

(12) STANDARD PATENT
(19) AUSTRALIAN PATENT OFFICE

(11) Application No. **AU 2018319637 B2**

(54) Title
Expression control using a regulatable intron

(51) International Patent Classification(s)
C12N 15/63 (2006.01)

(21) Application No: **2018319637**

(22) Date of Filing: **2018.08.22**

(87) WIPO No: **WO19/038544**

(30) Priority Data

(31) Number
1713545.0

(32) Date
2017.08.23

(33) Country
GB

(43) Publication Date: **2019.02.28**

(44) Accepted Journal Date: **2025.04.10**

(71) Applicant(s)
ASKBIO UK LTD

(72) Inventor(s)
WHYTESIDE, Graham

(74) Agent / Attorney
James & Wells Intellectual Property, GPO Box 1301, CANBERRA, ACT, 2601, AU

(56) Related Art
WO 2010/151827 A1
Shamu (Current Biology, 8, 4, R121 -R123, 1998).



(51) International Patent Classification:

C12N 15/63 (2006.01)

(21) International Application Number:

PCT/GB2018/052387

(22) International Filing Date:

22 August 2018 (22.08.2018)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

1713545.0 23 August 2017 (23.08.2017) GB

(71) Applicant: **SYNPROMICS LIMITED** [GB/GB]; Nine, Edinburgh Bioquarter, 9 Little France Road, Edinburgh EH16 4UX (GB).

(72) Inventor: **WHYTESIDE, Graham**; 3 Gartcows Drive, Falkirk Central Scotland FK1 5QQ (GB).

(74) Agent: **HGF LIMITED (GLASGOW)**; 1 City Walk, Leeds West Yorkshire LS11 9DX (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).

Published:

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

(54) Title: EXPRESSION CONTROL USING A REGULATABLE INTRON

(57) Abstract: The present invention relates to the use of a regulatory nucleic acid sequences that are able to regulate gene expression in eukaryotic cells and which are responsive to the unfolded protein response (UPR). There are disclosed regulatable introns and UPR-inducible promoters, which are able to regulate gene expression. There are also disclosed recombinant expression constructs comprising such regulatory nucleic acid sequences, whereby expression of the encoded expression product can be induced by invoking the unfolded protein response (UPR) in a eukaryotic cell containing the construct, methods of using such constructs and associated vectors, cells and suchlike.



WO 2019/038544 A1

Expression Control Using a Regulatable Intron

The present invention relates to the use of a regulatory nucleic acid sequences that are able to regulate gene expression in eukaryotic cells and which are responsive to the unfolded protein response (UPR). In some aspects, the invention relates to the use of regulatable introns to control gene expression. In other aspects it relates to UPR-inducible promoters. The invention also relates to recombinant expression constructs comprising such regulatory nucleic acid sequences, whereby expression of the encoded expression product can be induced by invoking the unfolded protein response (UPR) in a eukaryotic cell containing the construct, and to methods of using such constructs and associated vectors, cells and suchlike.

Background of the Invention

Regulatable gene expression is desirable in a great many circumstances, where it is beneficial or necessary to control the expression levels of an expression product. For example, in the case of industrial biotechnology, it can be highly advantageous to be able to induce production of an expression product (e.g. a protein) at the desired time in a fermentation process. In another example, in gene therapy it can be desirable to be able to induce expression of a therapeutic product (e.g. a therapeutic protein) at the desired time and/or location of treatment.

Inducible promoters are known in the art, for example the tetracycline (Tet)-on and -off inducible expression systems (Gossen M & Bujard H. *PNAS*. 1992 Jun 15; 89(12):5547-51; Gossen M, Freundlieb S, Bender G, Müller G, Hillen W, & Bujard H. *Science*. 1995 Jun 23; 268(5218)).

Inducible promoters work by regulating expression at the transcription level, controlling the amount of mRNA produced from the relevant expression system. While this can provide useful levels of gene expression, there is a need for additional, and ideally improved systems for controlling expression.

In particular, there is a need for improved expression control systems for the expression of toxic proteins, i.e. proteins that are toxic to the cell in which they are produced. In the case of toxic proteins, the expression of even small amounts of the protein can often cause cell death or very poor production.

Moreover, in the case of gene of cell therapy, there is a need for systems to allow expression of a therapeutic protein or RNA, or other expression product, to be induced at a desired time and/or location.

- 5 The present invention relates to regulation of expression that acts at the point of translation. In particular, it relates to the use of an intron that is spliced out as a result of the unfolded protein response (UPR) in eukaryotic cells to control expression.

The unfolded protein response (UPR) is a cellular coping mechanism for endoplasmic
10 reticulum stress. The UPR is activated in response to an accumulation of unfolded or misfolded proteins in the lumen of the endoplasmic reticulum (ER). The UPR aims to restore normal function of the cell through various mechanisms. If these objectives are not achieved within a certain time span or the disruption is prolonged, the UPR aims towards apoptosis. The UPR can be triggered by increased protein synthesis and folding such as heterologous
15 protein production or via other cellular stresses, e.g. chemically induced stresses such as blocking glycosylation pathways or disulphide bond formation. The UPR is highly conserved across all eukaryotes.

In mammalian cells there exists three mechanisms for ER stress to be sensed and the UPR
20 to be activated:

1) IRE-1 splicing of XBP-1 mRNA

ER-stress from increased protein folding demand or via chemical inhibition of ER-processes is detected by the ER transmembrane protein IRE1 via disassociation of the chaperone BiP.
25 This disassociation of BiP activates IRE1 by allowing oligomerisation and phosphorylation of the protein which leads to an ER lumen facing RNase domain being exposed. Subsequently, the RNase domain catalyses the removal of a non-canonical intron from XBP-1 mRNA in a spliceosome-independent manner. Under non-stress conditions XBP1 mRNA is un-spliced (XBP1u) which, when translated, forms a 261 amino acid ORF that is a non-functional
30 protein. However, when ER-stress is detected XBP1u is spliced by the IRE1 RNase to form XBP1s which encodes a functional 376 amino acid protein. This functional protein, XBP1 is a transcription factor that controls the expression of several genes involved in protein homeostasis such as chaperones, disulphide isomerases and enzymes involved in phospholipid biosynthesis. It regulates these processes by binding to specific sequences,
35 ER stress response elements (ERSEs) or unfolded protein response elements (UPREs), and enhancing gene expression. Therefore, the control of gene expression is the removal of the

mRNA intron from XBP1u by IRE1. A very similar system operates in non-mammalian eukaryotic cells.

2) *ATF6*

- 5 Activated transcription factor (ATF6) ATF6 is a type-2 transmembrane protein that has a transcription factor domain in the cytosolic region of the protein. It is synthesized as an inactive precursor and retained in the ER by association with BiP/GRP78. In response to stress conditions, ATF6 disassociates and is transported to the Golgi apparatus, where processing occurs to release the transcription factor domain. This domain is then transported
10 to the nucleus where it can bind to ERSEs and UPREs to enhance expression of genes involved in protein homeostasis.

3) *Double* stranded RNA-activated protein Kinase (PERK)

- PERK is a type-1 ER transmembrane protein that contains an ER luminal stress sensor and
15 cytosolic protein kinase domain. PERK is activated in response to ER stress and inhibits normal protein translation in the ER of mammalian cells by inactivating eukaryotic initiation factor (eIF2a) via phosphorylation.

Mechanisms 1) and 2) are of most relevance to the present invention.

20

Statements of the Invention

- According to the present invention there is provided a synthetic nucleic acid expression construct for producing an expression product in a cell, the nucleic acid expression construct
25 comprising a promoter sequence operably linked to a nucleic acid sequence encoding an expression product, the nucleic acid sequence encoding the expression product comprising a sequence which encodes a regulatable intron, said regulatable intron being an intron which comprises an excisable sequence which is capable of being spliced out of a transcript produced from the synthetic expression construct via the unfolded protein response (UPR)
30 system in the cell, thereby resulting in a transcript encoding a functional expression product.

- Thus, the present invention is based upon the use of a regulatable intron which is capable of being spliced out of an RNA transcript of a nucleic acid sequence encoding an expression product via the UPR mechanism to regulate expression. The UPR mechanism is ubiquitous
35 across eukaryotes, and the mechanism and relevant sequences are remarkably conserved. Thus, the invention can be practised across all eukaryotic cells. The regulatable intron is preferably capable of being spliced out by the IRE1 protein or a homologue or orthologue

thereof (homologues or orthologues of IRE1 are present in all eukaryotes, including fungi, plants and mammals).

The promoter is typically heterologous to the nucleic acid sequence encoding the expression product. That is to say, the promoter is not naturally found operatively linked to sequence encoding the expression product. For example, the promoter and nucleic acid sequence encoding the expression product are typically not found together in a naturally occurring gene. In some embodiments of the invention the promoter is a synthetic promoter, i.e. a promoter that is not naturally occurring. There are a wide range on constitutive and non-constitutive promoters known in the art suitable for use in eukaryotic cells, and mention can be made of CAG promoter, CMV promoter, SV40 as non-limiting examples.

In preferred embodiments the expression product is a protein. In such embodiments the un-spliced transcript produced from the nucleic acid sequence encoding the expression product encodes a truncated or otherwise defective version of the protein as a result of the presence of the regulatable intron, but when the transcript is processed by the UPR mechanism in the cell the excisable sequence of the intron is spliced out and the functional protein can be produced from the transcript. Thus, splicing out of the regulatable intron results in a functional mRNA encoding the functional protein expression product.

However, in some cases the expression product can be a product other than a protein. Suitably the expression product can be an RNA molecule, for example ribozyme, RNA aptamer, siRNA, antisense RNA, or miRNA. In such embodiments a non-functional form of the RNA molecule is produced as the un-spliced transcript, and splicing out of the excisable portion of the intron converts the RNA molecule to an active form.

In preferred embodiments of the present invention, splicing out of the excisable sequence of the regulatable intron permits correct translation of the transcript from the nucleic acid sequence encoding the expression product, thereby allowing the desired expression product (e.g. protein) to be produced. In such embodiments the presence of the intron in the transcript from the nucleic acid sequence encoding the expression product results in a protein being translated from the transcript which is non-functional. Typically, this results either from the insertion in the translated protein of amino acids coded for by the intron, or, more preferably, as a result of introduction of a stop codon or frame shift (relative to the normal transcript coding for the functional protein) in the transcript 3' of the intron. The protein encoded by the un-spliced transcript can be non-functional for many reasons, for example:

- The intron may result in a frame shift downstream (i.e. in a 3' direction) of the intron. This typically results when the excisable sequence of the intron is not a multiple of three nucleotides in length. Such a frame shift often results in introduction of a stop codon, resulting in a truncated protein. In other cases it can simply result in a complete alteration in the encoded amino acid sequence downstream of the intron.
- Introduction of a coding sequence which results in amino acids being present in the translated protein which are disruptive to function of the protein. In this case the amino acid sequence downstream of the intron is not altered, but the amino acid sequence encoded by the intron, which in the un-spliced form will be present in the translated protein, disrupts the function of the protein.
- Introduction of a stop codon in the intron sequence. In this case, the intron itself may comprise a stop codon, which will result in premature termination of translation, and the production of a truncated protein.

In certain preferred embodiments of the present invention the synthetic nucleic acid expression construct is configured such that splicing out of the excisable sequence of the intron eliminates a premature stop codon in the transcript. "Premature" in the present context means a stop codon which occurs upstream (i.e. in a 5' direction) of the stop codon in the normal transcript, i.e. the transcript encoding the functional protein (e.g. the wild type mRNA).

Preferably the regulatable intron is configured such that splicing out of the excisable sequence of the intron results in a shift of reading frame for sequences in the transcript located downstream (i.e. 3') of the regulatable intron.

In certain preferred embodiments of the present invention the regulatable intron is configured such that, when splicing occurs, a sequence having a length in nucleotides which is not a multiple of 3 is excised from the transcript. In other words, the excisable sequence has a length of n nucleotides, wherein n is not divisible by three.

In preferred embodiments of the present invention the regulatable intron comprises the sequence CNG/CNG- X_n -CNG/CNG, wherein X_n represents a sequence of length n bases, wherein / represents a cleavage site and wherein the sequence CNG- X_n -CNG is excised from the transcript.

Thus, in other words, the regulatable intron suitably comprises a central sequence (X_n) flanked by two splice site target sequences, each having the sequence CNG/CNG, wherein / represents a cleavage site.

5 CNG/CNG is a consensus splice site sequence targeted in a highly-conserved manner by the UPR system in eukaryotic cells. As is known in the art, this splice site consensus sequence is targeted by the IRE1 protein (homologues or orthologues of which are present in all eukaryotes, including fungi, plants and mammals) when the UPR response is induced. In nature, introns having this consensus splice site target sequence are found in the mRNA
10 encoding transcription factors that are activated in the UPR, e.g. the XBP1 protein (in metazoa), the Hac1 protein (in yeast), and the bZIP60 protein (in plants). The endoribonuclease activity of IRE1 or homologues or orthologues thereof acts to cleave the RNA transcript at the position indicated by /, to remove the excisable intron sequence, and the cleaved RNA is then ligated together by an RNA ligase protein (RNA ligase Rlg1p in *S.*
15 *cerevisiae*, RNA ligase RtcB in mammalian cells). The UPR systems in mammalian, yeast and plant cells have been widely investigated, and the mechanisms are comparatively well-characterised (see, for example, Yoshida et al., *Cell*, Vol. 107, 881–891, December 28, 2001; Lu, et al. *Mol Cell*. 2014 September 4; 55(5): 758–770; Samali et al., *International Journal of Cell Biology*, Vol. 2010, Article ID 830307, 11 pages doi:10.1155/2010/830307;
20 Chakraborty et al., *Appl Biol Chem* DOI, 10.1007/s13765-016-0167-6, Online ISSN 2468-0842, Print ISSN 2468-0834; and Nagashima et al., *Scientific Reports* 1, Article number: 29 (2011), DOI: 10.1038/srep00029). Accordingly, the mechanisms of the UPR will not be discussed in depth here. However, it is important to note that, in view of the high level of conservation of IRE1-mediated splicing, introns originating from one species can be
25 successfully spliced by another species which is evolutionary highly diversified; for example, Yoshida et al. (*id.*) explains that mammalian cells can successfully splice out the intron from the yeast Hac1 mRNA (Hac1 corresponding to XBP1 in mammalian cells).

It should be noted that the length of the regulatable intron can vary considerably. For
30 example, the XBP1 intron in mammals and plants is typically 26 nucleotides in length, though variants of 20 and 23 nucleotides have been observed. The Hac1 intron of yeast is considerably longer, typically in the order of 252 nucleotides in length (though, as mentioned above, this much longer Hac1 intron can still be spliced out in mammalian cells).

35 Accordingly, in various embodiments of the present invention the regulatable intron or X_n can be from 10 to 500 nucleotides in length, more preferably 15 to 350 nucleotides in length, yet more preferably 15 to 100 nucleotides in length, yet more preferably 15 to 35 nucleotides

in length, and yet more preferably 20 to 25 nucleotides in length. The excisable sequence of the regulatory intron thus suitably has a length of 16 to 506 nucleotides, more preferably 21 to 356 nucleotides in length, yet more preferably 21 to 106 nucleotides in length, yet more preferably 21 to 41 nucleotides in length, and yet more preferably 26 to 31 nucleotides in length. As mentioned above, in some embodiments of the present invention it is preferred that the length X_n is selected such that the length of the excisable sequence is not divisible by 3.

There is considerable freedom regarding the specific sequence of X_n . Preferred sequences are set out below, but many other variants could of course be used provided that the regulatable intron remains functional, i.e. it is spliced out of the transcript by the UPR system at suitable levels. Of course, some possible sequences may be sub-optimal or interfere with the splicing process, e.g. as the result of the formation of undesirable secondary structures, but the person skilled in the art can readily assess the effect of any given sequence to determine whether it has any adverse effect on splicing.

Assessment of the functionality of a regulatable intron, i.e. its ability of the regulatable intron to be successfully spliced out of a transcript upon induction of the UPR, can readily be assessed by the skilled person using a wide range of approaches, and these can be tailored for the particular expression system in which the construct is intended to be used. As one preferred example, the methodology described in the examples below can be used, e.g. Example 1. For example, the functionality of any candidate regulatable intron to be assessed can be substituted into the construct described in Example 1 (referred to as SYN-P-XBP-01) in place of the exemplary intron used in Example 1, and the ability of said intron to be successfully spliced out when the UPR is induced can be measured by assessing the level of EGFP expression before and after UPR induction by 2mM DTT, exactly as carried out in Example 1. A functional regulatable intron is one which is able to be successfully spliced out after induction of the UPR to result in the expression of functional EGFP. Preferably a functional intron confers at least a 5-fold increase in expression 24 hours after induction of the UPR with 2mM DTT, more preferably at least a 10-fold increase in expression, more preferably at least a 100-fold increase in expression, and yet more preferably at least a 1000-fold increase in expression of EGFP. It is preferred that before induction of the UPR the expression levels of EGFP are minimal, and preferably negligible. Minimal expression can be defined as, for example, less than 50% of the expression levels of a control construct as used in Example 1 (i.e. a construct without the regulatable intron in which expression of the sequence encoding the EGFP is driven by CMV-mp), preferably less than 20%, more preferably less than 10%, yet more preferably less than 5%, yet more

preferably less than 1%. Negligible expression levels are those that are essentially undetectable using the methodology of Example 1.

However, it will be appreciated the skilled person could readily modify the approach taken in Example 1. For example, this might involve using different cell types, using a different expression product, using a different indicator of successful splicing (e.g. measuring levels of spliced mRNA encoding the functional expression product, or using a different reporter protein), and using a different inducer of the UPR. In such modified approaches, it remains the case that a functional regulatable intron is one which is able to be successfully spliced out after induction of the UPR to result in expression of a functional expression product after induction of the UPR. Preferably this results in at least a 5-fold increase in expression of the transcript encoding a functional expression product 24 hours after induction of the UPR, more preferably a 10-fold increase, more preferably a 100-fold increase, and yet more preferably at least a 1000-fold increase. It is preferred that before induction of the UPR the expression levels of the transcript encoding the functional expression product are minimal or negligible.

CNG/CNG[CG] is a preferred splice site consensus target sequence for mammalian cells, in which the presence of a C or G at the indicated location is preferred (though not required).

The presence of a C at this location is typically preferred to a G. Accordingly, in some preferred embodiments of the invention, especially when the synthetic nucleic acid expression construct is intended for use in mammalian cells, the intron comprises the sequence CNG/CNG[CG] at one, other, or both (preferably both) ends of the intron.

Thus, in some preferred embodiments of the invention, the regulatable intron comprises the sequence CNG/CNG-Xn-CNG/CNG[CG], wherein Xn represents a sequence of length n nucleotides, wherein / represents the cleavage site such that the excisable sequence CNG-Xn-CNG is excised from the transcript upon splicing, and wherein the nucleotide at the 5' end of the sequence Xn is a C or G. Suitable lengths for Xn are set out above.

In some preferred embodiments of the invention Xn comprises the sequence CACUCAGACUACGUGCACCU (SEQ ID NO: 1) or a sequence which is at least 60% identical thereto, more preferably at least 70% identical thereto, yet more preferably at least 80% identical thereto, yet more preferably at least 90% identical thereto, and yet more preferably at least 95%, 96%, 97%, 98% or 99% identical thereto.

In some preferred embodiments of the invention X_n consists of the sequence CACUCAGACUACGUGCACCU (SEQ ID NO: 1) or a sequence which is at least 60% identical thereto, more preferably at least 70% identical thereto, yet more preferably at least 80% identical thereto, yet more preferably at least 90% identical thereto, and yet more preferably at least 95%, 96%, 97%, 98% or 99% identical thereto.

The sequence CACUCAGACUACGUGCACCU (SEQ ID NO: 1) corresponds to the region of the mammalian XBP1 intron lying inside of the IRE1 cleavage sites as set out above.

Accordingly, this represents a preferred embodiment of the invention, particularly when the synthetic nucleic acid expression construct is intended for use in mammalian cells. However, sequences which are highly similar to this are also found across a range of non-mammalian XBP1 introns.

In some embodiments of the invention X_n comprises or consists of one of the following sequences:

- CACUCAGACUACGUGCACCU (SEQ ID NO: 1);
- CACUCAGACUACGUGCUCU (SEQ ID NO: 2);
- CACUCAGACUACGUGCCCCU (SEQ ID NO: 3);
- CACUCAGACUACGUGCGCCU (SEQ ID NO: 4); and
- CACUCAGACUAUGUGCACCU (SEQ ID NO: 5).

In other embodiments of the invention X_n comprises or consists of the sequence ACGGGCAACUUUACACGACG (SEQ ID NO: 49) or a sequence which is at least 60% identical thereto, more preferably at least 70% identical thereto, yet more preferably at least 80% identical thereto, yet more preferably at least 90% identical thereto, and yet more preferably at least 95%, 96%, 97%, 98% or 99% identical thereto.

In a particularly preferred embodiment of the invention the regulatable intron comprises or consists of the sequence CNG/CNGCACUCAGACUACGUGCACCUCNG/CNGC (SEQ ID NO: 6), or a sequence which is at least 60% identical thereto, more preferably at least 70% identical thereto, yet more preferably at least 80% identical thereto, yet more preferably at least 90% identical thereto, and yet more preferably at least 95%, 96%, 97%, 98% or 99% identical thereto, wherein / represents a cleavage site. In variant sequences according to the above sequence identity levels, the splice site target sequence preferably remains as CNG/CNGC, and sequence variation occurs in the other regions.

Suitably the regulatable intron comprises or consists of the sequence

CAG/CAGCACUCAGACUACGUGCACCUCUG/CUGC (SEQ ID NO: 7), or a sequence

which is at least 60% identical thereto, more preferably at least 70% identical thereto, yet

more preferably at least 80% identical thereto, yet more preferably at least 90% identical

5 thereto, and yet more preferably at least 95%, 96%, 97%, 98% or 99% identical thereto,

wherein / represents a cleavage site. In variant sequences according to the above sequence

identity levels, the splice site target sequence preferably remains as CAG/CUGC, and

sequence variation occurs in the other regions.

10 In some preferred embodiments of the invention the regulatable intron comprises or consists of one of the following sequences:

- CNG/CAGCACUCAGACUACGUGCACCUCUG/CNG (SEQ ID NO: 8);

- CNG/CAGCACUCAGACUACGUGCUCUCUG/CNG (SEQ ID NO: 9);

- CNG/CAGCACUCAGACUACGUGCCCCUCUG/CNG (SEQ ID NO: 10);

15 - CNG/CAGCACUCAGACUACGUGCGCCUCUG/CNG (SEQ ID NO: 11); and

- CNG/CAGCACUCAGACUAUGUGCACCUCUG/CNG (SEQ ID NO: 12).

In further preferred embodiments of the invention the regulatable intron comprises or consists of one of the following sequences:

20 - CAG/CAGCACUCAGACUACGUGCACCUCUG/CUGC (SEQ ID NO: 7);

- CAG/CAGCACUCAGACUACGUGCUCUCUG/CUGC (SEQ ID NO: 13);

- CAG/CAGCACUCAGACUACGUGCCCCUCUG/CUGC (SEQ ID NO: 14);

- CAG/CAGCACUCAGACUACGUGCGCCUCUG/CUGC (SEQ ID NO: 15); and

- CAG/CAGCACUCAGACUAUGUGCACCUCUG/CUGC (SEQ ID NO: 16).

25

In another embodiment of the invention the regulatable intron comprises the

sequence CAG/CUGCAGCACUCAGACUACGUGCACCUCUG/CAG (SEQ ID NO:

17) or CAG/CUGCAGCACUCAGACUACGUGCACCUCUG/CUGG

(SEQ ID NO: 27), wherein / represents a cleavage site. This sequence results from the

30 addition of the trinucleotide CUG to the mammalian XBP1 intron sequence at the underlined

position. Addition of this trinucleotide is believed to slightly de-optimize splicing of the intron

to reduce any undesirable splicing (and hence background expression of the expression

product) in cells.

35 Accordingly, in some preferred embodiments of the invention Xn comprises or consists of

CAGCACUCAGACUACGUGCACCU (SEQ ID NO: 23).

In another embodiment of the invention the regulatable intron comprises the sequence: CNG/CAGACGGGCAACUUUACACGACGCUG/CNG (SEQ ID NO: 50), or a sequence which is at least 60% identical thereto, more preferably at least 70% identical thereto, yet more preferably at least 80% identical thereto, yet more preferably at least 90% identical thereto, and yet more preferably at least 95%, 96%, 97%, 98% or 99% identical thereto, wherein / represents a cleavage site. In variant sequences according to the above sequence identity levels, the splice site target sequence preferably remains as CNG/CNG, and sequence variation occurs in the other regions.

For the avoidance of doubt, it is noted that in the intron sequences of the present invention, it is preferred that the splice site target sequences at each end of the intron are constrained as CNG/CNG, and more preferably as CNG/CNG[CG]. Thus, while some variation in the splice site target sequences is provided by these sequences, further variation, if required, should be accommodated by the central sequence lying between the two splice site target sequences.

In some preferred embodiments of the invention the regulatable intron is the XBP1 intron, the Hac1 intron, the bZIP60 intron, or a homologue thereof. By this it is meant that the intron can be a wild type form of the XBP1, Hac1 or bZIP60 intron, or a naturally occurring homologue thereof.

In some embodiments of the invention it may be preferable that the splice site target sequence (i.e. comprising the sequence CNG/CNG) in the transcript is flanked by sequences which are able to interact to form a stem-loop structure. Thus, the splice site target sequence is preferably flanked by sequences which are complementary to one another, such that they will hybridise with each other to form a stem-loop structure in which the splice site target sequence is located at least partially, preferably entirely, within the loop region of the stem-loop structure that is formed in the transcript.

With regard to the wild type XBP1, HAC1 and bZIP60 introns, it has been hypothesised that a stem loop structure is formed at the splice sites in the mRNA transcripts. This involves hybridisation between sequences in the intron and the exon adjacent to the splice site target sequence that are complementary in nucleotide sequence when read in opposite directions. The experiments reported herein show that splicing is successfully carried out when the intron is inserted into a coding sequence in a situation where such a stem loop structure is not expected to form. Thus, formation of a stem-loop structure does not appear to be necessary for successful splicing of the regulatable intron of the present invention.

Nevertheless, in some cases it may be desirable that a stem-loop structure is formed as this may, for example, result in optimal splicing activity. In the alternative, it may in some cases be desirable to avoid providing a sequence which is amenable to forming a stem-loop structure, as that may lead to undesirably active splicing of the intron, thus potentially leading to expression leakage.

In certain embodiments of the present invention, where a stem-loop structure is to be formed, the stem-loop structure formed by the transcript preferably comprises a loop which comprises from 6 to 9 nucleotides, and a stem which is from 3 to 10 nucleotides in length.

More preferably the stem-loop structure comprises a loop which comprises from 7 to 8 nucleotides, and a stem which is from 4 to 8 nucleotides in length.

In certain embodiments of the present invention, where a stem-loop structure is to be formed, the intron may suitably comprise a sequence at the splice target site as follows:

-Yn-CNG/CNG-A-Zn-

wherein A is an sequence having a length of from 0 to 3 nucleotides (preferably 1 or 2 nucleotides),

wherein / represents the cleavage site,

and wherein Yn and Zn represent sequences that are complementary in nucleotide sequence when read in opposite directions, and are thus able to hybridise to form the stem of the stem-loop structure. Yn and Zn are preferably from 3 to 10 nucleotides in length, more preferably from 4 to 8 nucleotides in length.

In some embodiments the intron may suitably comprise a sequence at the splice target site as follows:

-Zn-CNG/CNG[CG]-A-Yn-, where the components have the same meaning as above. In this case A preferably has a length of 0, 1 or 2 nucleotides.

It will be apparent that providing suitable complementary sequences (e.g. Yn and Zn in the structures above) to provide the stem structure can be achieved by adapting the sequence of the intron to provide a suitable region which is complementary to the corresponding sequence in the adjacent coding (i.e. exon) sequence. It may also be possible or desirable to alter the sequence of the coding region to some extent, e.g. by utilising redundancy in the genetic code to alter the nucleic acid sequence without affecting the encoded amino acid sequence; typically an alteration in amino acid sequence of the expression product should be avoided.

In some embodiments of the invention the nucleic acid expression construct comprises an inducible promoter operably linked to the nucleic acid sequence encoding the expression product comprising a sequence which encodes a regulatable intron. As mentioned above, inducible promoters are known in the art. By combining an inducible promoter and the regulatable intron of the present invention, a dual level of expression can be achieved, i.e. control at both the transcription level and at the translation level. This can allow for very tight control of expression, e.g. to avoid any expression "leakage". This can be important, for example, during the expression of toxic proteins, or in any case where the amount of background expression has to be kept to an absolute minimum prior to induction of expression at the desired time.

In a preferred embodiment the inducible promoter is an unfolded protein response (UPR) inducible promoter, i.e. a promoter which itself is induced by the UPR. In such an embodiment induction of the UPR serves to both induce expression in terms of driving transcription and in permitting expression of a functional expression product as a result of splicing of the regulatable intron.

In embodiments of the present invention, the UPR inducible promoter suitably comprises at least one binding site for ATF6, XBP1, bZIP60, or homologous or otherwise equivalent transcription factors, which drive protein expression in the UPR.

Suitably the UPR inducible promoter comprises one or more copies of the at least one of the following sequences:

- TGACGTG (the ATF6 consensus sequence),
- TGACGTGCT (a variant of the above),
- TGACGTG[TG] (known as the UPRE site),
- CCAAT-N9-CCACG (known as the ERSE1 site) (SEQ ID NO: 18), and
- ATTGG-N-CCACG (known as the ERSE2 site) (SEQ ID NO: 19).

These sites are bound by ATF6, XBP1 and bZIP60.

Suitably the promoter comprises one or more copies of the sequence TGACGTG (optionally as part of TGACGTGCT or TGACGTG[TG]), preferably 3 or more copies of the sequence TGACGTG[TG] (optionally as part of TGACGTGCT or TGACGTG[TG]), and preferably 5 or more copies of the sequence TGACGTG[TG] (optionally as part of TGACGTGCT or TGACGTG[TG]).

An exemplary UPR inducible promoter sequence comprises the following sequence (SEQ ID NO: 20): TGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCT .

This sequence comprises 6 tandem copies of the UPRE site.

- 5 Another exemplary UPR inducible promoter sequence comprises the following sequence (SEQ ID NO: 47):

TGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTG
ATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTA
GCTAGTAGTTGACGTGCT . This sequence comprises 6 copies of the UPRE site, each

- 10 spaced by 20 nucleotides.

Suitably the UPR inducible promoter comprises said at least one binding site for ATF6, XBP1 or homologous or otherwise equivalent transcription factors which drive the UPR operatively linked to a minimal promoter sequence (e.g. the CMV-minimal promoter). Other
15 suitable minimal promoters are known in the art.

The CMV-minimal promoter has the following sequence (SEQ ID NO: 21):

AGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTAGATACGCCATCCACGCTGTTTT
GACCTCCATAGAAGAT

20

An exemplary inducible promoter thus comprises a nucleic acid comprising a sequence according to SEQ ID NO: 20 positioned upstream of, and operatively linked to, a nucleic acid having a sequence according to SEQ ID NO: 21.

- 25 For example, the inducible promoter can suitably comprise the following sequence (SEQ ID NO: 22):

TGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCTGGTACCGTCGACGA
TATCGGATCCAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTAGATACGCCATCC
ACGCTGTTTTGACCTCCATAGAAGATCGCCACC

30

Further details of UPR-inducible promoters that can be used in conjunction with the regulatable intron are provided below.

