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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 February 2015 (05.02.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/017146 A2

(51) International Patent Classification: Not classified

(21) International Application Number:
PCT/US2014/047287

(22) International Filing Date:
18 July 2014 (18.07.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/847,971 18 July 2013 (18.07.2013) US

(71) Applicant: **FABRUS, INC.** [US/US]; 11099 North Torrey Pines Road, Suite 230, La Jolla, CA 92037 (US).

(72) Inventors: **DE LOS RIOS, Miguel**; 11099 North Torrey Pines Road, Suite 230, La Jolla, CA 92037 (US). **BAZIRGAN, Omar**; 11099 North Torrey Pines Road, Suite 230, La Jolla, CA 92037 (US).

(74) Agents: **MCNICHOLAS, Janet, M.** et al.; Jones Day, 222 East 41st. Street, New York, NY 10017-6702 (US).

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

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

Published:

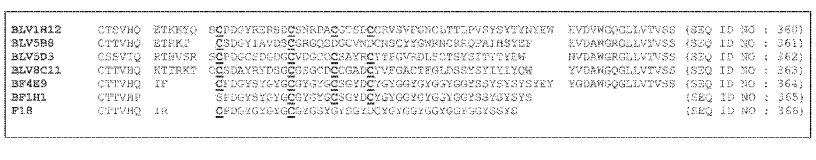
— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2015/017146 A2

(54) Title: ANTIBODIES WITH ULTRALONG COMPLEMENTARITY DETERMINING REGIONS

Figure 1



(57) Abstract: The present disclosure provides antibodies, including antibodies comprising ultralong CDR3 and uses thereof.

ANTIBODIES WITH ULTRALONG COMPLEMENTARITY DETERMINING REGIONS

FIELD

[0001] The present disclosure relates to antibodies, including antibodies comprising an ultralong CDR3.

CROSS REFERENCE

[0002] This application claims the benefit of priority of United States Provisional Application Serial No. 61/847,971, filed July 18, 2013, the entire contents of which is incorporated herein by reference.

[0003] The instant application contains a Sequence Listing which has been submitted in ASCII format via EFS-Web and is hereby incorporated by reference in its entirety. Said ASCII copy, created on July 18, 2014, is named 13379-007-228_SequenceListing.txt and is 475,777 bytes in size.

BACKGROUND

[0004] Antibodies are natural proteins that the vertebrate immune system forms in response to foreign substances (antigens), primarily for defense against infection. For over a century, antibodies have been induced in animals under artificial conditions and harvested for use in therapy or diagnosis of disease conditions, or for biological research. Each individual antibody producing cell produces a single type of antibody with a chemically defined composition, however, antibodies obtained directly from animal serum in response to antigen inoculation actually comprise an ensemble of non-identical molecules (*e.g.*, polyclonal antibodies) made from an ensemble of individual antibody producing cells.

[0005] Some bovine antibodies have unusually long VH CDR3 sequences compared to other vertebrates. For example, about 10% of IgM contains “ultralong” CDR3 sequences, which can be up to 61 amino acids long. These unusual CDR3s often have multiple cysteines. Functional VH genes form through a process called V(D)J recombination, wherein the D-region encodes a significant proportion of CDR3. A unique D-region encoding an ultralong sequence has been identified in cattle. Ultralong CDR3s are partially encoded in the cattle genome, and provide a unique characteristic of their antibody repertoire in comparison to humans. Kaushik *et al.* (U.S. Patent Nos. 6,740,747 and 7,196,185) disclose several bovine germline D-gene sequences unique to cattle stated to be useful as probes and a bovine VDJ cassette stated to be useful as a vaccine vector.

[0006] Human antibodies have heavy chain CDR3 regions that typically vary in size from 8-16 amino acid residues. Human antibodies with longer heavy chain CDR3 regions have been described with HIV-1 neutralization properties. McLellan et al., Nature 48:336-443 (2011); Walker et al., Nature, 477: 466470 (2011). Crystal structures of at least five such human antibodies have been described. These crystal structures indicate that the long CDR3 regions of such antibodies protrude from the antibody Ig fold.

[0007] Recently, the crystal structures of two bovine antibodies with ultralong CDRs have been published. Wang et al., Cell, 153:1379-93 (2013).

SUMMARY

[0008] The present disclosure provides antibody heavy chain variable regions comprising an ultralong CDR3, methods of making same, and uses thereof.

[0009] In some embodiments, the antibody heavy chain variable region comprises one or more human variable region framework sequences. In some aspects, the one or more human variable region framework sequences include the framework sequences of a human HIV-1 neutralizing antibody, such as PGT145, PG9, PG16, CHO4 or 2909.

[0010] In some embodiments, the antibody heavy chain variable region comprises both framework and CDR sequences of a human HIV-1 neutralizing antibody, such as PGT145, PG9, PG16, CHO4 or 2909, except a non-human sequence or a non-antibody sequence (e.g., a non-antibody human sequence) has been inserted into the CDR3 of the antibody, including, in some aspects, removing a portion of CDR3 (e.g., one or more amino acids of the CDR3) or the entire CDR3 (e.g., all or substantially all of the amino acids of the CDR3). In some embodiments, a non-human sequence or the non-antibody sequence (e.g., a non-antibody human sequence) replaces amino acids of the CDR3 antibody sequence.

[0011] In some embodiments, the present disclosure provides an antibody heavy chain variable region comprising a sequence of the formula V1-X-V2, wherein V1 comprises an amino acid sequence selected from the group consisting of:
(i) QVQLVQSGAEVKKPGSSVKVSCKASGNSFSNHDVHWVRQATGQGLEWMGWMSHE GDKTGLAQKFQGRVTITRDSGASTVYMELRGLTADDTAIYYCLT (SEQ ID NO: 496),
(ii) QRLVESGGGVVQPGSSLRLSCAASGFDFSRQGMHWVRQAPGQGLEWVAFIKYDGS EKYHADSVWGRLSISRDNSKDTLYLQMNSLRVEDTATYFCVR (SEQ ID NO: 497),
(iii) QEQLVESGGGVVQPGGSLRLSCLASGFTFHKYGMHWVRQAPGKGLEWVALISDDG MRKYHSDSMWGRVTISRDNSKNTLYLQFSSLKVEDTAMFFCAR (SEQ ID NO: 498),
(iv) EVQLVESGGGLIRPGGSLRLSCKGSGFIFENFGFGWVRQGPGKGLEWVSGTNWNG GDSRYGDSVKGRFTISRDNSNNFVYLQMNSLRPEDTAIYYCAR (SEQ ID NO: 499), and
(v) EVQLVESGGNVVQPGGSLRLSCTASGFSFDDSTMHWVRQAPGKGLQWVSLISWNG

GRTYYADSVKGRFTISRDNSKNSLYLQMNSLKTEDTAFYFCAK (SEQ ID NO: 500); wherein X comprises an ultralong CDR3, which can include a non-human sequence or a non-antibody sequence (e.g., a non-antibody human sequence) that has been inserted into the CDR3 sequence of the antibody, including optionally, removing a portion of CDR3 (e.g., one or more amino acids of the CDR3) or the entire CDR3 sequence (e.g., all or substantially all of the amino acids of the CDR3); and wherein V2 comprises an amino acid sequence selected from the group consisting of: (i) WGHGTAVTVSS (SEQ ID NO: 570), (ii) WGKGTTVTVSS (SEQ ID NO: 571), (iii) WGKGTTVTVSS (SEQ ID NO: 572), (iv) WGRGTLVTVSS (SEQ ID NO: 573), and (v) WGKGTTVTVSS (SEQ ID NO: 574) (*see*, Figure 6C).

[0012] In some embodiments of each or any of the above or below mentioned embodiments, the ultralong CDR3 comprises an amino acid sequence of: (i) any one of GSKHRLRDYFLYNE (SEQ ID NO: 501), GSKHRLRDYFLYN (SEQ ID NO: 502), GSKHRLRDYFLY (SEQ ID NO: 503), GSKHRLRDYFL (SEQ ID NO: 504), GSKHRLRDYF (SEQ ID NO: 505), GSKHRLRDY (SEQ ID NO: 506), or GSKHRLRD (SEQ ID NO: 507); (ii) any one of EAGGPDYRNGYNY (SEQ ID NO: 508), EAGGPDYRNGYN (SEQ ID NO: 509), EAGGPDYRNGY (SEQ ID NO: 510), EAGGPDYRNG (SEQ ID NO: 511), EAGGPDYRN (SEQ ID NO: 512), EAGGPDYR (SEQ ID NO: 513), EAGGPDY (SEQ ID NO: 514), or EAGGPD (SEQ ID NO: 515); (iii) any one of EAGGPIWHDDVKY (SEQ ID NO: 516), EAGGPIWHDDVK (SEQ ID NO: 517), EAGGPIWHDDV (SEQ ID NO: 518), EAGGPIWHDD (SEQ ID NO: 519), EAGGPIWHD (SEQ ID NO: 520), EAGGPIWH (SEQ ID NO: 521), EAGGPIW (SEQ ID NO: 522), or EAGGPI (SEQ ID NO: 523); (iv) any one of GTDYTIDDQGI (SEQ ID NO: 524), GTDYTIDDQG (SEQ ID NO: 525), GTDYTIDDQ (SEQ ID NO: 526), GTDYTIDD (SEQ ID NO: 527), GTDYTID (SEQ ID NO: 528), or GTDYTI (SEQ ID NO: 529); or (v) any one of DKGDSDYDYNL (SEQ ID NO: 530), DKGDSDYDYN (SEQ ID NO: 531), DKGDSDYDY (SEQ ID NO: 532), DKGDSDYD (SEQ ID NO: 533), DKGDSDY (SEQ ID NO: 534), DKGDSD (SEQ ID NO: 535) (*see*, Figure 6C).

[0013] In some embodiments of each or any of the above or below mentioned embodiments, the ultralong CDR3 comprises an amino acid sequence of: (i) any one of YGPNYEEWGDYLATLDV (SEQ ID NO: 536), GPNYEEWGDYLATLDV (SEQ ID NO: 537), PNYEEWGDYLATLDV (SEQ ID NO: 538), NYEEWGDYLATLDV (SEQ ID NO: 539), YEEWGDYLATLDV (SEQ ID NO: 540), or EEWGDYLATLDV (SEQ ID NO: 541); (ii) any one of YDFYDGYYNYHYMDV (SEQ ID NO: 542), DFYDGYYNYHYMDV (SEQ ID NO: 543), FYDGYYNYHYMDV (SEQ ID NO: 544), YDGYYNYHYMDV (SEQ ID NO: 545), DGYYNYHYMDV (SEQ ID NO: 546), GYYNYHYMDV (SEQ ID NO: 547), or YYNYHYMDV (SEQ ID NO: 548); (iii) any one of YDFNDGYYNYHYMDV (SEQ ID NO: 549),

DFYDGYYNYHYMDV (SEQ ID NO: 550), FYDGYYNYHYMDV (SEQ ID NO: 551), YDGYYNYHYMDV (SEQ ID NO: 552), DGYYNYHYMDV (SEQ ID NO: 553), or GYYNYHYMDV (SEQ ID NO: 554); (iv) any one of QGIRYQGSGTFWYFDV (SEQ ID NO: 555), GIRYQGSGTFWYFDV (SEQ ID NO: 556), IRYQGSGTFWYFDV (SEQ ID NO: 557), RYQGSGTFWYFDV (SEQ ID NO: 558), YQGSGTFWYFDV (SEQ ID NO: 559), QGSGTFWYFDV (SEQ ID NO: 560), GSGTFWYFDV (SEQ ID NO: 561), SGTFWYFDV (SEQ ID NO: 562), or GTFWYFDV (SEQ ID NO: 563); or (v) any one of YNLGYSYFYYMDG(SEQ ID NO: 564), NLGYSYFYYMDG (SEQ ID NO: 565), LGYSYFYYMDG (SEQ ID NO: 566), GYSYFYYMDG (SEQ ID NO: 567), YSYFYYMDG (SEQ ID NO: 568), or SYFYYMDG (SEQ ID NO: 569) (*see*, Figure 6C).

[0014] In some embodiments of each or any of the above or below mentioned embodiments, the ultralong CDR3 comprises an amino acid sequence of: (i) any one of GSKHRLRDYFLYNE (SEQ ID NO: 501), GSKHRLRDYFLYN (SEQ ID NO: 502), GSKHRLRDYFLY (SEQ ID NO: 503), GSKHRLRDYFL (SEQ ID NO: 504), GSKHRLRDYF (SEQ ID NO: 505), GSKHRLRDY (SEQ ID NO: 506), or GSKHRLRD (SEQ ID NO: 507), and any one of YGPNYEEWGDYLATLDV (SEQ ID NO: 536), GPNYEEWGDYLATLDV (SEQ ID NO: 537), PNYEEWGDYLATLDV (SEQ ID NO: 538), NYEEWGDYLATLDV (SEQ ID NO: 539), YEEWGDYLATLDV (SEQ ID NO: 540), or EEWGDYLATLDV (SEQ ID NO: 541); (ii) any one of EAGGPDYRNGYNY (SEQ ID NO: 508), EAGGPDYRNGYN (SEQ ID NO: 509), EAGGPDYRNGY (SEQ ID NO: 510), EAGGPDYRNG (SEQ ID NO: 511), EAGGPDYRN (SEQ ID NO: 512), EAGGPDYR (SEQ ID NO: 513), EAGGPDY (SEQ ID NO: 514), or EAGGPD (SEQ ID NO: 515), and any one of YDFYDGYYNYHYMDV (SEQ ID NO: 542), DFYDGYYNYHYMDV (SEQ ID NO: 543), FYDGYYNYHYMDV (SEQ ID NO: 544), YDGYYNYHYMDV (SEQ ID NO: 545), DGYYNYHYMDV (SEQ ID NO: 546), GYYNYHYMDV (SEQ ID NO: 547), or YYNYHYMDV (SEQ ID NO: 548); (iii) any one of EAGGPIWHDDVKY (SEQ ID NO: 516), EAGGPIWHDDVK (SEQ ID NO: 517), EAGGPIWHDDV (SEQ ID NO: 518), EAGGPIWHDD (SEQ ID NO: 519), EAGGPIWHD (SEQ ID NO: 520), EAGGPIWH (SEQ ID NO: 521), EAGGPIW (SEQ ID NO: 522), or EAGGPI (SEQ ID NO: 523), and any one of YDFNDGYYNYHYMDV (SEQ ID NO: 549), DFYDGYYNYHYMDV (SEQ ID NO: 550), FYDGYYNYHYMDV (SEQ ID NO: 551), YDGYYNYHYMDV (SEQ ID NO: 552), DGYYNYHYMDV (SEQ ID NO: 553), or GYYNYHYMDV (SEQ ID NO: 554); (iv) any one of GTDYTIDDQGI (SEQ ID NO: 524), GTDYTIDDQG (SEQ ID NO: 525), GTDYTIDDQ (SEQ ID NO: 526), GTDYTIDD (SEQ ID NO: 527), GTDYTID (SEQ ID NO: 528), or GTDYTI (SEQ ID NO: 529), and any one of QGIRYQGSGTFWYFDV (SEQ ID NO: 555), GIRYQGSGTFWYFDV (SEQ ID NO: 556), IRYQGSGTFWYFDV (SEQ ID NO: 557), RYQGSGTFWYFDV (SEQ ID NO: 558),

YQGSGTFWYFDV (SEQ ID NO: 559), QGSGTFWYFDV (SEQ ID NO: 560), GSGTFWYFDV (SEQ ID NO: 561), SGTFWYFDV (SEQ ID NO: 562), or GTFWYFDV (SEQ ID NO: 563); or (v) any one of YNLGYSYFYYMDG(SEQ ID NO: 564), NLGYSYFYYMDG (SEQ ID NO: 565), LGYSYFYYMDG (SEQ ID NO: 566), GYSYFYYMDG (SEQ ID NO: 567), YSYFYYMDG (SEQ ID NO: 568), or SYFYYMDG (SEQ ID NO: 569), and any one of YNLGYSYFYYMDG(SEQ ID NO: 564), NLGYSYFYYMDG (SEQ ID NO: 565), LGYSYFYYMDG (SEQ ID NO: 566), GYSYFYYMDG (SEQ ID NO: 567), YSYFYYMDG (SEQ ID NO: 568), or SYFYYMDG (SEQ ID NO: 569) (*see*, Figure 6C).

[0015] In some embodiments of each or any of the above or below mentioned embodiments, the antibody heavy chain variable region includes wherein V1 comprises an amino acid sequence of QVQLVQSGAEVKKPGSSVKVSCKASGNSFSNHDVHWVRQATG QGLEWMGWMSHEGDKTGLAQKFQGRVTITRDSGASTVYMELRGLTADDTAIYYCLT (SEQ ID NO: 496), wherein the ultralong CDR3 comprises an amino acid sequence of any one of GSKHRLRDYFLYNE (SEQ ID NO: 501), GSKHRLRDYFLYN (SEQ ID NO: 502), GSKHRLRDYFLY (SEQ ID NO: 503), GSKHRLRDYFL (SEQ ID NO: 504), GSKHRLRDYF (SEQ ID NO: 505), GSKHRLRDY (SEQ ID NO: 506), or GSKHRLRD (SEQ ID NO: 507), and an amino acid sequence of any one of YGPNYEEWGDYLATLDV (SEQ ID NO: 536), GPNYEEWGDYLATLDV (SEQ ID NO: 537), PNYEEWGDYLATLDV (SEQ ID NO: 538), NYEEWGDYLATLDV (SEQ ID NO: 539), YEEWGDYLATLDV (SEQ ID NO: 540), or EEWGDYLATLDV (SEQ ID NO: 541), and wherein V2 comprises an amino acid sequence selected of WGHGTAVTVSS (SEQ ID NO: 570) (*see*, Figure 6C).

[0016] In some embodiments of each or any of the above or below mentioned embodiments, the antibody heavy chain variable region includes wherein V1 comprises an amino acid sequence of QRLVESGGGVVQPGSSLRLSCAASGFDFSRQGMHWVRQAPG QGLEWVAFIKYDGSEKYHADSVWGRLSISRDNSKDTLYLQMNSLRVEDTATYFCVR (SEQ ID NO: 497), wherein the ultralong CDR3 comprises an amino acid sequence of any one of EAGGPDYRNGYNY (SEQ ID NO: 508), EAGGPDYRNGYN (SEQ ID NO: 509), EAGGPDYRNGY (SEQ ID NO: 510), EAGGPDYRNG (SEQ ID NO: 511), EAGGPDYRN (SEQ ID NO: 512), EAGGPDYR (SEQ ID NO: 513), EAGGPDY (SEQ ID NO: 514), or EAGGPD (SEQ ID NO: 515), and an amino acid sequence of any one of YDFYDGYYNYHYMDV (SEQ ID NO: 542), DFYDGYYNYHYMDV (SEQ ID NO: 543), FYDGYYNYHYMDV (SEQ ID NO: 544), YDGYYNYHYMDV (SEQ ID NO: 545), DGYYNYHYMDV (SEQ ID NO: 546), GYYNYHYMDV (SEQ ID NO: 547), or YYNYHYMDV (SEQ ID NO: 548), and wherein V2 comprises an amino acid sequence selected of WGKGTTVTVSS (SEQ ID NO: 571) (*see*, Figure 6C).

[0017] In some embodiments of each or any of the above or below mentioned embodiments, the antibody heavy chain variable region includes wherein V1 comprises an amino acid sequence of QEQLVESGGGVVQPGGSLRLSCLASGFTFHKYGMHWVRQAPG KGLEWVALISDDGMRKYHSDSMWGRVTISRDNSKNTLYLQFSSLKVEDTAMFFCAR (SEQ ID NO: 498), wherein the ultralong CDR3 comprises an amino acid sequence of any one of EAGGPIWHDDVKY (SEQ ID NO: 516), EAGGPIWHDDVK (SEQ ID NO: 517), EAGGPIWHDDV (SEQ ID NO: 518), EAGGPIWHDD (SEQ ID NO: 519), EAGGPIWHD (SEQ ID NO: 520), EAGGPIWH (SEQ ID NO: 521), EAGGPIW (SEQ ID NO: 522), or EAGGPI (SEQ ID NO: 523), and an amino acid sequence of any one of YDFNDGYYNYHYMDV (SEQ ID NO: 549), DFYDGYYNYHYMDV (SEQ ID NO: 550), FYDGYYNYHYMDV (SEQ ID NO: 551), YDGYYNYHYMDV (SEQ ID NO: 552), DGYYNYHYMDV (SEQ ID NO: 553), or GYYNYHYMDV (SEQ ID NO: 554) , and wherein V2 comprises an amino acid sequence selected of WGKGTTVTVSS (SEQ ID NO: 572) (*see*, Figure 6C).

[0018] In some embodiments of each or any of the above or below mentioned embodiments, the antibody heavy chain variable region includes wherein V1 comprises an amino acid sequence of EVQLVESGGGLIRPGGSLRLSCKGSGFIFENFGFGWVRQGPGK GLEWVSGTNWNGGDSRYGDSVKGRFTISRDNSNNFVYLQMNSLRPEDTAIYYCAR (SEQ ID NO: 499), wherein the ultralong CDR3 comprises an amino acid sequence of any one of GTDYTIDDQGI (SEQ ID NO: 524), GTDYTIDDQG (SEQ ID NO: 525), GTDYTIDDQ (SEQ ID NO: 526), GTDYTIDD (SEQ ID NO: 527), GTDYTID (SEQ ID NO: 528), or GTDYTI (SEQ ID NO: 529), and an amino acid sequence of any one of QGIRYQGSGTFWYFDV (SEQ ID NO: 555), GIRYQGSGTFWYFDV (SEQ ID NO: 556), IRYQGSGTFWYFDV (SEQ ID NO: 557), RYQGSGTFWYFDV (SEQ ID NO: 558), YQGSGTFWYFDV (SEQ ID NO: 559), QGSGTFWYFDV (SEQ ID NO: 560), GSGTFWYFDV (SEQ ID NO: 561), SGTFWYFDV (SEQ ID NO: 562), or GTFWYFDV (SEQ ID NO: 563), and wherein V2 comprises an amino acid sequence selected of WGRGTLVTVSS (SEQ ID NO: 573) (see, Figure 6C).

[0019] In some embodiments of each or any of the above or below mentioned embodiments, the antibody heavy chain variable region includes wherein V1 comprises an amino acid sequence of EVQLVESGGNVVQPGGSLRLSCTASGFSFDDSTMHWVRQAPG KGLQWVSLISWNGGRTYYADSVKGRFTISRDNSKNSLYLQMNSLKTEDTAFYFCAK (SEQ ID NO: 500), wherein the ultralong CDR3 comprises an amino acid sequence of any one of YNLGYSYFYYMDG(SEQ ID NO: 564), NLGYSYFYYMDG (SEQ ID NO: 565), LGYSYFYYMDG (SEQ ID NO: 566), GYSYFYYMDG (SEQ ID NO: 567), YSYFYYMDG (SEQ ID NO: 568), or SYFYYMDG (SEQ ID NO: 569), and any one of

YNLGYSYFYYMDG(SEQ ID NO: 564), NLGYSYFYYMDG (SEQ ID NO: 565), LGYSYFYYMDG (SEQ ID NO: 566), GYSYFYYMDG (SEQ ID NO: 567), YSYFYYMDG (SEQ ID NO: 568), or SYFYYMDG (SEQ ID NO: 569), wherein V2 comprises an amino acid sequence selected of WGKGTTVTVSS (SEQ ID NO: 574) (*see*, Figure 6C).

[0020] In some embodiments of each or any of the above or below mentioned embodiments, the ultralong CDR3 comprises an amino acid sequence selected from the group consisting of: (i) GSKHRLRDYFL (SEQ ID NO: 504) and YEEWGDYLATLDV (SEQ ID NO: 540); (ii) GTDYTID (SEQ ID NO: 528) and GIRYQGSGTFWYFDV (SEQ ID NO: 556); and (iii) DKGDSDYD (SEQ ID NO: 533) and GYSYFYYMDG (SEQ ID NO: 567) (*see*, Figure 6C).

[0021] In some embodiments of each or any of the above or below mentioned embodiments, the ultralong CDR3 is 35 amino acids in length or longer, 40 amino acids in length or longer, 45 amino acids in length or longer, 50 amino acids in length or longer, 55 amino acids in length or longer, or 60 amino acids in length or longer.

[0022] In some embodiments of each or any of the above or below mentioned embodiments, the ultralong CDR3 is 35 amino acids in length or longer.

[0023] In some embodiments of each or any of the above or below mentioned embodiments, the ultralong CDR3 comprises a cysteine motif.

[0024] In some embodiments of each or any of the above or below mentioned embodiments, the antibody heavy chain variable region includes wherein V1 comprises an amino acid sequence of QVQLVQSGAEVKKPGSSVKVSCKASGNSFSNHDVHWVRQATG QGLEWMGWMSHEGDKTGLAQKFQGRVTITRDSGASTVYMELRGLTADDTAIYYCLT (SEQ ID NO: 496), wherein the ultralong CDR3 comprises an amino acid sequence of any one of GSKHRLRDYFLYNE (SEQ ID NO: 501), GSKHRLRDYFLYN (SEQ ID NO: 502), GSKHRLRDYFLY (SEQ ID NO: 503), GSKHRLRDYFL (SEQ ID NO: 504), GSKHRLRDYF (SEQ ID NO: 505), GSKHRLRDY (SEQ ID NO: 506), or GSKHRLRD (SEQ ID NO: 507), a cysteine motif, and an amino acid sequence of any one of YGPNYEEWGDYLATLDV (SEQ ID NO: 536), GPNYEEWGDYLATLDV (SEQ ID NO: 537), PNYEEWGDYLATLDV (SEQ ID NO: 538), NYEEWGDYLATLDV (SEQ ID NO: 539), YEEWGDYLATLDV (SEQ ID NO: 540), or EEWGDYLATLDV (SEQ ID NO: 541), and wherein V2 comprises an amino acid sequence of WGHGTAVTVSS (SEQ ID NO: 570) (see, Figure 6C).

[0025] In some embodiments of each or any of the above or below mentioned embodiments, the antibody heavy chain variable region includes wherein V1 comprises an amino acid sequence of QRLVESGGGVVQPGSSLRLSCAASGFDFSRQGMHWVRQAPG QGLEWVAFIKYDGSEKYHADSVWGRLSISRDNSKDTLYLQMNSLRVEDTATYFCVR (SEQ ID NO: 497), wherein the ultralong CDR3 comprises an amino acid sequence of any

one of EAGGPDYRNGYNY (SEQ ID NO: 508), EAGGPDYRNGYN (SEQ ID NO: 509), EAGGPDYRNGY (SEQ ID NO: 510), EAGGPDYRNG (SEQ ID NO: 511), EAGGPDYRN (SEQ ID NO: 512), EAGGPDYR (SEQ ID NO: 513), EAGGPDY (SEQ ID NO: 514), or EAGGPD (SEQ ID NO: 515), a cysteine motif, and an amino acid sequence of any one of YDFYDGYYNYHYMDV (SEQ ID NO: 542), DFYDGYYNYHYMDV (SEQ ID NO: 543), FYDGYYNYHYMDV (SEQ ID NO: 544), YDGYYNYHYMDV (SEQ ID NO: 545), DGYYNYHYMDV (SEQ ID NO: 546), GYYNYHYMDV (SEQ ID NO: 547), or YYNYHYMDV (SEQ ID NO: 548), and wherein V2 comprises an amino acid sequence selected of WGKGTTVTVSS (SEQ ID NO: 571) (*see*, Figure 6C).

[0026] In some embodiments of each or any of the above or below mentioned embodiments, the antibody heavy chain variable region includes wherein V1 comprises an amino acid sequence of QEQLVESGGGVVQPGGSLRLSCLASGFTFHKYGMHWVRQAPG KGLEWVALISDDGMRKYHSDSMWGRVTISRDNSKNTLYLQFSSLKVEDTAMFFCAR (SEQ ID NO: 498), wherein the ultralong CDR3 comprises an amino acid sequence of any one of EAGGPIWHDDVKY (SEQ ID NO: 516), EAGGPIWHDDVK (SEQ ID NO: 517), EAGGPIWHDDV (SEQ ID NO: 518), EAGGPIWHDD (SEQ ID NO: 519), EAGGPIWHD (SEQ ID NO: 520), EAGGPIWH (SEQ ID NO: 521), EAGGPIW (SEQ ID NO: 522), or EAGGPI (SEQ ID NO: 523), a cysteine motif, and an amino acid sequence of any one of YDFNDGYYNYHYMDV (SEQ ID NO: 549), DFYDGYYNYHYMDV (SEQ ID NO: 550), FYDGYYNYHYMDV (SEQ ID NO: 551), YDGYYNYHYMDV (SEQ ID NO: 552), DGYYNYHYMDV (SEQ ID NO: 553), or GYYNYHYMDV (SEQ ID NO: 554), and wherein V2 comprises an amino acid sequence selected of WGKGTTVTVSS (SEQ ID NO: 572) (*see*, Figure 6C).

[0027] In some embodiments of each or any of the above or below mentioned embodiments, the antibody heavy chain variable region includes wherein V1 comprises an amino acid sequence of EVQLVESGGGLIRPGGSLRLSCKGSGFIFENFGFGWVRQGPGK GLEWVSGTNWNGGDSRYGDSVKGRFTISRDNSNNFVYLQMNSLRPEDTAIYYCAR (SEQ ID NO: 499), wherein the ultralong CDR3 comprises an amino acid sequence of any one of GTDYTIDDQGI (SEQ ID NO: 524), GTDYTIDDQG (SEQ ID NO: 525), GTDYTIDDQ (SEQ ID NO: 526), GTDYTIDD (SEQ ID NO: 527), GTDYTID (SEQ ID NO: 528), or GTDYTI (SEQ ID NO: 529), a cysteine motif, and an amino acid sequence of any one of QGIRYQGSGTFWYFDV (SEQ ID NO: 555), GIRYQGSGTFWYFDV (SEQ ID NO: 556), IRYQGSGTFWYFDV (SEQ ID NO: 557), RYQGSGTFWYFDV (SEQ ID NO: 558), YQGSGTFWYFDV (SEQ ID NO: 559), QGSGTFWYFDV (SEQ ID NO: 560), GSGTFWYFDV (SEQ ID NO: 561), SGTFWYFDV (SEQ ID NO: 562), or GTFWYFDV

(SEQ ID NO: 563), wherein V2 comprises an amino acid sequence selected of WGRGTLVTVSS (SEQ ID NO: 573) (*see*, Figure 6C).

[0028] In some embodiments of each or any of the above or below mentioned embodiments, the antibody heavy chain variable region includes wherein V1 comprises an amino acid sequence of EVQLVESGGNVVQPGGSLRLSCTASGFSFDDSTMHWVRQAPG KGLQWVSLISWNGGRTYYADSVKGRFTISRDNSKNSLYLQMNSLKTEDTAFYFCAK (SEQ ID NO: 500), wherein the ultralong CDR3 comprises an amino acid sequence of any one of YNLGYSYFYYMDG(SEQ ID NO: 564), NLGYSYFYYMDG (SEQ ID NO: 565), LGYSYFYYMDG (SEQ ID NO: 566), GYSYFYYMDG (SEQ ID NO: 567), YSYFYYMDG (SEQ ID NO: 568), or SYFYYMDG (SEQ ID NO: 569), a cysteine motif, and an amino acid sequence of any one of YNLGYSYFYYMDG(SEQ ID NO: 564), NLGYSYFYYMDG (SEQ ID NO: 565), LGYSYFYYMDG (SEQ ID NO: 566), GYSYFYYMDG (SEQ ID NO: 567), YSYFYYMDG (SEQ ID NO: 568), or SYFYYMDG (SEQ ID NO: 569), wherein V2 comprises an amino acid sequence selected of WGKGTTVTVSS (SEQ ID NO: 574) (*see*, Figure 6C).

[0029] In some embodiments of each or any of the above or below mentioned embodiments, the cysteine motif is selected from the group consisting of: $CX_{10}CX_5CX_5CXCX_7C$ (SEQ ID NO: 41), $CX_{10}CX_6CX_5CXCX_{15}C$ (SEQ ID NO: 42), $CX_{11}CXCX_5C$ (SEQ ID NO: 43), $CX_{11}CX_5CX_5CXCX_7C$ (SEQ ID NO: 44), $CX_{10}CX_6CX_5CXCX_{13}C$ (SEQ ID NO: 45), $CX_{10}CX_5CXCX_4CX_8C$ (SEQ ID NO: 46), $CX_{10}CX_6CX_6CXCX_7C$ (SEQ ID NO: 47), $CX_{10}CX_4CX_7CXCX_8C$ (SEQ ID NO: 48), $CX_{10}CX_4CX_7CXCX_7C$ (SEQ ID NO: 49), $CX_{13}CX_8CX_8C$ (SEQ ID NO: 50), $CX_{10}CX_6CX_5CXCX_7C$ (SEQ ID NO: 51), $CX_{10}CX_5CX_5C$ (SEQ ID NO: 52), $CX_{10}CX_5CX_6CXCX_7C$ (SEQ ID NO: 53), $CX_{10}CX_6CX_5CX_7CX_9C$ (SEQ ID NO: 54), $CX_9CX_7CX_5CXCX_7C$ (SEQ ID NO: 55), $CX_{10}CX_6CX_5CXCX_9C$ (SEQ ID NO: 56), $CX_{10}CXCX_4CX_5CX_{11}C$ (SEQ ID NO: 57), $CX_7CX_3CX_6CX_5CXCX_5CX_{10}C$ (SEQ ID NO: 58), $CX_{10}CXCX_4CX_5CXCX_2CX_3C$ (SEQ ID NO: 59), $CX_{16}CX_5CXC$ (SEQ ID NO: 60), $CX_6CX_4CXCX_4CX_5C$ (SEQ ID NO: 61), $CX_{11}CX_4CX_5CX_6CX_3C$ (SEQ ID NO: 62), $CX_8CX_2CX_6CX_5C$ (SEQ ID NO: 63), $CX_{10}CX_5CX_5CXCX_{10}C$ (SEQ ID NO: 64), $CX_{10}CXCX_6CX_4CXC$ (SEQ ID NO: 65), $CX_{10}CX_5CX_5CXCX_2C$ (SEQ ID NO: 66), $CX_{14}CX_2CX_3CXCXC$ (SEQ ID NO: 67), $CX_{15}CX_5CXC$ (SEQ ID NO: 68), $CX_4CX_6CX_9CX_2CX_{11}C$ (SEQ ID NO: 69), $CX_6CX_4CX_5CX_5CX_{12}C$ (SEQ ID NO: 70), $CX_7CX_3CXCXCX_4CX_5CX_9C$ (SEQ ID NO: 71), $CX_{10}CX_6CX_5C$ (SEQ ID NO: 72), $CX_7CX_3CX_5CX_5CX_9C$ (SEQ ID NO: 73), $CX_7CX_5CXCX_2C$ (SEQ ID NO: 74), $CX_{10}CXCX_6C$ (SEQ ID NO: 75), $CX_{10}CX_3CX_3CX_5CX_7CXCX_6C$ (SEQ ID NO: 76), $CX_{10}CX_4CX_5CX_{12}CX_2C$ (SEQ ID NO: 77), $CX_{12}CX_4CX_5CXCXCX_9CX_3C$ (SEQ ID NO: 78), $CX_{12}CX_4CX_5CX_{12}CX_2C$

(SEQ ID NO: 79), $CX_{10}CX_6CX_5CXCX_{11}C$ (SEQ ID NO: 80), $CX_{16}CX_5CXCXCX_{14}C$ (SEQ ID NO: 81), $CX_{10}CX_5CXCX_8CX_6C$ (SEQ ID NO: 82), $CX_{12}CX_4CX_5CX_8CX_2C$ (SEQ ID NO: 83), $CX_{12}CX_5CX_5CXCX_8C$ (SEQ ID NO: 84), $CX_{10}CX_6CX_5CXCX_4CXCX_9C$ (SEQ ID NO: 85), $CX_{11}CX_4CX_5CX_8CX_2C$ (SEQ ID NO: 86), $CX_{10}CX_6CX_5CX_8CX_2C$ (SEQ ID NO: 87), $CX_{10}CX_6CX_5CXCX_8C$ (SEQ ID NO: 88), $CX_{10}CX_6CX_5CXCX_3CX_8CX_2C$ (SEQ ID NO: 89), $CX_{10}CX_6CX_5CX_3CX_8C$ (SEQ ID NO: 90), $CX_{10}CX_6CX_5CXCX_2CX_6CX_5C$ (SEQ ID NO: 91), $CX_7CX_6CX_3CX_3CX_9C$ (SEQ ID NO: 92), $CX_9CX_8CX_5CX_6CX_5C$ (SEQ ID NO: 93), $CX_{10}CX_2CX_2CX_7CXCX_{11}CX_5C$ (SEQ ID NO: 94), and $CX_{10}CX_6CX_5CXCX_2CX_8CX_4C$ (SEQ ID NO: 95).

[0030] In some embodiments of each or any of the above or below mentioned embodiments, the cysteine motif is selected from the group consisting of: $CCX_3CXCX_3CX_2CCXCX_5CX_9CX_5CXC$ (SEQ ID NO: 96), $CX_6CX_2CX_5CX_4CCXCX_4CX_6CXC$ (SEQ ID NO: 97), $CX_7CXCX_5CX_4CCCX_4CX_6CXC$ (SEQ ID NO: 98), $CX_9CX_3CXCX_2CXCCCX_6CX_4C$ (SEQ ID NO: 99), $CX_5CX_3CXCX_4CX_4CCX_{10}CX_2CC$ (SEQ ID NO: 100), $CX_5CXCX_1CXCX_3CCX_3CX_4CX_{10}C$ (SEQ ID NO: 101), $CX_9CCCX_3CX_4CCCX_5CX_6C$ (SEQ ID NO: 102), $CCX_8CX_5CX_4CX_3CX_4CCXCX_1C$ (SEQ ID NO: 103), $CCX_6CCX_5CCCX_4CX_4CX_{12}C$ (SEQ ID NO: 104), $CX_6CX_2CX_3CCCX_4CX_5CX_3CX_3C$ (SEQ ID NO: 105), $CX_3CX_5CX_6CX_4CCXCX_5CX_4CXC$ (SEQ ID NO: 106), $CX_4CX_4CCX_4CX_4CXCX_{11}CX_2CXC$ (SEQ ID NO: 107), $CX_5CX_2CCX_5CX_4CCX_3CCX_7C$ (SEQ ID NO: 108), $CX_5CX_5CX_3CX_2CXCCX_4CX_7CXC$ (SEQ ID NO: 109), $CX_3CX_7CX_3CX_4CCXCX_2CX_5CX_2C$ (SEQ ID NO: 110), $CX_9CX_3CXCX_4CCX_5CCCX_6C$ (SEQ ID NO: 111), $CX_9CX_3CXCX_2CXCCX_6CX_3CX_3C$ (SEQ ID NO: 112), $CX_8CCXCX_3CCX_3CXCX_3CX_4C$ (SEQ ID NO: 113), $CX_9CCX_4CX_2CXCCXCX_4CX_3C$ (SEQ ID NO: 114), $CX_{10}CXCX_3CX_2CXCCX_4CX_5CXC$ (SEQ ID NO: 115), $CX_9CXCX_3CX_2CXCCX_4CX_5CXC$ (SEQ ID NO: 116), $CX_6CCXCX_5CX_4CCXCX_5CX_2C$ (SEQ ID NO: 117), $CX_6CCXCX_3CXCCX_3CX_4CC$ (SEQ ID NO: 118), $CX_6CCXCX_3CXCX_2CXCX_4CX_8C$ (SEQ ID NO: 119), $CX_4CX_2CCX_3CXCX_4CCX_2CX_3C$ (SEQ ID NO: 120), $CX_3CX_5CX_3CCCX_4CX_9C$ (SEQ ID NO: 121), $CCX_9CX_3CXCCX_3CX_5C$ (SEQ ID NO: 122), $CX_9CX_2CX_3CX_4CCX_4CX_5C$ (SEQ ID NO: 123), $CX_9CX_7CX_4CCXCX_7CX_3C$ (SEQ ID NO: 124), $CX_9CX_3CCCX_{10}CX_2CX_3C$ (SEQ ID NO: 125), $CX_3CX_5CX_5CX_4CCX_{10}CX_6C$ (SEQ ID NO: 126), $CX_9CX_5CX_4CCXCX_5CX_4C$ (SEQ ID NO: 127), $CX_7CXCX_6CX_4CCCX_{10}C$ (SEQ ID NO: 128), $CX_8CX_2CX_4CCX_4CX_3CX_3C$ (SEQ ID NO: 129), $CX_7CX_5CXCX_4CCX_7CX_4C$ (SEQ ID NO: 130), $CX_{11}CX_3CX_4CCCX_8CX_2C$ (SEQ ID NO: 131), $CX_2CX_3CX_4CCX_4CX_5CX_{15}C$ (SEQ ID NO: 132), $CX_9CX_5CX_4CCX_7C$ (SEQ ID NO: 133), $CX_9CX_7CX_3CX_2CX_6C$ (SEQ ID NO: 134), $CX_9CX_5CX_4CCX_{14}C$ (SEQ ID NO: 135), $CX_9CX_5CX_4CCX_8C$ (SEQ ID NO: 136), $CX_9CX_6CX_4CCXC$ (SEQ ID NO: 137),

$CX_5CCX_7CX_4CX_{12}$ (SEQ ID NO: 138), $CX_{10}CX_3CX_4CCX_4C$ (SEQ ID NO: 139), $CX_9CX_4CCX_5CX_4C$ (SEQ ID NO: 140), $CX_{10}CX_3CX_4CX_7CXC$ (SEQ ID NO: 141), $CX_7CX_7CX_2CX_2CX_3C$ (SEQ ID NO: 142), $CX_9CX_4CX_4CCX_6C$ (SEQ ID NO: 143), $CX_7CXCX_3CXCX_6C$ (SEQ ID NO: 144), $CX_7CXCX_4CXCX_4C$ (SEQ ID NO: 145), $CX_9CX_5CX_4C$ (SEQ ID NO: 146), $CX_3CX_6CX_8C$ (SEQ ID NO: 147), $CX_{10}CXCX_4C$ (SEQ ID NO: 148), $CX_{10}CCX_4C$ (SEQ ID NO: 149), $CX_{15}C$ (SEQ ID NO: 150), $CX_{10}C$ (SEQ ID NO: 151), and CX_9C (SEQ ID NO: 152).

[0031] In some embodiments of each or any of the above or below mentioned embodiments, the ultralong CDR3 comprises 2 to 6 disulfide bonds.

[0032] In some embodiments of each or any of the above or below mentioned embodiments, wherein the ultralong CDR3 comprises a non-antibody sequence.

[0033] In some embodiments of each or any of the above or below mentioned embodiments, the antibody heavy chain variable region includes wherein V1 comprises an amino acid sequence of QVQLVQSGAEVKKPGSSVKVSCKASGNSFSNHDVHWVRQATG QGLEWMGWMSHEGDKTGLAQKFQGRVTITRDSGASTVYMELRGLTADDTAIYYCLT (SEQ ID NO: 496), wherein the ultralong CDR3 comprises an amino acid sequence of any one of GSKHRLRDYFLYNE (SEQ ID NO: 501), GSKHRLRDYFLYN (SEQ ID NO: 502), GSKHRLRDYFLY (SEQ ID NO: 503), GSKHRLRDYFL (SEQ ID NO: 504), GSKHRLRDYF (SEQ ID NO: 505), GSKHRLRDY (SEQ ID NO: 506), or GSKHRLRD (SEQ ID NO: 507), a non-antibody sequence, and an amino acid sequence of any one of YGPNYEEWGDYLATLDV (SEQ ID NO: 536), GPNYEEWGDYLATLDV (SEQ ID NO: 537), PNYEEWGDYLATLDV (SEQ ID NO: 538), NYEEWGDYLATLDV (SEQ ID NO: 539), YEEWGDYLATLDV (SEQ ID NO: 540), or EEWGDYLATLDV (SEQ ID NO: 541), and wherein V2 comprises an amino acid sequence of WGHGTAVTVSS (SEQ ID NO: 570) (*see*, Figures 6C and 7A-7C).

[0034] In some embodiments of each or any of the above or below mentioned embodiments, the antibody heavy chain variable region includes wherein V1 comprises an amino acid sequence of QRLVESGGGVVQPGSSLRLSCAASGFDFSRQGMHWVRQAPG QGLEWVAFIKYDGSEKYHADSVWGRLSISRDNSKDTLYLQMNSLRVEDTATYFCVR (SEQ ID NO: 497), wherein the ultralong CDR3 comprises an amino acid sequence of any one of EAGGPDYRNGYNY (SEQ ID NO: 508), EAGGPDYRNGYN (SEQ ID NO: 509), EAGGPDYRNGY (SEQ ID NO: 510), EAGGPDYRNG (SEQ ID NO: 511), EAGGPDYRN (SEQ ID NO: 512), EAGGPDYR (SEQ ID NO: 513), EAGGPDY (SEQ ID NO: 514), or EAGGPD (SEQ ID NO: 515), a non-antibody sequence, and an amino acid sequence of any one of YDFYDGYYNYHYMDV (SEQ ID NO: 542), DFYDGYYNYHYMDV (SEQ ID NO: 543), FYDGYYNYHYMDV (SEQ ID NO: 544), YDGYYNYHYMDV (SEQ ID NO: 545),

DGYYNYHYMDV (SEQ ID NO: 546), GYYNYHYMDV (SEQ ID NO: 547), or YYNYHYMDV (SEQ ID NO: 548), wherein V2 comprises an amino acid sequence selected of WGKGTTVTVSS (SEQ ID NO: 571) (*see*, Figures 6C and 7A-7C).

[0035] In some embodiments of each or any of the above or below mentioned embodiments, the antibody heavy chain variable region includes wherein V1 comprises an amino acid sequence of QEQLVESGGGVVQPGGSLRLSCLASGFTFHKYGMHWVRQAPG KGLEWVALISDDGMRKYHSDSMWGRVTISRDNSKNTLYLQFSSLKVEDTAMFFCAR (SEQ ID NO: 498), wherein the ultralong CDR3 comprises an amino acid sequence of any one of EAGGPIWHDDVKY (SEQ ID NO: 516), EAGGPIWHDDVK (SEQ ID NO: 517), EAGGPIWHDDV (SEQ ID NO: 518), EAGGPIWHDD (SEQ ID NO: 519), EAGGPIWHD (SEQ ID NO: 520), EAGGPIWH (SEQ ID NO: 521), EAGGPIW (SEQ ID NO: 522), or EAGGPI (SEQ ID NO: 523), a non-antibody sequence, and an amino acid sequence of any one of YDFNDGYYNYHYMDV (SEQ ID NO: 549), DFYDGYYNYHYMDV (SEQ ID NO: 550), FYDGYYNYHYMDV (SEQ ID NO: 551), YDGYYNYHYMDV (SEQ ID NO: 552), DGYYNYHYMDV (SEQ ID NO: 553), or GYYNYHYMDV (SEQ ID NO: 554), wherein V2 comprises an amino acid sequence selected of WGKGTTVTVSS (SEQ ID NO: 572) (*see*, Figures 6C and 7A-7C).

[0036] In some embodiments of each or any of the above or below mentioned embodiments, the antibody heavy chain variable region includes wherein V1 comprises an amino acid sequence of EVQLVESGGGLIRPGGSLRLSCKGSGFIFENFGFGWVRQGPGK GLEWVSGTNWNGGDSRYGDSVKGRFTISRDNSNNFVYLQMNSLRPEDTAIYYCAR (SEQ ID NO: 499), wherein the ultralong CDR3 comprises an amino acid sequence of any one of GTDYTIDDQGI (SEQ ID NO: 524), GTDYTIDDQG (SEQ ID NO: 525), GTDYTIDDQ (SEQ ID NO: 526), GTDYTIDD (SEQ ID NO: 527), GTDYTID (SEQ ID NO: 528), or GTDYTI (SEQ ID NO: 529), a non-antibody sequence, and an amino acid sequence of any one of QGIRYQGSGTFWYFDV (SEQ ID NO: 555), GIRYQGSGTFWYFDV (SEQ ID NO: 556), IRYQGSGTFWYFDV (SEQ ID NO: 557), RYQGSGTFWYFDV (SEQ ID NO: 558), YQGSGTFWYFDV (SEQ ID NO: 559), QGSGTFWYFDV (SEQ ID NO: 560), GSGTFWYFDV (SEQ ID NO: 561), SGTFWYFDV (SEQ ID NO: 562), or GTFWYFDV (SEQ ID NO: 563), wherein V2 comprises an amino acid sequence selected of WGRGTLVTVSS (SEQ ID NO: 573) (*see*, Figures 6C and 7A-7C).

[0037] In some embodiments of each or any of the above or below mentioned embodiments, the antibody heavy chain variable region includes wherein V1 comprises an amino acid sequence of EVQLVESGGNVVQPGGSLRLSCTASGFSFDDSTMHWVRQAPG KGLQWVSLISWNGGRTYYADSVKGRFTISRDNSKNSLYLQMNSLKTEDTAFYFCAK (SEQ ID NO: 500), wherein the ultralong CDR3 comprises an amino acid sequence of any one of

YNLGYSYFYYMDG(SEQ ID NO: 564), NLGYSYFYYMDG (SEQ ID NO: 565), LGYSYFYYMDG (SEQ ID NO: 566), GYSYFYYMDG (SEQ ID NO: 567), YSYFYYMDG (SEQ ID NO: 568), or SYFYYMDG (SEQ ID NO: 569), a non-antibody sequence, and an amino acid sequence of any one of YNLGYSYFYYMDG(SEQ ID NO: 564), NLGYSYFYYMDG (SEQ ID NO: 565), LGYSYFYYMDG (SEQ ID NO: 566), GYSYFYYMDG (SEQ ID NO: 567), YSYFYYMDG (SEQ ID NO: 568), or SYFYYMDG (SEQ ID NO: 569), wherein V2 comprises an amino acid sequence selected of WGKGTTVTVSS (SEQ ID NO: 574) (*see*, Figures 6C and 7A-7C).

[0038] In some embodiments of each or any of the above or below mentioned embodiments, the non-antibody sequence is a synthetic sequence.

[0039] In some embodiments of each or any of the above or below mentioned embodiments, the non-antibody sequence is a cytokine sequence, a lymphokine sequence, a chemokine sequence, a growth factor sequence, a hormone sequence, or a toxin sequence.

[0040] In some embodiments of each or any of the above or below mentioned embodiments, the non-antibody sequence is an IL-8 sequence, an IL-21 sequence, an SDF-1 (alpha) sequence, a somatostatin sequence, a chlorotoxin sequence, a Pro-TxII sequence, a ziconotide sequence, an ADWX-1 sequence, an HsTx1 sequence, an OSK1 sequence, a Pi2 sequence, a Hongotoxin (HgTX) sequence, a Margatoxin sequence, an Agitoxin-2 sequence, a Pi3 sequence, a Kaliotoxin sequence, an Anuroctoxin sequence, a Charybdotoxin sequence, a Tityustoxin -K- alpha sequence, a Maurotoxin sequence, a Ceratotoxin 1 (CcoTx1) sequence, a CcoTx2 sequence, a CcoTx3 sequence, a Phrixotoxin 3 (PaurTx3) sequence, a Hanatoxin 1 sequence, a Phrixotoxin 1 sequence, a Huwentoxin-IV sequence, an α-conotoxin Iml sequence, an α-conotoxin Epl sequence, an α-conotoxin PnIA sequence, an α-conotoxin PnIB sequence, an α-conotoxin MII sequence, an α-conotoxin AuIA sequence, an α-conotoxin AuIB sequence, an α-conotoxin AuIC sequence, a conotoxin κ-PVIIA sequence, a charybdotoxin sequence, a neurotoxin B-IV sequence, a crotamine sequence, a ω-GVIA (conotoxin) sequence, a κ-hefutoxin 1 sequence, a Css4 sequence, a Bj-xtrIT sequence, a BclV sequence, a Hm-1 sequence, a Hm-2 sequence, a GsAF-I (β-theraphotoxin-Gr1b) sequence, a Protoxin I (ProTx-I sequence, a β-theraphotoxin-Tp1a) sequence, a Protoxin II (ProTx II) sequence, a Huwentoxin I sequence, a μ-Conotoxin PIIIA sequence, a Jingzhaotoxin-III (ß-TRTX-Cj1α) sequence, a GsAF-II (Kappa-theraphotoxin-Gr2c) sequence, a ShK (Stichodactyla toxin) sequence, a HsTx1 sequence, a Guangxitoxin 1E (GxTx-1E) sequence, a Maurotoxin sequence, a Charybdotoxin (ChTX) sequence, an Iberiotoxin (IbTx) sequence, a Leiurotoxin 1 (scyllatoxin) sequence, a Tamapin sequence, a Kaliotoxin-1 (KTX) sequence, a Purotoxin1

(PT-1) sequence, or a GpTx-1 sequence, a MOKA Toxin sequence, a OSK1 (P12, K16, D20) sequence, a OSK1 (K16, D20) sequence, a HmK sequence, a ShK (K16,Y26, K29) sequence, a ShK (K16) sequence, a ShK-A (K16) sequence, a ShK (K16,E30) sequence, a ShK (Q21) sequence, a ShK (L21) sequence, a ShK (F21) sequence, a ShK (I21) sequence, or a ShK (A21) sequence (*see*, Figures 7B-7C).

[0041] In some embodiments of each or any of the above or below mentioned embodiments, the present disclosure provides an antibody heavy chain variable region of comprising the amino acid sequence of QVQLVQSGAEVKKPGSSVKVSCKASGNSFSNHDVHWVRQATGQGLEWMGWMSHEGD KTGLAQKFQGRVTITRDSGASTVYMELRGLTADDTAIYYCLTGSKHRLRDYFLYNEYGPN YEEWGDYLATLDVWGHGTAVTVSS (SEQ ID NO: 656), wherein the heavy chain variable region further comprises a non-antibody sequence inserted between any one of: (i) the L at position 109 and the Y at position 110; (ii) the Y at position 110 and the N at position 111; (iii) the N at position 111 and the E at position 112; (iv) the E at position 112 and the Y at position 113; (v) the Y at position 113 and the G at position 114; (vi) the G at position 114 and the P at position 115; (vii) the P at position 115 and the N at position 116; or (viii) the N at position 116 and the Y at position 117, or wherein the amino acid sequence of YNEYGPN at positions 110 to 116 has been removed and replaced with a non-antibody sequence, or wherein the P at position 115 has been removed and replaced with a non-antibody sequence (*see*, Figures 6A-6B).

[0042] In some embodiments of each or any of the above or below mentioned embodiments, the present disclosure provides an antibody heavy chain variable region comprising the amino acid sequence of QRLVESGGGVVQPGSSLRLSCAASGFDFSRQGMHWVRQAPGQGLEWVAFIKYDGSEK YHADSVWGRLSISRDNSKDTLYLQMNSLRVEDTATYFCVREAGGPDYRNGYNYYDFYDG YYNYHYMDVWGKGTTVTVSS (SEQ ID NO: 657), wherein the heavy chain variable region further comprises a non-antibody sequence inserted between any one of: (i) the Y at position 104 and the R at position 105; (ii) the R at position 105 and the N at position 106; (iii) the N at position 106 and the G at position 107; (iv) the G at position 107 and the Y at position 108; (v) the Y at position 108 and the N at position 109; (vi) the N at position 109 and the Y at position 110; (vii) the Y at position 110 and the Y at position 111; (viii) the Y at position 111 and the D at position 112; (ix) the D at position 112 and the F at position 113; (x) the F at position 113 and the Y at position 114; (xi) the Y at position 114 and the D at position 115; (xi) the D at position 115 and the G at position 116, or wherein the amino acid sequence of NYYD at positions 109 to 112 has been removed and replaced with a non-

antibody sequence, or wherein the Y at position 110 has been removed and replaced with a non-antibody sequence (*see*, Figures 6A-6B).

[0043] In some embodiments of each or any of the above or below mentioned embodiments, the present disclosure provides an antibody heavy chain variable region comprising the amino acid sequence of QEQLVESGGGVVQPGGSLRLSCLASGFTFHKYGMHWVRQAPGKGLEWVALISDDGMRK YHSDSMWGRVTISRDNSKNTLYLQFSSLKVEDTAMFFCAREAGGPIWHDDVKYYDFNDG YYNYHYMDVWGKGTTVTVSS (SEQ ID NO: 658), wherein the heavy chain variable region further comprises a non-antibody sequence inserted between any on one of: (i) the W at position 105 and the H at position 106; (ii) the H at position 106 and the D at position 107; (iii) the D at position 107 and the D at position 108; (iv) the D at position 108 and the V at position 109; or (v) the V at position 109 and the K at position 110, or wherein the amino acid sequence of DD at positions 107 to 108 has been removed and replaced with a non-antibody sequence (*see*, Figures 6A-6B).

[0044] In some embodiments of each or any of the above or below mentioned embodiments, the present disclosure provides an antibody heavy chain variable region comprising the amino acid sequence of EVQLVESGGGLIRPGGSLRLSCKGSGFIFENFGFGWVRQGPGKGLEWVSGTNWNGGDS RYGDSVKGRFTISRDNSNNFVYLQMNSLRPEDTAIYYCARGTDYTIDDQGIRYQGSGTFW YFDVWGRGTLVTVSS (SEQ ID NO: 659), wherein the heavy chain variable region further comprises a non-antibody sequence inserted between any on one of: (i) the I at position 104 and the D at position 105; (ii) the D at position 105 and the D at position 106; (iii) the D at position 106 and the Q at position 107; (iv) the Q at position 107 and the G at position 108; or (v) the G at position 108 and the I at position 109, or wherein the amino acid sequence of DQ at positions 107 to 108 has been removed and replaced with a non-antibody sequence (*see*, Figures 6A-6B).

[0045] In some embodiments of each or any of the above or below mentioned embodiments, the present disclosure provides an antibody heavy chain variable region comprising the amino acid sequence of EVQLVESGGNVVQPGGSLRLSCTASGFSFDDSTMHWVRQAPGKGLQWVSLISWNGGR TYYADSVKGRFTISRDNSKNSLYLQMNSLKTEDTAFYFCAKDKGDSDYDYNLGYSYFYYM DGWGKGTTVTVSS (SEQ ID NO: 660), wherein the heavy chain variable region further comprises a non-antibody sequence inserted between any on one of: (i) the Y at position 105 and the D at position 106; (ii) the D at position 106 and the Y at position 107; (iii) the Y at position 107 and the N at position 108; (iv) the N at position 108 and the L at position 109; (v) the L at position 109 and the G at position 110; or (vi) the G at position 110 and the

Y at position 111, or wherein the amino acid sequence of YNL at positions 107 to 109 has been removed and replaced with a non-antibody sequence (*see*, Figures 6A-6B).

[0046] In some embodiments of each or any of the above or below mentioned embodiments, the ultralong CDR3 comprises a linker sequence.

[0047] In some embodiments of each or any of the above or below mentioned embodiments, the linker is linked to a N-terminus, a C-terminus, or both N-terminus and C-terminus of the non-antibody sequence.

[0048] In some embodiments of each or any of the above or below mentioned embodiments, the linker comprises one or more amino acid sequence selected from the group consisting of SEQ ID NO: 575 to 598, 699 to 726 and 756-773, or any combination thereof (*see*, Figure 7A).

[0049] In some embodiments of each or any of the above or below mentioned embodiments, the linkers linked to both N-terminus and C-terminus have the same or different amino acid sequence.

[0050] The present disclosure also provides an antibody or binding fragment thereof comprising a heavy chain variable region having an ultralong CDR3.

[0051] In some embodiments, an antibody or binding fragment thereof comprises the antibody heavy chain variable region of each or any of the above or below mentioned antibody heavy chain variable regions.

[0052] In some embodiments of each or any of the above or below mentioned embodiments, the heavy chain variable region further comprises a constant heavy chain 1 (CH1) region.

[0053] In some embodiments of each or any of the above or below mentioned embodiments, the CH1 region comprises an amino acid sequence selected from SEQ ID NO: 661 to 665.

[0054] In some embodiments of each or any of the above or below mentioned embodiments, the heavy chain variable region further comprises an amino acid sequence of SEQ ID NO: 390.

[0055] In some embodiments of each or any of the above or below mentioned embodiments, the antibody or binding fragment further comprises a light chain variable region.

[0056] In some embodiments of each or any of the above or below mentioned embodiments, the light chain variable region comprises an amino acid sequence selected from SEQ ID NO: 734 to 738.

[0057] In some embodiments of each or any of the above or below mentioned embodiments, the light chain variable region further comprising a constant light chain (CL) region.

[0058] In some embodiments of each or any of the above or below mentioned embodiments, the CL region comprises an amino acid sequence selected from SEQ ID NO: 739 to 743.

[0059] The present disclosure also provides polynucleotides encoding an antibody heavy chain variable region of each or any of the above or below mentioned antibody heavy chain variable regions.

[0060] The present disclosure also provides vectors comprising the polynucleotides disclosed herein.

[0061] The present disclosure also provides host cells comprising the vectors disclosed herein.

[0062] The present disclosure also provides a nucleic acid library comprising a plurality of polynucleotides comprising nucleic acid sequences encoding antibody heavy chain variable regions of each or any of the above or below mentioned heavy chain variable regions. Accordingly, in some embodiments, the nucleic acid library comprises a plurality of polynucleotides comprising nucleic acid sequences encoding for an antibody heavy chain comprising a sequence of the formula V1-X-V2, wherein V1 comprises an amino acid sequence selected from the group consisting of:
(i) QVQLVQSGAEVKKPGSSVKVSCKASGNSFSNHDVHWVRQATGQGLEWMGWMSHE GDKTGLAQKFQGRVTITRDSGASTVYMELRGLTADDTAIYYCLT (SEQ ID NO: 496),
(ii) QRLVESGGGVVQPGSSLRLSCAASGFDFSRQGMHWVRQAPGQGLEWVAFIKYDGS EKYHADSVWGRLSISRDNSKDTLYLQMNSLRVEDTATYFCVR (SEQ ID NO: 497),
(iii) QEQLVESGGGVVQPGGSLRLSCLASGFTFHKYGMHWVRQAPGKGLEWVALISDDG MRKYHSDSMWGRVTISRDNSKNTLYLQFSSLKVEDTAMFFCAR (SEQ ID NO: 498),
(iv) EVQLVESGGGLIRPGGSLRLSCKGSGFIFENFGFGWVRQGPGKGLEWVSGTNWNG GDSRYGDSVKGRFTISRDNSNNFVYLQMNSLRPEDTAIYYCAR (SEQ ID NO: 499), and
(v) EVQLVESGGNVVQPGGSLRLSCTASGFSFDDSTMHWVRQAPGKGLQWVSLISWNG GRTYYADSVKGRFTISRDNSKNSLYLQMNSLKTEDTAFYFCAK (SEQ ID NO: 500);
wherein X comprises an ultralong CDR3, which can include a non-human sequence or a non-antibody sequence (e.g., a non-antibody human sequence) that has been inserted into the CDR3 sequence of the antibody, including optionally, removing a portion of CDR3 (e.g., one or more amino acids of the CDR3) or the entire CDR3 sequence (e.g., all or substantially all of the amino acids of the CDR3); and wherein V2 comprises an amino acid sequence selected from the group consisting of: (i) WGHGTAVTVSS (SEQ ID NO: 570),

(ii) WGKGTTVTVSS (SEQ ID NO: 571), (iii) WGKGTTVTVSS (SEQ ID NO: 572), (iv) WGRGTLVTVSS (SEQ ID NO: 573), and (v) WGKGTTVTVSS (SEQ ID NO: 574) (*see*, Figure 6C).

[0063] The present disclosure also provides a library of antibodies comprising antibody heavy chain variable regions of each or any of the above or below mentioned heavy chain variable regions. Accordingly, in some embodiments, the library of antibodies comprise antibody heavy chain variable regions having a sequence of the formula V1-X-V2, wherein V1 comprises an amino acid sequence selected from the group consisting of: (i) QVQLVQSGAEVKKPGSSVKVSCKASGNSFSNHDVHWVRQATGQGLEW MGWMSHEGDKTGLAQKFQGRVTITRDSGASTVYMELRGLTADDTAIYYCLT (SEQ ID NO: 496), (ii) QRLVESGGGVVQPGSSLRLSCAASGFDFSRQGMHWVRQAPGQGLEWVA FIKYDGSEKYHADSVWGRLSISRDNSKDTLYLQMNSLRVEDTATYFCVR (SEQ ID NO: 497), (iii) QEQLVESGGGVVQPGGSLRLSCLASGFTFHKYGMHWVRQAPGKGLEWVALIS DDGMRKYHSDSMWGRVTISRDNSKNTLYLQFSSLKVEDTAMFFCAR (SEQ ID NO: 498), (iv) EVQLVESGGGLIRPGGSLRLSCKGSGFIFENFGFGWVRQGPGKGLEWVSGTN WNGGDSRYGDSVKGRFTISRDNSNNFVYLQMNSLRPEDTAIYYCAR (SEQ ID NO: 499), and (v) EVQLVESGGNVVQPGGSLRLSCTASGFSFDDSTMHWVRQAPGKGLQWVSLISW NGGRTYYADSVKGRFTISRDNSKNSLYLQMNSLKTEDTAFYFCAK (SEQ ID NO: 500); wherein X comprises an ultralong CDR3, which can include a non-human sequence or a non-antibody sequence (e.g., a non-antibody human sequence) that has been inserted into the CDR3 sequence of the antibody, including optionally, removing a portion of CDR3 (e.g., one or more amino acids of the CDR3) or the entire CDR3 sequence (e.g., all or substantially all of the amino acids of the CDR3); and wherein V2 comprises an amino acid sequence selected from the group consisting of: (i) WGHGTAVTVSS (SEQ ID NO: 570), (ii) WGKGTTVTVSS (SEQ ID NO: 571), (iii) WGKGTTVTVSS (SEQ ID NO: 572), (iv) WGRGTLVTVSS (SEQ ID NO: 573), and (v) WGKGTTVTVSS (SEQ ID NO: 574) (*see*, Figure 6C).

[0064] In some embodiments of each or any of the above or below mentioned embodiments, the ultralong CDR3 comprises a $X^1X^2X^3X^4X^5$ motif, wherein X^1 is threonine (T), glycine (G), alanine (A), serine (S), or valine (V), wherein X^2 is serine (S), threonine (T), proline (P), isoleucine (I), alanine (A), valine (V), or asparagine (N), wherein X^3 is valine (V), alanine (A), threonine (T), or aspartic acid (D), wherein X^4 is histidine (H), threonine (T), arginine (R), tyrosine (Y), phenylalanine (F), or leucine (L), and wherein X^5 is glutamine (Q).

[0065] In some embodiments of each or any of the above or below mentioned embodiments, the $X^1X^2X^3X^4X^5$ motif is TTVHQ (SEQ ID NO: 153), TSVHQ (SEQ ID NO:

154), SSVTQ (SEQ ID NO: 155), STVHQ (SEQ ID NO: 156), ATVRQ (SEQ ID NO: 157), TTVYQ (SEQ ID NO: 158), SPVHQ (SEQ ID NO: 159), ATVYQ (SEQ ID NO: 160), TAVYQ (SEQ ID NO: 161), TNVHQ (SEQ ID NO: 162), ATVHQ (SEQ ID NO: 163), STVYQ (SEQ ID NO: 164), TIVHQ (SEQ ID NO: 165), AIVYQ (SEQ ID NO: 166), TTVFQ (SEQ ID NO: 167), AAVFQ (SEQ ID NO: 168), GTVHQ (SEQ ID NO: 169), ASVHQ (SEQ ID NO: 170), TAVFQ (SEQ ID NO: 171), ATVFQ (SEQ ID NO: 172), AAAHQ (SEQ ID NO: 173), VVVYQ (SEQ ID NO: 174), GTVFQ (SEQ ID NO: 175), TAVHQ (SEQ ID NO: 176), ITVHQ (SEQ ID NO: 177), ITAHQ (SEQ ID NO: 178), VTVHQ (SEQ ID NO: 179); AAVHQ (SEQ ID NO: 180), GTVYQ (SEQ ID NO: 181), TTVLQ (SEQ ID NO: 182), TTTHQ (SEQ ID NO: 183), or TTDYQ (SEQ ID NO: 184).

[0066] In some embodiments of each or any of the above or below mentioned embodiments, the ultralong CDR3 comprises a $(X^aX^b)_z$ motif, wherein X^a is any amino acid residue, X^b is an aromatic amino acid selected from the group consisting of: tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H), and wherein z is 1 - 4.

[0067] In some embodiments of each or any of the above or below mentioned embodiments, the $(X^aX^b)_z$ motif is CYTYNYEF (SEQ ID NO: 217), HYTYTYDF (SEQ ID NO: 218), HYTYTYEW (SEQ ID NO: 219), KHRYTYEW (SEQ ID NO: 220), NYIYKYSF (SEQ ID NO: 221), PYIYTYQF (SEQ ID NO: 222), SFTYTYEW (SEQ ID NO: 223), SYIYIYQW (SEQ ID NO: 224), SYNYTYSW (SEQ ID NO: 225), SYSYSYEY (SEQ ID NO: 226), SYTYNYDF (SEQ ID NO: 227), SYTYNYEW (SEQ ID NO: 228), SYTYNYQF (SEQ ID NO: 229), SYVWTHNF (SEQ ID NO: 230), TYKYVYEW (SEQ ID NO: 231), TYTYTYEF (SEQ ID NO: 232), TYTYTYEW (SEQ ID NO: 233), VFTYTYEF (SEQ ID NO: 234), AYTYEW (SEQ ID NO: 235), DYIYTY (SEQ ID NO: 236), IHSYEF (SEQ ID NO: 237), SFTYEF (SEQ ID NO: 238), SHSYEF (SEQ ID NO: 239), THTYEF (SEQ ID NO: 240), TWTYEF (SEQ ID NO: 241), TYNYEW (SEQ ID NO: 242), TYSYEF (SEQ ID NO: 243), TYSYEH (SEQ ID NO: 244), TYTYDF (SEQ ID NO: 245), TYTYEF (SEQ ID NO: 246), TYTYEW (SEQ ID NO: 247), AYEF (SEQ ID NO: 248), AYSF (SEQ ID NO: 249), AYSY (SEQ ID NO: 250), CYSF (SEQ ID NO: 251), DYTY (SEQ ID NO: 252), KYEH (SEQ ID NO: 253), KYEW (SEQ ID NO: 254), MYEF (SEQ ID NO: 255), NWIY (SEQ ID NO: 256), NYDY (SEQ ID NO: 257), NYQW (SEQ ID NO: 258), NYSF (SEQ ID NO: 259), PYEW (SEQ ID NO: 260), RYNW (SEQ ID NO: 261), RYTY (SEQ ID NO: 262), SYEF (SEQ ID NO: 263), SYEH (SEQ ID NO: 264), SYEW (SEQ ID NO: 265), SYKW (SEQ ID NO: 266), SYTY (SEQ ID NO: 267), TYDF (SEQ ID NO: 268), TYEF (SEQ ID NO: 269), TYEW (SEQ ID NO: 270), TYQW (SEQ ID NO: 271), TYTY (SEQ ID NO: 272), or VYEW (SEQ ID NO: 273).

[0068] In some embodiments of each or any of the above or below mentioned embodiments, the $(X^aX^b)_z$ motif is YXYXYX.

[0069] In some embodiments of each or any of the above or below mentioned embodiments, the ultralong CDR3 comprises a $X^1X^2X^3X^4X^5X_n$ motif, wherein X_1 is threonine (T), glycine (G), alanine (A), serine (S), or valine (V), wherein X_2 is serine (S), threonine (T), proline (P), isoleucine (I), alanine (A), valine (V), or asparagine (N), wherein X_3 is valine (V), alanine (A), threonine (T), or aspartic acid (D), wherein X_4 is histidine (H), threonine (T), arginine (R), tyrosine (Y), phenylalanine (F), or leucine (L), wherein X_5 is glutamine (Q), and wherein n is 27 – 54.

[0070] In some embodiments of each or any of the above or below mentioned embodiments, the ultralong CDR3 comprises $X_n(X^aX^b)_z$ motif, wherein X^a is any amino acid residue, X^b is an aromatic amino acid selected from the group consisting of: tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H), wherein n is 27 – 54, and wherein z is 1 - 4.

[0071] In some embodiments of each or any of the above or below mentioned embodiments, the ultralong CDR3 comprises a $X^1X^2X^3X^4X^5X_n(X^aX^b)_z$ motif, wherein X^1 is threonine (T), glycine (G), alanine (A), serine (S), or valine (V), wherein X^2 is serine (S), threonine (T), proline (P), isoleucine (I), alanine (A), valine (V), or asparagine (N), wherein X^3 is valine (V), alanine (A), threonine (T), or aspartic acid (D), wherein X^4 is histidine (H), threonine (T), arginine (R), tyrosine (Y), phenylalanine (F), or leucine (L), and wherein X^5 is glutamine (Q), wherein X^a is any amino acid residue, X^b is an aromatic amino acid selected from the group consisting of: tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H), wherein n is 27 – 54, and wherein z is 1 - 4.

