CellPress

Matter
Previews

Machine learning makes magnificent macromolecules for medicine

Veronica Cunitz,[1,2,3] Evan Stacy,[1,3] Penelope Jankoski,[1,3] and Tristan Clemons[1,*]

At the University of Minnesota, scientists explore the application of machine learning to screen a multiparametric library of polymers to investigate the relationship between polymer attributes, payload type, and biological outcomes to optimize polymeric vector development for delivery of nucleic acid payloads.

With the incidence of global disease on the rise, there is a pressing need to advance treatment options that improve therapeutic bioavailability and efficacy. In recent years, gene therapeutics have been utilized for the targeted treatment of diseases including sickle cell anemia,[1] spinal muscular atrophy,[2] and cancer.[3] However, clinical translation of gene therapies has been slowed by a lack of safe and effective delivery vectors for gene-based therapeutics. In an effort to address concerns associated with these treatment methods, Professor Theresa Reineke and her team at the University of Minnesota have made great advances in the field of macromolecular polyplexes for gene delivery and protection. In a recent publication in the open access journal *JACS Au*, Kumar et al. applied machine learning methods to optimize polymer composition for the co-delivery of nucleic acid payloads, resulting in enhanced transfection rates and cellular uptake.[4] This involved the application of combinatorial polymer design, and high-throughput experimentation to develop an extensive polymer library. The multiparametric library incorporated cationic monomers of varying basicity copolymerized at different incorporation ratios (100%, 75%, 50%, and 25%) with neutral monomers of varying hydrophilicity. Machine learning was then used to systematically model several scenarios, allowing the most probable outcome to be determined based upon a subset of tests that gauge the interactions of each polymeric variable. By analyzing many experiments in tandem, it was possible to streamline the identification of polyplexes that are most likely to succeed in cell testing.

Currently, clinically available gene therapeutics heavily rely on engineered viral vectors for delivery of nucleic acid payloads. However, effective treatment requires the viral vector to first overcome a series of extracellular challenges such as enzymatic degradation, cellular binding and evasion of the immune system.[5] This is compounded by the presence of several intracellular barriers including endocytosis, endosomal escape, intracellular transport, unpackaging and nuclear uptake that must also be circumvented for efficacy.[6] To address some of these challenges associated with viral delivery, polymers have proven to be exciting candidates for delivering nucleic acid payloads due to their scalability, versatility, and degree of tuneability, while simultaneously minimizing immune response and cytotoxicity.

Polyplexes are one of the most extensively studied classes of nonviral gene delivery vectors, in which polycations condense DNA through electrostatic interactions to form particles on the order of 100 nm in size.[7] To enhance the efficacy of polycations for delivering nucleic acid payloads, many important design parameters must be considered. Of these parameters, special consideration must be given to the polymer architecture, the molecular weight, the identity of the cationic monomer, the inclusion of functionalized moieties, and the identity of the nucleic acid cargo. Nine intercorrelated physicochemical features have been determined to influence the activity of the polymeric vector: the molecular weight (Mn), the hydrophobicity, the pK_a, the Hill coefficient (n_{Hill}), the percent cationicity, the surface charge (ζ), the polyplex radius (R_h), the ratio of protonatable nitrogen atoms relative to phosphate groups in the gene cargo (N/P ratio), and the mobility of plasmids (pDNA) during gel electrophoresis. By applying machine learning, the structure-function relationships between these descriptors and transfection efficacy, cellular toxicity, and cellular uptake can be quantified to identify the predictive and causal importance of each polyplex descriptor on effective nucleic acid delivery.

Exogenous nucleic acids can be delivered in the form of mRNA, short interfering RNA (siRNA), plasmid DNA (pDNA), antisense oligonucleotides (ASO), ribonucleoproteins (RNP), self-amplifying RNA or replicon RNA (saRNA or repRNA), and microRNA. However, it has become apparent that current design frameworks fail to consider the biochemical differences between these varying nucleic acid modalities, which differ in their size, topology, and mechanism of action. Recent work in the field has challenged the "one-size-fits-all" approach to

[1]School of Polymer Science and Engineering, The University of Southern Mississippi, Hattiesburg, MS 39406, USA

[2]Department of Chemical Engineering, The University of Mississippi, Oxford, MS 38677, USA

[3]These authors contributed equally

*Correspondence: tristan.clemons@usm.edu

https://doi.org/10.1016/j.matt.2022.07.008

designing polyplexes, showing that polycations employed for mRNA delivery could not effectively deliver pDNA without making modular changes in monomer chemistry.[8] Hence, there is an inherent need to determine the principal criteria for creating polymeric gene delivery vectors based upon their nucleic acid payload.