- 35 The nucleic acid sequence encoding an expression product is suitably a transgene. The transgene typically encodes a gene expression product such as RNA or a polypeptide (protein). The transgene may be a full-length cDNA or genomic DNA sequence, or any fragment, subunit or mutant thereof that has at least some biological activity. The transgene

may suitably be a minigene, i.e. a gene sequence lacking part, most or all of its native intronic sequences. The transgene optionally may comprise conventional intron sequences (i.e. in addition to the regulatable intron). Optionally, the transgene may be a hybrid nucleic acid sequence, i.e., one constructed from homologous and/or heterologous cDNA and/or genomic DNA fragments. By 'mutant' is meant a nucleic acid sequence that contains one or more nucleotides that are different from the wild-type or naturally occurring sequence, i.e., the mutant nucleic acid sequence contains one or more nucleotide substitutions, deletions, and/or insertions. The nucleotide substitution, deletion, and/or insertion can give rise to a gene product (i.e., protein or nucleic acid) that is different in its amino acid/nucleic acid sequence from the wild type amino acid/nucleic acid sequence. In some cases, the transgene may also include a sequence encoding a leader peptide or signal sequence such that the transgene product will be secreted from the cell.

A conventional (i.e. non-regulatable) intron can be utilized in the nucleic acid sequence encoding an expression product in addition to the regulatable intron discussed above. The term "conventional intron" encompasses any portion of a whole intron that is large enough to be recognized and spliced by the nuclear splicing apparatus. Typically, short, functional, intron sequences are preferred in order to keep the size of the expression cassette as small as possible, which facilitates the construction and manipulation of the expression cassette. In some embodiments, the intron is obtained from a gene that encodes the protein that is encoded by the nucleic acid sequence encoding an expression product. The conventional intron can be located 5' to the sequence encoding an expression, 3' to the sequence encoding an expression, or within the sequence encoding an expression. Thus, in some embodiments, the nucleic acid sequence encoding an expression product further comprises a conventional intron. Non-limiting examples of suitable introns are Minute Virus of Mice (MVM) intron, beta-globin intron (betaIVS-II), factor IX (FIX) intron A, Simian virus 40 (SV40) small-t intron, and beta-actin intron. Introns can have benefits in terms of improved expression levels, as is well-known in the art.

In some preferred embodiments of the present invention the nucleic acid sequence encoding an expression product encodes a protein. Essentially any protein can be used, and by way of non-limiting example the protein can be an enzyme, an antibody or antibody fragment (e.g. a monoclonal antibody), a viral protein (e.g. REP-CAP, REV, VSV-G, or RD114), a therapeutic protein, or a toxic protein (e.g. Caspase 3, 8 or 9).

In some preferred embodiments of the present invention the nucleic acid sequence encoding an expression product encodes a toxic protein. A "toxic protein" in this case means a protein

that is toxic to the cells in which the expression product are, in use, produced. For example, the toxic protein might be one of the following: Caspase 3, caspase 8, caspase 9, and a toxic viral protein (such as VSV-G or the AAV REP protein).

5 In some preferred embodiments of the present invention the nucleic acid sequence encoding an expression product encodes a therapeutic expression product. The therapeutic expression product can be a protein, e.g. a secretable protein such as, e.g., a clotting factor, e.g., factor IX or factor VIII, a cytokine, a growth factor, an antibody or nanobody, a chemokine, a plasma factor, insulin, erythropoietin, lipoprotein lipase, or a toxic protein.

10 Alternatively, the therapeutic expression product may be RNA, such as siRNA or miRNA. Various therapeutic siRNAs have been described in the art, and, by way of non-limiting example, the siRNA may be on that is intended to treat to treat FTDP-17 (frontotemporal dementia), DYT1 dystonia, growth hormone deficiency, BACE1 in Alzheimer's, Leukaemia
15 (e.g. targeting c-raf, bcl-2), melanoma (e.g. targeting ATF2, BRAF), prostate cancer (e.g. targeting P110B), and pancreatic carcinoma (e.g. targeting K-Ras). SiRNA therapies are summarised in "Therapeutic potentials of short interfering RNAs", Appl Microbiol Biotechnol, DOI 10.1007/s00253-017-8433-z. Similarly, for miRNA, various miRNA therapeutic approaches that could be implemented according to the present invention are summarised in
20 "MicroRNA therapeutics: towards a new era for the management of cancer and other diseases", Nature Reviews Drug Discovery; 16, 203–222 (2017).

Suitably the nucleic acid expression construct comprises sequences providing or coding for one or more of, and preferably all of, a ribosomal binding site, a start codon, a stop codon,
25 and a transcription termination sequence.

In a further aspect the present invention provides a nucleic acid comprising a sequence encoding an expression product (e.g. a gene), the sequence encoding an expression product comprising a sequence which encodes a regulatable intron, said regulatable intron
30 being an intron which comprises an excisable sequence which is capable of being spliced out of a transcript produced from the synthetic expression construct via the unfolded protein response (UPR) system in the cell, thereby resulting in a transcript encoding a functional expression product, and wherein the expression product is not the XBP1 protein, Hac1 protein, bZIP60 protein or a homologue thereof.

35 Accordingly, the sequence encoding an expression product is not a gene which naturally contains a regulatable intron according to the present invention. In other words, the

regulatable intron is heterologous to the sequence encoding an expression product in which it is found.

Preferred features of the regulatable intron and the sequence encoding an expression product are as set out above.

Such a nucleic acid can be inserted into any suitable expression construct, e.g. an expression vector such that is operatively linked to a promoter and any other elements required to drive transcription of the nucleic acid. Expression of the functional expression product will be controlled by the regulatable intron.

In a further aspect, the invention provides a vector comprising a synthetic nucleic acid expression construct as set out above.

The term "vector" is well known in the art, and as used in the application refers to nucleic acid molecules, e.g. double-stranded DNA, which may have inserted into it a nucleic acid expression construct according to the present invention. A vector is suitably used to transport an inserted nucleic acid molecule into a suitable host cell. A vector typically contains all of the necessary elements that permit transcribing the insert nucleic acid molecule, and, preferably, translating the transcript into a polypeptide. A vector typically contains all of the necessary elements such that, once the vector is in a host cell, the vector can replicate independently of, or coincidental with, the host chromosomal DNA; several copies of the vector and its inserted nucleic acid molecule may be generated. Vectors of the present invention can be episomal vectors (i.e., that do not integrate into the genome of a host cell), or can be vectors that integrate into the host cell genome. This definition includes both non-viral and viral vectors. Non-viral vectors include but are not limited to plasmid vectors (e.g. pMA-RQ, pUC vectors, bluescript vectors (pBS) and pBR322 or derivatives thereof that are devoid of bacterial sequences (minicircles)) transposons-based vectors (e.g. PiggyBac (PB) vectors or Sleeping Beauty (SB) vectors), etc. Larger vectors such as artificial chromosomes (bacteria (BAC), yeast (YAC), or human (HAC)) may be used to accommodate larger inserts. Viral vectors are derived from viruses and include but are not limited to retroviral, lentiviral, adeno-associated viral, adenoviral, herpes viral, hepatitis viral vectors or the like. Typically, but not necessarily, viral vectors are replication-deficient as they have lost the ability to propagate in a given cell since viral genes essential for replication have been eliminated from the viral vector. However, some viral vectors can also be adapted to replicate specifically in a given cell, such as e.g. a cancer cell, and are typically used to trigger the (cancer) cell-specific (onco)lysis. Virosomes are a non-limiting

example of a vector that comprises both viral and non-viral elements, in particular they combine liposomes with an inactivated HIV or influenza virus (Yamada et al., 2003). Another example encompasses viral vectors mixed with cationic lipids.

- 5 In some preferred embodiments, the vector is a viral vector, such as a retroviral, lentiviral, adenoviral, or adeno-associated viral (AAV) vector, more preferably an AAV vector. AAV vectors are preferably used as self-complementary, double-stranded AAV vectors (scAAV) in order to overcome one of the limiting steps in AAV transduction (i.e. single-stranded to double-stranded AAV conversion) (McCarty, 2001, 2003; Nathwani et al., 2002, 2006, 2011; 10 Wu et al., 2008), although the use of single-stranded AAV vectors (ssAAV) are also encompassed herein.

- In some preferred embodiments, the vector is a plasmid. Such a plasmid may include a variety of other functional nucleic acid sequences, such as one or more selectable markers, 15 one or more origins of replication, polycloning sites and the like.

- In some preferred embodiments of the invention the vector is an expression vector for expression in eukaryotic cells. Examples of eukaryotic expression vectors include, but are not limited to, pW-LNEO, pSV2CAT, pOG44, pXTI and pSG available from Stratagene; 20 pSVK3, pBPV, pMSG and pSVL available from Amersham Pharmacia Biotech; and pCMVDsRed2-express, pIRES2-DsRed2, pDsRed2-Mito, pCMV-EGFP available from Clontech. Many other vectors are well-known and commercially available. For mammalian cells adenoviral vectors, the pSV and the pCMV series of vectors are particularly well-known non-limiting examples. There are many well-known yeast expression vectors including, 25 without limitation, yeast integrative plasmids (YIp) and yeast replicative plasmids (YRp). For plants the Ti plasmid of agrobacterium is an exemplary expression vector, and plant viruses also provide suitable expression vectors, e.g. tobacco mosaic virus (TMV), potato virus X, and cowpea mosaic virus.

- 30 In some preferred embodiments the vector is a gene therapy vector. Various gene therapy vectors are known in the art, and mention can be made of AAV vectors, adenoviral vectors, retroviral vectors and lentiviral vectors. Where the vector is a gene therapy vector the nucleic acid, sequence encoding an expression product suitably encodes a therapeutic protein. The therapeutic protein may be a secretable protein. Non-limiting examples of 35 secretable proteins, in particular secretable therapeutic proteins, include clotting factors, such as factor VII I or factor IX, insulin, erythropoietin, lipoprotein lipase, antibodies or nanobodies, growth factors, cytokines, chemokines, plasma factors, toxic proteins, etc.

The nucleic acid expression construct and vectors of the present invention may be formulated in a pharmaceutical composition with a pharmaceutically acceptable excipient, i.e., one or more pharmaceutically acceptable carrier substances and/or additives, e.g., buffers, carriers, excipients, stabilisers, etc. The pharmaceutical composition may be provided in the form of a kit. The term "pharmaceutically acceptable" as used herein is consistent with the art and means compatible with the other ingredients of the pharmaceutical composition and not deleterious to the recipient thereof.

10 Accordingly, a further aspect of the invention provides a pharmaceutical composition comprising a nucleic acid expression construct or a vector described herein.

In a further aspect of the invention there is provided the use of nucleic acid expression constructs and vectors according to various aspects of the present invention for the manufacture of a pharmaceutical composition.

According to a further aspect of the present invention there is provided a cell comprising a synthetic nucleic acid expression construct or vector according to the present invention.

20 Preferably the cell is a eukaryotic cell. The eukaryotic cell can suitably be a fungal cell (e.g. yeast cell), an animal (metazoan) cell (e.g. mammalian cells), or a plant cell.

In some embodiments of the invention the cell may be a prokaryotic cell; although prokaryotic cells do not possess the UPR, prokaryotic cells may nonetheless be useful in production of the synthetic nucleic acid expression construct or other steps in handling the synthetic nucleic acid expression construct.

In some preferred embodiments of the invention the cell is *ex vivo*, e.g. in cell culture. In other embodiments of the invention the cell may be part of a tissue or multicellular organism.

30

In some preferred embodiments, the expression product is toxic to the cell in which the construct or vector is present. In one such embodiment, the cell is a cell for use in cell therapy (e.g. a therapeutic immune cell, such as a therapeutic T cell) which comprises a synthetic nucleic acid expression construct or vector according to the present invention, wherein the expression product is toxic to the cell. In such an embodiment induction of the UPR can be used to induce death of the cell, i.e. as a kill switch. Suitable toxic expression products include caspases, e.g. caspase 3, caspase 8, or caspase 9.

The synthetic nucleic acid expression construct may be inserted into the genome of the cell, or may be present in an episomal vector.

- 5 In a further aspect the present invention provides a method for producing an expression product, the method comprising:
- a) providing a population of eukaryotic cells comprising a synthetic nucleic acid expression construct according to the present invention;
 - b) treating said population of cells so as to induce the unfolded protein response,
10 thereby inducing splicing of the excisable sequence out of the regulatable intron;
 - c) incubating said population of cells under suitable conditions for production of the expression product; and
 - d) isolating the expression product from said population of cells.
- 15 The method is suitable a cell culture method. The cells may therefore be provided under suitable cell culture condition for the cell type being used. Suitable cell culture conditions are well known to the skilled person.

20 The synthetic nucleic acid expression construct can be present in the genome or can be episomal.

It will be apparent the present invention allows for the production of the expression product (or the active form of the expression product) to be delayed until a desired point in a cell culture process. This can, for example, permit the population of cells to be expanded until
25 such time as a desired cell number or concentration is reached, or a desired growth phase is reached.

For example, in the case of toxic proteins, the production of a functional (i.e. toxic) expression product can be avoided until a cell culture system is at a desired stage. Once
30 the toxic protein is expressed the cells will of course be adversely affected or killed.

The method suitably comprises incubating said population of cells under conditions suitable for growth of the cells prior to step b) of treating said population of cells so as to induce the unfolded protein response (UPR).

Typically step b) comprises applying a stress to said cells, said stress being suitable to induce the UPR. There is a wide range of stresses that can be used to induce the UPR, and they are described extensively in the literature.

- 5 In some preferred embodiments the step of inducing the UPR suitably comprises administering a chemical agent that is able to induce the UPR in said cells (i.e. a UPR-inducing chemical agent).

10 The person skilled in the art can readily assess the ability of any particular stress (e.g. a chemical agent) to induce the UPR. For example, the person skilled in the art can suitably apply the stress (e.g. by administering a candidate chemical agent) in the method of Example 1 in place of DTT. The ability of the agent to induce the UPR will be identified by the effect that the application of stress (e.g. the chemical agent) has upon the expression of the functional expression product, i.e. EGFP in the case of Example 1. Of course, the
15 method of example 1 can be modified as required, e.g. using a different cell type or a different construct. Example 4, in particular, demonstrates how various candidate chemical agents can be tested for their ability to induce the UPR. Various other methods to assess the ability of a stress (e.g. a chemical agent) to induce the UPR will be apparent to the skilled person.

20

Many chemical agents that can induce ER stress and induce the UPR are known in the art. UPR-inducing chemical agents that are able to induce the IRE1 pathway are suitable for use in the present invention as they will lead to splicing of the regulatable intron.

25 Exemplary UPR-inducing chemical agents that can be used to induce the UPR include:

- Dithiothreitol (DTT) – this agent reduces the disulfide bridges of proteins. The denatured proteins accumulated inside the ER.
- Tunicamycin – this agent inhibits N-linked glycosylation.
- Brefeldin A – this is commonly used as an inducer of the unfolded protein
30 response.
- Thapsigargin – this agent leads to ER Ca^{2+} depletion due to inhibition of the Sarco/Endoplasmic Reticulum Ca^{2+} -ATPase (SERCA).
- 2-deoxyglucose
- A23187 (CAS Number 52665-69-7)
- 35 • Bortezomib (Velcade)
- Quercetin

- Agents that disrupts the lipid balance in the cells such that UPR is induced, e.g. saturated fatty acids (e.g. palmitic acid) – i.e. those agents that induce lipid-induced ER stress response/UPR.

5 Such UPR-inducing chemical agents can be administered at suitable concentrations, and such concentrations can readily be determined by the skilled person through routine experimentation and consultation of the literature. Suitable concentrations for administration of the various agents are as follows: DTT 2 mM; Tunicamycin 2.5-5 µg/ml; Brefeldin A 0.5 µg/ml; Thapsigargin 0.1-1 µM; 2-deoxyglucose 4 mM; A23187 (CAS Number 52665-69-7)
 10 0.5 µM; Bortezomib (Velcade) 5-30 nM; and palmitic acid (or other fatty acids) 100 µM. These concentrations refer to the concentration of the agent in the medium to which the cells are exposed.

Another UPR-inducing chemical agents for use in the present invention is forskolin.

15 Forskolin (coleonol) is a labdane diterpene that is produced by the Indian Coleus plant (*Plectranthus barbatus*). Other names for forskolin include pashanabhedi, Indian coleus, makandi, HL-362, and NKH477.

Dithiothreitol (DTT), tunicamycin and thapsigargin are used extensively in the literature to
 20 induce the UPR, and thus represent preferred UPR-inducing chemical agents in some embodiments of the invention. Forskolin is another preferred UPR-inducing chemical agent, particularly but not exclusively in view of its safety profile for use *in vivo*.

In a particularly preferred embodiment of the present invention, an UPR-inducing chemical
 25 agent that is able to disrupt the lipid balance in said cells is administered in order to induce the UPR. The role of lipids and lipid metabolism in inducing the UPR has been extensively reported in the literature, and the phenomenon is termed “lipid-induced ER stress response/UPR”. See, for example, Iwao and Shidoj, PLOS ONE | DOI:10.1371/journal.pone.0132761 July 17, 2015; Robblee et al. “Saturated Fatty Acids
 30 Engage an IRE1a-Dependent Pathway to Activate the NLRP3 Inflammasome in Myeloid Cells” - Cell Reports 14, 2611–2623, March 22, 2016; Ariyama et al., “Decrease in Membrane Phospholipid Unsaturation Induces Unfolded Protein Response” - THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 285, NO. 29, pp. 22027–22035, July 16, 2010; Basseri and Austin “Endoplasmic Reticulum Stress and Lipid Metabolism:
 35 Mechanisms and Therapeutic Potential” - Biochemistry Research International Volume 2012, Article ID 841362, 13 pages, doi:10.1155/2012/841362; Kitai et al. “Membrane lipid saturation activates IRE1a without inducing clustering”- Genes to Cells (2013) 18, 798–809.

Accordingly, suitably the UPR-inducing chemical agent is able to alter the lipid balance of the cells in such a way that UPR is induced. There are a wide range of agents that can achieve this. For example, it has been shown that disruption of the lipid balance in cells in such a way that lipid saturation levels are increased (and thus desaturated lipid levels are decreased) results in induction of the UPR. Accordingly, suitably the UPR-inducing chemical agent is able to alter the lipid balance of the cells in such a way that lipid saturation levels are increased.

Preferably the UPR-inducing chemical agent is able to alter the ratio of saturated fatty acids to unsaturated fatty acids in cell membranes such that the proportion of saturated fatty acids is increased. This can be achieved in several ways, e.g. introducing saturated fatty acids to the cells or inhibiting the activity of enzymes which convert saturated fatty acids to unsaturated fatty acids.

Thus, in one particularly preferred embodiment the UPR-inducing chemical agent comprises a saturated fatty acid, suitably a medium or long chain saturated fatty acid. In certain embodiments of the invention the fatty acid has an aliphatic chain length of from 6 to 26 carbons, more preferably from 9 to 22 carbons, yet more preferably from 12 to 20 carbons, and yet more preferably from 14 to 20 carbons.

Suitably the UPR-inducing chemical agent comprises at least one fatty acid selected from the list consisting of: Caproic acid, Enanthic acid, Caprylic acid, Pelargonic acid, Capric acid, Undecylic acid, Lauric acid, Tridecylic acid, Myristic acid, Pentadecylic acid, Palmitic acid, Margaric acid, Stearic acid, Nonadecylic acid, Arachidic acid, Heneicosylic acid, Behenic acid, Tricosylic acid, Lignoceric acid, Pentacosylic acid, and Cerotic acid. More preferably the UPR-inducing chemical agent comprises at least one fatty acid selected from the list consisting of: Pelargonic acid, Capric acid, Undecylic acid, Lauric acid, Tridecylic acid, Myristic acid, Pentadecylic acid, Palmitic acid, Margaric acid, Stearic acid, Nonadecylic acid, Arachidic acid, Heneicosylic acid, and Behenic acid. Yet more preferably the UPR-inducing chemical agent comprises at least one fatty acid selected from the list consisting of: Lauric acid, Tridecylic acid, Myristic acid, Pentadecylic acid, Palmitic acid, Margaric acid, Stearic acid, Nonadecylic acid and Arachidic acid. Yet more preferably the UPR-inducing chemical agent comprises at least one fatty acid selected from the list consisting of: Myristic acid, Pentadecylic acid, Palmitic acid, Margaric acid, Stearic acid, Nonadecylic acid and Arachidic acid.

In a particularly preferred embodiment the UPR-inducing chemical agent comprises palmitic acid or stearic acid.

5 It should be noted that where fatty acids are referred to, these can be provided in any suitable form, e.g. as a salt (palmitate, stearate or suchlike).

In another preferred embodiment, the UPR-inducing chemical agent comprises an agent that is able to downregulate stearoyl-CoA desaturase enzyme activity in the cell. Stearoyl-CoA desaturase is an ER resident enzyme which introduces a double-bond in saturated fatty
10 acids. For example, the UPR-inducing chemical agent suitably comprises an inhibitor of stearoyl-CoA desaturase. MF-43 is an example of a suitable inhibitor of Stearoyl-CoA desaturase. Alternatively, the UPR-inducing chemical agent may be able to downregulate expression of stearoyl-CoA desaturase, e.g. through knockdown of stearoyl-CoA desaturase expression (e.g. via RNA interference), or knockout of the stearoyl-CoA desaturase gene.
15 Other genes involved in desaturation of fatty acids can also be targeted.

Other exemplary UPR-inducing chemical agents include geranylgeranoic acid (GGA), 2,3-dihydro GGA, 9-cis retinoic acid and all-trans retinoic acid (see - Chieko Iwao and Yoshihiro Shidoj, PLOS ONE | DOI:10.1371/journal.pone.0132761 July 17, 2015).

20

Combinations of several UPR-inducing chemical agents can be used in some embodiments of the invention. For example, a saturated fatty acid (e.g. palmitic acid) can be used in combination with an agent that is able to downregulate stearoyl-CoA desaturase enzyme activity (e.g. MF-43).

25

In another embodiment of the present invention, inducing the UPR in step b) suitably comprises expressing an inducer protein in said population of eukaryotic cells so as to induce the UPR response in the population of cells. Such a protein is referred to as an "inducer protein" as when it is expressed in the population of cells it acts to induce the UPR
30 response, typically by creating ER stress. The inducer protein is typically a different protein to that encoded by the synthetic nucleic acid expression construct according to the first aspect of the invention. The inducer protein is suitably a heterologous protein, but in some embodiments it could be a homologous protein that is overexpressed. What matters is that expression of the inducer protein in the cells induces the UPR, which results in splicing of
35 the regulatable intron. There are many ways in which expression of an inducer protein can be achieved in the population of cells. In suitable embodiments, the population of cells is transfected with an expression vectors that is adapted to express the inducer protein in the

cells. In one embodiment, the cells can be infected with a virus which results in expression of viral proteins that lead to ER stress and induction of the UPR. A recombinant viral vector encoding viral or non-viral proteins could also be used, and an example is described below in which AAV is expressed in the cells to induce the UPR; this forms one preferred

embodiment of the present invention. Alternatively, essentially any other form of expression vector (e.g. a plasmid) could be introduced into the cells in order to express a heterologous protein. Suitable methods of transfecting cells in with suitable expression vectors are well-known in the art. The nature of the inducer protein is typically not of particular concern, though it is generally preferable that the protein is non-toxic; rather, it is the ER-stress that heterologous protein production creates in the cell that is of consequence.

In some embodiments step b) comprises transfecting said population of cells with an expression vector that is capable of expressing an inducer protein, preferably a heterologous protein, in said cells. Alternatively, step b) can comprise inducing expression of the inducer protein from an expression vector that was previously introduced into the cells, e.g. prior to step a).

Expression of the inducer protein can be under control of a constitutive or non-constitutive promoter. An exemplary non-constitutive promoter is an inducible promoter (in this case the inducible promoter would not be a UPR-inducible promoter).

Other methods of inducing the unfolded protein response comprise exposing the cells to hypoxia, or carbohydrate (e.g. glucose) deprivation.

As mentioned above, in view of the ubiquitous nature of IRE1-mediated intron slicing across the eukaryotes, the method can be performed with any type of eukaryotic cell. Accordingly the method can be carried out, for example, in fungal cells (e.g. yeast cells), animal (metazoan) cells (e.g. mammalian cells), and plant cells.

In certain preferred embodiments, the population of eukaryotic cells is a population of animal (metazoan) cells. Suitably the animal cells can be cells from an invertebrate animal or a vertebrate animal.

In some preferred embodiments, the population of eukaryotic cells is a population of mammalian cells. There a wide range of mammalian cells that can be used, including without limitation Chinese hamster ovary (CHO), human embryonic kidney (HEK) cells (e.g. HEK-293), human embryonic retinal cells, human amniocyte cells, Mouse myeloma

lymphoblastoid cells. In such embodiments, it may be preferred that the regulatable intron is an intron which comprises the sequence CNG/CNG[CG] at one, other, or both (preferably both) ends of the intron, i.e. the mammalian splice target consensus sequence. Suitably, for example, the regulatable intron has the sequence

- 5 CNG/CNGCACUCAGACUACGUGCACCUCNG/CNGC (SEQ ID NO: 6),
 CAG/CAGCACUCAGACUACGUGCACCUCUG/CUGC (SEQ ID NO: 7) or
 CAG/CUGCCAGCACUCAGACUACGUGCACCUCUG/CUGC (SEQ ID NO: 17).

In other embodiments, the population of eukaryotic cells is a population of insect cells.

- 10 Suitable insect cells for use in the method include baculovirus infected and non-infected cell, such as cells from the following insect species: *Spodoptera frugiperda* (e.g. Sf9 or Sf21), *Trichoplusia ni* (e.g. Hi-5), *Drosophila melanogaster* (e.g. Schneider 2 cells and Schneider 3 cells).

- 15 In other embodiments of the present invention the population of eukaryotic cells is suitably a population of fungal cells, preferably yeast cells. Suitable fungal cells for use in the present method include without limitation *Saccharomyces cerevisiae*, *Pichia pastoris*, *Aspergillus spp.*, *Trichoderma spp.* and *Myceliophthora thermophila*.

- 20 In other embodiments of the present invention, the population of eukaryotic cells is suitably a population of plant cells or plant protoplasts.

In other embodiments of the present invention the population of eukaryotic cells is suitably a population of protozoan cells, e.g. *Leishmania tarentolae*.

25

Step d), i.e. isolating the expression product from said population of cells, can be carried out using conventional techniques well-known in the art. Such techniques will of course vary depending on the nature of the expression product.

- 30 The method may suitably comprise the step of introducing the nucleic acid expression construct into the cells. There are many well-known methods of transfecting eukaryotic cells, and the skilled person could readily select a suitable method for any cell type. The nucleic acid expression construct can of course be provided in any suitable vector.

- 35 In a further aspect, the invention provides the nucleic acid expression constructs, vectors, cells or pharmaceutical compositions according to the various aspects of the present invention for use in a method of treatment or therapy.

As used herein, the terms "treat" or "treatment" refer to both therapeutic treatment and prophylactic or preventative measures. Beneficial or desired clinical results include, but are not limited to, prevention of an undesired clinical state or disorder, reducing the incidence of a disorder, alleviation of symptoms associated with a disorder, diminishment of extent of a disorder, stabilized (i.e., not worsening) state of a disorder, delay or slowing of progression of a disorder, amelioration or palliation of the state of a disorder, remission (whether partial or total), whether detectable or undetectable, or combinations thereof. "Treatment" can also mean prolonging survival as compared to expected survival if not receiving treatment.

As used herein, the terms "therapeutic treatment" or "therapy" and the like, refer to treatments wherein the object is to bring a subject's body or an element thereof from an undesired physiological change or disorder to a desired state, such as a less severe or unpleasant state (e.g., amelioration or palliation), or back to its normal, healthy state (e.g., restoring the health, the physical integrity and the physical well-being of a subject), to keep it at said undesired physiological change or disorder (e.g., stabilization, or not worsening), or to prevent or slow down progression to a more severe or worse state compared to said undesired physiological change or disorder.

As used herein the terms "prevention", "preventive treatment" or "prophylactic treatment" and the like encompass preventing the onset of a disease or disorder, including reducing the severity of a disease or disorder or symptoms associated therewith prior to affliction with said disease or disorder. Such prevention or reduction prior to affliction refers to administration of the nucleic acid expression constructs, vectors, or pharmaceutical compositions described herein to a patient that is not at the time of administration afflicted with clear symptoms of the disease or disorder. "Preventing" also encompasses preventing the recurrence or relapse-prevention of a disease or disorder for instance after a period of improvement. In embodiments, the nucleic acid expression constructs, vectors, or pharmaceutical compositions described herein may be for use in gene therapy.

The present invention also provides the use of the nucleic acid expression constructs, the vectors, or the pharmaceutical compositions described herein for the manufacture of a medicament for gene therapy.

Also disclosed herein is a method for gene therapy in a subject in need of said gene therapy comprising:

- introducing into the subject a gene therapy vector comprising a pharmaceutical composition comprising a nucleic acid expression construct according to the present invention, the nucleic acid expression construct comprising a sequence encoding a therapeutic expression product such that the gene therapy vector delivers the nucleic acid expression construct to target cells of the subject; and
- expressing a therapeutically effective amount of the functional therapeutic expression product in target cells of subject.

It will be apparent that the expression of a therapeutically effective amount of the functional therapeutic expression product will only occur when the expression product is expressed in a cell where the UPR is active. In many cases the UPR will be active in cells which are under stress, e.g. cells which are cancerous or which are infected with a pathogen (e.g. a virus). Accordingly, it is an advantage of the present invention that expression of the functional therapeutic product (which could be, for example, a toxic protein or other cytotoxic agent) will only occur in cells where the UPR is active. This can be useful to mitigate or avoid undesirable off-target expression of the functional therapeutic expression product.

Accordingly, in the various aspects of the invention directed to therapy, it is preferred that the condition to be treated is cancer or infection (e.g. infection with a virus).

Alternatively, UPR can be induced in cells using a suitable UPR-inducing agent. Various UPR-inducing agents are discussed above, and suitable UPR-inducing agents include pharmaceutically acceptable agents that induce the UPR (e.g. Bortezomib, forskolin) and agents that disturb the lipid balance in the cells to induce the UPR (e.g. saturated fatty acids as discussed above). The UPR-inducing agent can be delivered directly to a target site (e.g. by injection) or given systemically.

The therapeutic expression product may be a polypeptide/protein, e.g. a secretable protein or peptide such as, e.g., clotting factors, such as factor VII I or factor IX, insulin, erythropoietin, lipoprotein lipase, antibodies or nanobodies, growth factors, cytokines, chemokines, plasma factors, toxic proteins, etc. Alternatively, the therapeutic expression product may be RNA, such as siRNA or miRNA.

In some preferred embodiments, the therapeutic expression product may be toxic for the cells into which the gene therapy vector has been introduced. In such embodiments, induction of the UPR can be used to induce toxicity in cells, e.g. leading to cell death. Accordingly, induction of the UPR and consequent expression of the therapeutic expression

product can be used to induce a kill switch in said cells. Suitable toxic expression products, such as proteins, are well known in the art, and mention can be made of caspase 3, caspase 8, and caspase 9, for example.

5 Suitable gene therapy vectors for use in this aspect of the invention are discussed above.

Gene therapy protocols have been extensively described in the art. These include, but are not limited to, intramuscular injection of a suitable vector, hydrodynamic gene delivery in various tissues, including muscle, interstitial injection, instillation in airways, application to
10 endothelium, intra-hepatic parenchyme, and intravenous or intra-arterial administration. Various devices have been developed for enhancing the availability of DNA to the target cell. A simple approach is to contact the target cell physically with catheters or implantable materials containing DNA. Another approach is to utilize needle-free, jet injection devices which project a column of liquid directly into the target tissue under high pressure. These
15 delivery paradigms can also be used to deliver vectors. Another approach to targeted gene delivery is the use of molecular conjugates, which consist of protein or synthetic ligands to which a nucleic acid-or DNA-binding agent has been attached for the specific targeting of nucleic acids to cells (Cristiano et al., 1993).

20 The terms "subject" and "patient" are used interchangeably herein and refer to animals, preferably vertebrates, more preferably mammals, and specifically include human patients and non-human mammals. "Mammalian" subjects include, but are not limited to, humans. Preferred patients or subjects are human subjects.

