

**(12) STANDARD PATENT**  
**(19) AUSTRALIAN PATENT OFFICE**

**(11) Application No. AU 2016254498 B2**

**(54) Title**  
**Nucleic acid construct**

**(51) International Patent Classification(s)**  
**C07K 14/705** (2006.01)      **C12N 15/85** (2006.01)  
**C07K 16/28** (2006.01)

**(21) Application No:** **2016254498**      **(22) Date of Filing:** **2016.04.26**

**(87) WIPO No:** **WO16/174408**

**(30) Priority Data**

**(31) Number** **1507104.6**      **(32) Date** **2015.04.27**      **(33) Country** **GB**

**(43) Publication Date:** **2016.11.03**  
**(44) Accepted Journal Date:** **2020.10.22**

**(71) Applicant(s)**  
**UCL Business LTD**

**(72) Inventor(s)**  
**Pule, Martin;Cordoba, Shaun**

**(74) Agent / Attorney**  
**Pairman IP, PO Box 33252, Christchurch, 8244, NZ**

**(56) Related Art**

**A. EL AMRANI**, "Coordinate Expression and Independent Subcellular Targeting of Multiple Proteins from a Single Transgene", **PLANT PHYSIOLOGY**, US, (2004-05-01), vol. 135, no. 1, doi:10.1104/pp.103.032649, ISSN 0032-0889, pages 16 - 24

**DE FELIPE PABLO**, "Skipping the co-expression problem: the new 2A CHYSEL technology", **GENETIC VACCINES AND THERAPY**, BIOMED CENTRAL, LONDON, GB, (2004-09-13), vol. 2, no. 1, doi:10.1186/1479-0556-2-13, ISSN 1479-0556, page 13

**FELIPE DE P ET AL**, "Targeting of proteins derived from self-processing polyproteins containing multiple signal sequences", **TRAFFIC**, MUNKSGAARD, DK, (2004-08-01), vol. 5, no. 8, pages 616 - 626

**PABLO DE FELIPE ET AL**, "Inhibition of 2A-mediated 'cleavage' of certain artificial polyproteins bearing N-terminal signal sequences", **BIOTECHNOLOGY JOURNAL**, DE, (2009-11-27), vol. 5, no. 2, pages 213 - 223

**GARRY A. LUKE ET AL**, "Self-Processing Polyproteins: A Strategy for Co-expression of Multiple Proteins in Plants", **BIOTECHNOLOGY AND GENETIC ENGINEERING REVIEWS**, GB, (2006-12-01), vol. 23, no. 1, pages 239 - 252

**WO 2014079878 A1**

**WO 2013/185552 A1**

**WO 2007/014162 A2**

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

(19) World Intellectual Property Organization

International Bureau



(10) International Publication Number

WO 2016/174408 A1

(43) International Publication Date  
3 November 2016 (03.11.2016)

(51) International Patent Classification:  
*C07K 14/705* (2006.01)    *C12N 15/85* (2006.01)  
*C07K 16/28* (2006.01)

(21) International Application Number:  
PCT/GB2016/051164

(22) International Filing Date:  
26 April 2016 (26.04.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
1507104.6    27 April 2015 (27.04.2015)    GB

(71) Applicant: UCL BUSINESS PLC [GB/GB]; The Network Building, 97 Tottenham Court Road, London W1T 4TP (GB).

(72) Inventors: PULÉ, Martin; c/o Autolus Ltd, Forest House, Depot Road, London W12 7RP (GB). CORDOBA, Shaun; c/o Autolus Ltd, Forest House, Depot Road, London W12 7RP (GB).

(74) Agent: WILLIAMS, Aylsa; D Young & Co LLP, 120 Holborn, London EC1N 2DY (GB).

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

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

Published:

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



WO 2016/174408 A1

(54) Title: NUCLEIC ACID CONSTRUCT

(57) Abstract: The present invention provides a nucleic acid construct comprising the following structure: A – X – B in which A and B are nucleic acid sequences encoding a first and a second polypeptide of interest (POI); and X is a nucleic acid sequence which encodes a cleavage site, wherein either the first or second POI is a transmembrane protein which comprises an intracellular retention signal.

## NUCLEIC ACID CONSTRUCT

## FIELD OF THE INVENTION

The present invention relates to constructs and approaches for modulating the relative expression of polypeptides co-expressed from a single vector. In particular, the invention relates to modulating the expression of a transmembrane protein co-expressed from a single vector with a second polypeptide.

## BACKGROUND TO THE INVENTION

It is often desirable to express different proteins from the same vector since multiple transduction of the same cell is difficult, expensive and unpredictable. Different methods have therefore been developed to allow co-expression of two proteins from a single vector (see Figure 1).

Initial attempts used two different promoters within the same cassette. This results in two separate transcripts each of which code for a separate protein. This is a difficult approach for a number of reasons. A key problem is “promoter interference” whereby one promoter dominates and causes silencing of the second promoter. In addition, different promoters work differently in different cellular contexts and this makes consistent “tuning” of the relative expression of each transgene difficult to achieve.

An alternative approach is to use an Internal Ribosome Entry sequence (IRES). Here, a single transcript is generated. The IRES sequence in the transcript is placed between the open reading frames for the two transgenes and mimics an mRNA cap structure. Hence, the ribosome either initiates translation at the 5' cap or the IRES resulting in expression of two separate proteins. A key limitation with this approach is the inability to control relative expression. The 3' transcript is typically expressed less than the 5' one, but the ratio of expression is difficult to predict and tune.

A further approach has been provided following characterization of the role of foot-and-mouth-disease virus (FMDV) 2A peptide in allowing FMDV (and related viruses) to express multiple proteins from a single open reading frame (ORF) (Donnelly *et al*; *J. Gen. Virol.*; 82, 1027–1041 (2001)). The 2A peptide (and homologs) cleaves at very high efficiency immediately after translation of the ORF, enabling the expression of multiple peptides from a

single ORF. A problem with the use of the 2A peptide to cleave between different peptides in the same ORF is that expression is limited to a 1:1 ratio.

Thus there is a need for alternative methods for expressing more than one protein from a single vector which are not associated with the disadvantages described above.

#### SUMMARY OF ASPECTS OF THE INVENTION

The present invention is based on the determination that, when a transmembrane protein is co-expressed with a second protein as a polyprotein which after translation is subsequently cleaved to separate both proteins, the incorporation of an intracellular retention signal in the transmembrane protein allows the cell surface expression of the transmembrane protein to be modulated relative to the second protein by reducing its trafficking to the cell surface and/or by reducing its half-life at the cell surface. This need not be limited to a pair of transgenes, but may be used to allow control of the relative expression of multiple proteins initially translated as a polyprotein.

As used herein, 'polyprotein' refers to a polypeptide sequence translated from a single nucleic acid construct as a single entity, but which comprises polypeptide sequences which are subsequently separated and which function as discrete entities (e.g. separate proteins).

Thus in a first aspect the present invention provides a nucleic acid construct comprising the following structure:

A – X – B

in which

A and B are nucleic acid sequences encoding a first and a second protein of interest (POI); and

X is a nucleic acid sequence which encodes a cleavage site,  
wherein either the first or second POI is a transmembrane protein which comprises an intracellular retention signal.

The intracellular retention signal is heterologous to the transmembrane protein i.e. it does not occur in the "wild-type" version of the transmembrane protein (or part thereof) and is introduced by recombinant means. The intracellular retention signal may be synthetic or derivable from another organism or another protein.

The endodomain of the transmembrane protein may comprise the intracellular retention signal.

The intracellular retention signal may direct the transmembrane protein away from the secretory pathway and/or to a membrane-bound intracellular compartment such as a lysosomal, endosomal or Golgi compartment.

The intracellular retention signal may, for example, be a tyrosine-based sorting signal, a dileucine-based sorting signal, an acidic cluster signal, a lysosomal avoidance signal, an NPFX'(1,2)D-Type signal, a KDEL, a KXX'X' or a KX'KXX' signal (wherein X' is any amino acid).

The intracellular retention signal may comprise a sequence selected from the group of: NPX'Y, YXX'Z', [DE]X'X'X'L[LI], DX'X'LL, DP[FW], FX'DX'F, NPF, LZX'Z[DE], LLDLL, PWDLW, KDEL, KXX'X' or KX'KXX';  
wherein X' is any amino acid and Z' is an amino acid with a bulky hydrophobic side chain.

The intracellular retention signal may comprise any of the sequences shown in Tables 1 to 5.

The intracellular retention signal may comprise the Tyrosinase-related protein (TYRP)-1 intracellular retention signal. The intracellular retention signal may comprise the TYRP-1 intracellular domain. The intracellular retention signal may comprise the sequence NQPLLTD (SEQ ID No. 35).

The intracellular retention signal may comprise the Adenoviral E3/19K intracellular retention signal. The intracellular retention signal may comprise the E3/19K cytosolic domain. The intracellular retention signal may comprise the sequence KYKSRRSFIDEKKMP (SEQ ID No. 36); or DEKKMP (SEQ ID No. 37).

The intracellular retention signal may be proximal or distal to a transmembrane domain of the transmembrane protein.

X may be a nucleic acid sequence encoding a self-cleaving peptide, a furin cleavage site or a Tobacco Etch Virus cleavage site.

X may be a nucleic acid sequence encoding a 2A self-cleaving peptide from an aphtho- or a cardiovirus or a 2A-like peptide.

The transmembrane protein may be any transgenically expressed transmembrane protein.

The transmembrane protein may be selected from a list of: excitatory receptors such as 41BB, OX40, CD27, CD28 and related molecules; or inhibitory receptors such as PD1, CTLA4, LAIR1, CD22 and related molecules; or cytokine receptor molecules such as IL1R, IL2R, IL7R, IL15R and related molecules; or homing molecules such as N-CAM, V-CAM, L1-CAM, LFA-1, CDH1-3, Selectins or Integrins;

The transmembrane protein may be a synthetic protein such as a suicide gene or a marker gene.

The transmembrane protein may be or comprise the  $\alpha$  and/or  $\beta$  chains of a T-cell receptor.

The transmembrane protein may be a chimeric-antigen receptor (CAR).

Either or both of the first and second POIs may be a transmembrane protein; and either or both of the transmembrane proteins may comprise an intracellular retention signal as defined herein.

Either or both of the first and second POIs may be a single pass transmembrane protein, such a type I transmembrane protein.

The amount of a transmembrane protein which comprises an intracellular retention signal which is expressed at the cell surface may be, for example, less than 90%, 70%, 50% or 30% compared to a transmembrane protein expressed from the same nucleic acid construct which does not comprise an intracellular retention signal.

The invention also provides nucleic acid construct encoding more than two proteins of interest. For example, in one embodiment, the invention provides a nucleic acid construct comprising the following structure :

A – X – B – Y - C

in which

A, B and C are nucleic acid sequences encoding first, second and third polypeptides of interest (POIs); and

X and Y are nucleic acid sequences which may be the same or different, each of which encodes a cleavage site,

wherein at least two of the POIs are transmembrane proteins which comprise an intracellular retention signal.

In this embodiment, the at least two POIs may:

- (a) comprise different intracellular retention signals; and/or
- (b) have the intracellular retention signal located at a different position in the POI, such that when the nucleic acid is expressed in a cell, there is differential relative expression of the at least two POIs at the cell surface.

In a second aspect the present invention provides a vector comprising a nucleic acid construct according to the first aspect of the invention.

The vector may be a retroviral vector or a lentiviral vector or a transposon.

In a third aspect the present invention provides a cell comprising a nucleic acid construct according to the first aspect of the invention or a vector according to the second aspect of the invention.

The invention further relates to a method for modulating the relative cell surface expression of a transmembrane protein expressed as a polyprotein from a single nucleic acid with a second protein by including an intracellular retention signal in the nucleic acid sequence which encodes the transmembrane protein.

The inclusion of an intracellular retention signal in a transmembrane protein reduces the amount of the transmembrane protein expressed on the cell surface. As such, the relative expression level of a transmembrane protein derived from a polyprotein including a second polypeptide can be modulated. Where the transmembrane protein is only active at the cell surface (or predominantly active at the cell surface), reducing the relative cell surface expression of the protein also reduces its relative activity.

This invention can be extended to modulate the relative expression of three or more proteins expressed as a concatenated polypeptide, separated by cleavage sites and relative surface expression dictated by retention signals of differing activity.

#### DESCRIPTION OF THE FIGURES

**Figure 1:** Methods utilised to express different proteins from the same vector

(a) Two different promoters within the same cassette result in two different transcripts which each give rise to separate proteins. (b) Use of an Internal Ribosome Entry sequence (IRES) leads to a single transcript which is translated into two separate proteins. (c) Use of the FMDV 2A peptide results in a single transcript, and a single polyprotein which rapidly cleaves into two separate proteins.

**Figure 2:** TYRP1 endodomain is able to direct the retention of a transmembrane protein with a complex endodomain

TyRP1 is a type I transmembrane protein, 537aa long. The di-leucine motif, which retains the protein in the intracellular compartment, is indicated as a black rectangle on the cytoplasmic domain. (A) TyRP1 (wt). Wild type TyRP1 consists of a peptide signal, a luminal domain, a transmembrane domain, and a cytoplasmic domain. The cytoplasmic domain contains the di-leucine retention signal. (B) TyRP1 (wt)-SG Linker- eGFP. This construct contains the wild type TyRP1 simply fused to eGFP via a serine-glycine-glycine-glycine-serine linker. The TyRP1-L-eGFP represents the cytoplasmic- proximal TyRP1. (C) TyRP1 Luminal (LM)- Transmembrane (TM)- SG Linker- eGFP- TyRP1 Cytoplasmic (CP). This construct constitutes the cytoplasmic- distal TyRP1, since SG linker- eGFP interposes between the transmembrane and cytoplasmic domains. D: TyRP1 Luminal (LM)- Transmembrane (TM)- SG Linker- eGFP. This construct serves as the positive control, as the cytoplasmic domain containing the retention signal has been excluded. All constructs are co-expressed with IRES.CD34. Staining of transduced SupT1 cells is shown with intracellular and surface staining bottom left / right respectively.

**Figure 3:** Functionality of the TYRP1 retention signal in primary cells

A construct was generated which co-expresses an anti-CD19 and an anti-CD33 CAR using a FMD-2A like peptide. Two variants of this construct were also generated: in the first variant, the di-leucine motif from TYRP1 was inserted into the anti-CD19 CAR endodomain just proximal to the TM domain; in the second variant the same TYRP1 di-leucine motif was attached to the carboxy-terminus of the anti-CD19 CAR endodomain. PBMCs were isolated from blood and stimulated using PHA and IL-2. Two days later the cells were transduced on retronectin coated plates with retro virus containing the different CD19:CD33 CAR constructs. On day 5 the expression level of the two CARs translated by the construct was evaluated via flow cytometry using recombinant CD19-Fc and CD33-Fc fusions. A. Shows cartoon of the synthetic gene constructed to allow co-expression; B. Shows a cartoon of the subsequent pairs of proteins generated by the three constructs; C. Shows expression of the

two receptors by flow-cytometry. In the original construct, both CARs are equally expressed. With incorporation of the di-leucine motif distally in the endodomain of the anti-CD19 CAR, the CD33 CAR expression remains constant but the CD19 expression drops to intermediate levels. With incorporation of the di-leucine motif proximally in the endodomain of the anti-CD19 CAR, the CD33 CAR expression remains constant, but the CD19 expression drops to low levels.

**Figure 4:** Retention signal from cytosolic tail of E3/19K

A construct was generated which co-expresses an anti-CD19 and an anti-CD33 CAR using a FMD-2A like peptide. Two variants of this construct were also generated: in the first variant, the last 6aa from E3/19K (DEKKMP), which were found to be critical for its Golgi/ER retention ability, were attached to the carboxy-terminus of the anti-CD33 CAR endodomain; in the second variant, the entire cytosolic tail of adenovirus E3/19K protein was attached to the carboxy-terminus of the anti-CD33 CAR endodomain

**Figure 5:** Functionality of E3/19K retention signal

The constructs shown in Figure 4 were transfected into 293T cells and the expression level of the two CARs translated by the construct was evaluated via flow cytometry using recombinant CD19-Fc and CD33-Fc fusions. A clear retention was observed when the full length adenovirus E3/19K protein, or the DEKKMP motif was placed on the anti-CD33 receptor. The anti-CD19 receptor expression levels were unaffected.

#### DETAILED DESCRIPTION

The present invention provides a nucleic acid construct comprising the following structure:

A – X – B

in which;

A and B are nucleic acid sequences encoding a first and a second polypeptide of interest (POI); and

X is a nucleic acid sequence which encodes a cleavage site,

wherein either the first or second POI is a transmembrane protein which comprises an intracellular retention signal.

#### TRANSMEMBRANE PROTEIN

The present invention enables modulation of the relative expression of a transmembrane surface protein. The transmembrane surface protein is a protein which, in the absence of an

intracellular retention signal, is expressed at the cell surface. When expressed at the cell surface at least one domain of the transmembrane protein is exoplasmic (i.e. on the exterior of the cell).

The transmembrane protein may be a single-pass transmembrane protein, i.e. it may comprise a single transmembrane domain or it may comprise multiple transmembrane domains.

Transmembrane proteins may be classified by topology i.e. with reference to the position of the N- and C-terminal domains. Types I, II, and III transmembrane proteins are single-pass molecules, while type IV trans-membrane proteins are multiple-pass molecules. Type I transmembrane proteins are anchored to the lipid membrane with a stop-transfer anchor sequence and have their N-terminal domains targeted to the ER lumen during synthesis (and the extracellular space, when the mature form is located on the plasma membrane). Type II and III are anchored with a signal-anchor sequence, with type II being targeted to the ER lumen with its C-terminal domain, while type III have their N-terminal domains targeted to the ER lumen. Type IV is subdivided into IV-A, with their N-terminal domains targeted to the cytosol and IV-B, with an N-terminal domain targeted to the lumen.

The transmembrane protein(s) of the present invention may be any of the types I-IV.

The transmembrane domain may be any protein structure which is thermodynamically stable in a membrane. This is typically an alpha helix comprising of several hydrophobic residues. The transmembrane domain of any transmembrane protein can be used to supply the transmembrane portion. The presence and span of a transmembrane domain of a protein can be determined by those skilled in the art using the TMHMM algorithm (<http://www.cbs.dtu.dk/services/TMHMM-2.0/>). Further, given that the transmembrane domain of a protein is a relatively simple structure, i.e a polypeptide sequence predicted to form a hydrophobic alpha helix of sufficient length to span the membrane, an artificially designed TM domain may also be used (US7052906 B1 describes synthetic transmembrane components).

The transmembrane domain may be derived from CD28, which gives good stability.

The structure and processing of Type I transmembrane proteins is well known in the art. Such proteins typically comprise an extracellular domain, a transmembrane domain and an

intracellular endodomain and are single-pass molecules with a single  $\alpha$ -helix passing through the cell membrane.

Type I transmembrane proteins typically have a signal peptide which is quickly recognized by the endoplasmic reticulum (ER) and the protein in translation is therefore quickly redirected into the ER. A hydrophobic helix locks then anchors the protein in the membrane of the ER.

As mentioned above, Type I transmembrane proteins are anchored to the lipid membrane with a stop-transfer anchor sequence. The stop-transfer sequence halts the further translocation of the polypeptide and acts as a transmembrane anchor.

As used herein, the term Type I transmembrane protein encompasses any protein which comprises a Type I transmembrane domain and a stop-transfer anchor sequence and is, in the absence of an exogenous intracellular retention signal, targeted for expression on the cell surface.

Various type 1 transmembrane proteins which are suitable for use in the present invention are known in the art. Such proteins include, but are not limited to inhibitory receptors, stimulatory receptors, cytokine receptors and G-Proteins.

The transmembrane protein(s) may be a T-cell receptor  $\alpha$  or  $\beta$  chain.

The transmembrane protein(s) may be a Chimeric Antigen Receptor (CAR).

CARs are proteins which graft an antigen binding domain to the effector function of a T-cell. Their usual form is that of a type I transmembrane domain protein with an antigen recognizing amino terminus, a spacer, a transmembrane domain all connected to a compound endodomain which transmits T-cell survival and activation signals.

The antigen binding domain may be derived from an antibody or antibody mimetic, or it may be another entity which specifically binds the antigen, such as a ligand.

The most common form of these molecules are fusions of single-chain variable fragments (scFv) derived from monoclonal antibodies which recognize a target antigen, fused via a spacer and a trans-membrane domain to a signaling endodomain. Such molecules result in activation of the T-cell in response to recognition by the scFv of its target. When T cells

express such a CAR, they recognize and kill target cells that express the target antigen. Several CARs have been developed against tumour associated antigens, and adoptive transfer approaches using such CAR-expressing T cells are currently in clinical trial for the treatment of various cancers.

