

(19) **DANMARK**



Patent- og
Varemærkestyrelsen

(10) **DK/EP 2956477 T4**

(12) Oversættelse af ændret
europæisk patentskrift

-
- (51) Int.Cl.: **C 07 K 14/755 (2006.01)**
- (45) Oversættelsen bekendtgjort den: **2024-04-15**
- (80) Dato for Den Europæiske Patentmyndigheds
bekendtgørelse om opretholdelse af patentet i ændret form: **2024-01-24**
- (86) Europæisk ansøgning nr.: **14751254.5**
- (86) Europæisk indleveringsdag: **2014-02-14**
- (87) Den europæiske ansøgnings publiceringsdag: **2015-12-23**
- (86) International ansøgning nr.: **US2014016441**
- (87) Internationalt publikationsnr.: **WO2014127215**
- (30) Prioritet: **2013-02-15 US 201361765626 P**
- (84) Designerede stater: **AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV
MC MK MT NL NO PL PT RO RS SE SI SK SM TR**
- (73) Patenthaver: **Bioverativ Therapeutics Inc., 225 Second Avenue, Waltham, MA 02451, USA**
- (72) Opfinder: **TAN, Siyuan, 53 Buckman Drive, Lexington, Massachusetts 02421, USA
PETERS, Robert, T., 53 Parkinson Street, Needham, Massachusetts 02492, USA**
- (74) Fuldmægtig i Danmark: **RWS Group, Europa House, Chiltern Park, Chiltern Hill, Chalfont St Peter, Bucks SL9
9FG, Storbritannien**
- (54) Benævnelse: **OPTIMERET FAKTOR VIII-GEN**
- (56) Fremdragne publikationer:
WO-A1-2011/005968
WO-A2-2007/149852
WO-A2-2013/009627
US-A1- 2004 147 436
US-A1- 2009 017 533
US-A1- 2009 042 283
US-A1- 2010 284 971
US-A1- 2013 024 960
US-B1- 6 924 365
N. J. WARD ET AL: "Codon optimization of human factor VIII cDNAs leads to high-level expression", BLOOD,
vol. 117, no. 3, 20 January 2011 (2011-01-20), pages 798-807, XP055052195, ISSN: 0006-4971, DOI:
10.1182/blood-2010-05-282707
WO-A1-03/100053
WO-A1-2009/075772
WO-A1-2012/170289
WO-A2-2011/004361
US-B1- 6 316 226

Fortsættes ...

LIND, P. ET AL.: "Novel forms of B-domain-deleted recombinant factor VIII molecules. Construction and biochemical characterization.", *European Journal of Biochemistry*, Published by Springer-Verlag on behalf of the Federation of European Biochemical Societies, vol. 232., no. 01., 1 August 1995 (1995-08-01), pages 19-27., ISSN: 0014-2956, DOI: 10.1111/j.1432-1033.1995.tb20776.x

W. Cao et al: "Factor VIII accelerates proteolytic cleavage of von Willebrand factor by ADAMTS13", *Proceedings of the National Academy of Sciences*, National Academy of Sciences, vol. 105, no. 21, 27 May 2008 (2008-05-27) , pages 7416-7421, ISSN: 0027-8424, DOI: 10.1073/pnas.0801735105

SANDBERG H, ET AL.: "STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF B-DOMAIN DELETED RECOMBINANT FACTOR VIII", *Seminars in Hematology*, W.B. Saunders Co., US, vol. 38, no. 02, SUPPL. 04, 1 April 2001 (2001-04-01), pages 04-12, US ISSN: 0037-1963, DOI: 10.1053/shem.2001.25888

VEHAR et al.: *Nature*, vol. 312, no. 5992, 22 November 1984 (1984-11-22), pages 337-42,

N. J. Ward et al: "Codon optimization of human factor VIII cDNAs leads to high-level expression", *Blood*, vol. 117, no. 3, 20 January 2011 (2011-01-20), pages 798-807, ISSN: 00064971, DOI: 10.1182/blood-2010-05-282707

Sequence alignment of SEQ ID NOs: 1 and 3 of the opposed patent

Sequence alignment of SEQ ID NO: 1 of the opposed patent and SEQ IDNOs: 4-6 of D1

William R. Strohl: "Fusion Proteins for Half-Life Extension of Biologics as a Strategy to Make Biobetters", *BIODRUGS*, Adis International Ltd., NZ, vol. 29, no. 4, 1 August 2015 (2015-08-01) , pages 215-239, NZ ISSN: 1173-8804, DOI: 10.1007/s40259-015-0133-6

Anonymous: "European Commission Approves ReFacto AF(TM) as a Variation to the Refacto(R) Marketing Authorisation", *Europapress* Europapress, 11 March 2009 (2009-03-11), pages 1-3, Retrieved from the Internet: URL:<https://www.europapress.es/comunicados/empresas-00908/noticia-comunicado-europea-n-commission-approves-refacto-aftm-as-variation-to-the-refactor-marketing-authorisation-20090311172203.html> [retrieved on 2022-06-29]

Anonymous: "Search Orphan Drug Designations and Approvals", *FDA* FDA, 2 August 1996 (1996-08-02), pages 1-2, Retrieved from the Internet: URL:<https://www.accessdata.fda.gov/scripts/opdlisting/oopd/detailedIndex.cfm?cfgridkey=094495> [retrieved on 2022-06-29]

Comparison of in vivo FVIII activity after expression with codon optimised FVIII (SEQ ID: 1) and non-optimised FVIII (SEQ ID: 3)

Comparison of codon usage frequency in SEQ ID NO: 1 and SEQ ID: 3 of the Patent and SEQ ID: 5 of D1

DESCRIPTION

Description

BACKGROUND OF THE INVENTION

[0001] The blood coagulation pathway, in part, involves the formation of an enzymatic complex of Factor VIIIa (FVIIIa) and Factor IXa (FIXa) (Xase complex) on the surface of platelets. FIXa is a serine protease with relatively weak catalytic activity without its cofactor FVIIIa. The Xase complex cleaves Factor X (FX) into Factor Xa (FXa), which in turn interacts with Factor Va (FVa) to cleave prothrombin and generate thrombin. Hemophilia A is a bleeding disorder caused by mutations and/or deletions in the FVIII (FVIII) gene resulting in a deficiency of FVIII activity (Peyvandi *et al.* 2006). In some cases, patients have reduced levels of FVIII due to the presence of FVIII inhibitors, such as anti-FVIII antibodies.

[0002] Hemophilia A is characterized by spontaneous hemorrhage and excessive bleeding. Over time, the repeated bleeding into muscles and joints, which often begins in early childhood, results in hemophilic arthropathy and irreversible joint damage. This damage is progressive and can lead to severely limited mobility of joints, muscle atrophy and chronic pain (Rodriguez-Merchan, E.C., *Semin. Thromb. Hemost.* 29:87-96 (2003)).

[0003] The disease can be treated by replacement therapy targeting restoration of FVIII activity to 1 to 5 % of normal levels to prevent spontaneous bleeding (see, e.g., Mannucci, P.M., *et al.*, *N. Engl. J. Med.* 344:1773-9 (2001)). There are plasma-derived and recombinant FVIII products available to treat bleeding episodes on-demand or to prevent bleeding episodes from occurring by treating prophylactically. Based on the half-life of these products (10-12 hr) (White G.C., *et al.*, *Thromb. Haemost.* 77:660-7 (1997); Morfini, M., *Haemophilia* 9 (suppl 1):94-99;discussion 100 (2003)), treatment regimens require frequent intravenous administration, commonly two to three times weekly for prophylaxis and one to three times daily for on-demand treatment (Manco-Johnson, M.J., *et al.*, *N. Engl. J. Med.* 357:535-544 (2007)). Such frequent administration is inconvenient and costly.

[0004] A major impediment in providing a low-cost recombinant FVIII protein to patients is the high cost of commercial production. FVIII protein expresses poorly in heterologous expression systems, two to three orders of magnitude lower than similarly sized proteins. (Lynch *et al.*, *Hum. Gene. Ther.*; 4:259-72 (1993)). The poor expression of FVIII is due in part to the presence of cis-acting elements in the FVIII coding sequence that inhibit FVIII expression, such as transcriptional silencer elements (Hoebe *et al.*, *Blood* 85:2447-2454 (1995)), matrix attachment-like sequences (MARs) (Fallux *et al.*, *Mol. Cell. Biol.* 16:4264-4272 (1996)), and transcriptional elongation inhibitory elements (Koeberl *et al.*, *Hum. Gene. Ther.*; 6:469-479

(1995)).

[0005] Advances in our understanding of the biology of FVIII expression has led to the development of more potent FVIII variants. For instance, biochemical studies demonstrated that the FVIII B-domain was dispensable for FVIII cofactor activity. Deletion of the B-domain resulted in a 17-fold increase in mRNA levels over full-length wild-type FVIII and a 30% increase in secreted protein. (Toole et al., Proc Natl Acad Sci USA 83:5939-42 (1986)). This led to the development of B domain-deleted (BDD) FVIII protein concentrate, which is now widely used in the clinic. Recent studies, however, indicate that full length and BDD hFVIII misfold in the ER lumen, resulting in activation of the unfolded protein response (UPR) and apoptosis of murine hepatocytes. WO2011005968 and US2009042283 disclose codon optimized Factor VIII with deletion of the B-domain.

[0006] Thus, there exists a need in the art for FVIII sequences that express efficiently in heterologous systems.

SUMMARY OF THE INVENTION

[0007] The present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence at least 90% identical to SEQ ID NO:1, wherein the nucleotide sequence encodes a polypeptide with Factor VIII activity. In another embodiment, the invention provides an isolated nucleic acid molecule comprising a nucleotide sequence at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 100% identical to SEQ ID NO:1. In other embodiments, the invention provides an isolated nucleic acid molecule comprising SEQ ID NO:1.

[0008] The present invention also provides an isolated nucleic acid molecule comprising a nucleotide sequence at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 100% identical to SEQ ID NO:2, wherein the nucleotide sequence encodes a polypeptide with Factor VIII activity. In one embodiment, the invention provides an isolated nucleic acid molecule comprising SEQ ID NO:2.

[0009] In some embodiments, the isolated nucleic acid molecule of the invention has a human codon adaptation index that is increased relative to SEQ ID NO:3. In other embodiments, the isolated nucleic acid molecule of the invention has a human codon adaptation index that is at least about 0.75, at least about 0.76, at least about 0.77, at least about 0.78, at least about 0.79, or at least about 0.80. In still other embodiments, the isolated nucleic acid molecule of the invention has a human codon adaptation index that is at least about 0.80, at least about 0.81, at least about 0.82, at least about 0.83, at least about 0.84, at least about 0.85, at least about 0.86, at least about 0.87, or at least about 0.88.

[0010] In certain embodiments, the isolated nucleic acid molecule of the invention contains a higher percentage of G/C nucleotides compared to the percentage of G/C nucleotides in SEQ

ID NO:3. In other embodiments, the isolated nucleic acid molecule of the invention contains a percentage of G/C nucleotides that is at least about 45%, at least about 46%, at least about 47%, at least about 48%, at least about 49%, or at least about 50%.

[0011] In still other embodiments, the isolated nucleic acid molecule of the invention contains fewer MARS/ARS sequences (SEQ ID NOs:5 and 6) relative to SEQ ID NO:3. In yet other embodiments, the isolated nucleic acid molecule of the invention contains at most one MARS/ARS sequence. In some embodiments, the isolated nucleic acid molecule of the invention does not contain a MARS/ARS sequence.

[0012] In some embodiments, the isolated nucleic acid molecule of the invention does not contain the splice site GGTGAT (SEQ ID NO:7).

[0013] In certain embodiments, the isolated nucleic acid molecule of the invention contains fewer destabilizing sequences (SEQ ID NOs:8 and 9) relative to SEQ ID NO:3. In other embodiments, the isolated nucleic acid molecule of the invention contains at most 4 destabilizing sequences. In still other embodiments, the isolated nucleic acid molecule of the invention contains at most 2 destabilizing sequences. In yet other embodiments, the isolated nucleic acid molecule of the invention does not contain a destabilizing sequence.

[0014] In other embodiments, the isolated nucleic acid molecule of the invention does not contain a poly-T sequence (SEQ ID NO:10). In yet other embodiments, the isolated nucleic acid molecule of the invention does not contain a poly-A sequence (SEQ ID NO:11).

[0015] In one embodiment, the isolated nucleic acid molecule of the invention further comprises a heterologous nucleotide sequence. For example, the heterologous nucleotide sequence can encode a heterologous amino acid sequence that is a half-life extender. In some embodiments, the heterologous amino acid sequence is an immunoglobulin constant region or a portion thereof, transferrin, albumin, albumin-binding polypeptide, an XTEN sequence, Fc, the C-terminal peptide (CTP) of the β subunit of human chorionic gonadotropin, or a PAS sequence. In other embodiments, the heterologous amino acid sequence is an Fc region or an FcRn binding partner. In still other embodiments, the heterologous amino acid sequence is linked to the N-terminus or the C-terminus of the amino acid sequence encoded by the nucleotide sequence or inserted between two amino acids in the amino acid sequence encoded by the nucleotide sequence.

[0016] In a particular embodiment, the isolated nucleic acid molecule of the invention encodes a monomer-dimer hybrid molecule comprising Factor VIII.

[0017] In another embodiment, the isolated nucleic acid molecule of the invention is operatively linked to at least one transcription control sequence.

[0018] The present invention also provides a vector comprising the nucleic acid molecule of the invention.

[0019] The present invention also provides a host cell comprising the nucleic acid molecule of the invention. In some embodiments, the host cell is selected from the group consisting of: a CHO cell, a HEK293 cell, a BHK21 cell, a PER.C6 cell, a NS0 cell, and a CAP cell.

[0020] The present invention also provides a method of producing a polypeptide with Factor VIII activity, comprising: culturing the host cell of the invention under conditions whereby a polypeptide with Factor VIII activity is produced; and, recovering the polypeptide with Factor VIII activity. In other embodiments of the method of producing a polypeptide with Factor VIII activity, the expression of the polypeptide with Factor VIII activity is increased relative to a host cell cultured under the same conditions comprising a reference nucleotide sequence comprising SEQ ID NO: 3. In other embodiments of the method, the host cell is a CHO cell. In other embodiments of the method, the host cell is a HEK293 cell.

[0021] The present invention also provides a method of increasing expression of a polypeptide with Factor VIII activity in a subject comprising administering the isolated nucleic acid molecule of the invention or the vector of the invention to a subject in need thereof, wherein the expression of the polypeptide with Factor VIII activity is increased relative to a reference nucleic acid molecule comprising SEQ ID NO: 3 or the vector comprising the reference nucleic acid molecule.

[0022] The present invention also provides a method of increasing expression of a polypeptide with Factor VIII activity comprising culturing the host cell of the invention under conditions whereby a polypeptide with Factor VIII activity is expressed by the nucleic acid molecule, wherein the expression of the polypeptide with Factor VIII activity is increased relative to a host cell cultured under the same conditions comprising a reference nucleic acid sequence comprising SEQ ID NO: 3.

[0023] The present invention also provides a method of improving yield of a polypeptide with Factor VIII activity comprising culturing the host cell of the invention under conditions whereby a polypeptide with Factor VIII activity is produced by the nucleic acid molecule, wherein the yield of the polypeptide with Factor VIII activity is increased relative to a host cell cultured under the same conditions comprising a reference nucleic acid sequence comprising SEQ ID NO: 3.

[0024] The present invention also provides a method of treating a bleeding disorder comprising: administering to a subject in need thereof a nucleic acid molecule of the invention, a vector of the invention, or a polypeptide of the invention. In some embodiments of the method of treating a bleeding disorder, the bleeding disorder is characterized by a deficiency in Factor VIII. In some embodiments, the bleeding disorder is hemophilia. In some embodiments, the bleeding disorder is hemophilia A.

[0025] In some embodiments of the method of treating a bleeding disorder, plasma Factor VIII activity at 24 hours post administration is increased relative to a subject administered a

reference nucleic acid molecule comprising SEQ ID NO: 3, a vector comprising the reference nucleic acid molecule, or a polypeptide encoded by the reference nucleic acid molecule.

BRIEF DESCRIPTION OF THE DRAWINGS

[0026]

Figure 1 is a schematic diagram showing the location of various sites in the BDD Factor VIII coding sequence. These sites were removed during the codon optimization process.

Figure 2 is the nucleotide sequence of BDD Factor VIII (SEQ ID NO:1), codon optimized by a first codon optimization method, described in Example 1.

Figure 3 is the nucleotide sequence of BDD Factor VIII (SEQ ID NO:2), codon optimized by a second codon optimization method, described in Example 2.

Figures 4 A-B show the codon usage bias adjustment in the optimized BDD FVIII sequence (SEQ ID NO:1) Figure 4A shows the relative frequency of codons in the BDD FVIII sequence before codon optimization. The human codon adaptation index (CAI) of the starting BDD FVIII sequence is 0.74. Figure 4B shows the relative frequency of codons in the optimized BDD FVIII sequence (SEQ ID NO:1). The human CAI of the resulting optimized sequence is 0.88. The X-axis indicates the relative position of the codons along the length of the BDD FVIII nucleotide sequence. The Y-axis indicates the relative frequency of the codon at each position within the human genome.

Figures 5 A-B show the frequency of optimal human codons in the optimized BDD FVIII sequence (SEQ ID NO:1).

Figure 5A shows the frequency of optimal codons in the BDD FVIII sequence before codon optimization. Figure 5B shows the frequency of optimal codons in the BDD FVIII sequence after codon optimization (SEQ ID NO:1). The X-axis indicates codon frequency in the human genome. The Y-axis indicates the percentage of codons in the BDD FVIII sequence that fall into each category delineated on the X-axis.

Figures 6 A-B shows the G/C content of the optimized BDD FVIII sequence (SEQ ID NO:1). Figure 6A shows the G/C content of the BDD FVIII sequence before codon optimization. The G/C content of the starting BDD FVIII sequence is 46.16%. Figure 6B shows the G/C content of the BDD FVIII sequence after codon optimization (SEQ ID NO:1). The G/C content of the optimized BDD FVIII sequence is 51.56%. The X-axis indicates the relative position of the codons along the length of the BDD FVIII nucleotide sequence. The Y-axis indicates the percent G/C content.

Figure 7 is a histogram showing plasma FVIII activity in HemA mice 24 hours post hydrodynamic injection with plasmids containing either the starting BDD FVIII sequence (circles), optimized BDD FVIII sequence (SEQ ID NO:1) (squares), or optimized BDD FVIII

sequence (SEQ ID NO:2) (triangles).

DETAILED DESCRIPTION OF THE INVENTION

[0027] Exemplary constructs of the invention are illustrated in the accompanying Figures and sequence listing. In order to provide a clear understanding of the specification and claims, the following definitions are provided below.

I. Definitions

[0028] It is to be noted that the term "a" or "an" entity refers to one or more of that entity: for example, "a nucleotide sequence" is understood to represent one or more nucleotide sequences. As such, the terms "a" (or "an"), "one or more," and "at least one" can be used interchangeably herein.

[0029] The term "about" is used herein to mean approximately, roughly, around, or in the regions of. When the term "about" is used in conjunction with a numerical range, it modifies that range by extending the boundaries above and below the numerical values set forth. In general, the term "about" is used herein to modify a numerical value above and below the stated value by a variance of 10 percent, up or down (higher or lower).

[0030] The term "isolated" for the purposes of the present invention designates a biological material (cell, nucleic acid or protein) that has been removed from its original environment (the environment in which it is naturally present). For example, a polynucleotide present in the natural state in a plant or an animal is not isolated, however the same polynucleotide separated from the adjacent nucleic acids in which it is naturally present, is considered "isolated."

[0031] "Nucleic acid," "nucleic acid molecule," "oligonucleotide," and "polynucleotide" are used interchangeably and refer to the phosphate ester polymeric form of ribonucleosides (adenosine, guanosine, uridine or cytidine; "RNA molecules") or deoxyribonucleosides (deoxyadenosine, deoxyguanosine, deoxythymidine, or deoxycytidine; "DNA molecules"), or any phosphoester analogs thereof, such as phosphorothioates and thioesters, in either single stranded form, or a double-stranded helix. Double stranded DNA-DNA, DNA-RNA and RNA-RNA helices are possible. The term nucleic acid molecule, and in particular DNA or RNA molecule, refers only to the primary and secondary structure of the molecule, and does not limit it to any particular tertiary forms. Thus, this term includes double-stranded DNA found, *inter alia*, in linear or circular DNA molecules (e.g., restriction fragments), plasmids, supercoiled DNA and chromosomes. In discussing the structure of particular double-stranded DNA molecules, sequences can be described herein according to the normal convention of giving

only the sequence in the 5' to 3' direction along the non-transcribed strand of DNA (*i.e.*, the strand having a sequence homologous to the mRNA). A "recombinant DNA molecule" is a DNA molecule that has undergone a molecular biological manipulation. DNA includes, but is not limited to, cDNA, genomic DNA, plasmid DNA, synthetic DNA, and semi-synthetic DNA. A "nucleic acid composition" of the invention comprises one or more nucleic acids as described herein.

[0032] As used herein, a "coding region" or "coding sequence" is a portion of polynucleotide which consists of codons translatable into amino acids. Although a "stop codon" (TAG, TGA, or TAA) is typically not translated into an amino acid, it can be considered to be part of a coding region, but any flanking sequences, for example promoters, ribosome binding sites, transcriptional terminators, introns, and the like, are not part of a coding region. The boundaries of a coding region are typically determined by a start codon at the 5' terminus, encoding the amino terminus of the resultant polypeptide, and a translation stop codon at the 3' terminus, encoding the carboxyl terminus of the resulting polypeptide. Two or more coding regions can be present in a single polynucleotide construct, *e.g.*, on a single vector, or in separate polynucleotide constructs, *e.g.*, on separate (different) vectors. It follows, then, that a single vector can contain just a single coding region, or comprise two or more coding regions.

[0033] Certain proteins secreted by mammalian cells are associated with a secretory signal peptide which is cleaved from the mature protein once export of the growing protein chain across the rough endoplasmic reticulum has been initiated. Those of ordinary skill in the art are aware that signal peptides are generally fused to the N-terminus of the polypeptide, and are cleaved from the complete or "full-length" polypeptide to produce a secreted or "mature" form of the polypeptide. In certain embodiments, a native signal peptide or a functional derivative of that sequence that retains the ability to direct the secretion of the polypeptide that is operably associated with it. Alternatively, a heterologous mammalian signal peptide, *e.g.*, a human tissue plasminogen activator (TPA) or mouse β -glucuronidase signal peptide, or a functional derivative thereof, can be used.

[0034] The term "downstream" refers to a nucleotide sequence that is located 3' to a reference nucleotide sequence. In certain embodiments, downstream nucleotide sequences relate to sequences that follow the starting point of transcription. For example, the translation initiation codon of a gene is located downstream of the start site of transcription.

[0035] The term "upstream" refers to a nucleotide sequence that is located 5' to a reference nucleotide sequence. In certain embodiments, upstream nucleotide sequences relate to sequences that are located on the 5' side of a coding region or starting point of transcription. For example, most promoters are located upstream of the start site of transcription.

[0036] As used herein, the term "regulatory region" refers to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding region, and which influence the transcription, RNA processing, stability, or translation of the associated coding region. Regulatory regions can include promoters, translation leader

sequences, introns, polyadenylation recognition sequences, RNA processing sites, effector binding sites and stem-loop structures. If a coding region is intended for expression in a eukaryotic cell, a polyadenylation signal and transcription termination sequence will usually be located 3' to the coding sequence.

[0037] A polynucleotide which encodes a gene product, *e.g.*, a polypeptide, can include a promoter and/or other transcription or translation control elements operably associated with one or more coding regions. In an operable association a coding region for a gene product, *e.g.*, a polypeptide, is associated with one or more regulatory regions in such a way as to place expression of the gene product under the influence or control of the regulatory region(s). For example, a coding region and a promoter are "operably associated" if induction of promoter function results in the transcription of mRNA encoding the gene product encoded by the coding region, and if the nature of the linkage between the promoter and the coding region does not interfere with the ability of the promoter to direct the expression of the gene product or interfere with the ability of the DNA template to be transcribed. Other transcription control elements, besides a promoter, for example enhancers, operators, repressors, and transcription termination signals, can also be operably associated with a coding region to direct gene product expression.

[0038] "Transcriptional control sequences" refer to DNA regulatory sequences, such as promoters, enhancers, terminators, and the like, that provide for the expression of a coding sequence in a host cell. A variety of transcription control regions are known to those skilled in the art. These include, without limitation, transcription control regions which function in vertebrate cells, such as, but not limited to, promoter and enhancer segments from cytomegaloviruses (the immediate early promoter, in conjunction with intron-A), simian virus 40 (the early promoter), and retroviruses (such as Rous sarcoma virus). Other transcription control regions include those derived from vertebrate genes such as actin, heat shock protein, bovine growth hormone and rabbit β -globin, as well as other sequences capable of controlling gene expression in eukaryotic cells. Additional suitable transcription control regions include tissue-specific promoters and enhancers as well as lymphokine-inducible promoters (*e.g.*, promoters inducible by interferons or interleukins).

[0039] Similarly, a variety of translation control elements are known to those of ordinary skill in the art. These include, but are not limited to ribosome binding sites, translation initiation and termination codons, and elements derived from picornaviruses (particularly an internal ribosome entry site, or IRES, also referred to as a CITE sequence).

[0040] The term "expression" as used herein refers to a process by which a polynucleotide produces a gene product, for example, an RNA or a polypeptide. It includes without limitation transcription of the polynucleotide into messenger RNA (mRNA), transfer RNA (tRNA), small hairpin RNA (shRNA), small interfering RNA (siRNA) or any other RNA product, and the translation of an mRNA into a polypeptide. Expression produces a "gene product." As used herein, a gene product can be either a nucleic acid, *e.g.*, a messenger RNA produced by transcription of a gene, or a polypeptide which is translated from a transcript. Gene products

described herein further include nucleic acids with post transcriptional modifications, e.g., polyadenylation or splicing, or polypeptides with post translational modifications, e.g., methylation, glycosylation, the addition of lipids, association with other protein subunits, or proteolytic cleavage. The term "yield," as used herein, refers to the amount of a polypeptide produced by the expression of a gene.

[0041] A "vector" refers to any vehicle for the cloning of and/or transfer of a nucleic acid into a host cell. A vector can be a replicon to which another nucleic acid segment can be attached so as to bring about the replication of the attached segment. A "replicon" refers to any genetic element (e.g., plasmid, phage, cosmid, chromosome, virus) that functions as an autonomous unit of replication *in vivo*, i.e., capable of replication under its own control. The term "vector" includes both viral and nonviral vehicles for introducing the nucleic acid into a cell *in vitro*, *ex vivo* or *in vivo*. A large number of vectors are known and used in the art including, for example, plasmids, modified eukaryotic viruses, or modified bacterial viruses. Insertion of a polynucleotide into a suitable vector can be accomplished by ligating the appropriate polynucleotide fragments into a chosen vector that has complementary cohesive termini.

[0042] Vectors can be engineered to encode selectable markers or reporters that provide for the selection or identification of cells that have incorporated the vector. Expression of selectable markers or reporters allows identification and/or selection of host cells that incorporate and express other coding regions contained on the vector. Examples of selectable marker genes known and used in the art include: genes providing resistance to ampicillin, streptomycin, gentamycin, kanamycin, hygromycin, bialaphos herbicide, sulfonamide, and the like; and genes that are used as phenotypic markers, i.e., anthocyanin regulatory genes, isopentanyl transferase gene, and the like. Examples of reporters known and used in the art include: luciferase (Luc), green fluorescent protein (GFP), chloramphenicol acetyltransferase (CAT), -galactosidase (LacZ), -glucuronidase (Gus), and the like. Selectable markers can also be considered to be reporters.

[0043] The term "selectable marker" refers to an identifying factor, usually an antibiotic or chemical resistance gene, that is able to be selected for based upon the marker gene's effect, i.e., resistance to an antibiotic, resistance to a herbicide, colorimetric markers, enzymes, fluorescent markers, and the like, wherein the effect is used to track the inheritance of a nucleic acid of interest and/or to identify a cell or organism that has inherited the nucleic acid of interest. Examples of selectable marker genes known and used in the art include: genes providing resistance to ampicillin, streptomycin, gentamycin, kanamycin, hygromycin, bialaphos herbicide, sulfonamide, and the like; and genes that are used as phenotypic markers, i.e., anthocyanin regulatory genes, isopentanyl transferase gene, and the like.

[0044] The term "reporter gene" refers to a nucleic acid encoding an identifying factor that is able to be identified based upon the reporter gene's effect, wherein the effect is used to track the inheritance of a nucleic acid of interest, to identify a cell or organism that has inherited the nucleic acid of interest, and/or to measure gene expression induction or transcription. Examples of reporter genes known and used in the art include: luciferase (Luc), green

fluorescent protein (GFP), chloramphenicol acetyltransferase (CAT), β -galactosidase (LacZ), β -glucuronidase (Gus), and the like. Selectable marker genes can also be considered reporter genes.

[0045] "Promoter" and "promoter sequence" are used interchangeably and refer to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. Promoters can be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters can direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental or physiological conditions. Promoters that cause a gene to be expressed in most cell types at most times are commonly referred to as "constitutive promoters." Promoters that cause a gene to be expressed in a specific cell type are commonly referred to as "cell-specific promoters" or "tissue-specific promoters." Promoters that cause a gene to be expressed at a specific stage of development or cell differentiation are commonly referred to as "developmentally-specific promoters" or "cell differentiation-specific promoters." Promoters that are induced and cause a gene to be expressed following exposure or treatment of the cell with an agent, biological molecule, chemical, ligand, light, or the like that induces the promoter are commonly referred to as "inducible promoters" or "regulatable promoters." It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths can have identical promoter activity.

[0046] The promoter sequence is typically bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence will be found a transcription initiation site (conveniently defined for example, by mapping with nuclease S1), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase.

[0047] The terms "restriction endonuclease" and "restriction enzyme" are used interchangeably and refer to an enzyme that binds and cuts within a specific nucleotide sequence within double stranded DNA.

[0048] The term "plasmid" refers to an extra-chromosomal element often carrying a gene that is not part of the central metabolism of the cell, and usually in the form of circular double-stranded DNA molecules. Such elements can be autonomously replicating sequences, genome integrating sequences, phage or nucleotide sequences, linear, circular, or supercoiled, of a single- or double-stranded DNA or RNA, derived from any source, in which a number of nucleotide sequences have been joined or recombined into a unique construction which is capable of introducing a promoter fragment and DNA sequence for a selected gene product along with appropriate 3' untranslated sequence into a cell.

[0049] Eukaryotic viral vectors that can be used include, but are not limited to, adenovirus vectors, retrovirus vectors, adeno-associated virus vectors, poxvirus, e.g., vaccinia virus vectors, baculovirus vectors, or herpesvirus vectors. Non-viral vectors include plasmids, liposomes, electrically charged lipids (cytofectins), DNA-protein complexes, and biopolymers.

[0050] A "cloning vector" refers to a "replicon, " which is a unit length of a nucleic acid that replicates sequentially and which comprises an origin of replication, such as a plasmid, phage or cosmid, to which another nucleic acid segment can be attached so as to bring about the replication of the attached segment. Certain cloning vectors are capable of replication in one cell type, e.g., bacteria and expression in another, e.g., eukaryotic cells. Cloning vectors typically comprise one or more sequences that can be used for selection of cells comprising the vector and/or one or more multiple cloning sites for insertion of nucleic acid sequences of interest.

[0051] The term "expression vector" refers to a vehicle designed to enable the expression of an inserted nucleic acid sequence following insertion into a host cell. The inserted nucleic acid sequence is placed in operable association with regulatory regions as described above.

[0052] Vectors are introduced into host cells by methods well known in the art, e.g., transfection, electroporation, microinjection, transduction, cell fusion, DEAE dextran, calcium phosphate precipitation, lipofection (lysosome fusion), use of a gene gun, or a DNA vector transporter.

[0053] "Culture," "to culture" and "culturing," as used herein, means to incubate cells under *in vitro* conditions that allow for cell growth or division or to maintain cells in a living state. "Cultured cells," as used herein, means cells that are propagated *in vitro*.

[0054] As used herein, the term "polypeptide" is intended to encompass a singular "polypeptide" as well as plural "polypeptides," and refers to a molecule composed of monomers (amino acids) linearly linked by amide bonds (also known as peptide bonds). The term "polypeptide" refers to any chain or chains of two or more amino acids, and does not refer to a specific length of the product. Thus, peptides, dipeptides, tripeptides, oligopeptides, "protein," "amino acid chain," or any other term used to refer to a chain or chains of two or more amino acids, are included within the definition of "polypeptide," and the term "polypeptide" can be used instead of, or interchangeably with any of these terms. The term "polypeptide" is also intended to refer to the products of post-expression modifications of the polypeptide, including without limitation glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, or modification by non-naturally occurring amino acids. A polypeptide can be derived from a natural biological source or produced recombinant technology, but is not necessarily translated from a designated nucleic acid sequence. It can be generated in any manner, including by chemical synthesis.

[0055] The term "amino acid" includes alanine (Ala or A); arginine (Arg or R); asparagine (Asn

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

[0056] An "isolated" polypeptide or a fragment, variant, or derivative thereof refers to a polypeptide that is not in its natural milieu. No particular level of purification is required. For example, an isolated polypeptide can simply be removed from its native or natural environment. Recombinantly produced polypeptides and proteins expressed in host cells are considered isolated for the purpose of the invention, as are native or recombinant polypeptides which have been separated, fractionated, or partially or substantially purified by any suitable technique.

[0057] Also included in the present invention are fragments or variants of polypeptides, and any combination thereof. The term "fragment" or "variant" when referring to polypeptide binding domains or binding molecules of the present invention include any polypeptides which retain at least some of the properties (e.g., FcRn binding affinity for an FcRn binding domain or Fc variant, coagulation activity for an FVIII variant, or FVIII binding activity for the VWF fragment) of the reference polypeptide. Fragments of polypeptides include proteolytic fragments, as well as deletion fragments, in addition to specific antibody fragments discussed elsewhere herein, but do not include the naturally occurring full-length polypeptide (or mature polypeptide). Variants of polypeptide binding domains or binding molecules of the present invention include fragments as described above, and also polypeptides with altered amino acid sequences due to amino acid substitutions, deletions, or insertions. Variants can be naturally or non-naturally occurring. Non-naturally occurring variants can be produced using art-known mutagenesis techniques. Variant polypeptides can comprise conservative or non-conservative amino acid substitutions, deletions or additions.

[0058] A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid

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

[0059] The term "percent identity" as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case can be, as determined by the match between strings of such sequences. "Identity" can be readily calculated by known methods, including but not limited to those described in: Computational Molecular Biology (Lesk, A. M., ed.) Oxford University Press, New York (1988); Biocomputing: Informatics and Genome Projects (Smith, D. W., ed.) Academic Press, New York (1993); Computer Analysis of Sequence Data, Part I (Griffin, A. M., and Griffin, H. G., eds.) Humana Press, New Jersey (1994); Sequence Analysis in Molecular Biology (von Heinje, G., ed.) Academic Press (1987); and Sequence Analysis Primer (Gribskov, M. and Devereux, J., eds.) Stockton Press, New York (1991). Preferred methods to determine identity are designed to give the best match between the sequences tested. Methods to determine identity are codified in publicly available computer programs. Sequence alignments and percent identity calculations can be performed using sequence analysis software such as the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI), the GCG suite of programs (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI), BLASTP, BLASTN, BLASTX (Altschul et al., J Mol. Biol. 215:403 (1990)), and DNASTAR (DNASTAR, Inc. 1228 S. Park St. Madison, WI 53715 USA). Within the context of this application it will be understood that where sequence analysis software is used for analysis, that the results of the analysis will be based on the "default values" of the program referenced, unless otherwise specified. As used herein "default values" will mean any set of values or parameters which originally load with the software when first initialized. For the purposes of determining percent identity between an optimized BDD FVIII sequence of the invention and a reference sequence, only nucleotides in the reference sequence corresponding to nucleotides in the optimized BDD FVIII sequence of the invention are used to calculate percent identity. For example, when comparing a full length FVIII nucleotide sequence containing the B domain to an optimized B domain deleted (BDD) FVIII nucleotide sequence of the invention, the portion of the alignment including the A1, A2, A3, C1, and C2 domain will be used to calculate percent identity. The nucleotides in the portion of the full length FVIII sequence encoding the B domain (which will result in a large "gap" in the alignment) will not be counted as a mismatch.

[0060] As used herein, "nucleotides corresponding to nucleotides in the optimized BDD FVIII

sequence of the invention" are identified by alignment of the optimized BDD FVIII sequence of the invention to maximize the identity to the reference FVIII sequence. The number used to identify an equivalent amino acid in a reference FVIII sequence is based on the number used to identify the corresponding amino acid in the optimized BDD FVIII sequence of the invention.

[0061] A "fusion" or "chimeric" protein comprises a first amino acid sequence linked to a second amino acid sequence with which it is not naturally linked in nature. The amino acid sequences which normally exist in separate proteins can be brought together in the fusion polypeptide, or the amino acid sequences which normally exist in the same protein can be placed in a new arrangement in the fusion polypeptide, e.g., fusion of a Factor VIII domain of the invention with an Ig Fc domain. A fusion protein is created, for example, by chemical synthesis, or by creating and translating a polynucleotide in which the peptide regions are encoded in the desired relationship. A chimeric protein can further comprises a second amino acid sequence associated with the first amino acid sequence by a covalent, non-peptide bond or a non-covalent bond.

[0062] As used herein, the term "half-life" refers to a biological half-life of a particular polypeptide *in vivo*. Half-life can be represented by the time required for half the quantity administered to a subject to be cleared from the circulation and/or other tissues in the animal. When a clearance curve of a given polypeptide is constructed as a function of time, the curve is usually biphasic with a rapid α -phase and longer β -phase. The α -phase typically represents an equilibration of the administered Fc polypeptide between the intra- and extra-vascular space and is, in part, determined by the size of the polypeptide. The β -phase typically represents the catabolism of the polypeptide in the intravascular space. In some embodiments, FVIII and chimeric proteins comprising FVIII are monophasic, and thus do not have an alpha phase, but just the single beta phase. Therefore, in certain embodiments, the term half-life as used herein refers to the half-life of the polypeptide in the β -phase.

[0063] The term "linked" as used herein refers to a first amino acid sequence or nucleotide sequence covalently or non-covalently joined to a second amino acid sequence or nucleotide sequence, respectively. The first amino acid or nucleotide sequence can be directly joined or juxtaposed to the second amino acid or nucleotide sequence or alternatively an intervening sequence can covalently join the first sequence to the second sequence. The term "linked" means not only a fusion of a first amino acid sequence to a second amino acid sequence at the C-terminus or the N-terminus, but also includes insertion of the whole first amino acid sequence (or the second amino acid sequence) into any two amino acids in the second amino acid sequence (or the first amino acid sequence, respectively). In one embodiment, the first amino acid sequence can be linked to a second amino acid sequence by a peptide bond or a linker. The first nucleotide sequence can be linked to a second nucleotide sequence by a phosphodiester bond or a linker. The linker can be a peptide or a polypeptide (for polypeptide chains) or a nucleotide or a nucleotide chain (for nucleotide chains) or any chemical moiety (for both polypeptide and polynucleotide chains). The term "linked" is also indicated by a hyphen (-).

[0064] As used herein the term "associated with" refers to a covalent or non-covalent bond formed between a first amino acid chain and a second amino acid chain. In one embodiment, the term "associated with" means a covalent, non-peptide bond or a non-covalent bond. This association can be indicated by a colon, *i.e.*, (:). In another embodiment, it means a covalent bond except a peptide bond. For example, the amino acid cysteine comprises a thiol group that can form a disulfide bond or bridge with a thiol group on a second cysteine residue. In most naturally occurring IgG molecules, the CH1 and CL regions are associated by a disulfide bond and the two heavy chains are associated by two disulfide bonds at positions corresponding to 239 and 242 using the Kabat numbering system (position 226 or 229, EU numbering system). Examples of covalent bonds include, but are not limited to, a peptide bond, a metal bond, a hydrogen bond, a disulfide bond, a sigma bond, a pi bond, a delta bond, a glycosidic bond, an agnostic bond, a bent bond, a dipolar bond, a Pi backbond, a double bond, a triple bond, a quadruple bond, a quintuple bond, a sextuple bond, conjugation, hyperconjugation, aromaticity, hapticity, or antibonding. Non-limiting examples of non-covalent bond include an ionic bond (*e.g.*, cation-pi bond or salt bond), a metal bond, an hydrogen bond (*e.g.*, dihydrogen bond, dihydrogen complex, low-barrier hydrogen bond, or symmetric hydrogen bond), van der Waals force, London dispersion force, a mechanical bond, a halogen bond, aurophilicity, intercalation, stacking, entropic force, or chemical polarity.

[0065] The term "monomer-dimer hybrid" used herein refers to a chimeric protein comprising a first polypeptide chain and a second polypeptide chain, which are associated with each other by a disulfide bond, wherein the first chain comprises a clotting factor, *e.g.*, Factor VIII, and a first Fc region and the second chain comprises, consists essentially of, or consists of a second Fc region without the clotting factor. The monomer-dimer hybrid construct thus is a hybrid comprising a monomer aspect having only one clotting factor and a dimer aspect having two Fc regions.

[0066] Hemostasis, as used herein, means the stopping or slowing of bleeding or hemorrhage; or the stopping or slowing of blood flow through a blood vessel or body part.

[0067] Hemostatic disorder, as used herein, means a genetically inherited or acquired condition characterized by a tendency to hemorrhage, either spontaneously or as a result of trauma, due to an impaired ability or inability to form a fibrin clot. Examples of such disorders include the hemophilias. The three main forms are hemophilia A (factor VIII deficiency), hemophilia B (factor IX deficiency or "Christmas disease") and hemophilia C (factor XI deficiency, mild bleeding tendency). Other hemostatic disorders include, *e.g.*, von Willebrand disease, Factor XI deficiency (PTA deficiency), Factor XII deficiency, deficiencies or structural abnormalities in fibrinogen, prothrombin, Factor V, Factor VII, Factor X or factor XIII, Bernard-Soulier syndrome, which is a defect or deficiency in GPIb. GPIb, the receptor for vWF, can be defective and lead to lack of primary clot formation (primary hemostasis) and increased bleeding tendency), and thrombasthenia of Glanzman and Naegeli (Glanzmann thrombasthenia). In liver failure (acute and chronic forms), there is insufficient production of coagulation factors by the liver; this can increase bleeding risk.

[0068] The isolated nucleic acid molecules of the invention can be used prophylactically. As used herein the term "prophylactic treatment" refers to the administration of a molecule prior to a bleeding episode. In one embodiment, the subject in need of a general hemostatic agent is undergoing, or is about to undergo, surgery. The chimeric protein of the invention can be administered prior to or after surgery as a prophylactic. The chimeric protein of the invention can be administered during or after surgery to control an acute bleeding episode. The surgery can include, but is not limited to, liver transplantation, liver resection, dental procedures, or stem cell transplantation.

[0069] The isolated nucleic acid molecules of the invention are also used for on-demand treatment. The term "on-demand treatment" refers to the administration of an isolated nucleic acid molecule in response to symptoms of a bleeding episode or before an activity that can cause bleeding. In one aspect, the on-demand treatment can be given to a subject when bleeding starts, such as after an injury, or when bleeding is expected, such as before surgery. In another aspect, the on-demand treatment can be given prior to activities that increase the risk of bleeding, such as contact sports.

[0070] As used herein the term "acute bleeding" refers to a bleeding episode regardless of the underlying cause. For example, a subject can have trauma, uremia, a hereditary bleeding disorder (e.g., factor VII deficiency) a platelet disorder, or resistance owing to the development of antibodies to clotting factors.

[0071] Treat, treatment, treating, as used herein refers to, e.g., the reduction in severity of a disease or condition; the reduction in the duration of a disease course; the amelioration of one or more symptoms associated with a disease or condition; the provision of beneficial effects to a subject with a disease or condition, without necessarily curing the disease or condition, or the prophylaxis of one or more symptoms associated with a disease or condition. In one embodiment, the term "treating" or "treatment" means maintaining a FVIII trough level at least about 1 IU/dL, 2 IU/dL, 3 IU/dL, 4 IU/dL, 5 IU/dL, 6 IU/dL, 7 IU/dL, 8 IU/dL, 9 IU/dL, 10 IU/dL, 11 IU/dL, 12 IU/dL, 13 IU/dL, 14 IU/dL, 15 IU/dL, 16 IU/dL, 17 IU/dL, 18 IU/dL, 19 IU/dL, or 20 IU/dL in a subject by administering an isolated nucleic acid molecule of the invention. In another embodiment, treating or treatment means maintaining a FVIII trough level between about 1 and about 20 IU/dL, about 2 and about 20 IU/dL, about 3 and about 20 IU/dL, about 4 and about 20 IU/dL, about 5 and about 20 IU/dL, about 6 and about 20 IU/dL, about 7 and about 20 IU/dL, about 8 and about 20 IU/dL, about 9 and about 20 IU/dL, or about 10 and about 20 IU/dL. Treatment or treating of a disease or condition can also include maintaining FVIII activity in a subject at a level comparable to at least about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, or 20% of the FVIII activity in a non-hemophiliac subject. The minimum trough level required for treatment can be measured by one or more known methods and can be adjusted (increased or decreased) for each person.

[0072] "Administering," as used herein, means to give a pharmaceutically acceptable Factor VIII polypeptide to a subject via a pharmaceutically acceptable route. Routes of administration

can be intravenous, *e.g.*, intravenous injection and intravenous infusion. Additional routes of administration include, *e.g.*, subcutaneous, intramuscular, oral, nasal, and pulmonary administration. Chimeric polypeptides and hybrid proteins can be administered as part of a pharmaceutical composition comprising at least one excipient.

[0073] As used herein, the phrase "subject in need thereof" includes subjects, such as mammalian subjects, that would benefit from administration of a nucleic acid molecule of the invention, *e.g.*, to improve hemostasis. In one embodiment, the subjects include, but are not limited to, individuals with hemophilia. In another embodiment, the subjects include, but are not limited to, the individuals who have developed a FVIII inhibitor and thus are in need of a bypass therapy. The subject can be an adult or a minor (*e.g.*, under 12 years old).

[0074] As used herein, the term "clotting factor," refers to molecules, or analogs thereof, naturally occurring or recombinantly produced which prevent or decrease the duration of a bleeding episode in a subject. In other words, it means molecules having pro-clotting activity, *i.e.*, are responsible for the conversion of fibrinogen into a mesh of insoluble fibrin causing the blood to coagulate or clot. An "activatable clotting factor" is a clotting factor in an inactive form (*e.g.*, in its zymogen form) that is capable of being converted to an active form.

[0075] Clotting activity, as used herein, means the ability to participate in a cascade of biochemical reactions that culminates in the formation of a fibrin clot and/or reduces the severity, duration or frequency of hemorrhage or bleeding episode.

[0076] As used herein the terms "heterologous" or "exogenous" refer to such molecules that are not normally found in a given context, *e.g.*, in a cell or in a polypeptide. For example, an exogenous or heterologous molecule can be introduced into a cell and are only present after manipulation of the cell, *e.g.*, by transfection or other forms of genetic engineering or a heterologous amino acid sequence can be present in a protein in which it is not naturally found.

[0077] As used herein, the term "heterologous nucleotide sequence" refers to a nucleotide sequence that does not naturally occur with a given polynucleotide sequence. In one embodiment, the heterologous nucleotide sequence encodes a polypeptide capable of extending the half-life of FVIII. In another embodiment, the heterologous nucleotide sequence encodes a polypeptide that increases the hydrodynamic radius of FVIII. In other embodiments, the heterologous nucleotide sequence encodes a polypeptide that improves one or more pharmacokinetic properties of FVIII without significantly affecting its biological activity or function (*e.g.*, its procoagulant activity). In some embodiments, FVIII is linked or connected to the polypeptide encoded by the heterologous nucleotide sequence by a linker. Non-limiting examples of polypeptide moieties encoded by heterologous nucleotide sequences include an immunoglobulin constant region or a portion thereof, albumin or a fragment thereof, an albumin-binding moiety, a transferrin, the PAS polypeptides of U.S. Pat Application No. 20100292130, a HAP sequence, transferrin or a fragment thereof, the C-terminal peptide (CTP) of the β subunit of human chorionic gonadotropin, albumin-binding small molecule, an

XTEN sequence, FcRn binding moieties (e.g., complete Fc regions or portions thereof which bind to FcRn), single chain Fc regions (ScFc regions, e.g., as described in US 2008/0260738, WO 2008/012543, or WO 2008/1439545), polyglycine linkers, polyserine linkers, peptides and short polypeptides of 6-40 amino acids of two types of amino acids selected from glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P) with varying degrees of secondary structure from less than 50% to greater than 50%, amongst others, or two or more combinations thereof. In some embodiments, the polypeptide encoded by the heterologous nucleotide sequence is linked to a non-polypeptide moiety. Non-limiting examples of the non-polypeptide moieties include polyethylene glycol (PEG), albumin-binding small molecules, polysialic acid, hydroxyethyl starch (HES), a derivative thereof, or any combinations thereof.

[0078] As used herein, the term "Fc region" is defined as the portion of a polypeptide which corresponds to the Fc region of native Ig, *i.e.*, as formed by the dimeric association of the respective Fc domains of its two heavy chains. A native Fc region forms a homodimer with another Fc region. In contrast, the term "genetically-fused Fc region" or "single-chain Fc region" (scFc region), as used herein, refers to a synthetic dimeric Fc region comprised of Fc domains genetically linked within a single polypeptide chain (*i.e.*, encoded in a single contiguous genetic sequence).

[0079] In one embodiment, the "Fc region" refers to the portion of a single Ig heavy chain beginning in the hinge region just upstream of the papain cleavage site (*i.e.* residue 216 in IgG, taking the first residue of heavy chain constant region to be 114) and ending at the C-terminus of the antibody. Accordingly, a complete Fc domain comprises at least a hinge domain, a CH2 domain, and a CH3 domain.

[0080] The Fc region of an Ig constant region, depending on the Ig isotype can include the CH2, CH3, and CH4 domains, as well as the hinge region. Chimeric proteins comprising an Fc region of an Ig bestow several desirable properties on a chimeric protein including increased stability, increased serum half-life (see Capon et al., 1989, Nature 337:525) as well as binding to Fc receptors such as the neonatal Fc receptor (FcRn) (U.S. Pat. Nos. 6,086,875, 6,485,726, 6,030,613; WO 03/077834; US2003-0235536A1).

[0081] A "reference nucleotide sequence," when used herein as a comparison to a nucleotide sequence of the invention, is a polynucleotide sequence essentially identical to the nucleotide sequence of the invention except that the portions corresponding to FVIII sequence are not optimized. For example, the reference nucleotide sequence for a nucleic acid molecule consisting of the codon optimized BDD FVIII of SEQ ID NO:1 and a heterologous nucleotide sequence that encodes a single chain Fc region linked to SEQ ID NO:1 at its 3' end is a nucleic acid molecule consisting of the original (or "parent") BDD FVIII of SEQ ID NO:3 and the identical heterologous nucleotide sequence that encodes a single chain Fc region linked to SEQ ID NO:3 at its 3' end.

[0082] A "codon adaptation index," as used herein, refers to a measure of codon usage bias. A codon adaptation index (CAI) measures the deviation of a given protein coding gene sequence

with respect to a reference set of genes (Sharp PM and Li WH, Nucleic Acids Res. 15(3):1281-95 (1987)). CAI is calculated by determining the geometric mean of the weight associated to each codon over the length of the gene sequence (measured in codons):

$$CAI = \exp\left(\frac{1}{L} \sum_{l=1}^L \ln(w_i(l))\right), \quad (I)$$

[0083] For each amino acid, the weight of each of its codons, in CAI, is computed as the ratio between the observed frequency of the codon (f_i) and the frequency of the synonymous codon (f_j) for that amino acid:

Formula 2:

$$w_i = \frac{f_i}{\max(f_j)} \quad i, j \in [\text{synonymous codons for amino acid}] \quad (II)$$

[0084] As used herein, the term "optimized," with regard to nucleotide sequences, refers to a polynucleotide sequence that encodes a polypeptide, wherein the polynucleotide sequence has been mutated to enhance a property of that polynucleotide sequence. In some embodiments, the optimization is done to increase transcription levels, increase translation levels, increase steady-state mRNA levels, increase or decrease the binding of regulatory proteins such as general transcription factors, increase or decrease splicing, or increase the yield of the polypeptide produced by the polynucleotide sequence. Examples of changes that can be made to a polynucleotide sequence to optimize it include codon optimization, G/C content optimization, removal of repeat sequences, removal of AT rich elements, removal of cryptic splice sites, removal of cis-acting elements that repress transcription or translation, adding or removing poly-T or poly-A sequences, adding sequences around the transcription start site that enhance transcription, such as Kozak consensus sequences, removal of sequences that could form stem loop structures, removal of destabilizing sequences, and two or more combinations thereof.

[0085] The present invention is directed to optimized Factor VIII sequences, vectors and host cells comprising optimized Factor VIII sequences, and methods of producing polypeptides encoded by optimized Factor VIII sequences. The present invention is also directed to methods of treating bleeding disorders such as hemophilia comprising administering to the subject an optimized Factor VIII nucleic acid sequence. The present invention meets an important need in the art by providing optimized Factor VIII sequences that demonstrate increased expression in host cells, improved yield of Factor VIII protein in methods to produce recombinant Factor VIII, and potentially result in greater therapeutic efficacy when used in gene therapy methods.

[0086] In some embodiments, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide with Factor VIII (FVIII) activity, wherein the nucleotide sequence is at least 90% identical to SEQ ID NO:1. In other

embodiments, the nucleotide sequence is at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO:1 and encodes a polypeptide with FVIII activity. In still other embodiments, the nucleotide sequence comprises SEQ ID NO:1.

[0087] In some embodiments, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide with FVIII activity, wherein the nucleotide sequence is at least 95% identical to SEQ ID NO:2. In other embodiments, the the nucleotide sequence is at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO:2 and encodes a polypeptide with FVIII activity. In still other embodiments, the nucleotide sequence comprises SEQ ID NO:2.

[0088] SEQ ID NOs: 1 and 2 are optimized versions of SEQ ID NO:3, the starting or "parental" FVIII nucleotide sequence. SEQ ID NO:3 encodes a B domain-deleted human FVIII. While SEQ ID NOs:1 and 2 are derived from a specific B domain-deleted form of FVIII (SEQ ID NO:3), it is to be understood that the present invention is also directed to optimized versions of nucleic acids encoding other versions of FVIII. For example, other version of FVIII can include full length FVIII, other B-domain deletions of FVIII (described below), or other fragments of FVIII that retain FVIII activity.

