



(51) International Patent Classification:

C12N 15/63 (2006.01) *C12N 15/11* (2006.01)
C07K 14/78 (2006.01) *C12N 15/12* (2006.01)
C07K 16/00 (2006.01) *C12N 15/13* (2006.01)
C12N 15/00 (2006.01) *C12N 15/85* (2006.01)
C12N 15/09 (2006.01) *C40B 50/06* (2006.01)

(21) International Application Number:

PCT/CA2013/050203

(22) International Filing Date:

14 March 2013 (14.03.2013)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/610,774 14 March 2012 (14.03.2012) US

(71) Applicant: INNOVATIVE TARGETING SOLUTIONS INC. [CA/CA]; 309 - 4475 Wayburne Drive, Burnaby, British Columbia V5G 4X4 (CA).

(72) Inventors: **GALLO, Michael**; 2650 Colwood Drive, North Vancouver, British Columbia V7R 2R1 (CA). **KANG, Jaspal Singh**; 16727 109 Avenue, Surrey, British Columbia V4N 5H6 (CA). **PIGOTT, Craig Robin**; 210-1775 West 11th Avenue, Vancouver, British Columbia V6J 2C1 (CA). **LIN, Abby**; 1207-6611 Southoaks Crescent, Burnaby, British Columbia V5E 4L5 (CA).

(74) Agents: **SECHLEY, Konrad** et al.; Gowling Lafleur Henderson LLP, 550 Burrard Street, Suite 2300, Vancouver, British Columbia V6C 2B5 (CA).

(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, IS, JP, KE, KG, KM, 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, 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.

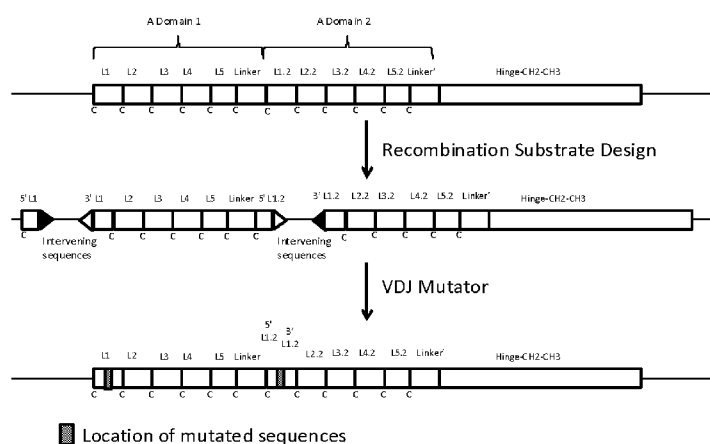
(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, 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, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: GENERATING TARGETED SEQUENCE DIVERSITY IN PROTEINS

FIG. 4 Double A-Domain Substrate Design
Mutagenesis of loop 1 of two separate A domains



(57) **Abstract:** Methods of generating sequence diversity in a protein, such as a ligand-binding protein, are provided. The methods comprise targeted introduction of two or more recombination signal sequences (RSSs) into the protein coding sequence and introduction of the modified protein coding sequence into a recombination-competent host cell, specifically a recombination-competent host cell that is capable of expressing at least RAG-1 and RAG-2, thereby allowing for recombination of the protein coding sequence and expression of variant proteins. Also provided are polynucleotides comprising a nucleic acid sequence encoding a target protein, such as a ligand-binding protein, and comprising two or more RSSs, and compositions and host cells comprising same.

GENERATING TARGETED SEQUENCE DIVERSITY IN PROTEINS

FIELD OF THE INVENTION

[0001] The present invention relates to the field of protein mutagenesis and, in particular, to methods and compositions for targeted protein mutagenesis.

BACKGROUND OF THE INVENTION

[0002] Protein function can be modified and improved *in vitro* by a variety of methods, including site-directed mutagenesis, combinatorial cloning and random
5 mutagenesis combined with an appropriate selection system.

[0003] The method of random mutagenesis together with selection has been used in a number of cases to improve protein function and generally follows one of two strategies. The first involves randomisation of the entire gene sequence in combination with the selection of a variant (mutant) protein with desired characteristics. This
10 process can be repeated on the selected variant until a protein variant is found which is considered optimal. Mutations are typically introduced by error-prone PCR (Leung *et al.*, 1989, *Technique*, 1:11-15) with a mutation rate of approximately 0.7%. The second strategy is to mutagenize defined regions of the gene with degenerate primers (“saturation mutagenesis”), which allows for mutation rates of up to 100% (Griffiths *et al.*, 1994, *EMBO. J.* 13:3245-3260; Yang *et al.*, 1995, *J. Mol. Biol.* 254:392-403),
15 followed by selection of variants with interesting characteristics. The mutated DNA regions from different variants, each with interesting characteristics, may subsequently be combined into one coding sequence (Yang *et al.*, *ibid*).

[0004] Another process for *in vitro* mutation of protein function is “DNA shuffling,”
20 which uses random fragmentation of DNA and assembly of fragments into a functional coding sequence (Stemmer, 1994, *Nature* 370:389-391). The DNA shuffling process generates diversity by recombination, combining useful mutations from individual genes. The genes are randomly fragmented using DNase I and then reassembled by recombination with each other. The starting material can be either a single gene (first

randomly mutated using error-prone PCR) or naturally occurring homologous sequences (so-called family shuffling).

[0005] V(D)J recombination is the process responsible for the assembly of antibody gene segments (V, D and J; or V and J in the case of the light chain) and as part of the assembly process creates the CDR3 of the respective antibody chain. V(D)J recombination can be considered conceptually as a segment shuffler for antibodies, *i.e.* it brings together the different VH segments, D segments and JH segments to create an antibody (similarly V(D)J recombination at the light chain assembles different combinations of light chain V and J segments at either the kappa or lambda locus). The recombination event results in large chromosomal deletions in order to bring the required segments together. V(D)J recombination is targeted by the presence of specific DNA sequences called the recombination signal sequences (RSSs). The recombination reaction involves the recombination proteins RAG-1 and RAG-2 and follows a 12/23 rule where an RSS with a 23 bp spacer is paired only with an RSS with 12 bp spacer and adjacent sequences are subsequently joined by double-stranded break repair proteins.

[0006] The V(D)J recombination reaction is responsible for the creation of CDR3, as it is the sole mechanism for gene segment assembly and antibody generation in the bone marrow. V(D)J recombination does not occur at CDR1 or CDR2. V(D)J recombination therefore is not involved in affinity maturation but in primary B cell development and antibody assembly.

[0007] U.S. Patent No. 8,012,714 describes compositions and methods for generating sequence diversity in the CDR3 region of *de novo* generated immunoglobulins *in vitro*. The methods comprise constructing nucleic acid molecules that comprise polynucleotide sequences encoding immunoglobulin V, D, J and C regions, together with recombination signal sequences (RSS), and subsequently introducing these nucleic acid molecules into suitable recombination-competent host cells. The methods provide for the assembly of gene segments to generate a functional antibody *in vitro*.

[0008] The use of "protein scaffolds" for the generation of novel binding proteins via combinatorial engineering has recently emerged as a powerful alternative to natural or recombinant antibodies. It has been found that novel binding sites can be introduced

into proteins from several protein families with non-Ig architectures by combinatorial engineering, such as site-directed random mutagenesis combined with phage display or other selection techniques (Rothe, A., *et al.*, 2006, *FASEB J.*, 20:1599-1610). This concept requires a stable protein architecture ("scaffold") tolerating multiple
5 substitutions or insertions at the primary structural level (see reviews by Binz, H.K., *et al.*, 2005, *Nature Biotechnology*, 23(10):1257-1268; Nygren, P-A. & Skerra, A., 2004, *J Immunol. Methods*, 290:3-28, and Gebauer, M. & Skerra, A., 2009, *Curr. Op. Chem. Biol.*, 13:245-255).

[0009] This background information is provided for the purpose of making known
10 information believed by the applicant to be of possible relevance to the present invention. No admission is necessarily intended, nor should be construed, that any of the preceding information constitutes prior art against the present invention.

SUMMARY OF THE INVENTION

[0010] An object of the present invention is to provide methods and compositions for generating sequence diversity in proteins. In accordance with an aspect of the
15 invention, there is provided a method of generating variants of a target protein comprising the steps of: (i) providing a polynucleotide comprising a nucleic acid sequence encoding the target protein and comprising a sequence cassette that comprises a first recombination signal sequence (RSS) linked by an intervening nucleotide sequence to a second RSS, the first RSS capable of functional recombination with the
20 second RSS, wherein the first and second RSS are located in a portion of the nucleic acid sequence that encodes a non-conformational region of the target protein, and wherein the first intervening nucleotide sequence is 100 base pairs or more in length; (ii) introducing the polynucleotide into a recombination-competent host cell, and (iii) culturing the host cell *in vitro* under conditions allowing recombination and expression
25 of the nucleic acid sequence, thereby generating variants of the target protein.

[0011] In accordance with another aspect of the invention, there is provided a polynucleotide comprising a nucleic acid sequence encoding a target protein and comprising a sequence cassette that comprises a first recombination signal sequence (RSS) linked by an intervening nucleotide sequence to a second RSS, the first RSS

capable of functional recombination with the second RSS, wherein the first and second RSS are located in a portion of the nucleic acid sequence that encodes a non-conformational region of the target protein, and wherein the intervening nucleotide sequence is 100 base pairs or more in length.

- 5 [0012] In accordance with another aspect, there is provided an isolated host cell comprising a polynucleotide of the invention.

[0013] In accordance with another aspect, there is provided a variant protein produced by the methods described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

- 10 [0014] These and other features of the invention will become more apparent in the following detailed description in which reference is made to the appended drawings.

[0015] **Figure 1** presents a partial nucleotide sequence of avimer construct E188 that comprises a single avimer A domain, a pair of RSSs introduced into loop 1 of the construct and a pair of RSSs introduced into loop 2 of the construct together with flanking sequences encoding GY amino acid residues [SEQ ID NO:28].

- 15 [0016] **Figure 2** presents a partial nucleotide sequence of avimer construct E189 that comprises double avimer A domains and a pair of RSSs in each loop 1 of the construct, as well as stop codons in other reading frames in the 3' loop 1.1 to 5' loop 1.2 region [SEQ ID NO:29].

- 20 [0017] **Figure 3** presents (A) a schematic representation of a single domain A avimer construct comprising a pair of RSSs in loop 1 and a pair of RSSs in loop 2, a selectable marker was included between the Tm domain and the poly A; (B) sequence details of the construct shown in (A) with arrows indicating the positions of insertion of the RSS cassettes, and (C) a schematic overview of the steps for mutagenesis of the single domain A avimer construct shown in (A).

- 25 [0018] **Figure 4** presents a schematic overview of the steps for mutagenesis of a double domain A avimer construct including RSS sequences in each loop 1.

[0019] **Figure 5** presents a schematic representation of single, double and triple A domain avimer constructs.

[0020] **Figure 6** depicts (A) the nucleotide and amino acid sequences of RSS flanked cassettes used to introduce sequence diversity into avimer sequences [SEQ ID NOs:86 & 87], and (B) the nucleotide and amino acid sequences of RSS flanked cassettes used to introduce sequence diversity into avimer sequences in which the CCA nucleotides have been changed to TGT introducing cysteines in two additional reading frames (nucleotide sequence: SEQ ID NO:88; amino acid sequences: SEQ ID NOs:89-91).

[0021] **Figure 7** depicts (A) the nucleotide sequence containing the 10Fn3 coding sequence [SEQ ID NO:34] used in the preparation of 10Fn3 constructs, (B) the amino acid sequence encoded by the nucleotide sequence shown in (A) [SEQ ID NO:39], (C) a schematic representation of the acceptor vector used in the construction of the 10Fn3 constructs and for CDR diversification, and (D) the nucleotide sequences for the vector represented in (C) [SEQ ID NO:35] (BsaI and KpnI restriction sites are bolded).

[0022] **Figure 8** presents (A) a schematic representation of the location of the loop regions (BC, DE and FG) of 10Fn3, and (B) the nucleotide and amino acid sequences of 10Fn3 with the loop regions indicated [SEQ ID NOs: 92 & 93].

[0023] **Figure 9** presents a schematic representation of (A) a bipartite 10Fn3 construct comprising a single pair of RSSs in the FG loop, (B) sequence details of the 23bp RSS shown in (A) [SEQ ID NO:94], (C) sequence details of the 12bp RSS shown in (A) [SEQ ID NO:95], and (D) an overview of the steps for mutagenesis of the 10Fn3 construct shown in (A).

[0024] **Figure 10** presents (A) the sequence of the construct shown in Figure 9A [SEQ ID NO:37], in which the 23bp RSS and the 12bp RSS are shown in bold, and (B) the sequence of the construct shown in Figure 11A [SEQ ID NO:38], in which the 23bp RSS and the 12bp RSS are shown in bold.

[0025] **Figure 11** presents a schematic representation of (A) a tripartite 10Fn3 construct comprising two pairs of RSSs in the FG loop, (B) sequence details of the 5' 23bp RSS shown in (A) [SEQ ID NO:96], (C) sequence details of the 5' 12bp RSS and

the 3' 12bp RSS shown in (A) [SEQ ID NO:97] and the encoded FG loop sequences [SEQ ID NOs:98-100], (D) sequence details of the 3' 23bp RSS shown in (A) [SEQ ID NO:101], and (E) an overview of the steps for mutagenesis of the 10Fn3 construct shown in (A).

- 5 [0026] **Figure 12** presents (A) a schematic overview of the steps for mutagenesis of a tripartite 10Fn3 construct that allow for simultaneous diversification of the DE and FG loops, and (B) a schematic overview of the steps for mutagenesis of a tripartite 10Fn3 construct that allow for simultaneous diversification of the BC and FG loops.

- 10 [0027] **Figure 13** presents (A) a schematic representation of the positions selected within CDR1 and CDR2 of an immunoglobulin heavy chain for insertion of pairs of RSSs, (B) sequence details for the 5' and 3' junctions selected for the CDR1 RSS placement [SEQ ID NO:102], and (C) sequence details for the 5' and 3' junctions selected for the CDR2 RSS placement [SEQ ID NO:103].

- 15 [0028] **Figure 14** presents (A) a schematic representation of the immunoglobulin heavy chain shown in Figure 13 including a pair of RSSs in CDR2, (B) sequence details of the 23bp RSS shown in (A) [SEQ ID NO:104], and (C) sequence details of the 12bp RSS shown in (A) [SEQ ID NO:105].

- 20 [0029] **Figure 15** presents (A) a schematic representation of the immunoglobulin heavy chain shown in Figure 13 including a pair of RSSs, together with 5' trinucleotide repeat flanking sequences, in each of CDR1 and CDR2, (B) sequence details of the 5' 23bp RSS shown in (A) [SEQ ID NO:106], (C) sequence details of the 5' 12bp RSS shown in (A) [SEQ ID NO:107], (D) sequence details of the 3' 12bp RSS shown in (A) [SEQ ID NO:108], and (E) sequence details of the 3' 23bp RSS shown in (A) [SEQ ID NO:109].

- 25 [0030] **Figure 16** shows (A) the nucleotide sequence of the unmodified immunoglobulin heavy chain depicted schematically in Figure 13 [SEQ ID NO:3] with the CDR1 and CDR2 regions shown in bold, (B) the nucleotide sequence of the immunoglobulin heavy chain including a pair of complementary RSSs positioned within CDR2 as depicted schematically in Figure 14A [SEQ ID NO:4] with the RSSs
30 shown in bold, and (C) the nucleotide sequence of the immunoglobulin heavy chain

including a pair of RSSs positioned, together with 5' trinucleotide repeat flanking sequences, within each of CDR1 and CDR2 as depicted schematically in Figure 15A [SEQ ID NO:5], with the RSSs shown in bold.

5 [0031] **Figure 17** presents the nucleotide sequence for the vector E188 [SEQ ID NO:1].

[0032] **Figure 18** presents the nucleotide sequence for the vector E189 [SEQ ID NO:40].

10 [0033] **Figure 19** presents (A) a schematic representation of a cassette for generating in-frame selection of a secreted protein (shown is Ig Kappa) showing from constant region to poly(A), and (B) the nucleotide sequence of the cassette shown in (A) [SEQ ID NO:41] with the furin cleavage site in bold.

[0034] **Figure 20** presents the nucleotide sequence for the vector ITS001-V655 [SEQ ID NO:74].

15 [0035] **Figure 21** presents (A) a schematic diagram of the light chain CDR2 optimization cassette from the vector ITS001-V655, and (B) the results from FACS-based analysis of HER2 binding versus antibody expression for a population of cells expressing light chain CDR2 optimized antibodies (ITS001-L145) as compared to cells expressing the original antibody.

20 [0036] **Figure 22** presents (A) a bar chart summarizing the results of a FACS-based assay of HER2 binding to cells expressing light chain CDR2 optimized antibodies cloned from ITS001-L145 relative to cells expressing the original antibody, and (B) results from the FACS-based assay for an individual clone (Clone 9) as compared to the original antibody.

25 [0037] **Figure 23** presents (A) the results from FACS-based analysis of HER2 binding versus antibody expression for a population of cells expressing light chain CDR1 optimized antibodies (ITS001-L167) as compared to cells expressing the original antibody, (B) a bar chart summarizing the results of a FACS-based assay of HER2 binding to cells expressing light chain CDR1 optimized antibodies cloned from ITS001-L167 relative to cells expressing the original antibody, and (C) results from the

FACS-based assay for an individual clone (Clone 15) as compared to the original antibody.

DETAILED DESCRIPTION OF THE INVENTION

[0038] The present invention is based on the finding, illustrated herein, that the use of components of the antibody V(D)J recombination system can be expanded outside their natural role of mediating assembly of antibody gene segments to their use to modify an existing protein sequence.

[0039] Accordingly, in certain embodiments, the invention relates to methods of generating sequence diversity in a known protein sequence, such as a ligand-binding protein sequence, by targeted introduction of two or more recombination signal sequences (RSSs) into the protein coding sequence and subsequent introduction of the modified protein coding sequence into a recombination-competent host cell, specifically a host cell that is capable of expressing at least RAG-1 and RAG-2, resulting in the generation and expression of variants of the protein. In certain embodiments, the present invention relates to polynucleotides comprising a nucleic acid sequence encoding a target protein, such as a ligand-binding protein, and comprising two or more RSSs, and compositions comprising same.

[0040] The present invention recognizes that the natural V(D)J reaction has inherent characteristics, specifically the imprecise junctions generated during the joining process, that make it useful as a general means to generate sequence diversity. The use of V(D)J recombination as a method to modify an existing protein sequence as opposed to assembly of a protein from gene segments, however, has a number of challenges, including a number of features of the reaction that are under-appreciated in the art.

[0041] The V(D)J recombination reaction is known to bring together different DNA sequences and result in large chromosomal deletions, which suggests that its utility to introduce sequence diversity would be limited to extended stretches of nucleic acid sequence that permit such large deletions. As demonstrated herein, however, the components of the V(D)J recombination system can be manipulated to allow the utility of this reaction to be extended to include targeted sequences within a restricted size of protein sequence, such as a small loop.

[0042] In addition, although the involvement of the enzyme TdT, which is responsible for non-template nucleotide additions (N-additions), is central to the reaction, the net size of the product following gene segment assembly is frequently less than would be predicted if no deletions or additions were to occur, *i.e.* the V(D)J reaction often results in a net loss of sequence. For example, the average size of the assembled germline V, D and J segments, without any additions or deletions, is 15 amino acids and yet the average CDR3 reported in humans is 12-13 amino acids, which includes N additions from TdT (Rock *et al.*, 1994, *J Exp Med*, 179:323-328).

[0043] Another feature of V(D)J recombination that is under-appreciated is that the additions introduced by TdT are small. *In vivo* and *in vitro* TdT additions have been reported to be typically an average of 2-4 nucleotides (Kallenbach *et al.*, 1992, *PNAS USA*, 89:2799-2903; Bentolila *et al.*, 1997, *J Immunol.*, 158:715-723). An larger number of amino acid changes per variant is generally preferred for mutagenesis techniques in order to allow for a greater amount of diversity to be sampled.

[0044] The above-noted features of V(D)J recombination can represent challenges to the application of V(D)J recombination to a non-antibody scaffold. The methods provided by the present invention, however, allow for this random deletional process to be used as a valuable tool for semi-rational protein engineering.

[0045] In some embodiments, for example, the methods employ flanking sequences adjacent to one or more of the RSSs to allow for incorporation of additional sequences into the final variant protein to minimise any net deletion effect of the V(D)J recombination reaction and/or to introduce additional functionality by way of addition of specific amino acid residues. By way of example, when the targeted location is within a small loop of a protein, flanking sequences may be used in conjunction with the RSSs to ensure that the loop retains a minimal length once sequence diversification has taken place.

[0046] The V(D)J reaction *in vivo* generates deletions and additions of different size and composition on either side of the junctions flanked by the RSSs. In certain embodiments, the methods of the present invention allow for control of the reaction so that deletions can be focused to one junction or the other through the use of flanking sequences. In some embodiments, the methods allow for specific heterologous

sequences to be incorporated into the final variant protein through the use of flanking sequences.

[0047] In addition, in some embodiments, the methods make use of a tripartite reaction that involves two pairs of RSSs so that diversity is generated at two junctions rather than a single junction. In accordance with those embodiments in which a tripartite reaction is employed, sequence diversity may be introduced at a single target location in the protein, or at two independent locations in the protein. Use of a tripartite reaction with an appropriately sized RSS flanked donor cassette sequence also allows for the incorporation of sequences from the donor cassette at the targeted location. In certain embodiments, the methods provide for sequence diversity to be introduced at a single location by way of a “bipartite” reaction that involves a single pair of RSSs, which may be used with or without flanking sequences.

[0048] The methods in accordance with the present invention have a number of features that make them attractive for generating sequence diversity. For example, the diversity can be targeted so that mutations are focused at one or more predetermined locations as opposed to being randomly distributed across a protein as would be the result of traditional approaches, such as PCR- or somatic hypermutation (SHM)-based approaches. The methods may also be used to simultaneously introduce mutations at two different target locations within the protein. These locations may be close together or distant, in terms of either sequence or structure. For example, in certain embodiments, the methods are used to simultaneously introduce sequence diversity simultaneously into two separate loops of a target protein.

[0049] In addition, in certain embodiments, the methods of the present invention allow for the generation of both composition and length diversity simultaneously. In some embodiments, the methods are entirely cell-mediated thus eliminating the requirement for cloning of variants and their subsequent introduction into cells as is required by other methods.

[0050] The methods of the present invention additionally allow for the generation of a very large number of protein variants such that, in certain embodiments, mutations imparting the desired functionality to the protein can be identified in a single round. For example, for binding proteins, the attainable affinity from a library of random

binding proteins is assumed to increase with its diversity (Griffiths, A.D., *et al.*, 1994, *EMBO J.*, 13:3245-3260). Accordingly, the methods in accordance with certain embodiments of the present invention provide for the generation of sufficient diversity within a target binding protein to allow for variants with high affinity for a selected
5 ligand to be generated in a single round.

[0051] The methods in accordance with certain embodiments of the invention include the use of flanking sequences adjacent to the RSSs and/or tripartite substrate structures to allow for the production of a large repertoire of functional variants.

[0052] In certain embodiments, the methods employ an inducible form of one or more
10 of the components of the recombination system to allow induction of sequence diversity generation to be controlled, for example to allow for expansion of the host cell prior to induction of sequence diversity generation.

[0053] In general, the methods comprise the steps of introducing a pair of RSSs at a selected location within the coding sequence for a target protein, introducing the
15 modified coding sequence into a cell that is capable of expressing at least RAG-1 and RAG-2 to allow for recombination and expression of the variant protein.

Definitions

[0054] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which
20 this invention belongs.

[0055] “Naturally occurring,” as used herein with reference to an object, refers to the fact that the object can be found in nature. For example, an organism, or a polypeptide or polynucleotide sequence that is present in an organism that can be isolated from a source in nature and which has not been intentionally modified by man in the
25 laboratory is naturally occurring.

[0056] The term “isolated,” as used herein with reference to a material, means that the material is removed from its original environment (for example, the natural environment if it is naturally occurring). For example, a naturally occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same

polynucleotide or polypeptide separated from some or all of the co-existing materials in the natural system, is isolated. Such polynucleotides could be part of a vector and/or such polynucleotides or polypeptides could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment.

- 5 **[0057]** The term “gene,” as used herein, refers to a segment of DNA involved in producing a polypeptide chain. The segment of DNA may include regions preceding and/or following the coding region, as well as intervening sequences (introns) between individual coding segments (exons), and may also include regulatory elements (for example, promoters, enhancers, repressor binding sites and the like).
- 10 **[0058]** The term “deletion” as used herein with reference to a polynucleotide, polypeptide or protein has its common meaning as understood by those familiar with the art and may refer to molecules that lack one or more of a portion of a sequence from either terminus or from a non-terminal region, relative to a corresponding full length molecule. For example, in certain embodiments, a deletion may be a deletion of
- 15 between 1 and about 1500 contiguous nucleotide or amino acid residues from the full length sequence.
- 20 **[0059]** The term “expression vector,” as used herein, refers to a vehicle used in a recombinant expression system for the purpose of expressing a polynucleotide sequence constitutively or inducibly in a host cell, including prokaryotic, yeast, fungal, plant, insect or mammalian host cells, either *in vitro* or *in vivo*. The term includes both linear and circular expression systems. The term includes expression systems that remain episomal and expression systems that integrate into the host cell genome. The expression systems can have the ability to self-replicate or they may not (for example, they may drive only transient expression in a cell).
- 25 **[0060]** The term “antigen-binding domain,” as used herein, refers to one or more fragments of an antibody that retain the ability to specifically bind to an antigen. Non-limiting examples of antibody fragments comprising antigen-binding domains include, but are not limited to, (i) a Fab fragment, a monovalent fragment consisting of the V_L, V_H, C_L and C_{H1} domains; (ii) a F(ab')₂ fragment, a bivalent fragment comprising two
- 30 Fab fragments linked by a disulphide bridge at the hinge region; (iii) a Fd fragment consisting of the V_H and C_{H1} domains; (iv) a Fv fragment consisting of the V_L and V_H

domains of a single arm of an antibody, (v) a dAb fragment, which consists of a V_H domain; and (vi) an isolated complementarity determining region (CDR). The term also encompasses single chain Fv (scFv) fragments, which comprise the two domains of the Fv fragment, V_L and V_H, joined using recombinant methods by a synthetic linker that
5 enables them to be made as a single protein chain in which the V_L and V_H regions pair to form monovalent molecules.

[0061] The term “bipartite reaction,” as used herein, refers to a recombination reaction that involves a single pair of RSSs (12 bp and 23 bp, or 23 bp and 12 bp). When V(D)J recombination occurs it generates a double-stranded break in the nucleic
10 acid sequence containing the RSSs. The double-stranded break is targeted as a result of the RSSs in that a 12 bp and 23 bp RSS are assembled with the RAG proteins to initiate the reaction. The ends of the DNA that will be subsequently rejoined will comprise the coding joint (or junction). An example of a bipartite reaction is *in vivo* immunoglobulin light chain recombination, which joins the Variable to the Joining
15 segment - these two segments comprise the “substrates” for the bipartite reaction. The bipartite reaction can occur in the presence or absence of TdT.

[0062] The term “tripartite reaction,” as used herein, refers to a recombination reaction that involves two pairs of RSSs (each 12 bp and 23 bp, or 23 bp and 12 bp). An example of a tripartite reaction is *in vivo* immunoglobulin heavy chain
20 recombination, which joins the V, the D and the J gene segments. A tripartite reaction generates two independent coding junctions. Two sequential bipartite reactions can be considered to be a tripartite reaction in that a tripartite reaction may comprise two bipartite reactions occurring in the same substrate, usually (but not always) in close temporal time. The tripartite reaction can occur in the presence or absence of TdT.

[0063] The term “recombination-competent” when used herein with reference to a host cell means that the host cell is capable of mediating RAG-1/RAG-2
25 recombination. The host cell may, therefore, express RAG-1 and RAG-2, or functional fragments thereof, or may be modified (for example, transformed or transfected with appropriate genetic constructs) such that it expresses RAG-1 and RAG-2, or functional
30 fragments thereof. The expression of one or both of RAG-1 and RAG-2 in the recombination-competent host cell may be constitutive or it may be inducible. A

recombination-competent host cell may optionally further express TdT, or a functional fragment thereof.

[0064] As used herein, the term “about” refers to an approximately +/-10% variation from a given value. It is to be understood that such a variation is always included in any
5 given value provided herein, whether or not it is specifically referred to.

[0065] The term “plurality” as used herein means more than one, for example, two or more, three or more, four or more, and the like.

METHODS OF GENERATING SEQUENCE DIVERSITY

[0066] The methods according to the present invention generally comprise the steps
10 of introducing a pair of RSSs at a selected location within the coding sequence for a target protein, and introducing the modified coding sequence into a recombination-competent host cell to allow for recombination and expression of variants of the target protein. Accordingly, in its simpler aspects, the present invention provides methods of generating variants of a target protein comprising the steps of: providing a
15 polynucleotide comprising a nucleic acid sequence encoding a target protein and comprising a complementary pair of RSSs, introducing the polynucleotide into a recombination-competent host cell, the host cell capable of expressing at least RAG-1 and RAG-2, and culturing the host cell *in vitro* under conditions allowing recombination and expression of the polynucleotide, thereby generating variants of the
20 target protein. In certain embodiments, the methods further comprise screening the variant proteins for variants having defined functional characteristics.

[0067] In certain embodiments of the present invention, the methods are applied to a target protein that is a ligand-binding protein. In some embodiments, the methods are applied to a ligand-binding protein in order to introduce sequence diversity into a loop
25 region involved in ligand-binding and comprise the steps of: providing a polynucleotide comprising a nucleic acid sequence encoding a target ligand-binding protein, the nucleic acid sequence comprising a complementary pair of RSSs in a region of the sequence encoding a ligand-binding loop of the protein, introducing the polynucleotide into a recombination-competent host cell, and culturing the host cell under conditions

allowing recombination and expression of the polynucleotide, thereby generating variants of the target ligand-binding protein.

[0068] The host cell may constitutively express RAG-1 and RAG-2, and optionally TdT, or one or more of these proteins may be under inducible control. In certain
5 embodiments, expression of one or more of RAG-1 and RAG-2, and optionally TdT, in the host cell is under inducible control allowing, for example, for expansion of the host cell prior to the induction of sequence diversity generation. Accordingly, in some embodiments, the method comprises the steps of: providing a polynucleotide comprising a nucleic acid sequence encoding a target protein and comprising a pair of
10 RSSs, introducing the polynucleotide into a recombination-competent host cell, the host cell capable of expressing at least RAG-1 and RAG-2 and optionally TdT, wherein expression of one or more of RAG-1, RAG-2 and TdT is under inducible control, culturing the host cell under conditions allowing expansion of the host cell, inducing expression of one or more of RAG-1, RAG-2 and TdT, culturing the expanded host
15 cells under conditions allowing recombination and expression of the polynucleotide, thereby generating variants of the target protein.

[0069] The polynucleotide may be introduced into the host cell on a suitable vector and may be, for example, stably integrated into the genome of the cell, stably maintained exogenously to the genome or transiently expressed.

20 [0070] In some embodiments, the nucleic acid encoding the target protein comprised by the polynucleotide is operably linked to a regulatable promoter, for example, an inducible promoter, such that expression of the target protein can be controlled.

[0071] In certain embodiments, the polynucleotide may comprise additional pairs of RSSs allowing for generation of additional sequence diversity in the protein. In some
25 embodiments, the polynucleotide comprises two complementary pairs of RSSs, each pair positioned to introduce sequence diversity into a different region of the target protein.

[0072] In some embodiments, the polynucleotide may also comprise additional coding sequences and thus may encode a fusion protein comprising the target protein
30 fused to a polypeptide that provides additional functionality to the protein. For

example, the polypeptide may localize the target protein to the cell membrane, nucleus or other organelle; provide for secretion of the target protein from the cell; introduce a detectable label, or the like.

[0073] In certain embodiments, the recombination is controlled. In some
5 embodiments, the host cell is capable of cell divisions without recombination. As
described herein, these and related embodiments permit expansion of the host cell
population prior to the initiation of recombination events that give rise to sequence
diversity in the target protein. Control of recombination in such host cells may be
achieved, for example, through the use of an operably linked recombination control
10 element (such as an inducible recombination control element, which may be a tightly
regulated inducible recombination control element), and/or through the use of one or
more low efficiency RSSs in the polynucleotide (as described in more detail below),
and/or through the use of low host cell expression levels of one or more of RAG-1 or
RAG-2, and/or through design of the polynucleotide to integrate at a chromosomal
15 integration site offering poor accessibility to host cell recombination mechanisms (for
example, RAG-1 and/or RAG-2).

[0074] In some embodiments, the methods further comprise selecting a variant having
the desired functional characteristics. In some embodiments, the methods further
comprise subjecting a selected variant to one or more additional rounds of sequence
20 diversity generation in order to obtain further variants having optimised functional
characteristics.

