Optimizing transgene expression for hemophilia gene therapy

Gene therapies for hemophilia are currently being studied to determine their safety and efficacy.
Approved gene therapies for hemophilia may have different labeling in different countries.

Pfizer

Contents

> Synopsis

Following transduction and delivery of the transgene to the liver cells, the transgene must undergo various cellular processes before the protein of interest is produced.[1,2] This involves conversion of recombinant adeno-associated virus (rAAV) vector-delivered single-stranded DNA into double-stranded DNA (dsDNA),[1–3] and transcription of the dsDNA into messenger RNA, which is then translocated to the cytoplasm to undergo the process of translation.[2] The rAAV vector expression cassette includes the transgene plus the essential regulatory elements – promoter, enhancer, polyadenylation signal sequence, and inverted terminal repeats – required for successful transgene expression.[2,4,5] This brochure will explore various approaches under investigation to optimize transgene expression of liver-directed hemophilia gene therapies.

The information included in this brochure is accurate as of January 2023. Please visit www.genetherapyscience.com for further information and check back regularly for updates.

Gene therapies for hemophilia are currently being studied to determine their safety and efficacy. Approved gene therapies for hemophilia may have different labelling in different countries.

Fundamentals of transgene expression

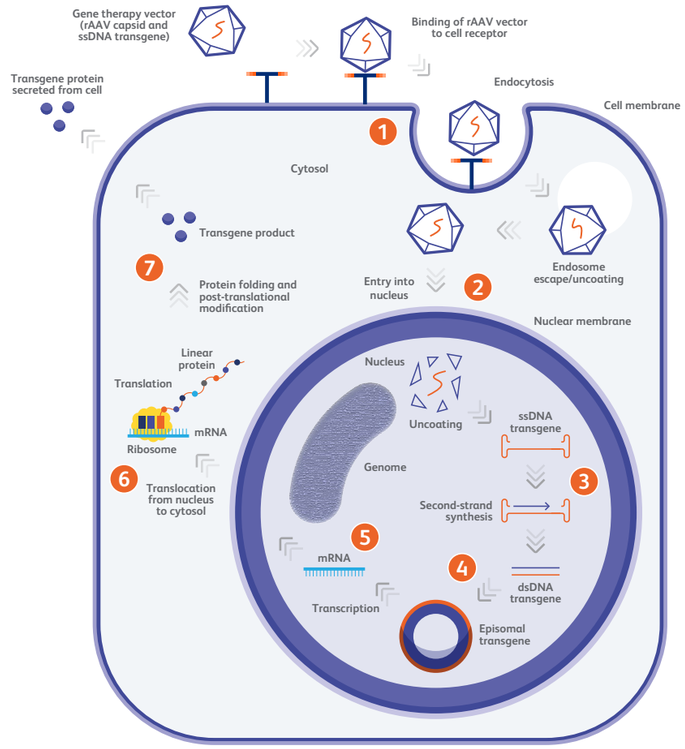
Gene therapy is the introduction, removal, or change in genetic material — specifically DNA or RNA — into the cells of a patient to treat a specific disease.[6,7] The success of gene therapy is, therefore, dependent on effective vehicles (vectors) to deliver the transgene to the nucleus of the target cell. The transgene will be subsequently expressed to produce a functional (healthy or working) protein (i.e., FVIII for hemophilia A or FIX for hemophilia B).[8,9]

> Recombinant adeno-associated virus (rAAV) vectors typically aim to deliver single-stranded DNA (ssDNA) to the nucleus of liver cells.[2,10] rAAV vectors are commonly used vectors for hemophilia gene therapies currently under investigation[11,12]

To explore the key principles of rAAV gene therapy and the importance of optimizing the rAAV vector, visit www.genetherapyscience.com/vector and access the brochures in the Learning Center

Gene therapies for hemophilia are currently being studied to determine their safety and efficacy. Approved gene therapies for hemophilia may have different labelling in different countries.

An overview of rAAV vector transgene expression



Not to scale, for illustrative purposes only

Figure adapted from Li C, Samulski RJ. 2020[2]
dsDNA: Double-stranded DNA; mRNA: Messenger RNA;
rAAV: Recombinant adeno-associated virus; ssDNA: Single-stranded DNA.

1

rAAV vectors bind to receptors on the cell surface and are internalized through endocytosis.[1,2]

2

Following endosomal escape, the rAAV vectors can be trafficked into the nucleus and uncoated, delivering ssDNA.[1,2,10]

3

ssDNA is converted into double-stranded DNA (dsDNA) through *de novo* synthesis of the complementary DNA (cDNA) strand – a process termed second-strand synthesis.[1] Conversion to dsDNA is essential to go from a transcriptionally inert to transcriptionally active transgene.[3]

4

rAAV vector-delivered transgenes exist as episomal DNA outside of the target cell genome. This involves circularization via genome recombination at the inverted terminal repeats (ITRs).[3]

5

The presence of transcription factor binding sites or transcription start sites within the expression cassette (promoter and enhancer regions) results in endogenous transcription factor binding and subsequent recruitment of RNA polymerase II.[13] Through the process of transcription, the transgene DNA is used as a template to produce a messenger (mRNA) molecule.[2,14]

> If introns are present in the transgene, transcription produces precursor-mRNA which is then spliced, removing the non-coding introns, to form mRNA[15]

7

The final step in the production of the functional protein requires the appropriate folding of the linear protein chain to form a functional protein structure, and any post-translational modifications that are required.[16]

6

The mRNA molecule is then translocated out of the nucleus into the cytoplasm, where the mRNA is translated by ribosomes to produce a linear protein chain.[15] Translation begins at the initiation codon – AUG – and terminates at one of the three stop codons – UAA, UAG, and UGA.[15]

rAAV is predominantly non-integrating[1,17–20]

Most rAAV vectors are predominantly non-integrating, meaning they generally do not insert into a person's own DNA or genome.[1,17–19] rAAV vector integration into the target cell genome can occur at a rate of 0.0001–1%[1,18,20]

Why does integration matter?

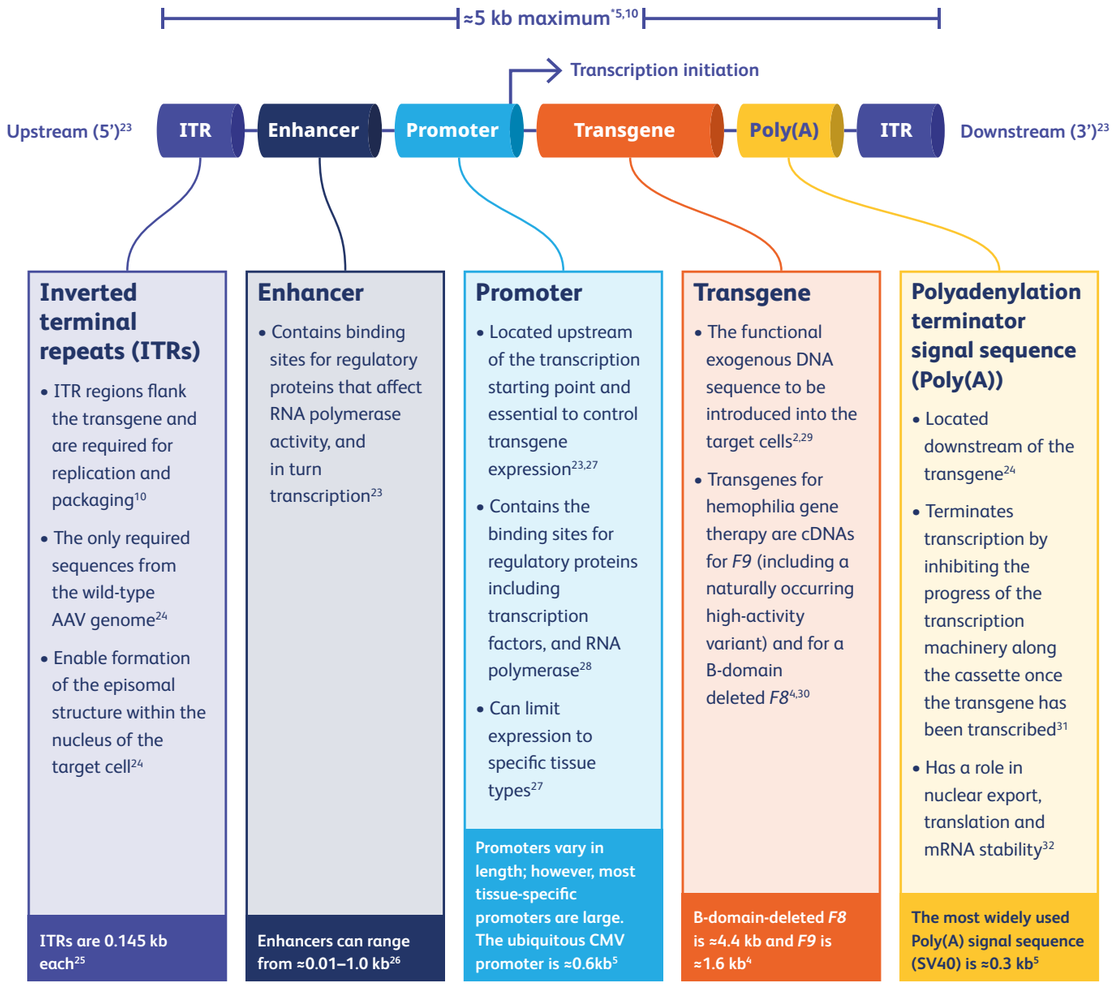
Insertion of rAAV DNA, or 'genome integration', does not necessarily lead to cancer development.[18] The infrequent insertion events with rAAV occur at random positions in the DNA but could lead to a risk of cancer if insertion occurred within a gene or region of the DNA related to cancer development[18]