[0089] "A polypeptide with FVIII activity" as used herein means a functional FVIII polypeptide in its normal role in coagulation, unless otherwise specified. The term a polypeptide with FVIII activity includes a functional fragment, variant, analog, or derivative thereof that retains the function of full-length wild-type Factor VIII in the coagulation pathway. "A polypeptide with FVIII activity" is used interchangeably with FVIII protein, FVIII polypeptide, or FVIII. Examples of FVIII functions include, but are not limited to, an ability to activate coagulation, an ability to act as a cofactor for factor IX, or an ability to form a tenase complex with factor IX in the presence of Ca^{2+} and phospholipids, which then converts Factor X to the activated form Xa. In one embodiment, a polypeptide having FVIII activity comprises two polypeptide chains, the first chain having the FVIII heavy chain and the second chain having the FVIII light chain. In another embodiment, the polypeptide having FVIII activity is single chain FVIII. Single chain FVIII can contain one or more mutation or substitutions at amino acid residue 1645 and/or 1648 corresponding to mature FVIII sequence. See International Application No. PCT/US2012/045784. The FVIII protein can be the human, porcine, canine, rat, or murine FVIII protein. In addition, comparisons between FVIII from humans and other species have identified conserved residues that are likely to be required for function (Cameron et al., *Thromb. Haemost.* 79:317-22 (1998); US 6,251,632).

[0090] The "B domain" of FVIII, as used herein, is the same as the B domain known in the art that is defined by internal amino acid sequence identity and sites of proteolytic cleavage by thrombin, e.g., residues Ser741-Arg1648 of full length human FVIII. The other human FVIII domains are defined by the following amino acid residues: A1, residues Ala1-Arg372; A2, residues Ser373-Arg740; A3, residues Ser1690-Ile2032; C1, residues Arg2033-Asn2172; C2, residues Ser2173-Tyr2332. The A3-C1-C2 sequence includes residues Ser1690-Tyr2332. The

remaining sequence, residues Glu1649-Arg1689, is usually referred to as the FVIII light chain activation peptide. The locations of the boundaries for all of the domains, including the B domains, for porcine, mouse and canine FVIII are also known in the art. An example of a BDD FVIII is REFACTO® recombinant BDD FVIII (Wyeth Pharmaceuticals, Inc.).

[0091] A "B domain deleted FVIII" can have the full or partial deletions disclosed in U.S. Patent Nos. 6,316,226, 6,346,513, 7,041,635, 5,789,203, 6,060,447, 5,595,886, 6,228,620, 5,972,885, 6,048,720, 5,543,502, 5,610,278, 5,171,844, 5,112,950, 4,868,112, and 6,458,563. In some embodiments, a B domain deleted FVIII sequence of the present invention comprises any one of the deletions disclosed at col. 4, line 4 to col. 5, line 28 and examples 1-5 of U.S. Patent No. 6,316,226 (also in US 6,346,513). In some embodiments, a B domain deleted FVIII of the present invention has a deletion disclosed at col. 2, lines 26-51 and examples 5-8 of U.S. Patent No. 5,789,203 (also US 6,060,447, US 5,595,886, and US 6,228,620). In some embodiments, a B domain deleted FVIII has a deletion described in col. 1, lines 25 to col. 2, line 40 of US Patent No. 5,972,885; col. 6, lines 1-22 and example 1 of U.S. Patent no. 6,048,720; col. 2, lines 17-46 of U.S. Patent No. 5,543,502; col. 4, line 22 to col. 5, line 36 of U.S. Patent no. 5,171,844; col. 2, lines 55-68, figure 2, and example 1 of U.S. Patent No. 5,112,950; col. 2, line 2 to col. 19, line 21 and table 2 of U.S. Patent No. 4,868,112; col. 2, line 1 to col. 3, line 19, col. 3, line 40 to col. 4, line 67, col. 7, line 43 to col. 8, line 26, and col. 11, line 5 to col. 13, line 39 of U.S. Patent no. 7,041,635; or col. 4, lines 25-53, of U.S. Patent No. 6,458,563. In some embodiments, a B domain deleted FVIII has a deletion of most of the B domain, but still contains amino-terminal sequences of the B domain that are essential for *in vivo* proteolytic processing of the primary translation product into two polypeptide chain, as disclosed in WO 9 1/9122. In some embodiments, a B domain deleted FVIII is constructed with a deletion of amino acids 747-1638, *i.e.*, virtually a complete deletion of the B domain. Hoebein R.C., et al. J. Biol. Chem. 265 (13): 7318-7323 (1990). A B domain deleted FVIII can also contain a deletion of amino acids 771-1666 or amino acids 868-1562 of FVIII. Meulien P., et al. Protein Eng. 2(4): 301-6 (1988). Additional B domain deletions that are part of the invention include, *e.g.*, deletion of amino acids 982 through 1562 or 760 through 1639 (Toole et al., Proc. Natl. Acad. Sci. U.S.A. (1986) 83, 5939-5942)), 797 through 1562 (Eaton, et al. Biochemistry (1986) 25:8343-8347)), 741 through 1646 (Kaufman (PCT published application No. WO 87/04187)), 747-1560 (Sarver, et al., DNA (1987) 6:553-564)), 741 through 1648 (Pasek (PCT application No.88/00831)), 816 through 1598 or 741 through 1689 (Lagner (Behring Inst. Mitt. (1988) No 82:16-25, EP 295597)). Each of the foregoing deletions can be made in any FVIII sequence.

[0092] A number of functional FVIII molecules, including B-domain deletions, are disclosed in the following patents US 6,316,226 and US 6,346,513, both assigned to Baxter; US 7,041,635 assigned to In2Gen; US 5,789,203, US 6,060,447, US 5,595,886, and US 6,228,620 assigned to Chiron; US 5,972,885 and US 6,048,720 assigned to Biovitrum, US 5,543,502 and US 5,610,278 assigned to Novo Nordisk; US 5,171,844 assigned to Immuno Ag; US 5,112,950 assigned to Transgene S.A.; US 4,868,112 assigned to Genetics Institute.

Codon Optimization

[0093] In one embodiment, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide with FVIII activity, wherein the nucleic acid sequence has been codon optimized. In another embodiment, the starting nucleic acid sequence that encodes a polypeptide with FVIII activity and that is subject to codon optimization is SEQ ID NO:3. In some embodiments, the sequence that encodes a polypeptide with FVIII activity is codon optimized for human expression. In other embodiments, the sequence that encodes a polypeptide with FVIII activity is codon optimized for murine expression. SEQ ID NOs:1 and 2 are codon optimized versions of SEQ ID NO:3, optimized for human expression.

[0094] The term "codon-optimized" as it refers to genes or coding regions of nucleic acid molecules for transformation of various hosts, refers to the alteration of codons in the gene or coding regions of the nucleic acid molecules to reflect the typical codon usage of the host organism without altering the polypeptide encoded by the DNA. Such optimization includes replacing at least one, or more than one, or a significant number, of codons with one or more codons that are more frequently used in the genes of that organism.

[0095] Deviations in the nucleotide sequence that comprises the codons encoding the amino acids of any polypeptide chain allow for variations in the sequence coding for the gene. Since each codon consists of three nucleotides, and the nucleotides comprising DNA are restricted to four specific bases, there are 64 possible combinations of nucleotides, 61 of which encode amino acids (the remaining three codons encode signals ending translation). The "genetic code" which shows which codons encode which amino acids is reproduced herein as Table 1. As a result, many amino acids are designated by more than one codon. For example, the amino acids alanine and proline are coded for by four triplets, serine and arginine by six, whereas tryptophan and methionine are coded by just one triplet. This degeneracy allows for DNA base composition to vary over a wide range without altering the amino acid sequence of the proteins encoded by the DNA.

Table 1. The Standard Genetic Code

	T	C	A	G
T	TTT Phe (F)	TCT Ser (S)	TAT Tyr (Y)	TGT Cys (C)
	TTC "	TCC "	TAC "	TGC
	TTA Leu (L)	TCA "	TAA Stop	TGA Stop
	TTG "	TCG "	TAG Stop	TGG Trp (W)
C	CTT Leu (L)	CCT Pro (P)	CAT His (H)	CGT Arg (R)
	CTC "	CCC "	CAC "	CGC "
	CTA "	CCA "	CAA Gln (Q)	CGA "
	CTG "	CCG "	CAG "	CGG "
A	ATT Ile (I)			

	T	C	A	G
	ATC "	ACT Thr (T)	AAT Asn (N)	AGT Ser (S)
		ACC "	AAC "	AGC "
	ATA "	ACA "	AAA Lys (K)	AGA Arg (R)
	ATG Met (M)	ACG "	AAG "	AGG "
G	GTT Val (V)	GCT Ala (A)	GAT Asp (D)	GGT Gly (G)
	GTC "	GCC "	GAC "	GGC "
	GTA "	GCA "	GAA Glu (E)	GGA "
	GTG "	GCG "	GAG "	GGG "

[0096] Many organisms display a bias for use of particular codons to code for insertion of a particular amino acid in a growing peptide chain. Codon preference, or codon bias, differences in codon usage between organisms, is afforded by degeneracy of the genetic code, and is well documented among many organisms. Codon bias often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, *inter alia*, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis. Accordingly, genes can be tailored for optimal gene expression in a given organism based on codon optimization.

[0097] Given the large number of gene sequences available for a wide variety of animal, plant and microbial species, the relative frequencies of codon usage have been calculated. Codon usage tables are available, for example, at the "Codon Usage Database" available at www.kazusa.or.jp/codon/ (visited June 18, 2012). See Nakamura, Y., et al. Nucl. Acids Res. 28:292 (2000).

[0098] Randomly assigning codons at an optimized frequency to encode a given polypeptide sequence can be done manually by calculating codon frequencies for each amino acid, and then assigning the codons to the polypeptide sequence randomly. Additionally, various algorithms and computer software programs can be used to calculate an optimal sequence.

[0099] In one embodiment, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide with FVIII activity, wherein the nucleotide sequence is at least 90% identical to SEQ ID NO:1, and wherein the human codon adaptation index is increased relative to SEQ ID NO:3. For example, the nucleotide sequence that encodes a polypeptide with FVIII activity and that is at least 90% identical to SEQ ID NO:1 can have a human codon adaptation index that is at least about 0.75, at least about 0.76, at least about 0.77, at least about 0.78, at least about 0.79, at least about 0.80, at least about 0.81, at least about 0.82, at least about 0.83, at least about 0.84, at least about 0.85, at least about 0.86, at least about 0.87, at least about 0.88, at least about 0.89, or at least about 0.90.

[0100] In another embodiment, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide with Factor VIII (FVIII) activity, wherein the nucleotide sequence is at least 95% identical to SEQ ID NO:2, and wherein the human codon adaptation index is increased relative to SEQ ID NO:3. For example, the nucleotide sequence that encodes a polypeptide with FVIII activity and that is at least 85% identical to SEQ ID NO:2 can have a human codon adaptation index that is at least about 0.75, at least about 0.76, at least about 0.77, at least about 0.78, at least about 0.79, at least about 0.80, at least about 0.81, at least about 0.82, at least about 0.83, at least about 0.84, at least about 0.85, at least about 0.86, at least about 0.87, at least about 0.88, at least about 0.89, or at least about 0.90.

[0101] In other embodiments, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide with Factor VIII activity, wherein the nucleotide sequence is at least 90% identical to SEQ ID NO:1 and has one or more of the following characteristics: (1) the nucleotide sequence contains a higher percentage of G/C nucleotides compared to SEQ ID NO:3, (2) the nucleotide sequence contains fewer MARS/ARS sequences compared to SEQ ID NO:3, (3) the nucleotide sequence does not contain the splice site GGTGAT, (4) the nucleotide sequence contains fewer destabilizing elements, (5) the nucleotide sequence does not contain a poly-T sequence, (6) the nucleotide sequence does not contain a poly-A sequence, (7) the nucleotide sequence has a codon adaptation index that is increased relative to SEQ ID NO:3, or a combination of two or more such characteristics. In a particular embodiment, the nucleotide sequence contains all of the characteristics (1) to (6).

[0102] In other embodiments, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide with Factor VIII activity, wherein the nucleotide sequence is at least 95% identical to SEQ ID NO:2 and has one or more of the following characteristics: (1) the nucleotide sequence contains a higher percentage of G/C nucleotides compared to SEQ ID NO:3, (2) the nucleotide sequence contains fewer MARS/ARS sequences, (3) the nucleotide sequence does not contain the splice site GGTGAT, (4) the nucleotide sequence contains fewer destabilizing elements, (5) the nucleotide sequence does not contain a poly-T sequence, (6) the nucleotide sequence does not contain a poly-A sequence, (7) the nucleotide sequence has a codon adaptation index that is increased relative to SEQ ID NO:3, or a combination of two or more such characteristics. In a particular embodiment, the nucleotide sequence contains all of the characteristics (1) to (6).

G/C Content Optimization

[0103] In some embodiments, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide with FVIII activity, wherein the nucleotide sequence is at least 90% identical to SEQ ID NO:1, and wherein the nucleotide sequence contains a higher percentage of G/C nucleotides compared to the percentage of G/C nucleotides in SEQ ID NO:3. In other embodiments, the nucleotide sequence that encodes a polypeptide with FVIII activity and that is at least 90% identical to SEQ ID NO:1 has a G/C

content that is at least about 45%, at least about 46%, at least about 47%, at least about 48%, at least about 49%, at least about 50%, at least about 51%, at least about 52%, at least about 53%, at least about 54%, or at least about 55%.

[0104] In some embodiments, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide with FVIII activity, wherein the nucleotide sequence is at least 95% identical to SEQ ID NO:2, and wherein the nucleotide sequence contains a higher percentage of G/C nucleotides compared to the percentage of G/C nucleotides in SEQ ID NO:3. In other embodiments, the nucleotide sequence that encodes a polypeptide with FVIII activity and that is at least 95% identical to SEQ ID NO:2 has a G/C content that is at least about 45%, at least about 46%, at least about 47%, at least about 48%, at least about 49%, at least about 50%, at least about 51%, at least about 52%, at least about 53%, at least about 54%, or at least about 55%.

[0105] "G/C content" (or guanine-cytosine content), or "percentage of G/C nucleotides," refers to the percentage of nitrogenous bases in a DNA molecule that are either guanine or cytosine. G/C content can be calculated using the following formula:

$$\frac{G + C}{A + T + G + C} \times 100 \quad (III)$$

[0106] Human genes are highly heterogeneous in their G/C content, with some genes having a G/C content as low as 20%, and other genes having a G/C content as high as 95%. In general, G/C rich genes are more highly expressed. In fact, it has been demonstrated that increasing the G/C content of a gene can lead to increased expression of the gene, due mostly to an increase in transcription and higher steady state mRNA levels. See Kudla et al., PLoS Biol., 4(6): e180 (2006).

Matrix Attachment Region-Like Sequences

[0107] In some embodiments, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide with FVIII activity, wherein the nucleotide sequence is at least 90% identical to SEQ ID NO:1, and wherein the nucleotide sequence contains fewer MARS/ARS sequences relative to SEQ ID NO:3. In other embodiments, the nucleotide sequence that encodes a polypeptide with FVIII activity and that is at least 90% identical to SEQ ID NO:1 contains at most 6, at most 5, at most 4, at most 3, or at most 2 MARS/ARS sequences. In other embodiments, the nucleotide sequence that encodes a polypeptide with FVIII activity and that is at least 90% identical to SEQ ID NO:1 contains at most 1 MARS/ARS sequence. In yet other embodiments, the nucleotide sequence that encodes a polypeptide with FVIII activity and that is at least 90% identical to SEQ ID NO:1 does not contain a MARS/ARS sequence.

[0108] In some embodiments, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide with FVIII activity, wherein the nucleotide sequence is at least 95% identical to SEQ ID NO:2, and wherein the nucleotide sequence contains fewer MARS/ARS sequences relative to SEQ ID NO:3. In other embodiments, the nucleotide sequence that encodes a polypeptide with FVIII activity and that is at least 95% identical to SEQ ID NO:2 contains at most 6, at most 5, at most 4, at most 3, or at most 2 MARS/ARS sequences. In other embodiments, the nucleotide sequence that encodes a polypeptide with FVIII activity and that is at least 95% identical to SEQ ID NO:2 contains at most 1 MARS/ARS sequence. In yet other embodiments, the nucleotide sequence that encodes a polypeptide with FVIII activity and that is at least 95% identical to SEQ ID NO:2 does not contain a MARS/ARS sequence.

[0109] AT-rich elements in the human FVIII nucleotide sequence that share sequence similarity with *Saccharomyces cerevisiae* autonomously replicating sequences (ARSs) and nuclear-matrix attachment regions (MARs) have been identified. (Fallux et al., Mol. Cell. Biol. 16:4264-4272 (1996). One of these elements has been demonstrated to bind nuclear factors *in vitro* and to repress the expression of a chloramphenicol acetyltransferase (CAT) reporter gene. *Id.* It has been hypothesized that these sequences can contribute to the transcriptional repression of the human FVIII gene. Thus, in one embodiment, all MAR/ARS sequences are abolished in the FVIII gene of the present invention. There are four MAR/ARS ATATTT sequences (SEQ ID NO:5) and three MAR/ARS AAATAT sequences (SEQ ID NO:6) in the parental FVIII sequence (SEQ ID NO:3). All of these sites were mutated to destroy the MAR/ARS sequences in the optimized FVIII sequences (SEQ ID NO:1 and SEQ ID NO:2). The location of each of these elements, and the sequence of the corresponding nucleotides in the optimized sequences are shown in Table 2, below.

Table 2: Summary of Changes to Repressive Elements

Location of Element	Starting BDD FVIII Sequence (SEQ ID NO: 3)	Optimized BDD FVIII Sequence (SEQ ID NO:1)	Optimized BDD FVIII Sequence (SEQ ID NO:2)
Destabilizing Sequences			
639 ATTTA		GTTCA	GTTCA
1338	ATTTA	GTTCA	GTTCA
1449	ATTTA	TTTCA	CTTCA
1590	TAAAT	CAAGT	CAAGT
1623	TAAAT	TAAGA	CAAGA
2410	ATTTA	ATCTA	ATCTA
2586	ATTTA	GTTTA	GTTCA
2630	TAAAT	TGAAC	TGAAC
3884	ATTTA	ATCTG	ACCTG
3887	TAAAT	TGAAC	TGAAC

Potential Promoter Binding Sites			
641	TTATA	TCATT	TCATC
1275	TATAA	TACAA	TACAA
1276	TTATA	CTACA	CTACA
1445	TTATA	TCATT	TCATC
1474	TATAA	TACAA	TACAA
1588	TATAA	TACAA	TACAA
2614	TTATA	CTGTA	CTGTA
2661	TATAA	TATCA	CATTA
3286	TATAA	TACAA	TACAA
3840	TTATA	TTATT	CTACA
Matrix Attachment-Like Sequences (MARS/ARS)			
1287 ATATTT		GTATCT	GTACCT
1447	ATATTT	ATTTTC	ATCTTC
1577	AAATAT	AAATCT	AGATCT
1585	AAATAT	AAGTAC	AAGTAC
2231	ATATTT	ACATCA	ACATCA
3054	AAATAT	AAACAT	GAACAT
3788	ATATTT	ACATTT	ACATCT
AU Rich Sequence Elements (AREs)			
2468	ATTTTATT	ACTTTATT	ACTTCATT
3790	ATTTTTAA	ATTTTCAA	ATCTTCAA
Poly A/Poly T Sequences			
3273	AAAAAAA	GAAGAAA	GAAGAAA
4195	IIIIII	TTCTTT	TTCTTT
Splice Sites			
2203	GGTGAT	GGGGAC	GGCGAC

Destabilizing Sequences

[0110] In some embodiments, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide with FVIII activity, wherein the nucleotide sequence is at least 90% identical to SEQ ID NO:1, and wherein the nucleotide sequence contains fewer destabilizing elements relative to SEQ ID NO:3. In other embodiments, the nucleotide sequence that encodes a polypeptide with FVIII activity and that

is at least 90% identical to SEQ ID NO:1 contains at most 9, at most 8, at most 7, at most 6, or at most 5 destabilizing elements. In other embodiments, the nucleotide sequence that encodes a polypeptide with FVIII activity and that is at least 90% identical to SEQ ID NO:1 contains at most 4, at most 3, at most 2, or at most 1 destabilizing elements. In yet other embodiments, the nucleotide sequence that encodes a polypeptide with FVIII activity and that is at least 90% identical to SEQ ID NO:1 does not contain a destabilizing element.

[0111] In some embodiments, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide with FVIII activity, wherein the nucleotide sequence is at least 95% identical to SEQ ID NO:2, and wherein the nucleotide sequence contains fewer destabilizing elements relative to SEQ ID NO:3. In other embodiments, the nucleotide sequence that encodes a polypeptide with FVIII activity and that is at least 95% identical to SEQ ID NO:2 contains at most 9, at most 8, at most 7, at most 6, or at most 5 destabilizing elements. In other embodiments, the nucleotide sequence that encodes a polypeptide with FVIII activity and that is at least 95% identical to SEQ ID NO:2 contains at most 4, at most 3, at most 2, or at most 1 destabilizing elements. In yet other embodiments, the nucleotide sequence that encodes a polypeptide with FVIII activity and that is at least 95% identical to SEQ ID NO:2 does not contain a destabilizing element.

[0112] There are ten destabilizing elements in the parental FVIII sequence (SEQ ID NO:3); six ATTTA sequences (SEQ ID NO:8) and four TAAAT sequences (SEQ ID NO:9). In one embodiment, sequences of these sites were mutated to destroy the destabilizing elements in optimized FVIII SEQ ID NO:1 and SEQ ID NO:2. The location of each of these elements, and the sequence of the corresponding nucleotides in the optimized sequences are shown in Table 2.

Potential Promoter Binding Sites

[0113] In some embodiments, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide with FVIII activity, wherein the nucleotide sequence is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:1, and wherein the nucleotide sequence contains fewer potential promoter binding sites relative to SEQ ID NO:3. In other embodiments, the nucleotide sequence that encodes a polypeptide with FVIII activity and that is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:1 contains at most 9, at most 8, at most 7, at most 6, or at most 5 potential promoter binding sites. In other embodiments, the nucleotide sequence that encodes a polypeptide with FVIII activity and that is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:1 contains at most 4, at most 3, at most 2, or at most 1 potential promoter binding sites. In yet other embodiments, the nucleotide sequence that encodes a polypeptide with FVIII activity and that is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:1 does not contain a potential promoter binding site.

[0114] In some embodiments, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide with FVIII activity, wherein the nucleotide sequence is at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:2, and wherein the nucleotide sequence contains fewer potential promoter binding sites relative to SEQ ID NO:3. In other embodiments, the nucleotide sequence that encodes a polypeptide with FVIII activity and that is at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:2 contains at most 9, at most 8, at most 7, at most 6, or at most 5 potential promoter binding sites. In other embodiments, the nucleotide sequence that encodes a polypeptide with FVIII activity and that is at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:2 contains at most 4, at most 3, at most 2, or at most 1 potential promoter binding sites. In yet other embodiments, the nucleotide sequence that encodes a polypeptide with FVIII activity and that is at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:2 does not contain a potential promoter binding site.

[0115] TATA boxes are regulatory sequences often found in the promoter regions of eukaryotes. They serve as the binding site of TATA binding protein (TBP), a general transcription factor. TATA boxes usually comprise the sequence TATAAA (SEQ ID NO:12) or a close variant. TATA boxes within a coding sequence, however, can inhibit the translation of full-length protein. There are ten potential promoter binding sequences in the wild type BDD FVIII sequence (SEQ ID NO:3); five TATAA sequences (SEQ ID NO:12) and five TTATA sequences (SEQ ID NO:13). In one embodiment, all promoter binding sites are abolished in the FVIII genes of the present invention. The location of each potential promoter binding site and the sequence of the corresponding nucleotides in the optimized sequences are shown in Table 2.

Other Cis Acting Negative Regulatory Elements

[0116] In addition to the MAR/ARS sequences, destabilizing elements, and potential promoter sites described above, several additional potentially inhibitory sequences can be identified in the wild type BDD FVIII sequence (SEQ ID NO:3). Two AU rich sequence elements (AREs) can be identified (SEQ ID NOs:14 and 15), along with a poly-A site (SEQ ID NO:11), a poly-T site (SEQ ID NO:10), and a splice site (SEQ ID NO:7) in the wild type BDD FVIII sequence. One or more of these elements can be removed from the optimized FVIII sequences. The location of each of these sites and the sequence of the corresponding nucleotides in the optimized sequences are shown in Table 2.

[0117] In certain embodiments, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide with FVIII activity, wherein the nucleotide sequence is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:1, wherein the nucleotide sequence does not contain one or more cis-acting negative regulatory elements, for example, a splice site, a poly-T sequence, a poly-A sequence, an ARE sequence, or any combinations thereof.

[0118] In certain embodiments, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide with FVIII activity, wherein the nucleotide sequence is at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:2, wherein the nucleotide sequence does not contain one or more cis-acting negative regulatory elements, for example, a splice site, a poly-T sequence, a poly-A sequence, an ARE sequence, or any combinations thereof.

[0119] In some embodiments, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide with FVIII activity, wherein the nucleotide sequence is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:1, and wherein the nucleotide sequence does not contain the splice site GGTGAT (SEQ ID NO:7). In some embodiments, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide with FVIII activity, wherein the nucleotide sequence is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:1, and wherein the nucleotide sequence does not contain a poly-T sequence (SEQ ID NO:10). In some embodiments, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide with FVIII activity, wherein the nucleotide sequence is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:1, and wherein the nucleotide sequence does not contain a poly-A sequence (SEQ ID NO:11). In some embodiments, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide with FVIII activity, wherein the nucleotide sequence is at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:1, and wherein the nucleotide sequence does not contain an ARE element (SEQ ID NO:14 or SEQ ID NO:15).

[0120] In some embodiments, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide with FVIII activity, wherein the nucleotide sequence is at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:2, and wherein the nucleotide sequence does not contain the splice site GGTGAT (SEQ ID NO:7). In some embodiments, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide with FVIII activity, wherein the nucleotide sequence is at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:2, and wherein the nucleotide sequence does not contain a poly-T sequence (SEQ ID NO:10). In some embodiments, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide with FVIII activity, wherein the nucleotide sequence is at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identical to SEQ ID NO:2, and wherein the nucleotide sequence does not contain a poly-A sequence (SEQ ID NO:11). In some embodiments, the present invention provides an isolated nucleic acid molecule comprising a nucleotide sequence that encodes a polypeptide with FVIII activity, wherein the nucleotide sequence is at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or

100% identical to SEQ ID NO:2, and wherein the nucleotide sequence does not contain an ARE element (SEQ ID NO:14 or SEQ ID NO:15).

[0121] In other embodiments, an optimized FVIII sequence of the invention does not comprise one or more of antiviral motifs, stem-loop structures, and repeat sequences.

[0122] In still other embodiments, the nucleotides surrounding the transcription start site are changed to a kozak consensus sequence (GCCGCCACCATGC, wherein the underlined nucleotides are the start codon; SEQ ID NO:16). In other embodiments, restriction sites can be added or removed to facilitate the cloning process.

Heterologous Nucleotide Sequences

[0123] In some embodiments, the isolated nucleic acid molecules of the invention further comprise a heterologous nucleotide sequence. In some embodiments, the isolated nucleic acid molecules of the invention further comprise at least one heterologous nucleotide sequence. The heterologous nucleotide sequence can be linked with the optimized BDD-FVIII nucleotide sequences of the invention at the 5' end, at the 3' end, or inserted into the middle of the optimized BDD-FVIII nucleotide sequence. Thus, in some embodiments, the heterologous amino acid sequence encoded by the heterologous nucleotide sequence is linked to the N-terminus or the C-terminus of the FVIII amino acid sequence encoded by the nucleotide sequence or inserted between two amino acids in the FVIII amino acid sequence. In other embodiments, the isolated nucleic acid molecules of the invention further comprise two, three, four, five, six, seven, or eight heterologous nucleotide sequences. In some embodiments, all the heterologous nucleotide sequences are identical. In some embodiments, at least one heterologous nucleotide sequence is different from the other heterologous nucleotide sequences. In some embodiments, the invention can comprise two, three, four, five, six, or more than seven heterologous nucleotide sequences in tandem.

[0124] In some embodiments, the heterologous nucleotide sequence encodes an amino acid sequence. In some embodiments, the amino acid sequence encoded by the heterologous nucleotide sequence is a heterologous moiety that can increase the half-life (a "half-life extender") of a FVIII molecule.

[0125] In some embodiments, the heterologous moiety is a peptide or a polypeptide with either unstructured or structured characteristics that are associated with the prolongation of *in vivo* half-life when incorporated in a protein of the invention. Non-limiting examples include albumin, albumin fragments, Fc fragments of immunoglobulins, the C-terminal peptide (CTP) of the β subunit of human chorionic gonadotropin, a HAP sequence, an XTEN sequence, a transferrin or a fragment thereof, a PAS polypeptide, polyglycine linkers, polyserine linkers, albumin-binding moieties, or any fragments, derivatives, variants, or combinations of these polypeptides. In some aspects, a heterologous moiety includes von Willebrand factor or a fragment thereof. In other related aspects a heterologous moiety can include an attachment

site (e.g., a cysteine amino acid) for a non-polypeptide moiety such as polyethylene glycol (PEG), hydroxyethyl starch (HES), polysialic acid, or any derivatives, variants, or combinations of these elements. In some aspects, a heterologous moiety comprises a cysteine amino acid that functions as an attachment site for a non-polypeptide moiety such as polyethylene glycol (PEG), hydroxyethyl starch (HES), polysialic acid, or any derivatives, variants, or combinations of these elements.

[0126] In one specific embodiment, a first heterologous nucleotide sequence encodes a first heterologous moiety that is a half-life extending molecule which is known in the art, and a second heterologous nucleotide sequence encodes a second heterologous moiety that can also be a half-life extending molecule which is known in the art. In certain embodiments, the first heterologous moiety (e.g., a first Fc moiety) and the second heterologous moiety (e.g., a second Fc moiety) are associated with each other to form a dimer. In one embodiment, the second heterologous moiety is a second Fc moiety, wherein the second Fc moiety is linked to or associated with the first heterologous moiety, e.g., the first Fc moiety. For example, the second heterologous moiety (e.g., the second Fc moiety) can be linked to the first heterologous moiety (e.g., the first Fc moiety) by a linker or associated with the first heterologous moiety by a covalent or non-covalent bond.

[0127] In some embodiments, the heterologous moiety is a polypeptide comprising, consisting essentially of, or consisting of at least about 10, at least about 100, at least about 200, at least about 300, at least about 400, at least about 500, at least about 600, at least about 700, at least about 800, at least about 900, at least about 1000, at least about 1100, at least about 1200, at least about 1300, at least about 1400, at least about 1500, at least about 1600, at least about 1700, at least about 1800, at least about 1900, at least about 2000, at least about 2500, at least about 3000, or at least about 4000 amino acids. In other embodiments, the heterologous moiety is a polypeptide comprising, consisting essentially of, or consisting of about 100 to about 200 amino acids, about 200 to about 300 amino acids, about 300 to about 400 amino acids, about 400 to about 500 amino acids, about 500 to about 600 amino acids, about 600 to about 700 amino acids, about 700 to about 800 amino acids, about 800 to about 900 amino acids, or about 900 to about 1000 amino acids.

[0128] In certain embodiments, a heterologous moiety improves one or more pharmacokinetic properties of the FVIII protein without significantly affecting its biological activity or function.

[0129] In certain embodiments, a heterologous moiety increases the *in vivo* and/or *in vitro* half-life of the FVIII protein. In other embodiments, a heterologous moiety facilitates visualization or localization of the FVIII protein or a fragment thereof (e.g., a fragment comprising a heterologous moiety after proteolytic cleavage of the FVIII protein). Visualization and/or location of the FVIII protein or a fragment thereof can be *in vivo*, *in vitro*, *ex vivo*, or combinations thereof.

[0130] In other embodiments, a heterologous moiety increases stability of the FVIII protein or a fragment thereof (e.g., a fragment comprising a heterologous moiety after proteolytic cleavage

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

[0131] In certain aspects, a FVIII protein comprises at least one half-life extender, *i.e.*, a heterologous moiety which increases the *in vivo* half-life of the FVIII protein with respect to the *in vivo* half-life of the corresponding FVIII protein lacking such heterologous moiety. *In vivo* half-life of a FVIII protein can be determined by any methods known to those of skill in the art, e.g., activity assays (chromogenic assay or one stage clotting aPTT assay), ELISA, ROTEM™, etc.

[0132] In some embodiments, the presence of one or more half-life extenders results in the half-life of the FVIII protein to be increased compared to the half-life of the corresponding protein lacking such one or more half-life extenders. The half-life of the FVIII protein comprising a half-life extender is at least about 1.5 times, at least about 2 times, at least about 2.5 times, at least about 3 times, at least about 4 times, at least about 5 times, at least about 6 times, at least about 7 times, at least about 8 times, at least about 9 times, at least about 10 times, at least about 11 times, or at least about 12 times longer than the *in vivo* half-life of the corresponding FVIII protein lacking such half-life extender.

[0133] In one embodiment, the half-life of the FVIII protein comprising a half-life extender is about 1.5-fold to about 20-fold, about 1.5 fold to about 15 fold, or about 1.5 fold to about 10 fold longer than the *in vivo* half-life of the corresponding protein lacking such half-life extender. In another embodiment, the half-life of FVIII protein comprising a half-life extender is extended about 2-fold to about 10-fold, about 2-fold to about 9-fold, about 2-fold to about 8-fold, about 2-fold to about 7-fold, about 2-fold to about 6-fold, about 2-fold to about 5-fold, about 2-fold to about 4-fold, about 2-fold to about 3-fold, about 2.5-fold to about 10-fold, about 2.5-fold to about 9-fold, about 2.5-fold to about 8-fold, about 2.5-fold to about 7-fold, about 2.5-fold to about 6-fold, about 2.5-fold to about 5-fold, about 2.5-fold to about 4-fold, about 2.5-fold to about 3-fold, about 3-fold to about 10-fold, about 3-fold to about 9-fold, about 3-fold to about 8-

fold, about 3-fold to about 7-fold, about 3-fold to about 6-fold, about 3-fold to about 5-fold, about 3-fold to about 4-fold, about 4-fold to about 6 fold, about 5-fold to about 7-fold, or about 6-fold to about 8 fold as compared to the *in vivo* half-life of the corresponding protein lacking such half-life extender.

[0134] In other embodiments, the half-life of the FVIII protein comprising a half-life extender is at least about 17 hours, at least about 18 hours, at least about 19 hours, at least about 20 hours, at least about 21 hours, at least about 22 hours, at least about 23 hours, at least about 24 hours, at least about 25 hours, at least about 26 hours, at least about 27 hours, at least about 28 hours, at least about 29 hours, at least about 30 hours, at least about 31 hours, at least about 32 hours, at least about 33 hours, at least about 34 hours, at least about 35 hours, at least about 36 hours, at least about 48 hours, at least about 60 hours, at least about 72 hours, at least about 84 hours, at least about 96 hours, or at least about 108 hours.

[0135] In still other embodiments, the half-life of the FVIII protein comprising a half-life extender is about 15 hours to about two weeks, about 16 hours to about one week, about 17 hours to about one week, about 18 hours to about one week, about 19 hours to about one week, about 20 hours to about one week, about 21 hours to about one week, about 22 hours to about one week, about 23 hours to about one week, about 24 hours to about one week, about 36 hours to about one week, about 48 hours to about one week, about 60 hours to about one week, about 24 hours to about six days, about 24 hours to about five days, about 24 hours to about four days, about 24 hours to about three days, or about 24 hours to about two days.

[0136] In some embodiments, the average half-life per subject of the FVIII protein comprising a half-life extender is about 15 hours, about 16 hours, about 17 hours, about 18 hours, about 19 hours, about 20 hours, about 21 hours, about 22 hours, about 23 hours, about 24 hours (1 day), about 25 hours, about 26 hours, about 27 hours, about 28 hours, about 29 hours, about 30 hours, about 31 hours, about 32 hours, about 33 hours, about 34 hours, about 35 hours, about 36 hours, about 40 hours, about 44 hours, about 48 hours (2 days), about 54 hours, about 60 hours, about 72 hours (3 days), about 84 hours, about 96 hours (4 days), about 108 hours, about 120 hours (5 days), about six days, about seven days (one week), about eight days, about nine days, about 10 days, about 11 days, about 12 days, about 13 days, or about 14 days.

1. An immunoglobulin Constant Region or a Portion Thereof

[0137] In another aspect, a heterologous moiety comprises one or more immunoglobulin constant regions or portions thereof (e.g., an Fc region). In one embodiment, an isolated nucleic acid molecule of the invention further comprises a heterologous nucleic acid sequence that encodes an immunoglobulin constant region or a portion thereof. In some embodiments, the immunoglobulin constant region or portion thereof is an Fc region.

[0138] An immunoglobulin constant region is comprised of domains denoted CH (constant heavy) domains (CH1, CH2, etc.). Depending on the isotype, (*i.e.* IgG, IgM, IgA IgD, or IgE), the constant region can be comprised of three or four CH domains. Some isotypes (*e.g.* IgG) constant regions also contain a hinge region. See Janeway et al. 2001, Immunobiology, Garland Publishing, N.Y., N.Y.

[0139] An immunoglobulin constant region or a portion thereof for producing the FVIII protein can be obtained from a number of different sources. In one embodiment, an immunoglobulin constant region or a portion thereof is derived from a human immunoglobulin. It is understood, however, that the immunoglobulin constant region or a portion thereof can be derived from an immunoglobulin of another mammalian species, including for example, a rodent (*e.g.* a mouse, rat, rabbit, guinea pig) or non-human primate (*e.g.* chimpanzee, macaque) species. Moreover, the immunoglobulin constant region or a portion thereof can be derived from any immunoglobulin class, including IgM, IgG, IgD, IgA and IgE, and any immunoglobulin isotype, including IgG1, IgG2, IgG3 and IgG4. In one embodiment, the human isotype IgG1 is used.

[0140] A variety of the immunoglobulin constant region gene sequences (*e.g.* human constant region gene sequences) are available in the form of publicly accessible deposits. Constant region domains sequence can be selected having a particular effector function (or lacking a particular effector function) or with a particular modification to reduce immunogenicity. Many sequences of antibodies and antibody-encoding genes have been published and suitable Ig constant region sequences (*e.g.* hinge, CH2, and/or CH3 sequences, or portions thereof) can be derived from these sequences using art recognized techniques. The genetic material obtained using any of the foregoing methods can then be altered or synthesized to obtain polypeptides as disclosed herein. It will further be appreciated that the scope of this invention encompasses alleles, variants and mutations of constant region DNA sequences.

[0141] The sequences of the immunoglobulin constant region or a portion thereof can be cloned, *e.g.*, using the polymerase chain reaction and primers which are selected to amplify the domain of interest. To clone a sequence of the immunoglobulin constant region or a portion thereof from an antibody, mRNA can be isolated from hybridoma, spleen, or lymph cells, reverse transcribed into DNA, and antibody genes amplified by PCR. PCR amplification methods are described in detail in U.S. Pat. Nos. 4,683,195; 4,683,202; 4,800,159; 4,965,188; and in, *e.g.*, "PCR Protocols: A Guide to Methods and Applications" Innis et al. eds., Academic Press, San Diego, CA (1990); Ho et al. 1989. *Gene* 77:51; Horton et al. 1993. *Methods Enzymol.* 217:270). PCR can be initiated by consensus constant region primers or by more specific primers based on the published heavy and light chain DNA and amino acid sequences. PCR also can be used to isolate DNA clones encoding the antibody light and heavy chains. In this case the libraries can be screened by consensus primers or larger homologous probes, such as mouse constant region probes. Numerous primer sets suitable for amplification of antibody genes are known in the art (*e.g.*, 5' primers based on the N-terminal sequence of purified antibodies (Benhar and Pastan. 1994. *Protein Engineering* 7:1509); rapid amplification of cDNA ends (Ruberti, F. et al. 1994. *J. Immunol. Methods* 173:33); antibody leader sequences (Larrick et al. 1989 *Biochem. Biophys. Res. Commun.* 160:1250). The cloning of

antibody sequences is further described in Newman et al., U.S. Pat. No. 5,658,570, filed January 25, 1995.

[0142] An immunoglobulin constant region used herein can include all domains and the hinge region or portions thereof. In one embodiment, the immunoglobulin constant region or a portion thereof comprises CH2 domain, CH3 domain, and a hinge region, *i.e.*, an Fc region or an FcRn binding partner.

[0143] As used herein, the term "Fc region" is defined as the portion of a polypeptide which corresponds to the Fc region of native Ig, *i.e.*, as formed by the dimeric association of the respective Fc domains of its two heavy chains. A native Fc region forms a homodimer with another Fc region. In contrast, the term "genetically-fused Fc region" or "single-chain Fc region" (scFc region), as used herein, refers to a synthetic dimeric Fc region comprised of Fc domains genetically linked within a single polypeptide chain (*i.e.*, encoded in a single contiguous genetic sequence). See International Publication No. WO 201 2/006635.

[0144] In one embodiment, the "Fc region" refers to the portion of a single Ig heavy chain beginning in the hinge region just upstream of the papain cleavage site (*i.e.* residue 216 in IgG, taking the first residue of heavy chain constant region to be 114) and ending at the C-terminus of the antibody. Accordingly, a complete Fc region comprises at least a hinge domain, a CH2 domain, and a CH3 domain.

[0145] An immunoglobulin constant region or a portion thereof can be an FcRn binding partner. FcRn is active in adult epithelial tissues and expressed in the lumen of the intestines, pulmonary airways, nasal surfaces, vaginal surfaces, colon and rectal surfaces (U.S. Pat. No. 6,485,726). An FcRn binding partner is a portion of an immunoglobulin that binds to FcRn.

[0146] The FcRn receptor has been isolated from several mammalian species including humans. The sequences of the human FcRn, monkey FcRn, rat FcRn, and mouse FcRn are known (Story et al. 1994, J. Exp. Med. 180:2377). The FcRn receptor binds IgG (but not other immunoglobulin classes such as IgA, IgM, IgD, and IgE) at relatively low pH, actively transports the IgG transcellularly in a luminal to serosal direction, and then releases the IgG at relatively higher pH found in the interstitial fluids. It is expressed in adult epithelial tissue (U.S. Pat. Nos. 6,485,726, 6,030,613, 6,086,875; WO 03/077834; US2003-0235536A1) including lung and intestinal epithelium (Israel et al. 1997, Immunology 92:69) renal proximal tubular epithelium (Kobayashi et al. 2002, Am. J. Physiol. Renal Physiol. 282:F358) as well as nasal epithelium, vaginal surfaces, and biliary tree surfaces.

[0147] FcRn binding partners useful in the present invention encompass molecules that can be specifically bound by the FcRn receptor including whole IgG, the Fc fragment of IgG, and other fragments that include the complete binding region of the FcRn receptor. The region of the Fc portion of IgG that binds to the FcRn receptor has been described based on X-ray crystallography (Burmeister et al. 1994, Nature 372:379). The major contact area of the Fc with the FcRn is near the junction of the CH2 and CH3 domains. Fc-FcRn contacts are all

within a single Ig heavy chain. The FcRn binding partners include whole IgG, the Fc fragment of IgG, and other fragments of IgG that include the complete binding region of FcRn. The major contact sites include amino acid residues 248, 250-257, 272, 285, 288, 290-291, 308-311, and 314 of the CH2 domain and amino acid residues 385-387, 428, and 433-436 of the CH3 domain. References made to amino acid numbering of immunoglobulins or immunoglobulin fragments, or regions, are all based on Kabat et al. 1991, Sequences of Proteins of Immunological Interest, U.S. Department of Public Health, Bethesda, Md.

[0148] Fc regions or FcRn binding partners bound to FcRn can be effectively shuttled across epithelial barriers by FcRn, thus providing a non-invasive means to systemically administer a desired therapeutic molecule. Additionally, fusion proteins comprising an Fc region or an FcRn binding partner are endocytosed by cells expressing the FcRn. But instead of being marked for degradation, these fusion proteins are recycled out into circulation again, thus increasing the *in vivo* half-life of these proteins. In certain embodiments, the portions of immunoglobulin constant regions are an Fc region or an FcRn binding partner that typically associates, via disulfide bonds and other non-specific interactions, with another Fc region or another FcRn binding partner to form dimers and higher order multimers.

[0149] Two FcRn receptors can bind a single Fc molecule. Crystallographic data suggest that each FcRn molecule binds a single polypeptide of the Fc homodimer. In one embodiment, linking the FcRn binding partner, e.g., an Fc fragment of an IgG, to a biologically active molecule provides a means of delivering the biologically active molecule orally, buccally, sublingually, rectally, vaginally, as an aerosol administered nasally or via a pulmonary route, or via an ocular route. In another embodiment, the FVIII protein can be administered invasively, e.g., subcutaneously, intravenously.

[0150] An FcRn binding partner region is a molecule or portion thereof that can be specifically bound by the FcRn receptor with consequent active transport by the FcRn receptor of the Fc region. Specifically bound refers to two molecules forming a complex that is relatively stable under physiologic conditions. Specific binding is characterized by a high affinity and a low to moderate capacity as distinguished from nonspecific binding which usually has a low affinity with a moderate to high capacity. Typically, binding is considered specific when the affinity constant K_A is higher than 10^6 M^{-1} , or higher than 10^8 M^{-1} . If necessary, non-specific binding can be reduced without substantially affecting specific binding by varying the binding conditions. The appropriate binding conditions such as concentration of the molecules, ionic strength of the solution, temperature, time allowed for binding, concentration of a blocking agent (e.g., serum albumin, milk casein), etc., can be optimized by a skilled artisan using routine techniques.

[0151] In certain embodiments, a FVIII protein of the invention comprises one or more truncated Fc regions that are nonetheless sufficient to confer Fc receptor (FcR) binding properties to the Fc region. For example, the portion of an Fc region that binds to FcRn (i.e., the FcRn binding portion) comprises from about amino acids 282-438 of IgG1, EU numbering (with the primary contact sites being amino acids 248, 250-257, 272, 285, 288, 290-291, 308-

311, and 314 of the CH2 domain and amino acid residues 385-387, 428, and 433-436 of the CH3 domain. Thus, an Fc region of the invention can comprise or consist of an FcRn binding portion. FcRn binding portions can be derived from heavy chains of any isotype, including IgG1, IgG2, IgG3 and IgG4. In one embodiment, an FcRn binding portion from an antibody of the human isotype IgG1 is used. In another embodiment, an FcRn binding portion from an antibody of the human isotype IgG4 is used.

[0152] The Fc region can be obtained from a number of different sources. In one embodiment, an Fc region of the polypeptide is derived from a human immunoglobulin. It is understood, however, that an Fc moiety can be derived from an immunoglobulin of another mammalian species, including for example, a rodent (e.g. a mouse, rat, rabbit, guinea pig) or non-human primate (e.g. chimpanzee, macaque) species. Moreover, the polypeptide of the Fc domains or portions thereof can be derived from any immunoglobulin class, including IgM, IgG, IgD, IgA and IgE, and any immunoglobulin isotype, including IgG1, IgG2, IgG3 and IgG4. In another embodiment, the human isotype IgG1 is used.

[0153] In certain embodiments, the Fc variant confers a change in at least one effector function imparted by an Fc moiety comprising said wild-type Fc domain (e.g., an improvement or reduction in the ability of the Fc region to bind to Fc receptors (e.g. FcγRI, FcγRII, or FcγRIII) or complement proteins (e.g. C1q), or to trigger antibody-dependent cytotoxicity (ADCC), phagocytosis, or complement-dependent cytotoxicity (CDCC)). In other embodiments, the Fc variant provides an engineered cysteine residue.

[0154] The Fc region of the invention can employ art-recognized Fc variants which are known to impart a change (e.g., an enhancement or reduction) in effector function and/or FcR or FcRn binding. Specifically, an Fc region of the invention can include, for example, a change (e.g., a substitution) at one or more of the amino acid positions disclosed in International PCT Publications WO88/07089A1, WO96/14339A1, WO98/05787A1, WO98/23289A1, WO99/51642A1, WO99/58572A1, WO00/09560A2, WO00/32767A1, WO00/42072A2, WO02/44215A2, WO02/060919A2, WO03/074569A2, WO04/016750A2, WO04/029207A2, WO04/035752A2, WO04/063351A2, WO04/074455A2, WO04/099249A2, WO05/040217A2, WO04/044859, WO05/070963A1, WO05/077981A2, WO05/092925A2, WO05/123780A2, WO06/019447A1, WO06/047350A2, and WO06/085967A2; US Patent Publication Nos. US2007/0231329, US2007/0231329, US2007/0237765, US2007/0237766, US2007/0237767, US2007/0243188, US2007/0248603, US2007/0286859, US2008/0057056 ; or US Patents 5,648,260; 5,739,277; 5,834,250; 5,869,046; 6,096,871; 6,121,022; 6,194,551; 6,242,195; 6,277,375; 6,528,624; 6,538,124; 6,737,056; 6,821,505; 6,998,253; 7,083,784; 7,404,956, and 7,317,091, . In one embodiment, the specific change (e.g., the specific substitution of one or more amino acids disclosed in the art) can be made at one or more of the disclosed amino acid positions. In another embodiment, a different change at one or more of the disclosed amino acid positions (e.g., the different substitution of one or more amino acid position disclosed in the art) can be made.

[0155] The Fc region or FcRn binding partner of IgG can be modified according to well

recognized procedures such as site directed mutagenesis and the like to yield modified IgG or Fc fragments or portions thereof that will be bound by FcRn. Such modifications include modifications remote from the FcRn contact sites as well as modifications within the contact sites that preserve or even enhance binding to the FcRn. For example, the following single amino acid residues in human IgG1 Fc (Fc γ 1) can be substituted without significant loss of Fc binding affinity for FcRn: P238A, S239A, K246A, K248A, D249A, M252A, T256A, E258A, T260A, D265A, S267A, H268A, E269A, D270A, E272A, L274A, N276A, Y278A, D280A, V282A, E283A, H285A, N286A, T289A, K290A, R292A, E293A, E294A, Q295A, Y296F, N297A, S298A, Y300F, R301A, V303A, V305A, T307A, L309A, Q311A, D312A, N315A, K317A, E318A, K320A, K322A, S324A, K326A, A327Q, P329A, A330Q, P331A, E333A, K334A, T335A, S337A, K338A, K340A, Q342A, R344A, E345A, Q347A, R355A, E356A, M358A, T359A, K360A, N361A, Q362A, Y373A, S375A, D376A, A378Q, E380A, E382A, S383A, N384A, Q386A, E388A, N389A, N390A, Y391F, K392A, L398A, S400A, D401A, D413A, K414A, R416A, Q418A, Q419A, N421A, V422A, S424A, E430A, N434A, T437A, Q438A, K439A, S440A, S444A, and K447A, where for example P238A represents wild type proline substituted by alanine at position number 238. As an example, a specific embodiment incorporates the N297A mutation, removing a highly conserved N-glycosylation site. In addition to alanine other amino acids can be substituted for the wild type amino acids at the positions specified above. Mutations can be introduced singly into Fc giving rise to more than one hundred Fc regions distinct from the native Fc. Additionally, combinations of two, three, or more of these individual mutations can be introduced together, giving rise to hundreds more Fc regions.

[0156] Certain of the above mutations can confer new functionality upon the Fc region or FcRn binding partner. For example, one embodiment incorporates N297A, removing a highly conserved N-glycosylation site. The effect of this mutation is to reduce immunogenicity, thereby enhancing circulating half-life of the Fc region, and to render the Fc region incapable of binding to Fc γ RI, Fc γ RIIA, Fc γ RIIB, and Fc γ RIIIA, without compromising affinity for FcRn (Routledge et al. 1995, Transplantation 60:847; Friend et al. 1999, Transplantation 68:1632; Shields et al. 1995, J. Biol. Chem. 276:6591). As a further example of new functionality arising from mutations described above affinity for FcRn can be increased beyond that of wild type in some instances. This increased affinity can reflect an increased "on" rate, a decreased "off" rate or both an increased "on" rate and a decreased "off" rate. Examples of mutations believed to impart an increased affinity for FcRn include, but not limited to, T256A, T307A, E380A, and N434A (Shields et al. 2001, J. Biol. Chem. 276:6591).

[0157] Additionally, at least three human Fc gamma receptors appear to recognize a binding site on IgG within the lower hinge region, generally amino acids 234-237. Therefore, another example of new functionality and potential decreased immunogenicity can arise from mutations of this region, as for example by replacing amino acids 233-236 of human IgG1 "ELLG" (SEQ ID NO:29) to the corresponding sequence from IgG2 "PVA" (with one amino acid deletion). It has been shown that Fc γ RI, Fc γ RII, and Fc γ RIII, which mediate various effector functions will not bind to IgG1 when such mutations have been introduced. Ward and Ghetie 1995, Therapeutic Immunology 2:77 and Armour et al. 1999, Eur. J. Immunol. 29:2613.

[0158] In another embodiment, the immunoglobulin constant region or a portion thereof comprises an amino acid sequence in the hinge region or a portion thereof that forms one or more disulfide bonds with a second immunoglobulin constant region or a portion thereof. The second immunoglobulin constant region or a portion thereof can be linked to a second polypeptide, bringing the FVIII protein and the second polypeptide together. In some embodiments, the second polypeptide is an enhancer moiety. As used herein, the term "enhancer moiety" refers to a molecule, fragment thereof or a component of a polypeptide which is capable of enhancing the procoagulant activity of FVIII. The enhancer moiety can be a cofactor, such as soluble tissue factor (sTF), or a procoagulant peptide. Thus, upon activation of FVIII, the enhancer moiety is available to enhance FVIII activity.

[0159] In certain embodiments, a FVIII protein of the invention comprises an amino acid substitution to an immunoglobulin constant region or a portion thereof (e.g., Fc variants), which alters the antigen-independent effector functions of the Ig constant region, in particular the circulating half-life of the protein.

2. scFc Regions

[0160] In another aspect, a heterologous moiety comprises a scFc (single chain Fc) region. In one embodiment, an isolated nucleic acid molecule of the invention further comprises a heterologous nucleic acid sequence that encodes a scFc region. The scFc region comprises at least two immunoglobulin constant regions or portions thereof (e.g., Fc moieties or domains (e.g., 2, 3, 4, 5, 6, or more Fc moieties or domains)) within the same linear polypeptide chain that are capable of folding (e.g., intramolecularly or intermolecularly folding) to form one functional scFc region which is linked by an Fc peptide linker. For example, in one embodiment, a polypeptide as disclosed herein is capable of binding, via its scFc region, to at least one Fc receptor (e.g. an FcRn, an FcγR receptor (e.g., FcγRIII), or a complement protein (e.g. C1q)) in order to improve half-life or trigger an immune effector function (e.g., antibody-dependent cytotoxicity (ADCC), phagocytosis, or complement-dependent cytotoxicity (CDCC) and/or to improve manufacturability).

3. CTP

[0161] In another aspect, a heterologous moiety comprises one C-terminal peptide (CTP) of the β subunit of human chorionic gonadotropin or fragment, variant, or derivative thereof. One or more CTP peptides inserted into a recombinant protein is known to increase the *in vivo* half-life of that protein. See, e.g., U.S. Patent No. 5,712,122.

[0162] Exemplary CTP peptides include DPRFQDSSSSKAPPPSLPSPSRLPGPSDTPIL (SEQ ID NO:17) or SSSSKAPPPSLPSPSRLPGPSDTPILPQ. (SEQ ID NO:18). See, e.g., U.S. Patent Application Publication No. US 2009/0087411 A1.

