(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
07 September 2023 (07.09.2023)

WIPO | PCT

(10) International Publication Number
WO 2023/168405 A3

(51) International Patent Classification:
C07K 14/605 (2006.01) *A61K 38/26* (2006.01)
C07K 19/00 (2006.01) *A61K 35/761* (2015.01)

(21) International Application Number:
PCT/US2023/063681

(22) International Filing Date:
03 March 2023 (03.03.2023)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/316,219 03 March 2022 (03.03.2022) US

(71) Applicant: **THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA** [US/US]; 3600 Civic Center Blvd., 9th Floor, Philadelphia, PA 19104 (US).

(72) Inventors: **WILSON, James, M.**; 1831 Delancey Street, Philadelphia, PA 19103 (US). **HINDERER, Christian**; 200 W. Lafayette Avenue, Baltimore, MD 21217 (US). **HORIUCHI, Makoto**; 1049 Putnam Boulevard, Wallingford, PA 19086 (US).

(74) Agent: **SCHALLER, Colleen, M.** et al.; Howson & Howson LLP, 325 Sentry Parkway East, Five Sentry East, Suite 160, Blue Bell, PA 19422 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*

(88) Date of publication of the international search report:
12 October 2023 (12.10.2023)

(54) Title: VIRAL VECTORS ENCODING GLP-2 RECEPTOR AGONIST FUSIONS AND USES THEREOF IN TREATING SHORT BOWEL SYNDROME

(57) Abstract: Compositions and methods for treating short bowel syndrome in a subject are provided. A viral vector is provided which includes a nucleic acid molecule comprising a sequence encoding a GLP-2 receptor agonist fusion protein and regulatory sequences which direct expression thereof.

WO 2023/168405 A3

VIRAL VECTORS ENCODING GLP-2 RECEPTOR AGONIST FUSIONS AND USES THEREOF IN TREATING SHORT BOWEL SYNDROME

REFERENCE TO AN ELECTRONIC SEQUENCE LISTING

The contents of the electronic sequence listing (22-10018PCT_Seq-Listing.xml; Size: 161 kb; and Date of Creation: March 3, 2023) is herein incorporated by reference in its entirety.

BACKGROUND OF THE INVENTION

Short bowel syndrome (SBS) is a rare organ failure caused by surgical resection of bowel due to congenital or acquired reasons. SBS is a major cause of intestinal failure (IF), the persistent reduction of gut function below the minimum necessary for the absorption of macro nutrients and/or water and electrolytes. Standard treatment of SBS/IF includes lifelong daily parenteral nutrition (PN), an intravenous infusion of special form of foods, to support daily required nutrient for patients. In addition to the burden and risk associated to conducting daily PN, PN causes complications including infection, gut hypoplasia, liver diseases, renal dysfunction, and bone diseases leading sever negative impact of the quality of life of SBS patients. Medications used for SBS had been those to treat only symptoms associated with SBS/IF until Teduglutide, a Glucagon-like peptide-2 (GLP-2) agonist, was developed and approved by FDA and EMA.

GLP-2 is a 33 amino acid intestine peptide hormone which has a potent intestinotrophic effect. GLP-2 indeed increases the length of small intestine and villi length of the intestinal epithelium leading to effective absorption of nutrient even with shorter remnant bowel of SBS patients and reduction of PN volume and numbers. However, native GLP-2 cannot be used as an effective injectable with its very short serum half-life due to the cleavage by DPP IV. DPP IV-resistant long acting GLP-2 agonists, including fusions with IgG Fc domain and serum albumin, have been developed as disease modifying agents for SBS. Teduglutide is a GLP-2 like peptide with the A2G mutation for DPP IV resistance, that extends serum half-life from 5 min to 1.5 hours. Daily 0.05 mg/kg subcutaneous Teduglutide injections results in sustained and continuous reduction of PN volume throughout 2 years of treatment. 90% of patients achieved >20% reduction of PN volume per week from the

baseline, and 70% obtained >1 additional days PN off per week from the baseline with this treatment with a great safety profile (Schwartz et al., 2016). Daily Teduglutide injections are required for lifelong to maintain reduced PN volume and numbers (Compher et al., 2011).

What is needed are improved treatments for SBS.

SUMMARY OF THE INVENTION

Viral vectors encoding glucagon-like peptide 2 (GLP-2) receptor agonist fusion protein constructs are provided herein. These viral vectors may achieve, in some embodiments, sustained expression of the GLP-2 receptor agonist in subjects and/or increased circulating half-life, as compared to vector-mediated delivery of a GLP-2 receptor agonist without a fusion partner. Further provided are methods of making and using such viral vectors.

In one aspect, a viral vector is provided which includes a nucleic acid comprising a polynucleotide sequence encoding a fusion protein. The fusion protein includes (a) a leader sequence comprising a secretion signal peptide, (b) a glucagon-like peptide-1 (GLP-2) receptor agonist, and (c) a fusion domain comprising (i) an IgG Fc or a functional variant thereof, (ii) an albumin or a functional variant thereof, or (iii) an XTEN polypeptide (Podust et al, 240:52-66 (Oct 2016). In one embodiment, the vector is an adeno-associated viral vector.

In one embodiment, the (i) the secretion signal peptide of the leader sequence comprises a thrombin signal peptide; (ii) the leader sequence comprises a thrombin propeptide; and/or (iii) the leader sequence comprises a thrombin leader sequence. In another embodiment, the leader sequence comprises an IL-2 leader sequence. In one embodiment, the GLP-fusion is selected from SEQ ID NO: 13, 15, 17, 19, 21, or 23, and functional variants thereof.

In one embodiment, the fusion domain is a human IgG4 Fc having the sequence of SEQ ID NO: 8, or a sequence sharing at least 90% identity therewith, or a functional variant thereof. In another embodiment, the fusion domain is a human albumin having the sequence of SEQ ID NO: 11, or a sequence sharing at least 90% identity therewith, or a functional variant thereof. In one embodiment, the fusion domain is a rhesus IgG4 Fc having the

sequence of SEQ ID NO: 9, or a sequence sharing at least 90% identity therewith, or a functional variant thereof.

In another aspect, the viral vector includes an AAV capsid, and a vector genome packaged in the AAV capsid, said vector genome comprising AAV inverted terminal repeats (ITRs), the polynucleotide sequence encoding the fusion protein, and regulatory sequences which direct expression of the fusion protein.

In another aspect, a pharmaceutical composition suitable for use in treating short bowel syndrome in a subject is provided. The composition includes an aqueous liquid and the viral vector as described herein. In one embodiment, the subject is a human.

In yet another aspect, use of a viral vector as described herein is provided for the manufacture of a medicament for treating a subject having short bowel syndrome, optionally diabetes.

In another aspect, a method of treating a subject having short bowel syndrome is provided. The method includes administering to the subject an effective amount of a viral vector or composition as described herein,

Other aspects and advantages of the invention will be readily apparent from the following detailed description of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a schematic of the processing of proglucagon in vivo. The sequences of GLP-1 (SEQ ID NO: 4) and GLP-2 (SEQ ID NO: 1) are shown.

FIG. 2 is a table showing sequences of GLP-2 and analogs.

FIG. 3A is a schematic of the GLP2.G2.Fc construct, which includes an A2G substitution in the GLP2 amino acid sequence, a thrombin leader, and an IgG4 Fc fusion.

FIG. 3B is a schematic of the GLP2.G2.SA construct, which includes an A2G substitution in the GLP2 amino acid sequence, a thrombin leader, and a serum albumim fusion.

FIG. 3C is an alignment of the human (SEQ ID NO: 1), cyno (SEQ ID NO: 5), and mouse (SEQ ID NO: 6) GLP2 sequences.

FIG. 4A shows an elution profile of hGLP2-SA protein using albupure column purification.

FIG. 4B shows a gel with stained hGLP2-SA protein from FIG. 4A.

FIG. 5A shows hGLP2-SA levels in RagKO mice dosed intramuscularly with AAVrh91.CI.hGLP2.G2.SA.rBG at 1e11 gc/mouse.

FIG. 5B are two graphs showing small intestine length and weight for vehicle and vector treated mice as described in FIG. 5A.

FIG. 5C shows intestine histology for vehicle and vector treated mice as described in FIG. 5A. Vector treated mice intestine show healthier villi as compared to vehicle treated mice.

FIG. 6A shows hGLP2-SA and hGLP2-Fc levels in mice treated with AAVrh91.CI.hGLP2.G2.SA.rBG or AAVrh91.CI.hGLP2.G2.Fc.rBG at a dosage of 1e11 gc/mouse. GLP-2 levels for the Fc construct were higher after about study day 7.

FIG. 6B shows a potency assay for hGLP2 and hGLP2.Fc. EC50 was determined to be 0.4nM for hGLP2 while EC=13.6nM for the fusion protein.

FIG. 7 shows the study design for an experiment as described in Example 4.

FIGs. 8A-8C show the results of the experiment as described in Example 4. 2 NHPs were administered AAVrh91.CB7.CI.hGLP-2-Fc.rBG via intramuscular injection (IM) at a dose of 1 x 10^{13} (1e13) GC/kg (E185NG) and a of dose 5 x 10^{10} (1e10) GC/kg (BM239H). FIG. 8A shows plasma level of GLP-2-Fc fusion protein. FIG. 8B shows serum citrulline, a biomarker of gut surface area. FIG. 8C shows detection of anti-GLP-2-Fc antibody in NHP serum at 1:100 dilution.

FIGs. 9A-9F show the results of the experiment as described in Example 5. Rag1KO female mice were treated with an injection of the vector AAVrh91.CB7.CI.hGLP-2-Fc.rBG at a dosage of 1 x 10^{10} GC/ mouse, 3 x 10^{10} GC/ mouse, or 1 x 10^{11} GC/ mouse via IM route of administration. The study design is shown in FIG. 9A. FIG. 9B shows serum GLP2 levels, FIG. 9C shows body weights over time. FIG. 9D shows body weights at day 56. FIG. 9E shows small intestine (SI) length, while FIG. 9F shows SI weight.

DETAILED DESCRIPTION OF THE INVENTION

Described herein are adeno-associated viral (AAV) vectors expressing GLP-2 agonists to treat SBS/IF patients with a single intramuscular vector administration. Transgene GLP-2 agonists include the A2G mutation for DPP IV resistance and fusions with human IgG Fc domain or serum albumin for further extended serum half-life. In combination with these

half-life extension technologies, the addition of a thrombin propeptide enables expression of GLP-2 agonists above the therapeutic level with remarkably lower vector doses (i.e., 1e10 to 1e12 GC/kg). Described herein are expression cassettes to express these proteins constitutively or in a controlled manner via administration of a small molecule drug that activates transcription of the GLP-2 agonist sequence.

Delivery of these constructs to subjects in need thereof via a number of routes, and particularly by expression *in vivo* mediated by a recombinant vector such as a rAAV vector, is described. Also provided are methods of using these constructs in regimens for treating short bowel syndrome in a subject in need thereof and increasing the half-life of GLP-2 in a subject. In addition, methods are provided for enhancing the activity of GLP-2 in a subject.

GLP-2 Fusion Proteins

Post-translational processing of proglucagon generates glucagon-like peptide-2 (GLP-2), a 33-amino acid intestinotrophic peptide hormone, together with GLP-1. GLP-2 acts to slow gastric emptying, reduce gastric secretions and increase intestinal blood flow. GLP-2 also stimulates growth of the large and small intestine at least by enhancing crypt cell proliferation and villus length so as to increase the surface area of the mucosal epithelium. These effects suggest that GLP-2 can be used to treat a wide variety of gastrointestinal conditions. However, administering GLP-2 by itself to human patients has not shown promise. GLP-2 has a short half-life that limits its use as a therapeutic because rapid in vivo cleavage of GLP-2 by dipeptidyl peptidase IV (DPP-IV) yields an essentially inactive peptide. The amino acid sequence of human GLP-2 is HADGSFSDEMNTILDNLAARDFINWLIQTKITD (SEQ ID NO: 1).

As discussed above, a GLP-2 analog, named teduglutide has been developed, in which amino acid residue 2 (alanine) has been substituted with glycine. The sequence of this GLP-2 analog is shown in SEQ ID NO: 2 HGDGSFSDEMNTILDNLAARDFINWLIQTKITD.

The disclosure provides fusion proteins comprising one or more copies of a GLP-2 receptor agonist, as well as polynucleotides and vectors encoding such fusion proteins. In some embodiments, the fusion protein comprises a polynucleotide sequence encoding a fusion protein comprising (a) a leader sequence comprising a secretion signal peptide, (b) a

glucagon-like peptide-2 (GLP-2) receptor agonist, and (c) a fusion domain. In one embodiment, the GLP-2 receptor agonist comprises a thrombin leader sequence, a GLP-2 receptor agonist, and an IgG Fc or functional variant thereof. In another embodiment, the fusion protein comprises a thrombin leader, a GLP-2 receptor agonist, and an albumin or functional variant thereof. In another embodiment, the fusion protein comprises a thrombin leader, two copies of a GLP-2 receptor agonist, and an albumin or functional variant thereof.

In some embodiments, GLP-2 receptor agonists include variants which may include up to about 10% variation from a GLP-2 nucleic acid or amino acid sequence described herein or known in the art, which retain the function of the wild-type sequence. As used herein, by "retain function" it is meant that the nucleic acid or amino acid functions in the same way as the wild-type sequence, although not necessarily at the same level of expression or activity. For example, in one embodiment, a functional variant has increased expression or activity as compared to the wild-type sequence. In another embodiment, the functional variant has decreased expression or activity as compared to the wild-type sequence. In one embodiment, the functional variant has 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% or greater increase or decrease in expression or activity as compared to the wild-type sequence.

Other GLP-2 analogs have been developed and include glepaglutide (SEQ ID NO: 3), apraglutide, and others shown below, and in FIG. 2.

Drug	Type	Company	Stage	T1/2 (h)	EC50 (/GLP2)	Comments
Teduglutide	Peptide	Shire	Approved	1.5 (0.1)	1	
Glepaglutide	Peptide	Zealand	P3	50	4	
Apraglutide	Peptide	Ferring	P3	30	0.5	
HM-15912	Fc fusion	Hanmi	P2	?	5	
GX-G8	Fc fusion	Genexine	P1	?	?	
TAK-681	Fc fusion	Takeda (Shire)	P1	100	5	
ZP7570	Peptide	Zealand	P1	?	10	Dual acting
NB-1002	XTEN fusion	Amunix	Pre	240	100	

The fusion comprises, in one embodiment, a GLP-2 analog in combination with heterologous sequences. By GLP-2 analog is meant a polypeptide sharing at least 90%, 95%, 97%, 98%, 99% or 100% identity with native human GLP-2 (SEQ ID NO: 1) or GLP-2-A2G (SEQ ID NO: 2). In one embodiment, the GLP-2 analog has at most 1, 2, or 3 amino acid substitutions as compared to the native sequence. In another embodiment, the GLP-2 sequence is derived from a species other than human. For example, the GLP-2 may be from a non-human primate, dog, cat, mouse, rat, sheep, cow, horse, etc. In some embodiments, it is

desirable to alter the native GLP-2 sequence to optimize one or more features thereof. In certain embodiments, the GLP-2 has a sequence which amino acid residue 2 (alanine) has been substituted with glycine, e.g., SEQ ID NO: 2. For example, in other embodiments, the GLP-2 analog contains one, two, three, 4, 5, 6, 7, 8 or up to 9 amino acid substitutions selected from A2G, D3E, S5T, D8S, M10L, N11A, N16A, N24A, Q28A as compared to the native sequence. These substitutions have been shown to improve efficacy of the clinical profile of GLP-2, including protection from DPP-4 inactivation (A2G). In one embodiment, the GLP-2 analog is a DPP-IV resistant variant of GLP-2. In one embodiment, the GLP-2 analog has a sequence comprising, or consisting of, SEQ ID NO: 2. In one embodiment, the variant shares at least 90% identity, 95% identity, 97% identity, 98% identity, 99% identity or 100% identity with SEQ ID NO: 2.

The fusion protein may comprise a leader sequence, which may comprise a secretion signal peptide. As used herein, the term "leader sequence" refers to any N-terminal sequence of a polypeptide.

The leader sequence may be derived from the same species for which administration is ultimately intended, e.g., a human. As used herein, the terms "derived" or "derived from" mean the sequence or protein is sourced from a specific subject species or shares the same sequence as a protein or sequence sourced from a specific subject species. For example, a leader sequence which is "derived from" a human, shares the same sequence (or a variant thereof, as defined herein) as the same leader sequence as expressed in a human. However, the specified nucleic acid or amino acid need not actually be sourced from a human. Various techniques are known in the art which are able to produce a desired sequence, including mutagenesis of a similar protein (e.g., a homolog) or artificial production of a nucleic acid or amino acid sequence. The "derived" nucleic acid or amino acid retains the function of the same nucleic acid or amino acid in the species from which it is "derived", regardless of actual source of the derived sequence.

The term "amino acid substitution" and its synonyms are intended to encompass modification of an amino acid sequence by replacement of an amino acid with another, substituting, amino acid. The substitution may be a conservative substitution. It may also be a non-conservative substitution. The term conservative, in referring to two amino acids, is intended to mean that the amino acids share a common property recognized by one of skill in

the art. For example, amino acids having hydrophobic nonacidic side chains, amino acids having hydrophobic acidic side chains, amino acids having hydrophilic nonacidic side chains, amino acids having hydrophilic acidic side chains, and amino acids having hydrophilic basic side chains. Common properties may also be amino acids having hydrophobic side chains, amino acids having aliphatic hydrophobic side chains, amino acids having aromatic hydrophobic side chains, amino acids with polar neutral side chains, amino acids with electrically charged side chains, amino acids with electrically charged acidic side chains, and amino acids with electrically charged basic side chains. Both naturally occurring and non-naturally occurring amino acids are known in the art and may be used as substituting amino acids in embodiments. Methods for replacing an amino acid are well known to the skilled in the art and include, but are not limited to, mutations of the nucleotide sequence encoding the amino acid sequence. Reference to "one or more" herein is intended to encompass the individual embodiments of, for example, 1, 2, 3, 4, 5, 6, or more.

In one embodiment, the leader is a human thrombin (Factor II) sequence. In one embodiment, the thrombin leader has the sequence shown in SEQ ID NO: 7: MAHVRGLQLPGCLALAALCSLVHSQHVFLAPQQARSLLQRVRR, or a functional variant thereof having at most 1, 2, or 3 amino acid substitutions.

In one embodiment, functional variants of the desired leader include variants which may include up to about 10% variation from a leader nucleic acid or amino acid sequence described herein or known in the art, which retain the function of the wild type sequence.

In some embodiments, the coding regions for both the propeptide and GLP-2 peptide are incorporated into a single nucleic acid sequence without a linker between the coding sequences of the propeptide and GLP-2.

The fusion protein further includes a fusion domain. The fusion domain, in one embodiment, is a human IgG Fc fragment or a functional variant thereof. Immunoglobulins typically have long circulating half-lives *in vivo*. By fusing the GLP-2 receptor agonist (and leader) to an IgG Fc, the circulation time of the fusion protein is prolonged, while the function of the GLP-2 is preserved. In another embodiment, the fusion domain is a rhesus IgG Fc fragment or functional variant thereof.

As used herein, the Fc portion of an immunoglobulin has the meaning commonly given to the term in the field of immunology. Specifically, this term refers to an antibody

fragment which does not contain the two antigen binding regions (the Fab fragments) from the antibody. The Fc portion consists of the constant region of an antibody from both heavy chains, which associate through non-covalent interactions and disulfide bonds. The Fc portion can include the hinge regions and extend through the CH2 and CH3 domains to the c-terminus of the antibody. The Fc portion can further include one or more glycosylation sites. In one embodiment, the fusion domain is a human IgG Fc. The four subclasses, IgG1, IgG2, IgG3, and IgG4, which are highly conserved, differ in their constant region, particularly in their hinges and upper CH2 domains. See, Vidarsson et al, IgG Subclasses and Allotypes: From Structure to Effector Functions, Front Immunol. Oct. 2014; 5: 520, which is incorporated herein by reference. The Fc domain can be derived from any human IgG, including human IgG1, human IgG2, human IgG3, or human IgG4. In one embodiment, the human IgG Fc is an IgG4 Fc. In one embodiment, the human IgG Fc is SEQ ID NO: 8:

AESKYGPPCPPCPAPEAAGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSQEDP EVQFNWYVDGVEVHNAKTKPREEQFNSTYRVVSVLTVLHQDWLNGKEYKCKVSN KGLPSSIEKTISKAKGQPREPQVYTLPPSQEEMTKNQVSLTCLVKGFYPSDIAVEWES NGQPENNYKTTPPVLDSDGSFFLYSRLTVDKSRWQEGNVFSCSVMHEALHNHYTQK SLSLSLG. In another embodiment, the human IgG Fc shares at least 90% identity, at least 95% identity, at least 99% identity, or at least 100% identity to SEQ ID NO: 8.

In another embodiment, the fusion domain is a rhesus IgG Fc. The Fc domain can be derived from any rhesus IgG, including rhesus IgG1, rhesus IgG2, rhesus IgG3, or rhesus IgG4. In one embodiment, the rhesus IgG Fc is an IgG4 Fc. In one embodiment, the rhesus IgG Fc is SEQ ID NO: 9:

PPCPPCPAPE LLGGPSVFLF PPKPKDTLMI SRTPEVTCVV VDVSQEDPEV QFNWYVDGVE VHNAQTKPRE RQFNSTYRVV SVLTVTHQDW LNGKEYTCKV SNKGLPAPIE KTISKAKGQP REPQVYILPP PQEELTKNQV SLTCLVTGFY PSDIAVEWES NGQPENTYKT TPPVLDSDGS YLLYSKLTVN KSRWQPGNIF TCSVMHEALH NHYTQKSLSV SPGK. In another embodiment, the rhesus IgG Fc shares at least 90% identity, at least 95% identity, at least 99% identity, or at least 100% identity to SEQ ID NO: 9. In one embodiment, the rhesus IgG further comprises a hinge sequence.

In another embodiment, the fusion domain is a human albumin or a functional variant thereof. In one embodiment, the human albumin is

SEQ ID NO: 10:

DAHKSEVAHRFKDLGEENFKALVLIAFAQYLQQCPFEDHVKLVNEVTEFAKTCVAD
ESAENCDKSLHTLFGDKLCTVATLRETYGEMADCCAKQEPERNECFLQHKDDNPNL
PRLVRPEVDVMCTAFHDNEETFLKKYLYEIARRHPYFYAPELLFFAKRYKAAFTECC
QAADKAACLLPKLDELRDEGKASSAKQRLKCASLQKFGERAFKAWAVARLSQRFP
KAEFAEVSKLVTDLTKVHTECCHGDLLECADDRADLAKYICENQDSISSKLKECCEK
PLLEKSHCIAEVENDEMPADLPSLAADFVESKDVCKNYAEAKDVFLGMFLYEYARR
HPDYSVVLLLRLAKTYETTLEKCCAAADPHECYAKVFDEFKPLVEEPQNLIKQNCEL
FEQLGEYKFQNALLVRYTKKVPQVSTPTLVEVSRNLGKVGSKCCKHPEAKRMPCAE
DYLSVVLNQLCVLHEKTPVSDRVTKCCTESLVNRRPCFSALEVDETYVPKEFNAETF
TFHADICTLSEKERQIKKQTALVELVKHKPKATKEQLKAVMDDFAAFVEKCCKADD
KETCFAEEGKKLVAASQAALGL. In another embodiment, the human albumin shares at least 90% identity, at least 95% identity, at least 99% identity, or at least 100% identity to SEQ ID NO: 10.

In another embodiment, the fusion domain is a rhesus albumin or a functional variant thereof. In one embodiment, the rhesus albumin is

SEQ ID NO: 11:

DTHKSEVAHRFKDLGEEHFKGLVLVAFSQYLQQCPFEEHVKLVNEVTEFAKTCVAD
ESAENCDKSLHTLFGDKLCTVATLRETYGEMADCCAKQEPERNECFLQHKDDNPNL
PPLVRPEVDVMCTAFHDNEATFLKKYLYEVARRHPYFYAPELLFFAARYKAAFAEC
CQAADKAACLLPKLDELRDEGKASSAKQRLKCASLQKFGDRAFKAWAVARLSQKF
PKAEFAEVSKLVTDLTKVHTECCHGDLLECADDRADLAKYMCENQDSISSKLKECC
DKPLLEKSHCLAEVENDEMPADLPSLAADYVESKDVCKNYAEAKDVFLGMFLYEY
ARRHPDYSVMLLLRLAKAYEATLEKCCAAADPHECYAKVFDEFQPLVEEPQNLVKQ
NCELFEQLGEYKFQNALLVRYTKKVPQVSTPTLVEVSRNLGKVGAKCCKLPEAKRM
PCAEDYLSVVLNRLCVLHEKTPVSEKVTKCCTESLVNRRPCFSALELDEAYVPKAFN
AETFTFHADMCTLSEKEKQVKKQTALVELVKHKPKATKEQLKGVMDNFAAFVEKC
CKADDKEACFAEEGPKFVAASQAALA. In another embodiment, the rhesus albumin shares at least 90% identity, at least 95% identity, at least 99% identity, or at least 100% identity to SEQ ID NO: 11.

The in vivo function and stability of the fusion proteins of the present disclosure may be optimized by adding small peptide linkers, e.g., to prevent potentially unwanted domain interactions or for other reasons. Further, a glycine-rich linker may provide some structural flexibility such that the GLP-2 analog portion can interact productively with the GLP-2 receptor on target cells such as the beta cells of the pancreas. Thus, the C- terminus of the GLP-2 analog and the N- terminus of the fusion domain of the fusion protein are, in one embodiment, fused via a linker. In one embodiment, the linker includes 1, 1.5 or 2 repeats of a G-rich peptide linker having the sequence GGGGSGGGGSGGGGS (SEQ ID NO: 12).

In one embodiment, the fusion protein comprises (a) human thrombin leader, (b) a DPP-IV resistant variant of GLP-2(A2G), a linker, and (c) a human IgG Fc. In one embodiment, the fusion protein has the sequence of SEQ ID NO: 13, or a sequence at least 90%, at least 95%, at least 98%, or at least 99% identical thereto.

SEQ ID NO: 13

MAHVRGLQLPGCLALAALCSLVHSQHVFLAPQQARSLLQRVRRHGDGSFSDEMNTI
LDNLAARDFINWLIQTKITDGGGGGGSGGGGSGGGGSAESKYGPPCPPCPAPEAAGG
PSVFLFPPKPKDTLMISRTPEVTCVVVDVSQEDPEVQFNWYVDGVEVHNAKTKPREE
QFNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKGLPSSIEKTISKAKGQPREPQVYTL
PPSQEEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYS
RLTVDKSRWQEGNVFSCSVMHEALHNHYTQKSLSLSLG*

In one embodiment, the sequence encoding the fusion protein is SEQ ID NO: 14 or a sequence at least 75%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identical thereto.

SEQ ID NO: 14:

ATGGCTCACGTTCGGGGACTGCAGCTGCCTGGATGTCTGGCTCTTGCCGCTCTGT
GTAGCCTGGTGCACAGCCAGCACGTGTTTCTGGCTCCTCAGCAAGCCAGATCACT
GCTGCAGAGAGTTAGAAGGCACGGCGACGGCAGCTTCAGCGACGAGATGAACAC
CATCCTGGACAACCTGGCCGCCAGAGACTTCATCAACTGGCTGATCCAGACCAA
GATCACCGACGGTGGCGGAGGCGGAGGATCTGGTGGTGGTGGATCTGGCGGCGG
AGGTTCTGCCGAGTCTAAGTACGGACCTCCTTGTCCTCCCTGTCCTGCTCCAGAA
GCTGCTGGCGGCCCATCCGTGTTTCTGTTCCCTCCAAAGCCTAAGGACACCCTGA
TGATCAGCAGAACCCCTGAAGTGACCTGCGTGGTGGTCGACGTGTCCCAAGAGG

ATCCTGAGGTGCAGTTCAATTGGTACGTGGACGGCGTGGAAGTGCACAACGCCA
AGACCAAGCCTAGAGAGGAACAGTTCAACAGCACCTACAGAGTGGTGTCCGTGC
TGACCGTGCTGCACCAGGATTGGCTGAACGGCAAAGAGTACAAGTGCAAGGTGT
CCAACAAGGGCCTGCCTAGCAGCATCGAGAAAACCATCAGCAAGGCCAAGGGCC
AGCCAAGAGAACCCCAGGTGTACACACTGCCTCCAAGCCAAGAGGAAATGACCA
AGAACCAGGTGTCCCTGACCTGCCTGGTCAAGGGCTTCTACCCTTCCGATATCGC
CGTGGAATGGGAGAGCAACGGCCAGCCTGAGAACAACTACAAGACCACACCTCC
TGTGCTGGACTCCGATGGCTCATTCTTCCTGTACAGCAGACTGACCGTGGACAAG
AGCAGGTGGCAAGAGGGCAACGTGTTCAGCTGCAGCGTGATGCACGAGGCCCTG
CACAACCACTACACCCAGAAAAGCCTGAGCCTGTCTCTGGGCTAA

In one embodiment, the fusion protein comprises (a) human thrombin leader, (b) a DPP-IV resistant variant of GLP-2(A2G), a linker, and (c) a human serum albumin. In one embodiment, the fusion protein has the sequence of SEQ ID NO: 15, or a sequence at least 90%, at least 95%, at least 98%, or at least 99% identical thereto.

SEQ ID NO: 15:

MAHVRGLQLPGCLALAALCSLVHSQHVFLAPQQARSLLQRVRRHGDGSFSDEMNTI
LDNLAARDFINWLIQTKITDGGGGGGSGGGGSGGGGSDAHKSEVAHRFKDLGEENF
KALVLIAFAQYLQQCPFEDHVKLVNEVTEFAKTCVADESAENCDKSLHTLFGDKLCT
VATLRETYGEMADCCAKQEPERNECFLQHKDDNPNLPRLVRPEVDVMCTAFHDNE
ETFLKKYLYEIARRHPYFYAPELLFFAKRYKAAFTECCQAADKAACLLPKLDELRDE
GKASSAKQRLKCASLQKFGERAFKAWAVARLSQRFPKAEFAEVSKLVTDLTKVHTE
CCHGDLLECADDRADLAKYICENQDSISSKLKECCEKPLLEKSHCIAEVENDEMPAD
LPSLAADFVESKDVCKNYAEAKDVFLGMFLYEYARRHPDYSVVLLLRLAKTYETTL
EKCCAAADPHECYAKVFDEFKPLVEEPQNLIKQNCELFEQLGEYKFQNALLVRYTK
KVPQVSTPTLVEVSRNLGKVGSKCCKHPEAKRMPCAEDYLSVVLNQLCVLHEKTPV
SDRVTKCCTESLVNRRPCFSALEVDETYVPKEFNAETFTFHADICTLSEKERQIKKQT
ALVELVKHKPKATKEQLKAVMDDFAAFVEKCCKADDKETCFAEEGKKLVAASQAA
LG*

In one embodiment, the sequence encoding the fusion protein is SEQ ID NO: 16 or a sequence at least 75%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identical thereto.

SEQ ID NO: 16:

ATGGCTCACGTTCGGGGACTGCAGCTGCCTGGATGTCTGGCTCTTGCCGCTCTGT
GTAGCCTGGTGCACAGCCAGCACGTGTTTCTGGCTCCTCAGCAAGCCAGATCACT
GCTGCAGAGAGTTAGAAGGCACGGCGACGGCAGCTTCAGCGACGAGATGAACAC
CATCCTGGACAACCTGGCCGCCAGAGACTTCATCAACTGGCTGATCCAGACCAA
GATCACCGACGGTGGCGGAGGCGGAGGATCTGGTGGTGGTGGATCTGGCGGCGG
AGGTTCTGACGCCCACAAATCTGAAGTGGCCCACCGGTTCAAGGACCTGGGCGA
AGAGAATTTCAAGGCCCTGGTGCTGATCGCCTTCGCTCAGTACCTGCAGCAGTGC
CCCTTCGAGGACCACGTGAAGCTGGTCAACGAAGTGACCGAGTTCGCCAAGACC
TGCGTGGCCGACGAGAGCGCCGAGAACTGTGATAAGAGCCTGCACACCCTGTTC
GGCGACAAGCTGTGTACAGTGGCCACACTGAGAGAAACCTACGGCGAGATGGCC
GACTGCTGCGCCAAGCAAGAGCCCGAGAGAAACGAGTGCTTCCTGCAGCACAAG
GACGACAACCCCAACCTGCCTAGACTCGTGCGGCCTGAAGTGGACGTGATGTGC
ACCGCCTTCCACGACAACGAGGAAACCTTCCTGAAGAAGTACCTGTACGAGATC
GCCAGACGGCACCCCTACTTTTACGCCCCTGAGCTGCTGTTCTTCGCCAAGCGGT
ATAAGGCCGCCTTCACCGAGTGTTGTCAGGCCGCTGATAAGGCTGCCTGCCTGCT
GCCTAAGCTGGACGAGCTTAGAGACGAGGGCAAAGCCAGCTCCGCCAAGCAGAG
ACTGAAGTGTGCCAGCCTGCAGAAGTTCGGCGAGAGAGCCTTTAAGGCCTGGGC
CGTTGCTAGACTGAGCCAGAGATTTCCCAAGGCCGAGTTTGCCGAGGTGTCCAAG
CTCGTGACCGACCTGACAAAGGTGCACACCGAGTGCTGCCACGGCGACCTGCTG
GAATGCGCCGACGATAGAGCCGACCTGGCCAAGTACATCTGCGAGAACCAGGAC
AGCATCAGCAGCAAGCTGAAAGAGTGCTGCGAGAAGCCTCTGCTGGAAAAGAGC
CACTGTATCGCCGAGGTGGAAAACGACGAGATGCCCGCCGATCTGCCTTCTCTGG
CCGCCGATTTTGTGGAAAGCAAGGACGTGTGCAAGAACTACGCCGAGGCCAAGG
ACGTGTTCCTGGGCATGTTTCTGTACGAGTACGCCCGCAGACACCCCGACTACTC
TGTTGTGCTGCTGCTGAGACTGGCCAAAACCTACGAGACAACCCTGGAAAAGTG
CTGTGCCGCCGCTGATCCTCACGAGTGTTACGCCAAGGTGTTCGACGAGTTCAAG
CCACTGGTGGAAGAACCCCAGAACCTGATCAAGCAGAACTGCGAGCTGTTCGAG
CAGCTGGGCGAGTACAAGTTCCAGAACGCCCTGCTCGTGCGGTACACCAAGAAG
GTGCCCCAGGTTTCCACACCTACACTGGTTGAGGTGTCCCGGAACCTGGGCAAAG
TGGGCAGCAAGTGTTGCAAGCACCCTGAGGCCAAGAGAATGCCCTGCGCCGAGG

ATTACCTGAGCGTGGTGCTGAATCAGCTGTGCGTGCTGCACGAGAAAACCCCTGT
GTCCGACAGAGTGACCAAGTGCTGTACCGAGAGCCTGGTCAACAGACGGCCTTG
CTTTAGCGCCCTCGAGGTGGACGAGACATACGTGCCCAAAGAGTTCAACGCCGA
GACATTCACCTTCCACGCCGACATCTGTACCCTGAGCGAGAAAGAGCGGCAGAT
CAAGAAACAGACTGCCCTGGTGGAACTGGTCAAGCACAAGCCCAAGGCCACCAA
AGAACAGCTGAAGGCCGTGATGGACGACTTCGCCGCCTTCGTGGAAAAGTGCTG
CAAGGCCGACGACAAAGAGACCTGCTTCGCCGAAGAGGGCAAGAAACTGGTGG
CCGCTTCTCAGGCTGCTCTGGGATAA

In one embodiment, the fusion protein comprises (a) human thrombin leader, (b) a DPP-IV resistant variant of GLP-2(A2G), a linker, and (c) an XTEN polypeptide. In one embodiment, the fusion protein has the sequence of SEQ ID NO: 17, or a sequence at least 90%, at least 95%, at least 98%, or at least 99% identical thereto.

SEQ ID NO: 17:

MAHVRGLQLPGCLALAALCSLVHSQHVFLAPQQARSLLQRVRRHGDGSFSDEMNTI
LDNLAARDFINWLIQTKITDGGGGGGSGGGGSGGGGSSPAGSPTSTEEGTSESATPES
GPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPES
GPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGS
APGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPES
GPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGS
APGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSE
TPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSA
PGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGP
GSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP
GTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGP
GTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETP
GTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETP
GTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGP
GTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEE
GTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP
G*

In one embodiment, the sequence encoding the fusion protein is SEQ ID NO: 18 or a sequence at least 75%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identical thereto.

SEQ ID NO: 18:

ATGGCTCACGTTCGGGGACTGCAGCTGCCTGGATGTCTGGCTCTTGCCGCTCTGT
GTAGCCTGGTGCACAGCCAGCACGTGTTTCTGGCTCCTCAGCAAGCCAGATCACT
GCTGCAGAGAGTTAGAAGGCACGGCGACGGCAGCTTCAGCGACGAGATGAACAC
CATCCTGGACAACCTGGCCGCCAGAGACTTCATCAACTGGCTGATCCAGACCAA
GATCACCGACGGTGGCGGAGGCGGAGGATCTGGTGGTGGTGGATCTGGCGGCGG
AGGAAGTTCTCCTGCTGGCAGCCCTACAAGCACCGAGGAAGGCACAAGCGAGTC
TGCCACACCTGAGTCTGGCCCTGGCACATCTACAGAGCCTAGCGAAGGATCTGCC
CCAGGATCTCCTGCCGGCTCTCCAACATCTACCGAAGAGGGAACCAGCACCGAG
CCATCTGAGGGATCTGCTCCCGGAACAAGCACAGAGCCTTCAGAAGGATCCGCT
CCTGGCACCTCTGAAAGCGCCACACCAGAAAGCGGACCTGGATCTGAGCCTGCC
ACAAGCGGATCTGAGACACCTGGAAGCGAGCCAGCCACATCTGGCAGCGAAACA
CCTGGTTCTCCAGCCGGATCTCCCACCAGCACAGAAGAGGGCACATCCGAATCTG
CTACCCCTGAATCTGGACCAGGCACCTCCACAGAACCTTCCGAGGGTTCTGCCCC
TGGAACCTCTACCGAACCATCAGAAGGCAGCGCTCCAGGTTCACCAGCTGGAAG
CCCAACCTCTACAGAGGAAGGGACATCCACTGAGCCAAGCGAGGGAAGCGCTCC
CGGCACTAGTACAGAACCGAGCGAGGGCAGTGCTCCTGGAACCAGCGAATCCGC
TACTCCAGAGAGTGGCCCAGGCACCAGTACTGAACCCTCTGAGGGTAGCGCACC
CGGAACATCTGAGAGCGCTACTCCCGAATCAGGCCCAGGCTCTGAACCTGCTACC
AGCGGAAGTGAAACACCCGGCACCTCTACTGAGCCCTCCGAAGGCTCAGCACCT
GGCACAAGCACTGAACCATCAGAGGGCTCCGCACCAGGCACCAGCGAAAGTGCT
ACACCAGAGTCAGGACCCGGAACCTCCGAAAGTGCAACTCCTGAGAGCGGACCA
GGCTCTCCCGCTGGATCTCCTACATCAACTGAAGAAGGGACCTCCGAGAGCGCA
ACCCCAGAGTCTGGTCCAGGATCAGAACCTGCCACCTCCGGCTCTGAAACCCCAG
GCACTTCTGAGTCCGCCACGCCAGAATCTGGTCCTGGGACTAGCACCGAACCGA
GTGAAGGTTCAGCTCCCGGGACTTCTACGGAACCCAGTGAAGGATCTGCACCCG
GCACATCAACCGAACCGTCAGAGGGATCAGCCCCTGGGACTTCCACAGAGCCGT
CTGAGGGCAGCGCCCCAGGGACGTCTACAGAACCATCTGAAGGATCAGCACCAG

GGACCTCTACCGAGCCAAGTGAAGGCAGTGCACCGGGAAGTCCAGCAGGCTCCC
CTACAAGTACTGAAGAGGGTACTAGCACGGAACCCAGCGAGGGTTCCGCTCCAG
GGACATCTGAATCCGCAACTCCGGAATCCGGACCTGGCAGTGAACCAGCTACAT
CCGGATCCGAGACTCCGGGAACCTCAGAATCAGCTACACCCGAGAGTGGACCTG
GCTCCGAACCAGCAACTAGCGGCTCAGAAACTCCTGGGACAAGCGAGAGTGCAA
CACCCGAATCTGGACCTGGAACAAGTACTGAGCCCAGCGAAGGCAGCGCCCCTG
GAACTTCTGAATCTGCCACTCCTGAAAGTGGCCCTGGAAGCCCTGCAGGCTCACC
CACATCCACAGAAGAAGGATCACCAGCAGGCAGCCCCACTTCAACGGAAGAGGG
ATCCCCAGCTGGATCCCCAACTAGTACGGAAGAAGGCACTTCAGAAAGCGCTAC
GCCCGAGTCCGGTCCTGGCACTTCTACTGAACCATCCGAGGGAAGTGCCCCTGGC
ACTTCCGAGAGTGCTACACCTGAAAGCGGTCCCGGCTCTGAACCAGCCACTTCTG
GATCTGAAACGCCCGGGACATCCGAGTCAGCAACGCCCGAAAGCGGCCCAGGTT
CCGAGCCGGCTACTAGTGGTTCAGAGACTCCAGGGACTTCCGAGTCTGCTACTCC
TGAGTCCGGACCGGGAACATCAACCGAGCCTTCCGAAGGATCTGCACCTGGAAG
CCCTGCCGGATCTCCTACCAGTACTGAGGAAGGCACCTCAGAGTCTGCCACTCCA
GAGTCAGGTCCTGGAAGCGAACCTGCAACAAGCGGCAGCGAAACTCCAGGCACT
AGCGAGTCAGCTACCCCAGAATCAGGACCTGGATCTCCAGCAGGGTCCCCAACA
TCTACTGAGGAAGGCTCTCCTGCTGGCTCCCCTACCTCTACCGAAGAGGGGACCT
CAACAGAGCCATCCGAGGGGAGCGCACCTGGTACATCAGAGTCCGCAACTCCCG
AGTCTGGCCCCGGAACTAGCGAATCTGCAACCCCGGAAAGTGGACCCGGGACGA
GTGAATCAGCCACACCTGAATCCGGTCCAGGATCCGAGCCTGCAACTTCTGGAA
GCGAGACACCAGGATCTGAGCCAGCTACGTCTGGCTCTGAGACTCCTGGATCTCC
TGCTGGTAGTCCCACCTCCACTGAAGAGGGAACTTCCACCGAACCGAGCGAGGG
ATCAGCACCAGGCACTAGCACAGAACCGTCCGAAGGATCTGCTCCAGGCTCTGA
ACCCGCAACCTCCGGATCAGAAACCCCTGGAACATCCGAAAGCGCTACACCGGA
AAGTGGCCCCGGAACCTCTACAGAACCTAGCGAGGGAAGCGCACCAGGATAA

In one embodiment, the fusion protein comprises (a) rhesus thrombin leader, (b) a DPP-IV resistant variant of GLP-2(A2G), a linker, and (c) a rhesus IgG Fc. In one embodiment, the fusion protein has the sequence of SEQ ID NO: 19, or a sequence at least 90%, at least 95%, at least 98%, or at least 99% identical thereto.

SEQ ID NO: 19

MAHVRGLQLPGCLALAALCSLVHSQHVFLAPQQALSLLQRVRRHGDGSFSDEMNTV
LVDNLATRDFINWLIQTKITDGGGGGGGSGGGGSGGGGSAEFTPPCPPCPAPELLGGP
SVFLFPPKPKDTLMISRTPEVTCVVVDVSQEDPEVQFNWYVDGVEVHNAQTKPRER
QFNSTYRVVSVLTVTHQDWLNGKEYTCKVSNKGLPAPIEKTISKAKGQPREPQVYIL
PPPQEELTKNQVSLTCLVTGFYPSDIAVEWESNGQPENTYKTTPPVLDSDGSYLLYSK
LTVNKSRWQPGNIFTCSVMHEALHNHYTQKSLSVSPG*

In one embodiment, the sequence encoding the fusion protein is SEQ ID NO: 20 or a sequence at least 75%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identical thereto.

SEQ ID NO: 20:

ATGGCTCACGTTCGGGGACTGCAGCTGCCTGGATGTCTGGCTCTTGCCGCTCTGT
GTAGCCTGGTGCACAGCCAGCATGTGTTTCTGGCTCCTCAGCAGGCCCTGAGCCT
GCTGCAAAGAGTTAGAAGGCACGGCGACGGCAGCTTCAGCGACGAGATGAATAC
CGTGCTGGTGGACAACCTGGCCACCAGAGACTTCATCAACTGGCTGATCCAGACC
AAGATCACCGACGGTGGTGGCGGAGGCGGAGGATCTGGTGGCGGTGGTTCTGGC
GGTGGCGGATCTGCTGAGTTTACCCCTCCTTGTCCTCCCTGTCCTGCTCCAGAACT
GCTCGGCGGACCTTCCGTGTTCCTGTTTCCTCCAAAGCCTAAGGACACCCTGATG
ATCAGCAGAACCCCTGAAGTGACCTGCGTGGTGGTGGACGTGTCCCAAGAGGAT
CCTGAGGTGCAGTTCAATTGGTACGTGGACGGCGTGGAAGTGCACAACGCCCAG
ACAAAGCCCAGAGAGCGGCAGTTCAACAGCACCTACAGAGTGGTGTCCGTGCTG
ACCGTGACACACCAGGATTGGCTGAACGGCAAAGAGTACACCTGTAAAGTCTCC
AACAAGGGCCTGCCTGCTCCTATCGAGAAAACCATCAGCAAGGCCAAGGGCCAG
CCTAGAGAACCCCAGGTGTACATCCTGCCTCCACCTCAAGAGGAACTGACCAAG
AACCAGGTGTCCCTGACCTGTCTGGTCACCGGCTTCTACCCTTCCGATATCGCCGT
GGAGTGGGAGAGCAACGGACAGCCCGAGAACACCTACAAGACCACACCTCCAGT
GCTGGACAGCGACGGCTCTTACCTGCTGTACTCCAAGCTGACAGTGAACAAGAG
CCGGTGGCAGCCCGGCAACATCTTCACCTGTTCTGTGATGCACGAGGCCCTGCAC
AACCACTACACCCAGAAAAGCCTGAGCGTGTCCCCTGGATAA

In one embodiment, the fusion protein comprises (a) rhesus thrombin leader, (b) a DPP-IV resistant variant of GLP-2(A2G), a linker, and (c) a rhesus serum albumin. In one

embodiment, the fusion protein has the sequence of SEQ ID NO: 21, or a sequence at least 90%, at least 95%, at least 98%, or at least 99% identical thereto.

SEQ ID NO: 21

MAHVRGLQLPGCLALAALCSLVHSQHVFLAPQQALSLLQRVRRHGDGSFSDEMNTV
LVDNLATRDFINWLIQTKITGGGGGGGSGGGGSGGGGSDTHKSEVAHRFKDLGEEH
FKGLVLVAFSQYLQQCPFEEHVKLVNEVTEFAKTCVADESAENCDKSLHTLFGDKL
CTVATLRETYGEMADCCAKQEPERNECFLQHKDDNPNLPPLVRPEVDVMCTAFHDN
EATFLKKYLYEVARRHPYFYAPELLFFAARYKAAFAECCQAADKAACLLPKLDELR
DEGKASSAKQRLKCASLQKFGDRAFKAWAVARLSQKFPKAEFAEVSKLVTDLTKV
HTECCHGDLLECADDRADLAKYMCENQDSISSKLKECCDKPLLEKSHCLAEVENDE
MPADLPSLAADYVESKDVCKNYAEAKDVFLGMFLYEYARRHPDYSVMLLLRLAKA
YEATLEKCCAAADPHECYAKVFDEFQPLVEEPQNLVKQNCELFEQLGEYKFQNALL
VRYTKKVPQVSTPTLVEVSRNLGKVGAKCCKLPEAKRMPCAEDYLSVVLNRLCVLH
EKTPVSEKVTKCCTESLVNRRPCFSALELDEAYVPKAFNAETFTFHADMCTLSEKEK
QVKKQTALVELVKHKPKATKEQLKGVMDNFAAFVEKCCKADDKEACFAEEGPKFV
AASQAALA

In one embodiment, the sequence encoding the fusion protein is SEQ ID NO: or a sequence at least 75%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identical thereto.

