Designing Therapeutic Recombinant AAV Vectors Using *In Silico* Vector Modeling

Ethan Waple[1], Jennifer C. Gifford[1], Sarath Mandava[1], Nicholas Ketz[2], Alpha B. Diallo[2], Amicia D. Elliott[2] and Adam S. Cockrell[1]

Solid Biosciences Inc.[1], Charlestown, MA, USA, Form Bio[2], Austin, TX/San Francisco, CA

Introduction

When designing safe, efficacious, and cost-effective AAV gene therapies, early R&D seeks to determine the best vector configuration with the intended full-length genetic payload from thousands or even millions of possibilities. Genetic cargo packaged within a single production lot can be a heterogenous population of recombinant AAV vector particles comprising a combination of different molecular species. We used long read sequencing to profile the molecular species in early discovery research grade rAAV therapeutic vector productions which revealed the presence of intended full-length vector species as well as varying amounts of unintended molecular species including vector truncations, chimeric molecules, molecules containing plasmid backbone sequences, and a variety of double-stranded vector molecules (i.e. snap-back genomes). The process for molecular species characterization is outlined in Figure 1. Long read sequencing was previously demonstrated to reveal novel insights into the heterogeneity of molecular species in rAAV preparations (Tran et al. 2022, Tai et al. 2018, Tran et al. 2020). Combining long read sequencing with *in silico* vector modeling may support early identification of *cis* elements in genetic cargos that may shift vector productions to the desired homogenous population of therapeutic molecules. *In silico* modeling of rAAV vectors using the FORMsight^{AI} artificial intelligence platform was performed (outlined in Figure 2). We demonstrate that truncation propensities predicted by *in silico* modeling were congruent with wet-lab results showing: 1. Vectors with higher truncation propensities led to less of the desired ssAAV vector product; 2. Vectors with higher truncation propensities had higher levels of molecular ssAAV subtypes; and 3. Vector modifications to mitigate truncations increased the percentage of the desired full-length vector product. Overall, utilizing *in silico* vector modeling to assess thousands and potentially millions of vector configurations prior to *in vitro* and *in vivo* testing could improve workflow efficiency resulting in a more cost-effective therapeutic for gene therapy applications and shorten timelines in early R&D pipelines.

Methods

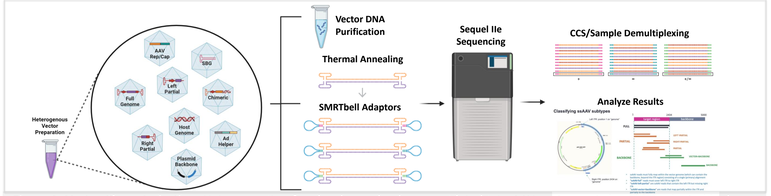
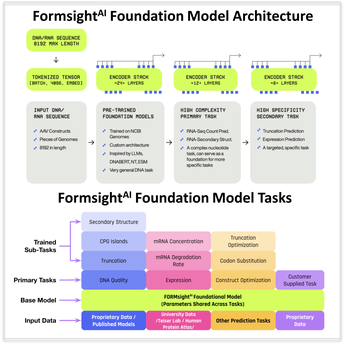


Figure 1. Vector Characterization. Heterogeneity of molecular payloads encapsidated into AAV-vector particles was evaluated by long-read sequencing and bioinformatics platforms that support circular consensus sequencing (CCS)/sample demultiplexing. The heterogenous mixture of molecular payloads after triple transfection (TT) can include the desired full-length vector genome, partial genomes, fragments of host genomic DNA, fragments of plasmid backbone DNA, fragments derived from Rep/Cap and/or Ad helper, DNA fragments comprising snap-back genomes (SBG), and DNA chimeras (combinations of different DNA fragments that have recombined during production). Total vector DNA was isolated and prepared for long read sequencing using Azenta workflows that were consistent with PacBio AAV genomic DNA preparation. Sequencing was executed on a PacBio Sequel IIe instrument followed by bioanalysis on the FormBio bioinformatics platform.



