

Miniaturized Dystrophins and Uses Thereof

REFERENCE TO RELATED APPLICATIONS

This application claims the priority benefit of U.S. Provisional Application No. 62/749982, filed October 24, 2018, which is incorporated herein by reference in its entirety.

FIELD

The presently disclosed subject matter generally relates to polynucleotides, polypeptides, cells, vectors, uses, and kits relating to miniaturized dystrophin.

BACKGROUND OF THE DISCLOSURE

Duchenne muscular dystrophy (DMD) is a recessively-inherited muscle wasting disorder afflicting approximately 1 in 3,500 males. DMD is caused by mutations in the dystrophin gene, which is located on the X chromosome. Mutations in this gene lead to aberrant or absent expression of the dystrophin protein.

Dystrophin is a key component of a protein complex that is responsible for regulating muscle cell integrity and function. DMD patients typically lose the ability to physically support themselves during childhood and become progressively weaker over time. This progressive wasting of skeletal muscles and cardiac dysfunction typically leads to loss of ambulation and premature death, primarily due to cardiac or respiratory failure.

Some attempts have been made in the past to treat DMD. However, the available treatment options were significantly limited due to the large size of the wild type dystrophin cDNA (approximately 13.9 kb) which cannot be administered to and expressed in DMD patients using standard viral vectors, including Adeno-associated virus (AAV), which cannot transfer more than approximately 4.9 kb of heterologous DNA. Therefore, there is a need to develop a recombinant dystrophin gene that can be efficiently packaged into a vector for gene therapy.

Adeno-associated viral (AAV) vectors have been shown to be useful in gene therapeutic approaches aimed at correcting genetic deficiencies that result in reduced or completely abolished levels of protein expression (Nathwani *et al* to A.M Keeler *et al.*), and are potentially useful for gene knockdown, genome editing or modification, and non-coding RNA modulation (Valdmanis *et al.*, 2017).

Packaging the entire cDNA of the muscle-specific isoform of dystrophin into a single rAAV capsid cannot be achieved easily because of the large size of the dystrophin cDNA. Previous studies have focused on the development of smaller genetic constructs that express only particular domains of dystrophin. *See* U.S. Patent Nos. 6,869,777 and 8,501,920, each of which is incorporated by reference. However, these approaches have had only limited success.

There remains a need for more precise and efficient gene therapy tools for treating patients with mutations in the dystrophin gene, and, in particular, a need to develop a recombinant dystrophin gene that can be efficiently packaged into a vector for gene therapy.

SUMMARY OF THE DISCLOSURE

The present disclosure provides a nucleic acid molecule comprising a nucleotide sequence, which encodes a miniaturized dystrophin polypeptide comprising from N terminus to C terminus hinge 1 (H1) domain, spectrin repeat 1 (R1) domain, spectrin repeat 3 (R3) domain, hinge 2 (H2) domain, spectrin repeat 16 (R16) domain, spectrin repeat 17 (R17) domain, spectrin repeat 24 (R24) domain, and hinge 4 (H4) domain of dystrophin, wherein the miniaturized dystrophin polypeptide does not comprise spectrin repeat 2 of dystrophin. In some embodiments, the miniaturized dystrophin polypeptide does not comprise spectrin repeat 2 (R2) domain, spectrin repeat 4 (R4) domain, spectrin repeat 5 (R5) domain, spectrin repeat 6 (R6) domain, spectrin repeat 7 (R7) domain, spectrin repeat 8 (R8) domain, spectrin repeat 9 (R9) domain, spectrin repeat 10 (R10) domain, spectrin repeat 11 (R11) domain, spectrin repeat 12 (R12) domain, spectrin repeat 13 (R13) domain, spectrin repeat 14 (R14) domain, spectrin repeat 15 (R15) domain, spectrin repeat 18 (R18) domain, spectrin repeat 19 (R19) domain, spectrin repeat 20 (R20) domain, spectrin repeat 21 (R21) domain, spectrin repeat 22 (R22) domain, or spectrin repeat 23 (R23) domain, or any combination thereof. In some embodiments, the R1 domain is directly fused to R3 domain by a peptide bond. In some embodiments, the R1 domain and the R3 domain are fused by amino acids ARG-VAL (RV). In some embodiments, the H2 domain and the R16 domain are fused by a linker.

In some embodiments, the linker comprises an amino acid sequence at least about 80%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 75 (IHTVREE

TMMVMTEDMP LEI), wherein the amino acid sequence is capable of improving nNOS signaling.

In some embodiments, disclosed herein is a nucleic acid molecule comprising a nucleotide sequence which encodes an amino acid sequence comprising SEQ ID NO: 75 (IHTVREE
5 TMMVMTEDMP LEI), wherein the nucleic acid molecule is at least about 80%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 100 (ATCCACACCGTGC GGGAAGAGACAATGATGGTCATGACAGAGGAC

ATGCCCCTGGAAATC), wherein the amino acid sequence is capable of improving
10 nNOS signaling. In some embodiments, the amino acid sequence is a linker connecting a first domain of dystrophin to a second domain of dystrophin. In some embodiments, the first domain of dystrophin is a H2 domain and the second domain of dystrophin is an R16 domain. In some embodiments, the amino acid sequence further comprises the sequence of SEQ ID NO: 74 at the N terminus.

15 In some embodiments, the nucleic acid molecule encodes a miniaturized dystrophin polypeptide. In some embodiments, the miniaturized dystrophin polypeptide comprises from N terminus to C terminus hinge 1 (H1) domain, spectrin repeat 1 (R1) domain, spectrin repeat 3 (R3) domain, hinge 2 (H2) domain, spectrin repeat 16 (R16) domain, spectrin repeat 17 (R17) domain, spectrin repeat 24 (R24) domain, and hinge 4 (H4) domain
20 of dystrophin. In some embodiments, the R1 domain and the R3 domain are fused by amino acids ARG-VAL (RV) and wherein the H2 domain and the R16 domain are fused by the amino acid sequence set forth in SEQ ID NOs: 74-75 in combination. In some embodiments, (i) the H1 domain and the R1 domain are fused directly, (ii) the R3 domain and the H2 domain are fused directly, (iii) the R16 and R17 domains are fused directly, (iv)
25 the R17 and R24 domains are fused directly, or (v) the R24 and H4 domains are fused directly, or (vi) any combination thereof. In some embodiments, the miniaturized dystrophin polypeptide further comprises ABD1 domain and/or CR domain. In some embodiments, the miniaturized dystrophin polypeptide consists essentially of or consists of, from N terminus to C terminus, the ABD1 domain, the H1 domain, the R1 domain,
30 amino acids RV, the R3 domain, the H2 domain, the amino acid sequence as set forth in SEQ ID NOs: 74-75, the R16 domain, the R17 domain, the R24 domain, the H4 domain, and the CR domain of dystrophin.

In some embodiments, the H1 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 69. In some embodiments, the R1 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 70. In some embodiments, the R3 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 72. In some embodiments, the H2 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 73. In some embodiments, the R16 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 76. In some embodiments, the R17 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 77. In some embodiments, the R24 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 78. In some embodiments, the H4 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 79. In some embodiments, the miniaturized dystrophin polypeptide further comprises at the N terminus an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 68. In some embodiments, the miniaturized dystrophin polypeptide further comprises at the C terminus an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 80. In some embodiments, the miniaturized dystrophin polypeptide comprises an

amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 118. In some embodiments, the nucleotide sequence comprises a sequence at least about 60%, at least about 65%, at least about 70%,
5 at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to the nucleotide sequence of SEQ ID NO: 133.

In some embodiments, the miniaturized dystrophin polypeptide exhibits one or more properties selected from the group consisting of (i) having a lower CD4 proliferation
10 compared to BXA-027741, (ii) having a lower CD8 proliferation compared to BXA-027741, (iii) having a higher expression of the miniaturized dystrophin polypeptide than BXA-027741, and (iv) any combination thereof.

In some embodiments, the nucleic acid molecule comprises a nucleotide sequence, which encodes a miniaturized dystrophin polypeptide comprising spectrin repeat 1 (R1) domain
15 and spectrin 16 (R16) domain, wherein the R1 domain and the R16 domain are fused by a linker comprising the amino acid sequence set forth in SEQ ID NO: 84 (IHTVREETMMVMTEDEMPLEI).

In some embodiments, the nucleic acid molecule comprises a nucleotide sequence, which encodes a miniaturized dystrophin polypeptide comprising from N terminus to C terminus
20 a hinge 1 (H1) domain, spectrin repeat 1 (R1) domain, spectrin repeat 16 (R16) domain, spectrin repeat 17 (R17) domain, a hinge 3 (H3) domain, spectrin repeat 23 (R23) domain, spectrin repeat 24 (R24) domain, and hinge 4 (H4) domain of dystrophin, wherein the R1 domain and the R16 domain are fused by the amino acid sequence set forth in SEQ ID NO: 84 (IHTVREETMMVMTEDEMPLEI).

25 In some embodiments, (i) the H1 domain and the R1 domain are fused directly, (ii) the R16 and R17 domains are fused directly, (iii) the R17 and the H3 domains are fused directly, (iv) the R23 and R24 domains are fused directly, or (v) the R24 and H4 domains are fused directly, or (vi) any combination thereof.

In some embodiments, the miniaturized dystrophin polypeptide does not comprise a
30 spectrin repeat 2 (R2) domain, spectrin repeat 3 (R3) domain, spectrin repeat 4 (R4) domain, spectrin repeat 5 (R5) domain, spectrin repeat 6 (R6) domain, spectrin repeat 7 (R7) domain, spectrin repeat 8 (R8) domain, spectrin repeat 9 (R9) domain, spectrin repeat

10 (R10) domain, spectrin repeat 11 (R11) domain, spectrin repeat 12 (R12) domain, spectrin repeat 13 (R13) domain, spectrin repeat 14 (R14) domain, spectrin repeat 15 (R15) domain, spectrin repeat 18 (R18) domain, spectrin repeat 19 (R19) domain, spectrin repeat 20 (R20) domain, spectrin repeat 21 (R21) domain, and/or spectrin repeat 22 (R22) domain.

5 In some embodiments, the miniaturized dystrophin polypeptide further comprises an ABD1 domain and/or a CR domain. In some embodiments, the miniaturized dystrophin polypeptide consists essentially of or consists of, from N terminus to C terminus, the ABD1 domain, the H1 domain, the R1 domain, the amino acid sequence set forth in SEQ ID NO: 84, the R16 domain, the R17 domain, the H3 domain, the R23 domain, the R24 domain, 10 the H4 domain, and the CR domain of dystrophin. In some embodiments, the H1 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 82. In some embodiments, the R1 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least 15 about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 83. In some embodiments, the R16 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 85. In some embodiments, the R17 domain is an 20 amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 86. In some embodiments, the H3 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, 25 or about 100% identical to SEQ ID NO: 87. In some embodiments, the R23 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 88. In some embodiments, the R24 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least 30 about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 89. In some embodiments, the H4 domain is an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least

about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 90.

In some embodiments, the miniaturized dystrophin polypeptide further comprises at the N terminus an amino acid sequence at least about 80%, at least about 85%, at least about 90%,

5 at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 81. In some embodiments, the miniaturized

dystrophin polypeptide further comprises at the C terminus an amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ

10 ID NO: 91. In some embodiments, the miniaturized dystrophin polypeptide comprises an

amino acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 132, or to SEQ ID NO: 132 wherein the C-terminal

domain is deleted. In some embodiments, the nucleotide sequence comprises a sequence at

15 least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about

80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to the nucleotide sequence of SEQ ID NO: 147 or SEQ ID NO: 148 or SEQ ID NO: 149.

In some embodiments, the miniaturized dystrophin polypeptide exhibits a higher
20 expression of the miniaturized dystrophin polypeptide than BXA-196481.

In some embodiments, the miniaturized dystrophin polypeptide expression is at least about 1.5 fold, at least about 1.6 fold, at least about 1.7 fold, at least about 1.8 fold, at least about

1.9 fold, at least about 2 fold, at least about 2.1 fold, at least about 2.2 fold, at least about 2.3 fold, at least about 2.4 fold, at least about 2.5 fold, at least about 2.6 fold, at least about

25 2.7 fold, at least about 2.8 fold, at least about 2.9 fold or at least about 3 fold higher than the BXA-196481 polypeptide expression.

In some embodiments, the nucleic acid molecule further comprises a promoter. In some embodiments, the promoter is a tissue-specific promoter. In some embodiments, the promoter drives expression of the therapeutic protein in hepatocytes, muscle cells,

30 endothelial cells, neuronal cells, sinusoidal cells, or any combination thereof. In some

embodiments, the promoter is selected from the group consisting of a mouse thyretin promoter (mTTR), an endogenous human factor VIII promoter (F8), a human alpha-1-

antitrypsin promoter (hAAT), a human albumin minimal promoter, a mouse albumin promoter, a tristetraproline (TTP) promoter, a CASI promoter, a synapsin 1 gene promoter, a CAG promoter, a cytomegalovirus (CMV) promoter, α 1-antitrypsin (AAT), muscle creatine kinase (MCK), myosin heavy chain alpha (α MHC), myoglobin (MB), desmin (DES), SPc5-12, 2R5Sc5-12, dMCK, tMCK, and a phosphoglycerate kinase (PGK) promoter.

In some embodiments, the nucleic acid molecule further comprises an intronic sequence. In some embodiments, the intronic sequence is positioned 5' to the nucleotide sequence encoding the miniaturized dystrophin polypeptide. In some embodiments, the intronic sequence is positioned 3' to the promoter. In some embodiments, the intronic sequence comprises a synthetic intronic sequence.

In some embodiments, the nucleic acid molecule further comprises a post-transcriptional regulatory element. In some embodiments, the post-transcriptional regulatory element is positioned 3' to the nucleotide sequence encoding the miniaturized dystrophin polypeptide. In some embodiments, the post-transcriptional regulatory element comprises a mutated woodchuck hepatitis virus post-transcriptional regulatory element (WPRE), a microRNA binding site, or a DNA nuclear targeting sequence, or any combination thereof.

In some embodiments, the nucleic acid molecule further comprises a 3'UTR poly(A) tail sequence. In some embodiments, the 3'UTR poly(A) tail sequence is selected from the group consisting of bGH poly(A), actin poly(A), hemoglobin poly(A), and any combination thereof. In some embodiments, the 3'UTR poly(A) tail sequence comprises bGH poly(A).

In some embodiments, the nucleic acid molecule further comprises an enhancer sequence. In some embodiments, the nucleic acid molecule further comprises a first ITR and/or a second ITR. In some embodiments, the first ITR and the second ITR are identical. In some embodiments, the first ITR and/or the second ITR are derived from adeno-associated virus.

In some embodiments, the nucleic acid molecule comprises a heterologous moiety. In some embodiments, the heterologous moiety is selected from the group consisting of albumin or a fragment thereof, an immunoglobulin Fc region, the C-terminal peptide (CTP) of the β subunit of human chorionic gonadotropin, a PAS sequence, a HAP sequence, a transferrin or a fragment thereof, an albumin-binding moiety or a derivative thereof, and any combination thereof.

In some embodiments, provided is a vector comprising a nucleic acid molecule disclosed herein. In some embodiments, the vector is selected from the group consisting of an adenoviral vector, a retroviral vector, a poxvirus vector, a baculovirus vector, a herpes viral vector. In some embodiments, the vector is an adeno-associated virus (AAV) vector. In some embodiments, the AAV vector is selected from AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, and AAV11. In some embodiments, the AAV vector is AAV9.

In some embodiments, the nucleic acid molecule or vector disclosed herein is formulated with a delivery agent. In some embodiments, the delivery agent comprises a lipid nanoparticle. In some embodiments, the delivery agent is selected from the group consisting of liposomes, non-lipid polymeric molecules, endosomes, and any combination thereof.

In some embodiments, the nucleic acid molecule or the vector disclosed herein is formulated for intravenous, transdermal, intradermal, subcutaneous, pulmonary, or oral delivery, or any combination thereof. In some embodiments, the nucleic acid molecule or vector disclosed herein is formulated for intravenous delivery.

In some embodiments, provided is a polypeptide encoded by the nucleic acid molecule or vector disclosed herein.

In some embodiments, provided is a host cell comprising the nucleic acid molecule disclosed herein. In some embodiments, the cell is a CHO cell, a HEK293 cell, a HBK cell, a COS cell, a NSO cell, or a HT1080 cell.

In some embodiments, provided is a pharmaceutical composition comprising (a) the nucleic acid disclosed herein, the vector disclosed herein, the polypeptide disclosed herein, or the host cell disclosed herein; and (b) a pharmaceutically acceptable excipient.

In some embodiments, provided is a kit, comprising the nucleic acid disclosed herein, the vector disclosed herein, the polypeptide disclosed herein, the host cell disclosed herein, or the pharmaceutical composition disclosed herein, and instructions for administering the nucleic acid molecule to a subject in need thereof.

In some embodiments, provided is a method of producing a miniaturized dystrophin polypeptide, comprising: culturing the host cell disclosed herein under suitable conditions and recovering the miniaturized dystrophin polypeptide.

In some embodiments, provided is a method of expressing a miniaturized dystrophin polypeptide in a subject in need thereof, comprising administering to the subject the nucleic acid disclosed herein, the vector disclosed herein, the host cell disclosed herein, or the pharmaceutical composition disclosed herein.

- 5 In some embodiments, provided is a method of treating a subject having a disease or condition comprising administering to the subject the nucleic acid disclosed herein, the vector disclosed herein, the polypeptide disclosed herein, the host cell disclosed herein, or the pharmaceutical composition disclosed herein. In some embodiments, the disease or condition is a disease caused by dystrophin deficiency. In some embodiments, the disease
- 10 is Sarcopenia, a heart disease, cachexia, Duchenne muscular dystrophy (DMD), Becker muscular dystrophy (BMD), X-linked dilated cardiomyopathy (XLDC), facioscapulohumeral muscular dystrophy, myotonic muscular dystrophy, limb-girdle muscular dystrophy, oculopharyngeal muscular dystrophy, Emery-Dreifuss muscular dystrophy, distal muscular dystrophy, and/or congenital muscular dystrophy. In some
- 15 embodiments, the method of any one of claims 86 to 89, wherein the nucleic acid molecule, the vector, the polypeptide, the host cell, or the pharmaceutical composition is administered intravenously, transdermally, intradermally, subcutaneously, orally, or pulmonarily, or any combination thereof. In some embodiments, the nucleic acid molecule, the vector, the polypeptide, the host cell, or the pharmaceutical composition is administered intravenously,
- 20 transdermally, intradermally, subcutaneously, orally, or pulmonarily, or any combination thereof.

In some embodiments, the method further comprises administering to the subject a second agent.

- In some embodiments, the subject is a human. In some embodiments, the administration of
- 25 the nucleic acid molecule, the vector, the polypeptide, the host cell, or the pharmaceutical composition to the subject results in increased dystrophin protein expression, relative to dystrophin protein expression in the subject prior to the administration, wherein the dystrophin protein expression is increased by at least about 2-fold, at least about 3-fold, at least about 4-fold, at least about 5-fold, at least about 6-fold, at least about 7-fold, at least
- 30 about 8-fold, at least about 9-fold, at least about 10-fold, at least about 11-fold, at least about 12-fold, at least about 13-fold, at least about 14-fold, at least about 15-fold, at least about 20-fold, at least about 25-fold, at least about 30-fold, at least about 35-fold, at least

about 40-fold, at least about 50-fold, at least about 60-fold, at least about 70-fold, at least about 80-fold, at least about 90-fold, or at least about 100-fold.

In some embodiments, disclosed herein is a nucleic acid molecule comprising a nucleotide sequence, which encodes a miniaturized dystrophin polypeptide comprising hinge 2 (H2) domain and spectrin 16 (R16) domain, wherein the H2 domain and the R16 domain are fused by a linker comprising the amino acid sequence set forth in SEQ ID NO: 84 (IHTVREETMMVMTE DMPL EI).

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 shows a schematic diagram of the full length human Dystrophin protein. ABD1: actin-binding domain-1; H# (e.g., H1): hinge region; R# (e.g., R1): spectrin-like repeat domains; ABD2: actin-binding domain-2; CR: cysteine-rich domain; C-term: C-terminal domain of the protein.

FIG. 2 shows schematic diagrams of miniaturized dystrophin polypeptides BXA-027741, BXA-027742, BXA-027743, BXA-027744, BXA-196480 and BXA-196481, as well as, as a point of reference, the full length human Dystrophin protein of FIG. 1.

FIG. 3 shows the ratio of miniaturized dystrophin polypeptide expression to miniaturized dystrophin mRNA expression in human isogenic induced-pluripotent stem cell (iPSC)-derived cardiac myocytes (iCMs) (carrying an E2035X premature stop codon in the dystrophin gene that prevented endogenous dystrophin expression) after transfection of plasmids expressing the indicated miniaturized dystrophin polypeptides.

FIG. 4 shows a schematic diagram of the miniaturized dystrophin polypeptide of BXA-027743 (SEQ ID NO: 132), as well as the amino acid sequence of its junctions J1 and J7, which have relatively low immunogenicity, as indicated.

FIG. 5 shows a schematic diagram of the miniaturized dystrophin polypeptide of BXA-027741 (SEQ ID NO: 129) with arrows pointing to its junctions J10, J 11 and J9.

FIG. 6A shows a histogram indicating the proportion of samples, among the 40-samples cell panel tested, that were pulsed with various junction peptides as indicated and had CD4+ proliferating cells (each colored square represents one patient sample). FIG. 6B shows a histogram indicating the proportion of samples, among the 40-samples cell panel tested, that were pulsed with various junction peptides as indicated and had CD8+ proliferating cells (each colored square represents one patient sample).

FIG. 7 shows a schematic diagram of the miniaturized dystrophin polypeptide of BXA-196473 (SEQ ID NO:119).

FIG. 8A shows a schematic diagram of the miniaturized dystrophin polypeptide of BXA-196477 (SEQ ID NO:118). RV: arginine-valine dipeptide between R1 and R3; SEAQ: serine-glutamic acid-alanine-glutamine peptide. FIG. 8B shows the three-dimensional folding around original junction 10 between R1 and R3 of miniaturized dystrophin polypeptide BXA-027741 (SEQ ID NO: 129) (left), and the three-dimensional folding around version 3 of the junction 10 of miniaturized dystrophin polypeptide BXA-196477 (SEQ ID NO: 118) (right).

FIG. 9 shows the ratio of miniaturized dystrophin polypeptide expression to miniaturized dystrophin mRNA in human isogenic induced-pluripotent stem cell (iPSC)-derived cardiac myocytes (iCMs) (carrying an E2035X premature stop codon in the dystrophin gene that prevented endogenous dystrophin expression) after transfection of plasmids expressing the indicated miniaturized dystrophin polypeptides.

FIG. 10 shows a schematic diagram of the miniaturized dystrophin polypeptide of BXA-196474 (SEQ ID NO:120).

FIG. 11 shows a schematic diagram of the miniaturized dystrophin polypeptide of BXA-196475 (SEQ ID NO:121).

FIG. 12 shows a schematic diagram of the miniaturized dystrophin polypeptide of BXA-196476 (SEQ ID NO:122).

FIG. 13 shows a schematic diagram of the miniaturized dystrophin polypeptide of BXA-196478 (SEQ ID NO:124).

FIG. 14 shows a schematic diagram of the miniaturized dystrophin polypeptide of BXA-196479 (SEQ ID NO:125).

FIG. 15 Conduction velocity in human DMD hiPSC CMs infected with AAV9-constructs expressing miniaturized dystrophin polypeptides of BXA-196477 and BXA-213788. FIG. 15A shows a photograph of a Multiple Electrode Array (MEA) used in the experiments (left) and a photograph of individual electrodes of the MEA (right). FIG. 15B shows the signal from the 28 electrodes of the MEA under experimental conditions. FIG. 15C shows a graphic illustrating impulse conduction across the array. FIG. 15D shows schematic diagrams of the miniaturized dystrophin polypeptides of BXA-196477 (SEQ ID NO: 118) and BXA-213788 (SEQ ID NO:152). FIG. 15E shows a graphic wherein the conduction

velocity of human DMD hiPSC CMs expressing miniaturized dystrophin polypeptides of BXA-196477 (SEQ ID NO: 118) and BXA-213788 is plotted as a function of time post transfection. FIG. 15F shows a histogram indicating the expression of miniaturized dystrophin polypeptides in cells in which conduction velocity was measured.

5 FIG. 16 shows a schematic summarizing the experimental set-up of in vivo studies that were conducted.

FIG. 17 Target engagement of AAV9-BXA-196477 or AAV9-BXA-213788. FIG. 17B shows immuno-fluorescence visualization of expression of miniaturized dystrophin polypeptides and wheat germ-agglutinin (WGA) in muscle tissue of mice treated with
10 AAV9-BXA-196477 or AAV9-BXA-213788. FIG. 17A shows histograms indicating the relative number of cells in various muscles positive for miniaturized dystrophin. Bars with diagonal pattern = % BXA-196477 miniaturized Dystrophin positive cells as a proportion of WGA positive cells; bars with longitudinal pattern = % BXA-213788 miniaturized Dystrophin as a proportion of WGA positive cells.

15 FIG. 18 Target engagement of AAV9-BXA-196477 or AAV9-BXA-213788 determined in *mdx*^{4cv} mice at 4-weeks of age. FIG. 18A shows schematic diagrams of the miniaturized dystrophin polypeptides of BXA-196477 (SEQ ID NO: 118) and BXA-213788 (SEQ ID NO:152). FIG. 18B shows a histogram indicating the relative amount of viral genomes in muscle tissue of *mdx*^{4cv} mice treated with AAV9-BXA-196477 or AAV9-BXA-213788.

20 FIG. 18C shows a histogram indicating the relative amount of miniaturized dystrophin mRNA in the heart of *mdx*^{4cv} mice treated with AAV9-BXA-196477 or AAV9-BXA-213788. FIG. 18D shows a histogram indicating the relative amount of miniaturized dystrophin protein in muscle tissue of mice treated with AAV9-BXA-196477 or AAV9-BXA-213788. Dia = diaphragm; Gas = gastrocnemius; TA = tibialis anterior. Treated
25 mice were compared to wildtype and *mdx*^{4cv} non-treated mice.

FIG. 19A shows a histogram indicating the amount of creatine kinase in the serum of *mdx*^{4cv} mice (4-weeks) treated with AAV9-BXA-196477 or AAV9-BXA-213788. FIG. 19B shows a histogram indicating the amount of creatine kinase in the serum of *mdx*^{4cv} mice (12-weeks) treated with AAV9-BXA-196477 or AAV9-BXA-213788. Treated mice were
30 compared to wildtype and *mdx*^{4cv} non-treated mice.

FIG. 20 Peak tetanic contraction at 2-months of age. FIG. 20A shows a graph indicating contractile properties (Force in Newton meters; Nm) in mice *mdx*^{4cv} treated with AAV9-

BXA-196477 or AAV9-BXA-213788. FIG. 20B shows a histogram indicating peak tetanic contraction in treated mice. FIG. 20C shows a histogram indicating TA mass in treated mice. FIG. 20D shows a histogram indicating force/TA mass in treated mice. Treated mice were compared to wildtype and *mdx*^{4cv} non-treated mice.

5 FIG. 21 Protection from contraction-induced injury. FIG. 21A shows a graphic outlining the experimental setup. FIG. 21B shows a graph indicating contractile properties (maximum isometric torque) of muscles in mice treated with AAV9-BXA-196477 or AAV9-BXA-213788. Treated mice were compared to wildtype and *mdx*^{4cv} non-treated mice.

10 FIG. 22A shows a schematic diagram of the miniaturized dystrophin polypeptides of BXA-027741 (SEQ ID NO: 129). FIG. 22B shows a histogram indicating immunogenicity of junction peptides 1 to 12.

FIG. 23A shows a histogram indicating the MHC I/CD8+ immune response to junction peptides. Peptide J11v3 here is the same peptide as peptide no. 3 in FIG. 22B.

15 FIG. 23B shows a histogram indicating the MHC II/CD4+ immune response to junction peptides. Peptide J11v3 here is the same peptide as peptide no. 3 in FIG. 22B.

DETAILED DESCRIPTION OF THE DISCLOSURE

Overview

The present disclosure relates to novel miniaturized dystrophins or the genes encoding the same. The miniaturized dystrophins can be operatively linked to a regulatory cassette. The present disclosure also relates to methods of treating a subject having muscular dystrophy, sarcopenia, heart failure, or cachexia. Further, the present disclosure relates to methods of prophylactically treating a subject at risk of developing muscular dystrophy, sarcopenia, heart failure, or cachexia. The methods for treating a subject having, or at risk of developing, muscular dystrophy, sarcopenia, heart failure, or cachexia can comprise administering a pharmaceutical composition including a miniaturized dystrophin gene and a delivery vehicle to the subject.

Definitions

In order that the present disclosure can be more readily understood, certain terms are first defined. As used in this application, except as otherwise expressly provided herein, each of

the following terms shall have the meaning set forth below. Additional definitions are set forth throughout the application.

The term "and/or" where used herein is to be taken as specific disclosure of each of the two specified features or components with or without the other. Thus, the term "and/or" as used in a phrase such as "A and/or B" herein is intended to include "A and B," "A or B," "A" (alone), and "B" (alone). Likewise, the term "and/or" as used in a phrase such as "A, B, and/or C" is intended to encompass each of the following aspects: A, B, and C; A, B, or C; A or C; A or B; B or C; A and C; A and B; B and C; A (alone); B (alone); and C (alone).

It is understood that wherever aspects are described herein with the language "comprising," otherwise analogous aspects described in terms of "consisting of" and/or "consisting essentially of" are also provided.

Unless defined otherwise, 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 disclosure is related. For example, the Concise Dictionary of Biomedicine and Molecular Biology, Juo, Pei-Show, 2nd ed., 2002, CRC Press; The Dictionary of Cell and Molecular Biology, 3rd ed., 1999, Academic Press; and the Oxford Dictionary Of Biochemistry And Molecular Biology, Revised, 2000, Oxford University Press, provide one of skill with a general dictionary of many of the terms used in this disclosure.

Units, prefixes, and symbols are denoted in their Système International de Unites (SI) accepted form. Numeric ranges are inclusive of the numbers defining the range. The headings provided herein are not limitations of the various aspects of the disclosure, which can be had by reference to the specification as a whole. Accordingly, the terms defined immediately below are more fully defined by reference to the specification in its entirety.

Dystrophin (DMD) is a large human X-linked gene that encodes Dystrophin. The protein Dystrophin is a 427 kDa cytoskeletal protein that localizes to the cytoplasmic face of the sarcolemma and is enriched at costameres in muscle fibers. The Dystrophin protein has four main functional domains: an actin-binding amino-terminal domain (ABD1); a central rod domain comprising a series of rods, called "spectrin repeat domains" and hinges; a cysteine-rich domain; and a carboxyl-terminus.

As used herein, the term "miniaturized dystrophin polypeptide" or "miniaturized dystrophin peptide" refers to a polypeptide that is smaller in size than the full-length wild-type dystrophin polypeptide. In some embodiments, the miniaturized dystrophin polypeptide is

capable of altering (increasing or decreasing, as the case may be) a measurable value of muscle physiology or anatomy in a DMD animal model by at least approximately 10 or 20% of the wild type value, such that the value is closer to the wild-type value (e.g., a mdx mouse has a measurable value of muscle physiology or anatomy that is 50% of the wild-type value, and this value is increased to at least 60% of the wild-type value; or a mdx mouse has a measurable value of muscle physiology or anatomy that is 150% of the wild-type value, and this value is decreased to at most 140% of the wild-type value). In certain embodiments, the miniaturized dystrophin polypeptide is capable of altering a measurable value of muscle physiology or anatomy in a DMD animal model by at least approximately 30% of the wild type value. In some embodiments, the miniaturized dystrophin polypeptide is capable of altering a measurable value of muscle physiology or anatomy in a DMD animal model to a level similar to the wild-type value (e.g., $\pm 4\%$).

As used herein, the term "spectrin repeats" or "spectrin-like repeats" refers to peptides composed of approximately 100 amino acids that are responsible for the rod-like shape of many structural proteins including, but not limited to, dystrophin, wherein the spectrin repeats are typically present in multiple copies. Spectrin repeats can include mutations of the natural peptide sequences, such as conservative and/or non-conservative changes in amino acid sequence, as well as the addition or deletion of up to 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acids to/from the end of a spectrin repeat or within the spectrin repeat. In some embodiments, each spectrin repeat (each of R1 to R24) has at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to the naturally occurring spectrin repeat (each of the naturally occurring R1 to R24).

As used herein, the term "spectrin repeat encoding sequences" refers to nucleic acid sequences encoding spectrin repeat peptides. This term includes natural and synthetic nucleic acid sequences encoding the spectrin repeats (e.g., both the naturally occurring and mutated spectrin repeat peptides).

As used herein, the term "spectrin repeat domain" refers to the region in a miniaturized dystrophin polypeptide that contains the spectrin repeats of the miniaturized dystrophin polypeptide.

The term "fused" refers to a first amino acid sequence that is linked in frame to a second amino acid sequence with which it is not normally linked in nature, forming a "fusion"

protein/polypeptide. These fused amino acid sequences which normally exist in separate proteins can be brought together in the fusion polypeptide, or the amino acid sequences which normally exist in the same protein can be placed in a new arrangement in the fusion polypeptide. A fusion protein is created, for example, by chemical peptide synthesis, or by
5 recombinant DNA technology whereby a polynucleotide is created, and then translated, in which the peptide regions are encoded in the desired relationship. A fusion protein can also comprise a second amino acid sequence associated with the first amino acid sequence by a covalent, non-peptide bond or by a non-covalent bond. In some embodiments, "fusion" between two polypeptides is achieved by a linker. Linkers can be amino acids or other
10 chemical structures. In some embodiments, linkers can be synthetic.

In some embodiments, "fusion" between two polypeptides is a direct fusion, i.e., without intervening linker. The term "fused directly" or "direct fusion" refers to a linkage between two polypeptide chains by a peptide bond. For example, a first amino acid is "fused directly" to a second amino acid when the first amino acid is "fused" to a second amino
15 acid by a peptide bond.

"Heterologous" and "heterologous moiety" in reference to a polypeptide moiety or polynucleotide moiety that is part of a larger polypeptide or polynucleotide, respectively, describes a polypeptide or polynucleotide that originates from a different polypeptide or polynucleotide than the remaining part of the polypeptide or polynucleotide molecule. The
20 additional heterologous component of the polypeptide or polynucleotide can originate from the same organism as the remaining polypeptide or polynucleotide, respectively, described herein, or the additional components can be from a different organism. For instance, a heterologous polypeptide can be synthetic, or derived from a different species, different cell type of an individual, or the same or different type of cell of distinct individuals. In one
25 aspect, a heterologous moiety is a polypeptide fused to another polypeptide to produce a polypeptide. In another aspect, a heterologous moiety is a non-polypeptide such as PEG conjugated to a polypeptide or protein.

As used herein, the terms "muscle cell" refers to a cell derived from muscle tissue, including, but not limited to, cells derived from skeletal muscle, smooth muscle (e.g. from
30 the digestive tract, urinary bladder, and blood vessels), and cardiac muscle. The term includes muscle cells in vitro, ex vivo, and in vivo. Thus, for example, an isolated cardiomyocyte would constitute a muscle cell, as would a cell as it exists in muscle tissue

present in a subject in vivo. This term also encompasses both terminally differentiated and nondifferentiated muscle cells, such as myocytes, myotubes, myoblasts, cardiomyocytes, and cardiomyoblasts.

As used herein, the term "muscle-specific" in reference to a gene regulatory element (e.g., enhancer sequence, promoter sequence) means that the regulatory element drives transcriptional activity primarily in muscle cells or tissue (e.g., 20:1) compared to the transcriptional activity driven by the regulatory element in other tissues. Assays to determine the muscle-specificity of a regulatory element are known in the art (e.g., in vitro assay using murine muscle cells and liver cells transfected with an expression vector comprising the regulatory element to be tested driving expression of a beta-galactoside reporter).

As used herein, the term "adeno-associated virus" or "AAV" includes but is not limited to, AAV type 1, AAV type 2, AAV type 3 (including types 3A and 3B), AAV type 4, AAV type 5, AAV type 6, AAV type 7, AAV type 8, AAV type 9, AAV type 10, AAV type 11, AAV type 12, AAV type 13, snake AAV, avian AAV, bovine AAV, canine AAV, equine AAV, ovine AAV, goat AAV, shrimp AAV, primate AAV, non-primate AAV, and ovine AAV, those AAV serotypes and clades disclosed by Gao *et al.* (J. Virol. 78:6381 (2004)) and Moris *et al.* (Virol. 33:375 (2004)), and any other AAV now known or later discovered. See, e.g., Fields *et al.* VIROLOGY, volume 2, chapter 69 (4th ed., Lippincott-Raven Publishers). AAV refers to a *Dependoparvovirus* within the *Parvoviridae* family of viruses. For example, the AAV can be an AAV derived from a naturally occurring "wild-type" virus, an AAV derived from a recombinant AAV (rAAV) genome packaged into a capsid derived from capsid proteins encoded by a naturally occurring *cap* gene and/or a rAAV genome packaged into a capsid derived from capsid proteins encoded by a non-natural capsid *cap* gene. As used herein, "AAV" can be used to refer to the virus itself or derivatives thereof. The term covers all subtypes and both naturally occurring and recombinant forms, except where expressly indicated otherwise. "Primate AAV" refers to AAV that infect primates, "non-primate AAV" refers to AAV that infects animals other than primates, "bovine AAV" refers to AAV that infect bovine mammals, etc. See, e.g., BERNARD N. FIELDS *et al.*, VIROLOGY, volume 2, chapter 69 (3 d ed., Lippincott-Raven Publishers).

The term "rAAV" refers to a "recombinant AAV." In some embodiments, a recombinant AAV has an AAV genome in which part or all of the *rep* and *cap* genes have been replaced with heterologous polynucleotide sequences.

5 An "AAV vector" or "adeno-associated virus vector" as used herein refers to an rAAV comprising a polynucleotide sequence not of AAV origin (i.e., a polynucleotide heterologous to AAV), typically a sequence of interest for the genetic transformation of a cell. In general, the heterologous polynucleotide is flanked by at least one, and generally by two, AAV inverted terminal repeat sequences (ITRs).

10 A "capsid-free" or "capsid-less" (or variations thereof) viral (e.g., AAV) genome or nucleic acid molecule refers to a genome or nucleic acid molecule free from a capsid. In some embodiments, the capsid-less genome or nucleic acid molecule does not contain sequences encoding, for example, an AAV Rep protein.

15 An "AAV" or "AAV viral particle" or "AAV vector" or "rAAV vector particle" refers to a viral particle composed of at least one AAV capsid protein (typically of all of the capsid proteins of a wild-type AAV) and an encapsidated polynucleotide. If the particle comprises a heterologous polynucleotide (i.e. a polynucleotide other than a wild-type AAV genome, such as a transgene to be delivered to a mammalian cell), it is typically referred to as an "rAAV vector particle" or simply an "AAV vector."

20 A "helper virus" for AAV refers to a virus that allows AAV (e.g., wild-type AAV) to be replicated and packaged by a mammalian cell. A variety of such helper viruses for AAV are known in the art, including adenoviruses, herpesviruses and poxviruses such as vaccinia. The adenoviruses encompass a number of different subgroups, although Adenovirus type 5 of subgroup C is most commonly used. Numerous adenoviruses of human, non-human mammalian and avian origin are known and available from depositories
25 such as the ATCC. Viruses of the herpes family include, for example, herpes simplex viruses (HSV) and Epstein-Barr viruses (EBV), as well as cytomegaloviruses (CMV) and pseudorabies viruses (PRV), all of which are also available from depositories such as ATCC.

30 As used herein, the term "inverted terminal repeat" (or "ITR") refers to a single stranded sequence of nucleotides followed downstream by its reverse complement. The intervening sequence of nucleotides between the initial sequence and the reverse complement can be

any length including zero. The AAV genome typically comprises inverted terminal repeats (ITRs) at both ends, wherein each end typically is palindromic and can form a hairpin.

The terms "polynucleotide" and "nucleic acid" are used interchangeably herein and refer to a biopolymer composed of a plurality of nucleotide monomers covalently bonded in a chain

The term "tropism" as used herein refers to a virus's (e.g., AAV's) ability to infect only one or more particular cell types and its ability to interact only with specific cell surface moieties to achieve cell entry, optionally and preferably followed by expression (e.g., transcription and, optionally, translation) of sequences carried by the virus (e.g., AAV) into the cell (e.g., for a recombinant virus, expression of the heterologous nucleotide sequence(s)).

As used herein, the term "transduction" refers to the entry of the virus (e.g., AAV) into the cell and the transfer of genetic material contained within the virus into the cell to obtain expression from the virus genome. Typically, a virus (e.g., AAV) enters cells in accordance with its tropism.

"Administering" refers to the physical introduction of a therapeutic agent to a subject, using any of the various methods and delivery systems known to those skilled in the art. Exemplary routes of administration, e.g., for an AAV therapy, include intravenous, intramuscular, intraarterial, intrathecal, intralymphatic, intralesional, intracapsular, intraorbital, intracardiac, intradermal, intraperitoneal, transtracheal, subcutaneous, subcuticular, intraarticular, subcapsular, subarachnoid, intraspinal, epidural, intrasternal, oral, rectal, topical, epidermal, mucosal, intranasal, vaginal, rectal, and sublingual administration. Administering can also be performed, for example, once, a plurality of times, and/or over one or more extended periods.

"Treatment" or "therapy" of a subject refers to any type of intervention or process performed on, or the administration of an active agent to, a subject with the objective of reversing, alleviating, ameliorating, inhibiting, slowing down, or preventing the onset, progression, development, severity, or recurrence of a symptom, complication, condition, or biochemical indicia associated with a disease.

A "therapeutically effective amount," "therapeutic dose," "effective dose," or "effective dosage," as used herein, means an amount or a dose that achieves a therapeutic goal, as described herein. One of ordinary skill in the art will further understand that a

therapeutically effective amount etc. can be administered in a single dose, or can be achieved by administration of multiple doses (*i.e.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more doses). The ability of a therapeutic agent to promote disease regression or inhibit the development or recurrence of the disease can be evaluated using a variety of methods known to the skilled practitioner, such as in human subjects during clinical trials, in animal model systems predictive of efficacy in humans, or by assaying the activity of the agent in *in vitro* assays.

A "subject" includes any human or non-human animal. The term "nonhuman animal" includes, but is not limited to, vertebrates such as nonhuman primates, sheep, dogs, and rodents such as mice, rats, and guinea pigs. In some embodiments, the subject is a human. The terms "subject" and "patient" are used interchangeably herein.

As used herein, the terms "ug" and "uM" are used interchangeably with "μg" and "μM," respectively.

The use of the alternative (*e.g.*, "or") should be understood to mean either one, both, or any combination thereof of the alternatives. As used herein, the indefinite articles "a" or "an" should be understood to refer to "one or more" of any recited or enumerated component or entity.

Approximately or about: As used herein, the term "approximately" or "about," as applied to one or more values of interest, refers to a value that is similar to a stated reference value and within a range of values that fall within 25%, 20%, 19%, 18%, 17%, 16%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, or less in either direction (greater than or less than) of the stated reference value unless otherwise stated or otherwise evident from the context (except where such number would exceed 100% of a possible value). When the term "approximately" or "about" is applied herein to a particular value, the value without the term "approximately" or "about" is also disclosed herein.

As described herein, any concentration range, percentage range, ratio range, or integer range is to be understood to include the value of any integer within the recited range and, when appropriate, fractions thereof (such as one tenth and one hundredth of an integer), unless otherwise indicated.

Various aspects of the disclosure are described in further detail in the following subsections.

Polynucleotides and Polypeptides

Miniaturized dystrophin

The present disclosure is directed to a nucleic acid molecule comprising a nucleotide sequence, which encodes a miniaturized dystrophin polypeptide. In some embodiments, the miniaturized dystrophin polypeptide comprises at least three hinge domains of dystrophin and at least five Spectrin repeat domains.

Dystrophin is a rod-shaped cytoplasmic protein that connects the cytoskeleton of a muscle fiber to the surrounding extracellular matrix through the cell membrane. This protein is located primarily in muscles used for movement (skeletal muscles) and in heart (cardiac) muscle. Small amounts of dystrophin are present in nerve cells in the brain. In skeletal and cardiac muscles, dystrophin is part of a group of proteins (a protein complex) that work together to strengthen muscle fibers and protect them from injury as muscles contract and relax. The dystrophin complex acts as an anchor, connecting each muscle cell's structural framework (cytoskeleton) with the lattice of proteins and other molecules outside the cell (extracellular matrix). The dystrophin complex can also play a role in cell signaling by interacting with proteins that send and receive chemical signals.

The DMD gene, encoding the full length dystrophin protein, is one of the longest human genes known, covering 2.3 megabases (0.08% of the human genome) at locus Xp21. The primary transcript in muscle measures about 2,100 kilobases and takes 16 hours to transcribe; the mature mRNA measures 14.0 kilobases. The 79-exon muscle transcript codes for a protein of 3685 amino acid residues.

Disclosed herein are amino acid and nucleotide sequences for dystrophin. The amino acid sequence constituting human wild type dystrophin, isoform Dp427m, is known as UniProt identifier No. NP_003997.1 and shown in Table 1.

Table 1. Amino Acids sequence of full-length Dystrophin Protein (NP_003997.1).

SEQ ID NO: 1
MLWWEIVEDCYEREDVQKKTFTKWVNAQFSKFGKQHIEENLFSDLQDGRRL LDLLEGLTGQKLPKEKGSTRVHALNNVNKALRVLQNNNVDLVNIGSTDIVD GNHKLTLGLIWNILHWQVKNVMKNIMAGLQQTNSEKILLSWVRQSTRNYP QNVNINFTTSWSDGLALNALIHSRDLFDWNSVVCQQSATQRLEHAFNIAR YQLGIEKLLDPEDVDTTYPDKKSILMYITSLFQVLPQQVSIEAIQEVEMLPRPP KVTKEEHFQLHHQMHYSQQITVSLAQGYERTSSPKPRFKSYAYTQAAYVTTS DPTRSPFPSQHLEAPEDKSGSSLMSEVNLDTRYQTAL EEVLSWLLSAEDTLQ AQGEISNDVEVVKDQFHTEGYMMDLTAHQGRVGNILQLGSKLIGTGKLSE

DEETEVQEQMNLNSRWECLRVASMEKQSNLHRVLMIDLQNKQLKELNDW
LTKTEERTRKMEEEEPLGPDLEDLKRQVQQHKVLQEDLEQEQVRVNSLTHMV
VVVDESSGDHATAALEEQLKVLGDRWANICRWTEDRWVLLQDILLKWQRL
TEEQCLFSAWLSEKEDAVNKIHTTGFKDQNEMLSSLQKLAVLKADLEKKKQ
SMGKLYSLKQDLLSTLKNKSVTQKTEAWLDNFARCWDNLVQKLEKSTAQIS
QAVTTTQPSLTQTTVMETVTTVTTREQILVKHAQEELPPPPQKKRQITVDSEI
RKRLDVDITELHSWITRSEAVLQSPEFAIFRKEGNFSDLKEKVNAIEREKAKEF
RKLQDASRSAQALVEQMVNEGVDNADSIKQASEQLNSRWIEFCQLLSERLNW
LEYQNNIAFYNNQLQQLQEQMTTTAENWLKIQPTTPSEPTAIKSQLKICKDEVN
RLSGLQPQIERLKIQSIALKEKGQGMFLDADFVAFTNHFKQVFSDVQAREKE
LQTFIDTLPPMRYQETMSAIRTWVQQSETKLSIPQLSVTDYEIMEQRLGELQA
LQSSLQEQQSGLYYLSTTVKEMSKKAPSEISRKYQSEFEEIEGRWKKLSSQLV
EHCQKLEEQMNKLRKIQNHIQTLKKWMAEVDVFLKEEWPALGDSEILKKQL
KQCRLLVSDIQTIPSLNSVNEGGQKIKNEAEPEFASRLETTELKELNTQWDH
MCQQVYARKEALKGGLEKTVSLQKDLSEMHEWMTQAEEEYLERDFEYKTP
DELQKAVEEMKRAKEEAQQKEAKVKLLTESVNSVIAQAPPVAQEALKKELE
TLTTNYQWLCTRLNGKCKTLEEVWACWHELLSYLEKANKWLNEVEFKLKT
TENIPGGAEIESEVLDSLENLMRHSEDNPNQIRILAQTLTDGGVMDELINEELE
TFNSRWRELHEEAARRQKLLQESIQSAQETEKSLHLIQESLTFIDKQLAAYIA
DKVDAAQMPQEAQKIQSDLTSHEISLEEMKKHNQGKEAAQRVLSQIDVAQK
KLQDVSMKFRLFQKPANFEQRLQESKMILDEVKMHLPALETKSVEQEVVQS
QLNHCNVNLYKSLSEVKSEVEMVIKTGRQIVQKKQTENPKELDERVTALKLH
YNELGAKVTERKQQLEKCLKLSRKMRKEMNVLTEWLAATDMELTKRSAVE
GMPSNLDSEVAWGKATQKEIEKQKVHLKSITEVGEALKTVLGKKETLVEDK
LSLLNSNWIAVTSRAEEWLNLLLEYQKHMETFQDQNVDHITKWIIQADTLLE
SEKKKPPQKEDVLKRLKAELNDIRPKVDSTRDQAANLMANRGDHCRLVEP
QISELNHRFAAISHRIKTGKASIPLKELEQFNSDIQKLEPLEAEIQQGVNLKEE
DFNKDMNEDNEGTVKELLQRGDNLQQRITDERKREEIKIKQQLLQTKHNAL
KDLRSQRRKKALEISHQWYQYKRQADDLLKCLDDIEKKLASLPEPRDERKIK
EIDRELQKKKEELNAVRRQAEGLSGDGAAMAVEPTQIQLSKRWREIESKFAQ
FRRLNFAQIHTVREETMMVMTEDMPLISYVPSTYLTEITHVSQALLEVEQLL
NAPDLCAKDFEDLQKQESLKNIKDSLQSSGRIDIIHSSKKTAAALQSATPVER
VKLQEALSQLDQFQWEKVNKMYKDRQGRFDRSVEKWRRFHYDIKIFNQWLT
EAEQFLRKQIPENWEHAKYKWYLKELQDGIGQRQTVVRTL NATGEEIIQQS
SKTDASILQEKLGSNLNRWQEVCKQLSDRKKRLEEQKNILSEFQRD LNEFVL
WLEEADNIASIPLEPGKEQQLEKLEQVKLLVEELPLRQGILKQLNETGGPVL
VSAPISPEEQDKLENKLKQTNLQWIKVSRAPEKQGEIEAQIKDLGQLEKKLE
DLEEQLNHLLLWLSPIRNQLEIYNQPNQEGPFDVQETEIAVQAKQPDVEEILS
KGQHLYKEKPATQPVKRKLEDLSSEWKA VNRLQLQELRAKQPD LAPGLTTIG
ASPTQTVTLVTQPVVTKETAISKLEMPSSLMLEVPALADFNRAWTELTDWLS
LLDQVIKSQRVMVGDL EDINEMIIKQKATMQDLEQRRPQLEELITAAQNLKN
KTSNQEARTIITDRIERIQNQWDEVQEHLQNRQQNLNEMLKDSTQWLEAKEE
AEQVLGQARAKLESWKEGPYTVDAIQKKITETKQLAKDLRQWQTNVDVAN
DLALKLLRDYSADDTRKVMHMITENINASWRSIHKRVSEREAALEETHRLLQQ
FPLDLEKFLAWLTEAETTANVLQDATRKERLLED SKGVKELMKQWQDLQG
EIEAHTDVYHNLDENSQKILRSLEGSDDAVLLQRRLDNMNFKWSELRKKS
NIRSHLEASSDQWKRLHLSLQELLVWLQLKDDELSRQAPIGGDFPAVQKQN
DVHRAFKRELKTKEPVMSTLETVRIFLTEQPLEGLEKLYQEPRELPPEERAQ

NVTRLLRKQAEENVTEWEKLNLSADWQRKIDETLERLQELQEATDELDLK
 LRQAEVIKGSWQPVGDLLIDSLQDHLEKVKALRGEIAPLKENVSHVNDLARQ
 LTTLGIQLSPYNLSTLEDLNTRWKLQVAVEDRVRQLHEAHRDFGPASQHFL
 STSVQGPWERAISPKNVPYYINHETQTTCDWHPKMTELYQSLADLNNVRFSA
 YRTAMKLRRLQKALCLDLSLSAACDALDQHNKQNDQPMQDILQIINCLTTI
 YDRLEQEHNNLVNPLCVDMCLNWLNVYDTGRTGRIRVLSFKTGIIISLCKA
 HLEDKYRYLQKQVASSTGFCDQRRLLGLLHDSIQIPRQLGEVASFGGSNIEPS
 VRSCFQFANNKPEIEAALFLDWMRLEPQSMVWLPVLHRVAAAETAKHQAK
 CNICKECPIIGFRYRSLKHFNYDICQSCFFSGRVAKGHKMHYPMVEYCTPTTS
 GEDVRDFAKVLKNKFRTKRYFAKHPRMGYLPVQTVLEGDNMETPVTLINF
 WPVDSAPASSPQLSHDDTHSRIEHYASRLAEMENSNGSYLNDISPNESIDDE
 HLLIQHYCQSLNQDSPLSQPRSPAQILISLESEERGELERILADLEENRNLQAE
 YDRLKQQHEHKGLSPLSPPEMMPTSPQSPRDAELIAEAKLLRQHKGRLEAR
 MQILEDHNKQLESQHLRLRQLLEQPQAEAKVNGTTVSSPSTSLQRSDSSQPM
 LLRVVGSQTSDSMGEEDLLSPPQDTSTGLEEVMEQLNNSFPSSRGRNTPGKP
 MREDTM

Various other dystrophin isoforms are known in the art that result from alternative splicing. In some embodiments, the constructs comprise the nucleotide sequences recited in Table 2.

- 5 Also disclosed herein is a nucleotide sequence encoding the full-length dystrophin protein.

Table 2. Nucleotide sequence of full-length Dystrophin Protein (isoform Dp427m).