25 A "therapeutic amount" or "therapeutically effective amount" as used herein refers to the amount of expression product effective to treat a disease or disorder in a subject, i.e., to obtain a desired local or systemic effect. The term thus refers to the quantity of an expression product that elicits the biological or medicinal response in a tissue, system, animal, or human that is being sought by a researcher, veterinarian, medical doctor or other
30 clinician. Such amount will typically depend on the gene product and the severity of the disease, but can be decided by the skilled person, possibly through routine experimentation.

Expression levels of the expression product (e.g. protein) can be measured by various conventional means, such as by antibody-based assays, e.g. a Western Blot or an ELISA
35 assay, for instance to evaluate whether therapeutic expression of the expression product is achieved. Expression of the expression product may also be measured in a bioassay that detects an enzymatic or biological activity of the gene product.

In a further aspect the present invention provides a synthetic UPR-inducible promoter comprising a synthetic UPR-responsive cis-regulatory element. Suitably the UPR-responsive cis-regulatory element comprises at least one binding site for ATF6, XBP1 or bZIP60, or homologous or otherwise equivalent transcription factors that drive gene expression as part of the UPR.

Such a promoter can be used to selectively drive expression of a desired expression product in eukaryotic cells upon induction of the UPR. While it may be advantageous in terms of tight control of gene expression that such a promoter is used in combination with a regulatable intron discussed above, in other circumstances a UPR-inducible promoter could be used without the regulatable intron.

A UPR-responsive cis-regulatory element is a sequence that contains functional transcription factor binding site(s) (TFBS) for one or more transcription factors that drive gene expression as part of the UPR. As discussed above, these include, but are not limited to, ATF6, XBP1 and bZIP60. UPR-responsive cis-regulatory elements comprising TFBS for ATF6 are of particular interest in the present invention. Preferably the UPR-responsive cis-regulatory element is UPR-specific, i.e. it only enhances expression during the UPR. For example, it is generally preferred that it does not contain any TFBS for transcription factors that are not involved in the UPR.

The synthetic UPR-inducible promoter typically comprises at least one synthetic UPR-responsive cis-regulatory element operably linked to a minimal promoter or proximal promoter. Where the cis-regulatory element is operably linked to a proximal promoter, the proximal promoter should itself be a UPR-inducible promoter that does not drive transcription of an operably linked gene in a eukaryotic cell when the UPR has not been induced.

Minimal promoters cannot typically drive expression without the presence of additional regulatory elements. Examples of suitable minimal promoters for use in the present invention include, but are not limited to the CMV-minimal (SEQ ID NO: 21) promoter and the MinTk minimal promoter. Other suitable minimal promoters are known in the art.

Suitably the UPR-inducible promoter comprises a UPR-responsive cis-regulatory element comprising one or more copies of the at least one of the following transcription factor target sequences:

- TGACGTG (the ATF6 transcription factor binding site consensus sequence),
- TGACGTGCT (a variant of the above)

- TGACGTG[TG] (known as the UPRE site),
- CCAAT-N9-CCACG (known as the ERSE1 site) (SEQ ID NO: 18), and
- ATTGG-N-CCACG (known as the ERSE2 site) (SEQ ID NO: 19).

5 Suitably the synthetic UPR-responsive cis-regulatory element comprises two or more, preferably three or more, suitably five or more, copies of at least one of the transcription factor target sequences listed above. Alternatively, or additionally, the synthetic UPR-responsive cis-regulatory element suitably comprises one or more copies of at least two of the transcription factor target sequences listed above.

10

The transcription factor target sequences may be directly adjacent to each other (tandem repeats) or may be spaced apart, for example by a spacer sequence or another functional sequence (e.g. another transcription factor target sequence). Typically, the spacer sequence, if present, is from 5-50 nucleotides in length, but it can be longer or shorter in some cases. For example, the spacer sequence is suitably from 2 to 50 nucleotides in length, suitably from 4 to 30 nucleotides in length, or suitably from 5 to 20 nucleotides in length. It may be preferred that the spacer sequence is a multiple of 5 nucleotides in length, as this provides an integer number of half-turns of the DNA double helix (a full turn corresponding to approximately 10 nucleotides in chromatin). A spacer sequence length that is a multiple of 10 nucleotides in length may be more preferable, as it provides an integer number of full-turns of the DNA double helix. The spacer sequence can have essentially any sequence, provided it does not prevent the UPR-responsive cis-regulatory element from functioning as desired (e.g. it includes a silencer sequence, prevents binding of the desired transcription factor, or suchlike). The spacer sequences between each transcription factor target sequence can be identical or they can be different.

25

In a preferred embodiment the UPR-responsive cis-regulatory element comprises one or more copies of the transcription factor target sequence TGACGTG (i.e. the ATF6 consensus sequence), preferably 3 or more copies of the transcription factor target sequence TGACGTG, and preferably 5 or more copies of the transcription factor target sequence TGACGTG, for example 6 or more copies of the transcription factor target sequence TGACGTG. As mentioned above, these transcription factor target sequences may be in tandem repeat or may be spaced from each other. Generally, it is preferred that at least two, and preferably all, of the transcription factor target sequences present in the UPR-responsive cis-regulatory element are spaced from each other, e.g. by a spacer sequence as discussed above.

35

Suitably the UPR-responsive cis-regulatory element comprises one or more copies of the transcription factor target sequence TGACGTGCT, preferably 3 or more copies of the transcription factor target sequence TGACGTGCT, preferably 5 or more copies of the transcription factor target sequence TGACGTGCT, for example 6 or more copies of the transcription factor target sequence TGACGTGCT. As mentioned above, these may be in tandem repeat, or may be spaced from each other. Generally, it is preferred that at least two, and preferably all, of the transcription factor target sequences present in the UPR-responsive cis-regulatory element are spaced from each other, e.g. by a spacer sequence as discussed above. The transcription factor target sequence TGACGTGCT has been found to be particularly effective when used in multiple copy number in a UPR-responsive cis-regulatory element, whether as a tandem repeat or including spacer sequences.

In some embodiments of the present invention the UPR-responsive cis-regulatory element comprises the sequence

TGACGTG-S-TGACGTG-S-TGACGTG-S-TGACGTG-S-TGACGTG-S-TGACGTG (SEQ ID NO: 54)

where S represents an optional spacer sequence as defined above. Preferably spacer sequences as defined above are present between at least two, and preferably all, of the transcription factor target sequences (TGACGTG).

In some embodiments of the present invention the UPR-responsive cis-regulatory element comprises the sequence

TGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCT (SEQ ID NO: 20).

In other embodiments of the present invention the UPR-responsive cis-regulatory element comprises the sequence

TGACGTGCT-S-TGACGTGCT-S-TGACGTGCT-S-TGACGTGCT-S-TGACGTGCT-S-TGACGTGCT (SEQ ID NO: 55)

where S represents an optional spacer sequence as defined above. Preferably spacer sequences as defined above are present between at least two, and preferably all, of the transcription factor target sequences.

In other embodiments of the present invention the UPR-responsive cis-regulatory element comprises the sequence

TGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTG
ATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTA
GCTAGTAGTTGACGTGCT (SEQ ID NO: 56)

or a sequence that is at least 50% identical thereto, yet more preferably at least 70%

- 5 identical thereto, yet more preferably at least 80% identical thereto, and yet more preferably at least 85%, 90%, 995%, 98% or 99% identical thereto. It is highly preferred that sequence variation only occurs in sequences which are not the transcription factor target sequences, i.e. those having the sequence TGACGTGCT. It is generally preferred that sequence variation only occurs in spacer sequences (i.e. those having the sequence
- 10 GATGATGCGTAGCTAGTAGT (SEQ ID NO: 61)).

In some embodiments of the present invention the UPR-inducible promoter comprises the following sequence

- TGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCTGGTACCGTCGACGA
15 TATCGGATCCAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTAGATACGCCATCC
ACGCTGTTTTGACCTCCATAGAAGATCGCCACC (SEQ ID NO: 22)

or a sequence that is at least 70% identical thereto, yet more preferably at least 80% identical thereto, yet more preferably at least 90% identical thereto, and yet more preferably at least 95%, 96%, 97%, 98% or 99% identical thereto. This UPR-inducible promoter

20 comprises the UPR-responsive cis-regulatory element of SEQ ID NO: 20 operably linked to the CMV-MP minimal promoter. It is highly preferred that sequence variation only occurs in sequences which are not the transcription factor target sequences, i.e. those having the sequence TGACGTGCT, nor in the CMV-MP sequence.

- 25 In other embodiments of the present invention the UPR-inducible promoter comprises the following sequence:

- TGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTG
ATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTA
GCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTGCAGTTAGCGTAGCTGAGGTACCGTCGACG
30 ATATCGGATCCAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGAT (SEQ ID NO: 57)

or a sequence that is at least 50% identical thereto, yet more preferably at least 70% identical thereto, yet more preferably at least 80% identical thereto, and yet more preferably at least 85%, 90%, 995%, 98% or 99% identical thereto. This UPR-inducible promoter comprises the UPR-responsive cis-regulatory element of SEQ ID NO: 56 operably linked to

35 the CMV-MP minimal promoter. It is highly preferred that sequence variation only occurs in sequences which are not the transcription factor target sequences, i.e. those having the sequence TGACGTGCT, nor in the CMV-MP sequence. It is generally preferred that

sequence variation only occurs in spacer sequences (i.e. those having the sequence GATGATGCGTAGCTAGTAGT (SEQ ID NO: 61)).

In other embodiments of the present invention the UPR-inducible promoter comprises the following sequence:

TGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTG
ATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTA
GCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTGACGTTAGCGTAGCTGAGGTACCGTCGACG
ATATCGGATCCTTCGCATATTAAGGTGACGCGTGTGGCCTCGAACACCGAG (SEQ ID NO: 58)

or a sequence that is at least 50% identical thereto, yet more preferably at least 70% identical thereto, yet more preferably at least 80% identical thereto, and yet more preferably at least 85%, 90%, 995%, 98% or 99% identical thereto. This UPR-inducible promoter comprises the UPR-responsive cis-regulatory element of SEQ ID NO: 56 operably linked to the MinTK minimal promoter. It is highly preferred that sequence variation only occurs in sequences which are not the transcription factor target sequences, i.e. those having the sequence TGACGTGCT, nor in the MinTK sequence. It is generally preferred that sequence variation only occurs in spacer sequences (i.e. those having the sequence GATGATGCGTAGCTAGTAGT (SEQ ID NO: 61)).

The UPR-inducible promoter preferably does not drive transcription of an operably linked gene when present in a eukaryotic cell in the absence of the UPR. The UPR-inducible promoter drives transcription of an operably linked gene when present in a eukaryotic cell when the UPR is occurring in the cell. Assessment of the ability of a UPR-inducible promoter to selectively drive transcription upon induction of the UPR can readily be assessed by the skilled person using a wide range of approaches, and these can be tailored for the particular expression system in which the construct is intended to be used. As one preferred example, the methodology described in the examples below can be used, e.g. Example 8. For example, any candidate UPR-inducible promoter to be assessed can be substituted into the construct described in Example 8 in place of the exemplary ATF6-containing UPR-inducible promoter used in Example 8, and the ability of said candidate UPR-inducible promoter to selectably drive transcription when the UPR is induced can be measured by assessing the level of luciferase expression before and after UPR induction by 2mM DTT, exactly as carried out in Example 8. A UPR-inducible promoter is one which is able to be successfully induced to significantly increase transcription of an operably linked gene (in the case of Example 8, the luciferase gene) upon induction of the UPR to result in the expression of the gene. Preferably a UPR-inducible promoter confers at least a 5-fold increase in expression 24 hours after induction of the UPR with 2mM DTT, more preferably

at least a 10-fold increase in expression, more preferably at least a 100-fold increase in expression, and yet more preferably at least a 1000-fold increase in expression of the gene (e.g. luciferase). It is preferred that before induction of the UPR the expression levels of the gene (e.g. luciferase) are minimal, and preferably negligible. Minimal expression can be defined as, for example, equal to or less than the expression levels of a negative control construct as used in Example 8 (i.e. a construct without in which expression of the sequence encoding the luciferase is driven by CMV-MP alone), preferably less than 50%, preferably less than 20%, more preferably less than 10%, yet more preferably less than 5%, yet more preferably less than 1% of the expression levels of a negative control construct. Negligible expression levels are, for example, those that are essentially undetectable using the methodology of Example 8.

The present invention also provides an expression construct or vector comprising a synthetic UPR-inducible promoter as set out above, operably linked to a nucleic acid sequence encoding an expression product. The expression construct or vector can be any expression construct or vector as discussed above for the other aspects of the invention. The expression product can be any expression product as discussed above for the other aspects of the invention (e.g. encoding a protein).

In preferred embodiments the expression product is not a reporter protein, i.e. it does not encode a protein that is used conventionally as an indicator of expression levels. Many reporter genes are known in the art, including, in particular, fluorescent, luminescent proteins and chromogenic proteins. Thus, in some preferred embodiments the expression product is not a fluorescent or luminescent protein, e.g. it is not a luciferase. As set out above, preferred expression products include therapeutic proteins and toxic proteins.

In a further aspect the present invention provides a method for producing an expression product, the method comprising:

- a) providing a population of eukaryotic cells comprising a synthetic nucleic acid expression construct comprising a UPR-inducible promoter operably linked to a nucleic acid sequence encoding an expression product according to the present invention;
- b) treating said population of cells so as to induce the unfolded protein response, thereby inducing transcription from the UPR-inducible promoter;
- c) incubating said population of cells under suitable conditions for production of the expression product; and
- d) isolating the expression product from said population of cells.

Further optional and preferred features of methods for producing an expression product are discussed above for the other aspects of the invention, and these apply to the present aspect *mutatis mutandis*. It is preferred that the expression product is a therapeutic protein or a toxic protein. It is preferred that the expression product is not a reporter protein

Accordingly, a further aspect of the invention provides a pharmaceutical composition comprising a nucleic acid expression construct or a vector comprising a UPR-inducible promoter operably linked to a nucleic acid sequence encoding an expression product according to the present invention. Further optional and preferred features of pharmaceutical composition are discussed above for the other aspects of the invention, and these apply to the present aspect *mutatis mutandis*.

In a further aspect of the invention there is provided the use of nucleic acid expression constructs and vectors comprising a UPR-inducible promoter operably linked to a nucleic acid sequence encoding an expression product according to the present invention for the manufacture of a pharmaceutical composition.

According to a further aspect of the present invention there is provided a cell comprising a synthetic nucleic acid expression construct or vector comprising a UPR-inducible promoter according to the present invention. Further optional and preferred features of such cells are discussed above for the other aspects of the invention, and these apply to the present aspect *mutatis mutandis*.

In a further aspect, the invention provides the nucleic acid expression constructs, vectors, cells or pharmaceutical compositions comprising a UPR-inducible promoter according to the present invention for use in a method of treatment or therapy. Further optional and preferred features of such methods are discussed above for the other aspects of the invention, and these apply to the present aspect *mutatis mutandis*.

To facilitate the understanding of this invention, a number of terms are defined below. Terms defined herein have meanings as commonly understood by a person of ordinary skill in the areas relevant to the present invention. The terminology herein is used to describe specific embodiments of the invention, but their usage does not delimit the invention, except as set out in the claims.

The discussion of the background to the invention herein is included to explain the context of the invention. This is not to be taken as an admission that any of the material referred to was published, known, or part of the common general knowledge in any country as of the priority date of any of the claims.

5

Throughout this disclosure, various publications, patents and published patent specifications are referenced by an identifying citation. All documents cited in the present specification are hereby incorporated by reference in their entirety. In particular, the teachings or sections of such documents herein specifically referred to are incorporated by reference.

10

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, Current Protocols in Molecular Biology (Ausubel, 2000, Wiley and son Inc, Library of Congress, USA); Molecular Cloning: A Laboratory Manual, Third Edition, (Sambrook et al, 2001, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press); Oligonucleotide Synthesis (M. J. Gait ed., 1984); U.S. Pat. No. 4,683,195; Nucleic Acid Hybridization (Harries and Higgins eds. 1984); Transcription and Translation (Hames and Higgins eds. 1984); Culture of Animal Cells (Freshney, Alan R. Liss, Inc., 1987); Immobilized Cells and Enzymes (IRL Press, 1986); Perbal, A Practical Guide to Molecular Cloning (1984); the series, Methods in Enzymology (Abelson and Simon, eds. -in-chief, Academic Press, Inc., New York), specifically, Vols. 154 and 155 (Wu et al. eds.) and Vol. 185, "Gene Expression Technology" (Goeddel, ed.); Gene Transfer Vectors For Mammalian Cells (Miller and Calos eds., 1987, Cold Spring Harbor Laboratory); Immunochemical Methods in Cell and Molecular Biology (Mayer and Walker, eds., Academic Press, London, 1987); Handbook of Experimental Immunology, Vols. I-IV (Weir and Blackwell, eds., 1986); and Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

30

Terms such as "a", "an" and "the" are not intended to refer to only a singular entity, but include the general class of which a specific example may be used for illustration.

35

The terms "comprising", "comprises" and "comprised of" as used herein are synonymous with "including", "includes" or "containing", "contains", and are inclusive or open-ended and do not exclude additional, non-recited features, elements or method steps.

The recitation of numerical ranges by endpoints includes all numbers and fractions subsumed within the respective ranges, as well as the recited endpoints.

The term "nucleic acid" as used herein typically refers to an oligomer or polymer (preferably a linear polymer) of any length composed essentially of nucleotides. A nucleotide unit commonly includes a heterocyclic base, a sugar group, and at least one, e.g. one, two, or three, phosphate groups, including modified or substituted phosphate groups. Heterocyclic bases may include inter alia purine and pyrimidine bases such as adenine (A), guanine (G), cytosine (C), thymine (T) and uracil (U) which are widespread in naturally-occurring nucleic acids, other naturally-occurring bases (e.g., xanthine, inosine, hypoxanthine) as well as chemically or biochemically modified (e.g., methylated), non-natural or derivatised bases. Sugar groups may include inter alia pentose (pentofuranose) groups such as preferably ribose and/or 2-deoxyribose common in naturally-occurring nucleic acids, or arabinose, 2-deoxyarabinose, threose or hexose sugar groups, as well as modified or substituted sugar groups. Nucleic acids as intended herein may include naturally occurring nucleotides, modified nucleotides or mixtures thereof. A modified nucleotide may include a modified heterocyclic base, a modified sugar moiety, a modified phosphate group or a combination thereof. Modifications of phosphate groups or sugars may be introduced to improve stability, resistance to enzymatic degradation, or some other useful property. The term "nucleic acid" further preferably encompasses DNA, RNA and DNA RNA hybrid molecules, specifically including hnRNA, pre-mRNA, mRNA, cDNA, genomic DNA, amplification products, oligonucleotides, and synthetic (e.g., chemically synthesised) DNA, RNA or DNA RNA hybrids. A nucleic acid can be naturally occurring, e.g., present in or isolated from nature; or can be non-naturally occurring, e.g., recombinant, i.e., produced by recombinant DNA technology, and/or partly or entirely, chemically or biochemically synthesised. A "nucleic acid" can be double-stranded, partly double stranded, or single-stranded. Where single-stranded, the nucleic acid can be the sense strand or the antisense strand. In addition, nucleic acid can be circular or linear.

"Nucleic acid expression construct" as used herein refers to nucleic acid molecules that include one or more transcriptional control elements (such as, but not limited to promoters, enhancers and/or regulatory elements, polyadenylation sequences, and introns) that direct expression in one or more desired cell types, tissues or organs. Nucleic acid expression constructs of the present invention are synthetic nucleic acid molecules.

"Synthetic" in the present application means a nucleic acid molecule that does not occur in nature. Synthetic nucleic acid expression constructs of the present invention are produced

artificially, typically by recombinant technologies. Such synthetic nucleic acids may contain naturally occurring sequences (e.g. promoter, enhancer, intron, and other such regulatory sequences), but these are present in a non-naturally occurring context. For example, a synthetic gene (or portion of a gene) typically contains one or more nucleic acid sequences that are not contiguous in nature (chimeric sequences), and/or may encompass substitutions, insertions, and deletions and combinations thereof.

The term “operably linked”, “operably connected” or equivalent expressions as used herein refer to the arrangement of various nucleic acid elements relative to each such that the elements are functionally connected and are able to interact with each other in the manner intended. Such elements may include, without limitation, a promoter, an enhancer and/or a regulatory element, a polyadenylation sequence, one or more introns and/or exons, and a coding sequence of a gene of interest to be expressed. The nucleic acid sequence elements, when properly oriented or operably linked, act together to modulate the activity of one another, and ultimately may affect the level of expression of an expression product. By modulate is meant increasing, decreasing, or maintaining the level of activity of a particular element. The position of each element relative to other elements may be expressed in terms of the 5' terminus and the 3' terminus of each element, and the distance between any particular elements may be referenced by the number of intervening nucleotides, or base pairs, between the elements. As understood by the skilled person, operably linked implies functional activity, and is not necessarily related to a natural positional link. Indeed, when used in nucleic acid expression cassettes, cis-regulatory elements will typically be located immediately upstream of the promoter (although this is generally the case, it should definitely not be interpreted as a limitation or exclusion of positions within the nucleic acid expression cassette), but this needs not be the case in vivo. E.g., a regulatory element sequence naturally occurring downstream of a gene whose transcription it affects is able to function in the same way when located upstream of the promoter. Hence, according to a specific embodiment, the regulatory or enhancing effect of the regulatory element is position-independent.

“Consensus sequence” – the meaning of consensus sequence is well-known in the art. In the present application, the following notation is used for the consensus sequences, unless the context dictates otherwise. Considering the following exemplary DNA sequence:

A[CT]N{A}YR

A means that an A is always found in that position; [CT] stands for either C or T in that position; N stands for any base in that position; and {A} means any base except A is found in that position. Y represents any pyrimidine, and R indicates any purine.

"Regulatable intron" in the present application, unless the context dictates otherwise, refers to a nucleic acid sequence occurring in an RNA molecule (typically a transcript) comprising an excisable sequence flanked by target sites for a ribonuclease, generally IRE1 or a
5 homologue or orthologue thereof, wherein the excisable sequence can be excised from the RNA molecule by the action of said ribonuclease as a result of the unfolded protein response. In some cases in the art the term intron is used to refer exclusively to the sequence which is excised (i.e. spliced out) from the RNA molecule, but in the case of the regulatable introns of the present invention that is less appropriate.

10 The terms "identity" and "identical" and the like refer to the sequence similarity between two polymeric molecules, e.g., between two nucleic acid molecules, e.g., two DNA molecules. Sequence alignments and determination of sequence identity can be done, e.g., using the Basic Local Alignment Search Tool (BLAST) originally described by Altschul et al. 1990 (J
15 Mol Biol 215: 403-10), such as the "Blast 2 sequences" algorithm described by Tatusova and Madden 1999 (FEMS Microbiol Lett 174: 247-250).

Methods for aligning sequences for comparison are well-known in the art. Various programs and alignment algorithms are described in, for example: Smith and Waterman (1981) Adv.
20 Appl. Math. 2:482; Needleman and Wunsch (1970) J. Mol. Biol. 48:443; Pearson and Lipman (1988) Proc. Natl. Acad. Sci. U.S.A. 85:2444; Higgins and Sharp (1988) Gene 73:237-44; Higgins and Sharp (1989) CABIOS 5:151-3; Corpet et al. (1988) Nucleic Acids Res. 16:10881-90; Huang et al. (1992) Comp. Appl. Biosci. 8:155-65; Pearson et al. (1994) Methods Mol. Biol. 24:307-31; Tatiana et al. (1999) FEMS Microbiol. Lett. 174:247-50. A
25 detailed consideration of sequence alignment methods and homology calculations can be found in, e.g., Altschul et al. (1990) J. Mol. Biol. 215:403-10.

The National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST™; Altschul et al. (1990)) is available from several sources, including the
30 National Center for Biotechnology Information (Bethesda, MD), and on the internet, for use in connection with several sequence analysis programs. A description of how to determine sequence identity using this program is available on the internet under the "help" section for BLAST™. For comparisons of nucleic acid sequences, the "Blast 2 sequences" function of the BLAST™ (Blastn) program may be employed using the default parameters. Nucleic acid
35 sequences with even greater similarity to the reference sequences will show increasing percentage identity when assessed by this method. Typically, the percentage sequence identity is calculated over the entire length of the sequence.

For example, a global optimal alignment is suitably found by the Needleman-Wunsch algorithm with the following scoring parameters: Match score: +2, Mismatch score: -3; Gap penalties: gap open 5, gap extension 2. The percentage identity of the resulting optimal
5 global alignment is suitably calculated by the ratio of the number of aligned bases to the total length of the alignment, where the alignment length includes both matches and mismatches, multiplied by 100.

“Cell culture”, as used herein, refers to a proliferating mass of cells that may be in either an
10 undifferentiated or differentiated state.

“Cis-regulatory element” or “CRE”, as used herein, is a term known to the skilled person; it relates to a region of non-coding DNA which regulates the transcription of a neighbouring gene (i.e. *in cis*). CREs typically regulate gene transcription by binding to transcription
15 factors. A CRE can be an enhancer, promoter, insulator or silence, for example. In the present case the UPR-inducible CRE is typically an enhancer element which binds to transcription factors that act to induce transcription as part of the UPR. In the present context, when the CRE is provided as part of a promoter and a gene encoding an expression product, it is preferred that the UPR-inducible CRE is located 1500 nucleotides or less from
20 the transcription start site (TSS), more preferably 1000 nucleotides or less from the TSS, more preferably 500 nucleotides or less from the TSS, and suitably 250, 200, 150, or 100 nucleotides or less from the TSS.

“Complementary” or “complementarity”, as used herein, refers to the Watson-Crick base-
25 pairing of two nucleic acid sequences. For example, for the sequence 5'-AGT-3' binds to the complementary sequence 3'-TCA-5'. Complementarity between two nucleic acid sequences may be “partial”, in which only some of the bases bind to their complement, or it may be complete as when every base in the sequence binds to its complementary base. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and
30 strength of hybridization between nucleic acid strands.

“Conventional intron” as used herein refers to introns that are conventionally present in pre-mRNA, typically spliceosomal introns, that do not encode information for protein synthesis and is removed from an mRNA molecule before translation of the mRNA. The term is used
35 to differentiate such introns from the non-canonical regulatory introns of the present invention.

"Transfection" in the present application refers broadly to any process of deliberately introducing nucleic acids into cells, and covers introduction of viral and non-viral vectors, and includes transformation, transduction and like terms. Examples include, but are not limited to: transfection with viral vectors; transformation with plasmid vectors; electroporation (Fromm et al. (1986) Nature 319:791-3); lipofection (Feigner et al. (1987) Proc. Natl. Acad. Sci. USA 84:7413-7); microinjection (Mueller et al. (1978) Cell 15:579-85); Agrobacterium-mediated transfer (Fraley et al. (1983) Proc. Natl. Acad. Sci. USA 80:4803-7); direct DNA uptake; whiskers-mediated transformation; and microprojectile bombardment (Klein et al. (1987) Nature 327:70).

As used herein, the phrase "transgene" refers to an exogenous nucleic acid sequence. In one example, a transgene is a gene sequence, a gene encoding an industrially or pharmaceutically useful compound, or a gene encoding a desirable trait. In yet another example, the transgene is an antisense nucleic acid sequence, wherein expression of the antisense nucleic acid sequence inhibits expression of a target nucleic acid sequence.

As used herein, the phrase "promoter" refers to a region of DNA that generally is located upstream of a nucleic acid sequence to be transcribed that is needed for transcription to occur. Promoters permit the proper activation or repression of transcription of sequence under their control. A promoter typically contains specific sequences that are recognized and bound by transcription factors, e.g. enhancer sequences. Transcription factors bind to the promoter DNA sequences and result in the recruitment of RNA polymerase, an enzyme that synthesizes RNA from the coding region of the gene. A great many promoters are known in the art.

As used herein, "minimal promoter" refers to a short DNA segment which is inactive or largely inactive by itself, but can mediate strong transcription when combined with other transcription regulatory elements. Minimum promoter sequence can be derived from various different sources, including prokaryotic and eukaryotic genes. Examples of minimal promoters are dopamine beta-hydroxylase gene minimum promoter and cytomegalovirus (CMV) immediate early gene minimum promoter (CMV-MP) and the herpes thymidine kinase minimal promoter (MinTK).

"RNA transcript" or "transcript" refers to the product resulting from RNA polymerase-catalysed transcription of a DNA sequence. When the RNA transcript is a, typically perfect, complementary copy of the DNA sequence, it is referred to as the primary transcript or it

may be a RNA sequence derived from post-transcriptional processing of the primary transcript and is referred to as the mature RNA.

"Messenger RNA" or "(mRNA)" refers to the processed form of the transcript RNA that is without introns and that can be translated into protein by the cell.

Embodiments of the present invention will now be described, by way of non-limiting example, with reference to the accompanying drawings.

10 Brief Description of the Figures

Fig. 1 shows ratio of GFP fluorescence from SYN-P-XBP-01 transfected cells before and after induction by 2mM DTT.

Fig. 2 shows in panel a) the ratio of GFP fluorescence positive control before and after induction. In panel b) there is shown fluorescence microscopy of SYN-P-XBP-01 with and without 2mM DTT.

Fig. 3 shows expression of GFP from both plasmids after induction by 2mM DTT. Figures are ratio to CMV-IE.

Fig. 4 shows raw data of expression from SYN-P-ATF6-01 before and after addition of 2mM DTT.

Fig. 5 shows raw data of expression from SYN-P-ATF6-02 before and after addition of 2mM DTT.

Fig. 6 shows fluorescence microscopy of HEK293F cells transfected with SYN-P-ATF6-01 or SYN-P-ATF6-02 before and after treatment with 2mM DTT.

Fig. 7 shows EGFP fluorescence from HEK293F cells before and after induction with wtAAV.

Zs Green is a recombinant AAV makes a variant of GFP that is used as a control, pre-AS-green is no transfection (i.e. blank).

Fig. 8 shows EGFP expression from HEK293F cells before and after wtAAV production.

Fig. 9 shows EGFP expression measured as a ratio to CMV-IE. Construct used was SYN-P-ATF6-02; contains intron and ATF6 inducible promoter. Transfections were performed on HEK293F cells using Maxcyte reagent. 24Hrs after transfection the inducers were added and GFP expression followed for the next 24hrs. Inducer concentrations; DTT 2mM, Palmitate 0.5mM, MF-43 1µM.

Fig. 10 shows a plasmid map of pMA-RQ.

Fig. 11 shows a graph of expression of SEAP after induction with DTT in CHO and HEK293 cells, compared with expression of SEAP under control of the CMV-IE promoter (results shown are in HEK293). (X-axis shows hours after induction)

Fig. 12 shows a graph of expression of luciferase after induction with DTT or palmitate in CHO (A) and HEK293 (B) cells (X-axis units are hours). For comparison, expression of luciferase under control of the CMV-IE promoter and CMV minimal promoter (CMV-MP) is shown.

5 Fig. 13 shows a plasmid map of SYN-P-CAS9_INT.

Fig. 14A shows a graph of HEK293 cell viability following induction of caspase 9 (CASP9) with DTT or forskolin, compared with a negative control and CASP9 under control of the CMV-IE promoter (X-axis shows hours after induction). Fig. 14B shows micrographs of the fluorescent marker Spark, which was fused to CASP9, in cells where the expression
10 construct did not contain the regulatable intron (left hand panels), and in cells where the expression construct contained the regulatable intron (right hand panels).

Fig. 15 shows a graph of HEK293 cell viability following induction of caspase 9 with DTT compared with a negative control (HEK293 cells plus DTT) and CASP9 under control of the CMV-IE promoter.

15 Fig. 16A-D shows the predicted secondary structures formed when the introns as used in:

- A. SEAP (CAG/CAGACGGGCAACTTTACACGACGCTG/CAG)
- B. EGFP (CAG/CTGGAGCACTCAGACTACGTGCACCTCTG/CTG)
- C. CASP9 (CAG/CAGACGGGCAACTTTACACGACGCTG/CTG)
- D. Luciferase (CCG/CAGACGGGCAACTTTACACGACGCTG/CAG)

20 Fig. 17 shows a graph of expression of luciferase after induction with DDT (X-axis units are hours after induction) when a UPR-responsive cis-regulatory element containing 5x ATF6 binding sites combined with two different minimal promoters. Controls were the CMV-IE promoter and the CMV-minimal promoter without any additional regulatory sequences.

