

VIRAL VECTORS ENCODING RECOMBINANT FVIII VARIANTS WITH INCREASED EXPRESSION FOR GENE THERAPY OF HEMOPHILIA A

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to United States Provisional Patent Application No. 62/255,323, filed November 13, 2015, the content of which is hereby incorporated by reference in its entirety for all purposes.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on November 7, 2016, is named 008073_5115_WO_Sequence_Listing.txt and is 183,311 bytes in size.

BACKGROUND OF THE DISCLOSURE

[0003] Blood coagulation proceeds through a complex and dynamic biological pathway of interdependent biochemical reactions, referred to as the coagulation cascade. Coagulation Factor VIII (FVIII) is a key component in the cascade. Factor VIII is recruited to bleeding sites, and forms a Xase complex with activated Factor IX (FIXa) and Factor X (FX). The Xase complex activates FX, which in turn activates prothrombin to thrombin, which then activates other components in the coagulation cascade to generate a stable clot (reviewed in Saenko *et al.*, *Trends Cardiovasc. Med.*, 9:185-192 (1999); Lenting *et al.*, *Blood*, 92:3983-3996 (1998)).

[0004] Hemophilia A is a congenital X-linked bleeding disorder characterized by a deficiency in Factor VIII activity. Diminished Factor VIII activity inhibits a positive feedback loop in the coagulation cascade. This causes incomplete coagulation, which manifests as bleeding episodes with increased duration, extensive bruising, spontaneous oral and nasal bleeding, joint stiffness and chronic pain, and possibly internal bleeding and anemia in severe cases (Zhang *et al.*, *Clinic. Rev. Allerg. Immunol.*, 37:114-124 (2009)).

[0005] Conventionally, hemophilia A is treated by Factor VIII replacement therapy, which consists of administering Factor VIII protein (e.g., plasma-derived or recombinantly-produced Factor VIII) to an individual with hemophilia A. Factor VIII is administered prophylactically to prevent or reduce frequency of bleeding episodes, in response to an acute bleeding episode, and/or perioperatively to manage bleeding during surgery. However, there are several undesirable features of Factor VIII replacement therapy.

[0006] First, Factor VIII replacement therapy is used to treat or manage hemophilia A, but does not cure the underlying Factor VIII deficiency. Because of this, individuals with hemophilia A require Factor VIII replacement therapy for the duration of their lives. Continuous treatment is expensive and requires the individual to maintain strict compliance, as missing only a few prophylactic doses can have serious consequences for individuals with severe hemophilia A.

[0007] Second, because Factor VIII has a relatively short half-life in vivo, conventional prophylactic Factor VIII replacement therapy requires administration every second or third day. This places a burden on the individual to maintain compliance throughout their life. While third generation “long-acting” Factor VIII drugs may reduce the frequency of administration, prophylactic Factor FVIII replacement therapy with these drugs still requires monthly, weekly, or more frequent administration in perpetuity. For example, prophylactic treatment with ELOCTATE™ [Antihemophilic Factor (Recombinant), Fc Fusion Protein] requires administration every three to five days (ELOCTATE™ Prescribing Information, Biogen Idec Inc., (2015)). Moreover, the long-term effects of chemically modified biologics (e.g., pegylated polypeptides) are not yet fully understood.

[0008] Third, between 15% and 30% of all individuals receiving Factor VIII replacement therapy form anti-Factor VIII inhibitor antibodies, rendering the therapy inefficient. Factor VIII bypass therapy (e.g., administration of plasma-derived or recombinantly-produced prothrombin complex concentrates) can be used to treat hemophilia in individuals that form inhibitor antibodies. However, Factor VIII bypass therapy is less effective than Factor VIII replacement therapy (Mannucci P.M., J Thromb Haemost., 1(7):1349-55 (2003)) and may be associated with an increased risk of cardiovascular complication (Luu and Ewenstein, Haemophilia, 10 Suppl. 2:10-16 (2004)).

[0009] Somatic gene therapy holds great promise for the treatment of hemophilia A because it would remedy the underlying under-expression functional Factor VIII activity (e.g., due to missense or nonsense mutations), rather than provide a one-time dose of Factor VIII activity to the individual. Because of this difference in the mechanism of action, as compared to Factor VIII replacement therapy, one-time administration of a Factor VIII gene therapy vector may provide an individual with Factor VIII for several years, reducing the cost of treatment and eliminating the need for continued patient compliance.

[0010] Coagulation Factor IX (FIX) gene therapy has been used effectively to treat individuals with hemophilia B, a related blood coagulation condition characterized by diminished Factor IX activity (Manno C.S., et al., Nat Med., 12(3):342-47 (2006)). However, Factor VIII gene therapy presents several unique challenges. For example, the full-length, wild-type Factor VIII polypeptide (2351 amino acids; UniProt accession number P00451) is five times larger than the full-length, wild-type Factor IX polypeptide (461 amino acids; UniProt accession number P00740). As such, the coding sequence of wild-type Factor VIII is 7053 base pairs, which is too large to be packaged in conventional AAV gene therapy vectors. Further, reported recombinant expression of B-domain deleted variants of Factor VIII (BDD-FVIII) has been poor. As such, several groups have attempted to alter the codon usage of BDD-FVIII constructs, with limited success.

SUMMARY

[0011] Accordingly, there is a need for Factor VIII variants whose coding sequences are more efficiently packaged into, and delivered via, gene therapy vectors. There is also a need for synthetic, codon-altered nucleic acids which express Factor VIII more efficiently. Such Factor VIII variants and codon-altered nucleic acids allow for improved treatment of Factor VIII deficiencies (e.g., hemophilia A). The above deficiencies and other problems associated with the treatment of Factor VIII deficiencies (e.g., hemophilia A) are reduced or eliminated by the disclosed codon-altered Factor VIII variants.

[0012] In accordance with some embodiments, the present disclosure provides nucleic acids encoding Factor VIII variants that have high sequence identity to the disclosed codon-altered sequences of the Factor VIII heavy chain (e.g., CS01-HC-NA, CS04-HC-NA, or CS23-HC-NA)

and light chain (CS01-LC-NA, CS04-LC-NA, or CS23-LC-NA). In some embodiments, these nucleic acids further include a sequence encoding a linker sequence that replaces the native Factor VIII B-domain (e.g., a linker sequences comprising a furin cleavage site), between the sequences coding for the Factor VIII heavy and light chains.

[0013] In one embodiment, the disclosure provides a polynucleotide including a nucleotide sequence encoding a Factor VIII polypeptide. The Factor VIII polypeptide includes a light chain, a heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having at least 95% identity to CS04-HC-NA (SEQ ID NO: 3). The light chain of the Factor FVIII polypeptide is encoded by a second nucleotide sequence having at least 95% identity to CS04-LC-NA (SEQ ID NO: 4). The polypeptide linker comprises a furin cleavage site.

[0014] In one embodiment of the polynucleotides described above, the polypeptide linker is encoded by a third nucleotide sequence having at least 95% identity to BDLO04 (SEQ ID NO: 6).

[0015] In one embodiment, the disclosure provides a polynucleotide including a nucleotide sequence encoding a Factor VIII polypeptide. The Factor VIII polypeptide includes a light chain, a heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having at least 95% identity to CS01-HC-NA (SEQ ID NO: 24). The light chain of the Factor FVIII polypeptide is encoded by a second nucleotide sequence having at least 95% identity to CS01-LC-NA (SEQ ID NO: 25). The polypeptide linker comprises a furin cleavage site.

[0016] In one embodiment of the polynucleotides described above, the polypeptide linker is encoded by a third nucleotide sequence having at least 95% identity to BDLO01 (SEQ ID NO: 5).

[0017] In one embodiment, the disclosure provides a polynucleotide including a nucleotide sequence encoding a Factor VIII polypeptide. The Factor VIII polypeptide includes a light chain, a heavy

chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having at least 95% identity to CS23-HC-NA (SEQ ID NO: 22). The light chain of the Factor FVIII polypeptide is encoded by a second nucleotide sequence having at least 95% identity to CS23-LC-NA (SEQ ID NO: 23). The polypeptide linker comprises a furin cleavage site.

[0018] In one embodiment of the polynucleotides described above, the polypeptide linker is encoded by a third nucleotide sequence having at least 95% identity to BDLO23 (SEQ ID NO: 7).

[0019] In one embodiment of the polynucleotides described above, the first nucleotide sequence encoding the heavy chain of the Factor VIII polypeptide has at least 96% identity to the respective heavy chain sequence (e.g., CS04-HC-NA (SEQ ID NO: 3), CS01-HC-NA (SEQ ID NO: 24), or CS23-HC-NA (SEQ ID NO: 22)), and the second nucleotide sequence encoding the light chain of the Factor FVIII polypeptide has at least 96% identity to the respective light chain sequence (e.g., CS04-LC-NA (SEQ ID NO: 4), CS01-LC-NA (SEQ ID NO: 25), or CS23-LC-NA (SEQ ID NO: 23)).

[0020] In one embodiment of the polynucleotides described above, the first nucleotide sequence encoding the heavy chain of the Factor VIII polypeptide has at least 97% identity to the respective heavy chain sequence (e.g., CS04-HC-NA (SEQ ID NO: 3), CS01-HC-NA (SEQ ID NO: 24), or CS23-HC-NA (SEQ ID NO: 22)), and the second nucleotide sequence encoding the light chain of the Factor FVIII polypeptide has at least 97% identity to the respective light chain sequence (e.g., CS04-LC-NA (SEQ ID NO: 4), CS01-LC-NA (SEQ ID NO: 25), or CS23-LC-NA (SEQ ID NO: 23)).

[0021] In one embodiment of the polynucleotides described above, the first nucleotide sequence encoding the heavy chain of the Factor VIII polypeptide has at least 98% identity to the respective heavy chain sequence (e.g., CS04-HC-NA (SEQ ID NO: 3), CS01-HC-NA (SEQ ID NO: 24), or CS23-HC-NA (SEQ ID NO: 22)), and the second nucleotide sequence encoding the light chain of the Factor FVIII polypeptide has at least 98% identity to the respective light chain

sequence (e.g., CS04-LC-NA (SEQ ID NO: 4), CS01-LC-NA (SEQ ID NO: 25), or CS23-LC-NA (SEQ ID NO: 23)).

[0022] In one embodiment of the polynucleotides described above, the first nucleotide sequence encoding the heavy chain of the Factor VIII polypeptide has at least 99% identity to the respective heavy chain sequence (e.g., CS04-HC-NA (SEQ ID NO: 3), CS01-HC-NA (SEQ ID NO: 24), or CS23-HC-NA (SEQ ID NO: 22)), and the second nucleotide sequence encoding the light chain of the Factor FVIII polypeptide has at least 99% identity to the respective light chain sequence (e.g., CS04-LC-NA (SEQ ID NO: 4), CS01-LC-NA (SEQ ID NO: 25), or CS23-LC-NA (SEQ ID NO: 23)).

[0023] In one embodiment of the polynucleotides described above, the first nucleotide sequence encoding the heavy chain of the Factor VIII polypeptide has at least 99.5% identity to the respective heavy chain sequence (e.g., CS04-HC-NA (SEQ ID NO: 3), CS01-HC-NA (SEQ ID NO: 24), or CS23-HC-NA (SEQ ID NO: 22)), and the second nucleotide sequence encoding the light chain of the Factor FVIII polypeptide has at least 99.5% identity to the respective light chain sequence (e.g., CS04-LC-NA (SEQ ID NO: 4), CS01-LC-NA (SEQ ID NO: 25), or CS23-LC-NA (SEQ ID NO: 23)).

[0024] In one embodiment of the polynucleotides described above, the first nucleotide sequence encoding the heavy chain of the Factor VIII polypeptide has at least 99.9% identity to the respective heavy chain sequence (e.g., CS04-HC-NA (SEQ ID NO: 3), CS01-HC-NA (SEQ ID NO: 24), or CS23-HC-NA (SEQ ID NO: 22)), and the second nucleotide sequence encoding the light chain of the Factor FVIII polypeptide has at least 99.9% identity to the respective light chain sequence (e.g., CS04-LC-NA (SEQ ID NO: 4), CS01-LC-NA (SEQ ID NO: 25), or CS23-LC-NA (SEQ ID NO: 23)).

[0025] In one embodiment of the polynucleotides described above, the first nucleotide sequence encoding the heavy chain of the Factor VIII polypeptide is CS04-HC-NA (SEQ ID NO: 3), and the second nucleotide sequence encoding the light chain of the Factor FVIII polypeptide is CS04-LC-NA (SEQ ID NO: 4).

[0026] In one embodiment of the polynucleotides described above, the first nucleotide sequence encoding the heavy chain of the Factor VIII polypeptide is CS01-HC-NA (SEQ ID NO: 24), and the second nucleotide sequence encoding the light chain of the Factor FVIII polypeptide is CS01-LC-NA (SEQ ID NO: 25).

[0027] In one embodiment of the polynucleotides described above, the first nucleotide sequence encoding the heavy chain of the Factor VIII polypeptide is CS23-HC-NA (SEQ ID NO: 22), and the second nucleotide sequence encoding the light chain of the Factor FVIII polypeptide is CS23-LC-NA (SEQ ID NO: 23).

[0028] In one embodiment, the disclosure provides a polynucleotide comprising a nucleotide sequence having at least 95% identity to CS04-FL-NA, wherein the polynucleotide encodes a Factor VIII polypeptide.

[0029] In one embodiment, the disclosure provides a polynucleotide comprising a nucleotide sequence having at least 95% identity to CS01-FL-NA, wherein the polynucleotide encodes a Factor VIII polypeptide.

[0030] In one embodiment, the disclosure provides a polynucleotide comprising a nucleotide sequence having at least 95% identity to CS23-FL-NA, wherein the polynucleotide encodes a Factor VIII polypeptide.

[0031] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 96% identity to the respective full-length polynucleotide sequence (e.g., CS04-FL-NA (SEQ ID NO: 1), CS01-FL-NA (SEQ ID NO: 13), or CS23-FL-NA (SEQ ID NO: 20)).

[0032] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 97% identity to the respective full-length polynucleotide sequence (e.g., CS04-FL-NA (SEQ ID NO: 1), CS01-FL-NA (SEQ ID NO: 13), or CS23-FL-NA (SEQ ID NO: 20)).

[0033] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 98% identity to the respective full-length polynucleotide sequence (e.g., CS04-FL-NA (SEQ ID NO: 1), CS01-FL-NA (SEQ ID NO: 13), or CS23-FL-NA (SEQ ID NO: 20)).

[0034] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 99% identity to the respective full-length polynucleotide sequence (e.g., CS04-FL-NA (SEQ ID NO: 1), CS01-FL-NA (SEQ ID NO: 13), or CS23-FL-NA (SEQ ID NO: 20)).

[0035] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 99.5% identity to the respective full-length polynucleotide sequence (e.g., CS04-FL-NA (SEQ ID NO: 1), CS01-FL-NA (SEQ ID NO: 13), or CS23-FL-NA (SEQ ID NO: 20)).

[0036] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 99.9% identity to the respective full-length polynucleotide sequence (e.g., CS04-FL-NA (SEQ ID NO: 1), CS01-FL-NA (SEQ ID NO: 13), or CS23-FL-NA (SEQ ID NO: 20)).

[0037] In one embodiment of the polynucleotides described above, the nucleotide sequence is CS04-FL-NA (SEQ ID NO: 1).

[0037a] In a first aspect, there is provided a polynucleotide comprising the nucleotide sequence of SEQ ID NO: 1, wherein the polynucleotide encodes a Factor VIII polypeptide.

[0038] In one embodiment of the polynucleotides described above, the nucleotide sequence is CS01-FL-NA (SEQ ID NO: 13).

[0039] In one embodiment of the polynucleotides described above, the nucleotide sequence is CS23-FL-NA (SEQ ID NO: 20).

[0040] In one embodiment of the polynucleotides described above, the polynucleotide encodes a Factor VIII polypeptide comprising an amino acid sequence having at least 95% identity to CS04-FL-AA (SEQ ID NO: 2).

[0041] In one embodiment of the polynucleotides described above, the polynucleotide encodes a Factor VIII polypeptide comprising an amino acid sequence having at least 96% identity to CS04-FL-AA (SEQ ID NO: 2).

