OPTIMIZED FACTOR VIII GENES

RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Patent Application No. 63/236,225, filed August 23, 2021, the disclosure of which is hereby incorporated by reference in its entirety.

REFERENCE TO SEQUENCE LISTING SUBMITTED ELECTRONICALLY

[0002] The content of the electronically submitted sequence listing in XML format (Name: SA9-484_SeqListing.xml; Size: 117,240 bytes; Date of Creation: August 22, 2022) is incorporated herein by reference in its entirety.

BACKGROUND OF THE DISCLOSURE

[0003] A major impediment in providing a low-cost recombinant FVIII protein to patients is the high cost of commercial production. FVIII protein expresses poorly in heterologous expression systems, two to three orders of magnitude lower than similarly sized proteins. (Lynch et al., *Hum. Gene. Ther.*; 4:259–72 (1993). The poor expression of FVIII is due in part to the presence of cis-acting elements in the FVIII coding sequence that inhibit FVIII expression, such as transcriptional silencer elements (Hoeben et al., *Blood* 85:2447-2454 (1995)), matrix attachment-like sequences (MARs) (Fallux et al., *Mol. Cell. Biol.* 16:4264-4272 (1996)), and transcriptional elongation inhibitory elements (Koeberl et al., *Hum. Gene. Ther.*; 6:469-479 (1995)).

[0004] Thus, there exists a need in the art for FVIII sequences that express efficiently in heterologous systems.

SUMMARY OF THE DISCLOSURE

[0005] Disclosed are codon optimized nucleic acid molecules encoding a polypeptide with FVIII activity.

[0006] In certain aspects, disclosed herein is an isolated nucleic acid molecule comprising a nucleotide sequence at least 85% identical to SEQ ID NO: 9, wherein the nucleotide sequence encodes a polypeptide with factor VIII (FVIII) activity. In some embodiments, the nucleotide sequence is at least 90% identical to SEQ ID NO: 9. In some embodiments, the nucleotide sequence is at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO: 9. In some embodiments, the nucleotide sequence is at least 50% identical to SEQ ID NO: 9.

[0007] Also disclosed herein is an isolated nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 9, wherein the nucleotide sequence encodes a polypeptide with Factor VIII activity.

[0008] Also disclosed herein is an isolated nucleic acid molecule comprising a nucleotide sequence at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to nucleotides 58-4824 of SEQ ID NO: 9. In some embodiments, the isolated nucleic acid molecule comprises nucleotides 58-4824 of SEQ ID NO: 9.

[0009] In certain aspects, disclosed herein is an isolated nucleic acid molecule comprising a nucleotide sequence at least 85% identical to SEQ ID NO: 33, wherein the nucleotide sequence encodes a polypeptide with factor VIII (FVIII) activity. In some embodiments, the nucleotide sequence is at least 90% identical to SEQ ID NO: 33. In some embodiments, the nucleotide sequence is at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO: 33. In some embodiments, the nucleotide sequence is at least 50% identical to SEQ ID NO: 33.

[0010] In some embodiments, the isolated nucleic acid molecule disclosed herein further comprises a nucleotide sequence encoding a signal peptide. In some embodiments, the nucleotide sequence encodes a signal peptide comprises the amino acid sequence of SEQ ID NO: 11.

[0011] In some embodiments, the isolated nucleic acid molecule disclosed herein is codon-optimized to contain fewer CpG motifs than SEQ ID NO: 32. In some embodiments, the isolated nucleic acid molecule disclosed herein has one or more CpG motifs depleted relative to SEQ ID NO: 32.

[0012] In another aspect, disclosed herein is an isolated nucleic acid molecule comprising a genetic cassette expressing a factor VIII (FVIII) polypeptide, wherein the genetic cassette comprises a nucleotide sequence at least 85% identical to SEQ ID NO: 14. In some embodiments, the genetic cassette comprises a nucleotide sequence that is at least 90% identical to SEQ ID NO: 14. In some embodiments, the genetic cassette comprises a nucleotide sequence that is at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO: 14. In some embodiments, the nucleotide sequence is at least 50% identical to SEQ ID NO: 14.

[0013] Also disclosed herein is an isolated nucleic acid molecule comprising a genetic cassette expressing a factor VIII (FVIII) polypeptide, wherein the genetic cassette comprises the nucleotide sequence of SEQ ID NO: 14.

[0014] In another aspect, disclosed herein is an isolated nucleic acid molecule comprising a genetic cassette expressing a factor VIII (FVIII) polypeptide, wherein the genetic cassette comprises a nucleotide sequence at least 85% identical to SEQ ID NO: 35. In some embodiments, the genetic cassette comprises a nucleotide sequence that is at least 90% identical to SEQ ID NO: 35. In some embodiments, the genetic cassette comprises a nucleotide sequence that is at

least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO: 35. In some embodiments, the nucleotide sequence is at least 50% identical to SEQ ID NO: 35.

[0015] Also disclosed herein is an isolated nucleic acid molecule comprising a genetic cassette expressing a factor VIII (FVIII) polypeptide, wherein the genetic cassette comprises the nucleotide sequence of SEQ ID NO: 35.

[0016] In another aspect, disclosed herein is an isolated nucleic acid molecule comprising a genetic cassette expressing a factor VIII (FVIII) polypeptide comprising: a nucleotide sequence encoding a FVIII protein comprising a nucleic acid sequence at least 85% identical to SEQ ID NO: 9 or SEQ ID NO: 33; a promoter controlling transcription of the nucleotide sequence, and a transcription termination sequence.

[0017] In some embodiments, the promoter is a liver-specific promoter. In some embodiments, the promoter is a mouse transthyretin (mTTR) promoter. In some embodiments, the promoter is a mTTR482 promoter. In some embodiments, the promoter comprises the nucleotide sequence of SEQ ID NO: 16.

[0018] In some embodiments, the transcription termination sequence is a polyadenylation (polyA) sequence. In some embodiments, the transcription termination sequence is a Bovine Growth Hormone Polyadenylation (bGHpA) signal sequence. In some embodiments, the transcription termination sequence comprises the nucleotide sequence of SEQ ID NO: 19.

[0019] In some embodiments, the isolated nucleic acid molecule further comprises an enhancer element. In some embodiments, the enhancer element is an A1MB2 enhancer element. In some embodiments, the A1MB2 enhancer element comprises the nucleotide sequence of SEQ ID NO: 15.

[0020] In some embodiments, the isolated nucleic acid molecule further comprises an intronic sequence. In some embodiments, the intronic sequence is a chimeric intron, a hybrid intron, or a synthetic intron. In some embodiments, the intronic sequence comprises the nucleotide sequence of SEQ ID NO: 17.

[0021] In some embodiments, the isolated nucleic acid molecule further comprises a post-transcriptional regulatory element. In some embodiments, the post-transcriptional regulatory element comprises a Woodchuck Posttranscriptional Regulatory Element (WPRE). In some embodiments, the WPRE comprises the nucleotide sequence of SEQ ID NO: 18.

[0022] In another aspect, disclosed herein is an isolated nucleic acid molecule comprising a genetic cassette expressing a factor VIII (FVIII) polypeptide, a first inverted terminal repeat (ITR), and a second ITR flanking the genetic cassette. In some embodiments, the first ITR and/or the second ITR are derived from a member of the viral family *Parvoviridae.* In some embodiments,

the first ITR and/or the second ITR are derived from human Bocavirus (HBoV1), human erythrovirus (B19), Goose Parvovirus (GPV), or a variant thereof. In some embodiments, the first ITR and/or the second ITR comprises a polynucleotide sequence at least about 75% identical to SEQ ID NOs: 1, 2, or 21-30. In some embodiments, the first ITR comprises a polynucleotide sequence at least about 75% identical to SEQ ID NO: 1, and the second ITR comprises a polynucleotide sequence at least about 75% identical to SEQ ID NO: 2. In some embodiments, the first ITR comprises a polynucleotide sequence at least about 50% identical to SEQ ID NO: 1, and the second ITR comprises a polynucleotide sequence at least about 50% identical to SEQ ID NO: 2. In some embodiments, the first ITR comprises the polynucleotide sequence of SEQ ID NO: 1, and the second ITR comprises the polynucleotide sequence of SEQ ID NO: 2.

[0023] In another aspect, disclosed herein is an isolated nucleic acid molecule comprising a genetic cassette expressing a factor VIII (FVIII) polypeptide, wherein the genetic cassette comprises, from 5' to 3': an A1MB2 enhancer element comprising the nucleotide sequence of SEQ ID NO: 15, a liver-specific modified mouse transthyretin (mTTR) promoter (mTTR) comprising the nucleotide sequence of SEQ ID NO: 16, a chimeric intron comprising the nucleotide sequence of SEQ ID NO: 17, a nucleotide sequence encoding a FVIII protein comprising a nucleic acid sequence at least 85% identical to SEQ ID NO: 9 or SEQ ID NO: 33; a Woodchuck Posttranscriptional Regulatory Element (WPRE) comprising the nucleotide sequence of SEQ ID NO: 18; and a Bovine Growth Hormone Polyadenylation (bGHpA) signal comprising the nucleotide sequence of SEQ ID NO: 19.

[0024] In another aspect, disclosed herein is a vector comprising a nucleic acid molecule disclosed herein.

[0025] In another aspect, disclosed herein is a host cell comprising a nucleic acid molecule disclosed herein. Also disclosed herein are polypeptides produced by the host cell. In some embodiments, the host cell is an insect cell.

[0026] In another aspect, disclosed herein is a baculovirus system for production of a nucleic acid molecule disclosed herein. In some aspects, the nucleic acid molecule is produced in insect cells.

[0027] In another aspect, disclosed herein is a pharmaceutical composition comprising a nucleic acid molecule disclosed herein. In some embodiments, the pharmaceutical composition comprises a vector comprising a nucleic acid molecule disclosed herein. In some embodiments, the pharmaceutical composition further comprises a pharmaceutically acceptable excipient.

[0028] In another aspect, disclosed herein is a kit comprising a nucleic acid molecule disclosed herein and instructions for administering the nucleic acid molecule to a subject in need thereof.

[0029] In another aspect, disclosed herein is a method of producing a polypeptide with FVIII activity, comprising: culturing the host cell disclosed herein under conditions whereby a polypeptide with FVIII activity is produced, and recovering the polypeptide with FVIII activity.

[0030] In another aspect, disclosed herein is a method of increasing expression of a polypeptide with FVIII activity in a subject comprising administering a nucleic acid molecule comprising a nucleotide sequence at least 85% identical to SEQ ID NO: 9, SEQ ID NO: 33, SEQ ID NO: 35, or SEQ ID NO: 14. In some embodiments, the nucleic acid molecule comprises the nucleotide sequence of SEQ ID NO: 9. In some embodiments, the nucleic acid molecule comprises the nucleotide sequence of SEQ ID NO: 33. In some embodiments, the nucleic acid molecule comprises the nucleotide sequence of SEQ ID NO: 14. In some embodiments, the nucleic acid molecule comprises the nucleotide sequence of SEQ ID NO: 35.

[0031] In another aspect, disclosed herein is a method of treating a bleeding disorder in a subject comprising administering a nucleic acid molecule comprising a nucleotide sequence at least 85% identical to SEQ ID NO: 9, SEQ ID NO: 33, SEQ ID NO: 35, or SEQ ID NO: 14. In some embodiments, the nucleic acid molecule comprises the nucleotide sequence of SEQ ID NO: 9. In some embodiments, the nucleic acid molecule comprises the nucleotide sequence of SEQ ID NO: 33. In some embodiments, the nucleic acid molecule comprises the nucleotide sequence of SEQ ID NO: 14.In some embodiments, the nucleic acid molecule comprises the nucleotide sequence of SEQ ID NO: 35.

[0032] In another aspect, disclosed herein is a method of treating a bleeding disorder in a subject comprising administering a pharmaceutical composition comprising a nucleotide sequence at least 85% identical to SEQ ID NO: 9, SEQ ID NO: 33, SEQ ID NO: 35, or SEQ ID NO: 14. In some embodiments, the nucleic acid molecule comprises the nucleotide sequence of SEQ ID NO: 9. In some embodiments, the nucleic acid molecule comprises the nucleotide sequence of SEQ ID NO: 33. In some embodiments, the nucleic acid molecule comprises the nucleotide sequence of SEQ ID NO: 14. In some embodiments, the nucleic acid molecule comprises the nucleotide sequence of SEQ ID NO: 35.

[0033] In another aspect, disclosed herein is a method of treating hemophilia A in a subject comprising administering a pharmaceutical composition comprising a nucleotide sequence at least 85% identical to SEQ ID NO: 9, SEQ ID NO: 33, SEQ ID NO: 35, or SEQ ID NO: 14. In some embodiments, the nucleic acid molecule comprises the nucleotide sequence of SEQ ID NO: 9. In some embodiments, the nucleic acid molecule comprises the nucleotide sequence of SEQ ID NO: 33. In some embodiments, the nucleic acid molecule comprises the nucleotide sequence of SEQ ID NO: 14. In some embodiments, the nucleic acid molecule comprises the nucleotide sequence of SEQ ID NO: 35.

BRIEF DESCRIPTION OF THE DRAWINGS

[0034] **FIG. 1** shows schematic linear maps of human FVIIIXTEN expression constructs according to embodiments of the invention. The V1.0 cassette comprises codon optimized cDNA clone#6 encoding B-domain deleted human Factor VIII (BDD-FVIIIco6) fused with XTEN 144 peptide (FVIIIco6XTEN) under the regulation of Tristetraprolin (TTP) promoter, intron, the Woodchuck Posttranscriptional Regulatory Element (WPRE), and the Bovine Growth Hormone Polyadenylation (bGHpA) signal (see U.S. Publication No. 20190185543). The V2.0 cassette (SEQ ID NO: 14) comprises a codon optimized cDNA with further removal of CpG motifs encoding a B-domain deleted (BDD) codon-optimized human Factor VIII (BDDcoFVIII) fused with XTEN 144 peptide (FVIIIXTEN) under the regulation of liver-specific modified mouse transthyretin (mTTR) promoter (mTTR482) with enhancer element (A1MB2), hybrid synthetic intron (Chimeric Intron), the Woodchuck Posttranscriptional Regulatory Element (WPRE), and the Bovine Growth Hormone Polyadenylation (bGHpA) signal. The V3.0 cassette (SEQ ID NO: 35) comprises a codon optimized cDNA with further removal of CpG motifs encoding a B-domain deleted (BDD) codon-optimized human Factor VIII (Co-BDD-FVIII) fused with XTEN 144 peptide (FVIIIXTEN) under the regulation of liver-specific alpha-1-antitrypsin (A1AT) promoter, hybrid synthetic intron (Chimeric Intron), the Woodchuck Posttranscriptional Regulatory Element (WPRE), and the Bovine Growth Hormone Polyadenylation (bGHpA) signal. FVIIIXTEN expression cassettes are flanked by parvoviral ITRs.

[0035] **FIG. 2** shows a schematic representation of approach used for ssDNA generation, where a FVIIIXTEN expression cassette flanked by the parvoviral ITRs was digested with restriction enzymes that recognize the ITR related sequence and produce blunt-end DNA, and heat denatured (denaturation) the double-stranded DNA products (FVIII expression cassette and plasmid backbone) of digestion at 95 ^{o}C followed by cooling down (renaturation) at 4 ^{o}C to allow the palindromic ITR sequences to fold. The resulting ssFVIIIXTEN (ssDNA) was used for systemic delivery *via* hydrodynamic tail-vein injections in HemA mice.

[0036] **FIG. 3** shows a graphical representation of plasma FVIII activity levels measured by the Chromogenix Coatest® SP Factor VIII chromogenic assays. The blood samples were collected at different intervals from hFVIIIR593C$^{+/+}$/HemA mice systemically injected *via* hydrodynamic tail-vein injection with 800 µg/kg of single-stranded V1.0 or V2.0 ssFVIIIXTEN (ssDNA) flanked by the B19 ITRs. Error bars represents standard deviation.

[0037] **FIG. 4** shows a graphical representation of plasma FVIII activity levels measured by the Chromogenix Coatest® SP Factor VIII chromogenic assays. The plasma samples were collected at different intervals from hFVIIIR593C$^{+/+}$/HemA mice systemically injected *via*

hydrodynamic tail-vein injection with 200, 800, or 1600 µg/kg of single-stranded V2.0 ssFVIIIXTEN (ssDNA) flanked by human Bocavirus (HBoV1), human erythrovirus (B19), Goose Parvovirus (GPV), or their variant ITRs or their combinations as indicated. Two hybrid ITR sets were also tested (5'B19-3'GPV and 5'GPV-3'B19). Error bars represent standard deviation. The ITR sequences and their variants were described in previous U.S. Patent Application No. 63/069,114.

[0038] **FIGs. 5A-5B** are representations of the purified ceFVIIIXTEN (ceDNA) obtained from the baculovirus system and their efficacies *in vivo*. **FIG. 5A** shows an image of agarose gel electrophoresis of the purified ceFVIIIXTEN (ceDNA) with AAV2 or HBoV1 ITRs obtained from the continuous-elution electrophoresis, as described in U.S. Patent Application No. 63/069,073. The purity is shown in comparison with the starting material (SM) with arrows indicating DNA bands corresponding to the size of FVIIIXTEN ceDNA vector (ceDNA), baculoviral DNA (vDNA) and Sf9 cell genomic DNA (gDNA). **FIG. 5B** shows a graphical representation of plasma FVIII activity levels measured by the Chromogenix Coatest® SP Factor VIII chromogenic assays. The plasma samples were collected at different intervals from hFVIIIR593C$^{+/+}$/HemA mice systemically injected *via* hydrodynamic tail-vein injection with 80, 40, or 12 µg/kg of ceFVIIIXTEN (ceDNA) flanked by the AAV2 or HBoV1 ITRs as indicated. Error bars represents standard deviation. The ITR sequences and their variants were described in previous U.S. Patent Application No. 63/069,073.

[0039] **FIG. 6A-6C** shows the testing of the liver-specific mTTR and human A1AT promoter driving expression of FVIIIXTEN in HBoV1 ITR constructs. **FIG. 6A** shows a schematic diagram of FVIIIXTEN expression cassettes with either the liver-specific mTTR (SEQ ID NO: 3) or the A1AT promoter flanked by HBoV1 WT ITRs. **FIG. 6B** is an agarose gel electrophoresis image of single-stranded DNA (ssDNA) FVIIIXTEN HBoV1 generated by restriction enzyme digestion as described. **FIG. 6C** shows the FVIII expression levels normalized to percent of normal in mice injected with the mTTR or A1AT promoter constructs depicted in FIG. 6A. Error bars represent standard deviation.

[0040] **FIG. 7A-7C** show the study results for the purified ceFVIIIXTEN AAV2 (ceDNA) species obtained from the baculovirus system. **FIG. 7A** depicts an agarose gel electrophoresis image showing of full-length (8.3kb) and truncated (6.0kb) species of purified ceFVIIIXTEN (ceDNA) with AAV2 WT ITRs obtained from continuous-elution electrophoresis. **FIG. 7B** shows next-generation sequence (NGS) analyses of full-length 8.3kb ceFVIIIXTEN (top panel) and of truncated 6.0kb ceFVIIIXTEN (bottom panel) with AAV2 WT ITRs. **FIG. 7C** shows the FVIII expression levels normalized to percent of normal in mice injected with either the full-length or truncated ceFVIIIXTEN AAV2 constructs at either 80 or 40 µg/kg. Error bars represent standard deviation.

[0041] **FIGs. 8A-8B** are representations of the purified ceFVIIIXTEN (ceDNA) obtained from the baculovirus system and their efficacies *in vivo*. **FIG. 8A** shows an image of an agarose gel electrophoresis of the purified ceFVIIIXTEN (ceDNA) with AAV2 or HBoV1 ITRs obtained from the continuous-elution electrophoresis, as described in U.S. Patent Application No. 63/069,073. The purity is shown in comparison with the starting material (SM) with arrows indicating DNA bands corresponding to the size of FVIIIXTEN ceDNA vector (ceDNA), baculoviral DNA (vDNA) and Sf9 cell genomic DNA (gDNA). **FIG. 8B** shows the FVIII expression levels normalized to percent of normal in mice injected with either 80 or 40 µg/kg of ceFVIIIXTEN (ceDNA) flanked by the AAV2 or HBoV1 ITRs as indicated. Error bars represent standard deviation.

DETAILED DESCRIPTION

[0042] The present disclosure describes codon-optimized genes encoding polypeptides with Factor VIII (FVIII) activity. The present disclosure is directed to codon optimized nucleic acid molecules encoding polypeptides with Factor VIII activity, vectors, and host cells comprising optimized nucleic acid molecules, polypeptides encoded by optimized nucleic acid molecules, and methods of producing such polypeptides. The present disclosure is also directed to methods of treating bleeding disorders such as hemophilia comprising administering to the subject an optimized Factor VIII nucleic acid sequence, a vector comprising the optimized nucleic acid sequence, or the polypeptide encoded thereby.

[0043] The present disclosure meets an important need in the art by providing optimized FVIII sequences that demonstrate increased expression in host cells, improved yield of FVIII protein in methods to produce recombinant FVIII, and potentially result in greater therapeutic efficacy when used in gene therapy methods. In certain embodiments, the disclosure describes an isolated nucleic acid molecule comprising a nucleotide sequence which has sequence homology to the nucleotide sequence of SEQ ID NO: 9. In certain embodiments, the disclosure describes an isolated nucleic acid molecule comprising a nucleotide sequence which has sequence homology to the nucleotide sequence of SEQ ID NO: 33. In certain embodiments, the disclosure describes an isolated nucleic acid molecule comprising a nucleotide sequence which has sequence homology to the nucleotide sequence of SEQ ID NO: 14. In certain embodiments, the disclosure describes an isolated nucleic acid molecule comprising a nucleotide sequence which has sequence homology to the nucleotide sequence of SEQ ID NO: 35. In some embodiments, the genetic cassette further comprises a nucleotide sequence encoding an XTEN polypeptide.

[0044] In order to provide a clear understanding of the specification and claims, the following definitions are provided below.

Definitions

[0045] It is to be noted that the term "a" or "an" entity refers to one or more of that entity: for example, "a nucleotide sequence" is understood to represent one or more nucleotide sequences. As such, the terms "a" (or "an"), "one or more," and "at least one" can be used interchangeably herein.

[0046] The term "about" is used herein to mean approximately, roughly, around, or in the regions of. When the term "about" is used in conjunction with a numerical range, it modifies that range by extending the boundaries above and below the numerical values set forth. In general, the term "about" is used herein to modify a numerical value above and below the stated value by a variance of 10 percent, up or down (higher or lower).

[0047] The term "isolated" for the purposes of the present disclosure designates a biological material (cell, polypeptide, polynucleotide, or a fragment, variant, or derivative thereof) that has been removed from its original environment (the environment in which it is naturally present). For example, a polynucleotide present in the natural state in a plant or an animal is not isolated, however the same polynucleotide separated from the adjacent nucleic acids in which it is naturally present, is considered "isolated." No particular level of purification is required. Recombinantly produced polypeptides and proteins expressed in host cells are considered isolated for the purpose of the disclosure, as are native or recombinant polypeptides which have been separated, fractionated, or partially or substantially purified by any suitable technique.

[0048] "Nucleic acids," "nucleic acid molecules," "oligonucleotide," and "polynucleotide" are used interchangeably and refer to the phosphate ester polymeric form of ribonucleosides (adenosine, guanosine, uridine or cytidine; "RNA molecules") or deoxyribonucleosides (deoxyadenosine, deoxyguanosine, deoxythymidine, or deoxycytidine; "DNA molecules"), or any phosphoester analogs thereof, such as phosphorothioates and thioesters, in either single stranded form, or a double-stranded helix. Double stranded DNA-DNA, DNA-RNA and RNA-RNA helices are possible. The term nucleic acid molecule, and in particular DNA or RNA molecule, refers only to the primary and secondary structure of the molecule, and does not limit it to any particular tertiary forms. Thus, this term includes double-stranded DNA found, *inter alia*, in linear or circular DNA molecules (*e.g.,* restriction fragments), plasmids, supercoiled DNA and chromosomes. In discussing the structure of particular double-stranded DNA molecules, sequences can be described herein according to the normal convention of giving only the sequence in the 5' to 3' direction along the non-transcribed strand of DNA (*i.e.,* the strand having a sequence homologous to the mRNA). A "recombinant DNA molecule" is a DNA molecule that has undergone a molecular biological manipulation. DNA includes, but is not limited to, cDNA,

genomic DNA, plasmid DNA, synthetic DNA, and semi-synthetic DNA. A "nucleic acid composition" of the disclosure comprises one or more nucleic acids as described herein.

[0049] As used herein, a "coding region" or "coding sequence" is a portion of polynucleotide which consists of codons translatable into amino acids. Although a "stop codon" (TAG, TGA, or TAA) is typically not translated into an amino acid, it can be considered to be part of a coding region, but any flanking sequences, for example promoters, ribosome binding sites, transcriptional terminators, introns, and the like, are not part of a coding region. The boundaries of a coding region are typically determined by a start codon at the 5' terminus, encoding the amino terminus of the resultant polypeptide, and a translation stop codon at the 3' terminus, encoding the carboxyl terminus of the resulting polypeptide. Two or more coding regions can be present in a single polynucleotide construct, *e.g.*, on a single vector, or in separate polynucleotide constructs, *e.g.*, on separate (different) vectors. It follows, then, that a single vector can contain just a single coding region or comprise two or more coding regions.

[0050] Certain proteins secreted by mammalian cells are associated with a secretory signal peptide which is cleaved from the mature protein once export of the growing protein chain across the rough endoplasmic reticulum has been initiated. Those of ordinary skill in the art are aware that signal peptides are generally fused to the N-terminus of the polypeptide and are cleaved from the complete or "full-length" polypeptide to produce a secreted or "mature" form of the polypeptide. In certain embodiments, a native signal peptide or a functional derivative of that sequence that retains the ability to direct the secretion of the polypeptide that is operably associated with it. Alternatively, a heterologous mammalian signal peptide, *e.g.*, a human tissue plasminogen activator (TPA) or mouse ß-glucuronidase signal peptide, or a functional derivative thereof, can be used.

[0051] The term "downstream" refers to a nucleotide sequence that is located 3' to a reference nucleotide sequence. In certain embodiments, downstream nucleotide sequences relate to sequences that follow the starting point of transcription. For example, the translation initiation codon of a gene is located downstream of the start site of transcription.

[0052] The term "upstream" refers to a nucleotide sequence that is located 5' to a reference nucleotide sequence. In certain embodiments, upstream nucleotide sequences relate to sequences that are located on the 5' side of a coding region or starting point of transcription. For example, most promoters are located upstream of the start site of transcription.

[0053] As used herein, the term "genetic cassette" means a DNA sequence capable of directing expression of a particular polynucleotide sequence in an appropriate host cell, comprising a promoter operably linked to a polynucleotide sequence of interest. A genetic cassette may encompass nucleotide sequences located upstream (5' non-coding sequences),

within, or downstream (3' non-coding sequences) of a coding region, and which influence the transcription**,** RNA processing, stability, or translation of the associated coding region. If a coding region is intended for expression in a eukaryotic cell, a polyadenylation signal and transcription termination sequence will usually be located 3' to the coding sequence. In some embodiments, the genetic cassette comprises a polynucleotide which encodes a gene product. In some embodiments, the genetic cassette comprises a polynucleotide which encodes a miRNA. In some embodiments, the genetic cassette comprises a heterologous polynucleotide sequence. A polynucleotide which encodes a product, *e.g.*, a miRNA or a gene product (*e.g.*, a polypeptide such as a therapeutic protein), can include a promoter and/or other expression (*e.g.*, transcription or translation) control sequences operably associated with one or more coding regions. In an operable association a coding region for a gene product, *e.g.*, a polypeptide, is associated with one or more regulatory regions in such a way as to place expression of the gene product under the influence or control of the regulatory region(s). For example, a coding region and a promoter are "operably associated" if induction of promoter function results in the transcription of mRNA encoding the gene product encoded by the coding region, and if the nature of the linkage between the promoter and the coding region does not interfere with the ability of the promoter to direct the expression of the gene product or interfere with the ability of the DNA template to be transcribed. Other expression control sequences, besides a promoter, for example enhancers, operators, repressors, and transcription termination signals, can also be operably associated with a coding region to direct gene product expression.

[0054] "Expression control sequences" refer to regulatory nucleotide sequences, such as promoters, enhancers, terminators, and the like, that provide for the expression of a coding sequence in a host cell. Expression control sequences generally encompass any regulatory nucleotide sequence which facilitates the efficient transcription and translation of the coding nucleic acid to which it is operably linked. Non-limiting examples of expression control sequences include include promoters, enhancers, translation leader sequences, introns, polyadenylation recognition sequences, RNA processing sites, effector binding sites, or stem-loop structures. A variety of expression control sequences are known to those skilled in the art. These include, without limitation, expression control sequences which function in vertebrate cells, such as, but not limited to, promoter and enhancer segments from cytomegaloviruses (the immediate early promoter, in conjunction with intron-A), simian virus 40 (the early promoter), and retroviruses (such as Rous sarcoma virus). Other expression control sequences include those derived from vertebrate genes such as actin, heat shock protein, bovine growth hormone and rabbit ß-globin, as well as other sequences capable of controlling gene expression in eukaryotic cells. Additional suitable expression control sequences include tissue-specific promoters and enhancers as well

as lymphokine-inducible promoters (*e.g.*, promoters inducible by interferons or interleukins). Other expression control sequences include intronic sequences, post-transcriptional regulatory elements, and polyadenylation signals. Additional exemplary expression control sequences are discussed elsewhere in the present disclosure.

[0055] Similarly, a variety of translation control elements are known to those of ordinary skill in the art. These include, but are not limited to ribosome binding sites, translation initiation and termination codons, and elements derived from picornaviruses (particularly an internal ribosome entry site, or IRES).

[0056] The term "expression" as used herein refers to a process by which a polynucleotide produces a gene product, for example, an RNA or a polypeptide. It includes without limitation transcription of the polynucleotide into messenger RNA (mRNA), transfer RNA (tRNA), small hairpin RNA (shRNA), small interfering RNA (siRNA) or any other RNA product, and the translation of an mRNA into a polypeptide. Expression produces a "gene product." As used herein, a gene product can be either a nucleic acid, *e.g.*, a messenger RNA produced by transcription of a gene, or a polypeptide which is translated from a transcript. Gene products described herein further include nucleic acids with post transcriptional modifications, *e.g.*, polyadenylation or splicing, or polypeptides with post translational modifications, *e.g.*, methylation, glycosylation, the addition of lipids, association with other protein subunits, or proteolytic cleavage. The term "yield," as used herein, refers to the amount of a polypeptide produced by the expression of a gene.

[0057] A "vector" refers to any vehicle for the cloning of and/or transfer of a nucleic acid into a host cell. A vector can be a replicon to which another nucleic acid segment can be attached so as to bring about the replication of the attached segment. A "replicon" refers to any genetic element (*e.g.*, plasmid, phage, cosmid, chromosome, virus) that functions as an autonomous unit of replication *in vivo*, *i.e.*, capable of replication under its own control. The term "vector" includes both viral and nonviral vehicles for introducing the nucleic acid into a cell *in vitro*, *ex vivo* or *in vivo*. A large number of vectors are known and used in the art including, for example, plasmids, modified eukaryotic viruses, or modified bacterial viruses. Insertion of a polynucleotide into a suitable vector can be accomplished by ligating the appropriate polynucleotide fragments into a chosen vector that has complementary cohesive termini.

[0058] Vectors can be engineered to encode selectable markers or reporters that provide for the selection or identification of cells that have incorporated the vector. Expression of selectable markers or reporters allows identification and/or selection of host cells that incorporate and express other coding regions contained on the vector. Examples of selectable marker genes known and used in the art include: genes providing resistance to ampicillin, streptomycin, gentamycin, kanamycin, hygromycin, bialaphos herbicide, sulfonamide, and the like; and genes

that are used as phenotypic markers, *i.e.*, anthocyanin regulatory genes, isopentanyl transferase gene, and the like. Examples of reporters known and used in the art include: luciferase (Luc), green fluorescent protein (GFP), chloramphenicol acetyltransferase (CAT), β-galactosidase (LacZ), β-glucuronidase (Gus), and the like. Selectable markers can also be considered to be reporters.

[0059] The term "selectable marker" refers to an identifying factor, usually an antibiotic or chemical resistance gene, that is able to be selected for based upon the marker gene's effect, *i.e.*, resistance to an antibiotic, resistance to a herbicide, colorimetric markers, enzymes, fluorescent markers, and the like, wherein the effect is used to track the inheritance of a nucleic acid of interest and/or to identify a cell or organism that has inherited the nucleic acid of interest. Examples of selectable marker genes known and used in the art include: genes providing resistance to ampicillin, streptomycin, gentamycin, kanamycin, hygromycin, bialaphos herbicide, sulfonamide, and the like; and genes that are used as phenotypic markers, *i.e.*, anthocyanin regulatory genes, isopentanyl transferase gene, and the like.

[0060] The term "reporter gene" refers to a nucleic acid encoding an identifying factor that is able to be identified based upon the reporter gene's effect, wherein the effect is used to track the inheritance of a nucleic acid of interest, to identify a cell or organism that has inherited the nucleic acid of interest, and/or to measure gene expression induction or transcription. Examples of reporter genes known and used in the art include: luciferase (Luc), green fluorescent protein (GFP), chloramphenicol acetyltransferase (CAT), β-galactosidase (LacZ), β-glucuronidase (Gus), and the like. Selectable marker genes can also be considered reporter genes.

[0061] "Promoter" and "promoter sequence" are used interchangeably and refer to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. Promoters can be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters can direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental or physiological conditions. Promoters that cause a gene to be expressed in most cell types at most times are commonly referred to as "constitutive promoters." Promoters that cause a gene to be expressed in a specific cell type are commonly referred to as "cell-specific promoters" or "tissue-specific promoters." Promoters that cause a gene to be expressed at a specific stage of development or cell differentiation are commonly referred to as "developmentally-specific promoters" or "cell differentiation-specific promoters." Promoters that are induced and cause a gene to be expressed following exposure or treatment of the cell with an agent, biological

molecule, chemical, ligand, light, or the like that induces the promoter are commonly referred to as "inducible promoters" or "regulatable promoters." It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths can have identical promoter activity. Additional exemplary promoters are discussed elsewhere in the present disclosure.

[0062] The promoter sequence is typically bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence will be found a transcription initiation site (conveniently defined for example, by mapping with nuclease S1), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase.

[0063] The term "plasmid" refers to an extra-chromosomal element often carrying a gene that is not part of the central metabolism of the cell, and usually in the form of circular double-stranded DNA molecules. Such elements can be autonomously replicating sequences, genome integrating sequences, phage or nucleotide sequences, linear, circular, or supercoiled, of a single- or double-stranded DNA or RNA, derived from any source, in which a number of nucleotide sequences have been joined or recombined into a unique construction which is capable of introducing a promoter fragment and DNA sequence for a selected gene product along with appropriate 3' untranslated sequence into a cell.

[0064] Eukaryotic viral vectors that can be used include, but are not limited to, adenovirus vectors, retrovirus vectors, adeno-associated virus vectors, poxvirus, *e.g.*, vaccinia virus vectors, baculovirus vectors, or herpesvirus vectors. Non-viral vectors include plasmids, liposomes, electrically charged lipids (cytofectins), DNA-protein complexes, and biopolymers.

[0065] A "cloning vector" refers to a "replicon," which is a unit length of a nucleic acid that replicates sequentially and which comprises an origin of replication, such as a plasmid, phage or cosmid, to which another nucleic acid segment can be attached so as to bring about the replication of the attached segment. Certain cloning vectors are capable of replication in one cell type, *e.g.*, bacteria and expression in another, *e.g.*, eukaryotic cells. Cloning vectors typically comprise one or more sequences that can be used for selection of cells comprising the vector and/or one or more multiple cloning sites for insertion of nucleic acid sequences of interest.

[0066] The term "expression vector" refers to a vehicle designed to enable the expression of an inserted nucleic acid sequence following insertion into a host cell. The inserted nucleic acid sequence is placed in operable association with regulatory regions as described above.

[0067] Vectors are introduced into host cells by methods well known in the art, *e.g.*, transfection, electroporation, microinjection, transduction, cell fusion, DEAE dextran, calcium

phosphate precipitation, lipofection (lysosome fusion), use of a gene gun, or a DNA vector transporter.

[0068] "Culture," "to culture" and "culturing," as used herein, means to incubate cells under *in vitro* conditions that allow for cell growth or division or to maintain cells in a living state. "Cultured cells," as used herein, means cells that are propagated *in vitro*.

[0069] As used herein, the term "polypeptide" is intended to encompass a singular "polypeptide" as well as plural "polypeptides," and refers to a molecule composed of monomers (amino acids) linearly linked by amide bonds (also known as peptide bonds). The term "polypeptide" refers to any chain or chains of two or more amino acids, and does not refer to a specific length of the product. Thus, peptides, dipeptides, tripeptides, oligopeptides, "protein," "amino acid chain," or any other term used to refer to a chain or chains of two or more amino acids, are included within the definition of "polypeptide," and the term "polypeptide" can be used instead of, or interchangeably with any of these terms. The term "polypeptide" is also intended to refer to the products of post-expression modifications of the polypeptide, including without limitation glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, or modification by non-naturally occurring amino acids. A polypeptide can be derived from a natural biological source or produced recombinant technology, but is not necessarily translated from a designated nucleic acid sequence. It can be generated in any manner, including by chemical synthesis.

[0070] The term "amino acid" includes alanine (Ala or A); arginine (Arg or R); asparagine (Asn or N); aspartic acid (Asp or D); cysteine (Cys or C); glutamine (Gln or Q); glutamic acid (Glu or E); glycine (Gly or G); histidine (His or H); isoleucine (Ile or I): leucine (Leu or L); lysine (Lys or K); methionine (Met or M); phenylalanine (Phe or F); proline (Pro or P); serine (Ser or S); threonine (Thr or T); tryptophan (Trp or W); tyrosine (Tyr or Y); and valine (Val or V). Non-traditional amino acids are also within the scope of the disclosure and include norleucine, omithine, norvaline, homoserine, and other amino acid residue analogues such as those described in Ellman *et al.* Meth. Enzym. 202:301-336 (1991). To generate such non-naturally occurring amino acid residues, the procedures of Noren *et al.* Science 244:182 (1989) and Ellman *et al.*, supra, can be used. Briefly, these procedures involve chemically activating a suppressor tRNA with a non-naturally occurring amino acid residue followed by in vitro transcription and translation of the RNA. Introduction of the non-traditional amino acid can also be achieved using peptide chemistries known in the art. As used herein, the term "polar amino acid" includes amino acids that have net zero charge, but have non-zero partial charges in different portions of their side chains (*e.g.*, M, F, W, S, Y, N, Q, C). These amino acids can participate in hydrophobic interactions and electrostatic interactions. As used herein, the term "charged amino acid" includes amino acids

that can have non-zero net charge on their side chains (*e.g.*, R, K, H, E, D). These amino acids can participate in hydrophobic interactions and electrostatic interactions.

[0071] Also included in the present disclosure are fragments or variants of polypeptides, and any combination thereof. The term "fragment" or "variant" when referring to polypeptide binding domains or binding molecules of the present disclosure include any polypeptides which retain at least some of the properties (*e.g.*, FcRn binding affinity for an FcRn binding domain or Fc variant, coagulation activity for an FVIII variant, or FVIII binding activity for the VWF fragment) of the reference polypeptide. Fragments of polypeptides include proteolytic fragments, as well as deletion fragments, in addition to specific antibody fragments discussed elsewhere herein, but do not include the naturally occurring full-length polypeptide (or mature polypeptide). Variants of polypeptide binding domains or binding molecules of the present disclosure include fragments as described above, and also polypeptides with altered amino acid sequences due to amino acid substitutions, deletions, or insertions. Variants can be naturally or non-naturally occurring. Non-naturally occurring variants can be produced using art-known mutagenesis techniques. Variant polypeptides can comprise conservative or non-conservative amino acid substitutions, deletions or additions.