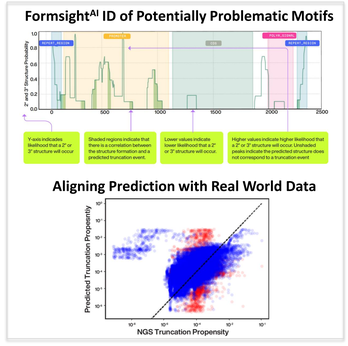


Figure 2. *In silico* vector modeling using the Formsight^{AI} foundation model. Potentially problematic motifs were identified early in the vector design process to allow for modifications that may mitigate down-stream vector heterogeneity. Aligning *in silico* predictions with real world NGS data suggests a correlation between the two.

Results

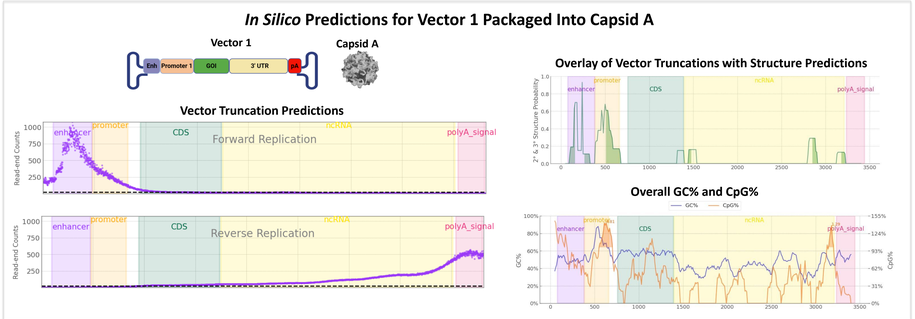


Figure 3. *In Silico* Predictions for Vector 1 Packaged Into Capsid A. Vector truncation predictions exhibited high read-end counts in the enhancer/promoter region in the + polarity strand (Forward Replication) and in the poly A signal in – polarity strand (reverse replication; purple dots represent predicted number of replication processes that terminate at a given nucleotide). Vector truncation predictions in the enhancer/promoter region overlay with strong secondary/tertiary structure probabilities (green shaded regions are hot spots for high correlation between a predicted truncation event and putative structural elements). The overall GC% is represented by the blue line and the %CpG regions are represented by the orange line. Regions with high %CpG content are shaded orange. **Summary: *In silico* predictions for vector 1 indicate high truncation potential in the enhancer/promoter region which may be associated with high levels of secondary/tertiary structures in this region and a high %CpG content.**

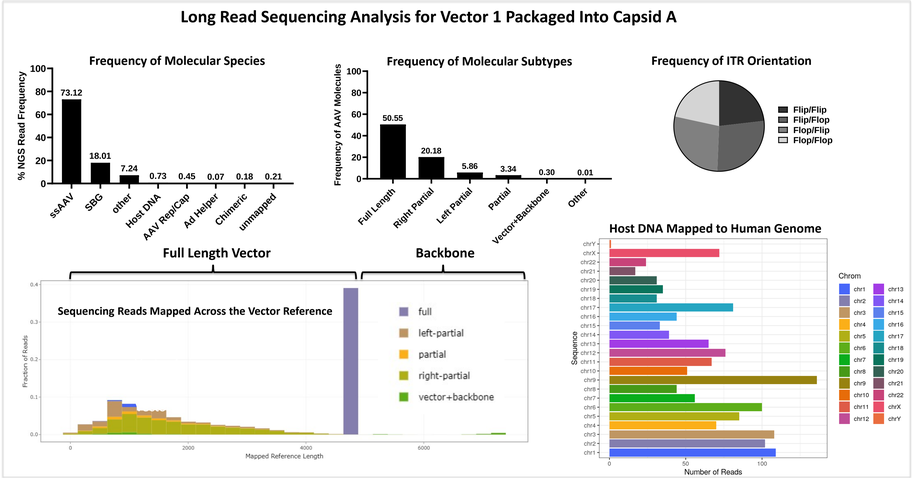


Figure 4. Long Read Sequencing Characterization of Vector 1 Packaged Into Capsid A. The predominant molecular species was single-stranded AAV (ssAAV) with snap-back genomes (SBGs) and other/undefined molecular species. Further investigation into the ssAAV subtypes revealed that ~50% of the heterogenous molecular population were full-length vector genomes with a large amount of the population comprising partial molecular species. The frequency of the ITR orientations in ssAAV molecules was as expected, about 1:1:1:1. Mapping the sequence reads across a linearized vector reference, including the plasmid backbone sequences, showed that many of the partial sequences mapped to the vector regions comprising the enhancer/promoter. Although less than 1% of the overall molecular species, the host DNA species were mapped to the human genome (vector production was in human embryonic kidney cells) and exhibited reads across all the chromosomes, except for the Y chromosome. **Summary: Long read sequencing characterization of vector 1 showed high levels of partials that mapped to the enhancer/promoter region, which is consistent with the *in silico* predictions in Figure 3.**