SEQ ID NO: 2
GGGATTCCCTCACTTTCCCCCTACAGGACTCAGATCTGGGAGGCAATTACCTTCGG AGAAAAACGAATAGGAAAAACTGAAGTGTTACTTTTTTTAAAGCTGCTGAAGTTT GTTGGTTTCTCATTGTTTTTAAGCCTACTGGAGCAATAAAGTTTGAAGAACTTTTA CCAGGTTTTTTTTATCGCTGCCTTGATATACACTTTTCAAATGCTTTGGTGGGAAG AAGTAGAGGACTGTTATGAAAGAGAAGATGTTCAAAGAAAACATTCACAAAAT GGGTAAATGCACAATTTTCTAAGTTTGGGAAGCAGCATATTGAGAACCTCTTCAGT GACCTACAGGATGGGAGGCGCCTCCTAGACCTCCTCGAAGGCCTGACAGGGCAAA AACTGCCAAAAGAAAAAGGATCCACAAGAGTTTCATGCCCTGAACAATGTCAACA AGGCACTGCGGGTTTTGCAGAACATAATGTTGATTTAGTGAATATTGGAAGTAC TGACATCGTAGATGGAAATCATAAACTGACTCTTGTTTGAATTTGGAATATAATCC TCCACTGGCAGGTCAAAAATGTAATGAAAAATATCATGGCTGGATTGCAACAAAC CAACAGTGAAAAGATTCTCCTGAGCTGGGTCCGACAATCAACTCGTAATTATCCA CAGGTTAATGTAATCAACTTCACCACCAGCTGGTCTGATGGCCTGGCTTTGAATGC TCTCATCCATAGTCATAGGCCAGACCTATTTGACTGGAATAGTGTGGTTTGCCAGC AGTCAGCCACACAACGACTGGAACATGCATTCAACATCGCCAGATATCAATTAGG CATAGAGAACTACTCGATCCTGAAGATGTTGATACCACCTATCCAGATAAGAAG TCCATCTTAATGTACATCATCACTCTTCCAAGTTTTGCCTCAACAAGTGAGCAT TGAAGCCATCCAGGAAGT GGAAATGTTGCCAAGGCCACCTAAAGTGACTAAAGAAGAACATTTTCAGTTACAT CATCA

AATGCACTATTCTCAACAGATCACGGTCAGTCTAGCACAGGGATATGAGAGAACT
 TCTTC
 CCCTAAGCCTCGATTCAAGAGCTATGCCTACACACAGGCTGCTTATGTCACCACCT
 CTGA
 CCCTACACGGAGCCCATTTCTTCACAGCATTTGGAAGCTCCTGAAGACAAGTCAT
 TTGG
 CAGTTCATTGATGGAGAGTGAAGTAAACCTGGACCGTTATCAAACAGCTTTAGAA
 GAAGT
 ATTATCGTGGCTTCTTTCTGCTGAGGACACATTGCAAGCACAAGGAGAGATTCTA
 ATGA
 TGTGGAAGTGGTGAAAGACCAGTTTCATACTCATGAGGGGTACATGATGGATTG
 ACAGC
 CCATCAGGGCCGGGTGGTAATATTCTACAATTGGGAAGTAAGCTGATTGGAACA
 GGAAA
 ATTATCAGAAGATGAAGAACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCA
 AGATG
 GGAATGCCTCAGGGTAGCTAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAA
 ATGGA
 TCTCCAGAATCAGAACTGAAAGAGTTGAATGACTGGCTAACAAAAACAGAAGA
 AAGAAC
 AAGGAAAATGGAGGAAGAGCCTCTTGGACCTGATCTTGAAGACCTAAAACGCCA
 AGTACA
 ACAACATAAGGTGCTTCAAGAAGATCTAGAACAAGAACAAGTCAGGGTCAATTCT
 CTCAC
 TCACATGGTGGTGGTAGTTGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGG
 AAGA
 ACAACTTAAGGTATTGGGAGATCGATGGGCAAACATCTGTAGATGGACAGAAGAC
 CGCTG
 GGTTCTTTTACAAGACATCCTTCTCAAATGGCAACGTCTTACTGAAGAACAGTGCC
 TTTT
 TAGTGCATGGCTTTCAGAAAAAGAAGATGCAGTGAACAAGATTCACACAACCTGGC
 TTAA
 AGATCAAAATGAAATGTTATCAAGTCTTCAAAAACCTGGCCGTTTTAAAAGCGGAT
 CTAGA
 AAAGAAAAAGCAATCCATGGGCAAACCTGTATTCACTCAAACAAGATCTTCTTTCA
 AACT
 GAAGAATAAGTCAGTGACCCAGAAGACGGAAGCATGGCTGGATAACTTTGCCCCG
 GTGTTG
 GGATAATTTAGTCCAAAAACTTGAAAAGAGTACAGCACAGATTTACAGGCTGTC
 ACCAC
 CACTCAGCCATCACTAACACAGACAACTGTAATGGAAACAGTAACTACGGTGACC
 ACAAG
 GGAACAGATCCTGGTAAAGCATGCTCAAGAGGAACTTCCACCACCACCTCCCCAA
 AAGAA
 GAGGCAGATTACTGTGGATTCTGAAATTAGGAAAAGGTTGGATGTTGATATAACT
 GAACT
 TCACAGCTGGATTACTCGCTCAGAAGCTGTGTTGCAGAGTCCTGAATTTGCAATCT
 TTCG

GAAGGAAGGCAACTTCTCAGACTTAAAAGAAAAAGTCAATGCCATAGAGCGAGA
 AAAAGC
 TGAGAAGTTCAGAAAAGTCAAGATGCCAGCAGATCAGCTCAGGCCCTGGTGGAA
 CAGAT
 GGTGAATGAGGGTGTTAATGCAGATAGCATCAAACAAGCCTCAGAACAAGTGAAC
 AGCCG
 GTGGATCGAATTCTGCCAGTTGCTAAGTGAGAGACTTAACTGGCTGGAGTATCAG
 ACAA
 CATCATCGCTTTCTATAATCAGCTACAACAATTGGAGCAGATGACAAGTACTGCTG
 AAAA
 CTGGTTGAAAATCCAACCCACCACCCCATCAGAGCCAACAGCAATTAAAAGTCAG
 TTAA
 AATTTGTAAGGATGAAGTCAACCGGCTATCAGGTCTTCAACCTCAAATTGAACGA
 TTAA
 AATTCAAAGCATAGCCCTGAAAGAGAAAGGACAAGGACCCATGTTCTGGATGCA
 GACTT
 TGTGGCCTTTACAAATCATTTTAAGCAAGTCTTTTCTGATGTGCAGGCCAGAGAGA
 AAGA
 GCTACAGACAATTTTTGACACTTTGCCACCAATGCGCTATCAGGAGACCATGAGT
 GCCAT
 CAGGACATGGGTCCAGCAGTCAGAAACCAAAGTCTCCATACCTCAACTTAGTGTC
 ACCGA
 CTATGAAATCATGGAGCAGAGACTCGGGGAATTGCAGGCTTTACAAAGTTCTCTG
 CAAGA
 GCAACAAAGTGGCCTATACTATCTCAGCACCAGTGTGAAAGAGATGTCGAAGAAA
 GCGCC
 CTCTGAAATTAGCCGGAAATATCAATCAGAATTTGAAGAAATTGAGGGACGCTGG
 AAGAA
 GCTCTCCTCCCAGCTGGTTGAGCATTGTCAAAGCTAGAGGAGCAAATGAATAAA
 CTCCG
 AAAAATTCAGAATCACATACAAACCCTGAAGAAATGGATGGCTGAAGTTGATGTT
 TTCT
 GAAGGAGGAATGGCCTGCCCTTGGGGATTGAGAAATTCTAAAAAGCAGCTGAA
 ACAGTG
 CAGACTTTTAGTCAGTGATATTCAGACAATTCAGCCCAGTCTAAACAGTGTCAATG
 AAGG
 TGGGCAGAAGATAAAGAATGAAGCAGAGCCAGAGTTTGCTTCGAGACTTGAGAC
 AGAAGT
 CAAAGAACTTAACACTCAGTGGGATCACATGTGCCAACAGGTCTATGCCAGAAAG
 GAGGC
 CTTGAAGGGAGGTTTGGAGAAAAGTGTAAAGCCTCCAGAAAGATCTATCAGAGATG
 CACGA
 ATGGATGACACAAGCTGAAGAAGAGTATCTTGAGAGAGATTTTGAATATAAAAGT
 CCAGA
 TGAATTACAGAAAGCAGTTGAAGAGATGAAGAGAGCTAAAGAAGAGGCCCAACA
 AAAAGA
 AGCGAAAGTGAAAGTCTTACTGAGTCTGTAAATAGTGTCATAGCTCAAGCTCCA
 CCTGT

AGCACAAAGAGGCCTTAAAAAAGGAACTTGAAACTCTAACCACCAACTACCAGTGG
CTCTG
CACTAGGCTGAATGGGAAATGCAAGACTTTGGAAGAAGTTTGGGCATGTTGGCAT
GAGTT
ATTGTCATACTTGGAGAAAGCAAACAAGTGGCTAAATGAAGTAGAATTTAAACTT
AAAAC
CACTGAAAACATTCCCTGGCGGAGCTGAGGAAATCTCTGAGGTGCTAGATTCACTT
GAAAA
TTTGATGCGACATTCAGAGGATAACCCAAATCAGATTCGCATATTGGCACAGACC
CTAAC
AGATGGCGGAGTCATGGATGAGCTAATCAATGAGGAACTTGAGACATTTAATTCT
CGTTG
GAGGGAACCTACATGAAGAGGCTGTAAGGAGGCAAAAGTTGCTTGAACAGAGCAT
CCAGTC
TGCCCAGGAGACTGAAAAATCCTTACACTTAATCCAGGAGTCCCTCACATTCATTG
ACAA
GCAGTTGGCAGCTTATATTGCAGACAAGGTGGACGCAGCTCAAATGCCTCAGGAA
GCCCA
GAAAATCCAATCTGATTTGACAAGTCATGAGATCAGTTTAGAAGAAATGAAGAAA
CATAA
TCAGGGGAAGGAGGCTGCCCAAAGAGTCCTGTCTCAGATTGATGTTGCACAGAAA
AAATT
ACAAGATGTCTCCATGAAGTTTCGATTATTCCAGAAACCAGCCAATTTTGAGCTGC
GTCT
ACAAGAAAGTAAGATGATTTTAGATGAAGTGAAGATGCACTTGCCTGCATTGGAA
ACAAA
GAGTGTGGAACAGGAAGTAGTACAGTCACAGCTAAATCATTGTGTGAACTTGTAT
AAAAG
TCTGAGTGAAGTGAAGTCTGAAGTGGAAATGGTGATAAAGACTGGACGTCAGATT
GTACA
GAAAAAGCAGACGGAAAATCCCAAAGAACTTGATGAAAGAGTAACAGCTTTGAA
ATTGCA
TTATAATGAGCTGGGAGCAAAGGTAACAGAAAGAAAGCAACAGTTGGAGAAATG
CTTGAA
ATTGTCCCGTAAGATGCGAAAGGAAATGAATGTCTTGACAGAATGGCTGGCAGCT
ACAGA
TATGGAATTGACAAAGAGATCAGCAGTTGAAGGAATGCCTAGTAATTTGGATTCT
GAAGT
TGCCTGGGGAAAGGCTACTCAAAAAGAGATTGAGAAACAGAAGGTGCACCTGAA
GAGTAT
CACAGAGGTAGGAGAGGCCTTGAAAACAGTTTTTGGGCAAGAAGGAGACGTTGGT
GGAAGA
TAAACTCAGTCTTCTGAATAGTAACTGGATAGCTGTCACCTCCCGAGCAGAAGAG
TGGTT
AAATCTTTTGTGGAATACCAGAAACACATGGAAACTTTTGACCAGAATGTGGAC
CACAT
CACAAAGTGGATCATTCAGGCTGACACACTTTTGGATGAATCAGAGAAAAAGAAA
CCCCA

GCAAAAAGAAGACGTGCTTAAGCGTTTAAAGGCAGAACTGAATGACATACGCCC
 AAAGGT
 GGACTCTACACGTGACCAAGCAGCAAACCTTGATGGCAAACCGCGGTGACCACTGC
 AGGAA
 ATTAGTAGAGCCCCAAATCTCAGAGCTCAACCATCGATTTGCAGCCATTTACACA
 GAAT
 TAAGACTGGAAAGGCCTCCATTCTTTGAAGGAATTGGAGCAGTTTAACTCAGAT
 ATACA
 AAAATTGCTTGAACCACTGGAGGCTGAAATTCAGCAGGGGGTGAATCTGAAAGAG
 GAAGA
 CTTCAATAAAGATATGAATGAAGACAATGAGGGTACTGTAAAAGAATTGTTGCAA
 AGAGG
 AGACAACCTACAACAAAGAATCACAGATGAGAGAAAGAGAGAGGAAATAAAGAT
 AAAACA
 GCAGCTGTTACAGACAAAACATAATGCTCTCAAGGATTTGAGGTCTCAAAGAAGA
 AAAAA
 GGCTCTAGAAATTTCTCATCAGTGGTATCAGTACAAGAGGCAGGCTGATGATCTC
 CTGAA
 ATGCTTGGATGACATTGAAAAAAAAATTAGCCAGCCTACCTGAGCCCAGAGATGAA
 AGGAA
 AATAAAGGAAATTGATCGGGAATTGCAGAAGAAGAAAGAGGAGCTGAATGCAGT
 GCGTAG
 GCAAGCTGAGGGCTTGTCTGAGGATGGGGCCGCAATGGCAGTGGAGCCAACTCAG
 ATCCA
 GCTCAGCAAGCGCTGGCGGGAAATTGAGAGCAAATTTGCTCAGTTTCGAAGACTC
 AACTT
 TGCACAAATTCACACTGTCCGTGAAGAAACGATGATGGTGATGACTGAAGACATG
 CCTTT
 GGAAATTTCTTATGTGCCTTCTACTTATTTGACTGAAATCACTCATGTCTCACAAG
 CCCT
 ATTAGAAGTGGAACAACTTCTCAATGCTCCTGACCTCTGTGCTAAGGACTTTGAAG
 ATCT
 CTTTAAGCAAGAGGAGTCTCTGAAGAATATAAAAGATAGTCTACAACAAAGCTCA
 GGTCG
 GATTGACATTATTCATAGCAAGAAGACAGCAGCATTGCAAAGTGCAACGCCTGTG
 GAAAG
 GGTGAAGCTACAGGAAGCTCTCTCCAGCTTGATTTCCAATGGGAAAAAGTTAAC
 AAAAT
 GTACAAGGACCGACAAGGGCGATTTGACAGATCTGTTGAGAAATGGCGGCGTTTT
 CATT
 TGATATAAAGATATTTAATCAGTGGCTAACAGAAGCTGAACAGTTTCTCAGAAAG
 ACACA
 AATTCCTGAGAATTGGGAACATGCTAAATACAAATGGTATCTTAAGGAACTCCAG
 GATGG
 CATTGGGCAGCGGCAAACCTGTTGTCAGAACATTGAATGCAACTGGGGAAGAAATA
 ATTCA
 GCAATCCTCAAAAACAGATGCCAGTATTCTACAGGAAAAATTGGGAAGCCTGAAT
 CTGCG

GTGGCAGGAGGTCTGCAAACAGCTGTCAGACAGAAAAAAGAGGCTAGAAGAACA
 AAAGAA
 TATCTTGTGAGAATTTCAAAGAGATTTAAATGAATTTGTTTTATGGTTGGAGGAAG
 CAGA
 TAACATTGCTAGTATCCCACTTGAACCTGGAAAAGAGCAGCAACTAAAAGAAAAG
 CTTGA
 GCAAGTCAAGTTACTGGTGGGAAGAGTTGCCCCTGCGCCAGGGAATTCTCAAACAA
 TTAAA
 TGAAACTGGAGGACCCGTGCTTGTAAGTGCTCCCATAAGCCCAGAAGAGCAAGAT
 AAACCT
 TGAAAATAAGCTCAAGCAGACAAATCTCCAGTGGATAAAGGTTTCCAGAGCTTTA
 CCTGA
 GAAACAAGGAGAAATTGAAGCTCAAATAAAAGACCTTGGGCAGCTTGAAAAAAA
 GCTTGA
 AGACCTTGAAGAGCAGTTAAATCATCTGCTGCTGTGGTTATCTCCTATTAGGAATC
 AGTT
 GGAAATTTATAACCAACCAAACCAAGAAGGACCATTTGACGTTTCAGGAAACTGAA
 ATAGC
 AGTTCAAGCTAAACAACCGGATGTGGAAGAGATTTTGTCTAAAGGGCAGCATTTG
 TACAA
 GGAAAAACCAGCCACTCAGCCAGTGAAGAGGAAGTTAGAAGATCTGAGCTCTGA
 GTGGAA
 GGCGGTAAACCGTTTACTTCAAGAGCTGAGGGCAAAGCAGCCTGACCTAGCTCCT
 GGACT
 GACCACTATTGGAGCCTCTCCTACTCAGACTGTTACTCTGGTGACACAACCTGTGG
 TTAC
 TAAGGAAACTGCCATCTCCAAACTAGAAATGCCATCTTCCTTGATGTTGGAGGTAC
 CTGC
 TCTGGCAGATTTCAACCGGGCTTGGACAGAACTTACCGACTGGCTTTCTCTGCTTG
 ATCA
 AGTTATAAAATCACAGAGGGTGATGGTGGGTGACCTTGAGGATATCAACGAGATG
 ATCAT
 CAAGCAGAAGGCAACAATGCAGGATTTGGAACAGAGGCGTCCCCAGTTGGAAGA
 ACTCAT
 TACCGCTGCCCAAATTTGAAAAACAAGACCAGCAATCAAGAGGCTAGAACAATC
 ATTAC
 GGATCGAATTGAAAGAATTCAGAATCAGTGGGATGAAGTACAAGAACACCTTCAG
 AACCG
 GAGGCAACAGTTGAATGAAATGTTAAAGGATTCAACACAATGGCTGGAAGCTAA
 GGAAGA
 AGCTGAGCAGGTCTTAGGACAGGCCAGAGCCAAGCTTGAGTCATGGAAGGAGGG
 TCCCTA
 TACAGTAGATGCAATCCAAAAGAAAATCACAGAAACCAAGCAGTTGGCCAAAGA
 CCTCCG
 CCAGTGGCAGACAAATGTAGATGTGGCAAATGACTTGGCCCTGAAACTTCTCCGG
 GATTA
 TTCTGCAGATGATACCAGAAAAGTCCACATGATAACAGAGAATATCAATGCCTCT
 TGGAG

AAGCATTTCATAAAAGGGTGAGTGAGCGAGAGGCTGCTTTGGAAGAACTCATAG
ATTACT
GCAACAGTTCCCCCTGGACCTGGAAAAGTTTCTTGCCTGGCTTACAGAAGCTGAA
ACAAC
TGCCAATGTCCTACAGGATGCTACCCGTAAGGAAAGGCTCCTAGAAGACTCCAAG
GGAGT
AAAAGAGCTGATGAAACAATGGCAAGACCTCCAAGGTGAAATTGAAGCTCACAC
AGATGT
TTATCACAACCTGGATGAAAACAGCCAAAAAATCCTGAGATCCCTGGAAGGTTCC
GATGA
TGCAGTCCTGTTACAAAGACGTTTGGATAACATGAACTTCAAGTGGAGTGAAGT
CGGAA
AAAGTCTCTCAACATTAGGTCCCATTGGAAGCCAGTTCTGACCAGTGAAGCGT
CTGCA
CCTTTCTCTGCAGGAACTTCTGGTGTGGCTACAGCTGAAAGATGATGAATTAAGCC
GGCA
GGCACCTATTGGAGGCGACTTTCCAGCAGTTCAGAAGCAGAACGATGTACATAGG
GCCTT
CAAGAGGGAATTGAAAACCTAAAGAACCTGTAATCATGAGTACTCTTGAGACTGTA
CGAAT
ATTTCTGACAGAGCAGCCTTTGGAAGGACTAGAGAACTCTACCAGGAGCCCAGA
GAGCT
GCCTCCTGAGGAGAGAGCCCAGAATGTCACTCGGCTTCTACGAAAGCAGGCTGAG
GAGGT
CAATACTGAGTGGGAAAAATTGAACCTGCACTCCGCTGACTGGCAGAGAAAAATA
GATGA
GACCCTTGAAAGACTCCAGGAACTTCAAGAGGCCACGGATGAGCTGGACCTCAAG
CTGCG
CCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGGCGATCTCCTCATTGAC
TCTCT
CCAAGATCACCTCGAGAAAGTCAAGGCACTTCGAGGAGAAATTGCGCCTCTGAAA
GAGAA
CGTGAGCCACGTCAATGACCTTGCTCGCCAGCTTACCACTTTGGGCATTGAGCTCT
CACC
GTATAACCTCAGCACTCTGGAAGACCTGAACACCAGATGGAAGCTTCTGCAGGTG
GCCGT
CGAGGACCGAGTCAGGCAGCTGCATGAAGCCCACAGGGACTTTGGTCCAGCATCT
CAGCA
CTTTCTTTCCACGTCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAG
TGCC
CTACTATATCAACCACGAGACTCAAACAACTTGCTGGGACCATCCCAAATGACA
GAGCT
CTACCAGTCTTTAGCTGACCTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCA
TGAA
ACTCCGAAGACTGCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCAT
GTGA
TGCCTTGGACCAGCACAACTCAAGCAAAATGACCAGCCCATGGATATCCTGCAG
ATTAT

TAATTGTTTGACCACTATTTATGACCGCCTGGAGCAAGAGCACAACAATTTGGTCA
 ACGT
 CCCTCTCTGCGTGGATATGTGTCTGAACTGGCTGCTGAATGTTTATGATACGGGAC
 GAAC
 AGGGAGGATCCGTGTCCTGTCTTTTAAAACTGGCATCATTTCCTGTGTAAAGCAC
 ATTT
 GGAAGACAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGT
 GACCA
 GCGCAGGCTGGGCCTCCTTCTGCATGATTCTATCCAAATTCCAAGACAGTTGGGTG
 AAGT
 TGCATCCTTTGGGGGCAGTAACATTGAGCCAAGTGTCCGGAGCTGCTTCCAATTTG
 CTAA
 TAATAAGCCAGAGATCGAAGCGGCCCTCTTCCTAGACTGGATGAGACTGGAACCC
 CAGTC
 CATGGTGTGGCTGCCCCGTCCTGCACAGAGTGGCTGCTGCAGAAACTGCCAAGCAT
 CAGGC
 CAAATGTAACATCTGCAAAGAGTGTCCAATCATTGGATTACAGGTACAGGAGTCTA
 AAGCA
 CTTTAATTATGACATCTGCCAAAGCTGCTTTTTTTCTGGTCGAGTTGCAAAAGGCC
 ATAA
 AATGCACTATCCCATGGTGGGAATATTGCACTCCGACTACATCAGGAGAAGATGTT
 CGAGA
 CTTTGCCAAGGTACTAAAAACAAATTTGAACCAAAGGTATTTTGCGAAGCAT
 CCCCCG
 AATGGGCTACCTGCCAGTGCAGACTGTCTTAGAGGGGGACAACATGGAAACTCCC
 GTTAC
 TCTGATCAACTTCTGGCCAGTAGATTCTGCGCCTGCCTCGTCCCCTCAGCTTTCAC
 ACGA
 TGATACTCATTACGCATTGAACATTATGCTAGCAGGCTAGCAGAAATGGAAAAC
 AGCAA
 TGGATCTTATCTAAATGATAGCATCTCTCCTAATGAGAGCATAGATGATGAACATT
 TGTT
 AATCCAGCATTACTGCCAAAGTTTGAACCAGGACTCCCCCCTGAGCCAGCCTCGT
 AGTCC
 TGCCCAGATCTTGATTTCTTAGAGAGTGAGGAAAGAGGGGAGCTAGAGAGAATC
 CTAGC
 AGATCTTGAGGAAGAAAACAGGAATCTGCAAGCAGAATATGACCGTCTAAAGCA
 GCAGCA
 CGAACATAAAGGCCTGTCCCCACTGCCGTCCCCTCCTGAAATGATGCCACCTCTC
 CCA
 GAGTCCCCGGGATGCTGAGCTCATTGCTGAGGCCAAGCTACTGCGTCAACACAAA
 GGCCG
 CCTGGAAGCCAGGATGCAAATCCTGGAAGACCACAATAAACAGCTGGAGTCACA
 GTTACA
 CAGGCTAAGGCAGCTGCTGGAGCAACCCCAGGCAGAGGCCAAAGTGAATGGCAC
 AACGGT
 GTCCTCTCCTTCTACCTCTCTACAGAGGTCCGACAGCAGTCAGCCTATGCTGCTCC
 GAGT

GGTTGGCAGTCAAACCTTCGGACTCCATGGGTGAGGAAGATCTTCTCAGTCCTCCCC
AGGA
CACAAGCACAGGGTTAGAGGAGGTGATGGAGCAACTCAACAACCTCCTTCCCTAGT
TCAAG
AGGAAGAAATACCCCTGGAAAGCCAATGAGAGAGGACACAATGTAGGAAGTCTT
TTCCAC
ATGGCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATG
AAGAA
GGAGCAGAATAAATGTTTTACAACCTCTGATTCCCGCATGGTTTTTATAATATTCA
TACA
ACAAAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAA
GGGTAA
GTGGTATTATACTGTAGATTTTCTAGTAGTTTCTAAGTCTGTTATTGTTTTGTAAACAA
TGG
CAGGTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACCTACATGTA
AAATC
TTGATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAA
AAAT
TTATAACAGTTATAAAGAAAGATTGTAAACTAAAGTGTGCTTTATAAAAAAAGT
TGTTT
ATAAAAACCCCTAAAAACAAAACAAACACACACACACACATACACACACACA
CACAAA
ACTTTGAGGCAGCGCATTGTTTTGCATCCTTTTGGCGTGATATCCATATGAAATTC
ATGG
CTTTTTCTTTTTTGCATATTAAAGATAAGACTTCCTCTACCACCACACCAAATGAC
TAC
TACACACTGCTCATTTGAGAACTGTCAGCTGAGTGGGGCAGGCTTGAGTTTTTCATT
TCAT
ATATCTATATGTCTATAAGTATATAAATACTATAGTTATATAGATAAAGAGATACG
AATT
TCTATAGACTGACTTTTTCCATTTTTTAAATGTTTCATGTCACATCCTAATAGAAAGA
AAT
TACTTCTAGTCAGTCATCCAGGCTTACCTGCTTGGTCTAGAATGGATTTTTCCCGG
AGCC
GGAAGCCAGGAGGAAACTACACCACACTAAAACATTGTCTACAGCTCCAGATGTT
TCTCA
TTTTAAACAACCTTCCACTGACAACGAAAGTAAAGTAAAGTATTGGATTTTTTTAA
AGGG
AACATGTGAATGAATACACAGGACTTATTATATCAGAGTGAGTAATCGGTTGGTT
GGTTG
ATTGATTGATTGATTGATACATTCAGCTTCCTGCTGCTAGCAATGCCACGATTTAG
ATTT
AATGATGCTTCAGTGGAATCAATCAGAAGGTATTCTGACCTTGTGAACATCAGA
AGGTA
TTTTTTAACTCCCAAGCAGTAGCAGGACGATGATAGGGCTGGAGGGCTATGGATT
CCCAG
CCCATCCCTGTGAAGGAGTAGGCCACTCTTTAAGTGAAGGATTGGATGATTGTTCA
TAAT

ACATAAAGTTCTCTGTAATTACAATAAATTATTATGCCCTCTTCTCACAGTCAAA
 AGGA
 ACTGGGTGGTTTGGTTTTTGTGCTTTTTTAGATTTATTGTCCCATGTGGGATGAGT
 TTT
 TAAATGCCACAAGACATAATTTAAAATAAATAAACTTTGGGAAAAGGTGTAAGAC
 AGTAG
 CCCCATCACATTTGTGATACTGACAGGTATCAACCCAGAAGCCCATGAACTGTGTT
 TCCA
 TCCTTTGCATTTCTCTGCGAGTAGTTCCACACAGGTTTGTAAGTAAGTAAGAAAGA
 AGGC
 AAATTGATTCAAATGTTACAAAAAAACCCTTCTTGGTGGATTAGACAGGTAAAT
 ATATA
 AACAAACAAACAAAAATTGCTCAAAAAAGAGGAGAAAAGCTCAAGAGGAAAAG
 CTAAGGA
 CTGGTAGGAAAAAGCTTTACTCTTTCATGCCATTTTATTTCTTTTTGATTTTTAAAT
 CAT
 TCATTCAATAGATAACACCGTGTGACCTATAATTTTGCAAATCTGTTACCTCTGAC
 ATCA
 AGTGTAAATTAGCTTTTGGAGAGTGGGCTGACATCAAGTGTAAATTAGCTTTTGGAGA
 GTGG
 GTTTTGTCCATTATTAATAATTAATTAATTAACATCAAACACGGCTTCTCATGCTAT
 TTC
 TACCTCACTTTGGTTTTGGGGTGTTCCTGATAATTGTGCACACCTGAGTTCACAGC
 TTCA
 CCACTTGTCCATTGCGTTATTTTCTTTTTCCTTTATAATTCTTTCTTTTTCCTTCATA
 AT
 TTTCAAAGAAAACCCAAAGCTCTAAGGTAACAAATTACCAAATTACATGAAGAT
 TTGGT
 TTTTGTCTTGCATTTTTTTCCTTTATGTGACGCTGGACCTTTTCTTTACCCAAGGATT
 TT
 TAAAACTCAGATTTAAAACAAGGGGTTACTTTACATCCTACTAAGAAGTTTAAGT
 AAGTA
 AGTTTCATTCTAAAATCAGAGGTAAATAGAGTGCATAAATAATTTTGTTTTAATCT
 TTTT
 GTTTTTCTTTTAGACACATTAGCTCTGGAGTGAGTCTGTCATAATATTTGAACAAA
 AATT
 GAGAGCTTTATTGCTGCATTTTAAGCATAATTAATTTGGACATTATTTTCGTGTTGTG
 TTC
 TTTATAACCAACCGAGTATTAACTGTAAATCATAATGTAAGTGAAGCATAAACAT
 CACAT
 GGCATGTTTTGTCATTGTTTTTCAGGTACTGAGTTCTTACTTGAGTATCATAATATAT
 TGT
 GTTTTAACACCAACACTGTAACATTTACGAATTATTTTTTTAACTTCAGTTTTACT
 GCA
 TTTTCACAACATATCAGACTTCACCAAATATATGCCTTACTATTGTATTATAGTACT
 GCT
 TTACTGTGTATCTCAATAAAGCACGCAGTTATGTTAC

The wild type, full length dystrophin protein (isoform Dp427m) contains 24 spectrin like repeats, at least four hinge regions, actin binding domain (ABD1), Cysteine rich domain (CR), and C terminal Domain (C-term.). The polypeptide sequence of each domain is shown in Table 3, and the nucleotide sequence of each domain is shown in Table 4.

5 Table 3. Amino Acid Sequences of Dystrophin Domains

Description and Sequence Identifier	Sequence
ABD1 (SEQ ID NO:3)	MLWWEEVEDCYEREDVQKKTFTKWVNAQFSKFGKQHLENLFSDLQDG RRLDLLEGLTGQKLPKEKGSTR VHALNNVNKALRVLQNNNVDLVNIGSTDIVDGNHKLTLGLIWNILHW QVKNVMKNIMAGLQQTNSEKIL LSWVRQSTRNYPQVNVINFTTSWSDGLALNALIHSRDLFDWNSVVC QQSATQRLEHAFNIARYQLGIE KLLDPEDVDTTYPDKKSILMYITSLFQVLPQQVSIEAIQEVE
Hinge 1 (SEQ ID NO:4)	MLPRPPKVTKEEHFQLHHQMHSYQQITVSLAQGYERTSSPKPRFKSYAY TQAAYVTTSDPTRSPFPSQHL EAPEDKSFGSSLMES
Spectrin repeat 1 (SEQ ID NO:5)	EVNLDTRYQTAL EEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTEGY MMDLTAHQGRVGNILQLGSKLI GTGKLSEDEETEVEQEQMNLNSRWECLRVASMEKQSNLH
Spectrin repeat 2 (SEQ ID NO:6)	RVLMDLQNQKLKELNDWLTKTEERTRKMEEEEPLGPDLEDLKRQVQQH KVLQEDLEQEVRVNSLTHMVVV VDESSGDHATAALEEQLKVLGDRWANICRWTEDRWVLLQDI
Spectrin repeat 3 (SEQ ID NO:7)	LLKWQRLTEEQCLFSAWLSEKEDAVNKIHTTGFKDQNEMLSSLQKLAV LKADLEKKKQSMGKLYSLKQDL LSTLKNKSVTQKTEAWLDNFARCWDNLVQKLEKSTAQISQA
Hinge 2 (SEQ ID NO:8)	VTTTQPSLTQTTVMETVTTVTTREQILVKHAQEELPPPPQKKRQITVD
Spectrin repeat 4 (SEQ ID NO:9)	SEIRKRLDVDITELHSWTRSEAVLQSPEFAIFRKEGNFSDLKEKVNAIER EKA EKFRKLQDASRSAQAL VEQMVNEGVNADSIKQASEQLNSRWIEFCQLLSERLNLWLEY
Spectrin repeat 5 (SEQ ID NO:10)	QNNIIAFYNQLQQL EQMTTTAENWLKIQPTTPSEPTAIKSQLKICKDEVN RLSGLQPQIERLKIQSIALK EKGQGP MFLDADFVAFTNHFKQVFSDVQAREKELQTIFD
Spectrin repeat 6 (SEQ ID NO:11)	TLPPMRYQETMSAIRTWVQQSETKLSIPQLSVTDYEIMEQRLGELQALQ SSLQEQQSGLYYLSTTVKEMS KKAPSEISRKYQSEFEEIEGRWKKLSSQLVEHCQKLEEQ
Spectrin repeat 7 (SEQ ID NO:12)	MNKL RKIQNHITLKKWMAEVDVFLKEEWPALGDSEILKKQLKQCRL VSDIQTIQPSLNSVNEGQKIK NEAEPEFASRLETTELKELNTQWDHMCQQVYARKEALKGG
Spectrin repeat 8 (SEQ ID NO:13)	LEKTVSLQKDLSEMHEWMTQAE EYLERDFEYKTPDELQKAVEEMKR AKEEAQQKEAKVKLLTESVNSVI AQAPPVAQEALKKELETLTNYQWLCTRLNGKCKTLEEV

Spectrin repeat 9 (SEQ ID NO:14)	WACWHELLSYLEKANKWLNEVEFKLKTENIPGGAEIESEVLDLENL MRHSEDNPNQIRILAQTLTDGG VMDELINEELETFSNRWRELHEEAARRQKLLEQS
Spectrin repeat 10 (SEQ ID NO:15)	IQSAQETEKSLHLIQESLTFIDKQLAAYIADKVDAAQMPQEAQKIQSDLT SHEISLEEMKKHNQGKEAAQ RVLSQIDVAQKKLQDVSMKFRL
Spectrin repeat 11 (SEQ ID NO:16)	FQKPANFEQRLQESKMILDEVKMHLPALETKSVEQEVVQSQLNHCVNL YKSLSEVKSEVEMVIKTGRQIV QKKQTENPKELDERVTALKLHYNELGAKVTERKQQLEKC
Spectrin repeat 12 (SEQ ID NO:17)	LKLSRKMRKEMNVLTEWLAATDMELTKRSAVEGMPSNLDSEVAWGK ATQKEIEKQKVHLKSITEVGEALK TVLGKKETLVEDKLSLLNSNWIAVTSRAEEWLNLLLEY
Spectrin repeat 13 (SEQ ID NO:18)	QKHMETFDQNVDHITKWIIQADTLLDESEKKKPQKEDVLKRLKAELN DIRPKVDSTRDQAANLMANRGD HCRKLVEPQISELNHRFAAISHRIKTGKASIPLK
Spectrin repeat 14 (SEQ ID NO:19)	ELEQFNSDIQKLLEPLEAEIQQGVNLKEEDFNKDMNEDNEGTVKELLQR GDNLQQRITDERKREEIKIKQ QLLQTKHNALKDLRSQRRKKALEI
Spectrin repeat 15 (SEQ ID NO:20)	SHQWYQYKRQADDLLKCLDDIEKKLASLPEPRDERKIKEIDRELQKKKE ELNAVRRQAEGLSGDGAAMAV EPTQIQLSKRWREIESKFAQFRRLNFAQ
L3 (20-mer linker) (SEQ ID NO:21)	IHTVREETMMVMTEDMPLFI
Spectrin repeat 16 (SEQ ID NO:22)	SYVPSTYLTEITHVSQALLEVEQLLNAPDLCAKDFEDLFKQEEESLKNIKD SLQQSSGRIDIIHSHKTAAL QSATPVERVKLQEALSQLDQWEKVNKMYKDRQGRFDRS
Spectrin repeat 17 (SEQ ID NO:23)	VEKWRRFHYDIKIFNQLWTEAEQFLRKTQIPENWEHAKYKWYKELQ DGIGQRQTVVRTLNTGEEIIQQ SSKTDASILQEKLGLSLNLRWQEVCKQLSDRKKRLEEQ
Spectrin repeat 18 (SEQ ID NO:24)	KNILSEFQRDLNEFVLWLEEADNIASIPLEPGKEQQLKEKLEQVKLLVEE LPLRQGILKQLNETGGPVLV SAPISPEEQDKLENKLLKQTNLQWIKVSRALPEKQGEIEAQIKDLGQL
Spectrin repeat 19 (SEQ ID NO:25)	EKKLEDLEEQLNHLLWLSPIRNQLEIYNQPNQEGPFDVQETEIADVQAK QPDVEEILSKGQHLYKEKPAT QPVKRKLEDLSSEWKAVNRLQLRAKQPD
Hinge 3 (SEQ ID NO:26)	APGLTTIGASPTQTVTLVTQPVVTKETAISKLEMPSSLMLE

Spectrin repeat (SEQ NO:27)	20	VPALADFNRAWTELTDWLSLLDQVIKSQRVMVGDLEDINEMIIKQKAT MQDLEQRRPQLEELITAAQNLK NKTSNQEARTIITDRIERIQNQWDEVQEHLQNRRLQQLNEM
Spectrin repeat (SEQ NO:28)	21	LKDSTQWLEAKEEAEQVLGQARAKLESWKEGPYTVDAIQKKITETKQL AKDLRQWQTNVDVANDLALKLL RDYSADDTRKVHMITENINASWRSIHKRVSEREAALLET
Spectrin repeat (SEQ NO:29)	22	HRLLQQFPLDLEKFLAWLTEAETTANVLQDATRKERLLEDKSGVKELM KQWQDLQGEIEAHTDVYHNLDE NSQKILRSLEGSDDAVLLQRRLDNMNFKWSELRKKSLNIRSHLEAS
Spectrin repeat (SEQ NO:30)	23	SDQWKRLHLSLQELLVWLQLKDDELSRQAPIGGDFPAVQKQNDVHRA FKRELKTKEPVIMSTLETVRIFL TEQPLEGLEKLYQEPRELPPEERAQNVTRLLRKQAEVNTWEKLNLS ADWQRKIDET
Spectrin repeat (SEQ NO:31)	24	LERLQELQEATDELDLKLRAEVIKGSWQPVGDLLIDSLQDHLEKVKA LRGEIAPLKENVSHVNDLARQL TTLGIQLSPYNLSTLEDLNTRWKLLQVAVEDRVRQLHE
Hinge 4 (SEQ ID NO:32)		AHRDFGPASQHFLSTSVQGPWERAISPKNKVPYYINHETQTTCWDHPKM TELYQSLADLNNVRFSAYRTAM KL
CR (SEQ ID NO:33)		RRLQKALCLDLLSLSAACDALDQHNLKQNDQPMILQIINCLTTIYDRL EQEHNVLNVPLCVDMLNWL LNVYDTGRTGRIRVLSFKTGIISLCKAHLEDKYRYLQVASTGFCDQ RRLGLLLHDSIQIPRQLGEVA SFGGSNIEPSVRSCFQFANNKPEIEAALFLDWMRLEPQSMVWLPVLHRV AAAEAKHQAKCNICKECPII GFRYRSLKHFNYDICQSCFFSGRVAKGHKMHYPMVEYCTPTTSGEDVR DFAKVLKNKFRTRKRYFAKHPRM GYLPVQTVLEGDNMET
C-term (SEQ ID NO:34)		PVTLINFWPVDSAPASSPQLSHDDTHSRIEHYASRLAEMENSNGSYLND SISPNESIDDEHLLIQHYCQS LNQDSPLSQPRSPAQILISLESEERGELERILADLEENRNQLAEYDRLKQ QHEHKGLSPLPSPPEMMPT SPQSPRDAELIAEAKLLRQHKGRLEARMQILEDHNLKQLESQHLRLRQLL EQPQAEAKVNGTTVSSPSTSL QRSDSSQPMLLRVVGSTSDSMGEEDLLSPPQDTSTGLEEVMEQLNNSF PSSRGRNTPGKPMREDTM

Table 4. Nucleotide Sequences of Dystrophin Domains

Description and Sequence Identifier	Sequence

5' untrans- lated region (SEQ ID NO:35)	gggattccct cactttcccc ctacaggact cagatctggg aggcaattac ctccggagaa 60 aaacgaatag gaaaaactga agtggttact ttttaaagc tgctgaagtt tgttggttc120 tcattgttt taagcctact ggagcaataa agtttgaaga acttttacca ggtttttt180 atcgctgcct tgatatacac tttcaaa 208
ABD1 (SEQ ID NO:36)	atgctttggt gggaagaagt agaggactgt tatgaaagag aagatgttca aaagaaaaca 60 ttcacaaaat gggtaaatgc acaattttct aagtttgga agcagcatat tgagaacctc120 ttcagtgaac tacaggatgg gaggcgcctc ctacacctcc tcgaaggcct gacagggcaa180 aaactgccaa aagaaaaagg atccacaaga gtcatgccc tgaacaatgt caacaaggca240 ctgcgggttt tgcagaacaa taatgttgat ttagtgaata ttggaagtac tgacatcgta300 gatggaaatc ataaactgac tcttggtttg atttggaata taatctcca ctggcaggtc360 aaaaatgtaa tgaaaaatat catggctgga ttgcaacaaa ccaacagtga aaagattctc420 ctgagctggg tccgacaatc aactcgtaat tatccacagg ttaatgtaat caactcacc480 accagctggt ctgatggcct ggctttgaat gctctcatcc atagtcatag gccagacctc540 tttgactgga atagtgtgtt ttgccagcag tcagccacac aacgactgga acatgcattc600 aacatcgcca gatatcaatt aggcataagag aaactactcg atcctgaaga tgttgatacc660 acctatccag ataagaagtc catctaatg tacatcacat cactctcca agtttgcct720 caacaagtga gcattgaagc catccaggaa gtggaa756
Hinge 1 (SEQ ID NO:37)	atgttgccaa ggccacctaa agtgactaaa gaagaacatt ttcagttaca tcataaatg 60 cactattctc aacagatcac ggctagctca gcacagggat atgagagaac ttctccct120 aagcctcgat tcaagagcta tgcttacaca caggctgctt atgtcaccac ctctgacct180 acacggagcc catttcttc acagcatttg gaagctctg aagacaagtc atttggcagt240 tcattgatgg agagt 255
Spectrin repeat 1 (SEQ ID NO:38)	gaagtaaacc tggaccgtta tcaaacagct ttagaagaag tattatctg gcttctttct 60 gctgaggaca cattgcaagc acaaggagag atttctaag atgtggaagt ggtgaaagac120 cagttcata ctcatgaggg gtacatgatg gatttgacag cccatcaggg cggggttggt180 aatattctac aattgggaag taagctgatt ggaacaggaa aattatcaga agatgaagaa240 actgaagtac aagagcagat gaatctccta aattcaagat gggaatgcct cagggtagct300 agcatggaaa aacaaagcaa ttacat327
Spectrin repeat 2 (SEQ ID NO:39)	agagttttaa tggatctcca gaatcagaaa ctgaaagagt tgaatgactg gctaacaaaa 60 acagaagaaa gaacaaggaa aatggaggaa gagcctcttg gacctgatct tgaagacctc120 aaacgccaag tacaacaaca taaggtgctt caagaagatc tagaacaaga acaagtcagg180 gtcaattctc tcaatcacat ggtgggtgta gttgatgaat ctagtggaga tcacgcaact240 gctgctttgg aagaacaact taaggtattg ggagatcgat gggcaaacat ctgtagatgg300 acagaagacc gctgggttct ttacaagac atc 333
Spectrin repeat 3 (SEQ ID NO:40)	cttctcaaat ggcaacgtct tactgaagaa cagtgccttt ttagtgcatt gctttcagaa 60 aaagaagatg cagtgaacaa gattcacaca actggcttta aagatcaaaa tgaaatgta120 tcaagtcttc aaaaactggc cgtttataaa gcggatctag aaaagaaaaa gcaatccatg180 ggcaactgt attcactcaa acaagatctt ctttaaacac tgaagaataa gtcagtgaac240 cagaagacgg aagcatggct ggataacttt gcccggtgtt gggataattt agtcaaaaa300 cttgaagaaga gtacagcaca gatttcacag gct 333
Hinge (SEQ ID NO:41)	2 gtcaccacca ctacgccatc actaacacag acaactgtaa tggaacagt aactacggtg 60 accacaaggg aacagatcct ggtaaagcat gtcaagagg aacttcacc accacctccc120 caaaagaaga ggcagattac tgtggat147

Spectrin repeat 4 (SEQ ID NO:42)	tctgaaatta ggaaaagggt ggatgttgat ataactgaac ttcacagctg gattactcgc 60 tcagaagctg tgttcagag tctgaattt gcaatcttc ggaaggaagg caacttctca120 gacttaaaag aaaaagtcaa tgccatagag cgagaaaaag ctgagaagtt cagaaaactg180 caagatgccca gcagatcagc tcaggccctg gtggaacaga tggatgaatga ggggtgtaat240 gcagatagca tcaacaagc ctcagaacaa ctgaacagcc ggtggatcga attctgccag300 ttgctaagt agagacttaa ctggctggag tat 333
Spectrin repeat 5 (SEQ ID NO:43)	cagaacaaca tcctcgctt ctataatcag ctacaacaat tggagcagat gacaactact 60 gctgaaaact ggtgaaaaa ccaaccacc accccatcag agccaacagc aattaaagt120 cagtaaaaaa ttgtaagga tgaagtcaac cggctatcag gtcttcaacc tcaattgaa180 cgattaataa ttcaagcat agccctgaaa gagaaaggac aaggacccat gtctctggat240 gcagactttg tggccttac aaatcattt aagcaagtct ttctgatgt gcaggccaga300 gagaaagagc tacagacaat tttgac327
Spectrin repeat 6 (SEQ ID NO:44)	actttgccac caatgcgcta tcaggagacc atgagtgcc tcaggacatg ggtccagcag 60 tcagaaacca aacttccat acctcaactt agtgtaccg actatgaaat catggagcag120 agactcgggg aattgcaggc ttacaaagt tctctgcaag agcaacaaag tggcctatac180 tatctcagca ccactgtgaa agagatgtcg aagaaagcgc cctctgaaat tagccggaaa240 tatcaatcag aattgaaga aattgaggga cgctggaaga agctctctc ccagctggt300 gagcattgtc aaaagctaga ggagcaa327
Spectrin repeat 7 (SEQ ID NO:45)	atgaataaac tccgaaaaa tcagaatcac atacaaccc tgaagaaatg gatggctgaa 60 gttgatgtt ttctgaagga ggaatggcct gcccttgggg attcagaaat ttaaaaaag120 cagctgaaac agtgcagact ttagtcagt gatattcaga caattcagcc cagtctaaac180 agtgtcaatg aaggtgggca gaagataaag aatgaagcag agccagagtt tgcttcgaga240 cttgagacag aactcaaaga acttaacact cagtgggatc acatgtgcc acaggtctat300 gccagaaagg aggccttgaa gggaggt327
Spectrin repeat 8 (SEQ ID NO:46)	ttggagaaaa ctgtaagcct ccagaaagat ctatcagaga tgcacgaatg gatgacacaa 60 gctgaagaag agtatcttga gagagattt gaataataaa ctccagatga attacagaaa120 gcagtgaag agatgaagag agctaaagaa gaggcccaac aaaaagaagc gaaagtgaag180 ctccttactg agtctgtaa tagtgtcata gctcaagctc cacctgtagc acaagaggcc240 ttaaaaaagg aactgaaac ttaaccacc aactaccagt ggctctgcac taggctgaat300 gggaaatgca agactttgga agaagtt327
Spectrin repeat 9 (SEQ ID NO:47)	tgggcatgtt ggcatgagtt attgtcatac ttggagaaag caaacaagtg gctaaatgaa 60 gtagaattta aacttaaac cactgaaaac attcctggcg gagctgagga aatctctgag120 gtgctagatt cactgaaaa ttgatgcga cattcagagg ataaccaaa tcagattcgc180 atattggcac agaccctaac agatggcgga gtcattgatg agctaatcaa tgaggaaact240 gagacattta attctgttg gagggaaact catgaagagg ctgtaaggag gcaaaagt300 cttgaaacaga gc312
Spectrin repeat 10 (SEQ ID NO:48)	atccagtctg cccaggagac tgaaaaatcc ttacacttaa tccaggagtc cctcacattc 60 attgacaagc agttggcagc ttatattgca gacaagggtg acgcagctca aatgcctcag120 gaagcccaga aaatccaatc tgattgaca agtcatgaga tcagtttaga agaaatgaag180 aaacataatc aggggaagga ggctgcccaa agagtctgt ctcagattga tgttcacag240 aaaaaattac aagatgtctc catgaagttt cgatta276

Spectrin repeat 11 (SEQ ID NO:49)	ttccagaaac cagccaattt tgagctgcgt ctacaagaaa gtaagatgat ttagatgaa 60 gtgaagatgc acttgcctgc attggaaaca aagagtgtgg aacaggaaagt agtacagtca120 cagctaaatc attgtgtgaa cttgtataaa agtctgagtg aagtgaagtc tgaagtggaa180 atgggtgataa agactggacg tcagattgta cagaaaaagc agacggaaaa tcccaaagaa240 cttcatgaaa gagtaacagc ttgaaattg cattataatg agctgggagc aaaggaaca300 gaaagaaagc aacagttgga gaaatgc327
Spectrin repeat 12 (SEQ ID NO:50)	ttgaaattgt cccgtaagat gcgaaaggaa atgaatgtct tgacagaatg gctggcagct 60 acagatatgg aattgacaaa gagatcagca gttgaaggaa tgcctagtaa tttgattct120 gaagtgcct ggggaaaggc tactcaaaaa gagattgaga aacagaaggt gcacgtgaag180 agtatcacag aggtaggaga ggccttgaaa acagttttgg gcaagaagga gacgttggtg240 gaagataaac tcagtcttct gaatagtaac tggatagctg tcacctcccg agcagaagag300 tggttaaate tttgttga atac 324
Spectrin repeat 13 (SEQ ID NO:51)	cagaaacaca tggaaacttt tgaccagaat gtggaccaca tcacaaagtg gatcattcag 60 gctgacacac ttttgatga atcagagaaa aagaaacccc agcaaaaaga agcgtgctt120 aagcgtttaa aggcagaact gaatgacata cgcccaagg tggactctac acgtgacca180 gcagcaaact tgatggcaaa ccgcggtgac cactgcagga aattagtaga gcccacaa240 tcagagctca accatcgatt tgcagccatt tcacacagaa ttaagactgg aaaggcctcc300 attcctttga ag312
Spectrin repeat 14 (SEQ ID NO:52)	gaattggagc agtttaactc agatatacaa aaattgcttg aaccactgga ggctgaaatt 60 cagcaggggg tgaatctgaa agaggaagac ttcaataaag atatgaatga agacaatgag120 ggtactgtaa aagaattgtt gcaagaggga gacaacttac aacaaagaat cacagatgag180 agaaagagag aggaataaaa gataaacag cagctgttac agacaaaaca taatgctctc240 aaggatttga ggtctcaaag aagaaaaaag gctctagaaa tt 282
Spectrin repeat 15 (SEQ ID NO:53)	tctcatcagt ggtatcagta caagaggcag gctgatgac tctgaaatg cttggatgac 60 attgaaaaaa aattagccag cctacctgag ccagagatg aaaggaaaa aaaggaaatt120 gatcggggat tgcagaagaa gaaagaggag ctgaatgcag tgcgtaggca agctgagggc180 ttgtctgagg atggggccgc aatggcagtg gagccaactc agatccagct cagcaagcgc240 tggcgggaaa ttgagagcaa attgctcag ttcgaagac tcaactttgc acaa294
L3 (20-mer linker)(SEQ ID NO:54)	attcacactg tccgtgaaga aacgatgatg gtgatgactg aagacatgcc tttgaaatt 60
Spectrin repeat 16 (SEQ ID NO:55)	tcttatgtgc ctttactta ttgactgaa atcactcatg ttcacaagc cctattagaa 60 gtggaacaac ttctaatgc tctgacctc tgtgctaagg actttgaaga tctctttaag120 caagaggagt ctctgaagaa tataaaagat agtctacaac aaagctcagg tcggattgac180 attattcata gcaagaagac agcagcattg caaagtgcaa cgctgttga aagggtgaag240 ctacaggaag ctctctcca gcttgatttc caatgggaaa aagtaacaa aatgtacaag300 gaccgacaag ggcgatttga cagatct327
Spectrin repeat 17 (SEQ ID NO:56)	gttgagaaat ggcggcgttt tcattatgat ataaagatat ttaatcagt gctaacagaa 60 gctgaacagt ttctcagaaa gacacaaatt cctgagaatt gggaacatgc taaatacaaa120 tggtatctta aggaactcca ggaatggcatt gggcagcggc aaactgttgt cagaacattg180 aatgcaactg gggaagaaat aattcagcaa tctcaaaaa cagatgccag tattctacag240

	gaaaaattgg gaagcctgaa tctgcggtgg caggaggtct gcaaacagct gtcagacaga300 aaaaagaggc tagaagaaca a321
Spectrin repeat 18 (SEQ ID NO:57)	aagaatatct tgtcagaatt tcaaagagat ttaatgaat ttgtttatg gttggaggaa 60 gcagataaca ttgctagtat cccacttgaa cctggaaaag agcagcaact aaaagaaaag120 cttgagcaag tcaagttact ggtggaagag ttgccctgc gccagggaat tctcaacaa180 ttaatgaaa ctggaggacc cgtgcttga agtgctcca taagcccaga agagcaagat240 aaactgaaa ataagctcaa gcagacaaat ctccagtga taaaggttc cagagcttta300 cctgagaaac aaggagaaat tgaagctcaa ataaaagacc ttgggcagct t 351
Spectrin repeat 19 (SEQ ID NO:58)	gaaaaaagc ttgaagacct tgaagagcag ttaatcattc tgctgctgtg gttatctcct 60 attaggaatc agttggaat ttataacaa ccaaccaag aaggaccatt tgacttcag120 gaaactgaaa tagcagttca agctaaacaa ccgatgtgg aagagatttt gtctaaagg180 cagcatttgt acaaggaaaa accagccact cagccagtga agaggaagtt agaagatctg240 agctctgagt ggaaggcggg aaaccgttta ctcaagagc tgagggcaaa gcagcctgac300 cta303
Hinge 3 (SEQ ID NO:59)	gctcctggac tgaccactat tggagcctct cctactcaga ctgttactct ggtgacacaa 60 cctgtggta ctaaggaaac tgccatctcc aaactagaaa tgccatcttc ctgtatgtg120 gag123
Spectrin repeat 20 (SEQ ID NO:60)	gtacctgctc tggcagattt caaccgggct tggacagaac ttaccgactg gttttctctg 60 cttgatcaag ttataaaac acagagggtg atggtgggtg acctgagga tatcaacgag120 atgatcatca agcagaaggc aacaatgcag gatttggaa agagggctcc ccagtggaa180 gaactcatta ccgtgcca aaatttga aaacaagacca gcaatcaaga ggctagaaca240 atcattacgg atcgaattga aagaattcag aatcagtggt atgaagtaca agaacacct300 cagaaccgga ggcaacagtt gaatgaaatg 330
Spectrin repeat 21 (SEQ ID NO:61)	ttaaaggatt caacacaatg gctggaagct aaggagaag ctgagcaggt cttaggacag 60 gccagagcca agcttgagtc atggaaggag ggtccctata cagtagatgc aatccaaaag120 aaaatcacag aaaccaagca gttggccaaa gacctccgc agtggcagac aatgtagat180 gtggcaaatg acttggcct gaaactctc cgggattatt ctgcagatga taccagaaaa240 gtccacatga taacagagaa tatcaatgcc tcttgagaa gcattcataa aagggtgagt300 gagcgagagg ctgcttggga agaaact327
Spectrin repeat 22 (SEQ ID NO:62)	catagattac tgcaacagtt cccctggac ctggaaaagt ttctgcctg gttacagaa 60 gctgaaacaa ctgcaaatgt cctacaggat gctaccgta aggaaggct cctagaagac120 tccaaggag taaaagagct gatgaacaa tggcaagacc tccaaggtga aattgaagct180 cacacagatg ttatcacia cctggatgaa aacagccaaa aaatctgag atccctggaa240 ggtccgatg atgcagtcct gttacaaaga cgttggata acatgaactt caagtggagt300 gaacttcgga aaaagtctct caacattagg tccatttgg aagccagt 348
Spectrin repeat 23 (SEQ ID NO:63)	tctgaccagt ggaagcgtct gcaccttct ctgcaggaac ttctggtgtg gctacagctg 60 aaagatgatg aattaagccg gcaggcacct attggaggcg acttccagc agttcagaag120 cagaacgatg tacatagggc ctcaagagg gaattgaaa cttaaagaacc tgtaatcatg180 agtactcttg agactgtacg aatattctg acagagcagc ctttggagg actagagaaa240 ctctaccagg agcccagaga gctgcctct gaggagagag cccagaatgt cactcggtt300 ctacgaaagc aggctgagga ggtcaatact gagtgggaaa aattgaacct gcactccgct360

	gactggcaga gaaaaataga tgagacc387
Spectrin repeat 24 (SEQ ID NO:64)	cttgaaagac tccaggaact tcaagaggcc acggatgagc tggacctcaa gctgcgcaa 60 gctgaggtga tcaagggatc ctggcagccc gtgggagcgc tctcattga ctctctcaa120 gatcacctcg agaaagtcaa ggcacttcga ggagaaattg cgctctgaa agagaacgtg180 agccacgtca atgaccttgc tggcagctt accactttgg gcattcagct ctaccgtat240 aacctcagca ctctggaaga cctgaacacc agatggaagc ttctgcaggt ggccgtcag300 gaccgagtca ggcagctgca tgaa 324
Hinge 4 (SEQ ID NO:65)	gccccacagg actttggtcc agcatctcag cactttctt ccaagtctgt ccagggtccc 60 tgaggagagag ccatctcgcc aaacaaagtg cctactata tcaaccacga gactcaaaac120 acttgctggg accatcccaa aatgacagag ctctaccagt cttagctga cctgaataat180 gtcagattct cagcttatag gactgcatg aaactc216
CR (SEQ ID NO:66)	cgaagactgc agaaggccct ttgcttgat ctcttgagcc tgtcagctgc atgtgatgcc 60 ttggaccagc acaacctcaa gcaaatgac cagcccatgg atatctgca gattattaat120 tgtttgacca ctatttatga cgcctggag caagagcaca acaatttgg caacgtccct180 ctctgcgtgg atatgtgtct gaactggctg ctgaatgtt atgatacggg acgaacagg240 aggatccgtg tctgtctt taaaactggc atcattccc tgtgtaaagc acatttgaa300 gacaagtaca gatactttt caagcaagt gcaagttcaa caggattttg tgaccagcgc360 aggctgggccc tcttctgca tgattctatc caaattccaa gacagttggg tgaagttgca420 tctttgggg gcagtaacat tgagccaagt gtccggagct gttccaatt tgtaataat480 aagccagaga tgaagcggc cctcttcta gactggatga gactggaacc ccagtcctg540 gtgtggctgc ccgtctgca cagagtggct gctgcagaaa ctgccaagca tcaggccaaa600 tgtaacatct gcaaagagtg tccaatcatt ggattcaggt acaggagtct aaagcacttt660 aattatgaca tctgcaaag ctgcttttt tctggtcgag ttgcaaagg ccataaaatg720 cactatccca tgggtggaata ttgactccg actacatcag gagaagatgt tcgagacttt780 gccaaggtac taaaaaaca atttgaacc aaaaggtatt ttgcaagca tcccgaatg840 ggctacctgc cagtgcagac tgtcttagag ggggacaaca tggaaact
C-term (SEQ ID NO:67)	cccgttactc tgatcaact ctggccagta gattctgcgc ctgctcgtc cctcagctt 60 tcacacgatg atactcattc acgcatgaa cattatgcta gcaggctagc agaatggaa120 aacagcaatg gatcttatct aaatgatagc atctctcta atgagagcat agatgatgaa180 catttgtaa tccagcatta ctgcaaaagt ttgaaccagg actccccct gagccagcct240 cgtagtctg cccagatctt gatttctta gagagtgagg aaagagggga gctagagaga300 atcctagcag atcttgagga agaaaacagg aatctgcaag cagaatatga ccgtctaaag360 cagcagcagc aacataaagg cctgtccca ctgcccgtcc ctctgaaat gatgccacc420 tctcccaga gtccccggga tctgagctc attgctgagg ccaagctact gctcaaac480 aaaggccgccc tggagccag gatgcaaatc ctggaagacc acaataaaca gctggagtca540 cagttacaca ggctaaggca gctgctggag caaccccagg cagaggccaa agtgaatggc600 acaacggtgt cctctcttc tacctctcta cagaggtccg acagcagtc gcctatgctg660 ctccgagtgg ttggcagtca aattcggac tccatgggtg aggaagatct tctagctct720 ccccaggaca caagcacagg gtagaggag gtgatggagc aactcaaca ctcttcct780 agttaagag gaagaaatac cctggaag ccaatgagag aggacacaat gtgag

The present disclosure is directed to a miniaturized dystrophin polypeptide that is smaller than the full-length dystrophin protein, i.e., isoform Dp427m, and that is not identical to

the naturally occurring dystrophin protein isoforms, or a nucleic acid molecule comprising a nucleotide sequence encoding the miniaturized dystrophin polypeptide. When the present disclosure discloses miniaturized dystrophin polypeptides, the present disclosure also discloses nucleic acid molecule comprising a nucleotide sequence encoding the corresponding disclosed miniaturized dystrophin polypeptide, and vice versa. In some embodiments, the nucleic acid molecule encoding the miniaturized dystrophin polypeptide is suitable for gene therapy. Accordingly, the nucleic acid molecule encoding the miniaturized dystrophin polypeptide is constructed not only to fit into a gene therapy vector, e.g., AAV vector, or to be suitable for recombinant expression, but also to reduce any unwanted immune response (e.g., humoral immune response and/or cellular immune response, e.g., CD4 and/or CD8) against the miniaturized dystrophin polypeptide when administered or expressed in vivo. In some embodiments, the miniaturized dystrophin polypeptide of the present disclosure comprises from N terminus to C terminus hinge 1 (H1) domain, spectrin repeat 1 (R1) domain, spectrin repeat 3 (R3) domain, hinge 2 (H2) domain, spectrin repeat 16 (R16) domain, spectrin repeat 17 (R17) domain, spectrin repeat 24 (R24) domain, and hinge 4 (H4) domain of dystrophin. In some embodiments, the miniaturized dystrophin polypeptide does not comprise spectrin repeat 2 (R2) domain of dystrophin. In other embodiments, the miniaturized dystrophin polypeptide does not comprise spectrin repeat 2 (R2) domain and (1) spectrin repeat 4 (R4) domain, (2) spectrin repeat 5 (R5) domain, (3) spectrin repeat 6 (R6) domain, (4) spectrin repeat 7 (R7) domain, (5) spectrin repeat 8 (R8) domain, (6) spectrin repeat 9 (R9) domain, (7) spectrin repeat 10 (R10) domain, (8) spectrin repeat 11 (R11) domain, (9) spectrin repeat 12 (R12) domain, (10) spectrin repeat 13 (R13) domain, (11) spectrin repeat 14 (R14) domain, (12) spectrin repeat 15 (R15) domain, (13) spectrin repeat 18 (R18) domain, (14) spectrin repeat 19 (R19) domain, (15) spectrin repeat 20 (R20) domain, (16) spectrin repeat 21 (R21) domain, (17) spectrin repeat 22 (R22) domain, or (18) spectrin repeat 23 (R23) domain, or (19) any combination thereof. In some embodiments, the miniaturized dystrophin polypeptide does not comprise spectrin repeat 2 (R2) domain, spectrin repeat 4 (R4) domain, spectrin repeat 5 (R5) domain, spectrin repeat 6 (R6) domain, spectrin repeat 7 (R7) domain, spectrin repeat 8 (R8) domain, spectrin repeat 9 (R9) domain, spectrin repeat 10 (R10) domain, spectrin repeat 11 (R11) domain, spectrin repeat 12 (R12) domain, spectrin repeat 13 (R13) domain, spectrin repeat 14 (R14) domain, spectrin repeat 15 (R15) domain, spectrin repeat

18 (R18) domain, spectrin repeat 19 (R19) domain, spectrin repeat 20 (R20) domain, spectrin repeat 21 (R21) domain, spectrin repeat 22 (R22) domain, and spectrin repeat 23 (R23) domain.