[0072] A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art, including basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, if an amino acid in a polypeptide is replaced with another amino acid from the same side chain family, the substitution is considered to be conservative. In another embodiment, a string of amino acids can be conservatively replaced with a structurally similar string that differs in order and/or composition of side chain family members.

[0073] The term "percent identity" as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case can be, as determined by the match between strings of such sequences. "Identity" can be readily calculated by known methods, including but not limited to those described in: *Computational Molecular Biology* (Lesk, A. M., ed.) Oxford University Press, New York (1988); *Biocomputing: Informatics and Genome Projects* (Smith, D. W., ed.) Academic Press, New York (1993); *Computer Analysis of Sequence Data, Part*

I (Griffin, A. M., and Griffin, H. G., eds.) Humana Press, New Jersey (1994); *Sequence Analysis in Molecular Biology* (von Heinje, G., ed.) Academic Press (1987); and *Sequence Analysis Primer* (Gribskov, M. and Devereux, J., eds.) Stockton Press, New York (1991). Preferred methods to determine identity are designed to give the best match between the sequences tested. Methods to determine identity are codified in publicly available computer programs. Sequence alignments and percent identity calculations can be performed using sequence analysis software such as the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI), the GCG suite of programs (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI), BLASTP, BLASTN, BLASTX (Altschul *et al*., *J. Mol. Biol. 215*:403 (1990)), and DNASTAR (DNASTAR, Inc. 1228 S. Park St. Madison, WI 53715 USA). Within the context of this application, it will be understood that where sequence analysis software is used for analysis, that the results of the analysis will be based on the "default values" of the program referenced, unless otherwise specified. As used herein "default values" will mean any set of values or parameters which originally load with the software when first initialized. For the purposes of determining percent identity between an optimized BDD FVIII sequence of the disclosure and a reference sequence, only nucleotides in the reference sequence corresponding to nucleotides in the optimized BDD FVIII sequence of the disclosure are used to calculate percent identity. For example, when comparing a full length FVIII nucleotide sequence containing the B domain to an optimized B domain deleted (BDD) FVIII nucleotide sequence of the disclosure, the portion of the alignment including the A1, A2, A3, C1, and C2 domain will be used to calculate percent identity. The nucleotides in the portion of the full length FVIII sequence encoding the B domain (which will result in a large "gap" in the alignment) will not be counted as a mismatch. In addition, in determining percent identity between an optimized BDD FVIII sequence of the disclosure, or a designated portion thereof (*e.g.*, nucleotides 2183-4474 and 4924-7006 of SEQ ID NO:14), and a reference sequence, percent identity will be calculated by aligning dividing the number of matched nucleotides by the total number of nucleotides in the complete sequence of the optimized BDD-FVIII sequence, or a designated portion thereof, as recited herein.

[0074] As used herein, the term "insertion site" refers to a position in a FVIII polypeptide, or fragment, variant, or derivative thereof, which is immediately upstream of the position at which a heterologous moiety can be inserted. An "insertion site" is specified as a number, the number corresponding to the number of the amino acid in mature native FVIII (SEQ ID NO: 20) to which the insertion site corresponds, which is immediately N-terminal to the position of the insertion. For example, the phrase "a3 comprises a heterologous moiety at an insertion site which corresponds to amino acid 1656 of SEQ ID NO: 24" indicates that the heterologous moiety is located between two amino acids corresponding to amino acid 1656 and amino acid 1657 of SEQ ID NO: 20.

[0075] The phrase "immediately downstream of an amino acid" as used herein refers to position right next to the terminal carboxyl group of the amino acid. Similarly, the phrase "immediately upstream of an amino acid" refers to the position right next to the terminal amine group of the amino acid.

[0076] The terms "inserted," "is inserted," "inserted into" or grammatically related terms, as used herein refers to the position of a heterologous moiety in a recombinant FVIII polypeptide, relative to the analogous position in native mature human FVIII (SEQ ID NO: 20).

[0077] As used herein, the term "half-life" refers to a biological half-life of a particular polypeptide *in vivo*. Half-life can be represented by the time required for half the quantity administered to a subject to be cleared from the circulation and/or other tissues in the animal. When a clearance curve of a given polypeptide is constructed as a function of time, the curve is usually biphasic with a rapid α-phase and longer β-phase. The α-phase typically represents an equilibration of the administered Fc polypeptide between the intra- and extra-vascular space and is, in part, determined by the size of the polypeptide. The β-phase typically represents the catabolism of the polypeptide in the intravascular space. In some embodiments, FVIII and chimeric proteins comprising FVIII are monophasic, and thus do not have an alpha phase, but just the single beta phase. Therefore, in certain embodiments, the term half-life as used herein refers to the half-life of the polypeptide in the β-phase.

[0078] The term "linked" as used herein refers to a first amino acid sequence or nucleotide sequence covalently or non-covalently joined to a second amino acid sequence or nucleotide sequence, respectively. The first amino acid or nucleotide sequence can be directly joined or juxtaposed to the second amino acid or nucleotide sequence or alternatively an intervening sequence can covalently join the first sequence to the second sequence. The term "linked" means not only a fusion of a first amino acid sequence to a second amino acid sequence at the C-terminus or the N-terminus, but also includes insertion of the whole first amino acid sequence (or the second amino acid sequence) into any two amino acids in the second amino acid sequence (or the first amino acid sequence, respectively). In one embodiment, the first amino acid sequence can be linked to a second amino acid sequence by a peptide bond or a linker. The first nucleotide sequence can be linked to a second nucleotide sequence by a phosphodiester bond or a linker. The linker can be a peptide or a polypeptide (for polypeptide chains) or a nucleotide or a nucleotide chain (for nucleotide chains) or any chemical moiety (for both polypeptide and polynucleotide chains). The term "linked" is also indicated by a hyphen (-).

[0079] As used herein the term "associated with" refers to a covalent or non-covalent bond formed between a first amino acid chain and a second amino acid chain. In one embodiment, the term "associated with" means a covalent, non-peptide bond or a non-covalent bond. This

association can be indicated by a colon, *i.e.,* (:). In another embodiment, it means a covalent bond except a peptide bond. For example, the amino acid cysteine comprises a thiol group that can form a disulfide bond or bridge with a thiol group on a second cysteine residue. In most naturally occurring IgG molecules, the CH1 and CL regions are associated by a disulfide bond and the two heavy chains are associated by two disulfide bonds at positions corresponding to 239 and 242 using the Kabat numbering system (position 226 or 229, EU numbering system). Examples of covalent bonds include, but are not limited to, a peptide bond, a metal bond, a hydrogen bond, a disulfide bond, a sigma bond, a pi bond, a delta bond, a glycosidic bond, an agnostic bond, a bent bond, a dipolar bond, a Pi backbond, a double bond, a triple bond, a quadruple bond, a quintuple bond, a sextuple bond, conjugation, hyperconjugation, aromaticity, hapticity, or antibonding. Non-limiting examples of non-covalent bond include an ionic bond (*e.g.,* cation-pi bond or salt bond), a metal bond, an hydrogen bond (*e.g.,* dihydrogen bond, dihydrogen complex, low-barrier hydrogen bond, or symmetric hydrogen bond), van der Walls force, London dispersion force, a mechanical bond, a halogen bond, aurophilicity, intercalation, stacking, entropic force, or chemical polarity.

[0080] “Hemostasis,” as used herein, means the stopping or slowing of bleeding or hemorrhage; or the stopping or slowing of blood flow through a blood vessel or body part.

[0081] “Hemostatic disorder,” as used herein, means a genetically inherited or acquired condition characterized by a tendency to hemorrhage, either spontaneously or as a result of trauma, due to an impaired ability or inability to form a fibrin clot. Examples of such disorders include the hemophilias. The three main forms are hemophilia A (factor VIII deficiency), hemophilia B (factor IX deficiency or "Christmas disease") and hemophilia C (factor XI deficiency, mild bleeding tendency). Other hemostatic disorders include, *e.g.,* von Willebrand disease, Factor XI deficiency (PTA deficiency), Factor XII deficiency, deficiencies or structural abnormalities in fibrinogen, prothrombin, Factor V, Factor VII, Factor X or factor XIII, Bernard-Soulier syndrome, which is a defect or deficiency in GPIb. GPIb, the receptor for vWF, can be defective and lead to lack of primary clot formation (primary hemostasis) and increased bleeding tendency), and thrombasthenia of Glanzman and Naegeli (Glanzmann thrombasthenia). In liver failure (acute and chronic forms), there is insufficient production of coagulation factors by the liver; this can increase bleeding risk.

[0082] The isolated nucleic acid molecules, isolated polypeptides, or vectors comprising the isolated nucleic acid molecule of the disclosure can be used prophylactically. As used herein the term "prophylactic treatment" refers to the administration of a molecule prior to a bleeding episode. In one embodiment, the subject in need of a general hemostatic agent is undergoing, or is about to undergo, surgery. A polynucleotide, polypeptide, or vector of the disclosure can be administered

prior to or after surgery as a prophylactic. The polynucleotide, polypeptide, or vector of the disclosure can be administered during or after surgery to control an acute bleeding episode. The surgery can include, but is not limited to, liver transplantation, liver resection, dental procedures, or stem cell transplantation.

[0083] The isolated nucleic acid molecules, isolated polypeptides, or vectors of the disclosure are also used for on-demand treatment. The term "on-demand treatment" refers to the administration of an isolated nucleic acid molecule, isolated polypeptide, or vector in response to symptoms of a bleeding episode or before an activity that can cause bleeding. In one aspect, the on-demand treatment can be given to a subject when bleeding starts, such as after an injury, or when bleeding is expected, such as before surgery. In another aspect, the on-demand treatment can be given prior to activities that increase the risk of bleeding, such as contact sports.

[0084] As used herein the term "acute bleeding" refers to a bleeding episode regardless of the underlying cause. For example, a subject can have trauma, uremia, a hereditary bleeding disorder (*e.g.,* factor VII deficiency) a platelet disorder, or resistance owing to the development of antibodies to clotting factors.

[0085] “Treat,” “treatment,” “treating,” as used herein refers to, *e.g.,* the reduction in severity of a disease or condition; the reduction in the duration of a disease course; the amelioration of one or more symptoms associated with a disease or condition; the provision of beneficial effects to a subject with a disease or condition, without necessarily curing the disease or condition, or the prophylaxis of one or more symptoms associated with a disease or condition. In one embodiment, the term "treating" or "treatment" means maintaining a FVIII trough level at least about 1 IU/dL, 2 IU/dL, 3 IU/dL, 4 IU/dL, 5 IU/dL, 6 IU/dL, 7 IU/dL, 8 IU/dL, 9 IU/dL, 10 IU/dL, 11 IU/dL, 12 IU/dL, 13 IU/dL, 14 IU/dL, 15 IU/dL, 16 IU/dL, 17 IU/dL, 18 IU/dL, 19 IU/dL, or 20 IU/dL in a subject by administering an isolated nucleic acid molecule, isolated polypeptide or vector of the disclosure. In another embodiment, treating or treatment means maintaining a FVIII trough level between about 1 and about 20 IU/dL, about 2 and about 20 IU/dL, about 3 and about 20 IU/dL, about 4 and about 20 IU/dL, about 5 and about 20 IU/dL, about 6 and about 20 IU/dL, about 7 and about 20 IU/dL, about 8 and about 20 IU/dL, about 9 and about 20 IU/dL, or about 10 and about 20 IU/dL. Treatment or treating of a disease or condition can also include maintaining FVIII activity in a subject at a level comparable to at least about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, or 20% of the FVIII activity in a non-hemophiliac subject. The minimum trough level required for treatment can be measured by one or more known methods and can be adjusted (increased or decreased) for each person.

[0086] "Administering," as used herein, means to give a pharmaceutically acceptable Factor VIII-encoding nucleic acid molecule, Factor VIII polypeptide, or vector comprising a Factor VIII-

encoding nucleic acid molecule of the disclosure to a subject via a pharmaceutically acceptable route. Routes of administration can be intravenous, *e.g.,* intravenous injection and intravenous infusion. Additional routes of administration include, *e.g.,* subcutaneous, intramuscular, oral, nasal, and pulmonary administration. The nucleic acid molecules, polypeptides, and vectors can be administered as part of a pharmaceutical composition comprising at least one excipient.

[0087] As used herein, the phrase "subject in need thereof" includes subjects, such as mammalian subjects, that would benefit from administration of a nucleic acid molecule, a polypeptide, or vector of the disclosure, *e.g.,* to improve hemostasis. In one embodiment, the subjects include, but are not limited to, individuals with hemophilia. In another embodiment, the subjects include, but are not limited to, the individuals who have developed a FVIII inhibitor and thus are in need of a bypass therapy. The subject can be an adult or a minor (*e.g.,* under 12 years old).

[0088] As used herein, the term "clotting factor," refers to molecules, or analogs thereof, naturally occurring or recombinantly produced which prevent or decrease the duration of a bleeding episode in a subject. In other words, it means molecules having pro-clotting activity, *i.e.,* are responsible for the conversion of fibrinogen into a mesh of insoluble fibrin causing the blood to coagulate or clot. An "activatable clotting factor" is a clotting factor in an inactive form (*e.g.,* in its zymogen form) that is capable of being converted to an active form.

[0089] “Clotting activity,” as used herein, means the ability to participate in a cascade of biochemical reactions that culminates in the formation of a fibrin clot and/or reduces the severity, duration or frequency of hemorrhage or bleeding episode.

[0090] As used herein the terms "heterologous" or "exogenous" refer to such molecules that are not normally found in a given context, *e.g.,* in a cell or in a polypeptide. For example, an exogenous or heterologous molecule can be introduced into a cell and are only present after manipulation of the cell, *e.g.,* by transfection or other forms of genetic engineering or a heterologous amino acid sequence can be present in a protein in which it is not naturally found.

[0091] As used herein, the term "heterologous nucleotide sequence" refers to a nucleotide sequence that does not naturally occur with a given polynucleotide sequence. In one embodiment, the heterologous nucleotide sequence encodes a polypeptide capable of extending the half-life of FVIII. In another embodiment, the heterologous nucleotide sequence encodes a polypeptide that increases the hydrodynamic radius of FVIII. In other embodiments, the heterologous nucleotide sequence encodes a polypeptide that improves one or more pharmacokinetic properties of FVIII without significantly affecting its biological activity or function (*e.g.,* its procoagulant activity). In some embodiments, FVIII is linked or connected to the polypeptide encoded by the heterologous nucleotide sequence by a linker.

[0092] A "reference nucleotide sequence," when used herein as a comparison to a nucleotide sequence of the disclosure, is a polynucleotide sequence essentially identical to the nucleotide sequence of the disclosure except that the portions corresponding to FVIII sequence are not optimized. In some embodiments, the reference nucleotide sequence for a nucleic acid molecule disclosed herein is SEQ ID NO: 32.

[0093] As used herein, the term "optimized," with regard to nucleotide sequences, refers to a polynucleotide sequence that encodes a polypeptide, wherein the polynucleotide sequence has been mutated to enhance a property of that polynucleotide sequence. In some embodiments, the optimization is done to increase transcription levels, increase translation levels, increase steady-state mRNA levels, increase or decrease the binding of regulatory proteins such as general transcription factors, increase or decrease splicing, or increase the yield of the polypeptide produced by the polynucleotide sequence. Examples of changes that can be made to a polynucleotide sequence to optimize it include codon optimization, G/C content optimization, removal of repeat sequences, removal of AT rich elements, removal of cryptic splice sites, removal of cis-acting elements that repress transcription or translation, adding or removing poly-T or poly-A sequences, adding sequences around the transcription start site that enhance transcription, such as Kozak consensus sequences, removal of sequences that could form stem loop structures, removal of destabilizing sequences, removal of CpG motifs, and two or more combinations thereof.

Polynucleotide Sequences

[0094] Certain aspects of the present disclosure aim to overcome deficiencies of AAV vectors for gene therapy. In particular, some aspects of the present disclosure are directed to a nucleic acid molecule comprising a genetic cassette, *e.g.*, encoding a therapeutic protein and/or a miRNA. In some embodiments, the genetic cassette encodes a therapeutic protein. In some embodiments, the therapeutic protein comprises a clotting factor. In some embodiments, the genetic cassette encodes a miRNA. In some embodiments, the nucleic acid molecule further comprises at least one noncoding region. In certain embodiments, the at least one non-coding region comprises a promoter sequence, an intron, a regulatory element, a 3'UTR poly(A) sequence, or any combination thereof. In some embodiments, the regulatory element is a post-transcriptional regulatory element.

[0095] In one embodiment, the genetic cassette is a single stranded nucleic acid. In another embodiment, the genetic cassette is a double stranded nucleic acid. In another embodiment, the genetic cassette is a closed-end double stranded nucleic acid (ceDNA).

[0096] In some embodiments, the genetic cassette comprises a nucleotide sequence encoding a FVIII polypeptide, wherein the nucleotide sequence is codon optimized. In some embodiments, the genetic cassette comprises a nucleotide sequence encoding a codon optimized FVIII driven by a mTTR promoter and synthetic intron. In some embodiments, the genetic cassette comprises a nucleotide sequence which is disclosed in International Application No. PCT/US2017/015879, which is incorporated by reference in its entirety. In some embodiments, the genetic cassette is a "hFVIIIco6XTEN" genetic cassette as described in PCT/US2017/015879. In some embodiments, the genetic cassette comprises SEQ ID NO: 32.

[0097] In some embodiments, the genetic cassette comprises codon optimized cDNA encoding B-domain deleted (BDD) codon-optimized human Factor VIII (BDDcoFVIII) fused with XTEN 144 peptide. In some embodiments, the genetic cassette comprises the nucleotide sequence set forth as SEQ ID NO: 9. In some embodiments, the genetic cassette comprises the nucleotide sequence set forth as SEQ ID NO: 14. In some embodiments, the genetic cassette has the nucleotide sequence of SEQ ID NO: 14. In some embodiments, the genetic cassette comprises the nucleotide sequence set forth as SEQ ID NO: 33. In some embodiments, the genetic cassette comprises the nucleotide sequence set forth as SEQ ID NO: 35. In some embodiments, the genetic cassette further comprises a nucleotide sequence encoding an XTEN polypeptide.

[0098] In some embodiments, the genetic cassette comprises a nucleotide sequence encoding a codon optimized FVIII driven by a mTTR promoter and synthetic intron. In some embodiments, the genetic cassette further comprises a a Woodchuck Posttranscriptional Regulatory Element (WPRE). In some embodiments, the genetic cassette further comprises the Bovine Growth Hormone Polyadenylation (bGHpA) signal.

[0099] In some embodiments, the present disclosure is directed to codon optimized nucleic acid molecules encoding a polypeptide with FVIII activity. In some embodiments, the polynucleotide encodes a full-length FVIII polypeptide. In other embodiments, the nucleic acid molecule encodes a B domain-deleted (BDD) FVIII polypeptide, wherein all or a portion of the B domain of FVIII is deleted. In one particular embodiment, the nucleic acid molecule encodes a polypeptide comprising an amino acid sequence having at least about 80%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to SEQ ID NO: 10 or a fragment thereof.

[0100] In some embodiments, the nucleic acid molecule of the disclosure encodes a FVIII polypeptide comprising a signal peptide or a fragment thereof. In other embodiments, the nucleic

acid molecule encodes a FVIII polypeptide which lacks a signal peptide. In some embodiments, the signal peptide comprises the amino acid sequence of SEQ ID NO: 11.

[0101] "A polypeptide with FVIII activity" as used herein means a functional FVIII polypeptide in its normal role in coagulation, unless otherwise specified. The term a polypeptide with FVIII activity includes a functional fragment, variant, analog, or derivative thereof that retains the function of full-length wild-type Factor VIII in the coagulation pathway. "A polypeptide with FVIII activity" is used interchangeably with FVIII protein, FVIII polypeptide, or FVIII. Examples of FVIII functions include, but are not limited to, an ability to activate coagulation, an ability to act as a cofactor for factor IX, or an ability to form a tenase complex with factor IX in the presence of Ca^{2+} and phospholipids, which then converts Factor X to the activated form Xa. In one embodiment, a polypeptide having FVIII activity comprises two polypeptide chains, the first chain having the FVIII heavy chain and the second chain having the FVIII light chain. In another embodiment, the polypeptide having FVIII activity is single chain FVIII. Single chain FVIII can contain one or more mutation or substitutions at amino acid residue 1645 and/or 1648 corresponding to mature human FVIII sequence (SEQ ID NO: 20). *See* International Application No. PCT/US2012/045784, incorporated herein by reference in its entirety. The FVIII protein can be the human, porcine, canine, rat, or murine FVIII protein. In addition, comparisons between FVIII from humans and other species have identified conserved residues that are likely to be required for function. See, e.g., Cameron et al. (1998) Thromb. Haemost. 79:317-22; and US Patent No. 6,251,632.

[0102] A number of tests are available to assess the FVIII activity of a polypeptide: activated partial thromboplastin time (aPTT) test, chromogenic assay, ROTEM® assay, prothrombin time (PT) test (also used to determine INR), fibrinogen testing (often by the Clauss method), platelet count, platelet function testing (often by PFA-100), TCT, bleeding time, mixing test (whether an abnormality corrects if the patient's plasma is mixed with normal plasma), coagulation factor assays, antiphosholipid antibodies, D-dimer, genetic tests (*e.g.,* factor V Leiden, prothrombin mutation G20210A), dilute Russell's viper venom time (dRVVT), miscellaneous platelet function tests, thromboelastography (TEG or Sonoclot), thromboelastometry (TEM®, e.g, ROTEM®), or euglobulin lysis time (ELT).

[0103] The aPTT test is a performance indicator measuring the efficacy of both the "intrinsic" (also referred to the contact activation pathway) and the common coagulation pathways. This test is commonly used to measure clotting activity of commercially available recombinant clotting factors, *e.g.*, FVIII or FIX. It is used in conjunction with prothrombin time (PT), which measures the extrinsic pathway.

[0104] ROTEM® analysis provides information on the whole kinetics of haemostasis: clotting time, clot formation, clot stability and lysis. The different parameters in thromboelastometry are

dependent on the activity of the plasmatic coagulation system, platelet function, fibrinolysis, or many factors which influence these interactions. This assay can provide a complete view of secondary haemostasis.

[0105] The "B domain" of FVIII, as used herein, is the same as the B domain known in the art that is defined by internal amino acid sequence identity and sites of proteolytic cleavage by thrombin, *e.g.,* residues Ser741-Arg1648 of full length human FVIII (SEQ ID NO: 20). The other human FVIII domains are defined by the following amino acid residues: A1, residues Ala1-Arg372; A2, residues Ser373-Arg740; A3, residues Ser1690-Ile2032; C1, residues Arg2033-Asn2172; C2, residues Ser2173-Tyr2332. The A3-C1-C2 sequence includes residues Ser1690-Tyr2332. The remaining sequence, residues Glu1649-Arg1689, is usually referred to as the FVIII light chain activation peptide. The locations of the boundaries for all of the domains, including the B domains, for porcine, mouse and canine FVIII are also known in the art. An example of a BDD FVIII is REFACTO® recombinant BDD FVIII (Wyeth Pharmaceuticals, Inc.).

[0106] A "B domain deleted FVIII" can have the full or partial deletions disclosed in U.S. Patent Nos. 6,316,226, 6,346,513, 7,041,635, 5,789,203, 6,060,447, 5,595,886, 6,228,620, 5,972,885, 6,048,720, 5,543,502, 5,610,278, 5,171,844, 5,112,950, 4,868,112, and 6,458,563, each of which is incorporated herein by reference in its entirety. Other examples of B domain deleted FVIII are disclosed in Hoeben R.C., *et al.* (1990) *J. Biol. Chem.* 265 (13): 7318-7323; Meulien *et al.* (1988), *Protein Eng.* 2(4): 301-6; Toole et al. (1986) Proc. Natl. Acad. Sci. U.S.A. 83, 5939-5942; Eaton, et al. (1986) Biochemistry 25:8343-8347; (Sarver, et al. (1987) DNA 6:553-564; European Patent No. 295597; and International Publication Nos. WO 91/09122, WO 88/00831, and WO 87/04187, each of which is incorporated herein by reference in its entirety. Each of the foregoing deletions can be made in any FVIII sequence.

Codon Optimization

[0107] In one embodiment, the present disclosure provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide with FVIII activity, wherein the nucleic acid sequence has been codon optimized. In another embodiment, the starting nucleic acid sequence that encodes a polypeptide with FVIII activity and that is subject to codon optimization is SEQ ID NO: 32. In some embodiments, the sequence that encodes a polypeptide with FVIII activity is codon optimized for human expression. In other embodiments, the sequence that encodes a polypeptide with FVIII activity is codon optimized for murine expression.

[0108] The term "codon-optimized" as it refers to genes or coding regions of nucleic acid molecules for transformation of various hosts, refers to the alteration of codons in the gene or coding regions of the nucleic acid molecules to reflect the typical codon usage of the host

organism without altering the polypeptide encoded by the DNA. Such optimization includes replacing at least one, or more than one, or a significant number, of codons with one or more codons that are more frequently used in the genes of that organism.

[0109] Deviations in the nucleotide sequence that comprises the codons encoding the amino acids of any polypeptide chain allow for variations in the sequence coding for the gene. Since each codon consists of three nucleotides, and the nucleotides comprising DNA are restricted to four specific bases, there are 64 possible combinations of nucleotides, 61 of which encode amino acids (the remaining three codons encode signals ending translation). As a result, many amino acids are designated by more than one codon. For example, the amino acids alanine and proline are coded for by four triplets, serine and arginine by six, whereas tryptophan and methionine are coded by just one triplet. This degeneracy allows for DNA base composition to vary over a wide range without altering the amino acid sequence of the proteins encoded by the DNA.

[0110] Many organisms display a bias for use of particular codons to code for insertion of a particular amino acid in a growing peptide chain. Codon preference, or codon bias, differences in codon usage between organisms, is afforded by degeneracy of the genetic code, and is well documented among many organisms. Codon bias often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, *inter alia*, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis. Accordingly, genes can be tailored for optimal gene expression in a given organism based on codon optimization.

[0111] Given the large number of gene sequences available for a wide variety of animal, plant and microbial species, the relative frequencies of codon usage have been calculated. Codon usage tables are available, for example, at the "Codon Usage Database" available at www.kazusa.or.jp/codon/ (visited June 18, 2012). *See* Nakamura, Y., et al. Nucl. Acids Res. 28:292 (2000).

[0112] Randomly assigning codons at an optimized frequency to encode a given polypeptide sequence can be done manually by calculating codon frequencies for each amino acid, and then assigning the codons to the polypeptide sequence randomly. Additionally, various algorithms and computer software programs can be used to calculate an optimal sequence.

[0113] In other embodiments, the nucleic acid molecules disclosed herein are further optimized by removal of one or more CpG motifs and/or the methylation of at least one CpG motif. As used herein, “CpG motif” refers to a dinucleotide sequence containing an unmethylated cytosine linked by a phosphate bond to a guanosine. The term “CpG motif” encompasses both methylated and unmethylated CpG dinucleotides. Unmethylated CpG motifs are common in

nucleic acid of bacterial and viral origin (e.g., plasmid DNA) but are suppressed and largely methylated in vertebrate DNA. Thus unmethylated CpG motifs stimulate the mammalian host to mount a rapid inflammatory response. Klinman, et al. (1996). PNAS 93:2879-2883. Exemplary methods of CpG removal are described in Yew, N. S., et al. (2002). Mol Ther. 5(6):731-738 and International Application No. PCT/US2001/010309. In some embodiments, the nucleic acid molecules disclosed herein have been modified to contain fewer CpG motifs (i.e. "CpG reduced" or "CpG depleted"). In one embodiment, the CpG motifs located within a codon triplet for a selected amino acid is changed to a codon triplet for the same amino acid lacking a CpG motif. In some embodiments, the nucleic acid molecules disclosed herein have been optimized to reduce innate immune response.

[0114] In some embodiments, disclosed herein is a nucleic acid molecule comprising a nucleotide sequence having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% sequence identity to SEQ ID NO: 9.

[0115] In some embodiments, disclosed herein is a nucleic acid molecule comprising a nucleotide sequence having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% sequence identity to SEQ ID NO: 33.

[0116] In some embodiments, disclosed herein is a nucleic acid molecule comprising a nucleotide sequence having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% sequence identity to SEQ ID NO: 14.

[0117] In some embodiments, disclosed herein is a nucleic acid molecule comprising a nucleotide sequence having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% sequence identity to SEQ ID NO: 35.

Heterologous Nucleotide Sequences

[0118] In some embodiments, the isolated nucleic acid molecules of the disclosure further comprise a heterologous nucleotide sequence. In some embodiments, the isolated nucleic acid molecules of the disclosure further comprise at least one heterologous nucleotide sequence. The heterologous nucleotide sequence can be linked with the optimized BDD-FVIII nucleotide sequences of the disclosure at the 5' end, at the 3' end, or inserted into the middle of the optimized BDD-FVIII nucleotide sequence. Thus, in some embodiments, the heterologous amino acid sequence encoded by the heterologous nucleotide sequence is linked to the N-terminus or the C-terminus of the FVIII amino acid sequence encoded by the nucleotide sequence or inserted between two amino acids in the FVIII amino acid sequence. In some embodiments, the heterologous amino acid sequence can be inserted between two amino acids at one or more insertion site. In some embodiments, the heterologous amino acid sequence can be inserted

within the FVIII polypeptide encoded by the nucleic acid molecule of the disclosure at any site disclosed in International Publication No. WO 2013/123457 A1, WO 2015/106052 A1 or U.S. Publication No. 2015/0158929 A1, each of which are incorporated by reference in their entirety.

[0119] In some embodiments, the heterologous amino acid sequence encoded by the heterologous nucleotide sequence is inserted within the B domain or a fragment thereof. In some embodiments, the heterologous amino acid sequence is inserted within the FVIII immediately downstream of an amino acid corresponding to amino acid 745 of wild type mature human FVIII (SEQ ID NO: 20). In one particular embodiment, the FVIII comprises a deletion of amino acids 746-1637, corresponding to wild type mature human FVIII (SEQ ID NO: 20), and the heterologous amino acid sequence encoded by the heterologous nucleotide sequence is inserted immediately downstream of amino acid 745, corresponding to wild type mature human FVIII (SEQ ID NO: 20). The insertion sites of FVIII referenced herein indicate the amino acid position corresponding to the amino acid position of wild type mature human FVIII (SEQ ID NO: 20).

[0120] In some embodiments, the heterologous moiety is a peptide or a polypeptide with either unstructured or structured characteristics that are associated with the prolongation of *in vivo* half-life when incorporated in a protein of the disclosure. Non-limiting examples include albumin, albumin fragments, Fc fragments of immunoglobulins, the C-terminal peptide (CTP) of the β subunit of human chorionic gonadotropin, a HAP sequence, an XTEN sequence, a transferrin or a fragment thereof, a PAS polypeptide, polyglycine linkers, polyserine linkers, albumin-binding moieties, or any fragments, derivatives, variants, or combinations of these polypeptides. In one particular embodiment, the heterologous amino acid sequence is an immunoglobulin constant region or a portion thereof, transferrin, albumin, or a PAS sequence. In other related aspects a heterologous moiety can include an attachment site (*e.g.*, a cysteine amino acid) for a non-polypeptide moiety such as polyethylene glycol (PEG), hydroxyethyl starch (HES), polysialic acid, or any derivatives, variants, or combinations of these elements. In some aspects, a heterologous moiety comprises a cysteine amino acid that functions as an attachment site for a non-polypeptide moiety such as polyethylene glycol (PEG), hydroxyethyl starch (HES), polysialic acid, or any derivatives, variants, or combinations of these elements.

[0121] In certain embodiments, a heterologous moiety improves one or more pharmacokinetic properties of the FVIII protein without significantly affecting its biological activity or function. In some embodiments, a heterologous moiety increases the *in vivo* and/or *in vitro* half-life of the FVIII protein of the disclosure. *In vivo* half-life of a FVIII protein can be determined by any methods known to those of skill in the art, *e.g.,* activity assays (chromogenic assay or one stage clotting aPTT assay), ELISA, ROTEM™, etc.

[0122] In other embodiments, a heterologous moiety increases stability of the FVIII protein of the disclosure or a fragment thereof (*e.g.,* a fragment comprising a heterologous moiety after proteolytic cleavage of the FVIII protein). As used herein, the term "stability" refers to an art-recognized measure of the maintenance of one or more physical properties of the FVIII protein in response to an environmental condition (*e.g.,* an elevated or lowered temperature). In certain aspects, the physical property can be the maintenance of the covalent structure of the FVIII protein (*e.g.,* the absence of proteolytic cleavage, unwanted oxidation or deamidation). In other aspects, the physical property can also be the presence of the FVIII protein in a properly folded state (*e.g.,* the absence of soluble or insoluble aggregates or precipitates). In one aspect, the stability of the FVIII protein is measured by assaying a biophysical property of the FVIII protein, for example thermal stability, pH unfolding profile, stable removal of glycosylation, solubility, biochemical function (*e.g.,* ability to bind to a protein, receptor or ligand), etc., and/or combinations thereof. In another aspect, biochemical function is demonstrated by the binding affinity of the interaction. In one aspect, a measure of protein stability is thermal stability, *i.e.,* resistance to thermal challenge. Stability can be measured using methods known in the art, such as, HPLC (high performance liquid chromatography), SEC (size exclusion chromatography), DLS (dynamic light scattering), etc. Methods to measure thermal stability include, but are not limited to differential scanning calorimetry (DSC), differential scanning fluorimetry (DSF), circular dichroism (CD), and thermal challenge assay.

[0123] In some embodiments, a heterologous moiety comprises one or more XTEN sequences, fragments, variants, or derivatives thereof. As used here "XTEN sequence" refers to extended length polypeptides with non-naturally occurring, substantially non-repetitive sequences that are composed mainly of small hydrophilic amino acids, with the sequence having a low degree or no secondary or tertiary structure under physiologic conditions. As a heterologous moiety, XTENs can serve as a half-life extension moiety. In addition, XTEN can provide desirable properties including but are not limited to enhanced pharmacokinetic parameters and solubility characteristics. Other advantageous properties which may be conferred by introducing an XTEN sequence include enhanced conformational flexibility, enhanced aqueous solubility, high degree of protease resistance, low immunogenicity, low binding to mammalian receptors, or increased hydrodynamic (or Stokes) radii.

[0124] XTEN can have varying lengths for insertion into or linkage to FVIII. In some embodiments, the XTEN sequence useful for the disclosure is a peptide or a polypeptide having greater than about 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1200, 1400, 1600, 1800, or 2000 amino acid residues. In certain embodiments, XTEN is a peptide or a polypeptide having greater than about

20 to about 3000 amino acid residues, greater than 30 to about 2500 residues, greater than 40 to about 2000 residues, greater than 50 to about 1500 residues, greater than 60 to about 1000 residues, greater than 70 to about 900 residues, greater than 80 to about 800 residues, greater than 90 to about 700 residues, greater than 100 to about 600 residues, greater than 110 to about 500 residues, or greater than 120 to about 400 residues. In one particular embodiment, the XTEN comprises an amino acid sequence of longer than 42 amino acids and shorter than 144 amino acids in length.

[0125] The XTEN sequence of the disclosure can comprise one or more sequence motif of 5 to 14 (e.g., 9 to 14) amino acid residues or an amino acid sequence at least 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to the sequence motif, wherein the motif comprises, consists essentially of, or consists of 4 to 6 types of amino acids (e.g., 5 amino acids) selected from the group consisting of glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P). *See* US 2010-0239554 A1.

[0126] Examples of XTEN sequences that can be used as heterologous moieties in chimeric proteins of the disclosure are disclosed, *e.g.*, in U.S. Patent Publication Nos. 2010/0239554 A1, 2010/0323956 A1, 2011/0046060 A1, 2011/0046061 A1, 2011/0077199 A1, or 2011/0172146 A1, or International Patent Publication Nos. WO 2010091122 A1, WO 2010144502 A2, WO 2010144508 A1, WO 2011028228 A1, WO 2011028229 A1, or WO 2011028344 A2, each of which is incorporated by reference herein in its entirety.

[0127] The one or more XTEN sequences can be inserted at the C-terminus or at the N-terminus of the amino acid sequence encoded by the nucleotide sequence or inserted between two amino acids in the amino acid sequence encoded by the nucleotide sequence. For example, the XTEN can be inserted between two amino acids at one or more insertion sites. Examples of sites within FVIII that are permissible for XTEN insertion can be found in, *e.g.*, International Publication No. WO 2013/123457 A1 or U.S. Publication No. 2015/0158929 A1, which are herein incorporated by reference in their entirety.

[0128] In certain embodiments, the heterologous moiety is a peptide linker.

[0129] As used herein, the terms "peptide linkers" or "linker moieties" refer to a peptide or polypeptide sequence (*e.g.,* a synthetic peptide or polypeptide sequence) which connects two domains in a linear amino acid sequence of a polypeptide chain.

[0130] In some embodiments, heterologous nucleotide sequences encoding peptide linkers can be inserted between the optimized FVIII polynucleotide sequences of the disclosure and a heterologous nucleotide sequence encoding, for example, one of the heterologous moieties described above, such as albumin. Peptide linkers can provide flexibility to the chimeric

polypeptide molecule. Linkers are not typically cleaved, however such cleavage can be desirable. In one embodiment, these linkers are not removed during processing.

[0131] A type of linker which can be present in a chimeric protein of the disclosure is a protease cleavable linker which comprises a cleavage site (*i.e.,* a protease cleavage site substrate, *e.g.,* a factor XIa, Xa, or thrombin cleavage site) and which can include additional linkers on either the N-terminal of C-terminal or both sides of the cleavage site. These cleavable linkers when incorporated into a construct of the disclosure result in a chimeric molecule having a heterologous cleavage site.

[0132] In one embodiment, an FVIII polypeptide encoded by a nucleic acid molecule of the instant disclosure comprises two or more Fc domains or moieties linked via a cscFc linker to form an Fc region comprised in a single polypeptide chain. The cscFc linker is flanked by at least one intracellular processing site, *i.e.,* a site cleaved by an intracellular enzyme. Cleavage of the polypeptide at the at least one intracellular processing site results in a polypeptide which comprises at least two polypeptide chains.

[0133] Other peptide linkers can optionally be used in a construct of the disclosure, *e.g.,* to connect an FVIII protein to an Fc region. Some exemplary linkers that can be used in connection with the disclosure include, *e.g.,* polypeptides comprising GlySer amino acids described in more detail below.

[0134] In one embodiment, the peptide linker is synthetic, *i.e.,* non-naturally occurring. In one embodiment, a peptide linker includes peptides (or polypeptides) (which can or cannot be naturally occurring) which comprise an amino acid sequence that links or genetically fuses a first linear sequence of amino acids to a second linear sequence of amino acids to which it is not naturally linked or genetically fused in nature. For example, in one embodiment the peptide linker can comprise non-naturally occurring polypeptides which are modified forms of naturally occurring polypeptides (*e.g.,* comprising a mutation such as an addition, substitution or deletion). In another embodiment, the peptide linker can comprise non-naturally occurring amino acids. In another embodiment, the peptide linker can comprise naturally occurring amino acids occurring in a linear sequence that does not occur in nature. In still another embodiment, the peptide linker can comprise a naturally occurring polypeptide sequence.

[0135] In another embodiment, a peptide linker comprises or consists of a gly-ser linker. As used herein, the term "gly-ser linker" refers to a peptide that consists of glycine and serine residues. In certain embodiments, said gly-ser linker can be inserted between two other sequences of the peptide linker. In other embodiments, a gly-ser linker is attached at one or both ends of another sequence of the peptide linker. In yet other embodiments, two or more gly-ser linker are incorporated in series in a peptide linker. In one embodiment, a peptide linker of the

disclosure comprises at least a portion of an upper hinge region (*e.g.,* derived from an IgG1, IgG2, IgG3, or IgG4 molecule), at least a portion of a middle hinge region (*e.g.,* derived from an IgG1, IgG2, IgG3, or IgG4 molecule) and a series of gly/ser amino acid residues.

