



(12) **DEMANDE DE BREVET CANADIEN**
CANADIAN PATENT APPLICATION
(13) **A1**

(86) Date de dépôt PCT/PCT Filing Date: 2018/05/02 (87) Date publication PCT/PCT Publication Date: 2018/11/08 (85) Entrée phase nationale/National Entry: 2019/10/22 (86) N° demande PCT/PCT Application No.: GB 2018/051166 (87) N° publication PCT/PCT Publication No.: 2018/203049 (30) Priorité/Priority: 2017/05/02 (GB1706945.1)	(51) Cl.Int./Int.Cl. <i>C12N 15/86</i> (2006.01), <i>A61K 35/76</i> (2015.01) (71) Demandeur/Applicant: UNITED KINGDOM RESEARCH AND INNOVATION, GB (72) Inventeurs/Inventors: TRIPODI, MARCO, GB; CIABATTI, ERNESTO, GB (74) Agent: BERESKIN & PARR LLP/S.E.N.C.R.L.,S.R.L.
---	--

(54) Titre : VECTEUR VIRAL A AUTO-INACTIVATION
(54) Title: SELF-INACTIVATING VIRAL VECTOR

(57) **Abrégé/Abstract:**

The invention relates to vectors based on a virus from the order Mononegavirales, and in particular a rabies virus. More specifically, it relates to a rabies virus vector which, having transfected a target cell, is switchable between replication-competent and replication-incompetent forms. Amongst other applications, the invention avoids the cytotoxicity associated with current vectors based on rabies virus.

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
08 November 2018 (08.11.2018)



(10) International Publication Number
WO 2018/203049 A1

- (51) International Patent Classification:
C12N 15/86 (2006.01) A61K 35/76 (2015.01)
Published:
— with international search report (Art. 21(3))
- (21) International Application Number:
PCT/GB2018/051166
- (22) International Filing Date:
02 May 2018 (02.05.2018)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
1706945.1 02 May 2017 (02.05.2017) GB
- (71) Applicant: UNITED KINGDOM RESEARCH AND IN-
NOVATION [GB/GB]; Polaris House, North Star Avenue,
Swindon SN2 1FL (GB).
- (72) Inventors: TRIPODI, Marco; MRC Laboratory of Mole-
cular Biology, Francis Crick Avenue, Cambridge Biomed-
ical Campus, Cambridge Cambridgeshire CB2 0QH (GB).
CIABATTI, Ernesto; MRC Laboratory of Molecular Bi-
ology, Francis Crick Avenue, Cambridge Biomedical Cam-
pus, Cambridge Cambridgeshire CB2 0QH (GB).
- (74) Agent: FORREST, Graham et al.; Mewburn Ellis LLP,
City Tower, 40 Basinghall Street, London Greater London
EC2V 5DE (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, DJ, DK, DM, DO,
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,
HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP,
KR, KW, 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).
- Declarations under Rule 4.17:
— of inventorship (Rule 4.17(iv))

(54) Title: SELF-INACTIVATING VIRAL VECTOR

(57) Abstract: The invention relates to vectors based on a virus from the order Mononegavirales, and in particular a rabies virus. More specifically, it relates to a rabies virus vector which, having transfected a target cell, is switchable between replication-competent and replication-incompetent forms. Amongst other applications, the invention avoids the cytotoxicity associated with current vectors based on rabies virus.

WO 2018/203049 A1

Self-inactivating viral vector

This application claims priority from GB 1706945.1, filed 2
May 2017, the contents and elements of which are herein
5 incorporated by reference for all purposes.

Field of the invention

The invention relates to a vector based on a virus from the
order Mononegavirales, such as a rhabdovirus, and in
10 particular a rabies virus vector. More specifically, it
relates to a vector which, having transfected a target cell,
is switchable between replication-competent and replication-
incompetent forms. Amongst other applications, the invention
avoids the cytotoxicity associated with current vectors based
15 on Mononegavirales such as rabies virus vectors.

Background to the invention

Vectors based on viruses of the order Mononegavirales have
considerable potential in various therapeutic, diagnostic and
20 research contexts. Although these vectors can be engineered
such that they do not propagate from a transfected cell, or
are able to propagate only in a tightly controlled manner,
they still show significant cytotoxicity to transfected cells
which limits their application.

25 For example, neurotropic viruses, particularly G-deleted
rabies (Δ G-Rabies) (12, 13), by spreading from neuron to
neuron along circuit paths, provide a potential tool to gain
genetic access to topologically defined neurons (13-16).

30 However, despite the transformative role of Δ G-Rabies based
approaches in the anatomical investigation of neuronal
circuits, viral induced cytotoxicity prevents their use to
follow network dynamics or manipulate functional properties of
neural networks *in vivo* for periods longer than 5-15 days
35 following the rabies infection (17-19). One possible solution
to eliminate viral cytotoxicity would be to silence viral

transcription after the primary infection. With DNA based viruses, such as Adeno-Associated Viruses (AAVs), this is traditionally achieved by inverting the viral genomic cassette in a CRE-recombinase dependent manner; flipping the genomic cassette effectively toggles the virus ON or OFF (20). Such an approach is currently precluded for RNA-based viruses, such as the rabies virus, due to the absence of reliable RNA recombinases.

Thus there remains a need for improved mononegaviral vectors which address the issue of cytotoxicity.

Summary of the invention

Given the coupled nature of transcription and replication in viruses of the order Mononegavirales, the inventors hypothesised that conditional modulation of viral protein stability might act as a switch for the viral transcription-replication cycle within infected cells, turning the virus on or off when specific conditions are met.

Thus the inventors have designed a viral vector which is capable of transfecting (and hence delivering a genetic payload to) a target cell, and which is switchable between an active state in which viral protein expression is possible and an inactive state in which viral protein expression is inhibited. Continued maintenance of the inactive state will typically result in the vector being eliminated from the host cell.

In particular, the vector encodes a "replication modulator protein" which comprises a mononegaviral protein required for replication of the vector genome, and which is capable of adopting two configurations.

The mononegaviral protein which forms part of the replication modulator protein is referred to here as the "viral protein moiety".

In one configuration, referred to as the "targeted" configuration, the replication modulator protein is targeted for degradation in the target cell. Thus the "targeted" configuration is unstable. While the replication modulator protein maintains this configuration, expression from the vector genome and/or replication of the vector genome is inhibited or completely suppressed.

The second configuration of the modulator protein, referred to as the "untargeted" configuration, is more stable than the "targeted" configuration. As a result, it supports higher levels of expression from the vector and/or replication of the vector genome, typically because the viral protein moiety accumulates in the host cell cytosol to higher levels than the targeted configuration.

References to "stability" in this context relate only to the half life of the relevant proteins in the target cell and their propensity to be degraded by the target cell's protein degradation pathways. The targeted configuration has a shorter half life than the untargeted configuration and is thus regarded as being less stable than the untargeted configuration.

Thus the replication modulator protein provides a switch which enables strictly regulated expression from the vector genome and/or replication of the vector genome.

Typically, the "targeted" configuration of the modulator protein displays a degradation targeting signal or "degron", i.e. a moiety that marks the protein for degradation within the cell. The terms "degron" and "degradation targeting signal" are used interchangeably in this specification.

The "untargeted" configuration of the modulator protein differs from the targeted configuration in that it either does

not comprise or does not display the degron responsible for the lower stability of the targeted configuration. As a result, the untargeted configuration typically has a longer half-life within the cell than the targeted configuration.

5

Degrans which may be employed in the context of the invention are discussed in more detail below.

10

The default state of the replication modulator protein may be the targeted configuration. That is to say, the targeted configuration may be encoded by the vector genome, such that the replication modulator protein is synthesised in the targeted configuration. Adoption of the untargeted configuration is stimulated by contact with an agent which may be referred to as an "activating agent" or "stabilising agent", since it supports production of sufficient functional protein to activate vector expression and/or replication. As result, may it permit the on-going persistence or survival of the vector within the cell.

15

20

A replication modulator protein having the targeted configuration as its default state may be referred to as an "inhibitory" replication modulator protein, since expression from a vector encoding such a modulator protein (or replication of such a vector) will tend to be inhibited except in the presence of the cognate activating agent. Thus an "inhibitory" modulator protein is constitutively unstable.

25

30

When a vector encoding a replication modulator protein having a default targeted configuration first infects a target cell, the initial level of protein synthesis may be sufficient to support a period of replication and potentially production of infectious progeny virions (depending on the payload of the vector and any other proteins provided in trans by the target cell). Thereafter, due to the instability of the replication modulator protein, replication is inhibited or completely suppressed in the absence of the activating agent.

35

Alternatively, the default state of the replication modulator protein may be the untargeted configuration. That is to say, the untargeted configuration of the replication modulator may be encoded by the vector genome, such that the replication modulator protein is synthesised in the untargeted configuration. Adoption of the targeted configuration is stimulated by contact with an agent which may be referred to as a "inhibitory agent" or "destabilising agent", since its presence will tend to inhibit or suppress expression from the vector and/or replication of the vector genome.

A replication modulator protein having the untargeted configuration as its default state may be referred to as an "inhibitable" replication modulator protein since expression from a vector comprising such a modulator protein will tend to proceed except in the presence of the cognate inhibitory agent.

In many embodiments, the replication modulator protein comprises a viral protein moiety and a regulator moiety. The viral protein moiety and the regulator moiety may form distinctly folded domains of the modulator protein, although this may not always be necessary. The viral protein moiety is typically capable of exerting all the biological functions of the native form of the relevant viral protein.

The nature and identity of the regulator moiety (when present) determines whether the modulator protein is an inhibitory modulator protein (i.e. vector expression or replication is inhibited in the absence of the cognate activating agent) or an inhibitable modulator protein (i.e. vector expression or replication proceeds as normal in the absence of the cognate inhibitory agent).

For an inhibitory modulator protein, the regulator moiety comprises or consists of the degron, and in the default state

of the modulator protein, the degron is displayed such that the protein is targeted for degradation. The cognate activating agent will therefore act to remove, mask or otherwise disable the degron, e.g. by covalent or non-covalent modification. For example, the cognate activating agent may cleave the regulator moiety from the viral protein moiety. Thus the activating agent may be a protease which is capable of acting on the modulator protein to cleave the regulator moiety from the viral protein moiety. This mechanism may be described as "cleavage-induced stabilisation" of the replication modulator protein (or of the viral protein moiety). Alternatively, the activating agent may bind to the modulator protein, e.g. to the regulator moiety, to mask or otherwise inactivate the degron. Thus the activating agent may be a ligand for the modulator protein, e.g. for the regulator moiety of the modulator protein. This mechanism may be described as "ligand-induced stabilisation" of the replication modulator protein (or of the viral protein moiety).

For an inhibitable modulator protein, the regulator moiety does not comprise or does not display the degron in its default state. Rather, the cognate inhibitory agent will interact with the modulator protein to create or reveal the degron, typically by covalent or non-covalent interaction with the regulator moiety. Thus the cognate inhibitory agent may cleave the regulator moiety from the viral protein moiety in order to reveal or create the degron in the viral protein moiety. This mechanism may be described as "cleavage-induced destabilisation" (or "cleavage-induced degradation") of the replication modulator protein. Alternatively the inhibitory agent may bind to the replication modulator protein, e.g. to the regulator moiety, so as to create or reveal the degron. Thus the inhibitory agent may be a ligand for the modulator protein, e.g. for the regulator moiety of the modulator protein. This mechanism may be described as "ligand-induced destabilisation" (or "ligand-induced degradation") of the

replication modulator protein. Thus, the regulator moiety and the inhibitory agent may respectively represent first and second components of an inducible degron system, wherein association of the first and second components generates a
5 degron.

Thus, the invention provides a mononegaviral vector genome comprising a gene encoding a replication modulator protein, wherein the replication modulator protein comprises a
10 mononegaviral protein moiety which is required for replication of the viral genome, the replication modulator protein being capable of adopting a targeted configuration displaying a degron, and an untargeted configuration which does not display the degron.

15 The vector genome is typically a negative-sense, single stranded RNA molecule.

As described above, the replication modulator protein encoded
20 by the vector genome may be an inhibitory modulator. It may comprise a viral protein moiety and a regulator moiety, wherein the regulator moiety comprises or consists of the degron.

25 The replication modulator protein may be switchable to an untargeted configuration on contact with a cognate activating agent. In some embodiments, the activating agent is capable of cleaving the regulator moiety from the modulator protein. In other embodiments, the activating agent is a ligand for the
30 modulator protein, e.g. for the regulator moiety. The ligand may mask or otherwise inactivate the degron, e.g. sterically or via a conformational change.

Alternatively, the replication modulator protein encoded by
35 the vector genome may be an inhibitable modulator which is switchable to a targeted configuration displaying a degron on

contact with a cognate inhibitory agent. It may comprise a viral protein moiety and a regulator moiety.

5 In some embodiments, the inhibitory agent is capable of cleaving the regulator moiety from the modulator protein to create or reveal the degron. In other embodiments, the inhibitory agent is a ligand for the regulator moiety and creates or reveals the degron on binding to the regulator moiety.

10 Thus the activating agent or inhibitory agent, as appropriate, may act by cleaving the regulator moiety from the viral protein moiety in order to remove or reveal the degron respectively. In such embodiments, the activating or
15 inhibitory agent is typically a protease. In such embodiments, the regulator moiety comprises a cleavage site for the protease.

20 The protease may be orthogonal to the target cell. That is to say, the protease recognises a cleavage site which is not found in the proteome of the target cell, i.e. the cleavage site is not found in native proteins encoded by and expressed in the target cell. Thus the particular protease may vary depending on the target cell. Examples of suitable proteases
25 include viral proteases, especially proteases from viruses which are not mononegaviruses (e.g. Tobacco Etch Virus protease (TEVp) and human rhinovirus (HRV) 3C protease), Factor Xa, enterokinase and thrombin. Any of these may be suitable when the target cell is a neural cell.

30 Preferably, the protease selected does not act on any other proteins encoded by the vector genome, e.g. amongst the proteins encoded by the vector genome, only the replication modulator comprises a cleavage site recognised by the relevant
35 protease.

The activating agent or inhibitory agent may itself be encoded by the vector genome. In such cases, expression or function of the agent will be inducible, typically by contacting a cell containing the vector genome with an appropriate inducer. For
5 example, the vector may comprise one or more genes encoding an agent which is expressed in functionally inactive form and wherein function is induced on contact with an inducer.

Thus, the agent may be expressed as two or more separate
10 protein moieties which require the presence of a functional inducer in order to associate into a functional form. Each of the component agent moieties may be expressed as a fusion with a partner protein, wherein the partner proteins associate on contact with the functional inducer.

15 More detail regarding inducible agents is provided below.

Any mononegavirus may be used as the basis for a vector as described. Those having non-segmented genomes are
20 particularly appropriate. For many applications, it may be desirable that the virus is capable of infecting and replicating in mammalian cells. For example, the mononegavirus may be a rhabdovirus (e.g. a lyssavirus, such as a rabies virus) or a vesiculovirus (such as a vesicular
25 stomatitis virus (VSV)).

The viral protein moiety of the replication modulator may comprise or consist of any one of the proteins common to the majority of mononegaviruses. These include the large protein
30 (L), nucleoprotein (N or NP), phosphoprotein (P), matrix protein (M), or glycoprotein (G; also referred to as simply the "envelope" protein). This terminology is commonly used for rhabdoviruses, and is used in the present specification also to refer to their equivalents by function or sequence
35 homology in other mononegaviruses. However, it is necessary to identify an appropriate viral protein moiety which provides adequate control over viral replication.

The viral protein moiety of the replication modulator protein is a protein which is required for expression from the viral genome (e.g. for transcription of mRNA from the viral genome) and/or for replication of the viral genome. Thus, the viral protein moiety is typically not the viral envelope protein (e.g. the glycoprotein or G protein) as, for most mononegaviruses, replication of the other viral proteins and the genome proceeds in the absence of the envelope protein.

Thus the large protein (L), nucleoprotein (N or NP), phosphoprotein (P) or matrix protein (M) may be preferred, e.g. the large protein (L), nucleoprotein (N or NP) or phosphoprotein (P).

The nucleoprotein (N or NP protein) may be particularly preferred, especially in rhabdoviral vectors such as rabies viral vectors. Indeed, the present inventors have found that only the N protein provides adequate control over viral replication in vectors based on the rabies virus.

Whichever mononegaviral protein represents the viral protein moiety of the replication modulator protein, it will be understood that the replication modulator protein represents the only copy of that mononegaviral protein in the vector genome. To put it another way, the vector genome does not contain a gene encoding a version of the same protein which exists only in the untargeted configuration.

Thus the vector genome may comprise genes encoding one, two, three or all four of an N protein, a P protein, an M protein and an L protein, wherein one of these proteins is provided as the viral protein moiety of a replication modulator protein.

The vector genome may comprise a gene encoding a replication modulator protein comprising an N protein as the viral protein

moiety, and may optionally further comprise one, two, or three of a P protein, an M protein and an L protein.

The vector genome may comprise genes encoding:

- 5 (i) a replication modulator protein comprising an N protein as the viral protein moiety;
(ii) a P protein;
(iii) an M protein; and
(iv) an L protein.

10

The vector genome may also comprise a gene encoding an envelope protein. The envelope protein may be native to the mononegaviral vector. For example, when the vector is a rabies virus, it may be a rabies virus G protein.

15

Alternatively, a different envelope protein may be incorporated to modulate infectivity of virions produced by the vector. Incorporation of alternative envelope proteins may be referred to as "pseudotyping". In such cases, the envelope protein may comprise an intracellular domain and

20

optionally also a trans-membrane domain of an envelope protein native to the vector, and a heterologous extracellular domain, i.e. the extracellular domain is not from the same mononegaviral envelope protein as the intracellular domain.

25

The extracellular domain may be from a different viral envelope protein (e.g. from an envelope protein from a different mononegavirus species, family or genus, or from an envelope protein from a different viral order), or may be any protein domain capable of binding to a receptor expressed on the surface of a target cell, as long as the envelope protein remains capable of mediating infection of the target cell by the virion.

30

For example, a rabies virus vector may comprise an envelope protein having at least the extracellular domain from a VSV envelope protein, avian sarcoma leukosis virus (ASLV) type A envelope protein, or ASLV type B envelope protein. It may comprise an entire envelope protein from VSV, ASLV type A, or

35

ASLV type B. Alternatively, it may comprise an intracellular domain and optionally also a trans-membrane domain from a rabies virus G protein.

5 In some embodiments, the vector genome does not encode an envelope protein. Such virions are particularly useful for so-called "monosynaptic tracing" in neuronal cells.

10 In addition to any pseudotyped envelope protein, the vector genome may comprise one or more heterologous genes, i.e. a gene encoding an expression product (typically a protein or RNA molecule) which is not native to the vector genome, i.e. is not present encoded by a wild type virus of the same virus type as the vector. The heterologous gene or genes may be
15 regarded as the "payload" which is to be delivered to the target cell by the vector. Thus the heterologous gene(s) may encode any expression product (whether RNA or protein) which it is desired to express in the target cell. The identity and function of the heterologous gene(s) will thus depend on the
20 intended role of the vector. In many embodiments, the heterologous gene does not encode a mononegaviral expression product, or a viral expression product.

25 The heterologous gene may be located at any suitable site within the vector genome. In some embodiments, it is located between the genes encoding the M and L proteins, e.g. replacing the gene encoding the endogenous G protein.

30 The heterologous gene(s) may, for example, encode one or more of the following:

- a marker protein, e.g. a fluorescent protein or a protein conferring antibiotic resistance;
- a protein against which it is desirable to raise an immune response;
- 35 - a recombinase, e.g. a CRE recombinase, e.g. where the genome of the target cell contains recognition sites for the recombinase;

- a nuclease, e.g. an RNA-guided endonuclease such as Cas9;
- a guide RNA (gRNA) molecule, e.g. for use in conjunction with an RNA-guided endonuclease;
- a repair template RNA, e.g. for use in conjunction with an RNA-guided endonuclease;
- a nucleic acid modulator of gene expression, such as a siRNA, RNAi, anti-sense RNA or ribozyme.

The heterologous gene or genes will be under the control of viral transcription regulatory sequences such as transcription initiator and terminator signals.

Where two or more heterologous genes are included, they will typically each have their own associated regulatory sequences. Heterologous gene expression may be affected by factors such as the transcriptional start site employed by the heterologous gene, and its spacing from the end of the gene located immediately upstream in the viral genome, i.e. the length of the intergenic region (IGR) between the end of the gene immediately upstream and the start of the heterologous gene. The N protein has the highest level of transcription in a rabies virus, so it may be desirable that the heterologous gene employs the transcriptional start site from the N protein. The transcriptional start site may, for example, have the sequence AACACCCCT (e.g. as seen in strains B19 and N2C) or AACACCTCT. Finke et al. (2000) have demonstrated that the length of IGR upstream of a gene affects its expression, and that shorter IGR sequences correlate with increased expression. Thus it may be desirable that the IGR sequence upstream of the heterologous gene is 5 nucleotides or fewer in length, e.g. 2 nucleotides. Where the IGR is 2 nucleotides in length, its sequence may be CT.

It may be desirable to include two or more heterologous genes, especially if they have complementary functions. For example, it may be desirable to include heterologous genes encoding an RNA-guided endonuclease, a guide RNA for that endonuclease (to

direct the endonuclease to the desired site of the target cell chromosome), and optionally also a repair template RNA to direct the particular modification to be achieved at the target site. The heterologous genes may be adjacent to one another in the viral genome. The IGR sequence between the heterologous genes may, for example, be 5 nucleotides or fewer in length, e.g. 2 nucleotides. It may be desirable that each heterologous gene employs the N protein transcriptional start site from the N protein.

The endonuclease may be a Cas9 enzyme, such as Cas9 from *Streptococcus aureus* (saCas9), *Streptococcus pyogenes* (SpCas9), *Neisseria meningitidis* (NM Cas9), *Streptococcus thermophilus* (ST Cas9), *Treponema denticola* (TD Cas9), or variants thereof such as the D1135E, VRER, EQR or VQR variants of SpCas9. Due to its relatively smaller size, saCas9 and its variants may be preferred.

The term Cas9 is used here to include functional variants of Cas9 endonucleases such as Cas9 nickase and nuclease-dead Cas9 (dCas9).

The size of the heterologous genes which can be incorporated into the vector will depend on various factors, such as the specific virus on which the vector is based, the packaging capacity of the virion and which (if any) of the native viral genes have been deleted from the vector.

For example, in a rabies viral vector in which the G gene has been deleted, it is possible to include at least 3.7kb of heterologous coding sequence without substantial effect on packaging efficiency, and even up to 4.0 kb, up to 4.5 kb, or up to 5.0 kb. See ref. 32.

The heterologous gene may also be heterologous to the target cell, i.e. the target cell does not naturally encode or express the product of the gene. Alternatively, the target

cell may encode or express the gene product but under different regulatory control, such that the expression pattern of the gene product in the target cell is modulated by introduction of the heterologous gene.

5

The invention further provides a ribonucleoprotein complex comprising the vector genome in association with one or more viral proteins, e.g. one or more of N, P and L proteins. The vector genome may be in association with all of N, P and L proteins. The ribonucleoprotein complex may be a functional viral nucleocapsid, capable of initiating transcription on introduction to the cytoplasm of a target cell. It may additionally comprise an M protein.

10

15

The invention further provides a mononegaviral vector virion comprising a mononegaviral vector genome as described herein. The virion typically comprises an L protein, an N protein, a P protein and an M protein, and also comprises an envelope protein. The envelope protein may be a native mononegaviral envelope protein. Alternatively, the virion may have been pseudotyped with a different envelope protein as described above to modulate its infectivity.

20

25

The virion typically takes the form of a ribonucleoprotein complex (comprising genome, L, N, M and P proteins) surrounded by a membrane envelope containing the envelope protein.

The virion is capable of delivering the genome to the target cell and of initiating transcription of the genome.

30

For the avoidance of doubt, the virion typically does not contain a replication modulator protein; i.e. the proteins in the virion are not switchable between targeted and untargeted configurations.

35

The invention further provides a positive sense nucleic acid molecule encoding a viral vector genome as described herein.

The positive sense nucleic acid may be an RNA or DNA molecule. When it is a DNA molecule, it may be provided as part of an expression construct, comprising transcriptional regulatory sequences capable of directing synthesis of viral genome molecules. The expression construct may be provided as part of an episome (e.g. a plasmid) or may be integrated into the chromosome of a host cell.

The invention further provides a packaging cell comprising a nucleic acid construct encoding a vector genome and capable of producing virions as described herein.

Thus the packaging cell expresses L, N, M and P proteins and an envelope protein. The proteins are capable of forming infective virions also comprising the vector genome.

Depending on the structure and coding content of the vector genome itself, these proteins may be encoded by the vector genome or provided *in trans* by other nucleic acid constructs within the cell.

The proteins in the virion particle itself are typically not linked to degrons. Typically, the packaging cell will encode (and be capable of expressing) functional versions of all of the virion proteins, without degrons. However, if one of the virion proteins is only expressed in the context of an inhibitory replication modulator as described here, then the packaging cell must also contain (e.g. express) the cognate activating agent so that viral replication can proceed.

The constructs encoding any proteins to be provided in trans are typically DNA constructs, and may independently be chromosomally integrated or present as episomes such as plasmids.

The nature of the packaging cell will depend on the particular viral vector, but may be a mammalian cell such as a primate or

rodent cell. It may be a fibroblast or any other suitable cell type, such as a HEK(Human embryonic kidney)293 cell or a BHK (baby hamster kidney) cell.

5 Construction of packaging cells for rabies viral vectors, and other aspects of viral vector design, are described in Osakada and Callaway (2013) Design and generation of rabies virus vectors, Nature Protocols 8(8): 1583-1601. See also
10 Wickersham et al. (2010) Production of glycoprotein-deleted rabies viruses for monosynaptic tracing and high level gene expression in neurons, Nature Protocols 5(3): 595-606.

The invention further provides a method of gene delivery to a target cell, comprising contacting the target cell with a
15 ribonucleoprotein complex or a virion of the invention.

When a ribonucleoprotein complex is employed, it may be desirable to deliver it directly into the cell cytosol, e.g. by microinjection, or in conjunction with a carrier, such as a
20 polymer or lipid, to facilitate transit into the cell.

In many embodiments, the target cell does not naturally express the agent responsible for the switch between the targeted and untargeted configurations of the modulator
25 protein. When the vector encodes an inhibitory replicator protein, the activating agent is required for continued replication and propagation of the vector after initial primary infection. Thus, when further replication and propagation is desirable, the vector will often be employed
30 alongside a delivery system for the activating agent, or the activating agent itself.

Where the vector encodes an inhibitable modulator protein, it may not be necessary to provide the inhibitory agent.
35 However, it may be desirable that vector expression and replication should not proceed immediately, in which case the

vector may be employed alongside a delivery system for the inhibitory agent, or the inhibitory agent itself.

Thus the method may comprise the step of contacting the target cell with the cognate activating agent or inhibitory agent, as appropriate. The target cell may be contacted with the cognate activating agent periodically, e.g. at repeated intervals of 1 day to 1 month, e.g. at intervals of 1 day to 14 days, e.g. at intervals of 5 to 10 days, for an indefinite period of time. Such periodic administration may be useful to maintain viable vector within a target cell without permitting sufficient transcription and viral replication to compromise the cell.

Where the agent is a protein, the method may further comprise introducing into the target cell a nucleic acid comprising a gene encoding a cognate activating agent or inhibitory agent, such that the agent is expressed in the target cell.

The agent may be introduced into the target cell before the vector, at substantially the same time as the vector, or after the vector. When the agent is introduced after the vector, it will typically be introduced within a month, e.g. within 2 weeks, within 1 week, within 1 day, within 12 hours, or within 1 hour of the vector. It may be introduced periodically, e.g. at repeated intervals of 1 day to 1 month, e.g. at intervals of 1 day to 14 days, e.g. at intervals of 5 to 10 days, for an indefinite period of time.

Expression of the agent may be inducible. For example, the gene encoding the agent may comprise an inducible promoter so that transcription of the agent requires the cell to be contacted with a transcriptional inducer. The inducer may be an antibiotic for example, such as doxycycline or tetracycline. Inducible promoters which are responsive to such antibiotics are well known to the skilled person.

Alternatively, the agent may be expressed in a functionally inactive form which requires the cell to be contacted with a functional inducer in order to restore function. For example, the agent may be expressed as two or more separate protein moieties which require the presence of a functional inducer in order to associate into a functional form. Each of the component agent moieties may be expressed as a fusion with a partner protein, wherein the partner proteins associate on contact with the functional inducer. For example, the TEV protease (TEVp) can be expressed as two separate subunits each fused to heterologous protein moieties (FRB and FKBP) which associate only in the presence of rapamycin (a so-called "SPLIT-TEV" system) (See Gray et al. (2010) Activation of Specific Apoptotic Caspases with an Engineered Small-Molecule-Activated Protease, Cell 142(4): 637-646.). Other pairs of fusion partners may also be used, which have different requirements for association.

In some embodiments, the agent may be encoded by the vector genome itself, or in the genome of the target cell. Thus, either the vector genome or the target cell may comprise one or more genes encoding the activating agent or inhibitory agent, wherein expression or function of the agent is inducible. In such circumstances, the method does not require the target cell to be contacted with the activating agent or inhibitory agent, but simply requires the target cell to be contacted with the inducer (of expression or function).

The method may therefore include the step of inducing expression and/or function of the agent in the target cell, typically by contacting the target cell with the inducer.

The target cell may be contacted with the inducer periodically, e.g. at repeated intervals of 1 day to 1 month, e.g. at intervals of 1 day to 14 days, e.g. at intervals of 5 to 10 days, for an indefinite period of time.

In some embodiments, the vector genome does not comprise a gene encoding an envelope protein. Thus, the method may (additionally or alternatively) comprise the step of introducing into the target cell a nucleic acid construct comprising a gene encoding an envelope protein, such that the envelope protein is expressed in the target cell and incorporated into the plasma membrane. The envelope protein may be introduced into the target cell before the vector, at substantially the same time as the vector, or after the vector. When the envelope protein is introduced after the vector, it will typically be introduced within a month, e.g. within 14 days, within 7 days, within 1 day, within 12 hours, or within 1 hour of the vector.

Expression of the envelope protein may be inducible. For example, the gene encoding the envelope protein may comprise an inducible promoter so that transcription of the envelope protein requires the cell to be contacted with a transcriptional inducer.

The nucleic acid containing the gene encoding the inhibitory or activating agent, and/or the envelope protein, may be delivered to the target cell by any appropriate means. For example, each nucleic acid may independently be introduced as DNA or as RNA, as naked nucleic acid, in association with a carrier such as a polymer or lipid (e.g. as a complex with a carrier), in encapsulated form (e.g. encapsulated in a liposome) or via a viral vector. Suitable viral vectors include retroviral vectors (including lentiviral vectors), adenoviral vectors, and adeno-associated virus (AAV) vectors. AAV vectors may be particularly suitable.

The genes encoding the agent and the envelope protein may be delivered separately or together, e.g. as part of the same nucleic acid construct.

A viral vector having an inhibitory modulator protein may find particular use as an immunostimulatory agent, e.g. as a vaccine, against the native form of the same virus. Delivery of the vector to a recipient will result in replication of the virus only in cells which also receive the activating agent. Progeny virions may be formed and released from such infected cells. However, they will be unable to replicate further in cells which do not receive the activating agent. Thus co-administration of the vector and the activating agent, to particular selected cells, may result in a short controlled cycle of viral replication and release, stimulating the immune system and priming it to recognise the virus.

The viral vector may be otherwise attenuated to reduce any risk to the recipient. Thus it may comprise one or more further mutations to reduce or eliminate infectivity or virulence.

Release of progeny virions may be useful to simulate an antibody response against the native virus.

In order to produce and release progeny virions, the viral vector must encode an envelope protein, or an envelope protein must be supplied separately in trans to the infected cells if the vector lacks an envelope protein. Use of a vector lacking an envelope protein may represent a useful safety measure.

If the vector lacks an envelope protein, and no envelope protein is supplied in trans, there may still be sufficient protein expression in the presence of the activating agent to result in display of viral antigens via the recipient cell's MHC molecules (especially MHC I), and consequent stimulation of a T cell response (especially a CTL response) against the virus.

Nevertheless, use of a vector encoding an envelope protein may still be desirable in many circumstances.

The vectors of the invention find particular use in gene delivery to neural cells, especially when based on rabies virus or vesicular stomatitis virus. Use of G-deleted rabies virus is well known for monosynaptic circuit tracing in both the central and peripheral nervous system. However, current methods of monosynaptic circuit tracing have limited use because the labelled cells typically only remain viable for 1-2 weeks because of the accumulation of viral protein within the cell. The present vectors provide the capacity to genetically label neural cells without affecting viability.

The invention further provides a kit comprising a vector genome of the invention and (a) a cognate activating or inhibitory agent, or a nucleic acid encoding a cognate activating or inhibitory agent, and/or (b) a nucleic acid encoding an envelope protein.

The kit may comprise a system for delivery of the nucleic acid, such as a vector as described elsewhere. The same system may be used to deliver both the cognate survival factor and the envelope protein as described above.

The invention further provides a composition comprising a vector virion of the invention, optionally admixed with an excipient or carrier. The composition may be a pharmaceutical composition and the carrier may be a pharmaceutically acceptable carrier.

Mononegavirales infect diverse types of cell including plant, insect, fish and mammalian cells. In principle, it is believed that the present invention can be applied to any type of mononegavirus because they conform to a similar pattern of genomic organisation and rely on a similar coupled system of transcription and replication. Rhabdoviruses may be particularly apt.

The vectors of the invention may find particular use in the rhabdoviral genera and species which infect mammalian cells, especially lyssaviruses, such as the rabies virus (*Rabies lyssavirus*; RABV) and vesiculoviruses, such as vesicular stomatitis virus (VSV). Rabies virus has a particular tropism for neurons, which makes it a valuable candidate for gene delivery to neural tissue.

When the vector is based on a rabies virus, any appropriate strain may be used, including CVS (Challenge virus standard) and variants thereof such as CVS-N2c, PV4 (Pasteur virus), PM (Pitman-Moore), Flury-LEP (low egg passage), Flury-HEP (high egg passage), ERA and SAD (Street-Alabama Dufferin). The CVS-N2c strain is described in reference 25 and a full genome sequence is available under GenBank accession no. HM535790, version HM535790.1, 29 December 2010.

Thus, the target cell type may be from a primate (e.g. Old World monkey, New World monkey, ape or human), rodent (e.g. mouse or rat), canine (e.g. domestic dog), feline (e.g. domestic cat), equine (e.g. horse), bovine (e.g. cow), caprine (e.g. goat), ovine (e.g. sheep) or lagomorph (e.g. rabbit).

The target cell type may be any desired cell type. In some examples, it may be a neural cell, e.g. a neuron or glial cell. The neural cell may be part of the peripheral or central nervous system.

Description of the Drawings

Fig. 1.