In certain embodiments, the miniaturized dystrophin polypeptide of the present disclosure comprises from N terminus to C terminus hinge 1 (H1) domain, spectrin repeat 1 (R1) domain, spectrin repeat 3 (R3) domain, hinge 2 (H2) domain, spectrin repeat 16 (R16) domain, spectrin repeat 17 (R17) domain, spectrin repeat 24 (R24) domain, and hinge 4 (H4) domain of dystrophin, wherein the R1 domain is directly fused to R3 domain by a peptide bond, e.g., See miniaturized dystrophin polypeptide BXA-027741.

In some embodiments, the miniaturized dystrophin polypeptide of the present disclosure comprises from N terminus to C terminus hinge 1 (H1) domain, spectrin repeat 1 (R1) domain, spectrin repeat 3 (R3) domain, hinge 2 (H2) domain, spectrin repeat 16 (R16) domain, spectrin repeat 17 (R17) domain, spectrin repeat 24 (R24) domain, and hinge 4 (H4) domain of dystrophin, wherein the R1 domain and the R3 domain are fused by amino acids ARG-VAL (RV).

In some embodiments, the miniaturized dystrophin polypeptide of the present disclosure comprises from N terminus to C terminus hinge 1 (H1) domain, spectrin repeat 1 (R1) domain, spectrin repeat 3 (R3) domain, hinge 2 (H2) domain, spectrin repeat 16 (R16) domain, spectrin repeat 17 (R17) domain, spectrin repeat 24 (R24) domain, and hinge 4 (H4) domain of dystrophin, wherein the H2 domain and the R16 domain are fused by a linker. The linker can be any linker known in the art. In other embodiments, the linker can be selected from any linker disclosed herein. In other embodiments, the linker can be a linker in Section 5.3.2. In some embodiments, the linker can comprise an amino acid sequence at least about 80%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 75 (IHTVREE TMMVMTEDMP LEI).

In some embodiments, the miniaturized dystrophin polypeptide of the present disclosure comprises from N terminus to C terminus hinge 1 (H1) domain, spectrin repeat 1 (R1) domain, spectrin repeat 3 (R3) domain, hinge 2 (H2) domain, spectrin repeat 16 (R16) domain, spectrin repeat 17 (R17) domain, spectrin repeat 24 (R24) domain, and hinge 4 (H4) domain of dystrophin, wherein the H2 domain and the R16 domain are fused by a linker. In some embodiments, the linker can comprise an amino acid sequence at least about

80%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NOs: 74-75 in combination (SEAQ IHTVREE TMMVMTEDMP LEI).

5 In some embodiments, the miniaturized dystrophin polypeptide of the present disclosure comprises from N terminus to C terminus hinge 1 (H1) domain, spectrin repeat 1 (R1) domain, spectrin repeat 3 (R3) domain, hinge 2 (H2) domain, spectrin repeat 16 (R16) domain, spectrin repeat 17 (R17) domain, spectrin repeat 24 (R24) domain, and hinge 4 (H4) domain of dystrophin, wherein the H2 domain and the R16 domain are fused by a linker. In some embodiments, the linker can comprise an amino acid sequence at least about
10 80%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 74 (SEAQ).

In other embodiments, the miniaturized dystrophin polypeptide of the present disclosure comprises from N terminus to C terminus hinge 1 (H1) domain, spectrin repeat 1 (R1) domain, spectrin repeat 3 (R3) domain, hinge 2 (H2) domain, spectrin repeat 16 (R16)
15 domain, spectrin repeat 17 (R17) domain, spectrin repeat 24 (R24) domain, and hinge 4 (H4) domain of dystrophin, wherein the R1 domain and the R3 domain are fused by amino acids ARG-VAL and wherein the H2 domain and the R16 domain are fused by a linker, e.g., SEQ ID NOs: 74-75 in combination.

In some embodiments, the miniaturized dystrophin polypeptide of the present disclosure
20 comprises from N terminus to C terminus hinge 1 (H1) domain, spectrin repeat 1 (R1) domain, spectrin repeat 3 (R3) domain, hinge 2 (H2) domain, spectrin repeat 16 (R16) domain, spectrin repeat 17 (R17) domain, spectrin repeat 24 (R24) domain, and hinge 4 (H4) domain of dystrophin, wherein the R1 domain and the R3 domain are fused by amino acids ARG-VAL, wherein the H2 domain and the R16 domain are fused by a linker, e.g.,
25 SEQ ID NOs: 74-75 in combination, and wherein (i) the H1 domain and the R1 domain are fused directly, (ii) the R3 domain and the H2 domain are fused directly, (iii) the R16 and R17 domains are fused directly, (iv) the R17 and R24 domains are fused directly, or (v) the R24 and H4 domains are fused directly, or (vi) any combination thereof. In some embodiments, the miniaturized dystrophin polypeptide useful for the present disclosure
30 comprises from N terminus to C terminus: hinge 1 (H1) domain, spectrin repeat 1 (R1) domain, spectrin repeat 3 (R3) domain, hinge 2 (H2) domain, spectrin repeat 16 (R16) domain, spectrin repeat 17 (R17) domain, spectrin repeat 24 (R24) domain, and hinge 4

(H4) domain of dystrophin, wherein the R1 domain and the R3 domain are fused by amino acids ARG-VAL, wherein the H2 domain and the R16 domain are fused by a linker, e.g., SEQ ID NOs: 74-75 in combination, and wherein (i) the H1 domain and the R1 domain are fused directly, (ii) the R3 domain and the H2 domain are fused directly, (iii) the R16 and R17 domains are fused directly, (iv) the R17 and R24 domains are fused directly, and (v) the R24 and H4 domains are fused directly. In some embodiments, the miniaturized dystrophin polypeptide useful for the present disclosure further comprises an ABD1 domain (optionally at the N terminus) and/or a CR domain (optionally at the N terminus). In some embodiments, the miniaturized dystrophin polypeptide consists essentially of or consists of, from N terminus to C terminus, an ABD1 domain, a H1 domain, an R1 domain, amino acids RV, an R3 domain, an H2 domain, the amino acid sequence as set forth in SEQ ID NOs: 74-75 in combination, an R16 domain, an R17 domain, an R24 domain, an H4 domain, and a CR domain of dystrophin.

Each domain in the miniaturized dystrophin polypeptides can have one or more changes from the corresponding wild-type domain.

For example, the miniaturized dystrophin BXA-196477 consists of the following protein domains in order:

Table 5A. Amino acid sequence and domain structure of miniaturized dystrophin polypeptide BXA-196477.

SEQ ID NO:	Description	Sequence
68	ABD1	MLWWEVEDCYEREDVQKKTFTKWVNAQFSKFGKQHIEN LFSDLQDGRRLDLLEGLTGQKLPKEKGSTRVHALNNVNK ALRVLQNNNVDLVNIGSTDIVDGNHKLTLGLIWNILHWQV KNVMKNIMAGLQQTNSEKILLSWVRQSTRNYPQVNVINFT TSWSDGLALNALIHSHRPDLFDWNSVVCQQSATQRLEHAF NIARYQLGIEKLLDPEDVDTTYPDKKSILMYITSLFQVLPQQ VSIEAIQEVE
69	Hinge 1	MLPRPPKVTKEEHFQLHHQMHYSQQITVSLAQGYERTSSP KPRFKSYAYTQAAYVTTSDPTRSPFPSQHLEAPEDKSFGSSL MES
70	Spectrin-1	EVNLDRYQTALEEVLSWLLSAEDTLQAQGEISNDVEVVKD QFHTHEGYMMDLTAHQGRVGNILQLGSKLIGTGKLSDEEE TEVQEQMNLNLSRWECLRVASMEKQSNLH
71	L1 (RV-Linker)	RV

72	Spectrin-3	LLKWQRLTEEQCLFSAWLSEKEDAVNKIHTTGFKDQNEML SSLQKLAVLKADLEKKKQSMGKLYSLKQDLLSTLKNKSVT QKTEAWLDNFARCWDNLVQKLEKSTAQISQA
73	Hinge 2	VTQTQPSLTQTTVMETVTTVTTREQILVKHAQEELPPPPQK KRQITVD
74	L2 (SEAQ linker)	SEAQ
75	L3 (20-mer linker)	IHTVREETMMVMTEDMPLFI
76	Spectrin-16	SYVPSTYLTEITHVSQALLEVEQLLNAPDLCAKDFEDLFKQ EESLKNIKDSLQSSGRIDIIHKKTAALQSATPVERVKLQE ALSQLDQWEKVNKMYKDRQGRFDRS
77	Spectrin-17	VEKWRRFHYDIKIFNQWLTEAEQFLRKTQIPENWEHAKYK WYLKELQDGIGQRQTVVRTLNATGEEIIQSSKTDASILQE KLGSLNLRWQEVCKQLSDRKKRLEEQ
78	Spectrin-24	LERLQELQEATDELDLKLRAEVIKGSWQPVGDLLIDSLQD HLEKVKALRGEIAPLKENVSHVNDLARQLTTGLIQLSPYNL STLEDLNRWKLQVAVEDRVRQLHE
79	Hinge 4	AHRDFGPASQHFLLSTSVQGPWERAISPKNVPYYINHETQTT CWDHPKMTLEYQSLADLNNVRFSAYRTAMKL
80	CR	RRLQKALCLDLLSLSAACDALDQHNLKQNDQPMQDILQIINC LTTIYDRLEQEHNNLVNVPLCVDMLNWLNNVYDTGRTG RIRVLSFKTGIIISLCKAHLEDKYRYLQKQVASTGFCQRRRL GLLLHDSIQIPRQLGEVASFGGSNIEPSVRSCFQFANNKPEIE AALFLDWMRLPQSMVWLPVLHRVAAAETAKHQAKCNIC KECPIIGFRYRSLKHFNIDICQSCFFSGRVAKGHKMHYPMV EYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGYLPV QTVLEGDNMETDTM

In some embodiments, the H1 domain in the miniaturized dystrophin polypeptide is an amino acid sequence at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 69. In some embodiments, the R1 domain is an amino acid sequence at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 70. In some embodiments, the R3 domain is an amino acid sequence at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 72. In some embodiments, the H2 domain is an amino acid sequence at least about 70%, at least about 75%, at least

about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 73. In some embodiments, the R16 domain is an amino acid sequence at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 76. In some embodiments, the R17 domain is an amino acid sequence at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 77. In some embodiments, the R24 domain is an amino acid sequence at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 78. In some embodiments, the H4 domain is an amino acid sequence at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 79. In some embodiments, the miniaturized dystrophin polypeptide further comprises at the N terminus a polypeptide sequence at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 68. In some embodiments, the miniaturized dystrophin polypeptide further comprises at the C terminus a polypeptide sequence at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 80.

Table 5B. Amino acid sequence and domain structure of miniaturized dystrophin polypeptide BXA-027743.

SEQ ID NO:	Description	Sequence
81	ABD1	MLWWEIVEDCYEREDVQKKTFTKWVNAQFSKFGKQHIE LFSDLQDGRRLDLLEGLTGQKLPKEKGSTRVHALNNVNK ALRVLQNNNVDLVNIGSTDIVDGNHKLTLGLIWNILHWQV KNVMKNIMAGLQQTNSEKILLSWVRQSTRNYPQVNVINFT TSWSDGLALNALIHSHRPDLFDWNSVVCQQSATQRLEHAF

		NIARYQLGIEKLLDPEDVDTTYPDKKSILMYITSLFQVLPQQ VSIEAIQEVE
82	Hinge 1	MLPRPPKVTKEEHFQLHHQMHYSQQITVSLAQQGYERTSSP KPRFKSYAYTQAAYVTTSDPTRSPFPSQHLEAPEDKSFSSS MES
83	Spectrin-1	EVNLDTRYQTAL EEVLSWLLSAEDTLQAQGEISNDVEVVKD QFHTHEGYMMDLTAHQGRVGNILQLGSKLIGTGKLSDEE TEVQEQMNNLSRWECRLVASMEKQSNLH
84	L1	IHTVREETMMVMTEDMPL EI
85	Spectrin-16	SYVPSTYLTEITHVSQALLEVEQLLNAPDLCAKDFEDLFKQ EESLKNIKDSLQSSGRIDIIHKKTAALQSATPVERVKLQE ALSQLD FQWEKVNKMYKDRQGRFDRS
86	Spectrin-17	VEKWRRFHYDIKIFNQWLTEAEQFLRKTQIPENWEHAKYK WYLKELQDGIGQRQTVVRTLNATGEEIIQQSSKTDASILQE KLGSLNLRWQEVCKQLSDRKKRLEEQ
87	Hinge 3	APGLTTIGASPTQTVTLVTQPVVTKETAISKLEMPSSLMLE
88	Spectrin-23	SDQWKRLHLSLQELLVWLQLKDDELSRQAPIGGDFPAVQK QNDVHRAFKRELKTKPEVIMSTLETVRIFLTEQPLEGLEKL YQEPRELPP EERAQNVTRLLRKQAE EVNTEWEKLNLSAD WQRKIDET
89	Spectrin-24	LERLQELQEATDELDLKL RQAEVIKGSWQPVGDLLIDSLQD HLEKVKALRGEIAPLKENVSHVNDLARQLTTLGIQLSPYNL STLEDLNTRWKLLQVAVEDRVRQLHE
90	Hinge 4	AHRDFGPASQHFLSTSVQGP WERAISP NKVPYYINHETQTT CWDHPKMT ELYQSLADLNNVRFSA YRTAMKL
91	CR/C-term	RRLQKALCLDLSLSAACDALDQHNLKQNDQPM DILQIINC LTTIYDRLEQEHNNLVN VPLCVD MCLNWLLNVYDTGR TG RIRVLSFKTGIISLCKAHLEDKYRYL FKQVASSTGFC DQRR L GLLLHDSIQIPRQLGEVASFGGSNIEPSVRSCFQFANNKPEIE AALFLDWMRLEPQSMVWLPVLHRVAAAETA KHQAKCNIC KECPIIGFRYRSLKHFN YDICQSCFFSGRVAKGHKMHPMV EYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGYLPV QTVLEGDNMETPAQILISLESEERGELERILADLEEENRN LQ AEYDRLKQQHEHKGLSPLSP PPTM

In some embodiments, the present disclosure provides a miniaturized dystrophin polypeptide comprising spectrin repeat 1 (R1) domain and spectrin 16 (R16) domain, wherein the R1 domain and the R16 domain are fused by a linker comprising the amino acid sequence set forth in SEQ ID NO: 84 (IHTVREETMMVMTEDMPL EI). In some

5 embodiments, the present disclosure provides a miniaturized dystrophin polypeptide comprising from N terminus to C terminus a hinge 1 (H1) domain, spectrin repeat 1 (R1) domain, spectrin repeat 16 (R16) domain, spectrin repeat 17 (R17) domain, a hinge 3 (H3) domain, spectrin repeat 23 (R23) domain, spectrin repeat 24 (R24) domain, and hinge 4

(H4) domain of dystrophin, wherein the R1 domain and the R16 domain are fused by the amino acid sequence set forth in SEQ ID NO: 84 (IHTVREETMMVMTEDEMPLEI).

In some embodiments, the present disclosure includes a nucleic acid molecule comprising a nucleotide sequence, which encodes a miniaturized dystrophin polypeptide comprising spectrin repeat 1 (R1) domain and spectrin 16 (R16) domain, wherein the R1 domain and the R16 domain are fused by a linker comprising the amino acid sequence set forth in SEQ ID NO: 84 (IHTVREETMMVMTEDEMPLEI). In some embodiments, the present

disclosure provides a nucleic acid molecule comprising a nucleotide sequence, which encodes a miniaturized dystrophin polypeptide comprising from N terminus to C terminus a hinge 1 (H1) domain, spectrin repeat 1 (R1) domain, spectrin repeat 16 (R16) domain, spectrin repeat 17 (R17) domain, a hinge 3 (H3) domain, spectrin repeat 23 (R23) domain, spectrin repeat 24 (R24) domain, and hinge 4 (H4) domain of dystrophin, wherein the R1 domain and the R16 domain are fused by the amino acid sequence set forth in SEQ ID NO: 84 (IHTVREETMMVMTEDEMPLEI).

In some embodiments, the miniaturized dystrophin polypeptide of the present disclosure has the following characteristics: (i) the H1 domain and the R1 domain are fused directly, (ii) the R16 and R17 domains are fused directly, (iii) the R17 and the H3 domains are fused directly, (iv) the R23 and R24 domains are fused directly, or (v) the R24 and H4 domains are fused directly, or (vi) any combination thereof. In some embodiments, the miniaturized dystrophin polypeptide for the present disclosure does not comprise a spectrin repeat 2 (R2) domain, spectrin repeat 3 (R3) domain, spectrin repeat 4 (R4) domain, spectrin repeat 5 (R5) domain, spectrin repeat 6 (R6) domain, spectrin repeat 7 (R7) domain, spectrin repeat 8 (R8) domain, spectrin repeat 9 (R9) domain, spectrin repeat 10 (R10) domain, spectrin repeat 11 (R11) domain, spectrin repeat 12 (R12) domain, spectrin repeat 13 (R13) domain, spectrin repeat 14 (R14) domain, spectrin repeat 15 (R15) domain, spectrin repeat 18 (R18) domain, spectrin repeat 19 (R19) domain, spectrin repeat 20 (R20) domain, spectrin repeat 21 (R21) domain, and/or spectrin repeat 22 (R22) domain. In some embodiments, the miniaturized dystrophin polypeptide further comprises an ABD1 domain and/or a CR domain.

In some embodiments, the miniaturized dystrophin polypeptide comprises, consists essentially of, or consists of, from N terminus to C terminus, the ABD1 domain, the H1 domain, the R1 domain, the amino acid sequence set forth in SEQ ID NO: 84

(IHTVREETMMVMTEDMPLEI), the R16 domain, the R17 domain, the H3 domain, the R23 domain, the R24 domain, the H4 domain, and the CR domain of dystrophin.

In some embodiments, the H1 domain useful for the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 70%, at least about 75%, at least about

5 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 82. In other embodiments, the R1 domain useful for the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about
10 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 83.

In some embodiments, the R16 domain useful for the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at
15 least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 85.

In some embodiments, the R17 domain useful for the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at
20 least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 86.

In other embodiments, the H3 domain useful for the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at
25 least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 87.

In some embodiments, the R23 domain useful for the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at
30 least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 88.

In other embodiments, the R24 domain useful for the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ

5 ID NO: 89.

In some embodiments, the H4 domain useful for the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ

10 ID NO: 90.

In some embodiments, the miniaturized dystrophin polypeptide further comprises at the N terminus an amino acid sequence at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 81.

15 In some embodiments, the miniaturized dystrophin polypeptide further comprises at the C terminus an amino acid sequence at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 91.

In some embodiments, a nucleic acid sequence encoding each domain can be the following:

20 Table 6A: Nucleotide sequence (and domain structure) encoding miniaturized dystrophin polypeptide BXA-196477.

SEQ ID NO:	Description	Nucleotide Sequence
92	5' UTR	CCGCCTTCGGCACCATTCTCACGACACCCAAATATGGCACG GGTGAGGAATGGTGGGGAGTTATTTTAGAGCGGTGAGGAAG GTGGGCAGGCAGCAGGTGTTGGCGCTCTAAAAATAACTCCCG GGAGTTATTTTAGAGCGGAGGAATGGTGGACACCCAAATAGG CGACGGTTCCTCACCCGTCGCCATATTTGGGTGTCCGCCCTCG GCCGGGGCCGCATTCTGGGGCCGGGCGGTGCTCCCGCCCGC CTCGATAAAAGGCTCCGGGGCCGGCGGCGGCCACGAGCTAC CCGGAGGAGCGGGAGGCTCTAGACCACC
93	ABD1	ATGCTGTGGTGGGAGGAAGTGAAGATTGCTACGAGCGCGA GGACGTGCAGAAGAAAACCTTCACCAAATGGGTGAACGCCC AGTTCAGCAAGTTCGGCAAGCAGCACATCGAGAACCTGTTCA GCGACCTGCAGGACGGCAGACGGCTGCTGGACCTGCTGGAA GGCCTGACCGGCCAGAAGCTGCCCAAAGAGAAGGGCAGCAC

		CAGAGTGCACGCCCTGAACAACGTGAACAAGGCCCTGCGGGT GCTGCAGAACAACAACGTGGACCTGGTGAACATCGGCAGCA CCGACATCGTGGACGGCAACCACAAGCTGACCCTGGGCCTGA TCTGGAACATCATCCTGCACTGGCAGGTCAAAAACGTGATGA AGAACATCATGGCCGGCCTGCAGCAGACCAACAGCGAGAAG ATCCTGCTGAGCTGGGTGCGCCAGAGCACCCGGAACCTACCCC CAGGTCAACGTGATCAACTTCACCACCTCTTGGAGCGACGGC CTGGCCCTGAACGCCCTGATCCACAGCCACCGGCCCGACCTG TTCGACTGGAACAGCGTGGTCTGCCAGCAGAGCGCCACCCAG CGGCTGGAACACGCCTTCAATATCGCCAGATACCAGCTGGGC ATCGAGAAGCTGCTGGATCCCGAGGACGTGGACACCACCTAC CCCGACAAGAAATCCATCCTGATGTATATCACCAGCCTGTTC CAGGTGCTGCCCCAGCAGGTGTCCATCGAGGCCATCCAGGAA GTGGAA
94	Hinge 1	ATGCTGCCCCAGACCCCCCAAAGTGACCAAAGAGGAACACTTC CAGCTGCACCACCAGATGCACTACAGCCAGCAGATCACCGTG TCCCTGGCCCAGGGCTACGAGAGAACCAGCAGCCCCAAGCCC CGGTTCAAGAGCTACGCCTATACCCAGGCCGCCTACGTGACC ACCAGCGACCCTACCAGAAGCCCATTCCCCAGCCAGCATCTG GAAGCCCCCGAGGACAAGAGCTTCGGCAGCAGCCTGATGGA AAGC
95	Spectrin-1	GAAGTGAACCTGGACAGATACCAGACCGCCCTGGAAGAGGT GCTGAGCTGGCTGCTGAGCGCCGAGGATACACTGCAGGCTCA GGGCGAGATCAGCAACGACGTGGAAGTCGTGAAGGACCACT TCCACACCCACGAGGGCTACATGATGGACCTGACAGCCCACC AGGGCAGAGTGGGCAACATCCTGCAGCTGGGCTCCAAGCTGA TCGGCACCGGCAAGCTGAGCGAGGACGAAGAGACAGAGGTG CAGGAACAGATGAACCTGCTGAACAGCAGATGGGAGTGCCT GCGGGTGGCCAGCATGGAAAAGCAGAGCAACCTGCAT
96	L1 (RV- Linker)	AGGGTC
97	Spectrin-3	CTGCTGAAGTGGCAGCGGCTGACCGAGGAACAGTGCCTGTTT AGCGCCTGGCTGTCCGAGAAAGAGGACGCCGTGAACAAGAT CCACACCACCGGCTTCAAGGACCAGAACGAGATGCTGAGCA GCCTGCAGAACTGGCCGTGCTGAAGGCCGACCTGGAAAAG AAAAAGCAGTCCATGGGCAAGCTGTACTCCCTGAAGCAGGAC CTGCTGTCCACCCTGAAGAACAAGAGCGTGACCCAGAAAACC GAGGCCTGGCTGGACAACTTCGCCCGGTGCTGGGACAACCTG GTGCAGAAGCTGGAAAAGTCCACCGCCCAGATCTCCAGGCC
98	Hinge 2	GTGACCACAACACAGCCCAGCCTGACCCAGACCACCGTGATG GAAACCGTGACAACAGTGACCACCCGGGAACAGATCCTCGTG AAGCACGCCCAGGAAGAACTGCCCCCTCACCCCCCAGAAG AAACGGCAGATCACAGTGGAC
99	L2 (SEAQ linker)	AGTGAAGCTCAG
100	L3 (20-mer linker)	ATCCACACCGTGCGGGAAGAGACAATGATGGTCATGACAGA GGACATGCCCCTGGAAATC

101	Spectrin-16	AGCTACGTGCCCAGCACCTACCTGACCGAGATCACCCATGTG TCCCAGGCCCTGCTGGAAGTGGAACAGCTGCTGAACGCCCC GACCTGTGCGCCAAGGATTTTCGAGGACCTGTTCAAGCAGGAA GAGAGCCTGAAGAATATCAAGGATAGCCTGCAGCAGAGCAG CGGCCGGATCGACATCATCCACAGCAAGAAAACAGCCGCCCT GCAGAGCGCCACCCCCGTGGAAAGAGTGAAACTGCAGGAAG CCCTGTCCCAGCTGGACTTCCAGTGGGAGAAAGTGAACAAAA TGTACAAGGACCGGCAGGGCAGATTTCGACCGCAGC
102	Spectrin-17	GTGGAAAAGTGGCGGCGGTTCCACTACGACATCAAGATCTTC AACCAGTGGCTGACAGAGGCCGAGCAGTTCCTGAGAAAGAC CCAGATCCCCGAGAACTGGGAGCACGCCAAGTACAAGTGGT ATCTGAAAGAGCTGCAGGACGGCATCGGCCAGAGACAGACA GTCGTGCGGACCCTGAATGCCACCGGCGAGGAAATCATCCAG CAGTCCAGCAAGACCGACGCCAGCATTCTGCAGGAAAAGCTG GGCAGCCTGAACCTGCGGTGGCAGGAAGTGTGCAAGCAGCT GTCCGACCGGAAGAAGCGGCTGGAAGAACAG
103	Spectrin-24	CTGGAACGGCTGCAGGAAGTGCAGGAGGCCACCGACGAGCT GGACCTGAAACTGAGACAGGCCGAAGTGATCAAGGGCAGCT GGCAGCCAGTGGGCGACCTGCTGATCGACTCCCTGCAGGACC ATCTGGAAAAAGTGAAGGCCCTGAGAGGCGAGATCGCCCC CTGAAAGAAAACGTGTCCACGTGAACGACCTGGCCCCGGCAG CTGACAACACTGGGCATTTCAGCTGAGCCCCCTACAACCTGAGC ACACTGGAAGATCTGAACACCCGGTGGAAGCTGCTGCAGGTG GCCGTGGAAGATAGAGTGCGGCAGCTGCACGAG
104	Hinge 4	GCCCACAGAGATTTTGCCCTGCCTCCCAGCACTTCCTGAGC ACCAGTGTGCAGGGCCCTTGGGAGAGAGCCATCTCCCCTAAC AAGGTGCCCTACTACATCAACCACGAGACACAGACCACCTGT TGGGACCACCCCAAGATGACCGAGCTGTACCAGAGCCTGGCT GACCTGAACAACGTGCGGTTTCAGGCCTACCGGACCGCCATGA AGCTG
105	CR/C-term	CGGAGACTGCAGAAAGCTCTGTGCCTGGATCTGCTGTCCCTG TCCGCCGCCTGTGATGCCCTGGACCAGCACAATCTGAAGCAG AACGACCAGCCCATGGATATCCTGCAGATCATCAACTGCCTG ACCACCATCTACGACCGGCTGGAACAGGAACACAACAATCTC GTGAACGTGCCCTGTGCGTGGACATGTGCCTGAATTGGCTG CTGAATGTGTACGACACCGGCCGGACAGGCCGGATCAGAGTG CTGTCCTTCAAGACCGGCATCATCAGCCTGTGCAAAGCCCAC CTGGAAGATAAGTACCGGTATCTGTTCAAACAGGTGGCCTCT AGCACCGGCTTTTTCGACACAGAGAAGGCTGGGCCTGCTGCTG CACGACAGCATCCAGATCCCTAGACAGCTGGGCGAAGTGGCC AGCTTTGGCGGCAGCAACATCGAGCCTAGCGTGCGGAGCTGC TTCCAGTTCGCCAACAACAAGCCCCGAGATCGAGGCCGCCCTG TTCCTGGACTGGATGAGACTGGAACCCAGAGCATGGTGTGG CTGCCCCTGCTGCATAGAGTGGCCGCTGCCGAGACAGCCAAG CACCAGGCCAAGTGCAACATCTGCAAAGAGTGCCCCATCATC GGCTTCCGGTACAGAAGCCTGAAGCACTTCAACTACGATATC TGCCAGAGCTGTTTCTTCAGCGGCAGGGTGGCCAAGGGCCAC AAAATGCACTACCCCATGGTGGAATACTGCACCCCCACCACA

		AGCGGCGAGGATGTGCGGGATTTCGCCAAGGTGCTGAAAAA CAAGTTCCGGACCAAGCGGTACTTCGCCAAACACCCCCGGAT GGGCTACCTGCCTGTGCAGACAGTGCTGGAAGGCGACAACAT GGAAACCGACACCATGTGATGATGATTTGGGCAGAGCGATGG AGTCCTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAAT AAATGTTTTACAACCTCCTGATTCCCGCATGCGGCCAGCTTATC GATACCGTCGAAATAAAAGATCCTTATTTTCATTGGATCTGTG TGTTGGTTTTTTTGTGTG
--	--	--

In some embodiments, a nucleic acid sequence encoding the H1 domain in the miniaturized dystrophin polypeptide is a sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 94. In some embodiments, a nucleic acid sequence encoding the R1 domain is a sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 95. In some embodiments, a nucleic acid sequence encoding the R3 domain is a sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 97. In some embodiments, a nucleic acid sequence encoding the H2 domain is a sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 98. In some embodiments, a nucleic acid sequence encoding the R16 domain is a sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 101. In some embodiments, a nucleic acid sequence encoding the R17 domain is a sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 102. In some embodiments, a nucleic acid sequence encoding the R24 domain is a sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 103. In some

embodiments, a nucleic acid sequence encoding the H4 domain is a sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 104. In some embodiments, a nucleic acid sequence

5 encoding the ABD1 domain in the miniaturized dystrophin polypeptide is a sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 93. In some embodiments, a nucleic acid sequence encoding the CR/C-term. polypeptide is a sequence at least about 60%, at

10 least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 105.

Table 6B: Nucleotide sequence (and domain structure) encoding miniaturized dystrophin polypeptide BXA-027743.

SEQ ID NO:	Description	Nucleotide Sequence
106	5' UTR	CCTGCGGCCGCTTCGGCACCATTCTCACGACACCCAAATA TGGCGACGGGTGAGGAATGGTGGGGAGTTATTTTTAGAGCGG TGAGGAAGGTGGGCAGGCAGCAGGTGTTGGCGCTCTAAAAA TAACTCCCGGGAGTTATTTTTAGAGCGGAGGAATGGTGGACA CCCAAATATGGCGACGGTTCCTACCCGTCGCCATATTTGGGT GTCCGCCCTCGGCCGGGGCCGCATTCTGGGGGCCGGGCGGT GCTCCCGCCCGCCTCGATAAAAGGCTCCGGGGCCGGCGGCGG CCCACGAGCTACCCGGAGGAGCGGGAGGCTCTAGACCACC
107	ABD1	ATGCTGTGGTGGGAGGAAGTGAAGATTGCTACGAGCGCGA GGACGTGCAGAAGAAAACCTTCACCAAATGGGTGAACGCCC AGTTCAGCAAGTTCGGCAAGCAGCACATCGAGAACCTGTTCA GCGACCTGCAGGACGGCAGACGGCTGCTGGACCTGCTGGAA GGCCTGACCGGCCAGAAGCTGCCCAAAGAGAAGGGCAGCAC CAGAGTGCACGCCCTGAACAACGTGAACAAGGCCCTGCGGGT GCTGCAGAACAACAACGTGGACCTGGTGAACATCGGCAGCA CCGACATCGTGGACGGCAACCACAAGCTGACCCTGGGCCTGA TCTGGAACATCATCCTGCACTGGCAGGTCAAAAACGTGATGA AGAACATCATGGCCGGCCTGCAGCAGACCAACAGCGAGAAG ATCCTGCTGAGCTGGGTGCGCCAGAGCACCCGGAACCTACCC CAGGTCAACGTGATCAACTTCACCACCTCTTGAGCGACGGC CTGGCCCTGAACGCCCTGATCCACAGCCACCGGCCCGACCTG TTCGACTGGAACAGCGTGGTCTGCCAGCAGAGCGCCACCCAG

		CGGCTGGAACACGCCTTCAATATCGCCAGATACCAGCTGGGC ATCGAGAAGCTGCTGGATCCCCGAGGACGTGGACACCACCTAC CCCGACAAGAAATCCATCCTGATGTATATCACCAGCCTGTTC CAGGTGCTGCCCCAGCAGGTGTCCATCGAGGCCATCCAGGAA GTGGAA
108	Hinge 1	ATGCTGCCCAGACCCCCCAAAGTGACCAAAGAGGAACACTTC CAGCTGCACCACCAGATGCACTACAGCCAGCAGATCACCCTG TCCCTGGCTCAGGGCTACGAGCGGACCAGCAGCCCCAAGCCC CGGTTCAAGAGCTACGCCTACACCCAGGCCGCCTACGTGACC ACCAGCGACCCCACCAGAAGCCCATTCCCCAGCCAGCATCTG GAAGCCCCCGAGGACAAGAGCTTCGGCAGCAGCCTGATGGA AAGC
108	Spectrin-1	GAAGTGAACCTGGACAGATACCAGACCGCCCTGGAAGAGGT GCTGTCCTGGCTGCTGAGCGCCGAGGATACACTGCAGGCCCA GGGCGAGATCAGCAACGACGTGGAAGTGGTGAAGACCAGT TCCACACCCACGAGGGCTACATGATGGACCTGACCGCCCACC AGGGCAGAGTGGGCAACATCCTGCAGCTGGGCAGCAAGCTG ATCGGCACCGGCAAGCTGAGCGAGGACGAAGAGACAGAGGT GCAGGAACAGATGAACCTGCTGAACAGCAGATGGGAGTGCC TGCGGGTGGCCAGCATGGAAAAGCAGAGCAACCTGCAC
110	L1	ATCCACACCGTGCGGGAAGAGACAATGATGGTGATGACCGA GGACATGCCCCTGGAAATC
111	Spectrin-16	AGCTACGTGCCCAGCACCTACCTGACCGAGATCACCACGTG TCCCAGGCTCTGCTGGAAGTGGAACAGCTGCTGAACGCCCC GACCTGTGCGCCAAGGACTTCGAGGATCTGTTCAAGCAGGAA GAGAGCCTGAAGAATATCAAGGACTCCCTGCAGCAGTCCAGC GGCCGGATCGACATCATCCACAGCAAGAAAACAGCCGCCCTG CAGTCCGCCACCCCCGTGGAAAGAGTGAAGCTGCAGGAAGC CCTGAGCCAGCTGGACTTCCAGTGGGAGAAAGTGAACAAGAT GTACAAGGACCGGCAGGGCAGATTTCGACCGCAGC
112	Spectrin-17	GTGGAAAAGTGGCGGCGGTTCCACTACGACATCAAGATCTTC AACCAGTGGCTGACCGAGGCCGAGCAGTTCCTGAGAAAGAC CCAGATCCCCGAGAACTGGGAGCACGCCAAGTACAAGTGGT ATCTGAAAGAACTGCAGGATGGCATCGGCCAGCGGCAGACC GTGGTGCGCACACTGAATGCCACCGGCGAGGAAATCATCCAG CAGAGCAGCAAGACCGACGCCAGCATCCTGCAGGAAAAGCT GGGCTCCCTGAACCTGCGGTGGCAGGAAGTGTGCAAGCAGCT GAGCGACCGGAAGAAGAGGCTGGAAGAACAG
113	Hinge 3	GCCCCTGGCCTGACCACCATCGGCGCCAGCCCTACCCAGACC GTGACCCTGGTGACACAGCCCGTGGTGACAAAAGAGACAGC CATCTCCAAGCTGGAAATGCCCAGCTCCCTGATGCTGGAA

114	Spectrin-23	TCTGACCAGTGGAAGCGTCTGCACCTTTCTCTGCAGGAACTTC TGGTGTGGCTACAGCTGAAAGATGATGAATTAAGCCGGCAGG CACCTATTGGAGGCGACTTTCCAGCAGTTCAGAAGCAGAACG ATGTACATAGGGCCTTCAAGAGGGAATTGAAAACATAAGAA CCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGA CAGAGCAGCCTTTGGAAGGACTAGAGAAACTCTACCAGGAG CCCAGAGAGCTGCCTCCTGAGGAGAGAGAGCCCAGAATGTCACT CGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGG GAAAAATTGAACCTGCACTCCGCTGACTGGCAGAGAAAAATA GATGAGACC
115	Spectrin-24	CTGGAAAGGCTGCAGGAACTGCAGGAGGCCACCGACGAGCT GGACCTGAAGCTGCGGCAGGCCGAAGTGATCAAGGGCAGCT GGCAGCCCGTGGGCGACCTGCTGATCGACAGCCTGCAGGACC ACCTGGAAAAAGTGAAGGCCCTGAGAGGCGAGATCGCCCCC CTGAAAGAAAACGTGTCCACGTGAACGACCTGGCCCCGGCAG CTGACCACCCTGGGCATCCAGCTGAGCCCCTACAACCTGAGC ACCCTGGAAGATCTGAACACCCGGTGGAAGCTGCTGCAGGTG GCCGTGGAAGATAGAGTGCGGCAGCTGCACGAG
116	Hinge 4	GCCCACAGAGACTTTGGCCCTGCCAGCCAGCACTTCCTGAGC ACCTCTGTGCAGGGACCCTGGGAGAGAGCCATCAGCCCCAAC AAGGTGCCCTACTACATCAACCACGAGACACAGACCACCTGT TGGGACCACCCCAAGATGACCGAGCTGTACCAGAGCCTGGCC GACCTGAACAATGTGCGGTTTCAGCGCCTACCGGACCGCCATG AAGCTG
117	CR/C-term	AGGCGGCTGCAGAAAGCTCTGTGCCTGGATCTGCTGAGCCTG AGCGCCGCCTGCGACGCCCTGGACCAGCACAACTGAAGCAG AACGACCAGCCCATGGATATCCTGCAGATCATCAACTGCCTG ACCACAATCTACGACAGGCTGGAACAGGAACACAACAATCT GGTCAACGTGCCCCTGTGCGTGGACATGTGCCTGAATTGGCT GCTGAATGTGTACGACACCGGCCGGACCGGCAGAATCCGGGT GCTGAGCTTCAAGACCGGCATCATCAGCCTGTGCAAGGCCCA CCTGGAAGATAAGTACCGCTACCTGTTCAAACAGGTGGCCAG CTCCACCGGCTTTTTCGACACGCGGAGACTGGGCCTGCTGCT GCACGACAGCATCCAGATCCCCAGACAGCTGGGCGAGGTGG CCTCCTTCGGCGGCAGCAACATTGAGCCCAGCGTGCGGAGCT GCTTCCAGTTCGCCAACAACAAGCCCGAGATCGAGGCCGCC TGTTCCCTGGACTGGATGAGACTGGAACCCAGAGCATGGTGT GGCTGCCCCTGCTGCATCGGGTGGCCGCTGCCGAGACAGCCA AGCACCAGGCCAAGTGCAACATCTGCAAAGAGTGCCCCATCA TCGGCTTCCGGTACAGAAGCCTGAAGCACTTCAACTACGATA TCTGCCAGAGCTGCTTCTTCAGCGGCAGAGTGGCCAAGGGCC ACAAAATGCACTACCCCATGGTGGAATACTGCACCCCCACCA CCAGCGGCGAGGATGTGCGGGACTTCGCCAAGGTGCTGAAA AACAAGTTCCGGACCAAGCGGTACTTTGCCAAGCACCCCCGG ATGGGCTACCTGCCCCTGCAGACAGTGCTGGAAGGCGACAAC

		ATGGAAACCCCTGCCAGATCCTGATCAGCCTGGAAAGCGAG GAACGGGGCGAGCTGGAACGGATCCTGGCCGATCTGGAAGA GGAAAACCGGAACCTGCAGGCCGAGTACGACCGGCTGAAGC AGCAGCACGAGCACAAGGGCCTGAGCCCCCTGCCTAGCCCCC CTGACACCATGTGA
--	--	---

The various miniaturized dystrophin polypeptides of the present disclosure are shown in Table 7.

Table 7. Amino Acid Sequences of miniaturized dystrophin constructs.

SEQ NO and Description	ID	Sequence
SEQ NO:118 BXA- 196477	-	MLWWEVEEDCYEREDVQKKTFTKWVNAQFSKFGKQHIEENLFSDLQDG RRLDLLEGLTGQKLPKEKGSTRVHALNNVNKALRVLQNNNVDLVNIG STDIVDGNHKLTLGLIWNILHWQVKNVMKNIMAGLQQTNSEKILLSWV RQSTRNYPQVNVINFTTSWSDGLALNALIHSRDLFDWNSVVCQQSAT QRLEHAFNIARYQLGIEKLLDPEDVDTTYPDKKSILMYITSLFQVLPQQV SIEAIQEVEMLPKPPKVTKEEHFQLHHQMHYSQQITVSLAQGYERTSSPK PRFKSYAYTQAAYVTTSDPTRSPFPSQHLEAPEDKSFSSLMSEVNLDL YQTALVEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTA HQGRVGNILQLGSKLIGTGKLSEDEETEVEQMNLLNSRWECLRVASM EKQSNLHRVLLKWQRLTEEQCLFSAWLSEKEDAVNKIHTTGFKDQNE LSSLQKLAVLKADLEKKKQSMGKLYSLKQDLLSTLKNKSVTQKTEAWL DNFARCWDNLVQKLEKSTAQISQAVTTTQPSLTQTTVMETVTTVTTR EQILVKHAQEELPPPPQKKRQITVDSEAIHTVREETMMVMTE DMPLEISYVPSTYLTEITHVSQALLEVEQLLNAPDLCAKDFEDLFKQ EESLKNIKDSLQSSGRIDIIHSSKTAALQSATPVERVKLQEALSQ LDFQWEKVNKMYKDRQGRFDRSVEKWRRFHYDIKIFNQWLTEAEQ FLRKTQIPENWEHAKYKWYLKELQDGIGQRQTVVRTL NATGEEIIQQSSKTDASILQEKLGSNLRWQEVCKQLSDRKKR LEEQLERLQELQEATDELDLKLRLQAEVIKGSWQPVGDLLID SLQDHLEKVKALRGEIAPLKENVSHVNDLARQLTTLGIQLSPY NLSTLEDLNTRWKLLQVAVEDRVRQLHEAHRDFGPASQHF LSTSVQGPWERAI SPNKVPYYINHETQTTCWDHPKMT ELYQSLADLNNVRFSAYRTAMKLRRLLQKALCLDLLSL SAACDALDQHNKQNDQPMQDILQIINCLTTIYDRLEQEH NNLVNVPLCVDMLNWLNVYDTGRTGRIRVLSFKTGIISL CKAHLEDKYRYLQKQVASTGFCDQRRLLGLLLHDSIQIP RQLGEVASFGGSNIEPSVRSCFQFANNKPEIEAALFLD WMLRLEPQSMVWLPVLRVAAAETAKHQAKCNICKECPI IGFRYRSLKHFNYDICQSCFFSGRVAKGHKMHYPMVEY CTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGYLPVQ TVLEGDNMETDTM
SEQ NO:119	-	MLWWEVEEDCYEREDVQKKTFTKWVNAQFSKFGKQHIEENLFSDLQDG RRLDLLEGLTGQKLPKEKGSTRVHALNNVNKALRVLQNNNVDLVNIG STDIVDGNHKLTLGLIWNILHWQVKNVMKNIMAGLQQTNSEKILLSWV RQSTRNYPQVNVINFTTSWSDGLALNALIHSRDLFDWNSVVCQQSAT

BXA- 196473	QRLEHAFNIARYQLGIEKLLDPEDVDTTYPDKKSILMYITSLFQVLPQQV SIEAIQEVEMLPKPPKVTKEEHFQLHHQMHYSQQITVSLAQGYERTSSPK PRFKSYAYTQAAYVTTSDPTRSPFPSQHLEAPEDKSFSSLMSEVNLD YQTALVEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTA HQGRVGNILQLGSKLIGTGKLSSEDEETEVEQEQMNLNLSRWECLRVASM EKQSNLHLLKWQRLTEEQCLFSAWLSEKEDAVNKIHTTGFKDQNEMLS SLQKLAVLKADLEKKKQSMGKLYSLKQDLLSTLKNKSVTQKTEAWLD NFARCWDNLVQKLEKSTAQISQAVTTTQPSLTQTTVMETVTTVTTREQI LVKHAQEELPPPPQKKRQITVDIHTVREETMMVMTEDMPLISYVPSTY LTEITHVSQALLEVEQLLNAPDLCAKDFEDLFKQEEESLKNIKDSLQSSG RIDIIHKKTAALQSATPVERVKLQEALSQDLDFQWEKVNKMYKDRQGRF DRSVEKWRRFHYDIKIFNQWLTEAEQFLRKTQIPENWEHAKYKWYLKE LQDGIGQRQTVVRTLNATGEEIIQQSSKTDASILQEKLGSNLNRWQEVCK QLSDRKKRLEEQLERLQELQEATDELDLKLRAEVIKGSWQPVGDLLID SLQDHLEKVKALRGEIAPLKENVSHVNDLARQLTTLGIQLSPYNLSTLED LNTRWKLLQVAVEDRVRQLHEAHRDFGPASQHFLSTSVQGPWERAI SPNKVPYYINHETQTTTCWDHPKMTLEYQSLADLNNVRFSAYRTAMKLRL QKALCLDLLSLSAACDALDQHNKQNDQPMILQIINCLTTIYDRLEQEH NNLVNVPLCVDMLNWLNNVYDTGRTGRIRVLSFKTGIISLCKAHLEDK YRYLFKQVASSTGFCDQRRLLGLLLHDSIQIPRQLGEVASFGGSNI EPSVRS CFQFANNKPEIEAALFLDWMRLPQSMVWLPVLRVAAAETAKHQAK CNICKECPIIGFRYRSLKHFNYDICQSCFFSGRVAKGHKMHYPMVEYCTP TTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGYLPVQTVLEGDNMETDT M
SEQ ID NO:120 - BXA- 196474	MLWWEVEEDCYEREDVQKKTFKWWNAQFSKFGKQHINLFSDLQDG RRLDLLEGLTGQKLPKEKGSTRVHALNNVNKALRVLQNNNVDLVNIG STDIVDGNHKLTLGLIWNILHWQVKNVMKNIMAGLQQTNSEKILLSWV RQSTRNYPQVNVINFTTSWSDGLALNALIHSRPLDFDWNSSVCQQSAT QRLEHAFNIARYQLGIEKLLDPEDVDTTYPDKKSILMYITSLFQVLPQQV SIEAIQEVEMLPKPPKVTKEEHFQLHHQMHYSQQITVSLAQGYERTSSPK PRFKSYAYTQAAYVTTSDPTRSPFPSQHLEAPEDKSFSSLMSEVNLD YQTALVEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTA HQGRVGNILQLGSKLIGTGKLSSEDEETEVEQEQMNLNLSRWECLRVASM EKQSNLHLLKWQRLTEEQCLFSAWLSEKEDAVNKIHTTGFKDQNEMLS SLQKLAVLKADLEKKKQSMGKLYSLKQDLLSTLKNKSVTQKTEAWLD NFARCWDNLVQKLEKSTAQISQAVTTTQPSLTQTTVMETVTTVTTREQI LVKHAQEELPPPPQKKRQITVDSEAQIHTVREETMMVMTEDMPLISY VPSTYLTEITHVSQALLEVEQLLNAPDLCAKDFEDLFKQEEESLKNIKDSL QSSGRIDIIHKKTAALQSATPVERVKLQEALSQDLDFQWEKVNKMYKD RQGRFDRSVEKWRRFHYDIKIFNQWLTEAEQFLRKTQIPENWEHAKYK WYLKELQDGIGQRQTVVRTLNATGEEIIQQSSKTDASILQEKLGSNLNR WQEVCKQLSDRKKRLEEQKNILQELQEATDELDLKLRAEVIKGSWQPV GDLLIDSLQDHLEKVKALRGEIAPLKENVSHVNDLARQLTTLGIQLSPY NLSTLEDLNTRWKLLQVAVEDRVRQLHEAHRDFGPASQHFLSTSVQGP WERAI SPNKVPYYINHETQTTTCWDHPKMTLEYQSLADLNNVRFSAYRT AMKLRLQKALCLDLLSLSAACDALDQHNKQNDQPMILQIINCLTTI YDRLEQEHNNLVNVPLCVDMLNWLNNVYDTGRTGRIRVLSFKTGIISL CKAHLEDKYRYLFKQVASSTGFCDQRRLLGLLLHDSIQIPRQLGEVASFG

		GSNIEPSVRSCFQFANNKPEIEAALFLDWMRLEPQSMVWLPVLHRVAAA ETAKHQAKCNICKECPIIGFRYRSLKHFNYDICQSCFFSGRVAKGHKMH YPMVEYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGYLPVQTVL EGDNMETTPCDDDLGRAMESLVSVMTDEEGAE
SEQ NO:121 BXA- 196475	ID -	MLWWEVEEDCYEREDVQKKTFTKWVNAQFSKFGKQHIEENLFSDLQDG RRLDLLEGLTGQKLPKEKGSTRVHALNNVNKALRVLQNNNVDLVNIG STDIVDGNHKLTLGLIWNILHWQVKNVMKNIMAGLQQTNSEKILLSWV RQSTRNYPQVNVINFTTSWSDGLALNALIHSHRPDLFDWNSVVCQQSAT QRLEHAFNIARYQLGIEKLLDPEDVDTTYPDKKSILMYITSLFQVLPQQV SIEAIQEVEMLPKPPKVTKEEHFQLHHQMHYSQQITVSLAQGYERTSSPK PRFKSYAYTQAAYVTTSDPTRSPFPSQHLEAPEDKSFSGSSLMESEVNDR YQTALEEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTA HQGRVGNILQLGSKLIGTGKLSEDEETEVEQEQMNLLNSRWECLRVASM EKQSNLHLLKWQRLTEEQCLFSAWLSEKEDAVNKIHTTGFKDQNEMLS SLQKLAVLKADLEKKKQSMGKLYSLKQDLLSTLKNKSVTQKTEAWLD NFARCWDNLVQKLEKSTAQISQAVTTTQPSLTQTTVMETVTTVTTREQI LVKHAQEELPPPPPQKKRQITVDSEAQIHTVREETMMVMTEDMPLISY VPSTYLTEITHVSQALLEVEQLLNAPDLCAKDFEDLFKQEEESLKNIKDSL QQSSGRIDIIHKKTAALQSATPVERVKLQEALSQDLDFQWEKVNKMYKD RQGRFDRSVEKWRRFHYDIKIFNQWLTEAEQFLRKTQIPENWEHAKYK WYLKELQDGIGQRQTVVRTLNATGEEIIQQSSKTDASILQEKLGSNLNR WQEVCKQLSDRKKRLEEQKLERLQELQEATDELDLKLRAEVIKGSWQ PVGDLLIDSLQDHLEKVKALRGEIAPLKENVSHVNDLARQLTTLGIQLSP YNLSTLEDLNTRWKLQVAVEDRVRQLHEAHRDFGPASQHFLSTSVQG PWERAIKPNKVPYYINHETQTTCDWHPKMTLEYQSLADLNNVRFSAYRT AMKLRRQLKALCLDLSLSAACDALDQHNKQNDQPMILQIINCLTTI YDRLEQEHNNLVNVPLCVDMLNWLNNVYDTGRTGRIRVLSFKTGIIISL CKAHLEDKYRYLFKQVASSTGFCDQRRLLGLLLHDSIQIPRQLGEVASFG GSNIEPSVRSCFQFANNKPEIEAALFLDWMRLEPQSMVWLPVLHRVAAA ETAKHQAKCNICKECPIIGFRYRSLKHFNYDICQSCFFSGRVAKGHKMH YPMVEYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGYLPVQTVL EGDNMETDTM
SEQ NO:122 BXA- 196476	ID -	MLWWEVEEDCYEREDVQKKTFTKWVNAQFSKFGKQHIEENLFSDLQDG RRLDLLEGLTGQKLPKEKGSTRVHALNNVNKALRVLQNNNVDLVNIG STDIVDGNHKLTLGLIWNILHWQVKNVMKNIMAGLQQTNSEKILLSWV RQSTRNYPQVNVINFTTSWSDGLALNALIHSHRPDLFDWNSVVCQQSAT QRLEHAFNIARYQLGIEKLLDPEDVDTTYPDKKSILMYITSLFQVLPQQV SIEAIQEVEMLPKPPKVTKEEHFQLHHQMHYSQQITVSLAQGYERTSSPK PRFKSYAYTQAAYVTTSDPTRSPFPSQHLEAPEDKSFSGSSLMESEVNDR YQTALEEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTA HQGRVGNILQLGSKLIGTGKLSEDEETEVEQEQMNLLNSRWECLRVASM EKQSNLHLLKWQRLTEEQCLFSAWLSEKEDAVNKIHTTGFKDQNEMLS SLQKLAVLKADLEKKKQSMGKLYSLKQDLLSTLKNKSVTQKTEAWLD NFARCWDNLVQKLEKSTAQISQAVTTTQPSLTQTTVMETVTTVTTREQI LVKHAQEELPPPPPQKKRQITVDSEAQIHTVREETMMVMTEDMPLISY VPSTYLTEITHVSQALLEVEQLLNAPDLCAKDFEDLFKQEEESLKNIKDSL QQSSGRIDIIHKKTAALQSATPVERVKLQEALSQDLDFQWEKVNKMYKD RQGRFDRSVEKWRRFHYDIKIFNQWLTEAEQFLRKTQIPENWEHAKYK