In long-term follow-up of hemophilic dogs, integration and clonal expansion of cells was seen with insertions near genes associated with cancer in humans.[21] However, to date, rAAV vectors have not been shown to cause tumors in humans and non-rodent species, although long-term monitoring is needed to assess this[22]

To optimize transgene expression, knowledge of the various components of the rAAV vector expression cassette is required, alongside an understanding of how each of these components can be modified

rAAV vector expression cassette components

Overview of the role of key components



rAAV vector expression cassette developed from Doshi BS, Arruda VR. 2018.[4]
*Maximum packaging capacity within the rAAV capsid.
AAV: Adeno-associated virus; CMV: Cytomegalovirus; cDNA: Complementary DNA; ITR: Inverted terminal repeat; mRNA: Messenger RNA; Poly(A): Polyadenylation signal sequence; rAAV: Recombinant adeno-associated virus.

Gene therapies for hemophilia are currently being studied to determine their safety and efficacy. Approved gene therapies for hemophilia may have different labelling in different countries.

Differences between wild-type adeno-associated virus and rAAV vectors

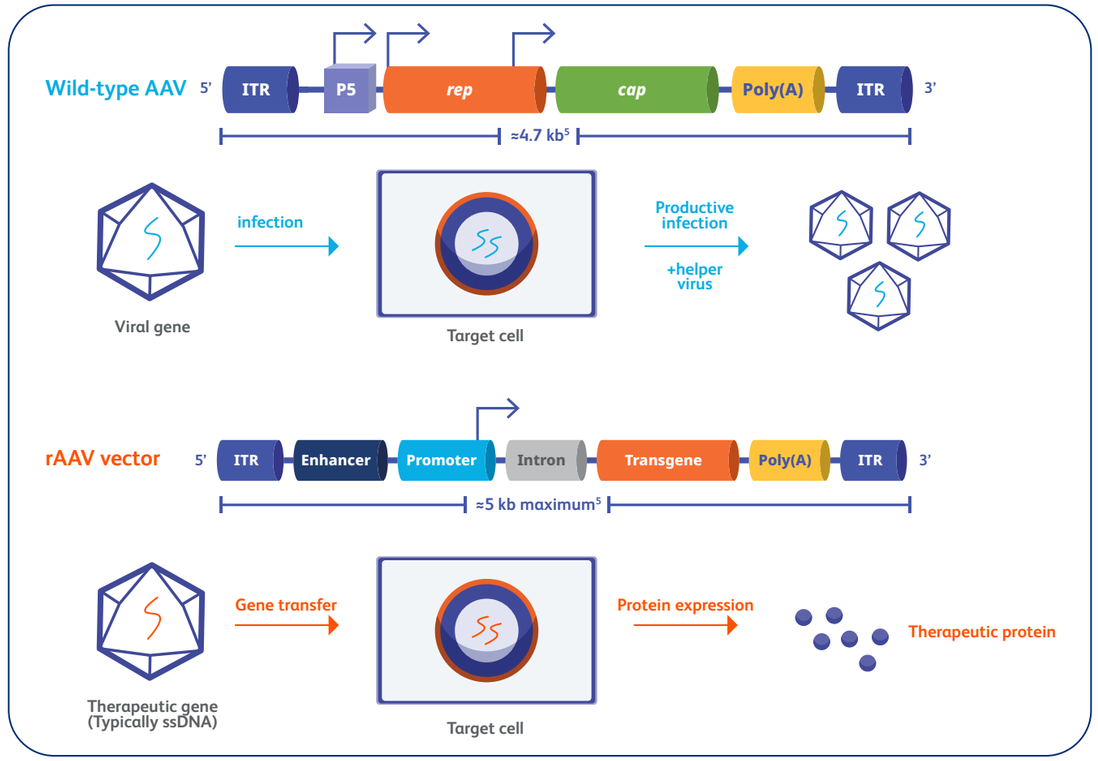


Figure adapted from Aponte-Ubillus JJ, et al. 2018[33] and Pfeifer A, Verma IM. 2001.[34]
AAV: Adeno-associated virus; ITR: Inverted terminal repeat; Poly (A): Polyadenylation signal sequence; P5: AAV promoter; rAAV: Recombinant AAV; ssDNA: Single-stranded DNA.

Visit www.genetherapyscience.com/vector for more information on the difference between wild-type AAV and rAAV vectors

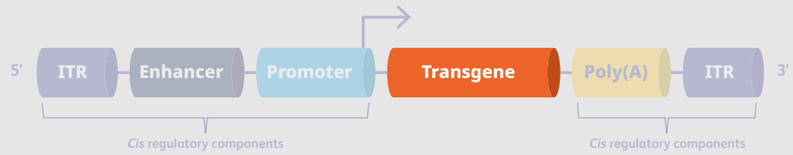
Each component of the rAAV vector expression cassette can be modified to optimize transgene expression

Why optimize the rAAV vector expression cassette?[2,3]

Overcome the packaging limitations of rAAV capsid	Establish transgene expression specifically in target issue, and limit vector uptake by non-target issue
Enhance protein expression and protein activity	Reduce the likelihood of triggering immune responses against the transgene, which could result in loss of transgene expression

To explore how the vector capsid can be optimized, visit www.genetherapyscience.com/vector and access the brochures and webinars available in the Learning Center

Focus on the transgene



Modifying transgene size to overcome rAAV vector packaging limitations

For successful transgene expression, the transgene and *cis* regulatory components must fit within the packaging capacity (~5 kb) of the rAAV capsid to enable delivery to the target cells[5]

- If a transgene exceeds the packaging capacity of the rAAV capsid, the cDNA may need to be altered to overcome this barrier[35]
- ~6% of all human proteins have a coding sequence >4 kb, meaning it is challenging to fit the cDNA for such proteins into a single rAAV vector genome[5]

What is cDNA?

cDNA is synthesized by reverse transcription from mRNA and so only contains the coding sequences or exons. 'Genomic' DNA by contrast is comprised of both coding and non-coding sequences[36]

For hemophilia:

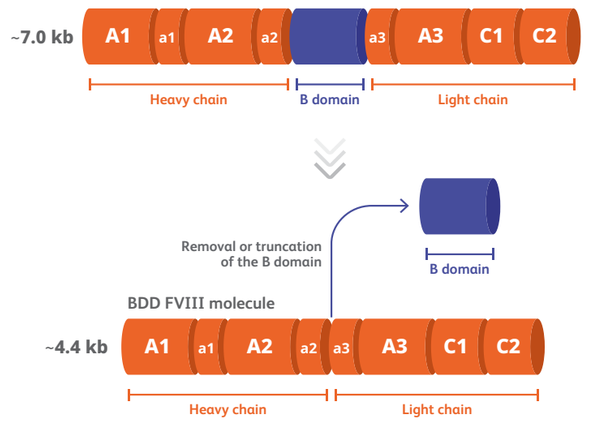
- The *F9* gene is relatively small at only ~1.6 kb and can be packaged into the rAAV expression cassette without modifications to its size[4]
- The *F8* gene is ~7 kb, which is too large to insert into an rAAV vector without modification[4,5]

Gene therapies for hemophilia are currently being studied to determine their safety and efficacy. Approved gene therapies for hemophilia may have different labelling in different countries.

