Supporting Information

**Machine Learning-Enabled Polymer Discovery for Enhanced Pulmonary siRNA Delivery**

Felix Sieber-Schäfer1#, Min Jiang1#, Adrian Kromer1, Anny Nguyen1, Müge Molbay1, Simone P. Carneiro1, David Jürgens1, Gerald Burgstaller2, Bastian Popper3, Benjamin Winkeljann1,4, Olivia M. Merkel1,4,5\*

1Ludwig-Maximilians-Universität München, Department of Pharmacy, Butenandtstraße 5

81377 Munich, Germany, Member of the German Center for Lung Research (DZL), Munich, Germany

2Helmholtz Munich, Comprehensive Pneumology Center (CPC-M), Max-Lebsche-Platz 31, 81377 Munich, Germany, Member of the German Center for Lung Research (DZL), Munich, Germany

3Ludwig-Maximilians-Universität München, Biomedical Center, Großhaderner Str. 9, 82152 Planegg- Martinsried

4Center for Nanoscience (CeNS) Munich,

5Cluster for Nucleic Acid Therapeutics Munich (CNATM), Munich, Germany

#Both authors contributed equally

\*Corresponding author:

Olivia Merkel

LMU Munich, Department of Pharmacy

Butenandtstraße 5

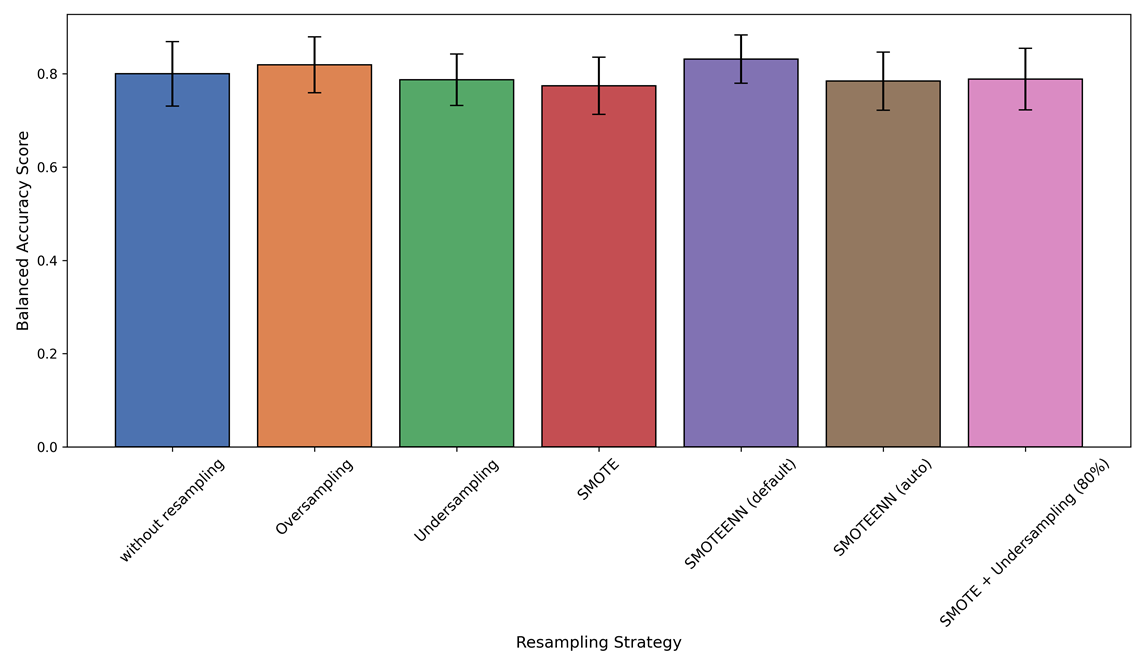
81377 Munich

Germany

[Olivia.merkel@lmu.de](mailto:Olivia.merkel@lmu.de)



**Figure S1**. **Selection of Machine Learning Algorithm** A) Comparison of different algorithms B) Averaged Confusion Matrix for default LightGBMClassifier