Screening viral amplification efficiency after systematic proteasome targeting of viral proteins. **(A)** Reversible viral protein destabilization via proteasome targeting PEST domain. **(B)** TEV_P-dependent viral amplification in HEKGG and HEKTGG. Scale bar: 100 μ m. **(C-H)** Quantification of amplification efficiency for all recombinant Rabies constructs (magenta) and Control Rabies (cyan) (mean $\pm$ SD; dashed line shows threshold

level). (**I-K**) Quantification of amplification efficiency in HEKTGG (cyan, +TEV) and HEGG (magenta, -TEV). x-axis, days post-transfection (p.t.), y-axis amplification efficiency.

5 **Fig. 2.**

Absence of cytotoxicity *in vivo*. (**A**) SiR life cycle. (**B**) SiR expression cassette and experimental procedure. (**C-E''**) Confocal images of hippocampal sections of *Rosa-LoxP-STOP-LoxP-YFP* mice infected with SiR^{CRE-mCherry} and imaged at 1-3-8 weeks p.i.. Scale bar: 25 μ m. (**F**) Number of YFP and mCherry positive neurons at 1-2-3 and 8 weeks normalized to 1 week time-point (mean $\pm$ SEM). (**G**) Levels of Viral RNA (magenta) and endogenous YFP expression (cyan) normalized to 1-week RNA level (mean $\pm$ SEM).

15 **Fig. 3.**

SiR transsynaptic and retrograde spread. (**A**) AAV-TVAmCherry-GLY was injected in CA1 of *Rosa-LoxSTOPLox-YFP* mice and the TVA expressing neurons were specifically targeted with an EnVA pseudotyped SiR^{CRE} 2 weeks later. (**A'**) In the site of injection YFP^{ON}/mCherry^{ON} starting neurons are detected (arrowheads) and (**A''-A'''**) the transsynaptic jump of SiR virus permanently labeled neurons in CA3 and lateral enthorinal cortex (LEC) with YFP expression. Scale bar: 25 μ m. (**B-B''**) Confocal images of CA1 pyramidal neurons infected with AAV-TVAmCherry-GLY and SiR^{CRE} at 3 weeks p.i.. Scale bar: 10 μ m. (**C-C''**) Retrograde tracing of neurons from CA3 and LEC projecting to CA1 by SiR^{mCherry} injection of CA1 of the hippocampus. (**D**) V1 cortical neurons infected by SiR injection in the superficial layers of the Superior Colliculus (SC). Scale bar: 25 μ m.

30 **Fig. 4.**

SiR infection has no long-term impact on neuronal physiology. (**A-B**) Membrane potential response to steps of positive and negative current of a CA1 pyramidal neuron 1 week post-infection (p.i.) and 8 weeks p.i.. (**C**) Input resistance, (**D**) resting membrane potential and action potential (AP)

threshold, (E) AP amplitude and width, (F) firing frequency at increasing steps of positive current for neurons 1 week p.i. ($n = 10$, magenta) and 8 weeks p. i. ($n = 8$, cyan) (mean $\pm$ SEM). (G) Membrane potential response to a 0.2 ms blue-light pulse of increasing intensity (by 1% in each sweep until spiking; 0% lighter grey, to 9%, black) of SiR^{CRE-mCherry} infected CA1 neuron expressing ChR2 (1 week p.i.). Insert, LED power delivered for each sweep. (H, I) Membrane potential response to a 800 ms light-pulse (2.17 mW) and to forty 1.5 ms long light-pulses at 20 Hz recorded from the same neuron. (J) Action potentials success rate following 40 light-stimulations at increasing frequencies, 1 (magenta) and 8 (cyan) weeks p.i.. (K) Light-evoked EPSPs recorded in non-ChR2-expressing neurons blocked by DNQX (20 μ M). Average traces for both conditions are shown in black. * $P < 0.05$.

Fig. 5.

Unaltered orientation tuning responses of SiR traced V1 neurons. (A) Schematic of SiR^{CRE} and AAV-GCaMP6s injection in *Rosa-LoxP-STOP-LoxP-tdTomato* mice in V2 and V1 respectively. (B-B'') Two-photon maximal projection of V1 neurons after SiR^{CRE} injection. In grey neurons expressing GCaMP6s (B), in magenta neurons expressing tdTomato (B') and in the merge neurons expressing both (B'', merge). Scale bar: 50 μ m. (C) Schematic of visual stimulation set up. (D) Outline of the active ROIs from the same field of view showed in panel B. (E) Representative Ca²⁺ traces of GCaMP6s (cyan) and GCaMP6s-tdTomato (magenta) neurons. Scale bars: 200 s, 20% dF/F₀. (F) Mean percentage of active neurons after 4 weeks from SiR injection ($n = 122$ GCaMP6 neurons (cyan), $n = 59$ GCaMP6s and GCaMP6s-tdTomato neurons (magenta)). (G) Changes in fluorescence over time reflecting visual responses to drifting gratings at the preferred direction of each neuron. (H) Example of tuning curve of V1 infected neurons. Scale bars: 5 s, 10% dF/F₀.

Fig. 6.

Testing cytotoxicity of Δ G-N^{PEST}Rabies^{SPLIT-TEV-mCherry} *in vitro* and *in vivo*. (A) hESCs derived neurons were infected with Δ G-N^{PEST}Rabies^{SPLIT-TEV-mCherry} and imaged longitudinally over 16 days. (B-B'') Δ G-N^{PEST}Rabies^{SPLIT-TEV-mCherry} and B19 Δ G-Rabies Control (C-C'') infected hESCs derived neurons imaged at 4-10 and 16 days post-infection (p.i.). (D) Percentage of infected cells after administration of CTR Δ G-Rabies or Δ G-N^{PEST}Rabies^{SPLIT-TEV-mCherry} in presence or absence of Rapamycin after 4-10 and 16 days normalized to day 4 time-point (mean $\pm$ SEM). (E) mCherry signal intensity of Δ G-N^{PEST}Rabies^{SPLIT-TEV-mCherry} and Δ G-Rabies infected neurons normalized to day 4 time-point (mean $\pm$ SEM., scale as in D). Scale bar: 50 μ m. (F-G'') Section of hippocampus infected *in vivo* with either Δ G Rabies (cyan) or Δ G-N^{PEST}Rabies^{SPLIT-TEV-CRE} (magenta, G-G') at 1, 2 or 3 weeks p.i.. Scale bar: 50 μ m (H) Percentage of infected neurons at 1, 2 or 3 weeks p.i. of Δ G Rabies (black) or Δ G-N^{PEST}Rabies^{SPLIT-TEV-CRE} (grey) in hippocampus normalized to day 7 time-point (mean $\pm$ SEM).

Fig. 7.

Rapamycin induced Split TEV reconstitution and cleavage of the PEST domain in HEK cells. (A) Strategy for the pharmacological stabilization of the tagged viral protein. Rapamycin induces the dimerization of the Split TEV proteins which cleave the degron domain rescuing the viral proteins. (B) Split TEV rapamycin dose response in HEK. The TEV dependent cleavage of a TEV reporter increase with incremental concentration of rapamycin (0-10-50 nM). (C) The Split TEV cassette was cloned into the glycoprotein locus in the Rabies genome. A clear rapamycin dependent TEV activity was observed in HEK293T infected with the Split TEV expressing Rabies and transfected with a TEV reporter.

Fig. 8.

Short-term SiR^{mCherry} kinetics *in vivo*. (A) SiR^{mCherry} cassette design. (B) SiR mCherry-CRE injection in CA1 of Rosa-*LoxSTOPLox-YFP* mice. (C-E'') Confocal images of CA1 pyramidal

neurons infected with SiR^{mCherry-CRE} at 3-6 and 9 days p.i. Scale bar: 25 μ m. (**F-F''**) Percentage of YFP^{ON}, mCherry^{ON} and YFP^{ON}mCherry^{ON} neurons at 3-6 and 9 days p.i..

5 **Fig. 9**

Pharmacological reactivation of SiR. (**A**) Design of the doxycycline inducible AAV. The rTTA trans-activator is constitutive expressed by the virus and in presence of doxycycline it drives the TEV protease expression. (**B**) Diagram of AAV-TRE-TEV injection in the hippocampus of *Rosa-LoxP-STOP-LoxP-YFP* mice follow by SiR^{CRE-mCherry} in the same region 1 week after and Doxycycline administration. (**C-F**) Hippocampal pyramidal neurons infected with SiR and AAV-TRE-TEV, reactivated with Doxycycline at 2 or 3 weeks p.i.. Scale bar: 50 μ m. (**G**) Quantification of mCherry^{ON} neurons over the total YFP^{ON} infected neurons.

Fig. 10.

Δ G-Rabies induced mortality in cortex and hippocampus. (**A-A''**) Confocal images of cortical neurons and (**B-B''**) CA1 pyramidal neurons infected with Δ G-Rabies^{GFP} at 1, 2 and 3 weeks p.i.. Scale bar: 50 μ m. (**C**) Percentage of infected neurons at 1, 2 or 3 weeks p.i. of Δ G Rabies in cortex (black) or hippocampus (grey) normalized to 1 week time-point (mean $\pm$ SEM) (hippocampus, $92 \pm 3\%$ cell death at 2 weeks, $n = 3$ per time-point, one-way ANOVA, $F = 101$, $P = 2.4 \times 10^{-5}$; cortex $85 \pm 2\%$ cell death at 3 weeks, $n = 3$ per time-point, one-way ANOVA, $F = 17$, $P = 3.2 \times 10^{-3}$)

Fig. 11.

(A) Map of plasmid encoding SiR CRE-mCherryPEST vector genome; (B) Sequence of the plasmid shown in (A) T7 promoter sequence is shown with double underlining. Open reading frames are shown with single underlining and represent in order N-TEV-PEST, P, M, iCRE-2A-mCherryPEST and L proteins.

Fig. 12.

Map of plasmid encoding SiR vector having multiple cloning site in place of gene for envelope (G) protein.

Fig. 13.

Full sequence of plasmid illustrated in Fig. 12. Backbone plasmid (i.e. non-SiR) sequence shown in upper case with shading. Vector open reading frames are underlined, in order (from 5'-3') N-TEV-PEST protein, P protein, M protein, L protein. Multiple cloning sites (SbfI - NheI and PacI-AscI) shown in italic upper case. Within the N-TEV-PEST protein, N sequence is shown in lower case regular font, TEV cleavage site in upper case italics, PEST sequence in upper case double-underlined, and linker peptide sequences in lower case italics.

Fig. 14.

Sequence of negative-sense vector RNA genome obtained by transcription from the plasmid illustrated in Figs. 12 and 13.

Detailed Description of the Invention

Mononegavirales

The order Mononegavirales contains the families Bornaviridae, Filoviridae, Mymonaviridae, Nyamiviridae, Paramyxoviridae, Pneumoviridae, Rhabdoviridae and Sunviridae, as well as five unclassified genera Anphevirus, Arlivirus, Chengtivirus, Crustavirus, and Wastrivirus. Genera and species within those families are shown in the table below. Asterisks "*" in the following table denote type species.

Order Mononegavirales: families, genera, species, and viruses			
Family	Genus	Species	Virus (Abbreviation)
Bornaviridae	Bornavirus	Elapid 1 bornavirus	Loveridge's garter snake virus 1 (LGSV-1)
		Mammalian 1 bornavirus*	Borna disease virus 1 (BoDV-1)
			Borna disease virus 2 (BoDV-2)
		Passeriform 1 bornavirus	canary bornavirus 1 (CnBV-1)
			canary bornavirus 2 (CnBV-2)
			canary bornavirus 3 (CnBV-3))
		Passeriform 2 bornavirus	estrildid finch bornavirus 1 (EsBV-1)
		Psittaciform 1 bornavirus	parrot bornavirus 1 (PaBV-1)
			parrot bornavirus 2 (PaBV-2)
			parrot bornavirus 3 (PaBV-3)
			parrot bornavirus 4 (PaBV-4)
			parrot bornavirus 7 (PaBV-7)
		Psittaciform 2 bornavirus	parrot bornavirus 5 (PaBV-5)
		Waterbird 1 bornavirus	aquatic bird bornavirus 1 (ABBV-1)
			aquatic bird bornavirus 2 (ABBV-2)
Filoviridae	Cuevavirus	Lloviu cuevavirus*	Lloviu virus (LLOV)
	Ebolavirus	Bundibugyo ebolavirus	Bundibugyo virus (BDBV)
		Reston ebolavirus	Reston virus (RESTV)

Order Mononegavirales: families, genera, species, and viruses			
Family	Genus	Species	Virus (Abbreviation)
		<i>Sudan ebolavirus</i>	Sudan virus (SUDV)
		<i>Tai Forest ebolavirus</i>	Tai Forest virus (TAFV)
		<i>Zaire ebolavirus*</i>	Ebola virus (EBOV)
	<i>Marburgvirus</i>	<i>Marburg marburgvirus*</i>	Marburg virus (MARV)
			Ravn virus (RAVV)
<i>Mymonaviridae</i>	<i>Sclerotimonavirus</i>	<i>Sclerotinia sclerotimonavirus*</i>	Sclerotinia sclerotiorum negative-stranded RNA virus 1 (SsNSRV-1)
<i>Nyamiviridae</i>	<i>Nyavirus</i>	<i>Midway nyavirus</i>	Midway virus (MIDWV)
		<i>Nyamanini nyavirus*</i>	Nyamanini virus (NYMV)
		<i>Sierra Nevada nyavirus</i>	Sierra Nevada virus (SNVV)
	<i>Socyvirus</i>	<i>Soybean cyst nematode socyvirus*</i>	soybean cyst nematode virus 1 (SbCNV-1)
<i>Paramyxoviridae</i>	<i>Aquaparamyxovirus</i>	<i>Atlantic salmon paramyxovirus*</i>	Atlantic salmon paramyxovirus (AsaPV)
	<i>Avulavirus</i>	<i>Avian paramyxovirus 2</i>	avian paramyxovirus 2 (APMV-2)
		<i>Avian paramyxovirus 3</i>	avian paramyxovirus 3 (APMV-3)
		<i>Avian paramyxovirus 4</i>	avian paramyxovirus 4 (APMV-4)
		<i>Avian paramyxovirus 5</i>	avian paramyxovirus 5 (APMV-5)
		<i>Avian paramyxovirus 6</i>	avian paramyxovirus 6 (APMV-6)
		<i>Avian paramyxovirus 7</i>	avian paramyxovirus 7 (APMV-7)
		<i>Avian paramyxovirus 8</i>	avian paramyxovirus 8 (APMV-8)
		<i>Avian paramyxovirus 9</i>	avian paramyxovirus 9

Order <i>Mononegavirales</i> : families, genera, species, and viruses			
Family	Genus	Species	Virus (Abbreviation)
			(APMV-9)
		<i>Avian paramyxovirus 10</i>	avian paramyxovirus 10 (APMV-10)
		<i>Avian paramyxovirus 11</i>	avian paramyxovirus 11 (APMV-11)
		<i>Avian paramyxovirus 12</i>	avian paramyxovirus 12 (APMV-12)
		<i>Newcastle disease virus*</i>	avian paramyxovirus 1 (APMV-1)
	<i>Ferlavirus</i>	<i>Fer-de-Lance paramyxovirus*</i>	Fer-de-Lance virus (FDLV)
	<i>Henipavirus</i>	<i>Cedar henipavirus</i>	Cedar virus (CedV)
		<i>Ghanaian bat henipavirus</i>	Kumasi virus (KV)
		<i>Hendra virus*</i>	Hendra virus (HeV)
		<i>Mojiang henipavirus</i>	Mòjiāng virus (MojV)
		<i>Nipah virus</i>	Nipah virus (NiV)
	<i>Morbillivirus</i>	<i>Canine distemper virus</i>	canine distemper virus (CDV)
		<i>Cetacean morbillivirus</i>	cetacean morbillivirus (CeMV)
		<i>Feline morbillivirus</i>	feline morbillivirus (FeMV)
		<i>Measles virus*</i>	measles virus (MeV)
		<i>Peste-des-petits-ruminants virus</i>	peste-des-petits-ruminants virus (PPRV)
		<i>Phocine distemper virus</i>	phocine distemper virus (PDV)
		<i>Rinderpest virus</i>	rinderpest virus (RPV)
	<i>Respirovirus</i>	<i>Bovine parainfluenza virus 3</i>	bovine parainfluenza virus 3 (BPIV-3)
		<i>Human parainfluenza virus 1</i>	human parainfluenza virus 1 (HPIV-1)
		<i>Human</i>	human

Order Mononegavirales: families, genera, species, and viruses			
Family	Genus	Species	Virus (Abbreviation)
		<i>parainfluenza virus 3</i>	parainfluenza virus 3 (HPIV-3)
		<i>Porcine parainfluenzavirus 1</i>	porcine parainfluenza virus 1 (PPIV-1)
		<i>Sendai virus*</i>	Sendai virus (SeV)
	<i>Rubulavirus</i>	<i>Human parainfluenza virus 2</i>	human parainfluenza virus 2 (HPIV-2)
		<i>Human parainfluenza virus 4</i>	human parainfluenza virus 4a (HPIV-4a)
			human parainfluenza virus 4b (HPIV-4b)
		<i>Mapuera virus</i>	Mapuera virus (MapV)
		<i>Mumps virus*</i>	mumps virus (MuV)
			bat mumps virus (BMV)
		<i>Parainfluenza virus 5</i>	parainfluenza virus 5 (PIV-5)
		<i>Porcine rubulavirus</i>	La Piedad Michoacán Mexico virus (LPMV)
		<i>Simian virus 41</i>	simian virus 41 (SV-41)
<i>Pneumoviridae</i>	<i>Metapneumovirus</i>	<i>Avian metapneumovirus*</i>	avian metapneumovirus (AMPV)
		<i>Human metapneumovirus</i>	human metapneumovirus (HMPV)
	<i>Orthopneumovirus</i>	<i>Bovine respiratory syncytial virus</i>	bovine respiratory syncytial virus (BRSV)
		<i>Human respiratory syncytial virus*</i>	human respiratory syncytial virus A2 (HRSV-A2)
			human respiratory syncytial virus B1 (HRSV-B1)
			human respiratory

Order Mononegavirales: families, genera, species, and viruses			
Family	Genus	Species	Virus (Abbreviation)
Rhabdoviridae			syncytial virus S2 (HRSV-S2)
		<i>Murine pneumonia virus</i>	murine pneumonia virus (MPV)
	<i>Cytorhabdovirus</i>	<i>Alfalfa dwarf cytorhabdovirus</i>	alfalfa dwarf virus (ADV)
		<i>Barley yellow striate mosaic cytorhabdovirus</i>	barley yellow striate mosaic virus (BYSMV)
		<i>Broccoli necrotic yellows cytorhabdovirus</i>	broccoli necrotic yellows virus (BNYV)
		<i>Festuca leaf streak cytorhabdovirus</i>	festuca leaf streak virus (FLSV)
		<i>Lettuce necrotic yellows cytorhabdovirus*</i>	lettuce necrotic yellows virus (LNYV)
		<i>Lettuce yellow mottle cytorhabdovirus</i>	lettuce yellow mottle virus (LYMoV)
		<i>Northern cereal mosaic cytorhabdovirus</i>	northern cereal mosaic virus (NCMV)
		<i>Sonchus cytorhabdovirus</i>	sonchus virus (SonV)
		<i>Strawberry crinkle cytorhabdovirus</i>	strawberry crinkle virus (SCV)
		<i>Wheat American striate mosaic cytorhabdovirus</i>	wheat American striate mosaic virus (WASMV)
	<i>Dichorhavirus</i>	<i>Coffee ringspot dichorhavirus</i>	coffee ringspot virus (CoRSV)
		<i>Orchid fleck dichorhavirus*</i>	orchid fleck virus (OFV)
	<i>Ephemerovirus</i>	<i>Adelaide River ephemerovirus</i>	Adelaide River virus (ARV)
		<i>Berrimah ephemerovirus</i>	Berrimah virus (BRMV)
		<i>Bovine fever ephemerovirus*</i>	bovine ephemeral fever virus (BEFV)
		<i>Kotonkan ephemerovirus</i>	kotonkan virus (KOTV)
		<i>Obodhiang ephemerovirus</i>	Obodhiang virus (OBOV)
	<i>Lyssavirus</i>	<i>Aravan lyssavirus</i>	Aravan virus (ARAV)
		<i>Australian bat</i>	Australian bat

Order Mononegavirales: families, genera, species, and viruses			
Family	Genus	Species	Virus (Abbreviation)
		lyssavirus	lyssavirus (ABLV)
		Bokeloh bat lyssavirus	Bokeloh bat lyssavirus (BBLV)
		Duvenhage lyssavirus	Duvenhage virus (DUVV)
		European bat 1 lyssavirus	European bat lyssavirus 1 (EBLV-1)
		European bat 2 lyssavirus	European bat lyssavirus 2 (EBLV-2)
		Ikoma lyssavirus	Ikoma lyssavirus (IKOV)
		Irkut lyssavirus	Irkut virus (IRKV)
		Khujand lyssavirus	Khujand virus (KHUV)
		Lagos bat lyssavirus	Lagos bat virus (LBV)
		Mokola lyssavirus	Mokola virus (MOKV)
		Rabies lyssavirus*	rabies virus (RABV)
		Shimoni bat lyssavirus	Shimoni bat virus (SHIBV)
		West Caucasian bat lyssavirus	West Caucasian bat virus (WCBV)
	Novirhabdovirus	Hirame novirhabdovirus	Hirame rhabdovirus (HIRV)
		Oncorhynchus 1 novirhabdovirus*	infectious hematopoietic necrosis virus (IHNV)
		Oncorhynchus 2 novirhabdovirus	viral hemorrhagic septicemia virus (VHSV)
		Snakehead novirhabdovirus	snakehead rhabdovirus (SHRV)
	Nucleorhabdovirus	Datura yellow vein nucleorhabdovirus	datura yellow vein virus (DYVV)
		Eggplant mottled dwarf nucleorhabdovirus	eggplant mottled dwarf virus (EMDV)
		Maize fine streak	maize fine

Order Mononegavirales: families, genera, species, and viruses			
Family	Genus	Species	Virus (Abbreviation)
		<i>nucleorhabdovirus</i>	streak virus (MSFV)
		<i>Maize Iranian mosaic nucleorhabdovirus</i>	maize Iranian mosaic virus (MIMV)
		<i>Maize mosaic nucleorhabdovirus</i>	maize mosaic virus (MMV)
		<i>Potato yellow dwarf nucleorhabdovirus*</i>	potato yellow dwarf virus (PYDV)
		<i>Rice yellow stunt nucleorhabdovirus</i>	rice yellow stunt virus (RYSV)
			rice transitory yellowing virus (RTYV)
		<i>Sonchus yellow net nucleorhabdovirus</i>	sonchus yellow net virus (SYNV)
		<i>Sowthistle yellow vein nucleorhabdovirus</i>	sowthistle yellow vein virus (SYVV)
		<i>Taro vein chlorosis nucleorhabdovirus</i>	taro vein chlorosis virus (TaVCV)
	<i>Perhabdovirus</i>	<i>Anguillid perhabdovirus</i>	eel virus European X (EVEX)
		<i>Perch perhabdovirus*</i>	perch rhabdovirus (PRV)
		<i>Sea trout perhabdovirus</i>	lake trout rhabdovirus (LTRV)
	<i>Sigmavirus</i>	<i>Drosophila affinis sigmavirus</i>	<i>Drosophila affinis sigmavirus</i> (DAffSV)
		<i>Drosophila ananassae sigmavirus</i>	<i>Drosophila ananassae sigmavirus</i> (DAAnaSV)
		<i>Drosophila immigrans sigmavirus</i>	<i>Drosophila immigrans sigmavirus</i> (DImmSV)
		<i>Drosophila melanogaster sigmavirus*</i>	<i>Drosophila melanogaster sigmavirus</i> (DMelSV)
		<i>Drosophila obscura sigmavirus</i>	<i>Drosophila obscura</i>

Order Mononegavirales: families, genera, species, and viruses			
Family	Genus	Species	Virus (Abbreviation)
			sigmavirus (DObsSV)
		<i>Drosophila tristis</i> <i>sigmavirus</i>	<i>Drosophila</i> <i>tristis</i> sigmavirus (DTrISV)
		<i>Muscina stabulans</i> <i>sigmavirus</i>	<i>Muscina</i> <i>stabulans</i> sigmavirus (MStaSV)
	<i>Sprivivirus</i>	<i>Carp sprivivirus</i> *	spring viremia of carp virus (SVCV)
		<i>Pike fry</i> <i>sprivivirus</i>	grass carp rhabdovirus (GrCRV)
			pike fry rhabdovirus (PFRV)
			Tench rhabdovirus (TenRV)
	<i>Tibrovirus</i>	<i>Coastal Plains</i> <i>tibrovirus</i>	Coastal Plains virus (CPV)
		<i>Tibrogargan</i> <i>tibrovirus</i> *	Bivens Arm virus (BAV)
			<i>Tibrogargan</i> virus (TIBV)
	<i>Tupavirus</i>	<i>Durham tupavirus</i> *	Durham virus (DURV)
		<i>Tupaia tupavirus</i>	tupaia virus (TUPV)
	<i>Varicosavirus</i>	<i>Lettuce big-vein</i> <i>associated</i> <i>varicosavirus</i> *	lettuce big-vein associated virus (LBVaV)
	<i>Vesiculovirus</i>	<i>Alagoas</i> <i>vesiculovirus</i>	vesicular stomatitis Alagoas virus (VSAV)
		<i>Carajas</i> <i>vesiculovirus</i>	Carajás virus (CJSV)
		<i>Chandipura</i> <i>vesiculovirus</i>	Chandipura virus (CHPV)
		<i>Cocal</i> <i>vesiculovirus</i>	Cocal virus (COCV)
		<i>Indiana</i> <i>vesiculovirus</i> *	vesicular stomatitis Indiana virus (VSIV)
		<i>Isfahan</i> <i>vesiculovirus</i>	Isfahan virus (ISFV)

Order <i>Mononegavirales</i> : families, genera, species, and viruses			
Family	Genus	Species	Virus (Abbreviation)
		<i>Maraba vesiculovirus</i>	Maraba virus (MARAV)
		<i>New Jersey vesiculovirus</i>	vesicular stomatitis New Jersey virus (VSNJV)
		<i>Piry vesiculovirus</i>	Piry virus (PIRYV)
	Unassigned	<i>Flanders virus</i>	Flanders virus (FLAV)
		<i>Ngaingan virus</i>	Ngaingan virus (NGAV)
		<i>Wongabel virus</i>	Wongabel virus (WONV)
<i>Sunviridae</i>	<i>Sunshinevirus</i>	<i>Reptile sunshinevirus 1*</i>	Sunshine Coast virus (SunCV)
Unassigned	<i>Anphevirus</i>	<i>Xincheng anphevirus*</i>	Xīnchéng mosquito virus (XcMV)
Unassigned	<i>Arlivirus</i>	<i>Lishi arlivirus*</i>	Líshí spider virus 2 (LsSV-2)
Unassigned	<i>Chengtivirus</i>	<i>Tacheng chengtivirus*</i>	Tǎchéng tick virus 6 (TcTV-6)
Unassigned	<i>Crustavirus</i>	<i>Wenzhou crustavirus*</i>	Wēnzhōu crab virus 1 (WzCV-1)
Unassigned	<i>Wastrivirus</i>	<i>Sanxia wastrivirus*</i>	Sānxiá water strider virus 4 (SxWSV-4)

For more detail, see "Taxonomy of the order Mononegavirales: update 2016", Afonso et al., Arch. Virol. (2016) 161:2351-2360.

5

The viruses within the order (referred to here as "mononegaviruses") are enveloped viruses possessing genomes of non-infectious, linear, single-stranded, negative sense RNA.

10

The majority of mononegaviruses have non-segmented genomes. Without wishing to be bound by theory, it is believed that the methods and compositions described in this specification are more applicable to non-segmented viruses than to segmented viruses.

15

The genome has inverse-complementary 3' and 5' termini and is not covalently linked to any proteins.

5 The genomes generally have a conserved layout of coding and non-coding elements in the order 3'UTR (untranslated region) - core protein-coding genes - envelope protein-coding gene(s) - polymerase-coding gene - 5'UTR.

10 The core proteins encoded by the genome include the nucleoprotein (often designated "N" or "NP"), a protein which is often phosphorylated and so is referred to as a phosphoprotein (designated "P"), and a matrix protein ("M").

15 The polymerase is an RNA-dependent RNA polymerase and often the largest protein encoded by the virus and so may be referred to as the large (or "L") protein. The polymerase is relatively highly conserved within the order. The P protein may represent a co-factor for the L protein, i.e. the L and P proteins may be required to be present together for polymerase activity.

20

The terms N (or NP), P, M and L proteins are used in this specification to refer to any proteins which fulfil the corresponding roles in mononegaviruses, regardless of their normal designations.

25

The envelope protein is a transmembrane protein which may be glycosylated. In some mononegaviruses, it is therefore referred to as the glycoprotein (or "G" protein).

30

The virion comprises a helical ribonucleoprotein nucleocapsid, in which the genome is associated with the N, P and L proteins. This nucleocapsid is surrounded by the matrix and the membrane envelope layer.

35

Infection of a host cell results in release of the viral nucleocapsid into the cytosol, where transcription and replication take place via the RNA-dependent RNA polymerase.

5 All transcription takes place from a single promoter at the 3' end of the genome. After transcription of each gene, the polymerase either terminates or continues to the next gene downstream, yielding a gradient of mRNA production, with those genes closest to the 3' end of the genome being transcribed in
10 the highest copy number, and increasingly fewer transcripts produced for the genes towards the 5' end. Typically each virus produces 5 to 10 different mRNAs. The level of the nucleoprotein (often designated "N" or "NP") determines the timing of a switch between mRNA generation and genome
15 replication. Replication involves the production by the polymerase of full-length positive-sense antigenomes which are subsequently transcribed into full-length negative sense genome copies for packaging into virions.

20 The term "gene" in this specification is used to refer to a sequence in the vector genome which encodes an expression product and which directs expression of that expression product, typically under the control of the vector's promoter. The ultimate expression product may be an RNA (in which case
25 expression requires only transcription from the vector genome) or a protein (in which case expression requires transcription into mRNA by viral proteins followed by translation into protein.

30 Features of Rhabdoviruses, and in particular of the rabies virus, are described in more detail below as illustrative examples of mononegaviral biology.

Rhabdoviruses typically have genomes of around 11-15kb in
35 length. They variously infect vertebrates (including mammals and fish), insects and plants.

The rabies genome has short non-coding regions at its termini, designated the 3' leader (le) and 5' trailer (tr), which respectively initiate and terminate genome transcription and replication. The very 3' and 5' ends are inversely
 5 complementary. The termini also contain promoter sequences for transcription and replication, and for encapsidation of genomic RNA.

The 5 structural genes are ordered (3' to 5') N
 10 (nucleoprotein), P (phosphoprotein), M (matrix), G (glycoprotein) and L (large), with short non-coding intergenic regions (IGRs). Each structural gene comprises a coding region flanked by a 3' transcription initiation signal (TIS) (consensus sequence 3'-U-U-G-U-R-R-n-G-A-5' and a 5'
 15 transcription termination polyadenylation (TTP) signal (consensus sequence 3'A/U-C-U-U-U-U-U-U-U-G-5')

Infection is mediated by binding of the G protein to its receptor on the surface of the target cell. The virus is then
 20 internalised via the cell's endosomal transport pathway. The low pH in the endosome induces membrane fusion (between the endosomal membrane and the viral envelope), also mediated by the G protein, releasing the ribonucleoprotein complex into the cytoplasm.

25 Transcription of mRNA encoding viral protein then begins, mediated by the viral polymerase (comprising P and L proteins). Transcription is believed to proceed via a stop-start mechanism, beginning at the 3' end of the genome and
 30 progressing towards the 5' end, producing 6 consecutive transcripts, firstly of the leader RNA, and then each of the N, P, M, G and L genes in turn. The polymerase is believed to dissociate from the template at each stop signal and re-initiate poorly at the next start signal. This results in a
 35 gradient of mRNA production with the amount of leader transcript being greatest, followed by N, P, M, G and L transcripts.

Later in infection, the polymerase switches from mRNA production to replication of the viral genome, which proceeds via a full length positive sense RNA intermediate. Both the positive sense intermediate and the progeny negative sense viral genomes are packaged with protein N to form nucleoprotein complexes. The M protein plays a regulatory role in determining the timing of the switch between transcription and replication, as well as being involved in recruitment of RNP nucleocapsids to the host cell plasma membrane, association with glycoprotein G, and budding of the progeny virion particles from the cell.

For a review, including more detail about the roles and functions of the individual proteins, see Albertini et al., Rabies Virus Transcription and Replication, Advances in Virus Research 79, December 2011 (ISSN 0065-3527; DOI: 10.1016/B978-0-12-387040-7.00001-9).