[0136] Peptide linkers of the disclosure are at least one amino acid in length and can be of varying lengths. In one embodiment, a peptide linker of the disclosure is from about 1 to about 50 amino acids in length. As used in this context, the term "about" indicates +/- two amino acid residues. Since linker length must be a positive integer, the length of from about 1 to about 50 amino acids in length, means a length of from 1-3 to 48-52 amino acids in length. In another embodiment, a peptide linker of the disclosure is from about 10 to about 20 amino acids in length. In another embodiment, a peptide linker of the disclosure is from about 15 to about 50 amino acids in length. In another embodiment, a peptide linker of the disclosure is from about 20 to about 45 amino acids in length. In another embodiment, a peptide linker of the disclosure is from about 15 to about 35 or about 20 to about 30 amino acids in length. In another embodiment, a peptide linker of the disclosure is from about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 40, 50, 60, 70, 80, 90, 100, 500, 1000, or 2000 amino acids in length. In one embodiment, a peptide linker of the disclosure is 20 or 30 amino acids in length.

[0137] In some embodiments, the peptide linker can comprise at least two, at least three, at least four, at least five, at least 10, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, or at least 100 amino acids. In other embodiments, the peptide linker can comprise at least 200, at least 300, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, or at least 1,000 amino acids. In some embodiments, the peptide linker can comprise at least about 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, or 2000 amino acids. The peptide linker can comprise 1-5 amino acids, 1-10 amino acids, 1-20 amino acids, 10-50 amino acids, 50-100 amino acids, 100-200 amino acids, 200-300 amino acids, 300-400 amino acids, 400-500 amino acids, 500-600 amino acids, 600-700 amino acids, 700-800 amino acids, 800-900 amino acids, or 900-1000 amino acids.

[0138] Peptide linkers can be introduced into polypeptide sequences using techniques known in the art. Modifications can be confirmed by DNA sequence analysis. Plasmid DNA can be used to transform host cells for stable production of the polypeptides produced.

Expression Control Sequences

[0139] In some embodiments, the nucleic acid molecule or vector of the disclosure further comprises at least one expression control sequence. For example, the isolated nucleic acid molecule of the disclosure can be operably linked to at least one expression control sequence.

The expression control sequence can, for example, be a promoter sequence or promoter-enhancer combination.

[0140] Constitutive mammalian promoters include, but are not limited to, the promoters for the following genes: hypoxanthine phosphoribosyl transferase (HPRT), adenosine deaminase, pyruvate kinase, beta-actin promoter, and other constitutive promoters. Exemplary viral promoters which function constitutively in eukaryotic cells include, for example, promoters from the cytomegalovirus (CMV), simian virus (*e.g.*, SV40), papilloma virus, adenovirus, human immunodeficiency virus (HIV), Rous sarcoma virus, cytomegalovirus, the long terminal repeats (LTR) of Moloney leukemia virus, and other retroviruses, and the thymidine kinase promoter of herpes simplex virus. Other constitutive promoters are known to those of ordinary skill in the art. The promoters useful as gene expression sequences of the disclosure also include inducible promoters. Inducible promoters are expressed in the presence of an inducing agent. For example, the metallothionein promoter is induced to promote transcription and translation in the presence of certain metal ions. Other inducible promoters are known to those of ordinary skill in the art.

[0141] In one embodiment, the disclosure includes expression of a transgene under the control of a tissue specific promoter and/or enhancer. In another embodiment, the promoter or other expression control sequence selectively enhances expression of the transgene in liver cells. In certain embodiments, the promoter or other expression control sequence selectively enhances expression of the transgene in hepatocytes, sinusoidal cells, and/or endothelial cells. In one particular embodiment, the promoter or other expression control sequence selective enhances expression of the transgene in endothelial cells. In certain embodiments, the promoter or other expression control sequence selective enhances expression of the transgene in muscle cells, the central nervous system, the eye, the liver, the heart, or any combination thereof. Examples of liver specific promoters include, but are not limited to, a mouse transthyretin promoter (mTTR), a native human factor VIII promoter, human alpha-1-antitrypsin promoter (hAAT), human albumin minimal promoter, and mouse albumin promoter. In some embodiments, the nucleic acid molecules disclosed herein comprise a mTTR promoter. The mTTR promoter is described in Costa et al. (1986) Mol. Cell. Biol. 6:4697. The FVIII promoter is described in Figueiredo and Brownlee, 1995, J. Biol. Chem. 270:11828-11838. In some embodiments, the promoter is selected from a liver specific promoter (*e.g.*, α1-antitrypsin (AAT)), a muscle specific promoter (*e.g.*, muscle creatine kinase (MCK), myosin heavy chain alpha (αMHC), myoglobin (MB), and desmin (DES)), a synthetic promoter (*e.g.*, SPc5-12, 2R5Sc5-12, dMCK, and tMCK), or any combination thereof.

[0142] In some embodiments, the transgene expression is targeted to the liver. In certain embodiments, the transgene expression is targeted to hepatocytes. In other embodiment, the transgene expression is targeted to endothelial cells. In one particular embodiment, the transgene

expression is targeted to any tissue that naturally expressed endogenous FVIII. In some embodiments, the transgene expression is targeted to the central nervous system. In certain embodiments, the transgene expression is targeted to neurons. In some embodiments, the transgene expression is targeted to afferent neurons. In some embodiments, the transgene expression is targeted to efferent neurons. In some embodiments, the transgene expression is targeted to interneurons. In some embodiments, the transgene expression is targeted to glial cells. In some embodiments, the transgene expression is targeted to astrocytes. In some embodiments, the transgene expression is targeted to oligodendrocytes. In some embodiments, the transgene expression is targeted to microglia. In some embodiments, the transgene expression is targeted to ependymal cells. In some embodiments, the transgene expression is targeted to Schwann cells. In some embodiments, the transgene expression is targeted to satellite cells. In some embodiments, the transgene expression is targeted to muscle tissue. In some embodiments, the transgene expression is targeted to smooth muscle. In some embodiments, the transgene expression is targeted to cardiac muscle. In some embodiments, the transgene expression is targeted to skeletal muscle. In some embodiments, the transgene expression is targeted to the eye. In some embodiments, the transgene expression is targeted to a photoreceptor cell. In some embodiments, the transgene expression is targeted to retinal ganglion cell.

[0143] Other promoters useful in the nucleic acid molecules disclosed herein include a mouse transthyretin promoter (mTTR), a native human factor VIII promoter, a human alpha-1-antitrypsin promoter (hAAT), a human albumin minimal promoter, a mouse albumin promoter, a tristetraprolin (TTP; also known as ZFP36) promoter, a CASI promoter, a CAG promoter, a cytomegalovirus (CMV) promoter, an α1-antitrypsin (AAT) promoter, a muscle creatine kinase (MCK) promoter, myosin heavy chain alpha (αMHC) promoter, a myoglobin (MB) promoter, desmin (DES) promoter, a SPc5-12 promoter, a 2R5Sc5-12 promoter, a dMCK promoter, and a tMCK promoter, a phosphoglycerate kinase (PGK) promoter, or any combinations thereof.

[0144] In some embodiments, the nucleic acid molecules disclosed herein comprise a transthyretin (TTR) promoter. In some embodiments, the promoter is a mouse transthyretin (mTTR) promoter. Non-limiting examples of mTTR promoters include the mTTR202 promoter, mTTR202opt promoter, and mTTR482 promoter, as disclosed in U.S. Publication No.. US2019/0048362, which is incorporated by reference herein in its entirety. In some embodiments, the promoter is a liver-specific modified mouse transthyretin (mTTR) promoter. In some embodiments, the promoter is the liver-specific modified mouse transthyretin (mTTR) promoter mTTR482. Examples of mTTR482 promoters are described in Kyostio-Moore et al. (2016) Mol Ther Methods Clin Dev. 3:16006, and Nambiar B. et al. (2017) Hum Gene Ther Methods, 28(1):23-

28. In some embodiments, the promoter is a liver-specific modified mouse transthyretin (mTTR) promoter comprising the nucleic acid sequence of SEQ ID NO: 16.

[0145] Expression levels can be further enhanced to achieve therapeutic efficacy using one or more enhancer elements. One or more enhancers can be provided either alone or together with one or more promoter elements. Typically, the expression control sequence comprises a plurality of enhancer elements and a tissue specific promoter. In one embodiment, an enhancer comprises one or more copies of the α-1-microglobulin/bikunin enhancer (Rouet et al. (1992) J. Biol. Chem. 267:20765-20773; Rouet et al. (1995), Nucleic Acids Res. 23:395-404; Rouet et al (1998) Biochem. J. 334:577-584; Ill et al. (1997) Blood Coagulation Fibrinolysis 8:S23-S30). In some embodiments, the enhancer is derived from liver specific transcription factor binding sites, such as EBP, DBP, HNF1, HNF3, HNF4, HNF6, with Enh1, comprising HNF1, (sense)-HNF3, (sense)-HNF4, (antisense)-HNF1, (antisense)-HNF6, (sense)-EBP, (antisense)-HNF4 (antisense).

[0146] In some embodiments, the enhancer element comprises one or two modified prothrombin enhancers (pPrT2), one or two alpha 1-microbikunin enhancers (A1MB2), a modified mouse albumin enhancer (mEalb), a hepatitis B virus enhancer II (HE11), or a CRM8 enhancer. In some embodiments, the A1MB2 enhancer is the enhancer disclosed in International Application No. PCT/US2019/055917. In some embodiments, the enhancer element is A1MB2. In some embodiments, the enhancer element includes multiple copies of the AIMB2 enhancer sequence. In some embodiments, the A1MB2 enhancer is positioned 5' to the nucleic acid sequence encoding the FVIII polypeptide. In some embodiments, the A1MB2 enhancer is positioned 5’ to the promoter sequence, such as the mTTR promoter. In some embodiments, the enhancer element is the A1MB2 enhancer comprising the nucleic acid sequence of SEQ ID NO: 15.

[0147] In some embodiments, the nucleic acid molecules disclosed herein comprise an intron or intronic sequence. In some embodiments, the intronic sequence is a naturally occurring intronic sequence. In some embodiments, the intronic sequence is a synthetic sequence. In some embodiments, the intronic sequence is derived from a naturally occurring intronic sequence. In some embodiments, the intronic sequence is a hybrid synthetic intron or chimeric intron. In some embodiments, the intronic sequence is a chimeric intron that consists of chicken beta-actin/rabbit beta-globin intron and has been modified to eliminate five existing ATG sequences to reduce false translation starts. In certain embodiments, the intronic sequence comprises the SV40 small T intron. In some embodiments, the intronic sequence is positioned 5' to the nucleic acid sequence encoding the FVIII polypeptide. In some embodiments, the chimeric intron is positioned 5’ to a promoter sequence, such as the mTTR promoter. In some embodiments, the chimeric intron comprises the nucleic acid sequence of SEQ ID NO: 17.

[0148] In some embodiments, the nucleic acid molecules disclosed herein comprise a post-transcriptional regulatory element. In certain embodiments, the regulatory element comprises a mutated woodchuck hepatitis virus regulatory element (WPRE). WPRE is believed to enhance the expression of viral vector-delivered transgenes. Examples of WPRE are described in Zufferey et al. (1999) J Virol., 73(4):2886-2892; Loeb et al. (1999) Hum Gene Ther. 10(14):2295-2305. In some embodiments, the WPRE is positioned 3' to the nucleic acid sequence encoding the FVIII polypeptide. In some embodiments, the WPRE comprises the nucleic acid sequence of SEQ ID NO: 18.

[0149] In some embodiments, the nucleic acid molecules disclosed herein comprise a transcription terminator. In some embodiments, the transcription terminator is a polyadenylation (poly(A)) sequence. Non-limiting examples of transcriptional terminators include those derived from the bovine growth hormone polyadenylation signal (BGHpA), the Simian virus 40 polyadenylation signal (SV40pA), or a synthetic polyadenylation signal. In one embodiment, the 3'UTR poly(A) tail comprises an actin poly(A) site. In one embodiment, the 3'UTR poly(A) tail comprises a hemoglobin poly(A) site. In some embodiments, the transcriptional terminator is BGHpA. Examples of BGHpA transcriptional terminators are described in Woychik et al.(1984) PNAS 81:3944-3948. In some embodiments, the transcriptionalo terminator is positioned at the 3' end of the genetic cassette encoding the nucleic acid sequence encoding the FVIII polypeptide. In some embodiments, the transcriptional terminator is a BGHpA comprising the nucleic acid sequence of SEQ ID NO: 19.

[0150] In some embodiments, the nucleic acid molecule disclosed herein comprises one or more DNA nuclear targeting sequences (DTSs). A DTS promotes translocation of DNA molecules containing such sequences into the nucleus. In certain embodiments, the DTS comprises an SV40 enhancer sequence. In certain embodiments, the DTS comprises a c-Myc enhancer sequence. In some embodiments, the nucleic acid molecule comprises DTSs that are located between the first ITR and the second ITR. In some embodiments, the nucleic acid molecule comprises a DTS located 3' to the first ITR and 5' to the transgene (e.g. FVIII protein). In some embodiments, the nucleic acid molecule comprises a DTS located 3' to the transgene and 5' to the second ITR on the nucleic acid molecule.

[0151] In some embodiments, the nucleic acid molecule disclosed herein comprises a toll-like receptor 9 (TLR9) inhibition sequence. Exemplary TLR9 inhibition sequences are described in, e.g., Trieu et al. (2006) Crit Rev Immunol. 26(6):527-44; Ashman et al. Int'l Immunology 23(3): 203-14.

Inverted Terminal Repeat (ITR) Sequences

[0152] Certain aspects of the present disclosure are directed to a nucleic acid molecule comprising a first ITR, *e.g.*, a 5' ITR, and second ITR, *e.g.*, a 3' ITR. Typically, ITRs are involved in parvovirus (*e.g.*, AAV) DNA replication and rescue, or excision, from prokaryotic plasmids (Samulski et al., 1983, 1987; Senapathy et al., 1984; Gottlieb and Muzyczka, 1988). In addition, ITRs appear to be the minimum sequences required for AAV proviral integration and for packaging of AAV DNA into virions (McLaughlin et al., 1988; Samulski et al., 1989). These elements are essential for efficient multiplication of a parvovirus genome. It is hypothesized that the minimal defining elements indispensable for ITR function are a Rep-binding site and a terminal resolution site plus a variable palindromic sequence allowing for hairpin formation. Palindromic nucleotide regions normally function together in *cis* as origins of DNA replication and as packaging signals for the virus. Complimentary sequences in the ITRs fold into a hairpin structure during DNA replication. In some embodiments, the ITRs fold into a hairpin T-shaped structure. In other embodiments, the ITRs fold into non-T-shaped hairpin structures, *e.g.*, into a U-shaped hairpin structure. Data suggests that the T-shaped hairpin structures of AAV ITRs may inhibit the expression of a transgene flanked by the ITRs. *See, e.g.*, Zhou et al. (2017) *Scientific Reports 7*:5432. By utilizing an ITR that does not form T-shaped hairpin structures, this form of inhibition may be avoided. Therefore, in certain aspects, a polynucleotide comprising a non-AAV ITR has an improved transgene expression compared to a polynucleotide comprising an AAV ITR that forms a T-shaped hairpin.

[0153] As used herein, an "inverted terminal repeat" (or "ITR") refers to a nucleic acid subsequence located at either the 5' or 3' end of a single stranded nucleic acid sequence, which comprises a set of nucleotides (initial sequence) followed downstream by its reverse complement, *i.e.,* palindromic sequence. The intervening sequence of nucleotides between the initial sequence and the reverse complement can be any length including zero. In one embodiment, the ITR useful for the present disclosure comprises one or more "palindromic sequences." An ITR can have any number of functions. In some embodiments, an ITR described herein forms a hairpin structure. In some embodiments, the ITR forms a T-shaped hairpin structure. In some embodiments, the ITR forms a non-T-shaped hairpin structure, *e.g.*, a U-shaped hairpin structure. In some embodiments, the ITR promotes the long-term survival of the nucleic acid molecule in the nucleus of a cell. In some embodiments, the ITR promotes the permanent survival of the nucleic acid molecule in the nucleus of a cell (*e.g.*, for the entire life-span of the cell). In some embodiments, the ITR promotes the stability of the nucleic acid molecule in the nucleus of a cell. In some embodiments, the ITR promotes the retention of the nucleic acid molecule in the nucleus of a cell. In some embodiments, the ITR promotes the persistence of the nucleic acid molecule in the nucleus of a cell. In some

embodiments, the ITR inhibits or prevents the degradation of the nucleic acid molecule in the nucleus of a cell.

[0154] Therefore, an "ITR" as used herein can fold back on itself and form a double stranded segment. For example, the sequence GATCXXXXGATC comprises an initial sequence of GATC and its complement (3'CTAG5') when folded to form a double helix. In some embodiments, the ITR comprises a continuous palindromic sequence (*e.g.*, GATCGATC) between the initial sequence and the reverse complement. In some embodiments, the ITR comprises an interrupted palindromic sequence (*e.g.*, GATCXXXXGATC) between the initial sequence and the reverse complement. In some embodiments, the complementary sections of the continuous or interrupted palindromic sequence interact with each other to form a "hairpin loop" structure. As used herein, a "hairpin loop" structure results when at least two complimentary sequences on a single-stranded nucleotide molecule base-pair to form a double stranded section. In some embodiments, only a portion of the ITR forms a hairpin loop. In other embodiments, the entire ITR forms a hairpin loop.

[0155] In the present disclosure, at least one ITR is an ITR of a non-adenovirus associated virus (non-AAV). In certain embodiments, the ITR is an ITR of a non-AAV member of the viral family *Parvoviridae*. In some embodiments, the ITR is an ITR of a non-AAV member of the genus *Dependovirus* or the genus *Erythrovirus*.

[0156] In some embodiments, the ITR is an ITR of a non-AAV genome from Bocavirus, Dependovirus, Erythrovirus, Amdovirus, Parvovirus, Densovirus, Iteravirus, Contravirus, Aveparvovirus, Copiparvovirus, Protoparvovirus, Tetraparvovirus, Ambidensovirus, Brevidensovirus, Hepandensovirus, Penstyldensovirus and any combination thereof. In certain embodiments, the ITR is derived from human Bocavirus (HBoV1). In certain embodiments, the ITR is derived from erythrovirus parvovirus B19 (human virus). In some embodiments, the ITR is derived from a Dependoparvovirus. In one embodiment, the Dependoparvovirus is a Dependovirus Goose parvovirus (GPV) strain. In a specific embodiment, the GPV strain is attenuated, e.g., GPV strain 82-0321V. In another specific embodiment, the GPV strain is pathogenic, e.g., GPV strain B. In some embodiments, the ITR is an ITR of a goose parvovirus (GPV) or a Muscovy duck parvovirus (MDPV).

[0157] In some embodiments, the ITR is an ITR of an erythrovirus parvovirus B19 (also known as parvovirus B19 – also referred to herein as “B19”, primate erythroparvovirus 1, B19 virus, and erythrovirus). In some embodiments, the ITR is an ITR of a human Bocavirus (HBoV1).

[0158] In certain embodiments, one ITR of two ITRs is an ITR of an AAV. In other embodiments, one ITR of two ITRs in the construct is an ITR of an AAV serotype selected from serotype 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and any combination thereof. In one particular embodiment, the ITR is derived from AAV serotype 2, *e.g.,* an ITR of AAV serotype 2.

[0159] In certain aspects of the present disclosure, the nucleic acid molecule comprises two ITRs, a 5' ITR and a 3' ITR, wherein the 5' ITR is located at the 5' terminus of the nucleic acid molecule, and the 3' ITR is located at the 3' terminus of the nucleic acid molecule. The first ITR and the second ITR of the nucleic acid molecule can be derived from the same genome, *e.g.*, from the genome of the same virus, or from different genomes, *e.g.*, from the genomes of two or more different virus genomes (also known as "hybrid" ITRs). In some embodiments, first ITR is derived from B19 and the second ITR is derived from GPV. In some embodiments, first ITR is derived from GPV and the second ITR is derived from B19.

[0160] In certain embodiments, the first ITR and/or the second ITR comprises or consists of all or a portion of an ITR derived from human Bocavirus (HBoV1). In certain embodiments, the first ITR and/or the second ITR comprises or consists of all or a portion of an ITR derived from HBoV1. In some embodiments, the second ITR is a reverse complement of the first ITR. In some embodiments, the first ITR is a reverse complement of the second ITR. In some embodiments, the first ITR and/or the second ITR derived from HBoV1 is capable of forming a hairpin structure. In certain embodiments, the hairpin structure does not comprise a T-shaped hairpin.

[0161] In some embodiments, the first ITR and/or the second ITR comprises or consists of a nucleotide sequence at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a nucleotide sequence set forth in SEQ ID NOs: SEQ ID NOs: 1, 2, 21-30, wherein the first ITR and/or the second ITR retains a functional property of the wild type ITR from which it is derived. In some embodiments, the first ITR and/or the second ITR is derived from a wild type HBoV1 ITR. In some embodiments, the first ITR and/or the second ITR is derived from a wild type B19 ITR. In some embodiments, the first ITR and/or the second ITR is derived from a wild type GPV ITR.

[0162] In some embodiments, the first ITR and/or the second ITR comprises or consists of a nucleotide sequence at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a nucleotide sequence set forth in SEQ ID NOs: 1, 2, 21-30, wherein the first ITR and/or the second ITR is capable of forming a hairpin structure. In certain embodiments, the hairpin structure does not comprise a T-shaped hairpin.

[0163] It will be appreciated to those of skill in the art that any of the first ITR sequences described herein can be matched with any of the second ITR sequences described herein. In some embodiments, the first ITR sequence described herein is a 5' ITR sequence. In some embodiments, the second ITR sequence described herein is a 3' ITR sequence. In some embodiments, the second ITR sequence described herein is a 5' ITR sequence. In some

embodiments, the first ITR sequence described herein is a 3' ITR sequence. Those of skill in the art will be able to determine the suitable orientation of the first and the second ITR described herein with respect to the architecture of a genetic cassette.

[0164] In another particular embodiment, the ITR is a synthetic sequence genetically engineered to include at its 5′ and 3′ ends ITRs not derived from an AAV genome. In another particular embodiment, the ITR is a synthetic sequence genetically engineered to include at its 5′ and 3′ ends ITRs derived from one or more of non-AAV genomes. The two ITRs present in the nucleic acid molecule of the invention can be the same or different non-AAV genomes. In particular, the ITRs can be derived from the same non-AAV genome. In a specific embodiment, the two ITRs present in the nucleic acid molecule of the invention are the same, and can in particular be AAV2 ITRs.

[0165] In some embodiments, the ITR sequence comprises one or more palindromic sequence. A palindromic sequence of an ITR disclosed herein includes, but is not limited to, native palindromic sequences (i.e., sequences found in nature), synthetic sequences (i.e., sequences not found in nature), such as pseudo palindromic sequences, and combinations or modified forms thereof.

[0166] In some embodiments, the ITRs form hairpin loop structures. In one embodiment, the first ITR forms a hairpin structure. In another embodiment, the second ITR forms a hairpin structure. Still in another embodiment, both the first ITR and the second ITR form hairpin structures. In some embodiments, the first ITR and/or the second ITR does not form a T-shaped hairpin structure. In certain embodiments, the first ITR and/or the second ITR forms a non-T-shaped hairpin structure. In some embodiments, the non-T-shaped hairpin structure comprises a U-shaped hairpin structure.

[0167] In some embodiments, an ITR in a nucleic acid molecule described herein may be a transcriptionally activated ITR. A transcriptionally-activated ITR can comprise all or a portion of a wild-type ITR that has been transcriptionally activated by inclusion of at least one transcriptionally active element. Various types of transcriptionally active elements are suitable for use in this context. In some embodiments, the transcriptionally active element is a constitutive transcriptionally active element. Constitutive transcriptionally active elements provide an ongoing level of gene transcription, and are preferred when it is desired that the transgene be expressed on an ongoing basis. In other embodiments, the transcriptionally active element is an inducible transcriptionally active element. Inducible transcriptionally active elements generally exhibit low activity in the absence of an inducer (or inducing condition), and are up-regulated in the presence of the inducer (or switch to an inducing condition). Inducible transcriptionally active elements may be preferred when expression is desired only at certain times or at certain locations, or when it is

desirable to titrate the level of expression using an inducing agent. Transcriptionally active elements can also be tissue-specific; that is, they exhibit activity only in certain tissues or cell types.

[0168] Transcriptionally active elements, can be incorporated into an ITR in a variety of ways. In some embodiments, a transcriptionally active element is incorporated 5′ to any portion of an ITR or 3′ to any portion of an ITR. In other embodiments, a transcriptionally active element of a transcriptionally-activated ITR lies between two ITR sequences. If the transcriptionally active element comprises two or more elements which must be spaced apart, those elements may alternate with portions of the ITR. In some embodiments, a hairpin structure of an ITR is deleted and replaced with inverted repeats of a transcriptional element. This latter arrangement would create a hairpin mimicking the deleted portion in structure. Multiple tandem transcriptionally active elements can also be present in a transcriptionally-activated ITR, and these may be adjacent or spaced apart. In addition, protein binding sites (*e.g.*, Rep binding sites) can be introduced into transcriptionally active elements of the transcriptionally-activated ITRs. A transcriptionally active element can comprise any sequence enabling the controlled transcription of DNA by RNA polymerase to form RNA, and can comprise, for example, a transcriptionally active element, as defined below.

[0169] Transcriptionally-activated ITRs provide both transcriptional activation and ITR functions to the nucleic acid molecule in a relatively limited nucleotide sequence length which effectively maximizes the length of a transgene which can be carried and expressed from the nucleic acid molecule. Incorporation of a transcriptionally active element into an ITR can be accomplished in a variety of ways. A comparison of the ITR sequence and the sequence requirements of the transcriptionally active element can provide insight into ways to encode the element within an ITR. For example, transcriptional activity can be added to an ITR through the introduction of specific changes in the ITR sequence that replicates the functional elements of the transcriptionally active element. A number of techniques exist in the art to efficiently add, delete, and/or change particular nucleotide sequences at specific sites (see, for example, Deng and Nickoloff (1992) Anal. Biochem. 200:81-88). Another way to create transcriptionally-activated ITRs involves the introduction of a restriction site at a desired location in the ITR. In addition, multiple transcriptionally activate elements can be incorporated into a transcriptionally-activated ITR, using methods known in the art.

[0170] By way of illustration, transcriptionally-activated ITRs can be generated by inclusion of one or more transcriptionally active elements such as: TATA box, GC box, CCAAT box, Sp1 site, Inr region, CRE (cAMP regulatory element) site, ATF-1/CRE site, APBβ box, APBα box, CArG box, CCAC box, or any other element involved in transcription as known in the art.

Vector Systems

[0171] Some embodiments of the present disclosure are directed to vectors comprising one or more codon optimized nucleic acid molecules encoding a polypeptide with FVIII activity described herein, host cells comprising the vectors, and methods of treating a bleeding disorder using the vectors. The present disclosure meets an important need in the art by providing a vector comprising an optimized FVIII sequence that demonstrates increased expression in a subject and potentially result in greater therapeutic efficacy when used in gene therapy methods.

[0172] Suitable vectors for the disclosure include expression vectors, viral vectors, and plasmid vectors. In one embodiment, the vector is a viral vector.

[0173] As used herein, an expression vector refers to any nucleic acid construct which contains the necessary elements for the transcription and translation of an inserted coding sequence, or in the case of an RNA viral vector, the necessary elements for replication and translation, when introduced into an appropriate host cell. Expression vectors can include plasmids, phagemids, viruses, and derivatives thereof.

[0174] Expression vectors of the disclosure will include optimized polynucleotides encoding the BDD FVIII protein described herein. In one embodiment, the optimized coding sequences for the BDD FVIII protein is operably linked to an expression control sequence. As used herein, two nucleic acid sequences are operably linked when they are covalently linked in such a way as to permit each component nucleic acid sequence to retain its functionality. A coding sequence and a gene expression control sequence are said to be operably linked when they are covalently linked in such a way as to place the expression or transcription and/or translation of the coding sequence under the influence or control of the gene expression control sequence. Two DNA sequences are said to be operably linked if induction of a promoter in the 5' gene expression sequence results in the transcription of the coding sequence and if the nature of the linkage between the two DNA sequences does not (1) result in the introduction of a frame-shift mutation, (2) interfere with the ability of the promoter region to direct the transcription of the coding sequence, or (3) interfere with the ability of the corresponding RNA transcript to be translated into a protein. Thus, a gene expression sequence would be operably linked to a coding nucleic acid sequence if the gene expression sequence were capable of effecting transcription of that coding nucleic acid sequence such that the resulting transcript is translated into the desired protein or polypeptide.

[0175] Viral vectors include, but are not limited to, nucleic acid sequences from the following viruses: retrovirus, such as Moloney murine leukemia virus, Harvey murine sarcoma virus, murine mammary tumor virus, and Rous sarcoma virus; lentivirus; adenovirus; adeno-associated virus; SV40-type viruses; polyomaviruses; Epstein-Barr viruses; papilloma viruses; herpes virus;

vaccinia virus; polio virus; and RNA virus such as a retrovirus. One can readily employ other vectors well-known in the art. Certain viral vectors are based on non-cytopathic eukaryotic viruses in which non-essential genes have been replaced with the gene of interest. In one embodiment, the virus is an adeno-associated virus, a double-stranded DNA virus. The adeno-associated virus can be engineered to be replication-deficient and is capable of infecting a wide range of cell types and species.

[0176] One or more of different AAV vector sequences derived from nearly any serotype can be used in accord with the present disclosure. Choice of a particular AAV vector sequence will be guided by known parameters such as tropism of interest, required vector yields, etc. Generally, the AAV serotypes have genomic sequences of significant homology at the amino acid and the nucleic acid levels, provide a related set of genetic functions, produce virions which are related, and replicate and assemble similarly. For the genomic sequence of the various AAV serotypes and an overview of the genomic similarities see, *e.g.*, GenBank Accession number U89790; GenBank Accession number J01901; GenBank Accession number AF043303; GenBank Accession number AF085716; Chlorini et al. (1997) J. Vir. 71: 6823-33; Srivastava et al. (1983) J. Vir. 45:555-64; Chlorini et al. (1999) J. Vir. 73:1309-1319; Rutledge et al. (1998), J. Vir. 72:309-319; or Wu et al. (2000) J. Vir. 74: 8635-47. AAV serotypes 1, 2, 3, 4 and 5 are an illustrative source of AAV nucleotide sequences for use in the context of the present disclosure. AAV6, AAV7, AAV8 or AAV9 or newly developed AAV-like particles obtained by e.g. capsid shuffling techniques and AAV capsid libraries, or from newly designed, developed or evolved ITR's are also suitable for certain disclosure applications. *See* Dalkara et al. (2013), Sci. Transl. Med. 5(189): 189ra76; Kotterman MA (2014) Nat. Rev. Genet. 15(7):455.

[0177] Other vectors include plasmid vectors. Plasmid vectors have been extensively described in the art and are well-known to those of skill in the art. *See*, *e.g.*, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press, 1989. In the last few years, plasmid vectors have been found to be particularly advantageous for delivering genes to cells *in vivo* because of their inability to replicate within and integrate into a host genome. These plasmids, however, having a promoter compatible with the host cell, can express a peptide from a gene operably encoded within the plasmid. Some commonly used plasmids available from commercial suppliers include pBR322, pUC18, pUC19, various pcDNA plasmids, pRC/CMV, various pCMV plasmids, pSV40, and pBlueScript. Additional examples of specific plasmids include pcDNA3.1, catalog number V79020; pcDNA3.1/hygro, catalog number V87020; pcDNA4/myc-His, catalog number V86320; and pBudCE4.1, catalog number V53220, all from Invitrogen (Carlsbad, CA.). Other plasmids are well-known to those of ordinary skill in the

art. Additionally, plasmids can be custom designed using standard molecular biology techniques to remove and/or add specific fragments of DNA.

[0178] In certain embodiments, it will be useful to include within the vector one or more miRNA target sequences which, for example, are operably linked to the optimized FVIII transgene. More than one copy of a miRNA target sequence included in the vector can increase the effectiveness of the system. For example, vectors which express more than one transgene can have the transgene under control of more than one miRNA target sequence, which can be the same or different. The miRNA target sequences can be in tandem, but other arrangements are also included. The transgene expression cassette, containing miRNA target sequences, can also be inserted within the vector in antisense orientation. Examples of the miRNA target sequences are described at WO2007/000668, WO2004/094642, WO2010/055413, or WO2010/125471, which are incorporated herein by reference in their entireties. However in certain other embodiments, the vector will not include any miRNA target sequence. Choice of whether or not to include an miRNA target sequence (and how many) will be guided by known parameters such as the intended tissue target, the level of expression required, etc.

Host Cells

[0179] The disclosure also provides a host cell comprising a nucleic acid molecule or vector of the disclosure. As used herein, the term "transformation" shall be used in a broad sense to refer to the introduction of DNA into a recipient host cell that changes the genotype and consequently results in a change in the recipient cell.

[0180] "Host cells" refers to cells that have been transformed with vectors constructed using recombinant DNA techniques and encoding at least one heterologous gene. The host cells of the present disclosure are preferably of mammalian origin; most preferably of human or mouse origin. Those skilled in the art are credited with ability to preferentially determine particular host cell lines which are best suited for their purpose. Exemplary host cell lines include, but are not limited to, CHO, DG44 and DUXB11 (Chinese Hamster Ovary lines, DHFR minus), HELA (human cervical carcinoma), CVI (monkey kidney line), COS (a derivative of CVI with SV40 T antigen), R1610 (Chinese hamster fibroblast) BALBC/3T3 (mouse fibroblast), HAK (hamster kidney line), SP2/O (mouse myeloma), P3.times.63-Ag3.653 (mouse myeloma), BFA-1c1BPT (bovine endothelial cells), RAJI (human lymphocyte), PER.C6®, NS0, CAP, BHK21, and HEK 293 (human kidney). In one particular embodiment, the host cell is selected from the group consisting of: a CHO cell, a HEK293 cell, a BHK21 cell, a PER.C6® cell, a NS0 cell, and a CAP cell. Host cell lines are typically available from commercial services, the American Tissue Culture Collection, or from published literature.

[0181] Introduction of the isolated nucleic acid molecules or vectors of the disclosure into the host cell can be accomplished by various techniques well known to those of skill in the art. These include, but are not limited to, transfection (including electrophoresis and electroporation), protoplast fusion, calcium phosphate precipitation, cell fusion with enveloped DNA, microinjection, and infection with intact virus. See, Ridgway, A. A. G. "*Mammalian Expression Vectors*" Chapter 24.2, pp. 470-472 Vectors, Rodriguez and Denhardt, Eds. (Butterworths, Boston, Mass. 1988). Plasmids can be introduced into the host via electroporation. The transformed cells are grown under conditions appropriate to the production of the light chains and heavy chains, and assayed for heavy and/or light chain protein synthesis. Exemplary assay techniques include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), or flourescence-activated cell sorter analysis (FACS), immunohistochemistry and the like.

[0182] Host cells comprising the isolated nucleic acid molecules or vectors of the disclosure are grown in an appropriate growth medium. As used herein, the term "appropriate growth medium" means a medium containing nutrients required for the growth of cells. Nutrients required for cell growth can include a carbon source, a nitrogen source, essential amino acids, vitamins, minerals, and growth factors. Optionally, the media can contain one or more selection factors. Optionally the media can contain bovine calf serum or fetal calf serum (FCS). In one embodiment, the media contains substantially no IgG. The growth medium will generally select for cells containing the DNA construct by, for example, drug selection or deficiency in an essential nutrient which is complemented by the selectable marker on the DNA construct or co-transfected with the DNA construct. Cultured mammalian cells are generally grown in commercially available serum-containing or serum-free media (*e.g.*, MEM, DMEM, DMEM/F12). In one embodiment, the medium is CDoptiCHO (Invitrogen, Carlsbad, CA.). In another embodiment, the medium is CD17 (Invitrogen, Carlsbad, CA.). Selection of a medium appropriate for the particular cell line used is within the level of those ordinary skilled in the art.

[0183] In some embodiments, host cells suitable for use in the present invention are of insect origin. In some embodiments, a suitable insect host cell includes, for example, a cell line isolated from *Spodoptera frugiperda* (Sf) or a cell line isolated from Trichoplusia ni (Tni). Those of skill in the art will readily be able to determine the suitability of any Sf or Tni cell line. Exemplary insect host cells include, without limitation, Sf9 cells, Sf21 cells, and High Five™ cells. Exemplary insect host cells also include, without limitation, any Sf or Tni cell line that is free from adventitious virus contamination, e.g., Sf-rhabdovirus-negative (Sf-RVN) and Tn-nodavirus-negative (Tn-NVN) cells. Other suitable host insect cells are known to those of skill in the art. In one particular embodiment, the insect host cells are Sf9 cells.

[0184] Aspects of the present disclosure provide a method of cloning a nucleic acid molecule described herein, comprising inserting a nucleic acid molecule capable of complex secondary structures into a suitable vector, and introducing the resulting vector into a suitable bacterial host strain. As known in the art, complex secondary structures (e.g., long palindromic regions) of nucleic acids may be unstable and difficult to clone in bacterial host strains. For example, nucleic acid molecules comprising a first ITR and a second ITR (e.g., non-AAV parvoviral ITRs, e.g., HBoV1 ITRs) of the present disclosure may be difficult to clone using conventional methodologies. Long DNA palindromes inhibit DNA replication and are unstable in the genomes of *E. coli, Bacillus, Streptococcus, Streptomyces, S. cerevisiae,* mice, and humans. These effects result from the formation of hairpin or cruciform structures by intrastrand base pairing. In *E. coli* the inhibition of DNA replication can be significantly overcome in SbcC or SbcD mutants. SbcD is the nuclease subunit, and SbcC is the ATPase subunit of the SbcCD complex. The *E. coli* SbcCD complex is an exonuclease complex responsible for preventing the replication of long palindromes. The SbcCD complex is a nuclear with ATP-dependent double-stranded DNA exonuclease activity and ATP-independent single-stranded DNA endonuclease activity. SbcCD may recognize DNA palindromes and collapse replication forks by attacking hairpin structures that arise.

[0185] In certain embodiments, a suitable bacterial host strain is incapable of resolving cruciform DNA structures. In certain embodiments, a suitable bacterial host strain comprises a disruption in the SbcCD complex. In some embodiments, the disruption in the SbcCD complex comprises a genetic disruption in the SbcC gene and/or SbcD gene. In certain embodiments, the disruption in the SbcCD complex comprises a genetic disruption in the SbcC gene. Various bacterial host strains that comprise a genetic disruption in the SbcC gene are known in the art. For example, without limitation, the bacterial host strain PMC103 comprises the genotype *sbcC, recD, mcrA, ΔmcrBCF*; the bacterial host strain PMC107 comprises the genotype *recBC, recJ, sbcBC, mcrA, ΔmcrBCF*; and the bacterial host strain SURE comprises the genotype *recB, recJ, sbcC, mcrA, ΔmcrBCF, umuC, uvrC*. Accordingly, in some embodiments a method of cloning a nucleic acid molecule described herein comprises inserting a nucleic acid molecule capable of complex secondary structures into a suitable vector, and introducing the resulting vector into host strain PMC103, PMC107, or SURE. In certain embodiments, the method of cloning a nucleic acid molecule described herein comprises inserting a nucleic acid molecule capable of complex secondary structures into a suitable vector, and introducing the resulting vector into host strain PMC103.

[0186] Suitable vectors are known in the art and described elsewhere herein. In certain embodiments, a suitable vector for use in a cloning methodology of the present disclosure is a

low copy vector. In certain embodiments, a suitable vector for use in a cloning methodology of the present disclosure is pBR322.

