive SHAP values for  $R_g$ , a parameter we aimed to minimize.

### ***Structural Characterization of Predicted SCNPs***

Once ML guided design was completed, individual copolymer backbones and copolymer conjugates were further inspected for interesting structural characteristics. **Table 1** contains a summary of BSA, EPO-SCNPs from the random screen, and a ML predicted EPO-SCNP with design features (DP, valence, molecular weight, and HLB). Other quantities included were those obtained by a combination of SAXS and DLS ( $R_g$ ,  $R_h$ ,  $R_g/R_h$ , and Porod exponent). BSA, widely used as a model compact and ordered protein, is presented alongside other copolymer conjugates that displayed evidence of compactness from the random screen (EPO119, EPO163, EPO166, and EPO172). We have also included the ML predicted compact EPO-SCNP (ML EPO24). Conversely, the extended and water-soluble polymer PEG 35 kDa is presented alongside other EPO SCNPs that exhibited extendedness (EPO18, EPO59, EPO229, and EPO308). In terms of the copolymer design features provided in **Table 1** (monomer composition, DP, valence, molecular weight, and HLB), there is no clear distinction in the physicochemical characteristics between these two groups, although extended copolymers and copolymer conjugates tended to exhibit greater hydrophilicity with a higher weighted HLB. This is further evidence that the problem of SCNP design does not contain a simple solution with a single parameter. We can, however, visualize a clear gap between these two groups of samples in terms of measured  $R_g/R_h$  and Porod exponent as compact macromolecules tended towards  $R_g/R_h \sim 0.775$  and Porod exponent from 3.0-4.0, while extended ones tended towards  $R_g/R_h > 1.00$  and Porod exponent  $< 2.0$ . The gap suggests that incorporation of experimental data may be useful in defining a framework for SCNP structural

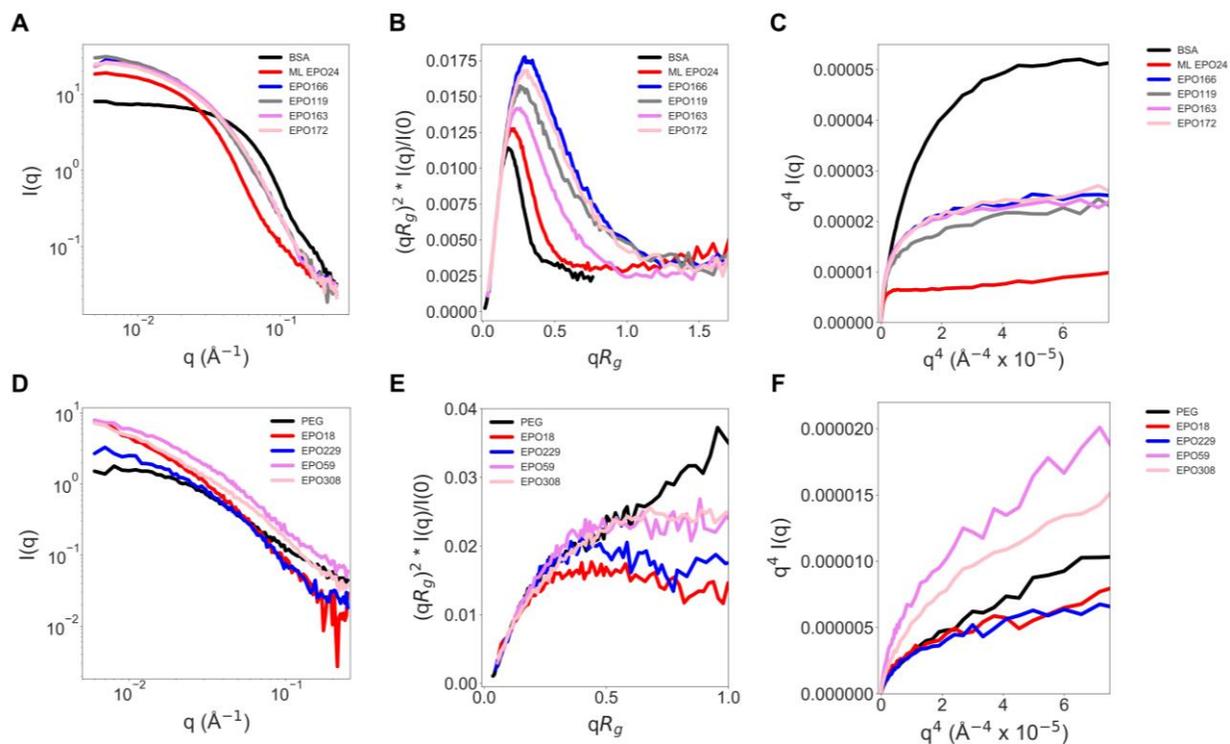
design. Also, SEC-MALS was completed to quantify molecular weight and  $D$  for copolymer backbones (**Table S1** and **Figure S10**) while UV-vis spectroscopy was conducted to determine DBCO incorporation which has implications for valency (**Table S2**).

**Table 1. Physical characterization data of highlighted random screen and ML predicted EPO-SCNPs along with BSA and PEG.**

ID	Polymer Composition	DP	Valence	MW <sup>theoretical</sup> (kDa)	HLB	$R_g$ (nm)	$R_h$ (nm)	$R_g/R_h$	Porod Exponent
BSA	--	--	--	66.5	--	3.07	4.00	0.775	3.9
ML EPO24	DMA 47.5% / TMAEMA 12.5% / BMA 35%	275	2.5%	53.5	--	7.31	12.10	0.604	4.0
EPO119	DMA 85% / TMAEMA 5% / MMA 5%	200	5%	48.6	14.0	8.38	10.90	0.769	3.2
EPO163	DMA 88% / MMA 10%	200	2%	33.1	12.7	7.81	9.70	0.805	3.6
EPO166	DMA 93% / BMA 5%	200	2%	31.3	10.5	8.16	10.39	0.785	3.7
EPO172	DMA 93% / MMA 5%	300	2%	46.4	11.9	9.17	11.76	0.780	3.8
PEG	PEG 100%	794	0%	35.0	10.9	5.10	4.50	1.13	1.5
EPO18	DMA 45% / TMAEMA 25% / MMA 17% / BMA 3%	300	10%	121.4	17.4	11.35	6.54	1.74	2.2
EPO59	DMA 45% / TMAEMA 25% / MMA 12% / BMA 13%	400	5%	108.2	14.6	9.37	7.26	1.29	1.5
EPO229	DMA 73% / TMAEMA 15% / BMA 10%	200	2%	35.0	12.6	6.75	3.60	1.88	2.3
EPO308	DMA 50% / TMAEMA 25% / BMA 20%	400	5%	109.3	14.6	9.89	9.40	1.05	2.1