It is also possible for the signalling endodomain to be present on a separate molecule. The term "CAR" in connection with the present invention also encompasses a molecule which comprises an antigen binding domain connected to a transmembrane domain. Such a CAR may be capable of interacting with an intracellular signalling domain in order to stimulate T-cell activation.

In the present invention, either of the nucleic acid sequences A or B may be a nucleic acid sequence which encodes a transmembrane protein comprising an intracellular retention signal.

Most transmembrane proteins of interest are only active, or are predominantly active when at the cell membrane. Therefore causing a proportion of the protein to be retained intracellularly reduces the relative expression of the protein at the cell surface and therefore reduces the relative activity of the protein.

#### SIGNAL SEQUENCE

The transmembrane protein may also comprise a signal sequence so that when the transmembrane protein is expressed inside a cell the nascent protein is directed to the endoplasmic reticulum (ER).

The core of the signal peptide may contain a long stretch of hydrophobic amino acids that has a tendency to form a single alpha-helix. The signal peptide may begin with a short positively charged stretch of amino acids, which helps to enforce proper topology of the polypeptide during translocation. At the end of the signal peptide there is typically a stretch of amino acids that is recognized and cleaved by signal peptidase. Signal peptidase may cleave either during or after completion of translocation to generate a free signal peptide and a mature protein. The free signal peptides are then digested by specific proteases.

The signal peptide may be at the amino terminus of the molecule.

#### CLEAVAGE SITE

The present nucleic acid construct comprises a sequence encoding a cleavage site positioned between nucleic acid sequences which encode first and second polypeptides, such that first and second polypeptides can be expressed as separate entities.

The cleavage site may be any sequence which enables the polypeptide comprising the first and second POIs to become separated.

The term "cleavage" is used herein for convenience, but the cleavage site may cause the first and second POIs to separate into individual entities by a mechanism other than classical cleavage. For example, for the Foot-and-Mouth disease virus (FMDV) 2A self-cleaving peptide (see below), various models have been proposed for to account for the "cleavage" activity: proteolysis by a host-cell proteinase, autoproteolysis or a translational effect (Donnelly et al (2001) *J. Gen. Virol.* 82:1027-1041). The exact mechanism of such "cleavage" is not important for the purposes of the present invention, as long as the cleavage site, when positioned between nucleic acid sequences which encode first and second polypeptides, causes the first and second polypeptides to be expressed as separate entities.

The cleavage site may be a furin cleavage site.

Furin is an enzyme which belongs to the subtilisin-like proprotein convertase family. The members of this family are proprotein convertases that process latent precursor proteins into their biologically active products. Furin is a calcium-dependent serine endoprotease that can efficiently cleave precursor proteins at their paired basic amino acid processing sites. Examples of furin substrates include proparathyroid hormone, transforming growth factor beta 1 precursor, proalbumin, pro-beta-secretase, membrane type-1 matrix metalloproteinase, beta subunit of pro-nerve growth factor and von Willebrand factor. Furin cleaves proteins just downstream of a basic amino acid target sequence (canonically, Arg-X-(Arg/Lys)-Arg') and is enriched in the Golgi apparatus.

The cleavage site may be a Tobacco Etch Virus (TEV) cleavage site.

TEV protease is a highly sequence-specific cysteine protease which is chymotrypsin-like proteases. It is very specific for its target cleavage site and is therefore frequently used for the controlled cleavage of fusion proteins both *in vitro* and *in vivo*. The consensus TEV cleavage site is ENLYFQ\S (where '\' denotes the cleaved peptide bond). Mammalian cells, such as human cells, do not express TEV protease. Thus in embodiments in which the

present nucleic acid construct comprises a TEV cleavage site and is expressed in a mammalian cell – exogenous TEV protease must also be expressed in the mammalian cell.

The cleavage site may encode a self-cleaving peptide.

A 'self-cleaving peptide' refers to a peptide which functions such that when the polypeptide comprising the first and second POIs and the self-cleaving peptide is produced, it is immediately "cleaved" or separated into distinct and discrete first and second polypeptides without the need for any external cleavage activity.

The self-cleaving peptide may be a 2A self-cleaving peptide from an aphtho- or a cardiovirus. The primary 2A/2B cleavage of the aphtho- and cardioviruses is mediated by 2A "cleaving" at its own C-terminus. In aphthoviruses, such as foot-and-mouth disease viruses (FMDV) and equine rhinitis A virus, the 2A region is a short section of about 18 amino acids, which, together with the N-terminal residue of protein 2B (a conserved proline residue) represents an autonomous element capable of mediating "cleavage" at its own C-terminus.

The C-terminal 19 amino acids of the longer cardiovirus protein, together with the N-terminal proline of 2B mediate "cleavage" with an efficiency approximately equal to the aphthovirus FMDV 2a sequence. Cardioviruses include encephalomyocarditis virus (EMCV) and Theiler's murine encephalitis virus (TMEV).

Mutational analysis of EMCV and FMDV 2A has revealed that the motif DxExNPGP is intimately involved in "cleavage" activity (Donnelly et al (2001) as above).

The cleavage site of the present invention may comprise the amino acid sequence: Dx<sub>1</sub>Ex<sub>2</sub>NPGP, where x<sub>1</sub> and x<sub>2</sub> are any amino acid. X<sub>1</sub> may be selected from the following group: I, V, M and S. X<sub>2</sub> may be selected from the following group: T, M, S, L, E, Q and F.

For example, the cleavage site may comprise one of the amino acid sequences shown in Table 1.

Table 1

Motif	Present in:	
DIETNPGP (SEQ ID No. 1)	Picornaviruses	EMCB,

	EMCD, EMCPV21
DVETNPGP (SEQ ID No. 2)	Picornaviruses MENG0 and TMEBEAN; Insect virus DCV, ABPV
DVEMNPGP (SEQ ID No. 3)	Picornaviruses TMEGD7 and TMEBEAN
DVESNPGP (SEQ ID No. 4)	Picornaviruses FMDA10, FMDA12, FMDC1, FMD01K, FMDSAT3, FMDVSAT2, ERAV; Insect virus CrPV
DMESNPGP (SEQ ID No. 5)	Picornavirus FMDV01G
DVELNPGP (SEQ ID No. 6)	Picornavirus ERBV; Porcine rotavirus
DVEENPGP (SEQ ID No. 7)	Picornavirus PTV-1; Insect virus TaV; Trypanosoma TSR1
DIELNPGP (SEQ ID No. 8)	Bovine Rotavirus, human rotavirus
DIEQNPGP (SEQ ID No. 9)	Trypanosoma AP endonuclease
DSEFNPGP (SEQ ID No. 10)	Bacterial sequence <i>T. maritima</i>

The cleavage site, based on a 2A sequence may be, for example 15-22 amino acids in length. The sequence may comprise the C-terminus of a 2A protein, followed by a proline residue (which corresponds to the N-terminal proline of 2B).

Mutational studies have also shown that, in addition to the naturally occurring 2A sequences, some variants are also active. The cleavage site may correspond to a variant sequence from a naturally occurring 2A polypeptide, have one, two or three amino acid substitutions, which retains the capacity to induce the “cleavage” of a polyprotein sequence into two or more separate proteins.

The cleavage sequence may be selected from the following which have all been shown to be active to a certain extent (Donnelly *et al* (2001) as above):

LLNFDLLKLAGDVESNPGP (SEQ ID No. 11)  
LLNFDLLKLAGDVQSNPGP (SEQ ID No. 12)  
LLNFDLLKLAGDVEINPGP (SEQ ID No. 13)  
LLNFDLLKLAGDVEFNPGP (SEQ ID No. 14)  
LLNFDLLKLAGDVESHGP (SEQ ID No. 15)  
LLNFDLLKLAGDVESEPGP (SEQ ID No. 16)  
LLNFDLLKLAGDVESQPGP (SEQ ID No. 17)  
LLNFDLLKLAGDVESNPGG (SEQ ID No. 18)

Based on the sequence of the DxExNPGP “a motif, “2A-like” sequences have been found in picornaviruses other than apho- or cardioviruses, ‘picornavirus-like’ insect viruses, type C rotaviruses and repeated sequences within *Trypanosoma spp* and a bacterial sequence (Donnelly et al (2001) as above). The cleavage site may comprise one of these 2A-like sequences, such as:

YHADYYKQRLIHDVEMNPGP (SEQ ID No. 19)  
HYAGYFADLLIHDIETNPGP (SEQ ID No. 20)  
QCTNYALLKLAGDVESNPGP (SEQ ID No. 21)  
ATNFSLLKQAGDVEENPGP (SEQ ID No. 22)  
AARQMLLLLSGDVETNPGP (SEQ ID No. 23)  
RAEGRGSLLTCGDVEENPGP (SEQ ID No. 24)  
TRAIEDELIRAGIESNPGP (SEQ ID No. 25)  
TRAIEDELIRADIESNPGP (SEQ ID No. 26)  
AKFQIDKILISGDVELNPGP (SEQ ID No. 27)  
SSIIRTKMLVSGDVEENPGP (SEQ ID No. 28)  
CDAQRQKLLLSDIEQNPGP (SEQ ID No. 29)  
YPIDFGGFLVKADSEFNPGP (SEQ ID No. 30)

The cleavage site may comprise the 2A-like sequence shown as SEQ ID No. 24 (RAEGRGSLLTCGDVEENPGP).

It has been shown that including an N-terminal “extension” of between 5 and 39 amino acids can increase activity (Donnelly et al (2001) as above). In particular, the cleavage sequence may comprise one of the following sequences or a variant thereof having, for example, up to 5 amino acid changes which retains cleavage site activity:

VTELLYRMKRAETYCPRPLAIHPTEARHKQKIVAPVKQTLNFDLLKLAGDVESNPGP (SEQ ID No. 31)

LLAIHPTEARHKQKIVAPVKQTLNFDLLKLAGDVESNPGP (SEQ ID No. 32)

EARHKQKIVAPVKQTLNFDLLKLAGDVESNPGP (SEQ ID No. 33)

APVKQTLNFDLLKLAGDVESNPGP (SEQ ID No. 34)

#### INTRACELLULAR RETENTION SIGNAL

The nucleic acid construct of the present invention comprises a sequence which encodes a transmembrane protein comprising an intracellular retention signal.

Protein targeting or protein sorting is the biological mechanism by which proteins are transported to the appropriate destinations in the cell or outside of it. Proteins can be targeted to the inner space of an organelle, different intracellular membranes, plasma membrane, or to exterior of the cell via secretion. This delivery process is carried out based on sequence information contained in the protein itself.

Proteins synthesised in the rough endoplasmic reticulum (ER) of eukaryotic cells use the exocytic pathway for transport to their final destinations. Proteins lacking special sorting signals are vectorially transported from the ER via the Golgi and the trans-Golgi network (TGN) to the plasma membrane. Other proteins have targeting signals for incorporation into specific organelles of the exocytic pathway, such as endosomes and lysosomes.

Lysosomes are acidic organelles in which endogenous and internalised macromolecules are degraded by luminal hydrolases. Endogenous macromolecules reach the lysosome by being sorted in the TGN from which they are transported to endosomes and then lysosomes.

The targeting signals used by a cell to sort proteins to the correct intracellular location may be exploited by the present invention. The signals may be broadly classed into the following types:

- i) endocytosis signals
- ii) Golgi retention signals
- iii) TGN recycling signals
- iv) ER retention signals
- v) lysosomal sorting signals

'Intracellular retention signal' refers to an amino acid sequence which directs the protein in which it is encompassed to a cellular compartment other than the cell surface membrane or to the exterior of the cell.

The intracellular retention signal causes a reduction in the amount of the transmembrane protein expressed on the surface of a cell compared to an equivalent, control transmembrane protein which does not comprise an intracellular retention signal.

In other words, the proportion of translated transmembrane protein comprising an intracellular retention signal which is expressed on at the cell surface is less than the proportion of an equivalent amount of an equivalent, translated control transmembrane protein which does not comprise an intracellular retention signal.

For example, the amount of the transmembrane protein comprising an intracellular retention signal which is expressed on the surface of a cell may be less than 75%, less than 50%, less than 25% or less than 10% of the amount of an equivalent control transmembrane protein which does not comprise an intracellular retention signal.

Constructs which express a polyprotein that is subsequently cleaved by a protease are generally limited by the fact the expression of the peptides from the polyprotein is limited to a 1:1 ratio. However, in the present invention, the inclusion of an intracellular retention signal in the transmembrane protein means that its expression on the cell surface can be modulated (e.g. reduced compared to an equivalent control transmembrane protein which does not comprise an intracellular retention signal). As such the ratio of the transmembrane protein which comprises the intracellular retention signal expressed on the cell surface compared to the expression of the second protein expressed in the polyprotein may be, for example about 1:1.5, of from 1:1.5-1:2, 1:2-1:3, 1:3-1:4, 1:4-1:5, or more than 1:5.

The amount of a transmembrane protein expressed on the surface of a cell may be determined using methods which are known in the art, for example flow cytometry or fluorescence microscopy.

The intracellular retention signal may direct the transmembrane protein away from the secretory pathway during translocation from the ER.

The intracellular retention signal may direct the transmembrane protein to an intracellular compartment or complex. The intracellular retention signal may direct the transmembrane protein to a membrane-bound intracellular compartment.

For example, the intracellular retention signal may direct the protein to a lysosomal, endosomal or Golgi compartment (trans-Golgi Network, 'TGN').

Within a normal cell, proteins arising from biogenesis or the endocytic pathway are sorted into the appropriate intracellular compartment following a sequential set of sorting decisions. At the plasma membrane, proteins can either remain at the cell surface or be internalised into endosomes. At the TGN, the choice is between going to the plasma membrane or being diverted to endosomes. In endosomes, proteins can either recycle to the plasma membrane or go to lysosomes. These decisions are governed by sorting signals on the proteins themselves.

Lysosomes are cellular organelles that contain acid hydrolase enzymes that break down waste materials and cellular debris. The membrane around a lysosome allows the digestive enzymes to work at the pH they require. Lysosomes fuse with autophagic vacuoles (phagosomes) and dispense their enzymes into the autophagic vacuoles, digesting their contents.

An endosome is a membrane-bounded compartment inside eukaryotic cells. It is a compartment of the endocytic membrane transport pathway from the plasma membrane to the lysosome and provides an environment for material to be sorted before it reaches the degradative lysosome. Endosomes may be classified as early endosomes, late endosomes, or recycling endosomes depending on the time it takes for endocytosed material to reach them. The intracellular retention signal used in the present invention may direct the protein to a late endosomal compartment.

The Golgi apparatus is part of the cellular endomembrane system, the Golgi apparatus packages proteins inside the cell before they are sent to their destination; it is particularly important in the processing of proteins for secretion.

There is a considerable body of knowledge which has arisen from studies investigating the sorting signals present in known proteins, and the effect of altering their sequence and/or position within the molecule (Bonifacino and Traub (2003) *Ann. Rev. Biochem.* 72:395-447; Braulke and Bonifacino (2009) *Biochimica and Biophysica Acta* 1793:605-614; Griffith (2001)

Current Biology 11:R226-R228; Mellman and Nelson (2008) Nat Rev Mol Cell Biol. 9:833-845; Dell'Angelica and Payne (2001) Cell 106:395-398; Schafer *et al* (1995) EMBO J. 14:2424-2435; Trejo (2005) Mol. Pharmacol. 67:1388-1390). Numerous studies have shown that it is possible to insert one or more sorting signals into a protein of interest in order to alter the intracellular location of a protein of interest (Pelham (2000) Meth. Enzymol. 327:279-283).

It is therefore perfectly possible to select a sorting signal having a desired localisation property and include it within a protein of interest in order to direct the intracellular location of that protein. In connection with the present application, it is therefore possible to select a sorting signal having the desired amount of reduction of expression at the plasma membrane.

The optimal position of the sorting signal in the nascent protein of interest may depend on the type of transmembrane protein (i.e. types I-IV) and whether the C-terminus is on the luminal or the cytoplasmic side of the membrane (Goder and Spiess (2001) FEBS Lett 504:87-93). This may readily be determined by considering the position of the sorting signal in its natural protein.

Examples of endocytosis signals include those from the transferrin receptor and the asialoglycoprotein receptor.

Examples of signals which cause TGN-endosome recycling include those from proteins such as the Cl- and CD-MPRs, sortilin, the LDL-receptor related proteins LRP3 and LRP10 and  $\beta$ -secretase, GGA1-3, LIMP-II, NCP1, mucolipin-1, sialin, GLUT8 and invariant chain.

Examples of TGN retention signals include those from the following proteins which are localized to the TGN: the prohormone processing enzymes furin, PC7, CPD and PAM; the glycoprotein E of herpes virus 3 and TGN38.

Examples of ER retention signals include C-terminal signals such as KDEL, KKXX or KXKXX and the RXR(R) motif of potassium channels. Known ER proteins include the adenovirus E19 protein and ERGIC53.

Examples of lysosomal sorting signals include those found in lysosomal membrane proteins, such as LAMP-1 and LAMP-2, CD63, CD68, endolyn, DC-LAMP, cystinosin, sugar phosphate exchanger 2 and acid phosphatase.

The intracellular retention signal may be from the adenovirus E19 protein. The intracellular retention signal may be from the protein E3/19K, which is also known as E3gp 19 kDa; E19 or GP19K. The intracellular retention signal may comprise the full cytosolic tail of E3/19K, which is shown as SEQ ID No. 36; or the last 6 amino acids of this tail, which is shown as SEQ ID No. 37. The present inventors have shown that the last 6 amino acids are particularly important for retention (Example 3 and Figure 5)

SEQ ID No. 36: KYKSRRSFIDEKKMP

SEQ ID No. 37: DEKKMP

#### TUNABILITY

The relative expression of one or more protein(s) may be fine tuned using the method of the invention by various methods, such as

- a) altering the position of the intracellular retention signal in the protein molecule; and/or
- b) selecting a particular intracellular retention signal.

Option a) is discussed in more detail below.

With regard to option b), a range of intracellular retention signals is available from the large number of naturally occurring proteins which are sorted to distinct cellular locations inside eukaryotic cells. It is also possible to use “synthetic” intracellular retention signals which comprise one of more of the motifs found in naturally occurring proteins (see next section) and have a similar sorting signal function.

A cascade of signal strength is available, depending on the intracellular location to which the sorting signal sends the relevant protein. Broadly speaking, the more “intracellular” the location directed by the sorting signal, the “stronger” the signal is in terms of lowering the relative expression of the protein.

When a sorting signal directs a protein to the lysosomal compartment, the protein is internalised and degraded by the cell, resulting in relatively little escape to the cell surface. The protein is degraded and lost from the system once it enters the lysosome. Therefore lysosomal sorting signals, such as LAMP1, are the “strongest” in terms of reducing relative expression at the cell surface.

When a sorting signal directs a protein to be retained in the ER, only a very small proportion of the protein gets to the cell surface. Hence ER retention or recycling signals, such as ER-GIC-53 and KKFF signal are the next most strong, in terms of reducing relative expression at the cell surface.

When a sorting signal directs a protein to the endosomal, Golgi or TGN compartments, then the protein is likely to recycle to some extent between the TGN, the endosomal compartment, and the plasma membrane. These signals provide a more limited level of reduction of expression as a significant proportion of the protein will still reach the plasma membrane.

In general the reduction in expression seen with known sorting signals can be summarised as follows:

Lysosomal sorting signals>ER retention/recycling signals>TGN retention/recycling signals>endocytosis signals.

The tunability using different sorting signals and/or different positions of sorting signals within the protein is especially useful when one considers the expression of multiple proteins, each with their own relative expression. For example, consider a nucleic acid construct having the following structure:

A – X – B – Y - C

in which

A, B and C are nucleic acid sequences encoding polypeptides; and  
X and Y are nucleic acid sequences encodes cleavage sites.

The nucleic acid construct will encode three proteins *A*, *B* and *C*, any or all of which may be transmembrane proteins. For example, *B* and *C* may be transmembrane proteins which comprise an intracellular retention signal. If it is desired for *A*, *B* and *C* to be expressed such that the relative levels are *A>B>C*, then the nucleic acid sequence *A* may have no intracellular retention signal, the nucleic acid sequence *B* may have an intracellular retention signal that causes a small proportion of protein *B* to be retained in the cell (i.e. not to be expressed at the cell surface), and the nucleic acid sequence *C* may have an intracellular retention signal that causes a large proportion of protein *C* to be retained in the cell.

As explained below, differential amounts of intracellular retention, leading to different amounts of cell surface expression may be achieved by:

- (a) using different intracellular retention signals for the proteins; and/or
- (b) having the intracellular retention signal located at a different position in the proteins.

#### SIGNAL TYPES

Numerous proteins which include an intracellular retention signal and are directed to an intracellular compartment are known in the art.

The intracellular retention signal may be a retention signal from a protein which resides in the lysosomal, endosomal or Golgi compartment.

Intracellular retention signals are well known in the art (see, for example, Bonifacino & Traub; *Annu. Rev. Biochem.*; 2003; 72; 395-447).