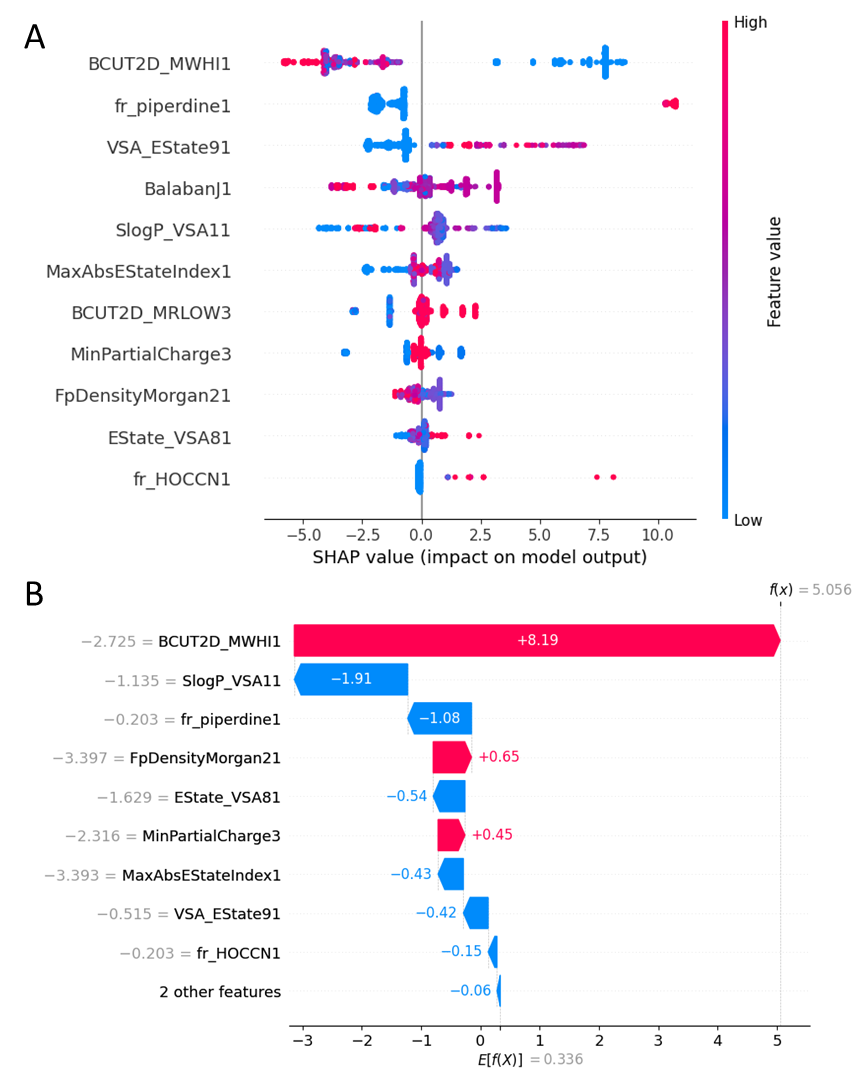


**Figure S2.** Comparison of different resampling strategies to handle the unbalanced dataset.

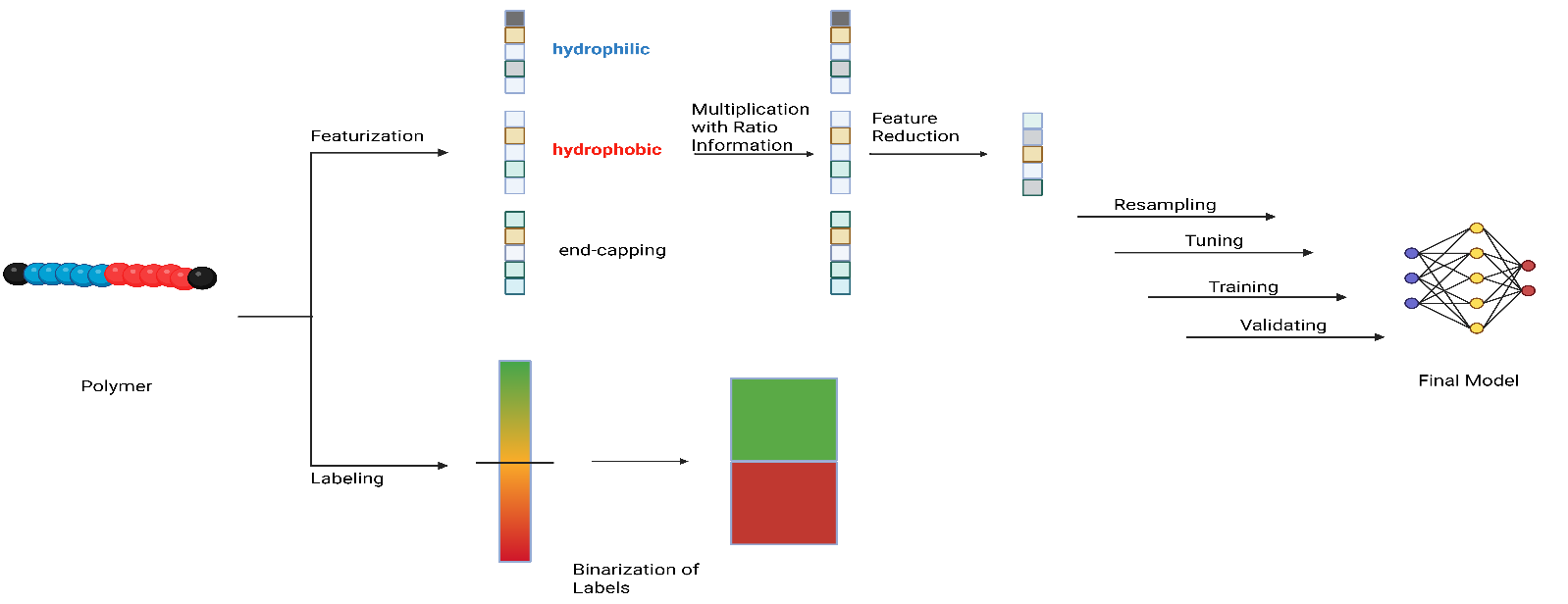
**Hyperparametertuning using hyperopt:**

