

Machine Learning-Enabled Polymer Discovery for Enhanced Pulmonary siRNA Delivery

Felix Sieber-Schäfer, Min Jiang, Adrian Kromer, Anny Nguyen, Müge Molbay, Simone Pinto Carneiro, David Jürgens, Gerald Burgstaller, Bastian Popper, Benjamin Winkeljann, and Olivia Monika Merkel*

Nucleic acid therapeutics are poised to revolutionize the clinical treatment of diseases once considered undruggable. Although these therapeutic approaches hold significant promise, delivering the nucleic acid cargo remains challenging due to susceptibility to nuclease degradation. Among all carrier systems, polymers stand out for their high tunability and cost-effectiveness. However, their flexible structure greatly expands the chemical space, making experimental exploration both costly and time-consuming. Leveraging published data and machine learning methods provides a valuable strategy to address these issues. The present study demonstrates a way to merge data from multiple sources and uses this information to identify new polyesters that effectively deliver siRNA into lung cells. One newly discovered polymer is further examined in *ex vivo* experiments and tested in a mouse model. The results indicate that a polymer capable of silencing specific genes *in vivo* can be discovered through machine learning, circumventing an extensive trial-and-error process in the search for novel materials.

1. Introduction

Therapeutic nucleic acids (NAs) are one of the most promising innovations in clinical research. A huge number of diseases that were previously considered undruggable, such as hypercholesterolemia^[1] or Huntington's disease^[2] can now be treated effectively through this groundbreaking approach to therapy. Since the discovery of NAs by Friedrich Miescher^[3] in 1868, extensive research has been conducted aiming to translate this technology into actual medicines. It was in 1998, when the first NA-based drug, vitravene, received approval by the U.S. Food & Drug Administration (FDA) for the treatment of cytomegalovirus (CMV) retinitis. However, as of 2024, only 20 further applications have been approved.^[4] One reason for the slow progress may be that NAs,

and particularly ribonucleic acid (RNA) is unstable in the bloodstream and rapidly degraded by ubiquitous RNases. To circumvent this limitation, it became common practice to encapsulate RNA into carrier systems that protect the cargo from enzymatic degradation and help to guide the NAs to the desired tissue. In this context, lipid nanoparticles (LNPs) have become increasingly popular. As of today, three LNP-based RNA therapeutics have received FDA market approval, namely the SARS-CoV-2 vaccines Comirnaty, and Spikevax as well as Onpattro, a therapy for hereditary transthyretin-mediated (hATTR) amyloidosis. However, LNPs have been associated with certain concerns, including their potential to trigger inflammation,^[5] immunogenicity,^[6] and challenges with long-term storage.^[7] Therefore, polymeric carrier systems have been proposed as an alternative to circumvent the limitations of lipid-based carriers.^[8] Polymer materials are generally very flexible and can be modified with optimized chemical structures. This enables simplified adaptation as, in contrast to LNPs, only a single component needs to be adapted. To condense polyanionic RNA via electrostatic interactions, polymers need to contain protonable groups. In many cases, amines are introduced in the polymer structure to fulfil this role.^[9] However, nanoparticles formulated with polycationic materials may cause safety issues if they are unable to be properly excreted, resulting in their accumulation within the body. Polycations remaining in the body can interact non-specifically with intracellular proteins and peptides, which may affect their functionality, and lead to cytotoxicity.

F. Sieber-Schäfer, M. Jiang, A. Kromer, A. Nguyen, M. Molbay, S. Pinto Carneiro, D. Jürgens, B. Winkeljann, O. M. Merkel

Department of Pharmacy

Ludwig-Maximilians-Universität München

Member of the German Center for Lung Research (DZL)

Butenandtstraße 5, 81377 Munich, Germany

E-mail: Olivia.merkel@lmu.de

G. Burgstaller

Helmholtz Munich

Comprehensive Pneumology Center (CPC-M)

Member of the German Center for Lung Research (DZL)

Max-Lebsche-Platz 31, 81377 Munich, Germany

B. Popper

Biomedical Center

Ludwig-Maximilians-Universität München

Großhaderner Str. 9, 82152 Planegg-Martinsried, Germany

B. Winkeljann, O. M. Merkel

Center for Nanoscience (CeNS) Munich

80799 Munich, Germany

O. M. Merkel

Cluster for Nucleic Acid Therapeutics Munich (CNATM)

81377 Munich, Germany

The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/adfm.202502805>

© 2025 The Author(s). Advanced Functional Materials published by Wiley-VCH GmbH. This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/adfm.202502805

This limitation can be addressed by introducing biodegradability into the polymer, which is very common in polyester structures. In addition to biodegradability, the inclusion of specific structural motifs, such as hydrophobic segments, is essential in enhancing the functional properties of the polymer. These hydrophobic motifs^[10] are usually included in forms of amphiphilic block copolymers^[11,12] to shield excessive electrostatic interactions. Additionally, these additions supplement the base polymer with hydrophobic properties for interactions with RNA and biological membranes, which support better performance in cellular uptake, endosomal escape, and many more.

Unfortunately, understanding the exact structure activity relationship between block copolymers and successful delivery of cargo is highly complex and far from trivial. This complexity is amplified by the thousands of potential variations in polymer architecture, composition, and environmental interactions, as well as the fact that synthesizing these polymers is both time-intensive and requires significant material resources, adding to the challenge of systematic exploration. Yet, it is exactly this understanding that is necessary to design new high-performing and safe carrier systems. In recent years, the development and application of artificial intelligence algorithms have significantly increased. These algorithms might help to uncover the underlying patterns differentiating successful from unsuccessful block copolymers and facilitate the virtual screening of potential candidates before synthesis. Machine learning (ML) models that could be used to make this possible, are highly data driven and therefore dependent on available experimental data. While ML is already broadly used for polymeric property predictions such as T_g ^[13] or dielectric constant,^[14] not much work is published on using ML models for the design of new amphiphilic polymeric nanocarriers. Pioneering work in this field was conducted by the groups of Green^[15] and Reineke.^[16] Both used high-throughput synthesis and screening methods to collect data and make predictions for unseen combinations. The need for the availability of high throughput screening opportunities is, however, limiting the wider use of these approaches. Furthermore, the authors relied solely on machine learning applied to a single type of polymer, which inherently limits the exploration of the broader chemical space and restricts the potential to uncover diverse structure-activity relationships.

Here, we show how the discovery of new polymeric nanocarriers can be guided with a prediction model trained on literature data for different kinds of polyesters. In this work, we emphasize pulmonary siRNA delivery to the lungs as a demonstration of our approach, while noting that it could equally be applied to other therapeutic cargo and targets, following a similar strategy. We collected >600 different polyester structures used for siRNA delivery from previous publications and trained multiple ML models with the corresponding gene silencing data. To obtain insights into polymeric siRNA delivery, we investigated key factors that drive successful delivery of cargo. Our lead model was then used to synthesize a novel amphiphilic polymer, which was subsequently tested for its performance of delivering siRNA. Starting with *in vitro* testing we progressively increased biological complexity by evaluating the polymer in an air-liquid-interface model followed by *ex vivo* human Precision-Cut-Lung-Slices (hPCLS). These models reflect critical challenges in pulmonary RNA delivery, including RNase activity, the mucus barrier and tight junc-

tions in respiratory epithelium. Finally, we evaluated the polymer's safety for pulmonary administration and its ability to facilitate gene knockdown in an *in vivo* murine model.

Our approach offers an easy-to-use method for designing new nanocarriers by utilizing historical data. Additionally, we demonstrate how data from a broader chemical space can be used to identify polymeric properties essential for successful delivery. To the best of our knowledge, we are the first to synthesize an amphiphilic polymer for siRNA delivery using ML, thereby contributing to a deeper understanding of RNA delivery via polymeric nanocarriers.

2. Results and Discussion

2.1. Generalizable Machine Learning Framework

A primary goal of this study is to empower researchers lacking HTS capabilities to employ ML on existing literature data. Our methodology achieves this by systematically integrating information from diverse sources into a unified dataset. However, compiling data from literature presents an inherent challenge: balancing the scope of chemical diversity. On the one hand, sufficient diversity is desirable for training models that yield generalizable insights into structure-property relationships. On the other hand, literature datasets are often sparse compared to HTS data. Including systems with widely divergent chemical structures or fundamentally different delivery mechanisms introduces significant noise. With limited data points, this can easily overwhelm the underlying patterns related to a specific delivery strategy, preventing the ML model from effectively learning the relevant mechanisms. Therefore, our approach necessitates carefully constraining the literature search to a comparable chemical space—focusing on systems sharing core structural similarities and presumed mechanisms. This focused scope enhances the signal-to-noise ratio, enabling the model to identify meaningful correlations even from limited data. We illustrate this methodology using a curated dataset of amphiphilic polyester structures, representing a class with comparable underlying chemistry.

Converting molecular structures into a format readable for a ML algorithm is a prerequisite for ML applications in the chemical space, and several methodologies have been proposed.^[17–19] Commonly employed fingerprints or SMILES rely on purely structural information, limiting their use for a generalization as required here. This limitation can be overcome using representation as molecular graphs or molecular descriptors.^[20] Unfortunately, however, using descriptors alone also does not necessarily lead to a good generalization since high-dimensional representations are prone to overfitting.^[21] Thus, we used a Tree-Based feature reduction to eliminate descriptors that did not contribute to the overall prediction of the model. To ensure valid representations of polymeric data, each of the polymer building blocks (hydrophilic, hydrophobic, endcapping), was separately encoded, and the ratio information was embedded by multiplying each descriptor with this ratio factor. The molecular weight and the cell type used in the original dataset were added to the sample. The latter was achieved using one-hot encoding, a method that converts categorical features into binary vectors, enabling their representation in machine learning models. To minimize the noise that is introduced by the experimental data and especially