[0187] Accordingly, the present disclosure provides a method of cloning a nucleic acid molecule, comprising inserting a nucleic acid molecule capable of complex secondary structures into a suitable vector, and introducing the resulting vector into a bacterial host strain comprising a disruption in the SbcCD complex, wherein the nucleic acid molecule comprises a first inverted terminal repeat (ITR) and a second ITR, wherein the first ITR and/or second ITR comprises a nucleotide sequence at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% identical to a nucleotide sequence set forth in SEQ ID NOs. 12-23 or a functional derivative thereof.

Production of Polypeptides

[0188] The disclosure also provides a polypeptide encoded by a nucleic acid molecule of the disclosure. In other embodiments, the polypeptide of the disclosure is encoded by a vector comprising the isolated nucleic molecules of the disclosure. In yet other embodiments, the polypeptide of the disclosure is produced by a host cell comprising the isolated nucleic molecules of the disclosure.

[0189] In other embodiments, the disclosure also provides a method of producing a polypeptide with FVIII activity, comprising culturing a host cell of the disclosure under conditions whereby a polypeptide with FVIII activity is produced, and recovering the polypeptide with FVIII activity. In some embodiments, the expression of the polypeptide with FVIII activity is increased relative to a host cell cultured under the same conditions but comprising a reference nucleotide sequence comprising SEQ ID NO: 32, a parental FVIII nucleotide sequence.

[0190] In other embodiments, the disclosure provides a method of increasing the expression of a polypeptide with FVIII activity comprising culturing a host cell of the disclosure under conditions whereby a polypeptide with FVIII activity is expressed by the nucleic acid molecule, wherein the expression of the polypeptide with FVIII activity is increased relative to a host cell cultured under the same conditions comprising a reference nucleic acid molecule comprising SEQ ID NO: 32.

[0191] In other embodiments, the disclosure provides a method of improving yield of a polypeptide with FVIII activity comprising culturing a host cell under conditions whereby a polypeptide with FVIII activity is produced by the nucleic acid molecule, wherein the yield of polypeptide with FVIII activity is increased relative to a host cell cultured under the same conditions comprising a reference nucleic acid sequence comprising SEQ ID NO: 32.

[0192] A variety of methods are available for recombinantly producing a FVIII protein from the optimized nucleic acid molecule of the disclosure. A polynucleotide of the desired sequence can be produced by *de novo* solid-phase DNA synthesis or by PCR mutagenesis of an earlier prepared polynucleotide. Oligonucleotide-mediated mutagenesis is one method for preparing a substitution, insertion, deletion, or alteration (*e.g.*, altered codon) in a nucleotide sequence. For example, the starting DNA is altered by hybridizing an oligonucleotide encoding the desired mutation to a single-stranded DNA template. After hybridization, a DNA polymerase is used to synthesize an entire second complementary strand of the template that incorporates the oligonucleotide primer. In one embodiment, genetic engineering, *e.g.*, primer-based PCR mutagenesis, is sufficient to incorporate an alteration, as defined herein, for producing a polynucleotide of the disclosure.

[0193] For recombinant protein production, an optimized polynucleotide sequence of the disclosure encoding the FVIII protein is inserted into an appropriate expression vehicle, *i.e.,* a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence, or in the case of an RNA viral vector, the necessary elements for replication and translation.

[0194] The polynucleotide sequence of the disclosure is inserted into the vector in proper reading frame. The expression vector is then transfected into a suitable target cell which will express the polypeptide. Transfection techniques known in the art include, but are not limited to, calcium phosphate precipitation (Wigler *et al.* 1978, *Cell* 14 : 725) and electroporation (Neumann et al. 1982, *EMBO, J.* 1 : 841). A variety of host-expression vector systems can be utilized to express the FVIII proteins described herein in eukaryotic cells. In one embodiment, the eukaryotic cell is an animal cell, including mammalian cells (e.g. HEK293 cells, PER.C6®, CHO, BHK, Cos, HeLa cells). A polynucleotide sequence of the disclosure can also code for a signal sequence that will permit the FVIII protein to be secreted. One skilled in the art will understand that while the FVIII protein is translated the signal sequence is cleaved by the cell to form the mature protein. Various signal sequences are known in the art, *e.g.,* native factor Vll signal sequence, native factor IX signal sequence and the mouse IgK light chain signal sequence. Alternatively, where a signal sequence is not included the FVIII protein can be recovered by lysing the cells.

[0195] The FVIII protein of the disclosure can be synthesized in a transgenic animal, such as a rodent, goat, sheep, pig, or cow. The term "transgenic animals" refers to non-human animals that have incorporated a foreign gene into their genome. Because this gene is present in germline tissues, it is passed from parent to offspring. Exogenous genes are introduced into single-celled embryos (Brinster et al. 1985, Proc. Natl. Acad.Sci. USA 82:4438). Methods of producing transgenic animals are known in the art including transgenics that produce immunoglobulin molecules (Wagner et al. 1981, Proc. Natl. Acad. Sci. USA 78: 6376; McKnight et al. 1983, Cell

34 : 335; Brinster et al. 1983, Nature 306: 332; Ritchie et al. 1984, Nature 312: 517; Baldassarre et al. 2003, Theriogenology 59 : 831 ; Robl et al. 2003, Theriogenology 59: 107; Malassagne et al. 2003, Xenotransplantation 10 (3): 267).

[0196] The expression vectors can encode for tags that permit for easy purification or identification of the recombinantly produced protein. Examples include, but are not limited to, vector pUR278 (Ruther et al. 1983, EMBO J. 2: 1791) in which the FVIII protein described herein coding sequence can be ligated into the vector in frame with the lac Z coding region so that a hybrid protein is produced; pGEX vectors can be used to express proteins with a glutathione S-transferase (GST) tag. These proteins are usually soluble and can easily be purified from cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The vectors include cleavage sites (*e.g.,* PreCission Protease (Pharmacia, Peapack, N. J.)) for easy removal of the tag after purification.

[0197] For the purposes of this disclosure, numerous expression vector systems can be employed. These expression vectors are typically replicable in the host organisms either as episomes or as an integral part of the host chromosomal DNA. Expression vectors can include expression control sequences including, but not limited to, promoters (*e.g.*, naturally-associated or heterologous promoters), enhancers, signal sequences, splice signals, enhancer elements, and transcription termination sequences. Preferably, the expression control sequences are eukaryotic promoter systems in vectors capable of transforming or transfecting eukaryotic host cells. Expression vectors can also utilize DNA elements which are derived from animal viruses such as bovine papilloma virus, polyoma virus, adenovirus, vaccinia virus, baculovirus, retroviruses (RSV, MMTV or MOMLV), cytomegalovirus (CMV), or SV40 virus. Others involve the use of polycistronic systems with internal ribosome binding sites.

[0198] Commonly, expression vectors contain selection markers (*e.g.*, ampicillin-resistance, hygromycin-resistance, tetracycline resistance or neomycin resistance) to permit detection of those cells transformed with the desired DNA sequences (see, *e.g.*, Itakura *et al.*, US Patent 4,704,362). Cells which have integrated the DNA into their chromosomes can be selected by introducing one or more markers which allow selection of transfected host cells. The marker can provide for prototrophy to an auxotrophic host, biocide resistance (*e.g.,* antibiotics) or resistance to heavy metals such as copper. The selectable marker gene can either be directly linked to the DNA sequences to be expressed, or introduced into the same cell by co-transformation.

[0199] An example of a vector useful for expressing an optimized FVIII sequence is NEOSPLA (U.S. Patent No. 6,159,730). This vector contains the cytomegalovirus promoter/enhancer, the mouse beta globin major promoter, the SV40 origin of replication, the bovine growth hormone polyadenylation sequence, neomycin phosphotransferase exon 1 and

exon 2, the dihydrofolate reductase gene and leader sequence. This vector has been found to result in very high level expression of antibodies upon incorporation of variable and constant region genes, transfection in cells, followed by selection in G418 containing medium and methotrexate amplification. Vector systems are also taught in U.S. Pat. Nos. 5,736,137 and 5,658,570, each of which is incorporated by reference in its entirety herein. This system provides for high expression levels, e.g., > 30 pg/cell/day. Other exemplary vector systems are disclosed *e.g.,* in U.S. Patent No. 6,413,777.

[0200] In other embodiments the polypeptides of the disclosure of the instant disclosure can be expressed using polycistronic constructs. In these expression systems, multiple gene products of interest such as multiple polypeptides of multimer binding protein can be produced from a single polycistronic construct. These systems advantageously use an internal ribosome entry site (IRES) to provide relatively high levels of polypeptides in eukaryotic host cells. Compatible IRES sequences are disclosed in U.S. Pat. No. 6,193,980 which is also incorporated herein.

[0201] More generally, once the vector or DNA sequence encoding a polypeptide has been prepared, the expression vector can be introduced into an appropriate host cell. That is, the host cells can be transformed. Introduction of the plasmid into the host cell can be accomplished by various techniques well known to those of skill in the art, as discussed above. The transformed cells are grown under conditions appropriate to the production of the FVIII polypeptide, and assayed for FVIII polypeptide synthesis. Exemplary assay techniques include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), or fluorescence-activated cell sorter analysis (FACS), immunohistochemistry and the like.

[0202] In descriptions of processes for isolation of polypeptides from recombinant hosts, the terms "cell" and "cell culture" are used interchangeably to denote the source of polypeptide unless it is clearly specified otherwise. In other words, recovery of polypeptide from the "cells" can mean either from spun down whole cells, or from the cell culture containing both the medium and the suspended cells.

[0203] The host cell line used for protein expression is preferably of mammalian origin; most preferably of human or mouse origin, as the isolated nucleic acids of the disclosure have been optimized for expression in human cells. Exemplary host cell lines have been described above. In one embodiment of the method to produce a polypeptide with FVIII activity, the host cell is a HEK293 cell. In another embodiment of the method to produce a polypeptide with FVIII activity, the host cell is a CHO cell.

[0204] Genes encoding the polypeptides of the disclosure can also be expressed in non-mammalian cells such as bacteria or yeast or plant cells. In this regard it will be appreciated that various unicellular non-mammalian microorganisms such as bacteria can also be transformed;

i.e., those capable of being grown in cultures or fermentation. Bacteria, which are susceptible to transformation, include members of the enterobacteriaceae, such as strains of Escherichia coli or Salmonella; Bacillaceae, such as Bacillus subtilis; Pneumococcus; Streptococcus, and Haemophilus influenzae. It will further be appreciated that, when expressed in bacteria, the polypeptides typically become part of inclusion bodies. The polypeptides must be isolated, purified and then assembled into functional molecules.

[0205] Alternatively, optimized nucleotide sequences of the disclosure can be incorporated in transgenes for introduction into the genome of a transgenic animal and subsequent expression in the milk of the transgenic animal (*see, e.g.*, Deboer *et al.*, US 5,741,957, Rosen, US 5,304,489, and Meade *et al.*, US 5,849,992). Suitable transgenes include coding sequences for polypeptides in operable linkage with a promoter and enhancer from a mammary gland specific gene, such as casein or beta lactoglobulin.

[0206] *In vitro* production allows scale-up to give large amounts of the desired polypeptides. Techniques for mammalian cell cultivation under tissue culture conditions are known in the art and include homogeneous suspension culture, *e.g.*, in an airlift reactor or in a continuous stirrer reactor, or immobilized or entrapped cell culture, *e.g.*, in hollow fibers, microcapsules, on agarose microbeads or ceramic cartridges. If necessary and/or desired, the solutions of polypeptides can be purified by the customary chromatography methods, for example gel filtration, ion-exchange chromatography, chromatography over DEAE-cellulose or (immuno-)affinity chromatography, *e.g.,* after preferential biosynthesis of a synthetic hinge region polypeptide or prior to or subsequent to the HIC chromatography step described herein. An affinity tag sequence (*e.g.* a His(6) tag) can optionally be attached or included within the polypeptide sequence to facilitate downstream purification.

[0207] Once expressed, the FVIII protein can be purified according to standard procedures of the art, including ammonium sulfate precipitation, affinity column chromatography, HPLC purification, gel electrophoresis and the like (*see generally* Scopes, Protein Purification (Springer-Verlag, N.Y., (1982)). Substantially pure proteins of at least about 90 to 95% homogeneity are preferred for pharmaceutical uses, with 98 to 99% or more homogeneity being most preferred.

Pharmaceutical Compositions

[0208] Compositions containing an isolated nucleic acid molecule, a polypeptide having FVIII activity encoded by the nucleic acid molecule, a vector, or a host cell of the present disclosure can contain a suitable pharmaceutically acceptable carrier. For example, they can contain excipients and/or auxiliaries that facilitate processing of the active compounds into preparations designed for delivery to the site of action.

[0209] The pharmaceutical composition can be formulated for parenteral administration (*i.e.* intravenous, subcutaneous, or intramuscular) by bolus injection. Formulations for injection can be presented in unit dosage form, *e.g.*, in ampoules or in multidose containers with an added preservative. The compositions can take such forms as suspensions, solutions, or emulsions in oily or aqueous vehicles, and contain formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the active ingredient can be in powder form for constitution with a suitable vehicle, *e.g.*, pyrogen free water.

[0210] Suitable formulations for parenteral administration also include aqueous solutions of the active compounds in water-soluble form, for example, water-soluble salts. In addition, suspensions of the active compounds as appropriate oily injection suspensions can be administered. Suitable lipophilic solvents or vehicles include fatty oils, for example, sesame oil, or synthetic fatty acid esters, for example, ethyl oleate or triglycerides. Aqueous injection suspensions can contain substances, which increase the viscosity of the suspension, including, for example, sodium carboxymethyl cellulose, sorbitol and dextran. Optionally, the suspension can also contain stabilizers. Liposomes also can be used to encapsulate the molecules of the disclosure for delivery into cells or interstitial spaces. Exemplary pharmaceutically acceptable carriers are physiologically compatible solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, water, saline, phosphate buffered saline, dextrose, glycerol, ethanol and the like. In some embodiments, the composition comprises isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, or sodium chloride. In other embodiments, the compositions comprise pharmaceutically acceptable substances such as wetting agents or minor amounts of auxiliary substances such as wetting or emulsifying agents, preservatives or buffers, which enhance the shelf life or effectiveness of the active ingredients.

[0211] Compositions of the disclosure can be in a variety of forms, including, for example, liquid (*e.g.*, injectable and infusible solutions), dispersions, suspensions, semi-solid and solid dosage forms. The preferred form depends on the mode of administration and therapeutic application.

[0212] The composition can be formulated as a solution, micro emulsion, dispersion, liposome, or other ordered structure suitable to high drug concentration. Sterile injectable solutions can be prepared by incorporating the active ingredient in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active ingredient into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-

drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution. The proper fluidity of a solution can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prolonged absorption of injectable compositions can be brought about by including in the composition an agent that delays absorption, for example, monostearate salts and gelatin.

[0213] The active ingredient can be formulated with a controlled-release formulation or device. Examples of such formulations and devices include implants, transdermal patches, and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, for example, ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for the preparation of such formulations and devices are known in the art. *See, e.g.*, Sustained and Controlled Release Drug Delivery Systems, J. R. Robinson, ed., Marcel Dekker, Inc., New York, 1978.

[0214] Injectable depot formulations can be made by forming microencapsulated matrices of the drug in biodegradable polymers such as polylactide-polyglycolide. Depending on the ratio of drug to polymer, and the nature of the polymer employed, the rate of drug release can be controlled. Other exemplary biodegradable polymers are polyorthoesters and polyanhydrides. Depot injectable formulations also can be prepared by entrapping the drug in liposomes or microemulsions.

[0215] Supplementary active compounds can be incorporated into the compositions. In one embodiment, the chimeric protein of the disclosure is formulated with another clotting factor, or a variant, fragment, analogue, or derivative thereof. For example, the clotting factor includes, but is not limited to, factor V, factor VII, factor VIII, factor IX, factor X, factor XI, factor XII, factor XIII, prothrombin, fibrinogen, von Willebrand factor or recombinant soluble tissue factor (rsTF) or activated forms of any of the preceding. The clotting factor of hemostatic agent can also include anti-fibrinolytic drugs, *e.g.*, epsilon-amino-caproic acid, tranexamic acid.

[0216] Dosage regimens can be adjusted to provide the optimum desired response. For example, a single bolus can be administered, several divided doses can be administered over time, or the dose can be proportionally reduced or increased as indicated by the exigencies of the therapeutic situation. It is advantageous to formulate parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. See, *e.g.*, Remington's Pharmaceutical Sciences (Mack Pub. Co., Easton, Pa. 1980).

[0217] In addition to the active compound, the liquid dosage form can contain inert ingredients such as water, ethyl alcohol, ethyl carbonate, ethyl acetate, benzyl alcohol, benzyl benzoate,

propylene glycol, 1,3-butylene glycol, dimethylformamide, oils, glycerol, tetrahydrofurfuryl alcohol, polyethylene glycols, and fatty acid esters of sorbitan.

[0218] Non-limiting examples of suitable pharmaceutical carriers are also described in Remington's Pharmaceutical Sciences by E. W. Martin. Some examples of excipients include starch, glucose, lactose, sucrose, gelatin, malt, rice, flour, chalk, silica gel, sodium stearate, glycerol monostearate, talc, sodium chloride, dried skim milk, glycerol, propylene, glycol, water, ethanol, and the like. The composition can also contain pH buffering reagents, and wetting or emulsifying agents.

[0219] For oral administration, the pharmaceutical composition can take the form of tablets or capsules prepared by conventional means. The composition can also be prepared as a liquid for example a syrup or a suspension. The liquid can include suspending agents (*e.g.*, sorbitol syrup, cellulose derivatives or hydrogenated edible fats), emulsifying agents (lecithin or acacia), non-aqueous vehicles (*e.g.*, almond oil, oily esters, ethyl alcohol, or fractionated vegetable oils), and preservatives (*e.g.*, methyl or propyl-p-hydroxybenzoates or sorbic acid). The preparations can also include flavoring, coloring and sweetening agents. Alternatively, the composition can be presented as a dry product for constitution with water or another suitable vehicle.

[0220] For buccal administration, the composition can take the form of tablets or lozenges according to conventional protocols.

[0221] For administration by inhalation, the compounds for use according to the present disclosure are conveniently delivered in the form of a nebulized aerosol with or without excipients or in the form of an aerosol spray from a pressurized pack or nebulizer, with optionally a propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoromethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit can be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, *e.g.*, gelatin for use in an inhaler or insufflator can be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

[0222] The pharmaceutical composition can also be formulated for rectal administration as a suppository or retention enema, *e.g.*, containing conventional suppository bases such as cocoa butter or other glycerides.

[0223] In one embodiment, a pharmaceutical composition comprises a polypeptide having Factor VIII activity, an optimized nucleic acid molecule encoding the polypeptide having Factor VIII activity, the vector comprising the nucleic acid molecule, or the host cell comprising the vector, and a pharmaceutically acceptable carrier. In some embodiments, the composition is administered by a route selected from the group consisting of topical administration, intraocular administration, parenteral administration, intrathecal administration, subdural administration and

oral administration. The parenteral administration can be intravenous or subcutaneous administration.

Methods of Treatment

[0224] In some aspects, the present disclosure is directed to methods of treating a disease or condition in a subject in need thereof, comprising administering a nucleic acid molecule, a vector, a polypeptide, or a pharmaceutical composition disclosed herein.

[0225] In some embodiments, the disclosure is directed to methods of treating a bleeding disorder. In some embodiments, the disclosure is directed to methods of treating hemophilia A.

[0226] The isolated nucleic acid molecule, vector, or polypeptide can be administered intravenously, subcutaneously, intramuscularly, or via any mucosal surface, *e.g.,* orally, sublingually, buccally, sublingually, nasally, rectally, vaginally or via pulmonary route. The isolated nucleic acid molecule, vector, or polypeptide can also be administered intraneurally, intraocularly, and intrathecally. The clotting factor protein can be implanted within or linked to a biopolymer solid support that allows for the slow release of the chimeric protein to the desired site.

[0227] In one embodiment, the route of administration of the isolated nucleic acid molecule, vector, or polypeptide is parenteral. The term parenteral as used herein includes intravenous, intraarterial, intraperitoneal, intramuscular, subcutaneous, rectal or vaginal administration. In some embodiments, the isolated nucleic acid molecule, vector, or polypeptide is administered intravenously. While all these forms of administration are clearly contemplated as being within the scope of the disclosure, a form for administration would be a solution for injection, in particular for intravenous or intraarterial injection or drip.

[0228] Effective doses of the compositions of the present disclosure, for the treatment of conditions vary depending upon many different factors, including means of administration, target site, physiological state of the patient, whether the patient is human or an animal, other medications administered, and whether treatment is prophylactic or therapeutic. Usually, the patient is a human but non-human mammals including transgenic mammals can also be treated. Treatment dosages can be titrated using routine methods known to those of skill in the art to optimize safety and efficacy.

[0229] The nucleic acid molecule, vector, or polypeptides of the disclosure can optionally be administered in combination with other agents that are effective in treating the disorder or condition in need of treatment (*e.g.,* prophylactic or therapeutic).

[0230] As used herein, the administration of isolated nucleic acid molecules, vectors, or polypeptides of the disclosure in conjunction or combination with an adjunct therapy means the

sequential, simultaneous, coextensive, concurrent, concomitant or contemporaneous administration or application of the therapy and the disclosed polypeptides. Those skilled in the art will appreciate that the administration or application of the various components of the combined therapeutic regimen can be timed to enhance the overall effectiveness of the treatment. A skilled artisan (*e.g.*, a physician) would be readily be able to discern effective combined therapeutic regimens without undue experimentation based on the selected adjunct therapy and the teachings of the instant specification.

[0231] It will further be appreciated that the isolated nucleic acid molecule, vector, or polypeptide of the instant disclosure can be used in conjunction or combination with an agent or agents (*e.g.*, to provide a combined therapeutic regimen). Exemplary agents with which a polypeptide or polynucleotide of the disclosure can be combined include agents that represent the current standard of care for a particular disorder being treated. Such agents can be chemical or biologic in nature. The term "biologic" or "biologic agent" refers to any pharmaceutically active agent made from living organisms and/or their products which is intended for use as a therapeutic.

[0232] The amount of agent to be used in combination with the polynucleotides or polypeptides of the instant disclosure can vary by subject or can be administered according to what is known in the art. *See, e.g.*, Bruce A Chabner *et al.*, *Antineoplastic Agents, in* GOODMAN & GILMAN'S THE PHARMACOLOGICAL BASIS OF THERAPEUTICS 1233-1287 ((Joel G. Hardman *et al.*, eds., 9th ed. 1996). In another embodiment, an amount of such an agent consistent with the standard of care is administered.

[0233] In one embodiment, also disclosed herein is a kit, comprising the nucleic acid molecule disclosed herein and instructions for administering the nucleic acid molecule to a subject in need thereof. In another embodiment, disclosed herein is a baculovirus system for production of the nucleic acid molecule provided herein. The nucleic acid molecule is produced in insect cells. In another embodiment, a nanoparticle delivery system for expression constructs is provided. The expression construct comprises the nucleic acid molecule disclosed herein.

Gene Therapy

[0234] In some embodiments, the nucleic acid molecule disclosed herein is used in gene therapy. The optimized FVIII nucleic acid molecules disclosed herein can be used in any context where expression of FVIII is required. In some embodiments, the nucleic acid molecules comprise the nucleotide sequence of SEQ ID NO: 9. In some embodiments, the nucleic acid molecules comprise the nucleotide sequence of SEQ ID NO: 33. In some embodiments, the nucleic acid molecules comprise the nucleotide sequence of SEQ ID NO: 14. In some embodiments, the nucleic acid molecules comprise the nucleotide sequence of SEQ ID NO: 35.

[0235] For example, somatic gene therapy has been explored as a possible treatment for hemophilia A. Gene therapy is a particularly appealing treatment for hemophilia because of its potential to cure the disease through continuous endogenous production of FVIII following a single administration of vector. Hemophilia A is well suited for a gene replacement approach because its clinical manifestations are entirely attributable to the lack of a single gene product (FVIII) that circulates in minute amounts (200ng/ml) in the plasma.

[0236] In one aspect, the nucleic acid molecule described herein may be used in AAV gene therapy. AAV is able to infect a number of mammalian cells. See, e.g., Tratschin et al. (1985) Mol. Cell Biol. 5:3251-3260 and Grimm et al. (1999) Hum. Gene Ther. 10:2445-2450. A rAAV vector carries a nucleic acid sequence encoding a gene of interest, or fragment thereof, under the control of regulatory sequences which direct expression of the product of the gene in cells. In some embodiments, the rAAV is formulated with a carrier and additional components suitable for administration.

[0237] In another aspect, the nucleic acid molecule described herein may be used in lentiviral gene therapy. Lentiviruses are RNA viruses wherein the viral genome is RNA. When a host cell is infected with a lentivirus, the genomic RNA is reverse transcribed into a DNA intermediate which is integrated very efficiently into the chromosomal DNA of infected cells. In some embodiments, the lentivirus is formulated with a carrier and additional components suitable for administration. In another aspect, the nucleic acid molecule described herein may be used in adenoviral therapy. A review of the use of adenovirus for gene therapy can be found e.g. in Wold et al. (2013) Curr Gene Ther. 13(6):421-33). In another aspect, the nucleic acid molecule described herein may be used in non-viral gene therapy.

[0238] An optimized FVIII protein of the disclosure can be produced *in vivo* in a mammal, *e.g.*, a human patient, using a gene therapy approach to treatment of a bleeding disease or disorder selected from the group consisting of a bleeding coagulation disorder, hemarthrosis, muscle bleed, oral bleed, hemorrhage, hemorrhage into muscles, oral hemorrhage, trauma, trauma capitis, gastrointestinal bleeding, intracranial hemorrhage, intra-abdominal hemorrhage, intrathoracic hemorrhage, bone fracture, central nervous system bleeding, bleeding in the retropharyngeal space, bleeding in the retroperitoneal space, and bleeding in the iliopsoas sheath would be therapeutically beneficial. In one embodiment, the bleeding disease or disorder is hemophilia. In another embodiment, the bleeding disease or disorder is hemophilia A. This involves administration of an optimized FVIII encoding nucleic acid operably linked to suitable expression control sequences. In certain embodiment, these sequences are incorporated into a viral vector. Suitable viral vectors for such gene therapy include adenoviral vectors, lentiviral vectors, baculoviral vectors, Epstein Barr viral vectors, papovaviral vectors, vaccinia viral vectors,

herpes simplex viral vectors, and adeno associated virus (AAV) vectors. The viral vector can be a replication-defective viral vector. In other embodiments, an adenoviral vector has a deletion in its E1 gene or E3 gene. In other embodiments, the sequences are incorporated into a non-viral vector known to those skilled in the art.

[0239] In another aspect, the methods disclosed herein provide techniques for the targeted, specific alteration of the genetic information (e.g. genome) of living organisms. As used herein, the term “alteration” or “alteration of genetic information” refers to any change in the genome of a cell. In the context of treating genetic disorders, alterations may include, but are not limited to, insertion, deletion and/or correction.

[0240] In some aspects, alterations may also include a gene knock-in, knock-out or knock down. As used herein, the term “knock-in” refers to an addition of a DNA sequence, or fragment thereof into a genome. Such DNA sequences to be knocked-in may include an entire gene or genes, may include regulatory sequences associated with a gene or any portion or fragment of the foregoing. For example, a cDNA encoding the wild-type protein may be inserted into the genome of a cell carrying a mutant gene. Knock-in strategies need not replace the defective gene, in whole or in part. In some cases, a knock-in strategy may further involve substitution of an existing sequence with the provided sequence, e.g., substitution of a mutant allele with a wildtype copy. The term “knock-out” refers to the elimination of a gene or the expression of a gene. For example, a gene can be knocked out by either a deletion or an addition of a nucleotide sequence that leads to a disruption of the reading frame. As another example, a gene may be knocked out by replacing a part of the gene with an irrelevant sequence. The term “knock-down” as used herein refers to reduction in the expression of a gene or its gene product(s). As a result of a gene knock-down, the protein activity or function may be attenuated or the protein levels may be reduced or eliminated.

[0241] In some embodiments, the nucleic acid sequences disclosed herein are used for genome editing. Genome editing generally refers to the process of modifying the nucleotide sequence of a genome, preferably in a precise or pre-determined manner. Examples of methods of genome editing described herein include methods of using site-directed nucleases to cut deoxyribonucleic acid (DNA) at precise target locations in the genome, thereby creating single-strand or double strand DNA breaks at particular locations within the genome. Such breaks can be and regularly are repaired by natural, endogenous cellular processes, such as homology-directed repair (HDR) and non-homologous end joining (NHEJ), as recently reviewed in Cox et al. (2015). Nature Medicine 21(2): 121-31. These two main DNA repair processes consist of a family of alternative pathways. NHEJ directly joins the DNA ends resulting from a double-strand break, sometimes with the loss or addition of nucleotide sequence, which may disrupt or enhance gene

expression. HDR utilizes a homologous sequence, or donor sequence, as a template for inserting a defined DNA sequence at the break point. The homologous sequence can be in the endogenous genome, such as a sister chromatid. Alternatively, the donor can be an exogenous nucleic acid, such as a plasmid, a single-strand oligonucleotide, a double-stranded oligonucleotide, a duplex oligonucleotide or a virus, that has regions of high homology with the nuclease-cleaved locus, but which can also contain additional sequence or sequence changes including deletions that can be incorporated into the cleaved target locus. A third repair mechanism can be microhomology-mediated end joining (MMEJ), also referred to as "Alternative NHEJ," in which the genetic outcome is similar to NHEJ in that small deletions and insertions can occur at the cleavage site. MMEJ can make use of homologous sequences of a few base pairs flanking the DNA break site to drive a more favored DNA end joining repair outcome, and recent reports have further elucidated the molecular mechanism of this process, see, e.g., Cho and Greenberg (2015). Nature 518, 174-76. In some instances, it may be possible to predict likely repair outcomes based on analysis of potential microhomologies at the site of the DNA break.

[0242] Each of these genome editing mechanisms can be used to create desired genomic alterations. A step in the genome editing process can be to create one or two DNA breaks, the latter as double-strand breaks or as two single-stranded breaks, in the target locus as near the site of intended mutation. This can be achieved via the use of site-directed polypeptides, such as the CRISPR endonuclease system and others.

[0243] In another aspect, the nucleic acid molecule described herein may be used in lipid nanoparticle (LNP)-mediated delivery of FVIII ceDNA. Lipid nanoparticles formed from cationic lipids with other lipid components, such as neutral lipids, cholesterol, PEG, PEGylated lipids, and oligonucleotides have been used to block degradation of nucleic acids in plasma and facilitate the cellular uptake of oligonucleotides. Such lipid nanoparticles may be used to deliver the nucleic acid molecule described herein to subjects.

[0244] The disclosure provides a method of increasing expression of a polypeptide with FVIII activity in a subject comprising administering the isolated nucleic acid molecule of the disclosure to a subject in need thereof, wherein the expression of the polypeptide is increased relative to a reference nucleic acid molecule comprising SEQ ID NO: 32. The disclosure also provides a method of increasing expression of a polypeptide with FVIII activity in a subject comprising administering a vector of the disclosure to a subject in need thereof, wherein the expression of the polypeptide is increased relative to a vector comprising a reference nucleic acid molecule.

[0245] All of the various aspects, embodiments, and options described herein can be combined in any and all variations.

[0246] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

[0247] Having generally described this disclosure, a further understanding can be obtained by reference to the examples provided herein. These examples are for purposes of illustration only and are not intended to be limiting.

EXAMPLES

Example 1: Modified FVIIIXTEN expression cassette

[0248] It was hypothesized that the transgene expression level can be increased by codon-optimizing the coding sequence for the targeted hosts. Higher level of FVIII expression has been demonstrated using a V1.0 FVIIIco6XTEN expression cassette (SEQ ID NO: 32)(**FIG. 1**) in previous studies as described in U.S. Publication No. 20190185543. However, to further improve the target specificity and reduce immunogenicity, the FVIIIXTEN expression cassette was codon-optimized with CpG motifs depleted to reduce the innate immune response raised against the DNA vector encoding FVIIIXTEN expression cassette with parvoviral ITRs. In this study, the modified V2.0 FVIIIXTEN expression cassette comprises of a codon optimized cDNA encoding B-domain deleted human Factor VIII (BDDcoFVIII) fused with XTEN 144 peptide (FVIIIXTEN) under the regulation of liver-specific modified mouse transthyretin (mTTR) promoter (mTTR482) with enhancer element (A1MB2), hybrid synthetic intron (Chimeric Intron), the Woodchuck Posttranscriptional Regulatory Element (WPRE), and the Bovine Growth Hormone Polyadenylation (bGHpA) signal (SEQ ID NO: 14).(**FIG. 1**). The *in vivo* functionality of the modified V2.0 FVIIIXTEN expression cassette has been demonstrated with different parvoviral ITRs in a form of single-stranded (ss) or closed-end (ce) DNA by systemic delivery *via* hydrodynamic tail-vein injections in hFVIIIR593C$^{+/+}$/HemA mice.

Example 2: Single-stranded FVIIIXTEN (ssFVIIIXTEN) DNA

Modified FVIIIXTEN shown significantly higher levels of activity in vivo

[0249] It was hypothesized that the hairpin formed within the ITR region drives the long-term persistent expression of transgene at higher level. To validate the functionality of modified FVIIIXTEN expression cassette *in vivo*, single-stranded DNA (ssDNA) comprising of V1.0 or V2.0 human FVIIIXTEN with preformed erythrovirus B19 ITRs was tested in hFVIIIR593C$^{+/+}$/HemA mice. These mice contain a human *FVIII-R593C* transgene, designed with the murine albumin (*Alb*) promoter driving expression of an altered human coagulation factor VIII (FVIII) cDNA harboring a mutation that is frequently observed in patients with mild hemophilia A. These mice

also carry a knock-out of the *FVIII* gene and are deficient for endogenous FVIIIprotein. These double mutant mice are tolerant of human FVIII injection and have no FVIII activity. They produce very little inhibitory antibodies and lack FVIII responsive T cells or B cells after treatment with human FVIII. The hFVIIIR593C$^{+/+}$/HemA mouse is further described in Bril, et al. (2006) Thromb. Haemost. 95(2): 341-7.

[0250] The ssFVIIIXTEN with preformed B19 ITRs was generated by denaturing the double-stranded DNA fragment products (FVIII expression cassette and plasmid backbone) of MscI digestion at 95 ºC (denaturation) and then cooling down at 4 ºC (renaturation) to allow the palindromic ITR sequences to fold (**FIG. 2**). The ssFVIIIXTEN was then systemically injected *via* hydrodynamic tail-vein injections at 800 µg/kg hFVIIIR593C$^{+/+}$/HemA mice. Plasma samples were collected from injected mice at indicated intervals for 5.5 months and the FVIII activity was measured by the Chromogenix Coatest® SP Factor VIII chromogenic assay, according to the manufacturer's instructions.

[0251] The plasma FVIII activity normalized to percent of normal for V1.0 and V2.0 ssFVIIIXTEN injected animals is shown in **FIG. 3**. The results showed significant improvement in FVIII activity in V2.0 injected cohorts in comparison to V1.0 ssFVIIIXTEN. However, there was initial drop in FVIII expression observed up to day 56 but then the levels were stabilized up to day 168 suggesting the persistent expression of parvoviral ITRs flanked V2.0 ssFVIIIXTEN from the liver of injected animals. Thus, these results validate the functionality of modified FVIIIXTEN with long-term persistent expression of FVIII activity in comparison to V1.0 *in vivo*.

Human Bocavirus (HBoV1) ITRs shown supraphysiological levels of FVIII expression in vivo

[0252] To determine the impact of ITRs on stability and long-term persistency of transgene expression, the improved version of FVIIIXTEN was tested with human Bocavirus (HBoV1), human erythrovirus B19, Goose Parvovirus (GPV), or their variant ITRs *in vivo*. These ITRs were engineered based on the thermostability and ITR-specific elements required for the long-term persistency of viral genome in their respective hosts. Tested ITR variants and predicted secondary structure is described in previous U.S. Patent Application No. 63/069,114. Individual variant ITR was cloned into the synthetic FVIIIXTEN expression construct using Golden Gate Assembly and verified by sequencing at the Genewiz sequencing facility. Sequence verified constructs were then used for generating ssFVIIIXTEN (ssDNA), as described above, and then systemically injected in hFVIIIR593C$^{+/+}$/HemA mice *via* hydrodynamic tail-vein injection at 200, 800, or 1600 µg/kg. Plasma samples were collected from injected mice at indicated interval for 5.5 months and the FVIII activity was measured by the Chromogenix Coatest® SP Factor VIII chromogenic assay, according to the manufacturer's instructions.

[0253] The plasma FVIII activity normalized to percent of normal for V2.0 ssFVIIIXTEN injected animals is shown in **FIG. 4**. The results showed long-term persistent FVIIIXTEN expression in all the parvoviral ITRs tested albeit with varying levels. All variants or hybrids of GPV ITRs tested showed continuous decline in the levels of FVIIIXTEN expression in comparison with other parvoviral ITRs. In contrary, HBoV1 and B19 ITRs showed initial decline in FVIIIXTEN up to day 56 and then stabilized through day 168 suggesting the ITR-dependent persistency of FVIIIXTEN transgene *in vivo*. Unlike GPV ITRs, both B19 and HBoV1 ITRs showed significantly higher levels of FVIII expression irrespective of the variant tested suggesting the ITR-dependent stability of FVIIIXTEN transgene *in vivo*.

[0254] Among different parvoviral ITRs tested, HBoV1 ITRs showed significantly higher levels (>1000%) of normal FVIII activity in hFVIIIR593C$^{+/+}$/HemA mice. (**FIG. 4**). These results validate the functionality of the modified FVIIIXTEN expression with different parvoviral ITRs and demonstrate the ITR-dependent stability as well as persistency of transgene expression *in vivo*.

Example 3: Closed end FVIIIXTEN (ceFVIIIXTEN) DNA

[0255] Though ssFVIIIXTEN (ssDNA) was effective in expressing a modified FVIIIXTEN expression cassette *in vivo*, there are several limitations associated with ssDNA to be used as a non-viral gene therapy vector. One of them is the level of endotoxin contamination due to the prokaryotic host (*E. coli*) used for generating plasmid DNA, which also contains the extraneous sequences, such as antibiotic resistance gene and prokaryotic origin of replication, needed for selection and amplification in *E. coli*. To address these challenges and limitations, a eukaryotic cell-based system was developed to generate DNA therapeutic drug substance in a form of closed-end DNA (ceDNA) comprising of the FVIIIXTEN expression cassette with parvoviral ITRs. The genetic organization of ceDNA, resembles recombinant AAV vector DNA, but differs in conformation.

[0256] To generate this DNA vector, the baculovirus insect cell system was leveraged, which is widely used for the biologics manufacturing and is the only platform approved by the FDA for recombinant influenza vaccine manufacturing. Three different approaches of ceDNA production were employed in the baculovirus system, as described in U.S. Patent Application No. 63/069,073. An exemplary purified ceDNA encoding modified FVIIIXTEN with AAV2 or HBoV1 ITRs in comparison with the starting material (SM) is shown in **FIG. 5A**.

[0257] To validate the functionality of modified FVIIIXTEN as expressed from ceDNA, purified ceFVIIIXTEN was injected systemically *via* hydrodynamic tail-vein injections in hFVIIIR593C$^{+/+}$/HemA mice at 0.3 µg, 1.0 µg, or 2.0 µg/mouse, which is equivalent to 12 µg, 40

µg, and 80 µg/kg, respectively. Plasma samples from injected mice were collected at indicated interval and FVIII activity was measured by the chromogenic assay, as described above.

[0258] The plasma FVIII activity normalized to percent of normal for ceFVIIIXTEN injected animals is shown in **FIG. 5B**. The results showed dose-dependent response in HemA mice with supraphysiological levels (>500% of normal) of FVIII expression observed in the highest dose tested up to day 56 post injection. Interestingly, similar level of expression was achieved when the mice were injected with ssFVIIIXTEN at 1600 µg/kg, which is at least 20x higher the dose of ceFVIIIXTEN (80 µg/kg) (**FIG. 4**). This data suggests that ceDNA provides higher level of FVIII expression in comparison to the ssDNA form. Thus, these studies validate the functionality of modified FVIIIXTEN as expressed from either ssDNA or ceDNA and confirms that codon optimization along with use of optimized ITRs can produce a functional transgene and improve its long-term persistency.

