

# Polymer Design via SHAP and Bayesian Machine Learning Optimizes pDNA and CRISPR Ribonucleoprotein Delivery

**Rishad Dalal**

University of Minnesota

**Michael Leyden**

University of Minnesota

**Felipe Oviedo**

Nanite, Inc.

**Theresa Reineke** (✉ [treineke@umn.edu](mailto:treineke@umn.edu))

University of Minnesota <https://orcid.org/0000-0001-7020-3450>

---

## Article

## Keywords:

**Posted Date:** August 8th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1785891/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

We present the facile synthesis of a clickable polymer library with systematic variations in length, binary composition,  $pK_a$ , and hydrophobicity (clogP) to optimize intracellular pDNA and CRISPR-Cas9 ribonucleoprotein (RNP) performance. We couple physiochemical characterization and machine learning to interpret quantitative structure-property relationships within the combinatorial design space. For the first time, we reveal unexpected disparate design parameters for nucleic acid carriers; via explainable machine learning on 432 formulations, we discover that lower polymer  $pK_a$  and higher percentages of benzimidazole ethanethiol enhance pDNA delivery, yet polymer length and captamine cation identity improve RNP delivery. Closed-loop Bayesian optimization of 552 formulation ratios further enhances *in vitro* performance. The top three polymers yield higher signal and stable transgene expression over 20-days *in vivo*, and a 1.7-fold enhancement over controls. Our facile coupling of synthesis, characterization, and machine analysis provide powerful tools to quantitate performance parameters accelerating next generation vehicles for nucleic acid medicines.

## Full Text

Nucleic acids are important therapeutics yet issues with delivery efficiency continue to hinder widespread advancement in the clinic. Delivery systems are crucial to encapsulate and protect these large and highly sensitive payloads and improve tissue internalization ensuring efficacy.<sup>1,2</sup> Current viral delivery methods have struggled to overcome obstacles of limited cargo capacity,<sup>3</sup> manufacturing costs,<sup>4</sup> and immunogenicity.<sup>5,6</sup> Nonviral delivery methods have been proven in commercial formulations and offer facile, tunable, and inexpensive vehicles for exogenous nucleic acid medicines. Polymers are established pharmaceutical formulation agents but have been under-utilized for carrying nucleic acids *in vivo* due to low performance.<sup>7</sup> However, there exists limitless potential for polymer delivery vehicles delivery agents due to ease of chemical and physical modulation along with affordable and scaled manufacturing.<sup>7,8</sup> The advancement of controlled radical polymerization,<sup>9–15</sup> post-polymerization modification,<sup>16–18</sup> and parallel synthetic techniques,<sup>19–21</sup> have rapidly advanced and offers powerful tools to accelerate the next generation of bioactive polymer libraries.<sup>17,22–24</sup> To this end, the field of machine learning<sup>25–27</sup> coupled with parallel experimentation is aiding analysis and understanding of data sets identifying the chemical, physical, and biological factors involved in performance enhancement of polymers.<sup>23,24,28–33</sup> However, we are limited by the vast chemical space and prediction of formulation chemistries indicating discrete selection and optimization of next generation systems. Indeed, new machine learning models aimed at selecting and predicting discrete parameters influencing biological efficacy are needed to advance the next frontier of personalized medicine.

To accelerate advancement of next generation nucleic acid medicines, we present a facile method to generate cationic polymers coupled with a machine learning workflow as a powerful tool to tailor nucleic acid delivery vehicles (**Figure 1**). Herein, we probe the effect of physicochemical properties on *in vitro* delivery of plasmid DNA (pDNA) and CRISPR-Cas9 ribonucleoprotein (RNP) payloads through Shapley

Additive ExPlanation (SHAP) analysis. We also utilize machine learning through Bayesian optimization (BO) to identify formulation prediction for optimizing *in vitro* delivery. We present three scaffold lengths and 36 copolymers containing systematic binary compositions of a hydrophobic cation benzimidazole ethanethiol (BET), along with the co-cations of cysteamine (Cys), captamine (Cap), or 2-(diethylamino) ethanethiol (DiE), to examine their effect on performance. Our models compare 552-formulations across numerous chemical and physical characteristics of the polymer scaffolds: repeat units (RU), incorporation of BET, type of co-cation, polymer  $pK_a$ , polymer clogP, polyplex size, formulation ratio, and relative binding, to identify overarching structure-property relationships benefitting cellular delivery (transfection) efficiency and cellular viability. Our predictive pipeline identifies three top performers *in vivo*, displaying higher and longer-term transgene expression in the mouse liver compared to a well-studied commercial control (JetPEI). Our powerful methods presented herein provide advanced understanding of quantitative structure-activity relationships important for rapid preclinical development of next generation nucleic acid medicines.

## Polymer Scaffold Synthesis

Pentafluorophenyl methacrylate (PFPMA) is known to polymerize under the reversible addition-fragmentation chain transfer (RAFT) mechanism to control degree of polymerization and dispersity of the backbone.<sup>16,34</sup> PFPMA allows efficient post-polymerization modification due to the labile nature of the pentafluoro group via amidation.<sup>16,17,34–36</sup> We performed synthesis to create three well-defined pPFPMA scaffold lengths yielding short (Sh), medium (Md) and Long (Lg) (N = 90, 190, 250) variants. Through quick amidation with allylamine, we obtain poly(allylmethacrylamide) (pAMAM), a polymer backbone decorated with pendent alkenes and responsive to the highly efficient thiol-ene click chemistry.<sup>37</sup> We create the library of copolymers in a stepwise conjugation of functional groups. First, four compositions of BET, ranging in incorporation from 0-45% are obtained, and then each sample is split into three and further saturated with either Cys, Cap, or DiE cations. Our modular process allows for consistent BET incorporation (high logP, possibly intercalating) and co-cation (promotes electrostatic binding) composition to be consistent across each sample ensuring structural uniformity. Full characterization of the polymer systems through <sup>1</sup>H NMR, <sup>19</sup>F NMR, ATR-FTIR, and SEC-MALLS is shown in **Figures S1-S4, S7-S12**. Polymers are named according to polymer RU\_Cation\_%BET (i.e., Sh\_DiE\_40).

Altering the protonation state and hydrophobicity of the polymer systems has shown to affect the ability of polymers to bind and release nucleic acids, interact with cell membranes, promote endosomal escape, and enable beneficial aggregation *in vitro* promoting particle settling onto cells.<sup>24,38–44</sup> Using an autotitrator, we measured the  $pK_a$  values (**Figure S13**) of the small molecule thiols. The values for Cys, Cap, DiE, and BET, respectively, were 8.10, 7.74, 7.68, and 5.90 (**Figure 2**). It should be noted that upon polymerization neighboring effects suppress amine ionization resulting in a common decrease of one  $pK_a$  unit from monomer to polymer, which is found after measuring  $pK_a$  of all polymers.<sup>45,46</sup> We also calculate the octanol to water partition coefficient (clogP) values of the monomers to be -2.70, -2.50, 1.74, and 2.40, respectively, for Cys, Cap, DiE, and BET; (the more negative the number, the more water soluble).

The molar averages of  $pK_a$  and  $\text{clogP}$  of the copolymer systems based on the repeat units and percent incorporation of the functional moieties are shown in **Figure 2** and **Equations S2, S3**.

## Polyplex Physical Characterization

For physical characterization and analysis of the polymer-payload complexes (polyplexes), we use dynamic light scattering (DLS) to determine the hydrodynamic radius ( $R_h$ ) and a dye exclusion assay to establish the relative binding affinity of each polymer with the pDNA or RNP (**Figure 3B**). We mixed polymers with either payload in a phosphate buffered saline (PBS, pH 7.4) and at a given nitrogen (N, on polymer) to phosphate (P, on nucleotide) N/P ratio (from either pDNA or guide RNA (gRNA)). We observed that pDNA polyplexes formed with the Md and Lg backbone were around 50 nm in size, while the Sh backbone polyplexes ranged in size from 50-250 nm (**Figures S17-S19**). When the polymers were bound with RNP, the polyplex sizes were considerably larger with all of the polyplex systems being larger than 100 nm, ranging to 1000 nm (**Figures S28-S30**). We also found that the RNP aggregates alone, which could be the cause for these larger aggregates.

To compare the relative binding affinity of the polymers to the payloads, we mixed a fluorescent dye, PicoGreen, with pDNA prior to forming polyplexes at various N/P ratios (**Figures S14-S16**). Lower fluorescence is linked to more dye excluded from the pDNA, indicative of higher relative binding affinity.<sup>47</sup> Similar to previous reports, we found that higher steric bulk displayed lower relative binding (bulk increases from Cys to Cap to DiE).<sup>17</sup> Additionally, we show that stronger binding correlates to BET composition, indicating possible intercalation within the pDNA.<sup>40,41</sup> Interestingly, Sh polymer backbones resulted in tighter binding compared to the Md and Lg scaffolds. For RNP binding (**Figure S27**), we used an OliGreen dye for RNP polyplexes. We found similar trends with regard to steric bulk, i.e., Cys functionalized polymers showed the tightest binding.<sup>17</sup> Additionally, we find that higher BET incorporation results in higher dye exclusion from the gRNA. We notice an opposite trend for polymer size however, where Lg scaffold variants promote stronger RNP binding. While efficient binding is important for payload encapsulation and protection, strong binding is not always positively correlated to biological delivery as payload release is important for performance. Overall, polyplex size and binding appeared to be correlated to BET composition, cation bulk, and scaffold length.