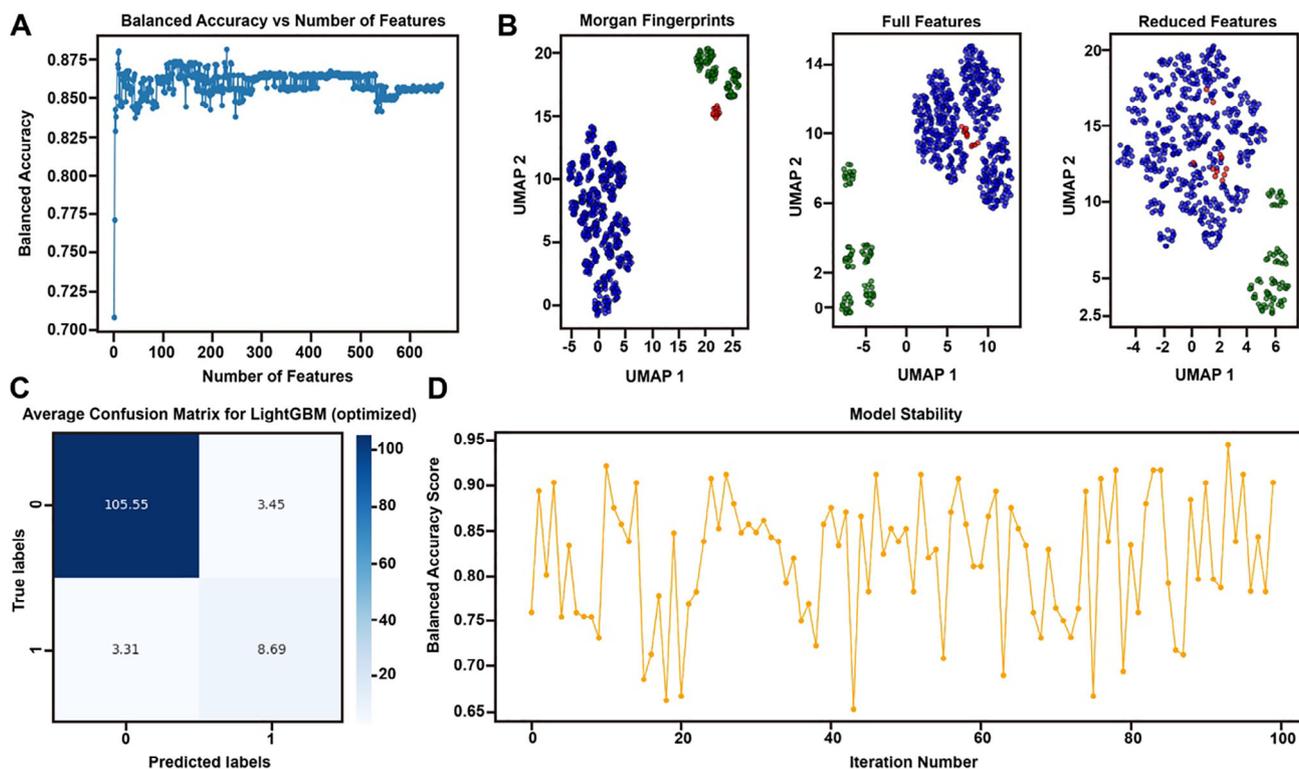


Figure 1. Insights into the ML Process A) Overview of the iterative feature reduction using LightGBM feature reduction. B) Comparison of feature space visualization with UMAP using featurization with MorganFingerprints (left), all RDkitDescriptors and self-defined features (middle), reduced feature set of the eleven most important features (right). C) Confusion matrix of the fully optimized LightGBM. Evaluated on the training set of 100 train-test splits and averaged accordingly. D) Evaluation of Model stability over 100 train-test splits, ranging from 0.65 to 0.95 balanced accuracy.

by merging datasets of different origin, we decided to use a binary binning approach to turn the regression problem, using the reported gene silencing percentages, into a classification problem. We selected a gene knockdown efficiency of 50% as threshold to separate the formulations into two different classes, reflecting our primary goal of assessing whether synthesizing a polymer is worthwhile rather than focusing on exact gene silencing values. Utilizing binary classification generally enhances interpretability, simplifies the analysis and effectively addresses data imbalance.

Using the prepared dataset, we first compared different ML algorithms (Figure S1A, Supporting Information). To address the imbalance in the dataset, balanced accuracy/mean recall was used to handle potential model biases toward the major class, and a RandomOverSampler was used to guarantee balanced training. The data was split into a train/test set at a ratio of 80:20 and 100 models were trained using each algorithm. The LGBMClassifier^[22] showed the best performance (0.8217 balanced accuracy) and was therefore selected for further optimization. We then compared different resampling strategies (Figure S2, Supporting Information), with SMOTEEN^[23] showing the best balanced accuracy (0.8309). After tuning using hyperopt (Figure S3, Supporting Information), additional feature reduction was performed, where eleven features lead to the best model performance (Figure 1A). This process was visualized using UMAP, revealing how feature reduction minimized gaps in the chemical properties space (Figure 1B). This approach was aimed to reduce the risk of overfitting while limiting the physico-

chemical information required to encode molecular structures. This ultimately facilitated the integration of different datasets and the generalization of unseen structures. The eleven most important features, using SHAP are shown in Figure S4 (Supporting Information). The tuned LGBMClassifier was finally evaluated using 100 stratified train-test splits of 80/20 and showed a mean balanced accuracy of 0.8462 on the validation sets (Figure 1C,D). Afterward, the model was trained on the entire dataset and used for the prediction task. The full workflow is also visualized in Figure S5 (Supporting Information).

2.2. Experimental Validation

To experimentally validate the trained classifier, novel polymers were rationally designed from available precursors via established synthetic routes. Given our group's significant expertise in synthesizing and characterizing poly(beta-amino ester)s (PBAEs), this class of polymers was selected as the focus for the validation set. To the best of our knowledge, all selected polymers are unpublished structures. The classifier predicted their potential knockdown efficiency. Based on these predictions, three polymers expected to exhibit low efficiency and three expected to exhibit high efficiency were selected for chemical synthesis and subsequent *in vitro* evaluation. Their schematic structures were shown in Figure S6 (Supporting Information), with specific chemical structures provided in Figures S7–S12

(Supporting Information). siRNA was formulated with these polymers at an N/P ratio of 10, and polyplexes were characterized regarding size, polydispersity, and zeta potential, as presented in Figure S13 (Supporting Information). Gene silencing efficiency was assessed in both enhanced green fluorescent protein (eGFP)-stably expressing H1299 cells (using siRNA targeting eGFP) and A549 cells (using siRNA targeting epidermal growth factor receptor (EGFR)). Consistent with the predictions, all three polymers anticipated to have low efficiency demonstrated negligible knockdown efficiencies (Figure S14, Supporting Information). However, the polymer OA-BG, comprising full oleylamine (OA) modification with bisphenol A glycerolate (BG) as its backbone, predicted as a high-efficiency candidate, failed to achieve the 50% knockdown threshold, reaching only 31.88% eGFP knockdown in H1299-eGFP cells and 24.62% eGFR knockdown in A549 cells. These results represented $\approx 30\%$ of the knockdown efficiency achieved by Lipofectamine 2000 and thus OA-BG was considered a false positive. In contrast, the other two polymers predicted to be high-performing, SP/TDA-BG (spermine/tetradecylamine with the BG backbone) and SP/OA-BG (spermine/oleylamine with the BG backbone), successfully demonstrated the predicted high knockdown efficiencies (91.82% and 96.17% eGFP knockdown, respectively). Overall, five out of six polymers were correctly classified, resulting in an experimental validation accuracy of 0.8333, which closely aligns with the classifier's estimated performance metric of 0.8462 (Section 2.1).

2.3. Characterization of Polymer and siRNA-Loaded Polyplexes

Following the experimental validation in Section 2.2, among the polymers tested, SP/TDA-BG demonstrated high transfection efficiency, in agreement with the classifier's prediction. Given its promising performance, we selected SP/TDA-BG as a model polymer for further systematic investigation into the relationship between its structural characteristics and biological activity. Although the machine learning model specifically suggested a 50:50 SP:TDA ratio as optimal, inspired by the transfection cliffs theory,^[24] we sought to investigate how minor deviations from this composition might impact transfection performance, as such effects are not necessarily captured by the machine learning model.^[25] Hence, we synthesized the corresponding PBAE polymers following the synthetic procedure shown in Figure 2A, adjusting the molar ratios of cationic monomer spermine and lipophilic monomer tetradecylamine from 40% to 60%, which were further confirmed by ¹H NMR analysis (Figure S15, Supporting Information). In addition, referring to our previous work on efficient siRNA delivery via amphiphilic PBAEs incorporating SP and OA with 1,4-butanediol diacrylate as the backbone,^[26,27] we also selected PBAE SP0.3/OA0.7 as a benchmark for comparative evaluation in our study.

The polymers were then complexed with siRNA at different N/P ratios. It is worth noting that, in the used dataset, N/P ratios were always set to at least 15 to ensure complete siRNA encapsulation and corresponding effectiveness. However, in our experimental work, we aimed to minimize polymer use, to particularly improve in vivo tolerability, based on our previous studies confirming efficient gene silencing and encapsulation at lower N/P ratios.^[26,28] Therefore, we initiated screening from an N/P ratio

of 3, increasing up to 10. Specifically, we assessed the physicochemical properties of the formed polyplexes, including size, size distribution, and zeta potential. Most polyplexes formed with diameters ranging from 50 to 300 nm and acceptable PDI values ≈ 0.2 (Figure 2B). Examining the zeta potential, a significant change was observed between N/P ratios of 3 and 5, particularly in case of polyplexes prepared with PBAEs SP0.5/TDA0.5 and SP0.4/TDA0.6, which displayed noticeable charge reversal (Figure 2C). Incomplete or unstable encapsulation of siRNA at N/P 3 (Figure S17, Supporting Information) could explain this observation. This near-neutral flipping zeta potential also revealed colloidal instability as evidenced by the extremely large size exceeding 2000 nm in case of polyplexes prepared with PBAE SP0.3/OA0.7 at N/P 5. When the ratio was increased to N/P 7 and N/P 10, the siRNA was completely encapsulated and the polyplexes were more stable in size.

Although stable formation of polyplexes is important for siRNA delivery, appropriate siRNA release is equally critical for successful gene silencing as the final action site will be in the cytoplasm, where the released siRNA cargo from polyplexes should bind with the RNA-induced silencing complex (RISC) to fulfill its function. Therefore, we investigated siRNA release from polyplexes in the presence of Triton X and heparin, which will competitively interfere hydrophobic and electrostatic interactions, respectively. After a non-linear fitting of released siRNA to the added interferents, EC50 values revealed that the release of equal amounts of siRNA from the polyplexes required higher concentrations of Triton X and heparin (6.2% vs 5.2%) when the spermine ratio in the polymers increased from 40% to 60% (Figure 2D). The EC50 value for SP0.3/OA0.7 polyplexes was even higher (12.1%), demonstrating the tightest binding between siRNA and the polymer in our study. The binding strength effectively protected siRNA from degradation by RNase, as all formulations retained >90% siRNA content after incubation with the enzyme. In contrast, free siRNA lost 99% of its integrity when treated with the same amount of RNase (Figure 2E).

2.4. In Vitro Performances: Cytotoxicity, Cell Uptake, and Knockdown Effects

We initially evaluated the safety profile of our polyplexes by assessing the viability of H1299 cells exposed to increasing polymer concentrations. The cell counting kit (CCK-8) assay showed a dose-dependent trend in cell viability. Notably, even at the highest N/P ratio of 20, cell viability remained above 80%. When the N/P ratio was reduced to 10, the viability of H1299 cells consistently reached 90–95% in all groups (Figure 3A). Therefore, all following experiments were conducted at an N/P ratio of 10 or lower. Next, we performed a wider uptake screening of polyplexes formulated from N/P 3 to N/P 10 in H1299 cells. With increasing N/P ratio, the uptake of all polyplex formulations was improved (Figure 3B). Quenching the fluorescent signal on the cell surface with trypan blue, only resulted in a negligible decrease in the detected mean fluorescence intensity (MFI), indicating internalization of the polyplexes rather than non-specific adsorption on the surface. Furthermore, the eGFP knockdown effects in H1299 cells stably expressing eGFP were consistently exceeding 94% in all polyplexes formulated at N/P ratios > 3 (Figure 3C).