[0072] The present disclosure also provides antibody heavy chain variable regions comprising a sequence of the formula V1-X, wherein V1 comprises an amino acid sequence selected from the group consisting of:

(i) QVQLVQSGAEVKKPGSSVKVSCKASGNSFSNHDVHWVRQATGQGLEWMGWMSHE GDKTGLAQKFQGRVTITRDSGASTVYMELRGLTADDTAIYYCLT (SEQ ID NO: 496),

(ii) QRLVESGGGVVQPGSSLRLSCAASGFDFSRQGMHWVRQAPGQGLEWVAFIKYDGS EKYHADSVWGRLSISRDNSKDTLYLQMNSLRVEDTATYFCVR (SEQ ID NO: 497),

(iii) QEQLVESGGGVVQPGGSLRLSCLASGFTFHKYGMHWVRQAPGKGLEWVALISDDG MRKYHSDSMWGRVTISRDNSKNTLYLQFSSLKVEDTAMFFCAR (SEQ ID NO: 498),

(iv) EVQLVESGGGLIRPGGSLRLSCKGSGFIFENFGFGWVRQGPGKGLEWVSGTNWNG GDSRYGDSVKGRFTISRDNSNNFVYLQMNSLRPEDTAIYYCAR (SEQ ID NO: 499),

(v) EVQLVESGGNVVQPGGSLRLSCTASGFSFDDSTMHWVRQAPGKGLQWVSLISWNG GRTYYADSVKGRFTISRDNSKNSLYLQMNSLKTEDTAFYFCAK (SEQ ID NO: 500);

(vi) QVQLVQSGAEVKKPGSSVKVSCKASGNSFSNHDVHWVRQATGQGLEWMGWMSHE GDKTGLAQKFQGRVTITRDSGASTVYMELRGLTADDTAIYYC (SEQ ID NO: 744), (vii) QRLVESGGGVVQPGSSLRLSCAASGFDFSRQGMHWVRQAPGQGLEWVAFIKYDGS EKYHADSVWGRLSISRDNSKDTLYLQMNSLRVEDTATYFC (SEQ ID NO: 745), (viii) QEQLVESGGGVVQPGGSLRLSCLASGFTFHKYGMHWVRQAPGKGLEWVALISDDG MRKYHSDSMWGRVTISRDNSKNTLYLQFSSLKVEDTAMFFC (SEQ ID NO: 746), (ix) EVQLVESGGGLIRPGGSLRLSCKGSGFIFENFGFGWVRQGPGKGLEWVSGTNWNG GDSRYGDSVKGRFTISRDNSNNFVYLQMNSLRPEDTAIYYC (SEQ ID NO: 747), (x) EVQLVESGGNVVQPGGSLRLSCTASGFSFDDSTMHWVRQAPGKGLQWVSLISWNG GRTYYADSVKGRFTISRDNSKNSLYLQMNSLKTEDTAFYFC (SEQ ID NO: 748), (xi) QVQLVQSGAEVKKPGSSVKVSCKASGNSFSNHDVHWVRQATGQGLEWMGWMSHE GDKTGLAQKFQGRVTITRDSGASTVYMELRGLTADDTAIYY (SEQ ID NO: 749), (xii) QRLVESGGGVVQPGSSLRLSCAASGFDFSRQGMHWVRQAPGQGLEWVAFIKYDGS EKYHADSVWGRLSISRDNSKDTLYLQMNSLRVEDTATYF (SEQ ID NO: 750), (xiii) QEQLVESGGGVVQPGGSLRLSCLASGFTFHKYGMHWVRQAPGKGLEWVALISDDG MRKYHSDSMWGRVTISRDNSKNTLYLQFSSLKVEDTAMFF (SEQ ID NO: 751), (xiv) EVQLVESGGGLIRPGGSLRLSCKGSGFIFENFGFGWVRQGPGKGLEWVSGTNWNG GDSRYGDSVKGRFTISRDNSNNFVYLQMNSLRPEDTAIYY (SEQ ID NO: 752), and (xv) EVQLVESGGNVVQPGGSLRLSCTASGFSFDDSTMHWVRQAPGKGLQWVSLISWNG GRTYYADSVKGRFTISRDNSKNSLYLQMNSLKTEDTAFYF (SEQ ID NO: 753); and wherein X comprises an ultralong CDR3, which can include a non-human sequence or a non-antibody sequence (e.g., a non-antibody human sequence) that has been inserted into the CDR3 sequence of the antibody, including optionally, removing a portion of CDR3 (e.g., one or more amino acids of the CDR3) or the entire CDR3 sequence (e.g., all or substantially all of the amino acids of the CDR3) (*see*, Figure 6D).

[0073] In some embodiments of each or any of the above or below mentioned embodiments, the ultralong CDR3 comprises a $X^1X^2X^3X^4X^5$ motif, wherein X^1 is threonine (T), glycine (G), alanine (A), serine (S), or valine (V), wherein X^2 is serine (S), threonine (T), proline (P), isoleucine (I), alanine (A), valine (V), or asparagine (N), wherein X^3 is valine (V), alanine (A), threonine (T), or aspartic acid (D), wherein X^4 is histidine (H), threonine (T), arginine (R), tyrosine (Y), phenylalanine (F), or leucine (L), and wherein X^5 is glutamine (Q).

[0074] In some embodiments of each or any of the above or below mentioned embodiments, the $X^1X^2X^3X^4X^5$ motif is TTVHQ (SEQ ID NO: 153), TSVHQ (SEQ ID NO: 154), SSVTQ (SEQ ID NO: 155), STVHQ (SEQ ID NO: 156), ATVRQ (SEQ ID NO: 157), TTVYQ (SEQ ID NO: 158), SPVHQ (SEQ ID NO: 159), ATVYQ (SEQ ID NO: 160), TAVYQ

(SEQ ID NO: 161), TNVHQ (SEQ ID NO: 162), ATVHQ (SEQ ID NO: 163), STVYQ (SEQ ID NO: 164), TIVHQ (SEQ ID NO: 165), AIVYQ (SEQ ID NO: 166), TTVFQ (SEQ ID NO: 167), AAVFQ (SEQ ID NO: 168), GTVHQ (SEQ ID NO: 169), ASVHQ (SEQ ID NO: 170), TAVFQ (SEQ ID NO: 171), ATVFQ (SEQ ID NO: 172), AAAHQ (SEQ ID NO: 173), VVVYQ (SEQ ID NO: 174), GTVFQ (SEQ ID NO: 175), TAVHQ (SEQ ID NO: 176), ITVHQ (SEQ ID NO: 177), ITAHQ (SEQ ID NO: 178), VTVHQ (SEQ ID NO: 179); AAVHQ (SEQ ID NO: 180), GTVYQ (SEQ ID NO: 181), TTVLQ (SEQ ID NO: 182), TTTHQ (SEQ ID NO: 183), or TTDYQ (SEQ ID NO: 184).

[0075] In some embodiments of each or any of the above or below mentioned embodiments, the ultralong CDR3 comprises a $CX^1X^2X^3X^4X^5$ motif.

[0076] In some embodiments of each or any of the above or below mentioned embodiments, the $CX^1X^2X^3X^4X^5$ motif is CTTVHQ (SEQ ID NO: 185), CTSVHQ (SEQ ID NO: 186), CSSVTQ (SEQ ID NO: 187), CSTVHQ (SEQ ID NO: 188), CATVRQ (SEQ ID NO: 189), CTTVYQ (SEQ ID NO: 190), CSPVHQ (SEQ ID NO: 191), CATVYQ (SEQ ID NO: 192), CTAVYQ (SEQ ID NO: 193), CTNVHQ (SEQ ID NO: 194), CATVHQ (SEQ ID NO: 195), CSTVYQ (SEQ ID NO: 196), CTIVHQ (SEQ ID NO: 197), CAIVYQ (SEQ ID NO: 198), CTTVFQ (SEQ ID NO: 199), CAAVFQ (SEQ ID NO: 200), CGTVHQ (SEQ ID NO: 201), CASVHQ (SEQ ID NO: 202), CTAVFQ (SEQ ID NO: 203), CATVFQ (SEQ ID NO: 204), CAAAHQ (SEQ ID NO: 205), CVVVYQ (SEQ ID NO: 206), CGTVFQ (SEQ ID NO: 207), CTAVHQ (SEQ ID NO: 208), CITVHQ (SEQ ID NO: 209), CITAHQ (SEQ ID NO: 210), CVTVHQ (SEQ ID NO: 211); CAAVHQ (SEQ ID NO: 212), CGTVYQ (SEQ ID NO: 213), CTTVLQ (SEQ ID NO: 214), CTTTHQ (SEQ ID NO: 215), or CTTDYQ (SEQ ID NO: 216).

[0077] In some embodiments of each or any of the above or below mentioned embodiments, the ultralong CDR3 comprises a $(X^aX^b)_z$ motif, wherein X^a is any amino acid residue, X^b is an aromatic amino acid selected from the group consisting of: tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H), and wherein z is 1 - 4.

[0078] In some embodiments of each or any of the above or below mentioned embodiments, the $(X^aX^b)_z$ motif is CYTYNYEF (SEQ ID NO: 217), HYTYTYDF (SEQ ID NO: 218), HYTYTYEW (SEQ ID NO: 219), KHRYTYEW (SEQ ID NO: 220), NYIYKYSF (SEQ ID NO: 221), PYIYTYQF (SEQ ID NO: 222), SFTYTYEW (SEQ ID NO: 223), SYIYIYQW (SEQ ID NO: 224), SYNYTYSW (SEQ ID NO: 225), SYSYSYEY (SEQ ID NO: 226), SYTYNYDF (SEQ ID NO: 227), SYTYNYEW (SEQ ID NO: 228), SYTYNYQF (SEQ ID NO: 229), SYVWTHNF (SEQ ID NO: 230), TYKYVYEW (SEQ ID NO: 231), TYTYTYEF (SEQ ID NO: 232), TYTYTYEW (SEQ ID NO: 233), VFTYTYEF (SEQ ID NO: 234), AYTYEW (SEQ ID NO: 235), DYIYTY (SEQ ID NO: 236), IHSYEF (SEQ ID NO: 237),

SFTYEF (SEQ ID NO: 238), SHSYEF (SEQ ID NO: 239), THTYEF (SEQ ID NO: 240), TWTYEF (SEQ ID NO: 241), TYNYEW (SEQ ID NO: 242), TYSYEF (SEQ ID NO: 243), TYSYEH (SEQ ID NO: 244), TYTYDF (SEQ ID NO: 245), TYTYEF (SEQ ID NO: 246), TYTYEW (SEQ ID NO: 247), AYEF (SEQ ID NO: 248), AYSF (SEQ ID NO: 249), AYSY (SEQ ID NO: 250), CYSF (SEQ ID NO: 251), DYTY (SEQ ID NO: 252), KYEH (SEQ ID NO: 253), KYEW (SEQ ID NO: 254), MYEF (SEQ ID NO: 255), NWIY (SEQ ID NO: 256), NYDY (SEQ ID NO: 257), NYQW (SEQ ID NO: 258), NYSF (SEQ ID NO: 259), PYEW (SEQ ID NO: 260), RYNW (SEQ ID NO: 261), RYTY (SEQ ID NO: 262), SYEF (SEQ ID NO: 263), SYEH (SEQ ID NO: 264), SYEW (SEQ ID NO: 265), SYKW (SEQ ID NO: 266), SYTY (SEQ ID NO: 267), TYDF (SEQ ID NO: 268), TYEF (SEQ ID NO: 269), TYEW (SEQ ID NO: 270), TYQW (SEQ ID NO: 271), TYTY (SEQ ID NO: 272), or VYEW (SEQ ID NO: 273).

[0079] In some embodiments of each or any of the above or below mentioned embodiments, the $(X^aX^b)_z$ motif is YXYXYX.

[0080] In some embodiments of each or any of the above or below mentioned embodiments, the ultralong CDR3 comprises a $X^1X^2X^3X^4X^5X_n$ motif, wherein X_1 is threonine (T), glycine (G), alanine (A), serine (S), or valine (V), wherein X_2 is serine (S), threonine (T), proline (P), isoleucine (I), alanine (A), valine (V), or asparagine (N), wherein X_3 is valine (V), alanine (A), threonine (T), or aspartic acid (D), wherein X_4 is histidine (H), threonine (T), arginine (R), tyrosine (Y), phenylalanine (F), or leucine (L), wherein X_5 is glutamine (Q), and wherein n is 27 – 54.

[0081] In some embodiments of each or any of the above or below mentioned embodiments, the ultralong CDR3 comprises $X_n(X^aX^b)_z$ motif, wherein X^a is any amino acid residue, X^b is an aromatic amino acid selected from the group consisting of: tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H), wherein n is 27 – 54, and wherein z is 1 - 4.

[0082] In some embodiments of each or any of the above or below mentioned embodiments, the ultralong CDR3 comprises a $X^1X^2X^3X^4X^5X_n(X^aX^b)_z$ motif, wherein X^1 is threonine (T), glycine (G), alanine (A), serine (S), or valine (V), wherein X^2 is serine (S), threonine (T), proline (P), isoleucine (I), alanine (A), valine (V), or asparagine (N), wherein X^3 is valine (V), alanine (A), threonine (T), or aspartic acid (D), wherein X^4 is histidine (H), threonine (T), arginine (R), tyrosine (Y), phenylalanine (F), or leucine (L), and wherein X^5 is glutamine (Q), wherein X^a is any amino acid residue, X^b is an aromatic amino acid selected from the group consisting of: tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H), wherein n is 27 – 54, and wherein z is 1 - 4.

BRIEF DESCRIPTION OF THE DRAWINGS

[0083] The foregoing summary, as well as the following detailed description of the disclosure, will be better understood when read in conjunction with the appended figures. For the purpose of illustrating the disclosure, shown in the figures are embodiments which are presently preferred. It should be understood, however, that the disclosure is not limited to the precise arrangements, examples and instrumentalities shown.

[0084] Figure 1 shows a sequence alignment of exemplary bovine-derived antibody variable region sequences designated BLV1H12, BLV5B8, BLV5D3, BLV8C11, BF4E9, BF1H1, or F18 that comprise an ultralong CDR3 sequence.

[0085] Figure 2A-2C depicts ultralong CDR3 sequences. (Top) Translation from the germline V_HBUL, D_H2, and J_H. The 5 full length ultralong CDR H3s reported in the literature contain between four and eight cysteines and are not highly homologous to one another; however, some conservation of cysteine residues with D_H2 could be found when the first cysteine of these CDR H3s was "fixed" prior to alignment. Four of the seven sequences (BLV1H12, BLV5D3, BLV8C11, and BF4E9) contain four cysteines in the same positions as D_H2, but also have additional cysteines. BLV5B8 has two cysteines in common with the germline D_H2. This limited homology with some cysteine conservation suggests that mutation of D_H2 could generate these sequences. B-L1 and B-L2 are from initial sequences from bovine spleen, and the remaining are selected ultralong CDR H3 sequences from deep sequencing data. The first group contains the longest CDR H3s identified, and appear clonally related. The * indicates a sequence represented 167 times, suggesting it was strongly selected for function. Several of the eight-cysteine sequences appear selected for function as they were represented multiple times, indicated in parentheses. Other representative sequences of various lengths are indicated in the last group. The framework cysteine and tryptophan residues that define the CDR H3 boundaries are double-underlined. The sequences BLV1H12 through UL-77 (left-most column) presented in Tables 2A-C are depicted broken apart into four segments to identify the segments of amino acid residues that are derived from certain germline sequences and V/D/J joining sequences. Moving from left to right, the first segment is derived from the V_H germline and is represented in the disclosure as a $X^1X^2X^3X^4X^5$ motif. The second segment represents sequences from V-D joining and is represented in the disclosure as X_n. The third segment is a string of amino acid residues derived from D_H2 germline, and the fourth segment is a string of amino acid residues derived from J_H1 germline region.

[0086] Figure 3 depicts a sequence alignment of exemplary bovine-derived ultralong CDR3 sequences designated BLV1H12, BLV5B8, BLV5D3, BLV8C11, BF4E9, BF1H1, or F18.

[0087] Figure 4 shows an exemplary bovine light chain variable region sequence designated BLV1H12 suitable for modification or use with an ultralong CDR 3 sequence (*e.g.*, a heavy chain variable region sequence comprising an ultralong CDR3 sequence).

[0088] Figure 5 shows exemplary light chain variable region sequences designated VI1-47, VI1-40*1, VI1-51*01, and VI2-18*02 that are suitable for modification or use with an ultralong CDR 3 sequence.

[0089] Figure 6A-6D shows exemplary heavy chain and light chain sequences of human HIV-1 neutralizing antibodies.

[0090] Figure 7A-7C shows exemplary linkers and non-antibody sequences.

DETAILED DESCRIPTION

[0091] The present disclosure provides antibody heavy chain variable regions comprising ultralong CDR3 sequences, and antibodies or binding fragments thereof comprising the antibody heavy chain variable regions, along with materials for (*e.g.*, protein sequences, genetic sequences, cells, libraries) and methods of making the antibodies (*e.g.*, humanizing methods, library methods). Such antibodies may be useful for the treatment or prevention of a variety of disease states or disorders.

[0092] The present disclosure also provides antibody heavy chain variable regions, antibodies or binding fragments thereof comprising a sequence of the formula V1-X-V2, wherein V1 comprises an amino acid sequence selected from the group consisting of:
(i) QVQLVQSGAEVKKPGSSVKVSCKASGNSFSNHDVHWVRQATGQGLEWMGWMSHE GDKTGLAQKFQGRVTITRDSGASTVYMELRGLTADDTAIYYCLT (SEQ ID NO: 496),
(ii) QRLVESGGGVVQPGSSLRLSCAASGFDFSRQGMHWVRQAPGQGLEWVAFIKYDGS EKYHADSVWGRLSISRDNSKDTLYLQMNSLRVEDTATYFCVR (SEQ ID NO: 497),
(iii) QEQLVESGGGVVQPGGSLRLSCLASGFTFHKYGMHWVRQAPGKGLEWVALISDDG MRKYHSDSMWGRVTISRDNSKNTLYLQFSSLKVEDTAMFFCAR (SEQ ID NO: 498),
(iv) EVQLVESGGGLIRPGGSLRLSCKGSGFIFENFGFGWVRQGPGKGLEWVSGTNWNG GDSRYGDSVKGRFTISRDNSNNFVYLQMNSLRPEDTAIYYCAR (SEQ ID NO: 499), and
(v) EVQLVESGGNVVQPGGSLRLSCTASGFSFDDSTMHWVRQAPGKGLQWVSLISWNG GRTYYADSVKGRFTISRDNSKNSLYLQMNSLKTEDTAFYFCAK (SEQ ID NO: 500);
wherein X comprises an ultralong CDR3, which can include a non-human sequence or a non-antibody sequence (e.g., a non-antibody human sequence) that has been inserted into the CDR3 sequence of the antibody, including optionally, removing a portion of CDR3 (e.g., one or more amino acids of the CDR3) or the entire CDR3 sequence (e.g., all or substantially all of the amino acids of the CDR3); and wherein V2 comprises an amino acid sequence selected from the group consisting of: (i) WGHGTAVTVSS (SEQ ID NO: 570),

(ii) WGKGTTVTVSS (SEQ ID NO: 571), (iii) WGKGTTVTVSS (SEQ ID NO: 572), (iv) WGRGTLVTVSS (SEQ ID NO: 573), and (v) WGKGTTVTVSS (SEQ ID NO: 574) (*see*, Figure 6C).

[0093] The present disclosure also provides antibody heavy chain variable regions, antibodies or binding fragments thereof comprising an ultralong CDR3, wherein the ultralong CDR3 comprises an amino acid sequence of: (i) any one of GSKHRLRDYFLYNE (SEQ ID NO: 501), GSKHRLRDYFLYN (SEQ ID NO: 502), GSKHRLRDYFLY (SEQ ID NO: 503), GSKHRLRDYFL (SEQ ID NO: 504), GSKHRLRDYF (SEQ ID NO: 505), GSKHRLRDY (SEQ ID NO: 506), or GSKHRLRD (SEQ ID NO: 507); (ii) any one of EAGGPDYRNGYNY (SEQ ID NO: 508), EAGGPDYRNGYN (SEQ ID NO: 509), EAGGPDYRNGY (SEQ ID NO: 510), EAGGPDYRNG (SEQ ID NO: 511), EAGGPDYRN (SEQ ID NO: 512), EAGGPDYR (SEQ ID NO: 513), EAGGPDY (SEQ ID NO: 514), or EAGGPD (SEQ ID NO: 515); (iii) any one of EAGGPIWHDDVKY (SEQ ID NO: 516), EAGGPIWHDDVK (SEQ ID NO: 517), EAGGPIWHDDV (SEQ ID NO: 518), EAGGPIWHDD (SEQ ID NO: 519), EAGGPIWHD (SEQ ID NO: 520), EAGGPIWH (SEQ ID NO: 521), EAGGPIW (SEQ ID NO: 522), or EAGGPI (SEQ ID NO: 523); (iv) any one of GTDYTIDDQGI (SEQ ID NO: 524), GTDYTIDDQG (SEQ ID NO: 525), GTDYTIDDQ (SEQ ID NO: 526), GTDYTIDD (SEQ ID NO: 527), GTDYTID (SEQ ID NO: 528), or GTDYTI (SEQ ID NO: 529); or (v) any one of DKGDSDYDYNL (SEQ ID NO: 530), DKGDSDYDYN (SEQ ID NO: 531), DKGDSDYDY (SEQ ID NO: 532), DKGDSDYD (SEQ ID NO: 533), DKGDSDY (SEQ ID NO: 534), DKGDSD (SEQ ID NO: 535) (*see*, Figure 6C).

[0094] The present disclosure also provides antibody heavy chain variable regions, antibodies or binding fragments thereof comprising an ultralong CDR3, wherein the ultralong CDR3 comprises an amino acid sequence of: (i) any one of YGPNYEEWGDYLATLDV (SEQ ID NO: 536), GPNYEEWGDYLATLDV (SEQ ID NO: 537), PNYEEWGDYLATLDV (SEQ ID NO: 538), NYEEWGDYLATLDV (SEQ ID NO: 539), YEEWGDYLATLDV (SEQ ID NO: 540), or EEWGDYLATLDV (SEQ ID NO: 541); (ii) any one of YDFYDGYYNYHYMDV (SEQ ID NO: 542), DFYDGYYNYHYMDV (SEQ ID NO: 543), FYDGYYNYHYMDV (SEQ ID NO: 544), YDGYYNYHYMDV (SEQ ID NO: 545), DGYYNYHYMDV (SEQ ID NO: 546), GYYNYHYMDV (SEQ ID NO: 547), or YYNYHYMDV (SEQ ID NO: 548); (iii) any one of YDFNDGYYNYHYMDV (SEQ ID NO: 549), DFYDGYYNYHYMDV (SEQ ID NO: 550), FYDGYYNYHYMDV (SEQ ID NO: 551), YDGYYNYHYMDV (SEQ ID NO: 552), DGYYNYHYMDV (SEQ ID NO: 553), or GYYNYHYMDV (SEQ ID NO: 554); (iv) any one of QGIRYQGSGTFWYFDV (SEQ ID NO: 555), GIRYQGSGTFWYFDV (SEQ ID NO: 556), IRYQGSGTFWYFDV (SEQ ID NO: 557), RYQGSGTFWYFDV (SEQ ID NO: 558), YQGSGTFWYFDV (SEQ ID NO: 559),

QGSGTFWYFDV (SEQ ID NO: 560), GSGTFWYFDV (SEQ ID NO: 561), SGTFWYFDV (SEQ ID NO: 562), or GTFWYFDV (SEQ ID NO: 563); or (v) any one of YNLGYSYFYYMDG(SEQ ID NO: 564), NLGYSYFYYMDG (SEQ ID NO: 565), LGYSYFYYMDG (SEQ ID NO: 566), GYSYFYYMDG (SEQ ID NO: 567), YSYFYYMDG (SEQ ID NO: 568), or SYFYYMDG (SEQ ID NO: 569) (*see*, Figure 6C).

[0095] The present disclosure also provides antibody heavy chain variable regions, antibodies or binding fragments thereof comprising an ultralong CDR3, wherein the ultralong CDR3 comprises an amino acid sequence of: (i) any one of GSKHRLRDYFLYNE (SEQ ID NO: 501), GSKHRLRDYFLYN (SEQ ID NO: 502), GSKHRLRDYFLY (SEQ ID NO: 503), GSKHRLRDYFL (SEQ ID NO: 504), GSKHRLRDYF (SEQ ID NO: 505), GSKHRLRDY (SEQ ID NO: 506), or GSKHRLRD (SEQ ID NO: 507), and any one of YGPNYEEWGDYLATLDV (SEQ ID NO: 536), GPNYEEWGDYLATLDV (SEQ ID NO: 537), PNYEEWGDYLATLDV (SEQ ID NO: 538), NYEEWGDYLATLDV (SEQ ID NO: 539), YEEWGDYLATLDV (SEQ ID NO: 540), or EEWGDYLATLDV (SEQ ID NO: 541); (ii) any one of EAGGPDYRNGYNY (SEQ ID NO: 508), EAGGPDYRNGYN (SEQ ID NO: 509), EAGGPDYRNGY (SEQ ID NO: 510), EAGGPDYRNG (SEQ ID NO: 511), EAGGPDYRN (SEQ ID NO: 512), EAGGPDYR (SEQ ID NO: 513), EAGGPDY (SEQ ID NO: 514), or EAGGPD (SEQ ID NO: 515), and any one of YDFYDGYYNYHYMDV (SEQ ID NO: 542), DFYDGYYNYHYMDV (SEQ ID NO: 543), FYDGYYNYHYMDV (SEQ ID NO: 544), YDGYYNYHYMDV (SEQ ID NO: 545), DGYYNYHYMDV (SEQ ID NO: 546), GYYNYHYMDV (SEQ ID NO: 547), or YYNYHYMDV (SEQ ID NO: 548); (iii) any one of EAGGPIWHDDVKY (SEQ ID NO: 516), EAGGPIWHDDVK (SEQ ID NO: 517), EAGGPIWHDDV (SEQ ID NO: 518), EAGGPIWHDD (SEQ ID NO: 519), EAGGPIWHD (SEQ ID NO: 520), EAGGPIWH (SEQ ID NO: 521), EAGGPIW (SEQ ID NO: 522), or EAGGPI (SEQ ID NO: 523), and any one of YDFNDGYYNYHYMDV (SEQ ID NO: 549), DFYDGYYNYHYMDV (SEQ ID NO: 550), FYDGYYNYHYMDV (SEQ ID NO: 551), YDGYYNYHYMDV (SEQ ID NO: 552), DGYYNYHYMDV (SEQ ID NO: 553), or GYYNYHYMDV (SEQ ID NO: 554); (iv) any one of GTDYTIDDQGI (SEQ ID NO: 524), GTDYTIDDQG (SEQ ID NO: 525), GTDYTIDDQ (SEQ ID NO: 526), GTDYTIDD (SEQ ID NO: 527), GTDYTID (SEQ ID NO: 528), or GTDYTI (SEQ ID NO: 529), and any one of QGIRYQGSGTFWYFDV (SEQ ID NO: 555), GIRYQGSGTFWYFDV (SEQ ID NO: 556), IRYQGSGTFWYFDV (SEQ ID NO: 557), RYQGSGTFWYFDV (SEQ ID NO: 558), YQGSGTFWYFDV (SEQ ID NO: 559), QGSGTFWYFDV (SEQ ID NO: 560), GSGTFWYFDV (SEQ ID NO: 561), SGTFWYFDV (SEQ ID NO: 562), or GTFWYFDV (SEQ ID NO: 563); or (v) any one of YNLGYSYFYYMDG(SEQ ID NO: 564), NLGYSYFYYMDG (SEQ ID NO: 565), LGYSYFYYMDG (SEQ ID NO: 566),

GYSYFYYMDG (SEQ ID NO: 567), YSYFYYMDG (SEQ ID NO: 568), or SYFYYMDG (SEQ ID NO: 569), and any one of YNLGYSYFYYMDG(SEQ ID NO: 564), NLGYSYFYYMDG (SEQ ID NO: 565), LGYSYFYYMDG (SEQ ID NO: 566), GYSYFYYMDG (SEQ ID NO: 567), YSYFYYMDG (SEQ ID NO: 568), or SYFYYMDG (SEQ ID NO: 569) (*see*, Figure 6C).

[0096] The present disclosure also provides antibody heavy chain variable regions, antibodies or binding fragments comprising a sequence of the formula V1-X-V2, wherein V1 comprises an amino acid sequence of QVQLVQSGAEVKKPGSSVKVSCKASGNSFSNHDVHWVRQATGQGLEWMGWMSHEGD KTGLAQKFQGRVTITRDSGASTVYMELRGLTADDTAIYYCLT (SEQ ID NO: 496), wherein the ultralong CDR3 comprises an amino acid sequence of any one of GSKHRLRDYFLYNE (SEQ ID NO: 501), GSKHRLRDYFLYN (SEQ ID NO: 502), GSKHRLRDYFLY (SEQ ID NO: 503), GSKHRLRDYFL (SEQ ID NO: 504), GSKHRLRDYF (SEQ ID NO: 505), GSKHRLRDY (SEQ ID NO: 506), or GSKHRLRD (SEQ ID NO: 507), and an amino acid sequence of any one of YGPNYEEWGDYLATLDV (SEQ ID NO: 536), GPNYEEWGDYLATLDV (SEQ ID NO: 537), PNYEEWGDYLATLDV (SEQ ID NO: 538), NYEEWGDYLATLDV (SEQ ID NO: 539), YEEWGDYLATLDV (SEQ ID NO: 540), or EEWGDYLATLDV (SEQ ID NO: 541), and wherein V2 comprises an amino acid sequence selected of WGHGTAVTVSS (SEQ ID NO: 570) (*see*, Figure 6C).

[0097] The present disclosure also provides antibody heavy chain variable regions, antibodies or binding fragments comprising a sequence of the formula V1-X-V2, wherein V1 comprises an amino acid sequence of QRLVESGGGVVQPGSSLRLSCAASGFDFSRQGMHWVRQAPGQGLEWVAFIKYDGSEK YHADSVWGRLSISRDNSKDTLYLQMNSLRVEDTATYFCVR (SEQ ID NO: 497), wherein the ultralong CDR3 comprises an amino acid sequence of any one of EAGGPDYRNGYNY (SEQ ID NO: 508), EAGGPDYRNGYN (SEQ ID NO: 509), EAGGPDYRNGY (SEQ ID NO: 510), EAGGPDYRNG (SEQ ID NO: 511), EAGGPDYRN (SEQ ID NO: 512), EAGGPDYR (SEQ ID NO: 513), EAGGPDY (SEQ ID NO: 514), or EAGGPD (SEQ ID NO: 515), and an amino acid sequence of any one of YDFYDGYYNYHYMDV (SEQ ID NO: 542), DFYDGYYNYHYMDV (SEQ ID NO: 543), FYDGYYNYHYMDV (SEQ ID NO: 544), YDGYYNYHYMDV (SEQ ID NO: 545), DGYYNYHYMDV (SEQ ID NO: 546), GYYNYHYMDV (SEQ ID NO: 547), or YYNYHYMDV (SEQ ID NO: 548), and wherein V2 comprises an amino acid sequence selected of WGKGTTVTVSS (SEQ ID NO: 571) (*see*, Figure 6C).

[0098] The present disclosure also provides antibody heavy chain variable regions, antibodies or binding fragments comprising a sequence of the formula V1-X-V2, wherein

V1 comprises an amino acid sequence of QEQLVESGGGVVQPGGSLRLSCLASGFTFHKYGMHWVRQAPGKGLEWVALISDDGMRK YHSDSMWGRVTISRDNSKNTLYLQFSSLKVEDTAMFFCAR (SEQ ID NO: 498), wherein the ultralong CDR3 comprises an amino acid sequence of any one of EAGGPIWHDDVKY (SEQ ID NO: 516), EAGGPIWHDDVK (SEQ ID NO: 517), EAGGPIWHDDV (SEQ ID NO: 518), EAGGPIWHDD (SEQ ID NO: 519), EAGGPIWHD (SEQ ID NO: 520), EAGGPIWH (SEQ ID NO: 521), EAGGPIW (SEQ ID NO: 522), or EAGGPI (SEQ ID NO: 523), and an amino acid sequence of any one of YDFNDGYYNYHYMDV (SEQ ID NO: 549), DFYDGYYNYHYMDV (SEQ ID NO: 550), FYDGYYNYHYMDV (SEQ ID NO: 551), YDGYYNYHYMDV (SEQ ID NO: 552), DGYYNYHYMDV (SEQ ID NO: 553), or GYYNYHYMDV (SEQ ID NO: 554) , and wherein V2 comprises an amino acid sequence selected of WGKGTTVTVSS (SEQ ID NO: 572) (*see*, Figure 6C).

[0099] The present disclosure also provides antibody heavy chain variable regions, antibodies or binding fragments comprising a sequence of the formula V1-X-V2, wherein V1 comprises an amino acid sequence of EVQLVESGGGLIRPGGSLRLSCKGSGFIFENFGFGWVRQGPGKGLEWVSGTNWNGGDS RYGDSVKGRFTISRDNSNNFVYLQMNSLRPEDTAIYYCAR (SEQ ID NO: 499), wherein the ultralong CDR3 comprises an amino acid sequence of any one of GTDYTIDDQGI (SEQ ID NO: 524), GTDYTIDDQG (SEQ ID NO: 525), GTDYTIDDQ (SEQ ID NO: 526), GTDYTIDD (SEQ ID NO: 527), GTDYTID (SEQ ID NO: 528), or GTDYTI (SEQ ID NO: 529), and an amino acid sequence of any one of QGIRYQGSGTFWYFDV (SEQ ID NO: 555), GIRYQGSGTFWYFDV (SEQ ID NO: 556), IRYQGSGTFWYFDV (SEQ ID NO: 557), RYQGSGTFWYFDV (SEQ ID NO: 558), YQGSGTFWYFDV (SEQ ID NO: 559), QGSGTFWYFDV (SEQ ID NO: 560), GSGTFWYFDV (SEQ ID NO: 561), SGTFWYFDV (SEQ ID NO: 562), or GTFWYFDV (SEQ ID NO: 563), and wherein V2 comprises an amino acid sequence selected of WGRGTLVTVSS (SEQ ID NO: 573) (*see*, Figure 6C).

[00100] The present disclosure also provides antibody heavy chain variable regions, antibodies or binding fragments comprising a sequence of the formula V1-X-V2, wherein V1 comprises an amino acid sequence of EVQLVESGGNVVQPGGSLRLSCTASGFSFDDSTMHWVRQAPGKGLQWVSLISWNGGR TYYADSVKGRFTISRDNSKNSLYLQMNSLKTEDTAFYFCAK (SEQ ID NO: 500), wherein the ultralong CDR3 comprises an amino acid sequence of any one of YNLGYSYFYYMDG(SEQ ID NO: 564), NLGYSYFYYMDG (SEQ ID NO: 565), LGYSYFYYMDG (SEQ ID NO: 566), GYSYFYYMDG (SEQ ID NO: 567), YSYFYYMDG (SEQ ID NO: 568), or SYFYYMDG (SEQ ID NO: 569), and any one of YNLGYSYFYYMDG(SEQ ID NO: 564), NLGYSYFYYMDG (SEQ ID NO: 565),

LGYSYFYYMDG (SEQ ID NO: 566), GYSYFYYMDG (SEQ ID NO: 567), YSYFYYMDG (SEQ ID NO: 568), or SYFYYMDG (SEQ ID NO: 569), wherein V2 comprises an amino acid sequence selected of WGKGTTVTVSS (SEQ ID NO: 574) (*see*, Figure 6C).

[00101] The present disclosure also provides antibody heavy chain variable regions, antibodies or binding fragments comprising an ultralong CDR3, wherein the ultralong CDR3 comprises an amino acid sequence selected from the group consisting of: (i) GSKHRLRDYFL (SEQ ID NO: 504) and YEEWGDYLATLDV (SEQ ID NO: 540); (ii) GTDYTID (SEQ ID NO: 528) and GIRYQGSGTFWYFDV (SEQ ID NO: 556); and (iii) DKGDSDYD (SEQ ID NO: 533) and GYSYFYYMDG (SEQ ID NO: 567) (*see*, Figure 6C).

[00102] The present disclosure also provides antibody heavy chain variable regions, antibodies or binding fragments thereof comprising ultralong CDR3 sequences, wherein the CDR3 sequences are 35 amino acids in length or longer (*e.g.*, 40 or longer, 45 or longer, 50 or longer, 55 or longer, 60 or longer) and/or wherein the CDR3 sequences have at least 3 cysteine residues or more (*e.g.*, 3 or more cysteine residues, 4 or more cysteine residues, 5 or more cysteine residues, 6 or more cysteine residues, 7 or more cysteine residues, 8 or more cysteine residues, 9 or more cysteine residues, 10 or more cysteine residues, 11 or more cysteine residues, or 12 or more cysteine residues). Such antibodies, as described herein, bind (*e.g.*, specifically or selectively bind) a variety of targets, including, for example protein targets such as transmembrane proteins (*e.g.*, GPCRs, ion channels, transporter, cell surface receptors).

[00103] The present disclosure also provides methods and materials for the preparation or making of antibodies comprising ultralong CDR3 sequences. Such materials include proteins, genetic sequences, cells and libraries. Such methods include methods of humanization and method of making and screening libraries.

[00104] The present disclosure provides an antibody heavy chain variable region, antibody or binding fragment thereof comprising an ultralong CDR3. In some embodiments, the ultralong CDR3 may be 35 amino acids in length or longer, 40 amino acids in length or longer, 45 amino acids in length or longer, 50 amino acids in length or longer, 55 amino acids in length or longer, or 60 amino acids in length or longer. In some embodiments, the ultralong CDR3 may comprise 3 or more cysteine residues, 4 or more cysteine residues, 5 or more cysteine residues, 6 or more cysteine residues, 7 or more cysteine residues, 8 or more cysteine residues, 9 or more cysteine residues, 10 or more cysteine residues, 11 or more cysteine residues, or 12 or more cysteine residues. The ultralong CDR3 may comprise a cysteine motif including, for example, where the cysteine motif is selected from the group consisting of: $CX_{10}CX_5CX_5CXCX_7C$ (SEQ ID NO: 41), $CX_{10}CX_6CX_5CXCX_{15}C$ (SEQ ID NO: 42), $CX_{11}CXCX_5C$ (SEQ ID NO: 43),

$CX_{11}CX_5CX_5CXCX_7C$ (SEQ ID NO: 44), $CX_{10}CX_6CX_5CXCX_{13}C$ (SEQ ID NO: 45), $CX_{10}CX_5CXCX_4CX_8C$ (SEQ ID NO: 46), $CX_{10}CX_6CX_6CXCX_7C$ (SEQ ID NO: 47), $CX_{10}CX_4CX_7CXCX_8C$ (SEQ ID NO: 48), $CX_{10}CX_4CX_7CXCX_7C$ (SEQ ID NO: 49), $CX_{13}CX_8CX_8C$ (SEQ ID NO: 50), $CX_{10}CX_6CX_5CXCX_7C$ (SEQ ID NO: 51), $CX_{10}CX_5CX_5C$ (SEQ ID NO: 52), $CX_{10}CX_5CX_6CXCX_7C$ (SEQ ID NO: 53), $CX_{10}CX_6CX_5CX_7CX_9C$ (SEQ ID NO: 54), $CX_9CX_7CX_5CXCX_7C$ (SEQ ID NO: 55), $CX_{10}CX_6CX_5CXCX_9C$ (SEQ ID NO: 56), $CX_{10}CXCX_4CX_5CX_{11}C$ (SEQ ID NO: 57), $CX_7CX_3CX_6CX_5CXCX_5CX_{10}C$ (SEQ ID NO: 58), $CX_{10}CXCX_4CX_5CXCX_2CX_3C$ (SEQ ID NO: 59), $CX_{16}CX_5CXC$ (SEQ ID NO: 60), $CX_6CX_4CXCX_4CX_5C$ (SEQ ID NO: 61), $CX_{11}CX_4CX_5CX_6CX_3C$ (SEQ ID NO: 62), $CX_8CX_2CX_6CX_5C$ (SEQ ID NO: 63), $CX_{10}CX_5CX_5CXCX_{10}C$ (SEQ ID NO: 64), $CX_{10}CXCX_6CX_4CXC$ (SEQ ID NO: 65), $CX_{10}CX_5CX_5CXCX_2C$ (SEQ ID NO: 66), $CX_{14}CX_2CX_3CXCXC$ (SEQ ID NO: 67), $CX_{15}CX_5CXC$ (SEQ ID NO: 68), $CX_4CX_6CX_9CX_2CX_{11}C$ (SEQ ID NO: 69), $CX_6CX_4CX_5CX_5CX_{12}C$ (SEQ ID NO: 70), $CX_7CX_3CXCXCX_4CX_5CX_9C$ (SEQ ID NO: 71), $CX_{10}CX_6CX_5C$ (SEQ ID NO: 72), $CX_7CX_3CX_5CX_5CX_9C$ (SEQ ID NO: 73), $CX_7CX_5CXCX_2C$ (SEQ ID NO: 74), $CX_{10}CXCX_6C$ (SEQ ID NO: 75), $CX_{10}CX_3CX_3CX_5CX_7CXCX_6C$ (SEQ ID NO: 76), $CX_{10}CX_4CX_5CX_{12}CX_2C$ (SEQ ID NO: 77), $CX_{12}CX_4CX_5CXCXCX_9CX_3C$ (SEQ ID NO: 78), $CX_{12}CX_4CX_5CX_{12}CX_2C$ (SEQ ID NO: 79), $CX_{10}CX_6CX_5CXCX_{11}C$ (SEQ ID NO: 80), $CX_{16}CX_5CXCXCX_{14}C$ (SEQ ID NO: 81), $CX_{10}CX_5CXCX_8CX_6C$ (SEQ ID NO: 82), $CX_{12}CX_4CX_5CX_8CX_2C$ (SEQ ID NO: 83), $CX_{12}CX_5CX_5CXCX_8C$ (SEQ ID NO: 84), $CX_{10}CX_6CX_5CXCX_4CXCX_9C$ (SEQ ID NO: 85), $CX_{11}CX_4CX_5CX_8CX_2C$ (SEQ ID NO: 86), $CX_{10}CX_6CX_5CX_8CX_2C$ (SEQ ID NO: 87), $CX_{10}CX_6CX_5CXCX_8C$ (SEQ ID NO: 88), $CX_{10}CX_6CX_5CXCX_3CX_8CX_2C$ (SEQ ID NO: 89), $CX_{10}CX_6CX_5CX_3CX_8C$ (SEQ ID NO: 90), $CX_{10}CX_6CX_5CXCX_2CX_6CX_5C$ (SEQ ID NO: 91), $CX_7CX_6CX_3CX_3CX_9C$ (SEQ ID NO: 92), $CX_9CX_8CX_5CX_6CX_5C$ (SEQ ID NO: 93), $CX_{10}CX_2CX_2CX_7CXCX_{11}CX_5C$ (SEQ ID NO: 94), and $CX_{10}CX_6CX_5CXCX_2CX_8CX_4C$ (SEQ ID NO: 95). Alternatively, the ultralong CDR3 may comprise a cysteine motif including, for example, where the cysteine motif is selected from the group consisting of: $CCX_3CXCX_3CX_2CCXCX_5CX_9CX_5CXC$ (SEQ ID NO: 96), $CX_6CX_2CX_5CX_4CCXCX_4CX_6CXC$ (SEQ ID NO: 97), $CX_7CXCX_5CX_4CCCX_4CX_6CXC$ (SEQ ID NO: 98), $CX_9CX_3CXCX_2CXCCCX_6CX_4C$ (SEQ ID NO: 99), $CX_5CX_3CXCX_4CX_4CCX_{10}CX_2CC$ (SEQ ID NO: 100), $CX_5CXCX_1CXCX_3CCX_3CX_4CX_{10}C$ (SEQ ID NO: 101), $CX_9CCCX_3CX_4CCCX_5CX_6C$ (SEQ ID NO: 102), $CCX_8CX_5CX_4CX_3CX_4CCXCX_1C$ (SEQ ID NO: 103), $CCX_6CCX_5CCCX_4CX_4CX_{12}C$ (SEQ ID NO: 104), $CX_6CX_2CX_3CCCX_4CX_5CX_3CX_3C$ (SEQ ID NO: 105), $CX_3CX_5CX_6CX_4CCXCX_5CX_4CXC$ (SEQ ID NO: 106), $CX_4CX_4CCX_4CX_4CXCX_{11}CX_2CXC$ (SEQ ID NO: 107), $CX_5CX_2CCX_5CX_4CCX_3CCX_7C$ (SEQ ID NO: 108), $CX_5CX_5CX_3CX_2CXCCX_4CX_7CXC$ (SEQ

ID NO: 109), $CX_3CX_7CX_3CX_4CCXCX_2CX_5CX_2C$ (SEQ ID NO: 110), $CX_9CX_3CXCX_4CCX_5CCCX_6C$ (SEQ ID NO: 111), $CX_9CX_3CXCX_2CXCCX_6CX_3CX_3C$ (SEQ ID NO: 112), $CX_8CCXCX_3CCX_3CXCX_3CX_4C$ (SEQ ID NO: 113), $CX_9CCX_4CX_2CXCCXCX_4CX_3C$ (SEQ ID NO: 114), $CX_{10}CXCX_3CX_2CXCCX_4CX_5CXC$ (SEQ ID NO: 115), $CX_9CXCX_3CX_2CXCCX_4CX_5CXC$ (SEQ ID NO: 116), $CX_6CCXCX_5CX_4CCXCX_5CX_2C$ (SEQ ID NO: 117), $CX_6CCXCX_3CXCCX_3CX_4CC$ (SEQ ID NO: 118), $CX_6CCXCX_3CXCX_2CXCX_4CX_8C$ (SEQ ID NO: 119), $CX_4CX_2CCX_3CXCX_4CCX_2CX_3C$ (SEQ ID NO: 120), $CX_3CX_5CX_3CCCX_4CX_9C$ (SEQ ID NO: 121), $CCX_9CX_3CXCCX_3CX_5C$ (SEQ ID NO: 122), $CX_9CX_2CX_3CX_4CCX_4CX_5C$ (SEQ ID NO: 123), $CX_9CX_7CX_4CCXCX_7CX_3C$ (SEQ ID NO: 124), $CX_9CX_3CCCX_{10}CX_2CX_3C$ (SEQ ID NO: 125), $CX_3CX_5CX_5CX_4CCX_{10}CX_6C$ (SEQ ID NO: 126), $CX_9CX_5CX_4CCXCX_5CX_4C$ (SEQ ID NO: 127), $CX_7CXCX_6CX_4CCCX_{10}C$ (SEQ ID NO: 128), $CX_8CX_2CX_4CCX_4CX_3CX_3C$ (SEQ ID NO: 129), $CX_7CX_5CXCX_4CCX_7CX_4C$ (SEQ ID NO: 130), $CX_{11}CX_3CX_4CCCX_8CX_2C$ (SEQ ID NO: 131), $CX_2CX_3CX_4CCX_4CX_5CX_{15}C$ (SEQ ID NO: 132), $CX_9CX_5CX_4CCX_7C$ (SEQ ID NO: 133), $CX_9CX_7CX_3CX_2CX_6C$ (SEQ ID NO: 134), $CX_9CX_5CX_4CCX_{14}C$ (SEQ ID NO: 135), $CX_9CX_5CX_4CCX_8C$ (SEQ ID NO: 136), $CX_9CX_6CX_4CCXC$ (SEQ ID NO: 137), $CX_5CCX_7CX_4CX_{12}$ (SEQ ID NO: 138), $CX_{10}CX_3CX_4CCX_4C$ (SEQ ID NO: 139), $CX_9CX_4CCX_5CX_4C$ (SEQ ID NO: 140), $CX_{10}CX_3CX_4CX_7CXC$ (SEQ ID NO: 141), $CX_7CX_7CX_2CX_2CX_3C$ (SEQ ID NO: 142), $CX_9CX_4CX_4CCX_6C$ (SEQ ID NO: 143), $CX_7CXCX_3CXCX_6C$ (SEQ ID NO: 144), $CX_7CXCX_4CXCX_4C$ (SEQ ID NO: 145), $CX_9CX_5CX_4C$ (SEQ ID NO: 146), $CX_3CX_6CX_8C$ (SEQ ID NO: 147), $CX_{10}CXCX_4C$ (SEQ ID NO: 148), $CX_{10}CCX_4C$ (SEQ ID NO: 149), $CX_{15}C$ (SEQ ID NO: 150), $CX_{10}C$ (SEQ ID NO: 151), and CX_9C (SEQ ID NO: 152).

[00105] The present disclosure provides an antibody heavy chain variable region, antibody or binding fragment thereof comprising an ultralong CDR3, wherein the ultralong CDR3 comprises a $X^1X^2X^3X^4X^5$ motif, wherein X^1 is threonine (T), glycine (G), alanine (A), serine (S), or valine (V), wherein X^2 is serine (S), threonine (T), proline (P), isoleucine (I), alanine (A), valine (V), or asparagine (N), wherein X^3 is valine (V), alanine (A), threonine (T), or aspartic acid (D), wherein X^4 is histidine (H), threonine (T), arginine (R), tyrosine (Y), phenylalanine (F), or leucine (L), and wherein X^5 is glutamine (Q). In some embodiments, the $X^1X^2X^3X^4X^5$ motif may be TTVHQ (SEQ ID NO: 153), TSVHQ (SEQ ID NO: 154), SSVTQ (SEQ ID NO: 155), STVHQ (SEQ ID NO: 156), ATVRQ (SEQ ID NO: 157), TTVYQ (SEQ ID NO: 158), SPVHQ (SEQ ID NO: 159), ATVYQ (SEQ ID NO: 160), TAVYQ (SEQ ID NO: 161), TNVHQ (SEQ ID NO: 162), ATVHQ (SEQ ID NO: 163), STVYQ (SEQ ID NO: 164), TIVHQ (SEQ ID NO: 165), AIVYQ (SEQ ID NO: 166), TTVFQ (SEQ ID NO: 167), AAVFQ (SEQ ID NO: 168), GTVHQ (SEQ ID NO: 169), ASVHQ (SEQ ID NO: 170),

TAVFQ (SEQ ID NO: 171), ATVFQ (SEQ ID NO: 172), AAAHQ (SEQ ID NO: 173), VVVYQ (SEQ ID NO: 174), GTVFQ (SEQ ID NO: 175), TAVHQ (SEQ ID NO: 176), ITVHQ (SEQ ID NO: 177), ITAHQ (SEQ ID NO: 178), VTVHQ (SEQ ID NO: 179); AAVHQ (SEQ ID NO: 180), GTVYQ (SEQ ID NO: 181), TTVLQ (SEQ ID NO: 182), TTTHQ (SEQ ID NO: 183), or TTDYQ (SEQ ID NO: 184).

[00106] The present disclosure provides an antibody heavy chain variable region, antibody or binding fragment thereof comprising an ultralong CDR3, wherein the ultralong CDR3 comprises a $(X^aX^b)_z$ motif, wherein X^a is any amino acid residue, X^b is an aromatic amino acid selected from the group consisting of: tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H), and wherein z is 1 - 4. In some embodiments, the $(X^aX^b)_z$ motif may be CYTYNYEF (SEQ ID NO: 217), HYTYTYDF (SEQ ID NO: 218), HYTYTYEW (SEQ ID NO: 219), KHRYTYEW (SEQ ID NO: 220), NYIYKYSF (SEQ ID NO: 221), PYIYTYQF (SEQ ID NO: 222), SFTYTYEW (SEQ ID NO: 223), SYIYIYQW (SEQ ID NO: 224), SYNYTYSW (SEQ ID NO: 225), SYSYSYEY (SEQ ID NO: 226), SYTYNYDF (SEQ ID NO: 227), SYTYNYEW (SEQ ID NO: 228), SYTYNYQF (SEQ ID NO: 229), SYVWTHNF (SEQ ID NO: 230), TYKYVYEW (SEQ ID NO: 231), TYTYTYEF (SEQ ID NO: 232), TYTYTYEW (SEQ ID NO: 233), VFTYTYEF (SEQ ID NO: 234), AYTYEW (SEQ ID NO: 235), DYIYTY (SEQ ID NO: 236), IHSYEF (SEQ ID NO: 237), SFTYEF (SEQ ID NO: 238), SHSYEF (SEQ ID NO: 239), THTYEF (SEQ ID NO: 240), TWTYEF (SEQ ID NO: 241), TYNYEW (SEQ ID NO: 242), TYSYEF (SEQ ID NO: 243), TYSYEH (SEQ ID NO: 244), TYTYDF (SEQ ID NO: 245), TYTYEF (SEQ ID NO: 246), TYTYEW (SEQ ID NO: 247), AYEF (SEQ ID NO: 248), AYSF (SEQ ID NO: 249), AYSY (SEQ ID NO: 250), CYSF (SEQ ID NO: 251), DYTY (SEQ ID NO: 252), KYEH (SEQ ID NO: 253), KYEW (SEQ ID NO: 254), MYEF (SEQ ID NO: 255), NWIY (SEQ ID NO: 256), NYDY (SEQ ID NO: 257), NYQW (SEQ ID NO: 258), NYSF (SEQ ID NO: 259), PYEW (SEQ ID NO: 260), RYNW (SEQ ID NO: 261), RYTY (SEQ ID NO: 262), SYEF (SEQ ID NO: 263), SYEH (SEQ ID NO: 264), SYEW (SEQ ID NO: 265), SYKW (SEQ ID NO: 266), SYTY (SEQ ID NO: 267), TYDF (SEQ ID NO: 268), TYEF (SEQ ID NO: 269), TYEW (SEQ ID NO: 270), TYQW (SEQ ID NO: 271), TYTY (SEQ ID NO: 272), or VYEW (SEQ ID NO: 273).

[00107] The present disclosure provides an antibody heavy chain variable region, antibody or binding fragment thereof comprising an ultralong CDR3, wherein the ultralong CDR3 comprises a $X^1X^2X^3X^4X^5X_n(X^aX^b)_z$ motif, wherein X^1 is threonine (T), glycine (G), alanine (A), serine (S), or valine (V), wherein X^2 is serine (S), threonine (T), proline (P), isoleucine (I), alanine (A), valine (V), or asparagine (N), wherein X^3 is valine (V), alanine (A), threonine (T), or aspartic acid (D), wherein X^4 is histidine (H), threonine (T), arginine (R), tyrosine (Y), phenylalanine (F), or leucine (L), and wherein X^5 is glutamine (Q),

wherein X^a is any amino acid residue, X^b is an aromatic amino acid selected from the group consisting of: tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H), wherein n is 27 – 54, and wherein z is 1 - 4.

[00108] The present disclosure provides an antibody heavy chain variable region, antibody or binding fragment thereof comprising an ultralong CDR3, wherein the ultralong CDR3 comprises: a $CX^1X^2X^3X^4X^5$ motif, wherein X^1 is threonine (T), glycine (G), alanine (A), serine (S), or valine (V), wherein X^2 is serine (S), threonine (T), proline (P), isoleucine (I), alanine (A), valine (V), or asparagine (N), wherein X^3 is valine (V), alanine (A), threonine (T), or aspartic acid (D), wherein X^4 is histidine (H), threonine (T), arginine (R), tyrosine (Y), phenylalanine (F), or leucine (L), and wherein X^5 is glutamine (Q), a cysteine motif selected from the group consisting of: $CX_{10}CX_5CX_5CXCX_7C$ (SEQ ID NO: 41), $CX_{10}CX_6CX_5CXCX_{15}C$ (SEQ ID NO: 42), $CX_{11}CXCX_5C$ (SEQ ID NO: 43), $CX_{11}CX_5CX_5CXCX_7C$ (SEQ ID NO: 44), $CX_{10}CX_6CX_5CXCX_{13}C$ (SEQ ID NO: 45), $CX_{10}CX_5CXCX_4CX_8C$ (SEQ ID NO: 46), $CX_{10}CX_6CX_6CXCX_7C$ (SEQ ID NO: 47), $CX_{10}CX_4CX_7CXCX_8C$ (SEQ ID NO: 48), $CX_{10}CX_4CX_7CXCX_7C$ (SEQ ID NO: 49), $CX_{13}CX_8CX_8C$ (SEQ ID NO: 50), $CX_{10}CX_6CX_5CXCX_7C$ (SEQ ID NO: 51), $CX_{10}CX_5CX_5C$ (SEQ ID NO: 52), $CX_{10}CX_5CX_6CXCX_7C$ (SEQ ID NO: 53), $CX_{10}CX_6CX_5CX_7CX_9C$ (SEQ ID NO: 54), $CX_9CX_7CX_5CXCX_7C$ (SEQ ID NO: 55), $CX_{10}CX_6CX_5CXCX_9C$ (SEQ ID NO: 56), $CX_{10}CXCX_4CX_5CX_{11}C$ (SEQ ID NO: 57), $CX_7CX_3CX_6CX_5CXCX_5CX_{10}C$ (SEQ ID NO: 58), $CX_{10}CXCX_4CX_5CXCX_2CX_3C$ (SEQ ID NO: 59), $CX_{16}CX_5CXC$ (SEQ ID NO: 60), $CX_6CX_4CXCX_4CX_5C$ (SEQ ID NO: 61), $CX_{11}CX_4CX_5CX_6CX_3C$ (SEQ ID NO: 62), $CX_8CX_2CX_6CX_5C$ (SEQ ID NO: 63), $CX_{10}CX_5CX_5CXCX_{10}C$ (SEQ ID NO: 64), $CX_{10}CXCX_6CX_4CXC$ (SEQ ID NO: 65), $CX_{10}CX_5CX_5CXCX_2C$ (SEQ ID NO: 66), $CX_{14}CX_2CX_3CXCXC$ (SEQ ID NO: 67), $CX_{15}CX_5CXC$ (SEQ ID NO: 68), $CX_4CX_6CX_9CX_2CX_{11}C$ (SEQ ID NO: 69), $CX_6CX_4CX_5CX_5CX_{12}C$ (SEQ ID NO: 70), $CX_7CX_3CXCXCX_4CX_5CX_9C$ (SEQ ID NO: 71), $CX_{10}CX_6CX_5C$ (SEQ ID NO: 72), $CX_7CX_3CX_5CX_5CX_9C$ (SEQ ID NO: 73), $CX_7CX_5CXCX_2C$ (SEQ ID NO: 74), $CX_{10}CXCX_6C$ (SEQ ID NO: 75), $CX_{10}CX_3CX_3CX_5CX_7CXCX_6C$ (SEQ ID NO: 76), $CX_{10}CX_4CX_5CX_{12}CX_2C$ (SEQ ID NO: 77), $CX_{12}CX_4CX_5CXCXCX_9CX_3C$ (SEQ ID NO: 78), $CX_{12}CX_4CX_5CX_{12}CX_2C$ (SEQ ID NO: 79), $CX_{10}CX_6CX_5CXCX_{11}C$ (SEQ ID NO: 80), $CX_{16}CX_5CXCXCX_{14}C$ (SEQ ID NO: 81), $CX_{10}CX_5CXCX_8CX_6C$ (SEQ ID NO: 82), $CX_{12}CX_4CX_5CX_8CX_2C$ (SEQ ID NO: 83), $CX_{12}CX_5CX_5CXCX_8C$ (SEQ ID NO: 84), $CX_{10}CX_6CX_5CXCX_4CXCX_9C$ (SEQ ID NO: 85), $CX_{11}CX_4CX_5CX_8CX_2C$ (SEQ ID NO: 86), $CX_{10}CX_6CX_5CX_8CX_2C$ (SEQ ID NO: 87), $CX_{10}CX_6CX_5CXCX_8C$ (SEQ ID NO: 88), $CX_{10}CX_6CX_5CXCX_3CX_8CX_2C$ (SEQ ID NO: 89), $CX_{10}CX_6CX_5CX_3CX_8C$ (SEQ ID NO: 90), $CX_{10}CX_6CX_5CXCX_2CX_6CX_5C$ (SEQ ID NO: 91), $CX_7CX_6CX_3CX_3CX_9C$ (SEQ ID NO: 92), $CX_9CX_8CX_5CX_6CX_5C$ (SEQ ID NO: 93),

$CX_{10}CX_2CX_2CX_7CXCX_{11}CX_5C$ (SEQ ID NO: 94), and $CX_{10}CX_6CX_5CXCX_2CX_8CX_4C$ (SEQ ID NO: 95), and a $(X^aX^b)_z$ motif, wherein X^a is any amino acid residue, X^b is an aromatic amino acid selected from the group consisting of: tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H), and wherein z is 1 - 4.

[00109] The present disclosure provides an antibody heavy chain variable region, antibody or binding fragment thereof comprising an ultralong CDR3, wherein the ultralong CDR3 comprises: a $CX^1X^2X^3X^4X^5$ motif, wherein X^1 is threonine (T), glycine (G), alanine (A), serine (S), or valine (V), wherein X^2 is serine (S), threonine (T), proline (P), isoleucine (I), alanine (A), valine (V), or asparagine (N), wherein X^3 is valine (V), alanine (A), threonine (T), or aspartic acid (D), wherein X^4 is histidine (H), threonine (T), arginine (R), tyrosine (Y), phenylalanine (F), or leucine (L), and wherein X^5 is glutamine (Q); a cysteine motif selected from the group consisting of: wherein the cysteine motif is selected from the group consisting of: $CCX_3CXCX_3CX_2CCXCX_5CX_9CX_5CXC$ (SEQ ID NO: 96), $CX_6CX_2CX_5CX_4CCXCX_4CX_6CXC$ (SEQ ID NO: 97), $CX_7CXCX_5CX_4CCCX_4CX_6CXC$ (SEQ ID NO: 98), $CX_9CX_3CXCX_2CXCCCX_6CX_4C$ (SEQ ID NO: 99), $CX_5CX_3CXCX_4CX_4CCX_{10}CX_2CC$ (SEQ ID NO: 100), $CX_5CXCX_1CXCX_3CCX_3CX_4CX_{10}C$ (SEQ ID NO: 101), $CX_9CCCX_3CX_4CCCX_5CX_6C$ (SEQ ID NO: 102), $CCX_8CX_5CX_4CX_3CX_4CCXCX_1C$ (SEQ ID NO: 103), $CCX_6CCX_5CCCX_4CX_4CX_{12}C$ (SEQ ID NO: 104), $CX_6CX_2CX_3CCCX_4CX_5CX_3CX_3C$ (SEQ ID NO: 105), $CX_3CX_5CX_6CX_4CCXCX_5CX_4CXC$ (SEQ ID NO: 106), $CX_4CX_4CCX_4CX_4CXCX_{11}CX_2CXC$ (SEQ ID NO: 107), $CX_5CX_2CCX_5CX_4CCX_3CCX_7C$ (SEQ ID NO: 108), $CX_5CX_5CX_3CX_2CXCCX_4CX_7CXC$ (SEQ ID NO: 109), $CX_3CX_7CX_3CX_4CCXCX_2CX_5CX_2C$ (SEQ ID NO: 110), $CX_9CX_3CXCX_4CCX_5CCCX_6C$ (SEQ ID NO: 111), $CX_9CX_3CXCX_2CXCCX_6CX_3CX_3C$ (SEQ ID NO: 112), $CX_8CCXCX_3CCX_3CXCX_3CX_4C$ (SEQ ID NO: 113), $CX_9CCX_4CX_2CXCCXCX_4CX_3C$ (SEQ ID NO: 114), $CX_{10}CXCX_3CX_2CXCCX_4CX_5CXC$ (SEQ ID NO: 115), $CX_9CXCX_3CX_2CXCCX_4CX_5CXC$ (SEQ ID NO: 116), $CX_6CCXCX_5CX_4CCXCX_5CX_2C$ (SEQ ID NO: 117), $CX_6CCXCX_3CXCCX_3CX_4CC$ (SEQ ID NO: 118), $CX_6CCXCX_3CXCX_2CXCX_4CX_8C$ (SEQ ID NO: 119), $CX_4CX_2CCX_3CXCX_4CCX_2CX_3C$ (SEQ ID NO: 120), $CX_3CX_5CX_3CCCX_4CX_9C$ (SEQ ID NO: 121), $CCX_9CX_3CXCCX_3CX_5C$ (SEQ ID NO: 122), $CX_9CX_2CX_3CX_4CCX_4CX_5C$ (SEQ ID NO: 123), $CX_9CX_7CX_4CCXCX_7CX_3C$ (SEQ ID NO: 124), $CX_9CX_3CCCX_{10}CX_2CX_3C$ (SEQ ID NO: 125), $CX_3CX_5CX_5CX_4CCX_{10}CX_6C$ (SEQ ID NO: 126), $CX_9CX_5CX_4CCXCX_5CX_4C$ (SEQ ID NO: 127), $CX_7CXCX_6CX_4CCCX_{10}C$ (SEQ ID NO: 128), $CX_8CX_2CX_4CCX_4CX_3CX_3C$ (SEQ ID NO: 129), $CX_7CX_5CXCX_4CCX_7CX_4C$ (SEQ ID NO: 130), $CX_{11}CX_3CX_4CCCX_8CX_2C$ (SEQ ID NO: 131), $CX_2CX_3CX_4CCX_4CX_5CX_{15}C$ (SEQ ID NO: 132), $CX_9CX_5CX_4CCX_7C$ (SEQ ID NO: 133), $CX_9CX_7CX_3CX_2CX_6C$ (SEQ ID NO: 134),

$CX_9CX_5CX_4CCX_{14}C$ (SEQ ID NO: 135), $CX_9CX_5CX_4CCX_8C$ (SEQ ID NO: 136), $CX_9CX_6CX_4CCXC$ (SEQ ID NO: 137), $CX_5CCX_7CX_4CX_{12}$ (SEQ ID NO: 138), $CX_{10}CX_3CX_4CCX_4C$ (SEQ ID NO: 139), $CX_9CX_4CCX_5CX_4C$ (SEQ ID NO: 140), $CX_{10}CX_3CX_4CX_7CXC$ (SEQ ID NO: 141), $CX_7CX_7CX_2CX_2CX_3C$ (SEQ ID NO: 142), $CX_9CX_4CX_4CCX_6C$ (SEQ ID NO: 143), $CX_7CXCX_3CXCX_6C$ (SEQ ID NO: 144), $CX_7CXCX_4CXCX_4C$ (SEQ ID NO: 145), $CX_9CX_5CX_4C$ (SEQ ID NO: 146), $CX_3CX_6CX_8C$ (SEQ ID NO: 147), $CX_{10}CXCX_4C$ (SEQ ID NO: 148), $CX_{10}CCX_4C$ (SEQ ID NO: 149), $CX_{15}C$ (SEQ ID NO: 150), $CX_{10}C$ (SEQ ID NO: 151), and CX_9C (SEQ ID NO: 152); and a $(X^aX^b)_z$ motif, wherein X^a is any amino acid residue, X^b is an aromatic amino acid selected from the group consisting of: tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H), and wherein z is 1 - 4.

[00110] The present disclosure also provides methods of generating a library of antibodies that comprises an ultralong CDR3, comprising: combining a nucleic acid sequence encoding an ultralong CDR3 with a nucleic acid sequence encoding an antibody heavy chain variable region disclosed herein to produce nucleic acids encoding for antibody heavy chain variable regions and/or antibodies that comprises an ultralong CDR3; and expressing the nucleic acids encoding for the antibody heavy chain variable region or antibody that comprises an ultralong CDR3 to generate a library of antibodies that comprises an ultralong CDR3.

[00111] The present disclosure also provides methods of generating a library of antibodies or binding fragments thereof comprising an ultralong CDR3 that comprises a non-antibody sequence, comprising: combining a nucleic acid sequence encoding an ultralong CDR3, a nucleic acid sequence encoding a human variable region framework (FR) sequence of a human HIV-1 neutralizing antibody, such as PGT145, PG9, PG16, CHO4 or 2909, and a nucleic acid sequence encoding a non-antibody sequence to produce nucleic acids encoding antibody heavy chain variable regions, antibodies or binding fragments thereof comprising an ultralong CDR3 and a non-antibody sequence, and expressing the nucleic acids encoding antibody heavy chain variable regions, antibodies or binding fragments thereof comprising an ultralong CDR3 and a non-antibody sequence to generate a library of antibodies or binding fragments thereof comprising an ultralong CDR3 and a non-antibody sequence.

[00112] The present disclosure also provides libraries of antibodies or binding fragments thereof comprising an ultralong CDR3 that comprises a non-antibody sequence.

[00113] The present disclosure also provides methods of generating a library of antibodies or binding fragments thereof comprising an ultralong CDR3 that comprises a cysteine motif, comprising: combining a human variable region framework (FR) sequence

of a human HIV-1 neutralizing antibody, such as PGT145, PG9, PG16, CHO4 or 2909, and a nucleic acid sequence encoding an ultralong CDR3 and a cysteine motif; introducing one or more nucleotide changes to the nucleic acid sequence encoding one or more amino acid residues that are positioned between one or more cysteine residues in the cysteine motif for nucleotides encoding different amino acid residues to produce nucleic acids encoding antibodies or binding fragments thereof comprising an ultralong CDR3 and a cysteine motif with one or more nucleotide changes introduced between one or more cysteine residues in the cysteine domain; and expressing the nucleic acids encoding antibodies or binding fragments thereof comprising an ultralong CDR3 and a cysteine motif with one or more nucleotide changes introduced between one or more cysteine residues in the cysteine domain to generate a library of antibodies or binding fragments thereof comprising an ultralong CDR3 and a cysteine motif with one or more amino acid changes introduced between one or more cysteine residues in the cysteine domain.

[00114] The present disclosure also provides libraries of antibodies or binding fragments thereof comprising an ultralong CDR3 that comprises a cysteine motif, wherein the antibodies or binding fragments comprise one or more substitutions of amino acid residues that are positioned between cysteine residues in the cysteine motif.

[00115] The present disclosure also provides methods of generating a library of antibodies or binding fragments thereof comprising a bovine ultralong CDR3, comprising: combining a nucleic acid sequence encoding a human variable region framework (FR) sequence of a human HIV-1 neutralizing antibody, such as PGT145, PG9, PG16, CHO4 or 2909, and a nucleic acid encoding a bovine ultralong CDR3, and expressing the nucleic acids encoding a human variable region framework (FR) sequence and a nucleic acid encoding a bovine ultralong CDR3 to generate a library of antibodies or binding fragments thereof comprising a bovine ultralong CDR3.

[00116] The present disclosure also provides a CDR3 scaffold comprising a non-antibody sequence and/or or cystein motif, and (i) any one of GSKHRLRDYFLYNE (SEQ ID NO: 501), GSKHRLRDYFLYN (SEQ ID NO: 502), GSKHRLRDYFLY (SEQ ID NO: 503), GSKHRLRDYFL (SEQ ID NO: 504), GSKHRLRDYF (SEQ ID NO: 505), GSKHRLRDY (SEQ ID NO: 506), or GSKHRLRD (SEQ ID NO: 507), and any one of YGPNYEEWGDYLATLDV (SEQ ID NO: 536), GPNYEEWGDYLATLDV (SEQ ID NO: 537), PNYEEWGDYLATLDV (SEQ ID NO: 538), NYEEWGDYLATLDV (SEQ ID NO: 539), YEEWGDYLATLDV (SEQ ID NO: 540), or EEWGDYLATLDV (SEQ ID NO: 541); (ii) any one of EAGGPDYRNGYNY (SEQ ID NO: 508), EAGGPDYRNGYN (SEQ ID NO: 509), EAGGPDYRNGY (SEQ ID NO: 510), EAGGPDYRNG (SEQ ID NO: 511), EAGGPDYRN (SEQ ID NO: 512), EAGGPDYR (SEQ ID NO: 513), EAGGPDY (SEQ ID NO: 514), or

EAGGPD (SEQ ID NO: 515), and any one of YDFYDGYYNYHYMDV (SEQ ID NO: 542), DFYDGYYNYHYMDV (SEQ ID NO: 543), FYDGYYNYHYMDV (SEQ ID NO: 544), YDGYYNYHYMDV (SEQ ID NO: 545), DGYYNYHYMDV (SEQ ID NO: 546), GYYNYHYMDV (SEQ ID NO: 547), or YYNYHYMDV (SEQ ID NO: 548); (iii) any one of EAGGPIWHDDVKY (SEQ ID NO: 516), EAGGPIWHDDVK (SEQ ID NO: 517), EAGGPIWHDDV (SEQ ID NO: 518), EAGGPIWHDD (SEQ ID NO: 519), EAGGPIWHD (SEQ ID NO: 520), EAGGPIWH (SEQ ID NO: 521), EAGGPIW (SEQ ID NO: 522), or EAGGPI (SEQ ID NO: 523), and any one of YDFNDGYYNYHYMDV (SEQ ID NO: 549), DFYDGYYNYHYMDV (SEQ ID NO: 550), FYDGYYNYHYMDV (SEQ ID NO: 551), YDGYYNYHYMDV (SEQ ID NO: 552), DGYYNYHYMDV (SEQ ID NO: 553), or GYYNYHYMDV (SEQ ID NO: 554); (iv) any one of GTDYTIDDQGI (SEQ ID NO: 524), GTDYTIDDQG (SEQ ID NO: 525), GTDYTIDDQ (SEQ ID NO: 526), GTDYTIDD (SEQ ID NO: 527), GTDYTID (SEQ ID NO: 528), or GTDYTI (SEQ ID NO: 529), and any one of QGIRYQGSGTFWYFDV (SEQ ID NO: 555), GIRYQGSGTFWYFDV (SEQ ID NO: 556), IRYQGSGTFWYFDV (SEQ ID NO: 557), RYQGSGTFWYFDV (SEQ ID NO: 558), YQGSGTFWYFDV (SEQ ID NO: 559), QGSGTFWYFDV (SEQ ID NO: 560), GSGTFWYFDV (SEQ ID NO: 561), SGTFWYFDV (SEQ ID NO: 562), or GTFWYFDV (SEQ ID NO: 563); or (v) any one of YNLGYSYFYYMDG(SEQ ID NO: 564), NLGYSYFYYMDG (SEQ ID NO: 565), LGYSYFYYMDG (SEQ ID NO: 566), GYSYFYYMDG (SEQ ID NO: 567), YSYFYYMDG (SEQ ID NO: 568), or SYFYYMDG (SEQ ID NO: 569), and any one of YNLGYSYFYYMDG(SEQ ID NO: 564), NLGYSYFYYMDG (SEQ ID NO: 565), LGYSYFYYMDG (SEQ ID NO: 566), GYSYFYYMDG (SEQ ID NO: 567), YSYFYYMDG (SEQ ID NO: 568), or SYFYYMDG (SEQ ID NO: 569) (*see*, Figure 6C).