4. XTEN Sequence

[0163] In some embodiments, a heterologous moiety comprises one or more XTEN sequences, fragments, variants, or derivatives thereof. As used here "XTEN sequence" refers to extended length polypeptides with non-naturally occurring, substantially non-repetitive sequences that are composed mainly of small hydrophilic amino acids, with the sequence having a low degree or no secondary or tertiary structure under physiologic conditions. As a heterologous moiety, XTENs can serve as a half-life extension moiety. In addition, XTEN can provide desirable properties including but are not limited to enhanced pharmacokinetic parameters and solubility characteristics.

[0164] The incorporation of a heterologous moiety comprising an XTEN sequence into a protein of the invention can confer to the protein one or more of the following advantageous properties: conformational flexibility, enhanced aqueous solubility, high degree of protease resistance, low immunogenicity, low binding to mammalian receptors, or increased hydrodynamic (or Stokes) radii.

[0165] In certain aspects, an XTEN sequence can increase pharmacokinetic properties such as longer *in vivo* half-life or increased area under the curve (AUC), so that a protein of the invention stays *in vivo* and has procoagulant activity for an increased period of time compared to a protein with the same but without the XTEN heterologous moiety.

[0166] Examples of XTEN sequences that can be used as heterologous moieties in chimeric proteins of the invention are disclosed, e.g., in U.S. Patent Publication Nos. 2010/0239554 A1, 2010/0323956 A1, 2011/0046060 A1, 2011/0046061 A1, 2011/0077199 A1, or 2011/0172146 A1, or International Patent Publication Nos. WO 2010091122 A1, WO 2010144502 A2, WO 2010144508 A1, WO 2011028228 A1, WO 2011028229 A1, or WO 2011028344 A2.

[0167] Exemplary XTEN sequences that can be used as heterologous moieties in chimeric protein of the invention include XTEN AE42-4 (SEQ ID NO:30, encoded by SEQ ID NO:31), XTEN 144-2A (SEQ ID NO:32, encoded by SEQ ID NO:33), XTEN A144-3B (SEQ ID NO:34, encoded by SEQ ID NO:35), XTEN AE144-4A (SEQ ID NO:36, encoded by SEQ ID NO:37), XTEN AE144-5A (SEQ ID NO:38, encoded by SEQ ID NO:39), XTEN AE144-6B (SEQ ID NO:40, encoded by SEQ ID NO:41), XTEN AG144-1 (SEQ ID NO:42, encoded by SEQ ID NO:43), XTEN AG144-A (SEQ ID NO:44, encoded by SEQ ID NO:45), XTEN AG144-B (SEQ ID NO:46, encoded by SEQ ID NO:47), XTEN AG144-C (SEQ ID NO:48, encoded by SEQ ID NO:49), and XTEN AG144-F (SEQ ID NO:50, encoded by SEQ ID NO:51).

5. Albumin or Fragment, Derivative, or Variant Thereof

[0168] In some embodiments, a heterologous moiety comprises albumin or a functional

fragment thereof. Human serum albumin (HSA, or HA), a protein of 609 amino acids in its full-length form, is responsible for a significant proportion of the osmotic pressure of serum and also functions as a carrier of endogenous and exogenous ligands. The term "albumin" as used herein includes full-length albumin or a functional fragment, variant, derivative, or analog thereof. Examples of albumin or the fragments or variants thereof are disclosed in US Pat. Publ. Nos. 2008/0194481A1, 2008/0004206 A1, 2008/0161243 A1, 2008/0261877 A1, or 2008/0153751 A1 or PCT Appl. Publ. Nos. 2008/033413 A2, 2009/058322 A1, or 2007/021494 A2.

[0169] In one embodiment, the FVIII protein of the invention comprises albumin, a fragment, or a variant thereof which is further linked to a second heterologous moiety selected from the group consisting of an immunoglobulin constant region or portion thereof (e.g., an Fc region), a PAS sequence, HES, and PEG.

6. Albumin-binding Moiety

[0170] In certain embodiments, the heterologous moiety is an albumin-binding moiety, which comprises an albumin-binding peptide, a bacterial albumin-binding domain, an albumin-binding antibody fragment, or any combinations thereof.

[0171] For example, the albumin-binding protein can be a bacterial albumin-binding protein, an antibody or an antibody fragment including domain antibodies (see U.S. Pat. No. 6,696,245). An albumin-binding protein, for example, can be a bacterial albumin-binding domain, such as the one of streptococcal protein G (Konig, T. and Skerra, A. (1998) J. Immunol. Methods 218, 73-83). Other examples of albumin-binding peptides that can be used as conjugation partner are, for instance, those having a Cys-Xaa₁-Xaa₂-Xaa₃-Xaa₄-Cys consensus sequence, wherein Xaa₁ is Asp, Asn, Ser, Thr, or Trp; Xaa₂ is Asn, Gln, His, Ile, Leu, or Lys; Xaa₃ is Ala, Asp, Phe, Trp, or Tyr; and Xaa₄ is Asp, Gly, Leu, Phe, Ser, or Thr as described in US patent application 2003/0069395 or Dennis et al. (Dennis et al. (2002) J. Biol. Chem. 277, 35035-35043).

[0172] Domain 3 from streptococcal protein G, as disclosed by Kraulis et al., FEBS Lett. 378:190-194 (1996) and Linhult et al., Protein Sci. 11:206-213 (2002) is an example of a bacterial albumin-binding domain. Examples of albumin-binding peptides include a series of peptides having the core sequence DICLPRWGCLW (SEQ ID NO:19). See, e.g., Dennis et al., J. Biol. Chem. 2002, 277: 35035-35043 (2002). Examples of albumin-binding antibody fragments are disclosed in Muller and Kontermann, Curr. Opin. Mol. Ther. 9:319-326 (2007); Roover et al., Cancer Immunol. Immunother. 56:303-317 (2007), and Holt et al., Prot. Eng. Design Sci., 21:283-288 (2008). An example of such albumin-binding moiety is 2-(3-maleimidopropanamido)-6-(4-(4-iodophenyl)butanamido) hexanoate ("Albu" tag) as disclosed by Trussetlet al., Bioconjugate Chem. 20:2286-2292 (2009).

[0173] Fatty acids, in particular long chain fatty acids (LCFA) and long chain fatty acid-like

albumin-binding compounds can be used to extend the *in vivo* half-life of FVIII proteins of the invention. An example of a LCFA-like albumin-binding compound is 16-(1-(3-(9-(((2,5-dioxopyrrolidin-1-yl)oxy) carbonyloxy)-methyl)-7-sulfo-9H-fluoren-2-ylamino)-3-oxopropyl)-2,5-dioxopyrrolidin-3-ylthio) hexadecanoic acid (see, e.g., WO 2010/140148).

7. PAS Sequence

[0174] In other embodiments, the heterologous moiety is a PAS sequence. A PAS sequence, as used herein, means an amino acid sequence comprising mainly alanine and serine residues or comprising mainly alanine, serine, and proline residues, the amino acid sequence forming random coil conformation under physiological conditions. Accordingly, the PAS sequence is a building block, an amino acid polymer, or a sequence cassette comprising, consisting essentially of, or consisting of alanine, serine, and proline which can be used as a part of the heterologous moiety in the chimeric protein. Yet, the skilled person is aware that an amino acid polymer also can form random coil conformation when residues other than alanine, serine, and proline are added as a minor constituent in the PAS sequence. The term "minor constituent" as used herein means that amino acids other than alanine, serine, and proline can be added in the PAS sequence to a certain degree, e.g., up to about 12%, *i.e.*, about 12 of 100 amino acids of the PAS sequence, up to about 10%, *i.e.* about 10 of 100 amino acids of the PAS sequence, up to about 9%, *i.e.*, about 9 of 100 amino acids, up to about 8%, *i.e.*, about 8 of 100 amino acids, about 6%, *i.e.*, about 6 of 100 amino acids, about 5%, *i.e.*, about 5 of 100 amino acids, about 4%, *i.e.*, about 4 of 100 amino acids, about 3%, *i.e.*, about 3 of 100 amino acids, about 2%, *i.e.*, about 2 of 100 amino acids, about 1%, *i.e.*, about 1 of 100 of the amino acids. The amino acids different from alanine, serine and proline can be selected from the group consisting of Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Thr, Trp, Tyr, and Val.

[0175] Under physiological conditions, the PAS sequence stretch forms a random coil conformation and thereby can mediate an increased *in vivo* and/or *in vitro* stability to the FVIII protein. Since the random coil domain does not adopt a stable structure or function by itself, the biological activity mediated by the FVIII protein is essentially preserved. In other embodiments, the PAS sequences that form random coil domain are biologically inert, especially with respect to proteolysis in blood plasma, immunogenicity, isoelectric point/electrostatic behaviour, binding to cell surface receptors or internalisation, but are still biodegradable, which provides clear advantages over synthetic polymers such as PEG.

[0176] Non-limiting examples of the PAS sequences forming random coil conformation comprise an amino acid sequence selected from the group consisting of ASPAAPAPASPAAPAPSAPA (SEQ ID NO: 20), AAPASPAPAAPSAPAPAAPS (SEQ ID NO: 21), APSSPSPSAPSSPSPASPSS (SEQ ID NO: 22), APSSPSPSAPSSPSPASPS (SEQ ID NO: 23), SSPSAPSPSSPASPSPSSPA (SEQ ID NO: 24), AASPAAPSAPPAAASPAAPSAPPA (SEQ ID NO: 25) and ASAAAPAAASAAASAPSAAA (SEQ ID NO: 26) or any combinations thereof. Additional examples of PAS sequences are known from, e.g., US Pat. Publ. No. 2010/0292130 A1 and

PCT Appl. Publ. No. WO 2008/155134 A1.

8. HAP Sequence

[0177] In certain embodiments, the heterologous moiety is a glycine-rich homo-aminoacid polymer (HAP). The HAP sequence can comprise a repetitive sequence of glycine, which has at least 50 amino acids, at least 100 amino acids, 120 amino acids, 140 amino acids, 160 amino acids, 180 amino acids, 200 amino acids, 250 amino acids, 300 amino acids, 350 amino acids, 400 amino acids, 450 amino acids, or 500 amino acids in length. In one embodiment, the HAP sequence is capable of extending half-life of a moiety fused to or linked to the HAP sequence. Non-limiting examples of the HAP sequence includes, but are not limited to $(\text{Gly})_n$, $(\text{Gly}_4\text{Ser})_n$ or $\text{S}(\text{Gly}_4\text{Ser})_n$, wherein n is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20. In one embodiment, n is 20, 21, 22, 23, 24, 25, 26, 26, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40. In another embodiment, n is 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, or 200.

9. Transferrin or Fragment thereof

[0178] In certain embodiments, the heterologous moiety is transferrin or a fragment thereof. Any transferrin can be used to make the FVIII proteins of the invention. As an example, wild-type human TF (TF) is a 679 amino acid protein, of approximately 75 KDa (not accounting for glycosylation), with two main domains, N (about 330 amino acids) and C (about 340 amino acids), which appear to originate from a gene duplication. See GenBank accession numbers NM001063, XM002793, M12530, XM039845, XM 039847 and S95936 (www.ncbi.nlm.nih.gov). Transferrin comprises two domains, N domain and C domain. N domain comprises two subdomains, N1 domain and N2 domain, and C domain comprises two subdomains, C1 domain and C2 domain.

[0179] In one embodiment, the transferrin heterologous moiety includes a transferrin splice variant. In one example, a transferrin splice variant can be a splice variant of human transferrin, e.g., Genbank Accession AAA61140. In another embodiment, the transferrin portion of the chimeric protein includes one or more domains of the transferrin sequence, e.g., N domain, C domain, N1 domain, N2 domain, C1 domain, C2 domain or any combinations thereof.

10. Clearance Receptors

[0180] In certain embodiments, the heterologous moiety is a clearance receptor, fragment, variant, or derivative thereof. LRP1 is a 600 kDa integral membrane protein that is implicated in the receptor-mediate clearance of a variety of proteins, such as Factor X. See, e.g., Narita et

al., Blood 91:555-560 (1998).

11. von Willebrand Factor or Fragments Thereof

[0181] In certain embodiments, the heterologous moiety is von Willebrand Factor (VWF) or fragments thereof.

[0182] VWF (also known as F8VWF) is a large multimeric glycoprotein present in blood plasma and produced constitutively in endothelium (in the Weibel-Palade bodies), megakaryocytes (α -granules of platelets), and subendothelial connective tissue. The basic VWF monomer is a 2813 amino acid protein. Every monomer contains a number of specific domains with a specific function, the D' and D3 domains (which together bind to Factor VIII), the A1 domain (which binds to platelet GPIb-receptor, heparin, and/or possibly collagen), the A3 domain (which binds to collagen), the C1 domain (in which the RGD domain binds to platelet integrin α IIb β 3 when this is activated), and the "cysteine knot" domain at the C-terminal end of the protein (which VWF shares with platelet-derived growth factor (PDGF), transforming growth factor- β (TGF β) and β -human chorionic gonadotropin (β HCG)).

[0183] The 2813 monomer amino acid sequence for human VWF is reported as Accession Number NP000543.2 in Genbank. The nucleotide sequence encoding the human VWF is reported as Accession Number NM000552.3 in Genbank. The nucleotide sequence of human VWF is designated as SEQ ID NO: 27. SEQ ID NO: 28 is the amino acid sequence encoded by SEQ ID NO: 27. The D' domain includes amino acids 764 to 866 of SEQ ID NO:28. The D3 domain includes amino acids 867 to 1240 of SEQ ID NO:28.

[0184] In plasma, 95-98% of FVIII circulates in a tight non-covalent complex with full-length VWF. The formation of this complex is important for the maintenance of appropriate plasma levels of FVIII *in vivo*. Lenting et al., Blood. 92(11): 3983-96 (1998); Lenting et al., J. Thromb. Haemost. 5(7): 1353-60 (2007). When FVIII is activated due to proteolysis at positions 372 and 740 in the heavy chain and at position 1689 in the light chain, the VWF bound to FVIII is removed from the activated FVIII.

[0185] In certain embodiments, the heterologous moiety is full length von Willebrand Factor. In other embodiments, the heterologous moiety is a von Willebrand Factor fragment. As used herein, the term "VWF fragment" or "VWF fragments" used herein means any VWF fragments that interact with FVIII and retain at least one or more properties that are normally provided to FVIII by full-length VWF, e.g., preventing premature activation to FVIIIa, preventing premature proteolysis, preventing association with phospholipid membranes that could lead to premature clearance, preventing binding to FVIII clearance receptors that can bind naked FVIII but not VWF-bound FVIII, and/or stabilizing the FVIII heavy chain and light chain interactions. In a specific embodiment, the heterologous moiety is a (VWF) fragment comprising a D' domain and a D3 domain of VWF. The VWF fragment comprising the D' domain and the D3 domain can further comprise a VWF domain selected from the group consisting of an A1 domain, an

A2 domain, an A3 domain, a D1 domain, a D2 domain, a D4 domain, a B1 domain, a B2 domain, a B3 domain, a C1 domain, a C2 domain, a CK domain, one or more fragments thereof, and any combinations thereof. Additional examples of the polypeptide having FVIII activity fused to the VWF fragment are disclosed in U.S. provisional patent application no. 61/667,901, filed July 3, 2012.

12. Linker Moieties

[0186] In certain embodiments, the heterologous moiety is a peptide linker.

[0187] As used herein, the terms "peptide linkers" or "linker moieties" refer to a peptide or polypeptide sequence (e.g., a synthetic peptide or polypeptide sequence) which connects two domains in a linear amino acid sequence of a polypeptide chain.

[0188] In some embodiments, heterologous nucleotide sequences encoding peptide linkers can be inserted between the optimized FVIII polynucleotide sequences of the invention and a heterologous nucleotide sequence encoding, for example, one of the heterologous moieties described above, such as albumin. Peptide linkers can provide flexibility to the chimeric polypeptide molecule. Linkers are not typically cleaved, however such cleavage can be desirable. In one embodiment, these linkers are not removed during processing.

[0189] A type of linker which can be present in a chimeric protein of the invention is a protease cleavable linker which comprises a cleavage site (i.e., a protease cleavage site substrate, e.g., a factor Xla, Xa, or thrombin cleavage site) and which can include additional linkers on either the N-terminal of or both sides of the cleavage site. These cleavable linkers when incorporated into a construct of the invention result in a chimeric molecule having a heterologous cleavage site.

[0190] In one embodiment, an FVIII polypeptide comprises two or more Fc domains or moieties linked via a cscFc linker to form an Fc region comprised in a single polypeptide chain. The cscFc linker is flanked by at least one intracellular processing site, i.e., a site cleaved by an intracellular enzyme. Cleavage of the polypeptide at the at least one intracellular processing site results in a polypeptide which comprises at least two polypeptide chains.

[0191] Other peptide linkers can optionally be used in a construct of the invention, e.g., to connect an FVIII protein to an Fc region. Some exemplary linkers that can be used in connection with the invention include, e.g., polypeptides comprising GlySer amino acids described in more detail below.

[0192] In one embodiment, the peptide linker is synthetic, i.e., non-naturally occurring. In one embodiment, a peptide linker includes peptides (or polypeptides) (which can or can not be naturally occurring) which comprise an amino acid sequence that links or genetically fuses a first linear sequence of amino acids to a second linear sequence of amino acids to which it is

not naturally linked or genetically fused in nature. For example, in one embodiment the peptide linker can comprise non-naturally occurring polypeptides which are modified forms of naturally occurring polypeptides (e.g., comprising a mutation such as an addition, substitution or deletion). In another embodiment, the peptide linker can comprise non-naturally occurring amino acids. In another embodiment, the peptide linker can comprise naturally occurring amino acids occurring in a linear sequence that does not occur in nature. In still another embodiment, the peptide linker can comprise a naturally occurring polypeptide sequence.

[0193] For example, in certain embodiments, a peptide linker can be used to fuse identical Fc moieties, thereby forming a homodimeric scFc region. In other embodiments, a peptide linker can be used to fuse different Fc moieties (e.g. a wild-type Fc moiety and an Fc moiety variant), thereby forming a heterodimeric scFc region.

[0194] In another embodiment, a peptide linker comprises or consists of a gly-ser linker. In one embodiment, a scFc or cscFc linker comprises at least a portion of an immunoglobulin hinge and a gly-ser linker. As used herein, the term "gly-ser linker" refers to a peptide that consists of glycine and serine residues. In certain embodiments, said gly-ser linker can be inserted between two other sequences of the peptide linker. In other embodiments, a gly-ser linker is attached at one or both ends of another sequence of the peptide linker. In yet other embodiments, two or more gly-ser linker are incorporated in series in a peptide linker. In one embodiment, a peptide linker of the invention comprises at least a portion of an upper hinge region (e.g., derived from an IgG1, IgG2, IgG3, or IgG4 molecule), at least a portion of a middle hinge region (e.g., derived from an IgG1, IgG2, IgG3, or IgG4 molecule) and a series of gly/ser amino acid residues.

[0195] Peptide linkers of the invention are at least one amino acid in length and can be of varying lengths. In one embodiment, a peptide linker of the invention is from about 1 to about 50 amino acids in length. As used in this context, the term "about" indicates +/-two amino acid residues. Since linker length must be a positive interger, the length of from about 1 to about 50 amino acids in length, means a length of from 1-3 to 48-52 amino acids in length. In another embodiment, a peptide linker of the invention is from about 10 to about 20 amino acids in length. In another embodiment, a peptide linker of the invention is from about 15 to about 50 amino acids in length. In another embodiment, a peptide linker of the invention is from about 20 to about 45 amino acids in length. In another embodiment, a peptide linker of the invention is from about 15 to about 35 or about 20 to about 30 amino acids in length. In another embodiment, a peptide linker of the invention is from about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 40, 50, 60, 70, 80, 90, 100, 500, 1000, or 2000 amino acids in length. In one embodiment, a peptide linker of the invention is 20 or 30 amino acids in length.

[0196] In some embodiments, the peptide linker can comprise at least two, at least three, at least four, at least five, at least 10, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, or at least 100 amino acids. In other embodiments, the peptide linker can comprise at least 200, at least 300, at least 400, at least 500, at least 600, at

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

[0197] Peptide linkers can be introduced into polypeptide sequences using techniques known in the art. Modifications can be confirmed by DNA sequence analysis. Plasmid DNA can be used to transform host cells for stable production of the polypeptides produced.

Monomer-Dimer Hybrids

[0198] In some embodiments, the isolated nucleic acid molecules of the invention which further comprise a heterologous nucleotide sequence encode a monomer-dimer hybrid molecule comprising FVIII.

[0199] The term "monomer-dimer hybrid" used herein refers to a chimeric protein comprising a first polypeptide chain and a second polypeptide chain, which are associated with each other by a disulfide bond, wherein the first chain comprises Factor VIII and a first Fc region and the second chain comprises, consists essentially of, or consists of a second Fc region without the FVIII. The monomer-dimer hybrid construct thus is a hybrid comprising a monomer aspect having only one clotting factor and a dimer aspect having two Fc regions.

Transcription Control Sequences

[0200] In some embodiments, the isolated nucleic acid molecules of the invention are operatively linked to at least one transcription control sequences. A transcription control sequences as used herein is any regulatory nucleotide sequence, such as a promoter sequence or promoter-enhancer combination, which facilitates the efficient transcription and translation of the coding nucleic acid to which it is operably linked. The gene expression control sequence can, for example, be a mammalian or viral promoter, such as a constitutive or inducible promoter. Constitutive mammalian promoters include, but are not limited to, the promoters for the following genes: hypoxanthine phosphoribosyl transferase (HPRT), adenosine deaminase, pyruvate kinase, beta-actin promoter, and other constitutive promoters. Exemplary viral promoters which function constitutively in eukaryotic cells include, for example, promoters from the cytomegalovirus (CMV), simian virus (e.g., SV40), papilloma virus, adenovirus, human immunodeficiency virus (HIV), Rous sarcoma virus, cytomegalovirus, the long terminal repeats (LTR) of Moloney leukemia virus, and other retroviruses, and the thymidine kinase promoter of herpes simplex virus. Other constitutive promoters are known to

those of ordinary skill in the art. The promoters useful as gene expression sequences of the invention also include inducible promoters. Inducible promoters are expressed in the presence of an inducing agent. For example, the metallothionein promoter is induced to promote transcription and translation in the presence of certain metal ions. Other inducible promoters are known to those of ordinary skill in the art.

[0201] In general, the transcription control sequences shall include, as necessary, 5' non-transcribing and 5' non-translating sequences involved with the initiation of transcription and translation, respectively, such as a TATA box, capping sequence, CAAT sequence, and the like. Especially, such 5' non-transcribing sequences will include a promoter region which includes a promoter sequence for transcriptional control of the operably joined coding nucleic acid. The gene expression sequences optionally include enhancer sequences or upstream activator sequences as desired.

Vectors

[0202] The invention also provides vectors comprising the isolated nucleic acid molecules of the invention. Suitable vectors include expression vectors, viral vectors, and plasmid vectors.

[0203] As used herein, an expression vector refers to any nucleic acid construct which contains the necessary elements for the transcription and translation of an inserted coding sequence, or in the case of an RNA viral vector, the necessary elements for replication and translation, when introduced into an appropriate host cell. Expression vectors can include plasmids, phagemids, viruses, and derivatives thereof.

[0204] Expression vectors of the invention will include optimized polynucleotides encoding the BDD FVIII protein described herein. In one embodiment, the optimized coding sequences for the BDD FVIII protein is operably linked to an expression control sequence. As used herein, two nucleic acid sequences are operably linked when they are covalently linked in such a way as to permit each component nucleic acid sequence to retain its functionality. A coding sequence and a gene expression control sequence are said to be operably linked when they are covalently linked in such a way as to place the expression or transcription and/or translation of the coding sequence under the influence or control of the gene expression control sequence. Two DNA sequences are said to be operably linked if induction of a promoter in the 5' gene expression sequence results in the transcription of the coding sequence and if the nature of the linkage between the two DNA sequences does not (1) result in the introduction of a frame-shift mutation, (2) interfere with the ability of the promoter region to direct the transcription of the coding sequence, or (3) interfere with the ability of the corresponding RNA transcript to be translated into a protein. Thus, a gene expression sequence would be operably linked to a coding nucleic acid sequence if the gene expression sequence were capable of effecting transcription of that coding nucleic acid sequence such that the resulting transcript is translated into the desired protein or polypeptide.

[0205] Viral vectors include, but are not limited to, nucleic acid sequences from the following viruses: retrovirus, such as Moloney murine leukemia virus, Harvey murine sarcoma virus, murine mammary tumor virus, and Rous sarcoma virus; adenovirus, adeno-associated virus; SV40-type viruses; polyomaviruses; Epstein-Barr viruses; papilloma viruses; herpes virus; vaccinia virus; polio virus; and RNA virus such as a retrovirus. One can readily employ other vectors well-known in the art. Certain viral vectors are based on non-cytopathic eukaryotic viruses in which non-essential genes have been replaced with the gene of interest. Non-cytopathic viruses include retroviruses, the life cycle of which involves reverse transcription of genomic viral RNA into DNA with subsequent proviral integration into host cellular DNA. Retroviruses have been approved for human gene therapy trials. Most useful are those retroviruses that are replication-deficient (*i.e.*, capable of directing synthesis of the desired proteins, but incapable of manufacturing an infectious particle). Such genetically altered retroviral expression vectors have general utility for the high efficiency transduction of genes *in vivo*. Standard protocols for producing replication-deficient retroviruses (including the steps of incorporation of exogenous genetic material into a plasmid, transfection of a packaging cell line with plasmid, production of recombinant retroviruses by the packaging cell line, collection of viral particles from tissue culture media, and infection of the target cells with viral particles) are provided in 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).

[0206] In one embodiment, the virus is an adeno-associated virus, a double-stranded DNA virus. The adeno-associated virus can be engineered to be replication-deficient and is capable of infecting a wide range of cell types and species. It further has advantages such as heat and lipid solvent stability; high transduction frequencies in cells of diverse lineages, including hematopoietic cells; and lack of superinfection inhibition thus allowing multiple series of transductions. Reportedly, the adeno-associated virus can integrate into human cellular DNA in a site-specific manner, thereby minimizing the possibility of insertional mutagenesis and variability of inserted gene expression characteristic of retroviral infection. In addition, wild-type adeno-associated virus infections have been followed in tissue culture for greater than 100 passages in the absence of selective pressure, implying that the adeno-associated virus genomic integration is a relatively stable event. The adeno-associated virus can also function in an extrachromosomal fashion.

[0207] Other vectors include plasmid vectors. Plasmid vectors have been extensively described in the art and are well-known to those of skill in the art. See, *e.g.*, Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, 1989. In the last few years, plasmid vectors have been found to be particularly advantageous for delivering genes to cells *in vivo* because of their inability to replicate within and integrate into a host genome. These plasmids, however, having a promoter compatible with the host cell, can express a peptide from a gene operably encoded within the plasmid. Some commonly used plasmids available from commercial suppliers include pBR322, pUC18, pUC19, various pcDNA plasmids, pRC/CMV, various pCMV plasmids, pSV40, and pBlueScript. Additional examples of specific plasmids include pcDNA3.1, catalog number V79020;

pcDNA3.1/hygro, catalog number V87020; pcDNA4/myc-His, catalog number V86320; and pBudCE4.1, catalog number V53220, all from Invitrogen (Carlsbad, CA.). Other plasmids are well-known to those of ordinary skill in the art. Additionally, plasmids can be custom designed using standard molecular biology techniques to remove and/or add specific fragments of DNA.

Host Cells

[0208] The invention also provides host cells comprising the isolated nucleic acid molecules of the invention. As used herein, the term "transformation" shall be used in a broad sense to refer to the introduction of DNA into a recipient host cell that changes the genotype and consequently results in a change in the recipient cell.

[0209] "Host cells" refers to cells that have been transformed with vectors constructed using recombinant DNA techniques and encoding at least one heterologous gene. The host cells of the present invention are preferably of mammalian origin; most preferably of human or mouse origin. Those skilled in the art are credited with ability to preferentially determine particular host cell lines which are best suited for their purpose. Exemplary host cell lines include, but are not limited to, CHO, DG44 and DUXB11 (Chinese Hamster Ovary lines, DHFR minus), HELA (human cervical carcinoma), CVI (monkey kidney line), COS (a derivative of CVI with SV40 T antigen), R1610 (Chinese hamster fibroblast) BALBC/3T3 (mouse fibroblast), HAK (hamster kidney line), SP2/O (mouse myeloma), P3.times.63-Ag3.653 (mouse myeloma), BFA-1c1BPT (bovine endothelial cells), RAJI (human lymphocyte), PER.C6[®], NS0, CAP, BHK21, and HEK 293 (human kidney). Host cell lines are typically available from commercial services, the American Tissue Culture Collection, or from published literature.

[0210] Introduction of the isolated nucleic acid molecules of the invention into the host cell can be accomplished by various techniques well known to those of skill in the art. These include, but are not limited to, transfection (including electrophoresis and electroporation), protoplast fusion, calcium phosphate precipitation, cell fusion with enveloped DNA, microinjection, and infection with intact virus. See, Ridgway, A. A. G. "Mammalian Expression Vectors" Chapter 24.2, pp. 470-472 Vectors, Rodriguez and Denhardt, Eds. (Butterworths, Boston, Mass. 1988). Most preferably, plasmid introduction into the host is via electroporation. The transformed cells are grown under conditions appropriate to the production of the light chains and heavy chains, and assayed for heavy and/or light chain protein synthesis. Exemplary assay techniques include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), or fluorescence-activated cell sorter analysis (FACS), immunohistochemistry and the like.

[0211] Host cells comprising the isolated nucleic acid molecules of the invention are grown in an appropriate growth medium. As used herein, the term "appropriate growth medium" means a medium containing nutrients required for the growth of cells. Nutrients required for cell growth can include a carbon source, a nitrogen source, essential amino acids, vitamins, minerals, and growth factors. Optionally, the media can contain one or more selection factors.

Optionally the media can contain bovine calf serum or fetal calf serum (FCS). In one embodiment, the media contains substantially no IgG. The growth medium will generally select for cells containing the DNA construct by, for example, drug selection or deficiency in an essential nutrient which is complemented by the selectable marker on the DNA construct or co-transfected with the DNA construct. Cultured mammalian cells are generally grown in commercially available serum-containing or serum-free media (e.g., MEM, DMEM, DMEM/F12). In one embodiment, the medium is CDoptiCHO (Invitrogen, Carlsbad, CA.). In another embodiment, the medium is CD17 (Invitrogen, Carlsbad, CA.). Selection of a medium appropriate for the particular cell line used is within the level of those ordinary skilled in the art.

Preparation of Polypeptides

[0212] A polypeptide may be encoded by the isolated nucleic acid molecules of the invention. In some embodiments, the polypeptide is encoded by a vector comprising the isolated nucleic molecules of the invention. In other embodiments, the polypeptide is produced by a host cell comprising the isolated nucleic molecules of the invention.

[0213] In one embodiment, the invention provides a method of producing a polypeptide with FVIII activity, comprising culturing a host cell of the invention under conditions whereby a polypeptide with FVIII activity is produced, and recovering the polypeptide with FVIII activity. In some embodiments, the expression of the polypeptide with FVIII activity is increased relative to a host cell cultured under the same conditions but containing a reference nucleotide sequence comprising SEQ ID NO:3, the parental FVIII gene sequence.

[0214] In other embodiments, the invention provides a method of increasing the expression of a polypeptide with FVIII activity comprising culturing a host cell of the invention under conditions whereby a polypeptide with FVIII activity is expressed by the nucleic acid molecule, wherein the expression of the polypeptide with FVIII activity is increased relative to a host cell cultured under the same conditions comprising a reference nucleic acid molecule comprising SEQ ID NO:3.

[0215] In other embodiments, the invention provides a method of improving yield of a polypeptide with Factor VIII activity comprising culturing a host cell under conditions whereby a polypeptide with Factor VIII activity is produced by the nucleic acid molecule, wherein the yield of polypeptide with Factor VIII activity is increased relative to a host cell cultured under the same conditions comprising a reference nucleic acid sequence comprising SEQ ID NO: 3.

[0216] A variety of methods are available for recombinantly producing a FVIII protein from the optimized nucleic acid molecule of the invention. A polynucleotide of the desired sequence can be produced by *de novo* solid-phase DNA synthesis or by PCR mutagenesis of an earlier prepared polynucleotide. Oligonucleotide-mediated mutagenesis is one method for preparing a substitution, insertion, deletion, or alteration (e.g., altered codon) in a nucleotide sequence. For example, the starting DNA is altered by hybridizing an oligonucleotide encoding the desired

mutation to a single-stranded DNA template. After hybridization, a DNA polymerase is used to synthesize an entire second complementary strand of the template that incorporates the oligonucleotide primer. In one embodiment, genetic engineering, e.g., primer-based PCR mutagenesis, is sufficient to incorporate an alteration, as defined herein, for producing a polynucleotide of the invention.

[0217] For recombinant protein production, an optimized polynucleotide sequence of the invention encoding the FVIII protein is inserted into an appropriate expression vehicle, *i.e.*, a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence, or in the case of an RNA viral vector, the necessary elements for replication and translation.

[0218] The polynucleotide sequence of the invention is inserted into the vector in proper reading frame. The expression vector is then transfected into a suitable target cell which will express the polypeptide. Transfection techniques known in the art include, but are not limited to, calcium phosphate precipitation (Wigler et al. 1978, Cell 14 : 725) and electroporation (Neumann et al. 1982, EMBO, J. 1 : 841). A variety of host-expression vector systems can be utilized to express the FVIII proteins described herein in eukaryotic cells. In one embodiment, the eukaryotic cell is an animal cell, including mammalian cells (e.g. HEK293 cells, PER.C6®, CHO, BHK, Cos, HeLa cells). A polynucleotide sequence of the invention can also code for a signal sequence that will permit the FVIII protein to be secreted. One skilled in the art will understand that while the FVIII protein is translated the signal sequence is cleaved by the cell to form the mature protein. Various signal sequences are known in the art, e.g., native factor VII signal sequence, native factor IX signal sequence and the mouse IgK light chain signal sequence. Alternatively, where a signal sequence is not included the FVIII protein can be recovered by lysing the cells.

[0219] The FVIII protein can be synthesized in a non-human transgenic animal, such as a rodent, goat, sheep, pig, or cow. The term "transgenic animals" refers to non-human animals that have incorporated a foreign gene into their genome. Because this gene is present in germline tissues, it is passed from parent to offspring. Exogenous genes are introduced into single-celled embryos (Brinster et al. 1985, Proc. Natl. Acad.Sci. USA 82:4438). Methods of producing transgenic animals are known in the art including transgenics that produce immunoglobulin molecules (Wagner et al. 1981, Proc. Natl. Acad. Sci. USA 78: 6376; McKnight et al. 1983, Cell 34 : 335; Brinster et al. 1983, Nature 306: 332; Ritchie et al. 1984, Nature 312: 517; Baldassarre et al. 2003, Theriogenology 59 : 831; Robl et al. 2003, Theriogenology 59: 107; Malassagne et al. 2003, Xenotransplantation 10 (3): 267).

[0220] The expression vectors can encode for tags that permit for easy purification or identification of the recombinantly produced protein. Examples include, but are not limited to, vector pUR278 (Ruther et al. 1983, EMBO J. 2: 1791) in which the FVIII protein described herein coding sequence can be ligated into the vector in frame with the lac Z coding region so that a hybrid protein is produced; pGEX vectors can be used to express proteins with a glutathione S-transferase (GST) tag. These proteins are usually soluble and can easily be

purified from cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The vectors include cleavage sites (e.g., PreCission Protease (Pharmacia, Peapack, N. J.)) for easy removal of the tag after purification.

[0221] For the purposes of this invention, numerous expression vector systems can be employed. These expression vectors are typically replicable in the host organisms either as episomes or as an integral part of the host chromosomal DNA. Expression vectors can include expression control sequences including, but not limited to, promoters (e.g., naturally-associated or heterologous promoters), enhancers, signal sequences, splice signals, enhancer elements, and transcription termination sequences. Preferably, the expression control sequences are eukaryotic promoter systems in vectors capable of transforming or transfecting eukaryotic host cells. Expression vectors can also utilize DNA elements which are derived from animal viruses such as bovine papilloma virus, polyoma virus, adenovirus, vaccinia virus, baculovirus, retroviruses (RSV, MMTV or MOMLV), cytomegalovirus (CMV), or SV40 virus. Others involve the use of polycistronic systems with internal ribosome binding sites.

[0222] Commonly, expression vectors contain selection markers (e.g., ampicillin-resistance, hygromycin-resistance, tetracycline resistance or neomycin resistance) to permit detection of those cells transformed with the desired DNA sequences (see, e.g., Itakura et al., US Patent 4,704,362). Cells which have integrated the DNA into their chromosomes can be selected by introducing one or more markers which allow selection of transfected host cells. The marker can provide for prototrophy to an auxotrophic host, biocide resistance (e.g., antibiotics) or resistance to heavy metals such as copper. The selectable marker gene can either be directly linked to the DNA sequences to be expressed, or introduced into the same cell by cotransformation.

[0223] An example of a vector useful for expressing an optimized FVIII sequence is NEOSPLA (U.S. Patent No. 6,159,730). This vector contains the cytomegalovirus promoter/enhancer, the mouse beta globin major promoter, the SV40 origin of replication, the bovine growth hormone polyadenylation sequence, neomycin phosphotransferase exon 1 and exon 2, the dihydrofolate reductase gene and leader sequence. This vector has been found to result in very high level expression of antibodies upon incorporation of variable and constant region genes, transfection in cells, followed by selection in G418 containing medium and methotrexate amplification. Vector systems are also taught in U.S. Pat. Nos. 5,736,137 and 5,658,570. This system provides for high expression levels, e.g., > 30 pg/cell/day. Other exemplary vector systems are disclosed e.g., in U.S. Patent No. 6,413,777.

[0224] In other embodiments the polypeptides can be expressed using polycistronic constructs. In these expression systems, multiple gene products of interest such as multiple polypeptides of multimer binding protein can be produced from a single polycistronic construct. These systems advantageously use an internal ribosome entry site (IRES) to provide relatively high levels of polypeptides in eukaryotic host cells. Compatible IRES sequences are disclosed in U.S. Pat. No. 6,193,980 which is also incorporated herein.

[0225] More generally, once the vector or DNA sequence encoding a polypeptide has been prepared, the expression vector can be introduced into an appropriate host cell. That is, the host cells can be transformed. Introduction of the plasmid into the host cell can be accomplished by various techniques well known to those of skill in the art, as discussed above. The transformed cells are grown under conditions appropriate to the production of the FVIII polypeptide, and assayed for FVIII polypeptide synthesis. Exemplary assay techniques include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), or fluorescence-activated cell sorter analysis (FACS), immunohistochemistry and the like.

[0226] In descriptions of processes for isolation of polypeptides from recombinant hosts, the terms "cell" and "cell culture" are used interchangeably to denote the source of polypeptide unless it is clearly specified otherwise. In other words, recovery of polypeptide from the "cells" can mean either from spun down whole cells, or from the cell culture containing both the medium and the suspended cells.

[0227] The host cell line used for protein expression is preferably of mammalian origin; most preferably of human or mouse origin, as the isolated nucleic acids of the invention have been optimized for expression in human cells. Exemplary host cell lines have been described above. In one embodiment of the method to produce a polypeptide with FVIII activity, the host cell is a HEK293 cell. In another embodiment of the method to produce a polypeptide with FVIII activity, the host cell is a CHO cell.

[0228] Genes encoding the polypeptides can also be expressed in non-mammalian cells such as bacteria or yeast or plant cells. In this regard it will be appreciated that various unicellular non-mammalian microorganisms such as bacteria can also be transformed; *i.e.*, those capable of being grown in cultures or fermentation. Bacteria, which are susceptible to transformation, include members of the enterobacteriaceae, such as strains of *Escherichia coli* or *Salmonella*; Bacillaceae, such as *Bacillus subtilis*; *Pneumococcus*; *Streptococcus*, and *Haemophilus influenzae*. It will further be appreciated that, when expressed in bacteria, the polypeptides typically become part of inclusion bodies. The polypeptides must be isolated, purified and then assembled into functional molecules.

[0229] Alternatively, optimized nucleotide sequences of the invention can be incorporated in transgenes for introduction into the genome of a non-human transgenic animal and subsequent expression in the milk of the transgenic animal (see, *e.g.*, Deboer et al., US 5,741,957, Rosen, US 5,304,489, and Meade et al., US 5,849,992). Suitable transgenes include coding sequences for polypeptides in operable linkage with a promoter and enhancer from a mammary gland specific gene, such as casein or beta lactoglobulin.

[0230] *In vitro* production allows scale-up to give large amounts of the desired polypeptides. Techniques for mammalian cell cultivation under tissue culture conditions are known in the art and include homogeneous suspension culture, *e.g.* in an airlift reactor or in a continuous stirrer reactor, or immobilized or entrapped cell culture, *e.g.* in hollow fibers, microcapsules, on agarose microbeads or ceramic cartridges. If necessary and/or desired, the solutions of

polypeptides can be purified by the customary chromatography methods, for example gel filtration, ion-exchange chromatography, chromatography over DEAE-cellulose or (immuno-)affinity chromatography, e.g., after preferential biosynthesis of a synthetic hinge region polypeptide or prior to or subsequent to the HIC chromatography step described herein. An affinity tag sequence (e.g. a His(6) tag) can optionally be attached or included within the polypeptide sequence to facilitate downstream purification.

[0231] Once expressed, the FVIII protein can be purified according to standard procedures of the art, including ammonium sulfate precipitation, affinity column chromatography, HPLC purification, gel electrophoresis and the like (see generally Scopes, Protein Purification (Springer-Verlag, N.Y., (1982))). Substantially pure proteins of at least about 90 to 95% homogeneity are preferred, and 98 to 99% or more homogeneity most preferred, for pharmaceutical uses.

Pharmaceutical Composition

[0232] Compositions containing the isolated nucleic acids of the present invention can contain a suitable pharmaceutically acceptable carrier. For example, they can contain excipients and/or auxiliaries that facilitate processing of the active compounds into preparations designed for delivery to the site of action.

[0233] The pharmaceutical composition can be formulated for parenteral administration (*i.e.* intravenous, subcutaneous, or intramuscular) by bolus injection. Formulations for injection can be presented in unit dosage form, e.g., in ampoules or in multidose containers with an added preservative. The compositions can take such forms as suspensions, solutions, or emulsions in oily or aqueous vehicles, and contain formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the active ingredient can be in powder form for constitution with a suitable vehicle, e.g., pyrogen free water.

[0234] Suitable formulations for parenteral administration also include aqueous solutions of the active compounds in water-soluble form, for example, water-soluble salts. In addition, suspensions of the active compounds as appropriate oily injection suspensions can be administered. Suitable lipophilic solvents or vehicles include fatty oils, for example, sesame oil, or synthetic fatty acid esters, for example, ethyl oleate or triglycerides. Aqueous injection suspensions can contain substances, which increase the viscosity of the suspension, including, for example, sodium carboxymethyl cellulose, sorbitol and dextran. Optionally, the suspension can also contain stabilizers. Liposomes also can be used to encapsulate the molecules of the invention for delivery into cells or interstitial spaces. Exemplary pharmaceutically acceptable carriers are physiologically compatible solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, water, saline, phosphate buffered saline, dextrose, glycerol, ethanol and the like. In some embodiments, the composition comprises isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, or sodium chloride. In other embodiments, the compositions comprise pharmaceutically

acceptable substances such as wetting agents or minor amounts of auxiliary substances such as wetting or emulsifying agents, preservatives or buffers, which enhance the shelf life or effectiveness of the active ingredients.

[0235] Compositions of the invention can be in a variety of forms, including, for example, liquid (e.g., injectable and infusible solutions), dispersions, suspensions, semisolid and solid dosage forms. The preferred form depends on the mode of administration and therapeutic application.

[0236] The composition can be formulated as a solution, micro emulsion, dispersion, liposome, or other ordered structure suitable to high drug concentration. Sterile injectable solutions can be prepared by incorporating the active ingredient in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active ingredient into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution. The proper fluidity of a solution can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prolonged absorption of injectable compositions can be brought about by including in the composition an agent that delays absorption, for example, monostearate salts and gelatin.

[0237] The active ingredient can be formulated with a controlled-release formulation or device. Examples of such formulations and devices include implants, transdermal patches, and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, for example, ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for the preparation of such formulations and devices are known in the art. See e.g., Sustained and Controlled Release Drug Delivery Systems, J. R. Robinson, ed., Marcel Dekker, Inc., New York, 1978.

[0238] Injectable depot formulations can be made by forming microencapsulated matrices of the drug in biodegradable polymers such as polylactide-polyglycolide. Depending on the ratio of drug to polymer, and the nature of the polymer employed, the rate of drug release can be controlled. Other exemplary biodegradable polymers are polyorthoesters and polyanhydrides. Depot injectable formulations also can be prepared by entrapping the drug in liposomes or microemulsions.

[0239] Supplementary active compounds can be incorporated into the compositions. In one embodiment, the chimeric protein is formulated with another clotting factor, or a variant, fragment, analogue, or derivative thereof. For example, the clotting factor includes, but is not limited to, factor V, factor VII, factor VIII, factor IX, factor X, factor XI, factor XII, factor XIII, prothrombin, fibrinogen, von Willebrand factor or recombinant soluble tissue factor (rsTF) or activated forms of any of the preceding. The clotting factor of hemostatic agent can also

include anti-fibrinolytic drugs, e.g., epsilon-amino-caproic acid, tranexamic acid.

[0240] Dosage regimens can be adjusted to provide the optimum desired response. For example, a single bolus can be administered, several divided doses can be administered over time, or the dose can be proportionally reduced or increased as indicated by the exigencies of the therapeutic situation. It is advantageous to formulate parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. See, e.g., Remington's Pharmaceutical Sciences (Mack Pub. Co., Easton, Pa. 1980).

[0241] In addition to the active compound, the liquid dosage form can contain inert ingredients such as water, ethyl alcohol, ethyl carbonate, ethyl acetate, benzyl alcohol, benzyl benzoate, propylene glycol, 1,3-butylene glycol, dimethylformamide, oils, glycerol, tetrahydrofurfuryl alcohol, polyethylene glycols, and fatty acid esters of sorbitan.

[0242] Non-limiting examples of suitable pharmaceutical carriers are also described in Remington's Pharmaceutical Sciences by E. W. Martin. Some examples of excipients include starch, glucose, lactose, sucrose, gelatin, malt, rice, flour, chalk, silica gel, sodium stearate, glycerol monostearate, talc, sodium chloride, dried skim milk, glycerol, propylene, glycol, water, ethanol, and the like. The composition can also contain pH buffering reagents, and wetting or emulsifying agents.

[0243] For oral administration, the pharmaceutical composition can take the form of tablets or capsules prepared by conventional means. The composition can also be prepared as a liquid for example a syrup or a suspension. The liquid can include suspending agents (e.g., sorbitol syrup, cellulose derivatives or hydrogenated edible fats), emulsifying agents (lecithin or acacia), non-aqueous vehicles (e.g., almond oil, oily esters, ethyl alcohol, or fractionated vegetable oils), and preservatives (e.g., methyl or propyl-p-hydroxybenzoates or sorbic acid). The preparations can also include flavoring, coloring and sweetening agents. Alternatively, the composition can be presented as a dry product for constitution with water or another suitable vehicle.

[0244] For buccal administration, the composition can take the form of tablets or lozenges according to conventional protocols.

[0245] For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of a nebulized aerosol with or without excipients or in the form of an aerosol spray from a pressurized pack or nebulizer, with optionally a propellant, e.g., dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoromethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit can be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, e.g., gelatin for use in an inhaler or insufflator can be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

[0246] The pharmaceutical composition can also be formulated for rectal administration as a suppository or retention enema, e.g., containing conventional suppository bases such as cocoa butter or other glycerides.

[0247] In one embodiment, a pharmaceutical composition comprises a FVIII protein, the optimized polynucleotide encoding the FVIII protein, the vector comprising the polynucleotide, or the host cell comprising the vector, and a pharmaceutically acceptable carrier. In some embodiments, the composition is administered by a route selected from the group consisting of topical administration, intraocular administration, parenteral administration, intrathecal administration, subdural administration and oral administration. The parenteral administration can be intravenous or subcutaneous administration,

[0248] In other embodiments, the composition is used to treat a bleeding disease or condition in a subject in need thereof. The bleeding disease or condition is selected from the group consisting of a bleeding coagulation disorder, hemarthrosis, muscle bleed, oral bleed, hemorrhage, hemorrhage into muscles, oral hemorrhage, trauma, trauma capitis, gastrointestinal bleeding, intracranial hemorrhage, intra-abdominal hemorrhage, intrathoracic hemorrhage, bone fracture, central nervous system bleeding, bleeding in the retropharyngeal space, bleeding in the retroperitoneal space, bleeding in the iliopsoas sheath and any combinations thereof. In still other embodiments, the subject is scheduled to undergo a surgery. In yet other embodiments, the treatment is prophylactic or on-demand.

Methods of Treatment

[0249] The invention provides a method of treating a bleeding disorder comprising administering to a subject in need thereof a nucleic acid molecule, or vector of the invention. In some embodiments, the bleeding disorder is characterized by a deficiency in Factor VIII. In some embodiments, the bleeding disorder is hemophilia. In some embodiments, the bleeding disorder is hemophilia A. In some embodiments of the method of treating a bleeding disorder, plasma Factor VIII activity at 24 hours post administration is increased relative to a subject administered a reference nucleic acid molecule comprising SEQ ID NO: 3, a vector comprising the reference nucleic acid molecule, or a polypeptide encoded by the reference nucleic acid molecule.

[0250] The invention also relates to a method of treating, ameliorating, or preventing a hemostatic disorder in a subject comprising administering a therapeutically effective amount of an isolated nucleic acid molecule of the invention. The treatment, amelioration, and prevention by the isolated nucleic acid molecule can be a bypass therapy. The subject receiving bypass therapy can have already developed an inhibitor to a clotting factor, e.g., Factor VIII, or is subject to developing a clotting factor inhibitor.

[0251] The nucleic acid molecules, or vectors of the invention treat or prevent a hemostatic disorder by promoting the formation of a fibrin clot. The FVIII protein can activate a member of

a coagulation cascade. The clotting factor can be a participant in the extrinsic pathway, the intrinsic pathway or both.

[0252] The nucleic acid molecules, or vectors of the invention can be used to treat hemostatic disorders known to be treatable with FVIII. The hemostatic disorders that can be treated using methods of the invention include, but are not limited to, hemophilia A, hemophilia B, von Willebrand's disease, Factor XI deficiency (PTA deficiency), Factor XII deficiency, as well as deficiencies or structural abnormalities in fibrinogen, prothrombin, Factor V, Factor VII, Factor X, or Factor XIII, hemarthrosis, muscle bleed, oral bleed, hemorrhage, hemorrhage into muscles, oral hemorrhage, trauma, trauma capitis, gastrointestinal bleeding, intracranial hemorrhage, intra-abdominal hemorrhage, intrathoracic hemorrhage, bone fracture, central nervous system bleeding, bleeding in the retropharyngeal space, bleeding in the retroperitoneal space, and bleeding in the iliopsoas sheath. Compositions for administration to a subject include nucleic acid molecules which comprise an optimized nucleotide sequence of the invention encoding a FVIII clotting factor (for gene therapy applications) as well as FVIII polypeptide molecules.

[0253] In some embodiments, the hemostatic disorder is an inherited disorder. In one embodiment, the subject has hemophilia A. In other embodiments, the hemostatic disorder is the result of a deficiency in Factor VIII. In other embodiments, the hemostatic disorder can be the result of a defective FVIII clotting factor.

[0254] In another embodiment, the hemostatic disorder can be an acquired disorder. The acquired disorder can result from an underlying secondary disease or condition. The unrelated condition can be, as an example, but not as a limitation, cancer, an autoimmune disease, or pregnancy. The acquired disorder can result from old age or from medication to treat an underlying secondary disorder (e.g. cancer chemotherapy).

[0255] The invention also relates to methods of treating a subject that does not have a hemostatic disorder or a secondary disease or condition resulting in acquisition of a hemostatic disorder. The invention thus relates to a method of treating a subject in need of a general hemostatic agent comprising administering a therapeutically effective amount of an isolated nucleic acid molecule of the invention. For example, in one embodiment, the subject in need of a general hemostatic agent is undergoing, or is about to undergo, surgery. An isolated nucleic acid molecule of the invention can be administered prior to or after surgery as a prophylactic. An isolated nucleic acid molecule of the invention can be administered during or after surgery to control an acute bleeding episode. The surgery can include, but is not limited to, liver transplantation, liver resection, or stem cell transplantation.

[0256] In another embodiment, an isolated nucleic acid molecule of the invention can be used to treat a subject having an acute bleeding episode who does not have a hemostatic disorder. The acute bleeding episode can result from severe trauma, e.g., surgery, an automobile accident, wound, laceration gun shot, or any other traumatic event resulting in uncontrolled bleeding.

[0257] The isolated nucleic acid molecules of the invention can be used to prophylactically treat a subject with a hemostatic disorder. The isolated nucleic acid molecules of the invention can be used to treat an acute bleeding episode in a subject with a hemostatic disorder.

[0258] A FVIII protein composition may be administered in combination with at least one other agent that promotes hemostasis. Said other agent that promotes hemostasis in a therapeutic with demonstrated clotting activity. As an example, but not as a limitation, the hemostatic agent can include Factor V, Factor VII, Factor IX, Factor X, Factor XI, Factor XII, Factor XIII, prothrombin, or fibrinogen or activated forms of any of the preceding. The clotting factor or hemostatic agent can also include anti-fibrinolytic drugs, e.g., epsilon-amino-caproic acid, tranexamic acid.

[0259] In one embodiment of the invention, the composition (e.g., the optimized nucleic acid molecule encoding the FVIII polypeptide) is one in which the FVIII is present in activatable form when administered to a subject. Such an activatable molecule can be activated *in vivo* at the site of clotting after administration to a subject.

[0260] The optimized nucleic acid molecule encoding the FVIII polypeptide can be administered intravenously, subcutaneously, intramuscularly, or via any mucosal surface, e.g., orally, sublingually, buccally, sublingually, nasally, rectally, vaginally or via pulmonary route.

[0261] For oral administration, the pharmaceutical composition can take the form of tablets or capsules prepared by conventional means. The composition can also be prepared as a liquid for example a syrup or a suspension. The liquid can include suspending agents (e.g. sorbitol syrup, cellulose derivatives or hydrogenated edible fats), emulsifying agents (lecithin or acacia), non-aqueous vehicles (e.g. almond oil, oily esters, ethyl alcohol, or fractionated vegetable oils), and preservatives (e.g. methyl or propyl-p-hydroxybenzoates or sorbic acid). The preparations can also include flavoring, coloring and sweetening agents. Alternatively, the composition can be presented as a dry product for constitution with water or another suitable vehicle.

[0262] For buccal and sublingual administration the composition can take the form of tablets, lozenges or fast dissolving films according to conventional protocols.