An exemplary sequence of a rabies N protein is as follows:

MDADKIVFKVNNQVVSLKPEIIVDQYEEKYPAIKDLKKPCITLGKAPDLNKAYKSVLSGM
 SAAKLNPDDVCSYLAAAMQFFEGTCPEDWTSYGIVIAARKGDKITPGSLVEIKRTDVEGNW
 ALTGGMELTRDPTVPEHASLVGLLLSLYRLSKI SGQNTGNYKTNIADRIEQIFETAPFVK
 IVEHHTLMTTHKMCANWSTIPNFRFLAGTYDMFFSRIEHLYSAIRVGT VVTAYEDCSGLV
 SFTGFIKQINLTAREAILYFFHKNFEEEEIRRMFEPGQETAVPHSYFIHFRSLGLSGKSPY
 SSNAVGHVFNLIH FVGCYMGQVRSLNATVIAACAPHEMSVLGGYLGEEFFGKGTFERRFF
 RDEKELQEYEAELTKTDVALADDGTVNSDDEDYFSGETRSP EAVYTRIMMNGGRLKRSH
 IRRYVSVSSNHQARPNSFAEFLNKTYSSDS

An exemplary sequence of a rabies P protein is as follows:

MSKIFVNPSAIRAGLADLEMAEETVDLINRNIEDNQAHLQGEPIEVDNLPEDMGRHLHDD
 GKSPNHGEIAKVGE GKYREDFQMDEGEDPSFLFQSYLENVGVQIVRQMRSGERFLKIWSQ
 TVEEIIISYVAVNFPNPPGKSSDKSTQTTGRELKKETTPTPSQRESQSSKARMAAQIASG
 PPALWSATNEEDDL SVEAEIAHQIAESFSKKYKFPSRSSGILLYNFEQLKMNLD DIVKE
 AKNVPGVTRLAHDGSKLPLRCVLGWVALANSKKFQLLVESDKLSKIMQDDLNRYTSC

An exemplary sequence of a rabies M protein is as follows:

MNLLRKIVKNRRDEDTQKSSPASAPLDDDDLWLPPPEYVPLKELTGKKNMRNFCINGRVK

VCSPNGYSFRILRHILKSFDEIYSGNHRMIGLVKVIGLALSGSPVPEGLNWVYKLRRTF
IFQWADSRGPLEGELEEYSQEITWDDDETFVGLQIRVIAKQCHIQGRVWCINMNPRACQL
WSDMSLQTQRSEEDKDSSILLE

5 An exemplary sequence of a rabies L protein is as follows:

MLDPGEVYDDPIDPIELEAEPRGTPIVPNILRNSDYNLNSPLIEDPARLMLEWLKTGNRP
YRMTLTDNCSRSFRVLKDYFKKVDLGLSKVGGMAAQSMISLWLYGAHSESNRSRRCITDL
AHFYSSKSSPIEKLLNLTGNGRLRIPPEGVLSCLERVDYDNAFGRYLANTYSSYLFFHVI
10 TLYMNALDWDEEKTILALWKDLTSVDIGKDLVKFKDQIWGLLIVTKDFVYSQSSNCLFDR
NYTLMKDLFLSRFNSLMVLLSPPEPRYSDDLISQLCQLYIAGDQVLSMCGNSGYEVIKI
LEPYVVNSLVQRAEKFRPLIHSLGDFPVFIKDKVSQLEETFGPCARRFFRALDQFDNIHD
LVFVFGCYRHWGHPYIDYRKGLSKLYDQVHLKKMIDKSYQECLASDLARRILRWGFDKYS
KWLDSRFLARDHPLTPYIKTQTWPPKHIVDLVGDTWHKLPITQIFEIPESMDPSEILDD
15 KSHSFTRTRLASWLSENRGGPVPSEKVIITALS KPPVNPREFLRSIDLGGLPDEDLIIGL
KPKERELKIEGRFFALMSWNLRLYFVITEKLLANYILPLFDALMTDNLNKVFKKLIDRV
TGQGLLDYSRVTYAFHLDYEKWNHQRLESTEDVFSVLDQVFGGLKRVFSRTHEFFQKAWI
YYSDRSDLIGLREDQIYCLDASNGPTCWNGQDGGLEGLRQKGWSLVSLLMIDRESQIRNT
RTKILAQGDNQVLCPTYMLSPGLSQEGLLYELERISRNALS IYRAVEEGASKLGLIIKKE
20 ETMCSYDFLIYGKTPFRGNILVPESKRWARVSCVSNDQIVNLANIMSTVSTNALTVAQH
SQSLIKPMRDFLLMSVQAVFHYLLFSPILKGRVYKILSAEGESFLLAMSRIIYLDPSLGG
ISGMSLGRFHIRQFSDPVSEGLSFWREIWLSSQESWIHALCQEAGNPDLGERTLESFTRL
LEDPTTLNIRGGASPTILLKDAIRKALYDEVKVENSEFREAILLSKTHRDNFILFLISV
EPLFPRFLSELFSSSFLGIPESIIGLIQNSRTIRRQFRKSLSKTLEESFYNSEIHGISRM
25 TQTPQRVGGVWPCSSERADLLREISWGRKVVGTTVPHVPSEMLGLLPKSSISCTCGATGGG
NPRVSVSVLPSFDQSFFSRGPLKGYLGSSTSMSTQLFHAWEKVTNVHVVKRALSLKESIN
WFITRDSNLAQALIRNIMSLTGPDFPLEEAPVFKRTGSALHREKFSARYSEGGSVCPNL
LSHISVSTDTMSDLTQDGKNYDFMFQPLMLYAQTWTSELVQRDTRLRDSTFHWHLRCNRC
VRPIDDVTLETSQIFEFDPVSKRISRMVSGAVPHFQRLPDIRLRPGDFESLSGREKSHHI
30 GSAQGLLYSILVAIHDSGYNDGTIFPVNIYGKVS PRDYLRGLARGVLIGSSICFLTRMTN
ININRPLELVSGVISYILLRLDNHPSLYIMLREPSLRGEIFSIPQKIPAAYPPTMKEGNR
SILCYLQHVLRYEREIITASPENDWLWIFSDFRSAKMTYLSLITYQSHLLLQRVERNLSK
SMRDNLRLQLSSLMRQVLGGHGEDTLESDDNIQRLLKDSLRRTRWVDQEV RHAARTMTGDY
SPNKKVSRKVGCSSEWVCSAQQVAVSTSANPAPVSELDIRALSKRFQNP LISGLRVVQWAT
35 GAHYKLPILDDLNVFPSLCLVVGDSGGISRVLNMFDAKLVFNSSLLEVNDLMASGTH
PLPPSAIMRGGNDIVSRVIDLDSIWEKPSDLRNLATWKYFQSVQKQVNMSYDLIICDAEV
TDIASINRITLLMSDFALSIDGPLYLVFKTYGTMLVNP NYKAIQHLSRAFPSTGFTITQV
TSSFSSELYLRFSKRGKFFRDAEYLTSSTLREMSLVLFNCSSPKSEMQRARSLNYQDLVR
GFPEEIIISNPYNEMIITLIDSDVESFLVHKMVDDLELQRTLSKVAIIIAIMIVFSNRVF
40 NVSKPLTDPSFYPPSPDKILRHFNICCSTMMYLSTALGDVPSFARLHDLYNRPITYYFRK
QVIRGNVYLSWSWSNDTSVFKRVACNSSLSLSSH WIRLIYKIVKTTRLVGSIKDLSREVE
RHLHRYNRWITLEDIRSRSSLLDYSC

Thus, when used in the vectors of the present invention, a
45 rabies N protein may have the sequence shown above, or may
have at least 70% identity thereto, e.g. at least 75%, at
least 80%, at least 85%, at least 90%, at least 95%, at least
96%, at least 97%, at least 98%, or at least 99% identity
thereto.

50

5 A rabies P protein may have the sequence shown above, or may have at least 70% identity thereto, e.g. at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity thereto.

10 A rabies M protein may have the sequence shown above, or may have at least 70% identity thereto, e.g. at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity thereto.

15 A rabies L protein may have the sequence shown above, or may have at least 70% identity thereto, e.g. at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity thereto.

20 An exemplary sequence of a rabies G protein (strain B19G) is as follows:

25 MVPQALLFVPLLVFPLCFGKFPIYTIPTDKLGPWSPIDIHHLSCPNNLVVEDEGCTNLSGFSY
MELKVGYILAIKVNGFTCTGVVTEAETYTNFVG YVTTTFKRKHFRPTPDACRAAYNWKMAGD
PRYEESLHNPYPDYRWLRTVKTTKESLVIISPSVADLDPYDRSLHSRVFPPSGKCSGVAVSST
YCSTNHDTIWMPEPNRLGMSCDIFTNSRGKRASKGSETCGFVDERGLYKSLKGACKLKLKG
VLGLRLMDGTWVSMQTSNETKWCPPDKLVNLHDFRSDEIEHLVVEELVRKREECLDALESIM
TTKSVSFRRLSHLRKLVPGFGKAYTIFNKTLMEADAHYKSVRTWNEILPSKGCLRVGGRCHP
HVNGVFFNGIILGPDGNVLIPEMQSSLLQQHMELLESSVIPLVHPLADPSTVFKDGDEAEDF
VEVHLDPDVHNQVSGVDLGLPNWGKYVLLSAGALTALMLIIIFLMTCCRRVNRSEPTQHNLRGTT
GREVSVTPQSGKIISSWESHKSGGETRL

30 where the extracellular domain is shown in regular font, the transmembrane domain underlined, and the intracellular domain in italics.

35 An alternative so-called "optimised" G protein, having B19 strain intracellular domain but the extracellular domain from a different strain, has the sequence:

MVPQALLFVPLLVFPLCFGKFPIYTIPDKLGPWSPIDIHHLSCPNNLVVEDEGCTNLSGFSY
 MELKVGYISAIKMNGFTCTGVVTEAETYTNFVGIVTTTFKRKHFRPTPDACRAAYNWKMAGD
 PRYEESLHNPYPDYHWLRTVKTTKESLVIISPSVADLDPYDRSLHSPVFPGGNC SGVAVSST
 YCSTNHDTIWMPENPRLGMSCDIFTNSRGKRASKGSETCGFVDERGLYKSLKGACKLKL CG
 5 VLGLRLMDGTWVAMQTSNETKWCPPGQLVNLHDFRSDEIEHLVVEELVKKREECLDALESIM
 TTKSVSFRRLSHLRKLVPGFGKAYTIFNKTLMEADAHYKSVRTWNEIIPSKGCLRVGGRCHP
 HVNGVFFNGIILGPDGNVLIPEMQSSLLQQHMELLVSSVIPLMHPLADPSTVFKNGDEAEDF
 VEVHLPDVHERISGVDLGLPNWGKYVLLSAGALTALMLIIFLMTCWRRVNRSEPTQHNL RGT
 GREVSVTPQSGKIISSWESHKSGGETRL

10 where the extracellular domain is shown in regular font, the transmembrane domain underlined, and the intracellular domain in italics.

15 A rabies G protein may have either of the sequences shown above, or may have at least 70% identity thereto, e.g. at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity thereto.

20 A pseudotyped envelope protein for use with a vector based on a rabies virus may have the intracellular domain and optionally the transmembrane domain from either of the G protein sequences shown above, or may have at least 70% identity thereto, e.g. at least 75%, at least 80%, at least
 25 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity thereto. The extracellular domain may have at least 70% identity, e.g. at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%
 30 identity to one of the extracellular domain sequences shown above or may be selected depending on the cell to which the vector will be targeted.

35 Each gene (or open reading frame) encoded by the vector is typically associated with appropriate regulatory signals to ensure expression (i.e. transcription and translation). This applies to any heterologous genes carried by the vector as

well as to the gene encoding the replication modulator protein and other viral proteins. Thus each gene may comprise a transcriptional start signal and a transcriptional stop signal. Transcriptional start signals are typically C-rich sequences of around 8 bases in length, such as ACATCCCT and ACACCCCT. Stop signals are typically stretches of poly(A), such as AAAAAA. Each gene may also comprise a Kozak sequence to ensure appropriate transcriptional initiation. Additional regulatory sequences, including non-mononegaviral elements, such as internal ribosomal entry sites (IRES) may also be present to ensure desired levels of gene expression, especially for any heterologous gene or genes.

Percent (%) amino acid sequence identity with respect to a reference sequence is defined as the percentage of amino acid residues in a candidate sequence that are identical with the amino acid residues in the reference sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity. % identity values may be determined by WU-BLAST-2 (Altschul et al., Methods in Enzymology, 266:460-480 (1996)). WU-BLAST-2 uses several search parameters, most of which are set to the default values. The adjustable parameters are set with the following values: overlap span = 1, overlap fraction = 0.125, word threshold (T) = 11. A % amino acid sequence identity value is determined by the number of matching identical residues as determined by WU-BLAST-2, divided by the total number of residues of the reference sequence (gaps introduced by WU-BLAST-2 into the reference sequence to maximize the alignment score being ignored), multiplied by 100.

Monosynaptic circuit tracing

Rabies virus vectors lacking a functional glycoprotein G are well known for mapping synaptic connections between neurons. See, for example, US 8,334,095 B2.

Rabies virus spreads selectively between synaptically connected neurons, exclusively in the retrograde direction (except for some sensory neurons). Rabies viruses which lack a G glycoprotein are unable to bud from the surface of an infected cell. Thus, a vector based on a rabies virus in which the G gene has been deleted or inactivated (Δ G-Rabies, or RVdG) can only propagate to adjacent cells if a gene encoding a G glycoprotein (or another pseudotyped envelope protein) is supplied in trans, e.g. via a helper virus, such as an adeno-associated viral (AAV) vector. If the Δ G-Rabies vector and the helper virus are carefully targeted to the same cell, the rabies virus is able to propagate to cells synaptically adjacent, but no further (as the adjacent cells do not express the envelope protein required for further propagation). This therefore enables synaptic connections to be visualised and mapped, e.g. via a fluorescent marker protein carried by the Δ G-Rabies vector.

Many variations of the basic labelling method are known. In some examples of mammalian use, Δ G-Rabies is pseudotyped with the envelope protein from avian sarcoma leukosis virus (ASLV) type A or type B (EnvA, EnvB), which are specific for the receptors TVA and TVB respectively. Since these proteins are not naturally expressed in mammals, transduction by the relevant pseudotyped vector can be limited to cells specifically engineered to express TVA or TVB.

Some approaches to synaptic mapping use animal models which have been engineered to express a viral recombinase which may target recognition sites in the rabies vector or helper virus genomes. Further levels of control can be achieved by placing the recombinase gene in the recipient animal under the transcriptional control of an inducible promoter. Alternatively, the recipient animal may have been engineered to possess genomic recognition sites for such a recombinase, in order that a cellular genomic recombination event takes

place after transduction with a vector encoding the recombinase protein.

5 Commonly, the Cre-Lox pairing of recombinase and recognition sites is used, where the Cre recombinase acts on the Lox recognition sites.

10 However, deletion or inactivation of the G glycoprotein does not interfere with the usual process of viral transcription, protein synthesis and replication of the core within the recipient cell. The accumulation of viral RNA and protein within the cell has the effect that the labelled cells typically remain viable for only 1-2 weeks, as a result of cell viability being directly compromised by viral proteins,
15 an immune response against the infected cell, or other mechanisms.

Switchable vectors

20 The vectors of the present invention exploit the coupled nature of mononegaviral transcription and replication to provide a switch for the viral transcription-replication cycle within infected cells, turning the virus on or off depending on the presence (and activity) of specific activating agents or inhibitory agents.

25 As already described above, the vector encodes a replication modulator protein switchable between a configuration displaying a degron which targets that protein for degradation (e.g. by the proteasome) and a configuration which does not display that degron and is hence more stable. These
30 configurations of the replication modulator protein are designated "targeted" and "untargeted" to reflect the presence or absence of the degron.

35 Thus, while the replication modulator protein exists primarily in the targeted configuration, the virus is unable to build up a significant quantity of the relevant protein in a functional

form. Although a short period of viral transcription and replication may be possible immediately after primary infection, viral transcription and replication will stall thereafter. If this configuration is maintained for a sufficient period of time, the virus will eventually be cleared from the cell, since the other viral proteins and the RNA genome itself will be broken down by the normal mechanisms of the cell.

The length of time for which the virus will persist in the infected cell in the targeted configuration will vary depending on the particular cell and virus in question. In general, though, a vector of this sort based on the rabies virus is likely to be cleared from a neural cell within a period of approximately two weeks to one month.

In general, the switch between configurations results from interaction between the regulator moiety and the activating agent or inhibitory agent, as appropriate. The interaction may involve covalent or non-covalent modification of the regulator moiety.

Covalent modification may be achieved by enzyme action, i.e. the activating agent or inhibitory agent is an enzyme and the regulator moiety is a substrate for that enzyme. An example of covalent modification is proteolytic cleavage.

Non-covalent modification may be achieved by binding of the activating agent or inhibitory agent to the regulator moiety, i.e. the activating agent or inhibitory agent may be a ligand for the regulator moiety.

Thus an activating agent may stabilise a replication modulator protein by "cleavage-induced stabilisation" or "ligand-induced stabilisation". Conversely, an inhibitory agent may destabilise a replication modulator protein by "cleavage-induced destabilisation" or "ligand-induced destabilisation".

(These may also be referred to as "cleavage-induced degradation" or "ligand-induced degradation".) Other mechanisms may be possible.

5 It will be apparent that the switch between targeted and untargeted configurations may be reversible or irreversible, depending on the nature of the replication modulator protein and the activating or inhibitory agent. For example, where the activating agent or inhibitory agent is a ligand for the
10 modulator protein, the switch between the two configurations may be reversible. Where the switch between the configurations is mediated by covalent modification such as proteolysis, the switch is likely to be irreversible.

15 In the context of the present invention the degron may be any feature which confers a particularly short half-life to a protein, e.g. by marking it for degradation. Many different types of degron are known. Some act by marking a protein for ubiquitinylation (ubiquitin-dependent degrons), while others
20 are ubiquitin-independent.

The degron may be a PEST sequence. A PEST sequence is a peptide sequence motif typically at least 12 amino acids in length, hydrophilic, and rich in proline, glutamic acid,
25 serine and threonine. Li et al. (Generation of Destabilized Green Fluorescent Protein as a Transcription Reporter, J. Biol. Chem. 1998; 273(52): 34970-5) describe a PEST sequence at residues 423-450 of mouse ornithine decarboxylase, having the sequence HGFPPEVEEQDDGTLPMSCAQESGMDRH and variants thereof
30 which also have destabilising activity. The region from residues 422 to 461 of this protein ("mODC(422-461)") is employed in the examples below and has the sequence SHGFPPEVEEQDDGTLPMSCAQESGMDRHPAACASARINV.

35 The C-terminal region of NPDC-1 (Neural proliferation and differentiation control protein-1) also possesses a PEST motif, having the sequence KELDTASSDEENEDGDFTVYECPLAPTGEMEVR.

See (NPDC-1, a Novel Regulator of Neuronal Proliferation, Is Degraded by the Ubiquitin/Proteasome System through a PEST Degradation Motif, Spencer et al., J. Biol. Chem. 2004; 279, 37069-37078)

5

Thus the replication modulator protein may comprise a PEST sequence. A replication modulator protein comprising a PEST sequence will typically be an inhibitory modulator protein since it will tend to be degraded until the PEST sequence is removed.

10

A PEST sequence will often be located C-terminal of the viral protein moiety, e.g. at the C-terminus of the modulator protein.

15

Thus, the replication modulator protein will typically comprise a viral protein moiety and a regulator moiety located C-terminal of the viral protein moiety, wherein the regulator moiety comprises a protease cleavage site and a PEST sequence.

20

The following amino acid sequence is an example of a replication modulator protein comprising a rabies virus N protein as viral protein moiety, and a regulator moiety comprising a TEV protease cleavage site and a PEST sequence ("N-TEV-PEST"):

25

MDADKIVFKVNNQVVSLKPEIIVDQYEEKYPAIKDLKKPCITLGKAPDLNKAYKSVLSGM
 SAAKLNPDDVCSYLAAMQFFEGTCPEDWTSYGIVIAARKGDKITPGSLVEIKRTDVEGNW
 ALTGGMELTRDPTVPEHASLVGLLLSLYRLSKISGQNTGNYKTNIADRIEQIFETAPFVK
 IVEHHTLMTTHKMCANWSTIPNFRFLAGTYDMFFSRIEHLYSAIRVGT VVTAYEDCSGLV
 SFTGFIKQINLTAREAILYFFHKNFEEIIRRMFEPGQETAVPHSYFIHFRSLGLSGKSPY
 SSNAVGHVFNLIHFVGCYMGQVRSLNATVIAACAPHEMSVLGGYLGEFFGKGTFERRFF
 RDEKELQEYEAELTKTDVALADDGTVNSDDEDFSGETRSP EAVYTRIMMNGGRLKRSH
 IRRYVSVSSNHQARPNSFAEFLNKTYSSDSGSGENLYFQSGSGSHGFPPEVEEQDDGTL
 P

30

35

The N protein viral protein moiety is shown in regular font. the regulator moiety is shown in italics, with single

underlining for the TEV sequence and double underlining for the PEST sequence.

5 The degron may be an N-terminal amino acid or short N-terminal sequence motif. It is well established that proteins having residues other than Met, Gly or Val at their N-terminus tend to be less stable than those having Met, Gly or Val. This is referred to as the N-end rule. See, for example, Varshavsky, A. (2011) "The N-end rule pathway and regulation by
10 proteolysis"; Protein Science 20: 1298-1345.

Various mechanisms are involved in determining the stability of a protein via its N-terminus, including the so-called Arg/N-end rule pathway (which involves N-terminal arginylation
15 of protein substrates) and the Ac/N-end rule pathway (which involves co-translational N-terminal acetylation of some proteins having Met, Ala, Val, Ser, Thr or Cys. In some cases, a protein's original N-terminus will be processed by endogenous cellular enzymes (e.g. Met-aminopeptidases). In
20 such cases, the N-terminal residue present after processing may represent the degron.

As a guideline, the order of stability in mammalian cells (least stable / shortest half life to most stable / longest
25 half life) is roughly Gln (Q), Arg (R), Glu (E), Phe (F), Asp (D), Cys (C), Lys (K), Asn (N), Ser (S), Tyr (Y), Trp (W), His (H), Ala (A), Leu (L), Thr (T), Ile (I), Pro (P), Gly (G), Met (M), Val (V).

30 Different alternatives may be tested for any given protein to identify a suitable combination of residues to be displayed at the N-terminus before and after cleavage, to achieve the required differential in protein stability.

35 Thus the modulator protein may comprise an N-terminal regulator moiety and a viral protein moiety downstream of the regulator moiety, wherein the regulator moiety comprises a

first residue at its N-terminus and a cleavage site for a cognate protease, and wherein, after cleavage by the cognate protease, the viral protein moiety has at its N-terminus a second residue which confers a different half life than the first residue.

Thus an inhibitory modulator protein may comprise a first N-terminal residue, and be cleavable (e.g. by a protease) to expose a second N-terminal residue which confers greater stability than the first N-terminal residue. For example, the first N-terminal residue may be Arg or Lys (or the sequence Arg-Lys), and the second N-terminal residue may be Val.

Conversely, an inhibitable modulator protein may be expressed with a first N-terminal residue, and be cleavable (e.g. by a protease) to expose a second N-terminal residue which confers lower stability than the first N-terminal residue. For example, the first N-terminal residue may be Met or Val, or the sequence Met-Val, and the second N-terminal residue may be Arg or Lys (or the sequence Arg-Lys).

In both cases, the regulator moiety will be located at the N-terminus of the molecule and may comprise a cleavage site for a cognate protease immediately upstream of the chosen second N-terminal residue.

Typically, the modulator protein will comprise one or more surface-exposed Lys residues to act as sites for ubiquitinylation.

The degron technologies described above provide examples of cleavage-induced stabilisation and cleavage-induced destabilisation.

Other conditional degron technologies are described in Kanemaki et al., Eur. J. Physiol. (2013) 465: 419-425 and

references cited therein, and details of some of these technologies are provided below.

5 An example of ligand-induced destabilisation is the auxin-inducible degron. This utilises an auxin to induce targeting for degradation. Binding of an auxin (such as indole-3-acetic acid (IAA) or 1-naphthaleneacetic acid (NAA)) to the plant protein TIR1 (a component of the ubiquitin ligase designated SCF) allows TIR1 to interact with proteins of the AUX/IAA family, resulting in ubiquitinylation of the AUX/IAA protein and its subsequent degradation by the proteasome. Other components of the SCF ubiquitin ligase are well conserved in most eukaryotes. Thus, an AUX/IAA protein such as IAA17 may be employed as a regulator moiety in cells which express TIR1 (e.g. eukaryotic non-plant cells which have been engineered to express TIR1). Addition of an auxin to the cell will then result in degradation of any protein comprising an AUX/IAA moiety. Thus an inhibitable regulator moiety may comprise an AUX/IAA protein.

20 The DD-FKBP and LID-FKBP degron systems are based on the FK506 binding protein 12 (FKBP12), and provide examples of ligand-induced stabilisation and ligand-induced destabilisation, respectively.

25 Destabilisation domains of FKBP12 ("DD-FKBP") are continuously degraded but are stabilised by the presence of a ligand called "Shield-1", which is a cell-permeable analogue of rapamycin. (See Banaszynski et al., Cell. 2006 September 8; 126(5): 995-1004.) Thus DD-FKBP can be used as an inhibitory regulator moiety, which can be stabilised by introduction of Shield-1. The DD-FKBP moiety can be located either N-terminal or C-terminal of the viral protein moiety in the modulator protein.

35 The LID-FKBP system is a ligand-induced destabilisation (or degradation) system in which a synthetic 19 amino acid degron having the sequence TRGVEEVAEGVLLRRRGN is fused to the C-

terminus of the FKBP12 protein (or a Phe36Val variant thereof designated FKBP12^{F36V}). The peptide binds to the ligand binding pocket of FKBP12 thus sequestering it. Addition of another ligand for the same binding site, such as Shield-1, induces release and exposure of the degron peptide, which consequently marks the entire protein for degradation. Thus, a regulator moiety may comprise a LID-FKBP moiety. A regulator moiety comprising a LID-FKBP moiety would typically be located C-terminal of the viral protein moiety in a replication modulator protein, and ideally at the C-terminus of the modulator protein.

Binding of a hydrophobic ligand to a regulator moiety may also be used to mark a protein for degradation. For example, the so-called "HaloTag" system employs a modified haloalkane dehydrogenase (the HaloTag) and a small molecule hydrophobic ligand (e.g. HyT13) which covalently binds to the active site of the modified haloalkane dehydrogenase. It appears that the binding of the ligand to the protein marks the protein for degradation by the proteasome. Thus an inhibitable regulator moiety may comprise a HaloTag, which and may be located either N-terminal or C-terminal of the viral protein moiety.

It will be clear from the discussion above that the switch between targeted and untargeted configurations of the replication modulator protein (or vice versa) may be implemented by cleavage of the regulator moiety from the viral protein moiety. This is typically accomplished by a protease. In such cases, the regulator moiety typically comprises a degron and a cleavage site for the protease, wherein the protease cleavage site is located between the viral protein moiety and the degron. The replication modulator protein may comprise a linker peptide located between the viral protein moiety and the regulator moiety. Additionally or alternatively, the regulator moiety may comprise a linker peptide between the protease cleavage site and the degron.

Unless one of the components of the protein have any specific functional requirements the moieties of the replication modulator may be in any appropriate orientation. For example, the viral protein may be located N-terminal of the regulator moiety, or the regulator moiety may be located N-terminal of the viral protein, with the linker (where present) between them. However, some degron sequences do have a requirement for being located at the N- or C-terminus of the protein.

A peptide linker is typically between 3 and 30 amino acids in length, with a high proportion of small and hydrophilic amino acid residues (e.g. glycine and serine) to provide the required flexibility without compromising aqueous solubility of the molecule. It may also contain the cleavage site on which the protease acts. The residues other than the cleavage site (and any other sequence required for recognition by the protease) may comprise at least 50% glycine and serine residues, at least 60% glycine and serine residues, at least 70% glycine and serine residues, at least 80% glycine and serine residues, or at least 90% glycine and serine residues.

Proteases

The protease may be orthogonal to the target cell, which is to say that the protease recognises a cleavage site not found in native proteins encoded by and expressed in the target cell (i.e. in the proteome of the target cell).

Thus the particular protease may vary depending on the intended target cell to which the vector is to be delivered. The target cell will typically be a neuron, in which case the protease should not act on any native cellular proteins expressed within that neuron.

It will also be apparent that the selected protease should not act on other proteins encoded by the vector.

Examples of suitable proteases include:

- Viral proteases, such as Tobacco Etch Virus protease (TEVp) and human rhinovirus (HRV) 3C protease;
- Factor Xa;
- Enterokinase;
- Granzyme B;
- Thrombin.

Consensus cleavage sites for these proteases are as follows, where "\ " indicates the position of the cleaved peptide bond:

Protease	Cognate cleavage site(s)
TEVp	ENLYFQ\G ENLYFQ\S
HRV 3C protease	LEVLFQ\GP
Factor Xa	IEGR\
Enterokinase	DDDDK\
Thrombin	LVPR\GS

Pharmaceutical compositions and methods of treatment

The agents described herein can be formulated in pharmaceutical compositions. These compositions may comprise, in addition to one of the above substances, a pharmaceutically acceptable excipient, carrier, buffer, stabiliser or other materials well known to those skilled in the art. Such materials should be non-toxic and should not interfere with the efficacy of the active ingredient. The precise nature of the carrier or other material may depend on the route of administration, which may be by any suitable route, and may be oral or parenteral. Because of the difficulties experienced with oral delivery of peptide agents, parenteral administration may prove the most suitable. Suitable parenteral routes include but are not limited to intravenous, intramuscular, intraperitoneal, cutaneous, subcutaneous, transdermal, and other mucosal routes such as nasal, buccal, rectal and vaginal routes. Examples of suitable compositions

and methods of administration are provided in Esseku and Adeyeye (2011) and Van den Mooter G. (2006).

Pharmaceutical compositions for oral administration may be in tablet, capsule, powder or liquid form. A tablet may include a solid carrier such as gelatin or an adjuvant. Liquid pharmaceutical compositions generally include a liquid carrier such as water, petroleum, animal or vegetable oils, mineral oil or synthetic oil. Physiological saline solution, dextrose or other saccharide solution or glycols such as ethylene glycol, propylene glycol or polyethylene glycol may be included.

For intravenous, cutaneous or subcutaneous injection, the active ingredient will be in the form of a parenterally acceptable aqueous solution which is pyrogen-free and has suitable pH, isotonicity and stability. Those of relevant skill in the art are well able to prepare suitable solutions using, for example, isotonic vehicles such as Sodium Chloride Injection, Ringer's Injection, Lactated Ringer's Injection. Preservatives, stabilisers, buffers, antioxidants and/or other additives may be included, as required.

Whatever the nature of the active agent that is to be given to an individual (e.g. a virion, encapsulated nucleic acid molecule, or other pharmaceutically useful agent according to the present invention), administration is preferably in a "prophylactically effective amount" or a "therapeutically effective amount" (as the case may be, although prophylaxis may be considered therapy), this being sufficient to show benefit to the individual. The actual amount administered, and rate and time-course of administration, will depend on the nature and severity of what is being treated. Prescription of treatment, e.g. decisions on dosage etc., is within the responsibility of general practitioners and other medical doctors, and typically takes account of the disorder to be treated, the condition of the individual patient, the site of

delivery, the method of administration and other factors known to practitioners. Examples of the techniques and protocols mentioned above can be found in Remington's Pharmaceutical Sciences, 20th Edition, 2000, pub. Lippincott, Williams & Wilkins.

Examples

We fused a proteasome-targeting domain to each protein of the Rabies virus (individually or in combinations) in order to target them to the proteasome (Fig. 1A-B, full list of constructs in Table 1). To implement a level conditional control on viral protein stability, the **Tobacco Etch Virus** cleavage site (TEVs) was interposed between the viral proteins and the proteasome-targeting domain. The **Tobacco Etch Virus** protease (TEVp) selectively cleaves the TEVs linker, separating the viral proteins from the proteasome-targeting domain, sparing them from degradation (Fig 1A). The binary system composed by TEVp and TEVs can be pharmacologically regulated to modulate on the extent and temporal window of viral protein degradation during viral production and *in vivo*. Namely, the virus is able to transcribe and replicate only when TEVp is present, giving origin to a system in which viral transcription and replication are constitutively OFF unless TEVp is provided.

We screened the suitability to viral production and TEV dependency for all generated viral constructs (Fig. 1). The result of the screen points to the N protein as the sole viral protein whose conditional destabilization is sufficient to reversibly suppress the viral transcription-replication cycle (Fig 1F, K) in a TEVp dependent manner. The destabilization of all other viral proteins, alone or in combination, gives origin to viruses that are either unable to amplify *in vitro* or that amplify both in presence or absence of TEVp (Fig 1). After a further round of improvement on the viral cassette design (Supplementary Text, Fig 6, 7) we were able to produce

a **Self-inactivating Rabies** (SiR) based on an N-protein destabilised Rabies cassette with the desired TEV-dependent ON-OFF kinetics.

5 We then generated an SiR encoding for CRE recombinase and mCherry^{PEST} (SiR^{CRE-mCherry}, Fig. 2B). We tested SiR transcription-replication kinetics and cytotoxicity *in vivo* by injecting SiR^{CRE-mCherry} in the CA1 pyramidal layer of *Rosa-LoxP-STOP-LoxP-YFP* mice (Fig. 2B). Transient expression of the CRE
10 recombinase driven by the SiR should ensure a permanent recombination of the Rosa locus and YFP expression even after a complete transcriptional shut down of the virus (Fig. 2A). Indeed, while 3 days post-infection (p.i.) only the virally encoded mCherry can be detected (Fig 8 C-C''), already by 6
15 days post-infection, the viral encoded CRE recombinase induces recombination of the conditional mouse reporter cassette and expression of YFP in all infected neurons (Fig 8 D-D''). The destabilised mCherry^{PEST} marks the presence of active virus with high temporal resolution (half life ~2hrs) (21).

20 We then assessed the survival of SiR^{CRE-mCherry} infected CA1 pyramidal neurons (by following YFP^{ON} neurons over a 8-weeks period) and monitor the switching OFF of mCherry in infected neurons as a proxy of the viral transcription-replication
25 cycle. One week post-infection SiR^{CRE-mCherry} begins to switch OFF in a fraction of infected neurons (29±3% mCherry^{OFF} YFP^{ON}, Fig. 2C-C''', F). By three weeks post-infection the transition is complete as virtually all YFP^{ON} neurons show no expression of the virally encoded mCherry (98±2% YFP^{ON} mCherry^{OFF}, Fig. 2E-E''',
30 F, N=3). More importantly, during the 8 weeks observation period we detected no significant neuronal loss following SiR infection (no significant decrease of YFP^{ON} neurons over the 8 weeks period, one-way ANOVA, $F = 0.19$, $P = 0.90$, $N=3$ per time-point, Fig. 2E-E'', F). On the contrary, upon ΔG-Rabies
35 infection, the majority of infected neurons in the hippocampus die within 2 weeks from the primary infection (hippocampus, $92 \pm 3\%$ cell death at 2 weeks, $n = 3$ per time-point, one-way

ANOVA, $F = 101$, $P = 2.4 \times 10^{-5}$, Fig. 92 $\pm$ 3%). In order to gain better temporal resolution of the viral transcription-replication cycle, without the caveat of the mCherry^{PEST} protein half life, we measured viral RNA titers by real time PCR on brain extracts of infected animals at 1,2,3 and 8 weeks post-infection. In agreement with the mCherry expression results, viral titer drops to near-background level by two weeks post-infection (Fig 2G). Overall, these results show that the SiR^{CRE-mCherry} transcription-replication kinetics provide enough time to generate an early CRE recombination event (Fig. 8D-D''') before the virus disappears (Fig 2G), which ensures permanent genetic access to the infected neurons without affecting neuronal survival (Fig. 2F).

Experiments *in vitro* indicate that modulation of viral stability by conditional proteasome degradation is sufficient to modulate the viral transcription-replication cycle. In order to assess whether conditional control over viral protein degradation can be achieved *in vivo* we designed an AAV virus to express TEVp under a doxycycline inducible promoter (AAV^{TRE::TEVp}, Fig. 7A) (22). This provides a means of pharmacological control on TEVp expression *in vivo* and therefore, on viral transcription-replication cycle.

To address if and until what stage after the infection SiR can be reactivated by doxycycline administration, we injected CA1 neurons in the hippocampus with AAV^{TRE::TEVp} followed, 1 week after, by the SiR^{CRE-mCherry} infection. Doxycycline (100 mg/Kg) was administered by gavage for 2 days at 2 time points, after a week or after 2 weeks (Fig. 9B). In agreement with the RNA analysis, the administration of doxy when the virus is transcriptionally active at 1 week post SiR infection, doubled the percentage of YFP^{ON} mCherry^{ON} neurons (mCherry^{ON}YFP^{ON} -doxy 22 $\pm$ 3%, +doxy 38 $\pm$ 2, $P = 0.02$, Fig. 9G) while no effect was detected by administering doxy 2 weeks post infection (mCherry^{ON}YFP^{ON} -doxy 1 $\pm$ 1%, +doxy 1 $\pm$ 1%, Fig. 9G).