[0042] In one embodiment of the polynucleotides described above, the polynucleotide encodes a Factor VIII polypeptide comprising an amino acid sequence having at least 97% identity to CS04-FL-AA (SEQ ID NO: 2).

[0043] In one embodiment of the polynucleotides described above, the polynucleotide encodes a Factor VIII polypeptide comprising an amino acid sequence having at least 98% identity to CS04-FL-AA (SEQ ID NO: 2).

[0044] In one embodiment of the polynucleotides described above, the polynucleotide encodes a Factor VIII polypeptide comprising an amino acid sequence having at least 99% identity to CS04-FL-AA (SEQ ID NO: 2).

[0045] In one embodiment of the polynucleotides described above, the polynucleotide encodes a Factor VIII polypeptide comprising an amino acid sequence having at least 99.5% identity to CS04-FL-AA (SEQ ID NO: 2).

[0046] In one embodiment of the polynucleotides described above, the polynucleotide encodes a Factor VIII polypeptide comprising an amino acid sequence having at least 99.9% identity to CS04-FL-AA (SEQ ID NO: 2).

[0047] In one embodiment of the polynucleotides described above, the polynucleotide encodes a Factor VIII polypeptide comprising the amino acid sequence of CS04-FL-AA (SEQ ID NO: 2).

[0048] In one embodiment, the disclosure provides a polynucleotide comprising a nucleotide sequence having at least 95% identity to CS04-SC1-NA (SEQ ID NO: 9), wherein the polynucleotide encodes a single-chain Factor VIII polypeptide.

[0049] In one embodiment, the disclosure provides a polynucleotide comprising a nucleotide sequence having at least 95% identity to CS04-SC2-NA (SEQ ID NO: 11), wherein the polynucleotide encodes a single-chain Factor VIII polypeptide.

[0050] In one embodiment, the disclosure provides a polynucleotide comprising a nucleotide sequence having at least 95% identity to CS01-SC1-NA (SEQ ID NO: 26), wherein the polynucleotide encodes a single-chain Factor VIII polypeptide.

[0051] In one embodiment, the disclosure provides a polynucleotide comprising a nucleotide sequence having at least 95% identity to CS01-SC2-NA (SEQ ID NO: 27), wherein the polynucleotide encodes a single-chain Factor VIII polypeptide.

[0052] In one embodiment, the disclosure provides a polynucleotide comprising a nucleotide sequence having at least 95% identity to CS23-SC1-NA (SEQ ID NO: 28), wherein the polynucleotide encodes a single-chain Factor VIII polypeptide.

[0053] In one embodiment, the disclosure provides a polynucleotide comprising a nucleotide sequence having at least 95% identity to CS23-SC2-NA (SEQ ID NO: 29), wherein the polynucleotide encodes a single-chain Factor VIII polypeptide.

[0054] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 96% identity to the respective full-length polynucleotide sequence (e.g., CS04-SC1-NA (SEQ ID NO: 9), CS04-SC2-NA (SEQ ID NO: 11), CS01-SC1-NA (SEQ ID NO: 26), CS01-SC2-NA (SEQ ID NO: 27), CS23-SC1-NA (SEQ ID NO: 28), or CS23-SC2-NA (SEQ ID NO: 29)).

[0055] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 97% identity to the respective full-length polynucleotide sequence (e.g., CS04-SC1-NA (SEQ ID NO: 9), CS04-SC2-NA (SEQ ID NO: 11), CS01-SC1-NA (SEQ ID NO: 26), CS01-SC2-NA (SEQ ID NO: 27), CS23-SC1-NA (SEQ ID NO: 28), or CS23-SC2-NA (SEQ ID NO: 29)).

[0056] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 98% identity to the respective full-length polynucleotide sequence (e.g., CS04-SC1-NA (SEQ ID NO: 9), CS04-SC2-NA (SEQ ID NO: 11), CS01-SC1-NA (SEQ ID NO: 26), CS01-SC2-NA (SEQ ID NO: 27), CS23-SC1-NA (SEQ ID NO: 28), or CS23-SC2-NA (SEQ ID NO: 29)).

[0057] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 99% identity to the respective full-length polynucleotide sequence (e.g., CS04-SC1-NA (SEQ ID NO: 9), CS04-SC2-NA (SEQ ID NO: 11), CS01-SC1-NA (SEQ ID NO: 26), CS01-SC2-NA (SEQ ID NO: 27), CS23-SC1-NA (SEQ ID NO: 28), or CS23-SC2-NA (SEQ ID NO: 29)).

[0058] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 99.5% identity to the respective full-length polynucleotide sequence (e.g., CS04-SC1-NA (SEQ ID NO: 9), CS04-SC2-NA (SEQ ID NO: 11), CS01-SC1-NA (SEQ ID NO: 26), CS01-SC2-NA (SEQ ID NO: 27), CS23-SC1-NA (SEQ ID NO: 28), or CS23-SC2-NA (SEQ ID NO: 29)).

[0059] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 99.9% identity to the respective full-length polynucleotide sequence (e.g., CS04-SC1-NA (SEQ ID NO: 9), CS04-SC2-NA (SEQ ID NO: 11), CS01-SC1-NA (SEQ ID NO: 26), CS01-SC2-NA (SEQ ID NO: 27), CS23-SC1-NA (SEQ ID NO: 28), or CS23-SC2-NA (SEQ ID NO: 29)).

[0060] In one embodiment of the polynucleotides described above, the nucleotide sequence is CS04-SC1-NA (SEQ ID NO: 9).

[0061] In one embodiment of the polynucleotides described above, the nucleotide sequence is CS04-SC2-NA (SEQ ID NO: 11).

[0062] In one embodiment of the polynucleotides described above, the nucleotide sequence is CS01-SC1-NA (SEQ ID NO: 26).

[0063] In one embodiment of the polynucleotides described above, the nucleotide sequence is CS01-SC2-NA (SEQ ID NO: 27).

[0064] In one embodiment of the polynucleotides described above, the nucleotide sequence is CS23-SC1-NA (SEQ ID NO: 28).

[0065] In one embodiment of the polynucleotides described above, the nucleotide sequence is CS23-SC2-NA (SEQ ID NO: 29).

[0066] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 95% identity to a sequence selected from the group consisting of CS01-FL-NA, CS01-HC-NA, CS01-LC-NA, CS04-FL-NA, CS04-HC-NA, CS04-LC-NA, CS23-FL-NA, CS23-HC-NA, CS23-LC-NA, CS01-SC1-NA, CS04-SC1-NA, CS23-SC1-NA, CS01-SC2-NA, CS04-SC2-NA, and CS23-SC2-NA.

[0067] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 96% identity to a sequence selected from the group consisting of CS01-FL-NA, CS01-HC-NA, CS01-LC-NA, CS04-FL-NA, CS04-HC-NA, CS04-LC-NA, CS23-FL-NA, CS23-HC-NA, CS23-LC-NA, CS01-SC1-NA, CS04-SC1-NA, CS23-SC1-NA, CS01-SC2-NA, CS04-SC2-NA, and CS23-SC2-NA.

[0068] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 97% identity to a sequence selected from the group consisting of CS01-FL-NA, CS01-HC-NA, CS01-LC-NA, CS04-FL-NA, CS04-HC-NA, CS04-LC-NA, CS23-FL-NA, CS23-HC-NA, CS23-LC-NA, CS01-SC1-NA, CS04-SC1-NA, CS23-SC1-NA, CS01-SC2-NA, CS04-SC2-NA, and CS23-SC2-NA.

[0069] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 98% identity to a sequence selected from the group consisting of CS01-FL-NA, CS01-HC-NA, CS01-LC-NA, CS04-FL-NA, CS04-HC-NA, CS04-LC-NA, CS23-FL-NA, CS23-HC-NA, CS23-LC-NA, CS01-SC1-NA, CS04-SC1-NA, CS23-SC1-NA, CS01-SC2-NA, CS04-SC2-NA, and CS23-SC2-NA.

[0070] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 99% identity to a sequence selected from the group consisting of CS01-FL-NA, CS01-HC-NA, CS01-LC-NA, CS04-FL-NA, CS04-HC-NA, CS04-LC-NA, CS23-FL-NA, CS23-HC-NA, CS23-LC-NA, CS01-SC1-NA, CS04-SC1-NA, CS23-SC1-NA, CS01-SC2-NA, CS04-SC2-NA, and CS23-SC2-NA.

[0071] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 99.5% identity to a sequence selected from the group consisting of CS01-FL-NA, CS01-HC-NA, CS01-LC-NA, CS04-FL-NA, CS04-HC-NA, CS04-LC-NA, CS23-FL-NA, CS23-HC-NA, CS23-LC-NA, CS01-SC1-NA, CS04-SC1-NA, CS23-SC1-NA, CS01-SC2-NA, CS04-SC2-NA, and CS23-SC2-NA.

[0072] In one embodiment of the polynucleotides described above, the nucleotide sequence has at least 99.5% identity to a sequence selected from the group consisting of CS01-FL-NA, CS01-HC-NA, CS01-LC-NA, CS04-FL-NA, CS04-HC-NA, CS04-LC-NA, CS23-FL-NA, CS23-HC-NA, CS23-LC-NA, CS01-SC1-NA, CS04-SC1-NA, CS23-SC1-NA, CS01-SC2-NA, CS04-SC2-NA, and CS23-SC2-NA.

[0073] In one embodiment of the polynucleotides described above, the nucleotide sequence is selected from the group consisting of CS01-FL-NA, CS01-HC-NA, CS01-LC-NA, CS04-FL-

NA, CS04-HC-NA, CS04-LC-NA, CS23-FL-NA, CS23-HC-NA, CS23-LC-NA, CS01-SC1-NA, CS04-SC1-NA, CS23-SC1-NA, CS01-SC2-NA, CS04-SC2-NA, and CS23-SC2-NA.

[0074] In one embodiment of the polynucleotides described above, the encoded Factor VIII polypeptide comprises a glycosylation polypeptide positioned between two consecutive amino acids.

[0075] In one embodiment of the polynucleotides described above, the polynucleotide also includes a promoter element operably linked to the polynucleotide encoding the Factor VIII polypeptide.

[0076] In one embodiment of the polynucleotides described above, the polynucleotide also includes an enhancer element operably linked to the polynucleotide encoding the Factor VIII polypeptide.

[0077] In one embodiment of the polynucleotides described above, the polynucleotide also includes a polyadenylation element operably linked to the polynucleotide encoding the Factor VIII polypeptide.

[0078] In one embodiment of the polynucleotides described above, the polynucleotide also includes an intron operatively linked to the nucleotide sequence encoding the Factor VIII polypeptide.

[0079] In one embodiment of the polynucleotides described above, the intron is positioned between a promoter element and the translation initiation site (e.g., the first coding ATG) of the nucleotide sequence encoding a Factor VIII polypeptide.

[0080] In another embodiment, the disclosure provides a mammalian gene therapy vector including a polynucleotide as described above.

[0081] In one embodiment of the mammalian gene therapy vector described above, the mammalian gene therapy vector is an adeno-associated virus (AAV) vector.

[0082] In one embodiment of the mammalian gene therapy vector described above, the AAV vector is an AAV-8 vector.

[0082a] In a second aspect, there is provided an adeno-associated virus (AAV) vector comprising any of the polynucleotide disclosed herein.

[0082b] In a third aspect, there is provided an adeno-associated virus (AAV) particle comprising any of the polynucleotide disclosed herein.

[0082c] In a fourth aspect, there is provided a host cell infected an adeno-associated virus (AAV) particle comprising any of the polynucleotide disclosed herein.

[0082d] In a fifth aspect, there is provided a method for producing an adeno-associated virus (AAV) particle comprising introducing any of the polynucleotide disclosed herein into a mammalian host cell, wherein the polynucleotide is competent for replication in the mammalian host cell.

[0083] In another embodiment, the disclosure provides a method for treating hemophilia A including administering, to a patient in need thereof, a mammalian gene therapy vector as described above.

[0084] In another embodiment, the disclosure provides a mammalian gene therapy vector as described above for treating hemophilia A.

[0085] In another embodiment, the disclosure provides the use of a mammalian gene therapy vector as described above for the manufacture of a medicament for treating hemophilia A.

[0085a] In a sixth aspect, there is provided an adeno-associated virus (AAV) particle according to the third aspect for use in the treatment of hemophilia A.

[0085b] In a seventh aspect, there is provided an in vitro method for transducing a host cell comprising contacting the host cell with an adeno-associated virus (AAV) particle according to the third aspect.

[0085c] In an eighth aspect, there is provided use of an adeno-associated virus (AAV) particle according to third aspect in the manufacture of a medicament for treating hemophilia A.

BRIEF DESCRIPTION OF DRAWINGS

[0086] Figure 1 shows schematic illustrations of the wild-type and ReFacto-type human Factor VIII protein constructs.

[0087] Figures 2A and 2B show the CS04 codon-altered nucleotide sequence (SEQ ID NO: 1) encoding a Factor VIII variant in accordance with some embodiments (“CS04-FL-NA” for full-length coding sequence).

[0088] Figure 3 shows the Factor VIII variant amino acid sequence (SEQ ID NO: 2) encoded by the CS04 codon-altered nucleotide sequence in accordance with some embodiments (“CS04-FL-AA” for full-length amino acid sequence).

[0089] Figure 4 shows the portion of the CS04 codon-altered nucleotide sequence (SEQ ID NO: 3) encoding the heavy chain of a Factor VIII variant in accordance with some embodiments (“CS04-HC-NA”).

[0090] Figure 5 shows the portion of the CS04 codon-altered nucleotide sequence (SEQ ID NO: 4) encoding the light chain of a Factor VIII variant in accordance with some embodiments (“CS04-LC-NA”).

[0091] Figure 6 shows exemplary coding sequences (SEQ ID NOS: 5-7) for B-domain substituted linkers in accordance with some embodiments. BDLO01 (SEQ ID NO: 5), BDLO04 (SEQ ID NO: 6), and BDLO23 (SEQ ID NO: 7) are the respective portions of the CS01, CS04, and CS23 codon-altered nucleotide sequences that encode a B-domain substituted linker, respectively.

[0092] Figures 7A, 7B, and 7C show an AAV vector sequence (SEQ ID NO: 8) containing an CS04 codon-altered nucleotide sequence in accordance with some embodiments (“CS04-AV-NA”).

[0093] Figures 8A and 8B show the CS04Δ(760-1667) (SPI; CS04Δ(741-1648), SPE) codon-altered nucleotide sequence (SEQ ID NO: 9) encoding a single-chain Factor VIII variant in accordance with some embodiments (“CS04-SC1-NA”).

[0094] Figure 9 shows the Factor VIII variant amino acid sequence (SEQ ID NO: 10) encoded by the CS01Δ(760-1667) (SPI; CS01Δ(741-1648), SPE), CS04Δ(760-1667) (SPI; CS04Δ(741-1648), SPE), and CS23Δ(760-1667) (SPI; CS23Δ(741-1648), SPE) codon-altered nucleotide sequences in accordance with some embodiments (“CS01-SC1-AA,” “CS04-SC1-AA,” and “CS23-SC1-AA,” respectively).

[0095] Figures 10A and 10B show the CS04Δ(772-1667) (SPI; CS04Δ(753-1648), SPE) codon-altered nucleotide sequence (SEQ ID NO: 11) encoding a single-chain Factor VIII variant in accordance with some embodiments (“CS04-SC2-NA”).

[0096] Figure 11 shows the Factor VIII variant amino acid sequence (SEQ ID NO: 12) encoded by the CS01Δ(772-1667) (SPI; CS01Δ(753-1648), SPE), CS04Δ(772-1667) (SPI; CS04Δ(753-1648), SPE), and CS23Δ(772-1667) (SPI; CS23Δ(753-1648), SPE) codon-altered nucleotide sequence in accordance with some embodiments (“CS01-SC2-AA,” “CS04-SC2-AA,” and “CS23-SC2-AA,” respectively).

[0097] Figures 12A and 12B show the CS01 codon-altered nucleotide sequence (SEQ ID NO: 13) encoding a Factor VIII variant in accordance with some embodiments (“CS01-FL-NA”).