In previous work, the Reineke group utilized Reversible Addition-Fragmentation Chain Transfer (RAFT) polymerization to develop a well-defined library of polymers with promising polymer architectures and compositions that improve complexation and delivery of nucleic acid payloads.[9] It was demonstrated through this work that variation of polymer pK_a and the balance between hydrophilic and hydrophobic groups through different copolymerization strategies alters the outcome of the polyplex in terms of size, RNP binding affinity, and successful cell transfection. Recently in *JACS Au*, the Reineke group revisited this library of polymers and focused on 43 distinct polymers to determine their efficiency in delivery of pDNA payloads (Figure 1).[4] It was initially predicted that the copolymer of 2-(diisopropylamino)ethyl methacrylate and 2-hydroxyethyl methacrylate, poly(DIPAEMA$_{52}$-*st*-HEMA$_{50}$), labeled as P38 in the study, would have optimized bioactivity and residence time for pDNA. The Reineke group applied machine learning to determine whether P38, along with the other polymers in the library, were suitable vectors for the co-delivery of RNP and pDNA to facilitate homology-directed repair (HDR) and to evaluate whether the design constraints restricting RNP payloads are equally applicable to pDNA payloads.

Polyplex characterization data was paired with pDNA delivery screening data to elucidate mechanistic insights and pDNA-specific structure-function relationships. The delivery of pDNA was investigated through evaluation of cellular internalization and quantitative confocal microscopy with Cy5-labeled pDNA. Cy5 fluorescence intensity was measured via flow cytometry after 24 h to determine the level of pDNA internalization, however, this was not indicative of effectual delivery of payloads. In contrast, green fluorescent protein (GFP) expression with cellular uptake for the polymer carriers in the library highlighted underlying cellular mechanistic differences. Visualization of the various payloads highlighted the differences in polyplex localization within the cell. From these experiments it was seen that P38 polyplexes are less likely to colocalize within lysosomal compartments, thus the payload is protected and effectively steered into the perinuclear region. Further, substantial elevation of transgene expression was found in eight of the library's polyplexes when delivering pDNA payloads, compared to three for effective RNP delivery. This suggested the structure-function relationship of pDNA and RNP payloads are fundamentally different.

Machine learning was applied to quantify the importance of nine physicochemical variables (*vide supre*) and their impact on delivery of pDNA and RNP payloads. The machine learning utilized applies both predictive and causal importance to interpret and to explain the biological outcomes based on variations in the polymeric delivery vectors. This methodology allows for decoupling of known features in the data to determine which have standalone causal effects. Furthermore, feature importance estimates were improved upon by utilizing a game-theoretic approach instead of forest clusters to reduce overestimates of importance. This important work highlights the primary differences in effective polymeric delivery of RNP and pDNA therapeutic payloads. The RNP delivery was correlated with the hydrophobic and electrostatic interactions of the polyplex, whereas the pDNA delivery efficiency is predicted primarily by the protonation of the polycation.

Despite RNP and pDNA having divergent constraints for cellular uptake, P38 was isolated to be a unique delivery system that satisfies the design criteria of both structures. This is an important result, as successful HDR requires the co-delivery of two nucleic acid payloads: single guide RNA (sgRNA) and the GFP pDNA donor. The feasibility of co-delivery was then examined by monitoring the effects of total nucleic acid dose, payload composition, and N/P ratio. The maximization of GFP expression is evident of higher HDR frequencies in the cell. This occurs when the payload contains the maximum dose of nucleic acid, has an intermediate payload composition (defined as 1:2 and 1:3 w/w sgRNA:pDNA), and has an N/P ratio of 2. The HDR performance was evaluated by comparing the percentage of GFP-positive cells when P38 was employed as the delivery vector and compared to the commercial reagents JetPEI and Lipofectamine 2000. It was observed that JetPEI results in nearly almost no HDR-edited cells, the expression of GFP does not exceed 0.7% when HDR constructs are delivered by P38, and only Lipofectamine 2000 results in more than 2% of the cell population becoming GFP-positive. The authors hypothesize that the low frequency of HDR when P38 is the delivery vector is a result of cellular processes rather than polymeric design. They believe this could be further improved by synchronizing transfection with the cell cycle or employing HDR-promoting drugs to bias editing pathways in favor of gene insertion. Despite the low levels of HDR experienced, a pool of HDR-edited cells were obtained which can be expanded to meet therapeutic demands, confirming P38 as viable in a proof-of-concept study for HDR.