The intracellular retention signal may be a tyrosine-based sorting signal, a dileucine-based sorting signal, an acidic cluster signal, a lysosomal avoidance signal, an NPFX'(1,2)D-Type signal, a KDEL, a KKXX' or a KX'KXX' signal (wherein X' is any amino acid).

Tyrosine-based sorting signals mediate rapid internalization of transmembrane proteins from the plasma membrane and the targeting of proteins to lysosomes (Bonifacino & Traub; as above). Two types of tyrosine-based sorting signals are represented by the NPX'Y and YXX'Z' consensus motifs (wherein Z' is an amino acid with a bulky hydrophobic side chain).

NPX'Y signals have been shown to mediate rapid internalization of type I transmembrane proteins, they occur in families such as members of the LDL receptor, integrin  $\beta$ , and  $\beta$ -amyloid precursor protein families.

Examples of NPX'Y signals are provided in Table 2.

**Table 2 – NPX'Y signals**

Protein	Species	Sequence
LDL receptor	Human	Tm-10- <b>INFDNPVYQKTT</b> -29
LRP1 (1)	Human	Tm-21- <b>VEIGNPPTYKMYE</b> -64
LRP1 (2)	Human	Tm-55- <b>TNFTNPVYATLY</b> -33
LRP1	<i>Drosophila</i>	Tm-43- <b>GNFANPVYESMY</b> -38
LRP1 (1)	<i>C. elegans</i>	Tm-54- <b>TTFTNPVYLED</b> -91

LRP1 (2)	<i>C. elegans</i>	Tm-140-LRVD <b>NPLYDPDS</b> -4
Megalin (1)	Human	Tm-70-IIFEN <b>NPMYSARD</b> -125
Megalin (2)	Human	Tm-144-TNFEN <b>NPIYAQME</b> -53
Integrin $\beta$ -1 (1)	Human	Tm-18-DTGEN <b>PIYKSAV</b> -11
Integrin $\beta$ -1 (2)	Human	Tm-30-TTVV <b>NPKYEGK</b>
Integrin $\beta$ (1)	<i>Drosophila</i>	Tm-26-WDTEN <b>PIYKQAT</b> -11
Integrin $\beta$ (2)	<i>Drosophila</i>	Tm-35-STFK <b>NPMYAGK</b>
APLPI	Human	Tm-33-HGYEN <b>PTYRFLE</b> -3
APP	Human	Tm-32-NGYEN <b>PTYKFFE</b> -4
APP-like	<i>Drosophila</i>	Tm-38-NGYEN <b>PTYKYFE</b> -3
Insulin receptor	Human	Tm-36-YASS <b>NPEYLSAS</b> -379
EGR receptor (1)	Human	Tm-434-GSVQ <b>NPVYHNQP</b> -96
EGR receptor (2)	Human	Tm-462-TAVG <b>NPEYLNTV</b> -68
EGR receptor (3)	Human	Tm-496-ISLD <b>NPDYQQDF</b> -34

Numbers in parentheses indicate motifs that are present in more than one copy within the same protein. The signals in this and other tables should be considered examples. Key residues are indicated in bold type. Numbers of amino acids before (i.e., amino-terminal) and after (i.e., carboxy-terminal) the signals are indicated. Abbreviations: Tm, transmembrane; LDL, low density lipoprotein; LRP1, LDL receptor related protein 1; APP,  $\beta$ -amyloid precursor protein; APLPI, APP-like protein 1.

YX'X'Z'-type signals are found in endocytic receptors such as the transferrin receptor and the asialoglycoprotein receptor, intracellular sorting receptors such as the Cl- and CD-MPRs, lysosomal membrane proteins such as LAMP-1 and LAMP-2, and TGN proteins such as TGN38 and furin, as well as in proteins localized to specialized endosomal-lysosomal organelles such as antigen-processing compartments (e.g., HLA-DM) and cytotoxic granules (e.g., GMP-17). The YX'X'Z'-type signals are involved in the rapid internalization of proteins from the plasma membrane. However, their function is not limited to endocytosis, since the same motifs have been implicated in the targeting of transmembrane proteins to lysosomes and lysosome-related organelles.

Examples of YX'X'Z'-type signals are provided in Table 3.

**Table 3 - YX'X'Z'-type signals**

Protein	Species	Sequence
LAMP-1	Human	Tm--RKRS <del>HAGYQTI</del>
LAMP-2a	Human	Tm--KHHAG <del>YEQF</del>
LAMP-2a	Chicken	Tm--KKHHNT <del>GYEQF</del>
LAMP-2b	Chicken	Tm--RRKSRTG <del>YQSV</del>
LAMP-2c	Chicken	Tm--RRKSY <del>ACYQTL</del>
LAMP	<i>Drosophila</i>	Tm--RRRSTS <del>RGYMSF</del>
LAMP	Earthworm	Tm--RKRS <del>RRGYESV</del>
CD63	Human	Tm--KSIRSG <del>YEV</del> <b>M</b>
GMP-17	Human	Tm--HCGGPRPG <del>YETL</del>
GMP-17	Mouse	Tm--HCRT <del>RRRAEYETL</del>
CD68	Human	Tm--RRRPSA <del>YQAL</del>
CD1b	Human	Tm--RRRS <del>YQNI</del> <b>P</b>
CD1c	Human	Tm--KKHCS <del>YQDIL</del>
CD1d	Mouse	Tm--RRRSAY <del>QDIR</del>
CD1	Rat	Tm--RKRRRS <del>YQDIM</del>
Endolyn	Rat	Tm--KFCKSKERN <del>YHTL</del>
Endolyn	<i>Drosophila</i>	Tm--KFYKARNERN <del>YHTL</del>
TSC403	Human	Tm--KIRLRCQSSG <del>YQRI</del>
TSC403	Mouse	Tm--KIRQRHQSSA <del>YQRI</del>
Cystinosin	Human	Tm--HFCLYRKRP <del>GYDQLN</del>
Putative solute carrier	Human	Tm--12-SLSRGSG <del>YKEI</del>
TRP-2	Human	Tm--RRLRKGY <del>TP</del> LMET-11
HLA-DM ♦	Human	Tm--RRAGHSS <del>YTP</del> LPGS-9
LmpA	Dictyostelium	Tm--KKLRQQKQQG <del>YQAI</del> INNE
Putative lysosomal protein	Dictyostelium	Tm--RSKSNQNQSY <del>NLI</del> QL
LIMP-II	Dictyostelium	Tm--RKT <del>FYNNNQYNGYNI</del> IN
Transferrin receptor	Human	16-PLSY <del>TRF</del> SLA-35-Tm
Asialoglycoprotein receptor H1	Human	MTKEY <del>QDI</del> QHL-29-Tm
CI-MPR	Human	Tm-22-SYK <del>YSKVNKE</del> -132
CD-MPR	Human	Tm-40-PAAY <del>RGV</del> GDD-16
CTLA-4	Human	Tm-10-TGV <del>YVKM</del> PPT-16
Furin	Human	Tm-17-LIS <del>YKGLP</del> PE-29
TGN38	Rat	Tm-23-ASDY <del>QRL</del> NLKL
gp41	HIV-1	Tm-13-RQGY <del>SPL</del> SFQT-144
Acid phosphatase	Human	Tm--RMQAQPPG <del>YRH</del> VADGEDHA

See legend to Table 1 for explanation of signal format

Dileucine-based sorting signals ([DE]X'X'X'LL[LI]) play critical roles in the sorting of many type I, type II, and multispanning transmembrane proteins. Dileucine-based sorting signals are involved in rapid internalization and lysosomal degradation of transmembrane proteins and the targeting of proteins to the late endosomal-lysosomal compartments. Transmembrane proteins that contain constitutively active forms of this signal are mainly localised to the late endosomes and lysosomes.

Examples of [DE]X'X'X'LL[LI] sorting signals are provided in Table 4.

**Table 4 - [DE]X'X'X'LL[LI] sorting signals**

Protein	Species	Signal
CD3-γ	Human	Tm-8-SDKQTLLPN-26
LIMP-II	Rat	Tm-11-DERAPLIRT
Nmb	Human	Tm-37-QEKDPLLKN-7
QNR-71	Quail	Tm-37-TERNPLLKS-5
Pmel17	Human	Tm-33-GENSPLLSG-3
Tyrosinase	Human	Tm-8-EEKQPLLME-12
Tyrosinase	Medaka fish	Tm-16-GERQPLLQS-13
Tyrosinase	Chicken	Tm-8-PEIQPLLTE-13
TRP-1	Goldfish	Tm-7-EGRQPLLGD-15
TRP-1	Human	Tm-7-EANQPLLTD-20
TRP-1	Chicken	Tm-7-ELHQPLLTD-20
TRP-2	Zebrafish	Tm-5-REFEPLLNA-11
VMAT2	Human	Tm-6-EEKMAILMD-29
VMAT1	Human	Tm-6-EEKLAILSQ-32
VAcT	Mouse	Tm-10-SERDVLLDE-42
VAMP4	Human	19-SERRNLLED-88-Tm
Neonatal FeR	Rat	Tm-16-DDSGDLLPG-19
CD4	Human	Tm-12-SQIKRLLSE-17
CD4	Cat	Tm-12-SHIKRLLSE-17
GLUT4	Mouse	Tm-17-RRTPSLLEQ-17
GLUT4	Human	Tm-17-HRTPSLLEQ-17
IRAP	Rat	46-EPRGSRLLVR-53-Tm
II	Human	MDDQRDLISNNEQLPMLGR-11-Tm
II	Mouse	MDDQRDLISNHEQLPILGN-10-Tm
II	Chicken	MAEEQRDLISSLGSSGVLP-12-Tm
II-1	Zebrafish	MEPDHQNESLIQRVPSAETILGR-12-Tm
II-2	Zebrafish	MSSEGNETPLISDQSSVNMGPQP-8-Tm
Lamp	Trypanosome	Tm-RPRRRTEEDELLPEEAEGLIDPQN
Menkes protein	Human	Tm-74-PDKHSLLVGDFREDDDTAL
NPC1	Human	Tm-13-TERERILLNF
AQP4	Human	Tm-32-VETDDLIL-29
RME-2	<i>C. elegans</i>	Tm-104-FENDSLL
Vam3p	<i>S. cerevisiae</i>	153-NEQSPLLHN-121-Tm
ALP	<i>S. cerevisiae</i>	7-SEQTRLVP-18-Tm
Gap1p	<i>S. cerevisiae</i>	Tm-23-EVDLDLLK-24

See legend to Table 1 for explanation of signal format.

DX'X'LL signals constitute a distinct type of dileucine-based sorting signals. These signals are present in several transmembrane receptors and other proteins that cycle between the TGN and endosomes, such as the Cl- and CD-MPRs, sortilin, the LDL-receptor-related proteins LRP3 and LRP10, and β-secretase.

Examples of DX'X'LL sorting signals are provided in Table 5.

**Table 5 - DX'X'LL sorting signals**

Protein	Species	Sequence
CI-MPR	Human	Tm-151- <u>S</u> FHDD <u>S</u> <b>DEDILL</b> HI
CI-MPR	Bovine	Tm-150-TFHDD <u>S</u> <b>DEDILL</b> HV
CI-MPR	Rabbit	Tm-151- <u>S</u> FHDD <u>S</u> <b>DEDILL</b> NI
CI-MPR	Chicken	Tm-148- <u>S</u> FHDD <u>S</u> <b>DEDILL</b> NV
CD-MPR	Human	Tm-54- <u>E</u> ESEER <u>D</u> DH <u>L</u> PM
CD-MPR	Chicken	Tm-54- <u>D</u> ESEER <u>D</u> DH <u>L</u> PM
Sortilin	Human	Tm-41-GYHDD <u>S</u> <b>DEDILLE</b>
SorLA	Human	Tm-41-ITGF <u>S</u> <b>D</b> DVP <u>M</u> VIA
Head-activator BP	Hydra	Tm-41-INRF <u>S</u> <b>D</b> EPI <u>L</u> VVA
LRP3	Human	Tm-237-MLEAS <u>S</u> <b>DEAILLVC</b>
ST7	Human	Tm-330-KNET <u>S</u> <b>D</b> EALLLC
LRP10	Mouse	Tm-235-WVVEAE <u>D</u> EPI <u>L</u> LA
LRP10	Human	Tm-237-WVAEA <u>E</u> D <u>E</u> PI <u>L</u> LT
Beta-secretase	Human	Tm-9-HDDF <u>A</u> D <u>S</u> ILK
Mucolipin-1	Mouse	Tm-43-GR <u>D</u> S <u>P</u> E <u>D</u> H <u>S</u> <u>L</u> LN
Nonclassical MHC-I	Deer mouse	Tm-6-VRCH <u>P</u> E <u>D</u> DR <u>L</u> G
FLJ30532	Human	Tm-83-HRVS <u>S</u> <u>D</u> D <u>L</u> <u>L</u> T <u>S</u>
GGA1	Human	350-ASV <u>S</u> <u>L</u> <u>D</u> E <u>L</u> <u>M</u> <u>S</u> U-275
GGA1	Human	415-ASS <u>G</u> L <u>D</u> D <u>L</u> <u>L</u> GK-211
GGA2	Human	408-VQNP <u>S</u> <u>A</u> DRN <u>L</u> <u>L</u> DL-192
GGA3	Human	384-NAL <u>S</u> WL <u>D</u> E <u>E</u> LL <u>C</u> L-326
GGA	<i>Drosophila</i>	447-TV <u>D</u> S <u>I</u> <u>D</u> D <u>V</u> <u>P</u> <u>L</u> <u>S</u> D-116

See legend to Table 1 for explanation of signal format. Serine and threonine residues are underlined.

Another family of sorting motifs is provided by clusters of acidic residues containing sites for phosphorylation by CKII. This type of motif is often found in transmembrane proteins that are localized to the TGN at steady state, including the prohormone-processing enzymes furin, PC6B, PC7, CPD, and PAM, and the glycoprotein E of herpes virus 3.

Examples of acidic cluster signals are provided in Table 6.

**Table 6 – Acidic cluster sorting signals**

Protein	Species	Sequence
Furin	Mouse	Tm-31- <u>QEECP<u>SD<u>SEE<u>DEG</u>-14</u></u></u>
PC6B (1) <sup>a</sup>	Mouse	Tm-39- <u>RDR<u>DY<u>DED<u>DEDD<u>I</u>-36</u></u></u></u>
PC6B (2)	Mouse	Tm-69- <u>LDE<u>TE<u>ED<u>DE<u>LEY<u>YD<u>DES</u>-4</u></u></u></u></u></u>
PC7	Human	Tm-38- <u>KDP<u>DEV<u>ETE<u>ES</u>-47</u></u></u>
CPD	Human	Tm-36- <u>HE<u>FQ<u>DET<u>TD<u>TE<u>EET</u>-6</u></u></u></u></u>
PAM	Human	Tm-59- <u>QE<u>KED<u>DG<u>SE<u>SEE<u>EY</u>-12</u></u></u></u></u>
VMAT2	Human	Tm-35- <u>GE<u>DE<u>E<u>ES<u>ES<u>SD</u>-</u></u></u></u></u>
VMAT1	Human	Tm-35- <u>GE<u>DS<u>DE<u>E<u>EP<u>DHEE</u>-</u></u></u></u></u>
VAMP4	Human	25- <u>LE<u>DD<u>S<u>DE<u>E<u>E<u>DF</u>-81-Tm</u></u></u></u></u></u>
Glycoprotein B	HCMV	Tm-125- <u>KDS<u>DE<u>E<u>ENV</u>-</u></u></u>
Glycoprotein E	Herpes virus 3	Tm-28- <u>FED<u>SE<u>ST<u>TD<u>TE<u>EE<u>E<u>F</u>-21</u></u></u></u></u></u></u>
Nef	HIV-1 (AA165476)	55- <u>LE<u>A<u>Q<u>EE<u>EE<u>E<u>E<u>V</u>-139</u></u></u></u></u></u></u>
Kex1p (1)	<i>S. cerevisiae</i>	Tm-29- <u>ADD<u>LE<u>S<u>GL<u>GA<u>E<u>DD<u>LE<u>Q<u>DE<u>Q<u>LEG</u>-40</u></u></u></u></u></u></u></u></u></u></u>
Kex1p (2)	<i>S. cerevisiae</i>	Tm-79- <u>TE<u>I<u>D<u>E<u>S<u>F<u>E<u>M<u>T<u>D<u>F</u>-</u></u></u></u></u></u></u></u></u></u>
Kex2p	<i>S. cerevisiae</i>	Tm-36- <u>TE<u>P<u>E<u>E<u>E<u>V<u>E<u>D<u>F<u>D<u>F<u>D<u>L<u>S<u>D<u>E<u>D<u>H</u>-61</u></u></u></u></u></u></u></u></u></u></u></u></u></u></u></u></u>
Vps10p	<i>S. cerevisiae</i>	Tm-112- <u>FE<u>I<u>E<u>DD<u>D<u>V<u>P<u>T<u>L<u>EE<u>H</u>-37</u></u></u></u></u></u></u></u></u></u>

See legend to Table 1 for explanation of signal format. Serine and threonine residues are underlined.

<sup>a</sup>The number in parentheses is the motif number.

The KDEL receptor binds protein in the ER-Golgi intermediate compartment, or in the early Golgi and returns them to the ER. Although the common mammalian signal is KDEL, it has been shown that the KDEL receptor binds the sequence HDEL more tightly (Scheel *et al*; *J. Biol. Chem.* 268; 7465 (1993)). The intracellular retention signal may be HDEL.

KKX'X' and KX'KX'X' signals are retrieval signals which can be placed on the cytoplasmic side of a type I membrane protein. Sequence requirements of these signals are provided in detail by Teasdale & Jackson (*Annu. Rev. Cell Dev. Biol.*; 12; 27 (1996)).

The intracellular retention signal may be selected from the group of: NPX'Y, YX'X'Z, [DE]X'X'X'L[LI], DX'X'LL, DP[FW], FX'DX'F, NPF, LZX'Z[DE], LLDLL, PWDLW, KDEL, HDEL, KKX'X' or KX'KX'X'; wherein X' is any amino acid and Z' is an amino acid with a bulky hydrophobic side chain.

The intracellular retention signal may be any sequence shown in Tables 2 to 6.

The intracellular retention signal may comprise the Tyrosinase-related protein (TYRP)-1 intracellular retention signal. The intracellular retention signal may comprise the TYRP-1

intracellular domain. The intracellular retention signal may comprise the sequence NQPLLTD (SEQ ID No. 35).

TYRP1 is a well-characterized melanosomal protein which is retained in the melanosome (a specialized lysosome) at >99% efficiency. TYRP1 is a 537 amino acid transmembrane protein with a luminal domain (1-477aa), a transmembrane domain (478-501), and a cytoplasmic domain (502-537). A di-leucine signal residing on the cytoplasmic domain causes retention of the protein. This di-leucine signal has the sequence shown as SEQ ID No. 35 (NQPLLTD).

The intracellular retention signal may be in the endodomain of the transmembrane protein. In other words, the intracellular retention signal may be in the domain of the transmembrane protein which would be on the intracellular side of the cell membrane if the protein was correctly expressed at the cell surface.

The endodomain of the transmembrane protein may comprise at least 100, at least 150, at least 200, at least 300 or at least 500 amino acids.

The endodomain comprising the intracellular retention signal may be located at the carboxy terminus of the transmembrane protein. In particular, where the transmembrane protein comprises a signal sequence at the amino terminus of the peptide the endodomain comprising the intracellular retention signal may be located at the carboxy terminus.

The intracellular retention signal may be proximal to the transmembrane domain, for instance being immediately connected to it. The intracellular retention signal may be distal to the transmembrane domain – for instance at the carboxy-terminus of the endodomain. The positioning of the retention signal modulates its activity allowing “tuning” of the relative expression of two proteins. For instance in the case of the TYRP1 di-leucine motif, proximal placement results in low-level surface expression, while distal placement results in intermediate surface expression, as shown in the Examples.

#### POLYPEPTIDE OF INTEREST

Any or all of A or B; or A, B or C of the nucleic acid sequences in the constructs defined herein may encode a transmembrane protein comprising an intracellular retention signal.

The other nucleic acid sequence may encode any polypeptide of interest (POI). For example, the other POI may be an intracellular protein such as a nucleic protein, a cytoplasmic protein or a protein localised to a membrane-bound compartment; a secretory protein or a transmembrane protein.

Any or all of A or B; or A, B or C of the nucleic acid sequences in the constructs defined herein may encode a chimeric antigen receptor (CAR). The nucleic acid constructs described in the Examples encode two chimeric antigen receptors.