## Transfection Performance and Toxicity

Reporter gene assays, via fluorescence output, are a facile tool to assess delivery of biological payloads. We employ two transfection assays (**Figure 3A**): i) delivery of pDNA to HEK293T cells exhibiting expression of a green fluorescent protein (GFP) and ii) delivery of RNP that upregulates mCherry expression in a modified HEK293T cell<sup>48</sup> type, where expression can be measured on a per cell count through flow cytometry. We monitor cell health and metabolism through a cell counting kit (CCK-8) viability assay. As an initial screen we transfect pDNA at N/P ratios of 10 and 20 with the entire polymer library, and GFP expression was probed forty-eight hours post-transfection (**Figure 3B, S24-S26**). We find that increasing BET in the Sh and Md scaffolds enhances transfection efficiency. The most effective

polymer is Sh\_DiE\_40 (polymer RU\_Cation\_%BET) yielding 75% +GFP cells (**Figure 3C**, left), similar to the JetPEI control (77% +GFP cells). We notice that at a formulation of N/P=10, the Sh backbone exhibits higher viability, yet at N/P=20, higher cell death occurs (**Figures S20-S22**).

An RNP is a complex formed with gRNA and CRISPR-Cas9 protein that targets a precise genome cut in HEK293T cells engineered with a traffic light reporter (TLR) gene, reporting nonhomologous end joining (NHEJ) via an indel frameshift upregulating mCherry expression.<sup>48</sup> We show RNP transfection and toxicity at N/P of 2.5 and 5 in **Figures 3B, S31-S33, S35**. We notice that while the Cys polymers have the strongest RNP binding, the Cap and DiE systems show significantly higher mCherry expression. Interestingly, Lg\_Cap\_0 at N/P 5 displays the highest mCherry expression (**Figure 3C**, right) with ~6% +mCherry (noting that only about 1/3 of the indels cause the +mCherry frameshift and expression).<sup>48</sup> The commercial control, JetPEI, exhibits only 1% +mCherry cells, representing a six-fold decrease from the highest performing polymer in our library. Unlike our findings for pDNA, BET did not improve binding to RNP nor did it correlate to improving mCherry expression (denoting possible intercalation with only DNA). We discover a further deviating trend that higher editing is found with the longest polymers. Further, we notice that cell viability was much higher in this assay due to the lower N/P ratios. Overall, our screening protocols identify promising formulations that are highly effective at delivering pDNA and RNP while balancing toxicity. Collectively, we show that polymer chemistry, physical properties, and payload identity significantly affects performance; indeed, quantitative discernment of the physicochemical drivers of performance is complex and difficult, supporting the need for quantitative multifactorial analysis techniques to more rapidly enable optimization.

## Polymer Feature Attribution Through SHAP Analysis

Shapley Additive ExPlanation<sup>49</sup> is a machine learning technique used to extrapolate predictive importance on a given model variable (or feature) on a particular output. A high and positive SHAP value correlates to high impact and a positive effect on the output variable. We start with our previous approach<sup>23,24</sup> and then fit a machine learning model (details in Supporting Information). To identify polymer and polyplex characteristics, our independent variables of scaffold RU, cation type, %BET,  $pK_a$ , clogP, polyplex size ( $R_h$ ), formulation (N/P) ratio, and binding strength are modeled to their effect on dependent variables of expression and viability. Based on our cross-validated model, we analyze SHAP values for both pDNA (**Figure 4A**) and RNP (**Figure 4B**) payloads. We find that for pDNA; polymer  $pK_a$ , BET incorporation, and scaffold RU have the highest absolute SHAP values across samples, which translate into a positive impact on GFP expression. We compare the average SHAP values in a spider plot (**Figure 4A**) and show the level of impact each polymer feature has on GFP expression and cell viability for pDNA polyplexes. We show that lowering the  $pK_a$  has a positive impact on GFP expression, likely due to a higher buffering capacity promoting endosomal escape.<sup>43,44</sup> It should be noted that the decrease in the polymer  $pK_a$  is a result of an increase in BET incorporation, where the co-cation aids in initial binding at physiological pH. We plot the SHAP values as a function of a given feature (SHAP dependency plots) for pDNA expression and find a linear dependence correlating lower  $pK_a$ , increasing BET incorporation

(**Figure 4C**: left), and increasing clogP value (**Figure 4C**: middle) to higher SHAP values. Interestingly, when we compare the polymer length, the Md length was the least effective. Higher percentages of BET incorporation positively correlated to GFP expression and cell viability. Often, higher incorporations of hydrophobic units can contribute to increased cellular internalization through membrane disruption but this often causes higher toxicity,<sup>23,38,39,50</sup> which is not found with our system. For pDNA delivery, we correlate the BET incorporation and improved performance to lowering polymer  $pK_a$  (5.90).

For RNP delivery, the mean SHAP values are shown in a spider plot (**Figure 4B**), detailing the impact of each polymer feature on mCherry expression and cell viability. We find a positive correlation between polymer length and increase in mCherry expression. We find that the Cys cation has the least importance for both pDNA and RNP delivery while, RNP delivery favors the Cap cation in the copolymer composition. Interestingly, with RNP we find that weaker binding polymers yield higher correlations to mCherry expression. We notice on the SHAP dependency plot that RNP delivery efficiency is negatively correlated to binding and %BET, an opposite trend to that found with pDNA (**Figure 4C**: left), demonstrating the intricacies of tuning polymer chemistry for specific biological payloads. Overall, we find that SHAP analysis quantifies and identifies predictive correlations and fundamental insight into the physicochemical components most influential for delivery of pDNA versus RNP. We show that lower  $pK_a$  in conjunction with higher BET incorporation in the Sh and Lg scaffold variants are important for pDNA delivery. However, for gene editing with RNP, the Lg polymers along with Cap and lower BET incorporation are most influential.

## Batch Bayesian Optimization

To efficiently explore the combinatorial design space and its impact on expression performance outputs, we use batch BO. According to this machine learning technique, we learn a probabilistic model “the optimization model” (**Figure 5A**), which relates design variables to our target delivery efficacy and predicts outcomes of discrete polymer formulations. Then, based on the predicted mean expression and variability of the target variable in the design space, we select a new batch, which we call a BO round, of promising polymers to formulate and measure transfection efficacy. Our Round 1 dataset is defined by sampling the entire polymer library at two N/P ratios, in triplicate, for each the pDNA and RNP payloads, resulting in sampling 432 formulations (216 per payload). We complete two additional rounds of transfections with the most promising polymer formulation subsets defined by the batch BO model. Rather than uniform sampling across the entire library for optimization, machine learning beneficially limits the number of experimental sample conditions required by predicting the most promising polymer formulations to sample for Round 2. This consists of 72 formulation ratios (36 per payload) for the model to improve predictive sampling. Round 3 predicts 48 more discrete sample formulations (24 per payload) to improve the mean expression output, while hovering above a certain viability threshold (viability  $\geq 0.30$ ). After three optimization rounds, we did not observe further improvement in the expression and we deemed our sequential optimization concluded.

The progression of **Figure 5B** presents our expression outputs for pDNA and RNP cargos respectively. Each box plot corresponds to the measured outputs after transfection. For pDNA (**Figure 5B**, top), our model suggests a subset of polymer formulations that results in the overall best polymer for the dataset in Round 2. The subset of polymer formulations chosen in Round 3 was able to increase in overall expression, however no more new optimal formulations were found. For RNP, our model explores a subset of promising polymers in Rounds 2 and 3, which outperform most of the initial polymers, but was similar to the performance collected in Round 1. This is likely due to the degree of error in the mCherry expression to noise ratio in the assay making it a bit more difficult to refine. The parallel coordinate plots (**Figure 5C**) summarize the progression of the sequential optimization to identify subsets of polymers that result in the highest expression. For pDNA (**Figure 5C**, top), our model directs experimental sampling towards the cations of Cap and DiE, higher BET incorporations, and intermediate N/P ratios than the initial screen of 10 and 20, with the top three performing polymers being Sh\_DiE\_40 (N/P 12.5), Sh\_Cap\_40 (N/P 10), and Lg\_Cap\_25 (N/P 12.5). For RNP (**Figure 5C**, bottom), our model suggests sampling cations with

Cap and DiE, lower BET incorporations, and polymers derived from the larger scaffolds, with the top three performing polymers being Lg\_Cap\_0 (N/P 5), Lg\_Cap\_15 (N/P 8.6), and Sh\_DiE\_0 (N/P 10). The trends we outline from the parallel coordinate plots align well with the SHAP analysis of the most distinct correlative features, displaying that our model is properly identifying the most important physiochemical variables for prediction of future polymer formulations. Overall, we reveal a two-fold benefit of our customized probabilistic BO model: i) BO allows for narrowing the amount of sampling and experimental time/expense from vast possibilities, and ii) BO identifies optimal chemical features and formulation ratios for given constructs unique to a biological payload of choice.