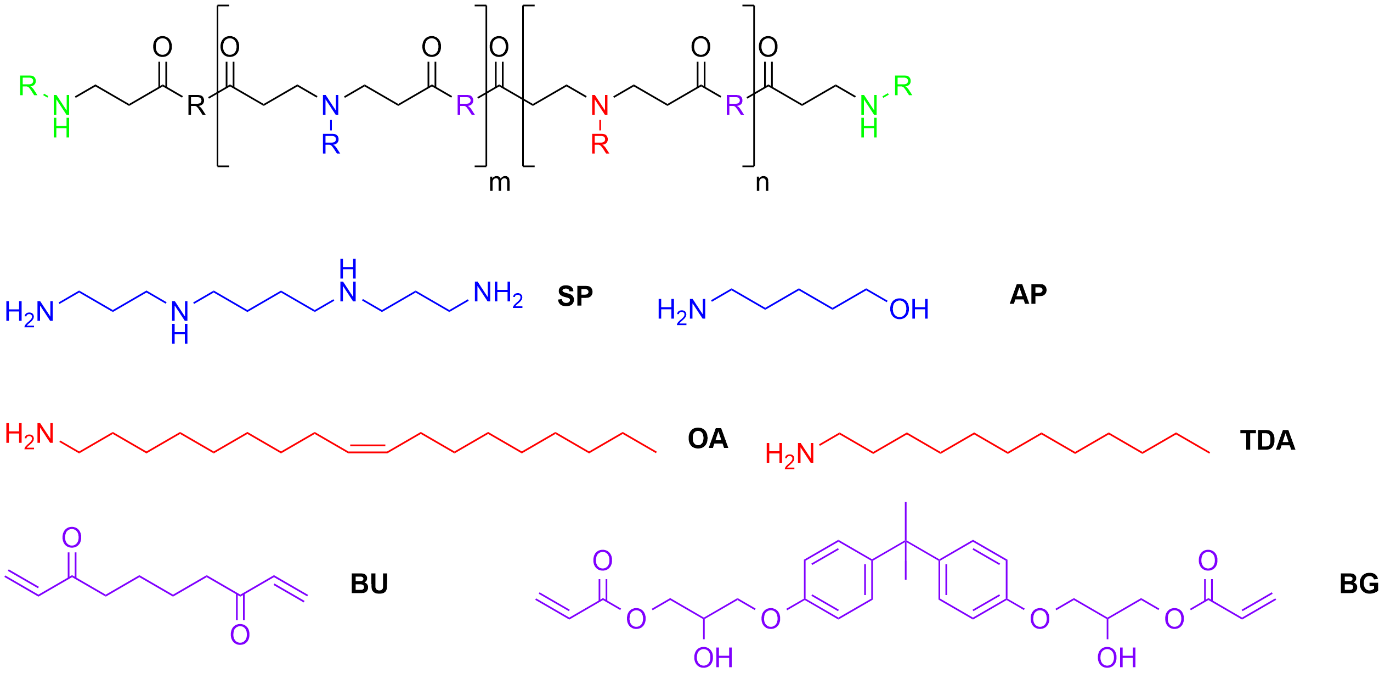
100 evaluations were performed using the mean of 10 train test splits with replacement as objective. The hyperparameters are the following {'colsample\_bytree': 0.9755088786798269, 'learning\_rate': 0.1827587842746705, 'max\_depth': 8, 'n\_estimators': 465, 'num\_leaves': 85, 'reg\_lambda': 0.8324249896997891, 'subsample': 0.9331718683905172}