SEQ ID NO: 22:

ATGGCTCACGTTCGGGGACTGCAGCTGCCTGGATGTCTGGCTCTTGCCGCTCTGT
GTAGCCTGGTGCACAGCCAGCATGTGTTTCTGGCTCCTCAGCAAGCCCTGAGCCT
GCTGCAAAGAGTTAGAAGGCACGGCGACGGCAGCTTCAGCGACGAGATGAATAC
CGTGCTGGTGGACAACCTGGCCACCAGAGACTTCATCAACTGGCTGATCCAGACC
AAGATCACCGGTGGTGGCGGAGGCGGAGGATCTGGTGGCGGTGGTTCTGGCGGT
GGCGGATCTGATACACACAAGTCTGAGGTGGCCCACCGGTTCAAGGACCTGGGC
GAAGAACACTTCAAAGGCCTGGTGCTGGTCGCCTTCAGCCAGTACCTGCAGCAGT
GCCCTTTCGAGGAACACGTGAAGCTGGTCAACGAAGTGACCGAGTTCGCCAAGA
CCTGCGTGGCCGACGAGAGCGCCGAGAACTGTGATAAGAGCCTGCACACCCTGT
TCGGCGACAAGCTGTGTACAGTGGCCACACTGAGAGAAACCTACGGCGAGATGG
CCGACTGCTGCGCCAAGCAAGAGCCCGAGAGAAACGAGTGCTTCCTGCAGCACA

AGGACGACAACCCCAACCTGCCTCCACTCGTCAGACCCGAAGTGGACGTGATGT
GCACCGCCTTCCACGACAATGAGGCCACCTTCCTGAAGAAATACCTGTACGAGGT
GGCCAGACGGCACCCCTACTTTTACGCCCCTGAACTGCTGTTCTTTGCCGCCAGG
TACAAGGCCGCCTTCGCCGAATGTTGTCAGGCCGCTGATAAGGCCGCTTGCCTGC
TGCCTAAGCTGGACGAGCTTAGAGACGAGGGCAAAGCCAGCTCCGCCAAGCAGA
GACTGAAGTGTGCCAGCCTGCAGAAGTTCGGCGATAGAGCCTTTAAGGCCTGGG
CCGTCGCTAGACTGAGCCAGAAGTTTCCCAAGGCCGAGTTTGCCGAGGTGTCCAA
GCTCGTGACCGACCTGACAAAGGTGCACACCGAGTGCTGTCACGGCGACCTGCT
GGAATGCGCCGACGATAGAGCCGACCTGGCCAAGTACATGTGCGAGAACCAGGA
CAGCATCAGCAGCAAGCTGAAAGAGTGCTGCGACAAGCCTCTGCTGGAAAAGAG
CCACTGTCTGGCCGAGGTGGAAAACGACGAGATGCCCGCCGATCTGCCTTCTCTG
GCCGCCGATTACGTGGAAAGCAAGGACGTGTGCAAGAACTACGCCGAGGCCAAG
GACGTGTTCCTGGGCATGTTTCTGTACGAGTACGCCCGCAGACACCCCGACTACT
CTGTTATGCTGCTGCTGAGACTGGCCAAGGCCTACGAGGCCACTCTGGAAAAGTG
TTGTGCCGCCGCTGATCCCCACGAGTGTTACGCCAAAGTGTTCGACGAGTTCCAG
CCACTGGTGGAAGAACCCCAGAACCTGGTCAAGCAGAACTGCGAGCTGTTCGAG
CAGCTGGGCGAGTACAAGTTCCAGAACGCCCTGCTCGTGCGGTACACCAAGAAG
GTGCCCCAGGTTTCCACACCTACACTGGTTGAGGTGTCCCGGAACCTGGGAAAAG
TGGGCGCCAAGTGTTGCAAGCTGCCTGAGGCCAAGAGAATGCCCTGCGCCGAGG
ATTACCTGAGCGTGGTGCTGAACAGACTGTGCGTGCTGCACGAGAAAACCCCTGT
GTCCGAGAAAGTGACCAAGTGCTGTACCGAGAGCCTGGTCAATCGGAGGCCTTG
CTTTAGCGCCCTGGAACTGGACGAGGCCTACGTGCCCAAGGCCTTCAACGCCGA
GACATTCACCTTCCACGCCGACATGTGTACCCTGAGCGAGAAAGAAAAGCAAGT
GAAGAAACAGACAGCCCTGGTCGAGCTGGTTAAGCACAAGCCTAAGGCCACCAA
AGAACAACTGAAGGGCGTGATGGACAACTTCGCCGCCTTTGTGGAAAAATGCTG
CAAGGCCGACGACAAAGAGGCCTGCTTCGCAGAAGAGGGCCCTAAGTTTGTGGC
CGCCTCTCAAGCTGCTCTGGCTTAA

In one embodiment, the fusion protein comprises (a) rhesus thrombin leader, (b) a DPP-IV resistant variant of GLP-2(A2G), a linker, and (c) an XTEN polypeptide. In one embodiment, the fusion protein has the sequence of SEQ ID NO: 23, or a sequence at least 90%, at least 95%, at least 98%, or at least 99% identical thereto.

SEQ ID NO: 23

MAHVRGLQLPGCLALAALCSLVHSQHVFLAPQQALSLLQRVRRHGDGSFSDEMNTV
LVDNLATRDFINWLIQTKITDGGGGGGSGGGGSGGGGSSPAGSPTSTEEGTSESATPE
SGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPES
GPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGS
APGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPES
GPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGS
APGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSE
TPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSA
PGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGP
GSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP
GTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGP
GTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETP
GTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETP
GTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGP
GTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEE
GTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP
G

In one embodiment, the sequence encoding the fusion protein is SEQ ID NO: 24 or a sequence at least 75%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identical thereto.

SEQ ID NO: 24:

ATGGCTCACGTTCGGGGACTGCAGCTGCCTGGATGTCTGGCTCTTGCCGCTCTGT
GTAGCCTGGTGCACAGCCAGCATGTGTTTCTGGCTCCTCAGCAAGCCCTGAGCCT
GCTGCAAAGAGTTAGAAGGCACGGCGACGGCAGCTTCAGCGACGAGATGAATAC
CGTGCTGGTGGACAACCTGGCCACCAGAGACTTCATCAACTGGCTGATCCAGACC
AAGATCACCGACGGTGGCGGAGGCGGAGGATCTGGTGGTGGTGGATCTGGCGGC
GGAGGAAGTTCTCCTGCTGGCAGCCCTACAAGCACCGAGGAAGGCACAAGCGAG
TCTGCCACACCTGAGTCTGGCCCTGGCACATCTACAGAGCCTAGCGAAGGATCTG
CCCCAGGATCTCCTGCCGGCTCTCCAACATCTACCGAAGAGGGAACCAGCACCG
AGCCATCTGAGGGATCTGCTCCCGGAACAAGCACAGAGCCTTCAGAAGGATCCG

CTCCTGGCACCTCTGAAAGCGCCACACCAGAAAGCGGACCTGGCTCTGAACCTG
CCACAAGCGGATCTGAGACACCTGGAAGCGAGCCAGCCACATCTGGCAGCGAAA
CACCTGGATCACCAGCCGGATCTCCCACCTCTACCGAGGAAGGGACATCCGAGA
GCGCTACCCCAGAATCTGGACCAGGCACCAGCACAGAACCCTCTGAAGGTTCAG
CCCCTGGAACCTCTACCGAACCATCAGAAGGCAGCGCTCCAGGTTCTCCCGCTGG
ATCCCCTACATCCACAGAAGAGGGCACCTCCACTGAACCTAGCGAGGGAAGTGC
TCCCGGCACTTCCACAGAACCATCCGAGGGCAGTGCACCTGGAACCAGCGAATC
TGCTACCCCTGAGAGTGGACCCGGAACATCCACTGAGCCCTCCGAGGGTTCAGCT
CCAGGCACATCAGAATCCGCCACTCCAGAGTCCGGACCAGGATCTGAGCCAGCT
ACCAGCGGCTCTGAAACACCCGGCACTAGTACCGAGCCAAGCGAGGGTAGCGCA
CCAGGGACAAGTACCGAACCGTCTGAGGGCTCCGCACCAGGCACTTCCGAAAGT
GCTACTCCTGAAAGCGGCCCAGGCACTAGCGAATCCGCAACACCCGAGAGCGGT
CCTGGAAGTCCTGCAGGTTCACCTACCAGCACTGAAGAGGGGACTAGCGAGAGC
GCAACTCCTGAATCAGGCCCTGGATCCGAACCTGCTACCTCCGGAAGTGAAACCC
CTGGGACAAGCGAAAGTGCAACGCCCGAGTCAGGACCCGGGACTAGCACGGAA
CCCAGTGAAGGATCTGCACCCGGGACATCTACCGAGCCGTCAGAAGGTTCTGCTC
CAGGGACTAGTACTGAGCCTTCCGAAGGTTCTGCACCTGGAACTTCCACAGAGCC
CAGTGAAGGCAGTGCCCCTGGCACAAGCACTGAACCGTCCGAAGGCAGTGCTCC
CGGGACCAGTACAGAACCGAGCGAGGGCTCTGCTCCTGGTAGTCCAGCAGGATC
TCCAACTAGCACCGAAGAAGGGACTTCCACCGAGCCTTCCGAGGGAAGCGCTCC
TGGAACATCCGAGTCCGCTACGCCAGAGAGTGGCCCAGGTTCTGAACCCGCTACT
TCCGGCTCAGAGACTCCTGGGACTTCTGAGTCTGCAACCCCGGAAAGTGGTCCTG
GTAGCGAACCAGCAACTAGCGGAAGCGAGACACCCGGAACCTCAGAGAGTGCTA
CACCGGAATCCGGTCCAGGGACGTCTACGGAACCGTCTGAAGGATCAGCTCCCG
GCACTAGCGAAAGCGCTACACCTGAAAGTGGTCCCGGATCTCCAGCAGGCAGCC
CAACCTCTACTGAAGAAGGTTCCCCAGCTGGAAGCCCCACTTCCACTGAGGAAG
GCTCTCCCGCAGGCTCACCCACTAGTACGGAAGAAGGCACGTCCGAGTCTGCTAC
TCCCGAATCCGGACCTGGAACTAGCACTGAGCCAAGCGAAGGATCAGCACCCGG
AACCTCTGAGTCCGCCACACCAGAATCTGGTCCTGGTTCCGAGCCTGCCACTTCA
GGATCAGAAACCCCGGGCACGAGTGAATCAGCAACGCCGGAATCTGGCCCCGGA
AGCGAACCGGCTACGTCTGGATCTGAAACGCCAGGGACCTCCGAATCAGCTACG

CCTGAGTCTGGTCCAGGGACATCCACCGAACCTAGTGAAGGCTCCGCACCTGGA
AGCCCTGCTGGAAGCCCAACGAGTACTGAAGAGGGCACTTCTGAGAGCGCTACG
CCTGAGTCAGGACCTGGAAGCGAACCTGCAACATCCGGCTCAGAAACACCAGGG
ACCAGCGAAAGCGCAACCCCAGAGAGTGGACCTGGATCTCCAGCTGGCTCTCCT
ACTAGTACAGAGGAAGGCAGCCCTGCTGGCTCCCCAACGTCAACAGAAGAAGGT
ACTAGCACAGAGCCCAGCGAGGGTTCCGCTCCGGGAACTTCTGAATCTGCTACAC
CCGAGTCAGGTCCTGGTACAAGCGAGTCAGCTACGCCCGAAAGTGGACCTGGCA
CCTCAGAGTCTGCAACTCCTGAGAGCGGTCCAGGATCAGAACCAGCCACCTCTG
GCTCTGAGACACCAGGTTCTGAGCCTGCAACGTCCGGAAGCGAAACACCAGGCA
GTCCTGCCGGAAGTCCTACTTCAACCGAAGAGGGGACCTCTACAGAGCCATCAG
AGGGCTCTGCACCGGGCACCTCAACAGAACCATCTGAAGGATCCGCACCGGGCT
CTGAGCCTGCTACTAGTGGAAGCGAAACTCCTGGCACCAGTGAATCCGCTACTCC
CGAGTCTGGCCCGGGAACGTCTACTGAACCATCTGAGGGAAGTGCCCCAGGCTA
A

When a variant or fragment of the leader sequence, GLP-2 receptor agonist, or fusion domain is desired, the coding sequences for these peptides may be generated using site-directed mutagenesis of the wild-type nucleic acid sequence. Alternatively or additionally, web-based or commercially available computer programs, as well as service based companies may be used to back translate the amino acids sequences to nucleic acid coding sequences, including both RNA and/or cDNA. See, e.g., backtranseq by EMBOSS, ebi.ac.uk/Tools/st/; Gene Infinity (geneinfinity.org/sms-/sms_backtranslation.html); ExPasy (expasy.org/tools/). In one embodiment, the RNA and/or cDNA coding sequences are designed for optimal expression in the subject species for which administration is ultimately intended, e.g., a human.

The coding sequences may be designed for optimal expression using codon optimization. Codon-optimized coding regions can be designed by various different methods. This optimization may be performed using methods which are available on-line, published methods, or a company which provides codon optimizing services. One codon optimizing method is described, e.g., in International Patent Application Pub. No. WO 2015/012924, which is incorporated by reference herein. Briefly, the nucleic acid sequence encoding the product is modified with synonymous codon sequences. Suitably, the entire length of the

open reading frame (ORF) for the product is modified. However, in some embodiments, only a fragment of the ORF may be altered. By using one of these methods, one can apply the frequencies to any given polypeptide sequence, and produce a nucleic acid fragment of a codon-optimized coding region which encodes the polypeptide.

In addition to the leader sequences, GLP-2 receptor agonists, fusion domains, and fusion proteins provided herein, nucleic acid sequences encoding these polypeptides are provided. In one embodiment, a nucleic acid sequence is provided which encodes for the GLP-2 peptides described herein. In some embodiments, this may include any nucleic acid sequence which encodes the GLP-2 sequence of SEQ ID NO: 1. In another embodiment, this includes any nucleic acid which includes the GLP-2 sequence of SEQ ID NO: 2.

In one embodiment, a nucleic acid sequence is provided which encodes for the GLP-2 fusion protein described herein. In another embodiment, this includes any nucleic acid sequence which encodes the GLP-2 fusion protein of SEQ ID NO: 13, 15, 17, 19, 21 or 23.

Expression Cassettes

Provided herein, in another aspect, is an expression cassette comprising a nucleic acid encoding a GLP-2 fusion protein as described herein. As used herein, an "expression cassette" refers to a nucleic acid molecule which comprises a biologically useful nucleic acid sequence (e.g., a gene cDNA encoding a protein, enzyme or other useful gene product, mRNA, etc.) and regulatory sequences operably linked thereto which direct or modulate transcription, translation, and/or expression of the nucleic acid sequence and its gene product. As used herein, "operably linked" sequences include both regulatory sequences (also referred to as elements) that are contiguous or non-contiguous with the nucleic acid sequence and regulatory sequences that act in trans or cis nucleic acid sequence. Such regulatory sequences typically include, e.g., one or more of a promoter, an enhancer, a transcription factor, transcription terminator, an intron, sequences that enhance translation efficiency (i.e., a Kozak consensus sequence), efficient RNA processing signals such as slicing and a polyadenylation sequence, sequences that stabilize cytoplasmic mRNA, for example Woodchuck Hepatitis Virus (WHP) posttranslational Regulatory Element (WPRE), and a TATA signal. The expression cassette may contain regulatory sequences upstream (5' to) of the gene sequence, e.g., one or more of a promoter, an enhancer, an intron, etc., and one or

more of an enhancer, or regulatory sequences downstream (3' to) a gene sequence, e.g., 3' untranslated region (3' UTR) comprising a polyadenylation site, among other elements. In certain embodiments, the regulatory sequences are operably linked to the nucleic acid sequence of a gene product, wherein the regulatory sequences are separated from nucleic acid sequence of a gene product by intervening nucleic acid sequences, i.e., 5'-untranslated regions (5'UTR). In certain embodiments, the expression cassette comprises nucleic acid sequence of one or more of gene products. In some embodiments, the expression cassette can be a monocistronic or a bicistronic expression cassette. In other embodiments, the term "transgene" refers to one or more DNA sequences from an exogenous source which are inserted into a target cell.

In one embodiment, the expression cassette refers to a nucleic acid molecule which comprises the GLP-2 construct coding sequences (e.g., coding sequences for the GLP-2 fusion protein), promoter, and may include other regulatory sequences therefor, which cassette may be engineered into a genetic element and/or packaged into the capsid of a viral vector (e.g., a viral particle). Typically, such an expression cassette for generating a viral vector contains the GLP-2 construct sequences described herein flanked by packaging signals of the viral genome (and is termed a "vector genome") and other expression control sequences such as those described herein. Any of the expression control sequences can be optimized for a specific species using techniques known in the art including, e.g., codon optimization, as described herein.

In certain embodiments, the expression cassette includes a constitutive promoter. In another embodiment, a CB7 promoter is used. CB7 is a chicken β-actin promoter with cytomegalovirus enhancer elements. In some embodiments, the CB7 promoter has the nucleic acid sequence of SEQ ID NO: 25. In one embodiment, the promoter is a CMV promoter. In some embodiments, the CMV promoter is a nucleic acid sequence of SEQ ID NO: 26.

In another embodiment, a tissue specific promoter is used. Alternatively, other liver-specific promoters may be used such as those listed in the Liver Specific Gene Promoter Database, Cold Spring Harbor, (rulai.schl.edu/LSPD), and including, but not limited to, alpha 1 anti-trypsin (A1AT); human albumin (Miyatake et al., J. Virol., 71:5124 32 (1997)), humAlb; hepatitis B virus core promoter (Sandig et al., Gene Ther., 3:1002 9 (1996)); a TTR minimal enhancer/promoter, alpha-antitrypsin promoter, liver-specific promoter (LSP) (Wu

et al. Mol Ther. 16:280-289 (2008)), TBG liver specific promoter. Other promoters, such as viral promoters, constitutive promoters, regulatable promoters (see, e.g., WO 2011/126808 and WO 2013/04943), or a promoter responsive to physiologic cues may be used may be utilized in the vectors described herein.

In one embodiment, the promoter is comprised in an inducible gene expression system. The inducible gene regulation/expression system contains at least the following components: a promoter operably linked to transgene encoding the GLP-2 fusion protein described herein (also referred to as the regulatable promoter), an activation domain, DNA binding domain, and zinc finger homeodomain binding site(s). In other embodiments, additional components may be included in the expression system, as further described herein.

The system comprises the promoter upstream of the coding sequence for the GLP-2 fusion protein. Promoters described herein, such as CMV and CB7 promoters may be used. In one embodiment, the promoter is a CMV promoter, such as that shown in SEQ ID NO: 26. In another embodiment, the promoter is the ubiquitous, inducible promoter Z12I which comprises 12 repeated copies of the binding site for ZFHD1 and the IL2 minimal promoter. See, e.g., Chen et al, Hum Gene Ther Methods. 2013 Aug; 24(4): 270–278, which is incorporated herein.

The expression system comprises an activation domain, which is preferably located upstream of the DNA binding domain. In one embodiment, the activation domain is a fusion of the carboxy terminus from the p65 subunit of NF-kappa B and FKBP12-rapamycin binding (FRB) domain of FKBP12-rapamycin-associated protein (FRAP). In one embodiment, the activation domain is a FKBP12-rapamycin binding (FRB) domain of human FKBP12-rapamycin-associated protein (FRAP) fused to a carboxy terminus from the p65 subunit of NF-kappa B from a human.

In one embodiment, there is a linker between the transactivation domain and DNA binding domain, which linker may be an F2A or an IRES. In one embodiment the linker is selected from an IRES or a 2A peptide.

The DNA binding domain is composed of a DNA-binding fusion of zinc finger homeodomain 1 (ZFHD1) joined to up to three copies of FK506 binding protein (FKBP). In the presence of an inducing agent, e.g., a rapalog such as rapamycin, the DNA binding domain and activation domain are dimerized through interaction of their FKBP and FRB

domains, leading to transcription activation of the transgene. In some embodiments, the ZFHD1 is included in frame with the GT2A or IRES.

The expression system is designed to have one, two or three copies of the FKBP sequence. These are termed herein FKBP subunits. In one embodiment, the subunits are designed to express the same protein, but to have nucleic acids which are divergent from one another in order to minimize recombination. For example, SEQ ID NO: 27 provides 3 "wobbled" coding sequences for FKBP, each of which encode the sequence shown in SEQ ID NO: 28:

GVQVETISPGDGRTFPKRGQTCVVHYTGMLEDGKKFDSSRDRNKPFKFMLGKQEVI RGWEEGVAQMSVGQRAKLTISPDYAYGATGHPGIIPPHATLVFDVELLKLE

The expression system further comprises zinc finger homeodomain binding sites. The nucleic acid molecule contains at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 or 12 binding sites for ZFHD. In one embodiment, the expression system contains 8 (eight) zinc finger homeodomains binding site (binding partners) (8XZFHD). However, the invention encompasses expression systems having from two to about twelve copies of the zinc finger binding site. An example of a single copy of a ZFHD binding site is: aatgatgggcgctcgagt (SEQ ID NO: 29)

In some embodiments, there is a minimal IL2 promoter downstream of the zinc finger homeodomain binding sites. An exemplary IL2 promoter is shown in SEQ ID NO: 30.

Such inducible systems are known in the art, and include, e.g., the rapamycin-inducible system described by e.g., Rivera et al, A humanized system for pharmacologic control of gene expression, Nature Medicine volume 2, pages 1028–1032 (September 1996) and Rivera et al, Long-term pharmacologically regulated expression of erythropoietin in primates following AAV-mediated gene transfer, Blood, 15 February 2005, volume 105, number 4, both of which are incorporated herein by reference. In one embodiment, the inducible gene expression system comprises a CMV promoter, the activation domain is a FKBP12-rapamycin binding (FRB) domain of human FKBP12-rapamycin-associated protein (FRAP) fused to a carboxy terminus from the p65 subunit of NF-kappa B from a human, GT2A peptide, ZFHD1 DNA binding domain, three FKBP subunits, an hGH poly A, 8XZFHD, and a minimal sIL2 promoter. These sequences are in addition to the coding sequence for the GLP-2 fusion protein and optionally other regulatory sequences.

In addition to a promoter, an expression cassette and/or a vector may contain other appropriate transcription initiation, termination, enhancer sequences, efficient RNA processing signals such as splicing and polyadenylation (polyA) signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (i.e., Kozak consensus sequence); sequences that enhance protein stability; and when desired, sequences that enhance secretion of the encoded product. Examples of suitable polyA sequences include, e.g., SV40, bovine growth hormone (bGH), human growth hormone (hGH), SV40, rabbit β-globin (also referred to as rabbit globin polyA; RGB), modified RGB (mRGB) and TK polyA. Examples of suitable enhancers include, e.g., the alpha fetoprotein enhancer, the TTR minimal promoter/enhancer, LSP (TH-binding globulin promoter/alpha1-microglobulin/bikunin enhancer), amongst others. In one embodiment, the polyA is a rabbit globin polyA.

These control sequences are "operably linked" to the GLP-2 construct sequences. As used herein, the term "operably linked" refers to both expression control sequences that are contiguous with the gene of interest and expression control sequences that act *in trans* or at a distance to control the gene of interest.

In one embodiment, a rAAV is provided which includes a 5' ITR, CB7 promoter, chicken beta-actin intron, coding sequence for the fusion protein of SEQ ID NO: 13, 15, 17, 19, 21, or 23, a rabbit globin poly A, and a 3' ITR. In another embodiment, the rAAV comprises a polynucleotide comprising a CMV promoter, the activation domain is a FKBP12-rapamycin binding (FRB) domain of human FKBP12-rapamycin-associated protein (FRAP) fused to a carboxy terminus from the p65 subunit of NF-kappa B from a human, IRES, ZFHD1 DNA binding domain, three FKBP subunits, an hGH poly A, 8XZFHD, a minimal sIL2 promoter, coding sequence for the GLP-2 fusion protein of SEQ ID NO: 13, 15, 17, 19, 21, or 23, and rabbit beta globin polyA.

In another embodiment, a two vector inducible system is provided. The first rAAV comprises 12XZFHD, a minimal IL2 promoter, coding sequence for the GLP-2 fusion protein of SEQ ID NO: 13, 15, 17, 19, 21, or 23, and rabbit beta globin polyA. A stuffer sequence may be included to increase the packaging size of the vector. The second rAAV comprises a polynucleotide comprising a CMV promoter, an intron, the activation domain is a FKBP12-rapamycin binding (FRB) domain of human FKBP12-rapamycin-associated protein (FRAP)

fused to a carboxy terminus from the p65 subunit of NF-kappa B from a human, IRES, ZFHD1 DNA binding domain, and a polyA.

In one embodiment, an expression cassette is provided that includes a polynucleotide comprising a CB7 promoter, chicken beta-actin intron, coding sequence for the fusion protein of SEQ ID NO: 13, 15, 17, 19, 21, or 23, and a rabbit globin poly A. In one embodiment, the expression cassette is that found in SEQ ID NO: X, or a sequence sharing at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% identity therewith. In another embodiment, a vector genome is provided wherein SEQ ID NO: X-X, or a sequence sharing at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% identity therewith is flanked by 5' and 3' AAV ITRs.

In another embodiment, an expression cassette is provided that includes a polynucleotide comprising a CB7 promoter, chicken beta-actin intron, coding sequence for the fusion protein of SEQ ID NO: 13, 15, 17, 19, 21, or 23, and a rabbit globin poly A.

Viral Vectors

In another aspect, viral vectors that include the expression cassettes described herein are provided. In certain embodiments of the viral vectors described herein, the viral vector is an adeno-associated virus (AAV) viral vector or recombinant AAV (rAAV). The term "recombinant AAV" or "rAAV" as used herein refers to naturally occurring adeno-associated viruses, adeno-associated viruses available to one of skill in the art and/or in light of the composition(s) and method(s) described herein, as well as artificial AAVs. An adeno-associated virus (AAV) viral vector is an AAV DNase-resistant particle having an AAV protein capsid into which is packaged an expression cassette flanked by AAV inverted terminal repeat sequences (ITRs) (together referred to as the "vector genome") for delivery to target cells. An AAV capsid is composed of 60 capsid (cap) protein subunits, VP1, VP2, and VP3, that are arranged in an icosahedral symmetry in a ratio of approximately 1:1:10 to 1:1:20, depending upon the selected AAV. Various AAVs may be selected as sources for capsids of AAV viral vectors as identified above. In one embodiment, the AAV capsid is an AAVrh91 capsid or variant thereof. In certain embodiments, the capsid protein is designated by a number or a combination of numbers and letters following the term "AAV" in the name of the rAAV vector. Unless otherwise specified, the AAV capsid, ITRs, and other selected AAV components described herein, may be readily selected from among any AAV,

including, without limitation, the AAVs identified as AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAVrh10, AAVhu37, AAVrh32.33, AAVAnc80, AAV10, AAV11, AAV12, AAVrh8, AAVrh74, AAV-DJ8, AAV-DJ, AAVhu.37, AAVrh.64R1, and AAVhu68. See, e.g., US Published Patent Application No. 2007-0036760-A1; US Published Patent Application No. 2009-0197338-A1; EP 1310571. See also, WO 2003/042397 (AAV7 and other simian AAV), US Patent 7790449 and US Patent 7282199 (AAV8), WO 2005/033321 and US 7,906,111 (AAV9), and WO 2006/110689, and WO 2003/042397 (rh.10), WO 2005/033321, WO 2018/160582 (AAVhu68), which are incorporated herein by reference. Other suitable AAVs may include, without limitation, AAVrh90 [PCT/US20/30273, filed April 28, 2020], AAVrh91 [PCT/US20/030266, filed April 28, 2020, now a publication WO 2020/223231, published November 5, 2020], AAVrh92, AAVrh93, AAVrh91.93 [PCT/US20/30281, filed April 28, 2020], which are incorporated by reference herein. Other suitable AAV include AAV3B variants which are described in US Provisional Patent Application No. 62/924,112, filed October 21, 2019, and US Provisional Patent Application No. 63/025,753, filed May 15, 2020, describing AAV3B.AR2.01, AAV3B.AR2.02, AAV3B.AR2.03, AAV3B.AR2.04, AAV3B.AR2.05, AAV3B.AR2.06, AAV3B.AR2.07, AAV3B.AR2.08, AAV3B.AR2.10, AAV3B.AR2.11, AAV3B.AR2.12, AAV3B.AR2.13, AAV3B.AR2.14, AAV3B.AR2.15, AAV3B.AR2.16, or AAV3B.AR2.17, which are incorporated herein by reference. See also, International Patent Application No. PCT/US21/45945, filed August 13, 2021, US Provisional Patent Application No. 63/065,616, filed August 14, 2020, and US Provisional Patent Application No. 63/109,734, filed November 4, 2020, which are all incorporated herein by reference in its entireties. These documents also describe other AAV capsids which may be selected for generating rAAV and are incorporated by reference. Among the AAVs isolated or engineered from human or non-human primates (NHP) and well characterized, human AAV2 is the first AAV that was developed as a gene transfer vector; it has been widely used for efficient gene transfer experiments in different target tissues and animal models.

As used herein, relating to AAV, the term "variant" means any AAV sequence which is derived from a known AAV sequence, including those with a conservative amino acid replacement, and those sharing at least 90%, at least 95%, at least 97%, at least 99% or greater sequence identity over the amino acid or nucleic acid sequence. In another

embodiment, the AAV capsid includes variants which may include up to about 10% variation from any described or known AAV capsid sequence. That is, the AAV capsid shares about 90% identity to about 99.9 % identity, about 95% to about 99% identity or about 97% to about 98% identity to an AAV capsid provided herein and/or known in the art. In one embodiment, the AAV capsid shares at least 95% identity with an AAV capsid. When determining the percent identity of an AAV capsid, the comparison may be made over any of the variable proteins (e.g., vp1, vp2, or vp3).

In one embodiment, the viral vector is an rAAV having the capsid of AAV8 or a functional variant thereof. In one embodiment, the viral vector is an rAAV having the capsid of AAVrh91 or a functional variant thereof. In one embodiment, the viral vector is an rAAV having the capsid of AAV3.AR.2.12 or a functional variant thereof. In one embodiment, the viral vector is an rAAV having a capsid selected from AAV9, AAVrh64R1, AAVhu37, or AAVrh10.

In certain embodiments, a novel isolated AAVrh91 capsid is provided. A nucleic acid sequence encoding the AAVrh91 capsid is provided in SEQ ID NO: 31 and the encoded amino acid sequence is provided in SEQ ID NO: 32. Provided herein is an rAAV comprising at least one of the vp1, vp2 and the vp3 of AAVrh91 (SEQ ID NO: 32). Also provided herein are rAAV comprising an AAV capsid encoded by at least one of the vp1, vp2 and the vp3 of AAVrh91 (SEQ ID NO: 31). In yet another embodiment, a nucleic acid sequence encoding the AAVrh91 amino acid sequence is provided in SEQ ID NO: 19 and the encoded amino acid sequence is provided in SEQ ID NO: 32. Also provided herein are rAAV comprising an AAV capsid encoded by at least one of the vp1, vp2 and the vp3 of AAVrh91eng (SEQ ID NO: 19). In certain embodiments, the vp1, vp2 and/or vp3 is the full-length capsid protein of AAVrh91 (SEQ ID NO: 32). In other embodiments, the vp1, vp2 and/or vp3 has an N-terminal and/or a C-terminal truncation (e.g., truncation(s) of about 1 to about 10 amino acids).

In certain embodiments, an AAVrh91 capsid is characterized by one or more of the following: (1) AAVrh91 capsid proteins comprising: a heterogeneous population of AAVrh91 vp1 proteins selected from: vp1 proteins produced by expression from a nucleic acid sequence which encodes the predicted amino acid sequence of 1 to 736 of SEQ ID NO: 32, vp1 proteins produced from SEQ ID NO: 31, or vp1 proteins produced from a nucleic

acid sequence at least 70% identical to SEQ ID NO: 31 which encodes the predicted amino acid sequence of 1 to 736 of SEQ ID NO: 32, a heterogeneous population of AAVrh91 vp2 proteins selected from: vp2 proteins produced by expression from a nucleic acid sequence which encodes the predicted amino acid sequence of at least about amino acids 138 to 736 of SEQ ID NO: 32, vp2 proteins produced from a sequence comprising at least nucleotides 412 to 2208 of SEQ ID NO: 31, or vp2 proteins produced from a nucleic acid sequence at least 70% identical to at least nucleotides 412 to 2208 of SEQ ID NO: 31 which encodes the predicted amino acid sequence of at least about amino acids 138 to 736 of SEQ ID NO: 32, a heterogeneous population of AAVrh91 vp3 proteins selected from: vp3 proteins produced by expression from a nucleic acid sequence which encodes the predicted amino acid sequence of at least about amino acids 203 to 736 of SEQ ID NO: 32, vp3 proteins produced from a sequence comprising at least nucleotides 607 to 2208 of SEQ ID NO: 31, or vp3 proteins produced from a nucleic acid sequence at least 70% identical to at least nucleotides 607 to 2208 of SEQ ID NO: 31 which encodes the predicted amino acid sequence of at least about amino acids 203 to 736 of SEQ ID NO: 32; and/or (2) a heterogeneous population of vp1 proteins which are the product of a nucleic acid sequence encoding the amino acid sequence of SEQ ID NO: 32, a heterogeneous population of vp2 proteins which are the product of a nucleic acid sequence encoding the amino acid sequence of at least about amino acids 138 to 736 of SEQ ID NO: 32, and a heterogeneous population of vp3 proteins which are the product of a nucleic acid sequence encoding at least amino acids 203 to 736 of SEQ ID NO: 32, wherein: the vp1, vp2 and vp3 proteins contain subpopulations with amino acid modifications comprising at least two highly deamidated asparagines (N) in asparagine – glycine pairs in SEQ ID NO: 32 and optionally further comprising subpopulations comprising other deamidated amino acids, wherein the deamidation results in an amino acid change; and (B) a vector genome in the AAVrh91 capsid, the vector genome comprising a nucleic acid molecule comprising AAV inverted terminal repeat sequences and a non-AAV nucleic acid sequence encoding a product operably linked to sequences which direct expression of the product in a host cell.

In certain embodiments, an AAVrh91 capsid is characterized by one or more of the following: (1) AAVrh91 capsid proteins comprising: a heterogeneous population of AAVrh91 vp1 proteins selected from: vp1 proteins produced by expression from a nucleic

acid sequence which encodes the predicted amino acid sequence of 1 to 736 of SEQ ID NO: 32, vp1 proteins produced from SEQ ID NO: 19, or vp1 proteins produced from a nucleic acid sequence at least 70% identical to SEQ ID NO: 19 which encodes the predicted amino acid sequence of 1 to 736 of SEQ ID NO: 32, a heterogeneous population of AAVrh91 vp2 proteins selected from: vp2 proteins produced by expression from a nucleic acid sequence which encodes the predicted amino acid sequence of at least about amino acids 138 to 736 of SEQ ID NO: 32, vp2 proteins produced from a sequence comprising at least nucleotides 412 to 2208 of SEQ ID NO: 19, or vp2 proteins produced from a nucleic acid sequence at least 70% identical to at least nucleotides 412 to 2208 of SEQ ID NO: 19 which encodes the predicted amino acid sequence of at least about amino acids 138 to 736 of SEQ ID NO: 32, a heterogeneous population of AAVrh91 vp3 proteins selected from: vp3 proteins produced by expression from a nucleic acid sequence which encodes the predicted amino acid sequence of at least about amino acids 203 to 736 of SEQ ID NO: 32, vp3 proteins produced from a sequence comprising at least nucleotides 607 to 2208 of SEQ ID NO: 19, or vp3 proteins produced from a nucleic acid sequence at least 70% identical to at least nucleotides 607 to 2208 of SEQ ID NO: 19 which encodes the predicted amino acid sequence of at least about amino acids 203 to 736 of SEQ ID NO: 32; and/or (2) a heterogeneous population of vp1 proteins which are the product of a nucleic acid sequence encoding the amino acid sequence of SEQ ID NO: 32, a heterogeneous population of vp2 proteins which are the product of a nucleic acid sequence encoding the amino acid sequence of at least about amino acids 138 to 736 of SEQ ID NO: 32, and a heterogeneous population of vp3 proteins which are the product of a nucleic acid sequence encoding at least amino acids 203 to 736 of SEQ ID NO: 32, wherein: the vp1, vp2 and vp3 proteins contain subpopulations with amino acid modifications comprising at least two highly deamidated asparagines (N) in asparagine – glycine pairs in SEQ ID NO: 32 and optionally further comprising subpopulations comprising other deamidated amino acids, wherein the deamidation results in an amino acid change; and (B) a vector genome in the AAVrh91 capsid, the vector genome comprising a nucleic acid molecule comprising AAV inverted terminal repeat sequences and a non-AAV nucleic acid sequence encoding a product operably linked to sequences which direct expression of the product in a host cell.

In certain embodiments, the AAVrh91 vp1, vp2 and vp3 proteins contain subpopulations with amino acid modifications comprising at least two highly deamidated asparagines (N) in asparagine – glycine pairs in SEQ ID NO: 32 and optionally further comprising subpopulations comprising other deamidated amino acids, wherein the deamidation results in an amino acid change. High levels of deamidation at N-G pairs N57, N383 and/or N512 are observed, relative to the number of SEQ ID NO: 32. Deamidation has been observed in other residues. In certain embodiments, AAVrh91 may have other residues deamidated, e.g., typically at less than 10% and/or may have other modifications, including phosphorylation (e.g., where present, in the range of about 2 to about 30%, or about 2 to about 20%, or about 2 to about 10%) (e.g., at S149), or oxidation (e.g., at one or more of ~W22, ~M211, W247, M403, M435, M471, W478, W503, ~M537, ~M541, ~M559, ~M599, M635, and/or, W695). Optionally the W may oxidize to kynurenine.

Table A – AAVrh91 Deamidation

AAVrh91 Deamidation based on VP1 numbering	% Deamidation
N57+Deamidation	65-90, 70-95, 80-95, 75-100, 80-100, or 90-100
N94+Deamidation	2-15 or 2-5
N303+Deamidation	2-15 or 5-10
N383+Deamidation	65-90, 70-95, 80-95, 75-100, 80-100, or 90-100
N497+Deamidation	2-15 or 5-10
N512+Deamidation	65-90, 70-95, 80-95, 75-100, 80-100, or 90-100
~N691+Deamidation	2-15, 2-10, or 5-10

In certain embodiments, an AAVrh91 capsid is modified in one or more of the positions identified in the preceding table, in the ranges provided, as determined using mass spectrometry with a trypsin enzyme. In certain embodiments, one or more of the positions, or

the glycine following the N is modified as described herein. Residue numbers are based on the AAVrh91 sequence provided herein. See, SEQ ID NO: 32.

In certain embodiments, an AAVrh91 capsid comprises: a heterogeneous population of vp1 proteins which are the product of a nucleic acid sequence encoding the amino acid sequence of SEQ ID NO: 32, a heterogeneous population of vp2 proteins which are the product of a nucleic acid sequence encoding the amino acid sequence of at least about amino acids 138 to 736 of SEQ ID NO: 32, and a heterogeneous population of vp3 proteins which are the product of a nucleic acid sequence encoding at least amino acids 203 to 736 of SEQ ID NO: 32.

In certain embodiments, the modified AAVrh91 nucleic acid sequences is be used to generate a mutant rAAV having a capsid with lower deamidation than the native AAVrh91 capsid. Such mutant rAAV may have reduced immunogenicity and/or increase stability on storage, particularly storage in suspension form.

In one aspect, a recombinant AAV (rAAV) is provided. The rAAV includes an AAV capsid from adeno-associated virus rh91, and a vector genome packaged in the AAV capsid, said vector genome comprising AAV inverted terminal repeats (ITRs), a coding sequence for the GLP-2 receptor agonist of SEQ ID NO: 14, and regulatory sequences which direct expression of the GLP-2 receptor agonist.

In certain embodiments, an AAV68 capsid is further characterized by one or more of the following. AAV hu68 capsid proteins comprise: AAVhu68 vp1 proteins produced by expression from a nucleic acid sequence which encodes the predicted amino acid sequence of 1 to 736 of SEQ ID NO: 55, vp1 proteins produced from SEQ ID NO: 53 or 54, or vp1 proteins produced from a nucleic acid sequence at least 70% identical to SEQ ID NO: 53 or 54 which encodes the predicted amino acid sequence of 1 to 736 of SEQ ID NO: 55; AAVhu68 vp2 proteins produced by expression from a nucleic acid sequence which encodes the predicted amino acid sequence of at least about amino acids 138 to 736 of SEQ ID NO: 55, vp2 proteins produced from a sequence comprising at least nucleotides 412 to 2211 of SEQ ID NO: 53 or 54, or vp2 proteins produced from a nucleic acid sequence at least 70% identical to at least nucleotides 412 to 2211 of SEQ ID NO: 53 or 54 which encodes the predicted amino acid sequence of at least about amino acids 138 to 736 of SEQ ID NO: 55, and/or AAVhu68 vp3 proteins produced by expression from a nucleic acid sequence which

encodes the predicted amino acid sequence of at least about amino acids 203 to 736 of SEQ ID NO: 55, vp3 proteins produced from a sequence comprising at least nucleotides 607 to 2211 of SEQ ID NO: 53 or 54, or vp3 proteins produced from a nucleic acid sequence at least 70% identical to at least nucleotides 607 to 2211 of SEQ ID NO: 53 or 54 which encodes the predicted amino acid sequence of at least about amino acids 203 to 736 of SEQ ID NO: 55.

Additionally or alternatively, an AAV capsid is provided which comprise a heterogenous population of vp1 proteins optionally comprising a valine at position 157, a heterogenous population of vp2 proteins optionally comprising a valine at position 157, and a heterogenous population of vp3 proteins, wherein at least a subpopulation of the vp1 and vp2 proteins comprise a valine at position 157 and optionally further comprising a glutamic acid at position 67 based on the numbering of the vp1 capsid of SEQ ID NO: 55. Additionally or alternatively, an AAVhu68 capsid is provided which comprises a heterogenous population of vp1 proteins which are the product of a nucleic acid sequence encoding the amino acid sequence of SEQ ID NO: 55, a heterogenous population of vp2 proteins which are the product of a nucleic acid sequence encoding the amino acid sequence of at least about amino acids 138 to 736 of SEQ ID NO: 55, and a heterogenous population of vp3 proteins which are the product of a nucleic acid sequence encoding at least amino acids 203 to 736 of SEQ ID NO: 55, wherein: the vp1, vp2 and vp3 proteins contain subpopulations with amino acid modifications

The AAVhu68 vp1, vp2 and vp3 proteins are typically expressed as alternative splice variants encoded by the same nucleic acid sequence which encodes the full-length vp1 amino acid sequence of SEQ ID NO: 55 (amino acid 1 to 736). Optionally the vp1-encoding sequence is used alone to express the vp1, vp2 and vp3 proteins. Alternatively, this sequence may be co-expressed with one or more of a nucleic acid sequence which encodes the AAVhu68 vp3 amino acid sequence of SEQ ID NO: 55 (about aa 203 to 736) without the vp1-unique region (about aa 1 to about aa 137) and/or vp2-unique regions (about aa 1 to about aa 202), or a strand complementary thereto, the corresponding mRNA or tRNA (about nt 607 to about nt 2211 of SEQ ID NO: 53 or 54), or a sequence at least 70% to at least 99% (e.g., at least 85%, at least 90%, at least 95%, at least 97%, at least 98% or at least 99%) identical to SEQ ID NO: 53 or 54 which encodes aa 203 to 736 of SEQ ID NO: 55. Additionally, or alternatively, the vp1-encoding and/or the vp2-encoding sequence may be

co-expressed with the nucleic acid sequence which encodes the AAVhu68 vp2 amino acid sequence of SEQ ID NO: 55 (about aa 138 to 736) without the vp1-unique region (about aa 1 to about 137), or a strand complementary thereto, the corresponding mRNA or tRNA (nt 412 to 2212 of SEQ ID NO: 53 or 54), or a sequence at least 70% to at least 99% (e.g., at least 85%, at least 90%, at least 95%, at least 97%, at least 98% or at least 99%) identical to SEQ ID NO: 53 or 54 which encodes about aa 138 to 736 of SEQ ID NO: 55.

As described herein, a rAAVhu68 has a rAAVhu68 capsid produced in a production system expressing capsids from an AAVhu68 nucleic acid which encodes the vp1 amino acid sequence of SEQ ID NO: 55, and optionally additional nucleic acid sequences, e.g., encoding a vp3 protein free of the vp1 and/or vp2-unique regions. The rAAVhu68 resulting from production using a single nucleic acid sequence vp1 produces the heterogenous populations of vp1 proteins, vp2 proteins and vp3 proteins. More particularly, the AAVhu68 capsid contains subpopulations within the vp1 proteins, within the vp2 proteins and within the vp3 proteins which have modifications from the predicted amino acid residues in SEQ ID NO: 55. These subpopulations include, at a minimum, deamidated asparagine (N or Asn) residues. For example, asparagines in asparagine - glycine pairs are highly deamidated.

In one embodiment, the AAVhu68 vp1 nucleic acid sequence has the sequence of SEQ ID NO: 53 or 54, or a strand complementary thereto, e.g., the corresponding mRNA or tRNA. In certain embodiments, the vp2 and/or vp3 proteins may be expressed additionally or alternatively from different nucleic acid sequences than the vp1, e.g., to alter the ratio of the vp proteins in a selected expression system. In certain embodiments, also provided is a nucleic acid sequence which encodes the AAVhu68 vp3 amino acid sequence of SEQ ID NO: 55 (about aa 203 to 736) without the vp1-unique region (about aa 1 to about aa 137) and/or vp2-unique regions (about aa 1 to about aa 202), or a strand complementary thereto, the corresponding mRNA or tRNA (about nt 607 to about nt 2211 of SEQ ID NO: 53 or 54). In certain embodiments, also provided is a nucleic acid sequence which encodes the AAVhu68 vp2 amino acid sequence of SEQ ID NO: 55 (about aa 138 to 736) without the vp1-unique region (about aa 1 to about 137), or a strand complementary thereto, the corresponding mRNA or tRNA (nt 412 to 2211 of SEQ ID NO: 53 or 54).

However, other nucleic acid sequences which encode the amino acid sequence of SEQ ID NO: 55 may be selected for use in producing rAAVhu68 capsids. In certain

embodiments, the nucleic acid sequence has the nucleic acid sequence of SEQ ID NO: 53 or 54 or a sequence at least 70% to 99% identical, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 99%, identical to SEQ ID NO: 53 or 54 which encodes SEQ ID NO: 55. In certain embodiments, the nucleic acid sequence has the nucleic acid sequence of SEQ ID NO: 53 or 54 or a sequence at least 70% to 99.%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 99%, identical to about nt 412 to about nt 2211 of SEQ ID NO: 53 or 54 which encodes the vp2 capsid protein (about aa 138 to 736) of SEQ ID NO: 55. In certain embodiments, the nucleic acid sequence has the nucleic acid sequence of about nt 607 to about nt 2211 of SEQ ID NO: 53 or 54 or a sequence at least 70% to 99.%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 99%, identical to nt SEQ ID NO: 53 or 54 which encodes the vp3 capsid protein (about aa 203 to 736) of SEQ ID NO: 55.

Deamidation Based on Predicted AAVHu68 [SEQ ID NO: 55]	Average % Based on VP1/VP2/VP3 Proteins in AAVhu68 Capsid	
Deamidated Residue + 1 (Neighboring AA)	Broad Range of Percentages (%)	Narrow Ranges (%)
N57 (N-G)	78 to 100%	80 to 100, 85 to 97
N66 (N-E)	0 to 5	0, 1 to 5
N94 (N-H)	0 to 15,	0, 1 to 15, 5 to 12, 8
N113 (N-L)	0 to 2	0, 1 to 2
~N253 (N-N)	10 to 25	15 to 22
Q259 (Q-I)	8 to 42	10 to 40, 20 to 35
~N270 (N-D)	12 to 30	15 to 28
~N304 (N-N) (position 303 also N)	0 to 5	1 to 4
N319 (N-I)	0 to 5	0, 1 to 5, 1 to 3
N329 * (N-G)*(position 328 also N)	65 to 100	70 to 95, 85 to 95, 80 to 100, 85 to 100,

Deamidation Based on Predicted AAVHu68 [SEQ ID NO: 55]	Average % Based on VP1/VP2/VP3 Proteins in AAVhu68 Capsid	
Deamidated Residue + 1 (Neighboring AA)	Broad Range of Percentages (%)	Narrow Ranges (%)
N336 (N-N)	0 to 100	0, 1 to 10, 25 to 100, 30 to 100, 30 to 95
~N409 (N-N)	15 to 30	20 to 25
N452 (N-G)	75 to 100	80 to 100, 90 to 100, 95 to 100,
N477 (N-Y)	0 to 8	0, 1 to 5
N512 (N-G)	65 to 100	70 to 95, 85 to 95, 80 to 100, 85 to 100,
~N515 (N-S)	0 to 25	0, 1 to 10, 5 to 25, 15 to 25
~Q599 (Asn-Q-Gly)	1 to 20	2 to 20, 5 to 15
N628 (N-F)	0 to 10	0, 1 to 10, 2 to 8
N651 (N-T)	0 to 3	0, 1 to 3
N663 (N-K)	0 to 5	0, 1 to 5, 2 to 4
N709 (N-N)	0 to 25	0,1 to 22, 15 to 25
N735	0 to 40	0. 1 to 35, 5 to 50, 20 to 35

In certain embodiments, the AAVhu68 capsid is characterized, by having, capsid proteins in which at least 45% of N residues are deamidated at least one of positions N57, N329, N452, and/or N512 based on the numbering of amino acid sequence of SEQ ID NO: 55. In certain embodiments, at least about 60%, at least about 70%, at least about 80%, or at least 90% of the N residues at one or more of these N-G positions (i.e., N57, N329, N452, and/or N512, based on the numbering of amino acid sequence of SEQ ID NO: 55) are deamidated. In these and other embodiments, an AAVhu68 capsid is further characterized by having a population of proteins in which about 1% to about 20% of the N residues have deamidations at one or more of positions: N94, N253, N270, N304, N409, N477, and/or Q599, based on the numbering of amino acid sequence of SEQ ID NO: 55. In certain

embodiments, the AAVhu68 comprises at least a subpopulation of vp1, vp2 and/or vp3 proteins which are deamidated at one or more of positions N35, N57, N66, N94, N113, N252, N253, Q259, N270, N303, N304, N305, N319, N328, N329, N336, N409, N410, N452, N477, N515, N598, Q599, N628, N651, N663, N709, N735, based on the numbering of amino acid sequence of SEQ ID NO: 55, or combinations thereof. In certain embodiments, the capsid proteins may have one or more amidated amino acids.