[0098] Figures 13A and 13B show the CS08 codon-altered nucleotide sequence (SEQ ID NO: 14) encoding a Factor VIII variant in accordance with some embodiments (“CS08-FL-NA”).

[0099] Figures 14A and 14B show the CS10 codon-altered nucleotide sequence (SEQ ID NO: 15) encoding a Factor VIII variant in accordance with some embodiments (“CS10-FL-NA”).

[00100] Figures 15A and 15B show the CS11 codon-altered nucleotide sequence (SEQ ID NO: 16) encoding a Factor VIII variant in accordance with some embodiments (“CS11-FL-NA”).

[00101] Figures 16A and 16B show the CS40 wild-type ReFacto coding sequence (SEQ ID NO: 17), in accordance with some embodiments (“CS40-FL-NA”).

[00102] Figures 17A and 17B show the CH25 codon-altered nucleotide sequence (SEQ ID NO: 18) encoding a Factor VIII variant in accordance with some embodiments (“CH25-FL-NA”).

[00103] Figure 18 shows a wild-type human Factor VIII amino acid sequence (SEQ ID NO: 19), in accordance with some embodiments (“FVIII-FL-AA”).

[00104] Figure 19 illustrates the scheme for cloning the pCS40, pCS01, pCS04, pCS08, pCS10, pCS11, and pCh25 constructs, by inserting synthetic Refacto-type BDD-FVIII DNA sequences into the vector backbone pCh-BB01 via *AscI* and *NotI* restriction sites.

[00105] Figure 20 shows the integrity of AAV vector genome preparations, as analyzed by agarose gel electrophoresis. Lane 1, DNA marker; lane 2, vCS40; lane 3, vCS01; lane 4, vCS04. The AAV vectors have all the same-sized genomes, migrating at approximately 5 kb (arrow, right side). The scale on the left side indicates size of the DNA fragments in kilobases (kb).

[00106] Figure 21 shows the protein analysis of AAV vector preparations by PAGE and silver staining. Lane 1, protein marker (M); lane 2, vCS40, lane 3, vCS01; and lane 4, vCS04. The constructs all have the same AAV8 capsids consisting of VP1, VP2, and VP3 (arrows right side). The scale on the left side indicates size of the protein marker in kilodaltons (kDa).

[00107] Figure 22A and 22B show the CS23 codon-altered nucleotide sequence (SEQ ID NO: 20) encoding a Factor VIII variant in accordance with some embodiments (“CS23-FL-NA”).

[00108] Figure 23 shows the Factor VIII variant amino acid sequence (SEQ ID NO: 21) encoded by the CS23 codon-altered nucleotide sequence in accordance with some embodiments (“CS23-FL-AA”).

[00109] Figure 24 shows the portion of the CS23 codon-altered nucleotide sequence (SEQ ID NO: 22) encoding the heavy chain of a Factor VIII variant in accordance with some embodiments ("CS23-HC-NA").

[00110] Figure 25 shows the portion of the CS23 codon-altered nucleotide sequence (SEQ ID NO: 23) encoding the light chain of a Factor VIII variant in accordance with some embodiments("CS23-LC-NA").

[00111] Figure 26 shows the portion of the CS01 codon-altered nucleotide sequence (SEQ ID NO: 24) encoding the heavy chain of a Factor VIII variant in accordance with some embodiments ("CS01-HC-NA").

[00112] Figure 27 shows the portion of the CS01 codon-altered nucleotide sequence (SEQ ID NO: 25) encoding the light chain of a Factor VIII variant in accordance with some embodiments("CS01-LC-NA").

[00113] Figures 28A and 28B show the CS01 Δ (760-1667) (SPI; CS01 Δ (741-1648), SPE) codon-altered nucleotide sequence (SEQ ID NO: 26) encoding a single-chain Factor VIII variant in accordance with some embodiments ("CS01-SC1-NA").

[00114] Figures 29A and 29B show the CS01 Δ (772-1667) (SPI; CS01 Δ (753-1648), SPE) codon-altered nucleotide sequence (SEQ ID NO: 27) encoding a single-chain Factor VIII variant in accordance with some embodiments ("CS01-SC2-NA").

[00115] Figures 30A and 30B show the CS23 Δ (760-1667) (SPI; CS23 Δ (741-1648), SPE) codon-altered nucleotide sequence (SEQ ID NO: 28) encoding a single-chain Factor VIII variant in accordance with some embodiments ("CS23-SC1-NA").

[00116] Figures 31A and 31B show the CS23 Δ (772-1667) (SPI; CS23 Δ (753-1648), SPE) codon-altered nucleotide sequence (SEQ ID NO: 29) encoding a single-chain Factor VIII variant in accordance with some embodiments ("CS23-SC2-NA").

DETAILED DESCRIPTION OF DISCLOSURE

I. Introduction

[00117] AAV-based gene therapy holds great promise for the treatment of hemophiliacs. For hemophilia B, first clinical data are encouraging in that FIX levels of about 10% can be maintained in at least some patients for more than 1 year. For hemophilia A however, achieving therapeutic expression levels of 5-10% with AAV vectors remains challenging for various reasons. First, the Factor VIII coding sequence is too large for conventional AAV-based vectors. Second, engineered B-domain deleted or truncated Factor VIII constructs suffer from poor expression in vivo, even when codon-optimized. Third, these B-domain deleted or truncated Factor VIII variant constructs have short half-lives in vivo, exacerbating the effects of poor expression. Fourth, even when expressed, FVIII is not efficiently secreted from cells, as are other coagulation factors, such as Factor IX.

[00118] Moreover, these challenges cannot be addressed by simply administering higher doses of the gene therapy construct. According to current knowledge, the vector dose of an AAV-based gene therapy vector should be increased above 2×10^{12} vg/kg bodyweight. This is because at such high doses a T cell immune response is triggered, which destroys transduced cells and, as a consequence, transgene expression is reduced or even eliminated. Therefore, strategies to improve the expression of FVIII are needed to make FVIII gene therapy a viable therapeutic option for hemophilia A patients.

[00119] The present disclosure relates to the discovery of codon-altered Factor VIII variant coding sequences that solve these and other problems associated with Factor VIII gene therapy. For example, the polynucleotides disclosed herein provide markedly improved expression in mammalian cells, and display improved virion packaging due to stabilized packing interactions. In some implementations, these advantages are realized by using coding sequences for the heavy and light chains of Factor VIII with high sequence identity to the codon altered CS01, CS04, and CS23 constructs (e.g., with high sequence identity to one of the CS01-HC, CS04-HC, and CS23-HC heavy chain coding sequences and high sequence identity to one of the CS01-LC, CS04-LC, and CS23-LC light chain coding sequences).

[00120] In some implementations, the Factor VIII molecules encoded by the polynucleotides described herein have been shortened by truncating, deleting, or replacing the wild-type B-domain. As such, the polynucleotides are better suited for expressing Factor VIII via conventional gene therapy vectors, which inefficiently express larger polypeptides, such as the wild-type Factor VIII.

[00121] Advantageously, it is shown herein that the CS01, CS04, and CS23 codon-altered Factor VIII variant coding sequences provide superior expression of a B-domain deleted Factor VIII construct in vivo. For example, it is demonstrated in Example 2 and Table 4 that intravenous administration of AAV-based gene therapy vectors having the CS01 (SEQ ID NO: 13), CS04 (SEQ ID NO: 1), and CS23 (SEQ ID NO: 20) coding sequence provide 18-fold, 74-fold, and 30-fold increases in Factor VIII expression, relative to the corresponding CS40 construct encoded with the wild-type polynucleotide sequence (SEQ ID NO: 17), in Factor VIII knock-out mice (**Table 4**).

[00122] Further, it also shown herein that the CS01 and CS04 codon-altered Factor VIII variant coding sequences provide superior virion packaging and virus production. For example, it is demonstrated in Example 1 that AAV vector constructs containing the CS01 and CS04 constructs provided 5 to 7-fold greater viral yield, relative to the corresponding CS40 construct encoded with the wild-type polynucleotide sequence, when isolated from the same amount of cell pellet.

II. Definitions

[00123] As used herein, the following terms have the meanings ascribed to them unless specified otherwise.

[00124] As used herein, the terms “Factor VIII” and “FVIII” are used interchangeably, and refer to any protein with Factor VIII activity (e.g., active FVIII, often referred to as FVIIIa) or protein precursor (e.g., pro-protein or pre-pro-protein) of a protein with Factor VIII activity, particularly Factor IXa cofactor activity. In an exemplary embodiment, a Factor VIII polypeptide refers to a polypeptide that has sequences with high sequence identity (e.g., at least 70%, 75%, 80%, 85%, 90%, 95%, 99%, or more) to the heavy and light chains of a wild type

Factor VIII polypeptide. In some embodiments, the B-domain of a Factor VIII polypeptide is deleted, truncated, or replaced with a linker polypeptide to reduce the size of the polynucleotide encoding the Factor VIII polypeptide. In an exemplary embodiment, amino acids 20-1457 of CS04-FL-AA constitute a Factor VIII polypeptide.

[00125] Non-limiting examples of wild type Factor VIII polypeptides include human pre-pro-Factor VIII (e.g., GenBank accession nos. AAA52485, CAA25619, AAA58466, AAA52484, AAA52420, AAV85964, BAF82636, BAG36452, CAI41660, CAI41666, CAI41672, CAI43241, CAO03404, EAW72645, AAH22513, AAH64380, AAH98389, AAI11968, AAI11970, or AAB61261), corresponding pro-Factor VIII, and natural variants thereof; porcine pre-pro-Factor VIII (e.g., UniProt accession nos. F1RZ36 or K7GSZ5), corresponding pro-Factor VIII, and natural variants thereof; mouse pre-pro-Factor VIII (e.g., GenBank accession nos. AAA37385, CAM15581, CAM26492, or EDL29229), corresponding pro-Factor VIII, and natural variants thereof; rat pre-pro-Factor VIII (e.g., GenBank accession no. AAQ21580), corresponding pro-Factor VIII, and natural variants thereof; rat pre-pro-Factor VIII; and other mammalian Factor VIII homologues (e.g., monkey, ape, hamster, guinea pig, etc.).

[00126] As used herein, a Factor VIII polypeptide includes natural variants and artificial constructs with Factor IX cofactor activity. As used in the present disclosure, Factor VIII encompasses any natural variants, alternative sequences, isoforms, or mutant proteins that retain some basal Factor IX cofactor activity (e.g., at least 5%, 10%, 25%, 50%, 75%, or more of the corresponding wild type activity). Examples of Factor VIII amino acid variations (relative to FVIII-FL-AA (SEQ ID NO: 19)) found in the human population include, without limitation, S19R, R22T, Y24C, Y25C, L26P/R, E30V, W33G, Y35C/H, G41C, R48C/K, K67E/N, L69P, E72K, D75E/V/Y, P83R, G89D/V, G92A/V, A97P, E98K, V99D, D101G/H/V, V104D, K108T, M110V, A111T/V, H113R/Y, L117F/R, G121S, E129V, G130R, E132D, Y133C, D135G/Y, T137A/I, S138R, E141K, D145H, V147D, Y155H, V159A, N163K, G164D/V, P165S, C172W, S176P, S179P, V181E/M, K185T, D186G/N/Y, S189L, L191F, G193R, L195P, C198G, S202N/R, F214V, L217H, A219D/T, V220G, D222V, E223K, G224W, T252I, V253F, N254I, G255V, L261P, P262L, G263S, G266F, C267Y, W274C, H275L, G278R, G280D, E284K, V285G, E291G/K, T294I, F295L, V297A, N299I, R301C/H/L, A303E/P, I307S, S308L, F312S,

T314A/I, A315V, G323E, L326P, L327P/V, C329F, I331V, M339T, E340K, V345A/L, C348R/S/Y, Y365C, R391C/H/P, S392L/P, A394S, W401G, I405F/S, E409G, W412G/R, K427I, L431F/S, R437P/W, I438F, G439D/S/V, Y442C, K444R, Y450D/N, T454I, F455C, G466E, P470L/R/T, G474E/R/V, E475K, G477V, D478N, T479R, F484C, A488G, R490G, Y492C/H, Y492H, I494T, P496R, G498R, R503H, G513S/V, I522Y, K529E, W532G, P540T, T541S, D544N, R546W, R550C/G/H, S553P, S554C/G, V556D, R560T, D561G/H/Y, I567T, P569R, S577F, V578A, D579A/H, N583S, Q584H/K/R, I585R/T, M586V, D588G/Y, L594Q, S596P, N601D/K, R602G, S603I/R, W604C, Y605H/S, N609I, R612C, N631K/S, M633I, S635N, N637D/I/S, Y639C, L644V, L650F, V653A/M, L659P, A663V, Q664P, F677L, M681I, V682F, Y683C/N, T686R, F698L, M699T/V, M701I, G705V, G710W, N713I, R717L/W, G720D/S, M721I/L, A723T, L725Q, V727F, E739K, Y742C, R795G, P947R, V1012L, E1057K, H1066Y, D1260E, K1289Q, Q1336K, N1460K, L1481P, A1610S, I1698T, Y1699C/F, E1701K, Q1705H, R1708C/H, T1714S, R1715G, A1720V, E1723K, D1727V, Y1728C, R1740G, K1751Q, F1762L, R1768H, G1769R, L1771P, L1775F/V, L1777P, G1779E/R, P1780L, I1782R, D1788H, M1791T, A1798P, S1799H, R1800C/G/H, P1801A, Y1802C, S1803Y, F1804S, L1808F, M1842I, P1844S, T1845P, E1848G, A1853T/V, S1858C, K1864E, D1865N/Y, H1867P/R, G1869D/V, G1872E, P1873R, L1875P, V1876L, C1877R/Y, L1882P, R1888I, E1894G, I1901F, E1904D/K, S1907C/R, W1908L, Y1909C, A1939T/V, N1941D/S, G1942A, M1945V, L1951F, R1960L/Q, L1963P, S1965I, M1966I/V, G1967D, S1968R, N1971T, H1973L, G1979V, H1980P/Y, F1982I, R1985Q, L1994P, Y1998C, G2000A, T2004R, M2007I, G2013R, W2015C, R2016P/W, E2018G, G2022D, G2028R, S2030N, V2035A, Y2036C, N2038S, 2040Y, G2045E/V, I2051S, I2056N, A2058P, W2065R, P2067L, A2070V, S2082N, S2088F, D2093G/Y, H2101D, T2105N, Q2106E/P/R, G2107S, R2109C, I2117F/S, Q2119R, F2120C/L, Y2124C, R2135P, S2138Y, T2141N, M2143V, F2145C, N2148S, N2157D, P2162L, R2169C/H, P2172L/Q/R, T2173A/I, H2174D, R2178C/H/L, R2182C/H/P, M2183R/V, L2185S/W, S2192I, C2193G, P2196R, G2198V, E2200D, I2204T, I2209N, A2211P, A2220P, P2224L, R2228G/L/P/Q, L2229F, V2242M, W2248C/S, V2251A/E, M2257V, T2264A, Q2265R, F2279C/I, I2281T, D2286G, W2290L, G2304V, D2307A, P2319L/S, R2323C/G/H/L, R2326G/L/P/Q, Q2330P, W2332R, I2336F, R2339T, G2344C/D/S, and C2345S/Y. Factor VIII proteins also include polypeptides containing post-translational modifications.

[00127] Generally, polynucleotides encoding Factor VIII encode for an inactive single-chain polypeptide (e.g., a pre-pro-protein) that undergoes post-translational processing to form an active Factor VIII protein (e.g., FVIIIa). For example, referring to Figure 1, the wild type human Factor VIII pre-pro-protein is first cleaved to release the encoded signal peptide (not shown), forming a first single-chain pro-protein (shown as “human wild-type FVIII). The pro-protein is then cleaved between the B and A3 domains to form a first polypeptide that includes the Factor VIII heavy chain (e.g., the A1 and A2 domains) and B-domain, and a second polypeptide that includes the Factor VIII light chain (e.g., including the A3, C1, and C3 domains). The first polypeptide is further cleaved to remove the B-domain, and also to separate the A1 and A2 domains, which remain associated with the Factor VIII light chain in the mature Factor VIIIa protein. For review of the Factor VIII maturation process, see Graw et al., *Nat Rev Genet.*, 6(6):488-501 (2005), the content of which is incorporated herein by reference in its entirety for all purposes.