Target Proteins

[0075] In accordance with the present invention, the methods of generating sequence
diversity may be applied to a wide variety of proteins for which a functional assay can
25 be designed for screening. In accordance with certain embodiments of the invention,
the target protein of the methods is preferably a ligand-binding protein, wherein the
ligand may be an antigen, another protein, a nucleic acid, a carbohydrate, a lipid, a
metal, a vitamin or the like. In the context of the present invention, the term “ligand-
binding protein” includes receptor-binding proteins. In some embodiments, the target
30 protein is a ligand-binding protein, wherein the ligand is another protein, a nucleic acid,
a carbohydrate, a lipid, a vitamin or a metal. In some embodiments, the target protein is

a ligand-binding protein, wherein the ligand is another protein. In certain embodiments, the target protein is a ligand-binding protein, wherein the ligand is an antigen. In some embodiments, the target protein is a receptor-binding protein.

[0076] In some embodiments, the target protein of the methods is an immunoglobulin, wherein the target location(s) for introduction of sequence diversity are the CDR1 and/or CDR2 region. The natural process of B cell development does not involve V(D)J recombination of CDR1 and/or CDR2. As described herein, however, the use of components of the antibody V(D)J recombination system can be expanded to introduce sequence diversity at CDR1 and/or CDR2 in order to generate antibodies with improved affinity and/or specificity.

[0077] In certain embodiments, the target protein is an existing immunoglobulin and the target location is CDR3, wherein the existing CDR3 can be sequence diversified to generate improved binding characteristics (for example improved affinity or specificity) over the original immunoglobulin.

[0078] The immunoglobulin may comprise a germline sequence or it may comprise a sequence that has already undergone affinity maturation or one or more artificial sequence optimization steps to improve the affinity of the immunoglobulin. Accordingly, in some embodiments, the methods can be used to improve the affinity of a germline immunoglobulin, and in some embodiments, the methods can be used to further improve the affinity of a known immunoglobulin.

[0079] In certain embodiments, the target protein of the methods is an immunoglobulin, wherein the target location(s) for introduction of sequence diversity is a non-CDR loop of the Ig molecule located in the constant region of the protein.

[0080] Immunoglobulins that may be used as target proteins in the methods of the present invention are antibodies or antibody fragments that comprise an antigen-binding domain including at least one of a CDR1, CDR2 or CDR3. Examples include, but are not limited to, IgA, IgA2, IgD, IgE, IgGs (*i.e.* IgG1, IgG2, IgG3 and/or IgG4) and IgM antibodies; camelid antibodies; HCAbs; single chain antibodies; shark antibodies; antibody fragments such as Fab, Fab', F(ab')₂, Fd, Fv and single-chain Fv (scFv) antibody fragments; diabodies, nanobodies and fluorobodies. In certain

embodiments, the target protein is an IgG antibody or fragment thereof. In some embodiments, the target protein is an immunoglobulin V_H chain or V_L chain.

[0081] Immunoglobulins suitable for use in the methods described herein may be derived from a variety of sources and technologies including, but not limited to, mammals including mice, transgenic mice and humans, phage display or yeast display, or they may be synthetically derived immunoglobulins or fragments thereof.

[0082] In one embodiment, the methods are applied to a target protein that is a non-immunoglobulin protein.

[0083] Non-immunoglobulin ligand-binding proteins that may be used as target proteins include naturally-occurring proteins and non-naturally occurring proteins. Naturally-occurring ligand-binding proteins include human proteins and non-human proteins, for example, proteins from a non-human animal, a plant, or a micro-organism. Examples of naturally-occurring ligand-binding proteins include, but are not limited to, biotin-binding proteins (such as avidin and streptavidin), lipid-binding proteins (such as β -lactoglobulin, α_1 -microglobulin and plasma transthyretin), periplasmic binding proteins, lectins, serum albumins, phosphate binding proteins, sulphate binding proteins, immunophilins, metal-binding proteins, DNA-binding proteins, GTP-binding proteins (G-proteins), transporter proteins and receptor proteins (soluble and non-soluble). Non-limiting examples of metal-binding proteins include transferrin, ferritin and metallothionein. Non-limiting examples of DNA-binding proteins include histones, transcription factors, single-stranded DNA-binding proteins and helicases. Non-limiting examples of transporter and receptor proteins include, haemoglobin, cytochromes, G-protein coupled receptors, adrenalin receptors, acetylcholine receptors, histamine receptors, dopamine receptors, serotonin receptors, glutamate receptors, serotonin transporters, oestrogen receptors, Ca²⁺ channels, Na⁺ channels and Cl⁻ channels. Non-limiting examples of soluble receptors include receptors for peptide hormones or cytokines, such as receptors for growth factors, lymphokines, monokines, interleukins, interferons, chemokines, colony-stimulating factors, hematopoietic factors, neurotrophic factors and differentiation-inhibiting factors. In certain embodiments, the ligand-binding protein may be a T-cell receptor.

[0084] Non-naturally occurring ligand-binding proteins include polypeptides that comprise one or more ligand-binding domains or fragments of naturally-occurring proteins capable of binding a ligand, such as fibronectin III domains (for example, FN3 and adnectins™), the immunoglobulin binding domain of *Staphylococcus aureus* protein A (“affibodies”), src homology domains 2 and 3 (SH2 and SH3, respectively) and PDZ domains. Non-naturally occurring ligand-binding proteins also include artificial ligand-binding proteins such as designed ankyrin repeat proteins (“DARPs”), avimers and aptamers.

[0085] In certain embodiments, the non-naturally occurring ligand-binding proteins are protein scaffolds consisting of a stably folded non-Ig protein capable of being equipped with a binding site as described in Binz, *et al.* (2005, *Nature Biotechnology*, 23(10):1257-1268), Nygren & Skerra (2004, *J Immunol. Methods*, 290:3-28) and Gebauer & Skerra (2009, *Curr. Op. Chem. Biol.*, 13:245-255). Examples of such protein scaffolds include, but are not limited to, cytotoxic lymphocyte-associated antigen-4 (CTLA-4), Tendamistat, 10th fibronectin type 3 domain (10FN3), carbohydrate-binding module 4 of family 2 of xylanase of *Rhodothermus marinus* (CBM4-2), lipocalins (“anticalins”), T-cell receptor, Protein A domain (protein Z), immunity protein 9 (Im9), designed ankyrin repeat proteins (DARPs), designed tetratricopeptide repeat (TPR) proteins, zinc finger proteins, protein VIII of filamentous bacteriophage (pVIII), avian pancreatic polypeptide, general control nonderepressible (yeast transcription factor) (GCN4), WW domain, Src homology domain 3 (SH3), Src homology domain 2 (SH2), PDZ domains, TEM-1, β -lactamase, green fluorescent protein (GFP), thioredoxin, staphylococcal nuclease, plant homeodomain finger protein (PHD-finger), chymotrypsin inhibitor 2 (CI-2), bovine pancreatic trypsin inhibitor (BPTI), Alzheimer amyloid β -protein precursor inhibitor (APPI), human pancreatic secretory trypsin inhibitor (hPSTI), ecotin, human lipoprotein-associated coagulation inhibitor domain 1 (LACI-D1), leech-derived trypsin inhibitor (LDTI), MTI-II, scorpion toxins, insect defensin A peptide, *Ecballium elaterium* trypsin inhibitor II (EETI-II), Min-23, cellulose-binding domain (CBD), periplasmic binding proteins (PBP), cytochrome b₅₆₂, low density lipoprotein (ldl) receptor domain A, γ -crystallin, ubiquitin, transferrin and C-type lectin-like domain.

[0086] Protein scaffolds can be considered as falling into two groups: a first group consisting of loop presenting scaffolds (which includes scaffolds presenting a single loop and scaffolds presenting a plurality of loops), and a second group consisting of interface presenting scaffolds, in which the binding site is presented on a secondary structure element. Examples of scaffolds in the first group include, but are not limited to, Kunitz domain inhibitors, hPSTI, APPI, LACI-D1, ecotin, members of the knottin family of proteins (such as EETI-II), thioredoxin, staphylococcal nuclease, immunoglobulins, CTLA-4, FN3, Tendamistat, GFP, members of the lipocalin family of proteins, and bilin binding protein (BBP) from *Pieris brassicae*. Examples in the second group include, but are not limited to, the immunoglobulin binding domain of Staphylococcal protein A (SPA) ("affibodies"), DARPin, leucine-rich repeat polypeptides, PDZ domains, cellulose binding domains (CBD), members of the lipocalin family of proteins, γ -crystallins, and Cys₂His₂ zinc-finger polypeptides. The binding domains of both of these groups of proteins have been studied and regions suitable for modification have been identified (see review by Nygren & Skerra, *ibid.*). The present invention contemplates that in various embodiments, the methods described herein may be applied to both loop presenting scaffolds and to interface presenting scaffolds. In certain embodiments, therefore, the target protein for the methods is a loop presenting scaffold protein, wherein sequence diversity is introduced into one or more loops. In some embodiments, the target protein for the methods is an interface presenting scaffold protein, in which sequence diversity is introduced into the binding site.

[0087] In certain embodiments, the methods are applied to target proteins that comprise a ligand-binding region that includes one or more loops, in which a loop can be defined as a region supported by a protein scaffold that can carry altered amino acids or sequence insertions without substantially compromising the structure of the scaffold, and wherein sequence diversity is introduced into one or more of the loops. In some embodiments, the methods are applied to target proteins that comprise a ligand-binding region that includes one or more surface-exposed loops, wherein one or more of the loops are targeted locations for introduction of sequence diversity. Examples of loop containing proteins are found within various categories of proteins described above and include, for example, immunoglobulins and loop presenting scaffold proteins.

[0088] While the term “target protein” is used herein, it is to be understood that the methods of the present invention are equally applicable to protein fragments and that the term “target protein” thus incorporates both the full length protein and fragments of the protein, for example, functional fragments, fragments comprising one or more domains, and the like. In certain embodiments, fragments include one or more deletions from either terminus of the protein or a deletion from a non-terminal region of the protein, for example, in some embodiments, deletions of between about 1 and about 500 contiguous amino acid residues. In some embodiments, the fragments may comprise a deletion of between about 1 and about 300 contiguous amino acid residues, for example, between 1 and about 250 contiguous amino acid residues, between 1 and about 200 contiguous amino acid residues, between 1 and about 150 contiguous amino acid residues, between 1 and about 100 contiguous amino acid residues, or between 1 and about 50 contiguous amino acid residues, including deletions of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50 contiguous amino acid residues. In some embodiments, deletions of between 1-30, 31-40, 41-50, 51-60, 61-70, 71-80, 81-90, 91-100, 101-150, 151-200, 201-250 or 251-300 contiguous amino acid residues are contemplated.

[0089] Certain embodiments of the invention also contemplate that the methods may be applied to variants of a target protein, for example, naturally-occurring variants, or variants that have been generated by conventional mutagenesis methods in order to modulate a property of the protein. Variants generated by the present methods are also suitable candidates for additional rounds of sequence diversity generation in order to further modulate one or more property of the protein.

25 *Polynucleotides*

[0090] Polynucleotides employed in the methods of the present invention comprise a nucleic acid sequence encoding a target protein (“coding sequence”). The polynucleotides may be in the form of RNA or in the form of DNA, which DNA includes cDNA, genomic DNA, and synthetic DNA. The DNA may be double-stranded or single-stranded, and if single stranded may be the coding strand or non-coding (anti-sense) strand. A nucleic acid sequence which encodes a target protein for use in the

methods of the present invention may be identical to the coding sequence known in the art for the target protein or may be a different coding sequence, which, as a result of the redundancy or degeneracy of the genetic code, encodes the same protein.

[0091] The polynucleotides may include only the coding sequence for the target protein; the coding sequence for the target protein and additional coding sequence; the coding sequence for the target protein (and optionally additional coding sequence) and non-coding sequence, such as introns or non-coding sequences 5' and/or 3' of the coding sequence. The coding sequence may be in the form of one or more exons, which may be contiguous or may be interspersed with one or more introns. The non-coding sequences may include, for example, one or more regulatory nucleic acid sequences that may be a regulated or regulatable promoter, enhancer, other transcription regulatory sequence, repressor binding sequence, translation regulatory sequence or other regulatory nucleic acid sequence. Thus, the term "polynucleotide encoding" a target protein encompasses a polynucleotide which includes only coding sequence for the target protein as well as polynucleotides that include additional coding and/or non-coding sequence(s).

[0092] The coding sequence for various proteins suitable for use as target proteins are known in the art and can be obtained from public databases such as GenBank. Many proteins have been cloned and polynucleotides comprising the coding sequences for these proteins may be obtained from commercial sources. Alternatively, coding sequences can be obtained from an appropriate source using standard molecular biology techniques, such as those described in *Molecular Cloning: A Laboratory Manual* (Third Edition) (Sambrook, *et al.*, 2001, Cold Spring Harbour Laboratory Press, NY) and *Current Protocols in Molecular Biology* (Ausubel *et al.* (Ed.), 1987 & Updates, J. Wiley & Sons, Inc., Hoboken, NJ). In addition, many companies offer custom gene synthesis and may be used as a source of coding sequences for a target protein.

[0093] As noted above, the term "target protein" as used herein includes both the full-length protein and fragments of the protein. Accordingly, the polynucleotides for use in the methods of the invention that encode a target protein may encode the full length protein or a fragment thereof. In some embodiments, the polynucleotide may be a

truncated nucleic acid molecule which has less than the full length nucleotide sequence of a known or described nucleic acid molecule, where such a known or described nucleic acid molecule may be a naturally occurring, a synthetic or a recombinant nucleic acid molecule, so long as one skilled in the art would regard it as a full length molecule. Thus, for example, truncated nucleic acid molecules that correspond to a gene sequence contain less than the full length gene where the gene may comprise coding sequences and optionally non-coding sequences, promoters, enhancers and other regulatory sequences, flanking sequences and the like, and other functional and non-functional sequences that are recognized as part of the gene. In certain embodiments, truncated nucleic acid molecules correspond to a mRNA sequence and contain less than the full length mRNA transcript, which may include various translated and non-translated regions as well as other functional and non-functional sequences. In certain embodiments, truncated nucleic acid molecules may include one or more deletions from either terminus of the polynucleotide or a deletion from a non-terminal region of the polynucleotide, for example, in some embodiments, deletions of between about 1 and about 1500 contiguous nucleotides. In some embodiments, truncated nucleic acid molecules may include deletions of between 1 and about 1200 contiguous nucleotides, for example, between 1 and about 1000 contiguous nucleotides, between 1 and about 750 contiguous nucleotides, between 1 and about 500 contiguous nucleotides, between 1 and about 300 contiguous nucleotides, including deletions of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31-40, 41-50, 51-74, 75-100, 101-150, 151-200, 201-250 or 251-299 contiguous nucleotides.

[0094] In certain embodiments, the polynucleotide may be codon-optimized according to standard codon usage preference tables, such that its expression in the chosen host cell is optimized.

[0095] Certain embodiments of the invention encompass the use of variant polynucleotides in the present methods, for example, polynucleotides that encode analogs and/or derivatives of a target protein. The polynucleotide variants may be, for example, naturally-occurring allelic variants of the polynucleotide or non-naturally occurring variants. As is known in the art, an allelic variant is an alternate form of a nucleic acid sequence which may have at least one of a substitution, a deletion or an addition of one or more nucleotides, any of which does not substantially alter the

function of the encoded protein. Non-naturally occurring polynucleotide variants may be accomplished by a number of conventional methods. For example, mutations can be introduced at particular loci by synthesizing oligonucleotides containing a mutant sequence, flanked by restriction sites enabling ligation to fragments of the native sequence. Following ligation, the resulting reconstructed sequence encodes an analog having the desired amino acid insertion(s), substitution(s), or deletion(s). Alternatively, oligonucleotide-directed site-specific mutagenesis procedures can be employed to provide an altered gene wherein predetermined codons can be altered by substitution, deletion or insertion. Exemplary methods of making such alterations are described, for example, in *Molecular Cloning: A Laboratory Manual* (Third Edition) (Sambrook, *et al.*, 2001, Cold Spring Harbour Laboratory Press, NY) and *Current Protocols in Molecular Biology* (Ausubel *et al.* (Ed.), 1987 & Updates, J. Wiley & Sons, Inc., Hoboken, NJ).

[0096] In certain embodiments, as noted above, the target protein may be a variant with modulated properties. Polynucleotides that encode such variants are thus also contemplated and may be generated by conventional mutagenesis methods as described above or by the present methods as candidates for additional rounds of sequence diversity generation.

Recombination Signal Sequences (RSSs)

[0097] The recombination signal sequence (RSS) in accordance with the present invention preferably consists of two conserved sequences (for example, heptamer, 5'-CACAGTG-3', and nonamer, 5'-ACAAAAACC-3'), separated by a spacer of either 12 +/- 1 bp (a "12-signal" RSS) or 23 +/- 1 bp (a "23-signal" RSS). Within the host cell, two RSSs (one 12-signal RSS and one 23-signal RSS) are selected and rearranged under the "12/23 rule." Recombination does not occur between two RSS signals with the same size spacer. As would be appreciated by one of skill in the art, the orientation of the RSS determines if recombination results in a deletion or inversion of the intervening sequence.

[0098] As a result of extensive investigations of RSS processes, it is known in the art which nucleotide positions within RSSs cannot be varied without compromising RSS functional activity in genetic recombination mechanisms, which nucleotide positions

within RSSs can be varied to alter (for example, increase or decrease in a statistically significant manner) the efficiency of RSS functional activity in genetic recombination mechanisms, and which positions within RSSs can be varied without having any significant effect on RSS functional activity in genetic recombination mechanisms (see, for example, Ramsden *et al.*, 1994, *Proc Natl Acad Sci USA* 88(23): 10721-10725; Akamatsu *et al.*, 1994, *J Immunol* 153:4520; Hesse *et al.*, 1989, *Genes Dev* 3:1053; Fanning *et al.*, 1996, *Immunogenetics* 44(2):146-150; Larijani *et al.*, 1999, *Nucleic Acids Res* 27(11):2304-2309; Nadel *et al.*, 1998, *J Exp Med* 187:1495; Lee *et al.*, 2003, *PLoS Biol* 1:E1; and Cowell *et al.*, 2004, *Immunol. Rev.* 200:57).

10 [0099] In certain embodiments, the RSS selected for inclusion in the target protein coding sequence is a RSS that is known to the art. Also contemplated in some embodiments are sequence variants of known RSSs that comprise one or more nucleotide substitutions (for example, about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 or more substitutions) relative to the known RSS sequence and which, by virtue of such substitutions, predictably have low efficiency (for example, about 1% or less, relative to a high efficiency RSS), medium efficiency (for example, about 10% to about 20%, relative to a high efficiency RSS) or high efficiency. Also contemplated in some embodiments are those RSS variants for which one or more nucleotide substitutions relative to a known RSS sequence will have no significant effect on the recombination efficiency of the RSS (for example, the success rate of the RSS in promoting formation of a recombination product, as known in the art).

[00100] In accordance with certain embodiments of the invention, RSSs selected for inclusion in the target protein coding sequence are pairs of RSSs in which the first RSS of the pair is capable of functional recombination with the second RSS of the pair (*i.e.* “complementary pairs”). It is to be understood that when a first RSS (for example present in a first polynucleotide or nucleic acid sequence) is described as being capable of functional recombination with a second RSS (for example present in a second polynucleotide or nucleic acid sequence), such capability includes compliance with the above-noted 12/23 rule for RSS spacers, such that if the first RSS comprises a 12-nucleotide spacer then the second RSS will comprise a 23-nucleotide spacer, and similarly if the first RSS comprises a 23-nucleotide spacer then the second RSS will comprise a 12-nucleotide spacer.

[00101] Examples of RSS sequences known in the art, including their characterization as high, medium or low efficiency RSSs, are presented in Table 1A and 1B.

Table 1A. EXEMPLARY RECOMBINATION SIGNAL SEQUENCES (12 NUCLEOTIDE SPACER)

5

	Heptamer	Spacer	Nonamer
	H12	S12	N12
Part I. Efficiency: HIGH			
1	CACAGTG	ATACAGACCTTA [SEQ ID NO:2]	ACAAAAACC
2	CACAGTG	CTACAGACTGGA [SEQ ID NO:9]	ACAAAAACC
3	CACAGTG	CTCCAGGGCTGA [SEQ ID NO:10]	ACAAAAACC
4	CACAGTG	CTACAGACTGGA [SEQ ID NO:9]	ACAAAAACC
5	CACAGTG	CTACAGACTGGA [SEQ ID NO:9]	ACAAAAACC
6	CACAGTG	CTACAGACTGGA [SEQ ID NO:9]	ACAAAAACC
7	CACAGTG	GTACAGACCAAT [SEQ ID NO:11]	ACAGAAACC
Part II. Efficiency: MEDIUM (~10-20% of High)			
8	CACGGTG	CTACAGACTGGA [SEQ ID NO:9]	ACAAAAACC
9	CACAATG	CTACAGACTGGA [SEQ ID NO:9]	ACAAAAACC
10	CACAGCG	CTACAGACTGGA [SEQ ID NO:9]	ACAAAAACC
11	CACAGTG	CTACAGACTGGA [SEQ ID NO:9]	ACAAAAACC
12	CACAGTG	CTACAGACTGGA [SEQ ID NO:9]	ACAAAAACC
13	CACAGTG	CTACAGACTGGA [SEQ ID NO:9]	ACAAAAACC
14	CACAGTG	CTACAGACTGGA [SEQ ID NO:9]	ACAAAAACC
15	CACAGTG	CTACAGACTGGA [SEQ ID NO:9]	ACAAAAACC
16	CACAGTG	CTACAGACTGGA [SEQ ID NO:9]	ACAAAAACC
17	CACAGTG	CTACAGACTGGA [SEQ ID NO:9]	CAAAAACCC
18	CACAGTG	CTACAGACTGGA [SEQ ID NO:9]	ACAAAAACC
19	CACAATG	CTACAGACTGGA [SEQ ID NO:9]	ACAAAAACC
20	CACAGCG	CTACAGACTGGA [SEQ ID NO:9]	ACAAAAACC

	Heptamer	Spacer	Nonamer
	H12	S12	N12
Part III. Efficiency: LOW (~1% or less of High)			
21	TACAGTG	CTACAGACTGGA [SEQ ID NO:9]	ACAAAAACC
22	GACAGTG	CTACAGACTGGA [SEQ ID NO:9]	ACAAAAACC
23	CATAGTG	CTACAGACTGGA [SEQ ID NO:9]	ACAAAAACC
24	CACAATG	CTACAGACTGGA [SEQ ID NO:9]	ACAAAAACC
25	CACAGTG	CTACAGACTGGA [SEQ ID NO:9]	ACAAAAACC
26	CAGAGTG	CTCCAGGGCTGA [SEQ ID NO:10]	ACAAAAACC
27	CACAGTG	CTCCAGGGCTGA [SEQ ID NO:10]	AAAAAAACC
28	CTCAGTG	CTCCAGGGCTGA [SEQ ID NO:10]	ACAAAAACC

Table 1B. EXEMPLARY RECOMBINATION SIGNAL SEQUENCES (23 NUCLEOTIDE SPACER)

	Heptamer	Spacer	Nonamer	Ref.*
	H23	S23	N23	
Part I. Efficiency: HIGH				
1	CACAGTG	GTAGTACTCCACTGTCTGGCTGT [SEQ ID NO:12]	ACAAAAACC	4
2	CACAGTG	GTAGTACTCCACTGTCTGGCTGT [SEQ ID NO:12]	ACAAAAACC	3
3	CACAGTG	GTAGTACTCCACTGTCTGGGTGT [SEQ ID NO:12]	ACAAAAACC	1
4	CACAGTG	TTGCAACCACATCCTGAGTGTGT [SEQ ID NO:14]	ACAAAAACC	2
5	CACAGTG	GTAGTACTCCACTGTCTGGCTGT [SEQ ID NO:12]	ACAAAAACC	2
6	CACAGTG	ACGGAGATAAAGGAGGAAGCAGG [SEQ ID NO:15]	ACAAAAACC	2

	Heptamer	Spacer	Nonamer	Ref.*
	H23	S23	N23	
7	CACAGTG	GCCGGGCCCCGCGGCCCGGCGGC [SEQ ID NO:13]	ACAAAAACC	5
Part II. Efficiency: MEDIUM (~10-20% of High)				
8	CACAGTG	GTAGTACTCCACTGTCTGGCTGT [SEQ ID NO:12]	ACAAAAACC	3
9	CACAGTG	GTAGTACTCCACTGTCTGGCTGT [SEQ ID NO:12]	ACAAAAACC	3
10	CACAGTG	GTAGTACTCCACTGTCTGGCTGT [SEQ ID NO:12]	ACAAAAACC	3
11	CACAATG	GTAGTACTCCACTGTCTGGCTGT [SEQ ID NO:12]	ACAAAAACC	3
12	CACAGCG	GTAGTACTCCACTGTCTGGCTGT [SEQ ID NO:12]	ACAAAAACC	3
13	CACAGTA	GTAGTACTCCACTGTCTGGCTGT [SEQ ID NO:12]	ACAAAAACC	3
14	CACAGTG	GTAGTACTCCACTGTCTGGCTGT [SEQ ID NO:12]	ACAATAACC	3
15	CACAGTG	GTAGTACTCCACTGTCTGGCTGT [SEQ ID NO:12]	ACAAGAACC	3
16	CACAGTG	GTAGTACTCCACTGTCTGGCTGT [SEQ ID NO:12]	ACACGAACC	3
17	CACAGTG	GTAGTACTCCACTGTCTGGCTGT [SEQ ID NO:12]	ACAAAAACC	3
18	CACAGTG	GTAGTACTCCACTGTCTGGCTGT [SEQ ID NO:12]	ACACGAACC	3
19	CACAATG	GTAGTACTCCACTGTCTGGCTGT [SEQ ID NO:12]	ACAAAAACC	3
20	CACAGCG	GTAGTACTCCACTGTCTGGCTGT [SEQ ID NO:12]	ACAAAAACC	3
Part III. Efficiency: LOW (~1% or less of High)				

	Heptamer	Spacer	Nonamer	Ref.*
	H23	S23	N23	
21	CACAGTA	GTAGTACTCCACTGTCTGGCTGT [SEQ ID NO:12]	ACAAAAACC	3
22	CACAGTG	GTAGTACTCCACTGTCTGGCTGT [SEQ ID NO:12]	ACAAAAACC	3
23	CACAATG	GTAGTACTCCACTGTCTGGCTGT [SEQ ID NO:12]	ACAAAAACC	3
24	CATAGTG	GTAGTACTCCACTGTCTGGCTGT [SEQ ID NO:12]	ACAAAAACC	3
25	CACAGTG	GTAGTACTCCACTGTCTGGCTGT [SEQ ID NO:12]	TGTCTCTGA	3
26	CACAGTG	GTAGTACTCCACTGTCTGGGTGT [SEQ ID NO:12]	ACAAAAACC	1
27	CACAGTG	GTAGTACTCCACTGTCTGGGTGT [SEQ ID NO:12]	ACAAAAACC	1
28	CACAGTG	GTAGTACTCCACTGTCTGGGTGT [SEQ ID NO:12]	ACAAAAACC	1

*(1) Akamatsu, 1994, *ibid*; (2) Cowell, 2004, *ibid*; (3) Hesse, 1989 *ibid*; (4) Lee, 2003, *ibid*; (5) Nadel, 1998, *ibid*.

Positioning the RSSs within the Protein Coding Sequence

[00102] In accordance with the present invention, the RSSs are positioned at a pre-determined (targeted) location or locations within the target protein coding sequence.

[00103] While various regions of the target protein may be selected as a targeted location for the introduction of sequence diversity, it is generally preferred that the targeted location is not within a region of the protein important for protein folding and/or adoption of the protein's functional conformation, although this does remain an alternative option. Selection of appropriate target locations can be readily made by one skilled in the art with reference to the target protein's known or predicted structure. For example, the crystal structures of many proteins are known and available, for example, from sources such as the Protein Data Bank (PDB) maintained by the Research

Collaboratory for Structural Bioinformatics (RCSB). In addition, many resources are publicly available allowing for the prediction of secondary and/or tertiary structure of a known protein, for example, RaptorX (University of Chicago), ESyPred3D (University of Namur), HHPred (Max Planck Institute for Experimental Biology), Phyre2 (Imperial College, London), ProtInfo ABCM, BAGHEERATH-H (Ministry of Science and Technology, India, Department of Biotechnology), Rosetta@home (University of Washington), JPred (University of Dundee), NetSurfP (Centre for Biological Sequence Analysis), PredictProtein, PSIPred (UCL), SymPred, and the like.

[00104] In accordance with certain embodiments, the targeted location is selected such that it is within a non-conformational region of the protein (*i.e.* a region of the protein that is not important for folding and/or adoption of the protein's functional conformation). In some embodiments, the targeted location is selected such that it is within an externally exposed region of the protein.

[00105] In certain embodiments in which the target protein is a ligand-binding protein, the targeted location is selected such that it is within or proximal to a ligand-binding domain, or in a region that otherwise impacts on ligand binding by the protein.

[00106] In some embodiments, the targeted location is selected such that it is within or proximal to a loop region, for example, an externally exposed loop. Loop regions of a target protein can be readily identified with reference to the target protein's known or predicted structure. In addition, various programs are available that allow loop regions of a protein to be identified based on the known amino acid sequence (for example, ArchPred, FREAD, ModLoop, RAPPER and SuperLooper). In some embodiments, the targeted location is selected such that it is within an externally exposed loop. In certain embodiments in which the targeted location is within a loop region of the protein, the location is selected such that it is substantially within the central portion of the loop.

[00107] For example, the 10th domain of FN3 (¹⁰FN3) is known to comprise three loops at one end of a β -sandwich, which can be sequence diversified to provide binding to a target ligand. Specifically, the three loops of ¹⁰FN3 that are analogous to the antigen binding loops of the IgG heavy chain are defined by amino acid residues 21-31, 51-56 and 76-88. Each of these loops is thus a suitable target for introduction of

sequence diversity according to the present methods. The second loop is only 6 amino acids in length and as such may be extended, for example by about 10 to 16 amino acids, at the same time as, or in addition to, introducing sequence diversity.

5 [00108] In immunoglobulins, both CDR1 and CDR2 are loop regions which can be readily identified and are suitable target locations for introduction of sequence diversity using the methods described herein. CDR3 is also a loop region of immunoglobulins that is a suitable target locations for introduction of sequence diversity in certain embodiments.

10 [00109] In those embodiments in which the targeted location is within a loop region of the protein, it is contemplated that the RSSs can be located at various positions within the loop. For example, one or both of the RSSs in a complementary pair may be positioned at or proximal to the start (*i.e.* the N-terminal, or 5', side) of the loop, or one or both of the RSSs may be positioned at or proximal to the end (*i.e.* the C-terminal, or 3', side) of the loop, or one or both of the RSSs may be positioned proximal to the
15 centre point of the loop. Other combinations for the positioning of the RSSs will be readily apparent to the skilled worker and are included. In one embodiment, the RSSs are located within a loop region of the target protein and at least one of the RSSs is positioned proximal to the centre point of the loop.

[00110] As noted above, in certain embodiments, complementary pairs of RSSs are
20 introduced into the coding sequence for the target protein, in which the first RSS of the pair is capable of functional recombination with the second RSS of the pair. In accordance with these embodiments, the two RSSs of the complementary pair are separated by an intervening sequence of about 100bp or more in length. The nucleotide sequence of the intervening sequence is not critical to the invention and may be
25 comprised of a sequence heterologous to the coding sequence or it may be comprised of part of the coding sequence. For example, in certain embodiments, the complementary pair of RSSs are introduced individually into the coding sequence such that part of the coding sequence forms the intervening sequence. In other embodiments, the complementary pair of RSSs is introduced together with a heterologous intervening
30 sequence into the coding sequence as a "cassette." The nucleotide sequence of the intervening sequence can accommodate a wide variety of sequences, including for

example some selectable markers, some promoters and other regulatory elements such as polyadenylation signals, but preferably does not include insulator-like elements as exemplified by cHS4 and AAV1.

5 [00111] In certain embodiments, the intervening sequence comprises an expression cassette, for example containing a promoter and optionally poly(A) sequences that drive expression of a marker such as GFP or a cell surface marker such that recombination can be monitored, or a selectable marker such as a drug resistance gene such that the cell can be maintained in the un-recombined state via drug selection.

10 [00112] Regardless of the composition of the intervening sequence, it is preferably selected to be at least 100bp in length, for example, at least 110bp, at least 120bp, at least 130bp, at least 140bp, at least 150bp, but may range up to several kilobases in size, for example up to about 5kb. One skilled in the art will understand that the exact upper limit for the intervening sequence will be dictated by the limitation of the vector system used. In certain embodiments, the intervening sequence is selected to be
15 between about 100bp and 5kb, for example, between about 150bp and 5kb, between about 180bp and 5kb, between about 180bp and 4kb, between about 180bp and 3kb or between about 180bp and 2kb. In some embodiments, the intervening sequence is selected to be between about 100bp and 1.5kb, for example, between about 110bp and 1.5kb, between about 120bp and 1.5kb, between about 130bp and 1.5kb, between
20 140bp and 1.5kb, or between 150bp and 1.5kb. In some embodiments, the intervening sequence is selected to be between about 180bp and 1.9kb, for example, between about 180bp and 1.8kb, between about 180bp and 1.7kb, between about 180bp and 1.6kb, or between 180bp and 1.5kb. Other exemplary embodiments include intervening sequences of between about 190bp and 1.5kb, between about 200bp and 1.5kb, between
25 about 210bp and 1.5kb, between about 220bp and 1.5kb, between about 230bp and 1.5kb, between about 240bp and 1.5kb, and between about 250bp and 1.5kb.

