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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 June 2022 (30.06.2022)

WIPO | PCT

(10) International Publication Number
WO 2022/140560 A1

(51) International Patent Classification:
*A61K 35/76* (2015.01) *C12N 15/11* (2006.01)
*C12N 7/00* (2006.01) *C12N 15/86* (2006.01)

(21) International Application Number:
PCT/US2021/064887

(22) International Filing Date:
22 December 2021 (22.12.2021)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/130,360 23 December 2020 (23.12.2020) US
63/147,064 08 February 2021 (08.02.2021) US

(71) Applicant: **FLAGSHIP PIONEERING INNOVATIONS V, INC.** [US/US]; 55 Cambridge Parkway, 8th Floor, Cambridge, MA 02142 (US).

(72) Inventors: **HAJJAR, Roger Joseph**; 55 Cambridge Parkway, 8th Floor, Cambridge, MA 02142 (US). **DELAGRAVE, Simon**; 55 Cambridge Parkway, 8th Floor, Cambridge, MA 02142 (US). **SWANSON, Kurt Adam**; 55 Cambridge Parkway, 8th Floor, Cambridge, MA 02142 (US).

(74) Agent: **KOYFMAN, Hannah, R.** et al.; K&L Gates LLP, One Lincoln Street, State Street Financial Center, Boston, MA 02111 (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, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, 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, GH, GM, KE, LR, LS, MW, MZ, NA, RW, 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, 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).

**Declarations under Rule 4.17:**
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

**Published:**
— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) **Title:** IN VITRO ASSEMBLY OF ANELLOVIRUS CAPSIDS ENCLOSING RNA

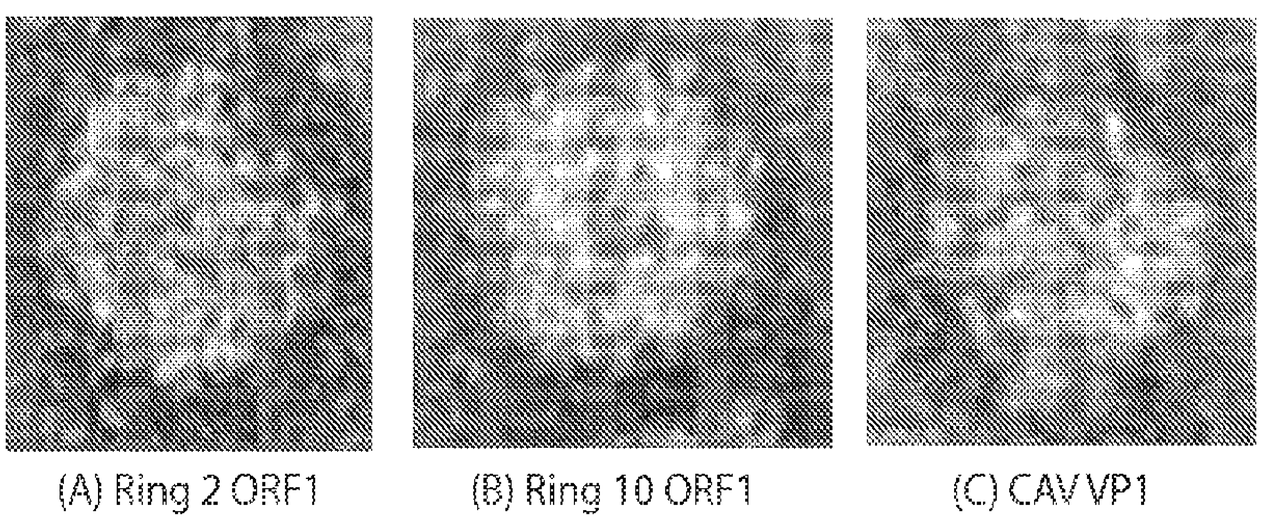



FIG. 1

(57) **Abstract:** This invention relates generally to compositions for making anellovectors and uses thereof.

WO 2022/140560 A1

# IN VITRO ASSEMBLY OF ANELLOVIRUS CAPSIDS ENCLOSING RNA

## CROSS REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application No. 63/130,360, filed December 23, 2020 and U.S. Provisional Application No. 63/147,064, filed February 8, 2021. The contents of the aforesaid applications are hereby incorporated by reference in their entirety.

## SEQUENCE LISTING

The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on, December 21, 2021, is named V2057-7017WO_SL.txt and is 286,720 bytes in size.

## BACKGROUND

There is an ongoing need to develop compositions and methods for making suitable vectors to deliver therapeutic effectors to patients.

## SUMMARY

The present disclosure provides compositions and methods for producing an anellovector (e.g., a synthetic anellovector) that can be used as a delivery vehicle, e.g., for delivering genetic material, for delivering an effector, e.g., a payload, or for delivering a therapeutic agent or a therapeutic effector to a eukaryotic cell (e.g., a human cell or a human tissue). While naturally occurring *Anellovirus* has a DNA genome, the present disclosure provides anellovectors with a genetic element that comprises RNA.

An anellovector (*e.g.*, produced using a composition or method as described herein) generally comprises a genetic element (*e.g.*, a genetic element comprising or encoding an effector, *e.g.*, an exogenous or endogenous effector, *e.g.*, a therapeutic effector) encapsulated in a proteinaceous exterior (*e.g.*, a proteinaceous exterior comprising an *Anellovirus* capsid protein, *e.g.*, an Anellovirus ORF1 protein or a polypeptide encoded by an *Anellovirus* ORF1 nucleic acid, *e.g.*, as described herein), which is capable of introducing the genetic element into a cell (*e.g.*, a mammalian cell, *e.g.*, a human cell). The genetic element may comprise RNA. In some embodiments, the anellovector is an infectious vehicle or particle comprising a proteinaceous exterior comprising a polypeptide encoded by an *Anellovirus* ORF1 nucleic acid (*e.g.*, an ORF1 nucleic acid of *Alphatorquevirus, Betatorquevirus,* or *Gammatorquevirus*, *e.g.*, an ORF1 of *Alphatorquevirus* clade 1, *Alphatorquevirus* clade 2, *Alphatorquevirus* clade 3, *Alphatorquevirus* clade 4, *Alphatorquevirus* clade 5, *Alphatorquevirus* clade 6, or *Alphatorquevirus* clade 7, *e.g.*, as described herein). The genetic element of an anellovector of the present disclosure is typically

a circular and/or single-stranded RNA molecule (*e.g.*, circular and single stranded) having a protein binding sequence that binds to the proteinaceous exterior enclosing it, or a polypeptide attached thereto, which may facilitate enclosure of the genetic element within the proteinaceous exterior and/or enrichment of the genetic element, relative to other nucleic acids, within the proteinaceous exterior. In some embodiments, the genetic element of an anellovector is produced using a baculovirus, nucleic acid construct (*e.g.*, a bacmid and/or donor vector), insect cell, and/or animal cell line, *e.g.*, as described herein.

In some instances, the genetic element comprises or encodes an effector (*e.g.*, comprises a nucleic acid effector, such as a non-coding RNA, or encodes a polypeptide effector, *e.g.*, a protein), *e.g.*, which can be expressed in a target cell. In some embodiments, the effector is a therapeutic agent or a therapeutic effector, *e.g.*, as described herein. In some embodiments, the effector is an endogenous effector or an exogenous effector, *e.g.*, exogenous to a wild-type *Anellovirus* or a target cell. In some embodiments, the effector is exogenous to a wild-type *Anellovirus* or a target cell. In some embodiments, the anellovector can deliver an effector into a cell by contacting the cell and introducing a genetic element encoding the effector into the cell, such that the effector is made or expressed by the cell. In certain instances, the effector is an endogenous effector (*e.g.*, endogenous to the target cell but, e.g., provided in increased amounts by the anellovector). In other instances, the effector is an exogenous effector. The effector can, in some instances, modulate a function of the cell or modulate an activity or level of a target molecule in the cell. For example, the effector can decrease levels of a target protein in the cell (*e.g.*, as described in Examples 3 and 4 of PCT/US19/65995). In another example, the anellovector can deliver and express an effector, *e.g.*, an exogenous protein, *in vivo* (*e.g.*, as described in Examples 19 and 28 of PCT/US19/65995). Anellovectors can be used, for example, to deliver genetic material to a target cell, tissue or subject; to deliver an effector to a target cell, tissue or subject; or for treatment of diseases and disorders, *e.g.*, by delivering an effector that can operate as a therapeutic agent to a desired cell, tissue, or subject.

In some embodiments, the compositions and methods described herein can be used to produce the genetic element of a synthetic anellovector, *e.g.*, in a host cell. A synthetic anellovector has at least one structural difference compared to a wild-type virus (*e.g.*, a wild-type *Anellovirus*, e.g., a described herein), *e.g.*, a deletion, insertion, substitution, modification (*e.g.*, enzymatic modification), relative to the wild-type virus. Generally, synthetic anellovectors include an exogenous genetic element enclosed within a proteinaceous exterior, which can be used for delivering the genetic element, or an effector (*e.g.*, an exogenous effector or an endogenous effector) encoded therein (*e.g.*, a polypeptide or nucleic acid effector), into eukaryotic (*e.g.*, human) cells. In embodiments, the anellovector does not cause a detectable and/or an unwanted immune or inflammarory response, *e.g.*, does not cause more than a 1%,

5%, 10%, 15% increase in a molecular marker(s) of inflammation, *e.g.*, TNF-alpha, IL-6, IL-12, IFN, as well as B-cell response *e.g.* reactive or neutralizing antibodies, *e.g.*, the anellovector may be substantially non-immunogenic to the target cell, tissue or subject.

In some embodiments, the compositions and methods described herein can be used to produce the genetic element of an anellovector comprising: (i) a proteinaceous exterior comprising an ORF1 molecule; and (ii) a genetic element comprising RNA; wherein the genetic element is enclosed within the proteinaceous exterior. In some embodients, the genetic element consists of at least 10%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% RNA. In some embodiments, the genetic element does not comprise DNA. In some embodiments, the genetic element does not comprise ssDNA. Alternatively, in some embodiments, the genetic element comprises a DNA region. In some embodiments, a DNA or RNA molecule described herein comprises one or more modified nucleotides (e.g., a base modification, sugar modification, or backbone modification). In some embodiments, the genetic element is a single-stranded. In some embodiments, the genetic element comprises a double stranded region. In some embodiments the genetic element is a linear polypeptide. Alternatively, in some embodiments, the genetic element is a circular polynucleotide. In some embodiments, the nucleic acid sequence is codon-optimized, e.g., for expression in an insect cell. In some embodiments, at least 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of the codons in the nucleic acid sequence are codon-optimized, e.g., for expression in an insect cell. In some embodiments, the nucleic acid sequence is codon-optimized, e.g., for expression in a mammalian (e.g., human) cell. In some embodiments, at least 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of the codons in the nucleic acid sequence are codon-optimized, e.g., for expression in a mammalian (e.g., human) cell. In some embodiments, the genetic element is about 10-20, 20-30, 30-40, 50-60 60-70, 70-80, 80-90, 90-100, 100-125, 125-150, 150-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, 900-1000, 1000-1500, 1500-2000, 2000-2500, 2500-3000, 3000-3500, 3500-4000, or 4000-4500 nucleotides in length. In some embodiments, the genetic element is at least about 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000, or 4500 nucleotides in length.

In some embodiments, the compositions and methods described herein can be used to produce the genetic element of an infectious (e.g., to a human cell) Annellovector, vehicle, or particle comprising a capsid (e.g., a capsid comprising an Anellovirus ORF, e.g., ORF1, polypeptide) encapsulating a genetic element comprising a protein binding sequence that binds to the capsid and a heterologous (to the Anellovirus) sequence encoding a therapeutic effector. In embodiments, the Anellovector is capable of delivering the genetic element into a mammalian, e.g., human, cell.

In an aspect, the invention features a method of making an anellovector by *in vitro* assembly. In some embodiments, a method of making an anellovector comprises: (a) providing a mixture comprising: (i) a genetic element comprising RNA, and (ii) an ORF1 molecule; and (b) incubating the mixture under conditions suitable for enclosing the genetic element within a proteinacous comprising the ORF1 molecule, thereby making an anellovector; optionally wherein the mixture is not comprised in a cell. In some embodiments, a method further comprises, prior to the providing of (a), expressing the ORF1 molecule, e.g., in a host cell (e.g., an insect cell or a mammalian cell). In some embodiments, the expressing comprises incubating a host cell (e.g., an insect cell or a mammalian cell) comprising a nucleic acid molecule (e.g., a baculovirus expression vector) encoding the ORF1 molecule under conditions suitable for producing the ORF1 molecule. In some embodiments, a method further comprises, prior to the providing of (a), purifying the ORF1 molecule expressed by the host cell.

In some embodiments, anellovectors, as described herein, can be used as effective delivery vehicles for introducing an agent, such as an effector described herein, to a target cell, e.g., a target cell in a subject to be treated therapeutically or prophylactically.

In an aspect, the invention features a pharmaceutical composition comprising an anellovector (e.g., a synthetic anellovector) as described herein. In some embodiments, the pharmaceutical composition further comprises a pharmaceutically acceptable carrier or excipient. In embodiments, the pharmaceutical composition comprises a unit dose comprising about $10^5$-$10^{14}$ (e.g., about $10^6$-$10^{13}$, $10^7$-$10^{12}$, $10^8$-$10^{11}$, or $10^9$-$10^{10}$) genome equivalents of the anellovector per kilogram of a target subject. In some embodiments, the pharmaceutical composition comprising the preparation will be stable over an acceptable period of time and temperature, and/or be compatible with the desired route of administration and/or any devices this route of administration will require, e.g., needles or syringes. In some embodiments, the pharmaceutical composition is formulated for administration as a single dose or multiple doses. In some embodiments, the pharmaceutical composition is formulated at the site of administration, e.g., by a healthcare professional. In some embodiments, the pharmaceutical composition comprises a desired concentration of anellovector genomes or genomic equivalents (e.g., as defined by number of genomes per volume).

In an aspect, the invention features a method of treating a disease or disorder in a subject, the method comprising administering to the subject an anellovector, e.g., a synthetic anellovector, e.g., as described herein.

In an aspect, the invention features a method of delivering an effector or payload (e.g., an endogenous or exogenous effector) to a cell, tissue or subject, the method comprising administering to the subject an anellovector, e.g., a synthetic anellovector, e.g., as described herein, wherein the anellovector

comprises a nucleic acid sequence encoding the effector. In some embodiments, the payload is a nucleic acid. In some embodiments, the payload is a polypeptide.

In an aspect, the invention features a method of delivering an anellovector to a cell, comprising contacting the anellovector, e.g., a synthetic anellovector, e.g., as described herein, with a cell, e.g., a eukaryotic cell, e.g., a mammalian cell, e.g., in vivo or ex vivo.

Additional features of any of the aforesaid anellovectors, compositions or methods include one or more of the following enumerated embodiments.

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following enumerated embodiments.

**Enumerated Embodiments**

1. An anellovector comprising:

a) proteinaceous exterior comprising an ORF1 molecule;

b) a genetic element comprising RNA,

wherein the genetic element is enclosed within the proteinaceous exterior.

2. The anellovector of embodiment 1, wherein the genetic element consists of at least 10%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% RNA.

3. The anellovector of embodiment 1 or 2, wherein the RNA comprises one or more chemical modifications.

4. The anellovector of any of the preceding embodiments, wherein the genetic element consists of or consists essentially of RNA.

5. The anellovector of any of the preceding embodiments, wherein the anellovector does not comprise DNA.

6. The anellovector of any of the preceding embodiments, wherein the anellovector does not comprise ssDNA.

7. The anellovector of any of the preceding embodiments, wherein the genetic element comprises a DNA region.

8. The anellovector of any of the preceding embodiments, wherein all nucleotides of the DNA region are chemically modified.

9. The anellovector of embodiment 7 or 8, wherein the DNA region is covalently linked to the RNA of the genetic element.

10. The anellovector of any of the preceding embodiments, wherein at least a portion of the DNA region hybridizes to at least a portion of the RNA of the genetic element.

11. The anellovector of any of the preceding embodiments, wherein the DNA region is 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, 150-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, or 900-1000 nucleotides in length.

12. The anellovector of any of the preceding embodiments, wherein the genetic element is single stranded.

13. The anellovector of any of the preceding embodiments, wherein the genetic element comprises a double stranded region (e.g., a region of RNA pairing with RNA or a DNA pairing with RNA).

14. The anellovector of any of the preceding embodiments, wherein the genetic element is linear.

15. The anellovector of any of the preceding embodiments, wherein the genetic element is circular.

16. The anellovector of any of the preceding embodiments, wherein the genetic element comprises a first region and a second region that can hybridize with the first region.

17. The anellovector of any of the preceding embodiments, wherein the genetic element does not comprise a 5' end or a 3' end.

18. The anellovector of any of embodiments 15-17, wherein the genetic element does not comprise one or both of a free phosphate and a free sugar.

19. The anellovector of any of embodiments 15-18, wherein every phosphate in the genetic element is covalently linked to a first sugar by a first oxygen atom comprised by the phosphate and a second sugar by a second oxygen atom comprised by the phosphate.

20. The anellovector of any of embodiments 15-19, wherein every sugar in the genetic element is covalently linked to a first phosphate by a first carbon atom comprised by the sugar and a second phosphate by a second carbon atom comprised by the sugar.

21. The anellovector of any of embodiments 15-20, wherein the genetic element was produced by circularizing a linear RNA.

22. The anellovector of any of the preceding embodiments, wherein the genetic element is about 10-20, 20-30, 30-40, 50-60 60-70, 70-80, 80-90, 90-100, 100-125, 125-150, 150-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, 900-1000, 1000-1500, 1500-2000, 2000-2500, 2500-3000, 3000-3500, 3500-4000, or 4000-4500 nucleotides in length.

23. The anellovector of any of the preceding embodiments, wherein the genetic element is at least about 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000, or 4500 nucleotides in length.

24. The anellovector of any of the preceding embodiments, wherein the genetic element binds the ORF1 molecule.

25. The anellovector of any of the preceding embodiments, wherein the genetic element binds a jelly roll domain comprised by the ORF1 molecule.

26. The anellovector of any of the preceding embodiments, wherein the genetic element binds an arginine-rich domain comprised by the ORF1 molecule.

27. The anellovector of any of the preceding embodiments, wherein the anellovector comprises a plurality of genetic elements, e.g., at least 2, 3, 4, 5, 10, 20, 30, 40, 50, or 60 genetic elements.

28. The anellovector of embodiment 27, wherein the genetic elements of the plurality each comprise the same sequence.

29. The anellovector of embodiment 27, wherein the genetic elements of the plurality comprise different sequences.

30. The anellovector of any of the preceding embodiments, wherein the genetic element encodes an exogenous effector.

31. The anellovector of embodiment 30, wherein the sequence encoding the exogenous effector is at least about 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, or 3000 nucleotides in length.

32. The anellovector of any of embodiments 30-31, wherein the exogenous effector comprises a therapeutic effector (e.g., a polypeptide or a nucleic acid molecule).

33. The anellovector of any of embodiments 30-32, wherein the exogenous effector comprises a human protein, or a polypeptide comprising an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

34. The anellovector of any of embodiments 30-33, wherein the exogenous effector comprises a nucleic acid molecule.

35. The anellovector of any of embodiments 30-34, wherein the exogenous effector comprises a noncoding nucleic acid molecule, e.g., a functional RNA, e.g., an mRNA, miRNA, or siRNA).

36. The anellovector of any of embodiments 30-35, wherein the genetic element is an mRNA molecule encoding the exogenous effector (e.g., a peptide or polypeptide, e.g., a therapeutic peptide or polypeptide).

37. The anellovector of embodiment 36, wherein the noncoding nucleic acid molecule is a human noncoding nucleic acid molecule, or a nucleic acid molecule comprising a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

38. The anellovector of any of embodiments 30-37, wherein the exogenous effector comprises a cytosolic polypeptide or cytosolic peptide (e.g., a DPP-4 inhibitor, an activator of GLP-1 signaling, or an inhibitor of neutrophil elastase, or a functional fragment thereof).

39. The anellovector of embodiment 38, wherein the exogenous effector comprises a regulatory intracellular polypeptide.

40. The anellovector of any of embodiments 30-39, wherein the exogenous effector comprises a secreted polypeptide or peptide (e.g., a cytokine, an antibody molecule, a hormone, a growth factor, or a clotting-associated factor, or a functional fragment thereof).

41. The anellovector of any of embodiments 30-40, wherein the exogenous effector comprises a protein replacement therapeutic.

42. The anellovector of any of embodiments 30-41, wherein the exogenous effector comprises an enzyme.

43. The anellovector of any of embodiments 30-42, wherein the exogenous effector comprises erythropoietin (EPO) or human growth hormone (hGH), or a functional fragment thereof.

44. The anellovector of any of embodiments 30-43, wherein the exogenous effector comprises a component of a gene editing system (e.g., a component of a CRISPR system, e.g., a Cas9, Cpf1, or a functional fragment thereof).

45. The anellovector of any of the preceding embodiments, wherein the RNA comprises chemically modified RNA, e.g., as described herein.

46. The anellovector of any of the preceding embodiments, wherein the RNA comprises a cap.

47. The anellovector of any of the preceding embodiments, wherein the RNA comprises a poly-A tail, e.g., at least about 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, or 100 adenosines in length.

48. The anellovector of any of the preceding embodiments, wherein the RNA lacks a poly-A tail, e.g., comprises no more than 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, or 100 sequential adenosines.

49. The anellovector of any of the preceding embodiments, wherein the proteinaceous exterior comprises about 60 (e.g., about 40, 50, 60, 70, or 80) copies of the ORF1 molecule.

50. The anellovector of any of the preceding embodiments, wherein the jelly roll domains of the ORF1 molecules face the interior of the proteinaceous exterior.

51. The anellovector of any of the preceding embodiments, wherein the ORF1 molecule comprises an amino acid sequence as listed in any of Tables N-S and 37A-37C, or an amino acid sequence having at least 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

52. The anellovector of any of the preceding embodiments, wherein the ORF1 molecule comprises an arginine-rich region, e.g., comprising the amino acid sequence of an arginine-rich region as listed in any of Tables N-S and 37A-37C, or an amino acid sequence having at least 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

53. The anellovector of embodiment 52, wherein the arginine-rich region comprises at least about 70% (e.g., at least about 70%, 75%, 80%, 85%, 90%, 95%, or 100%) basic residues (e.g., arginine or lysine).

54. The anellovector of any of the preceding embodiments, wherein the ORF1 molecule comprises a jelly roll domain, e.g., comprising the amino acid sequence of a jelly roll domain as listed in

any of Tables N-S and 37A-37C, or an amino acid sequence having at least 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

55. The anellovector of embodiment 54, wherein the jelly roll domain comprises one or more (e.g., 1, 2, 3, or 4) of the following characteristics:

(i) at least 30% (e.g., at least 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 90%, or more) of the amino acids of the jelly-roll domain are part of one or more β-sheets;

(ii) the secondary structure of the jelly-roll domain comprises at least four (e.g., at least 4, 5, 6, 7, 8, 9, 10, 11, or 12) β-strands; and/or

(iii) the tertiary structure of the jelly-roll domain comprises at least two (e.g., at least 2, 3, or 4) β-sheets; and/or

(iv) the jelly-roll domain comprises a ratio of β-sheets to α-helices of at least 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, or 10:1.

56. The anellovector of embodiment 54, wherein the jelly roll domain comprises two β-sheets, e.g., arranged in antiparallel orientation relative to each other.

57. The anellovector of embodiment 54, wherein the jelly roll domain comprises eight β-strands.

58. The anellovector of any of embodiments 52-57, wherein the jelly roll domain comprises a region having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the sequence of a D β-strand, e.g., as shown in FIG. 3.

59. The anellovector of embodiment 52, wherein the D β-strand comprises 1, 2, or 3, or more basic residues (e.g., arginine or lysine).

60. The anellovector of any of embodiments 52-59, wherein the jelly roll domain comprises a region having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the sequence of a G β-strand, e.g., as shown in FIG. 3.

61. The anellovector of embodiment 52, wherein the G β-strand comprises at least about 1, 2, or 3, or more basic residues (e.g., arginine or lysine).

62. The anellovector of any of the preceding embodiments, wherein the ORF1 molecule comprises an N22 domain, e.g., comprising the amino acid sequence of an N22 domain as listed in any of Tables N-S and 37A-37C, or an amino acid sequence having at least 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

63. The anellovector of any of the preceding embodiments, wherein the N22 domain comprises the amino acid sequence $YNPX^2DXGX^2N$, wherein $X^n$ is each independently a contiguous sequence of any *n* amino acids.

64. The anellovector of any of the preceding embodiments, wherein the N22 domain comprises a first beta strand and a second beta strand flanking the amino acid sequence $YNPX^2DXGX^2N$, e.g., wherein the first beta strand comprises the tyrosine (Y) residue of the amino acid sequence $YNPX^2DXGX^2N$ and/or wherein the second beta strand comprises the second asparagine (N) residue (from N to C) of the amino acid sequence $YNPX^2DXGX^2N$.

65. The anellovector of any of the preceding embodiments, wherein the ORF1 molecule comprises a C-terminal domain, e.g., comprising the amino acid sequence of a C-terminal domain as listed in any of Tables N-S and 37A-37C, or an amino acid sequence having at least 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

66. The anellovector of any of the preceding embodiments, wherein the genetic element lacks a sequence encoding an *Anellovirus* ORF1 protein (e.g., as described herein).

67. The anellovector of any of the preceding embodiments, wherein the genetic element lacks a sequence encoding an *Anellovirus* ORF2 protein (e.g., as described herein).

68. The anellovector of any of the preceding embodiments, wherein the genetic element lacks a sequence encoding an *Anellovirus* ORF3 protein (e.g., as described herein).

69. The anellovector of any of the preceding embodiments, wherein the anellovector is configured to deliver the genetic element to a cell (e.g., a eukaryotic cell, e.g., a mammalian cell, e.g., a human cell).

70. The anellovector of embodiment 69, wherein a population of at least 1000 of the anellovectors is capable of delivering at least about 100 copies (e.g., at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 5000, 10,000, 50,000, 100,000, 500,000, or 1,000,000 copies) of the genetic element into one or more of the cells.

71. The anellovector of embodiment 69 or 70, wherein a population of the anellovectors (e.g., at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 genome equivalents of the genetic element per cell) is capable of delivering the genetic element into at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, or more of a population of the cells.

72. The anellovector of any of embodiments 69-71, wherein a population of the anellovectors (e.g., at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 genome equivalents of the genetic element per cell) is capable of delivering at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, 8,000, $1 \times 10^4$, $1 \times 10^5$, $1 \times 10^6$, $1 \times 10^7$ or greater copies of the genetic element per cell to a population of the cells.

73. The anellovector of any of embodiments 69-72, wherein a population of the anellovectors (e.g., at least 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 genome equivalents of the genetic element per cell) is capable of delivering 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 5-10, 10-20, 20-50, 50-100, 100-1000, 1000-$10^4$, $1 \times 10^4$-$1 \times 10^5$, $1 \times 10^4$-$1 \times 10^6$, $1 \times 10^4$-$1 \times 10^7$, $1 \times 10^5$-$1 \times 10^6$, $1 \times 10^5$-$1 \times 10^7$, or $1 \times 10^6$-$1 \times 10^7$ copies of the genetic element per cell to a population of the cells.

74. The anellovector of any of the preceding embodiments, wherein the anellovector selectively delivers the effector to, or is present at higher levels in (e.g., preferentially accumulates in), a desired cell type, tissue, or organ (e.g., bone marrow, blood, heart, GI, skin, photoreceptors in the retina, epithelial linings, or pancreas).

75. The anellovector of any of the preceding embodiments, wherein the genetic element is protected from or resistant to digestion by an RNase (e.g., by the proteinaceous exterior).

76. The anellovector of any of the preceding embodiments, wherein the genetic element enclosed within the proteinaceous exterior is resistant to endonuclease digestion, e.g., to RNase digestion.

77. The anellovector of any of the preceding embodiments, wherein the genetic element comprises a promoter element.

78. The anellovector of any of the preceding embodiments, wherein the genetic element comprises a protein binding sequence.

79. The anellovector of embodiment 78, wherein the protein binding sequence is capable of binding to the ORF1 molecule.

80. A composition comprising a plurality of the anellovectors of any of the preceding embodiments.

81. The composition of embodiment 80, wherein at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of proteinaceous exteriors comprising an ORF1 molecule in the composition comprise at least one copy of the genetic element.

82. The composition of embodiment 80, wherein at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of proteinaceous exteriors comprising an ORF1 molecule in the composition comprise at least one copy of an anellovector genetic element.

83. The composition of any of embodiments 80-82, wherein the composition comprises at least $10^2$, $10^3$, $10^4$, $10^4$, $10^5$, $10^6$, or $10^7$ of the same anellovector.

84. The composition of any of embodiments 80-83, wherein the plurality comprises at least $10^3$, $10^4$, $10^5$, $10^6$, $10^7$, $10^8$, $10^9$, $10^{10}$, $10^{11}$, $10^{12}$, $10^{13}$, $10^{14}$, or $10^{15}$ anellovectors (e.g., copies of the anellovector); or wherein the composition comprises at least $10^5$, $10^6$, $10^7$, $10^8$, $10^9$, $10^{10}$, $10^{11}$, $10^{12}$, $10^{13}$, $10^{14}$, or $10^{15}$ anellovector genomes per mL.

85. The composition of any of embodiments 80-84, having one or more (e.g., 1, 2, 3, 4, 5, or 6) of the following characteristics:

a) the composition meets a pharmaceutical or good manufacturing practices (GMP) standard;

b) the composition was made according to good manufacturing practices (GMP);

c) the composition has a pathogen level below a predetermined reference value, e.g., is substantially free of pathogens;

d) the composition has a contaminant level below a predetermined reference value, e.g., is substantially free of contaminants (e.g., a denaturant, e.g., urea);

e) the composition has a predetermined level of non-infectious particles or a predetermined ratio of particles:infectious units (e.g., <300:1, < 200:1, <100:1, or <50:1), or

f) the pharmaceutical composition has low immunogenicity or is substantially non-immunogenic, e.g., as described herein;

optionally wherein the composition comprises urea at a concentration of less than 0.1M, 0.2M, 0.3M, 0.4M, 0.5M, 0.6M, 0.7M, 0.8M, 0.9M, 1M, 1.1M, 1.2M, 1.3M, 1.5M, 1.5M, 1.6M, 1.7M, 1.8M, 1.9M, or 2M.

86. The composition of any of embodiments 80-85, wherein the pharmaceutical composition has a contaminant level below a predetermined reference value, e.g., is substantially free of contaminants.

87. The composition of any of embodiments 80-86, wherein the composition comprises urea at a concentration of less than 0.1M, 0.2M, 0.3M, 0.4M, 0.5M, 0.6M, 0.7M, 0.8M, 0.9M, 1M, 1.1M, 1.2M, 1.3M, 1.5M, 1.5M, 1.6M, 1.7M, 1.8M, 1.9M, or 2M.

88. The composition of embodiment 86 or 87, wherein the contaminant comprises one or more of the following: mycoplasma, endotoxin, host cell nucleic acids (e.g., host cell DNA and/or host cell RNA), animal-derived process impurities (e.g., serum albumin or trypsin), replication-competent agents (RCA), e.g., replication-competent virus or unwanted anellovectors (e.g., an anellovector other than the desired anellovector, e.g., a synthetic anellovector as described herein), free viral capsid protein, adventitious agents, and/or aggregates.

89. The composition of any of embodiments 80-88, wherein the composition comprises less than 10% (e.g., less than about 10%, 5%, 4%, 3%, 2%, 1%, 0.5%, or 0.1%) contaminant by weight.

90. The composition of any of embodiments 80-89, wherein at least 90% of proteinaceous exteriors comprise the same genetic element (e.g., the genetic element of the anellovector).

91. The composition of any of embodiments 80-90, wherein at least 90% of proteinaceous exteriors comprise the same ORF1 molecule.

92. A pharmaceutical composition comprising the anellovector or composition of any of the preceding embodiments, and a pharmaceutically acceptable carrier or excipient.

93. A method of making an anellovector, the method comprising:

(a) providing a mixture comprising:

(i) a genetic element comprising RNA, and

(ii) an ORF1 molecule; and

(b) incubating the mixture under conditions suitable for enclosing the genetic element within a proteinaceous exterior comprising the ORF1 molecule, thereby making an anellovector;

optionally wherein the mixture is not comprised in a cell.

94. The method of embodiment 93, further comprising, prior to the providing of (a), expressing the ORF1 molecule, e.g., in a host cell (e.g., an insect cell or a mammalian cell).

95. The method of embodiment 94, wherein the expressing comprising incubating a host cell (e.g., an insect cell or a mammalian cell) comprising a nucleic acid molecule (e.g., a baculovirus expression vector) encoding the ORF1 molecule under conditions suitable for producing the ORF1 molecule.

96. The method of embodiment 94 or 95, further comprising, prior to the providing of (a), purifying the ORF1 molecule expressed by the host cell.

97. A method of purifying an anellovector, the method comprising:

(a) providing an anellovector (e.g., as described herein) comprising:

(i) a genetic element, e.g., a genetic element comprising RNA, and

(ii) a proteinaceous exterior comprising an ORF1 molecule, the proteinaceous exterior enclosing the genetic element; and

(b) purifying the anellovector.

98. The method of embodiment 96 or 97, wherein the purifying (e.g., the purifying of the ORF1 molecule or the purifying of the anelloevctor) comprises affinity purification, e.g., heparin affinity purification.

99. The method of any of embodiments 96-98, wherein the purifying (e.g., the purifying of the ORF1 molecule or the purifying of the anelloevctor) comprises size exclusion chromatography (e.g., using a Tris buffer mobile phase).

100. The method of any of embodiments 96-99, wherein the purifying (e.g., the purifying of the ORF1 molecule or the purifying of the anelloevctor) comprises affinity purification (e.g., heparin affinity purification) followed by size exclusion chromatography.

101. The method of any of embodiments 96-100, wherein the purifying (e.g., the purifying of the ORF1 molecule or the purifying of the anelloevctor) comprises anion exchange chromatography (e.g., Mustang Q membrane chromatography).

102. The method of any of embodiments 96-101, wherein the purifying (e.g., the purifying of the ORF1 molecule or the purifying of the anelloevctor) comprises mixed mode chromatography (e.g., using a mixed mode resin, e.g., a Cato700 resin).

103. The method of any of embodiments 96-102, wherein the purifying (e.g., the purifying of the ORF1 molecule or the purifying of the anelloevctor) produces a composition comprising one or more virus-like particles (VLPs) comprising at least about 20, 30, 40, 50, or 60 copies, or 20-30, 30-40, 40-50, or 50-60 copies, of the ORF1 molecule.

104. The method of embodiment 103, wherein at least 75%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of the virus-like particles comprise proteinaceous exteriors that are 60mers or are particles of at least 30, 31, 32, 33, 34, or 35 nm in diameter.

105. The method of embodiment 103, wherein the composition comprises at least $10^5$, $10^6$, $10^7$, $10^8$, $10^9$, $10^{10}$ particles/mL, or comprises $10^5$ - $10^6$, $10^6$ - $10^7$, $10^7$ - $10^9$, $10^8$ - $10^9$, $10^9$ - $10^{10}$, or $10^{10}$ - $10^{11}$ particles/mL (e.g., as measured by electron microscopy).

106. The method of any of embodiments 93-105, further comprising, prior to the providing of (a), incubating the ORF1 molecule under conditions suitable for disassembly of a proteinaceous exterior (e.g., a virus-like particle (VLP)) comprising the ORF1 molecule.

107. The method of embodiment 106, wherein the conditions suitable for disassembly of the proteinaceous exterior comprising the ORF1 molecule comprise incubation in the presence of a denaturant.

108. The method of any of embodiments 106-107, wherein the denaturant comprises a chaotropic agent (e.g., urea), or a detergent (e.g., SDS (e.g., 0.1% SDS), Tween, or Triton).

109. The method of any of embodiments 106-108, wherein the conditions suitable for disassembly of the proteinaceous exterior comprising the ORF1 molecule comprise a predetermined conductivity, a high salt solution (e.g., a solution comprising NaCl, e.g., at a concentration of at least about 1M, e.g., at least about 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 2, 3, 4, or 5M), heat (e.g., temperature above about 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, or 95°C), or pH (e.g., acidic pH or basic pH).

110. The method of any of embodiments 106-109, wherein the conditions suitable for disassembly of the proteinaceous exterior comprising the ORF1 moleceule comprise incubation in a solution comprising urea at a concentration of at least 0.1M, 0.2M, 0.3M, 0.4M, 0.5M, 0.6M, 0.7M, 0.8M, 0.9M, 1M, 1.1M, 1.2M, 1.3M, 1.5M, 1.5M, 1.6M, 1.7M, 1.8M, 1.9M, or 2M.

111. The method of any of embodiments 106-110, wherein the incubating of (b) results in at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 96%, or 100% of a population of particles comprising the ORF1 molecule or copies thereof being disassembled.

112. The method of any of embodiments 106-111, wherein the conditions suitable for disassembly of the proteinaceous exterior are sufficient to disassemble a complex (e.g., a proteinaceous exterior) comprising at least about 20, 30, 40, 50, or 60 copies, or 20-30, 30-40, 40-50, or 50-60 copies, of the ORF1 molecule.

113. The method of any of embodiments 106-112, wherein the conditions suitable for disassembly of the proteinaceous exterior result in fewer than $10^8$ remaining intact particles/mL.

114. The method of any of embodiments 106-113, further comprising, prior to the providing of (a), removing the ORF1 molecule from the conditions suitable for disassembly of the proteinaceous exterior (e.g., subjecting the ORF1 molecule to non-denaturing conditions).

115. The method of embodiment 114, wherein the removing of the ORF1 molecule from the conditions suitable for disassembly of the proteinaceous exterior comprises reducing the concentration of the denaturant, e.g., reducing the concentration of the denaturan (e.g., urea) to below 0.1M, 0.2M, 0.3M, 0.4M, 0.5M, 0.6M, 0.7M, 0.8M, 0.9M, 1M, 1.1M, 1.2M, 1.3M, 1.5M, 1.5M, 1.6M, 1.7M, 1.8M, 1.9M, or 2M.

116. The method of embodiment 114 or 115, wherein the removing results in the formation of one or more anellovectors each enclosing at least one copy (e.g., at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, or 1000 copies) of the genetic element (e.g., the mRNA).

117. The method of embodiment 116, wherein the number of anellovectors enclosing the at least one copy of the genetic element in the resulting solution is at least $10^5$, $10^6$, $10^7$, $10^8$, $10^9$, $10^{10}$, or $10^{11}$; or is between $10^5$ - $10^6$, $10^6$ - $10^7$, $10^7$ - $10^9$, $10^8$ - $10^9$, $10^9$ - $10^{10}$, or $10^{10}$ - $10^{11}$ (e.g., as measured by electron microscopy).

118. The method of embodiment 116, wherein the number of anellovectors enclosing the at least one copy of the genetic element in the resulting solution is at least $10^5$, $10^6$, $10^7$, $10^8$, $10^9$, $10^{10}$, or $10^{11}$ anellovectors/mL; or is between $10^5$ - $10^6$, $10^6$ - $10^7$, $10^7$ - $10^9$, $10^8$ - $10^9$, $10^9$ - $10^{10}$, or $10^{10}$ - $10^{11}$ anellovectors/mL (e.g., as measured by electron microscopy).

119. The method of any of embodiments 114-118, wherein the removing results in a solution comprising at least $10^5$, $10^6$, $10^7$, $10^8$, $10^9$, $10^{10}$, or $10^{11}$ anellovectors/mL; or between $10^5$ - $10^6$, $10^6$ - $10^7$, $10^7$ - $10^9$, $10^8$ - $10^9$, $10^9$ - $10^{10}$, or $10^{10}$ - $10^{11}$ anellovectors/mL (e.g., as measured by electron microscopy), wherein the anellovectors each enclose at least one copy (e.g., at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, or 1000 copies) of the genetic element (e.g., the mRNA)

120. The method of any of embodiments 114-119, wherein at least 75%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of the anellovectors comprise proteinaceous exteriors that are 60mers or particles or are particles of at least 30, 31, 32, 33, 34, or 35 nm in diameter.

121. The method of any of embodiments 93-120, wherein the genetic element of the anellovector is resistant to an endonuclease (e.g., an RNase).

122. The method of any of embodiments 93-121, wherein (a) comprises admixing (i) and (ii).

123. The method of any of embodiments 93-122, which is performed in a cell-free system.

124. A method of manufacturing an anellovector composition, comprising:

(a) providing a plurality of anellovectors or compositions according to any of the preceding embodiments;

(b) optionally evaluating the plurality for one or more of: a contaminant described herein, an optical density measurement (e.g., OD 260), particle number (e.g., by HPLC), infectivity (e.g., particle:infectious unit ratio, e.g., as determined by fluorescence and/or ELISA); and

(c) formulating the plurality of anellovectors, e.g., as a pharmaceutical composition suitable for administration to a subject, e.g., if one or more of the parameters of (b) meet a specified threshold.

125. A method of treating a disease or disorder in a subject in need thereof, the method comprising administering to the subject an anellovector or composition of any of the preceding embodiments, thereby treating a disease or disorder (e.g., as described herein) in the subject.

126. A method of modulating, e.g., enhancing or inhibiting, a biological function (e.g., as described herein) in a subject, the method comprising administering the anellovector or the composition of any of the preceding embodiments to the subject.

127. A method of delivering a genetic element to a cell, the method comprising contacting the anellovector or composition of any of the preceding embodiments with a cell, e.g., a eukaryotic cell, e.g., a mammalian cell, e.g., a human cell.

128. Use of the anellovector or composition of any of the preceding embodiments for treating a disease or disorder (e.g., as described herein) in a subject.

129. The anellovector or composition of any of the preceding embodiments for use in a method for treating a disease or disorder (e.g., as described herein) in a subject.

130. The anellovector or composition of any of the preceding embodiments for use in the manufacture of a medicament for treating a disease or disorder (e.g., as described herein) in a subject.

Other features, objects, and advantages of the invention will be apparent from the description and drawings, and from the claims.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

## BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a series of electron micrographs showing that recombinant capsid proteins from anellovirus strains form virus-like particles (VLPs) *in vitro*. Capsid proteins produced in different cell lines form VLPs *in vitro* as observed by negative staining electron microscopy. (A) Ring 2 ORF1 VLP purified from insect cells with an observed diameter of approximately 35 nm. (B) Ring 10 ORF1 VLP purified from insect cells with an observed diameter of approximately 35 nm. (C) CAV VP1 VLP purified from mammalian cells with an observed diameter approximately 20 nm.

FIG. 2 is a diagram showing the eight beta-strand jelly roll domain (or jelly roll fold) observed in the structure of Beak and Feather Disease Virus (BFDV). By convention, the beta strands are labeled B though I. The strands form four antiparallel beta sheets with an orientation of B-I-D-G and C-H-E-F. The B-I-D-G sheet forms the interior of the viral capsid.

FIG. 3 is an amino acid sequence alignment depicting the jelly roll sequences for Anellovirus ORF1 protein as compared to jelly roll domain of Beak and Feather Disease Virus (BFDV)/ Hepatitis E capsid protein (HepE).

FIGS. 4A-4E are a series of diagrams showing an exemplary method of producing Anellovirus ORF1 protein-based virus-like particles (VLPs) enclosing mRNA molecules encoding eGFP. (A) ORF1 protein was produced in cells and isolated as described herein. (B) VLPs that formed from the ORF1 proteins were then disassembled in a 2 M urea solution. (C) When the urea was removed in the absence of mRNA, few VLPs reformed (titer of less than $10^8$ particles/mL detected by electron microscopy). (D, E) When the urea was removed in the presence of mRNAs encoding eGFP, substantial quantities of VLPs (titer of $1x10^9$ - $1x10^{10}$ particles/mL) were detected by electron microscopy (EM).

The following detailed description of the embodiments of the invention will be better understood when read in conjunction with the appended drawings. For the purpose of illustrating the invention, there are shown in the drawings embodiments that are presently exemplified. It should be understood, however, that the invention is not limited to the precise arrangement and instrumentalities of the embodiments shown in the drawings.

## DETAILED DESCRIPTION OF THE INVENTION

Definitions

The present invention will be described with respect to particular embodiments and with reference to certain figures but the invention is not limited thereto but only by the claims. Terms as set forth hereinafter are generally to be understood in their common sense unless indicated otherwise.

Where the term "comprising" is used in the present description and claims, it does not exclude other elements. For the purposes of the present invention, the term "consisting of" is considered to be a preferred embodiment of the term "comprising of". If hereinafter a group is defined to comprise at least a certain number of embodiments, this is to be understood to preferably also disclose a group which consists only of these embodiments.

Where an indefinite or definite article is used when referring to a singular noun, e.g. "a", "an" or "the", this includes a plural of that noun unless something else is specifically stated.

The wording "compound, composition, product, etc. for treating, modulating, etc." is to be understood to refer a compound, composition, product, etc. *per se* which is suitable for the indicated purposes of treating, modulating, etc. The wording "compound, composition, product, etc. for treating, modulating, etc." additionally discloses that, as an embodiment, such compound, composition, product, etc. is for use in treating, modulating, etc.

The wording "compound, composition, product, etc. for use in …", "use of a compound, composition, product, etc in the manufacture of a medicament, pharmaceutical composition, veterinary composition, diagnostic composition, etc. for …", or "compound, composition, product, etc. for use as a medicament…" indicates that such compounds, compositions, products, etc. are to be used in therapeutic methods which may be practiced on the human or animal body. They are considered as an equivalent disclosure of embodiments and claims pertaining to methods of treatment, etc. If an embodiment or a claim thus refers to "a compound for use in treating a human or animal being suspected to suffer from a disease", this is considered to be also a disclosure of a "use of a compound in the manufacture of a medicament for treating a human or animal being suspected to suffer from a disease" or a "method of treatment by administering a compound to a human or animal being suspected to suffer from a disease".

The wording "compound, composition, product, etc. for treating, modulating, etc." is to be understood to refer a compound, composition, product, etc. *per se* which is suitable for the indicated purposes of treating, modulating, etc.

If hereinafter examples of a term, value, number, etc. are provided in parentheses, this is to be understood as an indication that the examples mentioned in the parentheses can constitute an embodiment. For example, if it is stated that "in embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1-encoding nucleotide sequence of Table 1 (e.g., nucleotides 571 – 2613 of the nucleic acid sequence of Table 1)", then some embodiments relate to nucleic acid molecules comprising a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to nucleotides 571 – 2613 of the nucleic acid sequence of Table 1.

The term "amplification," as used herein, refers to replication of a nucleic acid molecule or a portion thereof, to produce one or more additional copies of the nucleic acid molecule or a portion thereof (e.g., a genetic element or a genetic element region). In some embodiments, amplification results in partial replication of a nucleic acid sequence. In some embodiments, amplification occurs via rolling circle replication.

As used herein, the term "anellovector" refers to a vehicle comprising a genetic element, e.g., an RNA, e.g., a circular RNA, enclosed in a proteinaceous exterior. In some embodiments, the genetic element is substantially protected from digestion with RNase by a proteinaceous exterior. A "synthetic anellovector," as used herein, generally refers to an anellovector that is not naturally occurring, e.g., has a sequence that is different relative to a wild-type virus (e.g., a wild-type *Anellovirus* as described herein). In some embodiments, the synthetic anellovector is engineered or recombinant, e.g., comprises a genetic element that comprises a difference or modification relative to a wild-type viral genome (e.g., a wild-type *Anellovirus* genome as described herein). In some embodiments, enclosed within a proteinaceous exterior encompasses 100% coverage by a proteinaceous exterior, as well as less than 100% coverage, e.g., 95%, 90%, 85%, 80%, 70%, 60%, 50% or less. For example, gaps or discontinuities (e.g., that render the proteinaceous exterior permeable to water, ions, peptides, or small molecules) may be present in the proteinaceous exterior, so long as the genetic element is retained in the proteinaceous exterior or protected from digestion with an RNase, e.g., prior to entry into a host cell. In some embodiments, the anellovector is purified, e.g., it is separated from its original source and/or substantially free (>50%, >60%, >70%, >80%, >90%) of other components. In some embodiments, the anellovector is capable of introducing the genetic element into a target cell (e.g., via infection). In some embodiments, the anellovector is an

infective synthetic viral particle containing certain Anellovirus elements, such as an Anellovirus ORF1 molecule.

As used herein, the term "antibody molecule" refers to a protein, e.g., an immunoglobulin chain or fragment thereof, comprising at least one immunoglobulin variable domain sequence. The term "antibody molecule" encompasses full-length antibodies and antibody fragments (e.g., scFvs). In some embodiments, an antibody molecule is a multispecific antibody molecule, e.g., the antibody molecule comprises a plurality of immunoglobulin variable domain sequences, wherein a first immunoglobulin variable domain sequence of the plurality has binding specificity for a first epitope and a second immunoglobulin variable domain sequence of the plurality has binding specificity for a second epitope. In some embodiments, the multispecific antibody molecule is a bispecific antibody molecule. A bispecific antibody molecule is generally characterized by a first immunoglobulin variable domain sequence which has binding specificity for a first epitope and a second immunoglobulin variable domain sequence that has binding specificity for a second epitope.

The term "backbone" or "backbone region," as used herein, refers to a region within a nucleic acid molecule (e.g., within a bacmid or donor vector, e.g., as described herein) that comprises one or more elements involved in (e.g., necessary and/or sufficient for) replication and/or maintenance of the nucleic acid molecule in a host cell. In some embodiments, a backbone region, such as a "baculovirus backbone region," comprises one or more baculoviral elements (e.g., a baculovirus genome or a functional fragment thereof), e.g., suitable for replication of the nucleic acid construct in insect cells (e.g., Sf9 cells). In some embodiments, the backbone further comprises a selectable marker. In some embodiments, a nucleic acid molecule comprises a genetic element region and a backbone region (e.g., a baculovirus backbone region and/or a backbone region suitable for replication in bacterial cells).

The term "bacmid", as used herein, refers to a nucleic acid molecule comprising sufficient baculovirus backbone elements such that it is suitable for replication in insect cells, and furthermore is suitable for replication in bacterial cells. In some embodiments, the nucleic acid molecule is suitable for replication in bacterial cells (e.g., *E. coli* cells, e.g., DH 10Bac cells).

As used herein, a "circular" nucleic acid refers to a nucleic acid that forms a structure without free 5' or 3' ends. In some embodiments, the circular nucleic acid is closed through covalent or non-covalent bonds. For instance, the circular nucleic acid may be made by covalently linking the ends of a linear nucleic acid, e.g., with a phosphate-sugar bond or a synthetic linker moeity. In other embodiments, the circular nucleic acid comprises two ends that are in proximity and are not free (not substantially accessible to an exonuclease). For instance, the circular nucleic acid may be made by hybridizing the ends of a linear nucleic acid directly or through a nucleic acid splint.

As used herein, a "DNA region" refers to a portion of a polynucleotide strand comprising a plurality of DNA nucleotides. For example, in some embodiments a DNA region is a plurality of DNA nucleotides incorporated into an RNA strand. For example, a DNA region comprises about 5-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, or 90-100 DNA nucleotides within a polynucleotide strand.

As used herein, a nucleic acid "encoding" refers to a nucleic acid sequence encoding an amino acid sequence or a polynucleotide, e.g., an mRNA or functional polynucleotide (e.g., a non-coding RNA, e.g., an siRNA or miRNA).

An "exogenous" agent (e.g., an effector, a nucleic acid (e.g., RNA), a gene, payload, protein) as used herein refers to an agent that is either not comprised by, or not encoded by, a corresponding wild-type virus, e.g., an Anellovirus as described herein. In some embodiments, the exogenous agent does not naturally exist, such as a protein or nucleic acid that has a sequence that is altered (e.g., by insertion, deletion, or substitution) relative to a naturally occurring protein or nucleic acid. In some embodiments, the exogenous agent does not naturally exist in the host cell. In some embodiments, the exogenous agent exists naturally in the host cell but is exogenous to the virus. In some embodiments, the exogenous agent exists naturally in the host cell, but is not present at a desired level or at a desired time.

A "heterologous" agent or element (e.g., an effector, a nucleic acid sequence, an amino acid sequence), as used herein with respect to another agent or element (e.g., an effector, a nucleic acid sequence, an amino acid sequence), refers to agents or elements that are not naturally found together, e.g., in a wild-type virus, e.g., an Anellovirus. In some embodiments, a heterologous nucleic acid sequence may be present in the same nucleic acid as a naturally occurring nucleic acid sequence (e.g., a sequence that is naturally occurring in the Anellovirus). In some embodiments, a heterologous agent or element is exogenous relative to an Anellovirus from which other (e.g., the remainder of) elements of the anellovector are based.

As used herein, the term "genetic element" refers to a nucleic acid molecule that is or can be enclosed within (e.g, protected from RNase digestion by) a proteinaceous exterior, e.g., to form an anellovector as described herein. It is understood that the genetic element can be produced as naked RNA and optionally further assembled into a proteinaceous exterior. It is also understood that an anellovector can insert its genetic element into a cell, resulting in the genetic element being present in the cell and the proteinaceous exterior not necessarily entering the cell.

As used herein, "genetic element construct" refers to a nucleic acid construct (e.g., a plasmid, bacmid, donor vector, cosmid, or minicircle) comprising a genetic element sequence, or fragment thereof. In some embodiments, a bacmid or donor vector as described herein is a genetic element construct comprising a genetic element sequence, or fragment thereof.

The term "genetic element region," as used herein, refers to a region of a construct that comprises the sequence of a genetic element. In some embodiments, the genetic element region comprises a sequence having sufficient identity to a wild-type Anellovirus sequence, or a fragment thereof, to be enclosed by a proteinaceous exterior, thereby forming an anellovector (e.g., a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the wild-type Anellovirus sequence or fragment thereof). In embodiments, the genetic element region comprises a protein binding sequence, e.g., as described herein (e.g., a 5' UTR, 3' UTR, and/or a GC-rich region as described herein, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto). In some embodiments, the construct comprising a genetic element region is not enclosed in a proteinaceous exterior, but a genetic element produced from the construct can be enclosed in a proteinaceous exterior. In some embodiments, the construct comprising the genetic element region further comprises a vector backbone (e.g., a bacmid backbone or a donor vector backbone). In some embodiments, the construct (e.g., bacmid) comprises one or more baculovirus elements (e.g., a baculovirus genome, e.g., comprising the genetic element region).

As used herein, the term "mutant" when used with respect to a genome (e.g., an *Anellovirus* genome), or a fragment thereof, refers to a sequence having at least one change relative to a corresponding wild-type *Anellovirus* sequence. In some embodiments, the mutant genome or fragment thereof comprises at least one single nucleotide polymorphism, addition, deletion, or frameshift relative to the corresponding wild-type *Anellovirus* sequence. In some embodiments, the mutant genome or fragment thereof comprises a deletion of at least one Anellovirus ORF (e.g., one or more of ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, and/or ORF1/2) relative to the corresponding wild-type *Anellovirus* sequence. In some embodiments, the mutant genome or fragment thereof comprises a deletion of all Anellovirus ORFs (e.g., all of ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, and ORF1/2) relative to the corresponding wild-type *Anellovirus* sequence. In some embodiments, the mutant genome or fragment thereof comprises a deletion of at least one Anellovirus noncoding region (e.g., one or more of a 5' UTR, 3' UTR, and/or GC-rich region) relative to the corresponding wild-type *Anellovirus* sequence. In some embodiments, the mutant genome or fragment thereof comprises or encodes an exogenous effector.

As used herein the term "ORF molecule" refers to a polypeptide having an activity and/or a structural feature of an Anellovirus ORF protein (e.g., a polypeptide comprising an Anellovirus ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, and/or ORF1/2 protein), or a functional fragment thereof. When used generically (i.e., "ORF molecule"), the polypeptide may comprise an activity and/or structural feature of any of the Anellovirus ORFs described herein (e.g., any of an Anellovirus ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, and/or ORF1/2), or a functional fragment thereof. When used with a modifier to indicate a particular open reading frame (e.g., "ORF1 molecule," "ORF2 molecule," "ORF2/2 molecule,"

"ORF2/3 molecule," "ORF1/1 molecule," or "ORF1/2 molecule"), it is generally meant that the polypeptide comprises an activity and/or structural feature of the corresponding Anellovirus ORF protein, or a functional fragment thereof (for example, as defined below for "ORF1 molecule"). For example, an "ORF2 molecule" comprises an activity and/or structural feature of an Anellovirus ORF2 protein, or a functional fragment thereof.

As used herein, the term "ORF1 molecule" refers to a polypeptide having an activity and/or a structural feature of an Anellovirus ORF1 protein (e.g., an Anellovirus ORF1 protein as described herein), or a functional fragment thereof. An ORF1 molecule may, in some instances, comprise one or more of (e.g., 1, 2, 3 or 4 of): a first region comprising at least 60% basic residues (e.g., at least 60% arginine residues), a second region compising at least about six beta strands (e.g., at least 4, 5, 6, 7, 8, 9, 10, 11, or 12 beta strands), a third region comprising a structure or an activity of an Anellovirus N22 domain (e.g., as described herein, e.g., an N22 domain from an Anellovirus ORF1 protein as described herein), and/or a fourth region comprising a structure or an activity of an Anellovirus C-terminal domain (CTD) (e.g., as described herein, e.g., a CTD from an Anellovirus ORF1 protein as described herein). In some instances, the ORF1 molecule comprises, in N-terminal to C-terminal order, the first, second, third, and fourth regions. In some instances, an anellovector comprises an ORF1 molecule comprising, in N-terminal to C-terminal order, the first, second, third, and fourth regions. An ORF1 molecule may, in some instances, comprise a polypeptide encoded by an Anellovirus ORF1 nucleic acid. An ORF1 molecule may, in some instances, further comprise a heterologous sequence, e.g., a hypervariable region (HVR), e.g., an HVR from an Anellovirus ORF1 protein, e.g., as described herein. An "Anellovirus ORF1 protein," as used herein, refers to an ORF1 protein encoded by an Anellovirus genome (e.g., a wild-type Anellovirus genome, e.g., as described herein).

As used herein, the term "ORF2 molecule" refers to a polypeptide having an activity and/or a structural feature of an Anellovirus ORF2 protein (e.g., an Anellovirus ORF2 protein as described herein), or a functional fragment thereof. An "Anellovirus ORF2 protein," as used herein, refers to an ORF2 protein encoded by an Anellovirus genome (e.g., a wild-type Anellovirus genome, e.g., as described herein).

As used herein, the term "proteinaceous exterior" refers to an exterior component that is predominantly (e.g., >50%, >60%, > 70%, >80%, > 90%) protein.

As used herein, the term "regulatory nucleic acid" refers to a nucleic acid sequence that modifies expression, e.g., transcription and/or translation, of a DNA sequence that encodes an expression product. In some embodiments, the expression product comprises RNA or protein.

As used herein, the term "regulatory sequence" refers to a nucleic acid sequence that modifies transcription of a target gene product. In some embodiments, the regulatory sequence is a promoter or an enhancer.

As used herein, a "substantially non-pathogenic" organism, particle, or component, refers to an organism, particle (e.g., a virus or an anellovector, e.g., as described herein), or component thereof that does not cause or induce an unacceptable disease or pathogenic condition, e.g., in a host organism, e.g., a mammal, e.g., a human. In some embodiments, administration of an anellovector to a subject can result in minor reactions or side effects that are acceptable as part of standard of care.

As used herein, the term "non-pathogenic" refers to an organism or component thereof that does not cause or induce an undesirable condition (e.g., a disease or pathogenic condition), e.g., in a host organism, e.g., a mammal, e.g., a human.

As used herein, a "substantially non-integrating" genetic element refers to a genetic element, e.g., a genetic element in a virus or anellovector, e.g., as described herein, wherein less than about 0.01%, 0.05%, 0.1%, 0.5%, or 1% of the genetic element that enter into a host cell (e.g., a eukaryotic cell) or organism (e.g., a mammal, e.g., a human) integrate into the genome. In some embodiments the genetic element does not detectably integrate into the genome of, e.g., a host cell. In some embodiments, integration of the genetic element into the genome can be detected using techniques as described herein, e.g., nucleic acid sequencing, PCR detection and/or nucleic acid hybridization. In some embodiments, integration frequency is determined by quantitative gel purification assay of genomic DNA separated from free vector, e.g., as described in Wang et al. (2004, *Gene Therapy* 11: 711-721, incorporated herein by reference in its entirety).

As used herein, a "substantially non-immunogenic" organism, particle, or component, refers to an organism, particle (e.g., a virus or anellovector, e.g., as described herein), or component thereof, that does not cause or induce an undesired or untargeted immune response, e.g., in a host tissue or organism (e.g., a mammal, e.g., a human). In some embodiments, the substantially non-immunogenic organism, particle, or component does not produce a clinically significant immune response. In some embodiments, the substantially non-immunogenic anellovector does not produce a clinically significant immune response against a protein comprising an amino acid sequence or encoded by a nucleic acid sequence of an Anellovirus or anellovector genetic element. In some embodiments, an immune response (e.g., an undesired or untargeted immune response) is detected by assaying antibody (e.g., neutralizing antibody) presence or level (e.g., presence or level of an anti-anellovector antibody, e.g., presence or level of an antibody against an anellovector as described herein) in a subject, e.g., according to the anti-TTV antibody detection method described in Tsuda et al. (1999; *J. Virol. Methods* 77: 199-206; incorporated herein by reference) and/or the method for determining anti-TTV IgG levels described in Kakkola et al.

(2008; *Virology* 382: 182-189; incorporated herein by reference). Antibodies (e.g., neutralizing antibodies) against an *Anellovirus* or an anellovector based thereon can also be detected by methods in the art for detecting anti-viral antibodies, e.g., methods of detecting anti-AAV antibodies, e.g., as described in Calcedo et al. (2013; *Front. Immunol.* 4(341): 1-7; incorporated herein by reference).

A "subsequence" as used herein refers to a nucleic acid sequence or an amino acid sequence that is comprised in a larger nucleic acid sequence or amino acid sequence, respectively. In some instances, a subsequence may comprise a domain or functional fragment of the larger sequence. In some instances, the subsequence may comprise a fragment of the larger sequence capable of forming secondary and/or tertiary structures when isolated from the larger sequence similar to the secondary and/or tertiary structures formed by the subsequence when present with the remainder of the larger sequence. In some instances, a subsequence can be replaced by another sequence (e.g., a subseqence comprising an exogenous sequence or a sequence heterologous to the remainder of the larger sequence, e.g., a corresponding subsequence from a different *Anellovirus*).

This invention relates generally to anellovectors, e.g., synthetic anellovectors, and uses thereof. The present disclosure provides anellovectors, compositions comprising anellovectors, and methods of making or using anellovectors. Anellovectors are generally useful as delivery vehicles, e.g., for delivering a therapeutic agent to a eukaryotic cell. Generally, an anellovector described herein will include a genetic element comprising an RNA sequence (e.g., an RNA sequence encoding an effector, e.g., an exogenous effector or an endogenous effector) enclosed within a proteinaceous exterior. An anellovector may include one or more deletions of sequences (e.g., regions or domains as described herein) relative to an *Anellovirus* sequence (e.g., as described herein). Anellovectors can be used as a substantially non-immunogenic vehicle for delivering the genetic element, or an effector encoded therein (e.g., a polypeptide or nucleic acid effector, e.g., as described herein), into eukaryotic cells, e.g., to treat a disease or disorder in a subject comprising the cells.

**TABLE OF CONTENTS**

I. Compositions and Methods for Making Anellovectors by In Vitro Assembly

A. Components and Assembly of Anellovectors

i. ORF1 molecules for assembly of Anellovectors

ii. ORF2 molecules for assembly of Anellovectors

B. Genetic Elements

i. Genetic Elements comprising RNA

a. RNA-only genetic elements

b. Hybridized RNA-ssDNA genetic elements
c. RNA/DNA conjugates
C. Genetic Element Constructs
D. Production of RNA-based genetic elements
E. Production of protein components
i. Baculovirus expression systems
ii. Insect cell systems
iii. Mammalian cell systems
F. Effectors
G. In vitro assembly methods
H. Enrichment and Purification
II. Anellovectors
A. Anelloviruses
B. ORF1 molecules
C. ORF2 molecules
D. Genetic elements
E. Protein binding sequences
F. 5' UTR regions
G. GC-rich regions
H. Effectors
I. Regulatory Sequences
J. Other Sequences
K. Proteinaceous exterior
III. Methods of use
IV. Administration/Delivery

## I. Compositions and Methods for Making Anellovectors by In Vitro Assembly

The present disclosure provides, in some aspects, compositions and methods that can be used for producing anellovectors, e.g., anellovectors having a genetic element comprising RNA, as described herein. In some embodiments, the compositions and methods described herein can be used to produce a genetic element or a genetic element construct. In some embodiments, the compositions and methods described herein can be used to produce a genetic element or a genetic element construct by *in vitro* assembly. In some embodiments, the compositions and methods described herein can be used to produce

one or more Anellovirus ORF molecules (e.g., an ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, or ORF1/2 molecule, or a functional fragment or splice variant thereof). In some embodiments, the compositions and methods described herein can be used to produce a proteinaceous exterior or a component thereof (e.g., an ORF1 molecule), e.g., in a host cell (e.g., an insect cell, e.g., an Sf9 cell).

***Components and Assembly of Anellovectors***

The compositions and methods herein can be used to produce anellovectors. As described herein, an anellovector generally comprises a genetic element (e.g., an RNA molecule) enclosed within a proteinaceous exterior (e.g., comprising a polypeptide encoded by an Anellovirus ORF1 nucleic acid, e.g., as described herein). In some embodiments, the genetic element comprises one or more sequences encoding Anellovirus ORFs (e.g., one or more of an Anellovirus ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, or ORF1/2). As used herein, an Anellovirus ORF or ORF molecule (e.g., an Anellovirus ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, or ORF1/2) includes a polypeptide comprising an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to a corresponding Anellovirus ORF sequence, e.g., as described in PCT/US2018/037379 or PCT/US19/65995 (each of which is incorporated by reference herein in their entirety). In embodiments, the genetic element comprises a sequence encoding an Anellovirus ORF1, or a splice variant or functional fragment thereof (e.g., a jelly-roll region, e.g., as described herein). In some embodiments, the proteinaceous exterior comprises a polypeptide encoded by an Anellovirus ORF1 nucleic acid (e.g., an Anellovirus ORF1 molecule or a splice variant or functional fragment thereof).

In some embodiments, an anellovector is assembled by enclosing a genetic element (e.g., as described herein) within a proteinaceous exterior (e.g., as described herein). In some embodiments, the genetic element is enclosed within the proteinaceous exterior in a host cell (e.g., an insect cell, e.g., an Sf9 cell). In some embodiments, the host cell expresses one or more polypeptides comprised in the proteinaceous exterior (e.g., a polypeptide encoded by an Anellovirus ORF1 nucleic acid, e.g., an ORF1 molecule). For example, in some embodiments, the host cell comprises a nucleic acid sequence encoding an Anellovirus ORF1 molecule, e.g., a splice variant or a functional fragment of an Anellovirus ORF1 polypeptide (e.g., a wild-type Anellovirus ORF1 protein or a polypeptide encoded by a wild-type Anellovirus ORF1 nucleic acid, e.g., as described herein). In embodiments, the nucleic acid sequence encoding the Anellovirus ORF1 molecule is comprised in a nucleic acid construct (e.g., a plasmid, viral vector, virus, minicircle, bacmid, or artificial chromosome) comprised in the host cell. In embodiments, the nucleic acid sequence encoding the Anellovirus ORF1 molecule is integrated into the genome of the host cell.

In some embodiments, the host cell comprises the genetic element and/or a nucleic acid construct comprising the sequence of the genetic element. In some embodiments, the nucleic acid construct is selected from a plasmid, viral nucleic acid, minicircle, bacmid, or artificial chromosome. In some embodiments, the genetic element is excised from the nucleic acid construct (e.g., bacmid) and, optionally, converted from a double-stranded form to a single-stranded form (e.g., by denaturation). In some embodiments, the genetic element is generated by a polymerase based on a template sequence in the nucleic acid construct (e.g., bacmid). In some embodiments, the polymerase produces a single-stranded copy of the genetic element sequence, which can optionally be circularized to form a genetic element as described herein.

In some embodiments, the host cell comprises a genetic element construct (e.g., a bacmid, plasmid, or minicircle) and a bacmid comprising one or more sequences encoding Anellovirus ORF molecules (e.g., ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, and/or ORF1/2 ORF molecules), or functional fragments thereof. In some embodiments, proteinaceous exterior proteins are expressed from the bacmid. In embodiments, the proteinaceous exterior proteins expressed from the bacmid enclose a genetic element, thereby forming an anellovector. In some embodiments, the bacmid comprises a backbone suitable for replication of the nucleic acid construct in insect cells (e.g., Sf9 cells), e.g., a baculovirus backbone region. In some embodiments, the bacmid comprises a backbone region suitable for replication of the genetic element construct in bacterial cells (e.g., *E. coli* cells, e.g., DH 10Bac cells). In some embodiments, the genetic element construct comprises a backbone suitable for replication of the nucleic acid construct in insect cells (e.g., Sf9 cells), e.g., a baculovirus backbone region. In some embodiments, the genetic element construct comprises a backbone region suitable for replication of the genetic element construct in bacterial cells (e.g., *E. coli* cells, e.g., DH 10Bac cells). In some embodiments, the bacmid is introduced into the host cell via a baculovirus particle. In embodiments, the bacmid is produced by a producer cell, e.g., an insect cell (e.g., an Sf9 cell) or a bacterial cell (e.g., an *E. coli* cell, e.g., a DH 10Bac cell). In embodiments, the producer cell comprises a bacmid and/or a donor vector, e.g., as described herein. In embodiments, the producer cell further comprises sufficient cellular machinery for replication of the bacmid and/or donor vector.

***ORF1 Molecules, e.g., for assembly of Anellovectors***

An anellovector can be made, for example, by enclosing a genetic element within a proteinaceous exterior. In some embodiments, the enclosure occurs in a cell-free system or in a cell. The proteinaceous exterior of an Anellovector generally comprises a polypeptide encoded by an Anellovirus ORF1 nucleic acid (e.g., an Anellovirus ORF1 molecule or a splice variant or functional fragment thereof, e.g., as described herein). An ORF1 molecule may, in some embodiments, comprise one or more of: a first

region comprising an arginine rich region, e.g., a region having at least 60% basic residues (e.g., at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% basic residues; e.g., between 60%-90%, 60%-80%, 70%-90%, or 70-80% basic residues), and a second region comprising jelly-roll domain, e.g., at least six beta strands (e.g., 4, 5, 6, 7, 8, 9, 10, 11, or 12 beta strands). In embodiments, the proteinaceous exterior comprises one or more (e.g., 1, 2, 3, 4, or all 5) of an Anellovirus ORF1 arginine-rich region, jelly-roll region, N22 domain, hypervariable region, and/or C-terminal domain. In some embodiments, the proteinaceous exterior comprises an Anellovirus ORF1 jelly-roll region (e.g., as described herein). In some embodiments, the proteinaceous exterior comprises an Anellovirus ORF1 arginine-rich region (e.g., as described herein). In some embodiments, the proteinaceous exterior comprises an Anellovirus ORF1 N22 domain (e.g., as described herein). In some embodiments, the proteinaceous exterior comprises an Anellovirus hypervariable region (e.g., as described herein). In some embodiments, the proteinaceous exterior comprises an Anellovirus ORF1 C-terminal domain (e.g., as described herein).

In some embodiments, the anellovector comprises an ORF1 molecule and/or a nucleic acid encoding an ORF1 molecule. Generally, an ORF1 molecule comprises a polypeptide having the structural features and/or activity of an Anellovirus ORF1 protein (e.g., an Anellovirus ORF1 protein as described herein), or a functional fragment thereof. In some embodiments, the ORF1 molecule comprises a truncation relative to an Anellovirus ORF1 protein (e.g., an Anellovirus ORF1 protein as described herein). In some embodiments, the ORF1 molecule is truncated by at least 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, or 700 amino acids of the Anellovirus ORF1 protein. In some embodiments, an ORF1 molecule comprises an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity to an Alphatorquevirus, Betatorquevirus, or Gammatorquevirus ORF1 protein, e.g., as described herein. An ORF1 molecule can generally bind to a nucleic acid molecule, such as DNA (e.g., a genetic element, e.g., as described herein). In some embodiments, an ORF1 molecule localizes to the nucleus of a cell. In certain embodiments, an ORF1 molecule localizes to the nucleolus of a cell.

Without wishing to be bound by theory, an ORF1 molecule may be capable of binding to other ORF1 molecules, e.g., to form a proteinaceous exterior (e.g., as described herein). Such an ORF1 molecule may be described as having the capacity to form a capsid. In some embodiments, the proteinaceous exterior may enclose a nucleic acid molecule (e.g., a genetic element as described herein). In some embodiments, a plurality of ORF1 molecules may form a multimer, e.g., to produce a proteinaceous exterior. In some embodiments, the multimer may be a homomultimer. In other embodiments, the multimer may be a heteromultimer.

***ORF2 Molecules*, e.g., for assembly of Anellovectors**

Producing an anellovector using the compositions or methods described herein may involve expression of an Anellovirus ORF2 molecule (e.g., as described herein), or a splice variant or functional fragment thereof. In some embodiments, the anellovector comprises an ORF2 molecule, or a splice variant or functional fragment thereof, and/or a nucleic acid encoding an ORF2 molecule, or a splice variant or functional fragment thereof. In some embodiments, the anellovector does not comprise an ORF2 molecule, or a splice variant or functional fragment thereof, and/or a nucleic acid encoding an ORF2 molecule, or a splice variant or functional fragment thereof. In some embodiments, producing the anellovector comprises expression of an ORF2 molecule, or a splice variant or functional fragment thereof, but the ORF2 molecule is not incorporated into the anellovector.

***Genetic Elements***

***Genetic Elements comprising RNA***

In some embodiments, a genetic element is or comprises a nucleic acid. In some embodiments, a genetic element is a single-stranded polynucleotide. In some embodiments, a genetic element comprises one or more double stranded regions. In some embodiments, a genetic element comprises RNA. In some embodiments, the genetic element comprises an RNA hairpin structure. In some embodiments, the genetic element is an mRNA, e.g., a chemically modified mRNA. In some embodiments, a genetic element consists of at least 10%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% RNA. In some embodiments, the genetic element comprises a DNA strand and an RNA strand, e.g., wherein at least a portion of the DNA strand hybridizes to at least a portion of the RNA strand.

In some embodiments, the genetic element does not encode any of an *Anellovirus* ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, or ORF2t/3.

In some embodiments, the RNA genetic element encodes an effector, e.g., an effector protein.

In some embodiments, the RNA genetic element is or comprises an effector, e.g., a functional RNA. In some embodiments, RNA is selected from the group consisting of mRNA, rRNA, tRNA (e.g., a TREM), regulatory RNA, non-coding RNA, long non-coding RNA (lncRNA), circular RNA (circRNA), double stranded RNA (dsRNA), guide RNA (gRNA), small interfering RNA (siRNA), short hairpin RNA (shRNA), piwi-interacting RNA (piRNA), small nucleolar RNA (snoRNA), small nuclear RNA (snRNA), extracellular RNA (exRNA), small Cajal body-specific RNA (scaRNA), microRNA (miRNA), and other RNAi molecules.

In some embodiments, the genetic element comprises RNA, e.g., chemically modified RNA. In some embodiments, one or more nucleotides of RNA of a genetic element are chemically modified. In

some embodiments, RNA comprises one or more chemical modifications to one or more bases. In some embodiments, RNA comprises one or more chemical modifications to one or more sugars. In some embodiments, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides of RNA of a genetic element are chemically modified. In some embodiments, RNA comprises one or more backbone modifications. In some embodiments, a modification comprises a non-naturally occurring modification, e.g., a modification described in any one of Tables 5-9. A non-naturally occurring modification can be made according to methods known in the art.

In some embodiments, a genetic element described herein comprises a non-naturally occurring modification provided in Table 5, or a combination thereof.

Table 5: Exemplary non-naturally occurring modifications

| **Name** | **Symbol** | **Base** | **Naturally Occurring** |
|---|---|---|---|
| 7-deaza-adenosine | -- | A | NO |
| Nl-methyl-adenosine | -- | A | NO |
| N6, N6 (dimethyl)adenine | -- | A | NO |
| N6-cis-hydroxy-isopentenyl-adenosine | -- | A | NO |
| a-thio-adenosine | -- | A | NO |
| 2 (amino)adenine | -- | A | NO |
| 2 (aminopropyl)adenine | -- | A | NO |
| 2 (methylthio) N6 (isopentenyl)adenine | -- | A | NO |
| 2-(alkyl)adenine | -- | A | NO |
| 2-(aminoalkyl)adenine | -- | A | NO |
| 2-(aminopropyl)adenine | -- | A | NO |
| 2-(halo)adenine | -- | A | NO |
| 2-(halo)adenine | -- | A | NO |
| 2-(propyl)adenine | -- | A | NO |
| 2' -Amino-2' -deoxy-ATP | -- | A | NO |
| 2' -Azido-2'-deoxy-ATP | -- | A | NO |
| 2' -Deoxy-2'-a-aminoadenosine TP | -- | A | NO |
| 2' -Deoxy-2'-a-azidoadenosine TP | -- | A | NO |
| 6 (alkyl)adenine | -- | A | NO |
| 6 (methyl)adenine | -- | A | NO |

| | | | |
|---|---|---|---|
| 6-(alkyl)adenine | -- | A | NO |
| 6-(methyl)adenine | -- | A | NO |
| 7 (deaza)adenine | -- | A | NO |
| 8 (alkenyl)adenine | -- | A | NO |
| 8 (alkynyl)adenine | -- | A | NO |
| 8 (amino)adenine | -- | A | NO |
| 8 (thioalkyl)adenine | -- | A | NO |
| 8-(alkenyl)adenine | -- | A | NO |
| 8-(alkyl)adenine | -- | A | NO |
| 8-(alkynyl)adenine | -- | A | NO |
| 8-(amino)adenine | -- | A | NO |
| 8-(halo)adenine | -- | A | NO |
| 8-(hydroxyl)adenine | -- | A | NO |
| 8-(thioalkyl)adenine | -- | A | NO |
| 8-(thiol)adenine | -- | A | NO |
| 8-azido-adenosine | -- | A | NO |
| aza adenine | -- | A | NO |
| deaza adenine | -- | A | NO |
| N6 (methyl)adenine | -- | A | NO |
| N6-(isopentyl)adenine | -- | A | NO |
| 7-deaza-8-aza-adenosine | -- | A | NO |
| 7-methyladenine | -- | A | NO |
| 1-Deazaadenosine TP | -- | A | NO |
| 2'Fluoro-N6-Bz-deoxyadenosine TP | -- | A | NO |
| 2' -OMe-2-Amino-ATP | -- | A | NO |
| 2'O-methyl-N6-Bz-deoxyadenosine TP | -- | A | NO |
| 2' -a-Ethynyladenosine TP | -- | A | NO |
| 2-aminoadenine | -- | A | NO |
| 2-Aminoadenosine TP | -- | A | NO |
| 2-Amino-ATP | -- | A | NO |
| 2' -a-Trifluoromethyladenosine TP | -- | A | NO |
| 2-Azidoadenosine TP | -- | A | NO |

| | | | |
|---|---|---|---|
| 2' -b-Ethynyladenosine TP | -- | A | NO |
| 2-Bromoadenosine TP | -- | A | NO |
| 2'-b-Trifluoromethyladenosine TP | -- | A | NO |
| 2-Chloroadenosine TP | -- | A | NO |
| 2' -Deoxy-2',2'-difluoroadenosine TP | -- | A | NO |
| 2' -Deoxy-2'-a-mercaptoadenosine TP | -- | A | NO |
| 2' -Deoxy-2'-a-thiomethoxyadenosine TP | -- | A | NO |
| 2' -Deoxy-2'-b-aminoadenosine TP | -- | A | NO |
| 2' -Deoxy-2'-b-azidoadenosine TP | -- | A | NO |
| 2' -Deoxy-2'-b-bromoadenosine TP | -- | A | NO |
| 2' -Deoxy-2'-b-chloroadenosine TP | -- | A | NO |
| 2' -Deoxy-2'-b-fluoroadenosine TP | -- | A | NO |
| 2' -Deoxy-2'-b-iodoadenosine TP | -- | A | NO |
| 2' -Deoxy-2'-b-mercaptoadenosine TP | -- | A | NO |
| 2' -Deoxy-2'-b-thiomethoxyadenosine TP | -- | A | NO |
| 2-Fluoroadenosine TP | -- | A | NO |
| 2-Iodoadenosine TP | -- | A | NO |
| 2-Mercaptoadenosine TP | -- | A | NO |
| 2-methoxy-adenine | -- | A | NO |
| 2-methylthio-adenine | -- | A | NO |
| 2-Trifluoromethyladenosine TP | -- | A | NO |
| 3-Deaza-3-bromoadenosine TP | -- | A | NO |
| 3-Deaza-3-chloroadenosine TP | -- | A | NO |
| 3-Deaza-3-fluoroadenosine TP | -- | A | NO |
| 3-Deaza-3-iodoadenosine TP | -- | A | NO |
| 3-Deazaadenosine TP | -- | A | NO |
| 4' -Azidoadenosine TP | -- | A | NO |
| 4'-Carbocyclic adenosine TP | -- | A | NO |
| 4' -Ethynyladenosine TP | -- | A | NO |
| 5'-Homo-adenosine TP | -- | A | NO |
| 8-Aza-ATP | -- | A | NO |
| 8-bromo-adenosine TP | -- | A | NO |

| | | | |
|---|---|---|---|
| 8-Trifluoromethyladenosine TP | -- | A | NO |
| 9-Deazaadenosine TP | -- | A | NO |
| 2-aminopurine | -- | A/G | NO |
| 7-deaza-2,6-diaminopurine | -- | A/G | NO |
| 7-deaza-8-aza-2,6-diaminopurine | -- | A/G | NO |
| 7-deaza-8-aza-2-aminopurine | -- | A/G | NO |
| 2,6-diaminopurine | -- | A/G | NO |
| 7-deaza-8-aza-adenine, 7-deaza-2-aminopurine | -- | A/G | NO |
| 4-methylcytidine | -- | C | NO |
| 5-aza-cytidine | -- | C | NO |
| Pseudo-iso-cytidine | -- | C | NO |
| pyrrolo-cytidine | -- | C | NO |
| a-thio-cytidine | -- | C | NO |
| 2-(thio)cytosine | -- | C | NO |
| 2' -Amino-2' -deoxy-CTP | -- | C | NO |
| 2' -Azido-2'-deoxy-CTP | -- | C | NO |
| 2' -Deoxy-2'-a-aminocytidine TP | -- | C | NO |
| 2' -Deoxy-2'-a-azidocytidine TP | -- | C | NO |
| 3 (deaza) 5 (aza)cytosine | -- | C | NO |
| 3 (methyl)cytosine | -- | C | NO |
| 3-(alkyl)cytosine | -- | C | NO |
| 3-(deaza) 5 (aza)cytosine | -- | C | NO |
| 3-(methyl)cytidine | -- | C | NO |
| 4,2'-O-dimethylcytidine | -- | C | NO |
| 5 (halo)cytosine | -- | C | NO |
| 5 (methyl)cytosine | -- | C | NO |
| 5 (propynyl)cytosine | -- | C | NO |
| 5 (trifluoromethyl)cytosine | -- | C | NO |
| 5-(alkyl)cytosine | -- | C | NO |
| 5-(alkyny!)cytosine | -- | C | NO |
| 5-(halo)cytosine | -- | C | NO |
| 5-(propynyl)cytosine | -- | C | NO |

| | | | |
|---|---|---|---|
| 5-(trifluoromethyl)cytosine | -- | C | NO |
| 5-bromo-cytidine | -- | C | NO |
| 5-iodo-cytidine | -- | C | NO |
| 5-propynyl cytosine | -- | C | NO |
| 6-(azo)cytosine | -- | C | NO |
| 6-aza-cytidine | -- | C | NO |
| aza cytosine | -- | C | NO |
| deaza cytosine | -- | C | NO |
| N4 (acetyl)cytosine | -- | C | NO |
| l-methyl-1-deaza-pseudoisocytidine | -- | C | NO |
| 1-methyl-pseudoisocytidine | -- | C | NO |
| 2-methoxy-5-methyl-cytidine | -- | C | NO |
| 2-methoxy-cytidine | -- | C | NO |
| 2-thio-5-methyl-cytidine | -- | C | NO |
| 4-methoxy-1-methyl-pseudoisocytidine | -- | C | NO |
| 4-methoxy-pseudoisocytidine | -- | C | NO |
| 4-thio- l-methyl-1-deaza-pseudoisocytidine | -- | C | NO |
| 4-thio-1-methyl-pseudoisocytidine | -- | C | NO |
| 4-thio-pseudoisocytidine | -- | C | NO |
| 5-aza-zebularine | -- | C | NO |
| 5-methyl-zebularine | -- | C | NO |
| pyrrolo-pseudoisocytidine | -- | C | NO |
| zebularine | -- | C | NO |
| (E)-5-(2-Bromo-vinyl)cytidine TP | -- | C | NO |
| 2,2'-anhydro-cytidine TP hydrochloride | -- | C | NO |
| 2'Fluor-N4-Bz-cytidine TP | -- | C | NO |
| 2'Fluoro-N4-Acetyl-cytidine TP | -- | C | NO |
| 2' -O-Methyl-N4-Acetyl-cytidine TP | -- | C | NO |
| 2'O-methyl-N4-Bz-cytidine TP | -- | C | NO |
| 2' -a-Ethynylcytidine TP | -- | C | NO |
| 2' -a-Trifluoromethylcytidine TP | -- | C | NO |
| 2' -b-Ethynylcytidine TP | -- | C | NO |

| | | | |
|---|---|---|---|
| 2'-b-Trifluoromethylcytidine TP | -- | C | NO |
| 2' -Deoxy-2',2'-difluorocytidine TP | -- | C | NO |
| 2' -Deoxy-2'-a-mercaptocytidine TP | -- | C | NO |
| 2' -Deoxy-2'-a-thiomethoxycytidine TP | -- | C | NO |
| 2' -Deoxy-2'-b-aminocytidine TP | -- | C | NO |
| 2' -Deoxy-2'-b-azidocytidine TP | -- | C | NO |
| 2' -Deoxy-2'-b-bromocytidine TP | -- | C | NO |
| 2' -Deoxy-2'-b-chlorocytidine TP | -- | C | NO |
| 2' -Deoxy-2'-b-fluorocytidine TP | -- | C | NO |
| 2' -Deoxy-2'-b-iodocytidine TP | -- | C | NO |
| 2' -Deoxy-2'-b-mercaptocytidine TP | -- | C | NO |
| 2' -Deoxy-2'-b-thiomethoxycytidine TP | -- | C | NO |
| 2' -O-Methyl-5-(1-propynyl)cytidine TP | -- | C | NO |
| 3'-Ethynylcytidine TP | -- | C | NO |
| 4'-Azidocytidine TP | -- | C | NO |
| 4'-Carbocyclic cytidine TP | -- | C | NO |
| 4' -Ethynylcytidine TP | -- | C | NO |
| 5-(1-Propynyl)ara-cytidine TP | -- | C | NO |
| 5-(2-Chloro-phenyl)-2-thiocytidine TP | -- | C | NO |
| 5-(4-Amino-phenyl)-2-thiocytidine TP | -- | C | NO |
| 5-Aminoallyl-CTP | -- | C | NO |
| 5-Cyanocytidine TP | -- | C | NO |
| 5-Ethynylara-cytidine TP | -- | C | NO |
| 5-Ethynylcytidine TP | -- | C | NO |
| 5'-Homo-cytidine TP | -- | C | NO |
| 5-Methoxycytidine TP | -- | C | NO |
| 5-Trifluoromethyl-Cytidine TP | -- | C | NO |
| N4-Amino-cytidine TP | -- | C | NO |
| N4-Benzoyl-cytidine TP | -- | C | NO |
| pseudoisocytidine | -- | C | NO |
| 6-thio-guanosine | -- | G | NO |
| 7-deaza-guanosine | -- | G | NO |

| | | | |
|---|---|---|---|
| 8-oxo-guanosine | -- | G | NO |
| Nl-methyl-guanosine | -- | G | NO |
| a-thio-guanosine | -- | G | NO |
| 2 (propyl)guanine | -- | G | NO |
| 2-(alky1)guanine | -- | G | NO |
| 2' -Amino-2' -deoxy-GTP | -- | G | NO |
| 2' -Azido-2'-deoxy-GTP | -- | G | NO |
| 2' -Deoxy-2'-a-aminoguanosine TP | -- | G | NO |
| 2' -Deoxy-2'-a-azidoguanosine TP | -- | G | NO |
| 6 (methyl)guanine | -- | G | NO |
| 6-(alky1)guanine | -- | G | NO |
| 6-(methyl)guanine | -- | G | NO |
| 6-methyl-guanosine | -- | G | NO |
| 7 (alkyl)guanine | -- | G | NO |
| 7 (deaza)guanine | -- | G | NO |
| 7 (methyl)guanine | -- | G | NO |
| 7-(alkyl)guanine | -- | G | NO |
| 7-(deaza)guanine | -- | G | NO |
| 7-(methyl)guanine | -- | G | NO |
| 8 (alkyl)guanine | -- | G | NO |
| 8 (alkynyl)guanine | -- | G | NO |
| 8 (halo)guanine | -- | G | NO |
| 8 (thioalkyl)guanine | -- | G | NO |
| 8-(alkenyl)guanine | -- | G | NO |
| 8-(alkyl)guanine | -- | G | NO |
| 8-(alkynyl)guanine | -- | G | NO |
| 8-(amino)guanine | -- | G | NO |
| 8-(halo)guanine | -- | G | NO |
| 8-(hydroxyl)guanine | -- | G | NO |
| 8-(thioalkyl)guanine | -- | G | NO |
| 8-(thiol)guanine | -- | G | NO |
| azaguanine | -- | G | NO |

| | | | |
|---|---|---|---|
| deaza guanine | -- | G | NO |
| N (methyl)guanine | -- | G | NO |
| N-(methyl)guanine | -- | G | NO |
| l-methyl-6-thio-guanosine | -- | G | NO |
| 6-methoxy-guanosine | -- | G | NO |
| 6-thio-7-deaza-8-aza-guanosine | -- | G | NO |
| 6-thio-7-deaza-guanosine | -- | G | NO |
| 6-thio-7-methyl-guanosine | -- | G | NO |
| 7-deaza-8-aza-guanosine | -- | G | NO |
| 7-methyl-8-oxo-guanosine | -- | G | NO |
| N2,N2-dimethyl-6-thio-guanosine | -- | G | NO |
| N2-methyl-6-thio-guanosine | -- | G | NO |
| 1-Me-GTP | -- | G | NO |
| 2'Fluoro-N2-isobutyl-guanosine TP | -- | G | NO |
| 2'O-methyl-N2-isobutyl-guanosine TP | -- | G | NO |
| 2' -a-Ethynylguanosine TP | -- | G | NO |
| 2' -a-Trifluoromethylguanosine TP | -- | G | NO |
| 2' -b-Ethynylguanosine TP | -- | G | NO |
| 2'-b-Trifluoromethylguanosine TP | -- | G | NO |
| 2' -Deoxy-2',2'-difluoroguanosine TP | -- | G | NO |
| 2' -Deoxy-2'-a-mercaptoguanosine TP | -- | G | NO |
| 2' -Deoxy-2'-a-thiomethoxyguanosine TP | -- | G | NO |
| 2' -Deoxy-2'-b-aminoguanosine TP | -- | G | NO |
| 2' -Deoxy-2'-b-azidoguanosine TP | -- | G | NO |
| 2' -Deoxy-2'-b-bromoguanosine TP | -- | G | NO |
| 2' -Deoxy-2'-b-chloroguanosine TP | -- | G | NO |
| 2' -Deoxy-2'-b-fluoroguanosine TP | -- | G | NO |
| 2' -Deoxy-2'-b-iodoguanosine TP | -- | G | NO |
| 2' -Deoxy-2'-b-mercaptoguanosine TP | -- | G | NO |
| 2' -Deoxy-2'-b-thiomethoxyguanosine TP | -- | G | NO |
| 4' -Azidoguanosine TP | -- | G | NO |
| 4'-Carbocyclic guanosine TP | -- | G | NO |

| | | | |
|---|---|---|---|
| 4' -Ethynylguanosine TP | -- | G | NO |
| 5'-Homo-guanosine TP | -- | G | NO |
| 8-bromo-guanosine TP | -- | G | NO |
| 9-Deazaguanosine TP | -- | G | NO |
| N2-isobutyl-guanosine TP | -- | G | NO |
| 7-methylinosine | | A | NO |
| allyamino- thymidine | -- | T | NO |
| aza thymidine | -- | T | NO |
| deaza thymidine | -- | T | NO |
| deoxy-thymidine | -- | T | NO |
| 5-propynyl uracil | -- | U | NO |
| a-thio-uridine | -- | U | NO |
| 1 ( aminoalkylamino-carbonylethylenyl)-2(thio)-pseudouracil | -- | U | NO |
| 1 ( aminoalkylaminocarbonylethylenyl)-2,4-(dithio)pseudouracil | -- | U | NO |
| 1 ( aminoalkylaminocarbonylethylenyl)-4 (thio)pseudouracil | -- | U | NO |
| 1 ( aminoalkylaminocarbonylethylenyl)-pseudouracil | -- | U | NO |
| 1 ( aminocarbonylethylenyl)-2(thio)-pseudouracil | -- | U | NO |
| 1 ( aminocarbonylethylenyl)-2,4-( dithio)pseudouracil | -- | U | NO |
| 1 ( aminocarbonylethylenyl)-4 (thio)pseudouracil | -- | U | NO |
| 1 ( aminocarbonylethylenyl)-pseudouracil | -- | U | NO |
| 1 substituted 2(thio)-pseudouracil | -- | U | NO |
| 1 substituted 2,4-(dithio)pseudouracil | -- | U | NO |
| 1 substituted 4 (thio)pseudouracil | -- | U | NO |
| 1 substituted pseudouracil | -- | U | NO |
| 1-(aminoalkylamino-carbonylethylenyl)-2- (thio)-pseudouracil | -- | U | NO |
| 1 -Methyl-3-(3-amino-3-carboxypropyl) pseudouridine TP | -- | U | NO |
| 1-Methyl-3-(3-amino-3-carboxyproovl)pseudo-UTP | -- | U | NO |
| 1-Methyl-pseudo-UTP | -- | U | NO |

| | | | |
|---|---|---|---|
| 2 (thio)pseudouracil | -- | U | NO |
| 2' deoxy uridine | -- | U | NO |
| 2' fluorouridine | -- | U | NO |
| 2-(thio)uracil | -- | U | NO |
| 2,4-(dithio)psuedouracil | -- | U | NO |
| 2' methyl, 2'amino, 2'azido, 2'fluro-guanosine | -- | U | NO |
| 2' -Amino-2' -deoxy-UTP | -- | U | NO |
| 2' -Azido-2'-deoxy-UTP | -- | U | NO |
| 2' -Azido-deoxyuridine TP | -- | U | NO |
| 2'-O-methylpseudouridine | -- | U | NO |
| 2' deoxy uridine | 2'dU | U | NO |
| 2' fluorouridine | -- | U | NO |
| 2' -Deoxy-2'-a-aminouridine TP | -- | U | NO |
| 2' -Deoxy-2'-a-azidouridine TP | -- | U | NO |
| 2-methylpseudouridine | m3'P | U | NO |
| 3 (3 amino-3 carboxypropyl)uracil | -- | U | NO |
| 4 (thio)pseudouracil | -- | U | NO |
| 4-(thio )pseudouracil | -- | U | NO |
| 4-(thio)uracil | -- | U | NO |
| 4-thiouracil | -- | U | NO |
| 5 (1 ,3-diazole-1-alkyl)uracil | -- | U | NO |
| 5 (2-aminopropyl)uracil | -- | U | NO |
| 5 (aminoalkyl)uracil | -- | U | NO |
| 5 (dimethylaminoalkyl)uracil | -- | U | NO |
| 5 (guanidiniumalkyl)uracil | -- | U | NO |
| 5 (methoxycarbonylmethyl)-2-(thio)uracil | -- | U | NO |
| 5 (methoxycarbonyl-methyl)uracil | -- | U | NO |
| 5 (methyl) 2 (thio)uracil | -- | U | NO |
| 5 (methyl) 2,4 (dithio)uracil | -- | U | NO |
| 5 (methyl) 4 (thio)uracil | -- | U | NO |
| 5 (methylaminomethyl)-2 (thio)uracil | -- | U | NO |
| 5 (methylaminomethyl)-2,4 (dithio)uracil | -- | U | NO |

| | | | |
|---|---|---|---|
| 5 (methylaminomethyl)-4 (thio)uracil | -- | U | NO |
| 5 (propynyl)uracil | -- | U | NO |
| 5 (trifluoromethyl)uracil | -- | U | NO |
| 5-(2-aminopropyl)uracil | -- | U | NO |
| 5-(alky1)-2-(thio)pseudouracil | -- | U | NO |
| 5-(alkyl)-2,4 (dithio)pseudouracil | -- | U | NO |
| 5-(alky1)-4 (thio)pseudouracil | -- | U | NO |
| 5-(alkyl)pseudouracil | -- | U | NO |
| 5-(alkyl)uracil | -- | U | NO |
| 5-(alkynyl)uracil | -- | U | NO |
| 5-(allylamino)uracil | -- | U | NO |
| 5-(cyanoalkyl)uracil | -- | U | NO |
| 5-(dialkylaminoalkyl)uracil | -- | U | NO |
| 5-(dimethylaminoalkyl)uracil | -- | U | NO |
| 5-(guanidiniumalkyl)uracil | -- | U | NO |
| 5-(halo)uracil | -- | U | NO |
| 5-(1,3-diazole-l-alkyl)uracil | -- | U | NO |
| 5-(methoxy)uracil | -- | U | NO |
| 5-(methoxycarbonylmethyl)-2-(thio)uracil | -- | U | NO |
| 5-(methoxycarbonyl-methyl)uracil | -- | U | NO |
| 5-(methyl) 2(thio)uracil | -- | U | NO |
| 5-(methyl) 2,4 (dithio )uracil | -- | U | NO |
| 5-(methyl) 4 (thio)uracil | -- | U | NO |
| 5-(methyl)-2-(thio)pseudouracil | -- | U | NO |
| 5-(methyl)-2,4 (dithio)pseudouracil | -- | U | NO |
| 5-(methyl)-4 (thio)pseudouracil | -- | U | NO |
| 5-(methyl)pseudouracil | -- | U | NO |
| 5-(methylaminomethyl)-2 (thio)uracil | -- | U | NO |
| 5-(methylaminomethyl)-2,4(dithio )uracil | -- | U | NO |
| 5-(methylaminomethyl)-4-(thio)uracil | -- | U | NO |
| 5-(propyny1)uracil | -- | U | NO |
| 5-(trifluoromethyl)uracil | -- | U | NO |