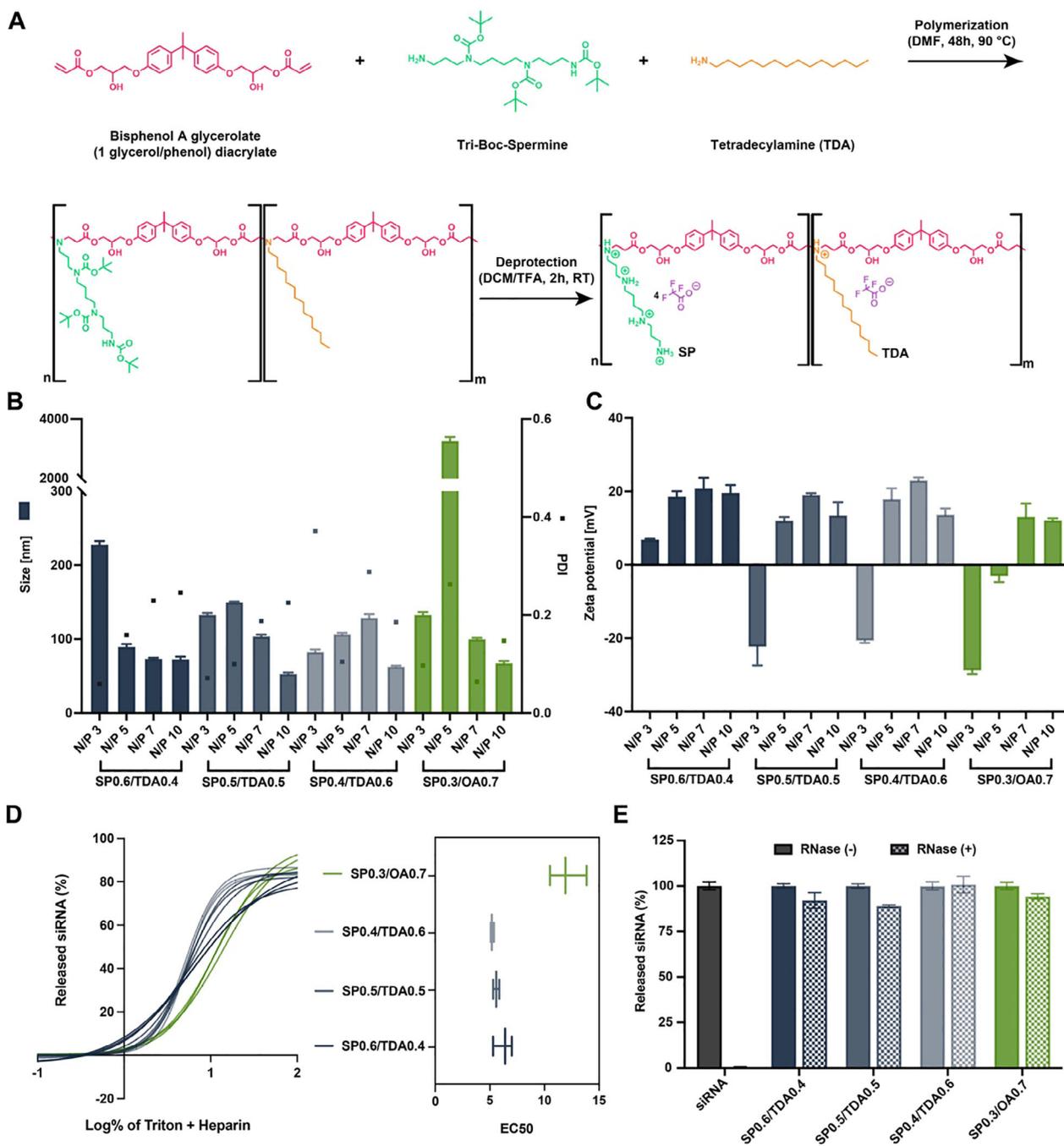


Figure 2. Characterization of synthetic polymers and siRNA-loaded polyplexes. A) Synthesis procedure and the structure of SP/TDA-BG PBAE polymers. B) Hydrodynamic diameter (represented by bar graph), polydispersity (represented by symbol), and C) zeta potential of siRNA-loaded polyplexes prepared at different N/P ratios. D) siRNA release from polyplexes at N/P ratio of 10 in the presence of Triton X and heparin using SYBR Gold assay, and EC50 values obtained by non-linear fitting analysis of released siRNA to added interfering substances. E) RNase protection assay of polyplexes prepared at an N/P ratio of 10. Polyplexes were first treated with RNase at 37 °C for 30 min, followed by RNase deactivation by heating to 70 °C for 30 min. After incubation with Triton X and heparin, released siRNA was quantified using SYBR Gold assay. Results are presented as mean ± SD, $n = 3$.

The uptake of polyplexes at N/P 10 in A549 cells mirrored the trends observed in H1299 cells, with reduced uptake observed when either SP or TDA proportions exceeded 60% (Figure 3D). This aligns with the mechanism of adsorptive endocytosis which is generally associated with polyplex uptake.^[29] For

highly hydrophilic cationic polymers such as poly(ethyleneimine) (PEI) and poly(L-lysine) (PLL), uptake primarily relies on electrostatic interaction with cell membrane.^[30,31] Hydrophobic modifications, however, have been shown to enhance uptake through interactions with lipids and membrane proteins.^[32,33]

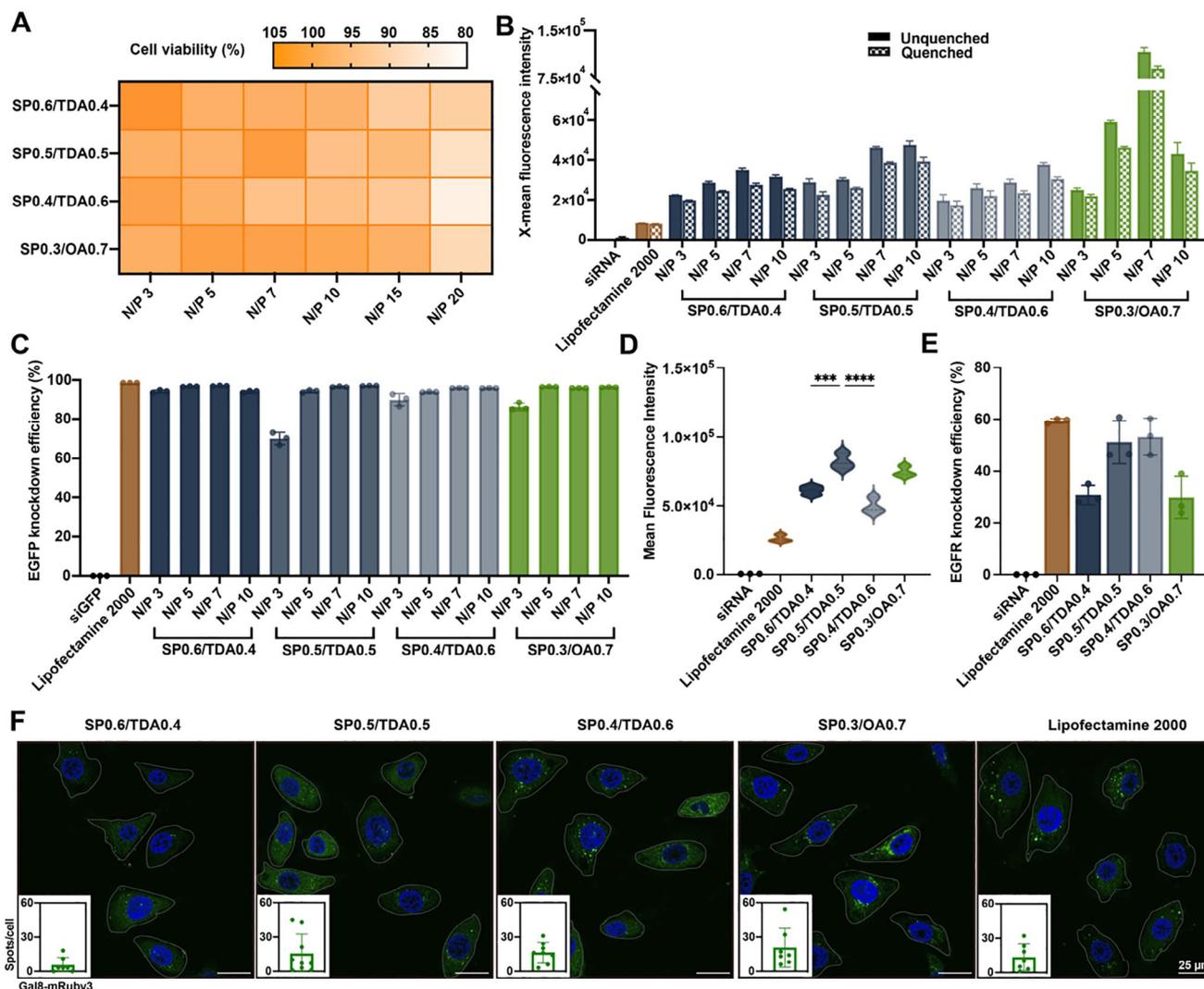


Figure 3. In Vitro performances of siRNA-loaded polyplexes. A) Viability of H1299 cells after treatment with polyplexes formulated at different N/P ratios. B) Cellular uptake of polyplexes containing Alexa Fluor 647-labeled siRNA in H1299 cells, presented as mean fluorescence intensity. C) EGFP knockdown efficiency of polyplexes in H1299/eGFP cells. (D) Cellular uptake of polyplexes containing Alexa Fluor 647-labeled siRNA in A549 cells. (Data are presented as mean \pm SD, $n = 3$; $***p < 0.001$, $****p < 0.0001$, one-way ANOVA) E) EGFR knockdown efficiency of polyplexes in A549 cells. F) Fluorescent spots of Gal8-mRuby3 in genetically modified Hela cells after 4 h of treatment with different polyplexes. Green color represents Gal8-mRuby3, nuclei are shown in blue. Scale bar, 25 μ m. Quantification of Gal8-mRuby3 dots was performed by the Fuji plug-in of Image J, and data are presented as mean \pm SD.

Similarly, Rui et al. reported that increasing PBAE hydrophobicity initially boosted uptake before declining, regardless of whether delivering siRNA, mRNA, or DNA.^[12] In our study, PBAE SP0.5/TDA0.5 polyplexes achieved the highest uptake, with an MFI > 80000. This indicates that a balance of electrostatic and hydrophobic interactions is crucial for optimal delivery.

Importantly, improved cellular uptake does not always correlate with stronger transfection. Although siRNA-loaded PBAE SP0.3/OA0.7 polyplexes showed superior internalization in A549 cells, transfection efficiency was lower than expected and inferior to the performance observed in H1299/eGFP cells (Figure 3E). This discrepancy may be attributed to the differences in siRNA lengths used for targeting eGFP (52 nucleotides) and EGFR (42 nucleotides) or differences in cell-type specific intracellular pro-

cessing. Meanwhile, the slower release of siRNA that we observed in SP0.3/OA0.7 polyplexes may be another reason (Figure 2D). Notably, despite lower uptake of PBAE SP0.4/TDA0.6 polyplexes, their knockdown efficiency (53.4%), was comparable to SP0.5/TDA0.5 (51.2%). This observation might be explained by the efficient endosomal escape, which we investigated utilizing the Galectin-8 (Gal8) assay.^[34] In brief, Gal8 binds glycans exposed upon endosomal membrane disruption, enabling quantification of endosomal escape using Gal8-mRuby-expressing cells.^[12,35] The average number of Gal8-mRuby3 punctate fluorescent spots increased from 5.43 to 16.25 per cell as the lipophilic TDA content was increased from 40% to 60% (Figure 3F). This finding underscored that lipophilic components enhanced hydrophobic interactions with membranes, leading to structural