Genetic engineering of cDNA

For larger genes (e.g., *F8*), reducing the size of the transgene – by removing sections of the cDNA that do not impair the function of the mature protein or impact on the level of expression of post-translational modification – has been used successfully to generate rAAV vector cassettes[5,35]

- The removal or truncation of the B domain (~2.6 kb) in *F8* can reduce the cDNA to ~4.4 kb[4,37]
- Removal or truncation of the *F8* B domain has been established in some approved recombinant FVIII products[37,38]



The structure of the *F8* gene is well understood[39]

- Three homologous A domains[39]
- Two homologous C domains[39]
- Unique B domain constituting ~38% of the cDNA[39]
  - Can be removed without loss of FVIII procoagulation activity.[39] Increased mRNA levels and secreted FVIII protein have been observed using B-domain-deleted (BDD) *F8* compared to wild-type *F8*[9*]
  - BDD FVIII is widely used as a replacement factor for the treatment of people with hemophilia A[9*]
  - Complete deletion has been associated with impaired post-translational trafficking and secretion.[40] A partial deletion of the B domain of *F8* can be used to compensate for this[40]

Truncation of the B domain of the *F8* gene is under investigation in hemophilia A gene therapy studies[4,19]

Figure adapted from Lheriteau E, et al. 2015[9] and Saenko EL, et al. 2003.[39]
*BDD-FVIII refers to FVIII variants where the B-domain has been replaced by short peptide linkers.[37]
BDD: B-domain-deleted; cDNA: Complementary DNA; mRNA: Messenger RNA.

Investigational approaches – dual vectors

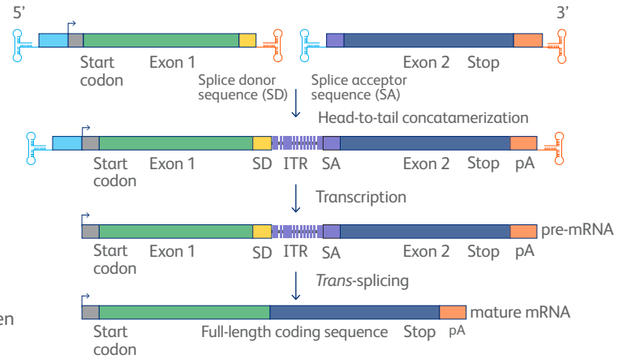
Another approach under investigation to overcome the packaging limitation of rAAV vectors is to use a dual-vector approach, where the transgene is delivered as two separate fragments within independent rAAV vectors[5,24,35]

- One rAAV vector includes the 5' transgene coding sequence with the promoter and enhancer elements[24,35]
- The other vector contains the 3' transgene coding sequence including the polyadenylation signal sequence – Poly(A) – which acts as the transcription terminator[24,35]

Different approaches to developing gene therapies for larger genes are currently at the pre-clinical stage; these include *trans*-splicing, overlapping, and hybrid mechanisms[35]

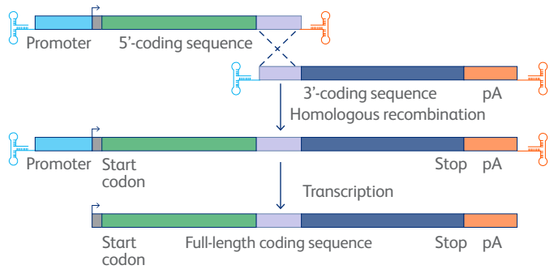
Trans-splicing[5,35]

- Takes advantage of the ITR-mediated intermolecular recombination and splicing ability of rAAV vector expression cassettes[5,35]
- Vector 1 encodes the promoter and 5' region of the transgene with a splice donor site[5,35]
- Vector 2 encodes the splice acceptor site, 3' region of the transgene and the Poly(A) terminator signal sequence[5,35]
- When both vector 1 and 2 co-transduce a cell, the viral genomes can form concatemers possibly through intermolecular recombination or ligation events[5]
- Pre-mRNA includes the recombined ITRs between the two coding regions. This region is removed through *trans*-splicing to result in a full-length coding sequence[5]



Overlapping[5,35]

- The 3' region of the first fragment and the 5' region of the second fragment are homologous[5,35]
- When delivered to the same nucleus, homologous recombination can occur generating a full-length coding sequence with all required regulatory elements[5,35]
- Following recombination, the expression cassette can then undergo transcription[5]



Hybrid[5,35]

- Combines the mechanisms of both the *trans*-splicing and the overlapping approaches[5,35]
- Vector 1 includes the promoter, 5' region of the transgene, a splice donor sequence, and a region that is homologous to a region just after the 5' ITR of vector 2[5,35]
- Vector 2 includes a homologous region (HR), splice acceptor sequence, the 3' region of the transgene, and the Poly(A) signal sequence[5,35]
- Both head-to-tail concatemerization and homologous recombination can occur[5]
- The splicing signals will allow splicing of the recombinogenic sequences, or the ITR structure to restore the large gene mRNA[35]

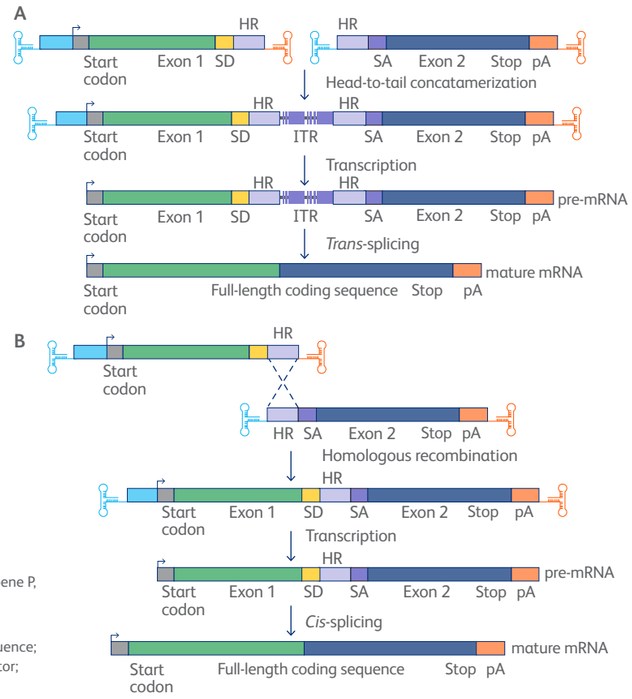


Figure adapted from Chamberlain K, et al. 2016[5] and Tornabene P, Trapani I. 2020.[35]

HR: homology region; ITR: Inverted terminal repeat; mRNA: Messenger RNA; Poly (A): Polyadenylation signal sequence; rAAV: Recombinant adeno-associated virus; SA: Splice acceptor; SD: Splice donor.

Gene therapies for hemophilia are currently being studied to determine their safety and efficacy. Approved gene therapies for hemophilia may have different labelling in different countries.

What is splicing and why does it matter?