25 **Specific Description of Embodiments of the Invention**

Example 1: Use of the XBP1 intron as regulation mechanism for protein expression.

A 26bp non-conventional intron (regulatable) is present in the mRNA of XBP1 which, prior to
30 splicing, codes for a nonsense protein. Once endoplasmic reticulum (ER) stress has been sensed via the IRE1 protein, the non-conventional intron is removed via unique splicing sites and the mRNA subsequently re-ligated to form an mRNA transcript that can be translated to make the XBP1 protein. ER stress can be induced in many ways, e.g. via chemical addition or via expression of heterologous proteins.

35

The inventors have recognised that this process can be adapted for regulating protein expression at the mRNA/translation level by including the regulatable intron, including

suitable splice sites in the sequence of a gene of interest. Under the control of a suitable promoter, the mRNA is made but cannot be processed to form the functional expression product of the gene of interest. However, once the UPR is induced, e.g. via administration of DTT or expression of another heterologous protein, the mRNA is processed, the regulatable intron is spliced out and the mRNA can be translated into a functional protein.

By using this mechanism, the production of proteins (e.g. toxic proteins) can be tailored to fit the manufacturing process as induction is required before the protein is made. Thus, for example, translation of the protein can be delayed until a desired stage in fermentation has been reached. This method can also be used for non-protein expression products, such as functional RNAs (e.g. siRNA or miRNAs).

Splice sites and intron

Consensus splice recognition site sequence for IRE1:
CNG/CNG

This consensus sequence is conserved across eukaryotes. The consensus sequence CNG/CNG[CG], with a preference for CNG/CNGC, is typically found in mammalian cells, though the CNG/CNG consensus sequence can be used.

WT mammalian XBP1 splice recognition site sequence:

5' site: CCG/CAGC

3' site: CUG/CAGC

WT intron sequence from mammals:

Excised intron sequence (SEQ ID NO: 23):

cagcacucagacuacgugcaccucug

WT mammal intron including excised sequence flanked by splice recognition site sequences: CCG/cagcacucagacuacgugcaccucug/CAGC (SEQ ID NO: 24)

Construction of the EGFP sequence including regulatable intron

The region chosen for insertion of the intron was due to sequence similarity to the splice recognition sites. Only a single silent mutation, CTCG to CTGG, was required at the 3' recognition site to allow for the insertion of the intron.

EGFP gene sequence with underlined region for intron insertion (SEQ ID NO: 25):

ATGGTGAGCAAGGGCGAGGAGCTGTTCAACGGGGTGGTGCCCATCCTGGTCGAGCTGGACG
GCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGG
CAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCCACCCTC
5 GTGACCACCCTGACCTACGGCGTGCAGTGTTCAGCCGCTACCCCGACCACATGAAGCAGC
ACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAA
GGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAAC
CGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGG
AGTACAACCTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAA
10 GGTGAACCTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAG/CTCGCCGACCACTA
CCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGC
ACCCAGTCCGCCCTGAGCAAAGACCCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGT
TCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAG

15 EGFP protein sequence (SEQ ID NO: 26):

MVSKGEELFT GVPILVELD GDVNGHKFSV SGEGEDATY GKLTLKFICT TGKLPVPWPT
LVTTLTYGVQ CFSRYPDHMK QHDFFKSAMP EGYVQERTIF FKDDGNYKTR AEVKFEGDTL
VNRIELKGID FKEDGNILGH KLEYNYNSHN VYIMADKQKN GIKVNFKIRH NIEDGSVQLA
DHYQQNTPIG DGPVLLPDNH YLSTQSALSK DPNEKRDHNV LLEFVTAAGI TLGMDELYK

20

The sequence used in this study is shown below. GFP is underlined and the intron and splice sites are indicated in bold. In the present example, an additional 3 bp sequence was added to 5' end of the intron to create an intron that was expected to have slightly sub-optimal splicing. The intention was to keep the background expression levels, i.e. when UPR has not yet been induced, to a minimum.

25

Suspension adapted cells are known to typically be in a slightly more stressed state than adherent cells, and thus this slight modification was expected to have benefits in avoiding background expression. The wild type intron can of course be used, and may be preferred in some cases. The intron sequence as used in the present

30

example was as follows (additional 3 bp sequence is underlined):

CAG/cugcagcacucagacuacgugcaccucug/CUGG (SEQ ID NO: 27)

EGFP sequence with regulatable XBP1 intron, CMV minimal promoter (CMV-MP) and SV40 polyA tail, the XBP1 intron sequence is in bold and the EGFP and the intron encoding sequence is underlined (this construct is named SYN-P-XBP-01) (SEQ ID NO: 28):

35

GGTACCGTCGACGATATCGGATCCAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCC
 TAGATACGCCATCCACGCTGTTTTGACCTCCATAGAAGATCGCCACCATGGTGAGCAAGGGCGAGGAG
 CTGTTACACGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGT
 GTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCA
 5 AGCTGCCCCGTGCCCTGGCCACCCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTAC
 CCCGACCACATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCAC
 CATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGG
 TGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAG
 TACAACACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAAGTT
 10 CAAGATCCGCCACAACATCGAGGACGGCAGCGTG**CAG/ctgcagcactcagactacgtgcacctctg/**
CTGGCCGACCCTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTA
 CCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTGCTGGAGT
 TCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTGAGACGATCCTTATCGGATT
 TTACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCACATCTCCCCCTGAACCTGAAACAT
 15 AAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAG
 CATCACAAATTTACAAATAAAGCATTTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCA
 ATGTATCTTATCATGTCTGCTCGAAGGCGCGCCCTGGGCCTCATGGGCCTTCCGCTCACTGCCC

It can be seen that EGFP with the regulatable XBP1 intron inserted as set out above codes
 20 for a truncated protein (SEQ ID NOs: 29, 51, 52, 53 – SEQ ID NOs: 51, 52, 53 refer to the
 fragmentary sequences falling after the initial stop codon, though these will not be
 translated):

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLLKFICTTGKLPVPWPFTLVTTLTYG
 VQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGLTLVNRIELKGIDFKEDGN
 25 ILGHKLEYNNSHNVIYIMADKQKNGIKVNFKIRHNIEDGSVQLEHSDYVHLCWPTTTSRTPPSATAPC
 CCPTTTTStopAPSPPTStopAKTPTRSAITWSCWSSStopPPPGSLSAWTSCT

DNA constructs used in this example:

- CMV-MP-GFP: CMV-MP controlling expression of EGFP.
- 30 - SYN-P-XBP-01: CMV-MP controlling expression of EGFP with XBP1 intron.

These constructs are identical except for the presence of the intron in the EGFP coding
 sequence, as set out above.

35 All of these constructs were synthesised by Geneart and were provided in plasmid pMA-RQ
 (see Fig. 10 for a plasmid map). Plasmid pMA-RQ containing these constructs was directly
 transfected into the relevant cells.

Cell lines used in this example:

- Freestyle HEK293F, Invitrogen cat no: R790-07.
- Freestyle CHO-S, Invitrogen cat no: R80007.

5

Growth of HEK293F and CHO-S cells:

- Cells were maintained according the manufacturer's instructions.
- Cells were transfected using MAX reagent, Invitrogen cat no: 16447100.

10 The standard protocol for transfection of these cells was modified for 24 well plates as follows:

- 40ml of cells were grown in a 250ml vented Erlenmeyer flask (Sigma-Aldrich CLS431144) at 37°C, 8% CO₂ with agitation at 100rpm. Cells were seeded as described in the manufacturer's instructions.
- 15 - 1 day before transfection cells were counted using a haemocytometer and split to 500,000 cells/ml.
- On the day of transfection cells are seeded to 1 million cells/ml in 500µl of appropriate medium in a 24 well plate.
- 0.625µg of DNA/well was then added to 10µl of OptiMem medium
- 20 (Thermofisher; 11058021) and incubated for 5 minutes at room temperature.
- Concurrently 0.625µl of Max reagent was made up to 10µl by addition of OptiMem and incubated for 5 minutes at room temperature.
- After this incubation both mixes were added to the same tube and incubated at room temperature for 25-30 minutes.
- 25 - The DNA/Max reagent mix (20µl/well) was then added directly to the cells and the cells incubated as described previously.
- Cells were then measured for GFP fluorescence after 24hrs.
- Intron splicing from the GFP construct was measured via addition of 2mM DTT and monitoring of GFP fluorescence after 1hr. DTT is a strong reducing agent
- 30 which induces ER stress by disrupting the formation of disulphide bonds within the ER.

Measurement of GFP fluorescence:

- Cell lysis buffer was prepared from the luciferase assay system (Promega; E1500) by
- 35 diluting 1 in 5 with sterile water.
- Cells were pelleted at 900xg for 5 minutes and re-suspended in 100µl/1x10⁶ cells of cell lysis buffer from the Luciferase kit.

- This was then incubated for 10-15 minutes at room temperature for lysis to occur.
- Cellular nuclei and debris was harvested at 900xg for 5 minutes.
- The supernatant was collected and a max of 100 µl/sample to a 96 well black plate.
- The supernatant was diluted 1 in 2 with PBS (this is to dilute out the mercaptoethanol from the lysis buffer which effects the GFP signal; GFP signal is decreased in the presence of reducing agents).
- Sample was incubated for 5 minutes at room temperature.
- GFP fluorescence was measured using a plate reader with excitation and emission set at 485nm and 520nm respectively.
- GFP fluorescence was also visualised directly under a fluorescent microscope.

Results

SYNP-XBP-01 experiments

15 *HEK293F experiments:*

- 24 well transfections were set-up as described above. Each condition was performed in duplicate/triplicate.
- 24hrs after transfection the cells were treated with either 2mM DTT or by mock treatment with an equivalent volume of water.
- 1hr after treatment the cells GFP fluorescence was measured as described above.
- The ratio of GFP fluorescence to the control plasmid was calculated before and after induction. Figure 1 shows the cumulative results of 3 independent experiments.

Fig 1 shows ratio of GFP fluorescence from SYNP-XBP-01 transfected cells before and after induction by 2mM DTT. The results clearly show that upon inducing ER-Stress GFP production is induced 4-10-fold with the resultant activity 2.5-7-fold higher than the positive control. In addition, the background is considerably lower than the control (dashed line), clearly showing that expression of EGFP is being controlled by the regulatable intron. This indicates that intron is functional when placed in a heterologous gene (i.e. a gene other than XBP1), and facilitates inducible gene expression and higher expression levels.

CHO-S experiments:

An experiment using CHO-S cells was performed in an identical manner as described for HEK293F cells. The result of the experiment can be seen in Fig. 2a and Fig. 2b.

Fig. 2 shows in panel a) the ratio of GFP fluorescence positive control before and after induction. In panel b) there is shown fluorescence microscopy of SYNP-XBP-01 with and

without 2mM DTT. These data show that in CHO-S cells there is no background at all before induction and GFP fluorescence is therefore induced by >1000-fold upon addition of 2mM DTT. Control in CHO-S cells may potentially be even tighter than that observed in HEK293F cells.

5

Example 2: Use of ATF6 response elements to enhance gene expression and to provide exquisite gene control in presence of the regulatable XBP1 intron

ATF6 is a transcription factor that is activated by ER-stress. Once activated this transcription factor binds to ERSEs or UPRs and activates gene transcription of important components of protein homeostasis (see Yoshida et al., Cell, Vol. 107, 881–891, December 28, 2001).

In this study, the inventors investigated the binding and gene expression enhancement of the ATF6 binding site, the mammalian UPR, in the presence and absence of the XBP1 intron. This UPR has a consensus sequence of TGACGTG and, in addition to ATF6, is also bound by XBP-1, therefore creating a potentially powerful feedback loop of gene expression based on ER-stress. The addition of the intron also allows the investigation of whether control at transcription and translation level confers better inducibility than each on its own.

20

For this study two constructs were prepared:

- 1) 6x ATF6 elements (having the sequence TGACGTGCT) upstream of the CMV-minimal promoter (CMV-MP) and EGFP; this construct was named SYN-ATF6-01 (see below, SEQ ID NO: 30).
- 2) 6x ATF6 elements upstream of the CMV-MP and EGFP with XBP1 intron insertion; this construct was named SYN-ATF6-02 (see below, SEQ ID NO: 31).

25

The intron was inserted into EGFP as described in Example 1.

HEK293F cells were used in this study. All growth conditions, transfections and analyses were carried out as described in Example 1.

30

Plasmids used in this study:

- pMA-RQ containing the CMV-IE-GFP construct.
- pMA-RQ containing the CMV-MP-GFP construct.
- pMA-RQ containing the SYN-ATF6-01 construct.
- pMA-RQ containing the SYN-ATF6-02 construct.

35

Sequence of SYN-P-ATF6-01, the 6x ATF6 sites are underlined (SEQ ID NO: 30):

TGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCTGGTACCGTCGACGA
 TATCGGATCCAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTAGATACGCCATCC
 5 ACGCTGTTTTGACCTCCATAGAAGATCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGT
 GGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCG
 AGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCCGTGCC
 TGGCCACCCCTCGTGACCACCTGACCTACGGCGTGAGTGCTTCAGCCGCTACCCCGACCACATGAA
 GCAGCACGACTTCTTCAAGTCCGCCATGCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGG
 10 ACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAG
 CTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACCTACAACAG
 CCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACCTCAAGATCCGCCACA
 ACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCC
 GTGCTGCTGCCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCG
 15 CGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACA
 AGTGA

Sequence of SYN-P-ATF6-02 the 6x ATF6 sites are underlined and the excised intron sequence is shown in lower case and bold (SEQ ID NO: 31):

20 TGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCTTGACGTGCTGGTACCGTCGACGATATCG
 GATCCAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTAGATACGCCATCCACGCTGTTTTG
 ACCTCCATAGAAGATCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGT
 CGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACG
 GCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCCTCGTGACCACCT
 25 GACCTACGGCGTGAGTGCTTCAGCCGCTACCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATG
 CCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTG
 AAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACAT
 CCTGGGGCACAAGCTGGAGTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGG
 CATCAAGGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGT**ctgcagcactcagactacgtgcac**
 30 **ctg**CTGGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCT
 GAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGA

Results

24 well transfections were set-up as described above. Each condition was performed in
 35 duplicate/triplicate. 24hrs after transfection the cells were treated with either 2mM DTT, to

induce the UPR, or by mock treatment with an equivalent volume of water. 1, 3, 5 and 24hrs after treatment of the cells GFP fluorescence was measured as previously described.

The ratio of GFP fluorescence to the control plasmid was calculated before and after induction. Fig. 3 shows the cumulative results of 3 independent experiments.

Fig. 3 shows Expression of GFP from both plasmids after induction by 2mM DTT. Figures are ratio to CMV-IE. This demonstrates that the 6x ATF6 elements are highly inducible by ER-stress, and can enhance gene expression to 3 times the level of CMV-IE. Furthermore, addition of the XBP1 intron to EGFP increases the expression to 4 times the level of CMV-IE. When the raw background data was analysed, it was revealed that SYN-ATF6-01 had background activity equivalent to CMV-MP and that addition of the intron to EGFP decreases the background level of GFP fluorescence to practically zero (Figs 4 and 5). This was confirmed via fluorescence microscopy (Fig. 6). This indicates that a dual-control approach to protein expression provided herein, i.e. transcriptional through the regulatable promoter and translational through the regulatable intron, can provide tight regulated control and high expression levels.

Example 3: Inducible expression of EGFP from SYN-ATF6-01 and -02 during wtAAV production

The purpose of this example was to determine whether the ER-stress response could be activated by production of heterologous proteins, as an alternative (or addition) to chemical ER-stress, as used previously. An experiment was performed whereby the GFP expression from SYN-ATF6-01 and SYN-ATF6-02 was measured during AAV production.

HEK293F cells were used in this study. All growth conditions, transfections and analyses were carried out as described in Example 1.

In this experiment transfections were carried out in 24-well plates as previously described. The ATF6 plasmids (SYN-ATF6-01 and SYN-ATF6-02 constructs in the pMA-RQ vector) were transfected into HEK293F cells (using MAX reagent, as above) and after 24hrs GFP was measured. After GFP measurement the same cells were transfected with the plasmids for wtAAV production pGRG25AAV2 (provided by Adrien Savvy) and pHelper plasmid from Takara/Clontech in a 1:1 ratio. EGFP fluorescence was then measured after 1, 3, 5 and 24hrs (Fig. 7 and 8 – not all data is shown). Both plasmids are required for wtAAV

production, with the pGRG25AAV2 providing the wtAAV2 genome and the pHelper plasmid providing the E2, VA and E4 helper functions required for virus replication.

The wtAAV2 virus genome is well-known in the art (see Srivastava et al. "Nucleotide sequence and organization of the adeno-associated virus 2 genome", J Virol. 1983 Feb; 45(2): 555–564), and expression systems for AAV are also well-known in the art. In the present case the wtAAV2 virus genome was inserted into plasmid pGRG25, which is described in McKenzie and Craig, "*Fast, easy and efficient: site-specific insertion of transgenes into Enterobacterial chromosomes using Tn7 without need for selection of the insertion event*"; BMC Microbiology 2006, 6:39. The pHelper plasmid is available from Takara/Clontech (catalogue # 6234 of the AAVpro system). Suitable AAV2 expression systems for inducing AAV expression are widely commercially available, for example from Takara (Clontech) as the "AAVpro Helper Free System (AAV2)" – see http://www.clontech.com/US/Products/Viral_Transduction/AAV_Vector_Systems/Helper_Free_Expression_System.

The plasmid pAAV-CMV-ZsGreen (catalogue # 6231 from Takara/Clontech AAV vector systems) was used as control to confirm AAV expression was achieved in the cells. pAAV-CMV-ZsGreen and pHelper plasmid, in a 1:1 ratio, were transfected into a separate population of HeK293F cells. This confirmed successful AAV expression. Zs green is a variant of eGFP and as such GFP measurement was carried out on these cells as previously described.

These data show that both plasmids can be induced to the level of CMV-IE by production of wtAAV. Furthermore, it can be observed that ATF6-01 has similar background level to previous experiments, whereas the background level of ATF6-02 is zero, and only the synthesis of wtAAV can induce the production of EGFP. This supports our initial findings that both ER-stress elements are required for complete control of expression.

Example 4: Induction of UPR with Various Inducers

Experiments were performed to assess the ability of several candidate agents to induce the UPR. The ability of these candidate inducers to induce UPR was assessed using the techniques essentially as described above. The additional candidates tested were 0.5mM palmitic acid, 1uM MF-43 (2-methyl-5-(6-(4-(2-(trifluoromethyl)phenoxy)piperidin-1-yl)pyridazin-3-yl)-1,3,4-thiadiazole), which is an inhibitor of stearyl-CoA desaturase, and a combination of both of these (for the same concentrations of each was used).

In these experiments transfections were done in 24-well plates as previously described. The SYN-P-ATF6-02 plasmid (i.e. SYN-P-ATF6-02 in pMA-RQ vector) was transfected into HEK293F cells and after 24hrs GFP was measured. Induction of the UPR was then stimulated using the aforementioned inducers and samples were taken and measured at 0, 1, 3, 5 and 24hrs after induction. GFP expression was measured as previously described. Results are the average of 3 individual experiments and the error bars represent the standard deviation.

Experimental procedure in brief:

- 40ml of cells were grown in a 250ml vented Erlenmeyer flask (Sigma-Aldrich CLS431144) at 37°C, 8% CO₂ with agitation at 100rpm. Cells were seeded as described in the manufacturer's instructions.
 - 1 day before transfection cells were counted using a haemocytometer and split to 500,000 cells/ml.
 - On the day of transfection cells are seeded to 1 million cells/ml in 500µl of appropriate medium in a 24 well plate.
 - 0.625µg of DNA/well was then added to 10µl of OptiMem medium (Thermofisher; 11058021) and incubated for 5 minutes at room temperature.
 - Concurrently 0.625µl of Max reagent was made up to 10µl by addition of OptiMem and incubated for 5 minutes at room temperature.
 - After this incubation both mixes were added to the same tube and incubated at room temperature for 25-30 minutes.
 - The DNA/Max reagent mix (20µl/well) was then added directly to the cells and the cells incubated as described previously.
 - Cells were then measured for GFP fluorescence after 24hrs.
 - Intron splicing from the GFP construct was measured after addition of either 2mM DTT, 0.5mM palmitate, 1µM MF-43, a combination of 0.5mM palmitate and 1µM MF-43, and AAV synthesis by monitoring of GFP fluorescence after 1, 3, 5 and 24hrs.
- AAV production on induction of UPR:** The ATF6 plasmid (i.e. SYN-P-ATF6-02 in pMA-RQ vector) was transfected into HEK293 cells as previously described. 24hrs later GFP was measured. After GFP measurement the same cells were transfected with the plasmids for wtAAV production pGRG25AAV2 and pHelper plasmid in a 1:1 ratio into the cells, as described above. EGFP fluorescence was then measured after 1, 3, 5 and 24hrs.

Measurement of GFP fluorescence

- Cell lysis buffer was prepared from the luciferase assay system (Promega; E1500) by diluting 1 in 5 with sterile water.
- Cells were pelleted at 900xg for 5 minutes and re-suspended in 100µl/1x10⁶ cells of cell lysis buffer from the Luciferase kit.
- 5 - This was then incubated for 10-15 minutes at room temperature for lysis to occur.
- Cellular nuclei and debris was harvested at 900xg for 5 minutes.
- The supernatant was collected and a max of 100 µl/sample to a 96well black plate.
- The supernatant was diluted 1 in 2 with PBS (this is to dilute out the mercaptoethanol from the lysis buffer which effects the GFP signal; GFP signal is decreased in the
- 10 presence of reducing agents).
- Sample was incubated for 5 minutes at room temperature.
- GFP fluorescence was measured using a plate reader with excitation and emission set at 485nm and 520nm respectively.
- GFP fluorescence was also visualised directly under a fluorescent microscope.

15

The results of these experiments are shown in Figure 9, from which it can be seen that all of the agents successfully induced UPR, as indicated by the expression of functional EGFP after addition of the agent. There was a range of the strengths of induction from each of the agents. DTT is a very potent inducer of the UPR. Both palmitate and MF-43, which affect

20 lipid balance in the cells, effectively induced the UPR individually, and when combined the level of induction was more potent. WtAAV production was an effective inducer, but the effect was less strong than the other inducers.

An interesting and potentially useful observation from this experiment is that different

25 inducers can be used to induce the UPR at different levels, thus allowing for control of the level of splicing and this expression of the functional expression product, in this case EGFP. Alternatively, or additionally, different dosage levels of the various agents can be used to adjust the level of expression of the functional expression product.

30 The procedure set out in this example can be used to assess the ability of any agent to induce the UPR.

Example 5 - Use of a regulatable intron for regulation of expression of Secreted Alkaline Phosphatase (SEAP)

35

Secreted alkaline phosphatase is a standard protein used in the bioprocessing industry as a marker. It is an ideal marker for secreted proteins as the protein passes through all of the

protein quality control (transcription, translation, post-translational modification and then secretion) checkpoints of cell.

Construction of the SEAP sequence including the regulatable intron.

- 5 Construct SEAP-ATF6-001 was synthesised at GeneART by chemical synthesis

The UPR-inducible cis-regulatory element (enhancer region) used was 6x ATF6 (TGACGTGCT), each ATF6 being spaced by 20bp, coupled to the CMV-MP. This is a modification of the tandem repeat 6x ATF6 promoter used above, and its sequence
10 underlined in the sequence below.

Intron was engineered between 2 CAG codons at position 1314 of the SEAP coding sequence. The DNA coding for the excised region of the inserted intron was as follows:
CAGACGGGCAACTTTACACGACGCTG (SEQ ID NO: 32)

- 15 Including the splice sites this gives rise to the following sequence:
CAG/CAGACGGGCAACTTTACACGACGCTG/CAG. (SEQ ID NO: 33)

This sequence does not lead to the intron secondary structure described for XBP1 wild type intron in the literature (see below in Example 6 for discussion of this point).

20

The SEAP expression construct sequence, including the promoter and regulatable intron, is as follows (the excised intron sequence is shown in bold, and the 6x ATF6 enhancer region is underlined):

- 25 TGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTG
ATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTA
GCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTGCAGTTAGCGTAGCTGAGGTACCGTCGACG
ATATCGGATCCAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTAGATACGCCATC
CACGCTGTTTTGACCTCCATAGAAGATCGCCACCATGCTGCTGCTGCTGCTGCTGGGCCTGAGGC
TACAGCTCTCCCTGGGCATCATCCCAGTTGAGGAGGAGAACCCGGACTTCTGGAACCGCGAGGCAGCC
30 GAGGCCCTGGGTGCCGCCAAGAAGCTGCAGCCTGCACAGACAGCCGCCAAGAACCTCATCATCTTCCT
GGGCGATGGGATGGGGGTGTCTACGGTGACAGCTGCCAGGATCCTAAAAGGGCAGAAGAAGGACAAAC
TGGGGCCTGAGATACCCCTGGCTATGGACCGCTTCCCATATGTGGCTCTGTCCAAGACATACAATGTA
GACAAACATGTGCCAGACAGTGGAGCCACAGCCACGGCCTACCTGTGCGGGGTCAAGGGCAACTTCCA
GACCATTGGCTTGAGTGCAGCCGCCCGCTTTAACCAGTGCAACACGACACGCGGCAACGAGGTTCATCT
35 CCGTGATGAATCGGGCCAAGAAAGCAGGGAAGTCAGTGGGAGTGGTAACCACCACACGAGTGCAGCAC
GCCTCGCCAGCCGGCACCTACGCCCACACGGTGAACCGCAACTGGTACTCGGACGCCGACGTGCCTGCG
CTCGGCCCGCCAGGAGGGGTGCCAGGACATCGCTACGCAGCTCATCTCCAACATGGACATTGACGTGA

TCCTAGGTGGAGGCCGAAAGTACATGTTTCGCATGGGAACCCAGACCCTGAGTACCCAGATGACTAC
 AGCCAAGGTGGGACCAGGCTGGACGGGAAGAATCTGGTGCAGGAATGGCTGGCGAAGCGCCAGGGTGC
 CCGGTATGTGTGGAACCGCACTGAGCTCATGCAGGCTTCCCTGGACCCGTCTGTGACCCATCTCATGG
 GCCTCTTTGAGCCTGGAGACATGAAATACGAGATCCACCGAGACTCCACACTGGACCCCTCCCTGATG
 5 GAGATGACAGAGGCTGCCCTGCGCCTGCTGAGCAGGAACCCCCGCGGCTTCTTCTCTTCGTGGAGGG
 TGGTTCGCATCGACCACGGTCATCATGAAAGCAGGGCTTACCGGGCACTGACTGAAACGATCATGTTTCG
 ACGACGCCATTGAGAGGGCGGGCCAGCTCACCAGCGAGGAGGACACGCTGAGCCTCGTCACTGCCGAC
 CACTCCACGTCTTCTCCTTCGGAGGCTACCCCTGCGAGGGAGCTCCATCTTCGGGCTGGCCCTGG
 CAAGGCCCCGGGACAGGAAGGCCTACACGGTCCTCCTATACGGAAACGGTCCAGGCTATGTGCTCAAGG
 10 ACGGCGCCCCGGCCGGATGTTACCGAGAGCGAGAGCGGGAGCCCCGAGTATCGGCAG**CAGACGGGCAAC**
TTTACACGACGCTGCAGTCAGCAGTCCCCCTGGACGAAGAGACACACGCAGGCGAGGACGTGGCGGTG
 TTCGCGCGCGGGCCCGCAGGCGCACCTGGTTCACGGCGTGCAGGAGCAGACCTTCATAGCGCACGTCAT
 GGCTTCGCCGCTGCCTGGAGCCCTACACCGCTGCGACCTGGCGCCCCCGCCGGCACCCACCGACG
 CCGCGCACCCGGGTTACTCTAGAGTCGGGGCGGCCGCGCTTCGAGCAGACATGA SEQ ID NO:
 15 34)

Translation of the coding sequence containing the intron leads to the following truncated protein sequence:

MLLLLLLLGLRLQLSLGIIPVEEENPDFWNREAAEALGAACKLQPAQTAAKNLIIIFLDGDMGVSTVTA
 20 ARILKGQKKDKLGPEIPLAMDRFPYVALSKTYNVDKHVPDSGATATAYLCGVKGNFQTIGLSAAARFN
 QCNTTRGNEVISVMNRACKAGKSVGVTTRVQHASPAGTYAHTVNRNWYSDADVPASARQEGCQDIA
 TQLISNMDIDVILGGGRKYMFRMGTPDPEYPDDYSQGGTRLDGKNLVQEWLAKRQGARYVWNRTELMQ
 ASLDPSVTHLMGLFEPGDMKYEIHRDSTLDPSLMEMTEAALRLLSRNPRGFFLFVEGGRIDHGHESR
 AYRALTETIMFDDAIERAGQLTSEEDTLLSLVTADHSHVFSFGGYPLRGSSIIFGLAPGKARDRKAYTVL
 25 LYGNPGYVLKDGARPDVTESESGSPEYRQQTGNFTRRCSQQCPWTKRHTQARTWRCRAARRRTWFT
 ACRSRPS (SEQ ID NO: 35)

Removal of the intron allows full translation of the SEAP protein:

MLLLLLLLGLRLQLSLGIIPVEEENPDFWNREAAEALGAACKLQPAQTAAKNLIIIFLDGDMGVSTVTA
 30 ARILKGQKKDKLGPEIPLAMDRFPYVALSKTYNVDKHVPDSGATATAYLCGVKGNFQTIGLSAAARFN
 QCNTTRGNEVISVMNRACKAGKSVGVTTRVQHASPAGTYAHTVNRNWYSDADVPASARQEGCQDIA
 TQLISNMDIDVILGGGRKYMFRMGTPDPEYPDDYSQGGTRLDGKNLVQEWLAKRQGARYVWNRTELMQ
 ASLDPSVTHLMGLFEPGDMKYEIHRDSTLDPSLMEMTEAALRLLSRNPRGFFLFVEGGRIDHGHESR
 AYRALTETIMFDDAIERAGQLTSEEDTLLSLVTADHSHVFSFGGYPLRGSSIIFGLAPGKARDRKAYTVL
 35 LYGNPGYVLKDGARPDVTESESGSPEYRQQSAVPLDEETHAGEDVAVFARGPQAHLVHGVQEQT FIA
 HVMAFAACLEPYTACDLAPPAGTTDAAHPGYSRVGAAGRFEQT (SEQ ID NO: 36)

Experiments were performed essentially as described in the above methods, namely, HEK or CHO-s cells were transfected with the aforementioned construct (pMA-RQ) and incubated

for 24hrs. After this time the activating compound DTT was added at 2mM final concentration, SEAP activity was then measured at 3, 5 and 24hrs post-induction.

In more detail:

- 5 The protocol for transfection for 24 well plates was as described in Example 1. Thereafter:
- Supernatant was measured for SEAP activity after 24hrs.
 - Intron splicing from the SEAP construct was measured after addition of either 2mM DTT by monitoring of SEAP activity after 3, 5 and 24hrs.
 - SEAP activity was measured as per manufacturer's instructions, Roche, SEAP reporter
- 10 gene assay kit, obtained from Sigma-Aldrich, Cat. No. 11 779 842 001.

As can be seen in Fig. 11, before induction there was no SEAP expression. However, upon addition of DTT there was a rapid increase in SEAP activity, indeed after 3h the level of SEAP activity was similar to CMV-IE (a strong constitutive promoter). This speed of

15 expression is too quick to be accounted for by transcription alone and suggests that mRNA is being made from the leaky promoter and being translated almost immediately by removal of the intron. Indeed, after 24h the level of expression in both cell types is ~2-fold higher than CMV-IE.