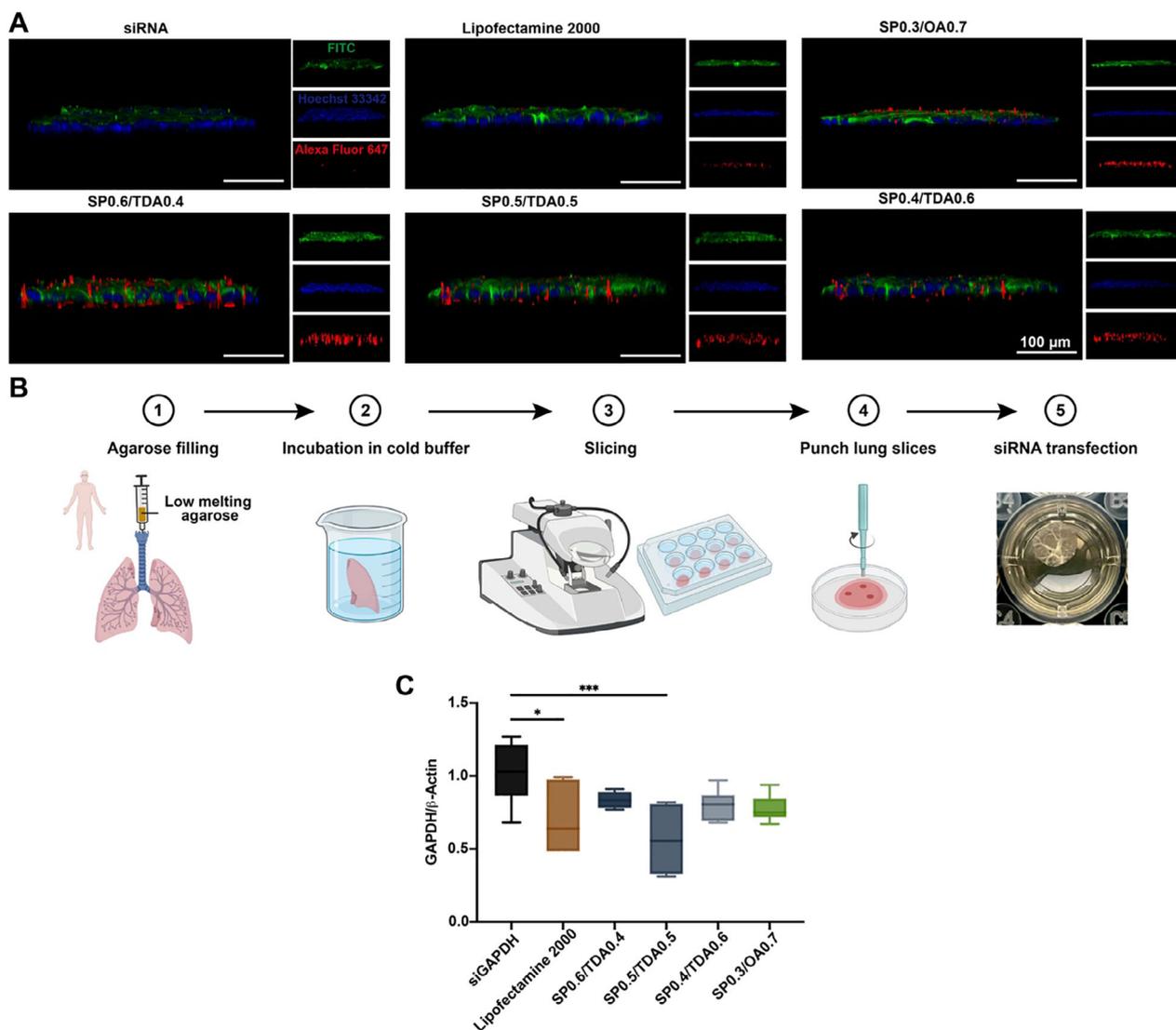


Figure 4. Mucus penetration assay and ex vivo knockdown of house-keeping gene GAPDH. A) Mucus penetration of polyplexes in air-liquid interface (ALI) culture of Calu-3 cells 24 h after transfection. Red color represents Alexa Fluor 647 labeled siRNA, nuclei are shown in blue and mucus layer in green. Scale bar, 100 μ m. B) Schematic diagram of preparation of human precision cut lung slices (hPCLS). C) GAPDH gene knockdown efficiency in hPCLS transfected with different formulations. The experiments were performed in technical triplicates and data are presented as mean \pm SD, $n = 2$; * $p < 0.05$, *** $p < 0.001$, one-way ANOVA.

instability of the membrane and disruption.^[36] As a result of this disruption, polyplexes were able to escape the endosome, releasing siRNA into the cytoplasm to bind RISC, cleave target mRNA, and achieve effective knockdown of protein translation.

2.5. Mucus Penetration on ALI Model and Gene Silencing in hPCLS

For pulmonary delivery, the mucus layer on the surface of the respiratory tract poses a significant barrier to effective siRNA delivery.^[37,38] If an RNA-carrier interacts excessively with mucus, it will not be able to penetrate this barrier during the time of mucus turnover, leading to its clearance from the lung

before cellular internalization. Additionally, tight junctions between respiratory epithelial cells further act as another barrier to paracellular transport of siRNA.^[39,40] To evaluate the ability of our PBAEs-siRNA formulations to overcome these lung-specific barriers, we used an air-liquid interface (ALI) culture of Calu-3 cells. Under ALI conditions, Calu-3 cells differentiate into a pseudostratified epithelium, produce mucus and cilia-like microvilli, and thus closely mimic the in vivo respiratory tract environment.^[41,42] As shown in **Figure 4A**, we obtained images by laser confocal laser scanning microscopy (CLSM), labelling mucus (green), cell nuclei (blue) and siRNA (red). Importantly, the mucus was largely distributed above the nuclei in all samples, confirming the successful establishment of a cell monolayer with mucus on the air-exposed side. When treated with free siRNA, signals from the siRNA were barely

detectable. In the Lipofectamine 2000 control group, a very weak red signal was observed across the mucus layer toward the cell layer. For the ALI cells treated with PBAE SP0.3/OA0.7 polyplexes, the red signal was significantly increased but mainly distributed within the mucus layer. In contrast, strong red signals were observed in the samples treated with PBAE SP/TDA polyplexes, with a wide distribution extending from the mucus layer to the cellular nuclei. However, a slight decrease in the red signal was observed across the cell monolayer as the lipophilic TDA ratio increased in the PBAE polymers. As previously reported, the long mucin proteoglycans chains present in mucus entangle, usually forming hydrophobic domains and hydrophilic channels in the network. This periodic hydrophobic domains have been shown to interact with hydrophobic particles or particles exhibiting hydrophobic moieties.^[43,44] For polyplexes with comparable electrical properties, this hydrophobic affinity may cause polyplexes with higher ratio of lipophilic monomers, either OA or TDA, to be restricted in diffusion. Overall, CLSM images showed that amphiphilic PBAEs consisting of SP/TDA were able to penetrate mucus and mediate sufficient uptake in epithelial cells.

Further increasing the biological complexity, we evaluated the gene silencing effects of our polyplexes in human Precision-Cut-Lung-Slices (hPCLS) (Figure 4B). hPCLS are widely recognized as a powerful tool for investigating drug responses in an environment that accurately reflects the complexity of the human lower respiratory tract. hPCLS maintain the native lung architecture, which includes the respiratory parenchyma and small airways, as well as a variety of lung-resident cells, including type I and II alveolar cells, bronchial epithelial cells, endothelial cells, and immune cells.^[45] After 48 h of siGAPDH transfection in hPCLS, the gene silencing effects were evaluated by measuring the downregulation of the housekeeping gene GAPDH as previously described.^[46] In this proof-of-concept study, GAPDH was chosen as a target gene only to evaluate the delivery efficiency, and it will be replaced with an aberrant gene for treating specific diseases in future applications. In addition, the hPCLS used in our study were derived from non-lesional regions and were in principle free of abnormal genes. As a result, qPCR analysis of the extracted RNA from the slices showed that the average GAPDH/ β -Actin ratio was ≈ 1.0 in the free siGAPDH-treated group, while in the Lipofectamine 2000-treated group, this ratio dropped significantly to 0.71 (Figure 4C). SP0.6/TDA0.4 and SP0.4/TDA0.6 polyplexes enabled a slight decrease of GAPDH gene expression in the hPCLS, with reductions of 16.3% and 20.1%, respectively. Overall, SP0.5/TDA0.5 polyplexes demonstrated the highest gene silencing efficiency, achieving a 43.7% reduction of the GAPDH level, confirming the need for balancing cationic and hydrophobic content in the PBAE nanocarriers for efficient pulmonary delivery.

2.6. In Vivo performance: Biodistribution, Biocompatibility, and Knockdown Effects after Pulmonary Delivery

Based on in vitro and ex vivo results, we selected PBAE SP0.5/TDA0.5 to move further to in vivo studies. Alexa Fluor 647-labeled siRNA was loaded into polyplexes and delivered via

intratracheal instillation (Figure 5A). Compared to free siRNA, polyplexes demonstrated significantly higher retention and internalization in the lung (Figure 5B; Figure S19, Supporting Information), with an 82.3-fold increase in fluorescence intensity (Figure 5C). As observed previously,^[47] when administered as polyplexes, some siRNA entered systemic circulation via pulmonary capillaries, accumulating in the liver before metabolism as evidenced by the signal detected in the liver and kidneys, respectively. Due to the complex architectural structure in the lung, polyplexes may face challenges in reaching the respiratory zone, which cannot be accurately evaluated by IVIS imaging. CLSM images revealed that polyplexes containing pHrodo red-labeled siRNA, represented in red color, have successfully reached not only the lower respiratory tract but also the respiratory zone (Figure 5D). Furthermore, polyplex uptake was observed in various cell types within the lung, and the corresponding flow cytometric gating strategy is shown in Figure S20 (Supporting Information). The higher MFI in polyplex-treated alive lung cells (average 2136) was consistent with IVIS imaging results (Figure 5E). High MFI detected in dendritic cells (average 4941), macrophages (average 7773), and eosinophils (average 1348) highlighted strong phagocytosis potential in the lung, which generally poses a challenge for pulmonary siRNA delivery. Interestingly, the uptake of polyplexes in both CD4⁺ and CD8⁺ T cells remained low, being beneficial for avoiding adverse immune activation and in line with the need for targeting ligands for efficient T cell transfection.^[48] Importantly, lung epithelial cells, particularly type II pneumocytes, are often related to the progression of respiratory diseases, such as chronic obstructive pulmonary disease (COPD), lung cancer, lung fibrosis, and pneumonia.^[49–51] The uptake of polyplexes in epithelial cells, particularly in type II pneumocytes was 9.46-fold and 7.61-fold higher, respectively, when compared to free siRNA. These results suggest the potential of siRNA therapy based on our carrier system for treating respiratory diseases in the future and underline the need for nanocarriers in pulmonary delivery.