[00128] However, in some embodiments, the Factor VIII polypeptide is a single-chain Factor VIII polypeptide. Single-chain Factor VIII polypeptides are engineered to remove natural cleavage sites, and optionally remove, truncate, or replace the B-domain of Factor VIII. As such, they are not matured by cleavage (other than cleavage of an optional signal and/or leader peptide), and are active as a single chain. Non-limiting examples of single-chain Factor VIII polypeptides are described in Zollner et al. (*Thromb Res*, 134(1):125-31 (2014)) and Donath et al. (*Biochem J.*, 312(1):49-55 (1995)), the disclosures of which are hereby incorporated by reference in their entireties for all purposes.

[00129] As used herein, the terms “Factor VIII heavy chain,” or simply “heavy chain,” refers to the aggregate of the A1 and A2 domains of a Factor VIII polypeptide. In an exemplary embodiment, amino acids 20-759 of CS04-FL-AA (SEQ ID NO: 2) constitute a Factor VIII heavy chain.

[00130] As used herein, the term “Factor VIII light chain,” or simply “light chain,” refers to the aggregate of the A3, C1, and C2 domains of a Factor VIII polypeptide. In an exemplary embodiment, amino acids 774-1457 CS04-FL-AA (SEQ ID NO: 2) constitute a Factor VIII light

chain. In some embodiments, a Factor VIII light chain excludes the acidic a3 peptide, which is released during maturation in vivo.

[00131] Generally, Factor VIII heavy and light chains are expressed as a single polypeptide chain, e.g., along with an optional B-domain or B-domain substituted linker. However, in some embodiments, a Factor VIII heavy chain and Factor VIII light chain are expressed as separate polypeptide chains (e.g., co-expressed), and reconstituted to form a Factor VIII protein (e.g., in vivo or in vitro).

[00132] As used herein, the terms “B-domain substituted linker” and “Factor VIII linker” are used interchangeably, and refer to truncated versions of a wild type Factor VIII B-domain (e.g., amino acids 760-1667 of FVIII-FL-AA (SEQ ID NO: 19)) or peptides engineered to replace the B-domain of a Factor VIII polypeptide. As used herein, a Factor VIII linker is positioned between the C-terminus of a Factor VIII heavy chain and the N-terminus of a Factor VIII light chain in a Factor VIII variant polypeptide in accordance with some embodiments. Non-limiting examples of B-domain substituted linkers are disclosed in U.S. Patent Nos. 4,868,112, 5,112,950, 5,171,844, 5,543,502, 5,595,886, 5,610,278, 5,789,203, 5,972,885, 6,048,720, 6,060,447, 6,114,148, 6,228,620, 6,316,226, 6,346,513, 6,458,563, 6,924,365, 7,041,635, and 7,943,374; U.S. Patent Application Publication Nos. 2013/024960, 2015/0071883, and 2015/0158930; and PCT Publication Nos. WO 2014/064277 and WO 2014/127215, the disclosures of which are hereby incorporated by reference, in their entireties, for all purposes.

[00133] Unless otherwise specified herein, the numbering of Factor VIII amino acids refers to the corresponding amino acid in the full-length, wild-type human Factor VIII sequence (FVIII-FL-AA), presented as SEQ ID NO: 19 in Figure 18. As such, when referring to an amino acid substitution in a Factor VIII variant protein disclosed herein, the recited amino acid number refers to the analogous (e.g., structurally or functionally equivalent) and/or homologous (e.g., evolutionarily conserved in the primary amino acid sequence) amino acid in the full-length, wild-type Factor VIII sequence. For example, a T2105N amino acid substitution refers to a T to N substitution at position 2105 of the full-length, wild-type human Factor VIII sequence (FVIII-

FL-AA; SEQ ID NO: 19) and a T to N substitution at position 1211 of the Factor VIII variant protein encoded by CS04 (CS04-FL-AA; SEQ ID NO: 2).

[00134] As described herein, the Factor VIII amino acid numbering system is dependent on whether the Factor VIII signal peptide (e.g., amino acids 1-19 of the full-length, wild-type human Factor VIII sequence) is included. Where the signal peptide is included, the numbering is referred to as “signal peptide inclusive” or “SPI”. Where the signal peptide is not included, the numbering is referred to as “signal peptide exclusive” or “SPE.” For example, F328S is SPI numbering for the same amino acid as F309S, in SPE numbering. Unless otherwise indicated, all amino acid numbering refers to the corresponding amino acid in the full-length, wild-type human Factor VIII sequence (FVIII-FL-AA), presented as SEQ ID NO: 19 in Figure 18.

[00135] As described herein, the codon-altered polynucleotides provide increased expression of transgenic Factor VIII in vivo (e.g., when administered as part of a gene therapy vector), as compared to the level of Factor VIII expression provided by a natively-coded Factor VIII construct (e.g., a polynucleotide encoding the same Factor VIII construct using the wild-type human codons). As used herein, the term “increased expression” refers to an increased level of transgenic Factor VIII activity in the blood of an animal administered the codon-altered polynucleotide encoding Factor VIII, as compared to the level of transgenic Factor VIII activity in the blood of an animal administered a natively-coded Factor VIII construct. The activity levels can be measured using any Factor VIII activity known in the art. An exemplary assay for determining Factor VIII activity is the Technochrome FVIII assay (Technoclone, Vienna, Austria).

[00136] In some embodiments, increased expression refers to at least 25% greater transgenic Factor VIII activity in the blood of an animal administered the codon-altered Factor VIII polynucleotide, as compared to the level of transgenic Factor VIII activity in the blood of an animal administered a natively coded Factor VIII polynucleotide. In some embodiments, increased expression refers to at least 50% greater, at least 75% greater, at least 100% greater, at least 3-fold greater, at least 4-fold greater, at least 5-fold greater, at least 6-fold greater, at least 7-fold greater, at least 8-fold greater, at least 9-fold greater, at least 10-fold greater, at least 15-fold greater, at least 20-fold greater, at least 25-fold greater, at least 30-fold greater, at least 40-

fold greater, at least 50-fold greater, at least 60-fold greater, at least 70-fold greater, at least 80-fold greater, at least 90-fold greater, at least 100-fold greater, at least 125-fold greater, at least 150-fold greater, at least 175-fold greater, at least 200-fold greater, at least 225-fold greater, or at least 250-fold greater transgenic Factor VIII activity in the blood of an animal administered the codon-altered Factor VIII polynucleotide, as compared to the level of transgenic Factor VIII activity in the blood of an animal administered a natively coded Factor VIII polynucleotide.

[00137] As described herein, the codon-altered polynucleotides provide increased vector production, as compared to the level of vector production provided by a natively-coded Factor VIII construct (e.g., a polynucleotide encoding the same Factor VIII construct using the wild-type human codons). As used herein, the term “increased virus production” refers to an increased vector yield in cell culture (e.g., titer per liter culture) inoculated with the codon-altered polynucleotide encoding Factor VIII, as compared to the vector yield in cell culture inoculated with a natively-coded Factor VIII construct. The vector yields can be measured using any vector titer assay known in the art. An exemplary assay for determining vector yield (e.g., of an AAV vector) is qPCR targeting the AAV2 inverted terminal repeats (Aurnhammer, Human Gene Therapy Methods: Part B 23:18–28 (2012)).

[00138] In some embodiments, increased virus production refers to at least 25% greater codon-altered vector yield, as compared to the yield of a natively-coded Factor VIII construct in the same type of culture. In some embodiments, increased vector production refers to at least 50% greater, at least 75% greater, at least 100% greater, at least 3-fold greater, at least 4-fold greater, at least 5-fold greater, at least 6-fold greater, at least 7-fold greater, at least 8-fold greater, at least 9-fold greater, at least 10-fold greater, at least 15-fold greater, or at least 20-fold greater codon-altered vector yield, as compared to the yield of a natively-coded Factor VIII construct in the same type of culture.

[00139] As used herein, the term "hemophilia" refers to a group of disease states broadly characterized by reduced blood clotting or coagulation. Hemophilia may refer to Type A, Type B, or Type C hemophilia, or to the composite of all three diseases types. Type A hemophilia (hemophilia A) is caused by a reduction or loss of factor VIII (FVIII) activity and is the most prominent of the hemophilia subtypes. Type B hemophilia (hemophilia B) results from the loss

or reduction of factor IX (FIX) clotting function. Type C hemophilia (hemophilia C) is a consequence of the loss or reduction in factor XI (FXI) clotting activity. Hemophilia A and B are X-linked diseases, while hemophilia C is autosomal. Conventional treatments for hemophilia include both prophylactic and on-demand administration of clotting factors, such as FVIII, FIX, including Bebulin®-VH, and FXI, as well as FEIBA-VH, desmopressin, and plasma infusions.

[00140] As used herein, the term "FVIII gene therapy" includes any therapeutic approach of providing a nucleic acid encoding Factor VIII to a patient to relieve, diminish, or prevent the reoccurrence of one or more symptoms (e.g., clinical factors) associated with hemophilia. The term encompasses administering any compound, drug, procedure, or regimen comprising a nucleic acid encoding a Factor VIII molecule, including any modified form of Factor VIII (e.g., Factor VIII variant), for maintaining or improving the health of an individual with hemophilia. One skilled in the art will appreciate that either the course of FVIII therapy or the dose of a FVIII therapeutic agent can be changed, e.g., based upon the results obtained in accordance with the present disclosure.

[00141] As used herein, the term "bypass therapy" includes any therapeutic approach of providing non-Factor VIII hemostatic agents, compounds or coagulation factors to a patient to relieve, diminish, or prevent the reoccurrence of one or more symptoms (e.g., clinical factors) associated with hemophilia. Non-Factor VIII compounds and coagulation factors include, but are not limited to, Factor VIII Inhibitor Bypass Activity (FEIBA), recombinant activated factor VII (FVIIa), prothrombin complex concentrates, and activated prothrombin complex concentrates. These non-Factor VIII compounds and coagulation factors may be recombinant or plasma-derived. One skilled in the art will appreciate that either the course of bypass therapy or the dose of bypass therapy can be changed, e.g., based upon the results obtained in accordance with the present disclosure.

[00142] As used herein, a "combination therapy" including administration of a nucleic acid encoding a Factor VIII molecule and a conventional hemophilia A therapeutic agent includes any therapeutic approach of providing both a nucleic acid encoding a Factor VIII molecule and a Factor VIII molecule and/or non-Factor VIII hemostatic agent (e.g., bypass therapeutic agent) to a patient to relieve, diminish, or prevent the reoccurrence of one or more

symptoms (e.g., clinical factors) associated with hemophilia. The term encompasses administering any compound, drug, procedure, or regimen including a nucleic acid encoding a Factor VIII molecule, including any modified form of factor VIII, which is useful for maintaining or improving the health of an individual with hemophilia and includes any of the therapeutic agents described herein.

[00143] The terms "therapeutically effective amount or dose" or "therapeutically sufficient amount or dose" or "effective or sufficient amount or dose" refer to a dose that produces therapeutic effects for which it is administered. For example, a therapeutically effective amount of a drug useful for treating hemophilia can be the amount that is capable of preventing or relieving one or more symptoms associated with hemophilia. The exact dose will depend on the purpose of the treatment, and will be ascertainable by one skilled in the art using known techniques (*see, e.g.,* Lieberman, *Pharmaceutical Dosage Forms* (vols. 1-3, 1992); Lloyd, *The Art, Science and Technology of Pharmaceutical Compounding* (1999); Pickar, *Dosage Calculations* (1999); and *Remington: The Science and Practice of Pharmacy*, 20th Edition, 2003, Gennaro, Ed., Lippincott, Williams & Wilkins).

[00144] As used herein, the term "gene" refers to the segment of a DNA molecule that codes for a polypeptide chain (e.g., the coding region). In some embodiments, a gene is positioned by regions immediately preceding, following, and/or intervening the coding region that are involved in producing the polypeptide chain (e.g., regulatory elements such as a promoter, enhancer, polyadenylation sequence, 5'-untranslated region, 3'-untranslated region, or intron).

[00145] As used herein, the term "regulatory elements" refers to nucleotide sequences, such as promoters, enhancers, terminators, polyadenylation sequences, introns, etc, that provide for the expression of a coding sequence in a cell.

[00146] As used herein, the term "promoter element" refers to a nucleotide sequence that assists with controlling expression of a coding sequence. Generally, promoter elements are located 5' of the translation start site of a gene. However, in certain embodiments, a promoter element may be located within an intron sequence, or 3' of the coding sequence. In some embodiments, a promoter useful for a gene therapy vector is derived from the native gene of the

target protein (e.g., a Factor VIII promoter). In some embodiments, a promoter useful for a gene therapy vector is specific for expression in a particular cell or tissue of the target organism (e.g., a liver-specific promoter). In yet other embodiments, one of a plurality of well characterized promoter elements is used in a gene therapy vector described herein. Non-limiting examples of well-characterized promoter elements include the CMV early promoter, the β -actin promoter, and the methyl CpG binding protein 2 (MeCP2) promoter. In some embodiments, the promoter is a constitutive promoter, which drives substantially constant expression of the target protein. In other embodiments, the promoter is an inducible promoter, which drives expression of the target protein in response to a particular stimulus (e.g., exposure to a particular treatment or agent). For a review of designing promoters for AAV-mediated gene therapy, see Gray et al. (Human Gene Therapy 22:1143-53 (2011)), the contents of which are expressly incorporated by reference in their entirety for all purposes.

[00147] As used herein, the term "vector" refers to any vehicle used to transfer a nucleic acid (e.g., encoding a Factor VIII gene therapy construct) into a host cell. In some embodiments, a vector includes a replicon, which functions to replicate the vehicle, along with the target nucleic acid. Non-limiting examples of vectors useful for gene therapy include plasmids, phages, cosmids, artificial chromosomes, and viruses, which function as autonomous units of replication in vivo. In some embodiments, a vector is a viral vehicle for introducing a target nucleic acid (e.g., a codon-altered polynucleotide encoding a Factor VIII variant). Many modified eukaryotic viruses useful for gene therapy are known in the art. For example, adeno-associated viruses (AAVs) are particularly well suited for use in human gene therapy because humans are a natural host for the virus, the native viruses are not known to contribute to any diseases, and the viruses illicit a mild immune response.

[00148] As used herein, the term "CpG island" refers to a region within a polynucleotide having a statistically elevated density of CpG dinucleotides. As used herein, a region of a polynucleotide (e.g., a polynucleotide encoding a codon-altered Factor VIII protein) is a CpG island if, over a 200-base pair window: (i) the region has GC content of greater than 50%, and (ii) the ratio of observed CpG dinucleotides per expected CpG dinucleotides is at least 0.6, as defined by the relationship:

$$\frac{N[CpG]*N[length\ of\ window]}{N[C]*N[G]} \geq 0.6.$$

For additional information on methods for identifying CpG islands, *see* Gardiner-Garden M. et al., J Mol Biol., 196(2):261-82 (1987), the content of which is expressly incorporated herein by reference, in its entirety, for all purposes.

[00149] As used herein, the term "nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form, and complements thereof. The term encompasses nucleic acids containing known nucleotide analogs or modified backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, which have similar binding properties as the reference nucleic acid, and which are metabolized in a manner similar to the reference nucleotides. Examples of such analogs include, without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, and peptide-nucleic acids (PNAs).

[00150] The term "amino acid" refers to naturally occurring and non-natural amino acids, including amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids include those encoded by the genetic code, as well as those amino acids that are later modified, e.g., hydroxyproline, γ -carboxyglutamate, and O-phosphoserine. Naturally occurring amino acids can include, e.g., D- and L-amino acids. The amino acids used herein can also include non-natural amino acids. Amino acid analogs refer to compounds that have the same basic chemical structure as a naturally occurring amino acid, i.e., any carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, e.g., homoserine, norleucine, methionine sulfoxide, or methionine methyl sulfonium. Such analogs have modified R groups (e.g., norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refer to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that function in a manner similar to a naturally occurring amino acid. Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

[00151] As to amino acid sequences, one of ordinary skill in the art will recognize that individual substitutions, deletions or additions to a nucleic acid or peptide sequence that alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a “conservatively modified variant” where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the disclosure.