[0263] For administration by inhalation, the polypeptide having FVIII activity are conveniently delivered in the form of an aerosol spray from a pressurized pack or nebulizer (e.g. in PBS), with a suitable propellant, e.g., dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoromethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit can be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, e.g., gelatin for use in an inhaler or insufflator can be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

[0264] In one embodiment, the route of administration of the optimized nucleic acid molecule encoding the FVIII polypeptide is parenteral. The term parenteral as used herein includes intravenous, intraarterial, intraperitoneal, intramuscular, subcutaneous, rectal or vaginal administration. The intravenous form of parenteral administration is preferred. While all these forms of administration are clearly contemplated as being within the scope of the invention, a form for administration would be a solution for injection, in particular for intravenous or intraarterial injection or drip. Usually, a suitable pharmaceutical composition for injection can comprise a buffer (e.g. acetate, phosphate or citrate buffer), a surfactant (e.g. polysorbate), optionally a stabilizer agent (e.g. human albumin), etc. However, in other methods compatible with the teachings herein, the optimized nucleic acid molecules encoding the FVIII polypeptides can be delivered directly to the site of the adverse cellular population thereby increasing the exposure of the diseased tissue to the therapeutic agent.

[0265] Preparations for parenteral administration include sterile aqueous or non-aqueous solutions, suspensions, and emulsions. Examples of non-aqueous solvents are propylene glycol, polyethylene glycol, vegetable oils such as olive oil, and injectable organic esters such as ethyl oleate. Aqueous carriers include water, alcoholic/aqueous solutions, emulsions or suspensions, including saline and buffered media. In the subject invention, pharmaceutically acceptable carriers include, but are not limited to, 0.01-0.1M and preferably 0.05M phosphate buffer or 0.8% saline. Other common parenteral vehicles include sodium phosphate solutions, Ringer's dextrose, dextrose and sodium chloride, lactated Ringer's, or fixed oils. Intravenous vehicles include fluid and nutrient replenishers, electrolyte replenishers, such as those based on Ringer's dextrose, and the like. Preservatives and other additives can also be present such as for example, antimicrobials, antioxidants, chelating agents, and inert gases and the like.

[0266] More particularly, pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions. In such cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It should be stable under the conditions of manufacture and storage and will preferably be preserved against the contaminating action of microorganisms, such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (e.g., glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants.

[0267] Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols, such as mannitol, sorbitol, or sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

[0268] In any case, sterile injectable solutions can be prepared by incorporating an active compound (e.g., a polypeptide by itself or in combination with other active agents) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated herein, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle, which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying, which yields a powder of an active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof. The preparations for injections are processed, filled into containers such as ampoules, bags, bottles, syringes or vials, and sealed under aseptic conditions according to methods known in the art. Further, the preparations can be packaged and sold in the form of a kit. Such articles of manufacture will preferably have labels or package inserts indicating that the associated compositions are useful for treating a subject suffering from, or predisposed to clotting disorders.

[0269] The pharmaceutical composition can also be formulated for rectal administration as a suppository or retention enema, e.g., containing conventional suppository bases such as cocoa butter or other glycerides.

[0270] Effective doses of the compositions of the present invention, for the treatment of conditions vary depending upon many different factors, including means of administration, target site, physiological state of the patient, whether the patient is human or an animal, other medications administered, and whether treatment is prophylactic or therapeutic. Usually, the patient is a human but non-human mammals including transgenic mammals can also be treated. Treatment dosages can be titrated using routine methods known to those of skill in the art to optimize safety and efficacy.

[0271] Dosages can range from 1000 ug/kg to 0.1 ng/kg body weight. In one embodiment, the dosing range is 1 ug/kg to 100 ug/kg. The FVIII polypeptide or the optimized nucleic acid molecule encoding the FVIII polypeptide can be administered continuously or at specific timed intervals. *In vitro* assays can be employed to determine optimal dose ranges and/or schedules for administration. *In vitro* assays that measure clotting factor activity are known in the art. Additionally, effective doses can be extrapolated from dose-response curves obtained from animal models, e.g., a hemophiliac dog (Mount et al. 2002, Blood 99 (8); 2670).

[0272] Doses intermediate in the above ranges are also intended to be within the scope of the invention. Subjects can be administered such doses daily, on alternative days, weekly or according to any other schedule determined by empirical analysis. An exemplary treatment entails administration in multiple dosages over a prolonged period, for example, of at least six months. In some methods, two or more polypeptides can be administered simultaneously, in which case the dosage of each polypeptide administered falls within the ranges indicated.

[0273] FVIII polypeptides or the optimized nucleic acid molecules encoding the FVIII polypeptides can be administered on multiple occasions. Intervals between single dosages can be daily, weekly, monthly or yearly. Intervals can also be irregular as indicated by measuring blood levels of modified polypeptide or antigen in the patient. Alternatively, polypeptides can be administered as a sustained release formulation, in which case less frequent administration is required. Dosage and frequency vary depending on the half-life of the polypeptide or polynucleotide in the patient.

[0274] The dosage and frequency of administration can vary depending on whether the treatment is prophylactic or therapeutic. In prophylactic applications, compositions containing the FVIII polypeptide or the optimized nucleic acid molecule encoding the FVIII polypeptide or a cocktail thereof are administered to a patient not already in the disease state to enhance the patient's resistance or minimize effects of disease. Such an amount is defined to be a "prophylactic effective dose." A relatively low dosage is administered at relatively infrequent intervals over a long period of time. Some patients continue to receive treatment for the rest of their lives.

[0275] FVIII polypeptides or the optimized nucleic acid molecules encoding the FVIII polypeptides can optionally be administered in combination with other agents that are effective in treating the disorder or condition in need of treatment (e.g., prophylactic or therapeutic).

[0276] As used herein, the administration of FVIII polypeptides or the optimized nucleic acid molecules encoding the FVIII polypeptides in conjunction or combination with an adjunct therapy means the sequential, simultaneous, coextensive, concurrent, concomitant or contemporaneous administration or application of the therapy and the disclosed polypeptides. Those skilled in the art will appreciate that the administration or application of the various components of the combined therapeutic regimen can be timed to enhance the overall effectiveness of the treatment. A skilled artisan (e.g. a physician) would be readily be able to discern effective combined therapeutic regimens without undue experimentation based on the selected adjunct therapy and the teachings of the instant specification.

[0277] It will further be appreciated that the FVIII polypeptide or the optimized nucleic acid molecule encoding the FVIII polypeptide can be used in conjunction or combination with an agent or agents (e.g. to provide a combined therapeutic regimen). Exemplary agents with which a polypeptide or polynucleotide can be combined include agents that represent the current standard of care for a particular disorder being treated. Such agents can be chemical or biologic in nature. The term "biologic" or "biologic agent" refers to any pharmaceutically active agent made from living organisms and/or their products which is intended for use as a therapeutic.

[0278] The amount of agent to be used in combination with the polynucleotides or polypeptides can vary by subject or can be administered according to what is known in the art. See for example, Bruce A Chabner et al., Antineoplastic Agents, in GOODMAN & GILMAN'S THE PHARMACOLOGICAL BASIS OF THERAPEUTICS 1233-1287 ((Joel G. Hardman et al.,

eds., 9th ed. 1996). In another embodiment, an amount of such an agent consistent with the standard of care is administered.

[0279] As previously discussed, the polynucleotides and polypeptides can be administered in a pharmaceutically effective amount for the *in vivo* treatment of clotting disorders. In this regard, it will be appreciated that the polypeptides or polynucleotides can be formulated to facilitate administration and promote stability of the active agent. Preferably, pharmaceutical compositions in accordance with the present invention comprise a pharmaceutically acceptable, non-toxic, sterile carrier such as physiological saline, non-toxic buffers, preservatives and the like. Of course, the pharmaceutical compositions of the present invention can be administered in single or multiple doses to provide for a pharmaceutically effective amount of the polypeptide.

[0280] A number of tests are available to assess the function of the coagulation system: activated partial thromboplastin time (aPTT) test, chromogenic assay, ROTEM[®] assay, prothrombin time (PT) test (also used to determine INR), fibrinogen testing (often by the Clauss method), platelet count, platelet function testing (often by PFA-100), TCT, bleeding time, mixing test (whether an abnormality corrects if the patient's plasma is mixed with normal plasma), coagulation factor assays, antiphospholipid antibodies, D-dimer, genetic tests (e.g., factor V Leiden, prothrombin mutation G20210A), dilute Russell's viper venom time (dRVVT), miscellaneous platelet function tests, thromboelastography (TEG or Sonoclot), thromboelastometry (TEM[®], e.g., ROTEM[®]), or euglobulin lysis time (ELT).

[0281] The aPTT test is a performance indicator measuring the efficacy of both the "intrinsic" (also referred to the contact activation pathway) and the common coagulation pathways. This test is commonly used to measure clotting activity of commercially available recombinant clotting factors, e.g., FVIII or FIX. It is used in conjunction with prothrombin time (PT), which measures the extrinsic pathway.

[0282] ROTEM[®] analysis provides information on the whole kinetics of haemostasis: clotting time, clot formation, clot stability and lysis. The different parameters in thromboelastometry are dependent on the activity of the plasmatic coagulation system, platelet function, fibrinolysis, or many factors which influence these interactions. This assay can provide a complete view of secondary haemostasis.

Gene Therapy

[0283] The invention provides a method of increasing expression of a polypeptide with Factor VIII activity in a subject comprising administering the isolated nucleic acid molecule of the invention to a subject in need thereof, wherein the expression of the polypeptide is increased relative to a reference nucleic acid molecule comprising SEQ ID NO: 3. The invention also provides a method of increasing expression of a polypeptide with Factor VIII activity in a

subject comprising administering a vector of the invention to a subject in need thereof, wherein the expression of the polypeptide is increased relative to a vector comprising a reference nucleic acid molecule.

[0284] Somatic gene therapy has been explored as a possible treatment for hemophilia A. Gene therapy is a particularly appealing treatment for hemophilia because of its potential to cure the disease through continuous endogenous production of FVIII following a single administration of vector. Haemophilia A is well suited for a gene replacement approach because its clinical manifestations are entirely attributable to the lack of a single gene product (FVIII) that circulates in minute amounts (200ng/ml) in the plasma.

[0285] A FVIII protein can be produced *in vivo* in a mammal, e.g., a human patient, using a gene therapy approach to treatment of a bleeding disease or disorder selected from the group consisting of a bleeding coagulation disorder, hemarthrosis, muscle bleed, oral bleed, hemorrhage, hemorrhage into muscles, oral hemorrhage, trauma, trauma capitis, gastrointestinal bleeding, intracranial hemorrhage, intra-abdominal hemorrhage, intrathoracic hemorrhage, bone fracture, central nervous system bleeding, bleeding in the retropharyngeal space, bleeding in the retroperitoneal space, and bleeding in the iliopsoas sheath would be therapeutically beneficial. In one embodiment, the bleeding disease or disorder is hemophilia. In another embodiment, the bleeding disease or disorder is hemophilia A. This involves administration of an optimized FVIII encoding nucleic acid operably linked to suitable expression control sequences. In certain embodiment, these sequences are incorporated into a viral vector. Suitable viral vectors for such gene therapy include adenoviral vectors, lentiviral vectors, baculoviral vectors, Epstein Barr viral vectors, papovaviral vectors, vaccinia viral vectors, herpes simplex viral vectors, and adeno associated virus (AAV) vectors. The viral vector can be a replication-defective viral vector. In other embodiments, an adenoviral vector has a deletion in its E1 gene or E3 gene. When an adenoviral vector is used, the mammal can not be exposed to a nucleic acid encoding a selectable marker gene. In other embodiments, the sequences are incorporated into a non-viral vector known to those skilled in the art.

[0286] All of the various aspects, embodiments, and options described herein can be combined in any and all variations.

[0287] Having generally described this invention, a further understanding can be obtained by reference to the examples provided herein. These examples are for purposes of illustration only and are not intended to be limiting.

EXAMPLES

[0288] Two codon-optimized BDD FVIII sequences were designed with the following goals:

1. 1. Remove all matrix attachment-like region (MAR) sequences (ATATTT and AAATAT; SEQ ID NOs:5 and 6, respectively);

2. 2. Remove all destabilizing sequences (ATTTA, SEQ ID NO:8 and TAAAT, SEQ ID NO:9);
3. 3. Remove promoter binding sequences (TATAA, SEQ ID NO:12 and TTATA, SEQ ID NO:13);
4. 4. Remove AU-rich sequence elements (AREs): ATTTTATT (nucleotide 2468) and ATTTTTAA (nucleotide 3790) (SEQ ID NOs: 14 and 15, respectively);
5. 5. Add kozak sequence (GCCGCCACCATGC, underlined indicate the translation start codon; SEQ ID NO:16) to increase translational initiation;
6. 6. Adjust restriction sites to facilitate cloning;
7. 7. Adapt codon usage to the codon bias of homo sapiens genes;
8. 8. Adjust to avoid regions of very high (>70%) or low (<30%) GC content, which can increase RNA stability or prolong RNA half-life.

Example 1: Codon Optimization by GENSCRIPT OPTIMUMGENE™

[0289] The BDD FVIII nucleotide sequence was codon optimized using GENSCRIPT OPTIMUMGENE™ codon optimization technology. (GenScript Corp., New Jersey, USA). The GENSCRIPT OPTIMUMGENE™ codon optimization technology is described in Burgess-Brown et al., Protein Expr Purif. 59(1):94-102 (2008).

[0290] The following human codon usage data was used for optimization:

CODON	AMINO ACID	FRACTION	FREQUENCY/THOUSAND
TTT	F	0.45	16.9
TCT	S	0.18	14.6
TAT	Y	0.43	12.0
TGT	C	0.45	9.9
TTC	F	0.55	20.4
TCC	S	0.22	17.4
TAC	Y	0.57	15.6
TGC	C	0.55	12.2
TTA	L	0.07	7.2
TCA	S	0.15	11.7
TAA	*	0.28	0.7
TGA	*	0.52	1.3
TTG	L	0.13	12.6
TCG	S	0.06	4.5
TAG	*	0.20	0.5

CODON	AMINO ACID	FRACTION	FREQUENCY/THOUSAND
TGG	W	1.00	12.8
CTT	L	0.13	12.8
CCT	P	0.28	17.3
CAT	H	0.41	10.4
CGT	R	0.08	4.7
CTC	L	0.20	19.4
CCC	P	0.33	20.0
CAC	H	0.59	14.9
CGC	R	0.19	10.9
CTA	L	0.07	6.9
CCA	P	0.27	16.7
CAA	Q	0.25	11.8
CGA	R	0.11	6.3
CTG	L	0.41	40.3
CCG	P	0.11	7.0
CAG	Q	0.75	34.6
CGG	R	0.21	11.9
ATT	I	0.36	15.7
ACT	T	0.24	12.8
AAT	N	0.46	16.7
AGT	S	0.15	11.9
ATC	I	0.48	21.4
ACC	T	0.36	19.2
AAC	N	0.54	19.5
AGC	S	0.24	19.4
ATA	I	0.16	7.1
ACA	T	0.28	14.8
AAA	K	0.42	24.0
AGA	R	0.20	11.5
ATG	M	1.00	22.3
ACG	T	0.12	6.2
AAG	K	0.58	32.9
AGG	R	0.20	11.4
GTT	V	0.18	10.9
GCT	A	0.26	18.6

CODON	AMINO ACID	FRACTION	FREQUENCY/THOUSAND
GAT	D	0.46	22.3
GGT	G	0.16	10.8
GTC	V	0.24	14.6
GCC	A	0.40	28.5
GAC	D	0.54	26.0
GGC	G	0.34	22.8
GTA	V	0.11	7.0
GCA	A	0.23	16.0
GAA	E	0.42	29.0
GGA	G	0.25	16.3
GTG	V	0.47	28.9
GCG	A	0.11	7.6
GAG	E	0.58	40.8
GGG	G	0.25	16.4

[0291] Codon usage was adjusted to human bias with the human codon adaption index (CAI) changing from 0.75 (wild type BDD FVIII) to 0.88 (GenScript optimized BDD FVIII). G/C content was increased from 46.16% to 51.56%. Peaks of G/C content in a 60bp window were removed. The resulting sequence of GenScript optimized BDD FVIII is disclosed herein as SEQ ID NO:1 and is shown in Figure 2.

Example 2: Codon Optimization by GENEART® GENEOPTIMIZER®

[0292] The BDD FVIII nucleotide sequence was codon optimized using GENEART® GENEOPTIMIZER® software. (Invitrogen LIFE TECHNOLOGIES™ Corp., Grand Island, NY). The GENEART® GENEOPTIMIZER® codon optimization technology is described in Graf et al., J. Virol. 74(22):10822-10826 (2000).

[0293] Codon usage was adjusted to human bias with the human codon adaption index (CAI) changing from 0.75 (wild type BDD FVIII) to 0.96 (GeneArt optimized BDD FVIII). G/C content was increased from 46.16% to 59%. The resulting sequence of GeneArt optimized BDD FVIII is disclosed herein as SEQ ID NO:2 and is shown in Figure 3.

Example 3: Expression constructs

[0294] All constructs were made in the Invitrogen pcDNA™ 4 vector backbone, which contains a human cytomegalovirus immediate-early (CMV) promoter, a QBI SP163 translation enhancer, and a ZEOCIN™ resistance gene for selection.

[0295] pSYN-FVIII-066 drives expression of wild-type BDD FVIII (SEQ ID NO:3) in pcDNA4 backbone (Invitrogen).

[0296] pSYN-FVIII-116 drives expression of codon-optimized BDD FVIII (SEQ ID NO:1) in pcDNA4 backbone. The construct is derived from pSYN-FVIII-066 by replacing wild-type BDD FVIII with codon-optimized BDD FVIII (SEQ ID NO:1) using BsiWI and XhoI sites.

[0297] pSYN-FVIII-115 drives expression of codon-optimized BDD FVIII (SEQ ID NO:2) in pcDNA4 backbone. The construct is derived from pSYN-FVIII-066 by replacing wild-type BDD FVIII with codon-optimized BDD FVIII (SEQ ID NO:2) using BsiWI and XhoI sites.

[0298] All constructs were confirmed by DNA sequencing.

Example 4: Codon optimization improves FVIII expression in a Hema mouse

[0299] To ask whether either codon optimized BDD FVIII constructs result in increased FVIII protein expression, expression plasmids pSYN-FVIII116, pSYN-FVIII115, and wild type control pSYN-FVIII-066 were introduced into Hema mice via hydrodynamic injection. Subsequently, FVIII expression levels were monitored in each injected mouse by plasma FVIII chromogenic assays.

[0300] Hydrodynamic injection is an efficient and safe non-viral method to deliver genes to the liver in small animals, such as mice and rats. The protein of interest is produced in the liver and can be detected within 24 hours post-injection.

[0301] Hema mice weighing 20-35 grams were injected via intravenous tail vein injection with either pSYN-FVIII116, pSYN-FVIII115, or wild type control pSYN-FVIII-066. Injections were made up with 10ug naked plasmid DNA free of endotoxin in 0.9% sterile saline solution, to a total volume of 2ml. Injections were performed rapidly, taking no more than 4-7 seconds to inject the full 2ml DNA solution. Mice were closely monitored for two hours after injection, or until normal activity resumed. At 24 hours post-injection, samples were collected via retro orbital blood collection, plasma was prepared and stored at -80°C for further analysis.

[0302] The FVIII activity was measured using the COATEST SP FVIII kit from DiaPharma (lot# N089019) and all incubations were performed on a 37°C plate heater with shaking. rFVIII standards ranged from 100 mIU/mL to 0.78 mIU/mL. A pooled normal human plasma assay control and plasma samples (diluted with IX Coatest buffer) were added into Immulon 2HB 96-well plates in duplicate (25 µL/well). Freshly prepared IXa/FX/Phospholipid mix (50 µL), 25 µL

of 25mM CaCl₂, and 50 µL of FXa substrate were added sequentially into each well with a 5 minutes incubation between each addition. After incubating with the substrate, 25 µL of 20% Acetic Acid was added to terminate the color reaction, and the absorbance at OD405 was measured with a SpectraMAX plus (Molecular Devices) instrument. Data were analyzed with SoftMax Pro software (version 5.2). The Lowest Level of Quantification (LLOQ) is 7.8 mIU/mL. Results are shown in Figure 7.

[0303] Low levels of BDD FVIII activity is detected upon administration of FVIII expression plasmid pSYN-FVIII-066 (wild type BDD FVIII control). Average BDD FVIII activity in control mice is about 4-5 IU/mL (Figure 7, circles). In contrast, on average about a three-fold increase in BDD FVIII activity is seen in the plasma of mice administered codon optimized pSYN-FVIII115 or pSYN-FVIII116 (Figure 7, squares and triangles). Therefore, codon optimization of BDD FVIII by the approaches described above improve FVIII expression in a HemA mouse model.

SEQUENCES

[0304]

SEQ ID NO:1 – optimized BDD FVIII

```
CGTACGGCCGCCACCATGCAGATTGAGCTGTCTACTTGCTTTTTCCTGTGCCGTGCTGAGGTTTGTCTTTCCGCTACACG
AAGGTATTATCTGGGGGCTGTGGAAGTGTCTTGGGATTACATGCAGAGTGACCTGGGAGAGCTGCCAGTGGACGCAAGGT
TTCCCCCTAGAGTCCCTAAGTCATTCCCCCTTCAACACTAGCGTGGTCTACAAGAAAACACTGTTCTGTGGAGTTTACTGAT
CACCTGTTCACATCGCAAAGCCTAGGCCACCCTGGATGGGACTGCTGGGGCCAACAATCCAGGCCGAGGTGTACGACAC
CGTGGTCATTACACTTAAGAACATGGCCTCACACCCCGTGAGCCTGCATGCTGTGGGCGTCAGCTACTGGAAGGCTTCCG
AAGGAGCAGAGTATGACGATCAGACTTCCAGAGAGAAAAAGAGGACGATAAGGTGTTTCCCTGGCGGATCTCATACCTAC
GTGTGGCAGGTCTGAAAGAGAATGGCCCTATGGCCCTCCGACCCCTCTGTGCCTGACCTACTCTTATCTGAGTCACGTGGA
CCTGGTCAAGGATCTGAACAGCGGCTGATCGGAGCCCTGCTGGTGTGCAGGAAGGAAGCCTGGCTAAGGAGAAAACCC
AGACACTGCATAAGTTCAATTCTGCTGTTTCGCCGTGTTTGACGAAGGGAATCATGGCACAGCGAGACAAAGAATAGTCTG
ATGCAGGACAGGGATGCCGCTTCAGCCAGAGCTTGGCCCAAAATGCACACTGTGAACGGCTACGTCAATCGCTCACTGCC
TGGGCTGATCGGCTGCCACCGAAAGAGCGTGTATTGGCATGTTCATCGGGATGGGCACCACACCTGAAGTGCATCCATTT
TCCTGGAGGGACATACCTTTCTGGTCCGCAACCACCGACAGGCTTCCCTGGAGATCTCTCCAATTACCTTCCTGACAGCA
CAGACTCTGCTGATGGACCTGGGGCAGTTCCTGCTGTTTTCGCCACATCAGCTCCCACCAGCATGATGGCATGGAGGCTTA
CGTGAAGTGGACTCTGTCTCCGAGGAACCTCAGCTGCGGATGAAGAACAATGAGGAAGCAGAAGACTATGACGATGACC
TGACCGACTCCGAGATGGATGTGGTCCGATTGATGACGATAACAGCCCTCCTTTATCCAGATTAGATCTGTGSCCAAG
AAACACCCTAAGACATGGGTCCATTACATCGCAGCCGAGGAAGAGGACTGGGATTATGCACCACTGGTGTGGCACCAGA
CGATCGCTCCTACAAATCTCAGTATCTGAACAATGGGCCACAGAGGATTGGCAGAAAGTACAAGAAAGTCCGGTTCATGG
CATATACCGATGAGACCTTCAAGACTCGCGAAGCCATCCAGCAGGAGCGGCATCCTGGGACCACTGCCTGACCGAGAA
GTGGGAGACACCCTGCTGATCATTTTCAAGAACCAGGCCAGCCGGCCTTACAATATCTATCCACATGGGATTACAGATGT
GCGCCCTCTGTACAGCAGGAGACTGCCAAGGGCGTCAAACACCTGAAGGACTTCCCAATCCTGCCCGGAGAAATCTTCA
AGTACAAGTGGACTGTACCCGTGAGGATGGCCCCACTAAGAGCGACCCCTCGGTGCCTGACCCGCTACTATTCTAGTTTC
GTGAATATGGAAGAGATCTGGCAAGCGGACTGATCGGACCACTGCTGATTGTGTTACAAAGAGAGCGTGGATCAGAGAGG
CAACCAGATCATGTCCGACAAGCGGAATGTGATTCTGTTTCAGTGTCTTTGACGAAAACAGGTCATGGTACCTGACCGAGA
ACATCCAGAGATTCCTGCCTAATCCAGCTGGGGTGACGCTGGAAGATCCTGAGTTTCAGGCATCTAACATCATGCATAGT
ATTAATGGCTACGTGTTTCGACAGTTTGCAGCTGAGCGTGTGCCTGCACGAGGTCGCTTACTGGTATATCCTGAGCATTGG
GGCACAGACAGATTTCTGAGCGTGTCTTTTCGGCTACACTTTTAAGCATAAAATGCTCTATGAGGACACACTGACTC
```

TGTTCCCTTCAGCGCGAAACCGTGTATTATGAGCATGGAGAATCCCGGACTGTGGATTCTGGGGTGCCACAACAGCGAT
 TTCAGAAATCGCGSAATGACTGCCCTGCTGAAAGTGTCAAGCTGTGACAAGAACACCGGGGACTACTATGAAGATTCTATA
 CGAGGACATCAGCGCATATCTGCTGTCCAAAAACAATGCCATTGAACCCCGGTCTTTTAGTCAGAAATCCTCCAGTGCTGA
 AGCGGCACAGCGCGAGATCACCCGCACTACCTGCAGAGTGATCAGGAAGAGATCGACTACGACGATACAATTTCTGTG
 GAAATGAAGAAAGAGGACTTCGATATCTATGACGAAGATGAGAACCAGAGTCCTCGATCATTCCAGAAGAAAACAGGGCA
 TTACTTTTATTGCCGAGTGAGCGGCTGTGGGATTATGGCATGTCTCTAGTCCTCAGTGCTGCGAAATAGGGCCAGT
 CAGGAAGCGTCCCACAGTTCAAGAAAGTGGTCTTCCAGGAGTTTACAGACGGGTCTTTACTCAGCCACTGTACAGGGGC
 GAACTGAACGAGCACCTGGGACTGCTGGGGCCCTATATCAGAGCAGAAGTGGAGGATAACATTATGGTCACCTTCAGAAA
 TCAGGCCTCTCGGCCTTACAGTTTTTATTCAAGCCTGATCTCTTACGAAGAGGACCAGCGACAGGGAGCTGAACACAGAA
 AAAACTTCGTGAAGCCTAATGAGACCAAAACATACTTTTGAAGGTGCAGCACCATATGGCCCCAACAAAAGACGAGTTT
 GATTGCAAGGCATGGGCCTATTTTTCTGACGTGGATCTGGAGAAGGACGTGCACAGTGGCCTGATTGGCCCACTGTGGT
 GTGCCATACTAACACCTGAATCCAGCCCACGGCCGCGAGGTCACTGTCCAGGAGTTCGCTCTGTTCTTTACCATCTTTG
 ATGAGACAAAGAGCTGGTACTTCACCGAAAACATGGAGCGAAATTCAGGGGCTCCATGTAACATTGAGTGAAGACCCC
 ACATTCAAGGAGAACTACCGCTTTTCATGCTATCAATGGATACATCATGGATACTCTGCCCGGGCTGGTCATGGCACAGGA
 CCAGAGAATCCGGTGGTATCTGCTGAGCATGGGCAGCAACGAGAATATCCACTCAATTCATTTCAGCGGGCACGTGTTTA
 CTGTCAGGAAGAAAGAGATACAAGATGGCCCTGTACAACCTGTATCCCGGCGTGTTCGAAACCGTCGAGATGCTGCCT
 AGCAAGCGCGGAATCTGGAGAGTGAATGCCTGATTGGAGAGCACCTGCATGCTGGGATGTCTACCCTGTTCTGTGTGTA
 CAGTAATAAGTGTGACACCCCTGGGAATGGCATCCGGGCATATCAGGGATTTCAGATTACCGCATCTGGACAGTACG
 GACAGTGGGCACCTAAGCTGGCTAGACTGCACTATTCCGGATCTATCAACGCTTGGTCCACAAAAGAGCCTTTCTCTTGG
 ATTAAGGTGGACCTGCTGGCCCCAATGATCATTGATGCGCATCAAAACTCAGGGAGCTCGGCAGAAAGTTCTCTCTCTGTGA
 CATCTCACAGTTTATCATCATGTACAGCCTGGATGGGAAGAAATGGCAGACATACCGCGGCAATAGCACAGGAACCTCTGA
 TGGTGTCTTTGGCAACGTGGACAGCAGCGGAATCAAGCACAACTTTTCAATCCCCCTATCATTGCTAGATACATCCGG
 CTGCACCCAAACCAATTATTCTATTCCAAGTACACTGAGGATGGAACCTGATGGGATGCGATCTGAACAGTTGTTCAATGCC
 CCTGGGGATGGAGTCCAAGGCAATCTCTGACGCCCAGATTACCGCTAGCTCCTACTTCACTAATATGTTTGCTACCTGGA
 GCCCTTCCAAAGCAAGACTGCACCTGCAAGGCCGAGCAACGCGATGGCGACACAGGTGAACAATCCCAAGGAGTGGTTG
 CAGGTGCGATTTTCAGAAAATATGAAGGTGACCGGGGTCACAACTCAGGGCGTGAAAAGTCTGCTGACCTCAATGTACGT
 CAAGGAGTTCCTGATCTCTAGTTTACAGGACGGACATCAGTGGACACTGTTCTTTTCAAGACGGGAAGGTGAAAGTCTTCC
 AGGGCAATCAGGATTCTTTTACACCTGTGGTCAACAGTCTAGACCCTCCACTGCTGACCAGATACCTGAGAATCCACCCT
 CAGTCTGGGTGCACCAGATTGCCCTGAGAATGGAAGTGTGGGATGCGAGGCCCAGGATCTGTACTGATAACTCGAGTC
 GACC

SEQ ID NO:2 – optimized BDD FVIII

GGCGCGCCCGTACGGCCGCCACCATGCGAGTCGAGCTGTCTACCTGCTTCTTCTGTGCCTGCTGCGGTTCTGCTTCAGC
 GCCACCCGGCGGTACTACCTGGGCGCGTGGAACTGAGCTGGGACTACATGCGAGCGACCTGGGGGAGCTGCCCGTGA
 CGCCAGATTCCCCCAAGAGTGCCCAAGAGCTTCCCTTCAACACCTCCGTGGTGTACAAGAAAACCTGTTCTGTCAGT
 TCACCGACCACTGTTCAATATCGCCAAGCCAGACCCCTGGATGGGCTGCTGGGCCCTACAAATCCAGGCCGAGGTG
 TACGACACCGTGGTCATCACCTTAAGAACATGCCAGCCACCCCGTGTCCCTGCACGCCGTGGGCGTGTCTACTGGAA
 GGCTCTGAGGGCGCTGAGTACGACGACGACGACGACGCGCGAGAAAGAGGACGACAAAGTCTTTCTGGCGGCAGCC
 ATACCTACGTGTGGCAGGTCTCTGAAAGAAAACGGCCCTATGGCCTCCGACCCCTGTGCCTGACCTACAGCTACCTGAGC
 CAGGTGGAACCTGTTCAAGGACCTGAACAGCGGCCTGATTGGCGCCTGCTCGTGTGTAGAGAGGGCAGCCTCGCCAAAGA
 GAAAACCCAGACCTGCAAGTTTATCCTGCTGTTCGCCGTGTTTCGACGAGGGCAAGAGCTGGCAGCAGGACAAAGA
 ACAGCCTGATGACGACCGGGACGCGCCTCTGCCAGAGCCTGGCCTAAGATGCACACCGTGAACGGCTACGTGAACAGA
 AGCCTGCCCGGACTGATCGGCTGCCACCGGAAGTCCGTGTAAGTGGCACGTGATCGGCATGGGCACCAACCCCGAGGTGCA

CAGCATCTTTCTGGAAGGCCACACCTTCTCTGTGCGGAACACAGACAGGCCAGCCTGGAATCAGCCCTATCACCTTCC
 TGACCGCCAGACACTGCTGATGGACCTGGGCCAGTTCTCTGCTGTTTTGCCACATCAGCAGCCACCAGCAGCAGCGCATG
 GAAGCCTACGTGAAGTGGACAGCTGCCCGAGGAACCCAGCTGCGGATGAAGAACAAAGAGGAAGCCGAGGACTACGA
 CGACGACCTGACCGACAGCGAGATGACGCTGCTGCGCTTCGACGACGACAACAGCCCGAGCTTATCCAGATCACAAGCG
 TGGCCAAACAAGCACCCCAAGACCTGCCCTGCATATATCGCCCCGAGGAACAGCACTGCCACTACGCCCTCTGGTCTGTG
 GCCCGGACGACAGAAGCTACAAGAGCCAGTACCTGAACAATGGCCCCAGCGGATCGGCCGGAAGTACAAGAAAGTGGC

GTTCATGGCCTACACCGACGAGACATTCAAGACCAGAGAGGCCATCCAGCACGAGAGCGGCATCCTGGGCCCCCTGCTGT
 ATGGCGAAGTGGGCGACACCTGCTGATCATCTTCAAGAACAGGCCAGCCGGCCCTACAACATCTACCCCCACGGCATC
 ACCGACGTGGGCCCCCTGTACAGCAGACGGGTGCCAAGGGCGTGAAGCACCTGAAGGACTTCCCCTCCTGCCCGGCGA
 GATCTTCAAGTACAAGTGGACCGTGACCGTGGAGATGGCCCCACCAAGAGCGACCCAGATGCCTGACCCGGTACTACA
 GCAGCTTCGTGAACATGGAACGGGACCTGGCCTCCGGGTGTATCGGCCCTCTGCTGATCTGCTACAAAGAAAGCGTGGAC
 CAGCGGGGCAACCAGATCATGAGCGACAAGCGGAACGTGATCCTGTTGAGCGTGTTCGATGAGAATCGGTCCCTGGTACCT
 GACCGAGAATATCCAGCGGTTCCTGCCCCAACCTGCCGGGTGCAGCTGGAAGATCCCGAGTTCAGGCCAGCAACATCA
 TGCATCCATCAATGGCTACGTGTTGCACAGCCTCCAGCTGAGCGTGTGCCTGCACGAGGTGGCCTACTGGTACATCCTG
 AGCATCGGCGCCAGACCGACTTCTGAGCGTGTCTTTCAGCGGTACACCTTCAAGCACAAGATGGTGTACGAGGATAC
 CCTGACCTCTGTCCCTTCTCGGCGCAACCGTGTTCATGAGCATGGAAAACCCGGCCTGTGGATTCTGGGCTGCCACA
 ACAGCGACTTCAGAAAACCGGGCATGACCGCCTGCTGAAGGTGTCCAGCTGCGACAAGAACACCGGCGACTACTACGAG
 GACAGCTATGAGGACATCAGCCTACCTGCTGAGCAAGAACACGCCATCGAGCCAGATCCTTCAGCCAGAACCCCCC
 CGTGCTGAAGCGGCAACCAGAGAGAGATCACCCGACACCTGCGAGTCCGACCAGGAAGAGATTGATTACGACGACACCA
 TCAGCGTCGAGATGAAGAAAGAGGATTCGACATCTACGACGAGGACGAGAACCAGAGCCCCCGGTCTTCCAGAAGAAA
 ACCCGGCACTACTTCATTGCGCGCGTGGAAAGACTGTGGACTACGGCATGAGCAGCAGCCCCACGTGCTGCCGAACAG
 AGCCAGAGCGGCGAGCTGCCCGAGTTCAAGAAAGTGGTGTTCAGGAGTTTACCGACGGCAGCTTACCCAGCCCCGT
 ATCGGGCGAGCTGAACGAGCACCTGGGACTCTGGGACCTTACATTAGAGCCGAGCTCGAAGATAACATCATCGCTCACC
 TTCAGAAAACCGGCTCCAGACCTACAGCTTCTACAGCAGCCTGATCAGCTACGAAGAGGACAGCGCGAGGGCGCCGA
 ACCCGGAAGAACTTCGTGAAGCCCAACGAGACTAAGACCTACTTCTGGAAGGTGCAGCACACATGGCCCCACAAGG
 ACGAGTTCGACTGCAAGGCTGGGCTACTTCTCCGATGTTGGACCTGGAAAAGGACGTGCATCTGGGCTGATTGGACCT
 CTGCTCGTCTGCCACACCAACACCTGAACCCCGCCACGGCCGGCAGGTACAGTGCAGGAATTTGCCCTGTCTTCAC
 CATCTTCGATGAGACAAAGAGCTGGTACTTACCCGAGAATGGAAGAAACTGTAGAGCCCCCTGCAACATCCAGATGG
 AAGATCCTACCTTCAAAGAGAACTATCGGTTCACGCCATCAACGGCTACATCATGGACACCTGCCCCGGCCTGGTCATG
 GCCCAGGATCAGAGAATCGGTGCTATCTGCTGAGCATGGGCAGCAACGAGAATCCACAGCATCCACTTCAGCGGCCA
 CGTGTTCAGTTCGGAAGAAAGAGTACAAGATGGCCCTGTACAACCTGTACCCCGGCGTGTTCGAGACAGTGGAAA
 TGCTGCCAGCAAGGCCGGCATCTGGCGGTGGAATGTCTGATCGGCGAGCATCTGCACGCGGAATGAGCACCCGTGTTT
 CTGGTGTACAGCAACAAGTGGCAGACCCCTCTGGGCATGGCCAGCGGCCACATCCGGGACTTCAGATCACCGCTCCGG
 CCAGTACGGCAGTGGGCCCTAAGCTGGCCCGGCTCCACTACTCCGGATCTATCAACGCCCTGGTCCACCAAAGAGCCCT
 TCAGCTGGATCAAGGTGGACCTGCTGGCCCTATGATCATCCACGGAATCAAGACCCAGGGCGCCAGACAGAAGTTCAGC
 AGCCTGTACATCAGCCAGTTCATCATCATGTACAGCCTGGACGGCAAGAAGTGGCAGACCTACCGGGGCAACAGCACCGG
 CACCCTGATGGTGTCTTTCGGCAACGTGGACAGCAGCGGCATCAAGCACAACATCTTCAACCCCCCATCATTCGCCGGT
 ACATCCGGCTGCACCCACCCACTACAGCATCCGGTCCACCTGCGGATGGAAGTGTGGGCTGCGACCTGAAGTCTTGC
 AGCATGCCCTCGGGATGGAAAGCAAGGCCATCAGCGACGCCAGATCACAGCCAGCAGTACTTCACCAACATGTTCCGC
 CACCTGGTCCCAAGCAAGCCCGCTGCATCTCCAAGGCAGAAGCAATGCCTGGCGGCTCAGGTCAACAACCCCAAG
 AATGGCTCCAGGTGGACTTTCAGAAAACCATGAAGGTCAAGCGGTGACCAACCCAGGGCGTGAAGAGCTGCTGACCTCT
 ATGTACGTGAAGAGTTCTGTATCAGCAGCAGCCAGGACGGGCAACAGTGGACCTGTCTTTTCAAGCGGCAAGTGA
 AGTGTTCAGGGCAACAGGACTCTTTTACCCCCGTGGTCAACTCTCTAGACCTCCACTGCTGACCAGATACCTGAGAA

TCCACCCCTAGTCTGGGTGCACCAGATGGCCCTGAGAATGGAAGTGTGGGATGCGAGGCCAGGATCTGTACTGATAA
 CTCGAGTCGACTTAATTAA

SEQ ID NO:3 – Nucleotide Sequence Encoding "Parental" BDD FVIII

ATGCAAAATAGAGCTCTCCACCTGCTTCTTCTGTGCCTTTTGGGATTCTGCTTTAGTGCCACCAGAAGATACTACCTGGG
 TGCAGTGGAACTGTATGGGACTATATGCAAAAGTATCTCGGTGAGCTGCCTGTGGACGCAAGATTTCCCTCCTAGAGTGC
 CAAAATCTTTTCCATTCAACACCTCAGTCGTGTACAAAAGACTCTGTTTGTAGAAATCACGGATCACCTTTTCAACATC
 GCTAAGCCAAGGCCACCTGGATGGGTCTGCTAGGTCTACCATCCAGGCTGAGGTATATGATACAGTGGTCATTACACT
 TAAGAATATGGCTTCCCATCTGTGAGTCTTCATGCTGTTGGTGTATCTACTGGAAAGCTTCTGAGGGAGCTGAATATG
 ATGATCAGACCACTCAAAGGGAGAAAGAGATGATAAAGTCTTCCCTGGTGGAAAGCCATACATATGTCTGGCAGGTCTTG
 AAAGAGAATGGTCCATGGCCTCTGACCCACTGTGCCTTACCTACTCATATCTTTCTCATGTGGACCTGGTAAAGACTT
 GAATTCAAGGCTCATTGGAGCCCTACTAGTATGTAGAGAAGGAGTCTGGCCAAGGAAAAGACACAGACCTTGCAAAAT
 TTATACTACTTTTGTGTATTGATGAAGGGAAAAGTTGGCACTCAGAAAACAAAGAACTCCTTGATGCAGGATAGGGAT

GCTGCATCTGCTCGGGCCTGGCCCTAAATGCACACAGTCAATGGTTATGTAAACAGGTCTCTGCCAGGTCTGATTGGATG
 CCACAGGAAATCAGTCTATTGGCATGTGATTGGAATGGGCACCACTCCTGAAGTGCACCTCAATATTCCTCGAAGGTCACA
 CATTCTTGTGAGGAACCATCGCCAGGCGTCTTGGAAATCTCGCCAATAACTTTCCTTACTGCTCAAACACTCTTGATG
 GACCTTGGACAGTTTCTACTGTTTTGTTCATATCTCTTCCCACCAACATGATGGCATGGAAGCTTATGTCAAAGTAGACAG
 CTGTCCAGAGGAACCCCACTACGAATGAAAAATAATGAAGAAGCGGAAGACTATGATGATGATCTTACTGATTCTGAAA
 TGGATGTGGTCAGGTTTGATGATGACAACTCTCCTTCCTTTATCCAAATTCGCTCAGTTGCCAAGAAGCATCCTAAAACT
 TGGGTACATTACATTGCTGCTGAAGAGGAGGACTGGGACTATGCTCCCTTAGTCCCTCGCCCCGATGACAGAAGTTATAA
 AAGTCAATATTTGAACATGGCCCTCAGCGGATTGGTAGGAAGTACAAAAAGTCCGATTTATGGCATACACAGATGAAA
 CCTTTAAGACTCCTGAAGCTATTGAGCATGAATCAGGAATCTGGGACCTTTACTTTATGGGAAGTTGGACACACTG
 TTGATTATATTTAAGAATCAAGCAAGCAGACCATATAACATCTACCCCTCAGGAATCACTGATGTCCGTCCTTTGTATTC
 AAGGAGATTACCAAAAGGTGTAAACATTGAAGGATTTTCCAATTCGCCAGGAGAAATATTCAAATATAAATGGACAG
 TGACTGTAGAAGATGGGCCAACTAAATCAGATCCTCGGTGCTGACCCGCTATTACTCTAGTTTCGTAAATATGGAGAGA
 GATCTAGCTTCAGGACTCATTGGCCCTCTCCTCATCTGCTACAAAGAATCTGTAGATCAAAGAGGAAACCAGATAATGTC
 AGACAAGAGGAATGTCATCCTGTTTTCTGTATTTGATGAGAACCAGAGCTGGTACCTCACAGAGAATATACAACGCTTTC
 TCCCCAATCCAGCTGGAGTGCAGCTTGAGGATCCAGAGTTCCAAGCCTCCAACATCATGCACAGCATCAATGGCTATGTT
 TTTGATAGTTTGCAGTTGTGAGTTGTTGTCATGAGGTGGCATACTGGTACATTCTAAGCATTGGAGCACAGACTGACTT
 CCTTCTGTCTTCTCTCTGGATATACCTTCAAACACAAAATGGTCTATGAAGACACACTCACCCTATCCCATTCTCAG
 GAGAACTGTCTTCATCTCGATGGAAAACCCAGGTCTATGGATTCTGGGGTGGCACAACTCAGACTTTCGGAACAGAGGC
 ATGACCGCCTTACTGAAGTTTCTAGTTGTGACAAGAACACTGGTGATTATTACGAGGACAGTTATGAAGATATTTCAGC
 ATACTTGCTGAGTAAAAACAATGCCATTGAACCAAGAAGCTTCTCTCAAAACCCACCAGTCTTGAAACGCCATCAACGGG
 AAATAACTCCTACTACTCTTCAGTCAGATCAAGAGGAAATGACTATGATGATACCATATCAGTTGAAATGAAGAAGGAA
 GATTTTGACATTTATGATGAGGATGAAAAATCAGAGCCCCCGAGCTTTCAAAAGAAACACGACACTATTTTATGCTGC
 AGTGGAGAGGCTCTGGGATTATGGGATGAGTAGCTCCCCACATGTTCTAAGAAACAGGGCTCAGAGTGGCAGTGTCCCTC
 AGTTCAAGAAAGTTGTTTTCCAGGAATTTACTGATGGCTCCTTTACTCAGCCCTTATACCGTGGAGAACTAAATGAACAT
 TTGGGACTCCTGGGGCCATATATAAGAGCAGAAGTTGAAGATAATATCATGTTAACTTTCAGAAATCAGGCCTCTCGTCC
 CTATTCCTTCTATTCTAGCCTTATTTCTTATGAGGAAGATCAGAGGCAAGGAGCAGAACCTAGAAAAAACTTTGTCAAGC
 CTAAATGAAACCAAACTTACTTTTGGAAAGTGCAACATCATATGGCACCCACTAAAGATGAGTTTGACTGCAAGCCTGG
 GCTTATTTCTCTGATGTTGACCTGGAAAAAGATGTGCACTCAGGCCTGATTGGACCCCTTCTGCTGTGCCACACTAACAC
 ACTGAACCCCTGCTCATGGGAGACAAGTGACAGTACAGGAATTTGCTCTGTTTTTACCATCTTTGATGAGACCAAAAGCT
 GGTACTTCACTGAAAAATATGGAAAGAACTCAGGGCTCCCTGCAATATCCAGATGGAAGATCCCACCTTTTAAAGAGAAT

TATCGCTTCCATGCAATCAATGGCTACATAATGGATACACTACCTGGCTTAGTAATGGCTCAGGATCAAAGGATTCGATG
 GTATCTGCTCAGCATGGGCAGCAATGAAAAATCCATTCTATTCTATTTCAGTGGACATGTGTTCACTGTACGAAAAAAG
 AGGAGTATAAAATGGCACTGTACAACTCTATCCAGGTGTTTTTGAGACAGTGGAAATGTTACCATCCAAAGCTGGAATT
 TGGCGGGTGAATGCCTTATTGGCGAGCATCTACATGCTGGGATGAGCACACTTTTCTGGTGTACAGCAATAAGTGTCA
 GACTCCCCTGGGAATGGCTTCTGGACACATTAGAGATTTTCAGATTACAGCTTCAGGACAAATATGGACAGTGGGCCCCAA
 AGCTGGCCAGACTTCATTATTCGGATCAATCAATGCCTGGAGCACCAAGGAGCCCTTTTCTTGATCAAGGTGGATCTG
 TTGGCACCATGATTATTCACGGCATCAAGACCAGGTTGCCGTCAGAAGTTCTCCAGCCTCTACATCTCTCAGTTTAT
 CATCATGTATAGTCTTGATGGGAAGAAGTGGCAGACTTATCGAGGAAATCCACTGGAACCTTAATGGTCTTCTTTGGCA
 ATGTGGATTCTCTGGGATAAAACACAATATTTTTAACCCCTCAATTATTGCTCGATACATCCGTTTGACCCCAACTCAT
 TATAGCATTGCGAGCACTCTTCGATGGAGTTGATGGGCTGTGATTTAAATAGTTGCAGCATGCCATTGGGAATGGAGAG
 TAAAGCAATATCAGATGCACAGATTACTGCTTCATCCTACTTTACCAATATGTTTGCCACCTGGTCTCCTTCAAAAGCTC
 GACTTCACCTCCAAGGGAGGAGTAATGCCTGGAGACCTCAGGTGAATAATCCAAAAGAGTGGCTGCAAGTGGACTTCCAG
 AAGACAATGAAAGTCACAGGAGTAACTACTCAGGGAGTAAATCTCTGCTTACCAGCATGTATGTGAAGGAGTTCCCTCAT
 CTCACAGCAGTCAAGATGGCCATCAGTGGACTCTCTTTTTCAGAAATGGCAAGTAAAGGTTTTTCAGGGAAATCAAGACT
 CCTTCACACCTGTGTTGAACTCTCTAGACCCACCGTTACTGACTCGCTACCTTCGAATTCACCCCCAGAGTTGGGTGCAC
 CAGATTGCCCTGAGGATGGAGTTCTGGGCTGCGAGGCACAGGACCTCTAC

SEQ ID NO:4 – Amino Acid Sequence of BDD FVIII

ATRRYYLGAVELSNDYMQSDLGELPVDARFPFPRVPKSFNFNTSVVYKKTLEFVEFTDHLFNIKPRPPWMGLLGPTIQAEV
 YDTVVIILKNMASHPVSLHAVGVSYWKASEGAEYDDQTSQREKEDDKVFPGGSHTYVWQVLEKENGPMASDPLCLITYSYLS

HVDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNR
 SLFGLIGCHRKSVYWHVIGMGTTPDEVHSIFLEGHTFLVRNHRQASLEISPTIFLTAQTLMDLGQFLIFCHISSHQHDGM
 EAYVKVDSCPEEPQLRMKNNEEAEDYDEDLTDEMDVVRFDNNSPSFIQIRSVAKKHPKTWVHYIAAEFEDWDYAPLVL
 APDDRSYKSYLNNGPQRIGRKYKKVRFMAYTDETFKTRAIQHESGILGPLLYGEVGDILLIIFKNQASRPYNIYPHGI
 TDVRPLYSRRLPKGKVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRYYSFVNMERDLASGLIGELLICYKESVD
 QRGNQ_MSDKRNVLFSVFDENRSWYLTENIQRFILPNPAGVQLDPEFQASNIMHSINGYVFDLSQLSVCLHEVAYWYIL
 SIGAQDFLSVFFSGYTFKHKMVEDTTLTFPFSGETVFMSENPGLWILGCHNSDFRNRGMTALLKVSSCDKNTGDYYE
 DSYEDISAYLLSKNNAIEPRSFQNPVPLKRHQREITRTLQSDQEEIDYDDTISVEMKKEDFDIYDEDENQSPRSFQKK
 TRHYFCAVERLWDYGMSSSPHVLNRNAQSGSVPPQFKKVVFEFTDGSFTQPLYRGELNEHLGLGPYIRAEVEDNIMVT
 FRNQASRPYSFYSSLISYEEDQRQGAEPFRKNFVKPNETKTYFWKVQHMAPTKDEFCKAWAYFSDVDLEKDVHSGLIGP
 LLVCHTNTLNPAHGRQVTVQEFALFTTIFDETKSWYFTENMERNCRAPCNIQMEDPTFKENYRFHAINGYIMDTLPGLVM
 AQDQIRIRWYLLSMGSNENIHSIHSGHVFTVRKKEEYKMALYNLYPGVFETVEMLPSKAGIWRVECLIGEHLHAGMSTLF
 LVYSNKCQTPLGMA SGHIRDFQITASGQYGQWAPKLARLHYSGSINAWSTKEPESWIKVDLLAPMIHGTQCARQKFS
 SLYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKHNIFNPPIIARYIRLHPHYSIRSTLRMELMGCDLNSC
 SMPLGMESKAISDAQITASSYFTNMFPATWSPSKARLHLQGRSNAWRPQVNNPKEWLQVDFQKTMKVTGVTTQGVKSLITS
 MYVKEFLISSSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVNSLDPPLLTRYLRHPQSWVHQIALRMEVLGCEAQDLY

SEQ ID NO:5 - MAR/ARS nucleotide sequence

ATATTT

SEQ ID NO:6 - MAR/ARS nucleotide sequence

AAATAT

SEQ ID NO:7 - Potential Splice Site

GGTGAT

SEQ ID NO:8 - Destabilizing Sequence

ATTTA

SEQ ID NO:9 - Destabilizing Sequence

TAAAT

SEQ ID NO:10 - poly-T Sequence

TTTTTT

SEQ ID NO:11 - poly-A Sequence

AAAAAAA

SEQ ID NO:12 - Promoter Binding Site

TATAA

SEQ ID NO:13 - Promoter Binding Site

TTATA

SEQ ID NO:14 - AU Rich Sequence Elements (ARE)

ATTTTATT

SEQ ID NO:15 - AU Rich Sequence Elements (ARE)

ATTTTTAA

SEQ ID NO:16 - Kozak Consensus Sequence

GCCGCCACCATGC

SEQ ID NO:17 - CTP peptide

DPRFQDSSSSKAPPPSLPSPSRLPGPSDTPIL

SEQ ID NO:18 - CTP peptide

SSSSKAPPPSLPSPSRLPGPSDTPILPQ

SEQ ID NO:19 - albumin-binding peptides core sequence

DICLPRWGCLW

SEQ ID NO:20 - PAS Sequence

ASPAAPAPASPAAPAPSAPA

SEQ ID NO:21 - PAS Sequence

AAPASPAPAAPSAPAPAAPS

SEQ ID NO:22 - PAS Sequence

APSSPSPSAPSSPSPASPSS

SEQ ID NO:23 - PAS Sequence

APSSPSPSAPSSPSPASPS

SEQ ID NO:24 - PAS Sequence

SSPSAPSPSSPASPSPPSPA

SEQ ID NO:25 - PAS Sequence

AASPAAPSAPPAAASPAAPSAPPA

SEQ ID NO:26 - PAS Sequence

ASAAAPAAASAAASAPSAAA

SEQ ID NO:27 – Full length von Willebrand Factor nucleotide sequence

ATGATTCCTGCCAGATTGCGGGGTGCTGCTTGCTCTGGCCCTCATTTTGCCAGGGACCCCTTTGTGCAGAAGGAACTCGCGGCAG
 GTCATCCACGCGCCCTACTAAGGACGGTCTAAACGGCCCCACGACGAAACGAGACCGGGAGTAAACGGTCCCTGGGAAACACGCTCTT
 CCTTGAGCGCCGTCAGTAGGTGCCGGGATGCAGCCTTTTCGGAAGTGACTTCGTCAACACCTTTGATGGGAGCATGTACAGCTT
 TCGCGGATACTGCAGTTACCTCCTGGCAGGGGCTGCCAGAACTACGTCCGAAAAGCCTTCACTGAAGCAGTTGTGGAACTACCC
 TCGTACATGTGCAACGCGCTATGACGTCAATGGAGGACCGTCCCCGACGGTCTTACGCTCCTTCTCGATTATTGGGGACTTCCA
 GAATGGCAAGAGAGTGAGCCTCTCCGTGTATCTTGGGGAATTTTGTGACATCCATTGTGTTGTCATGTTGCGAGGAAGAGCTAA
 TAACCCCTGAAGGTCTTACCGTTCTCTCACTCGGAGAGGCACATAGAACCCCTTAAAAAACTGTAGGTAAACAAACAGTTACCAAC
 CGTGACACAGGGGGACCAAGAGTCTCCATGCCCTATGCTCCAAAGGGCTGTATCTAGAAACTGAGGCTGGGTACTACAGCTGT
 CCGGTGAGGCTTGCCACTGTCTCCCCCTGGTTTCTCAGAGGTACGGGATACGGAGGTTTCCCGACATAGATCTTTGACTCCGACC
 CATCATCTTCCACAGGCCACTCCGCAATCGCTTCTGCGCAGGATCGATGGCAGCGGCAACTTTCAAGTCCTCCTGTCAGACACAT
 ACTTCAACAAGACCTGCGGGCTGTGTGCAACTTTAACATTACCGAAACACCGGTCCTAGCTACCGTCCGCGTTGAAAGTTCAGGA
 CGACAGTCTGTCTATGAAGTTGTTCTGGACGCCCGACACACCGTTGAAATTGTACTTTGCTGAAGATGACTTTATGACCCAGAAG
 GGACCTTGACCTCGGACCTTATGACTTTGCCAACTCATGGGCTCTGAGCAGTGGAGAACAGTGGTGTGAAACGACTTCTACTGAA
 ATACTGGGTTCTTCCCTGGAAGTGGAGCTGGGAATACTGAAACGGTTGAGTACCCGAGACTCGTCACCTCTTGTCAACACAGAAC
 GGGCATCTCTCCAGCAGCTGATGCAACATCTCCTCTGGGCAATGCAGAGGGCCCTGTTGGGAGCAGTGCCAGCTCTCTGAAGAGC
 ACCTCGGTGTCTTCCCGTAGAGGAGGTCGTGAGTACGTTGTAGAGGAGACCCCTTACGTCTTCCCGGACACCTCGTCAAGG
 TCGAAGACTTCTCGTGGAGCCACATGCCCCGCTGCCACCTCTGCTGGACCCGAGCCTTTTGTGGCCCTGTGTGAGAAGACTTTG
 TGTGAGTGTCTGGGGGCTGGAGTGGCCCTGCCCTGCAACGGGCGACGGTGGGAGACCACCTGGGGCTCGGAAAACACCGGGACA
 CACTCTTCTGAAACACACTCACACGACCCCCCGACCTCACGCGGACGGGACGCTCCTGGAGTACGCCCGACCTGTGCCAGGAG