Within the DNA sequence are exons, which code for amino acids, and introns, which are non-amino acid coding.[41] RNA splicing is the process by which introns are removed and exons joined together ("spliced") to create the full coding sequence of mRNA. [41,42] Mutations altering splicing can result in the development of different human diseases including cancer[42]

Studies exploring the efficacy of dual-vector approaches are often conflicting,[5,35] for the main part because several factors may impact the efficiency of dual vectors, including the transgene, target cell, the rAAV vector serotype used, and the doses chosen.[5,35] For the dual-vector approach to be successful, both parts of the expression cassette (i.e. vector 1 with the 5' region of the transgene and vector 2 with the 3' region of the transgene) must transduce the same cells.[5] Further research is required to understand the potential efficacy of different dual vector approaches when applied to hemophilia gene therapy[5]

Using natural gain-of-function gene variants to enhance protein activity

To optimize transgene expression and resultant protein activity, it is possible to use natural gain-of-function variants

- As wild-type *F9* is ~1.6 kb and can fit within an rAAV vector without modification to its size, optimization has focused on the use of a natural mutation that can increase resultant protein activity[4]
- The *F9* high-activity variant is a naturally occurring *F9* gene variant with an arginine to leucine substitution at position 338 of the gene (R338L),[30] resulting in a mature FIX-variant protein with an up to eight-fold increase in protein activity[4,40]
- By using a *F9* transgene with this gain-of-function point mutation in the *F9* gene, it may be possible to improve the activity of the resultant FIX protein[40]
- The single amino acid change coding for R338L has no impact on the size of the transgene, which therefore fits within the rAAV vector[5,40]
- The *F9* transgene with R338L gain-of-function point mutation is currently under investigation in clinical trials[29]

Codon optimization of the transgene to enhance mRNA translation

mRNA: Messenger RNA.

- mRNA nucleotides are translated into amino acids according to the genetic code. The message is read in groups of three nucleotides, resulting in a codon[15]
- There is a level of redundancy in the genetic code – most amino acids are encoded by more than one codon, with different organisms and cell types showing bias toward certain codons (i.e some codons are used more frequently than others – high abundance vs low abundance)[15,43,44]

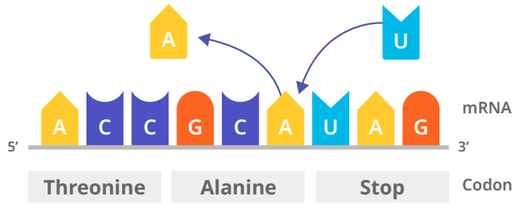
What is codon optimization?

- Rare codons are generally decoded by low-abundance transfer RNAs (tRNAs). The concentration of cognate tRNAs may be a key determinant influencing translation speed[44]
- The rate and efficiency of translation may be increased by matching codon usage in the transgene with the abundance of tRNAs for each codon in the target cell type (i.e., hepatocytes for hemophilia gene therapy) through the use of synonymous codon substitutions[44,45]

Hypothetical target cell tRNA usage

- Low-abundance alanine codons = GCA and GCC
- High-abundance alanine codon = GCU

Original mRNA from transgene



Codon optimization
Replacing the low-abundance codons with those that are in high abundance in the target cell

mRNA from codon-optimized transgene

May result in increased rate and efficiency of translation by using codons for the more abundant tRNAs in the target cell type[44,45]

mRNA: Messenger RNA; tRNA: Transfer RNA.

What are cognate tRNAs?

Cognate tRNAs are transfer RNA molecules that can be recognized by aminoacyl-tRNA synthetases,[46] the enzyme family that pairs tRNAs with the corresponding amino acids[47]

Gene therapies for hemophilia are currently being studied to determine their safety and efficacy. Approved gene therapies for hemophilia may have different labelling in different countries.

Transgene codon optimization is an approach being studied for hemophilia gene therapies to potentially maximize the expression and therapeutic potential of the *F8* or *F9* transgenes[8,40,44,45]

Codon optimization in hemophilia gene therapy

- Expression of the *F8* and *F9* transgenes has been improved through codon optimization[9,40,45]
  - Several different codon-optimized *F8* and *F9* transgenes have been demonstrated to improve expression levels of FVIII and FIX, respectively[40]

Considerations for codon optimization

Unintended changes to protein confirmation that may impact gene translation kinetics[44,45]

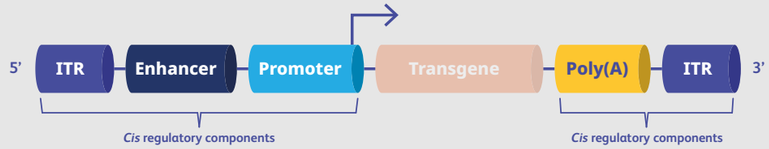
- Many proteins have complex structures dependent on proper folding and interactions with translocation machinery. Any mis-folding, potentially as a result of codon optimization, can result in loss-of-function of the protein[45]
- Codon optimization of *F9* may impact protein conformation and lead to different translation kinetics compared with wild-type *F9*[44]

Increase in CpG motifs and subsequent innate immune activation[48–51]

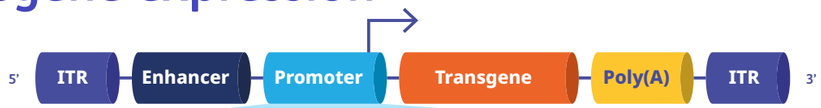
- CpG motifs – oligonucleotides of cytosine followed by guanine in a linear sequence of bases along the 5'-3' direction of the DNA strand – are pathogen-associated molecular patterns that can be recognized by the innate immune system via Toll-like receptors[48–50]
- Recognition of CpG motifs can result in the expression of pro-inflammatory cytokines or components of the type 1 interferon cascade, subsequently resulting in loss of transgene expression[48–50]
- CpG motifs generated during codon optimization should be carefully considered when developing transgene constructs for human gene therapy.[48] Loss of *F9* transgene expression, potentially due to CpG enrichment, has been observed in some clinical trial data[49]
- Codon optimization to remove CpG motifs may have a positive impact on transgene expression and avoid an innate immune response[51]

To learn more about capsid-triggered immune responses, see the brochure on "Immune responses associated with hemophilia gene therapy" or visit www.genetherapyscience.com

Focus on *cis* regulatory elements



Selecting the right promotor to optimize transgene expression



A sequence of DNA, typically in the 5' region of the expression cassette, where regulatory elements such as transcription factors bind and initiate transcription of the associated gene[28]

Various promoter types are available[27]

Constitutive / Ubiquitous	Tissue-specific	Inducible	Synthetic	Natural endogenous
• An unregulated promoter – constitutively active – that allows continual transcription of its associated gene in any cell type[32,52] • Examples: Immediate-early CMV, ubiquitin C, chicken beta-actin, etc.[32] • Used for approved drugs in other therapeutic areas[53]	• Used to limit expression to the desired cell types[27,32] • Often large and typically yield lower levels of expression compared with ubiquitous promoters[27,32] • Example: Endogenous liver-specific promoters – ApoE/hAAT and LP1 – have been used in hemophilia B gene therapies[27]	• Promoters designed with a layer of exogenous control, so they are induced under certain conditions[27] • Can be switched on and off when required[27] • Example: Tetracycline (Tet)-dependent system[27] • Evaluated in *in vivo* studies[54]	• Designed based on current knowledge and various computational methods[27] • Can be designed to be inducible under certain conditions[27] • May support enhanced tissue specificity[27] • Example: Synthetic liver-specific promoter – (HCB) – which has been investigated in animal models for use in hemophilia A[55] • Currently under clinical testing for hemophilia A[56]	• Natural endogenous promoters of the therapeutic gene offer the possibility of directing expression in physiologically relevant cells, at the appropriate level and time[57]

ApoE: Apolipoprotein E; CMV: Cytomegalovirus; hAAT: Human alpha-1-antitrypsin; HCB: Hepatic combinatorial bundle; ITR: Inverted terminal repeat; LP1: Liver-specific promoter 1; Poly (A): Polyadenylation signal sequence.

