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

(19) World Intellectual Property
Organization
International Bureau

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

WIPO | PCT

(10) International Publication Number
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
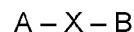
in which

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.

3

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 lysozomal, 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 KKX'X' or a KX'KX'X' signal (wherein X' is any amino acid).

The intracellular retention signal may comprise a sequence 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, 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 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 molcules; 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 α and/or β 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 :

$$A - X - B - Y - C$$
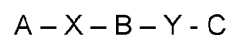
in which

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 Lumenal (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 Lumenal (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-lecuine 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 α-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 re-directed 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 α or β 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. Th 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

11

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

12

present nucleic acid construct comprises a TEV cleavage site and is expressed in a mammalian cell – exogenous TEV protease must also 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 aptho- and cardioviruses is mediated by 2A "cleaving" at its own C-terminus. In apthoviruses, 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 apthovirus 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 (Donelly et al (2001) as above).

The cleavage site of the present invention may comprise the amino acid sequence: Dx_1Ex_2NPGP, where x_1 and x_2 are any amino acid. X_1 may be selected from the following group: I, V, M and S. X_2 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,

13

	EMCD, EMCPV21
DVETNPGP (SEQ ID No. 2)	Picornaviruses MENGO 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 *T. maritima*

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)

LLNFDLLKLAGDVESHPGP (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 aptho- 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)

TRAEIEDELIRAGIESNPGP (SEQ ID No. 25)

TRAEIEDELIRADIESNPGP (SEQ ID No. 26)

AKFQIDKILISGDVELNPGP (SEQ ID No. 27)

SSIIRTKMLVSGDVEENPGP (SEQ ID No. 28)

CDAQRQKLLLSGDIEQNPGP (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 contain 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 hydolases. 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 form proteins such as the CI- and CD-MPRs, sortilin, the LDL-receptor related proteins LRP3 and LRP10 and β-secretase, GGA1-3, LIMP-II, NCP1, mucolipn-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.

19

The intracellular retention signal may be from the adenovirus E19 protein. The intracellular setention signal may be from the protein E3/19K, which is also known as E3gp 19 kDa; E19 or GP19K. The intrcellular 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:

$$A - X - B - Y - C$$
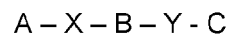
      in which

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 KKX'X' or a KX'KX'X' 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 YX'X'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 β, and β-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-INFDNPVYQKTT-29
LRP1 (1)	Human	Tm-21-VEIGNPTYKMYE-64
LRP1 (2)	Human	Tm-55-TNFTNPVYATLY-33
LRP1	Drosophila	Tm-43-GNFANPVYESMY-38
LRP1 (1)	C. elegans	Tm-54-TTFTNPVYELED-91

LRP1 (2)	*C. elegans*	Tm-140-LRVDNPLYDPDS-4
Megalin (1)	Human	Tm-70-IIFENPMYSARD-125
Megalin (2)	Human	Tm-144-TNFENPIYAQME-53
Integrin ß-1 (1)	Human	Tm-18-DTGENPIYKSAV-11
Integrin ß-1 (2)	Human	Tm-30-TTVVNPKYEGK
Integrin ß (1)	*Drosophila*	Tm-26-WDTENPIYKQAT-11
Integrin ß (2)	*Drosophila*	Tm-35-STFKNPMYAGK
APLP1	Human	Tm-33-HGYENPTYRFLE-3
APP	Human	Tm-32-NGYENPTYKFFE-4
APP-like	*Drosophila*	Tm-38-NGYENPTYKYFE-3
Insulin receptor	Human	Tm-36-YASSNPEYLSAS-379
EGR receptor (1)	Human	Tm-434-GSVQNPVYHNQP-96
EGR receptor (2)	Human	Tm-462-TAVGNPEYLNTV-68
EGR receptor (3)	Human	Tm-496-ISLDNPDYQQDF-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, ß-amyloid precursor protein; APLP1, 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 CI- 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.

23

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

Protein	Species	Sequence
LAMP-1	Human	Tm-RKRSHAGYQTI
LAMP-2a	Human	Tm-KHHHAGYEQF
LAMP-2a	Chicken	Tm-KKHHNTGYEQF
LAMP-2b	Chicken	Tm-RRKSRTGYQSV
LAMP-2c	Chicken	Tm-RRKSYAGYQTL
LAMP	*Drosophila*	Tm-RRRSTSRGYMSF
LAMP	Earthworm	Tm-RKRSRRGYESV
CD63	Human	Tm-KSIRSGYEVM
GMP-17	Human	Tm-HCGGPRPGYEIL
GMP-17	Mouse	Tm-HCRTRRAEYEIL
CD68	Human	Tm-RRRPSAYQAL
CD1b	Human	Tm-RRRSYQNIP
CD1c	Human	Tm-KKHCSYQDIL
CD1d	Mouse	Tm-RRRSAYQDIR
CD1	Rat	Tm-RKRRRSYQDIM
Endolyn	Rat	Tm-KFCKSKERNYHTL
Endolyn	*Drosophila*	Tm-KFYKARNERNYHTL
TSC403	Human	Tm-KIRLRCQSSGYQRI
TSC403	Mouse	Tm-KIRQRHQSSAYQRI
Cystinosin	Human	Tm-HFCLYRKRPGYDQLN
Putative solute carrier	Human	Tm-12-SLSRGSGYKEI
TRP-2	Human	Tm-RRLRKGYTPLMET-11
HLA-DM ◊	Human	Tm-RRAGHSSYTPLPGS-9
LmpA	Dictyostelium	Tm-KKLRQQKQQGYQAIINNE
Putative lysosomal protein	Dictyostelium	Tm-RSKSNQNQSYNLIQL
LIMP-II	Dictyostelium	Tm-RKTFYNNNQYNGYNIIN
Transferrin receptor	Human	16-PLSYTRFSLA-35-Tm
Asialoglycoprotein receptor H1	Human	MTKEYQDLQHL-29-Tm
CI-MPR	Human	Tm-22-SYKYSKVNKE-132
CD-MPR	Human	Tm-40-PAAYRGVGDD-16
CTLA-4	Human	Tm-10-TGVYVKMPPT-16
Furin	Human	Tm-17-LISYKGLPPE-29
TGN38	Rat	Tm-23-ASDYQRLNLKL
gp41	HIV-1	Tm-13-RQGYSPLSFQT-144
Acid phosphatase	Human	Tm-RMQAQPPGYRHVADGEDHA

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.

24

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
VAchT	Mouse	Tm-10-SERDVLLDE-42
VAMP4	Human	19-SERRNLLED-88-Tm
Neonatal FcR	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	MAEEQRDLISSDGSSGVLPI-12-Tm
Ii-1	Zebrafish	MEPDHQNESLIQRVPSAETILGR-12-Tm
Ii-2	Zebrafish	MSSEGNETPLISDQSSVNMGPQP-8-Tm
Lamp	Trypanosome	Tm-RPRRRIEEDELLPEEAEGLIDPQN
Menkes protein	Human	Tm-74-PDKHSLLVGDFREDDDTAL
NPC1	Human	Tm-13-TERERLLNF
AQP4	Human	Tm-32-VETDDLIL-29
RME-2	C. elegans	Tm-104-FENDSLL
Vam3p	S. cerevisiae	153-NEQSPLLHN-121-Tm
ALP	S. cerevisiae	7-SEQTRLVP-18-Tm
Gap1p	S. cerevisiae	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 CI- 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.