Example 4: Modified FVIIIXTEN expression cassette

[0259] The V2.0 FVIIIXTEN expression cassette contains a mTTR promoter and enhancer element (see **FIG. 1**). However, this promoter is mouse-liver specific and is not well-studied or characterized to determine the liver-specificity in large animal models or in human patients. Therefore, in this study V3.0 FVIIIXTEN expression cassette (SEQ ID NO: 35) was generated by replacing the mTTR promoter and enhancer element with human liver-specific alpha-1-antitrypsin (A1AT) promoter (SEQ ID NO: 36) in the V2.0 expression cassette (**FIG. 1**).

Example 5: FVIIIXTEN HBoV1 mTTR vs A1AT ssDNA *in vivo* efficacy

[0260] To validate the functionality of the mTTR versus the A1AT promoter *in vivo*, single-stranded DNA (ssDNA) comprising codon-optimized human FVIIIXTEN (ssFVIIIXTEN) with preformed HBoV1 ITRs, generating the constructs depicted in **FIG. 6A**. The ssFVIIIXTEN with preformed HBoV1 ITRs was generated by denaturing the double-stranded DNA (dsDNA) fragment products (mTTR or A1AT FVIII expression cassette and plasmid backbone) of PmlI digestion at 95 ºC and then cooling down at 4 ºC to allow the palindromic ITR sequences to fold. The resulting ssFVIIIXTEN was checked by 0.8 to 1.2% agarose gel electrophoresis. The gel analysis showed half the size of dsDNA for ssFVIIIXTEN suggesting efficient hairpin formation (**FIG. 6B**).

[0261] The ssFVIIIXTEN was systemically injected into hFVIIIR593C+/+/HemA mice via hydrodynamic tail-vein injections at 10 µg/mouse. Plasma samples were collected from injected mice at 7 day intervals for 5.5 months. Plasma FVIII activity was measured by the Chromogenix Coatest® SP Factor VIII chromogenic assay, according to the manufacturer's instructions.

[0262] The plasma FVIII activity normalized to percent of normal for ssFVIIIXTEN injected animals is shown in **FIG. 6C**. These results showed equivalent levels of FVIII expression up to day 21 post-injection, suggesting there is no significant difference in FVIIIXTEN levels expressed by the mTTR or A1AT promoter in hFVIIIR593C+/+/HemA mice animal model.

Example 6: FVIIIXTEN AAV2 full-length vs truncated ceDNA in vivo efficacy

[0263] Adeno-associated Virus (AAV) vector is known to produce different replicative forms of viral genome (e.g. monomer, dimer, or multimer) through ITR-ITR concatamerization.We previously observed that a closed-end DNA (ceDNA) vector comprising the V2.0 codon-optimized FVIIIXTEN (ceFVIIIXTEN) flanked by AAV2 WT ITRs produced a truncated species of ceFVIIIXTEN along with the monomeric and multimeric forms of vector genome in the baculovirus system. See, e.g., International Application No. PCT/US21/47218)

[0264] In this study, to further investigate the properties of the truncated species of ceFVIIIXTEN, we purified both full-length and truncated species of ceFVIIIXTEN by continuous-elution electrophoresis, as described in International Application No. PCT/US21/47218. The purity of both species of ceFVIIIXTEN was determined by agarose gel electrophoresis and the results showed major bands corresponding to the size of full-length (8.3kb) and truncated (6.0kb) species of ceFVIIIXTEN (**FIG. 7A**).

[0265] To further validate the nucleotide sequences of both species of ceFVIIIXTEN, we performed next-generation sequence (NGS) analyses on purified ceFVIIIXTEN materials using the MiSeq Illumina Sequence Analyzer. The NGS results, shown in **FIG. 7B**, showed >80% coverage for the full-length ceFVIIIXTEN sequence reads (top panel) and >75% coverage for the truncated ceFVIIIXTEN species (bottom panel) with some impurities coming from the host cell and/or baculoviral genome. Further analyses of NGS data revealed that the truncated ceFVIIIXTEN reads were missing a large portion of the chimeric intron region while retaining the ITRs sequences at the 5' end of the ceFVIIIXTEN (**FIG. 7B**, bottom panel).

[0266] To further validate the functionality of the truncated species of ceFVIIIXTEN, purified full-length or truncated species of ceFVIIIXTEN was systemically injected in hFVIIIR593C+/+/HemA mice via hydrodynamic tail-vein injections at either 40 or 80 µg/kg. Plasma samples were collected from injected mice at 7 day intervals and plasma FVIII activity was measured by the Chromogenix Coatest® SP Factor VIII chromogenic assay, according to the manufacturer's instructions. The plasma FVIII activity normalized to percent of normal for ceFVIIIXTEN injected animals is shown in **FIG. 7C**.

[0267] The results showed supraphysiological levels of FVIII expression in full-length ceFVIIIXTEN injected cohorts. However, animals injected with truncated ceFVIIIXTENshowed 2-

fold lower FVIII expression at both doses tested up to day 21 post injections (**FIG. 7C**). This data further supports the contribution of the chimeric intron to the improvement in the expression levels of the V2.0 codon-optimized FVIIIXTEN *in vivo* (**FIG. 7C**).

Example 7: FVIIIXTEN closed-end DNA (ceFVIIIXTEN) in vivo efficacy

[0268] In this study, we investigated the in vivo efficacy of ceDNAencoding modified FVIIIXTEN and flanked by either AAV2 or HBoV1 ITRs. ceFVIIIXTEN DNA was generated in the baculovirus system using either AAV2 or HBoV1 ITRs as described previously (see, e.g. International Application No. PCT/US21/47218). Agarose gel was used to analyze the purity of each ceDNA in comparison to the starting material (SM) is shown in **FIG. 8A**.

[0269] Purified ceFVIIIXTEN was injected systemically via hydrodynamic tail-vein injections in hFVIIIR593C+/+/HemA mice at either 1.0 µg or 2.0 µg/mouse, which is equivalent to either 40 µg or 80 µg/kg, respectively. Plasma samples from injected mice were collected at interval and FVIII activity was measured by the chromogenic assay, as described above.

[0270] The plasma FVIII activity normalized to percent of normal for ceFVIIIXTEN injected animals is shown in **FIG. 8B**.

[0271] The results showed comparable FVIII expression levels for ceDNA vectors flanked by either AAV2 or HBoV1 ITRs. As seen previously, FVIII expression levels gradually declined in treated animals up to day 256, suggesting the loss of vector over time in the liver hepatocytes. These studies validate the functionality and long-term persistence of modified V2.0 FVIIIXTEN as expressed from ceDNA vectors comprising AAV2 or HBoV1 ITRs.

[0272] The foregoing description of the specific embodiments will so fully reveal the general nature of the disclosure that others can, by applying knowledge within the skill of the art, readily modify and/or adapt for various applications such specific embodiments, without undue experimentation, without departing from the general concept of the present disclosure. Therefore, such adaptations and modifications are intended to be within the meaning and range of equivalents of the disclosed embodiments, based on the teaching and guidance presented herein. It is to be understood that the phraseology or terminology herein is for the purpose of description and not of limitation, such that the terminology or phraseology of the present specification is to be interpreted by the skilled artisan in light of the teachings and guidance.

[0273] Other embodiments of the disclosure will be apparent to those skilled in the art from consideration of the specification and practice of the disclosure disclosed herein. It is intended that the specification and examples be considered as exemplary only, with a true scope and spirit of the disclosure being indicated by the following claims.

[0274] All patents and publications cited herein are incorporated by reference herein in their entirety.

SEQUENCES

Table 1. Additional Nucleotide and Amino Acid Sequences

SEQ ID NO / Description	Nucleotide or amino acid sequence
SEQ ID NO. 1: HBoV1 5' ITR	GTGGTTGTACAGACGCCATCTTGGAATCCAATATGTCTGCCGGCTCAGTCATGCCTGCGCTGCGCGCAGCGCGCTGCG CGCGCGCATGATCTAATCGCCGGCAGACATATTGGATTCCAAGATGGCGTCTGTACAACCAC
SEQ ID NO. 2: HBoV1 3' ITR	TTGCTTATGCAATCGCGAAACTCTATATCTTTTAATGTGTTGTTGTTGTACATGCGCCATCTTAGTTTTATATCAGCT GGCGCCTTAGTTATATAACATGCATGTTATATAACTAAGGCGCCAGCTGATATAAAACTAAGATGGCGCATGTACAAC AACAACACATTAAAAGATATAGAGTTTCGCGATTGCATAAGCAA
SEQ ID NO. 3: HBoV1-5'ITR-mTTR482-Intron-coBDDFVIIIXTEN (V2.0)-WPRE-bGHPolyA-HBoV1-3'ITR	GTATACCTGCAGGCTAGCCACGTGTTGTTGTTGTACATGCGCCATCTTAGTTTTATATCAGCTGGCGCCTTAGTTATA TAACATGCATGTTATATAACTAAGGCGCCAGCTGATATAAAACTAAGATGGCGCATGTACAACAACAACACATTAAAA GATATAGAGTTTCGCGATTGCAAGCTTGGCCCCAGGTTAATTTTTAAAAAGCAGTCAAAGGTCAAAGTGGCCCTTGGC AGCATTTACTCTCTCTATTGACTTTGGTTAATAATCTCAGGAGCACAAACATTCCTGGAGGCAGGAGAAGAAATCAAC ATCCTGGACTTATCCTCTGGGCCTCTCCCCACCTTCGATGGCCCCAGGTTAATTTTTAAAAAGCAGTCAAAGGTCAAA GTGGCCCTTGGCAGCATTTACTCTCTCTATTGACTTTGGTTAATAATCTCAGGAGCACAAACATTCCTGGAGGCAGGA GAAGAAATCAACATCCTGGACTTATCCTCTGGGCCTCTCCCCACCGATATCTACCTGCTGATCGCCCGGCCCCTGTTC AAACATGTCCTAATACTCTGTCGGGGCAAAGGTCGGCAGTAGTTTTCCATCTTACTCAACATCCTCCCAGTGTACGTA GGATCCTGTCTGTCTGCACATTTCGTAGAGCGAGTGTTCCGATACTCTAATCTCCCGGGGCAAAGGTCGTATTGACTT AGGTTACTTATTCTCCTTTTGTTGACTAAGTCAATAATCAGAATCAGCAGGTTTGGAGTCAGCTTGGCAGGGATCAGC AGCCTGGGTTGGAAGGAGGGGGTATAAAAGCCCCTTCACCAGGAGAAGCCGTCACACAGATCCACAAGCTCCTGCTAG GAATTCTCAGGAGCACAAACATTCCTGGAGGCAGGAGAAGAAATCAACATCCTGGACTTATCCTCTGGGCCTCTCCCC ACCGATATCTACCTGCTGATCGCCCGGCCCCTGTTCAAACATGTCCTAATACTCTGTCGGGGCAAAGGTCGGCAGTAG TTTTCCATCTTACTCAACATCCTCCCAGTGTACGTAGGATCCTGTCTGTCTGCACATTTCGTAGAGCGAGTGTTCCGA TACTCTAATCTCCCGGGGCAAAGGTCGTATTGACTTAGGTTACTTATTCTCCTTTTGTTGACTAAGTCAATAATCAGA ATCAGCAGGTTTGGAGTCAGCTTGGCAGGGATCAGCAGCCTGGGTTGGAAGGAGGGGGTATAAAAGCCCCTTCACCAG GAGAAGCCGTCACACAGATCCACAAGCTCCTGCTAGAGTCGCTGCGCGCTGCCTTCGCCCCGTGCCCCGCTCCGCCGC CGCCTCGCGCCGCCCGCCCCGGCTCTGACTGACCGCGTTACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTCC GGGCTGTAATTAGCGCTTGGTTTATTGACGGCTTGTTTCTTTTCTGTGGCTGCGTGAAAGCCTTGAGGGGCTCCGGGA AGGCCCTTTGTGCGGGGGGAGCGGCTCGGGGGGTGCGTGCGTGTGTGTGTGCGTGGGGAGCGCCGCGTGCGGCTCCGC GCTGCCCGGCGGCTGTGAGCGCTGCGGGCGCGGCGCGGGGCTTTGTGCGCTCCGCAGTGTGCGCGAGGGGAGCGCGGC CGGGGGCGGTGCCCCGCGGTGCGGGGGGGGCTGCGAGGGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGGGGGTG AGCAGGGGGTGTGGGCGCGTCGGTCGGGCTGCAACCCCCCCTGCACCCCCCTCCCCGAGTTGCTGAGCACGGCCCGGC TTCGGGTGCGGGGCTCCGTACGGGGCGTGGCGCGGGGCTCGCCGTGCCGGGCGGGGGGTGGCGGCAGGTGGGGGTGCC GGGCGGGGCGGGGCCGCCTCGGGCCGGGGAGGGCTCGGGGGAGGGGCGCGGCGGCCCCCGGAGCGCCGGCGGCTGTCG AGGCGCGGCGAGCCGCAGCCATTGCCTTTTATGGTAATCGTGCGAGAGGGCGCAGGGACTTCCTTTGTCCCAAATCTG TGCGGAGCCGAAATCTGGGAGGCGCCGCCGCACCCCCTCTAGCGGGCGCGGGGCGAAGCGGTGCGGCGCCGGCAGGAA GGAAATGGGCGGGGAGGGCCTTCGTGCGTCGCCGCGCCGCCGTCCCCTTCTCCCTCTCCAGCCTCGGGGCTGTCCGCG GGGGGACGGCTGCCTTCGGGGGGGACGGGGCAGGGCGGGGTTCGGCTTCTGGCGTGTGACCGGCGGCTCTAGAGCCTC TGCTAACCTTGTTCTTGCCTTCTTCTTTTTCCTACAGCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCATCATTT TGGCAAAGAATTACTCGAGGCCACCATGCAGATTGAACTGTCCACTTGCTTCTTCCTGTGCCTCCTGCGGTTTTGCTT CTCGGCCACCCGCCGGTATTACTTAGGTGCTGTGGAACTGAGCTGGGACTACATGCAGTCCGACCTGGGAGAACTGCC GGTGGACGCGAGATTCCCACCTAGAGTCCCGAAGTCCTTCCCATTCAACACCTCCGTGGTCTACAAAAAGACCCTGTT CGTGGAGTTCACTGACCACCTTTTCAATATTGCCAAGCCGCGCCCCCCCTGGATGGGCCTGCTTGGTCCTACGATCCA AGCAGAGGTCTACGACACCGTGGTCATCACACTGAAGAACATGGCCTCACACCCCGTGTCGCTGCATGCTGTGGGAGT GTCCTACTGGAAGGCCTCAGAGGGTGCCGAATATGATGACCAGACCAGCCAGAGGGAAAAGGAGGATGACAAAGTGTT CCCGGGTGGCAGCCACACTTACGTGTGGCAAGTGCTGAAGGAAAACGGGCCTATGGCGTCGGACCCCCTATGCCTGAC CTACTCCTACCTGTCCCATGTGGACCTTGTGAAGGATCTCAACTCGGGACTGATCGGCGCCCTCTTGGTGTGCAGAGA AGGCAGCCTGGCGAAGGAAAAGACTCAGACCCTGCACAAGTTCATTCTGTTGTTTGCTGTGTTCGATGAAGGAAAGTC CTGGCACTCAGAAACCAAGAACTCGCTGATGCAGGATAGAGATGCGGCCTCGGCCAGAGCCTGGCCTAAAATGCACAC CGTCAACGGATATGTGAACAGGTCGCTCCCTGGCCTCATCGGCTGCCACAGAAAGTCCGTGTATTGGCATGTGATCGG CATGGGTACTACTCCGGAAGTGCATAGTATCTTTCTGGAGGGCCATACCTTCTTGGTGCGCAACCACAGACAGGCCTC GCTGGAAATCTCGCCTATCACTTTCTTGACTGCGCAGACCCTCCTTATGGACCTTGGACAGTTCCTGCTGTTCTGTCA

	CATCAGCTCCCATCAGCATGATGGGATGGAGGCCTATGTCAAAGTGGACTCCTGCCCTGAGGAGCCACAGCTCCGGAT GAAGAACAATGAGGAAGCGGAGGATTACGACGACGACCTGACTGACAGCGAAATGGACGTCGTGCGATTCGATGACGA CAACAGCCCGTCCTTCATCCAAATTAGATCAGTGGCGAAGAAGCACCCCAAGACCTGGGTGCACTACATTGCCGCCGA GGAAGAGGACTGGGACTACGCGCCGCTGGTGCTGGCGCCAGACGACAGGAGCTACAAGTCCCAGTACCTCAACAACGG GCCGCAGCGCATTGGCAGGAAGTACAAGAAAGTCCGCTTCATGGCCTACACTGATGAAACCTTCAAGACGAGGGAAGC CATCCAGCACGAGTCAGGCATCCTGGGACCGCTCCTTTACGGCGAAGTCGGGGATACCCTGCTCATCATTTTCAAGAA CCAGGCATCGCGGCCCTACAACATCTACCCTCACGGGATCACAGACGTGCGCCCGCTCTACTCCCGCCGGCTGCCCAA GGGAGTGAAGCACCTGAAGGATTTTCCCATCCTGCCGGGAGAAATCTTCAAGTACAAGTGGACCGTGACTGTGGAAGA TGGCCCTACCAAGTCGGACCCTCGCTGTCTGACCCGGTACTATTCCTCGTTTGTGAACATGGAGCGCGACCTGGCCTC GGGGCTGATTGGTCCGCTGCTGATCTGCTACAAGGAGTCCGTGGACCAGCGCGGGAACCAGATCATGTCCGACAAGCG CAACGTGATCCTGTTCTCTGTCTTTGATGAAAACAGATCGTGGTACTTGACTGAGAATATCCAGCGGTTCCTGCCCAA CCCAGCGGGAGTGCAACTGGAGGACCCGGAGTTCCAGGCCTCAAACATTATGCACTCTATCAACGGCTATGTGTTCGA CTCGCTCCAACTGAGCGTGTGCCTGCATGAAGTGGCATACTGGTACATTCTGTCCATCGGAGCCCAGACCGACTTCCT GTCCGTGTTCTTCTCCGGATACACCTTCAAGCATAAGATGGTGTACGAGGACACTCTGACCCTCTTCCCATTTTCCGG AGAAACTGTGTTCATGTCAATGGAAAACCCGGGCTTGTGGATTCTGGGTTGCCATAACTCGGACTTCCGGAATAGAGG GATGACCGCCCTGCTGAAAGTGTCCAGCTGTGACAAGAATACCGGCGATTACTACGAGGACAGCTATGAGGACATCTC CGCTTATCTGCTGTCCAAGAACAACGCCATTGAACCCAGGTCCTTCTCCCAAAACGGTGCACCGACCTCCGAAAGCGC CACCCCAGAGTCAGGACCTGGCTCGGAACCGGCTACCTCGGGCTCAGAGACACCGGGGACTTCCGAGTCCGCAACCCC CGAGAGTGGACCCGGATCCGAACCAGCAACCTCAGGATCAGAAACCCCGGGAACTTCGGAATCCGCCACTCCCGAGTC GGGACCAGGCACCTCCACTGAGCCTTCCGAGGGAAGCGCCCCCGGATCCCCTGCTGGATCCCCTACCAGCACTGAAGA AGGCACCTCAGAATCCGCGACCCCTGAGTCCGGCCCTGGAAGCGAACCCGCCACCTCCGGTTCCGAAACCCCTGGGAC TAGCGAGAGCGCCACTCCGGAATCGGGCCCAGGAAGCCCTGCCGGATCCCCGACCAGCACCGAGGAGGGAAGCCCCGC CGGGTCACCGACTTCCACTGAGGAGGGAGCCTCATCCCCCCCCGTGCTGAAGCGGCATCAAAGAGAGATCACCAGGAC CACTCTCCAGTCCGATCAGGAAGAAATTGACTACGACGATACTATCAGCGTGGAGATGAAGAAGGAGGACTTCGACAT CTACGATGAGGATGAGAACCAGTCCCCTCGGAGCTTTCAGAAGAAAACCCGCCACTACTTCATCGCTGCCGTGGAGCG GCTGTGGGATTACGGGATGTCCAGCTCACCGCATGTGCTGCGGAATAGAGCGCAGTCAGGATCGGTGCCCCAGTTCAA GAAGGTCGTGTTCCAAGAGTTCACCGACGGGTCCTTCACTCAACCCCTGTACCGGGGCGAACTCAACGAACACCTGGG ACTGCTTGGGCCGTATATCAGGGCAGAAGTGGAAGATAACATCATGGTCACCTTCCGCAACCAGGCCTCCCGGCCGTA CAGCTTCTACTCTTCACTGATCTCCTACGAGGAAGATCAGCGGCAGGGAGCCGAGCCCCGGAAGAACTTCGTCAAGCC TAACGAAACTAAGACCTACTTTTGGAAGGTCCAGCATCACATGGCCCCGACCAAAGACGAGTTCGACTGTAAAGCCTG GGCCTACTTCTCCGATGTGGACCTGGAGAAGGACGTGCACTCGGGACTCATTGGCCCGCTCCTTGTGTGCCATACTAA TACCCTGAACCCTGCTCACGGTCGCCAAGTCACAGTGCAGGAGTTCGCCCTCTTCTTCACCATCTTCGATGAAACAAA GTCCTGGTACTTTACTGAGAACATGGAACGCAATTGCAGGGCACCCTGCAACATCCAGATGGAAGATCCCACCTTCAA GGAAAACTACCGGTTTCATGCCATTAACGGCTACATAATGGACACGTTGCCAGGACTGGTCATGGCCCAGGACCAGAG AATCCGGTGGTATCTGCTCTCCATGGGCTCCAACGAAAACATTCACAGCATTCATTTTTCCGGCCATGTGTTCACCGT CCGGAAGAAGGAAGAGTACAAGATGGCTCTGTACAACCTCTACCCTGGAGTGTTCGAGACTGTGGAAATGCTGCCTAG CAAGGCCGGCATTTGGAGAGTGGAATGCCTGATCGGAGAGCATTTGCACGCCGGAATGTCCACCCTGTTTCTTGTGTA CTCCAACAAGTGCCAGACCCCGCTGGGAATGGCCTCAGGTCATATTAGGGATTTCCAGATCACTGCTTCGGGGCAGTA CGGGCAGTGGGCACCTAAGTTGGCCCGGCTGCACTACTCTGGCTCCATCAATGCCTGGTCCACCAAGGAACCCTTCTC CTGGATTAAGGTGGACCTCCTGGCCCCAATGATTATTCACGGTATTAAGACCCAGGGTGCCCGACAGAAGTTCTCCTC ACTCTACATCTCGCAATTCATCATAATGTACAGCCTGGATGGGAAGAAGTGGCAGACCTACCGGGGAAACTCCACTGG AACGCTCATGGTGTTTTTCGGCAACGTGGACTCCTCCGGCATTAAGCACAACATCTTCAACCCTCCGATCATTGCTCG GTACATCCGGCTGCACCCAACTCACTACAGCATCCGGTCCACCCTGCGGATGGAACTGATGGGTTGTGACCTGAACTC CTGCTCCATGCCCCTTGGGATGGAATCCAAGGCCATTAGCGATGCACAGATCACCGCCTCTTCATACTTCACCAACAT GTTCGCGACCTGGTCCCCGTCGAAGGCCCGCCTGCACCTCCAAGGTCGCTCCAATGCGTGGCGGCCTCAAGTGAACAA CCCCAAGGAGTGGCTCCAGGTCGACTTCCAAAAGACCATGAAGGTCACCGGAGTGACCACCCAGGGCGTGAAGTCCCT GCTGACCTCTATGTACGTTAAGGAGTTCCTCATCTCCTCAAGCCAAGACGGACATCAGTGGACCCTGTTCTTCCAAAA CGGAAAAGTCAAAGTATTCCAGGGCAACCAGGACTCCTTCACCCCTGTGGTCAACAGCCTGGACCCCCCATTGCTGAC CCGCTACCTCCGCATCCACCCCCAAAGCTGGGTCCACCAGATCGCACTGCGCATGGAGGTCCTTGGATGCGAAGCCCA AGATCTGTACTAAGCGGCCGCTCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTA TGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCAT TTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGT GTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGC TTTCCCCCTCCCTATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGG CACTGACAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCT GCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCT GCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCTGCCTAGGCG ACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCC ACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTG GGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGAAGACCATGGGCGCGCCAGGCCTGTC GACGCCCGGGCGGTACCGCGATCGCTCGCGACGCATAAAGTATATGTGACGTGGTTGTACAGACGCCATCTTGGAATC

	CAATATGTCTGCCGGCGATTAGATCATGCGCGCGCGCAGCGCGCTGCGCGCAGCGCAGGCATGACTGAGCCGGCAGAC ATATTGGATTCCAAGATGGCGTCTGTACAACCACGTGCTTAAGCTGCAGACTAGTGAGCTCGTTAAC
SEQ ID NO. 4: Sf codon optimized HBoV1 NS1	GCGGCCGCGGATCCGCCACCATGGCATTCAATCCGCCCGTAATACGCGCATTTTCACAACCCGCCTTTACGTATGTCT TTAAGTTTCCGTACCCTCAATGGAAAGAGAAAGAGTGGCTACTGCACGCGTTGCTTGCCCACGGCACCGAGCAGTCCA TGATTCAATTACGTAACTGTGCCCCACACCCGGACGAGGATATTATCCGGGACGATCTTCTAATAAGTTTGGAAGATA GGCATTTCGGGGCGGTCCTGTGTAAAGCGGTATACATGGCTACTACCACGTTGATGTCTCACAAGCAACGCAATATGT TCCCAAGGTGCGACATAATCGTTCAGTCAGAGTTAGGTGAAAAAAATTTACATTGTCATATTATCGTTGGAGGCGAAG GCCTATCAAAGAGAAACGCTAAGAGCTCTTGCGCTCAGTTTTACGGACTTATATTAGCAGAAATTATCCAGCGCTGTA AGAGTTTACTAGCCACCCGTCCGTTTGAGCCGGAAGAAGCGGATATATTTCATACGTTGAAGAAAGCGGAGCGCGAGG CCTGGGGTGGAGTTACTGGCGGTAACATGCAAATCTTACAATACAGGGACCGTCGGGGTGACCTGCATGCACAGACTG TTGATCCCCTCAGATTCTTCAAAAATTATTTGTTACCGAAGAACCGATGCATAAGTAGTTACAGCAAACCTGATGTCT GTACTAGCCCTGATAACTGGTTCATTCTGGCCGAAAAAACGTACTCGCATACACTTATCAATGGATTGCCGCTTCCCG AGCACTATCGAAAAAACTATCATGCCACCCTGGATAATGAAGTTATACCTGGACCACAGACTATGGCGTATGGAGGGA GAGGCCCTTGGGAACATTTACCCGAGGTGGGTGACCAGAGGCTTGCCGCAAGTTCCGTGAGCACTACGTATAAGCCAA ACAAGAAGGAGAAGCTAATGCTCAACCTCCTCGACAAGTGTAAGGAGTTGAATCTTCTAGTTTATGAGGATCTTGTAG CGAACTGCCCAGAGCTGCTGCTCATGCTAGAAGGCCAACCTGGAGGTGCTCGACTCATCGAGCAAGTACTAGGAATGC ATCACATCAATGTATGCTCGAATTTCACCGCGCTAACGTACCTCTTCCATCTGCATCCGGTGACATCGCTGGATAGTG ACAACAAAGCGTTACAGCTTTTACTAATTCAAGGGTACAACCCCCTGGCAGTGGGGCATGCTCTCTGTTGTGTGTTAA ACAAACAATTTGGTAAACAGAACACAGTCTGTTTTTACGGGCCAGCATCTACTGGGAAAACAAATATGGCAAAAGCGA TTGTGCAGGGAATCCGGCTATATGGCTGCGTCAACCATCTTAACAAGGGTTTTGTTTTCAATGATTGTCGACAACGCC TCGTAGTCTGGTGGGAGGAATGCCTAATGCACCAGGACTGGGTGGAGCCAGCAAAGTGTATTCTTGGCGGGACCGAAT GTCGTATCGACGTCAAGCACAGAGATTCTGTCCTATTGACACAAACGCCTGTAATAATTTCGACTAATCACGACATTT ACGCCGTCGTGGGAGGGAATTCGGTGTCTCACGTTCACGCTGCGCCTCTCAAAGAACGGGTTATTCAGCTGAATTTTA TGAAACAACTCCCCCAAACTTTTGGTGAGATAACCGCCACAGAAATCGCTGCTCTGCTACAGTGGTGCTTTAATGAAT ATGACTGCACCCTGACAGGTTTCAAACAGAAGTGGAATTTGGACAAGATACCTAACTCATTCCCGTTGGGGGTATTGT GCCCAACACATTCCCAAGATTTCACACTTCACGAAAATGGGTATTGCACGGACTGCGGGGGCTACCTTCCCCACTCCG CTGATAATTCAATGTATACCGATCGGGCTAGCGAAACATCCACCGGCGACATAACGCCCTCCAAATGATTCGAATCTA GAGCCTGCAGTCTCGAGGCATGCGGTACC
SEQ ID NO. 5: Outside Primer	GTGGACGTGAAAGAAACC
SEQ ID NO. 6: Inside Primer	GGTCATAGCTGTTTCCTGTG
SEQ ID NO. 7: hr5.ie1.neo.p10PAS	ATTAAGCTTCCGCGTAAAACACAATCAAGTATGAGTCATAAGCTGATGTCATGTTTTGCACACGGCTCATAACCGAAC TGGCTTTACGAGTAGAATTCTACTTGTAACGCACGATCAGTGGATGATGTCATTTGTTTTTCAAATCGAGATGATGTC ATGTTTTGCACACGGCTCATAAACTCGCTTTACGGGTAGAATTCTACGTGTAACGCACGATCGATTGATGAGTCATTT GTTTTGCAATATGATATCATACAATATGACTCATTTGTTTTTCAAAACCGAACTTGATTTACGGGTAGAATTCTACTT GTAAAGCACAATCAAAAAGATGATGTCATTTGTTTTTCAAAACTGAACTCGCTTTACGAGTAGAATTCTACGTGTAAA ACACAATCAAGAAATGATGTCATTTGTTATAAAAATAAAAGCTGATGTCATGTTTTGCACATGGCTCATAACTAAACT CGCTTTACGGGTAGAATTCTACGCGCGTCGATGTCTTTGTGATGCGCGCGACATTTTTGTAGGTTATTGATAAAATGA ACGGATACGTTGCCCGACATTATCATTAAATCCTTGGCGTAGAATTTGTCGGGTCCATTGTCCGTGTGCGCTAGCATG CCCGTAACGGACCTCGTACTTTTGGCTTCAAAGGTTTTGCGCACAGACAAAATGTGCCACACTTGCAGCTCTGCATGT GTGCGCGTTACCACAAATCCCAACGGCGCAGTGTACTTGTTGTATGCAAATAAATCTCGATAAAGGCGCGGCGCGCGA ATGCAGCTGATCACGTACGCTCCTCGTGTTCCGTTCAAGGACGGTGTTATCGACCTCAGATTAATGTTTATCGGCCGA CTGTTTTCGTATCCGCTCACCAAACGCGTTTTTGCATTAACATTGTATGTCGGCGGATGTTCTATATCTAATTTGAAT AAATAAACGATAACCGCGTTGGTTTTAGAGGGCATAATAAAAGAAATATTGTTATCGTGTTCGCCATTAGGGCAGTAT AAATTGACGTTCATGTTGGATATTGTTTCAGTTGCAAGTTGACACTGGCGGCGACAAGATCGTGAACAACCAAGTGAC GCGGCCGCATTTGTAAAAAAAAAATAAATAAAAATGATCGAGCAGGACGGCCTGCACGCTGGTTCTCCAGCTGCTTGG GTCGAGCGTCTGTTCGGTTACGACTGGGCTCAGCAGACCATCGGTTGCTCCGACGCTGCTGTGTTCCGTCTGTCCGCT CAGGGTCGTCCCGTGCTGTTCGTCAAGACCGACCTGTCCGGTGCTCTGAACGAGCTGCAGGACGAGGCTGCTCGTCTG TCCTGGCTGGCTACCACTGGTGTCCCTTGCGCTGCTGTCCTGGACGTGGTCACTGAGGCTGGTCGTGACTGGCTGCTG CTGGGAGAAGTGCCTGGACAGGACCTGCTGTCCAGCCACCTGGCTCCAGCTGAGAAGGTGTCCATCATGGCTGACGCT ATGCGTCGTCTGCACACCCTGGACCCTGCTACCTGCCCCTTCGACCACCAAGCTAAGCACCGTATCGAGCGTGCTCGT ACCCGTATGGAAGCTGGCCTGGTGGACCAGGACGACCTGGACGAAGAACACCAGGGACTGGCCCCTGCTGAGCTGTTC GCTCGTCTGAAGGCTCGTATGCCCGACGGCGAGGACCTGGTGGTTACTCACGGCGACGCTTGCCTGCCCAACATCATG

	GTCGAGAACGGTCGTTTCTCCGGTTTCATCGACTGCGGTCGTCTGGGTGTCGCTGACCGTTACCAGGATATCGCTCTG GCTACCCGTGATATCGCTGAGGAACTGGGTGGCGAGTGGGCTGACAGATTCCTGGTGCTGTACGGTATCGCTGCTCCC GACTCCCAGCGTATCGCTTTCTACCGTCTGCTGGACGAGTTCTTCTAAGCCCCTTGTAAACGCCACAATTGTGTTTGT TGCAAATAAACCCATGATTATTTGATTAAAATTGTTGTTTTCTTTGTTCATAGACAATAGTGTGTTTTGCCTAAACGG GTACC
SEQ ID NO. 8: hr5.ie1.eGFP. p10PAS	ATTAAGCTTCCGCGTAAAACACAATCAAGTATGAGTCATAAGCTGATGTCATGTTTTGCACACGGCTCATAACCGAAC TGGCTTTACGAGTAGAATTCTACTTGTAACGCACGATCAGTGGATGATGTCATTTGTTTTTCAAATCGAGATGATGTC ATGTTTTGCACACGGCTCATAAACTCGCTTTACGGGTAGAATTCTACGTGTAACGCACGATCGATTGATGAGTCATTT GTTTTGCAATATGATATCATACAATATGACTCATTTGTTTTTCAAAACCGAACTTGATTTACGGGTAGAATTCTACTT GTAAAGCACAATCAAAAAGATGATGTCATTTGTTTTTCAAAACTGAACTCGCTTTACGAGTAGAATTCTACGTGTAAA ACACAATCAAGAAATGATGTCATTTGTTATAAAAATAAAAGCTGATGTCATGTTTTGCACATGGCTCATAACTAAACT CGCTTTACGGGTAGAATTCTACGCGCGTCGATGTCTTTGTGATGCGCGCGACATTTTTGTAGGTTATTGATAAAATGA ACGGATACGTTGCCCGACATTATCATTAAATCCTTGGCGTAGAATTTGTCGGGTCCATTGTCCGTGTGCGCTAGCATG CCCGTAACGGACCTCGTACTTTTGGCTTCAAAGGTTTTGCGCACAGACAAAATGTGCCACACTTGCAGCTCTGCATGT GTGCGCGTTACCACAAATCCCAACGGCGCAGTGTACTTGTTGTATGCAAATAAATCTCGATAAAGGCGCGGCGCGCGA ATGCAGCTGATCACGTACGCTCCTCGTGTTCCGTTCAAGGACGGTGTTATCGACCTCAGATTAATGTTTATCGGCCGA CTGTTTTCGTATCCGCTCACCAAACGCGTTTTTGCATTAACATTGTATGTCGGCGGATGTTCTATATCTAATTTGAAT AAATAAACGATAACCGCGTTGGTTTTAGAGGGCATAATAAAAGAAATATTGTTATCGTGTTCGCCATTAGGGCAGTAT AAATTGACGTTCATGTTGGATATTGTTTCAGTTGCAAGTTGACACTGGCGGCGACAAGATCGTGAACAACCAAGTGAC GCGGCCGCATTTGTAAAAAAAAAATAAATAAAAATGGTGTCCAAGGGCGAGGAACTGTTCACCGGTGTCGTGCCCATC CTGGTCGAACTGGACGGCGACGTGAACGGTCACAAGTTCTCCGTGTCTGGCGAAGGCGAGGGCGACGCTACCTACGGA AAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCTTGGCCTACCCTGGTCACCACTCTGACCTAC GGTGTCCAGTGCTTCTCCCGTTACCCCGACCACATGAAGCAGCACGATTTCTTCAAGTCCGCTATGCCCGAGGGTTAC GTGCAAGAGCGTACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGTGCTGAAGTGAAGTTCGAAGGCGACACC CTCGTGAACCGTATCGAGCTGAAGGGTATCGACTTCAAGGAAGATGGAAACATCCTGGGCCACAAGCTCGAGTACAAC TACAACTCCCACAACGTGTACATCATGGCCGACAAGCAAAAGAACGGCATCAAAGTGAACTTCAAGATCCGCCACAAC ATCGAGGACGGTTCCGTGCAGCTGGCTGACCACTACCAGCAGAACACCCCCATCGGCGACGGTCCTGTGCTGCTGCCT GACAACCACTACCTGTCCACCCAGTCCGCTCTGTCCAAGGACCCCAACGAGAAGCGTGACCACATGGTGCTGCTCGAG TTCGTGACCGCTGCTGGTATCACCCTGGGCATGGACGAGCTGTACAAGTAAGCCCCTTGTAAACGCCACAATTGTGTT TGTTGCAAATAAACCCATGATTATTTGATTAAAATTGTTGTTTTCTTTGTTCATAGACAATAGTGTGTTTTGCCTAAA CGGGTACC
SEQ ID NO:9 Nucleotide sequence encoding coBDDFVIIIX TEN (V2.0)	ATGCAGATTGAACTGTCCACTTGCTTCTTCCTGTGCCTCCTGCGGTTTTGCTTCTCGGCCACCCGCCGGTATTACTTA GGTGCTGTGGAACTGAGCTGGGACTACATGCAGTCCGACCTGGGAGAACTGCCGGTGGACGCGAGATTCCCACCTAGA GTCCCGAAGTCCTTCCCATTCAACACCTCCGTGGTCTACAAAAAGACCCTGTTCGTGGAGTTCACTGACCACCTTTTC AATATTGCCAAGCCGCGCCCCCCCTGGATGGGCCTGCTTGGTCCTACGATCCAAGCAGAGGTCTACGACACCGTGGTC ATCACACTGAAGAACATGGCCTCACACCCCGTGTCGCTGCATGCTGTGGGAGTGTCCTACTGGAAGGCCTCAGAGGGT GCCGAATATGATGACCAGACCAGCCAGAGGGAAAAGGAGGATGACAAAGTGTTCCCGGGTGGCAGCCACACTTACGTG TGGCAAGTGCTGAAGGAAAACGGGCCTATGGCGTCGGACCCCCTATGCCTGACCTACTCCTACCTGTCCCATGTGGAC CTTGTGAAGGATCTCAACTCGGGACTGATCGGCGCCCTCTTGGTGTGCAGAGAAGGCAGCCTGGCGAAGGAAAAGACT CAGACCCTGCACAAGTTCATTCTGTTGTTTGCTGTGTTCGATGAAGGAAAGTCCTGGCACTCAGAAACCAAGAACTCG CTGATGCAGGATAGAGATGCGGCCTCGGCCAGAGCCTGGCCTAAAATGCACACCGTCAACGGATATGTGAACAGGTCG CTCCCTGGCCTCATCGGCTGCCACAGAAAGTCCGTGTATTGGCATGTGATCGGCATGGGTACTACTCCGGAAGTGCAT AGTATCTTTCTGGAGGGCCATACCTTCTTGGTGCGCAACCACAGACAGGCCTCGCTGGAAATCTCGCCTATCACTTTC TTGACTGCGCAGACCCTCCTTATGGACCTTGGACAGTTCCTGCTGTTCTGTCACATCAGCTCCCATCAGCATGATGGG ATGGAGGCCTATGTCAAAGTGGACTCCTGCCCTGAGGAGCCACAGCTCCGGATGAAGAACAATGAGGAAGCGGAGGAT TACGACGACGACCTGACTGACAGCGAAATGGACGTCGTGCGATTCGATGACGACAACAGCCCGTCCTTCATCCAAATT AGATCAGTGGCGAAGAAGCACCCCAAGACCTGGGTGCACTACATTGCCGCCGAGGAAGAGGACTGGGACTACGCGCCG CTGGTGCTGGCGCCAGACGACAGGAGCTACAAGTCCCAGTACCTCAACAACGGGCCGCAGCGCATTGGCAGGAAGTAC AAGAAAGTCCGCTTCATGGCCTACACTGATGAAACCTTCAAGACGAGGGAAGCCATCCAGCACGAGTCAGGCATCCTG GGACCGCTCCTTTACGGCGAAGTCGGGGATACCCTGCTCATCATTTTCAAGAACCAGGCATCGCGGCCCTACAACATC TACCCTCACGGGATCACAGACGTGCGCCCGCTCTACTCCCGCCGGCTGCCCAAGGGAGTGAAGCACCTGAAGGATTTT CCCATCCTGCCGGGAGAAATCTTCAAGTACAAGTGGACCGTGACTGTGGAAGATGGCCCTACCAAGTCGGACCCTCGC TGTCTGACCCGGTACTATTCCTCGTTTGTGAACATGGAGCGCGACCTGGCCTCGGGGCTGATTGGTCCGCTGCTGATC TGCTACAAGGAGTCCGTGGACCAGCGCGGGAACCAGATCATGTCCGACAAGCGCAACGTGATCCTGTTCTCTGTCTTT GATGAAAACAGATCGTGGTACTTGACTGAGAATATCCAGCGGTTCCTGCCCAACCCAGCGGGAGTGCAACTGGAGGAC CCGGAGTTCCAGGCCTCAAACATTATGCACTCTATCAACGGCTATGTGTTCGACTCGCTCCAACTGAGCGTGTGCCTG CATGAAGTGGCATACTGGTACATTCTGTCCATCGGAGCCCAGACCGACTTCCTGTCCGTGTTCTTCTCCGGATACACC TTCAAGCATAAGATGGTGTACGAGGACACTCTGACCCTCTTCCCATTTTCCGGAGAAACTGTGTTCATGTCAATGGAA AACCCGGGCTTGTGGATTCTGGGTTGCCATAACTCGGACTTCCGGAATAGAGGGATGACCGCCCTGCTGAAAGTGTCC

