DISCUSSION PAPER

Using AI and Machine Learning to Advance Nucleic Acid Therapeutic Delivery Beyond the Liver

Vikas Kumar, PhD, University of Connecticut; **Sai Pallavi Pradeep, MSc,** University of Connecticut; **W. Mark Saltzman, PhD,** Yale University and Yale University School of Medicine; **José E. Manautou, PhD,** University of Connecticut; and **Raman Bahal, PhD,** University of Connecticut

November 16, 2026

ABSTRACT | *Researchers can now design nucleic acid medicines for specific genetic targets, but delivering those medicines to the right cells and tissues remains a major challenge. Advances in nucleic acid therapeutics have greatly expanded treatment options for various genetic and metabolic diseases, especially those affecting the liver. Beyond chemical stabilization of nucleic acids, much of the progress has come from achieving efficient, cell-specific delivery. A key factor in this success is the development of chemically engineered bioconjugates, such as the N-acetylgalactosamine ligand, which binds selectively to hepatocyte receptors in the liver. This allows for highly efficient delivery of small interfering RNA and antisense oligonucleotides. However, replicating this delivery efficiency in tissues outside the liver (extrahepatic) remains a major challenge. The main obstacle is the lack of well-defined ligand–receptor pairs that mediate cell-specific recognition and uptake in extrahepatic organs. This discussion paper examines how artificial intelligence and machine learning can accelerate the discovery of new chemical ligands for targeted nucleic acid delivery to tissues outside the liver. Taken together, the methods these tools support could enable targeting of organs that are difficult to reach with nucleic acid-based therapies and broaden treatment options for certain cancers and metabolic disorders for which additional treatments are needed.*

Expanding the Reach of Nucleic Acid Therapeutics

Nucleic acid therapeutics are a rapidly growing class of medicines that modulate gene expression by targeting RNA or DNA sequences. Nucleic acid drugs (NADs), including small interfering RNAs (siRNAs) and antisense oligonucleotides (ASOs), have shown significant therapeutic potential by silencing or degrading disease-associated messenger RNAs, thereby reducing protein production. Advances in nucleic acid chemistry and delivery technologies have partly driven their clinical success by helping overcome major developmental challenges, including physiological instability, immune responses, and barriers to intracellular delivery. Delivering siRNAs and ASOs into cells is challenging because these molecules are too large and too polar to cross the cell membrane on their own. Unlike traditional small-molecule drugs, which are typically under 500 daltons and can diffuse across membranes, nucleic acids primarily enter cells via endocytic uptake (Dowdy, 2017; Juliano, 2016; Lipinski et al., 2001).

For safe and effective nucleic acid therapy, NADs must reach the appropriate organ and cell type, which necessitates the development of specialized, targeted delivery strategies. Naked or directly delivered siRNA is quickly eliminated, while chemically modified ASOs distribute widely across the body, which is not ideal for cell-selective

precision medicine (Geary et al., 2015). To overcome these challenges, delivery systems such as lipid nanoparticles (LNPs) and cell-targeting ligands that facilitate nucleic acid uptake have emerged as crucial tools. These approaches have proven successful, particularly in liver-targeted therapies, where both carbohydrate-based ligands and LNPs have already entered clinical use.

Ligand-mediated delivery has emerged as a promising strategy for targeted nucleic acid therapeutics, in which specific chemical ligands are conjugated to nucleic acids to form bioconjugates. Ligands act as molecular zip codes that help direct nucleic acid therapeutics and their carriers to specific cell types within an organ or tissue. These bioconjugates bind to receptors on specific cell types and are taken up by those cells. This bioconjugate-based approach has revolutionized hepatocyte-targeted delivery by using carbohydrate-based ligands. Drugs approved by the US Food and Drug Administration (FDA), such as givosiran, lumasiran, inclisiran, vutrisiran, eplontersen, and donidalorsen, use N-acetylgalactosamine (GalNAc) as a ligand for hepatic targeting. GalNAc binds to the asialoglycoprotein receptor (ASGPR) on hepatocytes with high affinity and is rapidly taken up via endocytosis with minimal off-target effects. This success was built on decades of sustained research (Ashwell and Harford, 1982; Nair et al., 2014). Since its discovery in 1965 and the subsequent recognition of its hepatocyte specificity, high affinity for GalNAc, and rapid endocytic uptake, ASGPR has served as a cornerstone for targeted liver delivery. This long arc of discovery and optimization ultimately culminated, nearly 50 years later, in the clinical translation of GalNAc nucleic acid therapeutics (Springer and Dowdy, 2018).

Most nucleic acid therapeutics are currently limited to liver-associated diseases or to local administration, largely because robust, generalizable delivery platforms for other organs are lacking. The success of hepatocyte-directed delivery through ASGPR targeting illustrates the potential of ligand-mediated, cell-specific uptake, but comparable ligand–receptor systems for most extrahepatic tissues remain largely underdeveloped. This bottleneck significantly constrains the therapeutic potential of RNA-based medicines across tissues such as the kidney, lung, brain, and heart, as well as diverse tumor types. Recognizing this challenge, nucleic acid–based scaffold design has advanced rapidly to enable delivery beyond the liver. Several bioconjugate-based strategies are being actively explored for their ability to target diseases of the brain, kidney, lung, and muscle, offering promising paths toward extrahepatic delivery. These approaches are being developed at both the clinical and preclinical stages to enhance extrahepatic organ delivery of nucleic acids through receptor-mediated targeting, peptide-facilitated cellular uptake, and lipophilicity-driven cellular uptake. Researchers have used carbohydrate-based ligands to target the lungs, kidneys, and specific cell types. Mannose-based ligands have been used to target mannose receptors expressed on pulmonary and pancreatic macrophages (Uehara et al., 2022; Yamazaki et al., 2023). For kidney-specific delivery, platforms like RENTAC and STRIKE are being developed; RENTAC targets renal tubules for nucleic acid and small-molecule delivery, while STRIKE enables selective siRNA delivery to the kidney (Ding et al., 2025; Kumar et al., 2024). These platforms target the megalin and cubilin receptors expressed in proximal convoluted tubule cells. Zeal Therapeutics, a University of Connecticut startup, is developing the RENTAC platform for kidney cell-specific delivery of siRNA. Judo Bio's STRIKE platform has demonstrated kidney-specific siRNA delivery and target gene knockdown in mice and non-human primates (Ding et al., 2025). Antibody-oligonucleotide conjugates (AOCs) are an emerging platform for cell-specific delivery of NADs. In preclinical studies, transferrin receptor 1 (TfR1)-targeting antibody-siRNA conjugates achieved over 15-fold greater accumulation in skeletal muscle and cardiac tissue than unconjugated siRNA, resulting in substantial target gene silencing (Malecova et al., 2023). The successful clinical advancement of