<u>Table 5 - DX'X'LL sorting signals</u>

Protein	Species	Sequence
CI-MPR	Human	Tm-151-SFHDDSDEDLLHI
CI-MPR	Bovine	Tm-150-TFHDDSDEDLLHV
CI-MPR	Rabbit	Tm-151-SFHDDSDEDLLNI
CI-MPR	Chicken	Tm-148-SFHDDSDEDLLNV
CD-MPR	Human	Tm-54-EESEERDDHLLPM
CD-MPR	Chicken	Tm-54-DESEERDDHLLPM
Sortilin	Human	Tm-41-GYHDDSDEDLLE
SorLA	Human	Tm-41-ITGFSDDVPMVIA
Head-activator BP	Hydra	Tm-41-INRFSDDEPLVVA
LRP3	Human	Tm-237-MLEASDDEALLVC
ST7	Human	Tm-330-KNETSDDEALLLC
LRP10	Mouse	Tm-235-WVVEAEDEPLLA
LRP10	Human	Tm-237-WVAEAEDEPLLT
Beta-secretase	Human	Tm-9-HDDFADDISLLK
Mucolipin-1	Mouse	Tm-43-GRDSPEDHSLLVN
Nonclassical MHC-I	Deer mouse	Tm-6-VRCHPEDDRLLG
FLJ30532	Human	Tm-83-HRVSQDDLDLLTS
GGA1	Human	350-ASVSLLDDELMSL-275
GGA1	Human	415-ASSGLDDLDLLGK-211
GGA2	Human	408-VQNPSADRNLLDL-192
GGA3	Human	384-NALSWLDEELLCL-326
GGA	*Drosophila*	447-TVDSIDDVPLLSD-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-QEECPSDSEEDEG-14
PC6B (1)[a]	Mouse	Tm-39-RDRDYDEDDEDDI-36
PC6B (2)	Mouse	Tm-69-LDETEDDELEYDDES-4
PC7	Human	Tm-38-KDPDEVETES-47
CPD	Human	Tm-36-HEFQDETDTEEET-6
PAM	Human	Tm-59-QEKEDDGSESEEEY-12
VMAT2	Human	Tm-35-GEDEESESD
VMAT1	Human	Tm-35-GEDSDEEPDHEE
VAMP4	Human	25-LEDDSDEEEDF-81-Tm
Glycoprotein B	HCMV	Tm-125-KDSDEEENV
Glycoprotein E	Herpes virus 3	Tm-28-FEDSESTDTEEEF-21
Nef	HIV-1 (AAL65476)	55-LEAQEEEEV-139
Kex1p (1)	S. cerevisiae	Tm-29-ADDLESGLGAEDDLEQDEQLEG-40
Kex1p (2)	S. cerevisiae	Tm-79-TEIDESFEMTDF
Kex2p	S. cerevisiae	Tm-36-TEPEEVEDFDFDLSDEDH-61
Vps10p	S. cerevisiae	Tm-112-FEIEEDDVPTLEEEH-37

See legend to Table 1 for explanation of signal format. Serine and threonine residues are underlined.
[a]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 melansomal protein which is retained in the melanosome (a specialized lysosome) at >99% efficiency. TYRP1 is a 537 amino acid transmembrane protein with a lumenal 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:

MSLPVTALLLPLALLLHAARP (SEQ ID No. 38)

scFv aCD19:

DIQMTQTTSSLSASLGDRVTISCRASQDISKYLNWYQQKPDGTVKLLIYHTSRLHSGVPSRF
SGSGSGTDYSLTISNLEQEDIATYFCQQGNTLPYTFGGGTKLEITKAGGGGSGGGGSGGG
GSGGGGSEVKLQESGPGLVAPSQSLSVTCTVSGVSLPDYGVSWIRQPPRKGLEWLGVIW
GSETTYYNSALKSRLTIIKDNSKSQVFLKMNSLQTDDTAIYYCAKHYYYGGSYAMDYWGQG
TSVTVS (SEQ ID No. 39)

Linker:

SD

Human CD8aSTK:

PTTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDFACDI (SEQ ID No. 40)

Human CD28TM:

FWVLVVVGGVLACYSLLVTVAFIIFWV (SEQ ID No. 41)

Human Typr-1 intracellular domain (retention signal):

RARRSMDEANQPLLTDQYQCYAEEYEKLQNPNQSVV (SEQ ID No. 42)

Human CD3zeta intracellular domain:

RRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDKRRGRDPEMGGKPRRKNPQEG
LYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLSTATKDTYDALHMQALPPR (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
GGGSGGGGSRSEVQLVESGGGLVQPGGSLRLSCAASGFTLSNYGMHWIRQAPGKGLEW
VSSISLNGGSTYYRDSVKGRFTISRDNAKSTLYLQMNSLRAEDTAVYYCAAQDAYTGGYFD
YWGQGTLVTVSSM (SEQ ID No. 45)

Linker:

DPA

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

EPKSPDKTHTCPPCPAPPVAGPSVFLFPPKPKDTLMIARTPEVTCVVVDVSHEDPEVKFNW
YVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISK
AKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLD
SDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK (SEQ ID No. 46)

Linker:

KDPK (SEQ ID No. 47)

Human CD148TM:

AVFGCIFGALVIVTVGGFIFW (SEQ ID No. 48)

Human CD148 intracellular domain:

RKKRKDAKNNEVSFSQIKPKKSKLIRVENFEAYFKKQQADSNCGFAEEYEDLKLVGISQPKY
AAELAENRGKNRYNNVLPYDISRVKLSVQTHSTDDYINANYMPGYHSKKDFIATQGPLPNTL
KDFWRMVWEKNVYAIIMLTKCVEQGRTKCEEYWPSKQAQDYGDITVAMTSEIVLPEWTIRD
FTVKNIQTSESHPLRQFHFTSWPDHGVPDTTDLLINFRYLVRDYMKQSPPESPILVHCSAGV
GRTGTFIAIDRLIYQIENENTVDVYGIVYDLRMHRPLMVQTEDQYVFLNQCVLDIVRSQKDSK
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:

MSLPVTALLLPLALLLHAARP (SEQ ID No. 38)

scFv aCD19:

DIQMTQTTSSLSASLGDRVTISCRASQDISKYLNWYQQKPDGTVKLLIYHTSRLHSGVPSRF
SGSGSGTDYSLTISNLEQEDIATYFCQQGNTLPYTFGGGTKLEITKAGGGGSGGGGSGGG
GSGGGGSEVKLQESGPGLVAPSQSLSVTCTVSGVSLPDYGVSWIRQPPRKGLEWLGVIW
GSETTYYNSALKSRLTIIKDNSKSQVFLKMNSLQTDDTAIYYCAKHYYYGGSYAMDYWGQG
TSVTVS (SEQ ID No. 39)

Linker:

SD

Human CD8aSTK:

PTTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDFACDI (SEQ ID No. 40)

Human CD28TM:

FWVLVVVGGVLACYSLLVTVAFIIFWV (SEQ ID No. 41)

Human CD3zeta intracellular domain:

RRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDKRRGRDPEMGGKPRRKNPQEG

LYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLSTATKDTYDALHMQALPPR   (SEQ