The nucleic acid constructs described in the Examples encode the following polyproteins which comprise the various components in the order they are listed:

**1. Polyprotein comprising anti-CD19 CAR and anti-CD33 CAR with proximal tyrp-1 retention on the anti-CD19 CAR**

**Signal peptide derived from Human CD8a:**

MSLPVTALLPLALLHAARP (SEQ ID No. 38)

**scFv aCD19:**

DIQMTQTTSSLASALGDRVTISCRASQDISKYLNWYQQKPDGTVKLLIYHTSRLHSGVPSRF  
SGSGSGTDYSLTISNLEQEDIATYFCQQGNTLPYTFGGGTKEITKAGGGGSGGGSGGG  
GSGGGGSEVKLQESGPGLVAPSQSLSVTCTVSGVSLPDYGVSWIRQPPRKGLEWLGVIV  
GSETTYYNSALKSRLTIKDNSKSQVFLKMNSLQTDDTAIYYCAKHYYYGGSYAMDYWGQG  
TSVTVS (SEQ ID No. 39)

**Linker:**

SD

**Human CD8aSTK:**

PTTTPAPRPPPTPAPTIASQPLSLRPEACRPAAGGAHVTRGLDFACDI (SEQ ID No. 40)

**Human CD28TM:**

FWVLVVGGVLACYSLLVTVAFIIFWV (SEQ ID No. 41)

**Human Typr-1 intracellular domain (retention signal):**

RARRSMDEANQPLTDQYQCYAEEYEKLQNPNSV (SEQ ID No. 42)

**Human CD3zeta intracellular domain:**

RRVKFSRSADAPAYQQGQNQLYNENLGRREEYDVLKDRRGRDPEMGKPRRKNPQEG  
LYNELQDKMAEAYSEIGMKGERRRGKGHDGLYQGLSTATKDTYDALHMQALPPR (SEQ  
ID No. 43)

**2A peptide:**

RAEGRGSLLTCGDVEENPGP (SEQ ID No. 24)

**Signal peptide derived from mouse Ig kappa:**

MAVPTQVLGLLLLWLTDA (SEQ ID No. 44)

**scFv aCD33:**

RCDIQMTQSPSSLSASVGDRVTITCRASEDIYFNLVWYQQKPGKAPKLLIYDTNRLADGVPS  
RFSGSGSGTQYTLTISSLQPEDFATYYCQHYKNYPLTFGQGTKLEIKRSGGGGSGGGGSG  
GGGGGGGSRSEVQLVESGGGVQPGGSLRLSCAASGFTLSNYGMHWIRQAPGKGLEW  
VSSISLNGGSTYYRDSVKGRFTISRDNAKSTLYLQMNSLRAEDTAVYYCAAQDAYTGGYFD  
YWGQGTLTVVSSM (SEQ ID No. 45)

**Linker:**

DPA

**Hinge and Fc derived from human IgG1 with mutations to prevent FcRg association (HCH2CH3pvaa):**

EPKSPDKTHTCPPCPAPPVAGPSVFLFPPKPKDTLMIAARTPEVTCVVVDVSHEDPEVKFNW  
YVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTIISK  
AKGQPREPQVTLPSSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPVLD  
SDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK (SEQ ID No. 46)

**Linker:**

KDPK (SEQ ID No. 47)

**Human CD148TM:**

AVFGCIFGALVIVTVGGFIFW (SEQ ID No. 48)

**Human CD148 intracellular domain:**

RKKRKDAKNNEVSFSQIKPKKSKLIRVENFEAYFKKQQADSNCFAEYEYEDLKLVGISQPKY  
AAELAENRGKNRYNNVLPYDISRVKLSVQTHSTDDYINANYMPGYHSKKDFIATQGPLPNTL  
KDFWRMVWEKNVYAIIMLTCKVEQGRTKCEEYWPSKQAQDYGDITVAMTSEIVLPEWTIRD  
FTVKNIQTSESHPLRFQHFTSWPDHGVPDTTDLLINFYLVRDYMKQSPPESPILVHCSAGV  
GRTGTFIAIDRLIYQIENENTVDVYGVYDLRMHRPLMVQTEDQYVFLNQCVLDIVRSQKDSK  
VDLIYQNTTAMTIYENLAPVTTFGKTNGYIA (SEQ ID No. 49)

2. Polyprotein comprising anti-CD19 CAR and anti-CD33 CAR with distal tyrp-1 retention on the anti-CD19 CAR

**Signal peptide derived from Human CD8a:**

MSLPVTALLPLALLHAARP (SEQ ID No. 38)

**scFv aCD19:**

DIQMTQTSSLSASLGDRVTISCRASQDISKYLWYQQKPDGTVKLLIYHTSRLHSGVPSRF  
SGSGSGTDYSLTISNLEQEDIATYFCQQGNTLPYTFFGGTKLEITKAGGGGGSGGGGGSGGG  
GSGGGGSEVKLQESGPGLVAPSQSLSVTCTVSGVSLPDYGVSWIRQPPRKGLEWLGVIVW  
GSETTYYNSALKSRLTIKDNSKSQVFLKMNSLQTDDTAIYYCAKHYYYGGSYAMDYWGQG  
TSVTVS (SEQ ID No. 39)

**Linker:**

SD

**Human CD8aSTK:**

PTTTPAPRPPPTPAPTIASQPLSLRPEACRPAAGGAHVTRGLDFACDI (SEQ ID No. 40)

**Human CD28TM:**

FWVLVVVGVLACYSLLVTVAIFIIFWV (SEQ ID No. 41)

**Human CD3zeta intracellular domain:**RRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDRGRDPEMGGKPRRKNPQEG  
LYNELQDKMAEAYSEIGMKGERRGKGHDGLYQGLSTATKDTYDALHMQALPPR (SEQ  
ID No. 43)**Human Typr-1 intracellular domain (retention signal):**

RARRSMDEANQPLTDQYQCYAEEYEKLQNPNQSVV (SEQ ID No. 42)

**2A peptide:**

RAEGRGSLLTCGDVEENPGP (SEQ ID No. 24)

**Signal peptide derived from mouse Ig kappa:**

MAVPTQVLGLLLWLTDA (SEQ ID No. 44)

**scFv aCD33:**RCDIQMTQSPSSLSASVGDRVTITCRASEDIYFNLVWYQQKPGKAPKLLIYDTNRLADGVPS  
RFSGSGSGTQYTLTSSLQPEDFATYYCQHYKNYPLTFGQGTKLEIKRSGGGGGGGGGSG  
GGGSGGGGSRSEVQLVESGGLVQPGGSLRLSCAASGFTLSNYGMHWIRQAPGKGLEW  
VSSISLNGGSTYYRDSVKGRFTISRDNAKSTLYLQMNSLRAEDTAVYYCAAQDAYTGGYFD  
YWGQGTLTVVSSM (SEQ ID No. 45)**Linker:**

DPA

**Hinge and Fc derived from human IgG1 with mutations to prevent FcRg association  
(HCH2CH3pvaa):**EPKSPDKHTCPPCPAPPVAGPSVLFPPKPKDTLMIAARTPEVTCVVVDVSHEDPEVKFNW  
YVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTS  
AKGQPREPQVTLPSSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPVLD  
SDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK (SEQ ID No. 46)**Linker:**

KDPK (SEQ ID No. 47)

**Human CD148TM:**

AVFGCIFGALIVTVGGFIFW (SEQ ID No. 48)

**Human CD148 intracellular domain:**RKKRKDAKNNEVSFSQIKPKKSKLIRVENFEAYFKKQQADSNCGFAEEYEDLKLVGISQPKY  
AAELAENRGKNRYNNVLPYDISRVKLSVQTHSTDDYINANYMPGYHSKKDFIATQGPLPNTL

KDFWRMVWEKNVYAIIMLTKCVEQGRTKCEEYWPSKQAQDYGDTVAMTSEIVLPEWTIRD  
FTVKNIQTSESHPLRQFHFTSWPDHGVPDTDLLINFRLVRDYMKQSPPESPILVHCSAGV  
GRTGTFIAIDRLIYQIENENTDVYGYIVYDLRMHRPLMVQTEDQYVFLNQCVL DIVRSQKDSK  
VDLIYQNTTAMTIYENLAPVTTFGKTNGYIA (SEQ ID No. 49)

3. Polyprotein comprising anti-CD19 CAR and anti-CD33 CAR with E3/19K retention on the anti-CD33 CAR

**Signal peptide derived from Human CD8a:**

MSLPVTALLLPLALLLHAARP (SEQ ID No. 38)

**scFv aCD19:**

DIQMTQTTSSLSASLGDRV TISCRASQDISKYLNWYQQKPDGTVKLLIYHTSRLHSGVPSRF  
SGSGSGTDYSLTISNLEQEDIATYFCQQGNTLPYTFGGGTKEITKAGGGGSGGGSGGG  
GSGGGGSEVKLQESGPGLVAPSQSLSVTCTVSGVSLPDYGVSWIRQPPRKGLEWLGV  
GSETTYYNSALKSRLTIKD NSKSQVFLKMNSLQTDDTAIYYCAKHYYYGGSYAMDYWGQG  
TSVTVS (SEQ ID No. 39)

**Linker:**

SD

**Human CD8aSTK:**

PTTTPAPRPPPTPAPTIASQPLSLRPEACRPAAGGA VHTRGLDFACDI (SEQ ID No. 40)

**Human CD28TM:**

FWVLVVVGVLACYSLLTVAFIIFWV (SEQ ID No. 41)

**Human CD3zeta intracellular domain:**

RRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVL DKRRGRDPEMGGKPRRKNPQEG  
LYNELQDKMAEAYSEIGMKGERRRGKGHDGLYQGLSTATKDTYDALHMQALPPR (SEQ  
ID No. 43)

**2A peptide:**

RAEGRGSLLTCGDVEENPGP (SEQ ID No. 24)

**Signal peptide derived from mouse Ig kappa:**

MAVPTQVLGLLLWLTDA (SEQ ID No. 44)

**scFv aCD33:**

RCDIQMTQSPSSLSASVGDRV TITCRASEDIYFNLVWYQQKPGKAPKLLIYDTNRLADGVPS  
RFSGSGSGTQYTLTISSLQPEDFATYYCQHYKNYPLTFGQGTKEIKRSGGGSGGGSG  
GGGSGGGGSRSEVQLVESGGLVQPGGSLRLSCAASGFTLSNYGMHWIRQAPGKGLEW  
VSSISLNGGSTYYRDSVKGRFTISRDNAKSTLYLQMNSLRAEDTAVYYCAAQDAYTGGYFD  
YWGQGTLTVSSM (SEQ ID No. 45)

**Linker:**

DPA

**Hinge and Fc derived from human IgG1 with mutations to prevent FcRg association (HCH2CH3pvaa):**

EPKSPDKTHTCPCPAPPVAGPSVFLFPPKPKDTLMIARTPEVTCVVVDVSHEDPEVKFNW  
YVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTSK  
AKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPVLD  
SDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK (SEQ ID No. 46)

**Linker:**

KDPK (SEQ ID No. 47)

**Human CD148TM:**

AVFGCIFGALVIVTVGGFIFW (SEQ ID No. 48)

**Human CD148 intracellular domain:**

RKKRKDAKNNEVSFSQIKPKKSKLIRVENFEAYFKKQQADSNCFAEEYEDLKLVGISQPKY  
AAELAENRGKNRYNNVLPYDISRVKLSVQTHSTDDYINANYMPGYHSKKDFIATQGPLPNTL  
KDFWRMVWEKNVYAIIMLTCKVEQGRTKCEEYWPSKQAQDYGDITVAMTSEIVLPEWTIRD  
FTVKNIQTSESHPLRQFHFTSWPDHGVPDTTDLLINFRLVRDYMKQSPPESPILVHCSAGV  
GRTGTFIAIDRLIYQIENENTDVYGVYDLRMHRPLMVQTEDQYVFLNQCVLIDRSQKDSK  
VDLIYQNTTAMTIYENLAPVTTFGKTNGYIA (SEQ ID No. 49)

**Adenoviral E3/19K cytosolic tail:**

KYKSRRSFIDEKKMP (SEQ ID No. 36)

In the above Polyprotein 3, the E3/19K cytosolic tail may be replaced with a truncated version having the sequence: DEKKMP

As shown in the Examples, it was found that, for the tyrp-1 retention signal, low levels of expression could be achieved by placing the retention signal between "Human CD28TM" and "Human CD3zeta intracellular domain" in the sequence given above, whereas medium levels of expression could be achieved by placing the retention signal between "Human CD3zeta intracellular domain" and "2A peptide".

The E3/19K retention signal caused reduction in expression of the anti-CD33 CAR when placed at the C-terminus of the anti-CD33 CAR.

**VECTOR**

The present invention also provides a vector comprising a nucleic acid construct according to the first aspect of the invention.

Such a vector may be used to introduce the nucleic acid construct into a host cell so that it expresses the first and second polypeptide.

The vector may, for example, be a plasmid or a viral vector, such as a retroviral vector or a lentiviral vector, or a transposon based vector or synthetic mRNA.

The vector may be capable of transfecting or transducing a mammalian cell, for example a T cell.

#### CELL

The present invention further provides a cell comprising a nucleic acid construct or vector of the present invention which expresses the first and second polypeptide encoded by the nucleic acid sequence.

The cell may be any eukaryotic cell capable of expressing a transmembrane protein at the cell surface, such as an immunological cell.

#### PROTEIN

The present invention also provides a transmembrane protein comprising an extracellular domain, a transmembrane domain and an endodomain wherein the endodomain comprises an exogenous intracellular retention signal as defined herein.

'Exogenous' means that the intracellular retention signal is not part of the wild type sequence of the transmembrane protein. Wild type sequence refers to the amino acid sequence of the protein which commonly occurs in nature.

The present inventors have demonstrated that the inclusion of an exogenous intracellular retention signal in the endodomain of a transmembrane protein causes the transmembrane protein to be directed to an intracellular compartment. As such the amount of the transmembrane protein expressed at the cell surface is reduced compared to an equivalent transmembrane protein which does not include an intracellular retention signal in the transmembrane domain.

#### METHOD

In a further aspect the present invention relates to a method for modulating the relative cell surface expression of a transmembrane protein expressed from a single nucleic acid construct as a polyprotein with a second protein; by including an intracellular retention signal in the nucleic acid sequence which encodes the transmembrane protein.

The invention will now be further described by way of Examples, which are meant to serve to assist one of ordinary skill in the art in carrying out the invention and are not intended in any way to limit the scope of the invention.

## EXAMPLES

### Example 1 – Dissection of TYRP1 lysosomal retention signals

The ability of the Tyrosinase-related protein 1 (TYRP1) retention signal to cause retention of a polypeptide when in the context of a more complex endodomain was determined using a number of constructs (Figure 2). The wild-type construct was compared with constructs where enhanced Green Fluorescent Protein (eGFP) was added or replaced the TYRP1 endodomain. Where eGFP was added, it was placed either after or before the native endodomain so the retention signal was either in its native location (just under the membrane), or distal to it.

All constructs are co-expressed with IRES.CD34. Staining of transduced SupT1 cells is shown with intracellular and surface staining in Figure 2.

It was found that replacement of the endodomain resulted in very bright surface expression, introduction of eGFP after the retention signal to almost no surface expression and introduction before the retention signal to intermediate surface expression (Figure 2).

### Example 2 – Modulation of the relative expression of a transmembrane protein co-expressed from a single expression cassette with a separate protein

An expression cassette encoding two CAR transmembrane proteins was modified such that one of the CAR proteins had the lysosomal retention signal from TYRP1 introduced either proximal or distal to the membrane. Expression of each of these two new variants at the cell surface was compared with that of the original unmodified CAR protein.

PBMCs were isolated from blood and stimulated using PHA and IL-2. Two days later the cells were transduced on retronectin coated plates with retro virus containing the CD19:CD33 CAR construct. On day 5 the expression level of the two CARs translated by the construct was evaluated via flow cytometry and the cells were depleted of CD56+ cells (predominantly NK cells). On day 6 the PBMCs were placed in a co-culture with target cells at a 1:2 effector to target cell ratio. On day 8 the supernatant was collected and analysed for IFN-gamma secretion via ELISA.

The pattern observed with Tyrp1-eGFP fusions was observed with some reduction of expression of modified transmembrane protein with the distal retention signal and marked reduction in the case of proximal retention signal. As expected, expression of the second transmembrane protein from the cassette was not altered (Figure 3).

**Example 3 – Modulation of expression using a retention signal from the Adenoviral E3/19K protein**

The human adenovirus E3/19K protein is a type I transmembrane glycoprotein of the Endoplasmic Reticulum/Golgi that abrogates cell surface transport of major histocompatibility complex class I (MHC-I) and MHC-I-related chain A and B (MICA/B) molecules. The retention motif was identified to be depended on the cytosolic tail of the adenovirus E3/19K protein. More specifically, the last 6aa DEKKMP was found to be the most important for retention. The optimal positioning was found to be at the c-terminus of the protein.

An expression cassette encoding two CAR transmembrane proteins, as described in Example 2, was modified such that one of the CAR proteins had the retention motif from adenovirus E3/19K protein. In this experiment, the retention motif on the second CAR in the expression cassette (the anti-CD33 inhibitory CAR).

Constructs were generated comprising either the entire cytosolic tail of adenovirus E3/19K protein or only the last 6aa from E3/19K (DEKKMP), which were found to be critical for its Golgi/ER retention ability (Figure 4). These constructs were transfected into 293T cells and stained primarily with a chimeric soluble CD19-Rabbit Fc and a chimeric soluble CD33-Mouse Fc proteins. These cells were then subsequently stained with an anti-Rabbit Fc-FITC and an anti-Mouse Fc-APC (Figure 5). These cells show a clear retention when the full length adenovirus E3/19K protein, or the DEKKMP motif, was placed on the anti-CD33 receptor but had no effect on anti-CD19 receptor expression levels.

All publications mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described methods and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology, cell biology or related fields are intended to be within the scope of the following claims.

CLAIMS

1. A nucleic acid construct comprising the following structure:

A – X – B

in which

A and B are nucleic acid sequences encoding first and second transmembrane proteins which, when the nucleic acid construct is introduced into a cell, are expressed at the cell surface; and

X is a nucleic acid sequence which encodes a cleavage site, wherein either the first or second transmembrane protein comprises an intracellular retention signal.

2. A nucleic acid construct according to claim 1 wherein the intracellular retention signal directs the transmembrane protein to a lysosomal, endosomal or Golgi compartment.
3. A nucleic acid construct according to claim 1 or claim 2, wherein the intracellular retention signal is selected from the following group: an endocytosis signal; a Golgi retention signal; a *trans*-Golgi network (TGN) recycling signal; an endoplasmic reticulum (ER) retention signal; and a lysosomal sorting signal.
4. A nucleic acid construct according to any preceding claim wherein the intracellular retention signal is proximal to a transmembrane domain of the transmembrane protein.
5. A nucleic acid construct according to any of claims 1 to 3 wherein the intracellular retention signal is distal to a transmembrane domain of the transmembrane protein.
6. A nucleic acid construct according to any preceding claim wherein X is a nucleic acid sequence encoding a self-cleaving peptide, a furin cleavage site or a Tobacco Etch Virus cleavage site.
7. A nucleic acid construct according to any preceding claim wherein the transmembrane protein is a Chimeric-antigen receptor (CAR).

8. A nucleic acid construct according to any preceding claim, wherein one or both of the first and second transmembrane proteins is a single pass transmembrane protein.
9. A nucleic acid construct according to claim 8 wherein one or both of the first and second transmembrane proteins is a type I transmembrane protein.

10. A nucleic acid construct comprising the following structure:

A – X – B – Y – C

in which

A, B and C are nucleic acid sequences encoding a first, second and third polypeptides of interest (POIs); and

X and Y are nucleic acid sequences which may be the same or different, each of which encodes a cleavage site,

wherein at least two of the POIs are transmembrane proteins which, when the nucleic acid construct is introduced into a cell, are expressed at the cell surface, and which comprise an intracellular retention signal.

11. A nucleic acid construct according to claim 10, wherein the at least two which are transmembrane proteins and which comprise an intracellular retention signal:
  - (a) comprise different intracellular retention signals; and/or
  - (b) have the intracellular retention signal located at a different position in the POI, such that when the nucleic acid is expressed in a cell, there is differential relative expression of the at least two POIs at the cell surface.
12. A vector comprising a nucleic acid construct according to any preceding claim.
13. A cell comprising a nucleic acid construct according to any one of claims 1 to 11 or a vector according to claim 12.