Next, we evaluated the siRNA knockdown efficiency in the lung and performed safety assessment. PEI25k, as well-established control, presents reliable transfection efficiency in gene delivery and has been widely used in previous studies focusing on polymer-based carriers.^[52,53] Due to its known cytotoxicity, PEI25k is also used as a positive control in safety evaluations and was therefore included in our in vivo test. RNA extracted from the lungs treated with different formulations was analyzed via qPCR. In the control group that received buffer only, the average GAPDH/ β -Actin ratio was 1.03 (Figure 6A). In mice treated with free siGAPDH, the ratio increased to 1.44, with a broader standard deviation of 0.39, demonstrating that free siRNA did not achieve GAPDH gene silencing. In contrast, siGAPDH-loaded PBAE polyplexes showed a significant 30.4% reduction of the GAPDH/ β -Actin ratio when compared to negative control siRNA-loaded PBAE polyplexes. In the mice treated with PEI-siGAPDH, the GAPDH/ β -Actin ratio oppositely increased to a broad range from 1.21 to 2.69, likely due to severe lung inflammation as hematoxylin and eosin (H&E) staining revealed noticeable inflammatory cell infiltration, alveolar wall thickening, and disruption of the alveolar architecture in these mice (Figure 6B). Conversely, lung tissue structures

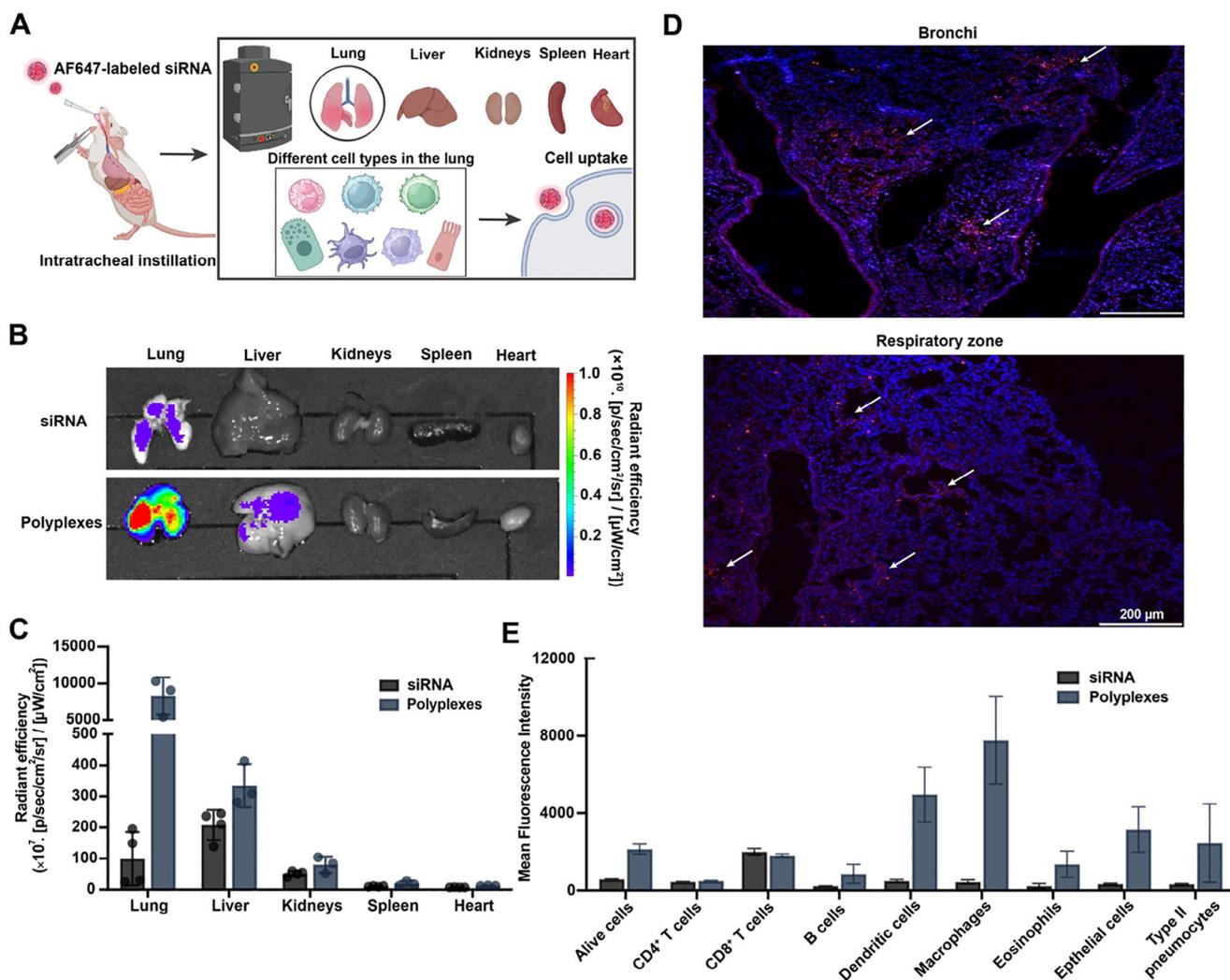


Figure 5. In vivo biodistribution in the organs and cellular uptake in the lung. A) Schematic diagram of in vivo distribution investigation after intratracheal instillation of polyplexes containing Alexa Fluor 647-labeled siRNA. B) Representative organ distributions in mice that received free siRNA or siRNA-loaded polyplexes, respectively. C) Quantification of fluorescence intensity of Alexa Fluor 647 labeled siRNA distributed in the organs. D) Distribution of polyplexes containing pHrodo red-labeled siRNA in different lung regions. White arrows indicate polyplexes. Scale bar, 200 μm . E) Mean fluorescence intensity of AF647-labeled siRNA in different cell types in the lung. Data are presented as mean \pm SD, $n = 3$.

in PBAE polyplex-treated mice were well-preserved, with clear alveolar spaces and negligible alveolar wall thickening as observed in buffer- and free siRNA-treated groups, which suggested minimal lung tissue damage or inflammation in these mice. Consistent with the H&E staining results, levels of inflammatory cytokines, i.e., IL-6, MCP-1, IFN- β , TNF- α in BALF were significantly higher in PEI-siGAPDH and PEI-siNC treated mice when compared to other groups (Figure 6C). In particular, IL-6 was detected at the highest concentration among all cytokine types, in the PEI-siGAPDH group, with an average value of 322 pg/mL, which was 9.4-fold and 64.4-fold higher than PBAE-siGAPDH and free siGAPDH groups, respectively. However, PBAE polyplexes treatment did not abnormally elevate cytokine levels, which remained comparable to blank and free siRNA-treated mice on most indicators.

3. Study Limitations and Data Scarcity

The study presented here demonstrates an elegant, literature-driven strategy for screening polymeric gene-delivery candidates and yields promising results on a newly synthesized validation set. Nevertheless, several limitations must be acknowledged so that readers can appreciate the scope of our conclusions.

First, although focusing on polyesters is a sensible starting point, essential details such as copolymerization patterns, block lengths, architecture, dispersity, and molecular weight variation are rarely reported, and even when they are, they are seldom provided in a standardized and machine-readable format. As a result, descriptors based on idealized repeat units capture only a fraction of the true physicochemical diversity. Ongoing standardization efforts that mandate sharing raw chromatograms and NMR spectra may eventually allow direct ingestion of this information

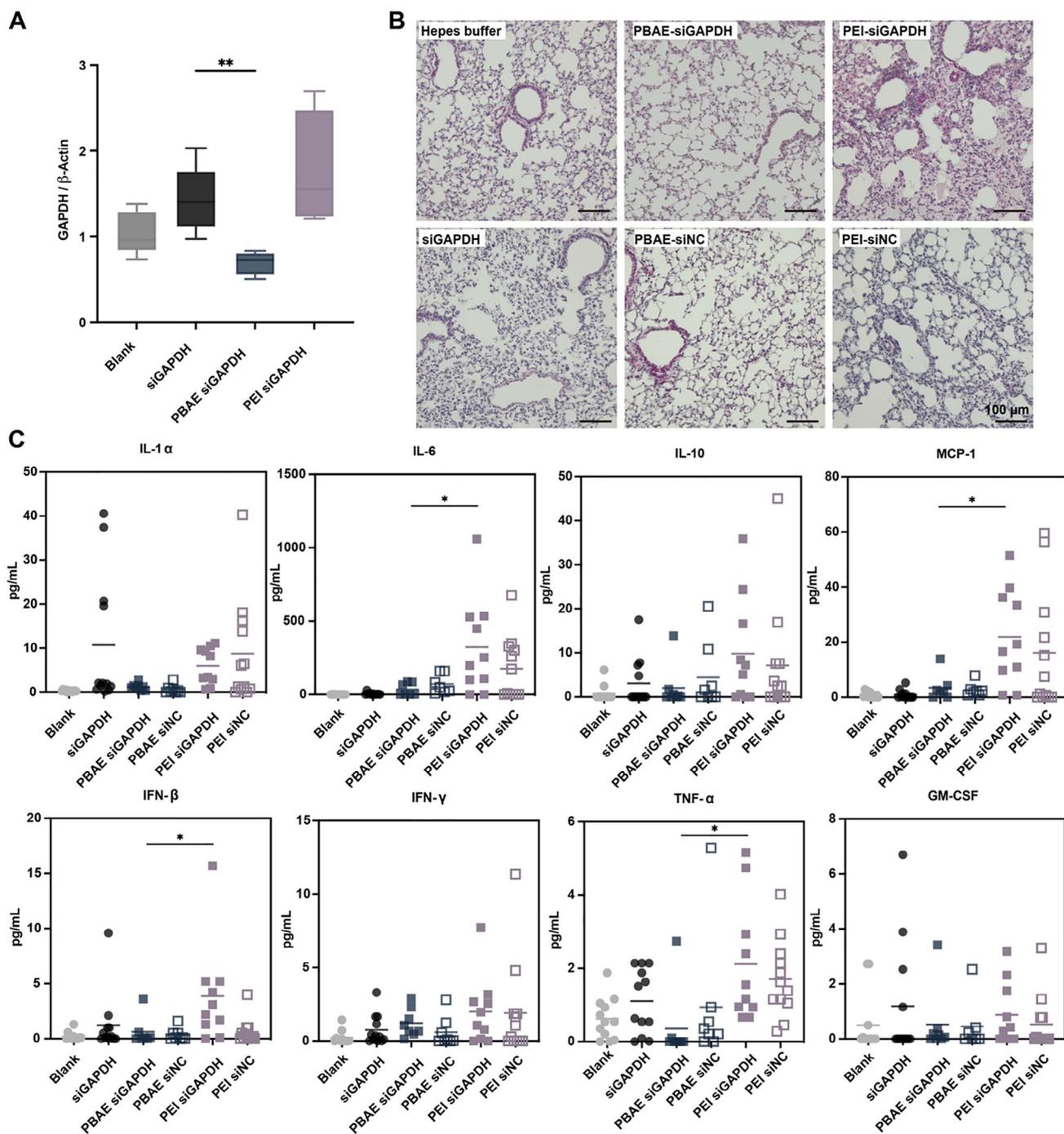


Figure 6. In vivo GAPDH gene silencing and safety evaluations. A) GAPDH gene silencing effects of polyplexes in the lung, data are presented as mean \pm SD, $n = 6$ for Blank and siGAPDH groups, $n = 4$ for PBAE siGAPDH and PEI siGAPDH groups, $**p < 0.01$, Student's *t*-test. B) H&E-stained lung sections collected from mice treated with different formulations. Scale bar, 100 μ m. C) Inflammatory cytokine levels in the bronchoalveolar lavage fluid (BALF) collected from mice treated with different formulations. The experiments were performed in technical duplicates and data are presented as mean \pm SD, $n = 6$ for Blank and siGAPDH groups, $n = 4$ for PBAE siGAPDH and PEI siGAPDH groups, $*p < 0.05$, one-way ANOVA.

into machine-learning pipelines, but such data are not yet widely available.

Second, data sparsity is a major hurdle. Whereas proteins and small molecules benefit from extensive databases, experimentally characterized polymeric gene-delivery systems are scarce.

We therefore limited the chemical space to structurally similar polyesters and used a UMAP projection solely as a qualitative coverage check. Predictions outside this region must be treated with caution, because extrapolating far from the training manifold typically yields unreliable results. A rigorous, quantitative

safeguard was not implemented here for three practical reasons: 1) no curated set of truly out-of-domain polymers yet exists for calibration; 2) distance estimates are highly sensitive to the chosen descriptor space; and 3) alternative distance metrics and thresholding schemes can give conflicting signals when data are sparse. As larger, standardized data sets emerge, these challenges should become tractable, enabling formal applicability-domain filters to accompany future models.

Third, biological context also matters. One-hot encoding of cell lines allows within-set predictions but offers no mechanistic insight and cannot guarantee accuracy for cell types absent from the training data. Future work could explore lineage- or transcriptome-derived embeddings to improve transferability.