In another embodiment a recombinant adeno-associated virus (rAAV) is provided that has an AAVhu68 capsid and a vector genome, wherein (a) the AAV hu68 capsid comprises a heterogenous population of AAVhu68 vp1 proteins, a heterogenous population of AAVhu68 vp2 proteins; and a heterogenous population of AAVhu68 vp3 proteins, wherein the heterogenous AAVhu68 vp1, AAVhu68 vp2 and AAVhu68 vp3 proteins contain subpopulations with amino acid modifications comprising 50% to 100% deamidation in at least two asparagines (N) in asparagine - glycine pairs in two or more of N57, N329, N452, N512 of SEQ ID NO: 55 as determined using mass spectrometry and optionally further comprising subpopulations comprising other deamidated amino acids, wherein the deamidation results in an amino acid change, wherein the deamidated asparagines are deamidated to aspartic acid, isoaspartic acid, an interconverting aspartic acid/isoaspartic acid pair, or combinations thereof, wherein the AAVhu68 capsid further comprises subpopulations having one or more of:

at least 65% of asparagines (N) in asparagine - glycine pairs located at positions N57 of the vp1 proteins are deamidated, based on the numbering of SEQ ID NO: 55;

at least 75% of N in asparagine - glycine pairs in position N329 of the vp1, v2 and vp3 proteins are deamidated, based on the residue numbering of the amino acid sequence of SEQ ID NO: 55,

at least 50% of N in asparagine - glycine pairs in position N452 of the vp1, v2 and vp3 proteins are deamidated, based on the residue numbering of the amino acid sequence of SEQ ID NO: 55; and/or

at least 75% of N in asparagine - glycine pairs in position N512 of the vp1, v2 and vp3 proteins are deamidated, based on the residue numbering of the amino acid sequence of SEQ ID NO: 55, and

a vector genome in the AAVhu68 capsid, the vector genome comprising a nucleic acid molecule comprising AAV inverted terminal repeat sequences and a non-AAV nucleic acid sequence encoding a GLP-2 fusion as described herein operably linked to sequences which direct expression of GLP-2 fusion in a target cell.

In one embodiment, the rAAV is an scAAV. The abbreviation "sc" refers to self-complementary. "Self-complementary AAV" refers a plasmid or vector having an expression cassette in which a coding region carried by a recombinant AAV nucleic acid sequence has been designed to form an intra-molecular double-stranded DNA template. Upon infection, rather than waiting for cell mediated synthesis of the second strand, the two complementary halves of scAAV will associate to form one double stranded DNA (dsDNA) unit that is ready for immediate replication and transcription. See, e.g., D M McCarty et al, "Self-complementary recombinant adeno-associated virus (scAAV) vectors promote efficient transduction independently of DNA synthesis", Gene Therapy, (August 2001), Vol 8, Number 16, Pages 1248-1254. Self-complementary AAVs are described in, *e.g.,* U.S. Patent Nos. 6,596,535; 7,125,717; and 7,456,683, each of which is incorporated herein by reference in its entirety.

In one embodiment, the nucleic acid sequences encoding the GLP-2 constructs described herein are engineered into any suitable genetic element, e.g., naked DNA, phage, transposon, cosmid, RNA molecule (e.g., mRNA), episome, etc., which transfers the GLP-2 sequences carried thereon to a host cell, e.g., for generating nanoparticles carrying DNA or RNA, viral vectors in a packaging host cell and/or for delivery to a host cell in a subject. In one embodiment, the genetic element is a plasmid. The selected genetic element may be delivered by any suitable method, including transfection, electroporation, liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion. The methods used to make such constructs are known to those with skill in nucleic acid manipulation and include genetic engineering, recombinant engineering, and synthetic techniques. See, e.g., Green and Sambrook, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring Harbor, NY (2012).

As used herein, the term "host cell" may refer to the packaging cell line in which a vector (e.g., a recombinant AAV or rAAV) is produced from a production plasmid. In the alternative, the term "host cell" may refer to any target cell in which expression of a gene

product described herein is desired. Thus, a "host cell," refers to a prokaryotic or eukaryotic cell (e.g., bacterial cell, human cell or insect cell) that contains exogenous or heterologous DNA that has been introduced into the cell by any means, e.g., electroporation, calcium phosphate precipitation, microinjection, transformation, viral infection, transfection, liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion. In certain embodiments herein, the term "host cell" refers to cultures of cells of various mammalian species for in vitro assessment of the compositions described herein. In other embodiments herein, the term "host cell" refers to the cells employed to generate and package the viral vector or recombinant virus. In a further embodiment, the term "host cell" is an intestine cell, a small intestine cell, a pancreatic cell, a liver cell.

As used herein, the term "target cell" refers to any target cell in which expression of a heterologous nucleic acid sequence or protein is desired. In certain embodiments, the target cell is a liver cell. In other embodiments, the target cell is a muscle cell.

In one embodiment, the rAAV is provided which comprises a vector genome comprising an expression cassette, wherein the expression cassette comprises a CMV promoter, the activation domain is a FKBP12-rapamycin binding (FRB) domain of human FKBP12-rapamycin-associated protein (FRAP) fused to a carboxy terminus from the p65 subunit of NF-kappa B from a human, GT2A_V1 peptide, ZFHD1 DNA binding domain, three FKBP subunits, an hGH poly A, 12XZFHD, a minimal sIL2 promoter, coding sequence for the GLP-2 fusion protein of SEQ ID NO: 13, 15, 17, 19, 21, or 23, and rabbit beta globin polyA. In another embodiment, the rAAV is provide which comprises a vector genome comprising an expression cassette, wherein the expression cassette comprises a CMV promoter, the activation domain is a FKBP12-rapamycin binding (FRB) domain of human FKBP12-rapamycin-associated protein (FRAP) fused to a carboxy terminus from the p65 subunit of NF-kappa B from a human, GT2A_V2 peptide, ZFHD1 DNA binding domain, three FKBP subunits, an hGH poly A, 12XZFHD, a minimal sIL2 promoter, coding sequence for the GLP-2 fusion protein of SEQ ID NO: 13, 15, 17, 19, 21, or 23, and rabbit beta globin polyA.

In one embodiment, the rAAV is provided which comprises a vector genome comprising an expression cassette, wherein the expression cassette comprises a CB7 promoter, chicken beta-actin intron, coding sequence for the fusion protein of SEQ ID NO:

13, 15, 17, 19, 21, or 23, a rabbit globin poly A, and a 3' ITR. In another embodiment, the rAAV is provided which comprises a vector genome comprising an expression cassette, wherein the expression cassette comprises a CMV promoter, the activation domain is a FKBP12-rapamycin binding (FRB) domain of human FKBP12-rapamycin-associated protein (FRAP) fused to a carboxy terminus from the p65 subunit of NF-kappa B from a human, IRES, ZFHD1 DNA binding domain, three FKBP subunits, an hGH poly A, 8XZFHD, a minimal sIL2 promoter, coding sequence for the GLP-2 fusion protein of SEQ ID NO: 13, 15, 17, 19, 21, or 23and rabbit beta globin polyA.

In another embodiment, a two vector inducible system is provided. In one embodiment, the first rAAV is provided which comprises a vector genome comprising an expression cassette, wherein the expression cassette comprises a 12XZFHD, a minimal IL2 promoter, coding sequence for the GLP-2 fusion protein of SEQ ID NO: 13, 15, 17, 19, 21, or 23, and rabbit beta globin polyA. A stuffer sequence may be included to increase the packaging size of the vector. The second rAAV comprises a vector genome comprising an expression cassette, wherein the expression cassette comprises a polynucleotide comprising a CMV promoter, an intron, the activation domain is a FKBP12-rapamycin binding (FRB) domain of human FKBP12-rapamycin-associated protein (FRAP) fused to a carboxy terminus from the p65 subunit of NF-kappa B from a human, IRES, ZFHD1 DNA binding domain, and a polyA.

In one embodiment, an rAAV includes a vector genome that includes an expression cassette that includes a polynucleotide comprising a CB7 promoter, chicken beta-actin intron, coding sequence for the fusion protein of SEQ ID NO: 13, 15, 17, 19, 21, or 23, and a rabbit globin poly A.

The minimal sequences required to package the expression cassette into an AAV viral particle are the AAV 5' and 3' ITRs, which may be of the same AAV origin as the capsid, or of a different AAV origin (to produce an AAV pseudotype). In one embodiment, the ITR sequences from AAV2, or the deleted version thereof (ΔITR), are used for convenience and to accelerate regulatory approval. However, ITRs from other AAV sources may be selected. Preferably, the source of the ITRs is the same as the source of the Rep protein, which is provided in trans for production. Typically, an expression cassette for an AAV vector comprises an AAV 5' ITR, the GLP-2 fusion protein coding sequences and any regulatory

sequences, and an AAV 3' ITR. However, other configurations of these elements may be suitable. A shortened version of the 5' ITR, termed ΔITR, has been described in which the D-sequence and terminal resolution site (trs) are deleted. In other embodiments, the full-length AAV 5' and 3' ITRs are used.

For packaging an expression cassette into virions, the ITRs are the only AAV components required in cis in the same construct as the gene. In one embodiment, the coding sequences for the replication (rep) and/or capsid (cap) are removed from the AAV genome and supplied *in trans* or by a packaging cell line in order to generate the AAV vector. For example, as described above, a pseudotyped AAV may contain ITRs from a source which differs from the source of the AAV capsid. In one embodiment, a chimeric AAV capsid may be utilized. Still other AAV components may be selected. Sources of such AAV sequences are described herein and may also be isolated or obtained from academic, commercial, or public sources (e.g., the American Type Culture Collection, Manassas, VA). The AAV sequences may be obtained through synthetic or other suitable means by reference to published sequences such as are available in the literature or in databases such as, e.g., GenBank®, PubMed®, or the like.

Methods for generating and isolating AAV viral vectors suitable for delivery to a subject are known in the art. See, e.g., US Patent 7790449; US Patent 7282199; WO 2003/042397; WO 2005/033321, WO 2006/110689; and US 7588772 B2]. In a one system, a producer cell line is transiently transfected with a construct that encodes the transgene flanked by ITRs and a construct(s) that encodes rep and cap. In a second system, a packaging cell line that stably supplies rep and cap is transiently transfected with a construct encoding the transgene flanked by ITRs. In each of these systems, AAV virions are produced in response to infection with helper adenovirus or herpesvirus, requiring the separation of the rAAVs from contaminating virus. More recently, systems have been developed that do not require infection with helper virus to recover the AAV - the required helper functions (*i.e.*, adenovirus E1, E2a, VA, and E4 or herpesvirus UL5, UL8, UL52, and UL29, and herpesvirus polymerase) are also supplied, *in trans*, by the system. In these newer systems, the helper functions can be supplied by transient transfection of the cells with constructs that encode the required helper functions, or the cells can be engineered to stably contain genes encoding the helper functions, the expression of which can be controlled at the transcriptional or

posttranscriptional level. In yet another system, the transgene flanked by ITRs and rep/cap genes are introduced into insect cells by infection with baculovirus-based vectors. For reviews on these production systems, see generally, e.g., Zhang et al., 2009, "Adenovirus-adeno-associated virus hybrid for large-scale recombinant adeno-associated virus production," Human Gene Therapy 20:922-929, the contents of each of which is incorporated herein by reference in its entirety. Methods of making and using these and other AAV production systems are also described in the following U.S. patents, the contents of each of which is incorporated herein by reference in its entirety: 5,139,941; 5,741,683; 6,057,152; 6,204,059; 6,268,213; 6,491,907; 6,660,514; 6,951,753; 7,094,604; 7,172,893; 7,201,898; 7,229,823; and 7,439,065. See generally, e.g., Grieger & Samulski, 2005, "Adeno-associated virus as a gene therapy vector: Vector development, production and clinical applications," Adv. Biochem. Engin/Biotechnol. 99: 119-145; Buning et al., 2008, "Recent developments in adeno-associated virus vector technology," J. Gene Med. 10:717-733; and the references cited below, each of which is incorporated herein by reference in its entirety. The methods used to construct any embodiment of this invention are known to those with skill in nucleic acid manipulation and include genetic engineering, recombinant engineering, and synthetic techniques. See, e.g., Green and Sambrook et al, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring Harbor, NY (2012). Similarly, methods of generating rAAV virions are well known and the selection of a suitable method is not a limitation on the present invention. See, e.g., K. Fisher et al, (1993) J. Virol., 70:520-532 and US Patent No. 5,478,745.

The rAAV described herein comprise a selected capsid with a vector genome packaged inside. The vector genome (or rAAV genome) comprises 5' and 3' AAV inverted terminal repeats (ITRs), the polynucleotide sequence encoding the fusion protein, and regulatory sequences which direct insertion of the polynucleotide sequence encoding the fusion protein to the genome of a host cell. In one embodiment, the vector genome is the sequence shown in SEQ ID NO: 16 or a sequence sharing at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% identity therewith.

As used herein, a "vector genome" refers to the nucleic acid sequence packaged inside a parvovirus (e.g., rAAV) capsid which forms a viral particle. Such a nucleic acid sequence contains AAV inverted terminal repeat sequences (ITRs). In the examples herein, a vector

genome contains, at a minimum, from 5' to 3', an AAV 5' ITR, coding sequence(s) (i.e., transgene(s)), and an AAV 3' ITR. ITRs from AAV2, a different source AAV than the capsid, or other than full-length ITRs may be selected. In certain embodiments, the ITRs are from the same AAV source as the AAV which provides the rep function during production or a transcomplementing AAV. Further, other ITRs, e.g., self-complementary (scAAV) ITRs, may be used. Both single-stranded AAV and self-complementary (sc) AAV are encompassed with the rAAV. The transgene is a nucleic acid coding sequence, heterologous to the vector sequences, which encodes a polypeptide, protein, functional RNA molecule (e.g., miRNA, miRNA inhibitor) or other gene product, of interest. The nucleic acid coding sequence is operatively linked to regulatory components in a manner which permits transgene transcription, translation, and/or expression in a cell of a target tissue. Suitable components of a vector genome are discussed in more detail herein. In one example, a "vector genome" contains, at a minimum, from 5' to 3', a vector-specific sequence, a nucleic acid sequence encoding GLP-2 constructs operably linked to regulatory control sequences (which direct their expression in a target cell), where the vector-specific sequence may be a terminal repeat sequence which specifically packages the vector genome into a viral vector capsid or envelope protein. For example, AAV inverted terminal repeats are utilized for packaging into AAV and certain other parvovirus capsids.

The AAV sequences of the vector typically comprise the cis-acting 5′ and 3′ inverted terminal repeat sequences (See, e.g., B. J. Carter, in "Handbook of Parvoviruses", ed., P. Tijsser, CRC Press, pp. 155 168 (1990)). The ITR sequences are about 145 bp in length. Preferably, substantially the entire sequences encoding the ITRs are used in the molecule, although some degree of minor modification of these sequences is permissible. The ability to modify these ITR sequences is within the skill of the art. (See, e.g., texts such as Sambrook et al, "Molecular Cloning. A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory, New York (1989); and K. Fisher et al., J. Virol., 70:520 532 (1996)). An example of such a molecule employed in the present invention is a "cis-acting" plasmid containing the transgene, in which the selected transgene sequence and associated regulatory elements are flanked by the 5′ and 3′ AAV ITR sequences. In one embodiment, the ITRs are from an AAV different than that supplying a capsid. In one embodiment, the ITR sequences from AAV2. However, ITRs from other AAV sources may be selected. A shortened version of the 5' ITR,

termed ΔITR, has been described in which the D-sequence and terminal resolution site (trs) are deleted. In certain embodiments, the vector genome includes a shortened AAV2 ITR of 130 base pairs, wherein the external A elements is deleted. Without wishing to be bound by theory, it is believed that the shortened ITR reverts back to the wild-type length of 145 base pairs during vector DNA amplification using the internal (A') element as a template. In other embodiments, full-length AAV 5' and 3' ITRs are used. Where the source of the ITRs is from AAV2 and the AAV capsid is from another AAV source, the resulting vector may be termed pseudotyped. However, other configurations of these elements may be suitable.

Optionally, the GLP-2 constructs described herein may be delivered via viral vectors other than rAAV. Such other viral vectors may include any virus suitable for gene therapy, including but not limited to adenovirus; herpes virus; lentivirus; retrovirus; etc. Suitably, where one of these other vectors is generated, it is produced as a replication-defective viral vector.

A "replication-defective virus" or "viral vector" refers to a synthetic or artificial viral particle in which an expression cassette containing a gene of interest is packaged in a viral capsid or envelope, where any viral genomic sequences also packaged within the viral capsid or envelope are replication-deficient; i.e., they cannot generate progeny virions but retain the ability to infect target cells. In one embodiment, the genome of the viral vector does not include genes encoding the enzymes required to replicate (the genome can be engineered to be "gutless"- containing only the transgene of interest flanked by the signals required for amplification and packaging of the artificial genome), but these genes may be supplied during production. Therefore, it is deemed safe for use in gene therapy since replication and infection by progeny virions cannot occur except in the presence of the viral enzyme required for replication.

Also provided are compositions which include the viral vector constructs described herein. The pharmaceutical compositions described herein are designed for delivery to subjects in need thereof by any suitable route or a combination of different routes. Direct delivery to the liver (optionally via intravenous, via the hepatic artery, or by transplant), oral, inhalation, intranasal, intratracheal, intraarterial, intraocular, intravenous, intramuscular, subcutaneous, intradermal, and other parental routes of administration. The viral vectors described herein may be delivered in a single composition or multiple compositions.

Optionally, two or more different AAV may be delivered, or multiple viruses [see, e.g., WO 2011/126808 and WO 2013/049493]. In another embodiment, multiple viruses may contain different replication-defective viruses (e.g., AAV and adenovirus). In one embodiment, administration is intramuscular. In another embodiment, administration is intravenous.

The replication-defective viruses can be formulated with a physiologically acceptable carrier for use in gene transfer and gene therapy applications. In the case of AAV viral vectors, quantification of the genome copies ("GC") may be used as the measure of the dose contained in the formulation. Any method known in the art can be used to determine the genome copy (GC) number of the replication-defective virus compositions of the invention. One method for performing AAV GC number titration is as follows: Purified AAV vector samples are first treated with DNase to eliminate un-encapsidated AAV genome DNA or contaminating plasmid DNA from the production process. The nuclease resistant particles are then subjected to heat treatment to release the genome from the capsid. The released genomes are then quantitated by real-time PCR using primer/probe sets targeting specific region of the viral genome (usually poly A signal). Another suitable method for determining genome copies is quantitative- PCR (qPCR), particularly the optimized qPCR or digital droplet PCR [Lock Martin, et al, Human Gene Therapy Methods. April 2014, 25(2): 115-125. doi:10.1089/hgtb.2013.131, published online ahead of editing December 13, 2013].

Also, the replication-defective virus compositions can be formulated in dosage units to contain an amount of replication-defective virus that is in the range of about 1.0 x 10^9 GC to about 1.0 x 10^{15} GC. In another embodiment, this amount of viral genome may be delivered in split doses. In one embodiment, the dose is about 1.0 x 10^{10} GC to about 3.0 x 10^{14} GC for an average human subject of about 70 kg. In another embodiment, the dose about 1 x 10^9 GC. For example, the dose of AAV virus may be about 1 x 10^{10} GC, 1 x 10^{11} GC, about 5 X 10^{11} GC, about 1 X 10^{12} GC, about 5 X 10^{12} GC, or about 1 X 10^{13} GC. In another embodiment, the dosage is about 1.0 x 10^9 GC/kg to about 3.0 x 10^{14} GC/kg for a human subject. In another embodiment, the dose about 1 x 10^9 GC/kg. For example, the dose of AAV virus may be about 1 x 10^{10} GC/kg, 1 x 10^{11} GC/kg, about 5 X 10^{11} GC/kg, about 1 X 10^{12} GC/kg, about 5 X 10^{12} GC/kg, or about 1 X 10^{13} GC/kg. In one embodiment, the constructs may be delivered in volumes from 1μL to about 100 mL. As used herein, the term "dosage" or "amount" can refer to the total dosage or amount delivered to the subject in the

course of treatment, or the dosage or amount delivered in a single unit (or multiple unit or split dosage) administration.

The above-described recombinant vectors may be delivered to host cells according to published methods. The rAAV, preferably suspended in a physiologically compatible carrier, may be administered to a desired subject including a human. Suitable carriers may be readily selected by one of skill in the art in view of the indication for which the transfer virus is directed. For example, one suitable carrier includes saline, which may be formulated with a variety of buffering solutions (e.g., phosphate buffered saline). Other exemplary carriers include sterile saline, lactose, sucrose, calcium phosphate, gelatin, dextran, agar, pectin, peanut oil, sesame oil, and water. The selection of the carrier is not a limitation of the present invention.

In another embodiment, the composition includes a carrier, diluent, excipient and/or adjuvant. In certain embodiments, for administration to a human patient, the rAAV is suitably suspended in an aqueous solution containing saline, a surfactant, and a pharmaceutically and/or physiologically compatible salt or mixture of salts. Suitably, the formulation is adjusted to a physiologically acceptable pH, e.g., in the range of pH 6 to 9, or pH 6.0 to 7.5, or pH 6.2 to 7.7, or pH 6.5 to 7.5, pH 7.0 to 7.7, or pH 7.2 to 7.8, or about 7.0. In certain embodiments, the formulation is adjusted to a pH of about 6.0, about 6.1, about 6.2, about 6.3, about 6.4, about 6.5, about 6.6, about 6.7, about 6.8, about 6.9, about 7.0, about 7.1, about 7.2, about 7.3 about 7.4, about 7.5, about 7.6, about 7.7, or about 7.8. In certain embodiments, a pH of about 7.28 to about 7.32, about 6.0 to about 7.5, about 6.2 to about 7.7, about 7.5 to about 7.8, about 6.0, about 6.1, about 6.2, about 6.3, about 6.4, about 6.5, about 6.6, about 6.7, about 6.8, about 6.9, about 7.0, about 7.1, about 7.2, about 7.3 about 7.4, about 7.5, about 7.6, about 7.7, or about 7.8 may be desired. In certain embodiments, for intravenous delivery, a pH of about 6.8 to about 7.2 may be desired. However, other pHs within the broadest ranges and these subranges may be selected for other route of delivery.

Optionally, the compositions of the invention may contain, in addition to the rAAV and/or variants and carrier(s), other conventional pharmaceutical ingredients, such as preservatives, or chemical stabilizers. Suitable exemplary preservatives include chlorobutanol, potassium sorbate, sorbic acid, sulfur dioxide, propyl gallate, the parabens,

ethyl vanillin, glycerin, phenol, and parachlorophenol. Suitable chemical stabilizers include gelatin and albumin.

As used herein, "carrier" includes any and all solvents, dispersion media, vehicles, coatings, diluents, antibacterial and antifungal agents, isotonic and absorption delaying agents, buffers, carrier solutions, suspensions, colloids, and the like. The use of such media and agents for pharmaceutical active substances is well known in the art. Supplementary active ingredients can also be incorporated into the compositions. The phrase "pharmaceutically-acceptable" refers to molecular entities and compositions that do not produce an allergic or similar untoward reaction when administered to a host. Delivery vehicles such as liposomes, nanocapsules, microparticles, microspheres, lipid particles, vesicles, and the like, may be used for the introduction of the compositions of the present invention into suitable host cells. In particular, the rAAV vector delivered transgenes may be formulated for delivery either encapsulated in a lipid particle, a liposome, a vesicle, a nanosphere, or a nanoparticle or the like.

In one embodiment, a composition includes a final formulation suitable for delivery to a subject, e.g., is an aqueous liquid suspension buffered to a physiologically compatible pH and salt concentration. Optionally, one or more surfactants are present in the formulation. In another embodiment, the composition may be transported as a concentrate which is diluted for administration to a subject. In other embodiments, the composition may be lyophilized and reconstituted at the time of administration.

A suitable surfactant, or combination of surfactants, may be selected from among non-ionic surfactants that are nontoxic. In one embodiment, a difunctional block copolymer surfactant terminating in primary hydroxyl groups is selected, e.g., such as Pluronic® F68 [BASF], also known as Poloxamer 188, which has a neutral pH, has an average molecular weight of 8400. Other surfactants and other Poloxamers may be selected, i.e., nonionic triblock copolymers composed of a central hydrophobic chain of polyoxypropylene (poly(propylene oxide)) flanked by two hydrophilic chains of polyoxyethylene (poly(ethylene oxide)), SOLUTOL HS 15 (Macrogol-15 Hydroxystearate), LABRASOL (Polyoxy caprylic glyceride), polyoxy 10 oleyl ether, TWEEN (polyoxyethylene sorbitan fatty acid esters), ethanol and polyethylene glycol. In one embodiment, the formulation contains a poloxamer. These copolymers are commonly named with the letter "P" (for poloxamer) followed by

three digits: the first two digits x 100 give the approximate molecular mass of the polyoxypropylene core, and the last digit x 10 gives the percentage polyoxyethylene content. In one embodiment Poloxamer 188 is selected. The surfactant may be present in an amount up to about 0.0005 % to about 0.001% of the suspension.

Dosages of the vector depends primarily on factors such as the condition being treated, the age, weight and health of the patient, and may thus vary among patients. For example, a therapeutically effective human dosage of viral vector is generally in the range of from about 25 to about 1000 microliters to about 100 mL of solution containing concentrations of from about 1 x 10^9 to 1 x 10^{16} genomes virus vector (to treat an average subject of 70 kg in body weight) including all integers or fractional amounts within the range, and preferably 1.0 x 10^{12} GC to 1.0 x 10^{13} GC for a human patient. The composition of the invention may be delivered in a volume of from about 0.1 µL to about 10 mL, including all numbers within the range, depending on the size of the area to be treated, the viral titer used, the route of administration, and the desired effect of the method. In one embodiment, the volume is about 50 µL. In another embodiment, the volume is about 70 µL. In another embodiment, the volume is about 100 µL. In another embodiment, the volume is about 125 µL. In another embodiment, the volume is about 150 µL. In another embodiment, the volume is about 175 µL. In yet another embodiment, the volume is about 200 µL. In another embodiment, the volume is about 250 µL. In another embodiment, the volume is about 300 µL. In another embodiment, the volume is about 450 µL. In another embodiment, the volume is about 500 µL. In another embodiment, the volume is about 600 µL. In another embodiment, the volume is about 750 µL. In another embodiment, the volume is about 850 µL. In another embodiment, the volume is about 1000 µL. In another embodiment, the volume is about 1.5 mL. In another embodiment, the volume is about 2 mL. In another embodiment, the volume is about 2.5 mL. In another embodiment, the volume is about 3 mL. In another embodiment, the volume is about 3.5 mL. In another embodiment, the volume is about 4 mL. In another embodiment, the volume is about 5 mL. In another embodiment, the volume is about 5.5 mL. In another embodiment, the volume is about 6 mL. In another embodiment, the volume is about 6.5 mL. In another embodiment, the volume is about 7 mL. In another embodiment, the volume is about 8 mL. In another embodiment, the volume is

about 8.5 mL. In another embodiment, the volume is about 9 mL. In another embodiment, the volume is about 9.5 mL. In another embodiment, the volume is about 10 mL.

In some embodiments, a concentration of a recombinant adeno-associated virus carrying a nucleic acid sequence encoding the desired transgene under the control of the regulatory sequences desirably ranges from about 10^7 and 10^{14} genome copies per milliliter (GC/mL) in a composition.

In one embodiment, the dosage of rAAV in a composition is from about 1.0 x 10^9 GC/kg of body weight to about 1.5 x 10^{13} GC/kg. In one embodiment, the dosage is about 1.0 x 10^{10} GC/kg. In one embodiment, the dosage is about 1.0 x 10^{11} GC/kg. In one embodiment, the dosage is about 1.0 x 10^{12} GC/kg. In one embodiment, the dosage is about 5.0 x 10^{12} GC/kg. In one embodiment, the dosage is about 1.0 x 10^{13} GC/kg. All ranges described herein are inclusive of the endpoints.

In one embodiment, the effective dosage (total genome copies delivered) is from about 10^7 to 10^{13} genome copies. In one embodiment, the total dosage is about 10^8 genome copies. In one embodiment, the total dosage is about 10^9 genome copies. In one embodiment, the total dosage is about 10^{10} genome copies. In one embodiment, the total dosage is about 10^{11} genome copies. In one embodiment, the total dosage is about 10^{12} genome copies. In one embodiment, the total dosage is about 10^{13} genome copies. In one embodiment, the total dosage is about 10^{14} genome copies. In one embodiment, the total dosage is about 10^{15} genome copies.

It is desirable that the lowest effective concentration of virus be utilized in order to reduce the risk of undesirable effects, such as toxicity. Still other dosages and administration volumes in these ranges may be selected by the attending physician, taking into account the physical state of the subject, preferably human, being treated, the age of the subject, the particular disorder and the degree to which the disorder, if progressive, has developed.

In certain embodiments, the composition comprises an rAAV comprising an inducible GLP-2 agonist construct. In certain embodiments, the inducing agent or molecule is a rapamycin or a rapalog. In certain embodiments, the inducing agent is rapamycin, and is administered at least one or more, at least two or more, at least three or more times following rAAV-comprising composition. In some embodiments the rapamycin is administered at dose at least about 4 to at least about 40 nM. In certain embodiments, the inducing agent (i.e.,

rapamycin) is administered at a dose at least about 0.1 mg/kg to at least about 3.0 mg/kg. In certain embodiments, the inducing agent (i.e., rapamycin) is administered at a dose at least about 0.5 mg/kg to at least about 2.0 mg/kg.

The viral vectors and other constructs described herein may be used in preparing a medicament for delivering a GLP-2 fusion protein construct to a subject in need thereof, supplying GLP-2 having an increased half-life to a subject, and/or for treating SBS in a subject. Thus, in another aspect a method of treating SBS is provided. The method includes administering a composition as described herein to a subject in need thereof. In one embodiment, the composition includes a viral vector containing a GLP-2 fusion protein expression cassette, as described herein.

As used herein, the term "treatment" or "treating" is defined encompassing administering to a subject one or more compounds or compositions described herein for the purposes of amelioration of one or more symptoms of short bowel syndrome. "Treatment" can thus include one or more of reducing progression of SBS, reducing the severity of the symptoms, removing the disease symptoms, delaying progression of disease, or increasing efficacy of therapy in a given subject.

In another embodiment, a method for treating SBS in a subject is provided. The method includes administering a viral vector comprising a nucleic acid molecule comprising a sequence encoding a fusion protein as described herein. In one embodiment, the subject is a human.

A course of treatment may optionally involve repeat administration of the same viral vector (e.g., an AAVrh91 vector) or a different viral vector (e.g., an AAVrh91 and an AAV3B.AR2.12). Still other combinations may be selected using the viral vectors described herein. Optionally, the composition described herein may be combined in a regimen involving nutritional therapy (enteral or parenteral nutrition), medications, such as those used to control stomach acid, reduce diarrhea, or improve intestinal absorption, or a GLP-2 analog, or surgery. In certain embodiments, the AAV vector and the combination therapy are administered essentially simultaneously. In other embodiments, the AAV vector is administered first. In other embodiments, the combination therapy is delivered first.

In one embodiment, the composition is administered in combination with an effective amount of a GLP-2 analog. Various commercially available GLP-2 products are known in the art, including, without limitation, teduglutide, glepaglutide, and apraglutide.

In some embodiments, combination of the rAAV described herein with GLP-2 analog decreases GLP-2 analog dose requirements in the subject, as compared to prior to treatment with the viral vector. Such dose requirements may be reduced by 10% or more, 20% or more, 30% or more, 40% or more, 50% or more, 60% or more, 70% or more, 80% or more, or 90% or more. The treating physician may determine the correct dosage of GLP-2 analog needed by the subject. For example, the subject may be being treated using GLP-2 analog or other therapy, which the treating physician may continue upon administration of the AAV vector. Such GLP-2 analog or other co-therapy may be continued, reduced, or discontinued as needed subsequently.

In one embodiment, composition comprising the expression cassette, vector genome, rAAV, or other composition described herein for gene therapy is delivered as a single dose per patient. In one embodiment, the subject is delivered a therapeutically effective amount of a composition described herein. As used herein, a "therapeutically effective amount" refers to the amount of the expression cassette or vector, or a combination thereof that delivers and expresses in the target cells an amount of GLP-2-Fc sufficient to reach therapeutic goal. The therapeutically effective amount may be selected by the treating physician, or guided based on previously determined guidelines. For example, teduglutide may be provided at an initial dose of 0.05 mg/kg subcutaneously daily. The dose may be increased in 0.025 mg/kg increments for subjects with moderate-to-severe renal impairment. The rAAV may be delivered to the subject and then supplemented with oral or subcutaneous teduglutide, or other medication as needed to reach the equivalent of the desired dosage of 0.05 mg/kg daily.

In certain embodiments, the therapeutic goal is to ameliorate or treat one or more of the symptoms of SBS. A therapeutically effective amount may be determined based on an animal model, rather than a human patient.

As used herein when used to refer to vp capsid proteins, the term "heterogenous" or any grammatical variation thereof, refers to a population consisting of elements that are not the same, for example, having vp1, vp2 or vp3 monomers (proteins) with different modified amino acid sequences. SEQ ID NO: 32 provides the encoded amino acid sequence of the

AAVrh91 vp1 protein. The term "heterogenous" as used in connection with vp1, vp2 and vp3 proteins (alternatively termed isoforms), refers to differences in the amino acid sequence of the vp1, vp2 and vp3 proteins within a capsid. The AAV capsid contains subpopulations within the vp1 proteins, within the vp2 proteins and within the vp3 proteins which have modifications from the predicted amino acid residues. These subpopulations include, at a minimum, certain deamidated asparagine (N or Asn) residues. For example, certain subpopulations comprise at least one, two, three or four highly deamidated asparagines (N) positions in asparagine - glycine pairs and optionally further comprising other deamidated amino acids, wherein the deamidation results in an amino acid change and other optional modifications.

As used herein, a "subpopulation" of vp proteins refers to a group of vp proteins which has at least one defined characteristic in common and which consists of at least one group member to less than all members of the reference group, unless otherwise specified. For example, a "subpopulation" of vp1 proteins is at least one (1) vp1 protein and less than all vp1 proteins in an assembled AAV capsid, unless otherwise specified. A "subpopulation" of vp3 proteins may be one (1) vp3 protein to less than all vp3 proteins in an assembled AAV capsid, unless otherwise specified. For example, vp1 proteins may be a subpopulation of vp proteins; vp2 proteins may be a separate subpopulation of vp proteins, and vp3 are yet a further subpopulation of vp proteins in an assembled AAV capsid. In another example, vp1, vp2 and vp3 proteins may contain subpopulations having different modifications, e.g., at least one, two, three or four highly deamidated asparagines, e.g., at asparagine - glycine pairs.

As used herein, a "stock" of rAAV refers to a population of rAAV. Despite heterogeneity in their capsid proteins due to deamidation, rAAV in a stock are expected to share an identical vector genome. A stock can include rAAV having capsids with, for example, heterogeneous deamidation patterns characteristic of the selected AAV capsid proteins and a selected production system. The stock may be produced from a single production system or pooled from multiple runs of the production system. A variety of production systems, including but not limited to those described herein, may be selected. As used herein the terms "GLP-2 construct", "GLP-2 expression construct" and synonyms include the GLP-2 sequence as described herein in combination with a leader and fusion domain. The terms "GLP-2 construct", "GLP-2 expression construct" and synonyms can be

used to refer to the nucleic acid sequences encoding the GLP-2 fusion protein or the expression products thereof.

The terms "percent (%) identity", "sequence identity", "percent sequence identity", or "percent identical" in the context of nucleic acid sequences refers to the bases in the two sequences which are the same when aligned for correspondence. The length of sequence identity comparison may be over the full-length of the genome, the full-length of a gene coding sequence, or a fragment of at least about 100 to 150 nucleotides, or as desired. However, identity among smaller fragments, e.g., of at least about nine nucleotides, usually at least about 20 to 24 nucleotides, at least about 28 to 32 nucleotides, at least about 36 or more nucleotides, may also be desired. Multiple sequence alignment programs are also available for nucleic acid sequences. Examples of such programs include, "Clustal W", "CAP Sequence Assembly", "BLAST", "MAP", and "MEME", which are accessible through Web Servers on the internet. Other sources for such programs are known to those of skill in the art. Alternatively, Vector NTI utilities are also used. There are also a number of algorithms known in the art that can be used to measure nucleotide sequence identity, including those contained in the programs described above. As another example, polynucleotide sequences can be compared using Fasta™, a program in GCG Version 6.1. Fasta™ provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences. For instance, percent sequence identity between nucleic acid sequences can be determined using Fasta™ with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) as provided in GCG Version 6.1, herein incorporated by reference.

By the term "highly conserved" is meant at least 80% identity, preferably at least 90% identity, and more preferably, over 97% identity. Identity is readily determined by one of skill in the art by resort to algorithms and computer programs known by those of skill in the art.

Unless otherwise specified by an upper range, it will be understood that a percentage of identity is a minimum level of identity and encompasses all higher levels of identity up to 100% identity to the reference sequence. Unless otherwise specified, it will be understood that a percentage of identity is a minimum level of identity and encompasses all higher levels of identity up to 100% identity to the reference sequence. For example, "95% identity" and

"at least 95% identity" may be used interchangeably and include 95%, 96%, 97%, 98%, 99%, and up to 100% identity to the referenced sequence, and all fractions therebetween.

The terms "percent (%) identity", "sequence identity", "percent sequence identity", or "percent identical" in the context of amino acid sequences refers to the residues in the two sequences which are the same when aligned for correspondence. Percent identity may be readily determined for amino acid sequences over the full-length of a protein, polypeptide, about 70 amino acids to about 100 amino acids, or a peptide fragment thereof or the corresponding nucleic acid sequence coding sequencers. A suitable amino acid fragment may be at least about 8 amino acids in length, and may be up to about 150 amino acids. Generally, when referring to "identity", "homology", or "similarity" between two different sequences, "identity", "homology" or "similarity" is determined in reference to "aligned" sequences. "Aligned" sequences or "alignments" refer to multiple nucleic acid sequences or protein (amino acids) sequences, often containing corrections for missing or additional bases or amino acids as compared to a reference sequence. Alignments are performed using any of a variety of publicly or commercially available Multiple Sequence Alignment Programs. Sequence alignment programs are available for amino acid sequences, e.g., the "Clustal X", "MAP", "PIMA", "MSA", "BLOCKMAKER", "MEME", and "Match-Box" programs. Generally, any of these programs are used at default settings, although one of skill in the art can alter these settings as needed. Alternatively, one of skill in the art can utilize another algorithm or computer program which provides at least the level of identity or alignment as that provided by the referenced algorithms and programs. See, e.g., J. D. Thomson et al, Nucl. Acids. Res., "A comprehensive comparison of multiple sequence alignments", 27(13):2682-2690 (1999).

It is to be noted that the term "a" or "an" refers to one or more. As such, the terms "a" (or "an"), "one or more," and "at least one" are used interchangeably herein.

The words "comprise", "comprises", and "comprising" are to be interpreted inclusively rather than exclusively. The words "consist", "consisting", and its variants, are to be interpreted exclusively, rather than inclusively. While various embodiments in the specification are presented using "comprising" language, under other circumstances, a related embodiment is also intended to be interpreted and described using "consisting of" or "consisting essentially of" language.

"Patient" or "subject" as used herein means a mammalian animal, including a human, a veterinary or farm animal, a domestic animal or pet, and animals normally used for clinical research. In one embodiment, the subject of these methods and compositions is a human. In another embodiment, the subject is not a feline.

As used herein, the term "about" means a variability of 10% (±10%, e.g., ±1, ±2, ±3, ±4, ±5, ±6, ±7, ±8, ±9, ±10, or values therebetween) from the reference given, unless otherwise specified.

In certain instances, the term "E+#" or the term "e+#" is used to reference an exponent. For example, "5E10" or "5e10" is 5 x 10^{10}. These terms may be used interchangeably.

The term "regulation" or variations thereof as used herein refers to the ability of a composition to inhibit one or more components of a biological pathway.

As used herein, "disease", "disorder" and "condition" are used interchangeably, to indicate an abnormal state in a subject.

Unless defined otherwise in this specification, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art and by reference to published texts, which provide one skilled in the art with a general guide to many of the terms used in the present application.

A reference to "one embodiment" or "another embodiment" in describing an embodiment does not imply that the referenced embodiment is mutually exclusive with another embodiment (e.g., an embodiment described before the referenced embodiment), unless expressly specified otherwise.

EXAMPLES

The following examples are provided to illustrate various embodiments of the present invention. The EXAMPLES are not intended to limit the present invention in any way.

Example 1 - Construction of GLP-2 vectors

GLP-2 agonists are challenging to express via adeno-associated virus (AAV). GLP-2 is normally expressed from the glucagon precursor protein, which requires tissue specific proteases and produces unwanted proteins. Expression systems using traditional heterologous

signal peptides yield low expression. Expression systems using heterologous propeptides with universal protease cleavage sites yield foreign protein sequences that could be targets for T cells.

More specifically, vectors were constructed in which a leader sequence was placed upstream of one of several GLP-2 receptor agonist amino acid sequences followed by a fusion domain. See, e.g., FIG. 3. The resulting protein sequence was back-translated, followed by addition of a kozak consensus sequence, stop codon, and cloning sites. The sequences were produced, and cloned into an expression vector containing a CMV promoter under the control of an inducible expression system. The expression construct was flanked by AAV2 ITRs. The resulting plasmid is called pAAV.Z121.hGLP2.G2.Fc.rBG.

In a second construct, the sequences were produced, and cloned into an expression vector containing a CB7 constitutive promoter.

Example 2 – *In vitro* expression

The following constructs were packaged in an AAVrh91 vector by triple transfection, as previously described.

AAVrh91.CB7.CI.hGLP-2-Fc.rBG

AAVrh91.CB7.CI.hGLP-2-SA.rBG

GLP-2-SA fusions were measured in culture supernatants of HEK293 cells transfected with vector for inducible human GLP-2-SA with human Thrombin signal sequence. GLP2-SA was identified by gel electrophoresis using spyroRuby stain. FIG. 4A and FIG. 4B.

Example 3 – Pilot expression in Rag1KO mice

Rag1KO female mice were treated with an injection of the vector AAVrh91.CB7.CI.hGLP-2-SA.rBG (1×10^{11} GC/ mouse) or AAVrh91.CB7.CI.hGLP-2-Fc.rBG (1×10^{11} GC/ mouse) via IM route of administration. Serum was serially collected by separating whole blood in serum separator tubes containing 5 microliters DPP-IV inhibitor (Millipore) and assayed for active GLP-2 expression and activity as above. Vector was injected at day 0 and mice were necropsied at day 56. Serum GLP-2 concentrations (nM) are shown in FIG. 5A and are an estimation based on Fc fusion standard. Serum expression levels reached increased through day 14 post dosing.

Small intestines were weighed and measured at necropsy. The length and weight of small intestine increased significantly as compared to control animals (FIG. 5B). In addition, vector treated intestine show healthy enterocyte growth as compared to vehicle treated animals. (FIG. 5C). Serum citrulline levels (uM), a biomarker of gut surface area, were measured. (FIG. 5D).

Expression of GLP-2 was compared with albumin fusion construct. Fc fusion construct showed greater GLP-2 expression (FIG. 6A). The GLP.G2.Fc fusion shows significantly lower potency (EC50=13.6nM) as compared to hGLP (EC50=0.4nM) (FIG. 6B). These data indicate that AAV-mediated expression of GLP-2 agonist demonstrates substantial and durable increases in gut surface area which should be therapeutic in short bowel syndromes.

Example 4 – Dosage study in NHPs

In this study, we examine expression of human GLP-2 in nonhuman primates (NHPs; i.e., rhesus macaques). FIG. 7 shows an outline of the study. Briefly, 2 NHPs were administered AAVrh91.CB7.CI.hGLP-2-Fc.rBG via intramuscular injection (IM) at a dose of 1 x 10^{13} (1e13) GC/kg (E185NG) and a of dose 5 x 10^{10} (1e10) GC/kg (BM239H).

GLP-2 expression and potency, liver enzymes, and citrulline levels were measured. Necropsy is performed at day 60. FIG. 8A shows plasma level of GLP-2-Fc fusion protein. FIG. 8B shows serum citrulline, a biomarker of gut surface area. FIG. 8C shows detection of anti-GLP-2-Fc antibody in NHP serum at 1:100 dilution. AAV-mediated expression of GLP-2-Fc fusion demonstrates substantial increase in gut surface area at the high dose until anti-GLP-2-Fc antibodies reduced its expression demonstrating its therapeutic efficacy on short bowel syndromes.

Example 5 – Long term efficacy and safety study in Rag1KO mice

Rag1KO female mice were treated with an injection of the vector AAVrh91.CB7.CI.hGLP-2-Fc.rBG at a dosage of 1 x 10^{10} GC/ mouse, 3 x 10^{10} GC/ mouse, or 1 x 10^{11} GC/ mouse via IM route of administration. The study design is shown in FIG. 9A. hGLP2-Fc levels and body weight were measured throughout the study. The highest vector dosage showed highest serum GLP2 levels (FIG. 9B), while body weights were relatively consistent amongst all groups, with vehicle trending lowest (FIG. 9C, D). At day 28

necropsy, small intestine length and weight were measured. While SI length showed modest increase with dosage increase as compared to vehicle (FIG. 9E), SI weight was significantly increased in the two highest dosage groups (FIG. 9F).

All documents cited in this specification, are incorporated herein by reference. US Provisional Patent Application No. 63/316,219, filed March 3, 2022, is incorporated herein by reference. While the invention has been described with reference to particular embodiments, it will be appreciated that modifications can be made without departing from the spirit of the invention. Such modifications are intended to fall within the scope of the appended claims.

WHAT IS CLAIMED IS:

1. A composition comprising a nucleic acid comprising a sequence encoding a fusion protein comprising a GLP-2 protein and a IgG4 Fc, albumin or XTEN polypeptide, wherein the fusion protein has the sequence of SEQ ID NO: 13, 15, 17, 19, 21, or 23, or a sequence at least 99% identical thereto.

2. The composition according to any one of claims 1 to 7, wherein the sequence encoding the fusion protein is SEQ ID NO: 14, 16, 18, 20, 22, or 24, or a sequence sharing at least 75% identical thereto.

3. The composition comprising a viral vector comprising:

(a) an adeno-associated virus (AAV) capsid , and

(b) a vector genome packaged in the AAV capsid, said vector genome comprising AAV inverted terminal repeats (ITRs), the coding sequence for comprising a GLP-2 protein and a IgG4 Fc, albumin or XTEN polypeptide, wherein the fusion protein has the sequence of SEQ ID NO: 13, 15, 17, 19, 21, or 23, or a sequence at least 99% identical thereto, and regulatory sequences which direct expression of the fusion protein.

4. The composition according to any one of claims 1 to 3, wherein the viral vector is an rAAV having the AAV capsid of AAVrh91 or AAVhu68.

5. The composition according to one of claims 1 to 4, wherein the fusion protein is under the control of an inducible gene expression system.

6. The composition according to claim 5, wherein the inducible gene expression system comprises a regulatable promoter, an activation domain, and a DNA binding domain.

7. The composition according to any one of claims 3 to 6, wherein the AAV inverted terminal repeats (ITRs) are an AAV2 5' ITR and an AAV2 3' ITR which flank the fusion protein coding sequence and regulatory sequences.

8. The composition according to any one of claims 3 to 7, wherein the vector genome comprises a CB7 promoter and a rabbit globin poly A.

9. The composition according to any one of claims 5 to 8, wherein the inducible gene expression system comprises

(a) an activation domain comprising a transactivation domain and a FKBP12-rapamycin binding (FRB) domain of FKBP12-rapamycin-associated protein (FRAP);

(b) a DNA binding domain comprising a zinc finger homeodomain (ZFHD) and one, two or three FK506 binding protein domain (FKBP) subunit genes; and

(c) at least one copy of the binding site for ZFHD followed by a minimal promoter, and

(d) a regulatable promoter.

10. The composition according to claim 9, wherein the inducible gene expression system is comprised in one vector.

11. The composition according to claim 9, wherein the inducible gene expression system is comprised in two vectors.

12. The composition according to any one of claims 9 to 11, wherein the transactivation domain comprises a portion of NF-κB p65.