### ***In vivo* Delivery of pDNA**

While polymeric vehicles are effective *in vitro*, the largest hurdle is *in vivo* performance. Predicted by SHAP and BO models, we select our three top performers, Sh\_DiE\_40, Sh\_Cap\_40, Md\_Cys\_40 for pDNA delivery *in vivo* (N/P=5 to minimize toxicity). Our polyplexes and controls (pDNA only and JetPEI) were administered to mice via hydrodynamic tail vein injections in triplicate and expression was compared over a 20-day span (**Figure 6A-C**). The pDNA only control commonly promotes luciferase expression through hydrodynamic injections due to the large injection volume (5-10 wt%) over a small time period (4-8 s), inducing a pressure plug localizing to the liver (but generally diminishes rapidly).<sup>51-53</sup> While these injections can induce stress to the mice, we find that all mice survive the injections with stable body weight over the 20-day span, indicating formulation tolerance (**Figure 6D**). We show that mice injected with JetPEI polyplexes have the greatest loss of weight but recover after ~6-days. The highest performing polymers yield a smaller dip in weight over the first 48 h, and an increase in weight was found prior to injection at 96 h, which remain stable over the 20-day span. While pDNA initially showed luciferase expression it rapidly decays, 1.7-fold faster than Md\_Cys\_40 and 1.6-fold faster than Sh\_DiE\_40 over a 6-day period (**Figure 6B**). All three of our optimized polymers show a slower decay rate and outperformed both controls. Polyplex formulations with Sh\_DiE\_40 and Sh\_Cap\_40 outperform the controls at all timepoints, however, Md\_Cys\_40 polyplexes decay in expression at a faster rate between

day 10 and 20, dipping below JetPEI. Sh\_DiE\_40 show the highest radiance throughout, outperforming all other formulations. Our data reveal the discovery of polymer formulations that can bind, protect, and deliver pDNA *in vivo* that show stable and long-term expression and stability over excretion over 20 days in the mouse liver.

## Conclusion

Here, we demonstrate a streamlined synthetic method via a facile post-polymerization modification yielding a polymer library systematically exploring chemical composition and physical parameters. Parallel assays combined with SHAP analysis and BO through progressive sampling allow modeling and quantitative understanding of features important to increasing effective transient expression of pDNA as well as gene editing through the delivery of CRISPR-Cas9 RNP. Features of lower polymer  $pK_a$  and higher %BET increase pDNA delivery, while, polymer length and Cap cation identity are more effective for RNP delivery. Additionally, our three top performing copolymers selected by SHAP and BO display high expression *in vivo* with a 1.7-fold kinetic enhancement of transgene expression over controls. Overall, facile tunable synthesis combined with screening and machine learning are powerful tools toward a data-driven materials discovery platform to identify candidates for *in vivo* screening and will aid selection for clinical nucleic acid therapeutic delivery.

## Declarations

### Acknowledgments

We acknowledge Nanite, Inc. for funding of this work. The resources and staff (Dr. Guillermo Marques) at the University of Minnesota University Imaging Centers (UIC, SCR\_020997) supported this work. We would also like to acknowledge Craig Flory for aid in the *in vivo* transfections and the technical assistance of Brenda Koniar, Joshua McCarra, Lia Coicou, Victoria Hoehn, and Kira Rolf from the Center for Translational Medicine at the University of Minnesota. We acknowledge Ramya Kumar, Leon Lillie, Derek Saxon and Cristiam Santa Chalarca for technical advice. Figures were partially made on Biorender.com.

### Author Contributions

Rishad J. Dalal designed, performed and analyzed experiments along with wrote the manuscript. Felipe Oviedo performed the machine learning analysis of SHAP and Bayesian Optimization. Michael C. Leyden performed DLS studies on the RNP complexes. Theresa M. Reineke helped design and supervised the research and helped write the manuscript.

### Conflict of Interest Statement

The authors declare the following competing financial interest(s): Theresa M. Reineke is one of the founders of Nanite, Inc. and has an equity interest. Nanite, Inc. is one of the sponsors of this research.



This interest has been reviewed and managed by the University of Minnesota in accordance with its conflict-of-interest policy. Felipe Oviedo consults for Nanite, Inc. and has an equity interest.

## Supporting Information

In addition to the methods section, materials, characterization through NMR, SEC-MALLS,  $pK_a$  titrations, DLS, dye exclusion, viability, supplemental data from flow cytometry, machine learning, and *in vivo* experimental setup are available within the Supporting Information.

## Research Animals

Project name: Polymeric Gene-Delivery Luciferase Kinetics in C57BL/6 Mice

Study name: 22-01-07-POLYMER-HDIV-IV-PV-LUCIFERASE KINETICS

Institutional Animal Care and Use Committee Protocol Number, PI: 2001-37736A (Schumacher, Robert)

Purpose: A kinetic study of polymeric gene-delivery and luciferase expression in the liver of mice after hydrodynamic, slow intravenous, and portal vein injection of vector polymers in C57BL/6 female mice.

C57BL/6NHsd female mice, strain code 044, from Envigo at 7-8 weeks of age at arrival. Mice were acclimated to the facility for a minimum of 3 days prior to performing any study procedures beyond ear tagging and weighing. As this was a pilot study, only females were used to reduce the number of animals required for the study. Definitive studies would require the use of both sexes to assess the potential for sex-related differences in vector delivery.

### *Ethical Compliance*

We have complied with all relevant ethical regulations regarding the care and use of animals in research. All animal studies were carried out under protocols approved by the University of Minnesota's Institutional Animal Care and Use Committee.

### *Ethics Committee*

The University of Minnesota Institutional Animal Care and Use Committee approved the study protocol 11/17/2021.

## Data Availability

The data of this study is intellectual property of Nanite, Inc. The datasets from this study are available from the corresponding author upon reasonable request. For peer-reviewers, we have made the data from this study available in the Supporting Information section and, the data relevant for machine learning modelling, available in our [private repository](#).

## Code Availability

The code used to generate the results in this study is Intellectual Property of Nanite, Inc; and thus, is not publicly available. However, the results may be reproduced using the descriptions in this paper. For peer-reviewers, we have made our code available in our [private repository](#).

## References

- (1) Ramamoorth, M.; Narvekar, A. Non Viral Vectors in Gene Therapy - An Overview. *J. Clin. Diagnostic Res.* **2015**, *9* (1), GE01–GE06. <https://doi.org/10.7860/JCDR/2015/10443.5394>.
- (2) Li, S. D.; Huang, L. Non-Viral Is Superior to Viral Gene Delivery. *J. Control. Release* **2007**, *123* (3), 181–183. <https://doi.org/10.1016/j.jconrel.2007.09.004>.
- (3) Trapani, I.; Tornabene, P.; Auricchio, A. Large Gene Delivery to the Retina with AAV Vectors: Are We There Yet? *Gene Ther.* **2021**, *28* (5), 220–222. <https://doi.org/10.1038/s41434-020-0174-4>.
- (4) May, M. Viral Vector Industry Boosts Product Quality and Quantity. *Genet. Eng. Biotechnol. News* **2020**, *40* (3), 42–44, 46. <https://doi.org/10.1089/gen.40.03.11>.
- (5) Nayak, S.; Herzog, R. W. Progress and Prospects: Immune Responses to Viral Vectors. *Gene Ther.* **2010**, *17* (3), 295–304. <https://doi.org/10.1038/gt.2009.148>.
- (6) Liu, C.; Zhang, L.; Liu, H.; Cheng, K. Delivery Strategies of the CRISPR-Cas9 Gene-Editing System for Therapeutic Applications. *J. Control. Release* **2017**, *266* (September), 17–26. <https://doi.org/10.1016/j.jconrel.2017.09.012>.
- (7) Kumar, R.; Santa Chalarca, C. F.; Bockman, M. R.; Bruggen, C. Van; Grimme, C. J.; Dalal, R. J.; Hanson, M. G.; Hexum, J. K.; Reineke, T. M. Polymeric Delivery of Therapeutic Nucleic Acids. *Chem. Rev.* **2021**, *121* (18), 11527–11652. <https://doi.org/10.1021/acs.chemrev.0c00997>.
- (8) Van Bruggen, C.; Hexum, J. K.; Tan, Z.; Dalal, R. J.; Reineke, T. M. Nonviral Gene Delivery with Cationic Glycopolymers. *Acc. Chem. Res.* **2019**. <https://doi.org/10.1021/acs.accounts.8b00665>.
- (9) Perrier, S. 50th Anniversary Perspective: RAFT Polymerization - A User Guide. *Macromolecules* **2017**, *50* (19), 7433–7447. <https://doi.org/10.1021/acs.macromol.7b00767>.
- (10) Chiefari, J.; Chong, Y. K. B.; Ercole, F.; Krstina, J.; Jeffery, J.; Le, T. P. T.; Mayadunne, R. T. A.; Meijs, G. F.; Moad, C. L.; Moad, G.; Rizzardo, E.; Thang, S. H.; South, C. Living Free-Radical Polymerization by Reversible Addition - Fragmentation Chain Transfer: The RAFT Process. *Macromolecules* **1998**, *31* (18), 5559–5562.
- (11) Moad, G. RAFT Polymerization – Then and Now. **2015**, 211–246. <https://doi.org/10.1021/bk-2015-1187.ch012>.