multiple TfR1-directed AOCs highlights the potential of receptor-mediated conjugate strategies to enable effective extrahepatic delivery of NADs (Johnson et al., 2026; Malecova et al., 2023). TfR overexpression on brain endothelial cells has been explored to enable central nervous system (CNS) delivery of nucleic acid therapeutics via receptor-mediated transcytosis across the blood–brain barrier (Desjardins et al., 2022; Lee et al., 2026). In addition, 2′-O-hexadecyl (C16)-siRNA conjugates have demonstrated promising CNS distribution and durable gene silencing up to 3 months after intrathecal administration. These conjugates also produce durable effects in the eye after intravitreal administration and in the lungs after intranasal administration, highlighting their potential for localized nucleic acid delivery (Brown et al., 2022). Attaching siRNAs to lipophilic compounds such as cholesterol, docosanoic acid, docosahexaenoic acid, and lithocholic acid uses the body's lipid transport pathways to improve their distribution and uptake in tissues such as the liver, kidneys, adrenal glands, and ovaries, which may help treat disorders that affect multiple systems (Osborn et al., 2019).

Despite these advances, progress in targeted nucleic acid delivery has been relatively slow. Researchers have largely tested one ligand at a time to see whether it helps cells take up nucleic acids. The ASGPR-GalNAc platform is highly successful, but it also shows how time-consuming this trial-and-error approach can be—even in the liver, where the target receptor is well understood and abundant (Prakash et al., 2014).

In tissues outside the liver, the receptors and uptake pathways needed for targeted delivery are less clear. Finding them through repeated experiments can hinder progress. More systematic, data-driven methods could speed discovery and reduce the uncertainty of traditional ligand screening. Advances in computational chemistry, large biological datasets, and information about receptor structures make it possible to study ligand–receptor interactions more effectively. Combining these resources with artificial intelligence (AI) and machine learning (ML) models could help identify new ligand–receptor pairs that enable selective uptake into cells, without years of testing. These methods could also support large-scale computer-based screening and help design delivery systems for specific cell types across different tissues, extending nucleic acid therapies beyond the liver.

Improving the Discovery Framework

Chemically, ligands may be composed of carbohydrates, peptides, aptamers, lipids, or antibodies. Among these, carbohydrates play an important role because their structural diversity and branched architecture enable cell-surface recognition by endocytic receptors. Carbohydrate–receptor interactions serve as key recognition mechanisms in cellular communication. Ligand interactions with cell membrane proteins or receptors form the basis for targeted bioconjugate-based delivery. To expand the range of targeted treatments, the field needs a more systematic discovery framework that identifies targeting ligands and accessible endocytic receptors, predicts ligand binding affinities, and maps receptor distribution across tissues. AI- and ML-guided approaches can improve these predictions and facilitate the identification of ligand–receptor pairs through more limited empirical experimentation (see *Figure 1*).

Criteria in Ligand Identification for Receptor Targeting

To design extrahepatic targeting ligands, a three-part framework for ligand and receptor screening is crucial (see *Figure 1*). First, methods for accurate binding site prediction on receptor protein structures are essential to identify a pool of candidate ligands. Second, advanced protein–ligand interaction models that can rank affinities across large receptor panels are essential for prioritizing key ligands based on receptor affinity. Third, comprehensive protein expression atlases must be available to map receptor prevalence at both the organ and cell-specific levels. This information will provide a clear understanding of where the ligand can effectively engage. These

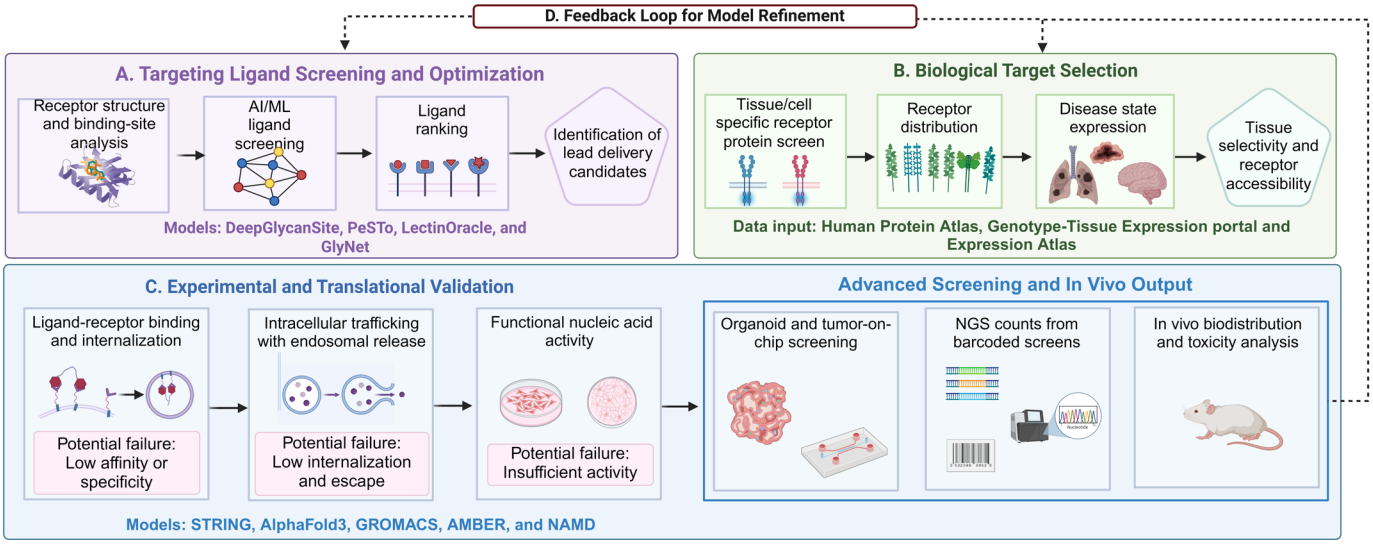


FIGURE 1 | AI/ML-Guided Screening and Optimization of Nucleic Acid Bioconjugates for Tissue- and Cell-Specific Delivery

SOURCE: The figure was created by the authors using BioRender software.