ID No. 43)

Human Typr-1 intracellular domain (retention signal):

RARRSMDEANQPLLTDQYQCYAEEYEKLQNPNQSVV (SEQ ID No. 42)

2A peptide:

RAEGRGSLLTCGDVEENPGP (SEQ ID No. 24)

Signal peptide derived from mouse Ig kappa:

MAVPTQVLGLLLLWLTDA (SEQ ID No. 44)

scFv aCD33:

RCDIQMTQSPSSLSASVGDRVTITCRASEDIYFNLVWYQQKPGKAPKLLIYDTNRLADGVPS

RFSGSGSGTQYTLTISSLQPEDFATYYCQHYKNYPLTFGQGTKLEIKRSGGGGSGGGGSG

GGGSGGGGSRSEVQLVESGGGLVQPGGSLRLSCAASGFTLSNYGMHWIRQAPGKGLEW

VSSISLNGGSTYYRDSVKGRFTISRDNAKSTLYLQMNSLRAEDTAVYYCAAQDAYTGGYFD

YWGQGTLVTVSSM (SEQ ID No. 45)

Linker:

DPA

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

EPKSPDKTHTCPPCPAPPVAGPSVFLFPPKPKDTLMIARTPEVTCVVVDVSHEDPEVKFNW

YVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISK

AKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLD

SDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK (SEQ ID No. 46)

Linker:

KDPK (SEQ ID No. 47)

Human CD148TM:

AVFGCIFGALVIVTVGGFIFW (SEQ ID No. 48)

Human CD148 intracellular domain:

RKKRKDAKNNEVSFSQIKPKKSKLIRVENFEAYFKKQQADSNCGFAEEYEDLKLVGISQPKY

AAELAENRGKNRYNNVLPYDISRVKLSVQTHSTDDYINANYMPGYHSKKDFIATQGPLPNTL

KDFWRMVWEKNVYAIIMLTKCVEQGRTKCEEYWPSKQAQDYGDITVAMTSEIVLPEWTIRD
FTVKNIQTSESHPLRQFHFTSWPDHGVPDTTDLLINFRYLVRDYMKQSPPESPILVHCSAGV
GRTGTFIAIDRLIYQIENENTVDVYGIVYDLRMHRPLMVQTEDQYVFLNQCVLDIVRSQKDSK
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:

DIQMTQTTSSLSASLGDRVTISCRASQDISKYLNWYQQKPDGTVKLLIYHTSRLHSGVPSRF
SGSGSGTDYSLTISNLEQEDIATYFCQQGNTLPYTFGGGTKLEITKAGGGGSGGGGSGGG
GSGGGGSEVKLQESGPGLVAPSQSLSVTCTVSGVSLPDYGVSWIRQPPRKGLEWLGVIW
GSETTYYNSALKSRLTIIKDNSKSQVFLKMNSLQTDDTAIYYCAKHYYYGGSYAMDYWGQG
TSVTVS (SEQ ID No. 39)

Linker:

SD

Human CD8aSTK:

PTTTPAPRPPTPAPTIASQPLSLRPEACRPAAGGAVHTRGLDFACDI (SEQ ID No. 40)

Human CD28TM:

FWVLVVVGGVLACYSLLVTVAFIIFWV (SEQ ID No. 41)

Human CD3zeta intracellular domain:

RRVKFSRSADAPAYQQGQNQLYNELNLGRREEYDVLDKRRGRDPEMGGKPRRKNPQEG
LYNELQKDKMAEAYSEIGMKGERRRGKGHDGLYQGLSTATKDTYDALHMQALPPR   (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
GGGSGGGGSRSEVQLVESGGGLVQPGGSLRLSCAASGFTLSNYGMHWIRQAPGKGLEW
VSSISLNGGSTYYRDSVKGRFTISRDNAKSTLYLQMNSLRAEDTAVYYCAAQDAYTGGYFD
YWGQGTLVTVSSM (SEQ ID No. 45)

Linker:

DPA

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

EPKSPDKTHTCPPCPAPPVAGPSVFLFPPKPKDTLMIARTPEVTCVVVDVSHEDPEVKFNW
YVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISK
AKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLD
SDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK (SEQ ID No. 46)

Linker:

KDPK (SEQ ID No. 47)

Human CD148TM:

AVFGCIFGALVIVTVGGFIFW (SEQ ID No. 48)

Human CD148 intracellular domain:

RKKRKDAKNNEVSFSQIKPKKSKLIRVENFEAYFKKQQADSNCGFAEEYEDLKLVGISQPKY
AAELAENRGKNRYNNVLPYDISRVKLSVQTHSTDDYINANYMPGYHSKKDFIATQGPLPNTL
KDFWRMVWEKNVYAIIMLTKCVEQGRTKCEEYWPSKQAQDYGDITVAMTSEIVLPEWTIRD
FTVKNIQTSESHPLRQFHFTSWPDHGVPDTTDLLINFRYLVRDYMKQSPPESPILVHCSAGV
GRTGTFIAIDRLIYQIENENTVDVYGIVYDLRMHRPLMVQTEDQYVFLNQCVLDIVRSQKDSK
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 furthers 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 lysozomal 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 lysozomal 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 decribed 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:
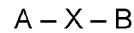
A – X – B

in which

A and B are nucleic acid sequences encoding 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.

2.   A nucleic acid construct according to claim 1 wherein the transmembrane protein comprises an endodomain which comprises the intracellular retention signal.

3.   A nucleic acid construct according to claim 1 or 2 wherein the intracellular retention signal directs the transmembrane protein away from the secretory pathway.

4.   A nucleic acid construct according to any preceding claim wherein the intracellular retention signal directs the transmembrane protein to a membrane-bound intracellular compartment.

5.   A nucleic acid construct according to any preceding claim wherein the intracellular retention signal directs the transmembrane protein to a lysozomal, endosomal or Golgi compartment.

6.   A nucleic acid consruct according to any preceding claim 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.

7.   A nucleic acid construct according to any preceding claim wherein the intracellular retention signal is 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 KKX'X' or a KX'KX'X' signal (wherein X' is any amino acid).

8.   A nucleic acid construct according to any preceding claim wherein the intracellular retention signal comprises a sequence 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.

9. A nucleic acid construct according to any preceding claim wherein the intracellular retention signal comprises any sequence shown in Tables 2 to 6.

10. A nucleic acid construct according to any preceding claim wherein the intracellular retention signal comprises the Tyrosinase-related protein (TYRP)-1 intracellular retention signal or the Adenoviral E3/19K retention signal.

11. A nucleic acid construct according to any preceding claim wherein the intracellular retention signal comprises the sequence NQPLLTD (SEQ ID No. 35), KYKSRRSFIDEKKMP (SEQ ID No. 36); or DEKKMP (SEQ ID No. 37).

12. A nucleic acid construct according to any preceding claim wherein the intracellular retention signal is proximal to a transmembrane domain of the transmembrane protein.

13. A nucleic acid constuct according to any of claims 1 to 11 wherein the intracellular retention signal is distal to a transmembrane domain of the transmembrane protein.

14. 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.

15. A nucleic acid construct according to claim 14, wherein X encodes a 2A self-cleaving peptide from an aphtho- or a cardiovirus or a 2A-like peptide.