Further, SAXS data are illustrated for the same compact and extended EPO-SCNPs from the random screen and ML-guided design alongside controls BSA and PEG 35 kDa (**Figure 4**). This includes the SAXS intensity profiles, normalized Kratky plot, and Porod plot of the compact copolymer conjugates (ML EPO24, EPO119, EPO163, EPO166, and EPO172) and BSA (**Figure 4A-C**). The same was included for extended copolymer conjugates (EPO18, EPO59, EPO229, and EPO308) and PEG 35 kDa (**Figure 4D-F**). There was reasonable overlap in SAXS intensity profiles of the synthesized copolymers and copolymer conjugates with the controls. Further, characteristic behavior of macromolecular compactness and flexibility were exhibited through respective intensity, Kratky, and normalized plots (**Figures S21-S23**). Compact and extended SCNPs displayed SAXS features similar to highly ordered protein BSA and the extended polymer PEG, respectively. While the SAXS intensity profiles show agreement between synthesized SCNPs and their respective compact and extended controls, normalized Kratky plots qualitatively

confirm that the degree of order is also in reasonable agreement. Interestingly, the ML EPO24 SCNP exhibits order more similar to that of BSA when compared against other SCNPs from the random screen (**Figure 4B**). These normalized Kratky plots can be useful in the case of combinatorial libraries that may contain various measured concentrations and thus differences in scattering intensity at zero angle, referred to as  $I(0)$ . Therefore, deviations from BSA in terms of peak broadness and intensity in the normalized Kratky plot reflect on the varying degrees of order. This work also illustrates that we can synthesize a wide range of SCNPs using ML-predicted compositions to obtain the desired compactness and flexibility utilizing PET-RAFT and click chemistry (**Figures 2 and 4**).



**Figure 4. SAXS summary of compact and extended EPO-SCNPs from the random screen and ML prediction.** (A) SAXS intensity plot of compact protein BSA, ML predicted compact EPO-SCNP (ML EPO24), and randomly screened compact EPO-SCNPs (EPO166, EPO119, EPO163, and EPO172). Compact behavior is revealed with characteristic (B) normalized Kratky and (C) Porod plots. (D) SAXS intensity plot of extended PEG 35 kDa and randomly screened extended EPO-SCNPs (EPO18, EPO229, EPO59, and EPO308). Extendedness is illustrated through characteristic (E) normalized Kratky and (F) Porod plots.

## CONCLUSION

While AlphaFold supplies protein structural information from specific amino acid sequences, a design-structure gap exists for synthetic macromolecules. As high-throughput polymer chemistry provides new opportunity for data-driven design, we aimed to bridge the gap between synthetic SCNP design features and structural characteristics (compactness and flexibility) where sequence-level control is not yet possible. It is clear from this large dataset of copolymers and copolymer conjugates that the synthesis and design of SCNPs is a multi-objective problem with high complexity as there are design features and measured characteristics that are interrelated.<sup>86</sup> SCNPs contain nearly an infinite chemical space with the potential to select features such as DP, monomer composition, architecture, and valency, but the question remains of how to select an initial polymer library. In this work, we intuitively selected copolymers based on theoretical  $\log P$  and attempted to prescreen these combinations using DLS. Employing our own logic can result in inefficiencies and may require a more iterative process which can be difficult to implement experimentally especially with a technique such as SAXS.<sup>86, 87</sup> Another challenge with complex copolymer conjugates involves deconvoluting physical contributions between polymer and peptide which would require other techniques such as NMR or small-angle neutron scattering (SANS).<sup>88</sup> In addition to all of these factors, it is likely difficult to synthesize structures with the same level of order and folding as proteins mainly because of the lack of sequence-level control.<sup>6, 18, 67, 87, 89, 90</sup> Until advances are made to achieve this level of control with synthetic ease, polymer chemists may need to rely on quantifying reactivity ratios,<sup>91-93</sup> illustrating compositional drift,<sup>94</sup> and potentially designing for molecular weight distribution.<sup>95</sup> Others are even making efforts to increase complexity with polymer design,<sup>25, 96</sup> biologically inspired monomers,<sup>97, 98</sup> and multi-orthogonal

approaches.<sup>53, 99</sup> Combining informative approaches such as ML, simulations, and structural characterization can only strengthen the SCNP design paradigm.

In summary, our objective was to demonstrate the utility of a data-driven design approach for identifying compact SCNPs that are challenging to design *de novo*. We first completed a random screen of over 1000 SCNP random heteropolymers and copolymer conjugates with varied DP, composition, valence, and functionality. In this random screen, we characterized SCNPs by DLS and SAXS to quantify macromolecular compactness and flexibility. Data visualization approaches (KDE and PCA) demonstrated that this is a complex multi-objective problem that cannot be understood by simple chemical design principles alone and seemed to indicate that polymer-peptide conjugates served as an intermediate between disordered copolymers and folded proteins. We proceeded to employ ENets to predict compact EPO-SCNPs with  $R_g < 7$  nm and Porod exponent from 3.7-4.0 and randomly sampled 30 EPO-SCNPs which were synthesized and characterized to validate the model. While the initial random screen only resulted in 5% of SCNPs with Porod exponent  $\geq 3.5$ , about 58% of the ENets predicted SCNPs exhibited a Porod exponent  $\geq 3.5$ . Not only did this ML-guided approach improve SCNP design efficiency, SHAP analysis uncovered interesting design features such as the importance of DMA and BMA monomer content on Porod exponent and  $R_g$ . In conclusion, our work has illustrated that data-driven design provides a much-needed solution to the often-unintuitive problem of nanoparticle design and provides an potential framework for understanding synthetic macromolecule design-structure relationships.

## MATERIALS AND METHODS

### *Materials*

The monomers DMA (Sigma-Aldrich), MMA (Alfa Aesar), BMA (Alfa Aesar), and TMAEMA (Fisher Scientific) were sourced from a variety of vendors. The chain transfer agent 4-cyano-4[(dodecylsulfanylthiocarbonyl)sulfanyl] pentanoic acid and initiator zinc tetraphenylporphyrin (ZnTPP) were purchased from Sigma-Aldrich and Fisher Scientific, respectively. Dibenzocyclooctyne-amine (DBCO-NH<sub>2</sub>) which was used for click chemistry was purchased from Click Chemistry Tools while 4-dimethylaminopyridine (DMAP) was purchased from Sigma-Aldrich. Traditional Fmoc amino acids (CEM Corporation), Fmoc-L-lysine-N<sub>3</sub> (Lys-N<sub>3</sub>) (Chem-Impex International), and the spacer Fmoc-*N*-amido-PEG<sub>2</sub>-acid (PEG<sub>2</sub>) (BroadPharm) were utilized for solid-phase peptide synthesis. Methoxypolyethylene glycol azide with M<sub>n</sub> = 2000 Da (PEG-N<sub>3</sub>) (Sigma-Aldrich) was the other molecule conjugated onto the synthesized copolymers.

### *Automated PET-RAFT*

Automated PET-RAFT was completed using oxygen-tolerant PET-RAFT chemistry<sup>41, 44</sup> coupled with adaptation of the Hamilton Microlab STARlet liquid handling robot to complete this chemistry.<sup>52</sup> The DMA, MMA, BMA, and TMAEMA monomers were prepared at a 2 M concentration in DMSO and deinhibited by passing solutions through columns containing inhibitor remover beads (Sigma-Aldrich), glass wool (Fisher Scientific), and aluminum oxide (Sigma-Aldrich). Meanwhile, the methacrylic acid NHS monomer was prepared in anhydrous DMSO at 0.4 M with 1 mol equivalent acetic acid/NHS. The degree of polymerization (DP) of the polymer was varied between 200, 300, and 400 by shifting the final monomer:CTA molar ratios (200:1,

300:1, and 400:1) and fixing the final CTA:ZnTPP molar ratio (50:1). All reagents were loaded into 1.5 mL centrifuge tubes, dispensed by the liquid handling robot into a 96-well polypropylene plate, and photoinitiated by a 560 nm LED light for 16 h. Representative monomer, copolymer, and copolymer conjugate structures were also included (**Figures S1-S4**).