**Figure S3.** Hyperparameter code for LGBMClassifier.

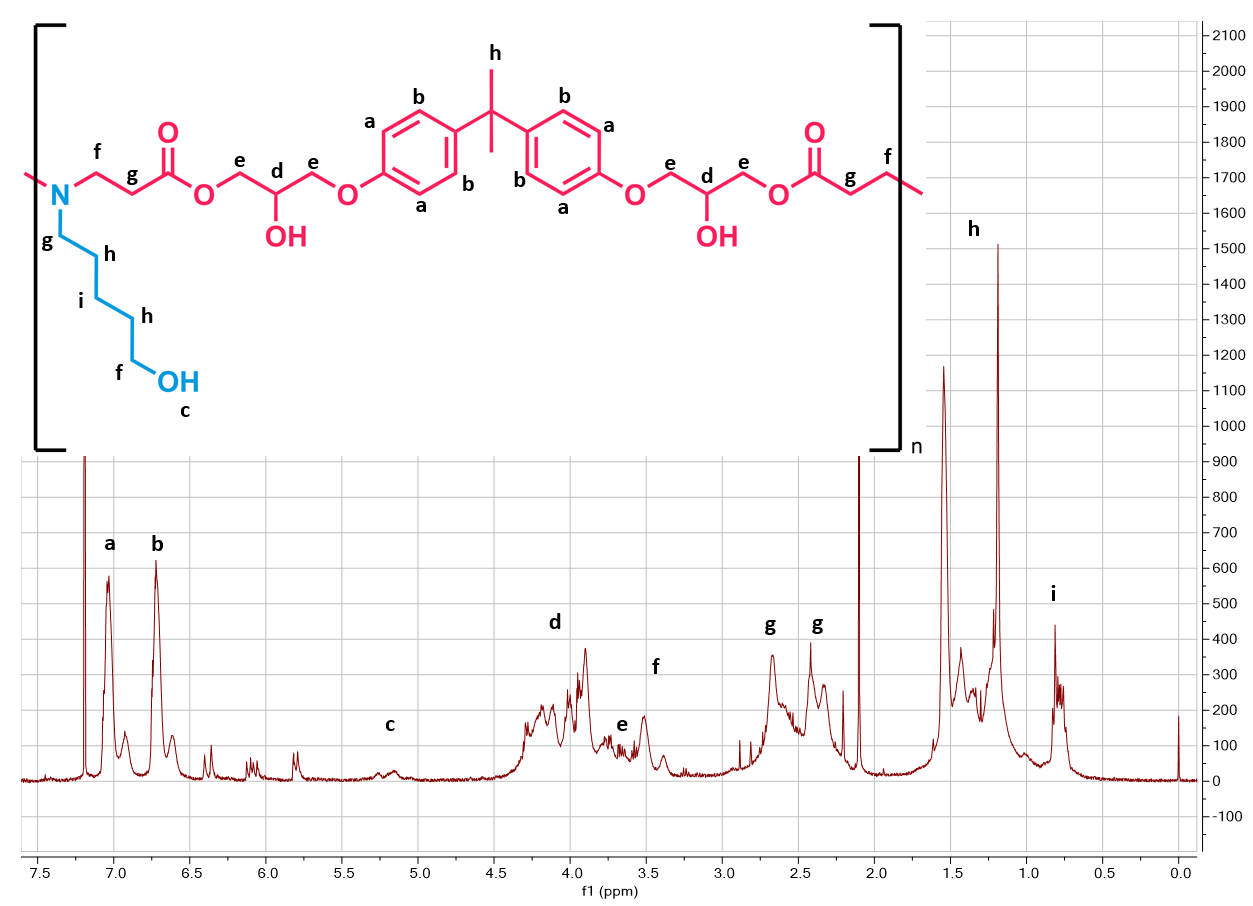


**Figure S4.** SHAP values of the optimized model on the whole dataset (A) and for the predicted polymer as single point prediction (B).

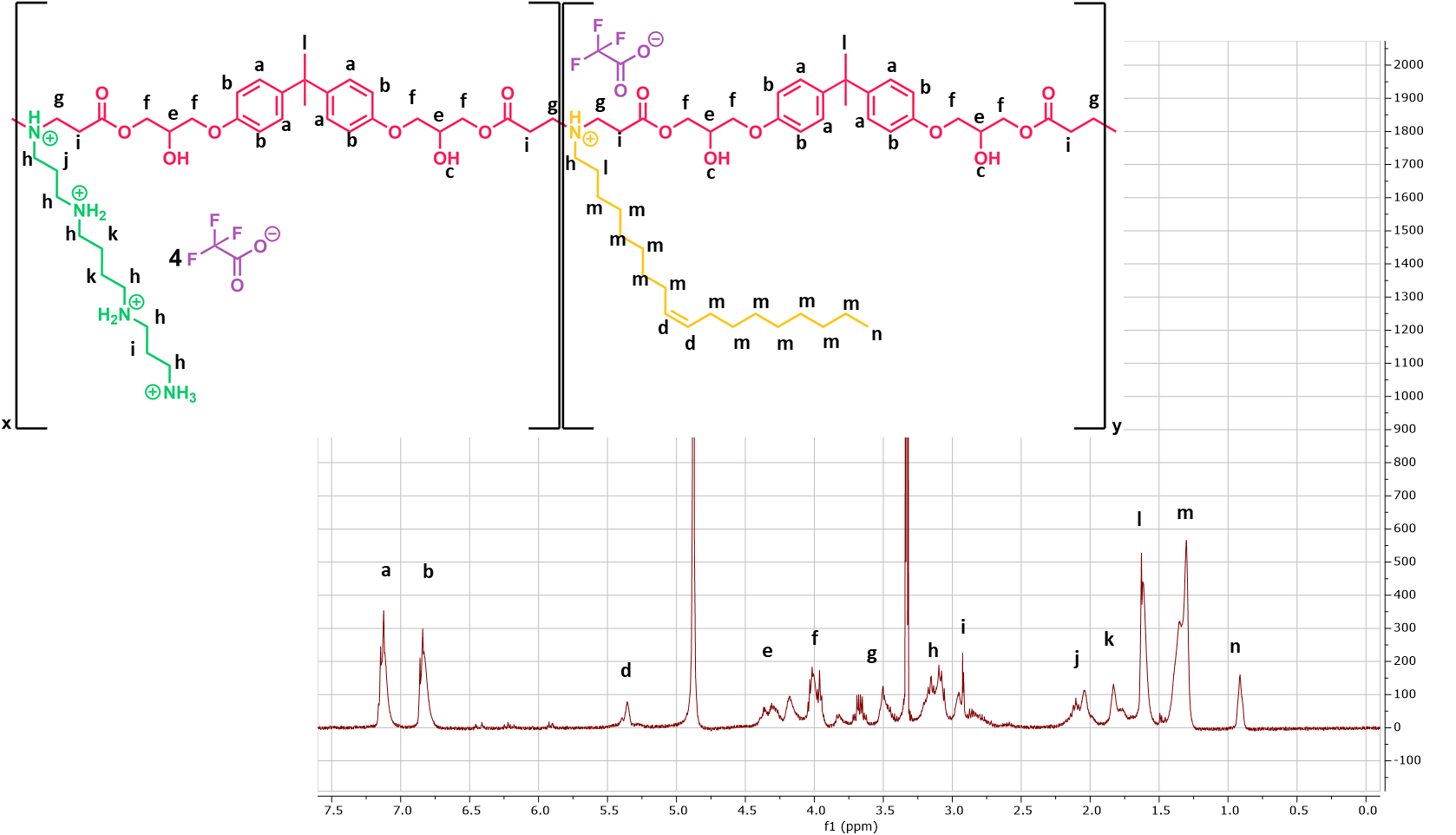
**Figure S5.** Overview of the machine learning workflow.