| | | | |
|---|---|---|---|
| 5-aminoallyl-uridine | -- | U | NO |
| 5-bromo-uridine | -- | U | NO |
| 5-iodo-uridine | -- | U | NO |
| 5-uracil | -- | U | NO |
| 6 (azo)uracil | -- | U | NO |
| 6-(azo)uracil | -- | U | NO |
| 6-aza-uridine | -- | U | NO |
| allyamino-uracil | -- | U | NO |
| aza uracil | -- | U | NO |
| deaza uracil | -- | U | NO |
| N3 (methyl)uracil | -- | U | NO |
| P seudo-UTP-1-2-ethanoic acid | -- | U | NO |
| pseudouracil | -- | U | NO |
| 4-Thio-pseudo-UTP | -- | U | NO |
| 1-carboxymethyl-pseudouridine | -- | U | NO |
| l-methyl-1-deaza-pseudouridine | -- | U | NO |
| 1-propynyl-uridine | -- | U | NO |
| l-taurinomethyl-1-methyl-uridine | -- | U | NO |
| l-taurinomethyl-4-thio-uridine | -- | U | NO |
| 1-taurinomethyl-pseudouridine | -- | U | NO |
| 2-methoxy-4-thio-pseudouridine | -- | U | NO |
| 2-thio- l-methyl-1-deaza-pseudouridine | -- | U | NO |
| 2-thio-1-methyl-pseudouridine | -- | U | NO |
| 2-thio-5-aza-uridine | -- | U | NO |
| 2-thio-dihydropseudouridine | -- | U | NO |
| 2-thio-dihydrouridine | -- | U | NO |
| 2-thio-pseudouridine | -- | U | NO |
| 4-methoxy-2-thio-pseudouridine | -- | U | NO |
| 4-methoxy-pseudouridine | -- | U | NO |
| 4-thio-1-methyl-pseudouridine | -- | U | NO |
| 4-thio-pseudouridine | -- | U | NO |
| 5-aza-uridine | -- | U | NO |

| | | | |
|---|---|---|---|
| dihydropseudouridine | -- | U | NO |
| (±)1-(2-Hydroxypropyl)pseudouridine TP | -- | U | NO |
| (2R)- 1 -(2-Hydroxypropyl)pseudouridine TP | -- | U | NO |
| (2S)- 1 -(2-Hydroxypropyl)pseudouridine TP | -- | U | NO |
| (E)-5-(2-Bromo-vinyl)ara-uridine TP | -- | U | NO |
| (E)-5-(2-Bromo-vinyl)uridine TP | -- | U | NO |
| (Z)-5-(2-Bromo-vinyl)ara-uridine TP | -- | U | NO |
| (Z)-5-(2-Bromo-vinyl)uridine TP | -- | U | NO |
| 1-(2,2,2-Trifluoroethyl)-pseudo-UTP | -- | U | NO |
| 1-(2,2,3,3,3-Pentafluoropropyl)pseudouridine TP | -- | U | NO |
| 1-(2,2-Diethoxyethy l)pseudouridine TP | -- | U | NO |
| 1-(2,4,6-Trimethylbenzyl)pseudouridine TP | -- | U | NO |
| 1-(2,4,6-Trimethyl-benzyl)pseudo-UTP | -- | U | NO |
| 1-(2,4,6-Trimethyl-phenyl)pseudo-UTP | -- | U | NO |
| 1-(2-Amino-2-carboxyethyl)pseudo-UTP | -- | U | NO |
| 1-(2-Amino-ethyl)pseudo-UTP | -- | U | NO |
| 1-(2-Hydroxyethyl)pseudouridine TP | -- | U | NO |
| 1-(2-Methoxyethyl)pseudouridine TP | -- | U | NO |
| 1-(3,4-Bis-trifluoromethoxybenzvl)pseudouridine TP | -- | U | NO |
| 1-(3,4-Dimethoxybenzyl)pseudouridine TP | -- | U | NO |
| 1-(3-Amino-3-carboxypropyl)pseudo-UTP | -- | U | NO |
| 1-(3-Amino-propyl)pseudo-UTP | -- | U | NO |
| 1-(3-Cyclopropyl-prop-2-ynyl)pseudouridine TP | -- | U | NO |
| 1-(4-Amino-4-carboxybutyl)pseudo-UTP | -- | U | NO |
| 1-(4-Amino-benzyl)pseudo-UTP | -- | U | NO |
| 1-(4-Amino-buty l)pseudo-UTP | -- | U | NO |
| 1-(4-Amino-phenyl)pseudo-UTP | -- | U | NO |
| 1-(4-Azidobenzyl)pseudouridine TP | -- | U | NO |
| 1-(4-Bromobenzyl)pseudouridine TP | -- | U | NO |

| | | | |
|---|---|---|---|
| 1-(4-Chlorobenzyl)pseudouridine TP | -- | U | NO |
| 1-(4-Fluorobenzyl)pseudouridine TP | -- | U | NO |
| 1-(4-Iodobenzyl)pseudouridine TP | -- | U | NO |
| 1-(4-Methanesulfonylbenzyl)pseudouridine TP | -- | U | NO |
| 1-(4-Methoxybenzy l)pseudouridine TP | -- | U | NO |
| 1-(4-Methoxy-benzyl)pseudo-UTP | -- | U | NO |
| 1-(4-Methoxy-phenyl)pseudo-UTP | -- | U | NO |
| 1-(4-Methylbenzyl)pseudouridine TP | -- | U | NO |
| 1-(4-Methyl-benzyl)pseudo-UTP | -- | U | NO |
| 1-(4-Nitrobenzyl)pseudouridine TP | -- | U | NO |
| 1-(4-Nitro-benzyl)pseudo-UTP | -- | U | NO |
| 1( 4-Nitro-phenyl)pseudo-UTP | -- | U | NO |
| 1-(4-Thiomethoxybenzyl)pseudouridine TP | -- | U | NO |
| 1-(4-Trifluoromethoxybenzyl)pseudouridine TP | -- | U | NO |
| 1-(4-Trifluoromethylbenzyl)pseudouridine TP | -- | U | NO |
| 1-(5-Amino-pentyl)pseudo-UTP | -- | U | NO |
| 1-(6-Amino-hexyl)pseudo-UTP | -- | U | NO |
| 1,6-Dimethyl-pseudo-UTP | -- | U | NO |
| l-[3-(2-{2-[2-(2-Aminoethoxy)-ethoxy]-ethoxy}-ethoxy)-propionyl]pseudouridine TP | -- | U | NO |
| 1-{3-[2-(2-Aminoethoxy)-ethoxy]-propionvl } pseudouridine TP | -- | U | NO |
| 1-Acetylpseudouridine TP | -- | U | NO |
| l-Alkyl-6-(1-propynyl)-pseudo-UTP | -- | U | NO |
| l-Alkyl-6-(2-propynyl)-pseudo-UTP | -- | U | NO |
| l-Alkyl-6-allyl-pseudo-UTP | -- | U | NO |
| l-Alkyl-6-ethynyl-pseudo-UTP | -- | U | NO |
| l -Alkyl-6-homoallyl-pseudo-UTP | -- | U | NO |
| l-Alkyl-6-vinyl-pseudo-UTP | -- | U | NO |
| 1-Allylpseudouridine TP | -- | U | NO |
| 1-Aminomethyl-pseudo-UTP | -- | U | NO |

| | | | |
|---|---|---|---|
| 1-Benzoylpseudouridine TP | -- | U | NO |
| 1-Benzyloxymethylpseudouridine TP | -- | U | NO |
| 1-Benzyl-pseudo-UTP | -- | U | NO |
| l-Biotinyl-PEG2-pseudouridine TP | -- | U | NO |
| 1-Biotinylpseudouridine TP | -- | U | NO |
| 1-Butyl-pseudo-UTP | -- | U | NO |
| 1-Cyanomethylpseudouridine TP | -- | U | NO |
| 1-Cyclobutylmethyl-pseudo-UTP | -- | U | NO |
| 1-Cyclobutyl-pseudo-UTP | -- | U | NO |
| 1-Cycloheptylmethyl-pseudo-UTP | -- | U | NO |
| 1-Cycloheptyl-pseudo-UTP | -- | U | NO |
| 1-Cyclohexylmethyl-pseudo-UTP | -- | U | NO |
| 1-Cyclohexyl-pseudo-UTP | -- | U | NO |
| 1-Cyclooctylmethyl-pseudo-UTP | -- | U | NO |
| 1-Cyclooctyl-pseudo-UTP | -- | U | NO |
| 1-Cyclopentylmethyl-pseudo-UTP | -- | U | NO |
| 1-Cyclopentyl-pseudo-UTP | -- | U | NO |
| 1-Cyclopropylmethyl-pseudo-UTP | -- | U | NO |
| 1-Cyclopropyl-pseudo-UTP | -- | U | NO |
| 1-Ethyl-pseudo-UTP | -- | U | NO |
| 1-Hexyl-pseudo-UTP | -- | U | NO |
| 1-Homoallylpseudouridine TP | -- | U | NO |
| 1-Hydroxymethylpseudouridine TP | -- | U | NO |
| 1-iso-propyl-pseudo-UTP | -- | U | NO |
| 1-Me-2-thio-pseudo-UTP | -- | U | NO |
| 1-Me-4-thio-pseudo-UTP | -- | U | NO |
| 1-Me-alpha-thio-pseudo-UTP | -- | U | NO |
| 1-Methanesulfonylmethylpseudouridine TP | -- | U | NO |
| 1-Methoxymethylpseudouridine TP | -- | U | NO |
| l -Methyl-6-(2,2,2-Trifluoroethyl)pseudo-UTP | -- | U | NO |
| l-Methyl-6-(4-morpholino )-pseudo-DTP | -- | U | NO |

| | | | |
|---|---|---|---|
| l -Methyl-6-(4-thiomorpholino)-pseudo-UTP | -- | U | NO |
| l-Methyl-6-(substituted phenyl)pseudo-UTP | -- | U | NO |
| 1-Methyl-6-amino-pseudo-UTP | -- | U | NO |
| l -Methyl-6-azido-pseudo-UTP | -- | U | NO |
| 1-Methyl-6-bromo-pseudo-UTP | -- | U | NO |
| l-Methyl-6-butyl-pseudo-UTP | -- | U | NO |
| l -Methyl-6-chloro-pseudo-UTP | -- | U | NO |
| 1-Methyl-6-cyano-pseudo-UTP | -- | U | NO |
| l -Methyl-6-dimethylamino-pseudo-UTP | -- | U | NO |
| l -Methyl-6-ethoxy-pseudo-UTP | -- | U | NO |
| l -Methyl-6-ethylcarboxylate-pseudo-UTP | -- | U | NO |
| l-Methyl-6-ethyl-pseudo-UTP | -- | U | NO |
| l -Methyl-6-fluoro-pseudo-UTP | -- | U | NO |
| l-Methyl-6-formyl-pseudo-UTP | -- | U | NO |
| 1-Methyl-6-hydroxyamino-pseudo-UTP | -- | U | NO |
| l -Methyl-6-hydroxy-pseudo-UTP | -- | U | NO |
| l -Methyl-6-iodo-pseudo-UTP | -- | U | NO |
| l-Methyl-6-iso-propyl-pseudo-UTP | -- | U | NO |
| l-Methyl-6-methoxy-pseudo-UTP | -- | U | NO |
| l -Methyl-6-methylamino-pseudo-UTP | -- | U | NO |
| l-Methyl-6-phenyl-pseudo-UTP | -- | U | NO |
| l-Methyl-6-propyl-pseudo-UTP | -- | U | NO |
| l-Methyl-6-tert-butyl-pseudo-UTP | -- | U | NO |
| 1-Methyl-6-trifluoromethoxy-pseudo-UTP | -- | U | NO |
| l-Methyl-6-trifluoromethyl-pseudo-UTP | -- | U | NO |
| 1-Morpholinomethylpseudouridine TP | -- | U | NO |
| 1-Pentyl-pseudo-UTP | -- | U | NO |
| 1-Phenyl-pseudo-UTP | -- | U | NO |
| 1-Piyaloylpseudouridine TP | -- | U | NO |
| 1-Propargylpseudouridine TP | -- | U | NO |
| 1-Propyl-pseudo-UTP | -- | U | NO |

| | | | |
|---|---|---|---|
| 1-propynyl-pseudouridine | -- | U | NO |
| 1-p-tolyl-pseudo-UTP | -- | U | NO |
| 1-tert-Butyl-pseudo-UTP | -- | U | NO |
| 1-Thiomethoxymethylpseudouridine TP | -- | U | NO |
| 1-Thiomorpholinomethylpseudouridine TP | -- | U | NO |
| 1-Trifluoroacetylpseudouridine TP | -- | U | NO |
| 1-Trifluoromethyl-pseudo-UTP | -- | U | NO |
| 1-Vinylpseudouridine TP | -- | U | NO |
| 2,2'-anhydro-uridine TP | -- | U | NO |
| 2' -bromo-deoxyuridine TP | -- | U | NO |
| 2' -F-5-Methyl-2'-deoxy-UTP | -- | U | NO |
| 2' -OMe-5-Me-UTP | -- | U | NO |
| 2' -OMe-pseudo-UTP | -- | U | NO |
| 2' -a-Ethynyluridine TP | -- | U | NO |
| 2' -a-Trifluoromethyluridine TP | -- | U | NO |
| 2' -b-Ethynyluridine TP | -- | U | NO |
| 2' -b-Trifluoromethyluridine TP | -- | U | NO |
| 2' -Deoxy-2',2'-difluorouridine TP | -- | U | NO |
| 2' -Deoxy-2'-a-mercaptouridine TP | -- | U | NO |
| 2' -Deoxy-2'-a-thiomethoxyuridine TP | -- | U | NO |
| 2' -Deoxy-2'-b-aminouridine TP | -- | U | NO |
| 2' -Deoxy-2'-b-azidouridine TP | -- | U | NO |
| 2' -Deoxy-2'-b-bromouridine TP | -- | U | NO |
| 2' -Deoxy-2'-b-chlorouridine TP | -- | U | NO |
| 2' -Deoxy-2'-b-fluorouridine TP | -- | U | NO |
| 2' -Deoxy-2'-b-iodouridine TP | -- | U | NO |
| 2' -Deoxy-2'-b-mercaptouridine TP | -- | U | NO |
| 2' -Deoxy-2'-b-thiomethoxyuridine TP | -- | U | NO |
| 2-methoxy-4-thio-uridine | -- | U | NO |
| 2-methoxyuridine | -- | U | NO |
| 2' -O-Methyl-5-(1-propynyl)uridine TP | -- | U | NO |
| 3-Alkyl-pseudo-UTP | -- | U | NO |

| | | | |
|---|---|---|---|
| 4' -Azidouridine TP | -- | U | NO |
| 4'-Carbocyclic uridine TP | -- | U | NO |
| 4' -Ethynyluridine TP | -- | U | NO |
| 5-(1-Propynyl)ara-uridine TP | -- | U | NO |
| 5-(2-Furanyl)uridine TP | -- | U | NO |
| 5-Cyanouridine TP | -- | U | NO |
| 5-Dimethylaminouridine TP | -- | U | NO |
| 5'-Homo-uridine TP | -- | U | NO |
| 5-iodo-2'-fluoro-deoxyuridine TP | -- | U | NO |
| 5-Phenylethynyluridine TP | -- | U | NO |
| 5-Trideuteromethyl-6-deuterouridine TP | -- | U | NO |
| 5-Trifluoromethyl-Uridine TP | -- | U | NO |
| 5-Vinylarauridine TP | -- | U | NO |
| 6-(2,2,2-Trifluoroethyl)-pseudo-UTP | -- | U | NO |
| 6-(4-Morpholino)-pseudo-DTP | -- | U | NO |
| 6-(4-Thiomorpholino)-pseudo-UTP | -- | U | NO |
| 6-(Substituted-Phenyl)-pseudo-UTP | -- | U | NO |
| 6-Amino-pseudo-UTP | -- | U | NO |
| 6-Azido-pseudo-UTP | -- | U | NO |
| 6-Bromo-pseudo-UTP | -- | U | NO |
| 6-Butyl-pseudo-UTP | -- | U | NO |
| 6-Chloro-pseudo-UTP | -- | U | NO |
| 6-Cyano-pseudo-UTP | -- | U | NO |
| 6-Dimethylamino-pseudo-UTP | -- | U | NO |
| 6-Ethoxy-pseudo-UTP | -- | U | NO |
| 6-Ethylcarboxylate-pseudo-UTP | -- | U | NO |
| 6-Ethyl-pseudo-UTP | -- | U | NO |
| 6-Fluoro-pseudo-UTP | -- | U | NO |
| 6-Formyl-pseudo-UTP | -- | U | NO |
| 6-Hydroxyamino-pseudo-UTP | -- | U | NO |
| 6-Hydroxy-pseudo-UTP | -- | U | NO |
| 6-Iodo-pseudo-UTP | -- | U | NO |

| | | | |
|---|---|---|---|
| 6-iso-Propyl-pseudo-UTP | -- | U | NO |
| 6-Methoxy-pseudo-UTP | -- | U | NO |
| 6-Methylamino-pseudo-UTP | -- | U | NO |
| 6-Methyl-pseudo-UTP | -- | U | NO |
| 6-Phenyl-pseudo-UTP | -- | U | NO |
| 6-Phenyl-pseudo-UTP | -- | U | NO |
| 6-Propyl-pseudo-UTP | -- | U | NO |
| 6-tert-Butyl-pseudo- UTP | -- | U | NO |
| 6-Trifluoromethoxy-pseudo-UTP | -- | U | NO |
| 6-Trifluoromethyl-pseudo-UTP | -- | U | NO |
| Alpha-thio-pseudo-UTP | -- | U | NO |
| Pseudouridine 1-(4-methylbenzenesulfonic acid) TP | -- | U | NO |
| Pseudouridine 1-(4-methylbenzoic acid) TP | -- | U | NO |
| Pseudouridine TP l-[3-(2-ethoxy)]propionic acid | -- | U | NO |
| Pseudouridine TP l-[3-{2-(2-[2-(2-ethoxy )-ethoxy]-ethoxy )-ethoxy}]propionic acid | -- | U | NO |
| Pseudouridine TP l-[3-{2-(2-[2-{2(2-ethoxy )-ethoxy}-ethoxy]-ethoxy )-ethoxy}]propionic acid | -- | U | NO |
| Pseudouridine TP l-[3-{2-(2-[2-ethoxy ]-ethoxy)-ethoxv}]propionic acid | -- | U | NO |
| Pseudouridine TP l-[3-{2-(2-ethoxy)-ethoxv}] propionic acid | -- | U | NO |
| Pseudouridine TP 1-methylphosphonic acid | -- | U | NO |
| Pseudouridine TP 1-methylphosphonic acid diethyl ester | -- | U | NO |
| Pseudo-UTP-N1-3-propionic acid | -- | U | NO |
| Pseudo-UTP-N1-4-butanoic acid | -- | U | NO |
| Pseudo-UTP-N 1-5-pentanoic acid | -- | U | NO |
| Pseudo-UTP-N1-6-hexanoic acid | -- | U | NO |
| Pseudo-UTP-Nl-7-heptanoic acid | -- | U | NO |
| Pseudo-UTP-N1-methy1-p-benzoic acid | -- | U | NO |
| Pseudo-UTP-N1-p-benzoic acid | -- | U | NO |

In some embodiments, a genetic element described herein comprises a modification provided in Table 6, or a combination thereof. The modifications provided in Table 6 occur naturally in RNAs, and may be used herein in a genetic element at a position that does not occur in nature.

Table 6: Additional exemplary modifications

| Name | Symbol | Base | Naturally Occurring |
|---|---|---|---|
| 2-methylthio-N6-(cis-hydroxvisopentenvl)adenosine | ms2i6A | A | YES |
| 2-methylthio-N6-methyladenosine | ms2m6A | A | YES |
| 2-methylthio-N6-threonyl carbamoyladenosine | ms2t6A | A | YES |
| N6-glycinylcarbamoyladenosine | g6A | A | YES |
| N6-isopentenyladenosine | i6A | A | YES |
| N6-methyladenosine | m6A | A | YES |
| N6-threonylcarbamoyladenosine | t6A | A | YES |
| 1,2'-O-dimethyladenosine | mlAm | A | YES |
| 1-methyladenosine | mlA | A | YES |
| 2'-O-methyladenosine | Am | A | YES |
| 2'-O-ribosyladenosine (phosphate) | Ar(p) | A | YES |
| 2-methyladenosine | m2A | A | YES |
| 2-methylthio-N6 isopentenyladenosine | ms2i6A | A | YES |
| 2-methylthio-N6-hydroxynorvalyl carbamoyladenosine | ms2hn6A | A | YES |
| 2' -O-methyladenosine | m6A | A | YES |
| 2' -O-ribosyladenosine (phosphate) | Ar(p) | A | YES |
| isopenteny ladenosine | Iga | A | YES |
| N6-(cis-hydroxyisopentenyl)adenosine | io6A | A | YES |
| N6,2'-O-dimethyladenosine | m6Am | A | YES |
| N6,2'-O-dimethyladenosine | $m^6Am$ | A | YES |
| N6,N6,2'-O-trimethyladenosine | m62Am | A | YES |
| N6,N6-dimethyladenosine | m62A | A | YES |
| N6-acetyladenosine | ac6A | A | YES |
| N6-hydroxynorvalylcarbamoyladenosine | hn6A | A | YES |
| N6-methyl-N6-threonylcarbamoyladenosine | m6t6A | A | YES |
| 2-methyladenosine | $m^2A$ | A | YES |
| 2-methylthio-$N^6$-isopenteny ladenosine | $ms^2i^6A$ | A | YES |
| 2-thiocytidine | s2C | C | YES |

| | | | |
|---|---|---|---|
| 3-methylcytidine | m3C | C | YES |
| 5-formylcytidine | f5C | C | YES |
| 5-hydroxymethylcytidine | hm5C | C | YES |
| 5-methylcytidine | m5C | C | YES |
| N4-acetylcytidine | ac4C | C | YES |
| 2'-O-methylcytidine | Cm | C | YES |
| 2' -O-methylcytidine | Cm | C | YES |
| 5,2'-O-dimethylcytidine | m5Cm | C | YES |
| 5-formyl-2'-O-methylcytidine | f5Cm | C | YES |
| lysidine | k2C | C | YES |
| N4,2'-O-dimethy lcytidine | m4Cm | C | YES |
| N4-acetyl-2'-O-methylcytidine | ac4Cm | C | YES |
| N4-methylcytidine | m4C | C | YES |
| N4,N4-Dimethyl-2'-OMe-Cytidine TP | -- | C | YES |
| 7-methylguanosine | m7G | G | YES |
| N2,2'-O-dimethylguanosine | m2Gm | G | YES |
| N2-methylguanosine | m2G | G | YES |
| wyosine | imG | G | YES |
| 1,2'-O-dimethylguanosine | mlGm | G | YES |
| 1-methylguanosine | mlG | G | YES |
| 2'-O-methylguanosine | Gm | G | YES |
| 2'-O-ribosylguanosine (phosphate) | Gr(p) | G | YES |
| 2' -O-methylguanosine | Gm | G | YES |
| 2'-O-ribosylguanosine (phosphate) | Gr(p) | G | YES |
| 7-aminomethyl-7-deazaguanosine | preQl | G | YES |
| 7-cyano-7-deazaguanosine | preQ0 | G | YES |
| archaeosine | G+ | G | YES |
| methylwyosine | mimG | G | YES |
| N2,7-dimethylguanosine | m2,7G | G | YES |
| N2,N2,2'-O-trimethylguanosine | m22Gm | G | YES |
| N2,N2,7-trimethylguanosine | m2,2,7G | G | YES |
| N2,N2-dimethylguanosine | m22G | G | YES |

| | | | |
|---|---|---|---|
| N2, 7,2 '-O-trimethylguanosine | m2 7 ' Gm | G | YES |
| 1-methylinosine | mll | A | YES |
| inosine | I | A | YES |
| 1,2'-O-dimethylinosine | mlim | A | YES |
| 2'-O-methylinosine | Im | A | YES |
| 2' -O-methylinosine | Im | A | YES |
| epoxyqueuosine | oQ | G | YES |
| galactosyl-queuosine | galQ | G | YES |
| mannosyl-queuosine | manQ | G | YES |
| 2' -O-methyluridine | -- | U | YES |
| 2-thiouridine | s2U | U | YES |
| 3-methyluridine | m3U | U | YES |
| 5-carboxymethyluridine | cm5U | U | YES |
| 5-hydroxyuridine | ho5U | U | YES |
| 5-methyluridine | m5U | U | YES |
| 5-taurinomethyl-2-thiouridine | rm5s2U | U | YES |
| 5-taurinomethyluridine | rm5U | U | YES |
| dihydrouridine | D | U | YES |
| pseudouridine | Q | U | YES |
| (3-(3-amino-3-carboxypropyl)uridine | acp3U | U | YES |
| l-methyl-3-(3-amino-5-carboxypropyl)pseudouridine | mlacp3'P | U | YES |
| | | | |
| 1-methylpseduouridine | ml'P | U | YES |
| 1-methyl-pseudouridine | -- | U | YES |
| 2'-0-methyluridine | Um | U | YES |
| 2'-0-methylpseudouridine | 'Pm | U | YES |
| 2' -0-methyluridine | Um | U | YES |
| 2-thio-2'-0-methyluridine | s2Um | U | YES |
| 3-(3-amino-3-carboxypropyl)uridine | acp3U | U | YES |
| 3,2'-0-dimethyluridine | m3Um | U | YES |
| 3-Methyl-pseudo-Uridine TP | -- | U | YES |
| 4-thiouridine | s4U | U | YES |

| | | | |
|---|---|---|---|
| 5-(carboxyhydroxymethyl)uridine | chm5U | U | YES |
| 5-(carboxyhydroxymethyl)uridine methyl ester | mchm5U | U | YES |
| 5,2'-0-dimethyluridine | m5Um | U | YES |
| 5,6-dihydro-uridine | -- | U | YES |
| 5-aminomethy1-2-thiouridine | nm5s2U | U | YES |
| 5-carbamoylmethyl-2'-0-methyluridine | ncm5Um | U | YES |
| 5-carbamoylmethyluridine | ncm5U | U | YES |
| 5-carboxyhydroxymethyluridine | -- | U | YES |
| 5-carboxyhydroxymethyluridine methyl ester | -- | U | YES |
| 5-carboxymethylaminomethyl-2'-0-methyluridine | cmnm5Um | U | YES |
| 5-carboxymethylaminomethyl-2-thiouridine | cmnm5s2U | U | YES |
| 5-carboxymethylaminomethyl-2-thiouridine | -- | U | YES |
| 5-carboxymethylaminomethyluridine | cmnm5U | U | YES |
| 5-carboxymethylaminomethyluridine | -- | U | YES |
| 5-Carbamoylmethyluridine TP | -- | U | YES |
| 5-methoxycarbonylmethyl-2'-0-methyluridine | mcm5Um | U | YES |
| 5-methoxycarbonylmethy1-2-thiouridine | mcm5s2U | U | YES |
| 5-methoxycarbonylmethyluridine | mcm5U | U | YES |
| 5-methoxyuridine | mo5U | U | YES |
| 5-methyl-2-thiouridine | m5s2U | U | YES |
| 5-methylaminomethyl-2-selenouridine | mnm5se2U | U | YES |
| 5-methylaminomethyl-2-thiouridine | mnm5s2U | U | YES |
| 5-methylaminomethyluridine | mnm5U | U | YES |
| 5-Methyldihydrouridine | -- | U | YES |
| 5-0xyacetic acid- Uridine TP | -- | U | YES |
| 5-0xyacetic acid-methyl ester-Uridine TP | -- | U | YES |
| Nl-methyl-pseudo-uridine | -- | U | YES |
| uridine 5-oxyacetic acid | cmo5U | U | YES |
| uridine 5-oxyacetic acid methyl ester | mcmo5U | U | YES |
| 3-(3-Amino-3-carboxypropyl)-Uridine | -- | U | YES |

| | | | |
|---|---|---|---|
| TP | | | |
| 5-(iso-Pentenylaminomethyl)- 2-thiouridine TP | -- | U | YES |
| 5-(iso-Pentenylaminomethyl)-2 '-0-methyluridine TP | -- | U | YES |
| 5-(iso-Pentenylaminomethyl)uridine TP | -- | U | YES |
| wybutosine | yW | A/T | YES |
| hydroxywybutosine | OHyW | A/T | YES |
| isowyosine | imG2 | A/T | YES |
| peroxywybutosine | o2yW | A/T | YES |
| undermodified hydroxywybutosine | OHyW* | A/T | YES |
| 4-demethylwyosine | imG-14 | A/T | YES |

In an embodiment, a genetic element described herein comprises a non-naturally occurring modification provided in Table 7, or a combination thereof.

Table 7: Additional exemplary non-naturally occurring modifications

| **Name** |
|---|
| 2,6-(diamino)purine |
| 1-(aza)-2-(thio)-3-(aza)-phenoxazin-1-yl |
| 1,3-( diaza)-2-( oxo )-phenthiazin-1-yl |
| 1,3-(diaza)-2-(oxo)-phenoxazin-1-yl |
| 1,3,5-(triaza)-2,6-(dioxa)-naphthalene |
| 2 (amino)purine |
| 2,4,5-(trimethyl)phenyl |
| 2' methyl, 2'amino, 2'azido, 2'fluro-cytidine |
| 2' methyl, 2'amino, 2'azido, 2'fluro-adenine |
| 2'methyl, 2'amino, 2'azido, 2'fluro-uridine |
| 2' -amino-2' -deoxyribose |
| 2-amino-6-Chloro-purine |
| 2-aza-inosinyl |
| 2' -azido-2'-deoxyribose |
| 2'fluoro-2 '-deoxyribose |
| 2' -fluoro-modified bases |
| 2' -O-methyl-ribose |
| 2-oxo-7-aminopyridopyrimidin-3-yl |
| 2-oxo-pyridopyrimidine-3-yl |
| 2-pyridinone |

| |
|---|
| 3 nitropyrrole |
| 3-(methyl)-7-(propynyl)isocarbostyrilyl |
| 3-(methyl)isocarbostyrilyl |
| 4-(fluoro)-6-(methyl)benzimidazole |
| 4-(methyl)benzimidazole |
| 4-(methyl)indolyl |
| 4,6-(dimethyl)indolyl |
| 5 nitroindole |
| 5 substituted pyrimidines |
| 5-(methyl)isocarbostyrilyl |
| 5-nitroindole |
| 6-(aza)pyrimidine |
| 6-(azo)thymine |
| 6-(methyl)-7-(aza)indolyl |
| 6-chloro-purine |
| 6-phenyl-pyrrolo-pyrimidin-2-on-3-yl |
| 7-(aminoalkylhydroxy)-1-(aza)-2-(thio )-3-(aza)-phenthiazin-1-yl |
| 7-(aminoalkylhydroxy)-1-(aza)-2-(thio)-3-(aza)-phenoxazin-1-yl |
| 7-(aminoalkylhydroxy)-1,3-(diaza)-2-(oxo)-phenoxazin-1-yl |
| 7-(aminoalkylhydroxy)-1,3-( diaza)-2-( oxo )-phenthiazin-1-yl |
| 7-(aminoalkylhydroxy)-1,3-( diaza)-2-(oxo)-phenoxazin-l-yl |
| 7-(aza)indolyl |
| 7-(guanidiniumalkylhydroxy)-1-(aza)-2-(thio )-3-(aza)-phenoxazinl-yl |
| 7-(guanidiniumalkylhydroxy)-1-(aza)-2-(thio )-3-(aza)-phenthiazin-1-yl |
| 7-(guanidiniumalkylhydroxy)-1-(aza)-2-(thio)-3-(aza)-phenoxazin-1-yl |
| 7-(guanidiniumalkylhydroxy)-1,3-(diaza)-2-(oxo)-phenoxazin-1-yl |
| 7-(guanidiniumalkyl-hydroxy)-1,3-( diaza)-2-( oxo )-phenthiazin-1-yl |
| 7-(guanidiniumalkylhydroxy)-1,3-(diaza)-2-( oxo )-phenoxazin-1-yl |
| 7-(propynyl)isocarbostyrilyl |
| 7-(propynyl)isocarbostyrilyl, propynyl-7-(aza)indolyl |
| 7-deaza-inosinyl |
| 7-substituted 1-(aza)-2-(thio)-3-(aza)-phenoxazin-1-yl |

| |
|---|
| 7-substituted 1,3-(diaza)-2-(oxo)-phenoxazin-1-yl |
| 9-(methyl)-imidizopyridinyl |
| aminoindolyl |
| anthracenyl |
| bis-ortho-(aminoalkylhydroxy)-6-phenyl-pyrrolo-nvrimidin-2-on-3-yl |
| bis-ortho-substituted-6-phenyl-pyrrolo-pyrimidin-2-on-3-yl |
| difluorotolyl |
| hypoxanthine |
| imidizopyridinyl |
| inosinyl |
| isocarbostyrilyl |
| isoguanine |
| N2-substituted purines |
| N6-methyl-2-amino-purine |
| N6-substituted purines |
| N-alkylated derivative |
| napthalenyl |
| nitrobenzimidazolyl |
| nitroimidazolyl |
| nitroindazolyl |
| nitropyrazolyl |
| nubularine |
| O6-substituted purines |
| O-alkylated derivative |
| ortho-(aminoalkylhydroxy)-6-phenyl-pyrrolo-pyrimidin-2-on-3-yl |
| ortho-substituted-6-phenyl-pyrrolo-pyrimidin-2-on-3-yl |
| Oxoformycin TP |
| para-(aminoalkylhydroxy)-6-phenyl-pyrrolo-pyrimidin-2-on-3-yl |
| para-substituted-6-phenyl-pyrrolo-pyrimidin-2-on-3-yl |
| pentacenyl |
| phenanthracenyl |
| phenyl |
| propynyl-7-(aza)indolyl |
| pyrenyl |
| pyridopyrimidin-3-yl |
| pyridopyrimidin-3-yl, 2-oxo-7-amino-pyridopyrimidin-3-yl |
| pyrrolo-pyrimidin-2-on-3-yl |

| |
|---|
| pyrrolopyrimidinyl |
| pyrrolopyrizinyl |
| stilbenzyl |
| substituted 1,2,4-triazoles |
| tetracenyl |
| tubercidine |
| xanthine |
| Xanthosine-5 '-TP |
| 2-thio-zebularine |
| 5-aza-2-thio-zebularine |
| 7-deaza-2-amino-purine |
| pyridin-4-one ribonucleoside |
| 2-Amino-riboside-TP |
| Formycin A TP |
| Formycin B TP |
| Pyrrolosine TP |
| 2' -OH-ara-adenosine TP |
| 2' -OH-ara-cytidine TP |
| 2' -OH-ara-uridine TP |
| 2' -OH-ara-guanosine TP |
| 5-(2-carbomethoxyvinyl)uridine TP |
| N6-(19-Amino-pentaoxanonadecyl)adenosine TP |

In an embodiment, a genetic element described herein comprises a non-naturally occurring modification provided in Table 8, or a combination thereof.

Table 8: Exemplary backbone modifications

| **Name** |
|---|
| 3'-alkylene phosphonates |
| 3'-amino phosphoramidate |
| alkene containing backbones |
| aminoalkylphosphoramidates |
| aminoalkylphosphotriesters |
| boranophosphates |
| -CH2-0-N(CH3)-CH2- |
| -CH2-N(CH3)-N(CH3)-CH2- |
| -CH2-NH-CH2- |
| chiral phosphonates |

| |
|---|
| chiral phosphorothioates |
| formacetyl and thioformacetyl backbones |
| methylene (methylimino) |
| methylene formacetyl and thioformacetyl backbones |
| methyleneimino and methylenehydrazino backbones |
| morpholino linkages |
| -N(CH3)-CH2-CH2- |
| oligonucleosides with heteroatom intemucleoside linkage |
| phosphinates |
| phosphoramidates |
| phosphorodithioates |
| phosphorothioate intemucleoside linkages |
| phosphorothioates |
| phosphotriesters |
| PNA |
| siloxane backbones |
| sulfamate backbones |
| sulfide sulfoxide and sulfone backbones |
| sulfonate and sulfonamide backbones |
| thionoalkylphosphonates |
| thionoalkylphosphotriesters |
| thionophosphoramidates |

In an embodiment, a genetic element described herein comprises a non-naturally occurring modification provided in Table 9, or a combination thereof.

Table 9: Exemplary non-naturally occurring backbone modificiations

| **Name of synthetic backbone modifications** |
|---|
| Phosphorothioate |
| Constrained nucleic acid (CNA) |
| 2' O'methylation |
| 2'-O-methoxyethylribose (MOE) |
| 2' Fluoro |
| Locked nucleic acid (LNA) |
| (*S*)-constrained ethyl (*cEt*) |
| Fluoro hexitol nucleic acid (FHNA) |

| |
|---|
| 5'phosphorothioate |
| Phosphorodiamidate Morpholino Oligomer (PMO) |
| Tricyclo-DNA (tcDNA) |
| (*S*) 5'-C-methyl |
| (*E*)-vinylphosphonate |
| Methyl phosphonate |
| (*S*) 5'-C-methyl with phosphate |

In some embodiments, the genetic element comprises a cap. A cap is typically placed at the 5' end of an mRNA, but a cap can also be positioned at the 3' end of an RNA. In some embodiments, a cap protects the genetic element from exonuclease degradation, and can help in delivery and/or localization within a cell. The cap can be present at the 5'-terminus (5'-cap) or at the 3'-terminal (3'-cap) or can be present on both termini. Non-limiting examples of a 5'-cap include, but are not limited to, glyceryl, inverted deoxy abasic residue (moiety); 4',5'-methylene nucleotide; 1-(beta-D-erythrofuranosyl) nucleotide, 4'-thio nucleotide; carbocyclic nucleotide; 1,5-anhydrohexitol nucleotide; L-nucleotides; alpha-nucleotides; modified base nucleotide; phosphorodithioate linkage; threo-pentofuranosyl nucleotide; acyclic 3',4'-seco nucleotide; acyclic 3,4-dihydroxybutyl nucleotide; acyclic 3,5-dihydroxypentyl nucleotide, 3'-3'-inverted nucleotide moiety; 3'-3'-inverted abasic moiety; 3'-2'-inverted nucleotide moiety; 3'-2'-inverted abasic moiety; 1,4-butanediol phosphate; 3'-phosphoramidate; hexylphosphate; aminohexyl phosphate; 3'-phosphate; 3'-phosphorothioate; phosphorodithioate; or bridging or non-bridging methylphosphonate moiety.

Non-limiting examples of the 3'-cap include, but are not limited to, glyceryl, inverted deoxy abasic residue (moiety), 4', 5'-methylene nucleotide; 1-(beta-D-erythrofuranosyl) nucleotide; 4'-thio nucleotide, carbocyclic nucleotide; 5'-amino-alkyl phosphate; 1,3-diamino-2-propyl phosphate; 3-aminopropyl phosphate; 6-aminohexyl phosphate; 1,2-aminododecyl phosphate; hydroxypropyl phosphate; 1,5-anhydrohexitol nucleotide; L-nucleotide; alpha-nucleotide; modified base nucleotide; phosphorodithioate; threo-pentofuranosyl nucleotide; acyclic 3',4'-seco nucleotide; 3,4-dihydroxybutyl nucleotide; 3,5-dihydroxypentyl nucleotide, 5'-5'-inverted nucleotide moiety; 5'-5'-inverted abasic moiety; 5'-phosphoramidate; 5'-phosphorothioate; 1,4-butanediol phosphate; 5'-amino; bridging and/or non-bridging 5'-phosphoramidate, phosphorothioate and/or phosphorodithioate, bridging or non bridging methylphosphonate and 5'-mercapto moieties (for more details see Beaucage and Iyer, 1993, Tetrahedron 49, 1925; incorporated by reference herein).

In some embodiments, the genetic element comprises a poly-A tail. In some embodiments, a poly-A tail comprises at least about 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, or 100 adenosines in length. In some embodiments, RNA lacks a poly-A tail. In some embodiments, wherein the RNA lacks a poly-A tail, the RNA comprises no more than about 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, or 100 sequential adenosines.

In some embodiments, a genetic element is linear. In some embodiments, a genetic element is circular. In some embodiments, a genetic element comprises a first region and a second region that can hybridize with the first region. In some embodiments, a genetic element comprises a first region and a second region that can hybridize with the first region to form a circle. In some embodiments, a genetic element does not comprise a 5' or a 3' end. In some embodiments, a genetic element does not comprise one or both of a free phosphate and a free sugar. In some embodiments, every phosphate in a genetic element is coalently linked to a first sugar by a first oxygen atom comprised by the phosphate and a second sugar by a second oxygen atom comprised by a phosphate. In some embodiments, every sugar in a genetic element is covalently linked to a first phosphate by a first carbon atom comprised by the sugar and a second phosphate by a second carbon atom comprised by the sugar. In some embodiments, a genetic element is produced by circularizing a linear RNA. Circular RNAs are described, e.g., in US Patent Publication 20200306286, which is herein incorporated by reference in its entirety.

In some embodiments, a genetic element is about 10-20, 20-30, 30-40, 50-60 60-70, 70-80, 80-90, 90-100, 100-125, 125-150, 150-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, 900-1000, 1000-1500, 1500-2000, 2000-2500, 2500-3000, 3000-3500, 3500-4000, or 4000-4500 nucleotides in length. In some embodiments a genetic element is at least about 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000, or 4500 nucleotides in length.

***RNA-only genetic elements***

In some embodiments, a genetic element consists of or consists essentially of RNA. For example, in some embodiments, a genetic element is substantially free of DNA. In some embodiments, a genetic element comprises single stranded RNA. In some embodiments, a genetic element comprises at least one double stranded region. In some embodiments, a double stranded region of a genetic element comprises a region of RNA pairing with RNA.

***Hybridized RNA-ssDNA genetic elements***

In some embodiments, a genetic element comprises a DNA region. In some embodiments, a genetic element comprising RNA further comprises a DNA region. For example, a genetic element may

be single stranded, wherein a first portion of the single strand comprises ribonucleotides and a second portion of the single strand comprises deoxyribonucleotides. In some embodiments, a genetic element comprising a DNA region comprises one or more DNA nucleotides with chemical modification. In some embodiments, a genetic element comprises a DNA region, wherein all nucleotides of the DNA region are chemically modified.

In some embodiments, at least a portion of a genetic element is single stranded. In some embodiments, a genetic element is single stranded. In some embodiments, a genetic element comprises ssDNA. In some embodiments, a genetic element comprises a double stranded region. In some embodiments, a double stranded region of a genetic element comprises a region of RNA pairing with RNA. In some such embodiments, a double stranded region of a genetic element comprises a region of DNA pairing with RNA. In some embodiments, at least a portion of the DNA region hybridizes to at least a portion of the RNA of the genetic element.

In some embodiments, a DNA region is about 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, 150-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, or 900-1000 nucleotides in length.

***RNA/DNA conjugates***

In some embodiments, a genetic element comprises a DNA region. In some embodiments, a genetic element comprising RNA further comprises a DNA region. In some embodiments, a genetic element comprising a DNA region comprises one or more DNA nucleotides with chemical modification. In some embodiments, a genetic element comprising a DNA region, wherein all nucleotides of the DNA region are chemically modified.

In some embodiments, at least a portion of a genetic element is single stranded. In some embodiments, a genetic element is single stranded. In some embodiments, a genetic element comprises ssDNA. In some embodiments, a genetic element comprises a double stranded region. In some embodiments, a double stranded region of a genetic element comprises a region of RNA pairing with RNA. In some such embodiments, a double stranded region of a genetic element comprises a region of DNA pairing with RNA. In some embodiments, wherein a genetic element comprises RNA, a DNA region is covalently linked to the RNA of the genetic element.

In some embodiments, a DNA region is about 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, 150-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, or 900-1000 nucleotides in length.

***Genetic Element Constructs***

In some embodiments, a genetic element is produced from a genetic element construct. For instance, in some embodiments, the genetic element construct is DNA, e.g., double stranded DNA, and the genetic element may be produced by transcription, generating an RNA genetic element.

The genetic element of an anellovector as described herein may be produced from a genetic element construct that comprises a genetic element region and optionally other sequence such as a bacmid (e.g., comprising a baculovirus genome or a fragment thereof, e.g., one or more baculovirus elements) or donor vector backbone. In some embodiments, the genetic element construct comprises an Anellovirus 5' UTR (e.g., as described herein). A genetic element construct may be any nucleic acid construct suitable for delivery of the sequence of the genetic element into a host cell or cell-free system in which the genetic element can be enclosed within a proteinaceous exterior. In some embodiments, the genetic element construct comprises a promoter. In some embodiments, transcription from the genetic element construct produces an RNA genetic element.

In some embodiments, the genetic element construct is a linear nucleic acid molecule. In some embodiments, the genetic element construct is a circular nucleic acid molecule (e.g., a plasmid, bacmid, donor vector, or a minicircle, e.g., as described herein). The genetic element construct may, in some embodiments, be double-stranded. In other embodiments, the genetic element is single-stranded. In some embodiments, the genetic element construct comprises DNA. In some embodiments, the genetic element construct comprises RNA. In some embodiments, the genetic element construct comprises one or more modified nucleotides.

*Plasmids*

In some embodiments, the genetic element construct is a plasmid. The plasmid will generally comprise the sequence of a genetic element as described herein as well as an origin of replication suitable for replication in a host cell (e.g., a bacterial origin of replication for replication in bacterial cells) and a selectable marker (e.g., an antibiotic resistance gene). In some embodiments, the sequence of the genetic element can be excised from the plasmid. In some embodiments, the plasmid is capable of replication in a bacterial cell. In some embodiments, the plasmid is capable of replication in a mammalian cell (e.g., a human cell). In some embodiments, a plasmid is at least 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, or 5000 bp in length. In some embodiments, the plasmid is less than 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, or 10,000 bp in length. In some embodiments, the plasmid has a length between 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, 900-1000, 1000-1500, 1500-2000, 2000-2500, 2500-3000, 3000-4000, or 4000-5000 bp.

*Small circular nucleic acid constructs*

In some embodiments, the genetic element construct is a circular nucleic acid construct, e.g., lacking a vector backbone (e.g., lacking a bacterial origin of replication and/or selectable marker). In

embodiments, the genetic element is a single- or double-stranded circular nucleic acid construct. In embodiments, the circular nucleic acid construct is produced by *in vitro* circularization (IVC), e.g., as described herein. In embodiments, the double-stranded circular nucleic acid construct can be introduced into a host cell, in which it can be converted into or used as a template for generating single-stranded circular genetic elements, e.g., as described herein. In some embodiments, the circular nucleic acid constructdoes not comprise a plasmid backbone or a functional fragment thereof. In some embodiments, the circular nucleic acid construct is at least 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, 3100, 3200, 3300, 3400, 3500, 3600, 3700, 3800, 3900, 4000, 4100, 4200, 4300, 4400, or 4500 bp in length. In some embodiments, the circular nucleic acid construct is less than 2900, 3000, 3100, 3200, 3300, 3400, 3500, 3600, 3700, 3800, 3900, 4000, 4100, 4200, 4300, 4400, 4500, 4600, 4700, 4800, 4900, 5000, 5500, or 6000 bp in length. In some embodiments, the circular nucleic acid constructis between 2000-2100, 2100-2200, 2200-2300, 2300-2400, 2400-2500, 2500-2600, 2600-2700, 2700-2800, 2800-2900, 2900-3000, 3000-3100, 3100-3200, 3200-3300, 3300-3400, 3400-3500, 3500-3600, 3600-3700, 3700-3800, 3800-3900, 3900-4000, 4000-4100, 4100-4200, 4200-4300, 4300-4400, or 4400-4500 bp in length. In some embodiments, the circular nucleic acid construct is a minicircle.

*Cis/Trans Constructs*

In some embodiments, a genetic element construct (e.g., a bacmid or donor vector) as described herein comprises one or more sequences encoding one or more Anellovirus ORFs, e.g., proteinaceous exterior components (e.g., polypeptides encoded by an Anellovirus ORF1 nucleic acid, e.g., as described herein). For example, the genetic element construct may comprise a nucleic acid sequence encoding an Anellovirus ORF1 molecule. Such genetic element constructs can be suitable for introducing the genetic element and the Anellovirus ORF(s) into a host cell in *cis*. In other embodiments, a genetic element construct as described herein does not comprise sequences encoding one or more Anellovirus ORFs, e.g., proteinaceous exterior components (e.g., polypeptides encoded by an Anellovirus ORF1 nucleic acid, e.g., as described herein). For example, the genetic element construct may not comprise a nucleic acid sequence encoding an Anellovirus ORF1 molecule. Such genetic element constructs can be suitable for introducing the genetic element into a host cell, with the one or more Anellovirus ORFs to be provided in *trans* (e.g., via introduction of a second nucleic acid construct encoding one or more of the Anellovirus ORFs, or via an Anellovirus ORF cassette integrated into the genome of the host cell). In some embodiments, the genetic element construct comprises a backbone suitable for replication of the nucleic acid construct in insect cells (e.g., Sf9 cells), e.g., a baculovirus backbone region. In some embodiments, the genetic element construct comprises a backbone region suitable for replication of the genetic element construct in bacterial cells (e.g., *E. coli* cells, e.g., DH 10Bac cells).

In some embodiments, the genetic element construct (e.g., bacmid or donor vector) comprises a sequence encoding an Anellovirus ORF1 molecule, or a splice variant or functional fragment thereof (e.g., a jelly-roll region, e.g., as described herein). In embodiments, the portion of the genetic element that does not comprise the sequence of the genetic element comprises the sequence encoding the Anellovirus ORF1 molecule, or splice variant or functional fragment thereof (e.g., in a cassette comprising a promoter and the sequence encoding the Anellovirus ORF1 molecule, or splice variant or functional fragment thereof). In further embodiments, the portion of the construct comprising the sequence of the genetic element comprises a sequence encoding an Anellovirus ORF1 molecule, or a splice variant or functional fragment thereof (e.g., a jelly-roll region, e.g., as described herein). In embodiments, enclosure of such a genetic element in a proteinaceous exterior (e.g., as described herein) produces a replication-component anellovector (e.g., an anellovector that upon infecting a cell, enables the cell to produce additional copies of the anellovector without introducing further nucleic acid constructs, e.g., encoding one or more Anellovirus ORFs as described herein, into the cell).

In other embodiments, the genetic element does not comprise a sequence encoding an Anellovirus ORF1 molecule, or a splice variant or functional fragment thereof (e.g., a jelly-roll region, e.g., as described herein). In embodiments, enclosure of such a genetic element in a proteinaceous exterior (e.g., as described herein) produces a replication-incompetent anellovector (e.g., an anellovector that, upon infecting a cell, does not enable the infected cell to produce additional anellovectors, e.g., in the absence of one or more additional constructs, e.g., encoding one or more Anellovirus ORFs as described herein).

*Expression Cassettes*

In some embodiments, a genetic element construct (e.g., bacmid or donor vector) comprises one or more cassettes for expression of a polypeptide or noncoding RNA (e.g., a miRNA or an siRNA). In some embodiments, the genetic element construct comprises a cassette for expression of an effector (e.g., an exogenous or endogenous effector), e.g., a polypeptide or noncoding RNA, as described herein. In some embodiments, the genetic element construct comprises a cassette for expression of an Anellovirus protein (e.g., an Anellovirus ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, or ORF1/2, or a functional fragment thereof). The expression cassettes may, in some embodiments, be located within the genetic element sequence. In embodiments, an expression cassette for an effector is located within the genetic element sequence. In embodiments, an expression cassette for an Anellovirus protein is located within the genetic element sequence. In other embodiments, the expression cassettes are located at a position within the genetic element construct outside of the sequence of the genetic element (e.g., in the backbone). In embodiments, an expression cassette for an Anellovirus protein is located at a position within the genetic element construct outside of the sequence of the genetic element (e.g., in the backbone).

A polypeptide expression cassette generally comprises a promoter and a coding sequence encoding a polypeptide, e.g., an effector (e.g., an exogenous or endogenous effector as described herein) or an Anellovirus protein (e.g., a sequence encoding an Anellovirus ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, or ORF1/2, or a functional fragment thereof). Exemplary promoters that can be included in an polypeptide expression cassette (e.g., to drive expression of the polypeptide) include, without limitation, constitutive promoters (e.g., CMV, RSV, PGK, EF1a, or SV40), cell or tissue-specific promoters (e.g., skeletal α-actin promoter, myosin light chain 2A promoter, dystrophin promoter, muscle creatine kinase promoter, liver albumin promoter, hepatitis B virus core promoter, osteocalcin promoter, bone sialoprotein promoter, CD2 promoter, immunoglobulin heavy chain promoter, T cell receptor a chain promoter, neuron-specific enolase (NSE) promoter, or neurofilament light-chain promoter), and inducible promoters (e.g., zinc-inducible sheep metallothionine (MT) promoter; the dexamethasone (Dex)-inducible mouse mammary tumor virus (MMTV) promoter; the T7 polymerase promoter system, tetracycline-repressible system, tetracycline-inducible system, RU486-inducible system, rapamycin-inducible system), e.g., as described herein. In some embodiments, the expression cassette further comprises an enhancer, e.g., as described herein.

***Production of RNA-based genetic elements***

An RNA-based genetic element may be produced by a variety of methods. For example, a genetic element construct comprising DNA can be transcribed to produce a genetic element that comprises RNA, e.g., as described above. The transcription may take place, e.g., in a cell or a cell-free system. RNA may be synthesized *in vitro,* for example, by solid phase synthesis.

***Production of protein components***

Protein components of an anellovector, e.g., ORF1, can be produced in a variety of ways described herein.

***Baculovirus expression systems***

A viral expression system, *e.g.*, a baculovirus expression system, may be used to express proteins (e.g., for production of anellovectors), e.g., as described herein. Baculoviruses are rod-shaped viruses with a circular, supercoiled double-stranded DNA genome. Genera of baculoviruses include: Alphabaculovirus (nucleopolyhedroviruses (NPVs) isolated from *Lepidoptera*), Betabaculoviruses (granuloviruses (GV) isolated from *Lepidoptera*), Gammabaculoviruses (NPVs isolated from *Hymenoptera*) and Deltabaculoviruses (NPVs isolated from *Diptera*). While GVs typically contain only one nucleocapsid per envelope, NPVs typically contain either single (SNPV) or multiple (MNPV)

nucleocapsids per envelope. The enveloped virions are further occluded in granulin matrix in GVs and polyhedrin in NPVs. Baculoviruses typically have both lytic and occluded life cycles. In some embodiments, the lytic and occluded life cycles manifest independently throughout the three phases of virus replication: early, late, and very late phase. In some embodiments, during the early phase, viral DNA replication takes place following viral entry into the host cell, early viral gene expression and shut-off of the host gene expression machinery. In some embodiments, in the late phase late genes that code for viral DNA replication are expressed, viral particles are assembled, and extracellular virus (EV) is produced by the host cell. In some embodiments, in the very late phase the polyhedrin and p10 genes are expressed, occluded viruses (OV) are produced by the host cell, and the host cell is lysed. Since baculoviruses infect insect species, they can be used as biological agents to produce exogenous proteins in baculoviruses-permissive insect cells or larvae. Different isolates of baculovirus, such as *Autographa californica* multiple nuclear polyhedrosis virus (AcMNPV) and *Bombyx mori* (silkworm) nuclear polyhedrosis virus (BmNPV) may be used in exogenous protein expression. Various baculoviral expression systems are commercially available, e.g., from ThermoFisher.

In some embodiments, the proteins described herein (e.g., an Anellovirus ORF molecule, e.g., ORF1, ORF2, ORF2/2, ORF2/3, ORF1/1, or ORF1/2, or a functional fragment or splice variant thereof) may be expressed using a baculovirus expression vector (e.g., a bacmid) that comprises one or more components described herein. For example, a baculovirus expression vector may include one or more of (e.g., all of) a selectable marker (e.g., kanR), an origin of replication (e.g., one or both of a bacterial origin of replication and an insect cell origin of replication), a recombinase recognition site (e.g., an att site), and a promoter. In some embodiments, a baculovirus expression vector (e.g., a bacmid as described herein) can be produced by replacing the naturally occurring wild-type polyhedrin gene, which encodes for baculovirus occlusion bodies, with genes encoding the proteins described herein. In some embodiments, the genes encoding the proteins described herein are cloned into a baculovirus expression vector (e.g., a bacmid as described herein) containing a baculovirus promoter. In some embodiments, the baculovirual vector comprises one or more non-baculoviral promoters, e.g., a mammalian promoter or an Anellovirus promoter. In some embodiments, the genes encoding the proteins described herein are cloned into a donor vector (e.g., as described herein), which is then contacted with an empty baculovirus expression vector (e.g., an empty bacmid) such that the genes encoding the proteins described herein are transferred (e.g., by homologous recombination or transposase activity) from the donor vector into the baculovirus expression vector (e.g., bacmid). In some embodiments, the baculovirus promoter is flanked by baculovirus DNA from the nonessential polyhedrin gene locus. In some embodiments, a protein described herein is under the transcriptional control of the AcNPV polyhedrin promoter in the very late phase of viral replication. In some embodiments, a strong promoter suitable for use in baculoviral expression in

insect cells include, but are not limited to, baculovirus p10 promoters, polyhedrin (polh) promoters, p6.9 promoters and capsid protein promoters. Weak promoters suitable for use in baculoviral expression in insect cells include ie1, ie2, ie0, et1, 39K (aka pp31) and gp64 promoters of baculoviruses.

In some embodiments, a recombinant baculovirus is produced by homologous recombination between a baculoviral genome (*e.g.*, a wild-type or mutant baculoviral genome), and a transfer vector. In some embodiments, one or more genes encoding a protein described herein are cloned into the transfer vector. In some embodiments, the transfer vector further contains a baculovirus promoter flanked by DNA from a nonessential gene locus, e.g., polyhedrin gene. In some embodiments, one or more genes encoding a protein described herein are inserted into the baculoviral genome by homologous recombination between the baculoviral genome and the transfer vector. In some embodiments, the baculoviral genome is linearized at one or more unique sites. In some embodiments, the linearized sites are located near the target site for insertion of genes encoding the proteins described herein into the baculoviral genome. In some embodiments, a linearized baculoviral genome missing a fragment of the baculoviral genome downstream from a gene, e.g., polyhedrin gene, can be used for homologous recombination. In some embodiments, the baculoviral genome and transfer vector are co-transfected into insect cells. In some embodiments, the method of producing the recombinant baculovirus comprises the steps of preparing the baculoviral genome for performing homologous recombination with a transfer vector containing the genes encoding one or more protein described herein and co-transfecting the transfer vector and the baculoviral genome DNA into insect cells. In some embodiments, the baculoviral genome comprises a region homologous to a region of the transfer vector. These homologous regions may enhance the probability of recombination between the baculoviral genome and the transfer vector. In some embodiments, the homology region in the transfer vector is located upstream or downstream of the promoter. In some embodiments, to induce homologous recombination, the baculoviral genome, and transfer vector are mixed at a weight ratio of about 1:1 to 10:1.

In some embodiments, a recombinant baculovirus is generated by a method comprising site-specific transposition with Tn7, e.g., whereby the genes encoding the proteins described herein are inserted into bacmid DNA, e.g., propagated in bacteria, *e.g.*, *E. coli* (e.g., DH 10Bac cells). In some embodiments, the genes encoding the proteins described herein are cloned into a pFASTBAC® vector and transformed into competent cells, *e.g.*, DH10BAC® competent cells, containing the bacmid DNA with a mini-*att*Tn7 target site. In some embodiments, the baculovirus expression vector, *e.g.*, pFASTBAC® vector, may have a promoter, *e.g.*, a dual promoter (*e.g.*, polyhedrin promoter, p10 promoter). Commercially available pFASTBAC® donor plasmids include: pFASTBAC 1, pFASTBAC HT, and pFASTBAC DUAL. In some embodiments, recombinant bacmid DNA containing-colonies are identified and bacmid DNA is isolated to transfect insect cells.

In some embodiments, a baculoviral vector is introduced into an insect cell together with a helper nucleic acid. The introduction may be concurrent or sequential. In some embodiments, the helper nucleic acid provides one or more baculoviral proteins, e.g., to promote packaging of the baculoviral vector. In some embodiments, recombinant baculovirus produced in insect cells (e.g., by homologous recombination) is expanded and used to infect insect cells (e.g., in the mid-logarithmic growth phase) for recombinant protein expression. In some embodiments, recombinant bacmid DNA produced by site-specific transposition in bacteria, *e.g.*, *E. coli*, is used to transfect insect cells with a transfection agent, e.g., Cellfectin® II. Additional information on baculovirus expression systems is discussed in US patent applications Nos. 14/447,341, 14/277,892, and 12/278,916, which are hereby incorporated by reference.

***Insect cell systems***

The proteins described herein may be expressed in insect cells infected or transfected with recombinant baculovirus or bacmid DNA, e.g., as described above. In some embodiments, insect cells include: the Sf9 and Sf21 cells derived from *Spodoptera frugiperda* and the Tn-368 and High Five™ BTI-TN-5B1-4 cells (also referred to as Hi5 cells) derived from *Trichoplusia ni*. In some embodiments, insect cell lines Sf21 and Sf9, derived from the ovaries of the pupal fall army worm *Spodoptera frugiperda*, can be used for the expression of recombinant proteins using the baculovirus expression system. In some embodiments, Sf21 and Sf9 insect cells may be cultured in commercially available serum-supplemented or serum-free media. Suitable media for culturing insect cells include: Grace's Supplemented (TNM-FH), IPL-41, TC-100, Schneider's Drosophila, SF-900 II SFM, and EXPRESS-FIVE™ SFM. In some embodiments, some serum-free media formulations utilize a phosphate buffer system to maintain a culture pH in the range of 6.0-6.4 (Licari et al. Insect cell hosts for baculovirus expression vectors contain endogenous exoglycosidase activity. Biotechnology Progress 9: 146-152 (1993) and Drugmand et al. Insect cells as factories for biomanufacturing. Biotechnology Advances 30:1140-1157 (2012)) for both cultivation and recombinant protein production. In some embodiments, a pH of 6.0-6.8 for cultivating various insect cell lines may be used. In some embodiments, insect cells are cultivated in suspension or as a monolayer at a temperature between 25° to 30°C with aeration. Additional information on insect cells is discussed, for example, in US Patent Application Nos. 14/564,512 and 14/775,154, each of which is hereby incorporated by reference.

***Mammalian cell systems***

In some embodiments, the proteins described herein may be expressed *in vitro* in animal cell lines infected or transfected with a vector encoding the protein, e.g., as described herein. Animal cell lines envisaged in the context of the present disclosure include porcine cell lines, e..g, immortalised porcine

cell lines such as, but not limited to the porcine kidney epithelial cell lines PK-15 and SK, the monomyeloid cell line 3D4/31 and the testicular cell line ST. Also, other mammalian cells lines are included, such as CHO cells (Chinese hamster ovaries), MARC-145, MDBK, RK-13, EEL. Additionally or alternatively, particular embodiments of the methods of the invention make use of an animal cell line which is an epithelial cell line, i.e. a cell line of cells of epithelial lineage. Cell lines suitable for expressing the proteins described herein include, but are not limited to cell lines of human or primate origin, such as human or primate kidney carcinoma cell lines.

***Effectors***

The compositions and methods described herein can be used to produce a genetic element of an anellovector comprising a sequence encoding an effector (e.g., an exogenous effector or an endogenous effector), e.g., as described herein. In some embodimenst, the genetic element is the effector, e.g., the genetic element is a functional RNA. The effector may be, in some instances, an endogenous effector or an exogenous effector. In some embodiments, the effector is a therapeutic effector. In some embodiments, the effector comprises a polypeptide (e.g., a therapeutic polypeptide or peptide, e.g., as described herein). In some embodiments, the effector comprises a non-coding RNA (e.g., an miRNA, siRNA, shRNA, mRNA, lncRNA, RNA, DNA, antisense RNA, or gRNA). In some embodiments, the effector comprises a regulatory nucleic acid, e.g., as described herein.

***In vitro assembly methods***

An anellovector may be produced, e.g., by *in vitro* assembly. In some embodiments, the genetic element is contacted to ORF1 *in vitro* under conditions that allow for assembly.

In some embodiments, baculovirus constructs are used to produce Anellovirus proteins. These proteins may then be used, e.g., for *in vitro* assembly to encapsidate a genetic element, e.g., a genetic element comprising RNA. In some embodiments, a polynucleotide encoding one or more Anellovirus protein is fused to a promoter for expression in a host cell, *e.g.*, an insect or animal cell. In some embodiments, the polynucleotide is cloned into a baculovirus expression system. In some embodiments, a host cell, *e.g.*, an insect cell is infected with the baculovirus expression system and incubated for a period of time. In some embodiments, an infected cell is incubated for about 1, 2, 3, 4, 5, 10, 15, or 20 days. In some embodiments, an infected cell is lysed to recover the Anellovirus protein.

In some embodiments, an isolated Anellovirus protein is purified. In some embodiments, an Anellovirus protein is purified using purification techniques including but not limited to chelating purification, heparin purification, gradient sedimentation purification and/or SEC purification. In some embodiments, a purified Anellovirus protein is mixed with a genetic element to encapsidate the genetic

element, e.g., a genetic element comprising RNA. In some embodiments, a genetic element is encapisdated using an ORF1 protein, ORF2 protein, or modified version thereof. In some embodiments two nucleic acids are encapsidated. For instance, the first nucleic acid may be an mRNA e.g., chemically modified mRNA, and the second nucleic acid may be DNA.

In some embodiments, DNA encoding Anellovirus (AV) ORF1 (*e.g.*, wildtype ORF1 protein, ORF1 proteins harboring mutations, *e.g.*, to improve assembly efficiency, yield or stability, chimeric ORF1 protein, or fragments thereof) are expressed in insect cell lines (*e.g.*, Sf9 and/or HighFive), animal cell lines (*e.g.*, chicken cell lines (MDCC)), bacterial cells (*e.g.*, E. coli) and/or mammalian cell lines (*e.g.*, 293expi and/or MOLT4). In some embodiments, DNA encoding AV ORF1 may be untagged. In some embodiments, DNA encoding AV ORF1 may contain tags fused N-terminally and/or C-terminally. In some embodiments, DNA encoding AV ORF1 may harbor mutations, insertions or deletions within the ORF1 protein to introduce a tag, *e.g.*, to aid in purification and/or identity determination, *e.g.*, through immunostaining assays (including but not limited to ELISA or Western Blot). In some embodiments, DNA encoding AV ORF1 may be expressed alone or in combination with any number of helper proteins. In some embodiments, DNA encoding AV ORF1 is expressed in combination with AV ORF2 and/or ORF3 proteins.

In some embodiments, ORF1 proteins harboring mutations to improve assembly efficiency may include, but are not limited to, ORF1 proteins that harbor mutations introduced into the N-terminal Arginine Arm (ARG arm) to alter the pI of the ARG arm permitting pH sensitive nucleic acid binding to trigger particle assembly (SEQ ID 3-5). In some embodiments, ORF1 proteins harboring mutations that improve stability may include mutations to an interprotomer contacting beta strands F and G of the canonical jellyroll beta-barrel to alter hydrophobic state of the protomer surface and improve thermodynamic favorability of capsid formation.

In some embodiments, chimeric ORF1 proteins may include, but are not limited to, ORF1 proteins which have a portion or portions of their sequence replaced with comparable portions from another capsid protein, *e.g.*, Beak and Feather Disease Virus (BFDV) capsid protein, or Hepatitis E capsid protein, *e.g.,* ARG arm or F and G beta strands of Ring 9 ORF1 replaced with the comparable components from BFDV capsid protein. In some embodiments, chimeric ORF1 proteins may also include ORF1 proteins which have a portion or portions of their sequence replaced with comparable portions of another AV ORF1 protein (*e.g.*, jellyroll fragments or the C-terminal portion of Ring 2 ORF1 replaced with comparable portions of Ring 9 ORF1).

In some embodiments, the present disclosure describes a method of making an anellovector, the method comprising: (a) providing a mixture comprising: (i) a genetic element comprising RNA, and (ii) an ORF1 molecule; and (b) incubating the mixture under conditions suitable for enclosing the genetic

element within a proteinaceous exterior comprising the ORF1 molecule, thereby making an anellovector; optionally wherein the mixture is not comprised in a cell. In some embodiments, the method further comprises, prior to the providing of (a), expressing the ORF1 molecule, e.g., in a host cell (e.g., an insect cell or a mammalian cell). In some embodiments, the expressing comprises incubating a host cell (e.g., an insect cell or a mammalian cell) comprising a nucleic acid molecule (e.g., a baculovirus expression vector) encoding the ORF1 molecule under conditions suitable for producing the ORF1 molecule. In some embodiments, the method further comprises, prior to the providing of (a), purifying the ORF1 molecule expressed by the host cell. In some embodiments, the method is performed in a cell-free system. In some embodiments, the present disclosure describes a method of manufacturing an anellovector composition, comprising: (a) providing a plurality of anellovectors or compositions according to any of the preceding embodiments; (b) optionally evaluating the plurality for one or more of: a contaminant described herein, an optical density measurement (e.g., OD 260), particle number (e.g., by HPLC), infectivity (e.g., particle:infectious unit ratio, e.g., as determined by fluorescence and/or ELISA); and (c) formulating the plurality of anellovectors, e.g., as a pharmaceutical composition suitable for administration to a subject, e.g., if one or more of the parameters of (b) meet a specified threshold.

***Enrichment and purification***

Harvested anellovectors can be purified and/or enriched, e.g., to produce an anellovector preparation. In some embodiments, the harvested anellovectors are isolated from other constituents or contaminants present in the harvest solution, e.g., using methods known in the art for purifying viral particles (e.g., purification by sedimentation, chromatography, and/or ultrafiltration). In some embodiments, the harvested anellovectors are purified by affinity purification (e.g., heparin affinity purification). In some embodiments, the harvested anellovectors are purified by size exclusion chromatography (e.g., using a Tris buffer mobile phase). In some embodiments, the harvested anellovectors are purified by anion exchange chromatography (e.g., Mustang Q membrane chromatography). In some embodiments, the harvested anellovectors are purified by mixed mode chromatography (e.g., using a mixed mode resin, e.g., a Cato700 resin). In some embodiments, the purification steps comprise removing one or more of serum, host cell DNA, host cell proteins, particles lacking the genetic element, and/or phenol red from the preparation. In some embodiments, the harvested anellovectors are enriched relative to other constituents or contaminants present in the harvest solution, e.g., using methods known in the art for enriching viral particles.

In some embodiments, the resultant preparation or a pharmaceutical composition comprising the preparation will be stable over an acceptable period of time and temperature, and/or be compatible with

the desired route of administration and/or any devices this route of administration will require, e.g., needles or syringes.

## II. Anellovectors

In some aspects, the the disclosure provides compositions and methods of using and making an anellovector, anellovector preparations, and therapeutic compositions. In some embodiments, the anellovector comprises one or more nucleic acids or polypeptides comprising a sequence, structure, and/or function that is based on an *Anellovirus* (e.g., an *Anellovirus* as described herein), or fragments or portions thereof, or other substantially non-pathogenic virus, e.g., a symbiotic virus, commensal virus, native virus. In some embodiments, an *Anellovirus*-based anellovector comprises at least one element exogenous to that *Anellovirus*, e.g., an exogenous effector or a nucleic acid sequence encoding an exogenous effector disposed within a genetic element of the anellovector. In some embodiments, an *Anellovirus*-based anellovector comprises at least one element heterologous to another element from that *Anellovirus*, e.g., an effector-encoding nucleic acid sequence that is heterologous to another linked nucleic acid sequence, such as a promoter element. In some embodiments, an anellovector comprises a genetic element (e.g., circular DNA, e.g., single stranded DNA), which comprise at least one element that is heterologous relative to the remainder of the genetic element and/or the proteinaceous exterior (e.g., an exogenous element encoding an effector, e.g., as described herein). An anellovector may be a delivery vehicle (e.g., a substantially non-pathogenic delivery vehicle) for a payload into a host, e.g., a human. In some embodiments, the anellovector is capable of replicating in a eukaryotic cell, e.g., a mammalian cell, e.g., a human cell. In some embodiments, the anellovector is substantially non-pathogenic and/or substantially non-integrating in the mammalian (e.g., human) cell. In some embodiments, the anellovector is substantially non-immunogenic in a mammal, e.g., a human. In some embodiments, the anellovector is replication-deficient. In some embodiments, the anellovector is replication-competent.

In some embodiments the anellovector comprises a curon, or a component thereof (e.g., a genetic element, e.g., comprising a sequence encoding an effector, and/or a proteinaceous exterior), e.g., as described in PCT Application No. PCT/US2018/037379, which is incorporated herein by reference in its entirety. In some embodiments the anellovector comprises an anellovector, or a component thereof (e.g., a genetic element, e.g., comprising a sequence encoding an effector, and/or a proteinaceous exterior), e.g., as described in PCT Application No. PCT/US19/65995, which is incorporated herein by reference in its entirety.

In an aspect, the invention includes an anellovector comprising (i) a genetic element comprising a promoter element, a sequence encoding an effector, (e.g., an endogenous effector or an exogenous effector, e.g., a payload), and a protein binding sequence (e.g., an exterior protein binding sequence, e.g.,

a packaging signal), wherein the genetic element is a single-stranded DNA, and has one or both of the following properties: is circular and/or integrates into the genome of a eukaryotic cell at a frequency of less than about 0.001%, 0.005%, 0.01%, 0.05%, 0.1%, 0.5%, 1%, 1.5%, or 2% of the genetic element that enters the cell; and (ii) a proteinaceous exterior; wherein the genetic element is enclosed within the proteinaceous exterior; and wherein the anellovector is capable of delivering the genetic element into a eukaryotic cell.

In some embodiments of the anellovector described herein, the genetic element integrates at a frequency of less than about 0.001%, 0.005%, 0.01%, 0.05%, 0.1%, 0.5%, 1%, 1.5%, or 2% of the genetic element that enters a cell. In some embodiments, less than about 0.01%, 0.05%, 0.1%, 0.5%, 1%, 2%, 3%, 4%, or 5% of the genetic elements from a plurality of the anellovectors administered to a subject will integrate into the genome of one or more host cells in the subject. In some embodiments, the genetic elements of a population of anellovectors, e.g., as described herein, integrate into the genome of a host cell at a frequency less than that of a comparable population of AAV viruses, e.g., at about a 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, or more lower frequency than the comparable population of AAV viruses.

In an aspect, the invention includes an anellovector comprising: (i) a genetic element comprising a promoter element and a sequence encoding an effector (e.g., an endogenous effector or an exogenous effector, e.g., a payload), and a protein binding sequence (e.g., an exterior protein binding sequence), wherein the genetic element has at least 75% (e.g., at least 75, 76, 77, 78, 79, 80, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100%) sequence identity to a wild-type Anellovirus sequence (e.g., a wild-type Torque Teno virus (TTV), Torque Teno mini virus (TTMV), or TTMDV sequence, e.g., a wild-type Anellovirus sequence as described herein); and (ii) a proteinaceous exterior; wherein the genetic element is enclosed within the proteinaceous exterior; and wherein the anellovector is capable of delivering the genetic element into a eukaryotic cell.

In one aspect, the invention includes an anellovector comprising:

a) a genetic element comprising (i) a sequence encoding an exterior protein (e.g., a non-pathogenic exterior protein), (ii) an exterior protein binding sequence that binds the genetic element to the non-pathogenic exterior protein, and (iii) a sequence encoding an effector (e.g., an endogenous or exogenous effector); and

b) a proteinaceous exterior that is associated with, e.g., envelops or encloses, the genetic element.

In some embodiments, the anellovector includes sequences or expression products from (or having >70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, 100% homology to) a non-enveloped, circular, single-stranded DNA virus. Animal circular single-stranded DNA viruses generally refer to a subgroup of single strand DNA (ssDNA) viruses, which infect eukaryotic non-plant hosts, and have a

circular genome. Thus, animal circular ssDNA viruses are distinguishable from ssDNA viruses that infect prokaryotes (i.e. Microviridae and Inoviridae) and from ssDNA viruses that infect plants (i.e. Geminiviridae and Nanoviridae). They are also distinguishable from linear ssDNA viruses that infect non-plant eukaryotes (i.e. Parvoviridiae).

In some embodiments, the anellovector modulates a host cellular function, e.g., transiently or long term. In certain embodiments, the cellular function is stably altered, such as a modulation that persists for at least about 1 hr to about 30 days, or at least about 2 hrs, 6 hrs, 12 hrs, 18 hrs, 24 hrs, 2 days, 3, days, 4 days, 5 days, 6 days, 7 days, 8 days, 9 days, 10 days, 11 days, 12 days, 13 days, 14 days, 15 days, 16 days, 17 days, 18 days, 19 days, 20 days, 21 days, 22 days, 23 days, 24 days, 25 days, 26 days, 27 days, 28 days, 29 days, 30 days, 60 days, or longer or any time therebetween. In certain embodiments, the cellular function is transiently altered, e.g., such as a modulation that persists for no more than about 30 mins to about 7 days, or no more than about 1 hr, 2 hrs, 3 hrs, 4 hrs, 5 hrs, 6 hrs, 7 hrs, 8 hrs, 9 hrs, 10 hrs, 11 hrs, 12 hrs, 13 hrs, 14 hrs, 15 hrs, 16 hrs, 17 hrs, 18 hrs, 19 hrs, 20 hrs, 21 hrs, 22 hrs, 24 hrs, 36 hrs, 48 hrs, 60 hrs, 72 hrs, 4 days, 5 days, 6 days, 7 days, or any time therebetween.

In some embodiments, the genetic element comprises a promoter element. In embodiments, the promoter element is selected from an RNA polymerase II-dependent promoter, an RNA polymerase III-dependent promoter, a PGK promoter, a CMV promoter, an EF-1α promoter, an SV40 promoter, a CAGG promoter, or a UBC promoter, TTV viral promoters, Tissue specific, U6 (pollIII), minimal CMV promoter with upstream DNA binding sites for activator proteins (TetR-VP16, Gal4-VP16, dCas9-VP16, etc). In embodiments, the promoter element comprises a TATA box. In embodiments, the promoter element is endogenous to a wild-type *Anellovirus*, e.g., as described herein.

In some embodiments, the genetic element comprises one or more of the following characteristics: single-stranded, circular, negative strand, and/or DNA. In embodiments, the genetic element comprises an episome. In some embodiments, the portions of the genetic element excluding the effector have a combined size of about 2.5-5 kb (e.g., about 2.8-4kb, about 2.8-3.2kb, about 3.6-3.9kb, or about 2.8-2.9kb), less than about 5kb (e.g., less than about 2.9kb, 3.2 kb, 3.6kb, 3.9kb, or 4kb), or at least 100 nucleotides (e.g., at least 1kb).

In some embodiments, an anellovector, or the genetic element comprised in the anellovector, is introduced into a cell (e.g., a human cell). In some embodiments, the effector (e.g., an RNA, e.g., an miRNA), e.g., encoded by the genetic element of an anellovector, is expressed in a cell (e.g., a human cell), e.g., once the anellovector or the genetic element has been introduced into the cell. In embodiments, introduction of the anellovector, or genetic element comprised therein, into a cell modulates (e.g., increases or decreases) the level of a target molecule (e.g., a target nucleic acid, e.g.,

RNA, or a target polypeptide) in the cell, e.g., by altering the expression level of the target molecule by the cell. In embodiments, introduction of the anellovector, or genetic element comprised therein, decreases level of interferon produced by the cell. In embodiments, introduction of the anellovector, or genetic element comprised therein, into a cell modulates (e.g., increases or decreases) a function of the cell. In embodiments, introduction of the anellovector, or genetic element comprised therein, into a cell modulates (e.g., increases or decreases) the viability of the cell. In embodiments, introduction of the anellovector, or genetic element comprised therein, into a cell decreases viability of a cell (e.g., a cancer cell).

In some embodiments, an anellovector (e.g., a synthetic anellovector) described herein induces an antibody prevalence of less than 70% (e.g., less than about 60%, 50%, 40%, 30%, 20%, or 10% antibody prevalence). In embodiments, antibody prevalence is determined according to methods known in the art. In embodiments, antibody prevalence is determined by detecting antibodies against an *Anellovirus* (e.g., as described herein), or an anellovector based thereon, in a biological sample, e.g., according to the anti-TTV antibody detection method described in Tsuda et al. (1999; *J. Virol. Methods* 77: 199-206; incorporated herein by reference) and/or the method for determining anti-TTV IgG seroprevalence described in Kakkola et al. (2008; *Virology* 382: 182-189; incorporated herein by reference). Antibodies against an *Anellovirus* or an anellovector based thereon can also be detected by methods in the art for detecting anti-viral antibodies, e.g., methods of detecting anti-AAV antibodies, e.g., as described in Calcedo et al. (2013; *Front. Immunol.* 4(341): 1-7; incorporated herein by reference).

In some embodiments, a replication deficient, replication defective, or replication incompetent genetic element does not encode all of the necessary machinery or components required for replication of the genetic element. In some embodiments, a replication defective genetic element does not encode a replication factor. In some embodiments, a replication defective genetic element does not encode one or more ORFs (e.g., ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, and/or ORF2t/3, e.g., as described herein). In some embodiments, the machinery or components not encoded by the genetic element may be provided in *trans* (e.g., encoded in a nucleic acid comprised by the host cell, e.g., integrated into the genome of the host cell), e.g., such that the genetic element can undergo replication in the presence of the machinery or components provided in *trans*.

In some embodiments, a packaging deficient, packaging defective, or packaging incompetent genetic element cannot be packaged into a proteinaceous exterior (e.g., wherein the proteinaceous exterior comprises a capsid or a portion thereof, e.g., comprising a polypeptide encoded by an ORF1 nucleic acid, e.g., as described herein). In some embodiments, a packaging deficient genetic element is packaged into a proteinaceous exterior at an efficiency less than 10% (e.g., less than 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.1%, 0.01%, or 0.001%) compared to a wild-type *Anellovirus* (e.g., as described

herein). In some embodiments, the packaging defective genetic element cannot be packaged into a proteinaceous exterior even in the presence of factors (e.g., ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, or ORF2t/3) that would permit packaging of the genetic element of a wild-type *Anellovirus* (e.g., as described herein). In some embodiments, a packaging deficient genetic element is packaged into a proteinaceous exterior at an efficiency less than 10% (e.g., less than 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.1%, 0.01%, or 0.001%) compared to a wild-type *Anellovirus* (e.g., as described herein), even in the presence of factors (e.g., ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, or ORF2t/3) that would permit packaging of the genetic element of a wild-type *Anellovirus* (e.g., as described herein).

In some embodiments, a packaging competent genetic element can be packaged into a proteinaceous exterior (e.g., wherein the proteinaceous exterior comprises a capsid or a portion thereof, e.g., comprising a polypeptide encoded by an ORF1 nucleic acid, e.g., as described herein). In some embodiments, a packaging competent genetic element is packaged into a proteinaceous exterior at an efficiency of at least 20% (e.g., at least 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 100%, or higher) compared to a wild-type *Anellovirus* (e.g., as described herein). In some embodiments, the packaging competent genetic element can be packaged into a proteinaceous exterior in the presence of factors (e.g., ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, or ORF2t/3) that would permit packaging of the genetic element of a wild-type *Anellovirus* (e.g., as described herein). In some embodiments, a packaging competent genetic element is packaged into a proteinaceous exterior at an efficiency of at least 20% (e.g., at least 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 100%, or higher) compared to a wild-type *Anellovirus* (e.g., as described herein) in the presence of factors (e.g., ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, or ORF2t/3) that would permit packaging of the genetic element of a wild-type *Anellovirus* (e.g., as described herein).

***Anelloviruses***

In some embodiments, an anellovector, e.g., as described herein, comprises sequences or expression products derived from an *Anellovirus*. In some embodiments, an anellovector includes one or more sequences or expression products that are exogenous relative to the *Anellovirus*. In some embodiments, an anellovector includes one or more sequences or expression products that are endogenous relative to the *Anellovirus*. In some embodiments, an anellovector includes one or more sequences or expression products that are heterologous relative to one or more other sequences or expression products in the anellovector. *Anelloviruses* generally have single-stranded circular DNA genomes with negative polarity.

In some embodiments, the genetic element comprises a nucleotide sequence encoding an amino acid sequence or a functional fragment thereof or a sequence having at least about 60%, 70% 80%, 85%,

90% 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of the amino acid sequences described herein, e.g., an *Anellovirus* amino acid sequence.

In some embodiments, an anellovector as described herein comprises one or more nucleic acid molecules (e.g., a genetic element as described herein) comprising a sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an *Anellovirus* sequence, e.g., as described herein, or a fragment thereof.

In some embodiments, an anellovector as described herein comprises one or more nucleic acid molecules (e.g., a genetic element as described herein) comprising a sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to one or more of a TATA box, cap site, initiator element, transcriptional start site, 5' UTR conserved domain, ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3, three open-reading frame region, poly(A) signal, GC-rich region, or any combination thereof, of an *Anellovirus,* e.g., as described herein. In some embodiments, the nucleic acid molecule comprises a sequence encoding a capsid protein, e.g., an ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3 sequence of any of the *Anelloviruses* described herein. In embodiments, the nucleic acid molecule comprises a sequence encoding a capsid protein comprising an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an *Anellovirus* ORF1 protein (or a splice variant or functional fragment thereof) or a polypeptide encoded by an *Anellovirus* ORF1 nucleic acid.

In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 nucleic acid sequence of Table A1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1/1 nucleotide sequence of Table A1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1/2 nucleotide sequence of Table A1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 nucleotide sequence of Table A1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2/2 nucleotide sequence of Table A1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2/3 nucleotide sequence of Table A1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%,

98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2t/3 nucleotide sequence of Table A1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* TATA box nucleotide sequence of Table A1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* initiator element nucleotide sequence of Table A1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* transcriptional start site nucleotide sequence of Table A1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* 5' UTR conserved domain nucleotide sequence of Table A1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* three open-reading frame region nucleotide sequence of Table A1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* poly(A) signal nucleotide sequence of Table A1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* GC-rich nucleotide sequence of Table A1.

In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 nucleic acid sequence of Table B1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1/1 nucleotide sequence of Table B1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1/2 nucleotide sequence of Table B1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 nucleotide sequence of Table B1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2/2 nucleotide sequence of Table B1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2/3 nucleotide sequence of Table B1. In embodiments, the nucleic acid molecule

comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* TATA box nucleotide sequence of Table B1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* initiator element nucleotide sequence of Table B1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* transcriptional start site nucleotide sequence of Table B1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* 5' UTR conserved domain nucleotide sequence of Table B1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* three open-reading frame region nucleotide sequence of Table B1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* poly(A) signal nucleotide sequence of Table B1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* GC-rich nucleotide sequence of Table B1.

In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 nucleic acid sequence of Table C1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1/1 nucleotide sequence of Table C1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1/2 nucleotide sequence of Table C1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 nucleotide sequence of Table C1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2/2 nucleotide sequence of Table C1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2/3 nucleotide sequence of Table C1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* TAIP nucleotide sequence of Table C1. In

embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* TATA box nucleotide sequence of Table C1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* initiator element nucleotide sequence of Table C1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* transcriptional start site nucleotide sequence of Table C1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* 5' UTR conserved domain nucleotide sequence of Table C1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* three open-reading frame region nucleotide sequence of Table C1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* poly(A) signal nucleotide sequence of Table C1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* GC-rich nucleotide sequence of Table C1.

In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 nucleic acid sequence of Table D1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1/1 nucleotide sequence of Table D1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1/2 nucleotide sequence of Table D1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 nucleotide sequence of Table D1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2/2 nucleotide sequence of Table D1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2/3 nucleotide sequence of Table D1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%,

98%, 99%, or 100% sequence identity to the *Anellovirus* TAIP nucleotide sequence of Table D1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* TATA box nucleotide sequence of Table D1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* initiator element nucleotide sequence of Table D1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* transcriptional start site nucleotide sequence of Table D1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* 5' UTR conserved domain nucleotide sequence of Table D1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* three open-reading frame region nucleotide sequence of Table D1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* poly(A) signal nucleotide sequence of Table D1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* GC-rich nucleotide sequence of Table D1.

In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1 nucleic acid sequence of Table E1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1/1 nucleotide sequence of Table E1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF1/2 nucleotide sequence of Table E1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2 nucleotide sequence of Table E1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2/2 nucleotide sequence of Table E1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* ORF2/3 nucleotide sequence of Table E1. In embodiments, the nucleic acid molecule

comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* TAIP nucleotide sequence of Table E1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* TATA box nucleotide sequence of Table E1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* initiator element nucleotide sequence of Table E1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* transcriptional start site nucleotide sequence of Table E1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* 5' UTR conserved domain nucleotide sequence of Table E1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* three open-reading frame region nucleotide sequence of Table E1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* poly(A) signal nucleotide sequence of Table E1. In embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the *Anellovirus* GC-rich nucleotide sequence of Table E1.

In some embodiments, the genetic element comprises a nucleotide sequence encoding an amino acid sequence or a functional fragment thereof or a sequence having at least about 60%, 70% 80%, 85%, 90% 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of the amino acid sequences described herein, e.g., an *Anellovirus* amino acid sequence.

In some embodiments, an anellovector as described herein comprises one or more nucleic acid molecules (e.g., a genetic element as described herein) comprising a sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an *Anellovirus* sequence, e.g., as described herein, or a fragment thereof. In embodiments, the anellovector comprises a nucleic acid sequence selected from a sequence as shown in any of Tables A1-M2, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. In embodiments, the anellovector comprises a polypeptide comprising a sequence as shown in any of Tables Tables A2-M2, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto.

In some embodiments, an anellovector as described herein comprises one or more nucleic acid molecules (e.g., a genetic element as described herein) comprising a sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to one or more of a TATA box, cap site, initiator element, transcriptional start site, 5' UTR conserved domain, ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3, three open-reading frame region, poly(A) signal, GC-rich region, or any combination thereof, of any of the *Anelloviruses* described herein (e.g., an *Anellovirus* sequence as annotated, or as encoded by a sequence listed, in any of Tables A-M). In some embodiments, the nucleic acid molecule comprises a sequence encoding a capsid protein, e.g., an ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3 sequence of any of the *Anelloviruses* described herein (e.g., an *Anellovirus* sequence as annotated, or as encoded by a sequence listed, in any of Tables A-M). In embodiments, the nucleic acid molecule comprises a sequence encoding a capsid protein comprising an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an *Anellovirus* ORF1 or ORF2 protein (e.g., an ORF1 or ORF2 amino acid sequence as shown in any of Tables A2-M2, or an ORF1 or ORF2 amino acid sequence encoded by a nucleic acid sequence as shown in any of Tables A1-M1). In embodiments, the nucleic acid molecule comprises a sequence encoding a capsid protein comprising an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an *Anellovirus* ORF1 protein (e.g., an ORF1 amino acid sequence as shown in any of Tables A2-M2, or an ORF1 amino acid sequence encoded by a nucleic acid sequence as shown in any of Tables A1-M1).

In some embodiments, an anellovector as described herein is a chimeric anellovector. In some embodiments, a chimeric anellovector further comprises one or more elements, polypeptides, or nucleic acids from a virus other than an *Anellovirus*.

In embodiments, the chimeric anellovector comprises a plurality of polypeptides (e.g., Anellovirus ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, and/or ORF2t/3) comprising sequences from a plurality of different Anelloviruses (e.g., as described herein). For example, a chimeric anellovector may comprise an ORF1 molecule from one Anellovirus (e.g., a Ring1 ORF1 molecule, or an ORF1 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto) and an ORF2 molecule from a different Anellovirus (e.g., a Ring2 ORF2 molecule, or an ORF2 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto). In another example, a chimeric anellovector may comprise a first ORF1 molecule from one Anellovirus (e.g., a Ring1 ORF1 molecule, or an ORF1 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto) and a second ORF1 molecule from a different Anellovirus (e.g., a Ring2 ORF1 molecule, or an ORF1 molecule

having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto).

In some embodiments, the anellovector comprises a chimeric polypeptide (e.g., Anellovirus ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, and/or ORF2t/3), e.g., comprising at least one portion from an Anellovirus (e.g., as described herein) and at least one portion from a different virus (e.g., as described herein).

In some embodiments, the anellovector comprises a chimeric polypeptide (e.g., Anellovirus ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, and/or ORF2t/3), e.g., comprising at least one portion from one Anellovirus (e.g., as described herein) and at least one portion from a different Anellovirus (e.g., as described herein). In embodiments, the anellovector comprises a chimeric ORF1 molecule comprising at least one portion of an ORF1 molecule from one Anellovirus (e.g., as described herein), or an ORF1 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto, and at least one portion of an ORF1 molecule from a different Anellovirus (e.g., as described herein), or an ORF1 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto. In embodiments, the chimeric ORF1 molecule comprises an ORF1 jelly-roll domain from one Anellovirus, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, and an ORF1 amino acid subsequence (e.g., as described herein) from a different Anellovirus, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In embodiments, the chimeric ORF1 molecule comprises an ORF1 arginine-rich region from one Anellovirus, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, and an ORF1 amino acid subsequence (e.g., as described herein) from a different Anellovirus, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In embodiments, the chimeric ORF1 molecule comprises an ORF1 hypervariable domain from one Anellovirus, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, and an ORF1 amino acid subsequence (e.g., as described herein) from a different Anellovirus, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In embodiments, the chimeric ORF1 molecule comprises an ORF1 N22 domain from one Anellovirus, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, and an ORF1 amino acid subsequence (e.g., as described herein) from a different Anellovirus, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In embodiments, the chimeric ORF1 molecule comprises an ORF1 C-terminal domain from one Anellovirus, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto, and an

ORF1 amino acid subsequence (e.g., as described herein) from a different Anellovirus, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

In embodiments, the anellovector comprises a chimeric ORF1/1 molecule comprising at least one portion of an ORF1/1 molecule from one Anellovirus (e.g., as described herein), or an ORF1/1 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto, and at least one portion of an ORF1/1 molecule from a different Anellovirus (e.g., as described herein), or an ORF1/1 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto. In embodiments, the anellovector comprises a chimeric ORF1/2 molecule comprising at least one portion of an ORF1/2 molecule from one Anellovirus (e.g., as described herein), or an ORF1/2 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto, and at least one portion of an ORF1/2 molecule from a different Anellovirus (e.g., as described herein), or an ORF1/2 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto. In embodiments, the anellovector comprises a chimeric ORF2 molecule comprising at least one portion of an ORF2 molecule from one Anellovirus (e.g., as described herein), or an ORF2 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto, and at least one portion of an ORF2 molecule from a different Anellovirus (e.g., as described herein), or an ORF2 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto. In embodiments, the anellovector comprises a chimeric ORF2/2 molecule comprising at least one portion of an ORF2/2 molecule from one Anellovirus (e.g., as described herein), or an ORF2/2 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto, and at least one portion of an ORF2/2 molecule from a different Anellovirus (e.g., as described herein), or an ORF2/2 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto. In embodiments, the anellovector comprises a chimeric ORF2/3 molecule comprising at least one portion of an ORF2/3 molecule from one Anellovirus (e.g., as described herein), or an ORF2/3 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto, and at least one portion of an ORF2/3 molecule from a different Anellovirus (e.g., as described herein), or an ORF2/3 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto. In embodiments, the anellovector comprises a chimeric ORF2T/3 molecule comprising at least one portion of an ORF2T/3 molecule from one Anellovirus (e.g., as described herein), or an ORF2T/3 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto, and at least one portion of an ORF2T/3 molecule from a different Anellovirus (e.g., as described herein), or an ORF2T/3 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity thereto.

Additional exemplary Anellovirus genomes, for which sequences or subsequences comprised therein can be utilized in the compositions and methods described herein are described, for example, in PCT Application Nos. PCT/US2018/037379 and PCT/US19/65995 (incorporated herein by reference in their entirety). In some embodiments, the exemplary Anellovirus sequences comprise a nucleic acid sequence as listed in any of Tables A1, A3, A5, A7, A9, A11, B1-B5, 1, 3, 5, 7, 9, 11, 13, 15, or 17 of PCT/US19/65995, incorporated herein by reference. In some embodiments, the exemplary Anellovirus sequences comprise an amino acid sequence as listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, or 18 of PCT/US19/65995, incorporated herein by reference. In some embodiments, the exemplary Anellovirus sequences comprise an ORF1 molecule sequence, or a nucleic acid sequence encoding same, e.g., as listed in any of Tables 21, 23, 25, 27, 29, 31, 33, 35, D2, D4, D6, D8, D10, or 37A-37C of PCT/US19/65995, incorporated herein by reference.