		WYLKELQDGIGQRQTVVRTLNATGEEIIQQSSKTDASILQEKLGSNLNR WQEVCKQLSDRKKRLEEQKNILERLQELQEATDELDLKLKRAEVIKGS WQPVGDLLIDSLQDHLEKVKALRGEIAPLKENVSHVNDLARQLTTLGIQ LSPYNLSTLEDLNTRWKLLQVAVEDRVRQLHEAHRDFGPASQHFLSTSV QGPWERAISPKNVPYYINHETQTTCWDHPKMTELYQSLADLNNVRFSA YRTAMKLRRRLQKALCLDLLSLSAACDALDQHNKQNDQPMQDILQINCL TTIYDRLEQEHNNLVNVPLCVDMLNWLNNVYDTGRTGRIRVLSFKTGI ISLCKAHLEDKYRYLFKQVASSTGFCDQRRRLGLLLHDSIQIPRQLGEVAS FGGSNIEPSVRSCFQFANNKPEIEAALFLDWMRLEPQSMVWLPVLHRVA AAETAKHQAKCNICKECPIIGFRYRSLKHFNYDICQSCFFSGRVAKGHK MHYPMVEYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGYLPVQ TVLEGDNMETDTM
SEQ ID NO:123 - BXA- 196477-Q		MLWWEVEDCYEREDVQKKTFTKWVNAQFSKFGKQHIEENLFSDLQDG RRLDLLEGLTGQKLPKEKGSTRVHALNNVNKALRVLQNNNVDLVNIG STDIVDGNHKLTLGLIWNILHWQVKNVMKNIMAGLQQTNSEKILLSWV RQSTRNYPQVNVINFTTSWSDGLALNALIHSRDLFDWNSVVCQQSAT QRLEHAFNIARYQLGIEKLLDPEDVDTTYPDKKSILMYITSLFQVLPQQV SIEAIQEVEMLPKPPKVTKEEHFQLHHQMHYSQQITVSLAQGYERTSSPK PRFKSYAYTQAAYVTTSDPTRSPFPSQHLEAPEDKSFSSLMSEVNLDL YQTALEEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTA HQGRVGNILQLGSKLIGTGKLSEDEETEVEQEMNLLNSRWECLRVASM EKQSNLHRVLLKWQRLTEEQCLFSAWLSEKEDAVNKIHTTGFKDQNE LSSLQKLAVLKADLEKKKQSMGKLYSLKQDLLSTLKNKSVTQKTEAWL DNFARCWDNLVQKLEKSTAQISQAVTTTQPSLTQTTVMETVTTVTREQ ILVKHAQEELPPPPQKKRQITVDSEAQIHTVREETMMVMTEMPLEISY VPSTYLTEITHVSQALLEVEQLLNAPDLCAKDFEDLFKQEEESLKNIKDSL QQSSGRIDIIHKKTAALQSATPVERVKLQEALSQLDQWEKVNKMYKD RQGRFDRSVEKWRFRHYDIKIFNQWLTEAEQFLRKTQIPENWEHAKYK WYLKELQDGIGQRQTVVRTLNATGEEIIQQSSKTDASILQEKLGSNLNR WQEVCKQLSDRKKRLEEQLERLQELQEATDELDLKLKRAEVIKGSWQP VGDLLIDSLQDHLEKVKALRGEIAPLKENVSHVNDLARQLTTLGIQVSPY NLSTLEDLNTRWKLLQVAVEDRVRQLHEAHRDFGPASQHFLSTSVQGP WERAISPKNVPYYINHETQTTCWDHPKMTELYQSLADLNNVRFSA YRTAMKLRRRLQKALCLDLLSLSAACDALDQHNKQNDQPMQDILQINCLTTI YDRLEQEHNNLVNVPLCVDMLNWLNNVYDTGRTGRIRVLSFKTGIISL CKAHLEDKYRYLFKQVASSTGFCDQRRRLGLLLHDSIQIPRQLGEVASFG GSNIEPSVRSCFQFANNKPEIEAALFLDWMRLEPQSMVWLPVLHRVAAA ETAKHQAKCNICKECPIIGFRYRSLKHFNYDICQSCFFSGRVAKGHKMH YPMVEYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGYLPVQTVL EGDNMETDTM
SEQ ID NO:124 - BXA- 196478		MLWWEVEDCYEREDVQKKTFTKWVNAQFSKFGKQHIEENLFSDLQDG RRLDLLEGLTGQKLPKEKGSTRVHALNNVNKALRVLQNNNVDLVNIG STDIVDGNHKLTLGLIWNILHWQVKNVMKNIMAGLQQTNSEKILLSWV RQSTRNYPQVNVINFTTSWSDGLALNALIHSRDLFDWNSVVCQQSAT QRLEHAFNIARYQLGIEKLLDPEDVDTTYPDKKSILMYITSLFQVLPQQV SIEAIQEVEMLPKPPKVTKEEHFQLHHQMHYSQQITVSLAQGYERTSSPK PRFKSYAYTQAAYVTTSDPTRSPFPSQHLEAPEDKSFSSLMSEVNLDL YQTALEEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTA

		HQGRVGNILQLGSKLIGTGKLSEDEETEVEQEQMNLNLSRWECRLVASM EKQSNLHRVLLQDILLKWQRLTEEQCLFSAWLSEKEDAVNKIHTTGFKD QNEMLSSLQKLAVLKADLEKKKQSMGKLYSLKQDLLSTLKNKSVTQKT EAWLDNFARCWDNLVQKLEKSTAQISQAVTTTQPSLTQTTVMETVTTV TTREQILVKHAQEELPPPPQKKRQITVDSEAQIHTVREETMMVMTEDM PLEISYVPSTYLTEITHVSQALLEVEQLLNAPDLCAKDFEDLFKQEEESLKN IKDSLQSSGRIDIIHSSKTAALQSATPVERVKLQEALSQLDQFQWEKVNK MYKDRQGRFDRSVEKWRRFHYDIKIFNQWLTEAEQFLRKTQIPENWEH AKYKWYLKELQDGIGQRQTVVRTLNATGEEIIQQSSKTDASILQEKLGS NLRWQEVCKQLSDRKKRLEEQLERLQELQEATDELDLKLKRAEVIKGS WQPVGDLIDSLQDHLEKVKALRGEIAPLKENVSHVNDLARQLTTLG LSPYNLSTLEDLNTRWKLLQVAVEDRVRQLHEAHRDFGPASQHFLSTSV QGPWERAI SPNKVPYYINHETQTTTCWDHPKMTELYQSLADLNNVRFSA YRTAMKLRRQLKALCLDLSLSAACDALDQHNKQNDQPMQDILQIINCL TTIYDRLEQEHNNLVNVPLCVDMLNWLNNVYDTGRTGRIRVLSFKTGI ISLCKAHLEDKYRYLKFQVASSTGFCDQRRLLGLLHDSIQIPRQLGEVAS FGGSNIEPSVRSCFQFANNKPEIEAALFLDWMRLEPQSMVWLPVLHRVA AAETAKHQAKCNICKECPIIGFRYRSLKHFNYDICQSCFFSGRVAKGHK MHYPMVEYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGYLPVQ TVLEGNMETDTM
SEQ ID NO:125 BXA- 196479	-	MLWWEVEDCYEREDVQKKTFTKWVNAQFSKFGKQHIEENLFSDLQDG RRLDLLEGLTGQKLPEKEGSTRVHALNNVNKALRVLQNNNVDLVNIG STDIVDGNHKLTLGLIWNILHWQVKNVMKNIMAGLQQTNSEKILLSWV RQSTRNYPQVNVINFTTSDSDGLALNALIHSRPLDFDWNSSVCCQSAT QRLEHAFNIARYQLGIEKLLDPEDVDTTYPDKKSILMYITSLFQVLPQQV SIEAIQEVEMLP RP PKVTKEEHFQLHHQM HYSQQITVSLAQGYERTSSPK PRFKSYAYTQAAYVTTSDPTRSPFPSQHLEAPEDKSFSSLMSEVNLD YQTAL EEVLSWLLSAEDTLQAQGEISNDVEVVKDQFH THEGYMMDLTA HQGRVGNILQLGSKLIGTGKLSEDEETEVEQEQMNLNLSRWECRLVASM EKQSNLHLLKWQRLTEEQCLFSAWLSEKEDAVNKIHTTGFKDQNEMLS SLQKLAVLKADLEKKKQSMGKLYSLKQDLLSTLKNKSVTQKTEAWLD NFARCWDNLVQKLEKSTAQISQAVTTTQPSLTQTTVMETVTTVTTREQI LVKHAQEELPPPPQKKRQITVDSEAQIHTVREETMMVMTEDMPLEISY VPSTYLTEITHVSQALLEVEQLLNAPDLCAKDFEDLFKQEEESLKNIKDSL QQSSGRIDIIHSSKTAALQSATPVERVKLQEALSQLDQFQWEKVNKMYKD RQGRFDRSVEKWRRFHYDIKIFNQWLTEAEQFLRKTQIPENWEHAKYK WYLKELQDGIGQRQTVVRTLNATGEEIIQQSSKTDASILQEKLGSNLR WQEVCKQLSDRKKRLEEQLERLQELQEATDELDLKLKRAEVIKGSWQP VGDLIDSLQDHLEKVKALRGEIAPLKENVSHVNDLARQLTTLGILSPY NLSTLEDLNTRWKLLQVAVEDRVRQLHEAHRDFGPASQHFLSTSVQGP WERAI SPNKVPYYINHETQTTTCWDHPKMTELYQSLADLNNVRFSA YRTAMKLRRQLKALCLDLSLSAACDALDQHNKQNDQPMQDILQIINCLTTI YDRLEQEHNNLVNVPLCVDMLNWLNNVYDTGRTGRIRVLSFKTGIISL CKAHLEDKYRYLKFQVASSTGFCDQRRLLGLLHDSIQIPRQLGEVASFG GSNIEPSVRSCFQFANNKPEIEAALFLDWMRLEPQSMVWLPVLHRVAAA ETAKHQAKCNICKECPIIGFRYRSLKHFNYDICQSCFFSGRVAKGHKMH YPMVEYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGYLPVQTVL EGDNMETDTM

SEQ ID NO:126 BXA- 196480	-	MLWWEEVEDCYEREDVQKKTFTKWVNAQFSKFGKQHIEENLFSDLQDG RRLDLLEGLTGQKLPEKEGSTRVHALNNVNKALRVLQNNNVDLVNIG STDIVDGNHKLTLGLIWNILHWQVKNVMKNIMAGLQQTNSEKILLSWV RQSTRNYPQVNVINFTTSWSDGLALNALIHSHRPDLFDWNSVVCQQSAT QRLEHAFNIARYQLGIEKLLDPEDVDTTYPDKKSILMYITSLFQVLPQQV SIEAIQEVEMLPKPPKVTKEEHFQLHHQMHSYQQITVSLAQGYERTSSPK PRFKSYAYTQAAYVTTSDPTRSPFSPQHLEAPEDKSFSSLMSEVNLDL YQTALEEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTA HQGRVGNILQLGSKLIGTGKLSEDEETEVEQEMNLLNSRWECLRVASM EKQSNLHIHTVREETMMVMTEDEMPLEISYVPSTYLTEITHVSQALLEVEQ LLNAPDLCAKDFEDLDFKQEESESLKNIKDSLQSSGRIDIIHSSKKTAAALQSAT PVERVKLQEALSQDFQWEKVNKMYKDRQGRFDRSVEKWRRFHYDIKI FNQWLTEAEQFLRKTQIPENWEHAKYKWYLKELQDGIGQRQTIVRTL ATGEEIIQQSSKTDASILQEKLSLNLRWQEVCKQLSDRKKRLEEQAAPGL TTIGASPTQTVTLVTQPVVTKETAISKLEMPSSLMLELERLQELQEATDE LDLKLRAEVIKGSWQPVGDLLIDSLQDHLEKVKALRGEIAPLKENVSH VNDLARQLTTGLIQLSPYNLSTLEDLNTRWKLQVAVEDRVRQLHEAH RDFGPASQHFSTSVQGPWERAIKPNKVPYYINHETQTTCDWHPKMTL YQSLADLNNVRFSAYRTAMKLRLQKALCLDLSLSAACDALDQHNK QNDQPMIDILQIINCLTTIYDRLEQEHNNLVNVPLCVDMLNWLNVYDT GRTGRIRVLSFKTGIIISLCKAHLEDKYRYLKFQVASTGFCQRRLLGLL HDSIQIPRQLGEVASFGGSNIEPSVRSCFQFANNKPEIEAALFLDWMRLEP QSMVWLPVLHRVAAAETAHQAACNICKECPIIGFRYRSLKHFNYDICQ SCFFSGRVAKGHKMHYPMVEYCTPTTSGEDVRDFAKVLKNKFRTKRYF AKHPRMGYLPVQTVLEGDNMETPAQILISLESEERGELEERILADLEEENR NLQAEYDRLKQQHEHKGSLPLSPPTDM
SEQ ID NO:127 BXA- 196480- Variant	-	MLWWEEVEDCYEREDVQKKTFTKWVNAQFSKFGKQHIEENLFSDLQDG RRLDLLEGLTGQKLPEKEGSTRVHALNNVNKALRVLQNNNVDLVNIG STDIVDGNHKLTLGLIWNILHWQVKNVMKNIMAGLQQTNSEKILLSWV RQSTRNYPQVNVINFTTSWSDGLALNALIHSHRPDLFDWNSVVCQQSAT QRLEHAFNIARYQLGIEKLLDPEDVDTTYPDKKSILMYITSLFQVLPQQV SIEAIQEVEMLPKPPKVTKEEHFQLHHQMHSYQQITVSLAQGYERTSSPK PRFKSYAYTQAAYVTTSDPTRSPFSPQHLEAPEDKSFSSLMSEVNLDL YQTALEEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTA HQGRVGNILQLGSKLIGTGKLSEDEETEVEQEMNLLNSRWECLRVASM EKQSNLHRVLIHTVREETMMVMTEDEMPLEISYVPSTYLTEITHVSQALLE VEQLLNAPDLCAKDFEDLDFKQEESESLKNIKDSLQSSGRIDIIHSSKKTAAAL QSATPVERVKLQEALSQDFQWEKVNKMYKDRQGRFDRSVEKWRRFH YDIKIFNQWLTEAEQFLRKTQIPENWEHAKYKWYLKELQDGIGQRQTV VRTLNATGEEIIQQSSKTDASILQEKLSLNLRWQEVCKQLSDRKKRLEE QKNDLAPGLTTIGASPTQTVTLVTQPVVTKETAISKLEMPSSLMLEVERL QELQEATDELDLKLRAEVIKGSWQPVGDLLIDSLQDHLEKVKALRGEI APLKENVSHVNDLARQLTTGLIQLSPYNLSTLEDLNTRWKLQVAVEDR VRQLHEAHRDFGPASQHFSTSVQGPWERAIKPNKVPYYINHETQTTCDW DHPKMTLYQSLADLNNVRFSAYRTAMKLRLQKALCLDLSLSAACD ALDQHNKQNDQPMIDILQIINCLTTIYDRLEQEHNNLVNVPLCVDMLN WLLNVYDTGRTGRIRVLSFKTGIIISLCKAHLEDKYRYLKFQVASTGFCQ RRLLGLLHDSIQIPRQLGEVASFGGSNIEPSVRSCFQFANNKPEIEAALF

		LDWMRLEPQSMVWLPVLHRVAAAETAKHQAKCNICKECPIIGFRYRSL KHFNYDICQSCFFSGRVAKGHKMHYPMVEYCTPTTSGEDVRDFAKVLK NKFRTKRYFAKHPRMGYLPVQTVLEGDNMETPAQILISLESEERGELERI LADLEENRNRLQAEYDRLKQQHEHKGLSPLPSPDPTM
SEQ ID NO:128 BXA- 196481	-	MLWWEVEDCYEREDVQKKTFTKWVNAQFSKFGKQHIEENLFSDLQDG RRLDLLEGLTGQKLPKEKGSTRVHALNNVNKALRVLQNNNVDLVNIG STDIVDGNHKLTLGLIWNILHWQVKNVMKNIMAGLQQTNSEKILLSWV RQSTRNYPQVNVINFTTSWSDGLALNALIHSHRPDLFDWNSVVCQQSAT QRLEHAFNIARYQLGIEKLLDPEDVDTTYPDKKSILMYITSLSFQVLPQQV SIEAIQEVEMLPKPPKVTKEEHFQLHHQMHYSQQITVSLAQGYERTSSPK PRFKSYAYTQAAYVTTSDPTRSPFPSQHLEAPEDKSFSGSSLMESEVNDR YQTALEEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTA HQGRVGNILQLGSKLIGTGKLSSEDEETEVEQEQMNLLNSRWECLRVASM EKQSNLHSYVPSTYLTEITHVSQALLEVEQLLNAPDLCAKDFEDLFKQEE SLKNIKDSLQQSSGRIDIIHKKTAALQSATPVERVKLQEALSQDLDFQWE KVNMKYKDRQGRFDRSVEKWRRFHYDIKIFNQWLTEAEQFLRKTQIPE NWEHAKYKWYLKELQDGIGQRQTVVRTL NATGEEIIQQSSKTDASILQE KLGSNLNRWQEVCKQLSDRKKRLEEQLGLEKLYQEPRELPEERAQNV TRLLRKQAEVNTWEKLNLSHADWQRKIDETLERLQELQEATDELDL KLRQAEVIKGSWQPVGDLLIDSLQDHLEKVKALRGEIAPLKENVSHVND LARQLTTLGIQLSPYNLSTLEDLNRWKLLQVAVEDRVRQLHEAHRDFG PASQHFLSTSVQGPWERAI SPNKVPYYINHETQTTCDWHPKMTELYQSL ADLNNVRFSAYRTAMKLRRLQKALCLDLSLSAACDALDQHNLKQND QPMIDILQIINCLTTIYDRLEQEHNLLVNVPLCVDMLCNWLLNVYDTGRT GRIRVLSFKTGIISLCKAHLEDKYRYLFKQVASSTGFCQRRLLGLLLHDSI QIPRQLGEVASFGGSNIEPSVRSCFQFANNKPEIEAALFLDWMRLEPQSM VWLPVLHRVAAAETAKHQAKCNICKECPIIGFRYRSLKHFNYDICQSCFF SGRVAKGHKMHYPMVEYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKH PRMGYLPVQTVLEGDNMETPAQILISLESEERGELERILADLEENRNRLQ AEYDRLKQQHEHKGLSPLPSPDPTM
SEQ ID NO:129 BXA- 027741	-	MLWWEVEDCYEREDVQKKTFTKWVNAQFSKFGKQHIEENLFSDLQDG RRLDLLEGLTGQKLPKEKGSTRVHALNNVNKALRVLQNNNVDLVNIG STDIVDGNHKLTLGLIWNILHWQVKNVMKNIMAGLQQTNSEKILLSWV RQSTRNYPQVNVINFTTSWSDGLALNALIHSHRPDLFDWNSVVCQQSAT QRLEHAFNIARYQLGIEKLLDPEDVDTTYPDKKSILMYITSLSFQVLPQQV SIEAIQEVEMLPKPPKVTKEEHFQLHHQMHYSQQITVSLAQGYERTSSPK PRFKSYAYTQAAYVTTSDPTRSPFPSQHLEAPEDKSFSGSSLMESEVNDR YQTALEEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTA HQGRVGNILQLGSKLIGTGKLSSEDEETEVEQEQMNLLNSRWECLRVASM EKQSNLHLLKWQRLTEEQCLFSAWLSEKEDAVNKIHTTGFKDQNEMLS SLQKLAVLKADLEKKKQSMGKLYSLKQDLLSTLKNKSVTQKTEAWLD NFARCWDNLVQKLEKSTAQISQAVTTTQPSLTQTTVMETVTTVTTREI LVKHAQEELPPPPPQKKRQITVDIHTVREETMMVMTEDMPLISYVPSTY LTEITHVSQALLEVEQLLNAPDLCAKDFEDLFKQEE SLKNIKDSLQQSSGRIDIIHKKTAALQSATPVERVKLQEALSQDLDFQWE KVNMKYKDRQGRFDRSVEKWRRFHYDIKIFNQWLTEAEQFLRKTQIPEN WEHAKYKWYLKELQDGIGQRQTVVRTL NATGEEIIQQSSKTDASILQE KLGSNLNRWQEVCKQLSDRKKRLEEQLERLQELQEATDELDL KLRQAEVIKGSWQPVGDLLID

		SLQDHLEKVKALRGEIAPLKENVSHVNDLARQLTTLGIQLSPYNLSTLED LNTRWKLLQVAVEDRVRQLHEAHRDFGPASQHFLSTSVQGPWERAISP NKVPYYINHETQTTTCWDHPKMTELYQSLADLNNVRFSAYRTAMKLRRRL QKALCLDLLSLSAACDALDQHNKQNDQPMQDILQIINCLTTIYDRLEQEH NNLVNVPLCVDMLCNWLLNVYDTGRTGRIRVLSFKTGIIISLCKAHLEDK YRYLFKQVASSTGFCDQRRRLGLLLHDSIQIPRQLGEVASFGGSNIEPSVRS CFQFANNKPEIEAALFLDWMRLEPQSMVWLPVLHRVAAAETAKHQAK CNICKECPIIGFRYRSLKHFNYDICQSCFFSGRVAKGHKMHYPMVEYCTP TTSGEDVRDFAKVLKNKFRTRKRYFAKHPRMGYLPVQTVLEGDNMETPA QILISLESEERGELERILADLEEENRNLQAEYDRLKQQHEHKGLSPLSPSP DTM
SEQ ID NO:130 - BXA- 027744		MLWWEVEEDCYEREDVQKKTFKWWNAQFSKFGKQHIEENLFSDLQDG RRLDLLEGLTGQKLPKEKGSTRVHALNNVNKALRVLQNNNVDLVNIG STDIVDGNHKLTLGLIWNILHWQVKNVMKNIMAGLQQTNSEKILLSWV RQSTRNYPQVNVINFTTSWSDGLALNALIHSHRPDLFDWNSVVCQQSAT QRLEHAFNIARYQLGIEKLLDPEDVDTTYPDKKSILMYITSLFQVLPQQV SIEAIQEVEMLPKPPKVTKEEHFQLHHQMHYSQQITVSLAQGYERTSSPK PRFKSYAYTQAAYVTTSDPTRSPFPSQHLEAPEDKSFSGSSLMSEVNDR YQTALEEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTA HQGRVGNILQLGSKLIGTGKLSEDEETEVEQEQMNLLNSRWECLRVASM EKQSNLHIHTVREETMMVMTEDEMPLEISYVPSTYLTEITHVSQALLEVEQ LLNAPDLCAKDFEDLFKQEEESLKNIKDSLQSSGRIDIIHSSKKTAAALQSAT PVERVKLQEALSQLDQWEKVNKMYKDRQGRFDRSVEKWRRFHYDIKI FNQWLTEAEQFLRKTQIPENWEHAKYKWLKELQDGIGQRQTIVRTLN ATGEEIIQQSSKTDASILQEKGLSLNLRWQEVCKQLSDRKKRLEEQAAPGL TTIGASPTQTVTLVTQPVVTKETAISKLEMPSSLMLEVPALADFNRAWTE LTDWLSLLDQVIKSQRVMVGDLNEMIIKQKATMQDLEQRRPQLEEL ITAAQNLKNKTSNQEARTIITDRIERIQNQWDEVQEHLQNRQQLNEML ERLQELQEATDELDELKLRQAEVIKGSWQPVGDLLIDSLQDHLEKVKALR GEIAPLKENVSHVNDLARQLTTLGIQLSPYNLSTLEDLNTRWKLLQVAV EDRVRQLHEAHRDFGPASQHFLSTSVQGPWERAISP NKVPYYINHETQT TCWDHPKMTELYQSLADLNNVRFSAYRTAMKLRRRLQKALCLDLLSLSA ACDALDQHNKQNDQPMQDILQIINCLTTIYDRLEQEHNNLVNVPLCVDML CLNWLNVYDTGRTGRIRVLSFKTGIIISLCKAHLEDK YRYLFKQVASST GFCDQRRRLGLLLHDSIQIPRQLGEVASFGGSNIEPSVRS CFQFANNKPEIE AALFLDWMRLEPQSMVWLPVLHRVAAAETAKHQAKCNICKECPIIGFR YRSLKHFNYDICQSCFFSGRVAKGHKMHYPMVEYCTPTTSGEDVRDFA KVLKNKFRTRKRYFAKHPRMGYLPVQTVLEGDNMETPAQILISLESEERG ELERILADLEEENRNLQAEYDRLKQQHEHKGLSPLSPSPDTM
SEQ ID NO:131 - BXA- 027742		MLWWEVEEDCYEREDVQKKTFKWWNAQFSKFGKQHIEENLFSDLQDG RRLDLLEGLTGQKLPKEKGSTRVHALNNVNKALRVLQNNNVDLVNIG STDIVDGNHKLTLGLIWNILHWQVKNVMKNIMAGLQQTNSEKILLSWV RQSTRNYPQVNVINFTTSWSDGLALNALIHSHRPDLFDWNSVVCQQSAT QRLEHAFNIARYQLGIEKLLDPEDVDTTYPDKKSILMYITSLFQVLPQQV SIEAIQEVEMLPKPPKVTKEEHFQLHHQMHYSQQITVSLAQGYERTSSPK PRFKSYAYTQAAYVTTSDPTRSPFPSQHLEAPEDKSFSGSSLMSEVNDR YQTALEEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTA HQGRVGNILQLGSKLIGTGKLSEDEETEVEQEQMNLLNSRWECLRVASM

	EKQSNLHRVLM DLQNQKLKELNDWLT KTEERTRKMEEEEPLGPDLEDLK RQVQQHKVLQEDLEQE QVRVNSLTHMVVVVDESSGDHATAALEEQLK VLGDRWANICRWTE DRWVLLQDILLKWQRLTEEQCLFSAWLSEKEDA VNKIHTTGFKDQNEMLSSLQKLAVLKADLEKKKQSMGKLYSLKQDLLS TLKNKSVTQKTEAWLDNFARCWDNLVQKLEKSTAQISQAIHTVREETM MVMTEDMPLEISYVPSTYLTEITHVSQALLEVEQLLNAPDLC AKDFEDLF KQEE SLKNIKDSLQSSGRIDIHHSKKTAAALQSATPVERVKLQEALSQ LDF QWEKVNKMYKDRQGRFDRSVEK WRRFH YDIKIFNQWLTEAEQFLRKT QIPENWEHAKYK WYLKELQDGIGQRQT VVRTLNATGEEIIQQSSKTDAS ILQEKLGS LNLRWQEVCKQLSDRKKRLEEQLERLQELQEATDELDLKL R QAEVIKGSWQPVGDLLIDSLQDHLEKVKALRGEIAPLKENVSHVNDLAR QLTTLGIQLSPYNLSTLEDLNTRWKLLQVAVEDRVRQLHEAHRDFGPAS QHFLSTSVQGPWER AISPNKVPYYINHETQTT CWDHPKMTELYQSLADL NNVRFSAYRTAMKLRR LQKALCLDLLSLSAACDALDQHNLKQNDQPM DILQIINCLTTIYDRLEQEHN NLVNVPLCVD MCLNWLLNVYDTGRTGRIR VLSFKTGIISLC KAHLEDKYRYL FKQVASSTGFCDQRR LGLLLHDSIQIPR QLGEVASFGGSNIEPSVRSCFQFANNKPEIEAALFDW MRLEPQSMVWL PVLHRVAAAETAKHQAKCNICKECPIIGFRYRSLKHFN YDICQSCFFSGR VAKGHKMHYPMVEYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRM GYLPVQTVLEGDNMETDTM
SEQ ID NO: 132 – BXA- 027743	MLWWEEVEDCYEREDVQKKTFTK WVNAAQFSKFGKQH IENLFSDLQDG RRLDLLLEGLTGQKL PKEKGSTRVHALNNVNKALRVLQNN NVDLVNIG STDIVDGNHKLTLGLIWN IILHWQVKNVMKNIMAGLQQTNSEKILLSWV RQSTRNYPQVNVINFTTSWSDGLALNALIHSR PDLFDWNSVVCQQSAT QRLEHAFNIARYQLGIEKLLDPEDVD TTPDKKSILMYITSLFQVLPQQV SIEAIQEVEMLP RPPKVTKEEHFQLHHQM HYSQQITVSLAQGYERTSSPK PRFKSYAYTQAAYVTTSDPTRSPFPSQH LEAPEDKSF GSSLMESVNLDR YQTAL EEVLSWLLSAEDTLQAQGEISNDVEVVKDQFH THEGYMMDLTA HQGRVGNILQLGSKLIGTGKLSEDEETE VQEQMNLNLSRWECLRVASM EKQSNLHIHTVREETM MVMTEDMPLEISYVPSTYLTEITHVSQALLEVEQ LLNAPDLC AKDFEDLFKQEE SLKNIKDSLQSSGRIDIHHSKKTAAALQSAT PVERVKLQEALSQ LDFQWEKVNKMYKDRQGRFDRSVEK WRRFH YDIKI FNQWLTEAEQFLRKTQIPENWEHAKYK WYLKELQDGIGQRQT VVRTLN ATGEEIIQQSSKTDASILQEKLGS LNLRWQEVCKQLSDRKKRLEEQA PGL TTIGASPTQTVTLVTQP VVTKET AISKLEMPSSLMLESDQWKRLHLSLQE LLVWLQLKDD ELSRQAPIGGDFPAVQKQNDVHRAFKRELKTKEPVMST LETVRIFLTEQPLEGLEKLYQEPREL PPEERAQNVTRLLRKQAE EVNTEW EKLNLHSADWQRKIDETLERLQELQEATDELDLKL RQAEVIKGSWQP VGDLLIDSLQDHLEKVKALRGEIAPLKENVSHVNDLARQLTTLGIQLSPYN LSTLEDLNTRWKLLQVAVEDRVRQLHEAHRDFGPASQHFLSTSVQGPW ER AISPNKVPYYINHETQTT CWDHPKMTELYQSLADLNNVRFSAYRTA MKLRR LQKALCLDLLSLSAACDALDQHNLKQNDQPM DILQIINCLTTIY DRLEQEHN NLVNVPLCVD MCLNWLLNVYDTGRTGRIRVLSFKTGIISLC KAHLEDKYRYL FKQVASSTGFCDQRR LGLLLHDSIQIPRQLGEVASFGGS NIEPSVRSCFQFANNKPEIEAALFDW MRLEPQSMVWL PVLHRVAAAET AKHQAKCNICKECPIIGFRYRSLKHFN YDICQSCFFSGR VAKGHKMHYPM MVEYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRM GYLPVQTVLEG DNMETPAQILISLESEERGELERILADLEEENRNLQAEYDRLKQQHEHKG

	LSPLPSPPDTM
--	-------------

In some embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NOs: 118-132, wherein the amino acid sequence when expressed has at least one dystrophin activity. In some embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to one of SEQ ID NO: 118 (BXA-196477), SEQ ID NO: 119 (BXA-196473), SEQ ID NO: 120 (BXA-196474), SEQ ID NO: 121 (BXA-196475), SEQ ID NO: 122 (BXA-196476), SEQ ID NO: 124 (BXA-196478), or SEQ ID NO: 125 (BXA-196479). In some embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 118, wherein the amino acid sequence when expressed has at least one dystrophin activity. In some embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 119, wherein the amino acid sequence when expressed has at least one dystrophin activity. In some embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 120, wherein the amino acid sequence when expressed has at least one dystrophin activity. In some embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 121, wherein the amino acid sequence when expressed has at least

one dystrophin activity. In some embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO:

5 122, wherein the amino acid sequence when expressed has at least one dystrophin activity.

In some embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 123, wherein the amino

10 acid sequence when expressed has at least one dystrophin activity. In some embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 124, wherein the amino acid sequence when

15 expressed has at least one dystrophin activity. In some embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 125, wherein the amino acid sequence when expressed has at least

20 one dystrophin activity.

In certain embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 126, wherein the amino

25 acid sequence when expressed has at least one dystrophin activity. In certain embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 127, wherein the amino acid sequence when

30 expressed has at least one dystrophin activity. In certain embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least

about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 128, wherein the amino acid sequence when expressed has at least one dystrophin activity. In certain embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 129, wherein the amino acid sequence when expressed has at least one dystrophin activity. In certain embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 130, wherein the amino acid sequence when expressed has at least one dystrophin activity. In certain embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 131, wherein the amino acid sequence when expressed has at least one dystrophin activity. In certain embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 132, wherein the amino acid sequence when expressed has at least one dystrophin activity.

In some embodiments, the miniaturized dystrophin polypeptide encoded by the nucleic acid molecule has formula (I):

H1-R1-L1-R3-H2-L2-L3-R16-R17-R24-H4

(I)

wherein: H1 is a hinge 1 domain of dystrophin; R1 is a spectrin repeat 1 domain of dystrophin; L1 is amino acids Arg-Val (RV); R3 is a spectrin repeat 3 domain of dystrophin; H2 is a hinge 2 domain of dystrophin; L2 is the amino acid sequence set forth in SEQ ID NO:74 (SEAQ); L3 is the amino acid sequence set forth in SEQ ID NO:75 (IHTVREE TMMVMTEDMP LEI); R16 is a spectrin repeat 16 of dystrophin; R17 is a

spectrin repeat 17 of dystrophin; R24 is a spectrin repeat 24 of dystrophin; H4 is a hinge 4 domain of dystrophin; and (-) is a peptide bond.

In some embodiments, the miniaturized dystrophin polypeptide encoded by the nucleic acid molecule exhibits one or more properties selected from the group consisting of (i) having
5 a lower CD4 proliferation compared to BXA-027741, (ii) having a lower CD8 proliferation compared to BXA-027741, and (iv) any combination thereof.

In some embodiments, the miniaturized dystrophin polypeptide encoded by the nucleic acid molecule has formula (II):

H1-R1-L-R16-R17-H3-R23-R24-H4

10 (II)

wherein: H1 is a hinge 1 domain of dystrophin; R1 is a spectrin repeat 1 domain of dystrophin; L is amino acids SEQ ID NO:84 (IHTVREETMMVMTEDEMPLEI); R16 is a spectrin repeat 16 of dystrophin; R17 is a spectrin repeat 17 of dystrophin; H3 is a hinge 3 domain of dystrophin; R23 is a spectrin repeat 23 of dystrophin; R24 is a spectrin repeat
15 24 of dystrophin; H4 is a hinge 4 domain of dystrophin; and (-) is a peptide bond.

In some embodiments, the miniaturized dystrophin polypeptide encoded by the nucleic acid molecule comprises an amino acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ
20 ID NO: 132.

In some embodiments, the nucleotide sequence comprises a sequence at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to the nucleotide sequence of SEQ
25 ID NO: 147.

In some embodiments, the miniaturized dystrophin polypeptide exhibits a higher expression of the miniaturized dystrophin polypeptide than BXA-196481. In some other embodiments, the miniaturized dystrophin polypeptide expression is at least about 1.5 fold, at least about 1.6 fold, at least about 1.7 fold, at least about 1.8 fold, at least about 1.9 fold,
30 at least about 2 fold, at least about 2.1 fold, at least about 2.2 fold, at least about 2.3 fold, at least about 2.4 fold, at least about 2.5 fold, at least about 2.6 fold, at least about 2.7 fold,

at least about 2.8 fold, at least about 2.9 fold or at least about 3 fold higher than the BXA-196481 polypeptide expression.

In some embodiments, the miniaturized dystrophin polypeptides can be encoded by nucleotide sequences. Some examples of the nucleotide sequences are shown in Table 8.

5 Table 8. Nucleotide Sequences of Dystrophin constructs.

SEQ NO and Description	ID	Sequence
SEQ NO:133 BXA-196477	-	ATGCTGTGGTGGGAGGAAGTGGAAAGATTGCTACGAGCGCGAGGACG TGCAGAAGAAAACCTTCACCAAATGGGTGAACGCCCAGTTCAGCAA GTTCGGCAAGCAGCACATCGAGAACCTGTTACGCGACCTGCAGGACG GCAGACGGCTGCTGGACCTGCTGGAAGGCCTGACCGGCCAGAAGCT GCCCCAAGAGAAGGGCAGCACCAGAGTGCACGCCCTGAACAACGTG AACAAGGCCCTGCGGGTGCTGCAGAACAACAACGTGGACCTGGTGA ACATCGGCAGCACCGACATCGTGGACGGCAACCACAAGCTGACCCTG GGCCTGATCTGGAACATCATCCTGCACTGGCAGGTCAAAAACGTGAT GAAGAACATCATGGCCGGCCTGCAGCAGACCAACAGCGAGAAGATC CTGCTGAGCTGGGTGCGCCAGAGCACCCGGAACCTACCCCCAGGTCAA CGTGATCAACTTCACCACCTCTTGGAGCGACGGCCTGGCCCTGAACG CCCTGATCCACAGCCACCGGCCCGACCTGTTTCGACTGGAACAGCGTG GTCTGCCAGCAGAGCGCCACCCAGCGGCTGGAACACGCCTTCAATAT CGCCAGATACCAGCTGGGCATCGAGAAGCTGCTGGATCCCGAGGAC GTGGACACCACCTACCCCGACAAGAAATCCATCCTGATGTATATCAC CAGCCTGTTCCAGGTGCTGCCCCAGCAGGTGTCCATCGAGGCCATCC AGGAAGTGGAAATGCTGCCCCAGACCCCCCAAAGTGACCAAAGAGGA ACACTTCCAGCTGCACCACCAGATGCACTACAGCCAGCAGATCACCG TGTCCCTGGCCCAGGGCTACGAGAGAACCAGCAGCCCCAAGCCCCGG TTCAAGAGCTACGCCTATACCCAGGCCGCTACGTGACCACCAGCGA CCCTACCAGAAGCCCATTCCCCAGCCAGCATCTGGAAGCCCCCGAGG ACAAGAGCTTCGGCAGCAGCCTGATGGAAAGCGAAGTGAACCTGGA CAGATACCAGACCGCCCTGGAAGAGGTGCTGAGCTGGCTGCTGAGCG CCGAGGATACACTGCAGGCTCAGGGCGAGATCAGCAACGACGTGGA AGTCGTGAAGGACCAGTTCCACACCCACGAGGGCTACATGATGGACC TGACAGCCCACCAGGGCAGAGTGGGCAACATCCTGCAGCTGGGCTCC AAGCTGATCGGCACCGGCAAGCTGAGCGAGGACGAAGAGACAGAGG TGCAGGAACAGATGAACCTGCTGAACAGCAGATGGGAGTGCCTGCG GGTGGCCAGCATGGAAAAGCAGAGCAACCTGCATAGGGTCCTGCTG AAGTGGCAGCGGCTGACCGAGGAACAGTGCCTGTTTAGCGCCTGGCT GTCCGAGAAAGAGGACGCCGTGAACAAGATCCACACCACCGGCTTC AAGGACCAGAACGAGATGCTGAGCAGCCTGCAGAAACTGGCCGTGC TGAAGGCCGACCTGGAAAAGAAAAAGCAGTCCATGGGCAAGCTGTA CTCCCTGAAGCAGGACCTGCTGTCCACCCTGAAGAACAAGAGCGTGA CCCAGAAAACCGAGGCCTGGCTGGACAACCTTCGCCCCGGTGCTGGGAC AACCTGGTGCAGAAGCTGGAAAAGTCCACCGCCCAGATCTCCAGGC CGTGACCACAACACAGCCCAGCCTGACCCAGACCACCGTGATGGAA

SEQ ID NO and Description	Sequence
	ACCGTGACAACAGTGACCACCCGGGAACAGATCCTCGTGAAGCACG CCAGGAAGAAGTCCCCCTCCACCCCCCAGAAGAAACGGCAGATC ACAGTGGACAGTGAAGCTCAGATCCACACCGTGCGGGAAGAGACAA TGATGGTCATGACAGAGGACATGCCCCCTGGAAATCAGCTACGTGCC AGCACCTACCTGACCGAGATCACCCATGTGTCCCAGGCCCTGCTGGA AGTGGAACAGCTGCTGAACGCCCCCGACCTGTGCGCCAAGGATTTCG AGGACCTGTTCAAGCAGGAAGAGAGCCTGAAGAATATCAAGGATAG CCTGCAGCAGAGCAGCGGCCGGATCGACATCATCCACAGCAAGAAA ACAGCCGCCCTGCAGAGCGCCACCCCCGTGGAAAGAGTGAAACTGC AGGAAGCCCTGTCCCAGCTGGACTTCCAGTGGGAGAAAGTGAACAA AATGTACAAGGACCGGCAGGGCAGATTTCGACCGCAGCGTGGAAGG TGGCGGCGGTTCCACTACGACATCAAGATCTTCAACCAGTGGCTGAC AGAGGCCGAGCAGTTCTGAGAAAGACCCAGATCCCCGAGAACTGG GAGCACGCCAAGTACAAGTGGTATCTGAAAGAGCTGCAGGACGGCA TCGGCCAGAGACAGACAGTCGTGCGGACCCTGAATGCCACCGGCGA GGAAATCATCCAGCAGTCCAGCAAGACCGACGCCAGCATTCTGCAGG AAAAGCTGGGCAGCCTGAACCTGCGGTGGCAGGAAGTGTGCAAGCA GCTGTCCGACCGGAAGAAGCGGCTGGAAGAACAGCTGGAACGGCTG CAGGAAGTGCAGGAGGCCACCGACGAGCTGGACCTGAAACTGAGAC AGGCCGAAGTGATCAAGGGCAGCTGGCAGCCAGTGGGCGACCTGCT GATCGACTCCCTGCAGGACCATCTGGAAAAAGTGAAGGCCCTGAGA GGCGAGATCGCCCCCTGAAAGAAAACGTGTCCACGTGAACGACCT GGCCCCGGCAGCTGACAACACTGGGCATTCAGCTGAGCCCCTACAACC TGAGCACACTGGAAGATCTGAACACCCGGTGGAAAGCTGCTGCAGGTG GCCGTGGAAGATAGAGTGCGGCAGCTGCACGAGGCCCACAGAGATT TTGGCCCTGCCTCCCAGCACTTCCTGAGCACCAGTGTGCAGGGCCCTT GGGAGAGAGCCATCTCCCCTAACAAGGTGCCCTACTACATCAACCAC GAGACACAGACCACCTGTTGGGACCACCCCAAGATGACCGAGCTGTA CCAGAGCCTGGCTGACCTGAACAACGTGCGGTTACAGCGCCTACCGGA CCGCCATGAAGCTGCGGAGACTGCAGAAAGCTCTGTGCCTGGATCTG CTGTCCCTGTCCGCCGCCTGTGATGCCCTGGACCAGCACAACTCTGAA GCAGAACGACCAGCCCATGGATATCCTGCAGATCATCAACTGCCTGA CCACCATCTACGACCGGCTGGAACAGGAACACAACAATCTCGTGAAC GTGCCCCCTGTGCGTGGACATGTGCCTGAATTGGCTGCTGAATGTGTA CGACACCGGCCGGACAGGCCGGATCAGAGTGCTGTCTTCAAGACCG GCATCATCAGCCTGTGCAAAGCCCACCTGGAAGATAAGTACCGGTAT CTGTTCAAACAGGTGGCCTCTAGCACCGGCTTTTGCAGCAGAGAAG GCTGGGCCTGCTGCTGCACGACAGCATCCAGATCCCTAGACAGCTGG GCGAAGTGGCCAGCTTTGGCGGCAGCAACATCGAGCCTAGCGTGCGG AGCTGCTTCCAGTTCGCCAACAAAGCCCGAGATCGAGGCCGCCCT GTTCTGGAAGTGGATGAGACTGGAACCCAGAGCATGGTGTGGCTGC CCGTGCTGCATAGAGTGGCCGCTGCCGAGACAGCCAAGCACCAGGCC AAGTGCAACATCTGCAAAGAGTGCCCCATCATCGGCTTCCGGTACAG AAGCCTGAAGCACTTCAACTACGATATCTGCCAGAGCTGTTTCTTCA GCGGCAGGGTGGCCAAGGGCCACAAAATGCACTACCCCATGGTGGA

SEQ ID NO and Description	Sequence
	ATACTGCACCCCCACCACAAGCGGCGAGGATGTGCGGGATTTCGCCA AGGTGCTGAAAAACAAGTTCGGGACCAAGCGGTACTTCGCCAAACAC CCCC GGATGGGCTACCTGCCTGTGCAGACAGTGCTGGAAGGCGACAA CATGGAAACCGACACCATGTGATGATGATTTGGGCAGAGCGATGGA GTCCTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGT TTTACAACCTCCTGATTCCCGCATGCGGCCAGCTTATCGATACCGTCGA AATAAAAGATCCTTATTTTCATTGGATCTGTGTGTTGGTTTTTTGTGT G
SEQ ID NO:134 - BXA- 196473	ATGCTGTGGTGGGAGGAAGTGGAAGATTGCTACGAGCGCGAGGACG TGCAGAAGAAAACCTTCACCAAATGGGTGAACGCCCAGTTCAGCAA GTTCGGCAAGCAGCACATCGAGAACCTGTTTCAGCGACCTGCAGGACG GCAGACGGCTGCTGGACCTGCTGGAAGGCCTGACCGGCCAGAAGCT GCCCAAAGAGAAGGGCAGCACCCAGAGTGCACGCCCTGAACAACGTG AACAAGGCCCTGCGGGTGCTGCAGAACAAACGTGGACCTGGTGA ACATCGGCAGCACCGACATCGTGGACGGCAACCACAAGCTGACCCTG GGCCTGATCTGGAACATCATCCTGCACTGGCAGGTCAAAAACGTGAT GAAGAACATCATGGCCGGCCTGCAGCAGACCAACAGCGAGAAGATC CTGCTGAGCTGGGTGCGCCAGAGCACCCGGAACCTACCCCCAGGTCAA CGTGATCAACTTCACCACCTCTTGAGCGACGGCCTGGCCCTGAACG CCCTGATCCACAGCCACCGGCCCGACCTGTTTCGACTGGAACAGCGTG GTCTGCCAGCAGAGCGCCACCCAGCGGCTGGAACACGCCTTCAATAT CGCCAGATACCAGCTGGGCATCGAGAAGCTGCTGGATCCCGAGGAC GTGGACACCACCTACCCCGACAAGAAATCCATCCTGATGTATATCAC CAGCCTGTTCCAGGTGCTGCCCCAGCAGGTGTCCATCGAGGCCATCC AGGAAGTGGAATGCTGCCCAGACCCCCCAAAGTGACCAAAGAGGA ACACTTCCAGCTGCACCACCAGATGCACTACAGCCAGCAGATCACCG TGTCCCTGGCCCAGGGCTACGAGAGAACCAGCAGCCCCAAGCCCCGG TTCAAGAGCTACGCCTATACCCAGGCCGCCTACGTGACCACCAGCGA CCCTACCAGAAGCCCATTCCCCAGCCAGCATCTGGAAGCCCCCGAGG ACAAGAGCTTCGGCAGCAGCCTGATGGAAAGCGAAGTGAACCTGGA CAGATACCAGACCGCCCTGGAAGAGGTGCTGAGCTGGCTGCTGAGCG CCGAGGATACACTGCAGGCTCAGGGCGAGATCAGCAACGACGTGGA AGTCGTGAAGGACCAGTTCCACACCCACGAGGGCTACATGATGGACC TGACAGCCCACCAGGGCAGAGTGGGCAACATCCTGCAGCTGGGCTCC AAGCTGATCGGCACCGGCAAGCTGAGCGAGGACGAAGAGACAGAGG TGCAGGAACAGATGAACCTGCTGAACAGCAGATGGGAGTGCCTGCG GGTGGCCAGCATGGAAAAGCAGAGCAACCTGCATCTGCTGAAGTGG CAGCGGCTGACCGAGGAACAGTGCCTGTTTAGCGCCTGGCTGTCCGA GAAAGAGGACGCCGTGAACAAGATCCACACCACCGGCTTCAAGGAC CAGAACGAGATGCTGAGCAGCCTGCAGAACTGGCCGTGCTGAAGG CCGACCTGGAAAAGAAAAAGCAGTCCATGGGCAAGCTGTACTCCCTG AAGCAGGACCTGCTGTCCACCCTGAAGAACAAGAGCGTGACCCAGA AAACCGAGGCCTGGCTGGACAACTTCGCCCCGGTGCTGGGACAACCTG GTGCAGAAGCTGGAAAAGTCCACCGCCCAGATCTCCAGGCCGTGAC CACAACACAGCCCAGCCTGACCCAGACCACCGTGATGGAAACCGTG

SEQ NO Description	ID and Sequence
	ACAACAGTGACCACCCGGGAACAGATCCTCGTGAAGCACGCCCAGG AAGAACTGCCCCCTCCACCCCCCAGAAGAAACGGCAGATCACAGTG GACATCCACACCGTGCGGGAAGAGACAATGATGGTCATGACAGAGG ACATGCCCCTGGAAATCAGCTACGTGCCCAGCACCTACCTGACCGAG ATCACCCATGTGTCCCAGGCCCTGCTGGAAGTGGAACAGCTGCTGAA CGCCCCCGACCTGTGCGCCAAGGATTTTCGAGGACCTGTTCAAGCAGG AAGAGAGCCTGAAGAATATCAAGGATAGCCTGCAGCAGAGCAGCGG CCGGATCGACATCATCCACAGCAAGAAAACAGCCGCCCTGCAGAGC GCCACCCCCGTGGAAAGAGTGAACTGCAGGAAGCCCTGTCCCAGCT GGACTTCCAGTGGGAGAAAGTGAACAAAATGTACAAGGACCGGCAG GGCAGATTCGACCGCAGCGTGGAAGAGTGGCGGCGGTTCCACTACG ACATCAAGATCTTCAACCAGTGGCTGACAGAGGCCGAGCAGTTCCTG AGAAAGACCCAGATCCCCGAGAACTGGGAGCACGCCAAGTACAAGT GGTATCTGAAAGAGCTGCAGGACGGCATCGGCCAGAGACAGACAGT CGTGCGGACCCTGAATGCCACCGGCGAGGAAATCATCCAGCAGTCCA GCAAGACCGACGCCAGCATTCTGCAGGAAAAGCTGGGCAGCCTGAA CCTGCGGTGGCAGGAAGTGTGCAAGCAGCTGTCCGACCGGAAGAAG CGGCTGGAAGAACAGCTGGAACGGCTGCAGGAACTGCAGGAGGCCA CCGACGAGCTGGACCTGAACTGAGACAGGCCGAAGTGATCAAGGG CAGCTGGCAGCCAGTGGGCGACCTGCTGATCGACTCCCTGCAGGACC ATCTGGAAAAAGTGAAGGCCCTGAGAGGCGAGATCGCCCCCTGAA AGAAAACGTGTCCCACGTGAACGACCTGGCCCCGGCAGCTGACAACA CTGGGCATTTCAGCTGAGCCCCTACAACCTGAGCACACTGGAAGATCT GAACACCCGGTGGAAAGCTGCTGCAGGTGGCCGTGGAAGATAGAGTG CGGCAGCTGCACGAGGCCACAGAGATTTTGGCCCTGCCTCCCAGCA CTTCCTGAGCACCAAGTGTGCAGGGCCCTTGGGAGAGAGCCATCTCCC CTAACAAGGTGCCCTACTACATCAACCACGAGACACAGACCACCTGT TGGGACCACCCCAAGATGACCGAGCTGTACCAGAGCCTGGCTGACCT GAACAACGTGCGGTTTCAGCGCCTACCGGACCGCCATGAAGCTGCGGA GACTGCAGAAAGCTCTGTGCCTGGATCTGCTGTCCCTGTCCGCCGCCT GTGATGCCCTGGACCAGCACAACTCTGAAGCAGAACGACCAGCCCATG GATATCCTGCAGATCATCAACTGCCTGACCACCATCTACGACCGGCT GGAACAGGAACACAACAATCTCGTGAACGTGCCCCCTGTGCGTGGACA TGTGCCTGAATTGGCTGCTGAATGTGTACGACACCGGCCGGACAGGC CGGATCAGAGTGCTGTCTTCAAGACCGGCATCATCAGCCTGTGCAA AGCCCACCTGGAAGATAAGTACCGGTATCTGTTCAAACAGGTGGCCT CTAGCACCGGCTTTTTCGACAGAGAAGGCTGGGCCTGCTGCTGCAC GACAGCATCCAGATCCCTAGACAGCTGGGCGAAGTGGCCAGCTTTGG CGGCAGCAACATCGAGCCTAGCGTGCGGAGCTGCTTCCAGTTCGCCA ACAACAAGCCCGAGATCGAGGCCGCCCTGTTCTTGGACTGGATGAGA CTGGAACCCAGAGCATGGTGTGGCTGCCCGTGCTGCATAGAGTGGC CGCTGCCGAGACAGCCAAGCACCAAGGCCAAGTGCAACATCTGCAAA GAGTGCCCCATCATCGGCTTCCGGTACAGAAGCCTGAAGCACTTCAA CTACGATATCTGCCAGAGCTGTTTCTTCAGCGGCAGGGTGGCCAAGG GCCACAAAATGCACTACCCCATGGTGGAATACTGCACCCCCACCACA

SEQ ID NO and Description	Sequence
	AGCGGCGAGGATGTGCGGGATTTGCGCAAAGGTGCTGAAAAACAAGT TCCGGACCAAGCGGTACTTCGCCAAACACCCCCGGATGGGCTACCTG CCTGTGCAGACAGTGCTGGAAGGCGACAACATGGAAACCGACACCA TGTGATGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCAT GACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACCTCCTGATTCC CGCATGCGGCCAGCTTATCGATACCGTCGAAATAAAAGATCCTTATT TTCATTGGATCTGTGTGTTGGTTTTTTGTGTG
SEQ ID NO:135 - BXA- 196474	ATGCTGTGGTGGGAGGAAGTGGAAGATTGCTACGAGCGCGAGGACG TGCAGAAGAAAACCTTCACCAAATGGGTGAACGCCCAGTTCAGCAA GTTCGGCAAGCAGCACATCGAGAACCTGTTACGCGACCTGCAGGACG GCAGACGGCTGCTGGACCTGCTGGAAGGCCTGACCGGCCAGAAGCT GCCCCAAGAGAAGGGCAGCACCAGAGTGCACGCCCTGAACAACGTG AACAAGGCCCTGCGGGTGCTGCAGAACAACAACGTGGACCTGGTGA ACATCGGCAGCACCGACATCGTGGACGGCAACCACAAGCTGACCCTG GGCCTGATCTGGAACATCATCCTGCACTGGCAGGTCAAAAACGTGAT GAAGAACATCATGGCCGGCCTGCAGCAGACCAACAGCGAGAAGATC CTGCTGAGCTGGGTGCGCCAGAGCACCCGGAACCTACCCCCAGGTCAA CGTGATCAACTTCACCACCTCTTGAGCGACGGCCTGGCCCTGAACG CCCTGATCCACAGCCACCGGCCCGACCTGTTTCGACTGGAACAGCGTG GTCTGCCAGCAGAGCGCCACCCAGCGGCTGGAACACGCCTTCAATAT CGCCAGATACCAGCTGGGCATCGAGAAGCTGCTGGATCCCGAGGAC GTGGACACCACCTACCCCGACAAGAAATCCATCCTGATGTATATCAC CAGCCTGTTCCAGGTGCTGCCCCAGCAGGTGTCCATCGAGGCCATCC AGGAAGTGGAATGCTGCCCCAGACCCCCCAAAGTGACCAAAGAGGA ACACTTCCAGCTGCACCACCAGATGCACTACAGCCAGCAGATCACCG TGTCCCTGGCCCAGGGCTACGAGAGAACCAGCAGCCCCAAGCCCCGG TTCAAGAGCTACGCCTATACCCAGGCCGCCTACGTGACCACCAGCGA CCCTACCAGAAGCCCATTCGCCAGCCAGCATCTGGAAGCCCCCGAGG ACAAGAGCTTCGGCAGCAGCCTGATGGAAAGCGAAGTGAACCTGGA CAGATACCAGACCGCCCTGGAAGAGGTGCTGAGCTGGCTGCTGAGCG CCGAGGATACACTGCAGGCTCAGGGCGAGATCAGCAACGACGTGGA AGTCGTGAAGGACCAGTTCCACACCCACGAGGGCTACATGATGGACC TGACAGCCCACCAGGGCAGAGTGGGCAACATCCTGCAGCTGGGCTCC AAGCTGATCGGCACCGGCAAGCTGAGCGAGGACGAAGAGACAGAGG TGCAGGAACAGATGAACCTGCTGAACAGCAGATGGGAGTGCCTGCG GGTGGCCAGCATGGAAAAGCAGAGCAACCTGCATCTGCTGAAGTGG CAGCGGCTGACCGAGGAACAGTGCCTGTTTAGCGCCTGGCTGTCCGA GAAAGAGGACGCCGTGAACAAGATCCACACCACCGGCTTCAAGGAC CAGAACGAGATGCTGAGCAGCCTGCAGAACTGGCCGTGCTGAAGG CCGACCTGGAAAAGAAAAAGCAGTCCATGGGCAAGCTGTACTCCCTG AAGCAGGACCTGCTGTCCACCCTGAAGAACAAGAGCGTGACCCAGA AAACCGAGGCCTGGCTGGACAACTTCGCCCCGGTGCTGGGACAACCTG GTGCAGAAGCTGGAAAAGTCCACCGCCCAGATCTCCAGGCCGTGAC CACAACACAGCCCAGCCTGACCCAGACCACCGTGATGGAAACCGTG ACAACAGTGACCACCCGGGAACAGATCCTCGTGAAGCACGCCCAGG

SEQ ID NO and Description	Sequence
	AAGAACTGCCCCCTCCACCCCCCAGAAGAAACGGCAGATCACAGTG GACAGTGAAGCTCAGATCCACACCGTGCGGGAAGAGACAATGATGG TCATGACAGAGGACATGCCCCTGGAAATCAGCTACGTGCCCAGCACC TACCTGACCGAGATCACCCATGTGTCCCAGGCCCTGCTGGAAGTGGA ACAGCTGCTGAACGCCCCCGACCTGTGCGCCAAGGATTTTCGAGGACC TGTTCAAGCAGGAAGAGAGCCTGAAGAATATCAAGGATAGCCTGCA GCAGAGCAGCGGCCGGATCGACATCATCCACAGCAAGAAAACAGCC GCCCTGCAGAGCGCCACCCCCGTGGAAAGAGTGAAACTGCAGGAAG CCCTGTCCCAGCTGGACTTCCAGTGGGAGAAAGTGAACAAAATGTAC AAGGACCGGCAGGGCAGATTTCGACCGCAGCGTGGAAGAGTGGCGGC GGTTCCACTACGACATCAAGATCTTCAACCAGTGGCTGACAGAGGCC GAGCAGTTCTTGAGAAAGACCCAGATCCCCGAGAACTGGGAGCACG CCAAGTACAAGTGGTATCTGAAAGAGCTGCAGGACGGCATCGGCCA GAGACAGACAGTCGTGCGGACCCTGAATGCCACCGGCGAGGAAATC ATCCAGCAGTCCAGCAAGACCGACGCCAGCATTTCTGCAGGAAAAGCT GGGCAGCCTGAACCTGCGGTGGCAGGAAGTGTGCAAGCAGCTGTCC GACCGGAAGAAGCGGCTGGAAGAACAGAAGAACATCCTGCAGGAAC TGCAGGAGGCCACCGACGAGCTGGACCTGAAACTGAGACAGGCCGA AGTGATCAAGGGCAGCTGGCAGCCAGTGGGCGACCTGCTGATCGACT CCCTGCAGGACCATCTGGAAAAAGTGAAGGCCCTGAGAGGCGAGAT CGCCCCCTGAAAGAAAACGTGTCCCACGTGAACGACCTGGCCCCGC AGCTGACAACACTGGGCATTACAGCTGAGCCCCCTACAACCTGAGCACA CTGGAAGATCTGAACACCCGGTGGAAGCTGCTGCAGGTGGCCGTGGA AGATAGAGTGCGGCAGCTGCACGAGGCCACAGAGATTTTGGCCCTG CCTCCCAGCACTTCCTGAGCACCAGTGTGCAGGGCCCTTGGGAGAGA GCCATCTCCCCTAACAAGGTGCCCTACTACATCAACCACGAGACACA GACCACCTGTTGGGACCACCCCAAGATGACCGAGCTGTACCAGAGCC TGGCTGACCTGAACAACGTGCGGTTTCAGCGCCTACCGGACCGCCATG AAGCTGCGGAGACTGCAGAAAGCTCTGTGCCTGGATCTGCTGTCCCT GTCCGCCGCCTGTGATGCCCTGGACCAGCACAATCTGAAGCAGAACG ACCAGCCCATGGATATCCTGCAGATCATCAACTGCCTGACCACCATC TACGACCGGCTGGAACAGGAACACAACAATCTCGTGAACGTGCCCT GTGCGTGGACATGTGCCTGAATTGGCTGCTGAATGTGTACGACACCG GCCGGACAGGCCGGATCAGAGTGCTGTCCTTCAAGACCGGCATCATC AGCCTGTGCAAAGCCCACCTGGAAGATAAGTACCGGTATCTGTTCAA ACAGGTGGCCTCTAGCACCGGCTTTTTCGACCCAGAGAAGGCTGGGCC TGCTGCTGCACGACAGCATCCAGATCCCTAGACAGCTGGGCGAAGTG GCCAGCTTTGGCGGCAGCAACATCGAGCCTAGCGTGCGGAGCTGCTT CCAGTTCGCCAACAACAAGCCCCGAGATCGAGGCCGCCCTGTTCTTG ACTGGATGAGACTGGAACCCCAGAGCATGGTGTGGCTGCCCCGTGCTG CATAGAGTGGCCGCTGCCGAGACAGCCAAGCACCAGGCCAAGTGCA ACATCTGCAAAGAGTGCCCCATCATCGGCTTCCGGTACAGAAGCCTG AAGCACTTCAACTACGATATCTGCCAGAGCTGTTTCTTCAGCGGCAG GGTGGCCAAGGGCCACAAAATGCACTACCCCATGGTGGAATACTGCA CCCCACCACAAGCGGCGAGGATGTGCGGGATTTTCGCCAAGGTGCTG