[00152] Conservative amino acid substitutions providing functionally similar amino acids are well known in the art. Dependent on the functionality of the particular amino acid, e.g., catalytic, structural, or sterically important amino acids, different groupings of amino acid may be considered conservative substitutions for each other. **Table 1** provides groupings of amino acids that are considered conservative substitutions based on the charge and polarity of the amino acid, the hydrophobicity of the amino acid, the surface exposure/structural nature of the amino acid, and the secondary structure propensity of the amino acid.

Table 1. Groupings of conservative amino acid substitutions based on the functionality of the residue in the protein.

Important Feature	Conservative Groupings
Charge/Polarity	1. H, R, and K 2. D and E 3. C, T, S, G, N, Q, and Y 4. A, P, M, L, I, V, F, and W
Hydrophobicity	1. D, E, N, Q, R, and K 2. C, S, T, P, G, H, and Y 3. A, M, I, L, V, F, and W
Structural/Surface Exposure	1. D, E, N, Q, H, R, and K 2. C, S, T, P, A, G, W, and Y

	3. M, I, L, V, and F
Secondary Structure Propensity	1. A, E, Q, H, K, M, L, and R 2. C, T, I, V, F, Y, and W 3. S, G, P, D, and N
Evolutionary Conservation	1. D and E 2. H, K, and R 3. N and Q 4. S and T 5. L, I, and V 6. F, Y, and W 7. A and G 8. M and C

[00153] The terms “identical” or percent “identity,” in the context of two or more nucleic acids or peptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same (i.e., about 60% identity, preferably 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or higher identity over a specified region, when compared and aligned for maximum correspondence over a comparison window or designated region) as measured using, e.g., a BLAST or BLAST 2.0 sequence comparison algorithms with default parameters described below, or by manual alignment and visual inspection.

[00154] As is known in the art, a number of different programs may be used to identify whether a protein (or nucleic acid as discussed below) has sequence identity or similarity to a known sequence. Sequence identity and/or similarity is determined using standard techniques known in the art, including, but not limited to, the local sequence identity algorithm of Smith & Waterman, *Adv. Appl. Math.*, 2:482 (1981), by the sequence identity alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.*, 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Natl. Acad. Sci. U.S.A.*, 85:2444 (1988), by computerized

implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Drive, Madison, WI), the Best Fit sequence program described by Devereux et al., Nucl. Acid Res., 12:387-395 (1984), preferably using the default settings, or by inspection. Preferably, percent identity is calculated by FastDB based upon the following parameters: mismatch penalty of 1; gap penalty of 1; gap size penalty of 0.33; and joining penalty of 30, "Current Methods in Sequence Comparison and Analysis," Macromolecule Sequencing and Synthesis, Selected Methods and Applications, pp 127-149 (1988), Alan R. Liss, Inc, all of which are incorporated by reference.

[00155] An example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pair wise alignments. It may also plot a tree showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle, J. Mol. Evol. 35:351-360 (1987); the method is similar to that described by Higgins & Sharp CABIOS 5:151-153 (1989), both incorporated by reference. Useful PILEUP parameters including a default gap weight of 3.00, a default gap length weight of 0.10, and weighted end gaps.

[00156] Another example of a useful algorithm is the BLAST algorithm, described in: Altschul et al., J. Mol. Biol. 215, 403-410, (1990); Altschul et al., Nucleic Acids Res. 25:3389-3402 (1997); and Karlin et al., Proc. Natl. Acad. Sci. U.S.A. 90:5873-5787 (1993), both incorporated by reference. A particularly useful BLAST program is the WU-BLAST-2 program which was obtained from Altschul et al., Methods in Enzymology, 266:460-480 (1996); [http://blast.wustl.edu/blast/ README.html](http://blast.wustl.edu/blast/README.html)]. WU-BLAST-2 uses several search parameters, most of which are set to the default values. The adjustable parameters are set with the following values: overlap span =1, overlap fraction = 0.125, word threshold (T) = 11. The HSP S and HSP S2 parameters are dynamic values and are established by the program itself depending upon the composition of the particular sequence and composition of the particular database against which the sequence of interest is being searched; however, the values may be adjusted to increase sensitivity.

[00157] An additional useful algorithm is gapped BLAST, as reported by Altschul et al., Nucl. Acids Res., 25:3389-3402, incorporated by reference. Gapped BLAST uses BLOSUM-62

substitution scores; threshold T parameter set to 9; the two-hit method to trigger ungapped extensions; charges gap lengths of k a cost of $10+k$; Xu set to 16, and Xg set to 40 for database search stage and to 67 for the output stage of the algorithms. Gapped alignments are triggered by a score corresponding to ~22 bits.

[00158] A % amino acid sequence identity value is determined by the number of matching identical residues divided by the total number of residues of the “longer” sequence in the aligned region. The “longer” sequence is the one having the most actual residues in the aligned region (gaps introduced by WU-Blast-2 to maximize the alignment score are ignored). In a similar manner, “percent (%) nucleic acid sequence identity” with respect to the coding sequence of the polypeptides identified is defined as the percentage of nucleotide residues in a candidate sequence that are identical with the nucleotide residues in the coding sequence of the cell cycle protein. A preferred method utilizes the BLASTN module of WU-BLAST-2 set to the default parameters, with overlap span and overlap fraction set to 1 and 0.125, respectively.

[00159] The alignment may include the introduction of gaps in the sequences to be aligned. In addition, for sequences which contain either more or fewer amino acids than the protein encoded by the sequence of Figure 2 (SEQ ID NO:1), it is understood that in one embodiment, the percentage of sequence identity will be determined based on the number of identical amino acids or nucleotides in relation to the total number of amino acids or nucleotides. Thus, for example, sequence identity of sequences shorter than that shown in Figure 2 (SEQ ID NO:1), as discussed below, will be determined using the number of nucleotides in the shorter sequence, in one embodiment. In percent identity calculations relative weight is not assigned to various manifestations of sequence variation, such as, insertions, deletions, substitutions, etc.

[00160] In one embodiment, only identities are scored positively (+1) and all forms of sequence variation including gaps are assigned a value of “0”, which obviates the need for a weighted scale or parameters as described below for sequence similarity calculations. Percent sequence identity may be calculated, for example, by dividing the number of matching identical residues by the total number of residues of the “shorter” sequence in the aligned region and multiplying by 100. The “longer” sequence is the one having the most actual residues in the aligned region.

[00161] The term “allelic variants” refers to polymorphic forms of a gene at a particular genetic locus, as well as cDNAs derived from mRNA transcripts of the genes, and the polypeptides encoded by them. The term “preferred mammalian codon” refers a subset of codons from among the set of codons encoding an amino acid that are most frequently used in proteins expressed in mammalian cells as chosen from the following list: Gly (GGC, GGG); Glu (GAG); Asp (GAC); Val (GTG, GTC); Ala (GCC, GCT); Ser (AGC, TCC); Lys (AAG); Asn (AAC); Met (ATG); Ile (ATC); Thr (ACC); Trp (TGG); Cys (TGC); Tyr (TAT, TAC); Leu (CTG); Phe (TTC); Arg (CGC, AGG, AGA); Gln (CAG); His (CAC); and Pro (CCC).

[00162] As used herein, the term codon-altered refers to a polynucleotide sequence encoding a polypeptide (e.g., a Factor VIII variant protein), where at least one codon of the native polynucleotide encoding the polypeptide has been changed to improve a property of the polynucleotide sequence. In some embodiments, the improved property promotes increased transcription of mRNA coding for the polypeptide, increased stability of the mRNA (e.g., improved mRNA half-life), increased translation of the polypeptide, and/or increased packaging of the polynucleotide within the vector. Non-limiting examples of alterations that can be used to achieve the improved properties include changing the usage and/or distribution of codons for particular amino acids, adjusting global and/or local GC content, removing AT-rich sequences, removing repeated sequence elements, adjusting global and/or local CpG dinucleotide content, removing cryptic regulatory elements (e.g., TATA box and CCAAT box elements), removing of intron/exon splice sites, improving regulatory sequences (e.g., introduction of a Kozak consensus sequence), and removing sequence elements capable of forming secondary structure (e.g., stem-loops) in the transcribed mRNA.

[00163] As discussed herein, there are various nomenclatures to refer to components of the disclosure herein. “CS-number” (e.g. “CS04”, “CS01”, “CS23”, etc.) refer to codon altered polynucleotides encoding FVIII polypeptides and/or the encoded polypeptides, including variants. For example, CS01-FL refers to the **F**ull **L**ength codon altered CS01 polynucleotide sequence or amino acid sequence (sometimes referred to herein as “CS01-FL-AA” for the **A**mino **A**cid sequence and “CS01-FL-NA” for the **N**ucleic **A**cid sequence) encoded by the CS01 polynucleotide sequence. Similarly, “CS01-LC” refers to either the codon altered nucleic acid sequence (“CS01-LC-NA”) encoding the light chain of a FVIII polypeptide or the amino acid

sequence (also sometimes referred to herein as “CS01-LC-AA”) of the FVIII light chain encoded by the CS01 polynucleotide sequence. Likewise, CS01-HC, CS01-HC-AA and CS01-HC-NA are the same for the FVIII heavy chain. As will be appreciated by those in the art, for constructs such as CS01, CS04, CS23, etc., that are only codon-altered (e.g. they do not contain additional amino acid substitutions as compared to Refacto), the amino acid sequences will be identical, as the amino acid sequences are not altered by the codon optimization. Thus, sequence constructs of the disclosure include, but are not limited to, CS01-FL-NA, CS01-FL-AA, CS01-LC-NA, CS01-LC-AA, CS01-HC-AA, CS01-HC-NA, CS04-FL-NA, CS04-FL-AA, CS04-LC-NA, CS04-LC-AA, CS04-HC-AA, CS04-HC-NA, CS23-FL-NA, CS23-FL-AA, CS23-LC-NA, CS23-LC-AA, CS23-HC-AA and CS23-HC-NA.

III. Codon-Altered Factor VIII Variants

[00164] In some embodiments, the present disclosure provides codon-altered polynucleotides encoding Factor VIII variants. These codon-altered polynucleotides provide markedly improved expression of Factor VIII when administered in an AAV-based gene therapy construct. The codon-altered polynucleotides also demonstrate improved AAV-virion packaging, as compared to conventionally codon-optimized constructs. As demonstrated in Example 2 and Table 4, Applicants have achieved these advantages through the discovery of three codon-altered polynucleotides (CS01-FL-NA, CS04-FL-NA, and CS23-FL-NA) encoding a Factor VIII polypeptide with human wild-type Factor VIII heavy and light chains, and a short, 14 amino acid, B-domain substituted linker (the “SQ” linker) containing a furin cleavage site to facilitate maturation of an active FVIIIa protein in vivo.

[00165] In one embodiment, a codon-altered polynucleotide provided herein has nucleotide sequences with high sequence identity to at least the sequences within CS01, CS04, or CS23 (SEQ ID NOS 13, 1, and 20, respectively) encoding the Factor VIII heavy chain and Factor VIII light chains. As known in the art, the B-domain of Factor VIII is dispensable for activity in vivo. Thus, in some embodiments, the codon-altered polynucleotides provided herein completely lack a Factor VIII B-domain. In some embodiments, the native Factor VIII B-domain is replaced with a short amino acid linker containing a furin cleavage site, e.g., the “SQ” linker consisting of amino acids 760-773 of the CS01, CS04, or CS23 (SEQ ID NOS 2, 2, and

21, respectively) constructs. The “SQ” linker is also referred to as BDLO04, (-AA for the amino acid sequence and -NA for the nucleotide sequence).

[00166] In one embodiment, the Factor VIII heavy and light chains encoded by the codon-altered polynucleotide are human Factor VIII heavy and light chains, respectively. In other embodiments, the Factor VIII heavy and light chains encoded by the codon-altered polynucleotide are heavy and light chain sequences from another mammal (e.g., porcine Factor VIII). In yet other embodiments, the Factor VIII heavy and light chains are chimeric heavy and light chains (e.g., a combination of human and a second mammalian sequence). In yet other embodiments, the Factor VIII heavy and light chains are humanized version of the heavy and light chains from another mammal, e.g., heavy and light chain sequences from another mammal in which human residues are substituted at select positions to reduce the immunogenicity of the resulting peptide when administered to a human.

[00167] The GC content of human genes varies widely, from less than 25% to greater than 90%. However, in general, human genes with higher GC contents are expressed at higher levels. For example, Kudla et al. (PLoS Biol., 4(6):80 (2006)) demonstrate that increasing a gene’s GC content increases expression of the encoded polypeptide, primarily by increasing transcription and effecting a higher steady state level of the mRNA transcript. Generally, the desired GC content of a codon-optimized gene construct is equal or greater than 60%. However, native AAV genomes have GC contents of around 56%.

[00168] Accordingly, in some embodiments, the codon-altered polynucleotides provided herein have a CG content that more closely matches the GC content of native AAV virions (e.g., around 56% GC), which is lower than the preferred CG contents of polynucleotides that are conventionally codon-optimized for expression in mammalian cells (e.g., at or above 60% GC). As outlined in Example 1, CS04-FL-NA (SEQ ID NO: 1), which has a GC content of about 56%, has improved virion packaging as compared to similarly codon-altered coding sequences with higher GC content.

[00169] Thus, in some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is less than 60%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is less

than 59%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is less than 58%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is less than 57%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is no more than 56%.

[00170] In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 54% to 59%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 55% to 59%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 56% to 59%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 54% to 58%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 55% to 58%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 56% to 58%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 54% to 57%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 55% to 57%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 56% to 57%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 54% to 56%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 55% to 56%.

[00171] In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is $56 \pm 0.5\%$. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is $56 \pm 0.4\%$. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is $56 \pm 0.3\%$. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is $56 \pm 0.2\%$. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is

56±0.1%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is 56%.

A. Factor VIII B-domain Substituted Linkers

[00172] In some embodiments, the linkage between the FVIII heavy chain and the light chain (e.g., the B-domain in wild-type Factor VIII) is further altered. Due to size constraints of AAV packaging capacity, B-domain deleted, truncated, and or linker substituted variants should improve the efficacy of the FVIII gene therapy construct. The most conventionally used B-domain substituted linker is that of SQ FVIII, which retains only 14 amino acids of the B domain as linker sequence. Another variant of porcine VIII ("OBI-1," described in U.S. Patent No. 6,458,563) is well expressed in CHO cells, and has a slightly longer linker of 24 amino acids. In some embodiments, the Factor VIII constructs encoded by the codon-altered polynucleotides described herein include an SQ-type B-domain linker sequence. In other embodiments, the Factor VIII constructs encoded by the codon-altered polynucleotides described herein include an OBI-1-type B-domain linker sequence.

[00173] In some embodiments, the encoded Factor VIII polypeptides described herein include an SQ-type B-domain linker (SFSQNPPVLKRHQR; BDL-SQ-AA; SEQ ID NO: 30), including amino acids 760-762/1657-1667 of the wild-type human Factor VIII B-domain (FVIII-FL-AA; SEQ ID NO: 19) (Sandberg et al. *Thromb. Haemost.* 85:93 (2001)). In some embodiments, the SQ-type B-domain linker has one amino acid substitution relative to the corresponding wild-type sequence. In some embodiments, the SQ-type B-domain linker has two amino acid substitutions relative to the corresponding wild-type sequence.

[00174] In some embodiments, the encoded Factor VIII polypeptides described herein include a Greengene-type B-domain linker, including amino acids 760/1582-1667 of the wild-type human Factor VIII B-domain (FVIII-FL-AA; SEQ ID NO: 19) (Oh et al., *Biotechnol. Prog.*, 17:1999 (2001)). In some embodiments, the Greengene-type B-domain linker has one amino acid substitution relative to the corresponding wild-type sequence. In some embodiments, the Greengene-type B-domain linker has two amino acid substitutions relative to the corresponding wild-type sequence.

[00175] In some embodiments, the encoded Factor VIII polypeptides described herein include an extended SQ-type B-domain linker, including amino acids 760-769/1657-1667 of the wild-type human Factor VIII B-domain (FVIII-FL-AA; SEQ ID NO: 19) (Thim et al., *Haemophilia*, 16:349 (2010)). In some embodiments, the extended SQ-type B-domain linker has one amino acid substitution relative to the corresponding wild-type sequence. In some embodiments, the extended SQ-type B-domain linker has two amino acid substitutions relative to the corresponding wild-type sequence.