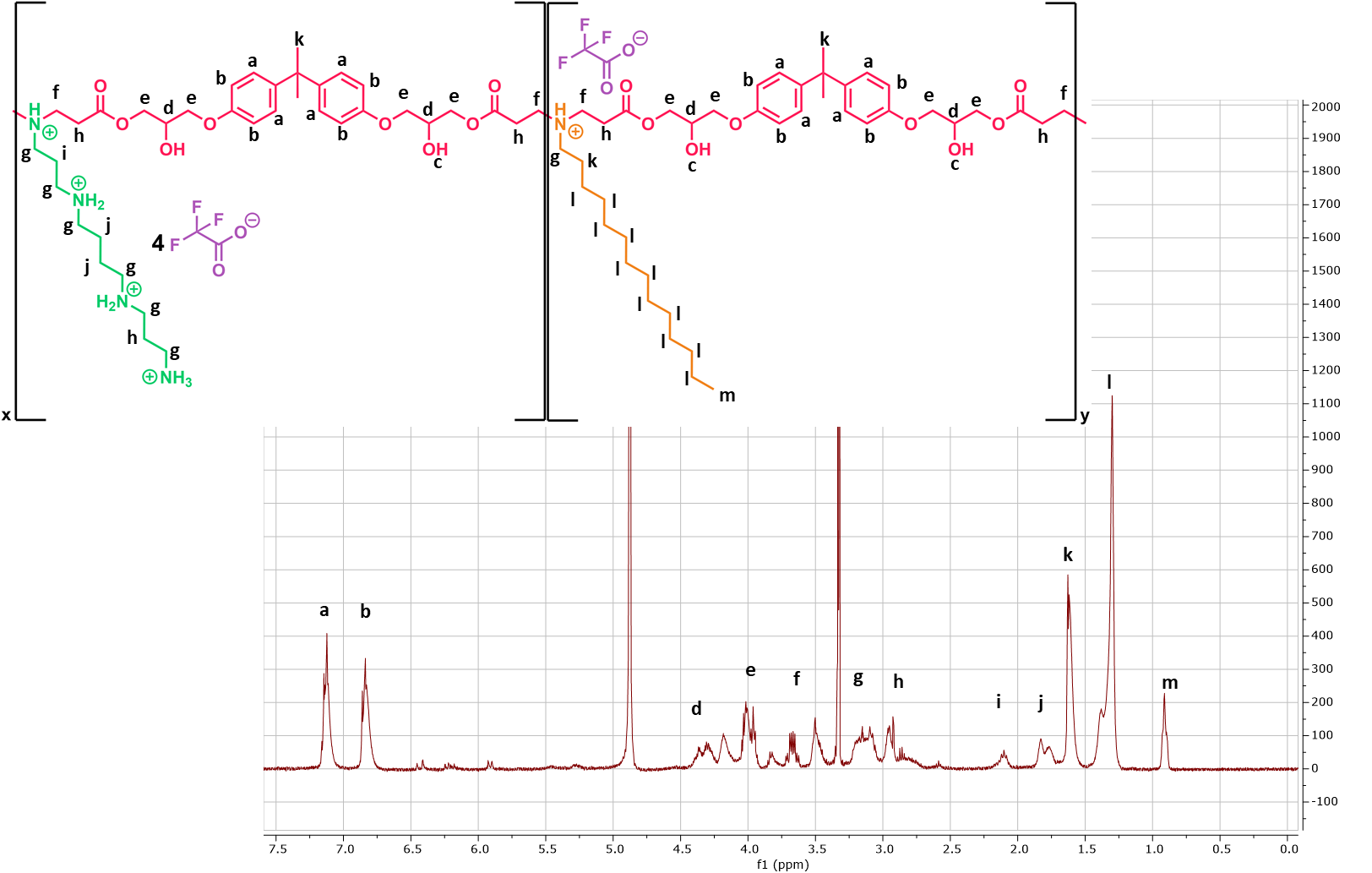
**Figure S6.** Monomers used to design the Validation Set. SP, spermine; AP, 4-Amino-1-butanol; OA, oleylamine; TDA, tetradecylamine; Bu, 1,4-butanediol diacrylate; BG, bisphenol A glycerolate.



