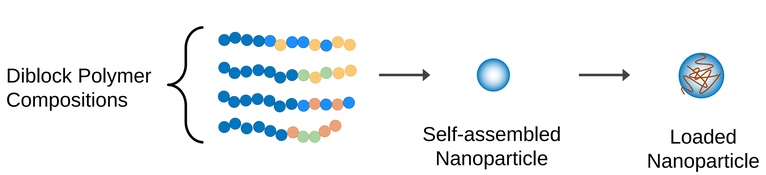
Machine Learning-derived Insights in Design of Small Polymeric Nanoparticles for Gene Delivery

Caleb Hillrich, Andrew Matas, Ken Sims, Emma Schmitz, Gerrit Bryan

INTRODUCTION

Polymer nanoparticles are promising gene delivery vehicles due to their high chemical diversity and tunability. A major challenge in the field is identifying novel compositions with properties of interest. In particular, small particle size (< 200 nm) is important for cellular uptake and transfection [1]. However, loading nanoparticles with a genetic payload can drastically change particle size. Predictive methods are needed to model which compositions remain small after introducing the payload.

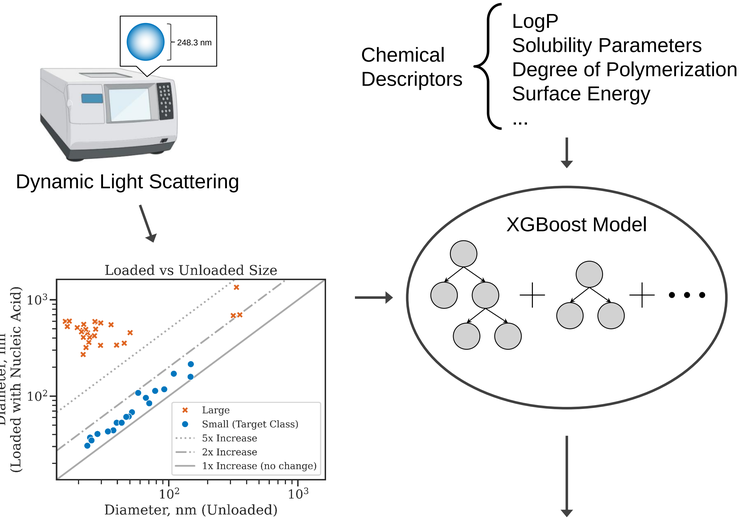


METHODS

A diverse library of 100 diblock polymer nanoparticles was synthesized using reversible addition-fragmentation transfer (RAFT) polymerization of 7 monomers ranging in hydrophobicity and pKa. Each composition consisted of one cationic block to enable binding to a genetic payload, and one core block consisting of hydrophobic and pH-responsive monomers. Polymers were synthesized in triplicate using a high-throughput screening platform. Particle size and polydispersity were characterized using dynamic light scattering both before and after loading with a genetic payload. An eXtreme Gradient Boosting (XGBoost) classifier [2] was trained to predict which compositions result in size < 200 nm pre-loading and less than doubling in size post-loading. Feature importance was used to identify valuable trends.

RESULTS

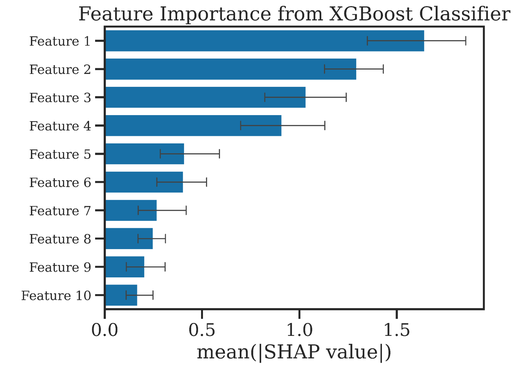
Clear separation was seen between composition groups that remained small after loading and those that increased in size by a factor of 2 or more. An XGBoost classifier was successfully trained using ten-fold cross-validation and achieved good fit (94% accuracy), indicating that selected features contain sufficient signal. Interpretation of Shapley Additive exPlanations (SHAP) values [3][4] successfully quantified how each chemical descriptor affected predictions.



Metric	Value
Accuracy	0.94
Matthews Correlation Coefficient	0.88
Binary F1 Score	0.93

CONCLUSIONS

Key physicochemical properties were identified which correlate with size classes, providing general rules for designing polymer nanoparticles with small diameter post-loading. These insights will accelerate the discovery of viable gene delivery vehicles. The present work demonstrates a powerful and interpretable approach to in silico prediction of polymer nanoparticle properties.



REFERENCES

1. Prabha S et. al. Artificial Cells, Nanomedicine, and Biotechnology, 44(1), 83–91.
2. Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 785-794).
3. Shapley, L. S. (1953). A value for n-person games. Contribution to the Theory of Games, 2.
4. Rozemberczki, B., Watson, L., Bayer, P., Yang, H. T., Kiss, O., Nilsson, S., & Sarkar, R. (2022). The shapley value in machine learning. arXiv preprint arXiv:2202.05594.