[00113] In certain embodiments, flanking sequences are included adjacent to the heptamer of the RSS. In accordance with this embodiment, the flanking sequences may be chosen to have a defined sequence (for example, to specifically encode one or more
30 amino acids) or they may have a random sequence. In some embodiments, the flanking sequences may be selected to introduce certain characteristics at the site of insertion,

for example, through the addition of one or more charged amino acids, histidine residues or cysteine residues. In certain embodiments, the flanking sequence may comprise a duplication of a part of the sequence into which the RSSs are to be introduced. In some embodiments, the position and length of the flanking sequences are
5 selected to bias diversification towards one side of the insertion point, or to provide a larger loop size prior to diversification.

[00114] When used, the length of the flanking sequence is selected such that it does not interfere with the structural integrity of the target protein. In certain embodiments in which flanking sequences accompany the RSSs and are introduced into a loop region of
10 the target protein, the flanking sequence(s) are selected such that their introduction into the loop results in an increase in loop length of about 150% or less, for example, 100% or less, or 50% or less. In some embodiments, the flanking sequence(s) are selected such that their introduction into the loop results in an increase in loop length of 0% to about 50%, for example, between about 1% and about 50%.

15 [00115] In some embodiments in which the targeted location is within a small loop of a protein, flanking sequences may be used in conjunction with the RSSs to ensure that the majority of the products include a loop that retains a minimal length once sequence diversification has taken place. In certain embodiments the RSSs are inserted into a region of the protein such that sequences are deleted resulting in a net smaller loop size
20 than the parent molecule, for example, a decrease in loop length of between about 1% and about 50%.

[00116] The methods of the present invention contemplate the introduction of one, or more than one, complementary pairs of RSSs into the coding sequence for a target protein. In certain embodiments, one complementary pair of RSSs is introduced into
25 the coding sequence in order to generate sequence diversity at a targeted location in the protein. In some embodiments, two complementary pairs of RSSs are introduced into the coding sequence in order to generate sequence diversity at more than one targeted location in the protein. In certain embodiments in which two pairs of RSSs are used, they may be oriented such that recombination between the distal members of the two
30 pairs cannot occur (*i.e.* the 5' RSS of the first pair and the 3' RSS of the second pair are both 23bp RSSs or are both 12bp RSSs).

[00117] In certain embodiments, two or more complementary pairs of RSSs are introduced into the coding sequence in order to generate sequence diversity at more than one targeted location in the protein.

5 [00118] In certain embodiments, a complementary pair of RSSs may be introduced into the coding sequence as a “cassette” that includes a heterologous intervening sequence spacing the two RSSs apart, such that upon recombination at the RSSs, the RSSs and the heterologous sequence are deleted and sequence diversity is introduced into the coding sequence at the site of recombination.

10 [00119] In certain embodiments, a complementary pair of RSSs may be introduced into the coding sequence as a “cassette” that includes a heterologous intervening sequence spacing the two RSSs apart as well as flanking sequences on the other side of one or both of the RSSs, such that upon recombination at the RSSs, the RSSs and the heterologous sequence are deleted and new sequences (from the flanking sequences) are added, in addition to sequence diversity being introduced into the coding sequence
15 at the site of recombination.

[00120] The RSSs can be introduced into the polynucleotide by standard genetic engineering techniques such as those described in *Molecular Cloning: A Laboratory Manual* (Third Edition) (Sambrook, *et al.*, 2001, Cold Spring Harbour Laboratory Press, NY) and *Current Protocols in Molecular Biology* (Ausubel *et al.* (Ed.), 1987 &
20 Updates, J. Wiley & Sons, Inc., Hoboken, NJ).

Additional Coding Sequences

[00121] In accordance with certain embodiments of the invention, the polynucleotide may comprise additional coding sequences and thus may encode a fusion protein that comprises the target protein fused to another peptide or polypeptide that provides
25 additional functionality to the protein. Examples of peptides and polypeptides that provide additional functionality include, but are not limited to, secretory signal sequences, leader sequences, plasma membrane anchor domain polypeptides such as hydrophobic transmembrane domains (see, for example, Heuck *et al.*, 2002, *Cell Biochem. Biophys.* 36:89; Sadlish *et al.*, 2002, *Biochem J.* 364:777; Phoenix *et al.*,
30 2002, *Mol. Membr. Biol.* 19:1; Minke *et al.*, 2002, *Physiol. Rev.* 82:429) or

glycosylphosphatidylinositol attachment sites (“glypiation” sites) (see, for example, Chatterjee *et al.*, 2001, *Cell Mol. Life Sci.* 58:1969; Hooper, 2001, *Proteomics* 1:748, and Spiro, 2002, *Glycobiol.* 12:43R), and other structural features that assist in localizing the fusion protein to the cell surface such as protein-protein association domains, lipid association domains, glycolipid association domains and proteoglycan association domains, for example, cell surface receptor binding domains, extracellular matrix binding domains, and lipid raft-associating domains (see, for example, Browman *et al.*, 2007, *Trends Cell Biol* 17:394-402; Harder, T., 2004, *Curr Opin Immunol* 16:353-9; Hayashi, T. and Su, T. P., 2005, *Life Sci* 77:1612-24; Holowka, D. and Baird, B., 2001, *Semin Immunol* 13:99-105, and Wollscheid *et al.*, 2004, *Subcell Biochem* 37:121- 52).

[00122] In some embodiments, the additional coding sequences may encode a “tag” to facilitate downstream screening and/or purification of the target protein. Examples of such heterologous nucleic acid sequences include, but are not limited to, affinity tags such as metal-affinity tags, histidine tags, protein A, glutathione S transferase, Glu-Glu affinity tag, substance P, FLAG peptide (Hopp *et al.*, 1988, *Biotechnology* 6:1204), streptavidin binding peptide, or other antigenic epitope or binding domain (see, in general, Ford *et al.*, 1991, *Protein Expression and Purification* 2:95). Nucleic acid sequences encoding affinity tags are available from commercial suppliers (for example, Pharmacia Biotech, Piscataway, N.J.).

[00123] In some embodiments, the polynucleotide comprises additional coding sequences that encode a plasma membrane anchor domain. For example, a transmembrane polypeptide domain typically comprising a membrane spanning domain (such as an $[\alpha]$ -helical domain) which includes a hydrophobic region capable of energetically favorable interaction with the phospholipid fatty acyl tails that form the interior of the plasma membrane bilayer, or a membrane-inserting domain polypeptide typically comprising a membrane-inserting domain which includes a hydrophobic region capable of energetically favorable interaction with the phospholipid fatty acyl tails that form the interior of the plasma membrane bilayer but that may not span the entire membrane. Well known examples of transmembrane proteins having one or more transmembrane polypeptide domains include members of the integrin family, CD44, glycophorin, MHC Class I and II glycoproteins, EGF receptor, G protein coupled

receptor (GPCR) family, receptor tyrosine kinases (such as insulin-like growth factor 1 receptor (IGFR) and platelet-derived growth factor receptor (PDGFR)), porin family and other transmembrane proteins. Certain embodiments of the invention contemplate using a portion of a transmembrane polypeptide domain such as a truncated polypeptide
5 having membrane-inserting characteristics as may be determined according to standard and well known methodologies.

[00124] In some embodiments of the invention, the polynucleotide comprises additional coding sequences that encode a specific protein-protein association domain, for example a protein-protein association domain that is capable of specifically
10 associating with an extracellularly disposed region of a cell surface protein or glycoprotein. In certain embodiments, the protein-protein association domain may result in an association that is initiated intracellularly, for instance, concomitant with the synthesis, processing, folding, assembly, transport and/or export to the cell surface of a cell surface protein. In some embodiments, the protein-protein association domain
15 is known to associate with another cell surface protein that is membrane anchored and exteriorly disposed on a cell surface. Non-limiting examples of such domains include, RGD-containing polypeptides including those that are capable of integrin binding (see, for example, Heckmann, D. and Kessler, H., 2007, *Methods Enzymol* 426:463-503 and Takada *et al.*, 2007, *Genome Biol* 8:215).

[00125] In some embodiments, the polynucleotide comprises a secretory signal sequence that encodes a secretory peptide. A secretory peptide is an amino acid sequence that acts to direct the secretion of a mature polypeptide or protein from a cell and is generally characterized by a core of hydrophobic amino acids. Secretory peptides are typically, but not exclusively, positioned at the amino termini of newly synthesized
25 proteins. The secretory peptide may be cleaved from the mature protein during secretion and may, therefore, contain processing sites that allow cleavage of the signal peptide from the mature protein as it passes through the secretory pathway. Examples of secretory peptides are known in the art and include, but are not limited to, alpha mating factor leader sequence, the secretory pre-peptide of IL-15, the tissue Plasminogen Activator (tPA) secretory leader peptide, transferrin (Tf) signal sequence,
30 IgE secretory peptides, IgHV and IgKV signal peptides and GM-CSF secretory peptides.

[00126] In certain embodiments, sequences encoding a transmembrane domain are included in the polynucleotide to provide surface expression of the variant protein. In some embodiments, the variant protein is cloned in-frame with a selectable marker to allow for the selection of productive in-frame products. In some embodiments, the polynucleotide comprises sequences encoding a transmembrane domain, a selectable marker and an enzyme cleavage site prior to the selectable marker to allow for cleavage of the variant protein from the transmembrane domain.

[00127] Additional sequences, when used, can be included in the polynucleotide by standard genetic engineering techniques such as those described in *Molecular Cloning: A Laboratory Manual* (Third Edition) (Sambrook, *et al.*, 2001, Cold Spring Harbour Laboratory Press, NY) and *Current Protocols in Molecular Biology* (Ausubel *et al.* (Ed.), 1987 & Updates, J. Wiley & Sons, Inc., Hoboken, NJ).

Vectors

[00128] Certain embodiments of the invention relate to vectors comprising the polynucleotide encoding the target protein, and also vectors comprising nucleic acid sequences encoding RAG-1, RAG-2 or TdT (or functional fragments or variants thereof), and vectors comprising regulatory constructs such as siRNA regulators of RAG-1, RAG-2 and/or TdT expression. A wide variety of suitable vectors are known in the art and may be employed as described or according to conventional procedures, including modifications, as described for example in Sambrook *et al.*, *ibid.*; Ausubel *et al.*, *ibid.*, and elsewhere.

[00129] One skilled in the art will appreciate that the precise vector used is not critical to the instant invention and suitable vectors can be readily selected by the skilled person. Examples of expression vectors and cloning vehicles include, but are not limited to, viral particles, baculovirus, phage, plasmids, phagemids, cosmids, fosmids, bacterial artificial chromosomes, retrovirus vectors, viral DNA (for example, vaccinia, adenovirus, fowl pox virus, pseudorabies and derivatives of SV40), P1-based artificial chromosomes, yeast plasmids, yeast artificial chromosomes, and other known vectors specific for specific host cells of interest.

[00130] Large numbers of suitable vectors are known to those of skill in the art, and many are commercially available. Exemplary commercially available vectors include the bacterial vectors: pcDNA (Invitrogen), pQE vectors (Qiagen), pBLUESCRIPT™ plasmids, pNH vectors, lambda-ZAP vectors (Stratagene); ptrc99a, pKK223-3, 5 pDR540, and pRIT2T (Pharmacia); and the eukaryotic vectors: pXT1, pSGS (Stratagene), pSVK3, pBPV, pMSG and pSVLSV40 (Pharmacia). Other vectors include, for example, adenovirus (Ad) vectors (such as, vectors based on non-replicating Ad5, or replication-competent Ad4 and Ad7 vectors), adeno-associated virus (AAV) vectors (such as, AAV type 5), alphavirus vectors (such as, Venezuelan equine encephalitis virus (VEE), sindbis virus (SIN), semliki forest virus (SFV), and 10 VEE-SIN chimeras), herpes virus vectors, measles virus vectors, pox virus vectors (such as, vaccinia virus, modified vaccinia virus Ankara (MVA), NYVAC (derived from the Copenhagen strain of vaccinia), and avipox vectors: canarypox (ALVAC) and fowlpox (FPV) vectors), and vesicular stomatitis virus vectors. Other suitable plasmids and vectors are known in the art and can readily be selected by the skilled worker. In 15 accordance with various embodiments of the invention, either low copy number or high copy number vectors may be employed.

[00131] One skilled in the art will understand that the vector may further include regulatory elements, such as transcriptional elements, required for efficient 20 transcription of the DNA sequence encoding the target protein. Examples of regulatory elements that can be incorporated into the vector include, but are not limited to, promoters, enhancers, terminators, alpha-factors, ribosome binding sites and polyadenylation signals. In various embodiments, the present invention, therefore, contemplates vectors comprising one or more regulatory elements operatively linked to 25 a polynucleotide encoding the target protein.

[00132] One skilled in the art will appreciate that selection of suitable regulatory elements is dependent on the host cell chosen for expression of the encoded protein and that such regulatory elements may be derived from a variety of sources, including bacterial, fungal, viral, mammalian or insect genes.

30 [00133] Mammalian expression vectors, for example, may comprise one or more of an origin of replication, any necessary ribosome binding sites, a polyadenylation site,

splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking non-transcribed sequences. DNA sequences derived from the SV40 splice and polyadenylation sites, for example, may be used to provide the required non-transcribed genetic elements. Eukaryotic expression vectors may also contain one or more
5 enhancers to increase expression levels of the protein. Enhancers are cis-acting elements of DNA, usually from about 10 to about 300 bp in length that act on a promoter to increase its transcription. Examples include, but are not limited to, the SV40 enhancer on the late side of the replication origin by 100 to 270, the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the
10 replication origin and the adenovirus enhancers.

[00134] Examples of typical promoters include, but are not limited to, the bacterial promoters: *lacZ*, *T3*, *T7*, *gpt*, λ P_R , P_L and *trp*; and the eukaryotic promoters: CMV immediate early, HSV thymidine kinase, early SV40, late SV40, LTRs from retrovirus and mouse metallothionein-I. Promoter regions can also be selected from a
15 desired gene using chloramphenicol transferase (CAT) vectors or other vectors with selectable markers.

[00135] In certain embodiments the vector comprises an expression control sequence which is a "regulated promoter," which may be a promoter as provided herein or may be a repressor binding site, an activator binding site or other regulatory sequence that
20 controls expression of a nucleic acid sequence. In some embodiments, the vector comprises a tightly regulated promoter that is specifically inducible and that permits little or no transcription of nucleic acid sequences under its control in the absence of an induction signal. Examples of such tightly regulated promoters are known to those familiar with the art and described, for example, in Guzman *et al.* (1995, *J. Bacteriol.*
25 177:4121), Carra *et al.* (1993, *EMBO J.* 12:35), Mayer (1995, *Gene* 163:41), Haldimann *et al.* (1998, *J. Bacteriol.* 180:1277), Lutz *et al.* (1997, *Nuc. Ac. Res.* 25:1203), Allgood *et al.* (1997, *Curr. Opin. Biotechnol.* 8:474) and Makrides (1996, *Microbiol. Rev.* 60:512). In other embodiments of the invention, the vector comprises a regulated promoter that is inducible but that may not be tightly regulated. Inducible
30 systems that include regulated promoters include, for example, the Tet system or other similar expression-regulating components, such as the Tet/on and Tet/off system (Clontech Inc., Palo Alto, CA), the Regulated Mammalian Expression system

(Promega, Madison, WI), and the GeneSwitch System (Invitrogen Life Technologies, Carlsbad, CA).

[00136] In certain embodiments, the vector comprises a promoter that is not a regulated promoter; such a promoter may include, for example, a constitutive promoter
5 such as an insect polyhedrin promoter.

[00137] In addition, vectors may contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells. Such selectable markers include for example genes encoding dihydrofolate reductase or genes conferring neomycin resistance, puromycin resistance, or hygromycin resistance, or the
10 use of xanthine-guanine phosphoribosyltransferase in eukaryotic host cells, genes conferring ampicillin, chloramphenicol, erythromycin, kanamycin, neomycin or tetracycline resistance in bacterial host cells, and the *S. cerevisiae* TRP1 gene. Promoter regions can be selected from any desired gene using chloramphenicol transferase (CAT) vectors or other vectors with selectable markers. Selectable markers
15 can also include biosynthetic genes, such as those in the histidine, tryptophan and leucine biosynthetic pathways.

[00138] In certain embodiments, the vector can have two replication systems to allow it to be maintained in two organisms, for example in mammalian or insect cells for expression and in a prokaryotic host for cloning and amplification.

[00139] Also contemplated are replicating and non-replicating episomal vectors for transient expression. Replicating vectors contain origin sequences that promote plasmid replication in the presence of the appropriate trans factors. The SV40 and polyoma origins and respective T-antigens are non-limiting examples. Also contemplated are stably maintained episomal expression vectors. Episomal plasmids are usually based
20 on sequences from DNA viruses, such as BK virus, bovine papilloma virus 1 and Epstein-Barr virus (see, for example, Van Craenenbroeck, K., *et al.*, 2000, *Eur. J. Biochem.* 267:5665-5678). These vectors contain a viral origin of DNA replication and a viral early gene(s), the product of which activates the viral origin and thus allows the episome to reside in the transfected host cell line in a well-controlled manner. Episomal

vectors are plasmid constructions that replicate in both eukaryotic and prokaryotic cells and can therefore also be “shuttled” from one host cell system to another.

[00140] In some embodiments the plasmid selected is a plasmid that can be integrated into the host chromosome. Integration can occur by random methods or can be targeted. In some embodiments in which integrating expression vectors are used, the expression vector can contain at least one sequence homologous to the host cell genome, for example, two homologous sequences which flank the expression construct. The integrating vector can thus be directed to a specific locus in the host cell by selecting the appropriate homologous sequence for inclusion in the vector. Constructs for integrating vectors are well known in the art. Alternatively, the use of recombination systems like Cre/Lox and Flp/Frt can be used to target integration. Other methods utilizing zinc-finger proteins as developed by Sangamo Inc. (Richmond, CA) provide another approach to targeting vector integration.

[00141] In certain embodiments, the methods described herein employ a vector or recombination system that allows for stable integration of the polynucleotide into the host cell genome. In some embodiments, the methods described herein employ a vector or recombination system that allows for stable integration of the polynucleotide into the host cell genome as a single copy.

[00142] In certain embodiments of the invention, the vector employed is a viral vector such as a retroviral vector. For example, retroviruses from which the retroviral plasmid vectors may be derived include, but are not limited to, Moloney Murine Leukemia Virus, spleen necrosis virus, retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma virus, avian leukosis virus, gibbon ape leukemia virus, human immunodeficiency virus, adenovirus, Myeloproliferative Sarcoma Virus, and mammary tumour virus. Suitable promoters for inclusion in viral vectors include, but are not limited to, the retroviral LTR; the SV40 promoter; and the human cytomegalovirus (CMV) promoter described in Miller, *et al.* (1989, *Biotechniques* 7:980-990), or other suitable promoter (for example, cellular promoters such as eukaryotic cellular promoters including, but not limited to, the histone, pol III, and β -actin promoters). Other viral promoters which may be employed include, but are not limited to, adenovirus promoters, thymidine kinase (TK) promoters, and B19 parvovirus promoters. The selection of a suitable

promoter will be apparent to those skilled in the art, and may be from among either regulated promoters or promoters as described above.

[00143] In those embodiments that employ a retroviral plasmid vector, the vector is used to transduce packaging cell lines to form producer cell lines. Examples of
5 packaging cells which may be transfected include, but are not limited to, the PE501, PA317, [psi]-2, [psi]-AM, PA12, T19-14X, VT-19-17-H2, [psi]CRE, [psi]CRIP, GP+E-86, GP+envAm12, and DAN cell lines as described in Miller (1990, *Human Gene Therapy*, 7:5-14). The packaging cells may be transduced with the vector using
10 various means known in the art such as, for example, electroporation, the use of liposomes, and CaPO₄ precipitation. The producer cell line generates infectious retroviral vector particles which include the polynucleotide encoding the protein. Such retroviral vector particles then may be employed to transduce eukaryotic cells, either *in vitro* or *in vivo*, and the transduced eukaryotic cells will express the polynucleotide encoding the protein. Eukaryotic cells which may be transduced include, but are not
15 limited to, embryonic stem cells, embryonic carcinoma cells, hematopoietic stem cells, hepatocytes, fibroblasts, myoblasts, keratinocytes, endothelial cells, and bronchial epithelial cells.

[00144] The appropriate DNA or polynucleotide sequences can be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into an
20 appropriate restriction endonuclease site(s) by procedures known in the art. Standard techniques for cloning, DNA isolation, amplification and purification, for enzymatic reactions involving DNA ligase, DNA polymerase, restriction endonucleases and the like, and various separation techniques are those known and commonly employed by those skilled in the art. A number of standard techniques are described, for example, in
25 Sambrook et al., *ibid.*; Ausubel et al., *ibid.*, and elsewhere.

[00145] The vector can be introduced into a suitable host cell by one of a variety of methods known in the art. Such methods can be found generally described in Ausubel et al. (*ibid.*) and include, for example, stable or transient transfection, lipofection, electroporation, the use of polyethylenimine (PEI) and infection with recombinant viral
30 vectors. One skilled in the art will understand that selection of the appropriate host cell for expression of the target protein will be dependent upon the vector chosen. The

polynucleotide may stably integrate into the genome of the host cell (for example, with retroviral introduction) or may exist either transiently or stably in the cytoplasm (for example, through the use of traditional plasmids or vectors, utilizing standard regulatory sequences, selection markers, and the like, as described above).

5 *Host cells*

[00146] In accordance with the present invention, the host cell employed in the methods described herein is a host cell capable of utilizing recombination signals and undergoing RAG-1/RAG-2 mediated recombination. Accordingly, host cells suitable for use in the methods described herein express or can be engineered to express at least
10 RAG-1 and RAG-2 or functional fragments thereof that allow the host cell to utilize recombination signals and undergo RAG-1/RAG-2 mediated recombination.

[00147] In certain embodiments, cell lines to be used as host cells may additionally contain a functional TdT gene. TdT is encoded by a single gene and expresses a nuclear enzyme whose expression *in vivo* is restricted to lymphoid progenitor cells. TdT has,
15 however, been expressed in non-lymphoid cells and shown to participate in V(D)J recombination using retroviral and transient recombination substrates. TdT has been shown to be expressed as a number of different splice variants, including long form and short form. Certain embodiments of the invention contemplate the use of different isoforms of TdT.

20 [00148] TdT has also been shown to have a 3' to 5' exonuclease activity and the different isoforms of TdT have been shown to have different amounts of exonuclease activity. TdT exonuclease activity can be modulated by substitutions at the conserved aspartic acid residue in the exonuclease motif. In addition, expression of both isoforms was shown to modulate nuclease activity. TdT is highly conserved among species.
25 While mice have two isoforms both human and bovine have three isoforms. In certain embodiments, TdT activity in the host cell can be modulated by altering the levels of TdT in the cell. In some embodiments, mutant forms of TdT or different combinations of isoforms may be used in the host cell to generate coding joints with different extents of deletion and addition.

[00149] Cell lines may in certain embodiments be pre-B cells or pre-T cells that express RAG-1 and RAG-2, and optionally TdT, proteins. Such pre-B and pre-T cells may be capable of being induced to express RAG-1 and/or RAG-2, and optionally TdT, or alternatively, may constitutively express RAG-1 and/or RAG-2, and optionally TdT, but can be modified to substantially impair the expression of one, two or all three of these enzymes.

[00150] In some embodiments, the cell lines are non-immune cells that have been transformed with genes encoding each of RAG-1 and RAG-2, and optionally TdT (see for example, for RAG-1/2: Schatz, DG *et al.*, 1989, *Cell* 59:1035-48; Oettinger, M.A. *et al.*, 1990, *Science* 248:1517-23; for TdT: Thai, T.H. & Kearney, J. F., 2004, *J Immunol* 173:4009-19; Koiwai, O. *et al.*, 1987, *Biochem Biophys Res Commun* 144:185-90; Peterson, R.C. *et al.*, 1984, *Proc Natl Acad Sci USA* 81:4363-7; for transfection of a host cell with all three of RAG-1, RAG-2 and TdT: U.S. Pat. No. 5,756,323). One skilled in the art can readily select an appropriate non-immune host cell. Examples of host cells include, but are not limited to, yeast and mammalian cells. Specific non-limiting examples include *Saccharomyces cerevisiae*, *Pichia pastoris*, green African monkey kidney (COS) cells, NIH 3T3 cells, Chinese hamster ovary (CHO) cells, BHK cells, human embryonic kidney (HEK 293) cells, Huh7.5 human hepatoma cells, Hep G2 human hepatoma cells, Hep 3B human hepatoma cells, HeLa cells and the like.

[00151] These and other host cells may be used according to contemplated embodiments of the present invention. For example, expression of RAG-1 and/or RAG-2 has been observed in mature B-cells *in vivo* and *in vitro* (Maes *et al.*, 2000, *J Immunol.* 165:703; Hikida *et al.*, 1998, *J Exp Med.* 187:795; Casillas *et al.*, 1995, *Mol Immunol.* 32:167; Rathbun *et al.*, 1993, *Int Immunol.* 5:997; Hikida *et al.*, 1996, *Science* 274:2092).

[00152] RAG-1 and RAG-2 have also been shown to be expressed in mature T-cell lines including Jurkat T-cells. CEM cells have been shown to have V(D)J recombination activity using extrachromosomal substrates (Gauss *et al.* 1998, *Eur J Immunol.* 28:351). Treatment of wild-type Jurkat T cells with chemical inhibitors of signaling components revealed that inhibition of Src family kinases using PP2, FK506

etc. overcame the repression of RAG-1 and resulted in increased RAG-1 expression. Mature T-cells have also been shown to reactivate recombination with treatment of anti-CD3/IL7 (Lantelme *et al.*, 2008, *Mol Immunol.* 45:328).

5 [00153] Tumor cells of non-lymphoid origin have also been shown to express RAG-1 and RAG-2 (Zheng *et al.*, 2007, *Mol Immunol.* 44: 2221, Chen *et al.*, 2007, *Faseb J.* 21:2931). Accordingly, in certain embodiments, these cells may also be suitable for use as host cells in the presently described methods. According to other embodiments that are contemplated herein, reactivation of V(D)J recombination would provide another approach to generating a suitable host cell with inducible recombinase expression.

10 [00154] Use of other host cells is contemplated according to certain embodiments, which may vary depending on the particular mammalian genes that are employed or for other reasons, including a human cell, a non-human primate cell, a camelid cell, a hamster cell, a mouse cell, a rat cell, a rabbit cell, a canine cell, a feline cell, an equine cell, a bovine cell and an ovine cell.

15 [00155] In certain embodiments, the host cell lines can also include added genetic elements that provide useful functionality. For example, Invitrogen provides a flip-in system in which the Frt recombination signal is integrated into different host cell lines (3T3, BHK, CHO, CV-1, 293). Equivalent cell lines incorporating LoxP sites or other sites for targeting integration can be used. Invitrogen also provides tet inducible systems (T-Rex) for 293 or HeLa cell lines. Other inducible systems are also available.

20 [00156] Alternatively, only one of the RAG-1 or RAG-2 genes may be stably integrated into a host cell, and the other gene can be introduced by transformation to regulate recombination. For example, a cell line that is stably transformed with TdT and RAG-2 would be recombinationally silent. Upon transient transformation with RAG-1, or viral infection with RAG-1, the cell lines would become recombinationally active. The skilled person will appreciate from these illustrative examples that other similar approaches may be used to control the onset of recombination in a host cell.

25 [00157] Another approach may be to use specific small interfering RNA (siRNA) to repress the expression in a host cell of RAG-1 and/or RAG-2 by RNA interference (RNAi) (including specific siRNAs the biosynthesis of which within a cell may be

directed by introduced encoding DNA vectors having regulatory elements for controlling siRNA production), and then to relieve such repression when it is desired to induce recombination. For instance, in certain such embodiments a cell line in which active RAG-1 and/or RAG-2 specific siRNA expression is present will be recombinationally silent. Activation of recombination occurs when RAG-1 and/or RAG-2 specific siRNA expression is shut off or repressed. Regulation of such siRNA expression may be achieved using inducible systems like the Tet system or other similar expression-regulating components. These include the Tet/on and Tet/off system (Clontech Inc., Palo Alto, CA), the Regulated Mammalian Expression system (Promega, Madison, WI), and the GeneSwitch System (Invitrogen Life Technologies, Carlsbad, CA). Alternatively, host cells may be transformed with an expression vector that encodes a repressing protein that prevents transcription of the inhibiting RNA.

[00158] In yet another alternative embodiment according to which RAG-1 and/or RAG-2 specific siRNA expression may regulate the recombination competence of the host cell, deletion of the introduced siRNA encoding sequences by use of the Cre/Lox recombinase system (see, for example, Sauer, 1998, *Methods* 14:381 ; Kaczmarczyk *et al.*, 2001, *Nucleic Acids Res* 29:E56; Sauer, 2002, *Endocrine* 19:221; Kondo *et al.*, 2003, *Nucleic Acids Res* 31:e76) may also permit activation of recombination mechanisms. Activation of recombination capability in a host cell may also be achieved by transforming or infecting the cell with an expression construct containing the repressed gene including modified codons so that the gene is not inhibited by the siRNA molecules.

[00159] Substantial impairment of the expression of one or more recombination control elements (for example, one or more of a RAG-1 gene, RAG-2 gene or TdT gene) may be achieved by a variety of methods that are well known in the art for blocking specific gene expression, including antisense inhibition of gene expression, ribozyme mediated inhibition of gene expression, siRNA mediated inhibition of gene expression, and Cre recombinase regulation of expression control elements using the Cre/Lox system. As used herein, expression of a gene encoding a recombination control element is substantially impaired by such methods for inhibiting when host cells are substantially but not necessarily completely depleted of functional DNA or functional mRNA encoding the recombination control element, or of the relevant polypeptide. In

certain embodiments, recombination control element expression is substantially impaired when cells are at least about 50% depleted of DNA or mRNA encoding the endogenous polypeptide (as detected using high stringency hybridization, for example) or at least about 50% depleted of detectable polypeptide (as measured by Western immunoblot, for example); for example, at least 75% depleted or at least 90% depleted.

Screening Assays

[00160] The methods according to the present invention may optionally include one or more screening steps, for example, to screen for expression of variant proteins by the host cells and/or to screen for variants having the desired functionality and/or improvements in functionality.

Expression Assays

[00161] In certain embodiments, the methods of the invention further comprise screening the transformed host cells for expression of variants of the target protein. Various protein expression assays are known in the art and include the use of UV/VS spectrophotometry, fluorescence-based readouts including spectrophotometry and FACS, mass spectrometry and the like. As noted above, in some embodiments, the protein variants may be expressed as fusion proteins comprising additional amino acid sequences to facilitate detection, for example, by localizing the protein to the cell surface or by incorporating a detectable label.

[00162] In certain embodiments in which the protein variants are not localized to the cell surface or secreted, the expression assay may further comprise a cell lysis step, or the protein variant may be assayed intracellularly.

[00163] In some embodiments of the invention, the methods generate high numbers of variants and in such embodiments high throughput screening or selection approaches are generally preferred. Many high throughput screening approaches are well known in the art and can be readily applied to identify and select variant proteins with modified functionality. FACS based and magnetic panning is also well known in the art.

Functional Assays

[00164] In certain embodiments, the methods of the invention further comprise submitting the variant protein(s) to a functional assay to identify those variants having the desired functionality and/or improvements in functionality. The specific assay used
5 will be dependent on the functionality being assessed. Exemplary functionalities that can be detected using a functional assay include, but are not limited to, ligand binding, enzymatic activities, and signalling activities. Various functional assays are known in the art and appropriate assays can be readily selected by one skilled in the art. Commonly used assays include, for example, binding assays, growth assays, reporter
10 gene assays and FACS-based assays.

[00165] The functionality of the variant proteins may be assessed by assaying the cells expressing the variants or the variants may be isolated from the host cells and assayed as isolated proteins.

POLYNUCLEOTIDE COMPOSITIONS

15 [00166] In certain embodiments, the invention provides for polynucleotides capable of undergoing RSS-mediated recombination when introduced into a recombination-competent host cell, and compositions comprising same. The polynucleotide preferably comprises a coding sequence that encodes a target protein as defined herein, such as a ligand-binding protein, into which two or more RSSs have been introduced.
20 The polynucleotide compositions may be provided as isolated polynucleotides or they may be provided as part of a vector, in which case they may be operatively linked to one or more regulatory elements, such as, promoters, enhancers, terminators, alpha-factors, ribosome binding sites, polyadenylation signals and the like, as described above. The present invention also contemplates that the compositions may be provided
25 as host cells that have been transformed with the polynucleotide or a vector comprising the polynucleotide. Examples of suitable host cells include those described above.