GGAATGGTSCGTACGGCTGACCGACACACGCGGTGCAGCCAGTGTGCCCTGCTGGTATGGAGGGAGGACCTCATGCGGGCCCT
 GGACACGGGICCTCCCTTACCACGACATGCCGACCTGGGTGGTGTGCGGCACGTGGGTACACGGGACGACCATACTCTATAGG
 CAGTGTGTGTCCCTTGGCCAGGACCTGCCAGAGCCTGCACATCAATGAAATGTGTGAGGACGATGGCTGGATGGCTGCAGCTG
 CCTGAGGATATCCGTACACACAGGGGAACGCGGTCTGGACGGTCTCGGACGTGTAGTTACTTTACACAGTCTCGCTACGCAC
 CTACCGACGTGACGGGACTCCGACAGCTCCTGGATGAAGGCTCTGCGTGGAGAGCACCGAGTGTCCCTGCGTGCATTCCGGAAA
 GCGCTACCCCTCCCGCACCTCCCTCTCTCGAGACGTCTGTCGAGGACCTACTTCCGAGACGACCTCTCGTGGCTCACAGGGACG
 CACGTAAGGCCCTTTCGCGATGGGAGGGCCGTGGAGGGAGAGAGCTCTGACCAACACCTGCATTTCGGAAACAGCCAGTGGATCTG
 CAGCAATGAAGAATGTCCAGGGGACTGCCTTGTCACTGGTCAATCCCACTTCAAGAGCTTTGACGTTGTGGACGTAAACGGCTTTC
 TCGGTCACTTAGACGTGCTTACTTCTTACAGGTCCCTCACGGAACAGTGACCACTTAGGGTGAAGTTCTCGAAACTGAACAGATA
 CTTCACTTTCAGTGGGATCTGCCAGTACCTGCTGGCCGGGATTGCCAGGACCACTCCTTCTCCATGTCATTGAGACTGTCCAGT
 GTGCTGTGTCTATGAAGTGAAGTCACTTAGACGGTCAATGACACACCGGGCCCTAACGGTCTGGTGGAGAGAGGTAACAGTA
 ACTCTGACAGGTACACGACATGACCGGACGCTGTGTGCACCGGCTCCGTCACTGCGGCTGCGTGGCTGCACACAGCCTTTC
 TGAAACTGAAGCATGGGGCAGGAGTTGCCATGGATACTGGCGCTGCGACACACCTGGGCGAGGAGTGGCAGGCGCAGCGGACCGGA
 CGTGTGTGTCGAACACTTTGACTTCGTACCCCGTCTCAACGGTACCTTGGCCAGGACATCCAGCTCCCGCTCCTGAAAGGTGACC
 TCCGCATCCAGCATAAGTGACGGCTCCGTGCGCTCAGCTACGGGGAGGACCTGCAGATGACCGCTCCTGTAGGTGAGGGGGA
 GGACTTTCCACTGGAGGCGTAGGTGTATGTCACTGCCGAGGACGCGGAGTGCATGCCCTCCTGACGTCTACGACTGGGATG
 GCCGCGGGAGGCTGTGTGTAAGCTGTCCCCGTCTATGCCGGGAAGACCTGCCGCTGTGTGGGAATTACAATGGCAACAGGGC
 GACGCTGACCTACCGCGCCCTCCGACGACCACTTCGACAGGGGGCAGATACGGCCCTTCTGGACGCGGACACACCTTAAATGT
 TACCGTGTGCTCCGCTGCACTTCTTACCCCTCTGGGCTGGCGAGCCCGGGTGGAGGACTTCGGGAACGCTGGAAGCTGCAC
 GGGGACTGCCAGGACCTGCAGAAGCAGCAGTGAAGGAATGGGGAGACCCGACCGYCTCGGGGCCACCTCCTGAAGCCCTTGC
 GACCTTCGACGTGCCCCGTGACGGTCTGGACGTCTTGTGTGTCCGATCCCTGCGCCCTCAACCGCGCATGACCAAGTTCTOC
 GAGGAGGCGTCCGCGTCTGACGTCCCCACATTCGAGGCTGCCATCGTCCGTGACGCTAGGGACGCGGGAGTTGGGCGCT
 ACTGGTCCAAGAGCTCTTCCGACGCGCCAGGACTGCAGGGGTGTAAGCTCCGGACGGTAGCACGGCAGTCCCGCTGCCCTAC
 CTGCCAACTCCCGTACACAGTGTCTCTCTCCGACGCGCCGACTGCCTCTCCGCGCCCTGGCCAGCTATGCCGCGCCCTG
 CGGGCGACGGGATGGACGCTTGACGCGATGCTGCACACGAGGACGAGCTGCGGGGCTCACGACACGCGCGGGACCGGTG
 ATACGGGCGCGACCGCGGGAGAGGCGTCCGCTGCGGTGGCGGAGCCAGGCGCTGTGAGCTGAAGTCCCGAAAGGCCAGGT
 GTACCTGCACTGCGGGACCCCTGCAACTGCCCTCTCCGACGCGCAGCGCACCGCGCTCGGTCCGGCGACACTGCAGTTGACG
 GCTTTCCGTTCCACATGGAGCTCACGCTTGGGGGACGTTGGAGACCTGCCGCTCTCTCTTACCCGGATGAGGAATGAATGA
 GGCCTGCTGAGGGCTGCTTCTGCCCCCGGGCTCTACATGGATGAGAGGGGGACCTGGACGGCGAGAGAGGAATGGGCTA
 CTCCTTACGTTACTCCGACGACCTCCCGACGAAGACGGGGGTCCCGAGATGTACCTACTCTCCCCCTGTGCGTGCCAAGGC
 CCAGTGCCCCCTTACTATGACGGTGGATCTTCCAGCCAGAAGACATCTTCTCAGACCATCACACCATGTGCTACTGTGAGGATG
 ACCGACGGGTTCGGGTCAAGGGGACAAATGATACTGCCACTCTAGAAGGTCCGTCTTCTGTAGAAGAGTCTGGTAGTGTGTACAC
 GATGACACTCCTACGCTTCAAGCACTGTACATGAGTGGAGTCCCGGAAGCTTGCCTGACGCTGTCTCAGCAGTCCCTGT
 CTCATCCGACAAAGAGGCTATCTGCGAAGTACGTGACATGCTACTCACCTCAGGGGCTTGAACGACGAGTCCGACAGGA
 GTGTCAGGGGACAGAGTAGCGTCTTTCTCGGATAGGACTCGGCCCCCATGGTCAAGCTGGTGTGTCCGCTGACAACTGC
 GCGCTGAAGGCTCGAGTGTACAAACGTGCCAGAACTATGACCTGAGTGCATGACCGGGGGGTACCAGTTCCAGCACACAGG
 GCGACTGTGACGCGCGACTTCCCGAGCTACATGGTTTTGACAGGTCTTGATACTGGACCTCACGTACAGCATGGGCTGTGTCT
 CTGGCTGCTCTGCCCGCGGCGATGCTCGGCATGAGAACAGATGTGTGCCCTGGAAGGTGTCCCTGCTTCCATCAGGGCATC
 GTACCGGACACAGAGACCGACGAGACGGGGGCGGCTACCGGCGGTACTCTTGTCTACACACCGGACCTTCCACAGGGACGA
 AGCTAGTCCCTAGGAGTATGCCCTGGAGAAACAGTGAAGATTGGCTGCAACACTTGTGTCTGTGGGACCGGAAGTGAAGTGC
 ACAGACCATGTGTGTGATGCCAGTGTCTCATACGGGGACCTTCTTGTCACTTCTAACCGACGTTGTGAACACAGACAGCCTGG
 CCTCACCTTACAGTGTCTGTACACACACTACGGTGCACCTCCAGATCGGCATGGCCACTACCTCACCTTCGACGGGCTCAAA
 TACCTGTCCCCGGGAGTGCCAGTACGTTCTGGTGCAGGATTACTGCGGCACTGAGGTGCTAGCCGTACCGGTGATGGAGTGA

AGCTGCCGAGTTTATGGACAAGGGGCCCTCACGGTATGCAAGACCAGTCTTAATGACCGCTCAAACCTGGGACCTTTCCG
 ATCCTAGTGGGGAATAAGGGATGCAGCCACCCCTCAGTGAATGAAGAAACGGGTACCATCCTGGTGGAGGGAGGAGATTGS
 GACCTGGAAAGCCTAGGATCACCCCTTATTCCTACGTGCGGTGGGAGTCACTTACGTTCTTTGCCAGTGGTAGGACCACTC
 CCTCCTCTCTTTGAGTGTGTTGACGGGAGGTGAATGTGAAGAGGCCATGAAGGATGAGACTCACTTTGAGGTGGTGGAGTCTGG
 CCGGTACATCATCTGCTGTGGGAACGACAACTGCCCTCCACTTACACTTCTCGGGTACTTCTACTCTGAGTGAAGTCTC
 CACCACTCAGACCGGCTATGTAGTAAGACGACGACCCAAAGCCCTCTCGTGGTCTGGGACCGCCACCTGAGCATCTCGGTGT
 CCTGAACGACATACAGGAGAAAGTGTGTGCCCTGTGTGGGAATTTGATTTTTCGGGAGAGGACACAGACCTGGCGGTGGAC
 TCGTAGAGGACACGAGTCTGTGTATGGTCTCTTTCACACACCGGACACACCTTAAACTAGSCATCCAGAACAATGACCT
 CACGAGCAGAACCTCCAAGTGGAGGAGACCTGTGCACTTTGGGAACCTTGGAAAGTGAAGTCCGAGTGTGCTGACACCGTAG
 GTCTTGTTACTGGAGTGGTGTGTTGGAGGTTACCTCCTTCTGGGACACTGAAACCTTGAGGACCTTCACTCGAGCGTCAC
 ACAGCTTCCAGAAAAGTGCCTCTGGACTCATCCCTGCCACCTGCCATAACAACATCATGAAGCAGACGATGGTGGATTCTCTCT

GTAGAATCCTTACCAGTGACGTGGTCTTTTACGGAGACCTGAGTAGGGGACGGTGGACGGTATTGTTGTAGTACTTGTCTGCTA
 CCACCTAAGGAGGACATCTTAGGAATGGTCACTGCACCTCCAGGACTGCAACAAGCTGGTGGACCCCGAGCCATATCTGGATGCT
 GCATTACGACACCTGCTCCTGTGAGTCCATTGGGGACTGCGCTGCTTCGAAGGTCTGACGTGTGTCGACCACCTGGGGCTCGG
 TATAGACCTACAGACGTAAATGCTGTGGACGAGGACACTCAGSIAACCCCTGACGCGGACGAAGTGGACACCATGCTGCTATG
 CCCACGTGTGTGCCAGCATGGCAAGGTGGTGAACCTGGAGGACGCCACATTGTGCCCCAGAGCTGGAGGAGAGGAACGCTGTG
 GTAACGACGGATACGGGTGCACACACGGGTCTTACCCTTCCACCTGGACCTCCTGCGGGTGTAAACACGGGGTCTCGACGCTCC
 TCTCTATCTCCGGGAGAACGGGTATGAGTGTGAGTGGCGCTATAACAGCTGTGCACCTGCTTETCAAGTCACGTGTGACACCCCT
 GAGCCACTGGCTTGCCTGTTAAGGGCTCTTGGCCATACCTACACTCACCGGATATTGTGACACGTGGACGGACAGTTCAGT
 GCACAGTGTGGGACTCGGTGACCGGACGGGACAGCAGTGTGTCGAGGGCTGCCATGCCACTCCCTCCAGGGAAAATCCTGGAT
 GAGCTTTTGACAGCTGCGTGGACCTGAAGACTGTCCAGTGTGTGAGCGTCACACACCTCCGACGGTACGGGTGACGGGAGGTG
 CCTTTTAGGACCTACTCGAAAAGCTGTGGACGCACTGGGACTTCTGACAGGTACACACTCTGTGGCTGGCGGGCTTTTGCCTCA
 GGAAAGAAAGTACCTTGAAATCCAGTGACCTGAGCACTGCCAGATTGGCCACTGTGATGTTGTCAACCTCACCTCACCGACCGG
 CCCCAAAACGGAGTCTTTCTTTTCACTGGAACCTTAGGCTCACTGGGACTCGTGACGGTCTAAACGGTGACACTACAACAGTTEGAG
 TGGAGTGAAGCTTGGCAGGAGCGGGAGGGCTGGTGGTCCCTCCACAGATGCCCGGTGAGCGCCACCACTCTGTATGTGGAGGA
 CATCTCGAAACCGGTTCACTTCCGACCGTCTCGGCTCTCGGACCAACCGAGGGTGTCTACGGGGCCACTCGGGGTGGTGA
 GACATACACCTCTGTAGAGCTTGGCGGCAAGCAGATTCTACTGCAGCAGGCTAGTGGACCTGGTCTTCTGCTGGATGGCTC
 CTCCAGGCTGTCCGAGCTGAGTTTGAAGTGTGAAGGCTTTGTGCGTGTCTAAAGATGACGTCTGCGATGACCTGGACCAAG
 GACGACCTACCGAGGAGTCCGACAGGCTCCGACTCAAACCTTACGACTTCCGAAACAGCTGGACATGATGGAGCGGCTGCGCAT
 CTCCAGAGTGGGTCCGCTGGCTGGTGGAGTACCACGAGGGCTCCACGCGTACATCGGGCTCAAGGACCACTCTGTACTAC
 CTCGCGGACGCTAGAGGCTCTACCCAGGCGCACCGGACCCCTCATGGTGTGCTGGAGGGTGGCGATGTAGCCGAGTTCTCT
 GGGGAAGCGAGCTCAGAGCTGCGGCGCATTGCAGCCAGGTGAAGTATGCGGGCAGCGAGGTGGCTCCACCAAGCGAGGTCTTGA
 AATACACACTGTTCCACTTTCGCTGGCAGTCTCGACGCGCGTAAACGGTCCGCTCCACTCATACGCGCGTCCGTCACCGGAGGTG
 GTGCTCCAGAACTTTATGTGTGACAGGTAACTCTCAGCAAGATCGACCGCCCTGAAGCTCCCGCATCGCCCTGCTCTGATGG
 CCAGCCAGGAGCCCAACGGATGTCCCGAACTTTGTCCGCTACTTAGAAGTCTTTCTAGCTGGCGGACTTCCGAGGGCGTAGCG
 GGACGAGGACTACCGGTGGTCTCGGGGTTCCTACAGGGCTTGAACAGGGCATGGTCCAGGGCTGAAGAAGAAGAGGTCA
 TTGTGATCCCGGTGGGCTTGGGCCCCATGCCAACCCTCAAGCAGATCGGCTCATCGAGAAGCAGGCCCCCTGCAGGTCCCGGACTT
 CTTCTCTTCCAGTAACATAGGGCCACCGCTAACCGGGGTACGGTTGGAGTTCGTCTAGGCGGAGTAGCTCTTCTGTCGGGGAC
 AGAACAAGGCTTCTGCTGAGCAGTGTGATGAGCTGGAGCAGCAAGGGGACAGATCTTAGCTACCTCTGTGACCTTGGCCCT
 GAAGCCCTCTCTCTCTTCTTCCGGAAGCAGACTCTGCACACCTACTCGACCTCGTCTGTTCCCTGCTCTAGCAATGATGGAGA
 CACTGGAAACGGGACTTCCGGGAGGAGTACTCTGCCGCCGACATGGCACAATCACTGTGGGCGGGGCTCTTGGGGTTTCC
 ACCCTGGGGCCCAAGAGGAACCTCATGGTCTGTGATGTGGCGATGAGACGGGGGCTGTACCGTGTTCAGTGACACCGGGCCCCG
 AGAACCCCAAGCTGGGACCCCGGTTCTCTTCAAGTACCAAGACCTACACCTCTGCTCTGGAAGATCGGACAAAATGCT
 GAAGCCGACTTCAACAGGAGCAAGAGTTGATGGAGAGGTGATTCAGCGATGATGTGGGCCAGGACAAAGCAGGACCTTCTTA

GCCTGTTTAAACACTTGGGTGAAGTTGTCTCGTTCCTCAAGTACCTCTTCACTAAGTGGCTTACCTACACCCGGTCTGTGCT
 ATCCAGCTCAGGTGCTGCAGTACTCTACATGCTGACCGTGGAGTACCCCTTCCAGCAGGACAGTCCAAAGGGGACATCTGCA
 GCGGGTGGCAGACGTAGGTGCAGTGCACACACCTCATGAGGATGTACCACTGGCACCTCATGGGGAAGTGGCTCCGTCTCAGGTT
 CCCCTGTAGGACGTGCCCCACGCTCTGATCCGCTACCAGGGCGGCAACAGGACCAACACTGGGCTGGCCCTGGGGTACCTCTCTGA
 CCACAGCTTCTTGGTCAGCCAGGCTGACCCGGAGCAGGCGCTAGGCGATGGTCCCBCGCTTGTCTGCTGTGACCCGACCGGGAC
 GCCATGGAGAGACTGGTGTGAGAGAACAGTGGTCCCACTGGCCCTCTGCGGCCCAACCTGGTCTACATGCTCACCGGAATCC
 TGCCTCTATGAGATCAAGAGGCTGCTTGGAGACATCCAGTGTGTCOCATTGGAGTGGGCCCTAATGGGTTGGACAGATGTAC
 CAGTGGCTTTTAGGACGGAGACTACTCTAGTTCTCCGACGGACCTCTGTAGTCCACCAGGGTAACCTCACCGGGATTACCCAA
 CGTGCAGGAGCTGGAGAGGATTGGCTGGCCCAATGCCCTATCCTCATCCAGGACTTTGAGACGCTCCCGGAGAGGCTCTGACC
 TGGTGTCTCAGGTTGCACGTCCTCGACCTCTCTTAACCGACCGGTTACGGGGATAGGAGTAGGTCCTGAAACTCTGCGAGGCGGC
 TCTCCGAGGACTGGACCAAGCTGAGGTGCTGCTCCGGAGAGGGCTGAGATCCCAACCTCTTCCCTGCACTTACTGACGCC
 AGCCCTGGAGCTGATCCTTCTCTGATGGCTCCTCCCTCCACGACGAGGCTCTCCCGACCTTAGGGGTGGGAGAGGAGGAG
 TGGACTGACCTCGGTCCGCCACCTGCACTAGGAACACACCTACCCAGGAGACTTCCACCTCTCTATTTTGTGAAATGAACA
 GTTTCGCCAAGGCTTTCATTTCAAAGCCATATAGGGCTCTGCTCACTCAGGTGTGAGTGTCTCAAGGGTCAAGAAATAAA
 ACTACTTTACTTCTCAAAGCGTTCCGAAAGTAAAGTTTTCGTTATATCCCGGAGCAGAGTGAAGTCCAGTCAAGACAGTATG
 GAAGCATCACCACTTACCTGTCATGGAACGTGGTCCCGGAGAAAGCCATTGCTGAGCTTGTGGACGTATGACGCGGAG
 GGAGGCCCTCATACCTTCTGATGTTGTTAACTGCACGTAACCTTGCACCGGGCTCTTTGGGGTAAACGACTCGGAACACCTGC
 AGTACCTGCCCCCTCCCTCCGGGACGCCAAATCGGGATGCTTGGGCTTGTCTGTGCGATACCTGACTTCAGAAATGCATGGTGC
 AGGCCGGGAGCTCAAAGCGGTGGTCACTGGTGTGCTGGTTTACCCCTACCGAACCAGGACGACGCTATGAAGTGAAGTC
 TTTACGTACCACTGCGGCTCGGCTCGGAGTTTCCGCCACAGTAGGACGAGCGGCTCTCTGTGGATTCACTGGATGCAGCAGCT
 GATGCGGCCAGGTCCAACAGAGTGACAGTGTCCCTATTGGAATTGGAGATCGCTACGATGACAGTCCCTGCAGAGACACCTAAGTC
 ACCTACGTCGTCGACTACGGCGGTCCAGGTTGTCTCACTGTCAACAGGATAACCTTAACCTCTAGCGATGCTACGTCGCCAGCTA

CGGATCTTGGCAGGCCAGCAGGCGACTCCAACGTGGTGAAGCTCCAGCGAATCGAAGACCTCCCTAOCATGGTCAOCTTGGGCAA
TTCTTGGGTCGATGCCTAGAACCGTCCGGGTCGTCCGCTGAGGTTGCACCACTTCGAGGTGCTTAGCTTCTGGAGGGATGGTAC
CAGTGGAAACCGTTAAGGAACCTCCACAACTGTGCTCTGGATTGTAGGATTGTCATGGATGAGGATGGGAATGAGAAGAGGCC
CGGGGACGTCTGGACCTTGGCAGACCACTGCCAGGAGGTGTTTGACACGAGACCTAAACAATCTAAACGTACCTACTCTACCC
TTACTCTTCTCCGGGCTCTGCAGACCTGGAACGCTGTGCTCACGTTGACCGTGACTTGGCAGCCAGATGGCCAGACCTTGTCTGAA
GAGTCATCGGGTCAACTCTGACCCGCGCTGAGCCCTTCTGCCCTAACACCCAGTCCCTCTGCGCACTCAACCCCTGGTCTACCG
GTCTGGAACGACTCTCTAGTAGCCAGTTGACACTGCCCCCGACTCCGGAAGCACGGGATTGTGGTCAAGGGACTTAAAGTGGAA
AGAGACCTGTGGCTGCCCTGGACCTGCCCTGYGTGTGACAGGCGAGCTCCACTCGGCACATCGTGACCTTTGATGGGCAGAAAT
TCAAAATTTTACCTTCTCTGGACACCGGCGGACCTGGACGGGACRCACACGTGTCCGTGAGGTGAGCCGTGTAGCACTGGAA
ACTACCCGTCTTAAAGTTGCTGACTGGCAGCTGTTCTATGTCTTATTTCAAAACAAGGAGCAGGACCTGGAGGTGATTCTCCATA
ATGGTGCTCGACCCCTGGAGCAAGGCGAGGCGGACTGACCGTCGACAAGAATACAGGATAAAGTTTGTCTCTCTCTGGACCT
CCACTAAGAGGTATTACCACGACGTCGGGACCTCGTTCCTGTCCTGTCATGAAATCCATCGAGGTGAAGCACAGTCCCTCTCCG
TCGAGSTGCACAGTCACATGGAGGTGACGCTGAATGCGAGACTGGTCTCTGTCTCTTACGACGTACTTTAGGTAGCTCCACTTCTGT
GTACGGGAGAGGCGAGTCSACGTGTACTGTACTTCCACTGCCACTTACCTCTGACCAGAGACAAGGAATGCTGGGTGGGAACA
TGGAAAGTCAACGTTTATGGTGCCATCATGATGAGGTGAGATTCAATCACCTTGGTCACATCTTACATTCAGTCCACAAACAAT
GAACCCACCCCTTGTACCTTCACTTGCATAATACACGCTAGTACGTACTCCAGTCTAAGTTAGTGAACCACTGTAGAAGTGAAGT
GAGGTGTTTTGTTACTGTTCCAACTGCAGCTCAGCCCAAGACTTTTGTCTCAAGACGTATGGTCTGTGTGGGATCTGTGATGAG
AAGCGGACCAATGACTTCATGCTGAGGATCAAGGTTGAGCTCGAGTCGGGTCTGAAAACGAACTTTCTGCATACCAAGACAC
CTTAGACACTACTCTTGCTCGGTTACTGAAGTACGACTCCCTAGGCACAGTCAACACAGACTGGAAAACACTTGTTCAGGAATGG
ACTGTGCAGCGGCCAGGCGAGAGCTGCCAGCCATCTTGAGGAGCAGTGTCTTGTCTCCCGTCTCAGTGGTGTGTGACCTTTTGTG
ACAAGTCTTACTGACACGTCGCGGTCTCCGTCTGCACGTCGGGTAGGACCTCTCTGTACAGAAACAGGCCGACAGCTCCGAC
TGCCAGGTCTCTCTTACCCTGTTTGTCTGAATGCCACAAGGTCCTGGCTCCAGCCACATTCTATGCCATCTGCCAGCAGGACAG

GGCTGTGAGGGTGACGCTCCAGGAGGAGAATGGTGACAAACGACTTACGGTGTTCAGGACCGAGGTGGGTGTAAGATACGGTAG
ACGGTCGTCTGTCTTGCACACGAGGACAAAGTGTGTGAGGTGATCGCCTCTTATGCCACCTCTGTGCGACCAACGGGGTCTGGT
TGACTGGAGGACACCTGATTTCTGTGCTAACGGTGGTCTCGTTTACACACTCCACTAGCGGAGAATACGGGTGGAGACAGCCTGG
TTGCCCCAGACCACTGACCTCTGTGTSACTAAAGACACGAATGTATGCCACCATCTCTGCTCTACAACCACTGTGAGCATGG
CTGTCCCCGGCACITGTATGGCAACGTGAGCTCTGTGGGGACCATCCCTCCGAAGTACAGTACGGGTGGTAGAGACAGATGTTG
GTGACACTCGTACCGACAGGGGCCGTGACACTACCGTTGCACTCGAGGACACCCCTGGTAGGGAGGCTTCGCTGTTTCTGCCCTCC
AGATAAAGTCATGTGGAAGGCAGCTGTGTCCCTGAAGAGGCTGCACTCAGTGCAATTGGTGAGGATGGAGTCCAGCACCACTTCG
ACAAAGACGGGAGGTCTATTTTCACTACAACCTTCCGTGACACAGGGAATCTCCGGACGTGAGTCACGTAACCACTCTTACCTCA
GGTCTGTGTCACCTGGAAGCTGGGTCCCGGACCAACGACCTGTGAGATCTGCACATGCCCTCAGCGGCGGGAAGGTCAACTGCA
CAACGACGCTCTGCCACCGGCCAAAGGACCTTCGGAACCGAGGCTGGTGGTGGGACAGTCTAGACGTGTACGGAGTCCGCGCG
CTTCCAGTTGACGTGTGCGTCCGGACGGGCTGCCGTTTGTCTCCACGTGTGGCTGTGTGAAGTACCGCGCTCCGCGAAGT
CAGACCACTGTGCCCCGAGTATGAGTGTGTGTGTGACCCAGTGAAGTGTGACCCGAGGCTGCACCCGACACACTTCATCGGSC
GGAGGCGGTCTTACGTCTGGTACAGCGGGGCTCATACTCACACACACTGGGTCACTCGACACTGTTGCCCCAGTGCCTCACT
GTGAACGTGGCTCCAGCCACACTGACCAACCTGGCGAGTGCAGACCAACTTCACCTGCGCTGCAGGAAGGAGGAGTGAACGG
GGGTACGGAGTGACACTTGACCGGAGGTGGGTGTGACTGGTTGGGACCGCTCACGTCTGGGTGAAGTGGACGCGGAGCTCT
TCCCTCTACCAAAAGACTGTCCCCACCTCTGCCCCCGCACCGTTTGGCCACCTTCGGAAGACCCAGTGTGTGATGAGTAT
CAGTGTGCTGCAACTGTGTCAACGTTTCTCACAGGGGTGGGAGGACGGGGGCGTGGCAACCGGTGGGAAGCCTTCTGGGTCA
CGACACTACTCATACTCACACGAGCTGTGACACAGTTGTCCACAGTGAAGTGTCTCCCTTGGGTACTTGGCTCAACCGCCACCAAT
GACTGTGGCTGTACCAACCACTGCTTCCGACAGGAGTGTGTGTCCACAGGTGTCACTCGACAGGGAACCCATGAACCGGA
GTTGGCGGTGGTTACTGACACCGACATGGTGTGGTGGACGGAAGGGCTGTTCCACACACAGGTGGGAAGCACCATCTACCTGTG
GGCCAGTTCTGGGAGGAGGCTGCGATGTGTGACCTGCACCGACATGGAGGATGCCGTGATGGGCTCCGCTGGCCCACTTCTGT
GGTAGATGGGACACCCCGTCAAGACCTCTCTCCGACGCTACACAGTGGAGTGGCTGTACCTCTACGGCACTACCCGGAGGCG
CACCGGTGTGCTCCAGAACCCCTGTGAGGACAGCTGTGGTGGGGCTTCACTTACGTTCTGCATGAAGGCGAGTGTCTGTGAAG
GTGCTGCGCATCTGCTGTGAGCAGGAGGCTCTTGGGACACTCTGTGTGACAGCCAGCCGAAGTGAATGCAAGACGTACTTCCG
CTCAGACACCTTCCACGGACGGTAGACGGACACTCGTGGTGAAGTGGCTCACCGCGGGGGACTCCAGTCTTCTCTGGAAGAGTGT
CGGCTCCAGTGGGCTCCCGGAGAACCCTGCCTCATCAATGAGTGTGACCACTGACCGAGTGGCGCCCCCTGAGGGTCAGA
AGGAACCTTCTCACAGCCGAGGCTCAACCGGAGGGGCTCTTGGGACGGAGTAGTTACTCACACTCCGAGTGAAGGAGGAGGTCTT
TATACAACAAAGSAACGTCTCTGCCCCAGCTGGAGGTCCCTGTCTGCCCTCGGGCTTTCAGCTGAGCTGTGAAGACAGGCTCAC
TTCCTCTCCAGAAATATGTGTTTCTTGCAGAGGACGGGGGTGACCTCCAGGGACAGACGGGGAGCCCGAAAGTTCGACTCGAC
ATTCTGCTCAGCGTGTGCCCCAGCTGTGCTGTGAGCGCATGGAGGCTGTCATGCTCAATGGCACTGTTCATTGGGCGCGGGAAGA
CTGTGATGATGATGTGTGCGAGTCCGACGAGGTTCCAGACGACACTCGCTTACCTCGGACGTACGAGTTACGCTGACAGTA
ACCCGGGCCCCCTTCTGACACTACTAGTACACACGACGACTGCCGCTGTCATGTTGACGTTGGGGTCACTCTGAGTCAAGCTGG
AGTGCAGGAAGACCACTGCAACCCCTGCCCTGGGTTACAAGGAAGTGTGAGCGGCGAGCTACCACTCCACCCCACTAGAG

ACCTAAGTTCGACCTCACGTCCTTCTGGTGGACGTTGGGGACGGGGACCCCAATGTTCCCTCAAATAACACAGSTGAATGTTGTG
GGAGATGTTTGCCCTACGGCTTGCACCAATTAGCTAAGAGGAGGACAGATCATGACACTGAAGCGTGATGAGACGCTTTTTATTGTG
TCCACTTACAACACCCCTCTACAAACGGATGCCGAACGTGGTAAGTCGATTCTCCTCCTGCTAGTACTGTGACTTCGCACTACTTT
GCGACCAGGATGSCCTGATAGTACTCACTTCTGCAAGGTCAATGAGAGAGGAGAGTACTTCTGGGAGAAGAGGGTCACAGGCTGCCA
CCCTTTGATGAACACAACGGCTCTACCGACACTATGAGTGAAGACGTTCCAGTTACTCTCTCCTCTCATGAAGACCTCTTCTCCC
AGTGTCCGACGGTGGGAAACTACTTGTGTTCTGTCTTGTCTGAGGGAGGTAAAATTATGAAAATCCAGGCACCTGCTGTGACACA
TGAGGAGGACCTGAGTGAACGACATCACTGCCAGGCTGCAGTATGACAGAACGACTCCCTCCATTTTAATACTTTAAGGTCCGT
GGACGACACTGTGTACTCTCGGACTACGTTGCTGTAGTGACGGTCCGACGTCTACTCAAGGTGGGAAGCTGTAAAGTCTGAA
GTAGAGGTGGATATCCACTACTGCCAGGGCAAATGTGCCAGCAAAGCCATGTACTCCATTGACATCAACGATGTAGTCCACCCCT
CGACATTGAGACTTCATCTCCACCTATAGGTGATGACGGTCCCGTTTACACGGTCGTTCCGTACATGAGTAACTGTAGTTGCTA
CAGCAGGACCAGTGCTCCTGCTGCTCTCCGACACGGACGGAGCCCATGCAAGTGGCCCTGCAGTGCACCAATGGCTCTGTTGTGTG

CCATCAGGTTCTCAATUGTCCCTGGTCACGAGGACGACGACAGSCTGTGCTCTGCTCGGGTACGTCACCGGGACGTGACGTGGTPTA
CCGAGACAAACACATGGTACTCCAAAGAGTTAGCCATGGAGTGCAAAATGCTCCCCCAGGAAGTGCAGCAAGTGA

SEQ ID NO:28 – Full length von Willebrand Factor peptide sequence (X is any natural

amino acid)

MIPARFAGVLLALALILPGT	LCAEGRTRGRSSTARCSLFGS	DFVNTFDGSMYSFAGYCSYL
LAGGCQKRSFSIIGDFQNGK	RVSLSVYLGEFFDIHLFVNG	TVTQGDQQRVSMFYASKGLYL
ETEAGYYKLSGEAYGFVARI	DGSGNFQVLLSDRYFNKTCG	LCGNFNIFAEDDFMTQEGTL
TSDPYDFANSWALSSGEQWC	ERASPPSSSCNLSGEMQKG	LWEQCQLLKSTSVFARCHPL
VDPEPFVALCEKTLCECAGG	LECACPALLEYARTCAQEGM	VLYGWTDHSAACSVCAPAGME
YRQCVSPCARTCQSLHINEM	CQERCVDGCSCPEGQLLDEG	LCVESTECPCVHSGKRYPPG
TLSLRDCNTCICRNSQWICS	NEECPCGECLVTGQSHFKSFD	NRYFTFSGICQYLLARDCQD
HSFSIVIETVQCADDRDAVC	TRSVTVRLPGLHNSLVKLKH	GAGVAMDGQDIQLPLLKGDL
RIQHTVTASVRLSYGEDLQM	DWDGRGRLLVKLSPVYAGKT	CGLCGNYNGNQGDDELTFSG
LAEPREDFGNAWKLHGDCQ	DLQKQHSDFPCALNFRMTRFS	EEACAVLTSPTEACHRAVS
PLPYLRNCRYDVCSCSDGRE	CLCGALASYAAACAGRGVRV	AWREPGRCELNCPKGQVYLQ
CGTPCNLTCSRSLYPDEECN	EACLEGCFCPPGLYMDERGD	CVPKAQCPCYIDGEIFQPED
IFSDHHTMCYCEDGFMHCTM	SGVPGSLLPDAVLSSPLSHR	SKRSLSCRPPMVKLVCPADN
LRAEGLECTKTQNYDLECM	SMGCVSGCLCPPGMVRHENR	CVALERCPCFHQGREYAPGE
TVKIGCNTCVCRDRKWNCTD	HVCDATCSTIGMAHYLTFDG	LKYLFPGECCQYVLVQDYCGS
NPGTFRILLVGNKGCSHPSVK	CKKRVITILVEGGEIELEFDGE	VNVKRPMKDETHFEVVESGR
YIILLGKALSVMWDRHLST	SVVLKQTYQEKVCGLCGNFD	GIQNNULTSSNLQVEEDPVD
FGNSWKVSSQCADTRKVPLD	SSPATCHNNIMKQTMVDSSC	RILTSDFVQDCNKLVDPEPY
LDVCIYDTCSCSIGDCACF	CDTIAAYAHVCAQHGVVVTW	RTATLCPQSCEERNLRENGY
ECEWRYNSCAPACQVTCQHP	EPLACPVQCVEGCHAHCPPG	KILDELLQTCVDPEDCPVCE
VAGRRFASGKKVTLNPSDPE	HCQICHCDVNVLTCEACQEP	GGLVVPPTDAPVSPTTLYVE
DISEPPLHDFYCSRLLDLVF	LLDGSSRLSEAEFEVLKAFV	VDMMERLRISQKWVRVAVVE
YHDGSHAYIGLKDRKRESEL	RRIASQVKYAGSQVASTSEV	LKYTLFQIFSKIDRPEASRI
ALLLMASQEPQRMRSRNFVRY	VOGLKKKKVIVIEVGIGPHA	NLKQIRLIEKQAPENKAFVL
SSVDELEQQORDEIVSYLCDL	APEAPPPTLPPDMAQVTVGF	GLLGVSTLGPKRNSMVLDA
FVLEGSCLKIGEADFNRSKEF	MEEVIQRMVDVGQDSIHVTVL	QYSYMTVVEYPFSEAQSKGD

ILQKRVKELRIQQGNKRNIGL ALKYLESDHSELVSQGDKEQA ENLVIMVIGNPFASDELIRKLE
GDIQVVPIGVGPANANVQELE RIGWPNAPILIQDFETLPRE APDLVLQRCCSGEGLQIPTL

SPAPDCSQPLDVILLILDGSS SFPASYFDEMKSFAKAFISK ANIGPRLTQVSVLQYGSITT
IDVPWNVVPEKAHLLSLVDV MQREGGPSQIGDALGFAVRY LTSEMHGARPGASKAVVILV
TDVSVDSVDAADAARSNRV TVFPIGIGDRYDAAQLRILA GPAGDSNVVKLQRIEDLPTM
VTLGNSFLHKLCSGFVRICM DEDGNEKRPGDVWTLPDQCH TVTCQPDGQTLLKSHRVNCD
RGLRPSCPNSQSEPVKVEETC GCRWTCPCVCTGSSTRHIVT FDGQNFKLTGSCSYVLFQNK
EQDLEVILHNGACSPGARQG CMKSIEVKHSALSVEXHSDM EVTVNGRLVSVPYVGGNMEV
NVYGAIMHEVRFNHLGHIETFTFPQNNFQLQLSPKT FASKTYGLCGICDENGANDF
MLRDGTVTTDWKTLVQEWTV QREGQTCQPILEEQCLVPDS SHCQVLLLPLFAECHKVLP
ATFYAICQODSCHQEQQVEV IASYAHLCRTNGVCVDWRTP DFCAMSCPPSLVYNHCEHGC
PRHCDGNVSSCGDHPSEGCF CPPDKVMLEGSCVPPEEACTQ CIGEDGVQHQFLEAWVPDHQ
PCQICTCLSGRKVNCTTQPC PTAKAPTCLCEVARLRQNA DQCCPEYECVCDPVSCDLFP
VPHCERGLQPTLTNPGECP NFCAACRKEECKRVSPPPSCP PHRLPTLRKTQCCDEYECAC
NCVNSTVSCPLGYLASTATN DCGCTTTTCLPDKVCVHRST IYPVGQFWEEGCDVCTCTDM
EDAVMGLRVAQCSQKPCEDS CRSGFTYVLHEGECCGRCLP SACEVVTGSPRGDSQSSWKS
VGSQWASPENPCLINECVRV KEEVFIQQRNVSCPQLEVPV CPSGFQLSCKTSACCPSCRC
ERMEACMLNGTVIGPGKTVM IDVCTTCRCMVQVGVISGFK LECRKTTCNCPCLGYKEENN
TGECCGRCLPTACTIQLRGG QIMTLKRDETLQDGDTHFC KVNERGEYFWEKRVTCPPFP
DEHKCLAEGGKIMKIPGTCC DTCEEPECNDITARLQYVKV GSCKSEVEVDIHYCQGKCAS
KAMYSIDINDVQDQSCCSP TRTEPMQVALHCTNGSVVYH EVLNAMECKCSPRKCSK

SEQ ID NO:29 - human IgG1 amino acids 233-236

ELLG

SEQ ID NO:30 - XTEN AE42-4, protein sequence

GAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPASS

SEQ ID NO:31 - XTEN AE42-4, DNA sequence

GGCGCGCCAGGTTCTCCTGCTGGCTCCCCACCTCAACAGAAGAGGGGACAAGCGAAAGC
GCTACGCCTGAGAGTGGCCCTGGCTCTGAGCCAGCCACCTCCGGCTCTGAAACCCCTGCC
TCGAGC

SEQ ID NO:32 - XTEN AE144-2A, protein sequence

TSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPG
TSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPG
TSESATPESGPGTSESATPESGPG

SEQ ID NO:33 - XTEN AE144-2A, DNA sequence

GGCGCGCCAACCACTACGGAGCCGTCGAGGGGAGCGCACCAGGAAGCCCGGCTGGGAGC
CCGACTTCTACCGAAGAGGGTACATCTACCGAACCAAGTGAAGGTTGAGCACCAGGCACC

TCAACAGAACCCTCTGAGGGCTCGGCGCCTGGTACAAGTGAGTCCGCCACCCAGAAATCC
GGGCCTGGGACAAGCACAGAACCTTCGGAAGGGAGTGCCCTGGAAACATCCGAATCGGCA
ACCCGAGAATCAGCGCGAGGATCTGAGCGCGGAGCTTCGGCTCCGAGACCGCTCGGACA

ACCCCGAGAGCCCTCCGAAGGATCAGCCCCAGGCACCGAGCCCTCTGAGGGAAGC
 GCACCTGGTACCAGCGAAAGCGCAACTCCCGAATCAGGTCCCGGTACGAGCGAGTCGGCG
 ACCCCGGAGAGCGGGCCAGGTGCCTCGAGC

SEQ ID NO:34 - XTEN AE144-3B, protein sequence

SPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPG
 TSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPG
 SPAGSPTSTEEGTSTEPSEGSAPG

SEQ ID NO:35 - XTEN AE144-3B, DNA sequence

GGCGCGCCAAGTCCCCTGGAAGCCCACTAGCACCGAAGAGGGGACCTCAGAGTCCGCC
 ACCCCCGAGTCCGGCCCTGGCTCTGAGCCTGCCACTAGCGGCTCCGAGACTCCTGGCACA
 TCCGAAAGCGCTACACCCGAGAGTGGACCCGGCACCTCTACCGAGCCCAGTGAGGGCTCC
 GCCCCTGGAACAAGCACCGAGCCCAGCGAAGGCAGCGCCCCAGGGACCTCCACAGAGCCC
 AGTGAAGGCAGTGCTCCTGGCACCAGCACCGAACCAAGCGAGGGCTCTGCACCCGGGACC
 TCCACCGAGCCAAGCGAAGGCTCTGCCCTGGCACTTCCACCGAGCCCAGCGAAGGCAGC
 GCCCCTGGGAGCCCCGCTGGCTCTCCACCAAGCACTGAGGAGGGCACATCTACCGAACCA
 AGTGAAGGCTCTGCACCAGGTGCCTCGAGC

SEQ ID NO:36 - XTEN AE144-4A, protein sequence

TSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPG
 TSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEG
 TSESATPESGPGTSTEPSEGSAPG

SEQ ID NO:37 - XTEN AE144-4A, DNA sequence

GGCGCGCCAACGTCCGAAAGTGCTACCCCTGAGTCAGGCCCTGGTAGTGAGCCTGCCACA
 AGCGGAAGCGAAACTCCGGGGACCTCAGAGTCTGCCACTCCCGAATCGGGGCCAGGCTCT
 GAACCGGCCACTTCAGGGAGCGAAACACCAGGAACATCGGAGAGCGCTACCCCGGAGAGC
 GGGCCAGGAAGTAGTACTGAGCCTAGCGAGGGAAGTGACCTGGTACAAGCGAGTCCGCC
 ACACCCGAGTCTGGCCCTGGCTCTCCAGCGGGCTCACCCACGAGCACTGAAGAGGGCTCT
 CCCGCTGGCAGCCCAACGTCGACAGAAGAAGGATCACCAGCAGGCTCCCCACATCAACA
 GAGGAGGGTACATCAGAATCTGCTACTCCCGAGAGTGGACCCGGTACCTCCACTGAGCCC
 AGCGAGGGGAGTGCACCAGGTGCCTCGAGC

SEQ ID NO:38 - XTEN AE144-5A, protein sequence

TSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPG
 TSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPG
 SPAGSPTSTEEGSPAGSPTSTEEG

SEQ ID NO:39 - XTEN AE144-5A, DNA sequence

GGCGCGCCAACATCAGAGAGCGCCACCCCTGAAAGTGGTCCCGGGAGCGAGCCAGCCACA
 TCTGGGTCCGAAACGCCAGGCACAAGTGAGTCTGCAACTCCCGAGTCCGGACCTGGGTCC
 GAGCCTGCCACTAGCGGCTCCGAGACTCCGGGAACTTCCGAGAGCGCTACACCAGAAAGC
 GGACCCGGAACCAAGTACCGAACCCTAGCGAGGGCTCTGCTCCGGGCAGCCAGCCGGCTCT
 CCTACATCCACGGAGGAGGGCACTTCCGAATCCGCCACCCCGAGTCAGGGCCAGGATCT
 GAACCCGCTACCTCAGGCAGTGAGACGCCAGGAACGAGCGAGTCCGCTACACCGGAGAGT

GGGCCAGGGAGCCCTGCTGGATCTCCTACGTCCACTGAGGAAGGGTCACCAGCGGGCTCG
 CCCACCAGCACTGAAGAAGGTGCCTCGAGC

SEQ ID NO:40 - XTEN AE144-6B, protein sequence

TSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPG
 SPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEG

SEFATGGDEIFGDFAGSFISIEGIDIEFDEGGAFGIDIEFDEGGAFGDEFAIGGDEIFG
TSESATPESGPGTSTEPSEGSAPG

SEQ ID NO:41 - XTEN AE144-6B, DNA sequence

GGCGCGCCAACATCTACCGAGCCTTCCGAAGGCTCTGCCCCCTGGGACCTCAGAATCTGCA
ACCCCTGAAAGCGGCCCTGGAACCTCCGAAAGTGCCACTCCCGAGAGCGGCCCAGGGACA
AGCGAGTCAGCAACCCCTGAGTCTGGACCCGGCAGCGAGCCTGCAACCTCTGGCTCAGAG
ACTCCCGGCTCAGAACCCGCTACCTCAGGCTCCGAGACACCCGGCTCTCCTGCTGGGAGT
CCCCTTCCACCGAGGAAGGAACATCCACTGAGCCTAGTGAGGGCTCTGCCCCCTGGAACC
AGCACAGAGCCAAGTGAGGGCAGTGACACAGGATCCGAGCCAGCAACCAGCGGGTCCGAG
ACTCCCGGGACCTCTGAGTCTGCCACCCAGAGAGCGGACCCGGCACTTCAACCGAGCCC
TCCGAAGGATCAGCACACAGGTGCCTCGAGC

SEQ ID NO:42 - XTEN AG144-1, protein sequence

PGSSPSASTGTGPGSSPSASTGTGPGTSGSTASSSPGSSTPSGATGSPGSSPSASTGTG
PGASPGTSSTGSPGTGSGTASSSPGSSTPSGATGSPGTGSGTASSSPGASPGTSSTGS
PGASPGTSSTGSPGTGSGTASSS

SEQ ID NO:43 - XTEN AG144-1, DNA sequence

GGCGCGCCACCCGGGTGCTCCCCGTCCGGCTCCACCGGAACAGGGCCAGGGTCATCCCCG
TCAGCGTCCACTGGGACGGGACCCGGGACACCCGGTTCGGGGACTGCATCCTCCTCGCCT
GGTTCGTCCACCCCGTCAGGAGCCACGGGTTCGCCGGGAAGCAGCCCAAGCGCATCCACT
GGTACAGGGCCTGGGGCTTCACCGGGTACTTCATCCACGGGGTCACCGGGAACGCCCGGA
TCGGGGACGGCTTCCTCATCACCAGGATCGTCAACACCCTCGGGCGCAACGGGCAGCCCC
GGAACCCCTGGTTCGGGTACGGCGTCTGTCAGCCCCGGTGCGAGCCCCGGGAACAAGCTCG
ACAGGATCGCCTGGGGCGTTCACCGGCACGTCCAGCACAGGCAGCCCCGGAACCCCTGGA
TCGGGAACCGCGTCTGTCAAGCGCCTCGAGC

SEQ ID NO:44 - XTEN AG144-A, protein sequence

GASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTSGSTASSSPGSSTPSGATGSP
GSSPSASTGTGPGASPGTSSTGSPGTGSGTASSSPGSSTPSGATGSPGTGSGTASSSP
GASPGTSSTGSPGASPGTSSTGSP

SEQ ID NO:45 - XTEN AG144-A, DNA sequence

GGCGCGCCAGGTGCCTCGCCGGGAACATCATCAACTGGTTCACCCGGGTGCATCCCCCTCG
GCCTCAACCGGGACGGGTCCCGGCTCATCCCCAGCGCCAGCACTGGAACAGGTCTCTGGC
ACTCCTGGTTCGGGTACGGCATCGTCATCCCCGGGAAGCTCAACACCGTCCGGAGCGACA
GGATCACCTGGCTCGTCACCTTCGGCGTCAACTGGAACGGGGCCAGGGGCCTCACCCGGA
ACGTCCCTCGACTGGGTTCGCTGGTACGCGGGGATCAGGAACGGCCTCATCCTCGCCTGGG
TCCTCAACGCCCTCGGGTGGGACTGGTTCGCCGGGAACCTCCTGGCTCGGGGACGGCCTCG
TCGTCGCTGGGGCATCACCGGGGACGAGCTCCACGGGGTCCCCTGGAGCGTCACCGGGG
ACCTCCTCGACAGGTAGCCCCGCCCTCGAGC

SEQ ID NO:46 - XTEN AG144-B, protein sequence

GTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGTGSGTASSSPGSSTPSGATGSP
GSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSP
GASPGTSSTGSPGASPGTSSTGSP

SEQ ID NO:47 - XTEN AG144-B, DNA sequence

GGCGCGCCAGGTACACCGGGCAGCGGCACGGCTTCGTCTGTCACCCGGCTCGTCCACACCG

TCGGGAGCTACGGGAAGCCCAGGAGCGTCACCGGGAACGTCGTCAACGGGGTCACCGGGT
 ACGCCAGGTAGCGGCACGGCCAGCAGCTCGCCAGGTTTCATCGACCCCGTCGGGAGCGACT
 GGGTCGCCCCGGATCAAGCCCGTCAGCTTCCACTGGAACAGGACCCGGGTCGTGCGCGTCA
 GCCTCAACGGGGACAGGACCTGGTTCATCGACGCCGTGAGGGGCGACAGGCTCGCCCGGA
 TCGTCAACACCCTCGGGGGCAACGGGGAGCCCTGGTGCGTCGCCTGGAACCTCATCCACC
 GGAAGCCCGGGGGCCTCGCCGGGTACGAGCTCCACGGGATCGCCCGGAGCGTCCCCCGGA
 ACTTCAAGCACAGGGAGCCCTGCCTCGAGC

SEQ ID NO:48 - XTEN AG144-C, protein sequence

GTPGSGTASSSPGASPGTSSSTGSPGASPGTSSSTGSPGASPGTSSSTGSPGSSPSASTGTGP
 GTPGSGTASSSPGASPGTSSSTGSPGASPGTSSSTGSPGASPGTSSSTGSPGSSSTPSGATGSP
 GSSTPSGATGSPGASPGTSSSTGSP

SEQ ID NO:49 - XTEN AG144-C, DNA sequence

GGCGCGCCAGGTACACCCGGATCGGGTACAGCGTCATCGAGCCCGGTGCGTCACCTGGT
 ACGTCGAGCACGGGGTCGCCAGGGGCGTCCCCTGGGACGTCCTCAACAGGCTCGCCCGGT
 GCGTCACCCGGCACGTCGTCCACGGGTTACCTGGTAGCTCCCCTCCGCGTCCACTGGC
 ACCGGGCTCGAACTCCGGGGAGCGGCACAGCGAGCTCGTCGCCGGGAGCATCGCCTGGG
 ACATCGAGCACCGGGTCGCCAGGAGCATCGCCCGGAACATCCAGCACAGGAAGCCCGGGC
 GCGTCGCCCGGGACATCAAGCACAGGTTCCC CGGGATCGAGCACGCCGTCGGGAGCCACT
 GGATCACAGGGAGCTCGACACCTTCCGGCGCAACGGGATCGCCCGGAGCCAGCCCGGGT
 ACGTCAAGCACTGGCTCCCCTGCCTCGAGC

SEQ ID NO:50 - XTEN AG144-F, protein sequence

GSSPSASTGTGPGSSPSASTGTGPGASPGTSSSTGSPGASPGTSSSTGSPGSSSTPSGATGSP
 GSSPSASTGTGPGASPGTSSSTGSPGSSPSASTGTGPGTSGTASSSPGSSSTPSGATGSP
 GSSTPSGATGSPGASPGTSSSTGSP

SEQ ID NO:51 - XTEN AG144-F, DNA sequence

GGCGCGCCAGGCTCCAGCCCCCTCCGCGAGCACGGGAACCGGACCAGGTTTCGTACCCCTCA
 GCATCAACGGGGACGGGACCGGGGGCGTCACCAGGAACGTCCTCCACCGGCTCGCCCGGT
 GCATCACCCGGAACGTTCATCGACCGGATCGCCAGGGAGCTCGACGCCATCAGGCGCAACA
 GGATCACCTGGCTCAAGCCCTAGCGCGTCAACCGGCACGGGTCCGGGTGCCTCCCCTGGC
 ACGTCCAGCACCGGATCAACCGGATCGAGCCCATCCGCCTCAACCGGAACCGGACCCGGT
 ACACCAGGGTCGGGAACAGCCTCCTCGTCACCAGGCTCCTCAACCCCTCGGGAGCCAAG
 GGTTCGCCCGGTTTCGTCAACGCCTTCCGGAGCAACTGGTAGCCCGGAGCATCGCCAGGA
 ACTTCGAGCACGGGGTCGCCCGCCTCGAGC

SEQUENCE LISTING

[0305]

<110> Biogen Idec MA Inc.