13. The composition according to any one of claims 9 to 12, wherein the regulatable promoter is a constitutive promoter.

14. The composition according to claim 12, wherein the regulatable promoter is a CMV promoter.

15. The composition according to any one of claims 9 to 14, further comprising an IRES or 2A.

16. The composition according to any one of claims 9 to 15, comprising at least 8 copies of the binding site for ZFHD.

17. A composition according to any one of claims 5 to 16 comprising a regulatable promoter; an activation domain comprising a p65 transactivation domain and a FKBP12-rapamycin binding (FRB) domain of FKBP12-rapamycin-associated protein (FRAP); a DNA binding domain comprising a zinc finger homeodomain (ZFHD) and three FK506 binding protein domain (FKBP) subunit genes; 12 copies of the binding site for ZFHD, and a sequence encoding a fusion protein comprising a GLP-2 analog and a human IgG4 Fc.

18. A pharmaceutical composition suitable for use in treating a metabolic disease in a subject comprising an aqueous liquid and the composition according to any one of claims 1 to 17.

19. The composition according to any one of claims 1 to 18, for use in a method for treating a subject having a metabolic disease.

20. Use of the composition according to any one of claims 1 to 19 in the manufacture of a medicament for treating a subject having a metabolic disease.

21. The composition or use according to any one of claims 1 to 20, wherein the composition is formulated to be administered a dose of 1 x 10^9 GC/kg to 5x 10^{13} GC/kg of the rAAV.

22. The composition or use according to any one of claims 1 to 20, wherein the patient is a human and is administered a dose of 1 x 10^{10} to 1.5 x 10^{15} GC of the rAAV.

23. The composition or use according to any one of claims 1 to 22, wherein the rAAV is delivered intramuscularly or intravenously.

24. A method of treating a subject having a metabolic disease, comprising delivering to the subject a recombinant adeno-associated virus (rAAV) having an AAV capsid from adeno-associated virus rh91 or hu68, and a vector genome packaged in the AAV capsid, said vector genome comprising AAV inverted terminal repeats (ITRs), a sequence encoding a fusion protein comprising a GLP-2 analog and a human IgG4 Fc, and regulatory sequences which direct expression of the fusion protein.

25. The method according to claim 24, wherein the patient is administered a composition according to any one of claims 1 to 18.

26. The method according to claim 24 or 25, wherein the patient is administered a dose of 1 x 10^9 GC/kg to 5x 10^{13} GC/kg body mass of the rAAV.

27. The method according to any one of claims 24 to 26, wherein the rAAV is delivered intramuscularly or intravenously.

28. The composition according to any one of claims 1 to 18, for treating diabetes in a human.

FIG. 1

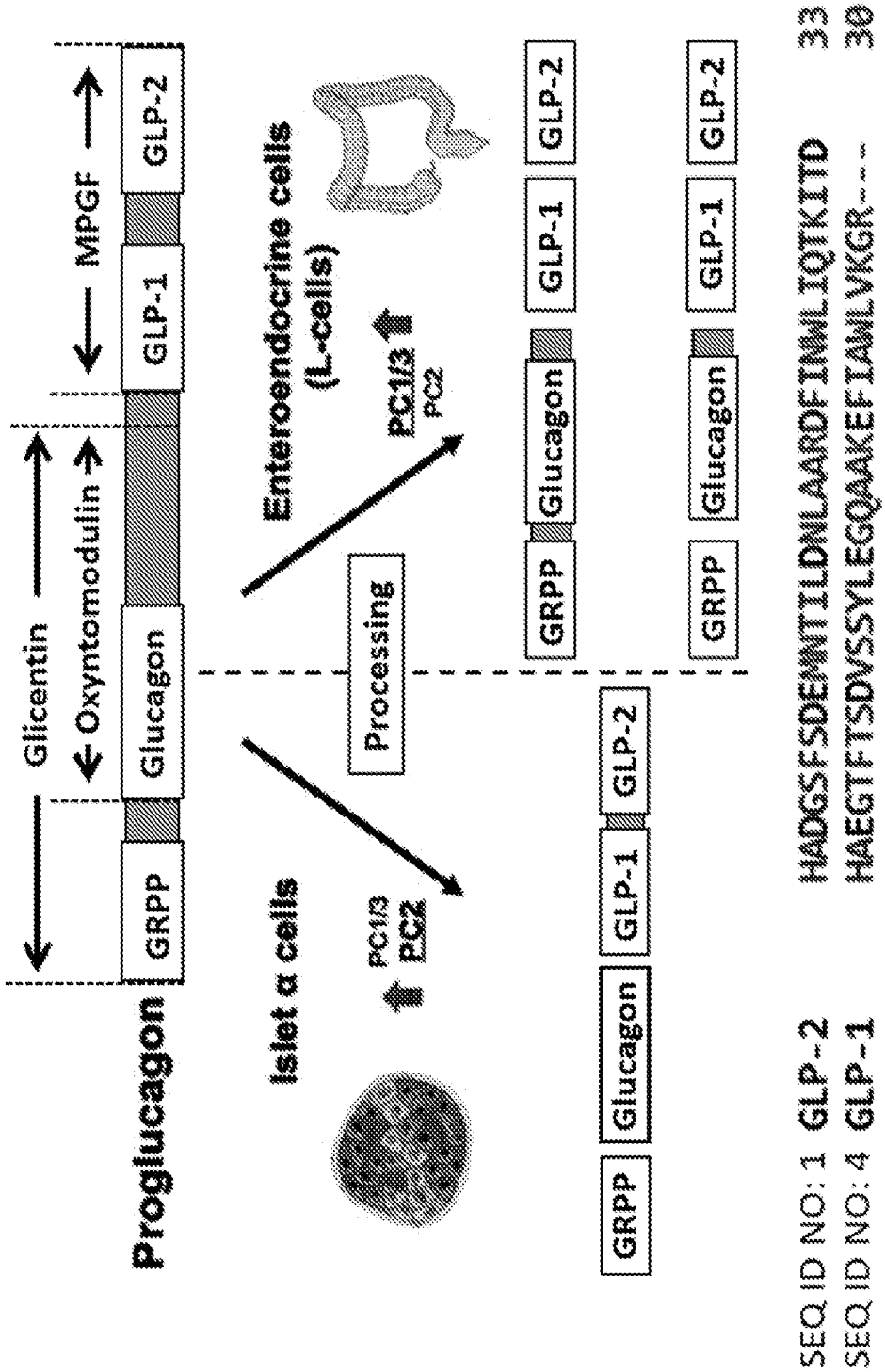

Glicentin
Oxyntomodulin
MPGF
GRPP
Glucagon
GLP-1
GLP-2
Proglucagon
Processing
PC1/3
PC2
Islet α cells
Enteroendocrine cells
(L-cells)
GRPP
Glucagon
GLP-1
GLP-2
SEQ ID NO: 1 GLP-2 HADGSFSDEMNTILDNLAARDFINWLIQTKITD 33
SEQ ID NO: 4 GLP-1 HAEGTFTSDVSSYLEGQAAKEFIAWLVKGR~~~ 30

FIG. 2

TABLE 1

Amino acid sequence and molecular weight of hGLP-2 and GLP-2 analogs

Amino acid substitutions relative to native hGLP-2 are shown in bold underlined letters

	Peptide	Mol. wt.	Amino Acid Sequence																
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
SEQ ID NO: 1	hGLP-2	3766	H	A	D	G	S	F	S	D	E	M	N	T	I	L	D	N	L
SEQ ID NO: 2	Glepaglutide	4316	H	**G**	**E**	G	**T**	F	S	**S**	E	**L**	**A**	T	I	L	D	**A**	L
SEQ ID NO: 3	Teduglutide	3752	H	**G**	D	G	S	F	S	D	E	M	N	T	I	L	D	N	L

	Peptide	Mol. wt.	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	
SEQ ID NO: 1	hGLP-2	3766	A	A	R	D	F	I	N	W	L	I	Q	T	K	I	T	D	OH
SEQ ID NO: 2	Glepaglutide	4316	A	A	R	D	F	I	**A**	W	L	I	**A**	T	K	I	T	D	**KKKKKK**NH_2
SEQ ID NO: 3	Teduglutide	3752	A	A	R	D	F	I	N	W	L	I	Q	T	K	I	T	D	OH

f, D-phenylalanine.

FIG. 3A

GLP2.G2.Fc

FIG. 3B

GLP2.G2.SA

Made human and cynomolgus versions

FIG. 3C

GLP2 is not identical across species

Human	**HADGSFSDEMNTILDNLAARDFINWLIQTKITD**	**33**	SEQ ID NO: 1
Cyno	**HADGSFSDEMNTVLDNLATRDFINWLIQTKITD**	**33**	SEQ ID NO: 5
Mouse	**HADGSFSDEMSTILDNLATRDFINWLIQTKITD**	**33**	SEQ ID NO: 6
	********** .*:***** :***************		

FIG. 4A

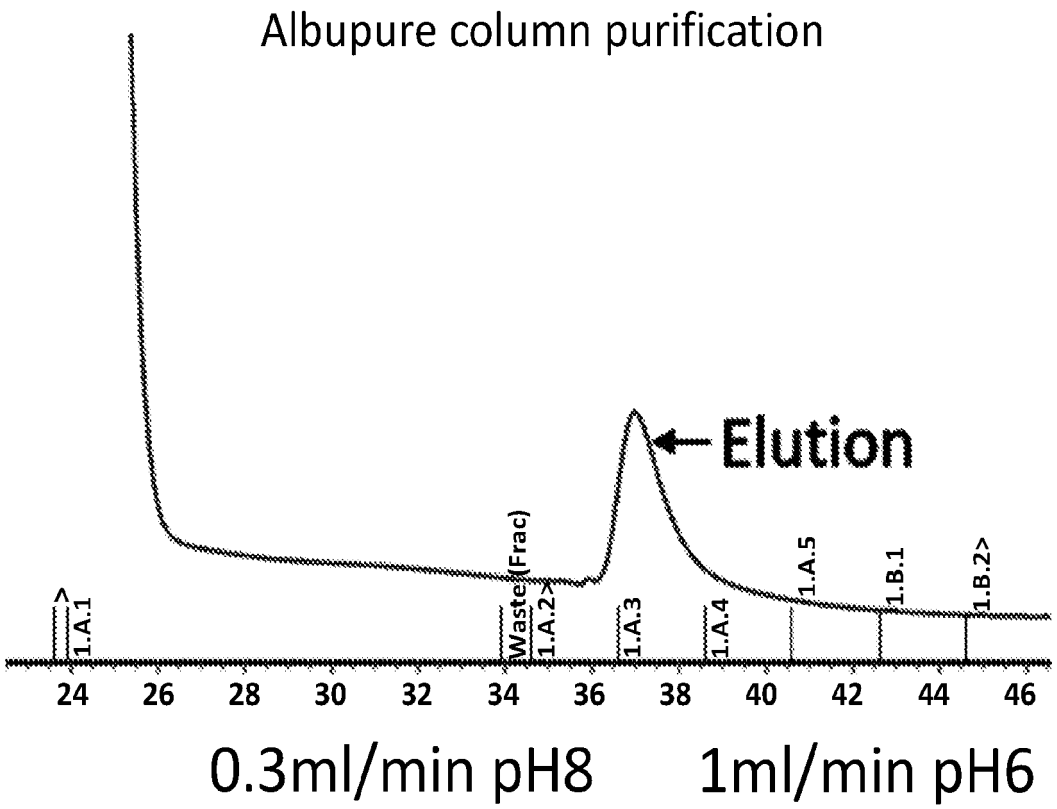

Albupure column purification
Elution
1.A.1
Waste(Frac)
1.A.2>
1.A.3
1.A.4
1.A.5
1.B.1
1.B.2>
24
26
28
30
32
34
36
38
40
42
44
46
0.3ml/min pH8
1ml/min pH6

FIG. 4B

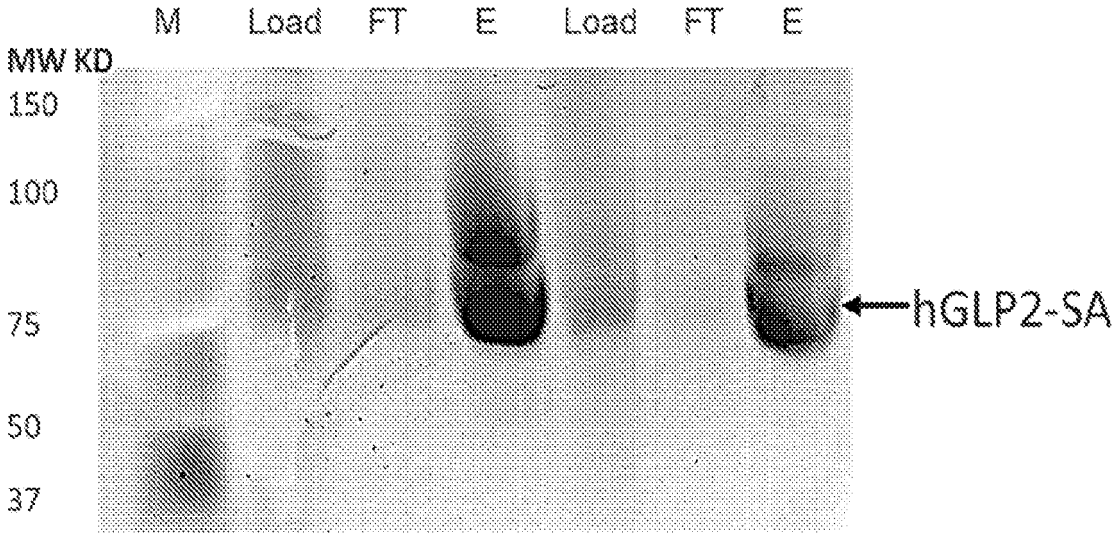

M
Load
FT
E
Load
FT
E
MW KD
150
100
75
50
37
hGLP2-SA

FIG. 5A AAVrh91.CB7.CI.hGLP2.G2.SA.rBG

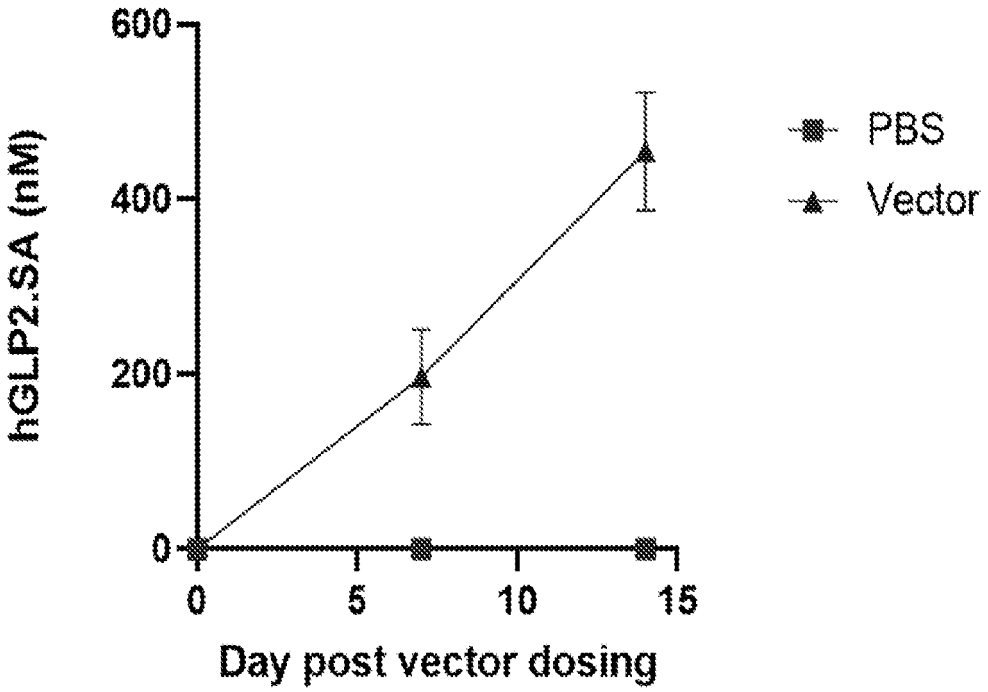
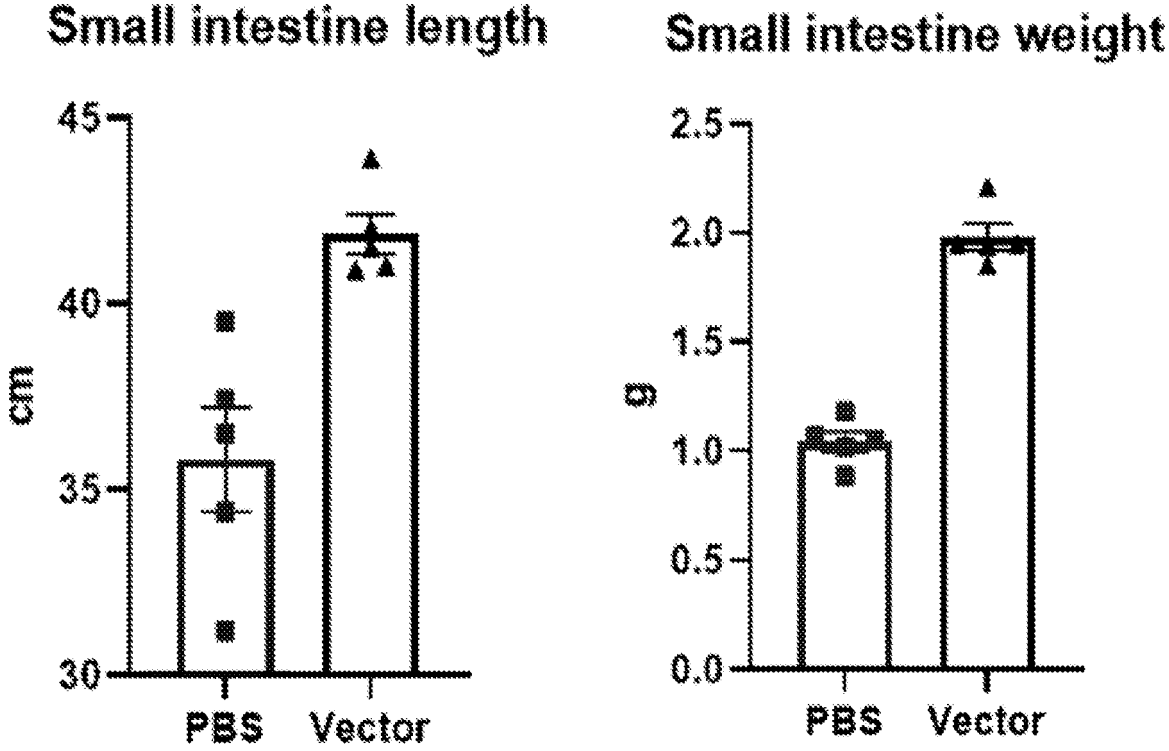


FIG. 5B **D56 post necropsy**



$p < 0.05$ $p < 0.001$

FIG. 5C

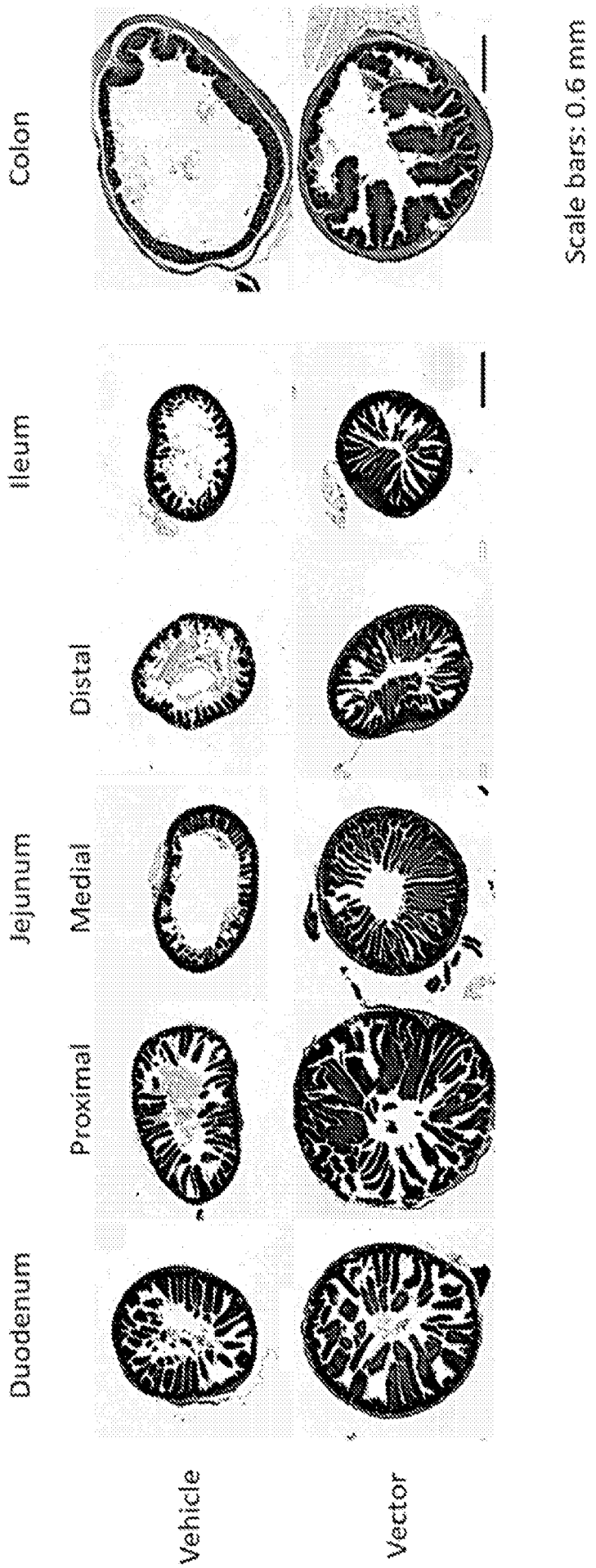
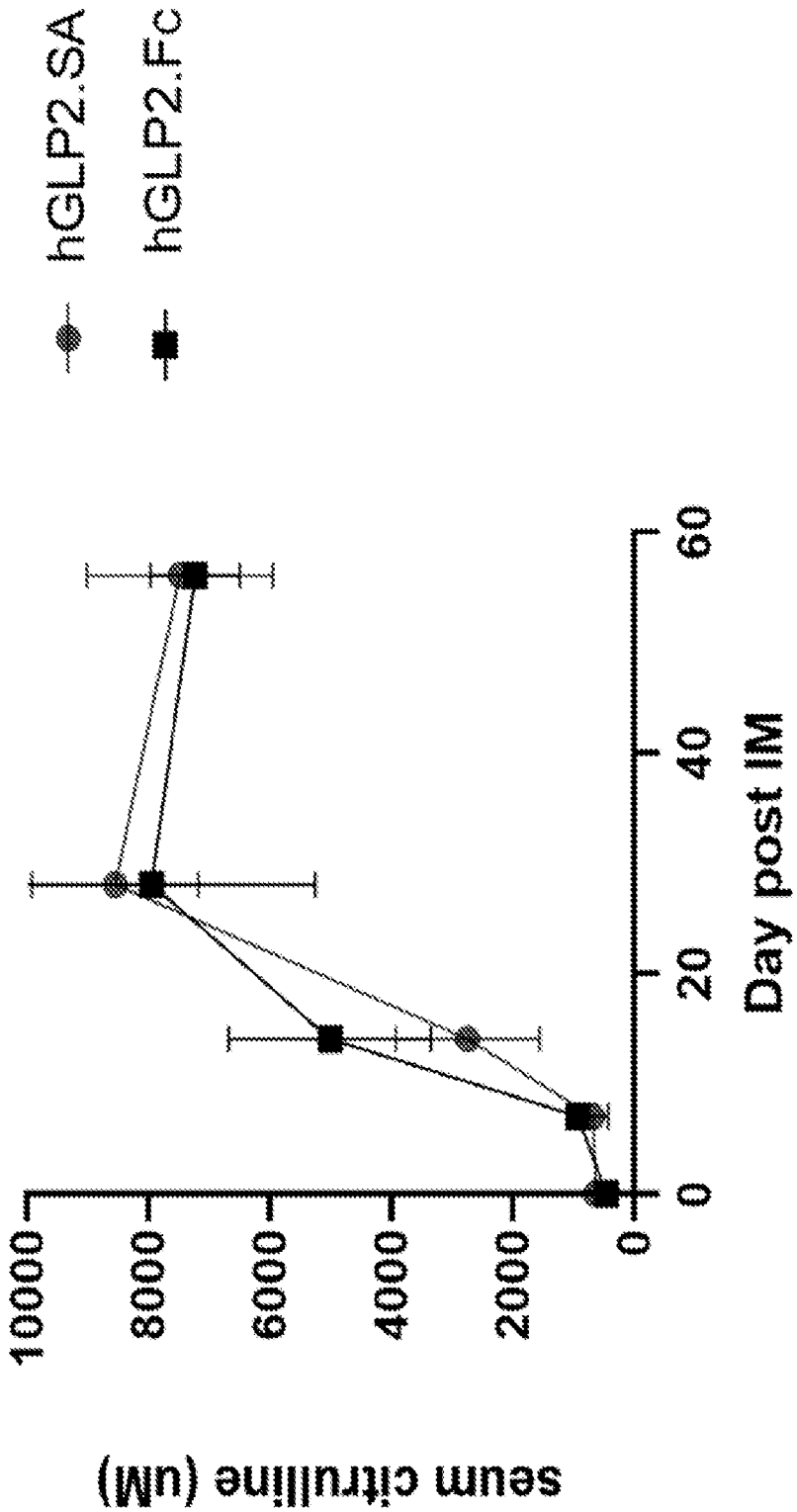

Intestine histology (H&E)
Duodenum
Jejunum
Ileum
Colon
Proximal
Medial
Distal
Vehicle
Vector
Scale bars: 0.6 mm

FIG. 5D


hGLP2.SA
hGLP2.Fc
serum citrulline (uM)
10000
8000
6000
4000
2000
0
0
20
40
60
Day post IM

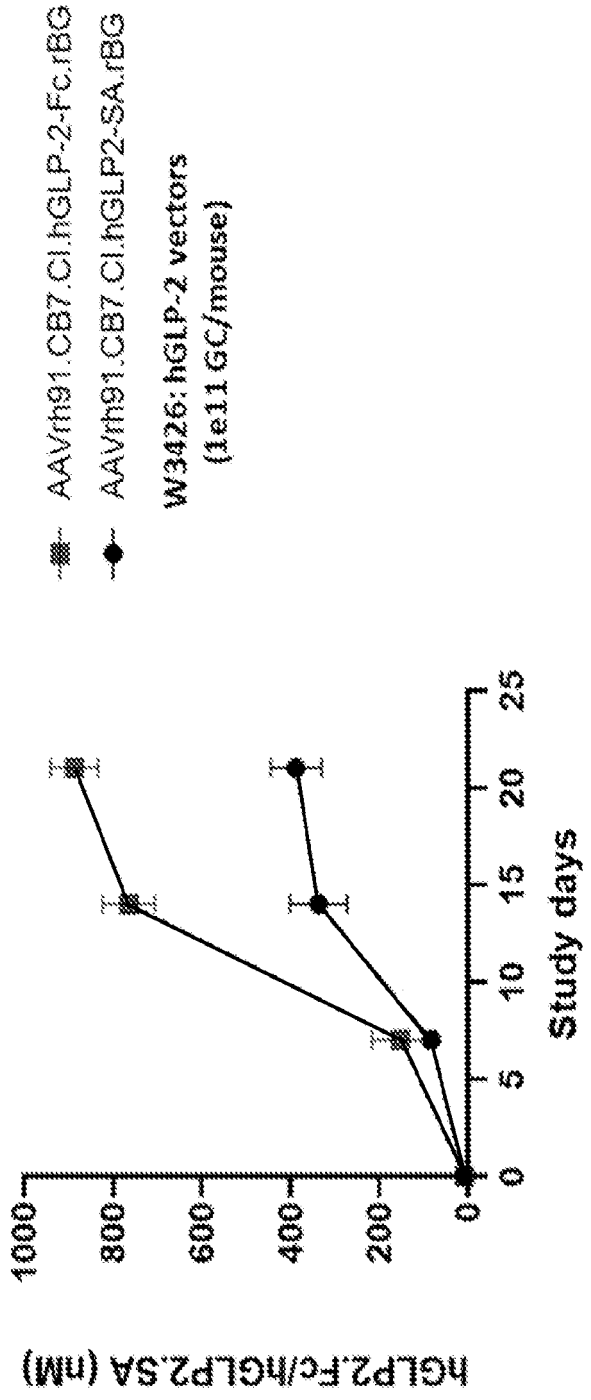

AAVrh91.CB7.CI.hGLP-2-Fc.rBG
AAVrh91.CB7.CI.hGLP2-SA.rBG
W3426: hGLP-2 vectors
(1e11 GC/mouse)
hGLP2.Fc/hGLP2.SA (nM)
1000
800
600
400
200
0
0
5
10
15
20
25
Study days

FIG. 6A

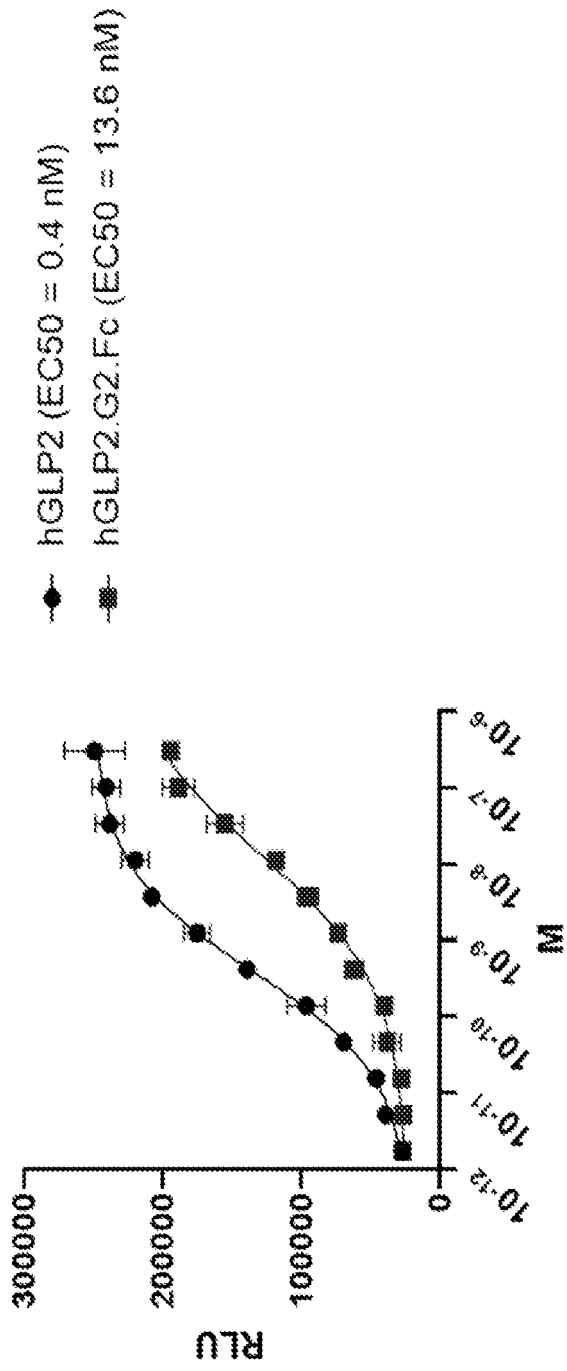

hGLP2 (EC50 = 0.4 nM)
hGLP2.G2.Fc (EC50 = 13.6 nM)
RLU
300000
200000
100000
0
10-12
10-11
10-10
10-9
10-8
10-7
10-6
M

FIG. 6B

FIG. 7

NHP study design

NHP	1	2
Vector	AAVrh91.CB7.CI.hGLP-2-Fc.rBG	
ROA	IM	
Dose (GC/kg)	1.00E+12	1.00E+10
Bleeds	Days 0, 7, 14, 21, 28, 60	
Readout	TG expression & potency, ADA, biomarker (citrulline*)	
Necropsy	Day 60 (intestine histology**)	

*: Plasma/serum citrulline correlates with enterocyte mass
**: H&E staining for villi length (NHPs from GLP-1-Fc studies will be used as control)

FIG. 8A

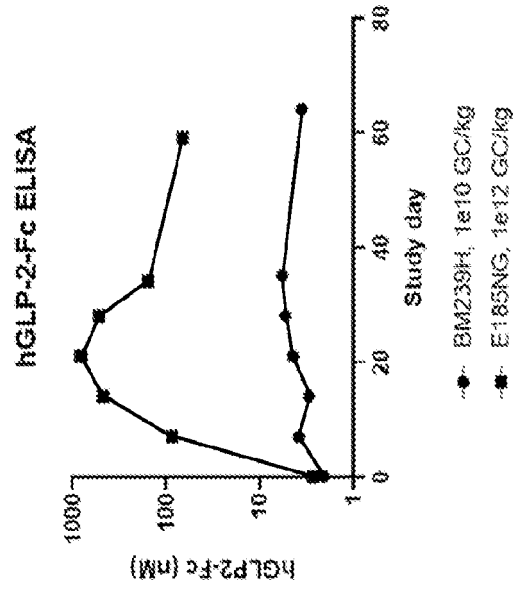


FIG. 8B

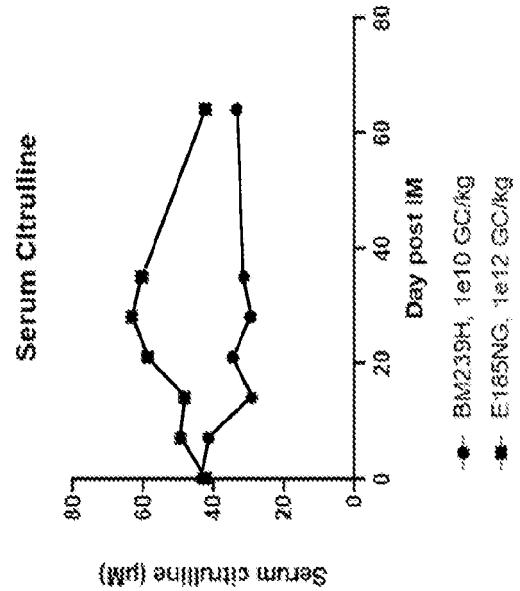


FIG. 8C

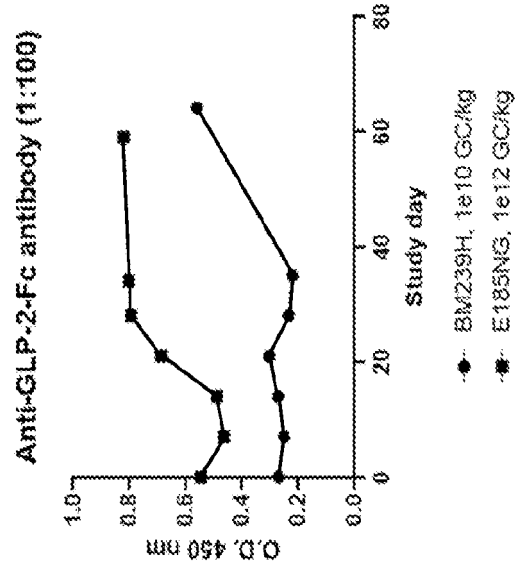


FIG. 9A

Group	Dose (GC/animal)	Necropsy days			
		1 mo	3 mo	6 mo	12 mo
1	0	5	5	10	10
2	1e10	5	5	10	10
3	3e10	5	5	10	10
4	1e11	5	5	10	10

FIG. 9C

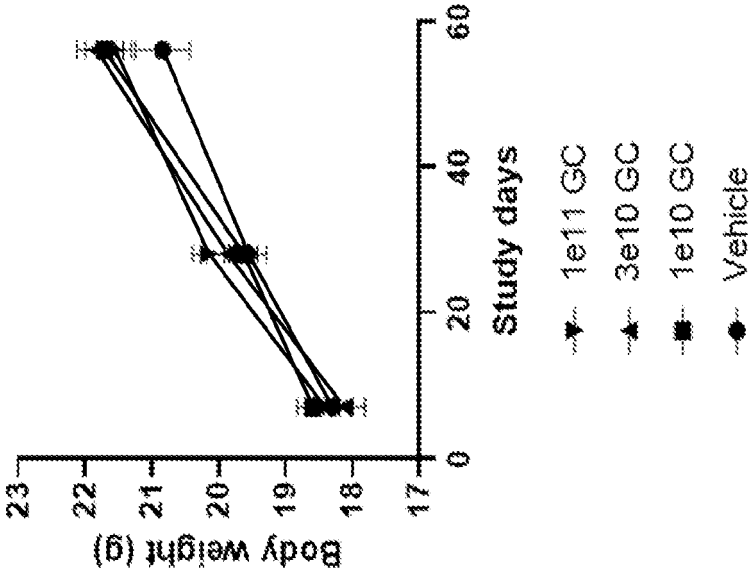

BW TC
Body weight (g)
23
22
21
20
19
18
17
0
20
40
60
Study days
1e11 GC
3e10 GC
1e10 GC
Vehicle

FIG. 9B

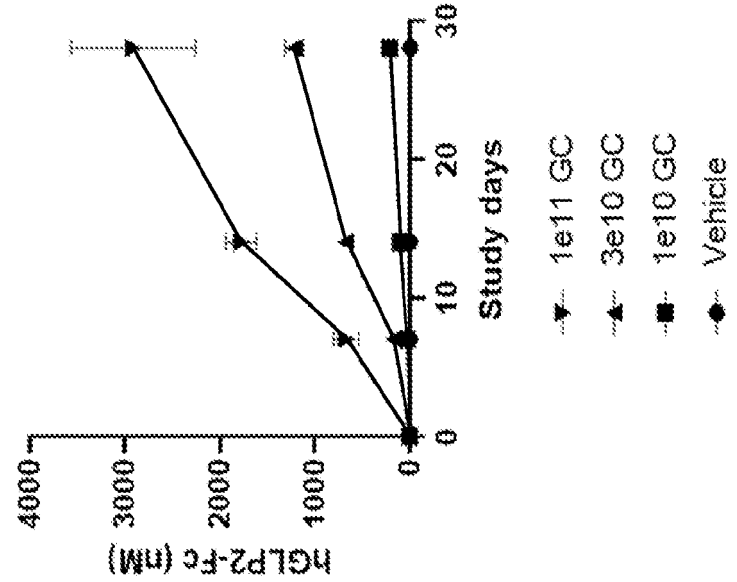

hGLP-2-Fc ELISA
hGLP2-Fc (nM)
4000
3000
2000
1000
0
0
10
20
30
Study days
1e11 GC
3e10 GC
1e10 GC
Vehicle

FIG. 9D

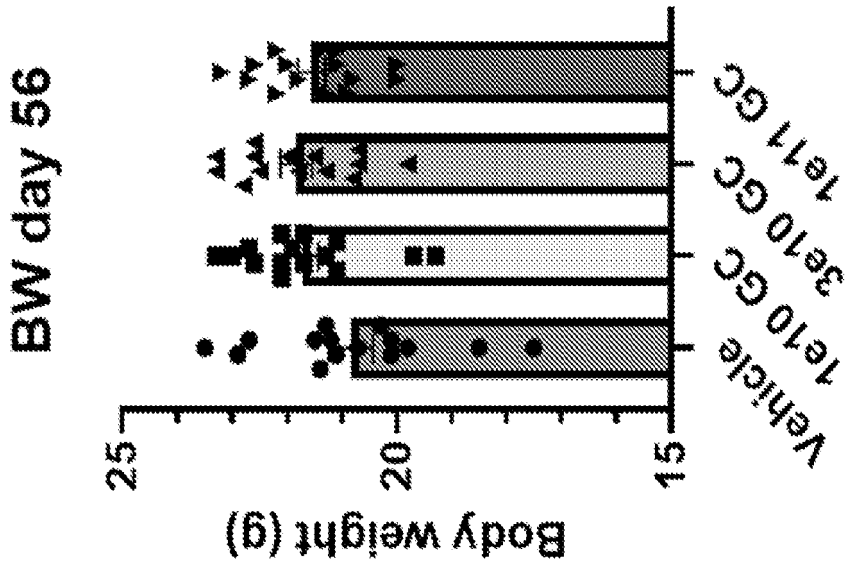


FIG. 9E

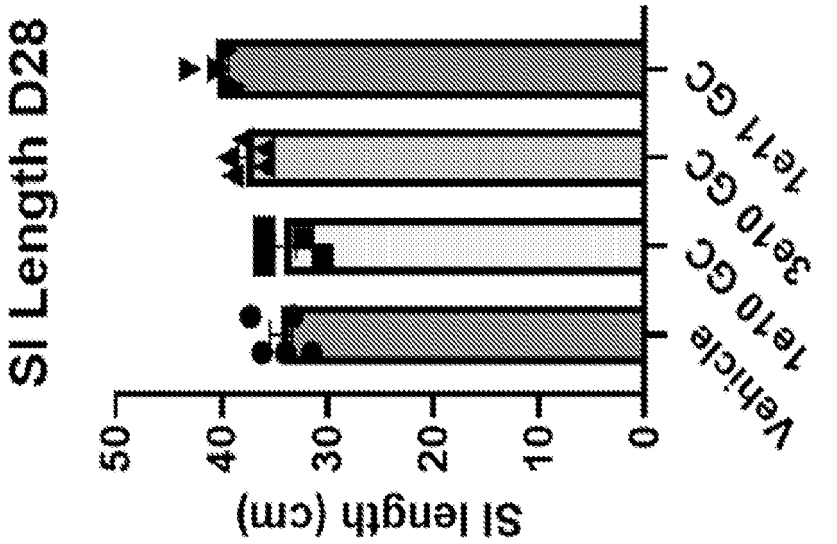
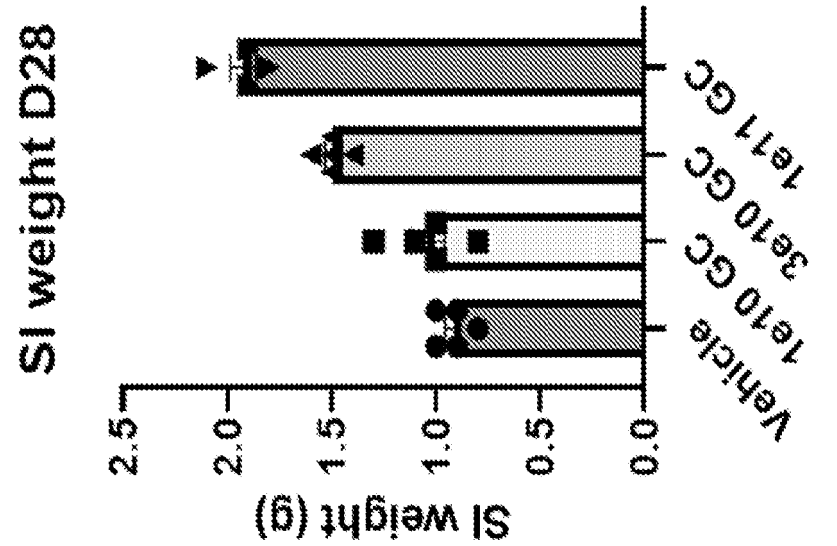


FIG. 9F



Sequence Listing

1	**Sequence Listing Information**	
1-1	File Name	22-10018PCT_Seq-Listing.xml
1-2	DTD Version	V1_3
1-3	Software Name	WIPO Sequence
1-4	Software Version	2.2.0
1-5	Production Date	2023-03-03
1-6	Original free text language code	
1-7	Non English free text language code	
2	**General Information**	
2-1	Current application: IP Office	US
2-2	Current application: Application number	
2-3	Current application: Filing date	
2-4	Current application: Applicant file reference	22-10018.PCT
2-5	Earliest priority application: IP Office	US
2-6	Earliest priority application: Application number	63/316,219
2-7	Earliest priority application: Filing date	2022-03-03
2-8en	Applicant name	The Trustees of the University of Pennsylvania
2-8	Applicant name: Name Latin	
2-9en	Inventor name	
2-9	Inventor name: Name Latin	
2-10en	Invention title	VIRAL VECTORS ENCODING GLP-2 RECEPTOR AGONIST FUSIONS AND USES THEREOF IN TREATING SHORT BOWEL SYNDROME
2-11	Sequence Total Quantity	55

3-1	**Sequences**		
3-1-1	Sequence Number [ID]	1	
3-1-2	Molecule Type	AA	
3-1-3	Length	33	
3-1-4	Features Location/	**source 1..33**	
	Qualifiers	mol_type=protein	
		organism=Homo sapiens	
	NonEnglishQualifier Value		
3-1-5	Residues	HADGSFSDEM NTILDNLAAR DFINWLIQTK ITD	33
3-2	**Sequences**		
3-2-1	Sequence Number [ID]	2	
3-2-2	Molecule Type	AA	
3-2-3	Length	33	
3-2-4	Features Location/	**source 1..33**	
	Qualifiers	mol_type=protein	
		organism=Homo sapiens	
	NonEnglishQualifier Value		
3-2-5	Residues	HGDGSFSDEM NTILDNLAAR DFINWLIQTK ITD	33
3-3	**Sequences**		
3-3-1	Sequence Number [ID]	3	
3-3-2	Molecule Type	AA	
3-3-3	Length	39	
3-3-4	Features Location/	**REGION 1..39**	
	Qualifiers	note=synthetic construct	
		source 1..39	
		mol_type=protein	
		organism=synthetic construct	
	NonEnglishQualifier Value		
3-3-5	Residues	HGEGTFSSEL ATILDALAAR DFIAWLIATK ITDKKKKKK	39
3-4	**Sequences**		
3-4-1	Sequence Number [ID]	4	
3-4-2	Molecule Type	AA	
3-4-3	Length	30	
3-4-4	Features Location/	**source 1..30**	
	Qualifiers	mol_type=protein	
		organism=Homo sapiens	
	NonEnglishQualifier Value		
3-4-5	Residues	HAEGTFTSDV SSYLEGQAAK EFIAWLVKGR	30
3-5	**Sequences**		
3-5-1	Sequence Number [ID]	5	
3-5-2	Molecule Type	AA	
3-5-3	Length	33	
3-5-4	Features Location/	**source 1..33**	
	Qualifiers	mol_type=protein	
		organism=Macaca fascicularis	
	NonEnglishQualifier Value		
3-5-5	Residues	HADGSFSDEM NTVLDNLATR DFINWLIQTK ITD	33
3-6	**Sequences**		
3-6-1	Sequence Number [ID]	6	
3-6-2	Molecule Type	AA	
3-6-3	Length	33	
3-6-4	Features Location/	**source 1..33**	
	Qualifiers	mol_type=protein	
		organism=Mus musculus	
	NonEnglishQualifier Value		
3-6-5	Residues	HADGSFSDEM STILDNLATR DFINWLIQTK ITD	33
3-7	**Sequences**		
3-7-1	Sequence Number [ID]	7	
3-7-2	Molecule Type	AA	
3-7-3	Length	43	
3-7-4	Features Location/	**source 1..43**	
	Qualifiers	mol_type=protein	
		organism=Homo sapiens	
	NonEnglishQualifier Value		
3-7-5	Residues	MAHVRGLQLP GCLALAALCS LVHSQHVFLA PQQARSLLQR VRR	43