Proteins

[00117] The present disclosure provides antibody heavy chain variable regions comprising ultralong CDR3 sequences.

[00118] In an embodiment, the present disclosure provides an antibody heavy chain variable region comprising an ultralong CDR3, wherein the CDR3 is 35 amino acids in length or more (*e.g.*, 40 or more, 45 or more, 50 or more, 55 or more, 60 or more). Such an antibody heavy chain variable region may comprise at least 3 cysteine residues or more (*e.g.*, 4 or more, 6 or more, 8 or more) within the ultralong CDR3.

[00119] In another embodiment, the present disclosure provides an antibody heavy chain variable region comprising an ultralong CDR3, wherein the CDR3 is 35 amino acids in length or more and is derived from or based on a non-human sequence. The

ultralong CDR3 sequence may be derived from any species that naturally produces ultralong CDR3 antibodies, including ruminants such as cattle (*Bos taurus*).

[00120] In another embodiment, the present disclosure provides an antibody heavy chain variable region comprising an ultralong CDR3, wherein the CDR3 is 35 amino acids in length or more and is derived from a non-antibody sequence. The non-antibody sequence may be derived from any protein family including, but not limited to, chemokines, growth factors, peptides, cytokines, cell surface proteins, serum proteins, toxins, extracellular matrix proteins, clotting factors, secreted proteins, etc. The non-antibody sequence may be of human or non-human origin and may comprise a portion of a non-antibody protein such as a peptide or domain. The non-antibody sequence of an ultralong CDR3 may contain mutations from its natural sequence, including amino acid changes (*e.g.*, substitutions), insertions or deletions. Engineering additional amino acids at the junction between the non-antibody sequence may be done to facilitate or enhance proper folding of the non-antibody sequence within the antibody heavy chain variable region.

[00121] In another embodiment, the present disclosure provides an antibody heavy chain variable region comprising an ultralong CDR3, wherein the CDR3 is 35 amino acids in length or more and comprises at least 3 cysteine residues or more, including, for example, 4 or more, 6 or more, and 8 or more.

[00122] In another embodiment, the present disclosure provides for an antibody heavy chain variable region comprising an ultralong CDR3 wherein the CDR3 is 35 amino acids in length or more and comprises at least 3 cysteine residues or more and wherein the ultralong CDR3 is a component of a multispecific antibody. The multispecific antibody may be bispecific or comprise greater valencies.

[00123] In another embodiment, the present disclosure provides an antibody heavy chain variable region comprising an ultralong CDR3, wherein the CDR3 is 35 amino acids in length or more and comprises at least 3 cysteine residues or more, wherein the partially human ultralong CDR3 is a component of an immunoconjugate.

[00124] In another embodiment, the present disclosure provides an antibody heavy chain variable region comprising an ultralong CDR3, wherein the CDR3 is 35 amino acids in length or more and comprises at least 3 cysteine residues or more, wherein the antibody heavy chain variable region comprising an ultralong CDR3 binds to a transmembrane protein target. Such transmembrane targets may include, but are not limited to, GPCRs, ion channels, transporters, and cell surface receptors.

Genetic Sequences

[00125] The present disclosure provides genetic sequences (*e.g.*, genes, nucleic acids, polynucleotides) encoding antibody heavy chain variable regions comprising ultralong CDR sequences.

[00126] The present disclosure also provides genetic sequences (*e.g.*, genes, nucleic acids, polynucleotides) encoding an ultralong CDR3.

[00127] In an embodiment, the present disclosure provides genetic sequences encoding an antibody heavy chain variable region comprising an ultralong CDR3, wherein the CDR3 is 35 amino acids in length or more (*e.g.*, 40 or more, 45 or more, 50 or more, 55 or more, 60 or more). Such an antibody heavy chain variable region may comprise at least 3 cysteine residues or more (*e.g.*, 4 or more, 6 or more, 8 or more) within the ultralong CDR3.

[00128] In another embodiment, the present disclosure provides genetic sequences encoding an antibody heavy chain variable region comprising an ultralong CDR3, wherein the CDR3 is 35 amino acids in length or more and is derived from or based on a non-human sequence. The genetic sequences encoding the ultralong CDR3 may be derived from any species that naturally produces ultralong CDR3 antibodies, including ruminants such as cattle (*Bos taurus*).

[00129] In another embodiment, the present disclosure provides genetic sequences encoding an antibody heavy chain variable region comprising an ultralong CDR3, wherein the CDR3 is 35 amino acids in length or more and is derived from a non-antibody protein sequence. The genetic sequences encoding the non-antibody protein sequences may be derived from any protein family including, but not limited to, chemokines, growth factors, peptides, cytokines, cell surface proteins, serum proteins, toxins, extracellular matrix proteins, clotting factors, secreted proteins, etc. The non-antibody protein sequence may be of human or non-human origin and may comprise a portion of a non-antibody protein such as a peptide or domain. The non-antibody protein sequence of an ultralong CDR3 may contain mutations from its natural sequence, including amino acid changes (*e.g.*, substitutions), insertions or deletions. Engineering additional amino acids at the junction between the non-antibody sequence may be done to facilitate or enhance proper folding of the non-antibody sequence within the antibody heavy chain variable region.

[00130] In another embodiment, the present disclosure provides genetic sequences encoding an antibody heavy chain variable region comprising an ultralong CDR3, wherein the CDR3 is 35 amino acids in length or more and comprises at least 3 cysteine residues or more, including, for example, 4 or more, 6 or more, and 8 or more.

[00131] In another embodiment, the present disclosure provides genetic sequences encoding an antibody heavy chain variable region comprising an ultralong CDR3 wherein the CDR3 is 35 amino acids in length or more and comprises at least 3 cysteine residues or more and wherein the ultralong CDR3 is a component of a multispecific antibody. The multispecific antibody may be bispecific or comprise greater valencies.

[00132] In another embodiment, the present disclosure provides genetic sequences encoding an antibody heavy chain variable region comprising an ultralong CDR3, wherein the CDR3 is 35 amino acids in length or more and comprises at least 3 cysteine residues or more, wherein the ultralong CDR3 is a component of an immunoconjugate.

[00133] In another embodiment, the present disclosure provides genetic sequences encoding an antibody heavy chain variable region comprising an ultralong CDR3 wherein the CDR3 is 35 amino acids in length or more and comprises at least 3 cysteine residues or more and wherein the antibody heavy chain variable region comprising an ultralong CDR3 binds to a transmembrane protein target. Such transmembrane targets may include, but are not limited to, GPCRs, ion channels, transporters, and cell surface receptors.

Libraries and Arrays

[00134] The present disclosure provides collections, libraries, and arrays of antibodies comprising ultralong CDR3 sequences.

[00135] In an embodiment, the present disclosure provides a library or an array of antibodies comprising ultralong CDR3 sequences wherein at least two members of the library or array differ in the positions of at least one of the cysteines in the ultralong CDR3 sequence. Structural diversity may be enhanced through different numbers of cysteines in the ultralong CDR3 sequence (*e.g.*, at least 3 or more cysteine residues such as 4 or more, 6 or more and 8 or more) and/or through different disulfide bond formation, and hence different loop structures.

[00136] In another embodiment, the present disclosure provides for a library or an array of antibodies comprising ultralong CDR3 sequences wherein at least two members of the library or the array differ in at least one amino acid located between cysteines in the ultralong CDR3. In this regard, members of the library or the array can contain cysteines in the same positions of CDR3, resulting in similar overall structural folds, but with fine

differences brought about through different amino acid side chains. Such libraries or arrays may be useful for affinity maturation.

[00137] In another embodiment, the present disclosure provides libraries or arrays of antibodies comprising ultralong CDR3 sequences wherein at least two of the ultralong CDR3 sequences differ in length (*e.g.*, 35 amino acids in length or more such as 40 or more, 45 or more, 50 or more, 55 or more and 60 or more). The amino acid and cysteine content may or may not be altered between the members of the library or the array. Different lengths of ultralong CDR3 sequences may provide for unique binding sites, including, for example, due to steric differences, as a result of altered length.

[00138] In another embodiment, the present disclosure provides libraries or arrays of antibodies comprising ultralong CDR3 sequences wherein at least two members of the library differ in the human framework used to construct the antibody heavy chain variable region, antibody or binding fragment thereof comprising an ultralong CDR3.

[00139] In another embodiment, the present disclosure provides libraries or arrays of antibodies comprising ultralong CDR3 sequences wherein at least two members of the library or the array differ in having a non-antibody protein sequence that comprises a portion of the ultralong CDR3. Such libraries or arrays may contain multiple non-antibody protein sequences, including for chemokines, growth factors, peptides, cytokines, cell surface proteins, serum proteins, toxins, extracellular matrix proteins, clotting factors, secreted proteins, viral or bacterial proteins, etc. The non-antibody protein sequence may be of human or non-human origin and may be comprised of a portion of a non-antibody protein such as a peptide or domain. The non-antibody protein sequence of the ultralong CDR3 may contain mutations from its natural sequence, including amino acid changes (*e.g.*, substitutions), or insertions or deletions. Engineering additional amino acids at the junction between the non-antibody sequence within the ultralong CDR3 may be done to facilitate or enhance proper folding of the non-antibody sequence within the antibody heavy chain variable region, antibody or binding fragment thereof.

[00140] The libraries or the arrays of the present disclosure may be in several formats well known in the art. The library or the array may be an addressable library or an addressable array. The library or array may be in display format, for example, the antibody sequences may be expressed on phage, ribosomes, mRNA, yeast, or mammalian cells.

Cells

[00141] The present disclosure provides cells comprising genetic sequences encoding antibody heavy chain variable regions, antibodies or binding fragments therof comprising ultralong CDR3 sequences.

[00142] In an embodiment, the present disclosure provides cells expressing an antibody heavy chain variable region, antibody or binding fragment thereof comprising an ultralong CDR3. The cells may be prokaryotic or eukaryotic, and an antibody heavy chain variable region, antibody or binding fragment thereof comprising an ultralong CDR3 may be expressed on the cell surface or secreted into the media. When displayed on the cell surface an antibody heavy chain variable region, antibody or binding fragment thereof preferentially contains a motif for insertion into the plasmid membrane such as a membrane spanning domain at the C-terminus or a lipid attachment site. For bacterial cells, an antibody heavy chain variable region, antibody or binding fragment thereof comprising an ultralong CDR3 may be secreted into the periplasm. When the cells are eukaryotic, they may be transiently transfected with genetic sequences encoding an antibody heavy chain variable region, antibody or binding fragment thereof comprising an ultralong CDR3. Alternatively, a stable cell line or stable pools may be created by transfecting or transducing genetic sequences encoding an antibody heavy chain variable region, antibody or binding fragment thereof comprising an ultralong CDR3 by methods well known to those of skill in the art. Cells can be selected by fluorescence activated cell sorting (FACS) or through selection for a gene encoding drug resistance. Cells useful for producing antibody heavy chain variable regions, antibodies or binding fragments thereof comprising ultralong CDR3 sequences include prokaryotic cells like E. *coli*, eukaryotic cells like the yeasts *Saccharomyces cerevisiae* and *Pichia pastoris*, chinese hamster ovary (CHO) cells, monkey cells like COS-1, or human cells like HEK-293, HeLa, SP-1.

Humanization Methods

[00143] The present disclosure provides methods for making antibodies comprising ultralong CDR3 sequences, comprising the steps of engineering an ultralong CDR3 sequence derived from a non-human CDR3 into a human framework. The human framework may be of germline origin, or may be derived from non-germline (*e.g.* mutated or affinity matured) sequences. Genetic engineering techniques well known to those in the art, including as disclosed herein, may be used to generate a hybrid DNA sequence containing a human framework and a non-human ultralong CDR3. Unlike human antibodies which may be encoded by V region genes derived from one of seven families, bovine antibodies which produce ultralong CDR3 sequences appear to utilize a single V region family which may be considered to be most homologous to the human VH4 family. In an embodiment where ultralong CDR3 sequences derived from cattle are to be humanized to produce an antibody comprising an ultralong CDR3, human V region

sequences derived from the VH4 family may be genetically fused to a bovine-derived ultralong CDR3 sequence.

[00144] In an embodiment, the present disclosure provides a fusion of a human VH4 framework sequence to a bovine-derived ultralong CDR3, for example, as may be accomplished through the following steps. First, the second cysteine of a V region genetic sequence is identified along with the nucleotide sequence encoding the second cysteine. Generally, the second cysteine marks the boundary of the framework and CDR3 two residues upstream (N-terminal) of the CDR3. Second, the second cysteine in a bovine-derived V region sequence is identified which similarly marks 2 residues upstream (N-terminal) of the CDR3. Third, the genetic material encoding the human V region is combined with the genetic sequence encoding the ultralong CDR3. Thus, a genetic fusion may be made, wherein the ultralong CDR3 sequence is placed in frame of the human V region sequence. Preferably an antibody comprising an ultralong CDR3 is as near to human in amino acid composition as possible. Optionally, a J region sequence may be mutated from bovine-derived sequence to a human sequence. Also optionally, a heavy chain may be paired with a human light chain.

[00145] In another embodiment, the present disclosure provides pairing of a human ultralong CDR3 heavy chain with a non-human light chain.

[00146] In another embodiment, the present disclosure provides pairing of a heavy chain comprising an ultralong CDR3 with a human light chain. Preferably the light chain is homologous to a bovine light chain known to pair with a bovine ultralong CDR3 heavy chain. Exemplary bovine light chains are shown in Figures 4 and 5 (*e.g.*, SEQ ID NO: 36-39; and 373-376).

Library Methods

[00147] The present disclosure provides methods for making libraries comprising antibodies comprising ultralong CDR3 sequences. Methods for making libraries of spatially addressed libraries are described in WO 2010/054007. Methods of making libraries in yeast, phage, *E. coli*, or mammalian cells are well known in the art.

[00148] The present disclosure also provides methods of screening libraries of antibodies comprising ultralong CDR3 sequences.

Definitions

[00149] An “ultralong CDR3” or an “ultralong CDR3 sequence”, used interchangeably herein, comprises a CDR3 or CDR3 sequence that is not derived from a human antibody sequence. An ultralong CDR3 may be 35 amino acids in length or longer, for example, 40 amino acids in length or longer, 45 amino acids in length or longer, 50

amino acids in length or longer, 55 amino acids in length or longer, or 60 amino acids in length or longer. The length of the ultralong CDR3 may include a non-antibody sequence. An ultralong CDR3 may comprise a non-antibody sequence, including, for example, an interleukin sequence, a hormone sequence, a cytokine sequence, a toxin sequence, a lymphokine sequence, a growth factor sequence, a chemokine sequence, a toxin sequence, or combinations thereof. Preferably, the ultralong CDR3 is a heavy chain CDR3 (CDR-H3 or CDRH3). Preferably, the ultralong CDR3 comprises an amino acid sequence of: (i) any one of GSKHRLRDYFLYNE (SEQ ID NO: 501), GSKHRLRDYFLYN (SEQ ID NO: 502), GSKHRLRDYFLY (SEQ ID NO: 503), GSKHRLRDYFL (SEQ ID NO: 504), GSKHRLRDYF (SEQ ID NO: 505), GSKHRLRDY (SEQ ID NO: 506), or GSKHRLRD (SEQ ID NO: 507); (ii) any one of EAGGPDYRNGYNY (SEQ ID NO: 508), EAGGPDYRNGYN (SEQ ID NO: 509), EAGGPDYRNGY (SEQ ID NO: 510), EAGGPDYRNG (SEQ ID NO: 511), EAGGPDYRN (SEQ ID NO: 512), EAGGPDYR (SEQ ID NO: 513), EAGGPDY (SEQ ID NO: 514), or EAGGPD (SEQ ID NO: 515); (iii) any one of EAGGPIWHDDVKY (SEQ ID NO: 516), EAGGPIWHDDVK (SEQ ID NO: 517), EAGGPIWHDDV (SEQ ID NO: 518), EAGGPIWHDD (SEQ ID NO: 519), EAGGPIWHD (SEQ ID NO: 520), EAGGPIWH (SEQ ID NO: 521), EAGGPIW (SEQ ID NO: 522), or EAGGPI (SEQ ID NO: 523); (iv) any one of GTDYTIDDQGI (SEQ ID NO: 524), GTDYTIDDQG (SEQ ID NO: 525), GTDYTIDDQ (SEQ ID NO: 526), GTDYTIDD (SEQ ID NO: 527), GTDYTID (SEQ ID NO: 528), or GTDYTI (SEQ ID NO: 529); or (v) any one of DKGDSDYDYNL (SEQ ID NO: 530), DKGDSDYDYN (SEQ ID NO: 531), DKGDSDYDY (SEQ ID NO: 532), DKGDSDYD (SEQ ID NO: 533), DKGDSDY (SEQ ID NO: 534), DKGDSD (SEQ ID NO: 535). Preferably, the ultralong CDR3 comprises an amino acid sequence of: (i) any one of YGPNYEEWGDYLATLDV (SEQ ID NO: 536), GPNYEEWGDYLATLDV (SEQ ID NO: 537), PNYEEWGDYLATLDV (SEQ ID NO: 538), NYEEWGDYLATLDV (SEQ ID NO: 539), YEEWGDYLATLDV (SEQ ID NO: 540), or EEWGDYLATLDV (SEQ ID NO: 541); (ii) any one of YDFYDGYYNYHYMDV (SEQ ID NO: 542), DFYDGYYNYHYMDV (SEQ ID NO: 543), FYDGYYNYHYMDV (SEQ ID NO: 544), YDGYYNYHYMDV (SEQ ID NO: 545), DGYYNYHYMDV (SEQ ID NO: 546), GYYNYHYMDV (SEQ ID NO: 547), or YYNYHYMDV (SEQ ID NO: 548); (iii) any one of YDFNDGYYNYHYMDV (SEQ ID NO: 549), DFYDGYYNYHYMDV (SEQ ID NO: 550), FYDGYYNYHYMDV (SEQ ID NO: 551), YDGYYNYHYMDV (SEQ ID NO: 552), DGYYNYHYMDV (SEQ ID NO: 553), or GYYNYHYMDV (SEQ ID NO: 554); (iv) any one of QGIRYQGSGTFWYFDV (SEQ ID NO: 555), GIRYQGSGTFWYFDV (SEQ ID NO: 556), IRYQGSGTFWYFDV (SEQ ID NO: 557), RYQGSGTFWYFDV (SEQ ID NO: 558), YQGSGTFWYFDV (SEQ ID NO: 559), QGSGTFWYFDV (SEQ ID NO: 560), GSGTFWYFDV (SEQ ID NO: 561), SGTFWYFDV (SEQ ID NO: 562), or GTFWYFDV

(SEQ ID NO: 563); or (v) any one of YNLGYSYFYYMDG(SEQ ID NO: 564), NLGYSYFYYMDG (SEQ ID NO: 565), LGYSYFYYMDG (SEQ ID NO: 566), GYSYFYYMDG (SEQ ID NO: 567), YSYFYYMDG (SEQ ID NO: 568), or SYFYYMDG (SEQ ID NO: 569). Preferably, the ultralong CDR3 comprises an amino acid sequence of: (i) any one of GSKHRLRDYFLYNE (SEQ ID NO: 501), GSKHRLRDYFLYN (SEQ ID NO: 502), GSKHRLRDYFLY (SEQ ID NO: 503), GSKHRLRDYFL (SEQ ID NO: 504), GSKHRLRDYF (SEQ ID NO: 505), GSKHRLRDY (SEQ ID NO: 506), or GSKHRLRD (SEQ ID NO: 507), and any one of YGPNYEEWGDYLATLDV (SEQ ID NO: 536), GPNYEEWGDYLATLDV (SEQ ID NO: 537), PNYEEWGDYLATLDV (SEQ ID NO: 538), NYEEWGDYLATLDV (SEQ ID NO: 539), YEEWGDYLATLDV (SEQ ID NO: 540), or EEWGDYLATLDV (SEQ ID NO: 541); (ii) any one of EAGGPDYRNGYNY (SEQ ID NO: 508), EAGGPDYRNGYN (SEQ ID NO: 509), EAGGPDYRNGY (SEQ ID NO: 510), EAGGPDYRNG (SEQ ID NO: 511), EAGGPDYRN (SEQ ID NO: 512), EAGGPDYR (SEQ ID NO: 513), EAGGPDY (SEQ ID NO: 514), or EAGGPD (SEQ ID NO: 515), and any one of YDFYDGYYNYHYMDV (SEQ ID NO: 542), DFYDGYYNYHYMDV (SEQ ID NO: 543), FYDGYYNYHYMDV (SEQ ID NO: 544), YDGYYNYHYMDV (SEQ ID NO: 545), DGYYNYHYMDV (SEQ ID NO: 546), GYYNYHYMDV (SEQ ID NO: 547), or YYNYHYMDV (SEQ ID NO: 548); (iii) any one of EAGGPIWHDDVKY (SEQ ID NO: 516), EAGGPIWHDDVK (SEQ ID NO: 517), EAGGPIWHDDV (SEQ ID NO: 518), EAGGPIWHDD (SEQ ID NO: 519), EAGGPIWHD (SEQ ID NO: 520), EAGGPIWH (SEQ ID NO: 521), EAGGPIW (SEQ ID NO: 522), or EAGGPI (SEQ ID NO: 523), and any one of YDFNDGYYNYHYMDV (SEQ ID NO: 549), DFYDGYYNYHYMDV (SEQ ID NO: 550), FYDGYYNYHYMDV (SEQ ID NO: 551), YDGYYNYHYMDV (SEQ ID NO: 552), DGYYNYHYMDV (SEQ ID NO: 553), or GYYNYHYMDV (SEQ ID NO: 554); (iv) any one of GTDYTIDDQGI (SEQ ID NO: 524), GTDYTIDDQG (SEQ ID NO: 525), GTDYTIDDQ (SEQ ID NO: 526), GTDYTIDD (SEQ ID NO: 527), GTDYTID (SEQ ID NO: 528), or GTDYTI (SEQ ID NO: 529), and any one of QGIRYQGSGTFWYFDV (SEQ ID NO: 555), GIRYQGSGTFWYFDV (SEQ ID NO: 556), IRYQGSGTFWYFDV (SEQ ID NO: 557), RYQGSGTFWYFDV (SEQ ID NO: 558), YQGSGTFWYFDV (SEQ ID NO: 559), QGSGTFWYFDV (SEQ ID NO: 560), GSGTFWYFDV (SEQ ID NO: 561), SGTFWYFDV (SEQ ID NO: 562), or GTFWYFDV (SEQ ID NO: 563); or (v) any one of YNLGYSYFYYMDG(SEQ ID NO: 564), NLGYSYFYYMDG (SEQ ID NO: 565), LGYSYFYYMDG (SEQ ID NO: 566), GYSYFYYMDG (SEQ ID NO: 567), YSYFYYMDG (SEQ ID NO: 568), or SYFYYMDG (SEQ ID NO: 569), and any one of YNLGYSYFYYMDG(SEQ ID NO: 564), NLGYSYFYYMDG (SEQ ID NO: 565), LGYSYFYYMDG (SEQ ID NO: 566), GYSYFYYMDG (SEQ ID NO: 567), YSYFYYMDG

(SEQ ID NO: 568), or SYFYYMDG (SEQ ID NO: 569). An ultralong CDR3 may comprise at least 3 or more cysteine residues, for example, 4 or more cysteine residues, 6 or more cysteine residues, 8 or more cysteine residues, 10 or more cysteine residues, or 12 or more cysteine residues (*e.g.*, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or more). An ultralong CDR3 may comprise one or more of the following motifs: a cysteine motif, a $X^1X^2X^3X^4X^5$ motif, a $CX^1X^2X^3X^4X^5$ motif, or a $(X^aX^b)_z$ motif. A "cysteine motif" is a segment of amino acid residues in an ultralong CDR3 that comprises 3 or more cysteine residues including, 4 or more cysteine residues, 5 or more cysteine residues, 6 or more cysteine residues, 7 or more cysteine residues, 8 or more cysteine residues, 9 or more cysteine residues, 10 or more cysteine residues, 11 or more cysteine residues, or 12 or more cysteine residues. A cysteine motif may comprise an amino acid sequence selected from the group consisting of: $CX_{10}CX_5CX_5CXCX_7C$ (SEQ ID NO: 41), $CX_{10}CX_6CX_5CXCX_{15}C$ (SEQ ID NO: 42), $CX_{11}CXCX_5C$ (SEQ ID NO: 43), $CX_{11}CX_5CX_5CXCX_7C$ (SEQ ID NO: 44), $CX_{10}CX_6CX_5CXCX_{13}C$ (SEQ ID NO: 45), $CX_{10}CX_5CXCX_4CX_8C$ (SEQ ID NO: 46), $CX_{10}CX_6CX_6CXCX_7C$ (SEQ ID NO: 47), $CX_{10}CX_4CX_7CXCX_8C$ (SEQ ID NO: 48), $CX_{10}CX_4CX_7CXCX_7C$ (SEQ ID NO: 49), $CX_{13}CX_8CX_8C$ (SEQ ID NO: 50), $CX_{10}CX_6CX_5CXCX_7C$ (SEQ ID NO: 51), $CX_{10}CX_5CX_5C$ (SEQ ID NO: 52), $CX_{10}CX_5CX_6CXCX_7C$ (SEQ ID NO: 53), $CX_{10}CX_6CX_5CX_7CX_9C$ (SEQ ID NO: 54), $CX_9CX_7CX_5CXCX_7C$ (SEQ ID NO: 55), $CX_{10}CX_6CX_5CXCX_9C$ (SEQ ID NO: 56), $CX_{10}CXCX_4CX_5CX_{11}C$ (SEQ ID NO: 57), $CX_7CX_3CX_6CX_5CXCX_5CX_{10}C$ (SEQ ID NO: 58), $CX_{10}CXCX_4CX_5CXCX_2CX_3C$ (SEQ ID NO: 59), $CX_{16}CX_5CXC$ (SEQ ID NO: 60), $CX_6CX_4CXCX_4CX_5C$ (SEQ ID NO: 61), $CX_{11}CX_4CX_5CX_6CX_3C$ (SEQ ID NO: 62), $CX_8CX_2CX_6CX_5C$ (SEQ ID NO: 63), $CX_{10}CX_5CX_5CXCX_{10}C$ (SEQ ID NO: 64), $CX_{10}CXCX_6CX_4CXC$ (SEQ ID NO: 65), $CX_{10}CX_5CX_5CXCX_2C$ (SEQ ID NO: 66), $CX_{14}CX_2CX_3CXCXC$ (SEQ ID NO: 67), $CX_{15}CX_5CXC$ (SEQ ID NO: 68), $CX_4CX_6CX_9CX_2CX_{11}C$ (SEQ ID NO: 69), $CX_6CX_4CX_5CX_5CX_{12}C$ (SEQ ID NO: 70), $CX_7CX_3CXCXCX_4CX_5CX_9C$ (SEQ ID NO: 71), $CX_{10}CX_6CX_5C$ (SEQ ID NO: 72), $CX_7CX_3CX_5CX_5CX_9C$ (SEQ ID NO: 73), $CX_7CX_5CXCX_2C$ (SEQ ID NO: 74), $CX_{10}CXCX_6C$ (SEQ ID NO: 75), $CX_{10}CX_3CX_3CX_5CX_7CXCX_6C$ (SEQ ID NO: 76), $CX_{10}CX_4CX_5CX_{12}CX_2C$ (SEQ ID NO: 77), $CX_{12}CX_4CX_5CXCXCX_9CX_3C$ (SEQ ID NO: 78), $CX_{12}CX_4CX_5CX_{12}CX_2C$ (SEQ ID NO: 79), $CX_{10}CX_6CX_5CXCX_{11}C$ (SEQ ID NO: 80), $CX_{16}CX_5CXCXCX_{14}C$ (SEQ ID NO: 81), $CX_{10}CX_5CXCX_8CX_6C$ (SEQ ID NO: 82), $CX_{12}CX_4CX_5CX_8CX_2C$ (SEQ ID NO: 83), $CX_{12}CX_5CX_5CXCX_8C$ (SEQ ID NO: 84), $CX_{10}CX_6CX_5CXCX_4CXCX_9C$ (SEQ ID NO: 85), $CX_{11}CX_4CX_5CX_8CX_2C$ (SEQ ID NO: 86), $CX_{10}CX_6CX_5CX_8CX_2C$ (SEQ ID NO: 87), $CX_{10}CX_6CX_5CXCX_8C$ (SEQ ID NO: 88), $CX_{10}CX_6CX_5CXCX_3CX_8CX_2C$ (SEQ ID NO: 89), $CX_{10}CX_6CX_5CX_3CX_8C$ (SEQ ID NO: 90), $CX_{10}CX_6CX_5CXCX_2CX_6CX_5C$ (SEQ ID NO: 91),

$CX_7CX_6CX_3CX_3CX_9C$ (SEQ ID NO: 92), $CX_9CX_8CX_5CX_6CX_5C$ (SEQ ID NO: 93), $CX_{10}CX_2CX_2CX_7CXCX_{11}CX_5C$ (SEQ ID NO: 94), and $CX_{10}CX_6CX_5CXCX_2CX_8CX_4C$ (SEQ ID NO: 95). Alternatively, a cysteine motif may comprise an amino acid sequence selected from the group consisting of: $CCX_3CXCX_3CX_2CCXCX_5CX_9CX_5CXC$ (SEQ ID NO: 96), $CX_6CX_2CX_5CX_4CCXCX_4CX_6CXC$ (SEQ ID NO: 97), $CX_7CXCX_5CX_4CCCX_4CX_6CXC$ (SEQ ID NO: 98), $CX_9CX_3CXCX_2CXCCCX_6CX_4C$ (SEQ ID NO: 99), $CX_5CX_3CXCX_4CX_4CCX_{10}CX_2CC$ (SEQ ID NO: 100), $CX_5CXCX_1CXCX_3CCX_3CX_4CX_{10}C$ (SEQ ID NO: 101), $CX_9CCCX_3CX_4CCCX_5CX_6C$ (SEQ ID NO: 102), $CCX_8CX_5CX_4CX_3CX_4CCXCX_1C$ (SEQ ID NO: 103), $CCX_6CCX_5CCCX_4CX_4CX_{12}C$ (SEQ ID NO: 104), $CX_6CX_2CX_3CCCX_4CX_5CX_3CX_3C$ (SEQ ID NO: 105), $CX_3CX_5CX_6CX_4CCXCX_5CX_4CXC$ (SEQ ID NO: 106), $CX_4CX_4CCX_4CX_4CXCX_{11}CX_2CXC$ (SEQ ID NO: 107), $CX_5CX_2CCX_5CX_4CCX_3CCX_7C$ (SEQ ID NO: 108), $CX_5CX_5CX_3CX_2CXCCX_4CX_7CXC$ (SEQ ID NO: 109), $CX_3CX_7CX_3CX_4CCXCX_2CX_5CX_2C$ (SEQ ID NO: 110), $CX_9CX_3CXCX_4CCX_5CCCX_6C$ (SEQ ID NO: 111), $CX_9CX_3CXCX_2CXCCX_6CX_3CX_3C$ (SEQ ID NO: 112), $CX_8CCXCX_3CCX_3CXCX_3CX_4C$ (SEQ ID NO: 113), $CX_9CCX_4CX_2CXCCXCX_4CX_3C$ (SEQ ID NO: 114), $CX_{10}CXCX_3CX_2CXCCX_4CX_5CXC$ (SEQ ID NO: 115), $CX_9CXCX_3CX_2CXCCX_4CX_5CXC$ (SEQ ID NO: 116), $CX_6CCXCX_5CX_4CCXCX_5CX_2C$ (SEQ ID NO: 117), $CX_6CCXCX_3CXCCX_3CX_4CC$ (SEQ ID NO: 118), $CX_6CCXCX_3CXCX_2CXCX_4CX_8C$ (SEQ ID NO: 119), $CX_4CX_2CCX_3CXCX_4CCX_2CX_3C$ (SEQ ID NO: 120), $CX_3CX_5CX_3CCCX_4CX_9C$ (SEQ ID NO: 121), $CCX_9CX_3CXCCX_3CX_5C$ (SEQ ID NO: 122), $CX_9CX_2CX_3CX_4CCX_4CX_5C$ (SEQ ID NO: 123), $CX_9CX_7CX_4CCXCX_7CX_3C$ (SEQ ID NO: 124), $CX_9CX_3CCCX_{10}CX_2CX_3C$ (SEQ ID NO: 125), $CX_3CX_5CX_5CX_4CCX_{10}CX_6C$ (SEQ ID NO: 126), $CX_9CX_5CX_4CCXCX_5CX_4C$ (SEQ ID NO: 127), $CX_7CXCX_6CX_4CCCX_{10}C$ (SEQ ID NO: 128), $CX_8CX_2CX_4CCX_4CX_3CX_3C$ (SEQ ID NO: 129), $CX_7CX_5CXCX_4CCX_7CX_4C$ (SEQ ID NO: 130), $CX_{11}CX_3CX_4CCCX_8CX_2C$ (SEQ ID NO: 131), $CX_2CX_3CX_4CCX_4CX_5CX_{15}C$ (SEQ ID NO: 132), $CX_9CX_5CX_4CCX_7C$ (SEQ ID NO: 133), $CX_9CX_7CX_3CX_2CX_6C$ (SEQ ID NO: 134), $CX_9CX_5CX_4CCX_{14}C$ (SEQ ID NO: 135), $CX_9CX_5CX_4CCX_8C$ (SEQ ID NO: 136), $CX_9CX_6CX_4CCXC$ (SEQ ID NO: 137), $CX_5CCX_7CX_4CX_{12}$ (SEQ ID NO: 138), $CX_{10}CX_3CX_4CCX_4C$ (SEQ ID NO: 139), $CX_9CX_4CCX_5CX_4C$ (SEQ ID NO: 140), $CX_{10}CX_3CX_4CX_7CXC$ (SEQ ID NO: 141), $CX_7CX_7CX_2CX_2CX_3C$ (SEQ ID NO: 142), $CX_9CX_4CX_4CCX_6C$ (SEQ ID NO: 143), $CX_7CXCX_3CXCX_6C$ (SEQ ID NO: 144), $CX_7CXCX_4CXCX_4C$ (SEQ ID NO: 145), $CX_9CX_5CX_4C$ (SEQ ID NO: 146), $CX_3CX_6CX_8C$ (SEQ ID NO: 147), $CX_{10}CXCX_4C$ (SEQ ID NO: 148), $CX_{10}CCX_4C$ (SEQ ID NO: 149), $CX_{15}C$ (SEQ ID NO: 150), $CX_{10}C$ (SEQ ID NO: 151), and CX_9C (SEQ ID NO: 152). A cysteine motif is preferably positioned within an ultralong CDR3 between a $X^1X^2X^3X^4X^5$

motif and a $(X^aX^b)_z$ motif. A "$X^1X^2X^3X^4X^5$ motif" is a series of five consecutive amino acid residues in an ultralong CDR3, wherein X^1 is threonine (T), glycine (G), alanine (A), serine (S), or valine (V), wherein X^2 is serine (S), threonine (T), proline (P), isoleucine (I), alanine (A), valine (V), or asparagine (N), wherein X^3 is valine (V), alanine (A), threonine (T), or aspartic acid (D), wherein X^4 is histidine (H), threonine (T), arginine (R), tyrosine (Y), phenylalanine (F), or leucine (L), and wherein X^5 is glutamine (Q). In some embodiments, the $X^1X^2X^3X^4X^5$ motif may be TTVHQ (SEQ ID NO: 153), TSVHQ (SEQ ID NO: 154), SSVTQ (SEQ ID NO: 155), STVHQ (SEQ ID NO: 156), ATVRQ (SEQ ID NO: 157), TTVYQ (SEQ ID NO: 158), SPVHQ (SEQ ID NO: 159), ATVYQ (SEQ ID NO: 160), TAVYQ (SEQ ID NO: 161), TNVHQ (SEQ ID NO: 162), ATVHQ (SEQ ID NO: 163), STVYQ (SEQ ID NO: 164), TIVHQ (SEQ ID NO: 165), AIVYQ (SEQ ID NO: 166), TTVFQ (SEQ ID NO: 167), AAVFQ (SEQ ID NO: 168), GTVHQ (SEQ ID NO: 169), ASVHQ (SEQ ID NO: 170), TAVFQ (SEQ ID NO: 171), ATVFQ (SEQ ID NO: 172), AAAHQ (SEQ ID NO: 173), VVVYQ (SEQ ID NO: 174), GTVFQ (SEQ ID NO: 175), TAVHQ (SEQ ID NO: 176), ITVHQ (SEQ ID NO: 177), ITAHQ (SEQ ID NO: 178), VTVHQ (SEQ ID NO: 179); AAVHQ (SEQ ID NO: 180), GTVYQ (SEQ ID NO: 181), TTVLQ (SEQ ID NO: 182), TTTHQ (SEQ ID NO: 183), or TTDYQ (SEQ ID NO: 184). A "$CX^1X^2X^3X^4X^5$ motif" is a series of six consecutive amino acid residues in an ultralong CDR3, wherein the first amino acid residue is cysteine, wherein X^1 is threonine (T), glycine (G), alanine (A), serine (S), or valine (V), wherein X^2 is serine (S), threonine (T), proline (P), isoleucine (I), alanine (A), valine (V), or asparagine (N), wherein X^3 is valine (V), alanine (A), threonine (T), or aspartic acid (D), wherein X^4 is histidine (H), threonine (T), arginine (R), tyrosine (Y), phenylalanine (F), or leucine (L), and wherein X^5 is glutamine (Q). In some embodiments, the $CX^1X^2X^3X^4X^5$ motif is CTTVHQ (SEQ ID NO: 185), CTSVHQ (SEQ ID NO: 186), CSSVTQ (SEQ ID NO: 187), CSTVHQ (SEQ ID NO: 188), CATVRQ (SEQ ID NO: 189), CTTVYQ (SEQ ID NO: 190), CSPVHQ (SEQ ID NO: 191), CATVYQ (SEQ ID NO: 192), CTAVYQ (SEQ ID NO: 193), CTNVHQ (SEQ ID NO: 194), CATVHQ (SEQ ID NO: 195), CSTVYQ (SEQ ID NO: 196), CTIVHQ (SEQ ID NO: 197), CAIVYQ (SEQ ID NO: 198), CTTVFQ (SEQ ID NO: 199), CAAVFQ (SEQ ID NO: 200), CGTVHQ (SEQ ID NO: 201), CASVHQ (SEQ ID NO: 202), CTAVFQ (SEQ ID NO: 203), CATVFQ (SEQ ID NO: 204), CAAAHQ (SEQ ID NO: 205), CVVVYQ (SEQ ID NO: 206), CGTVFQ (SEQ ID NO: 207), CTAVHQ (SEQ ID NO: 208), CITVHQ (SEQ ID NO: 209), CITAHQ (SEQ ID NO: 210), CVTVHQ (SEQ ID NO: 211); CAAVHQ (SEQ ID NO: 212), CGTVYQ (SEQ ID NO: 213), CTTVLQ (SEQ ID NO: 214), CTTTHQ (SEQ ID NO: 215), or CTTDYQ (SEQ ID NO: 216). A "$(X^aX^b)_z$" motif is a repeating series of two amino acid residues in an ultralong CDR3, wherein X^a is any amino acid residue, X^b is an aromatic amino acid selected from the group consisting of: tyrosine (Y), phenylalanine

(F), tryptophan (W), and histidine (H), and wherein z is 1 - 4. In some embodiments, the $(X^aX^b)_z$ motif may comprise CYTYNYEF (SEQ ID NO: 217), HYTYTYDF (SEQ ID NO: 218), HYTYTYEW (SEQ ID NO: 219), KHRYTYEW (SEQ ID NO: 220), NYIYKYSF (SEQ ID NO: 221), PYIYTYQF (SEQ ID NO: 222), SFTYTYEW (SEQ ID NO: 223), SYIYIYQW (SEQ ID NO: 224), SYNYTYSW (SEQ ID NO: 225), SYSYSYEY (SEQ ID NO: 226), SYTYNYDF (SEQ ID NO: 227), SYTYNYEW (SEQ ID NO: 228), SYTYNYQF (SEQ ID NO: 229), SYVWTHNF (SEQ ID NO: 230), TYKYVYEW (SEQ ID NO: 231), TYTYTYEF (SEQ ID NO: 232), TYTYTYEW (SEQ ID NO: 233), VFTYTYEF (SEQ ID NO: 234), AYTYEW (SEQ ID NO: 235), DYIYTY (SEQ ID NO: 236), IHSYEF (SEQ ID NO: 237), SFTYEF (SEQ ID NO: 238), SHSYEF (SEQ ID NO: 239), THTYEF (SEQ ID NO: 240), TWTYEF (SEQ ID NO: 241), TYNYEW (SEQ ID NO: 242), TYSYEF (SEQ ID NO: 243), TYSYEH (SEQ ID NO: 244), TYTYDF (SEQ ID NO: 245), TYTYEF (SEQ ID NO: 246), TYTYEW (SEQ ID NO: 247), AYEF (SEQ ID NO: 248), AYSF (SEQ ID NO: 249), AYSY (SEQ ID NO: 250), CYSF (SEQ ID NO: 251), DYTY (SEQ ID NO: 252), KYEH (SEQ ID NO: 253), KYEW (SEQ ID NO: 254), MYEF (SEQ ID NO: 255), NWIY (SEQ ID NO: 256), NYDY (SEQ ID NO: 257), NYQW (SEQ ID NO: 258), NYSF (SEQ ID NO: 259), PYEW (SEQ ID NO: 260), RYNW (SEQ ID NO: 261), RYTY (SEQ ID NO: 262), SYEF (SEQ ID NO: 263), SYEH (SEQ ID NO: 264), SYEW (SEQ ID NO: 265), SYKW (SEQ ID NO: 266), SYTY (SEQ ID NO: 267), TYDF (SEQ ID NO: 268), TYEF (SEQ ID NO: 269), TYEW (SEQ ID NO: 270), TYQW (SEQ ID NO: 271), TYTY (SEQ ID NO: 272), or VYEW (SEQ ID NO: 273). In some embodiments, the $(X^aX^b)_z$ motif is YXYXYX. An ultralong CDR3 may comprise an amino acid sequence that is derived from or based on SEQ ID NO: 40 (*see*, *e.g.*, amino acid residues 3-6 of SEQ ID NO: 1-4; see also, *e.g.*, VH germline sequences in Figures 2A-C). A variable region that comprises an ultralong CDR3 may include an amino acid sequence that is SEQ ID NO: 1 (CTTVHQ), SEQ ID NO:2 (CTSVHQ), SEQ ID NO:3 (CSSVTQ) or SEQ ID NO: 4 (CTTVHP). Such a sequence may be derived from or based on a bovine germline VH gene sequence (*e.g.*, SEQ ID NO: 1). An ultralong CDR3 may comprise a sequence derived from or based on a non-human DH gene sequence, for example, SEQ ID NO: 5 (*see also*, *e.g.*, Koti, *et al.* (2010) *Mol. Immunol.* 47: 2119-2128), or alternative sequences such as SEQ ID NO: 6, 7, 8, 9, 10, 11 or 12 (see also, *e.g.*, DH2 germline sequences in Figures 2A-C). An ultralong CDR3 may comprise a sequence derived from or based on a JH sequence, for example, SEQ ID NO: 13 (*see also*, *e.g.*, Hosseini, *et al.* (2004) *Int. Immunol.* 16: 843-852), or alternative sequences such as SEQ ID NO: 14, 15, 16 or 17 (see also, e.g., JH1 germline sequences in Figures 2A-C). In an embodiment, an ultralong CDR3 may comprise a sequence derived from or based on a non-human VH sequence (*e.g.*, SEQ ID NO: 1, 2, 3 or 4; alternatively VH sequences in Figures 2A-C)

and/or a sequence derived from or based on a non-human DH sequence (*e.g.*, SEQ ID NO: 5, 6, 7, 8, 9, 10, 11 or 12; alternatively DH sequences in Figures 2A-C) and/or a sequence derived from or based on a JH sequence (*e.g.*, SEQ ID NO: 13, 14, 15, 16, or 17; alternatively JH sequences in Figures 2A-C), and optionally an additional sequence comprising two to six amino acids or more (*e.g.*, IR, IF, SEQ ID NO: 18, 19, 20 or 21) such as, for example, between the VH derived sequence and the DH derived sequence. In another embodiment, an ultralong CDR3 may comprise a sequence derived from or based on SEQ ID NO: 22, 23, 24, 25, 26, 27, or 28 (see also, *e.g.*, SEQ ID NOs: 276-359 in Figures 2A-C).

[00150] An "isolated" biological molecule, such as the various polypeptides, polynucleotides, and antibodies disclosed herein, refers to a biological molecule that has been identified and separated and/or recovered from at least one component of its natural environment.

[00151] "Antagonist" refers to any molecule that partially or fully blocks, inhibits, or neutralizes an activity (*e.g.*, biological activity) of a polypeptide. Also encompassed by "antagonist" are molecules that fully or partially inhibit the transcription or translation of mRNA encoding the polypeptide. Suitable antagonist molecules include, *e.g.,* antagonist antibodies or antibody fragments; fragments or amino acid sequence variants of a native polypeptide; peptides; antisense oligonucleotides; small organic molecules; and nucleic acids that encode polypeptide antagonists or antagonist antibodies. Reference to "an" antagonist encompasses a single antagonist or a combination of two or more different antagonists.

[00152] "Agonist" refers to any molecule that partially or fully mimics a biological activity of a polypeptide. Also encompassed by "agonist" are molecules that stimulate the transcription or translation of mRNA encoding the polypeptide. Suitable agonist molecules include, *e.g.,* agonist antibodies or antibody fragments; a native polypeptide; fragments or amino acid sequence variants of a native polypeptide; peptides; antisense oligonucleotides; small organic molecules; and nucleic acids that encode polypeptides agonists or antibodies. Reference to "an" agonist encompasses a single agonist or a combination of two or more different agonists.

[00153] An "isolated" antibody refers to one which has been identified and separated and/or recovered from a component of its natural environment. Contaminant components of its natural environment are materials which would interfere with diagnostic or therapeutic uses for the antibody, and may include enzymes, hormones, and other proteinaceous or nonproteinaceous solutes. In preferred embodiments, the antibody will be purified (1) to greater than 95% by weight of antibody (*e.g.,* as determined by the Lowry

method), and preferably to more than 99% by weight, (2) to a degree sufficient to obtain at least 15 residues of N-terminal or internal amino acid sequence (*e.g.,* by use of a spinning cup sequenator), or (3) to homogeneity by SDS-PAGE under reducing or nonreducing conditions (*e.g.,* using Coomassie™ blue or, preferably, silver stain). Isolated antibody includes the antibody *in situ* within recombinant cells since at least one component of the antibody's natural environment will not be present. Similarly, isolated antibody includes the antibody in medium around recombinant cells. An isolated antibody may be prepared by at least one purification step.

[00154] An "isolated" nucleic acid molecule refers to a nucleic acid molecule that is identified and separated from at least one contaminant nucleic acid molecule with which it is ordinarily associated in the natural source of the antibody nucleic acid. An isolated nucleic acid molecule is other than in the form or setting in which it is found in nature. Isolated nucleic acid molecules therefore are distinguished from the nucleic acid molecule as it exists in natural cells. However, an isolated nucleic acid molecule includes a nucleic acid molecule contained in cells that express an antibody where, for example, the nucleic acid molecule is in a chromosomal location different from that of natural cells.

[00155] Variable domain residue numbering as in Kabat or amino acid position numbering as in Kabat, and variations thereof, refers to the numbering system used for heavy chain variable domains or light chain variable domains of the compilation of antibodies in Kabat *et al.*, Sequences of Proteins of Immunological Interest, 5th Ed. Public Health Service, National Institutes of Health, Bethesda, Md. (1991). Using this numbering system, the actual linear amino acid sequence may contain fewer or additional amino acids corresponding to a shortening of, or insertion into, a FR or CDR of the variable domain. For example, a heavy chain variable domain may include a single amino acid insert (*e.g.,* residue 52a according to Kabat) after residue 52 of H2 and inserted residues (*e.g.,* residues 82a, 82b, and 82c, etc according to Kabat) after heavy chain FR residue 82. The Kabat numbering of residues may be determined for a given antibody by alignment at regions of homology of the sequence of the antibody with a "standard" Kabat numbered sequence.

[00156] "Substantially similar," or "substantially the same", refers to a sufficiently high degree of similarity between two numeric values (generally one associated with an antibody disclosed herein and the other associated with a reference/comparator antibody) such that one of skill in the art would consider the difference between the two values to be of little or no biological and/or statistical significance within the context of the biological characteristic measured by said values (*e.g.,* Kd values). The difference between said two values is preferably less than about 50%, preferably less than about 40%, preferably less

than about 30%, preferably less than about 20%, preferably less than about 10% as a function of the value for the reference/comparator antibody.

[00157] "Binding affinity" generally refers to the strength of the sum total of noncovalent interactions between a single binding site of a molecule (*e.g.,* an antibody) and its binding partner (*e.g.,* an antigen). Unless indicated otherwise, "binding affinity" refers to intrinsic binding affinity which reflects a 1:1 interaction between members of a binding pair (*e.g.,* antibody and antigen). The affinity of a molecule X for its partner Y can generally be represented by the dissociation constant. Affinity can be measured by common methods known in the art, including those described herein. Low-affinity antibodies generally bind antigen slowly and tend to dissociate readily, whereas high-affinity antibodies generally bind antigen faster and tend to remain bound longer. A variety of methods of measuring binding affinity are known in the art, any of which can be used for purposes of the present disclosure.

[00158] An "on-rate" or "rate of association" or "association rate" or "k_{on}" can be determined with a surface plasmon resonance technique such as Biacore (*e.g.,* Biacore A100, Biacore™-2000, Biacore™-3000, Biacore, Inc., Piscataway, N.J.) carboxymethylated dextran biosensor chips (CM5, Biacore Inc.) and according to the supplier's instructions.

[00159] "Vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments may be ligated. Another type of vector is a phage vector. Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they 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, "recombinant vectors"). In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. Accordingly, "plasmid" and "vector" may, at times, be used interchangeably as the plasmid is a commonly used form of vector.

[00160] "Gene" refers to a nucleic acid (*e.g.,* DNA) sequence that comprises coding sequences necessary for the production of a polypeptide, precursor, or RNA (*e.g.,* rRNA, tRNA). The polypeptide can be encoded by a full length coding sequence or by any

portion of the coding sequence so long as the desired activity or functional properties (*e.g.,* enzymatic activity, ligand binding, signal transduction, immunogenicity, *etc.*) of the full-length or fragment are retained. The term also encompasses the coding region of a structural gene and the sequences located adjacent to the coding region on both the 5' and 3' ends for a distance of about 1 kb or more on either end such that the gene corresponds to the length of the full-length mRNA. Sequences located 5' of the coding region and present on the mRNA are referred to as 5' non-translated sequences. Sequences located 3' or downstream of the coding region and present on the mRNA are referred to as 3' non-translated sequences. The term "gene" encompasses both cDNA and genomic forms of a gene. A genomic form or clone of a gene contains the coding region interrupted with non-coding sequences termed "introns" or "intervening regions" or "intervening sequences." Introns are segments of a gene that are transcribed into nuclear RNA (hnRNA); introns can contain regulatory elements such as enhancers. Introns are removed or "spliced out" from the nuclear or primary transcript; introns therefore are absent in the messenger RNA (mRNA) transcript. The mRNA functions during translation to specify the sequence or order of amino acids in a nascent polypeptide. In addition to containing introns, genomic forms of a gene can also include sequences located on both the 5' and 3' end of the sequences that are present on the RNA transcript. These sequences are referred to as "flanking" sequences or regions (these flanking sequences are located 5' or 3' to the non-translated sequences present on the mRNA transcript). The 5' flanking region can contain regulatory sequences such as promoters and enhancers that control or influence the transcription of the gene. The 3' flanking region can contain sequences that direct the termination of transcription, post transcriptional cleavage and polyadenylation.

[00161] "Polynucleotide," or "nucleic acid," as used interchangeably herein, refers to polymers of nucleotides of any length, and include DNA and RNA. The nucleotides can be deoxyribonucleotides, ribonucleotides, modified nucleotides or bases, and/or their analogs, or any substrate that can be incorporated into a polymer by DNA or RNA polymerase, or by a synthetic reaction. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and their analogs. If present, modification to the nucleotide structure may be imparted before or after assembly of the polymer. The sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after synthesis, such as by conjugation with a label. Other types of modifications include, for example, "caps", substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as, for example, those with uncharged linkages (*e.g.,* methyl phosphonates, phosphotriesters, phosphoamidates, carbamates, *etc.*) and with charged linkages (*e.g.,* phosphorothioates,

phosphorodithioates, *etc.*), those containing pendant moieties, such as, for example, proteins (*e.g.,* nucleases, toxins, antibodies, signal peptides, poly-L-lysine, *etc.*), those with intercalators (*e.g.,* acridine, psoralen, *etc.*), those containing chelators (*e.g.,* metals, radioactive metals, boron, oxidative metals, *etc.*), those containing alkylators, those with modified linkages (*e.g.,* alpha anomeric nucleic acids, *etc.*), as well as unmodified forms of the polynucleotide(s). Further, any of the hydroxyl groups ordinarily present in the sugars may be replaced, for example, by phosphonate groups, phosphate groups, protected by standard protecting groups, or activated to prepare additional linkages to additional nucleotides, or may be conjugated to solid or semi-solid supports. The 5' and 3' terminal OH can be phosphorylated or substituted with amines or organic capping group moieties of from 1 to 20 carbon atoms. Other hydroxyls may also be derivatized to standard protecting groups. Polynucleotides can also contain analogous forms of ribose or deoxyribose sugars that are generally known in the art, including, for example, 2'-O-methyl-, 2'-O-allyl, 2'-fluoro- or 2'-azido-ribose, carbocyclic sugar analogs, alpha-anomeric sugars, epimeric sugars such as arabinose, xyloses or lyxoses, pyranose sugars, furanose sugars, sedoheptuloses, acyclic analogs and a basic nucleoside analogs such as methyl riboside. One or more phosphodiester linkages may be replaced by alternative linking groups. These alternative linking groups include, but are not limited to, embodiments wherein phosphate is replaced by P(O)S("thioate"), P(S)S ("dithioate"), "(O)NR_2 ("amidate"), P(O)R, P(O)OR', CO or CH_2 ("formacetal"), in which each R or R' is independently H or substituted or unsubstituted alkyl (1-20 C) optionally containing an ether (-O-) linkage, aryl, alkenyl, cycloalkyl, cycloalkenyl or araldyl. Not all linkages in a polynucleotide need be identical. The preceding description applies to all polynucleotides referred to herein, including RNA and DNA.

[00162] "Oligonucleotide" refers to short, generally single stranded, generally synthetic polynucleotides that are generally, but not necessarily, less than about 200 nucleotides in length. The terms "oligonucleotide" and "polynucleotide" are not mutually exclusive. The description above for polynucleotides is equally and fully applicable to oligonucleotides.

[00163] "Stringent hybridization conditions" refer to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acids, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally,

stringent conditions are selected to be about 5-10ºC lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength pH. The Tm is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at Tm, 50% of the probes are occupied at equilibrium). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is at least two times background, preferably 10 times background hybridization. Exemplary stringent hybridization conditions can be as following: 50% formamide, 5xSSC, and 1% SDS, incubating at 42º C, or, 5xSSC, 1% SDS, incubating at 65º C, with wash in 0.2xSSC, and 0.1% SDS at 65º C.

[00164] "Recombinant" when used with reference to a cell, nucleic acid, protein or vector indicates that the cell, nucleic acid, protein or vector has been modified by the introduction of a heterologous nucleic acid or protein, the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified. For example, recombinant cells express genes that are not found within the native (non-recombinant) form of the cell or express native genes that are overexpressed or otherwise abnormally expressed such as, for example, expressed as non-naturally occurring fragments or splice variants. By the term "recombinant nucleic acid" herein is meant nucleic acid, originally formed *in vitro*, in general, by the manipulation of nucleic acid, *e.g.,* using polymerases and endonucleases, in a form not normally found in nature. In this manner, operably linkage of different sequences is achieved. Thus an isolated nucleic acid, in a linear form, or an expression vector formed *in vitro* by ligating DNA molecules that are not normally joined, are both considered recombinant for the purposes of this disclosure. It is understood that once a recombinant nucleic acid is made and introduced into a host cell or organism, it will replicate non-recombinantly, *e.g.,* using the *in vivo* cellular machinery of the host cell rather than *in vitro* manipulations; however, such nucleic acids, once produced recombinantly, although subsequently replicated non-recombinantly, are still considered recombinant for the purposes disclosed herein. Similarly, a "recombinant protein" is a protein made using recombinant techniques, *e.g.,* through the expression of a recombinant nucleic acid as depicted above.

[00165] "Percent (%) amino acid sequence identity" with respect to a peptide or polypeptide sequence refers to the percentage of amino acid residues in a candidate sequence that are identical with the amino acid residues in the specific peptide or polypeptide sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative

substitutions as part of the sequence identity. Alignment for purposes of determining percent amino acid sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as BLAST, BLAST-2, ALIGN or MegAlign (DNASTAR) software. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared.

[00166] "Polypeptide," "peptide," "protein," and "protein fragment" may be used interchangeably to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymers.

[00167] "Amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function similarly to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, *e.g.,* hydroxyproline, gamma-carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, *e.g.,* an alpha carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, *e.g.,* homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs can have modified R groups (*e.g.,* norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions similarly to a naturally occurring amino acid.

[00168] "Conservatively modified variants" applies to both amino acid and nucleic acid sequences. "Amino acid variants" refers to amino acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical or associated (*e.g.,* naturally contiguous) sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode most proteins. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to another of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a

polypeptide also describes silent variations of the nucleic acid. One of skill will recognize that in certain contexts each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, silent variations of a nucleic acid which encodes a polypeptide is implicit in a described sequence with respect to the expression product, but not with respect to actual probe sequences. As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" including where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles disclosed herein. Typically conservative substitutions include: 1) Alanine (A), Glycine (G); 2) Aspartic acid (D), Glutamic acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W); 7) Serine (S), Threonine (T); and 8) Cysteine (C), Methionine (M) (see, *e.g.,* Creighton, Proteins (1984)).

[00169] "Antibodies" (Abs) and "immunoglobulins" (Igs) are glycoproteins having similar structural characteristics. While antibodies may exhibit binding specificity to a specific antigen, immunoglobulins may include both antibodies and other antibody-like molecules which generally lack antigen specificity. Polypeptides of the latter kind are, for example, produced at low levels by the lymph system and at increased levels by myelomas.

[00170] "Antibody" and "immunoglobulin" are used interchangeably in the broadest sense and include monoclonal antibodies (*e.g.,* full length or intact monoclonal antibodies), polyclonal antibodies, multivalent antibodies, multispecific antibodies (*e.g.,* bispecific antibodies so long as they exhibit the desired biological activity) and may also include certain antibody fragments (as described in greater detail herein). An antibody can be human, humanized and/or affinity matured. An antibody may refer to immunoglobulins and immunoglobulin portions, whether natural or partially or wholly synthetic, such as recombinantly produced, including any portion thereof containing at least a portion of the variable region of the immunoglobulin molecule that is sufficient to form an antigen binding site. Hence, an antibody or portion thereof includes any protein having a binding domain that is homologous or substantially homologous to an immunoglobulin antigen binding site. For example, an antibody may refer to an antibody that contains two heavy chains (which

can be denoted H and H') and two light chains (which can be denoted L and L'), where each heavy chain can be a full-length immunoglobulin heavy chain or a portion thereof sufficient to form an antigen binding site (*e.g.* heavy chains include, but are not limited to, VH, chains VH-CH1 chains and VH-CH1-CH2-CH3 chains), and each light chain can be a full-length light chain or a thereof sufficient to form an antigen binding site (*e.g.* light chains include, but are not limited to, VL chains and VL-CL chains). Each heavy chain (H and H') pairs with one light chain (L and L', respectively). Typically, antibodies minimally include all or at least a portion of the variable heavy (VH) chain and/or the variable light (VL) chain. The antibody also can include all or a portion of the constant region. For example, a full-length antibody is an antibody having two full-length heavy chains (*e.g.* VH-CH1-CH2-CH3 or VH-CH1-CH2-CH3- CH4) and two full-length light chains (VL-CL) and hinge regions, such as antibodies produced by antibody secreting B cells and antibodies with the same domains that are produced synthetically. Additionally, an "antibody" refers to a protein of the immunoglobulin family or a polypeptide comprising fragments of an immunoglobulin that is capable of noncovalently, reversibly, and in a specific manner binding a corresponding antigen. An exemplary antibody structural unit comprises a tetramer. Each tetramer is composed of two identical pairs of polypeptide chains, each pair having one "light" (about 25 kD) and one "heavy" chain (about 50-70 kD), connected through a disulfide bond. The recognized immunoglobulin genes include the κ, λ, α, γ, δ, ε, and μ constant region genes, as well as the myriad immunoglobulin variable region genes. Light chains are classified as either κ or λ. Heavy chains are classified as γ, μ, α, δ, or ε, which in turn define the immunoglobulin classes, IgG, IgM, IgA, IgD, and IgE, respectively. The N-terminus of each chain defines a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The terms variable light chain (VL) and variable heavy chain (VH) refer to these regions of light and heavy chains respectively.

[00171] "Variable" refers to the fact that certain portions of the variable domains (also referred to as variable regions) differ extensively in sequence among antibodies and are used in the binding and specificity of each particular antibody for its particular antigen. However, the variability is not evenly distributed throughout the variable domains of antibodies. It is concentrated in three segments called complementarity-determining regions (CDRs) or hypervariable regions (HVRs) both in the light-chain and the heavy-chain variable domains. CDRs include those specified as Kabat, Chothia, and IMGT as shown herein within the variable region sequences. The more highly conserved portions of variable domains are called the framework (FR). The variable domains of native heavy and light chains each comprise four FR regions, largely adopting a β-sheet configuration, connected by three CDRs, which form loops connecting, and in some cases forming part

of, the β-sheet structure. The CDRs in each chain are held together in close proximity by the FR regions and, with the CDRs from the other chain, contribute to the formation of the antigen-binding site of antibodies (see Kabat *et al.*, Sequences of Proteins of Immunological Interest, Fifth Edition, National Institute of Health, Bethesda, Md. (1991)). The constant domains are not involved directly in binding an antibody to an antigen, but exhibit various effector functions, such as participation of the antibody in antibody-dependent cellular toxicity.

[00172] Papain digestion of antibodies produces two identical antigen-binding fragments, called "Fab" fragments, each with a single antigen-binding site, and a residual "Fc" fragment, whose name reflects its ability to crystallize readily. Pepsin treatment yields an $F(ab')_2$ fragment that has two antigen-combining sites and is still capable of cross-linking antigen.

[00173] "Fv" refers to an antibody fragment which contains an antigen-recognition and antigen-binding site. In a two-chain Fv species, this region consists of a dimer of one heavy and one light chain variable domain in non-covalent association. In a single chain Fv (scFv) species, one heavy chain and one light chain variable domain can be covalently linked by a flexible peptide linker such that the light and heavy chains can associate in a "dimeric" structure analogous to that in a two-chain Fv (scFv) species. It is in this configuration that the three CDRs of each variable domain interact to define an antigen-binding site on the surface of the VH-VL dimer. Collectively, the six CDRs confer antigen-binding specificity to the antibody. However, even a single variable domain (or half of an Fv comprising only three CDRs specific for an antigen) has the ability to recognize and bind antigen, although at a lower affinity than the entire binding site.

[00174] The Fab fragment also contains the constant domain of the light chain and the first constant domain (CHI) of the heavy chain. Fab' fragments differ from Fab fragments by the addition of a few residues at the carboxy terminus of the heavy chain CH1 domain including one or more cysteines from the antibody hinge region. Fab'-SH is the designation herein for Fab' in which the cysteine residue(s) of the constant domains bear a free thiol group. $F(ab')_2$ antibody fragments originally were produced as pairs of Fab' fragments which have hinge cysteines between them. Other chemical couplings of antibody fragments are also known.

[00175] The "light chains" of antibodies (immunoglobulins) from any vertebrate species can be assigned to one of two clearly distinct types, called kappa (κ) and lambda (λ), based on the amino acid sequences of their constant domains.

[00176] Depending on the amino acid sequence of the constant domain of their heavy chains, immunoglobulins can be assigned to different classes. There are five major

classes of immunoglobulins: IgA, IgD, IgE, IgG, and IgM, and several of these can be further divided into subclasses (isotypes), *e.g.,* IgG_1, IgG_2, IgG_3, IgG_4, IgA_1, and IgA_2. The heavy-chain constant domains that correspond to the different classes of immunoglobulins are called α, δ, ε, γ, and μ, respectively. The subunit structures and three-dimensional configurations of different classes of immunoglobulins are well known.

[00177] "Antibody fragments" comprise only a portion of an intact antibody, wherein the portion preferably retains at least one, preferably most or all, of the functions normally associated with that portion when present in an intact antibody. Examples of antibody fragments include Fab, Fab', F(ab')2, single-chain Fvs (scFv), Fv, dsFv, diabody, Fd and Fd' fragments Fab fragments, Fd fragments, scFv fragments, linear antibodies, single-chain antibody molecules, and multispecific antibodies formed from antibody fragments (see, for example, Methods in Molecular Biology, Vol 207: Recombinant Antibodies for Cancer Therapy Methods and Protocols (2003); Chapter 1; p 3-25, Kipriyanov). Other known fragments include, but are not limited to, scFab fragments (Hust *et al.*, BMC Biotechnology (2007), 7:14). In one embodiment, an antibody fragment comprises an antigen binding site of the intact antibody and thus retains the ability to bind antigen. In another embodiment, an antibody fragment, for example one that comprises the Fc region, retains at least one of the biological functions normally associated with the Fc region when present in an intact antibody, such as FcRn binding, antibody half life modulation, ADCC function and complement binding. In one embodiment, an antibody fragment is a monovalent antibody that has an *in vivo* half life substantially similar to an intact antibody. For example, such an antibody fragment may comprise on antigen binding arm linked to an Fc sequence capable of conferring *in vivo* stability to the fragment. For another example, an antibody fragment or antibody portion refers to any portion of a full-length antibody that is less than full length but contains at least a portion of the variable region of the antibody sufficient to form an antigen binding site (*e.g.* one or more CDRs) and thus retains the a binding specificity and/or an activity of the full-length antibody; antibody fragments include antibody derivatives produced by enzymatic treatment of full-length antibodies, as well as synthetically, *e.g.* recombinantly produced derivatives.

[00178] A "dsFv" refers to an Fv with an engineered intermolecular disulfide bond, which stabilizes the VH-VL pair.

[00179] A "Fd fragment" refers to a fragment of an antibody containing a variable domain (VH) and one constant region domain (CH1) of an antibody heavy chain.

[00180] A "Fab fragment" refers to an antibody fragment that contains the portion of the full-length antibody that would results from digestion of a full-length immunoglobulin with papain, or a fragment having the same structure that is produced synthetically, *e.g.*

recombinantly. A Fab fragment contains a light chain (containing a VL and CL portion) and another chain containing a variable domain of a heavy chain (VH) and one constant region domain portion of the heavy chain (CH1); it can be recombinantly produced.

[00181] A "F(ab')2 fragment" refers to an antibody fragment that results from digestion of an immunoglobulin with pepsin at pH 4.0-4.5, or a synthetically, *e.g.* recombinantly, produced antibody having the same structure. The F(ab')2 fragment contains two Fab fragments but where each heavy chain portion contains an additional few amino acids, including cysteine residues that form disulfide linkages joining the two fragments; it can be recombinantly produced.

[00182] A "Fab' fragment" refers to a fragment containing one half (one heavy chain and one light chain) of the F(ab')2 fragment.

[00183] A "Fd' fragment refers to a fragment of an antibody containing one heavy chain portion of a F(ab')2 fragment.

[00184] A "Fv' fragment" refers to a fragment containing only the VH and VL domains of an antibody molecule.

[00185] A "scFv fragment" refers to an antibody fragment that contains a variable light chain (VL) and variable heavy chain (VH), covalently connected by a polypeptide linker in any order. The linker is of a length such that the two variable domains are bridged without substantial interference. Exemplary linkers are (Gly-Ser)n residues with some Glu or Lys residues dispersed throughout to increase solubility.

[00186] Diabodies are dimeric scFv; diabodies typically have shorter peptide linkers than scFvs, and they preferentially dimerize.

[00187] "HsFv" refers to antibody fragments in which the constant domains normally present in a Fab fragment have been substituted with a heterodimeric coiled-coil domain (see, *e.g.*, Arndt *et al.* (2001) J Mol Biol. 7:312:221-228).

[00188] "Hypervariable region", "HVR", or "HV", as well as "complementary determing region" or "CDR", may refer to the regions of an antibody variable domain which are hypervariable in sequence and/or form structurally defined loops. Generally, antibodies comprise six hypervariable or CDR regions; three in the VH (H1, H2, H3), and three in the VL (L1, L2, L3). A number of hypervariable region or CDR delineations are in use and are encompassed herein. The Kabat Complementarity Determining Regions (Kabat CDRs) are based on sequence variability and are the most commonly used (Kabat *et al.*, Sequences of Proteins of Immunological Interest, 5th Ed. Public Health Service, National Institutes of Health, Bethesda, Md. (1991)). Chothia refers instead to the location of the structural loops (Chothia and Lesk, J. Mol. Biol. 196:901-917 (1987)). The AbM hypervariable regions represent a compromise between the Kabat CDRs and Chothia structural loops, (Chothia

"CDRs") and are used by Oxford Molecular's AbM antibody modeling software. The "contact" hypervariable regions are based on an analysis of the available complex crystal structures. The residues from each of these hypervariable regions are noted below.

Loop	Kabat	AbM	Chothia	Contact
L1	L24-L34	L24-L34	L26-L32	L30-L36
L2	L50-L56	L50-L56	L50-L52	L46-L55
L3	L89-L97	L89-L97	L91-L96	L89-L96
H1	H31-H35B	H26-H35B	H26-H32	H30-H35B
		(Kabat Numbering)		
H1	H31-H35	H26-H35	H26-H32	H30-H35
		(Chothia Numbering)		
H2	H50-H65	H50-H58	H53-H55	H47-H58
H3	H95-H102	H95-H102	H96-H101	H93-H101

[00189] IMGT referes to the international ImMunoGeneTics Information System, as described by Lefrace *et al.,* Nucl. Acids, Res. 37; D1006-D1012 (2009), including for example, IMGT designated CDRs for antibodies.

[00190] Hypervariable regions may comprise "extended hypervariable regions" as follows: 24-36 or 24-34 (L1), 46-56 or 50-56 (L2) and 89-97 (L3) in the VL and 26-35 (H1), 50-65 or 49-65 (H2) and 93-102, 94-102 or 95-102 (H3) in the VH. The variable domain residues are numbered according to Kabat *et al.*, *Supra* for each of these definitions.

[00191] "Framework" or "FR" residues are those variable domain residues other than the hypervariable region residues as herein defined. "Framework regions" (FRs) are the domains within the antibody variable region domains comprising framework residues that are located within the beta sheets; the FR regions are comparatively more conserved, in terms of their amino acid sequences, than the hypervariable regions.