[00167] In certain embodiments, the polynucleotide of the composition comprises a nucleic acid sequence that encodes a target protein, such as a ligand-binding protein, and at least one complementary pair of RSSs, in which the two RSSs of the pair are (i)
30 capable of functional recombination with each other; (ii) positioned in a portion of the

nucleic acid sequence that encodes a non-conformational region of the protein, and (iii) spaced preferably 100 base pairs or more apart.

[00168] In some embodiments, the polynucleotide is a “tripartite substrate” in which the polynucleotide comprises two complementary pairs of RSSs, each pair positioned at a different location and each pair positioned in a portion of the nucleic acid sequence that encodes a non-conformational region of the protein.

[00169] In some embodiments, the polynucleotide comprises RSSs that are accompanied by flanking sequences adjacent to one or both of the heptamers of the RSS. In some embodiments, the polynucleotide comprises RSSs that are accompanied by flanking sequences that encode a specific amino acid, or amino acids, or peptide sequence.

[00170] In certain embodiments, the polynucleotide comprises a nucleic acid that encodes a ligand-binding protein and comprises at least one complementary pair of RSSs positioned in a portion of the nucleic acid sequence that encodes a loop region of the protein involved in ligand-binding.

[00171] In certain embodiments, the polynucleotide of the composition comprises a nucleic acid sequence that encodes an antibody or antibody fragment and at least one complementary pair of RSSs, in which the two RSSs of the pair are (i) capable of functional recombination with each other; (ii) positioned within a portion of the nucleic acid sequence that encodes CDR1 or CDR2, and (iii) spaced preferably 100 base pairs or more apart.

APPLICATIONS

[00172] In accordance with one aspect of the present invention, the methods can be used to generate variants of a target protein with a desired functionality or an improvement or modification in an existing functionality. In certain embodiments, the methods are employed to generate a large number of variants of the target protein for subsequent screening for a desired, improved or modified functionality. In certain embodiments of the invention, the functionality may be a binding affinity or specificity for a selected ligand.

[00173] In certain embodiments, the methods are used to generate variants of a ligand-binding protein with modulated binding properties, for example, modulated affinity. Modulated affinity may comprise a decrease in affinity or an increase in affinity over the native protein. In some embodiments, therefore, the methods are used to generate
5 variants of a ligand-binding protein with decreased affinity for a ligand. In some embodiments, the methods are used to generate variants of a ligand-binding protein with increased affinity for a ligand. Modulated binding properties also encompass in some embodiments a modified specificity, for example, an increase in specificity, a decrease in specificity or an altered specificity, for example, such that the protein binds
10 a ligand other than the native ligand(s).

[00174] In certain embodiments, the invention provides for the use of the methods to generate ligand-binding proteins with modified binding properties (for example, modified antibodies, avimers, adnectins, or other antibody mimetics) for therapeutic purposes, for diagnostic purposes, for drug targeting (for example, through the use of a
15 modified ligand-binding protein that targets a protein on a particular cell or tissue type as a targeting moiety for attachment to a therapeutic or diagnostic compound), or for research applications (such as screening assays, chromatography and the like).

KITS

[00175] In one aspect, the invention provides for kits comprising a polynucleotide
20 capable of undergoing RSS-mediated recombination when introduced into a recombination-competent host cell, or a composition comprising a polynucleotide capable of undergoing RSS-mediated recombination when introduced into such a host cell, as described above.

[00176] When the kit comprises a composition, the composition may comprise an
25 isolated polynucleotide, a polynucleotide comprised by a vector (in which case the polynucleotide may be operatively linked to one or more regulatory elements, such as, promoters, enhancers, terminators, alpha-factors, ribosome binding sites, polyadenylation signals and the like), or a host cell that has been transformed with the polynucleotide or a vector comprising the polynucleotide.

[00177] When the kit comprises an isolated polynucleotide, the kit may further comprise a vector suitable for expression of the polynucleotide and/or a recombination-competent host cell.

[00178] The kit may further comprise vectors encoding one or more of RAG-1, RAG-2 and TdT that are suitable for transforming a host cell such that the host cell expresses, or is capable of expressing, RAG-1 and/or RAG-2 and/or TdT.

[00179] The kit may further comprise one or more additional components to assist with cloning the polynucleotide and/or transformation of host cells, such as buffers, enzymes, selection reagents, growth media and the like.

[00180] One or more of the components of the kit may optionally be lyophilised and the kit may further comprise reagents suitable for the reconstitution of the lyophilised components. Individual components of the kit would be packaged in separate containers and, associated with such containers, can be instructions for use. The instructions for use may be provided in paper form or in computer-readable form, such as a disc, CD, DVD or the like.

[00181] To gain a better understanding of the invention described herein, the following examples are set forth. It will be understood that these examples are intended to describe illustrative embodiments of the invention and are not intended to limit the scope of the invention in any way.

EXAMPLES

EXAMPLE 1: Preparation Of Constructs For Introducing Sequence Diversity Into An Avimer

[00182] A domain or avimer-encoding DNA sequences were generated by gene synthesis by GeneArt® (Invitrogen, Carlsbad, CA). The sequences were codon-optimized and included RSSs in the appropriate positions, an IgG1 hinge region, CH2, CH3, a 5' hemagglutinin (HA) tag, a PDGFR transmembrane domain sequence and a selectable marker, as detailed in Tables 2 and 3 below.

[00183] E188 is a single A domain avimer construct and includes a pair of RSSs introduced into loop 1 of the construct and a pair of RSSs introduced into loop 2 of the construct together with flanking sequences encoding GY amino acid residues, which were selected to be a duplication of the naturally occurring residues, but could also have been non-endogenous sequences (see Figure 3A-C).

[00184] E189 is a double A domain avimer construct and includes a pair of RSSs in each loop 1 of the construct (see Figure 4). E189 also includes stop codons in other reading frames in the 3' loop 1 to 5'loop 1.2 region, but does not include flanking sequences.

10 [00185] Portions of the E188 and E189 sequences are shown in Figure 1 [SEQ ID NO:28] and Figure 2 [SEQ ID NO:29], respectively. The complete vector sequences are provided in Figure 17 [SEQ ID NO:1] and Figure 18 [SEQ ID NO:40], respectively.

[00186] Multiple A domain avimers can also be constructed (see Figure 5).

15 **Table 2: Sequence Annotation for [SEQ ID NO:28]**

Leader	10-66
HA-tag	67-93
Coding sequences 5' loop 1	94-102
Inserted flanking sequence	NA
23 bp RSS (>)	103-141
Intervening sequence	142-722
12 bp RSS (<)	723-750
Inserted flanking sequence	NA
Coding interverning sequence 3' Loop 1/5' Loop 2	751-771
inserted flanking sequence (GGCTAC)	772-777
12 bp RSS (>)	778-805
Intervening sequence	806-1429
23 bp RSS (<)	1430-1468
Inserted flanking sequence	NA
3' Loop 2 - Loop 5	1469-1501
Avimer linker	1502-1561
IgG hinge CH2-CH3	1562-2257
Transmembrane sequence	2258-2425

Table 3: Sequence Annotation for [SEQ ID NO:29]

Leader	10-66
HA-tag	67-93
Coding sequences 5' loop 1	94-102
Inserted flanking sequence	NA
23 bp RSS	103-141
Intervening sequence	142-722
12 bp RSS	723-750
Inserted flanking sequence	NA
Coding sequences 3' Loop 1 -Loop 5 linker 5' Loop 1.2	751-870
Inserted flanking sequence	NA
12 bp RSS	871-898
Intervening sequence	899-1522
23 bp RSS	1523-1561
Inserted flanking sequence	NA
Coding sequences 3' Loop 1.2 - loop 5.2	1562-1609
Avimer linker	1610-1669
IgG1 hinge CH2-CH3	1670-2365
Transmembrane sequence	2366-2533
Leader	10-66
HA-tag	67-93
Coding sequences 5' loop 1	94-102
Inserted flanking sequence	NA
23 bp RSS	103-141
Intervening sequence	142-722
12 bp RSS	723-750
Inserted flanking sequence	NA
Coding sequences 3' Loop 1 -Loop 5 linker 5' Loop 1.2	751-870
Inserted flanking sequence	NA
12 bp RSS	871-898
Intervening sequence	899-1522
23 bp RSS	1523-1561
Inserted flanking sequence	NA
Coding sequences 3' Loop 1.2 - loop 5.2	1562-1609
Avimer linker	1610-1669
IgG1 hinge CH2-CH3	1670-2365
Transmembrane sequence	2366-2533

- [00187] The synthesized DNA was cloned into a modified pcDNA (Invitrogen, Carlsbad, CA) that contains a consensus Kozak sequence and a mammalian leader signal sequence ([SEQ ID NO:36]; see Example 7) for efficient secretion or surface expression of the recombined avimers. The modified pcDNA acceptor vector allows for cloning of the avimer construct so that the 3' end is fused to the Fc portion of

human IgG1 followed by a PDGFR transmembrane domain and selectable marker such that the recombined molecules are surface expressed and can be selected for in-frame products. The nucleotide sequences for the IgG hinge through CH3 sequences and a transmembrane domain are shown in Figure 7D [SEQ ID NO:35]. The avimer scaffold
5 was cloned at the KpnI site (bolded in Figure 7D), which translates as a Gly-Thr prior to the hinge sequences of IgG1.

EXAMPLE 2: Generation of Surface Expressed Avimer Mutants

[00188] Avimer vectors containing E188 prepared as described in Example 1 were transfected into a recombination competent cell line (see Example 10) and stable
10 neomycin integrants were generated. The sequences of the expressed avimer mutants were obtained as described in Example 4 below.

EXAMPLE 3: Generation Of Libraries of Surface Expressed Avimer Mutants

[00189] Avimer vectors containing E188 prepared as described in Example 1 were stably integrated into a recombination competent cell line. Stable integrants were
15 expanded and then transfected with plasmids expressing RAG1/RAG2/TdT. The transfection was carried out using 1×10^7 stable integrants transfected with 8ug each of RAG1, RAG2 and TdT expression vectors using a 3:1 ratio of linear PEI (1mg/ml) to DNA. (See Example 10 for details).

[00190] RAG1/RAG2/TdT treated cells were then stained using anti-IgG Fc to confirm
20 surface expression of the recombined avimer molecules. Approximately 1×10^6 cells were stained with 1ug/ml Biotin conjugated anti-human IgG Fc (Jackson Laboratories) for 30min. The cells were then washed twice and stained with streptavidin-conjugated Alexa-647 for 30min. Samples were subsequently washed twice, resuspended in 300ul of PBS and analyzed using flow cytometry. The recombined population was shown to
25 have high uniform expression. The sequences of the expressed avimer mutants were obtained as described in Example 4 below.

EXAMPLE 4: Sequence Analysis Of Avimer Mutants (Single A Domain)

[00191] RNA samples obtained from FACS sorted cells (Example 3) were used for sequence analysis of the expressed avimer variants. mRNA from approximately 10^6

recombined cells was purified using Qiagen RNeasy RNA purification kit as per the manufacturer's recommendations. cDNA synthesis was carried out using Superscript enzyme (Invitrogen, Carlsbad, CA) as per the manufacturer's recommended protocol and primer MG59 (sequence 5'-TCTTGGCATTATGCACCTCCACGCCGTCC-3' [SEQ ID NO:30]).

[00192] The cDNA was then used as a template and amplified using primer MG301 (sequence 5'-GAGAGAGATTGGTCTCGAGAACCCACTGCTTACTGCTCGACGATCTGAT-3' [SEQ ID NO:31]), which anneals in the 5' UTR region, and primer MG58 (sequence 5'-GTCTTCGTGGCTCACGTCCACCACACGCA-3' [SEQ ID NO:32]), which anneals internal to the MG59 primer used in the RT reaction.

[00193] The amplified product was purified using a Qiagen PCR clean up kit as per the manufacturer's recommended protocol and eluted into 35ul of water. The purified PCR product was then digested with BsaI (NEB) and cloned into the modified pcDNA acceptor vector (Invitrogen, Carlsbad, CA) with corresponding compatible ends. Plasmid DNA from *E. coli* cultures was purified using Qiagen Miniprep kit and avimer sequences were analyzed using primer MG60 (sequence 5'-CTGACCTGGTTCTTGGTCAGCTCATCCCG-3' [SEQ ID NO:33]).

[00194] The results are presented in Tables 4A,B and 5 below.

20 **Table 4A: Nucleotide Sequence Analysis Of Single A Domain Avimer Variants (Loop 1)**

Mutant #	5' Deletions	Additions	3' Deletions
1	-1		-2
2	0	AGGGCCAAGA [SEQ ID NO:16]	-15
3	-1	GAG	-2
4	0	C	-1
5	-2	TAGGGGGTTCCAGT [SEQ ID NO:17]	-13
6	0	AGAA	-3
7	0	AGTGGGGAT	0
8	-1	CCC	-6
9	-1	CCT	-2

Mutant #	5' Deletions	Additions	3' Deletions
10	-2	T	0
11	-8	TCC	-4
12	0	AC	-3
13	0	AGAAGG	-3
14	-3	TTATTA	-1
15	-2	AAGAC	-12
16	0	CC	-5
17	-1	CTC	-3
18	0	AGG	0
19	0		-1
20	0	CG	-5
21	0	AGAC	-1

Table 4B: Nucleotide Sequence Analysis Of Single A Domain Avimer Variants (Loop 2)

Mutant #	5' Deletions	Additions	3' Deletions
1	0	GA	-2
2	-7	TGGGGTTAAGCCTC [SEQ ID NO:18]	-2
3	0		0
4	0	GGG	-6
5	-2	GAG	0
6	-12	CCCTCCGTCCTACCTC [SEQ ID NO:19]	-2
7	-12	C	-4
8	-14	TCCAGTGCGGCTCCGGGA [SEQ ID NO:20]	-24
9	-2	TC	0
10	-2		-3
11	-4	CTACA	-4
12	-4	CG	-3
13	0		-3
14	0		-2
15	0	GTC	-2
16	0		-6
17	-13		-4
18	-23	GGAGCCGCACTGGA [SEQ ID NO:21]	0
19	-2		-6
20	-2	CT	-6

Mutant #	5' Deletions	Additions	3' Deletions
21	-2	TCCC	-2

Table 5: Amino Acid Sequence Analysis Of Single A Domain Avimer Variants

Mutant #	Loop 1 (5')	Loop 1 (3') /Loop2 (5')	Loop 2 (3') and loop 3	Total aa Length (from CAP to GYC)	SEQ ID NO
Parent	DYACAP	SQFQCGSGY	GYCISQRWVCD	15	22
1	DYA	FQFQCGSGYN	CISQRWVCD	10	23
2	DYACAP	TSSSAAPAY	CISQRWVCD	13	24
3	DYACAP	RRQFQCGSGY	YCISQRWVCD	14	25
4	DYACA	LLASSAAPAT	YCISQRWVCD	13	26
5	DYACA	QDAAPATS	YCISQRWVCD	13	27
6	DYACAP	PQFQCGSGY	CISQRWVCD	13	42
7	DYACAP	SSSSD	CISQRWVCD	13	43
8	DYACAP	RSRSRTGT	GYCISQRWVCD	15	44
9	DYACAP	ASSAAPA	CISQRWVCD	13	45
10	DYACAP	RFQCGSGS	CISQRWVCD	13	46
11	DYACAP	RRQFQCGSGFP	YCISQRWVCD	14	47
12	DYACAP	QFQCGSGYD	YCISQRWVCD	14	48
13	DYACAP	RAKRLWGAS	YCISQRWVCD	14	49
14	DYACAP	SQFQCGSGY	GYCISQRWVCD	15	50
15	DYACAP	RQFQCGSGYG	CISQRWVCD	13	51
16	DYACA	LGGSSAAPAE	GYCISQRWVCD	14	52
17	DYACAP	RTVPVPLRPTS	YCISQRWVCD	14	53
18	DYACAP	SGDS QFQCH	CISQRWVCD	13	54
19	DYACAP	PSSSAAPG	VCD	7	55
20	DYACAP	LQFQCGSGF	GYCISQRWVCD	15	56
21	DYACA	LASSAAPA	YCISQRWVCD	13	57

- [00195] This data indicates that net size of the product is still smaller than the original product indicating that this is a situation in which additional flanking sequences may be beneficial. The data also demonstrated that a large fraction of products used the other reading frames for the RSS flanked cassette and as a result eliminated the cysteine residue. To counter this, an alternative cassette was designed as described in Example 6 below.

EXAMPLE 5: Large Libraries Of Avimer Mutants From An Inducible Cell Line

[00196] Avimer vectors prepared as described in Example 1 will be stably integrated into a cell line which constitutively expresses RAG2, Tdt and the Tet repressor (TetR, Life Technologies). The cell line will also have stably integrated RAG1 whose expression can be regulated via the addition of tetracycline (T-Rex System, Life Technologies). Stable avimer mutator integrants will be expanded to 100 million cells and RAG1 expression induced via the addition of 1ug/ml tetracycline (Life Technologies). Tetracycline will be maintained for a period of 2-14 days, or longer, to allow the V(D)J reaction to proceed.

10 **EXAMPLE 6: Alternative Construct For Introducing Sequence Diversity Into An Avimer**

[00197] The cassette used in Example 1 (see Figure 6A) was redesigned as shown in Figure 6B. The alternate cassette includes as additional flanking sequences, a TAC at both the 5' end and the 3' end (adding potential tyrosine if not deleted). The modified cassette also includes nucleotide changes that add cysteines in the other frames to help ensure retention of a cysteine in the final product.

EXAMPLE 7: Preparation Of Constructs For Introducing Sequence Diversity Into A Fibronectin Domain

[00198] Two constructs based on the 10Fn3 scaffold will be prepared as follows. The nucleotide sequences containing the 10Fn3 exon (Figure 7A [SEQ ID NO:34]) will be cloned behind the CMV promoter and a heterologous leader sequence and include a downstream BGH poly (A). The nucleotide sequence of the leader is:

5'-atggagtttgggctgagctggccttttctgtggctattttaaaaggtgtccagtgt-3' [SEQ ID NO:36]

[00199] The 10Fn3 scaffold will be cloned in frame with the IgG1 hinge through CH3 sequences and a transmembrane domain for cell surface expression, as well as a selectable marker cassette that allows for in-frame selection of recombination products. The nucleotide sequences for the IgG hinge through CH3 sequences and a transmembrane domain are shown in Figure 7D [SEQ ID NO:35]. The 10FN3 scaffold

will be cloned at the KpnI site (bolded in Figure 7D), which translates as a Gly-Thr prior to the hinge sequences of IgG1.

5 [00200] RSSs will be introduced into the region of the 10Fn3 nucleotide sequence that encodes the FG loop (see Figure 8, which depicts the location of the loop regions of 10Fn3). In the first construct, a 289bp (heptamer to heptamer) sequence cassette containing one pair of RSSs will be introduced (see Figure 9). The sequence of this construct is shown in Figure 10A [SEQ ID NO:37], in which the 23bp RSS and the 12bp RSS are shown in bold.

10 [00201] In the second construct, two pairs of RSSs will be introduced into the FG loop (see Figure 11) using a 289bp (heptamer to heptamer) sequence cassette for the 5' RSSs and a 427bp (heptamer to heptamer) sequence cassette for the 3' RSSs. The sequence of this construct is shown in Figure 10B [SEQ ID NO:38], in which the 23bp RSSs and the 12bp RSSs are shown in bold.

15 [00202] In the second construct, repeat sequences will be included with the RSSs as shown in Figure 11. As an alternative to repeat sequences, heterologous sequences could be included, for example, sequences encoding the amino acid histidine that could potentially allow for binders to be isolated that had pH dependent binding. Also in the second construct, stop codons will be eliminated from the sequence being flanked by the RSSs so that all three frames have the potential to generate a functional protein.

20 [00203] Additional examples of potential positioning of RSSs in the 10Fn3 sequence that allow for simultaneous diversification of two loops are shown in Figure 12A (diversification of the DE and FG loops) and 12B (diversification of the BC and FG loops). Diversification of the BC and DE loops is also contemplated.

25 **EXAMPLE 8: Introduction Of Sequence Diversity Into CDR2 Of An Immunoglobulin Heavy Chain**

[00204] The sequence of the heavy chain VDJ regions is shown in Figure 16A, with CDR1 and CDR2 indicated in bold. The following strategy (depicted in Figure 14) will be employed in order to introduce sequence diversity into CDR2 of this heavy chain.

[00205] A 289 bp (from heptamer to heptamer) sequence cassette including a 5' 23bp RSS and a 3' 12bp RSS (shown below [SEQ ID NO:6]) will be introduced into CDR2 and the resulting modified heavy chain coding sequence cloned behind the CMV promoter and a heterologous leader sequence and will include a downstream BGH poly (A), as described in Example 7 for the modified 10Fn3. The sequence of the modified heavy chain coding sequence is shown in Figure 16B.

[00206] 5'-

cacagtggtagtactccactgtctgggtgtacaaaaacctccctgcacgcctcttaacctcacaattctgtggcgccgct
ttgtagccagaccctcgggtcaactggatgtcacaactggcacctgagattggaacataaaaaacaaatattcttactattaatcat
 10 **gttatcagagaactccctgaagtccagtcagtactgtgactagctaattagtcagttacttaagcgtctatccaagtgcaaagg**
gacaggagggtttttgtaagggtgtatcactgtg-3' [SEQ ID NO:6]

EXAMPLE 9: Introduction Of Sequence Diversity Into Both CDR1 and CDR2 Of An Immunoglobulin Heavy Chain

[00207] The heavy chain described in Example 3 will be used and the following strategy (depicted in Figure 15) will be employed in order to introduce sequence diversity into CDR1 and CDR2. The following cassettes will be used.

[00208] CDR1 cassette, including a 5' 23bp RSS and a 3' 12bp RSS with a 222bp intervening sequence and a trinucleotide repeat sequence (AGC) located 5' of the 23bp RSS heptamer:

20 [00209] 5'-

agccacagtggtagtactccactgtctgggtgtacaaaaacctccctgcacgcctcttaacctcacaattctgtggcgcc
cgtttgtagccagaccctcgggtcaactggatgtcacaactggcacctgagattggaacataaaaaacaaatattcttactatta
atcatgttatcagagaactccctgaagtccagtcagtactgtgactagctaattagtcagttacttaagcgtctatccaagtgca
aaggacaggagggtttttgtaagggtgtatcactgtg-3' [SEQ ID NO:7]

25 [00210] CDR2 cassette, including a 5' 12bp RSS and a 3' 23bp RSS with a 360bp intervening sequence and a trinucleotide repeat sequence (ACC) located 5' of the 12bp RSS heptamer:

[00211] 5'-

accacagtgtatcagcccttaacaaaaaccctactgcaacctggcggttaatagacgtccggaagcacactggctgagt

aaattcctagtgttctccatccttacctcggagccagagtagcaggagccactagccagcttgggtcttctatcgcgagtcgta
 ttaatttcgataagccagcaagcagtggttctctagtagccagctgcctccttctctgggccagcgtcctctgtcctggagc
 tgggagataatgtccgggggctccttggtctgcgctgggcaaagggtgggcagagtcagcttgtgctggggacaaaatga
 ccttgggacacgggtcgacgggctggctgccacggccggccgggacagtcggagagtcagggtttttgtacaccagaca
 5 **gtggagtactaccactgtg-3'** [SEQ ID NO:8]

[00212] A modified heavy chain coding sequence including the CDR1 and CDR2 cassettes shown above will be gene synthesized and cloned behind the CMV promoter and a heterologous leader sequence. The construct will also include a downstream BGH poly(A), as described in Example 7 for the modified 10Fn3. The sequence of the
 10 modified heavy chain coding sequence is shown in Figure 16C [SEQ ID NO:5].

EXAMPLE 10: Recombination And Expression Of Recombination Substrates

[00213] In brief, HEK293 cells, containing an integrated LoxP sequence (Fukushige *et al.*, 1992, *PNAS USA*, 89:7905-7909; Baubonis *et al.*, 1993, *NAR*, 21(9):2025-2029; Thomson *et al.*, 2003, *Genesis*, 36:162-167) were maintained in DMEM media with
 15 10% FBS. Integration into the LoxP site was shown to support high protein expression and also support V(D)J recombination of inserted substrates and provides an easy method to generate integrants with the required properties. Vectors comprising the recombination substrate were designed to include a LoxP site for targeted integration which is in-frame with a codon-optimized hygromycin open reading frame. Bipartite
 20 vectors were also designed so that productive rearrangements will be in-frame with the selectable marker neomycin. The neomycin gene is cloned in-frame with a transmembrane domain both of which are positioned downstream of a furin cleavage site that allows for secretion of the encoded protein (see Figure 19 and SEQ ID NO:41, as an example).

[00214] For example, for bipartite substrates, HEK293 cells containing the LoxP site were co-transfected with the bipartite substrate containing the hygromycin gene for selection of stable integrants and a vector expressing the CRE protein at a ratio of 10:1 substrate to CRE expressing vector. Specifically, a 10 cm dish of cells was transfected using a polyethylenimine (PEI; 1mg/ml) to DNA ratio of 3:1. 21.6ug of substrate
 25 DNA was mixed with 2.4ug of CRE expression vector and placed in 1.5 ml OptiMEM™ media and mixed with an equal volume of OptiMEM™ containing the 72
 30

ul of PEI. The transfection was carried out for 24 hours and the following day the transfection media was removed and replaced with fresh DMEM media. The following day the transfected cells were split into ten 10 cm² dishes and selection was carried out for approximately 2 weeks. A pool of stable hygromycin resistant cells were selected.

- 5 The cell line was subsequently expanded in the un-recombined state to approximately 10 million cells and transfected with RAG-1, RAG-2 and TdT. 72 hours post-transfection the cells were placed in neomycin selection (1mg/ml).

[00215] Tripartite recombination substrates used vectors designed such that puromycin could be used for in-frame selection. Tripartite vectors also included a modified
10 neomycin cassette that allows maintenance of the unrecombined substrate during expansion.

EXAMPLE 11: Sequence Analysis Of Avimer Mutants (Double A Domain)

[00216] Avimer vectors containing E189 prepared as described in Example 1 were stably integrated into a recombination competent cell line as described in Example 3
15 and sequence analysis was conducted as described in Example 4. The results are shown in Tables 6-8, below.

Table 6: Nucleotide Sequence Analysis Of Double A Domain Avimer Variants

Mutant #	A1- L1 5' Deletions	A1-L1 Additions	A1-L1 3' Deletions	A2-L1 5' Deletions	A2-L1 Additions	A2-L1 3' Deletions
1	-4	T	0	-2	TCC	-1
2	0		0	-4	GAG	-5
3	0	GG	-2	-2	AC	0
4	-1	T	0	0	CTCCTT	0
5	-2		-4	-6	GG	-5
6	-3		0	-13	CGAGGGT	0
7	0	GGAACAGG	-2	-10	GGCC	0

Table 7: Amino Acid Sequence Analysis Of Double A Domain Avimer Variants (A Domain 1)

Mutant #	A1 Loop 1 (5')	A1-Loop 1 (3') - End A1	Total aa Length from CLP_V to DQFRC_D	SEQ ID NO
Parent	DYACLP	DQFRCGNGQCIPLDWVCDGVND CPDSD EEGC	8	58
1	DYACL	DQFRCGNGQCIPLDWVCDGVND CPDSD EEGC	7	59
2	DYACLP	DQFRCGNGQCIPLDWVCDGVND CPDSD EEGC	8	60
3	DYACLP	GQFRCGNGQCIPLDWVCDGVND CPDSD EEGC	8	61
4	DYACLP	DQFRCGNGQCIPLDWVCDGVND CPDSD EEGC	8	62
5	DYACL	QFRCGNGQCIPLDWVCDGVND CPDSD EGC	6	63
6	DYACL	DQFRCGNGQCIPLDWVCDGVND CPDSD EEGC	7	64
7	DYACLP	GTGQFRCGNGQCIPLDWVCDGVND CPD SDEEGC	10	65

Table 8: Amino Acid Sequence Analysis Of Double A Domain Avimer Variants (A Domain 2)

Mutant #	A1-A2 linker	A2 Loop 1 (5')	A2 Loop 1 (3')- Loop 2	Total aa Length from CAPSQ_D to FQC_J	SEQ ID NO
Parent	PPRT	CAPSQ	FQCGSGYCI	8	66
1	PPRT	CAPSL	QCGSGYCI	8	67
2	PPRT	CAPRR	CGSGYCI	6	68
3	PPRT	CAPSH	FQCGSGYCI	8	69
4	PPRT	CAPSLL	FQCGSGYCI	10	70
5	PPRT	CAPG	CGSGYCI	4	71
6	PPRT	CE	FQCGSGYCI	4	72
7	PPRT	CAA	FQCGSGYCI	5	73

EXAMPLE 12: Introduction Of Sequence Diversity Into CDR2 Of An Immunoglobulin Light Chain

[00217] A plasmid, ITS001-V655 (SEQ ID NO:74; Figure 20), was constructed that encodes sequences that permit light chain CDR2 optimization of a HER2-specific antibody. The plasmid consists of a light chain CDR2 optimization cassette, a
5 membrane-anchored heavy chain expression cassette and elements required for propagation in *E. coli* and targeted mammalian cell integration.

[00218] Details of the CDR2 optimization cassette are shown in Figure 21A. The light chain variable region is interrupted at CDR2 by three nucleotides of flanking sequence
10 ('CTG', position 5870 to 5872), a 23-bp RSS (position 5873 to 5911), a spacer region, a 12-bp RSS in the inverted orientation (position 6197 to 6224) and three nucleotides of flanking sequence ('TCC', position 6225 to 6227). The light chain variable region is followed by an intron, the kappa constant region, a furin cleavage site, a transmembrane domain and a G418 resistance marker that provides selection for in-
15 frame kappa genes following RAG-mediated recombination.

[00219] A stable cell line was generated by using Cre recombinase to target ITS001-V655 to a locus suitable for RAG-mediated recombination. The line was expanded, recombination induced and cells expressing recombined in-frame light chain genes were selected using G418. This population was enriched for binding to HER2 by
20 staining the recombined cells using 1ug/ml Biotin-HER2 ECD and then isolated using MACS MicroBeads as per manufacturer's suggested protocol. The enriched population, assigned the name ITS001-L145, was analyzed by flow cytometry for expression of cell surface antibody and for binding to HER2. The cells were stained with anti-human kappa-PE (1:5000, VENDOR), 1ug/ml Biotin HER2 ECD and 1ug/ml Streptavidin
25 Alexa 647 (Jackson Laboratories) and then analyzed on a C6 Accuri. As shown in Figure 21B, cells within the population had different ratios of HER2 binding to antibody expression, which suggested that antibodies with different affinities for HER2 had been generated. When the ITS001-L145 line was compared to cells expressing the original HER2 antibody, the majority of IT001-L145 events had a higher ratio of HER2
30 binding to antibody expression.

[00220] Flow cytometry was used to isolate cells from ITS001-L145 with the highest HER2 to antibody expression ratio. Total RNA was extracted followed by RT-PCR of the light chain variable gene. The PCR product was then cloned into an expression vector with the original heavy chain expression cassette. Isolated clones were transiently transfected into HEK-293 cells via PEI mediated transfection and compared to the original antibody in a FACS-based assay as described above for HER2 binding and antibody expression 48hrs post transfection. As shown in Figure 22A & B, clones with substantially higher ratios than the original antibody were found. These results demonstrate that V(D)J recombination can be used to precisely target mutations to a selected region of a protein and that sufficient diversity can be generated to create antibody variants with improved function (affinity).

[00221] CDR2 sequences from isolated clones are shown in Table 9.

Table 9: Light Chain CDR2 Sequences from Isolated Clones

Clone ID	Amino Acid Sequence at Light Chain CDR2	SEQ ID NO
Original	YAASS - - - - - LQS	75
2	FVAST - - - - - LQS	76
4	YAASSLQ - - - GLS	77
5	YAASSLPPS - LQS	78
7	FVASR - - - - - LQS	79
9	YAASSQAGLSLQS	80

EXAMPLE 13: Introduction Of Sequence Diversity Into CDR1 Of An Immunoglobulin Light Chain

[00222] ITS001-P126 consists of a library of plasmids that permit light chain CDR1 optimization of a HER2-specific antibody. Each plasmid consists of a light chain CDR1 optimization cassette, a membrane-anchored heavy chain expression cassette and elements required for propagation in *E. coli* and targeted mammalian cell integration.

[00223] In each plasmid, the light chain variable region is interrupted within CDR1 by the addition of flanking 23-bp RSS and 12-bp RSS. The light chain variable region is followed by an intron, the kappa constant region, a furin cleavage site, a transmembrane domain and a G418 resistance marker that provides selection for in-
5 frame kappa genes following RAG-mediated recombination. Additional variations of the CDR1 optimization vector were generated including different flanking sequences as well as additional mutations upstream or downstream of the break targeted by the RSSs. These changes introduce sequence diversity in addition to that resulting from RAG-mediated recombination.