	AGCTGTGACAAGAATACCGGCGATTACTACGAGGACAGCTATGAGGACATCTCCGCTTATCTGCTGTCCAAGAACAAC GCCATTGAACCCAGGTCCTTCTCCCAAAACGGTGCACCGACCTCCGAAAGCGCCACCCCAGAGTCAGGACCTGGCTCG GAACCGGCTACCTCGGGCTCAGAGACACCGGGGACTTCCGAGTCCGCAACCCCCGAGAGTGGACCCGGATCCGAACCA GCAACCTCAGGATCAGAAACCCCGGGAACTTCGGAATCCGCCACTCCCGAGTCGGGACCAGGCACCTCCACTGAGCCT TCCGAGGGAAGCGCCCCCGGATCCCCTGCTGGATCCCCTACCAGCACTGAAGAAGGCACCTCAGAATCCGCGACCCCT GAGTCCGGCCCTGGAAGCGAACCCGCCACCTCCGGTTCCGAAACCCCTGGGACTAGCGAGAGCGCCACTCCGGAATCG GGCCCAGGAAGCCCTGCCGGATCCCCGACCAGCACCGAGGAGGGAAGCCCCGCCGGGTCACCGACTTCCACTGAGGAG GGAGCCTCATCCCCCCCCGTGCTGAAGCGGCATCAAAGAGAGATCACCAGGACCACTCTCCAGTCCGATCAGGAAGAA ATTGACTACGACGATACTATCAGCGTGGAGATGAAGAAGGAGGACTTCGACATCTACGATGAGGATGAGAACCAGTCC CCTCGGAGCTTTCAGAAGAAAACCCGCCACTACTTCATCGCTGCCGTGGAGCGGCTGTGGGATTACGGGATGTCCAGC TCACCGCATGTGCTGCGGAATAGAGCGCAGTCAGGATCGGTGCCCCAGTTCAAGAAGGTCGTGTTCCAAGAGTTCACC GACGGGTCCTTCACTCAACCCCTGTACCGGGGCGAACTCAACGAACACCTGGGACTGCTTGGGCCGTATATCAGGGCA GAAGTGGAAGATAACATCATGGTCACCTTCCGCAACCAGGCCTCCCGGCCGTACAGCTTCTACTCTTCACTGATCTCC TACGAGGAAGATCAGCGGCAGGGAGCCGAGCCCCGGAAGAACTTCGTCAAGCCTAACGAAACTAAGACCTACTTTTGG AAGGTCCAGCATCACATGGCCCCGACCAAAGACGAGTTCGACTGTAAAGCCTGGGCCTACTTCTCCGATGTGGACCTG GAGAAGGACGTGCACTCGGGACTCATTGGCCCGCTCCTTGTGTGCCATACTAATACCCTGAACCCTGCTCACGGTCGC CAAGTCACAGTGCAGGAGTTCGCCCTCTTCTTCACCATCTTCGATGAAACAAAGTCCTGGTACTTTACTGAGAACATG GAACGCAATTGCAGGGCACCCTGCAACATCCAGATGGAAGATCCCACCTTCAAGGAAAACTACCGGTTTCATGCCATT AACGGCTACATAATGGACACGTTGCCAGGACTGGTCATGGCCCAGGACCAGAGAATCCGGTGGTATCTGCTCTCCATG GGCTCCAACGAAAACATTCACAGCATTCATTTTTCCGGCCATGTGTTCACCGTCCGGAAGAAGGAAGAGTACAAGATG GCTCTGTACAACCTCTACCCTGGAGTGTTCGAGACTGTGGAAATGCTGCCTAGCAAGGCCGGCATTTGGAGAGTGGAA TGCCTGATCGGAGAGCATTTGCACGCCGGAATGTCCACCCTGTTTCTTGTGTACTCCAACAAGTGCCAGACCCCGCTG GGAATGGCCTCAGGTCATATTAGGGATTTCCAGATCACTGCTTCGGGGCAGTACGGGCAGTGGGCACCTAAGTTGGCC CGGCTGCACTACTCTGGCTCCATCAATGCCTGGTCCACCAAGGAACCCTTCTCCTGGATTAAGGTGGACCTCCTGGCC CCAATGATTATTCACGGTATTAAGACCCAGGGTGCCCGACAGAAGTTCTCCTCACTCTACATCTCGCAATTCATCATA ATGTACAGCCTGGATGGGAAGAAGTGGCAGACCTACCGGGGAAACTCCACTGGAACGCTCATGGTGTTTTTCGGCAAC GTGGACTCCTCCGGCATTAAGCACAACATCTTCAACCCTCCGATCATTGCTCGGTACATCCGGCTGCACCCAACTCAC TACAGCATCCGGTCCACCCTGCGGATGGAACTGATGGGTTGTGACCTGAACTCCTGCTCCATGCCCCTTGGGATGGAA TCCAAGGCCATTAGCGATGCACAGATCACCGCCTCTTCATACTTCACCAACATGTTCGCGACCTGGTCCCCGTCGAAG GCCCGCCTGCACCTCCAAGGTCGCTCCAATGCGTGGCGGCCTCAAGTGAACAACCCCAAGGAGTGGCTCCAGGTCGAC TTCCAAAAGACCATGAAGGTCACCGGAGTGACCACCCAGGGCGTGAAGTCCCTGCTGACCTCTATGTACGTTAAGGAG TTCCTCATCTCCTCAAGCCAAGACGGACATCAGTGGACCCTGTTCTTCCAAAACGGAAAAGTCAAAGTATTCCAGGGC AACCAGGACTCCTTCACCCCTGTGGTCAACAGCCTGGACCCCCCATTGCTGACCCGCTACCTCCGCATCCACCCCCAA AGCTGGGTCCACCAGATCGCACTGCGCATGGAGGTCCTTGGATGCGAAGCCCAAGATCTGTACTAA
SEQ ID NO: 10 Amino acid sequence of coBDDFVIIIX TEN (V2.0)	ATRRYYLGAVELSWDYMQSDLGELPVDARFPPRVPKSFPFNTSVVYKKTLFVEFTDHLFNIAKPRPPWMGLLGPTIQA EVYDTVVITLKNMASHPVSLHAVGVSYWKASEGAEYDDQTSQREKEDDKVFPGGSHTYVWQVLKENGPMASDPLCLTY SYLSHVDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVFDEGKSWHSETKNSLMQDRDAASARAWPKMHTV NGYVNRSLPGLIGCHRKSVYWHVIGMGTTPEVHSIFLEGHTFLVRNHRQASLEISPITFLTAQTLLMDLGQFLLFCHI SSHQHDGMEAYVKVDSCPEEPQLRMKNNEEAEDYDDDLTDSEMDVVRFDDDNSPSFIQIRSVAKKHPKTWVHYIAAEE EDWDYAPLVLAPDDRSYKSQYLNNGPQRIGRKYKKVRFMAYTDETFKTREAIQHESGILGPLLYGEVGDTLLIIFKNQ ASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRYYSSFVNMERDLASG LIGPLLICYKESVDQRGNQIMSDKRNVILFSVFDENRSWYLTENIQRFLPNPAGVQLEDPEFQASNIMHSINGYVFDS LQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKHKMVYEDTLTLFPFSGETVFMSMENPGLWILGCHNSDFRNRGM TALLKVSSCDKNTGDYYEDSYEDISAYLLSKNNAIEPRSFSQNGAPTSESATPESGPGSEPATSGSETPGTSESATPE SGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTS ESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGASSPPVLKRHQREITRTTLQSDQEEIDYDDTISVEMKKEDFDIY DEDENQSPRSFQKKTRHYFIAAVERLWDYGMSSSPHVLRNRAQSGSVPQFKKVVFQEFTDGSFTQPLYRGELNEHLGL LGPYIRAEVEDNIMVTFRNQASRPYSFYSSLISYEEDQRQGAEPRKNFVKPNETKTYFWKVQHHMAPTKDEFDCKAWA YFSDVDLEKDVHSGLIGPLLVCHTNTLNPAHGRQVTVQEFALFFTIFDETKSWYFTENMERNCRAPCNIQMEDPTFKE NYRFHAINGYIMDTLPGLVMAQDQRIRWYLLSMGSNENIHSIHFSGHVFTVRKKEEYKMALYNLYPGVFETVEMLPSK AGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPLGMASGHIRDFQITASGQYGQWAPKLARLHYSGSINAWSTKEPFSW IKVDLLAPMIIHGIKTQGARQKFSSLYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKHNIFNPPIIARY IRLHPTHYSIRSTLRMELMGCDLNSCSMPLGMESKAISDAQITASSYFTNMFATWSPSKARLHLQGRSNAWRPQVNNP KEWLQVDFQKTMKVTGVTTQGVKSLLTSMYVKEFLISSSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVNSLDPPLLTR YLRIHPQSWVHQIALRMEVLGCEAQDLY
SEQ ID NO: 11 Signal peptide of	MQIELSTCFFLCLLRFCFS

coBDDFVIIIXTEN (V2.0)	
SEQ ID NO: 12 Amino acid sequence of BDD mature human FVIII	ATRRYYLGAVELSWDYMQSDLGELPVDARFPPRVPKSFPFNTSVVYKKTLFVEFTDHLFNIAKPRPPWMGLLGPTIQA EVYDTVVITLKNMASHPVSLHAVGVSYWKASEGAEYDDQTSQREKEDDKVFPGGSHTYVWQVLKENGPMASDPLCLTY SYLSHVDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVFDEGKSWHSETKNSLMQDRDAASARAWPKMHTV NGYVNRSLPGLIGCHRKSVYWHVIGMGTTPEVHSIFLEGHTFLVRNHRQASLEISPITFLTAQTLLMDLGQFLLFCHI SSHQHDGMEAYVKVDSCPEEPQLRMKNNEEAEDYDDDLTDSEMDVVRFDDDNSPSFIQIRSVAKKHPKTWVHYIAAEE EDWDYAPLVLAPDDRSYKSQYLNNGPQRIGRKYKKVRFMAYTDETFKTREAIQHESGILGPLLYGEVGDTLLIIFKNQ ASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRYYSSFVNMERDLASG LIGPLLICYKESVDQRGNQIMSDKRNVILFSVFDENRSWYLTENIQRFLPNPAGVQLEDPEFQASNIMHSINGYVFDS LQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKHKMVYEDTLTLFPFSGETVFMSMENPGLWILGCHNSDFRNRGM TALLKVSSCDKNTGDYYEDSYEDISAYLLSKNNAIEPRSFSQNPPVLKRHQREITRTTLQSDQEEIDYDDTISVEMKK EDFDIYDEDENQSPRSFQKKTRHYFIAAVERLWDYGMSSSPHVLRNRAQSGSVPQFKKVVFQEFTDGSFTQPLYRGEL NEHLGLLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLISYEEDQRQGAEPRKNFVKPNETKTYFWKVQHHMAPTKDEF DCKAWAYFSDVDLEKDVHSGLIGPLLVCHTNTLNPAHGRQVTVQEFALFFTIFDETKSWYFTENMERNCRAPCNIQME DPTFKENYRFHAINGYIMDTLPGLVMAQDQRIRWYLLSMGSNENIHSIHFSGHVFTVRKKEEYKMALYNLYPGVFETV EMLPSKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPLGMASGHIRDFQITASGQYGQWAPKLARLHYSGSINAWST KEPFSWIKVDLLAPMIIHGIKTQGARQKFSSLYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKHNIFNP PIIARYIRLHPTHYSIRSTLRMELMGCDLNSCSMPLGMESKAISDAQITASSYFTNMFATWSPSKARLHLQGRSNAWR PQVNNPKEWLQVDFQKTMKVTGVTTQGVKSLLTSMYVKEFLISSSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVNSLD PPLLTRYLRIHPQSWVHQIALRMEVLGCEAQDLY
SEQ ID NO: 13 Nucleotide sequence encoding BDD mature human FVIII	ATGCAAATAGAGCTCTCCACCTGCTTCTTTCTGTGCCTTTTGCGATTCTGCTTTAGTGCCACCAGAAGATACTACCTG GGTGCAGTGGAACTGTCATGGGACTATATGCAAAGTGATCTCGGTGAGCTGCCTGTGGACGCAAGATTTCCTCCTAGA GTGCCAAAATCTTTTCCATTCAACACCTCAGTCGTGTACAAAAAGACTCTGTTTGTAGAATTCACGGATCACCTTTTC AACATCGCTAAGCCAAGGCCACCCTGGATGGGTCTGCTAGGTCCTACCATCCAGGCTGAGGTTTATGATACAGTGGTC ATTACACTTAAGAACATGGCTTCCCATCCTGTCAGTCTTCATGCTGTTGGTGTATCCTACTGGAAAGCTTCTGAGGGA GCTGAATATGATGATCAGACCAGTCAAAGGGAGAAAGAAGATGATAAAGTCTTCCCTGGTGGAAGCCATACATATGTC TGGCAGGTCCTGAAAGAGAATGGTCCAATGGCCTCTGACCCACTGTGCCTTACCTACTCATATCTTTCTCATGTGGAC CTGGTAAAAGACTTGAATTCAGGCCTCATTGGAGCCCTACTAGTATGTAGAGAAGGGAGTCTGGCCAAGGAAAAGACA CAGACCTTGCACAAATTTATACTACTTTTTGCTGTATTTGATGAAGGGAAAAGTTGGCACTCAGAAACAAAGAACTCC TTGATGCAGGATAGGGATGCTGCATCTGCTCGGGCCTGGCCTAAAATGCACACAGTCAATGGTTATGTAAACAGGTCT CTGCCAGGTCTGATTGGATGCCACAGGAAATCAGTCTATTGGCATGTGATTGGAATGGGCACCACTCCTGAAGTGCAC TCAATATTCCTCGAAGGTCACACATTTCTTGTGAGGAACCATCGCCAGGCGTCCTTGGAAATCTCGCCAATAACTTTC CTTACTGCTCAAACACTCTTGATGGACCTTGGACAGTTTCTACTGTTTTGTCATATCTCTTCCCACCAACATGATGGC ATGGAAGCTTATGTCAAAGTAGACAGCTGTCCAGAGGAACCCCAACTACGAATGAAAAATAATGAAGAAGCGGAAGAC TATGATGATGATCTTACTGATTCTGAAATGGATGTGGTCAGGTTTGATGATGACAACTCTCCTTCCTTTATCCAAATT CGCTCAGTTGCCAAGAAGCATCCTAAAACTTGGGTACATTACATTGCTGCTGAAGAGGAGGACTGGGACTATGCTCCC TTAGTCCTCGCCCCCGATGACAGAAGTTATAAAAGTCAATATTTGAACAATGGCCCTCAGCGGATTGGTAGGAAGTAC AAAAAAGTCCGATTTATGGCATACACAGATGAAACCTTTAAGACTCGTGAAGCTATTCAGCATGAATCAGGAATCTTG GGACCTTTACTTTATGGGGAAGTTGGAGACACACTGTTGATTATATTTAAGAATCAAGCAAGCAGACCATATAACATC TACCCTCACGGAATCACTGATGTCCGTCCTTTGTATTCAAGGAGATTACCAAAAGGTGTAAAACATTTGAAGGATTTT CCAATTCTGCCAGGAGAAATATTCAAATATAAATGGACAGTGACTGTAGAAGATGGGCCAACTAAATCAGATCCTCGG TGCCTGACCCGCTATTACTCTAGTTTCGTTAATATGGAGAGAGATCTAGCTTCAGGACTCATTGGCCCTCTCCTCATC TGCTACAAAGAATCTGTAGATCAAAGAGGAAACCAGATAATGTCAGACAAGAGGAATGTCATCCTGTTTTCTGTATTT GATGAGAACCGAAGCTGGTACCTCACAGAGAATATACAACGCTTTCTCCCCAATCCAGCTGGAGTGCAGCTTGAGGAT CCAGAGTTCCAAGCCTCCAACATCATGCACAGCATCAATGGCTATGTTTTTGATAGTTTGCAGTTGTCAGTTTGTTTG CATGAGGTGGCATACTGGTACATTCTAAGCATTGGAGCACAGACTGACTTCCTTTCTGTCTTCTTCTCTGGATATACC TTCAAACACAAAATGGTCTATGAAGACACACTCACCCTATTCCCATTCTCAGGAGAAACTGTCTTCATGTCGATGGAA AACCCAGGTCTATGGATTCTGGGGTGCCACAACTCAGACTTTCGGAACAGAGGCATGACCGCCTTACTGAAGGTTTCT AGTTGTGACAAGAACACTGGTGATTATTACGAGGACAGTTATGAAGATATTTCAGCATACTTGCTGAGTAAAAACAAT GCCATTGAACCAAGAAGCTTCTCTCAAAACCCACCAGTCTTGAAACGCCATCAACGGGAAATAACTCGTACTACTCTT CAGTCAGATCAAGAGGAAATTGACTATGATGATACCATATCAGTTGAAATGAAGAAGGAAGATTTTGACATTTATGAT GAGGATGAAAATCAGAGCCCCCGCAGCTTTCAAAAGAAAACACGACACTATTTTATTGCTGCAGTGGAGAGGCTCTGG GATTATGGGATGAGTAGCTCCCCACATGTTCTAAGAAACAGGGCTCAGAGTGGCAGTGTCCCTCAGTTCAAGAAAGTT GTTTTCCAGGAATTTACTGATGGCTCCTTTACTCAGCCCTTATACCGTGGAGAACTAAATGAACATTTGGGACTCCTG GGGCCATATATAAGAGCAGAAGTTGAAGATAATATCATGGTAACTTTCAGAAATCAGGCCTCTCGTCCCTATTCCTTC TATTCTAGCCTTATTTCTTATGAGGAAGATCAGAGGCAAGGAGCAGAACCTAGAAAAAACTTTGTCAAGCCTAATGAA ACCAAAACTTACTTTTGGAAAGTGCAACATCATATGGCACCCACTAAAGATGAGTTTGACTGCAAAGCCTGGGCTTAT TTCTCTGATGTTGACCTGGAAAAAGATGTGCACTCAGGCCTGATTGGACCCCTTCTGGTCTGCCACACTAACACACTG AACCCTGCTCATGGGAGACAAGTGACAGTACAGGAATTTGCTCTGTTTTTCACCATCTTTGATGAGACCAAAAGCTGG

	TACTTCACTGAAAATATGGAAAGAAACTGCAGGGCTCCCTGCAATATCCAGATGGAAGATCCCACTTTTAAAGAGAAT TATCGCTTCCATGCAATCAATGGCTACATAATGGATACACTACCTGGCTTAGTAATGGCTCAGGATCAAAGGATTCGA TGGTATCTGCTCAGCATGGGCAGCAATGAAAACATCCATTCTATTCATTTCAGTGGACATGTGTTCACTGTACGAAAA AAAGAGGAGTATAAAATGGCACTGTACAATCTCTATCCAGGTGTTTTTGAGACAGTGGAAATGTTACCATCCAAAGCT GGAATTTGGCGGGTGGAATGCCTTATTGGCGAGCATCTACATGCTGGGATGAGCACACTTTTTCTGGTGTACAGCAAT AAGTGTCAGACTCCCCTGGGAATGGCTTCTGGACACATTAGAGATTTTCAGATTACAGCTTCAGGACAATATGGACAG TGGGCCCCAAAGCTGGCCAGACTTCATTATTCCGGATCAATCAATGCCTGGAGCACCAAGGAGCCCTTTTCTTGGATC AAGGTGGATCTGTTGGCACCAATGATTATTCACGGCATCAAGACCCAGGGTGCCCGTCAGAAGTTCTCCAGCCTCTAC ATCTCTCAGTTTATCATCATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATTCCACTGGAACCTTA ATGGTCTTCTTTGGCAATGTGGATTCATCTGGGATAAAACACAATATTTTTAACCCTCCAATTATTGCTCGATACATC CGTTTGCACCCAACTCATTATAGCATTCGCAGCACTCTTCGCATGGAGTTGATGGGCTGTGATTTAAATAGTTGCAGC ATGCCATTGGGAATGGAGAGTAAAGCAATATCAGATGCACAGATTACTGCTTCATCCTACTTTACCAATATGTTTGCC ACCTGGTCTCCTTCAAAAGCTCGACTTCACCTCCAAGGGAGGAGTAATGCCTGGAGACCTCAGGTGAATAATCCAAAA GAGTGGCTGCAAGTGGACTTCCAGAAGACAATGAAAGTCACAGGAGTAACTACTCAGGGAGTAAAATCTCTGCTTACC AGCATGTATGTGAAGGAGTTCCTCATCTCCAGCAGTCAAGATGGCCATCAGTGGACTCTCTTTTTTCAGAATGGCAAA GTAAAGGTTTTTCAGGGAAATCAAGACTCCTTCACACCTGTGGTGAACTCTCTAGACCCACCGTTACTGACTCGCTAC CTTCGAATTCACCCCCAGAGTTGGGTGCACCAGATTGCCCTGAGGATGGAGGTTCTGGGCTGCGAGGCACAGGACCTC TAC
SEQ ID NO: 14 V2.0 Expression cassette mTTR482-Intron-coBDDFVIIIXTEN (V2.0)-WPRE-bGHPolyA	GGCCCCAGGTTAATTTTTAAAAAGCAGTCAAAGGTCAAAGTGGCCCTTGGCAGCATTTACTCTCTCTATTGACTTTGG TTAATAATCTCAGGAGCACAAACATTCCTGGAGGCAGGAGAAGAAATCAACATCCTGGACTTATCCTCTGGGCCTCTC CCCACCTTCGATGGCCCCAGGTTAATTTTTAAAAAGCAGTCAAAGGTCAAAGTGGCCCTTGGCAGCATTTACTCTCTC TATTGACTTTGGTTAATAATCTCAGGAGCACAAACATTCCTGGAGGCAGGAGAAGAAATCAACATCCTGGACTTATCC TCTGGGCCTCTCCCCACCGATATCTACCTGCTGATCGCCCGGCCCCTGTTCAAACATGTCCTAATACTCTGTCGGGGC AAAGGTCGGCAGTAGTTTTCCATCTTACTCAACATCCTCCCAGTGTACGTAGGATCCTGTCTGTCTGCACATTTCGTA GAGCGAGTGTTCCGATACTCTAATCTCCCGGGGCAAAGGTCGTATTGACTTAGGTTACTTATTCTCCTTTTGTTGACT AAGTCAATAATCAGAATCAGCAGGTTTGGAGTCAGCTTGGCAGGGATCAGCAGCCTGGGTTGGAAGGAGGGGGTATAA AAGCCCCTTCACCAGGAGAAGCCGTCACACAGATCCACAAGCTCCTGCTAGGAATTCTCAGGAGCACAAACATTCCTG GAGGCAGGAGAAGAAATCAACATCCTGGACTTATCCTCTGGGCCTCTCCCCACCGATATCTACCTGCTGATCGCCCGG CCCCTGTTCAAACATGTCCTAATACTCTGTCGGGGCAAAGGTCGGCAGTAGTTTTCCATCTTACTCAACATCCTCCCA GTGTACGTAGGATCCTGTCTGTCTGCACATTTCGTAGAGCGAGTGTTCCGATACTCTAATCTCCCGGGGCAAAGGTCG TATTGACTTAGGTTACTTATTCTCCTTTTGTTGACTAAGTCAATAATCAGAATCAGCAGGTTTGGAGTCAGCTTGGCA GGGATCAGCAGCCTGGGTTGGAAGGAGGGGGTATAAAAGCCCCTTCACCAGGAGAAGCCGTCACACAGATCCACAAGC TCCTGCTAGAGTCGCTGCGCGCTGCCTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTCGCGCCGCCCGCCCCGGCTCTG ACTGACCGCGTTACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCGCTTGGTTTATTG ACGGCTTGTTTCTTTTCTGTGGCTGCGTGAAAGCCTTGAGGGGCTCCGGGAAGGCCCTTTGTGCGGGGGGAGCGGCTC GGGGGGTGCGTGCGTGTGTGTGTGCGTGGGGAGCGCCGCGTGCGGCTCCGCGCTGCCCGGCGGCTGTGAGCGCTGCGG GCGCGGCGCGGGGCTTTGTGCGCTCCGCAGTGTGCGCGAGGGGAGCGCGGCCGGGGGCGGTGCCCCGCGGTGCGGGGG GGGCTGCGAGGGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGGGGGTGAGCAGGGGGTGTGGGCGCGTCGGTCGG GCTGCAACCCCCCCTGCACCCCCCTCCCCGAGTTGCTGAGCACGGCCCGGCTTCGGGTGCGGGGCTCCGTACGGGGCG TGGCGCGGGGCTCGCCGTGCCGGGCGGGGGGTGGCGGCAGGTGGGGGTGCCGGGCGGGGCGGGGCCGCCTCGGGCCGG GGAGGGCTCGGGGGAGGGGCGCGGCGGCCCCCGGAGCGCCGGCGGCTGTCGAGGCGCGGCGAGCCGCAGCCATTGCCT TTTATGGTAATCGTGCGAGAGGGCGCAGGGACTTCCTTTGTCCCAAATCTGTGCGGAGCCGAAATCTGGGAGGCGCCG CCGCACCCCCTCTAGCGGGCGCGGGGCGAAGCGGTGCGGCGCCGGCAGGAAGGAAATGGGCGGGGAGGGCCTTCGTGC GTCGCCGCGCCGCCGTCCCCTTCTCCCTCTCCAGCCTCGGGGCTGTCCGCGGGGGGACGGCTGCCTTCGGGGGGGACG GGGCAGGGCGGGGTTCGGCTTCTGGCGTGTGACCGGCGGCTCTAGAGCCTCTGCTAACCTTGTTCTTGCCTTCTTCTT TTTCCTACAGCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCATCATTTTGGCAAAGAATTACTCGAGGCCACCAT GCAGATTGAACTGTCCACTTGCTTCTTCCTGTGCCTCCTGCGGTTTTGCTTCTCGGCCACCCGCCGGTATTACTTAGG TGCTGTGGAACTGAGCTGGGACTACATGCAGTCCGACCTGGGAGAACTGCCGGTGGACGCGAGATTCCCACCTAGAGT CCCGAAGTCCTTCCCATTCAACACCTCCGTGGTCTACAAAAAGACCCTGTTCGTGGAGTTCACTGACCACCTTTTCAA TATTGCCAAGCCGCGCCCCCCCTGGATGGGCCTGCTTGGTCCTACGATCCAAGCAGAGGTCTACGACACCGTGGTCAT CACACTGAAGAACATGGCCTCACACCCCGTGTCGCTGCATGCTGTGGGAGTGTCCTACTGGAAGGCCTCAGAGGGTGC CGAATATGATGACCAGACCAGCCAGAGGGAAAAGGAGGATGACAAAGTGTTCCCGGGTGGCAGCCACACTTACGTGTG GCAAGTGCTGAAGGAAAACGGGCCTATGGCGTCGGACCCCCTATGCCTGACCTACTCCTACCTGTCCCATGTGGACCT TGTGAAGGATCTCAACTCGGGACTGATCGGCGCCCTCTTGGTGTGCAGAGAAGGCAGCCTGGCGAAGGAAAAGACTCA GACCCTGCACAAGTTCATTCTGTTGTTTGCTGTGTTCGATGAAGGAAAGTCCTGGCACTCAGAAACCAAGAACTCGCT GATGCAGGATAGAGATGCGGCCTCGGCCAGAGCCTGGCCTAAAATGCACACCGTCAACGGATATGTGAACAGGTCGCT CCCTGGCCTCATCGGCTGCCACAGAAAGTCCGTGTATTGGCATGTGATCGGCATGGGTACTACTCCGGAAGTGCATAG TATCTTTCTGGAGGGCCATACCTTCTTGGTGCGCAACCACAGACAGGCCTCGCTGGAAATCTCGCCTATCACTTTCTT GACTGCGCAGACCCTCCTTATGGACCTTGGACAGTTCCTGCTGTTCTGTCACATCAGCTCCCATCAGCATGATGGGAT GGAGGCCTATGTCAAAGTGGACTCCTGCCCTGAGGAGCCACAGCTCCGGATGAAGAACAATGAGGAAGCGGAGGATTA CGACGACGACCTGACTGACAGCGAAATGGACGTCGTGCGATTCGATGACGACAACAGCCCGTCCTTCATCCAAATTAG

	ATCAGTGGCGAAGAAGCACCCCAAGACCTGGGTGCACTACATTGCCGCCGAGGAAGAGGACTGGGACTACGCGCCGCT GGTGCTGGCGCCAGACGACAGGAGCTACAAGTCCCAGTACCTCAACAACGGGCCGCAGCGCATTGGCAGGAAGTACAA GAAAGTCCGCTTCATGGCCTACACTGATGAAACCTTCAAGACGAGGGAAGCCATCCAGCACGAGTCAGGCATCCTGGG ACCGCTCCTTTACGGCGAAGTCGGGGATACCCTGCTCATCATTTTCAAGAACCAGGCATCGCGGCCCTACAACATCTA CCCTCACGGGATCACAGACGTGCGCCCGCTCTACTCCCGCCGGCTGCCCAAGGGAGTGAAGCACCTGAAGGATTTTCC CATCCTGCCGGGAGAAATCTTCAAGTACAAGTGGACCGTGACTGTGGAAGATGGCCCTACCAAGTCGGACCCTCGCTG TCTGACCCGGTACTATTCCTCGTTTGTGAACATGGAGCGCGACCTGGCCTCGGGGCTGATTGGTCCGCTGCTGATCTG CTACAAGGAGTCCGTGGACCAGCGCGGGAACCAGATCATGTCCGACAAGCGCAACGTGATCCTGTTCTCTGTCTTTGA TGAAAACAGATCGTGGTACTTGACTGAGAATATCCAGCGGTTCCTGCCCAACCCAGCGGGAGTGCAACTGGAGGACCC GGAGTTCCAGGCCTCAAACATTATGCACTCTATCAACGGCTATGTGTTCGACTCGCTCCAACTGAGCGTGTGCCTGCA TGAAGTGGCATACTGGTACATTCTGTCCATCGGAGCCCAGACCGACTTCCTGTCCGTGTTCTTCTCCGGATACACCTT CAAGCATAAGATGGTGTACGAGGACACTCTGACCCTCTTCCCATTTTCCGGAGAAACTGTGTTCATGTCAATGGAAAA CCCGGGCTTGTGGATTCTGGGTTGCCATAACTCGGACTTCCGGAATAGAGGGATGACCGCCCTGCTGAAAGTGTCCAG CTGTGACAAGAATACCGGCGATTACTACGAGGACAGCTATGAGGACATCTCCGCTTATCTGCTGTCCAAGAACAACGC CATTGAACCCAGGTCCTTCTCCCAAAACGGTGCACCGACCTCCGAAAGCGCCACCCCAGAGTCAGGACCTGGCTCGGA ACCGGCTACCTCGGGCTCAGAGACACCGGGGACTTCCGAGTCCGCAACCCCCGAGAGTGGACCCGGATCCGAACCAGC AACCTCAGGATCAGAAACCCCGGGAACTTCGGAATCCGCCACTCCCGAGTCGGGACCAGGCACCTCCACTGAGCCTTC CGAGGGAAGCGCCCCCGGATCCCCTGCTGGATCCCCTACCAGCACTGAAGAAGGCACCTCAGAATCCGCGACCCCTGA GTCCGGCCCTGGAAGCGAACCCGCCACCTCCGGTTCCGAAACCCCTGGGACTAGCGAGAGCGCCACTCCGGAATCGGG CCCAGGAAGCCCTGCCGGATCCCCGACCAGCACCGAGGAGGGAAGCCCCGCCGGGTCACCGACTTCCACTGAGGAGGG AGCCTCATCCCCCCCCGTGCTGAAGCGGCATCAAAGAGAGATCACCAGGACCACTCTCCAGTCCGATCAGGAAGAAAT TGACTACGACGATACTATCAGCGTGGAGATGAAGAAGGAGGACTTCGACATCTACGATGAGGATGAGAACCAGTCCCC TCGGAGCTTTCAGAAGAAAACCCGCCACTACTTCATCGCTGCCGTGGAGCGGCTGTGGGATTACGGGATGTCCAGCTC ACCGCATGTGCTGCGGAATAGAGCGCAGTCAGGATCGGTGCCCCAGTTCAAGAAGGTCGTGTTCCAAGAGTTCACCGA CGGGTCCTTCACTCAACCCCTGTACCGGGGCGAACTCAACGAACACCTGGGACTGCTTGGGCCGTATATCAGGGCAGA AGTGGAAGATAACATCATGGTCACCTTCCGCAACCAGGCCTCCCGGCCGTACAGCTTCTACTCTTCACTGATCTCCTA CGAGGAAGATCAGCGGCAGGGAGCCGAGCCCCGGAAGAACTTCGTCAAGCCTAACGAAACTAAGACCTACTTTTGGAA GGTCCAGCATCACATGGCCCCGACCAAAGACGAGTTCGACTGTAAAGCCTGGGCCTACTTCTCCGATGTGGACCTGGA GAAGGACGTGCACTCGGGACTCATTGGCCCGCTCCTTGTGTGCCATACTAATACCCTGAACCCTGCTCACGGTCGCCA AGTCACAGTGCAGGAGTTCGCCCTCTTCTTCACCATCTTCGATGAAACAAAGTCCTGGTACTTTACTGAGAACATGGA ACGCAATTGCAGGGCACCCTGCAACATCCAGATGGAAGATCCCACCTTCAAGGAAAACTACCGGTTTCATGCCATTAA CGGCTACATAATGGACACGTTGCCAGGACTGGTCATGGCCCAGGACCAGAGAATCCGGTGGTATCTGCTCTCCATGGG CTCCAACGAAAACATTCACAGCATTCATTTTTCCGGCCATGTGTTCACCGTCCGGAAGAAGGAAGAGTACAAGATGGC TCTGTACAACCTCTACCCTGGAGTGTTCGAGACTGTGGAAATGCTGCCTAGCAAGGCCGGCATTTGGAGAGTGGAATG CCTGATCGGAGAGCATTTGCACGCCGGAATGTCCACCCTGTTTCTTGTGTACTCCAACAAGTGCCAGACCCCGCTGGG AATGGCCTCAGGTCATATTAGGGATTTCCAGATCACTGCTTCGGGGCAGTACGGGCAGTGGGCACCTAAGTTGGCCCG GCTGCACTACTCTGGCTCCATCAATGCCTGGTCCACCAAGGAACCCTTCTCCTGGATTAAGGTGGACCTCCTGGCCCC AATGATTATTCACGGTATTAAGACCCAGGGTGCCCGACAGAAGTTCTCCTCACTCTACATCTCGCAATTCATCATAAT GTACAGCCTGGATGGGAAGAAGTGGCAGACCTACCGGGGAAACTCCACTGGAACGCTCATGGTGTTTTTCGGCAACGT GGACTCCTCCGGCATTAAGCACAACATCTTCAACCCTCCGATCATTGCTCGGTACATCCGGCTGCACCCAACTCACTA CAGCATCCGGTCCACCCTGCGGATGGAACTGATGGGTTGTGACCTGAACTCCTGCTCCATGCCCCTTGGGATGGAATC CAAGGCCATTAGCGATGCACAGATCACCGCCTCTTCATACTTCACCAACATGTTCGCGACCTGGTCCCCGTCGAAGGC CCGCCTGCACCTCCAAGGTCGCTCCAATGCGTGGCGGCCTCAAGTGAACAACCCCAAGGAGTGGCTCCAGGTCGACTT CCAAAAGACCATGAAGGTCACCGGAGTGACCACCCAGGGCGTGAAGTCCCTGCTGACCTCTATGTACGTTAAGGAGTT CCTCATCTCCTCAAGCCAAGACGGACATCAGTGGACCCTGTTCTTCCAAAACGGAAAAGTCAAAGTATTCCAGGGCAA CCAGGACTCCTTCACCCCTGTGGTCAACAGCCTGGACCCCCCATTGCTGACCCGCTACCTCCGCATCCACCCCCAAAG CTGGGTCCACCAGATCGCACTGCGCATGGAGGTCCTTGGATGCGAAGCCCAAGATCTGTACTAAGCGGCCGCTCATAA TCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATA CGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTT GCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCC CACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGA ACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGG GAAATCATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCTGCTACGTCCC TTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCG CCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCTGCCTAGGCGACTGTGCCTTCTAGTTGCCAGCCATCT GTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAA ATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGG GAAGACAATAGCAGGCATGCTGGGGAAGACCATGGGCGCGCCAGGCCTGTCGACGCCCGGGCGGTACCGCGATCGCTC GCGACGCATAAAG