- (12) Grubbs, R. B. Nitroxide-Mediated Radical Polymerization: Limitations and Versatility. *Polym. Rev.* **2011**, *51* (2), 104–137. <https://doi.org/10.1080/15583724.2011.566405>.
- (13) Hawker, C. J.; Bosman, A. W.; Harth, E. New Polymer Synthesis by Nitroxide Mediated Living Radical Polymerizations. *Chem. Rev.* **2001**, *101* (12), 3661–3688. <https://doi.org/10.1021/cr990119u>.
- (14) Matyjaszewski, K. Atom Transfer Radical Polymerization (ATRP): Current Status and Future Perspectives. *Macromolecules* **2012**, *45* (10), 4015–4039. <https://doi.org/10.1021/ma3001719>.
- (15) Matyjaszewski, K.; Xia, J. Atom Transfer Radical Polymerization. *Chem. Rev.* **2001**, *101* (9), 2921–2990. <https://doi.org/10.1021/cr940534g>.
- (16) Gibson, M. I.; Hlich, E. F.; Klok, H.-A. Postpolymerization Modification of Poly(Pentafluorophenyl Methacrylate): Synthesis of a Diverse Water-Soluble Polymer Library. *Am. Chem. Soc. Polym. Prepr. Div. Polym. Chem.* **2008**, *49* (1), 511–512. <https://doi.org/10.1002/pola>.
- (17) Santa Chalarca, C. F.; Dalal, R. J.; Chapa, A.; Hanson, M. G.; Reineke, T. M. Cation Bulk and p K a Modulate Diblock Polymer Micelle Binding to PDNA. *ACS Macro Lett.* **2022**, 588–594. <https://doi.org/10.1021/acsmacrolett.2c00015>.
- (18) Gauthier, M. A.; Gibson, M. I.; Klok, H.-A. Synthesis of Functional Polymers by Post-Polymerization Modification. *Angew. Chemie Int. Ed.* **2009**, *48* (1), 48–58. <https://doi.org/10.1002/anie.200801951>.
- (19) Akinc, A.; Lynn, D. M.; Anderson, D. G.; Langer, R. Parallel Synthesis and Biophysical Characterization of a Degradable Polymer Library for Gene Delivery. *J. Am. Chem. Soc.* **2003**, *125* (18), 5316–5323. <https://doi.org/10.1021/ja034429c>.
- (20) Barua, S.; Joshi, A.; Banerjee, A.; Matthews, D.; Sharfstein, S. T.; Cramer, S. M.; Kane, R. S.; Rege, K. Parallel Synthesis and Screening of Polymers for Nonviral Gene Delivery. *Mol. Pharm.* **2009**, *6* (1), 86–97. <https://doi.org/10.1021/mp800151j>.
- (21) Goldberg, M.; Mahon, K.; Anderson, D. Combinatorial and Rational Approaches to Polymer Synthesis for Medicine. *Adv. Drug Deliv. Rev.* **2008**, *60* (9), 971–978. <https://doi.org/10.1016/j.addr.2008.02.005>.
- (22) Anderson, D. G.; Akinc, A.; Hossain, N.; Langer, R. Structure/Property Studies of Polymeric Gene Delivery Using a Library of Poly( $\beta$ -Amino Esters). *Mol. Ther.* **2005**, *11* (3), 426–434. <https://doi.org/10.1016/j.ymthe.2004.11.015>.
- (23) Kumar, R.; Le, N.; Oviedo, F.; Brown, M. E.; Reineke, T. M. Combinatorial Polycation Synthesis and Causal Machine Learning Reveal Divergent Polymer Design Rules for Effective PDNA and Ribonucleoprotein Delivery. *JACS Au* **2022**, *2* (2), 428–442. <https://doi.org/10.1021/jacsau.1c00467>.

- (24) Kumar, R.; Le, N.; Tan, Z.; Brown, M. E.; Jiang, S.; Reineke, T. M. Efficient Polymer-Mediated Delivery of Gene-Editing Ribonucleoprotein Payloads through Combinatorial Design, Parallelized Experimentation, and Machine Learning. *ACS Nano* **2020**, *14* (12), 17626–17639. <https://doi.org/10.1021/acsnano.0c08549>.
- (25) Patra, T. K. Data-Driven Methods for Accelerating Polymer Design. *ACS Polym. Au* **2022**, *2* (1), 8–26. <https://doi.org/10.1021/acspolymersau.1c00035>.
- (26) Batra, R.; Song, L.; Ramprasad, R. Emerging Materials Intelligence Ecosystems Propelled by Machine Learning. *Nat. Rev. Mater.* **2021**, *6* (8), 655–678. <https://doi.org/10.1038/s41578-020-00255-y>.
- (27) Upadhyaya, 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. <https://doi.org/10.1016/j.addr.2020.11.009>.
- (28) Gormley, A. J.; Webb, M. A. Machine Learning in Combinatorial Polymer Chemistry. *Nat. Rev. Mater.* **2021**, *6* (8), 642–644. <https://doi.org/10.1038/s41578-021-00282-3>.
- (29) Ramprasad, R.; Batra, R.; Pilania, G.; Mannodi-Kanakkithodi, A.; Kim, C. Machine Learning and Materials Informatics: Recent Applications and Prospects. *npj Comput. Mater.* **2017**, *3* (1), 54. <https://doi.org/10.1038/s41524-017-0056-5>.
- (30) Gómez-Bombarelli, R.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Duvenaud, D.; Maclaurin, D.; Blood-Forsythe, M. A.; Chae, H. S.; Einzinger, M.; Ha, D.-G.; Wu, T.; Markopoulos, G.; Jeon, S.; Kang, H.; Miyazaki, H.; Numata, M.; Kim, S.; Huang, W.; Hong, S. I.; Baldo, M.; Adams, R. P.; Aspuru-Guzik, A. Design of Efficient Molecular Organic Light-Emitting Diodes by a High-Throughput Virtual Screening and Experimental Approach. *Nat. Mater.* **2016**, *15* (10), 1120–1127. <https://doi.org/10.1038/nmat4717>.
- (31) Kumar, J. N.; Li, Q.; Tang, K. Y. T.; Buonassisi, T.; Gonzalez-Oyarce, A. L.; Ye, J. Machine Learning Enables Polymer Cloud-Point Engineering via Inverse Design. *npj Comput. Mater.* **2019**, *5* (1), 73. <https://doi.org/10.1038/s41524-019-0209-9>.
- (32) Wang, Y.; Xie, T.; France-Lanord, A.; Berkley, A.; Johnson, J. A.; Shao-Horn, Y.; Grossman, J. C. Toward Designing Highly Conductive Polymer Electrolytes by Machine Learning Assisted Coarse-Grained Molecular Dynamics. *Chem. Mater.* **2020**, *32* (10), 4144–4151. <https://doi.org/10.1021/acs.chemmater.9b04830>.
- (33) 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**, *2201809*, 2201809. <https://doi.org/10.1002/adma.202201809>.
- (34) Das, A.; Theato, P. Multifaceted Synthetic Route to Functional Polyacrylates by Transesterification of Poly(Pentafluorophenyl Acrylates). *Macromolecules* **2015**, *48* (24), 8695–8707.

<https://doi.org/10.1021/acs.macromol.5b02293>.

(35) Günay, K. A.; Schüwer, N.; Klok, H. A. Synthesis and Post-Polymerization Modification of Poly(Pentafluorophenyl Methacrylate) Brushes. *Polym. Chem.* **2012**, *3* (8), 2186–2192.

<https://doi.org/10.1039/c2py20162c>.

(36) Blasco, E.; Sims, M. B.; Goldmann, A. S.; Sumerlin, B. S.; Barner-Kowollik, C. 50th Anniversary Perspective: Polymer Functionalization. *Macromolecules* **2017**, *50* (14), 5215–5252.

<https://doi.org/10.1021/acs.macromol.7b00465>.

(37) Hoyle, C. E.; Bowman, C. N. Thiol-Ene Click Chemistry. *Angew. Chemie - Int. Ed.* **2010**, *49* (9), 1540–1573. <https://doi.org/10.1002/anie.200903924>.