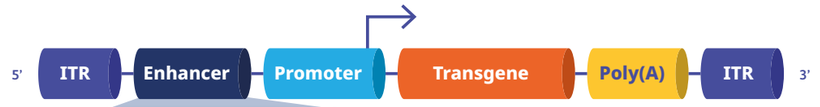
Promoter optimization in hemophilia gene therapy

> As the liver is the primary site of FVIII and FIX synthesis, current hemophilia gene therapy studies aim to establish expression in hepatocytes[40]
>
> – By using promoter sequences from genes naturally expressed in hepatocytes, the expression of the *F8* and *F9* genes can be targeted to the liver cells[2,28]

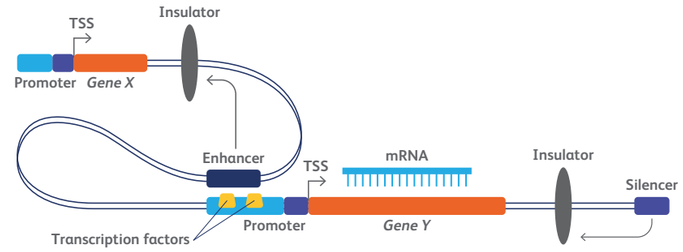
> Synthetic promoters to optimize transgene expression are also being investigated for hemophilia gene therapy – a minimal synthetic promoter has been used to construct an rAAV-*F8* vector genome for hepatocyte cell-specific expression.[27] AAV vectors utilizing synthetic liver-specific promoters are currently under clinical testing[56]

Gene therapies for hemophilia are currently being studied to determine their safety and efficacy. Approved gene therapies for hemophilia may have different labelling in different countries.

Introducing an enhancer element to the expression cassette



Supports promoter activity by providing binding sites for regulatory proteins involved in transcription[23]



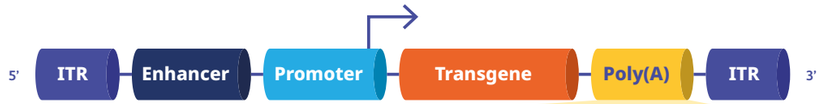
- Can be added upstream of the promoter sequence in the rAAV vector expression cassette[27]
- Supports enhanced transgene expression without impacting the selectivity of the promoter[27]
- Comprises binding sites for regulatory proteins that, when bound, trigger conformation changes that can affect RNA polymerase and transcription factor binding, thereby augmenting promoter activity[19]
- Can be added to circumvent weak promoter activity[27,32]
- Example: CMV enhancer[32]
  - CMV enhancers have been utilized in *in vivo* studies[27] and the first-in-human clinical trial for hemophilia B gene therapy[58]

CMV: Cytomegalovirus; ITR: Inverted terminal repeat; mRNA: Messenger RNA; Poly(A): Polyadenylation signal sequence; rAAV: Recombinant adeno-associated virus; TSS: Transcriptional start site.

Enhancer usage in hemophilia gene therapy

> Enhancer-promoter combinations are under investigation to optimize transgene expression in hemophilia gene therapies[58]

Engineering the Poly(A) terminator signal sequence



Polyadenylation via the Poly(A) terminator signal sequence is critical for mRNA stability, nuclear export, translation and subsequently the efficiency of transgene expression[32]

- Polyadenylation is signaled by the Poly(A) terminator signal sequence downstream of the transgene in the expression cassette[4,32]
  - The efficiency of transgene polyadenylation is key to transgene expression[32]
- Different Poly(A) signal sequences have different effects on transgene expression; for example, SV40 late or bovine growth hormone Poly(A) resulted in an increase in transgene expression compared with a minimal synthetic Poly(A) signal[32]
- Can be augmented by upstream signal enhancers (USE)
  - For example, the efficiency of polyadenylation can be increased by placing the SV40 late Poly(A) USE upstream of the Poly(A) signal[32]
- The Poly(A) signal sequence can also be modified to enhance transgene expression:
  - Synthetic Poly(A), which enhances polyadenylation, can be designed to increase nuclear export, translation and mRNA stability.[2] SV40, bovine growth hormone and modified synthetic Poly(A) sequences are being utilized in clinical trials in other therapy areas[13]
  - Reversed Poly(A) can support transgene expression by preventing transcription of the ITR.[2] Use of reversed Poly(A) has been suggested to decrease innate immune responses to the rAAV vector expression cassette;[2] however, at present, reversed Poly(A) sequences have not entered clinical evaluation[13]

ITR: Inverted terminal repeat; mRNA: Messenger RNA; Poly(A): Polyadenylation signal sequence; rAAV: Recombinant adeno-associated virus; USE: Upstream signal enhancers.

To learn more about the innate immune response to rAAV vector gene therapy, visit www.genetherapyscience.com/immunesystem

Engineering the Poly(A) terminator signal sequence in hemophilia gene therapy

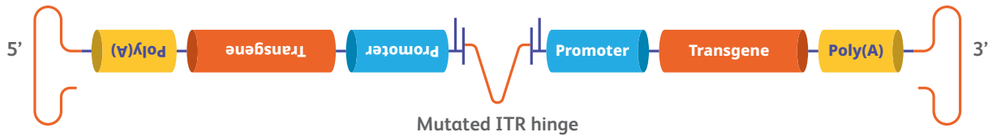
> For gene therapies where the transgene is large such as *F8* (hemophilia A gene therapy), the Poly(A) can be modified to help overcome the rAAV vector size limitations[55]
  – For example, replacing bovine growth hormone Poly(A) (0.26 kb) for a shorter synthetic Poly(A) (0.049 kb), to reduce the size of the rAAV-*F8* vector expression cassette[55]

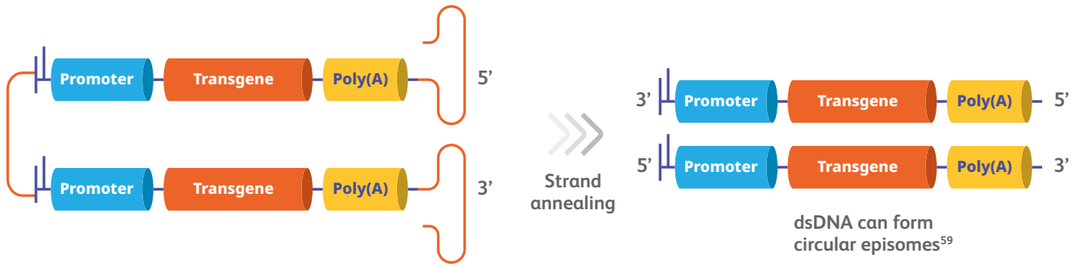
Gene therapies for hemophilia are currently being studied to determine their safety and efficacy. Approved gene therapies for hemophilia may have different labelling in different countries.

Modifying the ITRs to alter the expression cassette configuration

> In hemophilia gene therapy, rAAV vectors typically deliver ssDNA to the nucleus of liver cells[8]
  – To become transcriptionally active, ssDNA must be converted to dsDNA via a process called second-strand synthesis[3] – this is a rate-limiting step to gene expression[3,10]

Second-strand synthesis is a rate-limiting step to gene expression that can be bypassed by using self-complementary AAV vectors (scAAV vectors)[10]





dsDNA: Double-stranded DNA; ITR: Inverted terminal repeat; Poly(A): Polyadenylation signal sequence; scAAV: Self-complementary adeno-associated virus; scDNA: Self-complementary DNA.

scAAV vector expression cassette

- Mutation of one ITR terminal resolution site allows for the generation of scDNA[60,61]
  - Single DNA molecule with a hinge ITR at the 3’ end,[61] with a near-perfect complementarity[61]
- Includes both the coding and complementary sequence of the transgene expression cassette[3,10,59]
  - *De novo* DNA synthesis not required following delivery of the transgene[59]
  - Double-stranded by design so can undergo transcription after release in the target cell nucleus without need for second-strand synthesis[3]
- Designed as a single-stranded repeat, which can fold back upon itself to form dsDNA in the target cell nucleus, so bypassing the rate-limiting step[59,61]
  - Requires double-strand break repair proteins prior to circularization[59]

Gene therapies for hemophilia are currently being studied to determine their safety and efficacy. Approved gene therapies for hemophilia may have different labelling in different countries.