### *Solid-Phase Peptide Synthesis*

The WDCL (WDCLDNRIGRRQCV- $\{\text{Lys-N3}\}$ -L-amide) (Cyc 3, 13) (99% purity) and EPO (GGTYSCHFGPLTWVCKPQ- $\{\text{PEG2}\}$ -SS- $\{\text{Lys-N3}\}$ -amide) (Cyc 6, 15) (99% purity) peptides were synthesized by solid-phase peptide synthesis. HPLC chromatograms demonstrating peptide purity are provided (**Figure S11**). This was completed using a Liberty Blue automated microwave peptide synthesizer (CEM Corporation) on Rink Amide ProTide Resin with 0.16 mmol/g loading (CEM Corporation). 1 M Oxyma and 1 M DIC were used as the base and activator along with 20% (v/v) piperazine in DMF as the deprotection solution. Both cysteine-containing peptides were cleaved off the resin by bubbling for 4 h in a 95% (v/v) trifluoroacetic acid (TFA), 2.5% (v/v) triisopropylsilane (TIS) (Fisher Scientific), and 2.5% (v/v) ultrapure H<sub>2</sub>O with 2.5% (w/v) dithiothreitol (DTT) (Sigma-Aldrich). The solution was collected, the resin was washed three times with DCM, and the solution was rotary evaporated to remove TFA. Peptide was precipitated in cold diethyl ether and centrifuged at 3000 g for 10 min with a total of three wash steps. Overnight vacuum desiccation left solid pellets of peptide.

### *Peptide Purification and Cyclization*

The peptide pellets were dissolved in water/0.1% (v/v) TFA (at a concentration of about 5 mg/mL) and purified using the CombiFlash RF200 Flash Chromatography instrument (Teledyne Isco) equipped with a 50 g RediSep Rf Gold Reversed-Phase C18 column. 5 mL of peptide solution was loaded for each round of purification. At a 40 mL/min flow rate with H<sub>2</sub>O/0.1% TFA (solvent A) and acetonitrile (solvent B), the gradient used was: 0% solvent B from 0-3 min, ramped up to 100% solvent B at 9 min, ramped down to 0% solvent B at 11 min, and held at 0% solvent B until 15 min. Collected sample fractions were rotary evaporated to remove acetonitrile and trace amounts of TFA before lyophilizing. For cyclization, lyophilized peptides were reconstituted in a 10 mM borate buffer (pH 8.5) and stirred for 24 h. These peptides were then purified and lyophilized a second time.

### *Automated Post-Polymerization Modification*

DBCO-NH<sub>2</sub> addition was completed in an automated fashion using the Hamilton Microlab STARlet liquid handling robot.<sup>52</sup> Both DBCO-NH<sub>2</sub> and DMAP were prepared at 0.1 M in DMSO. Inputting custom user-defined reagent addition scripts allowed for addition of DBCO-NH<sub>2</sub> and DMAP directly into the 96-well plates based on methacrylic acid NHS content (molar ratio of DBCO-NH<sub>2</sub>:DMAP:NHS of 1:1:1). This reaction was left overnight to completion in sealed 96-well plates.

### *Purification, Peptide Addition, and Sample Preparation*

After DBCO-NH<sub>2</sub> and DMAP addition, all copolymers were purified using 96-well Zeba filter plates (Thermo Fisher Scientific) manually packed with Sephadex G-25 superfine, following a previously published protocol.<sup>50</sup> Incorporated DBCO concentration was quantified by measuring absorbance at the characteristic wavelength of 310 nm and comparing to a DBCO-NH<sub>2</sub> standard curve. Addition of WDCL and EPO peptides to conjugate onto copolymers by strain-promoted azide-alkyne cycloaddition (SPAAC) was performed using the same automated reagent addition scripts as the DBCO-NH<sub>2</sub> and DMAP addition step and left overnight for reaction completion. Polymer-peptide conjugates were then diluted 3x in DMSO and 10x in ultrapure H<sub>2</sub>O before being dialyzed into H<sub>2</sub>O using Spectra/Por 6 RC 18 mm dialysis tubing (Repligen). Dialysis was completed with three H<sub>2</sub>O changes over a period of two days. The remaining solutions were then lyophilized.

### *Size-Exclusion Chromatography-Multi-Angle Light Scattering (SEC-MALS)*

SEC-MALS was completed on two separate systems (organic and aqueous) to quantify polymer backbone molecular weight and dispersity ( $D$ ) depending on solubility. A 1200 series Agilent GPC was used to run all SEC-MALS. An organic system running on DMF with 50 mM LiBr mobile phase and an aqueous system running on 0.5x PBS (BioShop) with 0.02% (w/v) sodium azide were utilized depending on polymer solubility. The organic system separated copolymers by passing through two Agilent PLgel 5  $\mu\text{m}$  columns in series ( $10^3$  and  $10^4$  Å, 300 x 7.5 mm). The aqueous system utilized a Superose 12 10/300 GL 11  $\mu\text{m}$  column (300 x 10 mm). Both systems contained a pre-column stainless steel filter with a 2  $\mu\text{m}$  pore size (VWR). Following the columns

are an Agilent UV detector, Wyatt Technology miniDAWN TREOS MALS detector, and Agilent 1260 Series differential refractive index (RI) detector. SEC calibration and MALS detector normalization was carried out with polyethylene oxide standards (Agilent). All copolymers were prepared at approximately 2 mg/mL and filtered with a 0.45  $\mu\text{m}$  PTFE filter for organic SEC-MALS and a 0.45  $\mu\text{m}$  nylon filter for aqueous SEC-MALS.

### *High-Throughput Dynamic Light Scattering (DLS)*

Initial DLS screening of copolymers was conducted in 384-well plates using the DynaPro DLS Plate Reader III (Wyatt Technology) which uses a wavelength of 830 nm and scattering detector angle of  $173^\circ$ . For this, copolymers were diluted directly from the reaction volume into HBS-EP+ buffer (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.05% P20 at pH 7.6) (VWR) such that solutions contained 1% (v/v) DMSO. Hydrodynamic radius ( $R_h$ ) and diffusion interaction parameter ( $k_D$ ) of copolymers was quantified by cumulants analysis by using auto-attenuation with 20 acquisitions, 5 s/acquisition, and four concentrations depending on the molecular weight of the copolymer. This screening enabled selection of copolymers, polymer-PEG conjugates, and polymer-peptide conjugates for later SAXS characterization. For this, copolymers and copolymer conjugates had to be lyophilized to allow for proper SAXS buffer matching. DLS data was collected for lyophilized samples in 384-well plates using the same acquisition settings for concentrations of 1, 2, 5, and 10 mg/mL for polymer backbones and 1, 2, and 5 mg/mL for copolymer conjugates. All samples were centrifuged at 10,000 g for 10 min before loading the supernatant into 384-well plates for data collection. A data filter was applied to all data to assess data quality, following the Wyatt Technology default parameters (minimum amplitude = 0, maximum amplitude = 1, and baseline limit =  $1 \pm 0.01$ ).