(38) Van Bruggen, C.; Punihaole, D.; Keith, A. R.; Schmitz, A. J.; Tolar, J.; Frontiera, R. R.; Reineke, T. M. Quinine Copolymer Reporters Promote Efficient Intracellular DNA Delivery and Illuminate a Protein-Induced Unpackaging Mechanism. *Proc. Natl. Acad. Sci.* **2020**, *117* (52), 32919–32928.

<https://doi.org/10.1073/pnas.2016860117>.

(39) Pezzoli, D.; Giupponi, E.; Mantovani, D.; Candiani, G. Size Matters for in Vitro Gene Delivery: Investigating the Relationships among Complexation Protocol, Transfection Medium, Size and Sedimentation. *Sci. Rep.* **2017**, *7* (1), 44134. <https://doi.org/10.1038/srep44134>.

(40) Bhattacharya, S.; Chaudhuri, P. Medical Implications of Benzimidazole Derivatives as Drugs Designed for Targeting DNA and DNA Associated Processes. *Curr. Med. Chem.* **2008**, *15* (18), 1762–1777. <https://doi.org/10.2174/092986708785133013>.

(41) Kubota, Y.; Iwamoto, T.; Seki, T. The Interaction of Benzimidazole Compounds with DNA: Intercalation and Groove Binding Modes. *Nucleic Acids Symp. Ser.* **1999**, *42* (1), 53–54. <https://doi.org/10.1093/nass/42.1.53>.

(42) Shi, J.; Schellinger, J. G.; Johnson, R. N.; Choi, J. L.; Chou, B.; Anghel, E. L.; Pun, S. H. Influence of Histidine Incorporation on Buffer Capacity and Gene Transfection Efficiency of HPMA- Co -Oligolysine Brush Polymers. *Biomacromolecules* **2013**, *14* (6), 1961–1970. <https://doi.org/10.1021/bm400342f>.

(43) Nelson, A. M.; Pekkanen, A. M.; Forsythe, N. L.; Herlihy, J. H.; Zhang, M.; Long, T. E. Synthesis of Water-Soluble Imidazolium Polyesters as Potential Nonviral Gene Delivery Vehicles. *Biomacromolecules* **2017**, *18* (1), 68–76. <https://doi.org/10.1021/acs.biomac.6b01316>.

(44) Miyata, K.; Oba, M.; Nakanishi, M.; Fukushima, S.; Yamasaki, Y.; Koyama, H.; Nishiyama, N.; Kataoka, K. Polyplexes from Poly(Aspartamide) Bearing 1,2-Diaminoethane Side Chains Induce PH-Selective, Endosomal Membrane Destabilization with Amplified Transfection and Negligible Cytotoxicity. *J. Am. Chem. Soc.* **2008**, *130* (48), 16287–16294. <https://doi.org/10.1021/ja804561g>.

- (45) Sprouse, D.; Reineke, T. M. Investigating the Effects of Block versus Statistical Glycopolycations Containing Primary and Tertiary Amines for Plasmid DNA Delivery. *Biomacromolecules* **2014**, *15* (7), 2616–2628. <https://doi.org/10.1021/bm5004527>.
- (46) Dalal, R. J.; Kumar, R.; Ohnsorg, M.; Brown, M.; Reineke, T. M. Cationic Bottlebrush Polymers Outperform Linear Polycation Analogues for PDNA Delivery and Gene Expression. *ACS Macro Lett.* **2021**, *10* (7), 886–893. <https://doi.org/10.1021/acsmacrolett.1c00335>.
- (47) McLendon, P. M.; Fichter, K. M.; Reineke, T. M. Poly(Glycoamidoamine) Vehicles Promote PDNA Uptake through Multiple Routes and Efficient Gene Expression via Caveolae-Mediated Endocytosis. *Mol. Pharm.* **2010**, *7* (3), 738–750. <https://doi.org/10.1021/mp900282e>.
- (48) Certo, M. T.; Ryu, B. Y.; Annis, J. E.; Garibov, M.; Jarjour, J.; Rawlings, D. J.; Scharenberg, A. M. Tracking Genome Engineering Outcome at Individual DNA Breakpoints. *Nat. Methods* **2011**, *8* (8), 671–676. <https://doi.org/10.1038/nmeth.1648>.
- (49) Sayres, R.; Taly, A.; Rahimy, E.; Blumer, K.; Coz, D.; Hammel, N.; Krause, J.; Narayanaswamy, A.; Rastegar, Z.; Wu, D.; Xu, S.; Barb, S.; Joseph, A.; Shumski, M.; Smith, J.; Sood, A. B.; Corrado, G. S.; Peng, L.; Webster, D. R. A Unified Approach to Interpreting Model Predictions. *Proc. 31st Int. Conf. Neural Inf. Process. Syst.* **2017**, pp 4768–4777. <https://doi.org/10.1016/j.optha.2018.11.016>.
- (50) Tan, Z.; Jiang, Y.; Ganewatta, M. S.; Kumar, R.; Keith, A.; Twaroski, K.; Pengo, T.; Tolar, J.; Lodge, T. P.; Reineke, T. M. Block Polymer Micelles Enable CRISPR/Cas9 Ribonucleoprotein Delivery: Physicochemical Properties Affect Packaging Mechanisms and Gene Editing Efficiency. *Macromolecules* **2019**, *52* (21), 8197–8206. <https://doi.org/10.1021/acs.macromol.9b01645>.
- (51) Tolstyka, Z. P.; Phillips, H.; Cortez, M.; Wu, Y.; Ingle, N.; Bell, J. B.; Hackett, P. B.; Reineke, T. M. Trehalose-Based Block Copolycations Promote Polyplex Stabilization for Lyophilization and in Vivo PDNA Delivery. *ACS Biomater. Sci. Eng.* **2016**, *2* (1), 43–55. <https://doi.org/10.1021/acsbiomaterials.5b00312>.
- (52) Podetz-Pedersen, K. M.; Bell, J. B.; Steele, T. W. J.; Wilber, A.; Shier, W. T.; Belur, L. R.; McIvor, R. S.; Hackett, P. B. Gene Expression in Lung and Liver After Intravenous Infusion of Polyethylenimine Complexes of Sleeping Beauty Transposons. *Hum. Gene Ther.* **2010**, *21* (2), 210–220. <https://doi.org/10.1089/hum.2009.128>.
- (53) Bell, J. B.; Podetz-Pedersen, K. M.; Aronovich, E. L.; Belur, L. R.; McIvor, R. S.; Hackett, P. B. Preferential Delivery of the Sleeping Beauty Transposon System to Livers of Mice by Hydrodynamic Injection. *Nat. Protoc.* **2007**, *2* (12), 3153–3165. <https://doi.org/10.1038/nprot.2007.471>.

## Methods

More detailed methods are available in the Supporting Information

## General Copolymer Functionalization

To produce 9 copolymers starting from one pAMAm scaffold the following steps were completed.