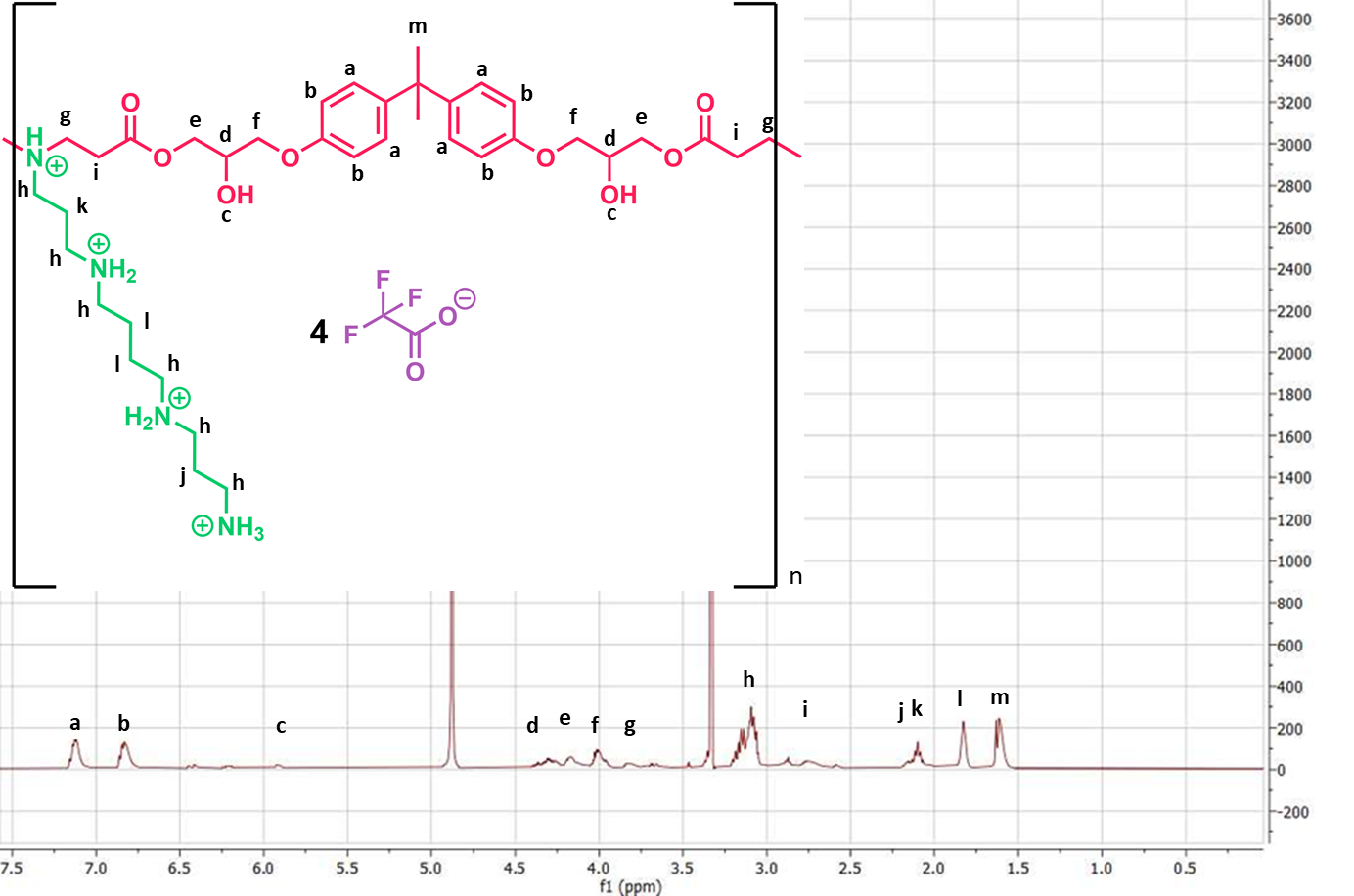
**Figure S7.** 1H-NMR measurement of validation polymer AP-BG.