20 These results indicate that the intron is suitable for use with secreted proteins. Furthermore, it suggests that there is considerable flexibility regarding secondary structures formed by the intron and regarding the central intron sequence that is flanked by the splice site target sequences (i.e. sequence referred to as X_n, as set out above), with the essential factors for successful splicing of the intron from mRNA being the splice site target sequences. The

25 regulatable intron sequence used in this experiment (CAGACGGGCAACUUUACACGACGCUG (SEQ ID NO: 37)) is quite different from the wild type XBP1 intron sequence (CAGCACUCAGACUACGUGCACCUCUG (SEQ ID NO: 23)), yet effective splicing was achieved. Furthermore, as discussed in more detail below, the intron used in this example is not expected to form a secondary structure resembling that

30 formed by the wild type XBP1 intron.

Example 6 - Use of a regulatable intron for regulation of protein expression using the firefly luciferase gene.

35 It has been asserted in the prior art describes that the secondary structure formed by the XBP1 intron is essential for splicing. However, all the experiments described above have used an intron structure that would not be expected to form such secondary structures, or

structures that resemble the wild type XBP1 intron, and yet the performance of the splicing system has been very robust.

Therefore, one aspect of this experiment was to determine the effect of an intron specifically designed to have the secondary structure described in the prior art. Luciferase, an intracellular protein, was selected as the reporter protein, which is also advantageous due to ease of assaying for expression levels.

Construction of the luciferase sequence including the regulatable intron.

The expression construct was synthesised at GeneART by chemical synthesis. The enhancer region is 6x ATF6 (TGACGTGCT) spaced by 20bp coupled to CMV-MP, as described above.

The intron was inserted between a CCG and a CAG codon at position 1447 of the luciferase coding sequence.

The DNA coding for the excised region of the inserted intron sequence was as follows:

CAGACGGGCAACTTTACACGACGCTG (SEQ ID NO: 32)

including splice sites this gives rise to the following sequence:

CCG/CAGACGGGCAACTTTACACGACGCTG/CAG (SEQ ID NO: 38)

This regulatable intron sequence is predicted to provide the same secondary structure, when inserted into the luciferase gene as described below, as has been described for the XBP1 wild type intron in its native location in the XBP1 gene (calculated using RNA fold web server at <http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>). The predicted secondary structure formed by the intron in luciferase is shown in Fig. 16D. For comparison the predicted secondary structures formed by the introns when inserted into the relevant genes as used in other experiments are also shown in Fig. 16:

- Fig 16B – Intron as used in EGFP (i.e. as used in Examples 1-4, encoded by CAG/CTGCAGCACTCAGACTACGTGCACCTCTG/CTG, SEQ ID NO: 48);
- Fig 16A – intron as used in SEAP (i.e. as used in Example 5, encoded by CAG/CAGACGGGCAACTTTACACGACGCTG/CAG, SEQ ID NO: 33);
- Fig. 16C – intron as used in CASP9 (i.e. as used in Example 7 encoded by CAG/CAGACGGGCAACTTTACACGACGCTG/CTG, SEQ ID NO: 43), see Fig. 16C

The luciferase expression construct sequence, including the promoter and regulatable intron, is as follows (the intron sequence is shown in bold, and the 6x ATF6 enhancer region is underlined):

Translation of the coding sequence containing the intron leads to the following truncated protein:

5 MEDAKNIKKGPAPFYPLEDGTAGEQLHKAMKRYALVPGTIAFTDAHIEVDITYAEYFEMSVR
 LAEAMKRYGLNTNHRIVVCSNSLQFFMPVLGALFIGVAVAPANDIYNERELLNSMGISQPT
 VVFSKKGGLQKILNVQKKLPPIQKIIIMDSKTDYQGFQSMYTFVTSHLPPGFNEYDFVPESFD
 RDKTIALIMNSSGSTGLPKGVALPHRTACVRFSHARDPIFGNQIIPDTAILSVVPFHHGFGMF
 TTLGYLICGFRVVLMYRFEEELFLRSLQDYKIQSALLVPTLFSFFAKSTLIDKYDLSNLHEIAS
 GGAPLSKEVGEAVAKRFHLPGRQGYGLTETTSAILITPEGDDKPGAVGKVPFFFEAKVVDL
 10 DTGKTLGVNQRGELCVRGPMIMSGYVNNPEATNALIDKDGWLHSGDIAYWDEDEHFFIVD
 RLKSLIKYKGYQVAPAELESILLQHPNIFDAGVAGLPDDDAGELPAADGQLYTTLQTSCWNT
 VKP (SEQ ID NO: 39)

Removal of the intron allows translation of the full luciferase protein:

15 MEDAKNIKKGPAPFYPLEDGTAGEQLHKAMKRYALVPGTIAFTDAHIEVDITYAEYFEMSVR
 LAEAMKRYGLNTNHRIVVCSNSLQFFMPVLGALFIGVAVAPANDIYNERELLNSMGISQPT
 VVFSKKGGLQKILNVQKKLPPIQKIIIMDSKTDYQGFQSMYTFVTSHLPPGFNEYDFVPESFD
 RDKTIALIMNSSGSTGLPKGVALPHRTACVRFSHARDPIFGNQIIPDTAILSVVPFHHGFGMF
 TTLGYLICGFRVVLMYRFEEELFLRSLQDYKIQSALLVPTLFSFFAKSTLIDKYDLSNLHEIAS
 20 GGAPLSKEVGEAVAKRFHLPGRQGYGLTETTSAILITPEGDDKPGAVGKVPFFFEAKVVDL
 DTGKTLGVNQRGELCVRGPMIMSGYVNNPEATNALIDKDGWLHSGDIAYWDEDEHFFIVD
 RLKSLIKYKGYQVAPAELESILLQHPNIFDAGVAGLPDDDAGELPAADVLEHGKTMTEKEIV
 DYVASQVTTAKKLRGGVVFVDEVPKGLTGKLDARKIREILIKAKKGGKIAV (SEQ ID NO: 40)

25 Experiments were performed essentially as described in Example 5, namely, HEK or CHO-s cells were transfected with the aforementioned construct and incubated for 24hrs. After this time the activating compounds DTT or palmitate was added. Luciferase activity was then measured at 3, 5 and 24h post-induction. Luciferase was measured as follows:

- Cell lysis buffer was prepared from the luciferase assay system (Promega; E1500) by
 30 diluting 1 in 5 with sterile water.
- Cells were pelleted at 900xg for 5 minutes and re-suspended in 100µl/1x10⁶ cells of cell lysis buffer from the Luciferase kit.
- This was then incubated for 10-15 minutes at room temperature for lysis to occur.
- Cellular nuclei and debris was harvested at 900xg for 5 minutes.
- 35 - The supernatant was collected, and 10µl/sample was added to a 96 well white plate.
- Bioluminescence was measured using a plate reader using an autoinjector and 50µl of substrate was injected. Substrate was prepared as per manufacturer's instructions (Promega; E1500).

As can be seen in Fig. 12, before induction there was no luciferase expression, however upon addition of either inducer there was a rapid increase in activity, indeed after 3hrs the level of luciferase activity was 2-2.5x that of CMV-IE (a strong constitutive promoter). This speed of expression is, again, too rapid to be accounted for by transcription alone and suggests that some mRNA is being already made from the promoter and being translated almost immediately after removal of the intron following induction. Indeed, after 24hrs the level of expression in both cell types is ~3-4-fold higher than CMV-IE.

These results are somewhat suggestive that the “perfect” intron secondary structure might give better induction levels or higher splicing efficiencies in some cases. However, the structure is clearly unimportant for overall function of the system. Furthermore, SEAP and luciferase are quite different proteins and a straightforward comparison is impossible; SEAP is secreted and therefore undergoes further modifications and another bottleneck for expression.

Example 7 - Control of Caspase 9 gene expression using a regulatable intron

This experiment was performed to confirm that expression control from the regulatable intron was suitably tight to control expression of toxic expression products that are lethal to the cells expressing them. The regulatable intron was engineered into the apoptosis causing protein caspase 9. Overexpression of this protein in HEK cells causes rapid death. Therefore, any expression of this protein from the unmodified sequence leads to a significant increase in the dead cells within a culture. This allows us to determine the tightness of control afforded by the intron. Caspase 9 was also fused to GFP-Spark which would not be visible if the intron was present in the transcript.

Construction of the luciferase sequence including the regulatable intron.

For expression the intron was coupled with the CMV-IE constitutive promoter. The intron construct was cloned into the plasmid vector SYNP-CASPSp-001 using the BsaI restriction site and 2 complimentary oligos. The plasmid was digested with BsaI (NEB R0535S) and the oligos annealed by heating to 98°C for 5mins, to melt the secondary structures, and then incubating at 55°C for 20 minutes. This allowed the oligos to form double stranded DNA. This DNA was designed with overhangs at both the 5' and 3' ends that would allow ligation to the digested SYNP-CASPSp-001. The double stranded DNA was then ligated into the plasmid and subsequently transformed into one shot top ten chemically competent cells (Thermofisher, C404003). Isolated DNA was sequenced to confirm the presence of the intron. The complete plasmid was referred to as SYNP-CASP9-INT, and a plasmid map is shown in Fig 13.

The intron oligos added into the vector had the following sequences:

INTC9FP: ACGTCAGACGGGCAACTTTACACGACGCTG (SEQ ID NO: 41)

INTC9RP: ACGTCAGCGTCGTGTAAAGTTGCCCGTCTG (SEQ ID NO: 42)

5

The regulatable intron was thus engineered between an added CCAG and a CTG codon at position 1087 of the Caspase 9 coding sequence.

The DNA coding for the excised region of the inserted intron sequence was as follows:

10 CAGACGGGCAACTTTACACGACGCTG (SEQ ID NO: 32)

Including splice sites this gives rise to the following sequence:

CAG/CAGACGGGCAACTTTACACGACGCTG/CTG (SEQ ID NO: 43)

15 This sequence does not lead to the "perfect" intron secondary structure described for XBP1 wild type intron.

The SYN-P-CASP9-INT vector including the CASP9-Intron construct sequence, is as follows (the intron sequence is shown in bold):

20 GACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACCTCTCAGTACAATCTGCTCTGATGCCGCATAG
TTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGC
TACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGC
TTCGCGAGTACATTTATATTGGCTCATGTCCAATATGACCGCCATGTTGACATTGATTATTGACTAGT
TATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCGCGGTTACATAACT
TACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATG
25 TTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCC
CACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATG
GCCCCGCTGGCATTATGCCAGTACATGACCTTACGGGACTTTCTTACTTGGCAGTACATCTACGTAT
TAGTCATCGCTATTACCATGGTGATGCGGTTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTTGAC
TCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTGTTTGGCACCAAAATCAACG
30 GGAATTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAATGGGCGGTAGGCGTGTACGGTGGG
AGGTCTATATAAGCAGAGCTCGTTTTAGTGAACCGTCAGATCCTCACTCTCTTCCGCATCGCTGTCTGC
GAGGGCCAGCTGTTGGGCTCGCGGTTGAGGACAACTCTTCGCGGTCTTTCCAGTACTCTTGGATCGG
AAACCCGTCGGCCTCCGAACGGTACTCCGCCACCGAGGGACCTGAGCGAGTCCGCATCGACCGGATCG
GAAAACCTCTCGAGAAAGGCGTCTAACCAGTCACAGTCGCAAGGTAGGCTGAGCACCGTGGCGGGCGG
35 CAGCGGGTGGCGGTTCGGGTTGTTTCTGGCGGAGGTGCTGCTGATGATGTAATTAAAGTAGGCGGTCT
TGAGACGGCGGATGGTCGAGGTGAGGTGTGGGTTTAGTGAACCGTCAGATCCTCACTCTCTTCCGCAT
CGCTGTCTGCGAGGGCCAGCTGTGAGGCTTGAATCCAGCTGTTGGGGTGAATCTCCCTCTCAAAAG
CGGGCATTACTTCTGCGCTAAGATTGTCAGTTTCCAAAACGAGGAGGATTTGATATTCACCTGGCCC
GATCTGGCCATACACTTGAGTGACAATGACATCCACTTTGCCTTTCTCTCCACAGGTGTCCACTCCCA
40 GGTCCAAGTTTAAACTTTAATACGACTCACTATAGGGGCCGCCACCAAGCTTGGTACATGGACGAAGC
GGATCGGCGGCTCCTGCGGCGGTGCCGGCTGCGGCTGGTGAAGAGCTGCAGGTGGACAGCTCTGGG
ACGCCCTGCTGAGCCGCGAGCTGTTCAAGCCCCATATGATCGAGGACATCCAGCGGGCAGGCTCTGGA

TCTCGGCGGGATCAGGCCAGGCAGCTGATCATAGATCTGGAGACTCGAGGGAGTCAGGCTCTTCCTTT
GTTTCATCTCCTGCTTAGAGGACACAGGCCAGGACATGCTGGCTTCGTTTCTGCGAACTAACAGGCAAG
CAGCAAAGTTGTGCAAGCCAACCTAGAAAACCTTACCCCAGTGGTGCTCAGACCAGAGATTTCGCAAA
CCAGAGGTTCTCAGACCGGAAACACCCAGACCAGTGGACATTGGTTCTGGAGGATTTCGGTGATGTCCG
5 TGCTCTTGAGAGTTTGAGGGGAAATGCAGATTTGGCTTACATCCTGAGCATGGAGCCCTGTGGCCACT
GCCTCATTATCAACAATGTGAACCTTCTGCCGTGAGTCCGGGCTCCGCACCCGCACTGGCTCCAACATC
GACTGTGAGAAGTTGCGGCGTCGCTTCTCCTCGCTGCATTTTCATGGTGGAGGTGAAGGGCGACCTGAC
TGCCAAGAAAATGGTGCTGGCTTTGCTGGAGCTGGCGCAGCAGGACCACGGTGCTCTGGACTGCTGCG
TGGTGGTCATTCTCTCTCACGGCTGTGAGGCCAGCCACCTGCAGTTCACAGGGGCTGTCTACGGCACA
10 GATGGATGCCCTGTGTCCGTGAGAAAGATTGTGAACATCTTCAATGGGACCAGCTGCCCCAGCCTGGG
AGGGAAGCCCAAGCTCTTTTTCATCCAGGCCTGTGGTGGGGAGCAGAAAGACCATGGGTTTGAGGTGG
CCTCCACTTCCCCTGAAGACGAGTCCCCTGGCAGTAACCCGAGCCAGATGCCACCCGTTCCAGGAA
GGTTTGAGGACCTTCGACCAGCTGGACGCCATATCTAGTTTGCCACACCCAGTGACATCTTTGTGTC
CTACTCTACTTTCCAGGTTTTGTTTCTGGAGGGACCCCAAGAGTGGCTCCTGGT**CAGACGGGCAAC**
15 **TTTACACGACGCTGG**AGACCCTGGACGACATCTTTGAGCAGTGGGCTCACTCTGAAGACCTGCAGTCC
CTCCTGCTTAGGGTCGCTAATGCTGTTTCGGTGAAAGGGATTTATAAACAGATGCCTGGTTGCTTTAA
TTTCTCCGAAAAAACTTTTCTTTAAACATCAGGGGGTGGAGGCTCTGTGAGCAAGGGCGAGGAGC
TGTTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTG
TCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAA
20 GCTGCCCGTGCCCTGGCCCACCTCGTGACCACCTGACCTACGGCGTGCAAGTTCAGCCGCTACC
CCGACCACATGAAGAAGCAGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACC
ATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGT
GAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGT
ACAACATAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGCTAACTTC
25 AAGGTTCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCAT
CGGCGACGGCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACC
CCAACGAGAAGCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATG
GACGAGCTGTACAAGTAAACTCGAGTCTAGAGCGCCGCCGAATTCGGGGCCGTTTAAACCCGCTGAT
CAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCCCCGTGCCCTTCCTTGACC
30 CTGGAAGGTGCCACTCCCACTGTCTTTCTTAATAAAATGAGGAAATGCATCGCATTGTCTGAGTAG
GTGTCAATCTATTCTGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCA
GGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGG
TATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGCGGGGTGTGGTGGTTACGCGCAGCGTGACCGC
TACACTTGCCAGCGCCCTAGCGCCCCGCTCCTTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTTCGAG
35 GCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTC
GACCCCAAAAAAAGTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCG
CCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACC
CTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTCGGCCTATTGGTTAAAAAATGAG
CTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCC
40 CCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAA
GTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCC
CGCCCCTAACCTCGCCCATCCCGCCCCTAACCTCGCCCAAGTTCGCCCCATTCTCGCCCCCATGGCTGA
CTAATTTTTTTTATTTATGAGAGGCGAGGCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGG
AGGCTTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCTCGGGAGCTTGTATATCCATTTTCGGATCTG
45 ATCAGCACGTGATGAAAAAGCCTGAACTCACCGCGACGTCTGTGAGAAGTTTCTGATCGAAAAGTTC
GACAGCGTCTCCGACCTGATGCAGCTCTCGAGGGCGAAGAATCTCGTGCTTTCAGCTTCGATGTAGG
AGGGCGTGATATGTCCTGCGGGTAAATAGCTGCGCCGATGGTTCACAAAGATCGTTATGTTTATC
GGCACTTTGCATCGGCCGCGCTCCCGATTCCGGAAGTGCTTGACATTGGGGAATTTCAGCGAGAGCCTG

ACCTATTGCATCTCCCGCCGTGCACAGGGTGTACGTTGCAAGACCTGCCTGAAACCGAACTGCCCGC
TGTTCTGCAGCCGGTCGCGGAGGCCATGGATGCGATCGCTGCGGCCGATCTTAGCCAGACGAGCGGGT
TCGGCCCATTCGGACCGCAAGGAATCGGTCAATACACTACATGGCGTGATTTTCATATGCGCGATTGCT
GATCCCCATGTGTATCACTGGCAAACGTGTATGGACGACACCGTCAGTGCGTCCGTGCGCGAGGCTCT
5 CGATGAGCTGATGCTTTGGGCCGAGGACTGCCCCGAAGTCCGGCACCTCGTGACGCGGATTTTCGGCT
CCAACAATGTCTTGACGACAATGGCCGCATAACAGCGGTCAATTGACTGGAGCGAGGCGATGTTTCGGG
GATTCCCAATACGAGGTGCGCAACATCTTCTTCTGGAGGCCGTGGTTGGCTTGTATGGAGCAGCAGAC
GCGCTACTTCGAGCGGAGGCATCCGGAGCTTGCGAGGATCGCCGCGGCTCCGGGCGTATATGCTCCGCA
TTGGTCTTGACCAACTCTATCAGAGCTTGGTTGACGGCAATTTTCGATGATGCAGCTTGGGCGCAGGGT
10 CGATGCGACGCAATCGTCCGATCCGGAGCCGGGACTGTCCGGCGTACACAAATCGCCCGCAGAAAGCGC
GGCCGTCTGGACCGATGGCTGTGTAGAAGTACTCGCCGATAGTGGAACCGACGCCCCAGCACTCGTC
CGAGGGCAAAGGAATAGCACGTGCTACGAGATTTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGC
TTCGGAATCGTTTTCCGGGACGCTGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTT
CGCCCCACCCCAACTTGTTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCA
15 CAAATAAAGCATTTTTTTTCACTGCATTCTAGTTGIGGTTTGTCCAAACTCATCAATGTATCTTATCAT
GTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTTCTGTGTGAAA
TTGTTATCCGCTCACAAATCCACACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCT
AATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCCGCTTTCAGTCGGGAAACCTGTCTG
TGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGC
20 TTCTCGCTCACTGACTCGCTGCGCTCGGTCTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGG
CGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAA
AAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCA
TCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTC
CCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCTTT
25 CTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGATAGGTCGT
TCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTTCAGCCCGACCGCTGCGCCTTATCCGGTAACT
ATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATT
AGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAG
AAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTT
30 GATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGA
AAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGAGCGCGGAACCCCTATTTGTTT
ATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAATTAATTCCTTAGAAAACTCATCGAGCATC
AAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAA
TGAAGGAGAAAACCTACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGA
35 CTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAAATAAGGTTATCAAGTGAGAAATCA
CCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAAC
AGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCTGTGATTGCG
CCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGG
CGCAGGAACACTGCCAGCGCATCAACAATATTTTACCTGAATCAGGATATTCTTCTAATACCTGGAA
40 TGCTGTTTTCCCAGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGA
TGGTCGGAAGAGGCATAAATTCGGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTTGGCA
ACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGT
CGCACCTGATTGCCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAAT
TTAATCGCGGCCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGATTACTGTTTT
45 ATGTAAGCAGACAGTTTTATTGTTTCATGACCAAAATCCCTTAACGTGAGTTTTTCGTTCCACTGAGCGT
CAGACCCCGTAGAAATCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTC (SEQ ID NO: 44)

The coding sequence containing the intron leads to the following truncated protein:

MDEADRRLRLRRCLRLVEELQVDQLWDALLSRELFRPHMIEDIQRAGSGSRRDQARQLIIDLETRGSQ
 ALPLFISCLEDTGQDMLASFLRTNRQAAKLSKPTLENLTPVVLRLPEIRKPEVLRPETPRPVDIGSGGF
 GDVGALESRLGNADLAYILSMEPCGHCLIIINNPNFCRESGLRTRTGSNIDCEKLRRRFSSSLHFMVEVK
 5 GDLTAKKMLALLELAQQDHGALDCCVVVILSHGCQASHLQFPGAVYGTGDCPVSVEKIVNIFNGTSC
 PSLGGKPKLFFIQACGGEQKDHGFEVASTSPEDESPGSNPEPDATPFQEGLRTFDQLDAISSLPSPD
 IFVSYSTFPGFVSWRDPKSGSWSAGDPGRHL (SEQ ID NO: 45)

Removal of the intron allows full translation of the Caspase 9 protein:

10 MDEADRRLRLRRCLRLVEELQVDQLWDALLSRELFRPHMIEDIQRAGSGSRRDQARQLIIDLETRGSQ
 ALPLFISCLEDTGQDMLASFLRTNRQAAKLSKPTLENLTPVVLRLPEIRKPEVLRPETPRPVDIGSGGF
 GDVGALESRLGNADLAYILSMEPCGHCLIIINNPNFCRESGLRTRTGSNIDCEKLRRRFSSSLHFMVEVK
 GDLTAKKMLALLELAQQDHGALDCCVVVILSHGCQASHLQFPGAVYGTGDCPVSVEKIVNIFNGTSC
 PSLGGKPKLFFIQACGGEQKDHGFEVASTSPEDESPGSNPEPDATPFQEGLRTFDQLDAISSLPSPD
 15 IFVSYSTFPGFVSWRDPKSGSWSADGQLYTTLLETLDLIFEQWAHSEDLSLLLRVANAVSVKGIYKQ
 MPGCFNFLRKKLFFKTSGGGGS (SEQ ID NO: 46)

Transfection was performed as previously described. DTT was added at 2mM and forskolin
 was added at 10 μ M. The latter was carried out to determine whether forskolin could induce
 20 the removal of the intron. Amongst other things, forskolin has potential gene therapy
 applications. Induction of intron removal was measured as function of cell death, with higher
 cell death an indicator of removal of the intron.

Experiments were performed as described in the methods namely, HEK or CHO-s cells were
 25 transfected with the aforementioned construct and incubated for 24h. After this time the
 activating compound DTT or Forskolin was added, % cell death was then measured at 3, 5
 and 24h post-induction.

Induction of intron was measured as function of cell death, using the Countess II cell counter
 30 (Thermofisher). Cells were stained with trypan blue to ascertain cell viability. 10 μ l of cell
 suspension was added to a Haemocytometer and analysed on the countess.

As can be seen in Fig. 14A, before induction there was no cell death, however upon addition
 of either inducer there was a rapid increase in the % of dead cells, indeed after 3h with DTT
 35 the number of dead cells was close to the control construct which contained no intron.

Indeed, 5h after DTT treatment the % of dead cells is equal to the positive control whereas
 the forskolin treatment only achieved this level after 24hrs, suggesting it is a weak inducer of
 splicing. Cells with no inducer added showed little or no cell death indicating that there was
 no expression of caspase 9 until addition of the inducing agent. This shows that control of

gene expression from this is very tight (indeed apparently absolute) and at the mRNA level as the CMV-IE promoter is constitutive and therefore continually making transcripts. Fig 14B also shows that the control of gene expression from the intron is very tight. This is because the Caspase 9 is fused to a GFP variant called SPARK, therefore expression of Caspase 9 will lead to expression of the GFP variant. This figure shows that there is no GFP expression until the addition of an inducer and cell death has begun.

In a further experiment, the same construct as described above (SYNP-CASP9-INT) was used and the rate of cell death was measured. From this figure it can be seen that ~50% of cells die within 1 hr of induction, showing that the response is rapid as well as incredibly tightly controlled. The results of this experiment are shown in Fig. 15.

Example 8 – UPR-Inducible Promoter using different minimal promoters

The purpose of this experiment was to test the ATF6-containing UPR-responsive cis-regulatory element (also referred to herein as an enhancer) as used in the examples above with different minimal promoters and to further assess their inducibility and tightness of control. To this end, the ATF6-containing UPR-responsive cis-regulatory element was operably linked to the CMV-MP and MinTK (herpes thymidine kinase minimal promoter) minimal promoters

The constructs were synthesised by GeneART by chemical synthesis, as above.

As described above, the enhancer sequence contains 6 repeats of the sequence TGACGTGCT (which contains the ATF6 consensus sequence, TGACGTG) spaced by 20bp spacer sequences. The enhancer sequence was coupled to either MinTK (in construct ATF06-MP001) or CMV-MP (in construct ATF06-MP002).

Transfection of HEK293-F cells was performed as previously described. DTT was added at a concentration of 2mM. Luciferase activity was measured as previously described. The results are shown in Fig. 17 – the DTT was added at time = 0h as shown on the X-axis (units are hours).

Both constructs showed good inducibility and negligible expression prior to addition of the DTT. Slightly higher expression was observed for the promoter comprising CMV-MP compared with the promoter comprising MinTK. However, both promoter constructs showed

high levels of expression after induction of the UPR by DTT, significantly higher than provided by the constitutive CMV-IE promoter.

ATF06-MP-001 (ATF06 and MinTK promoter sequence underlined)

5 GAAGTTCCTATTCCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCTGACGTGCTGATGATGCGTA
GCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTT
GACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGA
TGATGCGTAGCTAGTAGTGAGTTAGCGTAGCTGAGGTACCGTCGACGATATCGGATCCTTCGCATAT
TAAGGTGACGCGTGTGGCCTCGAACACCGAGCGACCCTGCAGCGACCCGCTTAAATGGAAGATGCCAA
10 AAACATTAAGAAGGGCCAGCGCCATTCTACCCACTCGAAGACGGGACCGCCGGCGAGCAGCTGCACA
AAGCCATGAAGCGCTACGCCCTGGTGCCCGGCACCATCGCCTTTACCGACGCACATATCGAGGTGGAC
ATTACCTACGCCGAGTACTTCGAGATGAGCGTTTCGGCTGGCAGAAGCTATGAAGCGCTATGGGCTGAA
TACAAACCATCGGATCGTGGTGTGCAGCGAGAATAGCTTGCAGTTCTTCATGCCCGTGTGGGTGCCC
TGTTTCATCGGTGTGGCTGTGGCCCCAGCTAACGACATCTACAACGAGCGCGAGCTGCTGAACAGCATG
15 GGCATCAGCCAGCCCACCGTCGTATTTCGTGAGCAAGAAAGGGCTGCAAAAGATCCTCAACGTGCAAAA
GAAGCTACCGATCATAAAAAGATCATCATCATGGATAGCAAGACCGACTACCAGGGCTTCCAAAGCA
TGTACACCTTCGTGACTTCCCATTTGCCACCCGGCTTCAACGAGTACGACTTCGTGCCCGAGAGCTTC
GACCGGGACAAAACCATCGCCCTGATCATGAACAGTAGTGGCAGTACCGGATTGCCCAAGGGCGTAGC
CCTACCGCACCGCACCGCTTGTGTCCGATTTCAGTCATGCCCGCGACCCCATCTTCGGCAACCAGATCA
20 TCCCCGACACCGCTATCCTCAGCGTGGTGCCATTTACACACGGCTTCGGCATGTTACACACGCTGGGC
TACTTGATCTGCGGCTTTCGGGTCTGCTCATGTACCGCTTCGAGGAGGAGCTATTCTTGCGCAGCTT
GCAAGACTATAAGATTCAATCTGCCCTGCTGGTGCCACACTATTTAGCTTCTTCGCTAAGAGCACTC
TCATCGACAAGTACGACCTAAGCAACTTGCACGAGATCGCCAGCGGGCGGGCGCGCTCAGCAAGGAG
GTAGGTGAGGCCGTGGCCAAACGCTTCCACCTACCAGGCATCCGCCAGGGCTACGGCCTGACAGAAAC
25 AACCAGCGCCATTCTGATCACCCCCGAAGGGGACGACAAGCCTGGCGCAGTAGGCAAGGTGGTGCCCT
TCTTCGAGGCTAAGGTGGTGGACTTGGACACCGGTAAGACACTGGGTGTGAACCAGCGCGGCGAGCTG
TGCGTCCGTGGCCCCATGATCATGAGCGGCTACGTTAACAACCCGAGGCTACAAACGCTCTCATCGA
CAAGGACGGCTGGCTGCACAGCGGCGACATCGCCTACTGGGACGAGGACGAGCACTTCTTCATCGTGG
ACCGGCTGAAGAGCCTGATCAAATACAAGGGCTACCAGGTAGCCCCAGCCGAAGTGGAGAGCATCCTG
30 CTGCAACACCCCAACATCTTCGACGCCGGGGTCGCCGGCCTGCCCGACGACGATGCCGGCGAGCTGCC
CGCCGCAGTCGTCTGTGGAACACGCTAAAACCATGACCGAGAAGGAGATCGTGGACTATGTGGCCA
GCCAGGTACAAACCGCCAAGAAGCTGCGCGGTGGTGTGTGTTTCGTGGACGAGGTGCCTAAAGGACTG
ACCGGCAAGTTGGACGCCCGCAAGATCCGCGAGATTCTCATTAAGGCCAAGAAGGGCGGCAAGATCGC
CGTGTAATGAAAGCTTGGTCTCTACGAGTAATAGACGCCAGTTGAATTCCTTCGAGCAGACATGATA
35 AGATACATTGATGAGTTTGGACAAACCACAACCTAGAATGCAGTGAAAAAATGCTTTATTTGTGAAAT
TTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCA

TTCATTTTATGTTTCAGGTTTCAGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAA
TGTGGTAAAATCGATAAGGATCCGT (SEQ ID NO: 59)

ATF06-MP-002 (ATF06 and CMV-MP promoter sequence underlined)

5 GCGACGTAATACGACTCACTATAGGGCGAATTGGCGGAAGGCCGTCAAGGCCGCATGCATAATAAAAT
ATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGTGACGTGCTGATGATGCGTAGCTAGTA
GTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGC
TGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCGTAGCTAGTAGTTGACGTGCTGATGATGCG
TAGCTAGTAGTGACGTTAGCGTAGCTGAGGTACCGTCGACGATATCGGATCCAGGTCTATATAAGCAG
10 AGCTCGTTTAGTGAACCGTCAGATCGCCTAGATACGCCATCCACGCTGTTTTGACCTCCATAGAAGAT
CGCCACCATGGAAGATGCCAAAAACATTAAAGAGGGCCAGCGCCATTCTACCCACTCGAAGACGGGA
CCGCCGGCGAGCAGCTGCACAAAGCCATGAAGCGCTACGCCCTGGTGCCCGGCACCATCGCCTTTACC
GACGCACATATCGAGGTGGACATTACCTACGCCGAGTACTTCGAGATGAGCGTTTCGGCTGGCAGAAGC
TATGAAGCGCTATGGGCTGAATACAAACCATCGGATCGTGGTGTGCAGCGAGAATAGCTTGCAGTTCT
15 TCATGCCCCTGTTGGGTGCCCTGTTTCATCGGTGTGGCTGTGGCCCCAGCTAACGACATCTACAACGAG
CGCGAGCTGCTGAACAGCATGGGCATCAGCCAGCCCACCGTCGTATTCGTGAGCAAGAAAGGGCTGCA
AAAGATCCTCAACGTGCAAAAGAAGCTACCGATCATACAAAGATCATCATCATGGATAGCAAGACCG
ACTACCAGGGCTTCCAAAGCATGTACACCTTCGTGACTTCCCATTGTCACCCGGCTTCAACGAGTAC
GACTTCGTGCCCAGAGCTTCGACCCGGGACAAAACCATCGCCCTGATCATGAACAGTAGTGGCAGTAC
20 CGGATTGCCCAAGGGCGTAGCCCTACCGCACCGCACCGCTTGTGTCCGATTCAGTCATGCCCAGCACC
CCATCTTCGGCAACCAGATCATCCCCGACACCGCTATCCTCAGCGTGGTGCCATTTACACACGGCTTC
GGCATGTTACACACGCTGGGCTACTTGATCTGCGGCTTTTCGGGTCGTGCTCATGTACCGCTTCGAGGA
GGAGCTATTCTTGCGCAGCTTGCAAGACTATAAGATTCAATCTGCCCTGCTGGTGCCACACTATTTA
GCTTCTTCGCTAAGAGCACTCTCATCGACAAGTACGACCTAAGCAACTTGCACGAGATCGCCAGCGGC
25 GGGGCGCCGCTCAGCAAGGAGGTAGGTGAGGCCGTGGCCAAACGCTTCCACCTACCAGGCATCCGCCA
GGGCTACGGCCTGACAGAAACAACCAGCGCCATTCTGATCACCCCCGAAGGGGACGACAAGCCTGGCG
CAGTAGGCAAGGTGGTGCCCTTCTTCGAGGCTAAGGTGGTGGACTTGACACCGGTAAGACACTGGGT
GTGAACCAGCGCGGCGAGCTGTGCGTCCGTGGCCCCATGATCATGAGCGGCTACGTTAACAACCCCGA
GGCTACAAACGCTCTCATCGACAAGGACGGCTGGCTGCACAGCGGCGACATCGCCTACTGGGACGAGG
30 ACGAGCACTTCTTCATCGTGGACCGGCTGAAGAGCCTGATCAAATACAAGGGCTACCAGGTAGCCCCA
GCCGAACCTGGAGAGCATCCTGCTGCAACACCCCCAACATCTTCGACGCCGGGGTCGCCGGCCTGCCGA
CGACGATGCCGCGGAGCTGCCCCGCCGAGTCGTCTGCTGGAACACGGTAAAACCATGACCGAGAAGG
AGATCGTGGACTATGTGGCCAGCCAGGTTACAACCGCCAAGAAGCTGCGCGGTGGTGTGTGTTGTTGCTG
GACGAGGTGCCTAAAGGACTGACCGGCAAGTTGGACGCCCGCAAGATCCGCGAGATTCTCATTAAGGC
35 CAAGAAGGGCGGCAAGATCGCCGTGTAATGAAAGCTTGGTCTCTACGAGTAATAGACGCCACAGTTGAA
TTCCTTCGAGCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAAGTAGAATGCAGTGAAA
AAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAAC

AAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTTCAGGGGGAGGTGTGGGAGGTTTTTTAA
AGCAAGTAAACCTCTACAAATGTGGTAAATCGATAAGGATCCGTCTGGGCCTCATGGGCCTTCCGC
TCACTGCCCCGCTTTCAGTCGGGAAACCTGTCGTCTGGAGCCAGCTGCAT (SEQ ID NO: 60)

While the making and using of various embodiments of the present invention are discussed in detail above, it should be appreciated that the present invention provides many applicable inventive concepts that can be embodied in a wide variety of specific contexts. The specific embodiments discussed herein are merely illustrative of specific ways to make and use the invention and do not delimit the scope of the invention.