NOTES: (A) Discovery and prioritization of new ligand candidates for known cell-surface receptors. (B) Identification of novel endocytic receptors and mapping of their organ- and tissue-level expression under healthy and disease conditions. (C) Experimental and translational validation; the failure mode associated with each stage is annotated, illustrating that affinity does not guarantee uptake and uptake does not guarantee functional delivery. (D) Feedback loop for model refinement, in which in vitro and in vivo data are returned to the models in panels A and B so that predictions are retrained on in vivo outcomes rather than on binding affinity and cellular uptake alone. NGS = next-generation sequencing.

three components help answer key translational questions, such as ligand selectivity, receptor accessibility, and receptor expression on the appropriate tissue or cell type, thereby making targeted delivery plausible. This approach can reduce trial-and-error efforts and accelerate the discovery of clinically translatable ligands.

Ligand Screening

The first step in developing targeted nucleic acid bioconjugates is determining whether a receptor possesses an accessible binding pocket that supports ligand interaction and internalization. Once candidate receptors are identified, the next step is to determine which ligands bind with high affinity and selectivity. Binding site prediction also saves time by allowing researchers to avoid occluded targets and helps guide ligand design by identifying the number of targeting moieties required to promote efficient receptor-mediated endocytosis. In addition, it helps researchers avoid low-confidence targets, saving time and resources and enabling them to focus on ligand–receptor pairs with the highest translational potential. One example of ligand screening tools is the use of computational models, such as DeepGlycanSite, to precisely identify carbohydrate-binding residues within target protein structures (see *Figure 1A*). This model predicts protein structures and interactions with different carbohydrate types and offers probability maps to aid in target selection (He et al., 2024). A complementary prediction model, PeSTo, uses geometric deep learning to predict protein–carbohydrate binding interfaces (Bibekar et al., 2024).

Ligand Optimization and Ranking

Selected carbohydrate-based ligands can be further screened or optimized using models such as LectinOracle and GlyNet. LectinOracle predicts

the binding affinity of previously uncharacterized targets such as PSE41-5 and the jacalin-related domain from OsJAC1 (a rice protein), and was trained on a large dataset of known protein–carbohydrate interactions (Lundstrom et al., 2022). GlyNet predicts the strength of protein–carbohydrate binding interactions and can be used to rank these interactions (Carpenter et al., 2022). Such tools help narrow the chemical design space and provide a more structural approach to ligand discovery by functionally categorizing candidates based on receptor recognition, affinity ranking, and uptake optimization. They help with rapid screening and narrowing down a list of carbohydrate ligand candidates for laboratory testing against a selected receptor (Carpenter et al., 2022).

Target Receptor Expression Profiling

For receptor-mediated delivery, ligands must bind receptors in the appropriate orientation, and those receptors must be sufficiently abundant in the target tissues. Selective extrahepatic delivery therefore depends on receptor expression matching the anatomical target, which helps predict the in vivo distribution of nucleic acid bioconjugates. Resources such as the Human Protein Atlas can be used to understand a receptor's tissue- and cell-specific preferences in a specific organ (Thul and Lindskog, 2018). The Genotype-Tissue Expression (GTEx) portal and the Expression Atlas (EMBL-EBI) can also be used to examine the tissue-specific distribution of coding RNA sequences (see *Figure 1B*; GTEx Consortium, 2020). Selectivity of the conjugate depends on the ratio of receptor expression in the target tissue to off-target receptor expression, rather than on absolute expression. This distinction matters most for the liver and kidney, which receive a systemically administered conjugate by default and therefore compete for it regardless of the targeting ligand (Geary et al., 2015). Receptor expression has also been shown to differ in diseased states compared with healthy-tissue reference datasets on which these atlases are built (Uhlen et al., 2017). ASGPR meets all these criteria, which is one reason the GalNAc platform succeeded (Nair et al., 2014). After ligand–receptor pair screening, results from the modeling studies can then be validated in cell culture and, if successful, evaluated in the preclinical setting for their pharmacokinetic and pharmacodynamic properties. The expression profile of receptors is essential, as even a strong ligand–receptor interaction may be of limited value if the receptor is broadly expressed, leading to off-target effects that divert the therapeutic away from its intended target. Integrating structural prediction, ligand ranking, and tissue-level expression data creates a more complete decision framework for selective in vivo targeting.

In this section, the authors describe a scenario that current AI models can test. Siglecs are receptors on immune cells that recognize sialic acid–containing carbohydrates. Sialic acid residues are widely present on cell surfaces, including those of extrahepatic tissues, making them accessible carbohydrate motifs for Siglec-mediated targeting and receptor-mediated endocytosis (Boelaars and van Kooyk, 2024). Computational models such as LectinOracle and GlyNet can rank sialylated glycan motifs by predicting how strongly they bind to different Siglec family members, providing a starting point for selecting promising ligands. DeepGlycanSite can then confirm that the receptor's binding pocket is accessible. The most promising candidates can be tested experimentally on a microarray to measure their specificity and binding strength. The number of targeting moieties in the ligand, such as two or three, can also be optimized to support effective cellular uptake and cytoplasmic delivery of the bioconjugates. Selecting a Siglec based on its expression can reduce the risk of off-target effects. Similar receptor binding screening methods can help identify new ligands from different types of small molecules and peptides (Wang et al., 2024). This example shows how AI/ML can guide a practical design process before experimental testing.

Identification of New Endocytic Receptors Is Key for Extrahepatic Delivery

The authors previously discussed how AI/ML approaches can be used to identify ligands for

known organ- or tissue-specific receptors that facilitate the cellular uptake of bioconjugates. But only a limited number of such receptors are currently known and can be efficiently exploited for delivery applications. Therefore, the next critical step in expanding extrahepatic delivery of nucleic acid therapeutics is the systematic discovery and characterization of the receptor landscape itself. Identifying new endocytic receptors requires a collaborative approach that integrates in vitro profiling, genome-wide screening, and computational modeling tools such as AlphaFold, LectinOracle, and GlyNet.

Receptor Screening Approach

In this section, the authors highlight the screening for receptors involved in endocytic uptake using a genome-wide in vitro approach. The Search Tool for the Retrieval of Interacting Genes (STRING) has been used to identify receptors involved in the uptake of serum-coated nanoparticles in cell culture (Ngo et al., 2022). This screening approach used clustered regularly interspaced short palindromic repeats (CRISPR) to disable 18,053 human protein-coding genes in epithelial cells with the Version 3 Toronto KnockOut library (Ngo et al., 2022). Each cell received a lentiviral particle containing a unique guide RNA, disrupting one gene in that cell. Screening thousands of these gene knockouts simultaneously can reveal which receptor proteins are responsible for nanoparticle uptake. This approach, which focuses on interactions between proteins on the nanoparticle surface and specific cell membrane receptors, could also help identify new ligands for cell-specific nucleic acid delivery. A similar method could be used in vitro to identify endocytic receptors specific to different tissues. Together, these advances suggest an opportunity to improve existing workflows and create more connected discovery pipelines. Genome-wide CRISPR screening can strengthen computational receptor-selection methods by providing direct functional evidence that identified targets or receptors support effective nucleic acid uptake and delivery. These unbiased screens can confirm computational predictions and reveal previously unknown receptors, trafficking pathways, and intracellular proteins involved in productive nucleic acid delivery. This approach could reduce the need for extensive empirical optimization and speed the development of nucleic acid therapies beyond the liver.