Given the shorter life cycle of the SiR^{CRE-mCherry} we asked whether it retains the ability to spread transsynaptically from the primary infected neurons before the virus switches off. In order to test this, we first injected an AAV
 5 expressing TVA and B19-G in the pyramidal layer of CA1 of *Rosa-LoxP-STOP-LoxP-YFP* mice followed by infection with an EnvA pseudo-typed SiR virus. As expected from the known anatomical connectivity between CA1 and CA3, we identified neurons labeled by the SiR in the pyramidal layer of CA3 (Fig.
 10 3A-A''), indicating specific transsynaptic spreading. Presynaptic neurons were also identified in the Entorhinal cortex (Fig. 3A'''). More importantly, SiR-infected starting cells expressing TVA and G remained viable throughout the infection period (Fig. 3B-B''). Furthermore, when pseudotyped
 15 with the recently developed optimized Rabies Glycoprotein (oG) (17) SiR also functions as a highly effective retrograde (non-transsynaptic) tracer (fig. 3C,D).

20 In order to confirm the absence of any long-term effect of the SiR on neuronal physiology, we injected SiR^{CRE-mCherry} in the pyramidal layer of CA1 in *Rosa-LoxP-STOP-LoxP-ChR2YFP*. We then compared the electrophysiological properties of the infected neurons one week and two months post-infection. YFP^{ON} neurons
 25 in CA1 were recorded in whole-cell patch-clamp mode in acute hippocampal slices. All pyramidal CA1 neurons recorded show regular spiking profiles (Fig. 4 A,B) with no significant difference in the input resistance ($226 \pm 13 \text{ M}\Omega$ at one week
 30 p.i., $n = 10$, versus $251 \pm 18 \text{ M}\Omega$ after two months, $n = 8$; Fig. 4C), resting membrane potential ($-57.0 \pm 1.9 \text{ mV}$ versus $-56.7 \pm 1.9 \text{ mV}$; Fig. 4D), action potential amplitude ($96.4 \pm 3.6 \text{ mV}$ versus $94.2 \pm 2.9 \text{ mV}$) and action potential half-width ($2.4 \pm 0.1 \text{ ms}$ versus $2.4 \pm 0.1 \text{ ms}$; Fig. 4E). Only a minor difference was found in the spike threshold ($-46.0 \pm 1.0 \text{ mV}$ versus $-41.3 \pm 2.0 \text{ mV}$, $P = 0.0394$; two-tailed two sample Student's *t*-test
 35 Fig. 4D); however, this difference does not affect the instantaneous firing frequency of the neurons (Fig. 4F). ChR2

is successfully expressed in neurons transduced with SiR^{CRE-mCherry} (Fig. 4G-I). Neurons can be activated at various frequencies with similar reliability both at one week and over two months post-infection (Fig. 4J). A key finding was that light-activation of SiR^{CRE-mCherry} infected neurons elicits DNQX-sensitive excitatory postsynaptic potentials (EPSPs) in their post-synaptic partners at two months after infection (Fig. 4K), indicating persistence of functional connectivity and no adverse effect on synaptic function.

Absence of cytotoxicity and unaltered electrophysiological responses support the use of SiR for long-term circuit manipulations. Presence of functional connectivity between SiR infected neurons and no adverse effect on synaptic function also indicate that network function is likely to be preserved upon SiR infection. In order to directly test whether network-dependent computations are indeed unaffected in SiR mapped circuits, we traced V1 neurons projecting to V2 with SiR and characterized their orientation tuning preferences, as a prototypical example of a network-dependent computation (18). We first targeted V2 projecting neurons by the injection of an oG pseudo-typed SiR^{CRE} in *Rosa-LoxP-STOP-LoxP-tdTomato* mice. At the same time, we injected an AAV::GCaMP6s in the ipsilateral V1. Retrograde spreading of SiR^{CRE} induces recombination of the Rosa locus permanently labeling V1 neurons projecting to V2 (V1^{>V2}) (Fig. 5B-B''). We then monitored the Ca²⁺ dynamics of SiR infected V1^{>V2} neurons *in vivo* 4 weeks p.i., under a two-photon microscope, while anesthetized animals were exposed to moving gratings of different orientations across the visual field (Fig. 5C) (19). Infected V1 neurons showed significant increase in fluorescence at particular grating orientation resulting in a tuning curve showing direction or orientation selectivity (Fig. 5G, H). Notably, recorded Ca²⁺ responses as well as the percentage of active neurons, were similar between SiR-traced neuron (GCaMP6s^{ON}-tdTomato^{ON}) and neighboring non-SiR V1 neurons (GCaMP6s^{ON}-tdTomato^{OFF}) (Fig. 5E, F). These data indicate that SiR traced networks preserve unaltered

computational properties and that SiR can be used in combination with GCaMP6s to monitor the Ca^{2+} dynamic with no upper bounds to the temporal window for the optical investigation.

5

The development of monosynaptically restricted Rabies viruses has had a transformative role in the study of neural circuits. However, until now, the cellular cytotoxicity that accompanies Rabies virus infection effectively limited its use, by and

10

large, to the anatomical mapping of neural circuits. The induced cytotoxicity is linked to the transcriptional activity of the virus, which hijacks of the cellular transcriptional machinery to sustain viral replication (24). Therefore, any

15

replicative competent Rabies virus will eventually compromise cellular physiology. To overcome this limitation and gain life-long genetic and functional access to topologically defined network elements we developed a **Self-inactivating**

20

Rabies virus, which transcriptionally switches off following the primary infection in a TEVp-dependent manner, both in human Embryonic Stem cells (hESCs) derived neurons

25

(Supplementary results, Fig. 6) and *in vivo* in mice. In line with the complete transcriptional silencing of the virus we observe no changes in the electrophysiological signature of the infected neurons months after infection and maintenance of synaptic function and circuit integrity, which is key for

30

physiological and behavioral studies. More importantly, we also show that higher level circuit-dependent computations of infected neurons, such as their orientation tuning to moving stimuli, remain unaffected *in vivo* for months after viral infection.

35

This, in turn, also shows that SiR can be used to monitor network activity *in vivo* using calcium imaging. With the canonical B19 ΔG -Rabies the optimal temporal window for optical imaging of neurons is, typically, 5-7 days from the Rabies infection (17). The use of the recently introduced ΔG -Rabies strain variant CVS-N2c ^{ΔG} can push the useful temporal

window for imaging further up to 17 days post-infection (25).
 With SiR there are no upper bounds to the temporal window for
 the optical investigation of network elements. These
 attributes make SiR the most valuable solution for the long-
 term monitoring of neural network activity as well as for the
 functional and genetic manipulation of neural circuits (26-
 29). In addition, the unique transient replicative nature of
 SiR, offers the possibility to follow circuit remodelling
 after physiological or pathological structural plasticity such
 as, upon learning, during neurodegenerative conditions or
 following traumatic brain injuries and it may pave the way to
 functional interventions at the network level in such cases.

Overall, the development of Self Inactivating Δ G-Rabies
 provides, for the first time, permanent genetic access to
 topologically defined network elements without adverse effects
 to neuronal physiology and circuit function.

Supplementary Text

First generation Δ G-N^{PEST}Rabies. *In vitro* and *in vivo* test of cytotoxicity

In order to obtain conditional regulation of viral protein
 stability a SPLIT-TEV cassette (30) was added at the C-
 terminal of each viral protein (Δ G-VP^{PEST}Rabies^{SPLIT-TEV-mCherry}). In
 addition a tag (myc, FLAG or V5) was fused to the N-terminal
 of each viral protein to monitor levels of protein expression.
 The SPLIT-TEV dimeric protease is only active in presence of
 rapamycin, and could potentially provide a tool for the
 exogenous regulation of viral protein stability during
 production and *in vivo*. We first tested the capability of the
 SPLIT-TEV expressed by plasmid to cleave a TEV activity
 reporter in HEK cells (Fig. 6B) then we tested the ability of
 the virally expressed cassette to cleave the TEV activity
 reporter (Fig. 6C).

In order to probe the effect of protein destabilization on neuronal survival, we infected human Embryonic Stem cells (hESCs) derived neurons with $\Delta G-N^{\text{PEST}}\text{Rabies}^{\text{SPLIT-TEV-mCherry}}$. We performed a longitudinal study of the survival of infected neurons and compared survival rate to a control ΔG Rabies. Neurons were infected and imaged longitudinally at 4-10-16 days to evaluate the cell death (Fig. 7 A, B-C"). Lentivirus expressing GFP was used to normalize infection rates in order to account for cell death due to the prolonged manipulation and repeated over-night imaging sessions. Only $26 \pm 4\%$ of control ΔG -Rabies infected neurons were still detectable at 16 days post-infection ($N = 3$, $n = 781$ for each condition; Fig. 7C-C'', D). On the contrary, the $\Delta G-N^{\text{PEST}}\text{Rabies}^{\text{SPLIT-TEV-mCherry}}$ virus showed no significant cell loss after 16 days $94 \pm 6\%$ ($N = 3$, $n = 917$ for each condition, Fig. 4B-B'', C) and a significant increase in cell survival compared to ΔG -Rabies controls ($P = 3.2 \times 10^{-5}$; paired two-tailed Student's t -test).

To understand if the reduced cytotoxicity of $\Delta G-N^{\text{PEST}}\text{Rabies}^{\text{SPLIT-TEV-mCherry}}$ is associated with a reduction of the viral transcription, we monitored the intensity of the reporter expressed in neurons infected with the control ΔG -Rabies and the $\Delta G-N^{\text{PEST}}\text{Rabies}^{\text{SPLIT-TEV-mCherry}}$ viruses, which share the same expression cassette (Fig. 7E). Over time the mean mCherry signal of $\Delta G-N^{\text{PEST}}\text{Rabies}^{\text{SPLIT-TEV-mCherry}}$ infected cells resulted to be significantly lower than controls (mCherry intensity at 10 days, ΔG -Rabies $138 \pm 3\%$, $\Delta G-N^{\text{PEST}}\text{Rabies}^{\text{SPLIT-TEV-mCherry}}$ $87 \pm 7\%$, $P = 0.01$; $N = 3$, $n = 584$ for each condition; paired two-tailed Student's t -test).

We then tested the performance of the $\Delta G-N^{\text{PEST}}\text{Rabies}^{\text{SPLIT-TEV-mCherry}}$ virus *in vivo*.

We replaced the mCherry gene in the $\Delta G-N^{\text{PEST}}\text{Rabies}^{\text{SPLIT-TEV-mCherry}}$ virus with the CRE recombinase ($\Delta G-N^{\text{PEST}}\text{Rabies}^{\text{SPLIT-TEV-CRE}}$). This ensures that infected neurons can be permanently labeled after

a complete transcriptional shut down of the virus, allowing to discriminate between viral silencing and cell death. We injected $\Delta G-N^{\text{PEST}}\text{Rabies}^{\text{SPLIT-TEV-CRE}}$ in CA1 of *Rosa-LoxP-STOP-LoxP-tdtomato* reporter mouse line in CA1 in the hippocampus. We observed a significant increase in neuronal survival upon $\Delta G-N^{\text{PEST}}\text{Rabies}^{\text{SPLIT-TEV-CRE}}$ infection compared to that observed upon infection with control ΔG Rabies ($25 \pm 2\%$, at 2 weeks for $\Delta G-N^{\text{PEST}}\text{Rabies}^{\text{SPLIT-TEV-CRE}}$ and $8 \pm 3\%$, at 2 weeks for ΔG Rabies $P = 7 \times 10^{-3}$, Fig. 7H). However prominent neuronal loss was still present upon $\Delta G-N^{\text{PEST}}\text{Rabies}^{\text{SPLIT-TEV-CRE}}$ infection ($76 \pm 3\%$, at 3 weeks, $P = 9 \times 10^{-4}$)

The residual cytotoxicity of $\Delta G-N^{\text{PEST}}\text{Rabies}^{\text{SPLIT-TEV-CRE}}$ might be linked to a constitutive low basal dimerization and activity of the SPLIT-TEV cassette and can give origin to transcriptionally active viral particles. Consistently with this hypothesis, we observed no significant effect on neuronal survival and mCherry expression levels in presence or absence of rapamycin (mCherry expression 10 days p.i. $\Delta G-N^{\text{PEST}}\text{Rabies}^{\text{SPLIT-TEV-RAP}}$ $87 \pm 7\%$, $\Delta G-N^{\text{PEST}}\text{Rabies}^{\text{SPLIT-TEV+RAP}}$ $85 \pm 6\%$, $P = 0.21$; $N = 3$, $n = 793$ for each condition; paired two-tailed Student's *t*-test; Fig. 7E) and no significant effects on cells survival were associated with the rapamycin administration in hESCs derived neurons (at 16 days; $\Delta G\text{-Rabies}^{\text{+RAP}}$ $26 \pm 4\%$, $\Delta G\text{-Rabies}^{\text{-RAP}}$ $30 \pm 7\%$, $P = 0.69$; $\Delta G-N^{\text{PEST}}\text{Rabies}^{\text{SPLIT-TEV+RAP}}$ $88 \pm 11\%$, $\Delta G-N^{\text{PEST}}\text{Rabies}^{\text{SPLIT-TEV-RAP}}$ $94 \pm 6\%$, $P = 0.79$; $N = 3$, $n = 833$ for each condition; paired two-tailed Student's *t*-test, Fig. 7D). Furthermore, we observe a constitutive low level of TEV activity in HEK cells in absence of Rapamycin (Fig. 6B, line2) indicating a basal level of Rapamycin-independent SPLIT-TEV dimerization. Overall these results indicate that $\Delta G-N^{\text{PEST}}\text{Rabies}^{\text{SPLIT-TEV}}$ has reduced cytotoxicity in hESCs derived neurons and *in vivo* when compared to $\Delta G\text{-Rabies}$. However, it fails to completely switch off following the infection, which leads to significantly delayed, yet still present, neuronal cytotoxicity and neuronal loss. For these reasons we generated a second generation of $\Delta G-N^{\text{PEST}}\text{Rabies}$ by removing the leaking

SPLIT-TEV and replacing it with an mCherry-CRE cassette giving origin to a **S**elf **i**nactivating **R**abies virus (SiR) with the desired ON-OFF and TEV dependent kinetics (main text).

5 Materials & Methods

Table 1.

Full list of constructs.

Rabies	N-terTAGs	C-terPEST
pSAD-F3-mCherry	-	-
pSAD-F3-N-terTAGs-mCherry	N ^{Myc} -pFLAG-M ^{V5}	-
pSAD-F3-N-terTAGs-N ^{PEST} -mCherry	N ^{Myc} -pFLAG-M ^{V5}	N ^{PEST}
pSAD-F3-N-terTAGs-M ^{PEST} -mCherry	N ^{Myc} -pFLAG-M ^{V5}	M ^{PEST}
pSAD-F3-N-terTAGs-P ^{PEST} -mCherry	N ^{Myc} -pFLAG-M ^{V5}	P ^{PEST}
pSAD-F3-N-terTAGs-L ^{PEST} -mCherry	N ^{Myc} -pFLAG-M ^{V5}	L ^{PEST}
pSAD-F3-N ^{PEST} -mCherry	-	N ^{PEST}
pSAD-F3-N-terTAGs-(P+L) ^{PEST} -mCherry	N ^{Myc} -pFLAG-M ^{V5}	P ^{PEST} L ^{PEST}
pSAD-F3-(P+L+N) ^{PEST} -mCherry	-	P ^{PEST} L ^{PEST} N ^{PEST}
pSAD-F3-N ^{PEST} -iCRE-2A-mCherryPEST	-	N ^{PEST}
pSAD-F3-N ^{PEST} -cTEV-FKBP-2A-FRB-NTEV-iCRE	-	N ^{PEST}
Lentiviruses	Gene A	Gene B
pLenti-H2BGFP-2A-GlySAD	H2BGFP	GlySAD
pLenti-puro-2A-TEV	puromycin	TEV
pLenti-GFP	GFP	-
pLenti-H2BGFP-2A-oG	H2BGFP	oG
AAVs	Gene A	Gene B
AAV-CMV-TVAmCherry-2A-Gly	TVAmCherry	Gly
AAV-TRE _{tight} -TEV-CMV-rTTA	TEV (doxy dependent)	rTTA
AAV-CAG-GCaMP6s	GCaMP6s	-

10 Animal strains

C57BL/6 wild type (WT) mice and the following transgenic lines were used: *Rosa-LoxP-STOP-LoxP-tdtomato* (Jackson: Gt(ROSA)26Sortm14 (CAG tdTomato, *Rosa-LoxP-STOP-LoxP-YFP* (Jackson: Gt(ROSA)26Sor<tm1(EYFP)Cos>)). All procedures were conducted in accordance with the UK Animals (Scientific procedures) Act 1986 and European Community Council Directive on Animal Care. Animals were housed in a 12 hours light/dark cycle with food and water ad libitum.

Design and generation of Δ G Rabies and lentivirus plasmids

All the attenuated Rabies plasmids, listed in Table 1, were generated by Gibson cloning using the pSAD- Δ G-F3 plasmid (21) as starting material. Briefly, the Rabies genome was PCR amplified in 2 fragments starting from the protein to be tagged. These fragments were then mixed with the tag and/or PEST domain obtained by oligonucleotides annealing and assembled using Gibson master mix (NEB).

The lentiviral vectors used to generate the packaging cells were derived from a 3rd generation lentivirus transfer vector (gift from Michael Hastings "361 polilinker", originally pCCL-SIN-18PPT.Pgk.EGFP-WPRE). All the lentiviral vectors were generated by Gibson assembly, opening the backbone by digestion with XbaI and KpnI and PCR amplifying the CMV promoter and the different inserts.

Cell lines

HEK293T cells and BHK were purchased from ATTC. HEK293T packaging cells expressing Rabies glycoprotein (HEK-GG) were generated by lentivirus infection with Lenti-H₂BGFP-2A-GlySAD (Table 1) and after 3 passages by fluorescent activated cell sorting (FACS) of GFP expressing cells. HEK293T packaging cells expressing Rabies glycoprotein and TEV protease (HEK-TGG) were generated from HEK-GG by lentivirus infection with Lenti-puro-2A-TEV and selected, after 3 passages, with 1 μ g/mL of puromycin added to the media for 1 week.

BHK packaging cells for pseudotyping Rabies virus with optimized G (BHK-TGoG) were generated with the same procedure as the HEK-TGG infecting first with pLenti-H₂BGFP-2A-oG and subsequently with pLenti-puro-2A-TEV. BHK packaging cells for pseudotyping Rabies virus with EnVA receptor (BHK-T-EnVA) were

obtained infecting BHK-EnVA with Lenti-puro-2A-TEV and selecting with puromycin.

Viral Screening

5 For screening of attenuated Δ G-Rabies viruses, HEK-GG or HEK-TGG cells were co-transfected with rabies genome vector, pcDNA-T7, pcDNA-SADB19N, pcDNA-SADB19P, pcDNA-SADB19L, and pcDNA-SADB19G (21) and maintained at 37 °C with 5% CO₂ humidified atmosphere in DMEM supplemented with 10% FBS
10 (Gibco) and 100 u/ml Penicillin-Streptomycin. The day after transfection and subsequently every 3 days, cells were washed with PBS, treated with 0.05% trypsin and replated in a new dish in a ratio 1:3. After splitting, cells were maintained for one day at 37°C and 5% CO₂ and then 2 days at 35°C and 3%
15 CO₂. Every 3 days cells were fixed and viral spreading was assessed by FACS sorting the cells for mCherry expression.

Viral productions

20 For the recovery of high titer Δ G-Rabies HEK-GG or HEK-TGG, for control or attenuated Rabies respectively, were infected in 10 cm dishes at 70-80% confluence with 1 ml of viral supernatant obtained as described in the viral screening section. Cells were split the day after infection and maintained for 1 or 2 days at 37°C and 5% CO₂ checking daily
25 the viral spreading. When 70-80% of cells expressed the viral marker, the media was replaced with 2% FBS DMEM and maintained for 2 days at 35 °C and 3% CO₂. Then, the viral supernatant was collected, cell debris removed by centrifugation at 2500 rpm for 10 minutes followed by filtration with 0.45 μ m filter and
30 then the virus concentrated by ultracentrifugation with sucrose cushion as described before (22).

Rabies viruses pseudotyped with oG were produced infecting BHK-T-oG cells in 10 cm dishes with 1 mL of viral supernatant.

Cells were split the day after infection and maintained for one or two days at 37°C and 5% checking daily the viral spreading. When 70-80% of cells expressed the viral marker, the media was replaced with 2% FBS DMEM and maintained for 2-3 days at 35°C and 3% CO₂. Then, the supernatant was collected and processed as previously described (20).

Rabies viruses pseudotyped with EnVA were produced as previously described (22) using BHK-T-EnVA cells instead of BHK-EnVA cells.

In vitro cytotoxicity analysis

Human Embryonic Stem cells (hESCs) derived neurons were kindly provided by Dr. Rick Livesey. Cells were plated in 24-wells glass bottom plates and infected over night with attenuated or control ΔG-Rabies viruses supernatants at comparable MOI to obtain ~5% of infected cells. Cells were imaged every 4 days post infection overnight in a 37 °C heated Leica SP8 confocal microscope in Hibernate®-A Medium (Invitrogen) with 5 random fields imaged for each well. Cell survival was calculated normalizing each condition to the mortality of control Lentivirus-GFP infected hESCs derived neurons imaged and processed in the same conditions.

Viral injections

All procedures using live animals were approved by the Home Office and the LMB Biosafety committee. For all experiments mice aged between 6-12 weeks were used. Mice were anesthetized with isofluorane delivered at a flow of 3% in 2L/min of oxygen for the initial induction and then maintained at 1-2% in 2 L/min of oxygen. The anesthetized animal was placed into a stereotaxic apparatus (David Kopf Instruments) and Rimadyl (2 mg/kg body weight) was administered subcutaneously (s.c.) as anti-inflammatory. A small hole (500 μm diameter) was drilled and viruses were injected using a Hamilton neurosyringe. The

syringe was left in the brain for 5 min before being retracted. Viruses were injected at the following titers: 3×10^8 infectious units/ml for Rabies viruses, 2×10^{12} genomic copies/ml for AAVs, 3×10^8 infectious units/ml for Lentiviruses.

Up to a max of 400 nl in volume of virus were injected in the following brain areas: CA1 (AP: 2.3 mm, ML: 1.65 mm and DV: 1.45mm from bregma), V1 (AP: 2.3 mm, ML: 1.65 mm and DV: 1.45 mm from bregma), V2 (AP: 3.6 mm, ML: 1.2 mm, DV: 0.6 mm from bregma).

In vivo cytotoxicity analysis

To test *in vivo* viral cytotoxicity 400 nl of same titer (3×10^8 infectious units/ml) attenuated and control ΔG-Rabies were injected in CA1 of hippocampus contralateral in the 2 hemispheres. At 1-2-3 or 8 weeks p.i. brains were sectioned at the cryostat (35 μm). Infected neurons were imaged sampling the entire hippocampus (acquiring one every 4 sections) using a robot assisted Nikon HCA microscope mounting a 10x (0.45NA) air objective and fluorescent hippocampal neurons counted using Nikon HCA analysis software. Cell survival for attenuated and control ΔG-Rabies was calculated normalizing each time point to the mortality of control Lentivirus-GFP infected hippocampi using the same injection protocol.

Drug induced reactivation of SiR virus *in vivo*

Rosa-LoxP-STOP-LoxP-YFP animals were injected in CA1 of hippocampus with an AAV constitutively expressing rTTA and TEV protease under the control of a doxycycline inducible promoter (Table 1). 1-week p.i. the same area was re-injected with SiR^{mCherry-CRE} and doxycycline (Santa Cruz Biotechnology, 100 mg/Kg) administered at 1 or 2 weeks post SiR injection. 1 week after drug administration brains were collected and sectioned at the cryostat (35 μm). Infected neurons were imaged and

counted sampling the entire hippocampus (acquiring one every 4 sections) using a robot assisted Nikon HCA microscope.

Analysis of SiR genomic copies *in vivo*

5 To evaluate the genomic copies of SiR virus in the infected animals over time SiR^{mCherry-CRE} was injected in CA1 region of hippocampus of *Rosa-LoxP-STOP-LoxP-YFP* animals. After 1, 2, 3 or 8 weeks, mice were culled and the injected hippocampi manually dissected immediately after. The hippocampi were
10 homogenised using a TissueLyser II (Qiagen) and processed accordingly to manufactory instruction with RNeasy kit (Qiagen) to extract total RNA. 500 ng of RNA per hippocampus were retrotranscribed using superscript IV kit (Invitrogen) and analysed for GADPH, YFP and mCherry expression by quantitative
15 PCR (rotorgene sybr-green). The Livak method was applied for quantification. The expression of YFP and mCherry was normalized to the expression of the GADPH housekeeping gene ($DCT = CT_{\text{gene}} - CT_{\text{GADPH}}$) and the variation over time as fold change (2^{-DDCT}) to the 1 week time point ($DDCT = DCT_{\text{Time point}} - DCT_1$
20 week).

Electrophysiology

For electrophysiological recordings, SiR^{mCherry-CRE} was injected bilaterally in the CA1 of one month-old *Rosa-LoxP-STOP-LoxP-ChR2YFP* mice. Recordings were made either one week or between
25 two and 3 months p.i..

Coronal hippocampal slices (350 μm) were prepared using a vibrating microtome (7000smz-2, Campden Instruments LTD, Loughborough, UK) in ice-cold sucrose-based cutting solution
30 oxygenated with carbogen gas (95% O₂, 5% CO₂) and with the following composition (in mM): KCl 3, NaH₂PO₄ 1.25, MgSO₄ 2, MgCl₂ 1, CaCl₂ 1, NaHCO₃ 26.4, glucose 10, sucrose 206, ascorbic acid 0.40, kynurenic acid 1. Slices were incubated at 37 °C

for 30 minutes in a submerged-style holding chamber with oxygenated artificial cerebrospinal fluid (aCSF; in mM: NaCl 126, KCl 3, NaH₂PO₄ 1.25, MgSO₄ 2, CaCl₂ 2, NaHCO₃ 26.4, glucose 10) with an osmolarity adjusted to 280-300 mOsm/L and stored thereafter in the same holding chamber at room temperature for at least 1 h. Slices were then individually transferred to the recording chamber and were superfused with oxygenated aCSF at room temperature at a flow-rate of approximately 2 mL/min.

Whole-cell current-clamp recordings were obtained from CA1 neurons using 6-9 MΩ pipettes pulled from borosilicate glass capillaries (1.5 mm OD x 0.86 mm ID). Pipettes were filled with artificial intracellular solution containing (in mM): K-gluconate 150, HEPES 10, NaCl 4, ATP-Mg 4, GTP-Na 0.3 and EGTA 0.2; adjusted to pH 7.2 and osmolarity 270-290 mOsm/L. Data were recorded using an Axon Multiclamp 700B amplifier (Molecular Devices, Union City, CA, USA) and signals were low-pass filtered at 2 kHz and acquired at 5 kHz using a digitizer (Axon Digidata 1550A, Molecular Devices, Union City, CA, USA) on a PC running pClamp. Light-evoked responses from neurons infected with SiR virus were elicited using a 450-490 nm LED light (pE-300 coolLED system, Scientifica Ltd, Uckfield, UK) through a 40X water immersion objective (0.8 NA).

Pharmacology

The AMPA receptor antagonist DNQX (20 μM; Sigma-Aldrich, Dorset, UK) was used in a subset of electrophysiological recordings in order to probe synaptic connectivity between neurons infected with SiR virus and neighbouring neurons.

In vivo 2-photon imaging

Injected *Rosa-LoxP-STOP-LoxP-tdTomato* mice (see *Viral injections* section) were anaesthetized with isofluorane 2%. Animal pinch withdrawal and eyelid reflex were tested to assay

the depth of anaesthesia. Rimadyl (2 mg/kg body weight) was injected subcutaneously as an anti-inflammatory. Both eyes were covered with an eye ointment to prevent corneal desiccation during the experiment. The animal was head-fixed and a metal head-post cemented to the skull. A craniotomy of 4 mm in diameter was drilled over the V1 cortex. After the removal of the skull, the cortical surface was kept moist with a cortex buffer, containing: 125 mM NaCl, 5 mM KCl, 10 mM Glucose, 10 mM HEPES, 2 mM MgSO₄ and 2 mM CaCl₂, adjusted to pH 7.3. The cortex was then covered with a custom made plug coverslip (23) and sealed with Super Glue and dental cement. Mice were anaesthetized with 2% of isofluorane and mounted under a two-photon laser-scanning microscope (Multiphoton Imaging System, Scientifica Ltd., Uckfield, United Kingdom) equipped with a Ti:sapphire mode-locked laser (Mai Tai-Series, Spectra Physics) tuned at 920 nm. Imaging was performed through a water-immersion lens (Nikon, 16X, 0.8 NA) at a resolution of 256 x 256 pixels with zoom 2 or 4, leading to a field of view of 390 x 390 µm and 195 x 195 µm respectively. Data were acquired at 3.5 Hz. The objective was shielded with a black fabric cone equipped with a plastic o-ring fixed onto the head plate (24). Visual stimulation was controlled using a custom-made GUI in Python (based on PsychoPy toolbox) and was performed with a LED screen positioned 15 cm from the left eye of the mouse. Moving square-wave gratings were presented at 12 directions in 30 degrees steps and a photodiode was used to detect the starting and the ending time of each stimulus. Each grating direction was presented 5 times in random order alternated with a blank condition. The spatial frequency of the grating was 0.04 cycles per degree (cpd) and the temporal frequency was 1 Hz. Imaging and visual stimulation were triggered together using Arduino micro-controller board. Imaging session lasted up to 2 hrs and the power at sample was controlled in the range 30-40 mW. Data analysis was performed in ImageJ and Matlab and was restricted to cell bodies. Detection of region of interest (ROI) was performed with Suite2p. The relative changes in fluorescence were calculated

as $dF/F_0 = (F(t) - F_0) / (F_0)$. Orientation tuning curves were generated by taking the mean response for each orientation during the entire stimulus period. Response amplitudes are presented as the relative change in fluorescence during the stimulus period compared to the pre-stimulus baseline (dF/F). All data are presented as mean $\pm$ SEM.

While the invention has been described in conjunction with the exemplary embodiments described above, many equivalent modifications and variations will be apparent to those skilled in the art when given this disclosure. Accordingly, the exemplary embodiments of the invention set forth are considered to be illustrative and not limiting. Various changes to the described embodiments may be made without departing from the spirit and scope of the invention. All documents cited herein are expressly incorporated by reference.

The teaching of all references in the present application, including patent applications and granted patents, are herein fully incorporated by reference. Any patent application to which this application claims priority is incorporated by reference herein in its entirety in the manner described herein for publications and references.

For the avoidance of doubt the terms 'comprising', 'comprise' and 'comprises' herein is intended by the inventors to be optionally substitutable with the terms 'consisting of', 'consist of', and 'consists of', respectively, in every instance. The term "about" (or "around") in all numerical values allows for a 5% variation, i.e. a value of about 1.25% would mean from between 1.19%-1.31%.

It will be understood that particular embodiments described herein are shown by way of illustration and not as limitations

of the invention. The principal features of this invention can be employed in various embodiments without departing from the scope of the invention. Those skilled in the art will recognize, or be able to ascertain using no more than routine study, numerous equivalents to the specific procedures described herein. Such equivalents are considered to be within the scope of this invention and are covered by the claims. All publications and patent applications mentioned in the specification are indicative of the level of skill of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

The use of the word "a" or "an" when used in conjunction with the term "comprising" in the claims and/or the specification may mean "one," but it is also consistent with the meaning of "one or more," "at least one," and "one or more than one." The use of the term "or" in the claims is used to mean "and/or" unless explicitly indicated to refer to alternatives only or the alternatives are mutually exclusive, although the disclosure supports a definition that refers to only alternatives and "and/or." Throughout this application, the term "about" is used to indicate that a value includes the inherent variation of error for the measurement, the method being employed to determine the value, or the variation that exists among the study subjects.

As used in this specification and claim(s), the words "comprising" (and any form of comprising, such as "comprise" and "comprises"), "having" (and any form of having, such as "have" and "has"), "including" (and any form of including, such as "includes" and "include") or "containing" (and any form of containing, such as "contains" and "contain") are inclusive or open-ended and do not exclude additional, unrecited elements or method steps.

The term "or combinations thereof" as used herein refers to all permutations and combinations of the listed items preceding the term. For example, "A, B, C, or combinations thereof is intended to include at least one of: A, B, C, AB, AC, BC, or ABC, and if order is important in a particular context, also BA, CA, CB, CBA, BCA, ACB, BAC, or CAB. Continuing with this example, expressly included are combinations that contain repeats of one or more item or term, such as BB, AAA, BBC, AAABCCCC, CBBAAA, CABABB, and so forth. The skilled artisan will understand that typically there is no limit on the number of items or terms in any combination, unless otherwise apparent from the context.

All of the compositions and/or methods disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the compositions and/or methods and in the steps or in the sequence of steps of the method described herein without departing from the concept, spirit and scope of the invention. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.

References

1. A. R. McIntosh, Towards a network theory of cognition. *Neural networks : the official journal of the International Neural Network Society* **13**, 861 (Oct-Nov, 2000).
- 5 2. G. Buzsaki, W. Freeman, Editorial overview: brain rhythms and dynamic coordination. *Curr Opin Neurobiol* **31**, v (Apr, 2015).
3. V. Mnih et al., Human-level control through deep reinforcement learning. *Nature* **518**, 529 (Feb 26, 2015).
- 10 4. D. Silver et al., Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484 (Jan 28, 2016).
5. M. Rubinov, E. Bullmore, Fledgling pathoconnectomics of psychiatric disorders. *Trends in cognitive sciences* **17**, 641
15 (Dec, 2013).
6. F. Roselli, P. Caroni, A circuit mechanism for neurodegeneration. *Cell* **151**, 250 (Oct 12, 2012).
7. E. M. Callaway, Transneuronal circuit tracing with neurotropic viruses. *Current Opinion in Neurobiology* **18**, 617
20 (Dec 01, 2008).
8. M. P. van den Heuvel, E. T. Bullmore, O. Sporns, Comparative Connectomics. *Trends in cognitive sciences* **20**, 345
(May, 2016).
9. A. Cardona et al., An integrated micro- and
25 macroarchitectural analysis of the Drosophila brain by computer-assisted serial section electron microscopy. *PLoS Biol* **8**, (2010).
10. J. Livet et al., Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system.
30 *Nature* **450**, 56 (Nov 01, 2007).

11. P. Osten, T. W. Margrie, Mapping brain circuitry with a light microscope. *Nat Methods* **10**, 515 (Jun, 2013).
12. T. Mebatsion, M. Konig, K. K. Conzelmann, Budding of rabies virus particles in the absence of the spike glycoprotein. *Cell* **84**, 941 (Apr 22, 1996).
13. I. R. Wickersham et al., Monosynaptic restriction of transsynaptic tracing from single, genetically targeted neurons. *Neuron* **53**, 639 (Apr 01, 2007).
14. M. Tripodi, A. E. Stepien, S. Arber, Motor antagonism exposed by spatial segregation and timing of neurogenesis. *Nature*, 1 (Oct 19, 2011).
15. A. E. Stepien, M. Tripodi, S. Arber, Monosynaptic rabies virus reveals premotor network organization and synaptic specificity of cholinergic partition cells. *Neuron* **68**, 456 (Nov 04, 2010).
16. K. Miyamichi et al., Cortical representations of olfactory input by trans-synaptic tracing. *Nature* **472**, 191 (Dec 22, 2010).
17. A. Wertz et al., PRESYNAPTIC NETWORKS. Single-cell-initiated monosynaptic tracing reveals layer-specific cortical network modules. *Science* **349**, 70 (Jul 3, 2015).
18. P. Namburi et al., A circuit mechanism for differentiating positive and negative associations. *Nature* **520**, 675 (Apr 30, 2015).
19. J. Tian et al., Distributed and Mixed Information in Monosynaptic Inputs to Dopamine Neurons. *Neuron* **91**, 1374 (Sep 21, 2016).
20. D. Atasoy, Y. Aponte, H. H. Su, S. M. Sternson, A FLEX switch targets Channelrhodopsin-2 to multiple cell types for imaging and long-range circuit mapping. *Journal of Neuroscience* **28**, 7025 (Jul 09, 2008).