SEQ ID NO and Description	Sequence
	AAAAACAAGTTCCGGACCAAGCGGTACTTCGCCAAACACCCCCGGAT GGGCTACCTGCCTGTGCAGACAGTGCTGGAAGGCGACAACATGGAA ACCACACCATGTGATGATGATTTGGGCAGAGCGATGGAGTCCTTAGT ATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAAC CCTGATTCCCGCATGCGGCCAGCTTATCGATACCGTCGAAATAAAAG ATCCTTATTTTCATTGGATCTGTGTGTTGGTTTTTTGTGTG
SEQ ID NO:136 - BXA- 196475	ATGCTGTGGTGGGAGGAAGTGGAAGATTGCTACGAGCGCGAGGACG TGCAGAAGAAAACCTTCACCAAATGGGTGAACGCCCAGTTCAGCAA GTTTCGGCAAGCAGCACATCGAGAACCTGTTTCAGCGACCTGCAGGACG GCAGACGGCTGCTGGACCTGCTGGAAGGCCTGACCGGCCAGAAGCT GCCCCAAGAGAAGGGCAGCACCAGAGTGCACGCCCTGAACAACGTG AACAAGGCCCTGCGGGTGCTGCAGAACAACAACGTGGACCTGGTGA ACATCGGCAGCACCGACATCGTGGACGGCAACCACAAGCTGACCCTG GGCCTGATCTGGAACATCATCCTGCACTGGCAGGTCAAAAACGTGAT GAAGAACATCATGGCCGGCCTGCAGCAGACCAACAGCGAGAAGATC CTGCTGAGCTGGGTGCGCCAGAGCACCCGGAACCTACCCCCAGGTCAA CGTGATCAACTTCACCACCTCTTGAGCGACGGCCTGGCCCTGAACG CCCTGATCCACAGCCACCGGCCCGACCTGTTTCGACTGGAACAGCGTG GTCTGCCAGCAGAGCGCCACCCAGCGGCTGGAACACGCCTTCAATAT CGCCAGATACCAGCTGGGCATCGAGAAGCTGCTGGATCCCGAGGAC GTGGACACCACCTACCCCGACAAGAAATCCATCCTGATGTATATCAC CAGCCTGTTCCAGGTGCTGCCCCAGCAGGTGTCCATCGAGGGCCATCC AGGAAGTGGAATGCTGCCCAGACCCCCCAAAGTGACCAAAGAGGA ACACTTCCAGCTGCACCACCAGATGCACTACAGCCAGCAGATCACCG TGTCCCTGGCCCCAGGGCTACGAGAGAACCAGCAGCCCCAAGCCCCGG TTCAAGAGCTACGCCTATACCCAGGCCCGCCTACGTGACCACCAGCGA CCCTACCAGAAGCCCATTCGCCAGCCAGCATCTGGAAGCCCCCGAGG ACAAGAGCTTCGGCAGCAGCCTGATGGAAAGCGAAGTGAACCTGGA CAGATACCAGACCGCCCTGGAAGAGGTGCTGAGCTGGCTGCTGAGCG CCGAGGATACACTGCAGGCTCAGGGCGAGATCAGCAACGACGTGGA AGTCGTGAAGGACCAGTTCCACACCCACGAGGGCTACATGATGGACC TGACAGCCCACCAGGGCAGAGTGGGCAACATCCTGCAGCTGGGCTCC AAGCTGATCGGCACCGGCAAGCTGAGCGAGGACGAAGAGACAGAGG TGCAGGAACAGATGAACCTGCTGAACAGCAGATGGGAGTGCCTGCG GGTGGCCAGCATGGAAAAGCAGAGCAACCTGCATCTGCTGAAGTGG CAGCGGCTGACCGAGGAACAGTGCCTGTTTAGCGCCTGGCTGTCCGA GAAAGAGGACGCCGTGAACAAGATCCACACCACCGGCTTCAAGGAC CAGAACGAGATGCTGAGCAGCCTGCAGAACTGGCCGTGCTGAAGG CCGACCTGGAAAAGAAAAAGCAGTCCATGGGCAAGCTGTACTCCCTG AAGCAGGACCTGCTGTCCACCCTGAAGAACAAGAGCGTGACCCAGA AAACCGAGGCCTGGCTGGACAACTTCGCCCCGGTGCTGGGACAACCTG GTGCAGAAGCTGGAAAAGTCCACCGCCCAGATCTCCAGGCCGTGAC CACAACACAGCCCAGCCTGACCCAGACCACCGTGATGGAAACCGTG ACAACAGTGACCACCGGGAACAGATCCTCGTGAAGCACGCCAGG AAGAAGTGGCCCCCTCCACCCCCCAGAAGAAACGGCAGATCACAGTG

SEQ ID NO and Description	Sequence
	GACAGTGAAGCTCAGATCCACACCGTGCGGGAAGAGACAATGATGG TCATGACAGAGGACATGCCCCCTGGAAATCAGCTACGTGCCCAGCACC TACCTGACCGAGATCACCCATGTGTCCCAGGCCCTGCTGGAAGTGGA ACAGCTGCTGAACGCCCCCGACCTGTGCGCCAAGGATTTTCGAGGACC TGTTCAAGCAGGAAGAGAGCCTGAAGAATATCAAGGATAGCCTGCA GCAGAGCAGCGGCCGGATCGACATCATCCACAGCAAGAAAACAGCC GCCCTGCAGAGCGCCACCCCCGTGGAAAGAGTGAAACTGCAGGAAG CCCTGTCCCAGCTGGACTTCCAGTGGGAGAAAGTGAACAAAATGTAC AAGGACCGGCAGGGCAGATTTCGACCGCAGCGTGGAAGAGTGGCGGC GGTTCCACTACGACATCAAGATCTTCAACCAGTGGCTGACAGAGGCC GAGCAGTTCCTGAGAAAGACCCAGATCCCCGAGAACTGGGAGCACG CCAAGTACAAGTGGTATCTGAAAGAGCTGCAGGACGGCATCGGCCA GAGACAGACAGTCGTGCGGACCCTGAATGCCACCGGCGAGGAAATC ATCCAGCAGTCCAGCAAGACCGACGCCAGCATTCTGCAGGAAAAGCT GGGCAGCCTGAACCTGCGGTGGCAGGAAGTGTGCAAGCAGCTGTCC GACCGGAAGAAGCGGCTGGAAGAACAGAAGCTGGAACGGCTGCAGG AACTGCAGGAGGCCACCGACGAGCTGGACCTGAAACTGAGACAGGC CGAAGTGATCAAGGGCAGCTGGCAGCCAGTGGGCGACCTGCTGATC GACTCCCTGCAGGACCATCTGGAAAAAGTGAAGGCCCTGAGAGGCG AGATCGCCCCCCTGAAAGAAAACGTGTCCACGTGAACGACCTGGCC CGGCAGCTGACAACACTGGGCATTCTAGCTGAGCCCCCTACAACCTGAG CACACTGGAAGATCTGAACACCCGGTGGAAGCTGCTGCAGGTGGCCG TGGAAGATAGAGTGCGGCAGCTGCACGAGGCCCCACAGAGATTTTGG CCCTGCCTCCCAGCACTTCCTGAGCACCAAGTGTGCAGGGGCCCTTGGG AGAGAGCCATCTCCCCTAACAAGGTGCCCTACTACATCAACCACGAG ACACAGACCACCTGTTGGGACCACCCCAAGATGACCGAGCTGTACCA GAGCCTGGCTGACCTGAACAACGTGCGGTTTCAGCGCCTACCGGACCG CCATGAAGCTGCGGAGACTGCAGAAAGCTCTGTGCCTGGATCTGCTG TCCCTGTCCGCCGCCTGTGATGCCCTGGACCAGCACAACTCTGAAGCA GAACGACCAGCCCATGGATATCCTGCAGATCATCAACTGCCTGACCA CCATCTACGACCGGCTGGAACAGGAACACAACAATCTCGTGAACGTG CCCCTGTGCGTGGACATGTGCCTGAATTGGCTGCTGAATGTGTACGA CACCGGCCGGACAGGCCGGATCAGAGTGCTGTCCTTCAAGACCGGCA TCATCAGCCTGTGCAAAGCCCACCTGGAAGATAAGTACCGGTATCTG TTCAAACAGGTGGCCTCTAGCACCGGCTTTTTCGACCAAGAGAAGGCT GGGCCTGCTGCTGCACGACAGCATCCAGATCCCTAGACAGCTGGGCG AAGTGGCCAGCTTTGGCGGCAGCAACATCGAGCCTAGCGTGCGGAGC TGCTTCCAGTTCGCCAACAACAAGCCCGAGATCGAGGCCGCCCTGTT CCTGGACTGGATGAGACTGGAACCCCAAGAGCATGGTGTGGCTGCCCCG TGCTGCATAGAGTGGCCGCTGCCGAGACAGCCAAGCACCAGGCCAA GTGCAACATCTGCAAAGAGTGCCCCATCATCGGCTTCCGGTACAGAA GCCTGAAGCACTTCAACTACGATATCTGCCAGAGCTGTTTCTTCAGCG GCAGGGTGGCCAAGGGCCACAAAATGCACTACCCCATGGTGGGAATA CTGCACCCCCACCACAAGCGGCGAGGATGTGCGGGATTTCGCCAAGG TGCTGAAAAACAAGTTCGGACCAAGCGGTACTTCGCCAACAACCCCC

SEQ ID NO and Description	Sequence
	CGGATGGGCTACCTGCCTGTGCAGACAGTGCTGGAAGGCGACAACAT GGAAACCGACACCATGTGATGATGATTTGGGCAGAGCGATGGAGTCC TTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTA CAACTCCTGATTCCCGCATGCGGCCAGCTTATCGATACCGTCGAAAT AAAAGATCCTTATTTTCATTGGATCTGTGTGTTGGTTTTTTGTGTG
SEQ ID NO:137 - BXA- 196476	ATGCTGTGGTGGGAGGAAGTGGAAGATTGCTACGAGCGCGAGGACG TGCAGAAGAAAACCTTCACCAAATGGGTGAACGCCCAGTTCAGCAA GTTTCGGCAAGCAGCACATCGAGAACCTGTTTCAGCGACCTGCAGGACG GCAGACGGCTGCTGGACCTGCTGGAAGGCCTGACCGGCCAGAAAGCT GCCCCAAGAGAAGGGCAGCACCAGAGTGCACGCCCTGAACAACGTG AACAAGGCCCTGCGGGTGTGTCAGAACAAACAACGTGGACCTGGTGA ACATCGGCAGCACCGACATCGTGGACGGCAACCACAAGCTGACCCTG GGCCTGATCTGGAACATCATCCTGCACTGGCAGGTCAAAAACGTGAT GAAGAACATCATGGCCGGCCTGCAGCAGACCAACAGCGAGAAGATC CTGCTGAGCTGGGTGCGCCAGAGCACCCGGAACACCCCCAGGTCAA CGTGATCAACTTCACCACCTCTTGAGCGACGGCCTGGCCCTGAACG CCCTGATCCACAGCCACCGGCCCGACCTGTTTCGACTGGAACAGCGTG GTCTGCCAGCAGAGCGCCACCCAGCGGCTGGAACACGCCTTCAATAT CGCCAGATACCAGCTGGGCATCGAGAAGCTGCTGGATCCCGAGGAC GTGGACACCACCTACCCCGACAAGAAATCCATCCTGATGTATATCAC CAGCCTGTTCCAGGTGCTGCCCCAGCAGGTGTCCATCGAGGCCATCC AGGAAGTGGAATGCTGCCCCAGACCCCCCAAAGTGACCAAAGAGGA ACACTTCCAGCTGCACCACCAGATGCACTACAGCCAGCAGATCACCG TGTCCCTGGCCCAGGGCTACGAGAGAACCAGCAGCCCCAAGCCCCGG TTCAAGAGCTACGCCTATACCCAGGCCCGCCTACGTGACCACCAGCGA CCCTACCAGAAGCCCATTCCCCAGCCAGCATCTGGAAGCCCCCGAGG ACAAGAGCTTCGGCAGCAGCCTGATGGAAAGCGAAGTGAACCTGGA CAGATACCAGACCGCCCTGGAAGAGGTGCTGAGCTGGCTGCTGAGCG CCGAGGATACACTGCAGGCTCAGGGCGAGATCAGCAACGACGTGGA AGTCGTGAAGGACCAGTTCCACACCCACGAGGGCTACATGATGGACC TGACAGCCCACCAGGGCAGAGTGGGCAACATCCTGCAGCTGGGCTCC AAGCTGATCGGCACCGGCAAGCTGAGCGAGGACGAAGAGACAGAGG TGCAGGAACAGATGAACCTGCTGAACAGCAGATGGGAGTGCCTGCG GGTGGCCAGCATGGAAAAGCAGAGCAACCTGCATCTGCTGAAGTGG CAGCGGCTGACCGAGGAACAGTGCCTGTTTAGCGCCTGGCTGTCCGA GAAAGAGGACGCCGTGAACAAGATCCACACCACCGGCTTCAAGGAC CAGAACGAGATGCTGAGCAGCCTGCAGAAACTGGCCGTGCTGAAGG CCGACCTGGAAAAGAAAAAGCAGTCCATGGGCAAGCTGTACTCCCTG AAGCAGGACCTGCTGTCCACCCTGAAGAACAAGAGCGTGACCCAGA AAACCGAGGCCTGGCTGGACAACTTCGCCCCGGTGTGGGACAACCTG GTGCAGAAGCTGGAAAAGTCCACCGCCCAGATCTCCAGGCCGTGAC CACAACACAGCCCAGCCTGACCCAGACCACCGTGATGGAAACCGTG ACAACAGTGACCACCCGGGAACAGATCCTCGTGAAGCACGCCAGG AAGAAGTGGCCCCCTCCACCCCCCAGAAGAAACGGCAGATCACAGTG GACAGTGAAGCTCAGATCCACACCGTGCGGGAAGAGACAATGATGG

SEQ ID NO and Description	Sequence
	TCATGACAGAGGACATGCCCCTGGAAATCAGCTACGTGCCCAGCACC TACCTGACCGAGATCACCCATGTGTCCCAGGCCCTGCTGGAAGTGGA ACAGCTGCTGAACGCCCCGACCTGTGCGCCAAGGATTTTCGAGGACC TGTTCAAGCAGGAAGAGAGCCTGAAGAATATCAAGGATAGCCTGCA GCAGAGCAGCGGCCGGATCGACATCATCCACAGCAAGAAAACAGCC GCCCTGCAGAGCGCCACCCCCGTGGAAAGAGTGAAACTGCAGGAAG CCCTGTCCCAGCTGGACTTCCAGTGGGAGAAAGTGAACAAAATGTAC AAGGACCGGCAGGGCAGATTTCGACCGCAGCGTGGAAGAGTGGCGGC GGTTCCACTACGACATCAAGATCTTCAACCAGTGGCTGACAGAGGCC GAGCAGTTCCTGAGAAAGACCCAGATCCCCGAGAACTGGGAGCACG CCAAGTACAAGTGGTATCTGAAAGAGCTGCAGGACGGCATCGGCCA GAGACAGACAGTCGTGCGGACCCTGAATGCCACCGGCGAGGAAATC ATCCAGCAGTCCAGCAAGACCGACGCCAGCATTCTGCAGGAAAAGCT GGGCAGCCTGAACCTGCGGTGGCAGGAAGTGTGCAAGCAGCTGTCC GACCGGAAGAAGCGGCTGGAAGAACAGAAGAACATCCTGGAACGGC TGCAGGAACTGCAGGAGGCCACCGACGAGCTGGACCTGAAACTGAG ACAGGCCGAAGTGATCAAGGGCAGCTGGCAGCCAGTGGGCGACCTG CTGATCGACTCCCTGCAGGACCATCTGGAAAAAGTGAAGGCCCTGAG AGGCGAGATCGCCCCCTGAAAGAAAACGTGTCCCACGTGAACGAC CTGGCCCGGCAGCTGACAACACTGGGCATTTCAGCTGAGCCCCTACAA CCTGAGCACACTGGAAGATCTGAACACCCGGTGGAAGCTGCTGCAGG TGGCCGTGGAAGATAGAGTGCGGCAGCTGCACGAGGCCCCACAGAGA TTTTGGCCCTGCCTCCCAGCACTTCTGAGCACCAAGTGTGCAGGGCCC TTGGGAGAGAGCCATCTCCCCTAACAAGGTGCCCTACTACATCAACC ACGAGACACAGACCACCTGTTGGGACCACCCCAAGATGACCGAGCT GTACCAGAGCCTGGCTGACCTGAACAACGTGCGGTTTCAGCGCCTACC GGACCGCCATGAAGCTGCGGAGACTGCAGAAAGCTCTGTGCCTGGAT CTGCTGTCCCTGTCCGCCGCTGTGATGCCCTGGACCAGCACAACTG AAGCAGAACGACCAGCCCATGGATATCCTGCAGATCATCAACTGCCT GACCACCATCTACGACCGGCTGGAACAGGAACACAACAATCTCGTGA ACGTGCCCCTGTGCGTGGACATGTGCCTGAATTGGCTGCTGAATGTG TACGACACCGGCCGGACAGGCCGGATCAGAGTGCTGTCCTTCAAGAC CGGCATCATCAGCCTGTGCAAAGCCCACCTGGAAGATAAGTACCGGT ATCTGTTCAAACAGGTGGCCTCTAGCACCGGCTTTTGCAGCCAGAGA AGGCTGGGCCTGCTGCTGCACGACAGCATCCAGATCCCTAGACAGCT GGGCGAAGTGGCCAGCTTTGGCGGCAGCAACATCGAGCCTAGCGTGC GGAGCTGCTTCCAGTTCGCCAACAACAAGCCCGAGATCGAGGCCGCC CTGTTCCCTGGACTGGATGAGACTGGAACCCAGAGCATGGTGTGGCT GCCCCGTGCTGCATAGAGTGGCCGCTGCCGAGACAGCCAAGCACCCAG GCCAAGTGCAACATCTGCAAAGAGTGCCCCATCATCGGCTTCCGGTA CAGAAGCCTGAAGCACTTCAACTACGATATCTGCCAGAGCTGTTTCT TCAGCGGCAGGGTGGCCAAGGGCCACAAAATGCACTACCCCATGGT GGAATACTGCACCCCCACCACAAGCGGCGAGGATGTGCGGGATTTCG CCAAGGTGCTGAAAAACAAGTTCCGGACCAAGCGGTACTTCGCCAAA CACCCCCGGATGGGCTACCTGCCTGTGCAGACAGTGCTGGAAGGCGA

SEQ ID NO and Description	Sequence
	CAACATGGAAACCGACACCATGTGATGATGATTTGGGCAGAGCGATG GAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAA TGTTTTACAACCTCCTGATTCCCGCATGCGGCCAGCTTATCGATACCGT CGAAATAAAAGATCCTTATTTTCATTGGATCTGTGTGTTGGTTTTTG TGTG
SEQ ID NO:138 - BXA- 196477-Q	ATGCTGTGGTGGGAGGAAGTGGAAGATTGCTACGAGCGCGAGGACG TGCAGAAGAAAACCTTCACCAAATGGGTGAACGCCCAGTTCAGCAA GTTCGGCAAGCAGCACATCGAGAACCTGTTTCAGCGACCTGCAGGACG GCAGACGGCTGCTGGACCTGCTGGAAGGCCTGACCGGCCAGAAGCT GCCCCAAGAGAAGGGCAGCACCAGAGTGCACGCCCTGAACAACGTG AACAAGGCCCTGCGGGTGCTGCAGAACAACAACGTGGACCTGGTGA ACATCGGCAGCACCGACATCGTGGACGGCAACCACAAGCTGACCCTG GGCCTGATCTGGAACATCATCCTGCACTGGCAGGTCAAAAACGTGAT GAAGAACATCATGGCCGGCCTGCAGCAGACCAACAGCGAGAAGATC CTGCTGAGCTGGGTGCGCCAGAGCACCCGGAACACCCCCAGGTCAA CGTGATCAACTTCACCACCTCTTGAGCGACGGCCTGGCCCTGAACG CCCTGATCCACAGCCACCGGCCCGACCTGTTTCGACTGGAACAGCGTG GTCTGCCAGCAGAGCGCCACCCAGCGGCTGGAACACGCCTTCAATAT CGCCAGATACCAGCTGGGCATCGAGAAGCTGCTGGATCCCGAGGAC GTGGACACCACCTACCCCGACAAGAAATCCATCCTGATGTATATCAC CAGCCTGTTCCAGGTGCTGCCCCAGCAGGTGTCCATCGAGGCCATCC AGGAAGTGGAATGCTGCCCCAGACCCCCCAAAGTGACCAAAGAGGA ACACTTCCAGCTGCACCACCAGATGCACTACAGCCAGCAGATCACCG TGTCCCTGGCCCAGGGCTACGAGAGAACCAGCAGCCCCAAGCCCCGG TTCAAGAGCTACGCCTATACCCAGGCCGCCTACGTGACCACCAGCGA CCCTACCAGAAGCCCATTCCCCAGCCAGCATCTGGAAGCCCCCGAGG ACAAGAGCTTCGGCAGCAGCCTGATGGAAAGCGAAGTGAACCTGGA CAGATACCAGACCGCCCTGGAAGAGGTGCTGAGCTGGCTGCTGAGCG CCGAGGATACACTGCAGGCTCAGGGCGAGATCAGCAACGACGTGGA AGTCGTGAAGGACCAGTTCCACACCCACGAGGGCTACATGATGGACC TGACAGCCCACCAGGGCAGAGTGGGCAACATCCTGCAGCTGGGCTCC AAGCTGATCGGCACCGGCAAGCTGAGCGAGGACGAAGAGACAGAGG TGCAGGAACAGATGAACCTGCTGAACAGCAGATGGGAGTGCCTGCG GGTGGCCAGCATGGAAAAGCAGAGCAACCTGCATAGGGTCCCTGCTG AAGTGGCAGCGGCTGACCGAGGAACAGTGCCTGTTTAGCGCCTGGCT GTCCGAGAAAGAGGACGCCGTGAACAAGATCCACACCACCGGCTTC AAGGACCAGAACGAGATGCTGAGCAGCCTGCAGAACTGGCCGTGC TGAAGGCCGACCTGGAAAAGAAAAAGCAGTCCATGGGCAAGCTGTA CTCCCTGAAGCAGGACCTGCTGTCCACCCTGAAGAACAAGAGCGTGA CCCAGAAAACCGAGGCCTGGCTGGACAACTTCGCCCCGGTGCTGGGAC AACCTGGTGCAGAAGCTGGAAAAGTCCACCGCCCAGATCTCCAGGC CGTGACCACAACACAGCCCAGCCTGACCCAGACCACCGTGATGGAA ACCGTGACAACAGTGACCACCCGGGAACAGATCCTCGTGAAGCACG CCCAGGAAGAACTGCCCCCTCCACCCCCCAGAAGAAACGGCAGATC ACAGTGGACAGTGAAGCTCAGATCCACACCGTGCGGGAAGAGACAA

SEQ ID NO and Description	Sequence
	TGATGGTCATGACAGAGGACATGCCCCCTGGAAATCAGCTACGTGCCC AGCACCTACCTGACCGAGATCACCCATGTGTCCCAGGCCCTGCTGGA AGTGGAACAGCTGCTGAACGCCCCCGACCTGTGCGCCAAGGATTTTCG AGGACCTGTTCAAGCAGGAAGAGAGCCTGAAGAATATCAAGGATAG CCTGCAGCAGAGCAGCGGCCGGATCGACATCATCCACAGCAAGAAA ACAGCCGCCCTGCAGAGCGCCACCCCCGTGGAAAGAGTGAAACTGC AGGAAGCCCTGTCCCAGCTGGACTTCCAGTGGGAGAAAGTGAACAA AATGTACAAGGACCGGCAGGGCAGATTTCGACCGCAGCGTGGAAAAG TGGCGGCGGTTCCACTACGACATCAAGATCTTCAACCAGTGGCTGAC AGAGGCCGAGCAGTTTCTGAGAAAGACCCAGATCCCCGAGAACTGG GAGCACGCCAAGTACAAGTGGTATCTGAAAGAGCTGCAGGACGGCA TCGGCCAGAGACAGACAGTCGTGCGGACCCTGAATGCCACCGGCGA GGAAATCATCCAGCAGTCCAGCAAGACCGACGCCAGCATTCTGCAGG AAAAGCTGGGCAGCCTGAACCTGCGGTGGCAGGAAGTGTGCAAGCA GCTGTCCGACCGGAAGAAGCGGCTGGAAGAACAGCTGGAACGGCTG CAGGAACTGCAGGAGGCCACCGACGAGCTGGACCTGAAACTGAGAC AGGCCGAAGTGATCAAGGGCAGCTGGCAGCCAGTGGGCGACCTGCT GATCGACTCCCTGCAGGACCATCTGGAAAAAGTGAAGGCCCTGAGA GGCGAGATCGCCCCCCTGAAAGAAAACGTGTCCCACGTGAACGACCT GGCCCCGGCAGCTGACAACACTGGGCATTTCAGCTGAGCCCCCTACAACC TGAGCACACTGGAAGATCTGAACACCCGGTGGAAAGCTGCTGCAGGTG GCCGTGGAAGATAGAGTGCGGCAGCTGCACGAGGCCACAGAGATT TTGGCCCTGCCTCCCAGCACTTCTGAGCACCAAGTGTGCAGGGCCCTT GGGAGAGAGCCATCTCCCCTAACAAGGTGCCCTACTACATCAACCAC GAGACACAGACCACCTGTTGGGACCACCCCAAGATGACCGAGCTGTA CCAGAGCCTGGCTGACCTGAACAACGTGCGGTTTCAGCGCCTACCGGA CCGCCATGAAGCTGCGGAGACTGCAGAAAGCTCTGTGCCTGGATCTG CTGTCCCTGTCCGCCGCCTGTGATGCCCTGGACCAGCACAACTCTGAA GCAGAACGACCAGCCCATGGATATCCTGCAGATCATCAACTGCCTGA CCACCATCTACGACCGGCTGGAACAGGAACACAACAATCTCGTGAAC GTGCCCCCTGTGCGTGGACATGTGCCTGAATTGGCTGCTGAATGTGTA CGACACCGGCCGGACAGGCCGGATCAGAGTGCTGTCTTCAAGACCG GCATCATCAGCCTGTGCAAAGCCCACCTGGAAGATAAGTACCGGTAT CTGTTCAAACAGGTGGCCTCTAGCACCGGCTTTTGCGACCAGAGAAG GCTGGGCCTGCTGCTGCACGACAGCATCCAGATCCCTAGACAGCTGG GCGAAGTGGCCAGCTTTGGCGGCAGCAACATCGAGCCTAGCGTGCGG AGCTGCTTCCAGTTCGCCAACAACAAGCCCGAGATCGAGGCCGCCCT GTTCTGACTGGATGAGACTGGAACCCCAAGAGCATGGTGTGGCTGC CCGTGCTGCATAGAGTGGCCGCTGCCGAGACAGCCAAGCACCAGGCC AAGTGCAACATCTGCAAAGAGTGCCCCATCATCGGCTTCCGGTACAG AAGCCTGAAGCACTTCAACTACGATATCTGCCAGAGCTGTTTCTTCA GCGGCAGGGTGGCCAAGGGCCACAAAATGCACTACCCCATGGTGGGA ATACTGCACCCCCACCACAAGCGGCGAGGATGTGCGGGATTTCGCCA AGGTGCTGAAAAACAAGTTCGGGACCAAGCGGTACTTCGCCAAACAC CCCCGGATGGGCTACCTGCCTGTGCAGACAGTGCTGGAAGGCGACAA

SEQ ID NO and Description	Sequence
	CATGGAAACCGACACCATGTGATGATGATTTGGGCAGAGCGATGGA GTCCTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGT TTACAACCTCCTGATTCCCGCATGCGGCCAGCTTATCGATAACCGTCGA AATAAAAGATCCTTATTTTCATTGGATCTGTGTGTTGGTTTTTTGTGT G
SEQ ID NO:139 - BXA- 196478	ATGCTGTGGTGGGAGGAAGTGGAAGATTGCTACGAGCGCGAGGACG TGCAGAAGAAAACCTTCACCAAATGGGTGAACGCCCAGTTCAGCAA GTTTCGGCAAGCAGCACATCGAGAACCTGTTTCAGCGACCTGCAGGACG GCAGACGGCTGCTGGACCTGCTGGAAGGCCTGACCGGCCAGAAGCT GCCCCAAGAGAAGGGCAGCACCAGAGTGCACGCCCTGAACAACGTG AACAAGGCCCTGCGGGTGCTGCAGAACAACAACGTGGACCTGGTGA ACATCGGCAGCACCGACATCGTGGACGGCAACCACAAGCTGACCCTG GGCCTGATCTGGAACATCATCCTGCACTGGCAGGTCAAAAACGTGAT GAAGAACATCATGGCCGGCCTGCAGCAGACCAACAGCGAGAAGATC CTGCTGAGCTGGGTGCGCCAGAGCACCCGGAACACCCCCAGGTCAA CGTGATCAACTTCACCACCTCTTGAGCGACGGCCTGGCCCTGAACG CCCTGATCCACAGCCACCGGCCCGACCTGTTTCGACTGGAACAGCGTG GTCTGCCAGCAGAGCGCCACCCAGCGGCTGGAACACGCCTTCAATAT CGCCAGATACCAGCTGGGCATCGAGAAGCTGCTGGATCCCGAGGAC GTGGACACCACCTACCCCGACAAGAAATCCATCCTGATGTATATCAC CAGCCTGTTCCAGGTGCTGCCCCAGCAGGTGTCCATCGAGGCCATCC AGGAAGTGGAATGCTGCCCCAGACCCCCCAAAGTGACCAAAGAGGA ACACTTCCAGCTGCACCACCAGATGCACTACAGCCAGCAGATCACCG TGTCCCTGGCCCAGGGCTACGAGAGAACCAGCAGCCCCAAGCCCCGG TTCAAGAGCTACGCCTATACCCAGGCCGCCTACGTGACCACCAGCGA CCCTACCAGAAGCCCATTCCCCAGCCAGCATCTGGAAGCCCCCGAGG ACAAGAGCTTCGGCAGCAGCCTGATGGAAAGCGAAGTGAACCTGGA CAGATACCAGACCGCCCTGGAAGAGGTGCTGAGCTGGCTGCTGAGCG CCGAGGATACACTGCAGGCTCAGGGCGAGATCAGCAACGACGTGGA AGTCGTGAAGGACCAGTTCCACACCCACGAGGGCTACATGATGGACC TGACAGCCCACCAGGGCAGAGTGGGCAACATCCTGCAGCTGGGCTCC AAGCTGATCGGCACCGGCAAGCTGAGCGAGGACGAAGAGACAGAGG TGCAGGAACAGATGAACCTGCTGAACAGCAGATGGGAGTGCCTGCG GGTGGCCAGCATGGAAAAGCAGAGCAACCTGCATAGAGTACTGCTC CAAGACATCCTGCTGAAGTGGCAGCGGCTGACCGAGGAACAGTGCCT GTTTAGCGCCTGGCTGTCCGAGAAAGAGGACGCCGTGAACAAGATCC ACACCACCGGCTTCAAGGACCAGAACGAGATGCTGAGCAGCCTGCA GAAACTGGCCGTGCTGAAGGCCGACCTGGAAAAGAAAAAGCAGTCC ATGGGCAAGCTGTACTCCCTGAAGCAGGACCTGCTGTCCACCCTGAA GAACAAGAGCGTGACCCAGAAAACCGAGGCCTGGCTGGACAACTTC GCCCCGTGCTGGGACAACCTGGTGCAGAAGCTGGAAAAGTCCACCG CCCAGATCTCCAGGCCGTGACCACAACACAGCCCAGCCTGACCCAG ACCACCGTGATGGAAACCGTGACAACAGTGACCACCCGGGAACAGA TCCTCGTGAAGCACGCCAGGAAGAACTGCCCCCTCCACCCCCCAG

SEQ ID NO and Description	Sequence
	AAGAAACGGCAGATCACAGTGGACAGTGAAGCTCAGATCCACACCG TGCGGGAAGAGACAATGATGGTCATGACAGAGGACATGCCCCCTGGA AATCAGCTACGTGCCCAGCACCTACCTGACCGAGATCACCCATGTGT CCCAGGCCCTGCTGGAAGTGGAACAGCTGCTGAACGCCCCCGACCTG TGCGCCAAGGATTTTCGAGGACCTGTTCAAGCAGGAAGAGAGCCTGA AGAATATCAAGGATAGCCTGCAGCAGAGCAGCGGCCGGATCGACAT CATCCACAGCAAGAAAACAGCCGCCCTGCAGAGCGCCACCCCCGTG GAAAGAGTGAAACTGCAGGAAGCCCTGTCCCAGCTGGACTTCCAGTG GGAGAAAGTGAACAAAATGTACAAGGACCGGCAGGGCAGATTCGAC CGCAGCGTGGAAGTGGCGGCGGTTCCACTACGACATCAAGATCTT CAACCAGTGGCTGACAGAGGCCGAGCAGTTCTGAGAAAGACCCAG ATCCCCGAGAACTGGGAGCACGCCAAGTACAAGTGGTATCTGAAAG AGCTGCAGGACGGCATCGGCCAGAGACAGACAGTCGTGCGGACCCT GAATGCCACCGGCGAGGAAATCATCCAGCAGTCCAGCAAGACCGAC GCCAGCATTCTGCAGGAAAAGCTGGGCAGCCTGAACCTGCGGTGGCA GGAAGTGTGCAAGCAGCTGTCCGACCGGAAGAAGCGGCTGGAAGAA CAGCTGGAACGGCTGCAGGAACTGCAGGAGGCCACCGACGAGCTGG ACCTGAAACTGAGACAGGCCGAAGTGATCAAGGGCAGCTGGCAGCC AGTGGGCGACCTGCTGATCGACTCCCTGCAGGACCATCTGGAAAAAG TGAAGGCCCTGAGAGGCGAGATCGCCCCCTGAAAGAAAACGTGTC CCACGTGAACGACCTGGCCCCGGCAGCTGACAACACTGGGCATTGAGC TGAGCCCCCTACAACCTGAGCACACTGGAAGATCTGAACACCCGGTGG AAGCTGCTGCAGGTGGCCGTGGAAGATAGAGTGCGGCAGCTGCACG AGGCCCCACAGAGATTTTGGCCCTGCCTCCCAGCACTTCTGAGCACC AGTGTGCAGGGCCCTTGGGAGAGAGCCATCTCCCCTAACAAGGTGCC CTACTACATCAACCACGAGACACAGACCACCTGTTGGGACCACCCCA AGATGACCGAGCTGTACCAGAGCCTGGCTGACCTGAACAACGTGCGG TTCAGCGCCTACCGGACCGCCATGAAGCTGCGGAGACTGCAGAAAGC TCTGTGCCTGGATCTGCTGTCCCTGTCCGCCGCTGTGATGCCCTGGA CCAGCACAATCTGAAGCAGAACGACCAGCCCATGGATATCCTGCAGA TCATCAACTGCCTGACCACCATCTACGACCGGCTGGAACAGGAACAC AACAATCTCGTGAACGTGCCCTGTGCGTGGACATGTGCCTGAATTG GCTGCTGAATGTGTACGACACCGGCCGGACAGGCCGGATCAGAGTGC TGTCTTCAAGACCGGCATCATCAGCCTGTGCAAAGCCCACCTGGAA GATAAGTACCGGTATCTGTTCAAACAGGTGGCCTCTAGCACCGGCTT TTGCGACCAGAGAAGGCTGGGCCTGCTGCTGCACGACAGCATCCAGA TCCCTAGACAGCTGGGCGAAGTGGCCAGCTTTGGCGGCAGCAACATC GAGCCTAGCGTGCGGAGCTGCTTCCAGTTCGCCAACAACAAGCCCGA GATCGAGGCCGCCCTGTTCTTGACTGGATGAGACTGGAACCCCAGA GCATGGTGTGGCTGCCCCGTGCTGCATAGAGTGGCCGCTGCCGAGACA GCCAAGCACCAGGCCAAGTGCAACATCTGCAAAGAGTGCCCCATCAT CGGCTTCCGGTACAGAAGCCTGAAGCACTTCAACTACGATATCTGCC AGAGCTGTTTCTTCAGCGGCAGGGTGGCCAAGGGGCCACAAAATGCAC TACCCCATGGTGAATACTGCACCCCCACCACAAGCGGCGAGGATGT GCGGGATTTCCGCAAGGTGCTGAAAAACAAGTTCCGGACCAAGCGGT

SEQ ID NO and Description	Sequence
	ACTTCGCCAAACACCCCCGGATGGGCTACCTGCCTGTGCAGACAGTG CTGGAAGGCGACAACATGGAAACCGACACCATGTGATGATGATTTGG GCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGG AGCAGAATAAATGTTTTACAACCTCCTGATTCCCGCATGCGGCCAGCT TATCGATACCGTCGAAATAAAAGATCCTTATTTTCATTGGATCTGTGT GTTGGTTTTTTGTGTG
SEQ ID NO:140 - BXA- 196479	ATGCTGTGGTGGGAGGAAGTGGAAGATTGCTACGAGCGCGAGGACG TGCAGAAGAAAACCTTCACCAAATGGGTGAACGCCCAGTTCAGCAA GTTCGGCAAGCAGCACATCGAGAACCTGTTTCAGCGACCTGCAGGACG GCAGACGGCTGCTGGACCTGCTGGAAGGCCTGACCGGCCAGAAGCT GCCCCAAGAGAAGGGCAGCACCAGAGTGCACGCCCTGAACAACGTG AACAAGGCCCTGCGGGTGCTGCAGAACAACAACGTGGACCTGGTGA ACATCGGCAGCACCGACATCGTGGACGGCAACCACAAGCTGACCCTG GGCCTGATCTGGAACATCATCCTGCACTGGCAGGTCAAAAACGTGAT GAAGAACATCATGGCCGGCCTGCAGCAGACCAACAGCGAGAAGATC CTGCTGAGCTGGGTGCGCCAGAGCACCCGGAACCTACCCCCAGGTCAA CGTGATCAACTTCACCACCTCTTGAGCGACGGCCTGGCCCTGAACG CCCTGATCCACAGCCACCGGCCCGACCTGTTTCGACTGGAACAGCGTG GTCTGCCAGCAGAGCGCCACCCAGCGGCTGGAACACGCCTTCAATAT CGCCAGATACCAGCTGGGCATCGAGAAGCTGCTGGATCCCGAGGAC GTGGACACCACCTACCCCGACAAGAAATCCATCCTGATGTATATCAC CAGCCTGTTCCAGGTGCTGCCCCAGCAGGTGTCCATCGAGGGCCATCC AGGAAGTGGAATGCTGCCCAGACCCCCCAAAGTGACCAAAGAGGA ACACTTCCAGCTGCACCACCAGATGCACTACAGCCAGCAGATCACCG TGTCCCTGGCCCCAGGGCTACGAGAGAACCAGCAGCCCCAAGCCCCGG TTCAAGAGCTACGCCTATACCCAGGCCCGCCTACGTGACCACCAGCGA CCCTACCAGAAGCCCATTCCCCAGCCAGCATCTGGAAGCCCCCGAGG ACAAGAGCTTCGGCAGCAGCCTGATGGAAAGCGAAGTGAACCTGGA CAGATACCAGACCGCCCTGGAAGAGGTGCTGAGCTGGCTGCTGAGCG CCGAGGATACACTGCAGGCTCAGGGCGAGATCAGCAACGACGTGGA AGTCGTGAAGGACCAGTTCCACACCCACGAGGGCTACATGATGGACC TGACAGCCCACCAGGGCAGAGTGGGCAACATCCTGCAGCTGGGCTCC AAGCTGATCGGCACCGGCAAGCTGAGCGAGGACGAAGAGACAGAGG TGCAGGAACAGATGAACCTGCTGAACAGCAGATGGGAGTGCCTGCG GGTGGCCAGCATGGAAAAGCAGAGCAACCTGCATCTGCTGAAGTGG CAGCGGCTGACCGAGGAACAGTGCCTGTTTAGCGCCTGGCTGTCCGA GAAAGAGGACGCCGTGAACAAGATCCACACCACCGGCTTCAAGGAC CAGAACGAGATGCTGAGCAGCCTGCAGAACTGGCCGTGCTGAAGG CCGACCTGGAAAAGAAAAAGCAGTCCATGGGCAAGCTGTACTCCCTG AAGCAGGACCTGCTGTCCACCCTGAAGAACAAGAGCGTGACCCAGA AAACCGAGGCCTGGCTGGACAACTTCGCCCGGTGCTGGGACAACCTG GTGCAGAAGCTGGAAAAGTCCACCGCCCAGATCTCCAGGGCCGTGAC CACAACACAGCCCAGCCTGACCCAGACCACCGTGATGGAAACCGTG ACAACAGTGACCACCGGGAACAGATCCTCGTGAAGCACGCCAGG AAGAAGTGGCCCCCTCCACCCCCCAGAAGAAACGGCAGATCACAGTG

SEQ ID NO and Description	Sequence
	GACAGTGAAGCTCAGATCCACACCGTGCGGGAAGAGACAATGATGG TCATGACAGAGGACATGCCCCCTGGAAATCAGCTACGTGCCCAGCACC TACCTGACCGAGATCACCCATGTGTCCCAGGCCCTGCTGGAAGTGGA ACAGCTGCTGAACGCCCCCGACCTGTGCGCCAAGGATTTTCGAGGACC TGTTCAAGCAGGAAGAGAGCCTGAAGAATATCAAGGATAGCCTGCA GCAGAGCAGCGGCCGGATCGACATCATCCACAGCAAGAAAACAGCC GCCCTGCAGAGCGCCACCCCCGTGGAAAGAGTGAAACTGCAGGAAG CCCTGTCCCAGCTGGACTTCCAGTGGGAGAAAGTGAAACAAAATGTAC AAGGACCGGCAGGGCAGATTTCGACCGCAGCGTGGAAGAGTGGCGGC GGTTCCACTACGACATCAAGATCTTCAACCAGTGGCTGACAGAGGCC GAGCAGTTCCTGAGAAAGACCCAGATCCCCGAGAACTGGGAGCACG CCAAGTACAAGTGGTATCTGAAAGAGCTGCAGGACGGCATCGGCCA GAGACAGACAGTCGTGCGGACCCTGAATGCCACCGGCGAGGAAATC ATCCAGCAGTCCAGCAAGACCGACGCCAGCATTCTGCAGGAAAAGCT GGGCAGCCTGAACCTGCGGTGGCAGGAAGTGTGCAAGCAGCTGTCC GACCGGAAGAAGCGGCTGGAAGAACAGCTGGAACGGCTGCAGGAAC TGCAGGAGGCCACCGACGAGCTGGACCTGAAACTGAGACAGGCCGA AGTGATCAAGGGCAGCTGGCAGCCAGTGGGCGACCTGCTGATCGACT CCCTGCAGGACCATCTGGAAAAAGTGAAGGCCCTGAGAGGCGAGAT CGCCCCCTGAAAGAAAACGTGTCCACGTGAACGACCTGGCCCCGC AGCTGACAACACTGGGCATTTCAGCTGAGCCCCCTACAACCTGAGCACA CTGGAAGATCTGAACACCCGGTGGAAGCTGCTGCAGGTGGCCGTGGA AGATAGAGTGCGGCAGCTGCACGAGGCCACAGAGATTTTGGCCCTG CCTCCCAGCACTTCCTGAGCACCAGTGTGCAGGGGCCCTTGGGAGAGA GCCATCTCCCCTAACAAGGTGCCCTACTACATCAACCACGAGACACA GACCACCTGTTGGGACCACCCCAAGATGACCGAGCTGTACCAGAGCC TGGCTGACCTGAACAACGTGCGGTTTCAGCGCCTACCGGACCGCCATG AAGCTGCGGAGACTGCAGAAAGCTCTGTGCCTGGATCTGCTGTCCCT GTCCGCCGCTGTGATGCCCTGGACCAGCACAATCTGAAGCAGAACG ACCAGCCCATGGATATCCTGCAGATCATCAACTGCCTGACCACCATC TACGACCGGCTGGAACAGGAACACAACAATCTCGTGAACGTGCCCT GTGCGTGGACATGTGCCTGAATTGGCTGCTGAATGTGTACGACACCG GCCGGACAGGCCGGATCAGAGTGCTGTCCTTCAAGACCGGCATCATC AGCCTGTGCAAAGCCCACCTGGAAGATAAGTACCGGTATCTGTTCAA ACAGGTGGCCTCTAGCACCGGCTTTTTCGACCGAGAGAAGGCTGGGCC TGCTGCTGCACGACAGCATCCAGATCCCTAGACAGCTGGGCGAAGTG GCCAGCTTTGGCGGCAGCAACATCGAGCCTAGCGTGCGGAGCTGCTT CCAGTTCGCCAACAACAAGCCCGAGATCGAGGCCGCCCTGTTCTTG ACTGGATGAGACTGGAACCCCAGAGCATGGTGTGGCTGCCCCGTGCTG CATAGAGTGGCCGCTGCCGAGACAGCCAAGCACCAGGCCAAGTGCA ACATCTGCAAAGAGTGCCCCATCATCGGCTTCCGGTACAGAAGCCTG AAGCACTTCAACTACGATATCTGCCAGAGCTGTTTCTTCAGCGGCAG GGTGGCCAAGGGCCACAAAATGCACTACCCCATGGTGGAAATACTGCA CCCCACCACAAGCGGCGAGGATGTGCGGGATTTTCGCCAAGGTGCTG AAAAACAAGTTCCGGACCAAGCGGTACTTCGCCAAACACCCCCGGAT

SEQ ID NO and Description	Sequence
	GGGCTACCTGCCTGTGCAGACAGTGCTGGAAGGCGACAACATGGAA ACCGACACCATGTGATGATGATTTGGGCAGAGCGATGGAGTCCTTAG TATCAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAAC TCCTGATTCCCGCATGCGGCCAGCTTATCGATACCGTCGAAATAAAA GATCCTTATTTTCATTGGATCTGTGTGTTGGTTTTTTGTGTG
SEQ ID NO:141 - BXA- 196480	ATGCTGTGGTGGGAGGAAGTGGAAGATTGCTACGAGCGCGAGGACG TGCAGAAGAAAACCTTCACCAAATGGGTGAACGCCCAGTTCAGCAA GTTTCGGCAAGCAGCACATCGAGAACCTGTTTCAGCGACCTGCAGGACG GCAGACGGCTGCTGGACCTGCTGGAAGGCCTGACCGGCCAGAAGCT GCCCCAAGAGAAGGGCAGCACCAGAGTGCACGCCCTGAACAACGTG AACAAGGCCCTGCGGGTGCTGCAGAACAACAACGTGGACCTGGTGA ACATCGGCAGCACCGACATCGTGGACGGCAACCACAAGCTGACCCTG GGCCTGATCTGGAACATCATCCTGCACTGGCAGGTCAAAAACGTGAT GAAGAACATCATGGCCGGCCTGCAGCAGACCAACAGCGAGAAGATC CTGCTGAGCTGGGTGCGCCAGAGCACCCGGAACACCCCCAGGTCAA CGTGATCAACTTCACCACCTCTTGAGCGACGGCCTGGCCCTGAACG CCCTGATCCACAGCCACCGGCCCGACCTGTTTCGACTGGAACAGCGTG GTCTGCCAGCAGAGCGCCACCCAGCGGCTGGAACACGCCTTCAATAT CGCCAGATACCAGCTGGGCATCGAGAAGCTGCTGGATCCCGAGGAC GTGGACACCACCTACCCCGACAAGAAATCCATCCTGATGTATATCAC CAGCCTGTTCCAGGTGCTGCCCCAGCAGGTGTCCATCGAGGCCATCC AGGAAGTGGAATGCTGCCCCAGACCCCCCAAAGTGACCAAAGAGGA ACACTTCCAGCTGCACCACCAGATGCACTACAGCCAGCAGATCACCG TGTCCCTGGCTCAGGGCTACGAGCGGACCAGCAGCCCCAAGCCCCGG TTCAAGAGCTACGCCTACACCCAGGCCGCCTACGTGACCACCAGCGA CCCCACCAGAAGCCCATTCCCCAGCCAGCATCTGGAAGCCCCCGAGG ACAAGAGCTTCGGCAGCAGCCTGATGGAAAGCGAAGTGAACCTGGA CAGATACCAGACCGCCCTGGAAGAGGTGCTGTCCTGGCTGCTGAGCG CCGAGGATACACTGCAGGCCCAGGGCGAGATCAGCAACGACGTGGA AGTGGTGAAAGACCAGTTCCACACCCACGAGGGCTACATGATGGACC TGACCGCCCACCAGGGCAGAGTGGGCAACATCCTGCAGCTGGGCAG CAAGCTGATCGGCACCGGCAAGCTGAGCGAGGACGAAGAGACAGAG GTGCAGGAACAGATGAACCTGCTGAACAGCAGATGGGAGTGCCTGC GGGTGGCCAGCATGGAAAAGCAGAGCAACCTGCACATCCACACCGT GCGGGAAGAGACAATGATGGTGATGACCGAGGACATGCCCCCTGGAA ATCAGCTACGTGCCCAGCACCTACCTGACCGAGATCACCCACGTGTC CCAGGCTCTGCTGGAAGTGGAACAGCTGCTGAACGCCCCCGACCTGT GCGCCAAGGACTTCGAGGATCTGTTCAAGCAGGAAGAGAGCCTGAA GAATATCAAGGACTCCCTGCAGCAGTCCAGCGGCCGGATCGACATCA TCCACAGCAAGAAAACAGCCGCCCTGCAGTCCGCCACCCCCGTGGAA AGAGTGAAGCTGCAGGAAGCCCTGAGCCAGCTGGACTTCCAGTGGG AGAAAGTGAACAAGATGTACAAGGACCGGCAGGGCAGATTTCGACCG CAGCGTGGAAGAGTGGCGGCGGTTCCACTACGACATCAAGATCTTCA ACCAAGTGGCTGACCGAGGCCGAGCAGTTCCTGAGAAAGACCCAGAT CCCCGAGAACTGGGAGCACGCCAAGTACAAGTGGTATCTGAAAGAA

SEQ ID NO and Description	Sequence
	CTGCAGGATGGCATCGGCCAGCGGCAGACCGTGGTGCGCACACTGA ATGCCACCGGCGAGGAAATCATCCAGCAGAGCAGCAAGACCGACGC CAGCATCCTGCAGGAAAAGCTGGGCTCCCTGAACCTGCGGTGGCAGG AAGTGTGCAAGCAGCTGAGCGACCGGAAGAAGAGGCTGGAAGAACA GGCCCCTGGCCTGACCACCATCGGCGCCAGCCCTACCCAGACCGTGA CCCTGGTGACACAGCCCCTGGTGACAAAAGAGACAGCCATCTCCAAG CTGGAAATGCCCAGCTCCCTGATGCTGGAAGTGGAAAGGCTGCAGGA ACTGCAGGAGGCCACCGACGAGCTGGACCTGAAGCTGCGGCAGGCC GAAGTGATCAAGGGCAGCTGGCAGCCCCGTGGGCGACCTGCTGATCG ACAGCCTGCAGGACCACCTGGAAAAAGTGAAGGCCCTGAGAGGCGA GATCGCCCCCTGAAAGAAAACGTGTCCACGTGAACGACCTGGCCC GGCAGCTGACCACCCTGGGCATCCAGCTGAGCCCCTACAACCTGAGC ACCCTGGAAGATCTGAACACCCGGTGGAAAGCTGCTGCAGGTGGCCGT GGAAGATAGAGTGCGGCAGCTGCACGAGGCCACAGAGACTTTGGC CCTGCCAGCCAGCACTTCCTGAGCACCTCTGTGCAGGGACCCTGGGA GAGAGCCATCAGCCCCAACAAGGTGCCCTACTACATCAACCACGAGA CACAGACCACCTGTTGGGACCACCCCAAGATGACCGAGCTGTACCAG AGCCTGGCCGACCTGAACAATGTGCGGTTTCAGCGCCTACCGGACCGC CATGAAGCTGAGGCGGCTGCAGAAAGCTCTGTGCCTGGATCTGCTGA GCCTGAGCGCCGCTGCGACGCCCTGGACCAGCACAACTGAAGCAG AACGACCAGCCCATGGATATCCTGCAGATCATCAACTGCCTGACCAC AATCTACGACAGGCTGGAACAGGAACACAACAATCTGGTCAACGTG CCCCTGTGCGTGGACATGTGCCTGAATTGGCTGCTGAATGTGTACGA CACCGGCCGGACCGGCAGAATCCGGGTGCTGAGCTTCAAGACCGGC ATCATCAGCCTGTGCAAGGCCACCTGGAAGATAAGTACCGCTACCT GTTCAAACAGGTGGCCAGCTCCACCGGCTTTTTCGACACGCGGAGAC TGGGCCTGCTGCTGCACGACAGCATCCAGATCCCCAGACAGCTGGGC GAGGTGGCCTCCTTCGGCGGCAGCAACATTGAGCCCAGCGTGCGGAG CTGCTTCCAGTTCGCCAACAACAAGCCCGAGATCGAGGCCGCCCTGT TCCTGGACTGGATGAGACTGGAACCCCAGAGCATGGTGTGGCTGCCC GTGCTGCATCGGGTGGCCGCTGCCGAGACAGCCAAGCACCAGGCCA AGTGCAACATCTGCAAAGAGTGCCCCATCATCGGCTTCCGGTACAGA AGCCTGAAGCACTTCAACTACGATATCTGCCAGAGCTGCTTCTTCAG CGGCAGAGTGGCCAAGGGCCACAAAATGCACTACCCCATGGTGGAA TACTGCACCCCCACCACCAGCGGCGAGGATGTGCGGGACTTCGCCAA GGTGCTGAAAAACAAGTTCCGGACCAAGCGGTACTTTGCCAAGCACC CCCGGATGGGCTACCTGCCCCTGCAGACAGTGCTGGAAGGCGACAAC ATGGAAACCCCTGCCCAGATCCTGATCAGCCTGGAAAGCGAGGAAC GGGGCGAGCTGGAACGGATCCTGGCCGATCTGGAAGAGGAAAACCG GAACCTGCAGGCCGAGTACGACCGGCTGAAGCAGCAGCACGAGCAC AAGGGCCTGAGCCCCCTGCCTAGCCCCCTGACACCATGTGATGATG ATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAA GAAGGAGCAGAATAAATGTTTTACAACCTCCTGATTCCCGCATGCGGC CAGCTTATCGATACCGTCGAAATAAAAGATCCTTATTTTCATTGGATC TGTGTGTTGGTTTTTTTGTGTG