Finally, although RDKit descriptors efficiently encode molecular structure, they are not optimized for human-interpretable structure-function insight. Graph-based neural network representations may provide traceable, learnable features and can be back-mapped to their structure^[54,55] once larger, standardized data sets become available.

4. Conclusion

This study provides an efficient approach for utilizing literature data to train a ML model for predicting suitable polymeric delivery systems. By employing straightforward strategies, we successfully merged multiple different datasets with different carrier systems. The trained *in silico* model was validated to be accurate by assessing *in vitro* gene silencing outcomes when siRNA was delivered using polymers that the ML model predicted to be effective or ineffective. Among the tested polymers, one candidate PBAE SP/TDA-BG was selected for detailed investigations of its structural characteristics and biological performance. This polymer, with its balanced hydrophilic and hydrophobic moieties combined with a biodegradable backbone, overcame key biological barriers in pulmonary siRNA delivery. Remarkably, it achieved efficient *in vivo* gene silencing without detectable adverse effects. These findings highlight the capability of the ML model to significantly reduce the need for extensive experimental screening efforts and associated resource costs and ethical considerations. Our study also provides conceptual insights into the complex processes of polymeric siRNA delivery, which emphasizes the transformative role of ML in optimizing delivery systems. While current limitations include a constrained dataset, which makes it difficult to extrapolate to novel polymer types, this challenge could be mitigated as more data becomes available. With expanded datasets, data-intensive methods, such as Deep Generative Models, could aid the design of entirely new materials for future nanomedicine applications.

5. Experimental Section

Data processing and Machine Learning: Structural data was collected from literature references^[12,56,57] on 605 polymers that had been employed for siRNA delivery before, reflecting a range of different polyester types. Chemical structures were created using ChemDraw (version 22.2.0). All data related tasks were performed using Python (version 3.11.5). Molecule sanitizing, embedding and MMFF force field optimization as well as Molecular Descriptor and Morgan Fingerprint calculation (radius = 2, nbits = 2048) were performed using the widely adopted cheminformatics

library RDKit (version 2024.09.1). Each block monomer was encoded separately and the respective component ratio was incorporated in the descriptors by multiplying them with the weighted ratio of copolymer blocks following Kim et al.^[58] Gene knockdown (KD) performance was categorized into two groups: KD < 50%, and > 50%. Additional data, including monomer ratios, cell types, and molecular weight (Mw), were incorporated. Data was cleaned removing multiple entries and columns that contain NaN (Not a number), followed by normalization of features using StandardScaler class from sklearn (version 1.6.0). Various models (SVM, KNN, RF, XGB, LGBM, NaiveBayes) with weighted sampling due to dataset imbalance, were evaluated. The lead model (LGBM) was fine-tuned with hyperopt (version 0.2.7). Important features were calculated using SHAP values and a TreeExplainer class. Irrelevant features were excluded from the dataset, using the integrated feature_importance method in LGBM. Data was split into training and test sets, stratified by KD classes (20% test set ratio). The trained model was applied to assess new, unpublished polymer formulas, identifying one high-performing polymer selected for synthesis. Additionally, waterfall plots were calculated for the predicted polymer using SHAP library version (version 0.46.0). The following Python libraries were used for data handling and plotting: Sklearn, Imblearn (0.13.0), Pandas (2.1.4), Numpy (1.26.4), Seaborn (0.13.2), Matplotlib (3.9.0).

Chemicals: Ethyl trifluoroacetate, tetradecylamine, oleylamine, 4-Amino-1-butanol, 1,4-butanediol diacrylate, and bisphenol A diglycidyl ether diacrylate were purchased from Sigma Aldrich (Taufkirchen, Germany). Di-tert-butyl dicarbonate, spermine, and SYBR Gold Nucleic Acid Gel Stain were bought from Fisher Scientific (Hampton, NH, USA).

Synthesis of Tri-boc-spermine: Tri-tert-butyl carbonyl spermine abbreviated as tri-Boc-spermine was synthesized as described elsewhere.^[28] In brief, spermine (1 eq) was dissolved in methanol and stirred at -78°C , ethyl trifluoroacetate (1 eq) was subsequently added dropwise and stirred at -78°C for 1 h, then at 0°C for 1 h. Without isolation, di-tert-butyl dicarbonate (4 eq) was added dropwise to the solution and stirred at room temperature (RT) for 2 days. Finally, the solution was adjusted to a pH > 11 by 25% ammonia and stirred overnight to cleave the trifluoroacetamide protecting group. The mixture was then evaporated under vacuum, and the residue was diluted with dichloromethane (DCM) and washed with distilled water and saturated sodium chloride aqueous solution. The DCM phase was finally dried by magnesium sulfate and concentrated to give the crude product. The crude product was purified by column chromatography ($\text{CH}_2\text{Cl}_2/\text{MeOH}/\text{NH}_3$, aq. 7:1:0.1, SiO_2 , KMnO_4 ; Rf = 0.413). Tri-Boc-spermine was isolated and characterized by ^1H nuclear magnetic resonance spectroscopy (^1H -NMR).

Synthesis of PBAEs: The synthesis involved dissolving hydrophilic amine in dimethylformamide (DMF) and adding lipophilic amine and the diacrylate backbone (1.2 eq). The reaction was sealed, heated and kept at 90°C for 48 h, then cooled to RT. DMF was evaporated, and the solid polymer was solubilized in DCM. Deprotection of the triboc-spermine containing polymers was achieved by the dropwise addition of Trifluoroacetic acid (TFA) to a final concentration of 5% v/v, cleaving the Boc groups. The reaction was stirred at RT for two h. To obtain the deprotected polymer, the solvent was evaporated. For all polymers the solid was purified by precipitating it in diethyl ether followed by a centrifugation step (1250 rpm for 2 min). The procedure was repeated three times. The final product was dried under vacuum and characterized using ^1H -NMR.

Gel Permeation Chromatography (GPC): GPC was performed with an Agilent aqueous GPC using a PSS Novema max Lux 100A followed by two PSS Novema max Lux 3000A columns. The chromatographic system and calibration standards were set up according to pre-analysis by Agilent Technologies. Measurements were performed at 40°C in 0.1 M sodium chloride solution supplemented with 0.3% formic acid. Samples were prepared at 4 g/L and measured at a flow of 1 mL min^{-1} . Molar Mass distributions were obtained through the Agilent WinGPC Software against pullulan calibration standards in the range of 180 Da to 1450 kDa. A daisy-chain detector setup of an Agilent 1260 VWD was followed by an Agilent 1260 GPC/SEC MDS and ended with an Agilent 1260 RID.

Preparation of Polyplexes: To prepare PBAE-siRNA polyplexes, the polymer stock solution was diluted to various concentrations with diethyl

pyrocarbonate (DEPC) treated water. Next, an equal volume of a specific amount of siRNA diluted in 10 mM HEPES buffer (pH 5.4) was added, and the mixture was incubated at RT for 30 min to obtain siRNA-loaded polyplexes at different N/P ratios. The N/P ratio represents the molar ratio between the polymer amine groups (N) and the siRNA phosphate groups (P), and the amount of polymer required for different N/P ratios was calculated using the following formula:

$$m \text{ (polymer in } \mu\text{g)} = n \text{ siRNA (pmol)} \times \text{N/P} \times \text{number of nucleotides siRNA} \times \text{M protonable unit (g/mol)} \quad (1)$$

The number of nucleotides for asymmetric 25/27mer siRNA was set to 52, while in EGFR siRNA, only 42 nucleotides were present. The protonable units for each polymer were calculated by dividing the molar mass of the repeating unit by the number of protonable amines within each repeating unit.

Characterization of Polyplexes: Particle size, polydispersity index (PDI) and zeta potential of PBAEs-siRNA polyplexes were determined using a Zetasizer Ultra (Malvern Instruments, Malvern, UK). All measurements were conducted using a 10 mM HEPES buffer as dispersant. Results are expressed as mean \pm standard deviation (SD) over three measurements.

The encapsulation efficiency of siRNA was determined using SYBR gold assays. In brief, 15 μL of PBAEs-siRNA polyplexes were added into a 384-well plate, then 5 μL of a 4X SYBR Gold solution were added to each well and incubated for 15 min protected from light at RT. Fluorescence intensity was measured using a plate reader (Tecan, Männedorf, Switzerland) with excitation and emission wavelength set at 492 and 555 nm, respectively. An equal amount of free siRNA was used as 100% value for calculating the unencapsulated siRNA in different polyplex samples.

siRNA Release Assay: SYBR Gold assay was performed to investigate siRNA release from polyplexes under different conditions. First, PBAEs-siRNA polyplexes at an N/P ratio of 10 were prepared as described under 5.6. Polyplexes containing 10 pmol of siRNA were incubated with serial dilutions of Triton X and heparin in a 384-well plate for 30 min at 37 °C. Then, 10 μL of a 4X SYBR Gold solution were added to each well and incubated for 15 min. The results were measured as described under 5.7.

RNase Protection Assay: PBAE-siRNA polyplexes at an N/P ratio of 10 were prepared as previously described. A total of 50 μL of the respective formulations containing 50 pmol of siRNA, was incubated with 1 μg RNase A (Sigma-Aldrich, Taufkirchen, Germany) for 30 min at 37 °C. As a control group, 50 pmol of free siRNA was included, either treated with 1 μg of RNase, or left untreated as a 100% reference value for calculating the degraded siRNA. Subsequently, the RNase was deactivated by heating to 70 °C for 30 min. To release the RNA, 1% Triton X and 4 IU heparin was added and incubated for 30 min at 37 °C. The released RNA was then quantified using SYBR Gold assay, and fluorescence was measured as described under 5.7.

In Vitro Cell Viability: H1299 cells seeded in 96-well plates at a density of 6000 cells per well were used to assess cytotoxicity. After incubation with PBAEs-siRNA polyplexes containing 20 pmol scrambled siRNA (siRNA negative control, siNC) ranging from N/P 3 to N/P 20 for 48 h, 10 μL of the Cell Counting Kit-8 (CCK-8, Sigma) reagent was added to develop color for 3–4 h. The optical density (OD) was measured on a Tecan plate reader at 450 nm and cell viability was calculated by dividing the values of groups treated with polyplexes by that obtained with the untreated group.

In Vitro Cellular Uptake: H1299 cells were seeded in 24-well plates at a density of 15 000 cells per well and incubated with PBAEs-siRNA polyplexes containing 50 pmol of siRNA with N/P ratios of 3 to 10, where 20% of the siRNA was Alexa Fluor 647-labeled. Free siRNA and Lipofectamine 2000 containing equal amounts of siRNA were used as controls. After 24 h of incubation, cells were divided equally. Half of the cells were measured directly with an Attune NxT flow cytometer (ThermoFisher Scientific, Waltham, MA USA), and the other half were pre-mixed with 0.4% Trypan blue solution and measured comparably.

A549 cells were seeded in 96-well plates at a density of 6000 cells per well and incubated with the same PBAEs-siRNA polyplexes contain-

ing 20 pmol of siRNA with an N/P ratio of 10. After 24 h of incubation, the cells were assessed on an Attune NxT flow cytometer (ThermoFisher Scientific).

In Vitro Endosomal Escape: HeLa-Gal8-mRuby3 cells were kindly provided by the lab of Professor Ernst Wagner (LMU Munich, Germany). HeLa-Gal8-mRuby3 cells were seeded in the 8-well chamber slide (ibidi, Gräfelfing, Germany) at a density of 10000 cells per well, and then incubated for 4 h with different PBAEs-siRNA polyplexes containing 40 pmol of siRNA (20% of which was Alexa Fluor 647-labeled). After incubation, the supernatant was discarded, and the chambers were rinsed with PBS for three times. The cells were first fixed with a 4% PFA solution at RT for 20 min and then stained with 0.5 $\mu\text{g mL}^{-1}$ of DAPI solution for 8 min. After rinsing the chambers with PBS for at least three times, the cells were imaged using a SP8 inverted confocal laser scanning microscope (Leica Camera, Wetzlar, Germany) equipped with a 63X objective. The fluorescent dots of Gal8-mRuby3 were quantified using the Fuji plug-in of Image J.