21. X. Li et al., Generation of Destabilized Green Fluorescent Protein as a Transcription Reporter *. **273**, 34970 (1998).
22. P. E. Shockett, D. G. Schatz, Diverse strategies for
5 tetracycline-regulated inducible gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 5173 (1996).
23. A. Benucci, D. L. Ringach, M. Carandini, Coding of
10 stimulus sequences by population responses in visual cortex. *Nat Neurosci* **12**, 1317 (Oct, 2009).
24. A. V. Komarova et al., Rabies virus matrix protein
interplay with eIF3, new insights into rabies virus
pathogenesis. *Nucleic Acids Res* **35**, 1522 (2007).
25. T. R. Reardon et al., Rabies Virus CVS-N2c Strain
15 Enhances Retrograde Synaptic Transfer and Neuronal Viability. *Neuron*, (Jan 19, 2016).
26. A. M. Packer, B. Roska, M. Hausser, Targeting neurons and
photons for optogenetics. *Nat Neurosci* **16**, 805 (Jul, 2013).
27. A. M. Packer, L. E. Russell, H. W. Dalglish, M. Hausser,
20 Simultaneous all-optical manipulation and recording of neural
circuit activity with cellular resolution in vivo. *Nat Methods*
12, 140 (Feb, 2015).
28. L. Luo, E. M. Callaway, K. Svoboda, Genetic dissection of
neural circuits. *Neuron* **57**, 634 (Apr 13, 2008).
29. V. Emiliani, A. E. Cohen, K. Deisseroth, M. Hausser, All-
25 Optical Interrogation of Neural Circuits. *J Neurosci* **35**, 13917
(Oct 14, 2015).
30. D. C. Gray, S. Mahrus, J. a. Wells, Activation of
specific apoptotic caspases with an engineered small-molecule-
30 activated protease. *Cell* **142**, 637 (2010).

5

31. S. Finke, J. H. Cox, K-K Conzelmann, Differential Transcription Attenuation of Rabies Virus Genes by Intergenic Regions: Generation of Recombinant Viruses Overexpressing the Polymerase Gene. J. Virol. 74(16), 7261 (2000)

32. Osakada and Callaway (2013) Design and generation of rabies virus vectors, Nature Protocols 8(8): 1583-1601.

Aspects of the invention

The following numbered paragraphs (paras.) contain statements of broad combinations of the inventive technical features herein disclosed:

5

1. A mononegaviral vector genome comprising a gene encoding a replication modulator protein, wherein the replication modulator protein comprises a mononegaviral protein moiety which is required for replication of the viral genome, the replication modulator protein being capable of adopting a targeted configuration displaying a degron, and an untargeted configuration which does not display the degron.

10

2. A vector genome according to para. 1 wherein the replication modulator protein encoded by the vector genome is an inhibitory modulator comprising a viral protein moiety and a regulator moiety, wherein the regulator moiety comprises or consists of the degron.

15

3. A vector genome according to para. 2 wherein the regulator moiety is switchable to an untargeted configuration on contact with a cognate activating agent.

20

4. A vector genome according to para. 3 wherein the activating agent cleaves the regulator moiety from the viral protein moiety.

25

5. A vector genome according to para. 4 wherein the activating agent is a protease.

30

6. A vector genome according to para. 5 wherein the regulator moiety comprises a cleavage site for the protease, located between the viral protein moiety and the degron.

7. A vector genome according to para. 5 or para. 6 wherein the protease does not act on any other proteins encoded by the vector genome.

35

8. A vector genome according to any one of paras. 5 to 7 wherein the protease is a viral protease, Factor Xa, enterokinase or thrombin.

5

9. A vector genome according to any one of paras. 2 to 8 wherein the degron is a PEST sequence.

10. A vector genome according to any one of paras. 2 to 9 wherein the modulator protein comprises a first N-terminal residue, and is cleavable by the cognate protease to expose a second N-terminal residue which confers greater stability than the first N-terminal residue.

10

11. A vector genome according to any one of paras. 3 to 10 wherein the activating agent is encoded by the vector genome and wherein expression or function of the agent is inducible.

15

12. A vector genome according to para. 3 wherein the activating agent is a ligand for the regulator moiety.

20

13. A vector genome according to para. 12 wherein the degron is a DD-FKBP sequence and the activating agent is a ligand therefor.

25

14. A vector genome according to para. 1 wherein the replication modulator protein encoded by the vector genome is an inhibitable modulator protein which is switchable to a targeted configuration displaying a degron on contact with a cognate inhibitory agent.

30

15. A vector genome according to para. 14 wherein the replication modulator protein comprises a viral protein moiety and a regulator moiety, and wherein the inhibitory agent cleaves the regulator moiety from the modulator protein to create or reveal the degron.

35

16. A vector genome according to para. 15 wherein the activating agent is a protease.

5 17. A vector genome according to para. 14 wherein the regulator moiety comprises a cleavage site for the protease, located between the viral protein moiety and the degron.

10 18. A vector genome according to para. 14 or para. 15 wherein the protease does not act on any other proteins encoded by the vector genome.

15 19. A vector genome according to any one of para. 16 to 18 wherein the protease is a viral protease, Factor Xa, enterokinase or thrombin.

20 20. A vector genome according to any one of para. 16 to 19 wherein the modulator protein comprises a first N-terminal residue, and is cleavable by the protease to expose a second N-terminal residue which confers lower stability than the first N-terminal residue.

25 21. A vector genome according to para. 14 wherein the replication modulator protein comprises a viral protein moiety and a regulator moiety, wherein the inhibitory agent is a ligand for the regulator moiety, and wherein the regulator moiety and inhibitory agent are components of an inducible degron system.

30 22. A vector genome according to para. 21 wherein the regulator moiety comprises a HaloTag sequence and the inhibitory agent is a ligand therefor.

35 23. A vector genome according to para. 21 wherein the regulator moiety comprises a LID-FKBP sequence and the inhibitory agent is a ligand therefor.

24. A vector genome according to para. 21 wherein the regulator moiety comprises an auxin-inducible degron sequence and the inhibitory agent is a ligand therefor.

5 25. A vector genome according to any one of paras. 14 to 24 wherein the inhibitory agent is encoded by the vector genome and wherein expression or function of the agent is inducible.

10 26. A vector genome according to any one of the preceding paras. which is a rhabdovirus vector genome.

15 27. A vector genome according to para. 26 wherein the vector genome is a lyssavirus vector genome or a vesiculovirus vector genome.

20 28. A vector genome according to para. 27 wherein the lyssavirus vector genome is a rabies virus vector genome, or wherein the vesiculovirus vector genome is a vesicular stomatitis virus vector genome.

25 29. A vector genome according to any one of the preceding paras. wherein the viral protein moiety of the replication modulator comprises or consists of mononegaviral N (Nucleoprotein) protein.

30 30. A vector genome according to para. 29 wherein the genome further comprises genes encoding a P protein, M protein, and/or L protein.

35 31. A vector genome according to any one of the preceding paras. wherein the genome further comprises a gene encoding an envelope protein.

32. A vector genome according to para. 31 wherein the envelope protein is native to the mononegaviral vector.

33. A vector genome according to para. 31 wherein the envelope protein is a pseudotyped envelope protein.

5 34. A vector genome according to any one of paras. 1 to 30 wherein the vector genome does not encode an envelope protein.

35. A vector genome according to any one of the preceding paras. wherein the vector genome further comprises one or more heterologous genes.

10

36. A vector genome according to para. 35 wherein the heterologous gene encodes a marker protein, a protein against which it is desirable to raise an immune response, a recombinase, a nuclease, a guide RNA (gRNA) molecule, a repair template RNA, or a nucleic acid modulator of gene expression.

15

37. A vector genome according to para. 36 wherein the nuclease is an RNA-guided endonuclease.

20

38. A vector genome according to para. 37 comprising heterologous genes encoding an RNA-guided endonuclease plus a guide RNA (gRNA) molecule and/or a repair template RNA.

25

39. A ribonucleoprotein complex comprising a vector genome according to any one of paras. 1 to 38 in association with one or more viral proteins.

30

40. A ribonucleoprotein complex according to para. 39 comprising the vector genome in association with N, P and L proteins.

35

41. A ribonucleoprotein complex according to para. 39 or para. 40 which is a functional viral nucleocapsid, capable of initiating transcription on introduction to the cytoplasm of a target cell.

42. A mononegaviral vector virion comprising a mononegaviral vector genome according to any one of paras. 1 to 38.

5 43. A vector virion according to para. 42 comprising a native mononegaviral envelope protein.

44. A vector virion according to para. 42 comprising a pseudotyped envelope protein.

10 45. A vector genome according to any one of paras. 1 to 38, a ribonucleoprotein complex according to any one of paras. 39 to 41, or a vector virion according to any one of paras. 42 to 44, for use in a method of medical treatment.

15 46. A vector genome according to any one of paras. 1 to 38, a ribonucleoprotein complex according to any one of paras. 39 to 41, or a vector virion according to any one of paras. 42 to 44, for use as an immunostimulatory agent.

20 47. A positive sense nucleic acid molecule encoding a viral vector genome according to any one of paras. 1 to 38.

25 48. A packaging cell comprising a nucleic acid construct encoding a vector genome according to any one of paras. 1 to 38 and capable of producing a virion according to any one of paras. 42 to 44.

30 49. A method of gene delivery to a target cell, comprising contacting the target cell with a ribonucleoprotein complex according to any one of paras. 39 to 41 or a virion according to any one of paras. 42 to 44.

35 50. A method according to para. 49 wherein the vector encodes an inhibitory modulator protein, and the method comprises contacting the cell with the cognate activating agent.

51. A method according to para. 50 wherein the activating agent is a protein, and the method comprises introducing into the target cell a nucleic acid comprising a gene encoding the activating agent, such that the activating agent is expressed in the target cell.

52. A method according to para. 49 wherein the vector encodes an inhibitable modulator protein, and the method comprises contacting the cell with the cognate inhibitory agent.

53. A method according to para. 52 wherein the inhibitory agent is a protein, and the method comprises introducing into the target cell a nucleic acid comprising a gene encoding the inhibitory agent, such that the inhibitory agent is expressed in the target cell.

54. A method according to any one of paraS. 50 to 53 wherein expression and/or function of the agent is inducible and the method comprises the step of inducing expression and/or function of the agent in the target cell.

55. A method according to para. 49 wherein:
(i) the vector encodes an inhibitory modulator protein and the activating agent, wherein expression or function of the activating agent is inducible, and wherein the method comprises the step of inducing expression and/or function of the activating agent in the target cell; or
(ii) the vector encodes an inhibitable modulator protein and the inhibitory agent, wherein expression or function of the inhibitory agent is inducible, and wherein the method comprises the step of inducing expression and/or function of the inhibitory agent in the target cell.

56. A method according to para. 55 comprising contacting the target cell with the cognate inducer.

57. A method according to any one of paras. 49 to 56 wherein the vector genome does not comprise a gene encoding an envelope protein and the method comprises the step of introducing into the target cell a nucleic acid construct comprising a gene encoding an envelope protein.

58. A method according to any one of paras. 49 to 57 wherein the target cell is a neural cell.

59. A kit comprising a vector genome according to any one of paras. 1 to 38, a ribonucleoprotein complex according to any one of paras. 39 to 41 or a virion according to any one of paras. 42 to 44 and (a) a cognate activating or inhibitory agent, or a nucleic acid encoding a cognate activating or inhibitory agent, and/or (b) a nucleic acid encoding an envelope protein.

60. A composition comprising ribonucleoprotein complex according to any one of paras. 39 to 41 or a virion according to any one of paras. 42 to 44, optionally admixed with an excipient or carrier.

61. A composition according to para. 60 wherein the composition is a pharmaceutical composition and the carrier is a pharmaceutically acceptable carrier.

Claims

1. A rabies virus vector genome comprising a gene encoding a replication modulator protein, wherein the replication modulator protein comprises a viral protein moiety which is required for replication of the viral genome, the replication modulator protein being capable of adopting a targeted configuration displaying a degron, and an untargeted configuration which does not display the degron, and wherein the viral protein moiety is an N (nucleoprotein) protein.
2. A vector genome according to claim 1 wherein the replication modulator protein encoded by the vector genome is an inhibitory modulator comprising a viral protein moiety and a regulator moiety, wherein the regulator moiety comprises or consists of the degron.
3. A vector genome according to claim 2 wherein the regulator moiety is switchable to an untargeted configuration on contact with a cognate activating agent.
4. A vector genome according to claim 3 wherein the activating agent cleaves the regulator moiety from the viral protein moiety.
5. A vector genome according to claim 4 wherein the activating agent is a protease.
6. A vector genome according to claim 5 wherein the regulator moiety comprises a cleavage site for the protease, located between the viral protein moiety and the degron.
7. A vector genome according to claim 5 or claim 6 wherein the protease does not act on any other proteins encoded by the vector genome.

8. A vector genome according to any one of claims 5 to 7 wherein the protease is a viral protease, Factor Xa, enterokinase or thrombin.

5 9. A vector genome according to any one of claims 2 to 8 wherein the degron is a PEST sequence.

10 10. A vector genome according to any one of claims 2 to 9 wherein the modulator protein comprises a first N-terminal residue, and is cleavable by the cognate protease to expose a second N-terminal residue which confers greater stability than the first N-terminal residue.

15 11. A vector genome according to any one of claims 3 to 10 wherein the activating agent is encoded by the vector genome and wherein expression or function of the agent is inducible.

20 12. A vector genome according to claim 3 wherein the activating agent is a ligand for the regulator moiety.

13. A vector genome according to claim 12 wherein the degron is a DD-FKBP sequence and the activating agent is a ligand therefor.

25 14. A vector genome according to claim 1 wherein the replication modulator protein encoded by the vector genome is an inhibitable modulator protein which is switchable to a targeted configuration displaying a degron on contact with a cognate inhibitory agent.

30 15. A vector genome according to claim 14 wherein the replication modulator protein comprises a viral protein moiety and a regulator moiety, and wherein the inhibitory agent cleaves the regulator moiety from the modulator protein to
35 create or reveal the degron.

16. A vector genome according to claim 15 wherein the activating agent is a protease.

17. A vector genome according to claim 14 wherein the regulator moiety comprises a cleavage site for the protease, located between the viral protein moiety and the degron.

18. A vector genome according to claim 14 or claim 15 wherein the protease does not act on any other proteins encoded by the vector genome.

19. A vector genome according to any one of claims 16 to 18 wherein the protease is a viral protease, Factor Xa, enterokinase or thrombin.

20. A vector genome according to any one of claims 16 to 19 wherein the modulator protein comprises a first N-terminal residue, and is cleavable by the protease to expose a second N-terminal residue which confers lower stability than the first N-terminal residue.

21. A vector genome according to claim 14 wherein the replication modulator protein comprises a viral protein moiety and a regulator moiety, wherein the inhibitory agent is a ligand for the regulator moiety, and wherein the regulator moiety and inhibitory agent are components of an inducible degron system.

22. A vector genome according to claim 21 wherein the regulator moiety comprises a HaloTag sequence and the inhibitory agent is a ligand therefor.

23. A vector genome according to claim 21 wherein the regulator moiety comprises a LID-FKBP sequence and the inhibitory agent is a ligand therefor.

24. A vector genome according to claim 21 wherein the regulator moiety comprises an auxin-inducible degron sequence and the inhibitory agent is a ligand therefor.

5 25. A vector genome according to any one of claims 14 to 24 wherein the inhibitory agent is encoded by the vector genome and wherein expression or function of the agent is inducible.

10 26. A vector genome according to any one of the preceding claims wherein the genome further comprises genes encoding a P protein, M protein, and/or L protein.

15 27. A vector genome according to any one of the preceding claims wherein the genome further comprises a gene encoding an envelope protein.

28. A vector genome according to claim 27 wherein the envelope protein is native to the mononegaviral vector.

20 29. A vector genome according to claim 27 wherein the envelope protein is a pseudotyped envelope protein.

25 30. A vector genome according to any one of claims 1 to 26 wherein the vector genome does not encode an envelope protein.

31. A vector genome according to any one of the preceding claims wherein the vector genome further comprises one or more heterologous genes.

30 32. A vector genome according to claim 35 wherein the heterologous gene encodes a marker protein, a protein against which it is desirable to raise an immune response, a recombinase, a nuclease, a guide RNA (gRNA) molecule, a repair template RNA, or a nucleic acid modulator of gene expression.

35 33. A vector genome according to claim 32 wherein the nuclease is an RNA-guided endonuclease.

34. A vector genome according to claim 33 further comprising heterologous genes encoding a guide RNA (gRNA) molecule and/or a repair template RNA.

5

35. A ribonucleoprotein complex comprising a vector genome according to any one of claims 1 to 34 in association with one or more viral proteins.

10

36. A ribonucleoprotein complex according to claim 35 comprising the vector genome in association with N, P and L proteins.

15

37. A ribonucleoprotein complex according to claim 35 or claim 36 which is a functional viral nucleocapsid, capable of initiating transcription on introduction to the cytoplasm of a target cell.

20

38. A rabies virus vector virion comprising a viral vector genome according to any one of claims 1 to 38.

39. A vector virion according to claim 38 comprising a rabies virus envelope protein.

25

40. A vector virion according to claim 38 comprising a pseudotyped envelope protein.

30

41. A vector genome according to any one of claims 1 to 34, a ribonucleoprotein complex according to any one of claims 35 to 37, or a vector virion according to any one of claims 38 to 40, for use in a method of medical treatment.

35

42. A vector genome according to any one of claims 1 to 34, a ribonucleoprotein complex according to any one of claims 35 to 37, or a vector virion according to any one of claims 38 to 40, for use as an immunostimulatory agent.

43. A positive sense nucleic acid molecule encoding a viral vector genome according to any one of claims 1 to 34.

5 44. A packaging cell comprising a nucleic acid construct encoding a vector genome according to any one of claims 1 to 34 and capable of producing a virion according to any one of claims 38 to 40.

10 45. A method of gene delivery to a target cell, comprising contacting the target cell with a ribonucleoprotein complex according to any one of claims 35 to 37 or a virion according to any one of claims 38 to 40.

15 46. A method according to claim 45 wherein the vector encodes an inhibitory modulator protein, and the method comprises contacting the cell with the cognate activating agent.

20 47. A method according to claim 46 wherein the activating agent is a protein, and the method comprises introducing into the target cell a nucleic acid comprising a gene encoding the activating agent, such that the activating agent is expressed in the target cell.

25 48. A method according to claim 45 wherein the vector encodes an inhibitable modulator protein, and the method comprises contacting the cell with the cognate inhibitory agent.

30 49. A method according to claim 48 wherein the inhibitory agent is a protein, and the method comprises introducing into the target cell a nucleic acid comprising a gene encoding the inhibitory agent, such that the inhibitory agent is expressed in the target cell.

35 50. A method according to any one of claims 46 to 49 wherein expression and/or function of the agent is inducible and the method comprises the step of inducing expression and/or function of the agent in the target cell.

51. A method according to claim 45 wherein:

(i) the vector encodes an inhibitory modulator protein and the activating agent, wherein expression or function of the activating agent is inducible, and wherein the method comprises the step of inducing expression and/or function of the activating agent in the target cell; or

(ii) the vector encodes an inhibitable modulator protein and the inhibitory agent, wherein expression or function of the inhibitory agent is inducible, and wherein the method comprises the step of inducing expression and/or function of the inhibitory agent in the target cell.

52. A method according to claim 51 comprising contacting the target cell with the cognate inducer.

53. A method according to any one of claims 45 to 52 wherein the vector genome does not comprise a gene encoding an envelope protein and the method comprises the step of introducing into the target cell a nucleic acid construct comprising a gene encoding an envelope protein.

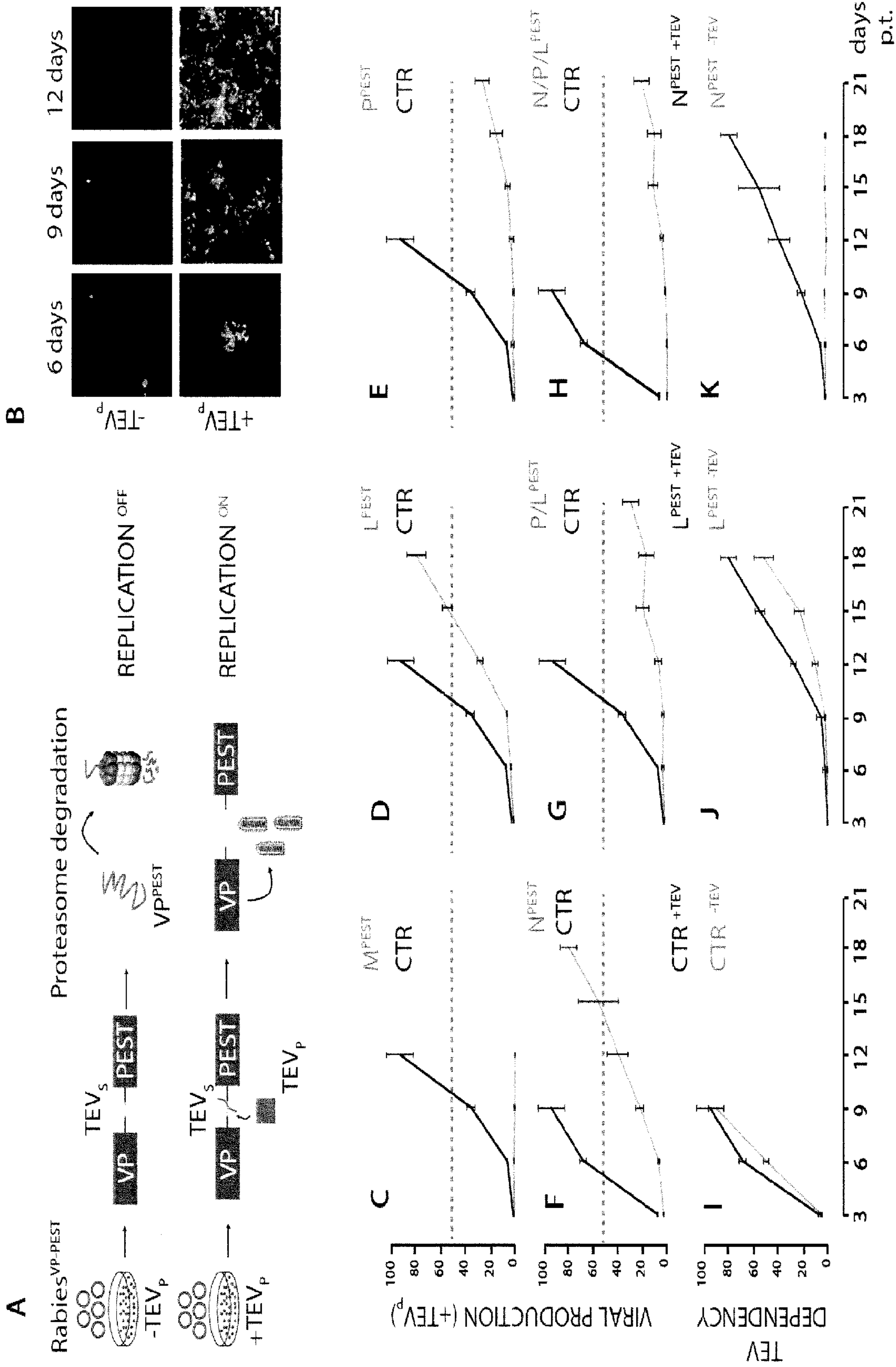
54. A method according to any one of claims 45 to 53 wherein the target cell is a neural cell.

55. A kit comprising a vector genome according to any one of claims 1 to 34, a ribonucleoprotein complex according to any one of claims 35 to 37 or a virion according to any one of claims 38 to 40 and (a) a cognate activating or inhibitory agent, or a nucleic acid encoding a cognate activating or inhibitory agent, and/or (b) a nucleic acid encoding an envelope protein.

56. A composition comprising ribonucleoprotein complex according to any one of claims 35 to 37 or a virion according to any one of claims 38 to 40, optionally admixed with an excipient or carrier.

57. A composition according to claim 56 wherein the composition is a pharmaceutical composition and the carrier is a pharmaceutically acceptable carrier.

5



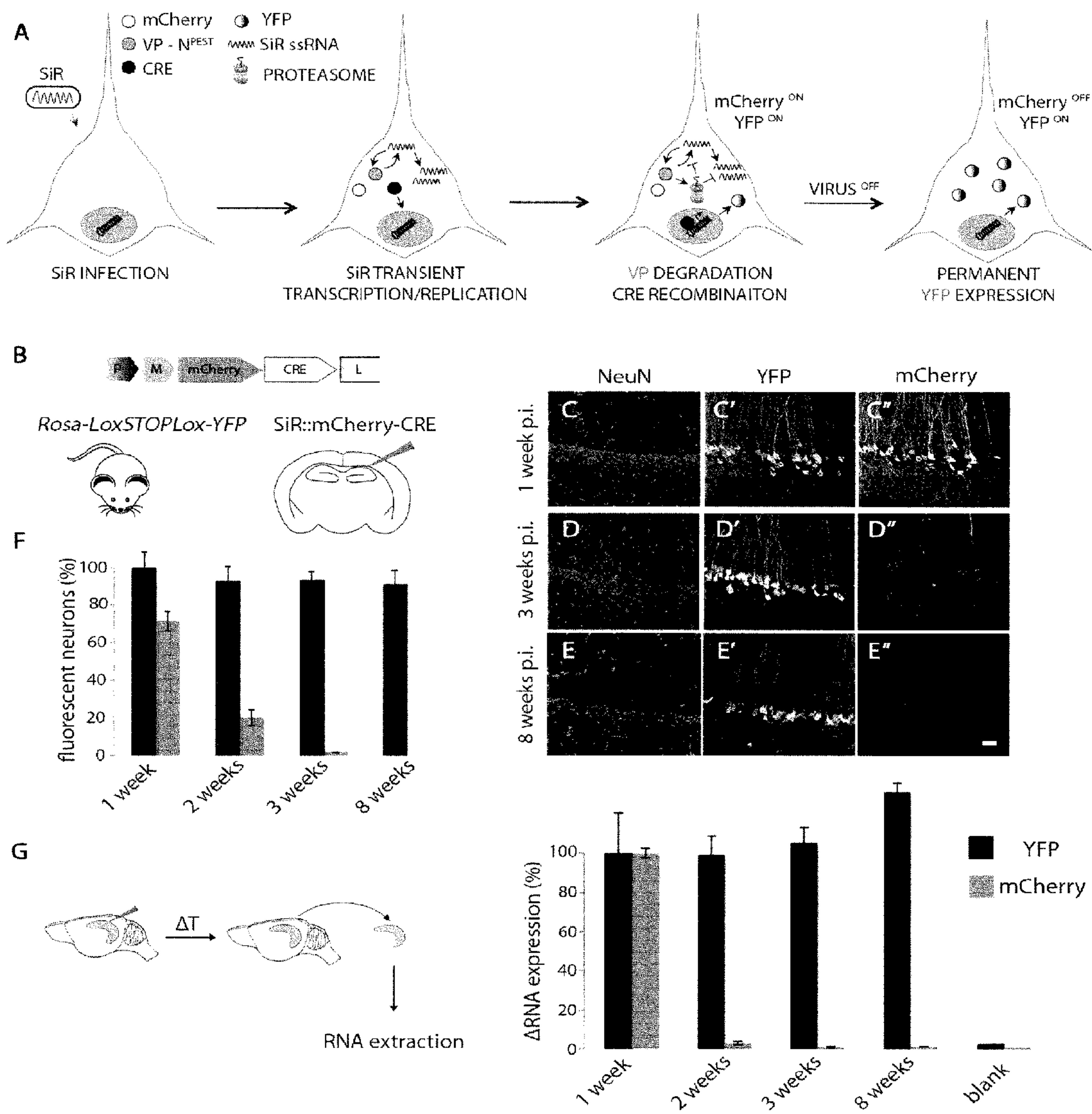
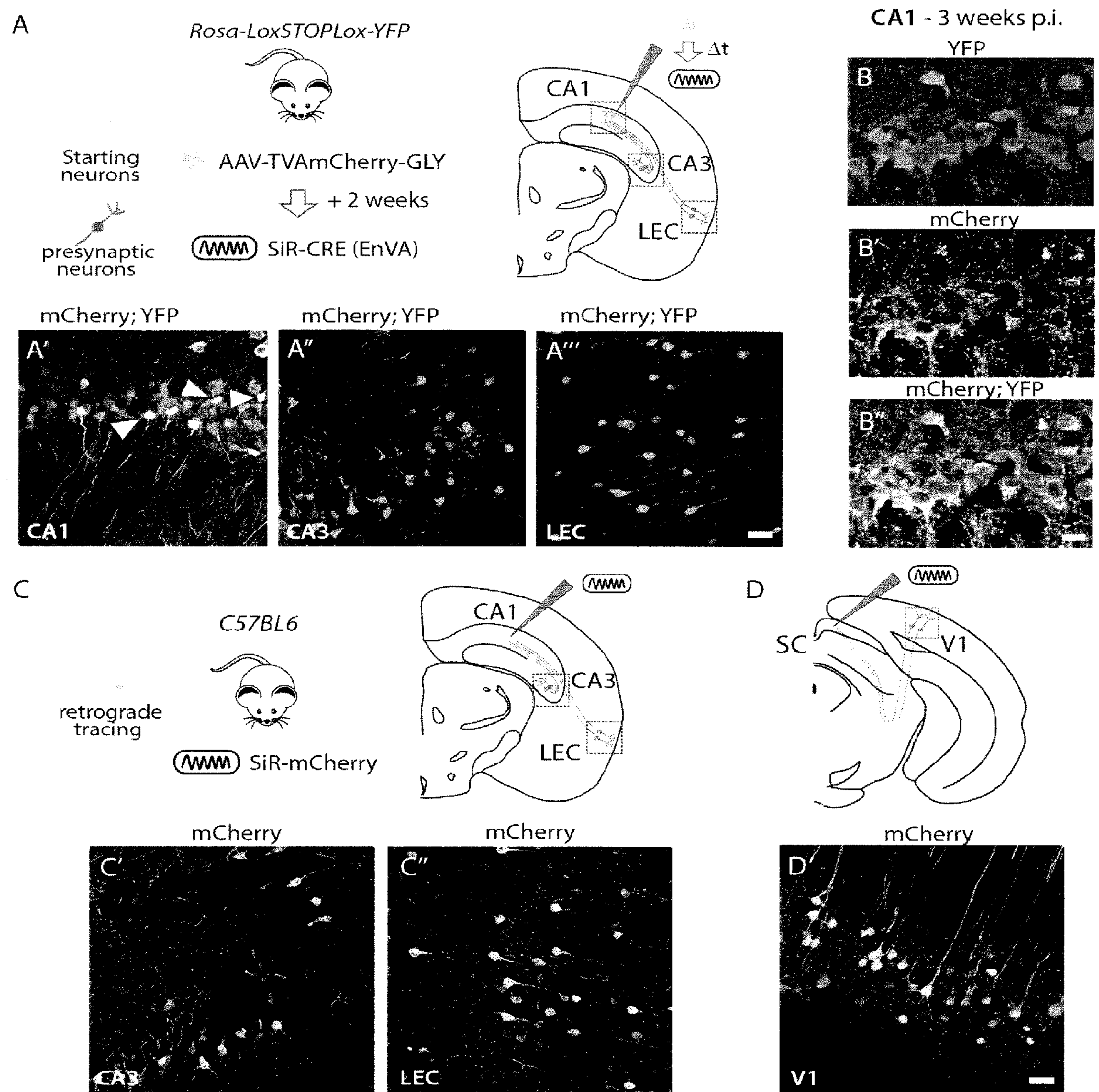
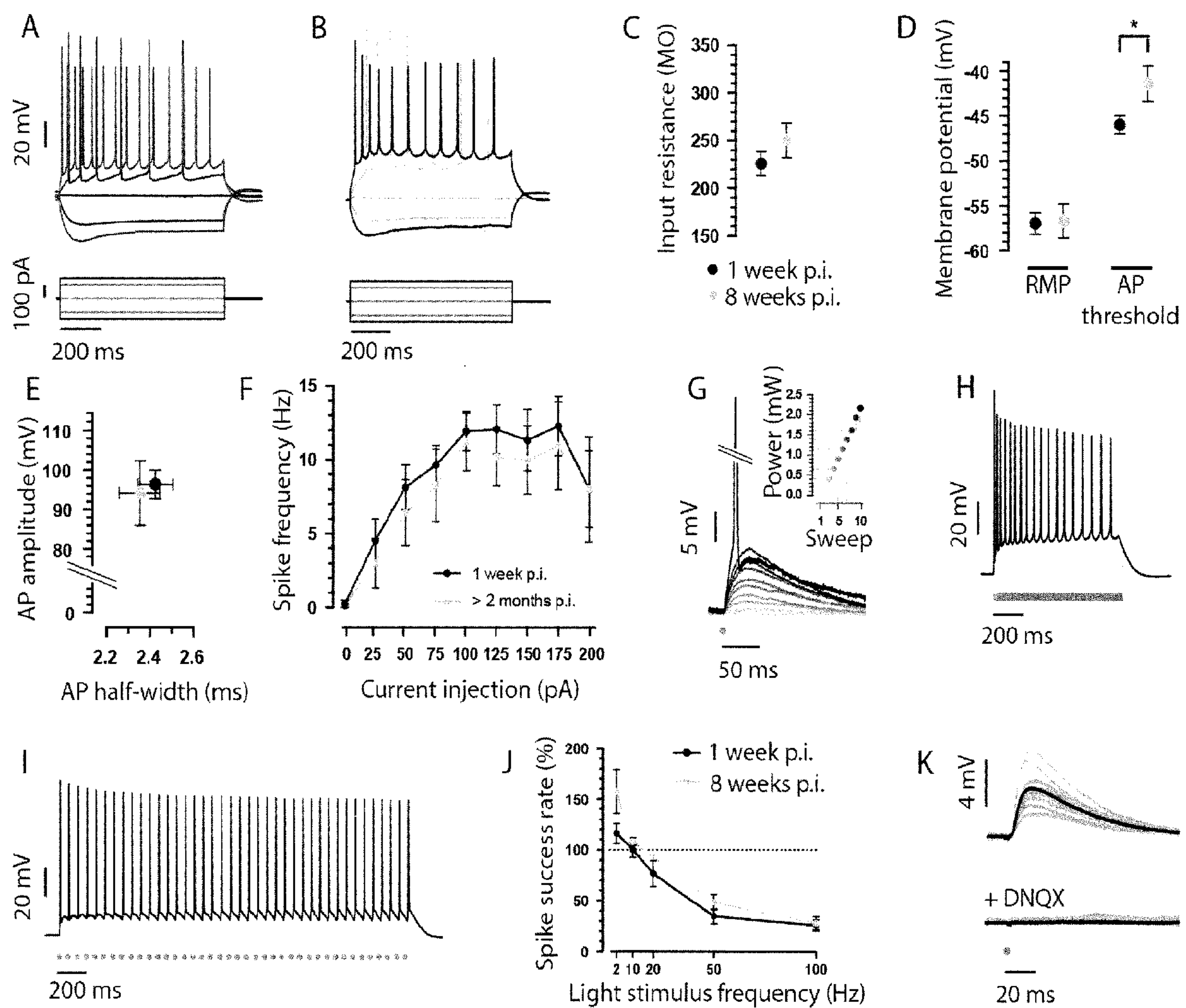
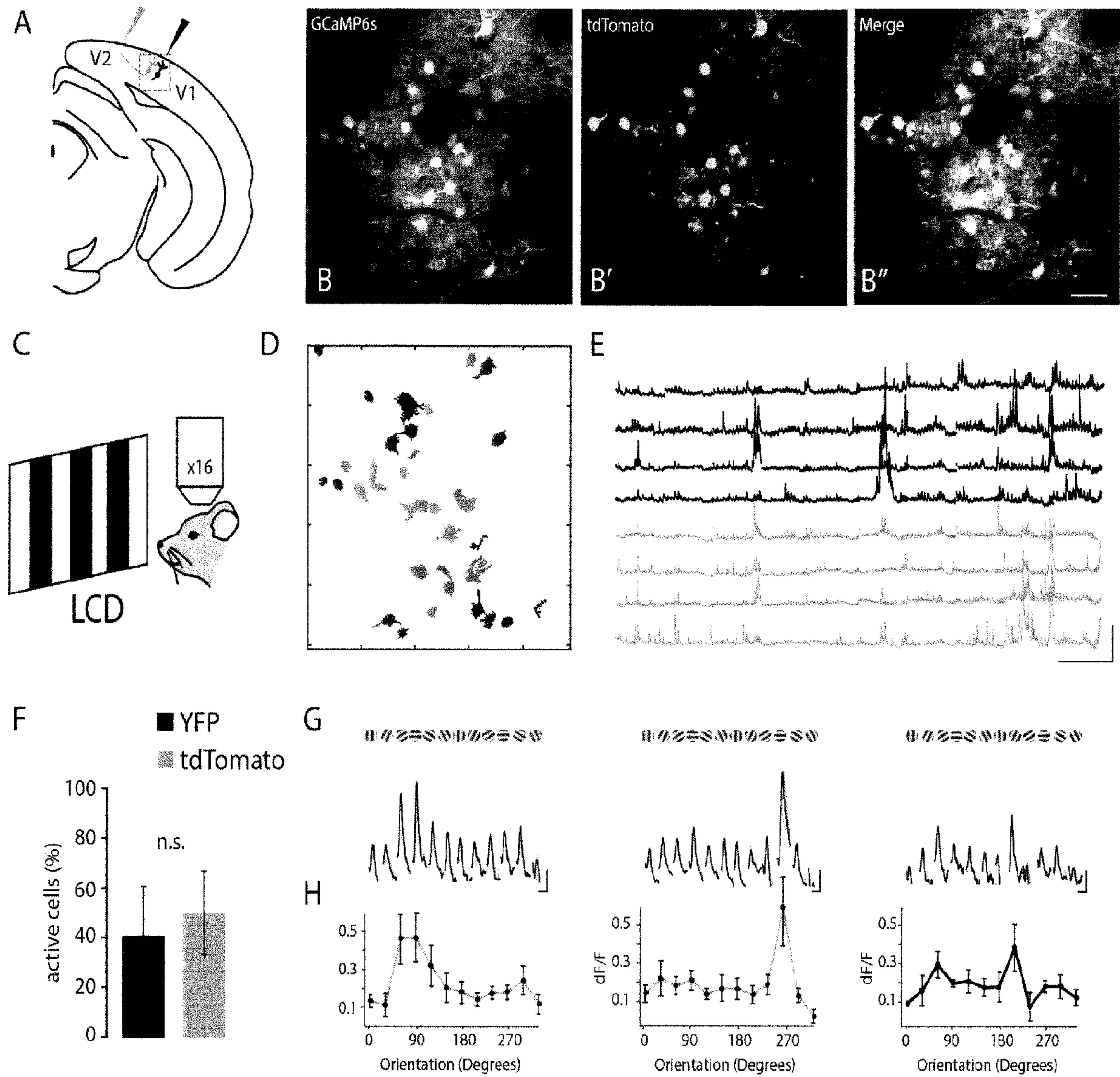
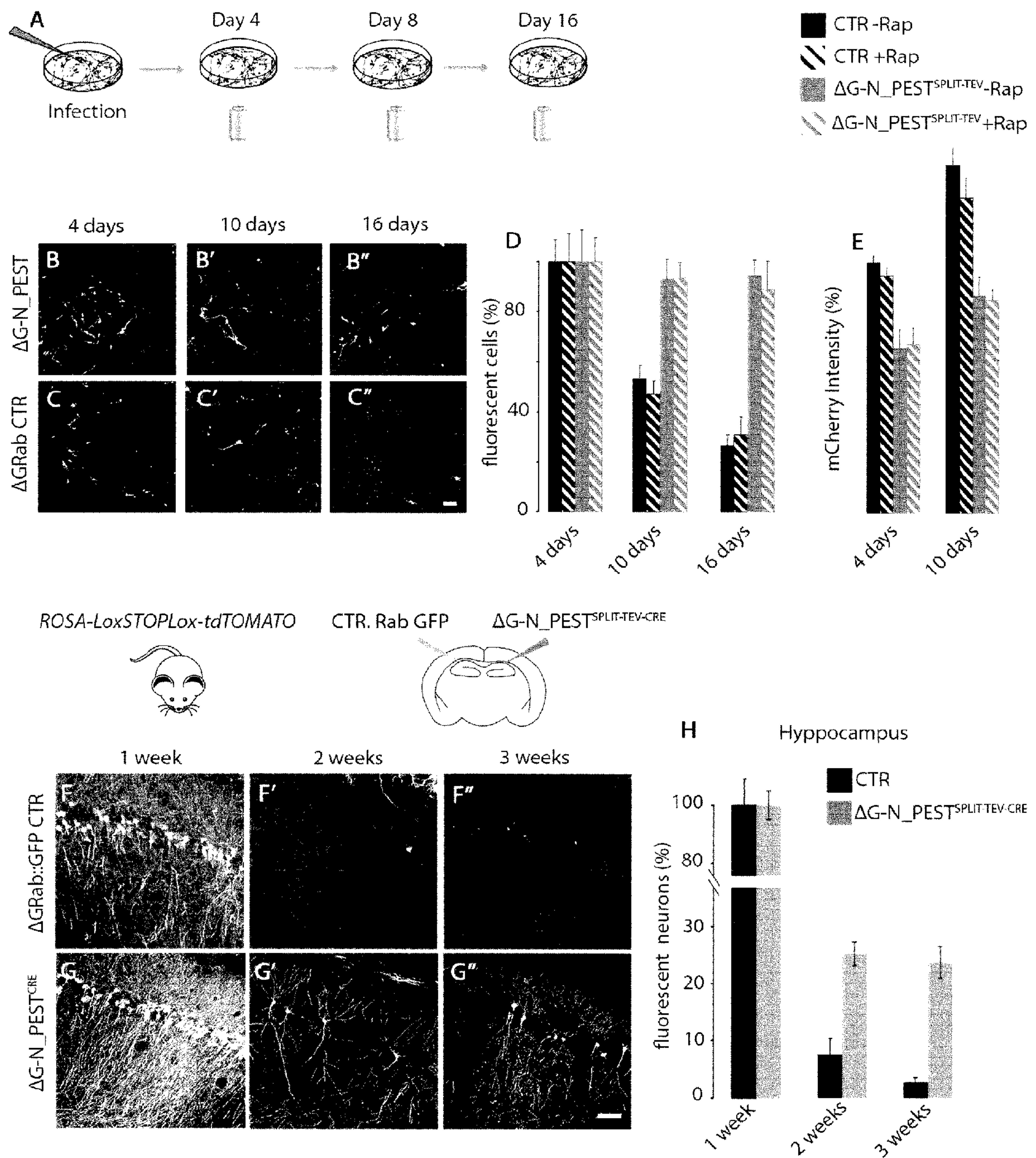