[00176] In some embodiments, the encoded Factor VIII polypeptides described herein include a porcine OBI-1-type B-domain linker, including the amino acids SFAQNSRPPSASAPKPPVLRHQ (SEQ ID NO: 31) from the wild-type porcine Factor VIII B-domain (Toschi et al., *Curr. Opin. Mol. Ther.* 12:517 (2010)). In some embodiments, the porcine OBI-1-type B-domain linker has one amino acid substitution relative to the corresponding wild-type sequence. In some embodiments, the porcine OBI-1-type B-domain linker has two amino acid substitutions relative to the corresponding wild-type sequence.

[00177] In some embodiments, the encoded Factor VIII polypeptides described herein include a human OBI-1-type B-domain linker, including amino acids 760-772/1655-1667 of the wild-type human Factor VIII B-domain (FVIII-FL-AA; SEQ ID NO: 19). In some embodiments, the human OBI-1-type B-domain linker has one amino acid substitution relative to the corresponding wild-type sequence. In some embodiments, the human OBI-1-type B-domain linker has two amino acid substitutions relative to the corresponding wild-type sequence.

[00178] In some embodiments, the encoded Factor VIII polypeptides described herein include an O8-type B-domain linker, including the amino acids SFSQNSRHQAYRYRRG (SEQ ID NO: 32) from the wild-type porcine Factor VIII B-domain (Toschi et al., *Curr. Opin. Mol. Ther.* 12:517 (2010)). In some embodiments, the porcine OBI-1-type B-domain linker has one amino acid substitution relative to the corresponding wild-type sequence. In some embodiments, the porcine OBI-1-type B-domain linker has two amino acid substitutions relative to the corresponding wild-type sequence.

[00179] Removal of the B-domain from Factor VIII constructs does not appear to affect the activity of the activated enzyme (e.g., FVIIIa), presumably because the B-domain is removed

during activation. However, the B-domain of Factor VIII contains several residues that are post-translationally modified, e.g., by N- or O-linked glycosylation. In silico analysis (Prediction of N-glycosylation sites in human proteins, R. Gupta, E. Jung and S. Brunak, in preparation (2004)) of the wild-type Factor VIII B-domain predicts that at least four of these sites are glycosylated in vivo. It is thought that these modifications within the B-domain contribute to the post-translational regulation and/or half-life of Factor VIII in vivo.

[00180] While the Factor VIII B-domain is absent in mature Factor VIIIa protein, glycosylation within the B-domain of the precursor Factor VIII molecule may increase the circulating half-life of the protein prior to activation. Thus, in some embodiments, the polypeptide linker of the encoded Factor VIII constructs described herein includes one or more glycosylation sequences, to allow for glycosylation in vivo. In some embodiments, the polypeptide linker includes at least one consensus glycosylation sequence (e.g., an N- or O-linked glycosylation consensus sequence). In some embodiments, the polypeptide linker includes at least two consensus glycosylation sequences. In some embodiments, the polypeptide linker includes at least three consensus glycosylation sequences. In some embodiments, the polypeptide linker includes at least four consensus glycosylation sequences. In some embodiments, the polypeptide linker includes at least five consensus glycosylation sequences. In some embodiments, the polypeptide linker includes at least 6, 7, 8, 9, 10, or more consensus glycosylation sequences.

[00181] In some embodiments, the polypeptide linker contains at least one N-linked glycosylation sequence N-X-S/T, where X is any amino acid other than P, S, or T. In some embodiments, the polypeptide linker contains at least two N-linked glycosylation sequences N-X-S/T, where X is any amino acid other than P, S, or T. In some embodiments, the polypeptide linker contains at least three N-linked glycosylation sequences N-X-S/T, where X is any amino acid other than P, S, or T. In some embodiments, the polypeptide linker contains at least four N-linked glycosylation sequences N-X-S/T, where X is any amino acid other than P, S, or T. In some embodiments, the polypeptide linker contains at least five N-linked glycosylation sequences N-X-S/T, where X is any amino acid other than P, S, or T. In some embodiments, the polypeptide linker contains at least 6, 7, 8, 9, 10, or more N-linked glycosylation sequences N-X-S/T, where X is any amino acid other than P, S, or T.

B. Codon-altered Polynucleotides Encoding a Factor VIII Variant with a Cleavable Linker

CS04 Codon Altered Polynucleotides

[00182] In one embodiment, the codon-altered polynucleotides provided herein include a nucleotide sequence encoding a Factor VIII variant polypeptide with a linker that is cleavable in vivo. The Factor VIII polypeptide includes a Factor VIII light chain, a Factor VIII heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having high sequence identity to CS04-HC-NA (SEQ ID NO: 3), which is the portion of CS04-FL-NA (SEQ ID NO: 1) encoding for a Factor VIII heavy chain. The light chain of the Factor VIII polypeptide is encoded by a second nucleotide sequence with high sequence identity to CS04-LC-NA (SEQ ID NO: 4), which is the portion of CS04-FL-NA (SEQ ID NO: 1) encoding for a Factor VIII light chain. The polypeptide linker includes a furin cleavage site, which allows for maturation in vivo (e.g., after expression in vivo or administration of the precursor polypeptide).

[00183] In some embodiments, the first and second nucleotide sequences have at least 95% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 96% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 97% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 98% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 99% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.5% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.9% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences are identical to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively.

[00184] In some embodiments, the polypeptide linker of the Factor VIII construct is encoded by a third nucleotide sequence having high sequence identity to BDLO04 (SEQ ID NO: 6), which encodes the 14-amino acid linker corresponding to amino acids 760-773 of CS04-FL-AA (SEQ ID NO: 2). In some embodiments, the third nucleotide sequence has at least 95% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence has at least 96% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence has at least 97% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence has at least 98% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence is identical to BDLO04 (SEQ ID NO: 6).

[00185] In some embodiments, the codon-altered polynucleotide has a nucleotide sequence with high sequence identity to CS04-FL-NA (SEQ ID NO: 1). In some embodiments, the nucleotide sequence has at least 95% identity to CS04-FL-NA (SEQ ID NO: 1). In some embodiments, the nucleotide sequence has at least 96% identity to CS04-FL-NA (SEQ ID NO: 1). In some embodiments, the nucleotide sequence has at least 97% identity to CS04-FL-NA (SEQ ID NO: 1). In some embodiments, the nucleotide sequence has at least 98% identity to CS04-FL-NA (SEQ ID NO: 1). In some embodiments, the nucleotide sequence has at least 99% identity to CS04-FL-NA (SEQ ID NO: 1). In some embodiments, the nucleotide sequence has at least 99.5% identity to CS04-FL-NA (SEQ ID NO: 1). In some embodiments, the nucleotide sequence has at least 99.9% identity to CS04-FL-NA (SEQ ID NO: 1). In some embodiments, the nucleotide sequence is identical to CS04-FL-NA (SEQ ID NO: 1).

[00186] In some embodiments, the Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS04-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 97% identity to CS04-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 98% identity to CS04-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 99% identity to CS04-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 99.5% identity to CS04-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 99.9% identity to CS04-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence is identical to CS04-FL-AA (SEQ ID NO: 2).

CS01 Codon Altered Polynucleotides

[00187] In one embodiment, the codon-altered polynucleotides provided herein include a nucleotide sequence encoding a Factor VIII variant polypeptide with a linker that is cleavable in vivo. The Factor VIII polypeptide includes a Factor VIII light chain, a Factor VIII heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having high sequence identity to CS01-HC-NA (SEQ ID NO: 24), which is the portion of CS01-FL-NA (SEQ ID NO: 13) encoding for a Factor VIII heavy chain. The light chain of the Factor VIII polypeptide is encoded by a second nucleotide sequence with high sequence identity to CS01-LC-NA (SEQ ID NO: 25), which is the portion of CS01-FL-NA (SEQ ID NO: 13) encoding for a Factor VIII light chain. The polypeptide linker includes a furin cleavage site, which allows for maturation in vivo (e.g., after expression in vivo or administration of the precursor polypeptide).

[00188] In some embodiments, the first and second nucleotide sequences have at least 95% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 96% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 97% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 98% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 99% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.5% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.9% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences are identical to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively.

[00189] In some embodiments, the polypeptide linker of the Factor VIII construct is encoded by a third nucleotide sequence having high sequence identity to BDLO04 (SEQ ID NO: 6), which encodes the 14-amino acid linker corresponding to amino acids 760-773 of CS01-FL-AA (SEQ ID NO: 2). In some embodiments, the third nucleotide sequence has at least 95% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence has at least 96% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence has at least 97% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence has at least 98% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence is identical to BDLO04 (SEQ ID NO: 6).

[00190] In some embodiments, the codon-altered polynucleotide has a nucleotide sequence with high sequence identity to CS01-FL-NA (SEQ ID NO: 13). In some embodiments, the nucleotide sequence has at least 95% identity to CS01-FL-NA (SEQ ID NO: 13). In some embodiments, the nucleotide sequence has at least 96% identity to CS01-FL-NA (SEQ ID NO: 13). In some embodiments, the nucleotide sequence has at least 97% identity to CS01-FL-NA (SEQ ID NO: 13). In some embodiments, the nucleotide sequence has at least 98% identity to CS01-FL-NA (SEQ ID NO: 13). In some embodiments, the nucleotide sequence has at least 99% identity to CS01-FL-NA (SEQ ID NO: 13). In some embodiments, the nucleotide sequence has at least 99.5% identity to CS01-FL-NA (SEQ ID NO: 13). In some embodiments, the nucleotide sequence has at least 99.9% identity to CS01-FL-NA (SEQ ID NO: 13). In some embodiments, the nucleotide sequence is identical to CS01-FL-NA (SEQ ID NO: 13).

[00191] In some embodiments, the Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS01-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 97% identity to CS01-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 98% identity to CS01-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 99% identity to CS01-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 99.5% identity to CS01-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 99.9% identity to CS01-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence is identical to CS01-FL-AA (SEQ ID NO: 2).

CS23 Codon Altered Polynucleotides

[00192] In one embodiment, the codon-altered polynucleotides provided herein include a nucleotide sequence encoding a Factor VIII variant polypeptide with a linker that is cleavable in vivo. The Factor VIII polypeptide includes a Factor VIII light chain, a Factor VIII heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having high sequence identity to CS23-HC-NA (SEQ ID NO: 22), which is the portion of CS23-FL-NA (SEQ ID NO: 20) encoding for a Factor VIII heavy chain. The light chain of the Factor VIII polypeptide is encoded by a second nucleotide sequence with high sequence identity to CS23-LC-NA (SEQ ID NO: 23), which is the portion of CS23-FL-NA (SEQ ID NO: 20) encoding for a Factor VIII light chain. The polypeptide linker includes a furin cleavage site, which allows for maturation in vivo (e.g., after expression in vivo or administration of the precursor polypeptide).

[00193] In some embodiments, the first and second nucleotide sequences have at least 95% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 96% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 97% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 98% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 99% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.5% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.9% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences are identical to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively.

[00194] In some embodiments, the polypeptide linker of the Factor VIII construct is encoded by a third nucleotide sequence having high sequence identity to BDLO04 (SEQ ID NO: 6), which encodes the 14-amino acid linker corresponding to amino acids 760-773 of CS23-FL-AA (SEQ ID NO: 21). In some embodiments, the third nucleotide sequence has at least 95% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence has at least 96% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence has at least 97% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence has at least 98% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence is identical to BDLO04 (SEQ ID NO: 6).

[00195] In some embodiments, the codon-altered polynucleotide has a nucleotide sequence with high sequence identity to CS23-FL-NA (SEQ ID NO: 20). In some embodiments, the nucleotide sequence has at least 95% identity to CS23-FL-NA (SEQ ID NO: 20). In some embodiments, the nucleotide sequence has at least 96% identity to CS23-FL-NA (SEQ ID NO: 20). In some embodiments, the nucleotide sequence has at least 97% identity to CS23-FL-NA (SEQ ID NO: 20). In some embodiments, the nucleotide sequence has at least 98% identity to CS23-FL-NA (SEQ ID NO: 20). In some embodiments, the nucleotide sequence has at least 99% identity to CS23-FL-NA (SEQ ID NO: 20). In some embodiments, the nucleotide sequence has at least 99.5% identity to CS23-FL-NA (SEQ ID NO: 20). In some embodiments, the nucleotide sequence has at least 99.9% identity to CS23-FL-NA (SEQ ID NO: 20). In some embodiments, the nucleotide sequence is identical to CS23-FL-NA (SEQ ID NO: 20).

[00196] In some embodiments, the Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS23-FL-AA (SEQ ID NO: 21). In some embodiments, the amino acid sequence has at least 97% identity to CS23-FL-AA (SEQ ID NO: 21). In some embodiments, the amino acid sequence has at least 98% identity to CS23-FL-AA (SEQ ID NO: 21). In some embodiments, the amino acid sequence has at least 99% identity to CS23-FL-AA (SEQ ID NO: 21). In some embodiments, the amino acid sequence has at least 99.5% identity to CS23-FL-AA (SEQ ID NO: 21). In some embodiments, the amino acid sequence has at least 99.9% identity to CS23-FL-AA (SEQ ID NO: 21). In some embodiments, the amino acid sequence is identical to CS23-FL-AA (SEQ ID NO: 21).

C. Codon-altered Polynucleotides Encoding a Single-chain Factor VIII Protein

[00197] Factor VIII constructs in which the furin cleavage site located at the C-terminal end of the B-domain is removed retain activity as a single chain polypeptide, despite that normal maturation of the Factor VIII molecule cannot occur (Leyte et al. (1991)). Similarly, a B-domain deleted Factor VIII construct with an attenuated furin site (containing an R1664H amino acid substitution) is more biologically active than the corresponding Factor VIII construct with a wild-type furin cleavage site (Siner et al. (2013)). Accordingly, in some embodiments, the codon-altered polynucleotides provided herein include a nucleotide sequence encoding a single-chain Factor VIII variant polypeptide. The single-chain Factor VIII polypeptide includes a Factor VIII light chain, a Factor VIII heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The polypeptide linker does not include a furin cleavage site.

Single-chain CS04 Codon Altered Polynucleotides

[00198] In one embodiment, the codon-altered polynucleotides provided herein include a nucleotide sequence encoding a single-chain Factor VIII variant polypeptide. The Factor VIII polypeptide includes a Factor VIII light chain, a Factor VIII heavy chain, and an optional polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having high sequence identity to CS04-HC-NA (SEQ ID NO: 3), which is the portion of CS04-FL-NA (SEQ ID NO: 1) encoding for a Factor VIII heavy chain. The light chain of the Factor VIII polypeptide is encoded by a second nucleotide sequence with high sequence identity to CS04-LC-NA (SEQ ID NO: 4), which is the portion of CS04-FL-NA (SEQ ID NO: 1) encoding for a Factor VIII light chain. The optional polypeptide linker does not include a furin cleavage site.

[00199] In some embodiments, the first and second nucleotide sequences have at least 95% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 96% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 97% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments,

the first and second nucleotide sequences have at least 98% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 99% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.5% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.9% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences are identical to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively.

[00200] In some embodiments, the codon-altered polynucleotide has a nucleotide sequence with high sequence identity to CS04-SC1-NA (SEQ ID NO: 9). In some embodiments, the nucleotide sequence has at least 95% identity to CS04-SC1-NA (SEQ ID NO: 9). In some embodiments, the nucleotide sequence has at least 96% identity to CS04-SC1-NA (SEQ ID NO: 9). In some embodiments, the nucleotide sequence has at least 97% identity to CS04-SC1-NA (SEQ ID NO: 9). In some embodiments, the nucleotide sequence has at least 98% identity to CS04-SC1-NA (SEQ ID NO: 9). In some embodiments, the nucleotide sequence has at least 99% identity to CS04-SC1-NA (SEQ ID NO: 9). In some embodiments, the nucleotide sequence has at least 99.5% identity to CS04-SC1-NA (SEQ ID NO: 9). In some embodiments, the nucleotide sequence has at least 99.9% identity to CS04-SC1-NA (SEQ ID NO: 9). In some embodiments, the nucleotide sequence is identical to CS04-SC1-NA (SEQ ID NO: 9).