**Table A1. Exemplary *Anellovirus* nucleic acid sequence (*Alphatorquevirus,* Clade 3)**

| | |
|---|---|
| Name | Ring1 |
| Genus/Clade | *Alphatorquevirus*, Clade 3 |
| Accession Number | AJ620231.1 |

**Full Sequence:** 3753 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
TGCTACGTCACTAACCCACGTGTCCTCTACAGGCCAATCGCAGTCTATGT
CGTGCACTTCCTGGGCATGGTCTACATAATTATATAAATGCTTGCACTTC
CGAATGGCTGAGTTTTTGCTGCCCGTCCGCGGAGAGGAGCCACGGCAGGG
GATCCGAACGTCCTGAGGGCGGGTGCCGGAGGTGAGTTTACACACCGAAG
TCAAGGGGCAATTCGGGCTCAGGACTGGCCGGGCTTTGGGCAAGGCTCTT
AAAAATGCACTTTTCTCGAATAAGCAGAAAGAAAAGGAAAGTGCTACTGC
TTTGCGTGCCAGCAGCTAAGAAAAAACCAACTGCTATGAGCTTCTGGAAA
CCTCCGGTACACAATGTCACGGGGATCCAACGCATGTGGTATGAGTCCTT
TCACCGTGGCCACGCTTCTTTTTGTGGTTGTGGGAATCCTATACTTCACA
TTACTGCACTTGCTGAAACATATGGCCATCCAACAGGCCCGAGACCTTCT
GGGCCACCGGGAGTAGACCCCAACCCCCACATCCGTAGAGCCAGGCCTGC
CCCGGCCGCTCCGGAGCCCTCACAGGTTGATTCGAGACCAGCCCTGACAT
GGCATGGGGATGGTGGAAGCGACGGAGGCGCTGGTGGTTCCGGAAGCGGT
GGACCCGTGGCAGACTTCGCAGACGATGGCCTCGATCAGCTCGTCGCCGC
CCTAGACGACGAAGAGTAAGGAGGCGCAGACGGTGGAGGAGGGGGAGACG
AAAAACAAGGACTTACAGACGCAGGAGACGCTTTAGACGCAGGGGACGAA
```

AAGCAAAACTTATAATAAAACTGTGGCAACCTGCAGTAATTAAAAGATGC

AGAATAAAGGGATACATACCACTGATTATAAGTGGGAACGGTACCTTTGC

CACAAACTTTACCAGTCACATAAATGACAGAATAATGAAAGGCCCCTTCG

GGGGAGGACACAGCACTATGAGGTTCAGCCTCTACATTTTGTTTGAGGAG

CACCTCAGACACATGAACTTCTGGACCAGAAGCAACGATAACCTAGAGCT

AACCAGATACTTGGGGGCTTCAGTAAAAATATACAGGCACCCAGACCAAG

ACTTTATAGTAATATACAACAGAAGAACCCCTCTAGGAGGCAACATCTAC

ACAGCACCCTCTCTACACCCAGGCAATGCCATTTTAGCAAAACACAAAAT

ATTAGTACCAAGTTTACAGACAAGACCAAAGGGTAGAAAAGCAATTAGAC

TAAGAATAGCACCCCCCACACTCTTTACAGACAAGTGGTACTTTCAAAAG

GACATAGCCGACCTCACCCTTTTCAACATCATGGCAGTTGAGGCTGACTT

GCGGTTTCCGTTCTGCTCACCACAAACTGACAACACTTGCATCAGCTTCC

AGGTCCTTAGTTCCGTTTACAACAACTACCTCAGTATTAATACCTTTAAT

AATGACAACTCAGACTCAAAGTTAAAAGAATTTTTAAATAAAGCATTTCC

AACAACAGGCACAAAAGGAACAAGTTTAAATGCACTAAATACATTTAGAA

CAGAAGGATGCATAAGTCACCCACAACTAAAAAAACCAAACCCACAAATA

AACAAACCATTAGAGTCACAATACTTTGCACCTTTAGATGCCCTCTGGGG

AGACCCCATATACTATAATGATCTAAATGAAAACAAAAGTTTGAACGATA

TCATTGAGAAAATACTAATAAAAAACATGATTACATACCATGCAAAACTA

AGAGAATTTCCAAATTCATACCAAGGAAACAAGGCCTTTTGCCACCTAAC

AGGCATATACAGCCCACCATACCTAAACCAAGGCAGAATATCTCCAGAAA

TATTTGGACTGTACACAGAAATAATTTACAACCCTTACACAGACAAAGGA

ACTGGAAACAAAGTATGGATGGACCCACTAACTAAAGAGAACAACATATA

TAAAGAAGGACAGAGCAAATGCCTACTGACTGACATGCCCCTATGGACTT

TACTTTTTGGATATACAGACTGGTGTAAAAAGGACACTAATAACTGGGAC

TTACCACTAAACTACAGACTAGTACTAATATGCCCTTATACCTTTCCAAA

ATTGTACAATGAAAAAGTAAAAGACTATGGGTACATCCCGTACTCCTACA

AATTCGGAGCGGGTCAGATGCCAGACGGCAGCAACTACATACCCTTTCAG

TTTAGAGCAAAGTGGTACCCCACAGTACTACACCAGCAACAGGTAATGGA

GGACATAAGCAGGAGCGGGCCCTTTGCACCTAAGGTAGAAAAACCAAGCA

CTCAGCTGGTAATGAAGTACTGTTTTAACTTTAACTGGGGCGGTAACCCT

ATCATTGAACAGATTGTTAAAGACCCCAGCTTCCAGCCCACCTATGAAAT

ACCCGGTACCGGTAACATCCCTAGAAGAATACAAGTCATCGACCCGCGGG

TCCTGGGACCGCACTACTCGTTCCGGTCATGGGACATGCGCAGACACACA

TTTAGCAGAGCAAGTATTAAGAGAGTGTCAGAACAACAAGAAACTTCTGA

CCTTGTATTCTCAGGCCCAAAAAAGCCTCGGGTCGACATCCCAAAACAAG

AAACCCAAGAAGAAAGCTCACATTCACTCCAAAGAGAATCGAGACCGTGG

GAGACCGAGGAAGAAAGCGAGACAGAAGCCCTCTCGCAAGAGAGCCAAGA

GGTCCCCTTCCAACAGCAGTTGCAGCAGCAGTACCAAGAGCAGCTCAAGC
TCAGACAGGGAATCAAAGTCCTCTTCGAGCAGCTCATAAGGACCCAACAA
GGGGTCCATGTAAACCCATGCCTACGGTAGGTCCCAGGCAGTGGCTGTTT
CCAGAGAGAAAGCCAGCCCCAGCTCCTAGCAGTGGAGACTGGGCCATGGA
GTTTCTCGCAGCAAAAATATTTGATAGGCCAGTTAGAAGCAACCTTAAAG
ATACCCCTTACTACCCATATGTTAAAAACCAATACAATGTCTACTTTGAC
CTTAAATTTGAATAAACAGCAGCTTCAAACTTGCAAGGCCGTGGGAGTTT
CACTGGTCGGTGTCTACCTCTAAAGGTCACTAAGCACTCCGAGCGTAAGC
GAGGAGTGCGACCCTCCCCCCTGGAACAACTTCTTCGGAGTCCGGCGCTA
CGCCTTCGGCTGCGCCGGACACCTCAGACCCCCCCTCCACCCGAAACGCT
TGCGCGTTTCGGACCTTCGGCGTCGGGGGGGTCGGGAGCTTTATTAAACG
GACTCCGAAGTGCTCTTGGACACTGAGGGGGTGAACAGCAACGAAAGTGA
GTGGGGCCAGACTTCGCCATAAGGCCTTTATCTTCTTGCCATTTGTCAGT
GTCCGGGGTCGCCATAGGCTTCGGGCTCGTTTTTAGGCCTTCCGGACTAC
AAAAATCGCCATTTTGGTGACGTCACGGCCGCCATCTTAAGTAGTTGAGG
CGGACGGTGGCGTGAGTTCAAAGGTCACCATCAGCCACACCTACTCAAAA
TGGTGGACAATTTCTTCCGGGTCAAAGGTTACAGCCGCCATGTTAAAACA
CGTGACGTATGACGTCACGGCCGCCATTTTGTGACACAAGATGGCCGACT
TCCTTCCTCTTTTTCAAAAAAAAGCGGAAGTGCCGCCGCGGCGGCGGGGG
GCGGCGCGCTGCGCGCGCCGCCCAGTAGGGGGAGCCATGCGCCCCCCCCC
GCGCATGCGCGGGGCCCCCCCCCGCGGGGGGCTCCGCCCCCCGGCCCCCC
CCG (SEQ ID NO: 16)

**Annotations:**

| *Putative Domain* | *Base range* |
|---|---|
| TATA Box | 83 – 88 |
| Cap Site | 104 – 111 |
| Transcriptional Start Site | 111 |
| 5’ UTR Conserved Domain | 170 – 240 |
| ORF2 | 336 – 719 |
| ORF2/2 | 336 – 715 ; 2363 – 2789 |
| ORF2/3 | 336 – 715 ; 2565 – 3015 |
| ORF2t/3 | 336 – 388 ; 2565 – 3015 |
| ORF1 | 599 – 2830 |
| ORF1/1 | 599 – 715 ; 2363 – 2830 |
| ORF1/2 | 599 – 715 ; 2565 – 2789 |

| | |
|---|---|
| Three open-reading frame region | 2551 – 2786 |
| Poly(A) Signal | 3011 – 3016 |
| GC-rich region | 3632 – 3753 |

**Table A2. Exemplary *Anellovirus* amino acid sequences (*Alphatorquevirus*, Clade 3)**

| Ring1 (*Alphatorquevirus* Clade 3) | |
|---|---|
| ORF2 | MSFWKPPVHNVTGIQRMWYESFHRGHASFCGCGNPILHITALAETYGHPTGPRPSG PPGVDPNPHIRRARPAPAAPEPSQVDSRPALTWHGDGGSDGGAGGSGSGGPVADFA DDGLDQLVAALDDEE (SEQ ID NO: 17) |
| ORF2/2 | MSFWKPPVHNVTGIQRMWYESFHRGHASFCGCGNPILHITALAETYGHPTGPRPSG PPGVDPNPHIRRARPAPAAPEPSQVDSRPALTWHGDGGSDGGAGGSGSGGPVADFA DDGLDQLVAALDDEELLKTPASSPPMKYPVPVTSLEEYKSSTRGSWDRTTRSGHGT CADTHLAEQVLRECQNNKKLLTLYSQAQKSLGSTSQNKKPKKKAHIHSKENRDRG RPRKKARQKPSRKRAKRSPSNSSCSSSTKSSSSSDRESKSSSSSS (SEQ ID NO: 18) |
| ORF2/3 | MSFWKPPVHNVTGIQRMWYESFHRGHASFCGCGNPILHITALAETYGHPTGPRPSG PPGVDPNPHIRRARPAPAAPEPSQVDSRPALTWHGDGGSDGGAGGSGSGGPVADFA DDGLDQLVAALDDEEPKKASGRHPKTRNPRRKLTFTPKRIETVGDRGRKRDRSPLA REPRGPLPTAVAAAVPRAAQAQTGNQSPLRAAHKDPTRGPCKPMPTVGPRQWLFP ERKPAPAPSSGDWAMEFLAAKIFDRPVRSNLKDTPYYPYVKNQYNVYFDLKFE (SEQ ID NO: 19) |
| ORF2t/3 | MSFWKPPVHNVTGIQRMWPKKASGRHPKTRNPRRKLTFTPKRIETVGDRGRKRDR SPLAREPRGPLPTAVAAAVPRAAQAQTGNQSPLRAAHKDPTRGPCKPMPTVGPRQ WLFPERKPAPAPSSGDWAMEFLAAKIFDRPVRSNLKDTPYYPYVKNQYNVYFDLK FE (SEQ ID NO: 20) |
| ORF1 | MAWGWWKRRRRWWFRKRWTRGRLRRRWPRSARRRPRRRRVRRRRRWRRGRRK TRTYRRRRRFRRRGRKAKLIIKLWQPAVIKRCRIKGYIPLIISGNGTFATNFTSHINDR IMKGPFGGGHSTMRFSLYILFEEHLRHMNFWTRSNDNLELTRYLGASVKIYRHPDQ DFIVIYNRRTPLGGNIYTAPSLHPGNAILAKHKILVPSLQTRPKGRKAIRLRIAPPTLFT DKWYFQKDIADLTLFNIMAVEADLRFPFCSPQTDNTCISFQVLSSVYNNYLSINTFN NDNSDSKLKEFLNKAFPTTGTKGTSLNALNTFRTEGCISHPQLKKPNPQINKPLESQ YFAPLDALWGDPIYYNDLNENKSLNDIIEKILIKNMITYHAKLREFPNSYQGNKAFC HLTGIYSPPYLNQGRISPEIFGLYTEIIYNPYTDKGTGNKVWMDPLTKENNIYKEGQS KCLLTDMPLWTLLFGYTDWCKKDTNNWDLPLNYRLVLICPYTFPKLYNEKVKDY |

| | |
|---|---|
| | GYIPYSYKFGAGQMPDGSNYIPFQFRAKWYPTVLHQQQVMEDISRSGPFAPKVEKP STQLVMKYCFNFNWGGNPIIEQIVKDPSFQPTYEIPGTGNIPRRIQVIDPRVLGPHYSF RSWDMRRHTFSRASIKRVSEQQETSDLVFSGPKKPRVDIPKQETQEESSHSLQRESR PWETEEESETEALSQESQEVPFQQQLQQQYQEQLKLRQGIKVLFEQLIRTQQGVHV NPCLR (SEQ ID NO: 21) |
| ORF1/1 | MAWGWWKRRRRWWFRKRWTRGRLRRRWPRSARRRPRRRRIVKDPSFQPTYEIPG TGNIPRRIQVIDPRVLGPHYSFRSWDMRRHTFSRASIKRVSEQQETSDLVFSGPKKPR VDIPKQETQEESSHSLQRESRPWETEEESETEALSQESQEVPFQQQLQQQYQEQLKL RQGIKVLFEQLIRTQQGVHVNPCLR (SEQ ID NO: 22) |
| ORF1/2 | MAWGWWKRRRRWWFRKRWTRGRLRRRWPRSARRRPRRRRAQKSLGSTSQNKK PKKKAHIHSKENRDRGRPRKKARQKPSRKRAKRSPSNSSCSSSTKSSSSSDRESKSSS SSS (SEQ ID NO: 23) |

**Table B1. Exemplary *Anellovirus* nucleic acid sequence (*Betatorquevirus*)**

| | |
|---|---|
| Name | Ring2 |
| Genus/Clade | *Betatorquevirus* |
| Accession Number | JX134045.1 |

**Full Sequence:** 2797 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
TAATAAATATTCAACAGGAAAACCACCTAATTTAAATTGCCGACCACAAA
CCGTCACTTAGTTCCCCTTTTTGCAACAACTTCTGCTTTTTTCCAACTGC
CGGAAAACCACATAATTTGCATGGCTAACCACAAACTGATATGCTAATTA
ACTTCCACAAAACAACTTCCCCTTTTAAAACCACACCTACAAATTAATTA
TTAAACACAGTCACATCCTGGGAGGTACTACCACACTATAATACCAAGTG
CACTTCCGAATGGCTGAGTTTATGCCGCTAGACGGAGAACGCATCAGTTA
CTGACTGCGGACTGAACTTGGGCGGGTGCCGAAGGTGAGTGAAACCACCG
AAGTCAAGGGGCAATTCGGGCTAGTTCAGTCTAGCGGAACGGGCAAGAAA
CTTAAAATTATTTTATTTTTCAGATGAGCGACTGCTTTAAACCAACATGC
TACAACAACAAAACAAAGCAAACTCACTGGATTAATAACCTGCATTTAAC
CCACGACCTGATCTGCTTCTGCCCAACACCAACTAGACACTTATTACTAG
CTTTAGCAGAACAACAAGAAACAATTGAAGTGTCTAAACAAGAAAAAGAA
AAAATAACAAGATGCCTTATTACTACAGAAGAAGACGGTACAACTACAGA
```

CGTCCTAGATGGTATGGACGAGGTTGGATTAGACGCCCTTTTCGCAGAAG

ATTTCGAAGAAAAAGAAGGGTAAGACCTACTTATACTACTATTCCTCTAA

AGCAATGGCAACCGCCATATAAAAGAACATGCTATATAAAAGGACAAGAC

TGTTTAATATACTATAGCAACTTAAGACTGGGAATGAATAGTACAATGTA

TGAAAAAAGTATTGTACCTGTACATTGGCCGGGAGGGGGTTCTTTTTCTG

TAAGCATGTTAACTTTAGATGCCTTGTATGATATACATAAACTTTGTAGA

AACTGGTGGACATCCACAAACCAAGACTTACCACTAGTAAGATATAAAGG

ATGCAAAATAACATTTTATCAAAGCACATTTACAGACTACATAGTAAGAA

TACATACAGAACTACCAGCTAACAGTAACAAACTAACATACCCAAACACA

CATCCACTAATGATGATGATGTCTAAGTACAAACACATTATACCTAGTAG

ACAAACAAGAAGAAAAAAGAAACCATACACAAAAATATTTGTAAAACCAC

CTCCGCAATTTGAAAACAAATGGTACTTTGCTACAGACCTCTACAAAATT

CCATTACTACAAATACACTGCACAGCATGCAACTTACAAAACCCATTTGT

AAAACCAGACAAATTATCAAACAATGTTACATTATGGTCACTAAACACCA

TAAGCATACAAAATAGAAACATGTCAGTGGATCAAGGACAATCATGGCCA

TTTAAAATACTAGGAACACAAAGCTTTTATTTTTACTTTTACACCGGAGC

AAACCTACCAGGTGACACAACACAAATACCAGTAGCAGACCTATTACCAC

TAACAAACCCAAGAATAAACAGACCAGGACAATCACTAAATGAGGCAAAA

ATTACAGACCATATTACTTTCACAGAATACAAAAACAAATTTACAAATTA

TTGGGGTAACCCATTTAATAAACACATTCAAGAACACCTAGATATGATAC

TATACTCACTAAAAAGTCCAGAAGCAATAAAAAACGAATGGACAACAGAA

AACATGAAATGGAACCAATTAAACAATGCAGGAACAATGGCATTAACACC

ATTTAACGAGCCAATATTCACACAAATACAATATAACCCAGATAGAGACA

CAGGAGAAGACACTCAATTATACCTACTCTCTAACGCTACAGGAACAGGA

TGGGACCCACCAGGAATTCCAGAATTAATACTAGAAGGATTTCCACTATG

GTTAATATATTGGGGATTTGCAGACTTTCAAAAAAACCTAAAAAAAGTAA

CAAACATAGACACAAATTACATGTTAGTAGCAAAAACAAAATTTACACAA

AAACCTGGCACATTCTACTTAGTAATACTAAATGACACCTTTGTAGAAGG

CAATAGCCCATATGAAAAACAACCTTTACCTGAAGACAACATTAAATGGT

ACCCACAAGTACAATACCAATTAGAAGCACAAAACAAACTACTACAAACT

GGGCCATTTACACCAAACATACAAGGACAACTATCAGACAATATATCAAT

GTTTTATAAATTTTACTTTAAATGGGGAGGAAGCCCACCAAAAGCAATTA

ATGTTGAAAATCCTGCCCACCAGATTCAATATCCCATACCCCGTAACGAG

CATGAAACAACTTCGTTACAGAGTCCAGGGGAAGCCCCAGAATCCATCTT

ATACTCCTTCGACTATAGACACGGGAACTACACAACAACAGCTTTGTCAC
GAATTAGCCAAGACTGGGCACTTAAAGACACTGTTTCTAAAATTACAGAG
CCAGATCGACAGCAACTGCTCAAACAAGCCCTCGAATGCCTGCAAATCTC
GGAAGAAACGCAGGAGAAAAAAGAAAAAGAAGTACAGCAGCTCATCAGCA
ACCTCAGACAGCAGCAGCAGCTGTACAGAGAGCGAATAATATCATTATTA
AAGGACCAATAACTTTTAACTGTGTAAAAAAGGTGAAATTGTTTGATGAT
AAACCAAAAAACCGTAGATTTACACCTGAGGAATTTGAAACTGAGTTACA
AATAGCAAAATGGTTAAAGAGACCCCCAAGATCCTTTGTAAATGATCCTC
CCTTTTACCCATGGTTACCACCTGAACCTGTTGTAAACTTTAAGCTTAAT
TTTACTGAATAAAGGCCAGCATTAATTCACTTAAGGAGTCTGTTTATTTA
AGTTAAACCTTAATAAACGGTCACCGCCTCCCTAATACGCAGGCGCAGAA
AGGGGGCTCCGCCCCCTTTAACCCCCAGGGGGCTCCGCCCCCTGAAACCC
CCAAGGGGGCTACGCCCCCTTACACCCCC (SEQ ID NO: 54)

**Annotations:**

| *Putative Domain* | *Base range* |
|---|---|
| TATA Box | 237– 243 |
| Cap Site | 260 – 267 |
| Transcriptional Start Site | 267 |
| 5’ UTR Conserved Domain | 323 – 393 |
| ORF2 | 424 – 723 |
| ORF2/2 | 424 – 719 ; 2274 – 2589 |
| ORF2/3 | 424 – 719 ; 2449 – 2812 |
| ORF1 | 612 – 2612 |
| ORF1/1 | 612 – 719 ; 2274 – 2612 |
| ORF1/2 | 612 – 719 ; 2449 – 2589 |
| Three open-reading frame region | 2441 – 2586 |
| Poly(A) Signal | 2808 – 2813 |
| GC-rich region | 2868 – 2929 |

**Table B2. Exemplary *Anellovirus* amino acid sequences (*Betatorquevirus*)**

| Ring2 (*Betatorquevirus*) | |
|---|---|
| ORF2 | MSDCFKPTCYNNKTKQTHWINNLHLTHDLICFCPTPTRHLLLALAEQQETIEVSKQE KEKITRCLITTEEDGTTTDVLDGMDEVGLDALFAEDFEEKEG (SEQ ID NO: 55) |

| | |
|---|---|
| ORF2/2 | MSDCFKPTCYNNKTKQTHWINNLHLTHDLICFCPTPTRHLLLALAEQQETIEVSKQE KEKITRCLITTEEDGTTTDVLDGMDEVGLDALFAEDFEEKEGFNIPYPVTSMKQLRY RVQGKPQNPSYTPSTIDTGTTQQQLCHELAKTGHLKTLFLKLQSQIDSNCSNKPSNA CKSRKKRRRKKKKKYSSSSATSDSSSSCTESE (SEQ ID NO: 56) |
| ORF2/3 | MSDCFKPTCYNNKTKQTHWINNLHLTHDLICFCPTPTRHLLLALAEQQETIEVSKQE KEKITRCLITTEEDGTTTDVLDGMDEVGLDALFAEDFEEKEGARSTATAQTSPRMP ANLGRNAGEKRKRSTAAHQQPQTAAAAAVQRANNIIIKGPITFNCVKKVKLFDDKPK NRRFTPEEFETELQIAKWLKRPPRSFVNDPPFYPWLPPEPVVNFKLNFTE (SEQ ID NO: 57) |
| ORF1 | MPYYYRRRRYNYRRPRWYGRGWIRRPFRRRFRRKRRVRPTYTTIPLKQWQPPYKR TCYIKGQDCLIYYSNLRLGMNSTMYEKSIVPVHWPGGGSFSVSMLTLDALYDIHKL CRNWWTSTNQDLPLVRYKGCKITFYQSTFTDYIVRIHTELPANSNKLTYPNTHPLM MMMSKYKHIIPSRQTRRKKKPYTKIFVKPPPQFENKWYFATDLYKIPLLQIHCTACN LQNPFVKPDKLSNNVTLWSLNTISIQNRNMSVDQGQSWPFKILGTQSFYFYFYTGA NLPGDTTQIPVADLLPLTNPRINRPGQSLNEAKITDHITFTEYKNKFTNYWGNPFNK HIQEHLDMILYSLKSPEAIKNEWTTENMKWNQLNNAGTMALTPFNEPIFTQIQYNP DRDTGEDTQLYLLSNATGTGWDPPGIPELILEGFPLWLIYWGFADFQKNLKKVTNID TNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQPLPEDNIKWYPQVQYQLEA QNKLLQTGPFTPNIQGQLSDNISMFYKFYFKWGGSPPKAINVENPAHQIQYPIPRNE HETTSLQSPGEAPESILYSFDYRHGNYTTTALSRISQDWALKDTVSKITEPDRQQLLK QALECLQISEETQEKKEKEVQQLISNLRQQQQLYRERIISLLKDQ (SEQ ID NO: 58) |
| ORF1/1 | MPYYYRRRRYNYRRPRWYGRGWIRRPFRRRFRRKRRIQYPIPRNEHETTSLQSPGE APESILYSFDYRHGNYTTTALSRISQDWALKDTVSKITEPDRQQLLKQALECLQISEE TQEKKEKEVQQLISNLRQQQQLYRERIISLLKDQ (SEQ ID NO: 59) |
| ORF1/2 | MPYYYRRRRYNYRRPRWYGRGWIRRPFRRRFRRKRRSQIDSNCSNKPSNACKSRK KRRRKKKKKYSSSSATSDSSSSCTESE (SEQ ID NO: 60) |

**Table C1. Exemplary *Anellovirus* nucleic acid sequence (*Gammatorquevirus*)**

| | |
|---|---|
| Name | Ring4 |
| Genus/Clade | *Gammatorquevirus* |
| Accession Number | |

**Full Sequence:** 3176 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
TAAAATGGCGGGAGCCAATCATTTTATACTTTCACTTTCCAATTAAAAAT
GGCCACGTCACAAACAAGGGGTGGAGCCATTTAAACTATATAACTAAGTG
GGGTGGCGAATGGCTGAGTTTACCCCGCTAGACGGTGCAGGGACCGGATC
GAGCGCAGCGAGGAGGTCCCCGGCTGCCCATGGGCGGGAGCCGAGGTGAG
TGAAACCACCGAGGTCTAGGGGCAATTCGGGCTAGGGCAGTCTAGCGGAA
CGGGCAAGAAACTTAAAACAATATTTGTTTTACAGATGGTTAGTATATCC
TCAAGTGATTTTTTTAAGAAAACGAAATTTAATGAGGAGACGCAGAACCA
AGTATGGATGTCTCAAATTGCTGACTCTCATGATAATATCTGCAGTTGCT
GGCATCCATTTGCTCACCTTCTTGCTTCCATATTTCCTCCTGGCCACAAA
GATCGTGATCTTACTATTAACCAAATTCTTCTAAGAGATTATAAAGAAAA
ATGCCATTCTGGTGGAGAAGAAGGAGAAAATTCTGGACCAACAACAGGTT
TAATTACACCAAAAGAAGAAGATATAGAAAAAGATGGCCCAGAAGGCGCC
GCAGAAGAAGACCATACAGACGCCCTGTTCGCCGCCGCCGTAGAAAACTT
CGAAAGGTAAAGAGAAAAAAAAAATCTTTAATTGTTAGACAATGGCAACC
AGACAGTATAAGAACTTGTAAAATTATAGGACAGTCAGCTATAGTTGTTG
GGGCTGAAGGAAAGCAAATGTACTGTTATACTGTCAATAAGTTAATTAAT
GTGCCCCCAAAAACACCATATGGGGGAGGCTTTGGAGTAGACCAATACAC
ACTGAAATACTTATATGAAGAATACAGATTTGCACAAAACATTTGGACAC
AATCTAATGTACTGAAAGACTTATGCAGATACATAAATGTTAAGCTAATA
TTCTACAGAGACAACAAAACAGACTTTGTCCTTTCCTATGACAGAAACCC
ACCTTTTCAACTAACAAAATTTACATACCCAGGAGCACACCCACAACAAA
TCATGCTTCAAAAACACCACAAATTCATACTATCACAAATGACAAAGCCT
AATGGAAGACTAACAAAAAAACTCAAAATTAAACCTCCTAAACAAATGCT
TTCTAAATGGTTCTTTTCAAAACAATTCTGTAAATACCCTTTACTATCTC
TTAAAGCTTCTGCACTAGACCTTAGGCACTCTTACCTAGGCTGCTGTAAT
GAAAATCCACAGGTATTTTTTTATTATTTAAACCATGGATACTACACAAT
AACAAACTGGGGAGCACAATCCTCAACAGCATACAGACCTAACTCCAAGG
TGACAGACACAACATACTACAGATACAAAAATGACAGAAAAAATATTAAC
ATTAAAAGCCATGAATACGAAAAAAGTATATCATATGAAAACGGTTATTT
TCAATCTAGTTTCTTACAAACACAGTGCATATATACCAGTGAGCGTGGTG
AAGCCTGTATAGCAGAAAAACCACTAGGAATAGCTATTTACAATCCAGTA
AAAGACAATGGAGATGGTAATATGATATACCTTGTAAGCACTCTAGCAAA
CACTTGGGACCAGCCTCCAAAAGACAGTGCTATTTTAATACAAGGAGTAC
CCATATGGCTAGGCTTATTTGGATATTTAGACTACTGTAGACAAATTAAA
GCTGACAAAACATGGCTAGACAGTCATGTACTAGTAATTCAAAGTCCTGC
TATTTTTACTTACCCAAATCCAGGAGCAGGCAAATGGTATTGTCCACTAT
CACAAAGTTTTATAAATGGCAATGGTCCGTTTAATCAACCACCTACACTG
CTACAAAAAGCAAAGTGGTTTCCACAAATACAATACCAACAAGAAATTAT
TAATAGCTTTGTAGAATCAGGACCATTTGTTCCCAAATATGCAAATCAAA
CTGAAAGCAACTGGGAACTAAAATATAAATATGTTTTTACATTTAAGTGG
GGTGGACCACAATTCCATGAACCAGAAATTGCTGACCCTAGCAAACAAGA
GCAGTATGATGTCCCCGATACTTTCTACCAAACAATACAAATTGAAGATC
CAGAAGGACAAGACCCCAGATCTCTCATCCATGATTGGGACTACAGACGA
```

GGCTTTATTAAAGAAAGATCTCTTAAAAGAATGTCAACTTACTTCTCAAC
TCATACAGATCAGCAAGCAACTTCAGAGGAAGACATTCCCAAAAAGAAAA
AGAGAATTGGACCCCAACTCACAGTCCCACAACAAAAAGAAGAGGAGACA
CTGTCATGTCTCCTCTCTCTCTGCAAAAAAGATACCTTCCAAGAAACAGA
GACACAAGAAGACCTCCAGCAGCTCATCAAGCAGCAGCAGGAGCAGCAGC
TCCTCCTCAAGAGAAACATCCTCCAGCTCATCCACAAACTAAAAGAGAAT
CAACAAATGCTTCAGCTTCACACAGGCATGTTACCTTAACCAGATTTAAA
CCTGGATTTGAAGAGCAAACAGAGAGAGAATTAGCAATTATATTTCATAG
GCCCCCTAGAACCTACAAAGAGGACCTTCCATTCTATCCCTGGCTACCAC
CTGCACCCCTTGTACAATTTAACCTTAACTTCAAAGGCTAGGCCAACAAT
GTACACTTAGTAAAGCATGTTTATTAAAGCACAACCCCCAAAATAAATGT
AAAAATAAAAAAAAAAAAAAAAAAATAAAAAATTGCAAAAATTCGGCGCT
CGCGCGCATGTGCGCCTCTGGCGCAAATCACGCAACGCTCGCGCGCCCGC
GTATGTCTCTTTACCACGCACCTAGATTGGGGTGCGCGCGCTAGCGCGCG
CACCCCAATGCGCCCCGCCCTCGTTCCGACCCGCTTGCGCGGGTCGGACC
ACTTCGGGCTCGGGGGGGCGCGCCTGCGGCGCTTTTTTACTAAACAGACT
CCGAGCCGCCATTTGGCCCCCTAAGCTCCGCCCCCCTCATGAATATTCAT
AAAGGAAACCACATAATTAGAATTGCCGACCACAAACTGCCATATGCTAA
TTAGTTCCCCTTTTACAAAGTAAAAGGGGAAGTGAACATAGCCCCACACC
CGCAGGGGCAAGGCCCCGCACCCCTACGTCACTAACCACGCCCCCGCCGC
CATCTTGGGTGCGGCAGGGCGGGGGC (SEQ ID NO: 886)

**Annotations:**

| *Putative Domain* | *Base range* |
|---|---|
| TATA Box | 87– 93 |
| Cap Site | 110 – 117 |
| Transcriptional Start Site | 117 |
| 5’ UTR Conserved Domain | 185 – 254 |
| ORF2 | 286 – 660 |
| ORF2/2 | 286 – 656 ; 1998 – 2442 |
| ORF2/3 | 286 – 656 ; 2209 – 2641 |
| TAIP | 385 - 484 |
| ORF1 | 501 – 2489 |
| ORF1/1 | 501 – 656 ; 1998 – 2489 |
| ORF1/2 | 501 – 656 ; 2209 – 2442 |
| Three open-reading frame region | 2209 – 2439 |
| Poly(A) Signal | 2672 – 2678 |
| GC-rich region | 3076 – 3176 |

**Table C2. Exemplary *Anellovirus* amino acid sequences (*Gammatorquevirus*)**

| Ring4 (*Gammatorquevirus*) | |
|---|---|
| ORF2 | MVSISSSDFFKKTKFNEETQNQVWMSQIADSHDNICSCWHPFAHLLASIFPPGHKDR DLTINQILLR DYKEKCHSGGEEGENSGPTTGLITPKEEDIEKDGPEGAAEEDHTDALFAAAVENFER (SEQ ID NO: 887) |
| ORF2/2 | MVSISSSDFFKKTKFNEETQNQVWMSQIADSHDNICSCWHPFAHLLASIFPPGHKDR DLTINQILLRDYKEKCHSGGEEGENSGPTTGLITPKEEDIEKDGPEGAAEEDHTDALF AAAVENFESGVDHNSMNQKLLTLANKSSMMSPILSTKQYKLKIQKDKTPDLSSMIG TTDEALLKKDLLKECQLTSQLIQISKQLQRKTFPKRKRELDPNSQSHNKKKRRHCH VSSLSAKKIPSKKQRHKKTSSSSSSSRSSSSSSRETSSSSTN (SEQ ID NO: 888) |
| ORF2/3 | MVSISSSDFFKKTKFNEETQNQVWMSQIADSHDNICSCWHPFAHLLASIFPPGHKDR DLTINQILLRDYKEKCHSGGEEGENSGPTTGLITPKEEDIEKDGPEGAAEEDHTDALF AAAVENFERSASNFRGRHSQKEKENWTPTHSPTTKRRGDTVMSPLSLQKRYLPRNR DTRRPPAAHQAAAGAAAPPQEKHPPAHPQTKRESTNASASHRHVTLTRFKPGFEEQ TERELAIIFHRPPRTYKEDLPFYPWLPPAPLVQFNLNFKG (SEQ ID NO: 889) |
| TAIP | MRRRRTKYGCLKLLTLMIISAVAGIHLLTFLLPYFLLATKIVILLLTKFF (SEQ ID NO: 890) |
| ORF1 | MPFWWRRRRKFWTNNRFNYTKRRRYRKRWPRRRRRRRPYRRPVRRRRRKLRKV KRKKKSLIVRQWQPDSIRTCKIIGQSAIVVGAEGKQMYCYTVNKLINVPPKTPYGGG FGVDQYTLKYLYEEYRFAQNIWTQSNVLKDLCRYINVKLIFYRDNKTDFVLSYDRN PPFQLTKFTYPGAHPQQIMLQKHHKFILSQMTKPNGRLTKKLKIKPPKQMLSKWFFS KQFCKYPLLSLKASALDLRHSYLGCCNENPQVFFYYLNHGYYTITNWGAQSSTAYR PNSKVTDTTYYRYKNDRKNINIKSHEYEKSISYENGYFQSSFLQTQCIYTSERGEACI AEKPLGIAIYNPVKDNGDGNMIYLVSTLANTWDQPPKDSAILIQGVPIWLGLFGYLD YCRQIKADKTWLDSHVLVIQSPAIFTYPNPGAGKWYCPLSQSFINGNGPFNQPPTLL QKAKWFPQIQYQQEIINSFVESGPFVPKYANQTESNWELKYKYVFTFKWGGPQFHE PEIADPSKQEQYDVPDTFYQTIQIEDPEGQDPRSLIHDWDYRRGFIKERSLKRMSTYF STHTDQQATSEEDIPKKKKRIGPQLTVPQQKEEETLSCLLSLCKKDTFQETETQEDL QQLIKQQQEQQLLLKRNILQLIHKLKENQQMLQLHTGMLP (SEQ ID NO: 891) |
| ORF1/1 | MPFWWRRRRKFWTNNRFNYTKRRRYRKRWPRRRRRRRPYRRPVRRRRRKLRKW GGPQFHEPEIADPSKQEQYDVPDTFYQTIQIEDPEGQDPRSLIHDWDYRRGFIKERSL KRMSTYFSTHTDQQATSEEDIPKKKKRIGPQLTVPQQKEEETLSCLLSLCKKDTFQE |

| | |
|---|---|
| | TETQEDLQQLIKQQQEQQLLLKRNILQLIHKLKENQQMLQLHTGMLP (SEQ ID NO: 892) |
| ORF1/2 | MPFWWRRRRKFWTNNRFNYTKRRRYRKRWPRRRRRRRPYRRPVRRRRRKLRKIS KQLQRKTFPKRKR ELDPNSQSHNKKKRRHCHVSSLSAKKIPSKKQRHKKTSSSSSSSSRSSSSSSRETSSSS STN (SEQ ID NO: 893) |

**Table D1. Exemplary *Anellovirus* nucleic acid sequence (*Betatorquevirus*)**

| | |
|---|---|
| Name | Ring9 |
| Genus/Clade | *Betatorquevirus* |
| Accession Number | MH649263.1 |

**Full Sequence:** 2845 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
TTATTAATATTCAACAGGAAAACCACCTAATTTAAATTGCCGACCACAAA
CCGTCACTAACTTCCTTATTTAACATTACTTCCCTTTTAACCAATGAATA
TTCATACAACACATCACACTTCCTGGGAGGAGACATAAAACTATATAACT
AACTACACAGACGAATGGCTGAGTTTATGCCGCTAGACGGAGGACGCACA
GCTACTGCTGCGACCTGAACTTGGGCGGGTGCCGAAGGTGAGTGTAACCA
CCGTAGTCAAGGGGCAATTCGGGCTAGTTCAGTCTAGCGGAACGGGCAAG
ATTATTAATACAAACTTATTTTTACAGATGAGCAAACAACTAAAACCAAC
TTTATACAAAGACAAATCATTGGAATTACAATGGCTAAACAACATTTTTA
GCTCTCACGACCTGTGCTGCGGCTGCAACGATCCAGTTTTACATTTACTG
ATTTTAATTAACAAAACCGGAGAAGCACCTAAACCAGAAGAAGACATTAA
AAATATAAAATGCCTCCTTACTGGCGCCAAAAATACTACCGAAGAAGATA
TAGACCTTTCTCCTGGAGAACTAGAAGAATTATTCAAAGAAGAAAAAGAT
GGAGATACCGCAAACCAAGAAAAACATACTGGAGAAGAAAACTGCGGGTA
AGAAAACGTTTTTATAAAAGAAAGTTAAAAAAAATTGTACTTAAACAGTT
TCAACCAAAAATTATTAGAAGATGTACAATATTTGGAACAATCTGCCTAT
TTCAAGGCTCTCCAGAAAGAGCCAACAATAATTATATTCAAACAATCTAC
TCCTACGTACCAGATAAAGAACCAGGAGGAGGGGGATGGACTTTAATAAC
TGAAAGCTTAAGTAGTTTATGGGAAGACTGGGAACATTTAAAAAATGTAT
GGACTCAAAGTAACGCTGGTTTACCACTTGTAAGATACGGGGGAGTAACA
TTATACTTTTATCAATCTGCCTATACTGACTATATTGCTCAAGTTTTCAA
CTGTTATCCTATGACAGACACAAAATACACACATGCAGACTCAGCACCAA
ACAGAATGTTATTAAAAAAACATGTAATAAGAGTACCTAGCAGAGAAACA
CGCAAAAAAAGAAAGCCATACAAAAGAGTTAGAGTAGGACCTCCTTCTCA
AATGCAAAACAAATGGTACTTTCAAAGAGACATATGTGAAATACCATTAA
TAATGATTGCAGCCACAGCCGTTGACTTTAGATATCCCTTTTGTGCAAGC
GACTGTGCTAGTAACAACTTAACTCTAACATGTTTAAACCCACTATTGTT
TCAAAACCAAGACTTTGACCACCCATCCGATACACAAGGCTACTTTCCAA
AACCTGGAGTATATCTATACTCAACACAAAGAAGTAACAAGCCAAGTTCT
TCAGACTGTATATACTTAGGAAACACAAAAGACAATCAAGAAGGTAAATC
TGCAAGTAGTCTAATGACTCTAAAAACACAAAAAATAACAGATTGGGGAA
ATCCATTTTGGCATTATTATATAGACGGTTCTAAAAAAATATTTTCTTAC
```

TTTAAACCCCCATCACAATTAGACAGCAGCGACTTTGAACACATGACAGA
ATTAGCAGAACCAATGTTTATACAAGTTAGATACAACCCAGAAAGAGACA
CAGGACAAGGAAACTTAATATACGTAACAGAAAACTTTAGAGGACAACAC
TGGGACCCTCCATCTAGTGACAACCTAAAATTAGATGGATTTCCCTTATA
TGACATGTGCTGGGGTTTCATAGACTGGATAGAAAAAGTTCATGAAACAG
AAAACTTACTTACCAACTACTGCTTCTGTATTAGAAGCAGCGCTTTCAAT
GAAAAAAAAACAGTTTTTATACCTGTAGATCATTCATTTTTAACAGGTTT
TAGCCCATATGAAACTCCAGTTAAATCATCAGACCAAGCTCACTGGCACC
CACAAATAAGATTTCAAACAAAATCAATAAATGACATTTGTTTAACAGGC
CCCGGTTGTGCTAGGTCCCCATATGGCAATTACATGCAGGCAAAAATGAG
TTATAAATTTCATGTAAAATGGGGAGGATGTCCAAAAACTTATGAAAAAC
CATATGATCCTTGTTCACAGCCCAATTGGACTATTCCCCATAACCTCAAT
GAAACAATACAAATCCAGAATCCAAACACATGCCCACAAACAGAACTCCA
AGAATGGGACTGGCGACGTGATATTGTTACAAAAAAAGCTATCGAAAGAA
TTAGACAACACACGGAACCTCATGAAACTTTGCAAATCTCTACAGGTTCC
AAACACAACCCACCAGTACACAGACAAACATCACCGTGGACGGACTCAGA
AACGGACTCGGAAGAGGAAAAAGACCAAACACAAGAGATCCAGATCCAGC
TCAACAAGCTCAGAAAGCATCAACAGCATCTCAAGCAGCAGCTCAAGCAG
TACCTGAAACCCCAAAATATAGAATAGTTGCAAGCAACATAAAAGTTGAA
CTTTTTCCTACTAAAAAACCTTTTAAAAACAGACGCTTTACTCCTTCTGA
AAGAGAAACAGAAAGACAATGTGCTAAAGCTTTTTGTAGACCAGAAAGAC
ATTTCTTTTATGATCCTCCTTTTTACCCTTACTGTGTACCTGAACCTATT
GTAAACTTTGCTTTGGGATATAAAATTTAAGGCCAACAAATTTCACTTAG
TGGTGTCTGTTTATTAAAGTTTAACCTTAATAAGCATACTCCGCCTCCCT
ACATTAAGGCGCCAAAAGGGGGCTCCGCCCCCTTAAACCCCAAGGGGGCT
CCGCCCCCTTAAACCCCCAAGGGGGCTCCGCCCCCTTACACCCCC (SEQ ID NO: 1001)

**Annotations:**

| *Putative Domain* | *Base range* |
|---|---|
| TATA Box | 142 – 148 |
| Initiation Element | 162 – 177 |
| Transcriptional Start Site | 172 |
| 5’ UTR Conserved Domain | 226 – 296 |
| ORF2 | 328 – 651 |
| ORF2/2 | 328 – 647; 2121 – 2457 |
| ORF2/3 | 328 – 647; 2296 – 2680 |
| ORF1 | 510 – 2477 |
| ORF1/1 | 510 – 647; 2121 – 2477 |
| ORF1/2 | 510 – 647; 2296 – 2457 |
| Three open-reading frame region | 2296 – 2454 |
| GC-rich region | 2734 – 2845 |

**Table D2. Exemplary *Anellovirus* amino acid sequences (*Betatorquevirus*)**

| Ring9 (*Betatorquevirus*) | |
|---|---|
| ORF2 | MSKQLKPTLYKDKSLELQWLNNIFSSHDLCCGCNDPVLHLLILINKTGEAPK PEEDIKNIKCLLTGAKNTTEEDIDLSPGELEELFKEEKDGDTANQEKHTGEEN CG (SEQ ID NO: 1002) |
| ORF2/2 | MSKQLKPTLYKDKSLELQWLNNIFSSHDLCCGCNDPVLHLLILINKTGEAPK PEEDIKNIKCLLTGAKNTTEEDIDLSPGELEELFKEEKDGDTANQEKHTGEEN CGPIGLFPITSMKQYKSRIQTHAHKQNSKNGTGDVILLQKKLSKELDNTRNL MKLCKSLQVPNTTHQYTDKHHRGRTQKRTRKRKKTKHKRSRSSSTSSESIN SISSSSSSST (SEQ ID NO: 1003) |
| ORF2/3 | MSKQLKPTLYKDKSLELQWLNNIFSSHDLCCGCNDPVLHLLILINKTGEAPK PEEDIKNIKCLLTGAKNTTEEDIDLSPGELEELFKEEKDGDTANQEKHTGEEN CGFQTQPTSTQTNITVDGLRNGLGRGKRPNTRDPDPAQQAQKASTASQAAA QAVPETPKYRIVASNIKVELFPTKKPFKNRRFTPSERETERQCAKAFCRPERH FFYDPPFYPYCVPEPIVNFALGYKI (SEQ ID NO: 1004) |
| ORF1 | MPPYWRQKYYRRRYRPFSWRTRRIIQRRKRWRYRKPRKTYWRRKLRVRKR FYKRKLKKIVLKQFQPKIIRRCTIFGTICLFQGSPERANNNYIQTIYSYVPDKE PGGGGWTLITESLSSLWEDWEHLKNVWTQSNAGLPLVRYGGVTLYFYQSA YTDYIAQVFNCYPMTDTKYTHADSAPNRMLLKKHVIRVPSRETRKKRKPYK RVRVGPPSQMQNKWYFQRDICEIPLIMIAATAVDFRYPFCASDCASNNLTLT CLNPLLFQNQDFDHPSDTQGYFPKPGVYLYSTQRSNKPSSSDCIYLGNTKDN QEGKSASSLMTLKTQKITDWGNPFWHYYIDGSKKIFSYFKPPSQLDSSDFEH MTELAEPMFIQVRYNPERDTGQGNLIYVTENFRGQHWDPPSSDNLKLDGFP LYDMCWGFIDWIEKVHETENLLTNYCFCIRSSAFNEKKTVFIPVDHSFLTGFS PYETPVKSSDQAHWHPQIRFQTKSINDICLTGPGCARSPYGNYMQAKMSYK FHVKWGGCPKTYEKPYDPCSQPNWTIPHNLNETIQIQNPNTCPQTELQEWD WRRDIVTKKAIERIRQHTEPHETLQISTGSKHNPPVHRQTSPWTDSETDSEEE KDQTQEIQIQLNKLRKHQQHLKQQLKQYLKPQNIE (SEQ ID NO: 1005) |
| ORF1/1 | MPPYWRQKYYRRRYRPFSWRTRRIIQRRKRWRYRKPRKTYWRRKLRPNW TIPHNLNETIQIQNPNTCPQTELQEWDWRRDIVTKKAIERIRQHTEPHETLQIS TGSKHNPPVHRQTSPWTDSETDSEEEKDQTQEIQIQLNKLRKHQQHLKQQL KQYLKPQNIE (SEQ ID NO: 1006) |

| | |
|---|---|
| ORF1/2 | MPPYWRQKYYRRRYRPFSWRTRRIIQRRKRWRYRKPRKTYWRRKLRVPNT THQYTDKHHRGRTQKRTRKRKKTKHKRSRSSSTSSESINSISSSSSSST (SEQ ID NO: 1007) |

**Table E1. Exemplary *Anellovirus* nucleic acid sequence (*Betatorquevirus*)**

| | |
|---|---|
| Name | Ring10 |
| Genus/Clade | *Betatorquevirus* |
| Accession Number | JX134044.1 |

**Full Sequence:** 2912 bp

```
1         10         20         30         40         50
|         |          |          |          |          |
TAATAAATATTCAACAGGAAAACCACCTAATTTAAATTGCCGACCACAAA
CCGTCACTTAGTTCCTCTTTTTCCACAACTTCCTCTTTTACTAATGAATA
TTCATGTAATTAATTAATAATCACCGTAATTCCGGGGAGGAGCCTTTAAA
CTATAAAACTAACTACACATTCGAATGGCTGAGTTTATGCCGCCAGACGG
AGACGGGATCACTTCAGTGACTCCAGGCTGATCAAGGGCGGGTGCCGAAG
GTGAGTGAAACCACCGTAGTCAAGGGGCAATTCGGGCTAGATCAGTCTGG
CGGAACGGGCAAGAAACTTAAAATGTACTTTATTTTACAGAAATGTTCAA
ATCTCCAACATACTTAACAACTAAAGGCAAAAACAATGCCTTAATCAACT
GCTTCGTTGGAGACCACGATCTTCTGTGCAGCTGTAACAATCCTGCCTAC
CATTGCCTCCAAATACTTGCAACTACCTTAGCACCTCAACTAAAACAAGA
AGAAAAACAACAAATAATACAATGCCTTGGTGGTACAGACGCCGTAGCTA
CAACCCGTGGAGACGAAGAAATTGGTTTAGAAGACCTAGAAAAACTATTT
ACAGAAGATACAGAAGAAGACGCCGCTGGGTAAGAAGAAAACCTTTTTAC
AAACGTAAAATTAAGAGACTAAATATAGTAGAATGGCAACCTAAATCAAT
TAGAAAATGTAGAATAAAAGGAATGCTATGCTTGTTTCAAACGACAGAAG
ACAGACTGTCATATAACTTTGATATGTATGAAGAGTCTATTATACCAGAA
AAACTGCCGGGAGGGGGGGGATTTAGCATTAAGAATATAAGCTTATATGC
CTTATACCAAGAACACATACATGCACACAACATATTTACACACACAAACA
CAGACAGACCACTAGCAAGATACACAGGCTGTTCTTTAAAATTCTACCAA
AGCAAAGACATAGACTACGTAGTAACATATTCTACATCACTCCCACTAAG
AAGCTCAATGGGAATGTACAACTCCATGCAACCATCCATACATCTAATGC
AACAAAACAAACTAATTGTACCAAGCAAACAAACACAAAAAAGAAGAAAA
CCATATATTAAAAAACATATATCACCACCAACACAAATGAAATCTCAATG
GTACTTTCAACATAACATTGCAAACATACCGCTACTAATGATAAGAACCA
CAGCATTAACATTAGATAATTACTATATAGGAAGCAGACAATTAAGTACA
AATGTCACTATACATACACTTAACACAACATACATCCAAAACAGAGACTG
GGGAGACAGAAATAAAACTTACTACTGCCAAACATTAGGAACACAAAGAT
ACTTCCTATATGGAACACATTCAACTGCACAAAATATTAATGACATAAAG
CTACAAGAACTAATACCTTTAACAAACACACAAGACTATGTACAAGGCTT
TGATTGGACAGAAAAAGACAAACATAACATAACAACCTACAAAGAATTCT
TAACTAAAGGAGCAGGAAATCCATTTCACGCAGAATGGATAACAGCACAA
AACCCAGTAATACACACAGCAAACAGTCCTACACAAATAGAACAAATATA
CACCGCTTCAACAACAACATTCCAAAACAAAAAACTAACAGACCTACCAA
CGCCAGGATATATATTTATAACTCCAACAGTAAGCTTAAGATACAACCCA
TACAAAGACCTAGCAGAAAGAAACAAATGCTACTTTGTAAGAAGCAAAAT
AAATGCACACGGGTGGGACCCAGAACAACACCAAGAATTAATAAACAGTG
```

ACCTACCACAATGGTTACTATTATTTGGCTACCCAGACTACATAAAAAGA
ACACAAAACTTTGCATTAGTAGACACAAATTACATACTAGTAGACCACTG
CCCATACACAAATCCAGAAAAAACACCATTTATACCTTTAAGCACATCAT
TTATAGAAGGTAGAAGCCCATACAGTCCTTCAGACACACATGAACCAGAT
GAAGAAGACCAAAACAGGTGGTACCCATGCTACCAATATCAACAAGAATC
AATAAATTCAATATGTCTTAGCGGTCCAGGCACACCAAAAATACCAAAAG
GAATAACAGCAGAAGCAAAAGTAAAATATTCCTTTAATTTTAAGTGGGGT
GGTGACCTACCACCAATGTCTACAATTACAAACCCGACAGACCAGCCAAC
ATATGTTGTTCCCAATAACTTCAATGAAACAACTTCGTTACAGAATCCAA
CCACCAGACCAGAGCACTTCTTGTACTCCTTTGACGAAAGGAGGGGACAA
CTTACAGAAAAAGCTACAAAACGCTTGCTTAAAGACTGGGAAACTAAAGA
AACTTCTTTATTGTCTACAGAATACAGATTCGCGGAGCCAACACAAACAC
AAGCCCCACAAGAGGACCCGTCCTCGGAAGAAGAAGAAGAGAGCAACCTC
TTCGAGCGACTCCTCCGACAGCGAACCAAGCAGCTCCAGCTCAAGCGCAG
AATAATACAAACATTGAAAGACCTACAAAAATTAGAATAACTAACAGCAA
AAACACCGTTTACCTATTTCCACCTGAACAAAAGAACAGAAGACTAACAC
CATGGGAAATACAAGAAGACAAAGAAATAGCCAATTTATTTGGCAGACCA
CATAGATACTTTTTAAAAGACATTCCTTTCTATTGGGATATACCCCCAGA
GCCTAAAGTAAACTTTGATTTAAATTTTCAATAAAGAAATAAAGGGCAAG
GCCCCATTAACTCAAAGTCGGTGTCTACCTCTTTAAGTTTAACTTTACTA
AACGGACTCCGCCTCCCTAAATTTGGGCGCCAAAAGGGGGCTCCGCCCCC
TTAAACCCCAGGGGGCTCCGCCCCCTAAAACCCCCAAGGGGGCTACGCCC
CCTTACACCCCC (SEQ ID NO: 1008)

**Annotations:**

| *Putative Domain* | *Base range* |
|---|---|
| TATA Box | 152 – 158 |
| Initiation Element | 172 – 187 |
| Transcriptional Start Site | 182 |
| 5’ UTR Conserved Domain | 239 – 309 |
| ORF2 | 343 – 633 |
| ORF2/2 | 343 – 629; 2196 – 2505 |
| ORF2/3 | 343 – 629; 2371 – 2734 |
| ORF1 | 522 – 2540 |
| ORF1/1 | 522 – 629; 2196 – 2540 |
| ORF1/2 | 522 – 629; 2371 – 2505 |
| Three open-reading frame region | 2276 – 2502 |
| GC-rich region | 2803 – 2912 |

**Table E2. Exemplary *Anellovirus* amino acid sequences (*Betatorquevirus*)**

| Ring10 (*Betatorquevirus*) |
|---|

| | |
|---|---|
| ORF2 | MFKSPTYLTTKGKNNALINCFVGDHDLLCSCNNPAYHCLQILATTLAPQLK QEEKQQIIQCLGGTDAVATTRGDEEIGLEDLEKLFTEDTEEDAAG (SEQ ID NO: 1009) |
| ORF2/2 | MFKSPTYLTTKGKNNALINCFVGDHDLLCSCNNPAYHCLQILATTLAPQLK QEEKQQIIQCLGGTDAVATTRGDEEIGLEDLEKLFTEDTEEDAAGQHMLFPI TSMKQLRYRIQPPDQSTSCTPLTKGGDNLQKKLQNACLKTGKLKKLLYCLQ NTDSRSQHKHKPHKRTRPRKKKKRATSSSDSSDSEPSSSSSSAE (SEQ ID NO: 1010) |
| ORF2/3 | MFKSPTYLTTKGKNNALINCFVGDHDLLCSCNNPAYHCLQILATTLAPQLK QEEKQQIIQCLGGTDAVATTRGDEEIGLEDLEKLFTEDTEEDAAGIQIRGANT NTSPTRGPVLGRRRREQPLRATPPTANQAAPAQAQNNTNIERPTKIRITNSKN TVYLFPPEQKNRRLTPWEIQEDKEIANLFGRPHRYFLKDIPFYWDIPPEPKVN FDLNFQ (SEQ ID NO: 1011) |
| ORF1 | MPWWYRRRSYNPWRRRNWFRRPRKTIYRRYRRRRRWVRRKPFYKRKIKR LNIVEWQPKSIRKCRIKGMLCLFQTTEDRLSYNFDMYEESIIPEKLPGGGGFSI KNISLYALYQEHIHAHNIFTHTNTDRPLARYTGCSLKFYQSKDIDYVVTYSTS LPLRSSMGMYNSMQPSIHLMQQNKLIVPSKQTQKRRKPYIKKHISPPTQMKS QWYFQHNIANIPLLMIRTTALTLDNYYIGSRQLSTNVTIHTLNTTYIQNRDW GDRNKTYYCQTLGTQRYFLYGTHSTAQNINDIKLQELIPLTNTQDYVQGFD WTEKDKHNITTYKEFLTKGAGNPFHAEWITAQNPVIHTANSPTQIEQIYTAS TTTFQNKKLTDLPTPGYIFITPTVSLRYNPYKDLAERNKCYFVRSKINAHGW DPEQHQELINSDLPQWLLLFGYPDYIKRTQNFALVDTNYILVDHCPYTNPEK TPFIPLSTSFIEGRSPYSPSDTHEPDEEDQNRWYPCYQYQQESINSICLSGPGTP KIPKGITAEAKVKYSFNFKWGGDLPPMSTITNPTDQPTYVVPNNFNETTSLQ NPTTRPEHFLYSFDERRGQLTEKATKRLLKDWETKETSLLSTEYRFAEPTQT QAPQEDPSSEEEEESNLFERLLRQRTKQLQLKRRIIQTLKDLQKLE (SEQ ID NO: 1012) |
| ORF1/1 | MPWWYRRRSYNPWRRRNWFRRPRKTIYRRYRRRRRWPTYVVPNNFNETT SLQNPTTRPEHFLYSFDERRGQLTEKATKRLLKDWETKETSLLSTEYRFAEP TQTQAPQEDPSSEEEEESNLFERLLRQRTKQLQLKRRIIQTLKDLQKLE |

| | |
|---|---|
| ORF1/2 | MPWWYRRRSYNPWRRRNWFRRPRKTIYRRYRRRRRWNTDSRSQHKHKPH KRTRPRKKKKRATSSSDSSDSEPSSSSSSAE (SEQ ID NO: 1013) |

In some embodiments, an anellovector comprises a nucleic acid comprising a sequence listed in PCT Application No. PCT/US2018/037379, incorporated herein by reference in its entirety. In some embodiments, an anellovector comprises a polypeptide comprising a sequence listed in PCT Application No. PCT/US2018/037379, incorporated herein by reference in its entirety. In some embodiments, an anellovector comprises a nucleic acid comprising a sequence listed in PCT Application No. PCT/US19/65995, incorporated herein by reference in its entirety. In some embodiments, an anellovector comprises a polypeptide comprising a sequence listed in PCT Application No. PCT/US19/65995, incorporated herein by reference in its entirety.

***ORF1 Molecules***

In some embodiments, the anellovector comprises an ORF1 molecule and/or a nucleic acid encoding an ORF1 molecule. Generally, an ORF1 molecule comprises a polypeptide having the structural features and/or activity of an Anellovirus ORF1 protein (e.g., an Anellovirus ORF1 protein as described herein). In some embodiments, the ORF1 molecule comprises a truncation relative to an Anellovirus ORF1 protein (e.g., an Anellovirus ORF1 protein as described herein). An ORF1 molecule may be capable of binding to other ORF1 molecules, e.g., to form a proteinaceous exterior (e.g., as described herein), e.g., a capsid. In some embodiments, the proteinaceous exterior may enclose a nucleic acid molecule (e.g., a genetic element as described herein). In some embodiments, a plurality of ORF1 molecules may form a multimer, e.g., to form a proteinaceous exterior. In some embodiments, the multimer may be a homomultimer. In other embodiments, the multimer may be a heteromultimer.

An ORF1 molecule may, in some embodiments, comprise one or more of: a first region comprising an arginine rich region, e.g., a region having at least 60% basic residues (e.g., at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% basic residues; e.g., between 60%-90%, 60%-80%, 70%-90%, or 70-80% basic residues), and a second region comprising jelly-roll domain, e.g., at least six beta strands (e.g., 4, 5, 6, 7, 8, 9, 10, 11, or 12 beta strands).

In some embodiments, an ORF1 molecule as described herein comprises one or more lysine-to-histidine mutations relative to a wild-type ORF1 protein sequence (e.g., as described herein). In certain embodiments, the ORF1 molecuel comprises one or more lysine-to-histidine mutations in the arginine-rich region and/or the first beta strand.

*Arginine-rich region*

An arginine rich region has at least 70% (e.g., at least about 70, 80, 90, 95, 96, 97, 98, 99, or 100%) sequence identity to an arginine-rich region sequence described herein or a sequence of at least about 40 amino acids comprising at least 60%, 70%, or 80% basic residues (e.g., arginine, lysine, or a combination thereof).

*Jelly Roll domain*

A jelly-roll domain or region comprises (e.g., consists of) a polypeptide (e.g., a domain or region comprised in a larger polypeptide) comprising one or more (e.g., 1, 2, or 3) of the following characteristics:

(i) at least 30% (e.g., at least 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 90%, or more) of the amino acids of the jelly-roll domain are part of one or more β-sheets;

(ii) the secondary structure of the jelly-roll domain comprises at least four (e.g., at least 4, 5, 6, 7, 8, 9, 10, 11, or 12) β-strands; and/or

(iii) the tertiary structure of the jelly-roll domain comprises at least two (e.g., at least 2, 3, or 4) β-sheets; and/or

(iv) the jelly-roll domain comprises a ratio of β-sheets to α-helices of at least 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, or 10:1.

In certain embodiments, a jelly-roll domain comprises two β-sheets.

In certain embodiments, one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) of the β-sheets comprises about eight (e.g., 4, 5, 6, 7, 8, 9, 10, 11, or 12) β-strands. In certain embodiments, one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) of the β-sheets comprises eight β-strands. In certain embodiments, one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) of the β-sheets comprises seven β-strands. In certain embodiments, one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) of the β-sheets comprises six β-strands. In certain embodiments, one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) of the β-sheets comprises five β-strands. In certain embodiments, one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) of the β-sheets comprises four β-strands.

In some embodiments, the jelly-roll domain comprises a first β-sheet in antiparallel orientation to a second β-sheet. In certain embodiments, the first β-sheet comprises about four (e.g., 3, 4, 5, or 6) β-strands. In certain embodiments, the second β-sheet comprises about four (e.g., 3, 4, 5, or 6) β-strands. In embodiments, the first and second β-sheet comprise, in total, about eight (e.g., 6, 7, 8, 9, 10, 11, or 12) β-strands.

In certain embodiments, a jelly-roll domain is a component of a capsid protein (e.g., an ORF1 molecule as described herein). In certain embodiments, a jelly-roll domain has self-assembly activity. In

some embodiments, a polypeptide comprising a jelly-roll domain binds to another copy of the polypeptide comprising the jelly-roll domain. In some embodiments, a jelly-roll domain of a first polypeptide binds to a jelly-roll domain of a second copy of the polypeptide.

*N22 Domain*

An ORF1 molecule may also include a third region comprising the structure or activity of an Anellovirus N22 domain (e.g., as described herein, e.g., an N22 domain from an Anellovirus ORF1 protein as described herein), and/or a fourth region comprising the structure or activity of an Anellovirus C-terminal domain (CTD) (e.g., as described herein, e.g., a CTD from an Anellovirus ORF1 protein as described herein). In some embodiments, the ORF1 molecule comprises, in N-terminal to C-terminal order, the first, second, third, and fourth regions.

*Hypervariable Region (HVR)*

The ORF1 molecule may, in some embodiments, further comprise a hypervariable region (HVR), e.g., an HVR from an Anellovirus ORF1 protein, e.g., as described herein. In some embodiments, the HVR is positioned between the second region and the third region. In some embodiments, the HVR comprises comprises at least about 55 (e.g., at least about 45, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, or 65) amino acids (e.g., about 45-160, 50-160, 55-160, 60-160, 45-150, 50-150, 55-150, 60-150, 45-140, 50-140, 55-140, or 60-140 amino acids).

*Exemplary ORF1 Sequences*

Exemplary Anellovirus ORF1 amino acid sequences, and the sequences of exemplary ORF1 domains, are provided in the tables below. In some embodiments, a polypeptide (e.g., an ORF1 molecule) described herein comprises an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to one or more *Anellovirus* ORF1 subsequences, e.g., as described in any of Tables N-Z). In some embodiments, an anellovector described herein comprises an ORF1 molecule comprising an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to one or more *Anellovirus* ORF1 subsequences, e.g., as described in any of Tables N-Z. In some embodiments, an anellovector described herein comprises a nucleic acid molecule (e.g., a genetic element) encoding an ORF1 molecule comprising an amino acid sequence having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to one or more *Anellovirus* ORF1 subsequences, e.g., as described in any of Tables N-Z.

In some embodiments, the one or more *Anellovirus* ORF1 subsequences comprises one or more of an arginine (Arg)-rich domain, a jelly-roll domain, a hypervariable region (HVR), an N22 domain, or a C-terminal domain (CTD) (e.g., as listed in any of Tables N-Z), or sequences having at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. In some embodiments, the ORF1 molecule comprises a plurality of subsequences from different *Anelloviruses* (e.g., any combination of ORF1 subsequences selected from the *Alphatorquevirus* Clade 1-7 subsequences listed in Tables N-Z). In embodiments, the ORF1 molecule comprises one or more of an Arg-rich domain, a jelly-roll domain, an N22 domain, and a CTD from one *Anellovirus*, and an HVR from another. In embodiments, the ORF1 molecule comprises one or more of a jelly-roll domain, an HVR, an N22 domain, and a CTD from one *Anellovirus*, and an Arg-rich domain from another. In embodiments, the ORF1 molecule comprises one or more of an Arg-rich domain, an HVR, an N22 domain, and a CTD from one *Anellovirus*, and a jelly-roll domain from another. In embodiments, the ORF1 molecule comprises one or more of an Arg-rich domain, a jelly-roll domain, an HVR, and a CTD from one *Anellovirus*, and an N22 domain from another. In embodiments, the ORF1 molecule comprises one or more of an Arg-rich domain, a jelly-roll domain, an HVR, and an N22 domain from one *Anellovirus*, and a CTD from another.

Additional exemplary Anelloviruses for which the ORF1 molecules, or splice variants or functional fragments thereof, can be utilized in the compositions and methods described herein (e.g., to form the proteinaceous exterior of an anellovector, e.g., by enclosing a genetic element) are described, for example, in PCT Application Nos. PCT/US2018/037379 and PCT/US19/65995 (incorporated herein by reference in their entirety).

**Table N. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Alphatorquevirus*, Clade 3)**

| | |
|---|---|
| Name | Ring1 |
| Genus/Clade | *Alphatorquevirus*, Clade 3 |
| Accession Number | AJ620231.1 |
| Protein Accession Number | CAF05750.1 |

**Full Sequence:** 743 AA

```
1          10         20         30         40         50
|          |          |          |          |          |
MAWGWWKRRRRWWFRKRWTRGRLRRRWPRSARRRPRRRRVRRRRRRWRRGR
RKTRTYRRRRRFRRRGRKAKLIIKLWQPAVIKRCRIKGYIPLIISGNGTF
```

ATNFTSHINDRIMKGPFGGGHSTMRFSLYILFEEHLRHMNFWTRSNDNLE
LTRYLGASVKIYRHPDQDFIVIYNRRTPLGGNIYTAPSLHPGNAILAKHK
ILVPSLQTRPKGRKAIRLRIAPPTLFTDKWYFQKDIADLTLFNIMAVEAD
LRFPFCSPQTDNTCISFQVLSSVYNNYLSINTFNNDNSDSKLKEFLNKAF
PTTGTKGTSLNALNTFRTEGCISHPQLKKPNPQINKPLESQYFAPLDALW
GDPIYYNDLNENKSLNDIIEKILIKNMITYHAKLREFPNSYQGNKAFCHL
TGIYSPPYLNQGRISPEIFGLYTEIIYNPYTDKGTGNKVWMDPLTKENNI
YKEGQSKCLLTDMPLWTLLFGYTDWCKKDTNNWDLPLNYRLVLICPYTFP
KLYNEKVKDYGYIPYSYKFGAGQMPDGSNYIPFQFRAKWYPTVLHQQQVM
EDISRSGPFAPKVEKPSTQLVMKYCFNFNWGGNPIIEQIVKDPSFQPTYE
IPGTGNIPRRIQVIDPRVLGPHYSFRSWDMRRHTFSRASIKRVSEQQETS
DLVFSGPKKPRVDIPKQETQEESSHSLQRESRPWETEEESETEALSQESQ
EVPFQQQLQQQYQEQLKLRQGIKVLFEQLIRTQQGVHVNPCLR
(SEQ ID NO: 185)

**Annotations:**

| *Putative Domain* | *AA range* |
|---|---|
| Arg-Rich Region | 1 – 68 |
| Jelly-roll domain | 69 - 280 |
| Hypervariable Region | 281 - 413 |
| N22 | 414 – 579 |
| C-terminal Domain | 580 - 743 |

**Table O. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Alphatorquevirus*, Clade 3)**

| Ring1 ORF1 (*Alphatorquevirus* Clade 3) | |
|---|---|
| Arg-Rich Region | MAWGWWKRRRRWWFRKRWTRGRLRRRWPRSARRRPRRRRVRRRRRRWRR GRRKTRTYRRRRRFRRRGRK (SEQ ID NO: 186) |
| Jelly-roll Domain | AKLIIKLWQPAVIKRCRIKGYIPLIISGNGTFATNFTSHINDRIMKGPFGGGHST MRFSLYILFEEHLRHMNFWTRSNDNLELTRYLGASVKIYRHPDQDFIVIYNRR TPLGGNIYTAPSLHPGNAILAKHKILVPSLQTRPKGRKAIRLRIAPPTLFTDKWY FQKDIADLTLFNIMAVEADLRFPFCSPQTDNTCISFQVLSSVYNNYLSI (SEQ ID NO: 187) |

| | |
|---|---|
| Hypervariable domain | NTFNNDNSDSKLKEFLNKAFPTTGTKGTSLNALNTFRTEGCISHPQLKKPNPQI NKPLESQYFAPLDALWGDPIYYNDLNENKSLNDIIEKILIKNMITYHAKLREFP NSYQGNKAFCHLTGIYSPPYLNQGR (SEQ ID NO: 188) |
| N22 | ISPEIFGLYTEIIYNPYTDKGTGNKVWMDPLTKENNIYKEGQSKCLLTDMPLW TLLFGYTDWCKKDTNNWDLPLNYRLVLICPYTFPKLYNEKVKDYGYIPYSYK FGAGQMPDGSNYIPFQFRAKWYPTVLHQQQVMEDISRSGPFAPKVEKPSTQL VMKYCFNFN (SEQ ID NO: 189) |
| C-terminal domain | WGGNPIIEQIVKDPSFQPTYEIPGTGNIPRRIQVIDPRVLGPHYSFRSWDMRRHT FSRASIKRVSEQQETSDLVFSGPKKPRVDIPKQETQEESSHSLQRESRPWETEEE SETEALSQESQEVPFQQQLQQQYQEQLKLRQGIKVLFEQLIRTQQGVHVNPCL R (SEQ ID NO: 190) |

**Table P. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Betatorquevirus*)**

| | |
|---|---|
| Name | Ring2 |
| Genus/Clade | *Betatorquevirus* |
| Accession Number | JX134045.1 |
| Protein Accession Number | AGG91484.1 |

**Full Sequence:** 666 AA

```
1         10         20         30         40         50
|         |          |          |          |          |
MPYYYRRRRYNYRRPRWYGRGWIRRPFRRRFRRKRRVRPTYTTIPLKQWQ
PPYKRTCYIKGQDCLIYYSNLRLGMNSTMYEKSIVPVHWPGGGSFSVSML
TLDALYDIHKLCRNWWTSTNQDLPLVRYKGCKITFYQSTFTDYIVRIHTE
LPANSNKLTYPNTHPLMMMMSKYKHIIPSRQTRRKKKPYTKIFVKPPPQF
ENKWYFATDLYKIPLLQIHCTACNLQNPFVKPDKLSNNVTLWSLNTISIQ
NRNMSVDQGQSWPFKILGTQSFYFYFYTGANLPGDTTQIPVADLLPLTNP
RINRPGQSLNEAKITDHITFTEYKNKFTNYWGNPFNKHIQEHLDMILYSL
KSPEAIKNEWTTENMKWNQLNNAGTMALTPFNEPIFTQIQYNPDRDTGED
TQLYLLSNATGTGWDPPGIPELILEGFPLWLIYWGFADFQKNLKKVTNID
TNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQPLPEDNIKWYPQV
QYQLEAQNKLLQTGPFTPNIQGQLSDNISMFYKFYFKWGGSPPKAINVEN
```

PAHQIQYPIPRNEHETTSLQSPGEAPESILYSFDYRHGNYTTTALSRISQ
DWALKDTVSKITEPDRQQLLKQALECLQISEETQEKKEKEVQQLISNLRQ
QQQLYRERIISLLKDQ (SEQ ID NO: 215)

**Annotations:**

| *Putative Domain* | *AA range* |
|---|---|
| Arg-Rich Region | 1 – 38 |
| Jelly-roll domain | 39 - 246 |
| Hypervariable Region | 247 - 374 |
| N22 | 375 – 537 |
| C-terminal Domain | 538 – 666 |

**Table Q. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Betatorquevirus*)**

| Ring2 ORF1 (*Betatorquevirus*) | |
|---|---|
| Arg-Rich Region | MPYYYRRRRYNYRRPRWYGRGWIRRPFRRRFRRKRRVR (SEQ ID NO: 216) |
| Jelly-roll Domain | PTYTTIPLKQWQPPYKRTCYIKGQDCLIYYSNLRLGMNSTMYEKSIVPVHWPG GGSFSVSMLTLDALYDIHKLCRNWWTSTNQDLPLVRYKGCKITFYQSTFTDYI VRIHTELPANSNKLTYPNTHPLMMMMSKYKHIIPSRQTRRKKKPYTKIFVKPP PQFENKWYFATDLYKIPLLQIHCTACNLQNPFVKPDKLSNNVTLWSLNT (SEQ ID NO: 217) |
| Hypervariable domain | ISIQNRNMSVDQGQSWPFKILGTQSFYFYFYTGANLPGDTTQIPVADLLPLTNP RINRPGQSLNEAKITDHITFTEYKNKFTNYWGNPFNKHIQEHLDMILYSLKSPE AIKNEWTTENMKWNQLNNAG (SEQ ID NO: 218) |
| N22 | TMALTPFNEPIFTQIQYNPDRDTGEDTQLYLLSNATGTGWDPPGIPELILEGFPL WLIYWGFADFQKNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVILNDTFVEGN SPYEKQPLPEDNIKWYPQVQYQLEAQNKLLQTGPFTPNIQGQLSDNISMFYKF YFK (SEQ ID NO: 219) |
| C-terminal domain | WGGSPPKAINVENPAHQIQYPIPRNEHETTSLQSPGEAPESILYSFDYRHGNYT TTALSRISQDWALKDTVSKITEPDRQQLLKQALECLQISEETQEKKEKEVQQLI SNLRQQQQLYRERIISLLKDQ (SEQ ID NO: 220) |

**Table R. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Gammatorquevirus*)**

| | |
|---|---|
| Name | Ring4 |
| Genus/Clade | *Gammatorquevirus* |
| Accession Number | |
| Protein Accession Number | |

**Full Sequence:** 662 AA

```
1         10         20         30         40         50
|         |          |          |          |          |
MPFWWRRRRKFWTNNRFNYTKRRRYRKRWPRRRRRRRRPYRRPVRRRRRKL
RKVKRKKKSLIVRQWQPDSIRTCKIIGQSAIVVGAEGKQMYCYTVNKLIN
VPPKTPYGGGFGVDQYTLKYLYEEYRFAQNIWTQSNVLKDLCRYINVKLI
FYRDNKTDFVLSYDRNPPFQLTKFTYPGAHPQQIMLQKHHKFILSQMTKP
NGRLTKKLKIKPPKQMLSKWFFSKQFCKYPLLSLKASALDLRHSYLGCCN
ENPQVFFYYLNHGYYTITNWGAQSSTAYRPNSKVTDTTYYRYKNDRKNIN
IKSHEYEKSISYENGYFQSSFLQTQCIYTSERGEACIAEKPLGIAIYNPV
KDNGDGNMIYLVSTLANTWDQPPKDSAILIQGVPIWLGLFGYLDYCRQIK
ADKTWLDSHVLVIQSPAIFTYPNPGAGKWYCPLSQSFINGNGPFNQPPTL
LQKAKWFPQIQYQQEIINSFVESGPFVPKYANQTESNWELKYKYVFTFKW
GGPQFHEPEIADPSKQEQYDVPDTFYQTIQIEDPEGQDPRSLIHDWDYRR
GFIKERSLKRMSTYFSTHTDQQATSEEDIPKKKKRIGPQLTVPQQKEEET
LSCLLSLCKKDTFQETETQEDLQQLIKQQQEQQLLLKRNILQLIHKLKEN
QQMLQLHTGMLP (SEQ ID NO: 925)
```

**Annotations:**

| *Putative Domain* | *AA range* |
|---|---|
| Arg-Rich Region | 1 – 58 |
| Jelly-roll domain | 59 - 260 |
| Hypervariable Region | 261 - 339 |
| N22 | 340 – 499 |
| C-terminal Domain | 500 – 662 |

**Table S. Exemplary *Anellovirus* ORF1 amino acid subsequence (*Gammatorquevirus*)**

| Ring4 (*Gammatorquevirus*) | |
|---|---|
| Arg-Rich Region | MPFWWRRRRKFWTNNRFNYTKRRRYRKRWPRRRRRRRRPYRRPVRRRRRKL<br>RKVKRKKK (SEQ ID NO: 926) |

| | |
|---|---|
| Jelly-roll Domain | SLIVRQWQPDSIRTCKIIGQSAIVVGAEGKQMYCYTVNKLINVPPKTPYGGGF GVDQYTLKYLYEEYRFAQNIWTQSNVLKDLCRYINVKLIFYRDNKTDFVLSY DRNPPFQLTKFTYPGAHPQQIMLQKHHKFILSQMTKPNGRLTKKLKIKPPKQM LSKWFFSKQFCKYPLLSLKASALDLRHSYLGCCNENPQVFFYYL (SEQ ID NO: 927) |
| Hypervariable domain | NHGYYTITNWGAQSSTAYRPNSKVTDTTYYRYKNDRKNINIKSHEYEKSISYE NGYFQSSFLQTQCIYTSERGEACIAE (SEQ ID NO: 928) |
| N22 | KPLGIAIYNPVKDNGDGNMIYLVSTLANTWDQPPKDSAILIQGVPIWLGLFGY LDYCRQIKADKTWLDSHVLVIQSPAIFTYPNPGAGKWYCPLSQSFINGNGPFN QPPTLLQKAKWFPQIQYQQEIINSFVESGPFVPKYANQTESNWELKYKYVFTF K (SEQ ID NO: 929) |
| C-terminal domain | WGGPQFHEPEIADPSKQEQYDVPDTFYQTIQIEDPEGQDPRSLIHDWDYRRGFI KERSLKRMSTYFSTHTDQQATSEEDIPKKKKRIGPQLTVPQQKEEETLSCLLSL CKKDTFQETETQEDLQQLIKQQQEQQLLLKRNILQLIHKLKENQQMLQLHTG MLP (SEQ ID NO: 930) |

In some embodiments, the first region can bind to a nucleic acid molecule (e.g., DNA). In some embodiments, the basic residues are selected from arginine, histidine, or lysine, or a combination thereof. In some embodiments, the first region comprises at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% arginine residues (e.g., between 60%-90%, 60%-80%, 70%-90%, or 70-80% arginine residues). In some embodiments, the first region comprises about 30-120 amino acids (e.g., about 40-120, 40-100, 40-90, 40-80, 40-70, 50-100, 50-90, 50-80, 50-70, 60-100, 60-90, or 60-80 amino acids). In some embodiments, the first region comprises the structure or activity of a viral ORF1 arginine-rich region (e.g., an arginine-rich region from an Anellovirus ORF1 protein, e.g., as described herein). In some embodiments, the first region comprises a nuclear localization sigal.

In some embodiments, the second region comprises a jelly-roll domain, e.g., the structure or activity of a viral ORF1 jelly-roll domain (e.g., a jelly-roll domain from an Anellovirus ORF1 protein, e.g., as described herein). In some embodiments, the second region is capable of binding to the second region of another ORF1 molecule, e.g., to form a proteinaceous exterior (e.g., capsid) or a portion thereof.

In some embodiments, the fourth region is exposed on the surface of a proteinaceous exterior (e.g., a proteinaceous exterior comprising a multimer of ORF1 molecules, e.g., as described herein).

In some embodiments, the first region, second region, third region, fourth region, and/or HVR each comprise fewer than four (e.g., 0, 1, 2, or 3) beta sheets.

In some embodiments, one or more of the first region, second region, third region, fourth region, and/or HVR may be replaced by a heterologous amino acid sequence (e.g., the corresponding region from a heterologous ORF1 molecule). In some embodiments, the heterologous amino acid sequence has a desired functionality, e.g., as described herein.

In some embodiments, the ORF1 molecule comprises a plurality of conserved motifs (e.g., motifs comprising about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, or more amino acids) (e.g., as shown in Figure 34 of PCT/US19/65995). In some embodiments, the conserved motifs may show 60, 70, 80, 85, 90, 95, or 100% sequence identity to an ORF1 protein of one or more wild-type Anellovirus clades (e.g., *Alphatorquevirus*, clade 1; *Alphatorquevirus*, clade 2; *Alphatorquevirus*, clade 3; *Alphatorquevirus*, clade 4; *Alphatorquevirus*, clade 5; *Alphatorquevirus*, clade 6; *Alphatorquevirus*, clade 7; *Betatorquevirus*; and/or *Gammatorquevirus*). In embodiments, the conserved motifs each have a length between 1-1000 (e.g., between 5-10, 5-15, 5-20, 10-15, 10-20, 15-20, 5-50, 5-100, 10-50, 10-100, 10-1000, 50-100, 50-1000, or 100-1000) amino acids. In certain embodiments, the conserved motifs consist of about 2-4% (e.g., about 1-8%, 1-6%, 1-5%, 1-4%, 2-8%, 2-6%, 2-5%, or 2-4%) of the sequence of the ORF1 molecule, and each show 100% sequence identity to the corresponding motifs in an ORF1 protein of the wild-type Anellovirus clade. In certain embodiments, the conserved motifs consist of about 5-10% (e.g., about 1-20%, 1-10%, 5-20%, or 5-10%) of the sequence of the ORF1 molecule, and each show 80% sequence identity to the corresponding motifs in an ORF1 protein of the wild-type Anellovirus clade. In certain embodiments, the conserved motifs consist of about 10-50% (e.g., about 10-20%, 10-30%, 10-40%, 10-50%, 20-40%, 20-50%, or 30-50%) of the sequence of the ORF1 molecule, and each show 60% sequence identity to the corresponding motifs in an ORF1 protein of the wild-type Anellovirus clade. In some embodiments, the conserved motifs comprise one or more amino acid sequences as listed in Table 19.

In some embodiments, an ORF1 molecule or a nucleic acid molecule encoding same comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type ORF1 protein, e.g., as described herein.

*Conserved ORF1 Motif in N22 Domain*

In some embodiments, a polypeptide (e.g., an ORF1 molecule) described herein comprises the amino acid sequence $YNPX^2DXGX^2N$ (SEQ ID NO: 829), wherein $X^n$ is a contiguous sequence of any *n* amino acids. For example, $X^2$ indicates a contiguous sequence of any two amino acids. In some embodiments, the $YNPX^2DXGX^2N$ (SEQ ID NO: 829) is comprised within the N22 domain of an ORF1 molecule, e.g., as described herein. In some embodiments, a genetic element described herein comprises a nucleic acid sequence (e.g., a nucleic acid sequence encoding an ORF1 molecule, e.g., as described

herein) encoding the amino acid sequence YNPX$^2$DXGX$^2$N (SEQ ID NO: 829), wherein X$^n$ is a contiguous sequence of any *n* amino acids.

In some embodiments, a polypeptide (e.g., an ORF1 molecule) comprises a conserved secondary structure, e.g., flanking and/or comprising a portion of the YNPX$^2$DXGX$^2$N (SEQ ID NO: 829) motif, e.g., in an N22 domain. In some embodiments, the conserved secondary structure comprises a first beta strand and/or a second beta strand. In some embodiments, the first beta strand is about 5-6 (e.g., 3, 4, 5, 6, 7, or 8) amino acids in length. In some embodiments, the first beta strand comprises the tyrosine (Y) residue at the N-terminal end of the YNPX$^2$DXGX$^2$N (SEQ ID NO: 829) motif. In some embodiments, the YNPX$^2$DXGX$^2$N (SEQ ID NO: 829) motif comprises a random coil (e.g., about 8-9 amino acids of random coil). In some embodiments, the second beta strand is about 7-8 (e.g., 5, 6, 7, 8, 9, or 10) amino acids in length. In some embodiments, the second beta strand comprises the asparagine (N) residue at the C-terminal end of the YNPX$^2$DXGX$^2$N (SEQ ID NO: 829) motif.

Exemplary YNPX$^2$DXGX$^2$N (SEQ ID NO: 829) motif-flanking secondary structures are described in Example 47 and Figure 48 of PCT/US19/65995; incorporated herein by reference in its entirety. In some embodiments, an ORF1 molecule comprises a region comprising one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or all) of the secondary structural elements (e.g., beta strands) shown in Figure 48 of PCT/US19/65995. In some embodiments, an ORF1 molecule comprises a region comprising one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or all) of the secondary structural elements (e.g., beta strands) shown in Figure 48 of PCT/US19/65995, flanking a YNPX$^2$DXGX$^2$N (SEQ ID NO: 829) motif (e.g., as described herein).

*Conserved Secondary Structural Motif in ORF1 Jelly-Roll Domain*

In some embodiments, a polypeptide (e.g., an ORF1 molecule) described herein comprises one or more secondary structural elements comprised by an Anellovirus ORF1 protein (e.g., as described herein). In some emboiments, an ORF1 molecule comprises one or more secondary structural elements comprised by the jelly-roll domain of an Anellovius ORF1 protein (e.g., as described herein). Generally, an ORF1 jelly-roll domain comprises a secondary structure comprising, in order in the N-terminal to C-terminal direction, a first beta strand, a second beta strand, a first alpha helix, a third beta strand, a fourth beta strand, a fifth beta strand, a second alpha helix, a sixth beta strand, a seventh beta strand, an eighth beta strand, and a ninth beta strand. In some embodiments, an ORF1 molecule comprises a secondary structure comprising, in order in the N-terminal to C-terminal direction, a first beta strand, a second beta strand, a first alpha helix, a third beta strand, a fourth beta strand, a fifth beta strand, a second alpha helix, a sixth beta strand, a seventh beta strand, an eighth beta strand, and/or a ninth beta strand.

In some embodiments, a pair of the conserved secondary structural elements (i.e., the beta strands and/or alpha helices) are separated by an interstitial amino acid sequence, e.g., comprising a random coil sequence, a beta strand, or an alpha helix, or a combination thereof. Interstitial amino acid sequences between the conserved secondary structural elements may comprise, for example, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more amino acids. In some embodiments, an ORF1 molecule may further comprise one or more additional beta strands and/or alpha helices (e.g., in the jelly-roll domain). In some embodiments, consecutive beta strands or consecutive alpha helices may be combined. In some embodiments, the first beta strand and the second beta strand are comprised in a larger beta strand. In some embodiments, the third beta strand and the fourth beta strand are comprised in a larger beta strand. In some embodiments, the fourth beta strand and the fifth beta strand are comprised in a larger beta strand. In some embodiments, the sixth beta strand and the seventh beta strand are comprised in a larger beta strand. In some embodiments, the seventh beta strand and the eighth beta strand are comprised in a larger beta strand. In some embodiments, the eighth beta strand and the ninth beta strand are comprised in a larger beta strand.

In some embodiments, the first beta strand is about 5-7 (e.g., 3, 4, 5, 6, 7, 8, 9, or 10) amino acids in length. In some embodiments, the second beta strand is about 15-16 (e.g., 13, 14, 15, 16, 17, 18, or 19) amino acids in length. In some embodiments, the first alpha helix is about 15-17 (e.g., 13, 14, 15, 16, 17, 18, 19, or 20) amino acids in length. In some embodiments, the third beta strand is about 3-4 (e.g., 1, 2, 3, 4, 5, or 6) amino acids in length. In some embodiments, the fourth beta strand is about 10-11 (e.g., 8, 9, 10, 11, 12, or 13) amino acids in length. In some embodiments, the fifth beta strand is about 6-7 (e.g., 4, 5, 6, 7, 8, 9, or 10) amino acids in length. In some embodiments, the second alpha helix is about 8-14 (e.g., 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, or 17) amino acids in length. In some embodiments, the second alpha helix may be broken up into two smaller alpha helices (e.g., separated by a random coil sequence). In some embodiments, each of the two smaller alpha helices are about 4-6 (e.g., 2, 3, 4, 5, 6, 7, or 8) amino acids in length. In some embodiments, the sixth beta strand is about 4-5 (e.g., 2, 3, 4, 5, 6, or 7) amino acids in length. In some embodiments, the seventh beta strand is about 5-6 (e.g., 3, 4, 5, 6, 7, 8, or 9) amino acids in length. In some embodiments, the eighth beta strand is about 7-9 (e.g., 5, 6, 7, 8, 9, 10, 11, 12, or 13) amino acids in length. In some embodiments, the ninth beta strand is about 5-7 (e.g., 3, 4, 5, 6, 7, 8, 9, or 10) amino acids in length.

Exemplary jelly-roll domain secondary structures are described in Example 47 and Figure 47 of PCT/US19/65995. In some embodiments, an ORF1 molecule comprises a region comprising one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or all) of the secondary structural elements (e.g., beta strands and/or alpha helices) of any of the jelly-roll domain secondary structures shown in Figure 47 of PCT/US19/65995.

*Consensus ORF1 Domain Sequences*

In some embodiments, an ORF1 molecule, e.g., as described herein, comprises one or more of a jelly-roll domain, N22 domain, and/or C-terminal domain (CTD). In some embodiments, the jelly-roll domain comprises an amino acid sequence having a jelly-roll domain consensus sequence as described herein (e.g., as listed in any of Tables 37A-37C). In some embodiments, the N22 domain comprises an amino acid sequence having a N22 domain consensus sequence as described herein (e.g., as listed in any of Tables 37A-37C). In some embodiments, the CTD domain comprises an amino acid sequence having a CTD domain consensus sequence as described herein (e.g., as listed in any of Tables 37A-37C). In some embodiments, the amino acids listed in any of Tables 37A-37C in the format "($X_{a-b}$)" comprise a contiguous series of amino acids, in which the series comprises at least *a*, and at most *b*, amino acids. In certain embodiments, all of the amino acids in the series are identical. In other embodiments, the series comprises at least two (e.g., at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 21) different amino acids.

**Table 37A. Alphatorquevius ORF1 domain consensus sequences**

| Domain | Sequence | SEQ ID NO: |
|---|---|---|
| Jelly-Roll | LVLTQWQPNTVRRCYIRGYLPLIICGEN($X_{0-3}$)TTSRNYATHS DDTIQKGPFGGGMSTTTFSLRVLYDEYQRFMNRWTYSNED LDLARYLGCKFTFYRHPDXDFIVQYNTNPPFKDTKLTAPSIH P($X_{1-5}$)GMLMLSKRKILIPSLKTRPKGKHYVKVRIGPPKLFED KWYTQSDLCDVPLVXLYATAADLQHPFGSPQTDNPCVTFQ VLGSXYNKHLSISP;<br><br>wherein X = any amino acid. | 227 |
| N22 | SNFEFPGAYTDITYNPLTDKGVGNMVWIQYLTKPDTIXDKT QS($X_{0-3}$)KCLIEDLPLWAALYGYVDFCEKETGDSAIIXNXGRV LIRCPYTKPPLYDKT($X_{0-4}$)NKGFVPYSTNFGNGKMPGGSGY VPIYWRARWYPTLFHQKEVLEDIVQSGPFAYKDEKPSTQLV MKYCFNFN;<br><br>wherein X = any amino acid. | 228 |
| CTD | WGGNPISQQVVRNPCKDSG($X_{0-3}$)SGXGRQPRSVQVVDPKY MGPEYTFHSWDWRRGLFGEKAIKRMSEQPTDDEIFTGGXPK RPRRDPPTXQXPEE($X_{1-4}$)QKESSSFR($X_{2-14}$)PWESSSQEXESES | 229 |

| | | |
|---|---|---|
| | QEEEE($X_{0-30}$)EQTVQQQLRQQLREQRRLRVQLQLLFQQLLKT($X_{0-4}$)QAGLHINPLLLSQA($X_{0-40}$)*;<br><br>wherein X = any amino acid. | |

**Table 37B. Betatorquevius ORF1 domain consensus sequences**

| Domain | Sequence | SEQ ID NO: |
|---|---|---|
| Jelly-Roll | LKQWQPSTIRKCKIKGYLPLFQCGKGRISNNYTQYKESIVPHHEPGGGGWSIQQFTLGALYEEHLKLRNWWTKSNDGLPLVRYLGCTIKLYRSEDTDYIVTYQRCYPMTATKLTYLSTQPSRMLMNKHKIIVPSKXT($X_{1-4}$)NKKKKPYKKIFIKPPSQMQNKWYFQQDIANTPLLQLTXTACSLDRMYLSSDSISNNITFTSLNTNFFQNPNFQ;<br><br>wherein X = any amino acid. | 230 |
| N22 | ($X_{4-10}$)TPLYFECRYNPFKDKGTGNKVYLVSNN($X_{1-8}$)TGWDPPTDPDLIIEGFPLWLLLWGWLDWQKKLGKIQNIDTDYILVIQSXYYIPP($X_{1-3}$)KLPYYVPLDXD($X_{0-2}$)FLHGRSPY($X_{3-16}$)PSDKQHWHPKVRFQXETINNIALTGPGTPKLPNQKSIQAHMKYKFYFK;<br><br>wherein X = any amino acid. | 231 |
| CTD | WGGCPAPMETITDPCKQPKYPIPNNLLQTTSLQXPTTPIETYLYKFDERRGLLTKKAAKRIKKDXTTETTLFTDTGXXTSTTLPTXXQTETTQEEXTSEEE($X_{0-5}$)ETLLQQLQQLRRKQKQLRXRILQLLQLLXLL($X_{0-26}$)*;<br><br>wherein X = any amino acid. | 232 |

**Table 37C. Gammatorquevius ORF1 domain consensus sequences**

| Domain | Sequence | SEQ ID NO: |
|---|---|---|
| Jelly-Roll | TIPLKQWQPESIRKCKIKGYGTLVLGAEGRQFYCYTNEKDEYTPPKAPGGGGFGVELFSLEYLYEQWKARNNIWTKSNXYK | 233 |

| | | |
|---|---|---|
| | DLCRYTGCKITFYRHPTTDFIVXYSRQPPFEIDKXTYMXXHP QXLLLRKHKKIILSKATNPKGKLKKKIKIKPPKQMLNKWFF QKQFAXYGLVQLQAAACBLRYPRLGCCNENRLITLYYLN;<br><br>wherein X = any amino acid. | |
| N22 | LPIVVARYNPAXDTGKGNKXWLXSTLNGSXWAPPTTDKDL IIEGLPLWLALYGYWSYJKKVKKDKGILQSHMFVVKSPAIQP LXTATTQXTFYPXIDNSFIQGKXPYDEPJTXNQKKLWYPTLE HQQETINAIVESGPYVPKLDNQKNSTWELXYXYTFYFK;<br><br>wherein X = any amino acid. | 234 |
| CTD | WGGPQIPDQPVEDPKXQGTYPVPDTXQQTIQIXNPLKQKPE TMFHDWDYRRGIITSTALKRMQENLETDSSFXSDSEETP($X_{0-2}$ )KKKKRLTXELPXPQEETEEIQSCLLSLCEESTCQEE($X_{1-6}$)ENL QQLIHQQQQQQQLKHNILKLLSDLKZKQRLLQLQTGILE($X_{1-10}$)*;<br><br>wherein X = any amino acid. | 235 |

In some embodiments, the jelly-roll domain comprises a jelly-roll domain amino acid sequence as listed in any of Tables 21, 23, 25, 27, 29, 31, 33, 35, D2, D4, D6, D8, D10, or 37A-37C, or an amino acid sequence having at least 70%, 75%, 80%, 8%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. In some embodiments, the N22 domain comprises a N22 domain amino acid sequence as listed in any of Tables 21, 23, 25, 27, 29, 31, 33, 35, D2, D4, D6, D8, D10, or 37A-37C, or an amino acid sequence having at least 70%, 75%, 80%, 8%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto. In some embodiments, the CTD domain comprises a CTD domain amino acid sequence as listed in any of Tables 21, 23, 25, 27, 29, 31, 33, 35, D2, D4, D6, D8, D10, or 37A-37C, or an amino acid sequence having at least 70%, 75%, 80%, 8%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity thereto.

*Identification of ORF1 protein sequences*

In some embodiments, an Anellovirus ORF1 protein sequence, or a nucleic acid sequence encoding an ORF1 protein, can be identified from the genome of an *Anellovirus* (e.g., a putative *Anellovirus* genome identified, for example, by nucleic acid sequencing techniques, e.g., deep sequencing

techniques). In some embodiments, an ORF1 protein sequence is identified by one or more (e.g., 1, 2, or all 3) of the following selection criteria:

*(i) Length Selection:* Protein sequences (e.g., putative Anellovirus ORF1 sequences passing the criteria described in (ii) or (iii) below) may be size-selected for those greater than about 600 amino acid residues to identify putative Anellovirus ORF1 proteins. In some embodiments, an Anellovirus ORF1 protein sequence is at least about 600, 650, 700, 750, 800, 850, 900, 950, or 1000 amino acid residues in length. In some embodiments, an Alphatorquevirus ORF1 protein sequence is at least about 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 900, or 1000 amino acid residues in length. In some embodiments, a Betatorquevirus ORF1 protein sequence is at least about 650, 660, 670, 680, 690, 700, 750, 800, 900, or 1000 amino acid residues in length. In some embodiments, a Gammatorquevirus ORF1 protein sequence is at least about 650, 660, 670, 680, 690, 700, 750, 800, 900, or 1000 amino acid residues in length. In some embodiments, a nucleic acid sequence encoding an Anellovirus ORF1 protein is at least about 1800, 1900, 2000, 2100, 2200, 2300, 2400, or 2500 nucleotides in length. In some embodiments, a nucleic acid sequence encoding an Alphatorquevirus ORF1 protein sequence is at least about 2100, 2150, 2200, 2250, 2300, 2400, or 2500 nucleotides in length. In some embodiments, a nucleic acid sequence encoding a Betatorquevirus ORF1 protein sequence is at least about 1900, 1950, 2000, 2500, 2100, 2150, 2200, 2250, 2300, 2400, or 2500 or 1000 nucleotides in length. In some embodiments, a nucleic acid sequence encoding a Gammatorquevirus ORF1 protein sequence is at least about 1900, 1950, 2000, 2500, 2100, 2150, 2200, 2250, 2300, 2400, or 2500 or 1000 nucleotides in length.

*(ii) Presence of ORF1 motif:* Protein sequences (e.g., putative Anellovirus ORF1 sequences passing the criteria described in (i) above or (iii) below) may be filtered to identify those that contain the conserved ORF1 motif in the N22 domain described above. In some embodiments, a putative Anellovirus ORF1 sequence comprises the sequence YNPXXDXGXXN. In some embodiments, a putative Anellovirus ORF1 sequence comprises the sequence Y[NCS]P*XX*D*X*[GASKR]*XX*[NTSVAK].