### *High-Throughput Small-Angle X-ray Scattering (SAXS)*

SAXS experiments were conducted at beamline 16-ID of the National Synchrotron Light Source II (NSLS-II) at Brookhaven National Laboratory (Upton, NY) which is for Life Science X-ray Scattering (LiX). Copolymers and copolymer conjugates were lyophilized and prepared at different concentration ranges (1-10 mg/mL and 1-5 mg/mL, respectively). All samples were centrifuged at 10,000 g for 10 min. Copolymers and copolymer conjugates were transferred from PCR plates to PCR tubes manually on-site using multichannel pipettes. In this format, samples and buffer blanks were loaded into custom sample holders where a robotic arm can handle a sequence of sample holders (1 buffer for every 3 samples). This beamline contains three Pilatus 1M detectors which have a  $q$  range of 0.005-3.13  $\text{\AA}^{-1}$  ( $q = 0.005\text{-}0.25 \text{\AA}^{-1}$  was taken as the small-angle region).<sup>100-</sup>  
<sup>102</sup> All background subtracted data were analyzed using BioXTAS RAW 2.1 with ATSAS 2.8.4 to perform Guinier analysis and indirect Fourier transform (IFT) in the same manner as previously done.<sup>51, 76, 103, 104</sup> Porod analysis was also utilized to quantify the flexibility of each copolymer chain in BIOSIS ScÅtter 4.0.<sup>105, 106</sup> Once the hyperbolic Porod region was identified, the linear region of the  $q^4 I(q)$  vs  $q^4$  Porod-Debye plot was selected to determine the Porod exponent.<sup>107-110</sup>

### *Simulated PDB Dataset*

Because proteins are well-characterized and this work takes inspiration from protein design, we compiled a set of 304 PDB files of single protein species. These PDB files were used as an input to obtain simulated SAXS  $I(q)$  vs  $q$ . The simulated data was generated in CRY SOL, a program in the ATSAS software suite.<sup>76, 77</sup> For a  $q$  range of 0-0.25  $\text{\AA}^{-1}$ , default parameters were used except for the number of points in the  $q$  range which was increased to 200 points. With the simulated SAXS data, the same analysis was completed as was done for experimental data of copolymers

and copolymer conjugates to compute  $R_g$ ,  $D_{\max}$ , Porod exponent, and Porod volume. These structure files were also input into HullRad which can calculate hydrodynamic size with a convex hull model.<sup>111</sup> The FASTA files associated with each PDB entry selected were then used to calculate chemical parameters such as molecular weight, total charge, total positive charge, total negative charge,  $\log P$ , and HLB.

### *ML Models*

Data obtained from random screening SAXS and DLS experiments were utilized to train evidential regression networks (ENets),<sup>71, 72</sup> a type of deep neural network for prediction of  $R_g$ ,  $R_h$ ,  $R_g/R_h$ , and Porod Exponent. In this approach, a neural network is trained to predict the hyperparameters ( $\gamma$ ,  $\nu$ ,  $\alpha$ ,  $\beta$ ) of a normal-inverse gamma (NIG) distribution representative of labeled data. This enables efficient prediction of complex data, while also learning uncertainty within data. To train the networks, relative incorporation of monomers, peptides, and degree of polymerization of SCNPs were represented as one-hot encoding vectors and normalized using Sci-Kit Learn's RobustScaler transform. ENets were trained to predict the normalized values of  $R_g$ ,  $R_h$ ,  $R_g/R_h$ , and Porod exponent after RobustScaler transformation. To determine effective ENet hyperparameters, we performed 10-Kfold cross-validation utilizing the Python KerasTuner package to minimize a combined loss function of negative log-likelihood (NLL) loss, a regularization coefficient, and mean squared error (MSE) as described elsewhere.<sup>112</sup> Optimal hyperparameters that were identified on nine of ten folds were then utilized to make predictions on target  $R_g$ ,  $R_h$ ,  $R_g/R_h$ , and Porod Exponent values from the remaining fold. This process was repeated for each fold until predictions on all held-out folds were made. These 10 sets of held-out predictions were then utilized to validate model performance and determine a final set of

hyperparameters for forward predictions of EPO-SCNPs (**Figure S5**). To determine an appropriate value for our loss function regularization coefficient ( $L^R$ ), values from  $1e-5:1$  were evaluated by observing model confidence on held-out data predictions. An  $L^R$  of 0.1 was selected for further hyperparameter optimization as it demonstrated effective calibration of uncertainty on held out data (**Figure S6**).

### *Prediction and Selection of Compact SCNPs*

To propose candidate SCNPs for synthesis and characterization, ENets trained on Porod exponent and  $R_g$  respectively were utilized to predict NIG distributions for  $\sim 750,000$  SCNP candidates representative of similar chemical constraints during our random screen. As we observed that predictions for EPO-SCNPs demonstrated an overall greater Porod Exponent than WDCL-SCNPs and PEG-SCNPs (**Figure S7**), we limited our selection to EPO-SCNPs for synthesis. From the remaining predictions, candidates were filtered to require predicted  $R_g < 7$  nm, Porod exponent of 3.7 - 4.0, and exist in the lowest quartile of predicted variance for Porod exponent predictions. These criteria filtered possible candidates to 336 SCNP chemistries, from which 30 were randomly sampled for final synthesis and characterization. Additionally, to ensure that proposed EPO-SCNPs were not highly similar to previously synthesized SCNPs in our random screen, three-dimensional PCA analysis was performed on ENet proposed EPO-SCNPs in comparison to EPO-SCNPs evaluated during random screen (**Figure S8**). Distinct differences between model proposed EPO-SCNPs and random screened samples are observed.

## **ASSOCIATED CONTENT**

Supporting Information for “Data-Driven Design of Protein-Like Single-Chain Polymer Nanoparticles”

## **AUTHOR INFORMATION**

### **Corresponding Author**

\*Adam J. Gormley: [adam.gormley@rutgers.edu](mailto:adam.gormley@rutgers.edu)

### **Author Contributions**

†These authors contributed equally to this work.

## **ACKNOWLEDGMENTS**

Funding support for this work was provided by the National Science Foundation under CBET Award Number 2009942 and by the National Institutes of Health under NIGMS MIRA Award R35GM138296. R.U. acknowledges support from the New Jersey Commission on Cancer Research (NJCCR) Pre-Doctoral Research Fellowship under the New Jersey Department of Health. R.U., M.J.T., and E.D. acknowledge support from the NIH Biotechnology Training program from The National Institute of General Medical Sciences (NIGMS) under grant T32 GM135141. We acknowledge James Byrnes, beamline scientist at NSLS-II beamline 16-ID for Life Science X-ray Scattering (LiX), for his assistance with conducting experiments at Brookhaven National Laboratory. The LiX beamline is part of the Center for BioMolecular Structure (CBMS), which is primarily supported by the National Institutes of Health, National Institute of General Medical Sciences (NIGMS) through a P30 Grant (P30GM133893), and by the DOE Office of

Biological and Environmental Research (KP1605010). LiX also received additional support from NIH Grant S10 OD012331. As part of NSLS-II, a national user facility at Brookhaven National Laboratory, work performed at the CBMS was supported in part by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences Program under contract number DE-SC0012704.