Limitation

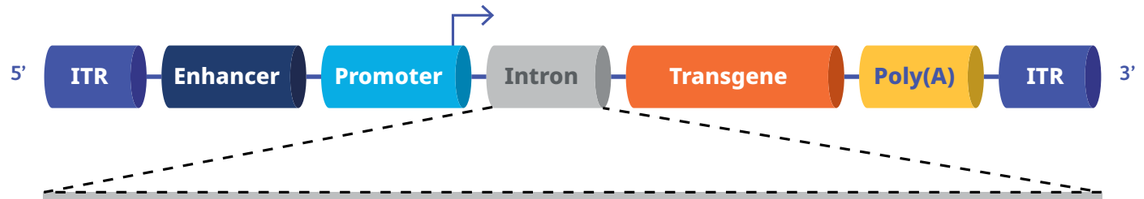
- Further restricts the packaging capacity to ~3.3 kb[3,10]
  - This approach is therefore applicable to gene therapies where the transgene is relatively small and is not currently an option for larger genes such as *F8*

scAAV vectors are under investigation for the delivery of *F9*[9,19]

Gene therapy trials with scAAV vectors are currently ongoing in other therapeutic areas[62]

> Additionally, modifying the ITRs may potentially reduce innate immune responses to the rAAV vector / genome[2]
  – The 5' and 3' ITR transcripts can form double-stranded RNA, which has been shown to activate the innate immune response via cytoplasmic RNA sensors (RIG-1 and MDA5), a response which may contribute to therapeutic failure[2]
  – Engineering ITRs that have no or weak promoter function could decrease the formation of dsRNA, potentially removing this concern[2]

Focus on introns



Non-coding regions of genes, introns,[63] can be placed between elements in the rAAV vector expression cassette to increase transgene expression[32]

> The first intron of *F9* has been shown to include an expression control sequence. Therefore, an intronic sequence can be inserted into the cDNA to impact transgene expression levels[28,63]

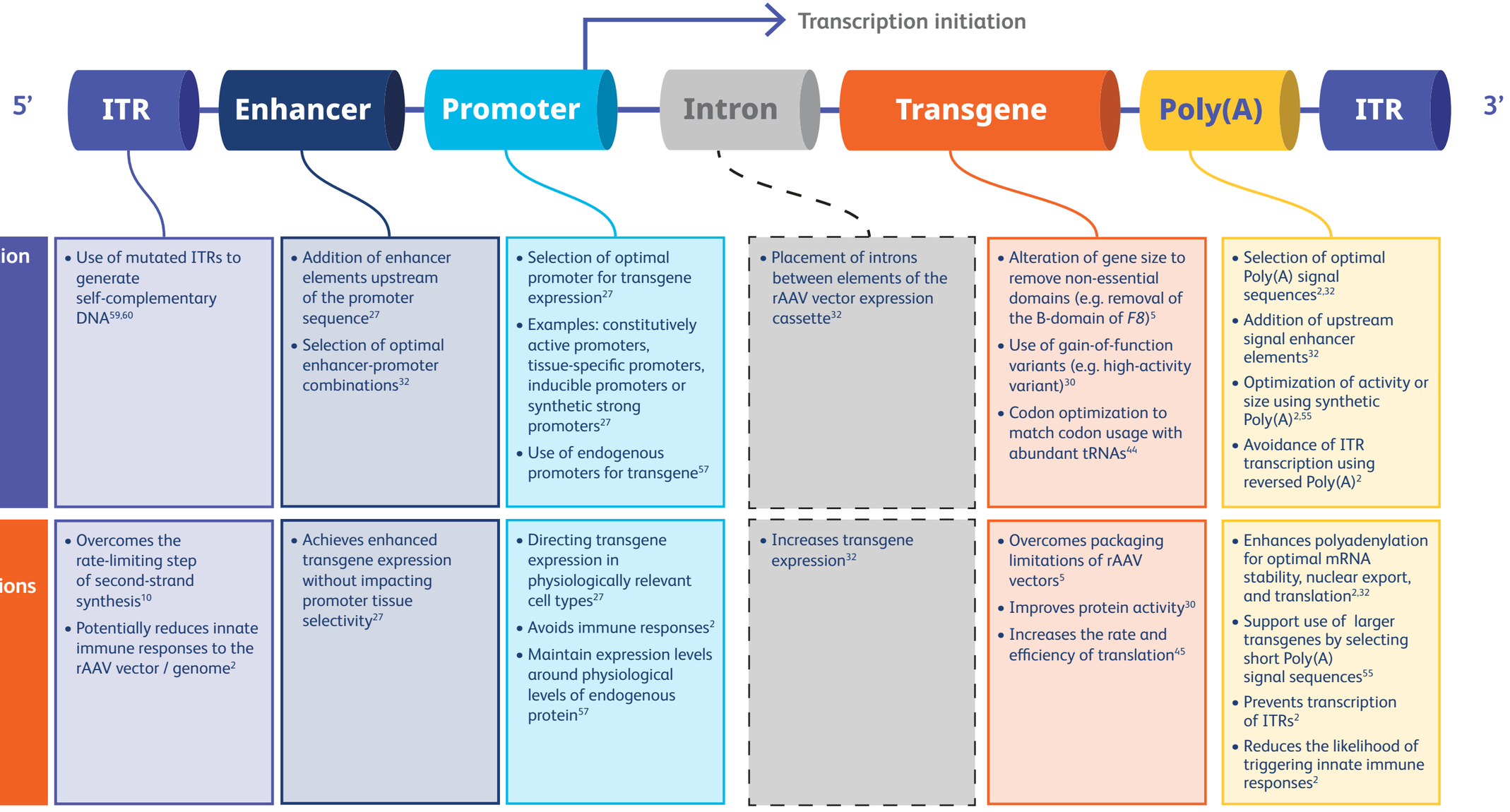
> In a study optimizing rAAV vectors for liver-directed gene therapy, it was shown that introducing an intron between the promoter and transgene increased transgene expression compared with when no intron was included. Of the introns investigated in the study, this effect was most apparent with the minute virus of mice (MVM) intron[32,64]

> Truncated versions of the first intron of *F9* have been evaluated in in vivo studies,[64] with one such vector under investigation in an ongoing clinical trial for hemophilia B[65]

cDNA: Complementary DNA; ITR: Inverted terminal repeat; Poly(A): Polyadenylation signal sequence; rAAV: Recombinant adeno-associated virus.

Key take-home messages

- The success of gene therapy is dependent not only on the development of an effective vector, but also on the development of an expression cassette that is optimized for protein production at the target site and to the appropriate level[32]
- The key elements of the rAAV vector expression cassette and their modifications are summarized in the figure below, along with their potential impact on transgene expression



rAAV vector expression cassette developed from Doshi BS, Arruda VR. 2018.[4]
ITR: Inverted terminal repeat; mRNA: Messenger RNA; Poly(A): Polyadenylation signal sequence; rAAV: Recombinant adeno-associated virus; tRNA: Transfer RNA.

Gene therapies for hemophilia are currently being studied to determine their safety and efficacy.
Approved gene therapies for hemophilia may have different labelling in different countries.