*(iii) Presence of arginine-rich region:* Protein sequences (e.g., putative Anellovirus ORF1 sequences passing the criteria described in (i) and/or (ii) above) may be filtered for those that include an arginine-rich region (e.g., as described herein). In some embodiments, a putative Anellovirus ORF1 sequence comprises a contiguous sequence of at least about 30, 35, 40, 45, 50, 55, 60, 65, or 70 amino acids that comprises at least 30% (e.g., at least about 20%, 25%, 30%, 35%, 40%, 45%, or 50%) arginine residues. In some embodiments, a putative Anellovirus ORF1 sequence comprises a contiguous sequence of about 35-40, 40-45, 45-50, 50-55, 55-60, 60-65, or 65-70 amino acids that comprises at least 30% (e.g., at least about 20%, 25%, 30%, 35%, 40%, 45%, or 50%) arginine residues. In some embodiments, the arginine-rich region is positioned at least about 30, 40, 50, 60, 70, or 80 amino acids downstream of the

start codon of the putative Anellovirus ORF1 protein. In some embodiments, the arginine-rich region is positioned at least about 50 amino acids downstream of the start codon of the putative Anellovirus ORF1 protein.

***ORF2 Molecules***

In some embodiments, the anellovector comprises an ORF2 molecule and/or a nucleic acid encoding an ORF2 molecule. Generally, an ORF2 molecule comprises a polypeptide having the structural features and/or activity of an Anellovirus ORF2 protein (e.g., an Anellovirus ORF2 protein as described herein, e.g., as listed in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, or 18), or a functional fragment thereof. In some embodiments, an ORF2 molecule comprises an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an Anellovirus ORF2 protein sequence as shown in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, or 18.

In some embodiments, an ORF2 molecule comprises an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity to an Alphatorquevirus, Betatorquevirus, or Gammatorquevirus ORF2 protein. In some embodiments, an ORF2 molecule (e.g., an ORF2 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity to an Alphatorquevirus ORF2 protein) has a length of 250 or fewer amino acids (e.g., about 150-200 amino acids). In some embodiments, an ORF2 molecule (e.g., an ORF2 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity to a Betatorquevirus ORF2 protein) has a length of about 50-150 amino acids. In some embodiments, an ORF2 molecule (e.g., an ORF2 molecule having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity to a Gammatorquevirus ORF2 protein) has a length of about 100-200 amino acids (e.g., about 100-150 amino acids). In some embodiments, the ORF2 molecule comprises a helix-turn-helix motif (e.g., a helix-turn-helix motif comprising two alpha helices flanking a turn region). In some embodiments, the ORF2 molecule does not comprise the amino acid sequence of the ORF2 protein of TTV isolate TA278 or TTV isolate SANBAN. In some embodiments, an ORF2 molecule has protein phosphatase activity. In some embodiments, an ORF2 molecule or a nucleic acid molecule encoding same comprises at least one difference (e.g., a mutation, chemical modification, or epigenetic alteration) relative to a wild-type ORF2 protein, e.g., as described herein (e.g., as shown in any of Tables A2, A4, A6, A8, A10, A12, C1-C5, 2, 4, 6, 8, 10, 12, 14, 16, or 18).

*Conserved ORF2 Motif*

In some embodiments, a polypeptide (e.g., an ORF2 molecule) described herein comprises the amino acid sequence [W/F]$X^7$H$X^3$C$X^1$C$X^5$H (SEQ ID NO: 949), wherein $X^n$ is a contiguous sequence of any *n* amino acids. In embodiments, $X^7$ indicates a contiguous sequence of any seven amino acids. In embodiments, $X^3$ indicates a contiguous sequence of any three amino acids. In embodiments, $X^1$ indicates any single amino acid. In embodiments, $X^5$ indicates a contiguous sequence of any five amino acids. In some embodiments, the [W/F] can be either tryptophan or phenylalanine. In some embodiments, the [W/F]$X^7$H$X^3$C$X^1$C$X^5$H (SEQ ID NO: 949) is comprised within the N22 domain of an ORF2 molecule, e.g., as described herein. In some embodiments, a genetic element described herein comprises a nucleic acid sequence (e.g., a nucleic acid sequence encoding an ORF2 molecule, e.g., as described herein) encoding the amino acid sequence [W/F]$X^7$H$X^3$C$X^1$C$X^5$H (SEQ ID NO: 949), wherein $X^n$ is a contiguous sequence of any *n* amino acids.

***Genetic Elements***

In some embodiments, the anellovector comprises a genetic element. In some embodiments, the genetic element has one or more of the following characteristics: is substantially non-integrating with a host cell's genome, is an episomal nucleic acid, is a single stranded RNA, is circular, is about 1 to 10 kb, exists within the nucleus of the cell, can be bound by endogenous proteins, produces an effector, such as a polypeptide or nucleic acid (e.g., an RNA, iRNA, microRNA) that targets a gene, activity, or function of a host or target cell. In one embodiment, the genetic element is a substantially non-integrating. In some embodiments, the genetic element comprises a packaging signal, e.g., a sequence that binds a capsid protein. In some embodiments, outside of the packaging or capsid-binding sequence, the genetic element has less than 70%, 60%, 50%, 40%, 30%, 20%, 10%, 5% sequence identity to a wild type Anellovirus nucleic acid sequence, e.g., has less than 70%, 60%, 50%, 40%, 30%, 20%, 10%, 5% sequence identity to an *Anellovirus* nucleic acid sequence, e.g., as described herein. In some embodiments, outside of the packaging or capsid-binding sequence, the genetic element has less than 500 450, 400, 350, 300, 250, 200, 150, or 100 contiguous nucleotides that are at least 70%, 75%, 80%, 8%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to an *Anellovirus* nucleic acid sequence.

In some embodiments, the genetic element has a length less than 20kb (e.g., less than about 19kb, 18kb, 17kb, 16kb, 15kb, 14kb, 13kb, 12kb, 11kb, 10kb, 9kb, 8kb, 7kb, 6kb, 5kb, 4kb, 3kb, 2kb, 1kb, or less). In some embodiments, the genetic element has, independently or in addition to, a length greater than 1000b (e.g., at least about 1.1kb, 1.2kb, 1.3kb, 1.4kb, 1.5kb, 1.6kb, 1.7kb, 1.8kb, 1.9kb, 2kb, 2.1kb, 2.2kb, 2.3kb, 2.4kb, 2.5kb, 2.6kb, 2.7kb, 2.8kb, 2.9kb, 3kb, 3.1kb, 3.2kb, 3.3kb, 3.4kb, 3.5kb, 3.6kb, 3.7kb, 3.8kb, 3.9kb, 4kb, 4.1kb, 4.2kb, 4.3kb, 4.4kb, 4.5kb, 4.6kb, 4.7kb, 4.8kb, 4.9kb, 5kb, or greater).

In some embodiments, the genetic element has a length of about 2.5-4.6, 2.8-4.0, 3.0-3.8, or 3.2-3.7 kb. In some embodiments, the genetic element has a length of about 1.5-2.0, 1.5-2.5, 1.5-3.0, 1.5-3.5, 1.5-3.8, 1.5-3.9, 1.5-4.0, 1.5-4.5, or 1.5-5.0 kb. In some embodiments, the genetic element has a length of about 2.0-2.5, 2.0-3.0, 2.0-3.5, 2.0-3.8, 2.0-3.9, 2.0-4.0, 2.0-4.5, or 2.0-5.0 kb. In some embodiments, the genetic element has a length of about 2.5-3.0, 2.5-3.5, 2.5-3.8, 2.5-3.9, 2.5-4.0, 2.5-4.5, or 2.5-5.0 kb. In some embodiments, the genetic element has a length of about 3.0-5.0, 3.5-5.0, 4.0-5.0, or 4.5-5.0 kb. In some embodiments, the genetic element has a length of about 1.5-2.0, 2.0-2.5, 2.5-3.0, 3.0-3.5, 3.1-3.6, 3.2-3.7, 3.3-3.8, 3.4-3.9, 3.5-4.0, 4.0-4.5, or 4.5-5.0 kb. In some embodiments, the genetic element has a length between about 3.6-3.9 kb. In some embodiments, the genetic element has a length between about 2.8-2.9 kb. In some embodiments, the genetic element has a length between about 2.0-3.2 kb.

In some embodiments, the genetic element comprises one or more of the features described herein, e.g., a sequence encoding a substantially non-pathogenic protein, a protein binding sequence, one or more sequences encoding a regulatory nucleic acid, one or more regulatory sequences, one or more sequences encoding a replication protein, and other sequences.

In embodiments, the genetic element was produced from a double-stranded circular DNA (e.g., produced by transcription).

In some embodiments, the genetic element does not comprise one or more bacterial plasmid elements (e.g., a bacterial origin of replication or a selectable marker, e.g., a bacterial resistance gene). In some embodiments, the genetic element does not comprise a bacterial plasmid backbone.

In one embodiment, the disclosure provides a genetic element comprising a nucleic acid sequence (e.g., an RNA sequence) encoding (i) a substantially non-pathogenic exterior protein, (ii) an exterior protein binding sequence that binds the genetic element to the substantially non-pathogenic exterior protein, and (iii) a regulatory nucleic acid. In such an embodiment, the genetic element may comprise one or more sequences with at least about 60%, 70% 80%, 85%, 90% 95%, 96%, 97%, 98% and 99% nucleotide sequence identity to any one of the nucleotide sequences to a native viral sequence (e.g., a native *Anellovirus* sequence, e.g., as described herein).

***Protein Binding Sequence***

In some embodiments, the genetic element encodes a protein binding sequence that binds to the substantially non-pathogenic protein. In some embodiments, the protein binding sequence facilitates packaging the genetic element into the proteinaceous exterior. In some embodiments, the protein binding sequence specifically binds an arginine-rich region of the substantially non-pathogenic protein. In some embodiments, the genetic element comprises a protein binding sequence as described in Example 8 of PCT/US19/65995.

In some embodiments, the genetic element comprises a protein binding sequence having at least 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to a 5' UTR conserved domain or GC-rich domain of an *Anellovirus* sequence, e.g., as described herein.

In embodiments, the protein binding sequence has at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an *Anellovirus* 5' UTR conserved domain nucleotide sequence, e.g., as described herein.

***5' UTR Regions***

In some embodiments, a nucleic acid molecule as described herein (e.g., a genetic element, genetic element construct, or genetic element region) comprises a 5' UTR sequence, e.g., a 5' UTR conserved domain sequence as described herein (e.g., in any of Tables A1, B1, or C1), or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

In some embodiments, the 5' UTR sequence comprises the nucleic acid sequence AGGTGAGTGAAACCACCGAAGTCAAGGGGCAATTCGGGCTAGGGX$_1$CAGTCT, or a nucleic acid sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the 5' UTR sequence comprises the nucleic acid sequence AGGTGAGTGAAACCACCGAAGTCAAGGGGCAATTCGGGCTAGGGX$_1$CAGTCT, or a nucleic acid sequence having no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotide differences (e.g., substitutions, deletions, or additions) relative thereto. In embodiments, $X_1$ is A. In embodiments, $X_1$ is absent.

In some embodiments, the 5' UTR sequence comprises the nucleic acid sequence of the 5' UTR of an Alphatorquevirus (e.g., Ring1), or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In embodiments, the 5' UTR sequence comprises the nucleic acid sequence of the 5' UTR conserved domain listed in Table A1, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 95% sequence identity to the 5' UTR conserved domain listed in Table A1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 95.775% sequence identity to the 5' UTR conserved domain listed in Table A1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 97% sequence identity to the 5' UTR conserved domain listed in Table A1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 97.183% sequence identity to the 5' UTR conserved domain listed in Table A1. In some embodiments, the 5' UTR sequence comprises the nucleic acid sequence AGGTGAGTTTACACACCGCAGTCAAGGGGCAATTCGGGCTCGGGACTGGC, or a nucleic acid sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some

embodiments, the 5' UTR sequence comprises the nucleic acid sequence AGGTGAGTTTACACACCGCAGTCAAGGGGCAATTCGGGCTCGGGACTGGC, or a nucleic acid sequence having no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotide differences (e.g., substitutions, deletions, or additions) relative thereto.

In some embodiments, the 5' UTR sequence comprises the nucleic acid sequence of the 5' UTR of an Betatorquevirus (e.g., Ring2), or a sequence having at least 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In embodiments, the 5' UTR sequence comprises the nucleic acid sequence of the 5' UTR conserved domain listed in Table B1, or a sequence having at least 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 85% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 87% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 87.324% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 88% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 88.732% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 91% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 91.549% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 92% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 92.958% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 94% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 94.366% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 95% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 95.775% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 97% sequence identity to the 5' UTR conserved domain listed in

Table B1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 97.183% sequence identity to the 5' UTR conserved domain listed in Table B1. In some embodiments, the 5' UTR sequence comprises the nucleic acid sequence AGGTGAGTGAAACCACCGAAGTCAAGGGGCAATTCGGGCTAGATCAGTCT, or a nucleic acid sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the 5' UTR sequence comprises the nucleic acid sequence AGGTGAGTGAAACCACCGAAGTCAAGGGGCAATTCGGGCTAGATCAGTCT, or a nucleic acid sequence having no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotide differences (e.g., substitutions, deletions, or additions) relative thereto.

In some embodiments, the 5' UTR sequence comprises the nucleic acid sequence of the 5' UTR of a Gammatorquevirus (e.g., Ring4), or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In embodiments, the 5' UTR sequence comprises the nucleic acid sequence of the 5' UTR conserved domain listed in Table C1, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 97% sequence identity to the 5' UTR conserved domain listed in Table C1. In some embodiments, the nucleic acid molecule comprises a nucleic acid sequence having at least 97.183% sequence identity to the 5' UTR conserved domain listed in Table C1. In some embodiments, the 5' UTR sequence comprises the nucleic acid sequence AGGTGAGTGAAACCACCGAGGTCTAGGGGCAATTCGGGCTAGGGCAGTCT, or a nucleic acid sequence having at least 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto. In some embodiments, the 5' UTR sequence comprises the nucleic acid sequence AGGTGAGTGAAACCACCGAGGTCTAGGGGCAATTCGGGCTAGGGCAGTCT, or a nucleic acid sequence having no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotide differences (e.g., substitutions, deletions, or additions) relative thereto.

In some embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to an Anellovirus 5' UTR sequence, e.g., a nucleic acid sequence shown in Table 38. In some embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence of the Consensus 5' UTR sequence shown in Table 38, wherein $X_1$, $X_2$, $X_3$, $X_4$, and $X_5$ are each independently any nucleotide, e.g., wherein $X_1$ = G or T, $X_2$ = C or A, $X_3$ = G or A, $X_4$ = T or C, and $X_5$ = A, C, or T). In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the Consensus 5' UTR sequence shown in Table 38. In embodiments, the genetic element (e.g., protein-

binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the exemplary TTV 5' UTR sequence shown in Table 38. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the TTV-CT30F 5' UTR sequence shown in Table 38. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the TTV-HD23a 5' UTR sequence shown in Table 38. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the TTV-JA20 5' UTR sequence shown in Table 38. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the TTV-TJN02 5' UTR sequence shown in Table 38. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the TTV-tth8 5' UTR sequence shown in Table 38.

In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the Alphatorquevirus Consensus 5' UTR sequence shown in Table 38. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the Alphatorquevirus Clade 1 5' UTR sequence shown in Table 38. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the Alphatorquevirus Clade 2 5' UTR sequence shown in Table 38. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the Alphatorquevirus Clade 3 5' UTR sequence shown in Table 38. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the Alphatorquevirus Clade 4 5' UTR sequence shown in Table 38. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%,

96%, 97%, 98%, 99%, or 100%) identity to the Alphatorquevirus Clade 5 5' UTR sequence shown in Table 38. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the Alphatorquevirus Clade 6 5' UTR sequence shown in Table 38. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to the Alphatorquevirus Clade 7 5' UTR sequence shown in Table 38.

**Table 38. Exemplary 5' UTR sequences from *Anelloviruses***

| Source | Sequence | SEQ ID NO: |
|---|---|---|
| Consensus | CGGGTGCCGX$_1$AGGTGAGTTTACACACCGX$_2$AGT CAAGGGGCAATTCGGGCTCX$_3$GGACTGGCCGGG CX$_4$X$_5$TGGG<br>X$_1$ = G or T<br>X$_2$ = C or A<br>X$_3$ = G or A<br>X$_4$ = T or C<br>X$_5$ = A, C, or T | 105 |
| Exemplary TTV Sequence | CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCT WTGGG | 106 |
| TTV-CT30F | CGGGTGCCGTAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCT ATGGG | 107 |
| TTV-HD23a | CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCC CTGGG | 108 |
| TTV-JA20 | CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCT TTGGG | 109 |

| | | |
|---|---|---|
| TTV-TJN02 | CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCT ATGGG | 110 |
| TTV-tth8 | CGGGTGCCGGAGGTGAGTTTACACACCGAAGTC AAGGGGCAATTCGGGCTCAGGACTGGCCGGGCT TTGGG | 111 |
| Alphatorquevirus Consensus 5' UTR | CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGC $X_1X_2$TGGG; wherein $X_1$ comprises T or C, and wherein $X_2$ comprises A, C, or T. | 112 |
| Alphatorquevirus Clade 1 5' UTR (e.g., TTV-CT30F) | CGGGTGCCGTAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCT ATGGG | 113 |
| Alphatorquevirus Clade 2 5' UTR (e.g., TTV-P13-1) | CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCC CGGG | 114 |
| Alphatorquevirus Clade 3 5' UTR (e.g., TTV-tth8) | CGGGTGCCGGAGGTGAGTTTACACACCGAAGTC AAGGGGCAATTCGGGCTCAGGACTGGCCGGGCT TTGGG | 115 |
| Alphatorquevirus Clade 4 5' UTR (e.g., TTV-HD20a) | CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGAGGCCGGGCCAT GGG | 116 |
| Alphatorquevirus Clade 5 5' UTR (e.g., TTV-16) | CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCC CCGGG | 117 |
| Alphatorquevirus Clade 6 5' UTR (e.g., TTV-TJN02) | CGGGTGCCGGAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCT ATGGG | 118 |
| Alphatorquevirus Clade 7 5' UTR (e.g., TTV-HD16d) | CGGGTGCCGAAGGTGAGTTTACACACCGCAGTC AAGGGGCAATTCGGGCTCGGGACTGGCCGGGCT ATGGG | 119 |

*Identification of 5' UTR sequences*

In some embodiments, an Anellovirus 5' UTR sequence can be identified within the genome of an *Anellovirus* (e.g., a putative *Anellovirus* genome identified, for example, by nucleic acid sequencing techniques, e.g., deep sequencing techniques). In some embodiments, an Anellovirus 5' UTR sequence is identified by one or both of the following steps:

*(i) Identification of circularization junction point:* In some embodiments, a 5' UTR will be positioned near a circularization junction point of a full-length, circularized *Anellovirus* genome. A circularization junction point can be identified, for example, by identifying overlapping regions of the sequence. In some embodiments, a overlapping region of the sequence can be trimmed from the sequence to produce a full-length *Anellovirus* genome sequence that has been circularized. In some embodiments, a genome sequence is circularized in this manner using software. Without wishing to be bound by theory, computationally circularizing a genome may result in the start position for the sequence being oriented in a non-biological. Landmarks within the sequence can be used to re-orient sequences in the proper direction. For example, landmark sequence may include sequences having substantial homology to one or more elements within an Anellovirus genome as described herein (e.g., one or more of a TATA box, cap site, initiator element, transcriptional start site, 5' UTR conserved domain, ORF1, ORF1/1, ORF1/2, ORF2, ORF2/2, ORF2/3, ORF2t/3, three open-reading frame region, poly(A) signal, or GC-rich region of an *Anellovirus*, e.g., as described herein).

*(ii) Identification of 5' UTR sequence:* Once a putative *Anellovirus* genome sequence has been obtained, the sequence (or portions thereof, e.g., having a length between about 40-50, 50-60, 60-70, 70-80, 80-90, or 90-100 nucleotides) can be compared to one or more Anellovirus 5' UTR sequences (e.g., as described herein) to identify sequences having substantial homology thereto. In some embodiments, a putative *Anellovirus* 5' UTR region has at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an *Anellovirus* 5' UTR sequence as described herein.

***GC-Rich Regions***

In some embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to a nucleic acid sequence shown in Table 39. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to a GC-rich sequence shown in Table 39.

In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to a 36-nucleotide GC-rich sequence as shown in Table 39 (e.g., 36-nucleotide consensus GC-rich region sequence 1, 36-nucleotide consensus GC-rich region sequence 2, TTV Clade 1 36-nucleotide region, TTV Clade 3 36-nucleotide region, TTV Clade 3 isolate GH1 36-nucleotide region, TTV Clade 3 sle1932 36-nucleotide region, TTV Clade 4 ctdc002 36-nucleotide region, TTV Clade 5 36-nucleotide region, TTV Clade 6 36-nucleotide region, or TTV Clade 7 36-nucleotide region). In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence comprising at least 10, 15, 20, 25, 30, 31, 32, 33, 34, 35, or 36 consecutive nucleotides of a 36-nucleotide GC-rich sequence as shown in Table 39 (e.g., 36-nucleotide consensus GC-rich region sequence 1, 36-nucleotide consensus GC-rich region sequence 2, TTV Clade 1 36-nucleotide region, TTV Clade 3 36-nucleotide region, TTV Clade 3 isolate GH1 36-nucleotide region, TTV Clade 3 sle1932 36-nucleotide region, TTV Clade 4 ctdc002 36-nucleotide region, TTV Clade 5 36-nucleotide region, TTV Clade 6 36-nucleotide region, or TTV Clade 7 36-nucleotide region).

In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to an Alphatorquevirus GC-rich region sequence, e.g., selected from TTV-CT30F, TTV-P13-1, TTV-tth8, TTV-HD20a, TTV-16, TTV-TJN02, or TTV-HD16d, e.g., as listed in Table 39. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence comprising at least 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 104, 105, 108, 110, 111, 115, 120, 122, 130, 140, 145, 150, 155, or 156 consecutive nucleotides of an Alphatorquevirus GC-rich region sequence, e.g., selected from TTV-CT30F, TTV-P13-1, TTV-tth8, TTV-HD20a, TTV-16, TTV-TJN02, or TTV-HD16d, e.g., as listed in Table 39.

In embodiments, the 36-nucleotide GC-rich sequence is selected from:

(i) CGCGCTGCGCGCGCCGCCCAGTAGGGGGAGCCATGC (SEQ ID NO: 160),

(ii) GCGCTX$_1$CGCGCGCGCGCCGGGGGGCTGCGCCCCCCC (SEQ ID NO: 164),

wherein $X_1$ is selected from T, G, or A;

(iii) GCGCTTCGCGCGCCGCCCACTAGGGGGCGTTGCGCG (SEQ ID NO: 165);

(iv) GCGCTGCGCGCGCCGCCCAGTAGGGGGCGCAATGCG (SEQ ID NO: 166);

(v) GCGCTGCGCGCGCGGCCCCCGGGGGAGGCATTGCCT (SEQ ID NO: 167);

(vi) GCGCTGCGCGCGCGCGCCGGGGGGGCGCCAGCGCCC (SEQ ID NO: 168);

(vii) GCGCTTCGCGCGCGCGCCGGGGGGCTCCGCCCCCCC (SEQ ID NO: 169);

(viii) GCGCTTCGCGCGCGCGCCGGGGGGCTGCGCCCCCCC (SEQ ID NO: 170);

(ix) GCGCTACGCGCGCGCGCCGGGGGGCTGCGCCCCCCC (SEQ ID NO: 171); or

(x) GCGCTACGCGCGCGCCGGGGGGCTCTGCCCCCCC (SEQ ID NO: 172).

In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises the nucleic acid sequence CGCGCTGCGCGCGCCGCCCAGTAGGGGGAGCCATGC (SEQ ID NO: 160).

In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence of the Consensus GC-rich sequence shown in Table 39, wherein $X_1$, $X_4$, $X_5$, $X_6$, $X_7$, $X_{12}$, $X_{13}$, $X_{14}$, $X_{15}$, $X_{20}$, $X_{21}$, $X_{22}$, $X_{26}$, $X_{29}$, $X_{30}$, and $X_{33}$ are each independently any nucleotide and wherein $X_2$, $X_3$, $X_8$, $X_9$, $X_{10}$, $X_{11}$, $X_{16}$, $X_{17}$, $X_{18}$, $X_{19}$, $X_{23}$, $X_{24}$, $X_{25}$, $X_{27}$, $X_{28}$, $X_{31}$, $X_{32}$, and $X_{34}$ are each independently absent or any nucleotide. In some embodiments, one or more of (e.g., all of) $X_1$ through $X_{34}$ are each independently the nucleotide (or absent) specified in Table 39. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to an exemplary TTV GC-rich sequence shown in Table 39 (e.g., the full sequence, Fragment 1, Fragment 2, Fragment 3, or any combination thereof, e.g., Fragments 1-3 in order). In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to a TTV-CT30F GC-rich sequence shown in Table 39 (e.g., the full sequence, Fragment 1, Fragment 2, Fragment 3, Fragment 4, Fragment 5, Fragment 6, Fragment 7, Fragment 8, or any combination thereof, e.g., Fragments 1-7 in order). In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to a TTV-HD23a GC-rich sequence shown in Table 39 (e.g., the full sequence, Fragment 1, Fragment 2, Fragment 3, Fragment 4, Fragment 5, Fragment 6, or any combination thereof, e.g., Fragments 1-6 in order). In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to a TTV-JA20 GC-rich sequence shown in Table 39 (e.g., the full sequence, Fragment 1, Fragment 2, or any combination thereof, e.g., Fragments 1 and 2 in order). In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to a TTV-TJN02 GC-rich sequence shown in Table 39 (e.g., the full sequence, Fragment 1, Fragment 2, Fragment 3, Fragment 4, Fragment 5, Fragment 6, Fragment 7, Fragment 8, or any combination thereof, e.g., Fragments 1-8 in order). In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to

a TTV-tth8 GC-rich sequence shown in Table 39 (e.g., the full sequence, Fragment 1, Fragment 2, Fragment 3, Fragment 4, Fragment 5, Fragment 6, Fragment 7, Fragment 8, Fragment 9, or any combination thereof, e.g., Fragments 1-6 in order). In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to Fragment 7 shown in Table 39. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to Fragment 8 shown in Table 39. In embodiments, the genetic element (e.g., protein-binding sequence of the genetic element) comprises a nucleic acid sequence having at least about 75% (e.g., at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to Fragment 9 shown in Table 39.

**Table 39. Exemplary GC-rich sequences from *Anelloviruses***

| **Source** | **Sequence** | **SEQ ID NO:** |
|---|---|---|
| Consensus | $CGGCGGGX_1GGX_2GX_3X_4X_5CGCGCTX_6CGCGC$ $GCX_7X_8X_9X_{10}CX_{11}X_{12}X_{13}X_{14}GGGGX_{15}X_{16}X_{17}X_{18}$ $X_{19}X_{20}X_{21}GCX_{22}X_{23}X_{24}X_{25}CCCCCCCCX_{26}CGCGC$ $ATX_{27}X_{28}GCX_{29}CGGGX_{30}CCCCCCCCCCX_{31}X_{32}X$ $_{33}GGGGGGCTCCGX_{34}CCCCCCGGCCCCCC$<br><br>$X_1$ = G or C<br>$X_2$ = G, C, or absent<br>$X_3$ = C or absent<br>$X_4$ = G or C<br>$X_5$ = G or C<br>$X_6$ = T, G, or A<br>$X_7$ = G or C<br>$X_8$ = G or absent<br>$X_9$ = C or absent<br>$X_{10}$ = C or absent<br>$X_{11}$ = G, A, or absent<br>$X_{12}$ = G or C<br>$X_{13}$ = C or T | 120 |

| | | | |
|---|---|---|---|
| | | $X_{14}$ = G or A<br>$X_{15}$ = G or A<br>$X_{16}$ = A, G, T, or absent<br>$X_{17}$ = G, C, or absent<br>$X_{18}$ = G, C, or absent<br>$X_{19}$ = C, A, or absent<br>$X_{20}$ = C or A<br>$X_{21}$ = T or A<br>$X_{22}$ = G or C<br>$X_{23}$ = G, T, or absent<br>$X_{24}$ = C or absent<br>$X_{25}$ = G, C, or absent<br>$X_{26}$ = G or C<br>$X_{27}$ = G or absent<br>$X_{28}$ = C or absent<br>$X_{29}$ = G or A<br>$X_{30}$ = G or T<br>$X_{31}$ = C, T, or absent<br>$X_{32}$ = G, C, A, or absent<br>$X_{33}$ = G or C<br>$X_{34}$ = C or absent | |
| Exemplary TTV Sequence | Full sequence | GCCGCCGCGGCGGCGGSGGNGNSGCGCGCT DCGCGCGCSNNNCRCCRGGGGGNNNNCWG CSNCNCCCCCCCCCGCGCATGCGCGGGKCC CCCCCCCNNCGGGGGGCTCCGCCCCCCGGC CCCCCCCCGTGCTAAACCCACCGCGCATGC GCGACCACGCCCCCGCCGCC | 121 |
| | Fragment 1 | GCCGCCGCGGCGGCGGSGGNGNSGCGCGCT DCGCGCGCSNNNCRCCRGGGGGNNNNCWG CSNCNCCCCCCCCCGCGCAT | 122 |
| | Fragment 2 | GCGCGGGKCCCCCCCCCCNNCGGGGGGCTC CG | 123 |

| | | | |
|---|---|---|---|
| | Fragment 3 | CCCCCCGGCCCCCCCCCGTGCTAAACCCAC CGCGCATGCGCGACCACGCCCCCGCCGCC | 124 |
| TTV-CT30F | Full sequence | GCGGCGG-GGGGGCG-GCCGCG-TTCGCGCGCCGCCCACCAGGGGGTG--CTGCG-CGCCCCCCCCCCGCGCAT GCGCGGGGCCCCCCCCCC--GGGGGGGCTCCGCCCCCCCCGGCCCCCCCCC GTGCTAAACCCACCGCGCATGCGCGACCAC GCCCCCCGCCGCC | 125 |
| | Fragment 1 | GCGGCGG | 126 |
| | Fragment 2 | GGGGGCG | 127 |
| | Fragment 3 | GCCGCG | 128 |
| | Fragment 4 | TTCGCGCGCCGCCCACCAGGGGGTG | 129 |
| | Fragment 5 | CTGCG | 130 |
| | Fragment 6 | CGCCCCCCCCCCGCGCAT | 131 |
| | Fragment 7 | GCGCGGGGCCCCCCCCC | 132 |
| | Fragment 8 | GGGGGGGCTCCGCCCCCCCCGGCCCCCCCCC GTGCTAAACCCACCGCGCATGCGCGACCAC GCCCCCCGCCGCC | 133 |
| TTV-HD23a | Full sequence | CGGCGGCGGCGGCG-CGCGCGCTGCGCGCGCG---CGCCGGGGGGGGCGCCAGCG-CCCCCCCCCCCCGCGCAT GCACGGGTCCCCCCCCCCCACGGGGGGCTCC G CCCCCCGGCCCCCCCCC | 134 |
| | Fragment 1 | CGGCGGCGGCGGCG | 135 |
| | Fragment 2 | CGCGCGCTGCGCGCGCG | 136 |
| | Fragment 3 | CGCCGGGGGGGCGCCAGCG | 137 |
| | Fragment 4 | CCCCCCCCCCCGCGCAT | 138 |
| | Fragment 5 | GCACGGGTCCCCCCCCCCCACGGGGGGCTCC G | 139 |
| | Fragment 6 | CCCCCCGGCCCCCCCCC | 140 |

| | | | |
|---|---|---|---|
| TTV-JA20 | Full sequence | CCGTCGGCGGGGGGGCCGCGCGCTGCGCG CGCGGCCC-CCGGGGGAGGCACAGCCTCCCCCCCCCGCG CGCATGCGCGCGGGTCCCCCCCCCCTCCGGG GGGCTCCGCCCCCCGGCCCCCCCC | 141 |
| | Fragment 1 | CCGTCGGCGGGGGGGCCGCGCGCTGCGCG CGCGGCCC | 142 |
| | Fragment 2 | CCGGGGGAGGCACAGCCTCCCCCCCCCGCG CGCATGCGCGCGGGTCCCCCCCCCCTCCGGG GGGCTCCGCCCCCCGGCCCCCCCC | 143 |
| TTV-TJN02 | Full sequence | CGGCGGCGGCG-CGCGCGCTACGCGCGCG---CGCCGGGGGG----CTGCCGC-CCCCCCCCCGCGCAT GCGCGGGGCCCCCCCCC-GCGGGGGGCTCCG CCCCCCGGCCCCCC | 144 |
| | Fragment 1 | CGGCGGCGGCG | 145 |
| | Fragment 2 | CGCGCGCTACGCGCGCG | 146 |
| | Fragment 3 | CGCCGGGGGG | 147 |
| | Fragment 4 | CTGCCGC | 148 |
| | Fragment 5 | CCCCCCCCCGCGCAT | 149 |
| | Fragment 6 | GCGCGGGGCCCCCCCCC | 150 |
| | Fragment 7 | GCGGGGGGCTCCG | 151 |
| | Fragment 8 | CCCCCCGGCCCCCC | 152 |
| TTV-tth8 | Full sequence | GCCGCCGCGGCGGCGGGGG-GCGGCGCGCTGCGCGCGCCGCCCAGTAGG GGGAGCCATGCG---CCCCCCCCCGCGCAT GCGCGGGGCCCCCCCCC-GCGGGGGGCTCCG CCCCCCGGCCCCCCCCG | 153 |
| | Fragment 1 | GCCGCCGCGGCGGCGGGGG | 154 |
| | Fragment 2 | GCGGCGCGCTGCGCGCGCCGCCCAGTAGG GGGAGCCATGCG | 155 |
| | Fragment 3 | CCCCCCCCCGCGCAT | 156 |

| | | | |
|---|---|---|---|
| | Fragment 4 | GCGCGGGGCCCCCCCCC | 157 |
| | Fragment 5 | GCGGGGGGCTCCG | 158 |
| | Fragment 6 | CCCCCCGGCCCCCCCG | 159 |
| | Fragment 7 | CGCGCTGCGCGCGCCGCCCAGTAGGGGGA GCCATGC | 160 |
| | Fragment 8 | CCGCCATCTTAAGTAGTTGAGGCGGACGGT GGCGTGAGTTCAAAGGTCACCATCAGCCAC ACCTACTCAAAATGGTGG | 161 |
| | Fragment 9 | CTTAAGTAGTTGAGGCGGACGGTGGCGTGA GTTCAAAGGTCACCATCAGCCACACCTACT CAAAATGGTGGACAATTTCTTCCGGGTCAA AGGTTACAGCCGCCATGTTAAAACACGTGA CGTATGACGTCACGGCCGCCATTTTGTGAC ACAAGATGGCCGACTTCCTTCC | 162 |
| Additional GC-rich Sequences | 36-nucleotide consensus GC-rich region sequence 1 | CGCGCTGCGCGCGCCGCCCAGTAGGGGGA GCCATGC | 163 |
| | 36-nucleotide region consensus sequence 2 | GCGCTX$_1$CGCGCGCGCGCCGGGGGGCTGCG CCCCCCC, wherein X$_1$ is selected from T, G, or A | 164 |
| | TTV Clade 1 36-nucleotide region | GCGCTTCGCGCGCCGCCCACTAGGGGGCGT TGCGCG | 165 |
| | TTV Clade 3 36-nucleotide region | GCGCTGCGCGCGCCGCCCAGTAGGGGGCG CAATGCG | 166 |
| | TTV Clade 3 isolate GH1 36-nucleotide region | GCGCTGCGCGCGCGGCCCCCGGGGGAGGC ATTGCCT | 167 |

| | | | |
|---|---|---|---|
| | TTV Clade 3 sle1932 36-nucleotide region | GCGCTGCGCGCGCGCCGGGGGGGCGCCAGCGCCC | 168 |
| | TTV Clade 4 ctdc002 36-nucleotide region | GCGCTTCGCGCGCGCCGGGGGGCTCCGCCCCCCC | 169 |
| | TTV Clade 5 36-nucleotide region | GCGCTTCGCGCGCGCCGGGGGGCTGCGCCCCCCC | 170 |
| | TTV Clade 6 36-nucleotide region | GCGCTACGCGCGCGCCGGGGGGCTGCGCCCCCCCC | 171 |
| | TTV Clade 7 36-nucleotide region | GCGCTACGCGCGCGCCGGGGGGCTCTGCCCCCCC | 172 |
| Additional Alphatorquevirus GC-rich region sequences | TTV-CT30F | GCGGCGGGGGGGCGGCCGCGTTCGCGCGCCGCCCACCAGGGGGTGCTGCGCGCCCCCCCCCGCGCATGCGCGGGGCCCCCCCCCCGGGGGGGCTCCGCCCCCCCGGCCCCCCCCCGTGCTAAACCCACCGCGCATGCGCGACCACGCCCCCGCCGCC | 801 |
| | TTV-P13-1 | CCGAGCGTTAGCGAGGAGTGCGACCCTACCCCCTGGGCCCACTTCTTCGGAGCCGCGCGCTACGCCTTCGGCTGCGCGCGGCACCTCAGACCCCCGCTCGTGCTGACACGCTTGCGCGTGTCAGACCACTTCGGGCTCGCGGGGGTCGGG | 802 |
| | TTV-tth8 | GCCGCCGCGGCGGCGGGGGGCGGCGCGCTGCGCGCCGCCCAGTAGGGGGAGCCATGCGCCCCCCCCCGCGCATGCGCGGGGCCCCCCCCCGCGGGGGGCTCCGCCCCCCGGCCCCCCCCCG | 803 |

| | | | |
|---|---|---|---|
| | TTV-HD20a | CGGCCCAGCGGCGGCGCGCGCGCTTCGCGC GCGCGCCGGGGGGCTCCGCCCCCCCCCGCG CATGCGCGGGGCCCCCCCCCGCGGGGGGCT CCGCCCCCCGGTCCCCCCCCG | 804 |
| | TTV-16 | CGGCCGTGCGGCGGCGCGCGCTTCGCGC GCGCGCCGGGGGCTGCCGCCCCCCCCCGCG CATGCGCGCGGGGCCCCCCCCCGCGGGGG GCTCCGCCCCCCGGCCCCCCCCCCG | 805 |
| | TTV-TJN02 | CGGCGGCGGCGCGCGCGCTACGCGCGCGC GCCGGGGGGCTGCCGCCCCCCCCCCGCGCA TGCGCGGGGCCCCCCCCCGCGGGGGGCTCC GCCCCCCGGCCCCCC | 806 |
| | TTV-HD16d | GGCGGCGGCGCGCGCGCTACGCGCGCGCG CCGGGGAGCTCTGCCCCCCCCCGCGCATGC GCGCGGGTCCCCCCCCCGCGGGGGGCTCCG CCCCCCGGTCCCCCCCCG | 807 |

***Effectors***

In some embodiments, the genetic element may include one or more sequences that are or encode an effector, e.g., a functional effector, e.g., an endogenous effector or an exogenous effector, e.g., a therapeutic polypeptide or nucleic acid, e.g., cytotoxic or cytolytic RNA or protein. In some embodiments, the functional nucleic acid is a non-coding RNA. In some embodiments, the functional nucleic acid is a coding RNA. The effector may modulate a biological activity, for example increasing or decreasing enzymatic activity, gene expression, cell signaling, and cellular or organ function. Effector activities may also include binding regulatory proteins to modulate activity of the regulator, such as transcription or translation. Effector activities also may include activator or inhibitor functions. For example, the effector may induce enzymatic activity by triggering increased substrate affinity in an enzyme, e.g., fructose 2,6-bisphosphate activates phosphofructokinase 1 and increases the rate of glycolysis in response to the insulin. In another example, the effector may inhibit substrate binding to a receptor and inhibit its activation, e.g., naltrexone and naloxone bind opioid receptors without activating them and block the receptors' ability to bind opioids. Effector activities may also include modulating protein stability/degradation and/or transcript stability/degradation. For example, proteins may be targeted for degradation by the polypeptide co-factor, ubiquitin, onto proteins to mark them for

degradation. In another example, the effector inhibits enzymatic activity by blocking the enzyme's active site, e.g., methotrexate is a structural analog of tetrahydrofolate, a coenzyme for the enzyme dihydrofolate reductase that binds to dihydrofolate reductase 1000-fold more tightly than the natural substrate and inhibits nucleotide base synthesis.

In some embodiments, the sequence encoding an effector comprises 100-2000, 100-1000, 100-500, 100-200, 200-2000, 200-1000, 200-500, 500-1000, 500-2000, or 1000-2000 nucleotides. In some embodiments, the effector is a nucleic acid or protein payload, e.g., as described herein.

*Regulatory Nucleic Acids*

In some embodiments, the effector is a regulatory nucleic acid. Regulatory nucleic acids modify expression of an endogenous gene and/or an exogenous gene. In one embodiment, the regulatory nucleic acid targets a host gene. The regulatory nucleic acids may include, but are not limited to, a nucleic acid that hybridizes to an endogenous gene (e.g., miRNA, siRNA, mRNA, lncRNA, RNA, DNA, an antisense RNA, gRNA as described herein elsewhere), nucleic acid that hybridizes to an exogenous nucleic acid such as a viral DNA or RNA, nucleic acid that hybridizes to an RNA, nucleic acid that interferes with gene transcription, nucleic acid that interferes with RNA translation, nucleic acid that stabilizes RNA or destabilizes RNA such as through targeting for degradation, and nucleic acid that modulates a DNA or RNA binding factor. In embodiments, the regulatory nucleic acid encodes an miRNA. In some embodiments, the regulatory nucleic acid is endogenous to a wild-type Anellovirus. In some embodiments, the regulatory nucleic acid is exogenous to a wild-type Anellovirus.

In some embodiments, the regulatory nucleic acid comprises RNA or RNA-like structures typically containing 5-500 base pairs (depending on the specific RNA structure, e.g., miRNA 5-30 bps, lncRNA 200-500 bps) and may have a nucleobase sequence identical (or complementary) or nearly identical (or substantially complementary) to a coding sequence in an expressed target gene within the cell, or a sequence encoding an expressed target gene within the cell.

In some embodiments, the regulatory nucleic acid comprises a nucleic acid sequence, e.g., a guide RNA (gRNA). In some embodiments, the DNA targeting moiety comprises a guide RNA or nucleic acid encoding the guide RNA. A gRNA short synthetic RNA can be composed of a "scaffold" sequence necessary for binding to the incomplete effector moiety and a user-defined ~20 nucleotide targeting sequence for a genomic target. In practice, guide RNA sequences are generally designed to have a length of between 17 – 24 nucleotides (e.g., 19, 20, or 21 nucleotides) and complementary to the targeted nucleic acid sequence. Custom gRNA generators and algorithms are available commercially for use in the design of effective guide RNAs. Gene editing has also been achieved using a chimeric "single guide RNA" ("sgRNA"), an engineered (synthetic) single RNA molecule that mimics a naturally

occurring crRNA-tracrRNA complex and contains both a tracrRNA (for binding the nuclease) and at least one crRNA (to guide the nuclease to the sequence targeted for editing). Chemically modified sgRNAs have also been demonstrated to be effective in genome editing; see, for example, Hendel et al. (2015) Nature Biotechnol., 985 – 991.

The regulatory nucleic acid comprises a gRNA that recognizes specific DNA sequences (e.g., sequences adjacent to or within a promoter, enhancer, silencer, or repressor of a gene).

Certain regulatory nucleic acids can inhibit gene expression through the biological process of RNA interference (RNAi). RNAi molecules comprise RNA or RNA-like structures typically containing 15-50 base pairs (such as about18-25 base pairs) and having a nucleobase sequence identical (complementary) or nearly identical (substantially complementary) to a coding sequence in an expressed target gene within the cell. RNAi molecules include, but are not limited to: short interfering RNAs (siRNAs), double-strand RNAs (dsRNA), micro RNAs (miRNAs), short hairpin RNAs (shRNA), meroduplexes, and dicer substrates (U.S. Pat. Nos. 8,084,599 8,349,809 and 8,513,207).

Long non-coding RNAs (lncRNA) are defined as non-protein coding transcripts longer than 100 nucleotides. This somewhat arbitrary limit distinguishes lncRNAs from small regulatory RNAs such as microRNAs (miRNAs), short interfering RNAs (siRNAs), and other short RNAs. In general, the majority (~78%) of lncRNAs are characterized as tissue-specific. Divergent lncRNAs that are transcribed in the opposite direction to nearby protein-coding genes (comprise a significant proportion ~20% of total lncRNAs in mammalian genomes) may possibly regulate the transcription of the nearby gene.

The genetic element may encode regulatory nucleic acids with a sequence substantially complementary, or fully complementary, to all or a fragment of an endogenous gene or gene product (e.g., mRNA). The regulatory nucleic acids may complement sequences at the boundary between introns and exons to prevent the maturation of newly-generated nuclear RNA transcripts of specific genes into mRNA for transcription. The regulatory nucleic acids that are complementary to specific genes can hybridize with the mRNA for that gene and prevent its translation. The antisense regulatory nucleic acid can be DNA, RNA, or a derivative or hybrid thereof.

The length of the regulatory nucleic acid that hybridizes to the transcript of interest may be between 5 to 30 nucleotides, between about 10 to 30 nucleotides, or about 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 or more nucleotides. The degree of identity of the regulatory nucleic acid to the targeted transcript should be at least 75%, at least 80%, at least 85%, at least 90%, or at least 95%.

The genetic element may encode a regulatory nucleic acid, e.g., a micro RNA (miRNA) molecule identical to about 5 to about 25 contiguous nucleotides of a target gene. In some embodiments, the miRNA sequence targets a mRNA and commences with the dinucleotide AA, comprises a GC-content of

about 30-70% (about 30-60%, about 40-60%, or about 45%-55%), and does not have a high percentage identity to any nucleotide sequence other than the target in the genome of the mammal in which it is to be introduced, for example as determined by standard BLAST search.

siRNAs and shRNAs resemble intermediates in the processing pathway of the endogenous microRNA (miRNA) genes (Bartel, Cell 116:281-297, 2004). In some embodiments, siRNAs can function as miRNAs and vice versa (Zeng et al., Mol Cell 9:1327-1333, 2002; Doench et al., Genes Dev 17:438-442, 2003). MicroRNAs, like siRNAs, use RISC to downregulate target genes, but unlike siRNAs, most animal miRNAs do not cleave the mRNA. Instead, miRNAs reduce protein output through translational suppression or polyA removal and mRNA degradation (Wu et al., Proc Natl Acad Sci USA 103:4034-4039, 2006). Known miRNA binding sites are within mRNA 3' UTRs; miRNAs seem to target sites with near-perfect complementarity to nucleotides 2-8 from the miRNA's 5' end (Rajewsky, Nat Genet 38 Suppl:S8-13, 2006; Lim et al., Nature 433:769-773, 2005). This region is known as the seed region. Because siRNAs and miRNAs are interchangeable, exogenous siRNAs downregulate mRNAs with seed complementarity to the siRNA (Birmingham et al., Nat Methods 3:199-204, 2006. Multiple target sites within a 3' UTR give stronger downregulation (Doench et al., Genes Dev 17:438-442, 2003).

Lists of known miRNA sequences can be found in databases maintained by research organizations, such as Wellcome Trust Sanger Institute, Penn Center for Bioinformatics, Memorial Sloan Kettering Cancer Center, and European Molecule Biology Laboratory, among others. Known effective siRNA sequences and cognate binding sites are also well represented in the relevant literature. RNAi molecules are readily designed and produced by technologies known in the art. In addition, there are computational tools that increase the chance of finding effective and specific sequence motifs (Lagana et al., Methods Mol. Bio., 2015, 1269:393-412).

The regulatory nucleic acid may modulate expression of RNA encoded by a gene. Because multiple genes can share some degree of sequence homology with each other, in some embodiments, the regulatory nucleic acid can be designed to target a class of genes with sufficient sequence homology. In some embodiments, the regulatory nucleic acid can contain a sequence that has complementarity to sequences that are shared amongst different gene targets or are unique for a specific gene target. In some embodiments, the regulatory nucleic acid can be designed to target conserved regions of an RNA sequence having homology between several genes thereby targeting several genes in a gene family (e.g., different gene isoforms, splice variants, mutant genes, etc.). In some embodiments, the regulatory nucleic acid can be designed to target a sequence that is unique to a specific RNA sequence of a single gene.

In some embodiments, the genetic element may include one or more sequences that encode regulatory nucleic acids that modulate expression of one or more genes.

In one embodiment, the gRNA described elsewhere herein are used as part of a CRISPR system for gene editing. For the purposes of gene editing, the anellovector may be designed to include one or multiple guide RNA sequences corresponding to a desired target DNA sequence; see, for example, Cong et al. (2013) Science, 339:819–823; Ran et al. (2013) Nature Protocols, 8:2281 – 2308. At least about 16 or 17 nucleotides of gRNA sequence generally allow for Cas9-mediated DNA cleavage to occur; for Cpf1 at least about 16 nucleotides of gRNA sequence is needed to achieve detectable DNA cleavage.

*Therapeutic effectors (e.g., peptides or polypeptides)*

In some embodiments, the genetic element comprises a therapeutic expression sequence, e.g., a sequence that encodes a therapeutic peptide or polypeptide, e.g., an intracellular peptide or intracellular polypeptide, a secreted polypeptide, or a protein replacement therapeutic. In some embodiments, the genetic element includes a sequence encoding a protein e.g., a therapeutic protein. Some examples of therapeutic proteins may include, but are not limited to, a hormone, a cytokine, an enzyme, an antibody (e.g., one or a plurality of polypeptides encoding at least a heavy chain or a light chain), a transcription factor, a receptor (e.g., a membrane receptor), a ligand, a membrane transporter, a secreted protein, a peptide, a carrier protein, a structural protein, a nuclease, or a component thereof.

In some embodiments, the genetic element includes a sequence encoding a peptide e.g., a therapeutic peptide. The peptides may be linear or branched. The peptide has a length from about 5 to about 500 amino acids, about 15 to about 400 amino acids, about 20 to about 325 amino acids, about 25 to about 250 amino acids, about 50 to about 200 amino acids, or any range there between.

In some embodiments, the polypeptide encoded by the therapeutic expression sequence may be a functional variant or fragment thereof of any of the above, e.g., a protein having at least 80%, 85%, 90%, 95%, 967%, 98%, 99% identity to a protein sequence which disclosed in a table herein by reference to its UniProt ID.

In some embodiments, the therapeutic expression sequence may encode an antibody or antibody fragment that binds any of the above, e.g., an antibody against a protein having at least 80%, 85%, 90%, 95%, 967%, 98%, 99% identity to a protein sequence which disclosed in a table herein by reference to its UniProt ID. The term "antibody" herein is used in the broadest sense and encompasses various antibody structures, including but not limited to monoclonal antibodies, polyclonal antibodies, multispecific antibodies (e.g., bispecific antibodies), and antibody fragments so long as they exhibit the desired antigen-binding activity. An "antibody fragment" refers to a molecule that includes at least one heavy chain or light chain and binds an antigen. Examples of antibody fragments include but are not limited to Fv, Fab, Fab', Fab'-SH, $F(ab')_2$; diabodies; linear antibodies; single-chain antibody molecules (e.g. scFv); and multispecific antibodies formed from antibody fragments.

**Exemplary intracellular polypeptide effectors**

In some embodiments, the effector comprises a cytosolic polypeptide or cytosolic peptide. In some embodiments, the effector comprises cytosolic peptide is a DPP-4 inhibitor, an activator of GLP-1 signaling, or an inhibitor of neutrophil elastase. In some embodiments, the effector increases the level or activity of a growth factor or receptor thereof (e.g., an FGF receptor, e.g., FGFR3). In some embodiments, the effector comprises an inhibitor of n-myc interacting protein activity (e.g., an n-myc interacting protein inhibitor); an inhibitor of EGFR activity (e.g., an EGFR inhibitor); an inhibitor of IDH1 and/or IDH2 activity (e.g., an IDH1 inhibitor and/or an IDH2 inhibitor); an inhibitor of LRP5 and/or DKK2 activity (e.g., an LRP5 and/or DKK2 inhibitor); an inhibitor of KRAS activity; an activator of HTT activity; or inhibitor of DPP-4 activity (e.g., a DPP-4 inhibitor).

In some embodiments, the effector comprises a regulatory intracellular polyeptpide. In some embodiments, the regulatory intracellular polypeptide binds one or more molecule (e.g., protein or nucleic acid) endogenous to the target cell. In some embodiments, the regulatory intracellular polypeptide increases the level or activity of one or more molecule (e.g., protein or nucleic acid) endogenous to the target cell. In some embodiments, the regulatory intracellular polypeptide decreases the level or activity of one or more molecule (e.g., protein or nucleic acid) endogenous to the target cell.

**Exemplary secreted polypeptide effectors**

Exemplary secreted therapeutics are described herein, e.g., in the tables below.

**Table 50. Exemplary cytokines and cytokine receptors**

| Cytokine | Cytokine receptor(s) | Entrez Gene ID | UniProt ID |
|---|---|---|---|
| IL-1α, IL-1β, or a heterodimer thereof | IL-1 type 1 receptor, IL-1 type 2 receptor | 3552, 3553 | P01583, P01584 |
| IL-1Ra | IL-1 type 1 receptor, IL-1 type 2 receptor | 3454, 3455 | P17181, P48551 |
| IL-2 | IL-2R | 3558 | P60568 |
| IL-3 | IL-3 receptor α + β c (CD131) | 3562 | P08700 |
| IL-4 | IL-4R type I, IL-4R type II | 3565 | P05112 |
| IL-5 | IL-5R | 3567 | P05113 |
| IL-6 | IL-6R (sIL-6R) gp130 | 3569 | P05231 |
| IL-7 | IL-7R and sIL-7R | 3574 | P13232 |

| | | | |
|---|---|---|---|
| IL-8 | CXCR1 and CXCR2 | 3576 | P10145 |
| IL-9 | IL-9R | 3578 | P15248 |
| IL-10 | IL-10R1/IL-10R2 complex | 3586 | P22301 |
| IL-11 | IL-11Rα 1 gp130 | 3589 | P20809 |
| IL-12 (e.g., p35, p40, or a heterodimer thereof) | IL-12Rβ1 and IL-12Rβ2 | 3593, 3592 | P29459, P29460 |
| IL-13 | IL-13R1α1 and IL-13R1α2 | 3596 | P35225 |
| IL-14 | IL-14R | 30685 | P40222 |
| IL-15 | IL-15R | 3600 | P40933 |
| IL-16 | CD4 | 3603 | Q14005 |
| IL-17A | IL-17RA | 3605 | Q16552 |
| IL-17B | IL-17RB | 27190 | Q9UHF5 |
| IL-17C | IL-17RA to IL-17RE | 27189 | Q9P0M4 |
| e | SEF | 53342 | Q8TAD2 |
| IL-17F | IL-17RA, IL-17RC | 112744 | Q96PD4 |
| IL-18 | IL-18 receptor | 3606 | Q14116 |
| IL-19 | IL-20R1/IL-20R2 | 29949 | Q9UHD0 |
| IL-20 | L-20R1/IL-20R2 and IL-22R1/ IL-20R2 | 50604 | Q9NYY1 |
| IL-21 | IL-21R | 59067 | Q9HBE4 |
| IL-22 | IL-22R | 50616 | Q9GZX6 |
| IL-23 (e.g., p19, p40, or a heterodimer thereof) | IL-23R | 51561 | Q9NPF7 |
| IL-24 | IL-20R1/IL-20R2 and IL-22R1/ IL-20R2 | 11009 | Q13007 |
| IL-25 | IL-17RA and IL-17RB | 64806 | Q9H293 |
| IL-26 | IL-10R2 chain and IL-20R1 chain | 55801 | Q9NPH9 |
| IL-27 (e.g., p28, EBI3, or a heterodimer thereof) | WSX-1 and gp130 | 246778 | Q8NEV9 |
| IL-28A, IL-28B, and IL29 | IL-28R1/IL-10R2 | 282617, 282618 | Q8IZI9, Q8IU54 |
| IL-30 | IL6R/gp130 | 246778 | Q8NEV9 |
| IL-31 | IL-31RA/OSMRβ | 386653 | Q6EBC2 |

| | | | |
|---|---|---|---|
| IL-32 | | 9235 | P24001 |
| IL-33 | ST2 | 90865 | O95760 |
| IL-34 | Colony-stimulating factor 1 receptor | 146433 | Q6ZMJ4 |
| IL-35 (e.g., p35, EBI3, or a heterodimer thereof) | IL-12Rβ2/gp130; IL-12Rβ2/IL-12Rβ2; gp130/gp130 | 10148 | Q14213 |
| IL-36 | IL-36Ra | 27179 | Q9UHA7 |
| IL-37 | IL-18Rα and IL-18BP | 27178 | Q9NZH6 |
| IL-38 | IL-1R1, IL-36R | 84639 | Q8WWZ1 |
| IFN-α | IFNAR | 3454 | P17181 |
| IFN-β | IFNAR | 3454 | P17181 |
| IFN-γ | IFNGR1/IFNGR2 | 3459 | P15260 |
| TGF-β | TβR-I and TβR-II | 7046, 7048 | P36897, P37173 |
| TNF-α | TNFR1, TNFR2 | 7132, 7133 | P19438, P20333 |

In some embodiments, an effector described herein comprises a cytokine of Table 50, or a functional variant thereof, e.g., a homolog (e.g., ortholog or paralog) or fragment thereof. In some embodiments, an effector described herein comprises a protein having at least 80%, 85%, 90%, 95%, 967%, 98%, 99% sequence identity to an amino acid sequence listed in Table 50 by reference to its UniProt ID. In some embodiments, the functional variant binds to the corresponding cytokine receptor with a Kd of no more than 10%, 20%, 30%, 40%, or 50% higher or lower than the Kd of the corresponding wild-type cytokine for the same receptor under the same conditions. In some embodiments, the effector comprises a fusion protein comprising a first region (e.g., a cytokine polypeptide of Table 50 or a functional variant or fragment thereof) and a second, heterologous region. In some embodiments, the first region is a first cytokine polypeptide of Table 50. In some embodiments, the second region is a second cytokine polypeptide of Table 50, wherein the first and second cytokine polypeptides form a cytokine heterodimer with each other in a wild-type cell. In some embodiments, the polypeptide of Table 50 or functional variant thereof comprises a signal sequence, e.g., a signal sequence that is endogenous to the effector, or a heterologous signal sequence. In some embodiments, an anellovector encoding a cytokine of Table 50, or a functional variant thereof, is used for the treatment of a disease or disorder described herein.

In some embodiments, an effector described herein comprises an antibody molecule (e.g., an scFv) that binds a cytokine of Table 50. In some embodiments, an effector described herein comprises an antibody molecule (e.g., an scFv) that binds a cytokine receptor of Table 50. In some embodiments, the antibody molecule comprises a signal sequence.

Exemplary cytokines and cytokine receptors are described, e.g., in Akdis et al., "Interleukins (from IL-1 to IL-38), interferons, transforming growth factor β, and TNF-α: Receptors, functions, and roles in diseases" October 2016 Volume 138, Issue 4, Pages 984–1010, which is herein incorporated by reference in its entirety, including Table I therein.

**Table 51. Exemplary polypeptide hormones and receptors**

| Hormone | Receptor | Entrez Gene ID | UniProt ID |
|---|---|---|---|
| Natriuretic Peptide, e.g., Atrial Natriuretic Peptide (ANP) | NPRA, NPRB, NPRC | 4878 | P01160 |
| Brain Natriuretic Peptide (BNP) | NPRA, NPRB | 4879 | P16860 |
| C-type natriuretic peptide (CNP) | NPRB | 4880 | P23582 |
| Growth hormone (GH) | GHR | 2690 | P10912 |
| Human growth hormone (hGH) | hGH receptor (human GHR) | 2690 | P10912 |
| Prolactin (PRL) | PRLR | 5617 | P01236 |
| Thyroid-stimulating hormone (TSH) | TSH receptor | 7253 | P16473 |
| Adrenocorticotropic hormone (ACTH) | ACTH receptor | 5443 | P01189 |
| Follicle-stimulating hormone (FSH) | FSHR | 2492 | P23945 |
| Luteinizing hormone (LH) | LHR | 3973 | P22888 |
| Antidiuretic hormone (ADH) | Vasopressin receptors, e.g., V2; AVPR1A; AVPR1B; AVPR3; AVPR2 | 554 | P30518 |
| Oxytocin | OXTR | 5020 | P01178 |

| | | | |
|---|---|---|---|
| Calcitonin | Calcitonin receptor (CT) | 796 | P01258 |
| Parathyroid hormone (PTH) | PTH1R and PTH2R | 5741 | P01270 |
| Insulin | Insulin receptor (IR) | 3630 | P01308 |
| Glucagon | Glucagon receptor | 2641 | P01275 |

In some embodiments, an effector described herein comprises a hormone of Table 51, or a functional variant thereof, e.g., a homolog (e.g., ortholog or paralog) or fragment thereof. In some embodiments, an effector described herein comprises a protein having at least 80%, 85%, 90%, 95%, 967%, 98%, 99% sequence identity to an amino acid sequence listed in Table 51 by reference to its UniProt ID. In some embodiments, the functional variant binds to the corresponding receptor with a Kd of no more than 10%, 20%, 30%, 40%, or 50% higher than the Kd of the corresponding wild-type hormone for the same receptor under the same conditions. In some embodiments, the polypeptide of Table 51 or functional variant thereof comprises a signal sequence, e.g., a signal sequence that is endogenous to the effector, or a heterologous signal sequence. In some embodiments, an anellovector encoding a hormone of Table 51, or a functional variant thereof, is used for the treatment of a disease or disorder described herein.

In some embodiments, an effector described herein comprises an antibody molecule (e.g., an scFv) that binds a hormone of Table 51. In some embodiments, an effector described herein comprises an antibody molecule (e.g., an scFv) that binds a hormone receptor of Table 51. In some embodiments, the antibody molecule comprises a signal sequence.

**Table 52. Exemplary growth factors**

| **Growth Factor** | | **Entrez Gene ID** | **UniProt ID** |
|---|---|---|---|
| PDGF family | | | |
| PDGF (e.g., PDGF-1, PDGF-2, or a heterodimer thereof) | PDGF receptor, e.g., PDGFRα, PDGFRβ | 5156 | P16234 |
| CSF-1 | CSF1R | 1435 | P09603 |
| SCF | CD117 | 3815 | P10721 |
| VEGF family | | | |
| VEGF (e.g., isoforms VEGF 121, VEGF 165, | VEGFR-1, VEGFR-2 | 2321 | P17948 |

| | | | |
|---|---|---|---|
| VEGF 189, and VEGF 206) | | | |
| VEGF-B | VEGFR-1 | 2321 | P17949 |
| VEGF-C | VEGFR-2 and VEGFR -3 | 2324 | P35916 |
| PlGF | VEGFR-1 | 5281 | Q07326 |
| EGF family | | | |
| EGF | EGFR | 1950 | P01133 |
| TGF-α | EGFR | 7039 | P01135 |
| amphiregulin | EGFR | 374 | P15514 |
| HB-EGF | EGFR | 1839 | Q99075 |
| betacellulin | EGFR, ErbB-4 | 685 | P35070 |
| epiregulin | EGFR, ErbB-4 | 2069 | O14944 |
| Heregulin | EGFR, ErbB-4 | 3084 | Q02297 |
| FGF family | | | |
| FGF-1, FGF-2, FGF-3, FGF-4, FGF-5, FGF-6, FGF-7, FGF-8, FGF-9 | FGFR1, FGFR2, FGFR3, and FGFR4 | 2246, 2247, 2248, 2249, 2250, 2251, 2252, 2253, 2254 | P05230, P09038, P11487, P08620, P12034, P10767, P21781, P55075, P31371 |
| Insulin family | | | |
| Insulin | IR | 3630 | P01308 |
| IGF-I | IGF-I receptor, IGF-II receptor | 3479 | P05019 |
| IGF-II | IGF-II receptor | 3481 | P01344 |
| HGF family | | | |
| HGF | MET receptor | 3082 | P14210 |
| MSP | RON | 4485 | P26927 |
| Neurotrophin family | | | |
| NGF | LNGFR, trkA | 4803 | P01138 |
| BDNF | trkB | 627 | P23560 |
| NT-3 | trkA, trkB, trkC | 4908 | P20783 |
| NT-4 | trkA, trkB | 4909 | P34130 |
| NT-5 | trkA, trkB | 4909 | P34130 |

| | | | |
|---|---|---|---|
| Angiopoietin family | | | |
| ANGPT1 | HPK-6/TEK | 284 | Q15389 |
| ANGPT2 | HPK-6/TEK | 285 | O15123 |
| ANGPT3 | HPK-6/TEK | 9068 | O95841 |
| ANGPT4 | HPK-6/TEK | 51378 | Q9Y264 |

In some embodiments, an effector described herein comprises a growth factor of Table 52, or a functional variant thereof, e.g., a homolog (e.g., ortholog or paralog) or fragment thereof. In some embodiments, an effector described herein comprises a protein having at least 80%, 85%, 90%, 95%, 967%, 98%, 99% sequence identity to an amino acid sequence listed in Table 52 by reference to its UniProt ID. In some embodiments, the functional variant binds to the corresponding receptor with a Kd of no more than 10%, 20%, 30%, 40%, or 50% higher than the Kd of the corresponding wild-type growth factor for the same receptor under the same conditions. In some embodiments, the polypeptide of Table 52 or functional variant thereof comprises a signal sequence, e.g., a signal sequence that is endogenous to the effector, or a heterologous signal sequence. In some embodiments, an anellovector encoding a growth factor of Table 52, or a functional variant thereof, is used for the treatment of a disease or disorder described herein.

In some embodiments, an effector described herein comprises an antibody molecule (e.g., an scFv) that binds a growth factor of Table 52. In some embodiments, an effector described herein comprises an antibody molecule (e.g., an scFv) that binds a growth factor receptor of Table 52. In some embodiments, the antibody molecule comprises a signal sequence.

Exemplary growth factors and growth factor receptors are described, e.g., in Bafico et al., "Classification of Growth Factors and Their Receptors" Holland-Frei Cancer Medicine. 6th edition, which is herein incorporated by reference in its entirety.

**Table 53. Clotting-associated factors**

| **Effector** | **Indication** | **Entrez Gene ID** | **UniProt ID** |
|---|---|---|---|
| Factor I (fibrinogen) | Afibrinogenomia | 2243, 2266, 2244 | P02671, P02679, P02675 |
| Factor II | Factor II Deficiency | 2147 | P00734 |
| Factor IX | Hemophilia B | 2158 | P00740 |
| Factor V | Owren's disease | 2153 | P12259 |
| Factor VIII | Hemophilia A | 2157 | P00451 |

| | | | |
|---|---|---|---|
| Factor X | Stuart-Prower Factor Deficiency | 2159 | P00742 |
| Factor XI | Hemophilia C | 2160 | P03951 |
| Factor XIII | Fibrin Stabilizing factor deficiency | 2162, 2165 | P00488, P05160 |
| vWF | von Willebrand disease | 7450 | P04275 |

In some embodiments, an effector described herein comprises a polypeptide of Table 53, or a functional variant thereof, e.g., a homolog (e.g., ortholog or paralog) or fragment thereof. In some embodiments, an effector described herein comprises a protein having at least 80%, 85%, 90%, 95%, 967%, 98%, 99% sequence identity to an amino acid sequence listed in Table 53 by reference to its UniProt ID. In some embodiments, the functional variant catalyzes the same reaction as the corresponding wild-type protein, e.g., at a rate no less than 10%, 20%, 30%, 40%, or 50% lower than the wild-type protein. In some embodiments, the polypeptide of Table 53 or functional variant thereof comprises a signal sequence, e.g., a signal sequence that is endogenous to the effector, or a heterologous signal sequence. In some embodiments, an anellovector encoding a polypeptide of Table 53, or a functional variant thereof is used for the treatment of a disease or disorder of Table 53.

**Exemplary protein replacement therapeutics**

Exemplary protein replacement therapeutics are described herein, e.g., in the tables below.

**Table 54. Exemplary enzymatic effectors and corresponding indications**

| **Effector** | **deficiency** | **Entrez Gene ID** | **UniProt ID** |
|---|---|---|---|
| 3-methylcrotonyl-CoA carboxylase | 3-methylcrotonyl-CoA carboxylase deficiency | 56922, 64087 | Q96RQ3, Q9HCC0 |
| Acetyl-CoA-glucosaminide N-acetyltransferase | Mucopolysaccharidosis MPS III (Sanfilippo's syndrome) Type III-C | 138050 | Q68CP4 |
| ADAMTS13 | Thrombotic Thrombocytopenic Purpura | 11093 | Q76LX8 |
| adenine phosphoribosyltransferase | Adenine phosphoribosyltransferase deficiency | 353 | P07741 |

| | | | |
|---|---|---|---|
| Adenosine deaminase | Adenosine deaminase deficiency | 100 | P00813 |
| ADP-ribose protein hydrolase | Glutamyl ribose-5-phosphate storage disease | 26119, 54936 | Q5SW96, Q9NX46 |
| alpha glucosidase | Glycogen storage disease type 2 (Pompe's disease) | 2548 | P10253 |
| Arginase | Familial hyperarginemia | 383, 384 | P05089, P78540 |
| Arylsulfatase A | Metachromatic leukodystrophy | 410 | P15289 |
| Cathepsin K | Pycnodysostosis | 1513 | P43235 |
| Ceramidase | Farber's disease (lipogranulomatosis) | 125981, 340485, 55331 | Q8TDN7, Q5QJU3, Q9NUN7 |
| Cystathionine B synthase | Homocystinuria | 875 | P35520 |
| Dolichol-P-mannose synthase | Congenital disorders of N-glycosylation CDG Ie | 8813, 54344 | O60762, Q9P2X0 |
| Dolicho-P-Glc:Man9GlcNAc2-PP-dolichol glucosyltransferase | Congenital disorders of N-glycosylation CDG Ic | 84920 | Q5BKT4 |
| Dolicho-P-Man:Man5GlcNAc2-PP-dolichol mannosyltransferase | Congenital disorders of N-glycosylation CDG Id | 10195 | Q92685 |
| Dolichyl-P-glucose:Glc-1-Man-9-GlcNAc-2-PP-dolichyl-α-3-glucosyltransferase | Congenital disorders of N-glycosylation CDG Ih | 79053 | Q9BVK2 |
| Dolichyl-P-mannose:Man-7-GlcNAc-2-PP-dolichyl-α-6-mannosyltransferase | Congenital disorders of N-glycosylation CDG Ig | 79087 | Q9BV10 |
| Factor II | Factor II Deficiency | 2147 | P00734 |

| | | | |
|---|---|---|---|
| Factor IX | Hemophilia B | 2158 | P00740 |
| Factor V | Owren's disease | 2153 | P12259 |
| Factor VIII | Hemophilia A | 2157 | P00451 |
| Factor X | Stuart-Prower Factor Deficiency | 2159 | P00742 |
| Factor XI | Hemophilia C | 2160 | P03951 |
| Factor XIII | Fibrin Stabilizing factor deficiency | 2162, 2165 | P00488, P05160 |
| Galactosamine-6-sulfate sulfatase | Mucopolysaccharidosis MPS IV (Morquio's syndrome) Type IV-A | 2588 | P34059 |
| Galactosylceramide β-galactosidase | Krabbe's disease | 2581 | P54803 |
| Ganglioside β-galactosidase | GM1 gangliosidosis, generalized | 2720 | P16278 |
| Ganglioside β-galactosidase | GM2 gangliosidosis | 2720 | P16278 |
| Ganglioside β-galactosidase | Sphingolipidosis Type I | 2720 | P16278 |
| Ganglioside β-galactosidase | Sphingolipidosis Type II (juvenile type) | 2720 | P16278 |
| Ganglioside β-galactosidase | Sphingolipidosis Type III (adult type) | 2720 | P16278 |
| Glucosidase I | Congenital disorders of N-glycosylation CDG IIb | 2548 | P10253 |
| Glucosylceramide β-glucosidase | Gaucher's disease | 2629 | P04062 |
| Heparan-S-sulfate sulfamidase | Mucopolysaccharidosis MPS III (Sanfilippo's syndrome) Type III-A | 6448 | P51688 |
| homogentisate oxidase | Alkaptonuria | 3081 | Q93099 |
| Hyaluronidase | Mucopolysaccharidosis MPS IX (hyaluronidase deficiency) | 3373, 8692, 8372, 23553 | Q12794, Q12891, O43820, Q2M3T9 |

| | | | |
|---|---|---|---|
| Iduronate sulfate sulfatase | Mucopolysaccharidosis MPS II (Hunter's syndrome) | 3423 | P22304 |
| Lecithin-cholesterol acyltransferase (LCAT) | Complete LCAT deficiency, Fish-eye disease, atherosclerosis, hypercholesterolemia | 3931 | 606967 |
| Lysine oxidase | Glutaric acidemia type I | 4015 | P28300 |
| Lysosomal acid lipase | Cholesteryl ester storage disease (CESD) | 3988 | P38571 |
| Lysosomal acid lipase | Lysosomal acid lipase deficiency | 3988 | P38571 |
| lysosomal acid lipase | Wolman's disease | 3988 | P38571 |
| Lysosomal pepstatin-insensitive peptidase | Ceroid lipofuscinosis Late infantile form (CLN2, Jansky-Bielschowsky disease) | 1200 | O14773 |
| Mannose (Man) phosphate (P) isomerase | Congenital disorders of N-glycosylation CDG Ib | 4351 | P34949 |
| Mannosyl-α-1,6-glycoprotein-β-1,2-N-acetylglucosminyltransferase | Congenital disorders of N-glycosylation CDG IIa | 4247 | Q10469 |
| Metalloproteinase-2 | Winchester syndrome | 4313 | P08253 |
| methylmalonyl-CoA mutase | Methylmalonic acidemia (vitamin b12 non-responsive) | 4594 | P22033 |
| N-Acetyl galactosamine α-4-sulfate sulfatase (arylsulfatase B) | Mucopolysaccharidosis MPS VI (Maroteaux-Lamy syndrome) | 411 | P15848 |
| N-acetyl-D-glucosaminidase | Mucopolysaccharidosis MPS III (Sanfilippo's syndrome) Type III-B | 4669 | P54802 |

| | | | |
|---|---|---|---|
| N-Acetyl-galactosaminidase | Schindler's disease Type I (infantile severe form) | 4668 | P17050 |
| N-Acetyl-galactosaminidase | Schindler's disease Type II (Kanzaki disease, adult-onset form) | 4668 | P17050 |
| N-Acetyl-galactosaminidase | Schindler's disease Type III (intermediate form) | 4668 | P17050 |
| N-acetyl-glucosaminine-6-sulfate sulfatase | Mucopolysaccharidosis MPS III (Sanfilippo's syndrome) Type III-D | 2799 | P15586 |
| N-acetylglucosaminyl-1-phosphotransferase | Mucolipidosis ML III (pseudo-Hurler's polydystrophy) | 79158 | Q3T906 |
| N-Acetylglucosaminyl-1-phosphotransferase catalytic subunit | Mucolipidosis ML II (I-cell disease) | 79158 | Q3T906 |
| N-acetylglucosaminyl-1-phosphotransferase, substrate-recognition subunit | Mucolipidosis ML III (pseudo-Hurler's polydystrophy) Type III-C | 84572 | Q9UJJ9 |
| N-Aspartylglucosaminidase | Aspartylglucosaminuria | 175 | P20933 |
| Neuraminidase 1 (sialidase) | Sialidosis | 4758 | Q99519 |
| Palmitoyl-protein thioesterase-1 | Ceroid lipofuscinosis Adult form (CLN4, Kufs' disease) | 5538 | P50897 |
| Palmitoyl-protein thioesterase-1 | Ceroid lipofuscinosis Infantile form (CLN1, Santavuori-Haltia disease) | 5538 | P50897 |
| Phenylalanine hydroxylase | Phenylketonuria | 5053 | P00439 |

| | | | |
|---|---|---|---|
| Phosphomannomutase-2 | Congenital disorders of N-glycosylation CDG Ia (solely neurologic and neurologic-multivisceral forms) | 5373 | O15305 |
| Porphobilinogen deaminase | Acute Intermittent Porphyria | 3145 | P08397 |
| Purine nucleoside phosphorylase | Purine nucleoside phosphorylase deficiency | 4860 | P00491 |
| pyrimidine 5' nucleotidase | Hemolytic anemia and/or pyrimidine 5' nucleotidase deficiency | 51251 | Q9H0P0 |
| Sphingomyelinase | Niemann-Pick disease type A | 6609 | P17405 |
| Sphingomyelinase | Niemann-Pick disease type B | 6609 | P17405 |
| Sterol 27-hydroxylase | Cerebrotendinous xanthomatosis (cholestanol lipidosis) | 1593 | Q02318 |
| Thymidine phosphorylase | Mitochondrial neurogastrointestinal encephalomyopathy (MNGIE) | 1890 | P19971 |
| Trihexosylceramide α-galactosidase | Fabry's disease | 2717 | P06280 |
| tyrosinase, e.g., OCA1 | albinism, e.g., ocular albinism | 7299 | P14679 |
| UDP-GlcNAc:dolichyl-P NAcGlc phosphotransferase | Congenital disorders of N-glycosylation CDG Ij | 1798 | Q9H3H5 |
| UDP-N-acetylglucosamine-2-epimerase/N-acetylmannosamine kinase, sialin | Sialuria French type | 10020 | Q9Y223 |
| Uricase | Lesch-Nyhan syndrome, gout | 391051 | No protein |

| | | | |
|---|---|---|---|
| uridine diphosphate glucuronyl-transferase (e.g., UGT1A1) | Crigler–Najjar syndrome | 54658 | P22309 |
| α-1,2-Mannosyltransferase | Congenital disorders of N-glycosylation CDG Il (608776) | 79796 | Q9H6U8 |
| α-1,2-Mannosyltransferase | Congenital disorders of N-glycosylation, type I (pre-Golgi glycosylation defects) | 79796 | Q9H6U8 |
| α-1,3-Mannosyltransferase | Congenital disorders of N-glycosylation CDG Ii | 440138 | Q2TAA5 |
| α-D-Mannosidase | α-Mannosidosis, type I (severe) or II (mild) | 10195 | Q92685 |
| α-L-Fucosidase | Fucosidosis | 4123 | Q9NTJ4 |
| α-l-Iduronidase | Mucopolysaccharidosis MPS I H/S (Hurler-Scheie syndrome) | 2517 | P04066 |
| α-l-Iduronidase | Mucopolysaccharidosis MPS I-H (Hurler's syndrome) | 3425 | P35475 |
| α-l-Iduronidase | Mucopolysaccharidosis MPS I-S (Scheie's syndrome) | 3425 | P35475 |
| β-1,4-Galactosyltransferase | Congenital disorders of N-glycosylation CDG IId | 3425 | P35475 |
| β-1,4-Mannosyltransferase | Congenital disorders of N-glycosylation CDG Ik | 2683 | P15291 |
| β-D-Mannosidase | β-Mannosidosis | 56052 | Q9BT22 |
| β-Galactosidase | Mucopolysaccharidosis MPS IV (Morquio's syndrome) Type IV-B | 4126 | O00462 |
| β-Glucuronidase | Mucopolysaccharidosis MPS VII (Sly's syndrome) | 2720 | P16278 |
| β-Hexosaminidase A | Tay-Sachs disease | 2990 | P08236 |
| β-Hexosaminidase B | Sandhoff's disease | 3073 | P06865 |

In some embodiments, an effector described herein comprises an enzyme of Table 54, or a functional variant thereof, e.g., a homolog (e.g., ortholog or paralog) or fragment thereof. In some embodiments, an effector described herein comprises a protein having at least 80%, 85%, 90%, 95%, 967%, 98%, 99% sequence identity to an amino acid sequence listed in Table 54 by reference to its UniProt ID. In some embodiments, the functional variant catalyzes the same reaction as the corresponding wild-type protein, e.g., at a rate no less than 10%, 20%, 30%, 40%, or 50% lower than the wild-type protein. In some embodiments, an anellovector encoding an enzyme of Table 54, or a functional variant thereof is used for the treatment of a disease or disorder of Table 54. In some embodiments, an anellovector is used to deliver uridine diphosphate glucuronyl-transferase or a functional variant thereof to a target cell, e.g., a liver cell. In some embodiments, an anellovector is used to deliver OCA1 or a functional variant thereof to a target cell, e.g., a retinal cell.

**Table 55. Exemplary non-enzymatic effectors and corresponding indications**

| Effector | Indication | Entrez Gene ID | UniProt ID |
|---|---|---|---|
| Survival motor neuron protein (SMN) | spinal muscular atrophy | 6606 | Q16637 |
| Dystrophin or micro-dystrophin | muscular dystrophy (e.g., Duchenne muscular dystrophy or Becker muscular dystrophy) | 1756 | P11532 |
| Complement protein, e.g., Complement factor C1 | Complement Factor I deficiency | 3426 | P05156 |
| Complement factor H | Atypical hemolytic uremic syndrome | 3075 | P08603 |
| Cystinosin (lysosomal cystine transporter) | Cystinosis | 1497 | O60931 |
| Epididymal secretory protein 1 (HE1; NPC2 protein) | Niemann-Pick disease Type C2 | 10577 | P61916 |
| GDP-fucose transporter-1 | Congenital disorders of N-glycosylation CDG | 55343 | Q96A29 |

| | | | |
|---|---|---|---|
| | IIc (Rambam-Hasharon syndrome) | | |
| GM2 activator protein | GM2 activator protein deficiency (Tay-Sachs disease AB variant, GM2A) | 2760 | Q17900 |
| Lysosomal transmembrane CLN3 protein | Ceroid lipofuscinosis Juvenile form (CLN3, Batten disease, Vogt-Spielmeyer disease) | 1207 | Q13286 |
| Lysosomal transmembrane CLN5 protein | Ceroid lipofuscinosis Variant late infantile form, Finnish type (CLN5) | 1203 | O75503 |
| Na phosphate cotransporter, sialin | Infantile sialic acid storage disorder | 26503 | Q9NRA2 |
| Na phosphate cotransporter, sialin | Sialuria Finnish type (Salla disease) | 26503 | Q9NRA2 |
| NPC1 protein | Niemann-Pick disease Type C1/Type D | 4864 | O15118 |
| Oligomeric Golgi complex-7 | Congenital disorders of N-glycosylation CDG IIe | 91949 | P83436 |
| Prosaposin | Prosaposin deficiency | 5660 | P07602 |
| Protective protein/cathepsin A (PPCA) | Galactosialidosis (Goldberg's syndrome, combined neuraminidase and β-galactosidase deficiency) | 5476 | P10619 |
| Protein involved in mannose-P-dolichol utilization | Congenital disorders of N-glycosylation CDG If | 9526 | O75352 |

| | | | |
|---|---|---|---|
| Saposin B | Saposin B deficiency (sulfatide activator deficiency) | 5660 | P07602 |
| Saposin C | Saposin C deficiency (Gaucher's activator deficiency) | 5660 | P07602 |
| Sulfatase-modifying factor-1 | Mucosulfatidosis (multiple sulfatase deficiency) | 285362 | Q8NBK3 |
| Transmembrane CLN6 protein | Ceroid lipofuscinosis Variant late infantile form (CLN6) | 54982 | Q9NWW5 |
| Transmembrane CLN8 protein | Ceroid lipofuscinosis Progressive epilepsy with intellectual disability | 2055 | Q9UBY8 |
| vWF | von Willebrand disease | 7450 | P04275 |
| Factor I (fibrinogen) | Afibrinogenomia | 2243, 2244, 2266 | P02671, P02675, P02679 |
| erythropoietin (hEPO) | | | |

In some embodiments, an effector described herein comprises an erythropoietin (EPO), e.g., a human erythropoietin (hEPO), or a functional variant thereof. In some embodiments, an anellovector encoding an erythropoietin, or a functional variant thereof is used for stimulating erythropoiesis. In some embodiments, an anellovector encoding an erythropoietin, or a functional variant thereof is used for the treatment of a disease or disorder, e.g., anemia. In some embodiments, an anellovector is used to deliver EPO or a functional variant thereof to a target cell, e.g., a red blood cell.

In some embodiments, an effector described herein comprises a polypeptide of Table 55, or a functional variant thereof, e.g., a homolog (e.g., ortholog or paralog) or fragment thereof. In some embodiments, an effector described herein comprises a protein having at least 80%, 85%, 90%, 95%, 967%, 98%, 99% sequence identity to an amino acid sequence listed in Table 55 by reference to its UniProt ID. In some embodiments, an anellovector encoding a polypeptide of Table 55, or a functional variant thereof is used for the treatment of a disease or disorder of Table 55. In some embodiments, an anellovector is used to deliver SMN or a functional variant thereof to a target cell, e.g., a cell of the spinal

cord and/or a motor neuron. In some embodiments, an anellovector is used to deliver a micro-dystrophin to a target cell, e.g., a myocyte.

Exemplary micro-dystrophins are described in Duan, "Systemic AAV Micro-dystrophin Gene Therapy for Duchenne Muscular Dystrophy." Mol Ther. 2018 Oct 3;26(10):2337-2356. doi: 10.1016/j.ymthe.2018.07.011. Epub 2018 Jul 17.

In some embodiments, an effector described herein comprises a clotting factor, e.g., a clotting factor listed in Table 54 or Table 55 herein. In some embodiments, an effector described herein comprises a protein that, when mutated, causes a lysosomal storage disorder, e.g., a protein listed in Table 54 or Table 55 herein. In some embodiments, an effector described herein comprises a transporter protein, e.g., a transporter protein listed in Table 55 herein.

In some embodiments, a functional variant of a wild-type protein comprises a protein that has one or more activities of the wild-type protein, e.g., the functional variant catalyzes the same reaction as the corresponding wild-type protein, e.g., at a rate no less than 10%, 20%, 30%, 40%, or 50% lower than the wild-type protein. In some embodiments, the functional variant binds to the same binding partner that is bound by the wild-type protein, e.g., with a Kd of no more than 10%, 20%, 30%, 40%, or 50% higher than the Kd of the corresponding wild-type protein for the same binding partner under the same conditions. In some embodiments, the functional variant has at a polyeptpide sequence at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to that of the wild-type polypeptide. In some embodiments, the functional variant comprises a homolog (e.g., ortholog or paralog) of the corresponding wild-type protein. In some embodiments, the functional variant is a fusion protein. In some embodiments, the fusion comprises a first region with at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity to the corresponding wild-type protein, and a second, heterologous region. In some embodiments, the functional variant comprises or consists of a fragment of the corresponding wild-type protein.

**Regeneration, Repair, and Fibrosis Factors**

Therapeutic polypeptides described herein also include growth factors, e.g., as disclosed in Table 56, or functional variants thereof, e.g., a protein having at least 80%, 85%, 90%, 95%, 967%, 98%, 99% identity to a protein sequence disclosed in Table 56 by reference to its UniProt ID. Also included are antibodies or fragments thereof against such growth factors, or miRNAs that promote regeneration and repair.

**Table 56. Exemplary regeneration, repair, and fibrosis factors**

| Target | Gene accession # | Protein accession # |
|---|---|---|
| VEGF-A | NG_008732 | NP_001165094 |
| NRG-1 | NG_012005 | NP_001153471 |
| FGF2 | NG_029067 | NP_001348594 |
| FGF1 | Gene ID: 2246 | NP_001341882 |
| miR-199-3p | MIMAT0000232 | |
| miR-590-3p | MIMAT0004801 | |
| mi-17-92 | MI0000071 | https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2732113/figure/F1/ |
| miR-222 | MI0000299 | |
| miR-302-367 | MIR302A And MIR367 | https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4400607/ |

**Transformation Factors**

Therapeutic polypeptides described herein also include transformation factors, e.g., protein factors that transform fibroblasts into differentiated cell e.g., factors disclosed in Table 57 or functional variants thereof, .g., a protein having at least 80%, 85%, 90%, 95%, 967%, 98%, 99% identity to a protein sequence disclosed in Table 57 by reference to its UniProt ID.

**Table 57. Exemplary transformation factors**

| Target | Indication | Gene accession # | Protein accession # |
|---|---|---|---|
| MESP1 | Organ Repair by transforming fibroblasts | Gene ID: 55897 | EAX02066 |
| ETS2 | Organ Repair by transforming fibroblasts | GeneID: 2114 | NP_005230 |
| HAND2 | Organ Repair by transforming fibroblasts | GeneID: 9464 | NP_068808 |
| MYOCARDIN | Organ Repair by transforming fibroblasts | GeneID: 93649 | NP_001139784 |
| ESRRA | Organ Repair by transforming fibroblasts | Gene ID: 2101 | AAH92470 |
| miR-1 | Organ Repair by transforming fibroblasts | MI0000651 | n/a |
| miR-133 | Organ Repair by transforming fibroblasts | MI0000450 | n/a |
| TGFb | Organ Repair by transforming fibroblasts | GeneID: 7040 | NP_000651.3 |
| WNT | Organ Repair by transforming fibroblasts | Gene ID: 7471 | NP_005421 |
| JAK | Organ Repair by transforming fibroblasts | Gene ID: 3716 | NP_001308784 |

| | | | |
|---|---|---|---|
| NOTCH | Organ Repair by transforming fibroblasts | GeneID: 4851 | XP_011517019 |

**Proteins that stimulate cellular regeneration**

Therapeutic polypeptides described herein also include proteins that stimulate cellular regeneration e.g., proteins disclosed in Table 58 or functional variants thereof, e.g., a protein having at least 80%, 85%, 90%, 95%, 967%, 98%, 99% identity to a protein sequence disclosed in Table 58 by reference to its UniProt ID.

**Table 58. Exemplary proteins that stimulate cellular regeneration**

| Target | Gene accession # | Protein accession # |
|---|---|---|
| MST1 | NG_016454 | NP_066278 |
| STK30 | Gene ID: 26448 | NP_036103 |
| MST2 | Gene ID: 6788 | NP_006272 |
| SAV1 | Gene ID: 60485 | NP_068590 |
| LATS1 | Gene ID: 9113 | NP_004681 |
| LATS2 | Gene ID: 26524 | NP_055387 |
| YAP1 | NG_029530 | NP_001123617 |
| CDKN2b | NG_023297 | NP_004927 |
| CDKN2a | NG_007485 | NP_478102 |

**STING modulator effectors**

In some embodiments, a secreted effector described herein modulates STING/cGAS signaling. In some embodiments, the STING modulator is a polypeptide, e.g., a viral polypeptide or a functional variant thereof. For instance, the effector may comprise a STING modulator (e.g., inhibitor) described in Maringer et al. "Message in a bottle: lessons learned from antagonism of STING signalling during RNA virus infection" Cytokine & Growth Factor Reviews Volume 25, Issue 6, December 2014, Pages 669-679, which is incorporated herein by reference in its entirety. Additional STING modulators (e.g., activators) are described, e.g., in Wang et al. "STING activator c-di-GMP enhances the anti-tumor effects of peptide vaccines in melanoma-bearing mice." Cancer Immunol Immunother. 2015 Aug;64(8):1057-66. doi: 10.1007/s00262-015-1713-5. Epub 2015 May 19; Bose "cGAS/STING Pathway in Cancer: Jekyll and Hyde Story of Cancer Immune Response" Int J Mol Sci. 2017 Nov; 18(11): 2456; and Fu et al. "STING agonist formulated cancer vaccines can cure established tumors resistant to PD-1 blockade" Sci Transl Med. 2015 Apr 15; 7(283): 283ra52, each of which is incorporated herein by reference in its entirety.

Some examples of peptides include, but are not limited to, fluorescent tag or marker, antigen, peptide therapeutic, synthetic or analog peptide from naturally-bioactive peptide, agonist or antagonist peptide, anti-microbial peptide, a targeting or cytotoxic peptide, a degradation or self-destruction peptide, and degradation or self-destruction peptides. Peptides useful in the invention described herein also include antigen-binding peptides, e.g., antigen binding antibody or antibody-like fragments, such as single chain antibodies, nanobodies (see, e.g., Steeland et al. 2016. Nanobodies as therapeutics: big opportunities for small antibodies. Drug Discov Today: 21(7):1076-113). Such antigen binding peptides may bind a cytosolic antigen, a nuclear antigen, or an intra-organellar antigen.

In some embodiments, the genetic element comprises a sequence that encodes small peptides, peptidomimetics (e.g., peptoids), amino acids, and amino acid analogs. Such therapeutics generally have a molecular weight less than about 5,000 grams per mole, a molecular weight less than about 2,000 grams per mole, a molecular weight less than about 1,000 grams per mole, a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds. Such therapeutics may include, but are not limited to, a neurotransmitter, a hormone, a drug, a toxin, a viral or microbial particle, a synthetic molecule, and agonists or antagonists thereof.

In some embodiments, the composition or anellovector described herein includes a polypeptide linked to a ligand that is capable of targeting a specific location, tissue, or cell.

*Gene Editing Components*

The genetic element of the anellovector may include one or more genes that encode a component of a gene editing system. Exemplary gene editing systems include the clustered regulatory interspaced short palindromic repeat (CRISPR) system, zinc finger nucleases (ZFNs), and Transcription Activator-Like Effector-based Nucleases (TALEN). ZFNs, TALENs, and CRISPR-based methods are described, e.g., in Gaj et al. Trends Biotechnol. 31.7(2013):397-405; CRISPR methods of gene editing are described, e.g., in Guan et al., Application of CRISPR-Cas system in gene therapy: Pre-clinical progress in animal model. DNA Repair 2016 Oct;46:1-8. doi: 10.1016/j.dnarep.2016.07.004; Zheng et al., Precise gene deletion and replacement using the CRISPR/Cas9 system in human cells. BioTechniques, Vol. 57, No. 3, September 2014, pp. 115–124.

CRISPR systems are adaptive defense systems originally discovered in bacteria and archaea. CRISPR systems use RNA-guided nucleases termed CRISPR-associated or "Cas" endonucleases (e. g., Cas9 or Cpf1) to cleave foreign DNA. In a typical CRISPR/Cas system, an endonuclease is directed to a target nucleotide sequence (e. g., a site in the genome that is to be sequence-edited) by sequence-specific, non-coding "guide RNAs" that target single- or double-stranded DNA sequences. Three classes (I-III) of CRISPR systems have been identified. The class II CRISPR systems use a single Cas endonuclease (rather than multiple Cas proteins). One class II CRISPR system includes a type II Cas endonuclease such as Cas9, a CRISPR RNA ("crRNA"), and a trans-activating crRNA ("tracrRNA"). The crRNA contains a "guide RNA", typically about 20-nucleotide RNA sequence that corresponds to a target DNA sequence. The crRNA also contains a region that binds to the tracrRNA to form a partially double-stranded structure which is cleaved by RNase III, resulting in a crRNA/tracrRNA hybrid. The crRNA/tracrRNA hybrid then directs the Cas9 endonuclease to recognize and cleave the target DNA sequence. The target DNA sequence must generally be adjacent to a "protospacer adjacent motif" ("PAM") that is specific for a given Cas endonuclease; however, PAM sequences appear throughout a given genome.

In some embodiments, the anellovector includes a gene for a CRISPR endonuclease. For example, some CRISPR endonucleases identified from various prokaryotic species have unique PAM sequence requirements; examples of PAM sequences include 5'-NGG (Streptococcus pyogenes), 5'-NNAGAA (Streptococcus thermophilus CRISPR1), 5'-NGGNG (Streptococcus thermophilus CRISPR3), and 5'-NNNGATT (Neisseria meningiditis). Some endonucleases, e. g., Cas9 endonucleases, are associated with G-rich PAM sites, e. g., 5'-NGG, and perform blunt-end cleaving of the target DNA at a location 3 nucleotides upstream from (5' from) the PAM site. Another class II CRISPR system includes the type V endonuclease Cpf1, which is smaller than Cas9; examples include AsCpf1 (from Acidaminococcus sp.) and LbCpf1 (from Lachnospiraceae sp.). Cpf1 endonucleases, are associated with

T-rich PAM sites, e. g., 5'-TTN. Cpf1 can also recognize a 5'-CTA PAM motif. Cpf1 cleaves the target DNA by introducing an offset or staggered double-strand break with a 4- or 5-nucleotide 5' overhang, for example, cleaving a target DNA with a 5-nucleotide offset or staggered cut located 18 nucleotides downstream from (3' from) from the PAM site on the coding strand and 23 nucleotides downstream from the PAM site on the complimentary strand; the 5-nucleotide overhang that results from such offset cleavage allows more precise genome editing by DNA insertion by homologous recombination than by insertion at blunt-end cleaved DNA. See, e. g., Zetsche et al. (2015) Cell, 163:759 – 771.

A variety of CRISPR associated (Cas) genes may be included in the anellovector. Specific examples of genes are those that encode Cas proteins from class II systems including Cas1, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9, Cas10, Cpf1, C2C1, or C2C3. In some embodiments, the anellovector includes a gene encoding a Cas protein, e.g., a Cas9 protein, may be from any of a variety of prokaryotic species. In some embodiments, the anellovector includes a gene encoding a particular Cas protein, e.g., a particular Cas9 protein, is selected to recognize a particular protospacer-adjacent motif (PAM) sequence. In some embodiments, the anellovector includes nucleic acids encoding two or more different Cas proteins, or two or more Cas proteins, may be introduced into a cell, zygote, embryo, or animal, e.g., to allow for recognition and modification of sites comprising the same, similar or different PAM motifs. In some embodiments, the anellovector includes a gene encoding a modified Cas protein with a deactivated nuclease, e.g., nuclease-deficient Cas9.

Whereas wild-type Cas9 protein generates double-strand breaks (DSBs) at specific DNA sequences targeted by a gRNA, a number of CRISPR endonucleases having modified functionalities are known, for example: a "nickase" version of Cas endonuclease (e.g., Cas9) generates only a single-strand break; a catalytically inactive Cas endonuclease, e.g., Cas9 ("dCas9") does not cut the target DNA. A gene encoding a dCas9 can be fused with a gene encoding an effector domain to repress (CRISPRi) or activate (CRISPRa) expression of a target gene. For example, the gene may encode a Cas9 fusion with a transcriptional silencer (e.g., a KRAB domain) or a transcriptional activator (e.g., a dCas9–VP64 fusion). A gene encoding a catalytically inactive Cas9 (dCas9) fused to FokI nuclease ("dCas9-FokI") can be included to generate DSBs at target sequences homologous to two gRNAs. See, e. g., the numerous CRISPR/Cas9 plasmids disclosed in and publicly available from the Addgene repository (Addgene, 75 Sidney St., Suite 550A, Cambridge, MA 02139; addgene.org/crispr/). A "double nickase" Cas9 that introduces two separate double-strand breaks, each directed by a separate guide RNA, is described as achieving more accurate genome editing by Ran et al. (2013) Cell, 154:1380 – 1389.

CRISPR technology for editing the genes of eukaryotes is disclosed in US Patent Application Publications 2016/0138008A1 and US2015/0344912A1, and in US Patents 8,697,359, 8,771,945, 8,945,839, 8,999,641, 8,993,233, 8,895,308, 8,865,406, 8,889,418, 8,871,445, 8,889,356, 8,932,814,

8,795,965, and 8,906,616. Cpf1 endonuclease and corresponding guide RNAs and PAM sites are disclosed in US Patent Application Publication 2016/0208243 A1.