The reference to any prior art in this specification is not, and should not be taken as, an acknowledgement or any form of suggestion that the prior art forms part of the common general knowledge in Australia.

Claims

1. A synthetic nucleic acid expression construct for producing an expression product in a cell, the nucleic acid expression construct comprising a promoter sequence operably linked to a nucleic acid sequence encoding the expression product, the nucleic acid sequence encoding the expression product comprising a sequence which encodes a regulatable intron, said regulatable intron being an intron which comprises an excisable sequence which is capable of being spliced out of a transcript produced from the synthetic expression construct via the unfolded protein response (UPR) system in the cell, thereby resulting in a transcript encoding a functional expression product.
2. The synthetic nucleic acid expression construct of claim 1, wherein the regulatable intron is capable of being spliced out by the IRE1 protein or a homologue or orthologue thereof; and/or wherein the splicing out of the excisable sequence of the regulatable intron permits correct translation of the transcript from the nucleic acid sequence encoding the expression product, thereby allowing the desired expression product to be produced.
3. The synthetic nucleic acid expression construct of claim 1 or claim 2, wherein the expression product is a protein, and/or wherein the presence of the intron in the transcript from the nucleic acid sequence encoding the expression product results in a protein being translated from the transcript which is non-functional.
4. The synthetic nucleic acid expression construct of any one of the preceding claims, wherein splicing out of the excisable sequence of the regulatable intron (a) eliminates a premature stop codon in the transcript; and/or (b) results in a shift of reading frame for sequences in the transcript located downstream of the regulatable intron; and/or (c) wherein the length in nucleotides of the excisable sequence is not divisible by 3.
5. The synthetic nucleic acid expression construct of any one of the preceding claims, wherein the regulatable intron comprises the sequence CNG/CNG-Xn-CNG/CNG, wherein Xn represents a sequence of length n bases, wherein / represents a cleavage site and wherein the sequence CNG-Xn-CNG is excised from the transcript; and preferably, wherein Xn has a length of from 10 to 500 nucleotides, more preferably 15 to 350 nucleotides, yet more preferably 15 to 100 nucleotides, yet more preferably 15 to 35 nucleotides, and yet more preferably 20 to 25 nucleotides, wherein Xn comprises the sequence CACUCAGACUACGUGCACCU (SEQ ID NO: 1) or a sequence which is at least 60% identical thereto.

6. The synthetic nucleic acid expression construct of any one of the preceding claims, wherein:
- (a) the regulatable intron comprises the sequence CNG/CNG-Xn-CNG/CNG[CG], wherein Xn represents a sequence of length n nucleotides, wherein / represents the cleavage site such that the excisable sequence CNG-Xn-CNG is excised from the transcript upon splicing, and wherein the nucleotide at the 5' end of the sequence Xn is a C or G; optionally wherein
- (b) Xn comprises the sequence CACUCAGACUACGUGCACCU (SEQ ID NO: 1) or a sequence which is at least 60% identical thereto, and further optionally wherein Xn comprises or consists of one of the following sequences:
- CACUCAGACUACGUGCACCU (SEQ ID NO: 1);
 - CACUCAGACUACGUGCUCU (SEQ ID NO: 2);
 - CACUCAGACUACGUGCCCCU (SEQ ID NO: 3);
 - CACUCAGACUACGUGCGCCU (SEQ ID NO: 4); and
 - CACUCAGACUAUGUGCACCU (SEQ ID NO: 5); and/or
- (c) the regulatable intron comprises the sequence CNG/CNGCACUCAGACUACGUGCACCUCNG/CNGC (SEQ ID NO: 6) or CAG/CAGCACUCAGACUACGUGCACCUCUG/CUGC (SEQ ID NO: 7), or a sequence which is at least 60% identical thereto;
- (d) the regulatable intron comprises or consists of one of the following sequences:
- CNG/CAGCACUCAGACUACGUGCACCUCUG/CNG (SEQ ID NO: 8);
 - CNG/CAGCACUCAGACUACGUGCUCUCCUCUG/CNG (SEQ ID NO: 9);
 - CNG/CAGCACUCAGACUACGUGCCCCUCUG/CNG (SEQ ID NO: 10);
 - CNG/CAGCACUCAGACUACGUGCGCCUCUG/CNG (SEQ ID NO: 11);
 - CNG/CAGCACUCAGACUAUGUGCACCUCUG/CNG (SEQ ID NO: 12);
 - CAG/CAGCACUCAGACUACGUGCACCUCUG/CUGC (SEQ ID NO: 7);
 - CAG/CAGCACUCAGACUACGUGCUCUCCUCUG/CUGC (SEQ ID NO: 13);
 - CAG/CAGCACUCAGACUACGUGCCCCUCUG/CUGC (SEQ ID NO: 14);
 - CAG/CAGCACUCAGACUACGUGCGCCUCUG/CUGC (SEQ ID NO: 15); and
 - CAG/CAGCACUCAGACUAUGUGCACCUCUG/CUGC (SEQ ID NO: 16); and/or
- (e) the regulatable intron comprises the sequence CAG/CUGCAGCACUCAGACUACGUGCACCUCUG/CUG (SEQ ID NO: 17) or CAG/CUGCAGCACUCAGACUACGUGCACCUCUG/CUGG (SEQ ID NO: 27), wherein / represents a cleavage site.
7. The synthetic nucleic acid expression construct of any one of the preceding claims, comprising an inducible promoter operably linked to the nucleic acid sequence encoding the expression product; optionally wherein the inducible promoter is an unfolded protein response (UPR) inducible promoter.

8. The synthetic nucleic acid expression construct of any one of the preceding claims,
wherein the nucleic acid sequence encoding an expression product is (a) a transgene,
and/or (b) encodes a protein, suitably wherein the protein is an enzyme, an antibody or
antibody fragment, a viral protein, a therapeutic protein, or a toxic protein.
9. A nucleic acid comprising a sequence encoding an expression product, the sequence
encoding an expression product comprising a sequence which encodes a regulatable
intron, said regulatable intron being an intron which comprises an excisable sequence
which is capable of being spliced out of a transcript produced from the synthetic
expression construct via the unfolded protein response (UPR) system in the cell, thereby
resulting in a transcript encoding a functional expression product, and wherein the
expression product is not the XBP1 protein, Had protein, bZIP60 protein or a homologue
thereof.
10. A vector comprising a synthetic nucleic acid expression construct according to any one of
claims 1 to 8 or a nucleic acid according to claim 9.
11. A pharmaceutical composition comprising a nucleic acid expression construct according
to any one of claims 1 to 8, a nucleic acid according to claim 9, or a vector according to
claim 10.
12. A cell comprising a nucleic acid expression construct according to any one of claims 1 to
8, a nucleic acid according to claim 9 or a vector according to claim 10; preferably
wherein the cell is a eukaryotic cell, and suitably a fungal cell (e.g., yeast cell), an animal
(metazoan) cell (e.g., mammalian cells), or a plant cell; and/or
wherein the nucleic acid expression construct, nucleic acid, or vector encodes an
expression product that is toxic to the cell.
13. A method for producing an expression product, the method comprising:
 - a) providing a population of eukaryotic cells comprising a synthetic nucleic acid
expression construct according to according to any one of claims 1 to 8;
 - b) treating said population of cells so as to induce the unfolded protein response,
thereby inducing splicing of the excisable sequence out of the regulatable intron;
 - c) incubating said population of cells under suitable conditions for production of the
expression product; and
 - d) isolating the expression product from said population of cells.
14. The method according to claim 13, wherein the method comprises incubating said
population of cells under conditions suitable for growth of the cells prior to step b) of treating
said population of cells so as to induce the unfolded protein response (UPR); and/or

wherein step b) comprises applying a stress to said cells, said stress being suitable to induce the UPR.

15. The method according to claim 13 or claim 14, and/or wherein (i) step b) comprises administering a chemical agent that is able to induce the UPR in said cells and preferably wherein the chemical agent is one or more of forskolin, dithiothreitol (DTT), tunicamycin, thapsigargin, a saturated fatty acid, an agent that is able to downregulate stearyl-CoA desaturase enzyme activity, optionally wherein step b) comprises transfecting said population of cells with an expression vector that is capable of expressing an inducer protein, preferably a heterologous protein, in said cells; or wherein step b) comprises inducing expression of the inducer protein from an expression vector that was previously introduced into the cells.
16. The method according to any one of claims 13 to 15, wherein step b) comprises:
 - (i) expressing an inducer protein in said population of eukaryotic cells so as to induce the UPR response in the population of cells; or wherein step b) comprises exposing the cells to hypoxia or carbohydrate deprivation.
17. The method according to any one of claims 13 to 16, wherein the population of cells comprises fungal cells, animal cells, or plant cells, preferably wherein the population of cells comprises mammalian cells.
18. The method according to any one of claims 13 to 17, wherein the method comprises, prior to step a), the step of introducing the nucleic acid expression construct into the cells.
19. Use of a nucleic acid expression construct according to any one of claims 1 to 8, a nucleic acid according to claim 9, a vector according to claim 10, or a pharmaceutical composition according to claim 11 for the manufacture of a medicament, optionally for gene therapy.
20. A method for gene therapy in a subject in need of said gene therapy comprising:
 - introducing into the subject a gene therapy vector comprising a nucleic acid expression construct according to any one of claims 1 to 8, the nucleic acid expression construct comprising a sequence encoding a therapeutic expression product such that the gene therapy vector delivers the nucleic acid expression construct to target cells of the subject; and
 - expressing a therapeutically effective amount of the functional therapeutic expression product in target cells of subject.

1/14

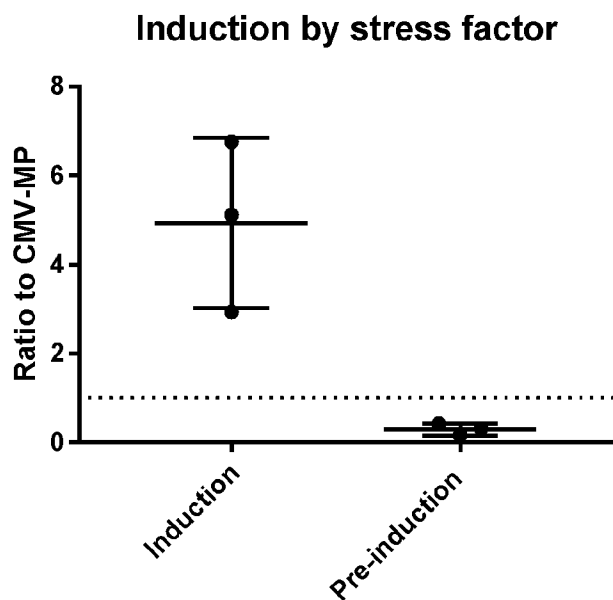


Fig. 1

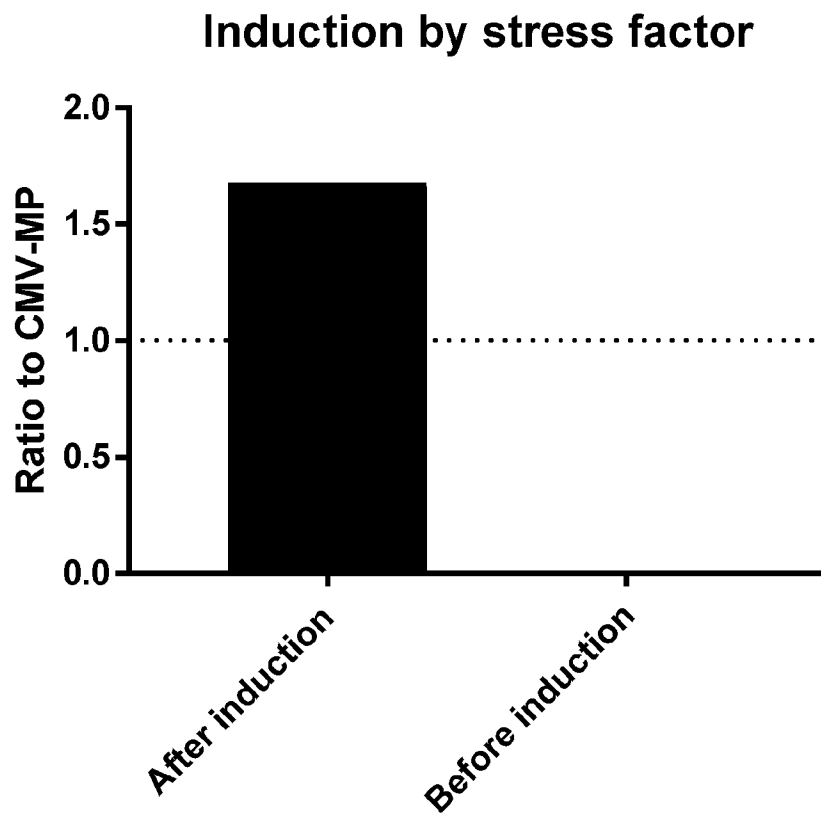


Fig. 2a

2/14

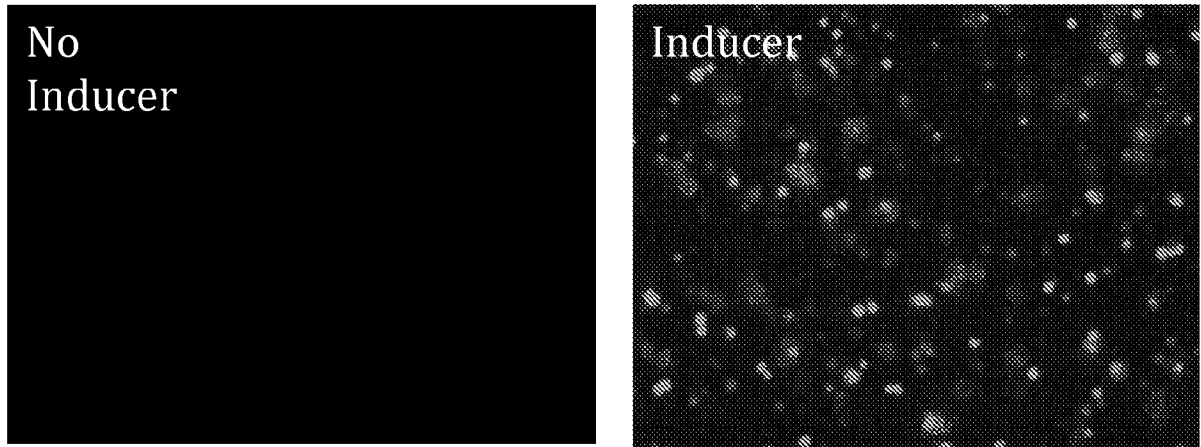


Fig. 2b

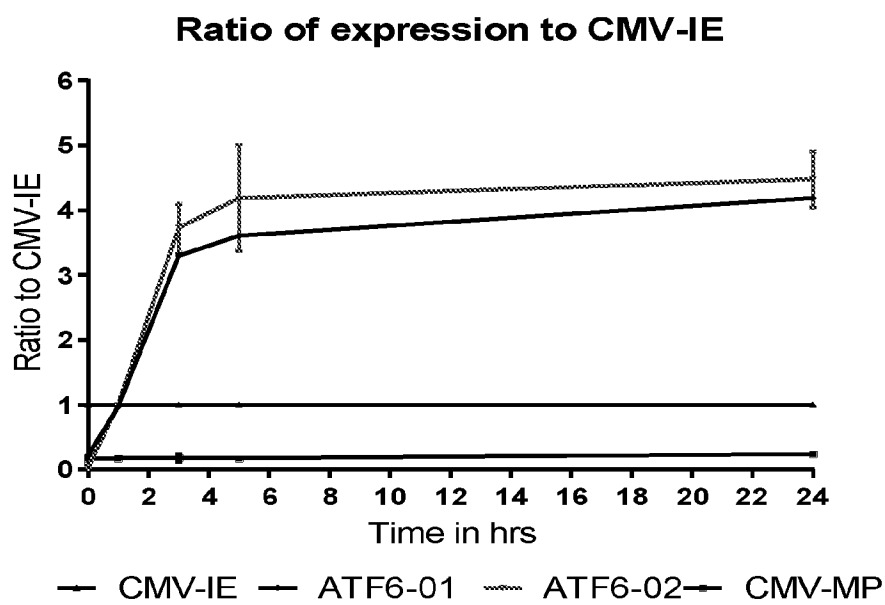


Fig. 3

3/14

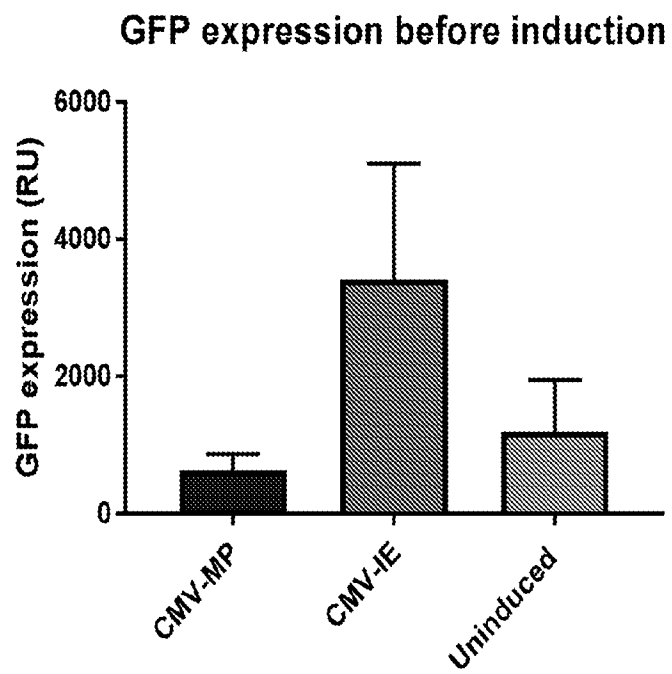


Fig. 4

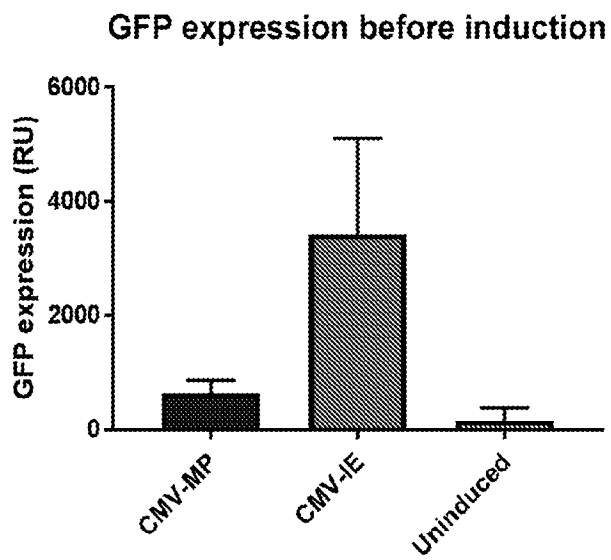
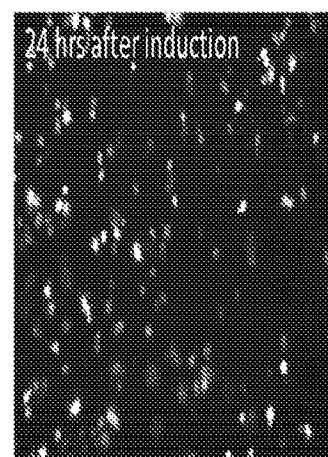
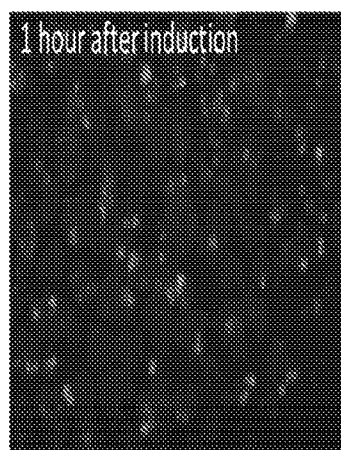
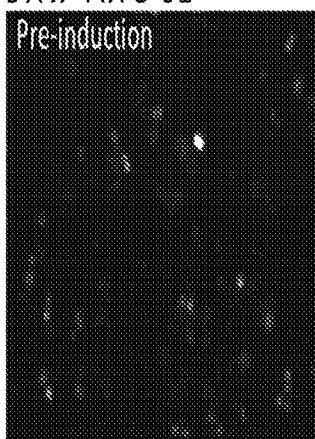


Fig. 5

4/14

SYNP-ATF6-01



SYNP-ATF6-02

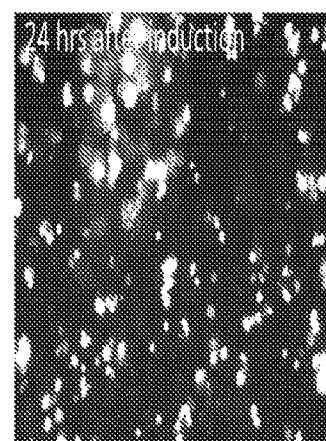
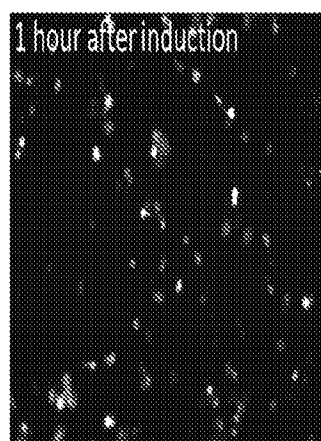
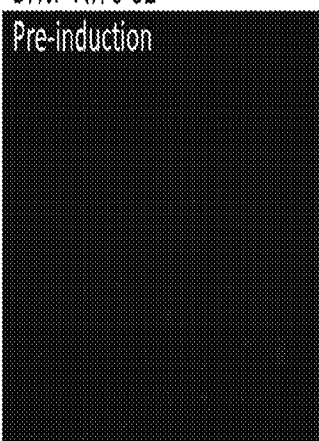