16. A nucleic acid construct according to claim 15 wherein the transmembrane protein is a Chimeric-antigen receptor (CAR).

17. A nucleic acid construct according to any preceding claim, wherein each of the first and second POIs is a transmembrane protein; wherein at least one of the transmembrane proteins comprises an intracellular retention signal as defined in any of claims 3 to 11.

18.   A nucleic acid construct according any preceding claim wherein one or both of the first and second POIs is a single pass transmembrane protein.

19.   A nucleic acid construct according to claim 18 wherein one or both of the first and second POIs is a type I transmembrane protein.

20.   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 comprise an intracellular retention signal.

21.   A nucleic acid construct according to claim 20, 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.

22.   A vector comprising a nucleic acid construct according to any preceding claim.

23.   A retroviral vector or a lentiviral vector or a transposon according to claim 22.

24.   A cell comprising a nucleic acid construct according to any one of claims 1 to 21 or a vector according to claim 22 or 23.

25.   A method for modulating the relative cell surface expression of a transmembrane protein expressed from a single nucleic acid consruct with a second protein by including an intracellular retention signal in the nucleic acid sequence which encodes the transmembrane protein.

26.   A method according to claim 25 wherein the nucleic acid sequence is as defined in any of claims 1 to 21.
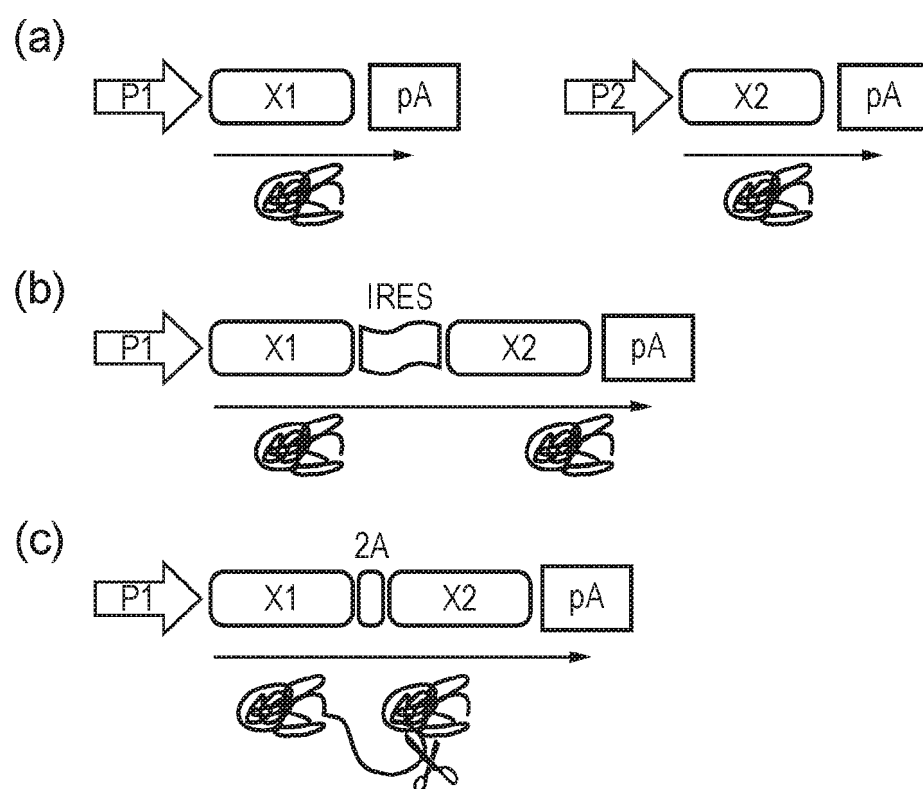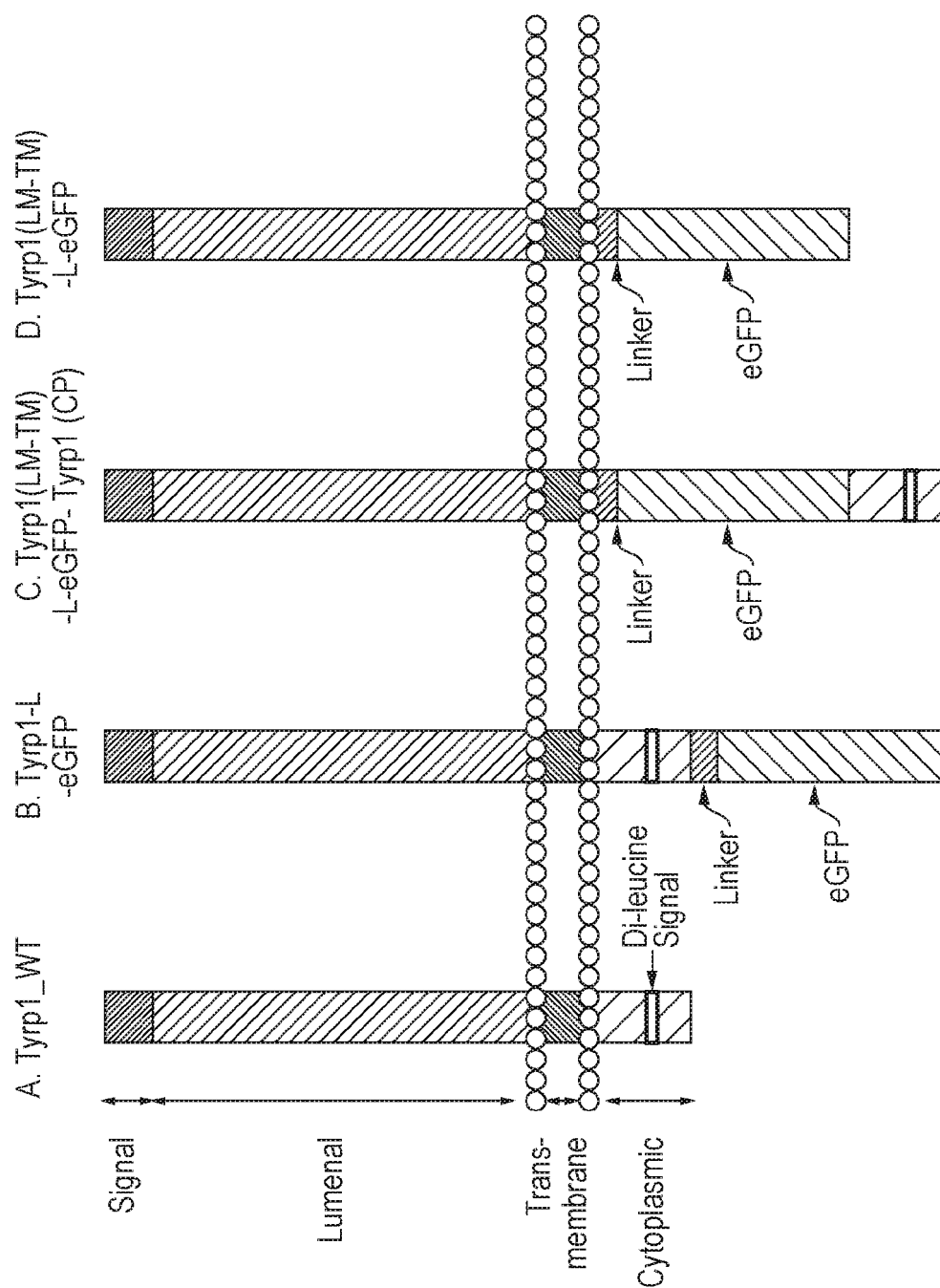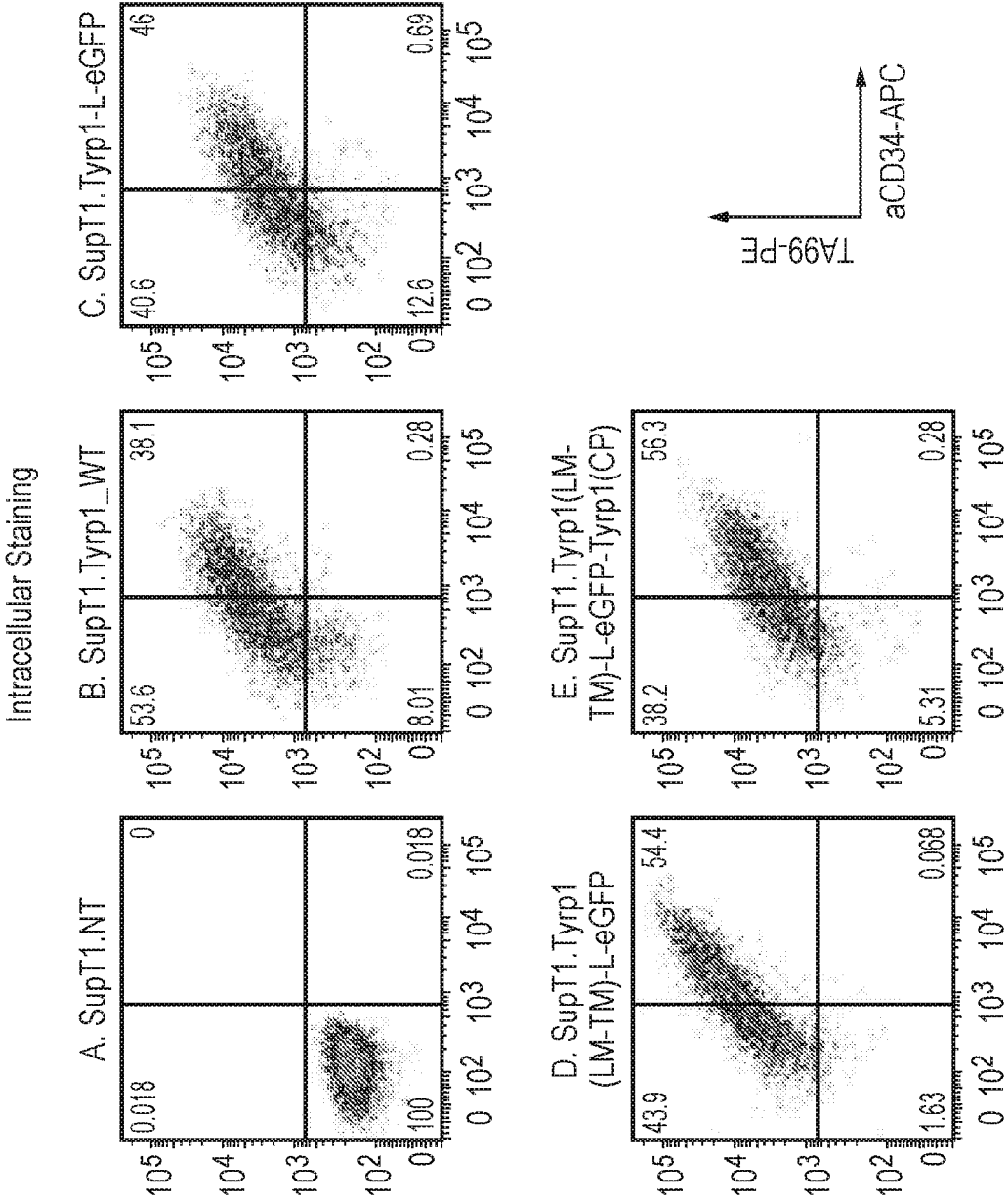
1/9