In some embodiments, the anellovector comprises a gene encoding a polypeptide described herein, e.g., a targeted nuclease, e.g., a Cas9, e.g., a wild type Cas9, a nickase Cas9 (e.g., Cas9 D10A), a dead Cas9 (dCas9), eSpCas9, Cpf1, C2C1, or C2C3, and a gRNA. The choice of genes encoding the nuclease and gRNA(s) is determined by whether the targeted mutation is a deletion, substitution, or addition of nucleotides, e.g., a deletion, substitution, or addition of nucleotides to a targeted sequence. Genes that encode a catalytically inactive endonuclease e.g., a dead Cas9 (dCas9, e.g., D10A; H840A) tethered with all or a portion of (e.g., biologically active portion of) an (one or more) effector domain (e.g., VP64) create chimeric proteins that can modulate activity and/or expression of one or more target nucleic acids sequences.

In some embodiments, the anellovector includes a gene encoding a fusion of a dCas9 with all or a portion of one or more effector domains (e.g., a full-length wild-type effector domain, or a fragment or variant thereof, e.g., a biologically active portion thereof) to create a chimeric protein useful in the methods described herein. Accordingly, in some embodiments, the anellovector includes a gene encoding a dCas9-methylase fusion. In other some embodiments, the anellovector includes a gene encoding a dCas9-enzyme fusion with a site-specific gRNA to target an endogenous gene.

In other aspects, the anellovector includes a gene encoding 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more effector domains (all or a biologically active portion) fused with dCas9.

***Regulatory Sequences***

In some embodiments, the genetic element comprises a regulatory sequence, e.g., a promoter or an enhancer, operably linked to the sequence encoding the effector. In some embodiments, e.g., wherein the genetic element is an mRNA, a promoter may be absent from the genetic element. In some embodiments, a genetic element construct comprises a promoter that is used to drive production of the RNA genetic element.

In some embodiments, a promoter includes a DNA sequence that is located adjacent to a DNA sequence that encodes an expression product. A promoter may be linked operatively to the adjacent DNA sequence. A promoter typically increases an amount of product expressed from the DNA sequence as compared to an amount of the expressed product when no promoter exists. A promoter from one organism can be utilized to enhance product expression from the DNA sequence that originates from another organism. For example, a vertebrate promoter may be used for the expression of jellyfish GFP in

vertebrates. Hence, one promoter element can enhance the expression of one or more products. Multiple promoter elements are well-known to persons of ordinary skill in the art.

In one embodiment, high-level constitutive expression is desired. Examples of such promoters include, without limitation, the retroviral Rous sarcoma virus (RSV) long terminal repeat (LTR) promoter/enhancer, the cytomegalovirus (CMV) immediate early promoter/enhancer (see, e.g., Boshart et al, Cell, 41:521-530 (1985)), the SV40 promoter, the dihydrofolate reductase promoter, the cytoplasmic .beta.-actin promoter and the phosphoglycerol kinase (PGK) promoter.

In another embodiment, inducible promoters may be desired. Inducible promoters are those which are regulated by exogenously supplied compounds, e.g., provided either in *cis* or in *trans*, including without limitation, the zinc-inducible sheep metallothionine (MT) promoter; the dexamethasone (Dex)-inducible mouse mammary tumor virus (MMTV) promoter; the T7 polymerase promoter system (WO 98/10088); the tetracycline-repressible system (Gossen et al, Proc. Natl. Acad. Sci. USA, 89:5547-5551 (1992)); the tetracycline-inducible system (Gossen et al., Science, 268:1766-1769 (1995); see also Harvey et al., Curr. Opin. Chem. Biol., 2:512-518 (1998)); the RU486-inducible system (Wang et al., Nat. Biotech., 15:239-243 (1997) and Wang et al., Gene Ther., 4:432-441 (1997)]; and the rapamycin-inducible system (Magari et al., J. Clin. Invest., 100:2865-2872 (1997); Rivera et al., Nat. Medicine. 2:1028-1032 (1996)). Other types of inducible promoters which may be useful in this context are those which are regulated by a specific physiological state, e.g., temperature, acute phase, or in replicating cells only.

In some embodiments, a native promoter for a gene or nucleic acid sequence of interest is used. The native promoter may be used when it is desired that expression of the gene or the nucleic acid sequence should mimic the native expression. The native promoter may be used when expression of the gene or other nucleic acid sequence must be regulated temporally or developmentally, or in a tissue-specific manner, or in response to specific transcriptional stimuli. In a further embodiment, other native expression control elements, such as enhancer elements, polyadenylation sites or Kozak consensus sequences may also be used to mimic the native expression.

In some embodiments, the genetic element comprises a gene operably linked to a tissue-specific promoter. For instance, if expression in skeletal muscle is desired, a promoter active in muscle may be used. These include the promoters from genes encoding skeletal α-actin, myosin light chain 2A, dystrophin, muscle creatine kinase, as well as synthetic muscle promoters with activities higher than naturally-occurring promoters. See Li et al., Nat. Biotech., 17:241-245 (1999). Examples of promoters that are tissue-specific are known for liver albumin, Miyatake et al. J. Virol., 71:5124-32 (1997); hepatitis B virus core promoter, Sandig et al., Gene Ther. 3:1002-9 (1996); alpha-fetoprotein (AFP), Arbuthnot et al., Hum. Gene Ther., 7:1503-14 (1996)], bone (osteocalcin, Stein et al., Mol. Biol. Rep., 24:185-96

(1997); bone sialoprotein, Chen et al., J. Bone Miner. Res. 11:654-64 (1996)), lymphocytes (CD2, Hansal et al., J. Immunol., 161:1063-8 (1998); immunoglobulin heavy chain; T cell receptor a chain), neuronal (neuron-specific enolase (NSE) promoter, Andersen et al. Cell. Mol. Neurobiol., 13:503-15 (1993); neurofilament light-chain gene, Piccioli et al., Proc. Natl. Acad. Sci. USA, 88:5611-5 (1991); the neuron-specific vgf gene, Piccioli et al., Neuron, 15:373-84 (1995)]; among others.

The genetic element construct may include an enhancer, e.g., a DNA sequence that is located adjacent to the DNA sequence that encodes a gene. Enhancer elements are typically located upstream of a promoter element or can be located downstream of or within a coding DNA sequence (e.g., a DNA sequence transcribed or translated into a product or products). Hence, an enhancer element can be located 100 base pairs, 200 base pairs, or 300 or more base pairs upstream or downstream of a DNA sequence that encodes the product. Enhancer elements can increase an amount of recombinant product expressed from a DNA sequence above increased expression afforded by a promoter element. Multiple enhancer elements are readily available to persons of ordinary skill in the art.

In some embodiments, the genetic element comprises one or more inverted terminal repeats (ITR) flanking the sequences encoding the expression products described herein. In some embodiments, the genetic element comprises one or more long terminal repeats (LTR) flanking the sequence encoding the expression products described herein. Examples of promoter sequences that may be used, include, but are not limited to, the simian virus 40 (SV40) early promoter, mouse mammary tumor virus (MMTV), human immunodeficiency virus (HIV) long terminal repeat (LTR) promoter, MoMuLV promoter, an avian leukemia virus promoter, an Epstein-Barr virus immediate early promoter, and a Rous sarcoma virus promoter.

***Other Sequences***

In some embodiments, the genetic element further includes a nucleic acid encoding a product (e.g., a ribozyme, a therapeutic mRNA encoding a protein, an exogenous gene).

In some embodiments, the genetic element includes one or more sequences that affect species and/or tissue and/or cell tropism (e.g. capsid protein sequences), infectivity (e.g. capsid protein sequences), immunosuppression/activation (e.g. regulatory nucleic acids), viral genome binding and/or packaging, immune evasion (non-immunogenicity and/or tolerance), pharmacokinetics, endocytosis and/or cell attachment, nuclear entry, intracellular modulation and localization, exocytosis modulation, propagation, and nucleic acid protection of the anellovector in a host or host cell.

In some embodiments, the genetic element may comprise other sequences that include DNA, RNA, or artificial nucleic acids. The other sequences may include, but are not limited to, genomic DNA,

cDNA, or sequences that encode tRNA, mRNA, rRNA, miRNA, gRNA, siRNA, or other RNAi molecules. In one embodiment, the genetic element includes a sequence encoding an siRNA to target a different loci of the same gene expression product as the regulatory nucleic acid. In one embodiment, the genetic element includes a sequence encoding an siRNA to target a different gene expression product as the regulatory nucleic acid.

In some embodiments, the genetic element further comprises one or more of the following sequences: a sequence that encodes one or more miRNAs, a sequence that encodes one or more replication proteins, a sequence that encodes an exogenous gene, a sequence that encodes a therapeutic, a regulatory sequence (e.g., a promoter, enhancer), a sequence that encodes one or more regulatory sequences that targets endogenous genes (siRNA, lncRNAs, shRNA), and a sequence that encodes a therapeutic mRNA or protein.

The other sequences may have a length from about 2 to about 5000 nts, about 10 to about 100 nts, about 50 to about 150 nts, about 100 to about 200 nts, about 150 to about 250 nts, about 200 to about 300 nts, about 250 to about 350 nts, about 300 to about 500 nts, about 10 to about 1000 nts, about 50 to about 1000 nts, about 100 to about 1000 nts, about 1000 to about 2000 nts, about 2000 to about 3000 nts, about 3000 to about 4000 nts, about 4000 to about 5000 nts, or any range therebetween.

*Encoded Genes*

For example, the genetic element may include a gene associated with a signaling biochemical pathway, e.g., a signaling biochemical pathway-associated gene or polynucleotide. Examples include a disease associated gene or polynucleotide. A "disease-associated" gene or polynucleotide refers to any gene or polynucleotide which is yielding transcription or translation products at an abnormal level or in an abnormal form in cells derived from a disease-affected tissues compared with tissues or cells of a non disease control. It may be a gene that becomes expressed at an abnormally high level; it may be a gene that becomes expressed at an abnormally low level, where the altered expression correlates with the occurrence and/or progression of the disease. A disease-associated gene also refers to a gene possessing mutation(s) or genetic variation that is directly responsible or is in linkage disequilibrium with a gene(s) that is responsible for the etiology of a disease.

Examples of disease-associated genes and polynucleotides are available from McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, Md.) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, Md.). Examples of disease-associated genes and polynucleotides are listed in Tables A and B of US Patent No.: 8,697,359, which are herein incorporated by reference in their entirety. Disease specific information is available from McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, Md.) and

National Center for Biotechnology Information, National Library of Medicine (Bethesda, Md.). Examples of signaling biochemical pathway-associated genes and polynucleotides are listed in Tables A-C of US Patent No.: 8,697,359, which are herein incorporated by reference in their entirety.

Moreover, the genetic elements can encode targeting moieties, as described elsewhere herein. This can be achieved, e.g., by inserting a polynucleotide encoding a sugar, a glycolipid, or a protein, such as an antibody. Those skilled in the art know additional methods for generating targeting moieties.

*Viral Sequence*

In some embodiments, the genetic element comprises at least one viral sequence. In some embodiments, the sequence has homology or identity to one or more sequence from a single stranded DNA virus, e.g., Anellovirus, Bidnavirus, Circovirus, Geminivirus, Genomovirus, Inovirus, Microvirus, Nanovirus, Parvovirus, and Spiravirus. In some embodiments, the sequence has homology or identity to one or more sequence from a double stranded DNA virus, e.g., Adenovirus, Ampullavirus, Ascovirus, Asfarvirus, Baculovirus, Fusellovirus, Globulovirus, Guttavirus, Hytrosavirus, Herpesvirus, Iridovirus, Lipothrixvirus, Nimavirus, and Poxvirus. In some embodiments, the sequence has homology or identity to one or more sequence from an RNA virus, e.g., Alphavirus, Furovirus, Hepatitis virus, Hordeivirus, Tobamovirus, Tobravirus, Tricornavirus, Rubivirus, Birnavirus, Cystovirus, Partitivirus, and Reovirus.

In some embodiments, the genetic element may comprise one or more sequences from a non-pathogenic virus, e.g., a symbiotic virus, e.g., a commensal virus, e.g., a native virus, e.g., an *Anellovirus*. Recent changes in nomenclature have classified the three *Anelloviruses* able to infect human cells into Alphatorquevirus (TT), Betatorquevirus (TTM), and Gammatorquevirus (TTMD) Genera of the *Anelloviridae* family of viruses. In some embodiments, the genetic element may comprise a sequence with homology or identity to a Torque Teno Virus (TT), a non-enveloped, single-stranded DNA virus with a circular, negative-sense genome. In some embodiments, the genetic element may comprise a sequence with homology or identity to a SEN virus, a Sentinel virus, a TTV-like mini virus, and a TT virus. Different types of TT viruses have been described including TT virus genotype 6, TT virus group, TTV-like virus DXL1, and TTV-like virus DXL2. In some embodiments, the genetic element may comprise a sequence with homology or identity to a smaller virus, Torque Teno-like Mini Virus (TTM), or a third virus with a genomic size in between that of TTV and TTMV, named Torque Teno-like Midi Virus (TTMD). In some embodiments, the genetic element may comprise one or more sequences or a fragment of a sequence from a non-pathogenic virus having at least about 60%, 70% 80%, 85%, 90% 95%, 96%, 97%, 98% and 99% nucleotide sequence identity to any one of the nucleotide sequences described herein.

In some embodiments, the genetic element may comprise one or more sequences or a fragment of a sequence from a substantially non-pathogenic virus having at least about 60%, 70% 80%, 85%, 90% 95%, 96%, 97%, 98% and 99% nucleotide sequence identity to any one of the nucleotide sequences described herein, e.g., Table 41.

**Table 41: Examples of *Anelloviruses* and their sequences.** Accessions numbers and related sequence information may be obtained at www.ncbi.nlm.nih.gov/genbank/, as referenced on December 11, 2018.

| **Accession #** | **Description** |
|---|---|
| AB017613.1 | Torque teno virus 16 DNA, complete genome, isolate: TUS01 |
| AB026345.1 | TT virus genes for ORF1 and ORF2, complete cds, isolate:TRM1 |
| AB026346.1 | TT virus genes for ORF1 and ORF2, complete cds, isolate:TK16 |
| AB026347.1 | TT virus genes for ORF1 and ORF2, complete cds, isolate:TP1-3 |
| AB028669.1 | TT virus gene for ORF1 and ORF2, complete genome, isolate:TJN02 |
| AB030487.1 | TT virus gene for pORF2a, pORF2b, pORF1, complete cds, clone:JaCHCTC19 |
| AB030488.1 | TT virus gene for pORF2a, pORF2b, pORF1, complete cds, clone:JaBD89 |
| AB030489.1 | TT virus gene for pORF2a, pORF2b, pORF1, complete cds, clone:JaBD98 |
| AB038340.1 | TT virus genes for ORF2s, ORF1, ORF3, complete cds |
| AB038622.1 | TT virus genes for ORF2, ORF1, ORF3, complete cds, isolate:TTVyon-LC011 |
| AB038623.1 | TT virus genes for ORF2, ORF1, ORF3, complete cds, isolate:TTVyon-KC186 |
| AB038624.1 | TT virus genes for ORF2, ORF1, ORF3, complete cds, isolate:TTVyon-KC197 |
| AB041821.1 | TT virus mRNA for VP1, complete cds |
| AB050448.1 | Torque teno virus genes for ORF1, ORF2, ORF3, ORF4, complete cds, isolate: TYM9 |
| AB060592.1 | Torque teno virus gene for ORF1, ORF2, ORF3, ORF4, clone: SAa-39 |
| AB060593.1 | Torque teno virus gene for ORF1, ORF2, ORF3, ORF4, complete cds, clone: SAa-38 |
| AB060595.1 | TT virus gene for ORF1, ORF2, ORF3, ORF4, complete cds, clone:SAj-30 |
| AB060596.1 | TT virus gene for ORF1, ORF2, ORF3, ORF4, complete cds, clone:SAf-09 |
| AB064596.1 | Torque teno virus DNA, complete genome, isolate: CT25F |
| AB064597.1 | Torque teno virus DNA, complete genome, isolate: CT30F |
| AB064599.1 | Torque teno virus DNA, complete genome, isolate: JT03F |
| AB064600.1 | Torque teno virus DNA, complete genome, isolate: JT05F |
| AB064601.1 | Torque teno virus DNA, complete genome, isolate: JT14F |

| | |
|---|---|
| AB064602.1 | Torque teno virus DNA, complete genome, isolate: JT19F |
| AB064603.1 | Torque teno virus DNA, complete genome, isolate: JT41F |
| AB064604.1 | Torque teno virus DNA, complete genome, isolate: CT39F |
| AB064606.1 | Torque teno virus DNA, complete genome, isolate: JT33F |
| AB290918.1 | Torque teno midi virus 1 DNA, complete genome, isolate: MD1-073 |
| AF079173.1 | TT virus strain TTVCHN1, complete genome |
| AF116842.1 | TT virus strain BDH1, complete genome |
| AF122914.3 | TT virus isolate JA20, complete genome |
| AF122917.1 | TT virus isolate JA4, complete genome |
| AF122919.1 | TT virus isolate JA10 unknown genes |
| AF129887.1 | TT virus TTVCHN2, complete genome |
| AF247137.1 | TT virus isolate TUPB, complete genome |
| AF254410.1 | TT virus ORF2 protein and ORF1 protein genes, complete cds |
| AF298585.1 | TT virus Polish isolate P/1C1, complete genome |
| AF315076.1 | TTV-like virus DXL1 unknown genes |
| AF315077.1 | TTV-like virus DXL2 unknown genes |
| AF345521.1 | TT virus isolate TCHN-G1 Orf2 and Orf1 genes, complete cds |
| AF345522.1 | TT virus isolate TCHN-E Orf2 and Orf1 genes, complete cds |
| AF345525.1 | TT virus isolate TCHN-D2 Orf2 and Orf1 genes, complete cds |
| AF345527.1 | TT virus isolate TCHN-C2 Orf2 and Orf1 genes, complete cds |
| AF345528.1 | TT virus isolate TCHN-F Orf2 and Orf1 genes, complete cds |
| AF345529.1 | TT virus isolate TCHN-G2 Orf2 and Orf1 genes, complete cds |
| AF371370.1 | TT virus ORF1, ORF3, and ORF2 genes, complete cds |
| AJ620212.1 | Torque teno virus, isolate tth6, complete genome |
| AJ620213.1 | Torque teno virus, isolate tth10, complete genome |
| AJ620214.1 | Torque teno virus, isolate tth11g2, complete genome |
| AJ620215.1 | Torque teno virus, isolate tth18, complete genome |
| AJ620216.1 | Torque teno virus, isolate tth20, complete genome |
| AJ620217.1 | Torque teno virus, isolate tth21, complete genome |
| AJ620218.1 | Torque teno virus, isolate tth3, complete genome |
| AJ620219.1 | Torque teno virus, isolate tth9, complete genome |
| AJ620220.1 | Torque teno virus, isolate tth16, complete genome |
| AJ620221.1 | Torque teno virus, isolate tth17, complete genome |

| | |
|---|---|
| AJ620222.1 | Torque teno virus, isolate tth25, complete genome |
| AJ620223.1 | Torque teno virus, isolate tth26, complete genome |
| AJ620224.1 | Torque teno virus, isolate tth27, complete genome |
| AJ620225.1 | Torque teno virus, isolate tth31, complete genome |
| AJ620226.1 | Torque teno virus, isolate tth4, complete genome |
| AJ620227.1 | Torque teno virus, isolate tth5, complete genome |
| AJ620228.1 | Torque teno virus, isolate tth14, complete genome |
| AJ620229.1 | Torque teno virus, isolate tth29, complete genome |
| AJ620230.1 | Torque teno virus, isolate tth7, complete genome |
| AJ620231.1 | Torque teno virus, isolate tth8, complete genome |
| AJ620232.1 | Torque teno virus, isolate tth13, complete genome |
| AJ620233.1 | Torque teno virus, isolate tth19, complete genome |
| AJ620234.1 | Torque teno virus, isolate tth22g4, complete genome |
| AJ620235.1 | Torque teno virus, isolate tth23, complete genome |
| AM711976.1 | TT virus sle1957 complete genome |
| AM712003.1 | TT virus sle1931 complete genome |
| AM712004.1 | TT virus sle1932 complete genome |
| AM712030.1 | TT virus sle2057 complete genome |
| AM712031.1 | TT virus sle2058 complete genome |
| AM712032.1 | TT virus sle2072 complete genome |
| AM712033.1 | TT virus sle2061 complete genome |
| AM712034.1 | TT virus sle2065 complete genome |
| AY026465.1 | TT virus isolate L01 ORF2 and ORF1 genes, complete cds |
| AY026466.1 | TT virus isolate L02 ORF2 and ORF1 genes, complete cds |
| DQ003341.1 | Torque teno virus clone P2-9-02 ORF2 (ORF2), ORF1A (ORF1A), and ORF1B (ORF1B) genes, complete cds |
| DQ003342.1 | Torque teno virus clone P2-9-07 ORF2 (ORF2), ORF1A (ORF1A), and ORF1B (ORF1B) genes, complete cds |
| DQ003343.1 | Torque teno virus clone P2-9-08 ORF2 (ORF2), ORF1A (ORF1A), and ORF1B (ORF1B) genes, complete cds |
| DQ003344.1 | Torque teno virus clone P2-9-16 ORF2 (ORF2), ORF1A (ORF1A), and ORF1B (ORF1B) genes, complete cds |

| | |
|---|---|
| DQ186994.1 | Torque teno virus clone P601 ORF2 (ORF2) and ORF1 (ORF1) genes, complete cds |
| DQ186995.1 | Torque teno virus clone P605 ORF2 (ORF2) and ORF1 (ORF1) genes, complete cds |
| DQ186996.1 | Torque teno virus clone BM1A-02 ORF2 (ORF2) and ORF1 (ORF1) genes, complete cds |
| DQ186997.1 | Torque teno virus clone BM1A-09 ORF2 (ORF2) and ORF1 (ORF1) genes, complete cds |
| DQ186998.1 | Torque teno virus clone BM1A-13 ORF2 (ORF2) and ORF1 (ORF1) genes, complete cds |
| DQ186999.1 | Torque teno virus clone BM1B-05 ORF2 (ORF2) and ORF1 (ORF1) genes, complete cds |
| DQ187000.1 | Torque teno virus clone BM1B-07 ORF2 (ORF2) and ORF1 (ORF1) genes, complete cds |
| DQ187001.1 | Torque teno virus clone BM1B-11 ORF2 (ORF2) and ORF1 (ORF1) genes, complete cds |
| DQ187002.1 | Torque teno virus clone BM1B-14 ORF2 (ORF2) and ORF1 (ORF1) genes, complete cds |
| DQ187003.1 | Torque teno virus clone BM1B-08 ORF2 (ORF2) gene, complete cds; and nonfunctional ORF1 (ORF1) gene, complete sequence |
| DQ187004.1 | Torque teno virus clone BM1C-16 ORF2 (ORF2) and ORF1 (ORF1) genes, complete cds |
| DQ187005.1 | Torque teno virus clone BM1C-10 ORF2 (ORF2) and ORF1 (ORF1) genes, complete cds |
| DQ187007.1 | Torque teno virus clone BM2C-25 ORF2 (ORF2) gene, complete cds; and nonfunctional ORF1 (ORF1) gene, complete sequence |
| DQ361268.1 | Torque teno virus isolate ViPi04 ORF1 gene, complete cds |
| EF538879.1 | Torque teno virus isolate CSC5 ORF2 and ORF1 genes, complete cds |
| EU305675.1 | Torque teno virus isolate LTT7 ORF1 gene, complete cds |
| EU305676.1 | Torque teno virus isolate LTT10 ORF1 gene, complete cds |
| EU889253.1 | Torque teno virus isolate ViPi08 nonfunctional ORF1 gene, complete sequence |
| FJ392105.1 | Torque teno virus isolate TW53A25 ORF2 gene, partial cds; and ORF1 gene, complete cds |

| | |
|---|---|
| FJ392107.1 | Torque teno virus isolate TW53A27 ORF2 gene, partial cds; and ORF1 gene, complete cds |
| FJ392108.1 | Torque teno virus isolate TW53A29 ORF2 gene, partial cds; and ORF1 gene, complete cds |
| FJ392111.1 | Torque teno virus isolate TW53A35 ORF2 gene, partial cds; and ORF1 gene, complete cds |
| FJ392112.1 | Torque teno virus isolate TW53A39 ORF2 gene, partial cds; and ORF1 gene, complete cds |
| FJ392113.1 | Torque teno virus isolate TW53A26 ORF2 gene, complete cds; and nonfunctional ORF1 gene, complete sequence |
| FJ392114.1 | Torque teno virus isolate TW53A30 ORF2 and ORF1 genes, complete cds |
| FJ392115.1 | Torque teno virus isolate TW53A31 ORF2 and ORF1 genes, complete cds |
| FJ392117.1 | Torque teno virus isolate TW53A37 ORF1 gene, complete cds |
| FJ426280.1 | Torque teno virus strain SIA109, complete genome |
| FR751500.1 | Torque teno virus complete genome, isolate TTV-HD23a (rheu215) |
| GU797360.1 | Torque teno virus clone 8-17, complete genome |
| HC742700.1 | Sequence 7 from Patent WO2010044889 |
| HC742710.1 | Sequence 17 from Patent WO2010044889 |
| JX134044.1 | TTV-like mini virus isolate TTMV_LY1, complete genome |
| JX134045.1 | TTV-like mini virus isolate TTMV_LY2, complete genome |
| KU243129.1 | TTV-like mini virus isolate TTMV-204, complete genome |
| KY856742.1 | TTV-like mini virus isolate zhenjiang, complete genome |
| LC381845.1 | Torque teno virus Human/Japan/KS025/2016 DNA, complete genome |
| MH648892.1 | Anelloviridae sp. isolate ctdc048, complete genome |
| MH648893.1 | Anelloviridae sp. isolate ctdh007, complete genome |
| MH648897.1 | Anelloviridae sp. isolate ctcb038, complete genome |
| MH648900.1 | Anelloviridae sp. isolate ctfc019, complete genome |
| MH648901.1 | Anelloviridae sp. isolate ctbb022, complete genome |
| MH648907.1 | Anelloviridae sp. isolate ctcf040, complete genome |
| MH648911.1 | Anelloviridae sp. isolate cthi018, complete genome |
| MH648912.1 | Anelloviridae sp. isolate ctea38, complete genome |
| MH648913.1 | Anelloviridae sp. isolate ctbg006, complete genome |
| MH648916.1 | Anelloviridae sp. isolate ctbg020, complete genome |

| | |
|---|---|
| MH648925.1 | Anelloviridae sp. isolate ctci019, complete genome |
| MH648932.1 | Anelloviridae sp. isolate ctid031, complete genome |
| MH648946.1 | Anelloviridae sp. isolate ctdb017, complete genome |
| MH648957.1 | Anelloviridae sp. isolate ctch017, complete genome |
| MH648958.1 | Anelloviridae sp. isolate ctbh011, complete genome |
| MH648959.1 | Anelloviridae sp. isolate ctbc020, complete genome |
| MH648962.1 | Anelloviridae sp. isolate ctif015, complete genome |
| MH648966.1 | Anelloviridae sp. isolate ctei055, complete genome |
| MH648969.1 | Anelloviridae sp. isolate ctjg000, complete genome |
| MH648976.1 | Anelloviridae sp. isolate ctcj064, complete genome |
| MH648977.1 | Anelloviridae sp. isolate ctbj022, complete genome |
| MH648982.1 | Anelloviridae sp. isolate ctbf014, complete genome |
| MH648983.1 | Anelloviridae sp. isolate ctbd027, complete genome |
| MH648985.1 | Anelloviridae sp. isolate ctch016, complete genome |
| MH648986.1 | Anelloviridae sp. isolate ctbd020, complete genome |
| MH648989.1 | Anelloviridae sp. isolate ctga035, complete genome |
| MH648990.1 | Anelloviridae sp. isolate cthf001, complete genome |
| MH648995.1 | Anelloviridae sp. isolate ctbd067, complete genome |
| MH648997.1 | Anelloviridae sp. isolate ctce026, complete genome |
| MH648999.1 | Anelloviridae sp. isolate ctfb058, complete genome |
| MH649002.1 | Anelloviridae sp. isolate ctjj046, complete genome |
| MH649006.1 | Anelloviridae sp. isolate ctcf030, complete genome |
| MH649008.1 | Anelloviridae sp. isolate ctbg025, complete genome |
| MH649011.1 | Anelloviridae sp. isolate ctbh052, complete genome |
| MH649014.1 | Anelloviridae sp. isolate ctba003, complete genome |
| MH649017.1 | Anelloviridae sp. isolate ctbb016, complete genome |
| MH649022.1 | Anelloviridae sp. isolate ctch023, complete genome |
| MH649023.1 | Anelloviridae sp. isolate ctbd051, complete genome |
| MH649028.1 | Anelloviridae sp. isolate ctbf9, complete genome |
| MH649038.1 | Anelloviridae sp. isolate ctbi030, complete genome |
| MH649039.1 | Anelloviridae sp. isolate ctca057, complete genome |
| MH649040.1 | Anelloviridae sp. isolate ctch033, complete genome |
| MH649042.1 | Anelloviridae sp. isolate ctjd005, complete genome |

| | |
|---|---|
| MH649045.1 | Anelloviridae sp. isolate ctdc021, complete genome |
| MH649051.1 | Anelloviridae sp. isolate ctdg044, complete genome |
| MH649056.1 | Anelloviridae sp. isolate ctcc062, complete genome |
| MH649061.1 | Anelloviridae sp. isolate ctid009, complete genome |
| MH649062.1 | Anelloviridae sp. isolate ctdc018, complete genome |
| MH649063.1 | Anelloviridae sp. isolate ctbf012, complete genome |
| MH649068.1 | Anelloviridae sp. isolate ctcc066, complete genome |
| MH649070.1 | Anelloviridae sp. isolate ctda011, complete genome |
| MH649077.1 | Anelloviridae sp. isolate ctbh034, complete genome |
| MH649083.1 | Anelloviridae sp. isolate ctdg028, complete genome |
| MH649084.1 | Anelloviridae sp. isolate ctii061, complete genome |
| MH649085.1 | Anelloviridae sp. isolate cteh021, complete genome |
| MH649092.1 | Anelloviridae sp. isolate ctbg012, complete genome |
| MH649101.1 | Anelloviridae sp. isolate ctif053, complete genome |
| MH649104.1 | Anelloviridae sp. isolate ctei657, complete genome |
| MH649106.1 | Anelloviridae sp. isolate ctca015, complete genome |
| MH649114.1 | Anelloviridae sp. isolate ctbf050, complete genome |
| MH649122.1 | Anelloviridae sp. isolate ctdc002, complete genome |
| MH649125.1 | Anelloviridae sp. isolate ctbb15, complete genome |
| MH649127.1 | Anelloviridae sp. isolate ctba013, complete genome |
| MH649137.1 | Anelloviridae sp. isolate ctbb000, complete genome |
| MH649141.1 | Anelloviridae sp. isolate ctbc019, complete genome |
| MH649142.1 | Anelloviridae sp. isolate ctid026, complete genome |
| MH649144.1 | Anelloviridae sp. isolate ctfj004, complete genome |
| MH649152.1 | Anelloviridae sp. isolate ctcj13, complete genome |
| MH649156.1 | Anelloviridae sp. isolate ctci006, complete genome |
| MH649157.1 | Anelloviridae sp. isolate ctbd025, complete genome |
| MH649158.1 | Anelloviridae sp. isolate ctbf005, complete genome |
| MH649161.1 | Anelloviridae sp. isolate ctcf045, complete genome |
| MH649165.1 | Anelloviridae sp. isolate ctcc29, complete genome |
| MH649169.1 | Anelloviridae sp. isolate ctib021, complete genome |
| MH649172.1 | Anelloviridae sp. isolate ctbh857, complete genome |
| MH649174.1 | Anelloviridae sp. isolate ctbj049, complete genome |

| | |
|---|---|
| MH649178.1 | Anelloviridae sp. isolate ctfc006, complete genome |
| MH649179.1 | Anelloviridae sp. isolate ctbe000, complete genome |
| MH649183.1 | Anelloviridae sp. isolate ctbb031, complete genome |
| MH649186.1 | Anelloviridae sp. isolate ctcb33, complete genome |
| MH649189.1 | Anelloviridae sp. isolate ctcc12, complete genome |
| MH649196.1 | Anelloviridae sp. isolate ctci060, complete genome |
| MH649199.1 | Anelloviridae sp. isolate ctbb017, complete genome |
| MH649203.1 | Anelloviridae sp. isolate cthc018, complete genome |
| MH649204.1 | Anelloviridae sp. isolate ctbj003, complete genome |
| MH649206.1 | Anelloviridae sp. isolate ctbg010, complete genome |
| MH649208.1 | Anelloviridae sp. isolate ctid008, complete genome |
| MH649209.1 | Anelloviridae sp. isolate ctbg056, complete genome |
| MH649210.1 | Anelloviridae sp. isolate ctda001, complete genome |
| MH649212.1 | Anelloviridae sp. isolate ctcf004, complete genome |
| MH649217.1 | Anelloviridae sp. isolate ctbe029, complete genome |
| MH649223.1 | Anelloviridae sp. isolate ctci016, complete genome |
| MH649224.1 | Anelloviridae sp. isolate ctce11, complete genome |
| MH649228.1 | Anelloviridae sp. isolate ctcf013, complete genome |
| MH649229.1 | Anelloviridae sp. isolate ctcb036, complete genome |
| MH649241.1 | Anelloviridae sp. isolate ctda027, complete genome |
| MH649242.1 | Anelloviridae sp. isolate ctbf003, complete genome |
| MH649254.1 | Anelloviridae sp. isolate ctjb007, complete genome |
| MH649255.1 | Anelloviridae sp. isolate ctbb023, complete genome |
| MH649256.1 | Anelloviridae sp. isolate ctca002, complete genome |
| MH649258.1 | Anelloviridae sp. isolate ctcg010, complete genome |
| MH649263.1 | Anelloviridae sp. isolate ctgh3, complete genome |
| MK012439.1 | Anelloviridae sp. isolate cthe000, complete genome |
| MK012440.1 | Anelloviridae sp. isolate ctjd008, complete genome |
| MK012448.1 | Anelloviridae sp. isolate ctch012, complete genome |
| MK012457.1 | Anelloviridae sp. isolate ctda009, complete genome |
| MK012458.1 | Anelloviridae sp. isolate ctcd015, complete genome |
| MK012485.1 | Anelloviridae sp. isolate ctfd011, complete genome |
| MK012489.1 | Anelloviridae sp. isolate ctba003, complete genome |

| | |
|---|---|
| MK012492.1 | Anelloviridae sp. isolate ctbb005, complete genome |
| MK012493.1 | Anelloviridae sp. isolate ctcj014, complete genome |
| MK012500.1 | Anelloviridae sp. isolate ctcb001, complete genome |
| MK012504.1 | Anelloviridae sp. isolate ctcj010, complete genome |
| MK012516.1 | Anelloviridae sp. isolate ctcf003, complete genome |
| NC_038336.1 | Torque teno virus 5 isolate TCHN-C1 Orf2 and Orf1 genes, complete cds |
| NC_038338.1 | Torque teno virus 11 isolate TCHN-D1 Orf2 and Orf1 genes, complete cds |
| NC_038339.1 | Torque teno virus 13 isolate TCHN-A Orf2 and Orf1 genes, complete cds |
| NC_038340.1 | Torque teno virus 20 ORF4, ORF3, ORF2, ORF1 genes, complete cds, clone: SAa-10 |
| NC_038341.1 | Torque teno virus 21 isolate TCHN-B ORF2 and ORF1 genes, complete cds |
| NC_038342.1 | Torque teno virus 23 ORF2, ORF1 genes, complete cds, isolate: s-TTV CH65-2 |
| NC_038343.1 | Torque teno virus 24 ORF4, ORF3, ORF2, ORF1 genes, complete cds, clone: SAa-01 |
| NC_038344.1 | Torque teno virus 29 ORF2, ORF1, ORF3 genes, complete cds, isolate: TTVyon-KC009 |
| NC_038345.1 | Torque teno mini virus 10 isolate LIL-y1 ORF2, ORF1, ORF3, and ORF4 genes, complete cds |
| NC_038346.1 | Torque teno mini virus 11 isolate LIL-y2 ORF2, ORF1, and ORF3 genes, complete cds |
| NC_038347.1 | Torque teno mini virus 12 isolate LIL-y3 ORF2, ORF1, ORF3, and ORF4 genes, complete cds |
| NC_038350.1 | Torque teno midi virus 3 isolate 2PoSMA ORF2 and ORF1 genes, complete cds |
| NC_038351.1 | Torque teno midi virus 4 isolate 6PoSMA ORF2, ORF1, and ORF3 genes, complete cds |
| NC_038352.1 | Torque teno midi virus 5 DNA, complete genome, isolate: MDJHem2 |
| NC_038353.1 | Torque teno midi virus 6 DNA, complete genome, isolate: MDJHem3-1 |
| NC_038354.1 | Torque teno midi virus 7 DNA, complete genome, isolate: MDJHem3-2 |
| NC_038355.1 | Torque teno midi virus 8 DNA, complete genome, isolate: MDJN1 |
| NC_038356.1 | Torque teno midi virus 9 DNA, complete genome, isolate: MDJN2 |
| NC_038357.1 | Torque teno midi virus 10 DNA, complete genome, isolate: MDJN14 |
| NC_038358.1 | Torque teno midi virus 11 DNA, complete genome, isolate: MDJN47 |
| NC_038359.1 | Torque teno midi virus 12 DNA, complete genome, isolate: MDJN51 |

| | |
|---|---|
| NC_038360.1 | Torque teno midi virus 13 DNA, complete genome, isolate: MDJN69 |
| NC_038361.1 | Torque teno midi virus 14 DNA, complete genome, isolate: MDJN97 |
| NC_038362.1 | Torque teno midi virus 15 DNA, complete genome, isolate: Pt-TTMDV210 |

In some embodiments, the genetic element comprises one or more sequences with homology or identity to one or more sequences from one or more non-*Anelloviruses*, e.g., adenovirus, herpes virus, pox virus, vaccinia virus, SV40, papilloma virus, an RNA virus such as a retrovirus, e.g., lentivirus, a single-stranded RNA virus, e.g., hepatitis virus, or a double-stranded RNA virus e.g., rotavirus. Since, in some embodiments, recombinant retroviruses are defective, assistance may be provided order to produce infectious particles. Such assistance can be provided, e.g., by using helper cell lines that contain plasmids encoding all of the structural genes of the retrovirus under the control of regulatory sequences within the LTR. Suitable cell lines for replicating the anellovectors described herein include cell lines known in the art, e.g., A549 cells, which can be modified as described herein. Said genetic element can additionally contain a gene encoding a selectable marker so that the desired genetic elements can be identified.

In some embodiments, the genetic element includes non-silent mutations, e.g., base substitutions, deletions, or additions resulting in amino acid differences in the encoded polypeptide, so long as the sequence remains at least about 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to the polypeptide encoded by the first nucleotide sequence or otherwise is useful for practicing the present invention. In this regard, certain conservative amino acid substitutions may be made which are generally recognized not to inactivate overall protein function: such as in regard of positively charged amino acids (and vice versa), lysine, arginine and histidine; in regard of negatively charged amino acids (and vice versa), aspartic acid and glutamic acid; and in regard of certain groups of neutrally charged amino acids (and in all cases, also vice versa), (1) alanine and serine, (2) asparagine, glutamine, and histidine, (3) cysteine and serine, (4) glycine and proline, (5) isoleucine, leucine and valine, (6) methionine, leucine and isoleucine, (7) phenylalanine, methionine, leucine, and tyrosine, (8) serine and threonine, (9) tryptophan and tyrosine, (10) and for example tyrosine, tryptophan and phenylalanine. Amino acids can be classified according to physical properties and contribution to secondary and tertiary protein structure. A conservative substitution is recognized in the art as a substitution of one amino acid for another amino acid that has similar properties.

Identity of two or more nucleic acid or polypeptide sequences having the same or a specified percentage of nucleotides or amino acid residues that are the same (e.g., about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or higher identity over a specified region, when compared and aligned for maximum correspondence over a comparison window or designated region) may be measured using a BLAST or BLAST 2.0 sequence comparison algorithms

with default parameters described below, or by manual alignment and visual inspection (see, e.g., NCBI web site www.ncbi.nlm.nih.gov/BLAST/ or the like). Identity may also refer to, or may be applied to, the compliment of a test sequence. Identity also includes sequences that have deletions and/or additions, as well as those that have substitutions. As described herein, the algorithms account for gaps and the like. Identity may exist over a region that is at least about 10 amino acids or nucleotides in length, about 15 amino acids or nucleotides in length, about 20 amino acids or nucleotides in length, about 25 amino acids or nucleotides in length, about 30 amino acids or nucleotides in length, about 35 amino acids or nucleotides in length, about 40 amino acids or nucleotides in length, about 45 amino acids or nucleotides in length, about 50 amino acids or nucleotides in length, or more. Since the genetic code is degenerate, a homologous nucleotide sequence can include any number of silent base changes, i.e., nucleotide substitutions that nonetheless encode the same amino acid.

***Proteinaceous Exterior***

In some embodiments, the anellovector, e.g., synthetic anellovector, comprises a proteinaceous exterior that encloses the genetic element. The proteinaceous exterior can comprise a substantially non-pathogenic exterior protein that fails to elicit an unwanted immune response in a mammal. The proteinaceous exterior of the anellovectors typically comprises a substantially non-pathogenic protein that may self-assemble into an icosahedral formation that makes up the proteinaceous exterior.

In some embodiments, the proteinaceous exterior protein is encoded by a sequence of the genetic element of the anellovector (e.g., is in cis with the genetic element). In other embodiments, the proteinaceous exterior protein is encoded by a nucleic acid separate from the genetic element of the anellovector (e.g., is in trans with the genetic element).

In some embodiments, the protein, e.g., substantially non-pathogenic protein and/or proteinaceous exterior protein, comprises one or more glycosylated amino acids, e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, or more.

In some embodiments, the protein, e.g., substantially non-pathogenic protein and/or proteinaceous exterior protein comprises at least one hydrophilic DNA-binding region, an arginine-rich region, a threonine-rich region, a glutamine-rich region, a N-terminal polyarginine sequence, a variable region, a C-terminal polyglutamine/glutamate sequence, and one or more disulfide bridges.

In some embodiments, the protein is a capsid protein, e.g., has a sequence having at least about 60%, 65%, 70%, 75%, 80%, 85%, 90% 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to a protein encoded by any one of the nucleotide sequences encoding a capsid protein described herein, e.g., an Anellovirus ORF1 molecule and/or capsid protein sequence, e.g., as described herein. In some embodiments, the protein or a functional fragment of a capsid protein is encoded by a nucleotide

sequence having at least about 60%, 70% 80%, 85%, 90% 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an Anellovirus ORF1 nucleic acid, e.g., as described herein.

In some embodiments, the anellovector comprises a nucleotide sequence encoding a capsid protein or a functional fragment of a capsid protein or a sequence having at least about 60%, 70% 80%, 85%, 90% 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an Anellovirus ORF1 molecule as described herein.

In some embodiments, the ranges of amino acids with less sequence identity may provide one or more of the properties described herein and differences in cell/tissue/species specificity (e.g. tropism).

In some embodiments, the anellovector lacks lipids in the proteinaceous exterior. In some embodiments, the anellovector lacks a lipid bilayer, e.g., a viral envelope. In some embodiments, the interior of the anellovector is entirely covered (e.g., 100% coverage) by a proteinaceous exterior. In some embodiments, the interior of the anellovector is less than 100% covered by the proteinaceous exterior, e.g., 95%, 90%, 85%, 80%, 70%, 60%, 50% or less coverage. In some embodiments, the proteinaceous exterior comprises gaps or discontinuities, e.g., permitting permeability to water, ions, peptides, or small molecules, so long as the genetic element is retained in the anellovector.

In some embodiments, the proteinaceous exterior comprises one or more proteins or polypeptides that specifically recognize and/or bind a host cell, e.g., a complementary protein or polypeptide, to mediate entry of the genetic element into the host cell.

In some embodiments, the proteinaceous exterior comprises one or more of the following: an arginine-rich region, jelly-roll region, N22 domain, hypervariable region, and/or C-terminal domain, e.g., of an ORF1 molecule, e.g., as described herein. In some embodiments, the proteinaceous exterior comprises one or more of the following: one or more glycosylated proteins, a hydrophilic DNA-binding region, an arginine-rich region, a threonine-rich region, a glutamine-rich region, a N-terminal polyarginine sequence, a variable region, a C-terminal polyglutamine/glutamate sequence, and one or more disulfide bridges. For example, the proteinaceous exterior comprises a protein encoded by an Anellovirus ORF1 nucleic acid, e.g., as described herein.

In some embodiments, the proteinaceous exterior comprises one or more of the following characteristics: an icosahedral symmetry, recognizes and/or binds a molecule that interacts with one or more host cell molecules to mediate entry into the host cell, lacks lipid molecules, lacks carbohydrates, is pH and temperature stable, is detergent resistant, and is substantially non-immunogenic or non-pathogenic in a host.

## III. Methods of use

The anellovectors and compositions comprising anellovectors described herein may be used in methods of treating a disease, disorder, or condition, e.g., in a subject (e.g., a mammalian subject, e.g., a human subject) in need thereof. Administration of a pharmaceutical composition described herein may be, for example, by way of parenteral (including intravenous, intratumoral, intraperitoneal, intramuscular, intracavity, and subcutaneous) administration. The anellovectors may be administered alone or formulated as a pharmaceutical composition.

The anellovectors may be administered in the form of a unit-dose composition, such as a unit dose parenteral composition. Such compositions are generally prepared by admixture and can be suitably adapted for parenteral administration. Such compositions may be, for example, in the form of injectable and infusable solutions or suspensions or suppositories or aerosols.

In some embodiments, administration of an anellovector or composition comprising same, e.g., as described herein, may result in delivery of a genetic element comprised by the anellovector to a target cell, e.g., in a subject.

An anellovector or composition thereof described herein, e.g., comprising an effector (e.g., an endogenous or exogenous effector), may be used to deliver the effector to a cell, tissue, or subject. In some embodiments, the anellovector or composition thereof is used to deliver the effector to bone marrow, blood, heart, GI or skin. Delivery of an effector by administration of an anellovector composition described herein may modulate (e.g., increase or decrease) expression levels of a noncoding RNA or polypeptide in the cell, tissue, or subject. Modulation of expression level in this fashion may result in alteration of a functional activity in the cell to which the effector is delivered. In some embodiments, the modulated functional activity may be enzymatic, structural, or regulatory in nature.

In some embodiments, the anellovector, or copies thereof, are detectable in a cell 24 hours (e.g., 1 day, 2 days, 3 days, 4 days, 5 days, 6 days, 1 week, 2 weeks, 3 weeks, 4 weeks, 30 days, or 1 month) after delivery into a cell. In embodiments, a anellovector or composition thereof mediates an effect on a target cell, and the effect lasts for at least 1, 2, 3, 4, 5, 6, or 7 days, 2, 3, or 4 weeks, or 1, 2, 3, 6, or 12 months. In some embodiments (e.g., wherein the anellovector or composition thereof comprises a genetic element encoding an exogenous protein), the effect lasts for less than 1, 2, 3, 4, 5, 6, or 7 days, 2, 3, or 4 weeks, or 1, 2, 3, 6, or 12 months.

Examples of diseases, disorders, and conditions that can be treated with the anellovector described herein, or a composition comprising the anellovector, include, without limitation: immune disorders, interferonopathies (e.g., Type I interferonopathies), infectious diseases, inflammatory disorders, autoimmune conditions, cancer (e.g., a solid tumor, e.g., lung cancer, non-small cell lung cancer, e.g., a tumor that expresses a gene responsive to mIR-625, e.g., caspase-3), and gastrointestinal disorders. In

some embodiments, the anellovector modulates (e.g., increases or decreases) an activity or function in a cell with which the anellovector is contacted. In some embodiments, the anellovector modulates (e.g., increases or decreases) the level or activity of a molecule (e.g., a nucleic acid or a protein) in a cell with which the anellovector is contacted. In some embodiments, the anellovector decreases viability of a cell, e.g., a cancer cell, with which the anellovector is contacted, e.g., by at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, or more. In some embodiments, the anellovector comprises an effector, e.g., an miRNA, e.g., miR-625, that decreases viability of a cell, e.g., a cancer cell, with which the anellovector is contacted, e.g., by at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, or more. In some embodiments, the anellovector modulates (e.g., increases or decreases) an activity or function in a cell with which the anellovector is contacted. In some embodiments, the anellovector modulates (e.g., increases or decreases) the level or activity of a molecule (e.g., a nucleic acid or a protein) in a cell with which the anellovector is contacted. In some embodiments, the anellovector decreases viability of a cell, e.g., a cancer cell, with which the anellovector is contacted, e.g., by at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, or more. In some embodiments, the anellovector comprises an effector, e.g., an miRNA, e.g., miR-625, that increases apoptosis of a cell, e.g., a cancer cell, e.g., by increasing caspase-3 activity, with which the anellovector is contacted, e.g., by at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, or more.

## IV. Administration/Delivery

The composition (e.g., a pharmaceutical composition comprising an anellovector as described herein) may be formulated to include a pharmaceutically acceptable excipient. Pharmaceutical compositions may optionally comprise one or more additional active substances, e.g. therapeutically and/or prophylactically active substances. Pharmaceutical compositions of the present invention may be sterile and/or pyrogen-free. General considerations in the formulation and/or manufacture of pharmaceutical agents may be found, for example, in Remington: The Science and Practice of Pharmacy 21st ed., Lippincott Williams & Wilkins, 2005 (incorporated herein by reference).

Although the descriptions of pharmaceutical compositions provided herein are principally directed to pharmaceutical compositions which are suitable for administration to humans, it will be understood by the skilled artisan that such compositions are generally suitable for administration to any other animal, e.g., to non-human animals, e.g. non-human mammals. Modification of pharmaceutical compositions suitable for administration to humans in order to render the compositions suitable for administration to various animals is well understood, and the ordinarily skilled veterinary pharmacologist can design and/or perform such modification with merely ordinary, if any, experimentation. Subjects to which administration of the pharmaceutical compositions is contemplated include, but are not limited to, humans and/or other primates; mammals, including commercially relevant mammals such as cattle, pigs, horses, sheep, cats, dogs, mice, and/or rats; and/or birds, including commercially relevant birds such as poultry, chickens, ducks, geese, and/or turkeys.

Formulations of the pharmaceutical compositions described herein may be prepared by any method known or hereafter developed in the art of pharmacology. In general, such preparatory methods include the step of bringing the active ingredient into association with an excipient and/or one or more other accessory ingredients, and then, if necessary and/or desirable, dividing, shaping and/or packaging the product.

In one aspect, the invention features a method of delivering an anellovector to a subject. The method includes administering a pharmaceutical composition comprising an anellovector as described herein to the subject. In some embodiments, the administered anellovector replicates in the subject (e.g., becomes a part of the virome of the subject).

The pharmaceutical composition may include wild-type or native viral elements and/or modified viral elements. The anellovector may include one or more Anellovirus sequences (e.g., nucleic acid sequences or nucleic acid sequences encoding amino acid sequences thereof) or a sequence with at least about 60%, 65%, 70%, 75%, 80%, 85%, 90% 95%, 96%, 97%, 98% and 99% nucleotide sequence identity thereto. The anellovector may comprise a nucleic acid molecule comprising a nucleic acid sequence with at least about 60%, 65%, 70%, 75%, 80%, 85%, 90% 95%, 96%, 97%, 98% and 99% sequence identity to one or more Anellovirus sequences (e.g., an Anellovirus ORF1 nucleic acid sequence). The anellovector may comprise a nucleic acid molecule encoding an amino acid sequence with at least about 60%, 65%, 70%, 75%, 80%, 85%, 90% 95%, 96%, 97%, 98% and 99% sequence identity to an Anellovirus amino acid sequence (e.g., the amino acid sequence of an Anellovirus ORF1 molecule). The anellovector may comprise a polypeptide comprising an amino acid sequence with at least about 60%, 65%, 70%, 75%, 80%, 85%, 90% 95%, 96%, 97%, 98% and 99% sequence identity to an Anellovirus amino acid sequence (e.g., the amino acid sequence of an Anellovirus ORF1 molecule).

In some embodiments, the anellovector is sufficient to increase (stimulate) endogenous gene and protein expression, e.g., at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, or more as compared to a reference, e.g., a healthy control. In certain embodiments, the anellovector is sufficient to decrease (inhibit) endogenous gene and protein expression, e.g., at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, or more as compared to a reference, e.g., a healthy control.

In some embodiments, the anellovector inhibits/enhances one or more viral properties, e.g., tropism, infectivity, immunosuppression/activation, in a host or host cell, e.g., at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, or more as compared to a reference, e.g., a healthy control.

In some embodiments, the subject is administered the pharmaceutical composition further comprising one or more viral strains that are not represented in the viral genetic information.

In some embodiments, the pharmaceutical composition comprising an anellovector described herein is administered in a dose and time sufficient to modulate a viral infection. Some non-limiting

examples of viral infections include adeno-associated virus, Aichi virus, Australian bat lyssavirus, BK polyomavirus, Banna virus, Barmah forest virus, Bunyamwera virus, Bunyavirus La Crosse, Bunyavirus snowshoe hare, Cercopithecine herpesvirus, Chandipura virus, Chikungunya virus, Cosavirus A, Cowpox virus, Coxsackievirus, Crimean-Congo hemorrhagic fever virus, Dengue virus, Dhori virus, Dugbe virus, Duvenhage virus, Eastern equine encephalitis virus, Ebolavirus, Echovirus, Encephalomyocarditis virus, Epstein-Barr virus, European bat lyssavirus, GB virus C/Hepatitis G virus, Hantaan virus, Hendra virus, Hepatitis A virus, Hepatitis B virus, Hepatitis C virus, Hepatitis E virus, Hepatitis delta virus, Horsepox virus, Human adenovirus, Human astrovirus, Human coronavirus, Human cytomegalovirus, Human enterovirus 68, Human enterovirus 70, Human herpesvirus 1, Human herpesvirus 2, Human herpesvirus 6, Human herpesvirus 7, Human herpesvirus 8, Human immunodeficiency virus, Human papillomavirus 1, Human papillomavirus 2, Human papillomavirus 16, Human papillomavirus 18, Human parainfluenza, Human parvovirus B19, Human respiratory syncytial virus, Human rhinovirus, Human SARS coronavirus, Human spumaretrovirus, Human T-lymphotropic virus, Human torovirus, Influenza A virus, Influenza B virus, Influenza C virus, Isfahan virus, JC polyomavirus, Japanese encephalitis virus, Junin arenavirus, KI Polyomavirus, Kunjin virus, Lagos bat virus, Lake Victoria marburgvirus, Langat virus, Lassa virus, Lordsdale virus, Louping ill virus, Lymphocytic choriomeningitis virus, Machupo virus, Mayaro virus, MERS coronavirus, Measles virus, Mengo encephalomyocarditis virus, Merkel cell polyomavirus, Mokola virus, Molluscum contagiosum virus, Monkeypox virus, Mumps virus, Murray valley encephalitis virus, New York virus, Nipah virus, Norwalk virus, O'nyong-nyong virus, Orf virus, Oropouche virus, Pichinde virus, Poliovirus, Punta toro phlebovirus, Puumala virus, Rabies virus, Rift valley fever virus, Rosavirus A, Ross river virus, Rotavirus A, Rotavirus B, Rotavirus C, Rubella virus, Sagiyama virus, Salivirus A, Sandfly fever sicilian virus, Sapporo virus, Semliki forest virus, Seoul virus, Simian foamy virus, Simian virus 5, Sindbis virus, Southampton virus, St. louis encephalitis virus, Tick-borne powassan virus, Torque teno virus, Toscana virus, Uukuniemi virus, Vaccinia virus, Varicella-zoster virus, Variola virus, Venezuelan equine encephalitis virus, Vesicular stomatitis virus, Western equine encephalitis virus, WU polyomavirus, West Nile virus, Yaba monkey tumor virus, Yaba-like disease virus, Yellow fever virus, and Zika Virus. In certain embodiments, the anellovector is sufficient to outcompete and/or displace a virus already present in the subject, e.g., at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, or more as compared to a reference. In certain embodiments, the anellovector is sufficient to compete with chronic or acute viral infection. In certain embodiments, the anellovector may be administered prophylactically to protect from viral infections (e.g. a provirotic). In some embodiments, the anellovector is in an amount sufficient to modulate (e.g., phenotype, virus levels, gene expression, compete with other viruses, disease state, etc. at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, or more).In some embodiments, treatment, treating, and cognates thereof

comprise medical management of a subject (e.g., by administering an anellovector, e.g., an anellovector made as described herein), e.g., with the intent to improve, ameliorate, stabilize, prevent or cure a disease, pathological condition, or disorder. In some embodiments, treatment comprises active treatment (treatment directed to improve the disease, pathological condition, or disorder), causal treatment (treatment directed to the cause of the associated disease, pathological condition, or disorder), palliative treatment (treatment designed for the relief of symptoms), preventative treatment (treatment directed to preventing, minimizing or partially or completely inhibiting the development of the associated disease, pathological condition, or disorder), and/or supportive treatment (treatment employed to supplement another therapy).

All references and publications cited herein are hereby incorporated by reference.

The following examples are provided to further illustrate some embodiments of the present invention, but are not intended to limit the scope of the invention; it will be understood by their exemplary nature that other procedures, methodologies, or techniques known to those skilled in the art may alternatively be used.

## EXAMPLES

**Table of Contents**

Example 1: *In vitro* assembly of anellovectors using ORF1 produced using baculovirus

Example 2: In vitro assembly of Ring2 ORF1-based anellovectors encapsidating mRNA

Example 3: In vitro assembly of an mRNA-encapsidating anellovector using a modified ORF1 protein

Example 4: In vitro assembly of an mRNA-encapsidating anellovector using a modified mRNA

Example 5: Structural analysis of anellovirus ORF1 capsid proteins

### Example 1: *In vitro* assembly of anellovectors using ORF1 produced using baculovirus

In this example, baculovirus constructs suitable for expression of Anellovirus proteins (e.g., ORF1) were generated by *in vitro* assembly.

In a first example, DNA encoding Ring2 ORF1 fused to an N-terminal $HIS_6$-tag (HIS-ORF1) was codon optimized for insect expression and cloned into the baculovirus expression vector pFASTbac system according to manufacturer instructions (ThermoFisher Scientific). 10 liters of insect cell culture

(Sf9) was infected with Ring2 HIS-ORF1 baculovirus and the cells were harvested 3-days post-infection by centrifugation. The cells were lysed and the lysis was purified using a chelating resin column using standard methods in the field. The elution fraction containing HIS-ORF1 was dialyzed and treated with DNAse to digest host cell DNA. The resulting material was purified again using a chelating resin column and fractions containing ORF1 were retained for nucleic acid encapsidation and viral vector purification. ORF1-containing fractions were also analyzed by negative staining electron microscopy.

In a second example, DNA encoding Ring10 ORF1 fused to an N-terminal $HIS_6$-tag (HIS-ORF1) was codon optimized for insect expression and cloned into the baculovirus expression vector pFASTbac system according to manufacturer instructions (ThermoFisher Scientific). Insect cells (Sf9) were infected with Ring HIS-ORF1 baculovirus and the cells were harvested 3-days post-infection by centrifugation. The cells were lysed and the protein was purified using a chelating resin affinity column (HisTrap, GE Healthcare) using standard methods in the field. The resulting material was purified again using a heparin affinity column (Heparin HiTrap, GE Healthcare) and fractions containing ORF1 were analyzed by negative staining electron microscopy.

In a third example, DNA encoding chicken anemia virus (CAV) capsid protein (Vp1) fused to an N-terminal $HIS_6$-Flag-tag (HIS-Flag-Vp1) and helper protein (Vp2) were codon optimized for mammalian expression and cloned into a mammalian expression vector using a CMV promoter. Mammalian cells (293expi) were transfected with CAV Vp1 and Vp2 expression vectors. The cells were harvested 3-days post-infection by centrifugation. The cells were lysed and the lysis was purified using a the chelation and heparin purification process. The elution fraction containing Chicken Anemia Virus (CAV) Vp1 were analyzed by negative staining electron microscopy.

As shown in FIG. 1, both Ring 2 ORF1 and Ring 10 ORF1 showed a propensity to form ~35 nm virus-like particles.

**Nucleic acid encapsidation and viral vector purification:** Ring ORF1 (wildtype protein, chimeric protein, or fragments thereof) will be treated with conditions sufficient to dissociate VLPs or viral capsids to enable reassembly with nucleic acid cargo. Nucleic acid cargo can be defined, for example, RNA which encodes a gene of interest that one wants to deliver as a therapeutic agent. Nucleic acid cargo of defined concentration will be combined with Ring ORF1 of defined concentration and treated with conditions sufficient to permit nucleic acid encapsidation and the resulting particle, defined as viral vector, will be subsequently purified using standard viral purification procedures.

**Example 2: In vitro assembly of Ring2 ORF1-based anellovectors encapsidating mRNA**

Ring2 ORF1 is purified by size exclusion chromatography (SEC) with mobile phases including Tris pH 8.0 with 500 mM NaCl, Tris pH 8.0 with 500 mM NaCl with 0.1% SDS, CAPS buffer pH 10.5

with 150 mM NaCl, CAPS buffer pH 10.5 with 500 mM NaCl or CAPS buffer pH 10.5 with 500 mM NaCl with 0.1% SDS to dissociate viral particles or VLPs into dispersed protein or capsomers.

In a first example, the ORF1 is mixed with mRNA, a fluorescently labeled mRNA or an mRNA transgene chemically conjugated to a segment of ssDNA shown in Example 1 to be competent for inducing vector formation. Viral vectors are formed through dialysis and SEC purification using Tris pH 8.0 buffer to isolate the anellovector encapsidating RNA (e.g., as measured by retained fluorescent absorption). Anellovector assembly is further evaluated by biophysical assessment such as DLS or electron microscopy.

In a second example, purified ORF1 is treated with 1 M NaCl with 0.1% SDS dissociate oligomers or VLPs into dispersed protein or capsomers. ORF1 is then mixed with mRNA, such as an mRNA that translate a gene of interest (e.g., a reporter gene, e.g., GFP, mCherry; or an effector of interest, e.g., EPO), and dialyzed against Tris pH 8.0 with 150 mM NaCl to permit VLP formation. The subsequent complex is purified by SEC using Tris pH 8.0 buffer to isolate the AV vector encapsidating mRNA. Anellovector vector assembly may be further evaluated by *in vitro* or *in vivo* readout, for example, by transducing cells and observing the expression of the reporter gene (e.g., mCherry or GFP) or through expression of an effector of interest (e.g., using an ELISA to detect the expression of a gene, such as EPO).

*In vitro assembly of Ring2 ORF1-based Anellovectors encapsidating GFP mRNA*

In a further example, Ring2 ORF1 protein was expressed as a full-length protein in insect cells and assembled VLPs were purified by a heparin affinity column followed by size exclusion chromatography (SEC) using a Tris buffer mobile phase. VLPs formed from the isolated Ring2 ORF1 proteins were observed with negative staining electron microscopy (EM) and had an estimated particle titer of $10^{10}$ particles/ml (pts/ml; FIG. 4A). The VLPs were treated with 2 molar (2M) urea to disassemble the VLPs. Reimaging by EM showed no VLPs observed (FIG. 4B). Urea-treated VLPs were then dialyzed to remove urea either in the absence of mRNA (FIG. 4C) or in the presence of ~10x excess mRNA encoding GFP (FIGS. 4D and 4E). For VLP samples treated with urea and dialyzed in absence of mRNA, few particles (less than $10^8$ particles/ml; FIG. 4C) were observed by EM. In contrast, dialysis in the presence of excess mRNA resulted in the observation of substantially higher titers of particles (~$10^9$-$10^{10}$ particles/ml; FIGS. 4D and eE) by EM. These data demonstrate that disassembly and reassembly of VLPs is more efficient in the presence of mRNA, and indicate that Anellovirus ORF1 protein can be used to encapsidate mRNA *in vitro* to form anellovectors.

**Example 3: In vitro assembly of an mRNA-encapsidating anellovector using a modified ORF1 protein**

In this example, packaging of an mRNA genetic element is improved by modifying the ORF1 protein to harbor contact residues that bind mRNA. In this example, ssDNA contact residues and/or jellyroll beta strands that contact ssDNA and/or the N-terminal arginine-rich motif (ARM) can be replaced with components of an mRNA binding viral protein or other mRNA-binding protein to permit efficient binding and packaging of mRNA. This mRNA-binding chimeric ORF1 is then treated with 1 M NaCl with 0.1% SDS to dissociate oligomers or VLPs into dispersed protein or capsomers. The chimeric ORF1 is then mixed with mRNA, such as an mRNA that encodes a gene of interest (e.g., a reporter gene, e.g., GFP, mCherry; or an effector of interest, e.g., EPO), and dialyzed against Tris pH 8.0 with 150 mM NaCl to permit VLP formation. The subsequent complex is purified by SEC using Tris pH 8.0 buffer to isolate the anellovector encapsidating mRNA. Anellovector assembly can be further evaluated by *in vitro* or *in vivo* readout, for example, by transducing cells and observing the expression of the reporter gene (e.g., mCherry or GFP) or through expression of an effector of interest (such as using an ELISA to detect the expression of a gene such as EPO).

**Exemplary modifications to ORF1 molecules:** Ring ORF1 molecules that may be used in the methods described herein include, for example, several wildtype Anellovirus ORF1 proteins; CAV capsid protein (VP1) variants; Anellovirus ORF1 proteins harboring mutations to improve assembly efficiency, yield or stability; and chimeric ORF1 strains or functional fragments thereof. In some instances, affinity tags are attached to the ORF1 molecule, e.g., at the N-terminus (SEQ ID NOs: 561-562). In some instances, the ORF1 molecules are untagged proteins. Ring ORF1 molecules may be expressed alone or in combination with any number of helper proteins including, but not limited to, Anellovirus ORF2 and/or ORF3 proteins.

Ring ORF1 proteins harboring mutations to improve assembly efficiency may include, but are not limited to, ORF1 proteins that harbor mutations introduced into the N-terminal arginine-rich motif (ARM), for example, to alter the pI of the ARG arm, which may permit pH sensitive nucleic acid binding to trigger particle assembly (SEQ ID NOs: 563-565). ORF1 mutations that improve stability may include, for example, mutations to the interprotomer contacting beta strands F and G of the canonical jellyroll beta-barrel (F and G beta strands), e.g., to alter the hydrophobic state of the protomer surface and/or to make capsid formation more thermodynamically favored.

Chimeric ORF1 proteins may include, but are not limited to, ORF1 proteins which have a portion or portions of their sequence replaced with comparable portions from another capsid protein, such as BFDV capsid protein, Hepatitis E capsid protein (e.g., the ARG arm and/or F and G beta strands, or comparable components thereof). Chimeric ORF1 proteins may also include ORF1 proteins which have a

portion or portions of their sequence replaced with comparable portions of another Anellovirus ORF1 protein (such as jelly roll fragments or the C-terminal portion of Ring 2 ORF1 replaced with comparable portions of Ring 9 ORF1; see, e.g., SEQ ID NOs: 568-575).

Generally, ORF1 molecules can be purified using purification techniques including, but not limited to, chelating purification, heparin purification, gradient sedimentation purification and/or SEC purification.

**Example 4: In vitro assembly of an mRNA-encapsidating anellovector using a modified mRNA**

In this example, encapsidation of an mRNA-based genetic element is optimized by binding the mRNA molecule to ssDNA or by modifying the mRNA transgene in such a way that that a section of the backbone would permit binding to the ssDNA contact residues of wildtype ORF1. The mRNA generally encodes a gene of interest, such as a reporter gene (e.g., GFP or mCherry), and/or an effector gene (e.g., EPO).

In one example, modified ssDNA that can bind ORF1 by virtue of its sugar-chain backbone, but which can also pair with mRNA non-covalently, is mixed with an mRNA of interest to produce an mRNA/DNA complex. This mRNA/DNA complex can then be encapsidated using a Ring ORF1 to form an anellovector, for example, as described below.

In another example, an mRNA molecule is synthesized with a section or sections of the mRNA molecule harboring a DNA backbone permitting binding and encapsidation with ORF1, while retaining the portion of the mRNA that encodes a gene (e.g., a reporter gene or an effector gene) to be delivered. This mRNA/DNA hybrid molecule can then be encapsidated using a Ring ORF1 to form an anellovector, for example, as described below.

**Encapsidation by *in vitro* assembly:** The mRNA/DNA genetic elements described above are then encapsidated by in vitro assembly. Briefly, anellovector ORF1 is then treated with 1 M NaCl with 0.1% SDS to dissociate oligomers or VLPs into dispersed protein or capsomers. The ORF1 is then mixed with the synthetic mRNA complexes or hybrid molecules and dialyzed against Tris pH 8.0 with 150 mM NaCl to permit VLP formation. The subsequent particle is purified by SEC using Tris pH 8.0 buffer to isolate the anellovector encapsidating mRNA. Anellovector assembly could be further evaluated by *in vitro* or *in vivo* readout by transducing cells and observing the expression of the reporter gene or effector gene, e.g., as described herein.

**Example 5: Structural analysis of anellovirus ORF1 capsid proteins**

Anelloviruses share predicted structural features to other well-characterized viruses such as the avian pathogens Beak and Feather Disease Virus (BFDV) or Chicken Anemia Virus (CAV).

Anellovirus ORF1 capsid proteins contain an N-terminal ARM sequence similar to that of BFDV. Secondary structure prediction showed that the first ~250 residues of ORF1, dependent on strain, includes 8 predicted beta strands (FIG. 3). When the 8 predicted beta strands of ORF1, named B through I following jelly roll domain naming conventions, are aligned guided by the secondary structure prediction to the capsid proteins of BFDV and Hep E, conserved lysine and arginine residues in ORF1 align with the known ssDNA contact residues of BFDV and Hep E capsid proteins (FIG. 3, denoted by asterisk).

**SEQUENCES**

*The sequences listed below are annotated as follows. Bolded and underlined text indicates a sequence comprising a His6 tag (HHHHHH) used for chelating purification and a Flag tag (DYKDDDDK), a strong epitope used, e.g., for Western blot detection of low-expressing proteins. Bolded and italicized sequences indicate Ring9 ORF1 sequence or portions thereof. Unbolded, non-underlined sequences are Ring2 sequences or portions thereof. Unbolded, underlined sequences are from Beak and Feather Disease Virus (BFDV). Gray highlighting indicates the positions of lysine-to-histidine mutations, e.g., in the arginine-rich region and the first beta strand of Ring 9 ORF1.*

SEQ ID NO: 561:

Ring 2 N-terminal HIS-FLAG-3CProtease-ORF1:

**MGSSHHHHHHGSDYKDDDDKSGSLEVLFQGPSG**MPYYYRRRRYNYRRPRWYGRGWIRRPFRRRFRRKRRVRPTYT
TIPLKQWQPPYKRTCYIKGQDCLIYYSNLRLGMNSTMYEKSIVPVHWPGGGSFSVSMLTLDALYDIHKLCRNWWTSTN
QDLPLVRYKGCKITFYQSTFTDYIVRIHTELPANSNKLTYPNTHPLMMMMSKYKHIIPSRQTRRKKKPYTKIFVKPPPQFE
NKWYFATDLYKIPLLQIHCTACNLQNPFVKPDKLSNNVTLWSLNTISIQNRNMSVDQGQSWPFKILGTQSFYFYFYTGA
NLPGDTTQIPVADLLPLTNPRINRPGQSLNEAKITDHITFTEYKNKFTNYWGNPFNKHIQEHLDMILYSLKSPEAIKNEWT
TENMKWNQLNNAGTMALTPFNEPIFTQIQYNPDRDTGEDTQLYLLSNATGTGWDPPGIPELILEGFPLWLIYWGFAD
FQKNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQPLPEDNIKWYPQVQYQLEAQNKLLQTGPFT
PNIQGQLSDNISMFYKFYFKWGGSPPKAINVENPAHQIQYPIPRNEHETTSLQSPGEAPESILYSFDYRHGNYTTTALSRI
SQDWALKDTVSKITEPDRQQLLKQALECLQISEETQEKKEKEVQQLISNLRQQQQLYRERIISLLKDQ

SEQ ID NO: 562:

Ring 9 N-terminal HIS-FLAG-3CProtease-ORF1:

**MGSSHHHHHHGSDYKDDDDKSGSLEVLFQGPSG***MPPYWRQKYYRRRYRPFSWRTRRIIQRRKRWRYRKPRKTY*
*WRRKLRVRKRFYKRKLKKIVLKQFQPKIIRRCTIFGTICLFQGSPERANNNYIQTIYSYVPDKEPGGGGWTLITESLSSL*
*WEDWEHLKNVWTQSNAGLPLVRYGGVTLYFYQSAYTDYIAQVFNCYPMTDTKYTHADSAPNRMLLKKHVIRVPS*
*RETRKKRKPYKRVRVGPPSQMQNKWYFQRDICEIPLIMIAATAVDFRYPFCASDCASNNLTLTCLNPLLFQNQDFDH*
*PSDTQGYFPKPGVYLYSTQRSNKPSSSDCIYLGNTKDNQEGKSASSLMTLKTQKITDWGNPFWHYYIDGSKKIFSYFK*
*PPSQLDSSDFEHMTELAEPMFIQVRYNPERDTGQGNLIYVTENFRGQHWDPPSSDNLKLDGFPLYDMCWGFIDWI*

***EKVHETENLLTNYCFCIRSSAFNEKKTVFIPVDHSFLTGFSPYETPVKSSDQAHWHPQIRFQTKSINDICLTGPGCARSP YGNYMQAKMSYKFHVKWGGCPKTYEKPYDPCSQPNWTIPHNLNETIQIQNPNTCPQTELQEWDWRRDIVTKKAI ERIRQHTEPHETLQISTGSKHNPPVHRQTSPWTDSETDSEEEKDQTQEIQIQLNKLRKHQQHLKQQLKQYLKPQNIE***

SEQ ID NO: 563:

Ring 2 ORF1 with ARG arm of Ring 9 (Ring 291)

**MGSSHHHHHHGSDYKDDDDKSGSLEVLFQGPSG*MPPYWRQKYYRRRYRPFSWRTRRIIQRRKRWRYRKPRKTY WRRKLRVRK***RRVRPTYTTIPLKQWQPPYKRTCYIKGQDCLIYYSNLRLGMNSTMYEKSIVPVHWPGGGSFSVSMLTLD ALYDIHKLCRNWWTSTNQDLPLVRYKGCKITFYQSTFTDYIVRIHTELPANSNKLTYPNTHPLMMMMSKYKHIIPSRQT RRKKKPYTKIFVKPPPQFENKWYFATDLYKIPLLQIHCTACNLQNPFVKPDKLSNNVTLWSLNTISIQNRNMSVDQGQS WPFKILGTQSFYFYFYTGANLPGDTTQIPVADLLPLTNPRINRPGQSLNEAKITDHITFTEYKNKFTNYWGNPFNKHIQE HLDMILYSLKSPEAIKNEWTTENMKWNQLNNAGTMALTPFNEPIFTQIQYNPDRDTGEDTQLYLLSNATGTGWDPPG IPELILEGFPLWLIYWGFADFQKNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQPLPEDNIKWYP QVQYQLEAQNKLLQTGPFTPNIQGQLSDNISMFYKFYFKWGGSPPKAINVENPAHQIQYPIPRNEHETTSLQSPGEAPE SILYSFDYRHGNYTTTALSRISQDWALKDTVSKITEPDRQQLLKQALECLQISEETQEKKEKEVQQLISNLRQQQQLYRERI ISLLKDQ

SEQ ID NO: 564:

Ring 2 ORF1 with ARG arm and Beta strands 1 + 2 722 epitope of Ring 9 (Ring 292)

**MGSSHHHHHHGSDYKDDDDKSGSLEVLFQGPSG*MPPYWRQKYYRRRYRPFSWRTRRIIQRRKRWRYRKPRKTY WRRKLRVRKRFYKRKLKKIVLKQFQPKIIRRCTIFGTICLFQG***SNLRLGMNSTMYEKSIVPVHWPGGGSFSVSMLTLDA LYDIHKLCRNWWTSTNQDLPLVRYKGCKITFYQSTFTDYIVRIHTELPANSNKLTYPNTHPLMMMMSKYKHIIPSRQTR RKKKPYTKIFVKPPPQFENKWYFATDLYKIPLLQIHCTACNLQNPFVKPDKLSNNVTLWSLNTISIQNRNMSVDQGQSW PFKILGTQSFYFYFYTGANLPGDTTQIPVADLLPLTNPRINRPGQSLNEAKITDHITFTEYKNKFTNYWGNPFNKHIQEHL DMILYSLKSPEAIKNEWTTENMKWNQLNNAGTMALTPFNEPIFTQIQYNPDRDTGEDTQLYLLSNATGTGWDPPGIP ELILEGFPLWLIYWGFADFQKNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQPLPEDNIKWYPQV QYQLEAQNKLLQTGPFTPNIQGQLSDNISMFYKFYFKWGGSPPKAINVENPAHQIQYPIPRNEHETTSLQSPGEAPESIL YSFDYRHGNYTTTALSRISQDWALKDTVSKITEPDRQQLLKQALECLQISEETQEKKEKEVQQLISNLRQQQQLYRERIISL LKDQ

SEQ ID NO: 565:

Ring 9 with LYS/HIS mutations in ARG arm and first beta strand

**MGSSHHHHHHGSDYKDDDDKSGSLEVLFQGPSG*MPPYWRQKYYRRRYRPFSWRTRRIIQRRHRWRYRKPRHTY WRRHLRVRHRFYHRHLHHIVLKQFQPKIIRRCTIFGTICLFQGSPERANNNYIQTIYSYVPDKEPGGGGWTLITESLSSL WEDWEHLKNVWTQSNAGLPLVRYGGVTLYFYQSAYTDYIAQVFNCYPMTDTKYTHADSAPNRMLLKKHVIRVPS RETRKKRKPYKRVRVGPPSQMQNKWYFQRDICEIPLIMIAATAVDFRYPFCASDCASNNLTLTCLNPLLFQNQDFDH***

***PSDTQGYFPKPGVYLYSTQRSNKPSSSDCIYLGNTKDNQEGKSASSLMTLKTQKITDWGNPFWHYYIDGSKKIFSYFK PPSQLDSSDFEHMTELAEPMFIQVRYNPERDTGQGNLIYVTENFRGQHWDPPSSDNLKLDGFPLYDMCWGFIDWI EKVHETENLLTNYCFCIRSSAFNEKKTVFIPVDHSFLTGFSPYETPVKSSDQAHWHPQIRFQTKSINDICLTGPGCARSP YGNYMQAKMSYKFHVKWGGCPKTYEKPYDPCSQPNWTIPHNLNETIQIQNPNTCPQTELQEWDWRRDIVTKKAI ERIRQHTEPHETLQISTGSKHNPPVHRQTSPWTDSETDSEEEKDQTQEIQIQLNKLRKHQQHLKQQLKQYLKPQNIE***

SEQ ID NO: 566:

Ring 9 with ARG arm of BFDV:

**MGSSHHHHHHGSDYKDDDDKSGSLEVLFQGPSG**MWGTSNCACAKFQIRRRYARPYRRRHIRRYRRRRRHFRR RRFTTNR***RKRFYKRKLKKIVLKQFQPKIIRRCTIFGTICLFQGSPERANNNYIQTIYSYVPDKEPGGGGWTLITESLSSL WEDWEHLKNVWTQSNAGLPLVRYGGVTLYFYQSAYTDYIAQVFNCYPMTDTKYTHADSAPNRMLLKKHVIRVPS RETRKKRKPYKRVRVGPPSQMQNKWYFQRDICEIPLIMIAATAVDFRYPFCASDCASNNLTLTCLNPLLFQNQDFDH PSDTQGYFPKPGVYLYSTQRSNKPSSSDCIYLGNTKDNQEGKSASSLMTLKTQKITDWGNPFWHYYIDGSKKIFSYFK PPSQLDSSDFEHMTELAEPMFIQVRYNPERDTGQGNLIYVTENFRGQHWDPPSSDNLKLDGFPLYDMCWGFIDWI EKVHETENLLTNYCFCIRSSAFNEKKTVFIPVDHSFLTGFSPYETPVKSSDQAHWHPQIRFQTKSINDICLTGPGCARSP YGNYMQAKMSYKFHVKWGGCPKTYEKPYDPCSQPNWTIPHNLNETIQIQNPNTCPQTELQEWDWRRDIVTKKAI ERIRQHTEPHETLQISTGSKHNPPVHRQTSPWTDSETDSEEEKDQTQEIQIQLNKLRKHQQHLKQQLKQYLKPQNIE***

SEQ ID NO: 567:

Ring 9 with beta strands F and G of BFDV capsid protein:

**MGSSHHHHHHGSDYKDDDDKSGSLEVLFQGPSG***MPPYWRQKYYRRRYRPFSWRTRRIIQRRKRWRYRKPRKTY WRRKLRVRKRFYKRKLKKIVLKQFQPKIIRRCTIFGTICLFQGSPERANNNYIQTIYSYVPDKEPGGGGWTLITESLSSL WEDWEHLKNVWTQSNAGLPLVRYGGVTLYFYQSAYTDYIAQVFNCYPMTDTKYTHADSAPNRMLLKKH*AKKWF***S** ***RETRKKRKP***GFKRLL***GPPSQMQNKWYFQRDICEIPLIMIAATAVDFRYPFCASDCASNNLTLTCLNPLLFQNQDFDHP SDTQGYFPKPGVYLYSTQRSNKPSSSDCIYLGNTKDNQEGKSASSLMTLKTQKITDWGNPFWHYYIDGSKKIFSYFKP PSQLDSSDFEHMTELAEPMFIQVRYNPERDTGQGNLIYVTENFRGQHWDPPSSDNLKLDGFPLYDMCWGFIDWIE KVHETENLLTNYCFCIRSSAFNEKKTVFIPVDHSFLTGFSPYETPVKSSDQAHWHPQIRFQTKSINDICLTGPGCARSPY GNYMQAKMSYKFHVKWGGCPKTYEKPYDPCSQPNWTIPHNLNETIQIQNPNTCPQTELQEWDWRRDIVTKKAIE RIRQHTEPHETLQISTGSKHNPPVHRQTSPWTDSETDSEEEKDQTQEIQIQLNKLRKHQQHLKQQLKQYLKPQNIE***

SEQ ID NO: 568:

Ring 2 with beta C of Ring 9:

**MGSSHHHHHHGSDYKDDDDKSGSLEVLFQGPSG**MPYYYRRRRYNYRRPRWYGRGWIRRPFRRRFRRKRRVRPTYT TIPLKQWQPPYKR***RCTIFGTICLFQG***SNLRLGMNSTMYEKSIVPVHWPGGGSFSVSMLTLDALYDIHKLCRNWWTSTN QDLPLVRYKGCKITFYQSTFTDYIVRIHTELPANSNKLTYPNTHPLMMMMSKYKHIIPSRQTRRKKKPYTKIFVKPPPQFE NKWYFATDLYKIPLLQIHCTACNLQNPFVKPDKLSNNVTLWSLNTISIQNRRNMSVDQGQSWPFKILGTQSFYFYFYTGA NLPGDTTQIPVADLLPLTNPRINRPGQSLNEAKITDHITFTEYKNKFTNYWGNPFNKHIQEHLDMILYSLKSPEAIKNEWT TENMKWNQLNNAGTMALTPFNEPIFTQIQYNPDRDTGEDTQLYLLSNATGTGWDPPGIPELILEGFPLWLIYWGFAD

FQKNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQPLPEDNIKWYPQVQYQLEAQNKLLQTGPFT
PNIQGQLSDNISMFYKFYFKWGGSPPKAINVENPAHQIQYPIPRNEHETTSLQSPGEAPESILYSFDYRHGNYTTTALSRI
SQDWALKDTVSKITEPDRQQLLKQALECLQISEETQEKKEKEVQQLISNLRQQQQLYRERIISLLKDQ

SEQ ID NO: 569:

Ring 2 with linker 1 of Ring 9:

**MGSSHHHHHHGSDYKDDDDKSGSLEVLFQGPSG**MPYYYRRRRYNYRRPRWYGRGWIRRPFRRRFRRKRRVRPTYT
TIPLKQWQPPYKRTCYIKGQDCLIYY***SPERANNNYIQTIYSYVPDKEPGGGGWTLITESLSSLWEDWEHLKNVWTQSN AGLP***LVRYKGCKITFYQSTFTDYIVRIHTELPANSNKLTYPNTHPLMMMMSKYKHIIPSRQTRRKKKPYTKIFVKPPPQFE
NKWYFATDLYKIPLLQIHCTACNLQNPFVKPDKLSNNVTLWSLNTISIQNRNMSVDQGQSWPFKILGTQSFYFYFYTGA
NLPGDTTQIPVADLLPLTNPRINRPGQSLNEAKITDHITFTEYKNKFTNYWGNPFNKHIQEHLDMILYSLKSPEAIKNEWT
TENMKWNQLNNAGTMALTPFNEPIFTQIQYNPDRDTGEDTQLYLLSNATGTGWDPPGIPELILEGFPLWLIYWGFAD
FQKNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQPLPEDNIKWYPQVQYQLEAQNKLLQTGPFT
PNIQGQLSDNISMFYKFYFKWGGSPPKAINVENPAHQIQYPIPRNEHETTSLQSPGEAPESILYSFDYRHGNYTTTALSRI
SQDWALKDTVSKITEPDRQQLLKQALECLQISEETQEKKEKEVQQLISNLRQQQQLYRERIISLLKDQ

SEQ ID NO: 570:

Ring 2 with beta strand D of Ring 9:

**MGSSHHHHHHGSDYKDDDDKSGSLEVLFQGPSG**MPYYYRRRRYNYRRPRWYGRGWIRRPFRRRFRRKRRVRPTYT
TIPLKQWQPPYKRTCYIKGQDCLIYYSNLRLGMNSTMYEKSIVPVHWPGGGSFSVSMLTLDALYDIHKLCRNWWTSTN
QDLP***LVRYGGVTLYF***YQSTFTDYIVRIHTELPANSNKLTYPNTHPLMMMMSKYKHIIPSRQTRRKKKPYTKIFVKPPPQF
ENKWYFATDLYKIPLLQIHCTACNLQNPFVKPDKLSNNVTLWSLNTISIQNRNMSVDQGQSWPFKILGTQSFYFYFYTG
ANLPGDTTQIPVADLLPLTNPRINRPGQSLNEAKITDHITFTEYKNKFTNYWGNPFNKHIQEHLDMILYSLKSPEAIKNE
WTTENMKWNQLNNAGTMALTPFNEPIFTQIQYNPDRDTGEDTQLYLLSNATGTGWDPPGIPELILEGFPLWLIYWGF
ADFQKNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQPLPEDNIKWYPQVQYQLEAQNKLLQTG
PFTPNIQGQLSDNISMFYKFYFKWGGSPPKAINVENPAHQIQYPIPRNEHETTSLQSPGEAPESILYSFDYRHGNYTTTAL
SRISQDWALKDTVSKITEPDRQQLLKQALECLQISEETQEKKEKEVQQLISNLRQQQQLYRERIISLLKDQ

SEQ ID NO: 571:

Ring 2 with linker 2 of Ring 9:

**MGSSHHHHHHGSDYKDDDDKSGSLEVLFQGPSG**MPYYYRRRRYNYRRPRWYGRGWIRRPFRRRFRRKRRVRPTYT
TIPLKQWQPPYKRTCYIKGQDCLIYYSNLRLGMNSTMYEKSIVPVHWPGGGSFSVSMLTLDALYDIHKLCRNWWTSTN
QDLPLVRYKGCKITFYQSTFTDYIVRIH***NCYPMTDTKYTHADSAPNRMLLKKH***KHIIPSRQTRRKKKPYTKIFVKPPPQFE
NKWYFATDLYKIPLLQIHCTACNLQNPFVKPDKLSNNVTLWSLNTISIQNRNMSVDQGQSWPFKILGTQSFYFYFYTGA
NLPGDTTQIPVADLLPLTNPRINRPGQSLNEAKITDHITFTEYKNKFTNYWGNPFNKHIQEHLDMILYSLKSPEAIKNEWT
TENMKWNQLNNAGTMALTPFNEPIFTQIQYNPDRDTGEDTQLYLLSNATGTGWDPPGIPELILEGFPLWLIYWGFAD
FQKNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQPLPEDNIKWYPQVQYQLEAQNKLLQTGPFT

PNIQGQLSDNISMFYKFYFKWGGSPPKAINVENPAHQIQYPIPRNEHETTSLQSPGEAPESILYSFDYRHGNYTTTALSRI
SQDWALKDTVSKITEPDRQQLLKQALECLQISEETQEKKEKEVQQLISNLRQQQQLYRERIISLLKDQ

SEQ ID NO: 572:

Ring 2 with beta strand G DNA binding of Ring 9:

**MGSSHHHHHHGSDYKDDDDKSGSLEVLFQGPSG**MPYYYRRRRYNYRRPRWYGRGWIRRPFRRRFRRKRRVRPTYT
TIPLKQWQPPYKRTCYIKGQDCLIYYSNLRLGMNSTMYEKSIVPVHWPGGGSFSVSMLTLDALYDIHKLCRNWWTSTN
QDLPLVRYKGCKITFYQSTFTDYIVRIHTELPANSNKLTYPNTHPLMMMMSKYKHIIPSRQTRRKKKP***YKRVRV***KPPPQF
ENKWYFATDLYKIPLLQIHCTACNLQNPFVKPDKLSNNVTLWSLNTISIQNRNMSVDQGQSWPFKILGTQSFYFYFYTG
ANLPGDTTQIPVADLLPLTNPRINRPGQSLNEAKITDHITFTEYKNKFTNYWGNPFNKHIQEHLDMILYSLKSPEAIKNE
WTTENMKWNQLNNAGTMALTPFNEPIFTQIQYNPDRDTGEDTQLYLLSNATGTGWDPPGIPELILEGFPLWLIYWGF
ADFQKNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQPLPEDNIKWYPQVQYQLEAQNKLLQTG
PFTPNIQGQLSDNISMFYKFYFKWGGSPPKAINVENPAHQIQYPIPRNEHETTSLQSPGEAPESILYSFDYRHGNYTTTAL
SRISQDWALKDTVSKITEPDRQQLLKQALECLQISEETQEKKEKEVQQLISNLRQQQQLYRERIISLLKDQ

SEQ ID NO: 573:

Ring 2 with beta strand F interprotomer contact of Ring 9:

**MGSSHHHHHHGSDYKDDDDKSGSLEVLFQGPSG**MPYYYRRRRYNYRRPRWYGRGWIRRPFRRRFRRKRRVRPTYT
TIPLKQWQPPYKRTCYIKGQDCLIYYSNLRLGMNSTMYEKSIVPVHWPGGGSFSVSMLTLDALYDIHKLCRNWWTSTN
QDLPLVRYKGCKITFYQSTFTDYIVRIHTELPANSNKLTYPNTHPLMMMMSK***HVIRVP***SRQTRRKKKPYTKIFVKPPPQF
ENKWYFATDLYKIPLLQIHCTACNLQNPFVKPDKLSNNVTLWSLNTISIQNRNMSVDQGQSWPFKILGTQSFYFYFYTG
ANLPGDTTQIPVADLLPLTNPRINRPGQSLNEAKITDHITFTEYKNKFTNYWGNPFNKHIQEHLDMILYSLKSPEAIKNE
WTTENMKWNQLNNAGTMALTPFNEPIFTQIQYNPDRDTGEDTQLYLLSNATGTGWDPPGIPELILEGFPLWLIYWGF
ADFQKNLKKVTNIDTNYMLVAKTKFTQKPGTFYLVILNDTFVEGNSPYEKQPLPEDNIKWYPQVQYQLEAQNKLLQTG
PFTPNIQGQLSDNISMFYKFYFKWGGSPPKAINVENPAHQIQYPIPRNEHETTSLQSPGEAPESILYSFDYRHGNYTTTAL
SRISQDWALKDTVSKITEPDRQQLLKQALECLQISEETQEKKEKEVQQLISNLRQQQQLYRERIISLLKDQ

SEQ ID NO: 574:

Ring 2 ORF1 with strand H and I with C-terminal fragment of Ring 9:

**MGSSHHHHHHGSDYKDDDDKSGSLEVLFQGPSG**MPYYYRRRRYNYRRPRWYGRGWIRRPFRRRFRRKRRVRPTYT
TIPLKQWQPPYKRTCYIKGQDCLIYYSNLRLGMNSTMYEKSIVPVHWPGGGSFSVSMLTLDALYDIHKLCRNWWTSTN
QDLPLVRYKGCKITFYQSTFTDYIVRIHTELPANSNKLTYPNTHPLMMMMSKYKHIIPSRQTRRKKKPYTKIFVKPPPQFE
NKWYFATDLYK***IPLIMIAATAVDFRYPFCASDCASNNLTLTCLNPLLFQNQDFDHPSDTQGYFPKPGVYLYSTQRSNK***
***PSSSDCIYLGNTKDNQEGKSASSLMTLKTQKITDWGNPFWHYYIDGSKKIFSYFKPPSQLDSSDFEHMTELAEPMFIQ***
***VRYNPERDTGQGNLIYVTENFRGQHWDPPSSDNLKLDGFPLYDMCWGFIDWIEKVHETENLLTNYCFCIRSSAFNEK***
***KTVFIPVDHSFLTGFSPYETPVKSSDQAHWHPQIRFQTKSINDICLTGPGCARSPYGNYMQAKMSYKFHVKWGGCP***
***KTYEKPYDPCSQPNWTIPHNLNETIQIQNPNTCPQTELQEWDWRRDIVTKKAIERIRQHTEPHETLQISTGSKHNPPV***
***HRQTSPWTDSETDSEEEKDQTQEIQIQLNKLRKHQQHLKQQLKQYLKPQNIE***

SEQ ID NO: 575:

Ring 2 ORF1 with strand I and C-terminal fragment of Ring 9:

**MGSSHHHHHHGSDYKDDDDKSGSLEVLFQGPSG**MPYYYRRRRYNYRRPRWYGRGWIRRPFRRRFRRKRRVRPTYT
TIPLKQWQPPYKRTCYIKGQDCLIYYSNLRLGMNSTMYEKSIVPVHWPGGGSFSVSMLTLDALYDIHKLCRNWWTSTN
QDLPLVRYKGCKITFYQSTFTDYIVRIHTELPANSNKLTYPNTHPLMMMMSKYKHIIPSRQTRRKKKPYTKIFVKPPPQFE
NKWYFATDLYKIPLLQIHCTACNLQNPFVKPDKLSN***NLTLTCLNPLLFQNQDFDHPSDTQGYFPKPGVYLYSTQRSNKP***
***SSSDCIYLGNTKDNQEGKSASSLMTLKTQKITDWGNPFWHYYIDGSKKIFSYFKPPSQLDSSDFEHMTELAEPMFIQV***
***RYNPERDTGQGNLIYVTENFRGQHWDPPSSDNLKLDGFPLYDMCWGFIDWIEKVHETENLLTNYCFCIRSSAFNEKK***
***TVFIPVDHSFLTGFSPYETPVKSSDQAHWHPQIRFQTKSINDICLTGPGCARSPYGNYMQAKMSYKFHVKWGGCPK***
***TYEKPYDPCSQPNWTIPHNLNETIQIQNPNTCPQTELQEWDWRRDIVTKKAIERIRQHTEPHETLQISTGSKHNPPV***
***HRQTSPWTDSETDSEEEKDQTQEIQIQLNKLRKHQQHLKQQLKQYLKPQNIE***

What is claimed is:

1. An anellovector comprising:

a) proteinaceous exterior comprising an ORF1 molecule;

b) a genetic element comprising RNA,

wherein the genetic element is enclosed within the proteinaceous exterior.

2. The anellovector of claim 1, wherein the genetic element consists of at least 10%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% RNA.

3. The anellovector of claim 1 or 2, wherein the RNA comprises one or more chemical modifications.

4. The anellovector of any of the preceding claims, wherein the genetic element consists of or consists essentially of RNA.

5. The anellovector of any of the preceding claims, wherein the genetic element comprises a DNA region.

6. The anellovector of any of the preceding claims, wherein at least a portion of the DNA region hybridizes to at least a portion of the RNA of the genetic element.

7. The anellovector of any of the preceding claims, wherein the genetic element is circular.

8. A method of making an anellovector, the method comprising:

(a) providing a mixture comprising:

(i) a genetic element comprising RNA, and

(ii) an ORF1 molecule; and

(b) incubating the mixture under conditions suitable for enclosing the genetic element within a proteinaceous exterior comprising the ORF1 molecule, thereby making an anellovector;

optionally wherein the mixture is not comprised in a cell.

9. A method of delivering a genetic element to a cell, the method comprising contacting the anellovector of any of claims 1-7 with a cell, e.g., a eukaryotic cell, e.g., a mammalian cell, e.g., a human cell.