3-8	**Sequences**	
3-8-1	Sequence Number [ID]	8
3-8-2	Molecule Type	AA
3-8-3	Length	229
3-8-4	Features Location/	**source 1..229**
	Qualifiers	mol_type=protein
		organism=Homo sapiens
	NonEnglishQualifier Value	
3-8-5	Residues	AESKYGPPCP PCPAPEAAGG PSVFLFPPKP KDTLMISRTP EVTCVVVDVS QEDPEVQFNW 60 YVDGVEVHNA KTKPREEQFN STYRVVSVLT VLHQDWLNGK EYKCKVSNKG LPSSIEKTIS 120 KAKGQPREPQ VYTLPPSQEE MTKNQVSLTC LVKGFYPSDI AVEWESNGQP ENNYKTTPPV 180 LDSDGSFFLY SRLTVDKSRW QEGNVFSCSV MHEALHNHYT QKSLSLSLG 229
3-9	**Sequences**	
3-9-1	Sequence Number [ID]	9
3-9-2	Molecule Type	AA
3-9-3	Length	224
3-9-4	Features Location/	**source 1..224**
	Qualifiers	mol_type=protein
		organism=Macaca mulatta
	NonEnglishQualifier Value	
3-9-5	Residues	PPCPPCPAPE LLGGPSVFLF PPKPKDTLMI SRTPEVTCVV VDVSQEDPEV QFNWYVDGVE 60 VHNAQTKPRE RQFNSTYRVV SVLTVTHQDW LNGKEYTCKV SNKGLPAPIE KTISKAKGQP 120 REPQVYILPP PQEELTKNQV SLTCLVTGFY PSDIAVEWES NGQPENTYKT TPPVLDSDGS 180 YLLYSKLTVN KSRWQPGNIF TCSVMHEALH NHYTQKSLSV SPGK 224
3-10	**Sequences**	
3-10-1	Sequence Number [ID]	10
3-10-2	Molecule Type	AA
3-10-3	Length	584
3-10-4	Features Location/	**source 1..584**
	Qualifiers	mol_type=protein
		organism=Homo sapiens
	NonEnglishQualifier Value	
3-10-5	Residues	DAHKSEVAHR FKDLGEENFK ALVLIAFAQY LQQCPFEDHV KLVNEVTEFA KTCVADESAE 60 NCDKSLHTLF GDKLCTVATL RETYGEMADC CAKQEPERNE CFLQHKDDNP NLPRLVRPEV 120 DVMCTAFHDN EETFLKKYLY EIARRHPYFY APELLFFAKR YKAAFTECCQ AADKAACLLP 180 KLDELRDEGK ASSAKQRLKC ASLQKFGERA FKAWAVARLS QRFPKAEFAE VSKLVTDLTK 240 VHTECCHGDL LECADDRADL AKYICENQDS ISSKLKECCE KPLLEKSHCI AEVENDEMPA 300 DLPSLAADFV ESKDVCKNYA EAKDVFLGMF LYEYARRHPD YSVVLLLRLA KTYETTLEKC 360 CAAADPHECY AKVFDEFKPL VEEPQNLIKQ NCELFEQLGE YKFQNALLVR YTKKVPQVST 420 PTLVEVSRNL GKVGSKCCKH PEAKRMPCAE DYLSVVLNQL CVLHEKTPVS DRVTKCCTES 480 LVNRRPCFSA LEVDETYVPK EFNAETFTFH ADICTLSEKE RQIKKQTALV ELVKHKPKAT 540 KEQLKAVMDD FAAFVEKCCK ADDKETCFAE EGKKLVAASQ AALG 584
3-11	**Sequences**	
3-11-1	Sequence Number [ID]	11
3-11-2	Molecule Type	AA
3-11-3	Length	584
3-11-4	Features Location/	**source 1..584**
	Qualifiers	mol_type=protein
		organism=Macaca mulatta
	NonEnglishQualifier Value	
3-11-5	Residues	DTHKSEVAHR FKDLGEEHFK GLVLVAFSQY LQQCPFEEHV KLVNEVTEFA KTCVADESAE 60 NCDKSLHTLF GDKLCTVATL RETYGEMADC CAKQEPERNE CFLQHKDDNP NLPPLVRPEV 120 DVMCTAFHDN EATFLKKYLY EVARRHPYFY APELLFFAAR YKAAFAECCQ AADKAACLLP 180 KLDELRDEGK ASSAKQRLKC ASLQKFGDRA FKAWAVARLS QKFPKAEFAE VSKLVTDLTK 240 VHTECCHGDL LECADDRADL AKYMCENQDS ISSKLKECCD KPLLEKSHCL AEVENDEMPA 300 DLPSLAADYV ESKDVCKNYA EAKDVFLGMF LYEYARRHPD YSVMLLLRLA KAYEATLEKC 360 CAAADPHECY AKVFDEFQPL VEEPQNLVKQ NCELFEQLGE YKFQNALLVR YTKKVPQVST 420 PTLVEVSRNL GKVGAKCCKL PEAKRMPCAE DYLSVVLNRL CVLHEKTPVS EKVTKCCTES 480 LVNRRPCFSA LELDEAYVPK AFNAETFTFH ADMCTLSEKE KQVKKQTALV ELVKHKPKAT 540 KEQLKGVMDN FAAFVEKCCK ADDKEACFAE EGPKFVAASQ AALA 584
3-12	**Sequences**	
3-12-1	Sequence Number [ID]	12
3-12-2	Molecule Type	AA
3-12-3	Length	15

3-12-4	Features Location/	**REGION 1..15**
	Qualifiers	note=synthetic construct
		source 1..15
		mol_type=protein
		organism=synthetic construct
	NonEnglishQualifier Value	
3-12-5	Residues	GGGGSGGGGS GGGGS 15
3-13	**Sequences**	
3-13-1	Sequence Number [ID]	13
3-13-2	Molecule Type	AA
3-13-3	Length	322
3-13-4	Features Location/	**REGION 1..322**
	Qualifiers	note=synthetic construct
		source 1..322
		mol_type=protein
		organism=synthetic construct
	NonEnglishQualifier Value	
3-13-5	Residues	MAHVRGLQLP GCLALAALCS LVHSQHVFLA PQQARSLLQR VRRHGDGSFS DEMNTILDNL 60
		AARDFINWLI QTKITDGGGG GGSGGGGSGG GGSAESKYGP PCPPCPAPEA AGGPSVFLFP 120
		PKPKDTLMIS RTPEVTCVVV DVSQEDPEVQ FNWYVDGVEV HNAKTKPREE QFNSTYRVVS 180
		VLTVLHQDWL NGKEYKCKVS NKGLPSSIEK TISKAKGQPR EPQVYTLPPS QEEMTKNQVS 240
		LTCLVKGFYP SDIAVEWESN GQPENNYKTT PPVLDSDGSF FLYSRLTVDK SRWQEGNVFS 300
		CSVMHEALHN HYTQKSLSLS LG 322
3-14	**Sequences**	
3-14-1	Sequence Number [ID]	14
3-14-2	Molecule Type	DNA
3-14-3	Length	969
3-14-4	Features Location/	**misc_feature 1..969**
	Qualifiers	note=synthetic construct
		source 1..969
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-14-5	Residues	atggctcacg ttcggggact gcagctgcct ggatgtctgg ctcttgccgc tctgtgtagc 60
		ctggtgcaca gccagcacgt gtttctggct cctcagcaag ccagatcact gctgcagaga 120
		gttagaaggc acggcgacgg cagcttcagc gacgagatga acaccatcct ggacaacctg 180
		gccgccagag acttcatcaa ctggctgatc cagaccaaga tcaccgacgg tggcggaggc 240
		ggaggatctg gtggtggtgg atctggcggc ggaggttctg ccgagtctaa gtacggacct 300
		ccttgtcctc cctgtcctgc tccagaagct gctggcggcc catccgtgtt tctgttccct 360
		ccaaagccta aggacaccct gatgatcagc agaacccctg aagtgacctg cgtggtggtc 420
		gacgtgtccc aagaggatcc tgaggtgcag ttcaattggt acgtggacgg cgtggaagtg 480
		cacaacgcca agaccaagcc tagagaggaa cagttcaaca gcacctacag agtggtgtcc 540
		gtgctgaccg tgctgcacca ggattggctg aacggcaaag agtacaagtg caaggtgtcc 600
		aacaagggcc tgcctagcag catcgagaaa accatcagca aggccaaggg ccagccaaga 660
		gaaccccagg tgtacacact gcctccaagc caagaggaaa tgaccaagaa ccaggtgtcc 720
		ctgacctgcc tggtcaaggg cttctaccct tccgatatcg ccgtggaatg ggagagcaac 780
		ggccagcctg agaacaacta caagaccaca cctcctgtgc tggactccga tggctcattc 840
		ttcctgtaca gcagactgac cgtggacaag agcaggtggc aagagggcaa cgtgttcagc 900
		tgcagcgtga tgcacgaggc cctgcacaac cactacaccc agaaaagcct gagcctgtct 960
		ctgggctaa 969
3-15	**Sequences**	
3-15-1	Sequence Number [ID]	15
3-15-2	Molecule Type	AA
3-15-3	Length	677
3-15-4	Features Location/	**REGION 1..677**
	Qualifiers	note=synthetic construct
		source 1..677
		mol_type=protein
		organism=synthetic construct
	NonEnglishQualifier Value	
3-15-5	Residues	MAHVRGLQLP GCLALAALCS LVHSQHVFLA PQQARSLLQR VRRHGDGSFS DEMNTILDNL 60
		AARDFINWLI QTKITDGGGG GGSGGGGSGG GGSDAHKSEV AHRFKDLGEE NFKALVLIAF 120
		AQYLQQCPFE DHVKLVNEVT EFAKTCVADE SAENCDKSLH TLFGDKLCTV ATLRETYGEM 180
		ADCCAKQEPE RNECFLQHKD DNPNLPRLVR PEVDVMCTAF HDNEETFLKK YLYEIARRHP 240
		YFYAPELLFF AKRYKAAFTE CCQAADKAAC LLPKLDELRD EGKASSAKQR LKCASLQKFG 300

		ERAFKAWAVA RLSQRFPKAE FAEVSKLVTD LTKVHTECCH GDLLECADDR ADLAKYICEN 360 QDSISSKLKE CCEKPLLEKS HCIAEVENDE MPADLPSLAA DFVESKDVCK NYAEAKDVFL 420 GMFLYEYARR HPDYSVVLLL RLAKTYETTL EKCCAAADPH ECYAKVFDEF KPLVEEPQNL 480 IKQNCELFEQ LGEYKFQNAL LVRYTKKVPQ VSTPTLVEVS RNLGKVGSKC CKHPEAKRMP 540 CAEDYLSVVL NQLCVLHEKT PVSDRVTKCC TESLVNRRPC FSALEVDETY VPKEFNAETF 600 TFHADICTLS EKERQIKKQT ALVELVKHKP KATKEQLKAV MDDFAAFVEK CCKADDKETC 660 FAEEGKKLVA ASQAALG 677
3-16	**Sequences**	
3-16-1	Sequence Number [ID]	16
3-16-2	Molecule Type	DNA
3-16-3	Length	2034
3-16-4	Features Location/	**misc_feature 1..2034**
	Qualifiers	note=synthetic construct
		source 1..2034
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-16-5	Residues	atggctcacg ttcggggact gcagctgcct ggatgtctgg ctcttgccgc tctgtgtagc 60 ctggtgcaca gccagcacgt gtttctggct cctcagcaag ccagatcact gctgcagaga 120 gttagaaggc acggcgacgg cagcttcagc gacgagatga acaccatcct ggacaacctg 180 gccgccagag acttcatcaa ctggctgatc cagaccaaga tcaccgacgg tggcggaggc 240 ggaggatctg gtggtggtgg atctggcggc ggaggttctg acgcccacaa atctgaagtg 300 gcccaccggt tcaaggacct gggcgaagag aatttcaagg ccctggtgct gatcgccttc 360 gctcagtacc tgcagcagtg ccccttcgag gaccacgtga agctggtcaa cgaagtgacc 420 gagttcgcca agacctgcgt ggccgacgag agcgccgaga actgtgataa gagcctgcac 480 accctgttcg gcgacaagct gtgtacagtg gccacactga gagaaaccta cggcgagatg 540 gccgactgct gcgccaagca agagcccgag agaaacgagt gcttcctgca gcacaaggac 600 gacaacccca acctgcctag actcgtgcgg cctgaagtgg acgtgatgtg caccgccttc 660 cacgacaacg aggaaacctt cctgaagaag tacctgtacg agatcgccag acggcacccc 720 tacttttacg cccctgagct gctgttcttc gccaagcggt ataaggccgc cttcaccgag 780 tgttgtcagg ccgctgataa ggctgcctgc ctgctgccta agctggacga gcttagagac 840 gagggcaaag ccagctccgc caagcagaga ctgaagtgtg ccagcctgca gaagttcggc 900 gagagagcct ttaaggcctg ggccgttgct agactgagcc agagatttcc caaggccgag 960 tttgccgagg tgtccaagct cgtgaccgac ctgacaaagg tgcacaccga gtgctgccac 1020 ggcgacctgc tggaatgcgc cgacgataga gccgacctgg ccaagtacat ctgcgagaac 1080 caggacagca tcagcagcaa gctgaaagag tgctgcgaga agcctctgct ggaaaagagc 1140 cactgtatcg ccgaggtgga aaacgacgag atgcccgccg atctgccttc tctggccgcc 1200 gattttgtgg aaagcaagga cgtgtgcaag aactacgccg aggccaagga cgtgttcctg 1260 ggcatgtttc tgtacgagta cgcccgcaga caccccgact actctgttgt gctgctgctg 1320 agactggcca aaacctacga gacaaccctg gaaaagtgct gtgccgccgc tgatcctcac 1380 gagtgttacg ccaaggtgtt cgacgagttc aagccactgg tggaagaacc ccagaacctg 1440 atcaagcaga actgcgagct gttcgagcag ctgggcgagt acaagttcca gaacgccctg 1500 ctcgtgcggt acaccaagaa ggtgccccag gtttccacac ctacactggt tgaggtgtcc 1560 cggaacctgg gcaaagtggg cagcaagtgt tgcaagcacc ctgaggccaa gagaatgccc 1620 tgcgccgagg attacctgag cgtggtgctg aatcagctgt gcgtgctgca cgagaaaacc 1680 cctgtgtccg acagagtgac caagtgctgt accgagagcc tggtcaacag acggccttgc 1740 tttagcgccc tcgaggtgga cgagacatac gtgcccaaag agttcaacgc cgagacattc 1800 accttccacg ccgacatctg taccctgagc gagaaagagc ggcagatcaa gaaacagact 1860 gccctggtgg aactggtcaa gcacaagccc aaggccacca aagaacagct gaaggccgtg 1920 atggacgact tcgccgcctt cgtggaaaag tgctgcaagg ccgacgacaa agagacctgc 1980 ttcgccgaag agggcaagaa actggtggcc gcttctcagg ctgctctggg ataa 2034
3-17	**Sequences**	
3-17-1	Sequence Number [ID]	17
3-17-2	Molecule Type	AA
3-17-3	Length	957
3-17-4	Features Location/	**REGION 1..957**
	Qualifiers	note=synthetic construct
		source 1..957
		mol_type=protein
		organism=synthetic construct
	NonEnglishQualifier Value	
3-17-5	Residues	MAHVRGLQLP GCLALAALCS LVHSQHVFLA PQQARSLLQR VRRHGDGSFS DEMNTILDNL 60 AARDFINWLI QTKITDGGGG GGSGGGGSGG GGSSPAGSPT STEEGTSESA TPESGPGTST 120 EPSEGSAPGS PAGSPTSTEE GTSTEPSEGS APGTSTEPSE GSAPGTSESA TPESGPGSEP 180 ATSGSETPGS EPATSGSETP GSPAGSPTST EEGTSESATP ESGPGTSTEP SEGSAPGTST 240

		EPSEGSAPGS PAGSPTSTEE GTSTEPSEGS APGTSTEPSE GSAPGTSESA TPESGPGTST 300 EPSEGSAPGT SESATPESGP GSEPATSGSE TPGTSTEPSE GSAPGTSTEP SEGSAPGTSE 360 SATPESGPGT SESATPESGP GSPAGSPTST EEGTSESATP ESGPGSEPAT SGSETPGTSE 420 SATPESGPGT STEPSEGSAP GTSTEPSEGS APGTSTEPSE GSAPGTSTEP SEGSAPGTST 480 EPSEGSAPGT STEPSEGSAP GSPAGSPTST EEGTSTEPSE GSAPGTSESA TPESGPGSEP 540 ATSGSETPGT SESATPESGP GSEPATSGSE TPGTSESATP ESGPGTSTEP SEGSAPGTSE 600 SATPESGPGS PAGSPTSTEE GSPAGSPTST EEGSPAGSPT STEEGTSESA TPESGPGTST 660 EPSEGSAPGT SESATPESGP GSEPATSGSE TPGTSESATP ESGPGSEPAT SGSETPGTSE 720 SATPESGPGT STEPSEGSAP GSPAGSPTST EEGTSESATP ESGPGSEPAT SGSETPGTSE 780 SATPESGPGS PAGSPTSTEE GSPAGSPTST EEGTSTEPSE GSAPGTSESA TPESGPGTSE 840 SATPESGPGT SESATPESGP GSEPATSGSE TPGSEPATSG SETPGSPAGS PTSTEEGTST 900 EPSEGSAPGT STEPSEGSAP GSEPATSGSE TPGTSESATP ESGPGTSTEP SEGSAPG 957
3-18	**Sequences**	
3-18-1	Sequence Number [ID]	18
3-18-2	Molecule Type	DNA
3-18-3	Length	2874
3-18-4	Features Location/	**misc_feature 1..2874**
	Qualifiers	note=synthetic construct
		source 1..2874
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-18-5	Residues	atggctcacg ttcggggact gcagctgcct ggatgtctgg ctcttgccgc tctgtgtagc 60 ctggtgcaca gccagcacgt gtttctggct cctcagcaag ccagatcact gctgcagaga 120 gttagaaggc acggcgacgg cagcttcagc gacgagatga acaccatcct ggacaacctg 180 gccgccagag acttcatcaa ctggctgatc cagaccaaga tcaccgacgg tggcggaggc 240 ggaggatctg gtggtggtgg atctggcggc ggaggaagtt ctcctgctgg cagccctaca 300 agcaccgagg aaggcacaag cgagtctgcc acacctgagt ctggccctgg cacatctaca 360 gagcctagcg aaggatctgc cccaggatct cctgccggct ctccaacatc taccgaagag 420 ggaaccagca ccgagccatc tgagggatct gctcccggaa caagcacaga gccttcagaa 480 ggatccgctc ctggcacctc tgaaagcgcc acaccagaaa gcggacctgg atctgagcct 540 gccacaagcg gatctgagac acctggaagc gagccagcca catctggcag cgaaacacct 600 ggttctccag ccggatctcc caccagcaca gaagagggca catccgaatc tgctaccct 660 gaatctggac caggcacctc cacagaacct tccgagggtt ctgcccctgg aacctctacc 720 gaaccatcag aaggcagcgc tccaggttca ccagctggaa gcccaacctc tacagaggaa 780 gggacatcca ctgagccaag cgagggaagc gctcccggca ctagtacaga accgagcgag 840 ggcagtgctc ctggaaccag cgaatccgct actccagaga gtggcccagg caccagtact 900 gaaccctctg agggtagcgc acccggaaca tctgagagcg ctactcccga atcaggccca 960 ggctctgaac ctgctaccag cggaagtgaa acacccggca cctctactga gccctccgaa 1020 ggctcagcac ctggcacaag cactgaacca tcagagggct ccgcaccagg caccagcgaa 1080 agtgctacac cagagtcagg acccggaacc tccgaaagtg caactcctga gagcggacca 1140 ggctctcccg ctggatctcc tacatcaact gaagaaggga cctccgagag cgcaacccca 1200 gagtctggtc caggatcaga acctgccacc tccggctctg aaaccccagg cacttctgag 1260 tccgccacgc cagaatctgg tcctgggact agcaccgaac cgagtgaagg ttcagctccc 1320 gggacttcta cggaacccag tgaaggatct gcacccggca catcaaccga accgtcagag 1380 ggatcagccc ctgggacttc cacagagccg tctgagggca gcgccccagg gacgtctaca 1440 gaaccatctg aaggatcagc accagggacc tctaccgagc caagtgaagg cagtgcaccg 1500 ggaagtccag caggctcccc tacaagtact gaagagggta ctagcacgga acccagcgag 1560 ggttccgctc cagggacatc tgaatccgca actccggaat ccggacctgg cagtgaacca 1620 gctacatccg gatccgagac tccgggaacc tcagaatcag ctacacccga gagtggacct 1680 ggctccgaac cagcaactag cggctcagaa actcctggga caagcgagag tgcaacaccc 1740 gaatctggac ctggaacaag tactgagccc agcgaaggca gcgcccctgg aacttctgaa 1800 tctgccactc ctgaaagtgg ccctggaagc cctgcaggct cacccacatc cacagaagaa 1860 ggatcaccag caggcagccc cacttcaacg gaagagggat ccccagctgg atccccaact 1920 agtacggaag aaggcacttc agaaagcgct acgcccgagt ccggtcctgg cacttctact 1980 gaaccatccg agggaagtgc ccctggcact tccgagagtg ctacacctga aagcggtccc 2040 ggctctgaac cagccacttc tggatctgaa acgcccggga catccgagtc agcaacgccc 2100 gaaagcggcc caggttccga gccggctact agtggttcag agactccagg gacttccgag 2160 tctgctactc ctgagtccgg accgggaaca tcaaccgagc cttccgaagg atctgcacct 2220 ggaagccctg ccggatctcc taccagtact gaggaaggca cctcagagtc tgccactcca 2280 gagtcaggtc ctggaagcga acctgcaaca agcggcagcg aaactccagg cactagcgag 2340 tcagctaccc cagaatcagg acctggatct ccagcagggt ccccaacatc tactgaggaa 2400 ggctctcctg ctggctcccc tacctctacc gaagagggga cctcaacaga gccatccgag 2460 gggagcgcac ctggtacatc agagtccgca actcccgagt ctggccccgg aactagcgaa 2520 tctgcaaccc cggaaagtgg acccgggacg agtgaatcag ccacacctga atccggtcca 2580

		ggatccgagc ctgcaacttc tggaagcgag acaccaggat ctgagccagc tacgtctggc 2640 tctgagactc ctggatctcc tgctggtagt cccacctcca ctgaagaggg aacttccacc 2700 gaaccgagcg agggatcagc accaggcact agcacagaac cgtccgaagg atctgctcca 2760 ggctctgaac ccgcaacctc cggatcagaa acccctggaa catccgaaag cgctacaccg 2820 gaaagtggcc ccggaacctc tacagaacct agcgagggaa gcgcaccagg ataa 2874
3-19	**Sequences**	
3-19-1	Sequence Number [ID]	19
3-19-2	Molecule Type	AA
3-19-3	Length	322
3-19-4	Features Location/	**REGION 1..322**
	Qualifiers	note=synthetic construct
		source 1..322
		mol_type=protein
		organism=synthetic construct
	NonEnglishQualifier Value	
3-19-5	Residues	MAHVRGLQLP GCLALAALCS LVHSQHVFLA PQQALSLLQR VRRHGDGSFS DEMNTVLVDN 60 LATRDFINWL IQTKITDGGG GGGGSGGGGS GGGGSAEFTP PCPPCPAPEL LGGPSVFLFP 120 PKPKDTLMIS RTPEVTCVVV DVSQEDPEVQ FNWYVDGVEV HNAQTKPRER QFNSTYRVVS 180 VLTVTHQDWL NGKEYTCKVS NKGLPAPIEK TISKAKGQPR EPQVYILPPP QEELTKNQVS 240 LTCLVTGFYP SDIAVEWESN GQPENTYKTT PPVLDSDGSY LLYSKLTVNK SRWQPGNIFT 300 CSVMHEALHN HYTQKSLSVS PG 322
3-20	**Sequences**	
3-20-1	Sequence Number [ID]	20
3-20-2	Molecule Type	DNA
3-20-3	Length	969
3-20-4	Features Location/	**misc_feature 1..969**
	Qualifiers	note=synthetic construct
		source 1..969
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-20-5	Residues	atggctcacg ttcggggact gcagctgcct ggatgtctgg ctcttgccgc tctgtgtagc 60 ctggtgcaca gccagcatgt gtttctggct cctcagcagg ccctgagcct gctgcaaaga 120 gttagaaggc acggcgacgg cagcttcagc gacgagatga ataccgtgct ggtggacaac 180 ctggccacca gagacttcat caactggctg atccagacca agatcaccga cggtggtggc 240 ggaggcggag gatctggtgg cggtggttct ggcggtggcg gatctgctga gtttacccct 300 ccttgtcctc cctgtcctgc tccagaactg ctcggcggac cttccgtgtt cctgtttcct 360 ccaaagccta aggacaccct gatgatcagc agaaacccctg aagtgacctg cgtggtggtg 420 gacgtgtccc aagaggatcc tgaggtgcag ttcaattggt acgtggacgg cgtggaagtg 480 cacaacgccc agacaaagcc cagagagcgg cagttcaaca gcacctacag agtggtgtcc 540 gtgctgaccg tgacacacca ggattggctg aacggcaaag agtacacctg taaagtctcc 600 aacaagggcc tgcctgctcc tatcgagaaa accatcagca aggccaaggg ccagcctaga 660 gaaccccagg tgtacatcct gcctccacct caagaggaac tgaccaagaa ccaggtgtcc 720 ctgacctgtc tggtcaccgg cttctaccct tccgatatcg ccgtggagtg ggagagcaac 780 ggacagcccg agaacaccta caagaccaca cctccagtgc tggacagcga cggctcttac 840 ctgctgtact ccaagctgac agtgaacaag agccggtggc agcccggcaa catcttcacc 900 tgttctgtga tgcacgaggc cctgcacaac cactacaccc agaaaagcct gagcgtgtcc 960 cctggataa 969
3-21	**Sequences**	
3-21-1	Sequence Number [ID]	21
3-21-2	Molecule Type	AA
3-21-3	Length	678
3-21-4	Features Location/	**REGION 1..678**
	Qualifiers	note=synthetic construct
		source 1..678
		mol_type=protein
		organism=synthetic construct
	NonEnglishQualifier Value	
3-21-5	Residues	MAHVRGLQLP GCLALAALCS LVHSQHVFLA PQQALSLLQR VRRHGDGSFS DEMNTVLVDN 60 LATRDFINWL IQTKITGGGG GGGSGGGGSG GGGSDTHKSE VAHRFKDLGE EHFKGLVLVA 120 FSQYLQQCPF EEHVKLVNEV TEFAKTCVAD ESAENCDKSL HTLFGDKLCT VATLRETYGE 180 MADCCAKQEP ERNECFLQHK DDNPNLPPLV RPEVDVMCTA FHDNEATFLK KYLYEVARRH 240 PYFYAPELLF FAARYKAAFA ECCQAADKAA CLLPKLDELR DEGKASSAKQ RLKCASLQKF 300 GDRAFKAWAV ARLSQKFPKA EFAEVSKLVT DLTKVHTECC HGDLLECADD RADLAKYMCE 360 NQDSISSKLK ECCDKPLLEK SHCLAEVEND EMPADLPSLA ADYVESKDVC KNYAEAKDVF 420

		LGMFLYEYAR RHPDYSVMLL LRLAKAYEAT LEKCCAAADP HECYAKVFDE FQPLVEEPQN 480 LVKQNCELFE QLGEYKFQNA LLVRYTKKVP QVSTPTLVEV SRNLGKVGAK CCKLPEAKRM 540 PCAEDYLSVV LNRLCVLHEK TPVSEKVTKC CTESLVNRRP CFSALELDEA YVPKAFNAET 600 FTFHADMCTL SEKEKQVKKQ TALVELVKHK PKATKEQLKG VMDNFAAFVE KCCKADDKEA 660 CFAEEGPKFV AASQAALA 678
3-22	**Sequences**	
3-22-1	Sequence Number [ID]	22
3-22-2	Molecule Type	DNA
3-22-3	Length	2037
3-22-4	Features Location/ Qualifiers	**misc_feature 1..2037** note=synthetic construct **source 1..2037** mol_type=other DNA organism=synthetic construct
	NonEnglishQualifier Value	
3-22-5	Residues	atggctcacg ttcggggact gcagctgcct ggatgtctgg ctcttgccgc tctgtgtagc 60 ctggtgcaca gccagcatgt gtttctggct cctcagcaag ccctgagcct gctgcaaaga 120 gttagaaggc acggcgacgg cagcttcagc gacgagatga ataccgtgct ggtggacaac 180 ctggccacca gagacttcat caactggctg atccagacca agatcaccgg tggtggcgga 240 ggcggaggat ctggtggcgg tggttctggc ggtggcggat ctgatacaca caagtctgag 300 gtggcccacc ggttcaagga cctgggcgaa gaacacttca aaggcctggt gctggtcgcc 360 ttcagccagt acctgcagca gtgcctttc gaggaacacg tgaagctggt caacgaagtg 420 accgagttcg ccaagacctg cgtggccgac gagagcgccg agaactgtga taagagcctg 480 cacaccctgt tcggcgacaa gctgtgtaca gtggccacac tgagagaaac ctacggcgag 540 atggccgact gctgcgccaa gcaagagccc gagagaaacg agtgcttcct gcagcacaag 600 gacgacaacc ccaacctgcc tccactcgtc agacccgaag tggacgtgat gtgcaccgcc 660 ttccacgaca atgaggccac cttcctgaag aaatacctgt acgaggtggc cagacggcac 720 ccctactttt acgcccctga actgctgttc tttgccgcca ggtacaaggc cgccttcgcc 780 gaatgttgtc aggccgctga taaggccgct tgcctgctgc ctaagctgga cgagcttaga 840 gacgagggca aagccagctc cgccaagcag agactgaagt gtgccagcct gcagaagttc 900 ggcgatagag cctttaaggc ctgggccgtc gctagactga gccagaagtt tcccaaggcc 960 gagtttgccg aggtgtccaa gctcgtgacc gacctgacaa aggtgcacac cgagtgctgt 1020 cacggcgacc tgctggaatg cgccgacgat agagccgacc tggccaagta catgtgcgag 1080 aaccaggaca gcatcagcag caagctgaaa gagtgctgcg acaagcctct gctggaaaag 1140 agccactgtc tggccgaggt ggaaaacgac gagatgcccg ccgatctgcc ttctctggcc 1200 gccgattacg tggaaagcaa ggacgtgtgc aagaactacg ccgaggccaa ggacgtgttc 1260 ctgggcatgt ttctgtacga gtacgcccgc agacaccccg actactctgt tatgctgctg 1320 ctgagactgg ccaaggccta cgaggccact ctggaaaagt gttgtgccgc cgctgatccc 1380 cacgagtgtt acgccaaagt gttcgacgag ttccagccac tggtggaaga accccagaac 1440 ctggtcaagc agaactgcga gctgttcgag cagctgggcg agtacaagtt ccagaacgcc 1500 ctgctcgtgc ggtacaccaa gaaggtgccc caggtttcca cacctacact ggttgaggtg 1560 tcccggaacc tgggaaaagt gggcgccaag tgttgcaagc tgcctgaggc caagagaatg 1620 ccctgcgccg aggattacct gagcgtggtg ctgaacagac tgtgcgtgct gcacgagaaa 1680 acccctgtgt ccgagaaagt gaccaagtgc tgtaccgaga gcctggtcaa tcggaggcct 1740 tgctttagcg ccctggaact ggacgaggcc tacgtgccca aggccttcaa cgccgagaca 1800 ttcaccttcc acgccgacat gtgtaccctg agcgagaaag aaaagcaagt gaagaaacag 1860 acagccctgg tcgagctggt taagcacaag cctaaggcca ccaaagaaca actgaagggc 1920 gtgatggaca acttcgccgc ctttgtggaa aaatgctgca aggccgacga caaagaggcc 1980 tgcttcgcag aagagggccc taagtttgtg gccgcctctc aagctgctct ggcttaa 2037
3-23	**Sequences**	
3-23-1	Sequence Number [ID]	23
3-23-2	Molecule Type	AA
3-23-3	Length	958
3-23-4	Features Location/ Qualifiers	**REGION 1..958** note=synthetic construct **source 1..958** mol_type=protein organism=synthetic construct
	NonEnglishQualifier Value	
3-23-5	Residues	MAHVRGLQLP GCLALAALCS LVHSQHVFLA PQQALSLLQR VRRHGDGSFS DEMNTVLVDN 60 LATRDFINWL IQTKITDGGG GGGSGGGGSG GGGSSPAGSP TSTEEGTSES ATPESGPGTS 120 TEPSEGSAPG SPAGSPTSTE EGTSTEPSEG SAPGTSTEPS EGSAPGTSES ATPESGPGSE 180 PATSGSETPG SEPATSGSET PGSPAGSPTS TEEGTSESAT PESGPGTSTE PSEGSAPGTS 240 TEPSEGSAPG SPAGSPTSTE EGTSTEPSEG SAPGTSTEPS EGSAPGTSES ATPESGPGTS 300 TEPSEGSAPG TSESATPESG PGSEPATSGS ETPGTSTEPS EGSAPGTSTE PSEGSAPGTS 360

		ESATPESGPG TSESATPESG PGSPAGSPTS TEEGTSESAT PESGPGSEPA TSGSETPGTS 420 ESATPESGPG TSTEPSEGSA PGTSTEPSEG SAPGTSTEPS EGSAPGTSTE PSEGSAPGTS 480 TEPSEGSAPG TSTEPSEGSA PGSPAGSPTS TEEGTSTEPS EGSAPGTSES ATPESGPGSE 540 PATSGSETPG TSESATPESG PGSEPATSGS ETPGTSESAT PESGPGTSTE PSEGSAPGTS 600 ESATPESGPG SPAGSPTSTE EGSPAGSPTS TEEGSPAGSP TSTEEGTSES ATPESGPGTS 660 TEPSEGSAPG TSESATPESG PGSEPATSGS ETPGTSESAT PESGPGSEPA TSGSETPGTS 720 ESATPESGPG TSTEPSEGSA PGSPAGSPTS TEEGTSESAT PESGPGSEPA TSGSETPGTS 780 ESATPESGPG SPAGSPTSTE EGSPAGSPTS TEEGTSTEPS EGSAPGTSES ATPESGPGTS 840 ESATPESGPG TSESATPESG PGSEPATSGS ETPGSEPATS GSETPGSPAG SPTSTEEGTS 900 TEPSEGSAPG TSTEPSEGSA PGSEPATSGS ETPGTSESAT PESGPGTSTE PSEGSAPG 958
3-24	**Sequences**	
3-24-1	Sequence Number [ID]	24
3-24-2	Molecule Type	DNA
3-24-3	Length	2877
3-24-4	Features Location/	**misc_feature 1..2877**
	Qualifiers	note=synthetic construct
		source 1..2877
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-24-5	Residues	atggctcacg ttcggggact gcagctgcct ggatgtctgg ctcttgccgc tctgtgtagc 60 ctggtgcaca gccagcatgt gtttctggct cctcagcaag ccctgagcct gctgcaaaga 120 gttagaaggc acggcgacgg cagcttcagc gacgagatga ataccgtgct ggtggacaac 180 ctggccacca gagacttcat caactggctg atccagacca agatcaccga cggtggcgga 240 ggcggaggat ctggtggtgg tggatctggc ggcggaggaa gttctcctgc tggcagccct 300 acaagcaccg aggaaggcac aagcgagtct gccacacctg agtctggccc tggcacatct 360 acagagccta gcgaaggatc tgccccagga tctcctgccg gctctccaac atctaccgaa 420 gagggaacca gcaccgagcc atctgaggga tctgctcccg gaacaagcac agagccttca 480 gaaggatccg ctcctggcac ctctgaaagc gccacaccag aaagcggacc tggctctgaa 540 cctgccacaa gcggatctga gacacctgga agcgagccag ccacatctgg cagcgaaaca 600 cctggatcac cagccggatc tcccacctct accgaggaag ggacatccga gagcgctacc 660 ccagaatctg gaccaggcac cagcacagaa ccctctgaag gttcagcccc tggaacctct 720 accgaaccat cagaaggcag cgctccaggt tctcccgctg gatcccctac atccacagaa 780 gagggcacct ccactgaacc tagcgaggga agtgctcccg gcacttccac agaaccatcc 840 gagggcagtg cacctggaac cagcgaatct gctacccctg agagtggacc cggaacatcc 900 actgagccct ccgagggttc agctccaggc acatcagaat ccgccactcc agagtccgga 960 ccaggatctg agccagctac cagcggctct gaaacacccg gcactagtac cgagccaagc 1020 gagggtagcg caccagggac aagtaccgaa ccgtctgagg gctccgcacc aggcacttcc 1080 gaaagtgcta ctcctgaaag cggcccaggc actagcgaat ccgcaacacc cgagagcggt 1140 cctggaagtc ctgcaggttc acctaccagc actgaagagg ggactagcga gagcgcaact 1200 cctgaatcag gccctggatc cgaacctgct acctccggaa gtgaaacccc tgggacaagc 1260 gaaagtgcaa cgcccgagtc aggacccggg actagcacgg aacccagtga aggatctgca 1320 cccgggacat ctaccgagcc gtcagaaggt tctgctccag ggactagtac tgagccttcc 1380 gaaggttctg cacctggaac ttccacagag cccagtgaag gcagtgcccc tggcacaagc 1440 actgaaccgt ccgaaggcag tgctcccggg accagtacag aaccgagcga gggctctgct 1500 cctggtagtc cagcaggatc tccaactagc accgaagaag ggacttccac cgagccttcc 1560 gagggaagcg ctcctggaac atccgagtcc gctacgccag agagtggccc aggttctgaa 1620 cccgctactt ccggctcaga gactcctggg acttctgagt ctgcaacccc ggaaagtggt 1680 cctggtagcg aaccagcaac tagcggaagc gagacacccg gaacctcaga gagtgctaca 1740 ccggaatccg gtccagggac gtctacggaa ccgtctgaag gatcagctcc cggcactagc 1800 gaaagcgcta cacctgaaag tggtcccgga tctccagcag gcagcccaac ctctactgaa 1860 gaaggttccc cagctggaag ccccacttcc actgaggaag gctctcccgc aggctcaccc 1920 actagtacgg aagaaggcac gtccgagtct gctactcccg aatccggacc tggaactagc 1980 actgagccaa gcgaaggatc agcacccgga acctctgagt ccgccacacc agaatctggt 2040 cctggttccg agcctgccac ttcaggatca gaaaccccgg gcacgagtga atcagcaacg 2100 ccggaatctg gccccggaag cgaaccggct acgtctggat ctgaaacgcc agggacctcc 2160 gaatcagcta cgcctgagtc tggtccaggg acatccaccg aacctagtga aggctccgca 2220 cctggaagcc ctgctggaag cccaacgagt actgaagagg gcacttctga gagcgctacg 2280 cctgagtcag gacctggaag cgaacctgca acatccggct cagaaacacc agggaccagc 2340 gaaagcgcaa ccccagagag tggacctgga tctccagctg gctctcctac tagtacagag 2400 gaaggcagcc ctgctggctc cccaacgtca acagaagaag gtactagcac agagcccagc 2460 gagggttccg ctccgggaac ttctgaatct gctacacccg agtcaggtcc tggtacaagc 2520 gagtcagcta cgcccgaaag tggacctggc acctcagagt ctgcaactcc tgagagcggt 2580 ccaggatcag aaccagccac ctctggctct gagacaccag gttctgagcc tgcaacgtcc 2640 ggaagcgaaa caccaggcag tcctgccgga agtcctactt caaccgaaga ggggacctct 2700

		acagagccat cagagggctc tgcaccgggc acctcaacag aaccatctga aggatccgca 2760 ccgggctctg agcctgctac tagtggaagc gaaactcctg gcaccagtga atccgctact 2820 cccgagtctg gcccgggaac gtctactgaa ccatctgagg gaagtgcccc aggctaa 2877
3-25	**Sequences**	
3-25-1	Sequence Number [ID]	25
3-25-2	Molecule Type	DNA
3-25-3	Length	665
3-25-4	Features Location/	**misc_feature 1..665**
	Qualifiers	note=synthetic construct
		source 1..665
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-25-5	Residues	ctagtcgaca ttgattattg actagttatt aatagtaatc aattacgggg tcattagttc 60 atagcccata tatggagttc cgcgttacat aacttacggt aaatggcccg cctggctgac 120 cgcccaacga cccccgccca ttgacgtcaa taatgacgta tgttcccata gtaacgccaa 180 tagggacttt ccattgacgt caatgggtgg agtatttacg gtaaactgcc cacttggcag 240 tacatcaagt gtatcatatg ccaagtacgc cccctattga cgtcaatgac ggtaaatggc 300 ccgcctggca ttatgcccag tacatgacct tatgggactt tcctacttgg cagtacatct 360 acgtattagt catcgctatt accatggtcg aggtgagccc cacgttctgc ttcactctcc 420 ccatctcccc cccctcccca cccccaattt tgtatttatt tatttttaa ttattttgtg 480 cagcgatggg ggcgggggg gggggggggc gcgcgccagg cggggcgggg cggggcgagg 540 ggcggggcgg ggcgaggcgg agaggtgcgg cggcagccaa tcagagcggc gcgctccgaa 600 agtttccttt tatggcgagg cggcggcggc ggcggcccta taaaaagcga agcgcgcggc 660 gggcg 665
3-26	**Sequences**	
3-26-1	Sequence Number [ID]	26
3-26-2	Molecule Type	DNA
3-26-3	Length	529
3-26-4	Features Location/	**misc_feature 1..529**
	Qualifiers	note=synthetic construct
		source 1..529
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-26-5	Residues	cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 60 gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 120 atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 180 aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 240 catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 300 catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 360 atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 420 ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 480 acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatc 529
3-27	**Sequences**	
3-27-1	Sequence Number [ID]	27
3-27-2	Molecule Type	DNA
3-27-3	Length	978
3-27-4	Features Location/	**misc_feature 1..978**
	Qualifiers	note=synthetic construct
		source 1..978
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-27-5	Residues	ggagtgcagg tggaaaccat ctccccagga gacgggcgca ccttccccaa gcgcggccag 60 acctgcgtgg tgcactacac cgggatgctt gaagatggaa agaaatttga ttcctcccgg 120 gacagaaaca agcctttaa gtttatgcta ggcaagcagg aggtgatccg aggctgggaa 180 gaaggggttg cccagatgag tgtgggtcag agagccaaac tgactatatc tccagattat 240 gcctatggtg ccactgggca cccaggcatc atcccaccac atgccactct cgtcttcgat 300 gtggagcttc taaaactgga aactagaggc gttcaggtgg aaaccatcag tccaggggat 360 ggccgaactt ttccaaagag agggcagact tgcgtcgtgc attatactgg tatgctggag 420 gatgggaaaa agttcgactc ttccagagat cggaacaaac cattcaaatt catgctcggg 480 aaacaggaag ttatccgcgg atgggaggag ggcgtggccc agatgtccgt gggccagcgc 540 gccaagctaa ccatctcccc agactacgcc tacggagcca ccggacaccc cggtatcata 600 cccccacacg ccacccttgt gtttgacgtg gaactgctta agctagagac tagaggcgtg 660

		caggtcgaga ccatcagccc cggcgacggc cgcacctttc ccaagagagg ccagacttgc 720
		gtggtccact acaccggcat gctggaggac ggcaagaagt tcgacagcag ccgcgaccgc 780
		aacaagccct tcaagttcat gctgggcaaa caggaagtga tccgcggctg ggaggaaggc 840
		gtggctcaga tgagcgtggg gcagcgggcc aagctgacca tcagccccga ctatgcctac 900
		ggcgccaccg gccaccccgg catcatcccc ccccacgcca ccctggtgtt cgacgtggag 960
		ctgctgaagc tggagtga 978
3-28	**Sequences**	
3-28-1	Sequence Number [ID]	28
3-28-2	Molecule Type	AA
3-28-3	Length	107
3-28-4	Features Location/	**REGION 1..107**
	Qualifiers	note=synthetic construct
		source 1..107
		mol_type=protein
		organism=synthetic construct
	NonEnglishQualifier Value	
3-28-5	Residues	GVQVETISPG DGRTFPKRGQ TCVVHYTGML EDGKKFDSSR DRNKPFKFML GKQEVIRGWE 60
		EGVAQMSVGQ RAKLTISPDY AYGATGHPGI IPPHATLVFD VELLKLE 107
3-29	**Sequences**	
3-29-1	Sequence Number [ID]	29
3-29-2	Molecule Type	DNA
3-29-3	Length	18
3-29-4	Features Location/	**misc_feature 1..18**
	Qualifiers	note=synthetic construct
		source 1..18
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-29-5	Residues	aatgatgggc gctcgagt 18
3-30	**Sequences**	
3-30-1	Sequence Number [ID]	30
3-30-2	Molecule Type	DNA
3-30-3	Length	121
3-30-4	Features Location/	**misc_feature 1..121**
	Qualifiers	note=synthetic construct
		source 1..121
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-30-5	Residues	aacattttga cacccccata atatttttcc agaattaaca gtataaattg catctcttgt 60
		tcaagagttc cctatcactc tctttaatca ctactcacag taacctcaac tcctgccaca 120
		a 121
3-31	**Sequences**	
3-31-1	Sequence Number [ID]	31
3-31-2	Molecule Type	DNA
3-31-3	Length	2211
3-31-4	Features Location/	**source 1..2211**
	Qualifiers	mol_type=unassigned DNA
		organism=Adeno-Associated Virus rh91
	NonEnglishQualifier Value	
3-31-5	Residues	atggctgctg acggttatct tccagattgg ctcgaggaca acctttctga aggcattcgt 60
		gagtggtggg ctctgaaacc tggagcccct aaacccaaag cgaaccaaca aaagcaggac 120
		gacggccggg gtcttgtgct tccgggttac aaatacctcg gacccttcaa cggactcgac 180
		aaaggagagc cggtcaacgc ggcggacgcg gcagccctcg aacacgacaa agcttacgac 240
		cagcagctca aggccggtga caacccgtac ctccggtaca accacgccga cgccgagttt 300
		caggagcgtc ttcaagaaga tacgtctttt gggggcaacc ttggcagagc agtcttccag 360
		gccaaaaaga gggttcttga gccttttggt ctggttgagg aagcagctaa aacggctcct 420
		ggaaagaaga ggcctgtaga gcagtctcct caggaaccgg actcatcatc tggtattggc 480
		aaatcgggcc agcagcctgc caaaaaaaga ctaaatttcg gtcagactgg cgactcagag 540
		tcagtccccg accctcaacc tctcggagaa cctccagaaa cccccgctgc tgtgggacct 600
		actacaatgg cttcaggcgg tggcgcacca atggcagaca ataacgaagg cgccgacgga 660
		gtgggtaatg cctcaggaaa ttggcattgc gattccacat ggctgggcga cagagtcatc 720
		accaccagca cccgaacctg ggcccttcct acctacaaca accacctcta caagcaaatc 780
		tccagcgctt caacgggggc cagtaacgac aaccactact ttggctacag cacccctgg 840
		gggtattttg atttcaacag attccactgc cacttctcac cacgtgactg gcagcgactc 900

		attaacaaca actggggatt ccggcccaag agactcaact tcaagctctt caacatccag 960 gtcaaggagg tcacgacgaa tgatggcgtc acaaccatcg ctaataacct taccagcacg 1020 gttcaagtgt tctcggactc ggagtaccag ctgccgtacg tcctcggttc tgcgcaccag 1080 ggctgcctcc ctccgttccc ggcggacgta ttcatgattc ctcagtatgg atacctcacc 1140 ctgaacaacg gaagtcaagc ggtgggacgc tcatcctttt actgcctgga gtacttccct 1200 tcgcagatgc taaggactgg aaataacttc accttcagct ataccttcga ggatgtacct 1260 tttcacagca gctacgctca cagccagagt ttggatcgct tgatgaatcc tcttattgat 1320 cagtatctgt actacctgaa cagaacgcaa aatcaatctg gaagtgcaca aaacaaggac 1380 ctgcttttta gccgggggtc tcctgctggc atgtctgttc agcccaaaaa ttggctacct 1440 gggccctgct accggcaaca gagagtttca aagactaaaa cagacaacaa caacagtaac 1500 tttacctgga caggtgccag caaatataat ctcaatggcc gcgaatcgat cattaatcca 1560 ggaaccgcta tggccagtca caaggacgat gaagacaaat ttttccctat gagcggcgtt 1620 atgatatttg gcaaagaaaa tgcaggagca agtaacactg cattagataa tgtaatgatt 1680 acggatgaag aagagattaa agctaccaat cctgtggcaa cagagagatt tggaactgtg 1740 gcagtcaact tgcagagctc aaatacagac cccgcaactg gagacgtcca tgtcatgggg 1800 gccttacctg gcatggtgtg gcaagatcgt gacgtgtacc ttcaaggacc tatctgggca 1860 aagattcctc acacggatgg acactttcat ccttctcctc tgatgggagg ctttggactg 1920 aaacatccgc ctcctcaaat cctcatcaaa aatactccgg taccggcaaa tcctccggca 1980 gagttcagcg ctacaaagtt tgcttcattt atcactcagt actccactgg acaggtcagc 2040 gtggaaattg agtgggagct acagaaagaa aacagcaaac gttggaatcc agaggtgcag 2100 tacacttcca actacgcgaa gtctgccaat gtggacttta ctgtagacaa caatggtctt 2160 tatactgaac ctcgccctat tggaacccgg tatctcacac gacccttgta a 2211
3-32	**Sequences**	
3-32-1	Sequence Number [ID]	32
3-32-2	Molecule Type	AA
3-32-3	Length	736
3-32-4	Features Location/	**source 1..736**
	Qualifiers	mol_type=protein
		organism=adeno-associated virus rh91
	NonEnglishQualifier Value	
3-32-5	Residues	MAADGYLPDW LEDNLSEGIR EWWALKPGAP KPKANQQKQD DGRGLVLPGY KYLGPFNGLD 60 KGEPVNAADA AALEHDKAYD QQLKAGDNPY LRYNHADAEF QERLQEDTSF GGNLGRAVFQ 120 AKKRVLEPFG LVEEAAKTAP GKKRPVEQSP QEPDSSSGIG KSGQQPAKKR LNFGQTGDSE 180 SVPDPQPLGE PPETPAAVGP TTMASGGGAP MADNNEGADG VGNASGNWHC DSTWLGDRVI 240 TTSTRTWALP TYNNHLYKQI SSASTGASND NHYFGYSTPW GYFDFNRFHC HFSPRDWQRL 300 INNNWGFRPK RLNFKLFNIQ VKEVTTNDGV TTIANNLTST VQVFSDSEYQ LPYVLGSAHQ 360 GCLPPFPADV FMIPQYGYLT LNNGSQAVGR SSFYCLEYFP SQMLRTGNNF TFSYTFEDVP 420 FHSSYAHSQS LDRLMNPLID QYLYYLNRTQ NQSGSAQNKD LLFSRGSPAG MSVQPKNWLP 480 GPCYRQQRVS KTKTDNNNSN FTWTGASKYN LNGRESIINP GTAMASHKDD EDKFFPMSGV 540 MIFGKENAGA SNTALDNVMI TDEEEIKATN PVATERFGTV AVNLQSSNTD PATGDVHVMG 600 ALPGMVWQDR DVYLQGPIWA KIPHTDGHFH PSPLMGGFGL KHPPPQILIK NTPVPANPPA 660 EFSATKFASF ITQYSTGQVS VEIEWELQKE NSKRWNPEVQ YTSNYAKSAN VDFTVDNNGL 720 YTEPRPIGTR YLTRPL 736
3-33	**Sequences**	
3-33-1	Sequence Number [ID]	33
3-33-2	Molecule Type	DNA
3-33-3	Length	3326
3-33-4	Features Location/	**misc_feature 1..3326**
	Qualifiers	note=synthetic construct
		source 1..3326
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-33-5	Residues	ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgtcg ggcgaccttt 60 ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggccaa ctccatcact 120 aggggttcct tgtagttaat gattaacccg ccatgctact tatctaccag ggtaatgggg 180 atcctctaga actatagcta gtcgacattg attattgact agttattaat agtaatcaat 240 tacggggtca ttagttcata gcccatatat ggagttccgc gttacataac ttacggtaaa 300 tggcccgcct ggctgaccgc ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt 360 tcccatagta acgccaatag ggactttcca ttgacgtcaa tgggtggagt atttacggta 420 aactgcccac ttggcagtac atcaagtgta tcatatgcca agtacgcccc ctattgacgt 480 caatgacggt aaatggcccg cctggcatta tgcccagtac atgaccttat gggactttcc 540 tacttggcag tacatctacg tattagtcat cgctattacc atggtcgagg tgagccccac 600 gttctgcttc actctcccca tctccccccc ctccccaccc ccaattttgt atttatttat 660 tttttaatta ttttgtgcag cgatgggggc ggggggggggg gggggcgcgc cgccaggcgg 720