SEQ NO Description	ID and Sequence
SEQ NO:142 BXA- 196480- Variant	ATGCTGTGGTGGGAGGAAGTGGAAGATTGCTACGAGCGCGAGGACG TGCAGAAGAAAACCTTCACCAAATGGGTGAACGCCAGTTCAGCAA GTTCGGCAAGCAGCACATCGAGAACCTGTTTCAGCGACCTGCAGGACG GCAGACGGCTGCTGGACCTGCTGGAAGGCCTGACCGGCCAGAAGCT GCCCCAAGAGAAGGGCAGCACCAGAGTGCACGCCCTGAACAACGTG AACAAGGCCCTGCGGGTGCTGCAGAACAACAACGTGGACCTGGTGA ACATCGGCAGCACCGACATCGTGGACGGCAACCACAAGCTGACCCTG GGCCTGATCTGGAACATCATCCTGCACTGGCAGGTCAAAAACGTGAT GAAGAACATCATGGCCGGCCTGCAGCAGACCAACAGCGAGAAGATC CTGCTGAGCTGGGTGCGCCAGAGCACCCGGAACCTACCCCCAGGTCAA CGTGATCAACTTCACCACCTCTTGAGCGACGGCCTGGCCCTGAACG CCCTGATCCACAGCCACCGGCCCGACCTGTTCGACTGGAACAGCGTG GTCTGCCAGCAGAGCGCCACCCAGCGGCTGGAACACGCCTTCAATAT CGCCAGATACCAGCTGGGCATCGAGAAGCTGCTGGATCCCGAGGAC GTGGACACCACCTACCCCGACAAGAAATCCATCCTGATGTATATCAC CAGCCTGTTCCAGGTGCTGCCCCAGCAGGTGTCCATCGAGGCCATCC AGGAAGTGGAATGCTGCCCAGACCCCCCAAAGTGACCAAAGAGGA ACACTTCCAGCTGCACCACCAGATGCACTACAGCCAGCAGATCACCG TGTCCTGGCTCAGGGCTACGAGCGGACCAGCAGCCCCAAGCCCCGG TTCAAGAGCTACGCCTACACCCAGGCCGCCTACGTGACCACCAGCGA CCCCACCAGAAGCCCATTCCCCAGCCAGCATCTGGAAGCCCCCGAGG ACAAGAGCTTCGGCAGCAGCCTGATGGAAAGCGAAGTGAACCTGGA CAGATACCAGACCGCCCTGGAAGAGGTGCTGTCCTGGCTGCTGAGCG CCGAGGATACACTGCAGGCCCGAGGGCGAGATCAGCAACGACGTGGA AGTGGTGAAAGACCAGTTCCACACCCACGAGGGCTACATGATGGACC TGACCGCCACCAGGGCAGAGTGGGCAACATCCTGCAGCTGGGCAG CAAGCTGATCGGCACCGGCAAGCTGAGCGAGGACGAAGAGACAGAG GTGCAGGAACAGATGAACCTGCTGAACAGCAGATGGGAGTGCCTGC GGGTGGCCAGCATGGAAAAGCAGAGCAACCTGCACCGGGTGCTGAT CCACACCGTGCGGGAAGAGACAATGATGGTGATGACCGAGGACATG CCCCTGGAAATCAGCTACGTGCCAGCACCTACCTGACCGAGATCAC CCACGTGTCCCAGGCTCTGCTGGAAGTGGAACAGCTGCTGAACGCC CCGACCTGTGCGCCAAGGACTTCGAGGATCTGTTCAAGCAGGAAGAG AGCCTGAAGAATATCAAGGACTCCCTGCAGCAGTCCAGCGGCCGGAT CGACATCATCCACAGCAAGAAAACAGCCGCCCTGCAGTCCGCCACCC CCGTGGAAAGAGTGAAGCTGCAGGAAGCCCTGAGCCAGCTGGACTT CCAGTGGGAGAAAGTGAACAAGATGTACAAGGACCGGCAGGGCAGA TTCGACCGCAGCGTGGAAGTGGCGGCGGTTCCACTACGACATCAA GATCTTCAACCAGTGGCTGACCGAGGCCGAGCAGTTCCTGAGAAAGA CCCAGATCCCCGAGAACTGGGAGCACGCCAAGTACAAGTGGTATCTG AAAGAACTGCAGGATGGCATCGGCCAGCGGCAGACCGTGGTGCGCA CACTGAATGCCACCGGCGAGGAAATCATCCAGCAGAGCAGCAAGAC CGACGCCAGCATCCTGCAGGAAAAGCTGGGCTCCCTGAACCTGCGGT GGCAGGAAGTGTGCAAGCAGCTGAGCGACCGGAAGAAGAGGCTGGA AGAACAGAAGAACGACCTGGCCCCCTGGCCTGACCACCATCGGCGCC

SEQ NO Description	ID and Sequence
	AGCCCTACCCAGACCGTGACCCTGGTGACACAGCCCGTGGTGACAAA AGAGACAGCCATCTCCAAGCTGGAAATGCCCAGCTCCCTGATGCTGG AAGTGGAAGGCTGCAGGAACTGCAGGAGGCCACCGACGAGCTGGA CCTGAAGCTGCGGCAGGCCGAAGTGATCAAGGGCAGCTGGCAGCCC GTGGGCGACCTGCTGATCGACAGCCTGCAGGACCACCTGGAAAAAGT GAAGGCCCTGAGAGGCGAGATCGCCCCCCTGAAAGAAAACGTGTCC CACGTGAACGACCTGGCCCGGCAGCTGACCACCCTGGGCATCCAGCT GAGCCCCTACAACCTGAGCACCTTGGAAGATCTGAACACCCGGTGGA AGCTGCTGCAGGTGGCCGTGGAAGATAGAGTGCGGCAGCTGCACGA GGCCACAGAGACTTTGGCCCTGCCAGCCAGCACTTCCTGAGCACCT CTGTGCAGGGACCCTGGGAGAGAGCCATCAGCCCCAACAAAGGTGCC CTACTACATCAACCACGAGACACAGACCACCTGTTGGGACCACCCCA AGATGACCGAGCTGTACCAGAGCCTGGCCGACCTGAACAATGTGCGG TTCAGCGCCTACCGGACCGCCATGAAGCTGAGGCGGCTGCAGAAAGC TCTGTGCCTGGATCTGCTGAGCCTGAGCGCCGCCTGCGACGCCCTGG ACCAGCACAACTGAAGCAGAACGACCAGCCCATGGATATCCTGCA GATCATCAACTGCCTGACCACAATCTACGACAGGCTGGAACAGGAAC ACAACAATCTGGTCAACGTGCCCCTGTGCGTGGACATGTGCCTGAAT TGGCTGCTGAATGTGTACGACACCGGCCGGACCGGCAGAATCCGGGT GCTGAGCTTCAAGACCGGCATCATCAGCCTGTGCAAGGCCACCTGG AAGATAAGTACCGCTACCTGTTCAAACAGGTGGCCAGCTCCACCGGC TTTTGCGACCAGCGGAGACTGGGCCTGCTGCTGCACGACAGCATCCA GATCCCCAGACAGCTGGGCGAGGTGGCCTCCTTCGGCGGCAGCAACA TTGAGCCCAGCGTGCGGAGCTGCTTCCAGTTCGCCAACAAAGCCC GAGATCGAGGCCGCCCTGTTCTGACTGGATGAGACTGGAACCCCA GAGCATGGTGTGGCTGCCCCTGCTGCATCGGGTGGCCGCTGCCGAGA CAGCCAAGCACCAGGCCAAGTGCAACATCTGCAAAGAGTGCCCCATC ATCGGCTTCCGGTACAGAAGCCTGAAGCACTTCAACTACGATATCTG CCAGAGCTGCTTCTCAGCGGCAGAGTGGCCAAGGGCCACAAAATGC ACTACCCCATGGTGGAACTACTGCACCCCCACCACCAGCGGCGAGGAT GTGCGGGACTTCGCCAAGGTGCTGAAAAACAAGTTCCGGACCAAGC GGTACTTTGCCAAGCACCCCCGGATGGGCTACCTGCCCCTGCAGACA GTGCTGGAAGGCGACAACATGGAAACCCCTGCCCAGATCCTGATCAG CCTGGAAAGCGAGGAACGGGGCGAGCTGGAACGGATCCTGGCCGAT CTGGAAGAGGAAAACCGGAACCTGCAGGCCGAGTACGACCGGCTGA AGCAGCAGCACGAGCACAAGGGCCTGAGCCCCCTGCCTAGCCCCCT GACACCATGTGATGATGATTTGGGCAGAGCGATGGAGTCCTTAGTAT CAGTCATGACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACCTCC TGATTCCCGCATGCGGCCAGCTTATCGATACCGTCGAAATAAAAGAT CCTTATTTTCATTGGATCTGTGTGTTGGTTTTTTGTGTG
SEQ NO:143	ID - ATGCTGTGGTGGGAGGAAGTGGAAGATTGCTACGAGCGCGAGGACG TGCAGAAGAAAACCTTCACCAAATGGGTGAACGCCCAGTTCAGCAA GTTCCGGCAAGCAGCACATCGAGAACCTGTTTCAGCGACCTGCAGGACG GCAGACGGCTGCTGGACCTGCTGGAAGGCCTGACCGGCCAGAAGCT GCCCAAAGAGAAGGGCAGCACCAGAGTGCACGCCCTGAACAACGTG

SEQ ID NO and Description	Sequence
BXA- 196481	AACAAAGGCCCTGCGGGTGCTGCAGAACAAACGTGGACCTGGTGA ACATCGGCAGCACCGACATCGTGGACGGCAACCACAAGCTGACCCTG GGCCTGATCTGGAACATCATCCTGCACTGGCAGGTCAAAAACGTGAT GAAGAACATCATGGCCGGCCTGCAGCAGACCAACAGCGAGAAGATC CTGCTGAGCTGGGTGCGCCAGAGCACCCGGAAC TACCCCCAGGTCAA CGTGATCAACTTCACCACCTCTTGGAGCGACGGCCTGGCCCTGAACG CCCTGATCCACAGCCACCGGCCCCGACCTGTTCGACTGGAACAGCGTG GTCTGCCAGCAGAGCGCCACCCAGCGGCTGGAACACGCCTTCAATAT CGCCAGATACCAGCTGGGCATCGAGAAGCTGCTGGATCCCGAGGAC GTGGACACCACCTACCCCGACAAGAAATCCATCCTGATGTATATCAC CAGCCTGTTCCAGGTGCTGCCCCAGCAGGTGTCCATCGAGGCCATCC AGGAAGTGGAATGCTGCCCAGACCCCCCAAAGTGACCAAAGAGGA ACACTTCCAGCTGCACCACCAGATGCACTACAGCCAGCAGATCACCG TGTCCCTGGCTCAGGGCTACGAGCGGACCAGCAGCCCCAAGCCCCGG TTCAAGAGCTACGCCTACACCCAGGCCGCCTACGTGACCACCAGCGA CCCCACCAGAAGCCCATTCCCCAGCCAGCATCTGGAAGCCCCCGAGG ACAAGAGCTTCGGCAGCAGCCTGATGGAAAGCGAAGTGAACCTGGA CAGATACCAGACCGCCCTGGAAGAGGTGCTGTCCTGGCTGCTGAGCG CCGAGGATACACTGCAGGCCCAGGGCGAGATCAGCAACGACGTGGA AGTGGTGAAAGACCAGTTCCACACCCACGAGGGCTACATGATGGACC TGACCGCCCACCAGGGCAGAGTGGGCAACATCCTGCAGCTGGGCAG CAAGCTGATCGGCACCGGCAAGCTGAGCGAGGACGAAGAGACAGAG GTGCAGGAACAGATGAACCTGCTGAACAGCAGATGGGAGTGCCTGC GGGTGGCCAGCATGGAAAAGCAGAGCAACCTGCACAGCTACGTGCC CAGCACCTACCTGACCGAGATCACCCACGTGTCCCAGGCTCTGCTGG AAGTGGAACAGCTGCTGAACGCCCCCGACCTGTGCGCCAAGGACTTC GAGGATCTGTTCAAGCAGGAAGAGAGCCTGAAGAATATCAAGGACT CCCTGCAGCAGTCCAGCGGCCGGATCGACATCATCCACAGCAAGAAA ACAGCCGCCCTGCAGTCCGCCACCCCCGTGGAAAGAGTGAAGCTGCA GGAAGCCCTGAGCCAGCTGGACTTCCAGTGGGAGAAAGTGAACAAG ATGTACAAGGACCGGCAGGGCAGATTTCGACCGCAGCGTGGAAGT GGCGGCGGTTCCACTACGACATCAAGATCTTCAACCAGTGGCTGACC GAGGCCGAGCAGTTCTGAGAAAGACCCAGATCCCCGAGA ACTGGG AGCACGCCAAGTACAAGTGGTATCTGAAAGAACTGCAGGATGGCAT CGGCCAGCGGCAGACCGTGGTGCGCACACTGAATGCCACCGGCGAG GAAATCATCCAGCAGAGCAGCAAGACCGACGCCAGCATCCTGCAGG AAAAGCTGGGCTCCCTGAACCTGCGGTGGCAGGAAGTGTGCAAGCA GCTGAGCGACCGGAAGAAGAGGCTGGAAGAACAGCTGGAAGGACTG GAAAAGCTGTACCAGGAACCCAGAGAGCTGCCCCCGAGGAACGGG CCCAGAACGTGACCCGGCTGCTGCGGAAACAGGCCGAAGAGGTGAA CACCGAGTGGGAGAAGCTGAACCTGCACTCCGCCGACTGGCAGCGG AAGATCGACGAGACACTGGAAAGGCTGCAGGAACTGCAGGAGGCCA CCGACGAGCTGGACCTGAAGCTGCGGCAGGCCGAAGTGATCAAGGG CAGCTGGCAGCCCGTGGGCGACCTGCTGATCGACAGCCTGCAGGACC ACCTGGAAAAAGTGAAGGCCCTGAGAGGCGAGATCGCCCCCCTGAA

SEQ ID NO and Description	Sequence
	AGAAAACGTGTCCACGTGAACGACCTGGCCCGGCAGCTGACCACCC TGGGCATCCAGCTGAGCCCCCTACAACCTGAGCACCCCTGGAAGATCTG AACACCCGGTGGAAGCTGCTGCAGGTGGCCGTGGAAGATAGAGTGC GGCAGCTGCACGAGGCCACAGAGACTTTGGCCCTGCCAGCCAGCAC TTCCTGAGCACCTCTGTGCAGGGACCCTGGGAGAGAGCCATCAGCCC CAACAAGGTGCCCTACTACATCAACCACGAGACACAGACCACCTGTT GGGACCACCCCAAGATGACCGAGCTGTACCAGAGCCTGGCCGACCTG AACAATGTGCGGTTTCAGCGCCTACCGGACCGCCATGAAGCTGAGGCG GCTGCAGAAAGCTCTGTGCCTGGATCTGCTGAGCCTGAGCGCCGCT GCGACGCCCTGGACCAGCACAACCTGAAGCAGAACGACCAGCCCAT GGATATCCTGCAGATCATCAACTGCCTGACCACAATCTACGACAGGC TGGAACAGGAACACAACAATCTGGTCAACGTGCCCTGTGCGTGGAC ATGTGCCTGAATTGGCTGCTGAATGTGTACGACACCGGCCGACCGG CAGAATCCGGGTGCTGAGCTTCAAGACCGGCATCATCAGCCTGTGCA AGGCCACCTGGAAGATAAGTACCGCTACCTGTTCAAACAGGTGGCC AGCTCCACCGGCTTTTTCGACCGAGCGGAGACTGGGCCTGCTGCTGCA CGACAGCATCCAGATCCCCAGACAGCTGGGCGAGGTGGCCTCCTTCG GCGGCAGCAACATTGAGCCCAGCGTGCGGAGCTGCTTCCAGTTCGCC AACAACAAGCCCGAGATCGAGGCCGCCCTGTTCTGGACTGGATGAG ACTGGAACCCCAAGAGCATGGTGTGGCTGCCCCTGCTGCATCGGGTGG CCGCTGCCGAGACAGCCAAGCACCAGGCCAAGTGCAACATCTGCAA AGAGTGCCCCATCATCGGCTTCCGGTACAGAAGCCTGAAGCACTTCA ACTACGATATCTGCCAGAGCTGCTTCTTCAGCGGCAGAGTGGCCAAG GGCCACAAAATGCACTACCCCATGGTGGAAATACTGCACCCCCACCAC CAGCGGCGAGGATGTGCGGGACTTCGCCAAGGTGCTGAAAAACAAG TTCCGGACCAAGCGGTACTTTGCCAAGCACCCCCGGATGGGCTACCT GCCCCTGCAGACAGTGCTGGAAGGCGACAACATGGAAACCCCTGCC CAGATCCTGATCAGCCTGGAAAGCGAGGAACGGGGCGAGCTGGAAC GGATCCTGGCCGATCTGGAAGAGGAAAACCGGAACCTGCAGGCCGA GTACGACCGGCTGAAGCAGCAGCACGAGCACAAGGGCCTGAGCCCC CTGCCTAGCCCCCCTGACACCATGTGATGATGATTTGGGCAGAGCGA TGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATA AATGTTTTACAACCTCCTGATTCCCGCATGCGGCCAGCTTATCGATACC GTCGAAATAAAAGATCCTTATTTTCATTGGATCTGTGTGTTGGTTTTT TGTGTG
SEQ ID NO:144 - BXA- 027741	ATGCTGTGGTGGGAGGAAGTGGAAGATTGCTACGAGCGCGAGGACG TGCAGAAGAAAACCTTCACCAAATGGGTGAACGCCCAGTTCAGCAA GTTTCGGCAAGCAGCACATCGAGAACCTGTTTCAGCGACCTGCAGGACG GCAGACGGCTGCTGGACCTGCTGGAAGGCCTGACCGGCCAGAAGCT GCCCAAAGAGAAGGGCAGCACCAGAGTGCACGCCCTGAACAACGTG AACAAGGCCCTGCGGGTGCTGCAGAACAAACGTGGACCTGGTGA ACATCGGCAGCACCGACATCGTGGACGGCAACCACAAGCTGACCCTG GGCCTGATCTGGAACATCATCCTGCACTGGCAGGTCAAAAACGTGAT GAAGAACATCATGGCCGGCCTGCAGCAGACCAACAGCGAGAAGATC CTGCTGAGCTGGGTGCGCCAGAGCACCCGGAACCTACCCCCAGGTCAA

SEQ ID NO and Description	Sequence
	CGTGATCAACTTCACCACCTCTTGGAGCGACGGCCTGGCCCTGAACG CCCTGATCCACAGCCACCGGCCCGACCTGTTCGACTGGAACAGCGTG GTCTGCCAGCAGAGCGCCACCCAGCGGCTGGAACACGCCTTCAATAT CGCCAGATACCAGCTGGGCATCGAGAAGCTGCTGGATCCCGAGGAC GTGGACACCACCTACCCCGACAAGAAATCCATCCTGATGTATATCAC CAGCCTGTTCCAGGTGCTGCCCCAGCAGGTGTCCATCGAGGCCATCC AGGAAGTGGAATGCTGCCCAGACCCCCAAAGTGACCAAAGAGGA ACACTTCCAGCTGCACCACCAGATGCACTACAGCCAGCAGATCACCG TGTCCCTGGCCCAGGGCTACGAGAGAACCAGCAGCCCCAAGCCCCGG TTCAAGAGCTACGCCTATACCCAGGCCGCCTACGTGACCACCAGCGA CCCTACCAGAAGCCCATTCCCCAGCCAGCATCTGGAAGCCCCCGAGG ACAAGAGCTTCGGCAGCAGCCTGATGGAAAGCGAAGTGAACCTGGA CAGATACCAGACCGCCCTGGAAGAGGTGCTGAGCTGGCTGCTGAGCG CCGAGGATACACTGCAGGCTCAGGGCGAGATCAGCAACGACGTGGA AGTCGTGAAGGACCAGTTCCACACCCACGAGGGCTACATGATGGACC TGACAGCCCACCAGGGCAGAGTGGGCAACATCCTGCAGCTGGGCTCC AAGCTGATCGGCACCGGCAAGCTGAGCGAGGACGAAGAGACAGAGG TGCAGGAACAGATGAACCTGCTGAACAGCAGATGGGAGTGCCTGCG GGTGGCCAGCATGGAAAAGCAGAGCAACCTGCATCTGCTGAAGTGG CAGCGGCTGACCGAGGAACAGTGCCTGTTTAGCGCCTGGCTGTCCGA GAAAGAGGACGCCGTGAACAAGATCCACACCACCGGCTTCAAGGAC CAGAACGAGATGCTGAGCAGCCTGCAGAACTGGCCGTGCTGAAGG CCGACCTGGAAAAGAAAAAGCAGTCCATGGGCAAGCTGTACTCCCTG AAGCAGGACCTGCTGTCCACCCTGAAGAACAAGAGCGTGACCCAGA AAACCGAGGCCTGGCTGGACAACTTCGCCCGGTGCTGGGACAACCTG GTGCAGAAGCTGGAAAAGTCCACCGCCCAGATCTCCCAGGCCGTGAC CACAACACAGCCCAGCCTGACCCAGACCACCGTGATGGAAACCGTG ACAACAGTGACCACCCGGGAACAGATCCTCGTGAAGCACGCCCAGG AAGAACTGCCCCCTCCACCCCCCAGAAGAAACGGCAGATCACAGTG GACATCCACACCGTGCGGGAAGAGACAATGATGGTCATGACAGAGG ACATGCCCCTGGAAATCAGCTACGTGCCCAGCACCTACCTGACCGAG ATCACCCATGTGTCCCAGGCCCTGCTGGAAGTGGAACAGCTGCTGAA CGCCCCCGACCTGTGCGCCAAGGATTTCGAGGACCTGTTCAAGCAGG AAGAGAGCCTGAAGAATATCAAGGATAGCCTGCAGCAGAGCAGCGG CCGGATCGACATCATCCACAGCAAGAAAACAGCCGCCCTGCAGAGC GCCACCCCCGTGGAAAGAGTGAAACTGCAGGAAGCCCTGTCCCAGCT GGACTTCCAGTGGGAGAAAGTGAACAAAATGTACAAGGACCGGCAG GGCAGATTCGACCGCAGCGTGGAAGAGTGGCGGCGGTTCCACTACG ACATCAAGATCTTCAACCAGTGGCTGACAGAGGCCGAGCAGTTCCTG AGAAAGACCCAGATCCCCGAGAACTGGGAGCACGCCAAGTACAAGT GGTATCTGAAAGAGCTGCAGGACGGCATCGGCCAGAGACAGACAGT CGTGCGGACCCTGAATGCCACCGGCGAGGAAATCATCCAGCAGTCCA GCAAGACCGACGCCAGCATTCTGCAGGAAAAGCTGGGCAGCCTGAA CCTGCGGTGGCAGGAAGTGTGCAAGCAGCTGTCCGACCGGAAGAAG CGGCTGGAAGAACAGCTGGAACGGCTGCAGGAAGTGCAGGAGGCCA

SEQ ID NO and Description	Sequence
	CCGACGAGCTGGACCTGAAACTGAGACAGGCCGAAGTGATCAAGGG CAGCTGGCAGCCAGTGGGCGACCTGCTGATCGACTCCCTGCAGGACC ATCTGGAAAAAGTGAAGGCCCTGAGAGGCGAGATCGCCCCCTGAA AGAAAACGTGTCCCACGTGAACGACCTGGCCCGGCAGCTGACAACA CTGGGCATTTCAGCTGAGCCCCTACAACCTGAGCACACTGGAAGATCT GAACACCCGGTGGGAAGCTGCTGCAGGTGGCCGTGGAAGATAGAGTG CGGCAGCTGCACGAGGCCCACAGAGATTTTGGCCCTGCCTCCCAGCA CTTCCTGAGCACCAGTGTGCAGGGCCCTTGGGAGAGAGCCATCTCCC CTAACAAGGTGCCCTACTACATCAACCACGAGACACAGACCACCTGT TGGGACCACCCCAAGATGACCGAGCTGTACCAGAGCCTGGCTGACCT GAACAACGTGCGGTTTACGCGCCTACCGGACCGCCATGAAGCTGCGGA GACTGCAGAAAGCTCTGTGCCTGGATCTGCTGTCCCTGTCCGCCGCCT GTGATGCCCTGGACCAGCACAATCTGAAGCAGAACGACCAGCCCATG GATATCCTGCAGATCATCAACTGCCTGACCACCATCTACGACCGGCT GGAACAGGAACACAACAATCTCGTGAACGTGCCCTGTGCGTGGACA TGTGCCTGAATTGGCTGCTGAATGTGTACGACACCGGCCGGACAGGC CGGATCAGAGTGCTGTCTTCAAGACCGGCATCATCAGCCTGTGCAA AGCCCACCTGGAAGATAAGTACCGGTATCTGTTCAAACAGGTGGCCT CTAGCACCGGCTTTTTCGACCAAGAGAAGGCTGGGCCTGCTGCTGCAC GACAGCATCCAGATCCCTAGACAGCTGGGCGAAGTGGCCAGCTTTGG CGGCAGCAACATCGAGCCTAGCGTGCGGAGCTGCTTCCAGTTCCGCA ACAACAAGCCCGAGATCGAGGCCGCCCTGTTCTTGGACTGGATGAGA CTGGAACCCCGAGAGCATGGTGTGGCTGCCCGTGCTGCATAGAGTGGC CGCTGCCGAGACAGCCAAGCACCAGGCCAAGTGCAACATCTGCAAA GAGTGCCCCATCATCGGCTTCCGGTACAGAAGCCTGAAGCACTTCAA CTACGATATCTGCCAGAGCTGTTTCTTCAGCGGCAGGGTGGCCAAGG GCCACAAAATGCACTACCCCATGGTGGAACTACTGCACCCCCACCACA AGCGGCGAGGATGTGCGGGATTTTCGCCAAGGTGCTGAAAAACAAGT TCCGGACCAAGCGGTACTTCGCCAAACACCCCCGGATGGGCTACCTG CCTGTGCAGACAGTGCTGGAAGGCGACAACATGGAAACCCCTGCCCA GATCCTGATCAGCCTGGAATCCGAGGAACGGGGCGAGCTGGAACGG ATCCTGGCCGATCTGGAAGAGGAAAACCGGAACCTGCAGGCCGAGT ACGACCGGCTGAAGCAGCAGCACGAGCACAAGGGCCTGAGCCCCCT GCCTAGCCCCCCTGACACCATGTGATGATGATTTGGGCAGAGCGATG GAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGAGCAGAATAAA TGTTTACAACCTCCTGATTCCCGCATGCGGCCAGCTTATCGATACCGT CGA
SEQ ID NO:145 – BXA- 027744	ATGCTGTGGTGGGAGGAAGTGGAAGATTGCTACGAGCGCGAGGACG TGCAGAAGAAAACCTTCACCAAATGGGTGAACGCCAGTTCAGCAA GTTTCGGCAAGCAGCACATCGAGAACCTGTTTCAGCGACCTGCAGGACG GCAGACGGCTGCTGGACCTGCTGGAAGGCCTGACCGGCCAGAAGCT GCCCAAAGAGAAGGGCAGCACCAGAGTGCACGCCCTGAACAACGTG AACAAGGCCCTGCGGGTGCTGCAGAACAACAACGTGGACCTGGTGA ACATCGGCAGCACCAGACATCGTGGACGGCAACCACAAGCTGACCCTG GGCCTGATCTGGAACATCATCCTGCACTGGCAGGTCAAAAACGTGAT

SEQ ID NO and Description	Sequence
	GAAGAACATCATGGCCGGCCTGCAGCAGACCAACAGCGAGAAGATC CTGCTGAGCTGGGTGCGCCAGAGCACCCGGAAC TACCCCCAGGTCAA CGTGATCAACTTCACCACCTCTTGGAGCGACGGCCTGGCCCTGAACG CCCTGATCCACAGCCACCGGCCCCGACCTGTTTCGACTGGAACAGCGTG GTCTGCCAGCAGAGCGCCACCCAGCGGCTGGAACACGCCTTCAATAT CGCCAGATACCAGCTGGGCATCGAGAAGCTGCTGGATCCCGAGGAC GTGGACACCACCTACCCCGACAAGAAATCCATCCTGATGTATATCAC CAGCCTGTTCCAGGTGCTGCCCCAGCAGGTGTCCATCGAGGCCATCC AGGAAGTGGAATGCTGCCCAGACCCCCCAAAGTGACCAAAGAGGA AACTTCCAGCTGCACCACCAGATGCACTACAGCCAGCAGATCACCG TGTCCTGGCTCAGGGCTACGAGCGGACCAGCAGCCCCAAGCCCCGG TTCAAGAGCTACGCCTACACCCAGGCCGCCTACGTGACCACCAGCGA CCCCACCAGAAGCCCATTCCCCAGCCAGCATCTGGAAGCCCCCGAGG ACAAGAGCTTCGGCAGCAGCCTGATGGAAAGCGAAGTGAACCTGGA CAGATACCAGACCGCCCTGGAAGAGGTGCTGTCCTGGCTGCTGAGCG CCGAGGATACACTGCAGGCCCCAGGGCGAGATCAGCAACGACGTGGA AGTGGTGAAAGACCAGTTCCACACCCACGAGGGCTACATGATGGACC TGACCGCCCACCAGGGCAGAGTGGGCAACATCCTGCAGCTGGGCAG CAAGCTGATCGGCACCGGCAAGCTGAGCGAGGACGAAGAGACAGAG GTGCAGGAACAGATGAACCTGCTGAACAGCAGATGGGAGTGCCTGC GGGTGGCCAGCATGGAAAAGCAGAGCAACCTGCACATCCACACCGT GCGGGAAGAGACAATGATGGTGATGACCGAGGACATGCCCTGGAA ATCAGCTACGTGCCCAGCACCTACCTGACCGAGATCACCCACGTGTC CCAGGCTCTGCTGGAAGTGGAACAGCTGCTGAACGCCCCCGACCTGT GCGCCAAGGACTTCGAGGATCTGTTCAAGCAGGAAGAGAGCCTGAA GAATATCAAGGACTCCCTGCAGCAGTCCAGCGGCCGGATCGACATCA TCCACAGCAAGAAAACAGCCGCCCTGCAGTCCGCCACCCCCGTGGAA AGAGTGAAGCTGCAGGAAGCCCTGAGCCAGCTGGACTTCCAGTGGG AGAAAGTGAACAAGATGTACAAGGACCGGCAGGGCAGATTTCGACCG CAGCGTGGA AAAAGTGGCGGCGGTTCCACTACGACATCAAGATCTTCA ACAGTGGCTGACCGAGGCCGAGCAGTTCCTGAGAAAGACCCAGAT CCCCGAGAACTGGGAGCACGCCAAGTACAAGTGGTATCTGAAAGAA CTGCAGGATGGCATCGGCCAGCGGCAGACCGTGGTGCGCACACTGA ATGCCACCGGCGAGGAAATCATCCAGCAGAGCAGCAAGACCGACGC CAGCATCCTGCAGGAAAAGCTGGGCTCCCTGAACCTGCGGTGGCAGG AAGTGTGCAAGCAGCTGAGCGACCGGAAGAAGAGGCTGGAAGAACA GGCCCCTGGCCTGACCACCATCGGCGCCAGCCCTACCCAGACCGTGA CCCTGGTGACACAGCCCGTGGTGACAAAAGAGACAGCCATCTCCAAG CTGGAAATGCCCAGCTCCCTGATGCTGGAAGTACCTGCTCTGGCAGA TTTCAACCGGGCTTGGACAGAACTTACCGACTGGCTTTCTCTGCTTGA TCAAGTTATAAAATCACAGAGGGTGATGGTGGGTGACCTTGAGGATA TCAACGAGATGATCATCAAGCAGAAGGCAACAATGCAGGATTTGGA ACAGAGGCGTCCCCAGTTGGAAGAACTCATTACCGCTGCCCAAATT TGAAAAACAAGACCAGCAATCAAGAGGCTAGAACAATCATTACGGA TCGAATTGAAAGAATTGAGAATCAGTGGGATGAAGTACAAGAACAC

SEQ ID NO and Description	Sequence
	CTTCAGAACCGGAGGCAACAGTTGAATGAAATGCTGGAAAGGCTGC AGGAACTGCAGGAGGCCACCGACGAGCTGGACCTGAAGCTGCGGCA GGCCGAAGTGATCAAGGGCAGCTGGCAGCCCGTGGGCGACCTGCTG ATCGACAGCCTGCAGGACCACCTGGAAAAAGTGAAGGCCCTGAGAG GCGAGATCGCCCCCTGAAAGAAAACGTGTCCACGTGAACGACCTG GCCCCGGCAGCTGACCACCCTGGGCATCCAGCTGAGCCCCTACAACCT GAGCACCCCTGGAAGATCTGAACACCCGGTGGAAAGCTGCTGCAGGTG GCCGTGGAAGATAGAGTGCGGCAGCTGCACGAGGCCCACAGAGACT TTGGCCCTGCCAGCCAGCACTTCCTGAGCACCTCTGTGCAGGGACCC TGGGAGAGAGCCATCAGCCCCAACAAGGTGCCCTACTACATCAACCA CGAGACACAGACCACCTGTTGGGACCACCCCAAGATGACCGAGCTGT ACCAGAGCCTGGCCGACCTGAACAATGTGCGGTTTCAGCGCCTACCGG ACCGCCATGAAGCTGAGGCGGCTGCAGAAAGCTCTGTGCCTGGATCT GCTGAGCCTGAGCGCCGCTGCGACGCCCTGGACCAGCACAACTGA AGCAGAACGACCAGCCCATGGATATCCTGCAGATCATCAACTGCCTG ACCACAATCTACGACAGGCTGGAACAGGAACACAACAATCTGGTCA ACGTGCCCCTGTGCGTGGACATGTGCCTGAATTGGCTGCTGAATGTG TACGACACCGGCCGGACCGGCAGAATCCGGGTGCTGAGCTTCAAGAC CGGCATCATCAGCCTGTGCAAGGCCCACCTGGAAGATAAGTACCGCT ACCTGTTCAAACAGGTGGCCAGCTCCACCGGCTTTTTCGACACGCGG AGACTGGGCCTGCTGCTGCACGACAGCATCCAGATCCCCAGACAGCT GGGCGAGGTGGCCTCCTTCGGCGGCAGCAACATTGAGCCCAGCGTGC GGAGCTGCTTCCAGTTCGCCAACAACAAGCCCGAGATCGAGGCCGCC CTGTTCTTGGACTGGATGAGACTGGAACCCAGAGCATGGTGTGGCT GCCCCGTGCTGCATCGGGTGGCCGCTGCCGAGACAGCCAAGCACCAGG CCAAGTGCAACATCTGCAAAGAGTGCCCCATCATCGGCTTCCGGTAC AGAAGCCTGAAGCACTTCAACTACGATATCTGCCAGAGCTGCTTCTT CAGCGGCAGAGTGGCCAAGGGCCACAAAATGCACTACCCCATGGTG GAATACTGCACCCCCACCACCAGCGGCGAGGATGTGCGGGGACTTCGC CAAGGTGCTGAAAAACAAGTTCCGGACCAAGCGGTACTTTGCCAAGC ACCCCCGGATGGGCTACCTGCCCCTGCAGACAGTGCTGGAAGGCGAC AACATGGAAACCCCTGCCAGATCCTGATCAGCCTGGAAAGCGAGG AACGGGGCGAGCTGGAACGGATCCTGGCCGATCTGGAAGAGGAAAA CCGGAACCTGCAGGCCGAGTACGACCGGCTGAAGCAGCAGCACGAG CACAAGGGCCTGAGCCCCCTGCCTAGCCCCCTGACACCATGTGATG ATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGAT GAAGAAGGAGCAGAATAAATGTTTTACAACCTCCTGATTCCTGCATGC GGCCAGCTTATCGATACCGTCGAAATAAAAGATCCTTATTTTCATTGG ATCTGTGTGTTGGTTTTTTGTGTG
SEQ ID NO:146 -	ATGCTGTGGTGGGAGGAAGTGAAGATTGCTACGAGCGCGAGGACG TGCAGAAAGAAAACCTTCACCAAATGGGTCAACGCCCAAGTTCAGCAAG TTCGGCAAGCAGCACATCGAGAACCTGTTTCAGCGACCTGCAGGACGG CAGACGGCTGCTGGATCTGCTGGAAGGCCTGACCGGACAGAAGCTGC CCAAAGAGAAGGGCAGCACCAGAGTGCACGCCCTGAACAACGTGAA

SEQ ID NO and Description	Sequence
BXA- 027742	CAAGGCCCTGCGGGTGCTGCAGAACAACAACGTGGACCTCGTGAAC ATCGGCAGCACCGACATCGTGGACGGCAACCACAAGCTGACCCTGG GCCTGATCTGGAACATCATCCTGCACTGGCAAGTGAAGAACGTGATG AAGAACATCATGGCCGGCCTGCAGCAGACCAACAGCGAGAAGATCC TGCTGAGCTGGGTGCGCCAGAGCACCAGAAACTACCCCCAAGTGAAC GTGATCAACTTCACCACCTCTTGGAGCGACGGCCTGGCCCTGAATGC CCTGATCCACAGCCACAGACCCGACCTGTTCGACTGGAACAGCGTCG TGTGTCAGCAGAGCGCCACCCAGAGGCTGGAACACGCCTTCAATATC GCCAGATACCAGCTGGGCATCGAGAAGCTGCTGGACCCCGAGGATGT GGACACCACCTACCCCGACAAGAAATCCATCCTGATGTATATCACCA GCCTGTTCCAGGTGCTGCCTCAGCAGGTGTCCATCGAGGCCATCCAG GAAGTGGAATGCTGCCCGAGACCCCCCAAAGTGACCAAAGAGGAAC ACTTCCAGCTGCACCACCAGATGCACTACTCTCAGCAGATCACCGTG TCCCTGGCCCAGGGCTACGAGAGAACCAGCAGCCCCAAGCCCCGGTT CAAGAGCTACGCCTATACCCAGGCCGCCTACGTGACCACCAGCGACC CTACCAGAAGCCCATTCCCCAGCCAGCATCTGGAAGCCCCCGAGGAC AAGAGCTTCGGCAGCAGCCTGATGGAAAGCGAAGTGAACCTGGATA GATACCAGACCGCCCTGGAAGAGGTGCTGTCCTGGCTGCTGAGCGCC GAGGATACACTGCAGGCTCAGGGCGAGATCAGCAACGACGTGGAAG TCGTGAAGGACCAGTTCCACACCCACGAGGGCTACATGATGGACCTG ACAGCCCACCAGGGCAGAGTGGGCAACATTCTGCAGCTGGGCTCCAA GCTGATCGGCACCGGCAAGCTGAGCGAGGACGAAGAGACAGAGGTG CAGGAACAGATGAACCTGCTGAACAGCAGATGGGAGTGCCTGAGAG TGGCCAGCATGGAAAAGCAGAGCAACCTGCACCGGGTGCTGATGGA TCTGCAGAACCAGAAGCTGAAAGAGCTGAACGACTGGCTGACCAAG ACCGAGGAACGGACCCGGAAGATGGAAGAGGAACCCCTGGGACCAG ACCTGGAAGATCTGAAAAGACAGGTGCAGCAGCACAAAGTGCTGCA GGAAGATCTGGAACAGGAACAGGTGCGCGTGAACAGCCTGACCCAC ATGGTCGTGGTGGTGGACGAGAGCAGCGGCGATCATGCCACAGCCG CTCTGGAAGAACAGCTGAAGGTGCTGGGCGACAGATGGGCCAACAT CTGCCGGTGGACCGAGGATAGATGGGTGCTGCTGCAGGACATCCTGC TGAAGTGGCAGCGGCTGACAGAGGAACAGTGCCTGTTCTCCGCCTGG CTGTCCGAGAAAGAGGACGCCGTGAACAAAATCCACACCACCGGCTT CAAGGACCAGAACGAGATGCTGAGCAGCCTGCAGAAACTGGCCGTG CTGAAGGCCGACCTGGAAAAGAAAAAGCAGTCCATGGGCAAGCTGT ACTCCCTGAAGCAGGACCTGCTGTCCACCCTGAAGAACAAGAGCGTG ACCCAGAAAACCGAGGCCTGGCTGGACAACTTCGCCCCGGTGCTGGGA CAACCTGGTGCAGAAGCTGGAAAAGTCCACCGCCCAGATCTCCCAGG CCATTCACACCGTGCGGGAAGAGACAATGATGGTCATGACTGAGGAC ATGCCCCCTGGAAATCAGCTACGTGCCCTCCACCTACCTGACCGAGAT CACCCATGTGTCCCAGGCCCTGCTGGAAGTGGAACAGCTGCTGAACG CCCCGATCTGTGCGCCAAGGACTTCGAGGATCTGTTCAAGCAGGAA GAGAGCCTGAAGAATATCAAGGACTCTCTGCAGCAGTCCAGCGGCA GAATCGACATCATCCACAGCAAGAAAACAGCCGCCCTGCAGTCCGCC ACCCCCGTGGAAAGAGTGAAACTGCAGGAAGCCCTGTCCCAGCTGG

SEQ ID NO and Description	Sequence
	ACTTCCAGTGGGAGAAAGTGAACAAGATGTACAAGGACCGGCAGGG CAGATTTCGACCGCAGCGTGGAAAAGTGGCGGCGGTTCCACTACGACA TCAAGATCTTCAACCAGTGGCTGACCGAGGCCGAACAGTTCCTGAGA AAGACCCAGATCCCCGAGAACTGGGAGCACGCCAAGTACAAGTGGT ATCTGAAAGAACTGCAGGATGGCATCGGCCAGAGACAGACCGTCGT GCGGACACTGAATGCCACCGGCGAGGAAATCATCCAGCAGAGCAGC AAGACCGACGCCAGCATCCTGCAGGAAAAGCTGGGCAGCCTGAACC TGAGATGGCAGGAAGTGTGCAAGCAGCTGTCCGACCGGAAGAAGCG GCTGGAAGAACAGCTGGAACGGCTGCAGGAACTGCAGGAGGCCACC GACGAGCTGGACCTGAAACTGAGACAGGCCGAAGTGATCAAGGGCA GCTGGCAGCCAGTGGGCGACCTGCTGATCGATTCTCTGCAGGACCAC CTGGAAAAAGTGAAGGCCCTGAGAGGCGAGATCGCCCCCTGAAAG AAAACGTGTCCACGTGAACGACCTGGCCCCGGCAGCTGACAACACTG GGCATCCAGCTGAGCCCCTACAACCTGAGCACACTGGAAGATCTGAA CACCCGGTGAAGCTGCTGCAGGTGGCCGTGGAAGATAGAGTGCGG CAGCTGCACGAGGCCCCACAGAGATTTTGGCCCTGCCTCCCAGCACTT CCTGAGCACCTCAGTGCAGGGCCCTTGGGAGAGAGCCATCTCCCCTA ACAAGGTGCCCTACTACATCAACCACGAGACACAGACCACCTGTTGG GACCACCCCAAGATGACCGAGCTGTACCAGAGCCTGGCCGACCTGAA CAATGTGCGGTTTCAGCGCCTACCGGACCGCCATGAAGCTGCGGAGAC TGCAGAAAGCTCTGTGCCTGGATCTGCTGTCCCTGAGCGCCGCTGT GATGCCCTGGACCAGCACAACTCTGAAGCAGAACGACCAGCCCATGG ATATCCTGCAGATCATTAACTGCCTGACCACCATCTACGACCGGCTG GAACAGGAACACAACAACCTCGTGAACTGCCCCCTGTGCGTGGACAT GTGCCTGAATTGGCTGCTGAATGTGTACGACACCGGCCGGACAGGCC GGATCAGAGTGCTGAGCTTCAAGACCGGCATCATCAGCCTGTGCAAA GCCCATCTGGAAGATAAGTACCGCTACCTGTTCAAACAGGTGGCCAG CTCTACCGGCTTCTGCGATCAGAGAAGGCTGGGCCTGCTGCTGCATG ACTCCATCCAGATCCCCAGACAGCTGGGCGAGGTGGCCTCTTTCGGC GGCAGCAATATCGAGCCTAGCGTGCGGAGCTGCTTCCAGTTCGCCAA CAACAAGCCCGAGATCGAGGCCGCCCTGTTTCTGGACTGGATGAGAC TGGAACCCCAAGAGCATGGTGTGGCTGCCCCGTGCTGCATAGAGTGGCC GCTGCCGAGACAGCCAAGCACCAGGCCAAGTGCAATATCTGCAAAG AGTGCCCCATCATCGGCTTCCGGTACAGAAGCCTGAAGCACTTCAAC TACGATATCTGCCAGAGCTGTTTCTTCAGCGGCAGGGTGGCCAAGGG CCACAAAATGCACTACCCCATGGTGAATACTGCACCCCCACCACAA GCGGAGAGGATGTGCGGGACTTCGCCAAGGTGCTGAAAAACAAGTT CCGGACCAAGCGGTATTTTCGCCAAACACCCCCGGATGGGCTACCTGC CTGTGCAGACAGTGCTGGAAGGCGACAACATGGAAACCGACACCAT GTGATGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATG ACAGATGAAGAAGGAGCAGAATAAATGTTTTACAACCTCCTGATTCCC GCATGCGGCCAGCTTATCGATACCGTCGAAATAAAAGATCCTTATTT TCATTGGATCTGTGTGTTGGTTTTTTGTGTG

SEQ NO Description	ID and Sequence
SEQ ID NO:147 – BXA- 027743	ATGCTGTGGTGGGAGGAAGTGGAAGATTGCTACGAGCGCGAGGACG TGCAGAAGAAAACCTTCACCAAATGGGTGAACGCCCAGTTCAGCAA GTTCGGCAAGCAGCACATCGAGAACCTGTTTCAGCGACCTGCAGGACG GCAGACGGCTGCTGGACCTGCTGGAAGGCCTGACCGGCCAGAAGCT GCCCCAAGAGAAGGGCAGCACCAGAGTGCACGCCCTGAACAACGTG AACAAGGC CCTGCGGGTGCTGCAGAACAAACAGTGGACCTGGTGAACATCGGCA GCACCGACATCGTGGACGGCAAC CACAAGCTGACCCTGGGCCTGATCTGGAACATCATCCTGCACTGGCA GGTCAAAAACGTGATGAAGAACA TCATGGCCGGCCTGCAGCAGACCAACAGCGAGAAGATCCTGCTGAGC TGGGTGCGCCAGAGCACCCGGAA CTACCCCCAGGTCAACGTGATCAACTTCACCACCTCTTGGAGCGACG GCCTGGCCCTGAACGCCCTGATC CACAGCCACCGGCCCGACCTGTTTCGACTGGAACAGCGTGGTCTGCCA GCAGAGCGCCACCCAGCGGCTGG AACACGCCTTCAATATCGCCAGATACCAGCTGGGCATCGAGAAGCTG CTGGATCCCGAGGACGTGGACAC CACCTACCCCGACAAGAAATCCATCCTGATGTATATCACCAGCCTGT TCCAGGTGCTGCCCCAGCAGGTG TCCATCGAGGCCATCCAGGAAGTGGAATGCTGCCAGACCCCCCAA AGTGACCAAAGAGGAACACTTCC AGCTGCACCACCAGATGCACTACAGCCAGCAGATCACCGTGTCCCTG GCTCAGGGCTACGAGCGGACCAG CAGCCCCAAGCCCCGGTTCAAGAGCTACGCCTACACCCAGGCCGCCT ACGTGACCACCAGCGACCCACC AGAAGCCCATTCCCCAGCCAGCATCTGGAAGCCCCCGAGGACAAGA GCTTCGGCAGCAGCCTGATGGAAA GCCAAGTGAACCTGGACAGATACCAGACCGCCCTGGAAGAGGTGCT GTCCTGGCTGCTGAGCGCCGAGGA TACACTGCAGGCCCGAGGCGAGATCAGCAACGACGTGGAAGTGGTG AAAGACCAGTTCCACACCCACGAG GGCTACATGATGGACCTGACCGCCCACCAGGGCAGAGTGGGCAACA TCCTGCAGCTGGGCAGCAAGCTGA TCGGCACCGGCAAGCTGAGCGAGGACGAAGAGACAGAGGTGCAGGA ACAGATGAACCTGCTGAACAGCAG ATGGGAGTGCCTGCGGGTGGCCAGCATGGAAAAGCAGAGCAACCTG CACATCCACACCGTGCGGGAAGAG ACAATGATGGTGATGACCGAGGACATGCCCCTGGAAATCAGCTACGT GCCAGCACCTACCTGACCGAGA TCACCCACGTGTCCAGGCTCTGCTGGAAGTGGAACAGCTGCTGAAC GCCCCGACCTGTGCGCCAAGGA CTTCGAGGATCTGTTCAAGCAGGAAGAGAGCCTGAAGAATATCAAG GACTCCCTGCAGCAGTCCAGCGGC CGGATCGACATCATCCACAGCAAGAAAACAGCCGCCCTGCAGTCCGC

SEQ ID NO and Description	Sequence
	CACCCCGTGGAAAGAGTGAAGC TGCAGGAAGCCCTGAGCCAGCTGGACTTCCAGTGGGAGAAAGTGAA CAAGATGTACAAGGACCGGCAGGG CAGATTTCGACCGCAGCGTGGAAAAGTGGCGGCGGTTCCACTACGACA TCAAGATCTTCAACCAGTGGCTG ACCGAGGCCGAGCAGTTCCTGAGAAAGACCCAGATCCCCGAGAACT GGGAGCACGCCAAGTACAAGTGGT ATCTGAAAGAACTGCAGGATGGCATCGGCCAGCGGCAGACCGTGGT GCGCACACTGAATGCCACCGGCGA GGAAATCATCCAGCAGAGCAGCAAGACCGACGCCAGCATCCTGCAG GAAAAGCTGGGCTCCCTGAACCTG CGGTGGCAGGAAGTGTGCAAGCAGCTGAGCGACCGGAAGAAGAGGC TGGAAGAACAGGCCCTGGCCTGA CCACCATCGGCGCCAGCCCTACCCAGACCGTGACCCTGGTGACACAG CCCGTGGTGACAAAAGAGACAGC CATCTCCAAGCTGGAAATGCCAGCTCCCTGATGCTGGAATCTGACC AGTGGAAGCGTCTGCACCTTTCT CTGCAGGAACTTCTGGTGTGGCTACAGCTGAAAGATGATGAATTAAG CCGGCAGGCACCTATTGGAGGCG ACTTTCAGCAGTTCAGAAGCAGAACGATGTACATAGGGCCTTCAAG AGGGAATTGAAAATAAAGAACC TGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGC AGCCTTTGGAAGGACTAGAGAAA CTCTACCAGGAGCCCAGAGAGCTGCCTCCTGAGGAGAGAGCCCAGA ATGTCACTCGGCTTCTACGAAAGC AGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTGAACCTGCACTCC GCTGACTGGCAGAGAAAAATAGA TGAGACCCTGGAAAGGCTGCAGGAACTGCAGGAGGCCACCGACGAG CTGGACCTGAAGCTGCGGCAGGCC GAAGTGATCAAGGGCAGCTGGCAGCCCGTGGGCGACCTGCTGATCG ACAGCCTGCAGGACCACCTGGAAA AAGTGAAGGCCCTGAGAGGCGAGATCGCCCCCTGAAAGAAAACGT GTCCACGTGAACGACCTGGCCCG GCAGCTGACCACCCTGGGCATCCAGCTGAGCCCCTACAACCTGAGCA CCCTGGAAGATCTGAACACCCGG TGGAAGCTGCTGCAGGTGGCCGTGGAAGATAGAGTGCGGCAGCTGC ACGAGGCCACAGAGACTTTGGCC CTGCCAGCCAGCACTTCCTGAGCACCTCTGTGCAGGGACCCTGGGAG AGAGCCATCAGCCCCAACAAGGT GCCCTACTACATCAACCACGAGACACAGACCACCTGTTGGGACCACC CCAAGATGACCGAGCTGTACCAG AGCCTGGCCGACCTGAACAATGTGCGGTTTCAGCGCCTACCGGACCGC CATGAAGCTGAGGCGGCTGCAGA AAGCTCTGTGCCTGGATCTGCTGAGCCTGAGCGCCGCCTGCGACGCC CTGGACCAGCACACCTGAAGCA

SEQ ID NO and Description	Sequence
	GAACGACCAGCCCATGGATATCCTGCAGATCATCAACTGCCTGACCA CAATCTACGACAGGCTGGAACAG GAACACAACAATCTGGTCAACGTGCCCTGTGCGTGGACATGTGCCT GAATTGGCTGCTGAATGTGTACG ACACCGGCCGGACCGGCAGAATCCGGGTGCTGAGCTTCAAGACCGG CATCATCAGCCTGTGCAAGGCCCA CCTGGAAGATAAGTACCGCTACCTGTTCAAACAGGTGGCCAGCTCCA CCGGCTTTTGCGACCAGCGGAGA CTGGGCCTGCTGCTGCACGACAGCATCCAGATCCCCAGACAGCTGGG CGAGGTGGCCTCCTTCGGCGGCA GCAACATTGAGCCCAGCGTGCGGAGCTGCTTCCAGTTCGCCAACAAC AAGCCCGAGATCGAGGCCGCCCT GTTCTGGACTGGATGAGACTGGAACCCAGAGCATGGTGTGGCTGC CCGTGCTGCATCGGGTGGCCGCT GCCGAGACAGCCAAGCACCAGGCCAAGTGCAACATCTGCAAAGAGT GCCCCATCATCGGCTTCCGGTACA GAAGCCTGAAGCACTTCAACTACGATATCTGCCAGAGCTGCTTCTTC AGCGGCAGAGTGGCCAAGGGCCA CAAAATGCACTACCCCATGGTGAATACTGCACCCCCACCACCAGCG GCGAGGATGTGCGGGACTTCGCC AAGGTGCTGAAAAACAAGTTCCGGACCAAGCGGTACTTTGCCAAGCA CCCCCGGATGGGCTACCTGCCCCG TGCAGACAGTGCTGGAAGGCGACAACATGGAAACCCCTGCCCAGAT CCTGATCAGCCTGGAAAGCGAGGA ACGGGGCGAGCTGGAACGGATCCTGGCCGATCTGGAAGAGGAAAAC CGGAACCTGCAGGCCGAGTACGAC CGGCTGAAGCAGCAGCACGAGCACAAGGGCCTGAGCCCCCTGCCTA GCCCCCTGACACCATGTGATGAT GATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGA AGAAGGAGCAGAATAAATGTTTT ACAACTCCTGATTCCCGCATGCGGCCAGCTTATCGATAACCGTCGAAA TAAAAGATCCTTATTTTCATTGG ATCTGTGTGTTGGTTTTTTGTGTG
SEQ ID NO:148 – BXA- 212371	ATGCTGTGGTGGGAGGAAGTGGAAGATTGCTACGAGCGCGAGGACG TGCAGAAGAAAACCTTCACCAAATGGGTGAACGCCAGTTCAGCAA GTTTCGGCAAGCAGCACATCGAGAACCTGTTTCAGCGACCTGCAGGACG GCAGACGGCTGCTGGACCTGCTGGAAGGCCTGACCGGCCAGAAGCT GCCCCAAGAGAAGGGCAGCACCAGAGTGACGCCCCTGAACAACGTG AACAAGGCCCTGCGGGTGCTGCAGAACAAACAGTGGACCTGGTGA ACATCGGCAGCACCGACATCGTGGACGGCAACCACAAGCTGACCCTG GGCCTGATCTGGAACATCATCCTGCACTGGCAGGTCAAAAACGTGAT GAAGAACATCATGGCCGGCCTGCAGCAGACCAACAGCGAGAAGATC CTGCTGAGCTGGGTGCGCCAGAGCACCCGGAACCTACCCCCAGGTCAA CGTGATCAACTTCACCACCTCTTGAGCGACGGCCTGGCCCTGAACG

SEQ ID NO and Description	Sequence
	CCCTGATCCACAGCCACCGGCCCGACCTGTTCTGACTGGAACAGCGTG GTCTGCCAGCAGAGCGCCACCCAGCGGCTGGAACACGCCTTCAATAT CGCCAGATAACCAGCTGGGCATCGAGAAGCTGCTGGATCCCGAGGAC GTGGACACCACCTACCCCGACAAGAAATCCATCCTGATGTATATCAC CAGCCTGTTCCAGGTGCTGCCCCAGCAGGTGTCCATCGAGGCCATCC AGGAAGTGGAAATGCTGCCCAGACCCCCCAAAGTGACCAAAGAGGA AACTTCCAGCTGCACCACCAGATGCACTACAGCCAGCAGATCACCG TGTCCCTGGCTCAGGGCTACGAGCGGACCAGCAGCCCCAAGCCCCGG TTCAAGAGCTACGCCTACACCCAGGCCCGCTACGTGACCACCAGCGA CCCCACCAGAAGCCCATTCCCCAGCCAGCATCTGGAAGCCCCCGAGG ACAAGAGCTTCGGCAGCAGCCTGATGGAAAGCGAAGTGAACCTGGA CAGATACCAGACCGCCCTGGAAGAGGTGCTGTCCTGGCTGCTGAGCG CCGAGGATACACTGCAGGCCCAGGGCGAGATCAGCAACGACGTGGA AGTGGTGAAAGACCAGTTCCACACCCACGAGGGCTACATGATGGACC TGACCGCCCACCAGGGCAGAGTGGGCAACATCCTGCAGCTGGGCAG CAAGCTGATCGGCACCGGCAAGCTGAGCGAGGACGAAGAGACAGAG GTGCAGGAACAGATGAACCTGCTGAACAGCAGATGGGAGTGCCTGC GGGTGGCCAGCATGGAAAAGCAGAGCAACCTGCACATCCACACCGT GCGGGAAGAGACAATGATGGTGATGACCGAGGACATGCCCCCTGGAA ATCAGCTACGTGCCCAGCACCTACCTGACCGAGATCACCCACGTGTC CCAGGCTCTGCTGGAAGTGGAACAGCTGCTGAACGCCCCCGACCTGT GCGCCAAGGACTTCGAGGATCTGTTCAAGCAGGAAGAGAGCCTGAA GAATATCAAGGACTCCCTGCAGCAGTCCAGCGGCCGGATCGACATCA TCCACAGCAAGAAAACAGCCGCCCTGCAGTCCGCCACCCCCGTGGAA AGAGTGAAGCTGCAGGAAGCCCTGAGCCAGCTGGACTTCCAGTGGG AGAAAGTGAACAAGATGTACAAGGACCGGCAGGGCAGATTTCGACCG CAGCGTGGAAGAGTGGCGGCGGTTCCACTACGACATCAAGATCTTCA ACAGTGGCTGACCGAGGCCGAGCAGTTCTTGAGAAAGACCCAGAT CCCCGAGAACTGGGAGCACGCCAAGTACAAGTGGTATCTGAAAGAG CTGCAGGACGGCATCGGCCAGCGGCAGACAGTGGTCCGCACCCTGA ATGCCACCGGCGAGGAAATCATCCAGCAGAGCAGCAAGACCGACGC CAGCATCCTGCAGGAAAAGCTGGGCAGCCTGAACCTGCGGTGGCAG GAAGTGTGCAAGCAGCTGAGCGACCGGAAGAAGCGGCTGGAAGAAC AGGCCCCCTGGCCTGACCACAATCGGCGCCAGCCCTACCCAGACCGTG ACCCTGGTGACACAGCCCGTGGTGACAAAAGAGACAGCCATCAGCA AGCTGGAAATGCCCAGCAGCCTGATGCTGGAAAGCGACCAGTGGAA GCGGCTGCACCTGAGCCTGCAGGAACTGCTGGTCTGGCTGCAGCTGA AGGACGACGAGCTGAGCAGACAGGCCCCCATCGGCGGCGATTTCGCC GCCGTGCAGAAACAGAACGACGTGCACCGGGCCTTCAAGCGCGAGC TGAAAACAAAAGAACCCGTGATCATGAGCACCTGGAAACCGTGCG GATCTTCCTGACCGAGCAGCCCCTGGAAGGCCTGGAAGAGCTGTACC AGGAACCCAGAGAGCTGCCCCCGAGGAACGGGCCAGAACGTGAC CAGACTGCTGCGGAAGCAGGCCGAAGAGGTCAACACCGAGTGGGAG AAGCTGAACCTGCACAGCGCCGACTGGCAGCGGAAGATCGACGAGA CACTGGAACGGCTGCAGGAACTGCAGGAGGCCACCGACGAGCTGGA