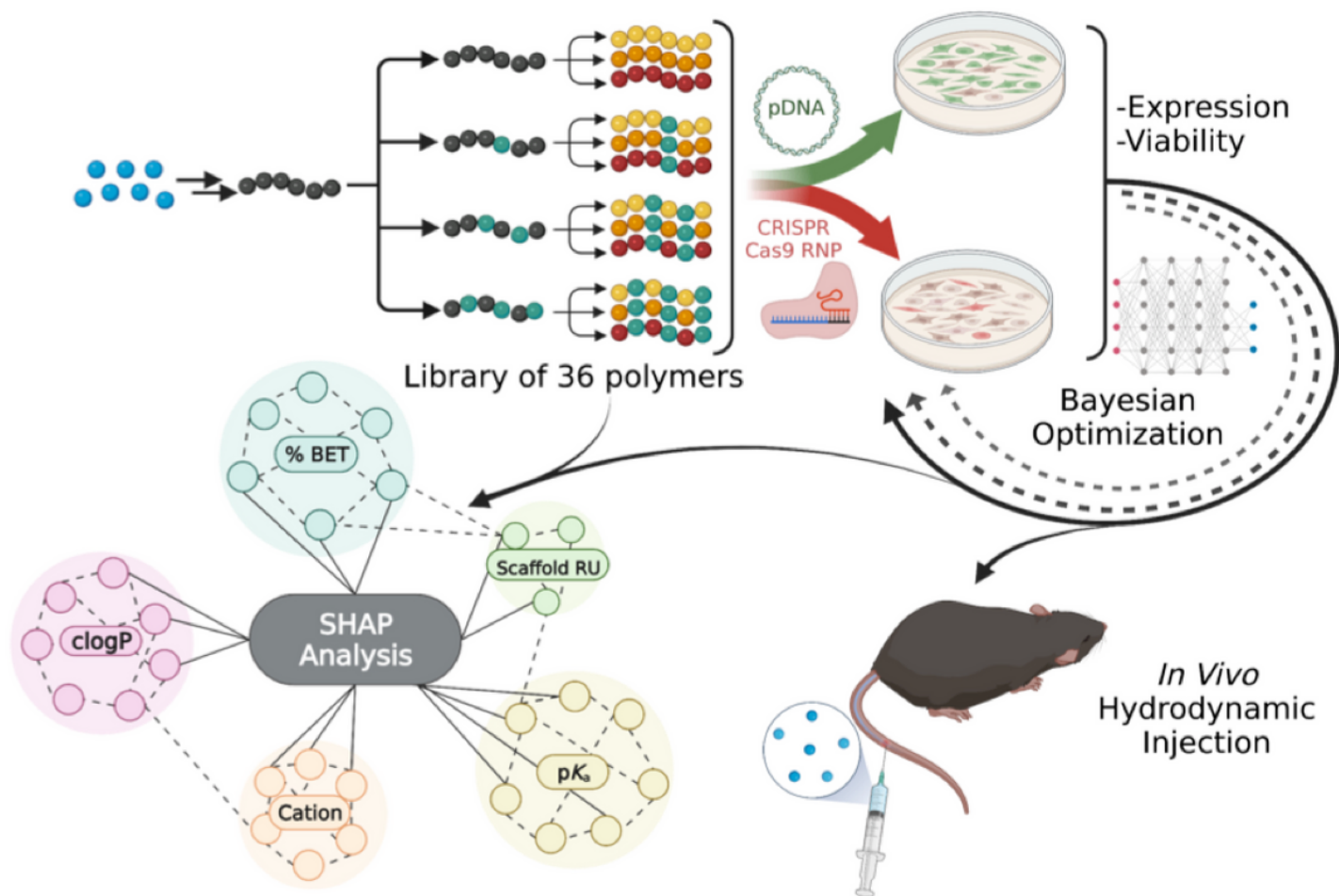
Each different length of pAMAm (RU: 90, 190, 250, 1 equiv., 1.2 mmol relative to monomeric form, 0.15 g) was dissolved in anhydrous DMSO (0.1 M, 12 mL) and mixed in three varying ratios (0.15 equiv., 0.18 mmol), (0.25 equiv., 0.3 mmol), and (0.5 equiv., 0.6 mmol) of BET. To further the reaction in an open capped scintillation vial, DMPA (0.05 equiv.) was added to the solution mixtures and stirred with a stir bar for 1 h at room temperature in a UV (~365 nm) gel nail curing box. The resulting partially BET-functionalized polymers were each split into 3 vials (4 mL each) and reacted first with one of three thiol-amine salts (cysteamine hydrochloride, capteamine hydrochloride or diamino ethanethiol hydrochloride (5 equiv., 2.0 mmol)) and then reacted in a UV nail curing box with DMPA (0.05 equiv.). The resulting polymer solution was acidified with 0.5 mL HCl (1 M) and purified in a 1 kDa dialysis bag in millipore water. The polymer solution was further concentrated via lyophilization to obtain a fluffy off-white powder. Polymers with higher incorporation of BET were fluffier. The mass of each polymer obtained was between 70-90 mg. Copolymer molar mass and dispersity measurements were conducted on H<sub>2</sub>O SEC-MALLS (0.1 M Na<sub>2</sub>SO<sub>4</sub>, 1% acetic acid).

## General pDNA and RNP Polymer Transfection Protocol

HEK293T cells were plated in polystyrene 48-well plates at a density of 50,000 cells/mL in DMEM (10% HI-FBS). After 24 h, polyplexes with pDNA were prepared in H<sub>2</sub>O by adding 80 µL polymer to 80 µL pDNA (0.02 µg/mL) at molar ratios of polymer to get N/P ratios. Polyplexes with RNP were prepared in PBS by adding 80 µL polymer to 80 µL RNP (gRNA (0.02 µg/mL) and spCas9 protein (0.1 µg/mL) complex) at molar ratios of polymer to get N/P ratios. Polyplexes were allowed to form at room temperature for 40 min. Opti-MEM in a 2:1 ratio (320 µL) was added to the polyplexes immediately before addition to cells.

Media was carefully aspirated from the well plate before addition of the polyplex samples. Each polyplex was split into triplicate adding 150 µL to each well. Well plates remained on the bench top at room temperature for 40 min before placing into the 37 °C incubator (5% CO<sub>2</sub>). 4 h after initial transfection, 1 mL of DMEM (10% HI-FBS) was added to each well. Media was further aspirated 24 hours after initial transfection, and fresh DMEM (10% HI-FBS, 1 mL) was added to each well. 48 hours after the initial transfection, the cells were analyzed using CCK-8 and flow cytometry, details are listed below.