Glossary

3' (3 prime): The end of a single-stranded nucleic acid chain to which a hydroxyl group (-OH) is attached to the 3'-carbon atom of the nucleotide[66]

5' (5 prime): The end of a single-stranded nucleic acid chain to which a phosphate is attached to the 5'-carbon atom of the nucleotide[66]

Capsid: The protein shell of a virus that protects the genetic material while interacting with the host environment.[66] Capsid proteins determine cell-type specificity[29]

***Cis* regulatory components:** Noncoding sequences of DNA such as promoters, enhancers, and silencers, which regulate the transcription of nearby genes[67]

Clonal expansion: The process by which a single cell divides to produce genetically identical daughter cells[68]

Codon: A sequence of three nucleotides that codes a specific amino acid. For DNA, there are four different nucleotides (A, T, C, or G) from which a codon can be composed[68]

Cognate tRNAs: Transfer RNA (tRNA) molecules that can be recognized by aminoacyl-tRNA synthetases[45] – the enzyme family that pairs tRNAs with the corresponding amino acids[47,69]

Complementary DNA (cDNA): A DNA copy of a messenger RNA (mRNA) sequence, typically made synthetically using the reverse transcriptase enzyme. cDNA is different to genomic DNA because it only contains coding sequences (exons)[70]

Concatemer: A composite DNA molecule made of multiple copies of the same DNA molecule joined together in tandem[71]

CpG motifs: Oligonucleotides of cytosine followed by guanine in a linear sequence of bases along the 5'–3' direction of the DNA strand[48]

Endocytosis: A cellular process by which substances are brought into a cell. The substance is surrounded by an area of cell membrane, which then buds off inside the cell to form a vesicle containing the ingested material[72]

Endosomal escape: The process by which a molecule can escape the endosomal pathway, exit the endosome, and re-enter the cell cytoplasm[73]

Enhancer: A regulatory DNA sequence that provides binding sites to regulatory proteins and can augment the activity of a promoter[23]

Episomal DNA: Exogenous DNA that remains physically independent of the cell's endogenous chromosome or complement of chromosomes[74]

Exons: Coding sections of DNA or the RNA transcript which are translated into amino acids[66]

Expression cassette: The part of the vector DNA that contains the functional copy of the gene to be expressed and the regulatory sequences that allow protein production to occur[11]

Genome recombination: The rearrangement of DNA sequences[75]

Homologous recombination: A type of genetic recombination where nucleotide sequences are exchanged between similar or identical DNA molecules[76]

Intermolecular recombination: A type of genetic recombination occurring between two different DNA molecules[77]

Intron: A portion of DNA that does not code for an amino acid[78]

Inverted terminal repeats (ITRs): 145-bp sequences that frame the expression cassette[2]

Ligation: Joining together of nucleic acid molecules using a ligase enzyme[79]

Messenger RNA (mRNA): A single molecule of RNA that works as a chemical map for a protein product[6]

Non-integrating viral vector: Viral vectors that do not insert into a person's DNA or genome[80]

Nucleotide: The building blocks of nucleic acids (RNA and DNA), consisting of a sugar molecule attached to a phosphate group and a nitrogen-containing base[81]

Point mutation: A mutation in a single nucleotide in a DNA molecule[66]

Polyadenylation: Addition of adenine residues to the 3' end of RNA molecules by the enzyme, poly(A) polymerase[82]

Polyadenylation signal sequence (poly(A)): Acts as the transcription terminator, halting transcription once the transgene is fully transcribed[83]

Precursor messenger RNA (pre-mRNA): Immature sequences of messenger RNA (mRNA) produced following transcription. Pre-mRNA molecules undergo splicing to create mature mRNA that can then be translated into amino acids[66]

Promoter: Sequence of DNA, typically at the 5' region, where regulatory elements such as transcription factors bind and initiate transcription of the associated gene[28]

Recombinant adeno-associated virus (rAAV) vector: Gene therapy vector created by removing the protein-producing genes from the adeno-associated virus (AAV) genome and replacing them with the expression cassette for the intended gene[84]

Ribosomes: The cellular organelle where messenger RNA (mRNA) is translated into protein[66]

RNA polymerase II: A type of RNA polymerase enzyme, which transcribes pre-messenger RNA (pre-mRNA) from a DNA template[66]

RNA splicing: The process by which introns are removed from RNA and exons are joined together to create the full coding sequence[41]

Second-strand synthesis: The process by which single-stranded DNA (ssDNA) is converted into double-stranded DNA (dsDNA) through *de novo* synthesis of the complementary DNA (cDNA) strand[1]

Self-complementary AAV vector: An AAV vector in which the single-stranded genome complements itself to form double-stranded DNA (dsDNA) in the nucleus. This bypasses the rate-limiting step of converting the single-stranded transgene to a double-stranded transgene before it can be transcribed[10]

Serotype: Group of closely related microorganisms distinguished by a characteristic set of antigens and detected by an antibody[6]

Splice acceptor site/sequence: The border between an intron and an exon, downstream of the intron (in the direction 5' to 3'), which is required for splicing[85]

Splice donor site/sequence: The border between an intron and an exon, upstream of the intron (in the direction 5' to 3'), which is required for splicing[85]

***Trans*-splicing:** The process by which exons from two or more pre-messenger RNAs (pre-mRNAs) are spliced together[66]

Transcription: The process of copying the information in a strand of DNA into a new molecule of messenger RNA (mRNA)[66]

Transcription factors: Proteins which initiate and regulate the transcription of genes. They contain DNA-binding domains to bind to specific sequences of DNA[66]

Transcriptionally active: Ready to undergo transcription; for example, double-stranded DNA (dsDNA)[3]

Transcriptionally inert: Requires further processing before transcription is possible; for example, single-stranded DNA (ssDNA)[3]

Transduction: Transfer of genetic material into the nucleus of a cell, such that elements of the newly transferred DNA are then expressed. This can be accomplished naturally by a virus or other vector or experimentally by augmenting the receptivity of the cell membrane of the recipient cell with chemicals or electricity[74]

Transfer RNAs (tRNAs): Act as adaptors between messenger RNA (mRNA) and amino acids during translation. tRNAs have a bound amino acid and an anticodon loop that binds to the complementary mRNA codon[86]

Transgene: The nucleic acid sequence encoding an artificially added gene[11]

Translation: The process of synthesizing a protein from the information contained in messenger RNA (mRNA)[66]

Vector: A gene therapy delivery vehicle, which encapsulates a therapeutic gene and delivers it to target cells. Vectors can be either virus-derived or non-viral[6]

The information included in this brochure is accurate as of January 2023.
Please visit **www.genetherapyscience.com** for further information and check back regularly for updates

Gene therapies for hemophilia are currently being studied to determine their safety and efficacy. Approved gene therapies for hemophilia may have different labelling in different countries.