SEQ ID NO and Description	Sequence
	CCTGAAGCTGAGACAGGCCGAAGTGATCAAGGGCAGCTGGCAGCCC GTGGGCGACCTGCTGATCGACTCCCTGCAGGACCACCTGGAAAAAGT GAAGGCCCTGCGGGGCGAGATCGCCCCCTGAAAGAAAACGTGTCC CACGTGAACGACCTGGCCCCGGCAGCTGACCACCCTGGGCATCCAGCT GAGCCCCCTACAACCTGTCCACCCTGGAAGATCTGAACACCCGGTGGA AGCTGCTGCAGGTGGCCGTGGAAGATAGAGTGCGGCAGCTGCACGA GGCCCACAGAGACTTTGGCCCTGCCAGCCAGCACTTCCTGAGCACCT CTGTGCAGGGACCCTGGGAGAGAGCCATCAGCCCCAACAAAGGTGCC CTACTACATCAACCACGAGACACAGACCACCTGTTGGGACCACCCCA AGATGACCGAGCTGTACCAGAGCCTGGCCGACCTGAACAATGTGCGG TTCAGCGCCTACCGGACCGCCATGAAGCTGAGGCGGCTGCAGAAAGC TCTGTGCCTGGATCTGCTGAGCCTGAGCGCCGCCTGCGACGCCCTGG ACCAGCACAACCTGAAGCAGAACGACCAGCCCATGGATATCCTGCA GATCATCAACTGCCTGACCACAATCTACGACAGGCTGGAACAGGAAC ACAACAATCTGGTCAACGTGCCCCCTGTGCGTGGACATGTGCCTGAAT TGGCTGCTGAATGTGTACGACACCGGCCGGACCGGCAGAATCCGGGT GCTGAGCTTCAAGACCGGCATCATCAGCCTGTGCAAGGCCACCTGG AAGATAAGTACCGCTACCTGTTCAAACAGGTGGCCAGCTCCACCGGC TTTTGCGACCAGCGGAGACTGGGCCTGCTGCTGCACGACAGCATCCA GATCCCCAGACAGCTGGGCGAGGTGGCCTCCTTCGGCGGCAGCAACA TTGAGCCCAGCGTGCGGAGCTGCTTCCAGTTCGCCAACAAACAGCCC GAGATCGAGGCCGCCCTGTTCTGGACTGGATGAGACTGGAACCCCA GAGCATGGTGTGGCTGCCCCGTGCTGCATCGGGTGGCCGCTGCCGAGA CAGCCAAGCACCAGGCCAAGTGCAACATCTGCAAAGAGTGCCCCATC ATCGGCTTCCGGTACAGAAGCCTGAAGCACTTCAACTACGATATCTG CCAGAGCTGCTTCTTCAGCGGCAGAGTGGCCAAGGGCCACAAAATGC ACTACCCCATGGTGGAAATACTGCACCCCCACCACCAGCGGCGAGGAT GTGCGGGACTTCGCCAAGGTGCTGAAAAACAAGTTCCGGACCAAGC GGTACTTTGCCAAGCACCCCCGGATGGGCTACCTGCCCCGTGCAGACA GTGCTGGAAGGCGACAACATGGAAACCGACACCATGTGATGATGATT TGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAA GGAGCAGAATAAATGTTTTACAACCTCCTGATTCCCGCATGCGGCCAG CTTATCGATACCGTCGAAATAAAAGATCCTTATTTTCATTGGATCTGT GTGTTGGTTTTTTGTGTG
SEQ ID NO:149 – BXA- 212372	ATGCTGTGGTGGGAGGAAGTGGAAGATTGCTACGAGCGCGAGGACG TGCAGAAGAAAACCTTCACCAAATGGGTGAACGCCCAGTTCAGCAA GTTTCGGCAAGCAGCACATCGAGAACCTGTTTCAGCGACCTGCAGGACG GCAGACGGCTGCTGGACCTGCTGGAAGGCCTGACCGGCCAGAAGCT GCCCAAAGAGAAGGGCAGCACCAGAGTGACGCCCCCTGAACAACGTG AACAAGGCCCTGCGGGTGCTGCAGAACAAACAGTGGACCTGGTGA ACATCGGCAGCACCGACATCGTGGACGGCAACCACAAGCTGACCCTG GGCCTGATCTGGAACATCATCCTGCACTGGCAGGTCAAAAACGTGAT GAAGAACATCATGGCCGGCCTGCAGCAGACCAACAGCGAGAAGATC CTGCTGAGCTGGGTGCGCCAGAGCACCCGGAACCTACCCCCAGGTCAA

SEQ ID NO and Description	Sequence
	CGTGATCAACTTCACCACCTCTTGGAGCGACGGCCTGGCCCTGAACG CCCTGATCCACAGCCACCGGCCCGACCTGTTCGACTGGAACAGCGTG GTCTGCCAGCAGAGCGCCACCCAGCGGCTGGAACACGCCTTCAATAT CGCCAGATACCAGCTGGGCATCGAGAAGCTGCTGGATCCCGAGGAC GTGGACACCACCTACCCCGACAAGAAATCCATCCTGATGTATATCAC CAGCCTGTTCCAGGTGCTGCCCCAGCAGGTGTCCATCGAGGCCATCC AGGAAGTGGAATGCTGCCCAGACCCCCAAAGTGACCAAAGAGGA ACACTTCCAGCTGCACCACCAGATGCACTACAGCCAGCAGATCACCG TGTCCCTGGCTCAGGGCTACGAGCGGACCAGCAGCCCCAAGCCCCGG TTCAAGAGCTACGCCTACACCCAGGCCGCCTACGTGACCACCAGCGA CCCCACCAGAAGCCCATTCCCCAGCCAGCATCTGGAAGCCCCCGAGG ACAAGAGCTTCGGCAGCAGCCTGATGGAAAGCGAAGTGAACCTGGA CAGATACCAGACCGCCCTGGAAGAGGTGCTGTCCTGGCTGCTGAGCG CCGAGGATACACTGCAGGCCCAGGGCGAGATCAGCAACGACGTGGA AGTGGTGAAAGACCAGTTCCACACCCACGAGGGCTACATGATGGACC TGACCGCCCACCAGGGCAGAGTGGGCAACATCCTGCAGCTGGGCAG CAAGCTGATCGGCACCGGCAAGCTGAGCGAGGACGAAGAGACAGAG GTGCAGGAACAGATGAACCTGCTGAACAGCAGATGGGAGTGCCTGC GGGTGGCCAGCATTGAAAAGCAGAGCAACCTGCACATCCACACCGT GCGGGAAGAGACAATGATGGTGATGACCGAGGACATGCCCCTGGAA ATCAGCTACGTGCCCAGCACCTACCTGACCGAGATCACCCACGTGTC CCAGGCTCTGCTGGAAGTGGAACAGCTGCTGAACGCCCCCGACCTGT GCGCCAAGGACTTCGAGGATCTGTTCAAGCAGGAAGAGAGCCTGAA GAATATCAAGGACTCCCTGCAGCAGTCCAGCGGCCGGATCGACATCA TCCACAGCAAGAAAACAGCCGCCCTGCAGTCCGCCACCCCCGTGGAA AGAGTGAAGCTGCAGGAAGCCCTGAGCCAGCTGGACTTCCAGTGGG AGAAAGTGAACAAGATGTACAAGGACCGGCAGGGCAGATTTCGACCG CAGCGTGGAAGAGTGGCGGCGGTTCCACTACGACATCAAGATCTTCA ACAGTGGCTGACCGAGGCCGAGCAGTTCCTGAGAAAGACCCAGAT CCCCGAGAACTGGGAGCACGCCAAGTACAAGTGGTATCTGAAAGAG CTGCAGGACGGCATCGGCCAGCGGCAGACAGTGGTCCGCACCCTGA ATGCCACCGGCGAGGAAATCATCCAGCAGAGCAGCAAGACCGACGC CAGCATCCTGCAGGAAAAGCTGGGCAGCCTGAACCTGCGGTGGCAG GAAGTGTGCAAGCAGCTGAGCGACCGGAAGAAGCGGCTGGAAGAAC AGGCCCCCTGGCCTGACCACAATCGGCGCCAGCCCTACCCAGACCGTG ACCCTGGTGACACAGCCCCGTGGTGACAAAAGAGACAGCCATCAGCA AGCTGGAAATGCCCAGCAGCCTGATGCTGGAAAGCGACCAGTGGAA GCGGCTGCACCTGAGCCTGCAGGAACTGCTGGTCTGGCTGCAGCTGA AGGACGACGAGCTGAGCAGACAGGCCCCCATCGGCGGCGATTTCCTCC GCCGTGCAGAAACAGAACGACGTGCACCGGGCCTTCAAGCGCGAGC TGAAAACAAAAGAACCCGTGATCATGAGCACCTGGAAACCGTGCG GATCTTCCTGACCGAGCAGCCCCCTGGAAGGCCTGGAAAAGCTGTACC AGGAACCCAGAGAGCTGCCCCCGAGGAACGGGCCCAGAACGTGAC CAGACTGCTGCGGAAGCAGGCCGAAGAGGTCAACACCGAGTGGGAG AAGCTGAACCTGCACAGCGCCGACTGGCAGCGGAAGATCGACGAGA

SEQ ID NO and Description	Sequence
	CACTGGAACGGCTGCAGGAAGTGCAGGAGGCCACCGACGAGCTGGA CCTGAAGCTGAGACAGGCCGAAGTGATCAAGGGCAGCTGGCAGCCC GTGGGCGACCTGCTGATCGACTCCCTGCAGGACCACCTGGAAAAAGT GAAGGCCCTGCGGGGCGAGATCGCCCCCTGAAAGAAAACGTGTCC CACGTGAACGACCTGGCCCGGCAGCTGACCACCCTGGGCATCCAGCT GAGCCCCTACAACCTGTCCACCCTGGAAGATCTGAACACCCGGTGGGA AGCTGCTGCAGGTGGCCGTGGAAGATAGAGTGCGGCAGCTGCACGA GGCCCACAGAGACTTTGGCCCTGCCAGCCAGCACTTCCTGAGCACCT CTGTGCAGGGACCCTGGGAGAGAGCCATCAGCCCCAACAAAGGTGCC CTACTACATCAACCACGAGACACAGACCACCTGTTGGGACCACCCCA AGATGACCGAGCTGTACCAGAGCCTGGCCGACCTGAACAATGTGCGG TTCAGCGCCTACCGGACCGCCATGAAGCTGAGGCGGCTGCAGAAAGC TCTGTGCCTGGATCTGCTGAGCCTGAGCGCCGCCTGCGACGCCCTGG ACCAGCACAACCTGAAGCAGAACGACCAGCCCATGGATATCCTGCA GATCATCAACTGCCTGACCACAATCTACGACAGGCTGGAACAGGAAC ACAACAATCTGGTCAACGTGCCCCCTGTGCGTGGACATGTGCCTGAAT TGGCTGCTGAATGTGTACGACACCGGCCGGACCGGCAGAATCCGGGT GCTGAGCTTCAAGACCGGCATCATCAGCCTGTGCAAGGCCACCTGG AAGATAAGTACCGCTACCTGTTCAAACAGGTGGCCAGCTCCACCGGC TTTTGCGACCAGCGGAGACTGGGCCTGCTGCTGCACGACAGCATCCA GATCCCCAGACAGCTGGGCGAGGTGGCCTCCTTCGGCGGCAGCAACA TTGAGCCCAGCGTGCGGAGCTGCTTCCAGTTCGCCAACAAACAAGCCC GAGATCGAGGCCGCCCTGTTCCCTGGACTGGATGAGACTGGAACCCCA GAGCATGGTGTGGCTGCCCCGTGCTGCATCGGGTGGCCGCTGCCGAGA CAGCCAAGCACCAGGCCAAGTGCAACATCTGCAAAGAGTGCCCCATC ATCGGCTTCCGGTACAGAAGCCTGAAGCACTTCAACTACGATATCTG CCAGAGCTGCTTCTTCAGCGGCAGAGTGGCCAAGGGCCACAAAATGC ACTACCCCATGGTGGGAATACTGCACCCCCACCACCAGCGGCGAGGAT GTGCGGGACTTCGCCAAGGTGCTGAAAAACAAGTTCCGGACCAAGC GGTACTTTGCCAAGCACCCCCGGATGGGCTACCTGCCCGTGCAGACA GTGCTGGAAGGCGACAACATGGAAACCGACACCATGTGATGATGAG AAGTCTTTTCCACATGGCAGATGAAATAAAAGATCCTTATTTTCATTG GATCTGTGTGTTGGTTTTTTGTGTG

SEQ ID NO:148 encodes construct BXA-212371, which is the same as construct BXA-027743, except that the C-terminus is deleted.

- 5 SEQ ID NO:149 encodes construct BXA-212372, which is the same as BXA-212371, but with a shorter 3'UTR.

In some embodiments, the miniaturized dystrophin polypeptide comprises or is the construct shown in Table 9.

Table 9. Amino acid sequence of miniaturized dystrophin polypeptide BXA-213788.

SEQ NO:	ID	Sequence
152		MLWWEVEEDCYEREDVQKKFTKWVNAQFSKFGKQHIEENLFSDLQDG RRLDLLLEGLTGQKLPEKEGSTRVHALNNVNKALRVLQNNNVDLVNIG STDIVDGNHKLTLGLIWNILHWQVKNVMKNIMAGLQQTNSEKILLSWV RQSTRNYPQVNVINFTTSWSDGLALNALIHSHRPDLFDWNSVVCQQSAT QRLEHAFNIARYQLGIEKLLDPEDVDTTYPDKKSILMYITSLFQVLPQQV SIEAIQEVEMLPKPPKVTKEEHFQLHHQMHYSQQITVSLAQGYERTSSPK PRFKSYAYTQAAYVTTSDPTRSPFPSQHLEAPEDKSFGSSLMSEVNDR YQTAL EEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTA HQGRVGNILQLGSKLIGTGKLSEDEETEVEQEQMNNLSNRWECLRVASM EKQSNLHIHTVREETMMVMTEDEMPLEISYVPSTYLTEITHVSQALLEVEQ LLNAPDLCAKDFEDLFKQEEESLKNIKDSLQSSGRIDIIHSSKKAALQSAT PVERVKLQEALSQDFQWEKVNKMYKDRQGRFDRSVEKWRRFHYDIKI FNQWLTEAEQFLRKTQIPENWEHAKYKWLKELQDGIGQRQTIVRTLN ATGEEIIQSSKTDASILQEKLGSNLNRWQEVCKQLSDRKKRLEEQKNDL APGLTTIGASPTQTVTLVTQPVVTKETAISKLEMPSSLMLESDQWKRLHL SLQELLVWLQLKDDELSRQAPIGGDFPAVQKQNDVHRAFKRELKTKEP VIMSTLETVRIFLTEQPLEGLEKLYQEPRELPEERAQNVTLLRKQAE VNTWEKLNLSADWQRKIDETLERLQELQEATDELDLKLRAEVIKGS WQPVGDLLIDSLQDHLEKVKALRGEIAPLKENVSHVNDLARQLTTLGIQ LSPYNLSTLEDLNTRWKLQVAVEDRVRQLHEAHRDFGPASQHFLSTSV QGPWERAISPKNVPYYINHETQTTCDHPKMTELYQSLADLNNVRFSA YRTAMKLRRRLQKALCLDLLSLSAACDALDQHNKQNDQPMQDILQIINCL TTIYDRLEQEHNNLVNPLCVDMLNWLNNVYDTGRTGRIRVLSFKTGI ISLCKAHLEDKYRYLQKQVASSTGFCQRRRLGLLLHDSIQIPRQLGEVAS FGGSNIEPSVRSCFQFANNKPEIEAALFLDWMRLEPQSMVWLPVLHRVA AAETAKHQAKCNICKECPIIGFRYRSLKHFNYDICQSCFFSGRVAKGHK MHYPMVEYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGYLPVQ TVLEGDNMET

In some embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 152, wherein the amino acid sequence when expressed has at least one dystrophin activity.

In some embodiments, the miniaturized dystrophin polypeptide comprises or is the construct shown in Table 10.

Table 10. Amino acid sequence of miniaturized dystrophin polypeptide BXA-213780J11V3.

SEQ NO:	ID	Sequence
------------	----	----------

153	MLWWEEVEDCYEREDVQKKTFTKWVNAQFSKFGKQHIEENLFSDLQDG RRLDLLEGLTGQKLPKEKGSTRVHALNNVNKALRVLQNNNVDLVNIG STDIVDGNHKLTLGLIWNILHWQVKNVMKNIMAGLQQTNSEKILLSWV RQSTRNYPQVNVINFTTSDGLALNALIHSRDLFDWNSVVCQQSAT QRLEHAFNIARYQLGIEKLLDPEDVDTTYDPDKKSILMYITSLFQVLPQQV SIEAIQEVEMLPKPPKVTKEEHFQLHHQMHSYQQITVSLAQGYERTSSPK PRFKSYAYTQAAYVTTSDPTRSPFPSQHLEAPEDKSFSSLMSEVNLDL YQTALKEVLSWLLSAEDTLQAQGEISNDVEVVKDQFHTHEGYMMDLTA HQGRVGNILQLGSKLIGTGKLSEDEETEVEQEMNLLNSRWECLRVASM EKQSNLHRVLLKWQRLTEEQCLFSAWLSEKEDAVNKIHTTGFKDQNE LSSLQKLAVLKADLEKKKQSMGKLYSLKQDLLSTLKNKSVTQKTEAWL DNFARCWDNLVQKLEKSTAQISQAVTTTQPSLTQTTVMETVTTVTREQ ILVKHAQEELPPPPQKKRQITVDSEIHTVREETMMVMTEDEMPLEISYVP STYLTEITHVSQALLEVEQLLNAPDLCAKDFEDLFKQEESESLKNKDSLQ SSGRIDIIHKKTAALQSATPVERVKLQEALSQDLDFQWEKVNKMYKDRQ GRFDRSVEKWRRFHDIKIFNQLWTEAEQFLRKTQIPENWEHAKYKWKY LKEQDQGGQRQTVVRTLNLATGEEIIQSSKTDASILQEKLGSLNLRWQE VCKQLSDRKKRLEEQLERLQELQEATDELDLKLRAEVKGSWQPVGD LLIDSLQDHLEKVKALRGEIAPLKENVSHVNDLARQLTTLGIQLSPYNLS TLEDLNRWKLQVAVEDRVRQLHEAHRDFGPASQHFLLSTSVQGPWER AISPKNVPYYINHETQTTCDWHPKMTELYQSLADLNNVRFSAYRTAMK LRLQKALCLDLSLSAACDALDQHNKQNDQPMQDILQIINCLTTIYDRL EQEHNNLVNVPLCVDMLNWLNVYDTGRTGRIRVLSFKTGIIISLCKAH LEDKYRYLFKQVASSTGFCDQRRLLGLLHDSIQIPRLGEVASFGGSNIE PSVRSCFQFANNKPEIEAALFLDWMRLPQSMVWLPVLRVAAAETAK HQAKCNICKECPIIGFRYRSLKHFNIDICQSCFFSGRVAKGHKMHYPMV EYCTPTTSGEDVRDFAKVLKNKFRTKRYFAKHPRMGYLPVQTVLEGDN MET
-----	--

SEQ ID NO:153 encodes construct BXA-213780J11V3, which is the same as construct BXA-196477, except that amino acids A and Q of the internal SEAG linker (SEQ ID NO. 74), as well as the last three amino acids at the C-terminus, are deleted in BXA-213780J11V3.

In some embodiments, the miniaturized dystrophin polypeptide comprises an amino acid sequence at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 153, wherein the amino acid sequence when expressed has at least one dystrophin activity.

In some embodiments, the miniaturized dystrophin BXA-213780J11V3 can be encoded by the nucleotide sequence shown in Table 11.

Table 11. Nucleotide sequence of miniaturized dystrophin polypeptide BXA-213780J11V3.

SEQ NO:	ID	Sequence
154		atgctgtggtgggaggaagtggaagattgctacgagcgcgaggacgtgcagaagaaaaccttcaccaaattgggt gaacgcccagttcagcaagttcggcaagcagcacatcgagaacctgttcagcgacctgcaggacggcagacgg ctgctggacctgctggaaggcctgaccggccagaagctgccccaaagagaagggcagcaccagagtgcacgcc ctgaacaacgtgaacaaggccctgcggtgctgcagaacaacaacgtggacctggtgaacatcggcagcaccg acatcgtggacggcaaccacaagctgacctgggctgatctggaacatcactcgtcactggcaggtcaaaaacg tgatgaagaacatcatggccggcctgcagcagaccaacagcgagaagatcctgctgagctgggtgcgccagag caccgggaactacccccaggtcaacgtgatcaacttcaccaccttggagcgacggcctggccctgaacgcct gatccacagccaccggcccagctgttcgactggaacagcgtggtctgccagcagagcgccaccacagcggtg gaacacgccttcaatacgcagataccagctgggcatcgagaagctgctggatcccaggacgtggacaccac ctaccccgacaagaaatccatcctgatgtatatcaccagcctgttcaggtgctgccccagcaggtgtccatcagg ccatccaggaagtggaaatgctgccagacccccaaagtgaccaaagaggaaacttccagctgcaccaccag atgcactacagccagcagatcacctgtccctggcccagggctacgagagaaccagcagcccaagccccggtt caagagctacgcctataccaggccgcctacgtgaccaccagcgacctaccagaagcccattccccagccagc atctggaagccccgaggacaagagcttcggcagcagcctgatggaaagcgaagtgaacctggacagatacca gaccgcccctggaagaggtgctgagctggctgctgagcgccgaggatacactgcaggctcagggcgagatcagc aacgacgtggaagtcgtgaaggaccagttccacaccacgagggctacatgatggacctgacagcccaccagg gcagagtgggcaacatcctgcagctgggctccaagctgatcggcaccggcaagctgagcgaggacgaagaga cagaggtgcaggaacagatgaacctgctgaacagcagatgggagtgctgcgggtggccagcatgaaaagca gagcaacctgcatagggtcctgctgaagtggcagcggctgaccgaggaacagtgctgtttagcgcttggtgtc cgagaaagaggacgccgtgaacaagatccacaccaccggctcaaggaccagaacgagatgctgagcagcctg cagaaactggcgtgctgaaggccgacctggaaaagaaaaagcagtcctatgggcaagctgtactccctgaagca ggacctgctgtccacctgaagaacaagagcgtgaccagaaaaaccgagggcctggctggacaacttcgccgggt gctgggacaacctggtgcagaagctggaaaagtccaccgcccagatctcccaggccgtgaccacaacacagcc cagcctgaccagaccaccgtgatggaaacctgacaacagtgaccaccgggaacagatcctcgtgaagcac gcccaggaagaactgccccctccacccccagaaagaaacggcagatcacagtggacagtgaatccacaccg tgcgggaagagacaatgatggtcatgacagaggacatgccccctggaaatcagctacgtgccagcacctacctg accgagatcacccatgtgtcccaggccctgctggaagtggaaacagctgctgaacgccccgacctgtgcgcaa ggatttcgaggacctgttcaagcaggaagagagcctgaagaatatcaaggatagcctgcagcagagcagcggcc ggatcgacatcatccacagcaagaaaacagccgccctgcagagcgccacccccgtggaaagagtgaactgca ggaagccctgtcccagctggacttccagtgaggagaaagtgaacaaaatgtacaaggaccggcagggcgagattcg accgacgctggaaaagtggcggcggttcactacgacatcaagatcttaaccagtggtgacagaggccgag cagttcctgagaaagaccagatccccgagaactgggagcacgccaagtacaagtgtatctgaaagagctgca ggacggcatcggccagagacagacagctgctgcggacctgaatgccaccggcgaggaaatcatccagcagtc agcaagaccgacgccagcattctgcaggaaaagctgggcagcctgaacctgcggtggcaggaagtgtgcaagc agctgtccgaccgggaagaagcggctggaagaacagctggaacggctgcaggaaactgcaggagggccaccgac gagctggacctgaaactgagacagggcgaagtgatcaagggcagctggcagccagtgggcgacctgctgatcg actccctgcaggaccatctggaaaaagtgaaggccctgagaggcgagatgccccctgaaagaaaacgtgtcc cacgtgaacgacctggcccggcagctgacaacactgggcattcagctgagccctacaacctgagcacactgga agatctgaacaccgggtggaagctgctgcaggtggcctgggaagatagagtgcggcagctgcacgaggccac agagattttggccctgcctcccagcacttctgagcaccagtgtgcaggggccctgggagagagccatctccccta acaagtgccctactacatcaaccacgagacacagaccacctgttgggaccaccccaagatgaccgagctgtacc agagcctggctgacctgaacaacgtgcgggtcagcgctaccggaccgccatgaagctgcggagactgcagaa agctctgtgcctggatctgctgtccctgtccgccgctgtgatgccctggaccagcacaatctgaagcagaacgac cagcccatggatctcctgcagatcatcaactgcctgaccacctctacgaccggctggaacaggaacacaacaat ctcgtgaacgtgcccctgtgctggacatgtgctgaattggctgctgaatgtgtacgacaccggccggacagggc cggatcagagtgtgtccttcaagaccggcatcatcagcctgtgcaagcccacctggaagataagtaccgggtatc

	tgttcaaacaggtggcctctagcaccggcttttgcgaccagagaaggctgggcctgctgctgcacgacagcatcc agatccctagacagctgggcgaagtggccagctttggcggcagcaacatcgagcctagcgtgcggagctgcttc cagttcgccaacaacaagcccagatcgaggccgccctgttcctggactggatgagactggaacccagagcat gggtggtgctgcccgtgctgcatagagtggcgcgtgccgagacagccaagcaccaggccaagtgaacatctgca aagagtgccccatcatcggcttcgggtacagaagcctgaagcacttcaactacgatatctgccagagctgtttctca gcggcaggggtggccaagggccacaaaatgcactaccccatgggtggaatactgcacccccaccacaagcggcga ggatgtgcgggatttcgccaaggtgctgaaaaacaagttccggaccaagcgggtacttcgccaacacccccggat gggctacctgcctgtgcagacagtgctggaaggcgacaacatggaaacctgaGAAGTCTTTTCCAC ATGGCAGATGAAATAAAAGATCCTTATTTTCATTGGATCTGTGTGTTG GTTTTTGTGTG
--	---

In some embodiments, the miniaturized dystrophin polypeptide has a junction between hinge 2 (H2) domain and R16 spectrin repeat 16 (R16) domain that comprises an amino acid sequence listed in Table 12.

5 Table 12. Amino Acid Sequences of Junctions

SEQ NO	ID	Sequence
SEQ NO:155	ID	LPPPPPQKKRQITVDIHTVREETMMVMTEDMPLEI
SEQ NO:156	ID	LPPPPPQKKRQITVDSIHTVREETMMVMTEDMPLEI
SEQ NO:157	ID	LPPPPPQKKRQITVDSEIHTVREETMMVMTEDMPLEI
SEQ NO:158	ID	LPPPPPQKKRQITVDQIHTVREETMMVMTEDMPLEI
SEQ NO:159	ID	LPPPPPQKKRQITVDAQIHTVREETMMVMTEDMPLEI
SEQ NO:160	ID	LPPPPPQKKRQITVIHTVREETMMVMTEDMPLEI
SEQ NO:161	ID	LPPPPPQKKRQITVDHTVREETMMVMTEDMPLEI
SEQ NO:162	ID	LPPPPPQKKRQITVHTVREETMMVMTEDMPLEI
SEQ NO:163	ID	LPPPPPQKKRQITVDSQIHTVREETMMVMTEDMPLEI

SEQ ID NO:164	LPPPPPQKKRQITVDSHTVREETMMVMTEDMPLEI
SEQ ID NO:165	LPPPPPQKKRQITVQIHTVREETMMVMTEDMPLEI
SEQ ID NO:166	QKKRQITVDSEAQIHTVREETMMV

In other embodiments, a nucleotide sequence encoding the miniaturized dystrophin polypeptide comprises a nucleic acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NOs: 133-149, wherein the miniaturized dystrophin polypeptide when expressed from the nucleotide sequence has at least one dystrophin activity. In other embodiments, a nucleotide sequence encoding the miniaturized dystrophin polypeptide comprises a nucleic acid sequence at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO:133 (BXA-196477), SEQ ID NO:134 (BXA-196473), SEQ ID NO:135 (BXA-196474), SEQ ID NO: 136 (BXA-196475), SEQ ID NO: 137 (BXA-196476), SEQ ID NO: 139 (BXA-196478), SEQ ID NO: 140 (BXA-196479), SEQ ID NO:148 (BXA-212371), or SEQ ID NO:149 (BXA-212372). In some embodiments, a nucleotide sequence encoding the miniaturized dystrophin polypeptide comprises a nucleic acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 133, wherein the miniaturized dystrophin polypeptide when expressed from the nucleotide sequence has at least one dystrophin activity. In some embodiments, a nucleotide sequence encoding the miniaturized dystrophin polypeptide comprises a nucleic acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 134, wherein the miniaturized dystrophin polypeptide when expressed from the nucleotide sequence has at least one dystrophin activity. In some embodiments, a nucleotide sequence encoding the miniaturized dystrophin polypeptide comprises a nucleic acid sequence at

least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 135. In certain embodiments, a nucleotide sequence encoding the miniaturized dystrophin polypeptide comprises a nucleic acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 136. In certain embodiments, a nucleotide sequence encoding the miniaturized dystrophin polypeptide comprises a nucleic acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 137. In certain embodiments, a nucleotide sequence encoding the miniaturized dystrophin polypeptide comprises a nucleic acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 138, wherein the miniaturized dystrophin polypeptide when expressed from the nucleotide sequence has at least one dystrophin activity. In certain embodiments, a nucleotide sequence encoding the miniaturized dystrophin polypeptide comprises a nucleic acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 139, wherein the miniaturized dystrophin polypeptide when expressed from the nucleotide sequence has at least one dystrophin activity. In certain embodiments, a nucleotide sequence encoding the miniaturized dystrophin polypeptide comprises a nucleic acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 140, wherein the miniaturized dystrophin polypeptide when expressed from the nucleotide sequence has at least one dystrophin activity. In certain embodiments, a nucleotide sequence encoding the miniaturized dystrophin polypeptide comprises a nucleic acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical

to SEQ ID NO: 141, wherein the miniaturized dystrophin polypeptide when expressed from the nucleotide sequence has at least one dystrophin activity. In certain embodiments, a nucleotide sequence encoding the miniaturized dystrophin polypeptide comprises a nucleic acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 142, wherein the miniaturized dystrophin polypeptide when expressed from the nucleotide sequence has at least one dystrophin activity. In certain embodiments, a nucleotide sequence encoding the miniaturized dystrophin polypeptide comprises a nucleic acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 143, wherein the miniaturized dystrophin polypeptide when expressed from the nucleotide sequence has at least one dystrophin activity. In certain embodiments, a nucleotide sequence encoding the miniaturized dystrophin polypeptide comprises a nucleic acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 144, wherein the miniaturized dystrophin polypeptide when expressed from the nucleotide sequence has at least one dystrophin activity. In certain embodiments, a nucleotide sequence encoding the miniaturized dystrophin polypeptide comprises a nucleic acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 145, wherein the miniaturized dystrophin polypeptide when expressed from the nucleotide sequence has at least one dystrophin activity. In certain embodiments, a nucleotide sequence encoding the miniaturized dystrophin polypeptide comprises a nucleic acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 146, wherein the miniaturized dystrophin polypeptide when expressed from the nucleotide sequence has at least one dystrophin activity. In certain embodiments, a nucleotide sequence encoding the miniaturized dystrophin polypeptide comprises a nucleic acid sequence at least about 60%, at least about 70%, at least about

80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 147, wherein the miniaturized dystrophin polypeptide when expressed from the nucleotide sequence has at least one dystrophin activity. In certain embodiments, a nucleotide sequence encoding the miniaturized dystrophin polypeptide comprises a nucleic acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 148, wherein the miniaturized dystrophin polypeptide when expressed from the nucleotide sequence has at least one dystrophin activity. In certain embodiments, a nucleotide sequence encoding the miniaturized dystrophin polypeptide comprises a nucleic acid sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 149, wherein the miniaturized dystrophin polypeptide when expressed from the nucleotide sequence has at least one dystrophin activity.

Linker

Also provided herein is an amino acid sequence or a nucleotide sequence encoding the amino acid sequence. In some embodiments, the present disclosure is directed to an amino acid sequence encoded by a nucleotide sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 100, wherein the amino acid sequence comprises SEQ ID NO:75 (IHTVREE TMMVMTEDEMP LEI). In some embodiments, the present disclosure is directed to an amino acid sequence encoded by a sequence at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NOs: 99-100 in combination, wherein the amino acid sequence comprises SEQ ID NOs: 74-75 in combination (SEAQIHTVREE TMMVMTEDEMP LEI). In some embodiments, a nucleic acid molecule comprises a nucleotide sequence which encodes an amino acid sequence comprising SEQ ID NO: 75 (IHTVREE TMMVMTEDEMP LEI), wherein the nucleic acid molecule is at least about 80%, at least

about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID NO: 100 (ATCCACACCGTGCGGGAAGAGACAATGATGGTCATGACAGAGGACATGCCCCTGGAAATC).

5 In some embodiments, the amino acid sequence can be used as a linker connecting one or more domains of dystrophin. In some embodiments, the linker connects a first domain of dystrophin and a second domain of dystrophin. In some embodiments, the first domain and the second domain of dystrophin that can be connected are an R1 domain and an R16 domain. In other embodiments, the first domain and the second domain of dystrophin that
10 can be connected are an R3 domain and an R16 domain. In other embodiments, the first domain and the second domain of dystrophin that can be connected are an H2 domain and an R16 domain.

In some embodiments, the disclosure includes a nucleic acid molecule comprising a nucleotide sequence which is at least about 60%, at least 70%, at least about 80%, at least
15 about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID No. 99 (AGTGAAGCTCAG).

In some embodiments, the disclosure includes a nucleic acid molecule comprising a nucleotide sequence which is at least about 60%, at least 70%, at least about 80%, at least
20 about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ ID No. 100 (ATCCACACCGTGCGGGAAGAGACAATGATGGTCATGACAGAGGACATGCCCCTGGAAATC).

In some embodiments, the disclosure includes a nucleic acid molecule comprising a
25 nucleotide sequence which encodes an amino acid sequence that is at least 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to SEQ
ID Nos. 99 and 100 in combination,
(AGTGAAGCTCAGATCCACACCGTGCGGGAAGAGACAATGATGGTCATGACA
30 GAGGACATGCCCCTGGAAATC).

In some embodiments, the linker is placed in a miniaturized dystrophin polypeptide disclosed herein by connecting two domains of the miniaturized dystrophin polypeptide

(e.g., H2 and R16 domains). In some embodiments, the linker is located between H2 and R16 domains in the miniaturized dystrophin polypeptide comprising ABD1-H1-R1-R3-H2-L-R16-R17-R24-H4-CR-C term, wherein ABD1 is actin binding domain 1, H1 is a hinge 1 domain, R1 is a spectrin repeat 1 domain, R3 is a spectrin repeat 3 domain, H2 is a hinge 2 domain, L is the linker, e.g., SEQ ID NO: 74 or SEQ ID NO: 75 (or both in combination) (or a sequence that is at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to these sequences), R16 is a spectrin repeat 16 domain, R17 is a spectrin repeat 17 domain, R24 is a spectrin 24 domain, H4 is a hinge 4 domain, CR is a cysteine rich domain, and C-term is an optional C-terminal domain (or portion thereof).

The present disclosure also provides a linker comprising IHTVREETMMVMTEDEMPLEI (SEQ ID NO: 84). In some embodiments, the linker comprising the sequence set forth in SEQ ID NO: 75 (IHTVREETMMVMTEDEMPLEI) (or a sequence that is at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or about 100% identical to the sequence set forth in SEQ ID NO: 84 (IHTVREETMMVMTEDEMPLEI)) is placed in a miniaturized dystrophin polypeptide disclosed herein by connecting two domains of the miniaturized dystrophin polypeptide (e.g., R1 and R16 domains). In some embodiments, the linker is located between R1 and R16 domains in the miniaturized dystrophin polypeptide comprising ABD1-H1-R1-L-R16-R17-H3-R23-R24-H4-CR-C term, wherein ABD1 is actin binding domain 1, H1 is a hinge 1 domain, R1 is a spectrin repeat 1 domain, L is the linker, e.g., SEQ ID NO: 84, R16 is a spectrin repeat 16 domain, R17 is a spectrin repeat 17 domain, H3 is a hinge 3 domain, R23 is a spectrin 23 domain, R24 is a spectrin 24 domain, H4 is a hinge 4 domain, CR is a cysteine rich domain, and C-term is a C-terminal domain (or portion thereof).

Non-Coding Polynucleotides

In some aspects, provided herein are nucleic acid molecules, e.g., DNA or RNA, comprising a nucleotide sequence encoding a miniaturized dystrophin polypeptide.

In some embodiments, the nucleic acid molecules disclosed herein comprise non-coding components. In some embodiments, the nucleic acid molecules disclosed herein comprise

promoters. Certain exemplary regulatory sequences for mammalian host cell expression include viral elements that direct high levels of protein expression in mammalian cells, such as promoters and/or enhancers derived from cytomegalovirus (CMV), Simian Virus 40 (SV40), adenovirus, (e.g., the adenovirus major late promoter (AdMLP) and polyoma.

5 Alternatively, nonviral regulatory sequences can be used, such as the ubiquitin promoter or β -globin promoter. Still further, regulatory elements composed of sequences from different sources may be used, such as the SRa promoter system, which contains sequences from the SV40 early promoter and the long terminal repeat of human T cell leukemia virus type 1 (Takebe, Y. *et al.* (1988) *Mol. Cell. Biol.* 8:466-472). In certain embodiments, the
10 regulatory sequence comprises a tissue specific promoter. In some embodiments, the tissue specific promoter drives expression of the gene of interest in a tissue selected from the group consisting of heart, liver, lungs, eyes, nervous system, lymphatic system, central nervous system, neuronal cells, muscle and stem cells.

In some embodiments, the promoters disclosed herein are tissue-specific promoters. In
15 some embodiments, the promoter drives expression of the therapeutic protein in hepatocytes, muscle cells, endothelial cells, sinusoidal cells, or neuronal cells, or any combination thereof. In some embodiments, the promoter is selected from the group consisting of a synapsin 1 gene promoter, a mouse thyretin promoter (mTTR), an endogenous human factor VIII promoter (F8), a human alpha-1-antitrypsin promoter
20 (hAAT), a human albumin minimal promoter, a mouse albumin promoter, a tristetraprolin (TTP) promoter, a CASI promoter, a CAG promoter, a cytomegalovirus (CMV) promoter, an α 1-antitrypsin (AAT) promoter, a muscle creatine kinase (MCK) promoter, a myosin heavy chain alpha (α MHC) promoter, a myoglobin (MB) promoter, a desmin (DES) promoter, a SPc5-12 promoter, a 2R5Sc5-12 promoter, a dMCK promoter, a tMCK
25 promoter, an α -synuclein promoter and a phosphoglycerate kinase (PGK) promoter.

In some embodiments, the nucleic acid molecules disclosed herein comprise an intronic sequence. In some embodiments, the intronic sequence is positioned 5' to the nucleotide sequence encoding the miniaturized dystrophin polypeptide. In some embodiments, the intronic sequence is positioned 3' to the promoter. In some embodiments, the intronic
30 sequence comprises a synthetic intronic sequence.

In some embodiments, the nucleic acid molecules disclosed herein comprise a post-transcriptional regulatory element. In some embodiments, the post-transcriptional

regulatory element is positioned 3' to the nucleotide sequence encoding the miniaturized dystrophin polypeptide. In some embodiments, the post-transcriptional regulatory element comprises a mutated woodchuck hepatitis virus post-transcriptional regulatory element (WPRE), a microRNA binding site, or a DNA nuclear targeting sequence, or any combination thereof.

In some embodiments, the nucleic acid molecules disclosed herein comprise a 3'UTR poly(A) tail sequence. In some embodiments, the 3'UTR poly(A) tail sequence is selected from the group consisting of bGH poly(A), actin poly(A), hemoglobin poly(A), and any combination thereof. In some embodiments, the 3'UTR poly(A) tail sequence comprises bGH poly(A).

In some embodiments, the nucleic acid molecules disclosed herein comprise an enhancer sequence. In some embodiments, the nucleic acid molecules disclosed herein comprise a first inverted terminal repeat (ITR) and/or a second ITR. In some embodiments, the first ITR and the second ITR are identical. In some embodiments, the first ITR and/or the second ITR are derived from adeno-associated virus. In some embodiments, the first ITR is derived from adeno-associated virus, and the second ITR is derived from adeno-associated virus.

It is further recognized that the nucleic acid molecule can comprise additional elements that aid in the translation of the polypeptide. Such sequences include, for example, Kozak sequences attached to the 5' end of the polynucleotide encoding polypeptide. The Kozak consensus sequence is a sequence which occurs on eukaryotic mRNA that plays a role in the initiation of the translation process and has the consensus (gee)gccRccAUGG (SEQ ID NO:150); wherein (1) a lower case letter denotes the most common base at a position where the base can nevertheless vary; (2) upper case letters indicate highly- conserved bases, i.e. the 'AUGG' sequence is constant or rarely, if ever, changes, with the exception being the IUPAC ambiguity code 'R' which indicates that a purine (adenine or guanine) is normally observed at this position; and (3) the sequence in brackets ((gee)) is of uncertain significance.

In one non-limiting embodiment, the nucleic acid molecule comprises a functional variant or fragment thereof of a Kozak sequence. A functional variant or fragment thereof of a Kozak sequence will retain the ability to increase translation of the protein when compared to the level of translation from a sequence lacking the leader. Such a functional fragment can comprise at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40 continuous

nucleotides of a Kozak sequence or the sequence set forth in SEQ ID NO:150 or SEQ ID NO:151 (gccaccATGG). Alternatively, a functional variant can comprise at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to the Kozak sequence or the sequence set forth in SEQ ID NO:150 or SEQ ID NO:151.

5 Heterologous Moieties

In some embodiments, the polypeptides of the present disclosure can further comprise an additional element, *e.g.*, heterologous moiety. Such elements can aid in the expression of the polypeptide, aid in the secretion of the polypeptide, improve the stability of the polypeptide, allow for more efficient purification of the polypeptide, and/or modulate the activity of the polypeptide. In some embodiments, the heterologous moiety is a polypeptide moiety. In other embodiments, the heterologous moiety is a non-polypeptide moiety.

In some embodiments, the polypeptide comprises a heterologous moiety fused to the polypeptide.

In some embodiments, the polypeptide disclosed herein comprises one or more additional heterologous moieties. In some embodiments, the heterologous moieties are half-life extending moieties. In some embodiments, the heterologous moiety comprises albumin or a fragment thereof, an immunoglobulin Fc region, the C-terminal peptide (CTP) of the β subunit of human chorionic gonadotropin, a PAS sequence, a HAP sequence, a transferrin or a fragment thereof, or an albumin-binding moiety or a derivative thereof, or any combination thereof.

In some embodiments, the polypeptides disclosed herein comprise one or more additional heterologous moieties. In some embodiments, the heterologous moieties are half-life extending moieties. In some embodiments, the heterologous moiety comprises albumin, an immunoglobulin constant region or a portion thereof, an immunoglobulin-binding polypeptide, an immunoglobulin G (IgG), albumin-binding polypeptide (ABP), a PASylation moiety, a HESylation moiety, XTEN, a PEGylation moiety, or an Fc region, or any combination thereof.

Cells

In certain aspects, provided herein are cells (*e.g.*, host cells) expressing (*e.g.*, recombinantly) proteins described herein and expression vectors comprising nucleotides that encode proteins described herein.

- 5 In some embodiments, the host cell comprises the nucleic acid molecules described herein.
In some embodiments, the host cell comprises the vectors described herein.

In some embodiments, the host cell is a eukaryotic cell. In some embodiments, the host cell is selected from the group consisting of a mammalian cell, an insect cell, a yeast cell, a transgenic mammalian cell, and a plant cell. In some embodiments, the host cell is a
10 prokaryotic cell. In some embodiments, the prokaryotic cell is a bacterial cell.

- In some embodiments, the host cell is a mammalian cell. Such mammalian host cells include but are not limited to CHO, VERO, BHK, Hela, MDCK, HEK 293, NIH 3T3, W138, BT483, Hs578T, HTB2, BT2O and T47D, NS0 (a murine myeloma cell line that does not endogenously produce any immunoglobulin chains), CRL7030, COS (*e.g.*, COS1
15 or COS), PER.C6, VERO, HsS78Bst, HEK-293T, HepG2, SP210, R1.1, B-W, L-M, BSC1, BSC40, YB/20, BMT10, HBK, NSO, HT1080 and HsS78Bst cells.

Vectors

Adeno-Associate Virus (AAV)

Overview

- 20 Provided herein are vectors (*e.g.*, expression vectors) comprising nucleic acid molecules comprising nucleotide sequences encoding a miniaturized dystrophin protein for recombinant expression in host cells and cells targeted for therapeutic intervention. The term "vector," as used herein, is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked; or an entity comprising such
25 a nucleic acid molecule capable of transporting another nucleic acid. One type of vector is a "plasmid," which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors, or polynucleotides that are part of vectors, are capable of autonomous replication in a host cell into which they
30 are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication, and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) can be

integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression vectors").

5 In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can sometimes be used interchangeably, depending on the context, as the plasmid is the most commonly used form of vector. However, also disclosed herein are other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

10 In some embodiments, the polynucleotides disclosed herein are expressed using an adeno-associated virus (AAV). AAV is a nonenveloped, single-stranded DNA virus of the *Parvoviridae* family. In contrast to most other members of the *Parvoviridae* family, AAV is replication defective and is only able to replicate efficiently in the presence of a helper virus such as adenovirus or herpes virus.

15 AAV was first discovered in the mid 1960's as a contaminant of viral preparations of adenovirus. *See* Atchison R W, Casto B C, HAMMON W M. *Science*. 149(3685), 754-756 (1965). Since then, progressively safer and more effective methods to use AAV as a recombinant DNA vector have been developed. *See, e.g.*, Hermonat P.L. and Muzyczka N. 20 *Proc Natl Acad Sci USA*. 81(20), 6466-6470 (1984). 3. Laughlin C.A., *et al. Gene*, 23(1), 65-73 (1983). Matsushita T., *et al. Gene Ther*. 5(7), 938-945 (1998). Xiao X., *et al. Journal of Virology*. 72(3), 2224-2232 (1998). Low numbers of AAV genomes have been shown to integrate into the host chromosome. *See* Cheung AK, Hoggan MD, Hauswirth WW, *et al.* Integration of the adeno-associated virus genome into cellular DNA in latently infected 25 human detroit 6 cells. *J Virol* 1980;33:739-748. AAV is immunologically distinct from any known adenovirus antigen. The AAV capsid contains a single-stranded DNA (ssDNA) genome. *See* Rose JA, Berns KI, Hoggan MD, *et al. Proc Natl Acad Sci USA* 1969;64:863-869.

30 AAV has a single stranded, 4.7 kb DNA genome encoding a replication (rep) gene and a capsid (cap) genes flanked by two inverted terminal repeats (ITRs). It is predominantly non-integrating, and forms stable episomes in non-dividing tissue. In spite of its high seroprevalence in the adult human population, AAV has not been associated with any

human disease. See Gonçalves, M. *Virol. J.* 2, 43 (2005). AAV's stable expression in tissues, its lack of pathogenicity, and its ease of high titer production have made it a very attractive vector and popular gene transfer platform.

5 A recombinant AAV (rAAV) is a genetically manipulated AAV in which typically part or all of the rep and cap genes have been replaced with heterologous transgene sequences. Recombinant AAVs too can trigger long-term transgene expression in postmitotic cells, most likely because the recombinant AAV genome persist as largely circular episomes within the nucleus. rAAVs' only DNA *cis*-element required for the production of rAAVs is the AAV inverted terminal repeats (ITRs), whereas *rep*, *cap*, and adenoviral helper genes
10 can be provided in *trans*. Thus, in some embodiments disclosed herein, rAAVs contain only heterologous transgene DNA flanked by the ITRs, and this genome is encapsidated within a serotype-specific AAV capsid.

AAV possesses unique features that make it attractive as a vector system for delivering foreign DNA into cells. AAV infection of cells in culture has generally been noncytopathic,
15 and natural infection of humans and other animals is silent and asymptomatic. Moreover, AAV infects many different types of mammalian cells allowing the possibility of targeting many different tissues *in vivo*. AAV also possesses additional advantages that make it a particularly attractive viral system for gene delivery, including the promotion of an immune response that is relatively mild compared to other forms of gene delivery, and persistent
20 expression in both dividing and quiescent cells based on non-integrating, episomal vector DNA. Also, AAV withstands the conditions used to inactivate adenovirus (56° to 65° C for several hours), making cold preservation of rAAV-based vaccines less critical.

Replication of the viral DNA is not required for integration into the host-cell genome, and thus helper virus is not required for this process. The AAV proviral genome is infectious
25 as cloned DNA in plasmids which makes construction of recombinant genomes feasible. Furthermore, because the signals directing AAV replication, genome encapsidation and integration are contained within the ITRs of the AAV genome, the internal approximately 4.7 kb of the genome (encoding the replication and structural capsid proteins, rep-cap) can thus be replaced with foreign DNA such as a gene cassette containing a promoter, a DNA
30 of interest and a polyadenylation signal.

AAV vectors can include additional elements that function in *cis* or in *trans*. In particular embodiments, an AAV vector that includes a vector genome also has one or more inverted

terminal repeat (ITR) sequences that flank the 5' or 3' terminus of the donor sequence; an expression control element that drives transcription (*e.g.*, a promoter or enhancer) of the donor sequence, such as a constitutive or regulatable control element, or tissue-specific expression control element; an intron sequence, a stuffer or filler polynucleotide sequence; and/or a poly-Adenine sequence located 3' of the donor sequence.

In some embodiments, AAV replicates using a helper virus. A variety of such helper viruses for AAV are known in the art, including adenoviruses, herpesviruses and poxviruses such as vaccinia. The adenoviruses encompass a number of different subgroups, although Adenovirus type 5 of subgroup C is most commonly used. Numerous adenoviruses of human, non-human mammalian and avian origin are known and available from depositories such as the ATCC. Viruses of the herpes family include, for example, herpes simplex viruses (HSV) and Epstein-Barr viruses (EBV), as well as cytomegaloviruses (CMV) and pseudorabies viruses (PRV); which are also available from depositories such as ATCC.

Exemplary AAV vectors include capsid sequences of any of AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, Rh10, Rh74 or AAV-2i8, or a capsid variant of AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, Rh10, Rh74 or AAV-2i8. Recombinant AAV vectors of the invention also include AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, Rh10, Rh74 or AAV-2i8, and variants thereof. Particular capsid variants include capsid variants of AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, Rh10, Rh74 or AAV-2i8, such as a capsid sequence with an amino acid substitution, deletion or insertion/addition. In one embodiment, the AAV vector is AAV9.

In some aspects the disclosure relates to AAVs having distinct tissue targeting capabilities (*e.g.*, tissue tropisms). In some embodiments, the variant AAV capsid polypeptides further exhibit increased transduction or tropism in one or more human stem cell types as compared to non-variant parent capsid polypeptides. In some embodiments, the human stem cell types include but are not limited to embryonic stem cells, adult tissue stem cells (*i.e.*, somatic stem cells), bone marrow stem cells, progenitor cells, induced pluripotent stem cells, and reprogrammed stem cells. In some embodiments, adult stem cells can include organoid stem cells (*i.e.*, stem cells derived from any organ or organ system of interest within the body). In some embodiments, the target tissue of an AAV is gonad, diaphragm, heart,

stomach, liver, spleen, pancreas, muscle or kidney. In some embodiments, the AAV targets organs of the body that include, but are not limited to, skin, hair, nails, sense receptors, sweat gland, oil glands, bones, muscles, brain, spinal cord, nerve, pituitary gland, pineal gland, hypothalamus, thyroid gland, parathyroid, thymus, adrenals, pancreas (islet tissue), heart, blood vessels, lymph nodes, lymph vessels, thymus, spleen, tonsils, nose, pharynx, larynx, trachea, bronchi, lungs, mouth, pharynx, esophagus, stomach, small intestine, large intestine, rectum, anal canal, teeth, salivary glands, tongue, liver, gallbladder, pancreas, appendix, kidneys, ureters, urinary bladder, urethra, testes, ductus (vas) deferens, urethra, prostate, penis, scrotum, ovaries, uterus, uterine (fallopian) tubes, vagina, vulva, and mammary glands (breasts). Organ systems of the body include but are not limited to the integumentary system, skeletal system, muscular system, nervous system, endocrine system, cardiovascular system, lymphatic system, respiratory system, digestive system, urinary system, and reproductive system. In some embodiments, transduction and/or tropism of an AAV with variant capsid polypeptides is increased by about 5%, about 10%, about 15%, about 20%, about 25%, about 30%, about 35%, about 40%, about 45%, about 50%, about 55%, about 60%, 65%, about 70%%, about 75%, about 80%, about 85%, about 90%, about 95%, about 99%, or about 100%, by comparison to an AAV having non-variant capsid polypeptides. In some embodiments, transduction and/or tropism is increased by about 5% to about 80%, about 10% to about 70%, about 20% to about 60% or about 30% to about 60%.

Replication, Capsid, and Assembly AAV Genes

The single-stranded genome of AAV comprises three genes, *rep* (Replication), *cap* (Capsid), and *aap* (Assembly). These three genes give rise to at least nine gene products through the use of three promoters, alternative translation start sites, and differential splicing.

The *rep* gene encodes four proteins (Rep78, Rep68, Rep52, and Rep40), which are required for viral genome replication and packaging.

Cap gene expression gives rise to the viral capsid proteins (VP1; VP2; VP3), which form the outer capsid shell that protects the viral genome, as well as being actively involved in cell binding and internalization. It is estimated that the viral coat is comprised of 60 proteins arranged into an icosahedral structure.

The *aap* gene encodes the assembly-activating protein (AAP) in an alternate reading frame overlapping the *cap* gene. This nuclear protein is thought to provide a scaffolding function for capsid assembly and plays a role in nucleolar localization of VP proteins in some AAV serotypes.

- 5 In some embodiments, one or more of the *rep*, *cap*, or *aap* genes are naturally occurring, e.g. the *rep*, *cap*, or *aap* genes comprise all or a portion of parvovirus *rep*, *cap*, or *aap* genes. In some embodiments, the one or more of the *rep*, *cap*, or *aap* genes comprise a synthetic sequence.

In one embodiment, the *rep* gene comprises a synthetic sequence. In one embodiment, the
10 *cap* gene comprises a synthetic sequence. In one embodiment, the *aap* gene comprises a synthetic sequence. In one embodiment, the *rep* and *cap* genes comprise a synthetic sequence. In one embodiment, the *rep* and *aap* genes comprise a synthetic sequence. In one embodiment, the *cap* and *aap* genes comprise a synthetic sequence. In one embodiment, the *rep*, *cap*, and *aap* genes comprise a synthetic sequence.

- 15 In some embodiments, *rep* is from an AAV genome selected from AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, and any combination thereof. In a particular embodiment, *rep* is from the AAV1 genome. In a particular embodiment, *rep* is from the AAV2 genome. In a particular embodiment, *rep* is from the AAV3 genome. In a particular embodiment, *rep* is from the AAV4 genome. In a particular
20 embodiment, *rep* is from the AAV5 genome. In a particular embodiment, *rep* is from the AAV6 genome. In a particular embodiment, *rep* is from the AAV7 genome. In a particular embodiment, *rep* is from the AAV8 genome. In a particular embodiment, *rep* is from the AAV9 genome. In a particular embodiment, *rep* is from the AAV10 genome. In a particular embodiment, *rep* is from the AAV11 genome.

- 25 In some embodiments, *cap* is from an AAV genome selected from AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, and any combination thereof. In a particular embodiment, *cap* is from the AAV1 genome. In a particular embodiment, *cap* is from the AAV2 genome. In a particular embodiment, *cap* is from the AAV3 genome. In a particular embodiment, *cap* is from the AAV4 genome. In a particular
30 embodiment, *cap* is from the AAV5 genome. In a particular embodiment, *cap* is from the AAV6 genome. In a particular embodiment, *cap* is from the AAV7 genome. In a particular embodiment, *cap* is from the AAV8 genome. In a particular embodiment, *cap* is from the

AAV9 genome. In a particular embodiment, *cap* is from the AAV10 genome. In a particular embodiment, *cap* is from the AAV11 genome.

In some embodiments, *aap* is from an AAV genome selected from AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, and any combination thereof. In a particular embodiment, *aap* is from the AAV1 genome. In a particular embodiment, *aap* is from the AAV2 genome. In a particular embodiment, *aap* is from the AAV3 genome. In a particular embodiment, *aap* is from the AAV4 genome. In a particular embodiment, *aap* is from the AAV5 genome. In a particular embodiment, *aap* is from the AAV6 genome. In a particular embodiment, *aap* is from the AAV7 genome. In a particular embodiment, *aap* is from the AAV8 genome. In a particular embodiment, *aap* is from the AAV9 genome. In a particular embodiment, *aap* is from the AAV10 genome. In a particular embodiment, *aap* is from the AAV11 genome.

It is to be understood that a particular AAV genome described herein could have genes derived from different AAV genomes (*e.g.*, genomes from AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, and AAV11). Thus, disclosed herein are AAVs that comprise any possible permutation/combination of *rep*, *cap*, or *aap*.