This work of Kumar et al. demonstrates that although there is a definitive

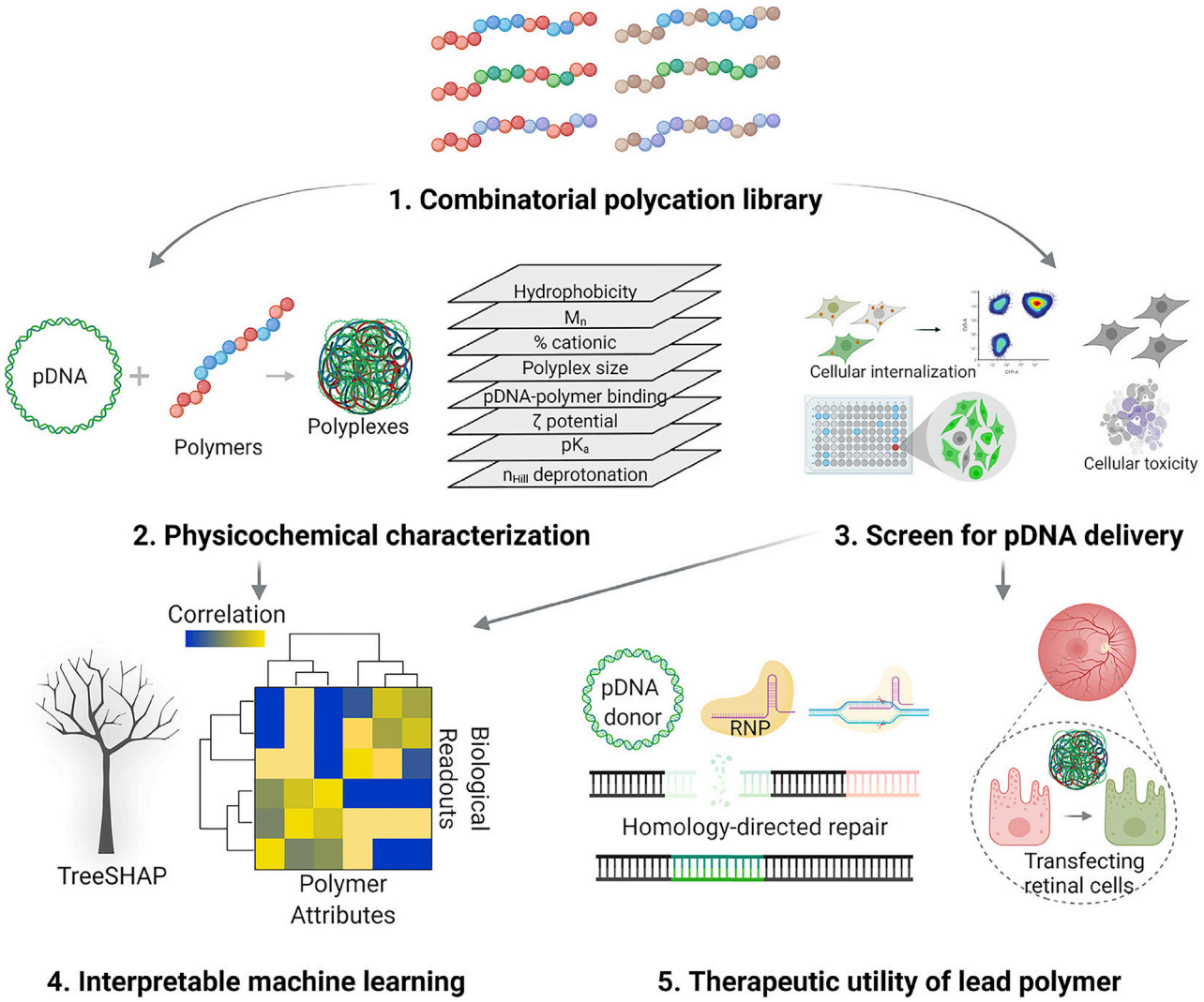

Figure 1. Using machine learning to assess polyplex efficacy

A combinatorial library of polymers was rapidly screened after complexation with pDNA payloads using interpretable machine learning approaches to further elucidate the structure-function relationships that enhance polyplex therapeutic utility. Reproduced from Kumar et al.[4]