10 [00224] A pool of stable cell lines incorporating different CDR1 optimization vectors was generated. The cell lines were generated by using Cre recombinase to target each CDR1 recombination substrate to a locus suitable for RAG-mediated recombination. The line was expanded, recombination induced and cells expressing recombined in-
15 frame light chain genes selected using G418. This population was enriched for binding to HER2 in two rounds using MACS MicroBeads. The enriched population, assigned the name ITS001-L167, was analyzed by flow cytometry for expression of cell surface antibody and for binding to HER2. As shown in Figure 23A, cells within the population had different ratios of HER2 binding to antibody expression, which suggested that antibodies with different affinities for HER2 had been generated. When the ITS001-
20 L167 line was compared to cells expressing the original HER2 antibody, the majority of IT001-L167 events had a higher ratio of HER2 binding to antibody expression.

[00225] Flow cytometry was used to isolate cells from ITS001-L167 with the highest HER2 to antibody expression ratio. Total RNA was extracted followed by the generation of light chain gene cDNA and cloning into an expression vector with the
25 original heavy chain expression cassette. Clones were isolated and compared to the original antibody in a FACS-based assay for HER2 binding and antibody expression. As shown in Figure 23B & C, clones with substantially higher ratios than the original antibody were found.

[00226] CDR1 sequences from isolated clones are shown in Table 10.

Table 10: Light Chain CDR1 Sequences from Isolated Clones

Clone ID	Amino Acid Sequence at Light Chain CDR1	SEQ ID NO
Original	RASQSI - - - SSYLN	81
1	RHSQRKSDVSGYAN	82
8	RHSQRKWDVSGYAN	83
10	RAPQPY - - IRGYLN	84
15	RHSQRKFDVSGYAN	85

[00227] The disclosures of all patents, patent applications, publications and database entries referenced in this specification are hereby specifically incorporated by reference in their entirety to the same extent as if each such individual patent, patent application, publication and database entry were specifically and individually indicated to be incorporated by reference.

[00228] Although the invention has been described with reference to certain specific embodiments, various modifications thereof will be apparent to those skilled in the art without departing from the spirit and scope of the invention. All such modifications as would be apparent to one skilled in the art are intended to be included within the scope of the following claims.

THE EMBODIMENTS OF THE INVENTION IN WHICH AN EXCLUSIVE PROPERTY OR PRIVILEGE IS CLAIMED ARE DEFINED AS FOLLOWS:

1. A method of generating variants of a target protein comprising the steps of:
 - (i) providing a polynucleotide comprising a nucleic acid sequence encoding the target protein and comprising a sequence cassette that comprises a first recombination signal sequence (RSS) linked by an intervening nucleotide sequence to a second RSS, the first RSS capable of functional recombination with the second RSS, wherein the first and second RSS are located in a portion of the nucleic acid sequence that encodes a non-conformational region of the target protein, and wherein the intervening nucleotide sequence is 100 base pairs or more in length;
 - (ii) introducing the polynucleotide into a recombination-competent host cell, and
 - (iii) culturing the host cell *in vitro* under conditions allowing recombination and expression of the nucleic acid sequence,thereby generating variants of the target protein.
2. The method according to claim 1, wherein the recombination-competent host cell is capable of expressing RAG-1 and RAG-2, wherein expression of at least one of RAG-1 and RAG-2 is under inducible control, and wherein step (iii) comprises inducing expression of the at least one of RAG-1 and RAG-2.
3. The method according to claim 2 further comprising the step of expanding the host cell prior to step (iii).
4. The method according to any one of claims 1, 2 or 3 further comprising the step of screening the variants for a desired functional activity.
5. The method according to any one of claims 1, 2, 3 or 4, wherein the target protein is a ligand-binding protein and the non-conformational region of the target protein is a region involved in ligand binding.

6. The method according to claim 5, wherein the region involved in ligand binding comprises a loop.
7. The method according to any one of claims 1 to 6, wherein the target protein is an immunoglobulin.
8. The method according to claim 7, wherein the sequence cassette is located within a portion of the nucleic acid sequence encoding a CDR of the immunoglobulin.
9. The method according to claim 8, wherein the CDR is a light chain CDR.
10. The method according to claim 8, wherein the CDR is a heavy chain CDR.
11. The method according to any one of claims 8, 9 or 10, wherein the CDR is CDR1 or CDR2.
12. The method according to any one of claims 8, 9 or 10, wherein the CDR is CDR3.
13. The method according to any one of claims 1 to 6, wherein the target protein is an avimer or a fibronectin domain.
14. The method according to any one of claims 1 to 13, wherein the sequence cassette further includes flanking sequences adjacent to one or both of the first and second RSSs.
15. The method according to any one of claims 1 to 14, wherein the nucleic acid sequence further comprises a second sequence cassette that comprises a third RSS linked by a second intervening nucleotide sequence to a fourth RSS, the third RSS capable of functional recombination with the fourth RSS.
16. The method according to claim 15, wherein the second sequence cassette further includes flanking sequences adjacent to one or both of the third and fourth RSSs.
17. The method according to claim 14 or 16, wherein the flanking sequences encode one or more histidine residues.

18. The method according to claim 15 or 16, wherein the target protein is a ligand-binding protein and the sequence cassette and the second sequence cassette are each positioned within a portion of the nucleic acid sequence that encodes a region involved in ligand binding.

19. The method according to claim 18, wherein the region involved in ligand binding comprises a loop.

20. The method according to claim 19, wherein the sequence cassette and the second sequence cassette are both positioned within a portion of the nucleic acid sequence that encodes the same loop.

21. The method according to claim 19, wherein the sequence cassette and the second sequence cassette are each positioned within a portion of the nucleic acid sequence that encodes a different loop.

22. The method according to any one of claims 1 to 21, wherein the polynucleotide further comprises additional coding sequences that encode a membrane anchor domain peptide.

23. The method according to any one of claims 1 to 22, wherein the recombination-competent host cell is capable of expressing TdT.

24. The method according to any one of claims 1 to 23, wherein the polynucleotide is stably integrated into the genome of the host cell.

25. The method according to claim 24, wherein the polynucleotide is stably integrated into the genome of the host cell as a single copy.

26. A polynucleotide comprising a nucleic acid sequence encoding a target protein and comprising a sequence cassette that comprises a first recombination signal sequence (RSS) linked by an intervening nucleotide sequence to a second RSS, the first RSS capable of functional recombination with the second RSS, wherein the first and second RSS are located in a portion of the nucleic acid sequence that encodes a non-conformational region of the target protein, and wherein the intervening nucleotide sequence is 100 base pairs or more in length.

27. The polynucleotide according to claim 26, wherein the target protein is a ligand-binding protein and the non-conformational region of the target protein is a region involved in ligand binding.
28. The polynucleotide according to claim 27, wherein the region involved in ligand binding comprises a loop.
29. The polynucleotide according to any one of claims 26, 27 or 28, wherein the target protein is an immunoglobulin.
30. The polynucleotide according to claim 29, wherein the sequence cassette is located within a region of the nucleic acid sequence encoding a CDR of the immunoglobulin.
31. The polynucleotide according to claim 30, wherein the CDR is a light chain CDR.
32. The polynucleotide according to claim 30, wherein the CDR is a heavy chain CDR.
33. The polynucleotide according to any one of claims 30, 31 or 32, wherein the CDR is CDR1 or CDR2.
34. The polynucleotide according to any one of claims 30, 31 or 32, wherein the CDR is CDR3.
35. The polynucleotide according to any one of claims 26, 27 or 28, wherein the target protein is an avimer or a fibronectin domain.
36. The polynucleotide according to any one of claims 26 to 35, wherein the sequence cassette further includes flanking sequences adjacent to one or both of the first and second RSSs.
37. The polynucleotide according to any one of claims 26 to 36, wherein the nucleic acid sequence further comprises a second sequence cassette that comprises a third RSS linked by a second heterologous nucleotide sequence to a fourth RSS, the third RSS capable of functional recombination with the fourth RSS.

38. The polynucleotide according to claim 37, wherein the second sequence cassette further includes flanking sequences adjacent to one or both of the third and fourth RSSs.
39. The polynucleotide according to claim 36 or 38, wherein the flanking sequences encode one or more histidine residues.
40. The polynucleotide according to claim 37 or 38, wherein the target protein is a ligand-binding protein and the sequence cassette and the second sequence cassette are each positioned within a portion of the nucleic acid sequence that encodes a region involved in ligand binding.
41. The polynucleotide according to claim 40, wherein the region involved in ligand binding comprises a loop.
42. The polynucleotide according to claim 41, wherein the sequence cassette and the second sequence cassette are both positioned within a portion of the nucleic acid sequence that encodes the same loop.
43. The polynucleotide according to claim 41, wherein the sequence cassette and the second sequence cassette are each positioned within a portion of the nucleic acid sequence that encodes a different loop.
44. The polynucleotide according to any one of claims 26 to 43, wherein the polynucleotide further comprises additional coding sequences that encode a membrane anchor domain peptide.
45. An isolated host cell comprising the polynucleotide of any one of claims 26 to 44.
46. The isolated host cell according to claim 45, wherein the host cell is recombination-competent.
47. A variant protein produced by the method according to any one of claims 1 to 25.

V84629WO

E188 PARTIAL SEQUENCE

gccgccaccatggagtttgggctgagctggctttttctgtggctattttaaaagggtgtccagtgttaccatacgaatgttccagattacgct
tgtccccctcacagtggtagtactccactgtctgggtgtacaaaaacctccctgcacgcctcttaacctcacaattctgtggcgccgc
gccgccaccatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaa
cagacaatcggctgctctgatgccgccgtgtccggctgtcagcgcagggggcgcccggttcttttgcagaccgacctgtccggctg
cctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacggcggttcttgcgcagctgtgtcgcagctgtgtcac
tgaagcgggaagggaactggctgctattggcggaagtgcggggcaggatctcctgtcatctcaccttgcctcgcgagaaagtatcc
atcatggctgatgcaatcggcggtgtcatcgttatccggctacctgcccattcgaccaccaagcgaacatcgcacgcagcga
gcacgtactcggatggaagccggctctgtcgtatcaggtgagtagcaggaggtggagagtagcgtaacacttaagcgtctctccaagt
caaagggaaggagggtttttgtaagggtgtatcactgtgagccagtgtccagtgcggctccggctaccacagtgtacagccctaac
aaaaaccctactgcaacctggcggtgaagagacgtccggaggccagcccttctcatgttcagagaacatgggttaactgggttaagtcatg
tcgtcccacaggatgatctggacgaagagcatcaggggctcgcgcagccgaactgttcgccaggctcaaggcgcgcatgccga
cggcgaggatctcgtctgacccatggcgatgcctgcttgcgaatatcatgggtggaaaatggccgcttttctggattcatcgaactgtgg
ccggctgggtgtggcggaccgctatcaggacatagcgttggctaccctgtatattgtgaagagcttggcggcgaaatgggtgaccg
cttctcgtgtttacgggtatcggctcccgattcgcagcgcacgccttctatgccttcttgacgagttcttctgagtcgactgcagga
gtccactgcacccccctccagcttctctgtccaggcaccaggccaggtatctgggtgtgagccggcctgggtctggcctgagg
ccacaagccccgggggtctgtgtggctggggacaggacggcggtgcctctgctctgtgttgggcatgtgacccattcgagtgtc
ctgcacggggcacagggttttctacaccagacagtggagtactaccactgtgggtactgcacagccagagatgggtgtgacgg
ggagaatgattgcgaggacggcagcgcagggccaattgtccggctctgtgcctaccgagcccaatctgtgacaaaactcacac
atgccaccgtgcccagcacctgaactcctggggggaccgtcagcttcttccccccaaaaccaaggacacctcatgatctcta
gaaccctgaggtcacatgcgtggtggtggacgtgagccacgaagaccctgaggtcaagttcaactggtacgtggacggcgtggag
gtgcataatgccaagacaaagccggggaggagcagtacaacagcacgtaccgtgtggtcagcgtcctcaccgtcctgcaccagga
ctggctgaatggcaaggagtacaagtgaaggtgtccaacaaagccctccagcccccatcgagaaaaccatcctcaaaagccaaag
ggcagccccgagaaccacagggtgtacacctgcccccatccgggatgagctgaccaagaaccagggtcagcctgacctgcctggt
caaaggcttctatcccagcgacatcgccgtggagtgggagagcaatgggcagccgggagaaactacaagaccacgcctccccgtg
ctggactccgacggctccttctctctacagcaagctaccgtggacaagttagatggcagcaggggaacgtcttctcatgctccgt
gatgcatgaggctctgcacaaccactacacgcagaagagccttccctgtctccgggcaactggctctcattgtcctggcgcgctg
gctggcctgtgtgtttattgggtgggcatcttcttctgtccgggtgtcggcataggaggcgccaaggaggtggcggtatctggagg
gggaggatctggagggggctcaggatcagggggaggatctggaggcggtatca

[SEQ ID NO:28]

FIG. 1

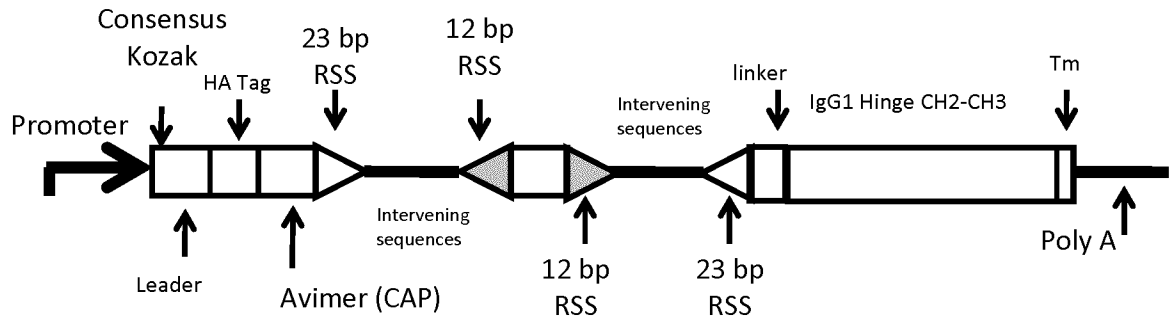
E189 PARTIAL SEQUENCE

gccgccaccatggagtttgggctgagctggcttttctgtggctattttaaagggtgtccagtgttaccatacgaatgtccagattacgct
tgcttgccccacagtggtagtactccactgtctgggtgtacaaaaacctccctgcacgcctcttaacctacaaattctgtggcgccgc
gccgccaccatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaa
cagacaatcggctgctctgatgccgccgtgttccggctgtcagcgcagggggcgcccggttcttttgaagaccgacctgtccggctg
cctgaatgaactgcaggacgaggcagcgcggctatcgtggctggccacgacggcggttcttgcgcagctgtgtcgcagctgtgtcac
tgaagcgggaagggaactggctgctattggcggaagtgcgggggcaggatctcctgtcatctcaccttgcctgccgagaaagtatcc
atcatggctgatgcaatcggcggtgtcatacgttgatccggctacctgcccattcgaccaccaagcgaaacatcgcacgcagcga
gcacgtactcggatggaagccggtcttctgatcaggtgagtacaggaggtggagagtacgcgtaacacttaagcgtctctccaagt
caaagggacaggaggtttttgtaagggtgtatcactgtggaccagttcagatgcggcaacggccagtgcatccccctggattgggt
gtgcgcagggcgtgaacgactgccccgattccgatgaggaaggctgccccctagaacctgtgcccclagccagcacagtatacag
cccttaacaaaaaacccctactgcaacctggcggttaagagacgtccggaggccagcccttctcatgttcagagaacatggttaactggt
aagtcattgtctcccacaggatgatctggacgaagagcatcaggggctcgccgagccgaactgttcgccaggctcaaggcgcgca
tgcccgcagggcgaggatctcgtcgtgacctatggcgatcctgcttccgaatatcatggtggaaaatggccgcttttctggattcatc
gactgtggccggctgggtgtggcgaccgctatcaggacatagcgttggctaccggtgataattgctgaagagcttggcgcgcaatgg
gctgaccgcttctcgtctttacggtatcgccgtcccgattcgcagcgcacgtccttctatgccttcttgacgagttcttctgagtcga
ctgcaggagtccactgcacccccctccagcttctctgtccaggcaccaggccaggtatctggggtgtgcagccggcctgggtctg
gcctgaggccacaagccccgggggtctgtgtggctggggacaggagacggcgctgcctctgctctgtgcttgggccatgtgacctatt
cgagtgtcctgcacgggcacagggttttgtacaccagacagtggagtactaccactgtgttccagtgcggctccggctactgcatcag
ccagagatgggtgtgcgacggggagaatgattgcgaggacggcagcgacgaggccaattgtccggctctgtgcctaccgagccc
aaatcttgtgacaaaactcacacatgccaccgtgccagcacctgaactcctggggggaccgtcagttcttcttcccccaaac
caaggacacctcatgatcttagaaccttgaggtcacatgctgtgtgtggacgtgagccacgaagacctgaggtcaagtcaa
ctgtgtacgtggacggcgtggaggtgcataatgccaagacaaagccgaggagagcagtagaacagcacgtaccgtgtgtgtcagc
gtcctcaccgtctcgcaccaggactggctgaatggcaaggagtagaagtgaagggtgtccaacaagccctcccagccccatcga
gaaaaccatctcaaagccaaagggcagccccgagaaccacaggtgtacacctgcccccatcccggtatgagctgaccaagaac
caggtcagcctgacctgcctggtcaaaggcttctatcccagcgacatcgccgtggagtgaggagagcaatgggcagccggagaacaa
ctacaagaccacgcctcccgtgtgactccgacggctccttcttctctacagcaagctcaccgtggacaagtctagatggcagcag
gggaacgtcttctcatgtccgtgatgcatgaggtctgcacaaccactacacgcagaagagccttccctgtctccgggcaactgg
ctctcattgtctggcggtgtggctgtgctgtttatgggtgggcacattcttctgttccgggtgtcgccataggaggcgcca
aggaggtggcggtatctggagggggaggatctggagggggctcaggatcagggggaggatctggaggcggtatca

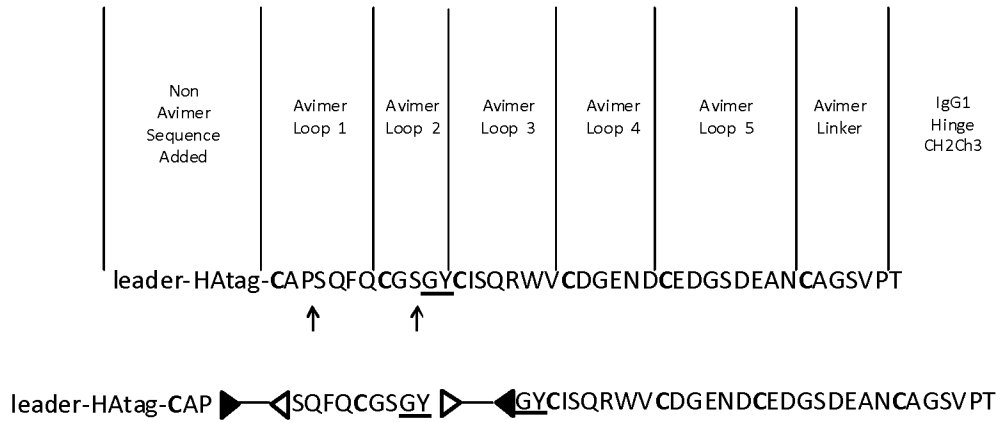
[SEQ ID NO:29]

FIG. 2

A.



B.



- Bold residues indicate conserved Cystiene Amino Acids
- Underlined Amino Acids indicate flanking residues which were duplicated in the mutator construct
- Arrows indicate location of inserted RSS cassette

FIG. 3

C.

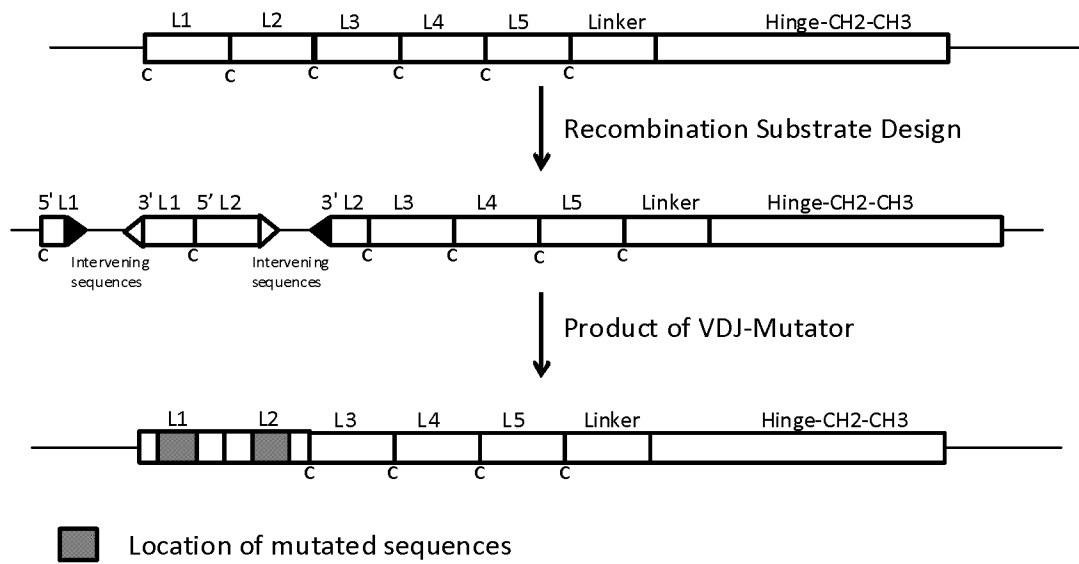


FIG. 3 (con.)

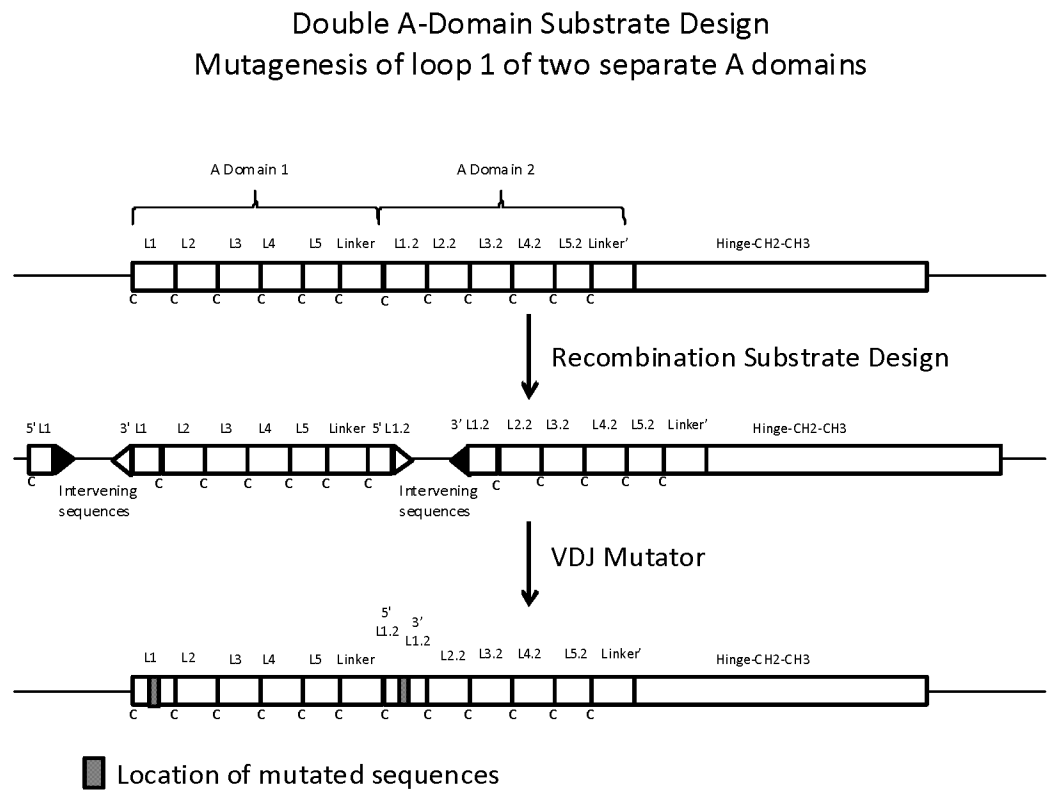


FIG. 4

Single A domain, double A domain, triple A domain fusion protein

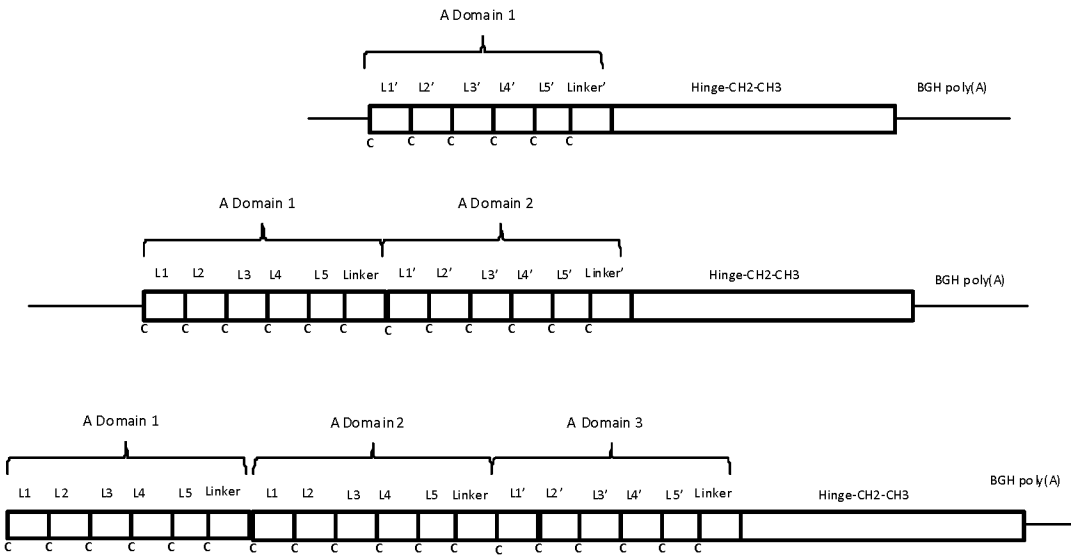


FIG. 5

A.

	+1	S	Q	F	Q	C	G	S	G	Y
1		AGCCAGTTCC	AGTGCGGCTC	CGGCTAC						
		TCGGTCAAGG	TCACGCCGAG	GCCGATG						

B.

+3	Q	P	V	C	V	R	L	R	L	L	
+2	T	A	S	L	C	A	A	P	A	T	
+1	Y	S	Q	F	V	C	G	S	G	Y	Y
1	TACAGCCAGT	TTGTGTGCGG	CTCCGGCTAC	TAC							
	ATGTCGGTCA	AACACACGCC	GAGGCCGATG	ATG							

FIG. 6

V84629WO

A.

tccgatgtgccaggacgtggaagtgggtggccgcccacctaccagcctgctgatctcttgggatgccctgccgtgaccgtgcgg
 tactacagaatcacctacggcgagacaggcggcaacagccccgtgcaggagtttacagtgcccggcagcaagagcaccgccacca
 tctctggactgaagccccggcgtggactacaccatcaccgtgtacgccgtgacaggcagaggcgacagccctgccagcagcaagcc
 catcagcatcaactaccggacc

[SEQ ID NO:34]

B.

SDVPRDLEVVAATPTSLLISWDAPAVTVRYRITYGETGGNSPVQEFTVPGSKSTATI
 SGLKPGVDYTITVYAVTGRGDSPASSKPISINYRT

[SEQ ID NO:39]

C.

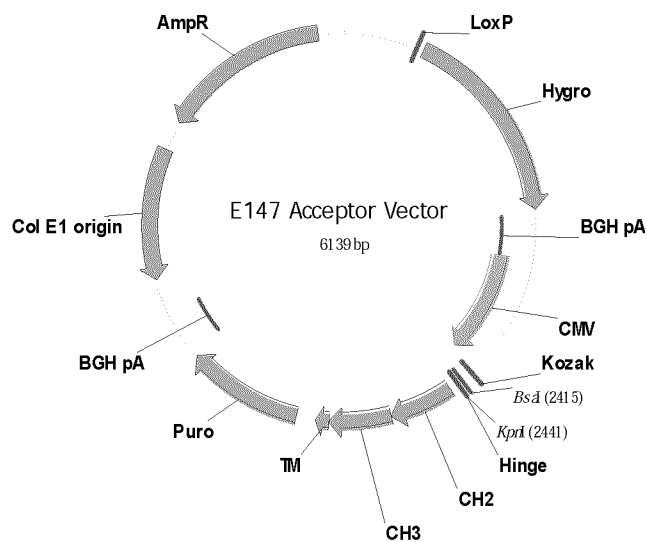


FIG. 7

V84629WO

D.

ctaaattgtaagcgftaataatgtttaaattcggttaaattttgttaaatcagctcatttttaaccaataggccgaaatcggaataatccc
 ttataaatcaaaagaatagaccgagatagggttgagtggccgctacagggcgctccattcgccattcaggctgcgcaactgttggga
 agggcggttccggtgcgggcctcttcgctattacgccagctggcgaaagggggatgtgctgcaaggcgattaagtgggtaacgccag
 gggtttccagtcacgacgttgaataacgacggccagtgagcgcgacgtaatacgaactactatagggcgaaattggcggaaggccgt
 caaggcctaggcgccctgaataacttcgtatagcatacattatagcaattatcgaaaaagcctgaactcaccgcgacatccgtggag
 aaattcctcatcgaaaaattcgactccgtgtccgatctatgcagctgtccgaggcgaggagagtagagcattctcattcgatgtggg
 cgggagaggctacgtgctgagagtgaactctgtgcccagggcttctacaaggaccgatacgtctaccggcattttgcttccgcccgtc
 tgcctattccagaagtccctggacattggggagttagcgagtcctcacttactgtattagccggcgagcccaggaggtagactccag
 gatctgctgaaactgaactgctgctgtgctccagcctgtcgctgaggcaatggatgctattgctgctgcccagctgagtcagactagc
 ggattcggcccatttggaccccaggcgattggccagtcacacaacatggcgagacttcatctgtgctatcgccgacctcacgtgtacca
 ttggcagactgtgatggacgatactgtgtctgcttctgtggcacaggcactcgacgaactcatgctgtgggctgaggactgtcctgaagt
 gagacatctggtccatgccgatttggctccaacaatgtgtctaccgataacgggagaatcactgcccgtgatcgactggagcgaggca
 atgtttggcgattccagtcgaagtggccaacatcttcttggcgcccttggctggtgtgatggaacagcagaccgggtactttgaac
 ggcgccaccctgagctggctgggagtcctagactgagagcctacatgctccgaattggcctggatcagctctaccagtcactgttggga
 tggcaatttcgacgatgctgcttgggacagggcgctgtgatgtattgtccgatccggcgctggaactgtggggagaacacagatc
 gctaggagatccgctgctgctggaccgatggatgtgtggaagtgtggccgatagtggaaaccggaggccttcaaccgaccccg
 ggcaaggagtaatgaccgtttaaaccgctgatcagcctcgactgtgcttctagtggcagccatctgttggccccctccccgtgc
 ctctctgaccctggaagggtgcccactcccactgtccttcttaataaaatgaggaaattgcatcgattgtctgagtaggtgtcattctattct
 ggggggtgggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgtggggatgctgggtgggtctatgg
 ggatcccgcgtgacattgatttactagttatataagtaataatcaggggctcattagttcatagcccatataggagtccgcgttac
 ataacttacggtaaatggcccgcctggctgaccgcccacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgc
 caatagggactttccattgacgtcaatgggtggagtattacggtaaatgcccacttggcagtcacatcaagtgtatcatatgccaagtac
 gccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtcacatgacctatgggacttctacttggcagtcacatc
 tactattagtcacgtattaccatgggtatgctgggtttggcagtcacatcaatggcgctggatagcggttgactcacggggatttccaa
 gtctccacccattgacgtcaatgggagttgtttggcaccaaaatcaacgggactttccaaaatgtcgtaaactccgccccattgac
 gcaaatgggcggtaggcggtgtacgggtgggaggtctatataagcagagctcttggcctaactagagaaccactgcttactgtcgcag
 atctgatcaagagacaggataaggagccgccaccatggagtgtggctgagctggcttttctgtggctattttaaagggtgtccaggt
agagaccggaagagatt**ggfacc**gagcccaaatctgtgacaaaactcacacatgcccaccgtgcccagcacctgaactcctgggg
 ggaccgtcagcttctcttcccccaaaacccaaggacacctcatgatcttagaaccctgaggtcacatgctgtgtgtgtggacgt
 gagccacgaagaccctgagggtcaagtcaactgggtacgtggacggcggtggagggtgcataatgccaagacaaagccgaggga
 gcagtacaacagcacgtaccgtgtgtcagcgtcctcaccgtcctgcaccaggactggctgaatggcaaggagtacaagtcaagggt
 gtccaacaaagccctcccagccccatcgagaaaaccatctccaaagccaaaggcgagccccgagaaccacaggtgtacaccctg
 cccccatcccgggatgagctgaccaagaaccaggtcagcctgacctgctgtaaggcttctatcccagcgacatcgccgtggag
 tgggagagcaatgggcagccggagaacaactacaagaccacgctcccgtgctggactccgacggctccttctctctacagcaag
 ctaccgtggacaagtctagatggcagcaggggaaacgtcttctcatgctccgtgatgcatgaggctctgcacaaccactacacgcaga
 agagcctctccctgtctccgggcaactggctctcattgtcctggcgggcggtggctgctgctgttattgggctgggcactctctt
 tgtgtccggtgtcggcataggaggcgccaaggaggtggcggtatctggaggggaggatctggagggggctcaggtacaggggga
 ggatctggaggcggtatcaactgagtacaaccactgtgaggctcgtactagagatgatgtgcctagagctgtccgaactctggctg
 ctgcttccgcatattaccctgccactcgccataccgtgatcccgatgccacattgaacgagtcaccgaactccaggagctgttctca
 cttagagtcgggctggtatattggcaagtctgggtggccgatgacggagccgctgtcgtgtgtggactacacctgagctgtgtggagg
 ctggcgccgtgttctgaaattggacctggatggctgaactgtctggatctcgactggctgcccagcagcagatggaggggactgct
 ggacccccatagaccaaggaacctgcctggttctggcaactgtgggaggtgcaccgatcatcaggggcaaggactgggatctgc
 cgtgtgtcctccgtggcggtggaggccgtgaacgagctggcgctccccgtttctgaaactctgcccccgaaactccttctacga
 acgactgggattcactgtcaccggcatgtcgaagtgcctgaggggcttagaacaatggtgtatgaccggaaacccggagcttaacc

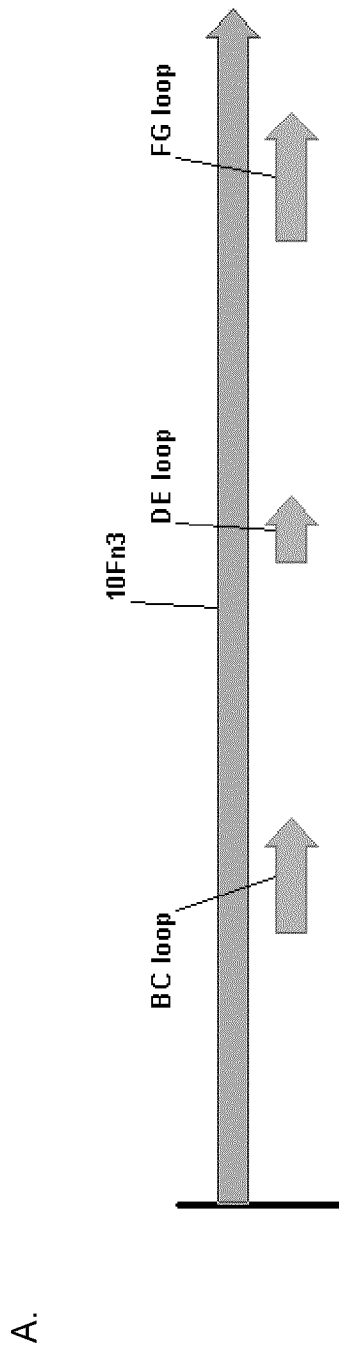
FIG. 7 (con.)