<120> Optimized Factor VIII Gene

<130> 2159.378PC01

<140> To be assigned

<141> 2014-02-14

<150> 61/765,626

<151> 2013-02-15

<160> 51

<170> PatentIn version 3.5

<210> 1

<211> 4404

<212> DNA

<213> Artificial Sequence

<220>

<223> optimized BDD FVIII

<400> 1

cgtacggccg ccaccatgca gattgagctg tctacttgct ttttcctgtg cctgctgagg	60
ttttgctttt ccgctacacg aaggtattat ctgggggctg tggaactgtc ttgggattac	120
atgcagagtg acctgggaga gctgccagtg gacgcaaggt ttccccctag agtccctaag	180
tcattcccct tcaacactag cgtggtctac aagaaaacac tgttcgtgga gtttactgat	240
cacctgttca acatcgcaaa gcctaggcca ccctggatgg gactgctggg gccacaacac	300
caggccgagg tgtacgacac cgtggtcatt acacttaaga acatggcctc acaccccggtg	360
agcctgcatg ctgtgggcgt cagctactgg aaggcttccg aaggagcaga gtatgacgat	420
cagacttccc agagagaaaa agaggacgat aagggttttc ctggcggatc tcatacctac	480
gtgtggcagg tcctgaaaga gaatggccct atggcctccg accctctgtg cctgacctac	540
tcttatctga gtcacgtgga cctggtcaag gatctgaaca gcggcctgat cggagccctg	600
ctggtgtgca ggaaggaag cctggctaag gagaaaaccc agacactgca taagttcatt	660
ctgctgttcg ccgtgtttga cgaagggaaa tcatggcaca gcgagacaaa gaatagtctg	720
atgcaggaca gggatgccgc ttcagccaga gcttggccca aaatgcacac tgtgaacggc	780
tacgtcaatc gctcactgcc tgggctgacg ggctgccacc gaaagagcgt gtattggcat	840
gtcatcgga tgggcaccac acctgaagtg cactccattt tcctggaggg acataccttt	900
ctggtccgca accaccgaca ggcttccctg gagatctctc caattacctt cctgacagca	960
cagactctgc tgatggacct ggggcagttc ctgctgtttt gccacatcag ctcccaccag	1020
catgatggca tggaggotta cgtgaaagtg gactcttgct ccgaggaacc tcagctgcgg	1080
atgaagaaca atgaggaagc agaagactat gacgatgacc tgaccgactc cgagatggat	1140
gtggtccgat tcgatgacga taacagcccc tccttttatcc agattagatc tgtggccaag	1200
aaacacccta agacatgggt ccattacatc gcagccgagg aagaggactg ggattatgca	1260
ccactggtgc tggcaccaga cgatcgctcc taaaaatctc agtatctgaa caatgggcca	1320
cagaggattg gcagaaagta caagaaagtg cggttcatgg catataccga tgagaccttc	1380
aagactcgcg aagccatcca gcacgagagc ggcacacctg gaccactgct gtacggagaa	1440

gtgggagaca ccctgctgat ctttttcaag aaccaggcca gccggcctta caatatctat	1500
ccacatggga ttacagatgt gcgccctctg tacagcagga gactgccaaa gggcgtcaaa	1560
cacctgaagg acttcccaat cctgcccgga gaaatcttca agtacaagtg gactgtcacc	1620
gtcgaggatg gcccactaa gagcgaccct cgggtgcctga cccgctacta ttctagtttc	1680
gtgaatatgg aaagagatct ggcaagcgga ctgacgcgac cactgctgat ttgttacaaa	1740
gagagcgtgg atcagagagg caaccagatc atgtccgaca agcgggaatgt gattctgttc	1800
agtgtctttg acgaaaacag gtcattgtac ctgaccgaga acatccagag attcctgcct	1860
aatccagctg ggggtgcagct ggaagatcct gagtttcagg catctaacat catgcatagt	1920
attaatggct acgtgttcga cagtttgcag ctgagcgtgt gcctgcacga ggtcgcttac	1980
tggtatatcc tgagcattgg ggcacagaca gatttcctga gcgtgttctt ttccggctac	2040
acttttaagc ataaaatggc ctatgaggac acaactgactc tgttccctt cagcggcgaa	2100
accgtgttta tgagcatgga gaatcccgga ctgtggattc tggggtgcc caacagcgat	2160
ttcagaaatc gcggaatgac tgcctgctg aaagtgtcaa gctgtgacaa gaacaccggg	2220
gactactatg aagattcata cgaggacatc agcgcatac tgetgtccaa aaacaatgcc	2280
attgaacccc ggtcttttag tcagaatcct ccagtgtga agcggcacca gcgcgagatc	2340
accgcacta ccctgcagag tgatcaggaa gagatcgact acgacgatac aatttctgtg	2400
gaaatgaaga aagaggactt cgatatctat gacgaagatg agaaccagag tcctcgatca	2460
ttccagaaga aaaccaggca ttactttatt gccgcagtgg agcggctgtg ggattatggc	2520
atgtcctcta gtctcacgt gctgcgaaat agggcccagt caggaagcgt cccacagttc	2580
aagaaagtgg tcttcaggga gtttacagac gggtccttta ctacgccact gtacaggggc	2640
gaactgaacg agcacctggg actgctgggg ccctatatca gagcagaagt ggaggataac	2700
attatggtca ccttcagaaa tcaggcctct cggccttaca gtttttattc aagcctgatc	2760
tcttacgaag aggaccagcg acagggagct gaaccacgaa aaaacttcgt gaagccta	2820
gagacaaaaa catacttttg gaaggtgcag caccatatgg ccccaacaaa agacgagttc	2880
gattgcaagg catgggccta tttttctgac gtggatctgg agaaggacgt gcacagtggc	2940
ctgattggcc cactgctggt gtgccatac aacacctga atccagccca cggccggcag	3000
gtcactgtcc aggagtctgc tctgttcttt accatctttg atgagacaaa gagctggtac	3060
ttcacgaaa acatggagcg aaattgcag gctccatgta acattcagat ggaagacccc	3120
acattcaagg agaactaccg ctttcatgct atcaatggat acatcatgga tactctgccc	3180
gggctggtca tggcacagga ccagagaatc cgggtgtatc tgctgagcat gggcagcaac	3240
gagaatatcc actcaattca tttcagcggg cacgtgttta ctgtcaggaa gaaagaagag	3300
tacaagatgg ccctgtacaa cctgtatccc ggcgtgttcg aaaccgtcga gatgctgcct	3360
agcaaggccg gaatctggag agtggaatgc ctgattggag agcacctgca tgctgggatg	3420
tctaccctgt ttctggtgta cagtaataag tgtcagacac ccctgggaat ggcatccggg	3480
catatcaggg atttccagat tacgcgatct ggacagtacg gacagtgggc acctaagctg	3540
gctagactgc actattccgg atctatcaac gcttgggtca caaaagagcc tttctcttgg	3600
attaagtggt acctgctggc cccaatgatc attcatggca tcaaaactca gggagctcgg	3660
gagaagtctg cctctctcta catctcagc tttatgata tctagagcct ccatgggagc	3720

cagaagctct cctctctgta catctccag cctatccca tgcacagctt ggaatgggaa	3720
aaatggcaga cataccgcgg caatagcaca ggaactctga tgggtgttctt tggcaacgtg	3780
gacagcagcg gaatcaagca caacattttc aatcccccta tcattgctag atacatccgg	3840
ctgcacccaa cccattattc tattcgaagt acactgagga tggaaactgat gggatgcgat	3900
ctgaacagtt gttcaatgcc cctggggatg gagtccaagg caatctctga cgccagatt	3960
accgctagct cctacttcac taatatgttt gctacctgga gcccttccaa agcaagactg	4020
cacctgcaag gccgcagcaa cgcattggcg ccacaggtga acaatcccaa ggagtgggtg	4080
caggtcgatt ttcagaaaac tatgaagggtg accgggggtc caactcaggg cgtgaaaagt	4140
ctgctgacct caatgtacgt caaggagttc ctgatctcta gttcacagga cggacatcag	4200
tggacactgt tctttcagaa cgggaagggtg aaagtcttcc agggcaatca ggattccttt	4260
acacctgtgg tcaacagtct agaccctcca ctgctgacca gataacctgag aatccaccct	4320
cagtcctggg tgcaccagat tgccctgaga atggaagtgc tgggatgcga ggccaggat	4380
ctgtactgat aactcgagtc gacc	4404

<210> 2

<211> 4419

<212> DNA

<213> Artificial Sequence

<220>

<223> optimized BDD FVIII

<400> 2

ggcgcgcccg tacggccgcc accatgcaga tegagctgtc tacctgcttc ttcctgtgcc	60
tgctgcgggt ctgcttcagc gccacccggc ggtactacct gggcgcctgt gaactgagct	120

gggactacat gcagagcgac ctgggggagc tgcccgtgga cgcacagatc ccccaagag	180
tgcccaagag cttcccttc aacacctccg tgggtgtaca gaaaaccctg ttcgtcgagt	240
tcaccgacca cctgttcaat atcgccaagc ccagaccccc ctggatgggc ctgctgggcc	300
ctacaatcca ggccgaggtg tacgacacgg tgggtcatcac ccttaagaac atggccagcc	360
acccctgtgc cctgcacgcc gtgggcgtgt cctactggaa ggctctgag ggcgtgagt	420
acgacgacca gaccagccag cgcgagaaag aggacgaca agtctttcct ggcggcagcc	480
atacctacgt gtggcaggtc ctgaaagaaa acggccctat ggctccgac ccctgtgcc	540
tgacctacag ctacctgagc cacgtggacc tgggtcaagga cctgaacagc ggctgattg	600
gcgcctgtct cgtgtgtaga gagggcagcc tcgccaaaga gaaaaccag accctgcaca	660
agttcatcct gctgttcgcc gtgttcgacg agggcaagag ctggcacagc gagacaaaga	720
acagcctgat gcaggaccgg gacgcgcct ctgccagagc ctggcctaag atgcacaccg	780
tgaacggcta cgtgaacaga agcctgccc gactgatcgg ctgccaccgg aagtcctgt	840
actggcacgt gatcggcatt ggcaccacc ccgaggtgca cagcatcttt ctggaaggcc	900
acaccttcct cgtgcggaac cacagacagg ccagcctgga aatcagcct atcaccttc	960
tgaccgcca gacactgctg atggacctgg gccagttcct gctgttttgc cacatcagca	1020

gccaccagca cgacggcatg gaagcctacg tgaagggtga cagctgcccc gaggaacccc	1080
agctgcggat gaagaacaac gaggaagccg aggactacga cgacgacctg accgacagcg	1140
agatggacgt cgtgcgcttc gacgacgaca acagccccag cttcatccag atcagaagcg	1200
tggccaagaa gcaccccaag acctgggtgc actatatcgc cgccgaggaa gaggactggg	1260
actacgcccc tctggtgctg gccccgacg acagaagcta caagagccag tacctgaaca	1320
atggccccc ggggatcggc cggaagtaca agaaagtgcg gttcatggcc tacaccgacg	1380
agacattcaa gaccagagag gccatccagc acgagagcgg catcctgggc ccctgctgt	1440
atggcgaagt gggcgacacc ctgctgatca tcttcaagaa ccaggccagc cgccctaca	1500
acatctaccc ccacggcatc accgacgtgc gggccctgta cagcagacgg ctgcccagg	1560
gcgtgaagca cctgaaggac ttccccatcc tgcccgcgca gatcttcaag tacaagtgga	1620
ccgtgaccgt ggaagatggc cccaccaaga gcgaccccag atgcctgacc cggtactaca	1680
gcagcttcgt gaacatggaa cgggacctgg cctccgggct gatcgccct ctgctgatct	1740
gctacaaaga aagcgtggac cagcggggca accagatcat gagcgacaag cggaacgtga	1800
tcctgttcag cgtgttcgat gagaatcggc cctggtacct gaccgagaat atccagcggc	1860
tcctgcccga ccctgccggc gtgcagctgg aagatcccga gttccaggcc agcaacatca	1920
tgcactccat caatggctac gtgttcgaca gcctccagct gagcgtgtgc ctgcacgagg	1980
tggcctactg gtacatcctg agcatcggcg ccagaccga cttcctgagc gtgttcttca	2040
goggctacac cttcaagcac aagatggtgt acgaggatac cctgaccctg tccccctct	2100
ccggcgaaac cgtgttcatg agcatggaaa accccggcct gtggattctg ggctgccaca	2160
acagcgactt cagaaaccgg ggcgatgacc ccctgctgaa ggtgtccagc tgcgacaaga	2220
acaccggcga ctactacgag gacagctatg aggacatcag cgcctacctg ctgagcaaga	2280
acaacgccat cgagcccaga tccttcagcc agaaccccc cgtgctgaag cggcaccaga	2340
gagagatcac ccggaccacc ctgcagtccg accaggaaga gattgattac gacgacacca	2400
tcagcgtcga gatgaagaaa gaggatttcg acatctacga cgaggacgag aaccagagcc	2460
cccggtcctt ccagaagaaa acccggcact acttcattgc cgcctggaa agactgtggg	2520
actacggcat gagcagcagc cccacgtgc tgcggaacag agcccagagc ggcagcgtgc	2580
cccagttcaa gaaagtgggt ttccaggagt tcaccgacgg cagcttcacc cagcccctgt	2640
atcggggcga gctgaacgag cacctgggac tgctgggacc ttacattaga gccgaggtgg	2700
aagataacat catggtcacc ttcagaaacc aggcctccag accctacagc ttctacagca	2760
gcctgatcag ctacgaagag gaccagcggc agggcgccga accccggaag aacttcgtga	2820
agcccaacga gactaagacc tacttctgga aggtgcagca ccacatggcc cccacaaagg	2880
acgagttcga ctgcaaggcc tgggcctact tctccgatgt ggacctggaa aaggacgtgc	2940
actctggcct gattggacct ctgctcgtct gccacaccaa caccctgaac cccgccacg	3000
gccggcaggc cacagtgcag gaatttgccc tgttcttcac catcttcgat gagacaaaga	3060
gctggtactt caccgagaac atggaaagaa actgtagagc cccctgcaac atccagatgg	3120
aagatcctac cttcaaagag aactatcggc tccacgccat caacggctac atcatggaca	3180
ccctgcccg cctggtcatg gcccaggatc agagaatccg gtggtatctg ctgagcatgg	3240
gcagcaacga gaacatccac agcatccact tcagcggcca cgtgttcaca gtgcggaaga	3300

aagaagagta caagatggcc ctgtacaacc tgtaccccg	3360
cggtgtcgag acagtggaaa	
tgctgcccag caaggccggc atctggcggg tggaatgtct gatcggcgag catctgcacg	3420
cgggaatgag caccctgttt ctggtgtaca gcaacaagtg ccagaccct ctgggcatgg	3480
ccagcgccca catccgggac ttccagatca ccgctccgg ccagtacggc cagtgggccc	3540
ctaagctggc ccggtccac tactccgat ctatcaacgc ctggtccacc aaagagccct	3600
tcagctggat caaggtggac ctgctggccc ctatgatcat ccacggaatc aagaccagg	3660
gcgccagaca gaagttcagc agcctgtaca tcagccagtt catcatcatg tacagcctgg	3720
acggcaagaa gtggcagacc taccggggca acagcaccgg caccctgatg gtgttcttcg	3780
gcaacgtgga cagcagcggc atcaagcaca acatcttcaa ccccccatc attgcccggc	3840
acatccggct gcacccacc cactacagca tccgggtccac cctgcggatg gaactgatgg	3900
gctgcgacct gaactcttgc agcatgcccc tggggatgga aagcaaggcc atcagcgacg	3960
cccagatcac agccagcagc tacttcacca acatgttcgc cacctggtcc ccaagcaaag	4020
cccgctgca tctccaaggc agaagcaatg cctggcggcc tcaggtcaac aaccccaaag	4080
aatggctcca ggtggacttt cagaaaacca tgaaggtcac aggcgtgacc acccaggcgc	4140
tgaaaagcct gctgacctct atgtacgtga aagagttcct gatcagcagc agccaggacg	4200
ggcaccagtg gaccctgttc ttccagaacg gcaaagtga agtggtccag ggcaaccagg	4260
actcctttac ccccggtgc aactctctag accctccact gctgaccaga tacctgagaa	4320
tccacctca gtccctgggtg caccagattg ccctgagaat ggaagtgtg ggatgcgagg	4380
cccaggatct gtactgataa ctcgagtcca cttaattaa	4419

<210> 3

<211> 4371

<212> DNA

<213> Artificial Sequence

<220>

<223> Parental BDD FVIII

<400> 3

atgcaaatag agctctccac ctgcttcttt ctgtgccttt tgcgattctg ctttagtgcc	60
accagaagat actacctggg tgcagtggaa ctgtcatggg actatatgca aagtgatctc	120
ggtgagctgc ctgtggacgc aagatttcct cctagagtgc caaaatcttt tccattcaac	180
acctcagtcg tgtacaaaaa gactctgttt gtagaattca cggatcacct tttcaacatc	240
gctaagccaa ggccaccctg gatgggtctg ctaggctcta ccatccaggc tgaggtttat	300
gatacagtgg tcattacact taagaacatg gcttcccatc ctgtcagtct tcatgctgtt	360
ggtgtatcct actggaaagc ttctgaggga gctgaatatg atgatcagac cagtcaaagg	420
gagaaagaag atgataaagt cttccctggt ggaagccata catatgtctg gcaggctctg	480
aaagagaatg gtccaatggc ctctgaccca ctgtgcctta cctactcata tctttctcat	540
gtggacctgg taaaagactt gaattcaggc ctcatggag ccctactagt atgtagagaa	600
ggagctctgg ccaaggaaaa gacacagacc ttgcacaaat ttatactact ttttactata	660

```

---
ttt gat gaag g g a a a g t t g g c a c t c a g a a a c a a g a a c t c c t t g a t g c a g g a t a g g g a t 720
g c t g c a t c t g c t c g g g c c t g g c c t a a a a t g c a c a c a g t c a a t g g t t a t g t a a c a g g t c t 780
c t g c c a g g t c t g a t t g g a t g c c a c a g g a a t c a g t c t a t t g g c a t g t g a t t g g a a t g g g c 840
a c c a c t c c t g a a g t g c a c t c a a t a t t c c t c g a a g g t c a c a c a t t t c t t g t g a g g a a c c a t 900
c g c c a g g c g t c c t t g g a a a t c t c g c c a a t a a c t t t c c t t a c t g c t c a a a c a c t c t t g a t g 960
g a c c t t g g a c a g t t t c t a c t g t t t t g t c a t a t c t c t t c c c a c c a a c a t g a t g g c a t g g a a 1020
g c t t a t g t c a a a g t a g a c a g c t g t c c a g a g g a c c c c a a c t a c g a a t g a a a a t a a t g a a 1080

g a a g c g g a a g a c t a t g a t g a t g a t g a t c t t a c t g a t t c t g a a a t g g a t g t g g t c a g g t t t g a t 1140
g a t g a c a a c t c t c c t t c c t t a t c c a a a t t c g t c a g t t g c c a a g a a g c a t c c t a a a a c t 1200
t g g g t a c a t t a c a t t g c t g c t g a a g a g g a g g a c t g g g a c t a t g c t c c c t t a g t c c t c g c c 1260
c c c g a t g a c a g a a g t t a t a a a a g t c a a t a t t t g a a c a a t g g c c c t c a g c g g a t t g g t a g g 1320
a a g t a c a a a a a a g t c c g a t t a t g g c a t a c a c a g a t g a a a c c t t t a a g a c t c g t g a a g c t 1380
a t t c a g c a t g a a t c a g g a a t c t t g g g a c c t t a c t t t a t g g g a a g t t g g a g a c a c a c t g 1440
t t g a t t a t a t t a a g a a t c a a g c a a g c a g a c c a t a t a c a t a c c c t c a c g g a a t c a c t 1500
g a t g t c c g t c c t t t g t a t t c a a g g a t t a c c a a a a g g t g t a a a a c a t t t g a a g a t t t t 1560
c c a a t t c t g c c a g g a g a a a t a t t c a a a t a t a a t g g a c a g t g a c t g t a g a a g a t g g g c c a 1620
a c t a a a t c a g a t c c t c g g t g c c t g a c c c g c t a t t a c t c t a g t t c g t t a a t a t g g a g a g a 1680
g a t c t a g c t t c a g g a c t c a t t g g c c c t c t c t c a t c t g c t a c a a a g a a t c t g t a g a t c a a 1740
a g a g g a a a c c a g a t a a t g t c a g a c a a g a g g a a t g t c a t c c t g t t t t c t g t a t t g a t g a g 1800
a a c c g a a g c t g g t a c c t c a c a g a g a a t a t a a a c g c t t t c t c c c c a a t c c a g c t g g a g t g 1860
c a g c t t g a g g a t c c a g a g t t c c a a g c c t c c a a c a t c a t g c a c a g c a t c a a t g g c t a t g t t 1920
t t t g a t a g t t t g c a g t t g t c a g t t t g t t t g c a t g a g g t g g c a t a c t g g t a c a t t c t a a g c 1980
a t t g g a g c a c a g a c t g a c t t c c t t t c t g t c t c t t c t c t g a t a t a c c t t c a a c a c a a a 2040
a t g g t c t a t g a a g a c a c a c t c a c c c t a t t c c a t t c t c a g g a g a a a c t g t c t t c a t g t c g 2100
a t g g a a a a c c c a g g t c t a t g g a t t c t g g g g t g c c a c a a c t c a g a c t t t c g g a a c a g a g g c 2160
a t g a c c g c c t t a c t g a a g t t t c t a g t t g t g a c a a g a a c a c t g g t g a t t a t t a c g a g g a c 2220
a g t t a t g a a g a t a t t t c a g c a t a c t t g c t g a g t a a a a c a a t g c c a t t g a a c c a a g a a g c 2280
t t c t c t c a a a a c c c a c c a g t c t t g a a a c g c c a t c a a c g g g a a t a a a c t c g t a c t a c t c t t 2340
c a g t c a g a t c a a g a g g a a a t g a c t a t g a t g a t a c c a t a t c a g t t g a a a t g a a g a g g a a 2400
g a t t t t g a c a t t a t g a t g a g g a t g a a a a t c a g a g c c c c c g c a g c t t t c a a a g a a a c a 2460
c g a c a c t a t t t a t t g c t g c a g t g g a g a g g c t c t g g g a t t a t g g g a t g a g t a g c t c c c c a 2520
c a t g t t c t a a g a a c a g g g c t c a g a t g g c a g t g t c c c t c a g t t c a a g a a a g t t g t t t t c 2580
c a g g a a t t t a c t g a t g g c t c c t t a c t c a g c c c t t a t a c c g t g g a g a a c t a a t g a a c a t 2640
t t g g g a c t c c t g g g g c c a t a t a a g a g c a g a a g t t g a a g a t a a t a t c a t g g t a a c t t t c 2700
a g a a a t c a g g c c t c t c g t c c c t a t t c c t t c t a t t c t a g c c t t a t t t c t t a t g a g g a a g a t 2760
c a g a g g c a a g g a g c a g a a c c t a g a a a a a c t t t g t c a a g c c t a a t g a a a c c a a a a c t t a c 2820
t t t t g a a a g t g a a c a t c a t a t g g a c c c a c t a a a g a t g a g t t t g a c t c a a a a c c t g g 2880

```

gcttatttct	ctgatgttga	cctggaaaaa	gatgtgcact	caggcctgat	tggaccctt	2940
ctggtctgcc	acactaacac	actgaaccct	gctcatggga	gacaagtgc	agtacaggaa	3000
tttgctctgt	ttttcaccat	ctttgatgag	accaaagct	ggtacttcac	tgaaaatatg	3060
gaaagaaact	gcagggtcc	ctgcaatatc	cagatggaag	atcccacttt	taaagagaat	3120
tatcgcttcc	atgcaatcaa	tggctacata	atggatacac	tacctggctt	agtaatggct	3180
caggatcaaa	ggattcgatg	gtatctgctc	agcatgggca	gcaatgaaaa	catccattct	3240
attcatttca	gtggacatgt	gttcaactga	cgaaaaaag	aggagtataa	aatggcactg	3300
tacaatctct	atccagggtg	ttttgagaca	gtggaaatgt	taccatccaa	agctggaatt	3360
tggcgggtgg	aatgccttat	tggcgagcat	ctacatgctg	ggatgagcac	actttttctg	3420
gtgtacagca	ataagtgtca	gactccctcg	ggaatggctt	ctggacacat	tagagatttt	3480
cagattacag	cttcaggaca	atatggacag	tgggccccaa	agctggccag	acttcattat	3540
tccggatcaa	tcaatgcctg	gagcaccaag	gagccctttt	cttggatcaa	ggtggatctg	3600
ttggcaccaa	tgattattca	cggcacccaag	acccagggtg	cccgtcagaa	gttctccagc	3660
ctctacatct	ctcagtttat	catcatgtat	agtcttgatg	ggaagaagtg	gcagacttat	3720
cgaggaaatt	ccactggaac	cttaatggtc	ttctttggca	atgtggattc	atctgggata	3780
aaacacaata	tttttaacct	tccaattatt	gctcgataca	tccgtttgca	cccaactcat	3840
tatagcattc	gcagcactct	tgcgatggag	ttgatgggct	gtgatttaaa	tagttgcagc	3900
atgccattgg	gaatggagag	taaagcaata	tcagatgcac	agattactgc	ttcatcctac	3960
tttaccata	tgtttgccac	ctgggtctcct	tcaaagctc	gacttcacct	ccaagggagg	4020
agtaatgcct	ggagacctca	ggtgaataat	ccaaaagagt	ggctgcaagt	ggacttccag	4080
aagacaatga	aagtcacagg	agtaactact	caggggagtaa	aatctctgct	taccagcatg	4140
tatgtgaagg	agttcctcat	ctccagcagt	caagatggcc	atcagtggac	tctctttttt	4200
cagaatggca	aagtaaagg	ttttcaggga	aatcaagact	ccttcacacc	tgtgggtgaac	4260
tctctagacc	caccgttact	gactcgctac	cttcgaattc	accccagag	ttgggtgcac	4320
cagattgcc	tgaggatgga	ggttctgggc	tgcgaggcac	aggacctcta	c	4371

 $\langle 210 \rangle_4$

<211> 1438

<212> PRT

<213> Artificial Sequence

 $\langle 220 \rangle$

<223> BDD FVIII

<400> 4

Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser Trp Asp Tyr
1 5 10 15

Met Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg Phe Pro Pro
20 25 30

Arg Val Pro Lys Ser Phe Pro Phe Asn Thr Ser Val Val Tyr Lys Lys
 35 40 45
 Thr Leu Phe Val Glu Phe Thr Asp His Leu Phe Asn Ile Ala Lys Pro
 50 55 60
 Arg Pro Pro Trp Met Gly Leu Leu Gly Pro Thr Ile Gln Ala Glu Val
 65 70 75 80
 Tyr Asp Thr Val Val Ile Thr Leu Lys Asn Met Ala Ser His Pro Val
 85 90 95
 Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys Ala Ser Glu Gly Ala
 100 105 110
 Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp Asp Lys Val
 115 120 125
 Phe Pro Gly Gly Ser His Thr Tyr Val Trp Gln Val Leu Lys Glu Asn
 130 135 140
 Gly Pro Met Ala Ser Asp Pro Leu Cys Leu Thr Tyr Ser Tyr Leu Ser
 145 150 155 160
 His Val Asp Leu Val Lys Asp Leu Asn Ser Gly Leu Ile Gly Ala Leu
 165 170 175
 Leu Val Cys Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr Gln Thr Leu
 180 185 190
 His Lys Phe Ile Leu Leu Phe Ala Val Phe Asp Glu Gly Lys Ser Trp
 195 200 205
 His Ser Glu Thr Lys Asn Ser Leu Met Gln Asp Arg Asp Ala Ala Ser
 210 215 220
 Ala Arg Ala Trp Pro Lys Met His Thr Val Asn Gly Tyr Val Asn Arg
 225 230 235 240
 Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys Ser Val Tyr Trp His
 245 250 255
 Val Ile Gly Met Gly Thr Thr Pro Glu Val His Ser Ile Phe Leu Glu
 260 265 270
 Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser Leu Glu Ile
 275 280 285
 Ser Pro Ile Thr Phe Leu Thr Ala Gln Thr Leu Leu Met Asp Leu Gly
 290 295 300
 Gln Phe Leu Leu Phe Cys His Ile Ser Ser His Gln His Asp Gly Met
 305 310 315 320
 Glu Ala Tyr Val Lys Val Asp Ser Cys Pro Glu Glu Pro Gln Leu Arg
 325 330 335
 Met Tyr Asp Asp Glu Glu Ala Glu Asp Thr Asp Asp Asp Tyr Thr Asp

```

Met Lys Asn Asn Glu Glu Ala Glu Asp Tyr Asp Asp Asp Leu Thr Asp
      340              345              350

Ser Glu Met Asp Val Val Arg Phe Asp Asp Asp Asn Ser Pro Ser Phe
      355              360              365

Ile Gln Ile Arg Ser Val Ala Lys Lys His Pro Lys Thr Trp Val His
      370              375              380

Tyr Ile Ala Ala Glu Glu Glu Asp Trp Asp Tyr Ala Pro Leu Val Leu
      385              390              395              400

Ala Pro Asp Asp Arg Ser Tyr Lys Ser Gln Tyr Leu Asn Asn Gly Pro
      405              410              415

Gln Arg Ile Gly Arg Lys Tyr Lys Lys Val Arg Phe Met Ala Tyr Thr
      420              425              430

Asp Glu Thr Phe Lys Thr Arg Glu Ala Ile Gln His Glu Ser Gly Ile
      435              440              445

Leu Gly Pro Leu Leu Tyr Gly Glu Val Gly Asp Thr Leu Leu Ile Ile
      450              455              460

Phe Lys Asn Gln Ala Ser Arg Pro Tyr Asn Ile Tyr Pro His Gly Ile
      465              470              475              480

Thr Asp Val Arg Pro Leu Tyr Ser Arg Arg Leu Pro Lys Gly Val Lys
      485              490              495

His Leu Lys Asp Phe Pro Ile Leu Pro Gly Glu Ile Phe Lys Tyr Lys
      500              505              510

Trp Thr Val Thr Val Glu Asp Gly Pro Thr Lys Ser Asp Pro Arg Cys
      515              520              525

Leu Thr Arg Tyr Tyr Ser Ser Phe Val Asn Met Glu Arg Asp Leu Ala
      530              535              540

Ser Gly Leu Ile Gly Pro Leu Leu Ile Cys Tyr Lys Glu Ser Val Asp
      545              550              555              560

Gln Arg Gly Asn Gln Ile Met Ser Asp Lys Arg Asn Val Ile Leu Phe
      565              570              575

Ser Val Phe Asp Glu Asn Arg Ser Trp Tyr Leu Thr Glu Asn Ile Gln
      580              585              590

Arg Phe Leu Pro Asn Pro Ala Gly Val Gln Leu Glu Asp Pro Glu Phe
      595              600              605

Gln Ala Ser Asn Ile Met His Ser Ile Asn Gly Tyr Val Phe Asp Ser
      610              615              620

Leu Gln Leu Ser Val Cys Leu His Glu Val Ala Tyr Trp Tyr Ile Leu
      625              630              635              640

Ser Ile Glu Ala Glu Thr Asp Phe Leu Ser Val Phe Phe Ser Glu Tyr

```

```

Ser Ile Gly Ala Gln Thr Asp Phe Leu Ser Val Phe Phe Ser Gly Tyr
645                                650                                655

Thr Phe Lys His Lys Met Val Tyr Glu Asp Thr Leu Thr Leu Phe Pro
660                                665                                670

Phe Ser Gly Glu Thr Val Phe Met Ser Met Glu Asn Pro Gly Leu Trp
675                                680                                685

Ile Leu Gly Cys His Asn Ser Asp Phe Arg Asn Arg Gly Met Thr Ala
690                                695                                700

Leu Leu Lys Val Ser Ser Cys Asp Lys Asn Thr Gly Asp Tyr Tyr Glu
705                                710                                715                                720

Asp Ser Tyr Glu Asp Ile Ser Ala Tyr Leu Leu Ser Lys Asn Asn Ala
725                                730                                735

Ile Glu Pro Arg Ser Phe Ser Gln Asn Pro Pro Val Leu Lys Arg His
740                                745                                750

Gln Arg Glu Ile Thr Arg Thr Thr Leu Gln Ser Asp Gln Glu Glu Ile
755                                760                                765

Asp Tyr Asp Asp Thr Ile Ser Val Glu Met Lys Lys Glu Asp Phe Asp
770                                775                                780

Ile Tyr Asp Glu Asp Glu Asn Gln Ser Pro Arg Ser Phe Gln Lys Lys
785                                790                                795                                800

Thr Arg His Tyr Phe Ile Ala Ala Val Glu Arg Leu Trp Asp Tyr Gly
805                                810                                815

Met Ser Ser Ser Pro His Val Leu Arg Asn Arg Ala Gln Ser Gly Ser
820                                825                                830

Val Pro Gln Phe Lys Lys Val Val Phe Gln Glu Phe Thr Asp Gly Ser
835                                840                                845

Phe Thr Gln Pro Leu Tyr Arg Gly Glu Leu Asn Glu His Leu Gly Leu
850                                855                                860

Leu Gly Pro Tyr Ile Arg Ala Glu Val Glu Asp Asn Ile Met Val Thr
865                                870                                875                                880

Phe Arg Asn Gln Ala Ser Arg Pro Tyr Ser Phe Tyr Ser Ser Leu Ile
885                                890                                895

Ser Tyr Glu Glu Asp Gln Arg Gln Gly Ala Glu Pro Arg Lys Asn Phe
900                                905                                910

Val Lys Pro Asn Glu Thr Lys Thr Tyr Phe Trp Lys Val Gln His His
915                                920                                925

Met Ala Pro Thr Lys Asp Glu Phe Asp Cys Lys Ala Trp Ala Tyr Phe
930                                935                                940

```

```

Ser Asp Val Asp Leu Glu Lys Asp Val His Ser Gly Leu Ile Gly Pro
945                      950                      955                      960

Leu Leu Val Cys His Thr Asn Thr Leu Asn Pro Ala His Gly Arg Gln
                      965                      970                      975

Val Thr Val Gln Glu Phe Ala Leu Phe Phe Thr Ile Phe Asp Glu Thr
                      980                      985                      990

Lys Ser Trp Tyr Phe Thr Glu Asn Met Glu Arg Asn Cys Arg Ala Pro
                      995                      1000                      1005

Cys Asn Ile Gln Met Glu Asp Pro Thr Phe Lys Glu Asn Tyr Arg
1010                      1015                      1020

Phe His Ala Ile Asn Gly Tyr Ile Met Asp Thr Leu Pro Gly Leu
1025                      1030                      1035

Val Met Ala Gln Asp Gln Arg Ile Arg Trp Tyr Leu Leu Ser Met
1040                      1045                      1050

Gly Ser Asn Glu Asn Ile His Ser Ile His Phe Ser Gly His Val
1055                      1060                      1065

Phe Thr Val Arg Lys Lys Glu Glu Tyr Lys Met Ala Leu Tyr Asn
1070                      1075                      1080

Leu Tyr Pro Gly Val Phe Glu Thr Val Glu Met Leu Pro Ser Lys
1085                      1090                      1095

Ala Gly Ile Trp Arg Val Glu Cys Leu Ile Gly Glu His Leu His
1100                      1105                      1110

Ala Gly Met Ser Thr Leu Phe Leu Val Tyr Ser Asn Lys Cys Gln
1115                      1120                      1125

Thr Pro Leu Gly Met Ala Ser Gly His Ile Arg Asp Phe Gln Ile
1130                      1135                      1140

Thr Ala Ser Gly Gln Tyr Gly Gln Trp Ala Pro Lys Leu Ala Arg
1145                      1150                      1155

Leu His Tyr Ser Gly Ser Ile Asn Ala Trp Ser Thr Lys Glu Pro
1160                      1165                      1170

Phe Ser Trp Ile Lys Val Asp Leu Leu Ala Pro Met Ile Ile His
1175                      1180                      1185

Gly Ile Lys Thr Gln Gly Ala Arg Gln Lys Phe Ser Ser Leu Tyr
1190                      1195                      1200

Ile Ser Gln Phe Ile Ile Met Tyr Ser Leu Asp Gly Lys Lys Trp
1205                      1210                      1215

Gln Thr Tyr Arg Gly Asn Ser Thr Gly Thr Leu Met Val Phe Phe
1220                      1225                      1230

```

```

-----
Gly Asn Val Asp Ser Ser Gly Ile Lys His Asn Ile Phe Asn Pro
1235          1240          1245

Pro Ile Ile Ala Arg Tyr Ile Arg Leu His Pro Thr His Tyr Ser
1250          1255          1260

Ile Arg Ser Thr Leu Arg Met Glu Leu Met Gly Cys Asp Leu Asn
1265          1270          1275

Ser Cys Ser Met Pro Leu Gly Met Glu Ser Lys Ala Ile Ser Asp
1280          1285          1290

Ala Gln Ile Thr Ala Ser Ser Tyr Phe Thr Asn Met Phe Ala Thr
1295          1300          1305

Trp Ser Pro Ser Lys Ala Arg Leu His Leu Gln Gly Arg Ser Asn
1310          1315          1320

Ala Trp Arg Pro Gln Val Asn Asn Pro Lys Glu Trp Leu Gln Val
1325          1330          1335

Asp Phe Gln Lys Thr Met Lys Val Thr Gly Val Thr Thr Gln Gly
1340          1345          1350

Val Lys Ser Leu Leu Thr Ser Met Tyr Val Lys Glu Phe Leu Ile
1355          1360          1365

Ser Ser Ser Gln Asp Gly His Gln Trp Thr Leu Phe Phe Gln Asn
1370          1375          1380

Gly Lys Val Lys Val Phe Gln Gly Asn Gln Asp Ser Phe Thr Pro
1385          1390          1395

Val Val Asn Ser Leu Asp Pro Pro Leu Leu Thr Arg Tyr Leu Arg
1400          1405          1410

Ile His Pro Gln Ser Trp Val His Gln Ile Ala Leu Arg Met Glu
1415          1420          1425

Val Leu Gly Cys Glu Ala Gln Asp Leu Tyr
1430          1435

```

<210> 5

<211> 6

<212> DNA

<213> Artificial Sequence

<220>

<223> MAR/ARS

<400> 5

atattt 6

<210> 6
<211> 6
<212> DNA
<213> Artificial Sequence

<220>
<223> MAR/ARS

<400> 6
aaatat 6

<210> 7
<211> 6
<212> DNA
<213> Artificial Sequence

<220>
<223> Potential Splice Site

<400> 7
ggtgat 6

<210> 8
<211> 5
<212> DNA
<213> Artificial Sequence

<220>
<223> Destabilizing Sequence

<400> 8
attta 5

<210> 9
<211> 5
<212> DNA
<213> Artificial Sequence

<220>
<223> Destabilizing Sequence

<400> 9
taaata 5

<210> 10
<211> 6
<212> DNA
<213> Artificial Sequence

<220>

<223> poly-T Sequence

<400> 10

tttttt 6

<210> 11

<211> 7

<212> DNA

<213> Artificial Sequence

<220>

<223> poly-A Sequence

<400> 11

aaaaaaaa 7

<210> 12

<211> 5

<212> DNA

<213> Artificial Sequence

<220>

<223> Promoter Binding Site

<400> 12

tataa 5

<210> 13

<211> 5

<212> DNA

<213> Artificial Sequence

<220>

<223> Promoter Binding Sequence

<400> 13

ttata 5

<210> 14

<211> 8

<212> DNA

<213> Artificial Sequence

<220>

<223> AU Rich Sequence Elements (ARE)

<400> 14

atatttatt 8

<210> 15

<211> 8

<212> DNA

<213> Artificial Sequence

<220>

<223> AU Rich Sequence Elements (ARE)

<400> 15

atattttaa 8

<210> 16

<211> 13

<212> DNA

<213> Artificial Sequence

<220>

<223> Kozak Consensus Sequence

<400> 16

gccgccacca tgc 13

<210> 17

<211> 32

<212> PRT

<213> Artificial Sequence

<220>

<223> CTP peptide

<400> 17

Asp	Pro	Arg	Phe	Gln	Asp	Ser	Ser	Ser	Ser	Lys	Ala	Pro	Pro	Pro	Ser
1				5					10					15	

Leu	Pro	Ser	Pro	Ser	Arg	Leu	Pro	Gly	Pro	Ser	Asp	Thr	Pro	Ile	Leu
			20					25					30		

<210> 18

<211> 28

<212> PRT

<213> Artificial Sequence

<220>

<223> CTP peptide

<400> 18

Ser	Ser	Ser	Ser	Lys	Ala	Pro	Pro	Pro	Ser	Leu	Pro	Ser	Pro	Ser	Arg
1				5					10					15	

Leu Pro Gly Pro Ser Asp Thr Pro Ile Leu Pro Gln
 20 25

<210> 19

<211> 11

<212> PRT

<213> Artificial Sequence

<220>

<223> albumin-binding peptides core sequence

<400> 19

Asp Ile Cys Leu Pro Arg Trp Gly Cys Leu Trp
 1 5 10

<210> 20

<211> 20

<212> PRT

<213> Artificial Sequence

<220>

<223> PAS Sequence

<400> 20

Ala Ser Pro Ala Ala Pro Ala Pro Ala Ser Pro Ala Ala Pro Ala Pro
 1 5 10 15

Ser Ala Pro Ala
 20

<210> 21

<211> 20

<212> PRT

<213> Artificial Sequence

<220>

<223> PAS Sequence

<400> 21

Ala Ala Pro Ala Ser Pro Ala Pro Ala Ala Pro Ser Ala Pro Ala Pro
 1 5 10 15

Ala Ala Pro Ser
 20

<210> 22

<211> 20

<212> PRT

<213> Artificial Sequence

<220>

<223> PAS Sequence

<400> 22

Ala Pro Ser Ser Pro Ser Pro Ser Ala Pro Ser Ser Pro Ser Pro Ala
1 5 10 15

Ser Pro Ser Ser
20

<210> 23

<211> 19

<212> PRT

<213> Artificial Sequence

<220>

<223> PAS Sequence

<400> 23

Ala Pro Ser Ser Pro Ser Pro Ser Ala Pro Ser Ser Pro Ser Pro Ala
1 5 10 15

Ser Pro Ser

<210> 24

<211> 20

<212> PRT

<213> Artificial Sequence

<220>

<223> PAS Sequence

<400> 24

Ser Ser Pro Ser Ala Pro Ser Pro Ser Ser Pro Ala Ser Pro Ser Pro
1 5 10 15

Ser Ser Pro Ala
20

<210> 25

<211> 24

<212> PRT

<213> Artificial Sequence

<220>

<223> PAS Sequence

<400> 25

Ala Ala Ser Pro Ala Ala Pro Ser Ala Pro Pro Ala Ala Ala Ser Pro
1 5 10 15

Ala Ala Pro Ser Ala Pro Pro Ala
20

<210> 26
 <211> 20
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> PAS Sequence

<400> 26
 Ala Ser Ala Ala Ala Pro Ala Ala Ala Ser Ala Ala Ala Ser Ala Pro
 1 5 10 15
 Ser Ala Ala Ala
 20

<210> 27
 <211> 16842
 <212> DNA
 <213> Homo sapiens

<400> 27
 atgattcctg ccagatttgc cggggtgctg cttgctctgg ccctcathtt gccagggacc 60
 ctttgtgcag aaggaactcg cggcaggtca tccacggccc tactaaggac ggtctaaacg 120
 gccccacgac gaacgagacc gggagtataa cgggtccctgg gaaacacgtc ttccttgagc 180
 gccgtccagt aggtgccggg gatgcagcct tttcggaagt gacttogtca acacctttga 240
 tgggagcatg tacagctttg cgggatactg cagttacctc ctggcagggg gctgccagaa 300
 ctacgtcgga aaagccttca ctgaagcagt tgtggaaact accctcgtac atgtcgaaac 360
 gccctatgac gtcaatggag gaccgtcccc cgacgggtctt acgctccttc tcgattattg 420
 gggacttcca gaatggcaag agagtgcgac tctccgtgta tcttggggaa ttttttgaca 480
 tccatttggt tgtcaatggt tgcgaggaag agctaataac ccctgaaggt cttaccgttc 540
 tctcactcgg agaggcacat agaaccctt aaaaaactgt aggtaaacaa acagttacca 600
 accgtgacac agggggacca aagagtctcc atgccctatg cctccaaagg gctgtatcta 660
 gaaactgagg ctgggtacta caagctgtcc ggtgaggcct tggcactgtg tccccctggc 720
 ttctcagagg tacgggatac ggaggtttcc cgacatagat ctttgactcc gacccatgat 780
 gttcgacagg ccactccgga atggctttgt ggccaggatc gatggcagcg gcaactttca 840
 agtcctgctg tcagacagat acttcaacaa gacctgcggg ctgtgtggca actttaacat 900
 taccgaacaa ccggtcctag ctaccgtcgc cgttgaaagt tcaggacgac agtctgtcta 960
 tgaagtgtt ctggacgccc gacacaccgt tgaaattgta ctttgctgaa gatgacttta 1020
 tgaccaaga agggaccttg acctcggacc cttatgactt tgccaactca tgggctctga 1080
 gcagtggaga acagtgggtg gaaacgactt ctactgaaat actgggttct tccctggaac 1140
 tggagcctgg gaatactgaa acggttgagt acccgagact cgtcacctct tgtcaccaca 1200
 gaacgggcat ctccctccag cagctcatgc aacatctcct ctggggaaat gcagaagggc 1260
 ctctccacac cctccacacg cctccacacg cctccacacg cctccacacg cctccacacg 1320

ctgtgggagc agtgcacgct tctgaagagc acctcgggtg cttgcccgtg gaggaggggc	1320
gtcgagtacg ttgtagagga gaccccttta cgtcttcccg gacacccctg tcacgggtcg	1380
agacttctcg tggagccaca ttgcccgctg ccacctctg gtggaccccg agccttttgt	1440
ggccctgtgt gagaagactt tgtgtgagtg tgctgggggg ctggagtgcg cctgccctgc	1500
aacgggagac ggtgggagac cacctggggc tcggaaaaca ccgggacaca ctcttctgaa	1560
acacactcac acgaccccc gacctcacgc ggacgggacg cctcctggag tacgcccgga	1620
cctgtgcccc ggagggaatg gtgctgtacg gctggaccga ccacagcgcg tgcagcccag	1680
tgtgccctgc tggatatggag ggaggacctc atgggggcct ggacacgggt cctcccttac	1740
cacgacatgc cgacctggct ggtgtcggcg acgtcgggtc acacgggacg accatacctc	1800
tataggcagt gtgtgtcccc ttgcgccagg acctgccaga gcctgcacat caatgaaatg	1860
tgtcaggagc gatgcgtgga tggctgcagc tgccctgagg atatccgtca cacacagggg	1920
aacgcgggtc tggacgggtc cggacgtgta gttactttac acagtccctg ctacgcacct	1980
accgacgtcg acgggactcc gacagctcct ggatgaaggc ctctgcgtgg agagcaccca	2040
gtgtccctgc gtgcattccg gaaagcgcta cctcccggc acctccctct ctcgagactg	2100
ctgtcgagga cctacttccg gagacgcacc tctcgtggct cacagggacg cacgtaaggc	2160
ctttcgcgat gggagggccg tggagggaga gagctctgac caacacctgc atttgccgaa	2220
acagccagtg gatctgcagc aatgaagaat gtccagggga gtgccttgtc actggtcaat	2280
cccacttcaa gagctttgac gttgtggacg taaacggctt tgtcggtcac ctgacgtcg	2340
ttacttctta caggccccct caccggaacag tgaccagtta ggggaagtct ctcgaaactg	2400
aacagatact tcaccttcag tgggatctgc cagtacctgc tggcccggga ttgccaggac	2460
cactccttct ccattgtcat tgagactgtc cagtgtgctg ttgtctatga agtggaagtc	2520
accttagacg gtcattggac accggggcct aacggtcctg gtgaggaaga ggtaacagta	2580
actctgacag gtcacacgac atgaccgcga cgtgtgtgc acccgctccg tcaccgtccg	2640
gctgcctggc ctgcacaaca gccttgtgaa actgaagcat ggggcaggag ttgccatgga	2700
tactggcgct gcgacacacg tgggcgaggc agtggcaggc cgacggaccg gacgtgttgt	2760
cggaacactt tgacttcgta ccccgctcctc aacggtaact tggccaggac atccagctcc	2820
ccctcctgaa aggtgacctc cgcattccagc atacagtgac ggctccgtg cgcctcagct	2880
acggggagga cctgcagatg accggtcctg taggtcgagg gggaggactt tccactggag	2940
gcgtaggctg tatgtcactg ccggaggcac gcggagtcca tgccctcctt ggacgtctac	3000
gactgggatg gccgcgggag gctgctggtg aagctgtccc ccgtctatgc cgggaagacc	3060
tgccgcctgt gtgggaatta caatggcaac cagggcgacg ctgaccctac cggcgcctc	3120
cgacgaccac ttcgacaggg ggcagatacg gcccttctgg acgccggaca cacccttaat	3180
gttacggttg gtcccgtgc acttccttac cccctctggg ctggcrgagc cccgggtgga	3240
ggacttcggg aacgcctgga agctgcacgg ggactgccag gacctgcaga agcagcacag	3300
tgaaggaatg ggggagaccc gaccgyctg gggccacct cctgaagccc ttgcggacct	3360
tcgacgtgcc cctgacggtc ctggacgtct tcgtcgtgtc cgatccctgc gccctcaacc	3420
cgcgcacgac cagggttctcc gaggaggcgt gcgcggctct gacgtccccc acattcgagg	3480
cctgccatcg tgccgtcagc gctagggacg cgggagttgg gcgcgtactg gtccaagagg	3540
.....	----

ctcctccgca cgcgccagga ctgcaggggg tgtaagctcc ggacggtagc acggcagtcg	3600
ccgctgccct acctgcggaa ctgccgtac gacgtgtgct cctgctcgga cggccgcgag	3660
tgctgtgctg ggcacctggc cagctatgcc gcggcctgct gcgcacggga tggacgcctt	3720
gacggcgatg ctgcacacga ggacgagcct gccggcgctc acggacacgc cgcgggaccg	3780
gtcgatacgg cgcgcgacgc cggggagagg cgtgcgcgtc gcgtggcgcg agccaggccg	3840
ctgtgagctg aactgccga aaggccaggt gtacctgcag tgcgggaccc cctgcaacct	3900
gccccctctcc gcacgcgcag cgcaccgcgc tcggtccggc gacactcgac ttgacgggct	3960
ttccggtcca catggacgtc acgccctggg ggacgttga gacctgccgc tctctctctt	4020
acccgatga ggaatgcaat gaggcctgcc tggagggctg cttctgcccc ccagggtctt	4080
acatggatga gaggggggac ctggacggcg agagagagaa tgggcctact ccttacgtta	4140
ctccggacgg acctccgcac gaagacgggg ggtcccgaga tgtacctact ctccccctg	4200
tgcgtgcca aggccagtg cccctgttac tatgacggtg agatcttcca gccagaagac	4260
atcttctcag accatcacac catgtgctac tgtgaggatg acgcacgggt tccgggtcac	4320
ggggacaatg atactgccac tctagaaggt cggctctctg tagaagagtc tggtagtgtg	4380
gtacacgatg aactcctac gcttcatgca ctgtaccatg agtggagtcc ccggaagctt	4440
gctgcctgac gctgtcctca gcagtcacct gtctcatgc agcaaaagga gcctatcctg	4500
cgaagtacgt gacatggtac tcacctcagg ggccttcgaa cgacggactg cgacaggagt	4560
cgtcagggga cagagtagcg tcgttttctt cggataggac tcggcccccc atggtcaagc	4620
tgggtgtgtcc cgtgacaac ctgcgggctg aagggctcga gtgtaccaa acgtgccaga	4680
actatgacct ggagtgcctg agccgggggg taccagttcg accacacagg gcgactgttg	4740
gacgcccgc tccccgagct cacatggttt tgcacggtct tgatactgga cctcacgtac	4800
agcatgggct gtgtctctgg ctgcctctgc cccccgggca tgggccggca tgagaacaga	4860
tgtgtggccc tggaaaggtg tccctgcttc catcagggca tcgtaccgca cacagagacc	4920
gacggagacg gggggcccgct accaggccgt actcttgtct acacacggg acctttccac	4980
agggacgaag gtagtcccgct aggagtatgc ccctggagaa acagtgaaga ttggctgcaa	5040
cacttgtgtc tgtcgggacc ggaagtggaa ctgcacagac catgtgtgtg atgccacgtg	5100
toctcatacg gggacctctt tgtcacttct aaccgacgtt gtgaacacag acagccctgg	5160
ccttcacctt gacgtgtctg gtacacacac tacgggtgcac ctccacgac gccatggccc	5220
actacctcac cttcgacggg ctcaaatacc tgttccccgg ggagtgccag tacgttctgg	5280
tgcaggatta ctgcggcagt gaggtgctag ccgtaccggg tgatggagtg gaagctgcc	5340
gagtttatgg acaaggggcc cctcacggtc atgcaagacc acgtcctaata gacgccgtca	5400
aaccctggga cctttcggat ctagtgggg aataagggat gcagccaccc ctacgtgaaa	5460
tgcaagaaac gggtcaccat cctggtggag ggaggagaga ttgggacctt ggaaagccta	5520
ggatcacccc ttattcccta cgtcggggg gagtcacttt acgttctttg cccagtggta	5580
ggaccacctc ctcctctctt ttgagctgtt tgacgggggag gtgaatgtga agaggcccat	5640
gaaggatgag actcactttg aggtggtgga gtctggcggg tacatcattc tgctgctggg	5700
aactcgacaa actgcccctc cacttacact tctccgggta cttcctactc tgagtgaac	5760