Two developments move candidate selection beyond two-dimensional cell cultures and provide data for predictive models. The first is direct in vivo screening with DNA-barcoded libraries. In this approach, researchers administer many candidate conjugates together and measure their distribution across tissues and cells by next-generation sequencing. This produces datasets based on biodistribution rather than binding affinity or uptake in a single immortalized cell line (Sago et al., 2018). The second approach combines ML-guided sequence prioritization with in vivo selection. In adeno-associated virus capsid engineering, models trained on large datasets of mutational fitness and in vivo enrichment have produced viable variants with altered tissue targeting more efficiently than random mutagenesis and with greater sequence diversity than sequential in vitro-to-in vivo methods (Bryant et al., 2021; Ogden et al., 2019). The same approach applies to bioconjugates: an in vivo screen using barcoded ligand libraries can measure outcomes such as delivery at relevant doses and specific biodistribution. Between computational predictions and animal studies, organ-on-chip and organoid systems offer an intermediate step by modeling barriers and interactions between different cell types (Ingber, 2022).

Understanding Endosomal Trafficking

Computational tools are highly developed for predicting protein–ligand interactions. Nucleic acid bioconjugates undergo endocytic uptake, which typically occurs via clathrin- or caveolae-mediated pathways. Integrating AlphaFold3 with sophisticated molecular dynamics (MD) simulations offers a promising approach for studying molecular interactions and the resulting conformational changes in receptors. Endocytosis driven

by ligand–receptor interactions depends not only on binding strength but also on how efficiently receptors are internalized, how readily ligands are released, and how quickly receptors return to the cell membrane. Understanding these biophysical interactions in chemical and cellular biology is essential for assessing whether a ligand–receptor pair is effective for delivery applications. Receptors that recycle quickly are especially promising for clinical use because they can repeatedly internalize bioconjugates and support sustained uptake. The clinical success of ASGPR-targeted GalNAc conjugates is a clear example, driven by the high expression and rapid recycling of ASGPR in hepatocytes. AlphaFold3 uses a diffusion architecture to predict how proteins interact with small-molecule ligands, nucleic acids, and antigen–antibody pairs. These features make it a powerful tool for predicting ligand–receptor interactions that initiate endocytosis. MD simulation studies can help researchers predict the specific endocytic pathways that follow ligand–receptor interaction. Advanced simulation platforms such as GROMACS, AMBER, and NAMD can support the investigation of these processes by modeling specific proteins involved in endocytosis and intracellular trafficking, including receptors and membrane-associated complexes (Abraham et al., 2015; Case et al., 2023; Phillips et al., 2020). Endosomes play an important role in trafficking conjugates within cells following uptake. After uptake, the conjugates may become trapped within a cellular organelle or compartment known as an endosome. Endosomal entrapment may lead to cargo degradation or provide a temporary storage compartment. After conjugates enter a cell through endocytosis, they move through different endosomal stages and may eventually reach lysosomes, where they can be degraded. However, some cargo avoids this pathway and instead remains temporarily in endosomal compartments, allowing more time for release into the cytosol.

These intracellular trafficking pathways are complex and not yet fully understood, especially in how they vary across cell types and disease states. Although receptor binding and effective internalization are important first steps, the ability to escape endosomes and successfully deliver nucleic acids is widely considered a key factor in determining efficacy. Combining insights from protein, membrane, and organelle simulations with AI- and ML-based screening may improve the understanding of cellular uptake and help design more effective bioconjugates. Yet these approaches can make accurate predictions only with a better understanding of complex cellular processes, such as intracellular trafficking, endosomal escape, subcellular localization, release, nucleic acid conjugate stability inside cells, and target engagement. Together, these factors determine how successfully nucleic acids are delivered. However, current AI/ML and MD simulation methods are better at identifying receptors and ranking ligands than at predicting the delivery pathway.

Analyzing large libraries of ligands, mRNA and protein sequences, structural data, and their complex biochemical interactions is a major challenge. Differences between datasets from healthy and diseased states make predictive modeling more difficult and require significant computational resources. Advances in AI/ML and emerging quantum computing technologies are helping researchers analyze complex biological data and design molecules. These developments could accelerate the development, optimization, and clinical translation of delivery platforms for extrahepatic nucleic acid therapies.

Bridging Nucleic Acid Delivery and Clinical Translation

AI/ML approaches can accelerate the discovery of new receptors and targeting ligands by enabling large-scale screening of ligand–receptor interactions under controlled, idealized conditions. When integrated with experimental in vitro validation, these approaches can further strengthen candidate selection and increase confidence in clinical application. Nevertheless, caution is advised because ligands that appear promising based on AI/ML approaches or cell culture studies may behave very differently in living organisms.

Successful in vivo translation of nucleic acid bioconjugates can be hindered by significant biological barriers, including premature metabolism of targeting ligands, systemic clearance before reaching target tissues, biomolecular interactions, immune recognition or immunogenicity, unintended passive uptake, sequestration by the reticuloendothelial system, hemodynamic shear forces, and broader pharmacokinetic and absorption, distribution, metabolism, and excretion constraints (Kielpinski and Schumacher, 2026; Roberts et al., 2020). These challenges are common across a wide range of targeted delivery platforms and remain major contributors to the high attrition rates observed from discovery-stage candidates to clinically successful therapeutics. A recent study using an in vivo endosomal escape assay highlighted the importance of endosomal escape in determining productive cellular delivery (Jozic et al., 2026). This highlights the disconnect between conventional in vitro uptake-based screening and actual in vivo nucleic acid delivery. Similar discrepancies may also arise in nucleic acid bioconjugates, where receptor-mediated uptake does not always translate to functional delivery. Screening adeno-associated virus capsids for gene delivery is a well-known example of the translational disconnect across species. Although several engineered capsids have been successfully tested in rodent models, their efficacy often declines in higher-order species due to differences in receptor biology, tissue tropism, and immune responses. This underscores the need for a more integrated and rational selection process to achieve translational success after targeted delivery (Hordeaux et al., 2018; Nisanov et al., 2025). Similar translational limitations have been observed for peptide-, antibody-, and carbohydrate-conjugated active-targeting nanoparticles. Although strong receptor engagement and cellular uptake in vitro may not translate into comparable in vivo efficacy, this may be due to altered biodistribution caused by passive distribution, biomolecular interactions, and limited target accessibility. For instance, transferrin-targeted nanoparticles lose receptor specificity in biological environments by interacting with other proteins, forming a protein corona (Salvati et al., 2013). These qualifications emphasize the importance of incorporating physiologically relevant models and cross-species biological data into AI/ML-guided discovery frameworks to improve predictions of in vivo delivery performance and bridge the gap between discovery and successful therapeutic translation.