Fig. 2

**Fig. 3**

**Fig. 4**

**Fig. 5**

**Fig. 6**

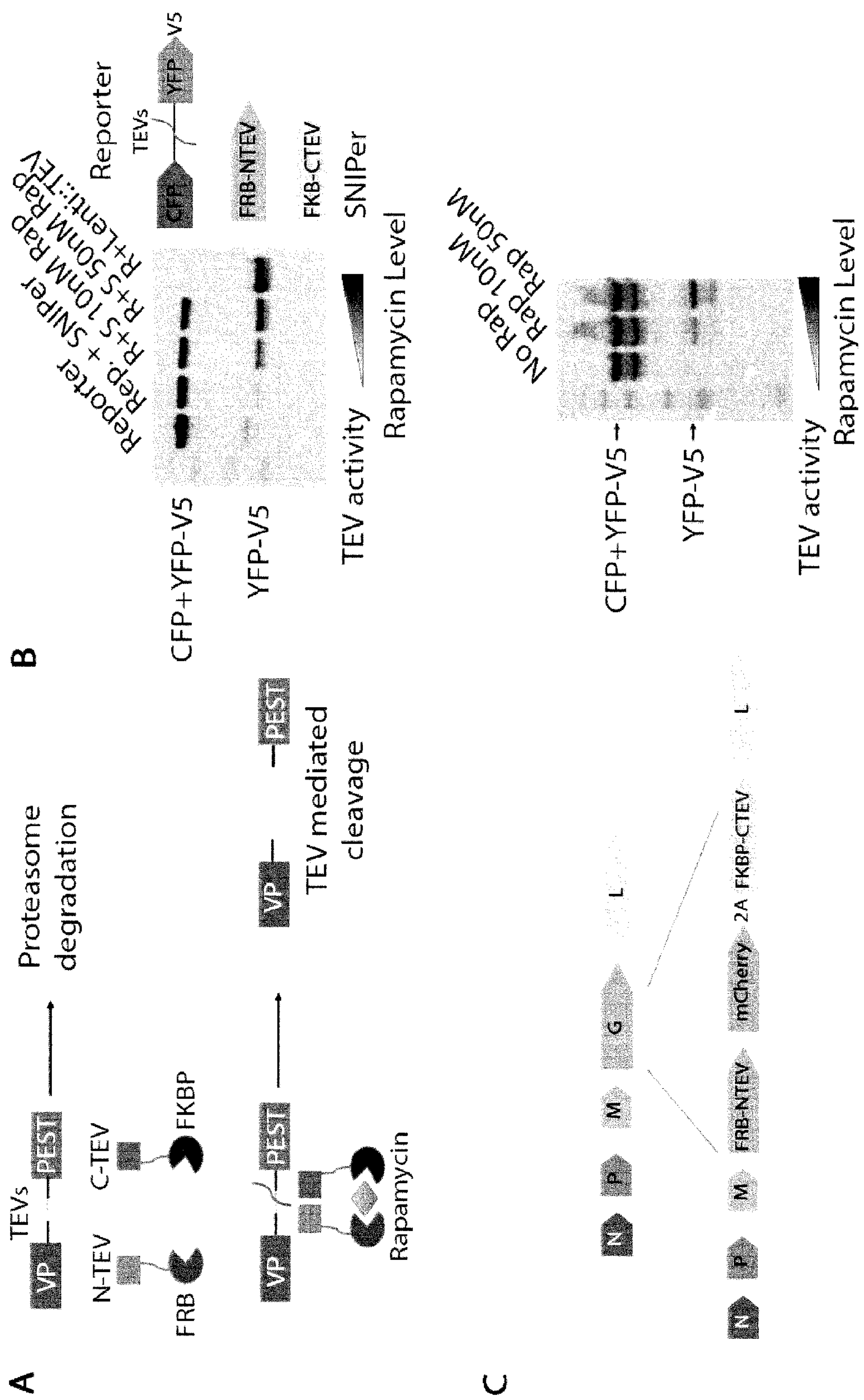


Fig. 7

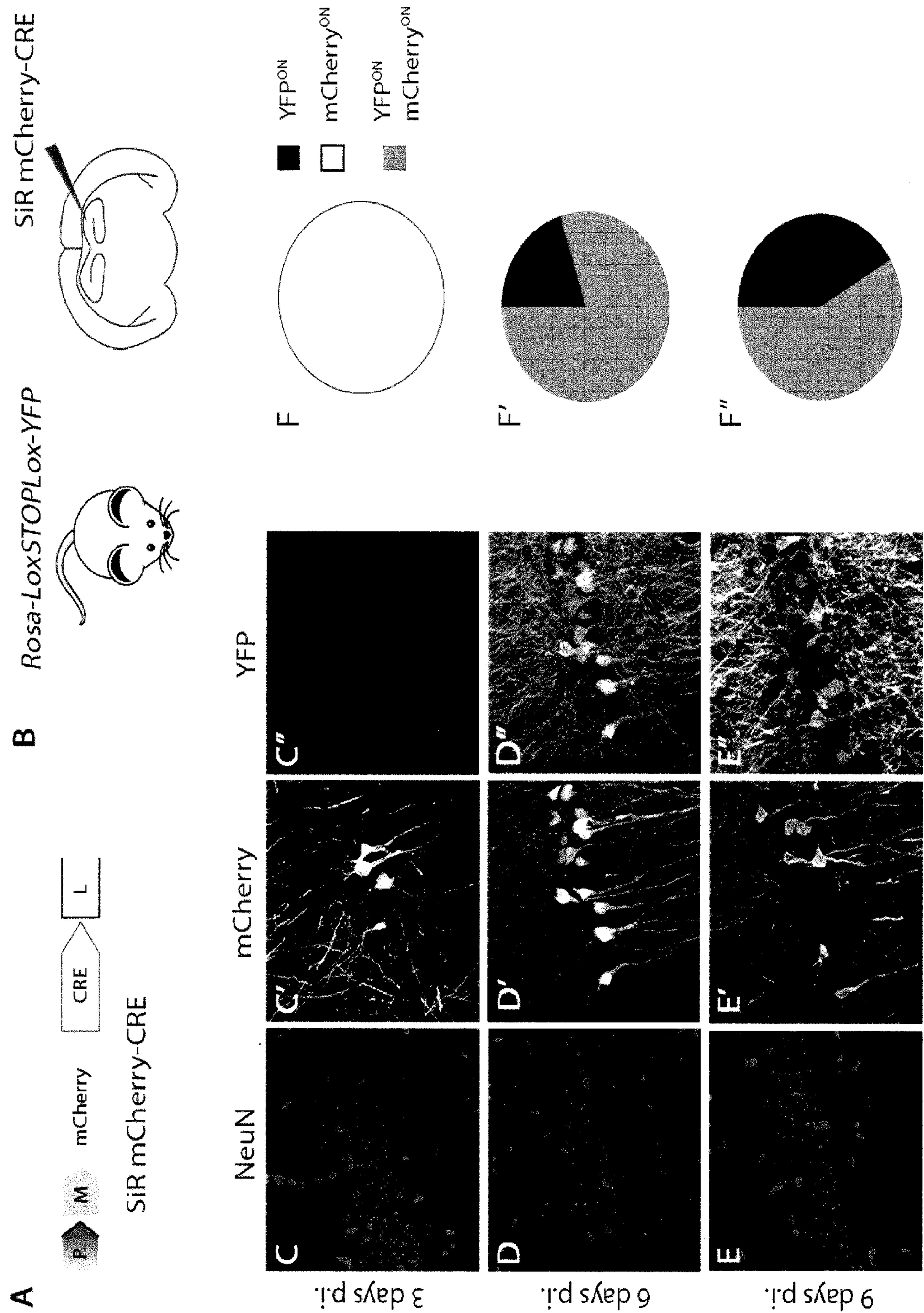


Fig. 8

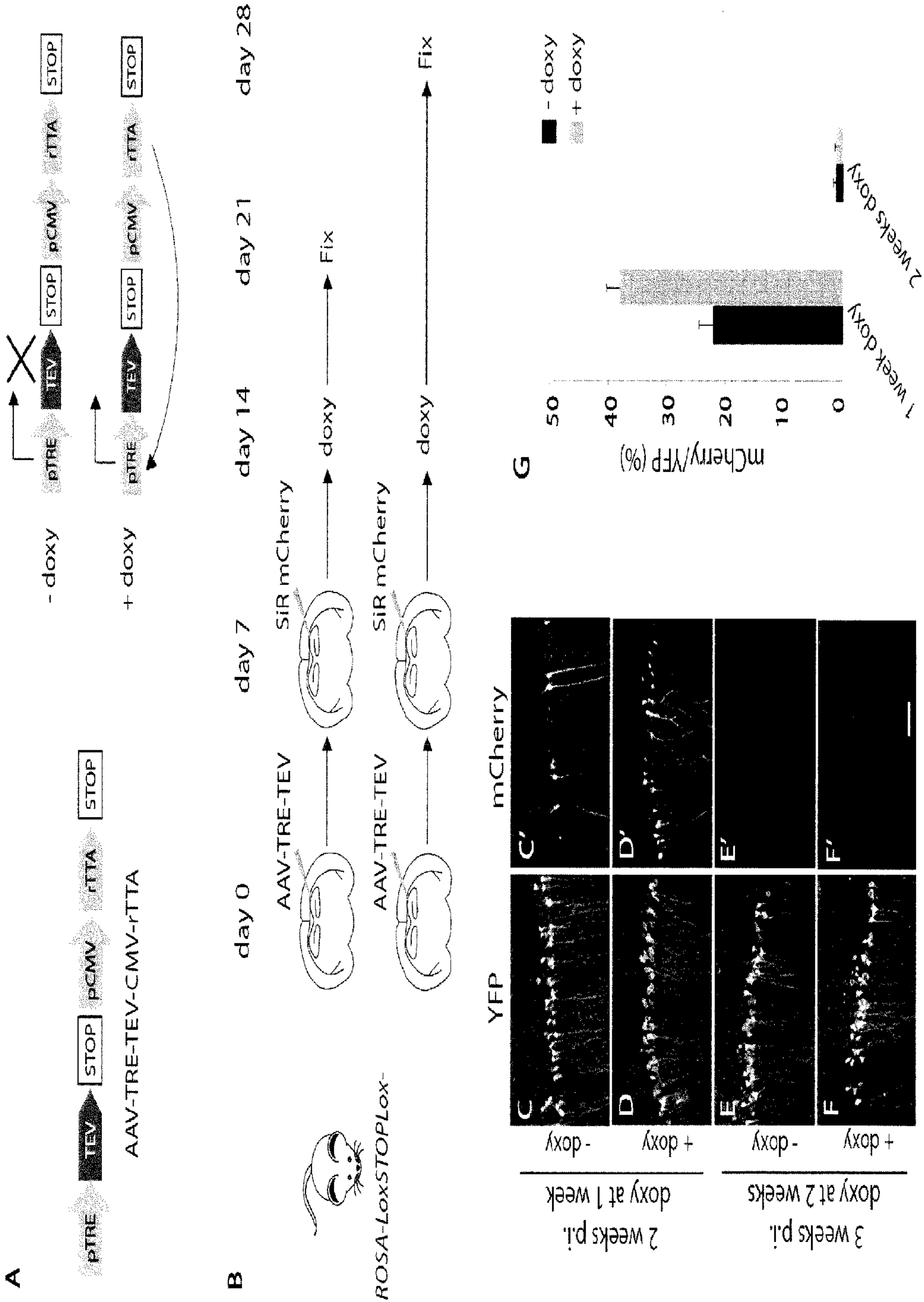


Fig. 9

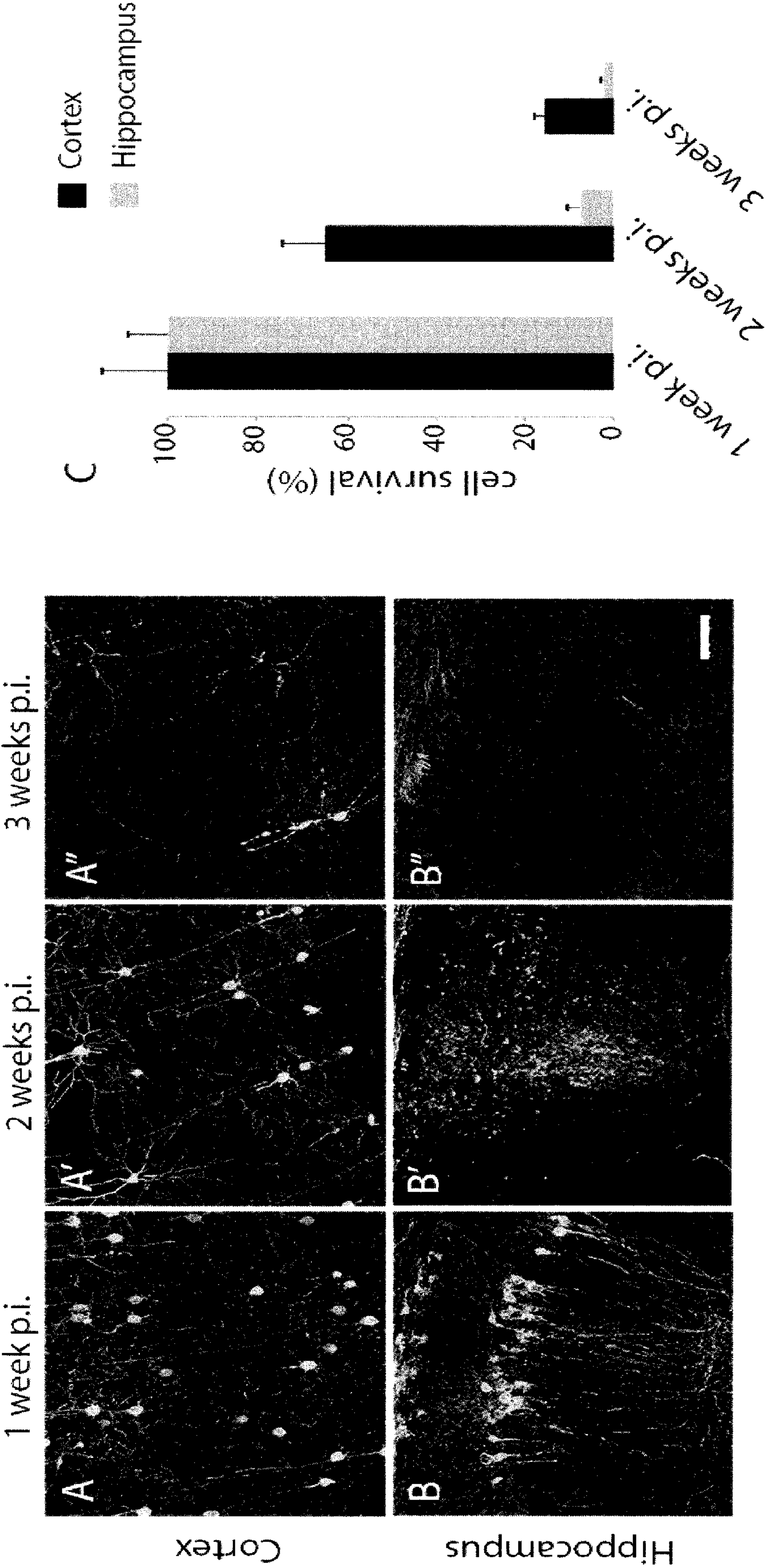


Fig. 10

(A)

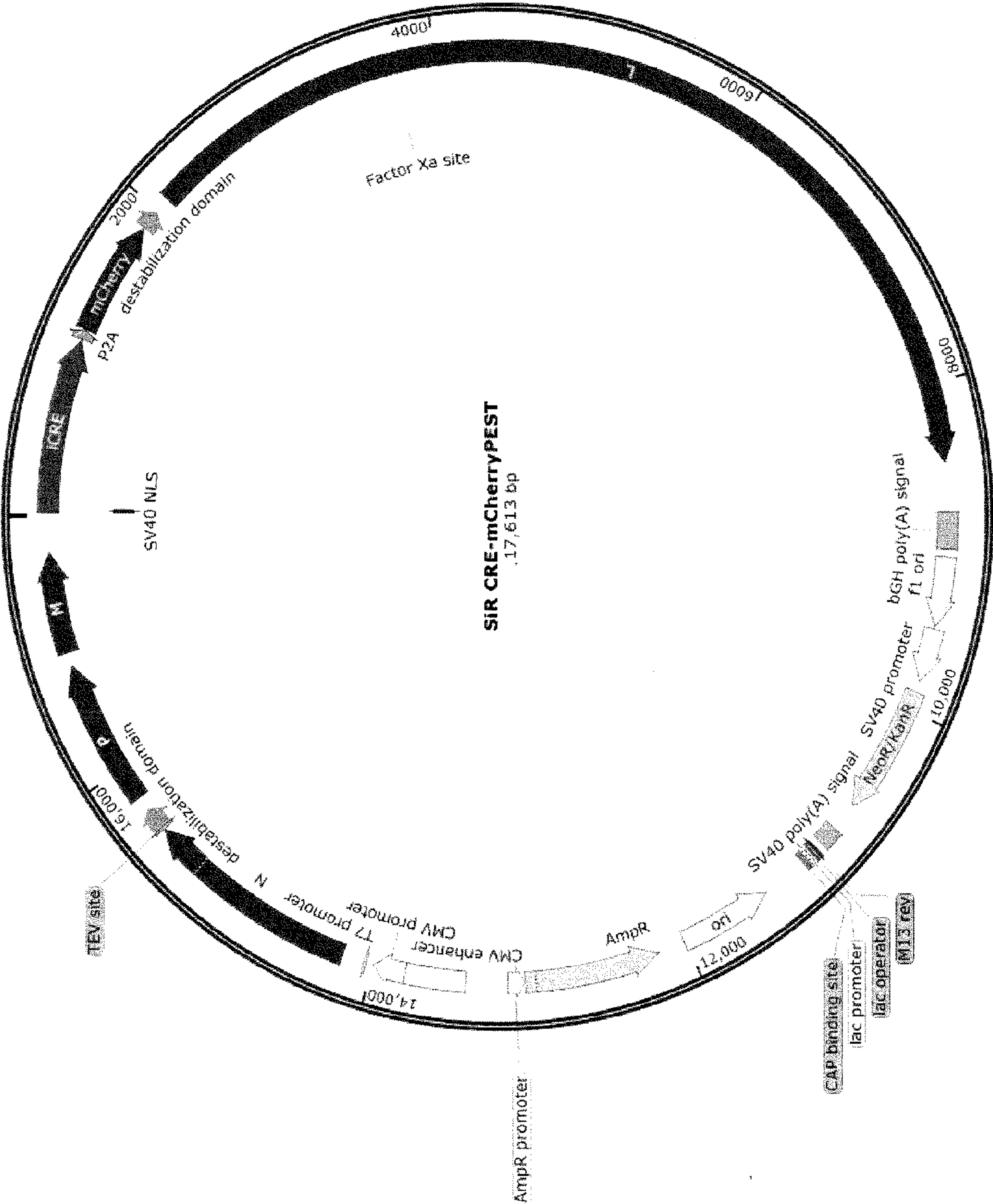


Fig. 11

(B)

GACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGA
 GTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTTGACG
 TCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATT
 TACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAA
 TGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTAC
 ATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGC
 GGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTTGACGTCAATGGGAGTTTGTTTTGGCACCAAAA
 TCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGG
 TGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTA
ATACGACTCACTATAGGGAGACCCAAGCTGGCTAGATTAAGCGTCTGATGAGTCCGTGAGGACGAAACCC
GGAGTCCCGGGTCACGCTTAACAACCAGATCAAAGAAAAAACAGACATTGTCAATTGCAAAGCAAAAATG
TAACACCCCTACAATGGATGCCGACAAGATTGTATTCAAAGTCAATAATCAGGTGGTCTCTTTGAAGCCT
GAGATTATCGTGGATCAATATGAGTACAAGTACCCTGCCATCAAAGATTTGAAAAAGCCCTGTATAACCC
TAGGAAAGGCTCCCGATTTAAATAAAGCATACAAGTCAGTTTTTGTGAGGCATGAGCGCCGCCAAACTTAA
TCCTGACGATGTATGTTCTTATTTGGCAGCGGCAATGCAGTTTTTTGAGGGGACATGTCCGGAAGACTGG
ACCAGCTATGGAATTGTGATTGCACGAAAAGGAGATAAGATCACCCAGGTTCTCTGGTGGAGATAAAAC
GTACTGATGTAGAAGGGAATTGGGCTCTGACAGGAGGCATGGAACTGACAAGAGACCCCACTGTCCCTGA
GCATGCGTCCTTAGTCGGTCTTCTCTTGAGTCTGTATAGGTTGAGCAAAATATCCGGGCAAAACACTGGT
AACTATAAGACAAACATTGCAGACAGGATAGAGCAGATTTTTGAGACAGCCCTTTTGTTAAATCGTG
AACACCATACTCTAATGACAACTCACAAAATGTGTGCTAATTGGAGTACTATACCAAACTTCAGATTTTT
GGCCGGAACCTATGACATGTTTTTCTCCCGGATTGAGCATCTATATTCAGCAATCAGAGTGGGCACAGTT
GTCACTGCTTATGAAGACTGTTCAGGACTGGTATCATTTACTGGGTTCATAAAACAAATCAATCTCACCG
CTAGAGAGGCAATACTATATTTCTTCCACAAGAACTTTGAGGAAGAGATAAGAAGAATGTTTGAGCCAGG
GCAGGAGACAGCTGTTCTCTACTCTTATTTTCATCCACTTCCGTTCACTAGGCTTGAGTGGGAAATCTCCT
TATTCATCAAATGCTGTTGGTCACGTGTTCAATCTCATTCACTTTGTAGGATGCTATATGGGTCAAGTCA
GATCCCTAAATGCAACGGTTATTGCTGCATGTGCTCCTCATGAAATGTCTGTTCTAGGGGGCTATCTGGG
AGAGGAATTCTTCGGGAAAGGGACATTTGAAAGAAGATTCTTCAGAGATGAGAAAGAACTTCAAGAATAC
GAGGCGGCTGAACTGACAAAGACTGACGTAGCACTGGCAGATGATGGAACTGTCAACTCTGACGACGAGG
ACTACTTTTCAGGTGAAACCAGAAGTCCGGAGGCTGTTTATACTCGAATCATGATGAATGGAGGTGCACT
AAAGAGATCTCACATACGGAGATATGTCTCAGTCAGTTCCAATCATCAAGCCCGTCCAAACTCATTCGCC
GAGTTTCTAAACAAGACATATTCGAGTGACTCAGGTTCCGGAGAGAACCTCTACTTCCAATCGGGATCCG
GTAGCCATGGCTTCCCGCCGGAGGTGGAGGAGCAGGATGATGGCACGCTGCCCATGTCTTGTGCCCAGGA
GAGCGGGATGGACCGTCACCCTGCAGCCTGTGCTTCTGCTAGGATCAATGTGTAAGAAGTTGAATAACAA
AATGCCGGAAATCTACGGATTGTGTATATCCATCATGAAAAAACTAACACCCCTCCTTTTGAACCATCC
CAAACATGAGCAAGATCTTTGTCAATCCTAGTGCTATTAGAGCCGGTCTGGCCGATCTTGAGATGGCTGA
AGAACTGTTGATCTGATCAATAGAAATATCGAAGACAATCAGGCTCATCTCCAAGGGGAACCCATAGAG
GTGGACAATCTCCCTGAGGATATGGGGCGACTTCACCTGGATGATGGAAAATCGCCCAACCATGGTGAGA
TAGCCAAGGTGGGAGAAGGCAAGTATCGAGAGGACTTTCAGATGGATGAAGGAGAGGATCCTAGCTTCCT
GTTCCAGTCATACCTGGAAAATGTTGGAGTCCAAATAGTCAGACAAATGAGGTCAGGAGAGAGATTTCTC
AAGATATGGTCACAGACCGTAGAAGAGATTATATCCTATGTGCGGGTCAACTTTCCCAACCCCTCAGGAA
AGTCTTCAGAGGATAAATCAACCCAGACTACTGGCCGAGAGCTCAAGAAGGAGACAACACCCACTCCTTC
TCAGAGAGAAAGCCAATCATCGAAAGCCAGGATGGCGGCTCAAATTGCTTCTGGCCCTCCAGCCCTTGAA
TGGTCGGCTACCAATGAAGAGGATGATCTATCAGTGGAGGCTGAGATCGCTCACCAGATTGCAGAAAGTT
TCTCCAAAAAATATAAGTTTTCCCTCTCGATCCTCAGGGATACTCTTGATAATTTTGAGCAATTGAAAAT
GAACCTTGATGATATAGTTAAAGAGGCAAAAAATGTACCAGGTGTGACCCGTTTAGCCCATGACGGGTCC
AAACTCCCCCTAAGATGTGTACTGGGATGGGTGCTTTGGCCAACTCTAAGAAATTCAGTTGTTAGTCG
AATCCGACAAGCTGAGTAAATCATGCAAGATGACTTGAATCGCTATACATCTTGCTAACCGAACCTCTC
CCCTCAGTCCCTCTAGACAATAAAATCCGAGATGTCCCAAAG

Fig. 11 (contd.)