In some embodiments disclosed herein, the AAV is recombinant AAV (rAAV). In some embodiments, the rAAV lacks one or more of the *rep* gene, the *cap* gene, and the *aap* gene. In some embodiments, the rAAV lacks a *rep* gene. In some embodiments, the rAAV lacks a *cap* gene. In some embodiments, the rAAV lacks an *aap* gene. In some embodiments, the rAAV lacks a *rep* gene and lacks a *cap* gene. In some embodiments, the rAAV lacks a *rep* gene and lacks an *aap* gene. In some embodiments, the rAAV lacks a *cap* gene and lacks an *aap* gene. In some embodiments, the rAAV lacks a *rep* gene, a *cap* gene, and an *aap* gene.

In some embodiments disclosed herein, the rAAV is modified so that one or more of the *rep* gene, the *cap* gene, and the *aap* gene is mutated so that expression of one or more of the AAV genes is modified. In some embodiments, the *rep* gene is mutated. In some embodiments, the *cap* gene is mutated. In some embodiments, the *aap* gene is mutated. In some embodiments, the *rep* gene and the *cap* gene are mutated. In some embodiments, the *rep* gene and the *aap* gene are mutated. In some embodiments, the *cap* gene and the *aap* gene are mutated. In some embodiments, the *cap* gene, the *rep* gene, and the *aap* gene are mutated.

Inverted Terminal Repeats

In certain embodiments, the AAV comprises a first ITR, *e.g.*, a 5' ITR, and second ITR, *e.g.*, a 3' ITR. Typically, ITRs are involved in parvovirus (*e.g.*, AAV) DNA replication and rescue, or excision, from prokaryotic plasmids (Samulski *et al.*, 1983, 1987; Senapathy *et al.*, 1984; Gottlieb and Muzyczka, 1988). In addition, ITRs have been reported to be the minimum sequences required for AAV proviral integration and for packaging of AAV DNA into virions (McLaughlin *et al.*, 1988; Samulski *et al.*, 1989). These elements are essential for efficient multiplication of a parvovirus genome.

In some embodiments, the ITR comprises a naturally occurring ITR, *e.g.*, the ITR comprises all or a portion of a parvovirus ITR. In some embodiments, the ITR comprises a synthetic sequence. In one embodiment, the first ITR or the second ITR comprises a synthetic sequence. In another embodiment, each of the first ITR and the second ITR comprises a synthetic sequence. In some embodiments, the first ITR or the second ITR comprises a naturally occurring sequence. In another embodiment, each of the first ITR and the second ITR comprises a naturally occurring sequence.

In some embodiments, the ITR comprises an ITR from an AAV genome. In some embodiments, the ITR is an ITR of an AAV genome selected from AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, and any combination thereof. In a particular embodiment, the ITR is an ITR of the AAV2 genome. In another embodiment, the ITR is a synthetic sequence genetically engineered to include at its 5' and 3' ends ITRs derived from one or more of AAV genomes. In some embodiments, the ITRs are derived from the same genome, *e.g.*, from the genome of the same virus, or from different genomes, *e.g.*, from the genomes of two or more different AAV genomes. In certain embodiments, the ITRs are derived from the same AAV genome. In a specific embodiment, the two ITRs present in the nucleic acid molecule of the invention are the same, and can in particular be AAV2 ITRs. In one particular embodiment, the first ITR and the second ITR are identical.

In some embodiments, the ITRs form hairpin loop structures. In one embodiment, the first ITR forms a hairpin structure. In another embodiment, the second ITR forms a hairpin structure. Still in another embodiment, both the first ITR and the second ITR form hairpin structures.

In some embodiments, an ITR in a nucleic acid molecule described herein is a transcriptionally activated ITR. A transcriptionally-activated ITR can comprise all or a portion of a wild-type ITR that has been transcriptionally activated by inclusion of at least one transcriptionally active element. Various types of transcriptionally active elements are suitable for use in this context. In some embodiments, the transcriptionally active element is a constitutive transcriptionally active element. Constitutive transcriptionally active elements provide an ongoing level of gene transcription, and can be used when it is desired that the transgene be expressed on an ongoing basis. In other embodiments, the transcriptionally active element is an inducible transcriptionally active element. Inducible transcriptionally active elements generally exhibit low activity in the absence of an inducer (or inducing condition), and are up-regulated in the presence of the inducer (or switch to an inducing condition). Inducible transcriptionally active elements can be used when expression is desired only at certain times or at certain locations, or when it is desirable to titrate the level of expression using an inducing agent. Transcriptionally active elements can also be tissue-specific; that is, they exhibit activity only in certain tissues or cell types. Transcriptionally active elements, can be incorporated into an ITR in a variety of ways. In some embodiments, a transcriptionally active element is incorporated 5' to any portion of an ITR or 3' to any portion of an ITR. In other embodiments, a transcriptionally active element of a transcriptionally-activated ITR lies between two ITR sequences. If the transcriptionally active element comprises two or more elements which must be spaced apart, those elements can alternate with portions of the ITR. In some embodiments, a hairpin structure of an ITR is deleted and replaced with inverted repeats of a transcriptional element. This latter arrangement would create a hairpin mimicking the deleted portion in structure. Multiple tandem transcriptionally active elements can also be present in a transcriptionally-activated ITR, and these can be adjacent or spaced apart. In addition, protein binding sites (*e.g.*, Rep binding sites) can be introduced into transcriptionally active elements of the transcriptionally-activated ITRs. A transcriptionally active element can comprise any sequence enabling the controlled transcription of DNA by RNA polymerase to form RNA, and can comprise, for example, a transcriptionally active element, as defined below.

Transcriptionally-activated ITRs provide both transcriptional activation and ITR functions to the nucleic acid molecule in a relatively limited nucleotide sequence length which

effectively maximizes the length of a transgene which can be carried and expressed from the nucleic acid molecule. Incorporation of a transcriptionally active element into an ITR can be accomplished in a variety of ways. A comparison of the ITR sequence and the sequence requirements of the transcriptionally active element can provide insight into ways

5 to encode the element within an ITR. For example, transcriptional activity can be added to an ITR through the introduction of specific changes in the ITR sequence that replicates the functional elements of the transcriptionally active element. A number of techniques exist in the art to efficiently add, delete, and/or change particular nucleotide sequences at specific sites (see, for example, Deng and Nickoloff (1992) *Anal. Biochem.* 200:81-88). Another
10 way to create transcriptionally-activated ITRs involves the introduction of a restriction site at a desired location in the ITR. In addition, multiple transcriptionally activate elements can be incorporated into a transcriptionally-activated ITR, using methods known in the art.

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

Gene of Interest and Other Sequences

Certain aspects of the present disclosure are directed to methods of administering to a
20 subject an AAV therapy. In some embodiments, the AAV comprises a gene of interest (GOI). In some embodiments, the GOI is a nucleic acid molecule comprising a nucleotide sequence as disclosed herein, which encodes a miniaturized dystrophin polypeptide as disclosed herein.

The GOI being expressed can be either a DNA segment encoding a protein, with any
25 necessary control elements (e.g., promoters, operators) desired by the user, or a non-coding DNA segment, the transcription of which produces all or part of some RNA-containing molecule, such as a ribozyme or an anti-sense molecule.

In some embodiments, the AAV comprises more than one GOI. In AAVs with more than one GOI, some embodiments include elements such as IRES or 2A, to co-express them
30 from one promoter. In some embodiments, the AAV comprises two genes of interest separated by an IRES element. In some embodiments, the AAV comprises two genes of

interest separated by a 2A element. In some embodiments, the AAV comprises three genes of interest separated by an IRES element between the genes of interest (e.g., GOI-IRES-GOI-IRES-GOI). In some embodiments, the AAV comprises three genes of interest separated by 2A elements between the genes of interest.

5 In some embodiments, the AAV comprises a regulatory sequence. In some embodiments, the AAV comprises non-coding regulatory DNA. In some embodiments, the AAV genome comprises regulatory sequences that control the expression of the antibody chain genes in a host cell. The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals) that control the
10 transcription or translation of the antibody chain genes. Such regulatory sequences are described, for example, in Goeddel (Gene Expression Technology. Methods in Enzymology 185, Academic Press, San Diego, CA (1990)). It will be appreciated by those skilled in the art that the design of the AAV, including the selection of regulatory sequences, can depend on such factors as the choice of the host cell to be transformed, the
15 level of expression of protein desired, etc. In some embodiments, the AAV genome comprises mRNA splice donor/splice acceptor sites. Certain regulatory sequences for mammalian host cell expression include viral elements that direct high levels of protein expression in mammalian cells, such as promoters and/or enhancers derived from cytomegalovirus (CMV), Simian Virus 40 (SV40), adenovirus, (e.g., the adenovirus major
20 late promoter (AdMLP) and polyoma. Alternatively, nonviral regulatory sequences can be used, such as the ubiquitin promoter or β -globin promoter. Still further, regulatory elements composed of sequences from different sources, such as the SRa promoter system, which contains sequences from the SV40 early promoter and the long terminal repeat of human T cell leukemia virus type 1 (Takebe, Y. *et al.* (1988) *Mol. Cell. Biol.* 8:466-472). In certain
25 embodiments, the regulatory sequence comprises a tissue specific promoter. In some embodiments, the tissue specific promoter drives expression of the gene of interest in a tissue selected from the group consisting of heart, liver, lungs, eyes, nervous system, lymphatic system, muscle and stem cells.

AAV Formulations

30 In some embodiments, the AAV vector is formulated with a delivery agent. In some embodiments, the delivery agent comprises a lipid nanoparticle. In some embodiments, the

delivery agent is selected from the group consisting of liposomes, non-lipid polymeric molecules, endosomes, and any combination thereof.

Non-AAV Vectors

5 A vector which comprises the above-described polynucleotides operably linked to a promoter is also provided herein. A nucleotide sequence is "operably linked" to an expression control sequence (*e.g.*, a promoter) when the expression control sequence controls and regulates the transcription and translation of that sequence. The term "operably linked" when referring to a nucleotide sequence includes having an appropriate start signal (*e.g.*, ATG) in front of the nucleotide sequence to be expressed and maintaining the correct
10 reading frame to permit expression of the sequence under the control of the expression control sequence and production of the desired product encoded by the sequence. If a gene that one desires to insert into a recombinant nucleic acid molecule does not contain an appropriate start signal, such a start signal can be inserted in front of the gene. A "vector" is a replicon, such as plasmid, phage or cosmid, to which another nucleic acid segment can
15 be attached so as to bring about the replication of the attached segment. The promoter can be, or is identical to, a bacterial, yeast, insect or mammalian promoter.

In some embodiments, the vector can be a plasmid, cosmid, yeast artificial chromosome (YAC), bacteriophage or eukaryotic viral DNA. Other numerous vector backbones known in the art as useful for expressing protein can be employed. Such vectors include, but are
20 not limited to:

adenoviral vector, a retroviral vector, poxvirus vector, a baculovirus vector, a herpes viral vector, simian virus 40 (SV40), cytomegalovirus (CMV), mouse mammary tumor virus (MMTV), and Moloney murine leukemia virus. Further, one class of vectors comprises DNA elements derived from viruses such as bovine papilloma virus, polyoma virus,
25 baculovirus, retroviruses, or Semliki Forest virus. Such vectors can be obtained commercially or assembled from the sequences described by methods well-known in the art.

In some embodiments, the vector described herein is formulated with a delivery agent. In some embodiments, the delivery agent comprises a lipid nanoparticle. In some
30 embodiments, the delivery agent is selected from the group consisting of liposomes, non-lipid polymeric molecules, endosomes, and any combination thereof.

Pharmaceutical Compositions

The various polypeptides and polynucleotides disclosed herein (also referred to herein as "active compounds") can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the polypeptide, or polynucleotides, and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active compounds is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

In some embodiments, disclosed is a pharmaceutical composition comprising (a) a polypeptide as described herein and (b) a pharmaceutically acceptable excipient.

In some embodiments, disclosed is a pharmaceutical composition comprising (a) a composition comprising a polypeptide as described herein and (b) a pharmaceutically acceptable excipient.

In some embodiments, disclosed is a pharmaceutical composition comprising (a) a polynucleotide as described herein and (b) a pharmaceutically acceptable excipient.

In some embodiments, disclosed is a pharmaceutical composition comprising (a) a vector (e.g., rAAV) as described herein and (b) a pharmaceutically acceptable excipient.

In some embodiments, disclosed is a pharmaceutical composition comprising (a) a host cell as described herein and (b) a pharmaceutically acceptable excipient.

A pharmaceutical composition of the disclosure is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral, transdermal (topical), and transmucosal, and any combination thereof. Another route of administration includes pulmonary administration. In addition, it can be desirable to administer a therapeutically effective amount of the pharmaceutical composition locally to an area in need of treatment. This can be achieved by, for example, local or regional infusion or perfusion during surgery, topical application, injection, catheter, suppository, or implant (for example, implants formed from porous, non-porous, or gelatinous materials, including membranes, such as sialastic

membranes or fibers), and the like. In another embodiment, the therapeutically effective amount of the pharmaceutical composition is delivered in a vesicle, such as liposomes (see, *e.g.*, Langer, *Science* 249:1527-33, 1990 and Treat *et al.*, in *Liposomes in the Therapy of Infectious Disease and Cancer*, Lopez Berestein and Fidler (eds.), Liss, N.Y., pp. 353-65, 1989).

In yet another embodiment, the therapeutically effective amount of the pharmaceutical composition can be delivered in a controlled release system. In one example, a pump can be used (see, *e.g.*, Langer, *Science* 249:1527-33, 1990; Sefton, *Crit. Rev. Biomed. Eng.* 14:201-40, 1987; Buchwald *et al.*, *Surgery* 88:507-16, 1980; Saudek *et al.*, *N Engl. J Med.* 321:574-79, 1989). In another example, polymeric materials can be used (see, *e.g.*, Levy *et al.*, *Science* 228:190-92, 1985; During *et al.*, *Ann. Neural.* 25:351-56, 1989; Howard *et al.*, *J Neurosurg.* 71:105-12, 1989). Other controlled release systems, such as those discussed by Langer (*Science* 249:1527-33, 1990), can also be used.

Acceptable carriers, excipients, or stabilizers are nontoxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, and other organic acids; antioxidants including ascorbic acid and methionine; preservatives (such as octadecyldimethylbenzyl ammonium chloride; hexamethonium chloride; benzalkonium chloride, benzethonium chloride; phenol, butyl or benzyl alcohol; alkyl parabens such as methyl or propyl paraben; catechol; resorcinol; cyclohexanol; 3-pentanol; and m-cresol); low molecular weight (less than about 10 residues) polypeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids such as glycine, glutamine, asparagine, histidine, arginine, or lysine; monosaccharides, disaccharides, and other carbohydrates including glucose, mannose, or dextrans; chelating agents such as EDTA; sugars such as sucrose, mannitol, trehalose or sorbitol; salt-forming counter-ions such as sodium; metal complexes (*e.g.*, Zn-protein complexes); and/or non-ionic surfactants such as TWEEN™, PLURONICS™ or polyethylene glycol (PEG).

Pharmaceutically acceptable carriers used in parenteral preparations include aqueous vehicles, nonaqueous vehicles, antimicrobial agents, isotonic agents, buffers, antioxidants, local anesthetics, suspending and dispersing agents, emulsifying agents, sequestering or chelating agents and other pharmaceutically acceptable substances. Examples of aqueous vehicles include Sodium Chloride Injection, Ringers Injection, Isotonic Dextrose Injection,

Sterile Water Injection, Dextrose and Lactated Ringers Injection. Nonaqueous parenteral vehicles include fixed oils of vegetable origin, cottonseed oil, corn oil, sesame oil and peanut oil. Antimicrobial agents in bacteriostatic or fungistatic concentrations can be added to parenteral preparations packaged in multiple-dose containers which include phenols or cresols, mercurials, benzyl alcohol, chlorobutanol, methyl and propyl p-hydroxybenzoic acid esters, thimerosal, benzalkonium chloride and benzethonium chloride. Isotonic agents include sodium chloride and dextrose. Buffers include phosphate and citrate. Antioxidants include sodium bisulfate. Local anesthetics include procaine hydrochloride. Suspending and dispersing agents include sodium carboxymethylcellulose, hydroxypropyl methylcellulose and polyvinylpyrrolidone. Emulsifying agents include Polysorbate 80 (TWEEN® 80). A sequestering or chelating agent of metal ions includes EDTA. Pharmaceutical carriers also include ethyl alcohol, polyethylene glycol and propylene glycol for water miscible vehicles; and sodium hydroxide, hydrochloric acid, citric acid or lactic acid for pH adjustment.

Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes, or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor ELS (BASF; Parsippany, NJ), or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol,

propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion, and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride, in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent that delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the methods of preparation can be vacuum drying and freeze-drying, which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from a pressurized container or dispenser that contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer. Systemic administration can also be by transmucosal or transdermal means.

For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art. The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated with each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the disclosure are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such a functional compound for the treatment of individuals. The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Uses and Methods

Methods of Producing Miniaturized Dystrophins

Also disclosed herein are methods of producing a miniaturized dystrophin polypeptide, comprising: culturing a host cell described herein under suitable conditions and recovering the miniaturized dystrophin polypeptide.

As used herein, an "isolated" polynucleotide or nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source (*e.g.*, in a mouse or a human) of the nucleic acid molecule. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of

chemical precursors or other chemicals when chemically synthesized. For example, the language "substantially free" includes preparations of polynucleotide or nucleic acid molecule having less than about 15%, 10%, 5%, 2%, 1%, 0.5%, or 0.1% (in particular less than about 10%) of other material, *e.g.*, cellular material, culture medium, other nucleic acid molecules, chemical precursors and/or other chemicals. In a specific embodiment, a

nucleic acid molecule(s) encoding a polypeptide described herein is isolated or purified. The polynucleotides can be obtained, and the nucleotide sequence of the polynucleotides determined, by any method known in the art. Nucleotide sequences encoding polypeptides described herein, *e.g.*, the polypeptides described in Tables 3 and 4, and modified versions of these polypeptides can be determined using methods well known in the art, *i.e.*, nucleotide codons known to encode particular amino acids are assembled in such a way to generate a nucleic acid that encodes the polypeptides. Such a polynucleotide encoding the polypeptide can be assembled from chemically synthesized oligonucleotides (*e.g.*, as described in Kutmeier G *et al.*, (1994), *BioTechniques* 17: 242-6), which, briefly, involves the synthesis of overlapping oligonucleotides containing portions of the sequence encoding the polypeptide, annealing and ligating of those oligonucleotides, and then amplification of the ligated oligonucleotides by PCR.

Alternatively, a polynucleotide encoding a polypeptide described herein can be generated from nucleic acid from a suitable source (*e.g.*, a hybridoma) using methods well known in the art (*e.g.*, PCR and other molecular cloning methods). For example, PCR amplification using synthetic primers hybridizable to the 3' and 5' ends of a known sequence can be performed using genomic DNA obtained from hybridoma cells producing the polypeptide of interest. Such PCR amplification methods can be used to obtain nucleic acids comprising the sequence encoding *e.g.*, IL2, a linker sequence, or IL2-R α . The amplified nucleic acids can be cloned into vectors for expression in host cells and for further cloning, for example, to generate polypeptides.

If a clone containing a nucleic acid encoding a particular polypeptide is not available, but the sequence of the polypeptide molecule is known, a nucleic acid encoding the polypeptide can be chemically synthesized or obtained from a suitable source (*e.g.*, a cDNA library or a cDNA library generated from, or nucleic acid, preferably poly A⁺ RNA, isolated from, any tissue or cells expressing the proteins of interest, such as hybridoma cells selected to express a polypeptide described herein) by PCR amplification using synthetic primers

hybridizable to the 3' and 5' ends of the sequence or by cloning using an oligonucleotide probe specific for the particular gene sequence to identify, *e.g.*, a cDNA clone from a cDNA library that encodes the polypeptides. Amplified nucleic acids generated by PCR can then be cloned into replicable cloning vectors using any method well known in the art.

- 5 DNA encoding polypeptides described herein can be readily isolated and sequenced using conventional procedures (*e.g.*, by using oligonucleotide probes that are capable of binding specifically to genes encoding the polypeptides disclosed herein). Hybridoma cells can serve as a source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as *E. coli* cells, simian COS cells, 10 Chinese hamster ovary (CHO) cells (*e.g.*, CHO cells from the CHO GS SYSTEM™ (Lonza)), or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of polypeptides in the recombinant host cells.

Therapeutic Uses and Methods

- The miniaturized dystrophin polypeptides, polynucleotides encoding miniaturized 15 dystrophin polypeptides, vectors (*e.g.*, rAAV) harboring polynucleotides encoding miniaturized dystrophin polypeptides and methods described herein have numerous *in vitro* and *in vivo* utilities. For example, the nucleotide sequence encoding a miniaturized dystrophin polypeptide, *e.g.*, a vector, *e.g.*, an AAV vector, or the polypeptides described herein can be administered to cells in culture, *in vitro* or *ex vivo*, or to human subjects, *e.g.*, 20 *in vivo*, to treat diseases.

Accordingly, disclosed herein are therapeutic methods using any of the miniaturized dystrophin nucleic acid molecules as disclosed herein, polypeptides as disclosed herein, host cells as disclosed herein, vectors as disclosed herein, or pharmaceutical compositions as disclosed herein, or any combination thereof.

- 25 In some embodiments, disclosed herein is a method of expressing a miniaturized dystrophin polypeptide in a subject in need thereof, comprising administering to the subject a nucleic acid as disclosed herein, a vector as disclosed herein, a host cell as disclosed herein, or a pharmaceutical composition as disclosed herein.

- In some embodiments, disclosed herein is a method of treating a subject having a disease 30 or condition comprising administering to the subject a nucleic acid as disclosed herein, a vector as disclosed herein, a polypeptide as disclosed herein, a host cell as disclosed herein,

or a pharmaceutical composition as disclosed herein. In some embodiments, the disease or condition is caused by dystrophin deficiency. In some embodiments, the disease is Duchene muscular dystrophy (DMD), Becker muscular dystrophy (BMD), X-linked dilated cardiomyopathy (XLDC), facioscapulohumeral muscular dystrophy, myotonic muscular dystrophy, limb-girdle muscular dystrophy, oculopharyngeal muscular dystrophy, Emery-Dreifuss muscular dystrophy, distal muscular dystrophy, and/or congenital muscular dystrophy. In other embodiments, the disease to be treated is Sarcopenia, heart disease, cachexia.

In some embodiments, a nucleic acid molecule as disclosed herein, a polypeptide as disclosed herein, a vector (e.g., rAAV) as disclosed herein, a host cell as disclosed herein, or a pharmaceutical composition as disclosed herein is administered intravenously, transdermally, intradermally, subcutaneously, orally, or pulmonarily, or any combination thereof. In some embodiments, the nucleic acid molecule as disclosed herein, the polypeptide as disclosed herein, the vector as disclosed herein, the host cell as disclosed herein, or the pharmaceutical composition as disclosed herein is administered via a topical, epidermal mucosal, intranasal, oral, vaginal, rectal, sublingual, topical, intravenous, intraperitoneal, intramuscular, intraarterial, intrathecal, intralymphatic, intralesional, intracapsular, intraorbital, intracardiac, intradermal, transtracheal, subcutaneous, subcuticular, intraarticular, subcapsular, subarachnoid, intraspinal, epidural or intrasternal route. In some embodiments, the nucleic acid molecule, the vector (e.g., rAAV), the host cell as disclosed herein, or the polypeptide is administered intravenously.

In some embodiments, the method of treatment further comprises administering to the subject a second agent.

As used herein, the term "subject" includes any human or non-human animal. For example, the methods and compositions described herein can be used to treat a subject having cancer. The term "non-human animal" includes all vertebrates, e.g., mammals and non-mammals, such as non-human primates, sheep, dog, cow, chickens, amphibians, reptiles, etc. In some embodiments, the subject is a human.

In some embodiments, the administration of the nucleic acid molecule, the vector (e.g., rAAV), the polypeptide, the host cell, or the pharmaceutical composition to the subject results in an increased dystrophin protein expression, relative to dystrophin protein expression in the subject prior to the administration, wherein the dystrophin protein

expression is increased by at least about 2-fold, at least about 3-fold, at least about 4-fold, at least about 5-fold, at least about 6-fold, at least about 7-fold, at least about 8-fold, at least about 9-fold, at least about 10-fold, at least about 11-fold, at least about 12-fold, at least about 13-fold, at least about 14-fold, at least about 15-fold, at least about 20-fold, at least about 25-fold, at least about 30-fold, at least about 35-fold, at least about 40-fold, at least about 50-fold, at least about 60-fold, at least about 70-fold, at least about 80-fold, at least about 90-fold, or at least about 100-fold.

In certain aspects of the disclosure, the method comprises, or further comprises, administering an AAV therapy to the subject. In some embodiments, the AAV therapy comprises administering a recombinant AAV. Any recombinant AAV known in the art and/or disclosed herein can be used in the methods of the present disclosure. In some embodiments, the AAV therapy comprises administering an AAV selected from the group consisting of AAV type 1, AAV type 2, AAV type 3 (including types 3A and 3B), AAV type 4, AAV type 5, AAV type 6, AAV type 7, AAV type 8, AAV type 9, AAV type 10, AAV type 11, AAV type 12, AAV type 13, snake AAV, avian AAV, bovine AAV, canine AAV, equine AAV, ovine AAV, goat AAV, shrimp AAV, and any combination thereof. In certain embodiments, the AAV therapy comprises administering an AAV type 1. In certain embodiments, the AAV therapy comprises administering an AAV type 2. In certain embodiments, the AAV therapy comprises administering an AAV type 3. In certain embodiments, the AAV therapy comprises administering an AAV type 4. In certain embodiments, the AAV therapy comprises administering an AAV type 5. In certain embodiments, the AAV therapy comprises administering an AAV type 6. In certain embodiments, the AAV therapy comprises administering an AAV type 7. In certain embodiments, the AAV therapy comprises administering an AAV type 8. In certain embodiments, the AAV therapy comprises administering an AAV type 9. In certain embodiments, the AAV therapy comprises administering an AAV type 10. In certain embodiments, the AAV therapy comprises administering an AAV type 11. In certain embodiments, the AAV therapy comprises administering an AAV type 12. In certain embodiments, the AAV therapy comprises administering an AAV type 13.

In some embodiments, treatment of a subject with the miniaturized dystrophin nucleic acid molecules as disclosed herein, polypeptides as disclosed herein, host cells as disclosed herein, vectors as disclosed herein, or pharmaceutical compositions as disclosed herein, or

any combination thereof, does not cause significant inflammatory reactions, *e.g.*, immune-mediated pneumonitis, immune-mediated colitis, immune mediated hepatitis, immune-mediated nephritis or renal dysfunction, immune-mediated hypophysitis, immune-mediated hypothyroidism and hyperthyroidism, or other immune-mediated adverse reactions. In some embodiments, treatment of a subject with the miniaturized dystrophin nucleic acid molecules as disclosed herein, polypeptides as disclosed herein, host cells as disclosed herein, vectors as disclosed herein, pharmaceutical compositions as disclosed herein, or any combination thereof does not cause significant cardiac disorders, *e.g.*, ventricular arrhythmia; eye disorders, *e.g.*, iridocyclitis; infusion-related reactions; increased amylase, increased lipase; nervous system disorders, *e.g.*, dizziness, peripheral and sensory neuropathy; skin and subcutaneous tissue disorders, *e.g.*, rash, pruritus, exfoliative dermatitis, erythema multiforme, vitiligo or psoriasis; respiratory, thoracic and mediastinal disorders, *e.g.*, cough; fatigue; nausea; decreased appetite; constipation; arthralgia; or diarrhea.

15 **Kits**

Also disclosed herein are kits comprising one or more nucleic acid molecules disclosed herein, one or more vectors (*e.g.*, rAAV) as disclosed herein, one or more polypeptides as disclosed herein, or one or more host cells as disclosed herein, or any combination thereof. In some embodiments, the kit also comprises instructions for administering any of the aforesaid, or a combination thereof, to a subject in need thereof.

20 The terms "kit" and "system," as used herein are intended to refer to at least one or more nucleic acid molecules disclosed herein, one or more vectors (*e.g.*, rAAV) as disclosed herein, one or more polypeptides as disclosed herein, or one or more host cells as disclosed herein, or any combination thereof, which, in specific embodiments, are in combination with one or more other types of elements or components (*e.g.*, other types of biochemical reagents, containers, packages, such as packaging intended for commercial sale, instructions of use, and the like).

25 In some embodiments, disclosed is a kit comprising (a) one or more of a miniaturized dystrophin polypeptide as described herein, a composition comprising a miniaturized dystrophin polypeptide as described herein, a nucleic acid encoding for a miniaturized dystrophin polypeptide as described herein, a vector (*e.g.*, rAAV), and/or a host cell; and

(b) and instructions for administering any of the aforesaid to a subject in need thereof. In some embodiments, disclosed is a kit comprising (a) a miniaturized dystrophin polypeptide as described herein and (b) and instructions for administering the miniaturized dystrophin polypeptide to a subject in need thereof. In some embodiments, disclosed is a kit comprising (a) a composition comprising a miniaturized dystrophin polypeptide as described herein and (b) and instructions for administering the composition to a subject in need thereof. In some embodiments, disclosed is a kit comprising (a) a nucleic acid encoding for a miniaturized dystrophin polypeptide as described herein and (b) and instructions for administering the nucleic to a subject in need thereof. In some embodiments, disclosed is a kit comprising (a) a vector as described herein and (b) and instructions for administering the vector to a subject in need thereof. In some embodiments, disclosed is a kit comprising (a) an AAV vector as described herein and (b) and instructions for administering the vector to a subject in need thereof. In some embodiments, disclosed is a kit comprising (a) a host cell as described herein and (b) and instructions for administering the host cell to a subject in need thereof.

In a specific embodiment, provided herein is a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions described herein, such as one or more miniaturized dystrophin peptides provided herein. In some embodiments, the kits contain a pharmaceutical composition described herein and any prophylactic or therapeutic agent, such as those described herein. In certain embodiments, the kits can contain a T cell mitogen, such as, *e.g.*, phytohaemagglutinin (PHA) and/or phorbol myristate acetate (PMA), or a TCR complex stimulating antibody, such as an anti-CD3 antibody and anti-CD28 antibody. Optionally associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration.

Also provided herein are kits that can be used in the above methods. In one embodiment, a kit comprises a miniaturized dystrophin polypeptide described herein, preferably a purified miniaturized dystrophin polypeptide, in one or more containers. In a specific embodiment, kits described herein contain a substantially isolated miniaturized dystrophin polypeptide as a control. In another specific embodiment, the kits described herein further comprise a

control protein which does not react with a miniaturized dystrophin polypeptide antigen. In another specific embodiment, kits described herein contain one or more elements for detecting the binding of the miniaturized dystrophin polypeptide to a dystrophin antigen (*e.g.*, the miniaturized dystrophin polypeptide can be conjugated to a detectable substrate such as a fluorescent compound, an enzymatic substrate, a radioactive compound or a luminescent compound, or a second antibody which recognizes the first antibody can be conjugated to a detectable substrate). In specific embodiments, a kit provided herein can include a recombinantly produced or chemically synthesized miniaturized dystrophin polypeptide. The antigen to a miniaturized dystrophin polypeptide disclosed herein as provided in the kit can also be attached to a solid support. In a more specific embodiment, the detecting means of the above described kit includes a solid support to which an antigen of the miniaturized dystrophin polypeptide is attached. Such a kit can also include a non-attached reporter-labeled anti-human antibody or anti-mouse/rat antibody. In this embodiment, binding of the miniaturized dystrophin polypeptide to an antigen can be detected by binding of the said reporter-labeled antibody.

The practice of the present disclosure will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, Sambrook *et al.*, ed. (1989) *Molecular Cloning A Laboratory Manual* (2nd ed.; Cold Spring Harbor Laboratory Press); Sambrook *et al.*, ed. (1992) *Molecular Cloning: A Laboratory Manual*, (Cold Springs Harbor Laboratory, NY); D. N. Glover ed., (1985) *DNA Cloning*, Volumes I and II; Gait, ed. (1984) *Oligonucleotide Synthesis*; Mullis *et al.* U.S. Pat. No. 4,683,195; Hames and Higgins, eds. (1984) *Nucleic Acid Hybridization*; Hames and Higgins, eds. (1984) *Transcription And Translation*; Freshney (1987) *Culture Of Animal Cells* (Alan R. Liss, Inc.); Immobilized Cells And Enzymes (IRL Press) (1986); Perbal (1984) *A Practical Guide To Molecular Cloning*; the treatise, *Methods In Enzymology* (Academic Press, Inc., N.Y.); Miller and Calos eds. (1987) *Gene Transfer Vectors For Mammalian Cells*, (Cold Spring Harbor Laboratory); Wu *et al.*, eds., *Methods In Enzymology*, Vols. 154 and 155; Mayer and Walker, eds. (1987) *Immunochemical Methods In Cell And Molecular Biology* (Academic Press, London); Weir and Blackwell, eds., (1986) *Handbook Of Experimental Immunology*, Volumes I-IV; *Manipulating the Mouse Embryo*, Cold Spring Harbor

Laboratory Press, Cold Spring Harbor, N.Y., (1986);); Crooks, Antisense drug Technology: Principles, strategies and applications, 2nd Ed. CRC Press (2007) and in Ausubel *et al.* (1989) Current Protocols in Molecular Biology (John Wiley and Sons, Baltimore, Md.).

- 5 All of the references cited above, as well as all references cited herein and the amino acid or nucleotide sequences (*e.g.*, GenBank numbers and/or Uniprot numbers), are incorporated herein by reference in their entireties.

The following examples are offered by way of illustration and not by way of limitation.

10 EXAMPLES

Example 1: Novel miniaturized dystrophins with reduced immunogenicity and enhanced stability

- 15 Mutations in the dystrophin gene often result in an impairment of the stability of the corresponding dystrophin protein, which in turn leads to proteosomal degradation of the unstable dystrophin protein, and dystrophic pathophysiology. Similarly, miniaturizing dystrophin-encoding DNA to accommodate the limited packaging capacity of AAV can impair the stability of the corresponding miniaturized dystrophin protein.

- 20 The stability of various miniaturized dystrophin proteins was examined by comparing dystrophin protein:mRNA ratios in cells transfected with the corresponding dystrophin expression vectors. Male human isogenic induced-pluripotent stem cell (iPSC)-derived cardiac myocytes (iCMs) were generated that carry an E2035X premature stop codon in the dystrophin gene that prevented endogenous dystrophin expression. These cells were transfected with various miniaturized cassettes expressing miniaturized dystrophin proteins and the dystrophin protein:mRNA ratios were examined after the transfected cells had been
- 25 cultured in vitro for 24 days. The protein levels were examined by a Meso Scale Discovery (MSD) ELISA assay, and the mRNA levels by qrtPCR.

The miniaturized dystrophins tested and the test results are shown in FIG. 2 and FIG. 3, respectively. The data indicated that miniaturized dystrophin peptides BXA-027741 (SEQ ID NO:129) and BXA-027743 (SEQ ID NO:132) provide the best protein stability.

- 30 Subsequently, the immunogenicity of the novel junctions created within these two most stable designs were tested using an in silico immunogenicity prediction tool. The novel junctions of the BXA-027743 design (R1/linkerR16 (junction J1), R17/H3 and H3/R23

(junction J7)) (see FIG. 4 and not shown) were determined to have minimal immunogenic risk, based on the aforesaid in silico approach.

The novel junctions of the BXA-027741 design (R1/R3 (junction J10), H2/linkerR16 (junction J11) and R17/R24 (junction J9)) (see FIG. 5) were analyzed in a similar fashion.

5 The junction J9 was determined to have minimal immunogenic risk, while junctions J10 and J11 could be improved.

The immunogenic potential of the aforesaid junctions of the BXA-027741 design, and junction variants thereof, were tested using an in vitro T cell proliferation assay as described below. Briefly, samples of peripheral blood mononuclear cells (PBMC) were isolated from
10 healthy volunteer human subjects by Ficoll (GE Healthcare) gradient centrifugation and characterized, regarding human lymphocyte antigen (HLA) Class I and II expression, using a combination of polymerase chain reaction (PCR) amplification and hybridization with oligonucleotide probes (ProImmune, Sarasota, FL).

A panel of PBMC samples from 40 donors, having an HLA expression profile closely
15 matching world population frequencies, was used for further analysis. PBMC samples were labeled with CFSE (Invitrogen, Carlsbad, CA) to monitor proliferation and plated in 96 well plates in six replicates at 200,000 cells per well in RPMI (Lonza, Basel, Switzerland) containing 10% human AB (Bioreclamation, Westbury, NY), non-essential amino acids and pen-strep (both Gibco/Fisher Scientific).

20 BXA-027741 junction peptides, variants thereof, and control peptides were each cultured with the panel of 40 PBMC samples at 1 μ M for 7 days, after which the media was washed away and cells were labeled with an anti-CD4 and an anti CD8 APC monoclonal antibody (BD Biosciences, Franklin Lake, NJ). After removal of the unbound antibodies through washing, cells were fixed with 3.7% formalin (Sigma, St. Louis, MO) in PBS and analyzed
25 by flow cytometry to determine the percentage of proliferating CD4⁺ cells or CD8⁺ cells. The percentage of samples (among the 40 donor samples) that showed a positive response after seven days in culture with the different BXA-027741 junction peptides and variants thereof - defined as a significant increase in the number of CD4⁺ or CD8⁺ T proliferating cells compared to PBMC incubated in media without junction peptides or control peptides
30 - is shown in FIG. 6A (CD4⁺) and FIG. 6B (CD8⁺). Control peptides used were: (1) Avastin Framework Peptide; (2) VL6-VL CDR3 Peptide; and (3) PADRE peptide-61309.

It was found that version 3 of junction J10 (J10v3) was superior to junction J10, and other tested versions thereof, and that version 12 of junction J11 (J11v12) was overall superior with respect to immunogenic risk to junction J11.

The C-terminal domain of miniaturized dystrophin BXA-027741 was then deleted to generate miniaturized dystrophin BXA-196473 (SEQ ID NO:119) (FIG. 7) to better accommodate the limited packaging capacity of AAV. J10v3 was then used to replace J10, and J11v12 was used to replace J11, in miniaturized dystrophin BXA-196473, resulting in miniaturized dystrophin BXA-196477 (SEQ ID NO:118) (FIGS. 8A and 8B). To create J10v3 and J11v12, the amino acids RV were inserted between amino acids 446 and 447, and the amino acids SEAQ were inserted between amino acids 606 and 607, respectively, in BXA-196473 (SEQ ID NO:119) (FIG. 8A and 8B).

BXA-196473 (SEQ ID NO:119) was additionally modified with the other junction versions tested as follows:

15 BXA-196474 (SEQ ID NO:120): Junction 9 Version 2 (J9v2) and Junction 11 Version 12 (J11v12)

20 BXA-196475 (SEQ ID NO:121): Junction 9 Version 5 (J9v5) and Junction 11 Version 12 (J11v12)

25 BXA-196476 (SEQ ID NO:122): Junction 9 Version 6 (J9v6) and Junction 11 Version 12 (J11v12)

30 BXA-196478 (SEQ ID NO:124): Junction 10 Version 6 (J10v6) and Junction 11 Version 12 (J11v12)

35 BXA-196479 (SEQ ID NO:125): Junction 11 Version 12 (J11v12)

J9v2 is a substitution of amino acids LER to KNI between the R17 domain and R24 domain, i.e., amino acids 843 to 845 in BXA-196473.

25 J9v5 is an insertion of K between the R17 domain and R24 domain, i.e., between amino acids 842 and 843 in BXA-196473.

J9v6 is an insertion of KNI between the R17 domain and R24 domain, i.e., between amino acids 842 and 843 in BXA-196473.

30 J10v3 is an insertion of RV between the R1 and R3 domain, i.e., between amino acids 446 and 447 in BXA-196473.

J10v6 is an insertion of RVLLQDI between the R1 and R3 domain, i.e., between amino acids 446 and 447 in BXA-196473.

J11v12 is an insertion of SEAQ between the H2 domain and the linker before the R16 domain, i.e., between amino acids 606 and 607 in BXA-196473.

The resulting miniaturized dystrophins were then tested for stability as described before (FIG. 9). The results indicated that the BXA-196477 miniaturized dystrophin design (that
5 contained J10V3 and J11V12) had not only the lowest immunogenic risk, but was also the most stable.

In addition, miniaturized dystrophin polypeptide design BXA-212371 was modified by deleting the last three C-terminal amino acids, and by inserting the amino acid junction variant sequence KNDL (J2V10; between spectrin repeat 17 (R17) domain and hinge 3
10 (H3) domain) after amino acid position 682 of BXA-212371 (resulting in miniaturized dystrophin BXA-213788, SEQ ID NO:152), which was found to have lower immunogenic risk.

Example 2: In Vitro Physiology

15 DMD iCMs have lower Na⁺ channel amplitude, prolonged cFPD (Q-T interval), and greater beat rate variability, when compared to isogenic wild-type iCMs using multi-electrode arrays. DMD iCMs also have a higher Ca²⁺ transduction and lower impedance than wild-type iCMs, with each method confirming the beat rate variability in non-paced cells.

20 It is examined whether expression of miniaturized dystrophin BXA-196477 in DMD iCMs carrying the E2035X mutation can mitigate the DMD phenotype and improve the physiological traits of the cells. Multi-electrode arrays, impedance contraction assay, and Ca²⁺ transients are used to measure the effect of miniaturized dystrophin expression. We also deleted the C-terminal domain of BXA-027743 to generate BXA-212371 to better
25 accommodate the limited packaging capacity of AAV. The functional capacity of BXA-212371 to mitigate the DMD phenotype and improve the physiological traits of the same cells is examined utilizing the same assays proposed for BXA196477.

The hiPSC CMs used in this study were DMD (E2035X) iCells purchased from Cellular
30 Dynamics International, Madison, Wisconsin. Human ventricular fibroblasts were purchased from Lonza, Walkersville, Maryland. iCell hiPSC CMs have been reported to have electrophysiological properties close to primary adult cardiomyocytes and to respond

similarly to a range of cardiac ion channel inhibitors as well as adrenergic and muscarinic receptor agonists and antagonists. Our own work showed that co-culturing hiPSC CMs with fibroblasts provided a more stable preparation for electrophysiologic studies on multi-electrode arrays (MEAs). hiPSC CMs are cultured with 7% CO₂ on 0.1 % gelatin treated 6-well culture plates for 7 days, then trypsinized and diluted with human adult cardiac fibroblasts at about a 5:1 ratio. Suspensions of hiPSC CMs and fibroblasts were then co-cultured on laminin-coated 9-well multi-electrode array (MEA) plates (256-9wellMEA300/30iR-ITO-mq; Multichannel Systems). hiPSC CMs were infected with AAV9-mDys expression constructs (expressing miniaturized dystrophins BXA-196477 and BXA-213788, FIG. 15D) at MOI of 1×10^6 for 48 hrs, and the hiPSC CMs were then evaluated for effects on extracellular field potentials (FPs) 7 days and 9 days after infection. Both BXA-196477 and BXA-213788 significantly improved conduction velocity by ~80% compared to the DMD CMs (Two-way ANOVA ***P<0.001 with Tukey post-test n = 6)(FIG. 15E). A dystrophin ELISA confirmed dystrophin expression in the cells (FIG. 15F).

MEA technology enables high content spatiotemporal analysis of excitable cells or tissues from an array of embedded substrate-integrated extracellular electrodes onto which cells can be cultured or tissues placed. Extracellular field potentials (FPs) are recorded by each electrode and correspond to cellular action potentials. Assessment of FP morphology, duration and conduction velocity provides a picture of ion channel activities of a treatment as well as effects on repolarization and conduction. After 7 days culture of hiPSC CMs on MEA plates, cells formed a spontaneously beating monolayer over recording electrodes imbedded in each well. Spontaneous FPs were recorded from 28 electrodes/well (30um diameter, 300 um center to center spacing) at a sampling frequency of 10 kHz using a USB-MEA256-System and MC Rack acquisition software (Multi Channel Systems). Following a 20-minute equilibration period in a humidified environment at 37°C with constant 5% CO₂ and 95% O₂ supply, each well was transfected with AAV9-mDys expression constructs in 300 µl maintenance medium. Effects of expression of miniaturized Dystrophins on inter-pulse intervals (IPIs) of FPs were monitored and recorded. Data were analyzed with custom software written in MatLab (Mathworks). Beat rate (beats/minute), a surrogate for heart rate, was calculated by using $BR = 60000 / IPI$, where the IPI is the averaged IPIs (msec) of 100 second recording at steady-state under each condition.

Electrophysiologic parameters to be measured are field potential duration, a surrogate for repolarization, and field potential conduction velocity. Field potential duration was corrected for beat rate changes (FPDc). All treatments had at least 3 replicates.

Conduction velocity was quantified by measuring field potential activation times for each electrode imbedded in an MEA well during a synchronized single propagated beat. The digitized recordings of field potentials from each electrode were smoothed using a 21-point least squares smoothing polynomial (Savitsky & Golay, 1964) with a window of 2.1 ms. The activation time is the value for the peak in the negative derivative of each field potential waveform. The time between two of the earliest and latest activation times is the conduction time for field potential propagation across a monolayer of cardiomyocytes and the distance between these two electrodes is the conduction distance. The conduction time divided by the conduction distance of each propagation is the conduction velocity of each beat of the monolayer cardiomyocytes on an MEA well (FIG. 15).

The data show that both miniaturized Dystrophin design significantly improved conduction velocity (FIG. 15).

Example 3: In vivo restoration of skeletal muscle physiology

Dystrophin-deficient skeletal muscles produce less specific force (force per area) and are highly susceptible to contraction-induced injury. Restoration of dystrophin expression can mitigate these disorders. Dystrophic mdx mice are systemically treated with either 2e14vg/kg AAV9-C5-12-BXA-196477 or AAV9-C5-12-BXA-212371 at 2 weeks of age by retro-orbital injection. The limb muscle physiology are examined at 8 weeks of age. Briefly, the mouse knee is clamped and the foot is placed in a stirrup, and the stirrup is moved while the muscles are maximally contracted with a needle electrode. This assay measures the peak twitch and tetanic muscle force production and contraction-induced injury.

Example 4: In vivo prevention of dystrophic pathology

The skeletal muscles in the mdx mouse model of DMD typically undergo necrosis and regeneration from ~3-4 weeks of age. The regenerated muscles are typically more variable in size and contain centrally located nuclei in frozen transverse sections. Also, fibrosis becomes more prevalent in regenerated muscles. The muscle fiber size, proportion of

centrally located nuclei, and fibrosis in untreated mdx muscles and mdx muscles treated with either AAV9-C5-12-BXA-196477 or AAV9-C5-12-BXA-212371 is measured. The proportion of muscle fibers expressing the miniaturized dystrophins is also quantified. The diaphragm, gastrocnemius, tibialis anterior, and cardiac muscles is examined using the HALO and Columbus imaging computer programs.

Example 5: In vivo study

Outline of Study

mdx^{4cv} mice were treated with 2e14vg/kg AAV9-BXA-196477 or AAV9-BXA-213788 at 2 weeks of age. Two-weeks after administration, n = 3 mice were terminated to examine target engagement of miniaturized dystrophin (biodistribution and expression levels). At 2 months of age the right tibialis anterior muscle was examined for strength and resistance to contraction-induced injury (n = 10) similar to previously described (Khairallah et. al., 2012). The mice were terminated at 3 months of age and examined for target engagement and prevention of dystrophy (FIG. 16).

Target Engagement

Immuno-fluorescence - Muscles were frozen in OCT in 2-methylbutane in liquid N₂. 10um frozen sections were immunostained with antibodies to dystrophin (Mandys106; DSHB; 1:200) and wheat-germ agglutinin conjugated to Alexa-647. Alexa-488 IgG2a antibody was used to label the dystrophin primary antibody. Sections were imaged using the Phoenix Opera High Content screening confocal microscope (Perkin-Elmer) and quantified using Columbus software. Briefly, the Alexa-647 images were inverted, M-mode was used to find the cells and calculate the morphology in a select population of cells in transverse sections <2000um². Next, the intensity of the Alexa-488 was calculated and the population with intensity greater than *mdx*^{4cv} control muscles was selected. Greater than 90% of the muscles were positive for miniaturized dystrophin in 4-weeks and 12-weeks mice (FIG. 17).

mRNA - For isolation of total RNA, tissue was homogenized using Qiagen Tissuelyzer and RNA was isolated using a Qiagen RNeasy 96 Universal tissue kit (Qiagen catalog number 74881). Tissue (~15) was placed in RNeasy kit collection microtubes containing 750ul of Qiazol reagent and one 5mm steel bead, homogenized using Tissuelyzer at 30hz for 2 min and repeat this step until homogenized followed by a centrifugation at 6000xg for 1min at

4°C. To each tube 150 µl of chloroform was added and samples vortexed vigorously for 15sec. Following a 3 min incubation at room temperature, samples were spun at 6000xg for 15min at 4°C. The aqueous phase was removed (~360µl) and transferred to a new tube containing 1 volume of RNase free 70% EtOH. All samples were transferred to a 96 well RNeasy 96 plate and plates were sealed with AirPore tape and centrifuge at 5600xg for 4 min at room temperature. 400 µl of RW1 buffer was added per well and plate resealed and spun for 4 min at 5600xg. During spin a DNaseI stock solution was prepared by adding 550 µl of RNase free water per DNase vial. 670 µl of the DNase I stock solution was diluted into 7.3mls RDD buffer, mixed and stored at 4°C. When centrifugation was completed the flow-through was discarded and 80µl of DNase I mix was added directly to center of each well and the plate was incubated at room temperature for 15 min. Following incubation, 400µl of RW1 was added to each well and the plate was sealed and centrifuged for 4 min at 5600xg. Flow-through was discarded and 800µl of RPE buffer was added per well the plate was re-sealed and spun for 4 min at 5600xg. This process was repeated and the plate was centrifuged for 10 min at 5600xg. Each sample was then eluted into a fresh tube by adding 60 µl of RNase free water to center of each well and centrifuging for 4 min at 5600xg. To improve recovery the 60 µL was re-applied back onto plate and centrifuged for an additional 4 min at 5600xg. RNA yield was quantitated using nanodrop 8000.

RT-PCR - For cDNA synthesis and subsequent PCR, 1 µg of RNA was added to 1 well of a 96 well plate in 10 µl H₂O (Plate-Axygen, PCR-96-C-S). To each well 10 µl of master mix (High Capacity cDNA Reverse Transcription kit, Applied Biosystem) was added and plate was centrifuged at 1000 rpm. cDNA synthesis was carried out at 25°C for 10 min, 37°C for 120 min, 85°C for 5 min followed by a hold at 4°C. For qPCR, each sample was run in duplicate with primer/probe sets (forward primer 5'-TGGAAGATTGCTACGAGCGC-3'; reverse primer 5'-CAGGTCGCTGAACAGGTTCT-3'; probe-6FAM-GCAAGTTCGGCAAGCAGCACAMGBNFG) in 384 well clear reaction plates (applied biosystem; catalog number 4483285) To each reaction 2µl of cDNA and 8 µL of master mix (5 µl fast advanced master mix, 0.5 µl 20X FAM primer probe mix and 2.5 µl water) was added and plates were centrifuged for 1 min at 1000 rpm. Samples were incubated at 95°C for 20 sec follow by 40 cycles at 95°C for 1 sec and 60°C for 20 sec using ViiA7 system and Quant Studio real-time PCR software for data analysis (Applied Biosystem).

Vector Genome/Genomic DNA isolation and qPCR – For genomic DNA isolation, tissue was homogenized using Qiagen TissueLyzer and genomic DNA was isolated using a Qiagen DNeasy 96 blood and tissue kit (Qiagen catalog number 69581). Tissue (~10 mg) was placed in 96 well plate (costar assay block 1 ml, catalog number 3958) containing 200 ul of proteinase K-buffer ATL and one 5mm steel bead, homogenized using TissueLyzer 30hz, 2 min, repeat until homogenized. Genomic DNA isolation followed by manufactural instructions. For qPCR, each sample was run in duplicate with primer/probe sets (forward primer 5'-TGGAAGATTGCTACGAGCGC-3'; reverse primer 5'-CAGGTCGCTGAACAGGTTCT-3'; probe-6FAM-GCAAGTTCGGCAAGCAGCACAMGBNFG) in 384 well clear reaction plates (applied biosystem; catalog number 4483285). To each reaction 2 ul of genomic DNA (80 ng) and 8 uL of master mix (5 ul fast advanced master mix, 0.5 ul 20X FAM primer probe mix and 2.5 ul water) was added and plates were centrifuged for 1 min at 1000 rpm. Samples were incubated at 95oC for 20 sec follow by 40 cycles at 95oC for 1 sec and 60oC for 20 sec using ViiA7 system and Quant Studio real-time PCR software for data analysis (Applied Biosystem).

MSD-ELISA - Protein expression was determined by Elisa assay. Multi-assay 384-well plates (Meso Scale Discovery, catalog number L21XB-4) were pre-coated with monoclonal antibody Manex 1011b at concentration of 2 ug/ml in bicarbonate buffer overnight. Plates were then blocked with blocking buffer (5% BSA in PBS) for 4 hrs with shaking at room temperature. Tissue (~20 mg) were homogenized in ripa buffer at concentration of 1mg tissue/10 ul lysis buffer (Sigma, catalog number R0278) with protease inhibitor cocktail tablet (Roche, catalog number 04 693 159 001) using Qiagen TissueLyzer at 30hz for 5 min and repeat this step until homogenized. Tissue ripa lysate were diluted 1 to 3 in binding buffer (1% BSA, 0.05% tween 20, 20 mM Tris pH 7.5 in PBS). Tissue lysates and sulfo-conjugated Mandys 106 (0.2 ug/ml) were added to the pre-coated 384 well plates and incubate at 40°C with shaking overnight. Plates were washed with PBS with 0.05% tween 20 and add 40 ul MSD read buffer T with surfactant (catalog number R92TC-1). Plates were then read on MSD Sector 6000 machine.

The data showed that sufficient virus was administered to achieve expression (mRNA and protein) in >90% striated muscle cells at 4-weeks of age (FIG. 18).

Creatine Kinase - Creatine kinase as an indicator of muscle damage was measured in serum using commercially available kits. The CK was measured at 4 weeks of age (2-weeks post

delivery). Also, the CK was measured at 3 months of age, which was 1 month after the contraction-induced injury protocol (see below). **** $P < 0.001$ one-way ANOVA with Tukey post-test, $n = 10$ compared to mdx^{4cv} . The data showed that in mice treated with BXA-196477 or BXA-213788, creatine kinase levels and thus muscle damage were significantly reduced (FIG. 19).

Functional Studies

The tibialis anterior (TA) muscle contractile properties were tested by an *in vivo* (foot plate) apparatus as per manufacturers instructions (Aurora Scientific). Briefly, the peak tetanic contraction was achieved at 150Hz in force frequency curve (force is measured in Torque as Newton Meters). The peak tetanic contraction was the same in wild-type, mdx^{4cv} and mdx^{4cv} mice treated with BXA-196477 or BXA-213788. However, the TA muscle mass is greater in mdx^{4cv} **** $P < 0.001$ one-way ANOVA with Tukey post-test, $n = 10$), such that normalization of peak tetanic force to TA mass was reduced in mdx^{4cv} mice and at wild-type levels in the treated mice *** $P < 0.001$ compared to mdx^{4cv} ; one-way ANOVA with Tukey post-test, $n = 10$) (FIG. 20).

The tibialis anterior (TA) muscle injury was measured by an *in vivo* (foot plate) apparatus as per manufacturers instructions (Aurora Scientific). During peak tetanic contraction at 150Hz (maximum isometric torque), the foot plate was rotated from 90° to 135° degrees to strain the muscles. This contraction was repeated every minute for 20 contractions as previously described (Khairallah et. al., 2012). The maximum isometric torque immediately prior to strain significantly reduced with each contraction in mdx^{4cv} mice ($P < 0.0001$ compared to wild-type; 2-way ANOVA; $n = 9-10$). In contrast both BXA-196477 and BXA-213788 prevented the contraction-induced injury similar to wild-type levels (**** $P < 0.0001$ BXA-213788 and ** $P < 0.01$ BXA-196477 compared to mdx^{4cv} with Tukey post-test $n = 9-10$). The data show that both miniaturized dystrophins designs protect the TA muscles from contraction-induced injury (FIG. 21).

In vitro an *in vivo* expression of miniaturized dystrophin constructs described in the Examples herein was under the control of a C5-12 promoter (see US 2004/0175727) wherein the first seven and last thirteen nucleotides were removed. The AAV used was AAV9, wherein the ITRs were AAV2.

Example 6: Further Immunogenicity Studies

[0256] The immunogenicity of novel junctions (see Table 12: SEQ ID NOs: 156 to 166; SEQ ID NO:155 was original sequence) created between hinge 2 (H2) domain and R16 spectrin repeat 16 (R16) within miniaturized dystrophin polypeptide BXA-027741 were tested using an in silico immunogenicity prediction tool. The novel junction SEQ ID NO:157 (see FIG. 22) was determined to have minimal immunogenic risk, based on the aforesaid in silico approach. The tested junction sequences are numbered as follows: SEQ ID NO:155 (1), SEQ ID NO:156 (2), SEQ ID NO:157 (3), SEQ ID NO:158 (4), SEQ ID NO:159 (5), SEQ ID NO:160 (6), SEQ ID NO:161 (7), SEQ ID NO:162 (8), SEQ ID NO:163 (9), SEQ ID NO:164 (10), SEQ ID NO:165 (11), SEQ ID NO:166 (12) (see FIG. 22B).

The immunogenic potential of the aforesaid junctions of the BXA-027741 design, and junction variants thereof, were tested using an in vitro T cell proliferation assay as described above. The data indicate that junction J11v3 (SEQ ID NO:157, No. 3 in FIG. 22B) elicited significantly less CD8⁺ (FIG. 23A) and CD4⁺ T (FIG. 23B) cell proliferation compared to other tested junctions and is therefore superior (FIG. 23). Miniaturized dystrophin polypeptide BXA-196477 was modified to carry junction J11v3, the last three amino acids at the C-terminus were deleted, and the 3'UTR was shortened, resulting in miniaturized dystrophin polypeptide BXA-213780 (SEQ ID NO:153) and nucleotide sequence SEQ ID NO:154.