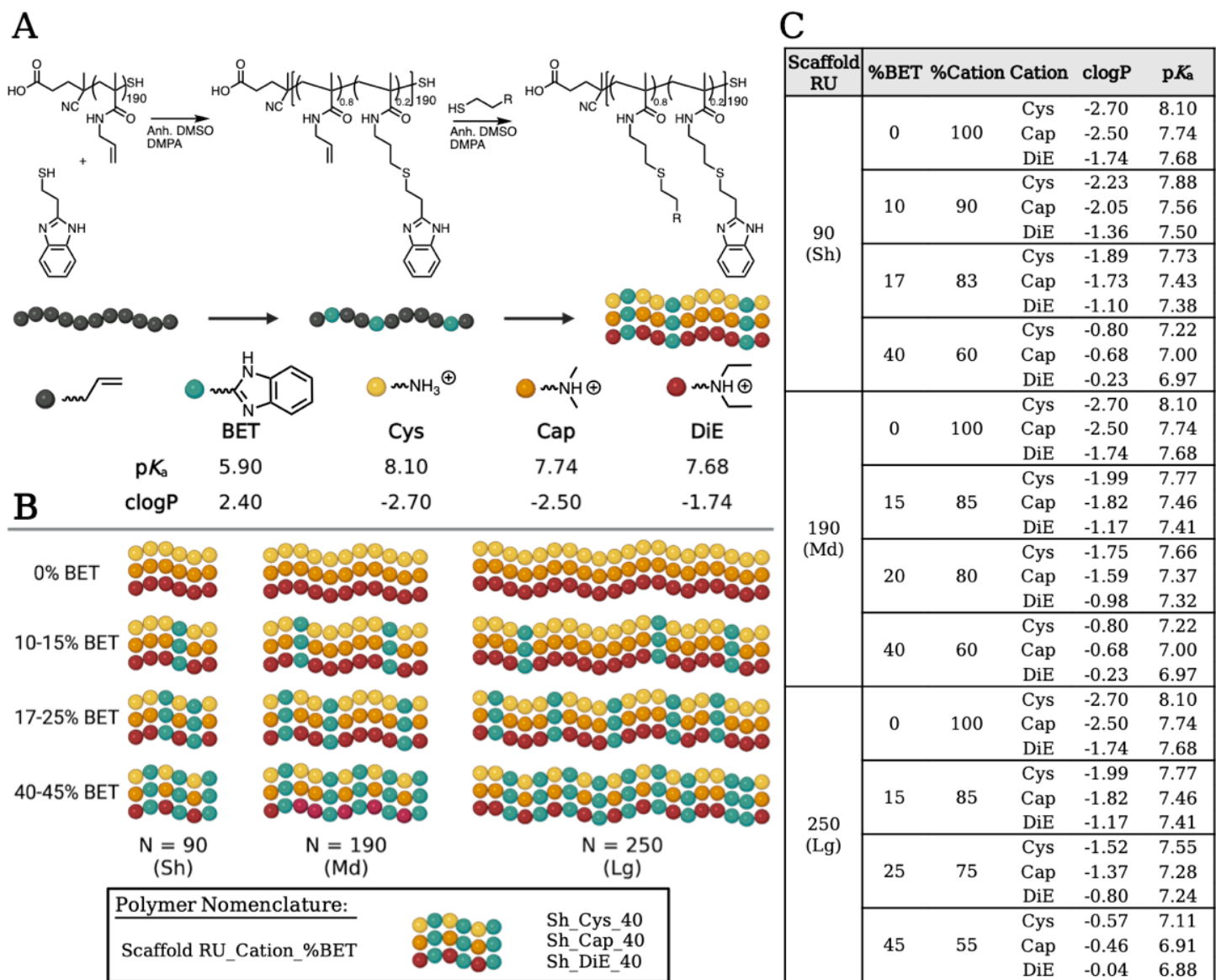
## Figures



**Figure 1**

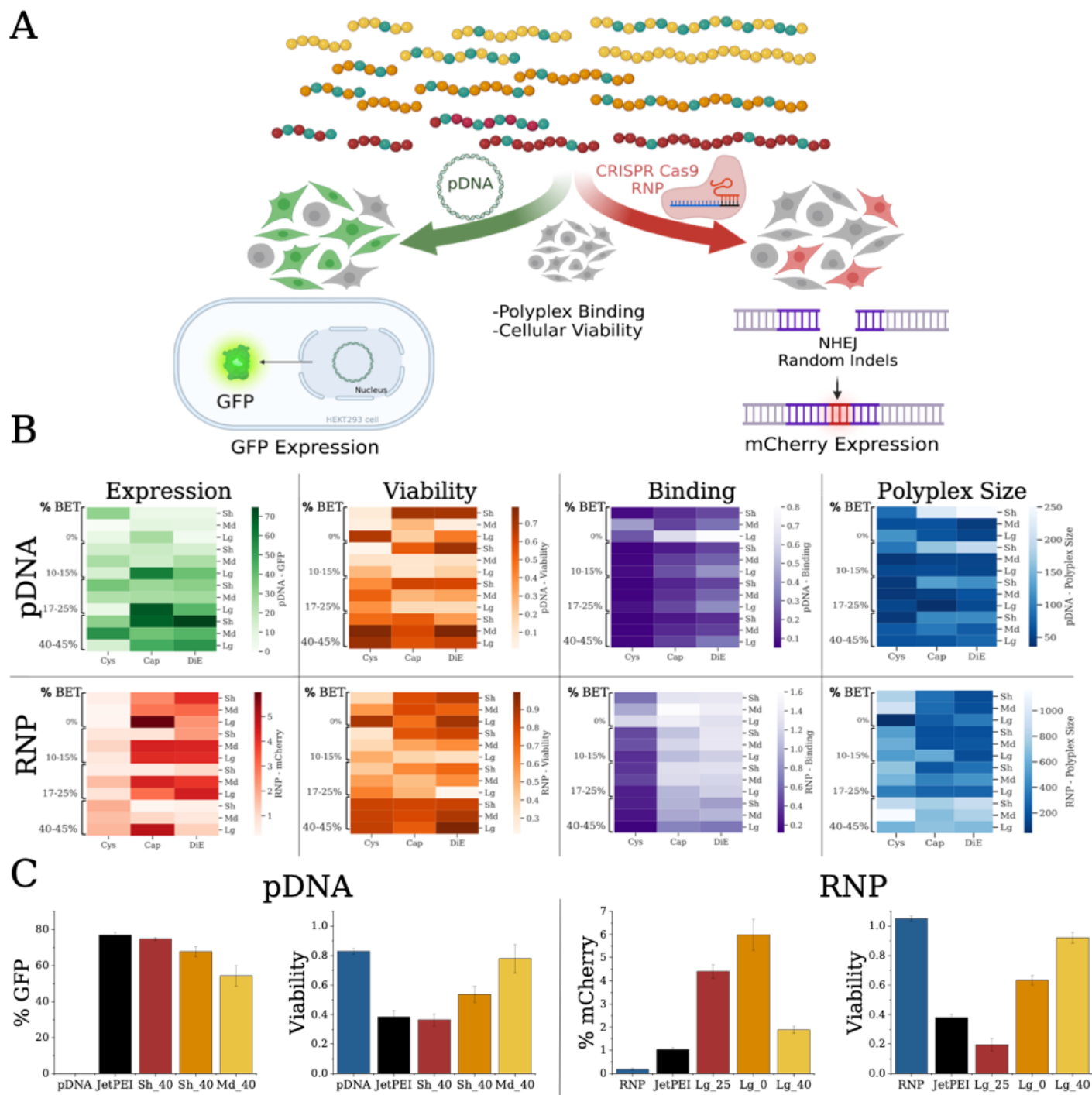
A tunable polymer scaffold for facile optimization of pDNA and RNP delivery performance. The polymers are modified with chemical functionality to modulate binding, delivery, and release. SHAP analysis, a machine learning technique, aids understand structure-activity relationships and identify polymer feature importance on expression and viability. A closed-loop Bayesian optimization further improves performance. Polymers of interest were evaluated for pDNA delivery to the mouse liver.





**Figure 2**

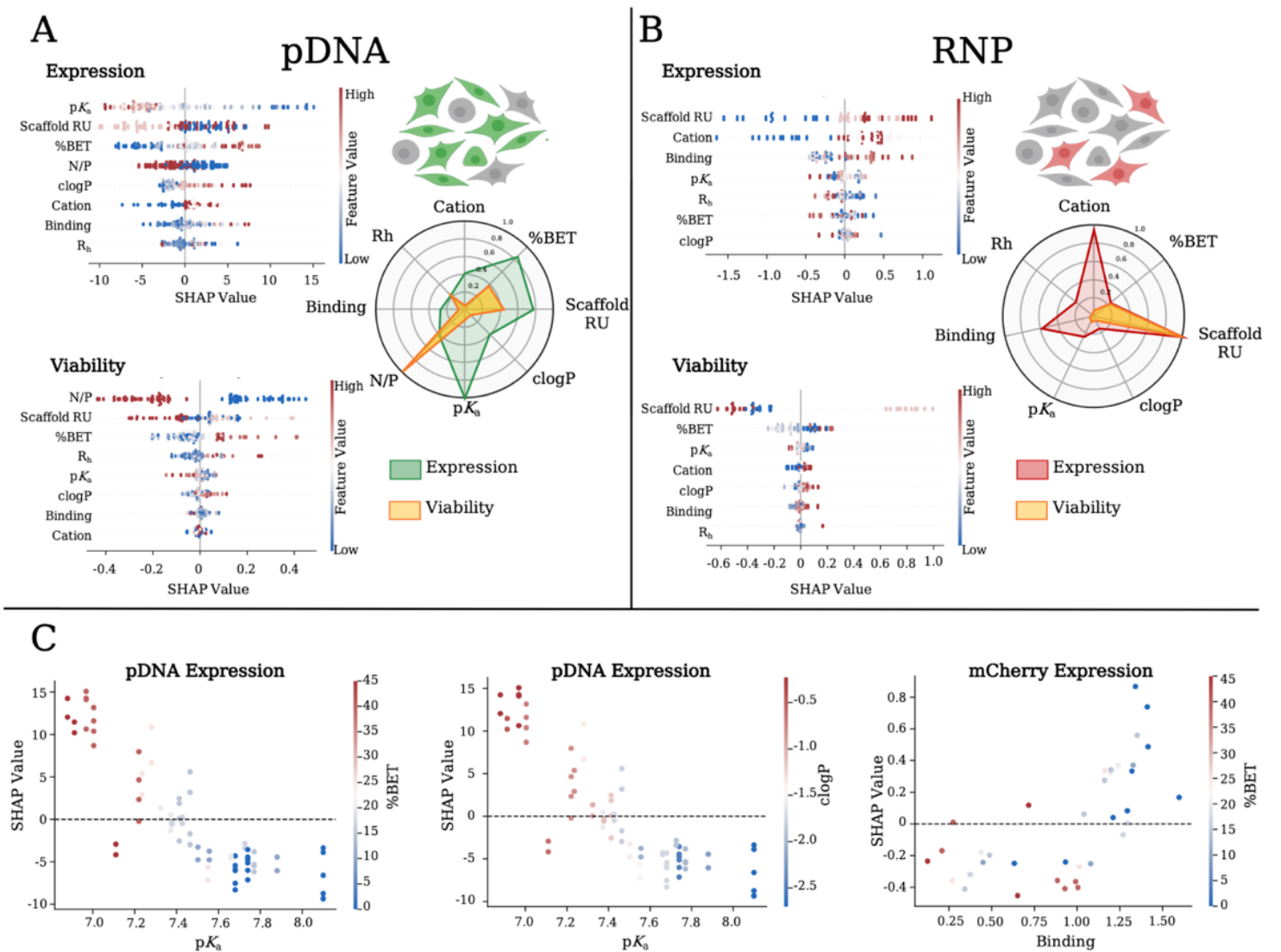
(A) Schematic of the poly(allylmethacrylamide) (pAMAm) polymer scaffold undergoing a stepwise thiol-ene post polymerization modification of the medium backbone with 20% BET and then further split into three to be saturated with the remaining cations (Cys, Cap, DiE). The functional amines have a range of charge state ( $pK_a$ ) and hydrophobicity (clogP) characteristics. (B) A visual representation of the 36 polymers in the library showing the repeat units in the parent backbone ( $N= 90, 190, 250$ ) and the range in BET incorporation within each polymer set. (C) Table of data displaying the chemical characterization of all 36 polymers.