Results

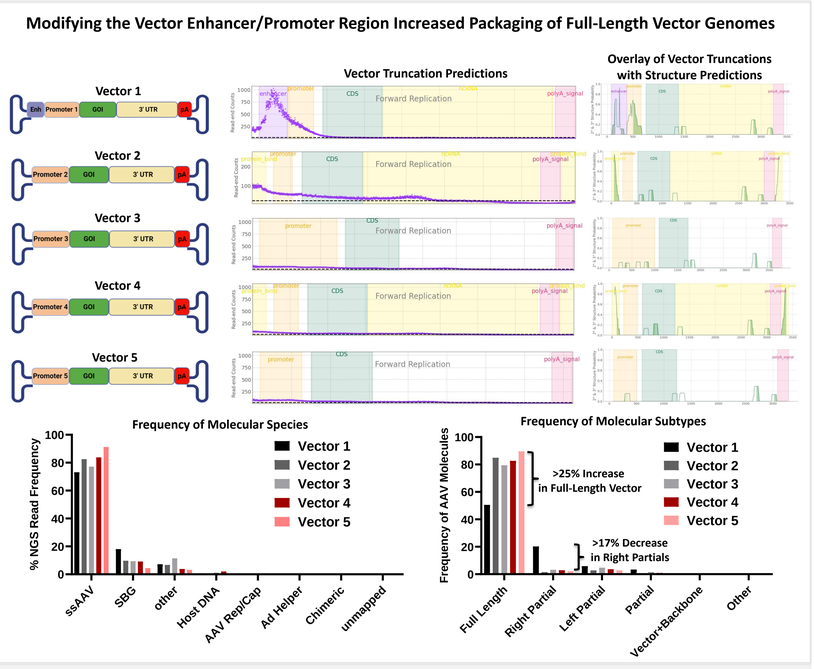


Figure 5. Vectors Modified in the Promoter Region and Individually Packaged Into the Same Capsid A. Vector 1 was modified by replacing the enhancer/promoter region with 4 different promoters (vectors 2-5). Each was individually packaged into the same capsid A. Vector truncation predictions indicated a nearly complete absence of truncations in the promoter regions, and throughout the entire genome, for vectors 2-5. Overlays with secondary/tertiary structure prediction also exhibited a near absence of complex structural elements in the promoters for vectors 2-5. Although the actual frequency of ssAAV molecular species was similar across all the vectors, the ssAAV molecular subtypes indicated vectors 2-5 have an increase (>25%) in full-length vector species packaged into vector particles compared to vector 1 and a decrease (>17%) in right partial molecular species was also observed. **Summary: Modification of the enhancer/promoter region in vector 1 led to an increase in the desired full-length molecular species and a decrease in the partial molecules, which was consistent with the *in silico* predictions.**

Conclusions

Table Summarizing Predicted Values vs. Real World Data

Vector	% Predicted Truncations	% Predicted Full-Length	% NGS AAV Full-Length
Vector 1	68	32	51
Vector 2	43	57	85
Vector 3	51	49	78
Vector 4	51	49	78
Vector 5	46	54	90

- **Predicted *in silico* values for truncations and full-length vector genomes are conservative compared to data acquired by long read sequencing; however, the trends across different vectors are similar.**
- ***In silico* truncation predictions that align with secondary/tertiary structure predictions indicate that reducing secondary/tertiary structures might mitigate vector truncations, resulting in a greater amount of vector product containing full-length recombinant AAV genomes.**
- **Consistent with *in silico* predictions we demonstrate that modifying enhancer/promoter regions with challenging structural elements can lead to a dramatic increase in full-length vector genomes while concomitantly reducing truncated molecular species**