difference in design requirements for delivering pDNA as compared to RNP, it is possible to achieve polymer compositions, such as P38, that fulfill the delivery requirements of both payloads. Utilization of machine learning allowed for optimization of design constraints for the narrowing of a chemically diverse library for novel polymeric vectors. Within polymeric delivery systems, this allows for the elucidation of payload-specific constraints and structure-function relationships. The isolation of factors that affect pDNA and RNP differently is a breakthrough that will aid the development of gene therapies. The understanding of the importance and the capability of precise tuneability of polymer physicochemical properties provides for a wide array of possibilities for further development of polyplexes. The results of this study opens the door for future exploration and optimization of various polycation-payload systems, as well as possibilities of applications of machine learning approaches for down-selecting to various systems with interdependent variables. Furthermore, optimization of the payload and relationship between polymer pDNA binding affinity will help propel polymeric vectors closer to clinical translation and continue to improve the efficacy of polyplexes for gene therapy, an important tool in modern medicine.

ACKNOWLEDGMENTS

V.C. and T.C. acknowledge funding support from the National Science Foundation under both Award No. 1757220 and 1950387. T.C. also thanks the Director of the School of Polymer Science and Engineering, the Dean of the College of

Arts and Sciences, and the vice President for Research, all at the University of Southern Mississippi, for their support with generous start-up funds.

REFERENCES

1. Wilson, R.C., and Gilbert, L.A. (2018). The Promise and challenge of *in Vivo* delivery for Genome therapeutics. ACS Chem. Biol. *13*, 376–382. https://doi.org/10.1021/acschembio.7b00680.

2. Cross, R. (2017). Gene therapy Came of Age. C&EN Global Enterprise *95*, 36. https://doi.org/10.1021/cen-09548-cover8.

3. Shen, T., Zhang, Y., Zhou, S., Lin, S., Zhang, X.B., and Zhu, G. (2020). Nucleic acid Immunotherapeutics for cancer. ACS Appl. Bio Mater. *3*, 2838–2849. https://doi.org/10.1021/acsabm.0c00101.

4. Kumar, R., Le, N., Oviedo, F., Brown, M.E., and Reineke, T.M. (2022). Combinatorial polycation Synthesis and causal machine learning Reveal divergent polymer design Rules for effective pDNA and ribonucleoprotein delivery. JACS Au *2*, 428–442. https://doi.org/10.1021/jacsau.1c00467.

5. Zhang, Y., Satterlee, A., and Huang, L. (2012). In vivo gene delivery by nonviral vectors: Overcoming Hurdles? Mol. Ther. *20*, 1298–1304. https://doi.org/10.1038/mt.2012.79.

6. Kumar, R., Santa Chalarca, C.F., Bockman, M.R., Bruggen, C.V., Grimme, C.J., Dalal, R.J., Hanson, M.G., Hexum, J.K., and Reineke, T.M. (2021). Polymeric delivery of therapeutic nucleic acids. Chem. Rev. *121*, 11527–11652. https://doi.org/10.1021/acs.chemrev.0c00997.

7. Burke, R.S., and Pun, S.H. (2008). Extracellular barriers to *in Vivo* PEI and PEGylated PEI polyplex-Mediated gene delivery to the Liver. Bioconjugate Chem. *19*, 693–704. https://doi.org/10.1021/bc700388u.

8. Kaczmarek, J.C., Patel, A.K., Rhym, L.H., Palmiero, U.C., Bhat, B., Heartlein, M.W., DeRosa, F., and Anderson, D.G. (2021). Systemic delivery of mRNA and DNA to the Lung using polymer-Lipid Nanoparticles. Biomaterials *275*, 120966. https://doi.org/10.1016/j.biomaterials.2021.120966.

9. Kumar, R., Le, N., Tan, Z., Brown, M.E., Jiang, S., and Reineke, T.M. (2020). Efficient polymer-Mediated delivery of gene-editing ribonucleoprotein payloads through combinatorial design, Parallelized experimentation, and machine learning. ACS Nano *14*, 17626–17639. https://doi.org/10.1021/acsnano.0c08549.