V84629WO

gtttaaacccgctgatcagccctgactgtgcccttctagttgccagccatctgttgtttgcccctccccctgccttcttgaccctggaagg
tgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcatgtctgagtaggtgtcatttctattctggggggtgggggtgggg
caggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggctcgagtaattaactggc
ctcatgggcttccgctcactgcccgtttccagtcgggaaacctgtcgtgccagctgcattaacatggcatagctgttcccttgcgtatt
ggggcgtctccgcttctcgtcactgactcgtcgtcgtcgttcgggttaaagcctgggggtgcctaatagcaaaaggccagcaa
aaggccaggaaccgtaaaaaggccgctgtgctggcgttttccataggctccgccccctgacgagcatcacaaaaatcgacgtca
agtcagaggtggcgaaaccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgtctcctgttccgacc
tgccgttaccggatacctgtccgctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggttag
gtcgttcgtccaagctgggctgtgtgcacgaacccccgttcagcccagccgtgcgcttatccggtaactatcgtcttgagtcaa
cccgttaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgtctacagagtt
cttgaagtgggtggcctaactacggctacactagaagaacagtatlttggtatctgcgctctgctgaagccagttacctcggaaaaagagt
tggttagctcttgatccggcaaaaccacccgctggtagcggtgtgtttttgttgaagcagcagattacgcgcagaaaaaaggat
ctcaagaagatccttgatctttctacggggtctgacgctcagtggaacgaaaactcacgttaagggaatttgggtcatgagattatcaaaa
aggatcttcacctagatccttttaataaaaaatgaagtttaaatcaatctaaagtatatatgagtaaaacttggctgacagttaccaatgctt
aatcagtgaggcacctatctcagcgatctgtctatlttcgttcacatagttgctgactccccgtcgtgtagataactacgatacgggag
ggcttaccatctggccccagtgctgcaatgataccgcgagaaccacgctcaccggctccagatttatcagcaataaaccagccagcc
ggaagggccgagcgcagaagtggctcctgcaactttatccgcctccatccagctctattaattgttggcggaagctagagtaagtagttc
gccagttaatagtttgcgcaacgttgttgcattgtctacaggcatcgtgggtgtcacgctcgtcgttgggtatggcttcattcagctccggttc
ccaacgatcaaggcgagttacatgatccccatgttgtcaaaaaagcgggttagctccttcggctcctccgatcgttgtcagaagtaagtt
ggccgcagtggtatcactcatggttatggcagcactgcataattcttactgtcatgccatccgtaagatgcttttctgtactgggtgagta
ctcaaccaagtcattctgagaatagtgatgcggcgaccgagttgctcttggccggcgtaatacgggataataccgcgccacatagca
gaactttaaagtgtcatcattggaaaacgttctcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacc
cactcgtgcaccaactgatcttcagcatcttttacttaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaa
gggaataaggcgacacggaaatgttgaatactcatacttctcttttcaatattatgaagcatttatcagggttattgtctcatgagcgg
atacataattgaatgtatttagaaaaataaacaataagggggtccgcgcacatttccccgaaaagtgccac

[SEQ ID NO:35]

FIG. 7 (con.)

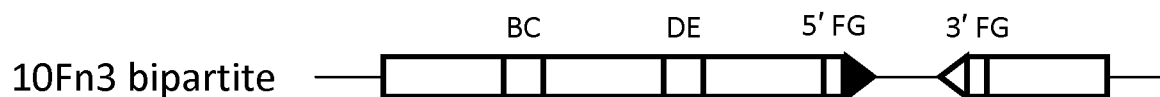


മ്

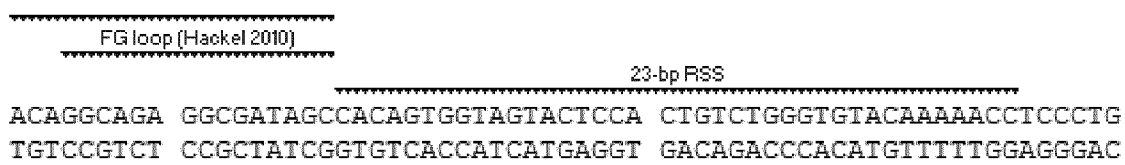
	10Fn3	BC loop	FG loop
1	S D V P R D L E V V A A T P T S L L I S V D A P A V T W R Y Y R I T Y G E T G G	TCCGATGTGCCAGG GACCTGGAGTGGTG GCCGCCACACCTACC AGCCTGCTGATCTCT TGGGATGCCCTGCC GTGACCGTGGGTAC TACGAATCACCTAC GCGGACACAGGCGGC	AGGCTACACGGGTCC CTGGACCTTCACCA CCGCGGTGTGGATGG TCGGACGACTAGACA ACCCTACGGGACGG CACTGGACGCCATG ATGCTTAGTGGATG CCGCTCTGTCCGCGG
121	N S P V Q E F T V P G S K S T A T I S G L K P G V D Y T I T V Y A V T G R G D S	DE loop	FG loop
241	P A S S K P I S I N Y R T	CCTGCCAGCAGCAG CCCATCAGCATCAAC TACCGGAAC	TTGTCCGGGCACGTC CTCAAATGTCACGGG CCGTCGTTCTCGTGG CCGTGTGTAGAGACCT GACTTCGGGCGCAC CTGATGTGTAGTGG CACATGGGCACTGT CCGTCTCCGCTGTGG

FIG. 8

A.



B.



C.

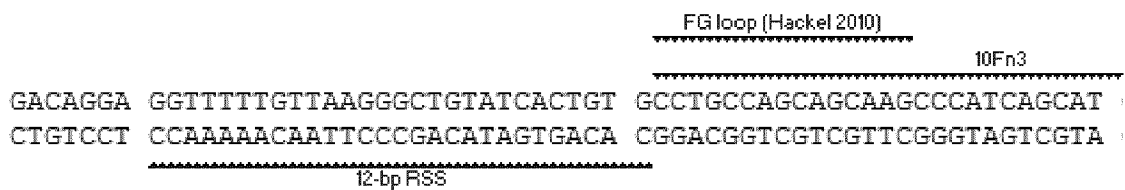
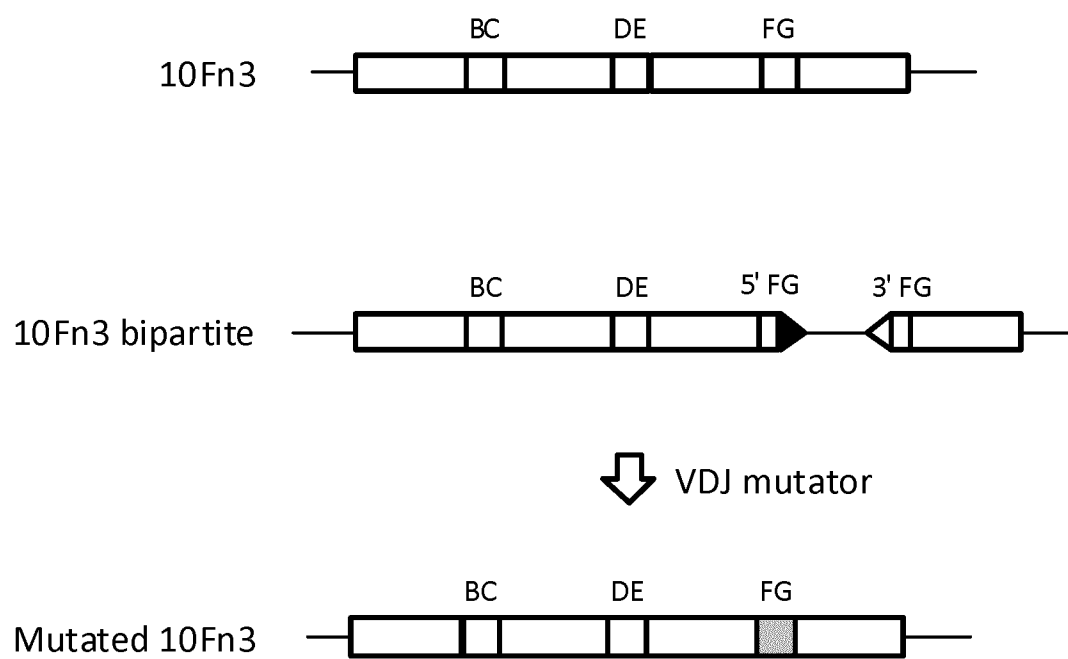


FIG. 9

V84629WO

D.

**FIG. 9 (con.)**

V84629WO

A.

tccgatgtgccaggacctggaagtgggtggccgcccacacctaccagcctgctgatctcttgggatgccctgccgtgaccgtgcgg
tactacagaatcacctacggcgagacaggcggcaacagccccgtgcaggagtttacagtgcccgagcaagagcaccgccacca
tctctggactgaagcccggtggactacaccatcacctgtacgccgtgacaggcagaggcgatag**ccacagtggtagtactcca**
ctgtctgggtgtacaaaaacctccctgcacgcctcttaacctcacaattctgtggcggccgctttagccagaccctcggtaactg
gatgtcacaactggcacctgagattggaacataaaaaacaatatcttactattaatcatgttatcagagaacttccctgaagtccagtc
agtactgtgactagctaattagtcagttacttaagcgtctatccaagtgcaggggacagagg**ggttttgttaagggtgtatcactgtg**
cctgccagcagcaagcccatcagcatcaactaccggacc

[SEQ ID NO:37]

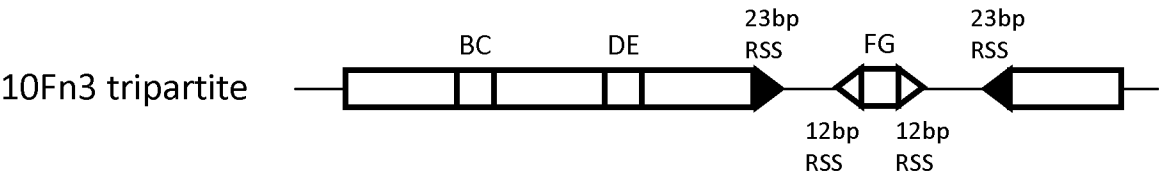
B.

tccgatgtgccaggacctggaagtgggtggccgcccacacctaccagcctgctgatctcttgggatgccctgccgtgaccgtgcgg
tactacagaatcacctacggcgagacaggcggcaacagccccgtgcaggagtttacagtgcccgagcaagagcaccgccacca
tctctggactgaagcccggtggactacaccatcacctgtacgccgtgacagg**ccacagtggtagtactccactgtctgggtgt**
acaaaaacctccctgcacgcctcttaacctcacaattctgtggcggccgctttagccagaccctcggtaactggatgtcacaact
ggcacctgagattggaacataaaaaacaatatcttactattaatcatgttatcagagaacttccctgaagtccagtcagtactgtgact
agctaattagtcagttacttaagcgtctatccaagtgcaggggacagagg**ggttttgttaagggtgtatcactgtgggcagaggcg**
acagccctgccagcagcaag**cacagtgatacagcccttaaaaaacct**actgcaacctggcggtaatagacgtccggaagc
acactggctgagtaattcctagtgttctccatccttacctcggagccagagtagcaggagccactagccagcttgggtcttctatcgc
gagtcgtattaattcgataagccagcaagcagtggttctctagttagccagctgcctccttctctgggccagcgtcctctgtcctgg
agctgggagataatgtccgggggtccttggctcgcgtgggcaaagggtgggcagagtcagcttgtgctggggacaaaatgacct
tgggacacggtcgacgggtggctgccacggccggccggggacagtcggagagtcag**ggttttgtacaccagacagtggagta**
ctaccactgtgaagcccatcagcatcaactaccggacc

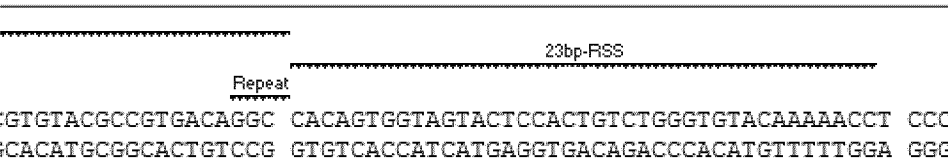
[SEQ ID NO:38]

FIG. 10

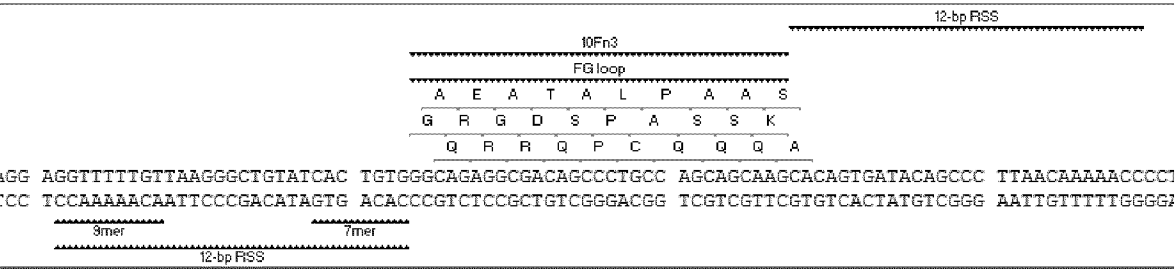
A.



B.



C.



D.

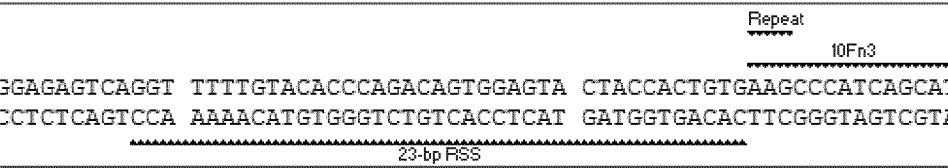
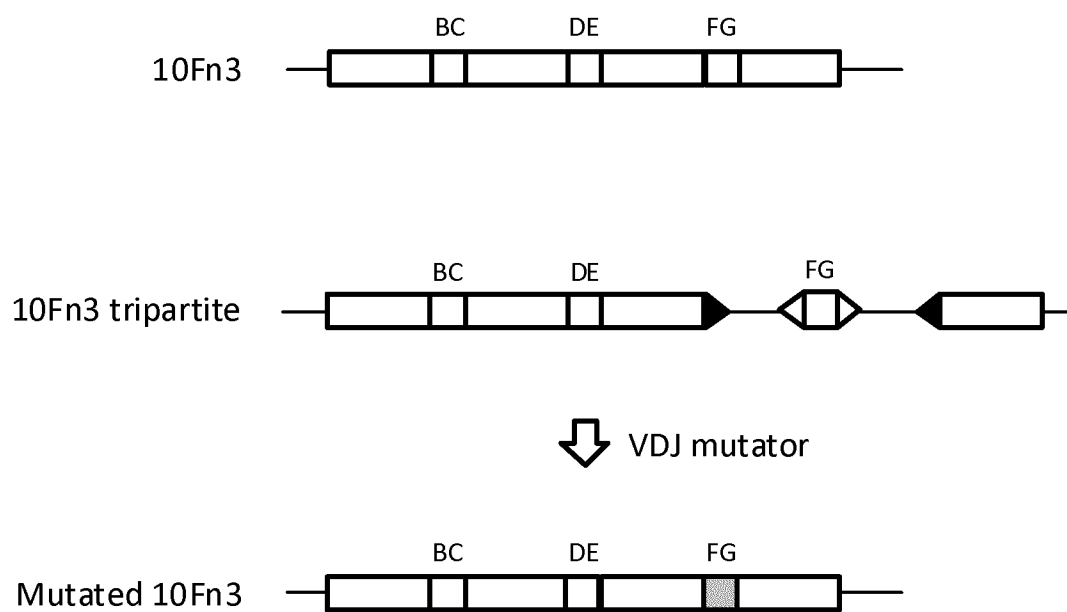


FIG. 11

V84629WO

E.

**FIG. 11 (con.)**

V84629WO

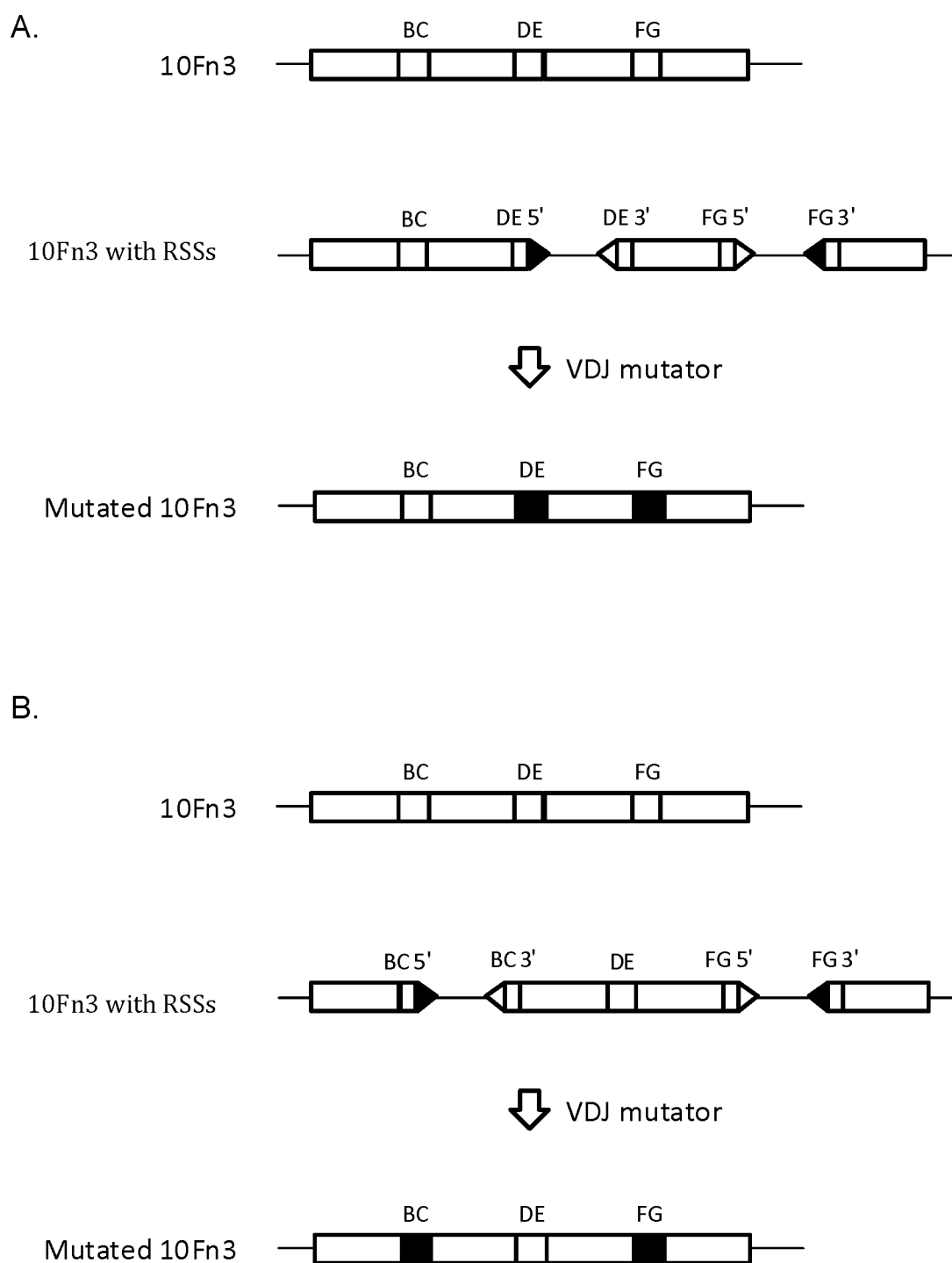
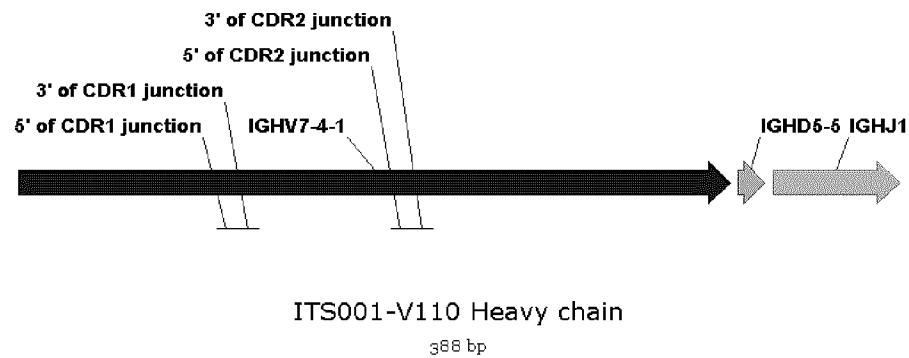


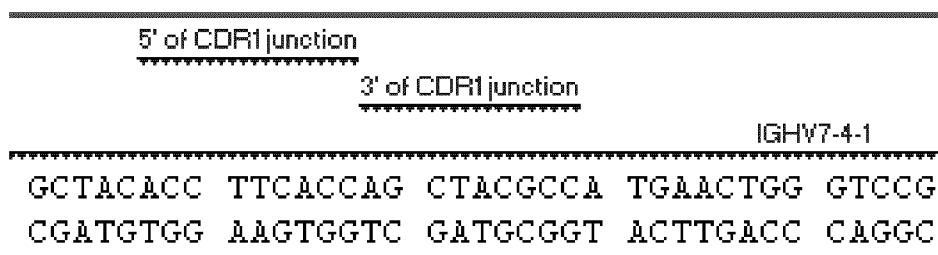
FIG. 12

V84629WO

A.



B.



C.

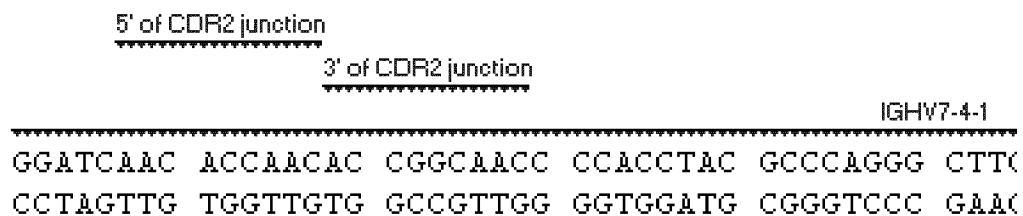
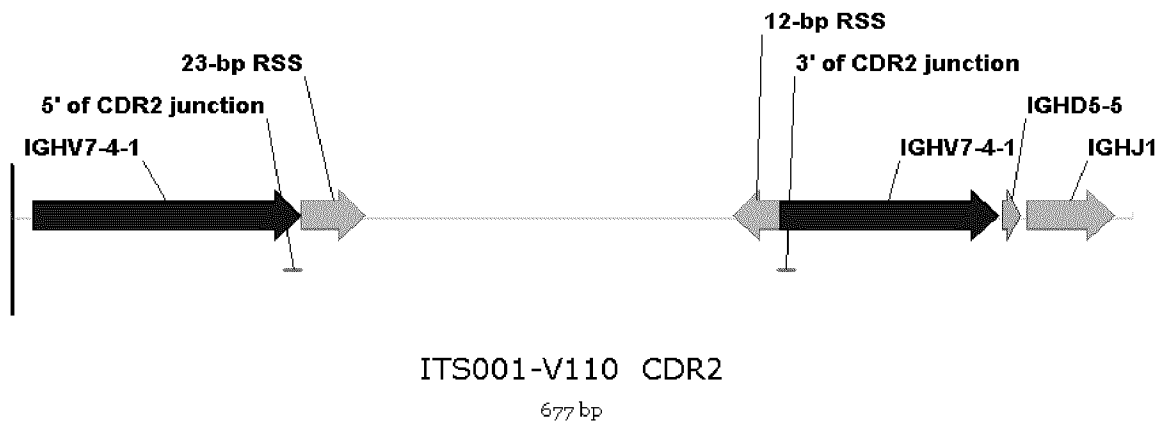


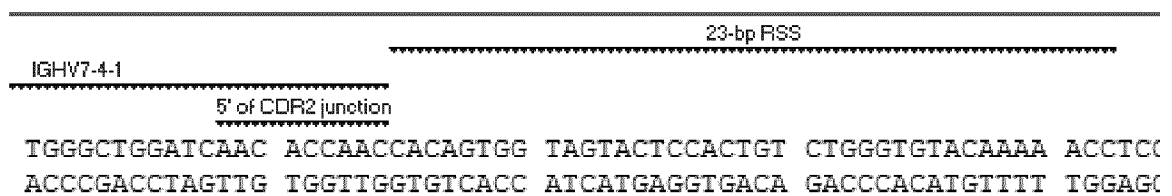
FIG. 13

V84629WO

A.



B.



C.

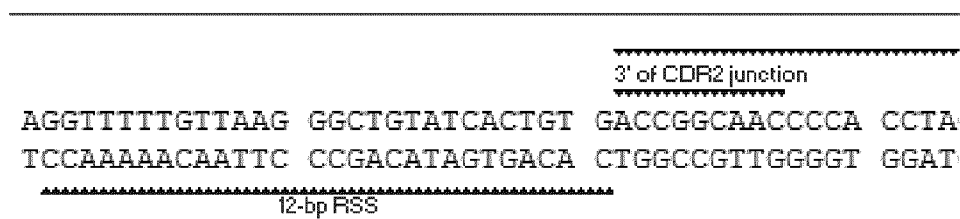
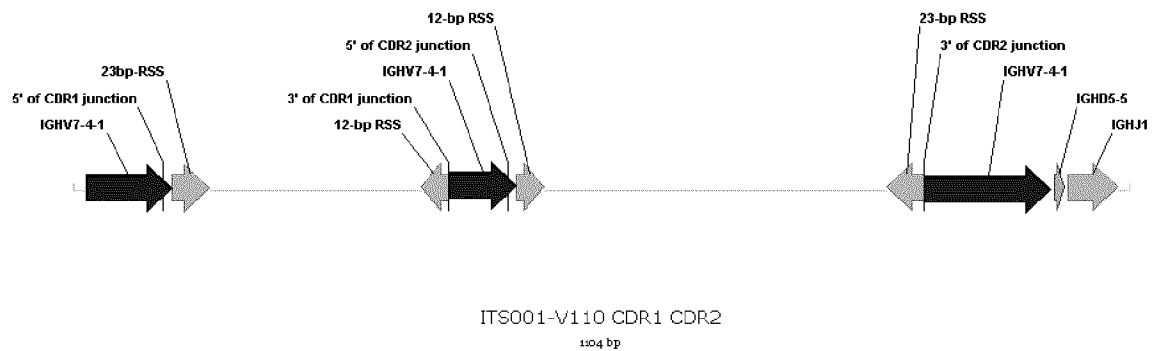


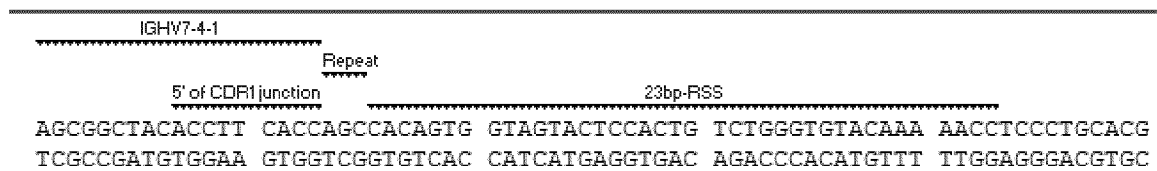
FIG. 14

V84629WO

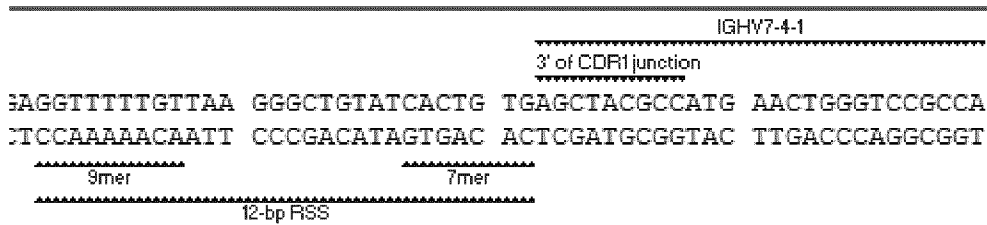
A.



B.



C.



D.

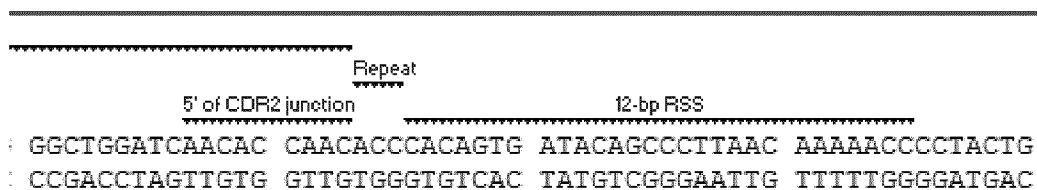


FIG. 15

V84629WO

E.

IGHV7-4-1

3' of CDR2 junction

TCAGGTTTTGTAC ACCCAGACAGTGGA GTACTACCACTGTG ACCGGCAACCCAC CTACGCCCAGGGCT
AGTCCAAAAACATG TGGGTCTGTCACCT CATGATGGTGACAC TGGCCGTTGGGGTG GATGCGGGTCCCGA

23-bp RSS

FIG. 15 (con.)