## DATA AVAILABILITY STATEMENT

The data that supports the findings of this study is openly available on Github at <https://github.com/GormleyLab/Data-Driven-Design-of-SCNPs>.<sup>113</sup>

## REFERENCES

1. Dill, K. A., Polymer principles and protein folding. *Protein Sci* **1999**, *8* (6), 1166-80.
2. Chan, H. S.; Dill, K. A., "Sequence space soup" of proteins and copolymers. *J Chem Phys* **1991**, *95* (5), 3775-3787.
3. Dill, K. A., Theory for the folding and stability of globular proteins. *Biochemistry* **1985**, *24* (6), 1501-9.
4. Dill, K. A.; MacCallum, J. L., The protein-folding problem, 50 years on. *Science* **2012**, *338* (6110), 1042-6.
5. Pomposo, J. A., Bioinspired single-chain polymer nanoparticles. *Polym Int* **2013**, *63* (4), 589-592.
6. Barbee, M. H. W., Z. M.; Allen, B. P.; Taylor, H. F.; Patteson, E. F.; Knight, A. S., Protein-Mimetic Self-Assembly with Synthetic Macromolecules. *Macromolecules* **2021**, *54* (8), 3585-3612.
7. Nassar, R.; Dignon, G. L.; Razban, R. M.; Dill, K. A., The Protein Folding Problem: The Role of Theory. *J Mol Biol* **2021**, *433* (20), 167126.
8. Warren, J. L.; Dykeman-Birmingham, 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.
9. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, 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.
10. Asenjo-Sanz, I.; Verde-Sesto, E.; Pomposo, J. A., A method to estimate the size of single-chain nanoparticles under severe crowding conditions. *RSC Adv* **2022**, *12* (3), 1571-1575.

11. Berda, E. B.; Foster, J.; Meijer, E. W., Toward Controlling Folding in Synthetic Polymers: Fabricating and Characterizing Supramolecular Single-Chain Nanoparticles. *Macromolecules* **2010**, *43* (3), 1430-1437.
12. Hosono, N.; Gillissen, M. A.; Li, Y.; Sheiko, S. S.; Palmans, A. R.; Meijer, E. W., Orthogonal self-assembly in folding block copolymers. *J Am Chem Soc* **2013**, *135* (1), 501-10.
13. Lo Verso, F.; Pomposo, J. A.; Colmenero, J.; Moreno, A. J., Multi-orthogonal folding of single polymer chains into soft nanoparticles. *Soft Matter* **2014**, *10* (27), 4813-4821.
14. Lo Verso, F.; Pomposo, J. A.; Colmenero, J.; Moreno, A. J., Simulation guided design of globular single-chain nanoparticles by tuning the solvent quality. *Soft Matter* **2015**, *11* (7), 1369-1375.
15. Moreno, A. J.; Lo Verso, F., *Single-Chain Polymer Nanoparticles*. Wiley-VCH: Weinheim, Germany, 2017.
16. Rabbel, H.; Breier, P.; Sommer, J. U., Swelling Behavior of Single-Chain Polymer Nanoparticles: Theory and Simulation. *Macromolecules* **2017**, *50* (18), 7410-7418.
17. Rosales, A. M.; Segalman, R. A.; Zuckermann, R. N., Polypeptoids: a model system to study the effect of monomer sequence on polymer properties and self-assembly. *Soft Matter* **2013**, *9* (35), 8400-8414.
18. Verde-Sesto, E. A., A.; Moreno, A. J.; Cangialosi, D.; Alegria, A.; Colmenero, J.; Pomposo, J. A., Single-chain nanoparticles: opportunities provided by internal and external confinement. *Mater. Horiz.* **2020**, *7*, 2292-2313.
19. Asenjo-Sanz, I.; Claros, T.; González, E.; Pinacho-Olaciregui, J.; Verde-Sesto, E.; Pomposo, J. A., Significant effect of intra-chain distribution of catalytic sites on catalytic activity in “clickase” single-chain nanoparticles. *Mater Lett* **2021**, *304* (1), 130622.
20. Perez-Baena, I.; Barroso-Bujans, F.; Gasser, U.; Arbe, A.; Moreno, A. J.; Colmenero, J.; Pomposo, J. A., Endowing Single-Chain Polymer Nanoparticles with Enzyme-Mimetic Activity. *ACS Macro Lett* **2013**, *2* (9), 775-779.
21. Cole, J. P.; Hanlon, A. M.; Rodriguez, K. J.; Berda, E. B., Protein-like structure and activity in synthetic polymers. *J Polym Sci A Polym Chem* **2017**, *55* (2), 191-206.
22. 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.
23. Besford, Q. A.; Yong, H.; Merlitz, H.; Christofferson, A. J.; Sommer, J. U.; Uhlmann, P.; Fery, A., FRET-Integrated Polymer Brushes for Spatially Resolved Sensing of Changes in Polymer Conformation. *Angew Chem Int Ed Engl* **2021**, *60* (30), 16600-16606.
24. Joseph, J. P. M., A.; Bhatt, A.; Ray, D.; Singh, A.; Gupta, D.; Ali, M. E.; Aswal, V. K.; Pal, A., Delineating synchronized control of dynamic covalent and non-covalent interactions for polymer chain collapse towards cargo localization and delivery†. *Polym. Chem.* **2021**, *12* (7), 1002-1013.
25. Murthy, N. S.; Zhang, Z.; Borsadia, S.; Kohn, J., Nanospheres with a smectic hydrophobic core and an amorphous PEG hydrophilic shell: structural changes and implications for drug delivery. *Soft Matter* **2018**, *14* (8), 1327-1335.
26. Upadhyaya, R.; Punia, A.; Kanagala, M. J.; Liu, L.; Lamm, M.; Rhodes, T. A.; Gormley, A. J., Automated PET-RAFT Polymerization Towards Pharmaceutical Amorphous Solid Dispersion Development. *ACS Appl Polym Mater* **2021**, *3* (3), 1525-1536.
27. Alvarez, Z.; Kolberg-Edelbrock, A. N.; Sasselli, I. R.; Ortega, J. A.; Qiu, R.; Syrgiannis, Z.; Mirau, P. A.; Chen, F.; Chin, S. M.; Weigand, S.; Kiskinis, E.; Stupp, S. I.,