		ggcggggcgg ggcgaggggc ggggcggggc gaggcggaga ggtgcggcgg cagccaatca 780 gagcggcgcg ctccgaaagt ttccttttat ggcgaggcgg cggcggcggc ggccctataa 840 aaagcgaagc gcgcggcggg cgggagtcgc tgcgcgctgc cttcgccccg tgccccgctc 900 cgccgccgcc tcgcgccgcc cgccccggct ctgactgacc gcgttactcc cacaggtgag 960 cgggcgggac ggcccttctc ctccgggctg taattagcgc ttggtttaat gacggcttgt 1020 ttcttttctg tggctgcgtg aaagccttga ggggctccgg gagggccctt tgtgcggggg 1080 gagcggctcg gggggtgcgt gcgtgtgtgt gtgcgtgggg agcgccgcgt gcggctccgc 1140 gctgcccggc ggctgtgagc gctgcgggcg cggcgcgggg ctttgtgcgc tccgcagtgt 1200 gcgcgagggg agcgcggccg ggggcggtgc cccgcggtgc gggggggggct gcgaggggaa 1260 caaaggctgc gtgcggggtg tgtgcgtggg ggggtgagca gggggtgtgg gcgcgtcggt 1320 cgggctgcaa ccccccctgc acccccctcc ccgagttgct gagcacggcc cggcttcggg 1380 tgcggggctc cgtacggggc gtggcgcggg gctcgccgtg ccgggcgggg ggtggcggca 1440 ggtgggggtg ccgggcgggg cggggccgcc tcgggccggg gagggctcgg gggaggggcg 1500 cggcggcccc cggagcgccg gcggctgtcg aggcgcggcg agccgcagcc attgcctttt 1560 atggtaatcg tgcgagaggg cgcagggact tcctttgtcc caaatctgtg cggagccgaa 1620 atctgggagg cgccgccgca ccccctctag cgggcgcggg gcgaagcggt gcggcgccgg 1680 caggaaggaa atgggcgggg agggccttcg tgcgtcgccg cgccgccgtc cccttctccc 1740 tctccagcct cggggctgtc cgcgggggga cggctgcctt cgggggggac ggggcagggc 1800 ggggttcggc ttctggcgtg tgaccggcgg ctctagagcc tctgctaacc atgttcatgc 1860 cttcttcttt ttcctacagc tcctgggcaa cgtgctggtt attgtgctgt ctcatcattt 1920 tggcaaagaa ttcgccctgc agcgggaatt gccaccatgg ctcacgttcg gggactgcag 1980 ctgcctggat gtctggctct tgccgctctg tgtagcctgg tgcacagcca gcacgtgttt 2040 ctggctcctc agcaagccag atcactgctg cagagagtta gaaggcacgg cgacggcagc 2100 ttcagcgacg agatgaacac catcctggac aacctggccg ccagagactt catcaactgg 2160 ctgatccaga ccaagatcac cgacggtggc ggaggcggag gatctggtgg tggtggatct 2220 ggcggcggag gttctgccga gtctaagtac ggacctcctt gtcctccctg tcctgctcca 2280 gaagctgctg gcggcccatc cgtgtttctg ttccctccaa agcctaagga caccctgatg 2340 atcagcagaa cccctgaagt gacctgcgtg gtggtcgacg tgtcccaaga ggatcctgag 2400 gtgcagttca attggtacgt ggacggcgtg gaagtgcaca acgccaagac caagcctaga 2460 gaggaacagt tcaacagcac ctacagagtg gtgtccgtgc tgaccgtgct gcaccaggat 2520 tggctgaacg gcaaagagta caagtgcaag gtgtccaaca agggcctgcc tagcagcatc 2580 gagaaaacca tcagcaaggc caagggccag ccaagagaac cccaggtgta cacactgcct 2640 ccaagccaag aggaaatgac caagaaccag gtgtccctga cctgcctggt caagggcttc 2700 taccccttccg atatcgccgt ggaatgggag agcaacggcc agcctgagaa caactacaag 2760 accacacctc ctgtgctgga ctccgatggc tcattcttcc tgtacagcag actgaccgtg 2820 gacaagagca ggtggcaaga gggcaacgtg ttcagctgca gcgtgatgca cgaggccctg 2880 cacaaccact acacccagaa aagcctgagc ctgtctctgg gctaataagg tacctctaga 2940 gcggccgcct cgaggacggg gtgaactacg cctgaggatc cgatcttttt ccctctgcca 3000 aaaattatgg ggacatcatg aagccccttg agcatctgac ttctggctaa taaaggaaat 3060 ttattttcat tgcaatagtg tgttggaatt ttttgtgtct ctcactcgga agcaattcgt 3120 tgatctgaat ttcgaccacc cataataccc attaccctgg tagataagta gcatggcggg 3180 ttaatcatta actacaagga acccctagtg atggagttgg ccactccctc tctgcgcgct 3240 cgctcgctca ctgaggccgg gcgaccaaag gtcgcccgac gcccgggctt tgcccgggcg 3300 gcctcagtga gcgagcgagc gcgcag 3326
3-34	**Sequences**	
3-34-1	Sequence Number [ID]	34
3-34-2	Molecule Type	DNA
3-34-3	Length	4379
3-34-4	Features Location/	**misc_feature 1..4379**
	Qualifiers	note=synthetic construct
		source 1..4379
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-34-5	Residues	ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgtcg ggcgaccttt 60 ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggccaa ctccatcact 120 aggggttcct tgtagttaat gattaacccg ccatgctact tatctaccag ggtaatgggg 180 atcctctaga actatagcta gtcgacattg attattgact agttattaat agtaatcaat 240 tacggggtca ttagttcata gcccatatat ggagttccgc gttacataac ttacggtaaa 300 tggcccgcct ggctgaccgc ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt 360 tcccatagta acgccaatag ggactttcca ttgacgtcaa tgggtggagt atttacggta 420 aactgcccac ttggcagtac atcaagtgta tcatatgcca agtacgcccc ctattgacgt 480 caatgacggt aaatggcccg cctggcatta tgcccagtac atgaccttat gggactttcc 540 tacttggcag tacatctacg tattagtcat cgctattacc atggtcgagg tgagccccac 600 gttctgcttc actctcccca tctccccccc ctccccaccc ccaattttgt atttatttat 660

		ttttttaatta ttttgtgcag cgatgggggc ggggggggggg gggggcgcgg cgccaggcgg 720 ggcggggcgg ggcgaggggc ggggcggggc gaggcggaga ggtgcggcgg cagccaatca 780 gagcggcgcg ctccgaaagt ttccttttat ggcgaggcgg cggcggcggc ggccctataa 840 aaagcgaagc gcgcggcggg cgggagtcgc tgcgcgctgc cttcgccccg tgccccgctc 900 cgccgccgcc tcgcgccgcc cgccccggct ctgactgacc gcgttactcc cacaggtgag 960 cgggcgggac ggcccttctc ctccgggctg taattagcgc ttggtttaat gacggcttgt 1020 ttcttttctg tggctgcgtg aaagccttga ggggctccgg gagggccctt tgtgcggggg 1080 gagcggctcg gggggtgcgt gcgtgtgtgt gtgcgtgggg agcgccgcgt gcggctccgc 1140 gctgcccggc ggctgtgagc gctgcgggcg cggcgcgggg ctttgtgcgc tccgcagtgt 1200 gcgcgagggg agcgcggccg ggggcggtgc cccgcggtgc gggggggggct gcgaggggaa 1260 caaaggctgc gtgcggggtg tgtgcgtggg ggggtgagca gggggtgtgg gcgcgtcggt 1320 cgggctgcaa ccccccctgc acccccctcc ccgagttgct gagcacggcc cggcttcggg 1380 tgcggggctc cgtacggggc gtggcgcggg gctcgccgtg ccgggcgggg ggtggcggca 1440 ggtgggggtg ccgggcgggg cggggccgcc tcgggccggg gagggctcgg gggaggggcg 1500 cggcggcccc cggagcgccg gcggctgtcg aggcgcggcg agccgcagcc attgcctttt 1560 atggtaatcg tgcgagaggg cgcagggact tcctttgtcc caaatctgtg cggagccgaa 1620 atctgggagg cgccgccgca ccccctctag cgggcgcggg gcgaagcggt gcggcgccgg 1680 caggaaggaa atgggcgggg agggccttcg tgcgtcgccg cgccgccgtc cccttctccc 1740 tctccagcct cggggctgtc cgcgggggga cggctgcctt cgggggggac ggggcagggc 1800 ggggttcggc ttctggcgtg tgaccggcgg ctctagagcc tctgctaacc atgttcatgc 1860 cttcttcttt ttcctacagc tcctgggcaa cgtgctggtt attgtgctgt ctcatcattt 1920 tggcaaagaa ttccagaagc caccatggct cacgttcggg gactgcagct gcctggatgt 1980 ctggctcttg ccgctctgtg tagcctggtg cacagccagc acgtgtttct ggctcctcag 2040 caagccagat cactgctgca gagagttaga aggcacggcg acggcagctt cagcgacgag 2100 atgaacacca tcctggacaa cctggccgcc agagacttca tcaactggct gatccagacc 2160 aagatcaccg acggtggcgg aggcggagga tctggtggtg gtggatctgg cggcggaggt 2220 tctgacgccc acaaatctga agtggcccac cggttcaagg acctgggcga agagaatttc 2280 aaggccctgg tgctgatcgc cttcgctcag tacctgcagc agtgcccctt cgaggaccac 2340 gtgaagctgg tcaacgaagt gaccgagttc gccaagacct gcgtggccga cgagagcgcc 2400 gagaactgtg ataagagcct gcacaccctg ttcggcgaca agctgtgtac agtggccaca 2460 ctgagagaaa cctacggcga gatggccgac tgctgcgcca agcaagagcc cgagagaaac 2520 gagtgcttcc tgcagcacaa ggacgacaac cccaacctgc ctagactcgt gcggcctgaa 2580 gtggacgtga tgtgcaccgc cttccacgac aacgaggaaa ccttcctgaa gaagtacctg 2640 tacgagatcg ccagacggca cccctacttt tacgcccctg agctgctgtt cttcgccaag 2700 cggtataagg ccgccttcac cgagtgttgt caggccgctg ataaggctgc ctgcctgctg 2760 cctaagctgg acgagcttag agacgagggc aaagccagct ccgccaagca gagactgaag 2820 tgtgccagcc tgcagaagtt cggcgagaga gcctttaagg cctgggccgt tgctagactg 2880 agccagagat ttcccaaggc cgagtttgcc gaggtgtcca agctcgtgac cgacctgaca 2940 aaggtgcaca ccgagtgctg ccacggcgac ctgctggaat gcgccgacga tagagccgac 3000 ctggccaagt acatctgcga gaaccaggac agcatcagca gcaagctgaa agagtgctgc 3060 gagaagcctc tgctggaaaa gagccactgt atcgccgagg tggaaaacga cgagatgccc 3120 gccgatctgc cttctctggc cgccgatttt gtggaaagca aggacgtgtg caagaactac 3180 gccgaggcca aggacgtgtt cctgggcatg tttctgtacg agtacgcccg cagacacccc 3240 gactactctg ttgtgctgct gctgagactg gccaaaacct acgagacaac cctggaaaag 3300 tgctgtgccg ccgctgatcc tcacgagtgt tacgccaagg tgttcgacga gttcaagcca 3360 ctggtggaag aaccccagaa cctgatcaag cagaactgcg agctgttcga gcagctgggc 3420 gagtacaagt tccagaacgc cctgctcgtg cggtacacca agaaggtgcc ccaggtttcc 3480 acacctacac tggttgaggt gtcccggaac ctgggcaaag tgggcagcaa gtgttgcaag 3540 caccctgagg ccaagagaat gccctgcgcc gaggattacc tgagcgtggt gctgaatcag 3600 ctgtgcgtgc tgcacgagaa aacccctgtg tccgacagag tgaccaagtg ctgtaccgag 3660 agcctggtca acagacggcc ttgctttagc gccctcgagg tggacgagac atacgtgccc 3720 aaagagttca acgccgagac attcaccttc cacgccgaca tctgtaccct gagcgagaaa 3780 gagcggcaga tcaagaaaca gactgccctg gtggaactgg tcaagcacaa gcccaaggcc 3840 accaaagaac agctgaaggc cgtgatggac gacttcgccg ccttcgtgga aaagtgctgc 3900 aaggccgacg acaaagagac ctgcttcgcc gaagagggca agaaactggt ggccgcttct 3960 caggctgctc tgggataata aggtacctct agagcggccg cctcgaggac ggggtgaact 4020 acgcctgagg atccgatctt tttccctctg ccaaaaatta tggggacatc atgaagcccc 4080 ttgagcatct gacttctggc taataaagga aatttatttt cattgcaata gtgtgttgga 4140 atttttgtg tctctcactc ggaagcaatt cgttgatctg aatttcgacc acccataata 4200 cccattaccc tggtagataa gtagcatggc gggttaatca ttaactacaa ggaaccccta 4260 gtgatggagt tggccactcc ctctctgcgc gctcgctcgc tcactgaggc cgggcgacca 4320 aaggtcgccc gacgcccggg ctttgcccgg gcggcctcag tgagcgagcg agcgcgcag 4379
3-35	**Sequences**	
3-35-1	Sequence Number [ID]	35
3-35-2	Molecule Type	DNA

3-35-3	Length	5219
3-35-4	Features Location/ Qualifiers	**misc_feature 1..5219** note=synthetic construct **source 1..5219** mol_type=other DNA organism=synthetic construct
	NonEnglishQualifier Value	
3-35-5	Residues	ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgtcg ggcgaccttt 60 ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggccaa ctccatcact 120 aggggttcct tgtagttaat gattaacccg ccatgctact tatctaccag ggtaatgggg 180 atcctctaga actatagcta gtcgacattg attattgact agttattaat agtaatcaat 240 tacggggtca ttagttcata gcccatatat ggagttccgc gttacataac ttacggtaaa 300 tggcccgcct ggctgaccgc ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt 360 tcccatagta acgccaatag ggactttcca ttgacgtcaa tgggtggagt atttacggta 420 aactgcccac ttggcagtac atcaagtgta tcatatgcca agtacgcccc ctattgacgt 480 caatgacggt aaatggcccg cctggcatta tgcccagtac atgaccttat gggactttcc 540 tacttggcag tacatctacg tattagtcat cgctattacc atggtcgagg tgagccccac 600 gttctgcttc actctcccca tctccccccc ctccccaccc ccaattttgt atttatttat 660 tttttaatta ttttgtgcag cgatgggggc ggggggggag gggggcgcgg cgccaggcgg 720 ggcggggcgg ggcgaggggc ggggcggggc gaggcggaga ggtgcggcgg cagccaatca 780 gagcggcgcg ctccgaaagt ttccttttat ggcgaggcgg cggcggcggc ggccctataa 840 aaagcgaagc gcgcggcggg cgggagtcgc tgcgcgctgc cttcgccccg tgccccgctc 900 cgccgccgcc tcgcgccgcc cgccccggct ctgactgacc gcgttactcc cacaggtgag 960 cgggcgggac ggcccttctc ctccgggctg taattagcgc ttggtttaat gacggcttgt 1020 ttcttttctg tggctgcgtg aaagccttga ggggctccgg gagggccctt tgtgcggggg 1080 gagcggctcg gggggtgcgt gcgtgtgtgt gtgcgtgggg agcgccgcgt gcggctccgc 1140 gctgcccggc ggctgtgagc gctgcgggcg cggcgcgggg ctttgtgcgc tccgcagtgt 1200 gcgcgagggg agcgcggccg ggggcggtgc cccgcggtgc gggggggggct gcgaggggaa 1260 caaaggctgc gtgcggggtg tgtgcgtggg ggggtgagca gggggtgtgg gcgcgtcggt 1320 cgggctgcaa ccccccctgc acccccctcc ccgagttgct gagcacggcc cggcttcggg 1380 tgcggggctc cgtacggggc gtggcgcggg gctcgccgtg ccgggcgggg ggtggcggca 1440 ggtgggggtg ccgggcgggg cggggccgcc tcgggccggg gagggctcgg gggaggggcg 1500 cggcggcccc cggagcgccg gcggctgtcg aggcgcggcg agccgcagcc attgcctttt 1560 atggtaatcg tgcgagaggg cgcagggact tcctttgtcc caaatctgtg cggagccgaa 1620 atctgggagg cgccgccgca ccccctctag cgggcgcggg gcgaagcggt gcggcgccgg 1680 caggaaggaa atgggcgggg agggccttcg tgcgtcgccg cgccgccgtc cccttctccc 1740 tctccagcct cggggctgtc cgcgggggga cggctgcctt cgggggggac ggggcagggc 1800 ggggttcggc ttctggcgtg tgaccggcgg ctctagagcc tctgctaacc atgttcatgc 1860 cttcttcttt ttcctacagc tcctgggcaa cgtgctggtt attgtgctgt ctcatcattt 1920 tggcaaagaa ttccagaagc caccatggct cacgttcggg gactgcagct gcctggatgt 1980 ctggctcttg ccgctctgtg tagcctggtg cacagccagc acgtgtttct ggctcctcag 2040 caagccagat cactgctgca gagagttaga aggcacggcg acggcagctt cagcgacgag 2100 atgaacacca tcctggacaa cctggccgcc agagacttca tcaactggct gatccagacc 2160 aagatcaccg acggtggcgg aggcggagga tctggtggtg gtggatctgg cggcggagga 2220 agttctcctg ctggcagccc tacaagcacc gaggaaggca caagcgagtc tgccacacct 2280 gagtctggcc ctggcacatc tacagagcct agcgaaggat ctgccccagg atctcctgcc 2340 ggctctccaa catctaccga agagggaacc agcaccgagc catctgaggg atctgctccc 2400 ggaacaagca cagagccttc agaaggatcc gctcctggca cctctgaaag cgccacacca 2460 gaaagcggac ctggatctga gcctgccaca agcggatctg agacacctgg aagcgagcca 2520 gccacatctg gcagcgaaac acctggttct ccagccggat ctcccaccag cacagaagag 2580 ggcacatccg aatctgctac ccctgaatct ggaccaggca cctccacaga accttccgag 2640 ggttctgccc ctggaacctc taccgaacca tcagaaggca gcgctccagg ttcaccagct 2700 ggaagcccaa cctctacaga ggaagggaca tccactgagc caagcgaggg aagcgctccc 2760 ggcactagta cagaaccgag cgagggcagt gctcctggaa ccagcgaatc cgctactcca 2820 gagagtggcc caggcaccag tactgaaccc tctgagggta gcgcacccgg aacatctgag 2880 agcgctactc ccgaatcagg cccaggctct gaacctgcta ccagcggaag tgaaacaccc 2940 ggcacctcta ctgagccctc cgaaggctca gcacctggca caagcactga accatcagag 3000 ggctccgcac caggcaccag cgaaagtgct acaccagagt caggacccgg aacctccgaa 3060 agtgcaactc ctgagagcgg accaggctct cccgctggat ctcctacatc aactgaagaa 3120 gggacctccg agagcgcaac cccagagtct ggtccaggat cagaacctgc cacctccggc 3180 tctgaaaccc caggcacttc tgagtccgcc acgccagaat ctggtcctgg gactagcacc 3240 gaaccgagtg aaggttcagc tcccgggact tctacggaac ccagtgaagg atctgcaccc 3300 ggcacatcaa ccgaaccgtc agagggatca gcccctggga cttccacaga gccgtctgag 3360 ggcagcgccc cagggacgtc tacagaacca tctgaaggat cagcaccagg gacctctacc 3420 gagccaagtg aaggcagtgc accgggaagt ccagcaggct cccctacaag tactgaagag 3480

		ggtactagca cggaacccag cgagggttcc gctccaggga catctgaatc cgcaactccg 3540 gaatccggac ctggcagtga accagctaca tccggatccg agactccggg aacctcagaa 3600 tcagctacac ccgagagtgg acctggctcc gaaccagcaa ctagcggctc agaaactcct 3660 gggacaagcg agagtgcaac acccgaatct ggacctggaa caagtactga gcccagcgaa 3720 ggcagcgccc ctggaacttc tgaatctgcc actcctgaaa gtggccctgg aagccctgca 3780 ggctcaccca catccacaga agaaggatca ccagcaggca gccccacttc aacggaagag 3840 ggatccccag ctggatcccc aactagtacg gaagaaggca cttcagaaag cgctacgccc 3900 gagtccggtc ctggcacttc tactgaacca tccgagggaa gtgcccctgg cacttccgag 3960 agtgctacac ctgaaagcgg tcccggctct gaaccagcca cttctggatc tgaaacgccc 4020 gggacatccg agtcagcaac gcccgaaagc ggcccaggtt ccgagccggc tactagtggt 4080 tcagagactc cagggacttc cgagtctgct actcctgagt ccggaccggg aacatcaacc 4140 gagccttccg aaggatctgc acctggaagc cctgccggat ctcctaccag tactgaggaa 4200 ggcacctcag agtctgccac tccagagtca ggtcctggaa gcgaacctgc aacaagcggc 4260 agcgaaactc caggcactag cgagtcagct accccagaat caggacctgg atctccagca 4320 gggtccccaa catctactga ggaaggctct cctgctggct cccctacctc taccgaagag 4380 gggacctcaa cagagccatc cgaggggagc gcacctggta catcagagtc cgcaactccc 4440 gagtctggcc ccggaactag cgaatctgca accccggaaa gtggacccgg gacgagtgaa 4500 tcagccacac ctgaatccgg tccaggatcc gagcctgcaa cttctggaag cgagacacca 4560 ggatctgagc cagctacgtc tggctctgag actcctggat ctcctgctgg tagtcccacc 4620 tccactgaag agggaacttc caccgaaccg agcgagggat cagcaccagg cactagcaca 4680 gaaccgtccg aaggatctgc tccaggctct gaacccgcaa cctccggatc agaaacccct 4740 ggaacatccg aaagcgctac accggaaagt ggccccggaa cctctacaga acctagcgag 4800 ggaagcgcac caggataata aggtacctct agagcggccg cctcgaggac ggggtgaact 4860 acgcctgagg atccgatctt tttccctctg ccaaaaatta tggggacatc atgaagcccc 4920 ttgagcatct gacttctggc taataaagga aatttatttt cattgcaata gtgtgttgga 4980 attttttgtg tctctcactc ggaagcaatt cgttgatctg aatttcgacc acccataata 5040 cccattaccc tggtagataa gtagcatggc gggttaatca ttaactacaa ggaaccccta 5100 gtgatggagt tggccactcc ctctctgcgc gctcgctcgc tcactgaggc cgggcgacca 5160 aaggtcgccc gacgcccggg ctttgcccgg gcggcctcag tgagcgagcg agcgcgcag 5219
3-36	**Sequences**	
3-36-1	Sequence Number [ID]	36
3-36-2	Molecule Type	DNA
3-36-3	Length	3326
3-36-4	Features Location/	**misc_feature 1..3326**
	Qualifiers	note=synthetic construct
		source 1..3326
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-36-5	Residues	ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgtcg ggcgaccttt 60 ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggccaa ctccatcact 120 aggggttcct tgtagttaat gattaacccg ccatgctact tatctaccag ggtaatgggg 180 atcctctaga actatagcta gtcgacattg attattgact agttattaat agtaatcaat 240 tacggggtca ttagttcata gcccatatat ggagttccgc gttacataac ttacggtaaa 300 tggcccgcct ggctgaccgc ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt 360 tcccatagta acgccaatag ggactttcca ttgacgtcaa tgggtggagt atttacggta 420 aactgcccac ttggcagtac atcaagtgta tcatatgcca agtacgcccc ctattgacgt 480 caatgacggt aaatggcccg cctggcatta tgcccagtac atgaccttat gggactttcc 540 tacttggcag tacatctacg tattagtcat cgctattacc atggtcgagg tgagccccac 600 gttctgcttc actctcccca tctccccccc ctccccaccc ccaattttgt atttatttat 660 tttttaatta ttttgtgcag cgatgggggc gggggggggg gggggcgcgg cgccaggcgg 720 ggcggggcgg ggcgaggggc ggggcggggc gaggcggaga ggtgcggcgg cagccaatca 780 gagcggcgcg ctccgaaagt ttccttttat ggcgaggcgg cggcggcggc ggccctataa 840 aaagcgaagc gcgcggcggg cgggagtcgc tgcgcgctgc cttcgccccg tgccccgctc 900 cgccgccgcc tcgcgccgcc cgccccggct ctgactgacc gcgttactcc cacaggtgag 960 cgggcgggac ggcccttctc ctccgggctg taattagcgc ttggtttaat gacggcttgt 1020 ttcttttctg tggctgcgtg aaagccttga ggggctccgg gagggccctt tgtgcggggg 1080 gagcggctcg gggggtgcgt gcgtgtgtgt gtgcgtgggg agcgccgcgt gcggctccgc 1140 gctgcccggc ggctgtgagc gctgcgggcg cggcgcgggg ctttgtgcgc tccgcagtgt 1200 gcgcgagggg agcgcggccg ggggcggtgc cccgcggtgc gggggggggct gcgaggggaa 1260 caaaggctgc gtgcggggtg tgtgcgtggg ggggtgagca gggggtgtgg gcgcgtcggt 1320 cgggctgcaa ccccccctgc acccccctcc ccgagttgct gagcacggcc cggcttcggg 1380 tgcggggctc cgtacggggc gtggcgcggg gctcgccgtg ccgggcgggg ggtggcggca 1440 ggtgggggtg ccgggcgggg cggggccgcc tcgggccggg gagggctcgg gggaggggcg 1500 cggcggcccc cggagcgccg gcggctgtcg aggcgcggcg agccgcagcc attgcctttt 1560

		atggtaatcg tgcgagaggg cgcagggact tcctttgtcc caaatctgtg cggagccgaa 1620 atctgggagg cgccgccgca ccccctctag cgggcgcggg gcgaagcggt gcggcgccgg 1680 caggaaggaa atgggcgggg agggccttcg tgcgtcgccg cgccgccgtc cccttctccc 1740 tctccagcct cggggctgtc cgcgggggga cggctgcctt cgggggggac ggggcagggc 1800 ggggttcggc ttctggcgtg tgaccggcgg ctctagagcc tctgctaacc atgttcatgc 1860 cttcttcttt ttcctacagc tcctgggcaa cgtgctggtt attgtgctgt ctcatcattt 1920 tggcaaagaa ttcgccctgc agcgggaatt gccaccatgg ctcacgttcg gggactgcag 1980 ctgcctggat gtctggctct tgccgctctg tgtagcctgg tgcacagcca gcatgtgttt 2040 ctggctcctc agcaggccct gagcctgctg caaagagtta gaaggcacgg cgacggcagc 2100 ttcagcgacg agatgaatac cgtgctggtg gacaacctgg ccaccagaga cttcatcaac 2160 tggctgatcc agaccaagat caccgacggt ggtggcggag gcggaggatc tggtggcggt 2220 ggttctggcg gtggcggatc tgctgagttt acccctcctt gtcctccctg tcctgctcca 2280 gaactgctcg gcggaccttc cgtgttcctg tttcctccaa agcctaagga caccctgatg 2340 atcagcagaa cccctgaagt gacctgcgtg gtggtggacg tgtcccaaga ggatcctgag 2400 gtgcagttca attggtacgt ggacggcgtg gaagtgcaca acgcccagac aaagcccaga 2460 gagcggcagt tcaacagcac ctacagagtg gtgtccgtgc tgaccgtgac acaccaggat 2520 tggctgaacg gcaaagagta cacctgtaaa gtctccaaca agggcctgcc tgctcctatc 2580 gagaaaacca tcagcaaggc caagggccag cctagagaac cccaggtgta catcctgcct 2640 ccacctcaag aggaactgac caagaaccag gtgtccctga cctgtctggt caccggcttc 2700 tacccttccg atatcgccgt ggagtgggag agcaacggac agcccgagaa cacctacaag 2760 accacacctc cagtgctgga cagcgacggc tcttacctgc tgtactccaa gctgacagtg 2820 aacaagagcc ggtggcagcc cggcaacatc ttcacctgtt ctgtgatgca cgaggccctg 2880 cacaaccact acacccagaa aagcctgagc gtgtcccctg gataataagg tacctctaga 2940 gcggccgcct cgaggacggg gtgaactacg cctgaggatc cgatcttttt ccctctgcca 3000 aaaattatgg ggacatcatg aagccccttg agcatctgac ttctggctaa taaaggaaat 3060 ttattttcat tgcaatagtg tgttggaatt ttttgtgtct ctcactcgga agcaattcgt 3120 tgatctgaat ttcgaccacc cataataccc attaccctgg tagataagta gcatggcggg 3180 ttaatcatta actacaagga acccctagtg atggagttgg ccactccctc tctgcgcgct 3240 cgctcgctca ctgaggccgg gcgaccaaag gtcgcccgac gcccgggctt tgcccgggcg 3300 gcctcagtga gcgagcgagc gcgcag 3326
3-37	**Sequences**	
3-37-1	Sequence Number [ID]	37
3-37-2	Molecule Type	DNA
3-37-3	Length	4382
3-37-4	Features Location/	**misc_feature 1..4382**
	Qualifiers	note=synthetic construct
		source 1..4382
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-37-5	Residues	ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgtcg ggcgaccttt 60 ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggccaa ctccatcact 120 aggggttcct tgtagttaat gattaacccg ccatgctact tatctaccag ggtaatgggg 180 atcctctaga actatagcta gtcgacattg attattgact agttattaat agtaatcaat 240 tacggggtca ttagttcata gcccatatat ggagttccgc gttacataac ttacggtaaa 300 tggcccgcct ggctgaccgc ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt 360 tcccatagta acgccaatag ggactttcca ttgacgtcaa tgggtggagt atttacggta 420 aactgcccac ttggcagtac atcaagtgta tcatatgcca agtacgcccc ctattgacgt 480 caatgacggt aaatggcccg cctggcatta tgcccagtac atgaccttat gggactttcc 540 tacttggcag tacatctacg tattagtcat cgctattacc atggtcgagg tgagccccac 600 gttctgcttc actctcccca tctccccccc ctccccaccc ccaattttgt atttatttat 660 tttttaatta ttttgtgcag cgatgggggc ggggggggag gggggcgcg cgccaggcgg 720 ggcggggcgg ggcgaggggc ggggcggggc gaggcggaga ggtgcggcgg cagccaatca 780 gagcggcgcg ctccgaaagt ttccttttat ggcgaggcgg cggcggcggc ggccctataa 840 aaagcgaagc gcgcggcggg cgggagtcgc tgcgcgctgc cttcgccccg tgccccgctc 900 cgccgccgcc tcgcgccgcc cgccccggct ctgactgacc gcgttactcc cacaggtgag 960 cgggcgggac ggcccttctc ctccgggctg taattagcgc ttggtttaat gacggcttgt 1020 ttcttttctg tggctgcgtg aaagccttga ggggctccgg gagggccctt tgtgcggggg 1080 gagcggctcg gggggtgcgt gcgtgtgtgt gtgcgtgggg agcgccgcgt gcggctccgc 1140 gctgcccggc ggctgtgagc gctgcgggcg cggcgcgggg ctttgtgcgc tccgcagtgt 1200 gcgcgagggg agcgcggccg ggggcggtgc cccgcggtgc gggggggggct gcgaggggaa 1260 caaaggctgc gtgcggggtg tgtgcgtggg ggggtgagca gggggtgtgg gcgcgtcggt 1320 cgggctgcaa cccccccctgc acccccctcc ccgagttgct gagcacggcc cggcttcggg 1380 tgcggggctc cgtacggggc gtggcgcggg gctcgccgtg ccgggcgggg ggtggcggca 1440 ggtgggggtg ccgggcgggg cggggccgcc tcgggccggg gagggctcgg gggaggggcg 1500

		cggcggcccc cggagcgccg gcggctgtcg aggcgcggcg agccgcagcc attgcctttt 1560 atggtaatcg tgcgagaggg cgcagggact tcctttgtcc caaatctgtg cggagccgaa 1620 atctgggagg cgccgccgca ccccctctag cgggcgcggg gcgaagcggt gcggcgccgg 1680 caggaaggaa atgggcgggg agggccttcg tgcgtcgccg cgccgccgtc cccttctccc 1740 tctccagcct cggggctgtc cgcgggggga cggctgcctt cgggggggac ggggcagggc 1800 ggggttcggc ttctggcgtg tgaccggcgg ctctagagcc tctgctaacc atgttcatgc 1860 cttcttcttt ttcctacagc tcctgggcaa cgtgctggtt attgtgctgt ctcatcattt 1920 tggcaaagaa ttccagaagc caccatggct cacgttcggg gactgcagct gcctggatgt 1980 ctggctcttg ccgctctgtg tagcctggtg cacagccagc atgtgtttct ggctcctcag 2040 caagccctga gcctgctgca aagagttaga aggcacggcg acggcagctt cagcgacgag 2100 atgaataccg tgctggtgga caacctggcc accagagact tcatcaactg gctgatccag 2160 accaagatca ccggtggtgg cggaggcgga ggatctggtg gcggtggttc tggcggtggc 2220 ggatctgata cacacaagtc tgaggtggcc caccggttca aggacctggg cgaagaacac 2280 ttcaaaggcc tggtgctggt cgccttcagc cagtacctgc agcagtgccc tttcgaggaa 2340 cacgtgaagc tggtcaacga agtgaccgag ttcgccaaga cctgcgtggc cgacgagagc 2400 gccgagaact gtgataagag cctgcacacc ctgttcggcg acaagctgtg tacagtggcc 2460 acactgagag aaacctacgg cgagatggcc gactgctgcg ccaagcaaga gcccgagaga 2520 aacgagtgct tcctgcagca caaggacgac aaccccaacc tgcctccact cgtcagaccc 2580 gaagtggacg tgatgtgcac cgccttccac gacaatgagg ccaccttcct gaagaaatac 2640 ctgtacgagg tggccagacg gcacccctac ttttacgccc ctgaactgct gttctttgcc 2700 gccaggtaca aggccgcctt cgccgaatgt tgtcaggccg ctgataaggc cgcttgcctg 2760 ctgcctaagc tggacgagct tagagacgag ggcaaagcca gctccgccaa gcagagactg 2820 aagtgtgcca gcctgcagaa gttcggcgat agagccttta aggcctgggc cgtcgctaga 2880 ctgagccaga agtttcccaa ggccgagttt gccgaggtgt ccaagctcgt gaccgacctg 2940 acaaaggtgc acaccgagtg ctgtcacggc gacctgctgg aatgcgccga cgatagagcc 3000 gacctggcca agtacatgtg cgagaaccag gacagcatca gcagcaagct gaaagagtgc 3060 tgcgacaagc ctctgctgga aaagagccac tgtctggccg aggtggaaaa cgacgagatg 3120 cccgccgatc tgccttctct ggccgccgat tacgtggaaa gcaaggacgt gtgcaagaac 3180 tacgccgagg ccaaggacgt gttcctgggc atgtttctgt acgagtacgc ccgcagacac 3240 cccgactact ctgttatgct gctgctgaga ctggccaagg cctacgaggc cactctggaa 3300 aagtgttgtg ccgccgctga tccccacgag tgttacgcca aagtgttcga cgagttccag 3360 ccactggtgg aagaacccca gaacctggtc aagcagaact gcgagctgtt cgagcagctg 3420 ggcgagtaca agttccagaa cgccctgctc gtgcggtaca ccaagaaggt gccccaggtt 3480 tccacaccta cactggttga ggtgtcccgg aacctgggaa aagtgggcgc caagtgttgc 3540 aagctgcctg aggccaagag aatgccctgc gccgaggatt acctgagcgt ggtgctgaac 3600 agactgtgcg tgctgcacga gaaaacccct gtgtccgaga aagtgaccaa gtgctgtacc 3660 gagagcctgg tcaatcggag gccttgcttt agcgccctgg aactggacga ggcctacgtg 3720 cccaaggcct tcaacgccga gacattcacc ttccacgccg acatgtgtac cctgagcgag 3780 aaagaaaagc aagtgaagaa acagacagcc ctggtcgagc tggttaagca caagcctaag 3840 gccaccaaag aacaactgaa gggcgtgatg gacaacttcg ccgcctttgt ggaaaaatgc 3900 tgcaaggccg acgacaaaga ggcctgcttc gcagaagagg gccctaagtt tgtggccgcc 3960 tctcaagctg ctctggctta ataaggtacc tctagagcgg ccgcctcgag gacggggtga 4020 actacgcctg aggatccgat ctttttccct ctgccaaaaa ttatggggac atcatgaagc 4080 cccttgagca tctgacttct ggctaataaa ggaaatttat tttcattgca atagtgtgtt 4140 ggaatttttt gtgtctctca ctcggaagca attcgttgat ctgaatttcg accacccata 4200 atacccatta ccctggtaga taagtagcat ggcgggttaa tcattaacta caaggaaccc 4260 ctagtgatgg agttggccac tccctctctg cgcgctcgct cgctcactga ggccgggcga 4320 ccaaaggtcg cccgacgccc gggctttgcc cgggcggcct cagtgagcga gcgagcgcgc 4380 ag 4382
3-38	**Sequences**	
3-38-1	Sequence Number [ID]	38
3-38-2	Molecule Type	DNA
3-38-3	Length	5222
3-38-4	Features Location/	**misc_feature 1..5222**
	Qualifiers	note=synthetic construct
		source 1..5222
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-38-5	Residues	ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgtcg ggcgaccttt 60 ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggccaa ctccatcact 120 aggggttcct tgtagttaat gattaacccg ccatgctact tatctaccag ggtaatgggg 180 atcctctaga actatagcta gtcgacattg attattgact agttattaat agtaatcaat 240 tacggggtca ttagttcata gcccatatat ggagttccgc gttacataac ttacggtaaa 300 tggcccgcct ggctgaccgc ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt 360

tcccatagta acgccaatag ggactttcca ttgacgtcaa tgggtggagt atttacggta 420
aactgcccac ttggcagtac atcaagtgta tcatatgcca agtacgcccc ctattgacgt 480
caatgacggt aaatggcccg cctggcatta tgcccagtac atgaccttat gggactttcc 540
tacttggcag tacatctacg tattagtcat cgctattacc atggtcgagg tgagccccac 600
gttctgcttc actctcccca tctccccccc ctccccaccc ccaattttgt atttatttat 660
tttttaatta ttttgtgcag cgatgggggc gggggggggg ggggggcgcg cgccaggcgg 720
ggcggggcgg ggcgaggggc ggggcggggc gaggcggaga ggtgcggcgg cagccaatca 780
gagcggcgcg ctccgaaagt ttccttttat ggcgaggcgg cggcggcggc ggccctataa 840
aaagcgaagc gcgcggcggg cgggagtcgc tgcgcgctgc cttcgccccg tgccccgctc 900
cgccgccgcc tcgcgccgcc cgccccggct ctgactgacc gcgttactcc cacaggtgag 960
cgggcgggac ggcccttctc ctccgggctg taattagcgc ttggtttaat gacggcttgt 1020
ttcttttctg tggctgcgtg aaagccttga ggggctccgg gagggccctt tgtgcggggg 1080
gagcggctcg gggggtgcgt gcgtgtgtgt gtgcgtgggg agcgccgcgt gcggctccgc 1140
gctgcccggc ggctgtgagc gctgcgggcg cggcgcgggg ctttgtgcgc tccgcagtgt 1200
gcgcgagggg agcgcggccg ggggcggtgc cccgcggtgc gggggggggct gcgaggggaa 1260
caaaggctgc gtgcggggtg tgtgcgtggg ggggtgagca gggggtgtgg gcgcgtcggt 1320
cgggctgcaa ccccccctgc acccccctcc ccgagttgct gagcacggcc cggcttcggg 1380
tgcggggctc cgtacggggc gtggcgcggg gctcgccgtg ccgggcgggg ggtggcggca 1440
ggtgggggtg ccgggcgggg cggggccgcc tcgggccggg gagggctcgg gggaggggcg 1500
cggcggcccc cggagcgccg gcggctgtcg aggcgcggcg agccgcagcc attgcctttt 1560
atggtaatcg tgcgagaggg cgcagggact tcctttgtcc caaatctgtg cggagccgaa 1620
atctgggagg cgccgccgca ccccctctag cgggcgcggg gcgaagcggt gcggcgccgg 1680
caggaaggaa atgggcgggg agggccttcg tgcgtcgccg cgccgccgtc cccttctccc 1740
tctccagcct cggggctgtc cgcgggggga cggctgcctt cgggggggac ggggcagggc 1800
ggggttcggc ttctggcgtg tgaccggcgg ctctagagcc tctgctaacc atgttcatgc 1860
cttcttcttt ttcctacagc tcctgggcaa cgtgctggtt attgtgctgt ctcatcattt 1920
tggcaaagaa ttccagaagc caccatggct cacgttcggg gactgcagct gcctggatgt 1980
ctggctcttg ccgctctgtg tagcctggtg cacagccagc atgtgtttct ggctcctcag 2040
caagccctga gcctgctgca aagagttaga aggcacggcg acggcagctt cagcgacgag 2100
atgaataccg tgctggtgga caacctggcc accagagact tcatcaactg gctgatccag 2160
accaagatca ccgacggtgg cggaggcgga ggatctggtg gtggtggatc tggcggcgga 2220
ggaagttctc ctgctggcag ccctacaagc accgaggaag gcacaagcga gtctgccaca 2280
cctgagtctg gccctggcac atctacagag cctagcgaag gatctgcccc aggatctcct 2340
gccggctctc caacatctac cgaagaggga accagcaccg agccatctga gggatctgct 2400
cccggaacaa gcacagagcc ttcagaagga tccgctcctg gcacctctga aagcgccaca 2460
ccagaaagcg gacctggctc tgaacctgcc acaagcggat ctgagacacc tggaagcgag 2520
ccagccacat ctggcagcga aacacctgga tcaccagccg gatctcccac ctctaccgag 2580
gaagggacat ccgagagcgc taccccagaa tctggaccag gcaccagcac agaaccctct 2640
gaaggttcag cccctggaac ctctaccgaa ccatcagaag gcagcgctcc aggttctccc 2700
gctggatccc ctacatccac agaagagggc acctccactg aacctagcga gggaagtgct 2760
cccggcactt ccacagaacc atccgagggc agtgcacctg gaaccagcga atctgctacc 2820
cctgagagtg gacccggaac atccactgag ccctccgagg gttcagctcc aggcacatca 2880
gaatccgcca ctccagagtc cggaccagga tctgagccag ctaccagcgg ctctgaaaca 2940
cccggcacta gtaccgagcc aagcgagggt agcgcaccag ggacaagtac cgaaccgtct 3000
gagggctccg caccaggcac ttccgaaagt gctactcctg aaagcggccc aggcactagc 3060
gaatccgcaa cacccgagag cggtcctgga agtcctgcag gttcacctac cagcactgaa 3120
gaggggacta gcgagagcgc aactcctgaa tcaggccctg gatccgaacc tgctacctcc 3180
ggaagtgaaa cccctgggac aagcgaaagt gcaacgcccg agtcaggacc cgggactagc 3240
acggaaccca gtgaaggatc tgcacccggg acatctaccg agccgtcaga aggttctgct 3300
ccagggacta gtactgagcc ttccgaaggt tctgcacctg gaacttccac agagcccagt 3360
gaaggcagtg cccctggcac aagcactgaa ccgtccgaag gcagtgctcc cgggaccagt 3420
acagaaccga gcgagggctc tgctcctggt agtccagcag gatctccaac tagcaccgaa 3480
gaagggactt ccaccgagcc ttccgaggga agcgctcctg gaacatccga gtccgctacg 3540
ccagagagtg gcccaggttc tgaacccgct acttccggct cagagactcc tgggacttct 3600
gagtctgcaa ccccggaaag tggtcctggt agcgaaccag caactagcgg aagcgagaca 3660
cccggaacct cagagagtgc tacaccggaa tccggtccag ggacgtctac ggaaccgtct 3720
gaaggatcag ctcccggcac tagcgaaagc gctacacctg aaagtggtcc cggatctcca 3780
gcaggcagcc caacctctac tgaagaaggt tccccagctg gaagccccac ttccactgag 3840
gaaggctctc ccgcaggctc acccactagt acggaagaag gcacgtccga gtctgctact 3900
cccgaatccg gacctggaac tagcactgag ccaagcgaag gatcagcacc cggaacctct 3960
gagtccgcca caccagaatc tggtcctggt tccgagcctg ccacttcagg atcagaaacc 4020
ccgggcacga gtgaatcagc aacgccggaa tctggccccg gaagcgaacc ggctacgtct 4080
ggatctgaaa cgccagggac ctccgaatca gctacgcctg agtctggtcc agggacatcc 4140
accgaaccta gtgaaggctc cgcacctgga agccctgctg gaagcccaac gagtactgaa 4200
gagggcactt ctgagagcgc tacgcctgag tcaggacctg gaagcgaacc tgcaacatcc 4260