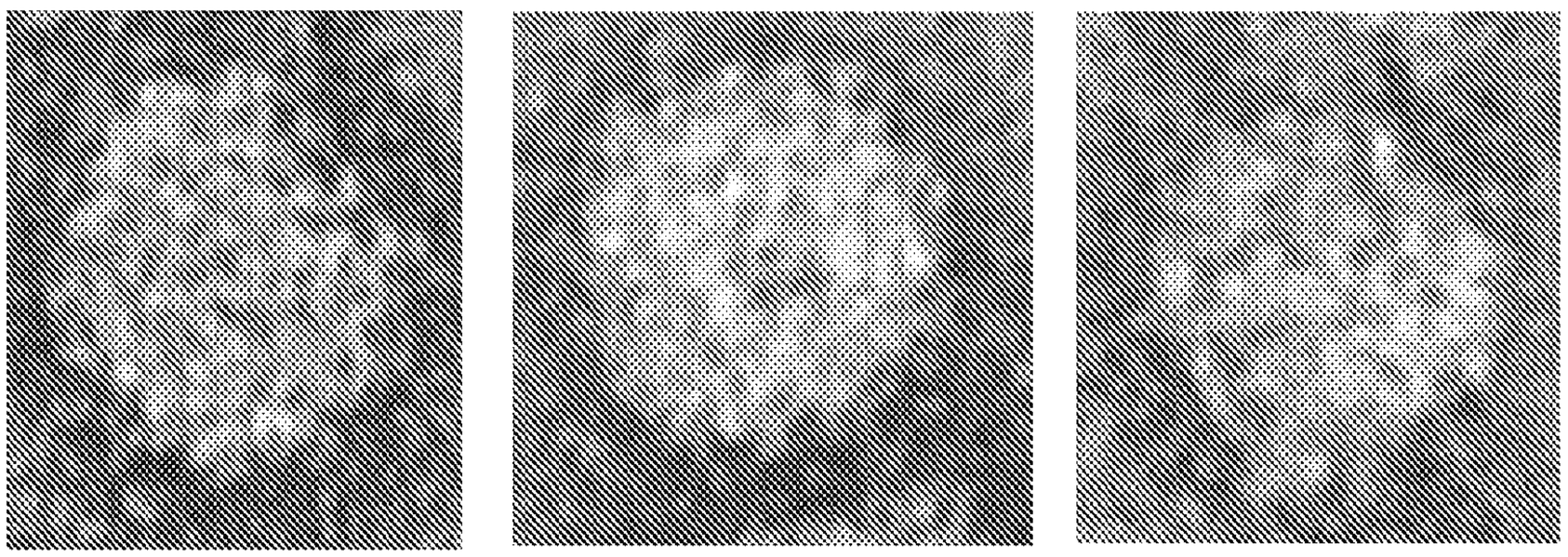



FIG. 1

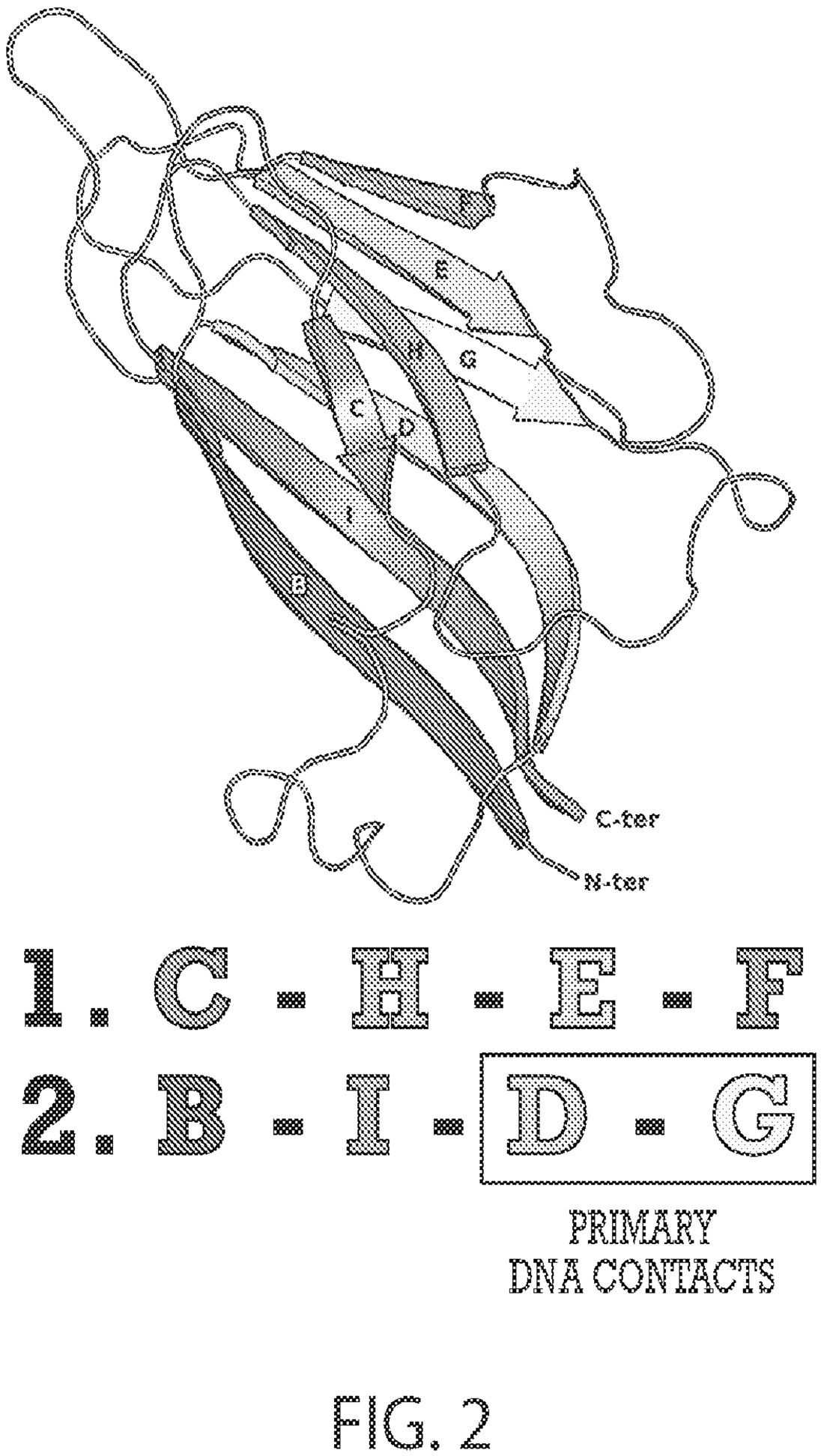



FIG. 2

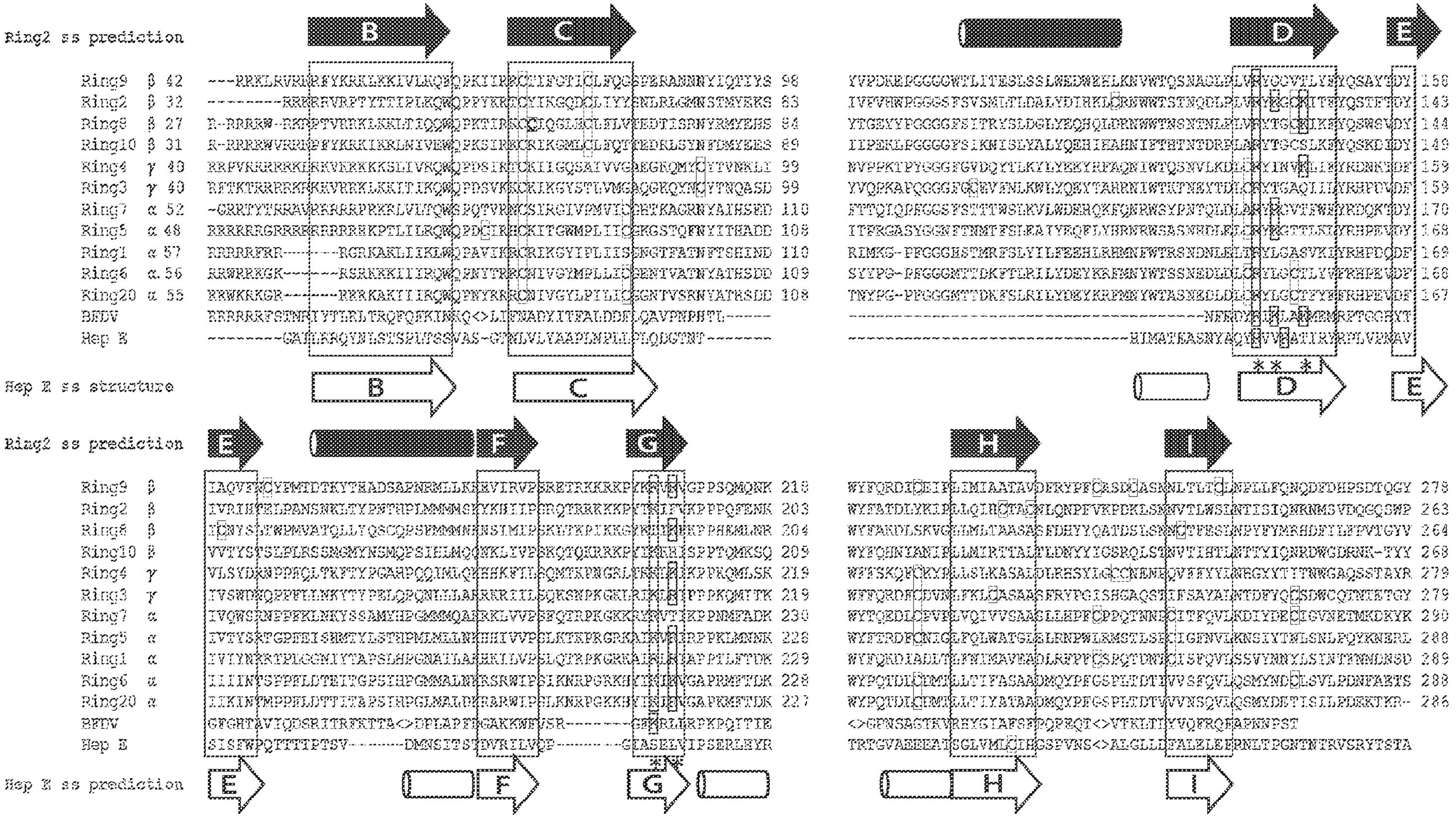



FIG. 3

SUBSTITUTE SHEET (RULE 26)

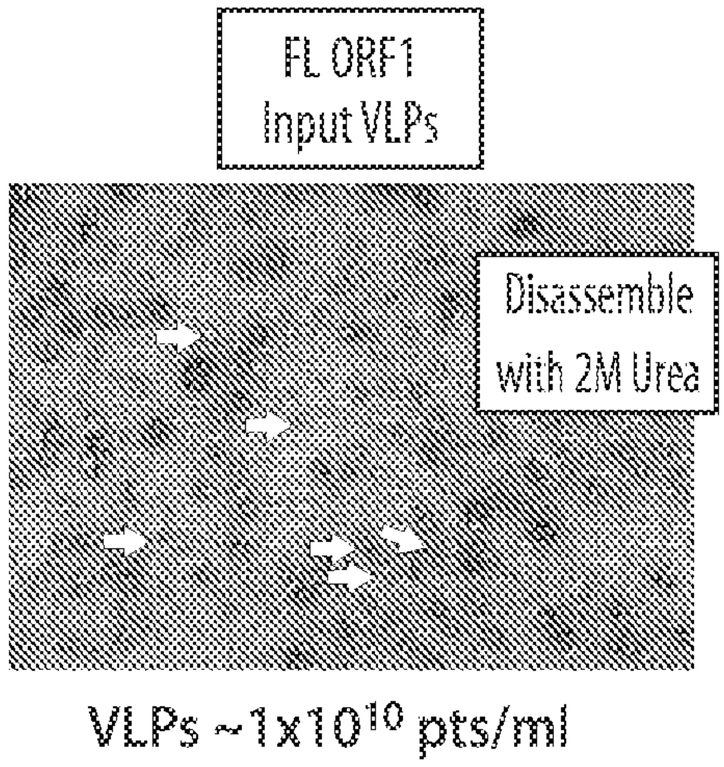



FIG. 4A

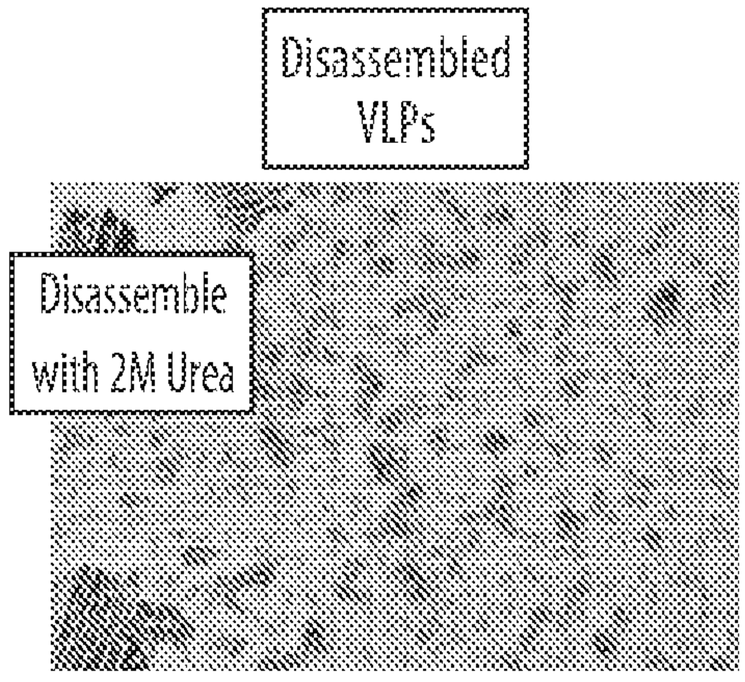



FIG. 4B

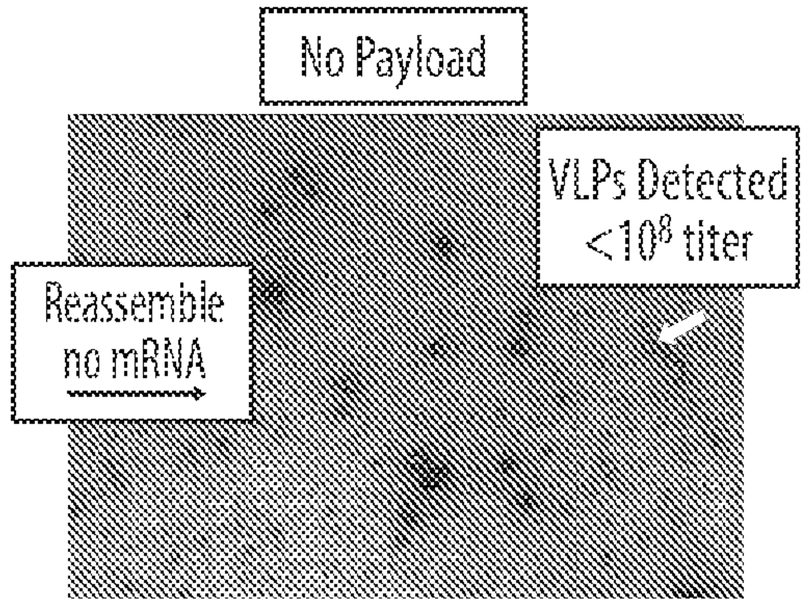



FIG. 4C

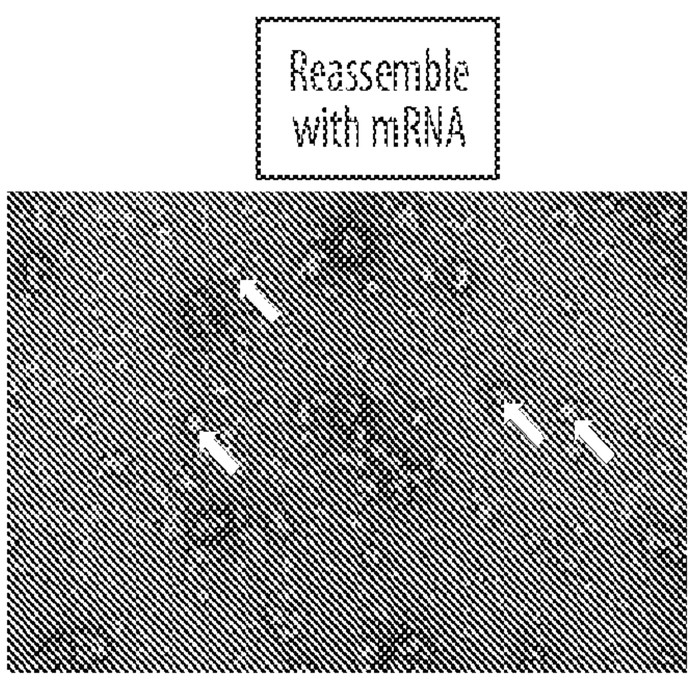



FIG. 4D

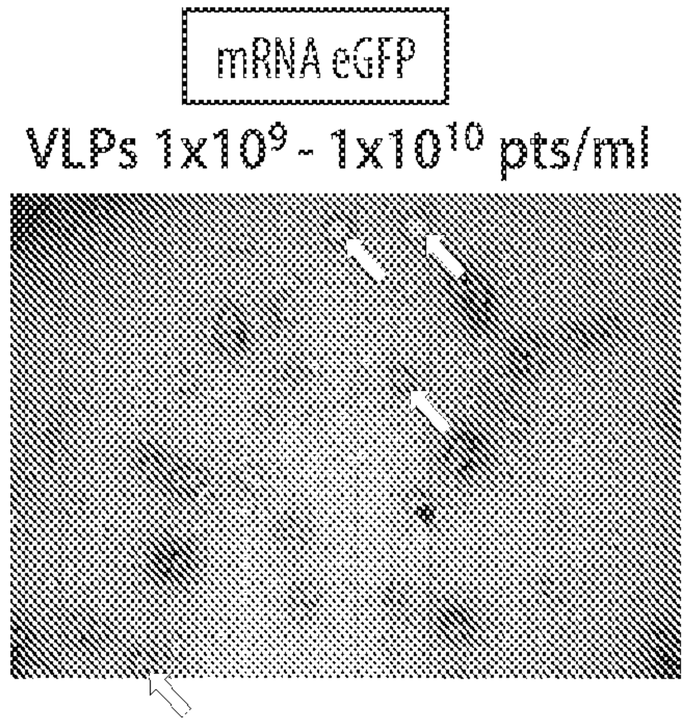



FIG. 4E

SEQUENCE LISTING

<110> FLAGSHIP PIONEERING INNOVATIONS V, INC.

<120> IN VITRO ASSEMBLY OF ANELLOVIRUS CAPSIDS ENCLOSING RNA

<130> V2057-7017WO

<140>
<141>

<150> 63/147,064
<151> 2021-02-08

<150> 63/130,360
<151> 2020-12-23

<160> 968

<170> PatentIn version 3.5

<210> 1
<211> 699
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 1
Met Gly Ser Ser His His His His His His Gly Ser Asp Tyr Lys Asp
1 5 10 15

Asp Asp Asp Lys Ser Gly Ser Leu Glu Val Leu Phe Gln Gly Pro Ser
20 25 30

Gly Met Pro Tyr Tyr Tyr Arg Arg Arg Arg Tyr Asn Tyr Arg Arg Pro
35 40 45

Arg Trp Tyr Gly Arg Gly Trp Ile Arg Arg Pro Phe Arg Arg Arg Phe
50 55 60

Arg Arg Lys Arg Arg Val Arg Pro Thr Tyr Thr Thr Ile Pro Leu Lys
65 70 75 80

```
Gln Trp Gln Pro Pro Tyr Lys Arg Thr Cys Tyr Ile Lys Gly Gln Asp
                85                  90                  95

Cys Leu Ile Tyr Tyr Ser Asn Leu Arg Leu Gly Met Asn Ser Thr Met
            100                 105                 110

Tyr Glu Lys Ser Ile Val Pro Val His Trp Pro Gly Gly Gly Ser Phe
        115                 120                 125

Ser Val Ser Met Leu Thr Leu Asp Ala Leu Tyr Asp Ile His Lys Leu
    130                 135                 140

Cys Arg Asn Trp Trp Thr Ser Thr Asn Gln Asp Leu Pro Leu Val Arg
145                 150                 155                 160

Tyr Lys Gly Cys Lys Ile Thr Phe Tyr Gln Ser Thr Phe Thr Asp Tyr
                165                 170                 175

Ile Val Arg Ile His Thr Glu Leu Pro Ala Asn Ser Asn Lys Leu Thr
            180                 185                 190

Tyr Pro Asn Thr His Pro Leu Met Met Met Met Ser Lys Tyr Lys His
        195                 200                 205

Ile Ile Pro Ser Arg Gln Thr Arg Arg Lys Lys Lys Pro Tyr Thr Lys
    210                 215                 220

Ile Phe Val Lys Pro Pro Pro Gln Phe Glu Asn Lys Trp Tyr Phe Ala
225                 230                 235                 240

Thr Asp Leu Tyr Lys Ile Pro Leu Leu Gln Ile His Cys Thr Ala Cys
                245                 250                 255

Asn Leu Gln Asn Pro Phe Val Lys Pro Asp Lys Leu Ser Asn Asn Val
            260                 265                 270

Thr Leu Trp Ser Leu Asn Thr Ile Ser Ile Gln Asn Arg Asn Met Ser
```

```
        275                 280                 285

Val Asp Gln Gly Gln Ser Trp Pro Phe Lys Ile Leu Gly Thr Gln Ser
    290                 295                 300

Phe Tyr Phe Tyr Phe Tyr Thr Gly Ala Asn Leu Pro Gly Asp Thr Thr
305                 310                 315                 320

Gln Ile Pro Val Ala Asp Leu Leu Pro Leu Thr Asn Pro Arg Ile Asn
                325                 330                 335

Arg Pro Gly Gln Ser Leu Asn Glu Ala Lys Ile Thr Asp His Ile Thr
            340                 345                 350

Phe Thr Glu Tyr Lys Asn Lys Phe Thr Asn Tyr Trp Gly Asn Pro Phe
        355                 360                 365

Asn Lys His Ile Gln Glu His Leu Asp Met Ile Leu Tyr Ser Leu Lys
    370                 375                 380

Ser Pro Glu Ala Ile Lys Asn Glu Trp Thr Thr Glu Asn Met Lys Trp
385                 390                 395                 400

Asn Gln Leu Asn Asn Ala Gly Thr Met Ala Leu Thr Pro Phe Asn Glu
                405                 410                 415

Pro Ile Phe Thr Gln Ile Gln Tyr Asn Pro Asp Arg Asp Thr Gly Glu
            420                 425                 430

Asp Thr Gln Leu Tyr Leu Leu Ser Asn Ala Thr Gly Thr Gly Trp Asp
        435                 440                 445

Pro Pro Gly Ile Pro Glu Leu Ile Leu Glu Gly Phe Pro Leu Trp Leu
    450                 455                 460

Ile Tyr Trp Gly Phe Ala Asp Phe Gln Lys Asn Leu Lys Lys Val Thr
465                 470                 475                 480
```

```
Asn Ile Asp Thr Asn Tyr Met Leu Val Ala Lys Thr Lys Phe Thr Gln
                485                 490                 495

Lys Pro Gly Thr Phe Tyr Leu Val Ile Leu Asn Asp Thr Phe Val Glu
            500                 505                 510

Gly Asn Ser Pro Tyr Glu Lys Gln Pro Leu Pro Glu Asp Asn Ile Lys
        515                 520                 525

Trp Tyr Pro Gln Val Gln Tyr Gln Leu Glu Ala Gln Asn Lys Leu Leu
    530                 535                 540

Gln Thr Gly Pro Phe Thr Pro Asn Ile Gln Gly Gln Leu Ser Asp Asn
545                 550                 555                 560

Ile Ser Met Phe Tyr Lys Phe Tyr Phe Lys Trp Gly Gly Ser Pro Pro
                565                 570                 575

Lys Ala Ile Asn Val Glu Asn Pro Ala His Gln Ile Gln Tyr Pro Ile
            580                 585                 590

Pro Arg Asn Glu His Glu Thr Thr Ser Leu Gln Ser Pro Gly Glu Ala
        595                 600                 605

Pro Glu Ser Ile Leu Tyr Ser Phe Asp Tyr Arg His Gly Asn Tyr Thr
    610                 615                 620

Thr Thr Ala Leu Ser Arg Ile Ser Gln Asp Trp Ala Leu Lys Asp Thr
625                 630                 635                 640

Val Ser Lys Ile Thr Glu Pro Asp Arg Gln Gln Leu Leu Lys Gln Ala
                645                 650                 655

Leu Glu Cys Leu Gln Ile Ser Glu Glu Thr Gln Glu Lys Lys Glu Lys
            660                 665                 670

Glu Val Gln Gln Leu Ile Ser Asn Leu Arg Gln Gln Gln Gln Leu Tyr
```

675 680 685

Arg Glu Arg Ile Ile Ser Leu Leu Lys Asp Gln
690 695

<210> 2
<211> 688
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 2
Met Gly Ser Ser His His His His His His Gly Ser Asp Tyr Lys Asp
1 5 10 15

Asp Asp Asp Lys Ser Gly Ser Leu Glu Val Leu Phe Gln Gly Pro Ser
20 25 30

Gly Met Pro Pro Tyr Trp Arg Gln Lys Tyr Tyr Arg Arg Arg Tyr Arg
35 40 45

Pro Phe Ser Trp Arg Thr Arg Arg Ile Ile Gln Arg Arg Lys Arg Trp
50 55 60

Arg Tyr Arg Lys Pro Arg Lys Thr Tyr Trp Arg Arg Lys Leu Arg Val
65 70 75 80

Arg Lys Arg Phe Tyr Lys Arg Lys Leu Lys Lys Ile Val Leu Lys Gln
85 90 95

Phe Gln Pro Lys Ile Ile Arg Arg Cys Thr Ile Phe Gly Thr Ile Cys
100 105 110

Leu Phe Gln Gly Ser Pro Glu Arg Ala Asn Asn Asn Tyr Ile Gln Thr
115 120 125

Ile Tyr Ser Tyr Val Pro Asp Lys Glu Pro Gly Gly Gly Gly Trp Thr

```
    130                 135                 140

Leu Ile Thr Glu Ser Leu Ser Ser Leu Trp Glu Asp Trp Glu His Leu
145                 150                 155                 160

Lys Asn Val Trp Thr Gln Ser Asn Ala Gly Leu Pro Leu Val Arg Tyr
                165                 170                 175

Gly Gly Val Thr Leu Tyr Phe Tyr Gln Ser Ala Tyr Thr Asp Tyr Ile
            180                 185                 190

Ala Gln Val Phe Asn Cys Tyr Pro Met Thr Asp Thr Lys Tyr Thr His
        195                 200                 205

Ala Asp Ser Ala Pro Asn Arg Met Leu Leu Lys Lys His Val Ile Arg
    210                 215                 220

Val Pro Ser Arg Glu Thr Arg Lys Lys Arg Lys Pro Tyr Lys Arg Val
225                 230                 235                 240

Arg Val Gly Pro Pro Ser Gln Met Gln Asn Lys Trp Tyr Phe Gln Arg
                245                 250                 255

Asp Ile Cys Glu Ile Pro Leu Ile Met Ile Ala Ala Thr Ala Val Asp
            260                 265                 270

Phe Arg Tyr Pro Phe Cys Ala Ser Asp Cys Ala Ser Asn Asn Leu Thr
        275                 280                 285

Leu Thr Cys Leu Asn Pro Leu Leu Phe Gln Asn Gln Asp Phe Asp His
    290                 295                 300

Pro Ser Asp Thr Gln Gly Tyr Phe Pro Lys Pro Gly Val Tyr Leu Tyr
305                 310                 315                 320

Ser Thr Gln Arg Ser Asn Lys Pro Ser Ser Ser Asp Cys Ile Tyr Leu
                325                 330                 335
```

```
Gly Asn Thr Lys Asp Asn Gln Glu Gly Lys Ser Ala Ser Ser Leu Met
            340                 345                 350

Thr Leu Lys Thr Gln Lys Ile Thr Asp Trp Gly Asn Pro Phe Trp His
        355                 360                 365

Tyr Tyr Ile Asp Gly Ser Lys Lys Ile Phe Ser Tyr Phe Lys Pro Pro
    370                 375                 380

Ser Gln Leu Asp Ser Ser Asp Phe Glu His Met Thr Glu Leu Ala Glu
385                 390                 395                 400

Pro Met Phe Ile Gln Val Arg Tyr Asn Pro Glu Arg Asp Thr Gly Gln
                405                 410                 415

Gly Asn Leu Ile Tyr Val Thr Glu Asn Phe Arg Gly Gln His Trp Asp
            420                 425                 430

Pro Pro Ser Ser Asp Asn Leu Lys Leu Asp Gly Phe Pro Leu Tyr Asp
        435                 440                 445

Met Cys Trp Gly Phe Ile Asp Trp Ile Glu Lys Val His Glu Thr Glu
    450                 455                 460

Asn Leu Leu Thr Asn Tyr Cys Phe Cys Ile Arg Ser Ser Ala Phe Asn
465                 470                 475                 480

Glu Lys Lys Thr Val Phe Ile Pro Val Asp His Ser Phe Leu Thr Gly
                485                 490                 495

Phe Ser Pro Tyr Glu Thr Pro Val Lys Ser Ser Asp Gln Ala His Trp
            500                 505                 510

His Pro Gln Ile Arg Phe Gln Thr Lys Ser Ile Asn Asp Ile Cys Leu
        515                 520                 525

Thr Gly Pro Gly Cys Ala Arg Ser Pro Tyr Gly Asn Tyr Met Gln Ala
```

```
    530                 535                 540

Lys Met Ser Tyr Lys Phe His Val Lys Trp Gly Gly Cys Pro Lys Thr
545                 550                 555                 560

Tyr Glu Lys Pro Tyr Asp Pro Cys Ser Gln Pro Asn Trp Thr Ile Pro
                565                 570                 575

His Asn Leu Asn Glu Thr Ile Gln Ile Gln Asn Pro Asn Thr Cys Pro
            580                 585                 590

Gln Thr Glu Leu Gln Glu Trp Asp Trp Arg Arg Asp Ile Val Thr Lys
        595                 600                 605

Lys Ala Ile Glu Arg Ile Arg Gln His Thr Glu Pro His Glu Thr Leu
    610                 615                 620

Gln Ile Ser Thr Gly Ser Lys His Asn Pro Pro Val His Arg Gln Thr
625                 630                 635                 640

Ser Pro Trp Thr Asp Ser Glu Thr Asp Ser Glu Glu Glu Lys Asp Gln
                645                 650                 655

Thr Gln Glu Ile Gln Ile Gln Leu Asn Lys Leu Arg Lys His Gln Gln
            660                 665                 670

His Leu Lys Gln Gln Leu Lys Gln Tyr Leu Lys Pro Gln Asn Ile Glu
        675                 680                 685


<210> 3
<211> 714
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 3
Met Gly Ser Ser His His His His His His Gly Ser Asp Tyr Lys Asp
```

```
1               5                   10                  15

Asp Asp Asp Lys Ser Gly Ser Leu Glu Val Leu Phe Gln Gly Pro Ser
            20                  25                  30

Gly Met Pro Pro Tyr Trp Arg Gln Lys Tyr Tyr Arg Arg Arg Tyr Arg
        35                  40                  45

Pro Phe Ser Trp Arg Thr Arg Arg Ile Ile Gln Arg Arg Lys Arg Trp
    50                  55                  60

Arg Tyr Arg Lys Pro Arg Lys Thr Tyr Trp Arg Arg Lys Leu Arg Val
65                  70                  75                  80

Arg Lys Arg Arg Val Arg Pro Thr Tyr Thr Thr Ile Pro Leu Lys Gln
                85                  90                  95

Trp Gln Pro Pro Tyr Lys Arg Thr Cys Tyr Ile Lys Gly Gln Asp Cys
            100                 105                 110

Leu Ile Tyr Tyr Ser Asn Leu Arg Leu Gly Met Asn Ser Thr Met Tyr
        115                 120                 125

Glu Lys Ser Ile Val Pro Val His Trp Pro Gly Gly Gly Ser Phe Ser
    130                 135                 140

Val Ser Met Leu Thr Leu Asp Ala Leu Tyr Asp Ile His Lys Leu Cys
145                 150                 155                 160

Arg Asn Trp Trp Thr Ser Thr Asn Gln Asp Leu Pro Leu Val Arg Tyr
                165                 170                 175

Lys Gly Cys Lys Ile Thr Phe Tyr Gln Ser Thr Phe Thr Asp Tyr Ile
            180                 185                 190

Val Arg Ile His Thr Glu Leu Pro Ala Asn Ser Asn Lys Leu Thr Tyr
        195                 200                 205
```

```
Pro Asn Thr His Pro Leu Met Met Met Met Ser Lys Tyr Lys His Ile
    210                 215                 220

Ile Pro Ser Arg Gln Thr Arg Arg Lys Lys Lys Pro Tyr Thr Lys Ile
225                 230                 235                 240

Phe Val Lys Pro Pro Pro Gln Phe Glu Asn Lys Trp Tyr Phe Ala Thr
                245                 250                 255

Asp Leu Tyr Lys Ile Pro Leu Leu Gln Ile His Cys Thr Ala Cys Asn
            260                 265                 270

Leu Gln Asn Pro Phe Val Lys Pro Asp Lys Leu Ser Asn Asn Val Thr
        275                 280                 285

Leu Trp Ser Leu Asn Thr Ile Ser Ile Gln Asn Arg Asn Met Ser Val
    290                 295                 300

Asp Gln Gly Gln Ser Trp Pro Phe Lys Ile Leu Gly Thr Gln Ser Phe
305                 310                 315                 320

Tyr Phe Tyr Phe Tyr Thr Gly Ala Asn Leu Pro Gly Asp Thr Thr Gln
                325                 330                 335

Ile Pro Val Ala Asp Leu Leu Pro Leu Thr Asn Pro Arg Ile Asn Arg
            340                 345                 350

Pro Gly Gln Ser Leu Asn Glu Ala Lys Ile Thr Asp His Ile Thr Phe
        355                 360                 365

Thr Glu Tyr Lys Asn Lys Phe Thr Asn Tyr Trp Gly Asn Pro Phe Asn
    370                 375                 380

Lys His Ile Gln Glu His Leu Asp Met Ile Leu Tyr Ser Leu Lys Ser
385                 390                 395                 400

Pro Glu Ala Ile Lys Asn Glu Trp Thr Thr Glu Asn Met Lys Trp Asn
```

```
                405                 410                 415

Gln Leu Asn Asn Ala Gly Thr Met Ala Leu Thr Pro Phe Asn Glu Pro
            420                 425                 430

Ile Phe Thr Gln Ile Gln Tyr Asn Pro Asp Arg Asp Thr Gly Glu Asp
        435                 440                 445

Thr Gln Leu Tyr Leu Leu Ser Asn Ala Thr Gly Thr Gly Trp Asp Pro
    450                 455                 460

Pro Gly Ile Pro Glu Leu Ile Leu Glu Gly Phe Pro Leu Trp Leu Ile
465                 470                 475                 480

Tyr Trp Gly Phe Ala Asp Phe Gln Lys Asn Leu Lys Lys Val Thr Asn
                485                 490                 495

Ile Asp Thr Asn Tyr Met Leu Val Ala Lys Thr Lys Phe Thr Gln Lys
            500                 505                 510

Pro Gly Thr Phe Tyr Leu Val Ile Leu Asn Asp Thr Phe Val Glu Gly
        515                 520                 525

Asn Ser Pro Tyr Glu Lys Gln Pro Leu Pro Glu Asp Asn Ile Lys Trp
    530                 535                 540

Tyr Pro Gln Val Gln Tyr Gln Leu Glu Ala Gln Asn Lys Leu Leu Gln
545                 550                 555                 560

Thr Gly Pro Phe Thr Pro Asn Ile Gln Gly Gln Leu Ser Asp Asn Ile
                565                 570                 575

Ser Met Phe Tyr Lys Phe Tyr Phe Lys Trp Gly Gly Ser Pro Pro Lys
            580                 585                 590

Ala Ile Asn Val Glu Asn Pro Ala His Gln Ile Gln Tyr Pro Ile Pro
        595                 600                 605
```

```
Arg Asn Glu His Glu Thr Thr Ser Leu Gln Ser Pro Gly Glu Ala Pro
    610                 615                 620

Glu Ser Ile Leu Tyr Ser Phe Asp Tyr Arg His Gly Asn Tyr Thr Thr
625                 630                 635                 640

Thr Ala Leu Ser Arg Ile Ser Gln Asp Trp Ala Leu Lys Asp Thr Val
                645                 650                 655

Ser Lys Ile Thr Glu Pro Asp Arg Gln Gln Leu Leu Lys Gln Ala Leu
            660                 665                 670

Glu Cys Leu Gln Ile Ser Glu Glu Thr Gln Glu Lys Lys Glu Lys Glu
        675                 680                 685

Val Gln Gln Leu Ile Ser Asn Leu Arg Gln Gln Gln Gln Leu Tyr Arg
    690                 695                 700

Glu Arg Ile Ile Ser Leu Leu Lys Asp Gln
705                 710
```

<210> 4
<211> 714
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 4

```
Met Gly Ser Ser His His His His His His Gly Ser Asp Tyr Lys Asp
1               5                   10                  15

Asp Asp Asp Lys Ser Gly Ser Leu Glu Val Leu Phe Gln Gly Pro Ser
            20                  25                  30

Gly Met Pro Pro Tyr Trp Arg Gln Lys Tyr Tyr Arg Arg Arg Tyr Arg
        35                  40                  45
```

```
Pro Phe Ser Trp Arg Thr Arg Arg Ile Ile Gln Arg Arg Lys Arg Trp
    50                  55                  60

Arg Tyr Arg Lys Pro Arg Lys Thr Tyr Trp Arg Arg Lys Leu Arg Val
65                  70                  75                  80

Arg Lys Arg Phe Tyr Lys Arg Lys Leu Lys Lys Ile Val Leu Lys Gln
                85                  90                  95

Phe Gln Pro Lys Ile Ile Arg Arg Cys Thr Ile Phe Gly Thr Ile Cys
            100                 105                 110

Leu Phe Gln Gly Ser Asn Leu Arg Leu Gly Met Asn Ser Thr Met Tyr
        115                 120                 125

Glu Lys Ser Ile Val Pro Val His Trp Pro Gly Gly Gly Ser Phe Ser
    130                 135                 140

Val Ser Met Leu Thr Leu Asp Ala Leu Tyr Asp Ile His Lys Leu Cys
145                 150                 155                 160

Arg Asn Trp Trp Thr Ser Thr Asn Gln Asp Leu Pro Leu Val Arg Tyr
                165                 170                 175

Lys Gly Cys Lys Ile Thr Phe Tyr Gln Ser Thr Phe Thr Asp Tyr Ile
            180                 185                 190

Val Arg Ile His Thr Glu Leu Pro Ala Asn Ser Asn Lys Leu Thr Tyr
        195                 200                 205

Pro Asn Thr His Pro Leu Met Met Met Met Ser Lys Tyr Lys His Ile
    210                 215                 220

Ile Pro Ser Arg Gln Thr Arg Arg Lys Lys Lys Pro Tyr Thr Lys Ile
225                 230                 235                 240

Phe Val Lys Pro Pro Pro Gln Phe Glu Asn Lys Trp Tyr Phe Ala Thr
```

```
                245                 250                 255

Asp Leu Tyr Lys Ile Pro Leu Leu Gln Ile His Cys Thr Ala Cys Asn
            260                 265                 270

Leu Gln Asn Pro Phe Val Lys Pro Asp Lys Leu Ser Asn Asn Val Thr
        275                 280                 285

Leu Trp Ser Leu Asn Thr Ile Ser Ile Gln Asn Arg Asn Met Ser Val
    290                 295                 300

Asp Gln Gly Gln Ser Trp Pro Phe Lys Ile Leu Gly Thr Gln Ser Phe
305                 310                 315                 320

Tyr Phe Tyr Phe Tyr Thr Gly Ala Asn Leu Pro Gly Asp Thr Thr Gln
                325                 330                 335

Ile Pro Val Ala Asp Leu Leu Pro Leu Thr Asn Pro Arg Ile Asn Arg
            340                 345                 350

Pro Gly Gln Ser Leu Asn Glu Ala Lys Ile Thr Asp His Ile Thr Phe
        355                 360                 365

Thr Glu Tyr Lys Asn Lys Phe Thr Asn Tyr Trp Gly Asn Pro Phe Asn
    370                 375                 380

Lys His Ile Gln Glu His Leu Asp Met Ile Leu Tyr Ser Leu Lys Ser
385                 390                 395                 400

Pro Glu Ala Ile Lys Asn Glu Trp Thr Thr Glu Asn Met Lys Trp Asn
                405                 410                 415

Gln Leu Asn Asn Ala Gly Thr Met Ala Leu Thr Pro Phe Asn Glu Pro
            420                 425                 430

Ile Phe Thr Gln Ile Gln Tyr Asn Pro Asp Arg Asp Thr Gly Glu Asp
        435                 440                 445
```

```
Thr Gln Leu Tyr Leu Leu Ser Asn Ala Thr Gly Thr Gly Trp Asp Pro
    450                 455                 460

Pro Gly Ile Pro Glu Leu Ile Leu Glu Gly Phe Pro Leu Trp Leu Ile
465                 470                 475                 480

Tyr Trp Gly Phe Ala Asp Phe Gln Lys Asn Leu Lys Lys Val Thr Asn
                485                 490                 495

Ile Asp Thr Asn Tyr Met Leu Val Ala Lys Thr Lys Phe Thr Gln Lys
            500                 505                 510

Pro Gly Thr Phe Tyr Leu Val Ile Leu Asn Asp Thr Phe Val Glu Gly
        515                 520                 525

Asn Ser Pro Tyr Glu Lys Gln Pro Leu Pro Glu Asp Asn Ile Lys Trp
    530                 535                 540

Tyr Pro Gln Val Gln Tyr Gln Leu Glu Ala Gln Asn Lys Leu Leu Gln
545                 550                 555                 560

Thr Gly Pro Phe Thr Pro Asn Ile Gln Gly Gln Leu Ser Asp Asn Ile
                565                 570                 575

Ser Met Phe Tyr Lys Phe Tyr Phe Lys Trp Gly Gly Ser Pro Pro Lys
            580                 585                 590

Ala Ile Asn Val Glu Asn Pro Ala His Gln Ile Gln Tyr Pro Ile Pro
        595                 600                 605

Arg Asn Glu His Glu Thr Thr Ser Leu Gln Ser Pro Gly Glu Ala Pro
    610                 615                 620

Glu Ser Ile Leu Tyr Ser Phe Asp Tyr Arg His Gly Asn Tyr Thr Thr
625                 630                 635                 640

Thr Ala Leu Ser Arg Ile Ser Gln Asp Trp Ala Leu Lys Asp Thr Val
```

645 650 655

Ser Lys Ile Thr Glu Pro Asp Arg Gln Gln Leu Leu Lys Gln Ala Leu
660 665 670

Glu Cys Leu Gln Ile Ser Glu Glu Thr Gln Glu Lys Lys Glu Lys Glu
675 680 685

Val Gln Gln Leu Ile Ser Asn Leu Arg Gln Gln Gln Gln Leu Tyr Arg
690 695 700

Glu Arg Ile Ile Ser Leu Leu Lys Asp Gln
705 710

<210> 5
<211> 688
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 5
Met Gly Ser Ser His His His His His His Gly Ser Asp Tyr Lys Asp
1 5 10 15

Asp Asp Asp Lys Ser Gly Ser Leu Glu Val Leu Phe Gln Gly Pro Ser
20 25 30

Gly Met Pro Pro Tyr Trp Arg Gln Lys Tyr Tyr Arg Arg Arg Tyr Arg
35 40 45

Pro Phe Ser Trp Arg Thr Arg Arg Ile Ile Gln Arg Arg His Arg Trp
50 55 60

Arg Tyr Arg Lys Pro Arg His Thr Tyr Trp Arg Arg His Leu Arg Val
65 70 75 80

Arg His Arg Phe Tyr His Arg His Leu His His Ile Val Leu Lys Gln

```
                85                  90                  95

Phe Gln Pro Lys Ile Ile Arg Arg Cys Thr Ile Phe Gly Thr Ile Cys
            100                 105                 110

Leu Phe Gln Gly Ser Pro Glu Arg Ala Asn Asn Asn Tyr Ile Gln Thr
        115                 120                 125

Ile Tyr Ser Tyr Val Pro Asp Lys Glu Pro Gly Gly Gly Gly Trp Thr
    130                 135                 140

Leu Ile Thr Glu Ser Leu Ser Ser Leu Trp Glu Asp Trp Glu His Leu
145                 150                 155                 160

Lys Asn Val Trp Thr Gln Ser Asn Ala Gly Leu Pro Leu Val Arg Tyr
                165                 170                 175

Gly Gly Val Thr Leu Tyr Phe Tyr Gln Ser Ala Tyr Thr Asp Tyr Ile
            180                 185                 190

Ala Gln Val Phe Asn Cys Tyr Pro Met Thr Asp Thr Lys Tyr Thr His
        195                 200                 205

Ala Asp Ser Ala Pro Asn Arg Met Leu Leu Lys Lys His Val Ile Arg
    210                 215                 220

Val Pro Ser Arg Glu Thr Arg Lys Lys Arg Lys Pro Tyr Lys Arg Val
225                 230                 235                 240

Arg Val Gly Pro Pro Ser Gln Met Gln Asn Lys Trp Tyr Phe Gln Arg
                245                 250                 255

Asp Ile Cys Glu Ile Pro Leu Ile Met Ile Ala Ala Thr Ala Val Asp
            260                 265                 270

Phe Arg Tyr Pro Phe Cys Ala Ser Asp Cys Ala Ser Asn Asn Leu Thr
        275                 280                 285
```

```
Leu Thr Cys Leu Asn Pro Leu Leu Phe Gln Asn Gln Asp Phe Asp His
    290                 295                 300

Pro Ser Asp Thr Gln Gly Tyr Phe Pro Lys Pro Gly Val Tyr Leu Tyr
305                 310                 315                 320

Ser Thr Gln Arg Ser Asn Lys Pro Ser Ser Ser Asp Cys Ile Tyr Leu
                325                 330                 335

Gly Asn Thr Lys Asp Asn Gln Glu Gly Lys Ser Ala Ser Ser Leu Met
            340                 345                 350

Thr Leu Lys Thr Gln Lys Ile Thr Asp Trp Gly Asn Pro Phe Trp His
        355                 360                 365

Tyr Tyr Ile Asp Gly Ser Lys Lys Ile Phe Ser Tyr Phe Lys Pro Pro
    370                 375                 380

Ser Gln Leu Asp Ser Ser Asp Phe Glu His Met Thr Glu Leu Ala Glu
385                 390                 395                 400

Pro Met Phe Ile Gln Val Arg Tyr Asn Pro Glu Arg Asp Thr Gly Gln
                405                 410                 415

Gly Asn Leu Ile Tyr Val Thr Glu Asn Phe Arg Gly Gln His Trp Asp
            420                 425                 430

Pro Pro Ser Ser Asp Asn Leu Lys Leu Asp Gly Phe Pro Leu Tyr Asp
        435                 440                 445

Met Cys Trp Gly Phe Ile Asp Trp Ile Glu Lys Val His Glu Thr Glu
    450                 455                 460

Asn Leu Leu Thr Asn Tyr Cys Phe Cys Ile Arg Ser Ser Ala Phe Asn
465                 470                 475                 480

Glu Lys Lys Thr Val Phe Ile Pro Val Asp His Ser Phe Leu Thr Gly
```

```
                485                 490                 495

Phe Ser Pro Tyr Glu Thr Pro Val Lys Ser Ser Asp Gln Ala His Trp
            500                 505                 510

His Pro Gln Ile Arg Phe Gln Thr Lys Ser Ile Asn Asp Ile Cys Leu
        515                 520                 525

Thr Gly Pro Gly Cys Ala Arg Ser Pro Tyr Gly Asn Tyr Met Gln Ala
    530                 535                 540

Lys Met Ser Tyr Lys Phe His Val Lys Trp Gly Gly Cys Pro Lys Thr
545                 550                 555                 560

Tyr Glu Lys Pro Tyr Asp Pro Cys Ser Gln Pro Asn Trp Thr Ile Pro
                565                 570                 575

His Asn Leu Asn Glu Thr Ile Gln Ile Gln Asn Pro Asn Thr Cys Pro
            580                 585                 590

Gln Thr Glu Leu Gln Glu Trp Asp Trp Arg Arg Asp Ile Val Thr Lys
        595                 600                 605

Lys Ala Ile Glu Arg Ile Arg Gln His Thr Glu Pro His Glu Thr Leu
    610                 615                 620

Gln Ile Ser Thr Gly Ser Lys His Asn Pro Pro Val His Arg Gln Thr
625                 630                 635                 640

Ser Pro Trp Thr Asp Ser Glu Thr Asp Ser Glu Glu Glu Lys Asp Gln
                645                 650                 655

Thr Gln Glu Ile Gln Ile Gln Leu Asn Lys Leu Arg Lys His Gln Gln
            660                 665                 670

His Leu Lys Gln Gln Leu Lys Gln Tyr Leu Lys Pro Gln Asn Ile Glu
        675                 680                 685
```

<210> 6
<211> 687
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 6
Met Gly Ser Ser His His His His His His Gly Ser Asp Tyr Lys Asp
1 5 10 15

Asp Asp Asp Lys Ser Gly Ser Leu Glu Val Leu Phe Gln Gly Pro Ser
20 25 30

Gly Met Trp Gly Thr Ser Asn Cys Ala Cys Ala Lys Phe Gln Ile Arg
35 40 45

Arg Arg Tyr Ala Arg Pro Tyr Arg Arg Arg His Ile Arg Arg Tyr Arg
50 55 60

Arg Arg Arg Arg His Phe Arg Arg Arg Arg Phe Thr Thr Asn Arg Arg
65 70 75 80

Lys Arg Phe Tyr Lys Arg Lys Leu Lys Lys Ile Val Leu Lys Gln Phe
85 90 95

Gln Pro Lys Ile Ile Arg Arg Cys Thr Ile Phe Gly Thr Ile Cys Leu
100 105 110

Phe Gln Gly Ser Pro Glu Arg Ala Asn Asn Asn Tyr Ile Gln Thr Ile
115 120 125

Tyr Ser Tyr Val Pro Asp Lys Glu Pro Gly Gly Gly Gly Trp Thr Leu
130 135 140

Ile Thr Glu Ser Leu Ser Ser Leu Trp Glu Asp Trp Glu His Leu Lys
145 150 155 160

```
Asn Val Trp Thr Gln Ser Asn Ala Gly Leu Pro Leu Val Arg Tyr Gly
                165                 170                 175

Gly Val Thr Leu Tyr Phe Tyr Gln Ser Ala Tyr Thr Asp Tyr Ile Ala
            180                 185                 190

Gln Val Phe Asn Cys Tyr Pro Met Thr Asp Thr Lys Tyr Thr His Ala
        195                 200                 205

Asp Ser Ala Pro Asn Arg Met Leu Leu Lys Lys His Val Ile Arg Val
    210                 215                 220

Pro Ser Arg Glu Thr Arg Lys Lys Arg Lys Pro Tyr Lys Arg Val Arg
225                 230                 235                 240

Val Gly Pro Pro Ser Gln Met Gln Asn Lys Trp Tyr Phe Gln Arg Asp
                245                 250                 255

Ile Cys Glu Ile Pro Leu Ile Met Ile Ala Ala Thr Ala Val Asp Phe
            260                 265                 270

Arg Tyr Pro Phe Cys Ala Ser Asp Cys Ala Ser Asn Asn Leu Thr Leu
        275                 280                 285

Thr Cys Leu Asn Pro Leu Leu Phe Gln Asn Gln Asp Phe Asp His Pro
    290                 295                 300

Ser Asp Thr Gln Gly Tyr Phe Pro Lys Pro Gly Val Tyr Leu Tyr Ser
305                 310                 315                 320

Thr Gln Arg Ser Asn Lys Pro Ser Ser Ser Asp Cys Ile Tyr Leu Gly
                325                 330                 335

Asn Thr Lys Asp Asn Gln Glu Gly Lys Ser Ala Ser Ser Leu Met Thr
            340                 345                 350

Leu Lys Thr Gln Lys Ile Thr Asp Trp Gly Asn Pro Phe Trp His Tyr
```

```
        355                 360                 365

Tyr Ile Asp Gly Ser Lys Lys Ile Phe Ser Tyr Phe Lys Pro Pro Ser
    370                 375                 380

Gln Leu Asp Ser Ser Asp Phe Glu His Met Thr Glu Leu Ala Glu Pro
385                 390                 395                 400

Met Phe Ile Gln Val Arg Tyr Asn Pro Glu Arg Asp Thr Gly Gln Gly
                405                 410                 415

Asn Leu Ile Tyr Val Thr Glu Asn Phe Arg Gly Gln His Trp Asp Pro
            420                 425                 430

Pro Ser Ser Asp Asn Leu Lys Leu Asp Gly Phe Pro Leu Tyr Asp Met
        435                 440                 445

Cys Trp Gly Phe Ile Asp Trp Ile Glu Lys Val His Glu Thr Glu Asn
    450                 455                 460

Leu Leu Thr Asn Tyr Cys Phe Cys Ile Arg Ser Ser Ala Phe Asn Glu
465                 470                 475                 480

Lys Lys Thr Val Phe Ile Pro Val Asp His Ser Phe Leu Thr Gly Phe
                485                 490                 495

Ser Pro Tyr Glu Thr Pro Val Lys Ser Ser Asp Gln Ala His Trp His
            500                 505                 510

Pro Gln Ile Arg Phe Gln Thr Lys Ser Ile Asn Asp Ile Cys Leu Thr
        515                 520                 525

Gly Pro Gly Cys Ala Arg Ser Pro Tyr Gly Asn Tyr Met Gln Ala Lys
    530                 535                 540

Met Ser Tyr Lys Phe His Val Lys Trp Gly Gly Cys Pro Lys Thr Tyr
545                 550                 555                 560
```

```
Glu Lys Pro Tyr Asp Pro Cys Ser Gln Pro Asn Trp Thr Ile Pro His
                565                 570                 575


Asn Leu Asn Glu Thr Ile Gln Ile Gln Asn Pro Asn Thr Cys Pro Gln
            580                 585                 590


Thr Glu Leu Gln Glu Trp Asp Trp Arg Arg Asp Ile Val Thr Lys Lys
        595                 600                 605


Ala Ile Glu Arg Ile Arg Gln His Thr Glu Pro His Glu Thr Leu Gln
    610                 615                 620


Ile Ser Thr Gly Ser Lys His Asn Pro Pro Val His Arg Gln Thr Ser
625                 630                 635                 640


Pro Trp Thr Asp Ser Glu Thr Asp Ser Glu Glu Glu Lys Asp Gln Thr
                645                 650                 655


Gln Glu Ile Gln Ile Gln Leu Asn Lys Leu Arg Lys His Gln Gln His
            660                 665                 670


Leu Lys Gln Gln Leu Lys Gln Tyr Leu Lys Pro Gln Asn Ile Glu
        675                 680                 685


<210> 7
<211> 688
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 7
Met Gly Ser Ser His His His His His His Gly Ser Asp Tyr Lys Asp
1               5                   10                  15


Asp Asp Asp Lys Ser Gly Ser Leu Glu Val Leu Phe Gln Gly Pro Ser
            20                  25                  30
```

```
Gly Met Pro Pro Tyr Trp Arg Gln Lys Tyr Tyr Arg Arg Arg Tyr Arg
        35                  40                  45

Pro Phe Ser Trp Arg Thr Arg Arg Ile Ile Gln Arg Arg Lys Arg Trp
    50                  55                  60

Arg Tyr Arg Lys Pro Arg Lys Thr Tyr Trp Arg Arg Lys Leu Arg Val
65                  70                  75                  80

Arg Lys Arg Phe Tyr Lys Arg Lys Leu Lys Lys Ile Val Leu Lys Gln
                85                  90                  95

Phe Gln Pro Lys Ile Ile Arg Arg Cys Thr Ile Phe Gly Thr Ile Cys
            100                 105                 110

Leu Phe Gln Gly Ser Pro Glu Arg Ala Asn Asn Asn Tyr Ile Gln Thr
        115                 120                 125

Ile Tyr Ser Tyr Val Pro Asp Lys Glu Pro Gly Gly Gly Gly Trp Thr
    130                 135                 140

Leu Ile Thr Glu Ser Leu Ser Ser Leu Trp Glu Asp Trp Glu His Leu
145                 150                 155                 160

Lys Asn Val Trp Thr Gln Ser Asn Ala Gly Leu Pro Leu Val Arg Tyr
                165                 170                 175

Gly Gly Val Thr Leu Tyr Phe Tyr Gln Ser Ala Tyr Thr Asp Tyr Ile
            180                 185                 190

Ala Gln Val Phe Asn Cys Tyr Pro Met Thr Asp Thr Lys Tyr Thr His
        195                 200                 205

Ala Asp Ser Ala Pro Asn Arg Met Leu Leu Lys Lys His Ala Lys Lys
    210                 215                 220

Trp Phe Ser Arg Glu Thr Arg Lys Lys Arg Lys Pro Gly Phe Lys Arg
```

```
225                 230                 235                 240

Leu Leu Gly Pro Pro Ser Gln Met Gln Asn Lys Trp Tyr Phe Gln Arg
                245                 250                 255

Asp Ile Cys Glu Ile Pro Leu Ile Met Ile Ala Ala Thr Ala Val Asp
            260                 265                 270

Phe Arg Tyr Pro Phe Cys Ala Ser Asp Cys Ala Ser Asn Asn Leu Thr
        275                 280                 285

Leu Thr Cys Leu Asn Pro Leu Leu Phe Gln Asn Gln Asp Phe Asp His
    290                 295                 300

Pro Ser Asp Thr Gln Gly Tyr Phe Pro Lys Pro Gly Val Tyr Leu Tyr
305                 310                 315                 320

Ser Thr Gln Arg Ser Asn Lys Pro Ser Ser Ser Asp Cys Ile Tyr Leu
                325                 330                 335

Gly Asn Thr Lys Asp Asn Gln Glu Gly Lys Ser Ala Ser Ser Leu Met
            340                 345                 350

Thr Leu Lys Thr Gln Lys Ile Thr Asp Trp Gly Asn Pro Phe Trp His
        355                 360                 365

Tyr Tyr Ile Asp Gly Ser Lys Lys Ile Phe Ser Tyr Phe Lys Pro Pro
    370                 375                 380

Ser Gln Leu Asp Ser Ser Asp Phe Glu His Met Thr Glu Leu Ala Glu
385                 390                 395                 400

Pro Met Phe Ile Gln Val Arg Tyr Asn Pro Glu Arg Asp Thr Gly Gln
                405                 410                 415

Gly Asn Leu Ile Tyr Val Thr Glu Asn Phe Arg Gly Gln His Trp Asp
            420                 425                 430
```

```
Pro Pro Ser Ser Asp Asn Leu Lys Leu Asp Gly Phe Pro Leu Tyr Asp
        435                 440                 445

Met Cys Trp Gly Phe Ile Asp Trp Ile Glu Lys Val His Glu Thr Glu
    450                 455                 460

Asn Leu Leu Thr Asn Tyr Cys Phe Cys Ile Arg Ser Ser Ala Phe Asn
465                 470                 475                 480

Glu Lys Lys Thr Val Phe Ile Pro Val Asp His Ser Phe Leu Thr Gly
                485                 490                 495

Phe Ser Pro Tyr Glu Thr Pro Val Lys Ser Ser Asp Gln Ala His Trp
            500                 505                 510

His Pro Gln Ile Arg Phe Gln Thr Lys Ser Ile Asn Asp Ile Cys Leu
        515                 520                 525

Thr Gly Pro Gly Cys Ala Arg Ser Pro Tyr Gly Asn Tyr Met Gln Ala
    530                 535                 540

Lys Met Ser Tyr Lys Phe His Val Lys Trp Gly Gly Cys Pro Lys Thr
545                 550                 555                 560

Tyr Glu Lys Pro Tyr Asp Pro Cys Ser Gln Pro Asn Trp Thr Ile Pro
                565                 570                 575

His Asn Leu Asn Glu Thr Ile Gln Ile Gln Asn Pro Asn Thr Cys Pro
            580                 585                 590

Gln Thr Glu Leu Gln Glu Trp Asp Trp Arg Arg Asp Ile Val Thr Lys
        595                 600                 605

Lys Ala Ile Glu Arg Ile Arg Gln His Thr Glu Pro His Glu Thr Leu
    610                 615                 620

Gln Ile Ser Thr Gly Ser Lys His Asn Pro Pro Val His Arg Gln Thr
```

625 630 635 640

Ser Pro Trp Thr Asp Ser Glu Thr Asp Ser Glu Glu Glu Lys Asp Gln
645 650 655

Thr Gln Glu Ile Gln Ile Gln Leu Asn Lys Leu Arg Lys His Gln Gln
660 665 670

His Leu Lys Gln Gln Leu Lys Gln Tyr Leu Lys Pro Gln Asn Ile Glu
675 680 685

<210> 8
<211> 699
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 8
Met Gly Ser Ser His His His His His His Gly Ser Asp Tyr Lys Asp
1 5 10 15

Asp Asp Asp Lys Ser Gly Ser Leu Glu Val Leu Phe Gln Gly Pro Ser
20 25 30

Gly Met Pro Tyr Tyr Tyr Arg Arg Arg Arg Tyr Asn Tyr Arg Arg Pro
35 40 45

Arg Trp Tyr Gly Arg Gly Trp Ile Arg Arg Pro Phe Arg Arg Arg Phe
50 55 60

Arg Arg Lys Arg Arg Val Arg Pro Thr Tyr Thr Thr Ile Pro Leu Lys
65 70 75 80

Gln Trp Gln Pro Pro Tyr Lys Arg Arg Cys Thr Ile Phe Gly Thr Ile
85 90 95

Cys Leu Phe Gln Gly Ser Asn Leu Arg Leu Gly Met Asn Ser Thr Met

```
            100                 105                 110

Tyr Glu Lys Ser Ile Val Pro Val His Trp Pro Gly Gly Gly Ser Phe
        115                 120                 125

Ser Val Ser Met Leu Thr Leu Asp Ala Leu Tyr Asp Ile His Lys Leu
    130                 135                 140

Cys Arg Asn Trp Trp Thr Ser Thr Asn Gln Asp Leu Pro Leu Val Arg
145                 150                 155                 160

Tyr Lys Gly Cys Lys Ile Thr Phe Tyr Gln Ser Thr Phe Thr Asp Tyr
                165                 170                 175

Ile Val Arg Ile His Thr Glu Leu Pro Ala Asn Ser Asn Lys Leu Thr
            180                 185                 190

Tyr Pro Asn Thr His Pro Leu Met Met Met Met Ser Lys Tyr Lys His
        195                 200                 205

Ile Ile Pro Ser Arg Gln Thr Arg Arg Lys Lys Lys Pro Tyr Thr Lys
    210                 215                 220

Ile Phe Val Lys Pro Pro Pro Gln Phe Glu Asn Lys Trp Tyr Phe Ala
225                 230                 235                 240

Thr Asp Leu Tyr Lys Ile Pro Leu Leu Gln Ile His Cys Thr Ala Cys
                245                 250                 255

Asn Leu Gln Asn Pro Phe Val Lys Pro Asp Lys Leu Ser Asn Asn Val
            260                 265                 270

Thr Leu Trp Ser Leu Asn Thr Ile Ser Ile Gln Asn Arg Asn Met Ser
        275                 280                 285

Val Asp Gln Gly Gln Ser Trp Pro Phe Lys Ile Leu Gly Thr Gln Ser
    290                 295                 300
```

```
Phe Tyr Phe Tyr Phe Tyr Thr Gly Ala Asn Leu Pro Gly Asp Thr Thr
305                 310                 315                 320

Gln Ile Pro Val Ala Asp Leu Leu Pro Leu Thr Asn Pro Arg Ile Asn
                325                 330                 335

Arg Pro Gly Gln Ser Leu Asn Glu Ala Lys Ile Thr Asp His Ile Thr
            340                 345                 350

Phe Thr Glu Tyr Lys Asn Lys Phe Thr Asn Tyr Trp Gly Asn Pro Phe
        355                 360                 365

Asn Lys His Ile Gln Glu His Leu Asp Met Ile Leu Tyr Ser Leu Lys
    370                 375                 380

Ser Pro Glu Ala Ile Lys Asn Glu Trp Thr Thr Glu Asn Met Lys Trp
385                 390                 395                 400

Asn Gln Leu Asn Asn Ala Gly Thr Met Ala Leu Thr Pro Phe Asn Glu
                405                 410                 415

Pro Ile Phe Thr Gln Ile Gln Tyr Asn Pro Asp Arg Asp Thr Gly Glu
            420                 425                 430

Asp Thr Gln Leu Tyr Leu Leu Ser Asn Ala Thr Gly Thr Gly Trp Asp
        435                 440                 445

Pro Pro Gly Ile Pro Glu Leu Ile Leu Glu Gly Phe Pro Leu Trp Leu
    450                 455                 460

Ile Tyr Trp Gly Phe Ala Asp Phe Gln Lys Asn Leu Lys Lys Val Thr
465                 470                 475                 480

Asn Ile Asp Thr Asn Tyr Met Leu Val Ala Lys Thr Lys Phe Thr Gln
                485                 490                 495

Lys Pro Gly Thr Phe Tyr Leu Val Ile Leu Asn Asp Thr Phe Val Glu
```

```
            500                 505                 510

Gly Asn Ser Pro Tyr Glu Lys Gln Pro Leu Pro Glu Asp Asn Ile Lys
        515                 520                 525

Trp Tyr Pro Gln Val Gln Tyr Gln Leu Glu Ala Gln Asn Lys Leu Leu
    530                 535                 540

Gln Thr Gly Pro Phe Thr Pro Asn Ile Gln Gly Gln Leu Ser Asp Asn
545                 550                 555                 560

Ile Ser Met Phe Tyr Lys Phe Tyr Phe Lys Trp Gly Gly Ser Pro Pro
                565                 570                 575

Lys Ala Ile Asn Val Glu Asn Pro Ala His Gln Ile Gln Tyr Pro Ile
            580                 585                 590

Pro Arg Asn Glu His Glu Thr Thr Ser Leu Gln Ser Pro Gly Glu Ala
        595                 600                 605

Pro Glu Ser Ile Leu Tyr Ser Phe Asp Tyr Arg His Gly Asn Tyr Thr
    610                 615                 620

Thr Thr Ala Leu Ser Arg Ile Ser Gln Asp Trp Ala Leu Lys Asp Thr
625                 630                 635                 640

Val Ser Lys Ile Thr Glu Pro Asp Arg Gln Gln Leu Leu Lys Gln Ala
                645                 650                 655

Leu Glu Cys Leu Gln Ile Ser Glu Glu Thr Gln Glu Lys Lys Glu Lys
            660                 665                 670

Glu Val Gln Gln Leu Ile Ser Asn Leu Arg Gln Gln Gln Gln Leu Tyr
        675                 680                 685

Arg Glu Arg Ile Ile Ser Leu Leu Lys Asp Gln
    690                 695
```

<210> 9
<211> 699
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 9
Met Gly Ser Ser His His His His His His Gly Ser Asp Tyr Lys Asp
1 5 10 15

Asp Asp Asp Lys Ser Gly Ser Leu Glu Val Leu Phe Gln Gly Pro Ser
20 25 30

Gly Met Pro Tyr Tyr Tyr Arg Arg Arg Arg Tyr Asn Tyr Arg Arg Pro
35 40 45

Arg Trp Tyr Gly Arg Gly Trp Ile Arg Arg Pro Phe Arg Arg Arg Phe
50 55 60

Arg Arg Lys Arg Arg Val Arg Pro Thr Tyr Thr Thr Ile Pro Leu Lys
65 70 75 80

Gln Trp Gln Pro Pro Tyr Lys Arg Thr Cys Tyr Ile Lys Gly Gln Asp
85 90 95

Cys Leu Ile Tyr Tyr Ser Pro Glu Arg Ala Asn Asn Asn Tyr Ile Gln
100 105 110

Thr Ile Tyr Ser Tyr Val Pro Asp Lys Glu Pro Gly Gly Gly Gly Trp
115 120 125

Thr Leu Ile Thr Glu Ser Leu Ser Ser Leu Trp Glu Asp Trp Glu His
130 135 140

Leu Lys Asn Val Trp Thr Gln Ser Asn Ala Gly Leu Pro Leu Val Arg
145 150 155 160

```
Tyr Lys Gly Cys Lys Ile Thr Phe Tyr Gln Ser Thr Phe Thr Asp Tyr
                165                 170                 175

Ile Val Arg Ile His Thr Glu Leu Pro Ala Asn Ser Asn Lys Leu Thr
            180                 185                 190

Tyr Pro Asn Thr His Pro Leu Met Met Met Met Ser Lys Tyr Lys His
        195                 200                 205

Ile Ile Pro Ser Arg Gln Thr Arg Arg Lys Lys Lys Pro Tyr Thr Lys
    210                 215                 220

Ile Phe Val Lys Pro Pro Pro Gln Phe Glu Asn Lys Trp Tyr Phe Ala
225                 230                 235                 240

Thr Asp Leu Tyr Lys Ile Pro Leu Leu Gln Ile His Cys Thr Ala Cys
                245                 250                 255

Asn Leu Gln Asn Pro Phe Val Lys Pro Asp Lys Leu Ser Asn Asn Val
            260                 265                 270

Thr Leu Trp Ser Leu Asn Thr Ile Ser Ile Gln Asn Arg Asn Met Ser
        275                 280                 285

Val Asp Gln Gly Gln Ser Trp Pro Phe Lys Ile Leu Gly Thr Gln Ser
    290                 295                 300

Phe Tyr Phe Tyr Phe Tyr Thr Gly Ala Asn Leu Pro Gly Asp Thr Thr
305                 310                 315                 320

Gln Ile Pro Val Ala Asp Leu Leu Pro Leu Thr Asn Pro Arg Ile Asn
                325                 330                 335

Arg Pro Gly Gln Ser Leu Asn Glu Ala Lys Ile Thr Asp His Ile Thr
            340                 345                 350

Phe Thr Glu Tyr Lys Asn Lys Phe Thr Asn Tyr Trp Gly Asn Pro Phe
```

```
        355                 360                 365

Asn Lys His Ile Gln Glu His Leu Asp Met Ile Leu Tyr Ser Leu Lys
    370                 375                 380

Ser Pro Glu Ala Ile Lys Asn Glu Trp Thr Thr Glu Asn Met Lys Trp
385                 390                 395                 400

Asn Gln Leu Asn Asn Ala Gly Thr Met Ala Leu Thr Pro Phe Asn Glu
                405                 410                 415

Pro Ile Phe Thr Gln Ile Gln Tyr Asn Pro Asp Arg Asp Thr Gly Glu
            420                 425                 430

Asp Thr Gln Leu Tyr Leu Leu Ser Asn Ala Thr Gly Thr Gly Trp Asp
        435                 440                 445

Pro Pro Gly Ile Pro Glu Leu Ile Leu Glu Gly Phe Pro Leu Trp Leu
    450                 455                 460

Ile Tyr Trp Gly Phe Ala Asp Phe Gln Lys Asn Leu Lys Lys Val Thr
465                 470                 475                 480

Asn Ile Asp Thr Asn Tyr Met Leu Val Ala Lys Thr Lys Phe Thr Gln
                485                 490                 495

Lys Pro Gly Thr Phe Tyr Leu Val Ile Leu Asn Asp Thr Phe Val Glu
            500                 505                 510

Gly Asn Ser Pro Tyr Glu Lys Gln Pro Leu Pro Glu Asp Asn Ile Lys
        515                 520                 525

Trp Tyr Pro Gln Val Gln Tyr Gln Leu Glu Ala Gln Asn Lys Leu Leu
    530                 535                 540

Gln Thr Gly Pro Phe Thr Pro Asn Ile Gln Gly Gln Leu Ser Asp Asn
545                 550                 555                 560
```

```
Ile Ser Met Phe Tyr Lys Phe Tyr Phe Lys Trp Gly Gly Ser Pro Pro
                565                 570                 575


Lys Ala Ile Asn Val Glu Asn Pro Ala His Gln Ile Gln Tyr Pro Ile
            580                 585                 590


Pro Arg Asn Glu His Glu Thr Thr Ser Leu Gln Ser Pro Gly Glu Ala
        595                 600                 605


Pro Glu Ser Ile Leu Tyr Ser Phe Asp Tyr Arg His Gly Asn Tyr Thr
    610                 615                 620


Thr Thr Ala Leu Ser Arg Ile Ser Gln Asp Trp Ala Leu Lys Asp Thr
625                 630                 635                 640


Val Ser Lys Ile Thr Glu Pro Asp Arg Gln Gln Leu Leu Lys Gln Ala
                645                 650                 655


Leu Glu Cys Leu Gln Ile Ser Glu Glu Thr Gln Glu Lys Lys Glu Lys
            660                 665                 670


Glu Val Gln Gln Leu Ile Ser Asn Leu Arg Gln Gln Gln Gln Leu Tyr
        675                 680                 685


Arg Glu Arg Ile Ile Ser Leu Leu Lys Asp Gln
    690                 695


<210> 10
<211> 699
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<400> 10
Met Gly Ser Ser His His His His His His Gly Ser Asp Tyr Lys Asp
1               5                   10                  15
```

```
Asp Asp Asp Lys Ser Gly Ser Leu Glu Val Leu Phe Gln Gly Pro Ser
            20                  25                  30

Gly Met Pro Tyr Tyr Tyr Arg Arg Arg Arg Tyr Asn Tyr Arg Arg Pro
        35                  40                  45

Arg Trp Tyr Gly Arg Gly Trp Ile Arg Arg Pro Phe Arg Arg Arg Phe
    50                  55                  60

Arg Arg Lys Arg Arg Val Arg Pro Thr Tyr Thr Thr Ile Pro Leu Lys
65                  70                  75                  80

Gln Trp Gln Pro Pro Tyr Lys Arg Thr Cys Tyr Ile Lys Gly Gln Asp
                85                  90                  95

Cys Leu Ile Tyr Tyr Ser Asn Leu Arg Leu Gly Met Asn Ser Thr Met
            100                 105                 110

Tyr Glu Lys Ser Ile Val Pro Val His Trp Pro Gly Gly Gly Ser Phe
        115                 120                 125

Ser Val Ser Met Leu Thr Leu Asp Ala Leu Tyr Asp Ile His Lys Leu
    130                 135                 140

Cys Arg Asn Trp Trp Thr Ser Thr Asn Gln Asp Leu Pro Leu Val Arg
145                 150                 155                 160

Tyr Gly Gly Val Thr Leu Tyr Phe Tyr Gln Ser Thr Phe Thr Asp Tyr
                165                 170                 175

Ile Val Arg Ile His Thr Glu Leu Pro Ala Asn Ser Asn Lys Leu Thr
            180                 185                 190

Tyr Pro Asn Thr His Pro Leu Met Met Met Met Ser Lys Tyr Lys His
        195                 200                 205

Ile Ile Pro Ser Arg Gln Thr Arg Arg Lys Lys Lys Pro Tyr Thr Lys
```

```
    210                 215                 220

Ile Phe Val Lys Pro Pro Pro Gln Phe Glu Asn Lys Trp Tyr Phe Ala
225                 230                 235                 240

Thr Asp Leu Tyr Lys Ile Pro Leu Leu Gln Ile His Cys Thr Ala Cys
                245                 250                 255

Asn Leu Gln Asn Pro Phe Val Lys Pro Asp Lys Leu Ser Asn Asn Val
            260                 265                 270

Thr Leu Trp Ser Leu Asn Thr Ile Ser Ile Gln Asn Arg Asn Met Ser
        275                 280                 285

Val Asp Gln Gly Gln Ser Trp Pro Phe Lys Ile Leu Gly Thr Gln Ser
    290                 295                 300

Phe Tyr Phe Tyr Phe Tyr Thr Gly Ala Asn Leu Pro Gly Asp Thr Thr
305                 310                 315                 320

Gln Ile Pro Val Ala Asp Leu Leu Pro Leu Thr Asn Pro Arg Ile Asn
                325                 330                 335

Arg Pro Gly Gln Ser Leu Asn Glu Ala Lys Ile Thr Asp His Ile Thr
            340                 345                 350

Phe Thr Glu Tyr Lys Asn Lys Phe Thr Asn Tyr Trp Gly Asn Pro Phe
        355                 360                 365

Asn Lys His Ile Gln Glu His Leu Asp Met Ile Leu Tyr Ser Leu Lys
    370                 375                 380

Ser Pro Glu Ala Ile Lys Asn Glu Trp Thr Thr Glu Asn Met Lys Trp
385                 390                 395                 400

Asn Gln Leu Asn Asn Ala Gly Thr Met Ala Leu Thr Pro Phe Asn Glu
                405                 410                 415
```

```
Pro Ile Phe Thr Gln Ile Gln Tyr Asn Pro Asp Arg Asp Thr Gly Glu
            420                 425                 430

Asp Thr Gln Leu Tyr Leu Leu Ser Asn Ala Thr Gly Thr Gly Trp Asp
        435                 440                 445

Pro Pro Gly Ile Pro Glu Leu Ile Leu Glu Gly Phe Pro Leu Trp Leu
    450                 455                 460

Ile Tyr Trp Gly Phe Ala Asp Phe Gln Lys Asn Leu Lys Lys Val Thr
465                 470                 475                 480

Asn Ile Asp Thr Asn Tyr Met Leu Val Ala Lys Thr Lys Phe Thr Gln
                485                 490                 495

Lys Pro Gly Thr Phe Tyr Leu Val Ile Leu Asn Asp Thr Phe Val Glu
            500                 505                 510

Gly Asn Ser Pro Tyr Glu Lys Gln Pro Leu Pro Glu Asp Asn Ile Lys
        515                 520                 525

Trp Tyr Pro Gln Val Gln Tyr Gln Leu Glu Ala Gln Asn Lys Leu Leu
    530                 535                 540

Gln Thr Gly Pro Phe Thr Pro Asn Ile Gln Gly Gln Leu Ser Asp Asn
545                 550                 555                 560

Ile Ser Met Phe Tyr Lys Phe Tyr Phe Lys Trp Gly Gly Ser Pro Pro
                565                 570                 575

Lys Ala Ile Asn Val Glu Asn Pro Ala His Gln Ile Gln Tyr Pro Ile
            580                 585                 590

Pro Arg Asn Glu His Glu Thr Thr Ser Leu Gln Ser Pro Gly Glu Ala
        595                 600                 605

Pro Glu Ser Ile Leu Tyr Ser Phe Asp Tyr Arg His Gly Asn Tyr Thr
```

610 615 620

Thr Thr Ala Leu Ser Arg Ile Ser Gln Asp Trp Ala Leu Lys Asp Thr
625 630 635 640

Val Ser Lys Ile Thr Glu Pro Asp Arg Gln Gln Leu Leu Lys Gln Ala
645 650 655

Leu Glu Cys Leu Gln Ile Ser Glu Glu Thr Gln Glu Lys Lys Glu Lys
660 665 670

Glu Val Gln Gln Leu Ile Ser Asn Leu Arg Gln Gln Gln Gln Leu Tyr
675 680 685

Arg Glu Arg Ile Ile Ser Leu Leu Lys Asp Gln
690 695

<210> 11
<211> 699
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 11
Met Gly Ser Ser His His His His His His Gly Ser Asp Tyr Lys Asp
1 5 10 15

Asp Asp Asp Lys Ser Gly Ser Leu Glu Val Leu Phe Gln Gly Pro Ser
20 25 30

Gly Met Pro Tyr Tyr Tyr Arg Arg Arg Arg Tyr Asn Tyr Arg Arg Pro
35 40 45

Arg Trp Tyr Gly Arg Gly Trp Ile Arg Arg Pro Phe Arg Arg Arg Phe
50 55 60

Arg Arg Lys Arg Arg Val Arg Pro Thr Tyr Thr Thr Ile Pro Leu Lys

```
65                  70                  75                  80

Gln Trp Gln Pro Pro Tyr Lys Arg Thr Cys Tyr Ile Lys Gly Gln Asp
                85                  90                  95

Cys Leu Ile Tyr Tyr Ser Asn Leu Arg Leu Gly Met Asn Ser Thr Met
            100                 105                 110

Tyr Glu Lys Ser Ile Val Pro Val His Trp Pro Gly Gly Gly Ser Phe
        115                 120                 125

Ser Val Ser Met Leu Thr Leu Asp Ala Leu Tyr Asp Ile His Lys Leu
    130                 135                 140

Cys Arg Asn Trp Trp Thr Ser Thr Asn Gln Asp Leu Pro Leu Val Arg
145                 150                 155                 160

Tyr Lys Gly Cys Lys Ile Thr Phe Tyr Gln Ser Thr Phe Thr Asp Tyr
                165                 170                 175

Ile Val Arg Ile His Asn Cys Tyr Pro Met Thr Asp Thr Lys Tyr Thr
            180                 185                 190

His Ala Asp Ser Ala Pro Asn Arg Met Leu Leu Lys Lys His Lys His
        195                 200                 205

Ile Ile Pro Ser Arg Gln Thr Arg Arg Lys Lys Lys Pro Tyr Thr Lys
    210                 215                 220

Ile Phe Val Lys Pro Pro Pro Gln Phe Glu Asn Lys Trp Tyr Phe Ala
225                 230                 235                 240

Thr Asp Leu Tyr Lys Ile Pro Leu Leu Gln Ile His Cys Thr Ala Cys
                245                 250                 255

Asn Leu Gln Asn Pro Phe Val Lys Pro Asp Lys Leu Ser Asn Asn Val
            260                 265                 270
```

```
Thr Leu Trp Ser Leu Asn Thr Ile Ser Ile Gln Asn Arg Asn Met Ser
        275                 280                 285

Val Asp Gln Gly Gln Ser Trp Pro Phe Lys Ile Leu Gly Thr Gln Ser
    290                 295                 300

Phe Tyr Phe Tyr Phe Tyr Thr Gly Ala Asn Leu Pro Gly Asp Thr Thr
305                 310                 315                 320

Gln Ile Pro Val Ala Asp Leu Leu Pro Leu Thr Asn Pro Arg Ile Asn
                325                 330                 335

Arg Pro Gly Gln Ser Leu Asn Glu Ala Lys Ile Thr Asp His Ile Thr
            340                 345                 350

Phe Thr Glu Tyr Lys Asn Lys Phe Thr Asn Tyr Trp Gly Asn Pro Phe
        355                 360                 365

Asn Lys His Ile Gln Glu His Leu Asp Met Ile Leu Tyr Ser Leu Lys
    370                 375                 380

Ser Pro Glu Ala Ile Lys Asn Glu Trp Thr Thr Glu Asn Met Lys Trp
385                 390                 395                 400

Asn Gln Leu Asn Asn Ala Gly Thr Met Ala Leu Thr Pro Phe Asn Glu
                405                 410                 415

Pro Ile Phe Thr Gln Ile Gln Tyr Asn Pro Asp Arg Asp Thr Gly Glu
            420                 425                 430

Asp Thr Gln Leu Tyr Leu Leu Ser Asn Ala Thr Gly Thr Gly Trp Asp
        435                 440                 445

Pro Pro Gly Ile Pro Glu Leu Ile Leu Glu Gly Phe Pro Leu Trp Leu
    450                 455                 460

Ile Tyr Trp Gly Phe Ala Asp Phe Gln Lys Asn Leu Lys Lys Val Thr
```

```
465                 470                 475                 480

Asn Ile Asp Thr Asn Tyr Met Leu Val Ala Lys Thr Lys Phe Thr Gln
                485                 490                 495

Lys Pro Gly Thr Phe Tyr Leu Val Ile Leu Asn Asp Thr Phe Val Glu
            500                 505                 510

Gly Asn Ser Pro Tyr Glu Lys Gln Pro Leu Pro Glu Asp Asn Ile Lys
        515                 520                 525

Trp Tyr Pro Gln Val Gln Tyr Gln Leu Glu Ala Gln Asn Lys Leu Leu
    530                 535                 540

Gln Thr Gly Pro Phe Thr Pro Asn Ile Gln Gly Gln Leu Ser Asp Asn
545                 550                 555                 560

Ile Ser Met Phe Tyr Lys Phe Tyr Phe Lys Trp Gly Gly Ser Pro Pro
                565                 570                 575

Lys Ala Ile Asn Val Glu Asn Pro Ala His Gln Ile Gln Tyr Pro Ile
            580                 585                 590

Pro Arg Asn Glu His Glu Thr Thr Ser Leu Gln Ser Pro Gly Glu Ala
        595                 600                 605

Pro Glu Ser Ile Leu Tyr Ser Phe Asp Tyr Arg His Gly Asn Tyr Thr
    610                 615                 620

Thr Thr Ala Leu Ser Arg Ile Ser Gln Asp Trp Ala Leu Lys Asp Thr
625                 630                 635                 640

Val Ser Lys Ile Thr Glu Pro Asp Arg Gln Gln Leu Leu Lys Gln Ala
                645                 650                 655

Leu Glu Cys Leu Gln Ile Ser Glu Glu Thr Gln Glu Lys Lys Glu Lys
            660                 665                 670
```

Glu Val Gln Gln Leu Ile Ser Asn Leu Arg Gln Gln Gln Gln Leu Tyr
675 680 685

Arg Glu Arg Ile Ile Ser Leu Leu Lys Asp Gln
690 695

<210> 12
<211> 699
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 12
Met Gly Ser Ser His His His His His His Gly Ser Asp Tyr Lys Asp
1 5 10 15

Asp Asp Asp Lys Ser Gly Ser Leu Glu Val Leu Phe Gln Gly Pro Ser
20 25 30

Gly Met Pro Tyr Tyr Tyr Arg Arg Arg Arg Tyr Asn Tyr Arg Arg Pro
35 40 45

Arg Trp Tyr Gly Arg Gly Trp Ile Arg Arg Pro Phe Arg Arg Arg Phe
50 55 60

Arg Arg Lys Arg Arg Val Arg Pro Thr Tyr Thr Thr Ile Pro Leu Lys
65 70 75 80

Gln Trp Gln Pro Pro Tyr Lys Arg Thr Cys Tyr Ile Lys Gly Gln Asp
85 90 95

Cys Leu Ile Tyr Tyr Ser Asn Leu Arg Leu Gly Met Asn Ser Thr Met
100 105 110

Tyr Glu Lys Ser Ile Val Pro Val His Trp Pro Gly Gly Gly Ser Phe
115 120 125

```
Ser Val Ser Met Leu Thr Leu Asp Ala Leu Tyr Asp Ile His Lys Leu
    130                 135                 140

Cys Arg Asn Trp Trp Thr Ser Thr Asn Gln Asp Leu Pro Leu Val Arg
145                 150                 155                 160

Tyr Lys Gly Cys Lys Ile Thr Phe Tyr Gln Ser Thr Phe Thr Asp Tyr
                165                 170                 175

Ile Val Arg Ile His Thr Glu Leu Pro Ala Asn Ser Asn Lys Leu Thr
            180                 185                 190

Tyr Pro Asn Thr His Pro Leu Met Met Met Met Ser Lys Tyr Lys His
        195                 200                 205

Ile Ile Pro Ser Arg Gln Thr Arg Arg Lys Lys Lys Pro Tyr Lys Arg
    210                 215                 220

Val Arg Val Lys Pro Pro Pro Gln Phe Glu Asn Lys Trp Tyr Phe Ala
225                 230                 235                 240

Thr Asp Leu Tyr Lys Ile Pro Leu Leu Gln Ile His Cys Thr Ala Cys
                245                 250                 255

Asn Leu Gln Asn Pro Phe Val Lys Pro Asp Lys Leu Ser Asn Asn Val
            260                 265                 270

Thr Leu Trp Ser Leu Asn Thr Ile Ser Ile Gln Asn Arg Asn Met Ser
        275                 280                 285

Val Asp Gln Gly Gln Ser Trp Pro Phe Lys Ile Leu Gly Thr Gln Ser
    290                 295                 300

Phe Tyr Phe Tyr Phe Tyr Thr Gly Ala Asn Leu Pro Gly Asp Thr Thr
305                 310                 315                 320

Gln Ile Pro Val Ala Asp Leu Leu Pro Leu Thr Asn Pro Arg Ile Asn
```

```
                325                 330                 335

Arg Pro Gly Gln Ser Leu Asn Glu Ala Lys Ile Thr Asp His Ile Thr
            340                 345                 350

Phe Thr Glu Tyr Lys Asn Lys Phe Thr Asn Tyr Trp Gly Asn Pro Phe
        355                 360                 365

Asn Lys His Ile Gln Glu His Leu Asp Met Ile Leu Tyr Ser Leu Lys
    370                 375                 380

Ser Pro Glu Ala Ile Lys Asn Glu Trp Thr Thr Glu Asn Met Lys Trp
385                 390                 395                 400

Asn Gln Leu Asn Asn Ala Gly Thr Met Ala Leu Thr Pro Phe Asn Glu
                405                 410                 415

Pro Ile Phe Thr Gln Ile Gln Tyr Asn Pro Asp Arg Asp Thr Gly Glu
            420                 425                 430

Asp Thr Gln Leu Tyr Leu Leu Ser Asn Ala Thr Gly Thr Gly Trp Asp
        435                 440                 445

Pro Pro Gly Ile Pro Glu Leu Ile Leu Glu Gly Phe Pro Leu Trp Leu
    450                 455                 460

Ile Tyr Trp Gly Phe Ala Asp Phe Gln Lys Asn Leu Lys Lys Val Thr
465                 470                 475                 480

Asn Ile Asp Thr Asn Tyr Met Leu Val Ala Lys Thr Lys Phe Thr Gln
                485                 490                 495

Lys Pro Gly Thr Phe Tyr Leu Val Ile Leu Asn Asp Thr Phe Val Glu
            500                 505                 510

Gly Asn Ser Pro Tyr Glu Lys Gln Pro Leu Pro Glu Asp Asn Ile Lys
        515                 520                 525
```

```
Trp Tyr Pro Gln Val Gln Tyr Gln Leu Glu Ala Gln Asn Lys Leu Leu
    530                 535                 540

Gln Thr Gly Pro Phe Thr Pro Asn Ile Gln Gly Gln Leu Ser Asp Asn
545                 550                 555                 560

Ile Ser Met Phe Tyr Lys Phe Tyr Phe Lys Trp Gly Gly Ser Pro Pro
                565                 570                 575

Lys Ala Ile Asn Val Glu Asn Pro Ala His Gln Ile Gln Tyr Pro Ile
            580                 585                 590

Pro Arg Asn Glu His Glu Thr Thr Ser Leu Gln Ser Pro Gly Glu Ala
        595                 600                 605

Pro Glu Ser Ile Leu Tyr Ser Phe Asp Tyr Arg His Gly Asn Tyr Thr
    610                 615                 620

Thr Thr Ala Leu Ser Arg Ile Ser Gln Asp Trp Ala Leu Lys Asp Thr
625                 630                 635                 640

Val Ser Lys Ile Thr Glu Pro Asp Arg Gln Gln Leu Leu Lys Gln Ala
                645                 650                 655

Leu Glu Cys Leu Gln Ile Ser Glu Glu Thr Gln Glu Lys Lys Glu Lys
            660                 665                 670

Glu Val Gln Gln Leu Ile Ser Asn Leu Arg Gln Gln Gln Gln Leu Tyr
        675                 680                 685

Arg Glu Arg Ile Ile Ser Leu Leu Lys Asp Gln
    690                 695
```

<210> 13
<211> 699
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 13
Met Gly Ser Ser His His His His His His Gly Ser Asp Tyr Lys Asp
1 5 10 15

Asp Asp Asp Lys Ser Gly Ser Leu Glu Val Leu Phe Gln Gly Pro Ser
20 25 30

Gly Met Pro Tyr Tyr Tyr Arg Arg Arg Arg Tyr Asn Tyr Arg Arg Pro
35 40 45

Arg Trp Tyr Gly Arg Gly Trp Ile Arg Arg Pro Phe Arg Arg Arg Phe
50 55 60

Arg Arg Lys Arg Arg Val Arg Pro Thr Tyr Thr Thr Ile Pro Leu Lys
65 70 75 80

Gln Trp Gln Pro Pro Tyr Lys Arg Thr Cys Tyr Ile Lys Gly Gln Asp
85 90 95

Cys Leu Ile Tyr Tyr Ser Asn Leu Arg Leu Gly Met Asn Ser Thr Met
100 105 110

Tyr Glu Lys Ser Ile Val Pro Val His Trp Pro Gly Gly Gly Ser Phe
115 120 125

Ser Val Ser Met Leu Thr Leu Asp Ala Leu Tyr Asp Ile His Lys Leu
130 135 140

Cys Arg Asn Trp Trp Thr Ser Thr Asn Gln Asp Leu Pro Leu Val Arg
145 150 155 160

Tyr Lys Gly Cys Lys Ile Thr Phe Tyr Gln Ser Thr Phe Thr Asp Tyr
165 170 175

Ile Val Arg Ile His Thr Glu Leu Pro Ala Asn Ser Asn Lys Leu Thr

```
            180                 185                 190

Tyr Pro Asn Thr His Pro Leu Met Met Met Met Ser Lys His Val Ile
        195                 200                 205

Arg Val Pro Ser Arg Gln Thr Arg Arg Lys Lys Lys Pro Tyr Thr Lys
    210                 215                 220

Ile Phe Val Lys Pro Pro Pro Gln Phe Glu Asn Lys Trp Tyr Phe Ala
225                 230                 235                 240

Thr Asp Leu Tyr Lys Ile Pro Leu Leu Gln Ile His Cys Thr Ala Cys
                245                 250                 255

Asn Leu Gln Asn Pro Phe Val Lys Pro Asp Lys Leu Ser Asn Asn Val
            260                 265                 270

Thr Leu Trp Ser Leu Asn Thr Ile Ser Ile Gln Asn Arg Asn Met Ser
        275                 280                 285

Val Asp Gln Gly Gln Ser Trp Pro Phe Lys Ile Leu Gly Thr Gln Ser
    290                 295                 300

Phe Tyr Phe Tyr Phe Tyr Thr Gly Ala Asn Leu Pro Gly Asp Thr Thr
305                 310                 315                 320

Gln Ile Pro Val Ala Asp Leu Leu Pro Leu Thr Asn Pro Arg Ile Asn
                325                 330                 335

Arg Pro Gly Gln Ser Leu Asn Glu Ala Lys Ile Thr Asp His Ile Thr
            340                 345                 350

Phe Thr Glu Tyr Lys Asn Lys Phe Thr Asn Tyr Trp Gly Asn Pro Phe
        355                 360                 365

Asn Lys His Ile Gln Glu His Leu Asp Met Ile Leu Tyr Ser Leu Lys
    370                 375                 380
```

```
Ser Pro Glu Ala Ile Lys Asn Glu Trp Thr Thr Glu Asn Met Lys Trp
385                 390                 395                 400

Asn Gln Leu Asn Asn Ala Gly Thr Met Ala Leu Thr Pro Phe Asn Glu
                405                 410                 415

Pro Ile Phe Thr Gln Ile Gln Tyr Asn Pro Asp Arg Asp Thr Gly Glu
            420                 425                 430

Asp Thr Gln Leu Tyr Leu Leu Ser Asn Ala Thr Gly Thr Gly Trp Asp
        435                 440                 445

Pro Pro Gly Ile Pro Glu Leu Ile Leu Glu Gly Phe Pro Leu Trp Leu
    450                 455                 460

Ile Tyr Trp Gly Phe Ala Asp Phe Gln Lys Asn Leu Lys Lys Val Thr
465                 470                 475                 480

Asn Ile Asp Thr Asn Tyr Met Leu Val Ala Lys Thr Lys Phe Thr Gln
                485                 490                 495

Lys Pro Gly Thr Phe Tyr Leu Val Ile Leu Asn Asp Thr Phe Val Glu
            500                 505                 510

Gly Asn Ser Pro Tyr Glu Lys Gln Pro Leu Pro Glu Asp Asn Ile Lys
        515                 520                 525

Trp Tyr Pro Gln Val Gln Tyr Gln Leu Glu Ala Gln Asn Lys Leu Leu
    530                 535                 540

Gln Thr Gly Pro Phe Thr Pro Asn Ile Gln Gly Gln Leu Ser Asp Asn
545                 550                 555                 560

Ile Ser Met Phe Tyr Lys Phe Tyr Phe Lys Trp Gly Gly Ser Pro Pro
                565                 570                 575

Lys Ala Ile Asn Val Glu Asn Pro Ala His Gln Ile Gln Tyr Pro Ile
```

580 585 590

Pro Arg Asn Glu His Glu Thr Thr Ser Leu Gln Ser Pro Gly Glu Ala
595 600 605

Pro Glu Ser Ile Leu Tyr Ser Phe Asp Tyr Arg His Gly Asn Tyr Thr
610 615 620

Thr Thr Ala Leu Ser Arg Ile Ser Gln Asp Trp Ala Leu Lys Asp Thr
625 630 635 640

Val Ser Lys Ile Thr Glu Pro Asp Arg Gln Gln Leu Leu Lys Gln Ala
645 650 655

Leu Glu Cys Leu Gln Ile Ser Glu Glu Thr Gln Glu Lys Lys Glu Lys
660 665 670

Glu Val Gln Gln Leu Ile Ser Asn Leu Arg Gln Gln Gln Gln Leu Tyr
675 680 685

Arg Glu Arg Ile Ile Ser Leu Leu Lys Asp Gln
690 695

<210> 14
<211> 673
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polypeptide

<400> 14
Met Gly Ser Ser His His His His His His Gly Ser Asp Tyr Lys Asp
1 5 10 15

Asp Asp Asp Lys Ser Gly Ser Leu Glu Val Leu Phe Gln Gly Pro Ser
20 25 30

Gly Met Pro Tyr Tyr Tyr Arg Arg Arg Arg Tyr Asn Tyr Arg Arg Pro

```
        35                  40                  45

Arg Trp Tyr Gly Arg Gly Trp Ile Arg Arg Pro Phe Arg Arg Arg Phe
    50                  55                  60

Arg Arg Lys Arg Arg Val Arg Pro Thr Tyr Thr Thr Ile Pro Leu Lys
65                  70                  75                  80

Gln Trp Gln Pro Pro Tyr Lys Arg Thr Cys Tyr Ile Lys Gly Gln Asp
                85                  90                  95

Cys Leu Ile Tyr Tyr Ser Asn Leu Arg Leu Gly Met Asn Ser Thr Met
            100                 105                 110

Tyr Glu Lys Ser Ile Val Pro Val His Trp Pro Gly Gly Gly Ser Phe
        115                 120                 125

Ser Val Ser Met Leu Thr Leu Asp Ala Leu Tyr Asp Ile His Lys Leu
    130                 135                 140

Cys Arg Asn Trp Trp Thr Ser Thr Asn Gln Asp Leu Pro Leu Val Arg
145                 150                 155                 160

Tyr Lys Gly Cys Lys Ile Thr Phe Tyr Gln Ser Thr Phe Thr Asp Tyr
                165                 170                 175

Ile Val Arg Ile His Thr Glu Leu Pro Ala Asn Ser Asn Lys Leu Thr
            180                 185                 190

Tyr Pro Asn Thr His Pro Leu Met Met Met Met Ser Lys Tyr Lys His
        195                 200                 205

Ile Ile Pro Ser Arg Gln Thr Arg Arg Lys Lys Lys Pro Tyr Thr Lys
    210                 215                 220

Ile Phe Val Lys Pro Pro Pro Gln Phe Glu Asn Lys Trp Tyr Phe Ala
225                 230                 235                 240
```

```
Thr Asp Leu Tyr Lys Ile Pro Leu Ile Met Ile Ala Ala Thr Ala Val
                245                 250                 255

Asp Phe Arg Tyr Pro Phe Cys Ala Ser Asp Cys Ala Ser Asn Asn Leu
            260                 265                 270

Thr Leu Thr Cys Leu Asn Pro Leu Leu Phe Gln Asn Gln Asp Phe Asp
        275                 280                 285

His Pro Ser Asp Thr Gln Gly Tyr Phe Pro Lys Pro Gly Val Tyr Leu
    290                 295                 300

Tyr Ser Thr Gln Arg Ser Asn Lys Pro Ser Ser Ser Asp Cys Ile Tyr
305                 310                 315                 320

Leu Gly Asn Thr Lys Asp Asn Gln Glu Gly Lys Ser Ala Ser Ser Leu
                325                 330                 335

Met Thr Leu Lys Thr Gln Lys Ile Thr Asp Trp Gly Asn Pro Phe Trp
            340                 345                 350

His Tyr Tyr Ile Asp Gly Ser Lys Lys Ile Phe Ser Tyr Phe Lys Pro
        355                 360                 365

Pro Ser Gln Leu Asp Ser Ser Asp Phe Glu His Met Thr Glu Leu Ala
    370                 375                 380

Glu Pro Met Phe Ile Gln Val Arg Tyr Asn Pro Glu Arg Asp Thr Gly
385                 390                 395                 400

Gln Gly Asn Leu Ile Tyr Val Thr Glu Asn Phe Arg Gly Gln His Trp
                405                 410                 415

Asp Pro Pro Ser Ser Asp Asn Leu Lys Leu Asp Gly Phe Pro Leu Tyr
            420                 425                 430

Asp Met Cys Trp Gly Phe Ile Asp Trp Ile Glu Lys Val His Glu Thr
```

```
        435                 440                 445

Glu Asn Leu Leu Thr Asn Tyr Cys Phe Cys Ile Arg Ser Ser Ala Phe
    450                 455                 460

Asn Glu Lys Lys Thr Val Phe Ile Pro Val Asp His Ser Phe Leu Thr
465                 470                 475                 480

Gly Phe Ser Pro Tyr Glu Thr Pro Val Lys Ser Ser Asp Gln Ala His
                485                 490                 495

Trp His Pro Gln Ile Arg Phe Gln Thr Lys Ser Ile Asn Asp Ile Cys
            500                 505                 510

Leu Thr Gly Pro Gly Cys Ala Arg Ser Pro Tyr Gly Asn Tyr Met Gln
        515                 520                 525

Ala Lys Met Ser Tyr Lys Phe His Val Lys Trp Gly Gly Cys Pro Lys
    530                 535                 540

Thr Tyr Glu Lys Pro Tyr Asp Pro Cys Ser Gln Pro Asn Trp Thr Ile
545                 550                 555                 560

Pro His Asn Leu Asn Glu Thr Ile Gln Ile Gln Asn Pro Asn Thr Cys
                565                 570                 575

Pro Gln Thr Glu Leu Gln Glu Trp Asp Trp Arg Arg Asp Ile Val Thr
            580                 585                 590

Lys Lys Ala Ile Glu Arg Ile Arg Gln His Thr Glu Pro His Glu Thr
        595                 600                 605

Leu Gln Ile Ser Thr Gly Ser Lys His Asn Pro Pro Val His Arg Gln
    610                 615                 620

Thr Ser Pro Trp Thr Asp Ser Glu Thr Asp Ser Glu Glu Glu Lys Asp
625                 630                 635                 640
```

Gln Thr Gln Glu Ile Gln Ile Gln Leu Asn Lys Leu Arg Lys His Gln
645 650 655

Gln His Leu Lys Gln Gln Leu Lys Gln Tyr Leu Lys Pro Gln Asn Ile
660 665 670

Glu

<210> 15
<211> 673
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 15
Met Gly Ser Ser His His His His His His Gly Ser Asp Tyr Lys Asp
1 5 10 15

Asp Asp Asp Lys Ser Gly Ser Leu Glu Val Leu Phe Gln Gly Pro Ser
20 25 30

Gly Met Pro Tyr Tyr Tyr Arg Arg Arg Arg Tyr Asn Tyr Arg Arg Pro
35 40 45

Arg Trp Tyr Gly Arg Gly Trp Ile Arg Arg Pro Phe Arg Arg Arg Phe
50 55 60

Arg Arg Lys Arg Arg Val Arg Pro Thr Tyr Thr Thr Ile Pro Leu Lys
65 70 75 80

Gln Trp Gln Pro Pro Tyr Lys Arg Thr Cys Tyr Ile Lys Gly Gln Asp
85 90 95

Cys Leu Ile Tyr Tyr Ser Asn Leu Arg Leu Gly Met Asn Ser Thr Met
100 105 110

```
Tyr Glu Lys Ser Ile Val Pro Val His Trp Pro Gly Gly Gly Ser Phe
        115                 120                 125

Ser Val Ser Met Leu Thr Leu Asp Ala Leu Tyr Asp Ile His Lys Leu
    130                 135                 140

Cys Arg Asn Trp Trp Thr Ser Thr Asn Gln Asp Leu Pro Leu Val Arg
145                 150                 155                 160

Tyr Lys Gly Cys Lys Ile Thr Phe Tyr Gln Ser Thr Phe Thr Asp Tyr
                165                 170                 175

Ile Val Arg Ile His Thr Glu Leu Pro Ala Asn Ser Asn Lys Leu Thr
            180                 185                 190

Tyr Pro Asn Thr His Pro Leu Met Met Met Met Ser Lys Tyr Lys His
        195                 200                 205

Ile Ile Pro Ser Arg Gln Thr Arg Arg Lys Lys Lys Pro Tyr Thr Lys
    210                 215                 220

Ile Phe Val Lys Pro Pro Pro Gln Phe Glu Asn Lys Trp Tyr Phe Ala
225                 230                 235                 240

Thr Asp Leu Tyr Lys Ile Pro Leu Leu Gln Ile His Cys Thr Ala Cys
                245                 250                 255

Asn Leu Gln Asn Pro Phe Val Lys Pro Asp Lys Leu Ser Asn Asn Leu
            260                 265                 270

Thr Leu Thr Cys Leu Asn Pro Leu Leu Phe Gln Asn Gln Asp Phe Asp
        275                 280                 285

His Pro Ser Asp Thr Gln Gly Tyr Phe Pro Lys Pro Gly Val Tyr Leu
    290                 295                 300

Tyr Ser Thr Gln Arg Ser Asn Lys Pro Ser Ser Ser Asp Cys Ile Tyr
```

```
305                 310                 315                 320

Leu Gly Asn Thr Lys Asp Asn Gln Glu Gly Lys Ser Ala Ser Ser Leu
                325                 330                 335

Met Thr Leu Lys Thr Gln Lys Ile Thr Asp Trp Gly Asn Pro Phe Trp
            340                 345                 350

His Tyr Tyr Ile Asp Gly Ser Lys Lys Ile Phe Ser Tyr Phe Lys Pro
        355                 360                 365

Pro Ser Gln Leu Asp Ser Ser Asp Phe Glu His Met Thr Glu Leu Ala
    370                 375                 380

Glu Pro Met Phe Ile Gln Val Arg Tyr Asn Pro Glu Arg Asp Thr Gly
385                 390                 395                 400

Gln Gly Asn Leu Ile Tyr Val Thr Glu Asn Phe Arg Gly Gln His Trp
                405                 410                 415

Asp Pro Pro Ser Ser Asp Asn Leu Lys Leu Asp Gly Phe Pro Leu Tyr
            420                 425                 430

Asp Met Cys Trp Gly Phe Ile Asp Trp Ile Glu Lys Val His Glu Thr
        435                 440                 445

Glu Asn Leu Leu Thr Asn Tyr Cys Phe Cys Ile Arg Ser Ser Ala Phe
    450                 455                 460

Asn Glu Lys Lys Thr Val Phe Ile Pro Val Asp His Ser Phe Leu Thr
465                 470                 475                 480

Gly Phe Ser Pro Tyr Glu Thr Pro Val Lys Ser Ser Asp Gln Ala His
                485                 490                 495

Trp His Pro Gln Ile Arg Phe Gln Thr Lys Ser Ile Asn Asp Ile Cys
            500                 505                 510
```

```
Leu Thr Gly Pro Gly Cys Ala Arg Ser Pro Tyr Gly Asn Tyr Met Gln
        515                 520                 525

Ala Lys Met Ser Tyr Lys Phe His Val Lys Trp Gly Gly Cys Pro Lys
    530                 535                 540

Thr Tyr Glu Lys Pro Tyr Asp Pro Cys Ser Gln Pro Asn Trp Thr Ile
545                 550                 555                 560

Pro His Asn Leu Asn Glu Thr Ile Gln Ile Gln Asn Pro Asn Thr Cys
                565                 570                 575

Pro Gln Thr Glu Leu Gln Glu Trp Asp Trp Arg Arg Asp Ile Val Thr
            580                 585                 590

Lys Lys Ala Ile Glu Arg Ile Arg Gln His Thr Glu Pro His Glu Thr
        595                 600                 605

Leu Gln Ile Ser Thr Gly Ser Lys His Asn Pro Pro Val His Arg Gln
    610                 615                 620

Thr Ser Pro Trp Thr Asp Ser Glu Thr Asp Ser Glu Glu Glu Lys Asp
625                 630                 635                 640

Gln Thr Gln Glu Ile Gln Ile Gln Leu Asn Lys Leu Arg Lys His Gln
                645                 650                 655

Gln His Leu Lys Gln Gln Leu Lys Gln Tyr Leu Lys Pro Gln Asn Ile
            660                 665                 670

Glu
```

<210> 16
<211> 3753
<212> DNA
<213> Alphatorquevirus sp.