**Figure S8.** 1H-NMR measurement of validation polymer SP-OA-BG



**Figure S9.** 1H-NMR measurement of validation polymer SP-TDA-BG

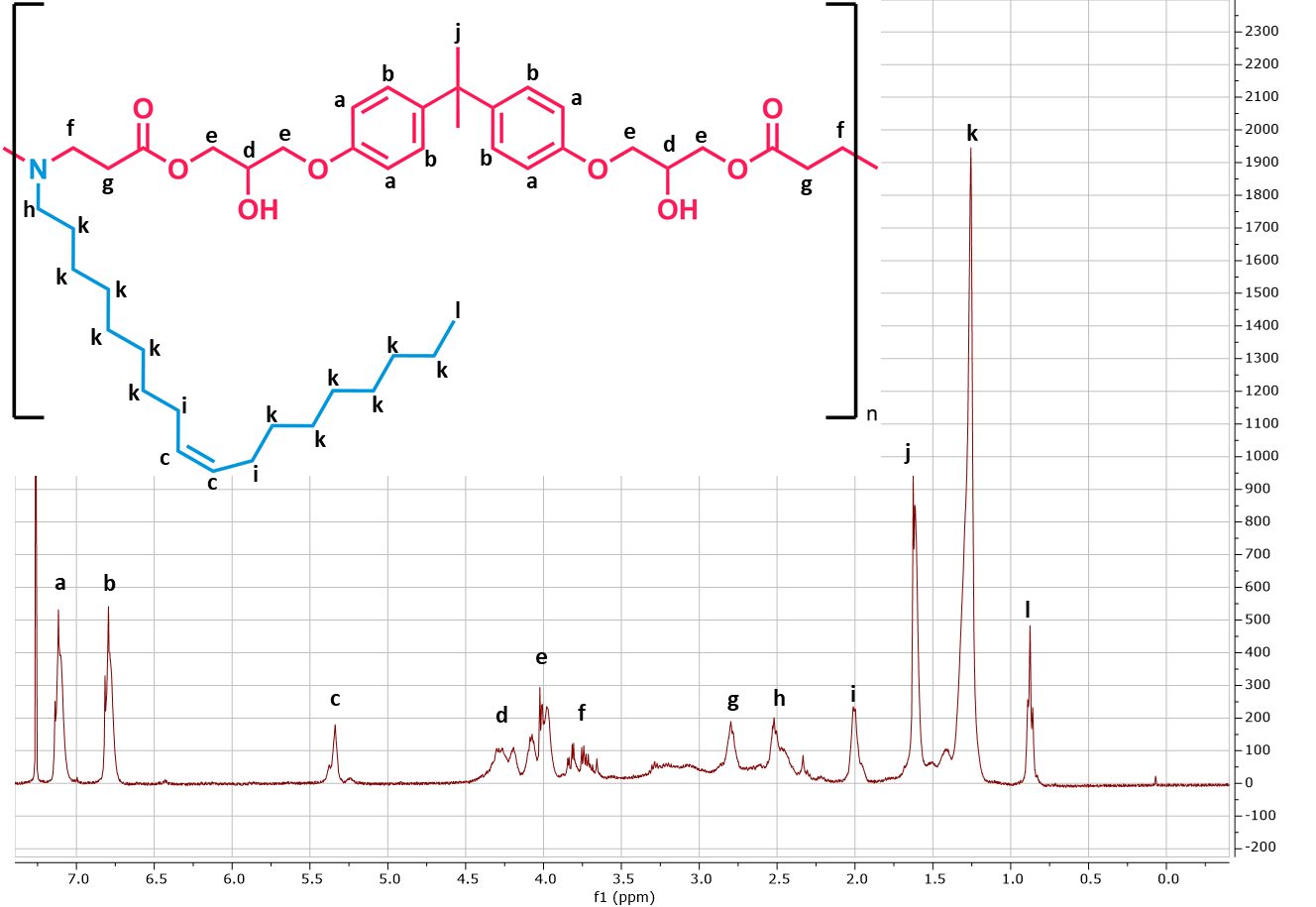


**Figure S10.** 1H-NMR measurement of validation polymer SP-BG

图示, 示意图

AI 生成的内容可能不正确。

**Figure S11.** 1H-NMR measurement of validation polymer SP-BU



**Figure S12.** 1H-NMR measurement of validation polymer OA-BG.

**图表, 条形图

AI 生成的内容可能不正确。**

**Figure S13. Characterization of siRNA PBAEs polyplexes.** (A) Hydrodynamic diameter (represented by bar graph), polydispersity (represented by symbol) and (B) Zeta potential of siRNA-loaded polyplexes formulated at N/P ratio of 10. (C) siRNA encapsulation efficiency in the polyplexes formulated at N/P 10 with different polymers. Data are presented as mean ± SD, n=3.

图表, 条形图

AI 生成的内容可能不正确。

**Figure S14. *In Vitro* gene silencing efficiency.** (A) Enhanced green fluorescent protein (eGFP) knockdown efficiency of siRNA polyplexes formulated at an N/P ratio of 10 in H1299/eGFP cells. (B) Epidermal growth factor receptor (EGFR) knockdown efficiency of siRNA polyplexes in A549 cells. Data are presented as mean ± SD, n=3.