[00192] "Monoclonal antibody" refers to an antibody from a population of substantially homogeneous antibodies, that is, for example, the individual antibodies comprising the population are identical and/or bind the same epitope(s), except for possible variants that may arise during production of the monoclonal antibody, such variants generally being present in minor amounts. Such monoclonal antibody typically includes an

antibody comprising a polypeptide sequence that binds a target, wherein the target-binding polypeptide sequence was obtained by a process that includes the selection of a single target binding polypeptide sequence from a plurality of polypeptide sequences. For example, the selection process can be the selection of a unique clone from a plurality of clones, such as a pool of hybridoma clones, phage clones or recombinant DNA clones. It should be understood that the selected target binding sequence can be further altered, for example, to improve affinity for the target, to humanize the target binding sequence, to improve its production in cell culture, to reduce its immunogenicity *in vivo*, to create a multispecific antibody, *etc.*, and that an antibody comprising the altered target binding sequence is also a monoclonal antibody of this disclosure. In contrast to polyclonal antibody preparations which typically include different antibodies directed against different determinants (*e.g.,* epitopes), each monoclonal antibody of a monoclonal antibody preparation is directed against a single determinant on an antigen. In addition to their specificity, the monoclonal antibody preparations are advantageous in that they are typically uncontaminated by other immunoglobulins. The modifier "monoclonal" indicates the character of the antibody as being obtained from a substantially homogeneous population of antibodies, and is not to be construed as requiring production of the antibody by any particular method. For example, the monoclonal antibodies to be used in accordance with the present disclosure may be made by a variety of techniques, including, for example, the hybridoma method (*e.g.,* Kohler *et al.*, Nature, 256:495 (1975); Harlow *et al.*, Antibodies: A Laboratory Manual, (Cold Spring Harbor Laboratory Press, 2nd ed. 1988); Hammerling *et al.*, in: Monoclonal Antibodies and T-Cell Hybridomas 563-681, (Elsevier, N.Y., 1981)), recombinant DNA methods (see, *e.g.,* U.S. Patent No. 4,816,567), phage display technologies (see, *e.g.,* Clackson *et al.*, Nature, 352:624-628 (1991); Marks *et al.*, J. Mol. Biol., 222:581-597 (1991); Sidhu *et al.*, J. Mol. Biol. 338(2):299-310 (2004); Lee *et al.*, J. Mol. Biol. 340(5):1073-1093 (2004); Fellouse, Proc. Nat. Acad. Sci. USA 101(34):12467-12472 (2004); and Lee *et al.* J. Immunol. Methods 284(1-2):119-132 (2004), and technologies for producing human or human-like antibodies in animals that have parts or all of the human immunoglobulin loci or genes encoding human immunoglobulin sequences (see, *e.g.,* WO 1998/24893; WO 1996/34096; WO 1996/33735; WO 1991/10741; Jakobovits *et al.*, Proc. Natl. Acad. Sci. USA, 90:2551 (1993); Jakobovits *et al.*, Nature, 362:255-258 (1993); Bruggemann *et al.*, Year in Immuno., 7:33 (1993); U.S. Patent Nos. 5,545,806; 5,569,825; 5,591,669; 5,545,807; WO 1997/17852; U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; and 5,661,016; Marks *et al.*, Bio/Technology, 10: 779-783 (1992); Lonberg *et al.*, Nature, 368: 856-859 (1994); Morrison, Nature, 368: 812-813 (1994); Fishwild *et al.*, Nature Biotechnology, 14: 845-851

(1996); Neuberger, Nature Biotechnology, 14: 826 (1996); and Lonberg and Huszar, Intern. Rev. Immunol., 13: 65-93 (1995)).

[00193] "Humanized" or “Human engineered” forms of non-human (*e.g.,* murine) antibodies are chimeric antibodies that contain amino acids represented in human immunoglobulin sequences, including, for example, wherein minimal sequence is derived from non-human immunoglobulin. For example, humanized antibodies may be human antibodies in which some hypervariable region residues and possibly some FR residues are substituted by residues from analogous sites in non-human (*e.g.,* rodent) antibodies. Alternatively, humanized or human engineered antibodies may be non-human (*e.g.,* rodent) antibodies in which some residues are substituted by residues from analogious sites in human antibodies (see, *e.g.,* U.S. Patent No. 5,766,886). Humanized antibodies include human immunoglobulins (recipient antibody) in which residues from a hypervariable region of the recipient are replaced by residues from a hypervariable region of a non-human species (donor antibody) such as mouse, rat, rabbit or nonhuman primate having the desired specificity, affinity, and capacity. In some instances, framework region (FR) residues of the human immunoglobulin are replaced by corresponding non-human residues. Furthermore, humanized antibodies may comprise residues that are not found in the recipient antibody or in the donor antibody, including, for example non-antibody sequences such as a chemokine, growth factor, peptide, cytokine, cell surface protein, serum protein, toxin, extracellular matrix protein, clotting factor, or secreted protein sequence. These modifications may be made to further refine antibody performance. Humanized antibodies include human engineered antibodies, for example, as described by U.S. Patent No. 5,766,886, including methods for preparing modified antibody variable domains. A humanized antibody may comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the hypervariable loops correspond to those of a non-human immunoglobulin and all or substantially all of the FRs are those of a human immunoglobulin sequence. A humanized antibody optionally may also comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin. For further details, see Jones *et al.*, Nature 321:522-525 (1986); Riechmann *et al.*, Nature 332:323-329 (1988); and Presta, Curr. Op. Struct. Biol. 2:593-596 (1992). See also the following review articles and references cited therein: Vaswani and Hamilton, Ann. Allergy, Asthma & Immunol. 1: 105-115 (1998); Harris, Biochem. Soc. Transactions 23:1035-1038 (1995); Hurle and Gross, Curr. Op. Biotech. 5:428-433 (1994).

[00194] "Hybrid antibodies" refer to immunoglobulin molecules in which pairs of heavy and light chains from antibodies with different antigenic determinant regions are

assembled together so that two different epitopes or two different antigens can be recognized and bound by the resulting tetramer.

[00195] "Chimeric" antibodies (immunoglobulins) have a portion of the heavy and/or light chain identical with or homologous to corresponding sequences in antibodies derived from a particular species or belonging to a particular antibody class or subclass, while the remainder of the chain(s) is identical with or homologous to corresponding sequences in antibodies derived from another species or belonging to another antibody class or subclass, as well as fragments of such antibodies, so long as they exhibit the desired biological activity (see *e.g.,* Morrison *et al.*, Proc. Natl. Acad. Sci. USA 81:6851-6855 (1984)). Humanized antibody refers to a subset of chimeric antibodies.

[00196] "Single-chain Fv" or "scFv" antibody fragments may comprise the VH and VL domains of antibody, wherein these domains are present in a single polypeptide chain. Generally, the scFv polypeptide further comprises a polypeptide linker between the VH and VL domains which enables the scFv to form the desired structure for antigen binding. For a review of scFv, see *e.g.,* Pluckthun, in The Pharmacology of Monoclonal Antibodies, vol. 113, Rosenburg and Moore eds., Springer-Verlag, New York, pp. 269-315 (1994).

[00197] An "antigen" refers to a predetermined antigen to which an antibody can selectively bind. The target antigen may be polypeptide, carbohydrate, nucleic acid, lipid, hapten or other naturally occurring or synthetic compound. Preferably, the target antigen is a polypeptide.

[00198] "Epitope" or "antigenic determinant", used interchangeably herein, refer to that portion of an antigen capable of being recognized and specifically bound by a particular antibody. When the antigen is a polypeptide, epitopes can be formed both from contiguous amino acids and noncontiguous amino acids juxtaposed by tertiary folding of a protein. Epitopes formed from contiguous amino acids are typically retained upon protein denaturing, whereas epitopes formed by tertiary folding are typically lost upon protein denaturing. An epitope typically includes at least 3, and more usually, at least 5 or 8-10 amino acids in a unique spatial conformation. Antibodies may bind to the same or a different epitope on an antigen. Antibodies may be characterized in different epitope bins. Whether an antibody binds to the same or different epitope as another antibody (*e.g.,* a reference antibody or benchmark antibody) may be determined by competition between antibodies in assays (*e.g.,* competitive binding assays).

[00199] Competition between antibodies may be determined by an assay in which the immunoglobulin under test inhibits specific binding of a reference antibody to a common antigen. Numerous types of competitive binding assays are known, for example: solid phase direct or indirect radioimmunoassay (RIA), solid phase direct or indirect

enzyme immunoassay or enzyme-linked immunosorbent assay (EIA or ELISA), sandwich competition assay including an ELISA assay (see Stahli *et al.*, Methods in Enzymology 9:242-253 (1983)); solid phase direct biotin-avidin EIA (see Kirkland *et al.*, J. Immunol. 137:3614-3619 (1986)); solid phase direct labeled assay, solid phase direct labeled sandwich assay (see Harlow and Lane, "Antibodies, A Laboratory Manual," Cold Spring Harbor Press (1988)); solid phase direct label RIA using I-125 label (see Morel *et al.*, Molec. Immunol. 25(1):7-15 (1988)); solid phase direct biotin-avidin EIA (Cheung *et al.*, Virology 176:546-552 (1990)); and direct labeled RIA (Moldenhauer *et al.*, Scand. J. Immunol., 32:77-82 (1990)). Competition binding assays may be performed using Surface Plasmon Resonance (SPR), for example, with a Biacore® instrument for kinetic analysis of binding interactions. In such an assay, an antibody comprising an ultralong CDR3 of unknown epitope specificity may be evaluated for its ability to compete for binding against a comparator antibody (*e.g.,* a BA1 or BA2 antibody as described herein). An assay may involve the use of purified antigen bound to a solid surface or cells bearing either of these, an unlabeled test immunoglobulin and a labeled reference immunoglobulin. Competitive inhibition may be measured by determining the amount of label bound to the solid surface or cells in the presence of the test immunoglobulin. Usually the test immunoglobulin is present in excess. An assay (competing antibodies) may include antibodies binding to the same epitope as the reference antibody and antibodies binding to an adjacent epitope sufficiently proximal to the epitope bound by the reference antibody for steric hindrance to occur. Usually, when a competing antibody is present in excess, it will inhibit specific binding of a reference antibody to a common antigen by at least 50%, or at least about 70%, or at least about 80%, or least about 90%, or at least about 95%, or at least about 99% or about 100% for a competitor antibody.

[00200] That an antibody "selectively binds" or "specifically binds" means that the antibody reacts or associates more frequently, more rapidly, with greater duration, with greater affinity, or with some combination of the above to an antigen or an epitope than with alternative substances, including unrelated proteins. "Selectively binds" or "specifically binds" may mean, for example, that an antibody binds to a protein with a K_D of at least about 0.1 mM, or at least about 1 μM or at least about 0.1 μM or better, or at least about 0.01 μM or better. Because of the sequence identity between homologous proteins in different species, specific binding can include an antibody that recognizes a given antigen in more than one species.

[00201] "Non-specific binding" and "background binding" when used in reference to the interaction of an antibody and a protein or peptide refer to an interaction that is not

dependent on the presence of a particular structure (*e.g.,* the antibody is binding to proteins in general rather that a particular structure such as an epitope).

[00202] "Diabodies" refer to small antibody fragments with two antigen-binding sites, which fragments comprise a heavy-chain variable domain (VH) connected to a light-chain variable domain (VL) in the same polypeptide chain (VH-VL). By using a linker that is too short to allow pairing between the two domains on the same chain, the domains are forced to pair with the complementary domains of another chain and create two antigen-binding sites. Diabodies are described more fully in, for example, EP 404,097; WO 93/11161; and Hollinger et. al., Proc. Natl. Acad. Sci. USA, 90:6444-6448 (1993).

[00203] A "human antibody" refers to one which possesses an amino acid sequence which corresponds to that of an antibody produced by a human and/or has been made using any of the techniques for making human antibodies as disclosed herein. This definition of a human antibody specifically excludes a humanized antibody comprising non-human antigen-binding residues.

[00204] An "affinity matured" antibody refers to one with one or more alterations in one or more CDRs thereof which result in an improvement in the affinity of the antibody for antigen, compared to a parent antibody which does not possess those alteration(s). Preferred affinity matured antibodies will have nanomolar or even picomolar affinities for the target antigen. Affinity matured antibodies are produced by procedures known in the art. Marks *et al.,* Bio/Technology 10:779-783 (1992) describes affinity maturation by VH and VL domain shuffling. Random mutagenesis of CDR and/or framework residues is described by: Barbas *et al.,* Proc Nat. Acad. Sci. USA 91:3809-3813 (1994); Schier *et al.*, Gene 169:147-155 (1995); Yelton *et al.,* J. Immunol. 155:1994-2004 (1995); Jackson *et al.*, J. Immunol. 154(7):3310-9 (1995); and Hawkins *et al.*, J. Mol. Biol. 226:889-896 (1992).

[00205] Antibody "effector functions" refer to those biological activities attributable to the Fc region (a native sequence Fc region or amino acid sequence variant Fc region) of an antibody, and vary with the antibody isotype. Examples of antibody effector functions include: Clq binding and complement dependent cytotoxicity; Fc receptor binding; antibody-dependent cell-mediated cytotoxicity (ADCC); phagocytosis; down regulation of cell surface receptors (*e.g.* B cell receptor); and B cell activation.

[00206] "Antibody-dependent cell-mediated cytotoxicity" or "ADCC" refers to a form of cytotoxicity in which secreted Ig bound onto Fc receptors (FcRs) present on certain cytotoxic cells (*e.g.,* Natural Killer (NK) cells, neutrophils, and macrophages) enable these cytotoxic effector cells to bind specifically to an antigen-bearing target cell and subsequently kill the target cell with cytotoxins. The antibodies "arm" the cytotoxic cells and are absolutely required for such killing. The primary cells for mediating ADCC, NK cells,

express FcγRIII only, whereas monocytes express FcγRI, FcγRII and FcγRIII. FcR expression on hematopoietic cells is summarized in Table 3 on page 464 of Ravetch and Kinet, Annu. Rev. Immunol 9:457-92 (1991). To assess ADCC activity of a molecule of interest, an *in vitro* ADCC assay, may be performed. Useful effector cells for such assays include peripheral blood mononuclear cells (PBMC) and Natural Killer (NK) cells. Alternatively, or additionally, ADCC activity of the molecule of interest may be assessed *in vivo*, *e.g.,* in a animal model such as that disclosed in Clynes *et al.* Proc. Natl. Acad. Sci. USA 95:652-656 (1998).

[00207] "Human effector cells" are leukocytes which express one or more FcRs and perform effector functions. Preferably, the cells express at least FcγRIII and perform ADCC effector function. Examples of human leukocytes which mediate ADCC include peripheral blood mononuclear cells (PBMC), natural killer (NK) cells, monocytes, cytotoxic T cells and neutrophils; with PBMCs and NK cells being preferred. The effector cells may be isolated from a native source, *e.g.,* from blood.

[00208] "Fc receptor" or "FcR" describes a receptor that binds to the Fc region of an antibody. The preferred FcR is a native sequence human FcR. Moreover, a preferred FcR is one which binds an IgG antibody (a gamma receptor) and includes receptors of the FcγRI, FcγRII, and FcγRIII subclasses, including allelic variants and alternatively spliced forms of these receptors. FcγRII receptors include FcγRIIA (an "activating receptor") and FcγRIIB (an "inhibiting receptor"), which have similar amino acid sequences that differ primarily in the cytoplasmic domains thereof. Activating receptor FcγRIIA contains an immunoreceptor tyrosine-based activation motif (ITAM) in its cytoplasmic domain. Inhibiting receptor FcγRIIB contains an immunoreceptor tyrosine-based inhibition motif (ITIM) in its cytoplasmic domain. (see review M. in Daeron, Annu. Rev. Immunol. 15:203-234 (1997)). FcRs are reviewed in Ravetch and Kinet, Annu. Rev. Immunol 9:457-92 (1991); Capel *et al.*, Immunomethods 4:25-34 (1994); and de Haas *et al.*, J. Lab. Clin. Med. 126:330-41 (1995). Other FcRs, including those to be identified in the future, are encompassed by the term "FcR" herein. The term also includes the neonatal receptor, FcRn, which is responsible for the transfer of maternal IgGs to the fetus (Guyer *et al.*, J. Immunol. 117:587 (1976) and Kim *et al.*, J. Immunol. 24:249 (1994)) and regulates homeostasis of immunoglobulins. For example, antibody variants with improved or diminished binding to FcRs have been described (see, *e.g.,* Shields *et al.* J. Biol. Chem. 9(2): 6591-6604 (2001)).

[00209] Methods of measuring binding to FcRn are known (see, *e.g.,* Ghetie 1997, Hinton 2004). Binding to human FcRn *in vivo* and serum half life of human FcRn high affinity binding polypeptides can be assayed, *e.g.,* in transgenic mice or transfected human

cell lines expressing human FcRn, or in primates administered with the Fc variant polypeptides.

[00210] "Complement dependent cytotoxicity" or "CDC" refers to the lysis of a target cell in the presence of complement. Activation of the classical complement pathway is initiated by the binding of the first component of the complement system (Clq) to antibodies (of the appropriate subclass) which are bound to their cognate antigen. To assess complement activation, a CDC assay, for example, as described in Gazzano-Santoro *et al.*, J. Immunol. Methods 202:163 (1996), may be performed.

[00211] Polypeptide variants with altered Fc region amino acid sequences and increased or decreased Clq binding capability have been described (*e.g.,* see, also, Idusogie *et al.* J. Immunol. 164: 4178-4184 (2000)).

[00212] "Fc region-comprising polypeptide" refers to a polypeptide, such as an antibody or immunoadhesin (see definitions below), which comprises an Fc region. The C-terminal lysine (residue 447 according to the EU numbering system) of the Fc region may be removed, for example, during purification of the polypeptide or by recombinant engineering the nucleic acid encoding the polypeptide.

[00213] "Blocking" antibody or an "antagonist" antibody refers to one which inhibits or reduces biological activity of the antigen it binds. Preferred blocking antibodies or antagonist antibodies substantially or completely inhibit the biological activity of the antigen.

[00214] "Agonist" antibody refers to an antibody which mimics (*e.g.,* partially or fully) at least one of the functional activities of a polypeptide of interest.

[00215] "Acceptor human framework" refers to a framework comprising the amino acid sequence of a VL or VH framework derived from a human immunoglobulin framework, or from a human consensus framework. An acceptor human framework "derived from" a human immunoglobulin framework or human consensus framework may comprise the same amino acid sequence thereof, or may contain pre-existing amino acid sequence changes. Where pre-existing amino acid changes are present, preferably no more than 5 and preferably 4 or less, or 3 or less, pre-existing amino acid changes are present.

[00216] A "human consensus framework" refers to a framework which represents the most commonly occurring amino acid residues in a selection of human immunoglobulin VL or VH framework sequences. Generally, the selection of human immunoglobulin VL or VH sequences is from a subgroup of variable domain sequences. Generally, the subgroup of sequences is a subgroup as in Kabat *et al.*, Sequences of Proteins of Immunological Interest, Fifth Edition, NIH Publication 91-3242, Bethesda Md. (1991), vols. 1-3. In one

embodiment, for the VL, the subgroup is subgroup kappa I as in Kabat *et al.*, supra. In one embodiment, for the VH, the subgroup is subgroup III as in Kabat *et al.*, supra.

[00217] "Disorder" or "disease" refers to any condition that would benefit from treatment with a substance/molecule (*e.g.*, an antibody comprising an ultralong CDR3 as dislosed herein) or method disclosed herein. This includes chronic and acute disorders or diseases including those pathological conditions which predispose the mammal to the disorder in question.

[00218] "Treatment" refers to clinical intervention in an attempt to alter the natural course of the individual or cell being treated, and can be performed either for prophylaxis or during the course of clinical pathology. Desirable effects of treatment include preventing occurrence or recurrence of disease, alleviation of symptoms, diminishment of any direct or indirect pathological consequences of the disease, preventing metastasis, decreasing the rate of disease progression, amelioration or palliation of the disease state, and remission or improved prognosis. In some embodiments, antibodies disclosed herein are used to delay development of a disease or disorder.

[00219] "Individual" (*e.g.,* a "subject") refers to a vertebrate, preferably a mammal, more preferably a human. Mammals include, but are not limited to, farm animals (such as cows), sport animals, pets (such as cats, dogs and horses), primates, mice and rats.

[00220] "Mammal" for purposes of treatment refers to any animal classified as a mammal, including humans, rodents (*e.g.,* mice and rats), and monkeys; domestic and farm animals; and zoo, sports, laboratory, or pet animals, such as dogs, cats, cattle, horses, sheep, pigs, goats, rabbits, *etc.* In some embodiments, the mammal is selected from a human, rodent, or monkey.

[00221] "Pharmaceutically acceptable" refers to approved or approvable by a regulatory agency of the Federal or a state government or listed in the U.S. Pharmacopeia or other generally recognized pharmacopeia for use in animals, including humans.

[00222] "Pharmaceutically acceptable salt" refers to a salt of a compound that is pharmaceutically acceptable and that possesses the desired pharmacological activity of the parent compound.

[00223] "Pharmaceutically acceptable excipient, carrier or adjuvant" refers to an excipient, carrier or adjuvant that can be administered to a subject, together with at least one antibody of the present disclosure, and which does not destroy the pharmacological activity thereof and is nontoxic when administered in doses sufficient to deliver a therapeutic amount of the compound.

[00224] "Pharmaceutically acceptable vehicle" refers to a diluent, adjuvant, excipient, or carrier with which at least one antibody of the present disclosure is administered.

[00225] "Providing a prognosis", "prognostic information", or "predictive information" refer to providing information, including for example the presence of cancer cells in a subject's tumor, regarding the impact of the presence of cancer (*e.g.,* as determined by the diagnostic methods of the present disclosure) on a subject's future health (*e.g.,* expected morbidity or mortality, the likelihood of getting cancer, and the risk of metastasis).

[00226] Terms such as "treating" or "treatment" or "to treat" or "alleviating" or "to alleviate" refer to both 1) therapeutic measures that cure, slow down, lessen symptoms of, and/or halt progression of a diagnosed pathologic condition or disorder and 2) prophylactic or preventative measures that prevent and/or slow the development of a targeted pathologic condition or disorder. Thus those in need of treatment include those already with the disorder; those prone to have the disorder; and those in whom the disorder is to be prevented.

[00227] "Providing a diagnosis" or "diagnostic information" refers to any information, including for example the presence of cancer cells, that is useful in determining whether a patient has a disease or condition and/or in classifying the disease or condition into a phenotypic category or any category having significance with regards to the prognosis of or likely response to treatment (either treatment in general or any particular treatment) of the disease or condition. Similarly, diagnosis refers to providing any type of diagnostic information, including, but not limited to, whether a subject is likely to have a condition (such as a tumor), whether a subject's tumor comprises cancer stem cells, information related to the nature or classification of a tumor as for example a high risk tumor or a low risk tumor, information related to prognosis and/or information useful in selecting an appropriate treatment. Selection of treatment can include the choice of a particular chemotherapeutic agent or other treatment modality such as surgery or radiation or a choice about whether to withhold or deliver therapy.

[00228] A “human consensus framework” refers to a framework which represents the most commonly occurring amino acid residues in a selection of human immunoglobulin VL or VH framework sequences. Generally, the selection of human immunoglobulin VL or VH sequences is from a subgroup of variable domain sequences. Generally, the subgroup of sequences is a subgroup as in Kabat *et al.*, Sequences of Proteins of Immunological Interest, Fifth Edition, NIH Publication 91-3242, Bethesda Md. (1991), vols. 1-3. In one

embodiment, for the VL, the subgroup is subgroup kappa I as in Kabat *et al.*, supra. In one embodiment, for the VH, the subgroup is subgroup III as in Kabat *et al.*, supra.

[00229] An "acceptor human framework" for the purposes herein refers to a framework comprising the amino acid sequence of a light chain variable domain (VL) framework or a heavy chain variable domain (VH) framework derived from a human immunoglobulin framework or a human consensus framework, as defined below. An acceptor human framework "derived from" a human immunoglobulin framework or a human consensus framework may comprise the same amino acid sequence thereof, or it may contain amino acid sequence changes. In some embodiments, the number of amino acid changes are 10 or less, 9 or less, 8 or less, 7 or less, 6 or less, 5 or less, 4 or less, 3 or less, or 2 or less. In some embodiments, the VL acceptor human framework is identical in sequence to the VL human immunoglobulin framework sequence or human consensus framework sequence.

[00230] "Antigen-binding site" refers to the interface formed by one or more complementary determining regions . An antibody molecule has two antigen combining sites, each containing portions of a heavy chain variable region and portions of a light chain variable region. The antigen combining sites can contain other portions of the variable region domains in addition to the CDRs.

[00231] An "antibody light chain" or an "antibody heavy chain" refers to a polypeptide comprising the VL or VH, respectively. The VL is encoded by the minigenes V (variable) and J (junctional), and the VH by minigenes V, D (diversity), and J. Each of VL or VH includes the CDRs as well as the framework regions. In this application, antibody light chains and/or antibody heavy chains may, from time to time, be collectively referred to as "antibody chains." These terms encompass antibody chains containing mutations that do not disrupt the basic structure of VL or VH, as one skilled in the art will readily recognize.

[00232] "Native antibodies" refer to naturally occurring immunoglobulin molecules with varying structures. For example, native IgG antibodies are heterotetrameric glycoproteins of about 150,000 daltons, composed of two identical light chains and two identical heavy chains that are disulfide bonded. From N- to C-terminus, each heavy chain has a variable region (V H), also called a variable heavy domain or a heavy chain variable domain, followed by three constant domains (CH1, CH2, and CH3). Similarly, from N- to C-terminus, each light chain has a variable region (V L), also called a variable light domain or a light chain variable domain, followed by a constant light (CL) domain. The light chain of an antibody may be assigned to one of two types, called kappa (K) and lambda (λ), based on the amino acid sequence of its constant domain.

[00233] "Combinatorial library" refers to collections of compounds formed by reacting different combinations of interchangeable chemical "building blocks" to produce a collection of compounds based on permutations of the building blocks. For an antibody combinatorial library, the building blocks are the component V, D and J regions (or modified forms thereof) from which antibodies are formed. For purposes herein, the terms "library" or "collection" are used interchangeably.

[00234] A "combinatorial antibody library" refers to a collection of antibodies (or portions thereof, such as Fabs), where the antibodies are encoded by nucleic acid molecules produced by the combination of V, D and J gene segments, particularly human V, D and J germline segments. The combinatorial libraries herein typically contain at least 50 different antibody (or antibody portions or fragment) members, typically at or about 50, 100, 500, 103, 1 x 103, 2 x 103, 3 x 103, 4x 103, 5 x 103, 6 x 103, 7 x 103, 8 x 103, 9 x 103, 1 x 104, 2 x 104, 3 x 104, 4x 104, 5 x 104, 6 x 104, 7 x 104, 8 x 104, 9 x 104 , 1 x 105, 2 x 105, 3 x 105, 4x 105, 5 x 105, 6 x 105, 7 x 105, 8 x 105, 9 x 105, 106, 107, 108, 109, 1010, or more different members. The resulting libraries or collections of antibodies or portions thereof, can be screened for binding to a target protein or modulation of a functional activity.

[00235] A "human combinatorial antibody library" refers to a collection of antibodies or portions thereof, whereby each member contains a VL and VH chains or a sufficient portion thereof to form an antigen binding site encoded by nucleic acid containing human germline segments produced as described herein.

[00236] A "variable germline segment" refers to V, D and J groups, subgroups, genes or alleles thereof. Gene segment sequences are accessible from known database (*e.g.*, National Center for Biotechnology Information (NCBI), the international ImMunoGeneTics information system® (IMGT), the Kabat database and the Tomlinson's VBase database (Lefranc (2003) Nucleic Acids Res., 31:307-310; Martin *et al.*, Bioinformatics Tools for Antibody Engineering in Handbook of Therapeutic Antibodies, Wiley-VCH (2007), pp. 104-107). Tables 3-5 list exemplary human variable germline segments. Sequences of exemplary VH, DH, JH, Vκ, Jκ, Vλ and or Jλ, germline segments are set forth in SEQ ID NOS: 10-451 and 868. For purposes herein, a germline segment includes modified sequences thereof, that are modified in accord with the rules of sequence compilation provided herein to permit practice of the method. For example, germline gene segments include those that contain one amino acid deletion or insertion at the 5' or 3' end compared to any of the sequences of nucleotides set forth in SEQ ID NOS:10-451, 868.

[00237] "Compilation," "compile," "combine," "combination," "rearrange," "rearrangement," or other similar terms or grammatical variations thereof refers to the process by which germline segments are ordered or assembled into nucleic acid sequences representing genes. For example, variable heavy chain germline segments are assembled such that the VH segment is 5' to the DH segment which is 5' to the JH segment, thereby resulting in a nucleic acid sequence encoding a VH chain. Variable light chain germline segments are assembled such that the VL segment is 5' to the JL segment, thereby resulting in a nucleic acid sequence encoding a VL chain. A constant gene segment or segments also can be assembled onto the 3' end of a nucleic acid encoding a VH or VL chain.

[00238] "Linked," or "linkage" or other grammatical variations thereof with reference to germline segments refers to the joining of germline segments. Linkage can be direct or indirect. Germline segments can be linked directly without additional nucleotides between segments, or additional nucleotides can be added to render the entire segment in-frame, or nucleotides can be deleted to render the resulting segment in-frame. It is understood that the choice of linker nucleotides is made such that the resulting nucleic acid molecule is in-frame and encodes a functional and productive antibody.

[00239] "In-frame" or "linked in-frame" with reference to linkage of human germline segments means that there are insertions and/or deletions in the nucleotide germline segments at the joined junctions to render the resulting nucleic acid molecule in-frame with the 5' start codon (ATG), thereby producing a "productive" or functional full-length polypeptide. The choice of nucleotides inserted or deleted from germline segments, particularly at joints joining various VD, DJ and VJ segments, is in accord with the rules provided in the method herein for V(D)J joint generation. For example, germline segments are assembled such that the VH segment is 5' to the DH segment which is 5' to the JH segment. At the junction joining the VH and the DH and at the junction joining the DH and JH segments, nucleotides can be inserted or deleted from the individual VH, DH or JH segments, such that the resulting nucleic acid molecule containing the joined VDJ segments are in-frame with the 5' start codon (ATG).

[00240] A portion of an antibody includes sufficient amino acids to form an antigen binding site.

[00241] A "reading frame" refers to a contiguous and non-overlapping set of three-nucleotide codons in DNA or RNA. Because three codons encode one amino acid, there exist three possible reading frames for given nucleotide sequence, reading frames 1, 2 or 3. For example, the sequence ACTGGTCA will be ACT GGT CA for reading frame 1, A CTG GTC A for reading frame 2 and AC TGG TCA for reading frame 3. Generally for

practice of the method described herein, nucleic acid sequences are combined so that the V sequence has reading frame 1.

[00242] A "stop codon" refers to a three-nucleotide sequence that signals a halt in protein synthesis during translation, or any sequence encoding that sequence (*e.g.* a DNA sequence encoding an RNA stop codon sequence), including the amber stop codon (UAG or TAG)), the ochre stop codon (UAA or TAA)) and the opal stop codon (UGA or TGA)). It is not necessary that the stop codon signal termination of translation in every cell or in every organism. For example, in suppressor strain host cells, such as amber suppressor strains and partial amber suppressor strains, translation proceeds through one or more stop codon (*e.g.* the amber stop codon for an amber suppressor strain), at least some of the time.

[00243] A "variable heavy" (VH) chain or a "variable light" (VL) chain (also termed VH domain or VL domain) refers to the polypeptide chains that make up the variable domain of an antibody. For purposes herein, heavy chain germline segments are designated as VH, DH and JH, and compilation thereof results in a nucleic acid encoding a VH chain. Light chain germline segments are designated as VL or JL, and include kappa and lambda light chains (Vκ and Jκ; Vλ and Jλ.) and compilation thereof results in a nucleic acid encoding a VL chain. It is understood that a light chain is either a kappa or lambda light chain, but does not include a kappa/lambda combination by virtue of compilation of a Vκ and Jλ.

[00244] A "degenerate codon" refers to three-nucleotide codon that specifies the same amino acid as a codon in a parent nucleotide sequence. One of skill in the art is familiar with degeneracy of the genetic code and can identify degenerate codons.

[00245] "Diversity" with respect to members in a collection refers to the number of unique members in a collection. Hence, diversity refers to the number of different amino acid sequences or nucleic acid sequences, respectively, among the analogous polypeptide members of that collection. For example, a collection of polynucleotides having a diversity of 104 contains 104 different nucleic acid sequences among the analogous polynucleotide members. In one example, the provided collections of polynucleotides and/or polypeptides have diversities of at least at or about 102, 103, 104, 105, 106, 107, 108, 109, 1010 or more.

[00246] "Sequence diversity" refers to a representation of nucleic acid sequence similarity and is determined using sequence alignments, diversity scores, and/or sequence clustering. Any two sequences can be aligned by laying the sequences side-by-side and analyzing differences within nucleotides at every position along the length of the sequences. Sequence alignment can be assessed in silico using Basic Local Alignment

Search Tool (BLAST), an NCBI tool for comparing nucleic acid and/or protein sequences. The use of BLAST for sequence alignment is well known to one of skill in the art. The Blast search algorithm compares two sequences and calculates the statistical significance of each match (a Blast score). Sequences that are most similar to each other will have a high Blast score, whereas sequences that are most varied will have a low Blast score.

[00247] A "polypeptide domain" refers to a part of a polypeptide (a sequence of three or more, generally 5 or 7 or more amino acids) that is a structurally and/or functionally distinguishable or definable. Exemplary of a polypeptide domain is a part of the polypeptide that can form an independently folded structure within a polypeptide made up of one or more structural motifs (*e.g.* combinations of alpha helices and/or beta strands connected by loop regions) and/or that is recognized by a particular functional activity, such as enzymatic activity or antigen binding. A polypeptide can have one, typically more than one, distinct domains. For example, the polypeptide can have one or more structural domains and one or more functional domains. A single polypeptide domain can be distinguished based on structure and function. A domain can encompass a contiguous linear sequence of amino acids. Alternatively, a domain can encompass a plurality of non-contiguous amino acid portions, which are non-contiguous along the linear sequence of amino acids of the polypeptide. Typically, a polypeptide contains a plurality of domains. For example, each heavy chain and each light chain of an antibody molecule contains a plurality of immunoglobulin (Ig) domains, each about 110 amino acids in length.

[00248] An "Ig domain" refers to a domain, recognized as such by those in the art, that is distinguished by a structure, called the Immunoglobulin (Ig) fold, which contains two beta-pleated sheets, each containing anti-parallel beta strands of amino acids connected by loops. The two beta sheets in the Ig fold are sandwiched together by hydrophobic interactions and a conserved intra-chain disulfide bond. Individual immunoglobulin domains within an antibody chain further can be distinguished based on function. For example, a light chain contains one variable region domain (VL) and one constant region domain (CL), while a heavy chain contains one variable region domain (VH) and three or four constant region domains (CH). Each VL, CL, VH, and CH domain is an example of an immunoglobulin domain.

[00249] A "variable domain" with reference to an antibody refers to a specific Ig domain of an antibody heavy or light chain that contains a sequence of amino acids that varies among different antibodies. Each light chain and each heavy chain has one variable region domain (VL, and, VH). The variable domains provide antigen specificity, and thus are responsible for antigen recognition. Each variable region contains CDRs that are part of the antigen binding site domain and framework regions (FRs).

[00250] A “constant region domain” refers to a domain in an antibody heavy or light chain that contains a sequence of amino acids that is comparatively more conserved among antibodies than the variable region domain. Each light chain has a single light chain constant region (CL) domain and each heavy chain contains one or more heavy chain constant region (CH) domains, which include, CH1, CH2, CH3 and CH4. Full-length IgA, IgD and IgG isotypes contain CH1, CH2 CH3 and a hinge region, while IgE and IgM contain CH1, CH2 CH3 and CH4. CH1 and CL domains extend the Fab arm of the antibody molecule, thus contributing to the interaction with antigen and rotation of the antibody arms. Antibody constant regions can serve effector functions, such as, but not limited to, clearance of antigens, pathogens and toxins to which the antibody specifically binds, *e.g.* through interactions with various cells, biomolecules and tissues.

[00251] An “antibody or portion thereof that is sufficient to form an antigen binding site” means that the antibody or portion thereof contains at least 1 or 2, typically 3, 4, 5 or all 6 CDRs of the VH and VL sufficient to retain at least a portion of the binding specificity of the corresponding full-length antibody containing all 6 CDRs. Generally, a sufficient antigen binding site at least requires CDR3 of the heavy chain (CDRH3). It typically further requires the CDR3 of the light chain (CDRL3). As described herein, one of skill in the art knows and can identify the CDRs based on Kabat or Chothia numbering (see, *e.g.*, Kabat, E.A. *et al.* (1991) Sequences of Proteins of Immunological Interest, Fifth Edition, U.S. Department of Health and Human Services, NIH Publication No. 91-3242, and Chothia, C. *et al.* (1987) J. Mol. Biol. 196:901-917). For example, based on Kabat numbering, CDR-Ll corresponds to residues L24-L34; CDR-L2 corresponds to residues L50-L56; CDR-L3 corresponds to residues L89-L97; CDR-H1 corresponds to residues H31 – H35, 35a or 35b depending on the length; CDR-H2 corresponds to residues H50-H65; and CDR-H3 corresponds to residues H95-H102.

[00252] A “peptide mimetic” refers to a peptide that mimics the activity of a polypeptide. For example, an erythropoietin (EPO) peptide mimetic is a peptide that mimics the activity of Epo, such as for binding and activation of the EPO receptor.

[00253] An “address” refers to a unique identifier for each locus in a collection whereby an addressed member (*e.g.* an antibody) can be identified. An addressed moiety is one that can be identified by virtue of its locus or location. Addressing can be effected by position on a surface, such as a well of a microplate. For example, an address for a protein in a microwell plate that is F9 means that the protein is located in row F, column 9 of the microwell plate. Addressing also can be effected by other identifiers, such as a tag encoded with a bar code or other symbology, a chemical tag, an electronic, such RF tag, a color-coded tag or other such identifier.

[00254] An “array” refers to a collection of elements, such as antibodies, containing three or more members.

[00255] A “spatial array” refers to an array where members are separated or occupy a distinct space in an array. Hence, spatial arrays are a type of addressable array. Examples of spatial arrays include microtiter plates where each well of a plate is an address in the array. Spacial arrays include any arrangement wherein a plurality of different molecules, *e.g.*, polypeptides, are held, presented, positioned, situated, or supported. Arrays can include microtiter plates, such as 48-well, 96-well, 144-well, 192-well, 240-well, 288-well, 336-well, 384-well, 432-well, 480-well, 576-well, 672-well, 768-well, 864-well, 960-well, 1056-well, 1152-well, 1248-well, 1344-well, 1440-well, or 1536-well plates, tubes, slides, chips, flasks, or any other suitable laboratory apparatus. Furthermore, arrays can also include a plurality of sub-arrays. A plurality of sub-arrays encompasses an array where more than one arrangement is used to position the polypeptides. For example, multiple 96-well plates could constitute a plurality of sub-arrays and a single array.

[00256] An “addressable library” or “spatially addressed library” refers to a collection of molecules such as nucleic acid molecules or protein agents, such as antibodies, in which each member of the collection is identifiable by virtue of its address.

[00257] An “addressable array” refers to one in which the members of the array are identifiable by their address, the position in a spatial array, such as a well of a microtiter plate, or on a solid phase support, or by virtue of an identifiable or detectable label, such as by color, fluorescence, electronic signal (*i.e.* RF, microwave or other frequency that does not substantially alter the interaction of the molecules of interest), bar code or other symbology, chemical or other such label. Hence, in general the members of the array are located at identifiable loci on the surface of a solid phase or directly or indirectly linked to or otherwise associated with the identifiable label, such as affixed to a microsphere or other particulate support (herein referred to as beads) and suspended in solution or spread out on a surface.

[00258] “An addressable combinatorial antibody library” refers to a collection of antibodies in which member antibodies are identifiable and all antibodies with the same identifier, such as position in a spatial array or on a solid support, or a chemical or RF tag, bind to the same antigen, and generally are substantially the same in amino acid sequence. For purposes herein, reference to an “addressable arrayed combinatorial antibody library” means that the antibody members are addressed in an array.

[00259] “In silico” refers to research and experiments performed using a computer. In silico methods include, but are not limited to, molecular modeling studies, biomolecular

docking experiments, and virtual representations of molecular structures and/or processes, such as molecular interactions. For purposes herein, the antibody members of a library can be designed using a computer program that selects component V, D and J germline segments from among those input into the computer and joins them in-frame to output a list of nucleic acid molecules for synthesis. Thus, the recombination of the components of the antibodies in the collections or libraries provided herein, can be performed in silico by combining the nucleotide sequences of each building block in accord with software that contains rules for doing so. The process could be performed manually without a computer, but the computer provides the convenience of speed.

[00260] A “database” refers to a collection of data items. For purposes herein, reference to a database is typically with reference to antibody databases, which provide a collection of sequence and structure information for antibody genes and sequences. Exemplary antibody databases include, but are not limited to, IMGT®, the international ImMunoGeneTics information system (imgt.cines.fr; see *e.g.*, Lefranc *et al.* (2008) Briefings in Bioinformatics, 9:263-275), National Center for Biotechnology Information (NCBI), the Kabat database and the Tomlinson’s VBase database (Lefranc (2003) Nucleic Acids Res., 31:307-310; Martin *et al.*, Bioinformatics Tools for Antibody Engineering in Handbook of Therapeutic Antibodies, Wiley-VCH (2007), pp. 104-107). A database also can be created by a user to include any desired sequences. The database can be created such that the sequences are inputted in a desired format (*e.g.*, in a particular reading frame; lacking stop codons; lacking signal sequences). The database also can be created to include sequences in addition to antibody sequences.

[00261] “Screening” refers to identification or selection of an antibody or portion thereof from a collection or library of antibodies and/or portions thereof, based on determination of the activity or property of an antibody or portion thereof. Screening can be performed in any of a variety of ways, including, for example, by assays assessing direct binding (*e.g.* binding affinity) of the antibody to a target protein or by functional assays assessing modulation of an activity of a target protein.

[00262] “Activity towards a target protein” refers to binding specificity and/or modulation of a functional activity of a target protein, or other measurements that reflects the activity of an antibody or portion thereof towards a target protein.

[00263] A “target protein” refers to candidate proteins or peptides that are specifically recognized by an antibody or portion thereof and/or whose activity is modulated by an antibody or portion thereof. A target protein includes any peptide or protein that contains an epitope for antibody recognition. Target proteins include proteins involved in

the etiology of a disease or disorder by virtue of expression or activity. Exemplary target proteins are described herein.

[00264] "Hit" refers to an antibody or portion thereof identified, recognized or selected as having an activity in a screening assay.

[00265] "Iterative" with respect to screening means that the screening is repeated a plurality of times, such as 2, 3, 4, 5 or more times, until a "Hit" is identified whose activity is optimized or improved compared to prior iterations.

[00266] "High-throughput" refers to a large-scale method or process that permits manipulation of large numbers of molecules or compounds, generally tens to hundred to thousands of compounds. For example, methods of purification and screening can be rendered high-throughput. High-throughput methods can be performed manually. Generally, however, high-throughput methods involve automation, robotics or software.

[00267] Basic Local Alignment Search Tool (BLAST) is a search algorithm developed by Altschul *et al.* (1990) to separately search protein or DNA databases, for example, based on sequence identity. For example, blastn is a program that compares a nucleotide query sequence against a nucleotide sequence database (*e.g.* GenBank). BlastP is a program that compares an amino acid query sequence against a protein sequence database.

[00268] A BLAST bit score is a value calculated from the number of gaps and substitutions associated with each aligned sequence. The higher the score, the more significant the alignment.

[00269] A "human protein" refers to a protein encoded by a nucleic acid molecule, such as DNA, present in the genome of a human, including all allelic variants and conservative variations thereof. A variant or modification of a protein is a human protein if the modification is based on the wildtype or prominent sequence of a human protein.

[00270] "Naturally occurring amino acids" refer to the 20 L-amino acids that occur in polypeptides. The residues are those 20 α-amino acids found in nature which are incorporated into protein by the specific recognition of the charged tRNA molecule with its cognate mRNA codon in humans.

[00271] "Non-naturally occurring amino acids" refer to amino acids that are not genetically encoded. For example, a non-natural amino acid is an organic compound that has a structure similar to a natural amino acid but has been modified structurally to mimic the structure and reactivity of a natural amino acid. Non-naturally occurring amino acids thus include, for example, amino acids or analogs of amino acids other than the 20 naturally-occurring amino acids and include, but are not limited to, the D-isostereomers of amino acids. Exemplary non-natural amino acids are known to those of skill in the art.

[00272] "Nucleic acids" include DNA, RNA and analogs thereof, including peptide nucleic acids (PNA) and mixtures thereof. Nucleic acids can be single or double-stranded. When referring to probes or primers, which are optionally labeled, such as with a detectable label, such as a fluorescent or radiolabel, single-stranded molecules are contemplated. Such molecules are typically of a length such that their target is statistically unique or of low copy number (typically less than 5, generally less than 3) for probing or priming a library. Generally a probe or primer contains at least 14, 16 or 30 contiguous nucleotides of sequence complementary to or identical to a gene of interest. Probes and primers can be 10, 20, 30, 50, 100 or more nucleic acids long.

[00273] A "peptide" refers to a polypeptide that is from 2 to 40 amino acids in length.

[00274] The amino acids which occur in the various sequences of amino acids provided herein are identified according to their known, three-letter or one-letter abbreviations (Table 1). The nucleotides which occur in the various nucleic acid fragments are designated with the standard single-letter designations used routinely in the art.

[00275] An "amino acid" is an organic compound containing an amino group and a carboxylic acid group. A polypeptide contains two or more amino acids. For purposes herein, amino acids include the twenty naturally-occurring amino acids, non-natural amino acids and amino acid analogs (*i.e.*, amino acids wherein the α-carbon has a side chain).

[00276] "Amino acid residue" refers to an amino acid formed upon chemical digestion (hydrolysis) of a polypeptide at its peptide linkages. The amino acid residues described herein are presumed to be in the "L" isomeric form. Residues in the "D" isomeric form, which are so designated, can be substituted for any L-amino acid residue as long as the desired functional property is retained by the polypeptide. NH2 refers to the free amino group present at the amino terminus of a polypeptide. COOH refers to the free carboxy group present at the carboxyl terminus of a polypeptide. In keeping with standard polypeptide nomenclature described in J. Biol. Chem., 243: 3552-3559 (1969), and adopted 37 C.F.R. §§ 1.821-1.822, abbreviations for amino acid residues are shown below:

SYMBOL		
1-Letter	**3-Letter**	**AMINO ACID**
Y	Tyr	Tyrosine
G	Gly	Glycine
F	Phe	Phenylalanine
M	Met	Methionine
A	Ala	Alanine
S	Ser	Serine
I	Ile	Isoleucine
L	Leu	Leucine

SYMBOL		
1-Letter	3-Letter	AMINO ACID
T	Thr	Threonine
V	Val	Valine
P	Pro	Proline
K	Lys	Lysine
H	His	Histidine
Q	Gln	Glutamine
E	Glu	Glutamic acid
Z	Glx	Glu and/or Gln
W	Trp	Tryptophan
R	Arg	Arginine
D	Asp	Aspartic acid
N	Asn	Asparagine
B	Asx	Asn and/or Asp
C	Cys	Cysteine
X	Xaa	Unknown or other

[00277] It should be noted that all amino acid residue sequences represented herein by formulae have a left to right orientation in the conventional direction of amino-terminus to carboxyl-terminus. In addition, the phrase “amino acid residue” is broadly defined to include the amino acids listed in the Table of Correspondence (Table 1) and modified and unusual amino acids, such as those referred to in 37 C.F.R. §§ 1.821-1.822, and incorporated herein by reference. Furthermore, it should be noted that a dash at the beginning or end of an amino acid residue sequence indicates a peptide bond to a further sequence of one or more amino acid residues, to an amino-terminal group such as NH2 or to a carboxyl-terminal group such as COOH. The abbreviations for any protective groups, amino acids and other compounds, are, unless indicated otherwise, in accord with their common usage, recognized abbreviations, or the IUPAC-IUB Commission on Biochemical Nomenclature (see, (1972) Biochem. 11:1726). Each naturally occurring L-amino acid is identified by the standard three letter code (or single letter code) or the standard three letter code (or single letter code) with the prefix "L-"; the prefix "D-" indicates that the stereoisomeric form of the amino acid is D.

[00278] An “immunoconjugate” refers to an antibody conjugated to one or more heterologous molecule(s), including but not limited to a cytotoxic agent. An immunoconjugate may include non-antibody sequences.

[00279] An "antibody-drug conjugate" or "ADC" refers to an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, a drug, a growth inhibitory agent, a toxin (*e.g.,* an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (*e.g.,* a radioconjugate).

General techniques

[00280] The present disclosure relies on routine techniques in the field of recombinant genetics. Basic texts disclosing the general methods of use in this present disclosure include Sambrook and Russell, Molecular Cloning: A Laboratory Manual 3d ed. (2001); Kriegler, Gene Transfer and Expression: A Laboratory Manual (1990); and Ausubel *et al.*, Current Protocols in Molecular Biology (1994).

[00281] For nucleic acids, sizes are given in either kilobases (Kb) or base pairs (bp). These are estimates derived from agarose or polyacrylamide gel electrophoresis, from sequenced nucleic acids, or from published DNA sequences. For proteins, sizes are given in kilo-Daltons (kD) or amino acid residue numbers. Proteins sizes are estimated from gel electrophoresis, from sequenced proteins, from derived amino acid sequences, or from published protein sequences.

[00282] Oligonucleotides that are not commercially available can be chemically synthesized according to the solid phase phosphoramidite triester method first described by Beaucage and Caruthers, Tetrahedron Letters, 22:1859-1862 (1981), using an automated synthesizer, as described in Van Devanter *et al.*, Nucleic Acids Res., 12:6159-6168 (1984). Purification of oligonucleotides is by either native polyacrylamide gel electrophoresis or by anion-exchange chromatography as described in Pearson & Reanier, J. Chrom., 255:137-149 (1983). The sequence of the cloned genes and synthetic oligonucleotides can be verified after cloning using, *e.g.*, the chain termination method for sequencing double-stranded templates of Wallace *et al.*, Gene, 16:21-26 (1981).

[00283] The nucleic acids encoding recombinant polypeptides of the present disclosure may be cloned into an intermediate vector before transformation into prokaryotic or eukaryotic cells for replication and/or expression. The intermediate vector may be a prokaryote vector such as a plasmid or shuttle vector.

Antibodies with Ultralong CDR3 Sequences

[00284] To date, cattle are the only species where ultralong CDR3 sequences have been identified. However, other species, for example other ruminants, may also possess antibodies with ultralong CDR3 sequences.

[00285] Exemplary antibody variable region sequences comprising an ultralong CDR3 sequence identified in cattle include those designated as: BLV1H12 (*see*, SEQ ID NO: 22), BLV5B8 (*see*, SEQ ID NO: 23), BLV5D3 (*see*, SEQ ID NO: 24) and BLV8C11 (*see*, SEQ ID NO: 25) (*see*, *e.g.*, Saini, *et al.* (1999) *Eur.J.Immunol*. 29: 2420-2426; and Saini and Kaushik (2002) *Scand. J. Immunol*. 55: 140-148); BF4E9 (*see*, SEQ ID NO: 26)

and BF1H1 (*see*, SEQ ID NO: 27) (*see*, *e.g.*, Saini and Kaushik (2002) Scand. J. Immunol. 55: 140-148); and F18 (*see*, SEQ ID NO: 28) (*see*, *e.g.*, Berens, *et al.* (1997) *Int. Immunol.* 9: 189-199) (see Figures 1 and 3).

[00286] Antibodies of the present disclosure may be isolated by screening combinatorial libraries for antibodies with the desired activity or activities. For example, a variety of methods are known in the art for generating phage display libraries and screening such libraries for antibodies possessing the desired binding characteristics. Such methods are reviewed, *e.g.*, in Hoogenboom *et al.* in Methods in Molecular Biology 178:1-37 (O'Brien *et al.*, ed., Human Press, Totowa, N.J., 2001) and further described, *e.g.*, in the McCafferty *et al.*, Nature 348:552-554; Clackson *et al.*, Nature 352: 624-628 (1991); Marks *et al.*, J. Mol. Biol. 222: 581-597 (1992); Marks and Bradbury, in Methods in Molecular Biology 248:161-175 (Lo, ed., Human Press, Totowa, N.J., 2003); Sidhu *et al.*, J. Mol. Biol. 338(2): 299-310 (2004); Lee *et al.*, J. Mol. Biol. 340(5): 1073-1093 (2004); Fellouse, Proc. Natl. Acad. Sci. USA 101(34): 12467-12472 (2004); and Lee *et al.*, J. Immunol. Methods 284(1-2): 119-132 (2004).

[00287] In certain phage display methods, repertoires of VH and VL genes are separately cloned by polymerase chain reaction (PCR) and recombined randomly in phage libraries, which can then be screened for antigen-binding phage as described in Winter *et al.*, Ann. Rev. Immunol., 12: 433-455 (1994). Phage typically display antibody fragments, either as single-chain Fv (scFv) fragments or as Fab fragments. Libraries from immunized sources provide high-affinity antibodies to the immunogen without the requirement of constructing hybridomas. Phage display libraries of bovine antibodies may be a source of bovine antibody gene sequences, including ultralong CDR3 sequences.

[00288] Typically, a non-human antibody is humanized to reduce imrnunogenicity to humans, while retaining the specificity and affinity of the parental non-human antibody. Generally, a humanized antibody comprises one or more variable domains in which CDRs (or portions thereof) are derived from a non-human antibody, and FRs (or portions thereof) are derived from human antibody sequences. A humanized antibody optionally will also comprise at least a portion of a human constant region. In some embodiments, some FR residues in a humanized antibody are substituted with corresponding residues from a non-human antibody (*e.g.*, the antibody from which the CDR residues are derived), *e.g.*, to restore or improve antibody specificity or affinity.

[00289] Humanized antibodies and methods of making them are reviewed, *e.g.*, in Almagro and Fransson, Front. Biosci. 13:1619-1633 (2008), and are further described, *e.g.*, in Riechmann *et al.*, Nature 332:323-329 (1988); Queen *et al.*, Proc. Nat'l Acad. Sci. USA 86:10029-10033 (1989); U.S. Pat. Nos. 5,821,337, 7,527,791, 6,982,321, and 7,087,409;

Kashmiri *et al.*, Methods 36:25-34 (2005) (describing SDR (a-CDR) grafting); Padlan, Mol. Immunol. 28:489-498 (1991) (describing "resurfacing"); Dall'Acqua *et al.*, Methods 36:43-60 (2005) (describing "FR shuffling"); and Osbourn *et al.*, Methods 36:61-68 (2005); Klimka *et al.*, Br. J. Cancer, 83:252-260 (2000) (describing the "guided selection" approach to FR shuffling); and Studnicka *et al.*, U.S. Patent No. 5,766,886.

[00290] Human variable region framework sequences that may be used for humanization include but are not limited to: framework sequences selected using the "best-fit" method (see, *e.g.*, Sims *et al.* J. Immunol. 151:2296 (1993)); framework sequences derived from the consensus sequence of human antibodies of a particular subgroup of light or heavy chain variable regions (see, *e.g.*, Carter *et al.* Proc. Natl. Acad. Sci. USA, 89:4285 (1992); and Presta *et al.* J. Immunol., 151:2623 (1993)); human mature (somatically mutated) framework sequences or human germline framework sequences (see, *e.g.*, Almagro and Fransson, Front. Biosci. 13:1619-1633 (2008)); and framework sequences derived from screening FR libraries (see, *e.g.*, Baca *et al.*, Biol. Chem. 272:10678-10684 (1997) and Rosok *et al.*, J. Biol. Chem. 271:22611-22618 (1996)).

[00291] Antibodies with ultralong CDR3 sequences may also include engineered non-antibody sequences, such as cytokines or growth factors, into the CDR3 region, such that the resultant antibody is effective, for example, in inhibiting tumor metastasis. Non-antibody sequences may include an interleukin sequence, a hormone sequence, a cytokine sequence, a toxin sequence, a lymphokine sequence, a growth factor sequence, a chemokine sequence, or combinations thereof. Non-antibody sequences may be human, non-human, or synthetic. In some embodiments, the cytokine or growth factor may be shown to have an antiproliferative effect on at least one cell population. Such cytokines, lymphokines, growth factors, or other hematopoietic factors include M-CSF, GM-CSF, TNF, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17, IL-18, IFN, TNFo1, TNF1, TNF2, G-CSF, Meg-CSF, GM-CSF, thrombopoietin, stem cell factor, and erythropoietin. Additional growth factors for use in antibodies and/or pharmaceutical compositions of the present disclosure include: angiogenin, bone morphogenic protein-1, bone morphogenic protein-2, bone morphogenic protein-3, bone morphogenic protein-4, bone morphogenic protein-5, bone morphogenic protein-6, bone morphogenic protein-7, bone morphogenic protein-8, bone morphogenic protein-9, bone morphogenic protein-10, bone morphogenic protein-11, bone morphogenic protein-12, bone morphogenic protein-13, bone morphogenic protein-14, bone morphogenic protein-15, bone morphogenic protein receptor IA, bone morphogenic protein receptor IB, brain derived neurotrophic factor, ciliary neutrophic factor, ciliary neutrophic factor receptor, cytokine-induced neutrophil chemotactic factor 1, cytokine-induced neutrophil, chemotactic

factor 2, cytokine-induced neutrophil chemotactic factor 2, endothelial cell growth factor, endothelin 1, epidermal growth factor, epithelial-derived neutrophil attractant, fibroblast growth factor 4, fibroblast growth factor 5, fibroblast growth factor 6, fibroblast growth factor 7, fibroblast growth factor 8, fibroblast growth factor 8b, fibroblast growth factor 8c, fibroblast growth factor 9, fibroblast growth factor 10, fibroblast growth factor acidic, fibroblast growth factor basic, glial cell line-derived neutrophic factor receptor-1, glial cell line-derived neutrophic factor receptor-2, growth related protein, growth related protein-1, growth related protein-2, growth related protein-3, heparin binding epidermal growth factor, hepatocyte growth factor, hepatocyte growth factor receptor, insulin-like growth factor I, insulin-like growth factor receptor, insulin-like growth factor II, insulin-like growth factor binding protein, keratinocyte growth factor, leukemia inhibitory factor, leukemia inhibitory factor receptor-1, nerve growth factor nerve growth factor receptor, neurotrophin-3, neurotrophin-4, placenta growth factor, placenta growth factor 2, platelet-derived endothelial cell growth factor, platelet derived growth factor, platelet derived growth factor A chain, platelet derived growth factor AA, platelet derived growth factor AB, platelet derived growth factor B chain, platelet derived growth factor BB, platelet derived growth factor receptor-1, platelet derived growth factor receptor-2, pre-B cell growth stimulating factor, stem cell factor, stem cell factor receptor, transforming growth factor-1, transforming growth factor-2, transforming growth factor-1, transforming growth factor-1.2, transforming growth factor-2, transforming growth factor-3, transforming growth factor-S, latent transforming growth factor-1, transforming growth factor-1 binding protein I, transforming growth factor-1 binding protein II, transforming growth factor-1 binding protein III, tumor necrosis factor receptor type I, tumor necrosis factor receptor type II, urokinase-type plasminogen activator receptor, vascular endothelial growth factor, and chimeric proteins and biologically or immunologically active fragments thereof. Exemplary toxin sequences include an ADWX-1 sequence, an HsTx1 sequence, an OSK1 sequence, a Pi2 sequence, a Hongotoxin (HgTX) sequence, a Margatoxin sequence, an Agitoxin-2 sequence, a Pi3 sequence, a Kaliotoxin sequence, an Anuroctoxin sequence, a Charybdotoxin sequence, a Tityustoxin -K- alpha sequence, a Maurotoxin sequence, a Ceratotoxin 1 (CcoTx1) sequence, a CcoTx2 sequence, a CcoTx3 sequence, a Phrixotoxin 3 (PaurTx3) sequence, a Hanatoxin 1 sequence, a Phrixotoxin 1 sequence, a Huwentoxin-IV sequence, an α-conotoxin Iml sequence, an α-conotoxin Epl sequence, an α-conotoxin PnIA sequence, an α-conotoxin PnIB sequence, an α-conotoxin MII sequence, an α-conotoxin AuIA sequence, an α-conotoxin AuIB sequence, an α-conotoxin AuIC sequence, a conotoxin κ-PVIIA sequence, a charybdotoxin sequence, a neurotoxin B-IV sequence, a crotamine sequence, a ω-GVIA (conotoxin) sequence, a κ-hefutoxin 1 sequence, a Css4 sequence, a Bj-xtrIT

sequence, a BclV sequence, a Hm-1 sequence, a Hm-2 sequence, a GsAF-I (β-theraphotoxin-Gr1b) sequence, a Protoxin I (ProTx-I sequence, a β-theraphotoxin-Tp1a) sequence, a Protoxin II (ProTx II) sequence, a Huwentoxin I sequence, a μ-Conotoxin PIIIA sequence, a Jingzhaotoxin-III (ß-TRTX-Cj1α) sequence, a GsAF-II (Kappa-theraphotoxin-Gr2c) sequence, a ShK (Stichodactyla toxin) sequence, a HsTx1 sequence, a Guangxitoxin 1E (GxTx-1E) sequence, a Maurotoxin sequence, a Charybdotoxin (ChTX) sequence, an Iberiotoxin (IbTx) sequence, a Leiurotoxin 1 (scyllatoxin) sequence, a Tamapin sequence, a Kaliotoxin-1 (KTX) sequence, a Purotoxin1 (PT-1) sequence, or a GpTx-1 sequence, a MOKA Toxin sequence, a OSK1 (P12, K16, D20) sequence, a OSK1 (K16, D20) sequence, a HmK sequence, a ShK (K16,Y26, K29) sequence, a ShK (K16) sequence, a ShK-A (K16) sequence, a ShK (K16,E30) sequence, a ShK (Q21) sequence, a ShK (L21) sequence, a ShK (F21) sequence, a ShK (I21) sequence, or a ShK (A21) sequence. Exemplary toxin sequences include SEQ ID NO: 599-655, 666-698, 727-733, 754, 755 or 774-778 (see, e.g., Figures 7B and 7C). Additionally, exemplary non-antibody sequences include interleukin 8 (IL-8, SEQ ID NO: 475), interleukin 21 (IL-21, SEQ ID NO: 480), CXCL12/SDF-1alpha (SEQ ID NO: 479), somatostain (SEQ ID NO: 477), ProTx-II (SEQ ID NO: 481), chlorotoxin (SEQ ID NO: 478), and ziconotide (SEQ ID NO: 476).

[00292] A non-human antibody may be humanized to reduce immunogenicity to humans, while retaining the specificity and affinity of the parental non-human antibody. Generally, a humanized antibody comprises one or more variable domains in which HVRs, *e.g.*, CDRs, (or portions thereof) are derived from a non-human antibody, and FRs (or portions thereof) are derived from human antibody sequences. A humanized antibody optionally will also comprise at least a portion of a human constant region. In some embodiments, some FR residues in a humanized antibody are substituted with corresponding residues from a non-human antibody (*e.g.*, the antibody from which the HVR residues are derived), *e.g.*, to restore or improve antibody specificity or affinity.

[00293] The antibodies comprising an ultralong CDR3 as disclosed herein are preferably monoclonal. Also encompassed within the scope of the disclosure are Fab, Fab', Fab'-SH and $F(ab')^2$ fragments of the antibodies comprising an ultralong CDR3 as provided herein. These antibody fragments can be created by traditional means, such as enzymatic digestion, or may be generated by recombinant techniques. Such antibody fragments may be chimeric or humanized. These fragments are useful for the diagnostic and therapeutic purposes set forth below.

[00294] Monoclonal antibodies are obtained from a population of substantially homogeneous antibodies, *e.g.,* the individual antibodies comprising the population are identical except for possible naturally occurring mutations that may be present in minor

amounts. Thus, the modifier "monoclonal" indicates the character of the antibody as not being a mixture of discrete antibodies.

[00295] The antibodies comprising an ultralong CDR3 as disclosed herein can be made using a hybridoma cell-based method first described by Kohler *et al.*, Nature, 256:495 (1975), or may be made by recombinant DNA methods.

[00296] Hybridoma cells can be generated by fusing B cells producing a desired antibody with an immortalized cell line, usually a myeloma cell line, so that the resulting fusion cells will be an immortalized cell line that secrets a particular antibody. By the same principle, myeloma cells can be first transfected with a nucleic acid encoding a germline antibody V region and can be screened for the expression of the germline V region. Those myeloma cells with highest level of proteolytic light chain expression can be subsequently fused with B cells that produce an antibody with desired target protein specificity. The fusion cells will produce two types of antibodies: one is a heterologous antibody containing an endogenous antibody chain (either heavy or light) operably joined to the recombinant germline V region (either heavy or light), and the other is the same antibody that the parental B cells would secrete (*e.g.* both endogenous heavy and light chains). The operably joined heterologous heavy and light chains can be isolated by conventional methods such as chromatography and identification can be confirmed by target protein binding assays, assays identifying a unique tag of the germline polypeptide, or endopeptidase activity assays described in other sections of this disclosure. In some cases, where the heterologous antibody is the predominant type in quantity among the two types of antibodies, such isolation may not be needed.

[00297] The hybridoma cells may be seeded and grown in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, parental myeloma cells. For example, if the parental myeloma cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine (HAT medium), which substances prevent the growth of HGPRT-deficient cells.

[00298] Preferred myeloma cells are those that fuse efficiently, support stable high-level production of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. Among these, myeloma cell lines may be murine myeloma lines, such as those derived from MOPC-21 and MPC-11 mouse tumors available from the Salk Institute Cell Distribution Center, San Diego, Calif. USA, and SP-2 or X63-Ag8-653 cells available from the American Type Culture Collection, Rockville, Md. USA. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, J. Immunol.,

133:3001 (1984); Brodeur *et al.*, Monoclonal Antibody Production Techniques and Applications, pp. 51-63 (Marcel Dekker, Inc., New York, 1987)).

[00299] Culture medium in which hybridoma cells are growing is assayed for production of antibodies comprising an ultralong CDR3. For example, the binding specificity of monoclonal antibodies produced by hybridoma cells may be determined by immunoprecipitation or by an *in vitro* binding assay, such as an enzyme-linked immunoadsorbent assay (ELISA).

[00300] The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson *et al.*, Anal. Biochem., 107:220 (1980).

[00301] After hybridoma cells are identified that produce antibodies of the desired specificity, affinity, and/or activity, the clones may be subcloned by limiting dilution procedures and grown by standard methods (Goding, Monoclonal Antibodies: Principles and Practice, pp. 59-103 (Academic Press, 1986)). Suitable culture media for this purpose include, for example, D-MEM or RPMI-1640 medium. In addition, the hybridoma cells may be grown *in vivo* as ascites tumors in an animal.

[00302] The monoclonal antibodies secreted by the subclones are suitably separated from the culture medium, ascites fluid, or serum by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

[00303] The antibodies comprising an ultralong CDR3 as disclosed herein may be made by using combinatorial libraries to screen for synthetic antibody clones with the desired activity or activities. For example, synthetic antibody clones are selected by screening phage libraries containing phage that display various fragments of antibody variable regions (*e.g.,* scFv or Fab) fused to phage coat protein. Such phage libraries may be panned, for example, by affinity chromatography against the desired antigen. Clones expressing antibody fragments capable of binding to the desired antigen may be adsorbed to the antigen and thus separated from the non-binding clones in the library. The binding clones may then be eluted from the antigen, and can be further enriched by additional cycles of antigen adsorption/elution. Any of the antibodies comprising an ultralong CDR3 as disclosed herein may be obtained by designing a suitable antigen screening procedure to select for the phage clone of interest followed by construction of a full length antibody comprising an ultralong CDR3 clone using the VH and VL (*e.g.,* from scFv or Fab) sequences from the phage clone of interest and suitable constant region (Fc) sequences described in Kabat *et al.*, Sequences of Proteins of Immunological Interest, Fifth Edition, NIH Publication 91-3242, Bethesda Md. (1991), vols. 1-3.

[00304] The antigen-binding domain of an antibody is formed from two variable (V) regions, one each from the light (VL) and heavy (VH) chains, that both present three hypervariable loops or complementarity-determining regions (CDRs). Variable domains may be displayed functionally on phage, either as single-chain Fv (scFv, also referred to as single-chain antibody (SCA)) fragments, in which VH and VL are covalently linked through a short, flexible peptide, or as Fab fragments, in which they are each fused to a constant domain and interact non-covalently, as described in Winter *et al.*, Ann. Rev. Immunol., 12: 433-455 (1994). scFv or SCA encoding phage clones and Fab encoding phage clones may be separately or collectively referred to as "Fv phage clones" or "Fv clones".

[00305] Repertoires of VH and VL genes may be separately cloned by polymerase chain reaction (PCR) and recombined randomly in phage libraries, which can then be searched for antigen-binding clones as described in Winter *et al.*, Ann. Rev. Immunol., 12: 433-455 (1994). Libraries from immunized sources provide high-affinity antibodies to the immunogen without the requirement of constructing hybridomas. Alternatively, the naive repertoire may be cloned to provide a single source of human antibodies to a wide range of non-self and also self antigens without any immunization as described by Griffiths *et al.*, EMBO J. 12: 725-734 (1993). Finally, naive libraries can also be made synthetically by cloning the unrearranged V-gene segments from stem cells, and using PCR primers containing random sequence to encode the highly variable CDR3 regions and to accomplish rearrangement *in vitro* as described by Hoogenboom and Winter, J. Mol. Biol., 227: 381-388 (1992).

[00306] Filamentous phage is used to display antibody fragments by fusion to the minor coat protein pIII. Protein pIII may include truncated forms of pIII. The antibody fragments can be displayed as single chain Fv fragments, in which VH and VL domains are connected on the same polypeptide chain by a flexible polypeptide spacer, (*e.g.,* as described by Marks *et al.*, J. Mol. Biol., 222: 581-597 (1991)), or as Fab fragments, in which one chain is fused to pIII (*e.g.,* a truncated pIII) and the other is secreted into the bacterial host cell periplasm where assembly of a Fab-coat protein structure which becomes displayed on the phage surface by displacing some of the wild type coat proteins, (*e.g.*, as described in Hoogenboom *et al.*, Nucl. Acids Res., 19: 4133-4137 (1991)).

[00307] Nucleic acid encoding antibody variable gene segments (including VH and VL segments) are recovered from the cells of interest and and may be amplified or copies made by recombinant DNA techniques (*e.g.,* Kunkel mutagenesis). For example, in the case of rearranged VH and VL gene libraries, the desired DNA may be obtained by isolating genomic DNA or mRNA from lymphocytes followed by polymerase chain reaction (PCR) with primers matching the 5' and 3' ends of rearranged VH and VL genes as

described in Orlandi *et al.*, Proc. Natl. Acad. Sci. (USA), 86: 3833-3837 (1989), thereby making diverse V gene repertoires for expression. The V genes may be amplified from cDNA and genomic DNA, with back primers at the 5' end of the exon encoding the mature V-domain and forward primers based within the J-segment as described in Orlandi *et al.* (1989) and in Ward *et al.*, Nature, 341: 544-546 (1989). For amplifying from cDNA, back primers can also be based in the leader exon as described in Jones *et al.*, Biotechnol., 9: 88-89 (1991), and forward primers within the constant region as described in Sastry *et al.*, Proc. Natl. Acad. Sci. (USA), 86: 5728-5732 (1989). To enhance or maximize complementarity, degeneracy may be incorporated in the primers as described in Orlandi *et al.* (1989) or Sastry *et al.* (1989). Library diversity may be enhanced or maximized by using PCR primers targeted to each V-gene family in order to amplify available VH and VL arrangements present in the immune cell nucleic acid sample, for example, as described in the method of Marks *et al.*, J. Mol. Biol., 222: 581-597 (1991) or as described in the method of Orum *et al.*, Nucleic Acids Res., 21: 4491-4498 (1993). For cloning of the amplified DNA into expression vectors, rare restriction may can be introduced within the PCR primer as a tag at one end as described in Orlandi *et al.* (1989), or by further PCR amplification with a tagged primer as described in Clackson *et al.*, Nature, 352: 624-628 (1991).

[00308] Repertoires of synthetically rearranged V genes may be derived *in vitro* from V gene segments. Most of the human VH-gene segments have been cloned and sequenced (*e.g.,* reported in Tomlinson *et al.*, J. Mol. Biol., 227: 776-798 (1992)), and mapped (*e.g.,* reported in Matsuda *et al.*, Nature Genet., 3: 88-94 (1993); these cloned segments (including all the major conformations of the H1 and H2 loop) may be used to generate diverse VH gene repertoires with PCR primers encoding H3 loops of diverse sequence and length as described in Hoogenboom and Winter, J. Mol. Biol., 227: 381-388 (1992). VH repertoires may also be made with all the sequence diversity focused in a long H3 loop of a single length as described in Barbas *et al.*, Proc. Natl. Acad. Sci. USA, 89: 4457-4461 (1992). Human Vκ and Vλ. segments have been cloned and sequenced (reported in Williams and Winter, Eur. J. Immunol., 23: 1456-1461 (1993)) and can be used to make synthetic light chain repertoires. Synthetic V gene repertoires, based on a range of VH and VL folds, and L3 and H3 lengths, will encode antibodies of considerable structural diversity. Following amplification of V-gene encoding DNAs, germline V-gene segments can be rearranged *in vitro* according to the methods of Hoogenboom and Winter, J. Mol. Biol., 227: 381-388 (1992).