<400> 16
tgctacgtca ctaacccacg tgtcctctac aggccaatcg cagtctatgt cgtgcacttc 60

ctgggcatgg tctacataat tatataaatg cttgcacttc cgaatggctg agtttttgct 120

gcccgtccgc ggagaggagc cacggcaggg gatccgaacg tcctgagggc gggtgccgga 180

ggtgagttta cacaccgaag tcaaggggca attcgggctc aggactggcc gggctttggg 240

caaggctctt aaaaatgcac ttttctcgaa taagcagaaa gaaaaggaaa gtgctactgc 300

tttgcgtgcc agcagctaag aaaaaaccaa ctgctatgag cttctggaaa cctccggtac 360

acaatgtcac ggggatccaa cgcatgtggt atgagtcctt tcaccgtggc cacgcttctt 420

tttgtggttg tgggaatcct atacttcaca ttactgcact tgctgaaaca tatggccatc 480

caacaggccc gagaccttct gggccaccgg gagtagaccc caacccccac atccgtagag 540

ccaggcctgc cccggccgct ccggagccct cacaggttga ttcgagacca gccctgacat 600

ggcatgggga tggtggaagc gacggaggcg ctggtggttc cggaagcggt ggacccgtgg 660

cagacttcgc agacgatggc ctcgatcagc tcgtcgccgc cctagacgac gaagagtaag 720

gaggcgcaga cggtggagga gggggagacg aaaaacaagg acttacagac gcaggagacg 780

ctttagacgc aggggacgaa aagcaaaact tataataaaa ctgtggcaac ctgcagtaat 840

taaaagatgc agaataaagg gatacatacc actgattata agtgggaacg gtacctttgc 900

cacaaacttt accagtcaca taaatgacag aataatgaaa ggccccttcg ggggaggaca 960

cagcactatg aggttcagcc tctacatttt gtttgaggag cacctcagac acatgaactt 1020

ctggaccaga agcaacgata acctagagct aaccagatac ttgggggctt cagtaaaaat 1080

atacaggcac ccagaccaag actttatagt aatatacaac agaagaaccc ctctaggagg 1140

caacatctac acagcaccct ctctacaccc aggcaatgcc attttagcaa aacacaaaat 1200

attagtacca agtttacaga caagaccaaa gggtagaaaa gcaattagac taagaatagc 1260

acccccaca ctctttacag acaagtggta ctttcaaaag gacatagccg acctcaccct 1320

tttcaacatc atggcagttg aggctgactt gcggtttccg ttctgctcac cacaaactga 1380

caacacttgc atcagcttcc aggtccttag ttccgtttac aacaactacc tcagtattaa 1440

tacctttaat aatgacaact cagactcaaa gttaaaagaa tttttaaata aagcatttcc 1500

aacaacaggc acaaaaggaa caagtttaaa tgcactaaat acatttagaa cagaaggatg 1560

cataagtcac ccacaactaa aaaaaccaaa cccacaaata aacaaaccat tagagtcaca 1620

atactttgca cctttagatg ccctctgggg agaccccata tactataatg atctaaatga 1680

aaacaaaagt ttgaacgata tcattgagaa aatactaata aaaaacatga ttacatacca 1740

tgcaaaacta agagaatttc caaattcata ccaaggaaac aaggcctttt gccacctaac 1800

aggcatatac agcccaccat acctaaacca aggcagaata tctccagaaa tatttggact 1860

gtacacagaa ataatttaca acccttacac agacaaagga actggaaaca aagtatggat 1920

ggacccacta actaaagaga acaacatata taaagaagga cagagcaaat gcctactgac 1980

tgacatgccc ctatggactt tacttttttgg atatacagac tggtgtaaaa aggacactaa 2040

taactgggac ttaccactaa actacagact agtactaata tgcccttata cctttccaaa 2100

attgtacaat gaaaaagtaa aagactatgg gtacatcccg tactcctaca aattcggagc 2160

gggtcagatg ccagacggca gcaactacat acccttttcag tttagagcaa agtggtaccc 2220

cacagtacta caccagcaac aggtaatgga ggacataagc aggagcgggc cctttgcacc 2280

taaggtagaa aaaccaagca ctcagctggt aatgaagtac tgttttaact ttaactgggg 2340

cggtaaccct atcattgaac agattgttaa agaccccagc ttccagccca cctatgaaat 2400

acccggtacc ggtaacatcc ctagaagaat acaagtcatc gacccgcggg tcctgggacc 2460

gcactactcg ttccggtcat gggacatgcg cagacacaca tttagcagag caagtattaa 2520

gagagtgtca gaacaacaag aaacttctga ccttgtattc tcaggcccaa aaaagcctcg 2580

ggtcgacatc ccaaaacaag aaacccaaga agaaagctca cattcactcc aaagagaatc 2640

gagaccgtgg gagaccgagg aagaaagcga gacagaagcc ctctcgcaag agagccaaga 2700

ggtccccttc caacagcagt tgcagcagca gtaccaagag cagctcaagc tcagacaggg 2760

aatcaaagtc ctcttcgagc agctcataag gacccaacaa ggggtccatg taaacccatg 2820

cctacggtag gtcccaggca gtggctgttt ccagagagaa agccagcccc agctcctagc 2880

agtggagact gggccatgga gtttctcgca gcaaaaatat ttgataggcc agttagaagc 2940

aaccttaaag atacccctta ctacccatat gttaaaaacc aatacaatgt ctactttgac 3000

cttaaatttg aataaacagc agcttcaaac ttgcaaggcc gtgggagttt cactggtcgg 3060

tgtctacctc taaaggtcac taagcactcc gagcgtaagc gaggagtgcg accctccccc 3120

ctggaacaac ttcttcggag tccggcgcta cgccttcggc tgcgccggac acctcagacc 3180

ccccctccac ccgaaacgct tgcgcgtttc ggaccttcgg cgtcgggggg gtcgggagct 3240

ttattaaacg gactccgaag tgctcttgga cactgagggg gtgaacagca acgaaagtga 3300

gtggggccag acttcgccat aaggccttta tcttcttgcc atttgtcagt gtccggggtc 3360

gccataggct tcgggctcgt ttttaggcct tccggactac aaaaatcgcc attttggtga 3420

cgtcacggcc gccatcttaa gtagttgagg cggacggtgg cgtgagttca aaggtcacca 3480

tcagccacac ctactcaaaa tggtggacaa tttcttccgg gtcaaaggtt acagccgcca 3540

tgttaaaaca cgtgacgtat gacgtcacgg ccgccatttt gtgacacaag atggccgact 3600

tccttcctct ttttcaaaaa aaagcggaag tgccgccgcg gcggcggggg gcggcgcgct 3660

gcgcgcgccg cccagtaggg ggagccatgc gccccccccc gcgcatgcgc ggggcccccc 3720

cccgcggggg gctccgcccc ccggcccccc ccg 3753

<210> 17
<211> 127
<212> PRT
<213> Alphatorquevirus sp.

<400> 17
Met Ser Phe Trp Lys Pro Pro Val His Asn Val Thr Gly Ile Gln Arg
1 5 10 15

Met Trp Tyr Glu Ser Phe His Arg Gly His Ala Ser Phe Cys Gly Cys
20 25 30

Gly Asn Pro Ile Leu His Ile Thr Ala Leu Ala Glu Thr Tyr Gly His
35 40 45

Pro Thr Gly Pro Arg Pro Ser Gly Pro Pro Gly Val Asp Pro Asn Pro
50 55 60

```
His Ile Arg Arg Ala Arg Pro Ala Pro Ala Ala Pro Glu Pro Ser Gln
65                  70                  75                  80


Val Asp Ser Arg Pro Ala Leu Thr Trp His Gly Asp Gly Gly Ser Asp
                85                  90                  95


Gly Gly Ala Gly Gly Ser Gly Ser Gly Gly Pro Val Ala Asp Phe Ala
            100                 105                 110


Asp Asp Gly Leu Asp Gln Leu Val Ala Ala Leu Asp Asp Glu Glu
        115                 120                 125


<210> 18
<211> 268
<212> PRT
<213> Alphatorquevirus sp.

<400> 18
Met Ser Phe Trp Lys Pro Pro Val His Asn Val Thr Gly Ile Gln Arg
1               5                   10                  15


Met Trp Tyr Glu Ser Phe His Arg Gly His Ala Ser Phe Cys Gly Cys
            20                  25                  30


Gly Asn Pro Ile Leu His Ile Thr Ala Leu Ala Glu Thr Tyr Gly His
        35                  40                  45


Pro Thr Gly Pro Arg Pro Ser Gly Pro Pro Gly Val Asp Pro Asn Pro
    50                  55                  60


His Ile Arg Arg Ala Arg Pro Ala Pro Ala Ala Pro Glu Pro Ser Gln
65                  70                  75                  80


Val Asp Ser Arg Pro Ala Leu Thr Trp His Gly Asp Gly Gly Ser Asp
                85                  90                  95


Gly Gly Ala Gly Gly Ser Gly Ser Gly Gly Pro Val Ala Asp Phe Ala
            100                 105                 110
```

```
Asp Asp Gly Leu Asp Gln Leu Val Ala Ala Leu Asp Asp Glu Glu Leu
        115                 120                 125

Leu Lys Thr Pro Ala Ser Ser Pro Pro Met Lys Tyr Pro Val Pro Val
    130                 135                 140

Thr Ser Leu Glu Glu Tyr Lys Ser Ser Thr Arg Gly Ser Trp Asp Arg
145                 150                 155                 160

Thr Thr Arg Ser Gly His Gly Thr Cys Ala Asp Thr His Leu Ala Glu
                165                 170                 175

Gln Val Leu Arg Glu Cys Gln Asn Asn Lys Lys Leu Leu Thr Leu Tyr
            180                 185                 190

Ser Gln Ala Gln Lys Ser Leu Gly Ser Thr Ser Gln Asn Lys Lys Pro
        195                 200                 205

Lys Lys Lys Ala His Ile His Ser Lys Glu Asn Arg Asp Arg Gly Arg
    210                 215                 220

Pro Arg Lys Lys Ala Arg Gln Lys Pro Ser Arg Lys Arg Ala Lys Arg
225                 230                 235                 240

Ser Pro Ser Asn Ser Ser Cys Ser Ser Ser Thr Lys Ser Ser Ser Ser
                245                 250                 255

Ser Asp Arg Glu Ser Lys Ser Ser Ser Ser Ser Ser Ser
            260                 265
```

<210> 19
<211> 276
<212> PRT
<213> Alphatorquevirus sp.

<400> 19

```
Met Ser Phe Trp Lys Pro Pro Val His Asn Val Thr Gly Ile Gln Arg
1               5                   10                  15
```

```
Met Trp Tyr Glu Ser Phe His Arg Gly His Ala Ser Phe Cys Gly Cys
            20                  25                  30

Gly Asn Pro Ile Leu His Ile Thr Ala Leu Ala Glu Thr Tyr Gly His
        35                  40                  45

Pro Thr Gly Pro Arg Pro Ser Gly Pro Pro Gly Val Asp Pro Asn Pro
    50                  55                  60

His Ile Arg Arg Ala Arg Pro Ala Pro Ala Ala Pro Glu Pro Ser Gln
65                  70                  75                  80

Val Asp Ser Arg Pro Ala Leu Thr Trp His Gly Asp Gly Gly Ser Asp
                85                  90                  95

Gly Gly Ala Gly Gly Ser Gly Ser Gly Gly Pro Val Ala Asp Phe Ala
            100                 105                 110

Asp Asp Gly Leu Asp Gln Leu Val Ala Ala Leu Asp Asp Glu Glu Pro
        115                 120                 125

Lys Lys Ala Ser Gly Arg His Pro Lys Thr Arg Asn Pro Arg Arg Lys
    130                 135                 140

Leu Thr Phe Thr Pro Lys Arg Ile Glu Thr Val Gly Asp Arg Gly Arg
145                 150                 155                 160

Lys Arg Asp Arg Ser Pro Leu Ala Arg Glu Pro Arg Gly Pro Leu Pro
                165                 170                 175

Thr Ala Val Ala Ala Ala Val Pro Arg Ala Ala Gln Ala Gln Thr Gly
            180                 185                 190

Asn Gln Ser Pro Leu Arg Ala Ala His Lys Asp Pro Thr Arg Gly Pro
        195                 200                 205

Cys Lys Pro Met Pro Thr Val Gly Pro Arg Gln Trp Leu Phe Pro Glu
    210                 215                 220
```

```
Arg Lys Pro Ala Pro Ala Pro Ser Ser Gly Asp Trp Ala Met Glu Phe
225                 230                 235                 240


Leu Ala Ala Lys Ile Phe Asp Arg Pro Val Arg Ser Asn Leu Lys Asp
                245                 250                 255


Thr Pro Tyr Tyr Pro Tyr Val Lys Asn Gln Tyr Asn Val Tyr Phe Asp
            260                 265                 270


Leu Lys Phe Glu
        275


<210> 20
<211> 167
<212> PRT
<213> Alphatorquevirus sp.

<400> 20
Met Ser Phe Trp Lys Pro Pro Val His Asn Val Thr Gly Ile Gln Arg
1               5                   10                  15


Met Trp Pro Lys Lys Ala Ser Gly Arg His Pro Lys Thr Arg Asn Pro
            20                  25                  30


Arg Arg Lys Leu Thr Phe Thr Pro Lys Arg Ile Glu Thr Val Gly Asp
        35                  40                  45


Arg Gly Arg Lys Arg Asp Arg Ser Pro Leu Ala Arg Glu Pro Arg Gly
    50                  55                  60


Pro Leu Pro Thr Ala Val Ala Ala Ala Val Pro Arg Ala Ala Gln Ala
65                  70                  75                  80


Gln Thr Gly Asn Gln Ser Pro Leu Arg Ala Ala His Lys Asp Pro Thr
                85                  90                  95


Arg Gly Pro Cys Lys Pro Met Pro Thr Val Gly Pro Arg Gln Trp Leu
            100                 105                 110
```

```
Phe Pro Glu Arg Lys Pro Ala Pro Ala Pro Ser Ser Gly Asp Trp Ala
        115                 120                 125


Met Glu Phe Leu Ala Ala Lys Ile Phe Asp Arg Pro Val Arg Ser Asn
    130                 135                 140


Leu Lys Asp Thr Pro Tyr Tyr Pro Tyr Val Lys Asn Gln Tyr Asn Val
145                 150                 155                 160


Tyr Phe Asp Leu Lys Phe Glu
                165
```

<210> 21
<211> 743
<212> PRT
<213> Alphatorquevirus sp.

<400> 21

```
Met Ala Trp Gly Trp Trp Lys Arg Arg Arg Arg Trp Trp Phe Arg Lys
1               5                   10                  15


Arg Trp Thr Arg Gly Arg Leu Arg Arg Arg Trp Pro Arg Ser Ala Arg
            20                  25                  30


Arg Arg Pro Arg Arg Arg Arg Val Arg Arg Arg Arg Arg Trp Arg Arg
        35                  40                  45


Gly Arg Arg Lys Thr Arg Thr Tyr Arg Arg Arg Arg Arg Phe Arg Arg
    50                  55                  60


Arg Gly Arg Lys Ala Lys Leu Ile Ile Lys Leu Trp Gln Pro Ala Val
65                  70                  75                  80


Ile Lys Arg Cys Arg Ile Lys Gly Tyr Ile Pro Leu Ile Ile Ser Gly
                85                  90                  95


Asn Gly Thr Phe Ala Thr Asn Phe Thr Ser His Ile Asn Asp Arg Ile
            100                 105                 110
```

```
Met Lys Gly Pro Phe Gly Gly Gly His Ser Thr Met Arg Phe Ser Leu
        115                 120                 125

Tyr Ile Leu Phe Glu Glu His Leu Arg His Met Asn Phe Trp Thr Arg
    130                 135                 140

Ser Asn Asp Asn Leu Glu Leu Thr Arg Tyr Leu Gly Ala Ser Val Lys
145                 150                 155                 160

Ile Tyr Arg His Pro Asp Gln Asp Phe Ile Val Ile Tyr Asn Arg Arg
                165                 170                 175

Thr Pro Leu Gly Gly Asn Ile Tyr Thr Ala Pro Ser Leu His Pro Gly
            180                 185                 190

Asn Ala Ile Leu Ala Lys His Lys Ile Leu Val Pro Ser Leu Gln Thr
        195                 200                 205

Arg Pro Lys Gly Arg Lys Ala Ile Arg Leu Arg Ile Ala Pro Pro Thr
    210                 215                 220

Leu Phe Thr Asp Lys Trp Tyr Phe Gln Lys Asp Ile Ala Asp Leu Thr
225                 230                 235                 240

Leu Phe Asn Ile Met Ala Val Glu Ala Asp Leu Arg Phe Pro Phe Cys
                245                 250                 255

Ser Pro Gln Thr Asp Asn Thr Cys Ile Ser Phe Gln Val Leu Ser Ser
            260                 265                 270

Val Tyr Asn Asn Tyr Leu Ser Ile Asn Thr Phe Asn Asn Asp Asn Ser
        275                 280                 285

Asp Ser Lys Leu Lys Glu Phe Leu Asn Lys Ala Phe Pro Thr Thr Gly
    290                 295                 300
```

```
Thr Lys Gly Thr Ser Leu Asn Ala Leu Asn Thr Phe Arg Thr Glu Gly
305                 310                 315                 320

Cys Ile Ser His Pro Gln Leu Lys Lys Pro Asn Pro Gln Ile Asn Lys
                325                 330                 335

Pro Leu Glu Ser Gln Tyr Phe Ala Pro Leu Asp Ala Leu Trp Gly Asp
            340                 345                 350

Pro Ile Tyr Tyr Asn Asp Leu Asn Glu Asn Lys Ser Leu Asn Asp Ile
        355                 360                 365

Ile Glu Lys Ile Leu Ile Lys Asn Met Ile Thr Tyr His Ala Lys Leu
    370                 375                 380

Arg Glu Phe Pro Asn Ser Tyr Gln Gly Asn Lys Ala Phe Cys His Leu
385                 390                 395                 400

Thr Gly Ile Tyr Ser Pro Pro Tyr Leu Asn Gln Gly Arg Ile Ser Pro
                405                 410                 415

Glu Ile Phe Gly Leu Tyr Thr Glu Ile Ile Tyr Asn Pro Tyr Thr Asp
            420                 425                 430

Lys Gly Thr Gly Asn Lys Val Trp Met Asp Pro Leu Thr Lys Glu Asn
        435                 440                 445

Asn Ile Tyr Lys Glu Gly Gln Ser Lys Cys Leu Leu Thr Asp Met Pro
    450                 455                 460

Leu Trp Thr Leu Leu Phe Gly Tyr Thr Asp Trp Cys Lys Lys Asp Thr
465                 470                 475                 480

Asn Asn Trp Asp Leu Pro Leu Asn Tyr Arg Leu Val Leu Ile Cys Pro
                485                 490                 495

Tyr Thr Phe Pro Lys Leu Tyr Asn Glu Lys Val Lys Asp Tyr Gly Tyr
            500                 505                 510
```

```
Ile Pro Tyr Ser Tyr Lys Phe Gly Ala Gly Gln Met Pro Asp Gly Ser
        515                 520                 525

Asn Tyr Ile Pro Phe Gln Phe Arg Ala Lys Trp Tyr Pro Thr Val Leu
    530                 535                 540

His Gln Gln Gln Val Met Glu Asp Ile Ser Arg Ser Gly Pro Phe Ala
545                 550                 555                 560

Pro Lys Val Glu Lys Pro Ser Thr Gln Leu Val Met Lys Tyr Cys Phe
                565                 570                 575

Asn Phe Asn Trp Gly Gly Asn Pro Ile Ile Glu Gln Ile Val Lys Asp
            580                 585                 590

Pro Ser Phe Gln Pro Thr Tyr Glu Ile Pro Gly Thr Gly Asn Ile Pro
        595                 600                 605

Arg Arg Ile Gln Val Ile Asp Pro Arg Val Leu Gly Pro His Tyr Ser
    610                 615                 620

Phe Arg Ser Trp Asp Met Arg Arg His Thr Phe Ser Arg Ala Ser Ile
625                 630                 635                 640

Lys Arg Val Ser Glu Gln Gln Glu Thr Ser Asp Leu Val Phe Ser Gly
                645                 650                 655

Pro Lys Lys Pro Arg Val Asp Ile Pro Lys Gln Glu Thr Gln Glu Glu
            660                 665                 670

Ser Ser His Ser Leu Gln Arg Glu Ser Arg Pro Trp Glu Thr Glu Glu
        675                 680                 685

Glu Ser Glu Thr Glu Ala Leu Ser Gln Glu Ser Gln Glu Val Pro Phe
    690                 695                 700
```

```
Gln Gln Gln Leu Gln Gln Gln Tyr Gln Glu Gln Leu Lys Leu Arg Gln
705                 710                 715                 720


Gly Ile Lys Val Leu Phe Glu Gln Leu Ile Arg Thr Gln Gln Gly Val
                725                 730                 735


His Val Asn Pro Cys Leu Arg
            740
```

<210> 22
<211> 194
<212> PRT
<213> Alphatorquevirus sp.

<400> 22

```
Met Ala Trp Gly Trp Trp Lys Arg Arg Arg Arg Trp Trp Phe Arg Lys
1               5                   10                  15


Arg Trp Thr Arg Gly Arg Leu Arg Arg Arg Trp Pro Arg Ser Ala Arg
            20                  25                  30


Arg Arg Pro Arg Arg Arg Arg Ile Val Lys Asp Pro Ser Phe Gln Pro
        35                  40                  45


Thr Tyr Glu Ile Pro Gly Thr Gly Asn Ile Pro Arg Arg Ile Gln Val
    50                  55                  60


Ile Asp Pro Arg Val Leu Gly Pro His Tyr Ser Phe Arg Ser Trp Asp
65                  70                  75                  80


Met Arg Arg His Thr Phe Ser Arg Ala Ser Ile Lys Arg Val Ser Glu
                85                  90                  95


Gln Gln Glu Thr Ser Asp Leu Val Phe Ser Gly Pro Lys Lys Pro Arg
            100                 105                 110


Val Asp Ile Pro Lys Gln Glu Thr Gln Glu Glu Ser Ser His Ser Leu
        115                 120                 125
```

```
Gln Arg Glu Ser Arg Pro Trp Glu Thr Glu Glu Glu Ser Glu Thr Glu
    130                 135                 140

Ala Leu Ser Gln Glu Ser Gln Glu Val Pro Phe Gln Gln Gln Leu Gln
145                 150                 155                 160

Gln Gln Tyr Gln Glu Gln Leu Lys Leu Arg Gln Gly Ile Lys Val Leu
                165                 170                 175

Phe Glu Gln Leu Ile Arg Thr Gln Gln Gly Val His Val Asn Pro Cys
            180                 185                 190

Leu Arg
```

<210> 23
<211> 113
<212> PRT
<213> Alphatorquevirus sp.

<400> 23

```
Met Ala Trp Gly Trp Trp Lys Arg Arg Arg Arg Trp Trp Phe Arg Lys
1               5                   10                  15

Arg Trp Thr Arg Gly Arg Leu Arg Arg Arg Trp Pro Arg Ser Ala Arg
            20                  25                  30

Arg Arg Pro Arg Arg Arg Arg Ala Gln Lys Ser Leu Gly Ser Thr Ser
        35                  40                  45

Gln Asn Lys Lys Pro Lys Lys Lys Ala His Ile His Ser Lys Glu Asn
    50                  55                  60

Arg Asp Arg Gly Arg Pro Arg Lys Lys Ala Arg Gln Lys Pro Ser Arg
65                  70                  75                  80

Lys Arg Ala Lys Arg Ser Pro Ser Asn Ser Ser Cys Ser Ser Ser Thr
                85                  90                  95
```

Lys Ser Ser Ser Ser Ser Asp Arg Glu Ser Lys Ser Ser Ser Ser Ser
100 105 110

Ser

<210> 24

<400> 24
000

<210> 25

<400> 25
000

<210> 26

<400> 26
000

<210> 27

<400> 27
000

<210> 28

<400> 28
000

<210> 29

<400> 29
000

<210> 30

<400> 30
000

<210> 31

<400> 31
000


<210> 32

<400> 32
000


<210> 33

<400> 33
000


<210> 34

<400> 34
000


<210> 35

<400> 35
000


<210> 36

<400> 36
000


<210> 37

<400> 37
000


<210> 38

<400> 38
000


<210> 39

<400> 39
000

<210> 40

<400> 40
000

<210> 41

<400> 41
000

<210> 42

<400> 42
000

<210> 43

<400> 43
000

<210> 44

<400> 44
000

<210> 45

<400> 45
000

<210> 46

<400> 46
000

<210> 47

<400> 47
000

<210> 48

<400> 48
000


<210> 49

<400> 49
000


<210> 50

<400> 50
000


<210> 51

<400> 51
000


<210> 52

<400> 52
000


<210> 53

<400> 53
000


<210> 54
<211> 2979
<212> DNA
<213> Betatorquevirus sp.

<400> 54
taataaatat tcaacaggaa aaccacctaa tttaaattgc cgaccacaaa ccgtcactta 60

gttccccttt ttgcaacaac ttctgctttt ttccaactgc cggaaaacca cataatttgc 120

atggctaacc acaaactgat atgctaatta acttccacaa aacaacttcc ccttttaaaa 180

ccacacctac aaattaatta ttaaacacag tcacatcctg ggaggtacta ccacactata 240

ataccaagtg cacttccgaa tggctgagtt tatgccgcta gacggagaac gcatcagtta 300

ctgactgcgg actgaacttg ggcgggtgcc gaaggtgagt gaaaccaccg aagtcaaggg 360

gcaattcggg ctagttcagt ctagcggaac gggcaagaaa cttaaaatta ttttattttt 420

cagatgagcg actgctttaa accaacatgc tacaacaaca aaacaaagca aactcactgg 480

attaataacc tgcatttaac ccacgacctg atctgcttct gcccaacacc aactagacac 540

ttattactag ctttagcaga acaacaagaa acaattgaag tgtctaaaca agaaaaagaa 600

aaaataacaa gatgccttat tactacagaa gaagacggta caactacaga cgtcctagat 660

ggtatggacg aggttggatt agacgccctt ttcgcagaag atttcgaaga aaaagaaggg 720

taagacctac ttatactact attcctctaa agcaatggca accgccatat aaaagaacat 780

gctatataaa aggacaagac tgtttaatat actatagcaa cttaagactg ggaatgaata 840

gtacaatgta tgaaaaaagt attgtacctg tacattggcc gggaggggggt tctttttctg 900

taagcatgtt aactttagat gccttgtatg atatacataa actttgtaga aactggtgga 960

catccacaaa ccaagactta ccactagtaa gatataaagg atgcaaaata acattttatc 1020

aaagcacatt tacagactac atagtaagaa tacatacaga actaccagct aacagtaaca 1080

aactaacata cccaaacaca catccactaa tgatgatgat gtctaagtac aaacacatta 1140

tacctagtag acaaacaaga agaaaaaaga aaccatacac aaaaatattt gtaaaaccac 1200

ctccgcaatt tgaaaacaaa tggtactttg ctacagacct ctacaaaatt ccattactac 1260

aaatacactg cacagcatgc aacttacaaa acccatttgt aaaaccagac aaattatcaa 1320

acaatgttac attatggtca ctaaacacca taagcataca aaatagaaac atgtcagtgg 1380

atcaaggaca atcatggcca tttaaaatac taggaacaca aagcttttat ttttacttt 1440

acaccggagc aaacctacca ggtgacacaa cacaaatacc agtagcagac ctattaccac 1500

taacaaaccc aagaataaac agaccaggac aatcactaaa tgaggcaaaa attacagacc 1560

atattacttt cacagaatac aaaaacaaat ttacaaatta ttggggtaac ccatttaata 1620

aacacattca agaacaccta gatatgatac tatactcact aaaaagtcca gaagcaataa 1680

aaaacgaatg gacaacagaa aacatgaaat ggaaccaatt aaacaatgca ggaacaatgg 1740

cattaacacc atttaacgag ccaatattca cacaaataca atataaccca gatagagaca 1800

caggagaaga cactcaatta tacctactct ctaacgctac aggaacagga tgggacccac 1860

caggaattcc agaattaata ctagaaggat ttccactatg gttaatatat tggggatttg 1920

cagactttca aaaaaaccta aaaaaagtaa caaacataga cacaaattac atgttagtag 1980

caaaaacaaa atttacacaa aaacctggca cattctactt agtaatacta aatgacacct 2040

ttgtagaagg caatagccca tatgaaaaac aacctttacc tgaagacaac attaaatggt 2100

acccacaagt acaataccaa ttagaagcac aaaacaaact actacaaact gggccattta 2160

caccaaacat acaaggacaa ctatcagaca atatatcaat gttttataaa ttttacttta 2220

aatggggagg aagcccacca aaagcaatta atgttgaaaa tcctgcccac cagattcaat 2280

atcccatacc ccgtaacgag catgaaacaa cttcgttaca gagtccaggg gaagccccag 2340

aatccatctt atactccttc gactatagac acgggaacta cacaacaaca gctttgtcac 2400

gaattagcca agactgggca cttaaagaca ctgtttctaa aattacagag ccagatcgac 2460

agcaactgct caaacaagcc ctcgaatgcc tgcaaatctc ggaagaaacg caggagaaaa 2520

aagaaaaaga agtacagcag ctcatcagca acctcagaca gcagcagcag ctgtacagag 2580

agcgaataat atcattatta aaggaccaat aacttttaac tgtgtaaaaa aggtgaaatt 2640

gtttgatgat aaaccaaaaa accgtagatt tacacctgag gaatttgaaa ctgagttaca 2700

aatagcaaaa tggttaaaga gacccccaag atcctttgta aatgatcctc ccttttaccc 2760

atggttacca cctgaacctg ttgtaaactt taagcttaat tttactgaat aaaggccagc 2820

attaattcac ttaaggagtc tgtttattta agttaaacct taataaacgg tcaccgcctc 2880

cctaatacgc aggcgcagaa agggggctcc gccccttta acccccaggg ggctccgccc 2940

cctgaaaccc ccaaggggggc tacgcccccc tacaccccc 2979

<210> 55
<211> 99
<212> PRT
<213> Betatorquevirus sp.

<400> 55
Met Ser Asp Cys Phe Lys Pro Thr Cys Tyr Asn Asn Lys Thr Lys Gln
1               5                  10                  15

Thr His Trp Ile Asn Asn Leu His Leu Thr His Asp Leu Ile Cys Phe
            20                  25                  30

Cys Pro Thr Pro Thr Arg His Leu Leu Leu Ala Leu Ala Glu Gln Gln
        35                  40                  45

Glu Thr Ile Glu Val Ser Lys Gln Glu Lys Glu Lys Ile Thr Arg Cys
    50                  55                  60

Leu Ile Thr Thr Glu Glu Asp Gly Thr Thr Thr Asp Val Leu Asp Gly
65                  70                  75                  80

Met Asp Glu Val Gly Leu Asp Ala Leu Phe Ala Glu Asp Phe Glu Glu
                85                  90                  95

Lys Glu Gly

<210> 56
<211> 203
<212> PRT
<213> Betatorquevirus sp.

<400> 56
Met Ser Asp Cys Phe Lys Pro Thr Cys Tyr Asn Asn Lys Thr Lys Gln
1               5                  10                  15

Thr His Trp Ile Asn Asn Leu His Leu Thr His Asp Leu Ile Cys Phe
            20                  25                  30

Cys Pro Thr Pro Thr Arg His Leu Leu Leu Ala Leu Ala Glu Gln Gln
        35                  40                  45

Glu Thr Ile Glu Val Ser Lys Gln Glu Lys Glu Lys Ile Thr Arg Cys
    50                  55                  60

```
Leu Ile Thr Thr Glu Glu Asp Gly Thr Thr Thr Asp Val Leu Asp Gly
65                  70                  75                  80


Met Asp Glu Val Gly Leu Asp Ala Leu Phe Ala Glu Asp Phe Glu Glu
                85                  90                  95


Lys Glu Gly Phe Asn Ile Pro Tyr Pro Val Thr Ser Met Lys Gln Leu
            100                 105                 110


Arg Tyr Arg Val Gln Gly Lys Pro Gln Asn Pro Ser Tyr Thr Pro Ser
        115                 120                 125


Thr Ile Asp Thr Gly Thr Thr Gln Gln Gln Leu Cys His Glu Leu Ala
    130                 135                 140


Lys Thr Gly His Leu Lys Thr Leu Phe Leu Lys Leu Gln Ser Gln Ile
145                 150                 155                 160


Asp Ser Asn Cys Ser Asn Lys Pro Ser Asn Ala Cys Lys Ser Arg Lys
                165                 170                 175


Lys Arg Arg Arg Lys Lys Lys Lys Lys Tyr Ser Ser Ser Ser Ala Thr
            180                 185                 190


Ser Asp Ser Ser Ser Ser Cys Thr Glu Ser Glu
        195                 200


<210> 57
<211> 219
<212> PRT
<213> Betatorquevirus sp.

<400> 57
Met Ser Asp Cys Phe Lys Pro Thr Cys Tyr Asn Asn Lys Thr Lys Gln
1               5                   10                  15


Thr His Trp Ile Asn Asn Leu His Leu Thr His Asp Leu Ile Cys Phe
            20                  25                  30
```

```
Cys Pro Thr Pro Thr Arg His Leu Leu Leu Ala Leu Ala Glu Gln Gln
        35                  40                  45

Glu Thr Ile Glu Val Ser Lys Gln Glu Lys Glu Lys Ile Thr Arg Cys
    50                  55                  60

Leu Ile Thr Thr Glu Glu Asp Gly Thr Thr Thr Asp Val Leu Asp Gly
65                  70                  75                  80

Met Asp Glu Val Gly Leu Asp Ala Leu Phe Ala Glu Asp Phe Glu Glu
                85                  90                  95

Lys Glu Gly Ala Arg Ser Thr Ala Thr Ala Gln Thr Ser Pro Arg Met
            100                 105                 110

Pro Ala Asn Leu Gly Arg Asn Ala Gly Glu Lys Arg Lys Arg Ser Thr
        115                 120                 125

Ala Ala His Gln Gln Pro Gln Thr Ala Ala Ala Ala Val Gln Arg Ala
    130                 135                 140

Asn Asn Ile Ile Ile Lys Gly Pro Ile Thr Phe Asn Cys Val Lys Lys
145                 150                 155                 160

Val Lys Leu Phe Asp Asp Lys Pro Lys Asn Arg Arg Phe Thr Pro Glu
                165                 170                 175

Glu Phe Glu Thr Glu Leu Gln Ile Ala Lys Trp Leu Lys Arg Pro Pro
            180                 185                 190

Arg Ser Phe Val Asn Asp Pro Pro Phe Tyr Pro Trp Leu Pro Pro Glu
        195                 200                 205

Pro Val Val Asn Phe Lys Leu Asn Phe Thr Glu
    210                 215
```

<210> 58

<211> 666
<212> PRT
<213> Betatorquevirus sp.

<400> 58
Met Pro Tyr Tyr Tyr Arg Arg Arg Arg Tyr Asn Tyr Arg Arg Pro Arg
1               5                  10                  15

Trp Tyr Gly Arg Gly Trp Ile Arg Arg Pro Phe Arg Arg Arg Phe Arg
            20                  25                  30

Arg Lys Arg Arg Val Arg Pro Thr Tyr Thr Thr Ile Pro Leu Lys Gln
        35                  40                  45

Trp Gln Pro Pro Tyr Lys Arg Thr Cys Tyr Ile Lys Gly Gln Asp Cys
    50                  55                  60

Leu Ile Tyr Tyr Ser Asn Leu Arg Leu Gly Met Asn Ser Thr Met Tyr
65                  70                  75                  80

Glu Lys Ser Ile Val Pro Val His Trp Pro Gly Gly Gly Ser Phe Ser
                85                  90                  95

Val Ser Met Leu Thr Leu Asp Ala Leu Tyr Asp Ile His Lys Leu Cys
            100                 105                 110

Arg Asn Trp Trp Thr Ser Thr Asn Gln Asp Leu Pro Leu Val Arg Tyr
        115                 120                 125

Lys Gly Cys Lys Ile Thr Phe Tyr Gln Ser Thr Phe Thr Asp Tyr Ile
    130                 135                 140

Val Arg Ile His Thr Glu Leu Pro Ala Asn Ser Asn Lys Leu Thr Tyr
145                 150                 155                 160

Pro Asn Thr His Pro Leu Met Met Met Met Ser Lys Tyr Lys His Ile
                165                 170                 175

Ile Pro Ser Arg Gln Thr Arg Arg Lys Lys Lys Pro Tyr Thr Lys Ile

```
            180                 185                 190

Phe Val Lys Pro Pro Pro Gln Phe Glu Asn Lys Trp Tyr Phe Ala Thr
        195                 200                 205

Asp Leu Tyr Lys Ile Pro Leu Leu Gln Ile His Cys Thr Ala Cys Asn
    210                 215                 220

Leu Gln Asn Pro Phe Val Lys Pro Asp Lys Leu Ser Asn Asn Val Thr
225                 230                 235                 240

Leu Trp Ser Leu Asn Thr Ile Ser Ile Gln Asn Arg Asn Met Ser Val
                245                 250                 255

Asp Gln Gly Gln Ser Trp Pro Phe Lys Ile Leu Gly Thr Gln Ser Phe
            260                 265                 270

Tyr Phe Tyr Phe Tyr Thr Gly Ala Asn Leu Pro Gly Asp Thr Thr Gln
        275                 280                 285

Ile Pro Val Ala Asp Leu Leu Pro Leu Thr Asn Pro Arg Ile Asn Arg
    290                 295                 300

Pro Gly Gln Ser Leu Asn Glu Ala Lys Ile Thr Asp His Ile Thr Phe
305                 310                 315                 320

Thr Glu Tyr Lys Asn Lys Phe Thr Asn Tyr Trp Gly Asn Pro Phe Asn
                325                 330                 335

Lys His Ile Gln Glu His Leu Asp Met Ile Leu Tyr Ser Leu Lys Ser
            340                 345                 350

Pro Glu Ala Ile Lys Asn Glu Trp Thr Thr Glu Asn Met Lys Trp Asn
        355                 360                 365

Gln Leu Asn Asn Ala Gly Thr Met Ala Leu Thr Pro Phe Asn Glu Pro
    370                 375                 380
```

```
Ile Phe Thr Gln Ile Gln Tyr Asn Pro Asp Arg Asp Thr Gly Glu Asp
385                 390                 395                 400

Thr Gln Leu Tyr Leu Leu Ser Asn Ala Thr Gly Thr Gly Trp Asp Pro
                405                 410                 415

Pro Gly Ile Pro Glu Leu Ile Leu Glu Gly Phe Pro Leu Trp Leu Ile
            420                 425                 430

Tyr Trp Gly Phe Ala Asp Phe Gln Lys Asn Leu Lys Lys Val Thr Asn
        435                 440                 445

Ile Asp Thr Asn Tyr Met Leu Val Ala Lys Thr Lys Phe Thr Gln Lys
    450                 455                 460

Pro Gly Thr Phe Tyr Leu Val Ile Leu Asn Asp Thr Phe Val Glu Gly
465                 470                 475                 480

Asn Ser Pro Tyr Glu Lys Gln Pro Leu Pro Glu Asp Asn Ile Lys Trp
                485                 490                 495

Tyr Pro Gln Val Gln Tyr Gln Leu Glu Ala Gln Asn Lys Leu Leu Gln
            500                 505                 510

Thr Gly Pro Phe Thr Pro Asn Ile Gln Gly Gln Leu Ser Asp Asn Ile
        515                 520                 525

Ser Met Phe Tyr Lys Phe Tyr Phe Lys Trp Gly Gly Ser Pro Pro Lys
    530                 535                 540

Ala Ile Asn Val Glu Asn Pro Ala His Gln Ile Gln Tyr Pro Ile Pro
545                 550                 555                 560

Arg Asn Glu His Glu Thr Thr Ser Leu Gln Ser Pro Gly Glu Ala Pro
                565                 570                 575

Glu Ser Ile Leu Tyr Ser Phe Asp Tyr Arg His Gly Asn Tyr Thr Thr
```

```
            580                 585                 590

Thr Ala Leu Ser Arg Ile Ser Gln Asp Trp Ala Leu Lys Asp Thr Val
        595                 600                 605

Ser Lys Ile Thr Glu Pro Asp Arg Gln Gln Leu Leu Lys Gln Ala Leu
    610                 615                 620

Glu Cys Leu Gln Ile Ser Glu Glu Thr Gln Glu Lys Lys Glu Lys Glu
625                 630                 635                 640

Val Gln Gln Leu Ile Ser Asn Leu Arg Gln Gln Gln Gln Leu Tyr Arg
                645                 650                 655

Glu Arg Ile Ile Ser Leu Leu Lys Asp Gln
            660                 665
```

<210> 59
<211> 148
<212> PRT
<213> Betatorquevirus sp.

<400> 59

```
Met Pro Tyr Tyr Tyr Arg Arg Arg Arg Tyr Asn Tyr Arg Arg Pro Arg
1               5                   10                  15

Trp Tyr Gly Arg Gly Trp Ile Arg Arg Pro Phe Arg Arg Arg Phe Arg
            20                  25                  30

Arg Lys Arg Arg Ile Gln Tyr Pro Ile Pro Arg Asn Glu His Glu Thr
        35                  40                  45

Thr Ser Leu Gln Ser Pro Gly Glu Ala Pro Glu Ser Ile Leu Tyr Ser
    50                  55                  60

Phe Asp Tyr Arg His Gly Asn Tyr Thr Thr Thr Ala Leu Ser Arg Ile
65                  70                  75                  80

Ser Gln Asp Trp Ala Leu Lys Asp Thr Val Ser Lys Ile Thr Glu Pro
```

```
                85                  90                  95

Asp Arg Gln Gln Leu Leu Lys Gln Ala Leu Glu Cys Leu Gln Ile Ser
            100                 105                 110

Glu Glu Thr Gln Glu Lys Lys Glu Lys Glu Val Gln Gln Leu Ile Ser
        115                 120                 125

Asn Leu Arg Gln Gln Gln Gln Leu Tyr Arg Glu Arg Ile Ile Ser Leu
    130                 135                 140

Leu Lys Asp Gln
145
```

<210> 60
<211> 82
<212> PRT
<213> Betatorquevirus sp.

<400> 60

```
Met Pro Tyr Tyr Tyr Arg Arg Arg Arg Tyr Asn Tyr Arg Arg Pro Arg
1               5                   10                  15

Trp Tyr Gly Arg Gly Trp Ile Arg Arg Pro Phe Arg Arg Arg Phe Arg
            20                  25                  30

Arg Lys Arg Arg Ser Gln Ile Asp Ser Asn Cys Ser Asn Lys Pro Ser
        35                  40                  45

Asn Ala Cys Lys Ser Arg Lys Lys Arg Arg Arg Lys Lys Lys Lys Lys
    50                  55                  60

Tyr Ser Ser Ser Ser Ala Thr Ser Asp Ser Ser Ser Ser Cys Thr Glu
65                  70                  75                  80

Ser Glu
```

<210> 61

<400> 61
000

<210> 62

<400> 62
000

<210> 63

<400> 63
000

<210> 64

<400> 64
000

<210> 65

<400> 65
000

<210> 66

<400> 66
000

<210> 67

<400> 67
000

<210> 68

<400> 68
000

<210> 69

<400> 69

000

<210> 70

<400> 70
000

<210> 71

<400> 71
000

<210> 72

<400> 72
000

<210> 73

<400> 73
000

<210> 74

<400> 74
000

<210> 75

<400> 75
000

<210> 76

<400> 76
000

<210> 77

<400> 77
000

<210> 78

<400> 78
000


<210> 79

<400> 79
000


<210> 80

<400> 80
000


<210> 81

<400> 81
000


<210> 82

<400> 82
000


<210> 83

<400> 83
000


<210> 84

<400> 84
000


<210> 85

<400> 85
000


<210> 86

<400> 86
000

<210> 87

<400> 87
000

<210> 88

<400> 88
000

<210> 89

<400> 89
000

<210> 90

<400> 90
000

<210> 91

<400> 91
000

<210> 92

<400> 92
000

<210> 93

<400> 93
000

<210> 94

<400> 94

000

<210> 95

<400> 95
000

<210> 96

<400> 96
000

<210> 97

<400> 97
000

<210> 98

<400> 98
000

<210> 99

<400> 99
000

<210> 100

<400> 100
000

<210> 101

<400> 101
000

<210> 102

<400> 102
000

<210> 103

<400> 103
000

<210> 104

<400> 104
000

<210> 105
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 105
cgggtgccgk aggtgagttt acacaccgma gtcaaggggc aattcgggct crggactggc 60

cgggcyhtgg g 71

<210> 106
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 106
cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60

cgggctwtgg g 71

<210> 107
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 107
cgggtgccgt aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60

cgggctatgg g 71

<210> 108
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 108
cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60

cgggccctgg g 71

<210> 109
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 109
cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60

cgggctttgg g 71

<210> 110
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 110
cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60

cgggctatgg g 71

<210> 111
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 111
cgggtgccgg aggtgagttt acacaccgaa gtcaaggggc aattcgggct caggactggc 60

cgggctttgg g 71

<210> 112
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 112
cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60

cgggcyhtgg g 71

<210> 113
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 113
cgggtgccgt aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60

cgggctatgg g 71

<210> 114
<211> 70

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 114
cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60

cgggcccggg 70

<210> 115
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 115
cgggtgccgg aggtgagttt acacaccgaa gtcaaggggc aattcgggct caggactggc 60

cgggctttgg g 71

<210> 116
<211> 69
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 116
cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggaggccg 60

ggccatggg 69

<210> 117
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 117
cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60

cgggccccgg g 71

<210> 118
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 118
cgggtgccgg aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60

cgggctatgg g 71

<210> 119
<211> 71
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 119
cgggtgccga aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 60

cgggctatgg g 71

<210> 120
<211> 117
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<220>
<221> misc_feature

<222> (10)..(10)
<223> May or may not be present

<220>
<221> misc_feature
<222> (12)..(12)
<223> May or may not be present

<220>
<221> misc_feature
<222> (30)..(32)
<223> May or may not be present

<220>
<221> misc_feature
<222> (34)..(34)
<223> May or may not be present

<220>
<221> misc_feature
<222> (43)..(46)
<223> May or may not be present

<220>
<221> misc_feature
<222> (52)..(54)
<223> May or may not be present

<220>
<221> misc_feature
<222> (70)..(71)
<223> May or may not be present

<220>
<221> misc_feature
<222> (89)..(90)
<223> May or may not be present

<220>
<221> misc_feature
<222> (103)..(103)
<223> May or may not be present

<400> 120
cggcggsggs gcsscgcgct dcgcgcgcsg cccrsyrggg grdssmmwgc skcsccccccc 60

cscgcgcatg cgcrcgggkc ccccccccyv sgggggggctc cgccccccccg gcccccc 117

<210> 121

<211> 169
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<220>
<221> modified_base
<222> (20)..(20)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (22)..(22)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (40)..(42)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (53)..(56)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (62)..(62)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (64)..(64)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (97)..(98)
<223> a, c, t, g, unknown or other

<400> 121
gccgccgcgg cggcggsggn gnsgcgcgct dcgcgcgcsn nncrccrggg ggnnnncwgc 60

sncnccccccc cccgcgcatg cgcgggkccc ccccccnncg gggggctccg ccccccggcc 120

ccccccgtg ctaaacccac cgcgcatgcg cgaccacgcc cccgccgcc 169

<210> 122
<211> 79
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<220>
<221> modified_base
<222> (20)..(20)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (22)..(22)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (40)..(42)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (53)..(56)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (62)..(62)
<223> a, c, t, g, unknown or other

<220>
<221> modified_base
<222> (64)..(64)
<223> a, c, t, g, unknown or other

<400> 122
gccgccgcgg cggcggsggn gnsgcgcgct dcgcgcgcsn nncrccrggg ggnnnncwgc 60

sncnccccccc cccgcgcat 79

<210> 123
<211> 31

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<220>
<221> modified_base
<222> (18)..(19)
<223> a, c, t, g, unknown or other

<400> 123
gcgcgggkcc ccccccnnc ggggggctcc g 31

<210> 124
<211> 59
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 124
cccccggcc cccccccgtg ctaaacccac cgcgcatgcg cgaccacgcc cccgccgcc 59

<210> 125
<211> 156
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 125
gcggcggggg ggcggccgcg ttcgcgcgcc gcccaccagg gggtgctgcg cgccccccc 60

cgcgcatgcg cggggccccc ccccgggggg gctccgcccc cccggccccc ccccgtgcta 120

aacccaccgc gcatgcgcga ccacgccccc gccgcc 156

<210> 126
<211> 7
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 126
gcggcgg 7

<210> 127
<211> 7
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 127
gggggcg 7

<210> 128
<211> 6
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 128
gccgcg 6

<210> 129
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 129
ttcgcgcgcc gcccaccagg gggtg 25

<210> 130

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

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 130
ctgcg 5

<210> 131
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 131
cgccccccc cgcgcat 17

<210> 132
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 132
gcgcggggcc ccccccc 17

<210> 133
<211> 72
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 133
gggggggctc cgcccccccg gccccccccc gtgctaaacc caccgcgcat gcgcgaccac 60

gcccccgccg cc 72

<210> 134
<211> 115
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 134
cggcggcggc ggcgcgcgcg ctgcgcgcgc gcgccggggg ggcgccagcg ccccccccc 60

cgcgcatgca cgggtccccc cccccacggg gggctccgcc ccccggcccc ccccc 115

<210> 135
<211> 14
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 135
cggcggcggc ggcg 14

<210> 136
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 136
cgcgcgctgc gcgcgcg 17

<210> 137
<211> 19
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 137
cgccgggggg gcgccagcg 19

<210> 138
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 138
ccccccccc cgcgcat 17

<210> 139
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 139
gcacgggtcc cccccccac gggggctcc g 31

<210> 140
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 140
ccccccggcc ccccccc 17

<210> 141
<211> 121
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 141
ccgtcggcgg gggggccgcg cgctgcgcgc gcggcccccg ggggaggcac agcctccccc 60

ccccgcgcgc atgcgcgcgg gtccccccccc ctccgggggg ctccgccccc cggccccccc 120

c 121

<210> 142
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 142
ccgtcggcgg gggggccgcg cgctgcgcgc gcggccc 37

<210> 143
<211> 84
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 143
ccgggggagg cacagcctcc ccccccccgcg cgcatgcgcg cgggtccccc cccctccggg 60

gggctccgcc ccccggcccc cccc 84

<210> 144
<211> 104
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 144
cggcggcggc gcgcgcgcta cgcgcgcgcg ccgggggggct gccgcccccc ccccgcgcat 60

gcgcggggcc cccccccgcg gggggctccg ccccccggcc cccc 104

<210> 145
<211> 11
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 145
cggcggcggc g 11

<210> 146
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 146
cgcgcgctac gcgcgcg 17

<210> 147
<211> 10
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 147
cgccgggggg 10

<210> 148
<211> 7
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 148
ctgccgc 7

<210> 149
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 149
ccccccccg cgcat 15

<210> 150
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 150
gcgcggggcc ccccccc 17

<210> 151
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 151
gcgggggct ccg 13

<210> 152
<211> 14
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 152
ccccccggcc cccc 14

<210> 153
<211> 122
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 153
gccgccgcgg cggcgggggg cggcgcgctg cgcgcgccgc ccagtagggg gagccatgcg 60

ccccccccg cgcatgcgcg gggcccccc ccgcgggggg ctccgccccc cggcccccc 120

cg 122

<210> 154
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 154
gccgccgcgg cggcggggg 19

<210> 155
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 155

gcggcgcgct gcgcgcgccg cccagtaggg ggagccatgc g 41

<210> 156
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 156
ccccccccg cgcat 15

<210> 157
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 157
gcgcggggcc ccccccc 17

<210> 158
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 158
gcgggggct ccg 13

<210> 159
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 159
ccccccggcc ccccccg 17

<210> 160
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 160
cgcgctgcgc gcgccgccca gtaggggggag ccatgc 36

<210> 161
<211> 78
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 161
ccgccatctt aagtagttga ggcggacggt ggcgtgagtt caaaggtcac catcagccac 60

acctactcaa aatggtgg 78

<210> 162
<211> 172
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 162
cttaagtagt tgaggcggac ggtggcgtga gttcaaaggt caccatcagc cacacctact 60

caaaatggtg gacaatttct tccgggtcaa aggttacagc cgccatgtta aaacacgtga 120

cgtatgacgt cacggccgcc attttgtgac acaagatggc cgacttcctt cc 172

<210> 163
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 163
cgcgctgcgc gcgccgccca gtagggggag ccatgc 36

<210> 164
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 164
gcgctdcgcg cgcgcgccgg ggggctgcgc cccccc 36

<210> 165
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 165
gcgcttcgcg cgccgcccac taggggcgt tgcgcg 36

<210> 166
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 166
gcgctgcgcg cgccgcccag taggggcgc aatgcg 36

<210> 167
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 167
gcgctgcgcg cgcggccccc gggggaggca ttgcct 36

<210> 168
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 168
gcgctgcgcg cgcgcgccgg gggggcgcca gcgccc 36

<210> 169
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 169
gcgcttcgcg cgcgcgccgg ggggctccgc cccccc 36

<210> 170
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 170
gcgcttcgcg cgcgcgccgg ggggctgcgc cccccc 36

<210> 171
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 171
gcgctacgcg cgcgcgccgg ggggctgcgc cccccc 36

<210> 172
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 172
gcgctacgcg cgcgcgccgg ggggctctgc cccccc 36

<210> 173

<400> 173
000

<210> 174

<400> 174
000

<210> 175

<400> 175
000

<210> 176

<400> 176
000

<210> 177

<400> 177
000

<210> 178

<400> 178
000

<210> 179

<400> 179
000

<210> 180

<400> 180
000

<210> 181

<400> 181
000

<210> 182

<400> 182
000

<210> 183

<400> 183
000

<210> 184

<400> 184
000

<210> 185
<211> 743
<212> PRT
<213> Alphatorquevirus sp.

<400> 185
Met Ala Trp Gly Trp Trp Lys Arg Arg Arg Arg Trp Trp Phe Arg Lys
1               5                  10                  15

Arg Trp Thr Arg Gly Arg Leu Arg Arg Arg Trp Pro Arg Ser Ala Arg
            20                  25                  30

Arg Arg Pro Arg Arg Arg Arg Val Arg Arg Arg Arg Arg Trp Arg Arg
        35                  40                  45

Gly Arg Arg Lys Thr Arg Thr Tyr Arg Arg Arg Arg Arg Phe Arg Arg
    50                  55                  60

Arg Gly Arg Lys Ala Lys Leu Ile Ile Lys Leu Trp Gln Pro Ala Val
65                  70                  75                  80

Ile Lys Arg Cys Arg Ile Lys Gly Tyr Ile Pro Leu Ile Ile Ser Gly
                85                  90                  95

Asn Gly Thr Phe Ala Thr Asn Phe Thr Ser His Ile Asn Asp Arg Ile
            100                 105                 110

Met Lys Gly Pro Phe Gly Gly Gly His Ser Thr Met Arg Phe Ser Leu
        115                 120                 125

Tyr Ile Leu Phe Glu Glu His Leu Arg His Met Asn Phe Trp Thr Arg
    130                 135                 140

Ser Asn Asp Asn Leu Glu Leu Thr Arg Tyr Leu Gly Ala Ser Val Lys
145                 150                 155                 160

Ile Tyr Arg His Pro Asp Gln Asp Phe Ile Val Ile Tyr Asn Arg Arg
                165                 170                 175

```
Thr Pro Leu Gly Gly Asn Ile Tyr Thr Ala Pro Ser Leu His Pro Gly
            180                 185                 190

Asn Ala Ile Leu Ala Lys His Lys Ile Leu Val Pro Ser Leu Gln Thr
        195                 200                 205

Arg Pro Lys Gly Arg Lys Ala Ile Arg Leu Arg Ile Ala Pro Pro Thr
    210                 215                 220

Leu Phe Thr Asp Lys Trp Tyr Phe Gln Lys Asp Ile Ala Asp Leu Thr
225                 230                 235                 240

Leu Phe Asn Ile Met Ala Val Glu Ala Asp Leu Arg Phe Pro Phe Cys
                245                 250                 255

Ser Pro Gln Thr Asp Asn Thr Cys Ile Ser Phe Gln Val Leu Ser Ser
            260                 265                 270

Val Tyr Asn Asn Tyr Leu Ser Ile Asn Thr Phe Asn Asn Asp Asn Ser
        275                 280                 285

Asp Ser Lys Leu Lys Glu Phe Leu Asn Lys Ala Phe Pro Thr Thr Gly
    290                 295                 300

Thr Lys Gly Thr Ser Leu Asn Ala Leu Asn Thr Phe Arg Thr Glu Gly
305                 310                 315                 320

Cys Ile Ser His Pro Gln Leu Lys Lys Pro Asn Pro Gln Ile Asn Lys
                325                 330                 335

Pro Leu Glu Ser Gln Tyr Phe Ala Pro Leu Asp Ala Leu Trp Gly Asp
            340                 345                 350

Pro Ile Tyr Tyr Asn Asp Leu Asn Glu Asn Lys Ser Leu Asn Asp Ile
        355                 360                 365
```

```
Ile Glu Lys Ile Leu Ile Lys Asn Met Ile Thr Tyr His Ala Lys Leu
    370                 375                 380

Arg Glu Phe Pro Asn Ser Tyr Gln Gly Asn Lys Ala Phe Cys His Leu
385                 390                 395                 400

Thr Gly Ile Tyr Ser Pro Pro Tyr Leu Asn Gln Gly Arg Ile Ser Pro
                405                 410                 415

Glu Ile Phe Gly Leu Tyr Thr Glu Ile Ile Tyr Asn Pro Tyr Thr Asp
            420                 425                 430

Lys Gly Thr Gly Asn Lys Val Trp Met Asp Pro Leu Thr Lys Glu Asn
        435                 440                 445

Asn Ile Tyr Lys Glu Gly Gln Ser Lys Cys Leu Leu Thr Asp Met Pro
    450                 455                 460

Leu Trp Thr Leu Leu Phe Gly Tyr Thr Asp Trp Cys Lys Lys Asp Thr
465                 470                 475                 480

Asn Asn Trp Asp Leu Pro Leu Asn Tyr Arg Leu Val Leu Ile Cys Pro
                485                 490                 495

Tyr Thr Phe Pro Lys Leu Tyr Asn Glu Lys Val Lys Asp Tyr Gly Tyr
            500                 505                 510

Ile Pro Tyr Ser Tyr Lys Phe Gly Ala Gly Gln Met Pro Asp Gly Ser
        515                 520                 525

Asn Tyr Ile Pro Phe Gln Phe Arg Ala Lys Trp Tyr Pro Thr Val Leu
    530                 535                 540

His Gln Gln Gln Val Met Glu Asp Ile Ser Arg Ser Gly Pro Phe Ala
545                 550                 555                 560

Pro Lys Val Glu Lys Pro Ser Thr Gln Leu Val Met Lys Tyr Cys Phe
                565                 570                 575
```

```
Asn Phe Asn Trp Gly Gly Asn Pro Ile Ile Glu Gln Ile Val Lys Asp
            580                 585                 590

Pro Ser Phe Gln Pro Thr Tyr Glu Ile Pro Gly Thr Gly Asn Ile Pro
        595                 600                 605

Arg Arg Ile Gln Val Ile Asp Pro Arg Val Leu Gly Pro His Tyr Ser
    610                 615                 620

Phe Arg Ser Trp Asp Met Arg Arg His Thr Phe Ser Arg Ala Ser Ile
625                 630                 635                 640

Lys Arg Val Ser Glu Gln Gln Glu Thr Ser Asp Leu Val Phe Ser Gly
                645                 650                 655

Pro Lys Lys Pro Arg Val Asp Ile Pro Lys Gln Glu Thr Gln Glu Glu
            660                 665                 670

Ser Ser His Ser Leu Gln Arg Glu Ser Arg Pro Trp Glu Thr Glu Glu
        675                 680                 685

Glu Ser Glu Thr Glu Ala Leu Ser Gln Glu Ser Gln Glu Val Pro Phe
    690                 695                 700

Gln Gln Gln Leu Gln Gln Gln Tyr Gln Glu Gln Leu Lys Leu Arg Gln
705                 710                 715                 720

Gly Ile Lys Val Leu Phe Glu Gln Leu Ile Arg Thr Gln Gln Gly Val
                725                 730                 735

His Val Asn Pro Cys Leu Arg
            740
```

<210> 186
<211> 68
<212> PRT
<213> Alphatorquevirus sp.

<400> 186
Met Ala Trp Gly Trp Trp Lys Arg Arg Arg Arg Trp Trp Phe Arg Lys
1               5                  10                  15

Arg Trp Thr Arg Gly Arg Leu Arg Arg Arg Trp Pro Arg Ser Ala Arg
            20                  25                  30

Arg Arg Pro Arg Arg Arg Arg Val Arg Arg Arg Arg Arg Trp Arg Arg
        35                  40                  45

Gly Arg Arg Lys Thr Arg Thr Tyr Arg Arg Arg Arg Arg Phe Arg Arg
    50                  55                  60

Arg Gly Arg Lys
65

<210> 187
<211> 212
<212> PRT
<213> Alphatorquevirus sp.

<400> 187
Ala Lys Leu Ile Ile Lys Leu Trp Gln Pro Ala Val Ile Lys Arg Cys
1               5                  10                  15

Arg Ile Lys Gly Tyr Ile Pro Leu Ile Ile Ser Gly Asn Gly Thr Phe
            20                  25                  30

Ala Thr Asn Phe Thr Ser His Ile Asn Asp Arg Ile Met Lys Gly Pro
        35                  40                  45

Phe Gly Gly Gly His Ser Thr Met Arg Phe Ser Leu Tyr Ile Leu Phe
    50                  55                  60

Glu Glu His Leu Arg His Met Asn Phe Trp Thr Arg Ser Asn Asp Asn
65                  70                  75                  80

Leu Glu Leu Thr Arg Tyr Leu Gly Ala Ser Val Lys Ile Tyr Arg His
                85                  90                  95

```
Pro Asp Gln Asp Phe Ile Val Ile Tyr Asn Arg Arg Thr Pro Leu Gly
            100                 105                 110


Gly Asn Ile Tyr Thr Ala Pro Ser Leu His Pro Gly Asn Ala Ile Leu
        115                 120                 125


Ala Lys His Lys Ile Leu Val Pro Ser Leu Gln Thr Arg Pro Lys Gly
    130                 135                 140


Arg Lys Ala Ile Arg Leu Arg Ile Ala Pro Pro Thr Leu Phe Thr Asp
145                 150                 155                 160


Lys Trp Tyr Phe Gln Lys Asp Ile Ala Asp Leu Thr Leu Phe Asn Ile
                165                 170                 175


Met Ala Val Glu Ala Asp Leu Arg Phe Pro Phe Cys Ser Pro Gln Thr
            180                 185                 190


Asp Asn Thr Cys Ile Ser Phe Gln Val Leu Ser Ser Val Tyr Asn Asn
        195                 200                 205


Tyr Leu Ser Ile
    210


<210> 188
<211> 133
<212> PRT
<213> Alphatorquevirus sp.

<400> 188
Asn Thr Phe Asn Asn Asp Asn Ser Asp Ser Lys Leu Lys Glu Phe Leu
1               5                   10                  15


Asn Lys Ala Phe Pro Thr Thr Gly Thr Lys Gly Thr Ser Leu Asn Ala
            20                  25                  30


Leu Asn Thr Phe Arg Thr Glu Gly Cys Ile Ser His Pro Gln Leu Lys
        35                  40                  45
```

```
Lys Pro Asn Pro Gln Ile Asn Lys Pro Leu Glu Ser Gln Tyr Phe Ala
    50                  55                  60

Pro Leu Asp Ala Leu Trp Gly Asp Pro Ile Tyr Tyr Asn Asp Leu Asn
65                  70                  75                  80

Glu Asn Lys Ser Leu Asn Asp Ile Ile Glu Lys Ile Leu Ile Lys Asn
                85                  90                  95

Met Ile Thr Tyr His Ala Lys Leu Arg Glu Phe Pro Asn Ser Tyr Gln
            100                 105                 110

Gly Asn Lys Ala Phe Cys His Leu Thr Gly Ile Tyr Ser Pro Pro Tyr
        115                 120                 125

Leu Asn Gln Gly Arg
    130
```

<210> 189
<211> 166
<212> PRT
<213> Alphatorquevirus sp.

<400> 189

```
Ile Ser Pro Glu Ile Phe Gly Leu Tyr Thr Glu Ile Ile Tyr Asn Pro
1               5                   10                  15

Tyr Thr Asp Lys Gly Thr Gly Asn Lys Val Trp Met Asp Pro Leu Thr
            20                  25                  30

Lys Glu Asn Asn Ile Tyr Lys Glu Gly Gln Ser Lys Cys Leu Leu Thr
        35                  40                  45

Asp Met Pro Leu Trp Thr Leu Leu Phe Gly Tyr Thr Asp Trp Cys Lys
    50                  55                  60

Lys Asp Thr Asn Asn Trp Asp Leu Pro Leu Asn Tyr Arg Leu Val Leu
65                  70                  75                  80
```

```
Ile Cys Pro Tyr Thr Phe Pro Lys Leu Tyr Asn Glu Lys Val Lys Asp
                85                  90                  95

Tyr Gly Tyr Ile Pro Tyr Ser Tyr Lys Phe Gly Ala Gly Gln Met Pro
            100                 105                 110

Asp Gly Ser Asn Tyr Ile Pro Phe Gln Phe Arg Ala Lys Trp Tyr Pro
        115                 120                 125

Thr Val Leu His Gln Gln Gln Val Met Glu Asp Ile Ser Arg Ser Gly
    130                 135                 140

Pro Phe Ala Pro Lys Val Glu Lys Pro Ser Thr Gln Leu Val Met Lys
145                 150                 155                 160

Tyr Cys Phe Asn Phe Asn
                165
```

<210> 190
<211> 164
<212> PRT
<213> Alphatorquevirus sp.

<400> 190

```
Trp Gly Gly Asn Pro Ile Ile Glu Gln Ile Val Lys Asp Pro Ser Phe
1               5                   10                  15

Gln Pro Thr Tyr Glu Ile Pro Gly Thr Gly Asn Ile Pro Arg Arg Ile
            20                  25                  30

Gln Val Ile Asp Pro Arg Val Leu Gly Pro His Tyr Ser Phe Arg Ser
        35                  40                  45

Trp Asp Met Arg Arg His Thr Phe Ser Arg Ala Ser Ile Lys Arg Val
    50                  55                  60

Ser Glu Gln Gln Glu Thr Ser Asp Leu Val Phe Ser Gly Pro Lys Lys
65                  70                  75                  80
```

```
Pro Arg Val Asp Ile Pro Lys Gln Glu Thr Gln Glu Glu Ser Ser His
                85                  90                  95

Ser Leu Gln Arg Glu Ser Arg Pro Trp Glu Thr Glu Glu Glu Ser Glu
            100                 105                 110

Thr Glu Ala Leu Ser Gln Glu Ser Gln Glu Val Pro Phe Gln Gln Gln
        115                 120                 125

Leu Gln Gln Gln Tyr Gln Glu Gln Leu Lys Leu Arg Gln Gly Ile Lys
    130                 135                 140

Val Leu Phe Glu Gln Leu Ile Arg Thr Gln Gln Gly Val His Val Asn
145                 150                 155                 160

Pro Cys Leu Arg
```

<210> 191

<400> 191
000

<210> 192

<400> 192
000

<210> 193

<400> 193
000

<210> 194

<400> 194
000

<210> 195

<400> 195
000


<210> 196

<400> 196
000


<210> 197

<400> 197
000


<210> 198

<400> 198
000


<210> 199

<400> 199
000


<210> 200

<400> 200
000


<210> 201

<400> 201
000


<210> 202

<400> 202
000


<210> 203

<400> 203
000

<210> 204

<400> 204
000

<210> 205

<400> 205
000

<210> 206

<400> 206
000

<210> 207

<400> 207
000

<210> 208

<400> 208
000

<210> 209

<400> 209
000

<210> 210

<400> 210
000

<210> 211

<400> 211
000

<210> 212

<400> 212
000

<210> 213

<400> 213
000

<210> 214

<400> 214
000

<210> 215
<211> 666
<212> PRT
<213> Betatorquevirus sp.

<400> 215
Met Pro Tyr Tyr Tyr Arg Arg Arg Arg Tyr Asn Tyr Arg Arg Pro Arg
1               5                   10                  15

Trp Tyr Gly Arg Gly Trp Ile Arg Arg Pro Phe Arg Arg Arg Phe Arg
            20                  25                  30

Arg Lys Arg Arg Val Arg Pro Thr Tyr Thr Thr Ile Pro Leu Lys Gln
        35                  40                  45

Trp Gln Pro Pro Tyr Lys Arg Thr Cys Tyr Ile Lys Gly Gln Asp Cys
    50                  55                  60

Leu Ile Tyr Tyr Ser Asn Leu Arg Leu Gly Met Asn Ser Thr Met Tyr
65                  70                  75                  80

Glu Lys Ser Ile Val Pro Val His Trp Pro Gly Gly Gly Ser Phe Ser
                85                  90                  95

```
Val Ser Met Leu Thr Leu Asp Ala Leu Tyr Asp Ile His Lys Leu Cys
            100                 105                 110

Arg Asn Trp Trp Thr Ser Thr Asn Gln Asp Leu Pro Leu Val Arg Tyr
        115                 120                 125

Lys Gly Cys Lys Ile Thr Phe Tyr Gln Ser Thr Phe Thr Asp Tyr Ile
    130                 135                 140

Val Arg Ile His Thr Glu Leu Pro Ala Asn Ser Asn Lys Leu Thr Tyr
145                 150                 155                 160

Pro Asn Thr His Pro Leu Met Met Met Met Ser Lys Tyr Lys His Ile
                165                 170                 175

Ile Pro Ser Arg Gln Thr Arg Arg Lys Lys Lys Pro Tyr Thr Lys Ile
            180                 185                 190

Phe Val Lys Pro Pro Pro Gln Phe Glu Asn Lys Trp Tyr Phe Ala Thr
        195                 200                 205

Asp Leu Tyr Lys Ile Pro Leu Leu Gln Ile His Cys Thr Ala Cys Asn
    210                 215                 220

Leu Gln Asn Pro Phe Val Lys Pro Asp Lys Leu Ser Asn Asn Val Thr
225                 230                 235                 240

Leu Trp Ser Leu Asn Thr Ile Ser Ile Gln Asn Arg Asn Met Ser Val
                245                 250                 255

Asp Gln Gly Gln Ser Trp Pro Phe Lys Ile Leu Gly Thr Gln Ser Phe
            260                 265                 270

Tyr Phe Tyr Phe Tyr Thr Gly Ala Asn Leu Pro Gly Asp Thr Thr Gln
        275                 280                 285

Ile Pro Val Ala Asp Leu Leu Pro Leu Thr Asn Pro Arg Ile Asn Arg
    290                 295                 300
```

```
Pro Gly Gln Ser Leu Asn Glu Ala Lys Ile Thr Asp His Ile Thr Phe
305                 310                 315                 320

Thr Glu Tyr Lys Asn Lys Phe Thr Asn Tyr Trp Gly Asn Pro Phe Asn
                325                 330                 335

Lys His Ile Gln Glu His Leu Asp Met Ile Leu Tyr Ser Leu Lys Ser
            340                 345                 350

Pro Glu Ala Ile Lys Asn Glu Trp Thr Thr Glu Asn Met Lys Trp Asn
        355                 360                 365

Gln Leu Asn Asn Ala Gly Thr Met Ala Leu Thr Pro Phe Asn Glu Pro
    370                 375                 380

Ile Phe Thr Gln Ile Gln Tyr Asn Pro Asp Arg Asp Thr Gly Glu Asp
385                 390                 395                 400

Thr Gln Leu Tyr Leu Leu Ser Asn Ala Thr Gly Thr Gly Trp Asp Pro
                405                 410                 415

Pro Gly Ile Pro Glu Leu Ile Leu Glu Gly Phe Pro Leu Trp Leu Ile
            420                 425                 430

Tyr Trp Gly Phe Ala Asp Phe Gln Lys Asn Leu Lys Lys Val Thr Asn
        435                 440                 445

Ile Asp Thr Asn Tyr Met Leu Val Ala Lys Thr Lys Phe Thr Gln Lys
    450                 455                 460

Pro Gly Thr Phe Tyr Leu Val Ile Leu Asn Asp Thr Phe Val Glu Gly
465                 470                 475                 480

Asn Ser Pro Tyr Glu Lys Gln Pro Leu Pro Glu Asp Asn Ile Lys Trp
                485                 490                 495
```

```
Tyr Pro Gln Val Gln Tyr Gln Leu Glu Ala Gln Asn Lys Leu Leu Gln
            500                 505                 510

Thr Gly Pro Phe Thr Pro Asn Ile Gln Gly Gln Leu Ser Asp Asn Ile
        515                 520                 525

Ser Met Phe Tyr Lys Phe Tyr Phe Lys Trp Gly Gly Ser Pro Pro Lys
    530                 535                 540

Ala Ile Asn Val Glu Asn Pro Ala His Gln Ile Gln Tyr Pro Ile Pro
545                 550                 555                 560

Arg Asn Glu His Glu Thr Thr Ser Leu Gln Ser Pro Gly Glu Ala Pro
                565                 570                 575

Glu Ser Ile Leu Tyr Ser Phe Asp Tyr Arg His Gly Asn Tyr Thr Thr
            580                 585                 590

Thr Ala Leu Ser Arg Ile Ser Gln Asp Trp Ala Leu Lys Asp Thr Val
        595                 600                 605

Ser Lys Ile Thr Glu Pro Asp Arg Gln Gln Leu Leu Lys Gln Ala Leu
    610                 615                 620

Glu Cys Leu Gln Ile Ser Glu Glu Thr Gln Glu Lys Lys Glu Lys Glu
625                 630                 635                 640

Val Gln Gln Leu Ile Ser Asn Leu Arg Gln Gln Gln Gln Leu Tyr Arg
                645                 650                 655

Glu Arg Ile Ile Ser Leu Leu Lys Asp Gln
            660                 665
```

<210> 216
<211> 38
<212> PRT
<213> Betatorquevirus sp.

<400> 216

```
Met Pro Tyr Tyr Tyr Arg Arg Arg Arg Tyr Asn Tyr Arg Arg Pro Arg
1               5                   10                  15


Trp Tyr Gly Arg Gly Trp Ile Arg Arg Pro Phe Arg Arg Arg Phe Arg
            20                  25                  30


Arg Lys Arg Arg Val Arg
        35
```

<210> 217
<211> 208
<212> PRT
<213> Betatorquevirus sp.

<400> 217

```
Pro Thr Tyr Thr Thr Ile Pro Leu Lys Gln Trp Gln Pro Pro Tyr Lys
1               5                   10                  15


Arg Thr Cys Tyr Ile Lys Gly Gln Asp Cys Leu Ile Tyr Tyr Ser Asn
            20                  25                  30


Leu Arg Leu Gly Met Asn Ser Thr Met Tyr Glu Lys Ser Ile Val Pro
        35                  40                  45


Val His Trp Pro Gly Gly Gly Ser Phe Ser Val Ser Met Leu Thr Leu
    50                  55                  60


Asp Ala Leu Tyr Asp Ile His Lys Leu Cys Arg Asn Trp Trp Thr Ser
65                  70                  75                  80


Thr Asn Gln Asp Leu Pro Leu Val Arg Tyr Lys Gly Cys Lys Ile Thr
                85                  90                  95


Phe Tyr Gln Ser Thr Phe Thr Asp Tyr Ile Val Arg Ile His Thr Glu
            100                 105                 110


Leu Pro Ala Asn Ser Asn Lys Leu Thr Tyr Pro Asn Thr His Pro Leu
        115                 120                 125
```

```
Met Met Met Met Ser Lys Tyr Lys His Ile Ile Pro Ser Arg Gln Thr
    130                 135                 140

Arg Arg Lys Lys Lys Pro Tyr Thr Lys Ile Phe Val Lys Pro Pro Pro
145                 150                 155                 160

Gln Phe Glu Asn Lys Trp Tyr Phe Ala Thr Asp Leu Tyr Lys Ile Pro
                165                 170                 175

Leu Leu Gln Ile His Cys Thr Ala Cys Asn Leu Gln Asn Pro Phe Val
            180                 185                 190

Lys Pro Asp Lys Leu Ser Asn Asn Val Thr Leu Trp Ser Leu Asn Thr
        195                 200                 205
```

<210> 218
<211> 128
<212> PRT
<213> Betatorquevirus sp.

<400> 218

```
Ile Ser Ile Gln Asn Arg Asn Met Ser Val Asp Gln Gly Gln Ser Trp
1               5                   10                  15

Pro Phe Lys Ile Leu Gly Thr Gln Ser Phe Tyr Phe Tyr Phe Tyr Thr
            20                  25                  30

Gly Ala Asn Leu Pro Gly Asp Thr Thr Gln Ile Pro Val Ala Asp Leu
        35                  40                  45

Leu Pro Leu Thr Asn Pro Arg Ile Asn Arg Pro Gly Gln Ser Leu Asn
    50                  55                  60

Glu Ala Lys Ile Thr Asp His Ile Thr Phe Thr Glu Tyr Lys Asn Lys
65                  70                  75                  80

Phe Thr Asn Tyr Trp Gly Asn Pro Phe Asn Lys His Ile Gln Glu His
                85                  90                  95
```

```
Leu Asp Met Ile Leu Tyr Ser Leu Lys Ser Pro Glu Ala Ile Lys Asn
            100                 105                 110

Glu Trp Thr Thr Glu Asn Met Lys Trp Asn Gln Leu Asn Asn Ala Gly
        115                 120                 125
```

<210> 219
<211> 163
<212> PRT
<213> Betatorquevirus sp.

<400> 219

```
Thr Met Ala Leu Thr Pro Phe Asn Glu Pro Ile Phe Thr Gln Ile Gln
1               5                   10                  15

Tyr Asn Pro Asp Arg Asp Thr Gly Glu Asp Thr Gln Leu Tyr Leu Leu
            20                  25                  30

Ser Asn Ala Thr Gly Thr Gly Trp Asp Pro Pro Gly Ile Pro Glu Leu
        35                  40                  45

Ile Leu Glu Gly Phe Pro Leu Trp Leu Ile Tyr Trp Gly Phe Ala Asp
    50                  55                  60

Phe Gln Lys Asn Leu Lys Lys Val Thr Asn Ile Asp Thr Asn Tyr Met
65                  70                  75                  80

Leu Val Ala Lys Thr Lys Phe Thr Gln Lys Pro Gly Thr Phe Tyr Leu
                85                  90                  95

Val Ile Leu Asn Asp Thr Phe Val Glu Gly Asn Ser Pro Tyr Glu Lys
            100                 105                 110

Gln Pro Leu Pro Glu Asp Asn Ile Lys Trp Tyr Pro Gln Val Gln Tyr
        115                 120                 125

Gln Leu Glu Ala Gln Asn Lys Leu Leu Gln Thr Gly Pro Phe Thr Pro
    130                 135                 140
```

```
Asn Ile Gln Gly Gln Leu Ser Asp Asn Ile Ser Met Phe Tyr Lys Phe
145                 150                 155                 160


Tyr Phe Lys
```

<210> 220
<211> 129
<212> PRT
<213> Betatorquevirus sp.