Fig. 6

5/14

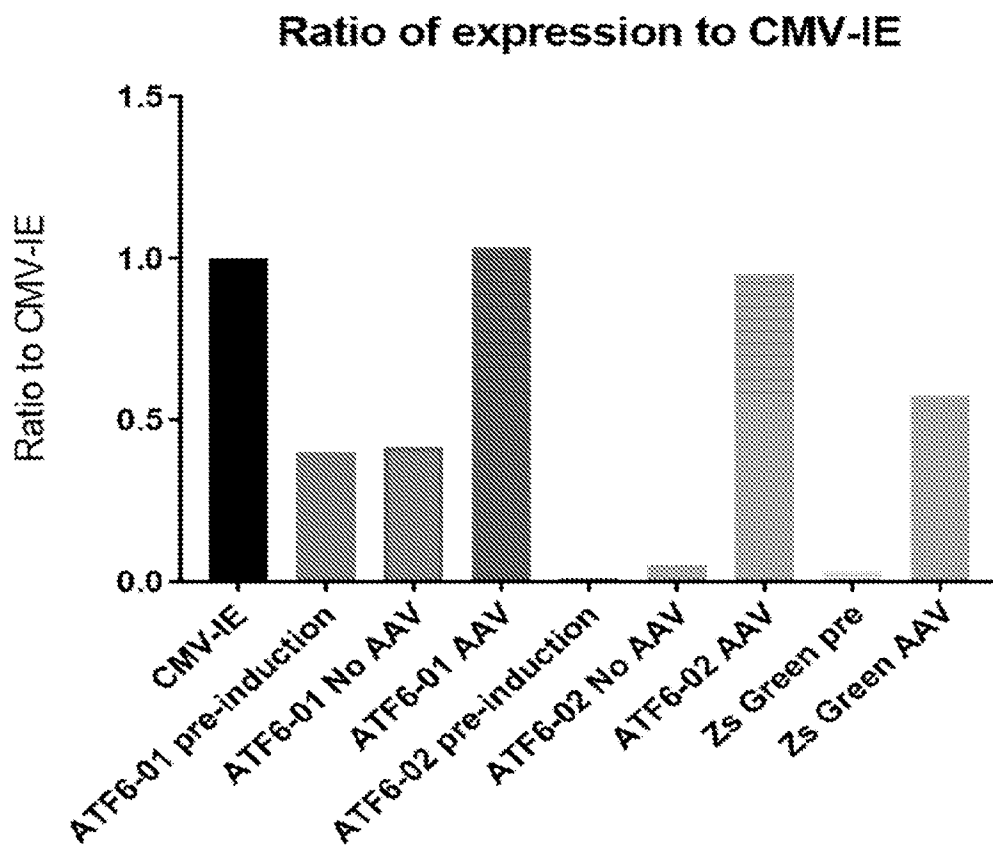


Fig. 7

6/14

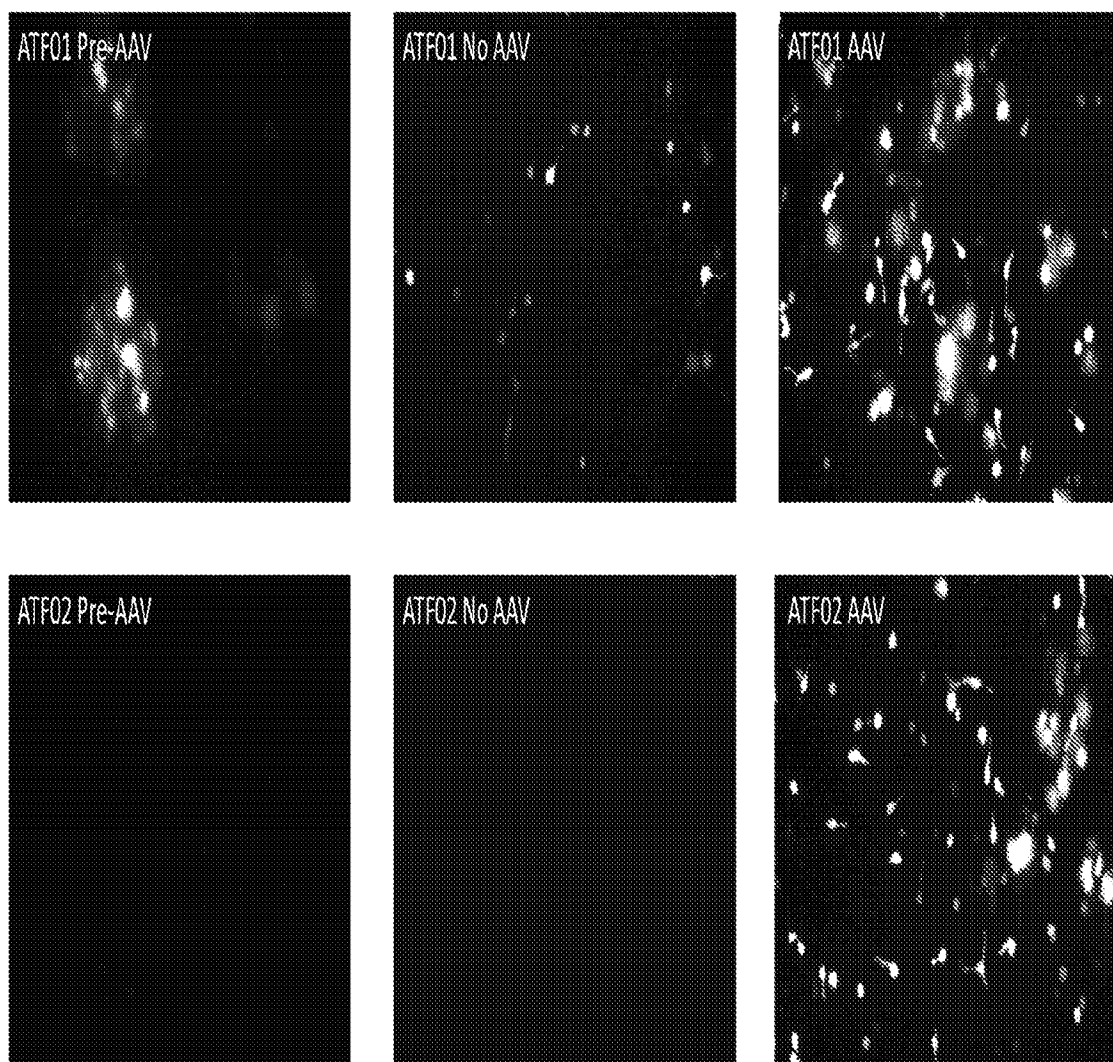


Fig. 8

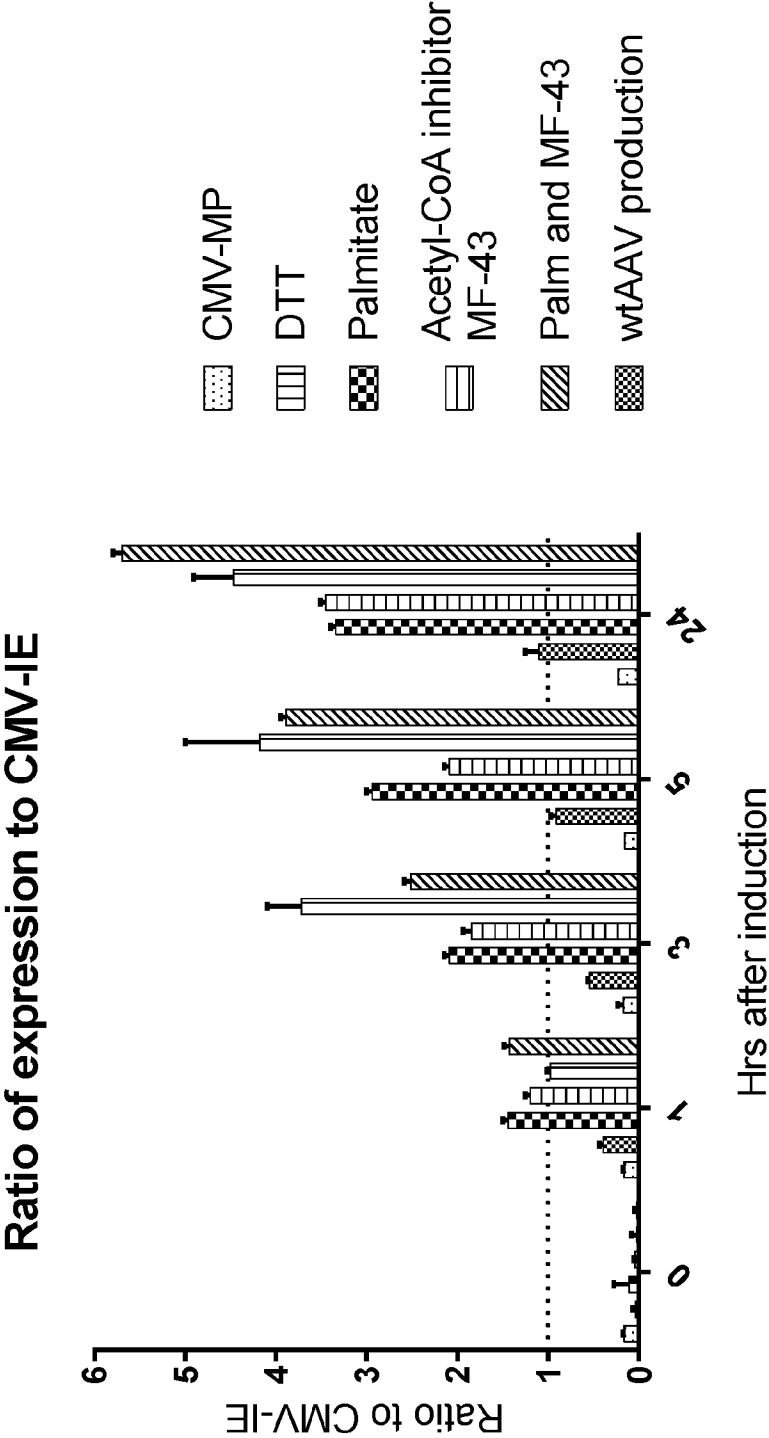


Fig. 9

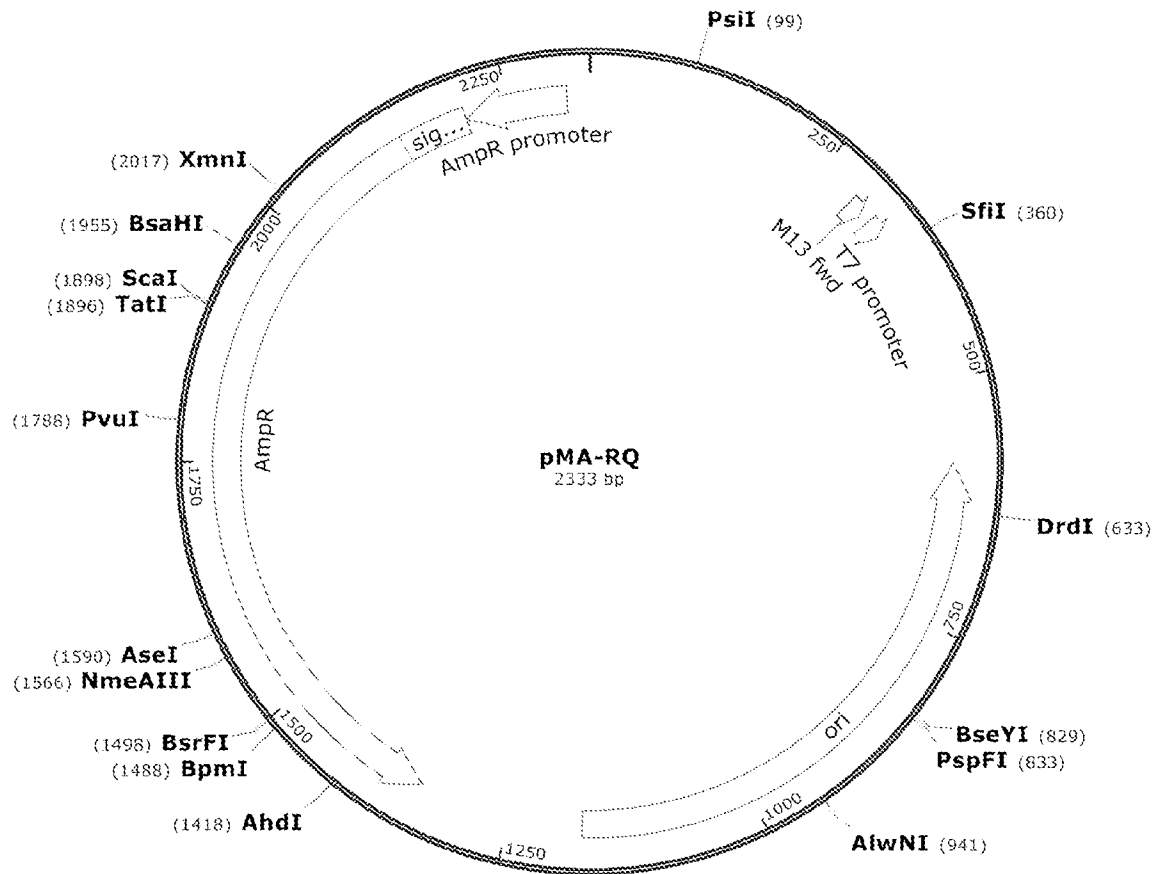


Fig. 10

9/14

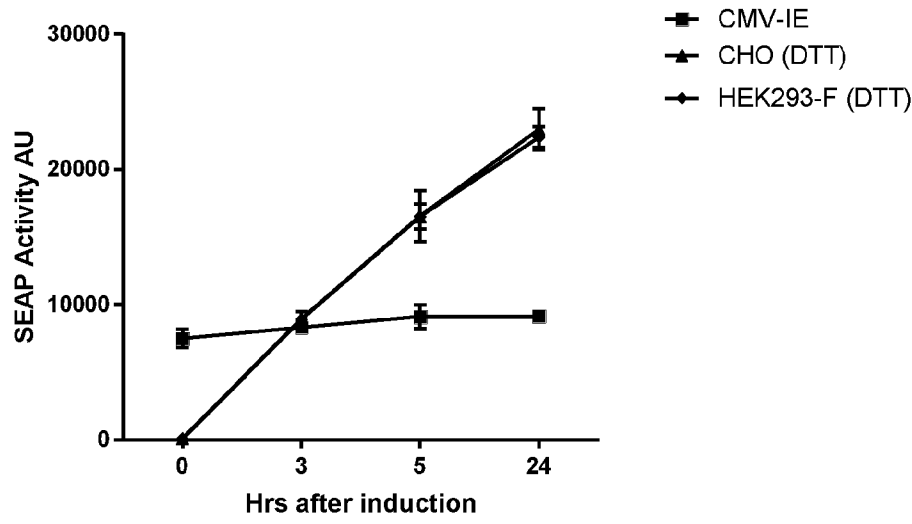
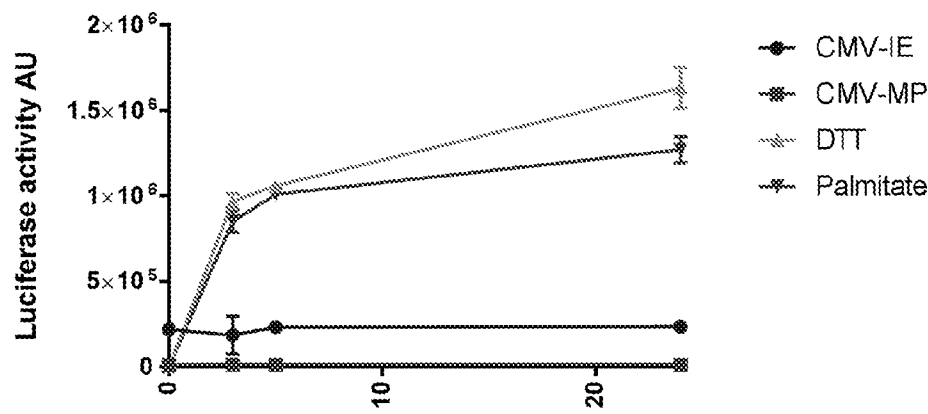


Fig. 11

A



B

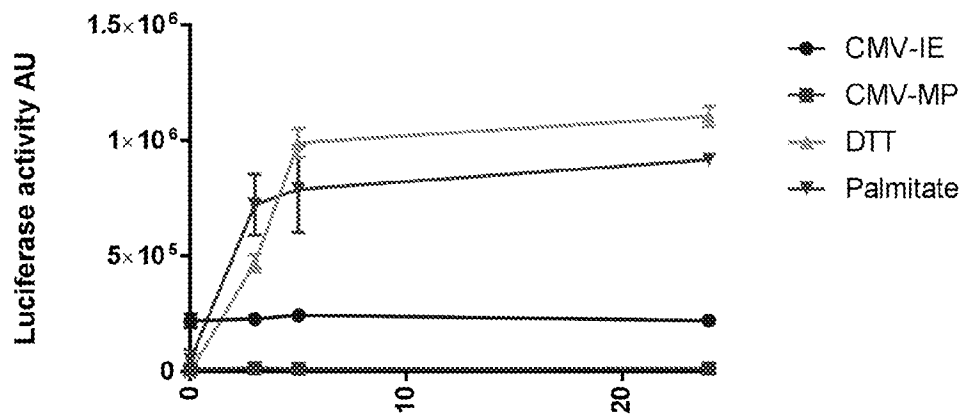


Fig. 12

10/14

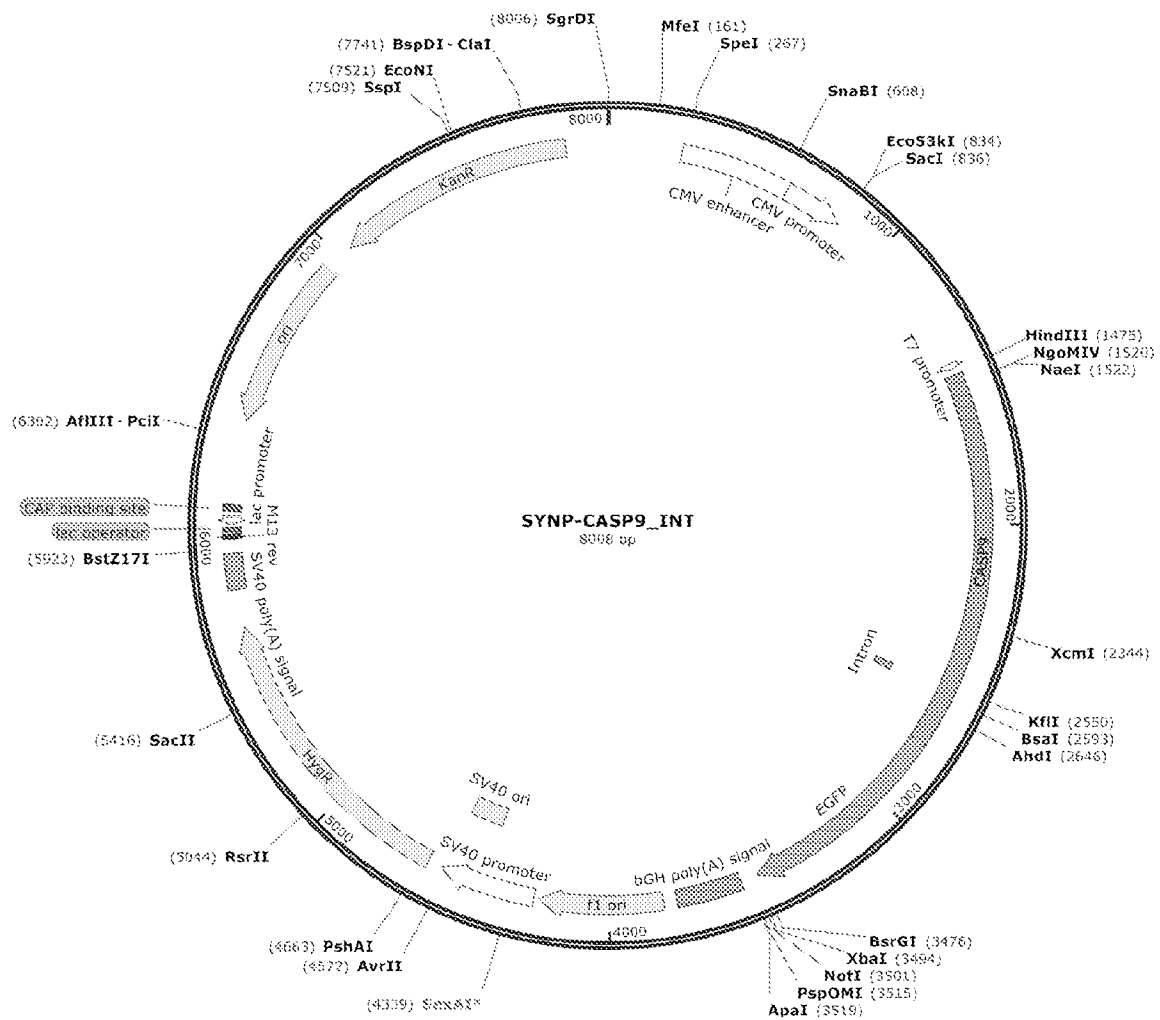


Fig. 13

11/14

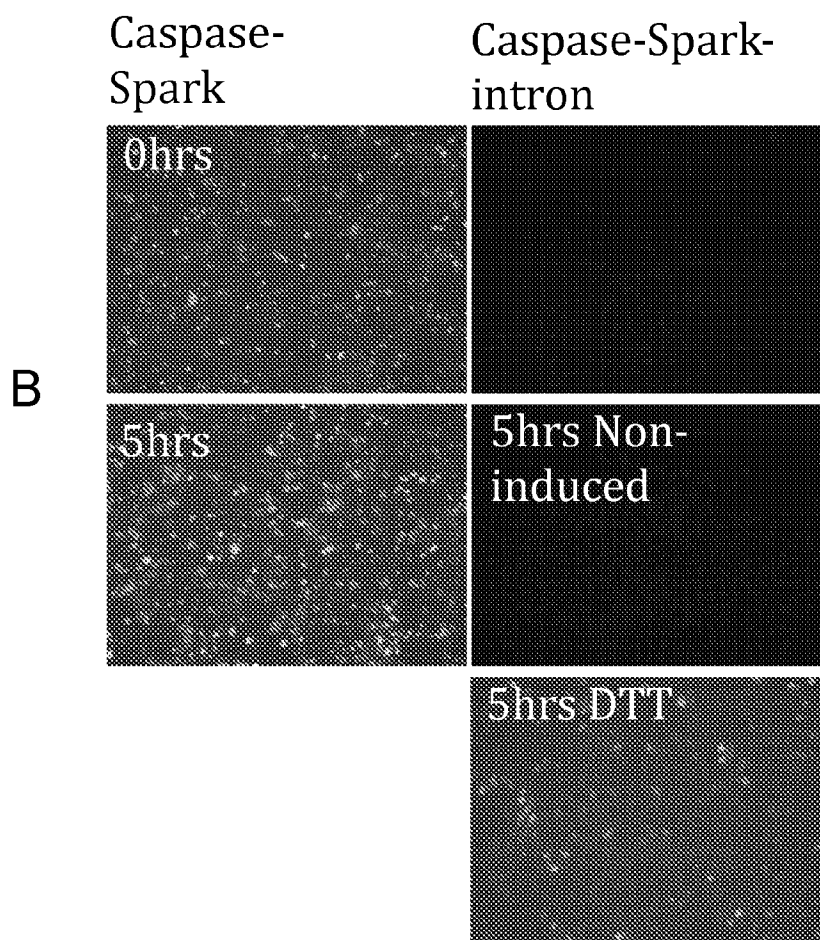
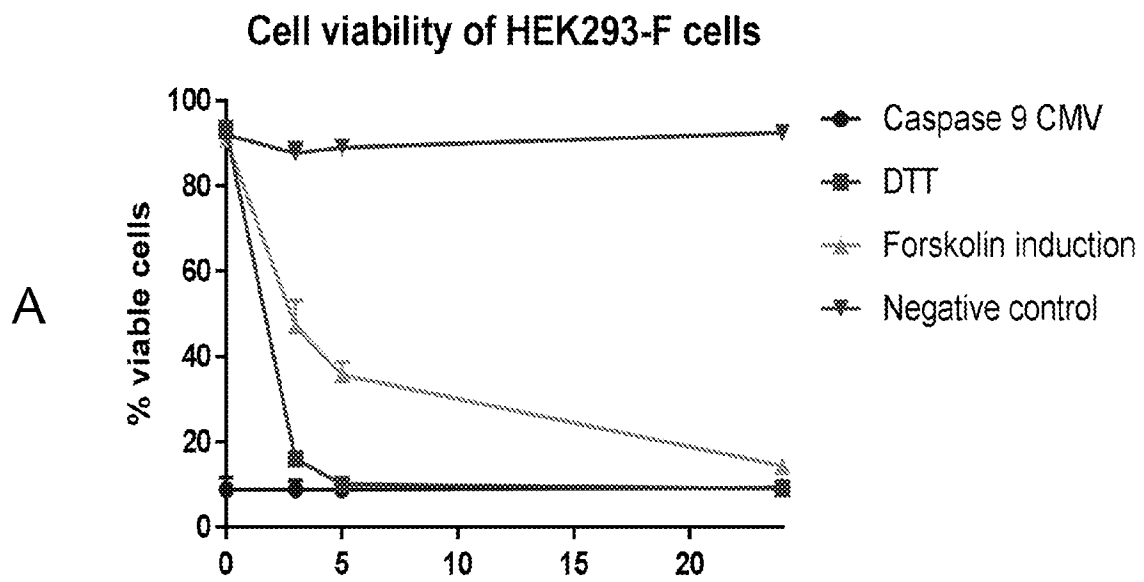


Fig. 14

12/14

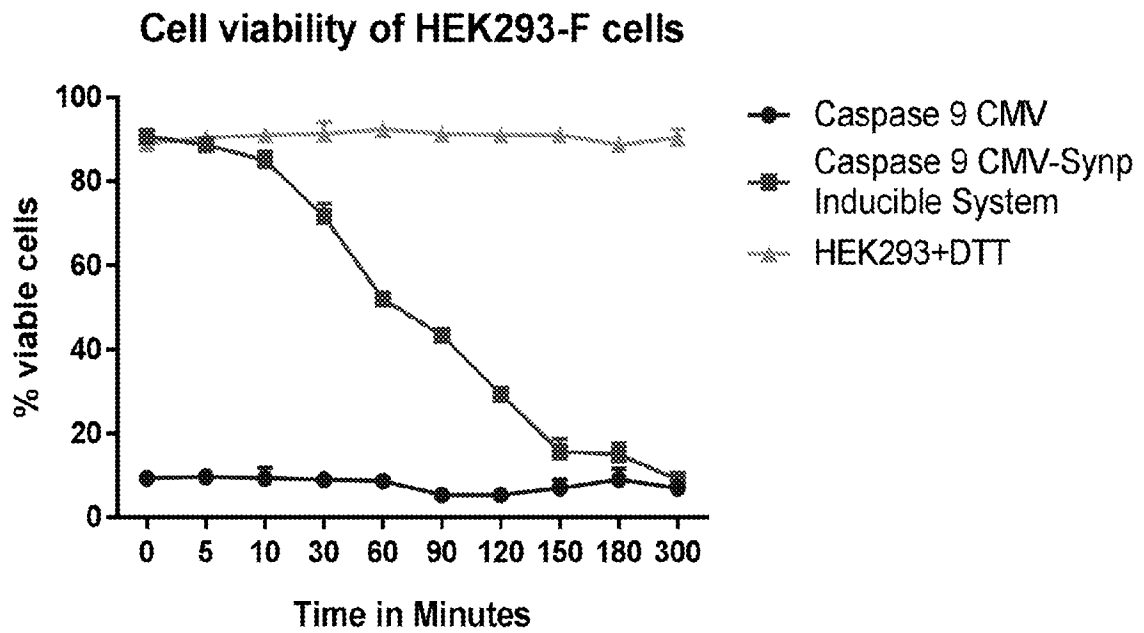


Fig. 15

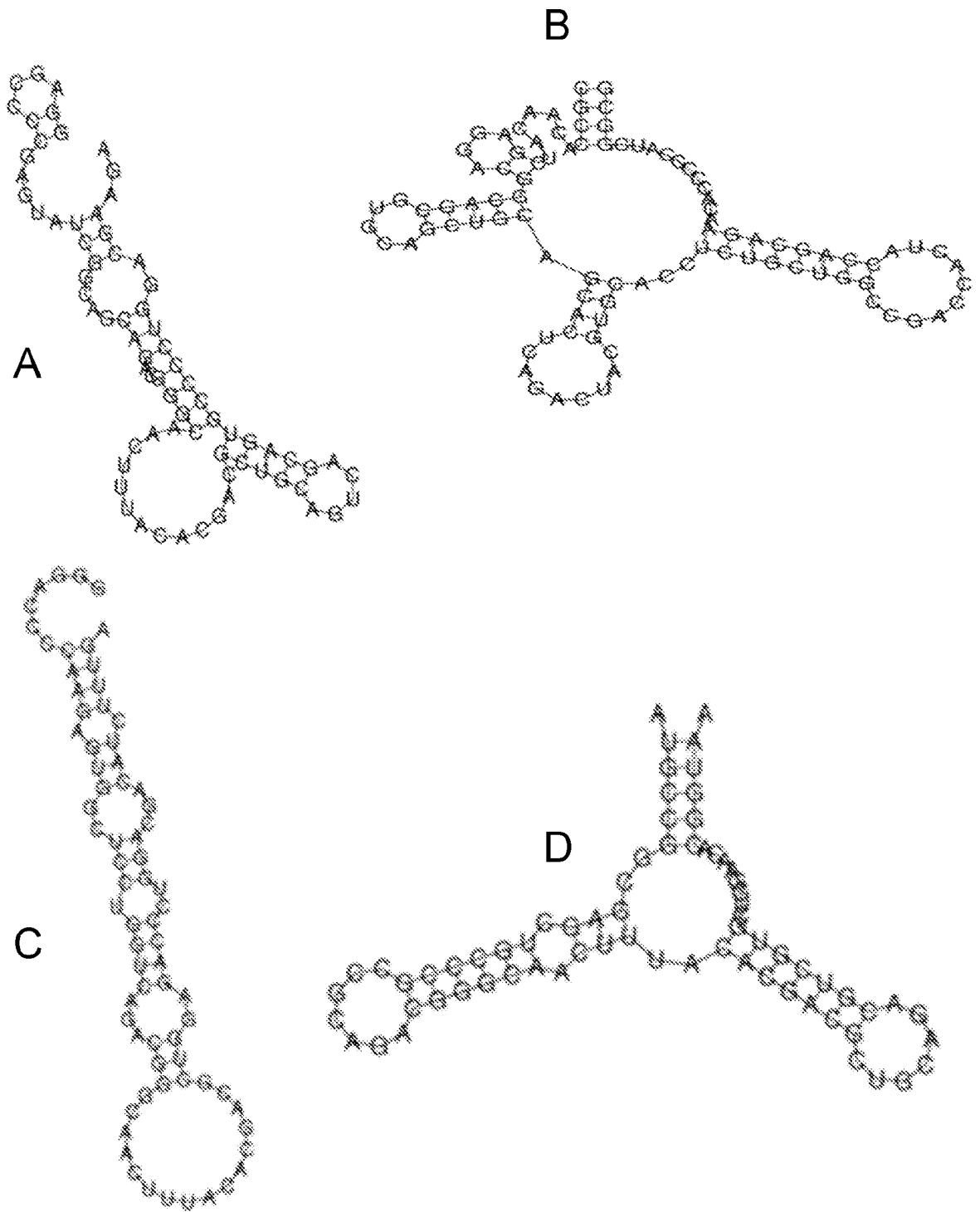


Fig. 16

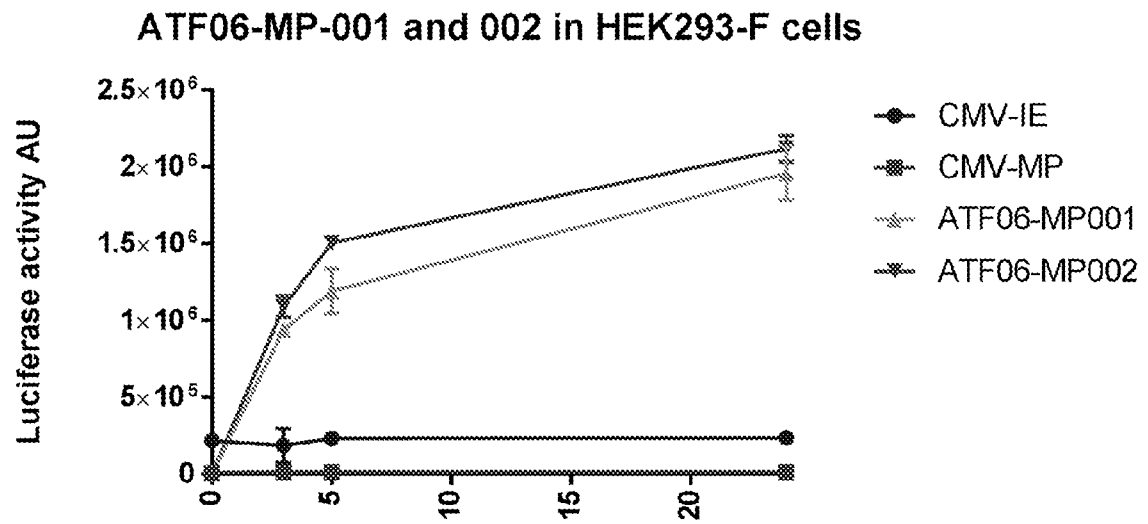


Fig. 17

pctgb2018052387-seq1.txt
SEQUENCE LISTING

<110> Synpromics Limited

<120> Expression Control Using Regulatable Intron

<130> P241689W0

<150> GB1713545.0

<151> 2017-08-23

<160> 61

<170> PatentIn version 3.5

<210> 1

<211> 20

<212> RNA

<213> Unknown

<220>

<223> sequence of mammalian XBP1 intron inside of the IRE 1 cleavage sites

<400> 1
cacucagacu acgugcaccu 20

<210> 2

<211> 20

<212> RNA

<213> Unknown

<220>

<223> variant of mammalian XBP1 intron inside of the IRE 1 cleavage sites

<400> 2
cacucagacu acgugcuccu 20

<210> 3

<211> 20

<212> RNA

<213> Unknown

<220>

<223> variant of mammalian XBP1 intron inside of the IRE 1 cleavage sites

<400> 3
cacucagacu acgugccccc 20

<210> 4
 <211> 20
 <212> RNA
 <213> Unknown

<220>
 <223> variant of mammalian XBP1 intron inside of the IRE 1 cleavage sites

<400> 4
 cacucagacu acgugcgccu 20

<210> 5
 <211> 20
 <212> RNA
 <213> Unknown

<220>
 <223> variant of mammalian XBP1 intron inside of the IRE 1 cleavage sites

<400> 5
 cacucagacu augugcaccu 20

<210> 6
 <211> 33
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> regulatable intron

<220>
 <221> misc_feature
 <222> (2)..(2)
 <223> n is a, c, g, or u

<220>
 <221> misc_feature
 <222> (5)..(5)
 <223> n is a, c, g, or u

<220>
 <221> misc_feature
 <222> (28)..(28)
 <223> n is a, c, g, or u

```

<220>
<221> misc_feature
<222> (31)..(31)
<223> n is a, c, g, or u

<400> 6
cngcngcacu cagacuacgu gcaccucngc ngc 33

<210> 7
<211> 33
<212> RNA
<213> Unknown

<220>
<223> mammalian XBP1 intron sequence

<400> 7
cagcagcacu cagacuacgu gcaccucugc ugc 33

<210> 8
<211> 32
<212> RNA
<213> Artificial Sequence

<220>
<223> regulatable intron

<220>
<221> misc_feature
<222> (2)..(2)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (31)..(31)
<223> n is a, c, g, or u

<400> 8
cngcagcacu cagacuacgu gcaccucugc ng 32

<210> 9
<211> 32
<212> RNA
<213> Artificial Sequence

<220>
<223> regulatable intron

```

<220>
 <221> misc_feature
 <222> (2)..(2)
 <223> n is a, c, g, or u

<220>
 <221> misc_feature
 <222> (31)..(31)
 <223> n is a, c, g, or u

<400> 9
 cngcagcacu cagacuacgu gcuccucugc ng

32

<210> 10
 <211> 32
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> regulatable intron

<220>
 <221> misc_feature
 <222> (2)..(2)
 <223> n is a, c, g, or u

<220>
 <221> misc_feature
 <222> (31)..(31)
 <223> n is a, c, g, or u

<400> 10
 cngcagcacu cagacuacgu gccccucugc ng

32

<210> 11
 <211> 32
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> regulatable intron

<220>
 <221> misc_feature
 <222> (2)..(2)
 <223> n is a, c, g, or u

<220>
 <221> misc_feature
 <222> (31)..(31)
 <223> n is a, c, g, or u

 <400> 11
 cngcagcacu cagacuacgu gcgccucugc ng 32

<210> 12
 <211> 32
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> regulatable intron

<220>
 <221> misc_feature
 <222> (2)..(2)
 <223> n is a, c, g, or u

<220>
 <221> misc_feature
 <222> (31)..(31)
 <223> n is a, c, g, or u

<400> 12
 cngcagcacu cagacuaugu gcaccucugc ng 32

<210> 13
 <211> 33
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> regulatable intron

<400> 13
 cagcagcacu cagacuacgu gcuccucugc ugc 33

<210> 14
 <211> 33
 <212> RNA
 <213> Artificial Sequence

<220>
 <223> regulatable intron

<400> 14
cagcagcacu cagacuacgu gccccucugc ugc 33

<210> 15
<211> 33
<212> RNA
<213> Artificial Sequence

<220>
<223> regulatable intron

<400> 15
cagcagcacu cagacuacgu gcgccucugc ugc 33

<210> 16
<211> 33
<212> RNA
<213> Artificial Sequence

<220>
<223> regulatable intron

<400> 16
cagcagcacu cagacuaugu gcaccucugc ugc 33

<210> 17
<211> 35
<212> RNA
<213> Artificial Sequence

<220>
<223> mammalian XBP1 intron sequence with addition of trinucleotide CUG

<400> 17
cagcugcagc acucagacua cgugcaccuc ugcag 35

<210> 18
<211> 19
<212> DNA
<213> Unknown

<220>
<223> ERSE1 site

<220>
<221> misc_feature
<222> (6)..(14)