Future Directions

AI/ML may help shift ligand discovery for site-directed delivery from a trial-and-error process toward a more systematic and predictive approach. However, experimental and clinical validation remain essential. Nucleic acid therapeutics are emerging as a powerful therapeutic modality, not only for genetic disorders but also for diseases driven by proteins that are challenging or impossible to target using conventional small-molecule or antibody approaches. RNA-targeted therapies are inherently adaptable because they intervene earlier in the biological hierarchy, modulating gene expression before disease-causing or disease-associated proteins are produced. With over 7,000 known single-gene disorders and a growing understanding that many disease-associated proteins lack suitable drug-binding pockets, RNA-targeted therapies offer a promising alternative approach (Xie et al., 2023). Advances in nucleic acid stabilization, sequence design, and delivery technologies have already enabled rapid progress in the preclinical and clinical development of siRNA, ASOs, mRNA medicines for vaccines, and gene editing (Egli and Manoharan, 2023; Liu et al., 2025; Baden et al., 2021). Disease-relevant RNAs can be readily identified through transcriptomic and proteomic analyses, while NADs provide a direct, programmable route to target biological pathways that have historically remained out of reach. To streamline the regulatory approval of personalized therapies, including nucleic acid–based treatments, the FDA released a draft guidance entitled "Considerations for the Use of the Plausible Mechanism Framework to Develop Individualized Therapies that Target

Specific Genetic Conditions with a Known Biological Cause" (FDA, 2026). This framework could inform regulatory evaluation of individualized therapies for ultrarare diseases by placing greater emphasis on mechanistic understanding, natural history data, and preclinical and clinical evidence of efficacy and safety from limited patient populations, rather than relying on large randomized controlled trials, which are not feasible for extremely small patient populations. This approach could help expedite access to nucleic acid therapeutics for patients with rare genetic disorders and bolster the pathway for advancing innovations in this field (FDA, 2026).

The major opportunity in the field is to extend the effective delivery of NADs beyond the liver. Over the next 5 to 10 years, success could be defined by establishing reproducible, tissue- and cell-specific bioconjugate-based delivery platforms for multiple organs beyond the liver, with durable efficacy and acceptable safety profiles. To achieve this, the application of AI and ML can provide a more realistic and integrative understanding of the biochemical space linking ligand design with critical biological features of target receptors, including expression levels, recycling kinetics, tissue perfusion, and disease-specific alterations. Once a viable tissue-specific endocytic receptor and corresponding ligand have been identified, bioconjugate-based delivery offers a well-characterized and reproducible delivery platform. Collectively, AI- and ML-driven frameworks have the potential to shift current empirical, trial-based ligand–receptor discovery toward more rational, generalizable, and predictive strategies. In parallel, ligand-functionalized nanoformulations for extrahepatic targeting are gaining momentum, while nanoparticle systems remain essential for the in vivo delivery of mRNA and gene editing technologies.

Despite these advances, several challenges remain that hinder the broad clinical translation of nucleic acid bioconjugates. Nonspecific interactions, off-target distribution, and unintended accumulation in non-diseased tissues and cells continue to pose safety and efficacy concerns. In addition, limited tissue perfusion, extracellular matrix barriers, inflammation, and pathological remodeling can significantly restrict delivery in chronic or advanced disease states. Addressing these challenges will require coordinated efforts across academia, industry, and clinical partners. Academic researchers are well-positioned to uncover new biological mechanisms, identify receptors, characterize disease-specific cellular changes, and generate mechanistic datasets, while industry provides expertise in scalable delivery platforms, advanced nucleic acid chemistries, integrating predictive models across portfolios, and navigating regulatory and clinical development. A key enabler of this precision-driven approach is access to high-quality, disease-specific biological data, particularly from clinical settings. Clinicians can play a pivotal role by facilitating access to patient-relevant data and helping contextualize biological data within real disease settings. Patient-derived transcriptomic, proteomic, and histopathological data are crucial for understanding alterations in cell type composition, receptor expression, and tissue architecture across organs and disease stages. These data are particularly significant because cellular targets and accessible receptors in diseased tissues often exhibit substantial differences in both expression and accessibility through systemic circulation compared to those in healthy organs. Integrating clinically informed insights into AI- and ML-driven extrahepatic delivery models can markedly improve the selection of target receptors, ligands, and dosing strategies, thereby making bioconjugate-based therapies more precise, effective, and disease-specific. In addition, incorporating emerging safety and toxicity data could improve the ability of these models to predict potential safety risks. Without this clinical perspective, delivery platforms risk being shaped by simplified biological models that do not fully capture the complexity of real disease.

If these scientific, clinical, and collaborative challenges can be overcome, AI-powered, machine learning–guided bioconjugates for nucleic acid delivery could become a predictive, effective,

and safe modular engineering platform. Such a shift would enable the rational design of delivery systems tailored to specific organs, cell types, and disease contexts, substantially expanding the extrahepatic therapeutic reach of nucleic acid–based medicines.

References

Abraham, M. J., T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl. 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1-2:19-25. https://doi.org/10.1016/j.softx.2015.06.001.

Ashwell, G., and J. Harford. 1982. Carbohydrate-specific receptors of the liver. *Annual Review of Biochemistry* 51:531-554. https://doi.org/10.1146/annurev.bi.51.070182.002531.

Baden, L. R., H. M. El Sahly, B. Essink, K. Kotloff, S. Frey, R. Novak, D. Diemert, S. A. Spector, N. Rouphael, C. B. Creech, J. McGettigan, S. Khetan, N. Segall, J. Solis, A. Brosz, C. Fierro, H. Schwartz, K. Neuzil, L. Corey, P. Gilbert, H. Janes, D. Follmann, M. Marovich, J. Mascola, L. Polakowski, J. Ledgerwood, B. S. Graham, H. Bennett, R. Pajon, C. Knightly, B. Leav, W. Deng, H. Zhou, S. Han, M. Ivarsson, J. Miller, T. Zaks, and COVE Study Group. 2021. Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *The New England Journal of Medicine* 384(5):403-416. https://doi.org/10.1056/NEJMoa2035389.