		ggctcagaaa caccagggac cagcgaaagc gcaaccccag agagtggacc tggatctcca 4320 gctggctctc ctactagtac agaggaaggc agccctgctg gctccccaac gtcaacagaa 4380 gaaggtacta gcacagagcc cagcgagggt tccgctccgg gaacttctga atctgctaca 4440 cccgagtcag gtcctggtac aagcgagtca gctacgcccg aaagtggacc tggcacctca 4500 gagtctgcaa ctcctgagag cggtccagga tcagaaccag ccacctctgg ctctgagaca 4560 ccaggttctg agcctgcaac gtccggaagc gaaacaccag gcagtcctgc cggaagtcct 4620 acttcaaccg aagaggggac ctctacagag ccatcagagg gctctgcacc gggcacctca 4680 acagaaccat ctgaaggatc cgcaccgggc tctgagcctg ctactagtgg aagcgaaact 4740 cctggcacca gtgaatccgc tactcccgag tctggcccgg gaacgtctac tgaaccatct 4800 gagggaagtg ccccaggcta ataaggtacc tctagagcgg ccgcctcgag gacggggtga 4860 actacgcctg aggatccgat ctttttccct ctgccaaaaa ttatggggac atcatgaagc 4920 cccttgagca tctgacttct ggctaataaa ggaaatttat tttcattgca atagtgtgtt 4980 ggaatttttt gtgtctctca ctcggaagca attcgttgat ctgaatttcg accacccata 5040 atacccatta ccctggtaga taagtagcat ggcgggttaa tcattaacta caaggaaccc 5100 ctagtgatgg agttggccac tccctctctg cgcgctcgct cgctcactga ggccgggcga 5160 ccaaaggtcg cccgacgccc gggctttgcc cgggcggcct cagtgagcga gcgagcgcgc 5220 ag 5222
3-39	**Sequences**	
3-39-1	Sequence Number [ID]	39
3-39-2	Molecule Type	DNA
3-39-3	Length	3611
3-39-4	Features Location/	**misc_feature 1..3611**
	Qualifiers	note=synthetic construct
		source 1..3611
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-39-5	Residues	ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgtcg ggcgaccttt 60 ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggccaa ctccatcact 120 aggggttcct tgtagttaat gattaacccg ccatgctact tatctacgta gccatgctct 180 agtacgcgtg ctagctaatg atgggcgctc gagtaatgat gggcggtcga ctaatgatgg 240 gcgctcgagt aatgatgggc gtctagctaa tgatgggcgc tcgagtaatg atgggcggtc 300 gactaatgat gggcgctcga gtaatgatgg gcgtctagct aatgatgggc gctcgagtaa 360 tgatgggcgg tcgactaatg atgggcgctc gagtaatgat gggcgtctag aacgcgaatt 420 aattcaacat tttgacaccc ccataatatt tttccagaat taacagtata aattgcatct 480 cttgttcaag agttccctat cactctcttt aatcactact cacagtaacc tcaactcctg 540 ccacaagctt gccctgcagc gggaattcgc cctgcagcgg gaattgccac catggctcac 600 gttcggggac tgcagctgcc tggatgtctg gctcttgccg ctctgtgtag cctggtgcac 660 agccagcatg tgtttctggc tcctcagcag gccctgagcc tgctgcaaag agttagaagg 720 cacggcgacg gcagcttcag cgacgagatg aataccgtgc tggtggacaa cctggccacc 780 agagacttca tcaactggct gatccagacc aagatcaccg acggtggtgg cggaggcgga 840 ggatctggtg gcggtggttc tggcggtggc ggatctgctg agtttacccc tccttgtcct 900 ccctgtcctg ctccagaact gctcggcgga ccttccgtgt tcctgtttcc tccaaagcct 960 aaggacaccc tgatgatcag cagaacccct gaagtgacct gcgtggtggt ggacgtgtcc 1020 caagaggatc ctgaggtgca gttcaattgg tacgtggacg gcgtggaagt gcacaacgcc 1080 cagacaaagc ccagagagcg gcagttcaac agcacctaca gagtggtgtc cgtgctgacc 1140 gtgacacacc aggattggct gaacggcaaa gagtacacct gtaaagtctc caacaagggc 1200 ctgcctgctc ctatcgagaa aaccatcagc aaggccaagg gccagcctag agaaccccag 1260 gtgtacatcc tgcctccacc tcaagaggaa ctgaccaaga accaggtgtc cctgacctgt 1320 ctggtcaccg gcttctaccc ttccgatatc gccgtggagt gggagagcaa cggacagccc 1380 gagaacacct acaagaccac acctccagtg ctggacagcg acggctctta cctgctgtac 1440 tccaagctga cagtgaacaa gagccggtgg cagcccggca acatcttcac ctgttctgtg 1500 atgcacgagg ccctgcacaa ccactacacc cagaaaagcc tgagcgtgtc ccctggataa 1560 taaggtaccg atctttttcc ctctgccaaa aattatgggg acatcatgaa gccccttgag 1620 catctgactt ctggctaata aaggaaattt attttcattg caatagtgtg ttggaatttt 1680 ttgtgtctct cactcggaag gcgcgaccgg ttgggcatgg ccaggtagcc tatgctgtgt 1740 ctggacgtcc tcctgctggt atagttattt taaaatcaga aggacaggga agggagcagt 1800 ggttcacgcc tgtaatccca gcaatttggg aggccaaggt gggtagatca cctgagatta 1860 ggagttggag accagcctgg ccaatatggt gaaaccccgt ctctaccaaa aaaacaaaaa 1920 ttagctgagc ctggtcatgc atgcctggaa tcccaacaac tcgggaggct gaggcaggag 1980 aatcgcttga acccaggagg cggagattgc agtgagccaa gattgtgcca ctgcactcat 2040 cgattccaat tcagcggggg ccacctgatg tcccggccag cagaggaagc aacgcgggga 2100 ccacgggtta acccgggtgc gcggcgtcgg tggtgccggc ggggggcgcc aggtcgcagg 2160 cggtgtaggg ctccaggcag gcggcgaagg ccatgacgtg cgctatgaag gtctgctcct 2220 gcacgccgtg aaccaggtgc gcctgcgggc cgcgcgcgaa caccgccacg tcctcgcctg 2280

		cgtgggtctc ttcgtccagg ggcactgctg actgctgccg atactcgggg ctcccgctct 2340 cgctctcggt aacatccggc cgggcgccgt ccttgagcac atagcctgga ccgtttccgt 2400 ataggaggac cgtgtaggcc ttcctgtccc gggccttgcc aggggccagc ccgaagatgg 2460 agctccctcg caggggtag cctccgaagg agaagacgtg ggagtggtcg gcagtgacga 2520 ggctcagcgt gtcctcctcg ctggtgagct ggcccgccct ctcaatggcg tcgtcgaaca 2580 tgatcgtctc agtcagtgcc cggtaagccc tgctttcatg atgaccatgg tcgatgcgac 2640 caccctccac gaagaggaag aagccgcggg ggttcctgct cagcaggcgc agggcagcct 2700 ctgtcatctc catcagggag gggtccagtg tggagtctcg gtggatctcg tatttcatgt 2760 ctccaggctc aaagagaccc atgagatggg tcacagacgg gtccagggaa gcctgcatga 2820 gctcagtgcg gttccacaca taccgggcac cctggcgctt cgccagccat tcctgcacca 2880 gattcttccc gtccagcctg gtcccacctt ggctgtagtc atctgggtac tcagggtctg 2940 gggttcccat gcggaacatg tactttcggc ctccacctag gatcacgtca atgtccatgt 3000 tggagatgag ctgcgtagcg atgtcctggc acccctcctg gcgggccgag gcaggcacgt 3060 cggcgtccga gtaccagttg cggttcaccg tgtgggcgta ggtgccggct ggcgaggcgt 3120 gctgcactcg tgtggtggtt accactccca ctgacttccc tgctttcttg gcccgattca 3180 tcacggagat gacctcgttg ccgcgtgtcg tgttgcactg gttaaagcgg gcggctgcac 3240 tcaagccaat ggtctggaag ttgcccttga ccccgcacag gtaggccgtg gctgtggctc 3300 cactgtctgg cacatgtttg tctacattgt atgtcttgga cagagccaca tatgggaagc 3360 ggtccatggc caggggtatc tcaggcccca gtttgtcctt cttctgccct tttaggatct 3420 acgtcgagac tagagcatgg ctacgtagat aagtagcatg gcgggttaat cattaactac 3480 aaggaacccc tagtgatgga gttggccact ccctctctgc gcgctcgctc gctcactgag 3540 gccgggcgac caaaggtcgc ccgacgcccg ggctttgccc gggcggcctc agtgagcgag 3600 cgagcgcgca g 3611
3-40	**Sequences**	
3-40-1	Sequence Number [ID]	40
3-40-2	Molecule Type	DNA
3-40-3	Length	4667
3-40-4	Features Location/	**misc_feature 1..4667**
	Qualifiers	note=synthetic construct
		source 1..4667
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-40-5	Residues	ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgtcg ggcgaccttt 60 ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggccaa ctccatcact 120 aggggttcct tgtagttaat gattaacccg ccatgctact tatctacgta gccatgctct 180 agtacgcgtg ctagctaatg atgggcgctc gagtaatgat gggcggtcga ctaatgatgg 240 gcgctcgagt aatgatgggc gtctagctaa tgatgggcgc tcgagtaatg atgggcggtc 300 gactaatgat gggcgctcga gtaatgatgg gcgtctagct aatgatgggc gctcgagtaa 360 tgatgggcgg tcgactaatg atgggcgctc gagtaatgat gggcgtctag aacgcgaatt 420 aattcaacat tttgacaccc ccataatatt tttccagaat taacagtata aattgcatct 480 cttgttcaag agttccctat cactctcttt aatcactact cacagtaacc tcaactcctg 540 ccacaagctt gccctgcagc gggaattcca gaagccacca tggctcacgt tcggggactg 600 cagctgcctg gatgtctggc tcttgccgct ctgtgtagcc tggtgcacag ccagcatgtg 660 tttctggctc ctcagcaagc cctgagcctg ctgcaaagag ttagaaggca cggcgacggc 720 agcttcagcg acgagatgaa taccgtgctg gtggacaacc tggccaccag agacttcatc 780 aactggctga tccagaccaa gatcaccggt ggtggcggag gcggaggatc tggtggcggt 840 ggttctggcg gtggcggatc tgatacacac aagtctgagg tggcccaccg gttcaaggac 900 ctgggcgaag aacacttcaa aggcctggtg ctggtcgcct tcagccagta cctgcagcag 960 tgccctttcg aggaacacgt gaagctggtc aacgaagtga ccgagttcgc caagacctgc 1020 gtggccgacg agagcgccga gaactgtgat aagagcctgc acaccctgtt cggcgacaag 1080 ctgtgtacag tggccacact gagagaaacc tacggcgaga tggccgactg ctgcgccaag 1140 caagagcccg agagaaacga gtgcttcctg cagcacaagg acgacaaccc caacctgcct 1200 ccactcgtca gacccgaagt ggacgtgatg tgcaccgcct tccacgacaa tgaggccacc 1260 ttcctgaaga aatacctgta cgaggtggcc agacggcacc cctactttta cgcccctgaa 1320 ctgctgttct ttgccgccag gtacaaggcc gccttcgccg aatgttgtca ggccgctgat 1380 aaggccgctt gcctgctgcc taagctggac gagcttagag acgagggcaa agccagctcc 1440 gccaagcaga gactgaagtg tgccagcctg cagaagttcg gcgatagagc ctttaaggcc 1500 tgggccgtcg ctagactgag ccagaagttt cccaaggccg agtttgccga ggtgtccaag 1560 ctcgtgaccg acctgacaaa ggtgcacacc gagtgctgtc acggcgacct gctggaatgc 1620 gccgacgata gagccgacct ggccaagtac atgtgcgaga accaggacag catcagcagc 1680 aagctgaaag agtgctgcga caagcctctg ctggaaaaga gccactgtct ggccgaggtg 1740 gaaaacgacg agatgcccgc cgatctgcct tctctggccg ccgattacgt ggaaagcaag 1800 gacgtgtgca agaactacgc cgaggccaag gacgtgttcc tgggcatgtt tctgtacgag 1860 tacgcccgca gacaccccga ctactctgtt atgctgctgc tgagactggc caaggcctac 1920

		gaggccactc tggaaaagtg ttgtgccgcc gctgatcccc acgagtgtta cgccaaagtg 1980
		ttcgacgagt tccagccact ggtggaagaa ccccagaacc tggtcaagca gaactgcgag 2040
		ctgttcgagc agctgggcga gtacaagttc cagaacgccc tgctcgtgcg gtacaccaag 2100
		aaggtgcccc aggtttccac acctacactg gttgaggtgt cccggaacct gggaaaagtg 2160
		ggcgccaagt gttgcaagct gcctgaggcc aagagaatgc cctgcgccga ggattacctg 2220
		agcgtggtgc tgaacagact gtgcgtgctg cacgagaaaa cccctgtgtc cgagaaagtg 2280
		accaagtgct gtaccgagag cctggtcaat cggaggcctt gctttagcgc cctggaactg 2340
		gacgaggcct acgtgcccaa ggccttcaac gccgagacat tcaccttcca cgccgacatg 2400
		tgtaccctga gcgagaaaga aaagcaagtg aagaaacaga cagccctggt cgagctggtt 2460
		aagcacaagc ctaaggccac caaagaacaa ctgaagggcg tgatggacaa cttcgccgcc 2520
		tttgtggaaa aatgctgcaa ggccgacgac aaagaggcct gcttcgcaga agagggccct 2580
		aagtttgtgg ccgcctctca agctgctctg gcttaataag gtaccgatct ttttccctct 2640
		gccaaaaatt atggggacat catgaagccc cttgagcatc tgacttctgg ctaataaagg 2700
		aaatttattt tcattgcaat agtgtgttgg aatttttgt gtctctcact cggaaggcgc 2760
		gaccggttgg gcatggccag gtagcctatg ctgtgtctgg acgtcctcct gctggtatag 2820
		ttattttaaa atcagaagga cagggaaggg agcagtggtt cacgcctgta atcccagcaa 2880
		tttgggaggc caaggtgggt agatcacctg agattaggag ttggagacca gcctggccaa 2940
		tatggtgaaa ccccgtctct accaaaaaaa caaaaattag ctgagcctgg tcatgcatgc 3000
		ctggaatccc aacaactcgg gaggctgagg caggagaatc gcttgaaccc aggaggcgga 3060
		gattgcagtg agccaagatt gtgccactgc actcatcgat tccaattcag cgggggccac 3120
		ctgatgtccc ggccagcaga ggaagcaacg cggggaccac gggttaaccc gggtgcgcgg 3180
		cgtcggtggt gccggcgggg ggcgccaggt cgcaggcggt gtagggctcc aggcaggcgg 3240
		cgaaggccat gacgtgcgct atgaaggtct gctcctgcac gccgtgaacc aggtgcgcct 3300
		gcgggccgcg cgcgaacacc gccacgtcct cgcctgcgtg ggtctcttcg tccaggggca 3360
		ctgctgactg ctgccgatac tcggggctcc cgctctcgct ctcggtaaca tccggccggg 3420
		cgccgtcctt gagcacatag cctggaccgt ttccgtatag gaggaccgtg taggccttcc 3480
		tgtcccgggc cttgccaggg gccagcccga agatggagct ccctcgcagg gggtagcctc 3540
		cgaaggagaa gacgtgggag tggtcggcag tgacgaggct cagcgtgtcc tcctcgctgg 3600
		tgagctggcc cgccctctca atggcgtcgt cgaacatgat cgtctcagtc agtgcccggt 3660
		aagccctgct ttcatgatga ccatggtcga tgcgaccacc ctccacgaag aggaagaagc 3720
		cgcgggggtt cctgctcagc aggcgcaggg cagcctctgt catctccatc agggaggggt 3780
		ccagtgtgga gtctcggtgg atctcgtatt tcatgtctcc aggctcaaag agacccatga 3840
		gatgggtcac agacgggtcc agggaagcct gcatgagctc agtgcggttc cacacatacc 3900
		gggcaccctg gcgcttcgcc agccattcct gcaccagatt cttcccgtcc agcctggtcc 3960
		caccttggct gtagtcatct gggtactcag ggtctggggt tcccatgcgg aacatgtact 4020
		ttcggcctcc acctaggatc acgtcaatgt ccatgttgga gatgagctgc gtagcgatgt 4080
		cctggcaccc ctcctggcgg gccgaggcag gcacgtcggc gtccgagtac cagttgcggt 4140
		tcaccgtgtg ggcgtaggtg ccggctggcg aggcgtgctg cactcgtgtg gtggttacca 4200
		ctcccactga cttccctgct ttcttggccc gattcatcac ggagatgacc tcgttgccgc 4260
		gtgtcgtgtt gcactggtta aagcgggcgg ctgcactcaa gccaatggtc tggaagttgc 4320
		ccttgacccc gcacaggtag gccgtggctg tggctccact gtctggcaca tgtttgtcta 4380
		cattgtatgt cttggacaga gccacatatg ggaagcggtc catggccagg ggtatctcag 4440
		gccccagttt gtccttcttc tgccctttta ggatctacgt cgagactaga gcatggctac 4500
		gtagataagt agcatggcgg gttaatcatt aactacaagg aacccctagt gatggagttg 4560
		gccactccct ctctgcgcgc tcgctcgctc actgaggccg ggcgaccaaa ggtcgcccga 4620
		cgcccgggct ttgcccgggc ggcctcagtg agcgagcgag cgcgcag 4667
3-41	**Sequences**	
3-41-1	Sequence Number [ID]	41
3-41-2	Molecule Type	DNA
3-41-3	Length	3611
3-41-4	Features Location/	**misc_feature 1..3611**
	Qualifiers	note=synthetic construct
		source 1..3611
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-41-5	Residues	ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgtcg ggcgaccttt 60
		ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggccaa ctccatcact 120
		aggggttcct tgtagttaat gattaacccg ccatgctact tatctacgta gccatgctct 180
		agtacgcgtg ctagctaatg atgggcgctc gagtaatgat gggcggtcga ctaatgatgg 240
		gcgctcgagt aatgatgggc gtctagctaa tgatgggcgc tcgagtaatg atgggcggtc 300
		gactaatgat gggcgctcga gtaatgatgg gcgtctagct aatgatgggc gctcgagtaa 360
		tgatgggcgg tcgactaatg atgggcgctc gagtaatgat gggcgtctag aacgcgaatt 420
		aattcaacat tttgacaccc ccataatatt tttccagaat taacagtata aattgcatct 480
		cttgttcaag agttccctat cactctcttt aatcactact cacagtaacc tcaactcctg 540

		ccacaagctt gccctgcagc gggaattcgc cctgcagcgg gaattgccac catggctcac 600 gttcggggac tgcagctgcc tggatgtctg gctcttgccg ctctgtgtag cctggtgcac 660 agccagcacg tgtttctggc tcctcagcaa gccagatcac tgctgcagag agttagaagg 720 cacggcgacg gcagcttcag cgacgagatg aacaccatcc tggacaacct ggccgccaga 780 gacttcatca actggctgat ccagaccaag atcaccgacg gtggcggagg cggaggatct 840 ggtggtggtg gatctggcgg cggaggttct gccgagtcta agtacggacc tccttgtcct 900 ccctgtcctg ctccagaagc tgctggcggc ccatccgtgt ttctgttccc tccaaagcct 960 aaggacaccc tgatgatcag cagaacccct gaagtgacct gcgtggtggt cgacgtgtcc 1020 caagaggatc ctgaggtgca gttcaattgg tacgtggacg gcgtggaagt gcacaacgcc 1080 aagaccaagc ctagagagga acagttcaac agcacctaca gagtggtgtc cgtgctgacc 1140 gtgctgcacc aggattggct gaacggcaaa gagtacaagt gcaaggtgtc caacaagggc 1200 ctgcctagca gcatcgagaa aaccatcagc aaggccaagg gccagccaag agaaccccag 1260 gtgtacacac tgcctccaag ccaagaggaa atgaccaaga accaggtgtc cctgacctgc 1320 ctggtcaagg gcttctaccc ttccgatatc gccgtggaat gggagagcaa cggccagcct 1380 gagaacaact acaagaccac acctcctgtg ctggactccg atggctcatt cttcctgtac 1440 agcagactga ccgtggacaa gagcaggtgg caagagggca acgtgttcag ctgcagcgtg 1500 atgcacgagg ccctgcacaa ccactacacc cagaaaagcc tgagcctgtc tctgggctaa 1560 taaggtaccg atctttttcc ctctgccaaa aattatgggg acatcatgaa gccccttgag 1620 catctgactt ctggctaata aaggaaattt attttcattg caatagtgtg ttggaatttt 1680 ttgtgtctct cactcggaag gcgcgaccgg ttgggcatgg ccaggtagcc tatgctgtgt 1740 ctggacgtcc tcctgctggt atagttattt taaaatcaga aggacaggga agggagcagt 1800 ggttcacgcc tgtaatccca gcaatttggg aggccaaggt gggtagatca cctgagatta 1860 ggagttggag accagcctgg ccaatatggt gaaaccccgt ctctaccaaa aaaacaaaaa 1920 ttagctgagc ctggtcatgc atgcctggaa tcccaacaac tcgggaggct gaggcaggag 1980 aatcgcttga acccaggagg cggagattgc agtgagccaa gattgtgcca ctgcactcat 2040 cgattccaat tcagcggggg ccacctgatg tcccggccag cagaggaagc aacgcgggga 2100 ccacgggtta acccgggtgc gcggcgtcgg tggtgccggc ggggggcgcc aggtcgcagg 2160 cggtgtaggg ctccaggcag gcggcgaagg ccatgacgtg cgctatgaag gtctgctcct 2220 gcacgccgtg aaccaggtgc gcctgcgggc cgcgcgcgaa caccgccacg tcctcgcctg 2280 cgtgggtctc ttcgtccagg ggcactgctg actgctgccg atactcgggg ctcccgctct 2340 cgctctcggt aacatccggc cgggcgccgt ccttgagcac atagcctgga ccgtttccgt 2400 ataggaggac cgtgtaggcc ttcctgtccc gggccttgcc aggggccagc ccgaagatgg 2460 agctccctcg caggggtag cctccgaagg agaagacgtg ggagtggtcg gcagtgacga 2520 ggctcagcgt gtcctcctcg ctggtgagct ggcccgccct ctcaatggcg tcgtcgaaca 2580 tgatcgtctc agtcagtgcc cggtaagccc tgctttcatg atgaccatgg tcgatgcgac 2640 caccctccac gaagaggaag aagccgcggg ggttcctgct cagcaggcgc agggcagcct 2700 ctgtcatctc catcagggag gggtccagtg tggagtctcg gtggatctcg tatttcatgt 2760 ctccaggctc aaagagaccc atgagatggg tcacagacgg gtccagggaa gcctgcatga 2820 gctcagtgcg gttccacaca taccgggcac cctggcgctt cgccagccat tcctgcacca 2880 gattcttccc gtccagcctg gtcccacctt ggctgtagtc atctgggtac tcagggtctg 2940 gggttcccat gcggaacatg tactttcggc ctccacctag gatcacgtca atgtccatgt 3000 tggagatgag ctgcgtagcg atgtcctggc acccctcctg gcgggccgag gcaggcacgt 3060 cggcgtccga gtaccagttg cggttcaccg tgtgggcgta ggtgccggct ggcgaggcgt 3120 gctgcactcg tgtggtggtt accactccca ctgacttccc tgctttcttg gcccgattca 3180 tcacggagat gacctcgttg ccgcgtgtcg tgttgcactg gttaaagcgg gcggctgcac 3240 tcaagccaat ggtctggaag ttgcccttga ccccgcacag gtaggccgtg gctgtggctc 3300 cactgtctgg cacatgtttg tctacattgt atgtcttgga cagagccaca tatgggaagc 3360 ggtccatggc caggggtatc tcaggcccca gtttgtcctt cttctgccct tttaggatct 3420 acgtcgagac tagagcatgg ctacgtagat aagtagcatg gcgggttaat cattaactac 3480 aaggaacccc tagtgatgga gttggccact ccctctctgc gcgctcgctc gctcactgag 3540 gccgggcgac caaaggtcgc ccgacgcccg ggctttgccc gggcggcctc agtgagcgag 3600 cgagcgcgca g 3611
3-42	**Sequences**	
3-42-1	Sequence Number [ID]	42
3-42-2	Molecule Type	DNA
3-42-3	Length	4664
3-42-4	Features Location/	**misc_feature 1..4664**
	Qualifiers	note=synthetic construct
		source 1..4664
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-42-5	Residues	ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgtcg ggcgaccttt 60 ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggccaa ctccatcact 120 aggggttcct tgtagttaat gattaacccg ccatgctact tatctacgta gccatgctct 180

agtacgcgtg ctagctaatg atgggcgctc gagtaatgat gggcggtcga ctaatgatgg 240
gcgctcgagt aatgatgggc gtctagctaa tgatgggcgc tcgagtaatg atgggcggtc 300
gactaatgat gggcgctcga gtaatgatgg gcgtctagct aatgatgggc gctcgagtaa 360
tgatgggcgg tcgactaatg atgggcgctc gagtaatgat gggcgtctag aacgcgaatt 420
aattcaacat tttgacaccc ccataatatt tttccagaat taacagtata aattgcatct 480
cttgttcaag agttccctat cactctcttt aatcactact cacagtaacc tcaactcctg 540
ccacaagctt gccctgcagc gggaattcca gaagccacca tggctcacgt tcggggactg 600
cagctgcctg gatgtctggc tcttgccgct ctgtgtagcc tggtgcacag ccagcacgtg 660
tttctggctc ctcagcaagc cagatcactg ctgcagagag ttagaaggca cggcgacggc 720
agcttcagcg acgagatgaa caccatcctg gacaacctgg ccgccagaga cttcatcaac 780
tggctgatcc agaccaagat caccgacggt ggcggaggcg gaggatctgg tggtggtgga 840
tctggcggcg gaggttctga cgcccacaaa tctgaagtgg cccaccggtt caaggacctg 900
ggcgaagaga atttcaaggc cctggtgctg atcgccttcg ctcagtacct gcagcagtgc 960
cccttcgagg accacgtgaa gctggtcaac gaagtgaccg agttcgccaa gacctgcgtg 1020
gccgacgaga gcgccgagaa ctgtgataag agcctgcaca ccctgttcgg cgacaagctg 1080
tgtacagtgg ccacactgag agaaacctac ggcgagatgg ccgactgctg cgccaagcaa 1140
gagcccgaga gaaacgagtg cttcctgcag cacaaggacg acaaccccaa cctgcctaga 1200
ctcgtgcggc ctgaagtgga cgtgatgtgc accgccttcc acgacaacga ggaaaccttc 1260
ctgaagaagt acctgtacga gatcgccaga cggcacccct acttttacgc ccctgagctg 1320
ctgttcttcg ccaagcggta taaggccgcc ttcaccgagt gttgtcaggc cgctgataag 1380
gctgcctgcc tgctgcctaa gctggacgag cttagagacg agggcaaagc cagctccgcc 1440
aagcagagac tgaagtgtgc cagcctgcag aagttcggcg agagagcctt taaggcctgg 1500
gccgttgcta gactgagcca gagatttccc aaggccgagt ttgccgaggt gtccaagctc 1560
gtgaccgacc tgacaaaggt gcacaccgag tgctgccacg gcgacctgct ggaatgcgcc 1620
gacgatagag ccgacctggc caagtacatc tgcgagaacc aggacagcat cagcagcaag 1680
ctgaaagagt gctgcgagaa gcctctgctg gaaaagagcc actgtatcgc cgaggtggaa 1740
aacgacgaga tgcccgccga tctgccttct ctggccgccg attttgtgga aagcaaggac 1800
gtgtgcaaga actacgccga ggccaaggac gtgttcctgg gcatgtttct gtacgagtac 1860
gcccgcagac accccgacta ctctgttgtg ctgctgctga gactggccaa aacctacgag 1920
acaaccctgg aaaagtgctg tgccgccgct gatcctcacg agtgttacgc caaggtgttc 1980
gacgagttca agccactggt ggaagaaccc cagaacctga tcaagcagaa ctgcgagctg 2040
ttcgagcagc tgggcgagta caagttccag aacgccctgc tcgtgcggta caccaagaag 2100
gtgccccagg tttccacacc tacactggtt gaggtgtccc ggaacctggg caaagtgggc 2160
agcaagtgtt gcaagcaccc tgaggccaag agaatgccct gcgccgagga ttacctgagc 2220
gtggtgctga atcagctgtg cgtgctgcac gagaaaaccc ctgtgtccga cagagtgacc 2280
aagtgctgta ccgagagcct ggtcaacaga cggccttgct ttagcgccct cgaggtggac 2340
gagacatacg tgcccaaaga gttcaacgcc gagacattca ccttccacgc cgacatctgt 2400
accctgagcg agaaagagcg gcagatcaag aaacagactg ccctggtgga actggtcaag 2460
cacaagccca aggccaccaa agaacagctg aaggccgtga tggacgactt cgccgccttc 2520
gtggaaaagt gctgcaaggc cgacgacaaa gagacctgct tcgccgaaga gggcaagaaa 2580
ctggtggccg cttctcaggc tgctctggga taataaggta ccgatctttt tccctctgcc 2640
aaaaattatg gggacatcat gaagcccctt gagcatctga cttctggcta ataaaggaaa 2700
tttattttca ttgcaatagt gtgttggaat tttttgtgtc tctcactcgg aaggcgcgac 2760
cggttgggca tggccaggta gcctatgctg tgtctggacg tcctcctgct ggtatagtta 2820
ttttaaaatc agaaggacag ggaagggagc agtggttcac gcctgtaatc ccagcaattt 2880
gggaggccaa ggtgggtaga tcacctgaga ttaggagttg gagaccagcc tggccaatat 2940
ggtgaaaccc cgtctctacc aaaaaaacaa aaattagctg agcctggtca tgcatgcctg 3000
gaatcccaac aactcgggag gctgaggcag gagaatcgct tgaacccagg aggcggagat 3060
tgcagtgagc caagattgtg ccactgcact catcgattcc aattcagcgg gggccacctg 3120
atgtcccggc cagcagagga agcaacgcgg ggaccacggg ttaacccggg tgcgcggcgt 3180
cggtggtgcc ggcgggggcc gccaggtcgc aggcggtgta gggctccagg caggcggcga 3240
aggccatgac gtgcgctatg aaggtctgct cctgcacgcc gtgaaccagg tgcgcctgcg 3300
ggccgcgcgc gaacaccgcc acgtcctcgc ctgcgtgggt ctcttcgtcc aggggcactg 3360
ctgactgctg ccgatactcg gggctcccgc tctcgctctc ggtaacatcc ggccgggcgc 3420
cgtccttgag cacatagcct ggaccgtttc cgtataggag gaccgtgtag gccttcctgt 3480
cccgggcctt gccaggggcc agcccgaaga tggagctccc tcgcaggggg tagcctccga 3540
aggagaagac gtgggagtgg tcggcagtga cgaggctcag cgtgtcctcc tcgctggtga 3600
gctggcccgc cctctcaatg gcgtcgtcga acatgatcgt ctcagtcagt gcccggtaag 3660
ccctgctttc atgatgacca tggtcgatgc gaccaccctc cacgaagagg aagaagccgc 3720
gggggttcct gctcagcagg cgcagggcag cctctgtcat ctccatcagg gaggggtcca 3780
gtgtggagtc tcggtggatc tcgtatttca tgtctccagg ctcaaagaga cccatgagat 3840
gggtcacaga cgggtccagg gaagcctgca tgagctcagt gcggttccac acataccggg 3900
caccctggcg cttcgccagc cattcctgca ccagattctt cccgtccagc ctggtcccac 3960
cttggctgta gtcatctggg tactcagggt ctggggttcc catgcggaac atgtactttc 4020
ggcctccacc taggatcacg tcaatgtcca tgttggagat gagctgcgta gcgatgtcct 4080

		ggcacccctc ctggcgggcc gaggcaggca cgtcggcgtc cgagtaccag ttgcggttca 4140 ccgtgtgggc gtaggtgccg gctggcgagg cgtgctgcac tcgtgtggtg gttaccactc 4200 ccactgactt ccctgctttc ttggcccgat tcatcacgga gatgacctcg ttgccgcgtg 4260 tcgtgttgca ctggttaaag cgggcggctg cactcaagcc aatggtctgg aagttgccct 4320 tgaccccgca caggtaggcc gtggctgtgg ctccactgtc tggcacatgt ttgtctacat 4380 tgtatgtctt ggacagagcc acatatggga agcggtccat ggccaggggt atctcaggcc 4440 ccagtttgtc cttcttctgc ccttttagga tctacgtcga gactagagca tggctacgta 4500 gataagtagc atggcgggtt aatcattaac tacaaggaac ccctagtgat ggagttggcc 4560 actccctctc tgcgcgctcg ctcgctcact gaggccgggc gaccaaaggt cgcccgacgc 4620 ccgggctttg cccgggcggc ctcagtgagc gagcgagcgc gcag 4664
3-43	**Sequences**	
3-43-1	Sequence Number [ID]	43
3-43-2	Molecule Type	DNA
3-43-3	Length	2911
3-43-4	Features Location/	**misc_feature 1..2911**
	Qualifiers	note=synthetic construct
		source 1..2911
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-43-5	Residues	ctagtcgaca ttgattattg actagttatt aatagtaatc aattacgggg tcattagttc 60 atagcccata tatggagttc cgcgttacat aacttacggt aaatggcccg cctggctgac 120 cgcccaacga cccccgccca ttgacgtcaa taatgacgta tgttcccata gtaacgccaa 180 tagggacttt ccattgacgt caatgggtgg agtatttacg gtaaactgcc cacttggcag 240 tacatcaagt gtatcatatg ccaagtacgc cccctattga cgtcaatgac ggtaaatggc 300 ccgcctggca ttatgcccag tacatgacct tatgggactt tcctacttgg cagtacatct 360 acgtattagt catcgctatt accatggtcg aggtgagccc cacgttctgc ttcactctcc 420 ccatctcccc cccctcccca cccccaattt tgtatttatt tatttttaa ttattttgtg 480 cagcgatggg ggcgggggg gggggggggc gcgcgccagg cggggcgggg cggggcgagg 540 ggcggggcgg ggcgaggcgg agaggtgcgg cggcagccaa tcagagcggc gcgctccgaa 600 agtttccttt tatggcgagg cggcggcggc ggcggcccta taaaaagcga agcgcgcggc 660 gggcgggagt cgctgcgcgc tgccttcgcc ccgtgccccg ctccgccgcc gcctcgcgcc 720 gcccgccccg gctctgactg accgcgttac tcccacaggt gagcgggcgg gacggccctt 780 ctcctccggg ctgtaattag cgcttggttt aatgacggct tgtttctttt ctgtggctgc 840 gtgaaagcct tgaggggctc cgggagggcc ctttgtgcgg ggggagcggc tcgggggtg 900 cgtgcgtgtg tgtgtgcgtg gggagcgccg cgtgcggctc cgcgctgccc ggcggctgtg 960 agcgctgcgg gcgcggcgcg gggctttgtg cgctccgcag tgtgcgcgag gggagcgcgg 1020 ccgggggcgg tgccccgcgg tgcgggggg gctgcgaggg gaacaaaggc tgcgtgcggg 1080 gtgtgtgcgt ggggggggtga gcaggggtg tgggcgcgtc ggtcgggctg caacccccc 1140 tgcacccccc tccccgagtt gctgagcacg gcccggcttc gggtgcgggg ctccgtacgg 1200 ggcgtggcgc ggggctcgcc gtgccgggcg gggggtggcg gcaggtgggg gtgccgggcg 1260 gggcggggcc gcctcgggcc ggggagggct cgggggaggg gcgcggcggc ccccggagcg 1320 ccggcggctg tcgaggcgcg gcgagccgca gccattgcct tttatggtaa tcgtgcgaga 1380 gggcgcaggg acttcctttg tcccaaatct gtgcggagcc gaaatctggg aggcgccgcc 1440 gcacccccctc tagcgggcgc ggggcgaagc ggtgcggcgc cggcaggaag gaaatgggcg 1500 gggagggcct tcgtgcgtcg ccgcgccgcc gtcccccttct ccctctccag cctcggggct 1560 gtccgcgggg ggacggctgc cttcgggggg gacggggcag ggcggggttc ggcttctggc 1620 gtgtgaccgg cggctctaga gcctctgcta accatgttca tgccttcttc tttttcctac 1680 agctcctggg caacgtgctg gttattgtgc tgtctcatca ttttggcaaa gaattcgccc 1740 tgcagcggga attgccacca tggctcacgt tcggggactg cagctgcctg gatgtctggc 1800 tcttgccgct ctgtgtagcc tggtgcacag ccagcacgtg tttctggctc ctcagcaagc 1860 cagatcactg ctgcagagag ttagaaggca cggcgacggc agcttcagcg acgagatgaa 1920 caccatcctg gacaacctgg ccgccagaga cttcatcaac tggctgatcc agaccaagat 1980 caccgacggt ggcggaggcg gaggatctgg tggtggtgga tctggcggcg gaggttctgc 2040 cgagtctaag tacggacctc cttgtcctcc ctgtcctgct ccagaagctg ctggcggccc 2100 atccgtgttt ctgttccctc caaagcctaa ggacaccctg atgatcagca gaacccctga 2160 agtgacctgc gtggtggtcg acgtgtccca agaggatcct gaggtgcagt tcaattggta 2220 cgtggacggc gtggaagtgc acaacgccaa gaccaagcct agagaggaac agttcaacag 2280 cacctacaga gtggtgtccg tgctgaccgt gctgcaccag gattggctga acggcaaaga 2340 gtacaagtgc aaggtgtcca acaagggcct gcctagcagc atcgagaaaa ccatcagcaa 2400 ggccaagggc cagccaagag aaccccaggt gtacacactg cctccaagcc aagaggaaat 2460 gaccaagaac caggtgtccc tgacctgcct ggtcaagggc ttctaccctt ccgatatcgc 2520 cgtggaatgg gagagcaacg gccagcctga gaacaactac aagaccacac ctcctgtgct 2580 ggactccgat ggctcattct tcctgtacag cagactgacc gtggacaaga gcaggtggca 2640 agagggcaac gtgttcagct gcagcgtgat gcacgaggcc ctgcacaacc actacaccca 2700

		gaaaagcctg agcctgtctc tgggctaata aggtacctct agagcggccg cctcgaggac 2760
		ggggtgaact acgcctgagg atccgatctt tttccctctg ccaaaaatta tggggacatc 2820
		atgaagcccc ttgagcatct gacttctggc taataaagga aatttatttt cattgcaata 2880
		gtgtgttgga atttttgtg tctctcactc g 2911
3-44	**Sequences**	
3-44-1	Sequence Number [ID]	44
3-44-2	Molecule Type	DNA
3-44-3	Length	3964
3-44-4	Features Location/	**misc_feature 1..3964**
	Qualifiers	note=synthetic construct
		source 1..3964
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-44-5	Residues	ctagtcgaca ttgattattg actagttatt aatagtaatc aattacgggg tcattagttc 60
		atagcccata tatggagttc cgcgttacat aacttacggt aaatggcccg cctggctgac 120
		cgcccaacga cccccgccca ttgacgtcaa taatgacgta tgttcccata gtaacgccaa 180
		tagggacttt ccattgacgt caatgggtgg agtatttacg gtaaactgcc cacttggcag 240
		tacatcaagt gtatcatatg ccaagtacgc cccctattga cgtcaatgac ggtaaatggc 300
		ccgcctggca ttatgcccag tacatgacct tatgggactt tcctacttgg cagtacatct 360
		acgtattagt catcgctatt accatggtcg aggtgagccc cacgttctgc ttcactctcc 420
		ccatctcccc cccctcccca cccccaattt tgtatttatt tatttttaa ttattttgtg 480
		cagcgatggg ggcgggggggg gggggggggc gcgcgccagg cggggcgggg cggggcgagg 540
		ggcggggcgg ggcgaggcgg agaggtgcgg cggcagccaa tcagagcggc gcgctccgaa 600
		agtttccttt tatggcgagg cggcggcggc ggcggcccta taaaaagcga agcgcgcggc 660
		gggcgggagt cgctgcgcgc tgccttcgcc ccgtgccccg ctccgccgcc gcctcgcgcc 720
		gcccgccccg gctctgactg accgcgttac tcccacaggt gagcgggcgg gacggccctt 780
		ctcctccggg ctgtaattag cgcttggttt aatgacggct tgtttctttt ctgtggctgc 840
		gtgaaagcct tgaggggctc cgggagggcc ctttgtgcgg ggggagcggc tcgggggggtg 900
		cgtgcgtgtg tgtgtgcgtg gggagcgccg cgtgcggctc cgcgctgccc ggcggctgtg 960
		agcgctgcgg gcgcggcgcg gggctttgtg cgctccgcag tgtgcgcgag gggagcgcgg 1020
		ccgggggcgg tgccccgcgg tgcgggggggg gctgcgaggg gaacaaaggc tgcgtgcggg 1080
		gtgtgtgcgt ggggggggtga gcaggggggtg tgggcgcgtc ggtcgggctg caacccccccc 1140
		tgcacccccc tccccgagtt gctgagcacg gcccggcttc gggtgcgggg ctccgtacgg 1200
		ggcgtggcgc ggggctcgcc gtgccgggcg gggggtggcg gcaggtgggg gtgccgggcg 1260
		gggcggggcc gcctcgggcc ggggagggct cgggggaggg gcgcggcggc ccccggagcg 1320
		ccggcggctg tcgaggcgcg gcgagccgca gccattgcct tttatggtaa tcgtgcgaga 1380
		gggcgcaggg acttcctttg tcccaaatct gtgcggagcc gaaatctggg aggcgccgcc 1440
		gcacccccctc tagcgggcgc ggggcgaagc ggtgcggcgc cggcaggaag gaaatgggcg 1500
		gggagggcct tcgtgcgtcg ccgcgccgcc gtccccttct ccctctccag cctcggggct 1560
		gtccgcgggg ggacggctgc cttcgggggg gacggggcag ggcggggttc ggcttctggc 1620
		gtgtgaccgg cggctctaga gcctctgcta accatgttca tgccttcttc tttttcctac 1680
		agctcctggg caacgtgctg gttattgtgc tgtctcatca ttttggcaaa gaattccaga 1740
		agccaccatg gctcacgttc ggggactgca gctgcctgga tgtctggctc ttgccgctct 1800
		gtgtagcctg gtgcacagcc agcacgtgtt tctggctcct cagcaagcca gatcactgct 1860
		gcagagagtt agaaggcacg gcgacggcag cttcagcgac gagatgaaca ccatcctgga 1920
		caacctggcc gccagagact tcatcaactg gctgatccag accaagatca ccgacggtgg 1980
		cggaggcgga ggatctggtg gtggtggatc tggcggcgga ggttctgacg cccacaaatc 2040
		tgaagtggcc caccggttca aggacctggg cgaagagaat ttcaaggccc tggtgctgat 2100
		cgccttcgct cagtacctgc agcagtgccc cttcgaggac cacgtgaagc tggtcaacga 2160
		agtgaccgag ttcgccaaga cctgcgtggc cgacgagagc gccgagaact gtgataagag 2220
		cctgcacacc ctgttcggcg acaagctgtg tacagtggcc acactgagag aaacctacgg 2280
		cgagatggcc gactgctgcg ccaagcaaga gcccgagaga aacgagtgct tcctgcagca 2340
		caaggacgac aaccccaacc tgcctagact cgtgcggcct gaagtggacg tgatgtgcac 2400
		cgccttccac gacaacgagg aaaccttcct gaagaagtac ctgtacgaga tcgccagacg 2460
		gcacccctac ttttacgccc ctgagctgct gttcttcgcc aagcggtata aggccgcctt 2520
		caccgagtgt tgtcaggccg ctgataaggc tgcctgcctg ctgcctaagc tggacgagct 2580
		tagagacgag ggcaaagcca gctccgccaa gcagagactg aagtgtgcca gcctgcagaa 2640
		gttcggcgag agagccttta aggcctgggc cgttgctaga ctgagccaga gatttcccaa 2700
		ggccgagttt gccgaggtgt ccaagctcgt gaccgacctg acaaaggtgc acaccgagtg 2760
		ctgccacggc gacctgctgg aatgcgccga cgatagagcc gacctggcca agtacatctg 2820
		cgagaaccag gacagcatca gcagcaagct gaaagagtgc tgcgagaagc ctctgctgga 2880
		aaagagccac tgtatcgccg aggtggaaaa cgacgagatg cccgccgatc tgccttctct 2940
		ggccgccgat tttgtggaaa gcaaggacgt gtgcaagaac tacgccgagg ccaaggacgt 3000
		gttcctgggc atgtttctgt acgagtacgc ccgcagacac cccgactact ctgttgtgct 3060

		gctgctgaga ctggccaaaa cctacgagac aaccctggaa aagtgctgtg ccgccgctga 3120 tcctcacgag tgttacgcca aggtgttcga cgagttcaag ccactggtgg aagaacccca 3180 gaacctgatc aagcagaact gcgagctgtt cgagcagctg ggcgagtaca agttccagaa 3240 cgccctgctc gtgcggtaca ccaagaaggt gccccaggtt tccacaccta cactggttga 3300 ggtgtcccgg aacctgggca aagtgggcag caagtgttgc aagcaccctg aggccaagag 3360 aatgccctgc gccgaggatt acctgagcgt ggtgctgaat cagctgtgcg tgctgcacga 3420 gaaaacccct gtgtccgaca gagtgaccaa gtgctgtacc gagagcctgg tcaacagacg 3480 gccttgcttt agcgccctcg aggtggacga gacatacgtg cccaaagagt tcaacgccga 3540 gacattcacc ttccacgccg acatctgtac cctgagcgag aaagagcggc agatcaagaa 3600 acagactgcc ctggtggaac tggtcaagca caagcccaag gccaccaaag aacagctgaa 3660 ggccgtgatg gacgacttcg ccgccttcgt ggaaaagtgc tgcaaggccg acgacaaaga 3720 gacctgcttc gccgaagagg gcaagaaact ggtggccgct tctcaggctg ctctgggata 3780 ataaggtacc tctagagcgg ccgcctcgag gacggggtga actacgcctg aggatccgat 3840 ctttttccct ctgccaaaaa ttatggggac atcatgaagc cccttgagca tctgacttct 3900 ggctaataaa ggaaatttat tttcattgca atagtgtgtt ggaatttttt gtgtctctca 3960 ctcg 3964
3-45	**Sequences**	
3-45-1	Sequence Number [ID]	45
3-45-2	Molecule Type	DNA
3-45-3	Length	2911
3-45-4	Features Location/	**misc_feature 1..2911**
	Qualifiers	note=synthetic construct
		source 1..2911
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-45-5	Residues	ctagtcgaca ttgattattg actagttatt aatagtaatc aattacgggg tcattagttc 60 atagcccata tatggagttc cgcgttacat aacttacggt aaatggcccg cctggctgac 120 cgcccaacga cccccgccca ttgacgtcaa taatgacgta tgttcccata gtaacgccaa 180 tagggacttt ccattgacgt caatgggtgg agtatttacg gtaaactgcc cacttggcag 240 tacatcaagt gtatcatatg ccaagtacgc cccctattga cgtcaatgac ggtaaatggc 300 ccgcctggca ttatgcccag tacatgacct tatgggactt tcctacttgg cagtacatct 360 acgtattagt catcgctatt accatggtcg aggtgagccc cacgttctgc ttcactctcc 420 ccatctcccc cccctcccca cccccaattt tgtatttatt tatttttaa ttattttgtg 480 cagcgatggg ggcgggggg gggggggggc gcgcgccagg cggggcgggg cggggcgagg 540 ggcggggcgg ggcgaggcgg agaggtgcgg cggcagccaa tcagagcggc gcgctccgaa 600 agtttccttt tatggcgagg cggcggcggc ggcggcccta taaaaagcga agcgcgcggc 660 gggcgggagt cgctgcgcgc tgccttcgcc ccgtgccccg ctccgccgcc gcctcgcgcc 720 gcccgccccg gctctgactg accgcgttac tcccacaggt gagcgggcgg gacggccctt 780 ctcctccggg ctgtaattag cgcttggttt aatgacggct tgtttctttt ctgtggctgc 840 gtgaaagcct tgaggggctc cgggagggcc ctttgtgcgg ggggagcggc tcgggggggtg 900 cgtgcgtgtg tgtgtgcgtg gggagcgccg cgtgcggctc cgcgctgccc ggcggctgtg 960 agcgctgcgg gcgcggcgcg gggctttgtg cgctccgcag tgtgcgcgag gggagcgcgg 1020 ccgggggcgg tgccccgcgg tgcgggggggg gctgcgaggg gaacaaaggc tgcgtgcggg 1080 gtgtgtgcgt gggggggtga gcaggggggtg tgggcgcgtc ggtcgggctg caacccccccc 1140 tgcacccccc tccccgagtt gctgagcacg gcccggcttc gggtgcgggg ctccgtacgg 1200 ggcgtggcgc ggggctcgcc gtgccgggcg gggggtggcg gcaggtgggg gtgccgggcg 1260 gggcggggcc gcctcgggcc ggggagggct cgggggaggg gcgcggcggc ccccggagcg 1320 ccggcggctg tcgaggcgcg gcgagccgca gccattgcct tttatggtaa tcgtgcgaga 1380 gggcgcaggg acttcctttg tcccaaatct gtgcggagcc gaaatctggg aggcgccgcc 1440 gcacccccctc tagcgggcgc ggggcgaagc ggtgcggcgc cggcaggaag gaaatgggcg 1500 gggagggcct tcgtgcgtcg ccgcgccgcc gtccccttct ccctctccag cctcggggct 1560 gtccgcgggg ggacggctgc cttcgggggg gacggggcag ggcggggttc ggcttctggc 1620 gtgtgaccgg cggctctaga gcctctgcta accatgttca tgccttcttc tttttcctac 1680 agctcctggg caacgtgctg gttattgtgc tgtctcatca ttttggcaaa gaattcgccc 1740 tgcagcggga attgccacca tggctcacgt tcggggactg cagctgcctg gatgtctggc 1800 tcttgccgct ctgtgtagcc tggtgcacag ccagcacgtg tttctggctc ctcagcaagc 1860 cagatcactg ctgcagagag ttagaaggca cggcgacggc agcttcagcg acgagatgaa 1920 caccatcctg gacaacctgg ccgccagaga cttcatcaac tggctgatcc agaccaagat 1980 caccgacggt ggcggaggcg gaggatctgg tggtggtgga tctggcggcg gaggttctgc 2040 cgagtctaag tacggacctc cttgtcctcc ctgtcctgct ccagaagctg ctggcggccc 2100 atccgtgttt ctgttccctc caaagcctaa ggacaccctg atgatcagca gaacccctga 2160 agtgacctgc gtggtggtcg acgtgtccca agaggatcct gaggtgcagt tcaattggta 2220 cgtggacggc gtggaagtgc acaacgccaa gaccaagcct agagaggaac agttcaacag 2280 cacctacaga gtggtgtccg tgctgaccgt gctgcaccag gattggctga acggcaaaga 2340