[00309] Repertoires of antibody fragments may be constructed by combining VH and VL gene repertoires together in several ways. Each repertoire may be created in different vectors, and the vectors recombined *in vitro*, for example, as described in Hogrefe

et al., Gene, 128: 119-126 (1993), or *in vivo* by combinatorial infection, for example, the loxP system described in Waterhouse *et al.*, Nucl. Acids Res., 21: 2265-2266 (1993). The *in vivo* recombination approach exploits the two-chain nature of Fab fragments to overcome the limit on library size imposed by *E. coli* transformation efficiency. Naive VH and VL repertoires are cloned separately, one into a phagemid and the other into a phage vector. The two libraries are then combined by phage infection of phagemid-containing bacteria so that each cell contains a different combination and the library size is limited only by the number of cells present (about 10^{12} clones). Both vectors contain *in vivo* recombination signals so that the VH and VL genes are recombined onto a single replicon and are co-packaged into phage virions. These large libraries may provide large numbers of diverse antibodies of good affinity (K_d^{-1} of about 10^{-8} M).

[00310] Alternatively, the repertoires may be cloned sequentially into the same vector, for example, as described in Barbas *et al.*, Proc. Natl. Acad. Sci. USA, 88: 7978-7982 (1991), or assembled together by PCR and then cloned, for example, as described in Clackson *et al.*, Nature, 352: 624-628 (1991). PCR assembly may also be used to join VH and VL DNAs with DNA encoding a flexible peptide spacer to form single chain Fv (scFv) repertoires. In yet another technique, "in cell PCR assembly" may be used to combine VH and VL genes within lymphocytes by PCR and then clone repertoires of linked genes as described in Embleton *et al.*, Nucl. Acids Res., 20: 3831-3837 (1992).

[00311] The antibodies produced by naive libraries (either natural or synthetic) can be of moderate affinity (K_d^{-1} of about 10^6 to $10^7 M^{-1}$), but affinity maturation may also be mimicked *in vitro* by constructing and reselecting from secondary libraries as described in Winter *et al.* (1994), *supra*. For example, mutation can be introduced at random *in vitro* by using error-prone polymerase (reported in Leung *et al.*, Technique, 1: 11-15 (1989)) in the method of Hawkins *et al.*, J. Mol. Biol., 226: 889-896 (1992) or in the method of Gram *et al.*, Proc. Natl. Acad. Sci. USA, 89: 3576-3580 (1992). Additionally, affinity maturation may be performed by randomly mutating one or more CDRs, for example, using PCR with primers carrying random sequence spanning the CDR of interest, in selected individual Fv clones and screening for higher affinity clones. WO 9607754 described a method for inducing mutagenesis in a complementarity determining region of an immunoglobulin light chain to create a library of light chain genes. Another effective approach is to recombine the VH or VL domains selected by phage display with repertoires of naturally occurring V domain variants obtained from unimmunized donors and screen for higher affinity in several rounds of chain reshuffling as described in Marks *et al.*, Biotechnol., 10: 779-783 (1992). This technique allows the production of antibodies and antibody fragments with affinities in the 10^{-9} M range.

[00312] The phage library samples are contacted with an immobilized protein under conditions suitable for binding of at least a portion of the phage particles with the adsorbent. Normally, the conditions, including pH, ionic strength, temperature and the like are selected to mimic physiological conditions. The phages bound to the solid phase are washed and then eluted by acid, *e.g.,* as described in Barbas *et al.*, Proc. Natl. Acad. Sci. USA, 88: 7978-7982 (1991), or by alkali, (*e.g.*, as described in Marks *et al.*, J. Mol. Biol., 222: 581-597 (1991)), or by antigen competition, (*e.g.*, in a procedure similar to the antigen competition method of Clackson *et al.*, Nature, 352: 624-628 (1991)). Phages may be enriched 20-1,000-fold in a single round of selection. Moreover, the enriched phages may be grown in bacterial culture and subjected to further rounds of selection.

[00313] The efficiency of selection depends on many factors, including the kinetics of dissociation during washing, and whether multiple antibody fragments on a single phage can simultaneously engage with antigen. Antibodies with fast dissociation kinetics (and weak binding affinities) may be retained by use of short washes, multivalent phage display and high coating density of antigen in solid phase. The high density not only stabilizes the phage through multivalent interactions, but favors rebinding of phage that has dissociated. The selection of antibodies with slow dissociation kinetics (and good binding affinities) may be promoted by use of long washes and monovalent phage display as described in Bass *et al.*, Proteins, 8: 309-314 (1990) and in WO 92/09690, and a low coating density of antigen as described in Marks *et al.*, Biotechnol., 10: 779-783 (1992).

[00314] DNA encoding the hybridoma-derived monoclonal antibodies or phage display Fv clones disclosed herein is readily isolated and sequenced using conventional procedures (*e.g.,* by using oligonucleotide primers designed to specifically amplify the heavy and light chain coding regions of interest from hybridoma or phage DNA template). 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, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of the desired monoclonal antibodies in the recombinant host cells. Recombinant expression in bacteria of antibody-encoding DNA has been described by Better *et al.,* U.S. Patent No. 6,204,023 (see also, *e.g.,* Skerra *et al.*, Curr. Opinion in Immunol., 5: 256 (1993) and Pluckthun, Immunol. Revs, 130: 151 (1992)).

[00315] DNA encoding Fv clones as disclosed herein may be combined with known DNA sequences encoding heavy chain and/or light chain constant regions (*e.g.*, the appropriate DNA sequences can be obtained from Kabat *et al.*, *supra*) to form clones encoding full or partial length heavy and/or light chains. It will be appreciated that constant regions of any isotype can be used for this purpose, including IgG, IgM, IgA, IgD, and IgE

constant regions, and that such constant regions may be obtained from any human or animal species. A Fv clone derived from the variable domain DNA of one animal (such as human) species and then fused to constant region DNA of another animal species to form coding sequence(s) for "hybrid", full length heavy chain and/or light chain is included in the definition of "chimeric" and "hybrid" antibody as used herein. In a preferred Fv clone embodiment, aFv clone derived from human variable DNA is fused to human constant region DNA to form coding sequence(s) for all human, full or partial length heavy and/or light chains.

[00316] DNA encoding an antibody comprising an ultralong CDR3 derived from a hybridoma disclosed herein may also be modified, for example, by substituting the coding sequence for human heavy- and light-chain constant domains in place of homologous murine sequences derived from the hybridoma clone (*e.g.,* as in the method of Morrison *et al.*, Proc. Natl. Acad. Sci. USA, 81: 6851-6855 (1984)). DNA encoding a hybridoma or Fv clone-derived antibody or fragment can be further modified by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. In this manner, "chimeric" or "hybrid" antibodies are prepared that have the binding specificity of the Fv clone or hybridoma clone-derived antibodies disclosed herein.

Antibody Genes and Proteins

[00317] The present disclosure provides antibody genes and proteins including, for example, antibody genes or proteins that comprise an ultralong CDR3 sequence and/or a CDR3 scaffold. The present disclosure additionally provides VH, DH, and JH sequences useful in the preparation of ultralong CDR3 sequences. Such sequences may comprise motifs (*e.g.*, cysteine motifs) as described herein including those as described in the many embodiments disclosed herein. In some embodiments, the antibodies disclosed herein may selectively or specifically bind to an epitope of a target protein. In some embodiments, the antibody may be an antagonist (*e.g.*, blocking) antibody or an agonist antibody.

[00318] The variable region of the heavy and light chains are encoded by multiple germline gene segments separated by non-coding regions, or introns, and often are present on different chromosomes. For example, the genes for the human immunoglobulin heavy chain region contains approximately 65 variable (VH) genes, 27 Diversity (DH) genes, and 6 Joining (JH) genes. The human kappa (κ) and lambda (λ) light chains are also each encoded by a similar number of VL and JL gene segments, but do not include any D gene segments. Exemplary VH, DH, JH and VL ($V\kappa$ or $V\lambda$) and JL ($J\kappa$ or $J\lambda$) germline gene segments are set forth in WO 2010/054007.

[00319] During B cell differentiation germline DNA is rearranged whereby one DH and one JH gene segment of the heavy chain locus are recombined, which is followed by the joining of one VH gene segment forming a rearranged VDJ gene that encodes a VH chain. The rearrangement occurs only on a single heavy chain allele by the process of allelic exclusion. Allelic exclusion is regulated by in-frame or "productive" recombination of the VDJ segments, which occurs in only about one-third of VDJ recombinations of the variable heavy chain. When such productive recombination events first occur in a cell, this result in production of a μ heavy chain that gets expressed on the surface of a pre- B cell and transmits a signal to shut off further heavy chain recombination, thereby preventing expression of the allelic heavy chain locus. The surface-expressed μ heavy chain also acts to activate the kappa (κ) locus for rearrangement. The lambda (λ) locus is only activated for rearrangement if the κ recombination is unproductive on both loci. The light chain rearrangement events are similar to the heavy chain, except that only the VL and JL segments are recombined. Before primary transcription of each, the corresponding constant chain gene is added. Subsequent transcription and RNA splicing leads to mRNA that is translated into an intact light chain or heavy chain.

[00320] The variable regions of antibodies confer antigen binding and specificity due to recombination events of individual germline V, D and J segments, whereby the resulting recombined nucleic acid sequences encoding the variable region domains differ among antibodies and confer antigen-specificity to a particular antibody. The variation, however, is limited to three complementarity determining regions (CDR1, CDR2, and CDR3) found within the N-terminal domain of the heavy (H) and (L) chain variable regions. The CDRs are interspersed with regions that are more conserved, termed "framework regions" (FR). The extent of the framework region and CDRs has been precisely defined (see *e.g.*, Kabat, E.A. *et al.* (1991) Sequences of Proteins of Immunological Interest, Fifth Edition, U.S. Department of Health and Human Services, NIH Publication No. 91-3242, and Chothia, C. *et al.* (1987) J. Mol. Biol. 196:901-917). Each VH and VL is typically composed of three CDRs and four FRs arranged from the amino terminus to carboxy terminus in the following order: FR1, CDR1, FR2, CDR2, FR3, CDR3 and FR4. Sequence variability among VL and VH domains is generally limited to the CDRs, which are the regions that form the antigen binding site. For example, for the heavy chain, generally, VH genes encode the N-terminal three framework regions, the first two complete CDRs and the first part of the third CDR), the DH gene encodes the central portion of the third CDR, and the JH gene encodes the last part of the third CDR and the fourth framework region. For the light chain, the VL genes encode the first CDR and second CDR. The third CDR (CDRL3) is formed by the joining of the VL and JL gene segments. Hence, CDRs 1 and 2 are

exclusively encoded by germline V gene segment sequences. The VH and VL chain CDR3s form the center of the Ag-binding site, with CDRs 1 and 2 form the outside boundaries; the FRs support the scaffold by orienting the H and L CDRs. On average, an antigen binding site typically requires at least four of the CDRs make contact with the antigen's epitope, with CDR3 of both the heavy and light chain being the most variable and contributing the most specificity to antigen binding (see, *e.g.*, Janis Kuby, Immunology, Third Edition, New York, W.H. Freeman and Company, 1998, pp. 115-118). CDRH3, which includes all of the D gene segment, is the most diverse component of the Ab-binding site, and typically plays a critical role in defining the specificity of the Ab. In addition to sequence variation, there is variation in the length of the CDRs between the heavy and light chains.

[00321] The constant regions, on the other hand, are encoded by sequences that are more conserved among antibodies. These domains confer functional properties to antibodies, for example, the ability to interact with cells of the immune system and serum proteins in order to cause clearance of infectious agents. Different classes of antibodies, for example IgM, IgD, IgG, IgE and IgA, have different constant regions, allowing them to serve distinct effector functions.

[00322] These natural recombination events of V, D, and J, can provide nearly $2x10^7$ different antibodies with both high affinity and specificity. Additional diversity is introduced by nucleotide insertions and deletions in the joining segments and also by somatic hypermutation of V regions. The result is that there are approximately 10^{10} antibodies present in an individual with differing antigen specificities.

[00323] Antibodies include bovine antibody BLVH12 (e.g., heavy chain variable region set forth in SEQ ID NO: 482, and light chain variable region set forth in SEQ ID NO: 483); and bovine antibody BLV5B8 (e.g., heavy chain variable region set forth in SEQ ID NO: 484, and light chain variable region set forth in SEQ ID NO: 485)

Antibody Fragments

[00324] The present disclosure encompasses antibody fragments. In certain circumstances there are advantages of using antibody fragments, rather than whole antibodies. The smaller size of the fragments allows for rapid clearance, and may lead to improved access to solid tumors. Antibody fragments include, but are not limited to, Fab, Fab', Fab'-SH, F(ab')2, Fv, and scFv fragments, and other fragments described below. For a review of certain antibody fragments, see Hudson *et al.* Nat. Med. 9:129-134 (2003). For a review of scFv fragments, see, *e.g.*, Pluckthun, in The Pharmacology of Monoclonal Antibodies, vol. 113, Rosenburg and Moore eds., (Springer-Verlag, New York), pp. 269-

315 (1994); see also WO 93/16185; and U.S. Pat. Nos. 5,571,894 and 5,587,458. For discussion of Fab and F(ab')2 fragments comprising salvage receptor binding epitope residues and having increased in vivo half-life, see U.S. Pat. No. 5,869,046.

[00325] Diabodies are antibody fragments with two antigen binding sites that may be bivalent or bispecific. See, for example, EP 404,097; WO 1993/01161; Hudson *et al.*, Nat. Med. 9:129-134 (2003); and Hollinger *et al.*, Proc. Natl. Acad. Sci. USA 90: 6444-6448 (1993). Triabodies and tetrabodies are also described in Hudson *et al.*, Nat. Med. 9: 129134 (2003).

[00326] Single-domain antibodies are antibody fragments comprising all or a portion of the heavy chain variable domain or all or a portion of the light chain variable domain of an antibody. In certain embodiments, a single-domain antibody is a human single-domain antibody (Domantis, Inc., Waltham, Mass.; see, *e.g.*, U.S. Pat. No. 6,248,516). Antibody fragments can be made by various techniques, including but not limited to proteolytic digestion of an intact antibody as well as production by recombinant host cells (*e.g. E. coli* or phage), as described herein.

[00327] Various techniques have been developed for the production of antibody fragments. Traditionally, these fragments were derived via proteolytic digestion of intact antibodies (see, *e.g.,* Morimoto *et al.*, Journal of Biochemical and Biophysical Methods 24:107-117 (1992); and Brennan *et al.*, Science, 229:81 (1985)). However, these fragments can now be produced directly by recombinant host cells. Fab, Fv and ScFv antibody fragments can all be expressed in and secreted from *E. coli*, thus allowing the facile production of large amounts of these fragments (see, *e.g.*, U.S. Patent No. 6,204,023). Antibody fragments can be isolated from antibody phage libraries as discussed above. Alternatively, Fab'-SH fragments can be directly recovered from *E. coli* and chemically coupled to form $F(ab')_2$ fragments (see, *e.g.,* Carter *et al.*, Bio/Technology 10: 163-167 (1992)). According to another approach, $F(ab')_2$ fragments can be isolated directly from recombinant host cell culture. Fab and $F(ab')_2$ fragment with increased *in vivo* half-life comprising a salvage receptor binding epitope residues (see, *e.g.,* in U.S. Patent No. 5,869,046). Other techniques for the production of antibody fragments will be apparent to the skilled practitioner. In other embodiments, the antibody of choice is a single chain Fv fragment (scFv or single chain antibody (SCA)). See WO 93/16185; U.S. Patent Nos. 5,571,894; and 5,587,458. Fv and sFv are the only species with intact combining sites that are devoid of constant regions; thus, they are suitable for reduced nonspecific binding during *in vivo* use. sFv fusion proteins may be constructed to yield fusion of an effector protein at either the amino or the carboxy terminus of an sFv. See Antibody Engineering, ed. Borrebaeck, *Supra*. The antibody fragment may also be a "linear antibody", for

example, as described in U.S. Patent No. 5,641,870. Such linear antibody fragments may be monospecific or bispecific.

Humanized Antibodies

[00328] The present disclosure provides antibodies comprising an ultralong CDR3. Antibodies may include human engineered antibodies (*see*, *e.g.,* Studnicka *et al.* (1994) Protein Eng.. 7(6) 805-814; and U.S. Patent No. 5,766,886). Various methods for humanizing non-human antibodies are known in the art. For example, a humanized antibody can have one or more amino acid residues introduced into it from a source which is human or non-human. Humanization may be performed following the method of Studnicka (see, *e.g*., Studnicka *et al.* (1994) Protein Eng.. 7(6) 805-814; and U.S. Patent No. 5,766,886), including the preparation of modified antibody variable domains. Humanization may alternatively be performed following the method of Winter and co-workers (Jones *et al.* (1986) Nature 321:522-525; Riechmann *et al.* (1988) Nature 332:323-327; Verhoeyen *et al.* (1988) Science 239:1534-1536), by substituting hypervariable region sequences for the corresponding sequences of a human antibody. Accordingly, such "humanized" or “human engineered” antibodies are chimeric antibodies, including wherein substantially less than an intact human variable domain has been substituted by or incorporated into the corresponding sequence from a non-human species. For example, humanized antibodies may be human antibodies in which some hypervariable region residues and possibly some FR residues are substituted by residues from analogous sites in rodent antibodies. Alternatively, humanized or human engineered antibodies may be non-human (*e.g,* rodent) antibodies in which some residues are substituted by residues from analogious sites in human antibodies (see, *e.g.,* Studnicka *et al.* (1994) Protein Eng.. 7(6) 805-814; and U.S. Patent No. 5,766,886).

[00329] The choice of human variable domains, both light and heavy, to be used in making the humanized antibodies is important to reduce antigenicity. For example, to the so-called "best-fit" method, the sequence of the variable domain of a rodent antibody is screened against the entire library of known human variable-domain sequences. The human sequence which is closest to that of the rodent is then accepted as the human framework for the humanized antibody (Sims *et al.* (1993) J. Immunol. 151:2296; Chothia *et al.* (1987) J. Mol. Biol. 196:901). Another method uses a particular framework derived from the consensus sequence of all human antibodies of a particular subgroup of light or heavy chains. The same framework may be used for several different humanized antibodies (Carter *et al.* (1992) Proc. Natl. Acad. Sci. USA, 89:4285; Presta *et al.* (1993) J. Immunol., 151:2623).

[00330] It is further important that antibodies be humanized with retention of high affinity for the antigen and other favorable biological properties. To achieve this goal, according to one method, humanized antibodies are prepared by a process of analysis of the parental sequences and various conceptual humanized products using three-dimensional models of the parental and humanized sequences. Three-dimensional immunoglobulin models are commonly available and are familiar to those skilled in the art. Computer programs are available which illustrate and display probable three-dimensional conformational structures of selected candidate immunoglobulin sequences. Inspection of these displays permits analysis of the likely role of the residues in the functioning of the candidate immunoglobulin sequence, *e.g.,* the analysis of residues that influence the ability of the candidate immunoglobulin to bind its antigen. In this way, FR residues can be selected and combined from the recipient and import sequences so that the desired antibody characteristic, such as increased affinity for the target antigen(s), is achieved. In general, the hypervariable region residues are directly and most substantially involved in influencing antigen binding.

[00331] In some embodiments, the humanized antibodies comprising an ultralong CDR3 may be deimmunized. Methods of deimmunizing an antibody or protein are well known in the art. The immunogenicity of therapeutic proteins such as antibodies is thought to result from the presence of T-cell epitopes which can bind MHC class II molecules and generate a proliferative and cytokine response in CD4+ helper T-cells. These CD4+ helper cells then collaborate with B-cells to generate an antibody response against the therapeutic protein. Removal of the T-cell epitopes are thought to be key steps in deimmunizing a recombinant protein. T-cell epitopes can be predicted by *in silico* algorithms that identify residues required for binding MHC. Alternatively, epitopes can be identified directly by utilizing peripheral blood mononuclear cells from panels of human donors and measuring their response against the therapeutic protein when incubated with antigen presenting cells. Such *in silico* and *in vitro* systems are well known in the art [Jones TD, Crompton LJ, Carr FJ, Baker MP. Methods Mol Biol. 2009;525:405-23,Deimmunization of monoclonal antibodies; and Baker M, and Jones TD. The identification and removal of immunogenicity in therapeutic proteins. *Curr. Opin. Drug Discovery Dev.* 2007; (2007); 10(2): 219-227]. When peptides are identified that bind MHC II or otherwise stimulate CD4+ cell activation, the residues of the peptide can be mutated one by one and tested for T-cell activation until a mutation is found which disrupts MHC II binding and T-cell activation. Such mutations, when found in an individual peptide, can be encoded directly in the recombinant therapeutic protein. Incubation of the whole protein

with antigen presenting cells will not induce a significant CD4+ response, indicating successful deimmunization.

Bispecific Antibodies

[00332] Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. For example, one of the binding specificities may be for a first antigen and the other may be for any other antigen. Exemplary bispecific antibodies may bind to two different epitopes of the same protein. Bispecific antibodies may also be used to localize cytotoxic agents to cells which express a particular protein. These antibodies possess a binding arm specific for the particular protein and an arm which binds the cytotoxic agent (*e.g.,* saporin, anti-interferon-α, vinca alkaloid, ricin A chain, methotrexate or radioactive isotope hapten). Bispecific antibodies may be prepared as full length antibodies or antibody fragments (*e.g.,* $F(ab')_2$ bispecific antibodies).

[00333] Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy chain-light chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, Nature, 305: 537 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of 10 different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule, which is usually done by affinity chromatography steps, is rather cumbersome, and the product yields are low. Similar procedures are disclosed in WO 93/08829, and in Traunecker *et al.*, EMBO J., 10: 3655 (1991).

[00334] According to a different approach, antibody variable domains with the desired binding specificities (antibody-antigen combining sites) are fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1), containing the site necessary for light chain binding, present in at least one of the fusions. DNAs encoding the immunoglobulin heavy chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. This provides for flexibility in adjusting the mutual proportions of the three polypeptide fragments in embodiments when unequal ratios of the three polypeptide chains used in the construction provide the optimum yields. It is, however, possible to insert the coding sequences for two or all three polypeptide chains in

one expression vector when the expression of at least two polypeptide chains in equal ratios results in high yields or when the ratios are not of particular significance.

[00335] In a preferred embodiment of this approach, the bispecific antibodies are composed of a hybrid immunoglobulin heavy chain with a first binding specificity in one arm, and a hybrid immunoglobulin heavy chain-light chain pair (providing a second binding specificity) in the other arm. This asymmetric structure may facilitate the separation of the desired bispecific compound from unwanted immunoglobulin chain combinations, as the presence of an immunoglobulin light chain in only one half of the bispecific molecule provides for a facile way of separation. This approach is disclosed in WO 94/04690. For further details of generating bispecific antibodies see, for example, Suresh *et al.*, Methods in Enzymology, 121:210 (1986).

[00336] According to another approach, the interface between a pair of antibody molecules may can be engineered to maximize the percentage of heterodimers which are recovered from recombinant cell culture. The preferred interface comprises at least a part of the C_{H3} domain of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (*e.g.,* tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (*e.g.,* alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

[00337] Bispecific antibodies include cross-linked or "heteroconjugate" antibodies. For example, one of the antibodies in the heteroconjugate may be coupled to avidin, the other to biotin. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360, WO 92/00373, and EP 03089). Heteroconjugate antibodies may be made using any convenient cross-linking methods. Suitable cross-linking agents are well known in the art, and are disclosed in U.S. Patent No. 4,676,980, along with a number of cross-linking techniques.

[00338] Techniques for generating bispecific antibodies from antibody fragments have also been described in the literature. For example, bispecific antibodies may be prepared using chemical linkage. Brennan *et al.*, Science, 229: 81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate $F(ab')_2$ fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives.

One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced may be used as agents for the selective immobilization of enzymes.

[00339] Recent progress has facilitated the direct recovery of Fab'-SH fragments from *E. coli*, which can be chemically coupled to form bispecific antibodies. Shalaby *et al.,* J. Exp. Med., 175: 217-225 (1992) describe the production of a fully humanized bispecific antibody $F(ab')_2$ molecule. Each Fab' fragment was separately secreted from *E. coli* and subjected to directed chemical coupling *in vitro* to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the HER2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

[00340] Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. See, *e.g.,* Kostelny *et al.*, J. Immunol., 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger *et al.*, Proc. Natl. Acad. Sci. USA, 90:6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (VH) connected to a light-chain variable domain (VL) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the VH and VL domains of one fragment are forced to pair with the complementary VL and VH domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber *et al.*, J. Immunol., 152:5368 (1994).

[00341] Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. See, *e.g.,* Tutt *et al.* J. Immunol. 147: 60 (1991).

Multivalent Antibodies

[00342] A multivalent antibody may be internalized (and/or catabolized) faster than a bivalent antibody by a cell expressing an antigen to which the antibodies bind. The antibodies of the present disclosure may be multivalent antibodies (which are other than of

the IgM class) with three or more antigen binding sites (*e.g.*, tetravalent antibodies), which may be produced by recombinant expression of nucleic acid encoding the polypeptide chains of the antibody. The multivalent antibody may comprise a dimerization domain and three or more antigen binding sites. A preferred dimerization domain may comprise (or consist of) an Fc region or a hinge region. In this scenario, the antibody will comprise an Fc region and three or more antigen binding sites amino-terminal to the Fe region. A preferred multivalent antibody may comprise (or consist of) three to about eight, but preferably four, antigen binding sites. The multivalent antibody comprises at least one polypeptide chain (and preferably two polypeptide chains), wherein the polypeptide chain(s) comprise two or more variable domains. For instance, the polypeptide chain(s) may comprise VD1-(X1)n-VD2-(X2)n-Fc, wherein VD1 is a first variable domain, VD2 is a second variable domain, Fc is one polypeptide chain of an Fc region, X1 and X2 represent an amino acid or polypeptide, and n is 0 or 1. For instance, the polypeptide chain(s) may comprise: VH-CH1-flexible linker-VH-CH1-Fc region chain; or VH-CH1-VH-CH1-Fc region chain. A multivalent antibody may preferably further comprises at least two (and preferably four) light chain variable domain polypeptides. A multivalent antibody may, for instance, comprise from about two to about eight light chain variable domain polypeptides. The light chain variable domain polypeptides may comprise a light chain variable domain and, optionally, further comprise a CL domain.

Antibody Variants

[00343] In some embodiments, amino acid sequence modification(s) of the antibodies comprising an ultralong CDR3 as described herein are contemplated. For example, it may be desirable to improve the binding affinity and/or other biological properties of the antibody. Amino acid sequence variants of the antibody are prepared by introducing appropriate nucleotide changes into the antibody nucleic acid, or by peptide synthesis. Such modifications include, for example, deletions from, and/or insertions into and/or substitutions of, residues within the amino acid sequences of the antibody. Any combination of deletion, insertion, and substitution is made to arrive at the final construct, provided that the final construct possesses the desired characteristics. The amino acid alterations may be introduced in the subject antibody amino acid sequence at the time that sequence is made.

[00344] A useful method for identification of certain residues or regions of the antibody that are preferred locations for mutagenesis is called "alanine scanning mutagenesis" as described by Cunningham and Wells (1989) Science, 244:1081-1085. Here, a residue or group of target residues are identified (*e.g.,* charged residues such as

arg, asp, his, lys, and glu) and replaced by a neutral or negatively charged amino acid (most preferably alanine or polyalanine) to affect the interaction of the amino acids with antigen. Those amino acid locations demonstrating functional sensitivity to the substitutions then are refined by introducing further or other variants at, or for, the sites of substitution. Thus, while the site for introducing an amino acid sequence variation is predetermined, the nature of the mutation per se need not be predetermined. For example, to analyze the performance of a mutation at a given site, ala scanning or random mutagenesis is conducted at the target codon or region and the expressed immunoglobulins are screened for the desired activity.

[00345] Amino acid sequence insertions include amino- and/or carboxyl-terminal fusions ranging in length from one residue to polypeptides containing a hundred or more residues, as well as intrasequence insertions of single or multiple amino acid residues. Examples of terminal insertions include an antibody with an N-terminal methionyl residue or the antibody fused to a cytotoxic polypeptide. Other insertional variants of the antibody molecule include the fusion to the N- or C-terminus of the antibody to an enzyme (*e.g.,* for ADEPT) or a polypeptide which increases the serum half-life of the antibody.

[00346] Glycosylation of polypeptides is typically either N-linked or O-linked. N-linked refers to the attachment of the carbohydrate moiety to the side chain of an asparagine residue. The tripeptide sequences asparagine-X-serine and asparagine-X-threonine, where X is any amino acid except proline, are the recognition sequences for enzymatic attachment of the carbohydrate moiety to the asparagine side chain. Thus, the presence of either of these tripeptide sequences in a polypeptide creates a potential glycosylation site. O-linked glycosylation refers to the attachment of one of the sugars N-aceylgalactosamine, galactose, or xylose to a hydroxyamino acid, most commonly serine or threonine, although 5-hydroxyproline or 5-hydroxylysine may also be used.

[00347] Addition of glycosylation sites to the antibody is conveniently accomplished by altering the amino acid sequence such that it contains one or more of the above-described tripeptide sequences (for N-linked glycosylation sites). The alteration may also be made by the addition of, or substitution by, one or more serine or threonine residues to the sequence of the original antibody (for O-linked glycosylation sites).

[00348] Where the antibody comprises an Fc region, the carbohydrate attached thereto may be altered. For example, antibodies with a mature carbohydrate structure that lacks fucose attached to an Fc region of the antibody have been described (see, *e.g.,* US 2003/0157108, US 2004/0093621. Antibodies with a bisecting N-acetylglucosamine (GlcNAc) in the carbohydrate attached to an Fc region of the antibody have been described (see, *e.g.,* WO 2003/011878, and U.S. Patent No. 6,602,684). Antibodies with at least one

galactose residue in the oligosaccharide attached to an Fc region of the antibody WO 1997/30087; see, also, WO 1998/58964 and WO 1999/22764 concerning antibodies with altered carbohydrate attached to the Fc region thereof). Antigen-binding molecules with modified glycosylation have been described (see, *e.g.*, WO 99/54342, U.S. Patent Nos. 6,602,684 and 7,517,670, and US 2004/0072290; see also, *e.g.*, U.S. Patent Nos. 7,214,775 and 7,682,610).

[00349] The preferred glycosylation variant herein comprises an Fc region, wherein a carbohydrate structure attached to the Fc region lacks fucose. Such variants have improved ADCC function. Optionally, the Fc region further comprises one or more amino acid substitutions therein which further improve ADCC, for example, substitutions at positions 298, 333, and/or 334 of the Fc region (Eu numbering of residues). Examples of publications related to "defucosylated" or "fucose-deficient" antibodies include: US 2003/0157108; WO 2000/61739; WO 2001/29246; US 2003/0115614 (now U.S. Patent No.6,946,292) US 2002/0164328 (now U.S. Patent No. 7,064,191); US 2004/0093621; US 2004/0132140; US 2004/0110704; US 2004/0110282 (now U.S. Patent No 7,749,753); US 2004/0109865; WO 2003/085119; WO 2003/084570; WO 2005/035586; WO 2005/035778; WO2005/053742; Okazaki *et al.* J. Mol. Biol. 336:1239-1249 (2004); Yamane-Ohnuki *et al.* Biotech. Bioeng. 87: 614 (2004). Examples of cell lines producing defucosylated antibodies include Lec13 CHO cells deficient in protein fucosylation (Ripka *et al.* Arch. Biochem. Biophys. 249:533-545 (1986); US Pat Appl No US 2003/0157108 A1, Presta, L; and WO 2004/056312 A1, Adams *et al.*, especially at Example 11), and knockout cell lines, such as alpha-1,6-fucosyltransferase gene, FUT8, knockout CHO cells (Yamane-Ohnuki *et al.* Biotech. Bioeng. 87: 614 (2004)).

[00350] Another type of variant is an amino acid substitution variant. These variants have at least one amino acid residue in the antibody molecule replaced by a different residue. The sites of greatest interest for substitutional mutagenesis include the hypervariable regions, but FR alterations are also contemplated. Conservative substitutions are shown in Table 2 under the heading of "preferred substitutions". If such substitutions result in a change in biological activity, then more substantial changes, denominated "exemplary substitutions", or as further described below in reference to amino acid classes, may be introduced and the products screened.

Original Residue	Exemplary Substitutions	Preferred Substitutions
Ala (A)	Val; Leu; Ile	Val
Arg (R)	Lys; Gln; Asn	Lys
Asn (N)	Gln; His; Asp, Lys; Arg	Gln
Asp (D)	Glu; Asn	Glu

Cys (C)	Ser; Ala	Ser
Gln (Q)	Asn; Glu	Asn
Glu (E)	Asp; Gln	Asp
Gly (G)	Ala	Ala
His (H)	Asn; Gln; Lys; Arg	Arg
Ile (I)	Leu; Val; Met; Ala; Phe; Norleucine	Leu
Leu (L)	Norleucine; Ile; Val; Met; Ala; Phe	Ile
Lys (K)	Arg; Gln; Asn	Arg
Met (M)	Leu; Phe; Ile	Leu
Phe (F)	Trp; Leu; Val; Ile; Ala; Tyr	Tyr
Pro (P)	Ala	Ala
Ser (S)	Thr	Thr
Thr (T)	Val; Ser	Ser
Trp (W)	Tyr; Phe	Tyr
Tyr (Y)	Trp; Phe; Thr; Ser	Phe
Val (V)	Ile; Leu; Met; Phe; Ala; Norleucine	Leu

[00351] Substantial modifications in the biological properties of the antibody are accomplished by selecting substitutions that differ significantly in their effect on maintaining (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site, or (c) the bulk of the side chain. Naturally occurring residues are divided into groups based on common side-chain properties: (1) hydrophobic: norleucine, met, ala, val, leu, ile; (2) neutral hydrophilic: Cys, Ser, Thr, Asn, Gln; (3) acidic: asp, glu; (4) basic: his, lys, arg; (5) residues that influence chain orientation: gly, pro; and (6) aromatic: trp, tyr, phe.

[00352] Non-conservative substitutions will entail exchanging a member of one of these classes for another class.

[00353] One type of substitutional variant involves substituting one or more hypervariable region residues of a parent antibody (*e.g.,* a humanized or human antibody). Generally, the resulting variant(s) selected for further development will have improved biological properties relative to the parent antibody from which they are generated. A convenient way for generating such substitutional variants involves affinity maturation using phage display. Briefly, several hypervariable region sites (*e.g.,* 6-7 sites) are mutated to generate all possible amino acid substitutions at each site. The antibodies thus generated are displayed from filamentous phage particles as fusions to the gene III product of M13 packaged within each particle. The phage-displayed variants are then screened for their

biological activity (*e.g.,* binding affinity) as herein disclosed. In order to identify candidate hypervariable region sites for modification, alanine scanning mutagenesis can be performed to identify hypervariable region residues contributing significantly to antigen binding. Alternatively, or additionally, it may be beneficial to analyze a crystal structure of the antigen-antibody complex to identify contact points between the antibody and antigen. Such contact residues and neighboring residues are candidates for substitution according to the techniques elaborated herein. Once such variants are generated, the panel of variants is subjected to screening as described herein and antibodies with superior properties in one or more relevant assays may be selected for further development.

[00354] Nucleic acid molecules encoding amino acid sequence variants of the antibody are prepared by a variety of methods known in the art. These methods include, but are not limited to, isolation from a natural source (in the case of naturally occurring amino acid sequence variants) or preparation by oligonucleotide-mediated (or site-directed) mutagenesis, PCR mutagenesis, and cassette mutagenesis of an earlier prepared variant or a non-variant version of the antibody.

[00355] It may be desirable to introduce one or more amino acid modifications in an Fc region of the immunoglobulin polypeptides disclosed herein, thereby generating a Fc region variant. The Fc region variant may comprise a human Fc region sequence (*e.g.,* a human IgG1, IgG2, IgG3 or IgG4 Fc region) comprising an amino acid modification (*e.g.*, a substitution) at one or more amino acid positions including that of a hinge cysteine.

[00356] In accordance with this description and the teachings of the art, it is contemplated that in some embodiments, an antibody used in methods disclosed herein may comprise one or more alterations as compared to the wild type counterpart antibody, *e.g,.* in the Fc region. These antibodies would nonetheless retain substantially the same characteristics required for therapeutic utility as compared to their wild type counterpart. For example, it is thought that certain alterations can be made in the Fc region that would result in altered (*e.g.,* either improved or diminished) Clq binding and/or Complement Dependent Cytotoxicity (CDC), *e.g.,* as described in WO99/51642. See also Duncan & Winter Nature 322:738-40 (1988); U.S. Patent No. 5,648,260; U.S. Patent No. 5,624,821; and WO94/29351 concerning other examples of Fc region variants. WO00/42072 and WO 2004/056312 describe antibody variants with improved or diminished binding to FcRs. See, also, Shields *et al.* J. Biol. Chem. 9(2): 6591-6604 (2001). Antibodies with increased half lives and improved binding to the neonatal Fc receptor (FcRn), which is responsible for the transfer of maternal IgGs to the fetus (Guyer *et al.*, J. Immunol. 117:587 (1976) and Kim *et al.*, J. Immunol. 24:249 (1994)), are described in US2005/0014934 (Hinton *et al.*). These antibodies comprise an Fc reg on with one or more substitutions therein which improve

binding of the Fc region to FcRn. Polypeptide variants with altered Fc region amino acid sequences and increased or decreased Clq binding capability are described in U.S. Patent No. 6,194,551, WO99/51642. See, also, Idusogie *et al.* J. Immunol. 164:4178-4184 (2000).

[00357] In certain embodiments, the present disclosure contemplates an antibody variant that possesses some but not all effector functions, which make it a desirable candidate for applications in which the half life of the antibody in vivo is important yet certain effector functions (such as complement and ADCC) are unnecessary or deleterious. In vitro and/or in vivo cytotoxicity assays can be conducted to confirm the reduction/ depletion of CDC and/ or ADCC activities. For example, Fc receptor (FcR) binding assays can be conducted to ensure that the antibody lacks FcyR binding (hence likely lacking ADCC activity), but retains FcRn binding ability. The primary cells for mediating ADCC, NK cells, express FcyRIII only, whereas monocytes express FcyRI, FcyRII and FcyRIII. FcR expression on hematopoietic cells is summarized in Table 3 on page 464 of Ravetch and Kinet, Annu. Rev. Immunol. 9:457-492 (1991). Non-limiting examples of in vitro assays to assess ADCC activity of a molecule of interest is described in U.S. Pat. No. 5,500,362 (see, *e.g.* Hellstrom, I. *et al.* Proc. Nat'! Acad. Sci. USA 83:7059-7063 (1986)) and Hellstrom, I *et al.*, Proc. Nat 'lAcad. Sci. USA 82:1499-1502 (1985); 5,821,337 (see, Bruggemann, M. *et al.*, Exp. Med. 166:1351-1361 (1987)). Alternatively, non-radioactive assays methods may be employed (see, for example, ACTITM non-radioactive cytotoxicity assay for flow cytometry (CellTecl1r1ology, Inc. Mountain View, Calif.; and CytoTox 96® non-radioactive cytotoxicity assay (Promega, Madison, Wis.). Useful effector cells for such assays include peripheral blood mononuclear cells (PBMC) and Natural Killer (NK) cells. Alternatively, or additionally, ADCC activity of the molecule of interest may be assessed in vivo, *e.g.*, in a animal model such as that disclosed in Clynes *et al.* Proc. Nat 'lAcad. Sci. USA 95:652-656 (1998). C1qbinding assays may also be carried out to confirm that the antibody is unable to bind Clq and hence lacks CDC activity. See, *e.g.*, C1q and C3c binding ELISA in WO 2006/029879 and WO 2005/ 100402. To assess complement activation, a CDC assay may be performed (see, for example, Gazzano-Santoro *et al.*, Immunol. Methods 202:163 (1996); Cragg, M. S. *et al.*, Blood 101:1045-1052 (2003); and Cragg, M. S, and M. J. Glennie, Blood 103:27382743 (2004)). FcRn binding and in vivo clearance/half life determinations can also be performed using methods known in the art (see, *e.g.*, Petkova, S. B. *et al.*, Int 'l Immunol. 18(12):1759-1769 (2006)).

[00358] Antibodies with reduced effector function include those with substitution of one or more of Fc region residues 238, 265, 269, 270, 297, 327 and 329 (U.S. Pat. No. 6,737, 056). Such Fc mutants include Fc mutants with substitutions at two or more of

amino acid positions 265, 269, 270, 297 and 327, including the so-called "DANA" Fc mutant with substitution of residues 265 and 297 to alanine (U .S. Pat. No. 7,332,581).

[00359] Certain antibody variants with improved or diminished binding to FcRs are described. (See, *e.g.*, U.S. Pat. No. 6,737,056; WO 2004/056312, and Shields *et al.*, Biol. Chem. 9(2): 6591-6604 (2001).)

[00360] In certain embodiments, an antibody variant comprises an Fc region with one or more amino acid substitutions which improve ADCC, *e.g.*, substitutions at positions 298, 333, and/or 334 of the Fc region (EU numbering of residues).

[00361] In some embodiments, alterations are made in the Fc region that result in altered (*i.e.*, either improved or diminished) C1q binding and/or Complement Dependent Cytotoxicity (CDC), *e.g.*, as described in U.S. Pat. No. 6,194,551, WO 99/51642, and Idusogie *et al.* Immunol. 164: 41784184 (2000).

[00362] Antibodies with increased half lives and improved binding to the neonatal Fc receptor (FcRn), which is responsible for the transfer of maternal IgGs to the fetus (Guyer *et al.*, Immunol. 117:587 (1976) and Kim *et al.*, Immunol. 24:249 (1994)), are described in US2005/0014934A1 (Hinton *et al.*). Those antibodies comprise an Fc region with one or more substitutions therein which improve binding of the Fc region to FcRn. Such Fc variants include those with substitutions at one or more of Fc region residues: 238, 256, 265, 272, 286,303,305, 307, 311,312, 317, 340,356, 360, 362, 376, 378, 380, 382, 413, 424 or 434, *e.g.*, substitution of Fc region residue 434 (U.S. Pat. No. 7,371,826).

[00363] See also Duncan & Winter, Nature 322:738-40 (1988); U.S. Pat. No. 5,648,260; U.S. Pat. No. 5,624,821; and WO 94/29351 concerning other examples of Fc region variants.

Antibody Derivatives

[00364] The antibodies comprising an ultralong CDR3 as disclosed herein may be further modified to contain additional nonproteinaceous moieties that are known in the art and readily available. Preferably, the moieties suitable for derivatization of the antibody are water soluble polymers. Non-limiting examples of water soluble polymers include, but are not limited to, polyethylene glycol (PEG), copolymers of ethylene glycol/propylene glycol, carboxymethylcellulose, dextran, polyvinyl alcohol, polyvinyl pyrrolidone, poly-1,3-dioxolane, poly-1,3,6-trioxane, ethylene/maleic anhydride copolymer, polyaminoacids (either homopolymers or random copolymers), and dextran or poly(n-vinyl pyrrolidone)polyethylene glycol, propropylene glycol homopolymers, prolypropylene oxide/ethylene oxide co-polymers, polyoxyethylated polyols (*e.g.,* glycerol), polyvinyl alcohol, and mixtures thereof. Polyethylene glycol propionaldehyde may have advantages

in manufacturing due to its stability in water. The polymer may be of any molecular weight, and may be branched or unbranched. The number of polymers attached to the antibody may vary, and if more than one polymers are attached, they can be the same or different molecules. In general, the number and/or type of polymers used for derivatization can be determined based on considerations including, but not limited to, the particular properties or functions of the antibody to be improved, whether the antibody derivative will be used in a therapy under defined conditions, *etc.*

Vectors, Host Cells and Recombinant Methods

[00365] For recombinant production of an antibody or fragment thereof as disclosed herein, the nucleic acid encoding it is isolated and inserted into a replicable vector for further cloning (amplification of the DNA) or for expression. DNA encoding the antibody is readily isolated and sequenced using conventional procedures (*e.g.,* by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of the antibody). In an exemplary embodiment, nucleic acid encoding an antibody heavy chain variable region, antibody or binding fragment thereof comprising an ultralong CDR3, a variable region comprising an ultralong CDR3, or an ultralong CDR3, is isolated and inserted into a replicable vector for further cloning (amplification of the DNA) or for expression. Many vectors are available. The choice of vector depends in part on the host cell to be used. Generally, preferred host cells are of either prokaryotic or eukaryotic (generally mammalian) origin. It will be appreciated that constant regions of any isotype can be used for this purpose, including IgG, IgM, IgA, IgD, and IgE constant regions, and that such constant regions can be obtained from any human or animal species.

[00366] Expression vectors containing regulatory elements from eukaryotic viruses are typically used in eukaryotic expression vectors, *e.g.*, SV40 vectors, papilloma virus vectors, and vectors derived from Epstein-Barr virus. Other exemplary eukaryotic vectors include pMSG, pAV009/A+, pMTO10/A+, pMAMneo-5, baculovirus pDSVE, and any other vector allowing expression of proteins under the direction of the CMV promoter, SV40 early promoter, SV40 later promoter, metallothionein promoter, murine mammary tumor virus promoter, Rous sarcoma virus promoter, polyhedrin promoter, or other promoters shown effective for expression in eukaryotic cells.

[00367] Some expression systems have markers that provide gene amplification such as thymidine kinase and dihydrofolate reductase. Alternatively, high yield expression systems not involving gene amplification are also suitable, such as using a baculovirus vector in insect cells, with a nucleic acid sequence encoding a partially human ultralong

CDR3 antibody chain under the direction of the polyhedrin promoter or other strong baculovirus promoters.

a. Generating Antibodies Using Prokaryotic or Eukaryotic Host Cells:

i. Vector Construction

[00368] Polynucleotide sequences encoding polypeptide components of the antibodies disclosed herein can be obtained using standard recombinant techniques. Desired polynucleotide sequences may be isolated and sequenced from antibody producing cells such as hybridoma cells. Alternatively, polynucleotides can be synthesized using nucleotide synthesizer or PCR techniques. Once obtained, sequences encoding the polypeptides are inserted into a recombinant vector capable of replicating and expressing heterologous polynucleotides in prokaryotic hosts. Many vectors that are available and known in the art can be used for the purpose of the present disclosure. Selection of an appropriate vector will depend mainly on the size of the nucleic acids to be inserted into the vector and the particular host cell to be transformed with the vector. Each vector contains various components, depending on its function (amplification or expression of heterologous polynucleotide, or both) and its compatibility with the particular host cell in which it resides. The vector components generally include, but are not limited to: an origin of replication, a selection marker gene, a promoter, a ribosome binding site (RBS), a signal sequence, the heterologous nucleic acid insert and a transcription termination sequence. Additionally, V regions comprising an ultralong CDR3 may optionally be fused to a C-region to produce an antibody comprising constant regions.

[00369] In general, plasmid vectors containing replicon and control sequences which are derived from species compatible with the host cell are used in connection with these hosts. The vector ordinarily carries a replication site, as well as marking sequences which are capable of providing phenotypic selection in transformed cells. For example, *E. coli* is typically transformed using pBR322, a plasmid derived from an *E. coli* species. pBR322 contains genes encoding ampicillin (Amp) and tetracycline (Tet) resistance and thus provides easy means for identifying transformed cells. pBR322, its derivatives, or other microbial plasmids or bacteriophage may also contain, or be modified to contain, promoters which can be used by the microbial organism for expression of endogenous proteins. Examples of pBR322 derivatives used for expression of particular antibodies have been described (see, *e.g.,* U.S. Patent No. 5,648,237).

[00370] In addition, phage vectors containing replicon and control sequences that are compatible with the host microorganism can be used as transforming vectors in connection with these hosts. For example, bacteriophage such as λGEM™-11 may be

utilized in making a recombinant vector which can be used to transform susceptible host cells such as *E. coli* LE392.

[00371] The expression vectors disclosed herein may comprise two or more promoter-cistron pairs, encoding each of the polypeptide components. A promoter is an untranslated regulatory sequence located upstream (5') to a cistron that modulates its expression. Prokaryotic promoters typically fall into two classes, inducible and constitutive. Inducible promoter is a promoter that initiates increased levels of transcription of the cistron under its control in response to changes in the culture condition, *e.g.,* the presence or absence of a nutrient or a change in temperature.

[00372] A large number of promoters recognized by a variety of potential host cells are well known. The selected promoter can be operably linked to cistron DNA encoding the light or heavy chain by removing the promoter from the source DNA via restriction enzyme digestion and inserting the isolated promoter sequence into the vector disclosed herein. Both the native promoter sequence and many heterologous promoters may be used to direct amplification and/or expression of the target genes. In some embodiments, heterologous promoters are utilized, as they generally permit greater transcription and higher yields of expressed target gene as compared to the native target polypeptide promoter.

[00373] Promoters suitable for use with prokaryotic hosts include: an ara B promoter, a PhoA promoter, β-galactamase and lactose promoter systems, a tryptophan (trp) promoter system and hybrid promoters such as the tac or the trc promoter. However, other promoters that are functional in bacteria (such as other known bacterial or phage promoters) are suitable as well. Their nucleotide sequences have been published, thereby enabling a skilled worker operably to ligate them to cistrons encoding the target light and heavy chains (*e.g.,* Siebenlist *et al.* (1980) Cell 20: 269) using linkers or adaptors to supply any required restriction sites.

[00374] Suitable bacterial promoters are well known in the art and fully described in scientific literature such as Sambrook and Russell, supra, and Ausubel et al, supra. Bacterial expression systems for expressing antibody chains of the recombinant catalytic polypeptide are available in, *e.g.*, E. coli, Bacillus sp., and Salmonella (Palva *et al.*, Gene, 22:229-235 (1983); Mosbach *et al.*, Nature, 302:543-545 (1983)).

[00375] In one aspect disclosed herein, each cistron within the recombinant vector comprises a secretion signal sequence component that directs translocation of the expressed polypeptides across a membrane. In general, the signal sequence may be a component of the vector, or it may be a part of the target polypeptide DNA that is inserted into the vector. The signal sequence should be one that is recognized and processed (*e.g.*,

cleaved by a signal peptidase) by the host cell. For prokaryotic host cells that do not recognize and process the signal sequences native to the heterologous polypeptides, the signal sequence is substituted by a prokaryotic signal sequence selected, for example PelB, OmpA, alkaline phosphatase, penicillinase, lpp, or heat-stable enterotoxin II (STII) leaders, LamB, PhoE, and MBP. In one embodiment disclosed herein, the signal sequences used in both cistrons of the expression system are STII signal sequences or variants thereof.

[00376] In another aspect, the production of the immunoglobulins according to the disclosure can occur in the cytoplasm of the host cell, and therefore does not require the presence of secretion signal sequences within each cistron. In that regard, immunoglobulin light and heavy chains are expressed, folded and assembled to form functional immunoglobulins within the cytoplasm. Certain host strains (*e.g.,* the *E. coli* trxB-strains) provide cytoplasm conditions that are favorable for disulfide bond formation, thereby permitting proper folding and assembly of expressed protein subunits (see *e.g.,* Proba and Pluckthun Gene, 159:203 (1995)).

[00377] Suitable host cells for cloning or expression of antibody-encoding vectors include prokaryotic or eukaryotic cells described herein. In one embodiment, the host cell is eukaryotic, *e.g.* a Chinese Hamster Ovary (CHO) cell, Human Embryonic Kidney (HEK) cell or lymphoid cell (*e.g.*, YO, NSO, Sp20 cell). For example, antibodies may be produced in bacteria, in particular when glycosylation and Fc effector function are not needed. For expression of antibody fragments and polypeptides in bacteria, see, *e.g.*, U.S. Pat. Nos. 5,648,237, 5,789,199, and 5,840,523. (See also Charlton, Methods in Molecular Biology, Vol. 248 (B.K.C. Lo, ed., Humana Press, Totowa, N.J., 2003), pp. 245-254, describing expression of antibody fragments in E. coli.) After expression, the antibody may be isolated from the bacterial cell paste in a soluble fraction and can be further purified. In addition to prokaryotes, eukaryotic microbes such as filamentous fungi or yeast are suitable cloning or expression hosts for antibody-encoding vectors, including fungi and yeast strains whose glycosylation pathways have been “humanized,” resulting in the production of an antibody with a partially or fully human glycosylation pattern. See Gemgross, Nat. Biotech. 22: 1409-1414 (2004), and Li *et al.*, Nat. Biotech. 24:210-215 (2006). Suitable host cells for the expression of glycosylated antibody are also derived from multicellular organisms (invertebrates and vertebrates). Examples of invertebrate cells include plant and insect cells. Numerous baculoviral strains have been identified which may be used in conjunction with insect cells, particularly for transfection of *Spodoptera frugiperda* cells. These examples are illustrative rather than limiting. Methods for constructing derivatives of any of the above-mentioned bacteria having defined genotypes are known in the art and

described in, for example, Bass *et al.*, Proteins, 8:309-314 (1990). It is generally necessary to select the appropriate bacteria taking into consideration replicability of the replicon in the cells of a bacterium. For example, *E. coli*, Serratia, or Salmonella species can be suitably used as the host when well known plasmids such as pBR322, pBR325, pACYC177, or pKN410 are used to supply the replicon. Typically the host cell should secrete minimal amounts of proteolytic enzymes, and additional protease inhibitors may desirably be incorporated in the cell culture.

[00378] Plant cell cultures can also be utilized as hosts. See, *e.g.* U.S. Pat. Nos. 5,959,177, 6,040,498, 6,420,548, 7,125, 978, and 6,417,429 (describing PLANTIBODIESTM technology for producing antibodies in transgenic plants). Vertebrate cells may also be used as hosts. For example, mammalian cell lines that are adapted to grow in suspension may be useful. Other examples of useful mammalian host cell lines are monkey kidney CV1 line transformed by SV40 (COS-7); human embryonic kidney line (293 or 293 cells as described, *e.g.*, in Graham *et al.*, Gen VII'0l. 36:59 (1977)); baby hamster kidney cells (BHK); mouse sertoli cells (TM4 cells as described, *e.g.*, in Mather, Biol. Reprod. 23:243-251 (1980)); monkey kidney cells (CV1); African green monkey kidney cells (V ERO-76); human cervical carcinoma cells (HELA); canine kidney cells (MDCK; buffalo rat liver cells (BRL 3A); human lung cells (W138); human liver cells (Hep G2); mouse mammary tumor (MMT 060562); TR1 cells, as described, *e.g.*, in Mather *et al.*, Annals NI'.Acad. Sci. 383:44-68 (1982); MRC 5 cells; and FS4 cells. Other useful mammalian host cell lines include Chinese hamster ovary (CHO) cells, including DHFR‘ CHO cells (Urlaub *et al.*, Proc. Natl. Acad. Sci. USA 77:4216 (1980)); and myeloma cell lines such as YO, NSO and Sp2/0. For a review of certain mammalian host cell lines suitable for antibody production, see, *e.g.*, Yazaki and Wu, Methods in Molecular Biology, Vol. 248 (B.K.C. Lo, ed., Humana Press, Totowa, N.].), pp. 255-268 (2003).

[00379] In one such embodiment, a host cell comprises (*e.g.*, has been transformed with): (1) a vector comprising a nucleic acid that encodes an amino acid sequence comprising the VL of the antibody and an amino acid sequence comprising the VH of the antibody, or (2) a first vector comprising a nucleic acid that encodes an amino acid sequence comprising the VL of the antibody and a second vector comprising a nucleic acid that encodes an amino acid sequence comprising the VH of the antibody.

ii. Antibody Production

[00380] For recombinant production of a partially human ultralong CDR3 antibody, nucleic acid encoding an antibody comprising an ultralong CDR3 is inserted into one or more expression vectors for further cloning and/or expression in a host cell. Such nucleic acid may be readily isolated and sequenced using conventional procedures (*e.g.*, by using

oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of the antibody). Host cells are transformed with such expression vectors and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences.

[00381] Transformation means introducing DNA into the prokaryotic host so that the DNA is replicable, either as an extrachromosomal element or by chromosomal integrant. Depending on the host cell used, transformation is done using standard techniques appropriate to such cells. The calcium treatment employing calcium chloride is generally used for bacterial cells that contain substantial cell-wall barriers. Another method for transformation employs polyethylene glycol/DMSO. Yet another technique used is electroporation.

[00382] Prokaryotic cells used to produce the polypeptides disclosed herein are grown in media known in the art and suitable for culture of the selected host cells. Examples of suitable media include luria broth (LB) plus necessary nutrient supplements. In some embodiments, the media also contains a selection agent, chosen based on the construction of the expression vector, to selectively permit growth of prokaryotic cells containing the expression vector. For example, ampicillin is added to media for growth of cells expressing ampicillin resistant gene.

[00383] Any necessary supplements besides carbon, nitrogen, and inorganic phosphate sources may also be included at appropriate concentrations introduced alone or as a mixture with another supplement or medium such as a complex nitrogen source. Optionally the culture medium may contain one or more reducing agents selected from the group consisting of glutathione, cysteine, cystamine, thioglycollate, dithioerythritol and dithiothreitol.

[00384] The prokaryotic host cells are cultured at suitable temperatures. For *E. coli* growth, for example, the preferred temperature ranges from about 20° C. to about 39° C., more preferably from about 25° C. to about 37° C., even more preferably at about 30° C. The pH of the medium may be any pH ranging from about 5 to about 9, depending mainly on the host organism. For *E. coli*, the pH is preferably from about 6.8 to about 7.4, and more preferably about 7.0.

[00385] If an inducible promoter is used in the expression vector disclosed herein, protein expression is induced under conditions suitable for the activation of the promoter. For example, an ara B or phoA promoter may be used for controlling transcription of the polypeptides. A variety of inducers may be used, according to the vector construct employed, as is known in the art.

[00386] The expressed polypeptides of the present disclosure are secreted into and recovered from the periplasm of the host cells or transported into the culture media. Protein recovery from the periplasm typically involves disrupting the microorganism, generally by such means as osmotic shock, sonication or lysis. Once cells are disrupted, cell debris or whole cells may be removed by centrifugation or filtration. The proteins may be further purified, for example, by affinity resin chromatography. Alternatively, proteins that are transported into the culture media may be isolated therein. Cells may be removed from the culture and the culture supernatant being filtered and concentrated for further purification of the proteins produced. The expressed polypeptides can be further isolated and identified using commonly known methods such as polyacrylamide gel electrophoresis (PAGE) and Western blot assay.

[00387] Antibody production may be conducted in large quantity by a fermentation process. Various large-scale fed-batch fermentation procedures are available for production of recombinant proteins. Large-scale fermentations have at least 1000 liters of capacity, preferably about 1,000 to 100,000 liters of capacity. These fermentors use agitator impellers to distribute oxygen and nutrients, especially glucose (a preferred carbon/energy source). Small scale fermentation refers generally to fermentation in a fermentor that is no more than approximately 100 liters in volumetric capacity, and can range from about 1 liter to about 100 liters.

[00388] In a fermentation process, induction of protein expression is typically initiated after the cells have been grown under suitable conditions to a desired density, *e.g.,* an OD550 of about 180-220, at which stage the cells are in the early stationary phase. A variety of inducers may be used, according to the vector construct employed, as is known in the art and described above. Cells may be grown for shorter periods prior to induction. Cells are usually induced for about 12-50 hours, although longer or shorter induction time may be used.

[00389] To improve the production yield and quality of the polypeptides disclosed herein, various fermentation conditions can be modified. For example, to improve the proper assembly and folding of the secreted antibody polypeptides, additional vectors overexpressing chaperone proteins, such as Dsb proteins (DsbA, DsbB, DsbC, DsbD and or DsbG) or FkpA (a peptidylprolyl cis,trans-isomerase with chaperone activity) may be used to co-transform the host prokaryotic cells. The chaperone proteins have been demonstrated to facilitate the proper folding and solubility of heterologous proteins produced in bacterial host cells. (see *e.g.,* Chen *et al.* (1999) J Bio Chem 274:19601-19605; U.S. Patent No. 6,083,715; U.S. Patent No. 6,027,888; Bothmann and Pluckthun

(2000) J. Biol. Chem. 275:17100-17105; Ramm and Pluckthun (2000) J. Biol. Chem. 275:17106-17113; Arie *et al.* (2001) Mol. Microbiol. 39:199-210).

[00390] To minimize proteolysis of expressed heterologous proteins (especially those that are proteolytically sensitive), certain host strains deficient for proteolytic enzymes can be used for the present disclosure. For example, host cell strains may be modified to effect genetic mutation(s) in the genes encoding known bacterial proteases such as Protease III, OmpT, DegP, Tsp, Protease I, Protease Mi, Protease V, Protease VI and combinations thereof. Some *E. coli* protease-deficient strains are available (see, *e.g.*, Joly *et al.* (1998), *supra*; U.S. Patent No. 5,264,365; U.S. Patent No. 5,508,192; Hara *et al.*, Microbial Drug Resistance, 2:63-72 (1996)).

[00391] *E. coli* strains deficient for proteolytic enzymes and transformed with plasmids overexpressing one or more chaperone proteins may be used as host cells in the expression systems disclosed herein.

iii. Antibody Purification

[00392] Standard protein purification methods known in the art can be employed. The following procedures are exemplary of suitable purification procedures: fractionation on immunoaffinity or ion-exchange columns, ethanol precipitation, reverse phase HPLC, chromatography on silica or on a cation-exchange resin such as DEAE, chromatofocusing, SDS-PAGE, ammonium sulfate precipitation, and gel filtration using, for example, Sephadex G-75.

[00393] In one aspect, Protein A immobilized on a solid phase is used for immunoaffinity purification of the full length antibody products disclosed herein. Protein A is a 41 kD cell wall protein from Staphylococcus aureas which binds with a high affinity to the Fc region of antibodies (see, *e.g.*, Lindmark *et al* (1983) J. Immunol. Meth. 62:1-13). The solid phase to which Protein A is immobilized is preferably a column comprising a glass or silica surface, more preferably a controlled pore glass column or a silicic acid column. In some applications, the column has been coated with a reagent, such as glycerol, in an attempt to prevent nonspecific adherence of contaminants.

[00394] As the first step of purification, the preparation derived from the cell culture as described above is applied onto the Protein A immobilized solid phase to allow specific binding of the antibody of interest to Protein A. The solid phase is then washed to remove contaminants non-specifically bound to the solid phase. Finally the antibody of interest is recovered from the solid phase by elution.

b. Generating Antibodies Using Eukaryotic Host Cells:

[00395] The vector components generally include, but are not limited to, one or more of the following: a signal sequence, an origin of replication, one or more marker genes, an enhancer element, a promoter, and a transcription termination sequence.

i. Signal Sequence Component

[00396] A vector for use in a eukaryotic host cell may also contain a signal sequence or other polypeptide having a specific cleavage site at the N-terminus of the mature protein or polypeptide of interest. The heterologous signal sequence selected preferably is one that is recognized and processed (*e.g.,* cleaved by a signal peptidase) by the host cell. In mammalian cell expression, mammalian signal sequences as well as viral secretory leaders, for example, the herpes simplex gD signal, are available.

[00397] The DNA for such precursor region is ligated in reading frame to DNA encoding the antibody.

ii. Origin of Replication

[00398] Generally, an origin of replication component is not needed for mammalian expression vectors. For example, the SV40 origin may be used only because it contains the early promoter.

iii. Selection Gene Component

[00399] Expression and cloning vectors may contain a selection gene, also termed a selectable marker. Typical selection genes encode proteins that (a) confer resistance to antibiotics or other toxins, *e.g.,* ampicillin, neomycin, methotrexate, or tetracycline, (b) complement auxotrophic deficiencies, where relevant, or (c) supply critical nutrients not available from complex media.

[00400] One example of a selection scheme utilizes a drug to arrest growth of a host cell. Those cells that are successfully transformed with a heterologous gene produce a protein conferring drug resistance and thus survive the selection regimen. Examples of such dominant selection use the drugs neomycin, mycophenolic acid and hygromycin.

[00401] Another example of suitable selectable markers for mammalian cells are those that enable the identification of cells competent to take up the antibody nucleic acid, such as DHFR, thymidine kinase, metallothionein-I and -II, preferably primate metallothionein genes, adenosine deaminase, ornithine decarboxylase, *etc.*

[00402] For example, cells transformed with the DHFR selection gene are first identified by culturing all of the transformants in a culture medium that contains methotrexate (Mtx), a competitive antagonist of DHFR. An appropriate host cell when wild-type DHFR is employed is the Chinese hamster ovary (CHO) cell line deficient in DHFR activity (*e.g.,* ATCC CRL-9096).

[00403] Alternatively, host cells (particularly wild-type hosts that contain endogenous DHFR) transformed or co-transformed with DNA sequences encoding an antibody, wild-type DHFR protein, and another selectable marker such as aminoglycoside 3'-phosphotransferase (APH) can be selected by cell growth in medium containing a selection agent for the selectable marker such as an aminoglycosidic antibiotic, *e.g.,* kanamycin, neomycin, or G418. See U.S. Patent No. 4,965,199.

iv. Promoter Component

[00404] Expression and cloning vectors usually contain a promoter that is recognized by the host organism and is operably linked to the antibody polypeptide nucleic acid. Promoter sequences are known for eukaryotes. Virtually alleukaryotic genes have an AT-rich region located approximately 25 to 30 bases upstream from the site where transcription is initiated. Another sequence found 70 to 80 bases upstream from the start of transcription of many genes is a CNCAAT region where N may be any nucleotide. At the 3' end of most eukaryotic genes is an AATAAA sequence that may be the signal for addition of the poly A tail to the 3' end of the coding sequence. All of these sequences are suitably inserted into eukaryotic expression vectors.

[00405] Antibody polypeptide transcription from vectors in mammalian host cells is controlled, for example, by promoters obtained from the genomes of viruses such as polyoma virus, fowlpox virus, adenovirus (such as Adenovirus 2), bovine papilloma virus, avian sarcoma virus, cytomegalovirus, a retrovirus, hepatitis-B virus and Simian Virus 40 (SV40), from heterologous mammalian promoters, *e.g.,* the actin promoter or an immunoglobulin promoter, from heat-shock promoters, provided such promoters are compatible with the host cell systems.

[00406] The early and late promoters of the SV40 virus are conveniently obtained as an SV40 restriction fragment that also contains the SV40 viral origin of replication. The immediate early promoter of the human cytomegalovirus is conveniently obtained as a HindIII E restriction fragment. A system for expressing DNA in mammalian hosts using the bovine papilloma virus as a vector is disclosed in U.S. Patent No. 4,419,446. A modification of this system is described in U.S. Patent No. 4,601,978. Alternatively, the Rous Sarcoma Virus long terminal repeat can be used as the promoter.

v. Enhancer Element Component

[00407] Transcription of DNA encoding the antibody polypeptide of this disclosure by higher eukaryotes is often increased by inserting an enhancer sequence into the vector. Many enhancer sequences are now known from mammalian genes (globin, elastase, albumin, α-fetoprotein, and insulin). An enhancer from a eukaryotic cell virus may also be used. Examples include the SV40 enhancer on the late side of the replication origin (bp

100-270), the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers. See also Yaniv, Nature 297:17-18 (1982) on enhancing elements for activation of eukaryotic promoters. The enhancer may be spliced into the vector at a position 5' or 3' to the antibody polypeptide-encoding sequence, but is preferably located at a site 5' from the promoter.

vi. Transcription Termination Component

[00408] Expression vectors used in eukaryotic host cells will typically also contain sequences necessary for the termination of transcription and for stabilizing the mRNA. Such sequences are commonly available from the 5' and, occasionally 3', untranslated regions of eukaryotic or viral DNAs or cDNAs. These regions contain nucleotide segments transcribed as polyadenylated fragments in the untranslated portion of the mRNA encoding an antibody. One useful transcription termination component is the bovine growth hormone polyadenylation region. See WO94/11026 and the expression vector disclosed therein.

vii. Selection and Transformation of Host Cells

[00409] Suitable host cells for cloning or expressing the DNA in the vectors herein include higher eukaryote cells described herein, including vertebrate host cells. Propagation of vertebrate cells in culture (tissue culture) has become a routine procedure. Examples of useful mammalian host cell lines are monkey kidney CV1 line transformed by SV40 (COS-7, ATCC CRL 1651); human embryonic kidney line (293 or 293 cells subcloned for growth in suspension culture, Graham *et al.*, J. Gen Virol. 36:59 (1977)); baby hamster kidney cells (BHK, ATCC CCL 10); Chinese hamster ovary cells/-DHFR (CHO, Urlaub *et al.*, Proc. Natl. Acad. Sci. USA 77:4216 (1980)); mouse sertoli cells (TM4, Mather, Biol. Reprod. 23:243-251 (1980); monkey kidney cells (CV1 ATCC CCL 70); African green monkey kidney cells (VERO-76, ATCC CRL-1587); human cervical carcinoma cells (HELA, ATCC CCL 2); canine kidney cells (MDCK, ATCC CCL 34); buffalo rat liver cells (BRL 3A, ATCC CRL 1442); human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); mouse mammary tumor (MMT 060562, ATCC CCL51); TRI cells (Mather *et al.*, Annals N.Y. Acad. Sci. 383:44-68 (1982)); MRC 5 cells; FS4 cells; and a human hepatoma line (Hep G2).

[00410] Any of the well-known procedures for introducing foreign nucleotide sequences into host cells may be used. These include the use of calcium phosphate transfection, polybrene, protoplast fusion, electroporation, biolistics, liposomes, microinjection, plasma vectors, viral vectors and any of the other well known methods for introducing cloned genomic DNA, cDNA, synthetic DNA, or other foreign genetic material into a host cell (see, *e.g.*, Sambrook and Russell, supra). It is only necessary that the

particular genetic engineering procedure used be capable of successfully introducing at least both genes into the host cell capable of expressing germline antibody polypeptide.

[00411] Host cells are transformed with the above-described expression or cloning vectors for antibody production and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants, or amplifying the genes encoding the desired sequences.

viii. Culturing the Host Cells

[00412] The host cells used to produce an antibody of this disclosure may be cultured in a variety of media. Commercially available media such as Ham's F10 (Sigma), Minimal Essential Medium ((MEM), (Sigma), RPMI-1640 (Sigma), and Dulbecco's Modified Eagle's Medium ((DMEM), Sigma) are suitable for culturing the host cells. In addition, any of the media described in Ham *et al.*, Meth. Enz. 58:44 (1979), Barnes *et al.*, Anal. Biochem. 102:255 (1980), U.S. Patent Nos. 4,767,704; 4,657,866; 4,927,762; 4,560,655; or 5,122,469; WO 90/03430; WO 87/00195; or U.S. Patent Reissue 30,985 may be used as culture media for the host cells. Any of these media may be supplemented as necessary with hormones and/or other growth factors (such as insulin, transferrin, or epidermal growth factor), salts (such as sodium chloride, calcium, magnesium, and phosphate), buffers (such as HEPES), nucleotides (such as adenosine and thymidine), antibiotics (such as GENTAMYCIN™ drug), trace elements (defined as inorganic compounds usually present at final concentrations in the micromolar range), and glucose or an equivalent energy source. Any other necessary supplements may also be included at appropriate concentrations that would be known to those skilled in the art. The culture conditions, such as temperature, pH, and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.

ix. Purification of Antibody

[00413] When using recombinant techniques, the antibody can be produced intracellularly, or directly secreted into the medium. If the antibody is produced intracellularly, as a first step, the particulate debris, either host cells or lysed fragments, are removed, for example, by centrifugation or ultrafiltration. Where the antibody is secreted into the medium, supernatants from such expression systems are generally first concentrated using a commercially available protein concentration filter, for example, an Amicon or Millipore Pellicon® ultrafiltration unit. A protease inhibitor such as PMSF may be included in any of the foregoing steps to inhibit proteolysis and antibiotics may be included to prevent the growth of adventitious contaminants.

[00414] The antibody composition prepared from the cells can be purified using, for example, hydroxylapatite chromatography, gel electrophoresis, dialysis, and affinity

chromatography, with affinity chromatography being the preferred purification technique. The suitability of protein A as an affinity ligand depends on the species and isotype of any immunoglobulin Fc domain that is present in the antibody. Protein A can be used to purify antibodies that are based on human γ1, γ2, or γ4 heavy chains (Lindmark *et al.*, J. Immunol. Meth. 62:1-13 (1983)). Protein G is recommended for all mouse isotypes and for human γ3 (Guss *et al.*, EMBO J. 5:15671575 (1986)). The matrix to which the affinity ligand is attached is most often agarose, but other matrices are available. Mechanically stable matrices such as controlled pore glass or poly(styrenedivinyl)benzene allow for faster flow rates and shorter processing times than can be achieved with agarose. Where the antibody comprises a CH3 domain, the Bakerbond ABX™ resin (J. T. Baker, Phillipsburg, N.J.) is useful for purification. Other techniques for protein purification such as fractionation on an ion-exchange column, ethanol precipitation, Reverse Phase HPLC, chromatography on silica, chromatography on heparin SEPHAROSE™ chromatography on an anion or cation exchange resin (such as a polyaspartic acid column), chromatofocusing, SDS-PAGE, and ammonium sulfate precipitation are also available depending on the antibody to be recovered.