References

1. Colella P, et al. *Mol Ther Methods Clin Dev* 2018;8:87–104.
2. Li C, Samulski RJ. *Nat Rev Genet* 2020;21(4):255–72.
3. Wang D, et al. *Nat Rev Drug Discov* 2019;18(5):358–78.
4. Doshi BS, Arruda VR. *Ther Adv Hematol* 2018;9(9):273–93.
5. Chamberlain K, et al. *Hum Gene Ther Methods* 2016;27(1):1–12.
6. ASGCT: Glossary. https://asgct.org/education/more-resources/glossary (Accessed January 2023).
7. Kumar SR, et al. *Mol Ther Methods Clin Dev* 2016;3:16034.
8. Nathwani AC. *Hematology Am Soc Hematol Educ Program* 2019;2019(1):1–8.
9. Lheriteau E, et al. *Blood Rev* 2015;29(5):321–8.
10. Naso MF, et al. *BioDrugs* 2017;31:317–34.
11. Sidonio RF Jr., et al. *Blood Rev* 2021;47:100759.
12. Lisowski L, et al. *Res Pract Thromb Haemost* 2021;5(6):e12586.
13. Buck TM, Wijnholds J. *Int J Mol Sci* 2020;21(12):4197.
14. Clancy, S. *Nature Education* 2008;1(1):41. www.nature.com/scitable/topicpage/dna-transcription-426/ (Accessed January 2023).
15. Alberts B, et al. How Cells Read the Genome: From DNA to Protein. Chapter 6. In: *Molecular Biology of the Cell*. 4th edition. New York: Garland Science; 2002.
16. Alberts B, et al. Proteins. Chapter 3. In: *Molecular Biology of the Cell*. 4th edition. New York: Garland Science; 2002.
17. Mingozzi F, High KA. *Blood* 2013;122:23–36.
18. Gil-Farina I, et al. *Mol Ther* 2016;24(6):1100–5.
19. Batty P, Lillicrap D. *Hum Mol Genet* 2019;28(R1):R95–101.
20. Deyle DR, Russell DW. *Curr Opin Mol Ther* 2009;11(4):442–7.
21. Nguyen GN, et al. *Nat Biotechnol* 2021;39(1):47–55.
22. Food and Drug Administration: CTGTAC 2020 Meeting #70. www.fda.gov/media/151599/download (Accessed January 2023).
23. Nature Scitable: Gene Expression. www.nature.com/scitable/topicpage/gene-expression-14121669/ (Accessed January 2023).
24. McClements ME, MacLaren RE. *Yale J Biol Med* 2017;90(4):611–23.
25. Zhou Q, et al. *Sci Rep* 2017;7(1):5432.
26. Li L, Wunderlich Z. *Front Genet* 2017;8:63.
27. Domenger C, Grimm D. *Hum Mol Genet* 2019;28(R1):R3–14.
28. Ohmori T. *Int J Hematol* 2020;111(3):31–41.
29. George LA. *Blood Adv* 2017;(26):2591–9.
30. Simioni P, et al. *N Engl J Med* 2009;361(17):1671–5.
31. Proudfoot NJ. *Science* 2016;352(6291):aad9926.
32. Powell SK, et al. *Discov Med* 2015;19(102):49–57.
33. Aponte-Ubillus JJ, et al. *Appl Microbiol Biotechnol* 2018;102(3):1045–54.
34. Pfeifer A, Verma M. *Annu Rev Genomics Hum Genet* 2001;2:177–211.
35. Tornabene P, Trapani I. *Hum Gene Ther* 2020;31(1–2):47–56.
36. BioChain: cDNA vs Genomic DNA. www.biochain.com/blog/cdna-vs-genomic-dna/ (Accessed January 2023).
37. Samelson-Jones BJ, Arruda VR. *Mol Ther Methods Clin Dev* 2018;12:184–201.
38. Santagostino E. *Drug Des Devel Ther* 2014;8:2507–15.
39. Saenko EL, et al. *J Thromb Haemost* 2003;1(5):922–30.
40. VandenDriessche T, Chuah MK. *Hum Gene Ther* 2017;28(11):1013–23.
41. Clancy, S. *Nature Education* 2008;1(1):31. www.nature.com/scitable/topicpage/rna-splicing-introns-exons-and-spliceosome-12375/ (Accessed January 2023).
42. Wong MS, et al. *Mol Cell* 2018;71(6):1012–26.
43. Ciesla M. RNA in cancer. Chapter 11. In: Pandey R (ed). *RNA-based Regulation in Human Health and Disease*. Academic Press; 2020.
44. Alexaki A, et al. *Sci Rep* 2019;9(1):15449.
45. Kattenhorn LM, et al. *Human Gene Ther* 2016;27(12):947–61.
46. King R, et al. cognate tRNAs. In: *A Dictionary of Genetics*. 7th edition. Oxford University Press; 2007. Retrieved from Oxford Reference: www.oxfordreference.com/view/10.1093/acref/9780195307610.001.0001/acref-9780195307610-e-1252 (Accessed January 2023).
47. Gomez MAR, Ibba M. *RNA* 2020;26:910–36.
48. Molocea E, Schneider J. BioPharma Excellence. www.biopharma-excellence.com/news/2021/3/2/aav-gene-therapy-development-immunogenicity-and-safety-aspects (Accessed January 2023).
49. Konkle BA, et al. *Blood* 2021;137(6):763–74.
50. Kawasaki T, Kawai T. *Front Immunol* 2014;5:461.
51. Wright JF. *Mol Ther* 2020;28(3):701–3.
52. GenScript: Constitutive promoter. www.genscript.com/biology-glossary/9201/Constitutive-promoter (Accessed January 2023).
53. Wagner HJ, et al. *Adv Sci* 2021;8(9):2004018.
54. Seitzer JL, et al. *Mol Ther* 2006;13(1):S126.
55. Brown HC, et al. *Mol Ther Methods Clin Dev* 2018;9:57–69.
56. Patel SR, et al. *Front Immunol* 2020;11:494.
57. Le Bec C, Douar AM. *Gene Ther* 2006;13(10):805–13.
58. Arruda VR, Doshi BS. *Mediterr J Hematol Infect Dis* 2020;12(1):e2020069.
59. Raj D, et al. *Expert Rev Hematol* 2011;4(5):539–49.
60. Xie J, et al. *Mol Ther* 2017;25(6):1363–74.
61. Hirsch ML, et al. *Gene Ther* 2010;17(9):1175–80.
62. Buck TM, Wijnholds J. *Int J Mol Sci* 2020;21(12):4197. (Supplementary information).
63. Kurachi S, et al. *J Biol Chem* 1995;270:5276–81.
64. Wu Z, et al. *Mol Ther* 2008;16(2):280–9.
65. Peyvandi F, Garagiola I. *Haemophilia* 2019;25:738–46.
66. Nature Scitable: Glossary. www.nature.com/scitable/glossary/ (Accessed January 2023).
67. Wang Q, et al. *BMC Genomics* 2019;20(1):933.
68. National Human Genome Research Institute (NIH): Codon. www.genome.gov/genetics-glossary/codon (Accessed January 2023).
69. Pang LNJ, et al. *Wiley Interdiscip Rev RNA* 2014;5(4):461–80.
70. Hastings PJ. Complementary DNA (cDNA). In: *Encyclopedia of Genetics*. Academic Press; 2001, page 433. Retrieved from Science Direct: www.sciencedirect.com/topics/neuroscience/complementary-dna (Accessed January 2023).
71. Concatemer. In: *Farlex Partner Medical Dictionary*; 2012. Retrieved from The Free Dictionary: https://medical-dictionary.thefreedictionary.com/concatemer (Accessed January 2023).
72. Cooper GM. Endocytosis. In: *The Cell: A Molecular Approach*. 2nd edition. Sunderland (MA): Sinauer Associates; 2000. www.ncbi.nlm.nih.gov/books/NBK9831/ (Accessed January 2023).
73. Varkouhi AK, et al. *J Control Release* 2011;151(3):220–8.
74. NEJM: Illustrated Glossary. https://illustrated-glossary.nejm.org (Accessed January 2023).

References

75. Stahl FW. Conversion Gradient. In: *Reference Module in Life Sciences*. Academic Press, 2013, page 181–82. Retrieved from Science Direct: www.sciencedirect.com/topics/medicine-and-dentistry/genetic-recombination (Accessed January 2023).
76. National Human Genome Research Institute (NIH): Homologous Recombination. www.genome.gov/genetics-glossary/homologous-recombination (Accessed January 2023).
77. SDSU College of Sciences: Genetic Recombination. www.sci.sdsu.edu/~smaloy/MicrobialGenetics/topics/genetic-analysis/recombination/rec-genetic.html (Accessed January 2023).
78. National Human Genome Research Institute (NIH): Intron. www.genome.gov/genetics-glossary/Intron (Accessed January 2023).
79. Ligation. In: *Farlex Partner Medical Dictionary*; 2012. Retrieved from The Free Dictionary: https://medical-dictionary.thefreedictionary.com/ligation (Accessed January 2023).
80. Nowakowski A, et al. *Acta Neurobiol* 2013;73(1):1–18.
81. National Human Genome Research Institute (NIH): Nucleotide. www.genome.gov/genetics-glossary/Nucleotide (Accessed January 2023).
82. Hajnsdorf E, Kaberdin VR. *Phil Trans R Soc B* 2018;373:20180166.
83. Tran DP, et al. *Mol Cell Biol* 2001;21(21):7495–508.
84. Ronzitti G, et al. *Front Immunol* 2020;11:670.
85. Baten AKMA, et al. *BMC Bioinformatics* 2006;7(Suppl 5):S15.
86. Cooper GM. Translation of mRNA. In: *The Cell: A Molecular Approach*. 2nd edition. Sunderland (MA): Sinauer Associates; 2000. https://www.ncbi.nlm.nih.gov/books/NBK9849/ (Accessed January 2023).

Gene therapies for hemophilia are currently being studied to determine their safety and efficacy. Approved gene therapies for hemophilia may have different labelling in different countries.

For more information

The Hemophilia Gene Therapy Webinar Series explores the complex science underpinning hemophilia gene therapy studies. Hosted by an expert hematologist, joined by a specialist, each webinar focuses on providing a high-science review of key areas of interest in hemophilia gene therapy research.

Gene Therapy **Webinar Series**

Scan the QR Code and watch the webinar series or visit **www.genetherapyscience.com/learning-center**