1/9

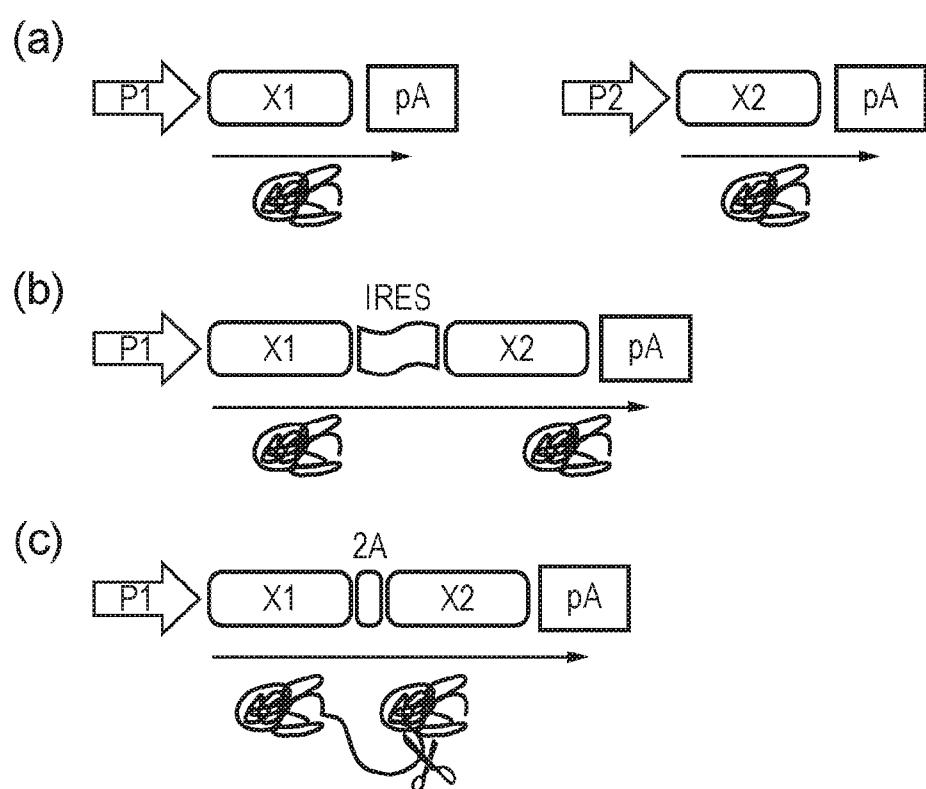


FIG. 1

2/9

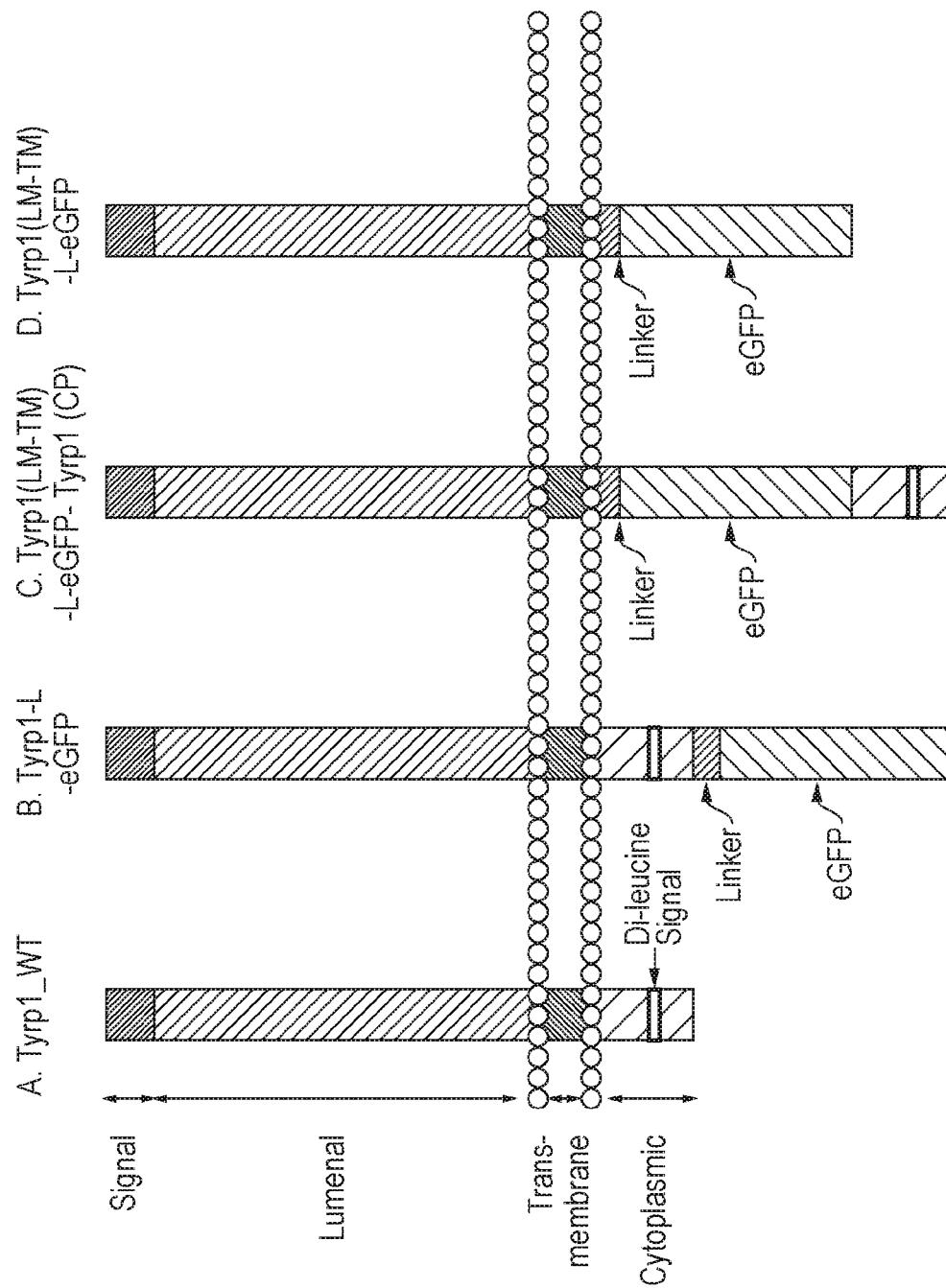


FIG. 2

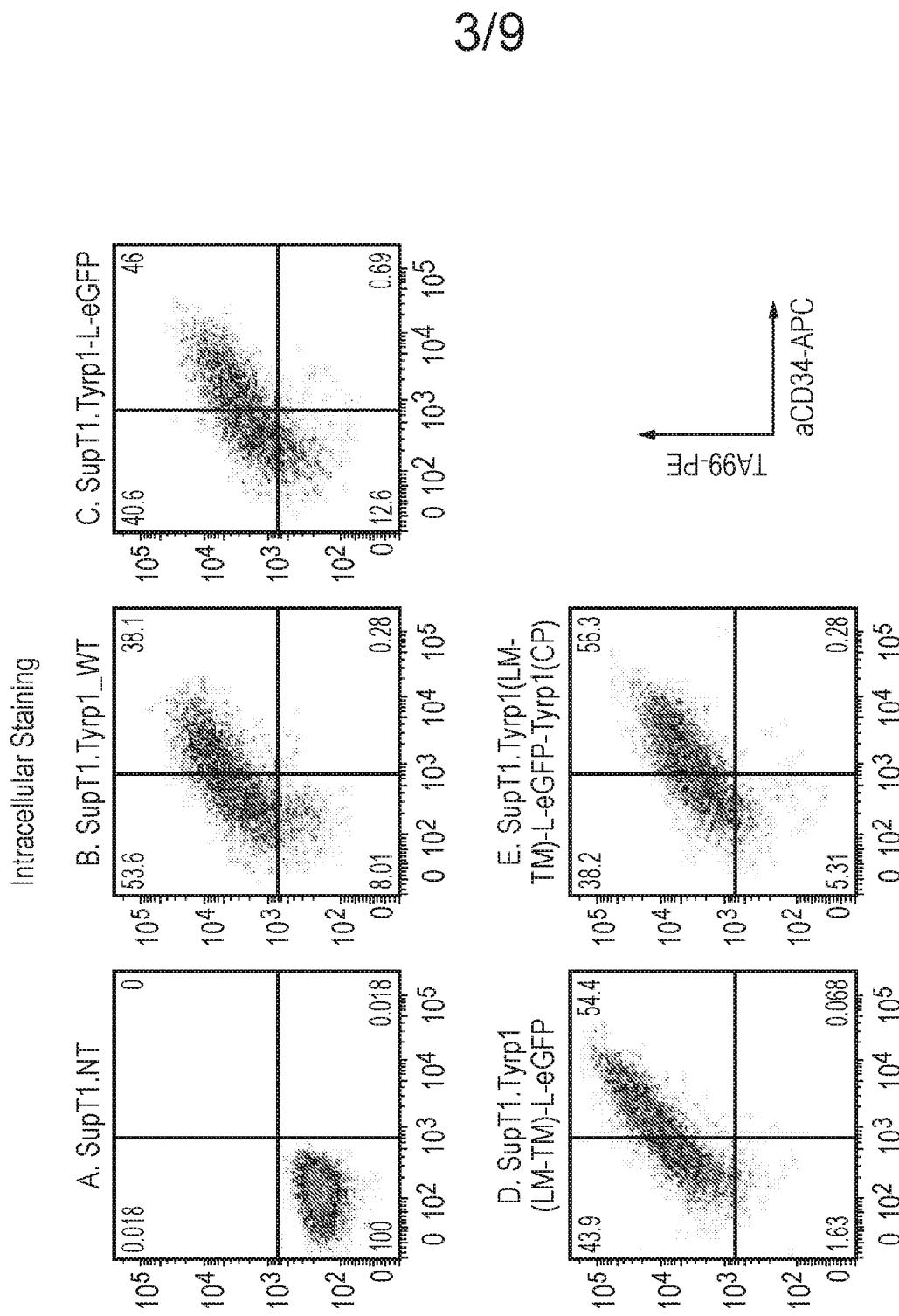


FIG. 2 (Continued)

4/9

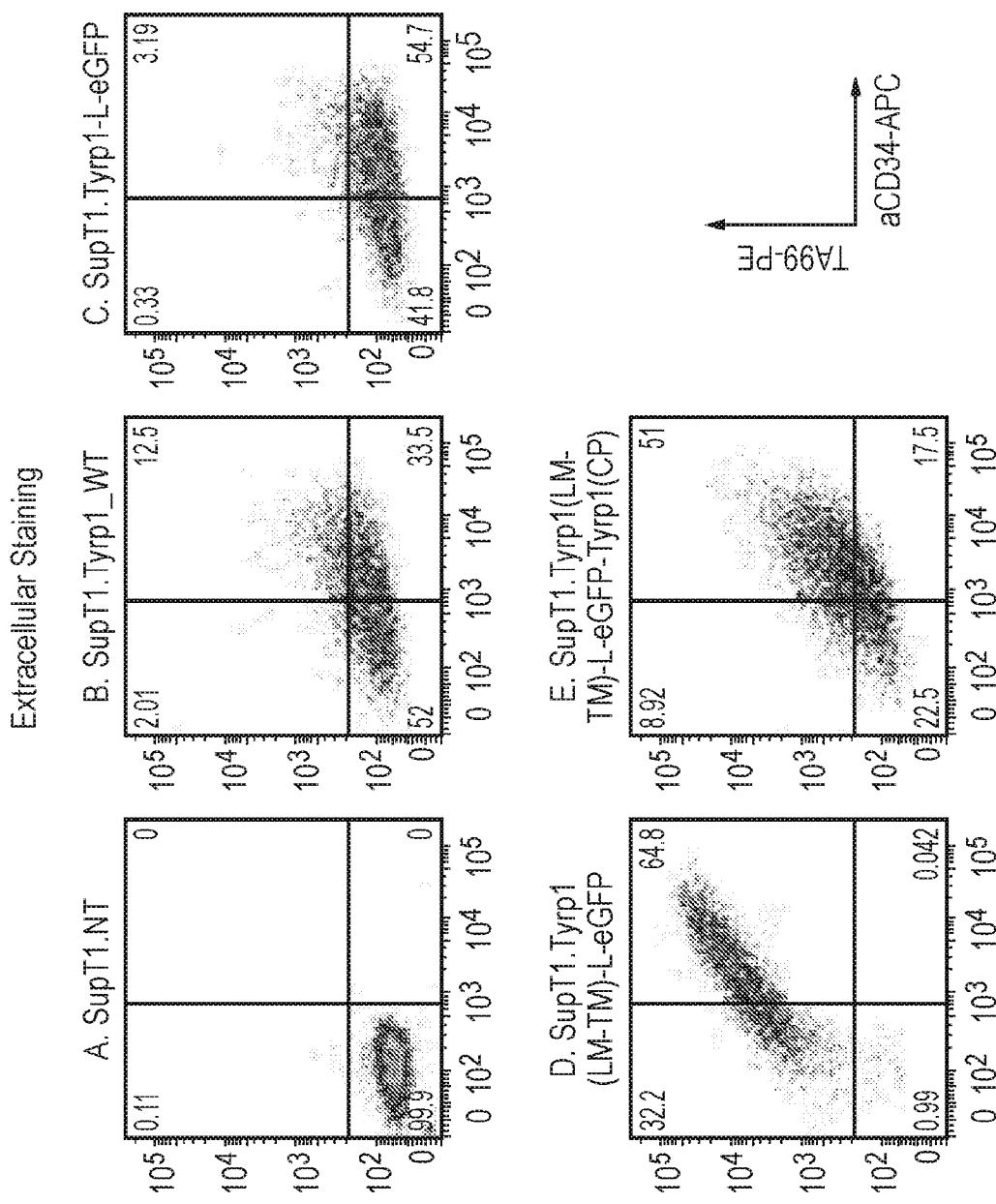


FIG. 2 (Continued)

5/9

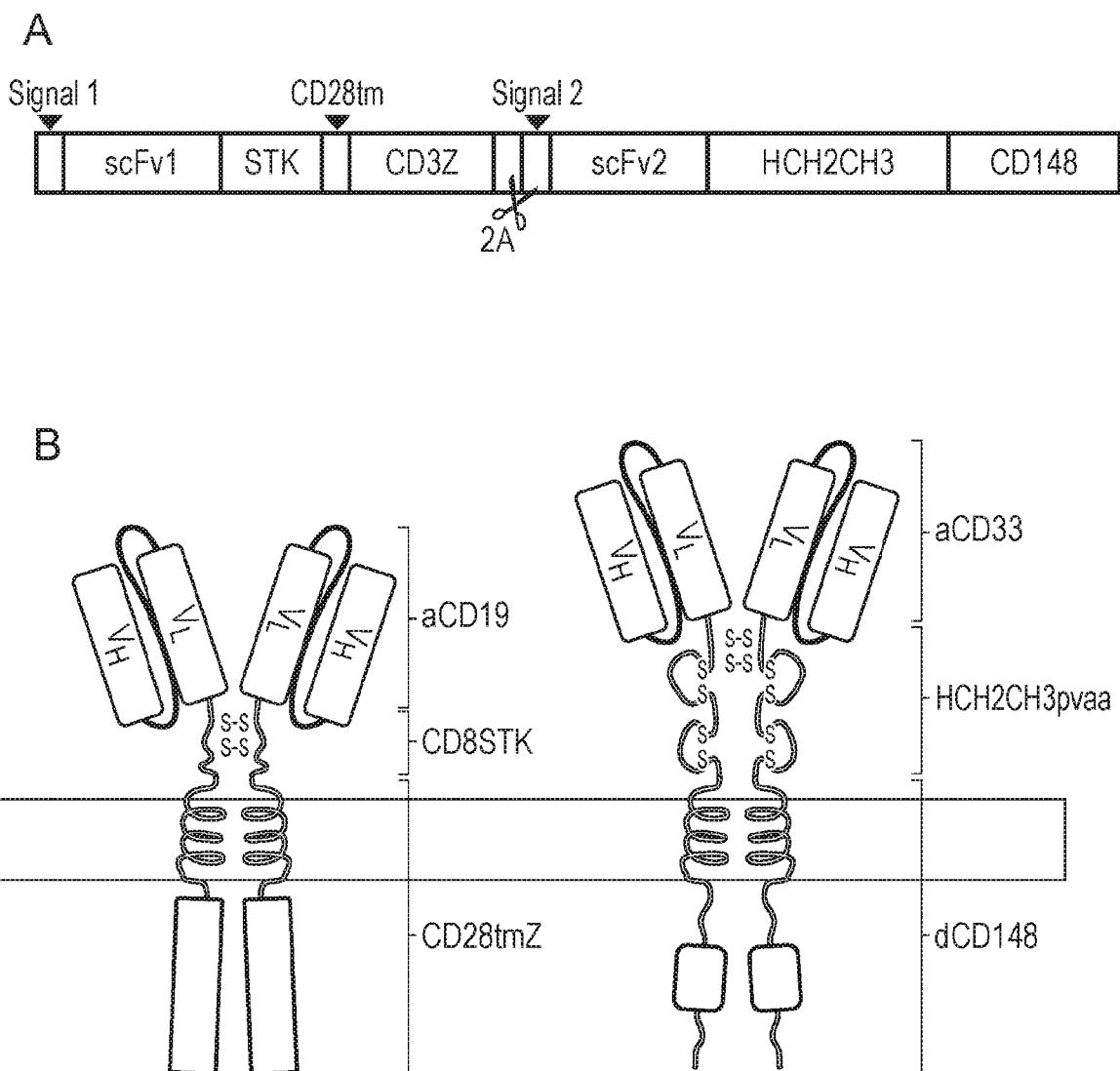
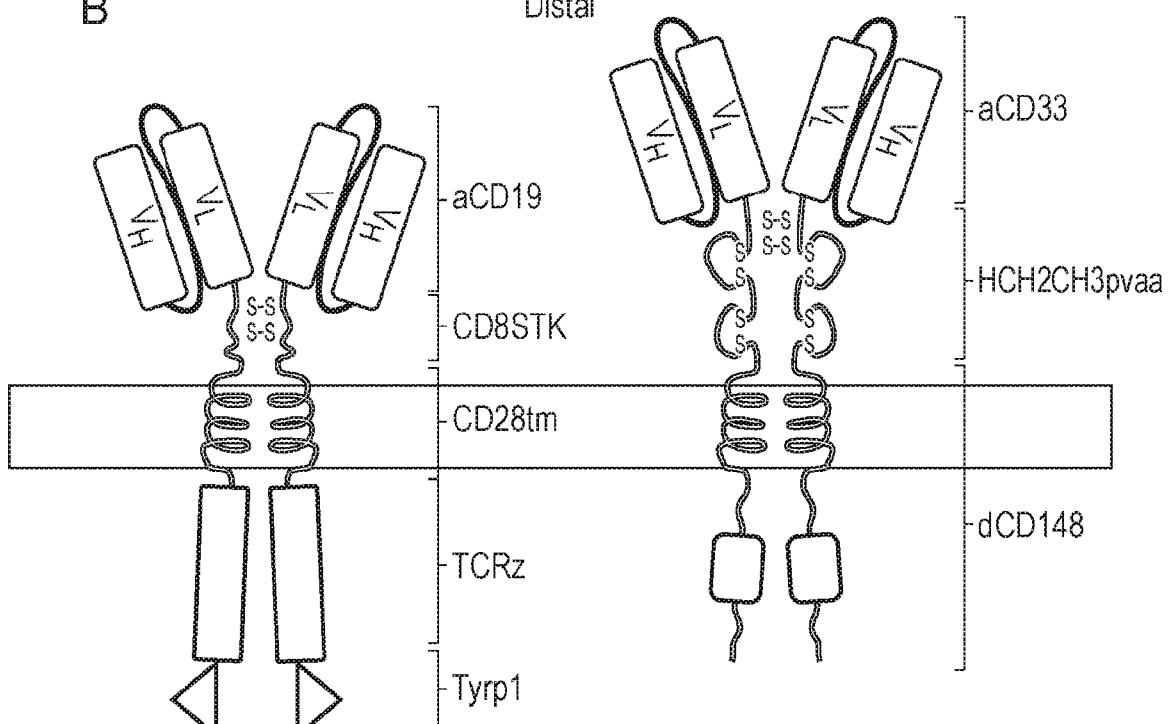