FIG. 1

FIG. 2

3/9

Intracellular Staining

A. SupT1.NT

B. SupT1.Tyrp1_WT

C. SupT1.Tyrp1-L-eGFP

D. SupT1.Tyrp1 (LM-TM)-L-eGFP

E. SupT1.Tyrp1(LM-TM)-L-eGFP-Tyrp1(CP)

aCD34-APC

TA99-PE

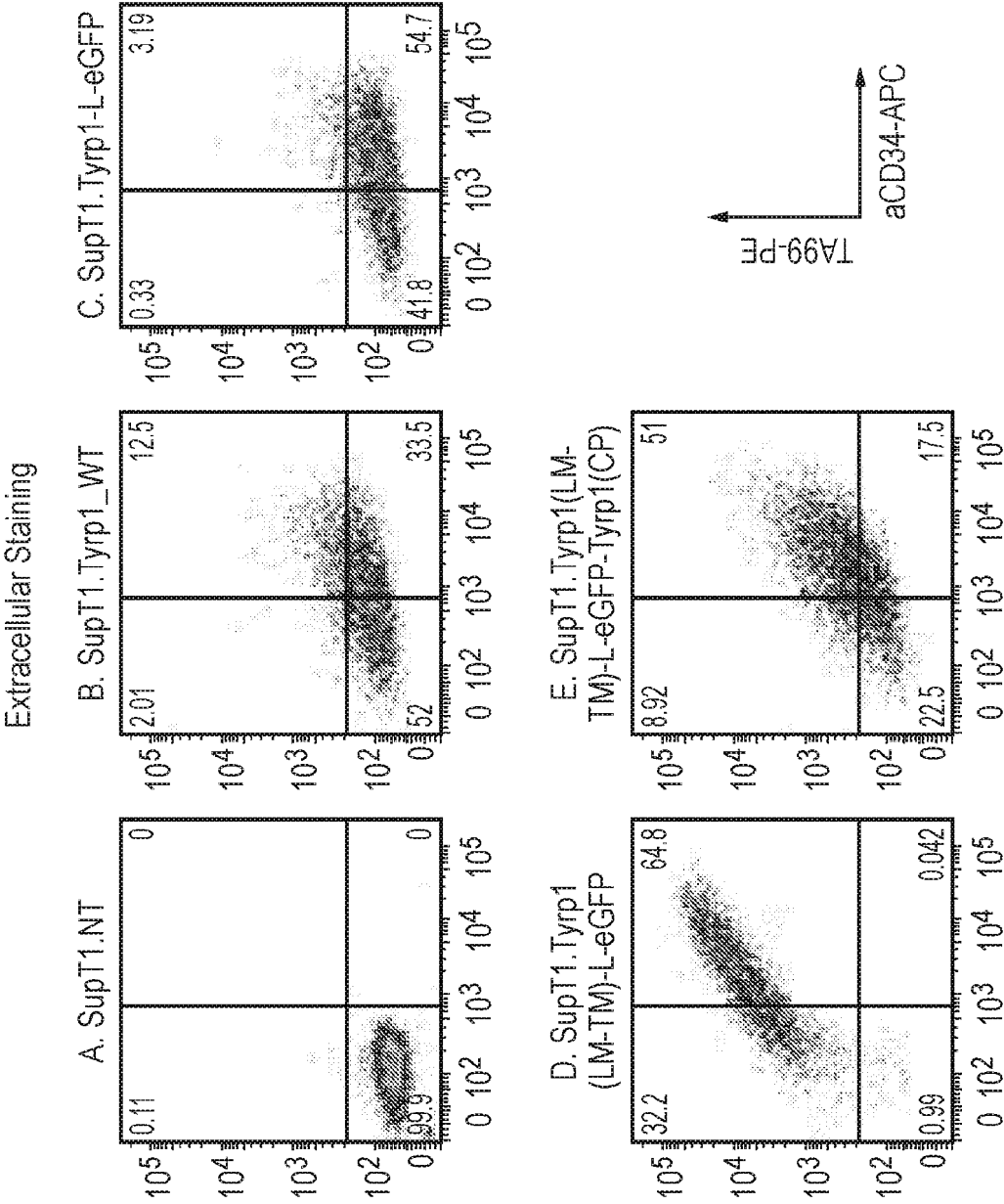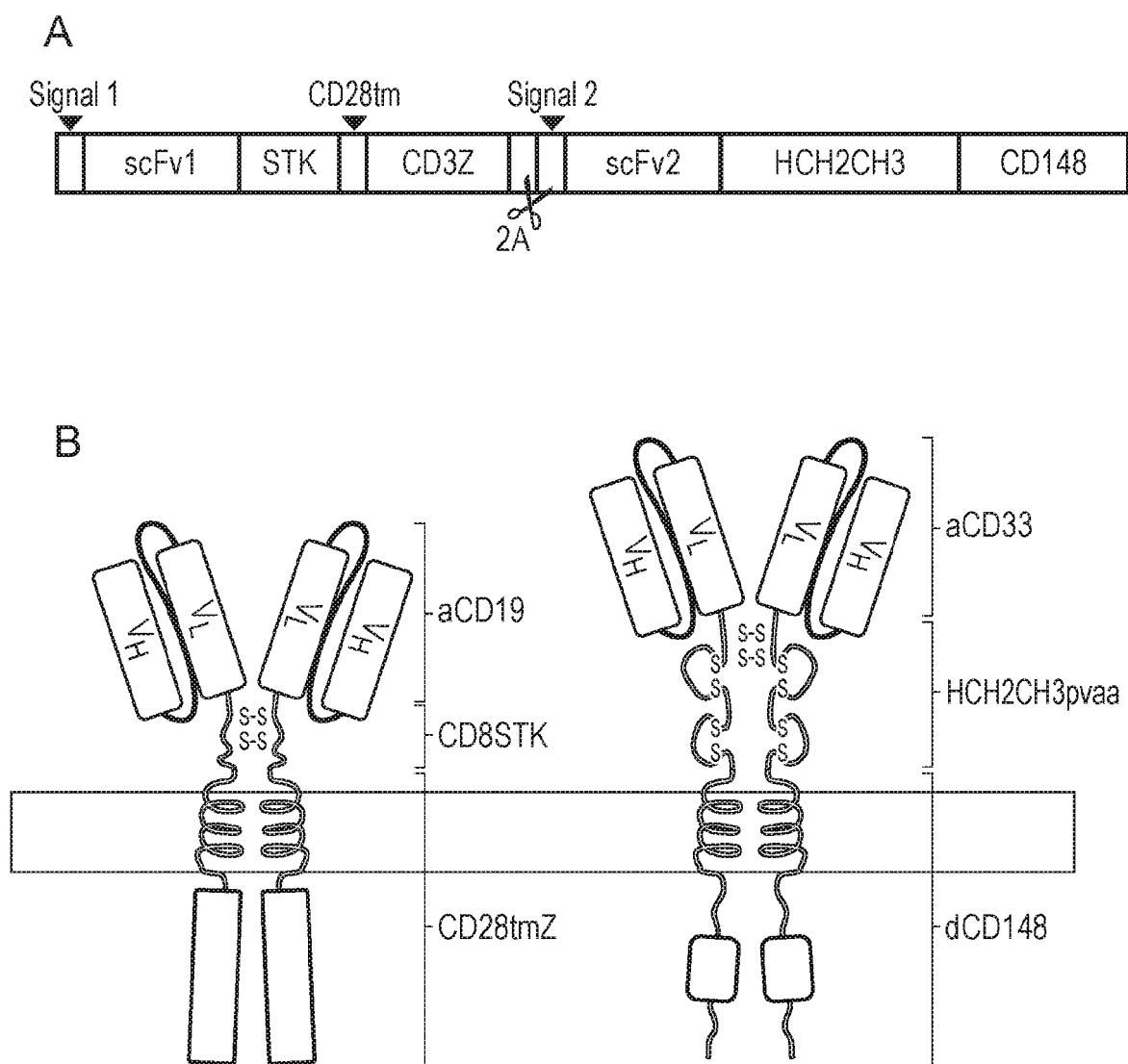FIG. 2 (Continued)

4/9

Extracellular Staining

FIG. 2 (Continued)