SEQ ID NO: 15 A1MB2 enhancer	GGCCCCAGGTTAATTTTTAAAAAGCAGTCAAAGGTCAAAGTGGCCCTTGGCAGCATTTACTCTCTCTATTGACTTTGG TTAATAATCTCAGGAGCACAAACATTCCTGGAGGCAGGAGAAGAAATCAACATCCTGGACTTATCCTCTGGGCCTCTC CCCACCTTCGATGGCCCCAGGTTAATTTTTAAAAAGCAGTCAAAGGTCAAAGTGGCCCTTGGCAGCATTTACTCTCTC TATTGACTTTGGTTAATAATCTCAGGAGCACAAACATTCCTGGAGGCAGGAGAAGAAATCAACATCCTGGACTTATCC TCTGGGCCTCTCCCCACC
SEQ ID NO: 16 mTTR promoter	GATATCTACCTGCTGATCGCCCGGCCCCTGTTCAAACATGTCCTAATACTCTGTCGGGGCAAAGGTCGGCAGTAGTTT TCCATCTTACTCAACATCCTCCCAGTGTACGTAGGATCCTGTCTGTCTGCACATTTCGTAGAGCGAGTGTTCCGATAC TCTAATCTCCCGGGGCAAAGGTCGTATTGACTTAGGTTACTTATTCTCCTTTTGTTGACTAAGTCAATAATCAGAATC AGCAGGTTTGGAGTCAGCTTGGCAGGGATCAGCAGCCTGGGTTGGAAGGAGGGGGTATAAAAGCCCCTTCACCAGGAG AAGCCGTCACACAGATCCACAAGCTCCTGCTAG
SEQ ID NO: 17 Chimeric Intron	TCAGGAGCACAAACATTCCTGGAGGCAGGAGAAGAAATCAACATCCTGGACTTATCCTCTGGGCCTCTCCCCACCGAT ATCTACCTGCTGATCGCCCGGCCCCTGTTCAAACATGTCCTAATACTCTGTCGGGGCAAAGGTCGGCAGTAGTTTTCC ATCTTACTCAACATCCTCCCAGTGTACGTAGGATCCTGTCTGTCTGCACATTTCGTAGAGCGAGTGTTCCGATACTCT AATCTCCCGGGGCAAAGGTCGTATTGACTTAGGTTACTTATTCTCCTTTTGTTGACTAAGTCAATAATCAGAATCAGC AGGTTTGGAGTCAGCTTGGCAGGGATCAGCAGCCTGGGTTGGAAGGAGGGGGTATAAAAGCCCCTTCACCAGGAGAAG CCGTCACACAGATCCACAAGCTCCTGCTAGAGTCGCTGCGCGCTGCCTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTC GCGCCGCCCGCCCCGGCTCTGACTGACCGCGTTACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGCTG TAATTAGCGCTTGGTTTATTGACGGCTTGTTTCTTTTCTGTGGCTGCGTGAAAGCCTTGAGGGGCTCCGGGAAGGCCC TTTGTGCGGGGGGAGCGGCTCGGGGGGTGCGTGCGTGTGTGTGTGCGTGGGGAGCGCCGCGTGCGGCTCCGCGCTGCC CGGCGGCTGTGAGCGCTGCGGGCGCGGCGCGGGGCTTTGTGCGCTCCGCAGTGTGCGCGAGGGGAGCGCGGCCGGGGG CGGTGCCCCGCGGTGCGGGGGGGGCTGCGAGGGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGGGGGTGAGCAGG GGGTGTGGGCGCGTCGGTCGGGCTGCAACCCCCCCTGCACCCCCCTCCCCGAGTTGCTGAGCACGGCCCGGCTTCGGG TGCGGGGCTCCGTACGGGGCGTGGCGCGGGGCTCGCCGTGCCGGGCGGGGGGTGGCGGCAGGTGGGGGTGCCGGGCGG GGCGGGGCCGCCTCGGGCCGGGGAGGGCTCGGGGGAGGGGCGCGGCGGCCCCCGGAGCGCCGGCGGCTGTCGAGGCGC GGCGAGCCGCAGCCATTGCCTTTTATGGTAATCGTGCGAGAGGGCGCAGGGACTTCCTTTGTCCCAAATCTGTGCGGA GCCGAAATCTGGGAGGCGCCGCCGCACCCCCTCTAGCGGGCGCGGGGCGAAGCGGTGCGGCGCCGGCAGGAAGGAAAT GGGCGGGGAGGGCCTTCGTGCGTCGCCGCGCCGCCGTCCCCTTCTCCCTCTCCAGCCTCGGGGCTGTCCGCGGGGGGA CGGCTGCCTTCGGGGGGGACGGGGCAGGGCGGGGTTCGGCTTCTGGCGTGTGACCGGCGGCTCTAGAGCCTCTGCTAA CCTTGTTCTTGCCTTCTTCTTTTTCCTACAGCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCATCATTTTGGCAA AGAATTA
SEQ ID NO: 18 WPRE	TCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATG TGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATC CTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGACGC AACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCAC GGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTT GTCGGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCTGCTA CGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCG CCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCTG
SEQ ID NO: 19 bGHpA	CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTC CCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGG TGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGA
SEQ ID NO: 20 Amino acid sequence of wild type human mature FVIII protein	ATRRYYLGAVELSWDYMQSDLGELPVDARFPPRVPKSFPFNTSVVYKKTLFVEFTDHLFNIAKPRPPWMGLLGPTIQA EVYDTVVITLKNMASHPVSLHAVGVSYWKASEGAEYDDQTSQREKEDDKVFPGGSHTYVWQVLKENGPMASDPLCLTY SYLSHVDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVFDEGKSWHSETKNSLMQDRDAASARAWPKMHTV NGYVNRSLPGLIGCHRKSVYWHVIGMGTTPEVHSIFLEGHTFLVRNHRQASLEISPITFLTAQTLLMDLGQFLLFCHI SSHQHDGMEAYVKVDSCPEEPQLRMKNNEEAEDYDDDLTDSEMDVVRFDDDNSPSFIQIRSVAKKHPKTWVHYIAAEE EDWDYAPLVLAPDDRSYKSQYLNNGPQRIGRKYKKVRFMAYTDETFKTREAIQHESGILGPLLYGEVGDTLLIIFKNQ ASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRYYSSFVNMERDLASG LIGPLLICYKESVDQRGNQIMSDKRNVILFSVFDENRSWYLTENIQRFLPNPAGVQLEDPEFQASNIMHSINGYVFDS LQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKHKMVYEDTLTLFPFSGETVFMSMENPGLWILGCHNSDFRNRGM TALLKVSSCDKNTGDYYEDSYEDISAYLLSKNNAIEPRSFSQNSRHPSTRQKQFNATTIPENDIEKTDPWFAHRTPMP KIQNVSSSDLLMLLRQSPTPHGLSLSDLQEAKYETFSDDPSPGAIDSNNSLSEMTHFRPQLHHSGDMVFTPESGLQLR LNEKLGTTAATELKKLDFKVSSTSNNLISTIPSDNLAAGTDNTSSLGPPSMPVHYDSQLDTTLFGKKSSPLTESGGPL SLSEENNDSKLLESGLMNSQESSWGKNVSSTESGRLFKGKRAHGPALLTKDNALFKVSISLLKTNKTSNNSATNRKTH

	IDGPSLLIENSPSVWQNILESDTEFKKVTPLIHDRMLMDKNATALRLNHMSNKTTSSKNMEMVQQKKEGPIPPDAQNP DMSFFKMLFLPESARWIQRTHGKNSLNSGQGPSPKQLVSLGPEKSVEGQNFLSEKNKVVVGKGEFTKDVGLKEMVFPS SRNLFLTNLDNLHENNTHNQEKKIQEEIEKKETLIQENVVLPQIHTVTGTKNFMKNLFLLSTRQNVEGSYDGAYAPVL QDFRSLNDSTNRTKKHTAHFSKKGEEENLEGLGNQTKQIVEKYACTTRISPNTSQQNFVTQRSKRALKQFRLPLEETE LEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKEKGAITQSPLSDCLTRSHSIPQANRSPLPIAKVSSFPSIRPIYL TRVLFQDNSSHLPAASYRKKDSGVQESSHFLQGAKKNNLSLAILTLEMTGDQREVGSLGTSATNSVTYKKVENTVLPK PDLPKTSGKVELLPKVHIYQKDLFPTETSNGSPGHLDLVEGSLLQGTEGAIKWNEANRPGKVPFLRVATESSAKTPSK LLDPLAWDNHYGTQIPKEEWKSQEKSPEKTAFKKKDTILSLNACESNHAIAAINEGQNKPEIEVTWAKQGRTERLCSQ NPPVLKRHQREITRTTLQSDQEEIDYDDTISVEMKKEDFDIYDEDENQSPRSFQKKTRHYFIAAVERLWDYGMSSSPH VLRNRAQSGSVPQFKKVVFQEFTDGSFTQPLYRGELNEHLGLLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLISYEE DQRQGAEPRKNFVKPNETKTYFWKVQHHMAPTKDEFDCKAWAYFSDVDLEKDVHSGLIGPLLVCHTNTLNPAHGRQVT VQEFALFFTIFDETKSWYFTENMERNCRAPCNIQMEDPTFKENYRFHAINGYIMDTLPGLVMAQDQRIRWYLLSMGSN ENIHSIHFSGHVFTVRKKEEYKMALYNLYPGVFETVEMLPSKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPLGMA SGHIRDFQITASGQYGQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIIHGIKTQGARQKFSSLYISQFIIMYS LDGKKWQTYRGNSTGTLMVFFGNVDSSGIKHNIFNPPIIARYIRLHPTHYSIRSTLRMELMGCDLNSCSMPLGMESKA ISDAQITASSYFTNMFATWSPSKARLHLQGRSNAWRPQVNNPKEWLQVDFQKTMKVTGVTTQGVKSLLTSMYVKEFLI SSSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWVHQIALRMEVLGCEAQDLY
SEQ ID NO: 21 B19 WT 5’	CCAAATCAGATGCCGCCGGTCGCCGCCGGTAGGCGGGACTTCCGGTACAAGATGGCGGACAATTACGTCATTTCCTGT GACGTCATTTCCTGTGACGTCACTTCCGGTGGGCGGGACTTCCGGAATTAGGGTTGGCTCTGGGCCAGCTTGCTTGGG GTTGCCTTGACACTAAGACAAGCGGCGCGCCGCTTGATCTTAGTGGCACGTCAACCCCAAGCGCTGGCCCAGAGCCAA CCCTAATTCCGGAAGTCCCGCCCACCGGAAGTGACGTCACAGGAAATGACGTCACAGGAAATGACGTAATTGTCCGCC ATCTTGTACCGGAAGTCCCGCCTACCGGCGGCGACCGGCGGCATCTGATTTGGTGTCTTCTTTTAAATTTT
SEQ ID NO: 22 B19 WT 3’	AAAATTTAAAAGAAGACACCAAATCAGATGCCGCCGGTCGCCGCCGGTAGGCGGGACTTCCGGTACAAGATGGCGGAC AATTACGTCATTTCCTGTGACGTCATTTCCTGTGACGTCACTTCCGGTGGGCGGGACTTCCGGAATTAGGGTTGGCTC TGGGCCAGCGCTTGGGGTTGACGTGCCACTAAGATCAAGCGGCGCGCCGCTTGTCTTAGTGTCAAGGCAACCCCAAGC AAGCTGGCCCAGAGCCAACCCTAATTCCGGAAGTCCCGCCCACCGGAAGTGACGTCACAGGAAATGACGTCACAGGAA ATGACGTAATTGTCCGCCATCTTGTACCGGAAGTCCCGCCTACCGGCGGCGACCGGCGGCATCTGATTTGG
SEQ ID NO: 23 5’_B19_minimal	CCAAATCAGATGCCGCCGGTCGCCGCCGGTAGGCGGGACTTCCGGTACAGCGCGCCGCTGTACCGGAAGTCCCGCCTA CCGGCGGCGACCGGCGGCATCTGATTTGGTGTCTTCTTTTAAATTTT
SEQ ID NO: 24 3’_B19_minimal	AAAATTTAAAAGAAGACACCAAATCAGATGCCGCCGGTCGCCGCCGGTAGGCGGGACTTCCGGTACAGCGGCGCGCTG TACCGGAAGTCCCGCCTACCGGCGGCGACCGGCGGCATCTGATTTGG
SEQ ID NO: 25 5’_GPV_minimal	CTCATTGGAGGGTTCGTTCGTTCGAACGTTCGTTCGCATGCGAACGAACGTTCGAACGAACGAACCCTCCAATGAGAC TCAAGGACAAGAGGATATTTTGCGCGCCAGGAAGTG
SEQ ID NO: 26 3’_GPV_minimal	CACTTCCTGGCGCGCAAAATATCCTCTTGTCCTTGAGTCTCATTGGAGGGTTCGTTCGTTCGAACGTTCGTTCGCATG CGAACGAACGTTCGAACGAACGAACCCTCCAATGAG
SEQ ID NO: 27 5’_GPV_Δ186	CTCATTGGAGGGTTCGTTCGTTCGAACCAGCCAATCAGGGGAGGGGGAAGTGACGCAAGTTCCGGTCACATGCTTCCG GTGACGCACATCCGGTGACGTAGTTCGCATGCCTGTCTATCGCCTACCCATCCCTGTCTGAGATCAAGGGCGTGATCG TGCACAGACTGGAGAGCGTGTCCTATAATATCGGCTCTCAGGAGTGGAGCACCACAGTGCCCAGATACGTGGCCACCC AGGGCTATCTGATCTCCAACTTCGACGCATGCGAACTACGTCACCGGATGTGCGTCACCGGAAGCATGTGACCGGAAC TTGCGTCACTTCCCCCTCCCCTGATTGGCTGGTTCGAACGAACGAACCCTCCAATGAGACTCAAGGACAAGAGGATAT TTTGCGCGCCAGGAAGTG

SEQ ID NO: 28 3'_GPV_Δ186	CACTTCCTGGCGCGCAAAATATCCTCTTGTCCTTGAGTCTCATTGGAGGGTTCGTTCGTTCGAACCAGCCAATCAGGG GAGGGGGAAGTGACGCAAGTTCCGGTCACATGCTTCCGGTGACGCACATCCGGTGACGTAGTTCGCATGCCTGTCTAT CGCCTACCCATCCCTGTCTGAGATCAAGGGCGTGATCGTGCACAGACTGGAGAGCGTGTCCTATAATATCGGCTCTCA GGAGTGGAGCACCACAGTGCCCAGATACGTGGCCACCCAGGGCTATCTGATCTCCAACTTCGACGCATGCGAACTACG TCACCGGATGTGCGTCACCGGAAGCATGTGACCGGAACTTGCGTCACTTCCCCCTCCCCTGATTGGCTGGTTCGAACG AACGAACCCTCCAATGAG
SEQ ID NO: 29 5'_GPV_Δ120	CTCATTGGAGGGTTCGTTCGTTCGAACCAGCCAATCAGGGGAGGGGGAAGTGACGCAAGTTCCGGTCACATGCTTCCG GTGACGCACATCCGGTGACGTAGTTCCGGTCACGTGCTTCCTGTCACGTGTTTCCGGTCGCATGCCTGTCTATCGCCT ACCCATCCCTGTCTGAGATCAAGGGCGTGATCGTGCACAGACTGGAGAGCGTGTCCTATAATATCGGCTCTCAGGAGT GGAGCACCACAGTGCCCAGATACGTGGCCACCCAGGGCTATCTGATCTCCAACTTCGACGCATGCTCACGTGACCGGA AACACGTGACAGGAAGCACGTGACCGGAACTACGTCACCGGATGTGCGTCACCGGAAGCATGTGACCGGAACTTGCGT CACTTCCCCCTCCCCTGATTGGCTGGTTCGAACGAACGAACCCTCCAATGAGACTCAAGGACAAGAGGATATTTTGCG CGCCAGGAAGTG
SEQ ID NO: 30 3'_GPV_Δ120	CACTTCCTGGCGCGCAAAATATCCTCTTGTCCTTGAGTCTCATTGGAGGGTTCGTTCGTTCGAACCAGCCAATCAGGG GAGGGGGAAGTGACGCAAGTTCCGGTCACATGCTTCCGGTGACGCACATCCGGTGACGTAGTTCCGGTCACGTGCTTC CTGTCACGTGTTTCCGGTCACGTGAGCATGCCTGTCTATCGCCTACCCATCCCTGTCTGAGATCAAGGGCGTGATCGT GCACAGACTGGAGAGCGTGTCCTATAATATCGGCTCTCAGGAGTGGAGCACCACAGTGCCCAGATACGTGGCCACCCA GGGCTATCTGATCTCCAACTTCGACGGCATGCGACCGGAAACACGTGACAGGAAGCACGTGACCGGAACTACGTCACC GGATGTGCGTCACCGGAAGCATGTGACCGGAACTTGCGTCACTTCCCCCTCCCCTGATTGGCTGGTTCGAACGAACGA ACCCTCCAATGAG
SEQ ID NO: 31 Nucleic acid sequence of wild type human FVIII	ATGCAAATAGAGCTCTCCACCTGCTTCTTTCTGTGCCTTTTGCGATTCTGCTTTAGTGCCACCAGAAGATACTACCTG GGTGCAGTGGAACTGTCATGGGACTATATGCAAAGTGATCTCGGTGAGCTGCCTGTGGACGCAAGATTTCCTCCTAGA GTGCCAAAATCTTTTCCATTCAACACCTCAGTCGTGTACAAAAAGACTCTGTTTGTAGAATTCACGGATCACCTTTTC AACATCGCTAAGCCAAGGCCACCCTGGATGGGTCTGCTAGGTCCTACCATCCAGGCTGAGGTTTATGATACAGTGGTC ATTACACTTAAGAACATGGCTTCCCATCCTGTCAGTCTTCATGCTGTTGGTGTATCCTACTGGAAAGCTTCTGAGGGA GCTGAATATGATGATCAGACCAGTCAAAGGGAGAAAGAAGATGATAAAGTCTTCCCTGGTGGAAGCCATACATATGTC TGGCAGGTCCTGAAAGAGAATGGTCCAATGGCCTCTGACCCACTGTGCCTTACCTACTCATATCTTTCTCATGTGGAC CTGGTAAAAGACTTGAATTCAGGCCTCATTGGAGCCCTACTAGTATGTAGAGAAGGGAGTCTGGCCAAGGAAAAGACA CAGACCTTGCACAAATTTATACTACTTTTTGCTGTATTTGATGAAGGGAAAAGTTGGCACTCAGAAACAAAGAACTCC TTGATGCAGGATAGGGATGCTGCATCTGCTCGGGCCTGGCCTAAAATGCACACAGTCAATGGTTATGTAAACAGGTCT CTGCCAGGTCTGATTGGATGCCACAGGAAATCAGTCTATTGGCATGTGATTGGAATGGGCACCACTCCTGAAGTGCAC TCAATATTCCTCGAAGGTCACACATTTCTTGTGAGGAACCATCGCCAGGCGTCCTTGGAAATCTCGCCAATAACTTTC CTTACTGCTCAAACACTCTTGATGGACCTTGGACAGTTTCTACTGTTTTGTCATATCTCTTCCCACCAACATGATGGC ATGGAAGCTTATGTCAAAGTAGACAGCTGTCCAGAGGAACCCCAACTACGAATGAAAAATAATGAAGAAGCGGAAGAC TATGATGATGATCTTACTGATTCTGAAATGGATGTGGTCAGGTTTGATGATGACAACTCTCCTTCCTTTATCCAAATT CGCTCAGTTGCCAAGAAGCATCCTAAAACTTGGGTACATTACATTGCTGCTGAAGAGGAGGACTGGGACTATGCTCCC TTAGTCCTCGCCCCCGATGACAGAAGTTATAAAAGTCAATATTTGAACAATGGCCCTCAGCGGATTGGTAGGAAGTAC AAAAAAGTCCGATTTATGGCATACACAGATGAAACCTTTAAGACTCGTGAAGCTATTCAGCATGAATCAGGAATCTTG GGACCTTTACTTTATGGGGAAGTTGGAGACACACTGTTGATTATATTTAAGAATCAAGCAAGCAGACCATATAACATC TACCCTCACGGAATCACTGATGTCCGTCCTTTGTATTCAAGGAGATTACCAAAAGGTGTAAAACATTTGAAGGATTTT CCAATTCTGCCAGGAGAAATATTCAAATATAAATGGACAGTGACTGTAGAAGATGGGCCAACTAAATCAGATCCTCGG TGCCTGACCCGCTATTACTCTAGTTTCGTTAATATGGAGAGAGATCTAGCTTCAGGACTCATTGGCCCTCTCCTCATC TGCTACAAAGAATCTGTAGATCAAAGAGGAAACCAGATAATGTCAGACAAGAGGAATGTCATCCTGTTTTCTGTATTT GATGAGAACCGAAGCTGGTACCTCACAGAGAATATACAACGCTTTCTCCCCAATCCAGCTGGAGTGCAGCTTGAGGAT CCAGAGTTCCAAGCCTCCAACATCATGCACAGCATCAATGGCTATGTTTTTGATAGTTTGCAGTTGTCAGTTTGTTTG CATGAGGTGGCATACTGGTACATTCTAAGCATTGGAGCACAGACTGACTTCCTTTCTGTCTTCTTCTCTGGATATACC TTCAAACACAAAATGGTCTATGAAGACACACTCACCCTATTCCCATTCTCAGGAGAAACTGTCTTCATGTCGATGGAA AACCCAGGTCTATGGATTCTGGGGTGCCACAACTCAGACTTTCGGAACAGAGGCATGACCGCCTTACTGAAGGTTTCT AGTTGTGACAAGAACACTGGTGATTATTACGAGGACAGTTATGAAGATATTTCAGCATACTTGCTGAGTAAAAACAAT GCCATTGAACCAAGAAGCTTCTCCCAGAATTCAAGACACCCTAGCACTAGGCAAAAGCAATTTAATGCCACCACAATT CCAGAAAATGACATAGAGAAGACTGACCCTTGGTTTGCACACAGAACACCTATGCCTAAAATACAAAATGTCTCCTCT AGTGATTTGTTGATGCTCTTGCGACAGAGTCCTACTCCACATGGGCTATCCTTATCTGATCTCCAAGAAGCCAAATAT GAGACTTTTTCTGATGATCCATCACCTGGAGCAATAGACAGTAATAACAGCCTGTCTGAAATGACACACTTCAGGCCA CAGCTCCATCACAGTGGGGACATGGTATTTACCCCTGAGTCAGGCCTCCAATTAAGATTAAATGAGAAACTGGGGACA ACTGCAGCAACAGAGTTGAAGAAACTTGATTTCAAAGTTTCTAGTACATCAAATAATCTGATTTCAACAATTCCATCA GACAATTTGGCAGCAGGTACTGATAATACAAGTTCCTTAGGACCCCCAAGTATGCCAGTTCATTATGATAGTCAATTA GATACCACTCTATTTGGCAAAAAGTCATCTCCCCTTACTGAGTCTGGTGGACCTCTGAGCTTGAGTGAAGAAAATAAT

	GATTCAAAGTTGTTAGAATCAGGTTTAATGAATAGCCAAGAAAGTTCATGGGGAAAAAATGTATCGTCAACAGAGAGT GGTAGGTTATTTAAAGGGAAAAGAGCTCATGGACCTGCTTTGTTGACTAAAGATAATGCCTTATTCAAAGTTAGCATC TCTTTGTTAAAGACAAACAAAACTTCCAATAATTCAGCAACTAATAGAAAGACTCACATTGATGGCCCATCATTATTA ATTGAGAATAGTCCATCAGTCTGGCAAAATATATTAGAAAGTGACACTGAGTTTAAAAAAGTGACACCTTTGATTCAT GACAGAATGCTTATGGACAAAAATGCTACAGCTTTGAGGCTAAATCATATGTCAAATAAAACTACTTCATCAAAAAAC ATGGAAATGGTCCAACAGAAAAAAGAGGGCCCCATTCCACCAGATGCACAAAATCCAGATATGTCGTTCTTTAAGATG CTATTCTTGCCAGAATCAGCAAGGTGGATACAAAGGACTCATGGAAAGAACTCTCTGAACTCTGGGCAAGGCCCCAGT CCAAAGCAATTAGTATCCTTAGGACCAGAAAAATCTGTGGAAGGTCAGAATTTCTTGTCTGAGAAAAACAAAGTGGTA GTAGGAAAGGGTGAATTTACAAAGGACGTAGGACTCAAAGAGATGGTTTTTCCAAGCAGCAGAAACCTATTTCTTACT AACTTGGATAATTTACATGAAAATAATACACACAATCAAGAAAAAAAAATTCAGGAAGAAATAGAAAAGAAGGAAACA TTAATCCAAGAGAATGTAGTTTTGCCTCAGATACATACAGTGACTGGCACTAAGAATTTCATGAAGAACCTTTTCTTA CTGAGCACTAGGCAAAATGTAGAAGGTTCATATGACGGGGCATATGCTCCAGTACTTCAAGATTTTAGGTCATTAAAT GATTCAACAAATAGAACAAAGAAACACACAGCTCATTTCTCAAAAAAAGGGGAGGAAGAAAACTTGGAAGGCTTGGGA AATCAAACCAAGCAAATTGTAGAGAAATATGCATGCACCACAAGGATATCTCCTAATACAAGCCAGCAGAATTTTGTC ACGCAACGTAGTAAGAGAGCTTTGAAACAATTCAGACTCCCACTAGAAGAAACAGAACTTGAAAAAAGGATAATTGTG GATGACACCTCAACCCAGTGGTCCAAAAACATGAAACATTTGACCCCGAGCACCCTCACACAGATAGACTACAATGAG AAGGAGAAAGGGGCCATTACTCAGTCTCCCTTATCAGATTGCCTTACGAGGAGTCATAGCATCCCTCAAGCAAATAGA TCTCCATTACCCATTGCAAAGGTATCATCATTTCCATCTATTAGACCTATATATCTGACCAGGGTCCTATTCCAAGAC AACTCTTCTCATCTTCCAGCAGCATCTTATAGAAAGAAAGATTCTGGGGTCCAAGAAAGCAGTCATTTCTTACAAGGA GCCAAAAAAAAATAACCTTTCTTTAGCCATTCTAACCTTGGAGATGACTGGTGATCAAAGAGAGGTTGGCTCCCTGGGG ACAAGTGCCACAAATTCAGTCACATACAAGAAAGTTGAGAACACTGTTCTCCCGAAACCAGACTTGCCCAAAACATCT GGCAAAGTTGAATTGCTTCCAAAAGTTCACATTTATCAGAAGGACCTATTCCCTACGGAAACTAGCAATGGGTCTCCT GGCCATCTGGATCTCGTGGAAGGGAGCCTTCTTCAGGGAACAGAGGGAGCGATTAAGTGGAATGAAGCAAACAGACCT GGAAAAGTTCCCTTTCTGAGAGTAGCAACAGAAAGCTCTGCAAAGACTCCCTCCAAGCTATTGGATCCTCTTGCTTGG GATAACCACTATGGTACTCAGATACCAAAAGAAGAGTGGAAATCCCAAGAGAAGTCACCAGAAAAAACAGCTTTTAAG AAAAAGGATACCATTTTGTCCCTGAACGCTTGTGAAAGCAATCATGCAATAGCAGCAATAAATGAGGGACAAAATAAG CCCGAAATAGAAGTCACCTGGGCAAAGCAAGGTAGGACTGAAAGGCTGTGCTCTCAAAACCCACCAGTCTTGAAACGC CATCAACGGGAAATAACTCGTACTACTCTTCAGTCAGATCAAGAGGAAATTGACTATGATGATACCATATCAGTTGAA ATGAAGAAGGAAGATTTTGACATTTATGATGAGGATGAAAATCAGAGCCCCCGCAGCTTTCAAAAGAAAACACGACAC TATTTTATTGCTGCAGTGGAGAGGCTCTGGGATTATGGGATGAGTAGCTCCCCACATGTTCTAAGAAACAGGGCTCAG AGTGGCAGTGTCCCTCAGTTCAAGAAAGTTGTTTTCCAGGAATTTACTGATGGCTCCTTTACTCAGCCCTTATACCGT GGAGAACTAAATGAACATTTGGGACTCCTGGGGCCATATATAAGAGCAGAAGTTGAAGATAATATCATGGTAACTTTC AGAAATCAGGCCTCTCGTCCCTATTCCTTCTATTCTAGCCTTATTTCTTATGAGGAAGATCAGAGGCAAGGAGCAGAA CCTAGAAAAAAACTTTGTCAAGCCTAATGAAACCAAAACTTACTTTTGGAAAGTGCAACATCATATGGCACCCACTAAA GATGAGTTTGACTGCAAAGCCTGGGCTTATTTCTCTGATGTTGACCTGGAAAAAGATGTGCACTCAGGCCTGATTGGA CCCCTTCTGGTCTGCCACACTAACACACTGAACCCTGCTCATGGGAGACAAGTGACAGTACAGGAATTTGCTCTGTTT TTCACCATCTTTGATGAGACCAAAAGCTGGTACTTCACTGAAAATATGGAAAGAAACTGCAGGGCTCCCTGCAATATC CAGATGGAAGATCCCACTTTTAAAGAGAATTATCGCTTCCATGCAATCAATGGCTACATAATGGATACACTACCTGGC TTAGTAATGGCTCAGGATCAAAGGATTCGATGGTATCTGCTCAGCATGGGCAGCAATGAAAACATCCATTCTATTCAT TTCAGTGGACATGTGTTCACTGTACGAAAAAAAGAGGAGTATAAAATGGCACTGTACAATCTCTATCCAGGTGTTTTT GAGACAGTGGAAATGTTACCATCCAAAGCTGGAATTTGGCGGGTGGAATGCCTTATTGGCGAGCATCTACATGCTGGG ATGAGCACACTTTTTCTGGTGTACAGCAATAAGTGTCAGACTCCCCTGGGAATGGCTTCTGGACACATTAGAGATTTT CAGATTACAGCTTCAGGACAATATGGACAGTGGGCCCCAAAGCTGGCCAGACTTCATTATTCCGGATCAATCAATGCC TGGAGCACCAAGGAGCCCTTTTCTTGGATCAAGGTGGATCTGTTGGCACCAATGATTATTCACGGCATCAAGACCCAG GGTGCCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTATCATCATGTATAGTCTTGATGGGAAGAAGTGGCAG ACTTATCGAGGAAATTCCACTGGAACCTTAATGGTCTTCTTTGGCAATGTGGATTCATCTGGGATAAAACACAATATT TTTAACCCTCCAATTATTGCTCGATACATCCGTTTGCACCCAACTCATTATAGCATTCGCAGCACTCTTCGCATGGAG TTGATGGGCTGTGATTTAAATAGTTGCAGCATGCCATTGGGAATGGAGAGTAAAGCAATATCAGATGCACAGATTACT GCTTCATCCTACTTTACCAATATGTTTGCCACCTGGTCTCCTTCAAAAGCTCGACTTCACCTCCAAGGGAGGAGTAAT GCCTGGAGACCTCAGGTGAATAATCCAAAAGAGTGGCTGCAAGTGGACTTCCAGAAGACAATGAAAGTCACAGGAGTA ACTACTCAGGGAGTAAAATCTCTGCTTACCAGCATGTATGTGAAGGAGTTCCTCATCTCCAGCAGTCAAGATGGCCAT CAGTGGACTCTCTTTTTTCAGAATGGCAAAGTAAAGGTTTTTCAGGGAAATCAAGACTCCTTCACACCTGTGGTGAAC TCTCTAGACCCACCGTTACTGACTCGCTACCTTCGAATTCACCCCCAGAGTTGGGTGCACCAGATTGCCCTGAGGATG GAGGTTCTGGGCTGCGAGGCACAGGACCTCTAC
SEQ ID NO: 32 Nucleotide sequence encoding	GCCACTCGCCGGTACTACCTTGGAGCCGTGGAGCTTTCATGGGACTACATGCAGAGCGACCTGGGCGAACTCCCCGTG GATGCCAGATTCCCCCCCCGCGTGCCAAAGTCCTTCCCCTTTAACACCTCCGTGGTGTACAAGAAAACCCTCTTTGTC GAGTTCACTGACCACCTGTTCAACATCGCCAAGCCGCGCCCACCTTGGATGGGCCTCCTGGGACCGACCATTCAAGCT GAAGTGTACGACACCGTGGTGATCACCCTGAAGAACATGGCGTCCCACCCCGTGTCCCTGCATGCGGTCGGAGTGTCC TACTGGAAGGCCTCCGAAGGAGCTGAGTACGACGACCAGACTAGCCAGCGGGAAAAGGAGGACGATAAAGTGTTCCCG GGCGGCTCGCATACTTACGTGTGGCAAGTCCTGAAGGAAAACGGACCTATGGCATCCGATCCTCTGTGCCTGACTTAC

BDD-co6FVIII (V1.0) (no XTEN)	TCCTACCTTTCCCATGTGGACCTCGTGAAGGACCTGAACAGCGGGCTGATTGGTGCACTTCTCGTGTGCCGCGAAGGT TCGCTCGCTAAGGAAAAGACCCAGACCCTCCATAAGTTCATCCTTTTGTTCGCTGTGTTCGATGAAGGAAAGTCATGG CATTCCGAAACTAAGAACTCGCTGATGCAGGACCGGGATGCCGCCTCAGCCCGCGCCTGGCCTAAAATGCATACAGTC AACGGATACGTGAATCGGTCACTGCCCGGGCTCATCGGTTGTCACAGAAAGTCCGTGTACTGGCACGTCATCGGCATG GGCACTACGCCTGAAGTGCACTCCATCTTCCTGGAAGGGCACACCTTCCTCGTGCGCAACCACCGCCAGGCCTCTCTG GAAATCTCCCCGATTACCTTTCTGACCGCCCAGACTCTGCTCATGGACCTGGGGCAGTTCCTTCTCTTCTGCCACATC TCCAGCCATCAGCACGACGGAATGGAGGCCTACGTGAAGGTGGACTCATGCCCGGAAGAACCTCAGTTGCGGATGAAG AACAACGAGGAGGCCGAGGACTATGACGACGATTTGACTGACTCCGAGATGGACGTCGTGCGGTTCGATGACGACAAC AGCCCCAGCTTCATCCAGATTCGCAGCGTGGCCAAGAAGCACCCCAAAACCTGGGTGCACTACATCGCGGCCGAGGAA GAAGATTGGGACTACGCCCCGTTGGTGCTGGCACCCGATGACCGGTCGTACAAGTCCCAGTATCTGAACAATGGTCCG CAGCGGATTGGCAGAAAGTACAAGAAAGTGCGGTTCATGGCGTACACTGACGAAACGTTTAAGACCCGGGAGGCCATT CAACATGAGAGCGGCATTCTGGGACCACTGCTGTACGGAGAGGTCGGCGATACCCTGCTCATCATCTTCAAAAACCAG GCCTCCCGGCCTTACAACATCTACCCTCACGGAATCACCGACGTGCGGCCACTCTACTCGCGGCGCCTGCCGAAGGGC GTCAAGCACCTGAAAGACTTCCCTATCCTGCCGGGCGAAATCTTCAAGTATAAGTGGACCGTCACCGTGGAGGACGGG CCCACCAAGAGCGATCCTAGGTGTCTGACTCGGTACTACTCCAGCTTCGTGAACATGGAACGGGACCTGGCATCGGGA CTCATTGGACCGCTGCTGATCTGCTACAAAGAGTCGGTGGATCAACGCGGCAACCAGATCATGTCCGACAAGCGCAAC GTGATCCTGTTCTCCGTGTTTGATGAAAACAGATCCTGGTACCTCACTGAAAACATCCAGAGGTTCCTCCCAAACCCC GCAGGAGTGCAACTGGAGGACCCTGAGTTTCAGGCCTCGAATATCATGCACTCGATTAACGGTTACGTGTTCGACTCG CTGCAGCTGAGCGTGTGCCTCCATGAAGTCGCTTACTGGTACATTCTGTCCATCGGCGCCCAGACTGACTTCCTGAGC GTGTTCTTTTCCGGTTACACCTTTAAGCACAAGATGGTGTACGAAGATACCCTGACCCTGTTCCCTTTCTCCGGCGAA ACGGTGTTCATGTCGATGGAGAACCCGGGTCTGTGGATTCTGGGATGCCACAACAGCGACTTTCGGAACCGCGGAATG ACTGCCCTGCTGAAGGTGTCCTCATGCGACAAGAACACCGGAGACTACTACGAGGACTCCTACGAGGATATCTCAGCC TACCTCCTGTCCAAGAACAACGCGATCGAGCCGCGCAGCTTCAGCCAGAACCCGCCTGTGCTGAAGAGGCACCAGCGA GAAATTACCCGGACCACCCTCCAATCGGATCAGGAGGAAATCGACTACGACGACACCATCTCGGTGGAAATGAAGAAG GAAGATTTCGATATCTACGACGAGGACGAAAATCAGTCCCCTCGCTCATTCCAAAAGAAAACTAGACACTACTTTATC GCCGCGGTGGAAAGACTGTGGGACTATGGAATGTCATCCAGCCCTCACGTCCTTCGGAACCGGGCCCAGAGCGGATCG GTGCCTCAGTTCAAGAAAGTGGTGTTCCAGGAGTTCACCGACGGCAGCTTCACCCAGCCGCTGTACCGGGGAGAACTG AACGAACACCTGGGCCTGCTCGGTCCCTACATCCGCGCGGAAGTGGAGGATAACATCATGGTGACCTTCCGTAACCAA GCATCCAGACCTTACTCCTTCTATTCCTCCCTGATCTCATACGAGGAGGACCAGCGCCAAGGCGCCGAGCCCCGCAAG AACTTCGTCAAGCCCAACGAGACTAAGACCTACTTCTGGAAGGTCCAACACCATATGGCCCCGACCAAGGATGAGTTT GACTGCAAGGCCTGGGCCTACTTCTCCGACGTGGACCTTGAGAAGGATGTCCATTCCGGCCTGATCGGGCCGCTGCTC GTGTGTCACACCAACACCCTGAACCCAGCGCATGGACGCCAGGTCACCGTCCAGGAGTTTGCTCTGTTCTTCACCATT TTTGACGAAACTAAGTCCTGGTACTTCACCGAGAATATGGAGCGAAACTGTAGAGCGCCCTGCAATATCCAGATGGAA GATCCGACTTTCAAGGAGAACTATAGATTCCACGCCATCAACGGGTACATCATGGATACTCTGCCGGGGCTGGTCATG GCCCAGGATCAGAGGATTCGGTGGTACTTGCTGTCAATGGGATCGAACGAAAACATTCACTCCATTCACTTCTCCGGT CACGTGTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCGCTGTACAATCTGTACCCCGGGGTGTTCGAAACTGTG GAGATGCTGCCGTCCAAGGCCGGCATCTGGAGAGTGGAGTGCCTGATCGGAGAGCACCTCCACGCGGGGATGTCCACC CTCTTCCTGGTGTACTCGAATAAGTGCCAGACCCCGCTGGGCATGGCCTCGGGCCACATCAGAGACTTCCAGATCACA GCAAGCGGACAATACGGCCAATGGGCGCCGAAGCTGGCCCGCTTGCACTACTCCGGATCGATCAACGCATGGTCCACC AAGGAACCGTTCTCGTGGATTAAGGTGGACCTCCTGGCCCCTATGATTATCCACGGAATTAAGACCCAGGGCGCCAGG CAGAAGTTCTCCTCCCTGTACATCTCGCAATTCATCATCATGTACAGCCTGGACGGGAAGAAGTGGCAGACTTACAGG GGAAACTCCACCGGCACCCTGATGGTCTTTTTCGGCAACGTGGATTCCTCCGGCATTAAGCACAACATCTTCAACCCA CCGATCATAGCCAGATATATTAGGCTCCACCCCACTCACTACTCAATCCGCTCAACTCTTCGGATGGAACTCATGGGG TGCGACCTGAACTCCTGCTCCATGCCGTTGGGGATGGAATCAAAGGCTATTAGCGACGCCCAGATCACCGCGAGCTCC TACTTCACTAACATGTTCGCCACCTGGAGCCCCTCCAAGGCCAGGCTGCACTTGCAGGGACGGTCAAATGCCTGGCGG CCGCAAGTGAACAATCCGAAGGAATGGCTTCAAGTGGATTTCCAAAAGACCATGAAAGTGACCGGAGTCACCACCCAG GGAGTGAAGTCCCTTCTGACCTCGATGTATGTGAAGGAGTTCCTGATTAGCAGCAGCCAGGACGGGCACCAGTGGACC CTGTTCTTCCAAAACGGAAAGGTCAAGGTGTTCCAGGGGAACCAGGACTCGTTCACACCCGTGGTGAACTCCCTGGAC CCCCCACTGCTGACGCGGTACTTGAGGATTCATCCTCAGTCCTGGGTCCATCAGATTGCATTGCGAATGGAAGTCCTG GGCTGCGAGGCCCAGGACCTGTACTGA
SEQ ID NO: 33 Nucleotide sequence encoding coBDDFVIII (V2.0) (no XTEN)	GCCACCCGCCGGTATTACTTAGGTGCTGTGGAACTGAGCTGGGACTACATGCAGTCCGACCTGGGAGAACTGCCGGTG GACGCGAGATTCCCACCTAGAGTCCCGAAGTCCTTCCCATTCAACACCTCCGTGGTCTACAAAAAGACCCTGTTCGTG GAGTTCACTGACCACCTTTTCAATATTGCCAAGCCGCGCCCCCCCTGGATGGGCCTGCTTGGTCCTACGATCCAAGCA GAGGTCTACGACACCGTGGTCATCACACTGAAGAACATGGCCTCACACCCCGTGTCGCTGCATGCTGTGGGAGTGTCC TACTGGAAGGCCTCAGAGGGTGCCGAATATGATGACCAGACCAGCCAGAGGGAAAAGGAGGATGACAAAGTGTTCCCG GGTGGCAGCCACACTTACGTGTGGCAAGTGCTGAAGGAAAACGGGCCTATGGCGTCGGACCCCCTATGCCTGACCTAC TCCTACCTGTCCCATGTGGACCTTGTGAAGGATCTCAACTCGGGACTGATCGGCGCCCTCTTGGTGTGCAGAGAAGGC AGCCTGGCGAAGGAAAAGACTCAGACCCTGCACAAGTTCATTCTGTTGTTTGCTGTGTTCGATGAAGGAAAGTCCTGG CACTCAGAAACCAAGAACTCGCTGATGCAGGATAGAGATGCGGCCTCGGCCAGAGCCTGGCCTAAAATGCACACCGTC AACGGATATGTGAACAGGTCGCTCCCTGGCCTCATCGGCTGCCACAGAAAGTCCGTGTATTGGCATGTGATCGGCATG GGTACTACTCCGGAAGTGCATAGTATCTTTCTGGAGGGCCATACCTTCTTGGTGCGCAACCACAGACAGGCCTCGCTG