In Vitro eGFP Knockdown: Protein knockdown experiments were conducted using H1299 cells stably expressing enhanced green fluorescent protein (eGFP). Polyplexes were formulated with siRNA targeting eGFP mRNA or scrambled siRNA (siNC) with the same length. H1299/eGFP cells were seeded in 96-well plates at a density of 6000 cells per well and then incubated with polyplexes containing 20 pmol siGFP or 20 pmol siNC for 48 h. Lipofectamine 2000 was used as a positive control, while free siRNA served as a negative control. After incubation, the cells were collected to perform the FACS analysis (Attune NxT Flow Cytometer, ThermoFisher Scientific). The eGFP knockdown efficiency was calculated by dividing the Median Fluorescence Intensity (MFI) of siRNA-treated group by that of the respective siNC-treated group.

In Vitro EGFR Knockdown: An EGFR knockdown experiment was conducted in A549 cells using polyplexes formulated with EGFR siRNA. Per well, 6000 A549 cells were seeded in 96-well plates and treated with polyplexes containing either 20 pmol of EGFR siRNA or 20 pmol of scrambled siRNA at an N/P ratio of 10 for 48 h. Following incubation, the cells were collected and stained with Vio R667 anti-human EGFR antibody for 10 min. After washing twice with PBS, the cells were analyzed using a flow cytometer (Attune NxT) to assess EGFR expression.

Mucus Penetration and Uptake Study: Air Liquid Interface (ALI) experiments were conducted utilizing Calu-3 cell culture. Specifically, Calu-3 cells were seeded at a density of 250000 cells per well onto uncoated Transwell polyester cell culture inserts (6.5 mm, 0.4 μm pore size) and were maintained in culture for three days until confluent. On day 4, the apical medium was removed to establish ALI conditions, and the medium in the basolateral chamber was replaced with 300 μL of PneumaCult ALI medium (STEMcell Technology, Vancouver, Canada). The medium was replaced every three days until the transepithelial electrical resistance (TEER) values stably reached 300 Ωcm^2 when monitoring with an EVOM epithelial volt/ Ω meter (World Precision Instruments, Sarasota, USA). Polyplexes and Lipofectamine 2000, each containing 100 pmol of siRNA, 20% of which was Alexa Fluor 647-labeled, were applied on top of Calu-3 monolayers without previous washing and incubated for 24 h. Free siRNA was employed as a negative control. Afterward, the cells were stained with 100 μL of diluted Hoechst 33342 (for nuclear staining) and Alexa Fluor 488-wheat germ agglutinin (for mucus staining) at 37 °C for 20 min. Cells were then gently washed twice with PBS and mounted on glass slides using FluorSave reagent. Fluorescent images were immediately captured using a 40X objective on the SP8 inverted confocal laser scanning microscope (Leica Camera) and were processed using the Fuji plug-in of Image J.

Ex Vivo Activity in Human Precision-Cut Lung Slices (hPCLS): Human tissue, ethics statement and human precision-cut lung slices (hPCLS): Human lung tissues were obtained from the University Hospital Großhadern of the Ludwig-Maximilians-Universität (Munich, Germany) and the Asklepios Biobank of Lung Diseases (Gauting Germany). Participants provided written informed consent to participate in this study, in accordance with approval by the local ethics committee of the Ludwig-Maximilians-Universität Munich, Germany (Project 19–630). In brief, hPCLS were prepared from tumor-free peri-tumor tissue. The lung tissue was inflated with 3% agarose solution and then solidified at 4 °C. The lung sections with a thickness of 500 μm were cut from the tissue blocks using a vibration

microtome (HyraxV50) (Karl Zeiss AG, Oberkochen, Germany). hPCLS were cultured in DMEM F-12 medium supplemented with 0.1% FBS. Prior to experiments, hPCLS were cut into 4 mm diameter circular pieces using a biopsy puncher.

GAPDH gene silencing in hPCLS: Each well containing three punches of hPCLS in a 24-well plate was treated with different formulations containing either 100 pmol of siGAPDH or 100 pmol of siNC. Lipofectamine 2000 was included as a positive control and free siGAPDH as a negative control. After 48 h of incubation, the tissue punches were submerged in 1 mL TRIzol within lysing matrix D tubes and homogenized using a FastPrep 24 Tissue Lyzer (M.P. Biomedicals, Irvine, CA, USA). Subsequently, 200 μ L of chloroform was added to each homogenized sample and mixed vigorously. The samples were then centrifuged at 11000 g for 15 min at 4 °C, after which the aqueous phase containing RNA was transferred to a new 1.5 mL Eppendorf tube. To precipitate the RNA, 500 μ L of isopropanol was added and mixed thoroughly. After 10 min incubation at RT, the samples were centrifuged at 11 000 g for 10 min. The supernatant was discarded, and the RNA pellet was washed with 1 mL of ice-cold 75% ethanol, followed by centrifugation at 7500 g for 5 min at 4 °C. The supernatant was discarded again, and the RNA pellet was resuspended in 30 μ L of RNase-free water. The extracted RNA was then processed for cDNA synthesis using a high-capacity cDNA synthesis kit (Applied Biosystems). Synthesized cDNA was diluted and subjected to quantitative PCR (qPCR) using SYBR Green PCR Master Mix (ThermoFisher Scientific), with Hs_GAPDH_2_SG primers specific for human GAPDH (Qiagen, Valencia, CA, US). Hs_ACTB_2_SG primers for human β -actin (Qiagen) were used as the normalization control.

In Vivo Distribution of Polyplexes after Pulmonary Delivery: All animal experiments were conducted according to the German law of animal protection and approved by the Government of Upper Bavaria (ROB-55.2-2532.Vet_0220-171) and the Committee for Animal Experimentation of the Ludwig Maximilian University Munich, Germany.

Eight-week-old female BALB/c mice were intratracheally instilled with polyplexes containing 1 nmol of Alexa Fluor 647 labeled siRNA under ketamine/xylazine anesthesia. The control group received free Alexa Fluor 647-siRNA. After 24 h, mice were sacrificed with an overdose of ketamine/xylazine anesthesia, and organs including the heart, lung, liver, spleen and kidneys were harvested for imaging. Fluorescence was measured at an excitation wavelength of 635 nm and an emission wavelength of 668 nm using an IVIS Lumina III (PerkinElmer, Shelton, CT, USA). After imaging, the lungs were homogenized to obtain single-cell suspensions, using the Mouse Lung Dissociation Kit (Miltenyi Biotec, Germany) according to the manufacturer's protocol. The lung cells were first incubated with PBS solution containing Zombie UV and later stained with FITC anti-mouse CD45, BUV395 anti-mouse CD3, Vioblu anti-mouse CD4, APC-Cyanine7 anti-mouse CD8, PE-Cyanine7 anti-mouse F4/80, BUV605 anti-mouse CD11c, BV785 anti-mouse CD326, PE/Dazzle594 anti-mouse CD170 and PerCP/Cyanine5.5 anti-mouse CD19 for 30 min at 4 °C. The stained cells were measured using a Cytex Aurora (San Diego, California, USA) implemented with autofluorescence extraction for the detection of cellular uptake.

Distribution of Polyplexes in the Lung: Eight-week-old female BALB/c mice were intratracheally instilled with polyplexes containing 1 nmol of pHrodo red-labeled siRNA under ketamine/xylazine anesthesia. After 24 h, the mice were sacrificed with an overdose of ketamine/xylazine anesthesia, and the lungs were harvested after lung perfusion. The lungs were then immersed in 4% PFA solution overnight. After PFA fixation, the lung tissues were embedded in paraffin and sliced into lung sections with thickness of 4 μ m. The obtained slices were deparaffinized by incubating in xylene, followed by a series of ethanol dilutions. After hydration, the slices were stained with 0.5 μ g mL⁻¹ DAPI solution for nuclear visualization and imaged using a 10X objective on an SP8 inverted confocal laser scanning microscope (Leica Camera).

In Vivo Transfection Evaluation of Polyplexes: Safety evaluation: Eight-week-old female BALB/c mice were intratracheally instilled with different formulations containing 1 nmol of siGAPDH or 1 nmol of siNC, including PBAEs-siRNA and PEI-siRNA polyplexes. Control groups received either free siGAPDH or buffer only. After 24 h, the mice were sacrificed, and their lungs were first perfused with 10 mL of saline. Bronchoalveolar lavage fluid

(BALF) was collected in a PBS/2 mM EDTA buffer containing protease inhibitor cocktail (cOmplete). The BALF was centrifuged at 500 g for 5 min at 4 °C, and the supernatant was used to measure the concentration of pro-inflammatory cytokines using the LEGENDplex Mouse Cytokine Panel 2 kit (Biolegend, San Diego, California, USA). The lungs were harvested, with one lobe fixed in 4% PFA overnight and then embedded in paraffin for histological analysis via H&E staining, while the remaining tissue was stored in 1 mL of RNA-later solution for further analysis.

In vivo GAPDH gene silencing efficacy of polyplexes: The lungs stored in RNA-later solution were transferred to lysing matrix D tubes and homogenized using a FastPrep 24 Tissue Lyzer (M.P. Biomedicals). RNA extraction was performed following the TRIzol-chloroform method as previously described under 5.16.2. The extracted RNA was then processed for cDNA synthesis using a high-capacity cDNA synthesis kit (Applied Biosystems). The synthesized cDNA was diluted and subjected to qPCR using SYBR Green PCR Master Mix (ThermoFisher Scientific) with Mm_GAPDH_3_SG primers (Qiagen) for GAPDH. Mm_ACTB_2_SG primer specific for mouse β -actin were used as the normalization control.

Statistical analysis: All data were expressed as means \pm standard deviation (SD). All statistical analyses were performed using one-way analysis of variance (ANOVA) in GraphPad Prism or Student's t-test when specifically stated. Levels of significant differences were expressed as follows, * p < 0.05, ** p < 0.01, *** p < 0.001, **** p < 0.0001.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

F.S.-S. and M.J. contributed equally to this work. These studies were funded by the European Research Council (ERC-2022-COG-101088587). Min Jiang is funded by the China Scholarship Council. The authors gratefully acknowledge the provision of human biomaterial (non-fibrotic fresh tissue from non-CLD patient) and clinical data from the CPC-M bioArchive and its partners at the Asklepios Biobank Gauting, the LMU Hospital and the Ludwig-Maximilians-Universität München. The authors would like to thank the patients and their families for their support. The authors additionally acknowledge ethris GmbH for whole body imaging resources and LMU's BMC flow cytometry core facility for expert advice and technical support. Figures were partially generated using biorender.com.