- Bioactive scaffolds with enhanced supramolecular motion promote recovery from spinal cord injury. *Science* **2021**, 374 (6569), 848-856.
28. Maynard, S. A.; Winter, C. W.; Cunnane, E. M.; Stevens, M. M., Advancing cell instructive biomaterials through increased understanding of cell receptor spacing and material surface functionalization. *Regen Eng Transl Med* **2021**, 7 (4), 553-547.
29. Woythe, L.; Tito, N. B.; Albertazzi, L., A quantitative view on multivalent nanomedicine targeting. *Adv Drug Deliv Rev* **2021**, 169, 1-21.
30. Johnson, D. L.; Farrell, F. X.; Barbone, F. P.; McMahon, F. J.; Tullai, J.; Hoey, K.; Livnah, O.; Wrighton, N. C.; Middleton, S. A.; Loughney, D. A.; Stura, E. A.; Dower, W. J.; Mulcahy, L. S.; Wilson, I. A.; Jolliffe, L. K., Identification of a 13 amino acid peptide mimetic of erythropoietin and description of amino acids critical for the mimetic activity of EMP1. *Biochemistry* **1998**, 37 (11), 3699-710.
31. Livnah, O.; Stura, E. A.; Johnson, D. L.; Middleton, S. A.; Mulcahy, L. S.; Wrighton, N. C.; Dower, W. J.; Jolliffe, L. K.; Wilson, I. A., Functional mimicry of a protein hormone by a peptide agonist: the EPO receptor complex at 2.8 Å. *Science* **1996**, 273 (5274), 464-71.
32. Wrighton, N. C.; Balasubramanian, P.; Barbone, F. P.; Kashyap, A. K.; Farrell, F. X.; Jolliffe, L. K.; Barrett, R. W.; Dower, W. J., Increased potency of an erythropoietin peptide mimetic through covalent dimerization. *Nat Biotechnol* **1997**, 15 (12), 1261-5.
33. Johnson, D. L.; Farrell, F. X.; Barbone, F. P.; McMahon, F. J.; Tullai, J.; Kroon, D.; Freedy, J.; Zivin, R. A.; Mulcahy, L. S.; Jolliffe, L. K., Amino-terminal dimerization of an erythropoietin mimetic peptide results in increased erythropoietic activity *Chem Biol* **1997**, 4 (12), 939-950.
34. Wrighton, N. C.; Farrell, F. X.; Chang, R.; Kashyap, A. K.; Barbone, F. P.; Mulcahy, L. S.; Johnson, D. L.; Barrett, R. W.; Jolliffe, L. K.; Dower, W. J., Small Peptides as Potent Mimetics of the Protein Hormone Erythropoietin. *Science* **1996**, 273 (5274), 458-463.
35. Ehrlich, P. H., The effect of multivalency on the specificity of protein and cell interactions. *J Theor Biol* **1979**, 81 (1), 123-7.
36. Mammen, M.; Choi, S. K.; Whitesides, G. M., Polyvalent Interactions in Biological Systems: Implications for Design and Use of Multivalent Ligands and Inhibitors. *Angew Chem Int Ed Engl* **1998**, 37 (20), 2754-2794.
37. Tjandra, K. C.; Thordarson, P., Multivalency in Drug Delivery-When Is It Too Much of a Good Thing? *Bioconjug Chem* **2019**, 30 (3), 503-514.
38. Moreno, A. J.; Lo Verso, F.; Sanchez-Sanchez, A.; Arbe, A.; Colmenero, J.; Pomposo, J. A., Advantages of Orthogonal Folding of Single Polymer Chains to Soft Nanoparticles. *Macromolecules* **2013**, 46 (24), 9748-9759.
39. Oliver, S. Z., 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.
40. Yeow, J.; Chapman, R.; Gormley, A. J.; Boyer, C., Up in the air: oxygen tolerance in controlled/living radical polymerisation. *Chem Soc Rev* **2018**, 47 (12), 4357-4387.
41. Gormley, A. J.; Yeow, J.; Ng, G.; Conway, O.; Boyer, C.; Chapman, R., An Oxygen-Tolerant PET-RAFT Polymerization for Screening Structure-Activity Relationships. *Angew Chem Int Ed Engl* **2018**, 57 (6), 1557-1562.
42. Ng, G. Y., J.; Chapman, R.; Isahak, N.; Wolvetang, E.; Cooper-White, J. J.; Boyer, C., Pushing the Limits of High Throughput PET-RAFT Polymerization. *Macromolecules* **2018**, 51 (19), 7600-7607.

43. Ng., G. Y., J.; Xu, J. T.; Boyer, C., Application of oxygen tolerant PET-RAFT to polymerization-induced self-assembly. *Polym. Chem.* **2017**, *8* (18), 2841-2851.
44. 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), 5508-19.
45. Yeow, J. C., R.; Xu, J. T.; Boyer, C., Oxygen tolerant photopolymerization for ultralow volumes. *Polym. Chem.* **2017**, *8* (34), 5012-5022.
46. Enciso, A. E.; Fu, L.; Russell, A. J.; Matyjaszewski, K., A Breathing Atom-Transfer Radical Polymerization: Fully Oxygen-Tolerant Polymerization Inspired by Aerobic Respiration of Cells. *Angew Chem Int Ed Engl* **2018**, *57* (4), 933-936.
47. Chapman, R.; Gormley, A. J.; Stenzel, M. H.; Stevens, M. M., Combinatorial Low-Volume Synthesis of Well-Defined Polymers by Enzyme Degassing. *Angew Chem Int Ed Engl* **2016**, *55* (14), 4500-3.
48. Chapman, R. G., A. J.; Herpoldt, K.; Stevens, M. M., Highly Controlled Open Vessel RAFT Polymerizations by Enzyme Degassing. *Macromolecules* **2014**, *47* (24), 8541-8547.
49. Atta, S.; Cohen, J.; Kohn, J.; Gormley, A. J., Ring opening polymerization of  $\epsilon$ -caprolactone through water. *Polym Chem* **2021**, *12* (2), 159-164.
50. Upadhy, R.; Kanagala, M. J.; Gormley, A. J., Purifying Low-Volume Combinatorial Polymer Libraries with Gel Filtration Columns. *Macromol Rapid Commun* **2019**, *40* (24), e1900528.
51. Upadhy, 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.
52. Tamasi, M.; Kosuri, S.; DiStefano, J.; Chapman, R.; Gormley, A. J., Automation of Controlled/Living Radical Polymerization. *Adv Intell Syst* **2020**, *2* (2).
53. Li, Z.; Kosuri, S.; Foster, H.; Cohen, J.; Jumeaux, C.; Stevens, M. M.; Chapman, R.; Gormley, A. J., A Dual Wavelength Polymerization and Bioconjugation Strategy for High Throughput Synthesis of Multivalent Ligands. *J Am Chem Soc* **2019**, *141* (50), 19823-19830.
54. Richards, S. J.; Jones, A.; Tomas, R. M. F.; Gibson, M. I., Photochemical "In-Air" Combinatorial Discovery of Antimicrobial Co-polymers. *Chemistry* **2018**, *24* (52), 13758-13761.
55. Stubbs, C.; Congdon, T.; Davis, J.; Lester, D.; Richards, S. J.; Gibson, M. I., High-Throughput Tertiary Amine Deoxygenated Photopolymerizations for Synthesizing Polymer Libraries. *Macromolecules* **2019**, *52* (20), 7603-7612.
56. Stubbs, C.; Murray, K. A.; Ishibe, T.; Mathers, R. T.; Gibson, M. I., Combinatorial Biomaterials Discovery Strategy to Identify New Macromolecular Cryoprotectants. *ACS Macro Lett* **2020**, *9* (2), 290-294.
57. Upadhy, R.; Kosuri, S.; Tamasi, M.; Meyer, T. A.; Atta, S.; Webb, M. A.; Gormley, A. J., Automation and data-driven design of polymer therapeutics. *Adv Drug Deliv Rev* **2021**, *171*, 1-28.
58. Chan, D.; Chien, J. C.; Axpe, E.; Blankemeier, L.; Baker, S. W.; Swaminathan, S.; Piunova, V. A.; Zubarev, D. Y.; Maikawa, C. L.; Grosskopf, A. K.; Mann, J. L.; Soh, H. T.; Appel, E. A., Combinatorial Polyacrylamide Hydrogels for Preventing Biofouling on Implantable Biosensors. *Adv Mater* **2022**, *34* (24), e2109764.
59. Georgiou, P. G.; Marton, H. L.; Baker, A. N.; Congdon, T. R.; Whale, T. F.; Gibson, M. I., Polymer Self-Assembly Induced Enhancement of Ice Recrystallization Inhibition. *J Am Chem Soc* **2021**, *143* (19), 7449-7461.