Bibekar, P., L. Krapp, and M. Dal Peraro. 2024. PeSTo-Carbs: Geometric deep learning for prediction of protein-carbohydrate binding interfaces. *Journal of Chemical Theory and Computation* 20(8):2985-2991. https://doi.org/10.1021/acs.jctc.3c01145.

Boelaars, K., and Y. van Kooyk. 2024. Targeting myeloid cells for cancer immunotherapy: Siglec-7/9/10/15 and their ligands. *Trends in Cancer* 10(3):230-241. https://doi.org/10.1016/j.trecan.2023.11.009.

Brown, K. M., J. K. Nair, M. M. Janas, Y. I. Anglero-Rodriguez, L. T. H. Dang, H. Peng, C. S. Theile, E. Castellanos-Rizaldos, C. Brown, D. Foster, J. Kurz, J. Allen, R. Maganti, J. Li, S. Matsuda, M. Stricos, T. Chickering, M. Jung, K. Wassarman, J. Rollins, L. Woods, A. Kelin, D. C. Guenther, M. W. Mobley, J. Petrulis, R. McDougall, T. Racie, J. Bombardier, D. Cha, S. Agarwal, L. Johnson, Y. Jiang, S. Lentini, J. Gilbert, T. Nguyen, S. Chigas, S. LeBlanc, U. Poreci, A. Kasper, A. B. Rogers, S. Chong, W. Davis, J. E. Sutherland, A. Castoreno, S. Milstein, M. K. Schlegel, I. Zlatev, K. Charisse, M. Keating, M. Manoharan, K. Fitzgerald, J. T. Wu, M. A. Maier, and V. Jadhav. 2022. Expanding RNAi therapeutics to extrahepatic tissues with lipophilic conjugates. *Nature Biotechnology* 40(10):1500-1508. https://doi.org/10.1038/s41587-022-01334-x.

Bryant, D. H., A. Bashir, S. Sinai, N. K. Jain, P. J. Ogden, P. F. Riley, G. M. Church, L. J. Colwell, and E. D. Kelsic. 2021. Deep diversification of an AAV capsid protein by machine learning. *Nature Biotechnology* 39(6):691-696. https://doi.org/10.1038/s41587-020-00793-4.

Carpenter, E. J., S. Seth, N. Yue, R. Greiner, and R. Derda. 2022. GlyNet: A multi-task neural network for predicting protein-glycan interactions. *Chemical Science* 13(22):6669-6686. https://doi.org/10.1039/d1sc05681f.

Case, D. A., H. M. Aktulga, K. Belfon, D. S. Cerutti, G. A. Cisneros, V. W. D. Cruzeiro, N. Forouzesh, T. J. Giese, A. W. Götz, H. Gohlke, S. Izadi, K. Kasavajhala, M. C. Kaymak, E. King, T. Kurtzman, T. S. Lee, P. Li, J. Liu, T. Luchko, R. Luo, M. Manathunga, M. R. Machado, H. M. Nguyen, K. A. O'Hearn, A. V. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, A. Risheh, S. Schott-Verdugo, A. Shajan, J. Swails, J. Wang, H. Wei, X. Wu, Y. Wu, S. Zhang, S. Zhao, Q. Zhu, T. E. Cheatham, III, D. R. Roe, A. Roitberg, C. Simmerling, D. M. York, M. C. Nagan, and K. M. Merz, Jr. 2023. AmberTools. *Journal of Chemical Information and Modeling* 63(20):6183-6191. https://doi.org/10.1021/acs.jcim.3c01153.

Desjardins, C. A., M. Yao, J. Hall, E. O'Donnell, R. Venkatesan, S. Spring, A. Wen, N. Hsia, P. Shen, R. Russo, B. Lan, T. Picariello, K. Tang, T. Weeden, S. Zanotti, R. Subramanian, and

O. Ibraghimov-Beskrovnaya. 2022. Enhanced exon skipping and prolonged dystrophin restoration achieved by TfR1-targeted delivery of antisense oligonucleotide using FORCE conjugation in mdx mice. *Nucleic Acids Research* 50(20):11401-11414. https://doi.org/10.1093/nar/gkac641.

Ding, S., Y. J. Jung, J. Lo, K. Liu, J. Cunniff, A. Belanger, H. Zhang, E. J. Comer, J. F. Lawrence, and A. Sehgal. 2025. STRIKE platform enables kidney-selective gene silencing via megalin-mediated uptake: SA-PO0610. *Journal of the American Society of Nephrology* 36(10S). https://doi.org/10.1681/ASN.2025yx5b0nt7.

Dowdy, S. F. 2017. Overcoming cellular barriers for RNA therapeutics. *Nature Biotechnology* 35(3):222-229. https://doi.org/10.1038/nbt.3802.

Egli, M., and M. Manoharan. 2023. Chemistry, structure and function of approved oligonucleotide therapeutics. *Nucleic Acids Research* 51(6):2529-2573. https://doi.org/10.1093/nar/gkad067.

FDA (US Food and Drug Administration). 2026. *Considerations for the use of the Plausible Mechanism Framework to develop individualized therapies that target specific genetic conditions with a known biological cause.* Silver Spring, MD: FDA. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-plausible-mechanism-framework-develop-individualized-therapies-target-specific (accessed November 5, 2026).

Geary, R. S., D. Norris, R. Yu, and C. F. Bennett. 2015. Pharmacokinetics, biodistribution and cell uptake of antisense oligonucleotides. *Advanced Drug Delivery Reviews* 87:46-51. https://doi.org/10.1016/j.addr.2015.01.008.

GTEx (Genotype-Tissue Expression) Consortium. 2020. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369(6509):1318-1330. https://doi.org/10.1126/science.aaz1776.

He, X., L. Zhao, Y. Tian, R. Li, Q. Chu, Z. Gu, M. Zheng, Y. Wang, S. Li, H. Jiang, Y. Jiang, L. Wen, D. Wang, and X. Cheng. 2024. Highly accurate carbohydrate-binding site prediction with DeepGlycanSite. *Nature Communications* 15(1):5163. https://doi.org/10.1038/s41467-024-49516-2.

Hordeaux, J., Q. Wang, N. Katz, E. L. Buza, P. Bell, and J. M. Wilson. 2018. The neurotropic properties of AAV-PHP.B are limited to C57BL/6J mice. *Molecular Therapy* 26(3):664-668. https://doi.org/10.1016/j.ymthe.2018.01.018.