		gtacaagtgc aaggtgtcca acaagggcct gcctagcagc atcgagaaaa ccatcagcaa 2400 ggccaagggc cagccaagag aaccccaggt gtacacactg cctccaagcc aagaggaaat 2460 gaccaagaac caggtgtccc tgacctgcct ggtcaagggc ttctaccctt ccgatatcgc 2520 cgtggaatgg gagagcaacg gccagcctga gaacaactac aagaccacac ctcctgtgct 2580 ggactccgat ggctcattct tcctgtacag cagactgacc gtggacaaga gcaggtggca 2640 agagggcaac gtgttcagct gcagcgtgat gcacgaggcc ctgcacaacc actacaccca 2700 gaaaagcctg agcctgtctc tgggctaata aggtacctct agagcggccg cctcgaggac 2760 ggggtgaact acgcctgagg atccgatctt tttccctctg ccaaaaatta tggggacatc 2820 atgaagcccc ttgagcatct gacttctggc taataaagga aatttatttt cattgcaata 2880 gtgtgttgga atttttgtg tctctcactc g 2911
3-46	**Sequences**	
3-46-1	Sequence Number [ID]	46
3-46-2	Molecule Type	DNA
3-46-3	Length	2911
3-46-4	Features Location/	**misc_feature 1..2911**
	Qualifiers	note=synthetic construct
		source 1..2911
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-46-5	Residues	ctagtcgaca ttgattattg actagttatt aatagtaatc aattacgggg tcattagttc 60 atagcccata tatggagttc cgcgttacat aacttacggt aaatggcccg cctggctgac 120 cgcccaacga cccccgccca ttgacgtcaa taatgacgta tgttcccata gtaacgccaa 180 tagggacttt ccattgacgt caatgggtgg agtatttacg gtaaactgcc cacttggcag 240 tacatcaagt gtatcatatg ccaagtacgc cccctattga cgtcaatgac ggtaaatggc 300 ccgcctggca ttatgcccag tacatgacct tatgggactt tcctacttgg cagtacatct 360 acgtattagt catcgctatt accatggtcg aggtgagccc cacgttctgc ttcactctcc 420 ccatctcccc cccctcccca cccccaattt tgtatttatt tatttttaa ttattttgtg 480 cagcgatggg ggcgggggg gggggggggc gcgcgccagg cggggcgggg cggggcgagg 540 ggcggggcgg ggcgaggcgg agaggtgcgg cggcagccaa tcagagcggc gcgctccgaa 600 agtttccttt tatggcgagg cggcggcggc ggcggcccta taaaaagcga agcgcgcggc 660 gggcgggagt cgctgcgcgc tgccttcgcc ccgtgccccg ctccgccgcc gcctcgcgcc 720 gcccgccccg gctctgactg accgcgttac tcccacaggt gagcgggcgg gacggccctt 780 ctcctccggg ctgtaattag cgcttggttt aatgacggct tgtttctttt ctgtggctgc 840 gtgaaagcct tgaggggctc cgggagggcc ctttgtgcgg ggggagcggc tcgggggtg 900 cgtgcgtgtg tgtgtgcgtg gggagcgccg cgtgcggctc cgcgctgccc ggcggctgtg 960 agcgctgcgg gcgcggcgcg gggctttgtg cgctccgcag tgtgcgcgag gggagcgcgg 1020 ccgggggcgg tgccccgcgg tgcgggggg gctgcgaggg gaacaaaggc tgcgtgcggg 1080 gtgtgtgcgt gggggggtga gcaggggtg tgggcgcgtc ggtcgggctg caacccccc 1140 tgcacccccc tccccgagtt gctgagcacg gcccggcttc gggtgcgggg ctccgtacgg 1200 ggcgtggcgc ggggctcgcc gtgccgggcg gggggtggcg gcaggtgggg gtgccgggcg 1260 gggcggggcc gcctcgggcc ggggagggct cgggggaggg gcgcggcggc ccccggagcg 1320 ccggcggctg tcgaggcgcg gcgagccgca gccattgcct tttatggtaa tcgtgcgaga 1380 gggcgcaggg acttcctttg tcccaaatct gtgcggagcc gaaatctggg aggcgccgcc 1440 gcacccccc tagcgggcgc ggggcgaagc ggtgcggcgc cggcaggaag gaaatgggcg 1500 gggagggcct tcgtgcgtcg ccgcgccgcc gtcccccttct ccctctccag cctcggggct 1560 gtccgcgggg ggacggctgc cttcgggggg gacggggcag ggcggggttc ggcttctggc 1620 gtgtgaccgg cggctctaga gcctctgcta accatgttca tgccttcttc tttttcctac 1680 agctcctggg caacgtgctg gttattgtgc tgtctcatca ttttggcaaa gaattcgccc 1740 tgcagcggga attgccacca tggctcacgt tcggggactg cagctgcctg gatgtctggc 1800 tcttgccgct ctgtgtagcc tggtgcacag ccagcatgtg tttctggctc ctcagcaggc 1860 cctgagcctg ctgcaaagag ttagaaggca cggcgacggc agcttcagcg acgagatgaa 1920 taccgtgctg gtggacaacc tggccaccag agacttcatc aactggctga tccagaccaa 1980 gatcaccgac ggtggtggcg gaggcggagg atctggtggc ggtggttctg gcggtggcgg 2040 atctgctgag tttacccctc cttgtcctcc ctgtcctgct ccagaactgc tcggcggacc 2100 ttccgtgttc ctgtttcctc caaagcctaa ggacaccctg atgatcagca gaacccctga 2160 agtgacctgc gtggtggtgg acgtgtccca agaggatcct gaggtgcagt tcaattggta 2220 cgtggacggc gtggaagtgc acaacgccca gacaaagccc agagagcggc agttcaacag 2280 cacctacaga gtggtgtccg tgctgaccgt gacacaccag gattggctga acggcaaaga 2340 gtacacctgt aaagtctcca acaagggcct gcctgctcct atcgagaaaa ccatcagcaa 2400 ggccaagggc cagcctagag aaccccaggt gtacatcctg cctccacctc aagaggaact 2460 gaccaagaac caggtgtccc tgacctgtct ggtcaccggc ttctaccctt ccgatatcgc 2520 cgtggagtgg gagagcaacg gacagcccga gaacacctac aagaccacac ctccagtgct 2580 ggacagcgac ggctcttacc tgctgtactc caagctgaca gtgaacaaga gccggtggca 2640 gcccggcaac atcttcacct gttctgtgat gcacgaggcc ctgcacaacc actacaccca 2700

		gaaaagcctg agcgtgtccc ctggataata aggtacctct agagcggccg cctcgaggac 2760
		ggggtgaact acgcctgagg atccgatctt tttccctctg ccaaaaatta tggggacatc 2820
		atgaagcccc ttgagcatct gacttctggc taataaagga aatttatttt cattgcaata 2880
		gtgtgttgga atttttgtg tctctcactc g 2911
3-47	**Sequences**	
3-47-1	Sequence Number [ID]	47
3-47-2	Molecule Type	DNA
3-47-3	Length	3967
3-47-4	Features Location/	**misc_feature 1..3967**
	Qualifiers	note=synthetic construct
		source 1..3967
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-47-5	Residues	ctagtcgaca ttgattattg actagttatt aatagtaatc aattacgggg tcattagttc 60
		atagcccata tatggagttc cgcgttacat aacttacggt aaatggcccg cctggctgac 120
		cgcccaacga cccccgccca ttgacgtcaa taatgacgta tgttcccata gtaacgccaa 180
		tagggacttt ccattgacgt caatgggtgg agtatttacg gtaaactgcc cacttggcag 240
		tacatcaagt gtatcatatg ccaagtacgc cccctattga cgtcaatgac ggtaaatggc 300
		ccgcctggca ttatgcccag tacatgacct tatgggactt tcctacttgg cagtacatct 360
		acgtattagt catcgctatt accatggtcg aggtgagccc cacgttctgc ttcactctcc 420
		ccatctcccc cccctcccca cccccaattt tgtatttatt tatttttaa ttattttgtg 480
		cagcgatggg ggcgggggggg gggggggggc gcgcgccagg cggggcgggg cggggcgagg 540
		ggcggggcgg ggcgaggcgg agaggtgcgg cggcagccaa tcagagcggc gcgctccgaa 600
		agtttccttt tatggcgagg cggcggcggc ggcggcccta taaaaagcga agcgcgcggc 660
		gggcgggagt cgctgcgcgc tgccttcgcc ccgtgccccg ctccgccgcc gcctcgcgcc 720
		gcccgccccg gctctgactg accgcgttac tcccacaggt gagcgggcgg gacggccctt 780
		ctcctccggg ctgtaattag cgcttggttt aatgacggct tgtttctttt ctgtggctgc 840
		gtgaaagcct tgaggggctc cgggagggcc ctttgtgcgg ggggagcggc tcgggggggtg 900
		cgtgcgtgtg tgtgtgcgtg gggagcgccg cgtgcggctc cgcgctgccc ggcggctgtg 960
		agcgctgcgg gcgcggcgcg gggctttgtg cgctccgcag tgtgcgcgag gggagcgcgg 1020
		ccgggggcgg tgccccgcgg tgcgggggggg gctgcgaggg gaacaaaggc tgcgtgcggg 1080
		gtgtgtgcgt ggggggggtga gcaggggggtg tgggcgcgtc ggtcgggctg caacccccccc 1140
		tgcacccccc tccccgagtt gctgagcacg gcccggcttc gggtgcgggg ctccgtacgg 1200
		ggcgtggcgc ggggctcgcc gtgccgggcg gggggtggcg gcaggtgggg gtgccgggcg 1260
		gggcggggcc gcctcgggcc ggggagggct cgggggaggg gcgcggcggc ccccggagcg 1320
		ccggcggctg tcgaggcgcg gcgagccgca gccattgcct tttatggtaa tcgtgcgaga 1380
		gggcgcaggg acttcctttg tcccaaatct gtgcggagcc gaaatctggg aggcgccgcc 1440
		gcacccccctc tagcgggcgc ggggcgaagc ggtgcggcgc cggcaggaag gaaatgggcg 1500
		gggagggcct tcgtgcgtcg ccgcgccgcc gtccccttct ccctctccag cctcggggct 1560
		gtccgcgggg ggacggctgc cttcgggggg gacggggcag ggcggggttc ggcttctggc 1620
		gtgtgaccgg cggctctaga gcctctgcta accatgttca tgccttcttc tttttcctac 1680
		agctcctggg caacgtgctg gttattgtgc tgtctcatca ttttggcaaa gaattccaga 1740
		agccaccatg gctcacgttc ggggactgca gctgcctgga tgtctggctc ttgccgctct 1800
		gtgtagcctg gtgcacagcc agcatgtgtt tctggctcct cagcaagccc tgagcctgct 1860
		gcaaagagtt agaaggcacg gcgacggcag cttcagcgac gagatgaata ccgtgctggt 1920
		ggacaacctg gccaccagag acttcatcaa ctggctgatc cagaccaaga tcaccggtgg 1980
		tggcggaggc ggaggatctg gtggcggtgg ttctggcggt ggcggatctg atacacacaa 2040
		gtctgaggtg gcccaccggt tcaaggacct gggcgaagaa cacttcaaag gcctggtgct 2100
		ggtcgccttc agccagtacc tgcagcagtg ccctttcgag gaacacgtga agctggtcaa 2160
		cgaagtgacc gagttcgcca agacctgcgt ggccgacgag agcgccgaga actgtgataa 2220
		gagcctgcac accctgttcg gcgacaagct gtgtacagtg gccacactga gagaaaccta 2280
		cggcgagatg gccgactgct gcgccaagca agagcccgag agaaacgagt gcttcctgca 2340
		gcacaaggac gacaacccca acctgcctcc actcgtcaga cccgaagtgg acgtgatgtg 2400
		caccgccttc cacgacaatg aggccacctt cctgaagaaa tacctgtacg aggtggccag 2460
		acggcacccc tacttttacg cccctgaact gctgttcttt gccgccaggt acaaggccgc 2520
		cttcgccgaa tgttgtcagg ccgctgataa ggccgcttgc ctgctgccta agctggacga 2580
		gcttagagac gagggcaaag ccagctccgc caagcagaga ctgaagtgtg ccagcctgca 2640
		gaagttcggc gatagagcct ttaaggcctg ggccgtcgct agactgagcc agaagtttcc 2700
		caaggccgag tttgccgagg tgtccaagct cgtgaccgac ctgacaaagg tgcacaccga 2760
		gtgctgtcac ggcgacctgc tggaatgcgc cgacgataga gccgacctgg ccaagtacat 2820
		gtgcgagaac caggacagca tcagcagcaa gctgaaagag tgctgcgaca agcctctgct 2880
		ggaaaagagc cactgtctgg ccgaggtgga aaacgacgag atgcccgccg atctgccttc 2940
		tctggccgcc gattacgtgg aaagcaagga cgtgtgcaag aactacgccg aggccaagga 3000
		cgtgttcctg ggcatgtttc tgtacgagta cgcccgcaga caccccgact actctgttat 3060

		gctgctgctg agactggcca aggcctacga ggccactctg gaaaagtgtt gtgccgccgc 3120 tgatccccac gagtgttacg ccaaagtgtt cgacgagttc cagccactgg tggaagaacc 3180 ccagaacctg gtcaagcaga actgcgagct gttcgagcag ctgggcgagt acaagttcca 3240 gaacgccctg ctcgtgcggt acaccaagaa ggtgccccag gtttccacac ctacactggt 3300 tgaggtgtcc cggaacctgg gaaaagtggg cgccaagtgt tgcaagctgc ctgaggccaa 3360 gagaatgccc tgcgccgagg attacctgag cgtggtgctg aacagactgt gcgtgctgca 3420 cgagaaaacc cctgtgtccg agaaagtgac caagtgctgt accgagagcc tggtcaatcg 3480 gaggccttgc tttagcgccc tggaactgga cgaggcctac gtgcccaagg ccttcaacgc 3540 cgagacattc accttccacg ccgacatgtg taccctgagc gagaaagaaa agcaagtgaa 3600 gaaacagaca gccctggtcg agctggttaa gcacaagcct aaggccacca aagaacaact 3660 gaagggcgtg atggacaact tcgccgcctt tgtggaaaaa tgctgcaagg ccgacgacaa 3720 agaggcctgc ttcgcagaag agggccctaa gtttgtggcc gcctctcaag ctgctctggc 3780 ttaataaggt acctctagag cggccgcctc gaggacgggg tgaactacgc ctgaggatcc 3840 gatctttttc cctctgccaa aaattatggg gacatcatga agcccccttga gcatctgact 3900 tctggctaat aaaggaaatt tattttcatt gcaatagtgt gttggaattt tttgtgtctc 3960 tcactcg 3967
3-48	**Sequences**	
3-48-1	Sequence Number [ID]	48
3-48-2	Molecule Type	DNA
3-48-3	Length	4807
3-48-4	Features Location/	**misc_feature 1..4807**
	Qualifiers	note=synthetic construct
		source 1..4807
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-48-5	Residues	ctagtcgaca ttgattattg actagttatt aatagtaatc aattacgggg tcattagttc 60 atagcccata tatggagttc cgcgttacat aacttacggt aaatggcccg cctggctgac 120 cgcccaacga cccccgccca ttgacgtcaa taatgacgta tgttcccata gtaacgccaa 180 tagggacttt ccattgacgt caatgggtgg agtatttacg gtaaactgcc cacttggcag 240 tacatcaagt gtatcatatg ccaagtacgc cccctattga cgtcaatgac ggtaaatggc 300 ccgcctggca ttatgcccag tacatgacct tatgggactt tcctacttgg cagtacatct 360 acgtattagt catcgctatt accatggtcg aggtgagccc cacgttctgc ttcactctcc 420 ccatctcccc cccctcccca cccccaattt tgtatttatt tatttttaa ttattttgtg 480 cagcgatggg ggcgggggg gggggggggc gcgcgccagg cggggcgggg cggggcgagg 540 ggcggggcgg ggcgaggcgg agaggtgcgg cggcagccaa tcagagcggc gcgctccgaa 600 agtttccttt tatggcgagg cggcggcggc ggcggcccta taaaaagcga agcgcgcggc 660 gggcgggagt cgctgcgcgc tgccttcgcc ccgtgccccg ctccgccgcc gcctcgcgcc 720 gcccgccccg gctctgactg accgcgttac tcccacaggt gagcgggcgg gacggccctt 780 ctcctccggg ctgtaattag cgcttggttt aatgacggct tgtttctttt ctgtggctgc 840 gtgaaagcct tgaggggctc cgggagggcc ctttgtgcgg ggggagcggc tcgggggggtg 900 cgtgcgtgtg tgtgtgcgtg gggagcgccg cgtgcggctc cgcgctgccc ggcggctgtg 960 agcgctgcgg gcgcggcgcg gggctttgtg cgctccgcag tgtgcgcgag gggagcgcgg 1020 ccgggggcgg tgccccgcgg tgcgggggggg gctgcgaggg gaacaaaggc tgcgtgcggg 1080 gtgtgtgcgt gggggggtga gcaggggggtg tgggcgcgtc ggtcgggctg caacccccccc 1140 tgcacccccc tccccgagtt gctgagcacg gcccggcttc gggtgcgggg ctccgtacgg 1200 ggcgtggcgc ggggctcgcc gtgccgggcg gggggtggcg gcaggtgggg gtgccgggcg 1260 gggcggggcc gcctcgggcc ggggagggct cgggggaggg gcgcggcggc ccccggagcg 1320 ccggcggctg tcgaggcgcg gcgagccgca gccattgcct tttatggtaa tcgtgcgaga 1380 gggcgcaggg acttcctttg tcccaaatct gtgcggagcc gaaatctggg aggcgccgcc 1440 gcacccccctc tagcgggcgc ggggcgaagc ggtgcggcgc cggcaggaag gaaatgggcg 1500 gggagggcct tcgtgcgtcg ccgcgccgcc gtcccccttct ccctctccag cctcggggct 1560 gtccgcgggg ggacggctgc cttcgggggg gacggggcag ggcggggttc ggcttctggc 1620 gtgtgaccgg cggctctaga gcctctgcta accatgttca tgccttcttc tttttcctac 1680 agctcctggg caacgtgctg gttattgtgc tgtctcatca ttttggcaaa gaattccaga 1740 agccaccatg gctcacgttc ggggactgca gctgcctgga tgtctggctc ttgccgctct 1800 gtgtagcctg gtgcacagcc agcatgtgtt tctggctcct cagcaagccc tgagcctgct 1860 gcaaagagtt agaaggcacg gcgacggcag cttcagcgac gagatgaata ccgtgctggt 1920 ggacaacctg gccaccagag acttcatcaa ctggctgatc cagaccaaga tcaccgacgg 1980 tggcggaggc ggaggatctg gtggtggtgg atctggcggc ggaggaagtt ctcctgctgg 2040 cagccctaca agcaccgagg aaggcacaag cgagtctgcc acacctgagt ctggccctgg 2100 cacatctaca gagcctagcg aaggatctgc cccaggatct cctgccggct ctccaacatc 2160 taccgaagag ggaaccagca ccgagccatc tgagggatct gctcccggaa caagcacaga 2220 gccttcagaa ggatccgctc ctggcacctc tgaaagcgcc acaccagaaa gcggacctgg 2280 ctctgaacct gccacaagcg gatctgagac acctggaagc gagccagcca catctggcag 2340

		cgaaacacct ggatcaccag ccggatctcc cacctctacc gaggaaggga catccgagag 2400 cgctacccca gaatctggac caggcaccag cacagaaccc tctgaaggtt cagcccctgg 2460 aacctctacc gaaccatcag aaggcagcgc tccaggttct cccgctggat cccctacatc 2520 cacagaagag ggcacctcca ctgaacctag cgagggaagt gctcccggca cttccacaga 2580 accatccgag ggcagtgcac ctggaaccag cgaatctgct acccctgaga gtggacccgg 2640 aacatccact gagccctccg agggttcagc tccaggcaca tcagaatccg ccactccaga 2700 gtccggacca ggatctgagc cagctaccag cggctctgaa acacccggca ctagtaccga 2760 gccaagcgag ggtagcgcac cagggacaag taccgaaccg tctgagggct ccgcaccagg 2820 cacttccgaa agtgctactc ctgaaagcgg cccaggcact agcgaatccg caacacccga 2880 gagcggtcct ggaagtcctg caggttcacc taccagcact gaagagggga ctagcgagag 2940 cgcaactcct gaatcaggcc ctggatccga acctgctacc tccggaagtg aaacccctgg 3000 gacaagcgaa agtgcaacgc ccgagtcagg acccgggact agcacggaac ccagtgaagg 3060 atctgcaccc gggacatcta ccgagccgtc agaaggttct gctccaggga ctagtactga 3120 gccttccgaa ggttctgcac ctggaacttc cacagagccc agtgaaggca gtgcccctgg 3180 cacaagcact gaaccgtccg aaggcagtgc tcccgggacc agtacagaac cgagcgaggg 3240 ctctgctcct ggtagtccag caggatctcc aactagcacc gaagaaggga cttccaccga 3300 gccttccgag ggaagcgctc ctggaacatc cgagtccgct acgccagaga gtggcccagg 3360 ttctgaaccc gctacttccg gctcagagac tcctgggact tctgagtctg caaccccgga 3420 aagtggtcct ggtagcgaac cagcaactag cggaagcgag acacccggaa cctcagagag 3480 tgctacaccg gaatccggtc cagggacgtc tacggaaccg tctgaaggat cagctcccgg 3540 cactagcgaa agcgctacac ctgaaagtgg tcccggatct ccagcaggca gcccaacctc 3600 tactgaagaa ggttccccag ctggaagccc cacttccact gaggaaggct ctcccgcagg 3660 ctcacccact agtacggaag aaggcacgtc cgagtctgct actcccgaat ccggacctgg 3720 aactagcact gagccaagcg aaggatcagc acccggaacc tctgagtccg ccacaccaga 3780 atctggtcct ggttccgagc ctgccacttc aggatcagaa accccgggca cgagtgaatc 3840 agcaacgccg gaatctggcc ccggaagcga accggctacg tctggatctg aaacgccagg 3900 gacctccgaa tcagctacgc ctgagtctgg tccagggaca tccaccgaac ctagtgaagg 3960 ctccgcacct ggaagccctg ctggaagccc aacgagtact gaagagggca cttctgagag 4020 cgctacgcct gagtcaggac ctggaagcga acctgcaaca tccggctcag aaacaccagg 4080 gaccagcgaa agcgcaaccc cagagagtgg acctggatct ccagctggct ctcctactag 4140 tacagaggaa ggcagccctg ctggctcccc aacgtcaaca gaagaaggta ctagcacaga 4200 gcccagcgag ggttccgctc cgggaacttc tgaatctgct acacccgagt caggtcctgg 4260 tacaagcgag tcagctacgc ccgaaagtgg acctggcacc tcagagtctg caactcctga 4320 gagcggtcca ggatcagaac cagccacctc tggctctgag acaccaggtt ctgagcctgc 4380 aacgtccgga agcgaaacac caggcagtcc tgccggaagt cctacttcaa ccgaagaggg 4440 gacctctaca gagccatcag agggctctgc accgggcacc tcaacagaac catctgaagg 4500 atccgcaccg ggctctgagc ctgctactag tggaagcgaa actcctggca ccagtgaatc 4560 cgctactccc gagtctggcc cgggaacgtc tactgaacca tctgagggaa gtgccccagg 4620 ctaataaggt acctctagag cggccgcctc gaggacgggg tgaactacgc ctgaggatcc 4680 gatctttttc cctctgccaa aaattatggg gacatcatga agcccccttga gcatctgact 4740 tctggctaat aaaggaaatt tattttcatt gcaatagtgt gttggaattt tttgtgtctc 4800 tcactcg 4807
3-49	**Sequences**	
3-49-1	Sequence Number [ID]	49
3-49-2	Molecule Type	DNA
3-49-3	Length	1513
3-49-4	Features Location/	**misc_feature 1..1513**
	Qualifiers	note=synthetic construct
		source 1..1513
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-49-5	Residues	acgcgtgcta gctaatgatg ggcgctcgag taatgatggg cggtcgacta atgatgggcg 60 ctcgagtaat gatgggcgtc tagctaatga tgggcgctcg agtaatgatg ggcggtcgac 120 taatgatggg cgctcgagta atgatgggcg tctagctaat gatgggcgct cgagtaatga 180 tgggcggtcg actaatgatg ggcgctcgag taatgatggg cgtctagaac gcgaattaat 240 tcaacatttt gacaccccca taatattttt ccagaattaa cagtataaat tgcatctctt 300 gttcaagagt tccctatcac tctctttaat cactactcac agtaacctca actcctgcca 360 caagcttgcc ctgcagcggg aattcgccct gcagcgggaa ttgccaccat ggctcacgtt 420 cggggactgc agctgcctgg atgtctggct cttgccgctc tgtgtagcct ggtgcacagc 480 cagcatgtgt ttctggctcc tcagcaggcc ctgagcctgc tgcaaagagt tagaaggcac 540 ggcgacggca gcttcagcga cgagatgaat accgtgctgg tggacaacct ggccaccaga 600 gacttcatca actggctgat ccagaccaag atcaccgacg gtggtggcgg aggcggagga 660 tctggtggcg gtggttctgg cggtggcgga tctgctgagt ttacccctcc ttgtcctccc 720 tgtcctgctc cagaactgct cggcggacct tccgtgttcc tgtttcctcc aaagcctaag 780

		gacaccctga tgatcagcag aacccctgaa gtgacctgcg tggtggtgga cgtgtcccaa 840 gaggatcctg aggtgcagtt caattggtac gtggacggcg tggaagtgca caacgcccag 900 acaaagccca gagagcggca gttcaacagc acctacagag tggtgtccgt gctgaccgtg 960 acacaccagg attggctgaa cggcaaagag tacacctgta aagtctccaa caagggcctg 1020 cctgctccta tcgagaaaac catcagcaag gccaagggcc agcctagaga accccaggtg 1080 tacatcctgc ctccacctca agaggaactg accaagaacc aggtgtccct gacctgtctg 1140 gtcaccggct tctacccttc cgatatcgcc gtggagtggg agagcaacgg acagcccgag 1200 aacacctaca agaccacacc tccagtgctg gacagcgacg gctcttacct gctgtactcc 1260 aagctgacag tgaacaagag ccggtggcag cccggcaaca tcttcacctg ttctgtgatg 1320 cacgaggccc tgcacaacca ctacacccag aaaagcctga gcgtgtcccc tggataataa 1380 ggtaccgatc tttttccctc tgccaaaaat tatggggaca tcatgaagcc ccttgagcat 1440 ctgacttctg gctaataaag gaaatttatt ttcattgcaa tagtgtgttg gaatttttg 1500 tgtctctcac tcg 1513
3-50	**Sequences**	
3-50-1	Sequence Number [ID]	50
3-50-2	Molecule Type	DNA
3-50-3	Length	2569
3-50-4	Features Location/	**misc_feature 1..2569**
	Qualifiers	note=synthetic construct
		source 1..2569
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-50-5	Residues	acgcgtgcta gctaatgatg ggcgctcgag taatgatggg cggtcgacta atgatgggcg 60 ctcgagtaat gatgggcgtc tagctaatga tgggcgctcg agtaatgatg ggcggtcgac 120 taatgatggg cgctcgagta atgatgggcg tctagctaat gatgggcgct cgagtaatga 180 tgggcggtcg actaatgatg ggcgctcgag taatgatggg cgtctagaac gcgaattaat 240 tcaacatttt gacaccccca taatattttt ccagaattaa cagtataaat tgcatctctt 300 gttcaagagt tccctatcac tctctttaat cactactcac agtaacctca actcctgcca 360 caagcttgcc ctgcagcggg aattccagaa gccaccatgg ctcacgttcg gggactgcag 420 ctgcctggat gtctggctct tgccgctctg tgtagcctgg tgcacagcca gcatgtgttt 480 ctggctcctc agcaagccct gagcctgctg caaagagtta gaaggcacgg cgacggcagc 540 ttcagcgacg agatgaatac cgtgctggtg gacaacctgg ccaccagaga cttcatcaac 600 tggctgatcc agaccaagat caccggtggt ggcggaggcg gaggatctgg tggcggtggt 660 tctggcggtg gcggatctga tacacacaag tctgaggtgg cccaccggtt caaggacctg 720 ggcgaagaac acttcaaagg cctggtgctg gtcgccttca gccagtacct gcagcagtgc 780 cctttcgagg aacacgtgaa gctggtcaac gaagtgaccg agttcgccaa gacctgcgtg 840 gccgacgaga gcgccgagaa ctgtgataag agcctgcaca ccctgttcgg cgacaagctg 900 tgtacagtgg ccacactgag agaaacctac ggcgagatgg ccgactgctg cgccaagcaa 960 gagcccgaga gaaacgagtg cttcctgcag cacaaggacg acaaccccaa cctgcctcca 1020 ctcgtcagac ccgaagtgga cgtgatgtgc accgccttcc acgacaatga ggccaccttc 1080 ctgaagaaat acctgtacga ggtggccaga cggcacccct acttttacgc ccctgaactg 1140 ctgttctttg ccgccaggta caaggccgcc ttcgccgaat gttgtcaggc cgctgataag 1200 gccgcttgcc tgctgcctaa gctggacgag cttagagacg agggcaaagc cagctccgcc 1260 aagcagagac tgaagtgtgc cagcctgcag aagttcggcg atagagcctt taaggcctgg 1320 gccgtcgcta gactgagcca gaagtttccc aaggccgagt ttgccgaggt gtccaagctc 1380 gtgaccgacc tgacaaaggt gcacaccgag tgctgtcacg gcgacctgct ggaatgcgcc 1440 gacgatagag ccgacctggc caagtacatg tgcgagaacc aggacagcat cagcagcaag 1500 ctgaaagagt gctgcgacaa gcctctgctg gaaaagagcc actgtctggc cgaggtggaa 1560 aacgacgaga tgcccgccga tctgccttct ctggccgccg attacgtgga aagcaaggac 1620 gtgtgcaaga actacgccga ggccaaggac gtgttcctgg gcatgtttct gtacgagtac 1680 gcccgcagac accccgacta ctctgttatg ctgctgctga gactggccaa ggcctacgag 1740 gccactctgg aaaagtgttg tgccgccgct gatccccacg agtgttacgc caaagtgttc 1800 gacgagttcc agccactggt ggaagaaccc cagaacctgg tcaagcagaa ctgcgagctg 1860 ttcgagcagc tgggcgagta caagttccag aacgccctgc tcgtgcggta caccaagaag 1920 gtgccccagg tttccacacc tacactggtt gaggtgtccc ggaacctggg aaaagtgggc 1980 gccaagtgtt gcaagctgcc tgaggccaag agaatgccct gcgccgagga ttacctgagc 2040 gtggtgctga acagactgtg cgtgctgcac gagaaaaccc ctgtgtccga gaaagtgacc 2100 aagtgctgta ccgagagcct ggtcaatcgg aggccttgct ttagcgccct ggaactggac 2160 gaggcctacg tgcccaaggc cttcaacgcc gagacattca ccttccacgc cgacatgtgt 2220 accctgagcg agaaagaaaa gcaagtgaag aaacagacag ccctggtcga gctggttaag 2280 cacaagccta aggccaccaa agaacaactg aagggcgtga tggacaactt cgccgccttt 2340 gtggaaaaat gctgcaaggc cgacgacaaa gaggcctgct tcgcagaaga gggccctaag 2400 tttgtggccg cctctcaagc tgctctggct taataaggta ccgatctttt tccctctgcc 2460

		aaaaattatg gggacatcat gaagcccctt gagcatctga cttctggcta ataaaggaaa 2520 tttattttca ttgcaatagt gtgttggaat tttttgtgtc tctcactcg 2569
3-51	**Sequences**	
3-51-1	Sequence Number [ID]	51
3-51-2	Molecule Type	DNA
3-51-3	Length	1513
3-51-4	Features Location/	**misc_feature 1..1513**
	Qualifiers	note=synthetic construct
		source 1..1513
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-51-5	Residues	acgcgtgcta gctaatgatg ggcgctcgag taatgatggg cggtcgacta atgatgggcg 60 ctcgagtaat gatgggcgtc tagctaatga tgggcgctcg agtaatgatg ggcggtcgac 120 taatgatggg cgctcgagta atgatgggcg tctagctaat gatgggcgct cgagtaatga 180 tgggcggtcg actaatgatg ggcgctcgag taatgatggg cgtctagaac gcgaattaat 240 tcaacatttt gacaccccca taatattttt ccagaattaa cagtataaat tgcatctctt 300 gttcaagagt tccctatcac tctctttaat cactactcac agtaacctca actcctgcca 360 caagcttgcc ctgcagcggg aattcgccct gcagcgggaa ttgccaccat ggctcacgtt 420 cggggactgc agctgcctgg atgtctggct cttgccgctc tgtgtagcct ggtgcacagc 480 cagcacgtgt ttctggctcc tcagcaagcc agatcactgc tgcagagagt tagaaggcac 540 ggcgacggca gcttcagcga cgagatgaac accatcctgg acaacctggc cgccagagac 600 ttcatcaact ggctgatcca gaccaagatc accgacggtg gcggaggcgg aggatctggt 660 ggtggtggat ctggcggcgg aggttctgcc gagtctaagt acggacctcc ttgtcctccc 720 tgtcctgctc cagaagctgc tggcggccca tccgtgtttc tgttccctcc aaagcctaag 780 gacaccctga tgatcagcag aacccctgaa gtgacctgcg tggtggtcga cgtgtcccaa 840 gaggatcctg aggtgcagtt caattggtac gtggacggcg tggaagtgca caacgccaag 900 accaagccta gagaggaaca gttcaacagc acctacagag tggtgtccgt gctgaccgtg 960 ctgcaccagg attggctgaa cggcaaagag tacaagtgca aggtgtccaa caagggcctg 1020 cctagcagca tcgagaaaac catcagcaag gccaagggcc agccaagaga accccaggtg 1080 tacacactgc ctccaagcca agaggaaatg accaagaacc aggtgtccct gacctgcctg 1140 gtcaagggct tctacccttc cgatatcgcc gtggaatggg agagcaacgg ccagcctgag 1200 aacaactaca agaccacacc tcctgtgctg gactccgatg gctcattctt cctgtacagc 1260 agactgaccg tggacaagag caggtggcaa gagggcaacg tgttcagctg cagcgtgatg 1320 cacgaggccc tgcacaacca ctacacccag aaaagcctga gcctgtctct gggctaataa 1380 ggtaccgatc tttttccctc tgccaaaaat tatggggaca tcatgaagcc ccttgagcat 1440 ctgacttctg gctaataaag gaaatttatt ttcattgcaa tagtgtgttg gaatttttg 1500 tgtctctcac tcg 1513
3-52	**Sequences**	
3-52-1	Sequence Number [ID]	52
3-52-2	Molecule Type	DNA
3-52-3	Length	2566
3-52-4	Features Location/	**misc_feature 1..2566**
	Qualifiers	note=synthetic construct
		source 1..2566
		mol_type=other DNA
		organism=synthetic construct
	NonEnglishQualifier Value	
3-52-5	Residues	acgcgtgcta gctaatgatg ggcgctcgag taatgatggg cggtcgacta atgatgggcg 60 ctcgagtaat gatgggcgtc tagctaatga tgggcgctcg agtaatgatg ggcggtcgac 120 taatgatggg cgctcgagta atgatgggcg tctagctaat gatgggcgct cgagtaatga 180 tgggcggtcg actaatgatg ggcgctcgag taatgatggg cgtctagaac gcgaattaat 240 tcaacatttt gacaccccca taatattttt ccagaattaa cagtataaat tgcatctctt 300 gttcaagagt tccctatcac tctctttaat cactactcac agtaacctca actcctgcca 360 caagcttgcc ctgcagcggg aattccagaa gccaccatgg ctcacgttcg gggactgcag 420 ctgcctggat gtctggctct tgccgctctg tgtagcctgg tgcacagcca gcacgtgttt 480 ctggctcctc agcaagccag atcactgctg cagagagtta gaaggcacgg cgacggcagc 540 ttcagcgacg agatgaacac catcctggac aacctggccg ccagagactt catcaactgg 600 ctgatccaga ccaagatcac cgacggtggc ggaggcggag gatctggtgg tggtggatct 660 ggcggcggag gttctgacgc ccacaaatct gaagtggccc accggttcaa ggacctgggc 720 gaagagaatt tcaaggccct ggtgctgatc gccttcgctc agtacctgca gcagtgcccc 780 ttcgaggacc acgtgaagct ggtcaacgaa gtgaccgagt tcgccaagac ctgcgtggcc 840 gacgagagcg ccgagaactg tgataagagc ctgcacaccc tgttcggcga caagctgtgt 900 acagtggcca cactgagaga aacctacggc gagatggccg actgctgcgc caagcaagag 960 cccgagagaa acgagtgctt cctgcagcac aaggacgaca accccaacct gcctagactc 1020

		gtgcggcctg aagtggacgt gatgtgcacc gccttccacg acaacgagga aaccttcctg 1080 aagaagtacc tgtacgagat cgccagacgg caccctact tttacgcccc tgagctgctg 1140 ttcttcgcca agcggtataa ggccgccttc accgagtgtt gtcaggccgc tgataaggct 1200 gcctgcctgc tgcctaagct ggacgagctt agagacgagg gcaaagccag ctccgccaag 1260 cagagactga agtgtgccag cctgcagaag ttcggcgaga gagcctttaa ggcctgggcc 1320 gttgctagac tgagccagag atttcccaag gccgagtttg ccgaggtgtc caagctcgtg 1380 accgacctga caaaggtgca caccgagtgc tgccacggcg acctgctgga atgcgccgac 1440 gatagagccg acctggccaa gtacatctgc gagaaccagg acagcatcag cagcaagctg 1500 aaagagtgct gcgagaagcc tctgctggaa aagagccact gtatcgccga ggtggaaaac 1560 gacgagatgc ccgccgatct gccttctctg gccgccgatt ttgtggaaag caaggacgtg 1620 tgcaagaact acgccgaggc caaggacgtg ttcctgggca tgtttctgta cgagtacgcc 1680 cgcagacacc ccgactactc tgttgtgctg ctgctgagac tggccaaaac ctacgagaca 1740 accctggaaa agtgctgtgc cgccgctgat cctcacgagt gttacgccaa ggtgttcgac 1800 gagttcaagc cactggtgga agaaccccag aacctgatca agcagaactg cgagctgttc 1860 gagcagctgg gcgagtacaa gttccagaac gccctgctcg tgcggtacac caagaaggtg 1920 ccccaggttt ccacacctac actggttgag gtgtcccgga acctgggcaa agtgggcagc 1980 aagtgttgca agcaccctga ggccaagaga atgccctgcg ccgaggatta cctgagcgtg 2040 gtgctgaatc agctgtgcgt gctgcacgag aaaaccctg tgtccgacag agtgaccaag 2100 tgctgtaccg agagcctggt caacagacgg ccttgcttta gcgccctcga ggtggacgag 2160 acatacgtgc ccaaagagtt caacgccgag acattcacct tccacgccga catctgtacc 2220 ctgagcgaga aagagcggca gatcaagaaa cagactgccc tggtggaact ggtcaagcac 2280 aagcccaagg ccaccaaaga acagctgaag gccgtgatgg acgacttcgc cgccttcgtg 2340 gaaaagtgct gcaaggccga cgacaaagag acctgcttcg ccgaagaggg caagaaactg 2400 gtggccgctt ctcaggctgc tctgggataa taaggtaccg atctttttcc ctctgccaaa 2460 aattatgggg acatcatgaa gccccttgag catctgactt ctggctaata aaggaaattt 2520 attttcattg caatagtgtg ttggaatttt ttgtgtctct cactcg 2566
3-53	**Sequences**	
3-53-1	Sequence Number [ID]	53
3-53-2	Molecule Type	DNA
3-53-3	Length	2211
3-53-4	Features Location/	**source 1..2211**
	Qualifiers	mol_type=unassigned DNA
		organism=Adeno-Associated Virus hu68
	NonEnglishQualifier Value	
3-53-5	Residues	atggctgccg atggttatct tccagattgg ctcgaggaca acctcagtga aggcattcgc 60 gagtggtggg ctttgaaacc tggagcccct caacccaagg caaatcaaca acatcaagac 120 aacgctcggg gtcttgtgct tccgggttac aaataccttg gacccggcaa cggactcgac 180 aaggggagc cggtcaacga agcagacgcg gcggccctcg agcacgacaa ggcctacgac 240 cagcagctca aggccggaga caacccgtac ctcaagtaca accacgccga cgccgagttc 300 caggagcggc tcaaagaaga tacgtctttt gggggcaacc tcgggcgagc agtcttccag 360 gccaaaaaga ggcttcttga acctcttggt ctggttgagg aagcggctaa gacggctcct 420 ggaaagaaga ggcctgtaga gcagtctcct caggaaccgg actcctccgt gggtattggc 480 aaatcgggtg cacagcccgc taaaaagaga ctcaatttcg gtcagactgg cgacacagag 540 tcagtccccg accctcaacc aatcggagaa cctcccgcag ccccctcagg tgtgggatct 600 cttacaatgg cttcaggtgg tggcgcacca gtggcagaca ataacgaagg tgccgatgga 660 gtgggtagtt cctcgggaaa ttggcattgc gattcccaat ggctgggga cagagtcatc 720 accaccagca cccgaacctg ggccctgccc acctacaaca atcacctcta caagcaaatc 780 tccaacagca catctggagg atcttcaaat gacaacgcct acttcggcta cagcaccccc 840 tgggggtatt ttgacttcaa cagattccac tgccacttct caccacgtga ctggcaaaga 900 ctcatcaaca acaactgggg attccggcct aagcgactca acttcaagct cttcaacatt 960 caggtcaaag aggttacgga caacaatgga gtcaagacca tcgctaataa ccttaccagc 1020 acggtccagg tcttcacgga ctcagactat cagctcccgt acgtgctcgg gtcggctcac 1080 gagggctgcc tcccgccgtt cccagcggac gttttcatga ttcctcagta cgggtatcta 1140 acgcttaatg atggaagcca agccgtgggt cgttcgtcct tttactgcct ggaatatttc 1200 ccgtcgcaaa tgctaagaac gggtaacaac ttccagttca gctacgagtt tgagaacgta 1260 cctttccata gcagctatgc tcacagccaa agcctggacc gactcatgaa tccactcatc 1320 gaccaatact tgtactatct ctcaaagact attaacggtt ctggacagaa tcaacaaacg 1380 ctaaaattca gtgtggccgg acccagcaac atggctgtcc agggaagaaa ctacatacct 1440 ggacccagct accgacaaca acgtgtctca accactgtga ctcaaaacaa caacagcgaa 1500 tttgcttggc ctggagcttc ttcttgggct ctcaatggac gtaatagctt gatgaatcct 1560 ggacctgcta tggccagcca caaagaagga gaggaccgtt tctttccttt gtctggatct 1620 ttaatttttg gcaaacaagg aactggaaga gacaacgtgg atgcggacaa agtcatgata 1680 accaacgaag aagaaattaa aactaccaac ccagtagcaa cggagtccta tggacaagtg 1740 gccacaaacc accagagtgc ccaagcacag gcgcagaccg gctgggttca aaaccaagga 1800 atacttccgg gtatggtttg gcaggacaga gatgtgtacc tgcaaggacc catttgggcc 1860

		aaaattcctc acacggacgg caactttcac ccttctccgc tgatgggagg gtttggaatg 1920 aagcacccgc ctcctcagat cctcatcaaa aacacacctg tacctgcgga tcctccaacg 1980 gctttcaaca aggacaagct gaactctttc atcacccagt attctactgg ccaagtcagc 2040 gtggagattg agtgggagct gcagaaggaa aacagcaagc gctggaaccc ggagatccag 2100 tacacttcca actattacaa gtctaataat gttgaatttg ctgttaatac tgaaggtgtt 2160 tattctgaac cccgccccat tggcaccaga tacctgactc gtaatctgta a 2211
3-54	**Sequences**	
3-54-1	Sequence Number [ID]	54
3-54-2	Molecule Type	DNA
3-54-3	Length	2211
3-54-4	Features Location/	**source 1..2211**
	Qualifiers	mol_type=unassigned DNA
		organism=Adeno-Associated Virus hu68
	NonEnglishQualifier Value	
3-54-5	Residues	atggctgccg atggttatct tccagattgg ctcgaggaca acctcagtga aggcattcgc 60 gagtggtggg ctttgaaacc tggagcccct caacccaagg caaatcaaca acatcaagac 120 aacgctcggg gtcttgtgct tccgggttac aaataccttg gacccggcaa cggactcgac 180 aaggggggagc cggtcaacga agcagacgcg gcggccctcg agcacgacaa ggcctacgac 240 cagcagctca aggccggaga caacccgtac ctcaagtaca accacgccga cgccgagttc 300 caggagcggc tcaaagaaga tacgtctttt gggggcaacc tcgggcgagc agtcttccag 360 gccaaaaaga ggcttcttga acctcttggt ctggttgagg aagcggctaa gacggctcct 420 ggaaagaaga ggcctgtaga gcagtctcct caggaaccgg actcctccgt gggtattggc 480 aaatcgggtg cacagcccgc taaaaagaga ctcaatttcg gtcagactgg cgacacagag 540 tcagtccccg accctcaacc aatcggagaa cctcccgcag ccccctcagg tgtgggatct 600 cttacaatgg cttcaggtgg tggcgcacca gtggcagaca ataacgaagg tgccgatgga 660 gtgggtagtt cctcgggaaa ttggcattgc gattcccaat ggctggggga cagagtcatc 720 accaccagca cccgaacctg ggccctgccc acctacaaca atcacctcta caagcaaatc 780 tccaacagca catctggagg atcttcaaat gacaacgcct acttcggcta cagcaccccc 840 tgggggtatt ttgacttcaa cagattccac tgccacttct caccacgtga ctggcaaaga 900 ctcatcaaca acaactgggg attccggcct aagcgactca acttcaagct cttcaacatt 960 caggtcaaag aggttacgga caacaatgga gtcaagacca tcgctaataa ccttaccagc 1020 acggtccagg tcttcacgga ctcagactat cagctcccgt acgtgctcgg gtcggctcac 1080 gagggctgcc tcccgccgtt cccagcggac gttttcatga ttcctcagta tggatacctc 1140 accctgaacg acggcagtca ggcggtgggc cgctcatcct tctactgcct ggagtacttc 1200 ccttcgcaga tgctgaggac tggcaacaac ttccagttca gctacgagtt cgagaacgtc 1260 cctttccaca gcagctacgc ccacagccag agtttggacc gcttgatgaa ccctctgatc 1320 gaccagtacc tgtactacct gtcaaagacg atcaacggtt ctggccagaa ccagcagacg 1380 ctgaagttca gcgtggccgg gcctagcaac atggccgtcc agggcagaaa ctacatccct 1440 gggcccagct accggcagca gagagtctca accactgtga ctcagaacaa caacagtgag 1500 ttcgcctggc ctggcgccag ctcttgggcc ctcaacggcc gcaactcgct gatgaaccca 1560 ggcccagcca tggccagtca caaggagggc gaggaccgtt tcttcccttt gtctggctct 1620 ctgatcttcg gcaagcaggg gaccggcaga gacaacgtgg acgcggacaa ggtcatgatc 1680 acgaacgagg aggagatcaa gaccaccaac cctgtggcaa ccgagtccta cggccaggtg 1740 gcaaccaacc accagagcgc ccaggcacag gcgcagactg gctgggtcca gaaccagggg 1800 atcctgcctg gcatggtgtg gcaggaccgt gacgtgtacc tgcagggccc tatctgggca 1860 aagatccctc acacggacgg caacttccac ccttctcctc tgatgggcgg cttcggcatg 1920 aagcacccgc ctcctcagat cctcatcaag aacactccgg tcccggcaga ccctccgacg 1980 gccttcaaca aggacaagct gaactcattc atcactcagt actccactgg ccaggtcagc 2040 gtggagatcg agtgggagct gcagaaggag aacagcaagc gttggaaccc agagatccag 2100 tacacttcca actactacaa gtctaacaac gtggagttcg ccgtcaacac tgagggtgtg 2160 tacagtgagc ctcgccctat cggcacccgg tacctcaccc gaaacttgtg a 2211
3-55	**Sequences**	
3-55-1	Sequence Number [ID]	55
3-55-2	Molecule Type	AA
3-55-3	Length	736
3-55-4	Features Location/	**source 1..736**
	Qualifiers	mol_type=protein
		organism=Adeno-Associated Virus hu68
	NonEnglishQualifier Value	
3-55-5	Residues	MAADGYLPDW LEDNLSEGIR EWWALKPGAP QPKANQQHQD NARGLVLPGY KYLGPGNGLD 60 KGEPVNEADA AALEHDKAYD QQLKAGDNPY LKYNHADAEF QERLKEDTSF GGNLGRAVFQ 120 AKKRLLEPLG LVEEAAKTAP GKKRPVEQSP QEPDSSVGIG KSGAQPAKKR LNFGQTGDTE 180 SVPDPQPIGE PPAAPSGVGS LTMASGGGAP VADNNEGADG VGSSSGNWHC DSQWLGDRVI 240 TTSTRTWALP TYNNHLYKQI SNSTSGGSSN DNAYFGYSTP WGYFDFNRFH CHFSPRDWQR 300 LINNNWGFRP KRLNFKLFNI QVKEVTDNNG VKTIANNLTS TVQVFTDSDY QLPYVLGSAH 360

		EGCLPPFPAD VFMIPQYGYL TLNDGSQAVG RSSFYCLEYF PSQMLRTGNN FQFSYEFENV	420
		PFHSSYAHSQ SLDRLMNPLI DQYLYYLSKT INGSGQNQQT LKFSVAGPSN MAVQGRNYIP	480
		GPSYRQQRVS TTVTQNNNSE FAWPGASSWA LNGRNSLMNP GPAMASHKEG EDRFFPLSGS	540
		LIFGKQGTGR DNVDADKVMI TNEEEIKTTN PVATESYGQV ATNHQSAQAQ AQTGWVQNQG	600
		ILPGMVWQDR DVYLQGPIWA KIPHTDGNFH PSPLMGGFGM KHPPPQILIK NTPVPADPPT	660
		AFNKDKLNSF ITQYSTGQVS VEIEWELQKE NSKRWNPEIQ YTSNYYKSNN VEFAVNTEGV	720
		YSEPRPIGTR YLTRNL	736