5/9

A

B
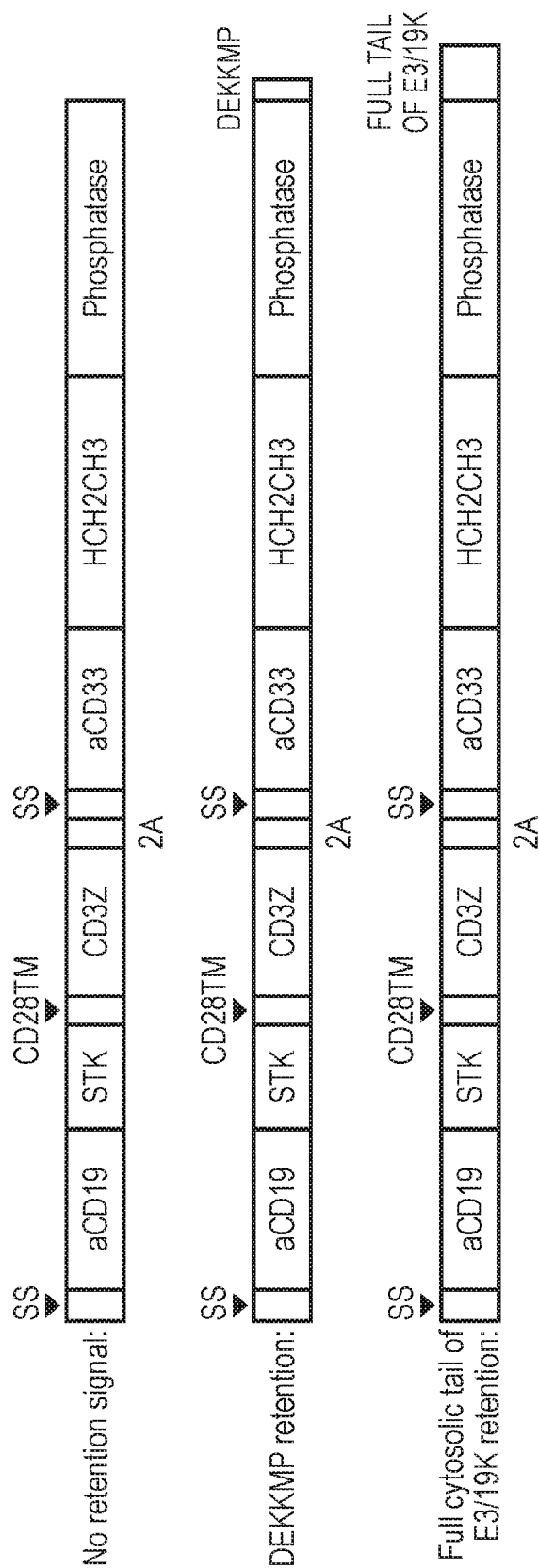

FIG. 3

6/9
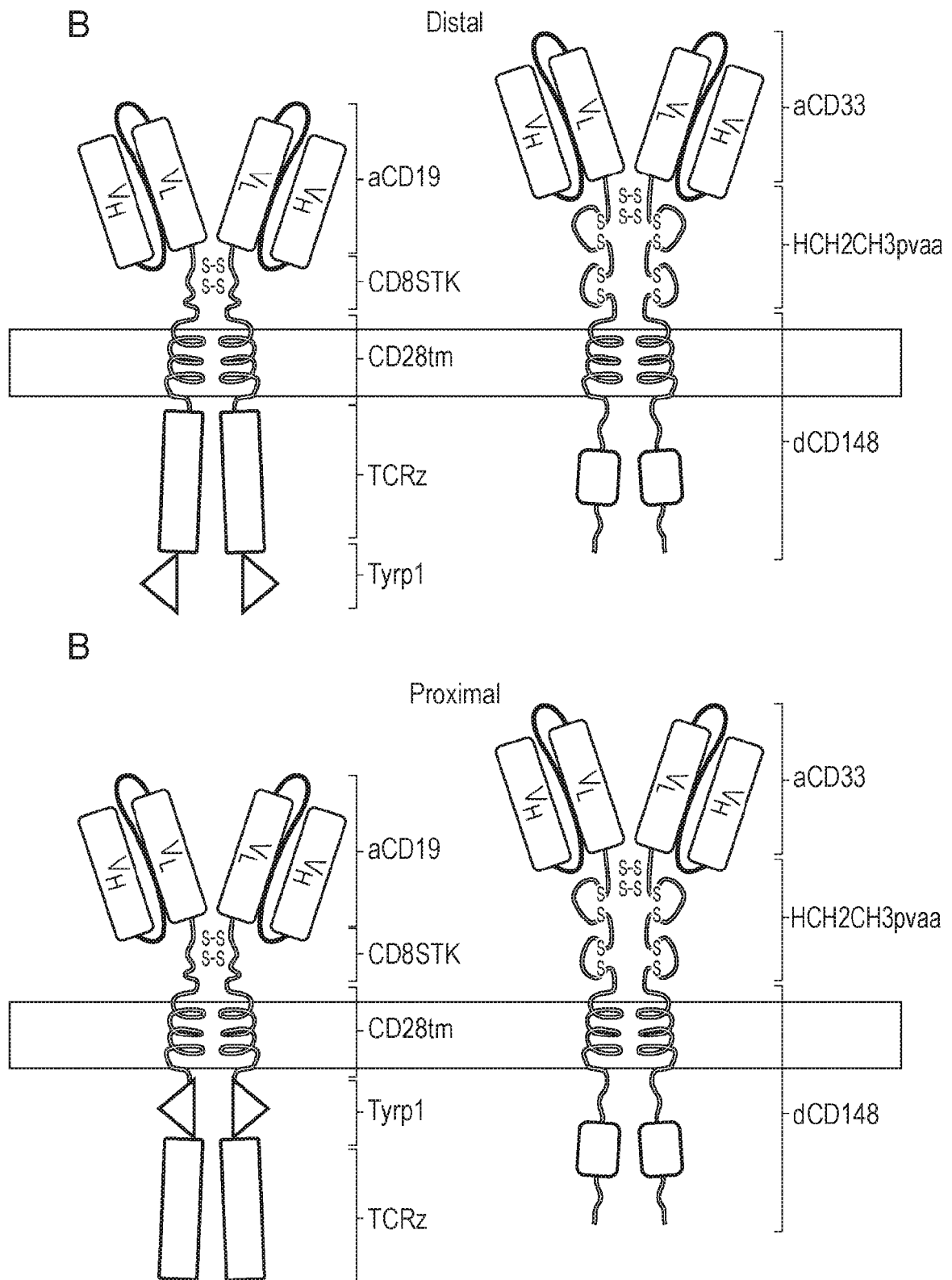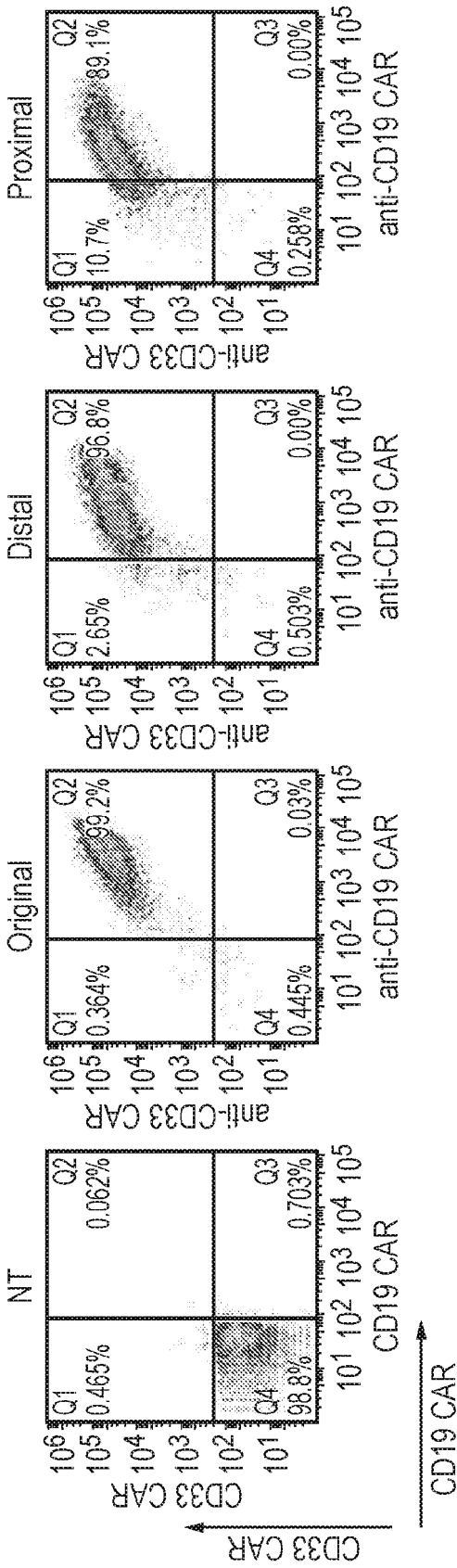
FIG. 3 (Continued)

FIG. 3 (Continued)