tccaccacct cagaccggcc atgtagtaag acgacgaccc caaagccctc tccgtggtct	5820
gggaccgcc a cctgagcatc tccgtggtcc tgaagcagac ataccaggag aaagtgtgtg	5880
gcctgtgtgg gaat tttgat gtttcgggag aggcaccaga ccttggcggg ggactcgtag	5940
aggcaccagg acttcgtctg tatggctcctc ttccacacac cggacacacc cttaaaacta	6000
ggcatccaga acaatgacct caccagcagc aacctccaag tggaggaaga ccctgtggac	6060
tttgggaact cctggaaagt gagctcgcag tgtgctgaca ccgtaggtct tgttactgga	6120
gtggtcgtcg ttggagggtc acctccttct gggacacctg aaaccttga ggacctttca	6180
ctcgagcgtc acacgactgt ccagaaaagt gcctctggac tcatccctg ccacctgcca	6240
taacaacatc atgaagcaga cgatggtgga ttccctcctgt agaataccta ccagtgcagt	6300
ggtcttttca cggagacctg agtaggggac ggtggacggt attgtttag tacttcgtct	6360
gctaccacct aaggaggaca tcttaggaat ggtcactgca cttccaggac tgcaacaagc	6420
tggtaggacc cgagccatat ctggatgtct gcatttacga cacctgctcc tgtgagtcca	6480
ttggggactg cgcctgcttc gaaggctcctg acgttgttcg accacctggg gctcgggtata	6540
gacctacaga cgtaaatgct gtggacgagg acaactcaggt aacctctgac gcggacgaag	6600
tgcgacacca ttgctgccta tgcccacgtg tgtgccacgc atggcaaggt ggtgacctg	6660
aggacggcca catttgtccc ccagagctgc gaggagagga acgctgtggt aacgacggat	6720
acgggtgcac acacgggtcg taccgttcca ccaactggacc tcttgcctgt gtaacacggg	6780
ggtctcgacg ctccctcctc atctccggga gaacgggtat gagtgtgagt ggcgtataa	6840
cagctgtgca cctgcctgtc aagtcacgtg tcagcacctc gagccactgg cctgccctgt	6900
tagaggccct cttgcccata ctccactca ccgcgatatt gtcgacacgt ggacggacag	6960
ttcagtgcac agtcgtggga ctgggtgacc ggacgggaca gcagtgtgtg gagggctgcc	7020
atgccactg cctccaggg aaaatcctgg atgagctttt gcagacctgc gttgacctg	7080
aagactgtcc agtgtgtgag cgtcacacac ctcccagcg tacgggtgac gggaggctcc	7140
ttttaggacc tactcgaaaa cgtctggacg caactgggac ttctgacagg tcacacactc	7200
gtggctggcc ggcgttttgc ctgaggaag aaagtcacct tgaatcccag tgacctgag	7260
cactgccaga tttgccactg tgatgttgtc aacctcactc caccgaccgg ccgcaaaacg	7320
gagtcctttc ttccagtgga acttagggtc actgggactc gtgacggtct aaacgggtgac	7380
actacaacag ttggagtgga gtgaagcctg ccaggagcgg ggaggcctgg tgggtgcctcc	7440
cacagatgcc ccgggtgagc ccaccactct gtatgtggag gacatctcgg aaccgctgt	7500
cacttcggac ggtcctcggc cctccggacc accacggagg gtgtctacgg ggccactcgg	7560
ggtggtgaga catacacctc ctgtagagcc ttggcggcaa gcacgatttc tactgcagca	7620
ggctactgga cctggctctc ctgctggatg gctcctccag gctgtccgag gctgagttt	7680
aagtgtgaa ggcctttgtg cgtgctaaag atgacgtcgt ccgatgacct ggaccagaag	7740
gacgacctac cgaggaggtc cgacaggctc cgactcaaac ttcacgactt ccgaaacac	7800
gtggacatga tggagcggct gcgcatctcc cagaagtggg tccgcgtggc cgtggtggag	7860
taccacgacg gctcccacgc ctacatcggg ctcaaggacc cacctgtact acctcgccga	7920
cgcgtagagg gtcttcaccc aggcgcaccg gcaccacctc atggtgctgc cgagggtgcg	7980
gatgtagccc gagttcctgg ggaagcgacc gtcagagctg cggcgcatg ccagccaggt	8040
ccactatgga cggggggggg tggggtggg cggggggggt tggggggggt gggggggggt	8100

gaagcatgcy ggcagccagg tggcctccac cagcgaggcc ttgaaataca cactgtccca	8100
ccttcgctgg cagtctcgac gccgcgtaac ggtcgggtcca cttcatacgc ccgtcgggtcc	8160
accggagggtg gtcgctccag aactttatgt gtgacaaggt aatcttcagc aagatcgacc	8220
gccctgaagc ctcccgcatc gccctgctcc tgatggccag ccaggagccc caacggatgt	8280
cccggaaactt tgtccgctac ttagaagtcg ttctagctgg cgggacttcg gagggcgtag	8340
cgggacgagg actaccggtc ggtcctcggg gttgcctaca gggccttgaa acaggcgatg	8400
gtccaggggc tgaagaagaa gaaggtcatt gtgatcccg tgggcattgg gccccatgcc	8460
aacctcaagc agatccgcct catcgagaag caggccctcg caggccccgg acttcttctt	8520
cttcagtaa cactagggcc acccgtaacc cggggtacgg ttggagtctg tctaggcgga	8580
gtagctcttc gtccggggac agaacaaggc cttcgtgctg agcagtgtgg atgagctgga	8640
gcagcaaagg gacgagatcg ttagctacct ctgtgacctt gccctgaag cccctcctcc	8700
tcttgttccg gaagcacgac tcgtcacacc tactcgacct cgtcgtttcc ctgctctagc	8760
aatcgatgga gacactggaa cggggacttc ggggaggagg tactctgccc cccgacatgg	8820
cacaagtcac tgtgggcccg gggctcttgg gggtttcgac cctggggccc aagaggaaact	8880
ccatggttct ggatgtggcg atgagacggg gggctgtacc gtgttcagtg acaccgggc	8940
cccgagaacc cccaaagctg ggaccccggg ttctccttga ggtaccaaga cctacaccgc	9000
ttcgtcctgg aaggatcgga caaaattggt gaagccgact tcaacaggag caaggagttc	9060
atggaggagg tgattcagcg gatggatgtg ggcaggaca aagcaggacc ttctagcct	9120
gttttaacca cttcggctga agttgtctc gttcctcaag tacctcctcc actaagtcgc	9180
ctacctacac ccggtcctgt gcatccacgt cacgggtgctg cagtactcct acatggtgac	9240
cgtggagtac cccttcagcg aggcacagtc caaaggggac atcctgcagc gggtgcgaga	9300
cgtaggtgca gtgccacgac gtcatgagga tgtaccactg gcacctcatg gggaagtgcg	9360
tccgtgtcag gtttcccctg taggacgtcg cccacgctct gatccgctac cagggcgga	9420
acaggacca cactgggctg gccctgcggt acctctctga ccacagcttc ttggtcagcc	9480
agggtgaccg ggagcaggcg ctaggcgatg gtcccgccgt tgcctgggt gtgaccgac	9540
cgggacgcca tggagagact ggtgtcgaag aaccagtcgg tccactggc cctcgtccgc	9600
cccaacctgg tctacatggt caccggaaat cctgcctctg atgagatcaa gaggtgcct	9660
ggagacatcc aggtggtgcc cattggagtg ggcctaatag gggttggacc agatgtacca	9720
gtggccttta ggacggagac tactctagtt ctccgacgga cctctgtagg tccaccacg	9780
gtaacctcac ccgggattac ccaacgtgca ggagctggag aggattggct ggcccaatgc	9840
ccctatcctc atccaggact ttgagacgct ccccgagag gctcctgacc tgggtctgca	9900
ggttgcaagt cctcgacctc tcccaacga ccgggttacg gggataggag taggtcctga	9960
aactctgcga gggggctctc cgaggactgg accacgacgt gaggtgctgc tccggagagg	10020
ggctgcagat cccaccctc tcccctgcac ctgactgcag ccagcccctg gacgtgatcc	10080
ttctcctgga tggctcctcc ctccacgac aggcctctcc ccgacgtcta ggggtgggag	10140
aggggacgtg gactgacgtc ggtcggggac ctgcactagg aagaggacct accgaggagg	10200
agtttccag cttcttattt tgatgaaatg aagagtttcg ccaaggcttt catttcaaaa	10260
gccaatatag ggcctcgtct cactcagggt tcagtgtgc tcaaagggtc gaagaataaa	10320

actactttac ttotcaaagc ggttccgaaa gtaaagtttt cggttatata ccggagcaga	10380
gtgagtcac agtcacgacg agtatggaag catcaccacc attgacgtgc catggaacgt	10440
gggtcccggag aaagcccatt tgctgagcct tgtggacgtc atgcagcggg agggaggccc	10500
tcataccttc gtagtggtgg taactgcacg gtaccttgca ccagggcctc tttcgggtaa	10560
acgactcgga acacctgcag tacgtcgccc tccctccggg cagccaaatc ggggatgcct	10620
tgggctttgc tgtgcgatac ttgacttcag aaatgcatgg tgccaggccg ggagcctcaa	10680
aggcgggtgt catcctggtc gtcggtttag cccctacgga acccgaaacg acacgctatg	10740
aactgaagtc tttacgtacc acgggtccggc cctcggagtt tccgccacca gtaggaccag	10800
acggacgtct ctgtggattc agtggaatgca gcagctgatg ccgccaggtc caacagagtg	10860
acagtgttcc ctattggaat tggagatcgc tacgatgcag tgccctgcaga gacacctaag	10920
tcacctacgt cgtcgactac ggcgggtccag gttgtctcac tgtcacaagg gataacctta	10980
acctctagcg atgctacgtc cccagctacg gatcttggca ggcgcagcag gcgactccaa	11040
cgtggtgaag ctccagcgaa tcgaagacct cctaccatg gtacaccttg gcaattcctt	11100
gggtcgtatc ctagaacctg ccgggtcgtc cgtcagaggt gcaccacttc gaggtcgtt	11160
agcttctgga gggatggtac cagtggaacc cgttaaggaa cctccacaaa ctgtgctctg	11220
gatttgtagt gatttgcatg gatgaggatg ggaatgagaa gagggccggg gacgtctgga	11280
ccttgccaga ccagtgcac ggaggtgttt gacacgagac ctaaacaatc ctaaacgtac	11340
ctactcctac ccttactctt ctccgggccc ctgcagacct ggaacgggtc ggtcacggtg	11400
accgtgactt gccagccaga tggccagacc ttgctgaaga gtcacgggt caactgtgac	11460
cgggggctga ggccttcgtg ccctaacagc cagtcccttg tggcactgaa cggtcgggtc	11520
accggtctgg aacgacttct cagtagccca gttgacactg gcccccgact ccggaagcac	11580
gggattgtcg gtcaggggac ttaaagtga agagacctgt ggctgccgct ggacctgccc	11640
ctgygtgtgc acaggcagct ccactcggca catogtgacc tttgatgggc agaatttcaa	11700
aatttcacct tctctggaca ccgacggoga cctggacggg gacrcacacg tgtccgtcga	11760
ggtgagccgt gtagcactgg aaactaccog tcttaaagtt gctgactggc agctgttctt	11820
atgtcctatt tcaaaacaag gagcaggacc tggaggtgat tctccataat ggtgcctgca	11880
gccctggagc aaggcagggc cgactgaccg tcgacaagaa tacaggataa agttttgttc	11940
ctcgtcctgg acctccacta agaggtatta ccacggacgt cgggacctcg ttccgtcccg	12000
tgcatgaaat ccacgaggt gaagcacagt gccctctccg tcgagstgca cagtgacatg	12060
gaggtgacgg tgaatgggag actggtctct gttocttacg acgtacttta ggtagctcca	12120
cttcgtgtca cgggagaggc agctcsacgt gtcactgtac ctccactgcc acttaccctc	12180
tgaccagaga caaggaatgc tgggtgggaa catggaagtc aacgtttatg gtgccatcat	12240
gcatgaggtc agattcaatc accttggta catcttcaca ttactccac aaaacaatga	12300
accaccctt gtaccttcag ttgcaaatac cacggtagta cgtactccag tctaagttag	12360
tggaaccagt gtagaagtg aagtgaggtg ttttgttact gttccaactg cagctcagcc	12420
ccaagacttt tgcctcaaag acgtatggtc tgtgtgggat ctgtgatgag aacggagcca	12480
atgacttcat gctgagggat caaggttgac gtcgagtcgg ggttctgaaa acgaagtttc	12540
tgcataccag acacacccta gacactactc ttgcctcggg tactgaagta cgactcccta	12600

ggcacagtca ccacagactg gaaaacactt gttcaggaat ggactgtgca gcggccaggg	12660
cagacgtgcc agcccatcct ggaggagcag tgtottgtcc ccgtgtcagt ggtgtctgac	12720
cttttgtgaa caagtccctta cctgacacgt cgcgggtccc gtctgcacgg tcgggtagga	12780
cctcctcgtc acagaacagg ccgacagctc cactgccag gtcctcctct taccactgtt	12840
tgtgtaatgc cacaaggtcc tggctccagc cacattctat gccatctgcc agcaggacag	12900
ggctgtcgag ggtgacggtc caggaggaga atggtgacaa acgacttacg gtgttccagg	12960
accgaggtcg gtgtaagata cggtagacgg tcgtcctgtc ttgccaccag gagcaagtgt	13020
gtgaggtgat cgcctcttat gccacctct gtccgaccaa cggggtctgc gttgactgga	13080
ggacacctga tttctgtgct aacggtggtc ctcttcaca cactccacta gcggagaata	13140
cgggtggaga cagcctggtt gcccagacg caactgacct cctgtggact aaagacacga	13200
atgtcatgcc caccatctct ggtctacaac cactgtgagc atggctgtcc ccggcactgt	13260
gatggcaacg tgagctcctg tggggaccat ccctccgaag tacagtacgg gtggtagaga	13320
ccagatgttg gtgacactcg taccgacagg ggccgtgaca ctaccgttg cactcaggac	13380
acccctggta gggaggcttc gctgtttctg ccctccagat aaagtcatgt tggaaaggcag	13440
ctgtgtccct gaagaggcct gcactcagtg cattgggtgag gatggagtcc agcaccagtt	13500
cgacaaagac gggagggtcta tttcagtaca accttcctgc gacacaggga cttctccgga	13560
cgtgagtcac gtaaccactc ctacctcagg tcgtgggtcaa cctggaagcc tgggtcccgg	13620
accaccagcc ctgtcagatc tgcacatgcc tcagcgggcg gaagggtcaac tgcacaacgc	13680
agccctgccc caccggccaaa ggaccttcgg acccagggcc tgggtggtcgg gacagtctag	13740
acgtgtacgg agtcgcccgc cttccagttg acgtgttgcg tcgggacggg gtgccggttt	13800
gctcccacgt gtggcctgtg tgaagtagcc cgcctccgcc agaattcaga ccagtgtctc	13860
cccagtatg agtgtgtgtg tgaccacgtg agctgtgacc cgagggtgca caccggacac	13920
acttcacagg gcggaggcgg tcttacgtct ggtcacgacg gggctcatac tcacacacac	13980
actgggtcac tcgacactgg tgccccagt gcctcactgt gaacgtggcc tccagcccac	14040
actgaccaac cctggcgagt gcagaccaa cttcacctgc gcctgcagga aggaggagtg	14100
acgggggtca cggagtgaca cttgcacagg aggtcgggtg tgactgggtg ggaccgctca	14160
cgtctggggt gaagtggacg cggacgtcct tcctcctcac caaaagagtg tccccacct	14220
cctgcccccc gcaccgtttg cccacccttc ggaagaccca gtgctgtgat gagtatgagt	14280
gtgcctgcaa ctgtgtcaac gttttctcac aggggtggga ggacggggg cgtggcaaac	14340
gggtgggaag ccttctgggt cacgacacta ctcatactca cacggacgtt gacacagttg	14400
tccacagtga gctgtccct tgggtacttg gcctcaacgg ccaccaatga ctgtggctgt	14460
accacaacca cctgccttc cgacaagggt tgtgtccacc aggtgtcact cgacaggga	14520
acccatgaac cggagtggc ggtggttact gacaccgaca tgggtgttgtt ggacggaagg	14580
gctgttcac acacagggtg gaagcaccat ctaccctgtg gccagttct gggaggagg	14640
ctgcgatgtg tgcacctgca ccgacatgga ggatgcctg atgggcctcc gcgtggccca	14700
cttcgtggta gatgggacac ccggtcaaga ccctcctccc gacgctacac acgtggacgt	14760
ggctgtacct cctacggcac taccggagg cgcaccgggt gtgctcccag aagccctgtg	14820
aaacacacta tcaatcaaac ttcacttaca ttctacatga aagcaactac tatcaaaact	14880

gcctgccatc tgcctgtgag cacgagggtc ttcgggacac tctgtctgac agccagcccg 14940
 aagtgaatgc aagacgtact tccgctcacg acaccttcca cggacggtag acggacactc 15000
 gtggtgactg gctcaccgcg gggggactcc cagtcttctc ggaagagtgt cggctcccag 15060
 tgggcctccc cggagaaccc ctgcctcatc aatgagtgtg caccactgac cgagtggcgc 15120
 cccctgagg gtcagaagga ccttctcaca gccgagggtc acccggaggg gcctcttggg 15180
 gacggagtag ttactcacac tccgagtga gaggagggtc tttatacaac aaaggaacgt 15240

 ctctgcccc cagctggagg tccctgtctg cccctcgggc tttcagctga gctgtaagac 15300
 aggctcactt cctcctccag aaatatgttg tttccttgca gaggacgggg gtcgacctcc 15360
 agggacagac ggggagcccg aaagtgcact cgacattctg ctacagcgtgc tgcccaagct 15420
 gtcgctgtga gcgcatggag gcctgcatgc tcaatggcac tgtcattggg ccgggaaga 15480
 ctgtgatgat cgatgtgtgc gagtcgcacg acgggttcga cagcgacact cgcgtacctc 15540
 cggacgtacg agttaccgtg acagtaaccc gggcccttct gacactacta gctacacacg 15600
 acgacctgcc gctgcatggt gcaggtgggg gtcctctctg gattcaagct ggagtgcagg 15660
 aagaccacct gcaacccctg cccctcgggt tacaaggaag tgctggacgg cgacgtacca 15720
 cgtccacccc cagtagagac ctaagttcga cctcacgtcc ttctgggtgga cgttggggac 15780
 gggggaccca atgttccttc aaaataacac aggtgaatgt tgtgggagat gtttgcctac 15840
 ggcttgacc attcagctaa gaggaggaca gatcatgaca ctgaagcgtg atgagacgct 15900
 ttttattgtg tccacttaca acacctcta caaacggatg ccgaacgtgg taagtcgatt 15960
 ctctcctgt ctagtactgt gacttcgcac tactctgcga ccaggatggc tgtgatactc 16020
 acttctgcaa ggtcaatgag agaggagagt acttctggga gaagaggggtc acaggctgcc 16080
 cacccttga tgaacacaag ggtcctaccg aactatgag tgaagacgtt ccagttactc 16140
 tctcctctca tgaagacct cttctcccag tgtccgacgg gtgggaaact acttgtgttc 16200
 tgtcttgctg agggaggtaa aattatgaaa attccaggca cctgctgtga cacatgtgag 16260
 gagcctgagt gcaacgacat cactgccagg ctgcagtatg acagaacgac tccctccatt 16320
 ttaatacttt taaggtcctg ggacgacact gtgtacactc ctcgactca cgttgctgta 16380
 gtgacgggtc gacgtcatac tcaagggtgg aagctgtaag tctgaagtag aggtggatat 16440
 ccactactgc cagggcaaat gtgccagcaa agccatgtac tccattgaca tcaacgatgt 16500
 agttccacco ttcgacattc agacttcac tccacctata ggtgatgacg gtcccgttta 16560
 cacggtcggt tcggtacatg aggttaactgt agttgctaca gcaggaccag tgcctctgct 16620
 gctctccgac acggacggag cccatgcagg tggcctgca ctgcaccaat ggctctgttg 16680
 tgtaccatga gggtctcaat cgtcctggtc acgaggacga cgagaggctg tgcctgcctc 16740
 gggtagctcc accgggacgt gacgtgggta ccgagacaac acatgggtact ccaagagtta 16800
 gccatggagt gcaaatgctc cccaggaag tgcagcaagt ga 16842

<210> 28

<211> 2813

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<222> (2016)..(2016)

<223> Xaa can be any naturally occurring amino acid

<400> 28

Met Ile Pro Ala Arg Phe Ala Gly Val Leu Leu Ala Leu Ala Leu Ile
1 5 10 15

Leu Pro Gly Thr Leu Cys Ala Glu Gly Thr Arg Gly Arg Ser Ser Thr
20 25 30

Ala Arg Cys Ser Leu Phe Gly Ser Asp Phe Val Asn Thr Phe Asp Gly
35 40 45

Ser Met Tyr Ser Phe Ala Gly Tyr Cys Ser Tyr Leu Leu Ala Gly Gly
50 55 60

Cys Gln Lys Arg Ser Phe Ser Ile Ile Gly Asp Phe Gln Asn Gly Lys
65 70 75 80

Arg Val Ser Leu Ser Val Tyr Leu Gly Glu Phe Phe Asp Ile His Leu
85 90 95

Phe Val Asn Gly Thr Val Thr Gln Gly Asp Gln Arg Val Ser Met Pro
100 105 110

Tyr Ala Ser Lys Gly Leu Tyr Leu Glu Thr Glu Ala Gly Tyr Tyr Lys
115 120 125

Leu Ser Gly Glu Ala Tyr Gly Phe Val Ala Arg Ile Asp Gly Ser Gly
130 135 140

Asn Phe Gln Val Leu Leu Ser Asp Arg Tyr Phe Asn Lys Thr Cys Gly
145 150 155 160

Leu Cys Gly Asn Phe Asn Ile Phe Ala Glu Asp Asp Phe Met Thr Gln
165 170 175

Glu Gly Thr Leu Thr Ser Asp Pro Tyr Asp Phe Ala Asn Ser Trp Ala
180 185 190

Leu Ser Ser Gly Glu Gln Trp Cys Glu Arg Ala Ser Pro Pro Ser Ser
195 200 205

Ser Cys Asn Ile Ser Ser Gly Glu Met Gln Lys Gly Leu Trp Glu Gln
210 215 220

Cys Gln Leu Leu Lys Ser Thr Ser Val Phe Ala Arg Cys His Pro Leu

225

230

235

240

Val Asp Pro Glu Pro Phe Val Ala Leu Cys Glu Lys Thr Leu Cys Glu
 245 250 255
 Cys Ala Gly Gly Leu Glu Cys Ala Cys Pro Ala Leu Leu Glu Tyr Ala
 260 265 270
 Arg Thr Cys Ala Gln Glu Gly Met Val Leu Tyr Gly Trp Thr Asp His
 275 280 285
 Ser Ala Cys Ser Pro Val Cys Pro Ala Gly Met Glu Tyr Arg Gln Cys
 290 295 300
 Val Ser Pro Cys Ala Arg Thr Cys Gln Ser Leu His Ile Asn Glu Met
 305 310 315 320
 Cys Gln Glu Arg Cys Val Asp Gly Cys Ser Cys Pro Glu Gly Gln Leu
 325 330 335
 Leu Asp Glu Gly Leu Cys Val Glu Ser Thr Glu Cys Pro Cys Val His
 340 345 350
 Ser Gly Lys Arg Tyr Pro Pro Gly Thr Ser Leu Ser Arg Asp Cys Asn
 355 360 365
 Thr Cys Ile Cys Arg Asn Ser Gln Trp Ile Cys Ser Asn Glu Glu Cys
 370 375 380
 Pro Gly Glu Cys Leu Val Thr Gly Gln Ser His Phe Lys Ser Phe Asp
 385 390 395 400
 Asn Arg Tyr Phe Thr Phe Ser Gly Ile Cys Gln Tyr Leu Leu Ala Arg
 405 410 415
 Asp Cys Gln Asp His Ser Phe Ser Ile Val Ile Glu Thr Val Gln Cys
 420 425 430
 Ala Asp Asp Arg Asp Ala Val Cys Thr Arg Ser Val Thr Val Arg Leu
 435 440 445
 Pro Gly Leu His Asn Ser Leu Val Lys Leu Lys His Gly Ala Gly Val
 450 455 460
 Ala Met Asp Gly Gln Asp Ile Gln Leu Pro Leu Leu Lys Gly Asp Leu
 465 470 475 480
 Arg Ile Gln His Thr Val Thr Ala Ser Val Arg Leu Ser Tyr Gly Glu
 485 490 495
 Asp Leu Gln Met Asp Trp Asp Gly Arg Gly Arg Leu Leu Val Lys Leu
 500 505 510
 Ser Pro Val Tyr Ala Gly Lys Thr Cys Gly Leu Cys Gly Asn Tyr Asn
 515 520 525
 Gly Asn Gln Gly Asp Asp Phe Leu Thr Pro Ser Gly Leu Ala Glu Pro
 530 535 540

Arg Val Glu Asp Phe Gly Asn Ala Trp Lys Leu His Gly Asp Cys Gln
 545 550 555 560
 Asp Leu Gln Lys Gln His Ser Asp Pro Cys Ala Leu Asn Pro Arg Met
 565 570 575
 Thr Arg Phe Ser Glu Glu Ala Cys Ala Val Leu Thr Ser Pro Thr Phe
 580 585 590
 Glu Ala Cys His Arg Ala Val Ser Pro Leu Pro Tyr Leu Arg Asn Cys
 595 600 605
 Arg Tyr Asp Val Cys Ser Cys Ser Asp Gly Arg Glu Cys Leu Cys Gly
 610 615 620
 Ala Leu Ala Ser Tyr Ala Ala Ala Cys Ala Gly Arg Gly Val Arg Val
 625 630 635 640
 Ala Trp Arg Glu Pro Gly Arg Cys Glu Leu Asn Cys Pro Lys Gly Gln
 645 650 655
 Val Tyr Leu Gln Cys Gly Thr Pro Cys Asn Leu Thr Cys Arg Ser Leu
 660 665 670
 Ser Tyr Pro Asp Glu Glu Cys Asn Glu Ala Cys Leu Glu Gly Cys Phe
 675 680 685
 Cys Pro Pro Gly Leu Tyr Met Asp Glu Arg Gly Asp Cys Val Pro Lys
 690 695 700
 Ala Gln Cys Pro Cys Tyr Tyr Asp Gly Glu Ile Phe Gln Pro Glu Asp
 705 710 715 720
 Ile Phe Ser Asp His His Thr Met Cys Tyr Cys Glu Asp Gly Phe Met
 725 730 735
 His Cys Thr Met Ser Gly Val Pro Gly Ser Leu Leu Pro Asp Ala Val
 740 745 750
 Leu Ser Ser Pro Leu Ser His Arg Ser Lys Arg Ser Leu Ser Cys Arg
 755 760 765
 Pro Pro Met Val Lys Leu Val Cys Pro Ala Asp Asn Leu Arg Ala Glu
 770 775 780
 Gly Leu Glu Cys Thr Lys Thr Cys Gln Asn Tyr Asp Leu Glu Cys Met
 785 790 795 800
 Ser Met Gly Cys Val Ser Gly Cys Leu Cys Pro Pro Gly Met Val Arg
 805 810 815
 His Glu Asn Arg Cys Val Ala Leu Glu Arg Cys Pro Cys Phe His Gln
 820 825 830
 Gly Lys Glu Tyr Ala Pro Gly Glu Thr Val Lys Ile Gly Cys Asn Thr
 835 840 845

Cys Val Cys Arg Asp Arg Lys Trp Asn Cys Thr Asp His Val Cys Asp
 850 855 860
 Ala Thr Cys Ser Thr Ile Gly Met Ala His Tyr Leu Thr Phe Asp Gly
 865 870 875 880
 Leu Lys Tyr Leu Phe Pro Gly Glu Cys Gln Tyr Val Leu Val Gln Asp
 885 890 895
 Tyr Cys Gly Ser Asn Pro Gly Thr Phe Arg Ile Leu Val Gly Asn Lys
 900 905 910
 Gly Cys Ser His Pro Ser Val Lys Cys Lys Lys Arg Val Thr Ile Leu
 915 920 925
 Val Glu Gly Gly Glu Ile Glu Leu Phe Asp Gly Glu Val Asn Val Lys
 930 935 940
 Arg Pro Met Lys Asp Glu Thr His Phe Glu Val Val Glu Ser Gly Arg
 945 950 955 960
 Tyr Ile Ile Leu Leu Leu Gly Lys Ala Leu Ser Val Val Trp Asp Arg
 965 970 975
 His Leu Ser Ile Ser Val Val Leu Lys Gln Thr Tyr Gln Glu Lys Val
 980 985 990
 Cys Gly Leu Cys Gly Asn Phe Asp Gly Ile Gln Asn Asn Asp Leu Thr
 995 1000 1005
 Ser Ser Asn Leu Gln Val Glu Glu Asp Pro Val Asp Phe Gly Asn
 1010 1015 1020
 Ser Trp Lys Val Ser Ser Gln Cys Ala Asp Thr Arg Lys Val Pro
 1025 1030 1035
 Leu Asp Ser Ser Pro Ala Thr Cys His Asn Asn Ile Met Lys Gln
 1040 1045 1050
 Thr Met Val Asp Ser Ser Cys Arg Ile Leu Thr Ser Asp Val Phe
 1055 1060 1065
 Gln Asp Cys Asn Lys Leu Val Asp Pro Glu Pro Tyr Leu Asp Val
 1070 1075 1080
 Cys Ile Tyr Asp Thr Cys Ser Cys Glu Ser Ile Gly Asp Cys Ala
 1085 1090 1095
 Cys Phe Cys Asp Thr Ile Ala Ala Tyr Ala His Val Cys Ala Gln
 1100 1105 1110
 His Gly Lys Val Val Thr Trp Arg Thr Ala Thr Leu Cys Pro Gln
 1115 1120 1125
 Ser Cys Glu Glu Arg Asn Leu Arg Glu Asn Gly Tyr Glu Cys Glu
 1130 1135 1140

Trp Arg	Tyr Asn Ser Cys	Ala	Pro Ala Cys Gln Val	Thr Cys Gln
1145		1150		1155
His Pro	Glu Pro Leu Ala Cys	Pro Val Gln Cys Val	Glu Gly Cys	
1160		1165	1170	
His Ala	His Cys Pro Pro Gly	Lys Ile Leu Asp Glu	Leu Leu Gln	
1175		1180	1185	
Thr Cys	Val Asp Pro Glu Asp	Cys Pro Val Cys Glu	Val Ala Gly	
1190		1195	1200	
Arg Arg	Phe Ala Ser Gly Lys	Lys Val Thr Leu Asn	Pro Ser Asp	
1205		1210	1215	
Pro Glu	His Cys Gln Ile Cys	His Cys Asp Val Val	Asn Leu Thr	
1220		1225	1230	
Cys Glu	Ala Cys Gln Glu Pro	Gly Gly Leu Val Val	Pro Pro Thr	
1235		1240	1245	
Asp Ala	Pro Val Ser Pro Thr	Thr Leu Tyr Val Glu	Asp Ile Ser	
1250		1255	1260	
Glu Pro	Pro Leu His Asp Phe	Tyr Cys Ser Arg Leu	Leu Asp Leu	
1265		1270	1275	
Val Phe	Leu Leu Asp Gly Ser	Ser Arg Leu Ser Glu	Ala Glu Phe	
1280		1285	1290	
Glu Val	Leu Lys Ala Phe Val	Val Asp Met Met Glu	Arg Leu Arg	
1295		1300	1305	
Ile Ser	Gln Lys Trp Val Arg	Val Ala Val Val Glu	Tyr His Asp	
1310		1315	1320	
Gly Ser	His Ala Tyr Ile Gly	Leu Lys Asp Arg Lys	Arg Pro Ser	
1325		1330	1335	
Glu Leu	Arg Arg Ile Ala Ser	Gln Val Lys Tyr Ala	Gly Ser Gln	
1340		1345	1350	
Val Ala	Ser Thr Ser Glu Val	Leu Lys Tyr Thr Leu	Phe Gln Ile	
1355		1360	1365	
Phe Ser	Lys Ile Asp Arg Pro	Glu Ala Ser Arg Ile	Ala Leu Leu	
1370		1375	1380	
Leu Met	Ala Ser Gln Glu Pro	Gln Arg Met Ser Arg	Asn Phe Val	
1385		1390	1395	
Arg Tyr	Val Gln Gly Leu Lys	Lys Lys Lys Val Ile	Val Ile Pro	
1400		1405	1410	
Val Glv	Ile Glv Pro His Ala	Asn Leu Lys Gln Ile	Arg Leu Ile	

```

1415          1420          1425

Glu Lys  Gln Ala Pro Glu Asn  Lys Ala Phe Val Leu  Ser Ser Val
1430          1435          1440

Asp Glu  Leu Glu Gln Gln Arg  Asp Glu Ile Val Ser  Tyr Leu Cys
1445          1450          1455

Asp Leu  Ala Pro Glu Ala Pro  Pro Pro Thr Leu Pro  Pro Asp Met
1460          1465          1470

Ala Gln  Val Thr Val Gly Pro  Gly Leu Leu Gly Val  Ser Thr Leu
1475          1480          1485

Gly Pro  Lys Arg Asn Ser Met  Val Leu Asp Val Ala  Phe Val Leu
1490          1495          1500

Glu Gly  Ser Asp Lys Ile Gly  Glu Ala Asp Phe Asn  Arg Ser Lys
1505          1510          1515

Glu Phe  Met Glu Glu Val Ile  Gln Arg Met Asp Val  Gly Gln Asp
1520          1525          1530

Ser Ile  His Val Thr Val Leu  Gln Tyr Ser Tyr Met  Val Thr Val
1535          1540          1545

Glu Tyr  Pro Phe Ser Glu Ala  Gln Ser Lys Gly Asp  Ile Leu Gln
1550          1555          1560

Arg Val  Arg Glu Ile Arg Tyr  Gln Gly Gly Asn Arg  Thr Asn Thr
1565          1570          1575

Gly Leu  Ala Leu Arg Tyr Leu  Ser Asp His Ser Phe  Leu Val Ser
1580          1585          1590

Gln Gly  Asp Arg Glu Gln Ala  Pro Asn Leu Val Tyr  Met Val Thr
1595          1600          1605

Gly Asn  Pro Ala Ser Asp Glu  Ile Lys Arg Leu Pro  Gly Asp Ile
1610          1615          1620

Gln Val  Val Pro Ile Gly Val  Gly Pro Asn Ala Asn  Val Gln Glu
1625          1630          1635

Leu Glu  Arg Ile Gly Trp Pro  Asn Ala Pro Ile Leu  Ile Gln Asp
1640          1645          1650

Phe Glu  Thr Leu Pro Arg Glu  Ala Pro Asp Leu Val  Leu Gln Arg
1655          1660          1665

Cys Cys  Ser Gly Glu Gly Leu  Gln Ile Pro Thr Leu  Ser Pro Ala
1670          1675          1680

Pro Asp  Cys Ser Gln Pro Leu  Asp Val Ile Leu Leu  Leu Asp Gly
1685          1690          1695

Ser Ser  Ser Phe Pro Ala Ser  Tyr Phe Asp Glu Met  Lys Ser Phe
1700          1705          1710

```

1700	1705	1710
Ala Lys 1715	Ala Phe Ile Ser 1720	Lys Ala Asn Ile Gly Pro Arg Leu Thr 1725
Gln Val 1730	Ser Val Leu Gln Tyr 1735	Gly Ser Ile Thr Thr Ile Asp Val 1740
Pro Trp 1745	Asn Val Val Pro Glu 1750	Lys Ala His Leu Leu Ser Leu Val 1755
Asp Val 1760	Met Gln Arg Glu Gly 1765	Gly Pro Ser Gln Ile Gly Asp Ala 1770
Leu Gly 1775	Phe Ala Val Arg Tyr 1780	Leu Thr Ser Glu Met His Gly Ala 1785
Arg Pro 1790	Gly Ala Ser Lys Ala 1795	Val Val Ile Leu Val Thr Asp Val 1800
Ser Val 1805	Asp Ser Val Asp Ala 1810	Ala Ala Asp Ala Ala Arg Ser Asn 1815
Arg Val 1820	Thr Val Phe Pro Ile 1825	Gly Ile Gly Asp Arg Tyr Asp Ala 1830
Ala Gln 1835	Leu Arg Ile Leu Ala 1840	Gly Pro Ala Gly Asp Ser Asn Val 1845
Val Lys 1850	Leu Gln Arg Ile Glu 1855	Asp Leu Pro Thr Met Val Thr Leu 1860
Gly Asn 1865	Ser Phe Leu His Lys 1870	Leu Cys Ser Gly Phe Val Arg Ile 1875
Cys Met 1880	Asp Glu Asp Gly Asn 1885	Glu Lys Arg Pro Gly Asp Val Trp 1890
Thr Leu 1895	Pro Asp Gln Cys His 1900	Thr Val Thr Cys Gln Pro Asp Gly 1905
Gln Thr 1910	Leu Leu Lys Ser His 1915	Arg Val Asn Cys Asp Arg Gly Leu 1920
Arg Pro 1925	Ser Cys Pro Asn Ser 1930	Gln Ser Pro Val Lys Val Glu Glu 1935
Thr Cys 1940	Gly Cys Arg Trp Thr 1945	Cys Pro Cys Val Cys Thr Gly Ser 1950
Ser Thr 1955	Arg His Ile Val Thr 1960	Phe Asp Gly Gln Asn Phe Lys Leu 1965
Thr Gly 1970	Ser Cys Ser Tyr Val 1975	Leu Phe Gln Asn Lys Glu Gln Asp 1980
Leu Glu 1985	Val Ile Leu His Asn 1990	Gly Ala Cys Ser Pro Gly Ala Arg 1995

Gln	Gly	Cys	Met	Lys	Ser	Ile	Glu	Val	Lys	His	Ser	Ala	Leu	Ser
2000						2005					2010			
Val	Glu	Xaa	His	Ser	Asp	Met	Glu	Val	Thr	Val	Asn	Gly	Arg	Leu
2015						2020					2025			
Val	Ser	Val	Pro	Tyr	Val	Gly	Gly	Asn	Met	Glu	Val	Asn	Val	Tyr
2030						2035					2040			
Gly	Ala	Ile	Met	His	Glu	Val	Arg	Phe	Asn	His	Leu	Gly	His	Ile
2045						2050					2055			
Phe	Thr	Phe	Thr	Pro	Gln	Asn	Asn	Glu	Phe	Gln	Leu	Gln	Leu	Ser
2060						2065					2070			
Pro	Lys	Thr	Phe	Ala	Ser	Lys	Thr	Tyr	Gly	Leu	Cys	Gly	Ile	Cys
2075						2080					2085			
Asp	Glu	Asn	Gly	Ala	Asn	Asp	Phe	Met	Leu	Arg	Asp	Gly	Thr	Val
2090						2095					2100			
Thr	Thr	Asp	Trp	Lys	Thr	Leu	Val	Gln	Glu	Trp	Thr	Val	Gln	Arg
2105						2110					2115			
Pro	Gly	Gln	Thr	Cys	Gln	Pro	Ile	Leu	Glu	Glu	Gln	Cys	Leu	Val
2120						2125					2130			
Pro	Asp	Ser	Ser	His	Cys	Gln	Val	Leu	Leu	Leu	Pro	Leu	Phe	Ala
2135						2140					2145			
Glu	Cys	His	Lys	Val	Leu	Ala	Pro	Ala	Thr	Phe	Tyr	Ala	Ile	Cys
2150						2155					2160			
Gln	Gln	Asp	Ser	Cys	His	Gln	Glu	Gln	Val	Cys	Glu	Val	Ile	Ala
2165						2170					2175			
Ser	Tyr	Ala	His	Leu	Cys	Arg	Thr	Asn	Gly	Val	Cys	Val	Asp	Trp
2180						2185					2190			
Arg	Thr	Pro	Asp	Phe	Cys	Ala	Met	Ser	Cys	Pro	Pro	Ser	Leu	Val
2195						2200					2205			
Tyr	Asn	His	Cys	Glu	His	Gly	Cys	Pro	Arg	His	Cys	Asp	Gly	Asn
2210						2215					2220			
Val	Ser	Ser	Cys	Gly	Asp	His	Pro	Ser	Glu	Gly	Cys	Phe	Cys	Pro
2225						2230					2235			
Pro	Asp	Lys	Val	Met	Leu	Glu	Gly	Ser	Cys	Val	Pro	Glu	Glu	Ala
2240						2245					2250			
Cys	Thr	Gln	Cys	Ile	Gly	Glu	Asp	Gly	Val	Gln	His	Gln	Phe	Leu
2255						2260					2265			

Glu Ala Trp Val Pro Asp His Gln Pro Cys Gln Ile Cys Thr Cys
 2270 2275 2280
 Leu Ser Gly Arg Lys Val Asn Cys Thr Thr Gln Pro Cys Pro Thr
 2285 2290 2295
 Ala Lys Ala Pro Thr Cys Gly Leu Cys Glu Val Ala Arg Leu Arg
 2300 2305 2310
 Gln Asn Ala Asp Gln Cys Cys Pro Glu Tyr Glu Cys Val Cys Asp
 2315 2320 2325
 Pro Val Ser Cys Asp Leu Pro Pro Val Pro His Cys Glu Arg Gly
 2330 2335 2340
 Leu Gln Pro Thr Leu Thr Asn Pro Gly Glu Cys Arg Pro Asn Phe
 2345 2350 2355
 Thr Cys Ala Cys Arg Lys Glu Glu Cys Lys Arg Val Ser Pro Pro
 2360 2365 2370
 Ser Cys Pro Pro His Arg Leu Pro Thr Leu Arg Lys Thr Gln Cys
 2375 2380 2385
 Cys Asp Glu Tyr Glu Cys Ala Cys Asn Cys Val Asn Ser Thr Val
 2390 2395 2400
 Ser Cys Pro Leu Gly Tyr Leu Ala Ser Thr Ala Thr Asn Asp Cys
 2405 2410 2415
 Gly Cys Thr Thr Thr Thr Cys Leu Pro Asp Lys Val Cys Val His
 2420 2425 2430
 Arg Ser Thr Ile Tyr Pro Val Gly Gln Phe Trp Glu Glu Gly Cys
 2435 2440 2445
 Asp Val Cys Thr Cys Thr Asp Met Glu Asp Ala Val Met Gly Leu
 2450 2455 2460
 Arg Val Ala Gln Cys Ser Gln Lys Pro Cys Glu Asp Ser Cys Arg
 2465 2470 2475
 Ser Gly Phe Thr Tyr Val Leu His Glu Gly Glu Cys Cys Gly Arg
 2480 2485 2490
 Cys Leu Pro Ser Ala Cys Glu Val Val Thr Gly Ser Pro Arg Gly
 2495 2500 2505
 Asp Ser Gln Ser Ser Trp Lys Ser Val Gly Ser Gln Trp Ala Ser
 2510 2515 2520
 Pro Glu Asn Pro Cys Leu Ile Asn Glu Cys Val Arg Val Lys Glu
 2525 2530 2535
 Glu Val Phe Ile Gln Gln Arg Asn Val Ser Cys Pro Gln Leu Glu
 2540 2545 2550

Val	Pro	Val	Cys	Pro	Ser	Gly	Phe	Gln	Leu	Ser	Cys	Lys	Thr	Ser
2555						2560					2565			
Ala	Cys	Cys	Pro	Ser	Cys	Arg	Cys	Glu	Arg	Met	Glu	Ala	Cys	Met
2570						2575					2580			
Leu	Asn	Gly	Thr	Val	Ile	Gly	Pro	Gly	Lys	Thr	Val	Met	Ile	Asp
2585						2590					2595			
Val	Cys	Thr	Thr	Cys	Arg	Cys	Met	Val	Gln	Val	Gly	Val	Ile	Ser
2600						2605					2610			
Gly	Phe	Lys	Leu	Glu	Cys	Arg	Lys	Thr	Thr	Cys	Asn	Pro	Cys	Pro
2615						2620					2625			
Leu	Gly	Tyr	Lys	Glu	Glu	Asn	Asn	Thr	Gly	Glu	Cys	Cys	Gly	Arg
2630						2635					2640			
Cys	Leu	Pro	Thr	Ala	Cys	Thr	Ile	Gln	Leu	Arg	Gly	Gly	Gln	Ile
2645						2650					2655			
Met	Thr	Leu	Lys	Arg	Asp	Glu	Thr	Leu	Gln	Asp	Gly	Cys	Asp	Thr
2660						2665					2670			
His	Phe	Cys	Lys	Val	Asn	Glu	Arg	Gly	Glu	Tyr	Phe	Trp	Glu	Lys
2675						2680					2685			
Arg	Val	Thr	Gly	Cys	Pro	Pro	Phe	Asp	Glu	His	Lys	Cys	Leu	Ala
2690						2695					2700			
Glu	Gly	Gly	Lys	Ile	Met	Lys	Ile	Pro	Gly	Thr	Cys	Cys	Asp	Thr
2705						2710					2715			
Cys	Glu	Glu	Pro	Glu	Cys	Asn	Asp	Ile	Thr	Ala	Arg	Leu	Gln	Tyr
2720						2725					2730			
Val	Lys	Val	Gly	Ser	Cys	Lys	Ser	Glu	Val	Glu	Val	Asp	Ile	His
2735						2740					2745			
Tyr	Cys	Gln	Gly	Lys	Cys	Ala	Ser	Lys	Ala	Met	Tyr	Ser	Ile	Asp
2750						2755					2760			
Ile	Asn	Asp	Val	Gln	Asp	Gln	Cys	Ser	Cys	Cys	Ser	Pro	Thr	Arg
2765						2770					2775			
Thr	Glu	Pro	Met	Gln	Val	Ala	Leu	His	Cys	Thr	Asn	Gly	Ser	Val
2780						2785					2790			
Val	Tyr	His	Glu	Val	Leu	Asn	Ala	Met	Glu	Cys	Lys	Cys	Ser	Pro
2795						2800					2805			
Arg	Lys	Cys	Ser	Lys										
2810														

<211> 4

<212> PRT

<213> Artificial Sequence

<220>

<223> human IgG1

<400> 29

Glu Leu Leu Gly
1

<210> 30

<211> 42

<212> PRT

<213> Artificial Sequence

<220>

<223> XTEN AE42-4

<400> 30

Gly Ala Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly
1 5 10 15

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala
20 25 30

Thr Ser Gly Ser Glu Thr Pro Ala Ser Ser
35 40

<210> 31

<211> 126

<212> DNA

<213> Artificial Sequence

<220>

<223> XTEN AE42-4

<400> 31

ggcgcgccag gttctcctgc tggctcccc acctcaacag aagaggggac aagcgaaagc 60
gctacgcctg agagtggccc tggctctgag ccagccacct ccggtcttga aaccctgcc 120
tcgagc 126

<210> 32

<211> 144

<212> PRT

<213> Artificial Sequence

<220>

<223> XTEN AE144-2A

<400> 32

Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Ser Pro Ala Gly
1 5 10 15

Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly
20 25 30

Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
35 40 45

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr Glu
50 55 60

Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
65 70 75 80

Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly
85 90 95

Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr Glu
100 105 110

Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
115 120 125

Ser Gly Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly
130 135 140

<210> 33

<211> 450

<212> DNA

<213> Artificial Sequence

<220>

<223> XTEN AE144-2A

<400> 33

ggcgcgcaa ccagtacgga gccgtccgag gggagcgcac caggaagccc ggctgggagc	60
ccgacttcta ccgaagaggg tacatctacc gaaccaagtg aaggttcagc accaggcacc	120
tcaacagaac cctctgaggg ctgggcgcct ggtacaagtg agtccgccac ccagagaatcc	180
gggcctggga caagcacaga accttcggaa gggagtgcgc ctggaacatc cgaatcggca	240
acccagaaat cagggccagg atctgagccc gcgacttcgg gctccgagac gcctgggaca	300
tccaccgagc cctccgaagg atcagcccca ggcaccagca cggagccctc tgaggggaagc	360
gcacctggta ccagcgaaag cgcaactccc gaatcaggtc ccggtacgag cgagtcggcg	420
accccgagga gcgggccagg tgcctcgagc	450

<210> 34

<211> 144

<212> PRT

<213> Artificial Sequence

<220>

<223> XTEN AE144-3B

<400> 34

Ser Pro Ala Gly Ser Ser Thr Ser Thr Glu Glu Gly Thr Ser Glu Ser
1 5 10 15

Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser
20 25 30

Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly
35 40 45

Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr Glu
50 55 60

Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly
65 70 75 80

Ser Ala Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
85 90 95

Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Thr Glu
100 105 110

Pro Ser Glu Gly Ser Ala Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser
115 120 125

Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
130 135 140

<210> 35

<211> 450

<212> DNA

<213> Artificial Sequence

<220>

<223> XTEN AE144-3B

<400> 35

ggcgcgccaa gtcccgtctgg aagcccaact agcaccgaag aggggacctc agagtccgcc	60
acccccgagt ccggccctgg ctctgagcct gccactagcg gctccgagac tcctggcaca	120
tccgaaagcg ctacaccga gagtggacct ggcacctcta ccgagcccag tgagggctcc	180
gcccctggaa caagcaccga gccagcgaa ggcagcgccc cagggacctc cacagagccc	240
agtgaaggca gtgctcctgg caccagcacc gaaccaagcg agggctctgc acccgggacc	300
tccaccgagc caagcgaagg ctctgcccct ggcacttcca ccgagcccag cgaaggcagc	360
gcccctggga gcccgtctgg ctctcccacc agcactgagg agggcacatc taccgaacca	420
agtgaaggct ctgcaccagg tgcctcgagc	450

<210> 36

<211> 144

<212> PRT

<213> Artificial Sequence

<220>

<223> XTEN AE144-4A

<400> 36

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala
1 5 10 15

Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
20 25 30

Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly
35 40 45

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr Glu
50 55 60

Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
65 70 75 80

Ser Gly Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly
85 90 95

Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly Ser Pro Ala Gly
100 105 110

Ser Pro Thr Ser Thr Glu Glu Gly Thr Ser Glu Ser Ala Thr Pro Glu
115 120 125

Ser Gly Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
130 135 140

<210> 37

<211> 450

<212> DNA

<213> Artificial Sequence

<220>

<223> XTEN AE144-4A

<400> 37

ggcgcgccaa cgtccgaaag tgctaccctt gagtcaggcc ctggtagtga gcttgccaca	60
agcgggaagcg aaactccggg gacctcagag tctgccactc ccgaatcggg gccaggctct	120
gaaccggcca cttcaggagg cgaaacacca ggaacatcgg agagcgctac cccggagagc	180
gggccaggaa ctagtactga gcctagcgag ggaagtgcac ctggtacaag cgagtccgcc	240
acacccgagt ctggccctgg ctctccagcg ggctcaccga cgagcactga agagggctct	300
cccgtggca gcccaacgtc gacagaagaa ggatcaccag caggctcccc cacatcaaca	360
caggagcta gtcagacta tctactgg cagactgg agctagctc gactagggc	420

gaggagggta caccagaatc tgcactccc gagagtggac ccggaacccc cactgagccc 420
 agcgagggga gtgcaccagg tgcctcgagc 450

<210> 38

<211> 144

<212> PRT

<213> Artificial Sequence

<220>

<223> XTEN AE144-5A

<400> 38

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Ser Glu Pro Ala
 1 5 10 15

Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
 20 25 30

Ser Gly Pro Gly Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly
 35 40 45

Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Thr Glu
 50 55 60

Pro Ser Glu Gly Ser Ala Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser
 65 70 75 80

Thr Glu Glu Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly
 85 90 95

Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
 100 105 110

Ala Thr Pro Glu Ser Gly Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser
 115 120 125

Thr Glu Glu Gly Ser Pro Ala Gly Ser Pro Thr Ser Thr Glu Glu Gly
 130 135 140

<210> 39

<211> 450

<212> DNA

<213> Artificial Sequence

<220>

<223> XTEN AE144-5A

<400> 39

ggcgcgccaa catcagagag cgccaccctt gaaagtggtc ccgggagcga gccagccaca 60

tctgggtcgg aaacgccagg cacaagttag tctgcaactc ccgagtcggg acctggctcc 120

gagcctgcca ctagcggctc cgagactccg ggaacttccg agagcgcctac accagaaagc 180

ggacccggaa ccactaccga acctaggag gactctgctc ggggcagccc acccagctct 240

```

ggccccggaa ccagccccga acccagcgag ggccccgccc cgggagccc agccggccc
cctacatcca cggaggaggg cacttccgaa tccgccaccc cggagtcagg gccaggatct 300
gaacccgcta cctcaggcag tgagacgcca ggaacgagcg agtccgctac accggagagt 360
gggccaggga gccctgctgg atctcctacg tccactgagg aagggtcacc agcgggctcg 420
cccaccagca ctgaagaagg tgcctcgagc 450

```

<210> 40

<211> 144

<212> PRT

<213> Artificial Sequence

<220>

<223> XTEN AE144-6B

<400> 40

```

Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Thr Ser Glu Ser
1           5           10           15

```

```

Ala Thr Pro Glu Ser Gly Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
          20           25           30

```

```

Ser Gly Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Ser Gly Pro Gly
          35           40           45

```

```

Ser Glu Pro Ala Thr Ser Gly Ser Glu Thr Pro Gly Ser Glu Pro Ala
50           55           60

```

```

Thr Ser Gly Ser Glu Thr Pro Gly Ser Pro Ala Gly Ser Pro Thr Ser
65           70           75           80

```

```

Thr Glu Glu Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
          85           90           95

```

```

Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly Ser Glu Pro Ala
100          105          110

```

```

Thr Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu
115          120          125

```

```

Ser Gly Pro Gly Thr Ser Thr Glu Pro Ser Glu Gly Ser Ala Pro Gly
130          135          140

```

<210> 41

<211> 450

<212> DNA

<213> Artificial Sequence

<220>

<223> XTEN AE144-6B

<400> 41

```

ggcgcgccaa catctaccga gccttccgaa ggctctgccc ctgggacctc agaatctgca 60

```

```

accctgaaa gcgccctgg aacctccgaa agtgccactc ccgagagcgg cccagggaca      120
agcgagtcag caaccctga gtctggacct ggcagcgagc ctgcaacctc tggctcagag      180
actcccggtc cagaaccgc tacctcaggc tccgagacac ccggctctcc tgctgggagt      240
cccaattcca ccgaggaagg aacatccact gagcctagtg agggctctgc ccctggaacc      300
agcacagagc caagtgaggg cagtgcacca ggatccgagc cagcaaccag cgggtccgag      360
actcccgga cctctgagtc tgccaccca gagagcggac ccggcacttc aaccgagccc      420
tccgaaggat cagcaccagg tgccctgagc                                     450

```

<210> 42

<211> 144

<212> PRT

<213> Artificial Sequence

<220>

<223> XTEN AG144-1

<400> 42

```

Pro Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Ser Ser
1           5           10          15

```

```

Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Thr Pro Gly Ser Gly Thr
20           25          30

```

```

Ala Ser Ser Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser
35           40          45

```

```

Pro Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Ala Ser
50           55          60

```

```

Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Thr Pro Gly Ser Gly Thr
65           70          75          80

```

```

Ala Ser Ser Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser
85           90          95

```

```

Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ala Ser
100          105         110

```

```

Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser
115          120         125

```

```

Ser Thr Gly Ser Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser

```

130

135

140

<210> 43

<211> 450

<212> DNA

<213> Artificial Sequence

<220>

<223> XTEN AG144-1

<400> 43

```

ggcgcgccac ccgggtcgtc ccgctcggcg tccaccggaa cagggccagg gtcacccccg      60
tcagcgctga ctgggacggg acccgggaca ccggttcgg ggactgcatc ctctcgcct      120
ggttcgtcca ccccgtcagg agccacgggt tcgccgggaa gcagcccaag cgcacccact      180
ggtacagggc ctggggcttc accgggtact tcatccacgg ggtcacggg aacgcccgga      240
tcggggacgg ctctctcatc accaggatcg tcaacacct cgggcgcaac gggcagcccc      300
ggaacccctg gttcgggtac ggcgtcgtcg agccccggtg cgagcccggg aacaagctcg      360
acaggatcgc ctggggcgtc acccggcacg tcgagcacag gcagccccgg aaccctgga      420
tcgggaaccg cgtcgtcaag cgctcgagc      450