[00415] Soluble forms of antibody or fragment present either in the cytoplasm or released from the periplasmic space may be further purified using methods known in the art, for example Fab fragments are released from the bacterial periplasmic space by osmotic shock techniques.

[00416] If inclusion bodies comprising an antibody or fragment have formed, they can often bind to the inner and/or outer cellular membranes and thus will be found primarily in the pellet material after centrifugation. The pellet material can then be treated at pH extremes or with chaotropic agent such as a detergent, guanidine, guanidine derivatives, urea, or urea derivatives in the presence of a reducing agent such as dithiothreitol at alkaline pH or tris carboxyethyl phosphine at acid pH to release, break apart, and solubilize the inclusion bodies. The soluble antibody or fragment can then be analyzed using gel electrophoresis, immunoprecipitation or the like. If it is desired to isolate a solublized antibody or antigen binding fragment isolation may be accomplished using standard methods such as those set forth below and in Marston *et al.* (Meth. Enz., 182:264-275 (1990)).

[00417] Following any preliminary purification step(s), the mixture comprising the antibody of interest and contaminants may be subjected to low pH hydrophobic interaction chromatography using an elution buffer at a pH between about 2.5-4.5, preferably performed at low salt concentrations (*e.g.,* from about 0-0.25 M salt).

[00418] In some cases, an antibody or fragment may not be biologically active upon isolation. Various methods for "refolding" or converting a polypeptide to its tertiary structure and generating disulfide linkages, can be used to restore biological activity. Such methods include exposing the solubilized polypeptide to a pH usually above 7 and in the presence of a particular concentration of a chaotrope. The selection of chaotrope is very similar to the choices used for inclusion body solubilization, but usually the chaotrope is used at a lower concentration and is not necessarily the same as chaotropes used for the solubilization. In most cases the refolding/oxidation solution will also contain a reducing agent or the reducing agent plus its oxidized form in a specific ratio to generate a particular redox potential allowing for disulfide shuffling to occur in the formation of the protein's cysteine bridge(s). Some of the commonly used redox couples include cysteine/cystamine, glutathione (GSH)/dithiobis GSH, cupric chloride, dithiothreitol(DTT)/dithiane DTT, and 2-mercaptoethanol(bME)/di-thio-b(ME). In many instances, a cosolvent may be used to increase the efficiency of the refolding, and common reagents used for this purpose include glycerol, polyethylene glycol of various molecular weights, arginine and the like.

Immunoconjugates

[00419] The disclosure also provides immunoconjugates (interchangeably termed "antibody-drug conjugates" or "ADC"), comprising any of the antibodies comprising an ultralong CDR3 as described herein conjugated to a cytotoxic agent such as a chemotherapeutic agent, a drug, a growth inhibitory agent, a toxin (*e.g.,* an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (*e.g.,* a radioconjugate).

[00420] The use of antibody-drug conjugates for the local delivery of cytotoxic or cytostatic agents. For example, drugs to kill or inhibit tumor cells in the treatment of cancer (Syrigos and Epenetos (1999) Anticancer Research 19:605-614; Niculescu-Duvaz and Springer (1997) Adv. Drg Del. Rev. 26:151-172; U.S. Patent No. 4,975,278) allows targeted delivery of the drug moiety to tumors, and intracellular accumulation therein, where systemic administration of these unconjugated drug agents may result in unacceptable levels of toxicity to normal cells as well as the tumor cells sought to be eliminated (Baldwin *et al.*, (1986) Lancet pp. (Mar. 15, 1986):603-05; Thorpe, (1985) "Antibody Carriers Of Cytotoxic Agents In Cancer Therapy: A Review," in Monoclonal Antibodies '84: Biological And Clinical Applications, A. Pinchera *et al.* (ed.s), pp. 475-506). Maximal efficacy with minimal toxicity is sought thereby. Both polyclonal antibodies and monoclonal antibodies have been reported as useful in these strategies (Rowland *et al.*, (1986) Cancer Immunol. Immunother., 21:183-87). Drugs used in these methods include

daunomycin, doxorubicin, methotrexate, and vindesine (Rowland *et al.*, (1986) *Supra*). Toxins used in antibody-toxin conjugates include bacterial toxins such as diphtheria toxin, plant toxins such as ricin, small molecule toxins such as geldanamycin (Mandler et al (2000) Jour. of the Nat. Cancer Inst. 92(19):1573-1581; Mandler et al (2000) Bioorganic & Med. Chem. Letters 10: 1025-1028; Mandler et al (2002) Bioconjugate Chem. 13:786-791), maytansinoids (EP 1391213; Liu *et al.*, (1996) Proc. Natl. Acad. Sci. USA 93:8618-8623), and calicheamicin (Lode et al (1998) Cancer Res. 58:2928; Hinman et al (1993) Cancer Res. 53:3336-3342). The toxins may effect their cytotoxic and cytostatic effects by mechanisms including tubulin binding, DNA binding, or topoisomerase inhibition. Some cytotoxic drugs tend to be inactive or less active when conjugated to large antibodies or protein receptor ligands.

[00421] ZEVALIN® (ibritumomab tiuxetan, Biogen/Idec) is an antibody-radioisotope conjugate composed of a murine IgG1 kappa monoclonal antibody directed against the CD20 antigen found on the surface of normal and malignant B lymphocytes and ^{111}In or ^{90}Y radioisotope bound by a thiourea linker-chelator (Wiseman et al (2000) Eur. Jour. Nucl. Med. 27(7):766-77; Wiseman et al (2002) Blood 99(12):4336-42; Witzig et al (2002) J. Clin. Oncol. 20(10):2453-63; Witzig et al (2002) J. Clin. Oncol. 20(15):3262-69). Although ZEVALIN has activity against B-cell non-Hodgkin's Lymphoma (NHL), administration results in severe and prolonged cytopenias in most patients. MYLOTARG™ (gemtuzumab ozogamicin, Wyeth Pharmaceuticals), an antibody drug conjugate composed of a hu CD33 antibody linked to calicheamicin, was approved in 2000 for the treatment of acute myeloid leukemia by injection (Drugs of the Future (2000) 25(7):686; U.S. Patent Nos. 4,970,198; 5,079,233; 5,585,089; 5,606,040; 5,693,762; 5,739,116; 5,767,285; 5,773,001). Cantuzumab mertansine (Immunogen, Inc.), an antibody drug conjugate composed of the huC242 antibody linked via the disulfide linker SPP to the maytansinoid drug moiety, DM1, is advancing into Phase II trials for the treatment of cancers that express CanAg, such as colon, pancreatic, gastric, and others. MLN-2704 (Millennium Pharm., BZL Biologics, Immunogen Inc.), an antibody drug conjugate composed of the anti-prostate specific membrane antigen (PSMA) monoclonal antibody linked to the maytansinoid drug moiety, DM1, is under development for the potential treatment of prostate tumors. The auristatin peptides, auristatin E (AE) and monomethylauristatin (MMAE), synthetic analogs of dolastatin, were conjugated to chimeric monoclonal antibodies cBR96 (specific to Lewis Y on carcinomas) and cAC10 (specific to CD30 on hematological malignancies) (Doronina et al (2003) Nature Biotechnology 21(7):778-784) and are under therapeutic development.

[00422] Chemotherapeutic agents useful in the generation of immunoconjugates are described herein. Enzymatically active toxins and fragments thereof that can be used

include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from Pseudomonas aeruginosa), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, Phytolaca americana proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. See, *e.g.,* WO 93/21232 published Oct. 28, 1993. A variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ^{212}Bi, ^{131}I, ^{131}In, ^{90}Y, and ^{186}Re. Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCl), active esters (such as disuccinimidyl suberate), aldehydes (such as glutaraldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as toluene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin may be prepared as described in Vitetta *et al.*, Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

[00423] Conjugates of an antibody and one or more small molecule toxins, such as a calicheamicin, maytansinoids, dolastatins, aurostatins, a trichothecene, and CC1065, and the derivatives of these toxins that have toxin activity, are also contemplated herein.

a. Maytansine and Maytansinoids

[00424] In some embodiments, the immunoconjugate comprises an antibody (full length or fragments) comprising an ultralong CDR3 as disclosed herein conjugated to one or more maytansinoid molecules.

[00425] Maytansinoids are mitototic inhibitors which act by inhibiting tubulin polymerization. Maytansine was first isolated from the east African shrub Maytenus serrata (U.S. Patent No. 3,896,111). Subsequently, it was discovered that certain microbes also produce maytansinoids, such as maytansinol and C-3 maytansinol esters (U.S. Patent No. 4,151,042). Synthetic maytansinol and derivatives and analogues thereof are disclosed, for example, in U.S. Patent Nos. 4,137,230; 4,248,870; 4,256,746; 4,260,608; 4,265,814; 4,294,757; 4,307,016; 4,308,268; 4,308,269; 4,309,428; 4,313,946; 4,315,929; 4,317,821; 4,322,348; 4,331,598; 4,361,650; 4,364,866; 4,424,219; 4,450,254; 4,362,663; and 4,371,533.

[00426] Maytansinoid drug moieties are attractive drug moieties in antibody drug conjugates because they are: (i) relatively accessible to prepare by fermentation or chemical modification, derivatization of fermentation products, (ii) amenable to derivatization with functional groups suitable for conjugation through the non-disulfide linkers to antibodies, (iii) stable in plasma, and (iv) effective against a variety of tumor cell lines.

[00427] Immunoconjugates containing maytansinoids, methods of making same, and their therapeutic use are disclosed, for example, in U.S. Patent Nos. 5,208,020, 5,416,064 and EP 0 425 235. Liu *et al.*, Proc. Natl. Acad. Sci. USA 93:8618-8623 (1996) described immunoconjugates comprising a maytansinoid designated DM1 linked to the monoclonal antibody C242 directed against human colorectal cancer. The conjugate was found to be highly cytotoxic towards cultured colon cancer cells, and showed antitumor activity in an *in vivo* tumor growth assay. Chari *et al.*, Cancer Research 52:127-131 (1992) describe immunoconjugates in which a maytansinoid was conjugated via a disulfide linker to the murine antibody A7 binding to an antigen on human colon cancer cell lines, or to another murine monoclonal antibody TA.1 that binds the HER-2/neu oncogene. The cytotoxicity of the TA.1-maytansinoid conjugate was tested *in vitro* on the human breast cancer cell line SK-BR-3, which expresses $3x10^5$ HER-2 surface antigens per cell. The drug conjugate achieved a degree of cytotoxicity similar to the free maytansinoid drug, which could be increased by increasing the number of maytansinoid molecules per antibody molecule. The A7-maytansinoid conjugate showed low systemic cytotoxicity in mice.

[00428] Antibody-maytansinoid conjugates are prepared by chemically linking an antibody to a maytansinoid molecule without significantly diminishing the biological activity of either the antibody or the maytansinoid molecule. See, *e.g.,* U.S. Patent No. 5,208,020. An average of 3-4 maytansinoid molecules conjugated per antibody molecule has shown efficacy in enhancing cytotoxicity of target cells without negatively affecting the function or solubility of the antibody, although even one molecule of toxin/antibody would be expected to enhance cytotoxicity over the use of naked antibody. Maytansinoids are well known in the art and can be synthesized by known techniques or isolated from natural sources. Suitable maytansinoids are disclosed, for example, in U.S. Patent No. 5,208,020 and in the other patents and nonpatent publications referred to hereinabove. Preferred maytansinoids are maytansinol and maytansinol analogues modified in the aromatic ring or at other positions of the maytansinol molecule, such as various maytansinol esters.

[00429] There are many linking groups known in the art for making antibody-maytansinoid conjugates, including, for example, those disclosed in U.S. Patent Nos.

5,208,020, 6,441,163, or EP Patent 0 425 235, Chari *et al.*, Cancer Research 52:127-131 (1992). Antibody-maytansinoid conjugates comprising the linker component SMCC may be prepared. The linking groups include disulfide groups, thioether groups, acid labile groups, photolabile groups, peptidase labile groups, or esterase labile groups, as disclosed in the above-identified patents, disulfide and thioether groups being preferred. Additional linking groups are described and exemplified herein.

[00430] Conjugates of the antibody and maytansinoid may be made using a variety of bifunctional protein coupling agents such as N-succinimidyl-3-(2-pyridyldithio) propionate (SPDP), succinimidyl-4-(N-maleimidomethyl)cyclohexane-1-carboxylate (SMCC), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCl), active esters (such as disuccinimidyl suberate), aldehydes (such as glutaraldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as toluene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). Particularly preferred coupling agents include N-succinimidyl-3-(2-pyridyldithio) propionate (SPDP) (Carlsson *et al.*, Biochem. J. 173:723-737 (1978)) and N-succinimidyl-4-(2-pyridylthio)pentanoate (SPP) to provide for a disulfide linkage.

[00431] The linker may be attached to the maytansinoid molecule at various positions, depending on the type of the link. For example, an ester linkage may be formed by reaction with a hydroxyl group using conventional coupling techniques. The reaction may occur at the C-3 position having a hydroxyl group, the C-14 position modified with hydroxymethyl, the C-15 position modified with a hydroxyl group, and the C-20 position having a hydroxyl group. In a preferred embodiment, the linkage is formed at the C-3 position of maytansinol or a maytansinol analogue.

b. Auristatins and Dolastatins

[00432] In some embodiments, the immunoconjugate comprises an antibody disclosed herein conjugated to dolastatins or dolostatin peptidic analogs and derivatives, the auristatins (U.S. Patent Nos. 5,635,483; 5,780,588). Dolastatins and auristatins have been shown to interfere with microtubule dynamics, GTP hydrolysis, and nuclear and cellular division (Woyke et al (2001) Antimicrob. Agents and Chemother. 45(12):3580-3584) and have anticancer (U.S. Patent No. 5,663,149) and antifungal activity (Pettit et al (1998) Antimicrob. Agents Chemother. 42:2961-2965). The dolastatin or auristatin drug moiety may be attached to the antibody through the N (amino) terminus or the C (carboxyl) terminus of the peptidic drug moiety (WO 02/088172).

[00433] Exemplary auristatin embodiments include the N-terminus linked monomethylauristatin drug moieties DE and DF, (see, *e.g.*, U.S. Patent No. 7,498,298).

[00434] Typically, peptide-based drug moieties can be prepared by forming a peptide bond between two or more amino acids and/or peptide fragments. Such peptide bonds can be prepared, for example, according to the liquid phase synthesis method (see, *e.g.*, E. Schroder and K. Lubke, "The Peptides", volume 1, pp 76-136, 1965, Academic Press) that is well known in the field of peptide chemistry. The auristatin/dolastatin drug moieties may be prepared according to the methods of: U.S. Patent No. 5,635,483; U.S. Patent No. 5,780,588; Pettit et al (1989) J. Am. Chem. Soc. 111:5463-5465; Pettit *et al.* (1998) Anti-Cancer Drug Design 13:243-277; Pettit, G. R., *et al.* Synthesis, 1996, 719-725; and Pettit *et al.* (1996) J. Chem. Soc. Perkin Trans. 1 5:859-863. See also Doronina (2003) Nat Biotechnol 21(7):778-784; U.S. Patent No. 7,498,289, (disclosing, linkers and methods of preparing monomethylvaline compounds such as MMAE and MMAF conjugated to linkers).

c. Calicheamicin

[00435] In other embodiments, the immunoconjugate comprises an antibody disclosed herein conjugated to one or more calicheamicin molecules. The calicheamicin family of antibiotics are capable of producing double-stranded DNA breaks at sub-picomolar concentrations. For the preparation of conjugates of the calicheamicin family, see U.S. Patent Nos. 5,712,374, 5,714,586, 5,739,116, 5,767,285, 5,770,701, 5,770,710, 5,773,001, 5,877,296. Structural analogues of calicheamicin which may be used include, but are not limited to, γ_1^I, α_2^I, α_3^I, N-acetyl-γ_1^I, PSAG and θ^I_1 (see, *e.g.*, Hinman *et al.*, Cancer Research 53:3336-3342 (1993), Lode *et al.*, Cancer Research 58:2925-2928 (1998) and the aforementioned U.S. patents). Another anti-tumor drug that the antibody can be conjugated is QFA which is an antifolate. Both calicheamicin and QFA have intracellular sites of action and do not readily cross the plasma membrane. Therefore, cellular uptake of these agents through antibody mediated internalization greatly enhances their cytotoxic effects.

d. Other Cytotoxic Agents

[00436] Other antitumor agents that can be conjugated to the antibodies disclosed herein include BCNU, streptozoicin, vincristine and 5-fluorouracil, the family of agents known collectively LL-E33288 complex described in U.S. Patent Nos. 5,053,394, 5,770,710, as well as esperamicins (U.S. Patent No. 5,877,296).

[00437] Enzymatically active toxins and fragments thereof which can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from Pseudomonas aeruginosa), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, Aleurites fordii proteins, dianthin proteins, Phytolaca americana proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria

officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin and the tricothecenes. See, for example, WO 93/21232 published Oct. 28, 1993.

[00438] The present disclosure further contemplates an immunoconjugate formed between an antibody and a compound with nucleolytic activity (*e.g.,* a ribonuclease or a DNA endonuclease such as a deoxyribonuclease; DNase).

[00439] For selective destruction of the tumor, the antibody may comprise a highly radioactive atom. A variety of radioactive isotopes are available for the production of radioconjugated antibodies. Examples include At^{211}, I^{131}, I^{125}, Y^{90}, Re^{186}, Re^{188}, Sm^{153}, Bi^{212}, P^{32}, Pb^{212} and radioactive isotopes of Lu. When the conjugate is used for detection, it may comprise a radioactive atom for scintigraphic studies, for example tc^{99m} or I^{123}, or a spin label for nuclear magnetic resonance (NMR) imaging (also known as magnetic resonance imaging, mri), such as iodine-123 again, iodine-131, indium-111, fluorine-19, carbon-13, nitrogen-15, oxygen-17, gadolinium, manganese or iron.

[00440] The radiolabels or other labels may be incorporated in the conjugate in known ways. For example, the peptide may be biosynthesized or may be synthesized by chemical amino acid synthesis using suitable amino acid precursors involving, for example, fluorine-19 in place of hydrogen. Labels such as tc^{99m} or I^{123}, Re^{186}, Re^{188} and In^{111} can be attached via a cysteine residue in the peptide. Yttrium-90 can be attached via a lysine residue. The IODOGEN method (Fraker et al (1978) Biochem. Biophys. Res. Commun. 80: 49-57) can be used to incorporate iodine-123. "Monoclonal Antibodies in Immunoscintigraphy" (Chatal, CRC Press 1989) describes other methods.

[00441] Conjugates of the antibody and cytotoxic agent may be made using a variety of bifunctional protein coupling agents such as N-succinimidyl-3-(2-pyridyldithio) propionate (SPDP), succinimidyl-4-(N-maleimidomethyl)cyclohexane-1-carboxylate (SMCC), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl adipimidate HCl), active esters (such as disuccinimidyl suberate), aldehydes (such as glutaraldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as toluene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as described in Vitetta *et al.*, Science 238:1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026. The linker may be a "cleavable linker" facilitating release of the cytotoxic drug in the cell. For example, an acid-labile linker, peptidase-sensitive linker, photolabile

linker, dimethyl linker or disulfide-containing linker (Chari *et al.*, Cancer Research 52:127-131 (1992); U.S. Patent No. 5,208,020) may be used.

[00442] The compounds disclosed herein expressly contemplate, but are not limited to, ADC prepared with cross-linker reagents: BMPS, EMCS, GMBS, HBVS, LC-SMCC, MBS, MPBH, SBAP, SIA, SIAB, SMCC, SMPB, SMPH, sulfo-EMCS, sulfo-GMBS, sulfo-KMUS, sulfo-MBS, sulfo-SIAB, sulfo-SMCC, and sulfo-SMPB, and SVSB (succinimidyl-(4-vinylsulfone)benzoate) which are commercially available (*e.g.,* from Pierce Biotechnology, Inc., Rockford, Ill., U.S.A). See pages 467-498, 2003-2004 Applications Handbook and Catalog.

e. Preparation of Antibody Drug Conjugates

[00443] In the antibody drug conjugates (ADC) disclosed herein, an antibody (Ab) is conjugated to one or more drug moieties (D), *e.g,.* about 1 to about 20 drug moieties per antibody, through a linker (L). An ADC of Formula I [Ab-(L-D)$_p$] may be prepared by several routes, employing organic chemistry reactions, conditions, and reagents known to those skilled in the art, including: (1) reaction of a nucleophilic group of an antibody with a bivalent linker reagent, to form Ab-L, via a covalent bond, followed by reaction with a drug moiety D; and (2) reaction of a nucleophilic group of a drug moiety with a bivalent linker reagent, to form D-L, via a covalent bond, followed by reaction with the nucleophilic group of an antibody. Additional methods for preparing ADC are described herein.

[00444] The linker may be composed of one or more linker components. Exemplary linker components include 6-maleimidocaproyl ("MC"), maleimidopropanoyl ("MP"), valine-citrulline ("val-cit"), alanine-phenylalanine ("ala-phe"), p-aminobenzyloxycarbonyl ("PAB"), N-Succinimidyl 4-(2-pyridylthio) pentanoate ("SPP"), N-Succinimidyl 4-(N-maleimidomethyl)cyclohexane-1 carboxylate ("SMCC"), and N-Succinimidyl (4-iodo-acetyl)aminobenzoate ("SIAB"). Additional linker components are known in the art and some are disclosed herein (see, *e.g.,* U.S. Patent No. 7,498,298).

[00445] In some embodiments, the linker may comprise amino acid residues. Exemplary amino acid linker components include a dipeptide, a tripeptide, a tetrapeptide or a pentapeptide. Exemplary dipeptides include: valine-citrulline (vc or val-cit), alanine-phenylalanine (af or ala-phe). Exemplary tripeptides include: glycine-valine-citrulline (gly-val-cit) and glycine-glycine-glycine (gly-gly-gly). Amino acid residues which comprise an amino acid linker component include those occurring naturally, as well as minor amino acids and non-naturally occurring amino acid analogs, such as citrulline. Amino acid linker components can be designed and optimized in their selectivity for enzymatic cleavage by a particular enzymes, for example, a tumor-associated protease, cathepsin B, C and D, or a plasmin protease.

[00446] Nucleophilic groups on antibodies include, but are not limited to: (i) N-terminal amine groups, (ii) side chain amine groups, *e.g.,* lysine, (iii) side chain thiol groups, *e.g.*, cysteine, and (iv) sugar hydroxyl or amino groups where the antibody is glycosylated. Amine, thiol, and hydroxyl groups are nucleophilic and capable of reacting to form covalent bonds with electrophilic groups on linker moieties and linker reagents including: (i) active esters such as NHS esters, HOBt esters, haloformates, and acid halides; (ii) alkyl and benzyl halides such as haloacetamides; (iii) aldehydes, ketones, carboxyl, and maleimide groups. Certain antibodies have reducible interchain disulfides, *e.g.,* cysteine bridges. Antibodies may be made reactive for conjugation with linker reagents by treatment with a reducing agent such as DTT (dithiothreitol). Each cysteine bridge will thus form, theoretically, two reactive thiol nucleophiles. Additional nucleophilic groups can be introduced into antibodies through the reaction of lysines with 2-iminothiolane (Traut's reagent) resulting in conversion of an amine into a thiol. Reactive thiol groups may be introduced into the antibody (or fragment thereof) by introducing one, two, three, four, or more cysteine residues (*e.g.,* preparing mutant antibodies comprising one or more non-native cysteine amino acid residues).

[00447] Antibody drug conjugates disclosed herein may also be produced by modification of the antibody to introduce electrophilic moieties, which can react with nucleophilic substituents on the linker reagent or drug. The sugars of glycosylated antibodies may be oxidized, *e.g.,* with periodate oxidizing reagents, to form aldehyde or ketone groups which may react with the amine group of linker reagents or drug moieties. The resulting imine Schiff base groups may form a stable linkage, or may be reduced, *e.g.,* by borohydride reagents to form stable amine linkages. In one embodiment, reaction of the carbohydrate portion of a glycosylated antibody with either glactose oxidase or sodium meta-periodate may yield carbonyl (aldehyde and ketone) groups in the protein that can react with appropriate groups on the drug (Hermanson, Bioconjugate Techniques). In another embodiment, proteins containing N-terminal serine or threonine residues can react with sodium meta-periodate, resulting in production of an aldehyde in place of the first amino acid (Geoghegan & Stroh, (1992) Bioconjugate Chem. 3:138-146; U.S. Patent No. 5,362,852). Such aldehyde can be reacted with a drug moiety or linker nucleophile.

[00448] Likewise, nucleophilic groups on a drug moiety include, but are not limited to: amine, thiol, hydroxyl, hydrazide, oxime, hydrazine, thiosemicarbazone, hydrazine carboxylate, and arylhydrazide groups capable of reacting to form covalent bonds with electrophilic groups on linker moieties and linker reagents including: (i) active esters such as NHS esters, HOBt esters, haloformates, and acid halides; (ii) alkyl and benzyl halides such as haloacetamides; (iii) aldehydes, ketones, carboxyl, and maleimide groups.

[00449] Alternatively, a fusion protein comprising the antibody and cytotoxic agent may be made, *e.g.,* by recombinant techniques or peptide synthesis. The length of DNA may comprise respective regions encoding the two portions of the conjugate either adjacent one another or separated by a region encoding a linker peptide which does not destroy the desired properties of the conjugate.

[00450] In yet another embodiment, the antibody may be conjugated to a "receptor" (such streptavidin) for utilization in tumor pre-targeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (*e.g.,* avidin) which is conjugated to a cytotoxic agent (*e.g.,* a radionucleotide).

Engineered Hybridomas

[00451] Hybridoma cells can be generated by fusing B cells producing a desired antibody with an immortalized cell line, usually a myeloma cell line, so that the resulting fusion cells will be an immortalized cell line that secrets a particular antibody. By the same principle, myeloma cells can be first transfected with a nucleic acid encoding a germline antibody V region and can be screened for the expression of the germline V region. Those myeloma cells with highest level of proteolytic light chain expression can be subsequently fused with B cells that produce an antibody with desired target protein specificity. The fusion cells will produce two types of antibodies: one is a heterologous antibody containing an endogenous antibody chain (either heavy or light) operably joined to the recombinant germline V region (either heavy or light), and the other is the same antibody that the parental B cells would secrete (*e.g.* both endogenous heavy and light chains). The operably joined heterologous heavy and light chains can be isolated by conventional methods such as chromatography and identification can be confirmed by target protein binding assays, assays identifying a unique tag of the germline polypeptide, or endopeptidase activity assays described in other sections of this disclosure. In some cases, where the heterologous antibody is the predominant type in quantity among the two types of antibodies, such isolation may not be needed. Hybridomas. Including bovine hybridomas, may be a source of bovine antibody gene sequences, including ultralong CDR3 sequences.

Transgenic Mammals

[00452] A nucleic acid sequence encoding a germline antibody polypeptide of the present disclosure can be introduced into a non-human mammal to generate a transgenic animal that expresses the germline antibody polypeptide. Unlike the transgenic animal models more commonly seen, the transgene expressed by the transgenic mammals of the

present disclosure need not replace at least one allele of the endogenous coding sequence responsible for the variable regions of antibody chains following somatic recombination. Due to allelic exclusion, the presence of an exogenous, post-somatic rearrangement version of the germline V region DNA will inhibit the endogenous alleles of pre-somatic rearrangement V minigenes from undergoing somatic rearrangement and contributing to the makeup of antibody chains this mammal may produce. Thus, when exposed to a particular antigen, the mammal will generate heterologous antibodies comprising one endogenously rearranged antibody chain, and one transgenic gene which was rearranged a priori. Such heterologous antibodies are invaluable in research and in treating certain conditions in live subjects. On the other hand, a method that directs the integration of the transgene to the locus of an endogenous allele will fully serve the purpose of practicing the present disclosure as well.

[00453] The general methods of generating transgenic animals have been well established and frequently practiced. For reviews and protocols for generating transgenic animals and related methods for genetic manipulations, see, *e.g.*, Mansour *et al.*, Nature 336:348-352 (1988); Capecchi *et al.*, Trends Genet. 5:70-76 (1989); Capecchi, Science 244:1288-1292 (1989); Capecchi *et al.*, Current Communications in Molecular Biology, pp45-52, Capecchi, M.R. (ed.), Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1989); Frohman *et al.*, Cell 56: 145-147 (1989); Brinster *et al.*, Proc. Natl. Acad. Sci. USA 82:4438-4442 (1985); Evans et. al., Nature 292:154-156 (1981); Bradley *et al.*, Nature 309:255-258 (1984); Gossler *et al.*, Proc. Natl. Acad. Sci. USA 83:9065-9069 (1986); Robertson *et al.*, Nature 322:445-448 (1986); Jaenisch Science 240:1468-1474 (1988); and Siedel, G. E., Jr., "Critical review of embryo transfer procedures with cattle" in Fertilization and Embryonic Development in Vitro, page 323, L. Mastroianni, Jr. and J. D. Biggers, ed., Plenum Press, New York, N.Y. (1981).

[00454] An exemplary transgenic animal of the present disclosure is mouse, whereas a number of other transgenic animals can also be produced using the same general method. These animals include, but are not limited to: rabbits, sheep, cattle, and pigs (Jaenisch Science 240:1468-1474 (1988); Hammer *et al.*, J. Animal. Sci. 63:269 (1986); Hammer *et al.* Nature 315:680 (1985); Wagner *et al.*, Theriogenology 21:29 (1984)).

Pharmaceutical Compositions

[00455] Antibodies comprising an ultralong CDR3, antibody fragments, nucleic acids, or vectors disclosed herein can be formulated in compositions, especially pharmaceutical compositions. Such compositions with antibodies comprising an ultralong CDR3 comprise a therapeutically or prophylactically effective amount of a antibodies comprising an ultralong CDR3, antibody fragment, nucleic acid, or vector disclosed herein in admixture with a suitable carrier, *e.g.,* a pharmaceutically acceptable agent. Typically, antibodies comprising an ultralong CDR3, antibody fragments, nucleic acids, or vectors disclosed herein are sufficiently purified for administration before formulation in a pharmaceutical composition.

[00456] Pharmaceutically acceptable agents for use in the present pharmaceutical compositions include carriers, excipients, diluents, antioxidants, preservatives, coloring, flavoring and diluting agents, emulsifying agents, suspending agents, solvents, fillers, bulking agents, buffers, delivery vehicles, tonicity agents, cosolvents, wetting agents, complexing agents, buffering agents, antimicrobials, and surfactants.

[00457] Neutral buffered saline or saline mixed with serum albumin are exemplary appropriate carriers. The pharmaceutical compositions may include antioxidants such as ascorbic acid; low molecular weight polypeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids such as glycine, glutamine, asparagine, arginine or lysine; monosaccharides, disaccharides, and other carbohydrates including glucose, mannose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; salt-forming counterions such as sodium; and/or nonionic surfactants such as Tween, pluronics, or polyethylene glycol (PEG). Also by way of example, suitable tonicity enhancing agents include alkali metal halides (preferably sodium or potassium chloride), mannitol, sorbitol, and the like. Suitable preservatives include benzalkonium chloride, thimerosal, phenethyl alcohol, methylparaben, propylparaben, chlorhexidine, sorbic acid and the like. Hydrogen peroxide also may be used as preservative. Suitable cosolvents include glycerin, propylene glycol, and PEG. Suitable complexing agents include caffeine, polyvinylpyrrolidone, beta-cyclodextrin or hydroxy-propyl-beta-cyclodextrin. Suitable surfactants or wetting agents include sorbitan esters, polysorbates such as polysorbate 80, tromethamine, lecithin, cholesterol, tyloxapal, and the like. The buffers may be conventional buffers such as acetate, borate, citrate, phosphate, bicarbonate, or Tris-HCl. Acetate buffer may be about pH 4-5.5, and Tris buffer can be about pH 7-8.5. Additional pharmaceutical agents are set forth in Remington's Pharmaceutical Sciences, 18th Edition, A. R. Gennaro, ed., Mack Publishing Company, 1990.

[00458] The composition may be in liquid form or in a lyophilized or freeze-dried form and may include one or more lyoprotectants, excipients, surfactants, high molecular weight structural additives and/or bulking agents (see, for example, U.S. Patent Nos. 6,685,940, 6,566,329, and 6,372,716). In one embodiment, a lyoprotectant is included, which is a non-reducing sugar such as sucrose, lactose or trehalose. The amount of lyoprotectant generally included is such that, upon reconstitution, the resulting formulation will be isotonic, although hypertonic or slightly hypotonic formulations also may be suitable. In addition, the amount of lyoprotectant should be sufficient to prevent an unacceptable amount of degradation and/or aggregation of the protein upon lyophilization. Exemplary lyoprotectant concentrations for sugars (*e.g.,* sucrose, lactose, trehalose) in the pre-lyophilized formulation are from about 10 mM to about 400 mM. In another embodiment, a surfactant is included, such as for example, nonionic surfactants and ionic surfactants such as polysorbates (*e.g.,* polysorbate 20, polysorbate 80); poloxamers (*e.g.,* poloxamer 188); poly(ethylene glycol) phenyl ethers (*e.g.*, Triton); sodium dodecyl sulfate (SDS); sodium laurel sulfate; sodium octyl glycoside; lauryl-, myristyl-, linoleyl-, or stearyl-sulfobetaine; lauryl-, myristyl-, linoleyl- or stearyl-sarcosine; linoleyl, myristyl-, or cetyl-betaine; lauroamidopropyl-, cocamidopropyl-, linoleamidopropyl-, myristamidopropyl-, palmidopropyl-, or isostearamidopropyl-betaine (*e.g.,* lauroamidopropyl); myristamidopropyl-, palmidopropyl-, or isostearamidopropyl-dimethylamine; sodium methyl cocoyl-, or disodium methyl ofeyl-taurate; and the MONAQUAT™. series (Mona Industries, Inc., Paterson, N.J.), polyethyl glycol, polypropyl glycol, and copolymers of ethylene and propylene glycol (*e.g.,* Pluronics, PF68 etc). Exemplary amounts of surfactant that may be present in the pre-lyophilized formulation are from about 0.001-0.5%. High molecular weight structural additives (*e.g.,* fillers, binders) may include for example, acacia, albumin, alginic acid, calcium phosphate (dibasic), cellulose, carboxymethylcellulose, carboxymethylcellulose sodium, hydroxyethylcellulose, hydroxypropylcellulose, hydroxypropylmethylcellulose, microcrystalline cellulose, dextran, dextrin, dextrates, sucrose, tylose, pregelatinized starch, calcium sulfate, amylose, glycine, bentonite, maltose, sorbitol, ethylcellulose, disodium hydrogen phosphate, disodium phosphate, disodium pyrosulfite, polyvinyl alcohol, gelatin, glucose, guar gum, liquid glucose, compressible sugar, magnesium aluminum silicate, maltodextrin, polyethylene oxide, polymethacrylates, povidone, sodium alginate, tragacanth microcrystalline cellulose, starch, and zein. Exemplary concentrations of high molecular weight structural additives are from 0.1% to 10% by weight. In other embodiments, a bulking agent (*e.g.,* mannitol, glycine) may be included.

[00459] Compositions may be suitable for parenteral administration. Exemplary compositions are suitable for injection or infusion into an animal by any route available to the skilled worker, such as intraarticular, subcutaneous, intravenous, intramuscular, intraperitoneal, intracerebral (intraparenchymal), intracerebroventricular, intramuscular, intraocular, intraarterial, or intralesional routes. A parenteral formulation typically will be a sterile, pyrogen-free, isotonic aqueous solution, optionally containing pharmaceutically acceptable preservatives.

[00460] Examples of non-aqueous solvents are propylene glycol, polyethylene glycol, vegetable oils such as olive oil, and injectable organic esters such as ethyl oleate. Aqueous carriers include water, alcoholic/aqueous solutions, emulsions or suspensions, including saline and buffered media. Parenteral vehicles include sodium chloride solution, Ringers' dextrose, dextrose and sodium chloride, lactated Ringer's, or fixed oils. Intravenous vehicles include fluid and nutrient replenishers, electrolyte replenishers, such as those based on Ringer's dextrose, and the like. Preservatives and other additives may also be present, such as, for example, anti-microbials, anti-oxidants, chelating agents, inert gases and the like. See generally, Remington's Pharmaceutical Science, 16th Ed., Mack Eds., 1980.

[00461] Pharmaceutical compositions described herein may be formulated for controlled or sustained delivery in a manner that provides local concentration of the product (*e.g.,* bolus, depot effect) and/or increased stability or half-life in a particular local environment. The compositions can include the formulation of antibodies comprising an ultralong CDR3, antibody fragments, nucleic acids, or vectors disclosed herein with particulate preparations of polymeric compounds such as polylactic acid, polyglycolic acid, *etc.*, as well as agents such as a biodegradable matrix, injectable microspheres, microcapsular particles, microcapsules, bioerodible particles beads, liposomes, and implantable delivery devices that provide for the controlled or sustained release of the active agent which then can be delivered as a depot injection. Techniques for formulating such sustained- or controlled-delivery means are known and a variety of polymers have been developed and used for the controlled release and delivery of drugs. Such polymers are typically biodegradable and biocompatible. Polymer hydrogels, including those formed by complexation of enantiomeric polymer or polypeptide segments, and hydrogels with temperature or pH sensitive properties, may be desirable for providing drug depot effect because of the mild and aqueous conditions involved in trapping bioactive protein agents (*e.g.,* antibodies comprising an ultralong CDR3). See, for example, the description of controlled release porous polymeric microparticles for the delivery of pharmaceutical compositions in WO 93/15722.

[00462] Suitable materials for this purpose include polylactides (see, *e.g.,* U.S. Patent No. 3,773,919), polymers of poly-(a-hydroxycarboxylic acids), such as poly-D-(-)-3-hydroxybutyric acid (EP 133,988A), copolymers of L-glutamic acid and gamma ethyl-L-glutamate (Sidman *et al.*, Biopolymers, 22: 547-556 (1983)), poly(2-hydroxyethyl-methacrylate) (Langer *et al.*, J. Biomed. Mater. Res., 15: 167-277 (1981), and Langer, Chem. Tech., 12: 98-105 (1982)), ethylene vinyl acetate, or poly-D(-)-3-hydroxybutyric acid. Other biodegradable polymers include poly(lactones), poly(acetals), poly(orthoesters), and poly(orthocarbonates). Sustained-release compositions also may include liposomes, which can be prepared by any of several methods known in the art (see, *e.g.,* Eppstein *et al.*, Proc. Natl. Acad. Sci. USA, 82: 3688-92 (1985)). The carrier itself, or its degradation products, should be nontoxic in the target tissue and should not further aggravate the condition. This can be determined by routine screening in animal models of the target disorder or, if such models are unavailable, in normal animals.

[00463] Microencapsulation of recombinant proteins for sustained release has been performed successfully with human growth hormone (rhGH), interferon-(rhIFN-), interleukin-2, and MN rgp120. Johnson *et al.*, Nat. Med., 2:795-799 (1996); Yasuda, Biomed. Ther., 27:1221-1223 (1993); Hora *et al.*, Bio/Technology. 8:755-758 (1990); Cleland, "Design and Production of Single Immunization Vaccines Using Polylactide Polyglycolide Microsphere Systems," in Vaccine Design: The Subunit and Adjuvant Approach, Powell and Newman, eds, (Plenum Press: New York, 1995), pp. 439-462; WO 97/03692, WO 96/40072, WO 96/07399; and U.S. Patent No. 5,654,010. The sustained-release formulations of these proteins were developed using poly-lactic-coglycolic acid (PLGA) polymer due to its biocompatibility and wide range of biodegradable properties. The degradation products of PLGA, lactic and glycolic acids can be cleared quickly within the human body. Moreover, the degradability of this polymer can be depending on its molecular weight and composition. Lewis, "Controlled release of bioactive agents from lactide/glycolide polymer," in: M. Chasin and R. Langer (Eds.), Biodegradable Polymers as Drug Delivery Systems (Marcel Dekker: New York, 1990), pp. 1-41. Additional examples of sustained release compositions include, for example, EP 58,481A, U.S. Patent No. 3,887,699, EP 158,277A, Canadian Patent No. 1176565, U. Sidman *et al.*, Biopolymers 22, 547 [1983], R. Langer *et al.*, Chem. Tech. 12, 98 [1982], Sinha *et al.*, J. Control. Release 90, 261 [2003], Zhu *et al.*, Nat. Biotechnol. 18, 24 [2000], and Dai *et al.*, Colloids Surf B Biointerfaces 41, 117 [2005].

[00464] Bioadhesive polymers are also contemplated for use in or with compositions of the present disclosure. Bioadhesives are synthetic and naturally occurring materials able to adhere to biological substrates for extended time periods. For example,

Carbopol and polycarbophil are both synthetic cross-linked derivatives of poly(acrylic acid). Bioadhesive delivery systems based on naturally occurring substances include for example hyaluronic acid, also known as hyaluronan. Hyaluronic acid is a naturally occurring mucopolysaccharide consisting of residues of D-glucuronic and N-acetyl-D-glucosamine. Hyaluronic acid is found in the extracellular tissue matrix of vertebrates, including in connective tissues, as well as in synovial fluid and in the vitreous and aqueous humor of the eye. Esterified derivatives of hyaluronic acid have been used to produce microspheres for use in delivery that are biocompatible and biodegradable (see, for example, Cortivo *et al.*, Biomaterials (1991) 12:727-730; EP 517,565; WO 96/29998; Illum *et al.*, J. Controlled Rel. (1994) 29:133-141). Exemplary hyaluronic acid containing compositions of the present disclosure comprise a hyaluronic acid ester polymer in an amount of approximately 0.1% to about 40% (w/w) of an antibody comprising an ultralong CDR3 to hyaluronic acid polymer.

[00465] Both biodegradable and non-biodegradable polymeric matrices may be used to deliver compositions of the present disclosure, and such polymeric matrices may comprise natural or synthetic polymers. Biodegradable matrices are preferred. The period of time over which release occurs is based on selection of the polymer. Typically, release over a period ranging from between a few hours and three to twelve months is most desirable. Exemplary synthetic polymers which may be used to form the biodegradable delivery system include: polymers of lactic acid and glycolic acid, polyamides, polycarbonates, polyalkylenes, polyalkylene glycols, polyalkylene oxides, polyalkylene terepthalates, polyvinyl alcohols, polyvinyl ethers, polyvinyl esters, poly-vinyl halides, polyvinylpyrrolidone, polyglycolides, polysiloxanes, polyanhydrides, polyurethanes and co-polymers thereof, poly(butic acid), poly(valeric acid), alkyl cellulose, hydroxyalkyl celluloses, cellulose ethers, cellulose esters, nitro celluloses, polymers of acrylic and methacrylic esters, methyl cellulose, ethyl cellulose, hydroxypropyl cellulose, hydroxypropyl methyl cellulose, hydroxybutyl methyl cellulose, cellulose acetate, cellulose propionate, cellulose acetate butyrate, cellulose acetate phthalate, carboxylethyl cellulose, cellulose triacetate, cellulose sulphate sodium salt, poly(methyl methacrylate), poly(ethyl methacrylate), poly(butylmethacrylate), poly(isobutyl methacrylate), poly(hexylmethacrylate), poly(isodecyl methacrylate), poly(lauryl methacrylate), poly(phenyl methacrylate), poly(methyl acrylate), poly(isopropyl acrylate), poly(isobutyl acrylate), poly(octadecyl acrylate), polyethylene, polypropylene, poly(ethylene glycol), poly(ethylene oxide), poly(ethylene terephthalate), poly(vinyl alcohols), polyvinyl acetate, poly vinyl chloride, polystyrene and polyvinylpyrrolidone. Exemplary natural polymers include alginate and other polysaccharides including dextran and cellulose, collagen, chemical derivatives thereof (substitutions, additions of chemical groups, for example, alkyl, alkylene,

hydroxylations, oxidations, and other modifications routinely made by those skilled in the art), albumin and other hydrophilic proteins, zein and other prolamines and hydrophobic proteins, copolymers and mixtures thereof. In general, these materials degrade either by enzymatic hydrolysis or exposure to water *in vivo*, by surface or bulk erosion. The polymer optionally is in the form of a hydrogel (see, for example, WO 04/009664, WO 05/087201, Sawhney, *et al.*, Macromolecules, 1993, 26, 581-587) that can absorb up to about 90% of its weight in water and further, optionally is cross-linked with multi-valent ions or other polymers.

[00466] Delivery systems also include non-polymer systems that are lipids including sterols such as cholesterol, cholesterol esters and fatty acids or neutral fats such as mono-di- and tri-glycerides; hydrogel release systems; silastic systems; peptide based systems; wax coatings; compressed tablets using conventional binders and excipients; partially fused implants; and the like. Specific examples include, but are not limited to: (a) erosional systems in which the product is contained in a form within a matrix such as those described in U.S. Patent Nos. 4,452,775, 4,675,189 and 5,736,152 and (b) diffusional systems in which a product permeates at a controlled rate from a polymer such as described in U.S. Patent Nos. 3,854,480, 5,133,974 and 5,407,686. Liposomes containing the product may be prepared by methods known methods, such as for example (DE 3,218,121; Epstein *et al.,* Proc. Natl. Acad. Sci. USA, 82: 3688-3692 (1985); Hwang *et al.,* Proc. Natl. Acad. Sci. USA, 77: 4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; JP 83-118008; U.S. Patent Nos. 4,485,045 and 4,544,545; and EP 102,324).

[00467] Alternatively or additionally, the compositions may be administered locally via implantation into the affected area of a membrane, sponge, or other appropriate material on to which a antibody comprising an ultralong CDR3, antibody fragment, nucleic acid, or vector disclosed herein has been absorbed or encapsulated. Where an implantation device is used, the device may be implanted into any suitable tissue or organ, and delivery of an antibody comprising an ultralong CDR3 antibody fragment, nucleic acid, or vector disclosed herein can be directly through the device via bolus, or via continuous administration, or via catheter using continuous infusion.

[00468] A pharmaceutical composition comprising an antibody comprising an ultralong CDR3, antibody fragment, nucleic acid, or vector disclosed herein may be formulated for inhalation, such as for example, as a dry powder. Inhalation solutions also may be formulated in a liquefied propellant for aerosol delivery. In yet another formulation, solutions may be nebulized. Additional pharmaceutical composition for pulmonary administration include, those described, for example, in WO 94/20069, which discloses

pulmonary delivery of chemically modified proteins. For pulmonary delivery, the particle size should be suitable for delivery to the distal lung. For example, the particle size may be from 1 μm to 5 μm; however, larger particles may be used, for example, if each particle is fairly porous.

[00469] Certain formulations containing antibodies comprising an ultralong CDR3, antibody fragments, nucleic acids, or vectors disclosed herein may be administered orally. Formulations administered in this fashion may be formulated with or without those carriers customarily used in the compounding of solid dosage forms such as tablets and capsules. For example, a capsule can be designed to release the active portion of the formulation at the point in the gastrointestinal tract when bioavailability is maximized and pre-systemic degradation is minimized. Additional agents may be included to facilitate absorption of a selective binding agent. Diluents, flavorings, low melting point waxes, vegetable oils, lubricants, suspending agents, tablet disintegrating agents, and binders also can be employed.

[00470] Another preparation may involve an effective quantity of an antibody comprising an ultralong CDR3, antibody fragment, nucleic acid, or vector disclosed herein in a mixture with non-toxic excipients which are suitable for the manufacture of tablets. By dissolving the tablets in sterile water, or another appropriate vehicle, solutions may be prepared in unit dose form. Suitable excipients include, but are not limited to, inert diluents, such as calcium carbonate, sodium carbonate or bicarbonate, lactose, or calcium phosphate; or binding agents, such as starch, gelatin, or acacia; or lubricating agents such as magnesium stearate, stearic acid, or talc.

[00471] Suitable and/or preferred pharmaceutical formulations may be determined in view of the present disclosure and general knowledge of formulation technology, depending upon the intended route of administration, delivery format, and desired dosage. Regardless of the manner of administration, an effective dose may be calculated according to patient body weight, body surface area, or organ size. Further refinement of the calculations for determining the appropriate dosage for treatment involving each of the formulations described herein are routinely made in the art and is within the ambit of tasks routinely performed in the art. Appropriate dosages may be ascertained through use of appropriate dose-response data.

[00472] In some embodiments, antibodies comprising an ultralong CDR3 or fragments thereof are provided with a modified Fc region where a naturally-occurring Fc region is modified to increase the half-life of the antibody or fragment in a biological environment, for example, the serum half-life or a half-life measured by an *in vitro* assay.

Methods for altering the original form of a Fc region of an IgG also are described in U.S. Patent No. 6,998,253.

[00473] In certain embodiments, it may be desirable to modify the antibody or fragment in order to increase its serum half-life, for example, adding molecules such as PEG or other water soluble polymers, including polysaccharide polymers, to antibody fragments to increase the half-life. This may also be achieved, for example, by incorporation of a salvage receptor binding epitope into the antibody fragment (*e.g.,* by mutation of the appropriate region in the antibody fragment or by incorporating the epitope into a peptide tag that is then fused to the antibody fragment at either end or in the middle, *e.g.,* by DNA or peptide synthesis) (see, International Publication No. WO96/32478). Salvage receptor binding epitope refers to an epitope of the Fc region of an IgG molecule (*e.g.,* IgG1, IgG2, IgG3, or IgG4) that is responsible for increasing the *in vivo* serum half-life of the IgG molecule.

[00474] A salvage receptor binding epitope may include a region wherein any one or more amino acid residues from one or two loops of a Fc domain are transferred to an analogous position of the antibody fragment. Even more preferably, three or more residues from one or two loops of the Fc domain are transferred. Still more preferred, the epitope is taken from the CH2 domain of the Fc region (*e.g.,* of an IgG) and transferred to the CH1, CH3, or VH region, or more than one such region, of the antibody. Alternatively, the epitope is taken from the CH2 domain of the Fc region and transferred to the C_L region or V_L region, or both, of the antibody fragment. See also WO 97/34631 and WO 96/32478 which describe Fc variants and their interaction with the salvage receptor.

[00475] Mutation of residues within Fc receptor binding sites may result in altered effector function, such as altered ADCC or CDC activity, or altered half-life. Potential mutations include insertion, deletion or substitution of one or more residues, including substitution with alanine, a conservative substitution, a non-conservative substitution, or replacement with a corresponding amino acid residue at the same position from a different IgG subclass (*e.g.*, replacing an IgG1 residue with a corresponding IgG2 residue at that position). For example, it has been reported that mutating the serine at amino acid position 241 in IgG4 to proline (found at that position in IgG1 and IgG2) led to the production of a homogeneous antibody, as well as extending serum half-life and improving tissue distribution compared to the original chimeric IgG4. (Angal *et al.*, Mol. Immunol. 30:105-8, 1993).

Kits/Articles of Manufacture

[00476] As an additional aspect, the present disclosure includes kits which comprise one or more compounds or compositions packaged in a manner which facilitates their use to practice methods of the present disclosure. In one embodiment, such a kit includes a compound or composition described herein (*e.g.*, a composition comprising an antibody comprising an ultralong CDR3 alone or in combination with a second agent), packaged in a container with a label affixed to the container or a package insert that describes use of the compound or composition in practicing the method. Suitable containers include, for example, bottles, vials, syringes, *etc.* The containers may be formed from a variety of materials such as glass or plastic. The container may have a sterile access port (for example the container may be an intravenous solution bag or a vial having a stopper pierceable by a hypodermic injection needle). The article of manufacture may comprise (a) a first container with a composition contained therein, wherein the composition comprises an antibody comprising an ultralong CDR3 as disclosed herein; and (b) a second container with a composition contained therein, wherein the composition comprises a further therapeutic agent. The article of manufacture in this embodiment disclosed herein may further comprise a package insert indicating that the first and second compositions can be used to treat a particular condition. Alternatively, or additionally, the article of manufacture may further comprise a second (or third) container comprising a pharmaceutically-acceptable buffer, such as bacteriostatic water for injection (BWFI), phosphate-buffered saline, Ringer's solution and dextrose solution. It may further include other materials desirable from a commercial and user standpoint, including other buffers, diluents, filters, needles, and syringes. Preferably, the compound or composition is packaged in a unit dosage form. The kit may further include a device suitable for administering the composition according to a specific route of administration or for practicing a screening assay. Preferably, the kit contains a label that describes use of the antibody comprising an ultralong CDR3 composition.

[00477] The following are examples of the methods and compositions of the disclosure. It is understood that various other embodiments may be practiced, given the general description provided above.

EXAMPLES

Example 1. Engineering non-human sequences into long CDR3's of human antibodies.

[00478] Human antibodies have heavy chain CDR3 regions that typically vary in size from 8-16 residues. Human antibodies with longer heavy chain CDR3 regions have been discovered with HIV-1 neutralization properties. McLellan *et al.*, Nature, 480:336-443 (2011); Walker *et al.*, Nature, 477:466-470 (2011). Crystal structures of at least 5 such

human antibodies (PGT145, PG9, PG16, CH04 and 2909, the VH-CH1 regions of which are provided in Figure 6A, SEQ ID NOs: 486-495) have been published. These crystal structures revealed that the longer heavy chain CDR3 regions for some antibodies protrude from the antibody Ig fold.

[00479] Ultralong CDR3s from bovine IgGs have been described, and recently the crystal structures of two have been published. Wang *et al.*, Cell, 153:1379-93 (2013). In addition to a beta-strand protrusion from the Ig fold, the cow ultralong CDR3s have a compact peptide domain supported by multiple disulfide bonds. The human HIV-1 neutralizing antibodies have little to no sequence homology to the cow ultralong CDR3s.

[00480] Described herein are exemplary methods for utilizing the human long CDR3 antibodies with longer heavy chain CDR3 regions as an IgG framework amenable to longer insertions comprising other functional domains, expressed within CDR3, thereby generating ultralong CDR3 containing antibodies.

[00481] In a first exemplary method to produce an antibody comprising an ultralong CDR3, the IgGs from mammalian cells are expressed. Nucleotide sequences encoding the heavy chain variable region from PGT145 (SEQ ID NO: 656) are amplified and inserted in-frame with pFuse HC vector (SEQ ID NO:458) using EcoRI and NheI, placing the PGT145 encoding sequence in frame between the secretion signal and human IgG1 CH1-CH2-CH3 in the vector. Amplification of the PGT145 encoding sequence can be accomplished by a variety of methods. In one method, the desired amino acid sequence is prepared as a synthetic sequence. A plasmid having a codon-optimized nucleotide sequence is used for expression of the submitted peptide sequence. In an alternative method, the desired peptide or nucleic acid sequence is prepared by automatic oligonucleotide design for PCR-based gene synthesis. For example, the method described in (Hoover, D. and Lubkowski, J., *Nucleic Acids Res.*, 30(10):e43 (2002)), is used to generate a series of nested, partially overlapping oligoprimers for subsequent PCR amplification of the entire desired sequence. A PGT145 amino acid sequence SEQ ID NO: 656 is prepared in which EcoRI, NheI, and BsaI sites are avoided. The following codon optimized nucleotide sequence is generated:

```
  1 CAGGTGCAGCTGGTGCAGTCAGGCGCCGAGGTGAAGAAGCCCGGCAGCAGCGTGAAGGTG
 61 AGCTGCAAAGCCAGCGGCAACAGCTTCAGCAACCACGACGTGCACTGGGTGAGGCAGGCC
121 ACCGGCCAGGGCCTGGAGTGGATGGGCTGGATGAGCCACGAGGGCGACAAGACCGGCCTG
181 GCCCAGAAGTTCCAGGGGAGGGTGACGATCACCAGAGACAGCGGAGCCAGCACCGTGTAC
241 ATGGAGCTGAGGGGCCTGACCGCCGACGACACCGCCATCTACTACTGCCTGACGGGCAGC
301 AAGCACAGGCTCAGGGACTACTTCCTGTACAACGAGTACGGCCCCAACTACGAGGAGTGG
361 GGCGACTACCTGGCCACCCTGGACGTGTGGGGCCACGGCACCGCCGTGACCGTGAGCAGT
421 GCCAGC
```

[00482] Appending gtcacgaattcg before and gctagcaccaag after the above sequence yields the following nucleotide sequence, where the EcoRI site and NheI site are underlined.

```
  1 GTCACGAATTCGCAGGTGCAGCTGGTGCAGTCAGGCGCCGAGGTGAAGAAGCCCGGCAGC
 61 AGCGTGAAGGTGAGCTGCAAAGCCAGCGGCAACAGCTTCAGCAACCACGACGTGCACTGG
121 GTGAGGCAGGCCACCGGCCAGGGCCTGGAGTGGATGGGCTGGATGAGCCACGAGGGCGAC
181 AAGACCGGCCTGGCCCAGAAGTTCCAGGGGAGGGTGACGATCACCAGAGACAGCGGAGCC
241 AGCACCGTGTACATGGAGCTGAGGGGCCTGACCGCCGACGACACCGCCATCTACTACTGC
301 CTGACGGGCAGCAAGCACAGGCTCAGGGACTACTTCCTGTACAACGAGTACGGCCCCAAC
361 TACGAGGAGTGGGGCGACTACCTGGCCACCCTGGACGTGTGGGGCCACGGCACCGCCGTG
421 ACCGTGAGCAGTGCCAGCGCTAGCACCAAG
```

[00483] This exemplary PGT145 heavy chain variable region sequence is used for in-frame insertion into pFuse HC vector (SEQ ID NO:458). The following list of oligonucleotides, when combined in a single PCR reaction, yields the desired nucleotide sequence encoding PGT145 VHC for insertion into pFuse HC vector using EcoRI and NheI.

```
1 GTCACGAATTCGCAGGTGCAGCTGGTGCAGTCAGGCGCCGAGGTGAAGAAGCCCGGC  57
  2 CTGAAGCTGTTGCCGCTGGCTTTGCAGCTCACCTTCACGCTGCTGCCGGGCTTCTTCACC  60
  3 AGCGGCAACAGCTTCAGCAACCACGACGTGCACTGGGTGAGGCAGGCCACCGGCCAGGGC  60
  4 GGCCGGTCTTGTCGCCCTCGTGGCTCATCCAGCCCATCCACTCCAGGCCCTGGCCGGTGG  60
  5 GGCGACAAGACCGGCCTGGCCCAGAAGTTCCAGGGGAGGGTGACGATCACCAGAGACAGC  60
  6 GCGGTCAGGCCCCTCAGCTCCATGTACACGGTGCTGGCTCCGCTGTCTCTGGTGATCGTC  60
  7 GAGGGGCCTGACCGCCGACGACACCGCCATCTACTACTGCCTGACGGGCAGCAAGCACAG  60
  8 CGTAGTTGGGGCCGTACTCGTTGTACAGGAAGTAGTCCCTGAGCCTGTGCTTGCTGCCCG  60
  9 AGTACGGCCCCAACTACGAGGAGTGGGGCGACTACCTGGCCACCCTGGACGTGTGGGGCC  60
 10 CTTGGTGCTAGCGCTGGCACTGCTCACGGTCACGGCGGTGCCGTGGCCCCACACGTCCAG  60
```

[00484] Next, DNA encoding the light chain of PGT145 is also prepared for insertion into a vector to facilitate the co-expression of both the heavy and light chains of PGT145. Nucleotide sequence encoding light chain PGT145 (SEQ ID NO: 491) may be obtained by any method well known in the art. Similar to the example above, the following nucleotide sequence is prepared that encodes PGT145 light chain variable and constant regions, and is devoid of BsaI, EcoRI, and NheI sites.

```
  1 GAGGTGGTCATCACGCAGAGCCCCCTCTTCCTGCCTGTGACACCCGGCGAGGCCGCCAGT
 61 CTGAGCTGCAAGTGCAGCCACAGCCTCCAGCACAGCACGGGTGCCAACTACCTGGCCTGG
121 TACCTGCAAAGGCCCGGTCAGACCCCCAGGCTGCTGATCCATCTGGCCACCCACAGGGCC
181 AGCGGCGTGCCCGACAGGTTCTCCGGCAGCGGCTCTGGCACCGACTTCACCCTGAAGATA
241 AGCAGGGTGGAGAGCGACGATGTGGGCACCTACTACTGTATGCAAGGCCTGCACAGCCCC
301 TGGACATTCGGCCAGGGCACCAAGGTAGAAATCAAGAGGACCGTGGCCGCACCCAGCGTG
361 TTCATCTTCCCACCCTCTGACGAGCAGCTTAAGAGCGGCACCGCCTCCGTTGTGTGCCTG
421 CTGAACAACTTCTACCCCAGGGAGGCCAAGGTGCAATGGAAAGTAGACAACGCCCTGCAG
481 AGCGGAAACAGCCAGGAAAGCGTGACCGAGCAAGACAGTAAGGACTCAACCTACAGCCTG
541 AGCAGCACGCTTACCCTCTCTAAGGCCGACTACGAGAAGCACAAGGTGTACGCCTGCGAG
601 GTGACCCACCAGGGCTTGTCTAGTCCCGTGACCAAGAGCTTCAACAGGGGCGAGTGC
```

[00485] Appending gtcacgaattcg before this sequence contributes EcoRI sites for cloning in frame with the pFuse LC (SEQ ID NO: 459). Adding taatgagctagctggcca after the above sequence adds stop codons at the end of the light chain and NheI site for cloning into the pFuse LC. This sequence is used for insertion of light chain PGT145 into pFuse LC vector: Note the underlined sequences indicate the restriction sites EcoRI and NheI.

```
   1 GTCACGAATTCGGAGGTGGTCATCACGCAGAGCCCCCTCTTCCTGCCTGTGACACCCGGC
  61 GAGGCCGCCAGTCTGAGCTGCAAGTGCAGCCACAGCCTCCAGCACAGCACGGGTGCCAAC
 121 TACCTGGCCTGGTACCTGCAAAGGCCCGGTCAGACCCCCAGGCTGCTGATCCATCTGGCC
 181 ACCCACAGGGCCAGCGGCGTGCCCGACAGGTTCTCCGGCAGCGGCTCTGGCACCGACTTC
 241 ACCCTGAAGATAAGCAGGGTGGAGAGCGACGATGTGGGCACCTACTACTGTATGCAAGGC
 301 CTGCACAGCCCCTGGACATTCGGCCAGGGCACCAAGGTAGAAATCAAGAGGACCGTGGCC
 361 GCACCCAGCGTGTTCATCTTCCCACCCTCTGACGAGCAGCTTAAGAGCGGCACCGCCTCC
 421 GTTGTGTGCCTGCTGAACAACTTCTACCCCAGGGAGGCCAAGGTGCAATGGAAAGTAGAC
 481 AACGCCCTGCAGAGCGGAAACAGCCAGGAAAGCGTGACCGAGCAAGACAGTAAGGACTCA
 541 ACCTACAGCCTGAGCAGCACGCTTACCCTCTCTAAGGCCGACTACGAGAAGCACAAGGTG
 601 TACGCCTGCGAGGTGACCCACCAGGGCTTGTCTAGTCCCGTGACCAAGAGCTTCAACAGG
 661 GGCGAGTGCTAATGAGCTAGCTGGCCA
```

[00486] The following oligonucleotides, when combined in a single PCR reaction, yield the desired nucleotide sequence encoding PGT145 light chain for insertion into pFuse LC vector using EcoRI and NheI.

```
  1 GTCACGAATTCGGAGGTGGTCATCACGCAGAGCCCCCTCTTCCTGCCTGTGACACCC  57
  2 TGGAGGCTGTGGCTGCACTTGCAGCTCAGACTGGCGGCCTCGCCGGGTGTCACAGGCAGG  60
  3 GCAGCCACAGCCTCCAGCACAGCACGGGTGCCAACTACCTGGCCTGGTACCTGCAAAGGC  60
  4 TGTGGGTGGCCAGATGGATCAGCAGCCTGGGGGTCTGACCGGGCCTTTGCAGGTACCAGG  60
  5 CCATCTGGCCACCCACAGGGCCAGCGGCGTGCCCGACAGGTTCTCCGGCAGCGGCTCTGG  60
  6 ACATCGTCGCTCTCCACCCTGCTTATCTTCAGGGTGAAGTCGGTGCCAGAGCCGCTGCCG  60
  7 GGTGGAGAGCGACGATGTGGGCACCTACTACTGTATGCAAGGCCTGCACAGCCCCTGGAC  60
  8 CGGCCACGGTCCTCTTGATTTCTACCTTGGTGCCCTGGCCGAATGTCCAGGGGCTGTGCA  60
  9 AAGAGGACCGTGGCCGCACCCAGCGTGTTCATCTTCCCACCCTCTGACGAGCAGCTTAAG  60
 10 AGAAGTTGTTCAGCAGGCACACAACGGAGGCGGTGCCGCTCTTAAGCTGCTCGTCAGAGG  60
 11 GTGCCTGCTGAACAACTTCTACCCCAGGGAGGCCAAGGTGCAATGGAAAGTAGACAACGC  60
 12 TGCTCGGTCACGCTTTCCTGGCTGTTTCCGCTCTGCAGGGCGTTGTCTACTTTCCATTGC  60
 13 GGAAAGCGTGACCGAGCAAGACAGTAAGGACTCAACCTACAGCCTGAGCAGCACGCTTAC  60
 14 CAGGCGTACACCTTGTGCTTCTCGTAGTCGGCCTTAGAGAGGGTAAGCGTGCTGCTCAGG  60
 15 GCACAAGGTGTACGCCTGCGAGGTGACCCACCAGGGCTTGTCTAGTCCCGTGACCAAGAG  60
 16 TGGCCAGCTAGCTCATTAGCACTCGCCCCTGTTGAAGCTCTTGGTCACGGGACTAGA  57
```

[00487] Further modifications to the resulting PGT145 vector may be made using any number of methods known in the art. For example, overlap PCR is used to join nucleotide sequences encoding PGT145 heavy chain variable region sequences with amino acid linkers, and toxins for expression within the IgG. In an exemplary method, two nucleotide fragments encoding sequences from PGT145 are generated. Also, a sequence of linkers and or toxins is made, for later insertion between the 2 PGT145 encoding fragments by PCR. The specific residues encoded in the PGT145 fragments specify the residues where the insertion will be positioned. For example, a nucleotide sequence encoding the first 109 amino acids of PGT145, equivalent to the residues in SEQ ID

NO:496 and SEQ ID NO:504 is amplified by PCR. The EcoRI sites described above are also incorporated for subsequent pFuse insertion.

```
GTCACGAATTCGcaggtgcagctggtgcagtcaggcgccgaggtgaagaagcccggcagcagcgtgaaggtg
             Q  V  Q  L  V  Q  S  G  A  E  V  K  K  P  G  S  S  V  K  V
agctgcaaagccagcggcaacagcttcagcaaccacgacgtgcactgggtgaggcaggcc
 S  C  K  A  S  G  N  S  F  S  N  H  D  V  H  W  V  R  Q  A
accggccagggcctggagtggatgggctggatgagccacgagggcgacaagaccggcctg
 T  G  Q  G  L  E  W  M  G  W  M  S  H  E  G  D  K  T  G  L
gcccagaagttccaggggagggtgacgatcaccagagacagcggagccagcaccgtgtac
 A  Q  K  F  Q  G  R  V  T  I  T  R  D  S  G  A  S  T  V  Y
atggagctgaggggcctgaccgccgacgacaccgccatctactactgcctgacgggcagc
 M  E  L  R  G  L  T  A  D  D  T  A  I  Y  Y  C  L  T  G  S
aagcacaggctcagggactacttcctg
 K  H  R  L  R  D  Y  F  L
```

[00488] The sequence encoding amino acids 117 through 140, equivalent to SEQ ID NO: 540 and SEQ ID NO:570 together, is also amplified by PCR, and the NheI site described above for pFuse insertion is incorporated as well.

```
tacgaggagtggggcgactacctggccaccctggacgtgtggggccacggcaccgccgtgaccgtg
 Y  E  E  W  G  D  Y  L  A  T  L  D  V  W  G  H  G  T  A  V  T  V
agcagtGCTAGCTGGCCA
 S  S  A  S
```

By inclusion of residues 1-109 in one fragment, and residues 117 through 140 in the other fragment of PGT145, the inserted sequence of linkers and or toxins are positioned between residues 109 and 117 of PGT145,effectively replacing residues 110-116, which are YNEYGPN.

[00489] The above PGT145 encoding nucleotides may be joined with any intervening sequence, for example, by overlap PCR, effecting an insertion in PGT145 ready for subloning into pFuse using EcoRI and NheI, as described above. Amino acids encoded by such an intervening nucleotide sequence may be expressed anywhere the boundaries of the two PGT145 encoding nucleotide fragments are chosen. For example, the boundary may be within CDR3, such as at the end of the ascending and descending stalk sequences in CDR3 of the PGT145 heavy chain variable region. Such insertions can encode any sequence, including sequences from other antibodies, non-antibody sequences, flexible linkers, toxins, or combinations thereof. For example, nucleotide sequences encoding amino acid sequences of any of the toxins identified as SEQ ID NO: 599-733, 754, 755, or 774-778 (see also, Figure 7B) may be joined with nucleotides encoding any of the amino acid linker sequence, such as SEQ ID NO: 575 to 598, 699 to 726, or 756-773 (see also, Figure 7A), or no linker at all, for insertion between the PGT145 variable region nucleotide fragments.

[00490] For example, the intervening sequence may encode a toxin flanked by flexible linkers. The sequence below encoding the $(G_4S)_3GG$ linker (SEQ ID NO: 724),

Ggtggaggaggttctggaggcggtggaagtggtggcggaggtagcggagga
G G G G S G G G G S G G G G S G G

the sequence encoding ShK toxin (SEQ ID NO: 648),

aggagctgcatcgacaccatccccaagagccgatgcaccgccttccagtgcaagcacagcatgaagtacaggct
gagcttctgcaggaagacctgcggcacctgc

and the sequence encoding the linker $(G_4S)_3$ (SEQ ID NO:577),

ggaggtggtggatctggtggaGgaggcagtggaggtggtggcagc
G G G G S G G G G S G G G G S

are all joined by overlap PCR. The new $(G_4S)_3GG$-ShK Toxin-$(G_4S)_3$ encoding sequence is then joined by, for example, overlap PCR with the two fragments described previously, encoding PGT145 1-109 and PGT145 117-140, to give the following sequence encoding $(G_4S)_3GG$-ShK Toxin-$(G_4S)_3$ in CDR3 of PGT145.

GTCACGAATTCGcaggtgcagctggtgcagtcaggcgccgaggtgaagaagcccggcagcagcgtgaaggtg
agctgcaaagccagcggcaacagcttcagcaaccacgacgtgcactgggtgaggcaggcc
accggccagggcctggagtggatgggctggatgagccacgagggcgacaagaccggcctg
gcccagaagttccaggggagggtgacgatcaccagagacagcggagccagcaccgtgtac
atggagctgaggggcctgaccgccgacgacaccgccatctactactgcctgacgggcagc
aagcacaggctcagggactacttcctg *Ggtggaggaggttctggaggcggtggaagtgg*
tggcggaggtagcggagga **aggagctgcatcgacaccatccccaagagccgatgcaccg**
ccttccagtgcaagcacagcatgaagtacaggctgagcttctgcaggaagacctgcggca
cctgc *ggaggtggtggatctggtggaGgaggcagtggaggtggtggcagc* tacgagga
gtggggcgactacctggccaccctggacgtgtggggcca cggcaccgccgtgaccgtga
gcagtGCTAGCTGGCCA

Underline = EcoRI or NheI for cloning into pFuse HC
Normal Font = Derived from PGT145
Italics = GGGGS Linkers
Bold = ShK Toxin

[00491] In addition, several other toxins derived from natural sources and recombinant libraries (see, e.g., Figures 7B and 7C) can be inserted into the vectors using the methods described above.

[00492] In addition to the example outlined above for insertion into PGT145 (SEQ ID NO:656) between amino acids 109 and 117, removing amino acids 110-116 (YNEYGPN) and replacing them with fusions (for example linkers and or toxins) in CDR3, these insertions may be positioned at other positions by modifying the initial PGT145 sequences appended to the insert. For example, to insert amino acids between the L position at 109 and the Y position at 110, the initial fragments of PGT145 encode amino acids 1-109, and 110-140 respectively. The following table illustrates other possible insertion points in PGT145, and the fragments of PGT145 to flank the insert.

PGT145 (SEQ ID NO:656)		Amino Acids encoded in	
Insertion Position:	Amino Acids removed	Fragment 1	Fragment 2
Between L109 and Y110	-	1-109	110-140
Between Y110 and N111	-	1-110	111-140
Between N111 and E112	-	1-111	112-140
Between E112 and Y113	-	1-112	113-140
Between Y113 and G114	-	1-113	114-140
Between G114 and P115	-	1-114	115-140
Between P115 and N116	-	1-115	116-140
Between N116 and Y117	-	1-116	117-140
Between L109 and Y117	YNEYGPN, 110-116	1-109	117-140
Between G114 and N116	P115	1-114	116-140

[00493] In addition, other VH antibody frameworks are modified with linker and/or toxin insertions in the manner described above. For example, the PG9 heavy chain variable region sequence (SEQ ID NO: 657) is modified with a linker and/or toxin insertion as described for the PGT145 antibody above. As described above, two nucleotide fragments encoding residues of PG9 are amplified, and the linker/toxin insert is positioned between them. The position of the insertion in the PG9 variable region sequence is determined by the residues encoded in the two nucleotide fragments appended to the toxin. The following table illustrates possible insertion positions in the PG9 variable region, and the fragments of PG9 variable region to flank the insert.