8/9

FIG. 4

9/9

FIG. 5
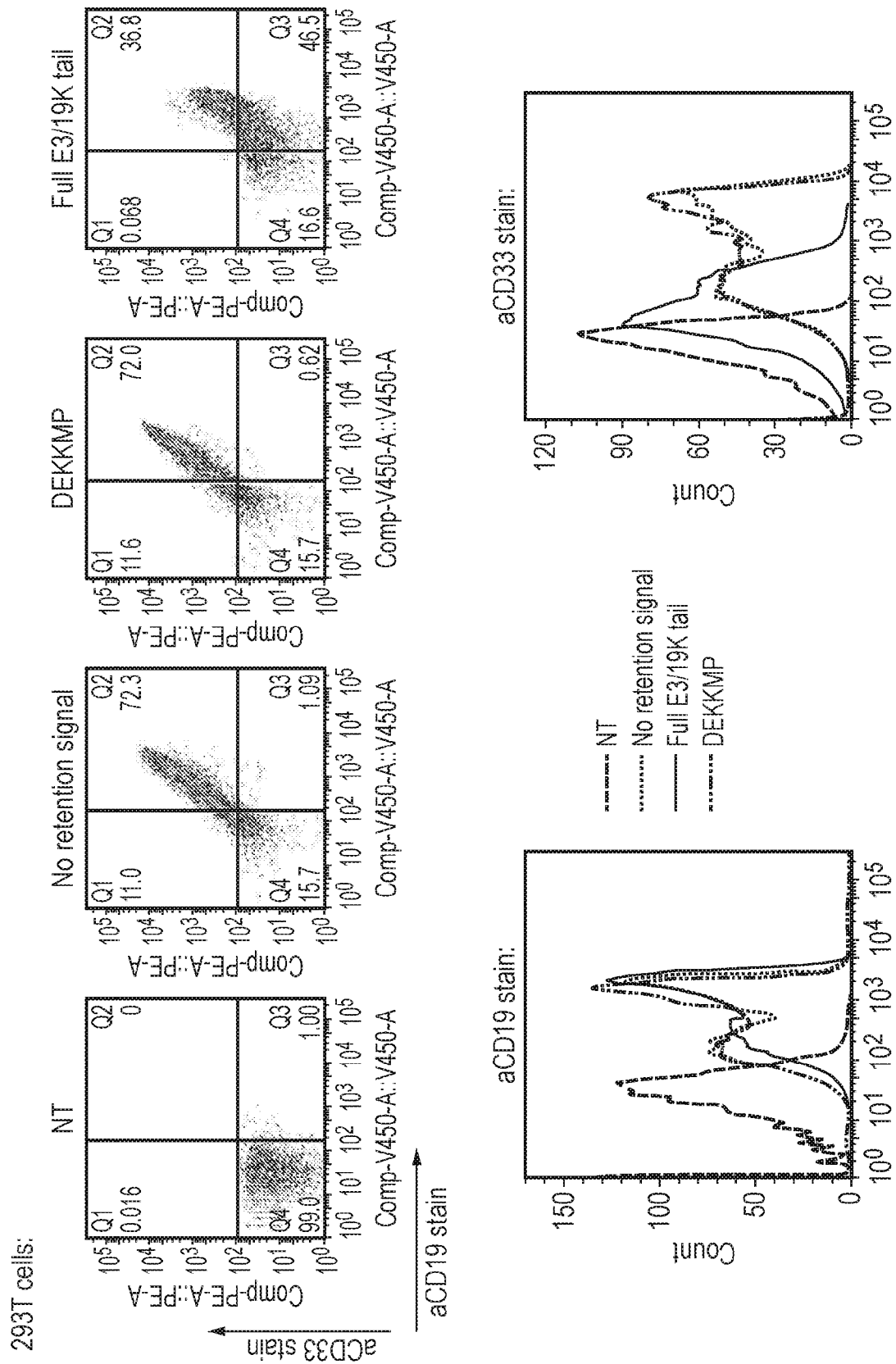
INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER

INV. C07K14/705   C07K16/28   C12N15/85
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, BIOSIS, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	A. EL AMRANI: "Coordinate Expression and Independent Subcellular Targeting of Multiple Proteins from a Single Transgene", PLANT PHYSIOLOGY., vol. 135, no. 1, 1 May 2004 (2004-05-01), pages 16-24, XP055284563, US ISSN: 0032-0889, DOI: 10.1104/pp.103.032649 figure 4 ----- -/--	1-26

[X] Further documents are listed in the continuation of Box C.   [X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 June 2016	11/07/2016

| Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016 | Authorized officer Lonnoy, Olivier |

1

C(Continuation).	DOCUMENTS CONSIDERED TO BE RELEVANT	
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DE FELIPE PABLO: "Skipping the co-expression problem: the new 2A CHYSEL technology", GENETIC VACCINES AND THERAPY, BIOMED CENTRAL, LONDON, GB, vol. 2, no. 1, 13 September 2004 (2004-09-13), page 13, XP021009743, ISSN: 1479-0556, DOI: 10.1186/1479-0556-2-13 figure 1 -----	1-26
X	FELIPE DE P ET AL: "Targeting of proteins derived from self-processing polyproteins containing multiple signal sequences", TRAFFIC, MUNKSGAARD, DK, vol. 5, no. 8, 1 August 2004 (2004-08-01), pages 616-626, XP002375856, ISSN: 1398-9219, DOI: 10.1111/J.1398-9219.2004.00205.X page 621, column 2, paragraph 2; figure 1 -----	1-26
X	PABLO DE FELIPE ET AL: "Inhibition of 2A-mediated 'cleavage' of certain artificial polyproteins bearing N-terminal signal sequences", BIOTECHNOLOGY JOURNAL, vol. 5, no. 2, 27 November 2009 (2009-11-27), pages 213-223, XP055284649, DE ISSN: 1860-6768, DOI: 10.1002/biot.200900134 page 221, column 1, paragraph 2; figure 2 -----	1-26
X	GARRY A. LUKE ET AL: "Self-Processing Polyproteins: A Strategy for Co-expression of Multiple Proteins in Plants", BIOTECHNOLOGY AND GENETIC ENGINEERING REVIEWS, vol. 23, no. 1, 1 December 2006 (2006-12-01), pages 239-252, XP055284655, GB ISSN: 0264-8725, DOI: 10.1080/02648725.2006.10648086 page 246, paragraph 2 - page 247, paragraph 2 -----	1-26
X	WO 2014/079878 A1 (MERZ PHARMA GMBH & CO KGAA [DE]) 30 May 2014 (2014-05-30) claims 1-3 ----- -/--	1-26

1

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JUAN S. BONIFACINO ET AL: "Signals for Sorting of Transmembrane Proteins to Endosomes and Lysosomes*", ANNUAL REVIEW OF BIOCHEMISTRY, vol. 72, no. 1, 1 June 2003 (2003-06-01), pages 395-447, XP055284679, US ISSN: 0066-4154, DOI: 10.1146/annurev.biochem.72.121801.161800 cited in the application abstract -----	1-26

1

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 2014079878 A1	30-05-2014	AU	2013349725 A1	30-04-2015
		CA	2885294 A1	30-05-2014
		CN	104837862 A	12-08-2015
		EP	2922866 A1	30-09-2015
		HK	1210190 A1	15-04-2016
		JP	2016502405 A	28-01-2016
		KR	20150087321 A	29-07-2015
		SG	11201502375Q A	28-05-2015
		WO	2014079878 A1	30-05-2014

摘要

本發明提供了包含以下結構的核酸構建體：A－X－B，其中A和B是編碼第一和第二目的多肽（POI）的核酸序列;並且X是編碼切割位點的核酸序列，其中第一或第二POI是包含細胞內保留信號的跨膜蛋白。