60. Pesenti, T.; Zhu, C.; Gonzalez-Martinez, N.; Tomas, R. M. F.; Gibson, M. I.; Nicolas, J., Degradable Polyampholytes from Radical Ring-Opening Copolymerization Enhance Cellular Cryopreservation. *ACS Macro Lett* **2022**, *11* (7), 889-894.
61. Coley, C. W.; Thomas, D. A., 3rd; Lummiss, J. A. M.; Jaworski, J. N.; Breen, C. P.; Schultz, V.; Hart, T.; Fishman, J. S.; Rogers, L.; Gao, H.; Hicklin, R. W.; Plehiers, P. P.; Byington, J.; Piotti, J. S.; Green, W. H.; Hart, A. J.; Jamison, T. F.; Jensen, K. F., A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science* **2019**, *365* (6453).
62. Guan, Y.; Coley, C. W.; Wu, H.; Ranasinghe, D.; Heid, E.; Struble, T. J.; Pattanaik, L.; Green, W. H.; Jensen, K. F., Regio-selectivity prediction with a machine-learned reaction representation and on-the-fly quantum mechanical descriptors. *Chem Sci* **2020**, *12* (6), 2198-2208.
63. Webb, M. A.; Jackson, N. E.; Gil, P. S.; de Pablo, J. J., Targeted sequence design within the coarse-grained polymer genome. *Sci Adv* **2020**, *6* (43).
64. Fong, A. Y.; Pellouchoud, L.; Davidson, M.; Walroth, R. C.; Church, C.; Tcareva, E.; Wu, L.; Peterson, K.; Meredig, B.; Tassone, C. J., Utilization of machine learning to accelerate colloidal synthesis and discovery. *J Chem Phys* **2021**, *154* (22), 224201.
65. Wu, S.; Kondo, Y.; Kakimoto, M. A.; Yang, B.; Yamada, H.; Kuwajima, I.; Lambard, G.; Hongo, K.; Xu, Y.; Shiomi, J.; Schick, C.; Morikawa, J.; Yoshida, R., Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm. *Npj Comput Mater* **2019**, *5* (66), 1-11.
66. Knox, S. T. W., N. J., Enabling technologies in polymer synthesis: accessing a new design space for advanced polymer materials. *React. Chem. Eng.* **2020**, *5* (3), 405-423.
67. Dijkstra, M.; Luijten, E., From predictive modelling to machine learning and reverse engineering of colloidal self-assembly. *Nat Mater* **2021**, *20* (6), 762-773.
68. Gormley, A. J.; Webb, M. A., Machine learning in combinatorial polymer chemistry. *Nat Rev Mater* **2021**, *6*, 642-644.
69. 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 Healthc Mater* **2022**, *11* (10), e2102101.
70. Tamasi, M. J.; Patel, R. A.; Borca, C. H.; Kosuri, S.; Mugnier, H.; Upadhyaya, 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**, e2201809.
71. Soleimany, A. P.; Amini, A.; Goldman, S.; Rus, D.; Bhatia, S. N.; Coley, C. W., Evidential Deep Learning for Guided Molecular Property Prediction and Discovery. *ACS Cent Sci* **2021**, *7* (8), 1356-1367.
72. Amini, A.; Schwarting, W.; Soleimany, A., Deep evidential regression. *Adv Neural Inf Process Syst* **2020**, *33*, 14927-14937.
73. Pavet, V.; Beyrath, J.; Pardin, C.; Morizot, A.; Lechner, M. C.; Briand, J. P.; Wendland, M.; Maison, W.; Fournel, S.; Micheau, O.; Guichard, G.; Gronemeyer, H., Multivalent DR5 peptides activate the TRAIL death pathway and exert tumoricidal activity. *Cancer Res* **2010**, *70* (3), 1101-10.
74. Henriques, J.; Arleth, L.; Lindorff-Larsen, K.; Skepö, M., On the Calculation of SAXS Profiles of Folded and Intrinsically Disordered Proteins from Computer Simulations. *J Mol Biol* **2018**, *16* (3), 2521-2539.

75. ww, P. D. B. c., Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res* **2019**, *47* (D1), D520-D528.
76. Franke, D.; Petoukhov, M. V.; Konarev, P. V.; Panjkovich, A.; Tuukkanen, A.; Mertens, H. D. T.; Kikhney, A. G.; Hajizadeh, N. R.; Franklin, J. M.; Jeffries, C. M.; Svergun, D. I., ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J Appl Crystallogr* **2017**, *50* (Pt 4), 1212-1225.
77. Svergun, D.; Barberato, C.; Koch, M. H. J., CRY SOL-a Program to Evaluation X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J Appl Crystallogr* **1995**, *28* (6), 768-773.
78. Gomes, G. W.; Krzeminski, M.; Namini, A.; Martin, E. W.; Mittag, T.; Head-Gordon, T.; Forman-Kay, J. D.; Gradinaru, C. C., Conformational Ensembles of an Intrinsically Disordered Protein Consistent with NMR, SAXS, and Single-Molecule FRET. *J Am Chem Soc* **2020**, *142* (37), 15697-15710.
79. Henriques, J.; Cragnell, C.; Skepo, M., Molecular Dynamics Simulations of Intrinsically Disordered Proteins: Force Field Evaluation and Comparison with Experiment. *J Chem Theory Comput* **2015**, *11* (7), 3420-31.
80. Receveur-Bréchet, V.; Durand, D., How Random are Intrinsically Disordered Proteins? A Small Angle Scattering Perspective. *Curr Protein Pept Sci* **2012**, *13* (1), 55-75.
81. Soheilmoghaddam, F.; Rumble, M.; Cooper-White, J., High-Throughput Routes to Biomaterials Discovery. *Chem Rev* **2021**, *121* (18), 10792-10864.
82. Hirzinger, B.; Helmstedt, M.; Stejskal, J., Light scattering studies on core-shell systems: determination of size parameters of sterically stabilized poly(methylmethacrylate) dispersions. *Polym* **2000**, *41* (8), 2883-2891.
83. Kujawa, P.; Tanaka, F.; Winnik, F. M., Temperature-Dependent Properties of Telechelic Hydrophobically Modified Poly(N-isopropylacrylamides) in Water: Evidence from Light Scattering and Fluorescence Spectroscopy for the Formation of Stable Mesoglobules at Elevated Temperatures. *Macromolecules* **2006**, *39* (8), 3048-3055.
84. Lundberg, S. M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J. M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S. I., From Local Explanations to Global Understanding with Explainable AI for Trees. *Nat Mach Intell* **2020**, *2* (1), 56-67.
85. Lundberg, S. M.; Lee, S. I. In *A unified approach to interpreting model predictions*, 31st Int Conf on Neural Information Processing Systems, Red Hook, NY, USA, Curran Associates Inc. : Red Hook, NY, USA, 2017; pp 4768-4777.
86. 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.
87. DeStefano, A. J.; Segalman, R. A.; Davidson, E. C., Where Biology and Traditional Polymers Meet: The Potential of Associating Sequence-Defined Polymers for Materials Science. *JACS Au* **2021**, *1* (10), 1556-1571.
88. Allen, B. P.; Wright, Z. M.; Taylor, H. F.; Oweida, T. J.; Kader-Pinky, S.; Patteson, E. F.; Bucci, K. M.; Cox, C. A.; Senthilvel, A. S.; Yingling, Y. G.; Knight, A. S., Mapping the Morphological Landscape of Oligomeric Di-block Peptide-Polymer Amphiphiles. *Angew Chem Int Ed Engl* **2022**, *61* (14), e202115547.
89. Heling, L. W. H. J.; Banijamali, S. E.; Satarifard, V.; Mashaghi, A., Programmed Polymer Folding. In *Topological Polymer Chemistry*, Tezuka, Y.; Deguchi, T., Eds. Springer: Singapore, 2022.