V84629WO

A.

caggtgcagctggtgcagagcggcagcgagctgaagaaacctggcgccctccgtgaaggtgtcctgcaaggccagcggctacacct
tcaccagctacgccaatgaactgggtccgcccaggccccaggccagggaactggaatggatgggctggatcaacaccaaccacggc
aacccccactacgcccagggttcacgggcagattcgtgttcagcttcgacaccagcgtgtccaccgcctacctgcagatctgtagc
ctgaaggccgaggacaccgccgtgtattactgtgcgagaggacagctatggtacgggctgaatactccagcactggggccaggg
cacctgtgcaccgtgtcctcag

[SEQ ID NO:3]

B.

caggtgcagctggtgcagagcggcagcgagctgaagaaacctggcgccctccgtgaaggtgtcctgcaaggccagcggctacacct
tcaccagctacgccaatgaactgggtccgcccaggccccaggccagggaactggaatggatgggctggatcaacaccaaccacagtg
tagtactccactgtctgggtgtacaaaaacctccctgcacgcctcttaacctcacaattctgtggcgccgctttgtagccagacct
cgggtcaactggatgtcacaactggcacctgagattggaacataaaaaacaaatattcttactattaatcatgtttatcagagaactccctga
agttccagtcagtagtgcactagctaattagtcagttacttaagcgtctatccaagtgcaggggacaggaggtttttgtaagggtg
tactactgtgaccggcaacccccactacgcccagggttcacgggcagattcgtgttcagcttcgacaccagcgtgtccaccgccta
cctgcaaatctgtagcctgaaggccgaggacaccgccgtgtattactgtgcgagaggacagctatggtccgggctgaatactccagc
cactggggccagggcacccctggtcaccgtgtcctcag

[SEQ ID NO:4]

C.

caggtgcagctggtgcagagcggcagcgagctgaagaaacctggcgccctccgtgaaggtgtcctgcaaggccagcggctacacct
tcaccagccacagtggttagtactccactgtctgggtgtacaaaaacctccctgcacgcctcttaacctcacaattctgtggcgcc
gctttgtagccagaccctcgggtcaactggatgtcacaactggcacctgagattggaacataaaaaacaaatattcttactattaatcatgtt
atcagagaactccctgaagttccagtcagtagtgcactagctaattagtcagttacttaagcgtctatccaagtgcaggggacagga
ggttttgttaagggtgtatcactgtgagctacgccatgaactgggtccgcccaggccccaggccagggaactgaatggatgggct
ggatcaacaccaacaccacagtgatacagcccttaacaaaaacctactgcaacctggcggtatagacgtccggaagcaca
ctggctgagtaaatcttagtgttctccatccttacctcgagccagagtagcaggagccactagccagctgggtcttctatcgag
tcgtattaatttcgataagccagcaagcagtggtgtcttagtagccagctgcctccttctctggggccagcgtcctctgtcctggagct
gggagataatgtccgggggtccttgggtctgcgtgggcaagggtgggcagagtcattgtgtctggggacaaaatgacctggg
acacggctgacgggtgggtgccacggccggccgggacagtcggagagtcagggtttttgtacaccagacagtgaggtagtactac
cactgtgaccggcaacccccactacgcccagggttcacgggcagattcgtgttcagcttcgacaccagcgtgtccaccgcctacct
gcaaatctgtagcctgaaggccgaggacaccgccgtgtattactgtgcgagaggacagctatggtccgggctgaatactccagca
ctggggccagggcacccctggtcaccgtgtcctcag

[SEQ ID NO:5]

FIG. 16

V84629WO

ctaaatgtaagcgtaaatatgtttaaattcggttaaattttgttaaatacagctcatttttaaccaataggccgaaatcggcaaaatccc
 ttataaatcaaaagaatagaccgagatagggtgagtgccgctacagggcgctccattcgccattcaggctgcgcaactgttggga
 agggcggttcggtgcgggaccttctgctattacgccagctggcgaaagggggatgtgctgcaaggcgattaagtgggtaacgccag
 gggtttccagtcacgacgttgtaaacgacggccagtgagcgcgacgtaatacgaactactataggcgaaattggcggaaggccgt
 caaggcctaggcgcgccgtgaataacttcgtatagcatacattatagcaatttatcgaaaaagcctgaactaccgcgacatccgtggag
 aaattctcatcgaaaaatcgcactccgtgtccgatctcatgcagctgtccgagggcgaggagagtagagcattctcattcgatgtggg
 cgggagagggtacgtgctgagagtgaactctgtgccgacggcttctacaaggaccgatacgtctaccggcattttgcttccgcccgtc
 tgccattccagaagtctggacattggggagtttagcgagtcctcacttactgtattagccggcgagcccaggggagtgacactccag
 gatctgctgaaactgaactgctgtgtctccagcctgtcgtgaggaatggatgtattgtgctgctgcccagctgagtcagactagc
 ggattcggccatttggaccccaggcgattggccagtcacacaactggcgagacttcatctgtgtatcgccgacccacgtgtacca
 ttggcagactgtgatggacgatactgtgtctgtctgtggcacaggcactcgacgaactcatgctgtgggctgaggactgtcctgaagt
 gagacatctggtccatgccgattttggctccaacaatgtgtctaccgataacggggagaatcactgccgtgatcgactggagcgaggca
 atgtttggcgattcccagtcgaagtggccaacatcttctttggcgcccttggctgctgtatggaacagcagaccgggtactttgaac
 ggcgccaccctgagctggctgggagtcctagactgagagcctacatgctccgaattggcctggatcagctctaccagtcactgttggga
 tggcaatttcgacgatgtgcttgggacagggggcgctgtgatgtattgtccgatccggcgctggaactgtggggagaacacagatc
 gctaggagatccgctgtgtctggaccgatggatgtgtggaagtgtggtggcgatagtggaaccggaggccctcaaccgaccccg
 ggcaagggagtaatacgctttaaaccgctgatcagcctcgactgtgcttctagtgtccagccatctgtgttggccctccccgtgc
 ctctctgaccctgggaagggtgccactcccactgtcttcttaataaaatgaggaaattgcatcgcatgtgtctgagtaggtgtcattctattct
 ggggggtgggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgtggggatgctgggtgggtctatgg
 ggatcccgctgtgacattgatttactagttatataagtaatacattacgggggtcattagttcatagcccatataggagttccgcgttac
 ataacttacggtaaatggcccgctggctgaccgccaacgaccccgccattgacgtcaataatgacgtatgttccatagtaacgc
 caatagggactttccattgacgtcaatgggtggagtatttacggtaaactgcccacttggcagtcacaaagtgtatcatatgccaagtac
 gccccattgacgtcaatgacggtaaatggcccgctggcattatgccagtcacatgacctatgggacttccacttggcagtcacac
 tacgtattagtcacgtattaccatgggtatgctggttttggcagtcacatgaatggcggtggatagcgggttgactcacggggatttccaa
 gtctccaccccatgacgtcaatgggagttgttttggcacaaaatcaacgggactttccaaaatgtcgtaaacactccgccccattgac
 gcaaatggcggttaggctgtacggtgggaggtctatataagcagagctctctggcctaactagagaaccactgcttactgtctgcag
 atctgatcaagagacaggataaggagccgccaccatggagtgggtgagctggttttctgtggttattttaaaggtgtccagtg
 taccatacagatgtccagattacgttgtgcccctcacagtggtagtactccactgtctgggtgtacaaaaacctccctgcacgcctctc
 taacctcacaattctgtggcgccgcccgcaccatgattgaacaagatggattgcacgcagggttctccggcggcttgggtggagag
 gctattcggctatgactgggcacacagacaatcggctgctgtatgcccggctgttccggctgtcagcgcaggggcgcccggttctt
 ttgtcaagaccgacctgtccggtgccctgaatgaactgcaggacgaggcagcgccgctatcgtggctggccacgacggcggttctt
 ggcagctgtgtcgcaggtgtcactgaagcgggaagggactggctgtattggcggaagtgcggggcaggatctcctgtcatctc
 acctgtcctgccgagaaaagtatccatcatggctgtatgcaatgcggcgctgcatacgttgatccggctacctgcccattcgaccac
 caagcgaacatcgcatcgagcagcagctactcggatggaagccggtctgtcgtatcaggtgagtcagggaggtggagagtacgc
 gtaacacttaagcgtctctcaagtcaaaaggacaggaggttttgaagggtgtatcactgtgagccagttccagtgccggtccg
 gctaccacagtgtacagcccttaacaaaaacccctactgcaacctggcggttaagagacgtccggaggccagcccttctcatgttcag
 agaacatggttaactggttaagtcatgtctgccacaggatgatctggacgaagagcatcaggggctcgcgcccagccgaactgttcgc
 caggctcaaggcgcatgcccacggcgaggatctctcgtgacctatggcgatgcctgcttgcgaatatcatggtggaaaatgg
 ccgcttttctgattcatcagactgtggccggctgggtgtggcgaccgctatcaggacatagcgttggctaccggtgatattgctgaaga
 gcttggcgccgaatgggctgaccgcttctctgtctttacggtatcggcgtcccgattcgcagcgcacgcttctatcgccttctgac
 gagtcttctgagtcgactgcaggagtccactgcacccccctccagcttctctgtccaggcaccaggccaggtatctggggtgtgc
 agccggcctgggtgtggcctgagccacaagcccggggtctgtgtggctggggacaggacgcccgtgctctgtctgtgtct
 gggccatgtgaccttcgagtgctcctgcacgggcacagggttttgtacaccagacagtggagtactaccactgtgggctactgcac
 agccagagatgggtgtgcgacggggagaatgattgcaggacggcagcgacgaggccaattgtgccggctctgtgcttaccgagc
 ccaactctgtgacaaaactcacacatgccaccgtgccagcacctgaactcctggggggaccgtcagttctcttccccccaaaa

FIG. 17

V84629WO

cccaaggacaccctcatgatctctagaacccctgaggtcacatgcgtgggtggacgtgagccacgaagaccctgaggtcaagttc
 aactggtagctggacggcggtggaggtgcataatgccaagacaaagccgaggaggagcagtacaacagcacgtaccgtgtgtca
 ggcgtcctcaccgtcctgcaccaggactggctgaatggcaaggagtacaagtgcagggtgtccaacaagccctccagccccatc
 gagaaaaccatctccaaagccaaaggcagccccgagaaccacaggtgtacaccctgccccatccgggatgagctgaccaaga
 accaggtcagcctgacctgctgggtcaaaggcttctatccagcgacatcgccgtggagtgaggagcaatgggcagccggagaac
 aactacaagaccacgcctcccgctgctggactccgacggctccttctctacagcaagctaccgtggacaagtctagatggcagc
 aggggaacgtcttctcatgctccgtgatgcatgaggctctgcacaaccactacacgcagaagagcctctccctgtctccggcaact
 ggctctcattgtcctggcggtgctggctggcctgctgctgtttattgggtgggcatcttcttgtgtccggtgtcggcataggaggcgc
 caaggaggtggcgatctggagggggaggtatctggagggggctcaggatcagggggaggtatctggaggcggtacactgagta
 aaaccactgtgaggctcgtactagagatgatgtgcttagagctgtccgaactctggctgctgcttccggtattaccctgccactg
 ccataccgtcgatcccgatcgccacattgaacgagtcaccgaactccaggagctgtttctactagagtcgggctgatatggcaag
 tctgggtggccgatgacggagccgctgtcgtgtgtggactacacctgagctgtggaggctggcgccgtgttctgaaattggacct
 cggatggctgaactgtctggatctcgactggctggccagcagcagatggagggtgctggtgaccccatagaccaaaaggaaactgc
 ctggttctggcaactgtgggagtgctacccgatcatcaggcgaaaggactgggatctgctggtgtctccctggcgtggaggccgt
 gaacgagctggcgtccccgttttctgaaacttctgccccgaaatctcccttctacgaacgactgggattcactgtcaccgctg
 gtcgaagtgcctgaggggcctagaacatggtgtatgaccggaaaccggagcttaaccgtttaaacccgtgatcagcctcgactgt
 gccttctagtgtccagccatctgtgttggccccctccccgtgcttcttacccttgaaggtgccactcccactgtccttctataaaaa
 tgaggaaattgcatcgactgtctgagtaggtgtcattctattctgggggtgggggtggggcaggacagcaagggggaggtatggga
 agacaatagcaggcatgctgggatgctgggtgggtctatggctcaggttaattaaactggcctcatgggccttccgctcactgcccgt
 tccagtgggaaacctgtctgcccagctgcattaacatggtcatagctgttcttctgctattggcgctcctccgttctcgtcactgac
 tctgctgctcgtgctgttccggtaaagcctgggtgcctaagtagcaaaaggccagcaaaaggccaggaaaccgtaaaaaggccgc
 gttgctggcgttttccataggctccgccccctgacgagcatcaaaaaatcgacgtcaagtcagaggtggcgaaccggacagg
 actataaagataccaggcggttccccctggaagctccctcgtgctcctcgttccgacctgcccgttaccggatacctgtccgcttt
 ctcccttccgggaagcgtggcgcttctcatagctcagctgtaggtatctcagttcgggtgtaggtcgttgcctcaagctgggctgtgtc
 acgaacccccgttccagccgaccgtgctgcttccgttaactatcgtctttagtccaaccggtaagacacgacttatcgccactg
 gcagcagccactgtaacaggattagcagagcgaggtatgtaggcggtgctacagagtcttgaagtgtggcctaactacggctac
 actagaagaacagtatttggatctgcgtctgtgaagccagttaccttccgaaaaagagttgtagctcttgatccggcaaaacc
 accgtggttagcgggtgtttttgttgaagcagcagattacgcgcagaaaaaaggatctcaagaagatccttgcattttctacgg
 ggtctgacgtcagtggaacgaaaactcacgttaagggttttggctatgagattatcaaaaaggatcttccactagatcctttaaataa
 aaatgaagttttaataatcaatctaaagtatatatagtaaaacttggctgacagttaccaatgcttaatcagttagggacactatctcagcgatc
 tgtctatttctgtatccatagttgcctgactccccgtcgttagataactacgatacgggagggttaccatctggccccagtgctgcaat
 gataccgcgagaaccacgtcaccggctccagatttatcagcaataaaccagccagccggaaggccgagcgagcaagtggtcct
 gcaactttatccgctccatccagctatattgttgcggggaagctagagtaagtagttcggcagttaatagtttgcgaacgttgtgc
 cattgtacaggcatcgtggtgtcacgtcgtcgttggatggttcttctcagctccgggttccaacgatcaaggcgagttacatgatcc
 ccatgtgtgcaaaaaagcgggttagctccttccgtcctccgatcgttgcagaagtaagttggccgaggttatcactcatggtatgg
 cagcactgcataattcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgat
 gggcgacccgagttgtcttggccggcgtcaataacgggataataccgcgccacatagcagaactttaaaagtgtcatcattggaaaa
 cgttcttccggggcgaactctcaaggatctaccgtgtttagatccagttcagatgaacccactcgtgcaccaactgatcttcagcat
 ctttactttaccagcgtttctgggtgagcaaaaacaggaaggcaaaatccgcgaaaaagggaataaggcgacacggaaatgttg
 aatactcatacttcttcttcaatatttgaagcatttatcagggttattgtctcatgagcggatacataattgaatgtatttagaaaaataaa
 caaatagggggtccgcgcacatttccccgaaaagtccac

[SEQ ID NO:1]

FIG. 17 (con.)

cttaattgtaagcgcttaataattttgttaaaaaltcgcgttaaaattttgttaaaatcagctcatttttaaaccaataggccgaatcggcaaaatccg
ttataaattcaaaagaatagaccgagatagggttgagtgccgctacagggcgctcccaattcgccattcaggctgcgcaactgttggga
agggcggttcggtgcgggcctcttcgctattacgcccagctggcgaaagggggatgtgctgcaaggcgattaagtgggtaacgccag
gggtttcccagtcacgacgttgtaaaacgacggccagtgagcgcgacgtaatacgaactactatagggcgaaftggcggaaggccgt
caaggcctaggcgcgccctgaataacttcgtatagcatacatattatagcaattatcgaaaagcctgaactcaccgcgacatccgtggag
aaattcctcatcgaaaaatcgaactccgtgtccgaltcatgcagctgtccgagggcgaggagagtagagcafttctatcgcgatgtggg
cgggagagggtacgtgctgagagtgaactctgtgcccagcggttctacaaggaccgatacgtctaccggcattttgcttccgcccgtc
tgcctattccagaagtcttggaacttggggagtttagcgagtcctcacttactgtattagccggcgagcccagggagtgacactccag
gatctgcctgaaactgaactgcctgctgtgctccagcctgtcgtgaggaatggatgtattgtgctgctccgatctgagtcagactagc
ggattcggcccatttggaccccagggcatttggccagtacacaacatggcgagacttcatctgtgctatcgccgatcctcactgtacca
ttggcagactgtgatggacgatactgtgtctgttctgtggcacaggcactcgacgaactcatgtgtgggctgaggactgtcctgaagt
gagacatctgggtccatgccgatttggctccaacaatgtgtcaccgataacgggagaaactcactgccgtgatcgactggagcgaggca
atgtttggcgattcccagtacgaagtggccaacatcttctttggcgcccttggctggtgtgatggaacagcagaccgggtactttgaac
ggcgccacctgagctggctgggagtcctagactgagagcctacatgtccgaattggcctggatcagctctaccagtcactggtgga
tggcaatttcgacgatgtgcttgggcacagggcgctgtgtatgtctattgtccgatccggcgctggaactgtggggagaacacagatc
gctaggagatccgctgctgttggaccgatggatgtgtggaagtgtggtccgatagtggaaaccggaggccctcaacccgaccccg
ggcaaggagtaatgaccgtttaaaccgcgtgatcagcctcgaactgtgcttctagtgtccagccatctgttgttggccctccccgtgc
cttcttgaccttggagggtgccactcccactgtccttcttaataaaatgaggaaattgcatcgcaattgtctgagtaggtgtcattctattct
gggggggtgggggtgggggcaggacagcaagggggaggattgggaagacaatagcagggcatgtctggggatgcgggtgggctctatgg
ggatcccgcggttgacattgatttactagttaataatagtaatacgaattacgggggtcattagtctatagccataataggaattccgcgttac
ataacttacggtaaatggcccgctggctgaccgccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaaacgc
caatagggactttccattgacgtcaatgggtggagtattacggtaaaactgccacttggcagttacatcaagtgtatcatatccaagtac
gccccctattgacgtcaatgacggtaaatggccgcctggcattatgccagttacatgacctatgggacttctacttggcagttacatc
acgtattagtcatcgctattaccatgggtgatcggttttggcagttacatcaatggcggtggatagcgggttgactacgggggatttccaa
gtctccacccttgcgtcaatgggagtttgttggcacaaaatcaacgggacttccaaaatgtcgtlaaactccgccccattgac
gcaaatggcggttaggcgtgtacggtgggaggtctatataagcagagctctctgggctaactagagaaccactgcttactgtctgcag
atctgatcaagagacaggataaggagccgccaccatggagtttgggctgagctggctttttctgtggctattttaaaaggtgtccaggtg
taccatacagatgttccagattacgcttgcctgccccacagtggtagtactccactgtctgggtgtacaaaaacctccctgcacgcctctc
taacctcacaattctgtggggggccgcgcgccaccatgattgaacaagatggattgcacgcaggttctccggccgcttgggtggagag
gctattcggctatgactgggcacacagacaatcggctgctctgatgccgcccgtgttccggctgtcagcgcagggggcgcccgggttctt
ttgtcaagaccgacctgtccggtgcccgtgaatgaactgcaggacgaggcgagcgcggtatcgtggcttggccacgacggggcgttctc
gcgagctgtgctgcaggttgcactgaagcgggaagggaactgggtgctatttggcggaagtgcgggggcaggatctctgtcatctc
acctgtctctgcgagaaaagtatccatcatggctgatgcaatgcggcggtgtcatagcctgatccgggtacctgcccattcgaccac
caagcgaacatcgcatcgagcgagcacgtactcggatggaagccgggtctgtgatcaggtgaglacaggaggtggagagtacgc
gtaacacttaagcgtctctccaagtgcgaaggggacaggaggttttgtlaagggtgtatcactgtggaccagttcagatgcggcaacg
gccagtgcacccccgtgattgggtgtgacgagcggtgaacgactgccccgattccgatgaggaaggctgccccctagaacctgtg
cccctagccagcacagtatacagcccctaacaaaaacccctactgcaacctggcggtlaagagacgtccggaggccagcccttctca
tgttcagagaacatggtaactggttaagtcattgtctcccacaggatgatctgacgaagagcatcaggggctcgcgccagccgaac
tgttcgccagggtcaaggcgcgcatcccgcacggcgaggatctgctgtgacccatggcgatgcctgcttgcggaatatcatggtgg
aaaatggccgcttttctgattcatcgactgtggccggctgggtgtggcgaccgctatcaggacatagcgttggctaccctgtatattg
ctgaagagcttggcggcgaatgggctgaccgcttctctgtctttacggtatcgccgctcccgaattcgagcgcacatcgcttctatcgc
cttcttgacgagttcttctgagtcgactgcaggagtcctcactgcacccccctccagcttctctgttccaggcaccaggccaggtatctg
gggtgtgcagccggcctgggtgtgacctgaggccacaagcccgggggtctgtgtggtgtggggacagggaacggcgctgcctctgc
tctgtgcttgggcatgtgacccattcgagtgtctgtcacgggcacaggttttgtacaccagacagtggagtactaccactgtgttcca
gtgcggctcgggtactgtatcagccagagatgggtgtgcgacggggagaatgattgcgaggacggcagcgacgagggccaattgt
gcccgtctgtgcttaccgagcccaaatcttgtgacaaaactcacacatgccaccgtgccagcacctgaactcctggggggagcg

FIG. 18

V84629WO

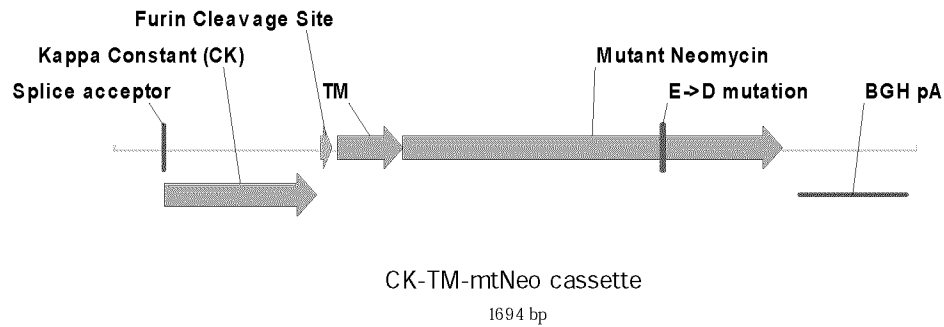
tcagtcttctcttcccccaaaaccaaggacacctcatgatctctagaaccttgaggtcacatgcgtgggtgggacgtgagcca
cgaagacctgaggtcaagtccaactggtacgtggacggcggtggaggtgcataatgccaagacaaagccgaggagcagtag
aacagcacgtaccgtgtggtcagcgtcctaccgtcctgaccaggactggctgaatggcaaggagtacaagtgaagggttccaac
aaagccctccagccccatcgagaaaaccatctccaaagccaaaggcgagccccgagaaccacaggtgtacacctgccccat
cccggtgatgagctgaccaagaaccaggtcagcctgacctgctggtcaaaaggcttctatccagcgacatcgccgtggagtgggag
agcaatgggcagccggagaaactacaagaccacgcctcccgctggtggactccgacggctccttctctacagcaagctacc
gtggacaagtctagatggcagcaggggaacgtcttctcatgtccgtgatgcatgaggctctgcacaaccactacacgcagaagagc
ctctccctgtctccgggcaactggtctcattgtctggcggtggctggcctgctgctgttattgggtgggcatcttcttctgtc
cgggtgcggcataggagcgccaaggaggtggcggtatctggagggggaggtatctggagggggctcaggatcagggggaggtatc
tggagcggtacaactgagtacaaaccactgtgaggctcgtactagagatgatgtgcctagagctgtccgaactctggctgctgcc
ttcgccgattaccctgccactcgccataccgtcgatcccgatcgccacattgaacgagtcaccgaactccaggagctgtttctactaga
gtcgggctggatattggcaagctggtgggtggccgatgacggagccgctgctgctgtgtggactacacctgagctgtggaggctggc
gccgtgttctgtaaattggacctggatggctgaactgtctggatctgactggctgcccagcagcagatggagggactgtctggcac
cccatagaccaaaggaaactgctgttctggcaactgtgggaggtgcacccgatcatcagggcaaaaggactggatctgcccgtgt
gtcctctggcggtggagggcggtgaacgagctggcgctccccgttctcgaactctgcccccgaaatctcccttctacgaacgac
tgggattcactgtcaccgccgatgtcgaagtgcctgagggggcctagaacatggtgtatgaccggaaaccggagcttaaccgtttaa
accgctgatcagcctcgactgtgcttctagtgtccagccatctgttctgccccctcccgctgcttctgacctggaaggtgcca
ctccactgtccttcttaataaaatgaggaaattgcatcgattgtctgagtaggtgtcattctattctgggggggtgggggtggggcagga
cagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggctcgagttaactgacctcatg
ggccttccgtcactgcccgttccagtcgggaaacctgtcgtgccagctgcattaacatggtcatagctgttcttgcgtattggcg
ctctccgttctcgtcactgactcgtcgtcgtcgttctgggttaaagcctgggggtgcctaatgagcaaaaggccagcaaaaggc
caggaaaccgtaaaaaggccggtgtgtggcggttctcataggtcggccccctgacgagcatcacaataacgacgtcaagtcag
aggtggcgaaaccggacaggactataaagataccaggcggttccccctggaagctccctcgtgcgtctcctgttccgacctgccc
ttaccggatacctgtccgcttctccttctcggaagcgtggcgcttctcatagctcacgctgtaggtatctcagttcgggttaggtcgtt
cgtccaagctgggctgtgtgcacgaacccccgttcagccccgaccgtgcgccttatccggtaactatcgtctttagtccaaccgggt
aagacacgacttatcgccactggcagcagccactggttaacaggattagcagagcgaggtatgtaggcgggtgtacagagttcttgaa
gtggtggcctaactacggctacactagaagaacagttatgtgtatctgcgtcgtgaagccagttaccttggaaaaagagttgtag
ctcttgatccggcaaaacaaccaccgctgtagcggtggtttttgttgaagcagcagattacgcgcagaaaaaaggatctcaag
aagatcctttagtcttctacggggtgtgacgtcagtggaacgaaaactcacgttaagggttttggatgagattatcaaaaggatc
ttcacctagatccttttaataaaatgaagttttaaatacaatctaaagtatatatgagtaaaactggtctgacagttaccaatgcttaacag
tgaggcacctatctcagcgatctgtctatttctgtcatccatagttgcctgactccccgtcgtgtagataactacgatacggagggttac
catctggccccagtgctgcaatgataccgcgagaaccacgctcaccggctccagatttatcagcaataaaccagccagccggaagg
gccgagcgagaagtggctctgcaacttatccgctccatccagcttattaattgttgcgggaagctagagtaagtagttccagtt
aatagtttgcgcaacgttgttgcattgctacaggcatcgtggtgtcacgctcgtcttggatggcttctcagctccggttcccaacg
atcaaggcgagttacatgatccccatgttgtgcaaaaagcggttagctccttggctcctccgatcgttgcagaagtaagttggccgc
agtgttatcactcatggttatggcagcactgcataattcttactgtcatgccatccgtaagatgcttttctgtactggtgagtactcaacc
aagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcggcgctcaatacgggataataccgcgccacatagcagaacttt
aaaagtgtcatcattggaaaacgttcttccggggcgaaaactctcaaggtatcttaccgctgttgagatccagttcgatgtaaccactcgt
gcaccaactgatcttcagcatcttttaccagcggttctgggtgagcaaaaacagggaaggcaaaatgccgcaaaaagggaat
aaggcgacacggaaatgtgaatactcactcttcttcaatattatgaagcatttatcagggttattgtctcatgagcggatacata
ttgaatgtatttagaaaaataaacaataggggtccgcacatttccccgaaaagtggcac

[SEQ ID NO:40]

FIG. 18 (con.)

V84629WO

A.



B.

aattcttctgtctgtccctaacaatgccctgtgattatccgcaacaacacaccaagggcagaactttgttacttaaacaccatcctgtttgc
ttcttctcaggaactgtggctgcacatctgtcttcatcttcccgccatctgatgagcagttgaaaagcgggaacagccagcgttgtgtg
cctgctgaataacttctatccagagaggccaaagtacagtggaaagtggaataacgccctccaatcgggtaactcccaggagagtgc
acagagcaggacagcaaggacagcacctacagcctcagcagcaccctgacgtgagcaaagcagactacgagaacacaaagtc
tacgctgcgaagtacccatcaggccctgagcagccccgtcacaaagagcttcaacaggggagagtgtggcggcgagc**agctccc**
ggcaccgcccagccctgggcgggcggcagcgacgtcccgctcaaatattgcaaaaattatcatcgccccctcatcttgtcttcttcttct
ccgttgtgattggaagtattatctattcctgagaaagaggcagccagatgggcccgtgggaccgctttacgcttctggaagcgctattg
aacaagatggattgcacgcaggttctccggccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggtgct
ctgatccgcccgtgtccggctgtcagcgaggggcgcccgggtcttttgtcaagaccgacctgtccggtgccctgaatgaactgcag
gacgaggcagcgcggtatcgtggctggccacgacggcgcttcttgcgcagctgtgtcgcacgttgcactgaagcgggaaggga
ctggctgtatgggcgaagtgcggggcaggatctcctgtcatctcaccctgtcctgcccagaaagtatccatcatggctgatgcaat
gcgggcggtgcatacgttgcacggctacgtgcccttcgaccaccaagcgaaacatcgcatcgagcgagcacgtactcggatgg
aagccggtcttgcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgccaggctcaaggcgcg
atgcccgacggcgatgatctcgtgacccatggcgatgcctgttgcgaatatcatggtggaaaatggccgcttttctggattcatc
gactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctaccgctgataattgctgaagagcttggcggcgaatgg
gctgaccgcttctcgtgctttacggtatcggcgtcccgaattgcagcgcatcgcttctatcgcttcttgacgagttcttctgatctaga
gggcccgtttaaaccgctgatcagcctcgaactgtgccttctagtgtccagccatctgtgtttgccccctccccctgcttcttctgacct
ggaaggtgccactcccactgtccttcttaataaaaataggaaattgcacgcattgtctgagtaggtgtcattctattctgggggtggg
gtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggcttctgaggcg
gctgcagttatg

[SEQ ID NO:41]

FIG. 19

V84629WO

CTAAATTGTAAGCGTTAATATTTTGTAAAATTTCGCGTTAAATTTTGTAAATCA
GCTCATTTTTTAACCAATAGGCCGAAATCGGCCAAAATCCCTTATAAATCAAAAGA
ATAGACCGAGATAGGGTTGAGTGGCCGCTACAGGGCGCTCCCATTCGCCATTCA
GGCTGCGCAACTGTTGGGAAGGGCGTTTCGGTGCGGGCCTCTTCGCTATTACGCC
AGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGT
TTTCCCAGTCACGACGTTGTAACGACGCGCCAGTGAGCGCGACGTAATACGAC
TCACTATAGGGCGAATTGGCGGAAGGCCGTCAAGGCCTACGCGCCGACCGAGTC
CACTAGTTAACTAGCTGAGGGCCGAGGCCGCCCTGCTGCAGTGCACCGCCGATA
CCCTGGCCGACGCCGTGCTGATCACCACCGCCCACGCTTGGCAGCACCAGGGCA
AGACACTGTTTCATCTCCAGAAAGACCTACCGGATCGACGGCAGCGGCCAGATGG
CCATCACAGTGGACGTGGAGGTGGCCTCCGACACCCCTCACCCCGCCAGAATCG
GCCTGAAGTGTGAGCTGGCCAGGTGGCCGAGAGAGTGAAGTGGCTGGGCCTGG
GCCCCCAGGAGAACTACCCCGACCGGCTGACCGCCGCCTGCTTCGACAGATGGG
ACCTGCCTCTGAGCGACATGTACACCCCTACGTGTTCCCCAGCGAGAATGGCCT
GAGATGCGGCACCCGGGAGCTGAACTACGGCCCCCACCAGTGGAGGGGCGACTT
CCAGTTCAACATCAGCCGGTACTCGGATCTGAAGTTCCTATTCCGAAGTTCCTAT
TCTCTAGAAAGTATAGGAACTTCGGGCCACATGGACAGAGGCCGGCTCGGCCCA
CCCTCTGCCCTGAGAGTGACCGCTGTACCAACCTCTGTCCCTACAGGGCAGCCCC
GAGAACCACAGGTGTACACCCTGCCTCCAAGCAGGGACGAACTGACAAAGAATC
AGGTGTCCCTCACCTGCCTCGTGAAGGGCTTTTACCCCTCCGATATCGCAGTGGA
ATGGGAGTCCAACGGCCAGCCCGAGAATAATTACAAAACAACCCCCCTGTGCT
GGACAGCGACGGCAGCTTCTTTCTGTACTCCAAGCTGACAGTGGATAAGTCCCGC
TGGCAGCAGGGCAATGTGTTGAGCTGCTCTGTGATGCACGAAGCCCTCCACAATC
ATTATACCCAGAAGTCCCTGAGCCTGAGCCCCGGCAAGTGACAATTCCAACGCC
GCCCCCTCTCCCTCCCCCCCCCTAACGTTACTGGCCGAAGCCGCTTGGAATAAGG
CCGGTGTGCGTTTGTCTATATGTTATTTTCCACCATATTGCCGTCTTTTGGCAATG
TGAGGGCCCCGAAACCTGGCCCTGTCTTCTTGACGAGCATTCTAGGGGTCTTTC
CCCTCTCGCCAAAGGAATGCAAGGTCTGTTGAATGTCGTGAAGGAAGCAGTTCCT
CTGGAAGCTTCTTGAAGACAAACAACGTCTGTAGCGACCCTTTGCAGGCAGCGG
AACCCCCCACCTGGCGACAGGTGCCTCTGCGGCCAAAAGCCACGTGTATAAGAT
ACACCTGCAAAGGCGGCACAACCCAGTGCCACGTTGTGAGTTGGATAGTTGTG
GAAAGAGTCAAATGGCTCTCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCC
CAGAAGGTACCCCATTTGTATGGGATCTGATCTGGGGCCTCGGTACACATGCTTTA
CATGTGTTTAGTCGAGGTTAAAAAACGTCTAGGCCCCCGAACCACGGGGACG
TGGTTTTCTTTGAAAAACACGATGATAATATGGCCACAAGATCTGCCACCACCG
CCGCCAACATGAGCGAAAAATACATCGTCACCTGGGACATGTTGCAGATCCATG
CACGTAACTCGCAAGCCGACTGATGCCTTCTGAACAATGGAAAGGCATTATTGC
CGTAAGCCGTGGCGGTCTGGTACCGGGTGC GTTACTGGCGCGTGAAGTGGGTATT
CGTCATGTGCGATACCGTTTGTATTTCCAGCTACGATCACGACAACCAGCGCGAGC
TTAAAGTGCTGAAACGCGCAGAAGGCGATGGCGAAGGCTTCATCGTTATTGATG
ACCTGGTGGATACCGGTGGTACTGCGGTGCGATTCTGTGAAATGTATCCAAAAGC
GCACTTTGTCACCATCTTCGCAAAACCGGCTGGTCGTCCGCTGGTTGATGACTAT
GTTGTTGATATCCCGCAAGATACCTGGATTGAACAGCCGTGGGATATGGGCGTCG
TATTCGTCCCGCCAATCTCCGGTCGCTAATCTTTTCAACGCCTGGCACTGCCGGG
CGTTGTTCTTTTAACTTCAGGCGGGTTACAATAGTTTCCAGTAAGTATTCTGGAG
GCTGCATCCATGACACAGGCAAACCTGAGCGAAACCTGTTCAAACCCCGCTTTA
AACATCCTGAAACCTCGACGCTAGTCCGCCGCTTTAATCACGGCGCACAAACGCC
TGTGCAGTCGGCCCTTGATGGTAAAACCATCCCTCACTGGTATCGCATGATTAAAC