PG9 (SEQ ID NO:657)		Amino Acids encoded in	
Insertion Position:	Amino Acids removed	Fragment 1	Fragment 2
Between Y104 and R105	-	1-104	105-136
Between R105 and N106	-	1-105	106-136
Between N106 and G107	-	1-106	107-136
Between G107 and Y108	-	1-107	108-136
Between Y108 and N109	-	1-108	109-136
Between N109 and Y110	-	1-109	110-136
Between Y110 and Y111	-	1-110	111-136
Between Y111 and D112	-	1-111	112-136
Between D112 and F113	-	1-112	113-136
Between F113 and Y114	-	1-113	114-136
Between Y114 and D115	-	1-114	115-136
Between D115 and G116	-	1-115	116-136
Between Y108 and F113	NYYD, 109-112	1-108	113-136
Between N109 and Y111	Y110	1-109	111-136

[00494] In addition, other VH antibody frameworks are modified with linker and/or toxin insertions in the manner described above. For example, the PG16 heavy chain variable region sequence (SEQ ID NO: 658) is modified with a linker and/or toxin insertion as described for the PGT145 antibody above. As described above, two nucleotide fragments encoding residues of PG16 are amplified, and the linker/toxin insert is positioned between them. The position of the insertion in the PG16 variable region sequence is determined by the residues encoded in the two nucleotide fragments appended to the toxin. The following table illustrates possible insertion positions in the PG16 variable region, and the fragments of PG16 variable region to flank the insert.

PG16 (SEQ ID NO:658)		Amino Acids encoded in	
Insertion Position:	Amino Acids removed	Fragment 1	Fragment 2
Between W105 and H106	-	1-105	106-137
Between H106 and D107	-	1-106	107-137
Between D107 and D108	-	1-107	108-137
Between D108 and V109	-	1-108	109-137
Between V109 and K110	-	1-109	110-137
Between H106 and V109	DD, 107-108	1-106	109-137

[00495] In addition, other VH antibody frameworks are modified with linker and/or toxin insertions in the manner described above. For example, the CHO4 heavy chain variable region sequence (SEQ ID NO: 659) is modified with a linker and/or toxin insertion as described for the PGT145 antibody above. As described above, two nucleotide fragments encoding residues of CHO4 are amplified, and the linker/toxin insert is positioned between them. The position of the insertion in the CHO4 variable region sequence is determined by the residues encoded in the two nucleotide fragments appended to the toxin. The following table illustrates possible insertion positions in the CHO4 variable region, and the fragments of CHO4 variable region to flank the insert.

CHO4 (SEQ ID NO: 659)		Amino Acids encoded in	
Insertion Position:	Amino Acids removed	Fragment 1	Fragment 2
Between I104 and D105	-	1-104	105-133
Between D105 and D106	-	1-105	106-133
Between D106 and Q107	-	1-106	107-133
Between Q107 and G108	-	1-107	108-133
Between G108 and I109	-	1-108	109-133
Between D106 and I109	D107, Q108	1-106	109-133

[00496] In addition, other VH antibody frameworks are modified with linker and/or toxin insertions in the manner described above. For example, the 2909 heavy chain variable region sequence (SEQ ID NO: 660) is modified with a linker and/or toxin insertion as described for the PGT145 antibody above. As described above, two nucleotide fragments encoding residues of 2909 are amplified, and the linker/toxin insert is positioned between them. The position of the insertion in the 2909 variable region sequence is determined by the residues encoded in the two nucleotide fragments appended to the toxin. The following table illustrates possible insertion positions in the 2909 variable region, and the fragments of 2909 variable region to flank the insert.

2909 (SEQ ID NO: 660)		Amino Acids encoded in	
Insertion Position:	Amino Acids removed	Fragment 1	Fragment 2
Between Y105 and D106	-	1-105	106-130
Between D106 and Y107	-	1-106	107-130
Between Y107 and N108	-	1-107	108-130
Between N108 and L109	-	1-108	109-130
Between L109 and G110	-	1-109	110-130
Between G110 and Y111	-	1-110	111-130
Between D106 and G110	YNL 107-109	1-106	110-130

Example 2. Expression of polynucleotides coding antibodies that include a long CDR3.

[00497] Polynucleotides coding for antibodies having a long CDR3 can be expressed including, transiently expressed, in a host cell by any method known in the art. In an exemplary method, the vectors comprising the heavy and light chains generated in Example 1 are transfected into either 293T cells or Freestyle™ 293-F cells. For example, 130,000 293T cells per well are being plated in 24 well plates and grown overnight in 500 µl DMEM media (Invitrogen) with 10% FBS (Invitrogen), and penicillin/streptomycin/glutamine (Invitrogen) at 37°C and 5% CO2. Next, 0.5 µg of Heavy chain-encoding pFuse vector and 0.5 µg of Light chain encoding pFuse vector (e.g., generated in Example 1) are added to 25 µl of Opti-MEM (Invitrogen). Subsequently, 1 µl of Lipofectamine 2000 or 293Fectin transfection reagent (Invitrogen) are added to 25 µl of Opti-MEM, and incubated 5 minutes. The dna-Opti- MEM solution and transfection reagent-Opti-MEM solution are combined and incubated for 15 minutes, added to 293T cells, and allowed to incubate on cells for 4-6 hours. Media is aspirated from wells and replaced with fresh media, and the cells are allowed to grow and secrete IgG into the media for 2-6 days.

[00498] Additionally, for example, $1x10^6$ 293Freestyle suspension cells/ml are grown overnight in Freestyle™293 Expression Medium (Invitrogen), and penicillin/streptomycin/ glutamine (Invitrogen) at 37°C and 5% CO_2. Each milliliter of cells, a 30µl solution is made in buffer PBS comprising 0.5 µg Heavy chain-encoding pFuse DNA and 0.5 µg of Lc-encoding pFuse. Additionally, for each milliliter of cells, a solution comprising 1 µl of 293Fectin transfection reagent is added to 30µl PBS and incubated 5 minutes. Subsequently, the DNA and lipofectamine solutions are combined and incubated 15 minutes before adding the mixture to the cells. The cells are then allowed to grow and secrete IgG into media for 2-6 days.

[00499] After the growth period for the 293T cells or Freestyle™ 293-F suspension cells, media is harvested, and IgG secreted into the supernatant is evaluated by sandwich ELISA. Briefly, Fc-specific anti Human IgG (Cat I2136, Sigma-Aldrich) is diluted 1:1000 in PBS and coated onto maxisorp plates (Nunc). Plates are blocked with 2% BSA in TBST, washed, and IgG secreted supernatants are incubated for 1 hour. After washing with TBST, the appropriate HRP conjugated anti-Light chain antibody is diluted 1:1000 in TBST and incubated 1 hour (anti kappa-HRP, Cat#A-7164, Sigma- Aldrich; anti-lambda-HRP, Cat# 2070-05, Southern Biotech). After washing with TBST, HRP was detected with TMB (Cat#TMBS-1000-01, BioFX), and neutralized with 0.6M H2SO4. Subsequently, absorbance (A450) is measured, and compared to a standard curve of known IgG concentration to determine the concentration of human antibody in cell culture supernatant. The yield of antibody from supernatants from cells transfected with heavy chain constructs encoding PGT145 with human CH1-CH2-CH3 and a replacement of at least a portion of CDR3 and a light chain construct encoding PGT145 LV-human lambda LC is determined by ELISA.

[00500] For the disclosure herein, the very least, and not as an attempt to limit the application of the doctrine of equivalents to the scope of the claims, each numerical parameter should at least be construed in light of the number of reported significant digits and by applying ordinary rounding techniques. Notwithstanding that the numerical ranges and parameters setting forth the broad scope of the disclosure are approximations, the numerical values set forth in the specific examples are reported as precisely as possible. Any numerical value, however, inherently contains certain errors necessarily resulting from the standard deviation found in their respective testing measurements.

[00501] The terms “a,” “an,” “the” and similar referents used in the context of describing the exemplary embodiments (especially in the context of the following claims) are to be construed to cover both the singular and the plural, unless otherwise indicated herein or clearly contradicted by context. Recitation of ranges of values herein is merely

intended to serve as a shorthand method of referring individually to each separate value falling within the range. Unless otherwise indicated herein, each individual value is incorporated into the specification as if it were individually recited herein. All methods described herein can be performed in any suitable order unless otherwise indicated herein or otherwise clearly contradicted by context. The use of any and all examples, or exemplary language (e.g., "such as") provided herein is intended merely to better illuminate the exemplary embodiments and does not pose a limitation on the scope of the exemplary embodiments otherwise claimed. No language in the specification should be construed as indicating any non-claimed element essential to the practice of the exemplary embodiments.

[00502] Groupings of alternative elements or embodiments disclosed herein are not to be construed as limitations. Each group member can be referred to and claimed individually or in any combination with other members of the group or other elements found herein. It is anticipated that one or more members of a group can be included in, or deleted from, a group for reasons of convenience and/or patentability. When any such inclusion or deletion occurs, the specification is deemed to contain the group as modified thus fulfilling the written description of all Markush groups used in the appended claims.

[00503] Certain embodiments are described herein, including the best mode known to the inventors for carrying out the exemplary embodiments. Of course, variations on these described embodiments will become apparent to those of ordinary skill in the art upon reading the foregoing description. The inventor expects skilled artisans to employ such variations as appropriate, and the inventors intend for the embodiments to be practiced otherwise than specifically described herein. Accordingly, this disclosure includes all modifications and equivalents of the subject matter recited in the claims appended hereto as permitted by applicable law. Moreover, any combination of the above-described elements in all possible variations thereof is encompassed by the disclosure unless otherwise indicated herein or otherwise clearly contradicted by context.

[00504] Furthermore, numerous references have been made to patents and printed publications. Each of the above-cited references is individually incorporated herein by reference in their entirety.

[00505] Specific embodiments disclosed herein can be further limited in the claims using consisting of or and consisting essentially of language. When used in the claims, whether as filed or added per amendment, the transition term "consisting of" excludes any element, step, or ingredient not specified in the claims. The transition term "consisting essentially of" limits the scope of a claim to the specified materials or steps and those that

do not materially affect the basic and novel characteristic(s). Exemplary embodiments so claimed are inherently or expressly described and enabled herein.

[00506] In closing, it is to be understood that the exemplary embodiments disclosed herein are illustrative of the principles of the present disclosure. Other modifications that can be employed are within the scope of the disclosure. Thus, by way of example, but not of limitation, alternative configurations of the present exemplary embodiments can be utilized in accordance with the teachings herein. Accordingly, the present exemplary embodiments are not limited to that precisely as shown and described.

CLAIMS

1. An antibody heavy chain variable region comprising a sequence of the formula V1-X-V2, wherein V1 comprises an amino acid sequence selected from the group consisting of:

(i) QVQLVQSGAEVKKPGSSVKVSCKASGNSFSNHDVHWVRQATGQGLEWMGW MSHEGDKTGLAQKFQGRVTITRDSGASTVYMELRGLTADDTAIYYCLT (SEQ ID NO: 496),

(ii) QRLVESGGGVVQPGSSLRLSCAASGFDFSRQGMHWVRQAPGQGLEWVAFIK YDGSEKYHADSVWGRLSISRDNSKDTLYLQMNSLRVEDTATYFCVR (SEQ ID NO: 497),

(iii) QEQLVESGGGVVQPGGSLRLSCLASGFTFHKYGMHWVRQAPGKGLEWVALI SDDGMRKYHSDSMWGRVTISRDNSKNTLYLQFSSLKVEDTAMFFCAR (SEQ ID NO: 498),

(iv) EVQLVESGGGLIRPGGSLRLSCKGSGFIFENFGFGWVRQGPGKGLEWVSGT NWNGGDSRYGDSVKGRFTISRDNSNNFVYLQMNSLRPEDTAIYYCAR (SEQ ID NO: 499), and

(v) EVQLVESGGNVVQPGGSLRLSCTASGFSFDDSTMHWVRQAPGKGLQWVSLI SWNGGRTYYADSVKGRFTISRDNSKNSLYLQMNSLKTEDTAFYFCAK (SEQ ID NO: 500);

wherein X comprises an ultralong CDR3; and

wherein V2 comprises an amino acid sequence selected from the group consisting of:

(i) WGHGTAVTVSS (SEQ ID NO: 570),

(ii) WGKGTTVTVSS (SEQ ID NO: 571),

(iii) WGKGTTVTVSS (SEQ ID NO: 572),

(iv) WGRGTLVTVSS (SEQ ID NO: 573), and

(v) WGKGTTVTVSS (SEQ ID NO: 574).

2. The antibody heavy chain variable region of claim 1, wherein the ultralong CDR3 comprises an amino acid sequence of:

(i) any one of GSKHRLRDYFLYNE (SEQ ID NO: 501), GSKHRLRDYFLYN (SEQ ID NO: 502), GSKHRLRDYFLY (SEQ ID NO: 503), GSKHRLRDYFL (SEQ ID NO:

504), GSKHRLRDYF (SEQ ID NO: 505), GSKHRLRDY (SEQ ID NO: 506), or GSKHRLRD (SEQ ID NO: 507);

(ii) any one of EAGGPDYRNGYNY (SEQ ID NO: 508), EAGGPDYRNGYN (SEQ ID NO: 509), EAGGPDYRNGY (SEQ ID NO: 510), EAGGPDYRNG (SEQ ID NO: 511), EAGGPDYRN (SEQ ID NO: 512), EAGGPDYR (SEQ ID NO: 513), EAGGPDY (SEQ ID NO: 514), or EAGGPD (SEQ ID NO: 515);

(iii) any one of EAGGPIWHDDVKY (SEQ ID NO: 516), EAGGPIWHDDVK (SEQ ID NO: 517), EAGGPIWHDDV (SEQ ID NO: 518), EAGGPIWHDD (SEQ ID NO: 519), EAGGPIWHD (SEQ ID NO: 520), EAGGPIWH (SEQ ID NO: 521), EAGGPIW (SEQ ID NO: 522), or EAGGPI (SEQ ID NO: 523);

(iv) any one of GTDYTIDDQGI (SEQ ID NO: 524), GTDYTIDDQG (SEQ ID NO: 525), GTDYTIDDQ (SEQ ID NO: 526), GTDYTIDD (SEQ ID NO: 527), GTDYTID (SEQ ID NO: 528), or GTDYTI (SEQ ID NO: 529); or

(v) any one of DKGDSDYDYNL (SEQ ID NO: 530), DKGDSDYDYN (SEQ ID NO: 531), DKGDSDYDY (SEQ ID NO: 532), DKGDSDYD (SEQ ID NO: 533), DKGDSDY (SEQ ID NO: 534), DKGDSD (SEQ ID NO: 535).

3. The antibody heavy chain variable region of claim 1, wherein the ultralong CDR3 comprises an amino acid sequence of:

(i) any one of YGPNYEEWGDYLATLDV (SEQ ID NO: 536), GPNYEEWGDYLATLDV (SEQ ID NO: 537), PNYEEWGDYLATLDV (SEQ ID NO: 538), NYEEWGDYLATLDV (SEQ ID NO: 539), YEEWGDYLATLDV (SEQ ID NO: 540), or EEWGDYLATLDV (SEQ ID NO: 541);

(ii) any one of YDFYDGYYNYHYMDV (SEQ ID NO: 542), DFYDGYYNYHYMDV (SEQ ID NO: 543), FYDGYYNYHYMDV (SEQ ID NO: 544), YDGYYNYHYMDV (SEQ ID NO: 545), DGYYNYHYMDV (SEQ ID NO: 546), GYYNYHYMDV (SEQ ID NO: 547), or YYNYHYMDV (SEQ ID NO: 548);

(iii) any one of YDFNDGYYNYHYMDV (SEQ ID NO: 549), DFYDGYYNYHYMDV (SEQ ID NO: 550), FYDGYYNYHYMDV (SEQ ID NO: 551), YDGYYNYHYMDV (SEQ ID NO: 552), DGYYNYHYMDV (SEQ ID NO: 553), or GYYNYHYMDV (SEQ ID NO: 554);

(iv) any one of QGIRYQGSGTFWYFDV (SEQ ID NO: 555), GIRYQGSGTFWYFDV (SEQ ID NO: 556), IRYQGSGTFWYFDV (SEQ ID NO: 557), RYQGSGTFWYFDV (SEQ ID NO: 558), YQGSGTFWYFDV (SEQ ID NO: 559), QGSGTFWYFDV (SEQ

ID NO: 560), GSGTFWYFDV (SEQ ID NO: 561), SGTFWYFDV (SEQ ID NO: 562), or GTFWYFDV (SEQ ID NO: 563); or

(v) any one of YNLGYSYFYYMDG(SEQ ID NO: 564), NLGYSYFYYMDG (SEQ ID NO: 565), LGYSYFYYMDG (SEQ ID NO: 566), GYSYFYYMDG (SEQ ID NO: 567), YSYFYYMDG (SEQ ID NO: 568), or SYFYYMDG (SEQ ID NO: 569).

4. The antibody heavy chain variable region of claim 1, wherein the ultralong CDR3 comprises an amino acid sequence of:

(i) any one of GSKHRLRDYFLYNE (SEQ ID NO: 501), GSKHRLRDYFLYN (SEQ ID NO: 502), GSKHRLRDYFLY (SEQ ID NO: 503), GSKHRLRDYFL (SEQ ID NO: 504), GSKHRLRDYF (SEQ ID NO: 505), GSKHRLRDY (SEQ ID NO: 506), or GSKHRLRD (SEQ ID NO: 507), and any one of YGPNYEEWGDYLATLDV (SEQ ID NO: 536), GPNYEEWGDYLATLDV (SEQ ID NO: 537), PNYEEWGDYLATLDV (SEQ ID NO: 538), NYEEWGDYLATLDV (SEQ ID NO: 539), YEEWGDYLATLDV (SEQ ID NO: 540), or EEWGDYLATLDV (SEQ ID NO: 541);

(ii) any one of EAGGPDYRNGYNY (SEQ ID NO: 508), EAGGPDYRNGYN (SEQ ID NO: 509), EAGGPDYRNGY (SEQ ID NO: 510), EAGGPDYRNG (SEQ ID NO: 511), EAGGPDYRN (SEQ ID NO: 512), EAGGPDYR (SEQ ID NO: 513), EAGGPDY (SEQ ID NO: 514), or EAGGPD (SEQ ID NO: 515), and any one of YDFYDGYYNYHYMDV (SEQ ID NO: 542), DFYDGYYNYHYMDV (SEQ ID NO: 543), FYDGYYNYHYMDV (SEQ ID NO: 544), YDGYYNYHYMDV (SEQ ID NO: 545), DGYYNYHYMDV (SEQ ID NO: 546), GYYNYHYMDV (SEQ ID NO: 547), or YYNYHYMDV (SEQ ID NO: 548);

(iii) any one of EAGGPIWHDDVKY (SEQ ID NO: 516), EAGGPIWHDDVK (SEQ ID NO: 517), EAGGPIWHDDV (SEQ ID NO: 518), EAGGPIWHDD (SEQ ID NO: 519), EAGGPIWHD (SEQ ID NO: 520), EAGGPIWH (SEQ ID NO: 521), EAGGPIW (SEQ ID NO: 522), or EAGGPI (SEQ ID NO: 523), and any one of YDFNDGYYNYHYMDV (SEQ ID NO: 549), DFYDGYYNYHYMDV (SEQ ID NO: 550), FYDGYYNYHYMDV (SEQ ID NO: 551), YDGYYNYHYMDV (SEQ ID NO: 552), DGYYNYHYMDV (SEQ ID NO: 553), or GYYNYHYMDV (SEQ ID NO: 554);

(iv) any one of GTDYTIDDQGI (SEQ ID NO: 524), GTDYTIDDQG (SEQ ID NO: 525), GTDYTIDDQ (SEQ ID NO: 526), GTDYTIDD (SEQ ID NO: 527), GTDYTID (SEQ ID NO: 528), or GTDYTI (SEQ ID NO: 529), and any one of QGIRYQGSGTFWYFDV (SEQ ID NO: 555), GIRYQGSGTFWYFDV (SEQ ID NO: 556), IRYQGSGTFWYFDV (SEQ ID NO: 557), RYQGSGTFWYFDV (SEQ ID NO:

558), YQGSGTFWYFDV (SEQ ID NO: 559), QGSGTFWYFDV (SEQ ID NO: 560), GSGTFWYFDV (SEQ ID NO: 561), SGTFWYFDV (SEQ ID NO: 562), or GTFWYFDV (SEQ ID NO: 563); or

(v) any one of YNLGYSYFYYMDG(SEQ ID NO: 564), NLGYSYFYYMDG (SEQ ID NO: 565), LGYSYFYYMDG (SEQ ID NO: 566), GYSYFYYMDG (SEQ ID NO: 567), YSYFYYMDG (SEQ ID NO: 568), or SYFYYMDG (SEQ ID NO: 569), and any one of YNLGYSYFYYMDG(SEQ ID NO: 564), NLGYSYFYYMDG (SEQ ID NO: 565), LGYSYFYYMDG (SEQ ID NO: 566), GYSYFYYMDG (SEQ ID NO: 567), YSYFYYMDG (SEQ ID NO: 568), or SYFYYMDG (SEQ ID NO: 569).

5. The antibody heavy chain variable region of claim 1,

-wherein V1 comprises an amino acid sequence of QVQLVQSGAEVKKPGSSVKVSCKASGNSFSNHDVHWVRQAT GQGLEWMGWMSHEGDKTGLAQKFQGRVTITRDSGASTVYMELRGLTADDTAIY YCLT (SEQ ID NO: 496),

-wherein the ultralong CDR3 comprises an amino acid sequence of any one of GSKHRLRDYFLYNE (SEQ ID NO: 501), GSKHRLRDYFLYN (SEQ ID NO: 502), GSKHRLRDYFLY (SEQ ID NO: 503), GSKHRLRDYFL (SEQ ID NO: 504), GSKHRLRDYF (SEQ ID NO: 505), GSKHRLRDY (SEQ ID NO: 506), or GSKHRLRD (SEQ ID NO: 507), and an amino acid sequence of any one of YGPNYEEWGDYLATLDV (SEQ ID NO: 536), GPNYEEWGDYLATLDV (SEQ ID NO: 537), PNYEEWGDYLATLDV (SEQ ID NO: 538), NYEEWGDYLATLDV (SEQ ID NO: 539), YEEWGDYLATLDV (SEQ ID NO: 540), or EEWGDYLATLDV (SEQ ID NO: 541),

-wherein V2 comprises an amino acid sequence selected of WGHGTAVTVSS (SEQ ID NO: 570).

6. The antibody heavy chain variable region of claim 1,

-wherein V1 comprises an amino acid sequence of QRLVESGGGVVQPGSSLRLSCAASGFDFSRQGMHWVRQAPG QGLEWVAFIKYDGSEKYHADSVWGRLSISRDNSKDTLYLQMNSLRVEDTATYFCV R (SEQ ID NO: 497),

-wherein the ultralong CDR3 comprises an amino acid sequence of any one of EAGGPDYRNGYNY (SEQ ID NO: 508), EAGGPDYRNGYN (SEQ ID NO: 509), EAGGPDYRNGY (SEQ ID NO: 510), EAGGPDYRNG (SEQ ID NO: 511),

EAGGPDYRN (SEQ ID NO: 512), EAGGPDYR (SEQ ID NO: 513), EAGGPDY (SEQ ID NO: 514), or EAGGPD (SEQ ID NO: 515), and an amino acid sequence of any one of YDFYDGYYNYHYMDV (SEQ ID NO: 542), DFYDGYYNYHYMDV (SEQ ID NO: 543), FYDGYYNYHYMDV (SEQ ID NO: 544), YDGYYNYHYMDV (SEQ ID NO: 545), DGYYNYHYMDV (SEQ ID NO: 546), GYYNYHYMDV (SEQ ID NO: 547), or YYNYHYMDV (SEQ ID NO: 548),

-wherein V2 comprises an amino acid sequence selected of WGKGTTVTVSS (SEQ ID NO: 571).

7. The antibody heavy chain variable region of claim 1,

-wherein V1 comprises an amino acid sequence of QEQLVESGGGVVQPGGSLRLSCLASGFTFHKYGMHWVRQAPGKGLEWVALISDDGMRKYHSDSMWGRVTISRDNSKNTLYLQFSSLKVEDTAMFFCAR (SEQ ID NO: 498),

-wherein the ultralong CDR3 comprises an amino acid sequence of any one of EAGGPIWHDDVKY (SEQ ID NO: 516), EAGGPIWHDDVK (SEQ ID NO: 517), EAGGPIWHDDV (SEQ ID NO: 518), EAGGPIWHDD (SEQ ID NO: 519), EAGGPIWHD (SEQ ID NO: 520), EAGGPIWH (SEQ ID NO: 521), EAGGPIW (SEQ ID NO: 522), or EAGGPI (SEQ ID NO: 523), and an amino acid sequence of any one of YDFNDGYYNYHYMDV (SEQ ID NO: 549), DFYDGYYNYHYMDV (SEQ ID NO: 550), FYDGYYNYHYMDV (SEQ ID NO: 551), YDGYYNYHYMDV (SEQ ID NO: 552), DGYYNYHYMDV (SEQ ID NO: 553), or GYYNYHYMDV (SEQ ID NO: 554) ,

-wherein V2 comprises an amino acid sequence selected of WGKGTTVTVSS (SEQ ID NO: 572).

8. The antibody heavy chain variable region of claim 1,

-wherein V1 comprises an amino acid sequence of EVQLVESGGGLIRPGGSLRLSCKGSGFIFENFGFGWVRQGPGKGLEWVSGTNWNGGDSRYGDSVKGRFTISRDNSNNFVYLQMNSLRPEDTAIYYCAR (SEQ ID NO: 499),

-wherein the ultralong CDR3 comprises an amino acid sequence of any one of GTDYTIDDQGI (SEQ ID NO: 524), GTDYTIDDQG (SEQ ID NO: 525), GTDYTIDDQ (SEQ ID NO: 526), GTDYTIDD (SEQ ID NO: 527), GTDYTID (SEQ ID NO: 528), or GTDYTI (SEQ ID NO: 529), and an amino acid sequence of any

one of QGIRYQGSGTFWYFDV (SEQ ID NO: 555), GIRYQGSGTFWYFDV (SEQ ID NO: 556), IRYQGSGTFWYFDV (SEQ ID NO: 557), RYQGSGTFWYFDV (SEQ ID NO: 558), YQGSGTFWYFDV (SEQ ID NO: 559), QGSGTFWYFDV (SEQ ID NO: 560), GSGTFWYFDV (SEQ ID NO: 561), SGTFWYFDV (SEQ ID NO: 562), or GTFWYFDV (SEQ ID NO: 563) ,

-wherein V2 comprises an amino acid sequence selected of WGRGTLVTVSS (SEQ ID NO: 573).

9. The antibody heavy chain variable region of claim 1,

-wherein V1 comprises an amino acid sequence of EVQLVESGGNVVQPGGSLRLSCTASGFSFDDSTMHWVRQAP GKGLQWVSLISWNGGRTYYADSVKGRFTISRDNSKNSLYLQMNSLKTEDTAFYFC AK (SEQ ID NO: 500),

-wherein the ultralong CDR3 comprises an amino acid sequence of any one of YNLGYSYFYYMDG(SEQ ID NO: 564), NLGYSYFYYMDG (SEQ ID NO: 565), LGYSYFYYMDG (SEQ ID NO: 566), GYSYFYYMDG (SEQ ID NO: 567), YSYFYYMDG (SEQ ID NO: 568), or SYFYYMDG (SEQ ID NO: 569), and any one of YNLGYSYFYYMDG(SEQ ID NO: 564), NLGYSYFYYMDG (SEQ ID NO: 565), LGYSYFYYMDG (SEQ ID NO: 566), GYSYFYYMDG (SEQ ID NO: 567), YSYFYYMDG (SEQ ID NO: 568), or SYFYYMDG (SEQ ID NO: 569) ,

-wherein V2 comprises an amino acid sequence selected of WGKGTTVTVSS (SEQ ID NO: 574).

10. The antibody heavy chain variable region of claim 1, wherein the ultralong CDR3 comprises an amino acid sequence selected from the group consisting of:

(i) GSKHRLRDYFL (SEQ ID NO: 504) and YEEWGDYLATLDV (SEQ ID NO: 540);

(ii) GTDYTID (SEQ ID NO: 528) and GIRYQGSGTFWYFDV (SEQ ID NO: 556); and

(iii) DKGDSDYD (SEQ ID NO: 533) and GYSYFYYMDG (SEQ ID NO: 567).

11. The antibody heavy chain variable region of any one of claims 1-10, wherein the ultralong CDR3 is 35 amino acids in length or longer, 40 amino acids in length or longer, 45 amino acids in length or longer, 50 amino acids in length or longer, 55 amino acids in length or longer, or 60 amino acids in length or longer.

12. The antibody heavy chain variable region of claim 6, wherein the ultralong CDR3 is 35 amino acids in length or longer.

13. The antibody heavy chain variable region of claim 1-12, wherein the ultralong CDR3 comprises a cysteine motif.

14. The antibody heavy chain variable region of claim 1,

-wherein V1 comprises an amino acid sequence of QVQLVQSGAEVKKPGSSVKVSCKASGNSFSNHDVHWVRQAT GQGLEWMGWMSHEGDKTGLAQKFQGRVTITRDSGASTVYMELRGLTADDTAIY YCLT (SEQ ID NO: 496),

-wherein the ultralong CDR3 comprises an amino acid sequence of any one of GSKHRLRDYFLYNE (SEQ ID NO: 501), GSKHRLRDYFLYN (SEQ ID NO: 502), GSKHRLRDYFLY (SEQ ID NO: 503), GSKHRLRDYFL (SEQ ID NO: 504), GSKHRLRDYF (SEQ ID NO: 505), GSKHRLRDY (SEQ ID NO: 506), or GSKHRLRD (SEQ ID NO: 507), a cysteine motif, and an amino acid sequence of any one of YGPNYEEWGDYLATLDV (SEQ ID NO: 536), GPNYEEWGDYLATLDV (SEQ ID NO: 537), PNYEEWGDYLATLDV (SEQ ID NO: 538), NYEEWGDYLATLDV (SEQ ID NO: 539), YEEWGDYLATLDV (SEQ ID NO: 540), or EEWGDYLATLDV (SEQ ID NO: 541), and

-wherein V2 comprises an amino acid sequence of WGHGTAVTVSS (SEQ ID NO: 570).

15. The antibody heavy chain variable region of claim 1,

-wherein V1 comprises an amino acid sequence of QRLVESGGGVVQPGSSLRLSCAASGFDFSRQGMHWVRQAPG QGLEWVAFIKYDGSEKYHADSVWGRLSISRDNSKDTLYLQMNSLRVEDTATYFCV R (SEQ ID NO: 497),

-wherein the ultralong CDR3 comprises an amino acid sequence of any one of EAGGPDYRNGYNY (SEQ ID NO: 508), EAGGPDYRNGYN (SEQ ID NO: 509), EAGGPDYRNGY (SEQ ID NO: 510), EAGGPDYRNG (SEQ ID NO: 511), EAGGPDYRN (SEQ ID NO: 512), EAGGPDYR (SEQ ID NO: 513), EAGGPDY (SEQ ID NO: 514), or EAGGPD (SEQ ID NO: 515), a cysteine motif, and an amino acid sequence of any one of YDFYDGYYNYHYMDV (SEQ ID NO: 542), DFYDGYYNYHYMDV (SEQ ID NO: 543), FYDGYYNYHYMDV (SEQ ID NO: 544), YDGYYNYHYMDV (SEQ ID NO: 545), DGYYNYHYMDV (SEQ ID NO: 546), GYYNYHYMDV (SEQ ID NO: 547), or YYNYHYMDV (SEQ ID NO: 548),

-wherein V2 comprises an amino acid sequence selected of WGKGTTVTVSS (SEQ ID NO: 571).

16. The antibody heavy chain variable region of claim 1,

-wherein V1 comprises an amino acid sequence of QEQLVESGGGVVQPGGSLRLSCLASGFTFHKYGMHWVRQAP GKGLEWVALISDDGMRKYHSDSMWGRVTISRDNSKNTLYLQFSSLKVEDTAMFF CAR (SEQ ID NO: 498),

-wherein the ultralong CDR3 comprises an amino acid sequence of any one of EAGGPIWHDDVKY (SEQ ID NO: 516), EAGGPIWHDDVK (SEQ ID NO: 517), EAGGPIWHDDV (SEQ ID NO: 518), EAGGPIWHDD (SEQ ID NO: 519), EAGGPIWHD (SEQ ID NO: 520), EAGGPIWH (SEQ ID NO: 521), EAGGPIW (SEQ ID NO: 522), or EAGGPI (SEQ ID NO: 523), a cysteine motif, and an amino acid sequence of any one of YDFNDGYYNYHYMDV (SEQ ID NO: 549), DFYDGYYNYHYMDV (SEQ ID NO: 550), FYDGYYNYHYMDV (SEQ ID NO: 551), YDGYYNYHYMDV (SEQ ID NO: 552), DGYYNYHYMDV (SEQ ID NO: 553), or GYYNYHYMDV (SEQ ID NO: 554) ,

-wherein V2 comprises an amino acid sequence selected of WGKGTTVTVSS (SEQ ID NO: 572).

17. The antibody heavy chain variable region of claim 1,

- wherein V1 comprises an amino acid sequence of EVQLVESGGGLIRPGGSLRLSCKGSGFIFENFGFGWVRQGPGK GLEWVSGTNWNGGDSRYGDSVKGRFTISRDNSNNFVYLQMNSLRPEDTAIYYCA R (SEQ ID NO: 499),

-wherein the ultralong CDR3 comprises an amino acid sequence of any one of GTDYTIDDQGI (SEQ ID NO: 524), GTDYTIDDQG (SEQ ID NO: 525), GTDYTIDDQ (SEQ ID NO: 526), GTDYTIDD (SEQ ID NO: 527), GTDYTID (SEQ ID NO: 528), or GTDYTI (SEQ ID NO: 529), a cysteine motif, and an amino acid sequence of any one of QGIRYQGSGTFWYFDV (SEQ ID NO: 555), GIRYQGSGTFWYFDV (SEQ ID NO: 556), IRYQGSGTFWYFDV (SEQ ID NO: 557), RYQGSGTFWYFDV (SEQ ID NO: 558), YQGSGTFWYFDV (SEQ ID NO: 559), QGSGTFWYFDV (SEQ ID NO: 560), GSGTFWYFDV (SEQ ID NO: 561), SGTFWYFDV (SEQ ID NO: 562), or GTFWYFDV (SEQ ID NO: 563),

-wherein V2 comprises an amino acid sequence selected of WGRGTLVTVSS (SEQ ID NO: 573).

18. The antibody heavy chain variable region of claim 1,

-wherein V1 comprises an amino acid sequence of EVQLVESGGNVVQPGGSLRLSCTASGFSFDDSTMHWVRQAP GKGLQWVSLISWNGGRTYYADSVKGRFTISRDNSKNSLYLQMNSLKTEDTAFYFC AK (SEQ ID NO: 500),

-wherein the ultralong CDR3 comprises an amino acid sequence of any one of YNLGYSYFYYMDG(SEQ ID NO: 564), NLGYSYFYYMDG (SEQ ID NO: 565), LGYSYFYYMDG (SEQ ID NO: 566), GYSYFYYMDG (SEQ ID NO: 567), YSYFYYMDG (SEQ ID NO: 568), or SYFYYMDG (SEQ ID NO: 569), a cysteine motif, and an amino acid sequence of any one of YNLGYSYFYYMDG(SEQ ID NO: 564), NLGYSYFYYMDG (SEQ ID NO: 565), LGYSYFYYMDG (SEQ ID NO: 566), GYSYFYYMDG (SEQ ID NO: 567), YSYFYYMDG (SEQ ID NO: 568), or SYFYYMDG (SEQ ID NO: 569) ,

-wherein V2 comprises an amino acid sequence selected of WGKGTTVTVSS (SEQ ID NO: 574).

19. The antibody heavy chain variable region of any one of claim 13-18, wherein the cysteine motif is selected from the group consisting of: $CX_{10}CX_5CX_5CXCX_7C$ (SEQ ID NO: 41), $CX_{10}CX_6CX_5CXCX_{15}C$ (SEQ ID NO: 42), $CX_{11}CXCX_5C$ (SEQ ID NO: 43), $CX_{11}CX_5CX_5CXCX_7C$ (SEQ ID NO: 44), $CX_{10}CX_6CX_5CXCX_{13}C$ (SEQ ID NO: 45), $CX_{10}CX_5CXCX_4CX_8C$ (SEQ ID NO: 46), $CX_{10}CX_6CX_6CXCX_7C$ (SEQ ID NO: 47), $CX_{10}CX_4CX_7CXCX_8C$ (SEQ ID NO: 48), $CX_{10}CX_4CX_7CXCX_7C$ (SEQ ID NO: 49), $CX_{13}CX_8CX_8C$ (SEQ ID NO: 50), $CX_{10}CX_6CX_5CXCX_7C$ (SEQ ID NO: 51), $CX_{10}CX_5CX_5C$ (SEQ ID NO: 52), $CX_{10}CX_5CX_6CXCX_7C$ (SEQ ID NO: 53), $CX_{10}CX_6CX_5CX_7CX_9C$ (SEQ ID NO: 54), $CX_9CX_7CX_5CXCX_7C$ (SEQ ID NO: 55), $CX_{10}CX_6CX_5CXCX_9C$ (SEQ ID NO: 56), $CX_{10}CXCX_4CX_5CX_{11}C$ (SEQ ID NO: 57), $CX_7CX_3CX_6CX_5CXCX_5CX_{10}C$ (SEQ ID NO: 58), $CX_{10}CXCX_4CX_5CXCX_2CX_3C$ (SEQ ID NO: 59), $CX_{16}CX_5CXC$ (SEQ ID NO: 60), $CX_6CX_4CXCX_4CX_5C$ (SEQ ID NO: 61), $CX_{11}CX_4CX_5CX_6CX_3C$ (SEQ ID NO: 62), $CX_8CX_2CX_6CX_5C$ (SEQ ID NO: 63), $CX_{10}CX_5CX_5CXCX_{10}C$ (SEQ ID NO: 64), $CX_{10}CXCX_6CX_4CXC$ (SEQ ID NO: 65), $CX_{10}CX_5CX_5CXCX_2C$ (SEQ ID NO: 66), $CX_{14}CX_2CX_3CXCXC$ (SEQ ID NO: 67), $CX_{15}CX_5CXC$ (SEQ ID NO: 68), $CX_4CX_6CX_9CX_2CX_{11}C$ (SEQ ID NO: 69), $CX_6CX_4CX_5CX_5CX_{12}C$ (SEQ ID NO: 70), $CX_7CX_3CXCXCX_4CX_5CX_9C$ (SEQ ID NO:

71), $CX_{10}CX_6CX_5C$ (SEQ ID NO: 72), $CX_7CX_3CX_5CX_5CX_9C$ (SEQ ID NO: 73), $CX_7CX_5CXCX_2C$ (SEQ ID NO: 74), $CX_{10}CXCX_6C$ (SEQ ID NO: 75), $CX_{10}CX_3CX_3CX_5CX_7CXCX_6C$ (SEQ ID NO: 76), $CX_{10}CX_4CX_5CX_{12}CX_2C$ (SEQ ID NO: 77), $CX_{12}CX_4CX_5CXCXCX_9CX_3C$ (SEQ ID NO: 78), $CX_{12}CX_4CX_5CX_{12}CX_2C$ (SEQ ID NO: 79), $CX_{10}CX_6CX_5CXCX_{11}C$ (SEQ ID NO: 80), $CX_{16}CX_5CXCXCX_{14}C$ (SEQ ID NO: 81), $CX_{10}CX_5CXCX_8CX_6C$ (SEQ ID NO: 82), $CX_{12}CX_4CX_5CX_8CX_2C$ (SEQ ID NO: 83), $CX_{12}CX_5CX_5CXCX_8C$ (SEQ ID NO: 84), $CX_{10}CX_6CX_5CXCX_4CXCX_9C$ (SEQ ID NO: 85), $CX_{11}CX_4CX_5CX_8CX_2C$ (SEQ ID NO: 86), $CX_{10}CX_6CX_5CX_8CX_2C$ (SEQ ID NO: 87), $CX_{10}CX_6CX_5CXCX_8C$ (SEQ ID NO: 88), $CX_{10}CX_6CX_5CXCX_3CX_8CX_2C$ (SEQ ID NO: 89), $CX_{10}CX_6CX_5CX_3CX_8C$ (SEQ ID NO: 90), $CX_{10}CX_6CX_5CXCX_2CX_6CX_5C$ (SEQ ID NO: 91), $CX_7CX_6CX_3CX_3CX_9C$ (SEQ ID NO: 92), $CX_9CX_8CX_5CX_6CX_5C$ (SEQ ID NO: 93), $CX_{10}CX_2CX_2CX_7CXCX_{11}CX_5C$ (SEQ ID NO: 94), and $CX_{10}CX_6CX_5CXCX_2CX_8CX_4C$ (SEQ ID NO: 95).

20. The antibody heavy chain variable region of any one of claim 13-18, wherein the cysteine motif is selected from the group consisting of: $CCX_3CXCX_3CX_2CCXCX_5CX_9CX_5CXC$ (SEQ ID NO: 96), $CX_6CX_2CX_5CX_4CCXCX_4CX_6CXC$ (SEQ ID NO: 97), $CX_7CXCX_5CX_4CCCX_4CX_6CXC$ (SEQ ID NO: 98), $CX_9CX_3CXCX_2CXCCCX_6CX_4C$ (SEQ ID NO: 99), $CX_5CX_3CXCX_4CX_4CCX_{10}CX_2CC$ (SEQ ID NO: 100), $CX_5CXCX_1CXCX_3CCX_3CX_4CX_{10}C$ (SEQ ID NO: 101), $CX_9CCCX_3CX_4CCCX_5CX_6C$ (SEQ ID NO: 102), $CCX_8CX_5CX_4CX_3CX_4CCXCX_1C$ (SEQ ID NO: 103), $CCX_6CCX_5CCCX_4CX_4CX_{12}C$ (SEQ ID NO: 104), $CX_6CX_2CX_3CCCX_4CX_5CX_3CX_3C$ (SEQ ID NO: 105), $CX_3CX_5CX_6CX_4CCXCX_5CX_4CXC$ (SEQ ID NO: 106), $CX_4CX_4CCX_4CX_4CXCX_{11}CX_2CXC$ (SEQ ID NO: 107), $CX_5CX_2CCX_5CX_4CCX_3CCX_7C$ (SEQ ID NO: 108), $CX_5CX_5CX_3CX_2CXCCX_4CX_7CXC$ (SEQ ID NO: 109), $CX_3CX_7CX_3CX_4CCXCX_2CX_5CX_2C$ (SEQ ID NO: 110), $CX_9CX_3CXCX_4CCX_5CCCX_6C$ (SEQ ID NO: 111), $CX_9CX_3CXCX_2CXCCX_6CX_3CX_3C$ (SEQ ID NO: 112), $CX_8CCXCX_3CCX_3CXCX_3CX_4C$ (SEQ ID NO: 113), $CX_9CCX_4CX_2CXCCXCX_4CX_3C$ (SEQ ID NO: 114), $CX_{10}CXCX_3CX_2CXCCX_4CX_5CXC$ (SEQ ID NO: 115), $CX_9CXCX_3CX_2CXCCX_4CX_5CXC$ (SEQ ID NO: 116), $CX_6CCXCX_5CX_4CCXCX_5CX_2C$ (SEQ ID NO: 117), $CX_6CCXCX_3CXCCX_3CX_4CC$ (SEQ ID NO: 118), $CX_6CCXCX_3CXCX_2CXCX_4CX_8C$ (SEQ ID NO: 119), $CX_4CX_2CCX_3CXCX_4CCX_2CX_3C$ (SEQ ID NO: 120), $CX_3CX_5CX_3CCCX_4CX_9C$ (SEQ ID NO: 121), $CCX_9CX_3CXCCX_3CX_5C$ (SEQ ID NO: 122),

$CX_9CX_2CX_3CX_4CCX_4CX_5C$ (SEQ ID NO: 123), $CX_9CX_7CX_4CCXCX_7CX_3C$ (SEQ ID NO: 124), $CX_9CX_3CCCX_{10}CX_2CX_3C$ (SEQ ID NO: 125), $CX_3CX_5CX_5CX_4CCX_{10}CX_6C$ (SEQ ID NO: 126), $CX_9CX_5CX_4CCXCX_5CX_4C$ (SEQ ID NO: 127), $CX_7CXCX_6CX_4CCCX_{10}C$ (SEQ ID NO: 128), $CX_8CX_2CX_4CCX_4CX_3CX_3C$ (SEQ ID NO: 129), $CX_7CX_5CXCX_4CCX_7CX_4C$ (SEQ ID NO: 130), $CX_{11}CX_3CX_4CCCX_8CX_2C$ (SEQ ID NO: 131), $CX_2CX_3CX_4CCX_4CX_5CX_{15}C$ (SEQ ID NO: 132), $CX_9CX_5CX_4CCX_7C$ (SEQ ID NO: 133), $CX_9CX_7CX_3CX_2CX_6C$ (SEQ ID NO: 134), $CX_9CX_5CX_4CCX_{14}C$ (SEQ ID NO: 135), $CX_9CX_5CX_4CCX_8C$ (SEQ ID NO: 136), $CX_9CX_6CX_4CCXC$ (SEQ ID NO: 137), $CX_5CCX_7CX_4CX_{12}$ (SEQ ID NO: 138), $CX_{10}CX_3CX_4CCX_4C$ (SEQ ID NO: 139), $CX_9CX_4CCX_5CX_4C$ (SEQ ID NO: 140), $CX_{10}CX_3CX_4CX_7CXC$ (SEQ ID NO: 141), $CX_7CX_7CX_2CX_2CX_3C$ (SEQ ID NO: 142), $CX_9CX_4CX_4CCX_6C$ (SEQ ID NO: 143), $CX_7CXCX_3CXCX_6C$ (SEQ ID NO: 144), $CX_7CXCX_4CXCX_4C$ (SEQ ID NO: 145), $CX_9CX_5CX_4C$ (SEQ ID NO: 146), $CX_3CX_6CX_8C$ (SEQ ID NO: 147), $CX_{10}CXCX_4C$ (SEQ ID NO: 148), $CX_{10}CCX_4C$ (SEQ ID NO: 149), $CX_{15}C$ (SEQ ID NO: 150), $CX_{10}C$ (SEQ ID NO: 151), and CX_9C (SEQ ID NO: 152).

21. The antibody heavy chain variable region of any one of claim 1-20, wherein the ultralong CDR3 comprises 2 to 6 disulfide bonds.

22. The antibody heavy chain variable region of any one of claim 1-20, wherein the ultralong CDR3 comprises a non-antibody sequence.

23. The antibody heavy chain variable region of claim 1,

-wherein V1 comprises an amino acid sequence of QVQLVQSGAEVKKPGSSVKVSCKASGNSFSNHDVHWVRQAT GQGLEWMGWMSHEGDKTGLAQKFQGRVTITRDSGASTVYMELRGLTADDTAIY YCLT (SEQ ID NO: 496),

-wherein the ultralong CDR3 comprises an amino acid sequence of any one of GSKHRLRDYFLYNE (SEQ ID NO: 501), GSKHRLRDYFLYN (SEQ ID NO: 502), GSKHRLRDYFLY (SEQ ID NO: 503), GSKHRLRDYFL (SEQ ID NO: 504), GSKHRLRDYF (SEQ ID NO: 505), GSKHRLRDY (SEQ ID NO: 506), or GSKHRLRD (SEQ ID NO: 507), a non-antibody sequence, and an amino acid sequence of any one of YGPNYEEWGDYLATLDV (SEQ ID NO: 536), GPNYEEWGDYLATLDV (SEQ ID NO: 537), PNYEEWGDYLATLDV (SEQ ID NO: 538), NYEEWGDYLATLDV (SEQ ID NO: 539), YEEWGDYLATLDV (SEQ ID NO: 540), or EEWGDYLATLDV (SEQ ID NO: 541), and

-wherein V2 comprises an amino acid sequence of WGHGTAVTVSS (SEQ ID NO: 570).

24. The antibody heavy chain variable region of claim 1,

-wherein V1 comprises an amino acid sequence of QRLVESGGGVVQPGSSLRLSCAASGFDFSRQGMHWVRQAPG QGLEWVAFIKYDGSEKYHADSVWGRLSISRDNSKDTLYLQMNSLRVEDTATYFCV R (SEQ ID NO: 497),

-wherein the ultralong CDR3 comprises an amino acid sequence of any one of EAGGPDYRNGYNY (SEQ ID NO: 508), EAGGPDYRNGYN (SEQ ID NO: 509), EAGGPDYRNGY (SEQ ID NO: 510), EAGGPDYRNG (SEQ ID NO: 511), EAGGPDYRN (SEQ ID NO: 512), EAGGPDYR (SEQ ID NO: 513), EAGGPDY (SEQ ID NO: 514), or EAGGPD (SEQ ID NO: 515), a non-antibody sequence, and an amino acid sequence of any one of YDFYDGYYNYHYMDV (SEQ ID NO: 542), DFYDGYYNYHYMDV (SEQ ID NO: 543), FYDGYYNYHYMDV (SEQ ID NO: 544), YDGYYNYHYMDV (SEQ ID NO: 545), DGYYNYHYMDV (SEQ ID NO: 546), GYYNYHYMDV (SEQ ID NO: 547), or YYNYHYMDV (SEQ ID NO: 548),

-wherein V2 comprises an amino acid sequence selected of WGKGTTVTVSS (SEQ ID NO: 571).

25. The antibody heavy chain variable region of claim 1,

-wherein V1 comprises an amino acid sequence of QEQLVESGGGVVQPGGSLRLSCLASGFTFHKYGMHWVRQAP GKGLEWVALISDDGMRKYHSDSMWGRVTISRDNSKNTLYLQFSSLKVEDTAMFF CAR (SEQ ID NO: 498),

-wherein the ultralong CDR3 comprises an amino acid sequence of any one of EAGGPIWHDDVKY (SEQ ID NO: 516), EAGGPIWHDDVK (SEQ ID NO: 517), EAGGPIWHDDV (SEQ ID NO: 518), EAGGPIWHDD (SEQ ID NO: 519), EAGGPIWHD (SEQ ID NO: 520), EAGGPIWH (SEQ ID NO: 521), EAGGPIW (SEQ ID NO: 522), or EAGGPI (SEQ ID NO: 523), a non-antibody sequence, and an amino acid sequence of any one of YDFNDGYYNYHYMDV (SEQ ID NO: 549), DFYDGYYNYHYMDV (SEQ ID NO: 550), FYDGYYNYHYMDV (SEQ ID NO: 551), YDGYYNYHYMDV (SEQ ID NO: 552), DGYYNYHYMDV (SEQ ID NO: 553), or GYYNYHYMDV (SEQ ID NO: 554) ,

-wherein V2 comprises an amino acid sequence selected of WGKGTTVTVSS (SEQ ID NO: 572).

26. The antibody heavy chain variable region of claim 1,

- wherein V1 comprises an amino acid sequence of EVQLVESGGGLIRPGGSLRLSCKGSGFIFENFGFGWVRQGPGKGLEWVSGTNWNGGDSRYGDSVKGRFTISRDNSNNFVYLQMNSLRPEDTAIYYCAR (SEQ ID NO: 499),

-wherein the ultralong CDR3 comprises an amino acid sequence of any one of GTDYTIDDQGI (SEQ ID NO: 524), GTDYTIDDQG (SEQ ID NO: 525), GTDYTIDDQ (SEQ ID NO: 526), GTDYTIDD (SEQ ID NO: 527), GTDYTID (SEQ ID NO: 528), or GTDYTI (SEQ ID NO: 529), a non-antibody sequence, and an amino acid sequence of any one of QGIRYQGSGTFWYFDV (SEQ ID NO: 555), GIRYQGSGTFWYFDV (SEQ ID NO: 556), IRYQGSGTFWYFDV (SEQ ID NO: 557), RYQGSGTFWYFDV (SEQ ID NO: 558), YQGSGTFWYFDV (SEQ ID NO: 559), QGSGTFWYFDV (SEQ ID NO: 560), GSGTFWYFDV (SEQ ID NO: 561), SGTFWYFDV (SEQ ID NO: 562), or GTFWYFDV (SEQ ID NO: 563),

-wherein V2 comprises an amino acid sequence selected of WGRGTLVTVSS (SEQ ID NO: 573).

27. The antibody heavy chain variable region of claim 1,

-wherein V1 comprises an amino acid sequence of EVQLVESGGNVVQPGGSLRLSCTASGFSFDDSTMHWVRQAPGKGLQWVSLISWNGGRTYYADSVKGRFTISRDNSKNSLYLQMNSLKTEDTAFYFCAK (SEQ ID NO: 500),

-wherein the ultralong CDR3 comprises an amino acid sequence of any one of YNLGYSYFYYMDG(SEQ ID NO: 564), NLGYSYFYYMDG (SEQ ID NO: 565), LGYSYFYYMDG (SEQ ID NO: 566), GYSYFYYMDG (SEQ ID NO: 567), YSYFYYMDG (SEQ ID NO: 568), or SYFYYMDG (SEQ ID NO: 569), a non-antibody sequence, and an amino acid sequence of any one of YNLGYSYFYYMDG(SEQ ID NO: 564), NLGYSYFYYMDG (SEQ ID NO: 565), LGYSYFYYMDG (SEQ ID NO: 566), GYSYFYYMDG (SEQ ID NO: 567), YSYFYYMDG (SEQ ID NO: 568), or SYFYYMDG (SEQ ID NO: 569) ,

-wherein V2 comprises an amino acid sequence selected of WGKGTTVTVSS (SEQ ID NO: 574).

28. The antibody heavy chain variable region of claim 22, wherein the non-antibody sequence is a synthetic sequence.

29. The antibody heavy chain variable region of claim 22, wherein the non-antibody sequence is a cytokine sequence, a lymphokine sequence, a chemokine sequence, a growth factor sequence, a hormone sequence, or a toxin sequence.

30. The antibody heavy chain variable region of claim 22, wherein the non-antibody sequence is an IL-8 sequence, an IL-21 sequence, an SDF-1 (alpha) sequence, a somatostatin sequence, a chlorotoxin sequence, a Pro-TxII sequence, a ziconotide sequence, an ADWX-1 sequence, an HsTx1 sequence, an OSK1 sequence, a Pi2 sequence, a Hongotoxin (HgTX) sequence, a Margatoxin sequence, an Agitoxin-2 sequence, a Pi3 sequence, a Kaliotoxin sequence, an Anuroctoxin sequence, a Charybdotoxin sequence, a Tityustoxin -K- alpha sequence, a Maurotoxin sequence, a Ceratotoxin 1 (CcoTx1) sequence, a CcoTx2 sequence, a CcoTx3 sequence, a Phrixotoxin 3 (PaurTx3) sequence, a Hanatoxin 1 sequence, a Phrixotoxin 1 sequence, a Huwentoxin-IV sequence, an α-conotoxin Iml sequence, an α-conotoxin Epl sequence, an α-conotoxin PnIA sequence, an α-conotoxin PnIB sequence, an α-conotoxin MII sequence, an α-conotoxin AuIA sequence, an α-conotoxin AuIB sequence, an α-conotoxin AuIC sequence, a conotoxin κ-PVIIA sequence, a charybdotoxin sequence, a neurotoxin B-IV sequence, a crotamine sequence, a ω-GVIA (conotoxin) sequence, a κ-hefutoxin 1 sequence, a Css4 sequence, a Bj-xtrIT sequence, a BcIV sequence, a Hm-1 sequence, a Hm-2 sequence, a GsAF-I (β-theraphotoxin-Gr1b) sequence, a Protoxin I (ProTx-I sequence, a β-theraphotoxin-Tp1a) sequence, a Protoxin II (ProTx II) sequence, a Huwentoxin I sequence, a μ-Conotoxin PIIIA sequence, a Jingzhaotoxin-III (ß-TRTX-Cj1α) sequence, a GsAF-II (Kappa-theraphotoxin-Gr2c) sequence, a ShK (Stichodactyla toxin) sequence, a HsTx1 sequence, a Guangxitoxin 1E (GxTx-1E) sequence, a Maurotoxin sequence, a Charybdotoxin (ChTX) sequence, an Iberiotoxin (IbTx) sequence, a Leiurotoxin 1 (scyllatoxin) sequence, a Tamapin sequence, a Kaliotoxin-1 (KTX) sequence, a Purotoxin1 (PT-1) sequence, or a GpTx-1 sequence, a MOKA Toxin sequence, a OSK1 (P12, K16, D20) sequence, a OSK1 (K16, D20) sequence, a HmK sequence, a ShK (K16,Y26, K29) sequence, a ShK (K16) sequence, a ShK-A (K16) sequence, a ShK (K16,E30) sequence, a ShK (Q21) sequence, a ShK (L21) sequence, a ShK (F21) sequence, a ShK (I21) sequence, or a ShK (A21) sequence.

31. The antibody heavy chain variable region of claim 1 comprising the amino acid sequence of QVQLVQSGAEVKKPGSSVKVSCKASGNSFSNHDVHWVRQATGQGLEWMGWM SHEGDKTGLAQKFQGRVTITRDSGASTVYMELRGLTADDTAIYYCLTGSKHRLRD YFLYNEYGPNYEEWGDYLATLDVWGHGTAVTVSS (SEQ ID NO: 656),

wherein the heavy chain variable region further comprises a non-antibody sequence inserted between any one of:

(i) the L at position 109 and the Y at position 110;

(ii) the Y at position 110 and the N at position 111;

(iii) the N at position 111 and the E at position 112 ;

(iv) the E at position 112 and the Y at position 113;

(v) the Y at position 113 and the G at position 114;

(vi) the G at position 114 and the P at position 115;

(vii) the P at position 115 and the N at position 116; or

(viii) the N at position 116 and the Y at position 117, or

wherein the amino acid sequence of YNEYGPN at positions 110 to 116 has been removed and replaced with a non-antibody sequence, or

wherein the P at position 115 has been removed and replaced with a non-antibody sequence.

32. The antibody heavy chain variable region of claim 1 comprising the amino acid sequence of QRLVESGGGVVQPGSSLRLSCAASGFDFSRQGMHWVRQAPGQGLEWVAFIKYD GSEKYHADSVWGRLSISRDNSKDTLYLQMNSLRVEDTATYFCVREAGGPDYRNG YNYYDFYDGYYNYHYMDVWGKGTTVTVSS (SEQ ID NO: 657),

wherein the heavy chain variable region further comprises a non-antibody sequence inserted between any one of:

(i) the Y at position 104 and the R at position 105;

(ii) the R at position 105 and the N at position 106;

(iii) the N at position 106 and the G at position 107;

(iv) the G at position 107 and the Y at position 108;

(v) the Y at position 108 and the N at position 109;

(vi) the N at position 109 and the Y at position 110;

(vii) the Y at position 110 and the Y at position 111 ;

(viii) the Y at position 111 and the D at position 112;

(ix) the D at position 112 and the F at position 113

(x) the F at position 113 and the Y at position 114;

(xi) the Y at position 114 and the D at position 115;

(xi) the D at position 115 and the G at position 116, or

wherein the amino acid sequence of NYYD at positions 109 to 112 has been removed and replaced with a non-antibody sequence, or

wherein the Y at position 110 has been removed and replaced with a non-antibody sequence.

33. The antibody heavy chain variable region of claim 1 comprising the amino acid sequence of QEQLVESGGGVVQPGGSLRLSCLASGFTFHKYGMHWVRQAPGKGLEWVALISD DGMRKYHSDSMWGRVTISRDNSKNTLYLQFSSLKVEDTAMFFCAREAGGPIWHD DVKYYDFNDGYYNYHYMDVWGKGTTVTVSS (SEQ ID NO: 658),

wherein the heavy chain variable region further comprises a non-antibody sequence inserted between any on one of:

(i) the W at position 105 and the H at position 106;

(ii) the H at position 106 and the D at position 107;

(iii) the D at position 107 and the D at position 108;

(iv) the D at position 108 and the V at position 109; or

(v) the V at position 109 and the K at position 110, or

wherein the amino acid sequence of DD at positions 107 to 108 has been removed and replaced with a non-antibody sequence.

34. The antibody heavy chain variable region of claim 1 comprising the amino acid sequence of EVQLVESGGGLIRPGGSLRLSCKGSGFIFENFGFGWVRQGPGKGLEWVSGTNW

NGGDSRYGDSVKGRFTISRDNSNNFVYLQMNSLRPEDTAIYYCARGTDYTIDDQG IRYQGSGTFWYFDVWGRGTLVTVSS (SEQ ID NO: 659),

wherein the heavy chain variable region further comprises a non-antibody sequence inserted between any on one of:

(i) the I at position 104 and the D at position 105;

(ii) the D at position 105 and the D at position 106;

(iii) the D at position 106 and the Q at position 107;

(iv) the Q at position 107 and the G at position 108; or

(v) the G at position 108 and the I at position 109, or

wherein the amino acid sequence of DQ at positions 107 to 108 has been removed and replaced with a non-antibody sequence.

35. The antibody heavy chain variable region of claim 1 comprising the amino acid sequence of EVQLVESGGNVVQPGGSLRLSCTASGFSFDDSTMHWVRQAPGKGLQWVSLISW NGGRTYYADSVKGRFTISRDNSKNSLYLQMNSLKTEDTAFYFCAKDKGDSDYDY NLGYSYFYYMDGWGKGTTVTVSS (SEQ ID NO: 660),

wherein the heavy chain variable region further comprises a non-antibody sequence inserted between any on one of:

(i) the Y at position 105 and the D at position 106;

(ii) the D at position 106 and the Y at position 107;

(iii) the Y at position 107 and the N at position 108 ;

(iv) the N at position 108 and the L at position 109;

(v) the L at position 109 and the G at position 110; or

(vi) the G at position 110 and the Y at position 111, or

wherein the amino acid sequence of YNL at positions 107 to 109 has been removed and replaced with a non-antibody sequence.

36. The antibody heavy chain variable region of any one of claims 1-35, wherein the ultralong CDR3 comprises a linker sequence.

37. The antibody heavy chain variable region of claim 36, wherein the linker is linked to a N-terminus, a C-terminus, or both N-terminus and C-terminus of the non-antibody sequence.

38. The antibody heavy chain variable region of claim 36 or 37, wherein the linker comprises one or more amino acid sequence selected from the group consisting of SEQ ID NO: 575 to 598, 699 to 726 and 756-773, or any combination thereof.

39. The antibody heavy chain variable region of claim 37, wherein the linkers linked to both N-terminus and C-terminus have the same or different amino acid sequence.

40. An antibody or binding fragment thereof comprising the antibody heavy chain variable region of any one of claims 1-39.

41. The antibody or binding fragment thereof of claim 40, wherein the heavy chain variable region further comprises a constant heavy chain 1 (CH1) region.

42. The antibody or binding fragment thereof claim 39, wherein the CH1 region comprises an amino acid sequence selected from SEQ ID NO: 661 to 665.

43. The antibody or binding fragment thereof claim 40, wherein the heavy chain variable region further comprises an amino acid sequence of SEQ ID NO: 390.

44. The antibody or binding fragment thereof of claim 40, wherein the antibody or binding fragment further comprises a light chain variable region.

45. The antibody or binding fragment thereof of claim 44, wherein the light chain variable region comprises an amino acid sequence selected from SEQ ID NO: 734 to 738.

46. The antibody or binding fragment thereof of claim 45, wherein the light chain variable region further comprising a constant light chain (CL) region.

47. The antibody or binding fragment thereof of claim 45, wherein the CL region comprises an amino acid sequence selected from SEQ ID NO: 739 to 743.

48. An isolated polynucleotide encoding the antibody heavy chain variable region of any one of claims 1-39.

49. A vector comprising the polynucleotide of claim 48.

50. A host cell comprising the vector of claim 49.

51. A nucleic acid library comprising a plurality of polynucleotides comprising nucleic acid sequences encoding for an antibody heavy chain variable region comprising a

sequence of the formula V1-X-V2, wherein V1 comprises an amino acid sequence selected from the group consisting of:

(i) QVQLVQSGAEVKKPGSSVKVSCKASGNSFSNHDVHWVRQATGQGLEWMGW MSHEGDKTGLAQKFQGRVTITRDSGASTVYMELRGLTADDTAIYYCLT (SEQ ID NO: 496),

(ii) QRLVESGGGVVQPGSSLRLSCAASGFDFSRQGMHWVRQAPGQGLEWVAFIK YDGSEKYHADSVWGRLSISRDNSKDTLYLQMNSLRVEDTATYFCVR (SEQ ID NO: 497),

(iii) QEQLVESGGGVVQPGGSLRLSCLASGFTFHKYGMHWVRQAPGKGLEWVALI SDDGMRKYHSDSMWGRVTISRDNSKNTLYLQFSSLKVEDTAMFFCAR (SEQ ID NO: 498),

(iv) EVQLVESGGGLIRPGGSLRLSCKGSGFIFENFGFGWVRQGPGKGLEWVSGT NWNGGDSRYGDSVKGRFTISRDNSNNFVYLQMNSLRPEDTAIYYCAR (SEQ ID NO: 499), and

(v) EVQLVESGGNVVQPGGSLRLSCTASGFSFDDSTMHWVRQAPGKGLQWVSLI SWNGGRTYYADSVKGRFTISRDNSKNSLYLQMNSLKTEDTAFYFCAK (SEQ ID NO: 500);

wherein X comprises an ultralong CDR3; and

wherein V2 comprises an amino acid sequence selected from the group consisting of:

(i) WGHGTAVTVSS (SEQ ID NO: 570),

(ii) WGKGTTVTVSS (SEQ ID NO: 571),

(iii) WGKGTTVTVSS (SEQ ID NO: 572),

(iv) WGRGTLVTVSS (SEQ ID NO: 573), and

(v) WGKGTTVTVSS (SEQ ID NO: 574).

52. A library of antibodies comprising antibody heavy chain variable regions comprising a sequence of the formula V1-X-V2, wherein V1 comprises an amino acid sequence selected from the group consisting of:

(i) QVQLVQSGAEVKKPGSSVKVSCKASGNSFSNHDVHWVRQATGQGLEWMGW MSHEGDKTGLAQKFQGRVTITRDSGASTVYMELRGLTADDTAIYYCLT (SEQ ID NO: 496),

(ii) QRLVESGGGVVQPGSSLRLSCAASGFDFSRQGMHWVRQAPGQGLEWVAFIK YDGSEKYHADSVWGRLSISRDNSKDTLYLQMNSLRVEDTATYFCVR (SEQ ID NO: 497),

(iii) QEQLVESGGGVVQPGGSLRLSCLASGFTFHKYGMHWVRQAPGKGLEWVALI SDDGMRKYHSDSMWGRVTISRDNSKNTLYLQFSSLKVEDTAMFFCAR (SEQ ID NO: 498),

(iv) EVQLVESGGGLIRPGGSLRLSCKGSGFIFENFGFGWVRQGPGKGLEWVSGT NWNGGDSRYGDSVKGRFTISRDNSNNFVYLQMNSLRPEDTAIYYCAR (SEQ ID NO: 499), and

(v) EVQLVESGGNVVQPGGSLRLSCTASGFSFDDSTMHWVRQAPGKGLQWVSLI SWNGGRTYYADSVKGRFTISRDNSKNSLYLQMNSLKTEDTAFYFCAK (SEQ ID NO: 500);

wherein X comprises an ultralong CDR3; and

wherein V2 comprises an amino acid sequence selected from the group consisting of:

(i) WGHGTAVTVSS (SEQ ID NO: 570),

(ii) WGKGTTVTVSS (SEQ ID NO: 571),

(iii) WGKGTTVTVSS (SEQ ID NO: 572),

(iv) WGRGTLVTVSS (SEQ ID NO: 573), and

(v) WGKGTTVTVSS (SEQ ID NO: 574).

53. The antibody heavy chain variable region of claim 1, wherein the ultralong CDR3 comprises a $X^1X^2X^3X^4X^5$ motif,

wherein X^1 is threonine (T), glycine (G), alanine (A), serine (S), or valine (V), wherein X^2 is serine (S), threonine (T), proline (P), isoleucine (I), alanine (A), valine (V), or asparagine (N), wherein X^3 is valine (V), alanine (A), threonine (T), or aspartic acid (D), wherein X^4 is histidine (H), threonine (T), arginine (R), tyrosine (Y), phenylalanine (F), or leucine (L), and wherein X^5 is glutamine (Q).

54. The antibody heavy chain variable region of claim 53, wherein the $X^1X^2X^3X^4X^5$ motif is TTVHQ (SEQ ID NO: 153), TSVHQ (SEQ ID NO: 154), SSVTQ (SEQ ID NO: 155), STVHQ (SEQ ID NO: 156), ATVRQ (SEQ ID NO: 157), TTVYQ (SEQ ID NO: 158), SPVHQ (SEQ ID NO: 159), ATVYQ (SEQ ID NO: 160), TAVYQ (SEQ ID NO: 161), TNVHQ (SEQ ID NO: 162), ATVHQ (SEQ ID NO: 163), STVYQ (SEQ ID NO:

164), TIVHQ (SEQ ID NO: 165), AIVYQ (SEQ ID NO: 166), TTVFQ (SEQ ID NO: 167), AAVFQ (SEQ ID NO: 168), GTVHQ (SEQ ID NO: 169), ASVHQ (SEQ ID NO: 170), TAVFQ (SEQ ID NO: 171), ATVFQ (SEQ ID NO: 172), AAAHQ (SEQ ID NO: 173), VVVYQ (SEQ ID NO: 174), GTVFQ (SEQ ID NO: 175), TAVHQ (SEQ ID NO: 176), ITVHQ (SEQ ID NO: 177), ITAHQ (SEQ ID NO: 178), VTVHQ (SEQ ID NO: 179); AAVHQ (SEQ ID NO: 180), GTVYQ (SEQ ID NO: 181), TTVLQ (SEQ ID NO: 182), TTTHQ (SEQ ID NO: 183), or TTDYQ (SEQ ID NO: 184).

55. The antibody heavy chain variable region of claim 1, wherein the ultralong CDR3 comprises a $(X^aX^b)_z$ motif,

wherein X^a is any amino acid residue, X^b is an aromatic amino acid selected from the group consisting of: tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H), and wherein z is 1 - 4.

56. The antibody heavy chain variable region of claim 55, wherein the $(X^aX^b)_z$ motif is CYTYNYEF (SEQ ID NO: 217), HYTYTYDF (SEQ ID NO: 218), HYTYTYEW (SEQ ID NO: 219), KHRYTYEW (SEQ ID NO: 220), NYIYKYSF (SEQ ID NO: 221), PYIYTYQF (SEQ ID NO: 222), SFTYTYEW (SEQ ID NO: 223), SYIYIYQW (SEQ ID NO: 224), SYNYTYSW (SEQ ID NO: 225), SYSYSYEY (SEQ ID NO: 226), SYTYNYDF (SEQ ID NO: 227), SYTYNYEW (SEQ ID NO: 228), SYTYNYQF (SEQ ID NO: 229), SYVWTHNF (SEQ ID NO: 230), TYKYVYEW (SEQ ID NO: 231), TYTYTYEF (SEQ ID NO: 232), TYTYTYEW (SEQ ID NO: 233), VFTYTYEF (SEQ ID NO: 234), AYTYEW (SEQ ID NO: 235), DYIYTY (SEQ ID NO: 236), IHSYEF (SEQ ID NO: 237), SFTYEF (SEQ ID NO: 238), SHSYEF (SEQ ID NO: 239), THTYEF (SEQ ID NO: 240), TWTYEF (SEQ ID NO: 241), TYNYEW (SEQ ID NO: 242), TYSYEF (SEQ ID NO: 243), TYSYEH (SEQ ID NO: 244), TYTYDF (SEQ ID NO: 245), TYTYEF (SEQ ID NO: 246), TYTYEW (SEQ ID NO: 247), AYEF (SEQ ID NO: 248), AYSF (SEQ ID NO: 249), AYSY (SEQ ID NO: 250), CYSF (SEQ ID NO: 251), DYTY (SEQ ID NO: 252), KYEH (SEQ ID NO: 253), KYEW (SEQ ID NO: 254), MYEF (SEQ ID NO: 255), NWIY (SEQ ID NO: 256), NYDY (SEQ ID NO: 257), NYQW (SEQ ID NO: 258), NYSF (SEQ ID NO: 259), PYEW (SEQ ID NO: 260), RYNW (SEQ ID NO: 261), RYTY (SEQ ID NO: 262), SYEF (SEQ ID NO: 263), SYEH (SEQ ID NO: 264), SYEW (SEQ ID NO: 265), SYKW (SEQ ID NO: 266), SYTY (SEQ ID NO: 267), TYDF (SEQ ID NO: 268), TYEF (SEQ ID NO: 269), TYEW (SEQ ID NO: 270), TYQW (SEQ ID NO: 271), TYTY (SEQ ID NO: 272), or VYEW (SEQ ID NO: 273).