[00201] In some embodiments, the codon-altered polynucleotide has a nucleotide sequence with high sequence identity to CS04-SC2-NA (SEQ ID NO: 11). In some embodiments, the nucleotide sequence has at least 95% identity to CS04-SC2-NA (SEQ ID NO: 11). In some embodiments, the nucleotide sequence has at least 96% identity to CS04-SC2-NA (SEQ ID NO: 11). In some embodiments, the nucleotide sequence has at least 97% identity to CS04-SC2-NA (SEQ ID NO: 11). In some embodiments, the nucleotide sequence has at least 98% identity to CS04-SC2-NA (SEQ ID NO: 11). In some embodiments, the nucleotide sequence has at least 99% identity to CS04-SC2-NA (SEQ ID NO: 11). In some embodiments, the nucleotide sequence has at least 99.5% identity to CS04-SC2-NA (SEQ ID NO: 11). In some embodiments, the nucleotide sequence has at least 99.9% identity to CS04-SC2-NA (SEQ

ID NO: 11). In some embodiments, the nucleotide sequence is identical to CS04-SC2-NA (SEQ ID NO: 11).

[00202] In some embodiments, the single-chain Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS04-SC1-AA (SEQ ID NO: 10; human Factor VIII Δ (760-1667) (SPI; *HsFVIII* Δ (741-1648), SPE)). In some embodiments, the Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS04-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 97% identity to CS04-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 98% identity to CS04-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 99% identity to CS04-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 99.5% identity to CS04-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 99.9% identity to CS04-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence is identical to CS04-SC1-AA (SEQ ID NO: 10).

[00203] In some embodiments, the single-chain Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS04-SC2-AA (SEQ ID NO: 12; human Factor VIII Δ (772-1667) (SPI; *HsFVIII* Δ (753-1648), SPE)). In some embodiments, the Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS04-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 97% identity to CS04-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 98% identity to CS04-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 99% identity to CS04-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 99.5% identity to CS04-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 99.9% identity to CS04-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence is identical to CS04-SC2-AA (SEQ ID NO: 12).

Single-chain CS01 Codon Altered Polynucleotides

[00204] In one embodiment, the codon-altered polynucleotides provided herein include a nucleotide sequence encoding a single-chain Factor VIII variant polypeptide. The Factor VIII

polypeptide includes a Factor VIII light chain, a Factor VIII heavy chain, and an optional polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having high sequence identity to CS01-HC-NA (SEQ ID NO: 24), which is the portion of CS01-FL-NA (SEQ ID NO: 13) encoding for a Factor VIII heavy chain. The light chain of the Factor VIII polypeptide is encoded by a second nucleotide sequence with high sequence identity to CS01-LC-NA (SEQ ID NO: 25), which is the portion of CS01-FL-NA (SEQ ID NO: 13) encoding for a Factor VIII light chain. The optional polypeptide linker does not include a furin cleavage site.

[00205] In some embodiments, the first and second nucleotide sequences have at least 95% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 96% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 97% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 98% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 99% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.5% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.9% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences are identical to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively.

[00206] In some embodiments, the codon-altered polynucleotide has a nucleotide sequence with high sequence identity to CS01-SC1-NA (SEQ ID NO: 26). In some embodiments, the nucleotide sequence has at least 95% identity to CS01-SC1-NA (SEQ ID NO: 26). In some embodiments, the nucleotide sequence has at least 96% identity to CS01-SC1-NA (SEQ ID NO: 26). In some embodiments, the nucleotide sequence has at least 97% identity to CS01-SC1-NA (SEQ ID NO: 26). In some embodiments, the nucleotide sequence has at least

98% identity to CS01-SC1-NA (SEQ ID NO: 26). In some embodiments, the nucleotide sequence has at least 99% identity to CS01-SC1-NA (SEQ ID NO: 26). In some embodiments, the nucleotide sequence has at least 99.5% identity to CS01-SC1-NA (SEQ ID NO: 26). In some embodiments, the nucleotide sequence has at least 99.9% identity to CS01-SC1-NA (SEQ ID NO: 26). In some embodiments, the nucleotide sequence is identical to CS01-SC1-NA (SEQ ID NO: 26).

[00207] In some embodiments, the codon-altered polynucleotide has a nucleotide sequence with high sequence identity to CS01-SC2-NA (SEQ ID NO: 27). In some embodiments, the nucleotide sequence has at least 95% identity to CS01-SC2-NA (SEQ ID NO: 27). In some embodiments, the nucleotide sequence has at least 96% identity to CS01-SC2-NA (SEQ ID NO: 27). In some embodiments, the nucleotide sequence has at least 97% identity to CS01-SC2-NA (SEQ ID NO: 27). In some embodiments, the nucleotide sequence has at least 98% identity to CS01-SC2-NA (SEQ ID NO: 27). In some embodiments, the nucleotide sequence has at least 99% identity to CS01-SC2-NA (SEQ ID NO: 27). In some embodiments, the nucleotide sequence has at least 99.5% identity to CS01-SC2-NA (SEQ ID NO: 27). In some embodiments, the nucleotide sequence has at least 99.9% identity to CS01-SC2-NA (SEQ ID NO: 27). In some embodiments, the nucleotide sequence is identical to CS01-SC2-NA (SEQ ID NO: 27).

[00208] In some embodiments, the single-chain Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS01-SC1-AA (SEQ ID NO: 10; human Factor VIII Δ (760-1667) (SPI; *HsFVIII* Δ (741-1648), SPE)). In some embodiments, the Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS01-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 97% identity to CS01-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 98% identity to CS01-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 99% identity to CS01-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 99.5% identity to CS01-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 99.9% identity to CS01-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence is identical to CS01-SC1-AA (SEQ ID NO: 10).

[00209] In some embodiments, the single-chain Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS01-SC2-AA (SEQ ID NO: 12; human Factor VIII Δ (772-1667) (SPI; HsFVIII Δ (753-1648), SPE)). In some embodiments, the Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS01-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 97% identity to CS01-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 98% identity to CS01-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 99% identity to CS01-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 99.5% identity to CS01-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 99.9% identity to CS01-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence is identical to CS01-SC2-AA (SEQ ID NO: 12).

Single-chain CS23 Codon Altered Polynucleotides

[00210] In one embodiment, the codon-altered polynucleotides provided herein include a nucleotide sequence encoding a single-chain Factor VIII variant polypeptide. The Factor VIII polypeptide includes a Factor VIII light chain, a Factor VIII heavy chain, and an optional polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having high sequence identity to CS23-HC-NA (SEQ ID NO: 22), which is the portion of CS23-FL-NA (SEQ ID NO: 20) encoding for a Factor VIII heavy chain. The light chain of the Factor VIII polypeptide is encoded by a second nucleotide sequence with high sequence identity to CS23-LC-NA (SEQ ID NO: 23), which is the portion of CS23-FL-NA (SEQ ID NO: 20) encoding for a Factor VIII light chain. The optional polypeptide linker does not include a furin cleavage site.

[00211] In some embodiments, the first and second nucleotide sequences have at least 95% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 96% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 97% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some

embodiments, the first and second nucleotide sequences have at least 98% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 99% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.5% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.9% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences are identical to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively.

[00212] In some embodiments, the codon-altered polynucleotide has a nucleotide sequence with high sequence identity to CS23-SC1-NA (SEQ ID NO: 28). In some embodiments, the nucleotide sequence has at least 95% identity to CS23-SC1-NA (SEQ ID NO: 28). In some embodiments, the nucleotide sequence has at least 96% identity to CS23-SC1-NA (SEQ ID NO: 28). In some embodiments, the nucleotide sequence has at least 97% identity to CS23-SC1-NA (SEQ ID NO: 28). In some embodiments, the nucleotide sequence has at least 98% identity to CS23-SC1-NA (SEQ ID NO: 28). In some embodiments, the nucleotide sequence has at least 99% identity to CS23-SC1-NA (SEQ ID NO: 28). In some embodiments, the nucleotide sequence has at least 99.5% identity to CS23-SC1-NA (SEQ ID NO: 28). In some embodiments, the nucleotide sequence has at least 99.9% identity to CS23-SC1-NA (SEQ ID NO: 28). In some embodiments, the nucleotide sequence is identical to CS23-SC1-NA (SEQ ID NO: 28).

[00213] In some embodiments, the codon-altered polynucleotide has a nucleotide sequence with high sequence identity to CS23-SC2-NA (SEQ ID NO: 29). In some embodiments, the nucleotide sequence has at least 95% identity to CS23-SC2-NA (SEQ ID NO: 29). In some embodiments, the nucleotide sequence has at least 96% identity to CS23-SC2-NA (SEQ ID NO: 29). In some embodiments, the nucleotide sequence has at least 97% identity to CS23-SC2-NA (SEQ ID NO: 29). In some embodiments, the nucleotide sequence has at least 98% identity to CS23-SC2-NA (SEQ ID NO: 29). In some embodiments, the nucleotide sequence has at least 99% identity to CS23-SC2-NA (SEQ ID NO: 29). In some embodiments,

the nucleotide sequence has at least 99.5% identity to CS23-SC2-NA (SEQ ID NO: 29). In some embodiments, the nucleotide sequence has at least 99.9% identity to CS23-SC2-NA (SEQ ID NO: 29). In some embodiments, the nucleotide sequence is identical to CS23-SC2-NA (SEQ ID NO: 29).

[00214] In some embodiments, the single-chain Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS23-SC1-AA (SEQ ID NO: 10; human Factor VIII Δ (760-1667) (SPI; *HsFVIII* Δ (741-1648), SPE)). In some embodiments, the Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS23-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 97% identity to CS23-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 98% identity to CS23-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 99% identity to CS23-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 99.5% identity to CS23-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 99.9% identity to CS23-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence is identical to CS23-SC1-AA (SEQ ID NO: 10).

[00215] In some embodiments, the single-chain Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS23-SC2-AA (SEQ ID NO: 12; human Factor VIII Δ (772-1667) (SPI; *HsFVIII* Δ (753-1648), SPE)). In some embodiments, the Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS23-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 97% identity to CS23-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 98% identity to CS23-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 99% identity to CS23-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 99.5% identity to CS23-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 99.9% identity to CS23-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence is identical to CS23-SC2-AA (SEQ ID NO: 12).

D. Factor VIII Expression Vectors

[00216] In some embodiments, the codon-altered polynucleotides described herein are integrated into expression vectors. Non-limiting examples of expression vectors include viral vectors (e.g., vectors suitable for gene therapy), plasmid vectors, bacteriophage vectors, cosmids, phagemids, artificial chromosomes, and the like.

[00217] Non-limiting examples of viral vectors include: retrovirus, e.g., Moloney murine leukemia virus (MMLV), Harvey murine sarcoma virus, murine mammary tumor virus, and Rous sarcoma virus; adenoviruses, adeno-associated viruses; SV40-type viruses; polyomaviruses; Epstein-Barr viruses; papilloma viruses; herpes viruses; vaccinia viruses; and polio viruses.

[00218] In some embodiments, the codon-altered polynucleotides described herein are integrated into a gene therapy vector. In some embodiments, the gene therapy vector is a retrovirus, and particularly a replication-deficient retrovirus. Protocols for the production of replication-deficient retroviruses are known in the art. For review, see Kriegler, M., *Gene Transfer and Expression, A Laboratory Manual*, W.H. Freeman Co., New York (1990) and Murry, E. J., *Methods in Molecular Biology*, Vol. 7, Humana Press, Inc., Clifton, N.J. (1991).

[00219] In one embodiment, the gene therapy vector is an adeno-associated virus (AAV) based gene therapy vector. AAV systems have been described previously and are generally well known in the art (Kelleher and Vos, *Biotechniques*, 17(6):1110-17 (1994); Cotten et al., *Proc Natl Acad Sci USA*, 89(13):6094-98 (1992); Curiel, *Nat Immun*, 13(2-3):141-64 (1994); Muzyczka, *Curr Top Microbiol Immunol*, 158:97-129 (1992); and Asokan A, et al., *Mol. Ther.*, 20(4):699-708 (2012), each incorporated herein by reference in their entireties for all purposes). Details concerning the generation and use of rAAV vectors are described, for example, in U.S. Patent Nos. 5,139,941 and 4,797,368, each incorporated herein by reference in their entireties for all purposes. In a particular embodiment, the AAV vector is an AAV-8 vector.

[00220] In some embodiments, the codon-altered polynucleotides described herein are integrated into a retroviral expression vector. These systems have been described previously, and are generally well known in the art (Mann *et al.*, *Cell*, 33:153-159, 1983; Nicolas and Rubinstein, In: *Vectors: A survey of molecular cloning vectors and their uses*, Rodriguez and

Denhardt, eds., Stoneham: Butterworth, pp. 494-513, 1988; Temin, In: Gene Transfer, Kucherlapati (ed.), New York: Plenum Press, pp. 149-188, 1986). In a specific embodiment, the retroviral vector is a lentiviral vector (*see*, for example, Naldini *et al.*, *Science*, 272(5259):263-267, 1996; Zufferey *et al.*, *Nat Biotechnol*, 15(9):871-875, 1997; Blomer *et al.*, *J Virol.*, 71(9):6641-6649, 1997; U.S. Pat. Nos. 6,013,516 and 5,994,136).

[00221] A wide variety of vectors can be used for the expression of a Factor VIII polypeptide from a codon-altered polypeptide in cell culture, including eukaryotic and prokaryotic expression vectors. In certain embodiments, a plasmid vector is contemplated for use in expressing a Factor VIII polypeptide in cell culture. In general, plasmid vectors containing replicon and control sequences which are derived from species compatible with the host cell are used in connection with these hosts. The vector can carry a replication site, as well as marking sequences which are capable of providing phenotypic selection in transformed cells. The plasmid will include the codon-altered polynucleotide encoding the Factor VIII polypeptide, operably linked to one or more control sequences, for example, a promoter.

[00222] Non-limiting examples of vectors for prokaryotic expression include plasmids such as pRSET, pET, pBAD, etc., wherein the promoters used in prokaryotic expression vectors include lac, trc, trp, recA, araBAD, etc. Examples of vectors for eukaryotic expression include: (i) for expression in yeast, vectors such as pAO, pPIC, pYES, pMET, using promoters such as AOX1, GAP, GAL1, AUG1, etc; (ii) for expression in insect cells, vectors such as pMT, pAc5, pIB, pMIB, pBAC, etc., using promoters such as PH, p10, MT, Ac5, OpIE2, gp64, polh, etc., and (iii) for expression in mammalian cells, vectors such as pSVL, pCMV, pRc/RSV, pcDNA3, pBPV, etc., and vectors derived from viral systems such as vaccinia virus, adeno-associated viruses, herpes viruses, retroviruses, etc., using promoters such as CMV, SV40, EF-1, UbC, RSV, ADV, BPV, and β -actin.

IV. Examples

Example 1 – Construction of a Codon-Altered Factor VIII Variant Expression Sequence

[00223] Two hurdles had to be overcome in order to create a Factor VIII coding sequence that is effective for gene therapy of hemophilia A. First, because of the genomic size limitations

of conventional gene therapy delivery vectors (e.g., AAV virions), the encoded Factor VIII polypeptide had to be shortened considerably. Second, the coding sequence had to be altered to: (i) stabilize packaging interactions within the delivery vector, (ii) stabilize the mRNA intermediary, and (iii) improve the robustness of transcription/translation of the mRNA.

[00224] To achieve the first objective, Applicants started with a B-domain deleted Factor VIII variant construct, referred to herein as “FVIII-BDD-SQ.” In this construct, the B-domain is replaced with a fourteen amino acid sequence referred to as the “SQ” sequence. Recombinant FVIII-BDD-SQ is sold under the trade name REFACTO®, and has been shown to be effective for the management of hemophilia A. However, the native coding sequence for FVIII-BDD-SQ, which includes human wild-type nucleic acid sequences for the Factor VIII heavy and light chains, is ineffectively expressed in gene therapy vectors.