Ingber, D. E. 2022. Human organs-on-chips for disease modelling, drug development and personalized medicine. *Nature Reviews Genetics* 23(8):467-491. https://doi.org/10.1038/s41576-022-00466-9.

Johnson, N. E., L. J. Tai, J. I. Hamel, J. W. Day, J. M. Statland, P. Soltanzadeh, S. H. Subramony, C. A. Thornton, W. D. Arnold, M. Wicklund, M. L. Freimer, K. Eichinger, J. Dekdebrun, C. Y. Chen, V. Goel, B. McEvoy, Y. Zhu, S. G. Hughes, E. J. Ackermann, and A. A. Levin. 2026. An antibody-oligonucleotide conjugate for myotonic dystrophy type 1. *The New England Journal of Medicine* 394(8):763-772. https://doi.org/10.1056/NEJMoa2407326.

Jozic, A., C. Le Roux, J. Kim, M. Berchel, D. K. Sahel, E. K. Bodi, M. Palumbo, A. Vasudevan, N. T. V. Murthy, Y. Eygeris, M. Gautam, E. Bloom, A. P. Barnes, P. A. Jaffres, and G. Sahay. 2026. In vivo endosomal escape assay identifies mechanisms for efficient hepatic LNP delivery. *Nature Biotechnology*. https://doi.org/10.1038/s41587-026-03022-6.

Juliano, R. L. 2016. The delivery of therapeutic oligonucleotides. *Nucleic Acids Research* 44(14):6518-6548. https://doi.org/10.1093/nar/gkw236.

Kielpinski, L. J., and F. F. Schumacher. 2026. A translational framework for extrahepatic oligonucleotide delivery. *Nature Reviews Drug Discovery*. https://doi.org/10.1038/d41573-026-00125-2.

Kumar, V., A. Wahane, M. S. Tham, S. Somlo, A. Gupta, and R. Bahal. 2024. Efficient and selective kidney targeting by chemically modified

carbohydrate conjugates. *Molecular Therapy* 32(12):4383-4400. https://doi.org/10.1016/j.ymthe.2024.10.020.

Lee, H. J., Y. Xie, C. F. Greineder, and P. M. Tessier. 2026. Bioconjugates for improved delivery of oligonucleotide therapeutics to the central nervous system. *Advanced Drug Delivery Reviews* 230:115778. https://doi.org/10.1016/j.addr.2026.115778.

Lipinski, C. A., F. Lombardo, B. W. Dominy, and P. J. Feeney. 2001. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews* 46(1-3):3-26. https://doi.org/10.1016/s0169-409x(00)00129-0.

Liu, M., Y. Wang, Y. Zhang, D. Hu, L. Tang, B. Zhou, and L. Yang. 2025. Landscape of small nucleic acid therapeutics: Moving from the bench to the clinic as next-generation medicines. *Signal Transduction and Targeted Therapy* 10(1):73. https://doi.org/10.1038/s41392-024-02112-8.

Lundstrom, J., E. Korhonen, F. Lisacek, and D. Bojar. 2022. LectinOracle: A generalizable deep learning model for lectin-glycan binding prediction. *Advanced Science* 9(1):e2103807. https://doi.org/10.1002/advs.202103807.

Malecova, B., R. S. Burke, M. Cochran, M. D. Hood, R. Johns, P. R. Kovach, V. R. Doppalapudi, G. Erdogan, J. D. Arias, B. Darimont, C. D. Miller, H. Huang, A. Geall, H. S. Younis, and A. A. Levin. 2023. Targeted tissue delivery of RNA therapeutics using antibody-oligonucleotide conjugates (AOCs). *Nucleic Acids Research* 51(12):5901-5910. https://doi.org/10.1093/nar/gkad415.

Nair, J. K., J. L. Willoughby, A. Chan, K. Charisse, M. R. Alam, Q. Wang, M. Hoekstra, P. Kandasamy, A. V. Kel'in, S. Milstein, N. Taneja, J. O'Shea, S. Shaikh, L. Zhang, R. J. van der Sluis, M. E. Jung, A. Akinc, R. Hutabarat, S. Kuchimanchi, K. Fitzgerald, T. Zimmermann, T. J. van Berkel, M. A. Maier, K. G. Rajeev, and M. Manoharan. 2014. Multivalent N-acetylgalactosamine-conjugated siRNA localizes in hepatocytes and elicits robust RNAi-mediated gene silencing. *Journal of the American Chemical Society* 136(49):16958-16961. https://doi.org/10.1021/ja505986a.

Ngo, W., J. L. Y. Wu, Z. P. Lin, Y. Zhang, B. Bussin, A. Granda Farias, A. M. Syed, K. Chan, A. Habsid, J. Moffat, and W. C. W. Chan. 2022. Identifying cell receptors for the nanoparticle protein corona using genome screens. *Nature Chemical Biology* 18(9):1023-1031. https://doi.org/10.1038/s41589-022-01093-5.

Nisanov, A. M., J. A. Rivera de Jesús, and D. V. Schaffer. 2025. Advances in AAV capsid engineering: Integrating rational design, directed evolution and machine learning. *Molecular Therapy* 33(5):1937-1945. https://doi.org/10.1016/j.ymthe.2025.03.056.

Ogden, P. J., E. D. Kelsic, S. Sinai, and G. M. Church. 2019. Comprehensive AAV capsid fitness landscape reveals a viral gene and enables machine-guided design. *Science* 366(6469):1139-1143. https://doi.org/10.1126/science.aaw2900.

Osborn, M. F., A. H. Coles, A. Biscans, R. A. Haraszti, L. Roux, S. Davis, S. Ly, D. Echeverria, M. R. Hassler, B. M. D. C. Godinho, M. Nikan, and A. Khvorova. 2019. Hydrophobicity drives the systemic distribution of lipid-conjugated siRNAs via lipid transport pathways. *Nucleic Acids Research* 47(3):1070-1081. https://doi.org/10.1093/nar/gky1232.

Phillips, J. C., D. J. Hardy, J. D. C. Maia, J. E. Stone, J. V. Ribeiro, R. C. Bernardi, R. Buch, G. Fiorin, J. Hénin, W. Jiang, R. McGreevy, M. C. R. Melo, B. K. Radak, R. D. Skeel, A. Singharoy, Y. Wang, B. Roux, A. Aksimentiev, Z. Luthey-Schulten, L. V. Kalé, K. Schulten, C. Chipot, and E. Tajkhorshid. 2020. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *The Journal of Chemical Physics* 153(4). https://doi.org/10.1063/5.0014475.