FIG. 20

CGTCTGATGTGGATCTGGCGCGGCATTGACCCACGCGAAATCCTCGACGTCCAGG
CACGTATTGTGATGAGCGATGCCGAACGTACCGACGATGATTTATACGATACGGT
GATTGGCTACCGTGGCGGCAACTGGATTTATGAGTGGGCCCCGGATCTTTGTGAA
GGAACCTTACTTCTGTGGTGTGACATAATTGGACAAACTACCTACAGAGATTTAA
AGCTCTAAGGTAAATATAAAATTTTTAAGTGTATAATGTGTTAAACTACTGATTC
TAATTGTTTGTGTATTTTAGATTCCAACCTATGGAAGTATGAATGGGAGCAGTG
GTGGAATGCCTTTAATGAGGAAAACCTGTTTTGCTCAGAAGAAATGCCATCTAGT
GATGATGAGGCTACTGCTGACTCTCAACATTCTACTCCTCCAAAAAAGAAGAGA
AAGGTAGAAGACCCCAAGGACTTTCCTTCAGAATTGCTAAGTTTTTTGAGTCATG
CTGTGTTTAGTAATAGAACTCTTGCTTGCTTTGCTATTTACACCACAAAGGAAAA
AGCTGCACTGCTATACAAGAAAATTATGGAAAAATATTCTGTAACCTTTATAAGT
AGGCATAACAGTTATAATCATAACATACTGTTTTTTCTTACTCCACACAGGCATA
GAGTGTCTGCTATTAATAACTATGCTCAAAAATTGTGTACCTTTAGCTTTTTAATT
TGTAAGGGGGTTAATAAGGAATATTTGATGTATAGTGCCTTGACTAGAGATCATA
ATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACC
TCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATT
GCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTACAAATAAA
GCATTTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAACTCATCAATGTATCTTA
TCATGTCTGGGCTAGCCTCGAAGAGATTAGAATCCGAGGGCGCGCCTGAATAAC
TTCGTATAGCATACATTATAGCAATTTATCGAAAAAGCCTGAACTCACCGCGACA
TCCGTGGAGAAATTCCTCATCGAAAAATTCGACTCCGTGTCCGATCTCATGCAGC
TGTCCGAGGGCGAGGAGAGTAGAGCATTCTCATTCGATGTGGGCGGGAGAGGCT
ACGTGCTGAGAGTGAACCTTTGTGCCGACGGCTTCTACAAGGACCGATACGTCTA
CCGGCATTTTGTCTCCGCCGCTCTGCCTATTCCAGAAGTCCTGGACATTGGGGAG
TTAGCGAGTCCCTCACTTACTGTATTAGCCGGCGAGCCCAGGGAGTGACACTCC
AGGATCTGCCTGAACTGAACTGCCTGCTGTGCTCCAGCCTGTCGCTGAGGCAAT
GGATGCTATTGCTGCTGCCGATCTGAGTCAGACTAGCGGATTCGGCCCATTGGA
CCCCAGGGCATTGGCCAGTACACAACATGGCGAGACTTCATCTGTGCTATCGCCC
ATCCTCACGTGTACCATTGGCAGACTGTGATGGACGATACTGTGTCTGCTTCTGT
GGCACAGGCACTCGACGAACTCATGCTGTGGGCTGAGGACTGTCCTGAAGTGAG
ACATCTGGTCCATGCCGATTTTGGCTCCAACAATGTGCTCACCGATAACGGGAGA
ATCACTGCCGTGATCGACTGGAGCGAGGCAATGTTTGGCGATTCCCAGTACGAA
GTGGCCAACATCTTCTTTTGGCGGCCTTGGCTGGCTTGTATGGAACAGCAGACCC
GGTACTTTGAACGGCGCCACCCTGAGCTGGCTGGGAGTCCTAGACTGAGAGCCT
ACATGCTCCGAATTGGCCTGGATCAGCTCTACCAGTCACTGGTGGATGGCAATTT
CGACGATGCTGCTTGGGCACAGGGGCGCTGTGATGCTATTGTCCGATCCGGCGCT
GGAAGTGTGGGGAGAACACAGATCGCTAGGAGATCCGCTGCTGTCTGGACCGAT
GGATGTGTGGAAGTGCTGGCCGATAGTGGAACCGGAGGCCTTCAACCCGACCC
CGGGCAAAGGAGTAATGACCGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTC
TAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAG
GTGCCACTCCCCTGTCTTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCT
GAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGA
GGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGGGAT
CCCGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCA
TTAGTTCATAGCCCATATATGGAGTTCGCGGTTACATAACTTACGGTAAATGGCC
CGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGT
TCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTA
CGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCC

FIG. 20

CTATTGACGTCAATGACGGTAAATGGCCCCGCTGGCATTATGCCCAGTACATGAC
CTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCA
TGGTGATGCGGTTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACG
GGGATTTCCAAGTCTCCACCCATTGACGTCAATGGGAGTTTGTGTTTGGCACCAA
AATCAACGGGACTTTCCAAAATGTCGTAACAACCTCCGCCCCATTGACGCAAATGG
GCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTA
GAGAACCCACTGCTTACTGCTCGACGATCTGATCAAGAGACAGGATAAGGAGCC
GCCACCATGGACATGAGGGTCCCCGCTCAGCTCCTGGGGCTCCTGCTACTCTGGC
TCCGAGGTGCCAGATGTGACATCCAGATGACCCAGAGCCCCAGCAGCCTGAGCG
CCAGCGTGGGCGACAGAGTGACCATCACCTGTGCGGCCAGCCAGTCGATCAGCA
GCTACCTGAACTGGTATCAGCAGAAGCCCGGCAAGGCCCCCAAGCTGCTGATCT
ACGCCGCCAGCTCCCTGCACAGTGGTAGTACTCCACTGTCTGGGTGTACAAAAC
CTCCCTGCACGCCTCTCTAACCTCACAATTCTGTGACGCGTTCCTGAATGCGGCC
GCCACTTCCGGAGTGCTGGATATCAGTCGACCGGTCTGAGTGTCACACCTACTGC
GAGCTCAACCTGGCGGTAACCTGTGACCTGGCGGTATGTGTCAGACCCAGATCTCG
TACTCCAATAGGTCCAAGCTTGGTTTGAGAGGAGAATAGGATTTCATGGGGGAA
ATGGGGAAGAAATAGCTAGATTTTTCTCTGAACAAGCAGCCTATCTCATATGATT
GGCTTCAAGAGAGGTTTTTGTATTGGTCTGTACCACTGTGTCCCTGCAGAGCGGC
GTGCCAAGCAGATTCAGCGGCAGCGGCTCCGGCACCGACTTCACCCTGACCATC
AGCAGCCTGCAGCCCCGAGGACTTCGCCACCTACTACTGCCAGCAGAGTTACAGT
ACCCCTCTCACTTTCGGCGGAGGGACCAAGGTGGAGATCAAACGTAAGTGCCT
TTCCTAATGCTTTTTCTTATAAGGTTTTAAATTTGGAGCGTTTTTGTGTTTGAGAT
ATTAGCTCAGGTCAATTCCAAAGAGTACCAGATGAATTCTTCTGTCTGTCCCTAA
CATGCCCTGTGATTATCCGCAAACAACACACCCAAGGGCAGAACTTTGTTACTTA
AACACCATCCTGTTTGCTTCTTTCTCAGGAACTGTGGCTGCACCATCTGTCTTCA
TCTTCCCGCCATCTGATGAGCAGTTGAAAAGCGGAACAGCCAGCGTTGTGTGCCT
GCTGAATAACTTCTATCCCAGAGAGGCCAAAGTACAGTGGAAAGGTGGATAACGC
CCTCCAATCGGGTAACCTCCAGGAGAGTGTACACAGAGCAGGACAGCAAGGACAG
CACCTACAGCCTCAGCAGCACCCCTGACGCTGAGCAAAGCAGACTACGAGAAACA
CAAAGTCTACGCCTGCGAAGTCACCCATCAGGGCCTGAGCAGCCCCGTCACAAA
GAGCTTCAACAGGGGAGAGTGTGGCGGCGGCAGCTCCCGGCACCGCCGAGCCCT
GGGCGGCGGCAGCGACGTCCCGTCAAATATTGCAAAAATTATCATCGGCCCCCT
CATCTTTGTCTTTCTCTTCTCCGTTGTGATTGGAAGTATTTATCTATTCCCTGAGAA
AGAGGCAGCCAGATGGGCGGCTGGGACCGCTTTACGCTTCTGGAAGCGCTATTG
AACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTG
GCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCT
GTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTG
AATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTT
CCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTAT
TGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAA
AGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACC
TGCCCATTGACCAACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATG
GAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAGGAGCATCAGGGGCTCGCG
CCAGCCGAACTGTTGCCAGGCTCAAGGCGCGCATGCCCGACGGCGATGATCTC
GTCGTGACCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAATAATGGCCGCT
TTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACAT
AGCGTTGGCTACCCGTGATATTGCTGAGGAGCTTGGCGGCGAATGGGCTGACCG
CTTCTCGTGCTTTACGGTATCGCCGCTCCCGATTTCGACGCGCATCGCCTTCTATC

FIG. 20 (con.)

V84629WO

GCCTTCTTGACGAGTTCTTCTGATCTAGAGGGCCCGTTTAAACCCGCTGATCAGC
CTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCCCTT
CCTTGACCCTGGAAGGTGCCACTCCCCTGTCCTTTCTAATAAAATGAGGAAAT
TGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAG
GACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGT
GGGCTCTATGGCTTCTGAGGCGGCTGCAGTTATGTCGACCGCGTTGACATTGATT
ATTGACTAGTTATTAATAGTAATCAATTACGGGGTTCATTAGTTCATAGCCCATAT
ATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCA
ACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAAT
AGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAACTGCCCCACTTG
GCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACG
GTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCTCTAC
TTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGC
AGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCA
CCCCATTGACGTCAATGGGAGTTTGTGTTTGGCACCAAAATCAACGGGACTTTCCA
AAATGTCGTAACAACCTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGG
TGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACT
GGCTTATCGAAATTAATACGACTCACTATAGGGAGACACAAGCTGGCGGCCGCT
AATAAAGGCGATCTGATCAAGAGACAGGATAAAGGAGCCGCCACCATGGAGTTTG
GGCTGAGCTGGCTTTTTCTTGTGGCTATTTTAAAGGTGTCCAGTGTGAGGTGCA
GCTGGTGCAGTCTGGCGCCGAAGTGAAGAAACCTGGCGCCTCCGTGAAGGTGTC
CTGCAAGGCCAGCGGCTACACCTTACCAGCTACGGCATCAGCTGGGTCCGCCA
GGCTCCTGGACAGGGACTGGAATGGATGGGCTGGATCAGCGCCTACAACGGCAA
CACCAACTACGCCCAGAACTGCAGGGCAGAGTGACCATGACCACCGACACCAG
CACCAGCACCGCCTACATGGAACCTCGAAGCCTGAGAAGCGACGACACCGCCGT
GTATTACTGTGCGAGAGAGCTAGCCTATGATGCTTTTGATATCTGGGGCCAAGGG
ACAATGGTCACCGTGTCTCAGCCTCCACCAAGGGCCCATCGGTCTTCCCCCTGG
CACCCTCCTCCAAGAGCACCTCTGGGGGCACAGCGGCCCTGGGCTGCCTGGTCA
AGGACTACTTCCCCGAACCGGTGACGGTGTCTGTGGAACCTCAGGCGCCCTGACCA
GCGGCGTGCATACCTTCCCGGCTGTCTTACAGTCCTCAGGACTCTACTCCCTCAG
CAGCGTGGTGACCGTGCCCTCCAGCAGCTTGGGCACCCAGACCTACATCTGCAAC
GTGAATCACAAAGCCCAGCAACACCAAGGTGGACAAGAAAGTTGAGCCCAAATCT
TGTGACAAAACCTCACACATGCCCACCGTGCCCAGCACCTGAACTCCTGGGGGGA
CCGTGAGTCTTCTCTTCCCCCAAAACCCAAGGACACCCTCATGATCTCTAGAA
CCCCTGAGGTACATGCGTGGTGGTGGACGTGAGCCACGAAGACCCTGAGGTCA
AGTTCAACTGGTACGTGGACGGCGTGGAGGTGCATAATGCCAAGACAAAGCCGC
GGGAGGAGCAGTACAACAGCACGTACCGTGTGGTTCAGCGTCCTACCGTCCTGC
ACCAGGACTGGCTGAATGGCAAGGAGTACAAGTGCAAGGTGTCCAACAAAGCCC
TCCAGCCCCCATCGAGAAAACCATCTCCAAAGCCAAAGGTGGGACCCGTGGGG
TGCGAATAGAAGTTCCTATTCCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTT
CGGGCCACATGGAATTAATTCAGAGGCCGGCTCGGCCACCCTCTGCCCTGAGA
GTGACCGCTGTACCAACCTCTGTCCCTACAGGGCAGCCCCGAGAACCACAGGTG
TACACCCTGCCCCCATCCCGGGATGAGCTGACCAAGAACCAGGTGAGCCTGACC
TGCCTGGTCAAAGGCTTCTATCCAGCGACATCGCCGTGGAGTGGGAGAGCAAT
GGGCAGCCGGAGAACAACCTACAAGACCACGCCTCCCGTGCTGGACTCCGACGGC
TCCTTCTTCTCTACAGCAAGCTCACCGTGGACAAGAGCAGGTGGCAGCAGGGG
AACGTCTTCTCATGCTCCGTGATGCATGAGGCTCTGCACAACCCTACACGCAGA
AGAACCTCTCCCTGTCTCCGGGCAAAGCTGTGGGCCAGGACACGCAGGAGGTCA

FIG. 20 (con.)

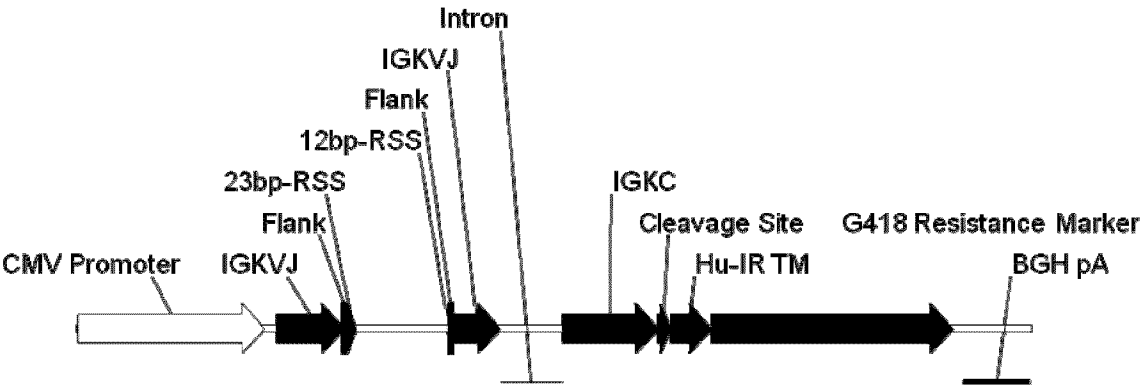
V84629WO

TCGTGGTGCCACACTCCTTGCCCTTTAAGGTGGTGGTGATCTCAGCCATCCTGGC
CCTGGTGGTGCTACCATCATCTCCCTTATCATCCTCATCATGCTTTGGCAGAAGA
AGCCACGTTAGGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCT
CATGCTGGAGTTCTTCGCCACCCCAACTTGTTTATTGCAGCTTATAATGGTTACA
AATAAAGCAATAGCATCACAAATTTACAAATAAAGCATTTTTTTTACTGCATTC
TAGTTGTGGTTTGTCCAACTCATCAATGTATCTTATCATGTCTGACCTCGAGTTA
ATTAAGTGGCCTCATGGGCCTTCCGCTCACTGCCCCGCTTTCCAGTCGGGAAACCT
GTCGTGCCAGCTGCATTAACATGGTCATAGCTGTTTCCTTGCGTATTGGGCGCTCT
CCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCCGGTAAAGCCTGGGGT
GCCTAATGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTT
GCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGC
TCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTTCC
CCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCT
GTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGT
ATCTCAGTTCGGTGTAGGTGCTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCC
CGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCG
GTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGA
GCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCT
ACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGG
AAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGG
TTTTTTTGTGTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGAT
CCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAACTCACGTTAAG
GGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTA
AAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAACTTGGTCTGACAGT
TACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTATC
CATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCA
TCTGGCCCCAGTGCTGCAATGATACCGCGAGAACCACGCTCACCGGCTCCAGATT
TATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAA
CTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGAAGCTAGAGTAAGTAG
TTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTG
TCACGCTCGTCGTTTGGTATGGCTTCATTACAGCTCCGGTTCCCAACGATCAAGGC
GAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCC
GATCGTTGTGCAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCA
CTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGA
GTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGC
CCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTC
ATCATTGGAAAACGTTCTTCGGGGCGAAAACCTCTCAAGGATCTTACCGCTGTTGA
GATCCAGTTCGATGTAACCCACTCGTGCACCCAACCTGATCTTCAGCATCTTTTACT
TTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAG
GGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTCAATATT
ATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTAT
TTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCGAAAAGTGCCAC

[SEQ ID NO:74]

FIG. 20 (con.)

A.



B.

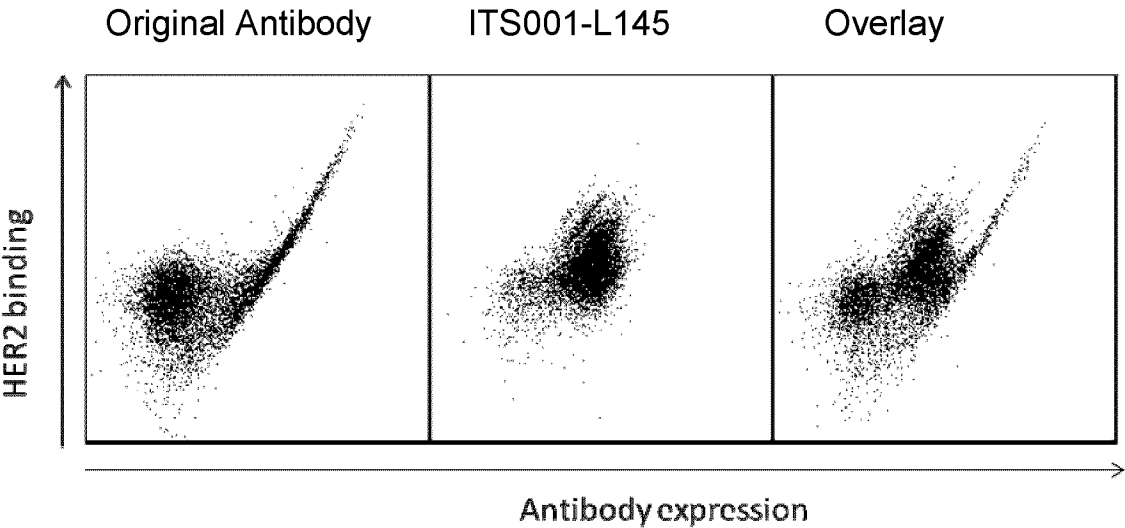
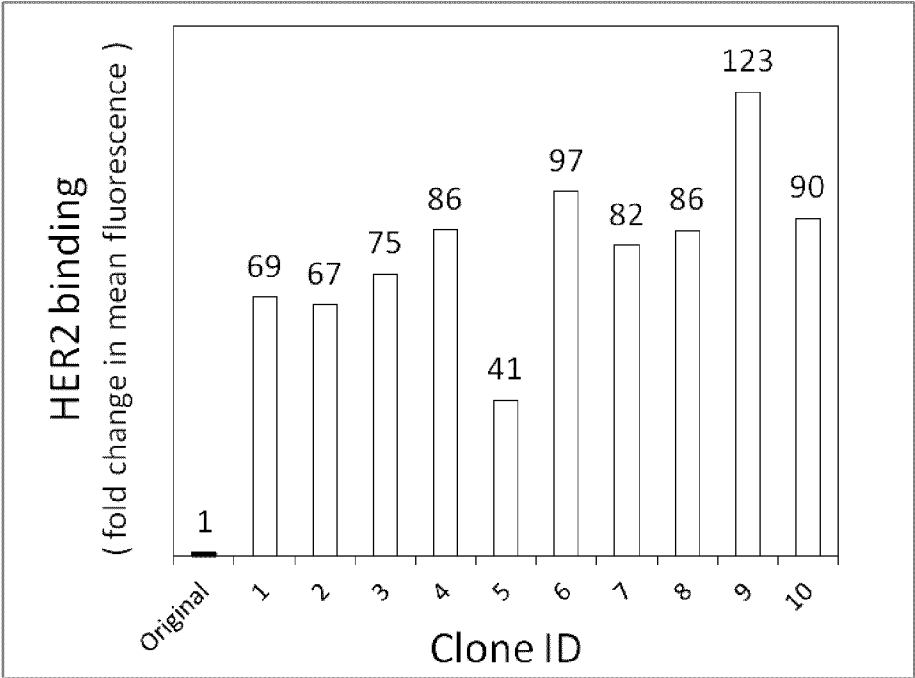


FIG. 21

A,



B.

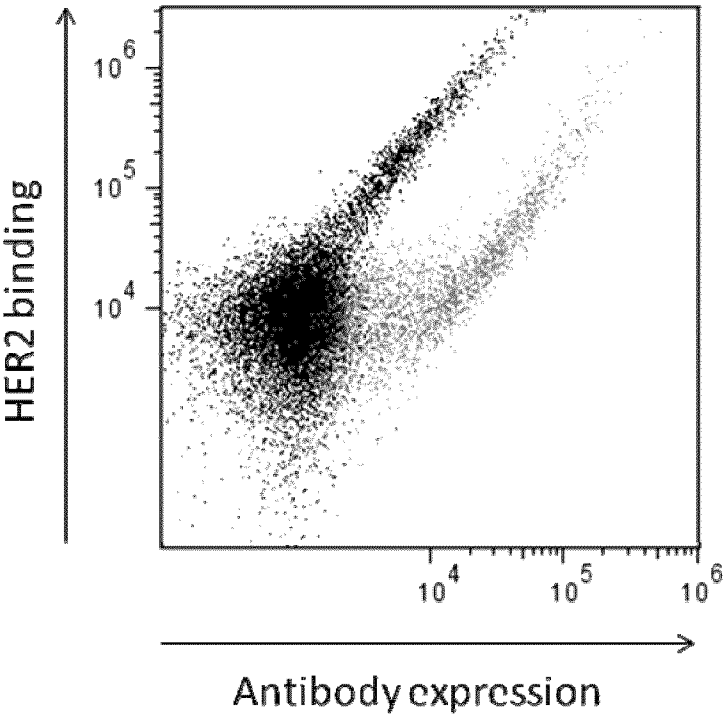
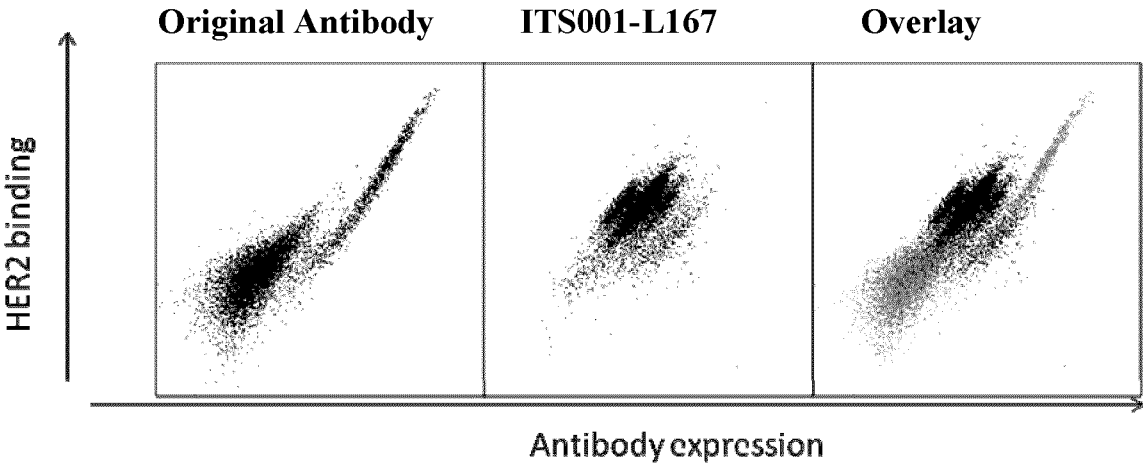
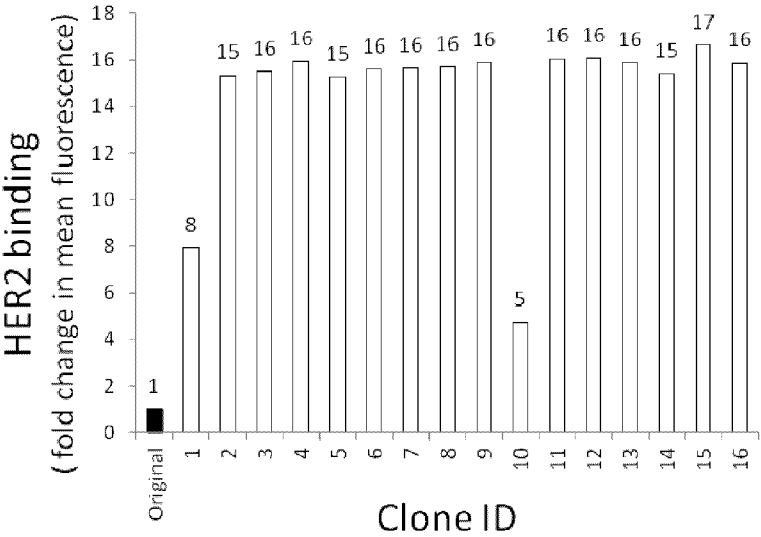


FIG. 22

A.



B.



C.

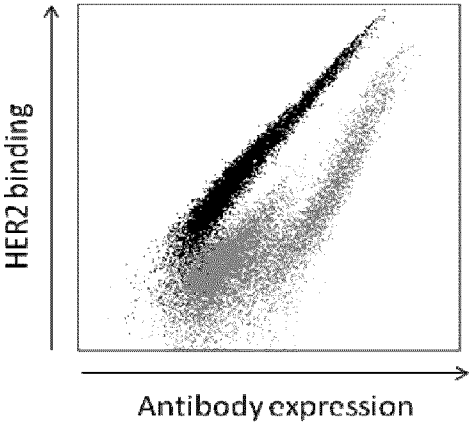


FIG. 23

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CA2013/050203

A. CLASSIFICATION OF SUBJECT MATTER IPC: <i>C12N 15/63</i> (2006.01), <i>C07K 14/78</i> (2006.01), <i>C07K 16/00</i> (2006.01), <i>C12N 15/00</i> (2006.01), <i>C12N 15/09</i> (2006.01), <i>C12N 15/11</i> (2006.01) (more IPCs on the last page) According to International Patent Classification (IPC) or to both national classification and IPC														
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC: <i>C12N 15/63</i> (2006.01), <i>C07K 14/78</i> (2006.01), <i>C07K 16/00</i> (2006.01), <i>C12N 15/00</i> (2006.01), <i>C12N 15/09</i> (2006.01), <i>C12N 15/11</i> (2006.01), <i>C12N 15/12</i> (2006.01), <i>C12N 15/13</i> (2006.01), <i>C12N 15/85</i> (2006.01), <i>C40B 50/06</i> (2006.01) Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic database(s) consulted during the international search (name of database(s) and, where practicable, search terms used) TotalPatent, Scopus, Medline, Canadian Patent Database (Keywords: recombination signal sequence, V(D)J, VDJ, somatic recombination)														
C. DOCUMENTS CONSIDERED TO BE RELEVANT <table border="1"> <thead> <tr> <th>Category*</th> <th>Citation of document, with indication, where appropriate, of the relevant passages</th> <th>Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>WO 2009/129247 A2 (GALLO et al.) 22 October 2009 (22-10-2009) *whole document*</td> <td>1-47</td> </tr> </tbody> </table>			Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.	X	WO 2009/129247 A2 (GALLO et al.) 22 October 2009 (22-10-2009) *whole document*	1-47						
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.												
X	WO 2009/129247 A2 (GALLO et al.) 22 October 2009 (22-10-2009) *whole document*	1-47												
☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex. <table border="1"> <tbody> <tr> <td>* Special categories of cited documents :</td> <td>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</td> </tr> <tr> <td>"A" document defining the general state of the art which is not considered to be of particular relevance</td> <td>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</td> </tr> <tr> <td>"E" earlier application or patent but published on or after the international filing date</td> <td>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</td> </tr> <tr> <td>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</td> <td>"&" document member of the same patent family</td> </tr> <tr> <td>"O" document referring to an oral disclosure, use, exhibition or other means</td> <td></td> </tr> <tr> <td>"P" document published prior to the international filing date but later than the priority date claimed</td> <td></td> </tr> </tbody> </table>			* Special categories of cited documents :	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention	"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone	"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art	"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family	"O" document referring to an oral disclosure, use, exhibition or other means		"P" document published prior to the international filing date but later than the priority date claimed	
* Special categories of cited documents :	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention													
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone													
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art													
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family													
"O" document referring to an oral disclosure, use, exhibition or other means														
"P" document published prior to the international filing date but later than the priority date claimed														
Date of the actual completion of the international search 21 May 2013 (21-05-2013)		Date of mailing of the international search report 19 June 2013 (19-06-2013)												
Name and mailing address of the ISA/CA Canadian Intellectual Property Office Place du Portage I, C114 - 1st Floor, Box PCT 50 Victoria Street Gatineau, Quebec K1A 0C9 Facsimile No.: 001-819-953-2476		Authorized officer Jeremy McLean (819) 994-9302												

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CA2013/050203

C12N 15/12 (2006.01) , *C12N 15/13* (2006.01) , *C12N 15/85* (2006.01) , *C40B 50/06* (2006.01)

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CA2013/050203

Patent Document Cited in Search Report	Publication Date	Patent Family Member(s)	Publication Date
WO2009129247A2	22 October 2009 (22-10-2009)	CA2721231A1	22 October 2009 (22-10-2009)
		EP2271758A2	12 January 2011 (12-01-2011)
		EP2271758A4	23 May 2012 (23-05-2012)
		US2009258392A1	15 October 2009 (15-10-2009)
		US8012714B2	06 September 2011 (06-09-2011)
		US2012258495A1	11 October 2012 (11-10-2012)