[00225] To address the poor expression of the native FVIII-BDD-SQ, the codon optimization algorithm described in Fath et al. (PLoS ONE, 6:e17596 (2011)), modified as described in Ward et al. (Blood, 117:798 (2011)) and in McIntosh et al. (Blood, 121, 3335-3344 (2013)) was applied to the FVIII-BDD-SQ sequence to create first intermediate coding sequence CS04a. However, Applicants recognized that the CS04a sequence created using the modified algorithm could be improved by further modifying the sequence. Accordingly, Applicants re-introduced CpG dinucleotides, re-introduced the CGC codon for arginine, changed the leucine and serine codon distributions, re-introduced highly conserved codon pairs, and removed cryptic TATA box, CCAAT box, and splice site elements, while avoiding CpG islands and local overrepresentation of AT-rich and GC-rich stretches.

[00226] First, the modified algorithm systematically replaces codons containing CpG-dinucleotides (e.g., arginine codons) with non-CpG-dinucleotide codons, and eliminates/avoids CpG-dinucleotides created by neighboring codons. This strict avoidance of CpG dinucleotides is usually done to prevent TLR-induced immunity after intramuscular injection of DNA vaccines. However, doing so limits the codon optimization possibilities. For example, the modified algorithm excludes use of the complete set of CGX arginine codons. This is particularly disruptive in the coding of genes for expression in human cells, because CGC is the most frequently used arginine codon in highly expressed human genes. Additionally, avoiding the

creation of CpGs by neighboring codons further limits the optimization possibilities (e.g., limits the number of codon pairs that may be used together).

[00227] Because TLR-induced immunity is not expected to be a problem associated with liver-directed, AAV-based gene therapy, codons including CpGs, and neighboring codons creating CpGs, were re-introduced into intermediate coding sequence CS04a, preferentially in the sequence coding for the Factor VIII light chain (e.g., at the 3' end of the FVIII-BDD-SQ coding sequence). This allowed for more frequent use of preferred human codons, particularly those for arginine. Care was taken, however, to avoid creation of CpG islands, which are regions of coding sequence having a high frequency of CpG sites. This is contrary to the teachings of Krinner et al. (Nucleic Acids Res., 42(6):3551-64 (2014)), which suggests that CpG domains downstream of transcriptional start sites promote high levels of gene expression.

[00228] Second, the modified algorithm applies certain codons exclusively, such as CTG for leucine, GTG for valine, and CAG for glutamine. However, this offends the principles of balanced codon use, for example, as proposed in Haas et al. (Current Biology, 6(3):315-24 (1996)). To account for the overuse of preferred codons by the modified algorithm, alternate leucine codons were re-introduced where allowed by the other rules applied to the codon alteration (e.g., CpG frequency and GC content).

[00229] Third, the modified algorithm replaces codon pairs without regard to how conserved they are in nature, when certain criteria (e.g., the presence of CG-dinucleotides) are met. To account for beneficial properties which may have been conserved by evolution, the most conserved codon pairs that were replaced by the algorithm and the most conserved preferred codon pairs, e.g., as described in Tats et al. (BMC Genomics 9:463 (2008)), were analyzed and adjusted where allowed by the other rules applied to the codon alteration (e.g., CpG frequency and GC content).

[00230] Fourth, serine codons used in the intermediate coding sequence were also re-engineered. Specifically, AGC, TCC, and TCT serine codons were introduced into the modified coding sequence with higher frequency, to better match overall for human codon usage (Haas et al., supra).

[00231] Fifth, TATA box, CCAAT box elements, and intron/exon splice sites were screened and removed from the modified coding sequence. When modifying the coding sequence, care was taken to avoid local overrepresentation of AT-rich or GC rich stretches.

[00232] Finally, in addition to optimizing the codon usage within the coding sequence, the structural requirements of the underlying AAV virion were considered when further refining the intermediate coding sequence CS04a. AAV vectors (e.g., the nucleic acid portion of an AAV virion) are packaged as single stranded DNA molecules into their capsids (for review, see, Daya and Berns, Clin. Microbiol Rev., 21(4):583-93 (2008)). The GC content of the vector is therefore likely to influence packaging of the genome and, thus, vector yields during production. Like many algorithms, the modified algorithm used here creates an optimized gene sequence with a GC content of at least 60% (*see*, Fath et al., PLoS One, 6(3):e17596 (2011) (erratum in: PLoS One, (6)3 (2011)). However, the AAV8 capsid protein is encoded by a nucleotide sequence having a lower GC content of about 56%. Thus, to better mimic the native AAV8 capsid protein coding sequence, the GC content of the intermediate coding sequence CS04a was reduced to 56%.

[00233] The resulting CS04 coding sequence, shown in Figure 2, has an overall GC content of 56%. The CpG-dinucleotide content of the sequence is moderate. However, CpG dinucleotides are predominantly in the downstream portion of the coding sequence, e.g., the portion coding for the Factor VIII light chain. The CS04 sequence has 79.77% nucleotide sequence identity to the corresponding coding sequences in wild-type Factor VIII (Genbank accession M14113).

[00234] For comparison purposes, several other codon-optimized, ReFacto constructs were prepared. CS01 was constructed by applying the codon-optimization algorithm of Fath et al., as modified by Ward et al., as done for CS04. However, unlike CS04, the CS01 construct does not contain any CpG islands. The CS08 ReFacto construct was codon-optimized as described in Radcliff P.M. et al., Gene Therapy, 15:289-97 (2008), the content of which is hereby expressly incorporated by reference herein, in its entirety, for all purposes. The CS10 codon-optimized ReFacto construct was obtained from Eurofins Genomics (Ebersberg, Germany). The CS11 codon-optimized ReFacto construct was obtained from Integrated DNA

Technologies, Inc. (Coralville, USA). The CH25 codon-optimized ReFacto construct was obtained from ThermoFischer Scientific's GeneArt services (Regensburg, Germany). The CS40 ReFacto construct consists of the wild type Factor VIII coding sequence. The algorithm used to construct CS23 is based on the JCAT tool (www.jcat.de), an on-line tool for codon-optimizations (Grote et al., 2005; Nucl. Acids Res. W526-31). The sequence was further modified to more reflect the codon usage of the albumin superfamily (Mirsafian et al. 2014: Sc. Word Journal 2014, ID 639682). The sequence identities shared between each of the ReFacto coding sequences is shown in **Table 2**, below.

Table 2 – Percent identity matrix for codon-altered Factor VIII constructs.

	CS01	CS04	CS08	CS10	CS11	CS40	CH25	CS23
CS01	100%							
CS04	93.0%	100%						
CS08	80.7%	82.2%	100%					
CS10	79.1%	79.4%	78.4%	100%				
CS11	78.3%	78.3%	78.1%	77.5%	100%			
CS40	79.6%	79.8%	76.7%	77.6%	75.4%	100%		
CH25	81.3%	85.1%	85.0%	79.9%	79.4%	75.8%	100%	
CS23	84.3%	89.2%	85.1%	80.3%	79.9	76.5%	93.2%	100%

[00235] Plasmids of each construct were constructed by cloning different synthetic DNA fragments into the same vector backbone plasmid (pCh-BB01). DNA synthesis of the Refacto-type BDD-FVIII fragments with flanking AscI and NotI enzyme restriction sites were done by ThermoFischer Scientific (Regensburg, Germany). The vector backbone contains two flanking AAV2-derived inverted terminal repeats (ITRs) that encompass a promoter/ enhancer sequence derived from the liver-specific murine transthyretin gene, AscI and NotI enzyme restriction sites for insertion of the respective Refacto-type BDD-FVIII and a synthetic polyA site. After ligation of the prepared vector backbone and inserts via the AscI and NotI sites, the resulting plasmids were amplified in milligram scale. The Refacto-type BDD-FVIII sequences of the constructs were verified by direct sequencing (Microsynth, Balgach, Switzerland). The cloning resulted in seven different plasmid constructs named pCS40, pCS01, pCS04, pCS08, pCS10, pCS11, pCh25, and pCS23 (Figure 19). The constructs have the same vector backbone and encode the

same B-domain deleted FVIII protein (Refacto-type BDD-FVIII), but differ in their FVIII coding sequence.

[00236] AAV8-based vectors were prepared by the three plasmid transfection method, as described in Grieger JC, et al. (Virus Vectors Using Suspension HEK293 Cells and Continuous Harvest of Vector From the Culture Media for GMP FIX and FLT1 Clinical Vector, *Mol Ther.*, Oct 6. (2015) doi: 10.1038/mt.2015.187. [Epub ahead of print]), the content of which is hereby expressly incorporated by reference herein, in its entirety, for all purposes. HEK293 suspensions cells were used for plasmid transfections using the corresponding FVIII vector plasmid, the helper plasmid pXX6-80 (carrying adenoviral helper genes), and the packaging plasmid pGSK2/8 (contributing the rep2 and cap8 genes). To isolate the AAV8 constructs, the cell pellets of one liter cultures were processed using iodixanol gradients, as described in Grieger et al. (2015, *Supra*). The procedure resulted in vector preparations called vCS01, vCS04, vCS08, vCS10, vCS11, and vCH25. Vectors were quantified by qPCR using the universal qPCR procedure targeting the AAV2 inverted terminal repeats (Aurnhammer, *Human Gene Therapy Methods: Part B* 23:18–28 (2012)). A control vector plasmid carrying AAV2 inverted terminal repeats served for preparing the standard curve. The resulting vCS04 construct is presented as SEQ ID NO: 8 in Figures 7A-7C.

[00237] The integrity of the vector genomes was analyzed by AAV agarose gel electrophoresis. The electrophoresis was performed as described in Fagone et al., *Human Gene Therapy Methods* 23:1-7 (2012). Briefly, AAV vector preparations were incubated at 75 °C for 10 minutes in the presence of 0.5% SDS and then cooled down to room temperature. Approximately 1.5E10 vector genomes (vg) were loaded per lane on a 1% 1xTAE agarose gel and electrophoresed for 60 min at 7 V/cm of gel length. The gel was then stained in 2x GelRed (Biotium Cat# 41003) solution and imaged by ChemiDocTMMP (Biorad). The results shown in Figure 20 demonstrate that the vCS01, vCS04, and vCS40 viral vectors have the same-sized genome, indicated by a distinct band in the 5kb range (Figure 20, lanes 2-4). Despite a vector size of approx. 5.2 kb, the genome is a homogenous band confirming correct packaging of the somewhat oversized genome (relative to an AAV wild-type genome of 4.7 kb). All other vCS vector preparations show the same genomic size (data not shown).

[00238] In order to confirm the expected pattern of capsid proteins, SDS PAGE followed by silver staining was performed with the vectors vCS01, vCS04, and vCS40 (Figure 21). As shown in the figure, the downstream purification procedure resulted in highly purified material displaying the expected protein pattern of VP1, VP2 and VP3 (Figure 21, lanes 2-4). The same pattern was seen with all other viral preparations (not shown). The SDS-PAGE procedure of AAV preparations was done according to standard procedures. Each lane contained 1E10 vg of the respective viral construct, and were separated on a 4-12% Bis-Tris (NuPAGE® Novex, Life Technologies) gel as per manufacturer's instructions. Silver staining was performed with a SilverQuest™ kit (Novex, Life Technologies) according to the manufacturer's instructions.

[00239] Surprisingly, AAV vectors vCS01 and vCS04 had higher virion packaging, measured by higher yields in AAV virus production, as compared to the vCS40 wild-type coding construct and the other codon-optimized constructs. As shown in **Table 3**, the vCS01 and vCS04 vectors replicated substantially better than vCS40, providing a 5-7 fold yield increase in AAV titer.

Table 3 – Yields per liter cell culture obtained with AAV vector constructs vCS01, vCS04, and vCD40, as purified from cell pellets.

Construct	Vector concentration [vg/ml] x10E12	Yields [vg /liter] x10E12	Fold increase vs wt
vCS40	2.0	11.0	-
vCS01	9.2	51.4	4.7
vCS04 – Sample 1	17.6	79.2	7.2
vCS04 – Sample 2	15.9	58.8	5.4

Example 2 – In Vivo Expression of Codon-Altered Factor VIII Variant Expression Sequences

[00240] To test the biological potency of the codon-altered Factor VIII variant sequences, the ReFacto-type FVIII constructs described in Example 1 were administered to mice lacking Factor VIII. Briefly, the assays were performed in C57Bl/6 FVIII knock-out (ko) mice (with 6-8 animals per group) by tail vein injection of 4E12 vector genomes (vg) per kilogram body weight

of mouse. Blood was drawn 14 days after injection by retroorbital puncture and plasma was prepared and frozen using standard procedures. Expression levels at day 14 were chosen because there is minimal influence of inhibitory antibodies at this time, which are seen in some animals of this mouse model at later times. FVIII activity in the mouse plasma was determined using the Technochrome FVIII assay performed, with only minor modifications, as suggested by the manufacture (Technoclone, Vienna, Austria). For the assay, the plasma samples were appropriately diluted and mixed with assay reagents, containing thrombin, activated factor IX (FIXa), phospholipids, factor X and calcium. Following FVIII activation by thrombin a complex with FIXa, phospholipids and calcium is formed. This complex activates FX to activated FX (FXa) which in turn cleaves para-nitroanilide (pNA) from the chromogenic substrate. The kinetics of pNA formation is measured at 405 nm. The rate is directly proportional to the FVIII concentration in the sample. FVIII concentrations are read from a reference curve and results are given in IU FVIII/milliliter.

[00241] The results, presented in **Table 4** below, demonstrate that the codon-altered sequences designed using commercial algorithms (CS10, CS11, and CH25) provided only a modest increase in BDD-Factor VIII (3-4 fold) as compared to the wild-type BDD-Factor VIII construct (CS40). Similarly, the codon-altered BDD-Factor VIII construct prepared as described in Radcliffe et al. (CS08), only provided a 3-4 fold increase in BDD-FVIII expression. This result is consistent with the results reported in Radcliff et al. Surprisingly, the CS01, CS04, and CS23 constructs provided much higher BDD-FVIII expression in the in-vivo biopotency assays (18-, 74-, and 30-fold increases, respectively).

Table 4 – Expression of FVIII in the plasma of FVIII-knock-out mice induced by the different AAV vector constructs.

Construct	Codon Algorithm	Average FVIII Expression at Day 14 [IU/ml]	Standard deviation	Number of mice	Fold increase vs wt
vCS40	Human wild-type	0.03	0.03	12	-
vCS01	Applicants'	0.55	0.28	22	18.3
vCS04	Applicants'	2.21	1.20	55	73.7
vCS08	Radcliffe et al.	0.11	0.01	6	3.6
vCS10	Eurofins	0.09	0.01	7	3.0
vCS11	IDT	0.08	0.02	8	2.7
vCH25	GeneArt	0.13	0.12	18	4.3
vCS23	Applicants'	0.91	0.32	5	30.3

[00242] It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes.

WHAT IS CLAIMED IS:

1. A polynucleotide comprising the nucleotide sequence of SEQ ID NO:1, wherein the polynucleotide encodes a Factor VIII polypeptide.
2. The polynucleotide of claim 1, further comprising a promoter element operably linked to the polynucleotide encoding the Factor VIII polypeptide.
3. The polynucleotide of claim 2, wherein the promoter element is a liver-specific promoter sequence upstream of the nucleotide sequence encoding the Factor VIII polypeptide.
4. The polynucleotide of claim 3, further comprising an intron sequence positioned between the liver-specific promoter sequence and the nucleotide sequence encoding the Factor VIII polypeptide.
5. An adeno-associated virus (AAV) vector comprising a polynucleotide according to any one of claims 1 to 4.
6. An adeno-associated virus (AAV) particle comprising a polynucleotide according to any one of claims 1 to 4.
7. A host cell infected with an adeno-associated virus (AAV) particle comprising a polynucleotide according to any one of claims 1 to 4.
8. A method for producing an adeno-associated virus (AAV) particle comprising introducing a polynucleotide according to any one of claims 1 to 4 into a mammalian host cell, wherein the polynucleotide is competent for replication in the mammalian host cell.
9. An adeno-associated virus (AAV) particle according to claim 6 for use in the treatment of hemophilia A.

10. An in vitro method for transducing a host cell comprising contacting the host cell with an adeno-associated virus (AAV) particle according to claim 6.

11. Use of an adeno-associated virus (AAV) particle according to claim 6 in the manufacture of a medicament for treating hemophilia A.