```

<210> 44

<211> 144

<212> PRT

<213> Artificial Sequence

<220>

<223> XTEN AG144-A

<400> 44

```

Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Ser Ser Pro
1          5          10          15

Ser Ala Ser Thr Gly Thr Gly Pro Gly Ser Ser Pro Ser Ala Ser Thr
          20          25          30

Gly Thr Gly Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro
          35          40          45

Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Pro
          50          55          60

Ser Ala Ser Thr Gly Thr Gly Pro Gly Ala Ser Pro Gly Thr Ser Ser
65          70          75          80

Thr Gly Ser Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro
          85          90          95

Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Thr Pro Gly
          100          105          110

Ser Gly Thr Ala Ser Ser Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser
          115          120          125

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
          130          135          140

```

<210> 45

<211> 450

<212> DNA

<213> Artificial Sequence

<220>

<223> XTEN AG144-A

<400> 45

```

ggcgcgccag gtgcctcgcc gggaacatca tcaactggtt cacccggtc atccccctcg      60
gcctcaaccg ggacgggtcc cggtcatcc cccagcgcca gactggaac aggtcctggc      120
actcctggtt ccggtacggc atcgtcatcc ccgggaagct caacaccgtc cggagcgaca      180
ggatcacctg gctcgtcacc ttggcgctca actggaacgg ggccaggggc ctccccgga      240
acgtcctcga ctgggtcgcc tggtagccg ggatcaggaa cggcctcatc ctgcctggg      300
tcctcaacgc cctcgggtgc gactgggtcg ccgggaactc ctggctcggg gacggcctcg      360
tcgtcgctg gggcatcacc ggggacgagc tccacggggt cccctggagc gtcaccgggg      420
acctcctcga caggtagccc ggctcgagc

```

<210> 46

<211> 144

<212> PRT

<213> Artificial Sequence

<220>

<223> XTEN AG144-B

<400> 46

```

Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ser Ser Thr
1          5          10          15

Pro Ser Gly Ala Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser
20          25          30

Thr Gly Ser Pro Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro
35          40          45

Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Pro
50          55          60

Ser Ala Ser Thr Gly Thr Gly Pro Gly Ser Ser Pro Ser Ala Ser Thr
65          70          75          80

Gly Thr Gly Pro Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro
85          90          95

Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ala Ser Pro
100         105         110

Gly Thr Ser Ser Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser
115         120         125

```

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
 130 135 140

<210> 47

<211> 450

<212> DNA

<213> Artificial Sequence

<220>

<223> XTEN AG144-B

<400> 47

```

ggcgcgccag gtacacoggg cagcggcacg gcttcgtcgt cacccggtc gtccacaccg      60
tcgggagcta cgggaagccc aggagcgtca ccgggaacgt cgtcaacggg gtcaccgggt      120
acgccaggtg gcggcacggc cagcagctcg ccaggttcat cgaccccgtc gggagcgact      180
gggtcgcccg gatcaagccc gtcagcttcc actggaacag gacccggtc gtcgccgtca      240
gcctcaacgg ggacaggacc tggttcatcg acgccgtcag gggcgacagg ctgcgccgga      300
tcgtcaacac cctcgggggc aacggggagc cctggtgcgt cgcctggaac ctcattcacc      360
ggaagcccgg gggcctcgcc gggtaggagc tccacgggat cgcccgagc gtcccccgga      420
acttcaagca cagggagccc tgcctcgagc                                     450

```

<210> 48

<211> 144

<212> PRT

<213> Artificial Sequence

<220>

<223> XTEN AG144-C

<400> 48

```

Gly Thr Pro Gly Ser Gly Thr Ala Ser Ser Ser Pro Gly Ala Ser Pro
1          5          10          15

Gly Thr Ser Ser Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser
          20          25          30

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
          35          40          45

Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Thr Pro Gly
          50          55          60

Ser Gly Thr Ala Ser Ser Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser
65          70          75          80

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
          85          90          95

Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro Gly Ser Ser Thr
          100          105          110

```

Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Thr Pro Ser Gly Ala
 115 120 125

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
 130 135 140

<210> 49

<211> 450

<212> DNA

<213> Artificial Sequence

<220>

<223> XTEN AG144-C

<400> 49

```

ggcgcgccag gtacacccgg atcgggtaca gcgtcatcga gccccgggtgc gtcacctggt      60
acgtcgagca cggggtcgcc aggggcgtcc cctgggacgt cctcaacagg ctgcccgggt      120
gcgtcaccgg gcacgtcgtc cacgggttca cctggtagct ccccttcgcg gtccactggc      180
accgggcctg gaactccggg gagcggcaca gcgagctcgt cgccgggagc atcgctctgg      240
acatcgagca ccgggtcgcc aggagcatcg cccggaacat ccagcacagg aagccccggc      300
gcgtcgcccc ggacatcaag cacaggttcc ccgggatcga gcacgccgtc cggagccact      360
ggatcaccag ggagctcgac accttcgggc gcaacgggat cgccgggagc cagcccgggt      420
acgtcaagca ctggctcccc tgcctcgagc                                     450

```

<210> 50

<211> 144

<212> PRT

<213> Artificial Sequence

<220>

<223> XTEN AG144-F

<400> 50

```

Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro Gly Ser Ser Pro
1          5          10          15

Ser Ala Ser Thr Gly Thr Gly Pro Gly Ala Ser Pro Gly Thr Ser Ser
          20          25          30

Thr Gly Ser Pro Gly Ala Ser Pro Gly Thr Ser Ser Thr Gly Ser Pro
          35          40          45

Gly Ser Ser Thr Pro Ser Gly Ala Thr Gly Ser Pro Gly Ser Ser Pro
50          55          60

Ser Ala Ser Thr Gly Thr Gly Pro Gly Ala Ser Pro Gly Thr Ser Ser
65          70          75          80

Thr Gly Ser Pro Gly Ser Ser Pro Ser Ala Ser Thr Gly Thr Gly Pro

```

85								90					95				
Gly	Thr	Pro	Gly	Ser	Gly	Thr	Ala	Ser	Ser	Ser	Pro	Gly	Ser	Ser	Thr		
			100				105						110				
Pro	Ser	Gly	Ala	Thr	Gly	Ser	Pro	Gly	Ser	Ser	Thr	Pro	Ser	Gly	Ala		
			115				120						125				
Thr	Gly	Ser	Pro	Gly	Ala	Ser	Pro	Gly	Thr	Ser	Ser	Thr	Gly	Ser	Pro		
			130				135						140				

<210> 51
<211> 450
<212> DNA
<213> Artificial Sequence

<220>
<223> XTEN AG144-F

<400>	51	
ggcgcgccag gctccagccc ctccgcgagc acgggaaccg gaccagggtc gtcaccctca		60
gcatcaacgg ggacgggacc gggggcggtca ccaggaacgt cctccaccgg ctcgccgggt		120
gcatcaccgg gaacgtcatc gaccggatcg ccagggagct cgacgccatc aggcgcaaca		180
ggatcacctg gctcaagccc tagcgcgtca accggcacgg gtccgggtgc ctccccctggc		240
acgtccagca ccggatcacc cggatcgagc ccatccgcct caaccggaac cggaccgggt		300
acaccagggt cggaacagc ctctctgtca ccaggctcct caacccccctc gggagccaag		360
ggttcgcccg gttcgtcaac gccttccgga gcaactggta gcccggagc atcgccagga		420
acttcgagca cggggtcgcc cgcctcgagc		450

REFERENCES CITED IN THE DESCRIPTION

Cited references

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- [WO2011005968A \[0005\]](#)
- [US2009042283A \[0005\]](#)
- [US20100292130A \[0077\]](#)
- [US20080260738A \[0077\]](#)
- [WO2008012543A \[0077\]](#)
- [WO20081439545A \[0077\]](#)
- [US6086875A \[0080\] \[0146\]](#)
- [US6485726B \[0080\] \[0145\] \[0146\]](#)
- [US6030613A \[0080\] \[0146\]](#)
- [WO03077834A \[0080\] \[0146\]](#)
- [US20030235536A1 \[0080\] \[0146\]](#)
- [US2012045784W \[0089\]](#)
- [US6251632B \[0089\]](#)
- [US6316226B \[0091\] \[0091\] \[0092\]](#)
- [US6346513B \[0091\] \[0091\] \[0092\]](#)
- [US7041635B \[0091\] \[0091\] \[0092\]](#)
- [US5789203A \[0091\] \[0091\] \[0092\]](#)
- [US6060447A \[0091\] \[0091\] \[0092\]](#)
- [US5595886A \[0091\] \[0091\] \[0092\]](#)
- [US6228620B \[0091\] \[0091\] \[0092\]](#)
- [US5972885A \[0091\] \[0091\] \[0092\]](#)
- [US6048720A \[0091\] \[0091\] \[0092\]](#)
- [US5543502A \[0091\] \[0091\] \[0092\]](#)
- [US5610278A \[0091\] \[0092\]](#)
- [US5171844A \[0091\] \[0091\] \[0092\]](#)
- [US5112950A \[0091\] \[0091\] \[0092\]](#)
- [US4868112A \[0091\] \[0091\] \[0092\]](#)
- [US6458563B \[0091\] \[0091\]](#)
- [WO919122A \[0091\]](#)
- [WO8704187A \[0091\]](#)
- [WO8800831A \[0091\]](#)
- [EP295597A \[0091\]](#)
- [US4683195A \[0141\]](#)
- [US4683202A \[0141\]](#)
- [US4800159A \[0141\]](#)
- [US4965188A \[0141\]](#)
- [US5658570A \[0141\] \[0223\]](#)
- [WO2012006635A \[0143\]](#)
- [WO8807089A1 \[0154\]](#)
- [WO9614339A1 \[0154\]](#)
- [WO9805787A1 \[0154\]](#)
- [WO9823289A1 \[0154\]](#)
- [WO9951642A1 \[0154\]](#)

- [WO9958572A1 \[0154\]](#)
- [WO0009560A2 \[0154\]](#)
- [WO0032767A1 \[0154\]](#)
- [WO0042072A2 \[0154\]](#)
- [WO0244215A2 \[0154\]](#)
- [WO02060919A2 \[0154\]](#)
- [WO03074569A2 \[0154\]](#)
- [WO04016750A2 \[0154\]](#)
- [WO04029207A2 \[0154\]](#)
- [WO04035752A2 \[0154\]](#)
- [WO04063351A2 \[0154\]](#)
- [WO04074455A2 \[0154\]](#)
- [WO04099249A2 \[0154\]](#)
- [WO05040217A2 \[0154\]](#)
- [WO04044859A \[0154\]](#)
- [WO05070963A1 \[0154\]](#)
- [WO05077981A2 \[0154\]](#)
- [WO05092925A2 \[0154\]](#)
- [WO05123780A2 \[0154\]](#)
- [WO06019447A1 \[0154\]](#)
- [WO06047350A2 \[0154\]](#)
- [WO06085967A2 \[0154\]](#)
- [US20070231329A \[0154\] \[0154\]](#)
- [US20070237765A \[0154\]](#)
- [US20070237766A \[0154\]](#)
- [US20070237767A \[0154\]](#)
- [US20070243188A \[0154\]](#)
- [US20070248603A \[0154\]](#)
- [US20070286859A \[0154\]](#)
- [US20080057056A \[0154\]](#)
- [US5648260A \[0154\]](#)
- [US5739277A \[0154\]](#)
- [US5834250A \[0154\]](#)
- [US5869046A \[0154\]](#)
- [US6096871A \[0154\]](#)
- [US6121022A \[0154\]](#)
- [US6194551B \[0154\]](#)
- [US6242195B \[0154\]](#)
- [US6277375B \[0154\]](#)
- [US6528624B \[0154\]](#)
- [US6538124B \[0154\]](#)
- [US6737056B \[0154\]](#)
- [US6821505B \[0154\]](#)
- [US6998253B \[0154\]](#)

- [US7083784B \[0154\]](#)
- [US7404956B \[0154\]](#)
- [US7317091B \[0154\]](#)
- [US57121A \[0161\]](#)
- [US20090087411A1 \[0162\]](#)
- [US20100239554A1 \[0166\]](#)
- [US20100323956A1 \[0166\]](#)
- [US20110046060A1 \[0166\]](#)
- [US20110046061A1 \[0166\]](#)
- [US20110077199A1 \[0166\]](#)
- [US20110172146A1 \[0166\]](#)
- [WO2010091122A1 \[0166\]](#)
- [WO2010144502A2 \[0166\]](#)
- [WO2010144508A1 \[0166\]](#)
- [WO2011028228A1 \[0166\]](#)
- [WO2011028229A1 \[0166\]](#)
- [WO2011028344A2 \[0166\]](#)
- [US20080194481A1 \[0168\]](#)
- [US20080004206A1 \[0168\]](#)
- [US20080161243A1 \[0168\]](#)
- [US20080261877A1 \[0168\]](#)
- [US20080153751A1 \[0168\]](#)
- [US2008033413A2W \[0168\]](#)
- [US2009058322A1 \[0168\]](#)
- [US2007021494A2 \[0168\]](#)
- [US6696245B \[0171\]](#)
- [US20030069395A \[0171\]](#)
- [WO2010140148A \[0173\]](#)
- [US20100292130A1 \[0176\]](#)
- [WO2008155134A1 \[0176\]](#)
- [US61667901A \[0185\]](#)
- [US4704362A \[0222\]](#)
- [US6159730A \[0223\]](#)
- [US5736137A \[0223\]](#)
- [US6413777B \[0223\]](#)
- [US6193980B \[0224\]](#)
- [US5741957A \[0229\]](#)
- [US5304489A \[0229\]](#)
- [US5849992A \[0229\]](#)
- [WO61765626A \[0305\]](#)

Non-patent literature cited in the description

- RODRIGUEZ-MERCHAN, E.C.Semin. Thromb. Hemost., 2003, vol. 29, 87-96 [0002]
- MANNUCCI, P.M. et al.N. Engl. J. Med., 2001, vol. 344, 1773-9 [0003]
- WHITE G.C. et al.Thromb. Haemost., 1997, vol. 77, 660-7 [0003]
- MORFINI, M.Haemophilia, 2003, vol. 9, 194-99 [0003]
- MANCO-JOHNSON, M.J. et al.N. Engl. J. Med., 2007, vol. 357, 535-544 [0003]
- LYNCH et al.Hum. Gene. Ther., 1993, vol. 4, 259-72 [0004]
- HOEBEN et al.Blood, 1995, vol. 85, 2447-2454 [0004]
- FALLUX et al.Mol. Cell. Biol., 1996, vol. 16, 4264-4272 [0004] [0109]
- KOEBERL et al.Hum. Gene. Ther., 1995, vol. 6, 469-479 [0004]
- TOOLE et al.Proc Natl Acad Sci USA, 1986, vol. 83, 5939-42 [0005]
- ELLMAN et al.Meth. Enzym., 1991, vol. 202, 301-336 [0055]
- NOREN et al.Science, 1989, vol. 244, 182- [0055]
- Computational Molecular BiologyOxford University Press19880000 [0059]
- Biocomputing: Informatics and Genome ProjectsAcademic Press19930000 [0059]
- Computer Analysis of Sequence DataHumana Press19940000 [0059]
- Sequence Analysis in Molecular BiologyAcademic Press19870000 [0059]
- Sequence Analysis PrimerStockton Press19910000 [0059]
- ALTSCHUL et al.J Mol. Biol., 1990, vol. 215, 403- [0059]
- CAPON et al.Nature, 1989, vol. 337, 525- [0080]
- SHARP PMLI WHNucleic Acids Res., 1987, vol. 15, 31281-95 [0082]
- CAMERON et al.Thromb. Haemost., 1998, vol. 79, 317-22 [0089]
- HOEBEN R.C. et al.J. Biol. Chem., 1990, vol. 265, 137318-7323 [0091]
- MEULIEN P. et al.Protein Eng., 1988, vol. 2, 4301-6 [0091]
- TOOLE et al.Proc. Natl. Acad. Sci. U.S.A., 1986, vol. 83, 5939-5942 [0091]
- EATON et al.Biochemistry, 1986, vol. 25, 8343-8347 [0091]
- SARVER et al.DNA, 1987, vol. 6, 553-564 [0091]
- LAGNERBehring Inst. Mitt., 1988, 8216-25 [0091]
- Codon Usage Database, 2012, [0097]
- NAKAMURA, Y. et al.Nucl. Acids Res., 2000, vol. 28, 292- [0097]
- KUDLA et al.PLoS Biol., 2006, vol. 4, 6e180- [0106]
- JANEWAY et al.ImmunobiologyGarland Publishing20010000 [0138]
- PCR Protocols: A Guide to Methods and ApplicationsAcademic Press19900000 [0141]
- HO et al.Gene, 1989, vol. 77, 51- [0141]
- HORTON et al.Methods Enzymol., 1993, vol. 217, 270- [0141]
- BENHARPASTANProtein Engineering, 1994, vol. 7, 1509- [0141]
- RUBERTI, F. et al.J. Immunol. Methods, 1994, vol. 173, 33- [0141]
- LARRICK et al.Biochem. Biophys. Res. Commun., 1989, vol. 160, 1250- [0141]
- STORY et al.J. Exp. Med., 1994, vol. 180, 2377- [0146]
- ISRAEL et al.Immunology, 1997, vol. 92, 69- [0146]
- KOBAYASHI et al.Am. J. Physiol. Renal Physiol., 2002, vol. 282, F358- [0146]
- BURMEISTER et al.Nature, 1994, vol. 372, 379- [0147]
- KABAT et al.Sequences of Proteins of Immunological InterestU.S. Department of Public

Health19910000 [0147]

- **ROUTLEDGE et al.**Transplantation, 1995, vol. 60, 847- [0156]
- **FRIEND et al.**Transplantation, 1999, vol. 68, 1632- [0156]
- **SHIELDS et al.**J. Biol. Chem., 1995, vol. 276, 6591- [0156]
- **SHIELDS et al.**J. Biol. Chem., 2001, vol. 276, 6591- [0156]
- **WARDGHETI**Therapeutic Immunology, 1995, vol. 2, 77- [0157]
- **ARMOUR et al.**Eur. J. Immunol., 1999, vol. 29, 2613- [0157]
- **KONIG, T.SKERRA, A.**J. Immunol. Methods, 1998, vol. 218, 73-83 [0171]
- **DENNIS et al.**J. Biol. Chem., 2002, vol. 277, 35035-35043 [0171] [0172]
- **KRAULIS et al.**FEBS Lett., 1996, vol. 378, 190-194 [0172]
- **LINHULT et al.**Protein Sci., 2002, vol. 11, 206-213 [0172]
- **MULLERKONTERMANN**Curr. Opin. Mol. Ther., 2007, vol. 9, 319-326 [0172]
- **ROOVER et al.**Cancer Immunol. Immunother., 2007, vol. 56, 303-317 [0172]
- **HOLT et al.**Prot. Eng. Design Sci., 2008, vol. 21, 283-288 [0172]
- **TRUSSEL et al.**Bioconjugate Chem., 2009, vol. 20, 2286-2292 [0172]
- **NARITA et al.**Blood, 1998, vol. 91, 555-560 [0180]
- **LENTING et al.**Blood, 1998, vol. 92, 113983-96 [0184]
- **LENTING et al.**J. Thromb. Haemost., 2007, vol. 5, 71353-60 [0184]
- **KRIEGLER, M.**Gene Transfer and Expression, A Laboratory ManualW.H. Freeman Co.19900000 [0205]
- **MURRY, E. J.**Methods in Molecular BiologyHumana Press, Inc.19910000vol. 7, [0205]
- **SAMBROOK et al.**Molecular Cloning: A Laboratory ManualCold Spring Harbor Laboratory Press19890000 [0207]
- **Mammalian Expression VectorsRIDGWAY, A. A. G.**VectorsButterworths19880000470-472 [0210]
- **WIGLER et al.**Cell, 1978, vol. 14, 725- [0218]
- **NEUMANN et al.**EMBO, J., 1982, vol. 1, 841- [0218]
- **BRINSTER et al.**Proc. Natl. Acad.Sci. USA, 1985, vol. 82, 4438- [0219]
- **WAGNER et al.**Proc. Natl. Acad. Sci. USA, 1981, vol. 78, 6376- [0219]
- **MCKNIGHT et al.**Cell, 1983, vol. 34, 335- [0219]
- **BRINSTER et al.**Nature, 1983, vol. 306, 332- [0219]
- **RITCHIE et al.**Nature, 1984, vol. 312, 517- [0219]
- **BALDASSARRE et al.**Theriogenology, 2003, vol. 59, 831- [0219]
- **ROBL et al.**Theriogenology, 2003, vol. 59, 107- [0219]
- **MALASSAGNE et al.**Xenotransplantation, 2003, vol. 10, 3267- [0219]
- **RUTHER et al.**EMBO J., 1983, vol. 2, 1791- [0220]
- **SCOPES**Protein PurificationSpringer-Verlag19820000 [0231]
- **Sustained and Controlled Release Drug Delivery Systems**Marcel Dekker, Inc.19780000 [0237]
- **Remington's Pharmaceutical Sciences**Mack Pub. Co.19800000 [0240]
- **MOUNT et al.**Blood, 2002, vol. 99, 82670- [0271]
- **Antineoplastic AgentsBRUCE A CHABNER et al.**GOODMAN & GILMAN'S THE PHARMACOLOGICAL BASIS OF THERAPEUTICS199600001233-1287 [0278]
- **BURGESS-BROWN et al.**Protein Expr Purif., 2008, vol. 59, 194-102 [0289]

- **GRAF et al.**J. Virol., 2000, vol. 74, 2210822-10826 [\[0292\]](#)

Parentkrav

1. Nukleinsyremolekyle, der omfatter en nukleotidsekvens, der er mindst 90 %, 95 %, 96 %, 97 %, 98 %, 99 % eller 100 % identisk med SEQ ID NO: 1, hvor nukleotidsekvensen koder for et polypeptid med faktor VIII-aktivitet.

2. Nukleinsyremolekyle ifølge krav 1, hvor nukleotidsekvensen koder for B-domænedeleteret FVIII.

3. Nukleinsyremolekyle ifølge krav 1 eller 2, hvor nukleotidsekvensen har en eller flere af følgende egenskaber:

(a) nukleotidsekvensen indeholder en højere procentdel af G/C-nukleotider sammenlignet med SEQ ID NO: 3;

(b) nukleotidsekvensen indeholder færre MARS/ARS-sekvenser sammenlignet med SEQ ID NO: 3;

(c) nukleotidsekvensen indeholder ikke splejsningsstedet GGTGAT (SEQ ID NO: 7);

(d) nukleotidsekvensen indeholder færre destabiliserende elementer (SEQ ID NO: 8 og SEQ ID NO: 9) sammenlignet med SEQ ID NO: 3;

(e) nukleotidsekvensen indeholder ikke en poly-T-sekvens (SEQ ID NO: 10);

(f) nukleotidsekvensen indeholder ikke en poly-A-sekvens (SEQ ID NO: 11); eller

(g) det humane kodontilpasningsindeks er forøget i forhold til SEQ ID NO: 3.

4. Nukleinsyremolekyle ifølge krav 3, hvor

(a) det humane kodontilpasningsindeks er mindst ca. 0,75, mindst ca. 0,76, mindst ca. 0,77, mindst ca. 0,78, mindst ca. 0,79, mindst ca. 0,80, mindst ca. 0,81, mindst ca. 0,82, mindst ca. 0,83, mindst ca. 0,84, mindst ca. 0,85, mindst ca. 0,86, mindst ca. 0,87 eller mindst ca. 0,88;

(b) procentdelen af G/C-nukleotider er mindst ca. 45 %, mindst ca. 46 %, mindst ca. 47 %, mindst ca. 48 %, mindst ca. 49 % eller mindst ca. 50 %;

(c) nukleotidsekvensen indeholder højst én MARS/ARS-sekvens;
eller

(d) nukleotidsekvensen indeholder højst 4 destabiliserende
elementer.

5

5. Nukleinsyremolekyle ifølge et hvilket som helst af kravene 1
til 4, som yderligere omfatter en heterolog nukleotidsekvens.

6. Nukleinsyremolekyle ifølge krav 5, hvor den heterologe
10 nukleotidsekvens koder for en heterolog aminosyresekvens, som
er en halveringstidsforlænger.

7. Nukleinsyremolekyle ifølge krav 6, hvor den heterologe
aminosyresekvens er en immunoglobulin-konstant region eller en
15 del deraf, transferrin, albumin eller en PAS-sekvens.

8. Nukleinsyremolekyle ifølge krav 7, hvor den heterologe
aminosyresekvens er en Fc-region.

20 9. Nukleinsyremolekyle ifølge krav 1, hvor nukleotidsekvensen
omfatter SEQ ID NO: 1.

10. Vektor, der omfatter nukleinsyremolekylet ifølge et hvilket
som helst af kravene 1 til 9.

25

11. Vektor ifølge krav 10, hvor vektoren er valgt blandt en
adenovirusvektor, en lentivirusvektor, en baculovirusvektor, en
Epstein-Barr-virusvektor, en papovavirusvektor, en
vacciniavirusvektor, en herpes simplex-virusvektor og en
30 adenoassocieret virus (AAV) vektor.

12. Værtscelle, der omfatter nukleinsyremolekylet ifølge et
hvilket som helst af kravene 1 til 9 eller vektoren ifølge krav
10 eller 11.

35

13. Fremgangsmåde til produktion af et polypeptid med
faktor VIII-aktivitet, som omfatter: dyrkning af værtscellen
ifølge krav 12 under betingelser, hvorved der produceres et

polypeptid med faktor VIII-aktivitet; og indvinding af polypeptidet med faktor VIII-aktivitet.

14. Nukleinsyremolekyle ifølge et hvilket som helst af kravene 1 til 9 eller vektor ifølge krav 10 eller 11 til anvendelse til behandling af en blødningsforstyrrelse hos et individ, der har brug for det.

DRAWINGS

Drawing

Figure 1

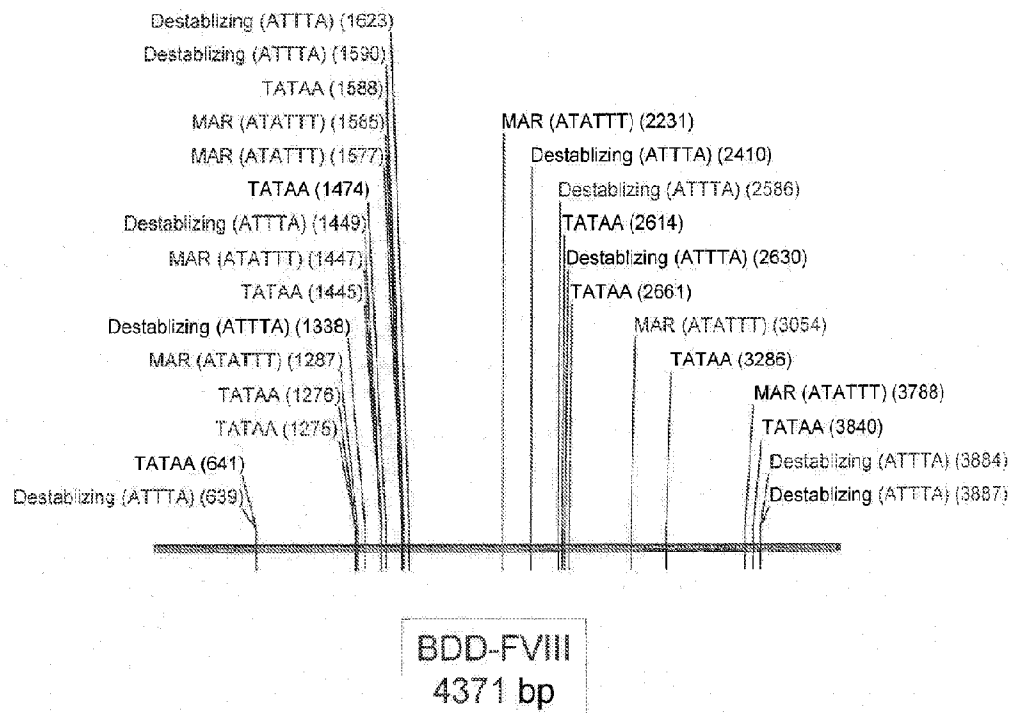


Figure 2

CGTACGGCCGCCACCATGCAGATTGAGCTGTCTACTTGGCTTTTCCGTGTGCCTGCTGAGGTTTTCGCTTTCCGCTACACG
 AAGGTATTATCTGGGGGCTGTGGAACTGTCTTGGGATTACATGCAGAGTGACCTGGGAGAGCTGCCAGTGGACGCAAGGT
 TTCCCCCTAGAGTCCCTAAGTCATTCCCTTCAACACTAGCGTGGTCTACAAGAAAACACGTTCGTGGAGTTTACTGAT
 CACCTGTTCAACATCGCAAGCCCTAGGCCACCTGGATGGGACTGCTGGGSCCAACAATCCAGGCCGAGGTGTACGACAC
 CGTGGTCATTACACTTAAGAACATGGCCTCACACCCCTGAGCCTGCATGCTGTGGGCGTCAGCTACTGGAAGGCTTCCG
 AAGGAGCAGAGTATCAGCATCAGACTTCCCAGAGAGAAAAAGAGGACGATAAGGTGTTTCCCTGGCGGATCTCATACCTAC
 GTGTGGCAGGTCCCTGAAAGAGAAATGGCCCTATGGCCTCCGACCCCTGTGTGCTGACCTACTCTTATCTGAGTCACGTGGA
 CCTGGTCAAGSATCTGAACAGCGGCTGATCGGAGCCCTGCTGGTGTGCAGGGAAGGAAGCCTGGCTAAGGAGAAAAACC
 AGACACTGCATAAGTTTCTTCTGCTGTTTCGCCGTGTTTGAAGCAAGGGAAATCATGGCACAGCGAGACAAAGAAATAGTCTG
 ATGACAGGACAGGGATGCGCCTTCAGCCAGAGCTTGGCCCCAAAATGCACACTGTGAACGGCTACGTCAATCGCTCACTGCC
 TGGGCTGATCGGCTGCCACCGAAAGAGCGTGTATTGGCATGTTCATCGGGATGGGCACACACCTGAAGTGCATCCATTT
 TCCTGGAGGACATACCTTTCTGGTCCGCAACCACCGCAGGCTTCCCTGGAGATCTCTCAATTACCTTCTGACAGCA
 CAGACTCTGCTGATGGACCTGGGCGAGTTCCTGCTGTTTGGCCATCAGCTCCCACAGCATGATGGCATGGAGGCTTA
 CGTAAATACCGATGAGACTCTTCCCGAGGAACCTCAGCTGCGCATGAAGAACATGAGGAAGCAGAGACTATGACCATGAGC
 TGACCGACTCCGAGATGGATGTGGTCCGATTGATGACGATAACAGCCCTCCTTTATCCAGATTAGATCTGTGGCCAG
 AAACACCCCTAAGACATGGGTCCATTACATCGCAGCCGAGGAAGAGGACTGGGATTATGCACCACTGGTGTGGCACCAGA
 CGATCGCTCCTACAAATCTCAGTATCTGAACAATGGGCCACAGAGGATTGGCAGAAAGTACAAGAAAGTGGCGTTCTATGG
 CATATAACCGATGAGACTCTTCCGCAAGCGACTGATCGGACCACTGCTGATTTGTTACAAGAGAGCGTGGATCAGAGAGG
 GTGGGAGACACCCCTGCTGATCATTTTCAAGAACCGAGGCCAGCCGCTTACAATATCTATCCACATGGGATTACAGATGT
 GGGCCCTCTGTACAGCAGGAGACTGCCAAAGGGCGTCAAACACCTGAAGGACTTCCCAATCCTGCCCGGAGAAATCTTCA
 AGTACAAGTGGACTGTACCGTCCGAGGATGGCCCCACTAAGAGCGACCCCTCGGTGCTGACCCGCTACTATTCTAGTTTC
 GTGAATATGGAAAGAGATCTCGCAAGCGACTGATCGGACCACTGCTGATTTGTTACAAGAGAGCGTGGATCAGAGAGG
 CAACCATCATGTCCGACAAGCGGAATGTGATTCTGTTCAAGTGTCTTTGACGAAAACAGGTCATGTTACCTGACCGAGA
 ACATCCAGAGATTCCCTGCCCTAATCCAGCTGGGGTGCAGCTGGAAGATCCTGAGTTTCAGGCATCTAACATCATGCATAGT
 ATTAATGGCTACGTGTTTCGACAGTTTGCAGCTGAGCGTGTGCTGCAAGAGGTGCTTACTGGTATATCTGAGCATTGG
 CATATACCGATGAGACTCTTCCGCGTGTCTTTCCGGCTACACTTTTAAGCATAAAATGGTCTATGAGGACACACTGACTC
 GTTTCCTTTCAGCGCGCAACCGTGTGTTATGAGCATGGAGAATCCCGACTGTGGATTCTGGGGTCCCAACAGGCGAT
 TTCAGAAATCGCGGAATGACTGCCCTGCTGAAAGTGTCAAGCTGTGACAAAGAACACCGGGGACTACTATGAAGATTGATA
 CGAGGACATCAGCGCATATCTGCTGTCCAAAACAAATGCCATTGAACCCCGCTCTTTAGTCAGAAATCCCTCAGTGTGTA
 AGCGGCACACGCTCGAGATCAACCGCACTACCTTCGAGAGTGATCAGGAAGAGATCGACTACGACGATACAAATTTCTGTG
 GAAATGAAGAAAGAGGACTTCGATATCTATGACGAAGATGAGAACCAGAGTCCCTCGATCATTCAGAAAGAAAACAGGCA
 TTACTTTTATTCGCGCAGTGGAGCGGCTGTGGGATTATGGCATGTCTCTAGTCTCTCAGCTGTGCGAAATAGGGCCCGAT
 CAGGAAGCGTCCACAGTTCAAGAAAGTGGTCTTCCAGGAGTTTACAGACGGGTCCCTTACTCAGCCACTGTACAGGGGC
 GAGCTGAACGAGAGCTGCTGGGACTGCTGGGCCCCATATCAGAGCAGAACTGGAGGATAACATTATGGTCACCTTCAGAAA
 TCAGGCTCTCGGCCCTTACAGTTTATTTCAAGCCTGATCTCTTACGAAGAGGACCGAGCAGGGGAGCTGAACACAGAA
 AAAACTTCTGTGAAGCCTAATGAGACCAAAACATACTTTTGAAGGTGCAAGCCATATGGCCCCAACAAAGAGCAGTTTC
 GATTGCAAGGCTAGGGCTATTTTCTCAGCTGGATCTGGAGAAGGAGCTGCACAGTGGCCCTGATTGGCCCACTCTCTGT
 GTGCCATACTAACACCTGAATCCAGCCCCAGGCGGCGGAGGTCACTGTCCAGGAGTTCTGCTCTGTTCTTTACCATCTTTG
 ATGAGACAAAGAGCTGGTACTTACCAGAAAACATGGAGCGAAATTCAGGGCTCCATGTAACATTGAGATGGAAGACCCCC
 ACATTCAAGGAGAACTACCGCTTTCTATGCTATCAATGGATACATCATGGATACTCTGCCCGGCTGGTCTATGGCACAGGA
 CCAGAGAAATCCGGTGGTATCTGCTGAGCATGGGCGAGCAACGAGAAATATCCACTCAATTCAATTCAGCGGGCACGTGTTTA
 CTGTCCAGGAAGAAAGAGAGTACAAGATGGCCCTGTACAACCTGTATCCCGGCGTGTTCGAAACCGTCCAGATGCTGCCCT
 AGCAACCCCGAAATCTGGAGAGTGGAAATGCTGATTGGAGAGCACCTGCATGCTGGGATGTCTACCTCTGTTCTGGTGTGA
 CAGTAATAAGTGTGACACCCCTGGGAATGGCATCCGGGCATATCAGGGATTTCAGATTACCGCATCTGGACAGTACG
 GACAGTGGGCACCTAAGCTGGCTAGACTGCATATTCCGGATCTATCAACGCTTGGTCCACAAAAGAGCCTTTCTCTTGG
 ATTAAGGTGGACCTGCTGGCCCCAATGATCATTTATGGCATCAAACTCAGGGAGCTCGGCAGAAAGTTCTCCCTCTCTGTA
 CATCTCACAGTTTATCATCATGTACAGCCTGGATGGGAAGAAATGGCAGACATACCGCGCAATAGCACAGGAACCTCTGA
 TGGTGTCTTTGGCAACGTGGACAGCAGCGGAATCAAGCACAAACATTTCAATCCCCCTATCATTTGCTAGATACATCCGG
 CTGCACCCCAACCATTTATCTATTGAAAGTACACTGAGGATGGAACTGATGGGATGCGATCTGAACAGTTGTTCAATGCC
 CCTGGGGATGGAGTCCAAGGCAATCTCTGACGCCAGATTACCGCTAGCTCCTACTTCACTAATATGTTTGGTACCTGGGA
 GCCCCTTCCAAGCAAGACTGCACCTGCAAGGCCGAGCAACCGCATGGCAGCACAGGTGAACAATCCAAGGAGTGGTTG
 CAGGTGCGATTTTCAGAAAATATGAAGGTGACCGGGTCACAACTCAGGGCGTGAAGGTCTGCTGACCTCAATGTACCT
 CAGGAGTTCTCTGATCTCTAGTTTCAGAGCAGGACATCAGTGGACACTGTTCTTTCAGAACGGGAAGGTGAAGCTCTTCC
 AGGGCAATCAGGATTCCTTTACACCTGTGGTCAACAGTCTAGACCTCCACTGCTGACAGATACCTGAGAATCCACCT
 CAGTCTGTGGTGCACAGATTGCTCTGAGAAATGGAAGTGTGGGATGCGAGGCCCAGGATCTGTACTGATACTCGAGTC
 GACC

SEQ ID NO:1

Figure 3

GGGCGCGCCGTAACGCGCCACCATGCAAGATCGAGCTGTCTACCTGCTTCTTCTGCTGCTGCTGCGGTTCTGCTTC
 AGCGCCACCCGCGGCTACTCTGCGCGCCGTGGAACTGAGCTGGGACTACATGCGAGAGCGACCTGGGCGAGCTGCC
 CGTGGACGCCAGATTCCCCCAAGAGTGGCCAAAGAGCTTCCCTTCAACACCTCCGTGGTGTACAAGAAAACCTGT
 TCGTGGAGTTCACCGACCCCTGTCAATATCGCCAAAGCCAGACCCCTCGGATGGGCTGCTGGGCTTACAAATC
 CAGGCGGAGGTGTACGACACCTGTGTCATCACCCTTAAGAACATGGCCAGGCCACCCGCTGTCCCTGACGCGCTGGG
 CGTGTCTTACTGGAAGGCTCTGAGGCGCTGAGTACGACGACAGACCCAGCCAGCGCGAGAAAGAGGACGACAAAG
 TCTTTCCTGGCGCAGCCATACCTACGTGTGGCAGGTCTGAAAGAAAACCGCCCTATGGCTCCGACCCCGTGTG
 CTGACCTACAGCTACTCTAGCCACCTGACCTGTGCAAGGACCTGAAACGCGCCCTGATTTGGCGCTGCTGCTGTG
 TAGAGAGGGCAGCTCCCAAAGAGAAAACCCAGACCTTGACAGATTATCTGCTGTTCGCCCTGTTCGACGAGG
 GCAAGAGCTGGCAGCGAGACAAAGAACAGCTGAGTGCAGGACCGGGACCGCGCTCTGCCAGAGCTGGGCTAAG
 ATGCAACCCGTGAAACCGCTACGTGAAACAGAGCTTGGCGGACTGATCGGCTGCCACCGAAGTCCGTGTACTGGCA
 CGTGATCGGCATGGGCACACCCCGAGGTGCACAGCATCTTCTGGAAGGCCACACCTTCTGTCGCGAACCACA
 GACAGGCCAGCTTGGAAATCAGCCCTATCACCCTTCTGACCCCGAGACACTGCTGATGGACCTGGGCTCAGTCTCTG
 CTGTTTGGCCACATCAGCAGCCACCCAGCAGCCGCTGGAAGCTTACGTGAAAGTGGACAGCTGCCCGGAGGAACC
 CCAGCTGGCGATGAAGAACACAGGAGGAGCCGAGGACTTACGACGACGACCTGACCGACGCGAGATGACCTGCTG
 GCTTCGACGACGACAAACAGCCCGAGCTTCATCCAGATCAGAAGCTGGCCAAAGAACACCCCAAGACCTGGGTGCAC
 TATATCGCGCGCGAGGAGAGGAGTCTGGGACTACGCTCTCTGGTCTGGCCCGGACGACAGAAAGTACAAAGGCCA
 GTAACCTGAACAAAGGCCCCAGCGGATCGCCCGGAAGTACAAGAAAGTGGCGTTCTATGCTTACACCGACGAGACAT
 TCAAGACAGAGAGGCCATCCAGCAGGAGAGGCTATCCCTGGGCGCCCTGTGTATGGCGGAAGTGGCGACACCTGT
 CTGATCATCTTCAAGAACACCGGACCGGCTTACAAACATCTACCCCAACGCTACCCGACCTGCGGCTCTGTA
 CAGCAGACCGCTGCCCAAGGCGTGAAGCAGCTGAAGGACTTCCCATCTCTGCCCGCGAGATCTTCAAGTACAGT
 GGACCGTGAACCTTGGAGATGGGCGCCACCAAGAGCGACCCAGATGCTGACCCGCTACACAGCAGCTTCTGTAAC
 ATGGAAACGGACCTGGCTCCGGCTGTGCGCTCTGCTGATCTGCTACAAAGAAAGCGGTGGACACCGCGGCAAA
 CCAGTCTATGAGCGGACAGCAAGCGGACCTGATCTGTTGAGCTGTTCGATGAGAATCGGTCTGCTGATCTGACCGAGA
 ATATCCAGCGGTTCTCTGCCAACCTGCGCGCTGCAGCTGGAAGATCCCGAGTTCCAGGCCAGCAACATCTATGCAC
 TCCATCAATGGCTTACCTGTTCGACAGCTTCCAGCTGAGCTGTGCTGCAACGAGGTGGCTTACTGGTACATCTGAG
 CATCGCGCGCCAGACCGACTTCTGAGCGTGTCTTTCAGCGCTTACACCTTCAAGTACAAGATGGTGTACGAGGATA
 CCGTGAACCTGTCTCCCTTCTCCCGCGGAAACCGTGTTCATGAGCATGGAAACCCCGGCTGTGTGATTTGGGCTGC
 CACAACACCGCATTCAGAAACCGGGCTGACCCGCTGTGTAAGGTGTCCAGCTGCGACAAAGAACACCGGCGACTA
 CTACAGGACAGCTATGAGGACATCAGCGCTTACCTGCTGAGCAGAAACACCCCATCGAGCCCGATCTTTCAGCC
 AGAACCCCGCGCTGTGTAAGCGGACCCAGAGAGAGATCAACCGGACACCTTGCAGTCCGACAGGAGAGATGTAT
 TACGACGACACCATCAGCGTCCAGATGAAGAAAGAGGATTTCCACATCTACGACGAGGACGAGAACCAAGCCCGCC
 GTCTTCCAGAAAGAAAACCGGCGCACTTCTATTGCGCGCTGGAAGAGCTGTGGACTACGCGATGACGAGCGCC
 CCGACGTGTGCGGACACAGCCACAGCGCGGAGGCTGCCAGTTCCAAAGAAAGTGTGTTCCAGGAGTTTCAACGAC
 GGCAGCTTTCACCGAGCCCTGTATCGGGCGGAGCTGAACGAGACCTGGGACTGCTGGGACCTTACATTAGAGCCCA
 GGTGGAAGATACATCATGGTCACTTTCAGAAACAGGCTTCCAGACCTTACAGCTTCTACAGCAGCTGATCAGCT
 ACGAAGAGGACACAGCGCGGCGGCGGAGAACCCGGAAGAACTTGGTGAAGCCCAAGGAGACTAAGACCTTCTCTGG
 AAGGTGCAGCACCATGCGCCCGACAAAGGAGGAGTTCGACTGCAAGGCTGGGCTACTTCTCCGATGTGGACCT
 GGAAGAGGAGCTGCACTCTGCGCTGATGGAGCTCTGCTGTCTGCGCACACCAACACCGGGAACCCCGCCAGGCG
 GGCAGGTTCAGTGCAGGAATTTGCCCTGTTCCTCAGCATCTTCGATGAGACAAAGAGCTGGTACTTCAACGAGAAC
 ATGGAAGAAACTGTAGAGCCCTTCCACATCCAGATGGAAGATCTTACCTTCAAGAGAACTATCCCTTCCACCC
 CATCAACGCGCTACATCATGGACACCTTGGCGGCTTGGTATGGCCAGGATCAGAGAAATCCGCTGGTATCTGCTGA
 GCATGGGCGAGCAGACACATCCACAGCATCCACTTCAAGCGGCGACCTGTTCACAGTTCGGAAGAAAGAGTAC
 AAGATGCCCTGTACAACTTATACCCCGGCTGTTCAGACAGTGGAAATGCTGCCAGCAAGCCCGGCACTTGGCG
 GGTGGAATGTCTGATCGCGGAGCATCTGCACCGCGGAATGAGCAGCTTCTGCTGTACAGCAACAGTGGCAGA
 CCGCTCTGGGCTATGGCGAGCGGCCACATCCCGGACTTCTAGATCACCAGCTTGGGCGAGTACGGCCAGTGGGCGCT
 AAGCTGGGCGGCTTCACTACTCCGATCTATCAACCTGCTGCCACCAAGAGCTTCAAGCTGGATCAAGCTGGA
 CTTGCTGGCCCTATGATCATCAAGGAATCAAGACCCAGGGCGCCAGACAGAAATTCAGCAGGCTGTATCATCAGCC
 AGTTCATCATCATGTACAGCTGGAGCGCAAGAGTGGCAAGCTTACCGGGCAACAGGACCGGCAACCTGATGGTG
 TTCTTCGGCAACCTGGACAGCGGCTATCAAGCACAAGATCTTCAACCCCGCATCTGCGCGTACATCCGGCT
 GCAACCCACCCACTACAGCATCCCGTCCACCTGCGGATGGAATGATGGGCTGCGACCTGAACCTTTCAGCATGC
 CCGTGGGATGGAAGCAAGCCCATCAGCGAGCGCCAGATCAGGCGGAGCAGCTACTTCAACCAACATCTTCCGCAAC
 TGGTCCCCAAGCAAGCCCGCTGCTCTTCAAGGAGAAATGCTTGGCGGCTCAGCTCAACAAACCCCAAGAA
 ATGGCTCCAGGTGAGCTTTCAAGAACCATGAGGCTCAGGCGTGAACCCAGGCGGTGAAAGCCTGCTGACCT
 CTATGTACGTGAAGAGATCTGATCAGCAGCAGCCAGGAGCGGCAACAGTGGACCTGTCTTTCAGAACGGCAAA
 GTGAAAGTGTTCAGGGCAACAGGACTCTTTACCCCGTGGTCAACTCTCTACACCTTCACTGCTGACAGATA
 CCGTGAAGATCCACCTCAGTCTCGGGTGCACAGATTGCCCTGAGAAATGGAAGTGTGGATGCGAGGCCAGGATC
 TGACTGATAACTCGAGTCACTTAATTAA

SEQ ID NO:2

Figure 4

4A: BDD FVIII Before Codon Optimization

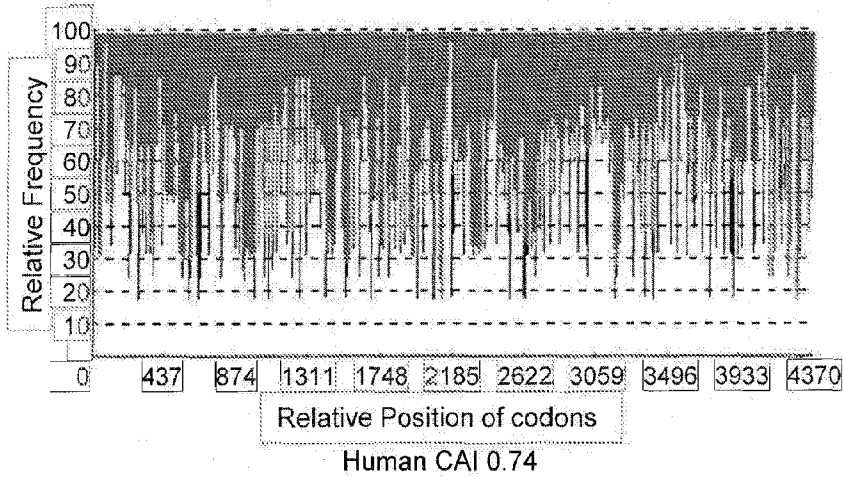
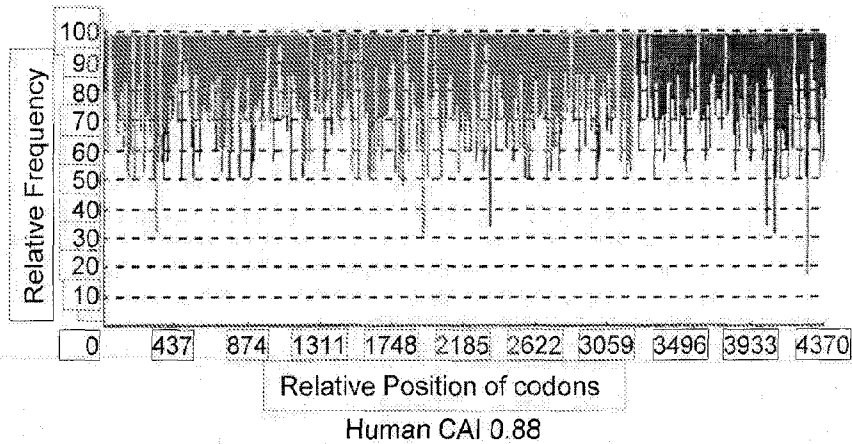
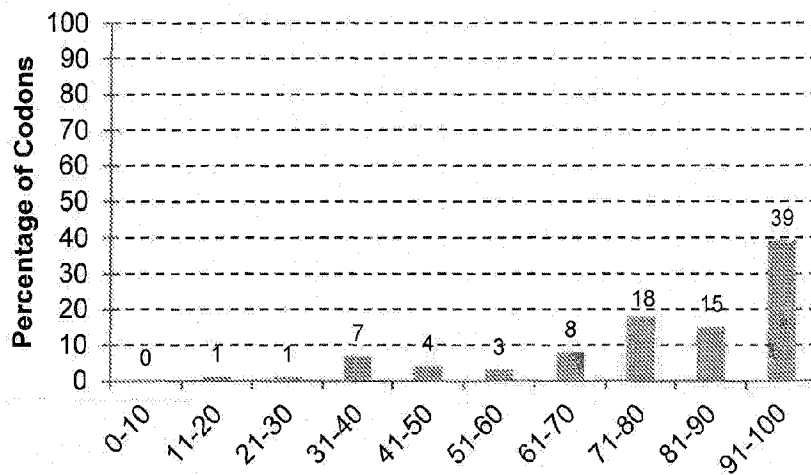
4B: BDD FVIII After Codon Optimization
with GENESCRIPTM OPTIMUMGENETM

Figure 5

5A: BDD FVIII Frequency of Optimal Codons
Before Codon Optimization



5B: BDD FVIII Frequency of Optimal Codons
After Codon Optimization with GENSCRIPT
OPTIMUMGENE™

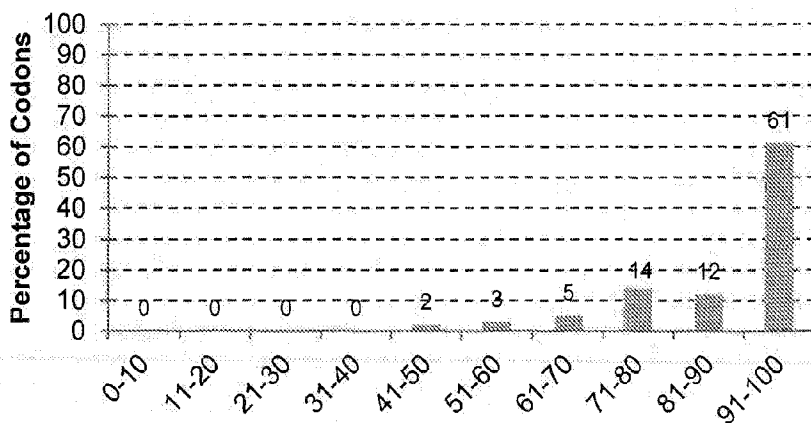
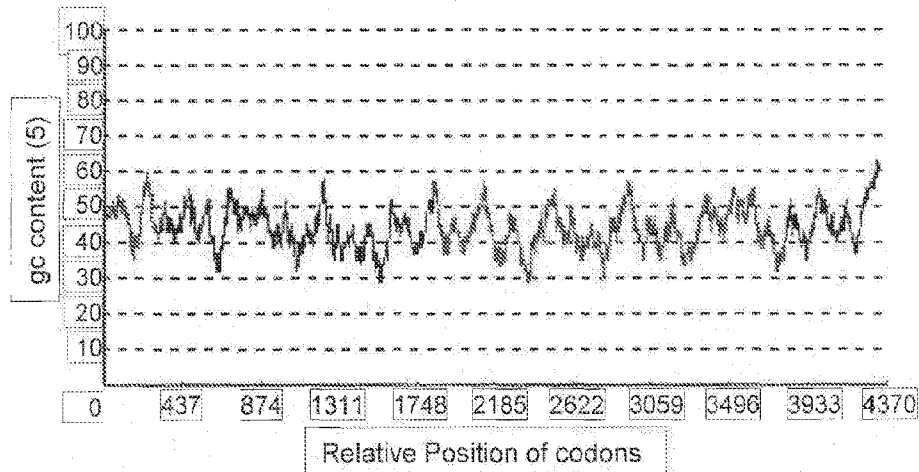


Figure 6

6A: BDD FVIII G/C Content
Before Codon Optimization



6B: BDD FVIII G/C Content After
Codon Optimization with GENSCRIPT
OPTIMUMGENE™

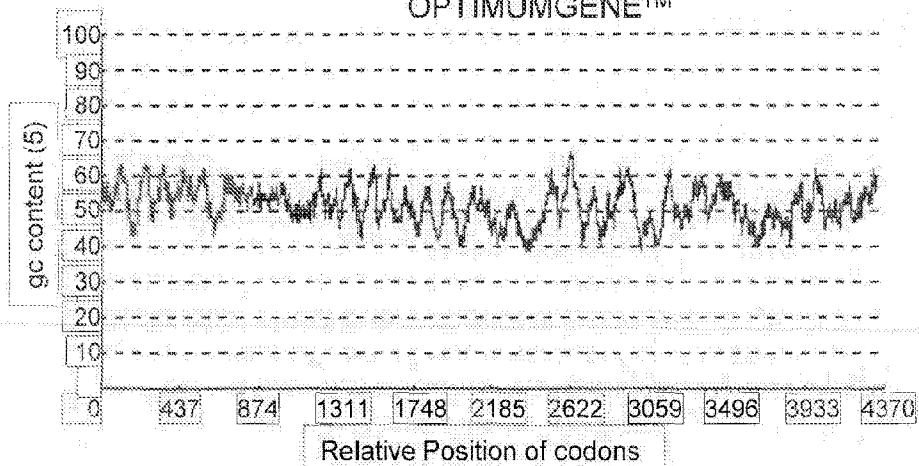


Figure 7