TCAACATGAAAAAACAGGCAACACCACTGATAAATCGATGAACCTCCTACGTAAGATAGTGAAAAACCG
CAGGGACGAGGACACTCAAAAATCCTCTCCCGCGTCAGCCCCCTCTGGATGACGATGACTTGTGGCTTCCA
CCCCCTGAATACGTCCCGCTGAAAGAACTTACAGGCAAGAAGAACATGAGGAACTTTTGTATCAACGGAA
GGGTTAAAGTGTGTAGCCCGAATGGTTACTCGTTCAGGATCCTGCGGCACATTCTGAAATCATTTCGACGA
GATATATTCTGGGAATCATAGGATGATCGGGTTAGTCAAAGTGGTTATTGGACTGGCTTTGTCAGGATCT
CCAGTCCCTGAGGGCCTGAACTGGGTATACAAATTGAGGAGAACCTTTATCTTCCAGTGGGCTGATTCCA
GGGGCCCTCTTGAAGGGGAGGAGTTGGAATACTCTCAGGAGATCACTTGGGATGATGATACTGAGTTCTG
CGGATTGCAAATAAGAGTGATTGCAAAACAGTGTCTATCCAGGGCAGAGTCTGGTGTATCAACATGAAC
CCGAGAGCATGTCAACTATGGTCTGACATGTCTCTTCAGACACAAAGGTCCGAAGAGGACAAAGATTCTT
CTCTGCTTCTAGATAACGTACGTCAGATTATATCCCGCAAATTTATCACTTGTTTACCTCTGGAGGAGA
GAACATATGGGCTCAACTCCAACCCCTTGGGAGCAATATAACAAAAAACATGTTATGGTGCCATTAAACCG
CTGCATTTTCATCAAAGTCAAGTTGATTACCTTTACATTTTGATCCTCTTGGATGTGAAAAAACTATTAA
CATCCCTCAAAGGACCTGCAGGTACGCGGCCGCGGTACCGCCACCATTGGTGCCCAAGAAGAAGAGGAAAG
TCTCCAACCTGCTGACTGTGCACCAAAACCTGCCTGCCCTCCCTGTGGATGCCACCTCTGATGAAGTCAG
GAAGAACCTGATGGACATGTTTCAGGGACAGGCAGGCCTTCTCTGAACACACCTGGAAGATGCTCCTGTCT
GTGTGCAGATCCTGGGCTGCCTGGTGCAAGCTGAACAACAGGAAATGGTTCCCTGCTGAACCTGAGGATG
TGAGGGACTACCTCCTGTACCTGCAAGCCAGAGGCCTGGCTGTGAAGACCATCCAACAGCACCTGGGCCA
GCTCAACATGCTGCACAGGAGATCTGGCCTGCCTCGCCCTTCTGACTCCAATGCTGTGTCCCTGGTGATG
AGGAGAATCAGAAAGGAGAATGTGGATGCTGGGGAGAGAGCCAAGCAGGCCCTGGCCTTTGAACGCACTG
ACTTTGACCAAGTCAGATCCCTGATGGAGAATCTGACAGATGCCAGGACATCAGGAACCTGGCCTTCCT
GGGCATTGCCTACAACACCCTGCTGCGCATTGCCGAAATTGCCAGAATCAGAGTGAAGGACATCTCCCGC
ACCGATGGTGGGAGAATGCTGATCCACATTGGCAGGACCAAGACCCTGGTGTCCACAGCTGGTGTGGAGA
AGGCCCTGTCCCTGGGGGTTACCAAGCTGGTGGAGAGATGGATCTCTGTGTCTGGTGTGGCTGATGACCC
CAACAACCTACCTGTTCTGCCGGGTGAGAAAGAAATGGTGTGGCTGCCCTTCTGCCACCTCCCAACTGTCC
ACCCGGGCCCTGGAAGGGATCTTTGAGGCCACCCACCGCCTGATCTATGGTGCCAAGGATGACTCTGGGC
AGAGATACCTGGCCTGGTCTGGCCACTCTGCCAGAGTGGGTGCTGCCAGGGACATGGCCAGGGCTGGTGT
GTCCATCCCTGAAATCATGCAGGCTGGTGGCTGGACCAATGTGAACATTGTGATGAACTACATCAGAAAC
CTGGACTCTGAGACTGGGGCCATGGTGAGGCTGCTCGAGGATGGGGACGGCAGTGGAGGATCCGGAGCCA
CGAACTTCTCTCTGTTAAAGCAAGCAGGAGACGTGGAAGAAAACCCCGGTCTACCGGTGTGAGCAAGGG
CGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAAC
GGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCTACGAGGGCACCCAGACCGCCAAGCTGA
AGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAA
GGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGG
GAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCG
AGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGAC
CATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAG
AGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCG
TGCAGCTGCCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCAT
CGTGGAACAGTACGAACGCGCCGAGGGCCGCACTCCACCGGCGGCATGGACGAGCTGTACAAGGGATAT
CTCAGCCATGGCTTCCCGCCGGAGGTGGAGGAGCAGGATGATGGCACGCTGCCCATGTCTTGTGCCCAGG
AGAGCGGGATGGACCGTCACCCTGCAGCCTGTGCTTCTGCTAGGATCAATGTGTGA

Fig. 11 (contd.)

CTCGAGGGCGCGCCTACCCGCGGTAGCTTTTCAGTCGAGAAAAAACATTAGATCAGAAGAACAACCTGGC
 AACACTTTCTCAACCTGAGACTTACTTCAAGATGCTCGATCCTGGAGAGGTCTATGATGACCCTATTGAC
 CCAATCGAGTTAGAGGCTGAACCCAGAGGAACCCCATTTGTCCCCAACATCTTGAGGAACTCTGACTACA
 ATCTCAACTCTCCTTTGATAGAAGATCCTGCTAGACTAATGTTAGAATGGTTAAAAACAGGGAATAGACC
 TTATCGGATGACTCTAACAGACAATTGCTCCAGGTCTTTCAGAGTTTTGAAAGATTATTTCAAGAAGGTA
 GATTTGGGTTCTCTCAAGGTGGGCGGAATGGCTGCACAGTCAATGATTTCTCTCTGGTTATATGGTGCCC
 ACTCTGAATCCAACAGGAGCCGGAGATGTATAACAGACTTGGCCCATTTCTATTCCAAGTCGTCCCCCAT
 AGAGAAGCTGTTGAATCTCACGCTAGGAAATAGAGGGCTGAGAATCCCCCAGAGGGAGTGTTAAGTTGC
 CTTGAGAGGGTTGATTATGATAATGCATTTGGAAGGTATCTTGCCAACACGTATTCCTCTTACTTGTCT
 TCCATGTAATCACCTTATACATGAACGCCCTAGACTGGGATGAAGAAAAGACCATCCTAGCATTATGGAA
 AGATTTAACCTCAGTGGACATCGGGAAGGACTTGGTAAAGTTCAAAGACCAAATATGGGGACTGCTGATC
 GTGACAAAGGACTTTGTTTACTCCCAAAGTTCCAATTGTCTTTTTTGACAGAACTACACACTTATGCTAA
 AAGATCTTTTCTTGTCTCGCTTCAACTCCTTAATGGTCTTGTCTCTCCCCCAGAGCCCCGATACTCAGA
 TGACTTGATATCTCAACTATGCCAGCTGTACATTGCTGGGGATCAAGTCTTGTCTATGTGTGGAACTCC
 GGCTATGAAGTCATCAAAATATTGGAGCCATATGTCGTGAATAGTTTAGTCCAGAGAGCAGAAAAGTTTA
 GGCCTCTCATTCATTCCTTGGGAGACTTTCCTGTATTTATAAAAGACAAGGTAAGTCAACTTGAAGAGAC
 GTTCGGTCCCTGTGCAAGAAGGTTCTTTAGGGCTCTGGATCAATTCGACAACATACATGACTTGGTTTTT
 GTGTTTGGCTGTTACAGGCATTGGGGGCACCCATATATAGATTATCGAAAGGGTCTGTCAAACTATATG
 ATCAGGTTACCTTAAAAAATGATAGATAAGTCTACCAGGAGTGCTTAGCAAGCGACCTAGCCAGGAG
 GATCCTTAGATGGGGTTTTGATAAGTACTCCAAGTGGTATCTGGATTCAAGATTCCTAGCCCGAGACCAC
 CCCTTGACTCCTTATATCAAAACCCAAACATGGCCACCCAAACATATTGTAGACTTGGTGGGGGATACAT
 GGCACAAGCTCCCGATCACGCAGATCTTTGAGATTCCTGAATCAATGGATCCGTCAGAAATATTGGATGA
 CAAATCACATTCTTTCACCAGAACGAGACTAGCTTCTTGCTGTGTCAGAAAACCGAGGGGGGCTGTTCT
 AGCGAAAAAGTTATTATCACGGCCCTGTCTAAGCCGCTGTCAATCCCCGAGAGTTTCTGAGGTCTATAG
 ACCTCGGAGGATTGCCAGATGAAGACTTGATAATTGGCCTCAAGCCAAAGGAACGGGAATTGAAGATTGA
 AGGTGATTCCTTTGCTCTAATGTCATGGAATCTAAGATTGTATTTTGTGTCATCACTGAAAACTCTTGGCC
 AACTACATCTTGCCACTTTTTGACGCGCTGACTATGACAGACAACCTGAACAAGGTGTTTAAAAAGCTGA
 TCGACAGGGTCACCGGGCAAGGGCTTTTGGACTATTCAAGGGTCACATATGCATTTACCTGGACTATGA
 AAAGTGGAACAACCATCAAAGATTAGAGTCAACAGAGGATGTATTTTCTGTCCTAGATCAAGTGTTTGGA
 TTGAAGAGAGTGTTTTCTAGAACACACGAGTTTTTTCAAAGGCCTGGATCTATTATTCAGACAGATCAG
 ACCTCATCGGGTTACGGGAGGATCAAATATACTGCTTAGATGCGTCCAACGGCCCAACCTGTTGGAATGG
 CCAGGATGGCGGGCTAGAAGGCTTACGGCAGAAGGGCTGGAGTCTAGTCAGCTTATTGATGATAGATAGA
 GAATCTCAAATCAGGAACACAAGAACC AAAATACTAGCTCAAGGAGACAACCAGGTTTTATGTCCGACAT
 ACATGTTGTGCGCCAGGGCTATCTCAAGAGGGGCTCCTCTATGAATTGGAGAGAATATCAAGGAATGCACT
 TTCGATATACAGAGCCGTGAGGAAGGGGCATCTAAGCTAGGGCTGATCATCAAGAAAGAAGAGACCATG
 TGAGTTATGACTTCCTCATCTATGGAAAAACCCCTTTGTTTAGAGGTAACATATTGGTGCCTGAGTCCA
 AAAGATGGGCCAGAGTCTCTTGCGTCTCTAATGACCAAATAGTCAACCTCGCCAATATAATGTGACAGT
 GTCCACCAATGCGCTAACAGTGGCACAACACTCTCAATCTTTGATCAAACCGATGAGGGATTTTCTGCTC
 ATGTCAGTACAGGCAGTCTTTCACCTGCTATTTAGCCCAATCTTAAAGGGAAGAGTTTACAAGATTC
 TGAGCGCTGAAGGGGAGAGCTTCTCCTAGCCATGTCAAGGATAATCTATCTAGATCCTTCTTTGGGAGG
 GATATCTGGAATGTCCCTCGGAAGATTCCATATACGACAGTTCTCAGACCCTGTCTCTGAAGGGTTATCC
 TTCTGGAGAGAGATCTGGTTAAGCTCCCAAGAGTCTTGATTACGCGTGTGTCAAGAGGCTGGAAACC
 CAGATCTTGAGAGAGAACTCGAGAGCTTCACTCGCCTTCTAGAAGATCCGACCACCTTAAATATCAG
 AGGAGGGGCCAGTCTACCATCTACTCAAGGATGCAATCAGAAAGGCTTTATATGACGAGGTGGACAAG
 GTGGAAAATTCAGAGTTTCGAGAGGCAATCCTGTTGTCCAAGACCCATAGAGATAATTTTATACTCTTCT
 TAATATCTGTTGAGCCTCTGTTTCTCGATTTCTCAGTGAGCTATTCAAGTTCGTCTTTTTTGGGAATCCC
 CGAGTCAATCATGGATTGATACAAACTCCCGAACGATAAGAAGGCAGTTTAGAAAGAGTCTCTCAAAA
 ACTTTAGAAGAATCCTTCTACAACTCAGAGATCCACGGGATTAGTCGGATGACCCAGACACCTCAGAGGG
 TTGGGGGGGTGTGGCCTTGCTCTTCAAGAGGGCAGATCTACTTAGGGAGATCTCTTGGGGAAGAAAAGT
 GGTAGGCACGACAGTTCCTCACCTTCTGAGATGTTGGGATTACTTCCAAGTCCTCTATTTCTTGCACT
 TGTGGAGCAACAGGAGGAGGCAATCCTAGAGTTTCT

Fig. 11 (contd.)

GTATCAGTACTCCCGTCCTTTGATCAGTCATTTTTTTTTCACGAGGCCCCCTAAAGGGATACTTGGGCTCGT
CCACCTCTATGTCGACCCAGCTATTCCATGCATGGGAAAAGTCACTAATGTTTCATGTGGTGAAGAGAGC
TCTATCGTTAAAAGAATCTATAAACTGGTTCATTACTAGAGATTCCAACCTGGCTCAAGCTCTAATTAGG
AACATTATGTCTCTGACAGGCCCTGATTTCCCTCTAGAGGAGGCCCTGTCTTCAAAGGACGGGGTCAG
CCTTGCCATAGGTTCAAGTCTGCCAGATACAGCGAAGGAGGGTATTCTTCTGTCTGCCCCGAACCTCCTCTC
TCATATTTCTGTTAGTACAGACACCATGTCTGATTTGACCCAAGACGGGAAGAACTACGATTTTCATGTTT
CAGCCATTGATGCTTTATGCACAGACATGGACATCAGAGCTGGTACAGAGAGACACAAGGCTAAGAGACT
CTACGTTTTCATTGGCACCTCCGATGCAACAGGTGTGTGAGACCCATTGACGACGTGACCCTGGAGACCTC
TCAGATCTTCGAGTTTCCGGATGTGTGCGAAAAGAATATCCAGAATGGTTTCTGGGGCTGTGCCTCACTTC
CAGAGGCTTCCCGATATCCGTCTGAGACCAGGAGATTTTGAATCTCTAAGCGGTAGAGAAAAGTCTCACC
ATATCGGATCAGCTCAGGGGCTCTTATACTCAATCTTAGTGGCAATTCACGACTCAGGATACAATGATGG
AACCATCTTCCCTGTCAACATATACGGCAAGGTTTCCCCTAGAGACTATTTGAGAGGGCTCGCAAGGGGA
GTATTGATAGGATCCTCGATTTGCTTCTTGACAAGAATGACAAATATCAATATTAATAGACCTCTTGAAT
TGGTCTCAGGGGTAAATCTCATATATTCTCCTGAGGCTAGATAACCATCCCTCCTTGTACATAATGCTCAG
AGAACCGTCTCTTAGAGGAGAGATATTTTCTATCCCTCAGAAAATCCCCGCCGCTTATCCAACCACTATG
AAAGAAGGCAACAGATCAATCTTGTGTTATCTCCAACATGTGCTACGCTATGAGCGAGAGATAATCACGG
CGTCTCCAGAGAATGACTGGCTATGGATCTTTTCAGACTTTAGAAGTGCCAAAATGACGTACCTATCCCT
CATTACTTACCAGTCTCATCTTCTACTCCAGAGGGTTGAGAGAAACCTATCTAAGAGTATGAGAGATAAC
CTGCGACAATTGAGTTCTTTGATGAGGCAGGTGCTGGGCGGGCACGGAGAAGATACCTTAGAGTCAGACG
ACAACATTCAACGACTGCTAAAAGACTCTTTACGAAGGACAAGATGGGTGGATCAAGAGGTGCGCCATGC
AGCTAGAACCATGACTGGAGATTACAGCCCCAACAAGAAGGTGTCCCGTAAGGTAGGATGTTTCAGAATGG
GTCTGCTCTGCTCAACAGGTTGCAGTCTCTACCTCAGCAAACCCGGCCCCCTGTCTCGGAGCTTGACATAA
GGGCCCTCTCTAAGAGGTTCCAGAACCCTTTGATCTCGGGCTTGAGAGTGGTTCAGTGGGCAACCGGTGC
TCATTATAAGCTTAAGCCTATTCTAGATGATCTCAATGTTTTCCCATCTCTCTGCCTTGTAGTTGGGGAC
GGGTGAGGGGGGATATCAAGGGCAGTCCCTCAACATGTTTCCAGATGCCAAGCTTGTGTTCAACAGTCTTT
TAGAGGTGAATGACCTGATGGCTTCCGGAACACATCCACTGCCTCCTTCAGCAATCATGAGGGGAGGAAA
TGATATCGTCTCCAGAGTGATAGATCTTGACTCAATCTGGGAAAACCGTCCGACTTGAGAACTTGCGA
ACCTGGAAATACTTCCAGTCAGTCCAAAAGCAGGTCAACATGTCCTATGACCTCATTATTTGCGATGCAG
AAGTTACTGACATTGCATCTATCAACCGGATCACCTGTAAATGTCCGATTTTGCATTGTCTATAGATGG
ACCACTCTATTTGGTCTTCAAACCTTATGGGACTATGCTAGTAAATCCAACTACAAGGCTATTCAACAC
CTGTCAAGAGCGTTCCCTCGGTCACAGGGTTTATCACCCAAGTAACTTCGTCTTTTTTCATCTGAGCTCT
ACCTCCGATTCTCCAAACGAGGGAAGTTTTTCAGAGATGCTGAGTACTTGACCTCTTCCACCCTTCGAGA
AATGAGCCTTGTGTTATTCAATTGTAGCAGCCCCAAGAGTGAGATGCAGAGAGCTCGTTCCTTGAACAT
CAGGATCTTGTGAGAGGATTTCTGAAGAAATCATATCAAATCCTTACAATGAGATGATCATAACTCTGA
TTGACAGTGATGTAGAATCTTTTCTAGTCCACAAGATGGTTGATGATCTTGAGTTACAGAGGGGAACTCT
GTCTAAAGTGGCTATCATTATAGCCATCATGATAGTTTTCTCCAACAGAGTCTTCAACGTTTCCAAACCC
CTAACTGACCCCTCGTTCTATCCACCGTCTGATCCCAAATCCTGAGGCACTTCAACATATGTTGCAGTA
CTATGATGTATCTATCTACTGCTTTAGGTGACGTCCCTAGCTTCGCAAGACTTCACGACCTGTATAACAG
ACCTATAACTTATTACTTCAGAAAGCAAGTCATTGAGGGGAACGTTTATCTATCTTGGAGTTGGTCCAAC
GACACCTCAGTGTTCAAAGGGTAGCCTGTAATTCTAGCCTGAGTCTGTCATCTCACTGGATCAGGTTGA
TTTACAAGATAGTGAAGACTACCAGACTCGTTGGCAGCATCAAGGATCTATCCAGAGAAGTGGAAGACA
CCTTCATAGGTACAACAGGTGGATCACCTAGAGGATATCAGATCTAGATCATCCCTACTAGACTACAGT
TGCCTGTGAACCGGATACTCCTGGAAGCCTGCCCATGCTAAGACTCTTGTGT

Fig. 11 (contd.)

GATGTATCTTGAAAAAACAAGATCACCGGATACTCCTGGAAGCCTGCCCATGCTAAGACTCTTGTGTGA
TGTATCTTGAAAAAACAAGATCCTAAATCTGAACCTTTGGTTGTTTGATTGTTTTCTCATTTTTGTTG
TTTATTTGTTAAGCGTGGGTGCGCATGGCATCTCCACCTCCTCGCGGTCCGACCTGGGCATCCGAAGGAG
GACGCACGTCCACTCGGATGGCTAAGGGAGAAGGGCGAATTCCAGCACACTGGCGGCCGTTACTAGTGGA
TCCGAGCTCGGTACCAAGCTTAAGTTTAAACCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCA
TCTGTTGTTTGCCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCCTGTCTTTCCTAAT
AAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGA
CAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAG
GCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGG
GTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTTCTT
CCCTTCCTTTCTCGCCACGTTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTC
CGATTTAGTGCTTTACGGCACCTCGACCCCAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCAT
CGCCCTGATAGACGGTTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCA
AACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTTCGGCC
TATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTT
AGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCA
ACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAG
CAACCATAGTCCCGCCCCCTAACTCCGCCCCATCCCGCCCCCTAACTCCGCCCCAGTTCCGCCCCATTCTCGCC
CCATGGCTGACTAATTTTTTTTTTATTTATGCAGAGGCCGAGGCCGCTCTGCCTCTGAGCTATTCCAGAAG
TAGTGAGGAGGCTTTTTTTGGAGGCCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGATATATCCATTTTCG
GATCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCC
GGCCGCTTGGGTGGAGAGGCTATTTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCC
GTGTTCCGGCTGTCAGCGCAGGGGCGCCCGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATG
AACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCTTGCGCAGCTGTGCTCGA
CGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCT
CACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGG
CTACCTGCCCATTTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTA CTGGATGGAAGCCGGTCT
TGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACCTGTTCCGACGGCTCAAG
GCGCGCATGCCCCAGGCGAGGATCTCGTCTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGG
AAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGC
GTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCTCGTGCTTTACGGT
ATCGCCGCTCCCGATTTCGACGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCT
GGGGTTGCAAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTTCGATTCACCGCCGCTTC
TATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCA
TGCTGGAGTTCTTCGCCCAACCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCAT
CACAAATTTACAAATAAAGCATTTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTA
TCTTATCATGTCTGTATAACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTG
TGTGAAATTGTTATCCGCTCACAAATCCACACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGG
TGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTG
TCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTTCGTATTGGGCGCTCTTCCG
CTTCTCGCTCACTGACTCGCTGCGCTCGGTGCTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGC
GGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAG
GCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACA
AAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATAACAGGCGTTTCCCCCTGG
AAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCG
GGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTTCGCTCCAAGC
TGGGCTGTGTGCACGAACCCCCGTTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTC
CAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTAT
GTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTA
TCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCAC
CGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAGCA

Fig. 11 (contd.)

GATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGG
AACGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAA
ATTAAAAATGAAGTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTT
AATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGTCGTG
TAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATAACGCGAGACCCACGCT
CACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAAC
TTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGT
TTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGGTATGGCTTCATTCA
GCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTT
CGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCAT
AATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCT
GAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCCGGCGTCAATACGGGATAATACGCGCCACATAG
CAGAACTTTAAAAGTGCTCATCATTTGAAAACGTTCTTCGGGGCGAAAACCTCTCAAGGATCTTACCGCTG
TTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACCTGATCTTCAGCATCTTTTACTTTACACGCG
TTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAGGGAATAAGGGCGACACGGAAATGTTG
AATACTCATACTCTTCCTTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATAC
ATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTG
ACGTCGACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACCTCTCAGTACAATCTGCTCTGATGCCGCA
TAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAG
CTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCT
TCGCGATGTACGGGCCAGATATACGCGT

Fig. 11 (contd.)

Fig. 12

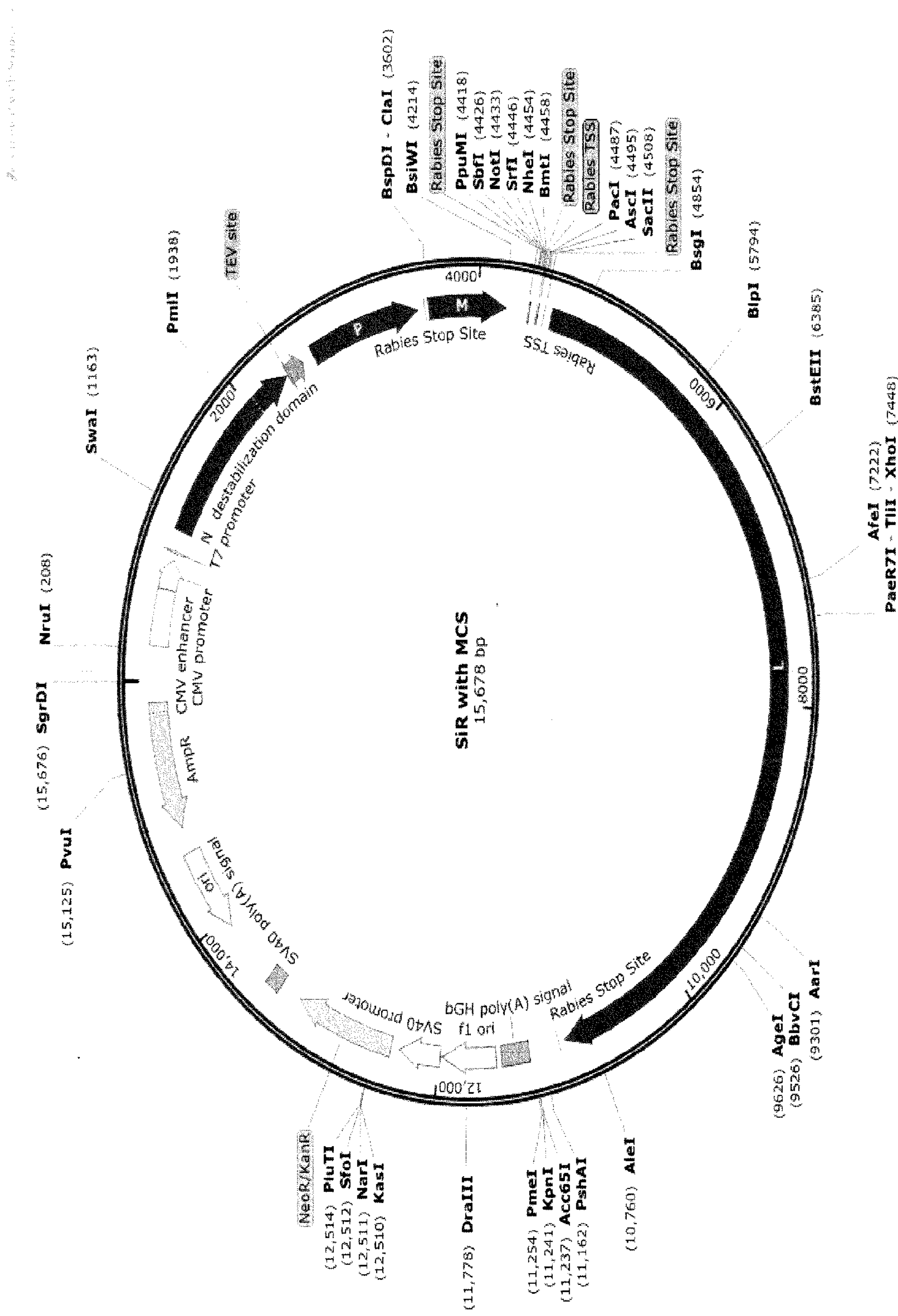


Fig 13

GACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACCTCTCAGTACAATCTGCTCTGATGCCGCATAGTT
 AAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACA
 ACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCG
 ATGTACGGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTC
 ATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCCGCTGGCTGACCG
 CCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCC
 ATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCC
 AAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCCGCTGGCATTATGCCCAGTACATGACCTTA
 TGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTTGGC
 AGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCGAAGTCTCCACCCCATTGACGTCAA
 TGGGAGTTTGTGTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACCTCCGCCCCATTGACG
 CAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCA
 CTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGATTAAGCGTCT
 GATGAGTCCGTGAGGACGAAACCCGGAGTCCCGGGTCacgcttaacaaccagatcaaagaaaaaacagac
 attgtcaattgcaaagcaaaaatgtaacacccctacaatggatgccgacaagattgtattcaaagtcaat
 aatcaggtggtctctttgaagcctgagattatcgtggatcaatatgagtacaagtaccctgccatcaaag
 atttgaaaaagccctgtataaccctaggaaaggctcccgatttaaataaagcatacaagtcagttttgtc
 aggcattgagcgcgcgcaaaacttaactcctgacgatgtatgttctctatttggcagcggcaatgcagtttttt
 gaggggacatgtccggaagactggaccagctatggaattgtgattgcacgaaaaggagataagatcaccc
 caggttctctggtggagataaaaacgtactgatgtagaagggaattgggctctgacaggaggcatggaact
 gacaagagacccccactgtccctgagcatgcgtccttagtcggtcttctcttgagtctgtataggttgagc
 aaaatatccgggcaaaacactggtaactataagacaaacattgcagacaggatagagcagatttttgaga
 cagccccttttgttaaaatcgtggaacaccatactctaatagacaactcacaataatgtgtgctaattggag
 tactataccaaacttcagatttttggccggaacctatgacatgtttttctcccgattgagcatctatat
 tcagcaatcagagtgggcacagttgtcactgcttatgaagactgttcaggactggtatcatttactgggt
 tcataaaacaaatcaatctcaccgctagagaggcaataactataatttcttcacaagaactttgaggaaga
 gataagaagaatgtttgagccagggcaggagacagctgttcctcactcttatttcatccacttccgttca
 ctaggcttgagtgggaaatctccttattcatcaaatactgtgttggtcacgtgttcaatctcattcaacttg
 taggatgctatatgggtcaagtcagatccctaaatgcaacgggttattgctgcatgtgctcctcatgaaat
 gtctgttctagggggctatctgggagaggaattcttcgggaaaggacatttgaaagaagattcttcaga
 gatgagaaagaacttcaagaatacagaggcggctgaactgacaaagactgacgtagcactggcagatgatg
 gaactgtcaactctgacgacgaggactacttttcagggtgaaaccagaagtccggaggctgtttatactcg
 aatcatgatgaatggaggtcgactaaagagatctcacatacggagatatgtctcagtcagttccaatcat
 caagcccgtccaaactcattcgccgagtttctaaacaagacatatctcagtgactcaGGTTCCGGAgaga
 acctctacttccaatcgGGATCCGGTAGCCATGGCTTCCCGCCGGAGGTGGAGGAGCAGGATGATGGCAC
 GCTGCCCATGTCTTGTGCCAGGAGAGCGGGATGGACCGTCACCCTGCAGCCTGTGCTTCTGCTAGGATC
 AATGTGtaagaagttgaataacaaaatgccggaaatctacggattgtgtatatccatcatgaaaaaaact
 aacacccctcctttcgaaccatcccaaacatgagcaagatctttgtcaatcctagtgtattagagccgg
 tctggccgatcttgagatggctgaagaaactgttgatctgatcaatagaaatatcgaagacaatcaggct
 catctccaagggggaacccatagagggtggacaatctccctgaggatatggggcgacttcacctggatgatg
 gaaaatcgcccaaccatggtgagatagccaagggtgggagaaggcaagtatcgagaggactttcagatgga
 tgaaggagaggatcctagcttctgttccagtcatacctggaaaatgttgagtcctaatagtcagacaa
 atgaggtcaggagagagattttctcaagatatggtcacagaccgtagaagagattatatcctatgtcgcgg
 tcaactttcccaaccctccaggaaagtcttcagaggataaatcaaccagactactggccgagagctcaa
 gaaggagacaaacaccactccttctcagagagaaagccaatcatcgaaagccaggatggcggctcaaatt
 gcttctggccctccagcccttgaatggctcggctaccaatgaagaggatgatctatcagtgagggtgaga
 tcgctcaccagattgcagaaagtttctccaaaaaatataagtttccctctcgcacccagggatactctt
 gtataattttgagcaattgaaaatgaaccttgatgatatagttaaagaggcaaaaaatgtaccaggtgtg
 acccgtttagcccatgacgggtccaaactccccctaagatgtgtactgggatgggtcgctttggccaact
 ctaagaaattccagttgttagtcgaatccgacaag

Fig. 13 (contd.)

ctgagtaaaatcatgcaagatgacttgaatcgctatacatcttgctaaccgaacctctccccctcagtc
 tctagacaataaaatccgagatgtcccaaagtcaacatgaaaaaacaggcaacaccactgataaaTCGa
 tgaacctcctacgtaagatagtgaaaaaaccgcagggacgaggacactcaaaaatcctctcccgcgtcagc
 cctctggatgacgatgacttgtggcttccacccccctgaatacgtcccgctgaaagaacttacaggcaag
 aagaacatgaggaacttttgtatcaacggaagggttaaagtgtgtagcccgaatgggttactcgttcagga
 tcctgcggcacattctgaaatcattcgacgagatatattctgggaatcataggatgatcgggttagtcaa
 agtgggttattggactggcttctgtcaggatctccagtccttgagggcctgaactgggtatacaaattgagg
 agaacctttatcttccagtgaggctgattccaggggccccctcttgaaggggaggagttggaatactctcagg
 agatcacttgggatgatgatactgagttcgtcggattgcaaataagagtgattgcaaaacagtggtcatat
 ccagggcagagtctggtgtatcaacatgaacccgagagcatgtcaactatgggtctgacatgtctcttcag
 acacaaagggtccgaagaggacaaagattcctctctgcttctagaataacgtacgtcagattatatcccg
 aaatttatcacttgtttacctctggaggagagaaacatatggggtcaactccaacccttgggagcaatata
 acaaaaaacatgttatggtgccattaaaccgctgcatttcacaaagtcaagttgattacctttacattt
 tgatcctcttggatgtgaaaaaaactattaacatccctcaaaggaCCTGCAGGTACGCGGCCGCTACGCC
 CGGGCTACGCTAGCatgaaaaaaactaacacccccctccTTAATTAATACGGCGCGCctaccgcggttagct
 tttcagtcgagaaaaaaacattagatcagaagaacaactggcaacactttctcaacctgagacttacttc
 aagatgctcgatcctggagaggtctatgatgaccctattgacccaatcgagtttagagggtgaacccagag
 gaacccccattgtccccaacatcttgaggaaactctgactacaatctcaactctcctttgatagaagatcc
 tgctagactaatgttagaatgggttaaaaaacagggaatagaccttatcggtgactctaacagacaattgc
 tccaggtctttcagagttttgaaagattatttcaagaaggtagatttgggttctctcaaggtgggcgga
 tggctgcacagtcaatgatttctctctggttatatgggtgccactctgaatccaacaggagccggagatg
 tataacagacttggccccatttctatttccaagtcgtcccccatagagaagctggtgaatctcacgctagga
 aatagagggctgagaatccccccagagggagtggttaagttgccttgagaggggttgattatgataatgcat
 ttggaagggtatcttgccaacacgtattcctcttacttgttcttccatgtaatcaccttatacatgaacgc
 cctagactgggatgaagaaaagaccatcctagcattatggaaagatttaacctcagtggaacatcggaag
 gacttggtaagttcaaagaccacaaatatggggactgctgatcgtgacaaaggactttgtttactcccaa
 gttccaattgtctttttgacagaaactacacacttatgctaaaagatcttttcttgtctcgttcaactc
 cttaatggtcttctctctccccagagccccgatactcagatgacttgatatctcaactatgccagctg
 tacattgctggggatcaagtcttgtctatgtgtggaaactccggctatgaagtcacaaaatattggagc
 catatgtcgtgaatagtttagtccagagagcagaaaaagtttaggcctctcattcattccttgggagactt
 tcctgtattttataaaagacaaggtaagtcaacttgaagagacgttcgggtccctgtgcaagaagggtcttt
 agggctctggatcaattcgacaacatacatgacttgggtttttgtgtttggctgttacaggcattgggggc
 acccatatatagattatcgaaagggtctgtcaaaactatatgatcaggttcaccttaaaaaaatgataga
 taagtcctaccaggagtgttagcaagcgacctagccaggaggatccttagatgggggttttgataagtac
 tccaagtggtatctggattcaagattcctagcccagagaccaccccttgactccttataatcaaaacccaaa
 catggccacccaaacatattgtagacttgggtgggggatacatggcacaagctcccgatcacgcagatctt
 tgagattcctgaatcaatggatccgtcagaaatattggatgacaaatcacattctttcaccagaacgaga
 ctagcttcttggctgtcagaaaaccgaggggggctgttcttagcgaaaaagttattatcacggccctgt
 ctaagccgctgtcaatccccgagagtttctgaggtctatagacctcggaggattgccagatgaagactt
 gataattggcctcaagccaaaggaacgggaattgaagattgaagggtcgattctttgctctaattgtcatgg
 aatctaagattgtattttgtcatcactgaaaaactcttggccaactacatcttggcactttttgacgcgc
 tgactatgacagacaacctgaacaagggtgtttaaaaagctgatcgacaggggtcacccgggcaagggtcttt
 ggactattcaagggtcacatatgcatttcacctggactatgaaaagtggacaacacatcaagatttagag
 tcaacagaggatgtattttctgtcctagatcaagtgtttggattgaagagagtggtttctagaacacacg
 agttttttcaaaaaggcctggatctattattcagacagatcagacctcatcgggttacgggaggatcaaat
 atactgcttagatgcttccaacggcccaacctgttggaaatggccaggatggcgggctagaaggcttacgg
 cagaagggttgaggtctagtcagcttattgatgatagatagagaatctcaaatcaggaacacaagaacca
 aaatactagctcaaggagacaaccagggttttatgtccgacatacatgttgtcgccagggttatctcaaga
 ggggtcctctatgaattggagagaatatcaaggaatgcactttcgatatacagagccgtcgaggaagg
 gcatctaagctagggtgatcatcaagaaagaagagaccatgtgtagttatgacttctcatctatggaa
 aaaccccccttgttttagaggtaacatattgggtgcctgagtcacaaagatgggcccagagtcctcttgcgtctc