Conflict of Interest

Olivia Merkel and Benjamin Winkeljann are co-founders of RNhale GmbH. Olivia Merkel is a Scientific Advisory Board Member of Coriolis Pharma, AMW, and Corden Pharma as well as a consultant for PARI Pharma, AbbVie Deutschland, and Boehringer-Ingelheim International.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

machine learning, poly(beta)aminoester, polyester, polymer discovery, polymeric nanoparticle, pulmonary delivery, siRNA delivery

Received: January 29, 2025

Revised: May 20, 2025

Published online:

- [1] C. E. Kosmas, A. Muñoz Estrella, A. Sourlas, D. Silverio, E. Hilario, P. D. Montan, E. Guzman, *Diseases* **2018**, *6*, 63.
- [2] L. Zhang, T. Wu, Y. Shan, G. Li, X. Ni, X. Chen, X. Hu, L. Lin, Y. Li, Y. Guan, J. Gao, D. Chen, Y. Zhang, Z. Pei, X. Chen, *Brain* **2021**, *144*, 3421.
- [3] R. Dahm, *Develop. Biol.* **2005**, *278*, 274.
- [4] W. Lou, L. Zhang, J. Wang, *Int. Immunopharmacol.* **2024**, *142*, 113157.
- [5] S. Omo-Lamai, Y. Wang, M. N. Patel, E.-O. Essien, M. Shen, A. Majumdar, C. Espy, J. Wu, B. Channer, M. Tobin, S. Murali, T. E. Papp, R. Maheshwari, L. Wang, L. S. Chase, M. E. Zamora, M. L. Arral, O. A. Marcos-Contreras, J. W. Myerson, C. A. Hunter, A. Tsourkas, V. Muzykantov, I. Brodsky, S. Shin, K. A. Whitehead, P. Gaskill, D. Discher, H. Parhiz, J. S. Brenner, *bioRxiv* **2024**.
- [6] P. Sharma, D. Hoorn, A. Aitha, D. Breier, D. Peer, *Adv. Drug Delivery Rev.* **2024**, *205*, 115175.
- [7] A. De, Y. T. Ko, *Exp. Opin. Drug Del.* **2023**, *20*, 175.
- [8] C. M. Jogdeo, K. Siddhanta, A. Das, L. Ding, S. Panja, N. Kumari, D. Oupický, *Adv. Mater.* **2024**, *36*, 2404608.
- [9] B. M. Cooper, D. Putnam, *ACS Biomater. Sci. Eng.* **2016**, *2*, 1837.
- [10] F. Xiao, Z. Chen, Z. Wei, L. Tian, *Adv. Sci.* **2020**, *7*, 2001048.
- [11] Z. Li, R. Guo, Z. Zhang, H. Yong, L. Guo, Z. Chen, D. Huang, D. Zhou, *J. Control Rel.* **2024**, *368*, 131.
- [12] Y. Rui, D. R. Wilson, S. Y. Tzeng, H. M. Yamagata, D. Sudhakar, M. Conge, C. A. Berlinicke, D. J. Zack, A. Tuesca, J. J. Green, *Sci. Adv.* **2022**, *8*, abk2855.
- [13] C. Yan, G. Li, *Adv. Intell. Syst.* **2023**, *5*, 2200243.
- [14] A. K. Chew, M. A. F. Afzal, A. Chandrasekaran, J. H. Kamps, V. Ramakrishnan, *Mach. Learn.: Sci. Technol.* **2024**, *5*, 045031.
- [15] D. Gong, D. Gong, E. Ben-Akiva, E. Ben-Akiva, A. Singh, A. Singh, H. Yamagata, H. M. Yamagata, S. Est-Witte, S. Est-Witte, J. K. Shade, J. K. Shade, N. A. Trayanova, N. A. Trayanova, J. J. Green, J. J. Green, *Acta Biomater.* **2022**, *154*, 349.
- [16] R. Kumar, N. Le, Z. Tan, M. E. Brown, S. Jiang, T. M. Reineke, *ACS Nano* **2020**, *14*, 17626.
- [17] D. Rogers, M. Hahn, *J. Chem. Inf. Model.* **2010**, *50*, 742.
- [18] D. Weininger, *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31.
- [19] S. Kearnes, K. McCloskey, M. Berndl, V. Pande, P. Riley, *J. Comput. Aided Mol. Des.* **2016**, *30*, 595.
- [20] R. Todeschini, V. Consonni, *Molecular Descriptors for Chemoinformatics: Volume I: Alphabetical Listing / Volume IVol. I: Appendices, References*, John Wiley & Sons, Hoboken, NJ **2009**.
- [21] T. Hastie, J. Friedman, R. Tibshirani, *In The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Eds.: T. Hastie, J. Friedman, R. Tibshirani), Springer, New York, NY **2001**, pp. 115.
- [22] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3146.
- [23] G. E. A. P. A. Batista, R. C. Prati, M. C. Monard, *SIGKDD Explor. Newsl.* **2004**, *6*, 20.
- [24] K. Wu, X. Yang, Z. Wang, N. Li, J. Zhang, L. Liu, *Briefings Bioinform.* **2024**, *25*, bbae186.
- [25] D. van Tilborg, A. Alenicheva, F. Grisoni, *J. Chem. Inf. Model.* **2022**, *62*, 5938.
- [26] K. M. Steinegger, L. Allmendinger, S. Sturm, F. Sieber-Schäfer, A. P. E. Kromer, K. Müller-Caspary, B. Winkeljann, O. M. Merkel, *Nano Lett.* **2024**, *24*, 15683.
- [27] Y. Jin, X. Wang, A. P. E. Kromer, J. T. Müller, C. Zimmermann, Z. Xu, A. Hartschuh, F. Adams, O. M. Merkel, *Biomacromolecules* **2024**, *25*, 4177.
- [28] Y. Jin, F. Adams, L. Isert, D. Baldassi, O. M. Merkel, *Mol. Pharmaceutics* **2023**, *20*, 4505.
- [29] C. J. Bishop, K. L. Kozielski, J. J. Green, *J. Controlled Release* **2015**, *219*, 488.
- [30] K. Itaka, A. Harada, Y. Yamasaki, K. Nakamura, H. Kawaguchi, K. Kataoka, *J. Gene Med.* **2004**, *6*, 76.
- [31] M. Breunig, C. Hozsa, U. Lungwitz, K. Watanabe, I. Umeda, H. Kato, A. Goepferich, *J. Controlled Release* **2008**, *130*, 57.
- [32] P. Y. Teo, C. Yang, J. L. Hedrick, A. C. Engler, D. J. Coady, S. Ghaem-Maghani, A. J. T. George, Y. Y. Yang, *Biomaterials* **2013**, *34*, 7971.
- [33] Z. Liu, Z. Zhang, C. Zhou, Y. Jiao, *Prog. Polym. Sci.* **2010**, *35*, 1144.
- [34] T. L. M. Thurston, M. P. Wandel, N. von Muhlinen, Á. Foeglein, F. Rando, *Nature* **2012**, *482*, 414.
- [35] S. Thalmayr, M. Grau, L. Peng, J. Pöhmerer, U. Wilk, P. Folda, M. Yazdi, E. Weidinger, T. Burghardt, M. Höhn, E. Wagner, S. Berger, *Adv. Mater.* **2023**, *35*, 2211105.
- [36] M. A. Beach, U. Nayanathara, Y. Gao, C. Zhang, Y. Xiong, Y. Wang, G. K. Such, *Chem. Rev.* **2024**, *124*, 5505.
- [37] Q. Wang, C. Bu, Q. Dai, J. Chen, R. Zhang, X. Zheng, H. Ren, X. Xin, X. Li, *Adv. Sci.* **2024**, *11*, 2309748.
- [38] W. Wang, Z. Huang, Y. Huang, X. Zhang, J. Huang, Y. Cui, X. Yue, C. Ma, F. Fu, W. Wang, C. Wu, X. Pan, *Adv. Drug Delivery Rev.* **2022**, *185*, 114309.
- [39] J. A. Whitsett, T. Alenghat, *Nat. Immunol.* **2015**, *16*, 27.
- [40] L. Ding, S. Tang, T. A. Wyatt, D. L. Knoell, D. Oupický, *J. Controlled Rel.* **2021**, *330*, 977.
- [41] O. M. Merkel, *J. Controlled Rel.* **2022**, *345*, 549.
- [42] C. De Souza Carvalho, N. Daum, C.-M. Lehr, *Adv. Drug Delivery Rev.* **2014**, *75*, 129.
- [43] P. Prasher, M. Sharma, S. K. Singh, M. Gulati, N. K. Jha, P. K. Gupta, G. Gupta, D. K. Chellappan, F. Zaccani, T. De Jesus Andreoli Pinto, Y. Chan, G. Liu, K. R. Paudel, P. M. Hansbro, B. G. George Oliver, K. Dua, *Chem.-Biol. Interact.* **2022**, *365*, 110048.
- [44] Q. Wang, C. Bu, Q. Dai, J. Chen, R. Zhang, X. Zheng, H. Ren, X. Xin, X. Li, *Adv. Sci.* **2024**, *11*, 2309748.
- [45] G. Liu, C. Betts, D. M. Cunoosamy, P. M. Åberg, J. J. Hornberg, K. B. Sivars, T. S. Cohen, *Respir. Res.* **2019**, *20*, 162.
- [46] C. M. Zimmermann, D. Baldassi, K. Chan, N. B. P. Adams, A. Neumann, D. L. Porras-Gonzalez, X. Wei, N. Kneidinger, M. G. Stoleriu, G. Burgstaller, D. Witzigmann, P. Luciani, O. M. Merkel, *J. Controlled Rel.* **2022**, *351*, 137.
- [47] O. M. Merkel, A. Beyerle, D. Librizzi, A. Pfestroff, T. M. Behr, B. Sproat, P. J. Barth, T. Kissel, *Mol. Pharmaceutics* **2009**, *6*, 1246.
- [48] Y. Xie, N. H. Kim, V. Nadithe, D. Schalk, A. Thakur, A. Kilic, L. G. Lum, D. J. P. Bassett, O. M. Merkel, *J. Controlled Rel.* **2016**, *229*, 120.
- [49] E. J. Chung, S. Kwon, J. L. Reedy, A. O. White, J. S. Song, I. Hwang, J. Y. Chung, K. Ylaya, S. M. Hewitt, D. E. Citrin, *Int. J. Radiation Oncol. Biol. Phys.* **2021**, *110*, 526.
- [50] C. Zhao, X. Fang, D. Wang, F. Tang, X. Wang, *Respiratory Med.* **2010**, *104*, 1391.
- [51] J. Kim, J. D. Minna, *Dev. Cell* **2022**, *57*, 292.
- [52] D. Baldassi, S. Ambike, M. Feuerherd, C.-C. Cheng, D. J. Peeler, D. P. Feldmann, D. L. Porras-Gonzalez, X. Wei, L.-A. Keller, N. Kneidinger, M. G. Stoleriu, A. Popp, G. Burgstaller, S. H. Pun, T. Michler, O. M. Merkel, *J. Controlled Rel.* **2022**, *345*, 661.
- [53] L. Xie, Y. Tan, Z. Wang, H. Liu, N. Zhang, C. Zou, X. Liu, G. Liu, J. Lu, H. Zheng, *ACS Appl. Mater. Interfaces* **2016**, *8*, 29261.
- [54] D. Duvenaud, D. Maclaurin, J. Aguilera-Iparraguirre, R. Gomez-Bombarelli, T. Hirzel, A. Aspuru-Guzik, R. P. Adams, *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 2224.
- [55] C. W. Coley, R. Barzilay, W. H. Green, T. S. Jaakkola, K. F. Jensen, *J. Chem. Inf. Model.* **2017**, *57*, 1757.
- [56] S. Y. Tzeng, J. J. Green, *Adv. Healthcare Mater.* **2013**, *2*, 468.
- [57] Y. Yan, K. Zhou, H. Xiong, J. B. Miller, E. A. Motea, D. A. Boothman, L. Liu, D. J. Siegwart, *Biomaterials* **2017**, *118*, 84.
- [58] C. Kim, A. Chandrasekaran, T. D. Huan, D. Das, R. Ramprasad, *J. Phys. Chem. C* **2018**, *122*, 17575.