57. The antibody heavy chain variable region of claim 56, wherein the $(X^aX^b)_z$ motif is YXYXYX.

58. The antibody heavy chain variable region of claim 1, wherein the ultralong CDR3 comprises a $X^1X^2X^3X^4X^5X_n$ motif,

wherein X_1 is threonine (T), glycine (G), alanine (A), serine (S), or valine (V), wherein X_2 is serine (S), threonine (T), proline (P), isoleucine (I), alanine (A), valine (V), or asparagine (N), wherein X_3 is valine (V), alanine (A), threonine (T), or aspartic acid (D), wherein X_4 is histidine (H), threonine (T), arginine (R), tyrosine (Y), phenylalanine (F), or leucine (L), wherein X_5 is glutamine (Q), and wherein n is 27 – 54.

59. The antibody heavy chain variable region of claim 1, wherein the ultralong CDR3 comprises $X_n(X^aX^b)_z$ motif,

wherein X^a is any amino acid residue, X^b is an aromatic amino acid selected from the group consisting of: tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H), wherein n is 27 – 54, and wherein z is 1 - 4.

60. The antibody heavy chain variable region of claim 1, wherein the ultralong CDR3 comprises a $X^1X^2X^3X^4X^5X_n(X^aX^b)_z$ motif,

wherein X^1 is threonine (T), glycine (G), alanine (A), serine (S), or valine (V), wherein X^2 is serine (S), threonine (T), proline (P), isoleucine (I), alanine (A), valine (V), or asparagine (N), wherein X^3 is valine (V), alanine (A), threonine (T), or aspartic acid (D), wherein X^4 is histidine (H), threonine (T), arginine (R), tyrosine (Y), phenylalanine (F), or leucine (L), and wherein X^5 is glutamine (Q), wherein X^a is any amino acid residue, X^b is an aromatic amino acid selected from the group consisting of: tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H), wherein n is 27 – 54, and wherein z is 1 - 4.

61. An antibody heavy chain variable region comprising a sequence of the formula V1-X, wherein V1 comprises an amino acid sequence selected from the group consisting of:

(i) QVQLVQSGAEVKKPGSSVKVSCKASGNSFSNHDVHWVRQATGQGLEWMGW MSHEGDKTGLAQKFQGRVTITRDSGASTVYMELRGLTADDTAIYYCLT (SEQ ID NO: 496),

(ii) QRLVESGGGVVQPGSSLRLSCAASGFDFSRQGMHWVRQAPGQGLEWVAFIK YDGSEKYHADSVWGRLSISRDNSKDTLYLQMNSLRVEDTATYFCVR (SEQ ID NO: 497),

(iii) QEQLVESGGGVVQPGGSLRLSCLASGFTFHKYGMHWVRQAPGKGLEWVALI SDDGMRKYHSDSMWGRVTISRDNSKNTLYLQFSSLKVEDTAMFFCAR (SEQ ID NO: 498),

(iv) EVQLVESGGGLIRPGGSLRLSCKGSGFIFENFGFGWVRQGPGKGLEWVSGT NWNGGDSRYGDSVKGRFTISRDNSNNFVYLQMNSLRPEDTAIYYCAR (SEQ ID NO: 499),

(v) EVQLVESGGNVVQPGGSLRLSCTASGFSFDDSTMHWVRQAPGKGLQWVSLI SWNGGRTYYADSVKGRFTISRDNSKNSLYLQMNSLKTEDTAFYFCAK (SEQ ID NO: 500);

(vi) QVQLVQSGAEVKKPGSSVKVSCKASGNSFSNHDVHWVRQATGQGLEWMG WMSHEGDKTGLAQKFQGRVTITRDSGASTVYMELRGLTADDTAIYYC (SEQ ID NO: 744),

(vii) QRLVESGGGVVQPGSSLRLSCAASGFDFSRQGMHWVRQAPGQGLEWVAFI KYDGSEKYHADSVWGRLSISRDNSKDTLYLQMNSLRVEDTATYFC (SEQ ID NO: 745),

(viii) QEQLVESGGGVVQPGGSLRLSCLASGFTFHKYGMHWVRQAPGKGLEWVAL ISDDGMRKYHSDSMWGRVTISRDNSKNTLYLQFSSLKVEDTAMFFC (SEQ ID NO: 746),

(ix) EVQLVESGGGLIRPGGSLRLSCKGSGFIFENFGFGWVRQGPGKGLEWVSGT NWNGGDSRYGDSVKGRFTISRDNSNNFVYLQMNSLRPEDTAIYYC (SEQ ID NO: 747),

(x) EVQLVESGGNVVQPGGSLRLSCTASGFSFDDSTMHWVRQAPGKGLQWVSLI SWNGGRTYYADSVKGRFTISRDNSKNSLYLQMNSLKTEDTAFYFC (SEQ ID NO: 748),

(xi) QVQLVQSGAEVKKPGSSVKVSCKASGNSFSNHDVHWVRQATGQGLEWMG WMSHEGDKTGLAQKFQGRVTITRDSGASTVYMELRGLTADDTAIYY (SEQ ID NO: 749),

(xii) QRLVESGGGVVQPGSSLRLSCAASGFDFSRQGMHWVRQAPGQGLEWVAFI KYDGSEKYHADSVWGRLSISRDNSKDTLYLQMNSLRVEDTATYF (SEQ ID NO: 750),

(xiii) QEQLVESGGGVVQPGGSLRLSCLASGFTFHKYGMHWVRQAPGKGLEWVAL ISDDGMRKYHSDSMWGRVTISRDNSKNTLYLQFSSLKVEDTAMFF (SEQ ID NO: 751),

(xiv) EVQLVESGGGLIRPGGSLRLSCKGSGFIFENFGFGWVRQGPGKGLEWVSGT NWNGGDSRYGDSVKGRFTISRDNSNNFVYLQMNSLRPEDTAIYY (SEQ ID NO: 752), and

(xv) EVQLVESGGNVVQPGGSLRLSCTASGFSFDDSTMHWVRQAPGKGLQWVSL ISWNGGRTYYADSVKGRFTISRDNSKNSLYLQMNSLKTEDTAFYF (SEQ ID NO: 753); and

wherein X comprises an ultralong CDR3.

62. The antibody heavy chain variable region of claim 61, wherein the ultralong CDR3 comprises a $X^1X^2X^3X^4X^5$ motif,

wherein X^1 is threonine (T), glycine (G), alanine (A), serine (S), or valine (V), wherein X^2 is serine (S), threonine (T), proline (P), isoleucine (I), alanine (A), valine (V), or asparagine (N), wherein X^3 is valine (V), alanine (A), threonine (T), or aspartic acid (D), wherein X^4 is histidine (H), threonine (T), arginine (R), tyrosine (Y), phenylalanine (F), or leucine (L), and wherein X^5 is glutamine (Q).

63. The antibody heavy chain variable region of claim 62, wherein the $X^1X^2X^3X^4X^5$ motif is TTVHQ (SEQ ID NO: 153), TSVHQ (SEQ ID NO: 154), SSVTQ (SEQ ID NO: 155), STVHQ (SEQ ID NO: 156), ATVRQ (SEQ ID NO: 157), TTVYQ (SEQ ID NO: 158), SPVHQ (SEQ ID NO: 159), ATVYQ (SEQ ID NO: 160), TAVYQ (SEQ ID NO: 161), TNVHQ (SEQ ID NO: 162), ATVHQ (SEQ ID NO: 163), STVYQ (SEQ ID NO: 164), TIVHQ (SEQ ID NO: 165), AIVYQ (SEQ ID NO: 166), TTVFQ (SEQ ID NO: 167), AAVFQ (SEQ ID NO: 168), GTVHQ (SEQ ID NO: 169), ASVHQ (SEQ ID NO: 170), TAVFQ (SEQ ID NO: 171), ATVFQ (SEQ ID NO: 172), AAAHQ (SEQ ID NO: 173), VVVYQ (SEQ ID NO: 174), GTVFQ (SEQ ID NO: 175), TAVHQ (SEQ ID NO: 176), ITVHQ (SEQ ID NO: 177), ITAHQ (SEQ ID NO: 178), VTVHQ (SEQ ID NO: 179); AAVHQ (SEQ ID NO: 180), GTVYQ (SEQ ID NO: 181), TTVLQ (SEQ ID NO: 182), TTTHQ (SEQ ID NO: 183), or TTDYQ (SEQ ID NO: 184).

64. The antibody heavy chain variable region of claim 61, wherein the ultralong CDR3 comprises a $CX^1X^2X^3X^4X^5$ motif.

65. The antibody heavy chain variable region of claim 64, wherein the $CX^1X^2X^3X^4X^5$ motif is CTTVHQ (SEQ ID NO: 185), CTSVHQ (SEQ ID NO: 186), CSSVTQ (SEQ

ID NO: 187), CSTVHQ (SEQ ID NO: 188), CATVRQ (SEQ ID NO: 189), CTTVYQ (SEQ ID NO: 190), CSPVHQ (SEQ ID NO: 191), CATVYQ (SEQ ID NO: 192), CTAVYQ (SEQ ID NO: 193), CTNVHQ (SEQ ID NO: 194), CATVHQ (SEQ ID NO: 195), CSTVYQ (SEQ ID NO: 196), CTIVHQ (SEQ ID NO: 197), CAIVYQ (SEQ ID NO: 198), CTTVFQ (SEQ ID NO: 199), CAAVFQ (SEQ ID NO: 200), CGTVHQ (SEQ ID NO: 201), CASVHQ (SEQ ID NO: 202), CTAVFQ (SEQ ID NO: 203), CATVFQ (SEQ ID NO: 204), CAAAHQ (SEQ ID NO: 205), CVVVYQ (SEQ ID NO: 206), CGTVFQ (SEQ ID NO: 207), CTAVHQ (SEQ ID NO: 208), CITVHQ (SEQ ID NO: 209), CITAHQ (SEQ ID NO: 210), CVTVHQ (SEQ ID NO: 211); CAAVHQ (SEQ ID NO: 212), CGTVYQ (SEQ ID NO: 213), CTTVLQ (SEQ ID NO: 214), CTTTHQ (SEQ ID NO: 215), or CTTDYQ (SEQ ID NO: 216).

66. The antibody heavy chain variable region of claim 61, wherein the ultralong CDR3 comprises a $(X^aX^b)_z$ motif,

wherein X^a is any amino acid residue, X^b is an aromatic amino acid selected from the group consisting of: tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H), and wherein z is 1 - 4.

67. The antibody heavy chain variable region of claim 66, wherein the $(X^aX^b)_z$ motif is CYTYNYEF (SEQ ID NO: 217), HYTYTYDF (SEQ ID NO: 218), HYTYTYEW (SEQ ID NO: 219), KHRYTYEW (SEQ ID NO: 220), NYIYKYSF (SEQ ID NO: 221), PYIYTYQF (SEQ ID NO: 222), SFTYTYEW (SEQ ID NO: 223), SYIYIYQW (SEQ ID NO: 224), SYNYTYSW (SEQ ID NO: 225), SYSYSYEY (SEQ ID NO: 226), SYTYNYDF (SEQ ID NO: 227), SYTYNYEW (SEQ ID NO: 228), SYTYNYQF (SEQ ID NO: 229), SYVWTHNF (SEQ ID NO: 230), TYKYVYEW (SEQ ID NO: 231), TYTYTYEF (SEQ ID NO: 232), TYTYTYEW (SEQ ID NO: 233), VFTYTYEF (SEQ ID NO: 234), AYTYEW (SEQ ID NO: 235), DYIYTY (SEQ ID NO: 236), IHSYEF (SEQ ID NO: 237), SFTYEF (SEQ ID NO: 238), SHSYEF (SEQ ID NO: 239), THTYEF (SEQ ID NO: 240), TWTYEF (SEQ ID NO: 241), TYNYEW (SEQ ID NO: 242), TYSYEF (SEQ ID NO: 243), TYSYEH (SEQ ID NO: 244), TYTYDF (SEQ ID NO: 245), TYTYEF (SEQ ID NO: 246), TYTYEW (SEQ ID NO: 247), AYEF (SEQ ID NO: 248), AYSF (SEQ ID NO: 249), AYSY (SEQ ID NO: 250), CYSF (SEQ ID NO: 251), DYTY (SEQ ID NO: 252), KYEH (SEQ ID NO: 253), KYEW (SEQ ID NO: 254), MYEF (SEQ ID NO: 255), NWIY (SEQ ID NO: 256), NYDY (SEQ ID NO: 257), NYQW (SEQ ID NO: 258), NYSF (SEQ ID NO: 259), PYEW (SEQ ID NO: 260), RYNW (SEQ ID NO: 261), RYTY (SEQ ID NO: 262), SYEF (SEQ ID NO: 263), SYEH (SEQ ID NO: 264), SYEW (SEQ ID NO: 265), SYKW (SEQ ID NO:

266), SYTY (SEQ ID NO: 267), TYDF (SEQ ID NO: 268), TYEF (SEQ ID NO: 269), TYEW (SEQ ID NO: 270), TYQW (SEQ ID NO: 271), TYTY (SEQ ID NO: 272), or VYEW (SEQ ID NO: 273).

68. The antibody heavy chain variable region of claim 66, wherein the $(X^aX^b)_z$ motif is YXYXYX.

69. The antibody heavy chain variable region of claim 61, wherein the ultralong CDR3 comprises a $X^1X^2X^3X^4X^5X_n$ motif,

wherein X_1 is threonine (T), glycine (G), alanine (A), serine (S), or valine (V), wherein X_2 is serine (S), threonine (T), proline (P), isoleucine (I), alanine (A), valine (V), or asparagine (N), wherein X_3 is valine (V), alanine (A), threonine (T), or aspartic acid (D), wherein X_4 is histidine (H), threonine (T), arginine (R), tyrosine (Y), phenylalanine (F), or leucine (L), wherein X_5 is glutamine (Q), and wherein n is 27 – 54.

70. The antibody heavy chain variable region of claim 61, wherein the ultralong CDR3 comprises $X_n(X^aX^b)_z$ motif,

wherein X^a is any amino acid residue, X^b is an aromatic amino acid selected from the group consisting of: tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H), wherein n is 27 – 54, and wherein z is 1 - 4.

71. The antibody heavy chain variable region of claim 61, wherein the ultralong CDR3 comprises a $X^1X^2X^3X^4X^5X_n(X^aX^b)_z$ motif,

wherein X^1 is threonine (T), glycine (G), alanine (A), serine (S), or valine (V), wherein X^2 is serine (S), threonine (T), proline (P), isoleucine (I), alanine (A), valine (V), or asparagine (N), wherein X^3 is valine (V), alanine (A), threonine (T), or aspartic acid (D), wherein X^4 is histidine (H), threonine (T), arginine (R), tyrosine (Y), phenylalanine (F), or leucine (L), and wherein X^5 is glutamine (Q), wherein X^a is any amino acid residue, X^b is an aromatic amino acid selected from the group consisting of: tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H), wherein n is 27 – 54, and wherein z is 1 - 4.

Figure 1

BLV1H12	CTSVHQ	ETMKYQ	SCPDGYRERSDCSNRFACGFSDCCRVSVFGNCLTTLPVSYSYTYNYEW	HVDVWGGGLLVTVSS	(SEQ ID NO : 360)
BLV5B8	CTTVHQ	ETRKT	CSDGYIAVDSCGRGQSDGCVNDCNSCYYGWRNCRFQPAIHSYEF	HVDAWGRGLLVTVSS	(SEQ ID NO : 361)
BLV5D3	CSSVTQ	RTHVSR	SCPDGCSDGEDGCVDGCCCSAYRCYTFGVRDLSCTSYSITTYTYEW	NVDAWGRGLLVTVSS	(SEQ ID NO : 362)
BLV8C11	CTTVHQ	KTTRKT	CCSTAYRYDSGCGSGCDCCGADCYVFGACTFGLDSSYSYIYIYQW	YVDAWGQGLLVTVSS	(SEQ ID NO : 363)
BF4E9	CTPVHQ	IF	CPDGYSYGYGCGYGYGCSGYDCYGYGGYGGYSSYSYSYSYEY	YGDAWGQGLLVTVSS	(SEQ ID NO : 364)
BF1H1	CTPVHP		SPDGYSYGYGCGYGYGCSGYDCYGYGGYGGYSSYSYSYS		(SEQ ID NO : 365)
F18	CTTVHQ	IR	CPDGYGYGYGCGYGSYGYSGYDCYGYGGYGGYGGYSSYS		(SEQ ID NO : 366)

Figure 2A



V_H Germ	CTTVHQ				(SEQ ID NO : 185)
D_H2 Germ		SCPDGYSYGYGCGYGYGCSGYDCYGYGGYGGYGGYGYSSYSYSYTYEY			(SEQ ID NO : 274)
J_H1 Germ			YVDAWGQGLLVTVSS		(SEQ ID NO : 275)
	Published sequences:			**Length**	
BLV1H12	CTSVHQ ETKKYQ	SCPDGYRERSDCSNRPACGTSDCCRVSVFGNCLTTLPVSYSYTYNYEW	HVDVW	61	(SEQ ID NO : 276)
BLV5B8	CTTVHQ ETRAT	CSDGYIAVDSCGRGQSDGCVNDCNSCYYGWRNCRRQPAIHSYEF	HVDAW	56	(SEQ ID NO : 277)
BLV5D3	CSSVTQ RTHVSR	SCPDGCSDGDGCVDGCCCSAYRCYTFGVRDLSCTSYSITYTYEW	NVDAW	57	(SEQ ID NO : 278)
BLV8C11	CTTVHQ KTTRKT	CCSDAYRYDSGCGSGCDCCGADCYVFGACTFGLDSSYSYIYIYQW	YVDAW	58	(SEQ ID NO : 279)
BF4E9	CTTVHQ IF	CPDGYSYGYGCGYGYGCSGYDCYGYGGYGYGGYGGYSSYSYSYSYEY	YGDAW	56	(SEQ ID NO : 280)
B-L1	CSTVHQ KTRTTQGNT	CPDGYTLKDDCPRCRGGCDGYDCCWGDACRSSGLCWGHNPLVTETYTYEF	YIDAW	66	(SEQ ID NO : 281)
B-L2	CATVRQ TTLRD	CPGGYTEDRSCVNTYSCGADDCCGRGDVGYPALYGYRCAAHIQRYNW	HADAW	59	(SEQ ID NO : 282)
	Longest CDR H3s:				
UL-1	CSTVHQ KTRTTQGEY	LSLMVTLLKDDCPRCRGGCDGYDCCWGDACRSSGLCWGHNPLVTETYTYEF	YIDAW	67	(SEQ ID NO : 283)
UL-2	CSTVHQ KERTTQGNN	LSLMVTLLKDDCPRCRGGCDGYDCCWGDACRSSGLCWGHNPLVTETYTYEF	YIDAW	67	(SEQ ID NO : 284)
UL-3	CSTVHQ ATRTTQGN	TCPDGYTLKDDCPRCRGGCDGYDCCWGDACRSSGLCWGHNPLVTETYTYEF	YIDAW	66*	(SEQ ID NO : 285)
UL-4	CSTVHQ KTRTTQGN	TCPDGYTFKDDCPRCRGGCDGYDCCWGDACRSSGLCWGHNPLVTETYTYEF	YIDAW	66	(SEQ ID NO : 286)
UL-5	CTTVHQ KERTTQGN	TCPDGYTLKDDCPRCRGGCDGYDCCWGDACRSSGLCWGHNPLVTETYTYEF	YIDAW	66	(SEQ ID NO : 287)
UL-6	CSTVHQ ATRTTQGN	TCPDGYTLKNDCPRCRGGCDGYDCCWGDACRSSGLCWGHNPLVTETYTYEF	YIDAW	66	(SEQ ID NO : 288)
UK-7	CTTVYQ KTRTTQGN	TCPDGYTLKDDCPRCRGGCDGYDCCWGDACRSSGLCWGHNPLVTETYTYEF	YIDAW	66	(SEQ ID NO : 289)
UL-8	CSTVHQ KPGQHKGI	LVLMVTLLKDDCPRCRGGCDGYDCCWGDACRSSGLCWGHNPLVTETYTYEF	YIDAW	66	(SEQ ID NO : 290)
UL-9	CSTVHQ KTRTTQGI	LVLMVTLLKDDCPRCRGGCDGYDCCWGDACRSSGLCWGHNPLVTETYTYEF	YIDAW	66	(SEQ ID NO : 291)
	12 Cysteines:				
UL-10	CSPVHQ EIRK	CCPAGCQCGRSCGACCGCAGDEFCGINVYGYVTCGGYRTCSCIDTYDF	YVDAW	59	(SEQ ID NO : 292)
UL-11	CSFVHQ QTRK	CCPAGCQCGRSCGACCGCAGDEFCGINVYGYITCGGYRTCSCIDTYDF	YVEAW	59	(SEQ ID NO : 293)
	10 Cysteines:				
UL-12	CATVYQ KTNQSK	NCPEGSAWCRSCDGGAGCADYECCRCGWSGCSWRNGACECSSLSSSYTYEL	HVDAW	64	(SEQ ID NO : 294)
UL-13	CSTVHQ TTHQIH	TCPNGWTGGCVCSSRFNCRGNNCCCRTAYCSVDRYVCACFTVTYTYEF	NVDSW	61	(SEQ ID NO : 295)
UL-14	CTAVYQ KTTSIR	SCPGGTTLRNGCRSACGCNDCDCCCGSSWDICYMSKCTSAPETYTYEL	HIDAW	61	(SEQ ID NO : 296)
UL-15	CTNVHQ KTKK	TCPDDYTCGVSCSCSSSGCADYGCCSYITYGVPGDCGGCCSYARRYTYEW	NVDAW	61	(SEQ ID NO : 297)
UL-16	CTTVHQ KTKK	LCPNGRTCGCGCDCGSGCCTSYCDSFGCWGGRDTFGSSCTSAIYTYEW	GVDAW	59	(SEQ ID NO : 298)
UL-17	CATVHQ HTNKK	PCPDGYEFSAGCCGEGCSGSDCCCNSRLRCSWYEIYCSVSPSDTYEF	HVDAW	60	(SEQ ID NO : 299)
UL-18	CTTVHQ RTNKK	RCPDGYRFSAACCGEGCSGNECCCNTRLRCSWYEIYCSVSPSDTYEF	HVDAW	60	(SEQ ID NO : 300)
UL-19	CTTVHQ HTNQN	RCPTGYSHSAGCCGVGCSGNDCCCNSRLRCSWYETYCSLSPTDMYEF	YVDAW	60	(SEQ ID NO : 301)
UL-20	CSTVHQ HTNQN	RCPAGYKHSAGCCGVGCSGNDCCCNSRLRCSWYETYCSLSPTDMYEF	YVDAW	60	(SEQ ID NO : 302)

Figure 2B

UL-21	CTTVHQ KTNE	RCCRVVSDDGECGDGNSCHRWLCSDYCYSGDCCACGCRAYHYTYTYEW	NIDAW	59	(SEQ ID NO : 303)
UL-22	CTTVHQ KTNER	CCRVVSDDGECGDGNSCHRWLCSDYCYSGDCCACGCRAYHYTYTYDF	RIDVW	59	(SEQ ID NO : 304)
UL-23	CTTVHQ KTNRER	CCPDGYYYCCRSVSDCCCSTRACVGDSCGWTDFGSTHNVDCSFTYEF	HVDAW	60	(SEQ ID NO : 305)
UL-24	CTTVHQ QTRK	SCPDGYTYCHDCGYGCCCGASFCRDYGGCGSLCGRYCTSFDYIYTYEN	YVEIW	59	(SEQ ID NO : 306)
UL-25	CTTVHQ ETKR	NCPDNCYYENSCGDYGSGCNGGDCCRCGTWLTCSVSGCTCIRATNTYQW	YVNAW	60	(SEQ ID NO : 307)
UL-26	CTTVHQ STNKK	SCPDRVCWAVGCCFGEDCTSSDCTCYASPGNFYRHDCGNCDCRSSYEH	HVDAW	60	(SEQ ID NO : 308)
UL-27	CTTVRQ ETLI	RCRDGPSCAACCRSGRRCSGYGCCTDGCCSDNDYADCIRGEFVDVYEW	NVDAW	59	(SEQ ID NO : 309)
UL-28	CSTVYQ KTRT	TCPDGYTCGDGARCEKACRGCDCCRTTVCDTVWSSYCSCYSFTDSYEF	YVDAW	59	(SEQ ID NO : 310)
UL-29	CATVYQ KTNREM	SCPDGCRIHNARLCLSGCSGSDCCSCGDCVSDARCYNCRSAVFTYTYEF	HVDAW	62	(SEQ ID NO : 311)
UL-30	CTIVHQ ETKR	SCPDGYNTGTRCFGSCGCIGSNCCRSTTSCCCAGIYSQCTTSTLTYEW	HADVW	59	(SEQ ID NO : 312)
UL-31	CAIVYQ RTRQ	RCPDGYNTGTRCFGTCGCNGSNCCRFTTSCCCAGVYSQCTTSTLTYEW	HADVW	59	(SEQ ID NO : 313)
UL-32	CTTVHQ KTET	RCPDGYSSTNGCDARCGCSDCDCCNVGRWGCPLICSRNCRSFTYTYEW	YADAW	59	(SEQ ID NO : 314)
UL-33	CTTVHQ KTNKKE	SCPDGYTMNECCGCGYGCCRGGCVCSAYCSRPNCWRELTYTYTYEF	YVDTW	59	(SEQ ID NO : 315)
UL-34	CTTVYQ KSRKES	SCPNGWIYGKDCCSWSYCTDCDCCLCGDLHCYDGCSSFGVTWTYEF	HVDAW	59	(SEQ ID NO : 316)
UL-35	CTTVFQ ETRK	SCPTGFYVDGSTCGCATYCRTCDCCGGYRCSGGGSCACSSYTYNYDF	HVDAW	58	(SEQ ID NO : 317)
UL-36	CAAVFQ ETRT	NCPSGYGNAFSCGCPIACRDCDCCGGYWCSGGADCHCVSYNYTYSW	HVDAW	57	(SEQ ID NO : 318)
UL-37	CATVYQ KTEK	HCPLFHSICCHCGEGVGCSGGDCCGCERRSGCVVCIMRNSYTYNYQF	HVDAW	58	(SEQ ID NO : 319)
UL-38	CGTVHQ KTKE	LCPDDSTYCCGCVSGCACCTYGCDGVGCCRVSLWTTYIKDIVGVSYEW	HVDAW	59	(SEQ ID NO : 320)
UL-39	CASVHQ HTEP	TCPAGYTYCCGCLYKCNCGDCGCYNVGCGSGWLGKACGDYRETYEW	YVDAW	57	(SEQ ID NO : 321)
UL-40	CASVHQ HTEP	TCPAGYTYCCGCLYKCNCGDCGCYNAGCGSGWLGKACGDYRETYEW	YVDAW	57	(SEQ ID NO : 322)
UL-41	CTTVFQ ETRK	SCFSGFRDRDACGCAVTCRNCDCCGGGPCNGGGSCRCNNYIYKYSF	HVDAW	57	(SEQ ID NO : 323)
UL-42	CTAVFQ ETRK	DCPSGYGSAFTCGCLAACHGCDCCGGGWCSGGGDCRCRSYSTAYSF	HIDAW	57	(SEQ ID NO : 324)
UL-43	CATVFQ ETRK	SCPSGYADRFTCDCVYYCQTCDCCGGNRCSGGGPCRCSSYSINYSF	HVDIW	57	(SEQ ID NO : 325)
UL-44	CAAAHQ ETKR	SCPDGTCRQCCGGVCRCHASGCCYWCTTGCVGRALSESHSYEF	HVDTW	54	(SEQ ID NO : 326)
	8 Cysteines:				
UL-45	CSTVHQ KTRTTQGN	TCPDGYTLKDDCPRCRGGCDGYDCCWGDACRSSGLCWGHNPLVTETYTYEF	YIDAW	66 (167)	(SEQ ID NO : 327)
UL-46	CVVVYQ KTNSQK	SCPRGYTERETCNRRYGWGCGRYDCCDCDRWVSGNCANICTDYTDTHTYEF	HADAW	64 (23)	(SEQ ID NO : 328)
UL-47	CGTVFQ QTHKVR	DCPDGFTAAPRCGGECCCSNVNSRSGGWCRYCGRDCTAPTETSTYEF	HVDAW	60 (29)	(SEQ ID NO : 329)
UL-48	CTAVYQ RTGQ	KCPEGCESRNTCLYSRNCGDYTCCGGSRASGSGACGWNSVDCKNKYEH	HVDAW	59 (30)	(SEQ ID NO : 330)
UL-49	CTTVYQ KTKQ	NCPDGYDFRDTCGSQSYCSGYDCCRCSRFGGCSIGTCISYSDAYTYEW	YVDAW	59 (16)	(SEQ ID NO : 331)
UL-50	CTTVHQ QTHEKR	SCPESYSYSCSCASGVVGCGPDDCCCTYRISIRGYTCSSLSNSYEW	YVDAW	59 (15)	(SEQ ID NO : 332)
UL-51	CTAVHQ QTRRKS	GCPDGYSDESCSYCGSSWCCPVYWCGSPCSYRCLRHTDTYSYEH	HVDAW	57 (12)	(SEQ ID NO : 333)
UL-52	CATVYQ ETKR	TCAGGHSVECDSPYDCNCRGGDCCRSPIFNDCWAASCSATRTYEW	HVESW	56 (16)	(SEQ ID NO : 334)
UL-53	CITVHQ ETQK	SCPDDYTYYGDGTCAYVCSIDKCCCGRTWLSSGCLPCRYTYNL	HVDAW	54 (155)	(SEQ ID NO : 335)
UL-54	CITVHQ ETQK	SCPDDYTSYGDATCAYVCSTDECCCGRTWLSAGCRPCRYTYNL	HVDAW	54 (61)	(SEQ ID NO : 336)
UL-55	CITVHQ ETQK	SCFDDYTYYGDASCAYVCSTDECCCGRTWLSAGCRPCRYTYNL	HVDAW	54 (20)	(SEQ ID NO : 337)
UL-56	CITAHQ ETQK	SCSDDYTYYGDATCAYVCSTDECCCGRTWLSAGCRPCRYTYNL	HVDAW	54 (20)	(SEQ ID NO : 338)
UL-57	CITVHQ ETQK	SCPDDYTYYGDGTCAYVCSIDNCCCGRTWLSSGCLPCRYTYNL	HVDAW	54 (10)	(SEQ ID NO : 339)

Figure 2C

Other Representative sequences:

UL-58	CVTVEQ QTHATR	RCPDNGYSDSYACKSNYGCSAEGCCPRWGPGSGACTGAIYTTSPYEW	YVDAW	57	(SEQ ID NO : 340)
UL-59	CAAVEQ RTEGQQ	SCPDGYLETRVCPYRMYRCTGWDCCKCSDGSRDNYIMTYSYEF	HVDVW	56	(SEQ ID NO : 341)
UL-60	CTTVYQ ETNTKS	GCPDGYSCCYNGRSRSCRPNDCSTYGEVKSLSRSCYTYNYEF	YVDAW	55	(SEQ ID NO : 342)
UL-61	CGTVYQ HTKEIK	TCPDGYSDVFTYCPVTCPGWDCCRPNDCGRTRYTVAYSYAL	HVDVW	54	(SEQ ID NO : 343)
UL-62	CTTVLQ ETHQQR	GCPAGYQVVDGCPYGDCCSTSYVCGELTCTSNTATRNYQW	YVDAW	53	(SEQ ID NO : 344)
UL-63	CSTVYQ KTEK	RCPDGYTDRRDECPNTCKNFDCENEGGLRCLCSAYISAYEF	HVDAW	52	(SEQ ID NO : 345)
UL-64	CTTTHQ RTGK	SCPDYASYDCGSFDDEECSSCRSCTPWCAPTAPYIYTYQF	YIDAW	51	(SEQ ID NO : 346)
UL-65	CTTVEQ QTNK	RCPTGYNSGTLCNMTGCSUDECCNYGRVECTSYVWTHNF	YVDAW	50	(SEQ ID NO : 347)
UL-66	CTTVEQ ETQRT	SCPSGWTYTCNCRNGCCCYRPSQLCGAYVAVTHTYEF	HVDAW	49	(SEQ ID NO : 348)
UL-67	CATVEQ KDK	BCPAGYRSGTLCRMIGCTGDDCCNYDSVECTNYDYTNNF	YVDAW	49	(SEQ ID NO : 349)
UL-68	CTAVEQ QTTEAGK	TCPFRSRDMGTRCSDDRYYPWRYSDYAYTYTYEW	HVDAW	48	(SEQ ID NO : 350)
UL-69	CTSVEQ NTDV	TCPSGATYRCDCGGKGCGCYTPWCSTTYRGTYTYDF	HVETW	47	(SEQ ID NO : 351)
UL-70	CGTVEQ ETHTQR	TCPDACDVTGDNCKVERNGDWCGRASKTDTYDF	YVDAW	46	(SEQ ID NO : 352)
UL-71	CTTDYQ KTEK	SCPENYYAETGYCMCGSWRCGYGSTTSLIVSYKW	YVDAW	45	(SEQ ID NO : 353)
UL-72	CTTVEQ NTNQKW	GCPDGYVHMSGSCCKGSICTNGLFKNTYTYEF	NVEAW	45	(SEQ ID NO : 354)
UL-73	CTTVYQ ETRT	NCPDGYNYRSGDCKRWNHWLGEQRVSPTTYNYEW	YVDSW	44	(SEQ ID NO : 355)
UL-74	CTTVYQ KTTTTK	SCPGGFDNGRKCIMSLGDLRDYTYKFNKYEW	YVETW	43	(SEQ ID NO : 356)
UL-75	CSTVEQ NTEQ	RCLDGYDDRGAYCYDSVRGLMSWTYKNYYIEW	RVDTW	42	(SEQ ID NO : 357)
UL-76	CTNVHQ MTIK	TCPDGGSYGWYWFNSYGCSNGGVSANYTYEF	YVDAW	41	(SEQ ID NO : 358)
UL-77	CTTVYQ KTESVR	SCPNGSMDGWRCRLGTNNWTYSNTYEF	YVDAW	40	(SEQ ID NO : 359)

Figure 3

BLV1H12

SVHQETKKYQSCPDGYRERSDCSNRPACGTSDCCRVSVFGNCLTTLPVSYSYTYNYEWHVD
(SEQ ID NO : 22)

BLV5B8
TVHQETRKTCSDGYIAVDSCGRGQSDGCVNDCNSCYYGWRNCRRQPAIHSYEFHVD
(SEQ ID NO : 23)

BLV5D3
SVTQRTHVSRSCPDGCSDGDGCVDGCCCSAYRCYTPGVRDLSCTSYSITYTYEWNVD
(SEQ ID NO : 24)

BLV6C11
TVHQKTTRKTCCSDAYRYDSGCGSGCDCCGADCYVFGACTFGLDSSYSYIYIYQWYVD
(SEQ ID NO : 25)

BF4E9
TVHQIFCPDGYSYGYGCGYGYGCSGYDCYGYGGYGYGGYGGYSSYSYSYSYEYYGD
(SEQ ID NO : 26)

BF1H1
TVHPSPDGYSYGYGCGYGYGCSGYDCYGYGGYGYGGYGGYSSYSYSYS
(SEQ ID NO : 27)

F18
TVHQIRCPDGYGYGYGCGYGSYGYSGYDCYGYGGYGGYGGYGGYSSYS
(SEQ ID NO : 28)

BLV1H12 VL:

Caggctgtgctgaatcagccatcatccgtgtccgggtccctgggccagagggtctccatcacctgctctggaagcag
cagcaatgttggaaatggatatgtgagctggtaccaactgatcccaggatcggccccccagaaccctcatctatggtg
acaccagtcgagcctcgggggtcccccgaccgattctccggctccaggtctgggaacacagccaccctgaccatcagc
tcgctccaggctgaggacgaggcagattatttctgtgcatctgctgaggatagtagcagtaatgctgttttcggcag
cgggaccacactgaccgtcctg
(SEQ ID NO: 372)

QAVLNQPSSVSGSLGQRVSITCSGSSSNVGNGYVSWYQLIPGSAPRTLIYGDTSRASGVPDRFSGSRSGNTATLTIS
SLQAEDEADYFCASAEDSSSNAVFGSGTTLTVL
(SEQ ID NO: 35)

Vl1-47:

Cagtctgtgctgactcagccaccctcagcgtctgggacccccgggcagagggtcaccatctcttgttctggaagcag
ctccaacatcggaagtaattatgtatactggtaccagcagctcccaggaacggcccccaaactcctcatctatagga
ataatcagcggccctcaggggtccctgaccgattctctggctccaagtctggcacctcagcctccctggccatcagt
gggctccggtccgaggatgaggctgattattactgtgcagcatgggatgacagcctgagtggtcc
(SEQ ID NO: 373)

QSVLTQPPSASGTPGQRVTISCSGSSSNIGSNYVYWYQQLPGTAPKLLIYRNNQRPSGVPDRFSGSKSGTSASLAIS
GLRSEDEADYYCAAWDDSLSG
(SEQ ID NO: 36)

Vl1-40*1:
Cagtctgtgctgacgcagccgccctcagtgtctggggccccagggcagagggtcaccatctcctgcactgggagcag
ctccaacatcggggcaggttatgatgtacactggtaccagcagcttccaggaacagcccccaaactcctcatctatg
gtaacagcaatcggccctcagggtccctgaccgattctctggctccaagtctggcacctcagcctccctggccatc
actgggctccaggctgaggatgaggctgattattactgccagtcctatgacagcagcctgagtggttc
(SEQ ID NO: 374)

QSVLTQPPSVSGAPGQRVTISCTGSSSNIGAGYDVHWYQQLPGTAPKLLIYGNSNRPSGVPDRFSGSKSGTSASLAI
TGLQAEDEADYYCQSYDSSLSG
(SEQ ID NO: 37)

Vl1-51 *01:
Cagtctgtgttgacgcagccgccctcagtgtctgcggccccaggacagaaggtcaccatctcctgctctggaagcag
ctccaacattgggaataattatgtatcctggtaccagcagctcccaggaacagcccccaaactcctcatttatgaca
ataataagcgaccctcagggattcctgaccgattctctggctccaagtctggcacgtcagccaccctgggcatcacc
ggactccagactggggacgaggccgattattactgcggaacatgggatagcagcctgagtgctgg
SEQ ID NO: 375)

QSVLTQPPSVSAAPGQKVTISCSGSSSNIGNNYVSWYQQLPGTAPKLLIYDNNKRPSGIPDRFSGSKSGTSATLGIT
GLQTGDEADYYCGTWDSSLSA
SEQ ID NO: 38)

Vl2-18*02:
Cagtctgccctgactcagcctccctccgtgtcccgggtctcctggacagtcagtcaccatctcctgcactggaaccagcagtgacgt
tggtagttataaccgtgtctcctggtaccagcagccccaggcacagccccaaactcatgatttatgaggtcagtaatcggccct
caggggtccctgatcgcttctctgggtccaagtctggcaacacggcctccctgaccatctctgggctccaggctgaggacgaggct
gattattactgcagctcatatacaagcagcagcacttc
(SEQ ID NO: 376)

QSALTQPPSVSGSPGQSVTISCTGTSSDVGSYNRVSWYQQPPGTAPKLMIYEVSNRPSGVPDRFSGSKSGNTASLTI
SGLQAEDEADYYCSSYTSSSTF
(SEQ ID NO: 39)

Figure 6A

Ab Name	VH-CH1	VL-CL
PGT145	QVQLVQSGAEVKKPGSSVKVSCKASGNSFSNHDVH WVRQATGQGLEWMGWMSHEGDKTGLAQKFQGRVT ITRDSGASTVYMELRGLTADDTAIYYCLTGSKHRLRDY FLYNEYGPNYEEWGDYLATLDVWGHGTAVTVSSAST KGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVS WNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLG TQTYICNVNHKPSNTKVDKKVEPKSCD (SEQ ID NO: 486)	EVVITQSPLFLPVTPGEAASLSCKCSHSLQHSTGANYLAWYLQRP GQTPRLLIHLATHRASGVPDRFSGSGSGTDFTLKISRVESDDVGTY YCMQGLHSPWTFGQGTKVEIKRTVAAPSVFIFPPSDEQLKSGTAS VVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSL SSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC (SEQ ID NO: 491)
PG9	QRLVESGGGVVQPGSSLRLSCAASGFDFSRQGMHW VRQAPGQGLEWVAFIKYDGSEKYHADSVWGRLSISR DNSKDTLYLQMNSLRVEDTATYFCVREAGGPDYRNG YNYYDFYDGYYNYHYMDVWGKGTTVTVSSASTKGP SVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNS GALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQT YICNVNHKPSNTKVDKKVEPKSCDKGLEVLFQ (SEQ ID NO: 487)	QSALTQPASVSGSPGQSITISCQGTSNDVGGYESVSWYQQHPGK APKVVIYDVSKRPSGVSNRFSGSKSGNTASLTISGLQAEDEGDYY CKSLTSTRRRVFGTGTKLTVLGQPKAAPSVTLFPPSSEELQANKA TLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAAS SYLSLTPEQWKSHKSYSCQVTHEGSTVEKTVAPTECS (SEQ ID NO: 492)
PG16	QEQLVESGGGVVQPGGSLRLSCLASGFTFHKYGMH WVRQAPGKGLEWVALISDDGMRKYHSDSMWGRVTI SRDNSKNTLYLQFSSLKVEDTAMFFCAREAGGPIWH DDVKYYDFNDGYYNYHYMDVWGKGTTVTVSSASTK GPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSW NSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGT QTYICNVNHKPSNTKVDKRVEPK (SEQ ID NO: 488)	QSALTQPASVSGSPGQTITISCNGTSSDVGGFDSVSWYQQSPGK APKVMVFDVSHRPSGISNRFSGSKSGNTASLTISGLHIEDEGDYFC SSLTDRSHRIFGGGTKVTVLGQPKAAPSVTLFPPSSEELQANKATL VCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSY LSLTPEQWKSHKSYSCQVTHEGSTVEKTVAPTECS (SEQ ID NO: 493)
CHO4	EVQLVESGGGLIRPGGSLRLSCKGSGFIFENFGFGWV RQGPGKGLEWVSGTNWNGGDSRYGDSVKGRFTISR DNSNNFVYLQMNSLRPEDTAIYYCARGTDYTIDDQGI RYQGSGTFWYFDVWGRGTLVTVSSASTKGPSVFPLA PSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTS GVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNV NHKPSNTKVDKKVEPKSCD (SEQ ID NO: 489)	EIVLTQSPDTLSLSPGERATLSCRASQSVHSRYFAWYQHKPGQPP RLLIYGGSTRATGIPNRFSAGGSGTQFTLTVNRLEAEDFAVYYCQ QYGRSPYTFGQGTKVEIRRTVAAPSVFIFPPSDEQLKSGTASVVCL LNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTL TLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC (SEQ ID NO: 494)

Figure 6A (Continued)

Ab Name	VH-CH1	VL-CL
2909	EVQLVESGGNVVQPGGSLRLSCTASGFSFDDSTMH WVRQAPGKGLQWVSLISWNGGRTYYADSVKGRFTIS RDNSKNSLYLQMNSLKTEDTAFYFCAKDKGDSDYDY NLGYSYFYYMDGWGKGTTVTVSSASTKGPSVFPLAP SSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSG VHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVN HKPSNTKVDKRVEPK (SEQ ID NO: 490)	SYVLTQPPSVSVAPGKTARITCGGNNIANKNVHWYQQKPGQAPVL VIYYDDDRPSGIPDRFSGSNSGNTATLTISRVEAGDEADYYCQVW DSNSDHVVFGGGTQLTVLGQPKAAPSVTLFPPSSEELQANKATLV CLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYL SLTPEQWKSHRSYSCQVTHEGSTVEKTVAPT (SEQ ID NO: 495)

Figure 6B

Ab Name	VH	CH1	VL	CL
PGT145	QVQLVQSGAEVKKPGSSVKVSCKASGNSFSNHDVHWVRQATGQGLEWMGWMSHEGDKTGLAQKFQGRVTITRDSGASTVYMELRGLTADDTAIYYCLT**GSKHRLRDYFLYNEYGPNYEEWGDYLATLDV**WGHGTAVTVSS (SEQ ID NO: 656)	ASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCD (SEQ ID NO: 661)	EVVITQSPLFLPVTPGEAASLSCKCSHSLQHSTGANYLAWYLQRPGQTPRLLIHLATHRASGVPDRFSGSGSGTDFTLKISRVESDDVGTYYCMQGLHSPWTF (SEQ ID NO: 734)	GQGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC (SEQ ID NO: 739)
PG9	QRLVESGGGVVQPGSSLRLSCAASGFDFSRQGMHWVRQAPGQGLEWVAFIKYDGSEKYHADSVWGRLSISRDNSKDTLYLQMNSLRVEDTATYFCVR**EAGGPDYRNGYNYYDFYDGYYNYHYMDV**WGKGTTVTVSS (SEQ ID NO: 657)	ASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCDKGLEVLFQ (SEQ ID NO: 662)	QSALTQPASVSGSPGQSITISCQGTSNDVGGYESVSWYQQHPGKAPKVVIYDVSKRPSGVSNRFSGSKSGNTASLTISGLQAEDEGDYYCKSLTSTRRRVF (SEQ ID NO: 735)	GTGTKLTVLGQPKAAPSVTLFPPSSEELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQWKSHKSYSCQVTHEGSTVEKTVAPTECS (SEQ ID NO: 740)
PG16	QEQLVESGGGVVQPGGSLRLSCLASGFTFHKYGMHWVRQAPGKGLEWVALISDDGMRKYHSDSMWGRVTISRDNSKNTLYLQFSSLKVEDTAMFFCAR**EAGGPIWHDDVKYYDFNDGYYNYHYMDV**WGKGTTVTVSS (SEQ ID NO: 658)	TKGPSVFPLAPSSKSTSGGASTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKRVEPK (SEQ ID NO: 663)	QSALTQPASVSGSPGQTITISCNGTSSDVGGFDSVSWYQQSPGKAPKVMVFDVSHRPSGISNRFSGSKSGNTASLTISGLHIEDEGDYFCSSLTDRSHRIF (SEQ ID NO: 736)	GGGTKVTVLGQPKAAPSVTLFPPSSEELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTPEQWKSHKSYSCQVTHEGSTVEKTVAPTECS (SEQ ID NO: 741)
CHO4	EVQLVESGGGLIRPGGSLRLSCKGSGFIFENFGFGWVRQGPGKGLEWVSGTNWNGGDSRYGDSVKGRFTISRDNSNNFVYLQMNSLRPEDTAIYYCAR**GTDYTIDDQGIRYQGSGTFWYFDV**WGRGTLVTVSS (SEQ ID NO: 659)	ASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCD (SEQ ID NO: 664)	EIVLTQSPDTLSLSPGERATLSCRASQSVHSRYFAWYQHKPGQPPRLLIYGGSTRATGIPNRFSAGGSGTQFTLTVNRLEAEDFAVYYCQQYGRSPYTF (SEQ ID NO: 737)	GQGTKVEIRRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC (SEQ ID NO: 742)

Figure 6B (Continued)

Ab Name	VH	CH1	VL	CL
2909	EVQLVESGGNVVQPGGSLR LSCTASGFSFDDSTMHWVR QAPGKGLQWVSLISWNGG RTYYADSVKGRFTISRDNSK NSLYLQMNSLKTEDTAFYFC AK**DKGDSDYDYNLGYSYFY YMDG**WGKGTTVTVSS (SEQ ID NO: 660)	ASTKGPSVFPLAPSSKSTSGG TAALGCLVKDYFPEPVTVSW NSGALTSGVHTFPAVLQSSGL YSLSSVVTVPSSSLGTQTYIC NVNHKPSNTKVDKRVEPK (SEQ ID NO: 665)	SYVLTQPPSVSVAPGKTARIT CGGNNIANKNVHWYQQKPG QAPVLVIYYDDDRPSGIPDRF SGSNSGNTATLTISRVEAGDE ADYYCQVWDSNSDHVVF (SEQ ID NO: 738)	GGGTQLTVLGQPKAAPSVTL FPPSSEELQANKATLVCLISDF YPGAVTVAWKADSSPVKAGV ETTTPSKQSNNKYAASSYLSL TPEQWKSHRSYSCQVTHEGS TVEKTVAPT (SEQ ID NO: 743)

*Bold residues represent CDR3.

Figure 6C

		Ultralong CDR3			
Ab Name	V1 region	A region	Insert	D region	V2 region
PGT145	QVQLVQSGAEVKKPGSSVKVSC KASGNSFSNHDVHWVRQATGQG LEWMGWMSHEGDKTGLAQKFQ GRVTITRDSGASTVYMELRGLTA DDTAIYYCLT (SEQ ID NO: 496)	GSKHRLRDYFLYNE (SEQ ID NO: 501)		YGPNYEEWGDYLA TLDV (SEQ ID NO: 536)	WGHGTAVTVSS (SEQ ID NO: 570)
		GSKHRLRDYFLYN (SEQ ID NO: 502)		GPNYEEWGDYLAT LDV (SEQ ID NO: 537)	
		GSKHRLRDYFLY (SEQ ID NO: 503)		PNYEEWGDYLATLD V (SEQ ID NO: 538)	
		GSKHRLRDYFL (SEQ ID NO: 504)		NYEEWGDYLATLDV (SEQ ID NO: 539)	
		GSKHRLRDYF (SEQ ID NO: 505)		YEEWGDYLATLDV (SEQ ID NO: 540)	
		GSKHRLRDY (SEQ ID NO: 506)		EEWGDYLATLDV (SEQ ID NO: 541)	
		GSKHRLRD (SEQ ID NO: 507)			
PG9	QRLVESGGGVVQPGSSLRLSCA ASGFDFSRQGMHWVRQAPGQG LEWVAFIKYDGSEKYHADSVWGR LSISRDNSKDTLYLQMNSLRVEDT ATYFCVR (SEQ ID NO: 497)	EAGGPDYRNGYNY (SEQ ID NO: 508)		YDFYDGYYNYHYM DV (SEQ ID NO: 542)	WGKGTTVTVSS (SEQ ID NO: 571)
		EAGGPDYRNGYN (SEQ ID NO: 509)		DFYDGYYNYHYMD V (SEQ ID NO: 543)	
		EAGGPDYRNGY (SEQ ID NO: 510)		FYDGYYNYHYMDV (SEQ ID NO: 544)	
		EAGGPDYRNG (SEQ ID NO: 511)		YDGYYNYHYMDV (SEQ ID NO: 545)	
		EAGGPDYRN (SEQ ID NO: 512)		DGYYNYHYMDV (SEQ ID NO: 546)	

Figure 6C (Continued)

		Ultralong CDR3			
Ab Name	**V1 region**	**A region**	**Insert**	**D region**	**V2 region**
		EAGGPDYR (SEQ ID NO: 513)		GYYNYHYMDV (SEQ ID NO: 547)	
		EAGGPDY (SEQ ID NO: 514)		YYNYHYMDV (SEQ ID NO: 548)	
		EAGGPD (SEQ ID NO: 515)			
PG16	QEQLVESGGGVVQPGGSLRLSCLASGFTFHKYGMHWVRQAPGKGLEWVALISDDGMRKYHSDSMWGRVTISRDNSKNTLYLQFSSLKVEDTAMFFCAR (SEQ ID NO: 498)	EAGGPIWHDDVKY (SEQ ID NO: 516)		YDFNDGYYNYHYMDV (SEQ ID NO: 549)	WGKGTTVTVSS (SEQ ID NO: 572)
		EAGGPIWHDDVK (SEQ ID NO: 517)		DFYDGYYNYHYMDV (SEQ ID NO: 550)	
		EAGGPIWHDDV (SEQ ID NO: 518)		FYDGYYNYHYMDV (SEQ ID NO: 551)	
		EAGGPIWHDD (SEQ ID NO: 519)		YDGYYNYHYMDV (SEQ ID NO: 552)	
		EAGGPIWHD (SEQ ID NO: 520)		DGYYNYHYMDV (SEQ ID NO: 553)	
		EAGGPIWH (SEQ ID NO: 521)		GYYNYHYMDV (SEQ ID NO: 554)	
		EAGGPIW (SEQ ID NO: 522)			
		EAGGPI (SEQ ID NO: 523)			
CHO4	EVQLVESGGGLIRPGGSLRLSCKGSGFIFENFGFGWVRQGPGKGLEWVSGTNWNGGDSRYGDSVKGRFTISRDNSNNFVYLQMNSLRPEDTAIYYCAR (SEQ ID NO: 499)	GTDYTIDDQGI (SEQ ID NO: 524)		QGIRYQGSGTFWYFDV (SEQ ID NO: 555)	WGRGTLVTVSS (SEQ ID NO: 573)
		GTDYTIDDQG (SEQ ID		GIRYQGSGTFWYFD	

Figure 6C (Continued)

		Ultralong CDR3			
Ab Name	V1 region	A region	Insert	D region	V2 region
		NO: 525)		V (SEQ ID NO: 556)	
		GTDYTIDDQ (SEQ ID NO: 526)		IRYQGSGTFWYFDV (SEQ ID NO: 557)	
		GTDYTIDD (SEQ ID NO: 527)		RYQGSGTFWYFDV (SEQ ID NO: 558)	
		GTDYTID (SEQ ID NO: 528)		YQGSGTFWYFDV (SEQ ID NO: 559)	
		GTDYTI (SEQ ID NO: 529)		QGSGTFWYFDV (SEQ ID NO: 560)	
				GSGTFWYFDV (SEQ ID NO: 561)	
				SGTFWYFDV (SEQ ID NO: 562)	
				GTFWYFDV (SEQ ID NO: 563)	
2909	EVQLVESGGNVVQPGGSLRLSCT ASGFSFDDSTMHWVRQAPGKGL QWVSLISWNGGRTYYADSVKGR FTISRDNSKNSLYLQMNSLKTEDT AFYFCAK (SEQ ID NO: 500)	DKGDSDYDYNL (SEQ ID NO: 530)		YNLGYSYFYYMDG (SEQ ID NO: 564)	WGKGTTVTVSS (SEQ ID NO: 574)
		DKGDSDYDYN (SEQ ID NO: 531)		NLGYSYFYYMDG (SEQ ID NO: 565)	
		DKGDSDYDY (SEQ ID NO: 532)		LGYSYFYYMDG (SEQ ID NO: 566)	
		DKGDSDYD (SEQ ID NO: 533)		GYSYFYYMDG (SEQ ID NO: 567)	
		DKGDSDY (SEQ ID NO: 534)		YSYFYYMDG (SEQ ID NO: 568)	
		DKGDSD (SEQ ID NO: 535)		SYFYYMDG (SEQ ID NO: 569)	

Figure 6D

Ab Name	V1 Alternative A	V1 Alternative B
PGT145	QVQLVQSGAEVKKPGSSVKVSCKASGN SFSNHDVHWVRQATGQGLEWMGWMS HEGDKTGLAQKFQGRVTITRDSGASTV YMELRGLTADDTAIYYC (SEQ ID NO: 744)	QVQLVQSGAEVKKPGSSVKVSCKASGN SFSNHDVHWVRQATGQGLEWMGWMS HEGDKTGLAQKFQGRVTITRDSGASTV YMELRGLTADDTAIYY (SEQ ID NO: 749)
PG9	QRLVESGGGVVQPGSSLRLSCAASGFD FSRQGMHWVRQAPGQGLEWVAFIKYD GSEKYHADSVWGRLSISRDNSKDTLYL QMNSLRVEDTATYFC (SEQ ID NO: 745)	QRLVESGGGVVQPGSSLRLSCAASGFD FSRQGMHWVRQAPGQGLEWVAFIKYD GSEKYHADSVWGRLSISRDNSKDTLYL QMNSLRVEDTATYF (SEQ ID NO: 750)
PG16	QEQLVESGGGVVQPGGSLRLSCLASGF TFHKYGMHWVRQAPGKGLEWVALISD DGMRKYHSDSMWGRVTISRDNSKNTL YLQFSSLKVEDTAMFFC (SEQ ID NO: 746)	QEQLVESGGGVVQPGGSLRLSCLASGF TFHKYGMHWVRQAPGKGLEWVALISD DGMRKYHSDSMWGRVTISRDNSKNTL YLQFSSLKVEDTAMFF (SEQ ID NO: 751)
CHO4	EVQLVESGGGLIRPGGSLRLSCKGSGFI FENFGFGWVRQGPGKGLEWVSGTNW NGGDSRYGDSVKGRFTISRDNSNNFVY LQMNSLRPEDTAIYYC (SEQ ID NO: 747)	EVQLVESGGGLIRPGGSLRLSCKGSGFI FENFGFGWVRQGPGKGLEWVSGTNW NGGDSRYGDSVKGRFTISRDNSNNFVY LQMNSLRPEDTAIYY (SEQ ID NO: 752)
2909	EVQLVESGGNVVQPGGSLRLSCTASGF SFDDSTMHWVRQAPGKGLQWVSLISW NGGRTYYADSVKGRFTISRDNSKNSLYL QMNSLKTEDTAFYFC (SEQ ID NO: 748)	EVQLVESGGNVVQPGGSLRLSCTASGF SFDDSTMHWVRQAPGKGLQWVSLISW NGGRTYYADSVKGRFTISRDNSKNSLYL QMNSLKTEDTAFYF (SEQ ID NO: 753)

Figure 7A

SEQ ID NO:	Amino Acid Sequence
575	GGGGS
576	GGGGS GGGGS
577	GGGGS GGGGS GGGGS
578	GGGGS GGGGS GGGGS GGGGS
579	GGS
580	GGS GGS
581	GGS GGS GGS
582	GGS GGS GGS GGS
583	ASG
584	ASG ASG
585	ASG ASG ASG
586	ASG ASG ASG ASG
587	GCGGGGS
588	GCGGGGS GGGGS
589	GCGGGGS GGGGS GGGGS
590	GCGGGGS GGGGS GGGGS GGGGS
591	GCGGS
592	GCGGS GGS
593	GCGGS GGS GGS
594	GCGGS GGS GGS GGS
595	GCASG
596	GCGCASG ASG
597	GCASG ASG ASG
598	GCASG ASG ASG ASG
699	SGGGG
700	SGGGG SGGGG
701	SGGGG SGGGG SGGGG
702	SGGGG SGGGG SGGGG SGGGG
703	SGG
704	SGG SGG
705	SGG SGG SGG
706	SGG SGG SGG SGG
707	GSA

Figure 7A (Continued)

SEQ ID NO:	Amino Acid Sequence
708	GSA GSA
709	GSA GSA GSA
710	GSA GSA GSA GSA
711	SGGGGCG
712	SGGGG SGGGGCG
713	SGGGG SGGGG SGGGGCG
714	SGGGG SGGGG SGGGG SGGGGCG
715	SGGCG
716	SGG SGGCG
717	SGG SGG SGGCG
718	SGG SGG SGG SGGCG
719	GSACG
720	GSA GSACGCG
721	GSA GSA GSACG
722	GSA GSA GSA GSACG
723	GGGGS GG
724	GGGGS GGGGS GGGGS GG
725	GG SGGGG
726	GG SGGGG SGGGG GGGGS
756	G
757	GG
758	GGG
759	GGGG
760	GGGGGSGGS
761	GGGGSGGGGSGG
762	GGGGSGGS
763	GGGGSGGSGGS
764	GGGGSGGSGGSGGS
765	GGSG
766	GGSGG
767	GGSGGSG
768	GGSGGSGG
769	GGSGGSGGSG

Figure 7A (Continued)

SEQ ID NO:	Amino Acid Sequence
770	GGSGGSGGSGSG
771	GGSGGSGGSGGSGSG
772	GSG
773	GSGG

Figure 7B

Peptide Name	Amino Acid Sequence	Target
ADWX-1	VGINVKCKHSRQCLKPCKDAGMRFGKCTNGKCHCTPK (SEQ ID NO: 599)	Kv1.3
HsTx1	ASCRTPKDCADPCRKETGCPYGKCMNRKCKCNRC (SEQ ID NO: 600)	Kv1.3
OSK1	GVIINVKCKISRQCLEPCKKAGMRFGKCMNGKCHCTPK (SEQ ID NO: 601)	Kv1.3
Pi2	TISCTNPKQCYPHCKKETGYPNAKCMNRKCKCFGR (SEQ ID NO: 602)	Kv1.3
Hongotoxin (HgTX)	TVIDVKCTSPKQCLPPCKAQFGIRAGAKCMNGKCKCYPH (SEQ ID NO: 603)	Kv1.3
Margatoxin	TIINVKCTSPKQCLPPCKAQFGQSAGAKCMNGKCKCYPH (SEQ ID NO: 604)	Kv1.3
Agitoxin-2	GVPINVSCTGSPQCIKPCKDAGMRFGKCMNRKCHCTPK (SEQ ID NO: 605)	Kv1.3
Pi3	TISCTNEKQCYPHCKKETGYPNAKCMNRKCKCFGR (SEQ ID NO: 606)	Kv1.3
Kaliotoxin	GVEINVKCSGSPQCLKPCKDAGMRFGKCMNRKCHCTPK (SEQ ID NO: 607)	Kv1.3
Anuroctoxin	ZKECTGPQHCTNFCRKNKCTHGKCMNRKCKCFNCK (SEQ ID NO: 608)	Kv1.3
Charybdotoxin	ZFTNVSCTTSKECWSVCQRLHNTSRGKCMNKKCRCYS (SEQ ID NO: 609)	Kv1.3
Tityustoxin -K- alpha	VFINAKCRGSPECLPKCKEAIGKAAGKCMNGKCKCYP (SEQ ID NO: 610)	Kv1.3
Maurotoxin	VSCTGSKDCYAPCRKQTGCPNAKCINKSCKCYGC (SEQ ID NO: 611)	Kv1.3
Ceratotoxin 1 (CcoTx1)	DCLGWFKSCDPKNDKCCKNYTCSRRDRWCKYDL (SEQ ID NO: 612)	
CcoTx2	DCLGWFKSCDPKNDKCCKNYTCSRRDRWCKYYL (SEQ ID NO: 613)	
CcoTx3	GVDKEGCRKLLGGCTIDDDCCPHLGCNKKYWHCGWDGTF (SEQ ID NO: 614)	
Phrixotoxin 3 (PaurTx3)	DCLGFLWKCNPSNDKCCRPNLVCSRKDKWCKYQI (SEQ ID NO: 615)	

Figure 7B (Continued)

Peptide Name	Amino Acid Sequence	Target
Hanatoxin 1	ECRYLFGGCKTTSDCCKHLGCKFRDKYCAWDFTFS (SEQ ID NO: 616)	
Phrixotoxin 1	YCQKWMWTCDSARKCCEGLVCRLWCKKII (SEQ ID NO: 617)	
Huwentoxin-IV	ECLEIFKACNPSNDQCCKSSKLVCSRKTRWCKYQI (SEQ ID NO: 618)	
α-conotoxin Iml	GCCSDPRCAWRC (SEQ ID NO: 619)	
α-conotoxin Epl	GCCSDPRCNMNNPDYC (SEQ ID NO: 620)	
α-conotoxin PnIA	GCCSLPPCAANNPDYC (SEQ ID NO: 621)	
α-conotoxin PnIB	GCCSLPPCALSNPDYC (SEQ ID NO: 622)	
α-conotoxin MII	GCCSNPVCHLEHSNLC (SEQ ID NO: 623)	
α-conotoxin AuIA	GCCSYPPCFATNSDYC (SEQ ID NO: 624)	
α-conotoxin AuIB	GCCSYPPCFATNPDC (SEQ ID NO: 625)	
α-conotoxin AuIC	GCCSYPPCFATNSGYC (SEQ ID NO: 626)	
conotoxin κ-PVIIA	CRIPNQKCFQHLDDCCSRKCNRFNKCV (SEQ ID NO: 627)	
charybdotoxin	ZFTNVSCTTSKECWSVCQRLHNTSRGKCMNKKCRCYS (SEQ ID NO: 628)	
neurotoxin B-IV	ASATWGAAYPACENNCRKKYDLCIRCQGKWAGKRGKCAAHCIIQKNNCKG KCKKE (SEQ ID NO: 629)	
crotamine	YKQCHKKGGHCFPKEKICLPPSSDFGKMDCCRWRWKCCKKGSG (SEQ ID NO: 630)	
ω-GVIA (conotoxin)	CKSPGSSCSPTSYNCCRSCNPYTKRCY (SEQ ID NO: 631)	
κ-hefutoxin 1	GHACYRNCWREGNDEETCKERC (SEQ ID NO: 632)	

Figure 7B (Continued)

Peptide Name	Amino Acid Sequence	Target
Css4	KEGYLVNSYTGCKFECFKLGDNDYCLRECRQQYGKGSGGYCYAFGCWCT HLYEQAVVWPLPNKTCN (SEQ ID NO: 633)	
Bj-xtrIT	KKNGYPLDRNGKTTECSGVNAIAPHYCNSECTKVYVAESGYCCWGACYCF GLEDDKPIGPMKDITKKYCDVQIIPS (SEQ ID NO: 634)	
BclV	GLPCDCHGHTGTYWLNYYSKCPKGYGYTGRCRYLVGSCCYK (SEQ ID NO: 635)	
Hm-1	GCIPYGKTCEFWSGPWCCAGKCKLNVWSMTLSCTRNF (SEQ ID NO: 636)	
Hm-2	GCIPSFGECAWFSGESCCTGICKWVFFTSKFMCRRVWGKD (SEQ ID NO: 637)	
GsAF-I (β-theraphotoxin-Gr1b)	YCQKWLWTCDSERKCCEDMVCRLWCKKRL (SEQ ID NO: 638)	
Protoxin I (ProTx-I, β-theraphotoxin-Tp1a)	ECRYWLGGCSAGQTCCKHLVCSRRHGWCVWDGTFS (SEQ ID NO: 639)	
Protoxin II (ProTx II)	YCQKWMWTCDSERKCCEGMVCRLWCKKKLW (SEQ ID NO: 640)	
Huwentoxin I	ACKGVFDACTPGKNECCPNRVCSDKHKWCKWKL (SEQ ID NO: 641)	
μ-Conotoxin PIIIA	ERLCCGFPKSCRSRQCKPHRCC (SEQ ID NO: 642)	
Jingzhaotoxin-III (ß-TRTX-Cj1α)	DGECGGFWWKCGRGKPPCCKGYACSKTWGWCAVEAP (SEQ ID NO: 643)	
GsAF-II (Kappa-theraphotoxin-Gr2c)	YCQKWMWTCDEERKCCEGLVCRLWCKKKIEW (SEQ ID NO: 644)	
ShK, K16,E30	RSCIDTIPKSRCTAFKCKHSMKYRLSFCRETCGTC (SEQ ID NO: 645)	
ShK (Stichodactyla toxin)	RSCIDTIPKSRCTAFQCKHSMKYRLSFCRKTCGTC (SEQ ID NO: 646)	Kv1.3
HsTx1	ASCRTPKDCADPCRKETGCPYGKCMNRKCKCNRC (SEQ ID NO: 647)	
Guangxitoxin 1E, GxTx-1E	EGECGGFWWKCGSGKPACCPKYVCSPKWGLCNFPMP (SEQ ID NO: 648)	
Charybdotoxin, ChTX	EFTNVSCTTSKECWSVCQRLHNTSRGKCMNKKCRCYS (SEQ ID NO: 649)	

Figure 7B (Continued)

Peptide Name	Amino Acid Sequence	Target
Iberiotoxin, IbTx	EFTDVDCSVSKECWSVCKDLFGVDRGKCMGKKCRCYQ (SEQ ID NO: 650)	
Leiurotoxin 1, scyllatoxin	AFCNLRMCQLSCRSLGLLGKCIGDKCECVKH (SEQ ID NO: 651)	
Tamapin	AFCNLRRCELSCRSLGLLGKCIGEECKCVPY (SEQ ID NO: 652)	
Kaliotoxin-1, KTX	GVEINVKCSGSPQCLKPCKDAGMRFGKCMNRKCHCTPK (SEQ ID NO: 653)	
Purotoxin1, PT-1	GYCAEKGIRCDDIHCCTGLKCKCNASGYNCVCRKK (SEQ ID NO: 654)	
GpTx-1	DCLGFMRKCIPDNDKCCRPNLVCSRTHKWCKYVF (SEQ ID NO: 655)	Nav1.7
MOKA Toxin	INVKCSLPQQCIKPCKDAGMRFGKCMNKKCRCYS (SEQ ID NO: 727)	
OSK1, P12, K16, D20	GVIINVKCKISPQCLKPCKDAGMRFGKCMNGKCHCTPK (SEQ ID NO: 728)	Kv1.3
OSK1 K16, D20	GVIINVKCKISRQCLKPCKDAGMRFGKCMNGKCHCTPK (SEQ ID NO: 729)	Kv1.3
HmK	RTCKDLIPVSECTDIRCRTSMKYRLNLCRKTCGSC (SEQ ID NO: 730)	Kv1.3
ShK, K16,Y26, K29	RSCIDTIPKSRCTAFKCKHSMKYRLYFCKKTCGTC (SEQ ID NO: 731)	Kv1.3
ShK, K16	RSCIDTIPKSRCTAFKCKHSMKYRLSFCRKTCGTC (SEQ ID NO: 732)	Kv1.3
ShK-A, K16	RSCIDTIPKSRCTAFKCKHSMKYRLSFCRKTCGTCA (SEQ ID NO: 733)	Kv1.3
ProTxII toxin	SVTLFPPSSEELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPS KQSNNKYA (SEQ ID NO: 754)	
GPTX toxin	ASSYLSLTPEQWKSHRSYSCQVTHEGSTVEKTVAPTECS (SEQ ID NO: 755)	
ShK, Q21	RSCIDTIPKSRCTAFQCKHSQKYRLSFCRKTCGTC (SEQ ID NO: 774)	Kv1.3
ShK, L21	RSCIDTIPKSRCTAFQCKHSLKYRLSFCRKTCGTC (SEQ ID NO: 775)	Kv1.3

Figure 7B (Continued)

Peptide Name	Amino Acid Sequence	Target
ShK, F21	RSCIDTIPKSRCTAFQCKHSFKYRLSFCRKTCGTC (SEQ ID NO: 776)	Kv1.3
ShK, I21	RSCIDTIPKSRCTAFQCKHSIKYRLSFCRKTCGTC (SEQ ID NO: 777)	Kv1.3
ShK, A21	RSCIDTIPKSRCTAFQCKHSAKYRLSFCRKTCGTC (SEQ ID NO: 778)	Kv1.3

Figure 7C

Toxin	Block 1 Seq	Linker A Seq	Block 2 Seq	Linker B Seq	Block 3 Seq
ADWX-1	VGINVKCKHSR (SEQ ID NO: 666)	QC	LKPCKDAGMRFG (SEQ ID NO: 677)	KCM	NGKCHCTPK (SEQ ID NO: 688)
HsTx1	ASCRTPK (SEQ ID NO: 667)	QC	ADPCRKETGCPYG (SEQ ID NO: 678)	KCM	NRKCKCNRC (SEQ ID NO: 689)
OSK1	GVIINVKCKISR (SEQ ID NO: 668)	QC	LEPCKKAGMRFG (SEQ ID NO: 679)	KCM	NGKCHCTPK (SEQ ID NO: 690)
Pi2	TISCTNPK (SEQ ID NO: 669)	QC	YPHCKKETGYPNA (SEQ ID NO: 680)	KCM	NRKCKCFGR (SEQ ID NO: 691)
Hongotoxin (HgTX)	TVIDVKCTSPK (SEQ ID NO: 670)	QC	LPPCKAQFGIRAGA (SEQ ID NO: 681)	KCM	NGKCKCYPH (SEQ ID NO: 692)
Margatoxin	TIINVKCTSPK (SEQ ID NO: 671)	QC	LPPCKAQFGQSAGA (SEQ ID NO: 682)	KCM	NGKCKCYPH (SEQ ID NO: 693)
Agitoxin-2	GVPINVSCTGSP (SEQ ID NO: 672)	QC	IKPCKDAGMRFG (SEQ ID NO: 683)	KCM	NRKCHCTPK (SEQ ID NO: 694)
Pi3	TISCTNEK (SEQ ID NO: 673)	QC	YPHCKKETGYPNA (SEQ ID NO: 684)	KCM	NRKCKCFGR (SEQ ID NO: 695)
Kaliotoxin	GVEINVKCSGSP (SEQ ID NO: 674)	QC	LKPCKDAGMRFG (SEQ ID NO: 685)	KCM	NRKCHCTPK (SEQ ID NO: 696)
Anuroctoxin	ZKECTGPQ (SEQ ID NO: 675)	QC	TNFCRKNKCTHG (SEQ ID NO: 686)	KCM	NRKCKCFNCK (SEQ ID NO: 697)
Charybdotoxin	ZFTNVSCTTSK (SEQ ID NO: 676)	QC	WSVCQRLHNTSRG (SEQ ID NO: 687)	KCM	NKKCRCYS (SEQ ID NO: 698)