	GAAATCTCGCCTATCACTTTCTTGACTGCGCAGACCCTCCTTATGGACCTTGGACAGTTCCTGCTGTTCTGTCACATC AGCTCCCATCAGCATGATGGGATGGAGGCCTATGTCAAAGTGGACTCCTGCCCTGAGGAGCCACAGCTCCGGATGAAG AACAATGAGGAAGCGGAGGATTACGACGACGACCTGACTGACAGCGAAATGGACGTCGTGCGATTCGATGACGACAAC AGCCCGTCCTTCATCCAAATTAGATCAGTGGCGAAGAAGCACCCCAAGACCTGGGTGCACTACATTGCCGCCGAGGAA GAGGACTGGGACTACGCGCCGCTGGTGCTGGCGCCAGACGACAGGAGCTACAAGTCCCAGTACCTCAACAACGGGCCG CAGCGCATTGGCAGGAAGTACAAGAAAGTCCGCTTCATGGCCTACACTGATGAAACCTTCAAGACGAGGGAAGCCATC CAGCACGAGTCAGGCATCCTGGGACCGCTCCTTTACGGCGAAGTCGGGGATACCCTGCTCATCATTTTCAAGAACCAG GCATCGCGGCCCTACAACATCTACCCTCACGGGATCACAGACGTGCGCCCGCTCTACTCCCGCCGGCTGCCCAAGGGA GTGAAGCACCTGAAGGATTTTCCCATCCTGCCGGGAGAAATCTTCAAGTACAAGTGGACCGTGACTGTGGAAGATGGC CCTACCAAGTCGGACCCTCGCTGTCTGACCCGGTACTATTCCTCGTTTGTGAACATGGAGCGCGACCTGGCCTCGGGG CTGATTGGTCCGCTGCTGATCTGCTACAAGGAGTCCGTGGACCAGCGCGGGAACCAGATCATGTCCGACAAGCGCAAC GTGATCCTGTTCTCTGTCTTTGATGAAAACAGATCGTGGTACTTGACTGAGAATATCCAGCGGTTCCTGCCCAACCCA GCGGGAGTGCAACTGGAGGACCCGGAGTTCCAGGCCTCAAACATTATGCACTCTATCAACGGCTATGTGTTCGACTCG CTCCAACTGAGCGTGTGCCTGCATGAAGTGGCATACTGGTACATTCTGTCCATCGGAGCCCAGACCGACTTCCTGTCC GTGTTCTTCTCCGGATACACCTTCAAGCATAAGATGGTGTACGAGGACACTCTGACCCTCTTCCCATTTTCCGGAGAA ACTGTGTTCATGTCAATGGAAAACCCGGGCTTGTGGATTCTGGGTTGCCATAACTCGGACTTCCGGAATAGAGGGATG ACCGCCCTGCTGAAAGTGTCCAGCTGTGACAAGAATACCGGCGATTACTACGAGGACAGCTATGAGGACATCTCCGCT TATCTGCTGTCCAAGAACAACGCCATTGAACCCAGGTCCTTCTCCCAAAACGGTGCACCGGCCTCATCCCCCCCCGTG CTGAAGCGGCATCAAAGAGAGATCACCAGGACCACTCTCCAGTCCGATCAGGAAGAAATTGACTACGACGATACTATC AGCGTGGAGATGAAGAAGGAGGACTTCGACATCTACGATGAGGATGAGAACCAGTCCCCTCGGAGCTTTCAGAAGAAA ACCCGCCACTACTTCATCGCTGCCGTGGAGCGGCTGTGGGATTACGGGATGTCCAGCTCACCGCATGTGCTGCGGAAT AGAGCGCAGTCAGGATCGGTGCCCCAGTTCAAGAAGGTCGTGTTCCAAGAGTTCACCGACGGGTCCTTCACTCAACCC CTGTACCGGGGCGAACTCAACGAACACCTGGGACTGCTTGGGCCGTATATCAGGGCAGAAGTGGAAGATAACATCATG GTCACCTTCCGCAACCAGGCCTCCCGGCCGTACAGCTTCTACTCTTCACTGATCTCCTACGAGGAAGATCAGCGGCAG GGAGCCGAGCCCCGGAAGAACTTCGTCAAGCCTAACGAAACTAAGACCTACTTTTGGAAGGTCCAGCATCACATGGCC CCGACCAAAGACGAGTTCGACTGTAAAGCCTGGGCCTACTTCTCCGATGTGGACCTGGAGAAGGACGTGCACTCGGGA CTCATTGGCCCGCTCCTTGTGTGCCATACTAATACCCTGAACCCTGCTCACGGTCGCCAAGTCACAGTGCAGGAGTTC GCCCTCTTCTTCACCATCTTCGATGAAACAAAGTCCTGGTACTTTACTGAGAACATGGAACGCAATTGCAGGGCACCC TGCAACATCCAGATGGAAGATCCCACCTTCAAGGAAAACTACCGGTTTCATGCCATTAACGGCTACATAATGGACACG TTGCCAGGACTGGTCATGGCCCAGGACCAGAGAATCCGGTGGTATCTGCTCTCCATGGGCTCCAACGAAAACATTCAC AGCATTCATTTTTCCGGCCATGTGTTCACCGTCCGGAAGAAGGAAGAGTACAAGATGGCTCTGTACAACCTCTACCCT GGAGTGTTCGAGACTGTGGAAATGCTGCCTAGCAAGGCCGGCATTTGGAGAGTGGAATGCCTGATCGGAGAGCATTTG CACGCCGGAATGTCCACCCTGTTTCTTGTGTACTCCAACAAGTGCCAGACCCCGCTGGGAATGGCCTCAGGTCATATT AGGGATTTCCAGATCACTGCTTCGGGGCAGTACGGGCAGTGGGCACCTAAGTTGGCCCGGCTGCACTACTCTGGCTCC ATCAATGCCTGGTCCACCAAGGAACCCTTCTCCTGGATTAAGGTGGACCTCCTGGCCCCAATGATTATTCACGGTATT AAGACCCAGGGTGCCCGACAGAAGTTCTCCTCACTCTACATCTCGCAATTCATCATAATGTACAGCCTGGATGGGAAG AAGTGGCAGACCTACCGGGGAAACTCCACTGGAACGCTCATGGTGTTTTTCGGCAACGTGGACTCCTCCGGCATTAAG CACAACATCTTCAACCCTCCGATCATTGCTCGGTACATCCGGCTGCACCCAACTCACTACAGCATCCGGTCCACCCTG CGGATGGAACTGATGGGTTGTGACCTGAACTCCTGCTCCATGCCCCTTGGGATGGAATCCAAGGCCATTAGCGATGCA CAGATCACCGCCTCTTCATACTTCACCAACATGTTCGCGACCTGGTCCCCGTCGAAGGCCCGCCTGCACCTCCAAGGT CGCTCCAATGCGTGGCGGCCTCAAGTGAACAACCCCAAGGAGTGGCTCCAGGTCGACTTCCAAAAGACCATGAAGGTC ACCGGAGTGACCACCCAGGGCGTGAAGTCCCTGCTGACCTCTATGTACGTTAAGGAGTTCCTCATCTCCTCAAGCCAA GACGGACATCAGTGGACCCTGTTCTTCCAAAACGGAAAAGTCAAAGTATTCCAGGGCAACCAGGACTCCTTCACCCCT GTGGTCAACAGCCTGGACCCCCCATTGCTGACCCGCTACCTCCGCATCCACCCCCAAAGCTGGGTCCACCAGATCGCA CTGCGCATGGAGGTCCTTGGATGCGAAGCCCAAGATCTGTACTAA
SEQ ID NO: 34 V1.0 Expression cassette TTP-Intron-BDDFVIIIco6 XTEN (V1.0)-WPRE-bGHPolyA	ATGCAGATTGAGCTGTCCACTTGTTTCTTCCTGTGCCTCCTGCGCTTCTGTTTCTCCGCCACTCGCCGGTACTACCTT GGAGCCGTGGAGCTTTCATGGGACTACATGCAGAGCGACCTGGGCGAACTCCCCGTGGATGCCAGATTCCCCCCCCGC GTGCCAAAGTCCTTCCCCTTTAACACCTCCGTGGTGTACAAGAAAACCCTCTTTGTCGAGTTCACTGACCACCTGTTC AACATCGCCAAGCCGCGCCCACCTTGGATGGGCCTCCTGGGACCGACCATTCAAGCTGAAGTGTACGACACCGTGGTG ATCACCCTGAAGAACATGGCGTCCCACCCCGTGTCCCTGCATGCGGTCGGAGTGTCCTACTGGAAGGCCTCCGAAGGA GCTGAGTACGACGACCAGACTAGCCAGCGGGAAAAGGAGGACGATAAAGTGTTCCCGGGCGGCTCGCATACTTACGTG TGGCAAGTCCTGAAGGAAAACGGACCTATGGCATCCGATCCTCTGTGCCTGACTTACTCCTACCTTTCCCATGTGGAC CTCGTGAAGGACCTGAACAGCGGGCTGATTGGTGCACTTCTCGTGTGCCGCGAAGGTTCGCTCGCTAAGGAAAAGACC CAGACCCTCCATAAGTTCATCCTTTTGTTCGCTGTGTTCGATGAAGGAAAGTCATGGCATTCCGAAACTAAGAACTCG CTGATGCAGGACCGGGATGCCGCCTCAGCCCGCGCCTGGCCTAAAATGCATACAGTCAACGGATACGTGAATCGGTCA CTGCCCGGGCTCATCGGTTGTCACAGAAAGTCCGTGTACTGGCACGTCATCGGCATGGGCACTACGCCTGAAGTGCAC TCCATCTTCCTGGAAGGGCACACCTTCCTCGTGCGCAACCACCGCCAGGCCTCTCTGGAAATCTCCCCGATTACCTTT CTGACCGCCCAGACTCTGCTCATGGACCTGGGGCAGTTCCTTCTCTTCTGCCACATCTCCAGCCATCAGCACGACGGA ATGGAGGCCTACGTGAAGGTGGACTCATGCCCGGAAGAACCTCAGTTGCGGATGAAGAACAACGAGGAGGCCGAGGAC TATGACGACGATTTGACTGACTCCGAGATGGACGTCGTGCGGTTCGATGACGACAACAGCCCCAGCTTCATCCAGATT CGCAGCGTGGCCAAGAAGCACCCCAAAACCTGGGTGCACTACATCGCGGCCGAGGAAGAAGATTGGGACTACGCCCCG

	TTGGTGCTGGCACCCGATGACCGGTCGTACAAGTCCCAGTATCTGAACAATGGTCCGCAGCGGATTGGCAGAAAGTAC AAGAAAGTGCGGTTCATGGCGTACACTGACGAAACGTTTAAGACCCGGGAGGCCATTCAACATGAGAGCGGCATTCTG GGACCACTGCTGTACGGAGAGGTCGGCGATACCCTGCTCATCATCTTCAAAAACCAGGCCTCCCGGCCTTACAACATC TACCCTCACGGAATCACCGACGTGCGGCCACTCTACTCGCGGCGCCTGCCGAAGGGCGTCAAGCACCTGAAAGACTTC CCTATCCTGCCGGGCGAAATCTTCAAGTATAAGTGGACCGTCACCGTGGAGGACGGGCCCACCAAGAGCGATCCTAGG TGTCTGACTCGGTACTACTCCAGCTTCGTGAACATGGAACGGGACCTGGCATCGGGACTCATTGGACCGCTGCTGATC TGCTACAAAGAGTCGGTGGATCAACGCGGCAACCAGATCATGTCCGACAAGCGCAACGTGATCCTGTTCTCCGTGTTT GATGAAAACAGATCCTGGTACCTCACTGAAAACATCCAGAGGTTCCTCCCAAACCCCGCAGGAGTGCAACTGGAGGAC CCTGAGTTTCAGGCCTCGAATATCATGCACTCGATTAACGGTTACGTGTTCGACTCGCTGCAACTGAGCGTGTGCCTC CATGAAGTCGCTTACTGGTACATTCTGTCCATCGGCGCCCAGACTGACTTCCTGAGCGTGTTCTTTTCCGGTTACACC TTTAAGCACAAGATGGTGTACGAAGATACCCTGACCCTGTTCCCTTTCTCCGGCGAAACGGTGTTCATGTCGATGGAG AACCCGGGTCTGTGGATTCTGGGATGCCACAACAGCGACTTTCGGAACCGCGGAATGACTGCCCTGCTGAAGGTGTCC TCATGCGACAAGAACACCGGAGACTACTACGAGGACTCCTACGAGGATATCTCAGCCTACCTCCTGTCCAAGAACAAC GCGATCGAGCCGCGCAGCTTCAGCCAGAACGGCGCGCCAACATCAGAGAGCGCCACCCCTGAAAGTGGTCCCGGGAGC GAGCCAGCCACATCTGGGTCGGAAACGCCAGGCACAAGTGAGTCTGCAACTCCCGAGTCCGGACCTGGCTCCGAGCCT GCCACTAGCGGCTCCGAGACTCCGGGAACTTCCGAGAGCGCTACACCAGAAAGCGGACCCGGAACCAGTACCGAACCT AGCGAGGGCTCTGCTCCGGGCAGCCCAGCCGGCTCTCCTACATCCACGGAGGAGGGCACTTCCGAATCCGCCACCCCG GAGTCAGGGCCAGGATCTGAACCCGCTACCTCAGGCAGTGAGACGCCAGGAACGAGCGAGTCCGCTACACCGGAGAGT GGGCCAGGGAGCCCTGCTGGATCTCCTACGTCCACTGAGGAAGGGTCACCAGCGGGCTCGCCCACCAGCACTGAAGAA GGTGCCTCGAGCCCGCCTGTGCTGAAGAGGCACCAGCGAGAAATTACCCGGACCACCCTCCAATCGGATCAGGAGGAA ATCGACTACGACGACACCATCTCGGTGGAAATGAAGAAGGAAGATTTCGATATCTACGACGAGGACGAAAATCAGTCC CCTCGCTCATTCCAAAAGAAAACTAGACACTACTTTATCGCCGCGGTGGAAAGACTGTGGGACTATGGAATGTCATCC AGCCCTCACGTCCTTCGGAACCGGGCCCAGAGCGGATCGGTGCCTCAGTTCAAGAAAGTGGTGTTCCAGGAGTTCACC GACGGCAGCTTCACCCAGCCGCTGTACCGGGGAGAACTGAACGAACACCTGGGCCTGCTCGGTCCCTACATCCGCGCG GAAGTGGAGGATAACATCATGGTGACCTTCCGTAACCAAGCATCCAGACCTTACTCCTTCTATTCCTCCCTGATCTCA TACGAGGAGGACCAGCGCCAAGGCGCCGAGCCCCGCAAGAACTTCGTCAAGCCCAACGAGACTAAGACCTACTTCTGG AAGGTCCAACACCATATGGCCCCGACCAAGGATGAGTTTGACTGCAAGGCCTGGGCCTACTTCTCCGACGTGGACCTT GAGAAGGATGTCCATTCCGGCCTGATCGGGCCGCTGCTCGTGTGTCACACCAACACCCTGAACCCAGCGCATGGACGC CAGGTCACCGTCCAGGAGTTTGCTCTGTTCTTCACCATTTTTGACGAAACTAAGTCCTGGTACTTCACCGAGAATATG GAGCGAAACTGTAGAGCGCCCTGCAATATCCAGATGGAAGATCCGACTTTCAAGGAGAACTATAGATTCCACGCCATC AACGGGTACATCATGGATACTCTGCCGGGGCTGGTCATGGCCCAGGATCAGAGGATTCGGTGGTACTTGCTGTCAATG GGATCGAACGAAAACATTCACTCCATTCACTTCTCCGGTCACGTGTTCACTGTGCGCAAGAAGGAGGAGTACAAGATG GCGCTGTACAATCTGTACCCCGGGGTGTTCGAAACTGTGGAGATGCTGCCGTCCAAGGCCGGCATCTGGAGAGTGGAG TGCCTGATCGGAGAGCACCTCCACGCGGGGATGTCCACCCTCTTCCTGGTGTACTCGAATAAGTGCCAGACCCCGCTG GGCATGGCCTCGGGCCACATCAGAGACTTCCAGATCACAGCAAGCGGACAATACGGCCAATGGGCGCCGAAGCTGGCC CGCTTGCACTACTCCGGATCGATCAACGCATGGTCCACCAAGGAACCGTTCTCGTGGATTAAGGTGGACCTCCTGGCC CCTATGATTATCCACGGAATTAAGACCCAGGGCGCCAGGCAGAAGTTCTCCTCCCTGTACATCTCGCAATTCATCATC ATGTACAGCCTGGACGGGAAGAAGTGGCAGACTTACAGGGGAAACTCCACCGGCACCCTGATGGTCTTTTTCGGCAAC GTGGATTCCTCCGGCATTAAGCACAACATCTTCAACCCACCGATCATAGCCAGATATATTAGGCTCCACCCCACTCAC TACTCAATCCGCTCAACTCTTCGGATGGAACTCATGGGGTGCGACCTGAACTCCTGCTCCATGCCGTTGGGGATGGAA TCAAAGGCTATTAGCGACGCCCAGATCACCGCGAGCTCCTACTTCACTAACATGTTCGCCACCTGGAGCCCCTCCAAG GCCAGGCTGCACTTGCAGGGACGGTCAAATGCCTGGCGGCCGCAAGTGAACAATCCGAAGGAATGGCTTCAAGTGGAT TTCCAAAAGACCATGAAAGTGACCGGAGTCACCACCCAGGGAGTGAAGTCCCTTCTGACCTCGATGTATGTGAAGGAG TTCCTGATTAGCAGCAGCCAGGACGGGCACCAGTGGACCCTGTTCTTCCAAAACGGAAAGGTCAAGGTGTTCCAGGGG AACCAGGACTCGTTCACACCCGTGGTGAACTCCCTGGACCCCCCACTGCTGACGCGGTACTTGAGGATTCATCCTCAG TCCTGGGTCCATCAGATTGCATTGCGAATGGAAGTCCTGGGCTGCGAGGCCCAGGACCTGTACTGA
SEQ ID NO: 35 V3.0 Expression cassette Human-codon optimized A1AT-Intron-BDDFVIIIXTEN-WPRE-bGHPolyA	ATCGATGGCCCCAGGTTAATTTTTAAAAAGCAGTCAAAAGTCCAAGTGGCCCTTGGCAGCATTTACTCTCTCTGTTTG CTCTGGTTAATAATCTCAGGAGCACAAACATTCCTGGAGGCAGGAGAAGAAATCAACATCCTGGACTTATCCTCTGGG CCTCTCCCCACCTTCGATGGCCCCAGGTTAATTTTTAAAAAGCAGTCAAAAGTCCAAGTGGCCCTTGGCAGCATTTAC TCTCTCTGTTTGCTCTGGTTAATAATCTCAGGAGCACAAACATTCCTGGAGGCAGGAGAAGAAATCAACATCCTGGAC TTATCCTCTGGGCCTCTCCCCACCTTCGAACTAGCCACTAGCCTGAGGCTGGTCAAAATTGAACCTCCTCCTGCTCTG AGCAGCCTGGGGGGCAGACTAAGCAGAGGGCTGTGCAGACCCACATAAAGAGCCTACTGTGTGCCAGGCACTTCACCC GAGGCACTTCACAAGCATGCTTGGGAATGAAACTTCCAACTCTTTGGGATGCAGGTGAAACAGTTCCTGGTTCAGAGA GGTGAAGCGGCCTGCCTGAGGCAGCACAGCTCTTCTTTACAGATGTGCTTCCCCACCTCTACCCTGTCTCACGGCCCC CCATGCCAGCCTGACGGTTGTGTCTGCCTCAGTCATGCTCCATTTTTCCATCGGGACCATCAAGAGGGTGTTTGTGTC TAAGGCTGACTGGGTAACTTTGGATGAGCGGTCTCTCCGCTCTGAGCCTGTTTCCTCATCTGTCAAATGGGCTCTAAC CCACTCTGATCTCCCAGGGCGGCAGTAAGTCTTCAGCATCAGGCATTTTGGGGTGACTCAGTAAATGGTAGATCTTGC TACCAGTGGAACAGCCACTAAGGATTCTGCAGTGAGAGCAGAGGGCCAGCTAAGTGGTACTCTCCCAGAGACTGTCTG ACTCACGCCACCCCCTCCACCTTGGACACAGGACGCTGTGGTTTCTGAGCCAGGTACAATGACTCCTTTCGGTAAGTG CAGTGGAAGCTGTACACTGCCCAGGCAAAGCGTCCGGGCAGCGTAGGCGGGCGACTCAGATCCCAGCCAGTGGACTTA GCCCCTGTTTGCTCCTCCGATAACTGGGGTGACCTTGGTTAATATTCACCAGCAGCCTCCCCCGTTGCCCCTCTGGAT

CCACTGCTTAAATACGGACGAGGACAGGGCCCTGTCTCCTCAGCTTCAGGCACCACCACTGACCTGGGACAGGAATTC
TCAGGAGCACAAACATTCCTGGAGGCAGGAGAAGAAATCAACATCCTGGACTTATCCTCTGGGCCTCTCCCCACCGAT
ATCTACCTGCTGATCGCCCGGCCCCTGTTCAAACATGTCCTAATACTCTGTCGGGGCAAAGGTCGGCAGTAGTTTTCC
ATCTTACTCAACATCCTCCCAGTGTACGTAGGATCCTGTCTGTCTGCACATTTCGTAGAGCGAGTGTTCCGATACTCT
AATCTCCCGGGGCAAAGGTCGTATTGACTTAGGTTACTTATTCTCCTTTTGTTGACTAAGTCAATAATCAGAATCAGC
AGGTTTGGAGTCAGCTTGGCAGGGATCAGCAGCCTGGGTTGGAAGGAGGGGGTATAAAAGCCCCTTCACCAGGAGAAG
CCGTCACACAGATCCACAAGCTCCTGCTAGAGTCGCTGCGCGCTGCCTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTC
GCGCCGCCCGCCCCGGCTCTGACTGACCGCGTTACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGCTG
TAATTAGCGCTTGGTTTATTGACGGCTTGTTTCTTTTCTGTGGCTGCGTGAAAGCCTTGAGGGGCTCCGGGAAGGCCC
TTTGTGCGGGGGGAGCGGCTCGGGGGGTGCGTGCGTGTGTGTGTGCGTGGGGAGCGCCGCGTGCGGCTCCGCGCTGCC
CGGCGGCTGTGAGCGCTGCGGGCGCGGCGCGGGGCTTTGTGCGCTCCGCAGTGTGCGCGAGGGGAGCGCGGCCGGGGG
CGGTGCCCCGCGGTGCGGGGGGGGCTGCGAGGGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGGGGGTGAGCAGG
GGGTGTGGGCGCGTCGGTCGGGCTGCAACCCCCCCTGCACCCCCCTCCCCGAGTTGCTGAGCACGGCCCGGCTTCGGG
TGCGGGGCTCCGTACGGGGCGTGGCGCGGGGCTCGCCGTGCCGGGCGGGGGGTGGCGGCAGGTGGGGGTGCCGGGCGG
GGCGGGGCCGCCTCGGGCCGGGGAGGGCTCGGGGGAGGGGCGCGGCGGCCCCCGGAGCGCCGGCGGCTGTCGAGGCGC
GGCGAGCCGCAGCCATTGCCTTTTATGGTAATCGTGCGAGAGGGCGCAGGGACTTCCTTTGTCCCAAATCTGTGCGGA
GCCGAAATCTGGGAGGCGCCGCCGCACCCCCTCTAGCGGGCGCGGGGCGAAGCGGTGCGGCGCCGGCAGGAAGGAAAT
GGGCGGGGAGGGCCTTCGTGCGTCGCCGCGCCGCCGTCCCCTTCTCCCTCTCCAGCCTCGGGGCTGTCCGCGGGGGGA
CGGCTGCCTTCGGGGGGGACGGGGCAGGGCGGGGTTCGGCTTCTGGCGTGTGACCGGCGGCTCTAGAGCCTCTGCTAA
CCTTGTTCTTGCCTTCTTCTTTTTCCTACAGCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCATCATTTTGGCAA
AGAATTACTCGAGGCCACCATGCAGATTGAACTGTCCACTTGCTTCTTCCTGTGCCTCCTGCGGTTTTGCTTCTCGGC
CACCCGCCGGTATTACTTAGGTGCTGTGGAACTGAGCTGGGACTACATGCAGTCCGACCTGGGAGAACTGCCGGTGGA
CGCGAGATTCCCACCTAGAGTCCCGAAGTCCTTCCCATTCAACACCTCCGTGGTCTACAAAAAGACCCTGTTCGTGGA
GTTCACTGACCACCTTTTCAATATTGCCAAGCCGCGCCCCCCCTGGATGGGCCTGCTTGGTCCTACGATCCAAGCAGA
GGTCTACGACACCGTGGTCATCACACTGAAGAACATGGCCTCACACCCCGTGTCGCTGCATGCTGTGGGAGTGTCCTA
CTGGAAGGCCTCAGAGGGTGCCGAATATGATGACCAGACCAGCCAGAGGGAAAAGGAGGATGACAAAGTGTTCCCGGG
TGGCAGCCACACTTACGTGTGGCAAGTGCTGAAGGAAAACGGGCCTATGGCGTCGGACCCCCTATGCCTGACCTACTC
CTACCTGTCCCATGTGGACCTTGTGAAGGATCTCAACTCGGGACTGATCGGCGCCCTCTTGGTGTGCAGAGAAGGCAG
CCTGGCGAAGGAAAAGACTCAGACCCTGCACAAGTTCATTCTGTTGTTTGCTGTGTTCGATGAAGGAAAGTCCTGGCA
CTCAGAAACCAAGAACTCGCTGATGCAGGATAGAGATGCGGCCTCGGCCAGAGCCTGGCCTAAAATGCACACCGTCAA
CGGATATGTGAACAGGTCGCTCCCTGGCCTCATCGGCTGCCACAGAAAGTCCGTGTATTGGCATGTGATCGGCATGGG
TACTACTCCGGAAGTGCATAGTATCTTTCTGGAGGGCCATACCTTCTTGGTGCGCAACCACAGACAGGCCTCGCTGGA
AATCTCGCCTATCACTTTCTTGACTGCGCAGACCCTCCTTATGGACCTTGGACAGTTCCTGCTGTTCTGTCACATCAG
CTCCCATCAGCATGATGGGATGGAGGCCTATGTCAAAGTGGACTCCTGCCCTGAGGAGCCACAGCTCCGGATGAAGAA
CAATGAGGAAGCGGAGGATTACGACGACGACCTGACTGACAGCGAAATGGACGTCGTGCGATTCGATGACGACAACAG
CCCGTCCTTCATCCAAATTAGATCAGTGGCGAAGAAGCACCCCAAGACCTGGGTGCACTACATTGCCGCCGAGGAAGA
GGACTGGGACTACGCGCCGCTGGTGCTGGCGCCAGACGACAGGAGCTACAAGTCCCAGTACCTCAACAACGGGCCGCA
GCGCATTGGCAGGAAGTACAAGAAAGTCCGCTTCATGGCCTACACTGATGAAACCTTCAAGACGAGGGAAGCCATCCA
GCACGAGTCAGGCATCCTGGGACCGCTCCTTTACGGCGAAGTCGGGGATACCCTGCTCATCATTTTCAAGAACCAGGC
ATCGCGGCCCTACAACATCTACCCTCACGGGATCACAGACGTGCGCCCGCTCTACTCCCGCCGGCTGCCCAAGGGAGT
GAAGCACCTGAAGGATTTTCCCATCCTGCCGGGAGAAATCTTCAAGTACAAGTGGACCGTGACTGTGGAAGATGGCCC
TACCAAGTCGGACCCTCGCTGTCTGACCCGGTACTATTCCTCGTTTGTGAACATGGAGCGCGACCTGGCCTCGGGGCT
GATTGGTCCGCTGCTGATCTGCTACAAGGAGTCCGTGGACCAGCGCGGGAACCAGATCATGTCCGACAAGCGCAACGT
GATCCTGTTCTCTGTCTTTGATGAAAACAGATCGTGGTACTTGACTGAGAATATCCAGCGGTTCCTGCCCAACCCAGC
GGGAGTGCAACTGGAGGACCCGGAGTTCCAGGCCTCAAACATTATGCACTCTATCAACGGCTATGTGTTCGACTCGCT
CCAACTGAGCGTGTGCCTGCATGAAGTGGCATACTGGTACATTCTGTCCATCGGAGCCCAGACCGACTTCCTGTCCGT
GTTCTTCTCCGGATACACCTTCAAGCATAAGATGGTGTACGAGGACACTCTGACCCTCTTCCCATTTTCCGGAGAAAC
TGTGTTCATGTCAATGGAAAACCCGGGCTTGTGGATTCTGGGTTGCCATAACTCGGACTTCCGGAATAGAGGGATGAC
CGCCCTGCTGAAAGTGTCCAGCTGTGACAAGAATACCGGCGATTACTACGAGGACAGCTATGAGGACATCTCCGCTTA
TCTGCTGTCCAAGAACAACGCCATTGAACCCAGGTCCTTCTCCCAAAACGGTGCACCGACCTCCGAAAGCGCCACCCC
AGAGTCAGGACCTGGCTCGGAACCGGCTACCTCGGGCTCAGAGACACCGGGGACTTCCGAGTCCGCAACCCCCGAGAG
TGGACCCGGATCCGAACCAGCAACCTCAGGATCAGAAACCCCGGGAACTTCGGAATCCGCCACTCCCGAGTCGGGACC
AGGCACCTCCACTGAGCCTTCCGAGGGAAGCGCCCCCGGATCCCCTGCTGGATCCCCTACCAGCACTGAAGAAGGCAC
CTCAGAATCCGCGACCCCTGAGTCCGGCCCTGGAAGCGAACCCGCCACCTCCGGTTCCGAAACCCCTGGGACTAGCGA
GAGCGCCACTCCGGAATCGGGCCCAGGAAGCCCTGCCGGATCCCCGACCAGCACCGAGGAGGGAAGCCCCGCCGGGTC
ACCGACTTCCACTGAGGAGGGAGCCTCATCCCCCCCCGTGCTGAAGCGGCATCAAAGAGAGATCACCAGGACCACTCT
CCAGTCCGATCAGGAAGAAATTGACTACGACGATACTATCAGCGTGGAGATGAAGAAGGAGGACTTCGACATCTACGA
TGAGGATGAGAACCAGTCCCCTCGGAGCTTTCAGAAGAAAACCCGCCACTACTTCATCGCTGCCGTGGAGCGGCTGTG
GGATTACGGGATGTCCAGCTCACCGCATGTGCTGCGGAATAGAGCGCAGTCAGGATCGGTGCCCCAGTTCAAGAAGGT
CGTGTTCCAAGAGTTCACCGACGGGTCCTTCACTCAACCCCTGTACCGGGGCGAACTCAACGAACACCTGGGACTGCT
TGGGCCGTATATCAGGGCAGAAGTGGAAGATAACATCATGGTCACCTTCCGCAACCAGGCCTCCCGGCCGTACAGCTT
CTACTCTTCACTGATCTCCTACGAGGAAGATCAGCGGCAGGGAGCCGAGCCCCGGAAGAACTTCGTCAAGCCTAACGA

	AACTAAGACCTACTTTTGGAAGGTCCAGCATCACATGGCCCCGACCAAAGACGAGTTCGACTGTAAAGCCTGGGCCTA CTTCTCCGATGTGGACCTGGAGAAGGACGTGCACTCGGGACTCATTGGCCCGCTCCTTGTGTGCCATACTAATACCCT GAACCCTGCTCACGGTCGCCAAGTCACAGTGCAGGAGTTCGCCCTCTTCTTCACCATCTTCGATGAAACAAAGTCCTG GTACTTTACTGAGAACATGGAACGCAATTGCAGGGCACCCTGCAACATCCAGATGGAAGATCCCACCTTCAAGGAAAA CTACCGGTTTCATGCCATTAACGGCTACATAATGGACACGTTGCCAGGACTGGTCATGGCCCAGGACCAGAGAATCCG GTGGTATCTGCTCTCCATGGGCTCCAACGAAAACATTCACAGCATTCATTTTTCCGGCCATGTGTTCACCGTCCGGAA GAAGGAAGAGTACAAGATGGCTCTGTACAACCTCTACCCTGGAGTGTTCGAGACTGTGGAAATGCTGCCTAGCAAGGC CGGCATTTGGAGAGTGGAATGCCTGATCGGAGAGCATTTGCACGCCGGAATGTCCACCCTGTTTCTTGTGTACTCCAA CAAGTGCCAGACCCCGCTGGGAATGGCCTCAGGTCATATTAGGGATTTCCAGATCACTGCTTCGGGGCAGTACGGGCA GTGGGCACCTAAGTTGGCCCGGCTGCACTACTCTGGCTCCATCAATGCCTGGTCCACCAAGGAACCCTTCTCCTGGAT TAAGGTGGACCTCCTGGCCCCAATGATTATTCACGGTATTAAGACCCAGGGTGCCCGACAGAAGTTCTCCTCACTCTA CATCTCGCAATTCATCATAATGTACAGCCTGGATGGGAAGAAGTGGCAGACCTACCGGGGAAACTCCACTGGAACGCT CATGGTGTTTTTCGGCAACGTGGACTCCTCCGGCATTAAGCACAACATCTTCAACCCTCCGATCATTGCTCGGTACAT CCGGCTGCACCCAACTCACTACAGCATCCGGTCCACCCTGCGGATGGAACTGATGGGTTGTGACCTGAACTCCTGCTC CATGCCCCTTGGGATGGAATCCAAGGCCATTAGCGATGCACAGATCACCGCCTCTTCATACTTCACCAACATGTTCGC GACCTGGTCCCCGTCGAAGGCCCGCCTGCACCTCCAAGGTCGCTCCAATGCGTGGCGGCCTCAAGTGAACAACCCCAA GGAGTGGCTCCAGGTCGACTTCCAAAAGACCATGAAGGTCACCGGAGTGACCACCCAGGGCGTGAAGTCCCTGCTGAC CTCTATGTACGTTAAGGAGTTCCTCATCTCCTCAAGCCAAGACGGACATCAGTGGACCCTGTTCTTCCAAAACGGAAA AGTCAAAGTATTCCAGGGCAACCAGGACTCCTTCACCCCTGTGGTCAACAGCCTGGACCCCCCATTGCTGACCCGCTA CCTCCGCATCCACCCCCAAAGCTGGGTCCACCAGATCGCACTGCGCATGGAGGTCCTTGGATGCGAAGCCCAAGATCT GTACTAAGCGGCCGCTCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGC TCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTC CTCCTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCAC TGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCC CCTCCCTATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGA CAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGG GACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCC TCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCTGCCTAGGCGACTGTG CCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTC CTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAG GACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGAAGACCATGGGCGCGCCAGGCCTGTCGACGCC CGGGCGGTACCGCGATCGCTCGCGACGCATAAAG
SEQ ID NO: 36 human liver-specific alpha-1-antitrypsin (A1AT) promoter	ATCGATGGCCCCAGGTTAATTTTTAAAAAGCAGTCAAAAGTCCAAGTGGCCCTTGGCAGCATTTACTCTCTCTGTTTG CTCTGGTTAATAATCTCAGGAGCACAAACATTCCTGGAGGCAGGAGAAGAAATCAACATCCTGGACTTATCCTCTGGG CCTCTCCCCACCTTCGATGGCCCCAGGTTAATTTTTAAAAAGCAGTCAAAAGTCCAAGTGGCCCTTGGCAGCATTTAC TCTCTCTGTTTGCTCTGGTTAATAATCTCAGGAGCACAAACATTCCTGGAGGCAGGAGAAGAAATCAACATCCTGGAC TTATCCTCTGGGCCTCTCCCCACCTTCGAACTAGCCACTAGCCTGAGGCTGGTCAAAATTGAACCTCCTCCTGCTCTG AGCAGCCTGGGGGGCAGACTAAGCAGAGGGCTGTGCAGACCCACATAAAGAGCCTACTGTGTGCCAGGCACTTCACCC GAGGCACTTCACAAGCATGCTTGGGAATGAAACTTCCAACTCTTTGGGATGCAGGTGAAACAGTTCCTGGTTCAGAGA GGTGAAGCGGCCTGCCTGAGGCAGCACAGCTCTTCTTTACAGATGTGCTTCCCCACCTCTACCCTGTCTCACGGCCCC CCATGCCAGCCTGACGGTTGTGTCTGCCTCAGTCATGCTCCATTTTTCCATCGGGACCATCAAGAGGGTGTTTGTGTC TAAGGCTGACTGGGTAACTTTGGATGAGCGGTCTCTCCGCTCTGAGCCTGTTTCCTCATCTGTCAAATGGGCTCTAAC CCACTCTGATCTCCCAGGGCGGCAGTAAGTCTTCAGCATCAGGCATTTTGGGGTGACTCAGTAAATGGTAGATCTTGC TACCAGTGGAACAGCCACTAAGGATTCTGCAGTGAGAGCAGAGGGCCAGCTAAGTGGTACTCTCCCAGAGACTGTCTG ACTCACGCCACCCCCTCCACCTTGGACACAGGACGCTGTGGTTTCTGAGCCAGGTACAATGACTCCTTTCGGTAAGTG CAGTGGAAGCTGTACACTGCCCAGGCAAAGCGTCCGGGCAGCGTAGGCGGGCGACTCAGATCCCAGCCAGTGGACTTA GCCCCTGTTTGCTCCTCCGATAACTGGGGTGACCTTGGTTAATATTCACCAGCAGCCTCCCCCGTTGCCCCTCTGGAT CCACTGCTTAAATACGGACGAGGACAGGGCCCTGTCTCCTCAGCTTCAGGCACCACCACTGACCTGGGACAG

CLAIMS

1. An isolated nucleic acid molecule comprising a nucleotide sequence at least 85% identical to SEQ ID NO: 9, wherein the nucleotide sequence encodes a polypeptide with factor VIII (FVIII) activity.