**Figure 3**

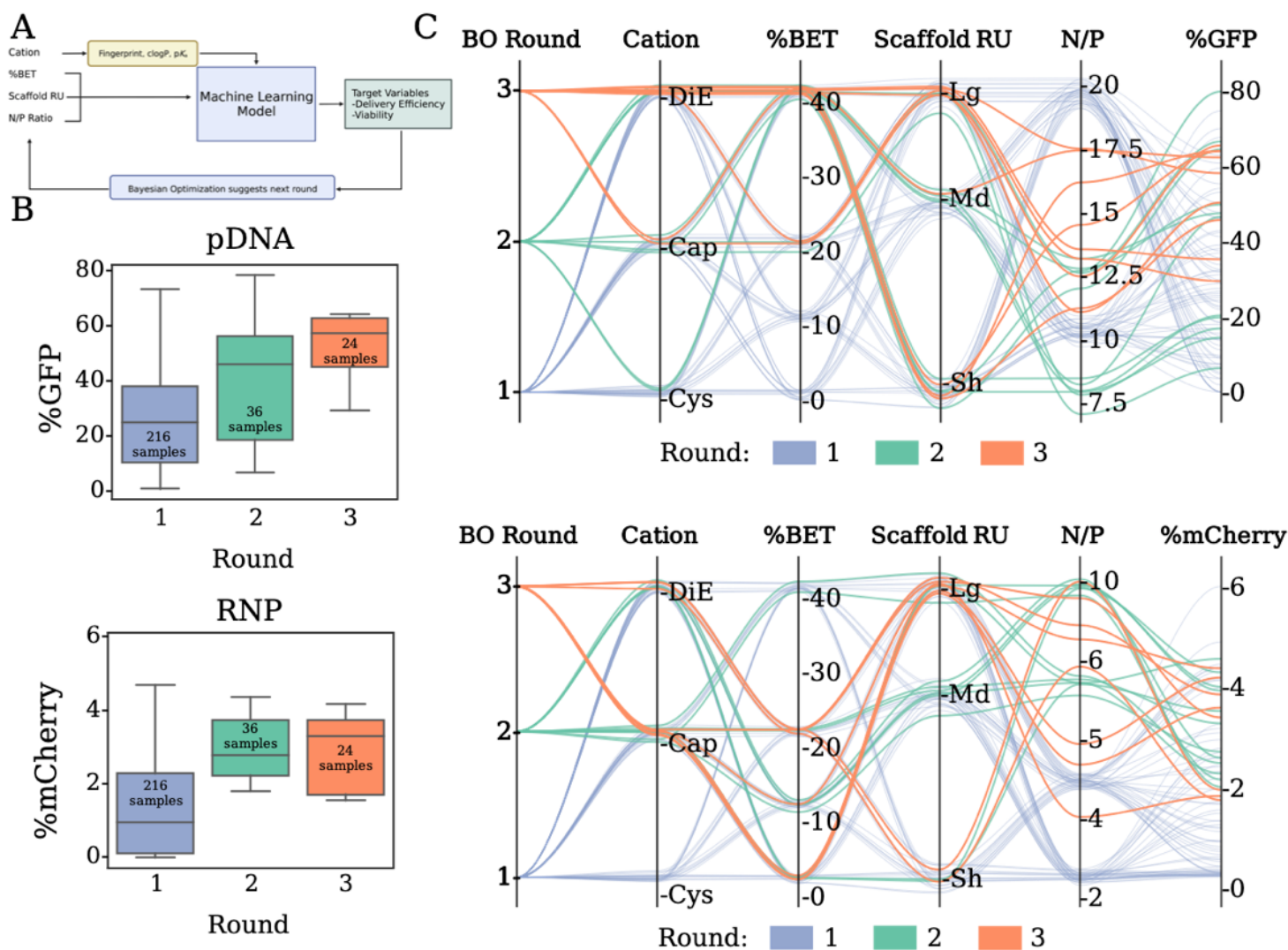
(A) Data schematic showing the parallel characterization workflow on polymer formulations with pDNA and CRISPR RNP displaying relative polymer-payload binding affinity, polyplex size ( $R_n$ ), and *in vitro* assays with HEK293T cells monitoring GFP or mCherry expression. (B) Heat maps of compiled data to visualize binding, polyplex size, % cells with GFP (pDNA) or mCherry (RNP) expression, and toxicity reported as % cell viability. Diagrams show the polymer library with pDNA (top row, N/P 10) and RNP (bottom row, N/P 5). (C) Top expressing polymer of each cation (DiE-red, Cap-orange, Cys-yellow) in

comparison to the payload pDNA or RNP only (blue) and JetPEI polyplex (black) controls. Data also shows the respective viability of each system.



**Figure 4**

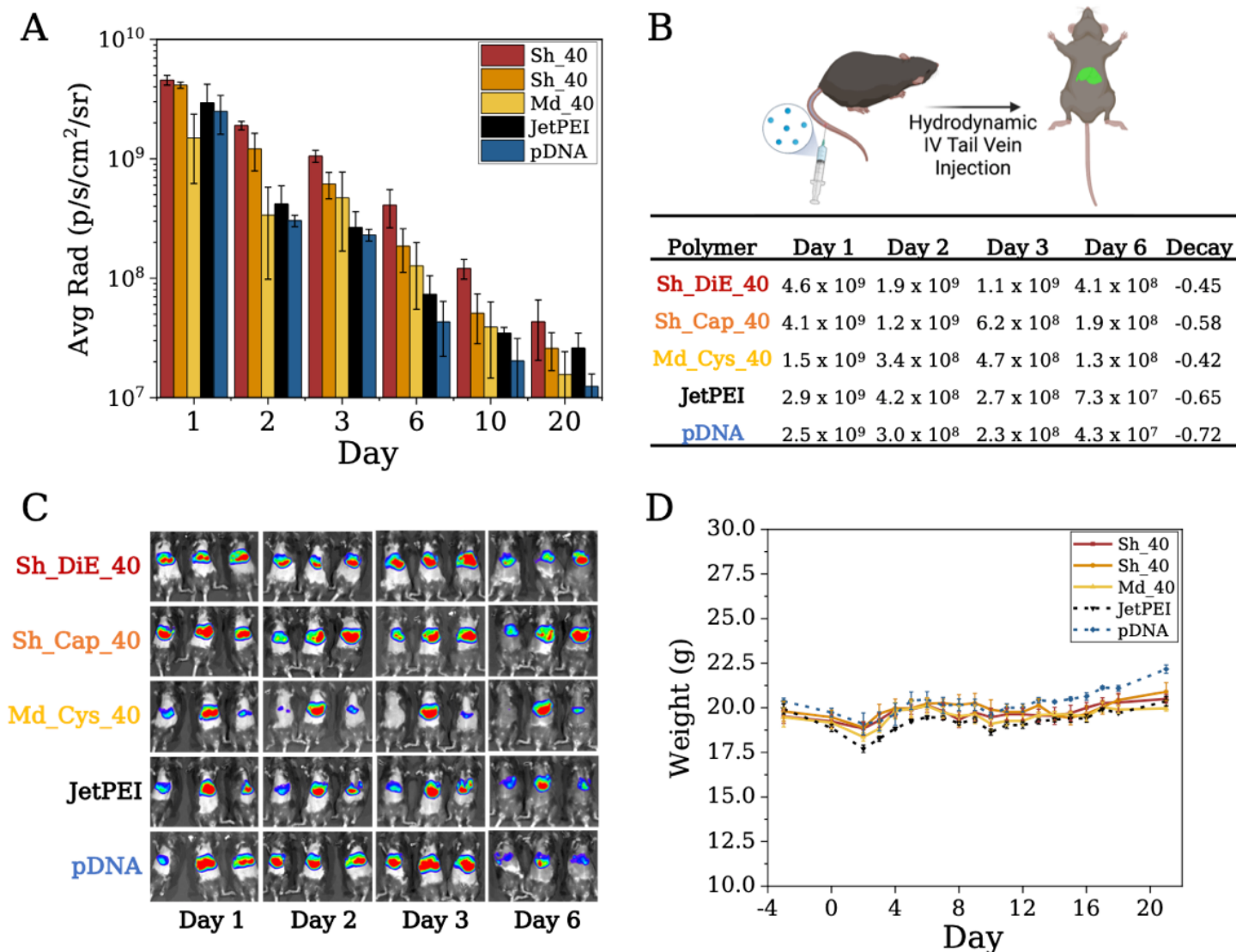
SHAP values for physicochemical features related to expression and cell viability when delivering (A) pDNA or (B) RNP. Higher SHAP values correlate to higher impact on the output variable. The feature value color bar corresponds to the normalized value of the feature of interest (where low = blue; moderate = white; high = red). Each dot represents a polymer formulation. (A) Overlay spider plot showing average impact of individual polymer variables on expression and viability when delivering (A) pDNA and (B) RNP. The spider web plot is constructed by taking the mean SHAP value for a given feature across all samples and normalizing to the maximum SHAP value for each output variable. (C) SHAP dependency plots values across two variables relating to expression.



**Figure 5**

(A) Looped machine learning approach by tuning N/P ratio based on polymer characteristics to effect delivery efficiency and cell viability through Bayesian optimization. (B) Sequential optimization data from the 3 rounds of experimental expression of GFP (top) and mCherry (bottom). (C) Parallel coordinate plots showing the copolymer compositions selected in Rounds 2 and 3 and how it relates to the effective expression of GFP (top) and mCherry (bottom).





**Figure 6**

(A) Kinetic hydrodynamic tail vein study showing the average radiance (p/s/cm<sup>2</sup>/sr) emitted after polyplex administration to mice over a 20-day study after delivery of a luciferase expressing pDNA (n=3). (B) Table displaying average radiance values (p/s/cm<sup>2</sup>/sr) and a decay rate that displayed first-order kinetics of triplicate mice on days 1-6. (C) Images of mice in triplicate showing heat map expression on days 1, 2, 3, and 6 post injection. (D) Average weight of mice per sample group over 22 days. Day 0 is injection day.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [floatimage1.png](#)
- [SIThiolene8TMR20220729.docx](#)