<400> 220

```
Trp Gly Gly Ser Pro Pro Lys Ala Ile Asn Val Glu Asn Pro Ala His
1               5                   10                  15


Gln Ile Gln Tyr Pro Ile Pro Arg Asn Glu His Glu Thr Thr Ser Leu
            20                  25                  30


Gln Ser Pro Gly Glu Ala Pro Glu Ser Ile Leu Tyr Ser Phe Asp Tyr
        35                  40                  45


Arg His Gly Asn Tyr Thr Thr Thr Ala Leu Ser Arg Ile Ser Gln Asp
    50                  55                  60


Trp Ala Leu Lys Asp Thr Val Ser Lys Ile Thr Glu Pro Asp Arg Gln
65                  70                  75                  80


Gln Leu Leu Lys Gln Ala Leu Glu Cys Leu Gln Ile Ser Glu Glu Thr
                85                  90                  95


Gln Glu Lys Lys Glu Lys Glu Val Gln Gln Leu Ile Ser Asn Leu Arg
            100                 105                 110


Gln Gln Gln Gln Leu Tyr Arg Glu Arg Ile Ile Ser Leu Leu Lys Asp
        115                 120                 125


Gln
```

<210> 221

<400> 221
000


<210> 222

<400> 222
000


<210> 223

<400> 223
000


<210> 224

<400> 224
000


<210> 225

<400> 225
000


<210> 226

<400> 226
000


<210> 227
<211> 220
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polypeptide


<220>
<221> MOD_RES
<222> (29)..(31)
<223> Any amino acid

<220>
<221> SITE
<222> (29)..(31)
<223> This region may encompass 0-3 residues

<220>
<221> MOD_RES
<222> (100)..(100)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (125)..(129)
<223> Any amino acid

<220>
<221> SITE
<222> (125)..(129)
<223> This region may encompass 1-5 residues

<220>
<221> MOD_RES
<222> (181)..(181)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (211)..(211)
<223> Any amino acid

<400> 227
Leu Val Leu Thr Gln Trp Gln Pro Asn Thr Val Arg Arg Cys Tyr Ile
1               5                   10                  15

Arg Gly Tyr Leu Pro Leu Ile Ile Cys Gly Glu Asn Xaa Xaa Xaa Thr
            20                  25                  30

Thr Ser Arg Asn Tyr Ala Thr His Ser Asp Asp Thr Ile Gln Lys Gly
        35                  40                  45

Pro Phe Gly Gly Gly Met Ser Thr Thr Thr Phe Ser Leu Arg Val Leu
    50                  55                  60

Tyr Asp Glu Tyr Gln Arg Phe Met Asn Arg Trp Thr Tyr Ser Asn Glu
65                  70                  75                  80

```
Asp Leu Asp Leu Ala Arg Tyr Leu Gly Cys Lys Phe Thr Phe Tyr Arg
                85                  90                  95

His Pro Asp Xaa Asp Phe Ile Val Gln Tyr Asn Thr Asn Pro Pro Phe
            100                 105                 110

Lys Asp Thr Lys Leu Thr Ala Pro Ser Ile His Pro Xaa Xaa Xaa Xaa
        115                 120                 125

Xaa Gly Met Leu Met Leu Ser Lys Arg Lys Ile Leu Ile Pro Ser Leu
    130                 135                 140

Lys Thr Arg Pro Lys Gly Lys His Tyr Val Lys Val Arg Ile Gly Pro
145                 150                 155                 160

Pro Lys Leu Phe Glu Asp Lys Trp Tyr Thr Gln Ser Asp Leu Cys Asp
                165                 170                 175

Val Pro Leu Val Xaa Leu Tyr Ala Thr Ala Ala Asp Leu Gln His Pro
            180                 185                 190

Phe Gly Ser Pro Gln Thr Asp Asn Pro Cys Val Thr Phe Gln Val Leu
        195                 200                 205

Gly Ser Xaa Tyr Asn Lys His Leu Ser Ile Ser Pro
    210                 215                 220
```

<210> 228
<211> 172
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<220>
<221> MOD_RES

<222> (38)..(38)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (44)..(46)
<223> Any amino acid

<220>
<221> SITE
<222> (44)..(46)
<223> This region may encompass 0-3 residues

<220>
<221> MOD_RES
<222> (77)..(77)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (79)..(79)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (98)..(101)
<223> Any amino acid

<220>
<221> SITE
<222> (98)..(101)
<223> This region may encompass 0-4 residues

<400> 228
Ser Asn Phe Glu Phe Pro Gly Ala Tyr Thr Asp Ile Thr Tyr Asn Pro
1               5                   10                  15

Leu Thr Asp Lys Gly Val Gly Asn Met Val Trp Ile Gln Tyr Leu Thr
            20                  25                  30

Lys Pro Asp Thr Ile Xaa Asp Lys Thr Gln Ser Xaa Xaa Xaa Lys Cys
        35                  40                  45

Leu Ile Glu Asp Leu Pro Leu Trp Ala Ala Leu Tyr Gly Tyr Val Asp
    50                  55                  60

```
Phe Cys Glu Lys Glu Thr Gly Asp Ser Ala Ile Ile Xaa Asn Xaa Gly
65                  70                  75                  80


Arg Val Leu Ile Arg Cys Pro Tyr Thr Lys Pro Pro Leu Tyr Asp Lys
                85                  90                  95


Thr Xaa Xaa Xaa Xaa Asn Lys Gly Phe Val Pro Tyr Ser Thr Asn Phe
            100                 105                 110


Gly Asn Gly Lys Met Pro Gly Gly Ser Gly Tyr Val Pro Ile Tyr Trp
        115                 120                 125


Arg Ala Arg Trp Tyr Pro Thr Leu Phe His Gln Lys Glu Val Leu Glu
    130                 135                 140


Asp Ile Val Gln Ser Gly Pro Phe Ala Tyr Lys Asp Glu Lys Pro Ser
145                 150                 155                 160


Thr Gln Leu Val Met Lys Tyr Cys Phe Asn Phe Asn
                165                 170


<210> 229
<211> 258
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide


<220>
<221> MOD_RES
<222> (20)..(22)
<223> Any amino acid

<220>
<221> SITE
<222> (20)..(22)
<223> This region may encompass 0-3 residues

<220>
<221> MOD_RES
```

<222> (25)..(25)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (78)..(78)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (89)..(89)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (91)..(91)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (95)..(98)
<223> Any amino acid

<220>
<221> SITE
<222> (95)..(98)
<223> This region may encompass 1-4 residues

<220>
<221> MOD_RES
<222> (107)..(120)
<223> Any amino acid

<220>
<221> SITE
<222> (107)..(120)
<223> This region may encompass 2-14 residues

<220>
<221> MOD_RES
<222> (129)..(129)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (139)..(168)
<223> Any amino acid

<220>
<221> SITE

<222> (139)..(168)
<223> This region may encompass 0-30 residues

<220>
<221> MOD_RES
<222> (201)..(204)
<223> Any amino acid

<220>
<221> SITE
<222> (201)..(204)
<223> This region may encompass 0-4 residues

<220>
<221> MOD_RES
<222> (219)..(258)
<223> Any amino acid

<220>
<221> SITE
<222> (219)..(258)
<223> This region may encompass 0-40 residues

<400> 229
Trp Gly Gly Asn Pro Ile Ser Gln Gln Val Val Arg Asn Pro Cys Lys
1               5                   10                  15

Asp Ser Gly Xaa Xaa Xaa Ser Gly Xaa Gly Arg Gln Pro Arg Ser Val
            20                  25                  30

Gln Val Val Asp Pro Lys Tyr Met Gly Pro Glu Tyr Thr Phe His Ser
        35                  40                  45

Trp Asp Trp Arg Arg Gly Leu Phe Gly Glu Lys Ala Ile Lys Arg Met
    50                  55                  60

Ser Glu Gln Pro Thr Asp Asp Glu Ile Phe Thr Gly Gly Xaa Pro Lys
65                  70                  75                  80

Arg Pro Arg Arg Asp Pro Pro Thr Xaa Gln Xaa Pro Glu Glu Xaa Xaa
                85                  90                  95

Xaa Xaa Gln Lys Glu Ser Ser Ser Phe Arg Xaa Xaa Xaa Xaa Xaa Xaa
            100                 105                 110

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Pro Trp Glu Ser Ser Ser Gln Glu
        115                 120                 125

Xaa Glu Ser Glu Ser Gln Glu Glu Glu Glu Xaa Xaa Xaa Xaa Xaa Xaa
    130                 135                 140

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
145                 150                 155                 160

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Glu Gln Thr Val Gln Gln Gln Leu
                165                 170                 175

Arg Gln Gln Leu Arg Glu Gln Arg Arg Leu Arg Val Gln Leu Gln Leu
            180                 185                 190

Leu Phe Gln Gln Leu Leu Lys Thr Xaa Xaa Xaa Xaa Gln Ala Gly Leu
        195                 200                 205

His Ile Asn Pro Leu Leu Leu Ser Gln Ala Xaa Xaa Xaa Xaa Xaa Xaa
    210                 215                 220

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
225                 230                 235                 240

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
                245                 250                 255

Xaa Xaa

<210> 230
<211> 214
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<220>
<221> MOD_RES
<222> (136)..(136)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (138)..(141)
<223> Any amino acid

<220>
<221> SITE
<222> (138)..(141)
<223> This region may encompass 1-4 residues

<220>
<221> MOD_RES
<222> (179)..(179)
<223> Any amino acid

<400> 230
Leu Lys Gln Trp Gln Pro Ser Thr Ile Arg Lys Cys Lys Ile Lys Gly
1               5                   10                  15

Tyr Leu Pro Leu Phe Gln Cys Gly Lys Gly Arg Ile Ser Asn Asn Tyr
            20                  25                  30

Thr Gln Tyr Lys Glu Ser Ile Val Pro His His Glu Pro Gly Gly Gly
        35                  40                  45

Gly Trp Ser Ile Gln Gln Phe Thr Leu Gly Ala Leu Tyr Glu Glu His
    50                  55                  60

Leu Lys Leu Arg Asn Trp Trp Thr Lys Ser Asn Asp Gly Leu Pro Leu
65                  70                  75                  80

Val Arg Tyr Leu Gly Cys Thr Ile Lys Leu Tyr Arg Ser Glu Asp Thr
                85                  90                  95

Asp Tyr Ile Val Thr Tyr Gln Arg Cys Tyr Pro Met Thr Ala Thr Lys
            100                 105                 110

Leu Thr Tyr Leu Ser Thr Gln Pro Ser Arg Met Leu Met Asn Lys His
        115                 120                 125

Lys Ile Ile Val Pro Ser Lys Xaa Thr Xaa Xaa Xaa Xaa Asn Lys Lys
    130                 135                 140

Lys Lys Pro Tyr Lys Lys Ile Phe Ile Lys Pro Pro Ser Gln Met Gln
145                 150                 155                 160

Asn Lys Trp Tyr Phe Gln Gln Asp Ile Ala Asn Thr Pro Leu Leu Gln
                165                 170                 175

Leu Thr Xaa Thr Ala Cys Ser Leu Asp Arg Met Tyr Leu Ser Ser Asp
            180                 185                 190

Ser Ile Ser Asn Asn Ile Thr Phe Thr Ser Leu Asn Thr Asn Phe Phe
        195                 200                 205

Gln Asn Pro Asn Phe Gln
    210

<210> 231
<211> 187
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
      polypeptide

<220>
<221> MOD_RES
<222> (1)..(10)
<223> Any amino acid

<220>
<221> SITE
<222> (1)..(10)
<223> This region may encompass 4-10 residues

<220>

<221> MOD_RES
<222> (38)..(45)
<223> Any amino acid

<220>
<221> SITE
<222> (38)..(45)
<223> This region may encompass 1-8 residues

<220>
<221> MOD_RES
<222> (94)..(94)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (100)..(102)
<223> Any amino acid

<220>
<221> SITE
<222> (100)..(102)
<223> This region may encompass 1-3 residues

<220>
<221> MOD_RES
<222> (112)..(112)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (114)..(115)
<223> Any amino acid

<220>
<221> SITE
<222> (114)..(115)
<223> This region may encompass 0-2 residues

<220>
<221> MOD_RES
<222> (124)..(139)
<223> Any amino acid

<220>
<221> SITE
<222> (124)..(139)
<223> This region may encompass 3-16 residues

<220>

<221> MOD_RES
<222> (154)..(154)
<223> Any amino acid

<400> 231
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Thr Pro Leu Tyr Phe Glu
1               5                  10                  15

Cys Arg Tyr Asn Pro Phe Lys Asp Lys Gly Thr Gly Asn Lys Val Tyr
            20                  25                  30

Leu Val Ser Asn Asn Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Thr Gly Trp
        35                  40                  45

Asp Pro Pro Thr Asp Pro Asp Leu Ile Ile Glu Gly Phe Pro Leu Trp
    50                  55                  60

Leu Leu Leu Trp Gly Trp Leu Asp Trp Gln Lys Lys Leu Gly Lys Ile
65                  70                  75                  80

Gln Asn Ile Asp Thr Asp Tyr Ile Leu Val Ile Gln Ser Xaa Tyr Tyr
                85                  90                  95

Ile Pro Pro Xaa Xaa Xaa Lys Leu Pro Tyr Tyr Val Pro Leu Asp Xaa
            100                 105                 110

Asp Xaa Xaa Phe Leu His Gly Arg Ser Pro Tyr Xaa Xaa Xaa Xaa Xaa
        115                 120                 125

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Pro Ser Asp Lys Gln
    130                 135                 140

His Trp His Pro Lys Val Arg Phe Gln Xaa Glu Thr Ile Asn Asn Ile
145                 150                 155                 160

Ala Leu Thr Gly Pro Gly Thr Pro Lys Leu Pro Asn Gln Lys Ser Ile
                165                 170                 175

Gln Ala His Met Lys Tyr Lys Phe Tyr Phe Lys

180 185

<210> 232
<211> 163
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<220>
<221> MOD_RES
<222> (34)..(34)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (65)..(65)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (77)..(78)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (86)..(87)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (96)..(96)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (102)..(106)
<223> Any amino acid

<220>
<221> SITE
<222> (102)..(106)
<223> This region may encompass 0-5 residues

<220>
<221> MOD_RES

<222> (125)..(125)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (135)..(135)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (138)..(163)
<223> Any amino acid

<220>
<221> SITE
<222> (138)..(163)
<223> This region may encompass 0-26 residues

<400> 232
Trp Gly Gly Cys Pro Ala Pro Met Glu Thr Ile Thr Asp Pro Cys Lys
1               5                   10                  15

Gln Pro Lys Tyr Pro Ile Pro Asn Asn Leu Leu Gln Thr Thr Ser Leu
            20                  25                  30

Gln Xaa Pro Thr Thr Pro Ile Glu Thr Tyr Leu Tyr Lys Phe Asp Glu
        35                  40                  45

Arg Arg Gly Leu Leu Thr Lys Lys Ala Ala Lys Arg Ile Lys Lys Asp
    50                  55                  60

Xaa Thr Thr Glu Thr Thr Leu Phe Thr Asp Thr Gly Xaa Xaa Thr Ser
65                  70                  75                  80

Thr Thr Leu Pro Thr Xaa Xaa Gln Thr Glu Thr Thr Gln Glu Glu Xaa
                85                  90                  95

Thr Ser Glu Glu Glu Xaa Xaa Xaa Xaa Xaa Glu Thr Leu Leu Gln Gln
            100                 105                 110

Leu Gln Gln Leu Arg Arg Lys Gln Lys Gln Leu Arg Xaa Arg Ile Leu
        115                 120                 125

Gln Leu Leu Gln Leu Leu Xaa Leu Leu Xaa Xaa Xaa Xaa Xaa Xaa Xaa
130 135 140

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
145 150 155 160

Xaa Xaa Xaa

<210> 233
<211> 203
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<220>
<221> MOD_RES
<222> (79)..(79)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (104)..(104)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (116)..(116)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (120)..(121)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (125)..(125)
<223> Any amino acid

<220>
<221> MOD_RES

<222> (170)..(170)
<223> Any amino acid

<400> 233
Thr Ile Pro Leu Lys Gln Trp Gln Pro Glu Ser Ile Arg Lys Cys Lys
1               5                   10                  15

Ile Lys Gly Tyr Gly Thr Leu Val Leu Gly Ala Glu Gly Arg Gln Phe
            20                  25                  30

Tyr Cys Tyr Thr Asn Glu Lys Asp Glu Tyr Thr Pro Pro Lys Ala Pro
        35                  40                  45

Gly Gly Gly Gly Phe Gly Val Glu Leu Phe Ser Leu Glu Tyr Leu Tyr
    50                  55                  60

Glu Gln Trp Lys Ala Arg Asn Asn Ile Trp Thr Lys Ser Asn Xaa Tyr
65                  70                  75                  80

Lys Asp Leu Cys Arg Tyr Thr Gly Cys Lys Ile Thr Phe Tyr Arg His
                85                  90                  95

Pro Thr Thr Asp Phe Ile Val Xaa Tyr Ser Arg Gln Pro Pro Phe Glu
            100                 105                 110

Ile Asp Lys Xaa Thr Tyr Met Xaa Xaa His Pro Gln Xaa Leu Leu Leu
        115                 120                 125

Arg Lys His Lys Lys Ile Ile Leu Ser Lys Ala Thr Asn Pro Lys Gly
    130                 135                 140

Lys Leu Lys Lys Lys Ile Lys Ile Lys Pro Pro Lys Gln Met Leu Asn
145                 150                 155                 160

Lys Trp Phe Phe Gln Lys Gln Phe Ala Xaa Tyr Gly Leu Val Gln Leu
                165                 170                 175

Gln Ala Ala Ala Cys Asx Leu Arg Tyr Pro Arg Leu Gly Cys Cys Asn
            180                 185                 190

Glu Asn Arg Leu Ile Thr Leu Tyr Tyr Leu Asn
195 200

<210> 234
<211> 162
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic polypeptide

<220>
<221> MOD_RES
<222> (12)..(12)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (20)..(20)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (23)..(23)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (30)..(30)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (58)..(58)
<223> I or L

<220>
<221> MOD_RES
<222> (84)..(84)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (90)..(90)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (95)..(95)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (105)..(105)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (111)..(111)
<223> I or L

<220>
<221> MOD_RES
<222> (113)..(113)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (154)..(154)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (156)..(156)
<223> Any amino acid

<400> 234
Leu Pro Ile Val Val Ala Arg Tyr Asn Pro Ala Xaa Asp Thr Gly Lys
1               5                   10                  15

Gly Asn Lys Xaa Trp Leu Xaa Ser Thr Leu Asn Gly Ser Xaa Trp Ala
            20                  25                  30

Pro Pro Thr Thr Asp Lys Asp Leu Ile Ile Glu Gly Leu Pro Leu Trp
        35                  40                  45

Leu Ala Leu Tyr Gly Tyr Trp Ser Tyr Xaa Lys Lys Val Lys Lys Asp
    50                  55                  60

Lys Gly Ile Leu Gln Ser His Met Phe Val Val Lys Ser Pro Ala Ile
65                  70                  75                  80

```
Gln Pro Leu Xaa Thr Ala Thr Thr Gln Xaa Thr Phe Tyr Pro Xaa Ile
                85                  90                  95


Asp Asn Ser Phe Ile Gln Gly Lys Xaa Pro Tyr Asp Glu Pro Xaa Thr
            100                 105                 110


Xaa Asn Gln Lys Lys Leu Trp Tyr Pro Thr Leu Glu His Gln Gln Glu
        115                 120                 125


Thr Ile Asn Ala Ile Val Glu Ser Gly Pro Tyr Val Pro Lys Leu Asp
    130                 135                 140


Asn Gln Lys Asn Ser Thr Trp Glu Leu Xaa Tyr Xaa Tyr Thr Phe Tyr
145                 150                 155                 160


Phe Lys
```

<210> 235
<211> 177
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polypeptide

<220>
<221> MOD_RES
<222> (16)..(16)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (26)..(26)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (33)..(33)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (73)..(73)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (81)..(82)
<223> Any amino acid

<220>
<221> SITE
<222> (81)..(82)
<223> This region may encompass 0-2 residues

<220>
<221> MOD_RES
<222> (90)..(90)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (94)..(94)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (119)..(124)
<223> Any amino acid

<220>
<221> SITE
<222> (119)..(124)
<223> This region may encompass 1-6 residues

<220>
<221> MOD_RES
<222> (168)..(177)
<223> Any amino acid

<220>
<221> SITE
<222> (168)..(177)
<223> This region may encompass 1-10 residues

<400> 235
Trp Gly Gly Pro Gln Ile Pro Asp Gln Pro Val Glu Asp Pro Lys Xaa
1               5                   10                  15

Gln Gly Thr Tyr Pro Val Pro Asp Thr Xaa Gln Gln Thr Ile Gln Ile
            20                  25                  30

Xaa Asn Pro Leu Lys Gln Lys Pro Glu Thr Met Phe His Asp Trp Asp
        35                  40                  45

Tyr Arg Arg Gly Ile Ile Thr Ser Thr Ala Leu Lys Arg Met Gln Glu
    50                  55                  60

Asn Leu Glu Thr Asp Ser Ser Phe Xaa Ser Asp Ser Glu Glu Thr Pro
65                  70                  75                  80

Xaa Xaa Lys Lys Lys Lys Arg Leu Thr Xaa Glu Leu Pro Xaa Pro Gln
                85                  90                  95

Glu Glu Thr Glu Glu Ile Gln Ser Cys Leu Leu Ser Leu Cys Glu Glu
            100                 105                 110

Ser Thr Cys Gln Glu Glu Xaa Xaa Xaa Xaa Xaa Xaa Glu Asn Leu Gln
        115                 120                 125

Gln Leu Ile His Gln Gln Gln Gln Gln Gln Gln Gln Leu Lys His Asn
    130                 135                 140

Ile Leu Lys Leu Leu Ser Asp Leu Lys Glx Lys Gln Arg Leu Leu Gln
145                 150                 155                 160

Leu Gln Thr Gly Ile Leu Glu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
                165                 170                 175

Xaa

<210> 236

<400> 236
000

<210> 237

<400> 237
000


<210> 238

<400> 238
000


<210> 239

<400> 239
000


<210> 240

<400> 240
000


<210> 241

<400> 241
000


<210> 242

<400> 242
000


<210> 243

<400> 243
000


<210> 244

<400> 244
000


<210> 245

<400> 245
000

<210> 246

<400> 246
000

<210> 247

<400> 247
000

<210> 248

<400> 248
000

<210> 249

<400> 249
000

<210> 250

<400> 250
000

<210> 251

<400> 251
000

<210> 252

<400> 252
000

<210> 253

<400> 253

000

<210> 254

<400> 254
000

<210> 255

<400> 255
000

<210> 256

<400> 256
000

<210> 257

<400> 257
000

<210> 258

<400> 258
000

<210> 259

<400> 259
000

<210> 260

<400> 260
000

<210> 261

<400> 261
000

<210> 262

<400> 262
000


<210> 263

<400> 263
000


<210> 264

<400> 264
000


<210> 265

<400> 265
000


<210> 266

<400> 266
000


<210> 267

<400> 267
000


<210> 268

<400> 268
000


<210> 269

<400> 269
000


<210> 270

<400> 270
000

<210> 271

<400> 271
000

<210> 272

<400> 272
000

<210> 273

<400> 273
000

<210> 274

<400> 274
000

<210> 275

<400> 275
000

<210> 276

<400> 276
000

<210> 277

<400> 277
000

<210> 278

<400> 278

000

<210> 279

<400> 279
000

<210> 280

<400> 280
000

<210> 281

<400> 281
000

<210> 282

<400> 282
000

<210> 283

<400> 283
000

<210> 284

<400> 284
000

<210> 285

<400> 285
000

<210> 286

<400> 286
000

<210> 287

<400> 287
000


<210> 288

<400> 288
000


<210> 289

<400> 289
000


<210> 290

<400> 290
000


<210> 291

<400> 291
000


<210> 292

<400> 292
000


<210> 293

<400> 293
000


<210> 294

<400> 294
000


<210> 295

<400> 295
000

<210> 296

<400> 296
000

<210> 297

<400> 297
000

<210> 298

<400> 298
000

<210> 299

<400> 299
000

<210> 300

<400> 300
000

<210> 301

<400> 301
000

<210> 302

<400> 302
000

<210> 303

<400> 303

000

<210> 304

<400> 304
000

<210> 305

<400> 305
000

<210> 306

<400> 306
000

<210> 307

<400> 307
000

<210> 308

<400> 308
000

<210> 309

<400> 309
000

<210> 310

<400> 310
000

<210> 311

<400> 311
000

<210> 312

<400> 312
000


<210> 313

<400> 313
000


<210> 314

<400> 314
000


<210> 315

<400> 315
000


<210> 316

<400> 316
000


<210> 317

<400> 317
000


<210> 318

<400> 318
000


<210> 319

<400> 319
000


<210> 320

<400> 320
000

<210> 321

<400> 321
000

<210> 322

<400> 322
000

<210> 323

<400> 323
000

<210> 324

<400> 324
000

<210> 325

<400> 325
000

<210> 326

<400> 326
000

<210> 327

<400> 327
000

<210> 328

<400> 328

000

<210> 329

<400> 329
000

<210> 330

<400> 330
000

<210> 331

<400> 331
000

<210> 332

<400> 332
000

<210> 333

<400> 333
000

<210> 334

<400> 334
000

<210> 335

<400> 335
000

<210> 336

<400> 336
000

<210> 337

<400> 337
000


<210> 338

<400> 338
000


<210> 339

<400> 339
000


<210> 340

<400> 340
000


<210> 341

<400> 341
000


<210> 342

<400> 342
000


<210> 343

<400> 343
000


<210> 344

<400> 344
000


<210> 345

<400> 345
000

<210> 346

<400> 346
000

<210> 347

<400> 347
000

<210> 348

<400> 348
000

<210> 349

<400> 349
000

<210> 350

<400> 350
000

<210> 351

<400> 351
000

<210> 352

<400> 352
000

<210> 353

<400> 353

000

<210> 354

<400> 354
000

<210> 355

<400> 355
000

<210> 356

<400> 356
000

<210> 357

<400> 357
000

<210> 358

<400> 358
000

<210> 359

<400> 359
000

<210> 360

<400> 360
000

<210> 361

<400> 361
000

<210> 362

<400> 362
000


<210> 363

<400> 363
000


<210> 364

<400> 364
000


<210> 365

<400> 365
000


<210> 366

<400> 366
000


<210> 367

<400> 367
000


<210> 368

<400> 368
000


<210> 369

<400> 369
000


<210> 370

<400> 370
000

<210> 371

<400> 371
000

<210> 372

<400> 372
000

<210> 373

<400> 373
000

<210> 374

<400> 374
000

<210> 375

<400> 375
000

<210> 376

<400> 376
000

<210> 377

<400> 377
000

<210> 378

<400> 378

000

<210> 379

<400> 379
000

<210> 380

<400> 380
000

<210> 381

<400> 381
000

<210> 382

<400> 382
000

<210> 383

<400> 383
000

<210> 384

<400> 384
000

<210> 385

<400> 385
000

<210> 386

<400> 386
000

<210> 387

<400> 387
000


<210> 388

<400> 388
000


<210> 389

<400> 389
000


<210> 390

<400> 390
000


<210> 391

<400> 391
000


<210> 392

<400> 392
000


<210> 393

<400> 393
000


<210> 394

<400> 394
000


<210> 395

<400> 395
000

<210> 396

<400> 396
000

<210> 397

<400> 397
000

<210> 398

<400> 398
000

<210> 399

<400> 399
000

<210> 400

<400> 400
000

<210> 401

<400> 401
000

<210> 402

<400> 402
000

<210> 403

<400> 403

000

<210> 404

<400> 404
000

<210> 405

<400> 405
000

<210> 406

<400> 406
000

<210> 407

<400> 407
000

<210> 408

<400> 408
000

<210> 409

<400> 409
000

<210> 410

<400> 410
000

<210> 411

<400> 411
000

<210> 412

<400> 412
000


<210> 413

<400> 413
000


<210> 414

<400> 414
000


<210> 415

<400> 415
000


<210> 416

<400> 416
000


<210> 417

<400> 417
000


<210> 418

<400> 418
000


<210> 419

<400> 419
000


<210> 420

<400> 420
000

<210> 421

<400> 421
000

<210> 422

<400> 422
000

<210> 423

<400> 423
000

<210> 424

<400> 424
000

<210> 425

<400> 425
000

<210> 426

<400> 426
000

<210> 427

<400> 427
000

<210> 428

<400> 428

000

<210> 429

<400> 429
000

<210> 430

<400> 430
000

<210> 431

<400> 431
000

<210> 432

<400> 432
000

<210> 433

<400> 433
000

<210> 434

<400> 434
000

<210> 435

<400> 435
000

<210> 436

<400> 436
000

<210> 437

<400> 437
000


<210> 438

<400> 438
000


<210> 439

<400> 439
000


<210> 440

<400> 440
000


<210> 441

<400> 441
000


<210> 442

<400> 442
000


<210> 443

<400> 443
000


<210> 444

<400> 444
000


<210> 445

<400> 445
000

<210> 446

<400> 446
000

<210> 447

<400> 447
000

<210> 448

<400> 448
000

<210> 449

<400> 449
000

<210> 450

<400> 450
000

<210> 451

<400> 451
000

<210> 452

<400> 452
000

<210> 453

<400> 453

000

<210> 454

<400> 454
000

<210> 455

<400> 455
000

<210> 456

<400> 456
000

<210> 457

<400> 457
000

<210> 458

<400> 458
000

<210> 459

<400> 459
000

<210> 460

<400> 460
000

<210> 461

<400> 461
000

<210> 462

<400> 462
000


<210> 463

<400> 463
000


<210> 464

<400> 464
000


<210> 465

<400> 465
000


<210> 466

<400> 466
000


<210> 467

<400> 467
000


<210> 468

<400> 468
000


<210> 469

<400> 469
000


<210> 470

<400> 470
000

<210> 471

<400> 471
000

<210> 472

<400> 472
000

<210> 473

<400> 473
000

<210> 474

<400> 474
000

<210> 475

<400> 475
000

<210> 476

<400> 476
000

<210> 477

<400> 477
000

<210> 478

<400> 478

000

<210> 479

<400> 479
000

<210> 480

<400> 480
000

<210> 481

<400> 481
000

<210> 482

<400> 482
000

<210> 483

<400> 483
000

<210> 484

<400> 484
000

<210> 485

<400> 485
000

<210> 486

<400> 486
000

<210> 487

<400> 487
000


<210> 488

<400> 488
000


<210> 489

<400> 489
000


<210> 490

<400> 490
000


<210> 491

<400> 491
000


<210> 492

<400> 492
000


<210> 493

<400> 493
000


<210> 494

<400> 494
000


<210> 495

<400> 495
000

<210> 496

<400> 496
000

<210> 497

<400> 497
000

<210> 498

<400> 498
000

<210> 499

<400> 499
000

<210> 500

<400> 500
000

<210> 501

<400> 501
000

<210> 502

<400> 502
000

<210> 503

<400> 503

000

<210> 504

<400> 504
000

<210> 505

<400> 505
000

<210> 506

<400> 506
000

<210> 507

<400> 507
000

<210> 508

<400> 508
000

<210> 509

<400> 509
000

<210> 510

<400> 510
000

<210> 511

<400> 511
000

<210> 512

<400> 512
000


<210> 513

<400> 513
000


<210> 514

<400> 514
000


<210> 515

<400> 515
000


<210> 516

<400> 516
000


<210> 517

<400> 517
000


<210> 518

<400> 518
000


<210> 519

<400> 519
000


<210> 520

<400> 520
000

<210> 521

<400> 521
000

<210> 522

<400> 522
000

<210> 523

<400> 523
000

<210> 524

<400> 524
000

<210> 525

<400> 525
000

<210> 526

<400> 526
000

<210> 527

<400> 527
000

<210> 528

<400> 528

000

<210> 529

<400> 529
000

<210> 530

<400> 530
000

<210> 531

<400> 531
000

<210> 532

<400> 532
000

<210> 533

<400> 533
000

<210> 534

<400> 534
000

<210> 535

<400> 535
000

<210> 536

<400> 536
000

<210> 537

<400> 537
000


<210> 538

<400> 538
000


<210> 539

<400> 539
000


<210> 540

<400> 540
000


<210> 541

<400> 541
000


<210> 542

<400> 542
000


<210> 543

<400> 543
000


<210> 544

<400> 544
000


<210> 545

<400> 545
000

<210> 546

<400> 546
000

<210> 547

<400> 547
000

<210> 548

<400> 548
000

<210> 549

<400> 549
000

<210> 550

<400> 550
000

<210> 551

<400> 551
000

<210> 552

<400> 552
000

<210> 553

<400> 553

000

<210> 554

<400> 554
000

<210> 555

<400> 555
000

<210> 556

<400> 556
000

<210> 557

<400> 557
000

<210> 558

<400> 558
000

<210> 559

<400> 559
000

<210> 560

<400> 560
000

<210> 561

<400> 561
000

<210> 562

<400> 562
000


<210> 563

<400> 563
000


<210> 564

<400> 564
000


<210> 565

<400> 565
000


<210> 566

<400> 566
000


<210> 567

<400> 567
000


<210> 568

<400> 568
000


<210> 569

<400> 569
000


<210> 570

<400> 570
000

<210> 571

<400> 571
000

<210> 572

<400> 572
000

<210> 573

<400> 573
000

<210> 574

<400> 574
000

<210> 575

<400> 575
000

<210> 576

<400> 576
000

<210> 577

<400> 577
000

<210> 578

<400> 578

000

<210> 579

<400> 579
000

<210> 580

<400> 580
000

<210> 581

<400> 581
000

<210> 582

<400> 582
000

<210> 583

<400> 583
000

<210> 584

<400> 584
000

<210> 585

<400> 585
000

<210> 586

<400> 586
000

<210> 587

<400> 587
000


<210> 588

<400> 588
000


<210> 589

<400> 589
000


<210> 590

<400> 590
000


<210> 591

<400> 591
000


<210> 592

<400> 592
000


<210> 593

<400> 593
000


<210> 594

<400> 594
000


<210> 595

<400> 595
000

<210> 596

<400> 596
000

<210> 597

<400> 597
000

<210> 598

<400> 598
000

<210> 599

<400> 599
000

<210> 600

<400> 600
000

<210> 601

<400> 601
000

<210> 602

<400> 602
000

<210> 603

<400> 603

000

<210> 604

<400> 604
000

<210> 605

<400> 605
000

<210> 606

<400> 606
000

<210> 607

<400> 607
000

<210> 608

<400> 608
000

<210> 609

<400> 609
000

<210> 610

<400> 610
000

<210> 611

<400> 611
000

<210> 612

<400> 612
000


<210> 613

<400> 613
000


<210> 614

<400> 614
000


<210> 615

<400> 615
000


<210> 616

<400> 616
000


<210> 617

<400> 617
000


<210> 618

<400> 618
000


<210> 619

<400> 619
000


<210> 620

<400> 620
000

<210> 621

<400> 621
000

<210> 622

<400> 622
000

<210> 623

<400> 623
000

<210> 624

<400> 624
000

<210> 625

<400> 625
000

<210> 626

<400> 626
000

<210> 627

<400> 627
000

<210> 628

<400> 628

000

<210> 629

<400> 629
000

<210> 630

<400> 630
000

<210> 631

<400> 631
000

<210> 632

<400> 632
000

<210> 633

<400> 633
000

<210> 634

<400> 634
000

<210> 635

<400> 635
000

<210> 636

<400> 636
000

<210> 637

<400> 637
000


<210> 638

<400> 638
000


<210> 639

<400> 639
000


<210> 640

<400> 640
000


<210> 641

<400> 641
000


<210> 642

<400> 642
000


<210> 643

<400> 643
000


<210> 644

<400> 644
000


<210> 645

<400> 645
000

<210> 646

<400> 646
000

<210> 647

<400> 647
000

<210> 648

<400> 648
000

<210> 649

<400> 649
000

<210> 650

<400> 650
000

<210> 651

<400> 651
000

<210> 652

<400> 652
000

<210> 653

<400> 653

000

<210> 654

<400> 654
000

<210> 655

<400> 655
000

<210> 656

<400> 656
000

<210> 657

<400> 657
000

<210> 658

<400> 658
000

<210> 659

<400> 659
000

<210> 660

<400> 660
000

<210> 661

<400> 661
000

<210> 662

<400> 662
000


<210> 663

<400> 663
000


<210> 664

<400> 664
000


<210> 665

<400> 665
000


<210> 666

<400> 666
000


<210> 667

<400> 667
000


<210> 668

<400> 668
000


<210> 669

<400> 669
000


<210> 670

<400> 670
000

<210> 671

<400> 671
000

<210> 672

<400> 672
000

<210> 673

<400> 673
000

<210> 674

<400> 674
000

<210> 675

<400> 675
000

<210> 676

<400> 676
000

<210> 677

<400> 677
000

<210> 678

<400> 678

000

<210> 679

<400> 679
000

<210> 680

<400> 680
000

<210> 681

<400> 681
000

<210> 682

<400> 682
000

<210> 683

<400> 683
000

<210> 684

<400> 684
000

<210> 685

<400> 685
000

<210> 686

<400> 686
000

<210> 687

<400> 687
000


<210> 688

<400> 688
000


<210> 689

<400> 689
000


<210> 690

<400> 690
000


<210> 691

<400> 691
000


<210> 692

<400> 692
000


<210> 693

<400> 693
000


<210> 694

<400> 694
000


<210> 695

<400> 695
000

<210> 696

<400> 696
000

<210> 697

<400> 697
000

<210> 698

<400> 698
000

<210> 699

<400> 699
000

<210> 700

<400> 700
000

<210> 701

<400> 701
000

<210> 702

<400> 702
000

<210> 703

<400> 703

000

<210> 704

<400> 704
000

<210> 705

<400> 705
000

<210> 706

<400> 706
000

<210> 707

<400> 707
000

<210> 708

<400> 708
000

<210> 709

<400> 709
000

<210> 710

<400> 710
000

<210> 711

<400> 711
000

<210> 712

<400> 712
000


<210> 713

<400> 713
000


<210> 714

<400> 714
000


<210> 715

<400> 715
000


<210> 716

<400> 716
000


<210> 717

<400> 717
000


<210> 718

<400> 718
000


<210> 719

<400> 719
000


<210> 720

<400> 720
000

<210> 721

<400> 721
000

<210> 722

<400> 722
000

<210> 723

<400> 723
000

<210> 724

<400> 724
000

<210> 725

<400> 725
000

<210> 726

<400> 726
000

<210> 727

<400> 727
000

<210> 728

<400> 728

000

<210> 729

<400> 729
000

<210> 730

<400> 730
000

<210> 731

<400> 731
000

<210> 732

<400> 732
000

<210> 733

<400> 733
000

<210> 734

<400> 734
000

<210> 735

<400> 735
000

<210> 736

<400> 736
000

<210> 737

<400> 737
000


<210> 738

<400> 738
000


<210> 739

<400> 739
000


<210> 740

<400> 740
000


<210> 741

<400> 741
000


<210> 742

<400> 742
000


<210> 743

<400> 743
000


<210> 744

<400> 744
000


<210> 745

<400> 745
000

<210> 746

<400> 746
000

<210> 747

<400> 747
000

<210> 748

<400> 748
000

<210> 749

<400> 749
000

<210> 750

<400> 750
000

<210> 751

<400> 751
000

<210> 752

<400> 752
000

<210> 753

<400> 753

000

<210> 754

<400> 754
000

<210> 755

<400> 755
000

<210> 756

<400> 756
000

<210> 757

<400> 757
000

<210> 758

<400> 758
000

<210> 759

<400> 759
000

<210> 760

<400> 760
000

<210> 761

<400> 761
000

<210> 762

<400> 762
000


<210> 763

<400> 763
000


<210> 764

<400> 764
000


<210> 765

<400> 765
000


<210> 766

<400> 766
000


<210> 767

<400> 767
000


<210> 768

<400> 768
000


<210> 769

<400> 769
000


<210> 770

<400> 770
000

<210> 771

<400> 771
000

<210> 772

<400> 772
000

<210> 773

<400> 773
000

<210> 774

<400> 774
000

<210> 775

<400> 775
000

<210> 776

<400> 776
000

<210> 777

<400> 777
000

<210> 778

<400> 778

000

<210> 779

<400> 779
000

<210> 780

<400> 780
000

<210> 781

<400> 781
000

<210> 782

<400> 782
000

<210> 783

<400> 783
000

<210> 784

<400> 784
000

<210> 785

<400> 785
000

<210> 786

<400> 786
000

<210> 787

<400> 787
000


<210> 788

<400> 788
000


<210> 789

<400> 789
000


<210> 790

<400> 790
000


<210> 791

<400> 791
000


<210> 792

<400> 792
000


<210> 793

<400> 793
000


<210> 794

<400> 794
000


<210> 795

<400> 795
000

<210> 796

<400> 796
000

<210> 797

<400> 797
000

<210> 798

<400> 798
000

<210> 799

<400> 799
000

<210> 800

<400> 800
000

<210> 801
<211> 156
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 801
gcggcggggg ggcggccgcg ttcgcgcgcc gcccaccagg gggtgctgcg cgccccccc 60

cgcgcatgcg cggggccccc ccccgggggg gctccgcccc cccggccccc ccccgtgcta 120

aacccaccgc gcatgcgcga ccacgccccc gccgcc 156

<210> 802
<211> 150
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 802
ccgagcgtta gcgaggagtg cgaccctacc ccctgggccc acttcttcgg agccgcgcgc 60

tacgccttcg gctgcgcgcg gcacctcaga cccccgctcg tgctgacacg cttgcgcgtg 120

tcagaccact tcgggctcgc gggggtcggg 150

<210> 803
<211> 122
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 803
gccgccgcgg cggcgggggg cggcgcgctg cgcgcgccgc ccagtagggg gagccatgcg 60

ccccccccg cgcatgcgcg gggcccccc ccgcgggggg ctccgccccc cggcccccc 120

cg 122

<210> 804
<211> 111
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 804
cggcccagcg gcggcgcgcg cgcttcgcgc gcgcgccggg gggctccgcc ccccccgcg 60

catgcgcggg gcccccccc gcgggggct ccgcccccg gtcccccccc g 111

<210> 805
<211> 115
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 805
cggccgtgcg gcggcgcgcg cgcttcgcgc gcgcgccggg ggctgccgcc ccccccgcg 60

catgcgcgcg gggcccccc ccgcgggggg ctccgccccc cggcccccc ccccg 115

<210> 806
<211> 104
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 806
cggcggcggc gcgcgcgcta cgcgcgcgcg ccgggggggct gccgcccccc ccccgcgcat 60

gcgcggggcc ccccccgcg gggggctccg ccccccggcc cccc 104

<210> 807
<211> 108
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 807
ggcggcggcg cgcgcgctac gcgcgcgcgc cggggagctc tgccccccc cgcgcatgcg 60

cgcgggtccc ccccccgcgg ggggctccgc cccccggtcc cccccccg 108

<210> 808

<400> 808
000

<210> 809

<400> 809
000


<210> 810

<400> 810
000


<210> 811

<400> 811
000


<210> 812

<400> 812
000


<210> 813

<400> 813
000


<210> 814

<400> 814
000


<210> 815

<400> 815
000


<210> 816

<400> 816
000

<210> 817

<400> 817
000


<210> 818

<400> 818
000


<210> 819

<400> 819
000


<210> 820

<400> 820
000


<210> 821

<400> 821
000


<210> 822

<400> 822
000


<210> 823

<400> 823
000


<210> 824

<400> 824
000


<210> 825

<400> 825
000

<210> 826

<400> 826
000

<210> 827

<400> 827
000

<210> 828

<400> 828
000

<210> 829
<211> 11
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<220>
<221> MOD_RES
<222> (4)..(5)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (7)..(7)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (9)..(10)
<223> Any amino acid

<400> 829
Tyr Asn Pro Xaa Xaa Asp Xaa Gly Xaa Xaa Asn
1 5 10

<210> 830

<400> 830
000


<210> 831

<400> 831
000


<210> 832

<400> 832
000


<210> 833

<400> 833
000


<210> 834

<400> 834
000


<210> 835

<400> 835
000


<210> 836

<400> 836
000


<210> 837

<400> 837
000

<210> 838

<400> 838
000


<210> 839

<400> 839
000


<210> 840

<400> 840
000


<210> 841

<400> 841
000


<210> 842

<400> 842
000


<210> 843

<400> 843
000


<210> 844

<400> 844
000


<210> 845

<400> 845
000


<210> 846

<400> 846
000

<210> 847

<400> 847
000

<210> 848

<400> 848
000

<210> 849

<400> 849
000

<210> 850

<400> 850
000

<210> 851

<400> 851
000

<210> 852

<400> 852
000

<210> 853

<400> 853
000

<210> 854

<400> 854
000

<210> 855

<400> 855
000


<210> 856

<400> 856
000


<210> 857

<400> 857
000


<210> 858

<400> 858
000


<210> 859

<400> 859
000


<210> 860

<400> 860
000


<210> 861

<400> 861
000


<210> 862

<400> 862
000

<210> 863

<400> 863
000


<210> 864

<400> 864
000


<210> 865

<400> 865
000


<210> 866

<400> 866
000


<210> 867

<400> 867
000


<210> 868

<400> 868
000


<210> 869

<400> 869
000


<210> 870

<400> 870
000


<210> 871

<400> 871
000

<210> 872

<400> 872
000

<210> 873

<400> 873
000

<210> 874

<400> 874
000

<210> 875

<400> 875
000

<210> 876

<400> 876
000

<210> 877

<400> 877
000

<210> 878

<400> 878
000

<210> 879

<400> 879
000

<210> 880

<400> 880
000


<210> 881

<400> 881
000


<210> 882

<400> 882
000


<210> 883

<400> 883
000


<210> 884

<400> 884
000


<210> 885

<400> 885
000


<210> 886
<211> 3176
<212> DNA
<213> Gammatorquevirus sp.

<400> 886
taaaatggcg ggagccaatc attttatact ttcactttcc aattaaaaat ggccacgtca 60

caaacaaggg gtggagccat ttaaactata taactaagtg gggtggcgaa tggctgagtt 120

taccccgcta gacggtgcag ggaccggatc gagcgcagcg aggaggtccc cggctgccca 180

tgggcgggag ccgaggtgag tgaaaccacc gaggtctagg ggcaattcgg gctagggcag 240

tctagcggaa cgggcaagaa acttaaaaca atatttgttt tacagatggt tagtatatcc 300

tcaagtgatt tttttaagaa aacgaaattt aatgaggaga cgcagaacca agtatggatg 360

tctcaaattg ctgactctca tgataatatc tgcagttgct ggcatccatt tgctcacctt 420

cttgcttcca tatttcctcc tggccacaaa gatcgtgatc ttactattaa ccaaattctt 480

ctaagagatt ataaagaaaa atgccattct ggtggagaag aaggagaaaa ttctggacca 540

acaacaggtt taattacacc aaaagaagaa gatatagaaa aagatggccc agaaggcgcc 600

gcagaagaag accatacaga cgccctgttc gccgccgccg tagaaaactt cgaaaggtaa 660

agagaaaaaa aaaatcttta attgttagac aatggcaacc agacagtata agaacttgta 720

aaattatagg acagtcagct atagttgttg gggctgaagg aaagcaaatg tactgttata 780

ctgtcaataa gttaattaat gtgcccccaa aaacaccata tgggggaggc tttggagtag 840

accaatacac actgaaatac ttatatgaag aatacagatt tgcacaaaac atttggacac 900

aatctaatgt actgaaagac ttatgcagat acataaatgt taagctaata ttctacagag 960

acaacaaaac agactttgtc ctttcctatg acagaaaccc acctttttcaa ctaacaaaat 1020

ttacataccc aggagcacac ccacaacaaa tcatgcttca aaaacaccac aaattcatac 1080

tatcacaaat gacaaagcct aatggaagac taacaaaaaa actcaaaatt aaacctccta 1140

aacaaatgct ttctaaatgg ttcttttcaa aacaattctg taaataccct ttactatctc 1200

ttaaagcttc tgcactagac cttaggcact cttacctagg ctgctgtaat gaaaatccac 1260

aggtattttt ttattattta aaccatggat actacacaat aacaaactgg ggagcacaat 1320

cctcaacagc atacagacct aactccaagg tgacagacac aacatactac agatacaaaa 1380

atgacagaaa aaatattaac attaaaagcc atgaatacga aaaaagtata tcatatgaaa 1440

acggttattt tcaatctagt ttcttacaaa cacagtgcat atataccagt gagcgtggtg 1500

aagcctgtat agcagaaaaa ccactaggaa tagctattta caatccagta aaagacaatg 1560

gagatggtaa tatgatatac cttgtaagca ctctagcaaa cacttgggac cagcctccaa 1620

aagacagtgc tattttaata caaggagtac ccatatggct aggcttattt ggatatttag 1680

actactgtag acaaattaaa gctgacaaaa catggctaga cagtcatgta ctagtaattc 1740

aaagtcctgc tatttttact tacccaaatc caggagcagg caaatggtat tgtccactat 1800

cacaaagttt tataaatggc aatggtccgt ttaatcaacc acctacactg ctacaaaaag 1860

caaagtggtt tccacaaata caataccaac aagaaattat taatagcttt gtagaatcag 1920

gaccatttgt tcccaaatat gcaaatcaaa ctgaaagcaa ctgggaacta aaatataaat 1980

atgtttttac atttaagtgg ggtggaccac aattccatga accagaaatt gctgacccta 2040

gcaaacaaga gcagtatgat gtccccgata ctttctacca aacaatacaa attgaagatc 2100

cagaaggaca agaccccaga tctctcatcc atgattggga ctacagacga ggctttatta 2160

aagaaagatc tcttaaaaga atgtcaactt acttctcaac tcatacagat cagcaagcaa 2220

cttcagagga agacattccc aaaaagaaaa agagaattgg accccaactc acagtcccac 2280

aacaaaaaga agaggagaca ctgtcatgtc tcctctctct ctgcaaaaaa gataccttcc 2340

aagaaacaga gacacaagaa gacctccagc agctcatcaa gcagcagcag gagcagcagc 2400

tcctcctcaa gagaaacatc ctccagctca tccacaaact aaaagagaat caacaaatgc 2460

ttcagcttca cacaggcatg ttaccttaac cagatttaaa cctggatttg aagagcaaac 2520

agagagagaa ttagcaatta tatttcatag gccccctaga acctacaaag aggaccttcc 2580

attctatccc tggctaccac ctgcacccct tgtacaattt aaccttaact tcaaaggcta 2640

ggccaacaat gtacacttag taaagcatgt ttattaaagc acaaccccca aaataaatgt 2700

aaaaataaaa aaaaaaaaaa aaaaataaaa aattgcaaaa attcggcgct cgcgcgcatg 2760

tgcgcctctg gcgcaaatca cgcaacgctc gcgcgcccgc gtatgtctct ttaccacgca 2820

cctagattgg ggtgcgcgcg ctagcgcgcg caccccaatg cgccccgccc tcgttccgac 2880

ccgcttgcgc gggtcggacc acttcgggct cgggggggcg cgcctgcggc gctttttac 2940

taaacagact ccgagccgcc atttggcccc ctaagctccg ccccctcat gaatattcat 3000

aaaggaaacc acataattag aattgccgac cacaaactgc catatgctaa ttagttcccc 3060

ttttacaaag taaaagggga agtgaacata gccccacacc cgcaggggca aggccccgca 3120

cccctacgtc actaaccacg cccccgccgc catcttgggt gcggcagggc gggggc 3176

<210> 887
<211> 124
<212> PRT
<213> Gammatorquevirus sp.

<400> 887
Met Val Ser Ile Ser Ser Ser Asp Phe Phe Lys Lys Thr Lys Phe Asn
1               5                   10                  15

Glu Glu Thr Gln Asn Gln Val Trp Met Ser Gln Ile Ala Asp Ser His
            20                  25                  30

Asp Asn Ile Cys Ser Cys Trp His Pro Phe Ala His Leu Leu Ala Ser
        35                  40                  45

Ile Phe Pro Pro Gly His Lys Asp Arg Asp Leu Thr Ile Asn Gln Ile
    50                  55                  60

Leu Leu Arg Asp Tyr Lys Glu Lys Cys His Ser Gly Gly Glu Glu Gly
65                  70                  75                  80

Glu Asn Ser Gly Pro Thr Thr Gly Leu Ile Thr Pro Lys Glu Glu Asp
                85                  90                  95

Ile Glu Lys Asp Gly Pro Glu Gly Ala Ala Glu Glu Asp His Thr Asp
            100                 105                 110

Ala Leu Phe Ala Ala Ala Val Glu Asn Phe Glu Arg
        115                 120

<210> 888
<211> 271
<212> PRT
<213> Gammatorquevirus sp.

<400> 888
Met Val Ser Ile Ser Ser Ser Asp Phe Phe Lys Lys Thr Lys Phe Asn
1               5                   10                  15

Glu Glu Thr Gln Asn Gln Val Trp Met Ser Gln Ile Ala Asp Ser His

```
            20                  25                  30

Asp Asn Ile Cys Ser Cys Trp His Pro Phe Ala His Leu Leu Ala Ser
        35                  40                  45

Ile Phe Pro Pro Gly His Lys Asp Arg Asp Leu Thr Ile Asn Gln Ile
    50                  55                  60

Leu Leu Arg Asp Tyr Lys Glu Lys Cys His Ser Gly Gly Glu Glu Gly
65                  70                  75                  80

Glu Asn Ser Gly Pro Thr Thr Gly Leu Ile Thr Pro Lys Glu Glu Asp
                85                  90                  95

Ile Glu Lys Asp Gly Pro Glu Gly Ala Ala Glu Glu Asp His Thr Asp
            100                 105                 110

Ala Leu Phe Ala Ala Ala Val Glu Asn Phe Glu Ser Gly Val Asp His
        115                 120                 125

Asn Ser Met Asn Gln Lys Leu Leu Thr Leu Ala Asn Lys Ser Ser Met
    130                 135                 140

Met Ser Pro Ile Leu Ser Thr Lys Gln Tyr Lys Leu Lys Ile Gln Lys
145                 150                 155                 160

Asp Lys Thr Pro Asp Leu Ser Ser Met Ile Gly Thr Thr Asp Glu Ala
                165                 170                 175

Leu Leu Lys Lys Asp Leu Leu Lys Glu Cys Gln Leu Thr Ser Gln Leu
            180                 185                 190

Ile Gln Ile Ser Lys Gln Leu Gln Arg Lys Thr Phe Pro Lys Arg Lys
        195                 200                 205

Arg Glu Leu Asp Pro Asn Ser Gln Ser His Asn Lys Lys Lys Arg Arg
    210                 215                 220
```

```
His Cys His Val Ser Ser Leu Ser Ala Lys Lys Ile Pro Ser Lys Lys
225                 230                 235                 240


Gln Arg His Lys Lys Thr Ser Ser Ser Ser Ser Ser Ser Ser Arg Ser
                245                 250                 255


Ser Ser Ser Ser Ser Arg Glu Thr Ser Ser Ser Ser Ser Thr Asn
            260                 265                 270


<210> 889
<211> 267
<212> PRT
<213> Gammatorquevirus sp.

<400> 889
Met Val Ser Ile Ser Ser Ser Asp Phe Phe Lys Lys Thr Lys Phe Asn
1               5                   10                  15


Glu Glu Thr Gln Asn Gln Val Trp Met Ser Gln Ile Ala Asp Ser His
            20                  25                  30


Asp Asn Ile Cys Ser Cys Trp His Pro Phe Ala His Leu Leu Ala Ser
        35                  40                  45


Ile Phe Pro Pro Gly His Lys Asp Arg Asp Leu Thr Ile Asn Gln Ile
    50                  55                  60


Leu Leu Arg Asp Tyr Lys Glu Lys Cys His Ser Gly Gly Glu Glu Gly
65                  70                  75                  80


Glu Asn Ser Gly Pro Thr Thr Gly Leu Ile Thr Pro Lys Glu Glu Asp
                85                  90                  95


Ile Glu Lys Asp Gly Pro Glu Gly Ala Ala Glu Glu Asp His Thr Asp
            100                 105                 110


Ala Leu Phe Ala Ala Ala Val Glu Asn Phe Glu Arg Ser Ala Ser Asn
        115                 120                 125
```

```
Phe Arg Gly Arg His Ser Gln Lys Glu Lys Glu Asn Trp Thr Pro Thr
    130                 135                 140

His Ser Pro Thr Thr Lys Arg Arg Gly Asp Thr Val Met Ser Pro Leu
145                 150                 155                 160

Ser Leu Gln Lys Arg Tyr Leu Pro Arg Asn Arg Asp Thr Arg Arg Pro
                165                 170                 175

Pro Ala Ala His Gln Ala Ala Ala Gly Ala Ala Ala Pro Pro Gln Glu
            180                 185                 190

Lys His Pro Pro Ala His Pro Gln Thr Lys Arg Glu Ser Thr Asn Ala
        195                 200                 205

Ser Ala Ser His Arg His Val Thr Leu Thr Arg Phe Lys Pro Gly Phe
    210                 215                 220

Glu Glu Gln Thr Glu Arg Glu Leu Ala Ile Ile Phe His Arg Pro Pro
225                 230                 235                 240

Arg Thr Tyr Lys Glu Asp Leu Pro Phe Tyr Pro Trp Leu Pro Pro Ala
                245                 250                 255

Pro Leu Val Gln Phe Asn Leu Asn Phe Lys Gly
            260                 265
```

<210> 890
<211> 50
<212> PRT
<213> Gammatorquevirus sp.

<400> 890

```
Met Arg Arg Arg Arg Thr Lys Tyr Gly Cys Leu Lys Leu Leu Thr Leu
1               5                   10                  15

Met Ile Ile Ser Ala Val Ala Gly Ile His Leu Leu Thr Phe Leu Leu
            20                  25                  30
```

```
Pro Tyr Phe Leu Leu Ala Thr Lys Ile Val Ile Leu Leu Leu Thr Lys
        35                  40                  45


Phe Phe
    50


<210> 891
<211> 662
<212> PRT
<213> Gammatorquevirus sp.

<400> 891
Met Pro Phe Trp Trp Arg Arg Arg Arg Lys Phe Trp Thr Asn Asn Arg
1               5                   10                  15


Phe Asn Tyr Thr Lys Arg Arg Arg Tyr Arg Lys Arg Trp Pro Arg Arg
            20                  25                  30


Arg Arg Arg Arg Arg Pro Tyr Arg Arg Pro Val Arg Arg Arg Arg Arg
        35                  40                  45


Lys Leu Arg Lys Val Lys Arg Lys Lys Lys Ser Leu Ile Val Arg Gln
    50                  55                  60


Trp Gln Pro Asp Ser Ile Arg Thr Cys Lys Ile Ile Gly Gln Ser Ala
65                  70                  75                  80


Ile Val Val Gly Ala Glu Gly Lys Gln Met Tyr Cys Tyr Thr Val Asn
                85                  90                  95


Lys Leu Ile Asn Val Pro Pro Lys Thr Pro Tyr Gly Gly Gly Phe Gly
            100                 105                 110


Val Asp Gln Tyr Thr Leu Lys Tyr Leu Tyr Glu Glu Tyr Arg Phe Ala
        115                 120                 125


Gln Asn Ile Trp Thr Gln Ser Asn Val Leu Lys Asp Leu Cys Arg Tyr
    130                 135                 140
```

```
Ile Asn Val Lys Leu Ile Phe Tyr Arg Asp Asn Lys Thr Asp Phe Val
145                 150                 155                 160

Leu Ser Tyr Asp Arg Asn Pro Pro Phe Gln Leu Thr Lys Phe Thr Tyr
                165                 170                 175

Pro Gly Ala His Pro Gln Gln Ile Met Leu Gln Lys His His Lys Phe
            180                 185                 190

Ile Leu Ser Gln Met Thr Lys Pro Asn Gly Arg Leu Thr Lys Lys Leu
        195                 200                 205

Lys Ile Lys Pro Pro Lys Gln Met Leu Ser Lys Trp Phe Phe Ser Lys
    210                 215                 220

Gln Phe Cys Lys Tyr Pro Leu Leu Ser Leu Lys Ala Ser Ala Leu Asp
225                 230                 235                 240

Leu Arg His Ser Tyr Leu Gly Cys Cys Asn Glu Asn Pro Gln Val Phe
                245                 250                 255

Phe Tyr Tyr Leu Asn His Gly Tyr Tyr Thr Ile Thr Asn Trp Gly Ala
            260                 265                 270

Gln Ser Ser Thr Ala Tyr Arg Pro Asn Ser Lys Val Thr Asp Thr Thr
        275                 280                 285

Tyr Tyr Arg Tyr Lys Asn Asp Arg Lys Asn Ile Asn Ile Lys Ser His
    290                 295                 300

Glu Tyr Glu Lys Ser Ile Ser Tyr Glu Asn Gly Tyr Phe Gln Ser Ser
305                 310                 315                 320

Phe Leu Gln Thr Gln Cys Ile Tyr Thr Ser Glu Arg Gly Glu Ala Cys
                325                 330                 335

Ile Ala Glu Lys Pro Leu Gly Ile Ala Ile Tyr Asn Pro Val Lys Asp
```

```
            340                 345                 350

Asn Gly Asp Gly Asn Met Ile Tyr Leu Val Ser Thr Leu Ala Asn Thr
        355                 360                 365

Trp Asp Gln Pro Pro Lys Asp Ser Ala Ile Leu Ile Gln Gly Val Pro
    370                 375                 380

Ile Trp Leu Gly Leu Phe Gly Tyr Leu Asp Tyr Cys Arg Gln Ile Lys
385                 390                 395                 400

Ala Asp Lys Thr Trp Leu Asp Ser His Val Leu Val Ile Gln Ser Pro
                405                 410                 415

Ala Ile Phe Thr Tyr Pro Asn Pro Gly Ala Gly Lys Trp Tyr Cys Pro
            420                 425                 430

Leu Ser Gln Ser Phe Ile Asn Gly Asn Gly Pro Phe Asn Gln Pro Pro
        435                 440                 445

Thr Leu Leu Gln Lys Ala Lys Trp Phe Pro Gln Ile Gln Tyr Gln Gln
    450                 455                 460

Glu Ile Ile Asn Ser Phe Val Glu Ser Gly Pro Phe Val Pro Lys Tyr
465                 470                 475                 480

Ala Asn Gln Thr Glu Ser Asn Trp Glu Leu Lys Tyr Lys Tyr Val Phe
                485                 490                 495

Thr Phe Lys Trp Gly Gly Pro Gln Phe His Glu Pro Glu Ile Ala Asp
            500                 505                 510

Pro Ser Lys Gln Glu Gln Tyr Asp Val Pro Asp Thr Phe Tyr Gln Thr
        515                 520                 525

Ile Gln Ile Glu Asp Pro Glu Gly Gln Asp Pro Arg Ser Leu Ile His
    530                 535                 540
```

```
Asp Trp Asp Tyr Arg Arg Gly Phe Ile Lys Glu Arg Ser Leu Lys Arg
545                 550                 555                 560


Met Ser Thr Tyr Phe Ser Thr His Thr Asp Gln Gln Ala Thr Ser Glu
                565                 570                 575


Glu Asp Ile Pro Lys Lys Lys Lys Arg Ile Gly Pro Gln Leu Thr Val
            580                 585                 590


Pro Gln Gln Lys Glu Glu Glu Thr Leu Ser Cys Leu Leu Ser Leu Cys
        595                 600                 605


Lys Lys Asp Thr Phe Gln Glu Thr Glu Thr Gln Glu Asp Leu Gln Gln
    610                 615                 620


Leu Ile Lys Gln Gln Gln Glu Gln Gln Leu Leu Leu Lys Arg Asn Ile
625                 630                 635                 640


Leu Gln Leu Ile His Lys Leu Lys Glu Asn Gln Gln Met Leu Gln Leu
                645                 650                 655


His Thr Gly Met Leu Pro
            660


<210> 892
<211> 215
<212> PRT
<213> Gammatorquevirus sp.

<400> 892
Met Pro Phe Trp Trp Arg Arg Arg Arg Lys Phe Trp Thr Asn Asn Arg
1               5                   10                  15


Phe Asn Tyr Thr Lys Arg Arg Arg Tyr Arg Lys Arg Trp Pro Arg Arg
            20                  25                  30


Arg Arg Arg Arg Arg Pro Tyr Arg Arg Pro Val Arg Arg Arg Arg Arg
        35                  40                  45
```

```
Lys Leu Arg Lys Trp Gly Gly Pro Gln Phe His Glu Pro Glu Ile Ala
    50                  55                  60

Asp Pro Ser Lys Gln Glu Gln Tyr Asp Val Pro Asp Thr Phe Tyr Gln
65                  70                  75                  80

Thr Ile Gln Ile Glu Asp Pro Glu Gly Gln Asp Pro Arg Ser Leu Ile
                85                  90                  95

His Asp Trp Asp Tyr Arg Arg Gly Phe Ile Lys Glu Arg Ser Leu Lys
            100                 105                 110

Arg Met Ser Thr Tyr Phe Ser Thr His Thr Asp Gln Gln Ala Thr Ser
        115                 120                 125

Glu Glu Asp Ile Pro Lys Lys Lys Lys Arg Ile Gly Pro Gln Leu Thr
    130                 135                 140

Val Pro Gln Gln Lys Glu Glu Glu Thr Leu Ser Cys Leu Leu Ser Leu
145                 150                 155                 160

Cys Lys Lys Asp Thr Phe Gln Glu Thr Glu Thr Gln Glu Asp Leu Gln
                165                 170                 175

Gln Leu Ile Lys Gln Gln Gln Glu Gln Gln Leu Leu Leu Lys Arg Asn
            180                 185                 190

Ile Leu Gln Leu Ile His Lys Leu Lys Glu Asn Gln Gln Met Leu Gln
        195                 200                 205

Leu His Thr Gly Met Leu Pro
    210                 215
```

<210> 893
<211> 129
<212> PRT
<213> Gammatorquevirus sp.

<400> 893
Met Pro Phe Trp Trp Arg Arg Arg Arg Lys Phe Trp Thr Asn Asn Arg
1               5                  10                  15

Phe Asn Tyr Thr Lys Arg Arg Arg Tyr Arg Lys Arg Trp Pro Arg Arg
            20                  25                  30

Arg Arg Arg Arg Arg Pro Tyr Arg Arg Pro Val Arg Arg Arg Arg Arg
        35                  40                  45

Lys Leu Arg Lys Ile Ser Lys Gln Leu Gln Arg Lys Thr Phe Pro Lys
    50                  55                  60

Arg Lys Arg Glu Leu Asp Pro Asn Ser Gln Ser His Asn Lys Lys Lys
65                  70                  75                  80

Arg Arg His Cys His Val Ser Ser Leu Ser Ala Lys Lys Ile Pro Ser
                85                  90                  95

Lys Lys Gln Arg His Lys Lys Thr Ser Ser Ser Ser Ser Ser Ser Ser
            100                 105                 110

Arg Ser Ser Ser Ser Ser Ser Arg Glu Thr Ser Ser Ser Ser Ser Thr
        115                 120                 125

Asn

<210> 894

<400> 894
000

<210> 895

<400> 895
000

<210> 896

<400> 896
000

<210> 897

<400> 897
000

<210> 898

<400> 898
000

<210> 899

<400> 899
000

<210> 900

<400> 900
000

<210> 901

<400> 901
000

<210> 902

<400> 902
000

<210> 903

<400> 903
000

<210> 904

<400> 904

000

<210> 905

<400> 905
000

<210> 906

<400> 906
000

<210> 907

<400> 907
000

<210> 908

<400> 908
000

<210> 909

<400> 909
000

<210> 910

<400> 910
000

<210> 911

<400> 911
000

<210> 912

<400> 912
000

<210> 913

<400> 913
000


<210> 914

<400> 914
000


<210> 915

<400> 915
000


<210> 916

<400> 916
000


<210> 917

<400> 917
000


<210> 918

<400> 918
000


<210> 919

<400> 919
000


<210> 920

<400> 920
000


<210> 921

<400> 921
000

<210> 922

<400> 922
000

<210> 923

<400> 923
000

<210> 924

<400> 924
000

<210> 925
<211> 662
<212> PRT
<213> Gammatorquevirus sp.

<400> 925
Met Pro Phe Trp Trp Arg Arg Arg Arg Lys Phe Trp Thr Asn Asn Arg
1  5  10  15

Phe Asn Tyr Thr Lys Arg Arg Arg Tyr Arg Lys Arg Trp Pro Arg Arg
            20  25  30

Arg Arg Arg Arg Arg Pro Tyr Arg Arg Pro Val Arg Arg Arg Arg Arg
        35  40  45

Lys Leu Arg Lys Val Lys Arg Lys Lys Lys Ser Leu Ile Val Arg Gln
    50  55  60

Trp Gln Pro Asp Ser Ile Arg Thr Cys Lys Ile Ile Gly Gln Ser Ala
65  70  75  80

Ile Val Val Gly Ala Glu Gly Lys Gln Met Tyr Cys Tyr Thr Val Asn

```
                85                  90                  95

Lys Leu Ile Asn Val Pro Pro Lys Thr Pro Tyr Gly Gly Gly Phe Gly
            100                 105                 110

Val Asp Gln Tyr Thr Leu Lys Tyr Leu Tyr Glu Glu Tyr Arg Phe Ala
        115                 120                 125

Gln Asn Ile Trp Thr Gln Ser Asn Val Leu Lys Asp Leu Cys Arg Tyr
    130                 135                 140

Ile Asn Val Lys Leu Ile Phe Tyr Arg Asp Asn Lys Thr Asp Phe Val
145                 150                 155                 160

Leu Ser Tyr Asp Arg Asn Pro Pro Phe Gln Leu Thr Lys Phe Thr Tyr
                165                 170                 175

Pro Gly Ala His Pro Gln Gln Ile Met Leu Gln Lys His His Lys Phe
            180                 185                 190

Ile Leu Ser Gln Met Thr Lys Pro Asn Gly Arg Leu Thr Lys Lys Leu
        195                 200                 205

Lys Ile Lys Pro Pro Lys Gln Met Leu Ser Lys Trp Phe Phe Ser Lys
    210                 215                 220

Gln Phe Cys Lys Tyr Pro Leu Leu Ser Leu Lys Ala Ser Ala Leu Asp
225                 230                 235                 240

Leu Arg His Ser Tyr Leu Gly Cys Cys Asn Glu Asn Pro Gln Val Phe
                245                 250                 255

Phe Tyr Tyr Leu Asn His Gly Tyr Tyr Thr Ile Thr Asn Trp Gly Ala
            260                 265                 270

Gln Ser Ser Thr Ala Tyr Arg Pro Asn Ser Lys Val Thr Asp Thr Thr
        275                 280                 285
```

```
Tyr Tyr Arg Tyr Lys Asn Asp Arg Lys Asn Ile Asn Ile Lys Ser His
    290                 295                 300

Glu Tyr Glu Lys Ser Ile Ser Tyr Glu Asn Gly Tyr Phe Gln Ser Ser
305                 310                 315                 320

Phe Leu Gln Thr Gln Cys Ile Tyr Thr Ser Glu Arg Gly Glu Ala Cys
                325                 330                 335

Ile Ala Glu Lys Pro Leu Gly Ile Ala Ile Tyr Asn Pro Val Lys Asp
            340                 345                 350

Asn Gly Asp Gly Asn Met Ile Tyr Leu Val Ser Thr Leu Ala Asn Thr
        355                 360                 365

Trp Asp Gln Pro Pro Lys Asp Ser Ala Ile Leu Ile Gln Gly Val Pro
    370                 375                 380

Ile Trp Leu Gly Leu Phe Gly Tyr Leu Asp Tyr Cys Arg Gln Ile Lys
385                 390                 395                 400

Ala Asp Lys Thr Trp Leu Asp Ser His Val Leu Val Ile Gln Ser Pro
                405                 410                 415

Ala Ile Phe Thr Tyr Pro Asn Pro Gly Ala Gly Lys Trp Tyr Cys Pro
            420                 425                 430

Leu Ser Gln Ser Phe Ile Asn Gly Asn Gly Pro Phe Asn Gln Pro Pro
        435                 440                 445

Thr Leu Leu Gln Lys Ala Lys Trp Phe Pro Gln Ile Gln Tyr Gln Gln
    450                 455                 460

Glu Ile Ile Asn Ser Phe Val Glu Ser Gly Pro Phe Val Pro Lys Tyr
465                 470                 475                 480

Ala Asn Gln Thr Glu Ser Asn Trp Glu Leu Lys Tyr Lys Tyr Val Phe
```

```
                485                 490                 495

Thr Phe Lys Trp Gly Gly Pro Gln Phe His Glu Pro Glu Ile Ala Asp
            500                 505                 510

Pro Ser Lys Gln Glu Gln Tyr Asp Val Pro Asp Thr Phe Tyr Gln Thr
        515                 520                 525

Ile Gln Ile Glu Asp Pro Glu Gly Gln Asp Pro Arg Ser Leu Ile His
    530                 535                 540

Asp Trp Asp Tyr Arg Arg Gly Phe Ile Lys Glu Arg Ser Leu Lys Arg
545                 550                 555                 560

Met Ser Thr Tyr Phe Ser Thr His Thr Asp Gln Gln Ala Thr Ser Glu
                565                 570                 575

Glu Asp Ile Pro Lys Lys Lys Lys Arg Ile Gly Pro Gln Leu Thr Val
            580                 585                 590

Pro Gln Gln Lys Glu Glu Glu Thr Leu Ser Cys Leu Leu Ser Leu Cys
        595                 600                 605

Lys Lys Asp Thr Phe Gln Glu Thr Glu Thr Gln Glu Asp Leu Gln Gln
    610                 615                 620

Leu Ile Lys Gln Gln Gln Glu Gln Gln Leu Leu Leu Lys Arg Asn Ile
625                 630                 635                 640

Leu Gln Leu Ile His Lys Leu Lys Glu Asn Gln Gln Met Leu Gln Leu
                645                 650                 655

His Thr Gly Met Leu Pro
            660
```

<210> 926
<211> 58
<212> PRT

<213> Gammatorquevirus sp.

<400> 926
Met Pro Phe Trp Trp Arg Arg Arg Arg Lys Phe Trp Thr Asn Asn Arg
1               5                   10                  15

Phe Asn Tyr Thr Lys Arg Arg Arg Tyr Arg Lys Arg Trp Pro Arg Arg
            20                  25                  30

Arg Arg Arg Arg Arg Pro Tyr Arg Arg Pro Val Arg Arg Arg Arg Arg
        35                  40                  45

Lys Leu Arg Lys Val Lys Arg Lys Lys Lys
    50                  55

<210> 927
<211> 202
<212> PRT
<213> Gammatorquevirus sp.

<400> 927
Ser Leu Ile Val Arg Gln Trp Gln Pro Asp Ser Ile Arg Thr Cys Lys
1               5                   10                  15

Ile Ile Gly Gln Ser Ala Ile Val Val Gly Ala Glu Gly Lys Gln Met
            20                  25                  30

Tyr Cys Tyr Thr Val Asn Lys Leu Ile Asn Val Pro Pro Lys Thr Pro
        35                  40                  45

Tyr Gly Gly Gly Phe Gly Val Asp Gln Tyr Thr Leu Lys Tyr Leu Tyr
    50                  55                  60

Glu Glu Tyr Arg Phe Ala Gln Asn Ile Trp Thr Gln Ser Asn Val Leu
65                  70                  75                  80

Lys Asp Leu Cys Arg Tyr Ile Asn Val Lys Leu Ile Phe Tyr Arg Asp
                85                  90                  95

Asn Lys Thr Asp Phe Val Leu Ser Tyr Asp Arg Asn Pro Pro Phe Gln

```
            100                 105                 110

Leu Thr Lys Phe Thr Tyr Pro Gly Ala His Pro Gln Gln Ile Met Leu
        115                 120                 125

Gln Lys His His Lys Phe Ile Leu Ser Gln Met Thr Lys Pro Asn Gly
    130                 135                 140

Arg Leu Thr Lys Lys Leu Lys Ile Lys Pro Pro Lys Gln Met Leu Ser
145                 150                 155                 160

Lys Trp Phe Phe Ser Lys Gln Phe Cys Lys Tyr Pro Leu Leu Ser Leu
                165                 170                 175

Lys Ala Ser Ala Leu Asp Leu Arg His Ser Tyr Leu Gly Cys Cys Asn
            180                 185                 190

Glu Asn Pro Gln Val Phe Phe Tyr Tyr Leu
        195                 200


<210> 928
<211> 79
<212> PRT
<213> Gammatorquevirus sp.

<400> 928
Asn His Gly Tyr Tyr Thr Ile Thr Asn Trp Gly Ala Gln Ser Ser Thr
1               5                   10                  15

Ala Tyr Arg Pro Asn Ser Lys Val Thr Asp Thr Thr Tyr Tyr Arg Tyr
            20                  25                  30

Lys Asn Asp Arg Lys Asn Ile Asn Ile Lys Ser His Glu Tyr Glu Lys
        35                  40                  45

Ser Ile Ser Tyr Glu Asn Gly Tyr Phe Gln Ser Ser Phe Leu Gln Thr
    50                  55                  60

Gln Cys Ile Tyr Thr Ser Glu Arg Gly Glu Ala Cys Ile Ala Glu
```

65 70 75

<210> 929
<211> 160
<212> PRT
<213> Gammatorquevirus sp.

<400> 929
Lys Pro Leu Gly Ile Ala Ile Tyr Asn Pro Val Lys Asp Asn Gly Asp
1 5 10 15

Gly Asn Met Ile Tyr Leu Val Ser Thr Leu Ala Asn Thr Trp Asp Gln
20 25 30

Pro Pro Lys Asp Ser Ala Ile Leu Ile Gln Gly Val Pro Ile Trp Leu
35 40 45

Gly Leu Phe Gly Tyr Leu Asp Tyr Cys Arg Gln Ile Lys Ala Asp Lys
50 55 60

Thr Trp Leu Asp Ser His Val Leu Val Ile Gln Ser Pro Ala Ile Phe
65 70 75 80

Thr Tyr Pro Asn Pro Gly Ala Gly Lys Trp Tyr Cys Pro Leu Ser Gln
85 90 95

Ser Phe Ile Asn Gly Asn Gly Pro Phe Asn Gln Pro Pro Thr Leu Leu
100 105 110

Gln Lys Ala Lys Trp Phe Pro Gln Ile Gln Tyr Gln Gln Glu Ile Ile
115 120 125

Asn Ser Phe Val Glu Ser Gly Pro Phe Val Pro Lys Tyr Ala Asn Gln
130 135 140

Thr Glu Ser Asn Trp Glu Leu Lys Tyr Lys Tyr Val Phe Thr Phe Lys
145 150 155 160

<210> 930

<211> 163
<212> PRT
<213> Gammatorquevirus sp.

<400> 930
Trp Gly Gly Pro Gln Phe His Glu Pro Glu Ile Ala Asp Pro Ser Lys
1               5                  10                  15

Gln Glu Gln Tyr Asp Val Pro Asp Thr Phe Tyr Gln Thr Ile Gln Ile
            20                  25                  30

Glu Asp Pro Glu Gly Gln Asp Pro Arg Ser Leu Ile His Asp Trp Asp
        35                  40                  45

Tyr Arg Arg Gly Phe Ile Lys Glu Arg Ser Leu Lys Arg Met Ser Thr
    50                  55                  60

Tyr Phe Ser Thr His Thr Asp Gln Gln Ala Thr Ser Glu Glu Asp Ile
65                  70                  75                  80

Pro Lys Lys Lys Lys Arg Ile Gly Pro Gln Leu Thr Val Pro Gln Gln
                85                  90                  95

Lys Glu Glu Glu Thr Leu Ser Cys Leu Leu Ser Leu Cys Lys Lys Asp
            100                 105                 110

Thr Phe Gln Glu Thr Glu Thr Gln Glu Asp Leu Gln Gln Leu Ile Lys
        115                 120                 125

Gln Gln Gln Glu Gln Gln Leu Leu Leu Lys Arg Asn Ile Leu Gln Leu
    130                 135                 140

Ile His Lys Leu Lys Glu Asn Gln Gln Met Leu Gln Leu His Thr Gly
145                 150                 155                 160

Met Leu Pro

<210> 931

<400> 931
000

<210> 932

<400> 932
000

<210> 933

<400> 933
000

<210> 934

<400> 934
000

<210> 935

<400> 935
000

<210> 936

<400> 936
000

<210> 937

<400> 937
000

<210> 938

<400> 938
000

<210> 939

<400> 939

000

<210> 940

<400> 940
000

<210> 941

<400> 941
000

<210> 942

<400> 942
000

<210> 943

<400> 943
000

<210> 944

<400> 944
000

<210> 945

<400> 945
000

<210> 946

<400> 946
000

<210> 947

<400> 947
000

<210> 948

<400> 948
000

<210> 949
<211> 21
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic peptide

<220>
<221> MOD_RES
<222> (1)..(1)
<223> W or F

<220>
<221> MOD_RES
<222> (2)..(8)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (10)..(12)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (14)..(14)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (16)..(20)
<223> Any amino acid

<400> 949
Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Cys Xaa Cys Xaa
1               5                   10                  15

Xaa Xaa Xaa Xaa His
            20

<210> 950
<211> 22
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
peptide

<220>
<221> MOD_RES
<222> (6)..(7)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (9)..(9)
<223> Any amino acid

<220>
<221> MOD_RES
<222> (15)..(16)
<223> Any amino acid

<400> 950
Tyr Asn Cys Ser Pro Xaa Xaa Asp Xaa Gly Ala Ser Lys Arg Xaa Xaa
1 5 10 15

Asn Thr Ser Val Ala Lys
20

<210> 951
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<220>
<221> modified_base
<222> (45)..(45)
<223> a, c, t, g, unknown or other

<400> 951
aggtgagtga aaccaccgaa gtcaaggggc aattcgggct agggncagtc t 51

<210> 952
<211> 50
<212> DNA
<213> Alphatorquevirus sp.

<400> 952
aggtgagttt acacaccgca gtcaaggggc aattcgggct cgggactggc 50

<210> 953
<211> 50
<212> DNA
<213> Betatorquevirus sp.

<400> 953
aggtgagtga aaccaccgaa gtcaaggggc aattcgggct agatcagtct 50

<210> 954
<211> 50
<212> DNA
<213> Gammatorquevirus sp.

<400> 954
aggtgagtga aaccaccgag gtctaggggc aattcgggct agggcagtct 50

<210> 955
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
6xHis tag

<400> 955
His His His His His His
1 5

<210> 956
<211> 237
<212> PRT
<213> Unknown

<220>
<223> Description of Unknown:
Anelloviridae family sequence

<400> 956
Arg Arg Lys Leu Arg Val Arg Lys Arg Phe Tyr Lys Arg Lys Leu Lys
1 5 10 15

Lys Ile Val Leu Lys Gln Phe Gln Pro Lys Ile Ile Arg Arg Cys Thr
20 25 30

Ile Phe Gly Thr Ile Cys Leu Phe Gln Gly Ser Pro Glu Arg Ala Asn
35 40 45

Asn Asn Tyr Ile Gln Thr Ile Tyr Ser Tyr Val Pro Asp Lys Glu Pro
50 55 60

Gly Gly Gly Gly Trp Thr Leu Ile Thr Glu Ser Leu Ser Ser Leu Trp
65 70 75 80

Glu Asp Trp Glu His Leu Lys Asn Val Trp Thr Gln Ser Asn Ala Gly
85 90 95

Leu Pro Leu Val Arg Tyr Gly Gly Val Thr Leu Tyr Phe Tyr Gln Ser
100 105 110

Ala Tyr Thr Asp Tyr Ile Ala Gln Val Phe Asn Cys Tyr Pro Met Thr
115 120 125

Asp Thr Lys Tyr Thr His Ala Asp Ser Ala Pro Asn Arg Met Leu Leu
130 135 140

Lys Lys His Val Ile Arg Val Pro Ser Arg Glu Thr Arg Lys Lys Arg
145 150 155 160

Lys Pro Tyr Lys Arg Val Arg Val Gly Pro Pro Ser Gln Met Gln Asn
165 170 175

Lys Trp Tyr Phe Gln Arg Asp Ile Cys Glu Ile Pro Leu Ile Met Ile

180 185 190

Ala Ala Thr Ala Val Asp Phe Arg Tyr Pro Phe Cys Ala Ser Asp Cys
195 200 205

Ala Ser Asn Asn Leu Thr Leu Thr Cys Leu Asn Pro Leu Leu Phe Gln
210 215 220

Asn Gln Asp Phe Asp His Pro Ser Asp Thr Gln Gly Tyr
225 230 235

<210> 957
<211> 232
<212> PRT
<213> Unknown

<220>
<223> Description of Unknown:
Anelloviridae family sequence

<400> 957
Arg Arg Lys Arg Arg Val Arg Pro Thr Tyr Thr Thr Ile Pro Leu Lys
1 5 10 15

Gln Trp Gln Pro Pro Tyr Lys Arg Thr Cys Tyr Ile Lys Gly Gln Asp
20 25 30

Cys Leu Ile Tyr Tyr Ser Asn Leu Arg Leu Gly Met Asn Ser Thr Met
35 40 45

Tyr Glu Lys Ser Ile Val Pro Val His Trp Pro Gly Gly Gly Ser Phe
50 55 60

Ser Val Ser Met Leu Thr Leu Asp Ala Leu Tyr Asp Ile His Lys Leu
65 70 75 80

Cys Arg Asn Trp Trp Thr Ser Thr Asn Gln Asp Leu Pro Leu Val Arg
85 90 95

Tyr Lys Gly Cys Lys Ile Thr Phe Tyr Gln Ser Thr Phe Thr Asp Tyr

```
            100                 105                 110

Ile Val Arg Ile His Thr Glu Leu Pro Ala Asn Ser Asn Lys Leu Thr
        115                 120                 125

Tyr Pro Asn Thr His Pro Leu Met Met Met Met Ser Lys Tyr Lys His
    130                 135                 140

Ile Ile Pro Ser Arg Gln Thr Arg Arg Lys Lys Lys Pro Tyr Thr Lys
145                 150                 155                 160

Ile Phe Val Lys Pro Pro Pro Gln Phe Glu Asn Lys Trp Tyr Phe Ala
                165                 170                 175

Thr Asp Leu Tyr Lys Ile Pro Leu Leu Gln Ile His Cys Thr Ala Cys
            180                 185                 190

Asn Leu Gln Asn Pro Phe Val Lys Pro Asp Lys Leu Ser Asn Asn Val
        195                 200                 205

Thr Leu Trp Ser Leu Asn Thr Ile Ser Ile Gln Asn Arg Asn Met Ser
    210                 215                 220

Val Asp Gln Gly Gln Ser Trp Pro
225                 230


<210> 958
<211> 238
<212> PRT
<213> Unknown

<220>
<223> Description of Unknown:
      Anelloviridae family sequence

<400> 958
Arg Arg Arg Arg Arg Trp Arg Lys Arg Pro Thr Val Arg Arg Lys Leu
1               5                   10                  15

Lys Lys Leu Thr Ile Gln Gln Trp Gln Pro Lys Thr Ile Arg Lys Cys
```

```
            20                  25                  30

Cys Ile Gln Gly Leu His Cys Leu Phe Leu Val Thr Glu Asp Thr Ile
        35                  40                  45

Ser Arg Asn Tyr Arg Met Tyr Glu His Ser Tyr Thr Gly Glu Tyr Tyr
    50                  55                  60

Pro Gly Gly Gly Gly Phe Ser Ile Thr Arg Tyr Ser Leu Asp Gly Leu
65                  70                  75                  80

Tyr Glu Gln His Gln Leu Asp Arg Asn Trp Trp Thr Asn Ser Asn Thr
                85                  90                  95

Asn Leu Pro Leu Val Arg Tyr Thr Gly Cys Lys Ile Lys Phe Tyr Gln
            100                 105                 110

Ser Trp Ser Val Asp Tyr Ile Cys Asn Tyr Ser Leu Thr Trp Pro Met
        115                 120                 125

Val Ala Thr Gln Leu Leu Tyr Gln Ser Cys Gln Pro Ser Phe Met Met
    130                 135                 140

Met Asn Lys Asn Ser Ile Met Ile Pro Ser Lys Leu Thr Lys Pro Ile
145                 150                 155                 160

Lys Lys Gly Tyr Lys Thr Ile Lys Glu Lys Pro Pro His Glu Met Leu
                165                 170                 175

Asn Arg Trp Tyr Phe Ala Lys Asp Leu Ser Lys Val Gly Leu Leu Met
            180                 185                 190

Leu Thr Ala Ala Ser Ala Ser Phe Asp His Tyr Tyr Gln Ala Thr Asp
        195                 200                 205

Ser Leu Ser Asn Asn Cys Thr Phe Glu Ser Leu Asn Pro Tyr Phe Tyr
    210                 215                 220
```

Met Arg His Asp Phe Ile Leu Phe Pro Val Thr Gly Tyr Val
225 230 235

<210> 959
<211> 238
<212> PRT
<213> Unknown

<220>
<223> Description of Unknown:
Anelloviridae family sequence

<400> 959
Arg Arg Arg Arg Arg Trp Val Arg Arg Lys Pro Phe Tyr Lys Arg Lys
1 5 10 15

Ile Lys Arg Leu Asn Ile Val Glu Trp Gln Pro Lys Ser Ile Arg Lys
20 25 30

Cys Arg Ile Lys Gly Met Leu Cys Leu Phe Gln Thr Thr Glu Asp Arg
35 40 45

Leu Ser Tyr Asn Phe Asp Met Tyr Glu Glu Ser Ile Ile Pro Glu Lys
50 55 60

Leu Pro Gly Gly Gly Gly Phe Ser Ile Lys Asn Ile Ser Leu Tyr Ala
65 70 75 80

Leu Tyr Gln Glu His Ile His Ala His Asn Ile Phe Thr His Thr Asn
85 90 95

Thr Asp Arg Pro Leu Ala Arg Tyr Thr Gly Cys Ser Leu Lys Phe Tyr
100 105 110

Gln Ser Lys Asp Ile Asp Tyr Val Val Thr Tyr Ser Thr Ser Leu Pro
115 120 125

Leu Arg Ser Ser Met Gly Met Tyr Asn Ser Met Gln Pro Ser Ile His
130 135 140

```
Leu Met Gln Gln Asn Lys Leu Ile Val Pro Ser Lys Gln Thr Gln Lys
145                 150                 155                 160


Arg Arg Lys Pro Tyr Ile Lys Lys His Ile Ser Pro Pro Thr Gln Met
                165                 170                 175


Lys Ser Gln Trp Tyr Phe Gln His Asn Ile Ala Asn Ile Pro Leu Leu
            180                 185                 190


Met Ile Arg Thr Thr Ala Leu Thr Leu Asp Asn Tyr Tyr Ile Gly Ser
        195                 200                 205


Arg Gln Leu Ser Thr Asn Val Thr Ile His Thr Leu Asn Thr Thr Tyr
    210                 215                 220


Ile Gln Asn Arg Asp Trp Gly Asp Arg Asn Lys Thr Tyr Tyr
225                 230                 235


<210> 960
<211> 240
<212> PRT
<213> Unknown

<220>
<223> Description of Unknown:
      Anelloviridae family sequence

<400> 960
Arg Arg Pro Val Arg Arg Arg Arg Arg Lys Leu Arg Lys Val Lys Arg
1               5                   10                  15


Lys Lys Lys Ser Leu Ile Val Arg Gln Trp Gln Pro Asp Ser Ile Arg
            20                  25                  30


Thr Cys Lys Ile Ile Gly Gln Ser Ala Ile Val Val Gly Ala Glu Gly
        35                  40                  45


Lys Gln Met Tyr Cys Tyr Thr Val Asn Lys Leu Ile Asn Val Pro Pro
    50                  55                  60
```

```
Lys Thr Pro Tyr Gly Gly Gly Phe Gly Val Asp Gln Tyr Thr Leu Lys
65                  70                  75                  80

Tyr Leu Tyr Glu Glu Tyr Arg Phe Ala Gln Asn Ile Trp Thr Gln Ser
                85                  90                  95

Asn Val Leu Lys Asp Leu Cys Arg Tyr Ile Asn Val Lys Leu Ile Phe
            100                 105                 110

Tyr Arg Asp Asn Lys Thr Asp Phe Val Leu Ser Tyr Asp Arg Asn Pro
        115                 120                 125

Pro Phe Gln Leu Thr Lys Phe Thr Tyr Pro Gly Ala His Pro Gln Gln
    130                 135                 140

Ile Met Leu Gln Lys His His Lys Phe Ile Leu Ser Gln Met Thr Lys
145                 150                 155                 160

Pro Asn Gly Arg Leu Thr Lys Lys Leu Lys Ile Lys Pro Pro Lys Gln
                165                 170                 175

Met Leu Ser Lys Trp Phe Phe Ser Lys Gln Phe Cys Lys Tyr Pro Leu
            180                 185                 190

Leu Ser Leu Lys Ala Ser Ala Leu Asp Leu Arg His Ser Tyr Leu Gly
        195                 200                 205

Cys Cys Asn Glu Asn Pro Gln Val Phe Phe Tyr Tyr Leu Asn His Gly
    210                 215                 220

Tyr Tyr Thr Ile Thr Asn Trp Gly Ala Gln Ser Ser Thr Ala Tyr Arg
225                 230                 235                 240

<210> 961
<211> 240
<212> PRT
<213> Unknown
```

<220>
<223> Description of Unknown:
Anelloviridae family sequence

<400> 961
Arg Phe Thr Lys Thr Arg Arg Arg Arg Lys Arg Lys Lys Val Arg Arg
1 5 10 15

Lys Leu Lys Lys Ile Thr Ile Lys Gln Trp Gln Pro Asp Ser Val Lys
20 25 30

Lys Cys Lys Ile Lys Gly Tyr Ser Thr Leu Val Met Gly Ala Gln Gly
35 40 45

Lys Gln Tyr Asn Cys Tyr Thr Asn Gln Ala Ser Asp Tyr Val Gln Pro
50 55 60

Lys Ala Pro Gln Gly Gly Gly Phe Gly Cys Glu Val Phe Asn Leu Lys
65 70 75 80

Trp Leu Tyr Gln Glu Tyr Thr Ala His Arg Asn Ile Trp Thr Lys Thr
85 90 95

Asn Glu Tyr Thr Asp Leu Cys Arg Tyr Thr Gly Ala Gln Ile Ile Leu
100 105 110

Tyr Arg His Pro Asp Val Asp Phe Ile Val Ser Trp Asp Asn Gln Pro
115 120 125

Pro Phe Leu Leu Asn Lys Tyr Thr Tyr Pro Glu Leu Gln Pro Gln Asn
130 135 140

Leu Leu Leu Ala Arg Arg Lys Arg Ile Ile Leu Ser Gln Lys Ser Asn
145 150 155 160

Pro Lys Gly Lys Leu Arg Ile Lys Leu Arg Ile Pro Pro Pro Lys Gln
165 170 175

Met Ile Thr Lys Trp Phe Phe Gln Arg Asp Phe Cys Asp Val Asn Leu

180 185 190

Phe Lys Leu Cys Ala Ser Ala Ala Ser Phe Arg Tyr Pro Gly Ile Ser
195 200 205

His Gly Ala Gln Ser Thr Ile Phe Ser Ala Tyr Ala Leu Asn Thr Asp
210 215 220

Phe Tyr Gln Cys Ser Asp Trp Cys Gln Thr Asn Thr Glu Thr Gly Tyr
225 230 235 240

<210> 962
<211> 239
<212> PRT
<213> Unknown

<220>
<223> Description of Unknown:
Anelloviridae family sequence

<400> 962
Gly Arg Arg Thr Tyr Thr Arg Arg Ala Val Arg Arg Arg Arg Arg Pro
1 5 10 15

Arg Lys Arg Leu Val Leu Thr Gln Trp Ser Pro Gln Thr Val Arg Asn
20 25 30

Cys Ser Ile Arg Gly Ile Val Pro Met Val Ile Cys Gly His Thr Lys
35 40 45

Ala Gly Arg Asn Tyr Ala Ile His Ser Glu Asp Phe Thr Thr Gln Ile
50 55 60

Gln Pro Phe Gly Gly Ser Phe Ser Thr Thr Thr Trp Ser Leu Lys Val
65 70 75 80

Leu Trp Asp Glu His Gln Lys Phe Gln Asn Arg Trp Ser Tyr Pro Asn
85 90 95

Thr Gln Leu Asp Leu Ala Arg Tyr Arg Gly Val Thr Phe Trp Phe Tyr

```
            100                 105                 110

Arg Asp Gln Lys Thr Asp Tyr Ile Val Gln Trp Ser Arg Asn Pro Pro
        115                 120                 125

Phe Lys Leu Asn Lys Tyr Ser Ser Ala Met Tyr His Pro Gly Met Met
    130                 135                 140

Met Gln Ala Lys Arg Lys Leu Val Val Pro Ser Phe Gln Thr Arg Pro
145                 150                 155                 160

Lys Gly Lys Lys Arg Tyr Arg Val Thr Ile Lys Pro Pro Asn Met Phe
                165                 170                 175

Ala Asp Lys Trp Tyr Thr Gln Glu Asp Leu Cys Pro Val Pro Leu Val
            180                 185                 190

Gln Ile Val Val Ser Ala Ala Ser Leu Leu His Pro Phe Cys Pro Pro
        195                 200                 205

Gln Thr Asn Asn Pro Cys Ile Thr Phe Gln Val Leu Lys Asp Ile Tyr
    210                 215                 220

Asp Glu Cys Ile Gly Val Asn Glu Thr Met Lys Asp Lys Tyr Lys
225                 230                 235


<210> 963
<211> 240
<212> PRT
<213> Unknown

<220>
<223> Description of Unknown:
      Anelloviridae family sequence

<400> 963
Arg Arg Arg Arg Arg Arg Gly Arg Arg Arg Arg Arg Arg Arg Arg Arg
1               5                   10                  15

His Lys Pro Thr Leu Ile Leu Arg Gln Trp Gln Pro Asp Cys Ile Arg
```

```
            20                  25                  30

His Cys Lys Ile Thr Gly Trp Met Pro Leu Ile Ile Cys Gly Lys Gly
        35                  40                  45

Ser Thr Gln Phe Asn Tyr Ile Thr His Ala Asp Asp Ile Thr Pro Arg
    50                  55                  60

Gly Ala Ser Tyr Gly Gly Asn Phe Thr Asn Met Thr Phe Ser Leu Glu
65                  70                  75                  80

Ala Ile Tyr Glu Gln Phe Leu Tyr His Arg Asn Arg Trp Ser Ala Ser
                85                  90                  95

Asn His Asp Leu Glu Leu Cys Arg Tyr Lys Gly Thr Thr Leu Lys Leu
            100                 105                 110

Tyr Arg His Pro Glu Val Asp Tyr Ile Val Thr Tyr Ser Arg Thr Gly
        115                 120                 125

Pro Phe Glu Ile Ser His Met Thr Tyr Leu Ser Thr His Pro Met Leu
    130                 135                 140

Met Leu Leu Asn Lys His His Ile Val Val Pro Ser Leu Lys Thr Lys
145                 150                 155                 160

Pro Arg Gly Arg Lys Ala Ile Lys Val Arg Ile Arg Pro Pro Lys Leu
                165                 170                 175

Met Asn Asn Lys Trp Tyr Phe Thr Arg Asp Phe Cys Asn Ile Gly Leu
            180                 185                 190

Phe Gln Leu Trp Ala Thr Gly Leu Glu Leu Arg Asn Pro Trp Leu Arg
        195                 200                 205

Met Ser Thr Leu Ser Pro Cys Ile Gly Phe Asn Val Leu Lys Asn Ser
    210                 215                 220
```

Ile Tyr Thr Asn Leu Ser Asn Leu Pro Gln Tyr Lys Asn Glu Arg Leu
225 230 235 240

<210> 964
<211> 233
<212> PRT
<213> Unknown

<220>
<223> Description of Unknown:
Anelloviridae family sequence

<400> 964
Arg Arg Arg Arg Arg Phe Arg Arg Arg Gly Arg Lys Ala Lys Leu Ile
1 5 10 15

Ile Lys Leu Trp Gln Pro Ala Val Ile Lys Arg Cys Arg Ile Lys Gly
20 25 30

Tyr Ile Pro Leu Ile Ile Ser Gly Asn Gly Thr Phe Ala Thr Asn Phe
35 40 45

Thr Ser His Ile Asn Asp Arg Ile Met Lys Gly Pro Phe Gly Gly Gly
50 55 60

His Ser Thr Met Arg Phe Ser Leu Tyr Ile Leu Phe Glu Glu His Leu
65 70 75 80

Arg His Met Asn Phe Trp Thr Arg Ser Asn Asp Asn Leu Glu Leu Thr
85 90 95

Arg Tyr Leu Gly Ala Ser Val Lys Ile Tyr Arg His Pro Asp Gln Asp
100 105 110

Phe Ile Val Ile Tyr Asn Arg Arg Thr Pro Leu Gly Gly Asn Ile Tyr
115 120 125

Thr Ala Pro Ser Leu His Pro Gly Asn Ala Ile Leu Ala Lys His Lys
130 135 140

```
Ile Leu Val Pro Ser Leu Gln Thr Arg Pro Lys Gly Arg Lys Ala Ile
145                 150                 155                 160


Arg Leu Arg Ile Ala Pro Pro Thr Leu Phe Thr Asp Lys Trp Tyr Phe
                165                 170                 175


Gln Lys Asp Ile Ala Asp Leu Thr Leu Phe Asn Ile Met Ala Val Glu
            180                 185                 190


Ala Asp Leu Arg Phe Pro Phe Cys Ser Pro Gln Thr Asp Asn Thr Cys
        195                 200                 205


Ile Ser Phe Gln Val Leu Ser Ser Val Tyr Asn Asn Tyr Leu Ser Ile
    210                 215                 220


Asn Thr Phe Asn Asn Asp Asn Ser Asp
225                 230


<210> 965
<211> 233
<212> PRT
<213> Unknown

<220>
<223> Description of Unknown:
      Anelloviridae family sequence

<400> 965
Arg Arg Trp Arg Arg Lys Gly Lys Arg Ser Arg Lys Lys Lys Ile Ile
1               5                   10                  15


Ile Arg Gln Trp Gln Pro Asn Tyr Thr Arg Arg Cys Asn Ile Val Gly
            20                  25                  30


Tyr Met Pro Leu Leu Ile Cys Gly Glu Asn Thr Val Ala Thr Asn Tyr
        35                  40                  45


Ala Thr His Ser Asp Asp Ser Tyr Tyr Pro Gly Pro Phe Gly Gly Gly
    50                  55                  60
```

```
Met Thr Thr Asp Lys Phe Thr Leu Arg Ile Leu Tyr Asp Glu Tyr Lys
65                  70                  75                  80

Arg Phe Met Asn Tyr Trp Thr Ser Ser Asn Glu Asp Leu Asp Leu Cys
                85                  90                  95

Arg Tyr Leu Gly Cys Thr Leu Tyr Val Phe Arg His Pro Glu Val Asp
            100                 105                 110

Phe Ile Ile Ile Ile Asn Thr Ser Pro Pro Phe Leu Asp Thr Glu Ile
        115                 120                 125

Thr Gly Pro Ser Ile His Pro Gly Met Met Ala Leu Asn Lys Arg Ser
    130                 135                 140

Arg Trp Ile Pro Ser Ile Lys Asn Arg Pro Gly Arg Lys His Tyr Ile
145                 150                 155                 160

Lys Ile Lys Val Gly Ala Pro Arg Met Phe Thr Asp Lys Trp Tyr Pro
                165                 170                 175

Gln Thr Asp Leu Cys Asp Met Thr Leu Leu Thr Ile Phe Ala Ser Ala
            180                 185                 190

Ala Asp Met Gln Tyr Pro Phe Gly Ser Pro Leu Thr Asp Thr Ile Val
        195                 200                 205

Val Ser Phe Gln Val Leu Gln Ser Met Tyr Asn Asp Cys Leu Ser Val
    210                 215                 220

Leu Pro Asp Asn Phe Ala Glu Thr Ser
225                 230
```

<210> 966
<211> 232
<212> PRT
<213> Unknown

<220>
<223> Description of Unknown:
Anelloviridae family sequence

<400> 966
Arg Arg Trp Lys Arg Lys Gly Arg Arg Arg Arg Lys Ala Lys Ile Ile
1 5 10 15

Ile Arg Gln Trp Gln Pro Asn Tyr Arg Arg Arg Cys Asn Ile Val Gly
20 25 30

Tyr Leu Pro Ile Leu Ile Cys Gly Gly Asn Thr Val Ser Arg Asn Tyr
35 40 45

Ala Thr His Ser Asp Asp Thr Asn Tyr Pro Gly Pro Phe Gly Gly Gly
50 55 60

Met Thr Thr Asp Lys Phe Ser Leu Arg Ile Leu Tyr Asp Glu Tyr Lys
65 70 75 80

Arg Phe Met Asn Tyr Trp Thr Ala Ser Asn Glu Asp Leu Asp Leu Cys
85 90 95

Arg Tyr Leu Gly Cys Thr Phe Tyr Phe Phe Arg His Pro Glu Val Asp
100 105 110

Phe Ile Ile Lys Ile Asn Thr Met Pro Pro Phe Leu Asp Thr Thr Ile
115 120 125

Thr Ala Pro Ser Ile His Pro Gly Leu Met Ala Leu Asp Lys Arg Ala
130 135 140

Arg Trp Ile Pro Ser Leu Lys Asn Arg Pro Gly Lys Lys His Tyr Ile
145 150 155 160

Lys Ile Arg Val Gly Ala Pro Lys Met Phe Thr Asp Lys Trp Tyr Pro
165 170 175

Gln Thr Asp Leu Cys Asp Met Thr Leu Leu Thr Ile Tyr Ala Thr Ala

```
            180                 185                 190

Ala Asp Met Gln Tyr Pro Phe Gly Ser Pro Leu Thr Asp Thr Val Val
        195                 200                 205

Val Asn Ser Gln Val Leu Gln Ser Met Tyr Asp Glu Thr Ile Ser Ile
    210                 215                 220

Leu Pro Asp Glu Lys Thr Lys Arg
225                 230
```

<210> 967
<211> 170
<212> PRT
<213> Beak and feather disease virus

<400> 967

```
Arg Arg Arg Arg Arg Arg Phe Ser Thr Asn Arg Ile Tyr Thr Leu Arg
1               5                   10                  15

Leu Thr Arg Gln Phe Gln Phe Lys Ile Asn Lys Gln Leu Ile Phe Asn
            20                  25                  30

Ala Asp Tyr Ile Thr Phe Ala Leu Asp Asp Phe Leu Gln Ala Val Pro
        35                  40                  45

Asn Pro His Thr Leu Asn Phe Glu Asp Tyr Arg Ile Lys Leu Ala Lys
    50                  55                  60

Met Glu Met Arg Pro Thr Gly Gly His Tyr Thr Gly Phe Gly His Thr
65                  70                  75                  80

Ala Val Ile Gln Asp Ser Arg Ile Thr Arg Phe Lys Thr Thr Ala Asp
                85                  90                  95

Pro Leu Ala Pro Phe Asp Gly Ala Lys Lys Trp Phe Val Ser Arg Gly
            100                 105                 110

Phe Lys Arg Leu Leu Arg Pro Lys Pro Gln Ile Thr Ile Glu Gly Pro
```

```
        115                 120                 125

Asn Ser Ala Gly Thr Lys Val Arg His Tyr Gly Ile Ala Phe Ser Phe
    130                 135                 140

Pro Gln Pro Glu Gln Thr Val Thr Lys Leu Thr Leu Tyr Val Gln Phe
145                 150                 155                 160

Arg Gln Phe Ala Pro Asn Asn Pro Ser Thr
                165                 170
```

<210> 968
<211> 179
<212> PRT
<213> Orthohepevirus A

<400> 968

```
Gly Ala Ile Leu Arg Arg Gln Tyr Asn Leu Ser Thr Ser Pro Leu Thr
1               5                   10                  15

Ser Ser Val Ala Ser Gly Thr Asn Leu Val Leu Tyr Ala Ala Pro Leu
            20                  25                  30

Asn Pro Leu Leu Pro Leu Gln Asp Gly Thr Asn Thr His Ile Met Ala
        35                  40                  45

Thr Glu Ala Ser Asn Tyr Ala Gln Tyr Arg Val Val Arg Ala Thr Ile
    50                  55                  60

Arg Tyr Arg Pro Leu Val Pro Asn Ala Val Ser Ile Ser Phe Trp Pro
65                  70                  75                  80

Gln Thr Thr Thr Thr Pro Thr Ser Val Asp Met Asn Ser Ile Thr Ser
                85                  90                  95

Thr Asp Val Arg Ile Leu Val Gln Pro Gly Ile Ala Ser Glu Leu Val
            100                 105                 110

Ile Pro Ser Glu Arg Leu His Tyr Arg Thr Arg Thr Gly Val Ala Glu
```

```
        115                 120                 125

Glu Glu Ala Thr Ser Gly Leu Val Met Leu Cys Ile His Gly Ser Pro
    130                 135                 140

Val Asn Ser Ala Leu Gly Leu Leu Asp Phe Ala Leu Glu Leu Glu Phe
145                 150                 155                 160

Arg Asn Leu Thr Pro Gly Asn Thr Asn Thr Arg Val Ser Arg Tyr Thr
                165                 170                 175

Ser Thr Ala
```