90. Jung, K.; Corrigan, N.; Wong, E. H. H.; Boyer, C., Bioactive Synthetic Polymers. *Adv Mater* **2022**, *34* (2), e2105063.
91. Hirano, T.; Kamiike, R.; Yuki, T.; Matsumoto, D.; Ute, K., Determination of monomer reactivity ratios from a single sample using multivariate analysis of the <sup>1</sup>H NMR spectra of poly[(methyl methacrylate)-co-(benzyl methacrylate)]. *Polym J* **2022**, *54* (5).
92. Ting, J. M.; Navale, T. S.; Bates, F. S.; Reineke, T. M., Precise Compositional Control and Systematic Preparation of Multimonomeric Statistical Copolymers. *ACS Macro Lett* **2013**, *2* (9), 770-774.
93. Ting, J. M.; Navale, T. S.; Jones, S. D.; Bates, F. S.; Reineke, T. M., Deconstructing HPMCAS: Excipient Design to Tailor Polymer-Drug Interactions for Oral Drug Delivery. *ACS Biomater Sci Eng* **2015**, *1* (10), 978-990.
94. Smith, A. A. A.; Hall, A.; Wu, V.; Xu, T., Practical Prediction of Heteropolymer Composition and Drift. *ACS Macro Lett* **2019**, *8* (1), 36-40.
95. Liu, K.; Corrigan, N.; Postma, A.; Moad, G.; Boyer, C., A Comprehensive Platform for the Design and Synthesis of Polymer Molecular Weight Distributions. *Macromolecules* **2020**, *53* (20), 8867-8882.
96. Corrigan, N.; Trujillo, F. J.; Xu, J.; Moad, G.; Hawker, C. J.; Boyer, C., Divergent Synthesis of Graft and Branched Copolymers through Spatially Controlled Photopolymerization in Flow Reactors. *Macromolecules* **2021**, *54* (7), 3430-3446.
97. Berger, O.; Battistella, C.; Chen, Y.; Oktawiec, J.; Siwicka, Z. E.; Tullman-Ercek, D.; Wang, M.; Gianneschi, N. C., Mussel Adhesive-Inspired Proteomimetic Polymer. *J Am Chem Soc* **2022**, *144* (10), 4383-4392.
98. Sun, H.; Cao, W.; Zang, N.; Clemons, T. D.; Scheutz, G. M.; Hu, Z.; Thompson, M. P.; Liang, Y.; Vratsanos, M.; Zhou, X.; Choi, W.; Sumerlin, B. S.; Stupp, S. I.; Gianneschi, N. C., Proapoptotic Peptide Brush Polymer Nanoparticles via Photoinitiated Polymerization-Induced Self-Assembly. *Angew Chem Int Ed Engl* **2020**, *59* (43), 19136-19142.
99. Blazquez-Martin, A.; Verde-Sesto, E.; Moreno, A. J.; Arbe, A.; Colmenero, J.; Pomposo, J. A., Advances in the Multi-Orthogonal Folding of Single Polymer Chains into Single-Chain Nanoparticles. *Polymers (Basel)* **2021**, *13* (2).
100. Yang, L.; Antonelli, S.; Chodankar, S.; Byrnes, J.; Lazo, E.; Qian, K., Solution scattering at the Life Science X-ray Scattering (LiX) beamline. *J Synchrotron Radiat* **2020**, *27* (Pt 3), 804-812.
101. Yang, L.; Lazo, E.; Byrnes, J.; Chodankar, S.; Antonelli, S.; Rakitin, M., Tools for supporting solution scattering during the COVID-19 pandemic. *J Synchrotron Radiat* **2021**, *28* (Pt 4), 1237-1244.
102. Yang, L.; Liu, J.; Chodankar, S.; Antonelli, S.; DiFabio, J., Scanning structural mapping at the Life Science X-ray Scattering Beamline. *J Synchrotron Radiat* **2022**, *29* (Pt 2), 540-548.
103. Hopkins, J. B.; Gillilan, R. E.; Skou, S., BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *J Appl Crystallogr* **2017**, *50* (Pt 5), 1545-1553.
104. Nielsen, S. S.; Toft, K. N.; Snakenborg, D.; Jeppesen, M. G.; Jacobsen, J. K.; Vestergaard, B.; Kutter, J. P.; Arleth, L., BioXTAS RAW, a software program for high-throughput automated small-angle X-ray scattering data reduction and preliminary analysis. *J Appl Crystallogr* **2009**, *42* (5), 959-964.

105. Hura, G. L.; Menon, A. L.; Hammel, M.; Rambo, R. P.; Poole, F. L., 2nd; Tsutakawa, S. E.; Jenney, F. E., Jr.; Classen, S.; Frankel, K. A.; Hopkins, R. C.; Yang, S. J.; Scott, J. W.; Dillard, B. D.; Adams, M. W.; Tainer, J. A., Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat Methods* **2009**, *6* (8), 606-12.
106. Rambo, R. P.; Tainer, J. A., Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* **2011**, *95* (8), 559-71.
107. Besselink, R.; Stawski, T. M.; Van Driessche, A. E. S.; Benning, L. G., Not just fractal surfaces, but surface fractal aggregates: Derivation of the expression for the structure factor and its applications. *J Chem Phys* **2016**, *145* (21), 1-11.
108. Martin, J. E.; Hurd, A. J., Scattering from Fractals. *J Appl Crystallogr* **1987**, *20* (2), 61-78.
109. Murthy, N. S.; Knox, J. R., On Solu e- Porod Plots of Protein X-Ray Scattering Data. *J Appl Crystallogr* **1977**, *10* (3), 137-140.
110. Teixeira, J., Small-Angle Scattering by Fractal Systems. *J Appl Crystallogr* **1988**, *21* (6), 781-785.
111. Fleming, P. J.; Fleming, K. G., HullRad: Fast Calculations of Folded and Disordered Protein and Nucleic Acid Hydrodynamic Properties. *Biophys J* **2018**, *114* (4), 856-869.
112. Oh, D.; Shin, B., Improving Evidential Deep Learning via Multi-task Learning. *Proc Innov Appl Artif Intell Conf* **2022**, *36* (7), 7895-7903.
113. Gormley, A. J., Data-Driven-Design-of-SCNPs on GitHub.