FIG. 3

6/9

B

Distal



B

Proximal

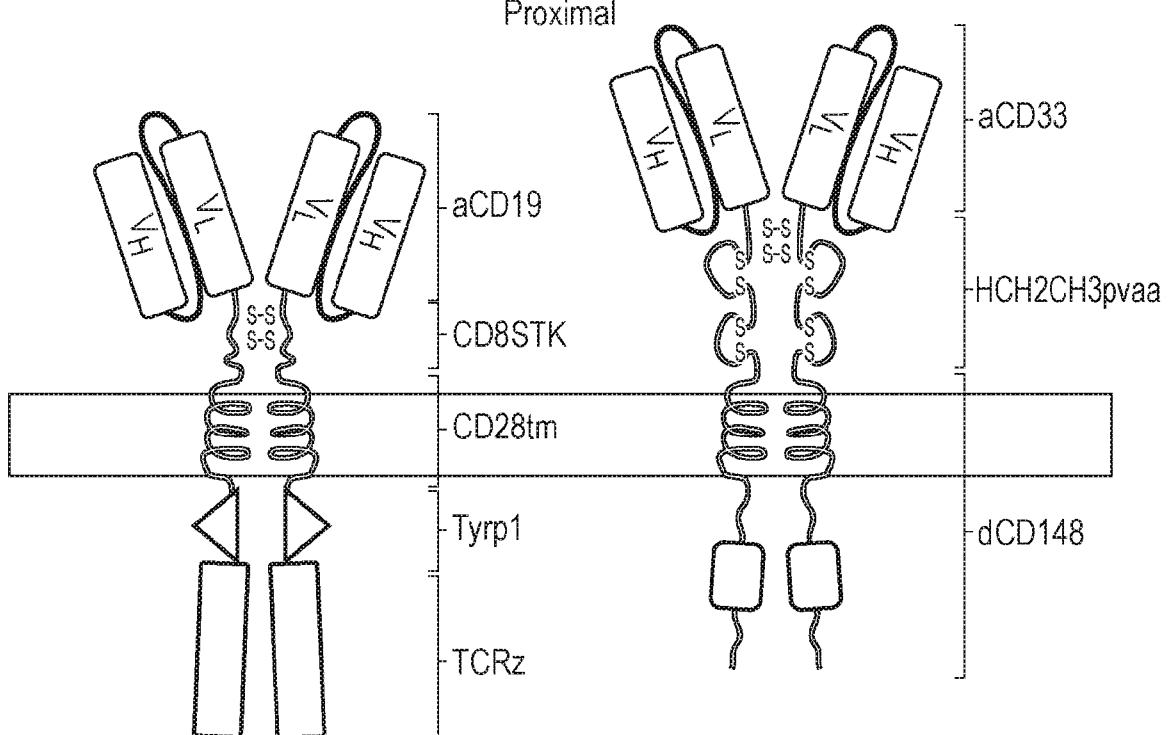


FIG. 3 (Continued)

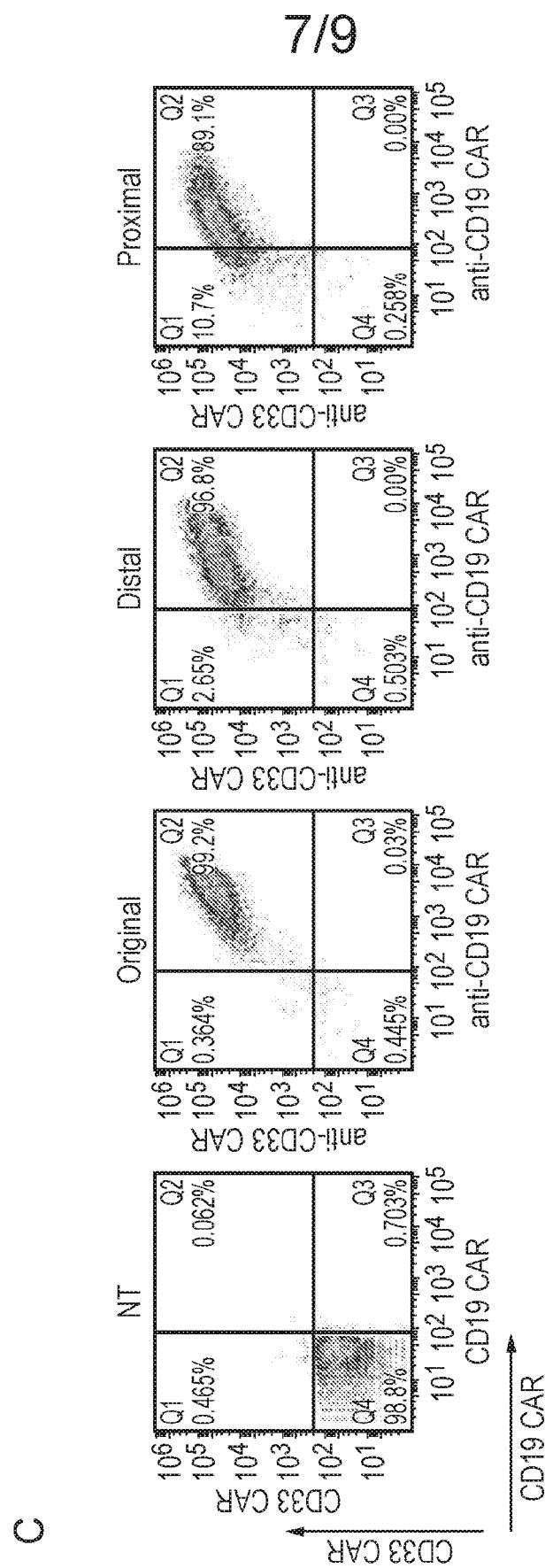


FIG. 3 (Continued)

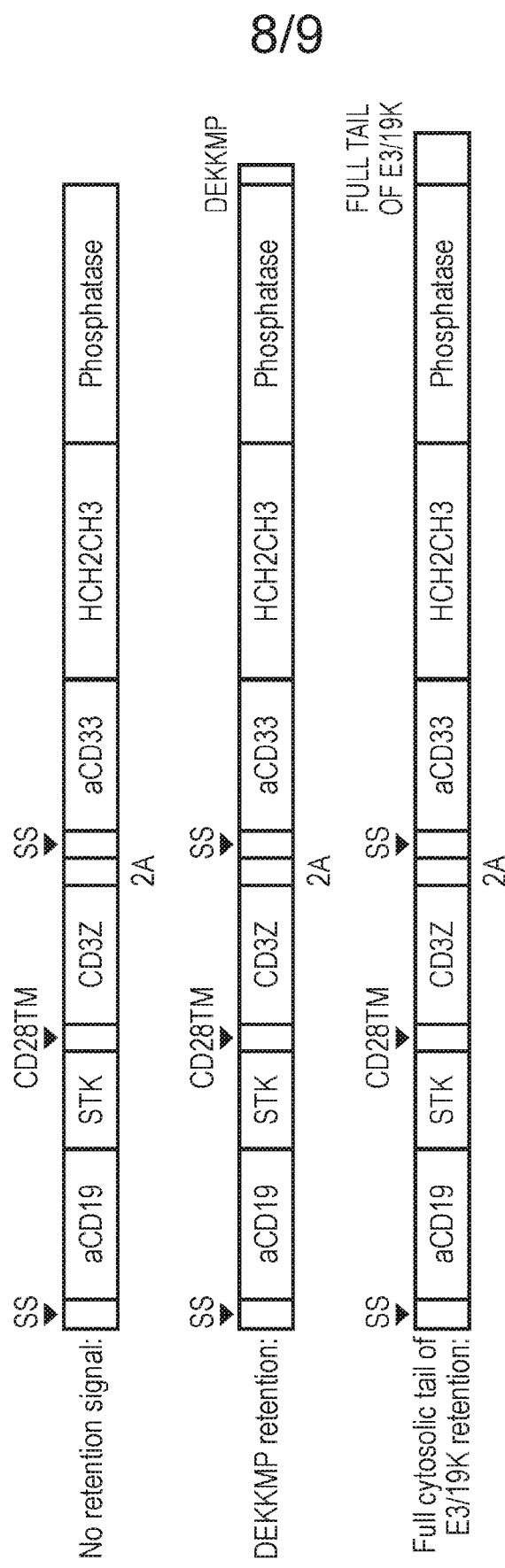
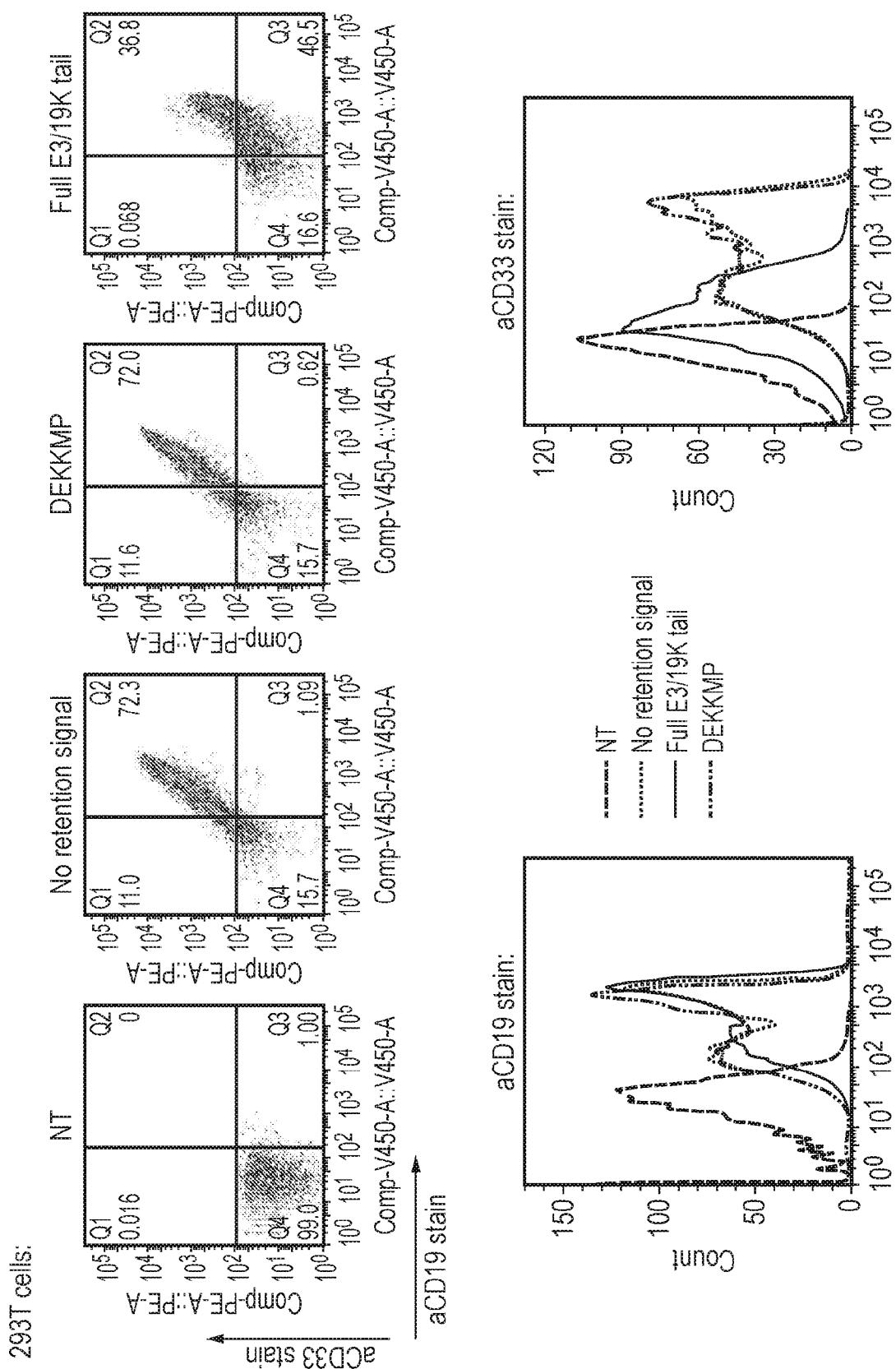


FIG. 4

9/9



50

SEQUENCE LISTING

<110> UCL Business PLC

<120> Nucleic acid construct

<130> P106298PCT

<150> GB1507104.6

<151> 2015-04-27

<160> 197

<170> PatentIn version 3.5

<210> 1

<211> 8

<212> PRT

<213> Artificial Sequence

<220>

<223> Cleavage site

<400> 1

Asp Ile Glu Thr Asn Pro Gly Pro

1 5

<210> 2

<211> 8

<212> PRT

<213> Artificial Sequence

<220>

<223> Cleavage site

<400> 2

Asp Val Glu Thr Asn Pro Gly Pro

1 5

<210> 3

<211> 8

<212> PRT

<213> Artificial Sequence

<220>

<223> Cleavage site

<400> 3

Asp Val Glu Met Asn Pro Gly Pro

1 5

<210> 4

<211> 8

<212> PRT

<213> Artificial Sequence

<220>

<223> Cleavage site

<400> 4

Asp Val Glu Ser Asn Pro Gly Pro

1 5

<210> 5

<211> 8

<212> PRT

<213> Artificial Sequence

<220>

<223> Cleavege site

<400> 5

Asp Met Glu Ser Asn Pro Gly Pro

1 5

<210> 6

<211> 8

<212> PRT

<213> Artificial Sequence

<220>

<223> Cleavage site

<400> 6

Asp Val Glu Leu Asn Pro Gly Pro

1 5

<210> 7

<211> 8

<212> PRT

<213> Artificial Sequence

<220>

<223> Cleavage site

<400> 7

Asp Val Glu Glu Asn Pro Gly Pro

1 5

<210> 8

<211> 8

<212> PRT

<213> Artificial Sequence

<220>

<223> Cleavage site

<400> 8

Asp Ile Glu Leu Asn Pro Gly Pro

1 5

<210> 9

<211> 8

<212> PRT

<213> Artificial Sequence

<220>

<223> Cleavage site

<400> 9

Asp Ile Glu Gln Asn Pro Gly Pro

1 5

<210> 10

<211> 8

<212> PRT

<213> Artificial Sequence

<220>

<223> Cleavage site

<400> 10

Asp Ser Glu Phe Asn Pro Gly Pro

1 5

<210> 11

<211> 19

<212> PRT

<213> Artificial Sequence

<220>

<223> Cleavage sequence

<400> 11

Leu Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly Asp Val Glu Ser Asn

1 5 10 15

Pro Gly Pro

<210> 12

<211> 19

<212> PRT

<213> Artificial Sequence

<220>

<223> Cleavage sequence

<400> 12

Leu Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly Asp Val Gln Ser Asn  
1 5 10 15

Pro Gly Pro

<210> 13

<211> 19

<212> PRT

<213> Artificial Sequence

<220>

<223> Cleavage sequence

<400> 13

Leu Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly Asp Val Glu Ile Asn  
1 5 10 15

Pro Gly Pro

<210> 14

<211> 19

<212> PRT

<213> Artificial Sequence

<220>

<223> Cleavage sequence

<400> 14

Leu Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly Asp Val Glu Phe Asn  
1 5 10 15

Pro Gly Pro

<210> 15  
<211> 19  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Cleavage sequence

<400> 15

Leu Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly Asp Val Glu Ser His  
1 5 10 15

Pro Gly Pro

<210> 16  
<211> 19  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Cleavage sequence

<400> 16

Leu Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly Asp Val Glu Ser Glu  
1 5 10 15

Pro Gly Pro

<210> 17  
<211> 19  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Cleavage sequence

<400> 17

Leu Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly Asp Val Glu Ser Gln  
1 5 10 15

Pro Gly Pro

<210> 18

<211> 19

<212> PRT

<213> Artificial Sequence

<220>

<223> Cleavage sequence

<400> 18

Leu Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly Asp Val Glu Ser Asn  
1 5 10 15

Pro Gly Gly

<210> 19

<211> 20

<212> PRT

<213> Artificial Sequence

<220>

<223> Cleavage site, 2A-like sequence

<400> 19

Tyr His Ala Asp Tyr Tyr Lys Gln Arg Leu Ile His Asp Val Glu Met  
1 5 10 15

Asn Pro Gly Pro

20

<210> 20  
<211> 20  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Cleavage site, 2A-like sequence

<400> 20

His Tyr Ala Gly Tyr Phe Ala Asp Leu Leu Ile His Asp Ile Glu Thr  
1 5 10 15

Asn Pro Gly Pro  
20

<210> 21  
<211> 20  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Cleavage site, 2A-like sequence

<400> 21

Gln Cys Thr Asn Tyr Ala Leu Leu Lys Leu Ala Gly Asp Val Glu Ser  
1 5 10 15

Asn Pro Gly Pro  
20

<210> 22  
<211> 19  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Cleavage site, 2A-like sequence

<400> 22

Ala Thr Asn Phe Ser Leu Leu Lys Gln Ala Gly Asp Val Glu Glu Asn

1 5 10 15

Pro Gly Pro

<210> 23

<211> 19

<212> PRT

<213> Artificial Sequence

<220>

<223> Cleavage site, 2A-like sequence

<400> 23

Ala Ala Arg Gln Met Leu Leu Leu Ser Gly Asp Val Glu Thr Asn

1 5 10 15

Pro Gly Pro

<210> 24

<211> 20

<212> PRT

<213> Artificial Sequence

<220>

<223> Cleavage site, 2A-like sequence

<400> 24

Arg Ala Glu Gly Arg Gly Ser Leu Leu Thr Cys Gly Asp Val Glu Glu

1 5 10 15

Asn Pro Gly Pro

20

<210> 25  
<211> 20  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Cleavage site, 2A-like sequence

<400> 25

Thr Arg Ala Glu Ile Glu Asp Glu Leu Ile Arg Ala Gly Ile Glu Ser  
1 5 10 15

Asn Pro Gly Pro  
20

<210> 26  
<211> 20  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Cleavage site, 2A-like sequence

<400> 26

Thr Arg Ala Glu Ile Glu Asp Glu Leu Ile Arg Ala Asp Ile Glu Ser  
1 5 10 15

Asn Pro Gly Pro  
20

<210> 27  
<211> 20  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Cleavage site, 2A-like sequence

<400> 27

Ala Lys Phe Gln Ile Asp Lys Ile Leu Ile Ser Gly Asp Val Glu Leu

1 5 10 15

Asn Pro Gly Pro

20

<210> 28

<211> 20

<212> PRT

<213> Artificial Sequence

<220>

<223> Cleavage site, 2A-like sequence

<400> 28

Ser Ser Ile Ile Arg Thr Lys Met Leu Val Ser Gly Asp Val Glu Glu

1 5 10 15

Asn Pro Gly Pro

20

<210> 29

<211> 20

<212> PRT

<213> Artificial Sequence

<220>

<223> Cleavage site, 2A-like sequence

<400> 29

Cys Asp Ala Gln Arg Gln Lys Leu Leu Leu Ser Gly Asp Ile Glu Gln

1 5 10 15

Asn Pro Gly Pro

20

<210> 30  
<211> 20  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Cleavage site, 2A-like sequence

<400> 30

Tyr Pro Ile Asp Phe Gly Gly Phe Leu Val Lys Ala Asp Ser Glu Phe  
1 5 10 15

Asn Pro Gly Pro  
20

<210> 31  
<211> 57  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Cleavage sequence

<400> 31

Val Thr Glu Leu Leu Tyr Arg Met Lys Arg Ala Glu Thr Tyr Cys Pro  
1 5 10 15

Arg Pro Leu Ala Ile His Pro Thr Glu Ala Arg His Lys Gln Lys Ile  
20 25 30

Val Ala Pro Val Lys Gln Thr Leu Asn Phe Asp Leu Leu Lys Leu Ala  
35 40 45

Gly Asp Val Glu Ser Asn Pro Gly Pro  
50 55

<210> 32  
<211> 40  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Cleavage sequence

<400> 32

Leu Leu Ala Ile His Pro Thr Glu Ala Arg His Lys Gln Lys Ile Val  
1 5 10 15

Ala Pro Val Lys Gln Thr Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly  
20 25 30

Asp Val Glu Ser Asn Pro Gly Pro  
35 40

<210> 33  
<211> 33  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Cleavage sequence

<400> 33

Glu Ala Arg His Lys Gln Lys Ile Val Ala Pro Val Lys Gln Thr Leu  
1 5 10 15

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

Pro

<210> 34  
<211> 24  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Cleavage sequence

<400> 34

Ala Pro Val Lys Gln Thr Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly

1 5 10 15

Asp Val Glu Ser Asn Pro Gly Pro  
20

<210> 35  
<211> 7  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Intracellular retention signal

<400> 35

Asn Gln Pro Leu Leu Thr Asp  
1 5

<210> 36  
<211> 15  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Intracellular retention signal

<400> 36

Lys Tyr Lys Ser Arg Arg Ser Phe Ile Asp Glu Lys Lys Met Pro  
1 5 10 15

<210> 37  
<211> 6  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Intracellular retention signal

<400> 37

Asp Glu Lys Lys Met Pro  
1 5

<210> 38  
<211> 21  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Signal peptide derived from Human CD8a

<400> 38

Met Ser Leu Pro Val Thr Ala Leu Leu Leu Pro Leu Ala Leu Leu Leu  
1 5 10 15

His Ala Ala Arg Pro  
20

<210> 39  
<211> 248  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> scFv aCD19

<400> 39

Asp Ile Gln Met Thr Gln Thr Thr Ser Ser Leu Ser Ala Ser Leu Gly  
1 5 10 15

Asp Arg Val Thr Ile Ser Cys Arg Ala Ser Gln Asp Ile Ser Lys Tyr  
20 25 30

Leu Asn Trp Tyr Gln Gln Lys Pro Asp Gly Thr Val Lys Leu Ile  
35 40 45

Tyr His Thr Ser Arg Leu His Ser Gly Val Pro Ser Arg Phe Ser Gly  
50 55 60

Ser Gly Ser Gly Thr Asp Tyr Ser Leu Thr Ile Ser Asn Leu Glu Gln  
65 70 75 80

Glu Asp Ile Ala Thr Tyr Phe Cys Gln Gln Gly Asn Thr Leu Pro Tyr  
85 90 95

Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile Thr Lys Ala Gly Gly Gly  
100 105 110

Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Gly  
115 120 125

Ser Glu Val Lys Leu Gln Glu Ser Gly Pro Gly Leu Val Ala Pro Ser  
130 135 140

Gln Ser Leu Ser Val Thr Cys Thr Val Ser Gly Val Ser Leu Pro Asp  
145 150 155 160

Tyr Gly Val Ser Trp Ile Arg Gln Pro Pro Arg Lys Gly Leu Glu Trp  
165 170 175

Leu Gly Val Ile Trp Gly Ser Glu Thr Thr Tyr Tyr Asn Ser Ala Leu  
180 185 190

Lys Ser Arg Leu Thr Ile Ile Lys Asp Asn Ser Lys Ser Gln Val Phe  
195            200            205

Leu Lys Met Asn Ser Leu Gln Thr Asp Asp Thr Ala Ile Tyr Tyr Cys  
210            215            220

Ala Lys His Tyr Tyr Gly Gly Ser Tyr Ala Met Asp Tyr Trp Gly  
225            230            235            240

Gln Gly Thr Ser Val Thr Val Ser  
245

<210> 40  
<211> 47  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Human CD8aSTK

<400> 40

Pro Thr Thr Thr Pro Ala Pro Arg Pro Pro Thr Pro Ala Pro Thr Ile  
1            5            10            15

Ala Ser Gln Pro Leu Ser Leu Arg Pro Glu Ala Cys Arg Pro Ala Ala  
20            25            30

Gly Gly Ala Val His Thr Arg Gly Leu Asp Phe Ala Cys Asp Ile  
35            40            45

<210> 41  
<211> 27  
<212> PRT  
<213> Artificial Sequence

<220>

<223> Human CD28TM

<400> 41

Phe Trp Val Leu Val Val Val Gly Gly Val Leu Ala Cys Tyr Ser Leu  
1 5 10 15

Leu Val Thr Val Ala Phe Ile Ile Phe Trp Val  
20 25

<210> 42

<211> 36

<212> PRT

<213> Artificial Sequence

<220>

<223> Human Typr-1 intracellular domain (retention signal)

<400> 42

Arg Ala Arg Arg Ser Met Asp Glu Ala Asn Gln Pro Leu Leu Thr Asp  
1 5 10 15

Gln Tyr Gln Cys Tyr Ala Glu Glu Tyr Glu Lys Leu Gln Asn Pro Asn  
20 25 30

Gln Ser Val Val  
35

<210> 43

<211> 113

<212> PRT

<213> Artificial Sequence

<220>

<223> Human CD3zeta intracellular domain

<400> 43

Arg Arg Val Lys Phe Ser Arg Ser Ala Asp Ala Pro Ala Tyr Gln Gln

1 5 10 15

Gly Gln Asn Gln Leu Tyr Asn Glu Leu Asn Leu Gly Arg Arg Glu Glu  
20 25 30

Tyr Asp Val Leu Asp Lys Arg Arg Gly Arg Asp Pro Glu Met Gly Gly  
35 40 45

Lys Pro Arg Arg Lys Asn Pro Gln Glu Gly Leu Tyr Asn Glu Leu Gln  
50 55 60

Lys Asp Lys Met Ala Glu Ala Tyr Ser Glu Ile Gly Met Lys Gly Glu  
65 70 75 80

Arg Arg Arg Gly Lys Gly His Asp Gly Leu Tyr Gln Gly Leu Ser Thr  
85 90 95

Ala Thr Lys Asp Thr Tyr Asp Ala Leu His Met Gln Ala Leu Pro Pro  
100 105 110

Arg

<210> 44  
<211> 18  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Signal peptide derived from mouse Ig kappa

<400> 44

Met Ala Val Pro Thr Gln Val Leu Gly Leu Leu Leu Leu Trp Leu Thr  
1 5 10 15

Asp Ala

<210> 45  
<211> 254  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> scFv aCD33

<400> 45

Arg Cys Asp Ile Gln Met Thr Gln Ser Pro Ser Ser Leu Ser Ala Ser  
1 5 10 15

Val Gly Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Glu Asp Ile Tyr  
20 25 30

Phe Asn Leu Val Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro Lys Leu  
35 40 45

Leu Ile Tyr Asp Thr Asn Arg Leu Ala Asp Gly Val Pro Ser Arg Phe  
50 55 60

Ser Gly Ser Gly Ser Gly Thr Gln Tyr Thr Leu Thr Ile Ser Ser Leu  
65 70 75 80

Gln Pro Glu Asp Phe Ala Thr Tyr Tyr Cys Gln His Tyr Lys Asn Tyr  
85 90 95

Pro Leu Thr Phe Gly Gln Gly Thr Lys Leu Glu Ile Lys Arg Ser Gly  
100 105 110

Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly Ser Gly Gly  
115 120 125

Gly Gly Ser Arg Ser Glu Val Gln Leu Val Glu Ser Gly Gly Leu  
130            135            140

Val Gln Pro Gly Gly Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe  
145            150            155            160

Thr Leu Ser Asn Tyr Gly Met His Trp Ile Arg Gln Ala Pro Gly Lys  
165            170            175

Gly Leu Glu Trp Val Ser Ser Ile Ser Leu Asn Gly Gly Ser Thr Tyr  
180            185            190

Tyr Arg Asp Ser Val Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ala  
195            200            205

Lys Ser Thr Leu Tyr Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr  
210            215            220

Ala Val Tyr Tyr Cys Ala Ala Gln Asp Ala Tyr Thr Gly Gly Tyr Phe  
225            230            235            240

Asp Tyr Trp Gly Gln Gly Thr Leu Val Thr Val Ser Ser Met  
245            250

<210> 46  
<211> 231  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Hinge and Fc derived from human IgG1 with mutations to prevent  
FcRg association (HCH2CH3pva)

<400> 46

Glu Pro Lys Ser Pro Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala

1 5 10 15

Pro Pro Val Ala Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys  
20 25 30

Asp Thr Leu Met Ile Ala Arg Thr Pro Glu Val Thr Cys Val Val Val  
35 40 45

Asp Val Ser His Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp  
50 55 60

Gly Val Glu Val His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln Tyr  
65 70 75 80

Asn Ser Thr Tyr Arg Val Val Ser Val Leu Thr Val Leu His Gln Asp  
85 90 95

Trp Leu Asn Gly Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Ala Leu  
100 105 110

Pro Ala Pro Ile Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro Arg  
115 120 125

Glu Pro Gln Val Tyr Thr Leu Pro Pro Ser Arg Asp Glu Leu Thr Lys  
130 135 140

Asn Gln Val Ser Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp  
145 150 155 160

Ile Ala Val Glu Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys  
165 170 175

Thr Thr Pro Pro Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser  
180 185 190

Lys Leu Thr Val Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe Ser  
195 200 205

Cys Ser Val Met His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser  
210 215 220

Leu Ser Leu Ser Pro Gly Lys  
225 230

<210> 47  
<211> 4  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Linker

<400> 47

Lys Asp Pro Lys  
1

<210> 48  
<211> 21  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Human CD148TM

<400> 48

Ala Val Phe Gly Cys Ile Phe Gly Ala Leu Val Ile Val Thr Val Gly  
1 5 10 15

Gly Phe Ile Phe Trp  
20

<210> 49  
<211> 341  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Human CD148 intracellular domain

<400> 49

Arg Lys Lys Arg Lys Asp Ala Lys Asn Asn Glu Val Ser Phe Ser Gln  
1 5 10 15

Ile Lys Pro Lys Lys Ser Lys Leu Ile Arg Val Glu Asn Phe Glu Ala  
20 25 30

Tyr Phe Lys Lys Gln Gln Ala Asp Ser Asn Cys Gly Phe Ala Glu Glu  
35 40 45

Tyr Glu Asp Leu Lys Leu Val Gly Ile Ser Gln Pro Lys Tyr Ala Ala  
50 55 60

Glu Leu Ala Glu Asn Arg Gly Lys Asn Arg Tyr Asn Asn Val Leu Pro  
65 70 75 80

Tyr Asp Ile Ser Arg Val Lys Leu Ser Val Gln Thr His Ser Thr Asp  
85 90 95

Asp Tyr Ile Asn Ala Asn Tyr Met Pro Gly Tyr His Ser Lys Lys Asp  
100 105 110

Phe Ile Ala Thr Gln Gly Pro Leu Pro Asn Thr Leu Lys Asp Phe Trp  
115 120 125

Arg Met Val Trp Glu Lys Asn Val Tyr Ala Ile Ile Met Leu Thr Lys  
130 135 140

Cys Val Glu Gln Gly Arg Thr Lys Cys Glu Tyr Trp Pro Ser Lys  
145 150 155 160

Gln Ala Gln Asp Tyr Gly Asp Ile Thr Val Ala Met Thr Ser Glu Ile  
165 170 175

Val Leu Pro Glu Trp Thr Ile Arg Asp Phe Thr Val Lys Asn Ile Gln  
180 185 190

Thr Ser Glu Ser His Pro Leu Arg Gln Phe His Phe Thr Ser Trp Pro  
195 200 205

Asp His Gly Val Pro Asp Thr Thr Asp Leu Leu Ile Asn Phe Arg Tyr  
210 215 220

Leu Val Arg Asp Tyr Met Lys Gln Ser Pro Pro Glu Ser Pro Ile Leu  
225 230 235 240

Val His Cys Ser Ala Gly Val Gly Arg Thr Gly Thr Phe Ile Ala Ile  
245 250 255

Asp Arg Leu Ile Tyr Gln Ile Glu Asn Glu Asn Thr Val Asp Val Tyr  
260 265 270

Gly Ile Val Tyr Asp Leu Arg Met His Arg Pro Leu Met Val Gln Thr  
275 280 285

Glu Asp Gln Tyr Val Phe Leu Asn Gln Cys Val Leu Asp Ile Val Arg  
290 295 300

Ser Gln Lys Asp Ser Lys Val Asp Leu Ile Tyr Gln Asn Thr Thr Ala  
305 310 315 320

Met Thr Ile Tyr Glu Asn Leu Ala Pro Val Thr Thr Phe Gly Lys Thr  
325            330            335

Asn Gly Tyr Ile Ala  
340

<210> 50  
<211> 4  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Intracellular retention signal

<220>  
<221> misc\_feature  
<222> (4)..(4)  
<223> Xaa can be any naturally occurring amino acid

<400> 50

Asn Pro Phe Xaa  
1

<210> 51  
<211> 4  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Intracellular retention signal

<400> 51

Lys Asp Glu Leu  
1

<210> 52  
<211> 4

<212> PRT

<213> Artificial Sequence

<220>

<223> Intracellular retention signal

<220>

<221> misc\_feature

<222> (3)..(4)

<223> Xaa can be any naturally occurring amino acid

<400> 52

Lys Lys Xaa Xaa

1

<210> 53

<211> 5

<212> PRT

<213> Artificial Sequence

<220>

<223> Intracellular retention signal

<220>

<221> misc\_feature

<222> (2)..(2)

<223> Xaa can be any naturally occurring amino acid

<220>

<221> misc\_feature

<222> (4)..(5)

<223> Xaa can be any naturally occurring amino acid

<400> 53

Lys Xaa Lys Xaa Xaa

1 5

<210> 54

<211> 4

<212> PRT

<213> Artificial Sequence

<220>

<223> Intracellular retention signal

<220>

<221> misc\_feature

<222> (3)..(3)

<223> Xaa can be any naturally occurring amino acid

<400> 54

Asn Pro Xaa Tyr

1

<210> 55

<211> 4

<212> PRT

<213> Artificial Sequence

<220>

<223> Intracellular retention signal

<220>

<221> MISC\_FEATURE

<222> (2)..(2)

<223> Xaa can be any naturally occurring amino acid.

<220>

<221> MISC\_FEATURE

<222> (3)..(3)

<223> Xaa can be any naturally occurring amino acid.

<220>

<221> MISC\_FEATURE

<222> (4)..(4)

<223> Xaa is an amino acid with a bulky hydrophobic side chain.

<400> 55

Tyr Xaa Xaa Xaa

1

<210> 56  
<211> 8  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Intracellular retention signal

<220>  
<221> misc\_feature  
<222> (3)..(5)  
<223> Xaa can be any naturally occurring amino acid

<400> 56

Asp Glu Xaa Xaa Xaa Leu Leu Ile  
1 5

<210> 57  
<211> 5  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Intracellular retention signal

<220>  
<221> misc\_feature  
<222> (2)..(3)  
<223> Xaa can be any naturally occurring amino acid

<400> 57

Asp Xaa Xaa Leu Leu  
1 5

<210> 58  
<211> 4

<212> PRT

<213> Artificial Sequence

<220>

<223> Intracellular retention signal

<400> 58

Asp Pro Phe Trp

1

<210> 59

<211> 5

<212> PRT

<213> Artificial Sequence

<220>

<223> Intracellular retention signal

<220>

<221> misc\_feature

<222> (2)..(2)

<223> Xaa can be any naturally occurring amino acid

<220>

<221> misc\_feature

<222> (4)..(4)

<223> Xaa can be any naturally occurring amino acid

<400> 59

Phe Xaa Asp Xaa Phe

1 5

<210> 60

<211> 6

<212> PRT

<213> Artificial Sequence

<220>

<223> Intracellular retention signal

<220>  
<221> MISC\_FEATURE  
<222> (2)..(2)  
<223> Xaa is an amino acid with a bulky hydrophobic side chain.

<220>  
<221> MISC\_FEATURE  
<222> (3)..(3)  
<223> Xaa can be any naturally occurring amino acid.

<220>  
<221> MISC\_FEATURE  
<222> (4)..(4)  
<223> Xaa is an amino acid with a bulky hydrophobic side chain.

<400> 60

Leu Xaa Xaa Xaa Asp Glu  
1 5

<210> 61  
<211> 5  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Intracellular retention signal

<400> 61

Leu Leu Asp Leu Leu  
1 5

<210> 62  
<211> 5  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Intracellular retention signal

<400> 62

Pro Trp Asp Leu Trp

1 5

<210> 63

<211> 4

<212> PRT

<213> Artificial Sequence

<220>

<223> Furin cleavage site

<220>

<221> MISC\_FEATURE

<222> (2)..(2)

<223> Xaa can be any naturally occurring amino acid.

<220>

<221> MISC\_FEATURE

<222> (3)..(3)

<223> Xaa can be ARG or LYS

<400> 63

Arg Xaa Xaa Arg

1

<210> 64

<211> 7

<212> PRT

<213> Artificial Sequence

<220>

<223> Tobacco Etch Virus (TEV) cleavage site

<400> 64

Glu Asn Leu Tyr Phe Gln Ser

1 5

<210> 65

<211> 8  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Cleavage site

<220>  
<221> misc\_feature  
<222> (2)..(2)  
<223> Xaa can be any naturally occurring amino acid

<220>  
<221> misc\_feature  
<222> (4)..(4)  
<223> Xaa can be any naturally occurring amino acid

<400> 65

Asp Xaa Glu Xaa Asn Pro Gly Pro  
1 5

<210> 66  
<211> 8  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Cleavage site

<220>  
<221> MISC\_FEATURE  
<222> (2)..(2)  
<223> Xaa may be selected from the following group:ILE, VAL, MET and  
SER

<220>  
<221> MISC\_FEATURE  
<222> (4)..(4)  
<223> Xaa may be selected from the following group: THR, MET, SER, LEU,  
GLU, GLN and PHE.

<400> 66

Asp Xaa Glu Xaa Asn Pro Gly Pro

1 5

<210> 67

<211> 4

<212> PRT

<213> Artificial Sequence

<220>

<223> Endoplasmic reticulum (ER) retention signal

<220>

<221> misc\_feature

<222> (2)..(2)

<223> Xaa can be any naturally occurring amino acid

<400> 67

Arg Xaa Arg Arg

1

<210> 68

<211> 12

<212> PRT

<213> Homo sapiens

<400> 68

Ile Asn Phe Asp Asn Pro Val Tyr Gln Lys Thr Thr

1 5 10

<210> 69

<211> 12

<212> PRT

<213> Homo sapiens

<400> 69

Val Glu Ile Gly Asn Pro Thr Tyr Lys Met Tyr Glu

1 5 10

<210> 70  
<211> 12  
<212> PRT  
<213> Homo sapiens

<400> 70

Thr Asn Phe Thr Asn Pro Val Tyr Ala Thr Leu Tyr

1 5 10

<210> 71  
<211> 12  
<212> PRT  
<213> Drosophila sp.

<400> 71

Gly Asn Phe Ala Asn Pro Val Tyr Glu Ser Met Tyr

1 5 10

<210> 72  
<211> 12  
<212> PRT  
<213> Caenorhabditis elegans

<400> 72

Thr Thr Phe Thr Asn Pro Val Tyr Glu Leu Glu Asp

1 5 10

<210> 73  
<211> 12  
<212> PRT  
<213> Caenorhabditis elegans

<400> 73

Leu Arg Val Asp Asn Pro Leu Tyr Asp Pro Asp Ser

1 5 10

<210> 74  
<211> 12  
<212> PRT  
<213> Homo sapiens

<400> 74

Ile Ile Phe Glu Asn Pro Met Tyr Ser Ala Arg Asp

1 5 10

<210> 75  
<211> 12  
<212> PRT  
<213> Homo sapiens

<400> 75

Thr Asn Phe Glu Asn Pro Ile Tyr Ala Gln Met Glu

1 5 10

<210> 76  
<211> 12  
<212> PRT  
<213> Homo sapiens

<400> 76

Asp Thr Gly Glu Asn Pro Ile Tyr Lys Ser Ala Val

1 5 10

<210> 77  
<211> 11  
<212> PRT  
<213> Homo sapiens

<400> 77

Thr Thr Val Val Asn Pro Lys Tyr Glu Gly Lys

1 5 10

<210> 78  
<211> 12  
<212> PRT  
<213> Drosophila sp.

<400> 78

Trp Asp Thr Glu Asn Pro Ile Tyr Lys Gln Ala Thr  
1 5 10

<210> 79  
<211> 11  
<212> PRT  
<213> Drosophila sp.

<400> 79

Ser Thr Phe Lys Asn Pro Met Tyr Ala Gly Lys  
1 5 10

<210> 80  
<211> 12  
<212> PRT  
<213> Homo sapiens

<400> 80

His Gly Tyr Glu Asn Pro Thr Tyr Arg Phe Leu Glu  
1 5 10

<210> 81  
<211> 12  
<212> PRT  
<213> Homo sapiens

<400> 81

Asn Gly Tyr Glu Asn Pro Thr Tyr Lys Phe Phe Glu  
1 5 10

<210> 82  
<211> 12  
<212> PRT  
<213> Drosophila sp.

<400> 82

Asn Gly Tyr Glu Asn Pro Thr Tyr Lys Tyr Phe Glu  
1 5 10

<210> 83  
<211> 12  
<212> PRT  
<213> Homo sapiens

<400> 83

Tyr Ala Ser Ser Asn Pro Glu Tyr Leu Ser Ala Ser  
1 5 10

<210> 84  
<211> 12  
<212> PRT  
<213> Homo sapiens

<400> 84

Gly Ser Val Gln Asn Pro Val Tyr His Asn Gln Pro  
1 5 10

<210> 85  
<211> 12  
<212> PRT  
<213> Homo sapiens

<400> 85

Thr Ala Val Gly Asn Pro Glu Tyr Leu Asn Thr Val  
1 5 10

<210> 86

<211> 12  
<212> PRT  
<213> Homo sapiens

<400> 86

Ile Ser Leu Asp Asn Pro Asp Tyr Gln Gln Asp Phe  
1 5 10

<210> 87  
<211> 11  
<212> PRT  
<213> Homo sapiens

<400> 87

Arg Lys Arg Ser His Ala Gly Tyr Gln Thr Ile  
1 5 10

<210> 88  
<211> 10  
<212> PRT  
<213> Homo sapiens

<400> 88

Lys His His His Ala Gly Tyr Glu Gln Phe  
1 5 10

<210> 89  
<211> 11  
<212> PRT  
<213> Gallus gallus

<400> 89

Lys Lys His His Asn Thr Gly Tyr Glu Gln Phe  
1 5 10

<210> 90  
<211> 11

<212> PRT

<213> Gallus gallus

<400> 90

Arg Arg Lys Ser Arg Thr Gly Tyr Gln Ser Val

1 5 10

<210> 91

<211> 11

<212> PRT

<213> Gallus gallus

<400> 91

Arg Arg Lys Ser Tyr Ala Gly Tyr Gln Thr Leu

1 5 10

<210> 92

<211> 12

<212> PRT

<213> Drosophila sp.