Prakash, T. P., M. J. Graham, J. Yu, R. Carty, A. Low, A. Chappell, K. Schmidt, C. Zhao, M. Aghajan, H. F. Murray, S. Riney, S. L. Booten, S. F. Murray,

H. Gaus, J. Crosby, W. F. Lima, S. Guo, B. P. Monia, E. E. Swayze, and P. P. Seth. 2014. Targeted delivery of antisense oligonucleotides to hepatocytes using triantennary N-acetyl galactosamine improves potency 10-fold in mice. *Nucleic Acids Research* 42(13):8796-8807. https://doi.org/10.1093/nar/gku531.

Roberts, T. C., R. Langer, and M. J. A. Wood. 2020. Advances in oligonucleotide drug delivery. *Nature Reviews Drug Discovery* 19(10):673-694. https://doi.org/10.1038/s41573-020-0075-7.

Sago, C. D., M. P. Lokugamage, K. Paunovska, D. A. Vanover, C. M. Monaco, N. N. Shah, M. Gamboa Castro, S. E. Anderson, T. G. Rudoltz, G. N. Lando, P. Munnilal Tiwari, J. L. Kirschman, N. Willett, Y. C. Jang, P. J. Santangelo, A. V. Bryksin, and J. E. Dahlman. 2018. High-throughput in vivo screen of functional mRNA delivery identifies nanoparticles for endothelial cell gene editing. *Proceedings of the National Academy of Sciences* 115(42):E9944-E9952. https://doi.org/10.1073/pnas.1811276115.

Salvati, A., A. S. Pitek, M. P. Monopoli, K. Prapainop, F. Baldelli Bombelli, D. R. Hristov, P. M. Kelly, C. Åberg, E. Mahon, and K. A. Dawson. 2013. Transferrin-functionalized nanoparticles lose their targeting capabilities when a biomolecule corona adsorbs on the surface. *Nature Nanotechnology* 8(2):137-143. https://doi.org/10.1038/nnano.2012.237.

Springer, A. D., and S. F. Dowdy. 2018. GalNAc-siRNA conjugates: Leading the way for delivery of RNAi therapeutics. *Nucleic Acid Therapeutics* 28(3):109-118. https://doi.org/10.1089/nat.2018.0736.

Thul, P. J., and C. Lindskog. 2018. The human protein atlas: A spatial map of the human proteome. *Protein Science* 27(1):233-244. https://doi.org/10.1002/pro.3307.

Uehara, K., T. Harumoto, A. Makino, Y. Koda, J. Iwano, Y. Suzuki, M. Tanigawa, H. Iwai, K. Asano, K. Kurihara, A. Hamaguchi, H. Kodaira, T. Atsumi, Y. Yamada, and K. Tomizuka. 2022. Targeted delivery to macrophages and dendritic cells by chemically modified mannose ligand-conjugated siRNA. *Nucleic Acids Research* 50(9):4840-4859. https://doi.org/10.1093/nar/gkac308.

Uhlen, M., C. Zhang, S. Lee, E. Sjostedt, L. Fagerberg, G. Bidkhori, R. Benfeitas, M. Arif, Z. Liu, F. Edfors, K. Sanli, K. von Feilitzen, P. Oksvold, E. Lundberg, S. Hober, P. Nilsson, J. Mattsson, J. M. Schwenk, H. Brunnstrom, B. Glimelius, T. Sjoblom, P. H. Edqvist, D. Djureinovic, P. Micke, C. Lindskog, A. Mardinoglu, and F. Ponten. 2017. A pathology atlas of the human cancer transcriptome. *Science* 357(6352). https://doi.org/10.1126/science.aan2507.

Wang, W., K. Chen, T. Jiang, Y. Wu, Z. Wu, H. Ying, H. Yu, J. Lu, J. Lin, and D. Ouyang. 2024. Artificial intelligence-driven rational design of ionizable lipids for mRNA delivery. *Nature Communications* 15(1):10804. https://doi.org/10.1038/s41467-024-55072-6.

Xie, X., T. Yu, X. Li, N. Zhang, L. J. Foster, C. Peng, W. Huang, and G. He. 2023. Recent advances in targeting the "undruggable" proteins: From drug discovery to clinical trials. *Signal Transduction and Targeted Therapy* 8(1):335. https://doi.org/10.1038/s41392-023-01589-z.

Yamazaki, K., K. Kubara, Y. Suzuki, T. Hihara, D. Kurotaki, T. Tamura, M. Ito, and K. Tsukahara. 2023. Multivalent mannose-conjugated siRNA causes robust gene silencing in pancreatic macrophages in vivo. *European Journal of Pharmaceutics and Biopharmaceutics* 183:61-73. https://doi.org/10.1016/j.ejpb.2022.12.017.

DOI

https://doi.org/10.31478/202611a

Suggested Citation

Kumar, V., S. P. Pradeep, W. M. Saltzman, J. E. Manautou, and R. Bahal. 2026. Using AI and Machine Learning to Advance Nucleic Acid Therapeutic Delivery Beyond the Liver. *NAM Perspectives*. Discussion Paper, National Academy of Medicine, Washington, DC. https://doi.org/10.31478/202611a.

Author Information

Vikas Kumar, PhD, is Assistant Research Professor, University of Connecticut. **Sai Pallavi Pradeep, MSc,** is a PhD candidate at the University of Connecticut. **W. Mark Saltzman, PhD,** is Sterling Professor of Biomedical Engineering, Yale University. **José E. Manautou, PhD,** is Boehringer Ingelheim Pharmaceuticals Inc. Chair in Mechanistic Toxicology and Professor of Toxicology, Department of Pharmaceutical Sciences, University of Connecticut. **Raman Bahal, PhD,** is Professor of Pharmaceutics, Department of Pharmaceutical Sciences, University of Connecticut.

Acknowledgments

Work in the Bahal lab was supported by the SPARK Technology Commercialization Fund provided by the University of Connecticut and the National Institutes of Health.

Conflict of Interest Disclosures

Dr. Saltzman discloses serving as a co-founder, consultant, and member of the Board of Directors of Xanadu Bio and as a consultant for Stradefy Bio, Cranius, and CMC Pharma. Dr. Bahal discloses serving as a co-founder and member of the Board of Directors of Zeal Therapeutics and as a co-founder of Teucerian Therapeutics. No other conflicts were disclosed.

Disclaimer

The views expressed in this paper are those of the authors and not necessarily of the authors' organizations, the National Academy of Medicine (NAM), or the National Academies of Sciences, Engineering, and Medicine (the National Academies). The paper is intended to help inform and stimulate discussion. It is not a report of the NAM or the National Academies. Copyright by the National Academy of Sciences. All rights reserved.