图片包含 船, 小, 不同, 充满

AI 生成的内容可能不正确。

**Figure S15. 1H-NMR of the synthesized structures** A) SP0.6/0.4TDA B) SP0.5/0.5TDA C) SP0.4/TDA0.6.

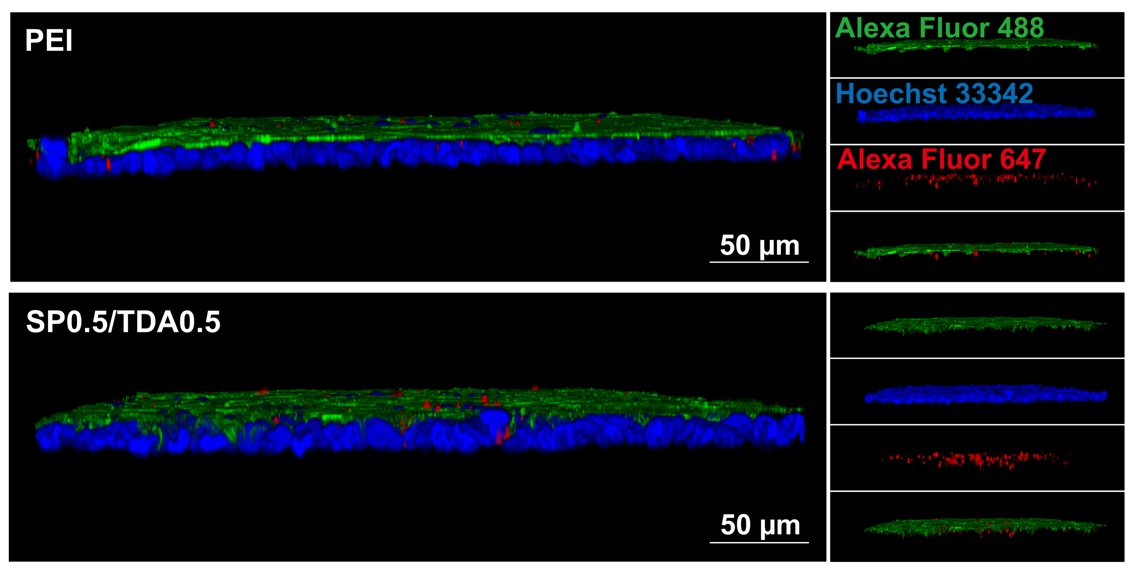
图表, 折线图, 直方图

AI 生成的内容可能不正确。

**Figure S16.** GPC measurement of SP0.5/TDA0.5 which was tested *in vivo*.

**图表, 直方图

描述已自动生成Figure S17.** siRNA encapsulation efficiency in the polyplexes prepared at different N/P ratios. Data are presented as mean ± SD, n=3.

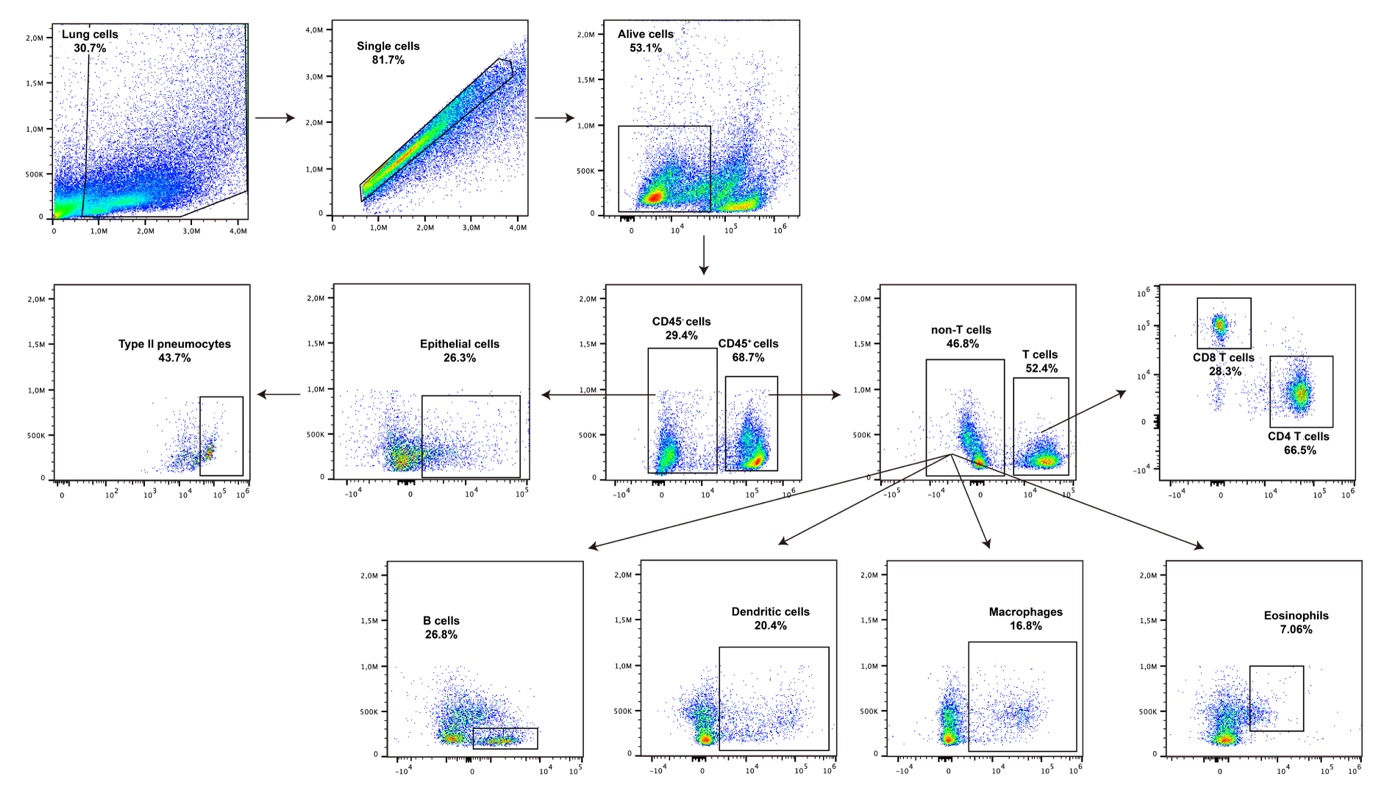


**Figure S18**. Mucus penetration assay of siRNA-loaded PEI 25kDa polyplexes and PBAE SP0.5/TDA0.5 polyplexes in air-liquid interface (ALI) culture of Calu-3 cells. Scale bar, 50 μm.

电脑游戏的截图

中度可信度描述已自动生成

**Figure S19.** Organ distribution after intratracheal instillation of free Alexa Fluor 647-labeled siRNA or siRNA-loaded polyplexes.



**Figure S20.** Gating strategy of different cell types in the lung.