<400> 92

Arg Arg Arg Ser Thr Ser Arg Gly Tyr Met Ser Phe

1 5 10

<210> 93

<211> 11

<212> PRT

<213> Lumbricus terrestris

<400> 93

Arg Lys Arg Ser Arg Arg Gly Tyr Glu Ser Val

1 5 10

<210> 94

<211> 10

<212> PRT

<213> Homo sapiens

<400> 94

Lys Ser Ile Arg Ser Gly Tyr Glu Val Met

1 5 10

<210> 95

<211> 12

<212> PRT

<213> Homo sapiens

<400> 95

His Cys Gly Gly Pro Arg Pro Gly Tyr Glu Thr Leu

1 5 10

<210> 96

<211> 12

<212> PRT

<213> Mus musculus

<400> 96

His Cys Arg Thr Arg Arg Ala Glu Tyr Glu Thr Leu

1 5 10

<210> 97

<211> 10

<212> PRT

<213> Homo sapiens

<400> 97

Arg Arg Arg Pro Ser Ala Tyr Gln Ala Leu

1 5 10

<210> 98

<211> 9

<212> PRT

<213> Homo sapiens

<400> 98

Arg Arg Arg Ser Tyr Gln Asn Ile Pro

1 5

<210> 99

<211> 10

<212> PRT

<213> Homo sapiens

<400> 99

Lys Lys His Cys Ser Tyr Gln Asp Ile Leu

1 5 10

<210> 100

<211> 10

<212> PRT

<213> Mus musculus

<400> 100

Arg Arg Arg Ser Ala Tyr Gln Asp Ile Arg

1 5 10

<210> 101

<211> 11

<212> PRT

<213> Rattus norvegicus

<400> 101

Arg Lys Arg Arg Arg Ser Tyr Gln Asp Ile Met

1 5 10

<210> 102

<211> 13

<212> PRT

<213> Rattus norvegicus

<400> 102

Lys Phe Cys Lys Ser Lys Glu Arg Asn Tyr His Thr Leu  
1 5 10

<210> 103

<211> 14

<212> PRT

<213> Drosophila sp.

<400> 103

Lys Phe Tyr Lys Ala Arg Asn Glu Arg Asn Tyr His Thr Leu  
1 5 10

<210> 104

<211> 14

<212> PRT

<213> Homo sapiens

<400> 104

Lys Ile Arg Leu Arg Cys Gln Ser Ser Gly Tyr Gln Arg Ile  
1 5 10

<210> 105

<211> 14

<212> PRT

<213> Mus musculus

<400> 105

Lys Ile Arg Gln Arg His Gln Ser Ser Ala Tyr Gln Arg Ile  
1 5 10

<210> 106

<211> 15

<212> PRT

<213> Homo sapiens

<400> 106

His Phe Cys Leu Tyr Arg Lys Arg Pro Gly Tyr Asp Gln Leu Asn  
1 5 10 15

<210> 107  
<211> 11  
<212> PRT  
<213> Homo sapiens

<400> 107

Ser Leu Ser Arg Gly Ser Gly Tyr Lys Glu Ile  
1 5 10

<210> 108  
<211> 13  
<212> PRT  
<213> Homo sapiens

<400> 108

Arg Arg Leu Arg Lys Gly Tyr Thr Pro Leu Met Glu Thr  
1 5 10

<210> 109  
<211> 14  
<212> PRT  
<213> Homo sapiens

<400> 109

Arg Arg Ala Gly His Ser Ser Tyr Thr Pro Leu Pro Gly Ser  
1 5 10

<210> 110  
<211> 18  
<212> PRT  
<213> Dictyostelium discoideum

<400> 110

Lys Lys Leu Arg Gln Gln Lys Gln Gln Gly Tyr Gln Ala Ile Ile Asn

1 5 10 15

Asn Glu

<210> 111

<211> 15

<212> PRT

<213> Dictyostelium discoideum

<400> 111

Arg Ser Lys Ser Asn Gln Asn Gln Ser Tyr Asn Leu Ile Gln Leu

1 5 10 15

<210> 112

<211> 17

<212> PRT

<213> Dictyostelium discoideum

<400> 112

Arg Lys Thr Phe Tyr Asn Asn Asn Gln Tyr Asn Gly Tyr Asn Ile Ile

1 5 10 15

Asn

<210> 113

<211> 10

<212> PRT

<213> Homo sapiens

<400> 113

Pro Leu Ser Tyr Thr Arg Phe Ser Leu Ala

1 5 10

<210> 114  
<211> 11  
<212> PRT  
<213> Homo sapiens

<400> 114

Met Thr Lys Glu Tyr Gln Asp Leu Gln His Leu  
1 5 10

<210> 115  
<211> 10  
<212> PRT  
<213> Homo sapiens

<400> 115

Ser Tyr Lys Tyr Ser Lys Val Asn Lys Glu  
1 5 10

<210> 116  
<211> 10  
<212> PRT  
<213> Homo sapiens

<400> 116

Pro Ala Ala Tyr Arg Gly Val Gly Asp Asp  
1 5 10

<210> 117  
<211> 10  
<212> PRT  
<213> Homo sapiens

<400> 117

Thr Gly Val Tyr Val Lys Met Pro Pro Thr  
1 5 10

<210> 118

<211> 10  
<212> PRT  
<213> Homo sapiens

<400> 118

Leu Ile Ser Tyr Lys Gly Leu Pro Pro Glu  
1 5 10

<210> 119  
<211> 11  
<212> PRT  
<213> Rattus norvegicus

<400> 119

Ala Ser Asp Tyr Gln Arg Leu Asn Leu Lys Leu  
1 5 10

<210> 120  
<211> 11  
<212> PRT  
<213> Human immunodeficiency virus type 1

<400> 120

Arg Gln Gly Tyr Ser Pro Leu Ser Phe Gln Thr  
1 5 10

<210> 121  
<211> 19  
<212> PRT  
<213> Homo sapiens

<400> 121

Arg Met Gln Ala Gln Pro Pro Gly Tyr Arg His Val Ala Asp Gly Glu  
1 5 10 15

Asp His Ala

<210> 122  
<211> 9  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Dileucine-based sorting signal

<220>  
<221> misc\_feature  
<222> (3)..(5)  
<223> Xaa can be any naturally occurring amino acid

<400> 122

Asp Glu Xaa Xaa Xaa Leu Leu Leu Ile  
1 5

<210> 123  
<211> 9  
<212> PRT  
<213> Homo sapiens

<400> 123

Ser Asp Lys Gln Thr Leu Leu Pro Asn  
1 5

<210> 124  
<211> 9  
<212> PRT  
<213> Rattus norvegicus

<400> 124

Asp Glu Arg Ala Pro Leu Ile Arg Thr  
1 5

<210> 125

<211> 9  
<212> PRT  
<213> Homo sapiens

<400> 125

Gln Glu Lys Asp Pro Leu Leu Lys Asn  
1 5

<210> 126  
<211> 9  
<212> PRT  
<213> Coturnix coturnix

<400> 126

Thr Glu Arg Asn Pro Leu Leu Lys Ser  
1 5

<210> 127  
<211> 9  
<212> PRT  
<213> Homo sapiens

<400> 127

Gly Glu Asn Ser Pro Leu Leu Ser Gly  
1 5

<210> 128  
<211> 9  
<212> PRT  
<213> Homo sapiens

<400> 128

Glu Glu Lys Gln Pro Leu Leu Met Glu  
1 5

<210> 129  
<211> 9

<212> PRT

<213> Oryzias latipes

<400> 129

Gly Glu Arg Gln Pro Leu Leu Gln Ser

1 5

<210> 130

<211> 9

<212> PRT

<213> Gallus gallus

<400> 130

Pro Glu Ile Gln Pro Leu Leu Thr Glu

1 5

<210> 131

<211> 9

<212> PRT

<213> Carassius auratus

<400> 131

Glu Gly Arg Gln Pro Leu Leu Gly Asp

1 5

<210> 132

<211> 9

<212> PRT

<213> Homo sapiens

<400> 132

Glu Ala Asn Gln Pro Leu Leu Thr Asp

1 5

<210> 133

<211> 9

<212> PRT

<213> Gallus gallus

<400> 133

Glu Leu His Gln Pro Leu Leu Thr Asp

1 5

<210> 134

<211> 9

<212> PRT

<213> Danio rerio

<400> 134

Arg Glu Phe Glu Pro Leu Leu Asn Ala

1 5

<210> 135

<211> 9

<212> PRT

<213> Homo sapiens

<400> 135

Glu Glu Lys Met Ala Ile Leu Met Asp

1 5

<210> 136

<211> 9

<212> PRT

<213> Homo sapiens

<400> 136

Glu Glu Lys Leu Ala Ile Leu Ser Gln

1 5

<210> 137

<211> 9

<212> PRT

<213> Mus musculus

<400> 137

Ser Glu Arg Asp Val Leu Leu Asp Glu

1 5

<210> 138

<211> 9

<212> PRT

<213> Homo sapiens

<400> 138

Ser Glu Arg Arg Asn Leu Leu Glu Asp

1 5

<210> 139

<211> 9

<212> PRT

<213> Rattus norvegicus

<400> 139

Asp Asp Ser Gly Asp Leu Leu Pro Gly

1 5

<210> 140

<211> 9

<212> PRT

<213> Homo sapiens

<400> 140

Ser Gln Ile Lys Arg Leu Leu Ser Glu

1 5

<210> 141

<211> 9

<212> PRT

<213> Felis catus

<400> 141

Ser His Ile Lys Arg Leu Leu Ser Glu  
1 5

<210> 142

<211> 9

<212> PRT

<213> *Mus musculus*

<400> 142

Arg Arg Thr Pro Ser Leu Leu Glu Gln  
1 5

<210> 143

<211> 9

<212> PRT

<213> *Homo sapiens*

<400> 143

His Arg Thr Pro Ser Leu Leu Glu Gln  
1 5

<210> 144

<211> 10

<212> PRT

<213> *Rattus norvegicus*

<400> 144

Glu Pro Arg Gly Ser Arg Leu Leu Val Arg  
1 5 10

<210> 145

<211> 19

<212> PRT

<213> *Homo sapiens*

<400> 145

Met Asp Asp Gln Arg Asp Leu Ile Ser Asn Asn Glu Gln Leu Pro Met  
1 5 10 15

Leu Gly Arg

<210> 146  
<211> 19  
<212> PRT  
<213> *Mus musculus*

<400> 146

Met Asp Asp Gln Arg Asp Leu Ile Ser Asn His Glu Gln Leu Pro Ile  
1 5 10 15

Leu Gly Asn

<210> 147  
<211> 20  
<212> PRT  
<213> *Gallus gallus*

<400> 147

Met Ala Glu Glu Gln Arg Asp Leu Ile Ser Ser Asp Gly Ser Ser Gly  
1 5 10 15

Val Leu Pro Ile  
20

<210> 148  
<211> 23  
<212> PRT  
<213> *Danio rerio*

<400> 148

Met Glu Pro Asp His Gln Asn Glu Ser Leu Ile Gln Arg Val Pro Ser  
1 5 10 15

Ala Glu Thr Ile Leu Gly Arg  
20

<210> 149  
<211> 23  
<212> PRT  
<213> Danio rerio

<400> 149

Met Ser Ser Glu Gly Asn Glu Thr Pro Leu Ile Ser Asp Gln Ser Ser  
1 5 10 15

Val Asn Met Gly Pro Gln Pro  
20

<210> 150  
<211> 24  
<212> PRT  
<213> Trypanosoma brucei

<400> 150

Arg Pro Arg Arg Arg Thr Glu Glu Asp Glu Leu Leu Pro Glu Glu Ala  
1 5 10 15

Glu Gly Leu Ile Asp Pro Gln Asn  
20

<210> 151  
<211> 19  
<212> PRT  
<213> Homo sapiens

<400> 151

Pro Asp Lys His Ser Leu Leu Val Gly Asp Phe Arg Glu Asp Asp Asp  
1 5 10 15

Thr Ala Leu

<210> 152  
<211> 9  
<212> PRT  
<213> Homo sapiens

<400> 152

Thr Glu Arg Glu Arg Leu Leu Asn Phe  
1 5

<210> 153  
<211> 8  
<212> PRT  
<213> Homo sapiens

<400> 153

Val Glu Thr Asp Asp Leu Ile Leu  
1 5

<210> 154  
<211> 7  
<212> PRT  
<213> Caenorhabditis elegans

<400> 154

Phe Glu Asn Asp Ser Leu Leu  
1 5

<210> 155  
<211> 9  
<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 155

Asn Glu Gln Ser Pro Leu Leu His Asn

1 5

<210> 156

<211> 8

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 156

Ser Glu Gln Thr Arg Leu Val Pro

1 5

<210> 157

<211> 8

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 157

Glu Val Asp Leu Asp Leu Leu Lys

1 5

<210> 158

<211> 5

<212> PRT

<213> Artificial Sequence

<220>

<223> Dileucine-based sorting signal

<220>

<221> misc\_feature

<222> (2)..(3)

<223> Xaa can be any naturally occurring amino acid

<400> 158

Asp Xaa Xaa Leu Leu

1 5

<210> 159

<211> 13

<212> PRT

<213> Homo sapiens

<400> 159

Ser Phe His Asp Asp Ser Asp Glu Asp Leu Leu His Ile

1 5 10

<210> 160

<211> 13

<212> PRT

<213> Bos taurus

<400> 160

Thr Phe His Asp Asp Ser Asp Glu Asp Leu Leu His Val

1 5 10

<210> 161

<211> 13

<212> PRT

<213> Oryctolagus cuniculus

<400> 161

Ser Phe His Asp Asp Ser Asp Glu Asp Leu Leu Asn Ile

1 5 10

<210> 162

<211> 13

<212> PRT

<213> Gallus gallus

<400> 162

Ser Phe His Asp Asp Ser Asp Glu Asp Leu Leu Asn Val

1 5 10

<210> 163

<211> 13

<212> PRT

<213> Homo sapiens

<400> 163

Glu Glu Ser Glu Glu Arg Asp Asp His Leu Leu Pro Met

1 5 10

<210> 164

<211> 13

<212> PRT

<213> Gallus gallus

<400> 164

Asp Glu Ser Glu Glu Arg Asp Asp His Leu Leu Pro Met

1 5 10

<210> 165

<211> 12

<212> PRT

<213> Homo sapiens

<400> 165

Gly Tyr His Asp Asp Ser Asp Glu Asp Leu Leu Glu

1 5 10

<210> 166

<211> 13

<212> PRT

<213> Homo sapiens

<400> 166

Ile Thr Gly Phe Ser Asp Asp Val Pro Met Val Ile Ala

1 5 10

<210> 167

<211> 13

<212> PRT

<213> *Hydra vulgaris*

<400> 167

Ile Asn Arg Phe Ser Asp Asp Glu Pro Leu Val Val Ala

1 5 10

<210> 168

<211> 13

<212> PRT

<213> *Homo sapiens*

<400> 168

Met Leu Glu Ala Ser Asp Asp Glu Ala Leu Leu Val Cys

1 5 10

<210> 169

<211> 13

<212> PRT

<213> *Homo sapiens*

<400> 169

Lys Asn Glu Thr Ser Asp Asp Glu Ala Leu Leu Leu Cys

1 5 10

<210> 170

<211> 12

<212> PRT

<213> *Mus musculus*

<400> 170

Trp Val Val Glu Ala Glu Asp Glu Pro Leu Leu Ala

1 5 10

<210> 171  
<211> 12  
<212> PRT  
<213> Homo sapiens

<400> 171

Trp Val Ala Glu Ala Glu Asp Glu Pro Leu Leu Thr  
1 5 10

<210> 172  
<211> 12  
<212> PRT  
<213> Homo sapiens

<400> 172

His Asp Asp Phe Ala Asp Asp Ile Ser Leu Leu Lys  
1 5 10

<210> 173  
<211> 13  
<212> PRT  
<213> Mus musculus

<400> 173

Gly Arg Asp Ser Pro Glu Asp His Ser Leu Leu Val Asn  
1 5 10

<210> 174  
<211> 12  
<212> PRT  
<213> Peromyscus maniculatus

<400> 174

Val Arg Cys His Pro Glu Asp Asp Arg Leu Leu Gly  
1 5 10

<210> 175  
<211> 13  
<212> PRT  
<213> Homo sapiens

<400> 175

His Arg Val Ser Gln Asp Asp Leu Asp Leu Leu Thr Ser  
1 5 10

<210> 176  
<211> 13  
<212> PRT  
<213> Homo sapiens

<400> 176

Ala Ser Val Ser Leu Leu Asp Asp Glu Leu Met Ser Leu  
1 5 10

<210> 177  
<211> 13  
<212> PRT  
<213> Homo sapiens

<400> 177

Ala Ser Ser Gly Leu Asp Asp Leu Asp Leu Leu Gly Lys  
1 5 10

<210> 178  
<211> 13  
<212> PRT  
<213> Homo sapiens

<400> 178

Val Gln Asn Pro Ser Ala Asp Arg Asn Leu Leu Asp Leu  
1 5 10

<210> 179  
<211> 13  
<212> PRT  
<213> Homo sapiens

<400> 179

Asn Ala Leu Ser Trp Leu Asp Glu Glu Leu Leu Cys Leu  
1 5 10

<210> 180  
<211> 13  
<212> PRT  
<213> Drosophila sp.

<400> 180

Thr Val Asp Ser Ile Asp Asp Val Pro Leu Leu Ser Asp  
1 5 10

<210> 181  
<211> 13  
<212> PRT  
<213> Mus musculus

<400> 181

Gln Glu Glu Cys Pro Ser Asp Ser Glu Glu Asp Glu Gly  
1 5 10

<210> 182  
<211> 13  
<212> PRT  
<213> Mus musculus

<400> 182

Arg Asp Arg Asp Tyr Asp Glu Asp Asp Glu Asp Asp Ile  
1 5 10

<210> 183

<211> 15  
<212> PRT  
<213> Mus musculus

<400> 183

Leu Asp Glu Thr Glu Asp Asp Glu Leu Glu Tyr Asp Asp Glu Ser  
1 5 10 15

<210> 184  
<211> 10  
<212> PRT  
<213> Homo sapiens

<400> 184

Lys Asp Pro Asp Glu Val Glu Thr Glu Ser  
1 5 10

<210> 185  
<211> 13  
<212> PRT  
<213> Homo sapiens

<400> 185

His Glu Phe Gln Asp Glu Thr Asp Thr Glu Glu Glu Thr  
1 5 10

<210> 186  
<211> 14  
<212> PRT  
<213> Homo sapiens

<400> 186

Gln Glu Lys Glu Asp Asp Gly Ser Glu Ser Glu Glu Glu Tyr  
1 5 10

<210> 187  
<211> 9

<212> PRT

<213> Homo sapiens

<400> 187

Gly Glu Asp Glu Glu Ser Glu Ser Asp

1 5

<210> 188

<211> 12

<212> PRT

<213> Homo sapiens

<400> 188

Gly Glu Asp Ser Asp Glu Glu Pro Asp His Glu Glu

1 5 10

<210> 189

<211> 11

<212> PRT

<213> Homo sapiens

<400> 189

Leu Glu Asp Asp Ser Asp Glu Glu Glu Asp Phe

1 5 10

<210> 190

<211> 9

<212> PRT

<213> Human cytomegalovirus

<400> 190

Lys Asp Ser Asp Glu Glu Asn Val

1 5

<210> 191

<211> 13

<212> PRT

<213> Human herpesvirus 3

<400> 191

Phe Glu Asp Ser Glu Ser Thr Asp Thr Glu Glu Glu Phe

1 5 10

<210> 192

<211> 9

<212> PRT

<213> Human immunodeficiency virus type 1

<400> 192

Leu Glu Ala Gln Glu Glu Glu Glu Val

1 5

<210> 193

<211> 22

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 193

Ala Asp Asp Leu Glu Ser Gly Leu Gly Ala Glu Asp Asp Leu Glu Gln

1 5 10 15

Asp Glu Gln Leu Glu Gly

20

<210> 194

<211> 12

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 194

Thr Glu Ile Asp Glu Ser Phe Glu Met Thr Asp Phe

1 5 10

<210> 195  
<211> 18  
<212> PRT  
<213> *Saccharomyces cerevisiae*

<400> 195

Thr Glu Pro Glu Glu Val Glu Asp Phe Asp Asp Phe Asp Leu Ser Asp Glu  
1 5 10 15

Asp His

<210> 196  
<211> 15  
<212> PRT  
<213> *Saccharomyces cerevisiae*

<400> 196

Phe Glu Ile Glu Glu Asp Asp Val Pro Thr Leu Glu Glu His  
1 5 10 15

<210> 197  
<211> 4  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Intracellular retention signal

<400> 197

His Asp Glu Leu  
1