Fig. 13 (contd.)

taatgaccaaatagtcacacctcgccaatataatgtcgacagtggtccaccaatgctgctaacagtggcaca
 cactctcaatctttgatcaaaccgatgagggattttctgctcatgtcagtcaggtacagggcagtcctttcactacc
 tgctattttagcccaatcttaaagggaagagtttacaagattctgagcgctgaaggggagagctttctcct
 agccatgtcaaggataatctatctagatccttctttgggagggatctctggaatgtccctcggaagattc
 catatacgacagttctcagaccctgtctctgaaggggttatccttctggagagagatctgggttaagctccc
 aagagtcctggattcacgcgttggtgtcaagaggctggaaaccagatcttgagagagagaacactcgagag
 cttcactcgccttctagaagatccgaccaccttaaatatcagaggagggggccagtcctaccattctactc
 aaggatgcaatcagaaaggctttatatgacgaggtggacaagggtggaaaattcagagtttcgagaggcaa
 tcctgttggtccaagaccatagagataattttataactcttcttaatatctgttgagcctctgtttcctcg
 atttctcagtgagctattcagttcgtcttttttgggaatccccgagtcattcattggattgatacaaaac
 tcccgaacgataagaaggcagtttagaaagagttctctcaaaaactttagaagaatccttctacaactcag
 agatccacgggattagtcggatgaccagacacctcagagggttgggggggtgtggccttgctcttcaga
 gagggcagatctacttagggagatctcttggggaagaaaagtggtaggcacgacagttcctcacccttct
 gagatggttgaggattacttcccaagtcctctattttcttgacttggtgagcaacaggaggaggcaatccta
 gagtttctgtatcagtaactccgctcctttgatcagtcatttttttcacgaggccccctaaagggatactt
 gggctcgtccacctctatgtcgacccagctattccatgcatgggaaaaagtactaatgttcatgtggtg
 aagagagctctatcgtttaaagaatctataaactgggttcattactagagattccaacttgggtcaagctc
 taattaggaacattatgtctctgacaggccctgatttccctctagaggaggccccctgtcttcaaaaggac
 ggggtcagccttgcataggttcaagtctgccagatacagcgaaggagggtatttcttctgtctgccgaac
 ctctctctcatattttctgttagtacagacacccatgtctgatttgacccaagacgggaagaactacgatt
 tcatgttccagccattgatgctttatgcacagacatggacatcagagctggtacagagagacacaaggct
 aagagactctacgtttcattggcacctccgatgcaacagggtgtgtgagacccattgacgacgtgacctg
 gagacctctcagatcttcgagtttccggatgtgtcgaaaagaatatccagaatgggttctggggctgtgc
 ctacttccagaggcttcccgatatccgtctgagaccaggagattttgaatctctaaagcggtagagaaaa
 gtctcaccatatcggtcagctcaggggctcttataactcaatcttagtggaattcacgactcaggatac
 aatgatggaaccatcttccctgtcaacatatacggcaagggttcccttagagactatttgagagggtcgtc
 caaggggaggtattgataggatcctcgatttgcttcttgacaagaatgacaaatatcaatattaatagacc
 tcttgaattgggtctcaggggtaatctcatatattctcctgaggctagataaccatccctccttgtaata
 atgctcagagaaccgtctcttagaggagagatattttctatccctcagaaaatccccgcccgttatccaa
 ccactatgaaagaaggcaacagatcaatcttgtgttatctccaacatgtgctacgctatgagcgagagat
 aatcacggcgtctccagagaatgactggctatggatcttttcagactttagaagtgccaaaatgacgtac
 ctatccctcattacttaccagttctcatcttctactccagagggttgagagaaacctatctaagagtatga
 gagataacctgcgacaattgagttctttgatgaggcaggtgctgggcgggcacggagaagataacctaga
 gtcagacgacaacattcaacgactgctaaaagactctttacgaaggacaagatgggtggatcaagagggtg
 cgccatgcagctagaaccatgactggagattacagccccaacaagaagggtgtcccgtaaggtaggatgtt
 cagaatgggtctgctctgctcaacagggttgagctctctacctcagcaaacccggccccctgtctcggagct
 tgacataagggccctctctaagagggtccagaacctttgatctcgggcttgagagtggttcagtgggca
 accggtgctcattataagcttaagcctattctagatgatctcaatgttttcccatctctctgccttgtag
 ttggggacgggtcaggggggatatacaagggcagtcctcaacatgtttccagatgccaaagcttggttcaa
 cagtccttttagagggtgaatgacctgatggcttccggaacacatccactgcctccttcagcaatcatgagg
 ggaggaaatgatatcgtctccagagtgatagatcttgactcaatctgggaaaaaccgtccgacttgagaa
 acttggcaacctggaaatacttccagtcagtcacaaaagcaggtcaacatgtcctatgacctcattatttg
 cgatgcagaagttactgacattgcatctatcaaccggatcacctgttaatgtccgatttttgcatgtct
 atagatggaccactctatttggtcttcaaaacttatgggactatgctagtaaatccaaactacaaggcta
 ttcaacacctgtcaagagcgttccctcgggtcacagggtttatcacccaagtaacttcgtctttttcctc
 tgagctctacctccgattctccaaacgaggggaagtttttcagagatgctgagtaacttgacctcttccacc
 ctctcgagaaatgagccttggtgttattcaattgtagcagccccaagagtgagatgcagagagctcgttcct
 tgaactatcaggatcttgtgagaggatttctgaagaaatcatatcaaatccttacaatgagatgatcat
 aactctgattgacagtgatgtagaatcttttctagtcacaaagatgggttgatgatcttgagttacagagg
 ggaactctgtctaaagtggctatcattatagccatcatgatagttttctccaacagagtccttcaacggtt
 ccaaacccctaactgacctctgttctatccacgtctgatcccaaatcctgaggcacttcaacatatg

Fig. 13 (contd.)

ttgcagtactatgatgtatctatctactgcttttaggtgacgtccctagcttcgcaagacttcacgacctg
tataacagacctataaacttattacttcagaaagcaagtcattcgagggaacgtttatctatcttggagtt
ggtccaacgacacctcagtggttcaaaagggtagcctgtaattctagcctgagtcgtcatctcactggat
caggttgattttacaagatagtgaaagactaccagactcgttggcagcatcaaggatctatccagagaagtg
gaaagacaccttcataaggtacaacaggtggatcaccctagaggatatcagatctagatcatccctactag
actacagttgcctgtgaaccggatactcctggaagcctgcccattgctaagactccttggtgtgatgtatctt
gaaaaaaacaagatcctaatactgaacctttggttggttgattgtttttctcatttttggttggttatttg
ttaagcgtGGGTCGGCATGGCATCTCCACCTCCTCGCGGTCCGACCTGGGCAICCGAAGGAGGACGCACG
TCCACTCGGATGGCTAAGGGAGAAGGGCGAATTCCAGCACACTGGCGGCCGTTACTAGTGGATCCGAGCT
CGGTACCAAGCTTAAGTTTAAACCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGT
TTGCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGIGCCACTCCCACTGTCCTTTTCTAATAAAATGAG
GAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGG
GGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAG
AACCAGCTGGGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGCGGGTGTGGTG
GTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTTCTTCCCTTCCT
TTCTCGCCACGTTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAG
TGCTTTACGGCACCTCGACCCCCAAAAAACTTGATTAGGGTGATGGTTTACGTAGTGGGCCATCGCCCTGA
TAGACGGTTTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAACTGGAA
CAACACTCAACCCTATCTCGGTCTATTCTTTTGATTATAAGGGATTTTGCCGATTTTCGGCCTATTGGTT
AAAAAATGAGCTGATTTAACAAAAAATTAACGCGAATTAATTCTGTGGAATGTGTCTCAGTTAGGGTGTG
GAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTG
TGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATA
GTCCCCGCCCTAACTCCGCCCATCCCCGCCCTAACTCCGCCCAGTTCCGCCCATTTCTCCGCCCATGGCT
GACTAATTTTTTTTATTTATGAGAGGCCGAGGCCGCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGG
AGGCTTTTTTTGGAGGCCCTAGGCTTTTGCAAAAAGCTCCCGGGAGCTTGTATATCCATTTTCGGATCTGAT
CAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTT
GGGTGGAGAGGCTATTCCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCG
GCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAG
GACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCCTTGCGCAGCTGTGCTCGACGTTGTCA
CTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGC
TCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGC
CCATTGACCAACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATC
AGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAAGTGTTCGCCAGGCTCAAGGCGCGCAT
GCCCCGACGGCGAGGATCTCGTCTGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGC
CGCTTTTCTGGATTCTGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTA
CCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCTCGTGCTTTACGGTATCGCCGC
TCCCGATTTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCTG
AAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTGATTCCACCGCCGCCTTCTATGAAAG
GTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAG
TTCTTCGCCCACCCCAACTTGTTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATT
TCACAAATAAAGCATTTTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAACTCATCAATGTATCTTATCA
TGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAAT
TGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAAT
GAGTGAGCTAACTCACATTAAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTGCTGCCA
GCTGCATTAAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTTCGCTATTGGGCGCTCTTCCGCTTCTCTG
CTCACTGACTCGCTGCGCTCGGTCTGCTTGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATAC
GGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAA
CCGTAAAAAGGCCGCGTTGCTGGCGTTTTTTCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGA
CGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCC
TCGTGCGCTCTCCTGTTCCGACCTTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGT
GGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCA

Fig. 13 (contd.)

GTTCGGTGTAGGTCGTTTCGCTCCAAGCTGGGGCTGTGTGCACGAACCCCCCGTTTCAGCCCGACCGCTGCGC
CTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACT
GGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACG
GCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGG
TAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTggtTTTTTTGTTTGCAAGCAGCAGATTACG
CGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAA
ACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAA
ATGAAGTTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGT
GAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGTGCTGTAGATAA
CTACGATACGGGAGGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGC
TCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCCGAGAAGTGGTCCTGCAACTTTATCC
GCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCA
ACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGGTATGGCTTCATTTCAGCTCCGG
TTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCT
CCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTC
TTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATA
GTGTATGCGGCGACCGAGTTGCTCTTGCCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACT
TTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACCTCTCAAGGATCTTACCGCTGTTGAGAT
CCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTTCACCAGCGTTTCTGG
GTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTC
ATACTCTTCCTTTTTTCAATATTATTGAAGCATTATCAGGGTTATTGTCTCATGAGCGGATACATATTTG
AATGTATTTAGAAAAATAAACAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTC

Fig. 14

5' –

ACGCUUAACAAAUAAACAACAAAAAUGAGAAAAACAAUCAAACAACCAAAGGUUCAGAUUUAGGAUCUUG
 UUUUUUUUCAAGAUACAUCACACAAGAGUCUUAGCAUGGGGCAGGCUUCCAGGAGUAUCCGGUUCACAGGCA
 ACUGUAGUCUAGUAGGGAUGAUCUAGAUCUGAUAUCCUCUAGGGUGAUCCACCUGUUGUACCUAUGAAGG
 UGUCUUUCCACUUCUCUGGAUAGAUCUUGAUGCUGCCAACGAGUCUGGUAGUCUUCACUAUCUUGUAAA
 UCAACCUGAUCCAGUGAGAUGACAGACUCAGGCUAGAAUUACAGGCUACCCUUUUGAACACUGAGGGUGUC
 GUUGGACCAACUCCAAGAUAGAUAAACGUUCCCUAGAAUGACUUGCUUUCUGAAGUAAUAAGUUUAUAGGU
 CUGUUUAUACAGGUCGUGAAGUCUUGCGAAGCUAGGGACGUCACCUAAAGCAGUAGAUAGAUACAUCAUAG
 UACUGCAACAUAUGUUGAAGUGCCUCAGGAUUUUGGGAUCAGACGGUGGAUAGAACGAGGGGUCAGUUAG
 GGGUUUGGAAACGUUGAAGACUCUGUUGGAGAAAACUAUCAUGAUGGCUAUAAUGAUAGCCACUUUAGAC
 AGAGUUCUUUCUGUAACUCAAGAUCUAACCAUCUUGUGGACUAGAAAAGAUUCUACAUCACUGUCA
 UCAGAGUUUAUGAUCAUCUCAUUGUAAGGAUUUGAUUUGAUUUUCUUCAGGAAAUCCUCUCACAAGAUCUG
 AUAGUUCAAGGAACGAGCUCUCUGCAUCUCACUCUUGGGGCGUCUACAAUUGAAUAACACAAGGCUCAU
 UCUCGAAGGGUGGAAGAGGUCAAGUACUCAGCAUCUCUGAAAAACUUCUUUGGAGAAUCGGAGGU
 AGAGCUCAGAUAAAAAGACGAAGUUACUUGGGUGAUAAACCCUGUGACCGAGGGGAACGCUCUUGACAG
 GUGUUGAAUAGCCUUGUAGUUUGGAUUUACUAGCAUAGUCCCAUAAGUUUUGAAGACCAAUAGAGUGGU
 CCAUCUAUAGACAAUGCAAAAUCCGACAUUAACAGGGUGAUCCGGUUGAUAGAUGCAAUGUCAGUAACUU
 CUGCAUCGCAAUAAUGAGGUCAUAGGACAUGUUGACCUGCUUUUGGACUGACUGGAAGUAUUUCCAGGU
 UGCCAAGUUUCUCAAGUCGGACGGUUUUUCCAGAUUGAGUCAAGAUCUAUCACUCUGGAGACGAUAUCA
 UUUCUCCCCUCAUGAUUGCUGAAGGAGGCAGUGGAUGUGUUCGGAAGCCAUCAGGUCAUUCACCUCUA
 AAAGACUGUUGAACACAAGCUUGGCAUCUGGAAACAUGUUGAGGACUGCCCUUGAUAUCCCCCUGACCC
 GUCCCCAACUACAAGGCAGAGAGAUGGGAAAACAUUGAGAUCAUCUAGAAUAGGCUUAAGCUUAUAUGA
 GCACCGGUUGCCCACUGAACCAUCUCUAAAGCCGAGAUCAAAGGGUUCUGGAACCUCUAGAGAGGGCCC
 UUAUGUCAAGCUCCGAGACAGGGGCCGGGUUUGCUGAGGUAGAGACUGCAACCUGUUGAGCAGAGCAGAC
 CCAUUCUGAACAUCCUACCUUACGGGACACCUUCUUGUUGGGGCGUGAAUCUCCAGUCAUGGUUCUAGCU
 GCAUGGCGCACCUCUUGAUCCACCCAUCUUGUCCUUCGUAAAGAGUCUUUUAGCAGUCGUUGAAUGUUGU
 CGUCUGACUCUAAGGUUAUCUUCUCCGUGCCCCGCCAGCACCUGCCUCAUCAAAGAACUCAAUUGUCGCAG
 GUUAUCUCUCAUACUCUAGAUAGGUUUCUCUCAACCCUCUGGAGUAGAAGAUGAGACUGGUAAGUAAUG
 AGGGAUAGGUACGUCAUUUUGGCACUUCUAAAGUCUGAAAAGAUCCAUAAGCCAGUCAUUCUCUGGAGACG
 CCGUGAUUAUCUCUCGCUCAUAGCGUAGCACAUGUUGGAGAUAAACACAAGAUUGAUCUGUUGCCUUCUUU
 CAUAGUGGUUGGAUAAGCGGCGGGGAUUUUCUGAGGGAUAGAAAUAUCUCUCCUCUAAGAGACGGUUCU
 CUGAGCAUUAUGUACAAGGAGGGAUGGUUAUCUAGCCUCAGGAGAAUAUAUGAGAUUACCCUGAGACCA
 AUUCAAGAGGUCUAUUAAUAUUGAUUUUGUCAUUCUUGUCAAGAAGCAAUUCGAGGAUCCUAUCAUAC
 UCCCCUUGCGAGCCCUCUCAAAUAGUCUCUAGGGGAAACCUUGCCGUUAUUGUUGACAGGGAAGAUGGUU
 CCAUCAUUGUAUCCUGAGUCGUGAAUUGCCACUAAGAUAUGAUUAAGAGCCCCUGAGCUGAUCCGAUAU
 GGUGAGACUUUUUCUACCGCUUAGAGAUUCAAUAUCUCCUGGUCUCAGACGGAUUUCGGGAAGCCUCUG
 GAAGUGAGGCACAGCCCCAGAAACCAUUCUGGAUAUUCUUUUCGACACAUCCGGAAACUCGAAGAUCUGA
 GAGGUCUCCAGGGUCACGUCGUCAAUGGGGUCUCACACACCUGUUGCAUCGGAGGUGCCAAUGAAACGUAG
 AGUCUCUAGCCUUGUGUCUCUCUGUACCAGCUCUGAUGUCCAUGUCUGUGCAUAAAGCAUCAUAGGCUG
 GAACAUGAAAUCGUAGUUCUUCGUCUUGGGUCAAAUCAGACAUGGUGUCUGUACUAAACAGAAAUAUGA
 GAGAGGAGGUUCGGGCAGACAGAAGAAUACCCUCCUUCGUGUAUCUGGCAGACUUGAACCUAUGCAAGG
 CUGACCCCGUCCUUUUGAAGACAGGGGCCUCCUCUAGAGGGGAAUACAGGGCCUGUCAGAGACAUAAUGUU
 CCUAAUAGAGCUUGAGCCAAGUUGGAUUCUCUAGUAAUGAACCAAGUUUAUAGAUUCUUUUAAACGAUAGA
 GCUCUCUACCAUCAUGAACAUUAGUGACUUUUUCCCAUGCAUGGAUAGCUGGGUCGACAUAGAGGUGG
 ACGAGCCCAAGUAUCCCUUUAGGGGGCCUCGUGAAAAAAUAGACUGAUCAAAGGACGGGAGUACUGAUAC
 AGAAACUCUAGGAUUGCCUCCUCCUGUUGCUCACAAAGUGCAAGAAUAGAGGACUUGGGAAGUAAUCCC
 AACAUUCAGAAAGGGUGAGGAACUGUCGUGCCUACCACUUUUCUUCCCCAAGAGAUUCUCCUAAGUAGAU
 CUGCCUCUCUGAAGAGCAAGGCCACACCCCCCAACCCUCUGAGGUGUCUGGGUCAUCCGACUAAUCCC
 GUGGAUCUCUGAGUUGUAGAAGGAUUCUUCUAAAGUUUUGAGAGACUCUUCUAAACUGCCUUCUUAUC
 GUUCGGGAGUUUUGUAUCAAUCCAUGAUUGACUCGGGGAUUCCCAAAAAAGACGAACUGAAUAGCUCAC
 UGAGAAAUCGAGGAAACAGAGGCUCAACAGAUUU

Fig. 14 (contd.)

AAGAAGAGUAUAAAAUUAUCUCUAUGGGUCUUGGACAACAGGAUUGCCUCUCGAAACUCUGAAUUUUCCA
 CCUUGUCCACCUCGUCAUUAAAGCCUUCUGAUUGCAUCCUUGAGUAGAAUGGUAGGACUGGCCCCUCC
 UCUGAUUUUAAGGUGGUCGGAUCUUCUAGAAGGCGAGUGAAGCUCUCGAGUGUUCUCUCUCCAAGAUCU
 GGGUUUCCAGCCUCUUGACACAACGCGUGAAUCCAGGACUCUUGGGAGCUUAACCAGAUUCUCUCUCCAGA
 AGGAUAACCCUUCAGAGACAGGGUCUGAGAACUGUCGUUAUGGAAUCUUCCGAGGGACAUUCCAGAUAU
 CCCUCCCAAAGAAGGAUCUAGAUAGAUUAUCCUUGACAUGGCUAGGAGAAAGCUCUCCCCUUCAGCGCUC
 AGAAUCUUGUAAACUCUUCUCCUUAAGAUUGGGCUAAAUAGCAGGUAGUGAAAGACUGCCUGUACUGACA
 UGAGCAGAAAAUCCCUCAUCGGUUGAUCAAAGAUUGAGAGUGUUGGCCACUGUUAGCGCAUUGGUGGA
 CACUGUCGACAUUAUUAUUGGCGAGGUUGACUUAUUGGUCAUUAGAGACGCAAGAGACUCUGGCCCAUCUU
 UUGGACUCAGGCACCAAUAUGUUACCUCUAAACAAAGGGGUUUUCCAUAGAUGAGGAAGUCAUAACUAC
 ACAUGGUCUCUUCUUCUUGAUUGAUCAGCCCUAGCUUAGAUGCCCCUCCUUGACGGCUCUGUAUAUCGA
 AAGUGCAUUCUUGAUUUCUCUCCAAUUCAUAGAGGAGCCCCUCUUGAGAUAGCCCUGGCGACAACAUG
 UAUGUCGGACAUAUAAACCUGGUUGUCUCCUUGAGCUAGUAUUUUGGUUCUUGUGUCCUGAUUUUGAGAUU
 CUCUAUCUAUCAUAUAAGCUGACUAGACUCCAGCCCUUCUGCCGUAAGCCUUCUAGCCCGCCAUCUUG
 GCCAUUCCAACAGGUUGGGCCGUUGGACGCAUCUAAGCAGUAUAUUUGAUCCUCCCGUAACCCGAUGAGG
 UCUGAUCUGUCUGAAUAAUAGAUCCAGGCCUUUUGAAAAACUCGUGUGUUCUAGAAAAACACUCUCUUC
 AUCCAAACACUUGAUCUAGGACAGAAAAUACAUCUUCUGUUGACUCUAAUCUUUGAUGGUUGUCCACUU
 UUCAUAGUCCAGGUGAAAUGCAUAUGUGACCCUUGAAUAGUCCAAAAGCCCUUGCCCGGUGACCCUGUCG
 AUCAGCUUUUUAACACCUUGUUCAGGUUGUCUGUCAUAGUCAGCGCGUCAAAAAGUGGCAAGAUGUAGU
 UGGCCAAGAGUUUUUCAGUGAUGACAAAUAACAUCUUAAGAUUCCAUGACAUAUAGAGCAAAGAAUCGACC
 UUCAUUCUCAAUUCUCCGUUCCUUGGCUUGAGGCCAAUUAUCAAGUCUUCUAGGCAAUCCUCCGAGG
 UCUAUAGACCUCAGAAACUCUCGGGGAUUGACAGGCGGCUUAGACAGGGCCGUGAUAAUAACUUUUUCGC
 UAGGAACAGGCCCCCCUCGGUUUUCUGACAGCCAAGAAGCUAGUCUCGUUUCUGGUGAAAGAAUGUGAUUU
 GUCAUCCAUAUUUCUGACGGAUCCAUAUUGAUUCAGGAAUCUCAAGAUUCGCGUGAUCGGGAGCUUGUGC
 CAUGUAUCCCCACCAAGUCUACAUAUGUUUGGGUGGCAUGUUUGGGUUUUGAUUAUAGGAGUCAAGG
 GGUGGUCUCGGGCUAGGAAUCUUGAAUCCAGAUACCACUUGGAGUACUUAUCAAAACCCCAUCUAAGGAU
 CCUCCUGGCUAGGUCGCUUGCUAAGCACUCCUGGUAGGACUUAUCUAUCAUUUUUUUAAGGUGAACCUGA
 UCAUAUAGUUUUGACAGACCCUUCGAUAAUCUAUAUAUGGGUGCCCCCAAUGCCUGUAACAGCCAAACA
 CAAAAACCAAGUCAUGUAUGUUGUCGAAUUGAUCCAGAGCCCUAAAGAACCUCUUGCACAGGGACCGAA
 CGUCUCUUCUAAAGUUGACUUAACCUUGUCUUUUUAUAAUACAGGAAAGUCUCCCAAGGAUUGAAUGAGAGGC
 CUAAACUUUUCUGCUCUCUGGACUAAACUAUUCACGACAUUUGGCUCCAUAUUUUGAUGACUUCAUAGC
 CGGAGUUUCCACACAUAGACAAGACUUGAUCCCCAGCAAUGUACAGCUGGCAUAGUUGAGAUUAUAGUC
 AUCUGAGUAUCGGGGCUCUGGGGGGAGAGAGCAAGACCAUUAAGGAGUUGAAGCGAGACAAGAAAAGAUUCU
 UUUAGCAUAAGUGUGUAGUUUCUGUCAAAAAGACAAUUGGAACUUUGGGAGUAAACAAAGUCCUUUGUCA
 CGAUCAGCAGUCCCCAUUUUUGGUCUUUGAACUUUACCAAGUCCUUCUCCGAUGUCCACUGAGGUUAAAUC
 UUUCCAUAUAGCUAGGAUGGUCUUUUUCUUAUCCAGUCUAGGGCGUUAUGUAUAAGGUGAUUACAUGG
 AAGAACAAGUAAGAGGAAUACGUGUUGGCAAGAUACCUUCCAAAUGCAUUAUCAUAAUCAACCCUCUCAA
 GGCAACUUAACACUCCUCUGGGGGGAUUCUCAGCCUCUAUUUCCUAGCGUGAGAUUCAACAGCUUCUC
 UAUGGGGGGACGACUUGGAAUAGAAAUGGGCCAAGUCUGUUAUACAUCUCCGGCUCCUGUUGGAUUCAGAG
 UGGGCACCAUAUAACCAGAGAGAAAUCAUUGACUGUGCAGCCAUUCGCCCCACCUUGAGAGAACCCAAAU
 CUACCUUCUUGAAUAUUCUUUCAAACUCUGAAAGACCUGGAGCAAUUGUCUGUUAGAGUCAUCCGAUA
 AGGUCUAUUCUCCUGUUUUUAACCAUUCUAACAUUAGUCUAGCAGGAUCUUCUAUCAAGGAGAGUUGAGA
 UUGUAGUCAGAGUCCUCAAGAUGUUGGGGACAAUGGGGGUUCUUGGUGUCCUUAACUCGAUUG
 GGUCAUAGGGUCAUCAUAGACCUCUCCAGGAUCGAGCAUCUUGAAGUAAGUCUCAGGUUGAGAAAGUGU
 UGCCAGUUGUUCUUCUGAUCUAAUGUUUUUUCUCGACUGAAAAGCUACCGCGGGUAGGCGCGCCGUUU
 AAUUAAGGAGGGGUGUUAGUUUUUUUCAUGCUAGCGUAGCCCGGGCGUAGCGGCCGCGUACCUGCAGGUC
 CUUUGAGGGGAUGUUAUAGUUUUUUUCAUCCAAGAGGAUCAAAAUGUAAAGGUAAUCAACUUGACUUU
 GAUGAAAUGCAGCGGUUUAAUGGCACCAUAACAUGUUUUUUGUUUAUUAUUGCUCCCAAGGGUUGGAGUUGA
 GCCCAUAUGUUCUCUCCUCCAGAGGUAAACAAGUGAUAAAUUGCGGGGAUUAUAUCUGACGUACGUUAU
 CUAGAAGCAGAGAGGAUUCUUGUCCUCUUCGGACCUUUGUGUCUGAAGAGACAUGUCAGACCAUAGUUG
 ACAUGCUCUCGGGUUCAUGUUGAUACACCAGACUCUGCCCUGGAUAUGACACUGUUUUGCAAUCACUCU
 AUUUGCAAUCCGACGAACUCAGUAUCAUCAUCCCAAGUGAUCUCCUGAGAGUAUCCAACUCCUCCCUU

Fig. 14 (contd.)

CAAGAGGGCCCCUGGAAUCAGCCCACUGGAAGAUAAAGGUUCUCCUCAAUUUGUAUACCCAGUUCAGGCC
CUCAGGGACUGGAGAUCUGACAAAGCCAGUCCAAUAACCACUUGACUAACCCGAUCAUCCUAUGAUUC
CCAGAAUAUAUCUCGUCGAAUGAUUUUCAGAAUGUGCCGCAGGAUCCUGAACGAGUAACCAUUCGGGCUAC
ACACUUUAACCCUUCGUGUGAUACAAAAGUCCUCAUGUUCUUCUUGCCUGUAAGUUCUUUCAGCGGGAC
GUAUUCAGGGGGUGGAAGCCACAAGUCAUCGUCAUCCAGAGGGGCUGACGCGGGAGAGGAUUUUUGAGUG
UCCUCGUCCCUGCGGUUUUUCACUAUCUUACGUAGGAGGUUCAUCGAUUUAUCAGUGGUGUUGCCUGUUU
UUUUCAUUGUUGACUUUGGGACAUCUCGGAUUUUUAUUGUCUAGAGGGACUGAGGGGAGAGGUUCGGUUAGC
AAGAUGUAUAGCGAUUCAAGUCAUCUUGCAUGAUUUUACUCAGCUUGUCGGAUUCGACUAACAACUGGAA
UUUCUUAGAGUUGGCCAAAGCGACCCAUCCCAGUACACAUCUAGGGGGAGUUUGGACCCGUC AUGGGCU
AAACGGGUCACACCUGGUACAUUUUUUGCCUCUUUAACUAUAUCAUCAAGGUUCAUUUUCAAUUGCUCAA
AAUUAUACAAGAGUAUCCCUGAGGAUCGAGAGGGAAACUUUAUUUUUUGGAGAAACUUUCUGCAAUCUG
GUGAGCGAUUCUCAGCCUCCACUGAUAGAUAUCCUCUUCAUUGGUAGCCGACCAUUC AAGGGCUGGAGGG
CCAGAAGCAAUUUGAGCCGCCAUCCUGGCUUUCGAUGAUUGGCUUUCUCUCUGAGAAGGAGUGGGUGUUG
UCUCCUUCUUGAGCUCUCGGCCAGUAGUCUGGGUUGAUUUUAUCCUCUGAAGACUUUCCUGGAGGGUUGG
AAAGUUGACCGCGACAUAAGGAUAUAUCUCUUCUACGGUCUGUGACCAUAUCUUGAGAAAUCUCUCUCCU
GACCUCAUUUUGUCUGACUAUUUGGACUCCAACAUUUCCAGGU AUGACUGGAACAGGAAGCUAGGAUCCU
CUCCUUCAUCCAUCUGAAAGUCCUCUCGAUACUUGCCUUCUCCCAUUGGCUAUCUCACCAUGGUUGGG
CGAUUUUCCAUCAUCCAGGUGAAGUCGCCCCAUUCCUCAGGGAGAUUGUCCACCUCUAUGGGUUCUUUU
UGGAGAUAGAGCCUGAUUGUCUUCGAUAUUUCUAUUUGAUCAAGAUCAACAGUUUCUUCAGCCAUCUCAAGAU
CGGCCAGACCGGCUCUAAUAGCACUAGGAUUGACAAAGAUUCUUGCUCAUGUUUGGGAUGGUUCGAAAGGA
GGGGUGUUAGUUUUUUUCAUGAUGGAUAUACACAAUCCGUAGAUUUCCGGCAUUUUUGUUAUUC AACUUCU
UACACAUUGAUCCUAGCAGAAGCACAGGCUGCAGGGUGACGGUCCAUCCCGCUCUCCUGGGCACAAGACA
UGGGCAGCGUGCCAUAUCCUGCUCCUCCACCUCGCGGGGAAGCCAUGGCUACCGGAUCCCGAUUGGAA
GUAGAGGUUCUCUCCGGAACCUGAGUCACUCGAAUAUGUCUUGUUUAGAAACUCGGCGAAUGAGUUUGGA
CGGGCUUGAUUGGAACUGACUGAGACAUUUCUCCGU AUGUGAGAUUCUCUUUAGUCGACCUCCAUUCA
UCAUGAUUCGAGUAUAAACAGCCUCCGGACUUCUGGUUUCACCUGAAAAGUAGUCCUCGUCGUCAGAGUU
GACAGUUCCAUCAUUCUGCCAGUGCUACGUCAGUCUUUGUCAGUUCAGCCGCCUCGU AUUCUUGAAGUUCU
UUCUCAUCUCUGAAGAAUCUUCUUUCAAUGUCCCUUUCCCGAAGAAUUCUUCUCCAGAUAGCCCCUA
GAACAGACAUUUCAUGAGGAGCACAUUGCAGCAAUAACCGUUGCAUUUAGGGAUCUGACUUGACCCAUAUA
GCAUCCUACAAAGUGAAUGAGAUUGAACACGUGACCAACAGCAUUUGAUGAAUAAGGAGAUUUCCACUC
AAGCCUAGUGAACGGAAGUGGAUGAAUAAGAGUGAGGAACAGCUGUCUCCUGCCCUGGCUC AAACAUC
UUCUUUAUCUCUCCUCAAAGUUCUUGUGGAAGAAUAUAGUAUUGCCUCUCUAGCGGUGAGAUUGAUUUG
UUUUUAUGAACCCAGUAAAUGAUACCAGUCCUGAACAGUCUUCAUAAGCAGUGACAACUGUGCCCACUCUG
AUUGCUGAAUAUAGAUGCUCAAUCCGGGAGAAAAACAUGUCAUAGGUUCCGGCCAAAAAUCUGAAGUUUG
GUUAUGUACUCCA AUUAGCACACAUUUUGUGAGUUUGUCAUUAAGAGUAUGGUGUCCACGAUUUUAAACAAA
AGGGGCUGUCUCAAAAAUCUGCUCUAUCCUGUCUGCAAUGUUUGUCUUAUAGUUACCAGUGUUUUGCCCG
GAUAUUUUGCUCAACCUAUACAGACUCAAGAGAAGACCGACUAAGGACGCAUGCUCAGGGACAGUGGGGU
CUCUUGUCAGUUC CAUGCCUCCUGUCAGAGCCCAAUCCCUUCUACAUCAGUACGUUUUAUCUCCACCAG
AGAACCUGGGGUGAUUCUUAUCUCCUUUUCGUGCAAUCACAAUCCAUAAGCUGGUCCAGUCUUCGGGACAU
GUCCCCUCAAAAAACUGCAUUGCCGCUGCCAAUAGGAACAUACAUCGUCAGGAUUAAGUUUGGCGGCGC
UCAUGCCUGACAAAACUGACUUGUAUGCUUUUAUUUAAAUCGGGAGCCUUUCCUAGGGUUUAUACAGGGCUU
UUUCAAAUCUUUGAUGGCAGGGUACUUGUACUCAUAUUGAUCCACGAUAAUCUCAGGCUUCAAGAGACC
ACCUGAUUAUUGACUUUGAAUACAAUCUUGUCGGCAUCCAUGUAGGGGUGUUACA UUUUUGCUUUGCAA
UUGACAAUGUCUGUUUUUUCUUUGAUUCUGGUUGUUAAGCGU-3'