<223> n is a, c, g, or t

<400> 18

ccaatnnnnn nnnnccacg

19

<210> 19

<211> 11

<212> DNA

<213> Unknown

<220>

<223> ERSE2 site

<220>

<221> misc_feature

<222> (6)..(6)

<223> n is a, c, g, or t

<400> 19

attggnccac g

11

<210> 20

<211> 54

<212> DNA

<213> Unknown

<220>

<223> UPR inducible promoter sequence

<400> 20

tgacgtgctt gacgtgcttg acgtgcttga cgtgcttgac gtgcttgacg tgct

54

<210> 21

<211> 84

<212> DNA

<213> Unknown

<220>

<223> CMV-minimal promoter

<400> 21

aggtctatat aagcagagct cgtttagtga accgtcagat cgcctagata cgccatccac

60

gctgttttga cctccataga agat

84

<210> 22

<211> 169

pctgb2018052387-seq1.txt

<212> DNA
<213> Unknown

<220>
<223> inducible promoter

<400> 22
tgacgtgctt gacgtgcttg acgtgcttga cgtgcttgac gtgcttgacg tgctggtacc 60
gtcgacgata tcggatccag gtctatataa gcagagctcg tttagtgaac cgtcagatcg 120
cctagatacg ccatccacgc tgttttgacc tccatagaag atcgccacc 169

<210> 23
<211> 26
<212> RNA
<213> Unknown

<220>
<223> mammalian WT intron sequence

<400> 23
cagcacucag acuacgugca ccucug 26

<210> 24
<211> 33
<212> RNA
<213> Unknown

<220>
<223> mammalian WT intron sequence flanked by splice recognition site
sequences

<400> 24
ccgcagcacu cagacuacgu gcaccucugc agc 33

<210> 25
<211> 717
<212> DNA
<213> Artificial Sequence

<220>
<223> EGFP gene with intron insertion

<400> 25
atggtgagca agggcgagga gctgttcacc ggggtggtgc ccatcctggt cgagctggac 60
ggcgacgtaa acggccacaa gttcagcgtg tccggcgagg gcgagggcga tgccacctac 120

pctgb2018052387-seq1.txt

```

ggcaagctga ccctgaagtt catctgcacc accggcaagc tgcccgtgcc ctggcccacc      180
ctcgtgacca ccctgaccta cggcgtgcag tgcttcagcc gctaccccga ccacatgaag      240
cagcacgact tcttcaagtc cgccatgccc gaaggctacg tccaggagcg caccatcttc      300
ttcaaggacg acggcaacta caagaccgc gccgaggtga agttcgaggg cgacaccctg      360
gtgaaccgca tcgagctgaa gggcatcgac ttcaaggagg acggcaacat cctggggcac      420
aagctggagt acaactacaa cagccacaac gtctatatca tggccgacaa gcagaagaac      480
ggcatcaagg tgaacttcaa gatccgccac aacatcgagg acggcagcgt gcagctcgcc      540
gaccactacc agcagaacac ccccatcggc gacggccccg tgctgctgcc cgacaaccac      600
tacctgagca ccagtcctgc cctgagcaaa gaccccaacg agaagcgcgga tcacatggtc      660
ctgctggagt tcgtgaccgc cgccgggatc actctcggca tggacgagct gtacaag      717

```

<210> 26
 <211> 239
 <212> PRT
 <213> Unknown

<220>
 <223> EGFP protein sequence

<400> 26

```

Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu
1           5           10           15

```

```

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

```

```

Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile
          35           40           45

```

```

Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
          50           55           60

```

```

Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys
65           70           75           80

```

```

Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu

```

Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
 100 105 110

Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
 115 120 125

Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
 130 135 140

Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn
 145 150 155 160

Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser
 165 170 175

Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
 180 185 190

Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu
 195 200 205

Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe
 210 215 220

Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys
 225 230 235

<210> 27

<211> 36

<212> RNA

<213> Artificial Sequence

<220>

<223> mammalian XBP1 intron sequence with addition of trinucleotide CUG

<400> 27

cagcugcagc acucagacua cgugcaccuc ugcugg

36

<210> 28

pctgb2018052387-seq1.txt

<211> 1150

<212> DNA

<213> Artificial Sequence

<220>

<223> EGFP sequence with regulatable XBP1 intron, CMV-MP and SV40 polyA tail

<400> 28

```

ggtaccgtcg acgatatcgg atccaggtct atataagcag agctcgttta gtgaaccgtc      60
agatcgcccta gatacgccat ccacgtgtt ttgacctcca tagaagatcg ccaccatggt      120
gagcaagggc gaggagctgt tcaccggggt ggtgcccata ctggtcgagc tggacggcga      180
cgtaaaccggc cacaagttca gcgtgtccgg cgagggcgag ggcgatgcca cctacggcaa      240
gctgaccctg aagttcatct gcaccaccgg caagctgccc gtgccctggc ccaccctcgt      300
gaccaccctg acctacggcg tgcagtgtt cagccgtac cccgaccaca tgaagcagca      360
cgacttcttc aagtccgcca tgcccgaagg ctacgtccag gagcgcacca tcttcttcaa      420
ggacgacggc aactacaaga cccgcgccga ggtgaagttc gagggcgaca ccctggtgaa      480
ccgcatcgag ctgaagggca tcgacttcaa ggaggacggc aacatcctgg ggcacaagct      540
ggagtacaac tacaacagcc acaacgtcta tatcatggcc gacaagcaga agaacggcat      600
caaggtgaac ttcaagatcc gccacaacat cgaggacggc agcgtgcagc tgcagcactc      660
agactacgtg cacctctgct ggccgaccac taccagcaga acacccccat cggcgacggc      720
cccgtgctgc tgcccgacaa ccactacctg agcaccacgt cgcacctgag caaagacccc      780
aacgagaagc gcgatcacat ggtcctgctg gagttcgtga ccgccgccgg gatcactctc      840
ggcatggacg agctgtacaa gtgagacgat cttatcgga ttttaccaca tttgtagagg      900
ttttacttgc tttaaaaaac ctcccacatc tccccctgaa cctgaaacat aaaatgaatg      960
caattgttgt tgtaacttg tttattgcag cttataatgg ttacaaataa agcaatagca     1020
tcacaaatth cacaataaaa gcattttttt cactgcattc tagttgtggt ttgtccaaac     1080
tcatcaatgt atcttatcat gtctgctcga aggcgcgccc tgggcctcat gggccttccg     1140
ctcactgccc                                     1150

```

<210> 29

<211> 211

<212> PRT

<213> Artificial Sequence

<220>

<223> truncated EGFP protein - start to first stop codon

<400> 29

Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu
 1 5 10 15

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

Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile
 35 40 45

Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
 50 55 60

Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys
 65 70 75 80

Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
 85 90 95

Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
 100 105 110

Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
 115 120 125

Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
 130 135 140

Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn
 145 150 155 160

Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser
 165 170 175

pctgb2018052387-seql.txt

Val Gln Leu Glu His Ser Asp Tyr Val His Leu Cys Trp Pro Thr Thr
180 185 190

Thr Ser Arg Thr Pro Pro Ser Ala Thr Ala Pro Cys Cys Cys Pro Thr
195 200 205

Thr Thr Thr
210

<210> 30
<211> 889
<212> DNA
<213> Artificial Sequence

<220>
<223> SYN-P-ATF6-02 construct

<400> 30
tgacgtgctt gacgtgcttg acgtgcttga cgtgcttgac gtgcttgacg tgctgggtacc 60
gtcgacgata tcggatccag gtctatataa gcagagctcg tttagtgaac cgtcagatcg 120
cctagatacg ccatccacgc tgttttgacc tccatagaag atcgccacca tggtagagcaa 180
gggagaggag ctgttcaccg ggggtggtgcc catcctgggtc gagctggacg gcgacgtaaa 240
cggccacaag ttcagcgtgt ccggcgaggg cgagggcgat gccacctacg gcaagctgac 300
cctgaagttc atctgcacca ccggcaagct gcccgtgccc tggccccaccc tcgtgaccac 360
cctgacctac ggcgtgcagt gcttcagccg ctaccccgac cacatgaagc agcacgactt 420
cttcaagtcc gccatgcccg aaggctacgt ccaggagcgc accatcttct tcaaggacga 480
cggcaactac aagaccgacg ccgaggtgaa gttcgagggc gacaccctgg tgaaccgcat 540
cgagctgaag ggcacgact tcaaggagga cggcaacatc ctggggcaca agctggagta 600
caactacaac agccacaacg tctatatcat ggccgacaag cagaagaacg gcatcaaggt 660
gaacttcaag atccgccaca acatcgagga cggcagcgtg cagctcgccg accactacca 720
gcagaacacc cccatcggcg acggccccgt gctgctgccc gacaaccact acctgagcac 780
ccagtccgcc ctgagcaaag accccaacga gaagcgcgat cacatgggtcc tgctggagtt 840
cgtgaccgcc gccgggatca ctctcgcat ggacgagctg tacaagtga 889

pctgb2018052387-seq1.txt

<210> 31
 <211> 845
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> SYN-ATF6-02 with excised intron sequence

<400> 31
 tgacgtgctt gacgtgcttg acgtgcttga cgtgcttgac gtgcttgacg tgctgggtacc 60
 gtcgacgata tcggatccag gtctatataa gcagagctcg tttagtgaac cgtcagatcg 120
 cctagatacg ccatccacgc tgttttgacc tccatagaag atcgccacca tgggtgagcaa 180
 gggcgaggag ctgttcaccg ggggtgggtgcc catcctgggtc gagctggacg gcgacgtaaa 240
 cggccacaag ttcagcgtgt ccggcgaggg cgagggcgat gccacctacg gcaagctgac 300
 cctgaagttc atctgcacca ccggcaagct gcccggtgccc tggcccaccc tcgtgaccac 360
 cctgacctac ggcgtgcagt gcttcagccg ctaccccgac cacatgaagc agcacgactt 420
 cttcaagtcc gccatgcccc aaggctacgt ccaggagcgc accatcttct tcaaggacga 480
 cggcaactac aagaccgcg ccgaggtgaa gtctgagggc gacaccctgg tgaaccgcat 540
 cgagctgaag ggcacgact tcaaggagga cggcaacatc ctggggcaca agctggagta 600
 caactacaac agccacaacg tctatatcat ggccgacaag cagaagaacg gcatcaaggt 660
 gaacttcaag atccgccaca acatcgagga cggcagcgtg cagctggccg accactacca 720
 gcagaacacc cccatcggcg acggccccgt gctgctgccc gacaaccact acctgagcac 780
 ccagtccgcc ctgagcaaag accccaacga gaagcgcgat cacatggtcc tgctggagtt 840
 cgtga 845

<210> 32
 <211> 26
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> DNA coding for excised region of intron

<400> 32
 cagacgggca actttacacg acgctg 26

pctgb2018052387-seq1.txt

<210> 33

<400> 33

000

<210> 34

<211> 1892

<212> DNA

<213> Artificial Sequence

<220>

<223> SEAP expression construct

<400> 34

tgacgtgctg atgatgcgta gctagtagtt gacgtgctga tgatgcgtag ctagtagttg	60
acgtgctgat gatgcgtagc tagtagttga cgtgctgatg atgcgtagct agtagttgac	120
gtgctgatga tgcgtagcta gtagttgacg tgctgatgat gcgtagctag tagtgagtt	180
agcgtagctg aggtaccgtc gacgatatcg gatccaggtc tatataagca gagctcgttt	240
agtgaaccgt cagatcgctt agatacgcca tccacgtgtg ttgacctcc atagaagatc	300
gccacatgac tgctgctgct gctgctgctg ggcctgaggc tacagctctc cctgggcatc	360
atcccagttg aggaggagaa cccggacttc tggaaccgcg aggcagccga ggccctgggt	420
gccgccaaga agctgcagcc tgcacagaca gccgccaaga acctcatcat cttcctgggc	480
gatgggatgg ggggtgtctac ggtgacagct gccaggatcc taaaaggga gaagaaggac	540
aaactggggc ctgagatacc cctggctatg gaccgcttc catatgtggc tctgtccaag	600
acatacaatg tagacaaaca tgtgccagac agtggagcca cagccacggc ctacctgtgc	660
ggggtcaagg gcaacttcca gaccattggc ttgagtgcag ccgcccgtt taaccagtgc	720
aacacgacac gcggcaacga ggtcatctcc gtgatgaatc gggccaagaa agcaggggaag	780
tcagtgggag tggttaaccac cacacgagtg cagcacgcct cgccagccgg cacctacgcc	840
cacacggtga accgcaactg gtactcggac gccgacgtgc ctgcctcggc ccgccaggag	900
gggtgccagg acatcgctac gcagctcatc tccaacatgg acattgacgt gatcctaggt	960
ggaggccgaa agtacatgtt tcgcatggga accccagacc ctgagtaccc agatgactac	1020
agccaagggtg ggaccaggct ggacgggaag aatctggtgc aggaatggct ggccaagcgc	1080
cagggtgccc ggtatgtgtg gaaccgcact gagctcatgc aggtttccct ggaccctct	1140

pctgb2018052387-seq1.txt

gtgacccatc tcatgggcct ctttgagcct ggagacatga aatacgagat ccaccgagac	1200
tccacactgg acccctccct gatggagatg acagaggctg ccctgcgctt gctgagcagg	1260
aacccccgcg gcttcttcct cttcgtggag ggtggtcgca tcgaccacgg tcatcatgaa	1320
agcagggctt accgggcact gactgaaacg atcatgttcg acgacgcat tgagagggcg	1380
ggccagctca ccagcgagga ggacacgctg agcctcgtca ctgccgacca ctcccacgtc	1440
ttctccttcg gaggctaccc cctgagagg agctccatct tcgggctggc ccctggcaag	1500
gcccgggaca ggaaggccta cacggctctc ctatacgaa acggtccagg ctatgtgctc	1560
aaggacggcg cccggccgga tgttaccgag agcgagagcg ggagccccga gtatcggcag	1620
cagacgggca actttacacg acgctgcagt cagcagtgcc cctggacgaa gagacacacg	1680
caggcgagga cgtggcggtg ttcgcgcgcg gcccgaggc gcacctggtt cacggcgtgc	1740
aggagcagac cttcatagcg cacgtcatgg ccttcgccgc ctgcctggag ccctacaccg	1800
cctgcgacct ggcgcccccc gccggcacca ccgacgccgc gcacccgggt tactctagag	1860
tcggggcggc cggccgcttc gagcagacat ga	1892

<210> 35

<211> 483

<212> PRT

<213> Artificial Sequence

<220>

<223> truncated SEAP protein

<400> 35

Met	Leu	Leu	Leu	Leu	Leu	Leu	Leu	Gly	Leu	Arg	Leu	Gln	Leu	Ser	Leu
1				5				10					15		

Gly	Ile	Ile	Pro	Val	Glu	Glu	Glu	Asn	Pro	Asp	Phe	Trp	Asn	Arg	Glu
			20					25					30		

Ala	Ala	Glu	Ala	Leu	Gly	Ala	Ala	Lys	Lys	Leu	Gln	Pro	Ala	Gln	Thr
		35				40						45			

Ala	Ala	Lys	Asn	Leu	Ile	Ile	Phe	Leu	Gly	Asp	Gly	Met	Gly	Val	Ser
		50			55						60				

pctgb2018052387-seq1.txt

Thr Val Thr Ala Ala Arg Ile Leu Lys Gly Gln Lys Lys Asp Lys Leu
65 70 75 80

Gly Pro Glu Ile Pro Leu Ala Met Asp Arg Phe Pro Tyr Val Ala Leu
85 90 95

Ser Lys Thr Tyr Asn Val Asp Lys His Val Pro Asp Ser Gly Ala Thr
100 105 110

Ala Thr Ala Tyr Leu Cys Gly Val Lys Gly Asn Phe Gln Thr Ile Gly
115 120 125

Leu Ser Ala Ala Ala Arg Phe Asn Gln Cys Asn Thr Thr Arg Gly Asn
130 135 140

Glu Val Ile Ser Val Met Asn Arg Ala Lys Lys Ala Gly Lys Ser Val
145 150 155 160

Gly Val Val Thr Thr Thr Arg Val Gln His Ala Ser Pro Ala Gly Thr
165 170 175

Tyr Ala His Thr Val Asn Arg Asn Trp Tyr Ser Asp Ala Asp Val Pro
180 185 190

Ala Ser Ala Arg Gln Glu Gly Cys Gln Asp Ile Ala Thr Gln Leu Ile
195 200 205

Ser Asn Met Asp Ile Asp Val Ile Leu Gly Gly Gly Arg Lys Tyr Met
210 215 220

Phe Arg Met Gly Thr Pro Asp Pro Glu Tyr Pro Asp Asp Tyr Ser Gln
225 230 235 240

Gly Gly Thr Arg Leu Asp Gly Lys Asn Leu Val Gln Glu Trp Leu Ala
245 250 255

Lys Arg Gln Gly Ala Arg Tyr Val Trp Asn Arg Thr Glu Leu Met Gln
260 265 270

pctgb2018052387-seql.txt

Ala Ser Leu Asp Pro Ser Val Thr His Leu Met Gly Leu Phe Glu Pro
275 280 285

Gly Asp Met Lys Tyr Glu Ile His Arg Asp Ser Thr Leu Asp Pro Ser
290 295 300

Leu Met Glu Met Thr Glu Ala Ala Leu Arg Leu Leu Ser Arg Asn Pro
305 310 315 320

Arg Gly Phe Phe Leu Phe Val Glu Gly Gly Arg Ile Asp His Gly His
325 330 335

His Glu Ser Arg Ala Tyr Arg Ala Leu Thr Glu Thr Ile Met Phe Asp
340 345 350

Asp Ala Ile Glu Arg Ala Gly Gln Leu Thr Ser Glu Glu Asp Thr Leu
355 360 365

Ser Leu Val Thr Ala Asp His Ser His Val Phe Ser Phe Gly Gly Tyr
370 375 380

Pro Leu Arg Gly Ser Ser Ile Phe Gly Leu Ala Pro Gly Lys Ala Arg
385 390 395 400

Asp Arg Lys Ala Tyr Thr Val Leu Leu Tyr Gly Asn Gly Pro Gly Tyr
405 410 415

Val Leu Lys Asp Gly Ala Arg Pro Asp Val Thr Glu Ser Glu Ser Gly
420 425 430

Ser Pro Glu Tyr Arg Gln Gln Thr Gly Asn Phe Thr Arg Arg Cys Ser
435 440 445

Gln Gln Cys Pro Trp Thr Lys Arg His Thr Gln Ala Arg Thr Trp Arg
450 455 460

Cys Ser Arg Ala Ala Arg Arg Arg Thr Trp Phe Thr Ala Cys Arg Ser
465 470 475 480

Arg Pro Ser

<210> 36
 <211> 519
 <212> PRT
 <213> Unknown

<220>
 <223> full length SEAP protein

<400> 36

Met Leu Leu Leu Leu Leu Leu Leu Gly Leu Arg Leu Gln Leu Ser Leu
 1 5 10 15

Gly Ile Ile Pro Val Glu Glu Glu Asn Pro Asp Phe Trp Asn Arg Glu
 20 25 30

Ala Ala Glu Ala Leu Gly Ala Ala Lys Lys Leu Gln Pro Ala Gln Thr
 35 40 45

Ala Ala Lys Asn Leu Ile Ile Phe Leu Gly Asp Gly Met Gly Val Ser
 50 55 60

Thr Val Thr Ala Ala Arg Ile Leu Lys Gly Gln Lys Lys Asp Lys Leu
 65 70 75 80

Gly Pro Glu Ile Pro Leu Ala Met Asp Arg Phe Pro Tyr Val Ala Leu
 85 90 95

Ser Lys Thr Tyr Asn Val Asp Lys His Val Pro Asp Ser Gly Ala Thr
 100 105 110

Ala Thr Ala Tyr Leu Cys Gly Val Lys Gly Asn Phe Gln Thr Ile Gly
 115 120 125

Leu Ser Ala Ala Ala Arg Phe Asn Gln Cys Asn Thr Thr Arg Gly Asn
 130 135 140

Glu Val Ile Ser Val Met Asn Arg Ala Lys Lys Ala Gly Lys Ser Val

pctgb2018052387-seql.txt

145 150 155 160
 Gly Val Val Thr Thr Thr Arg Val Gln His Ala Ser Pro Ala Gly Thr
 165 170 175
 Tyr Ala His Thr Val Asn Arg Asn Trp Tyr Ser Asp Ala Asp Val Pro
 180 185 190
 Ala Ser Ala Arg Gln Glu Gly Cys Gln Asp Ile Ala Thr Gln Leu Ile
 195 200 205
 Ser Asn Met Asp Ile Asp Val Ile Leu Gly Gly Gly Arg Lys Tyr Met
 210 215 220
 Phe Arg Met Gly Thr Pro Asp Pro Glu Tyr Pro Asp Asp Tyr Ser Gln
 225 230 235 240
 Gly Gly Thr Arg Leu Asp Gly Lys Asn Leu Val Gln Glu Trp Leu Ala
 245 250 255
 Lys Arg Gln Gly Ala Arg Tyr Val Trp Asn Arg Thr Glu Leu Met Gln
 260 265 270
 Ala Ser Leu Asp Pro Ser Val Thr His Leu Met Gly Leu Phe Glu Pro
 275 280 285
 Gly Asp Met Lys Tyr Glu Ile His Arg Asp Ser Thr Leu Asp Pro Ser
 290 295 300
 Leu Met Glu Met Thr Glu Ala Ala Leu Arg Leu Leu Ser Arg Asn Pro
 305 310 315 320
 Arg Gly Phe Phe Leu Phe Val Glu Gly Gly Arg Ile Asp His Gly His
 325 330 335
 His Glu Ser Arg Ala Tyr Arg Ala Leu Thr Glu Thr Ile Met Phe Asp
 340 345 350
 Asp Ala Ile Glu Arg Ala Gly Gln Leu Thr Ser Glu Glu Asp Thr Leu

355

360

365

Ser Leu Val Thr Ala Asp His Ser His Val Phe Ser Phe Gly Gly Tyr
 370 375 380

Pro Leu Arg Gly Ser Ser Ile Phe Gly Leu Ala Pro Gly Lys Ala Arg
 385 390 395 400

Asp Arg Lys Ala Tyr Thr Val Leu Leu Tyr Gly Asn Gly Pro Gly Tyr
 405 410 415

Val Leu Lys Asp Gly Ala Arg Pro Asp Val Thr Glu Ser Glu Ser Gly
 420 425 430

Ser Pro Glu Tyr Arg Gln Gln Ser Ala Val Pro Leu Asp Glu Glu Thr
 435 440 445

His Ala Gly Glu Asp Val Ala Val Phe Ala Arg Gly Pro Gln Ala His
 450 455 460

Leu Val His Gly Val Gln Glu Gln Thr Phe Ile Ala His Val Met Ala
 465 470 475 480

Phe Ala Ala Cys Leu Glu Pro Tyr Thr Ala Cys Asp Leu Ala Pro Pro
 485 490 495

Ala Gly Thr Thr Asp Ala Ala His Pro Gly Tyr Ser Arg Val Gly Ala
 500 505 510

Ala Gly Arg Phe Glu Gln Thr
 515

<210> 37

<211> 26

<212> RNA

<213> Artificial Sequence

<220>

<223> regulatable intron

<400> 37

cagacgggca acuuuacacg acgcug

26

<210> 38

<211> 32

<212> DNA

<213> Artificial Sequence

<220>

<223> intron inserted in luciferase coding sequence

<400> 38

ccgcagacgg gcaactttac acgacgctgc ag

32

<210> 39

<211> 501

<212> PRT

<213> Artificial Sequence

<220>

<223> truncated luciferase protein

<400> 39

Met	Glu	Asp	Ala	Lys	Asn	Ile	Lys	Lys	Gly	Pro	Ala	Pro	Phe	Tyr	Pro
1				5					10					15	

Leu	Glu	Asp	Gly	Thr	Ala	Gly	Glu	Gln	Leu	His	Lys	Ala	Met	Lys	Arg
			20					25					30		

Tyr	Ala	Leu	Val	Pro	Gly	Thr	Ile	Ala	Phe	Thr	Asp	Ala	His	Ile	Glu
		35					40					45			

Val	Asp	Ile	Thr	Tyr	Ala	Glu	Tyr	Phe	Glu	Met	Ser	Val	Arg	Leu	Ala
	50					55					60				

Glu	Ala	Met	Lys	Arg	Tyr	Gly	Leu	Asn	Thr	Asn	His	Arg	Ile	Val	Val
65						70				75				80	

Cys	Ser	Glu	Asn	Ser	Leu	Gln	Phe	Phe	Met	Pro	Val	Leu	Gly	Ala	Leu
				85					90					95	

Phe	Ile	Gly	Val	Ala	Val	Ala	Pro	Ala	Asn	Asp	Ile	Tyr	Asn	Glu	Arg
			100					105					110		

pctgb2018052387-seq1.txt

Glu Leu Leu Asn Ser Met Gly Ile Ser Gln Pro Thr Val Val Phe Val
115 120 125

Ser Lys Lys Gly Leu Gln Lys Ile Leu Asn Val Gln Lys Lys Leu Pro
130 135 140

Ile Ile Gln Lys Ile Ile Ile Met Asp Ser Lys Thr Asp Tyr Gln Gly
145 150 155 160

Phe Gln Ser Met Tyr Thr Phe Val Thr Ser His Leu Pro Pro Gly Phe
165 170 175

Asn Glu Tyr Asp Phe Val Pro Glu Ser Phe Asp Arg Asp Lys Thr Ile
180 185 190

Ala Leu Ile Met Asn Ser Ser Gly Ser Thr Gly Leu Pro Lys Gly Val
195 200 205

Ala Leu Pro His Arg Thr Ala Cys Val Arg Phe Ser His Ala Arg Asp
210 215 220

Pro Ile Phe Gly Asn Gln Ile Ile Pro Asp Thr Ala Ile Leu Ser Val
225 230 235 240

Val Pro Phe His His Gly Phe Gly Met Phe Thr Thr Leu Gly Tyr Leu
245 250 255

Ile Cys Gly Phe Arg Val Val Leu Met Tyr Arg Phe Glu Glu Glu Leu
260 265 270

Phe Leu Arg Ser Leu Gln Asp Tyr Lys Ile Gln Ser Ala Leu Leu Val
275 280 285

Pro Thr Leu Phe Ser Phe Phe Ala Lys Ser Thr Leu Ile Asp Lys Tyr
290 295 300

Asp Leu Ser Asn Leu His Glu Ile Ala Ser Gly Gly Ala Pro Leu Ser
305 310 315 320

pctgb2018052387-seql.txt

Lys Glu Val Gly Glu Ala Val Ala Lys Arg Phe His Leu Pro Gly Ile
325 330 335

Arg Gln Gly Tyr Gly Leu Thr Glu Thr Thr Ser Ala Ile Leu Ile Thr
340 345 350

Pro Glu Gly Asp Asp Lys Pro Gly Ala Val Gly Lys Val Val Pro Phe
355 360 365

Phe Glu Ala Lys Val Val Asp Leu Asp Thr Gly Lys Thr Leu Gly Val
370 375 380

Asn Gln Arg Gly Glu Leu Cys Val Arg Gly Pro Met Ile Met Ser Gly
385 390 395 400

Tyr Val Asn Asn Pro Glu Ala Thr Asn Ala Leu Ile Asp Lys Asp Gly
405 410 415

Trp Leu His Ser Gly Asp Ile Ala Tyr Trp Asp Glu Asp Glu His Phe
420 425 430

Phe Ile Val Asp Arg Leu Lys Ser Leu Ile Lys Tyr Lys Gly Tyr Gln
435 440 445

Val Ala Pro Ala Glu Leu Glu Ser Ile Leu Leu Gln His Pro Asn Ile
450 455 460

Phe Asp Ala Gly Val Ala Gly Leu Pro Asp Asp Asp Ala Gly Glu Leu
465 470 475 480

Pro Ala Ala Asp Gly Gln Leu Tyr Thr Thr Leu Gln Thr Ser Cys Trp
485 490 495

Asn Thr Val Lys Pro
500

<210> 40
<211> 550
<212> PRT

<213> Unknown

<220>

<223> full length luciferase protein

<400> 40

Met Glu Asp Ala Lys Asn Ile Lys Lys Gly Pro Ala Pro Phe Tyr Pro
 1 5 10 15

Leu Glu Asp Gly Thr Ala Gly Glu Gln Leu His Lys Ala Met Lys Arg
 20 25 30

Tyr Ala Leu Val Pro Gly Thr Ile Ala Phe Thr Asp Ala His Ile Glu
 35 40 45

Val Asp Ile Thr Tyr Ala Glu Tyr Phe Glu Met Ser Val Arg Leu Ala
 50 55 60

Glu Ala Met Lys Arg Tyr Gly Leu Asn Thr Asn His Arg Ile Val Val
 65 70 75 80

Cys Ser Glu Asn Ser Leu Gln Phe Phe Met Pro Val Leu Gly Ala Leu
 85 90 95

Phe Ile Gly Val Ala Val Ala Pro Ala Asn Asp Ile Tyr Asn Glu Arg
 100 105 110

Glu Leu Leu Asn Ser Met Gly Ile Ser Gln Pro Thr Val Val Phe Val
 115 120 125

Ser Lys Lys Gly Leu Gln Lys Ile Leu Asn Val Gln Lys Lys Leu Pro
 130 135 140

Ile Ile Gln Lys Ile Ile Ile Met Asp Ser Lys Thr Asp Tyr Gln Gly
 145 150 155 160

Phe Gln Ser Met Tyr Thr Phe Val Thr Ser His Leu Pro Pro Gly Phe
 165 170 175

Asn Glu Tyr Asp Phe Val Pro Glu Ser Phe Asp Arg Asp Lys Thr Ile

pctgb2018052387-seq1.txt

180

185

190

Ala Leu Ile Met Asn Ser Ser Gly Ser Thr Gly Leu Pro Lys Gly Val
 195 200 205

Ala Leu Pro His Arg Thr Ala Cys Val Arg Phe Ser His Ala Arg Asp
 210 215 220

Pro Ile Phe Gly Asn Gln Ile Ile Pro Asp Thr Ala Ile Leu Ser Val
 225 230 235 240

Val Pro Phe His His Gly Phe Gly Met Phe Thr Thr Leu Gly Tyr Leu
 245 250 255

Ile Cys Gly Phe Arg Val Val Leu Met Tyr Arg Phe Glu Glu Glu Leu
 260 265 270

Phe Leu Arg Ser Leu Gln Asp Tyr Lys Ile Gln Ser Ala Leu Leu Val
 275 280 285

Pro Thr Leu Phe Ser Phe Phe Ala Lys Ser Thr Leu Ile Asp Lys Tyr
 290 295 300

Asp Leu Ser Asn Leu His Glu Ile Ala Ser Gly Gly Ala Pro Leu Ser
 305 310 315 320

Lys Glu Val Gly Glu Ala Val Ala Lys Arg Phe His Leu Pro Gly Ile
 325 330 335

Arg Gln Gly Tyr Gly Leu Thr Glu Thr Thr Ser Ala Ile Leu Ile Thr
 340 345 350

Pro Glu Gly Asp Asp Lys Pro Gly Ala Val Gly Lys Val Val Pro Phe
 355 360 365

Phe Glu Ala Lys Val Val Asp Leu Asp Thr Gly Lys Thr Leu Gly Val
 370 375 380

Asn Gln Arg Gly Glu Leu Cys Val Arg Gly Pro Met Ile Met Ser Gly

pctgb2018052387-seq1.txt

385 390 395 400

Tyr Val Asn Asn Pro Glu Ala Thr Asn Ala Leu Ile Asp Lys Asp Gly
405 410 415

Trp Leu His Ser Gly Asp Ile Ala Tyr Trp Asp Glu Asp Glu His Phe
420 425 430

Phe Ile Val Asp Arg Leu Lys Ser Leu Ile Lys Tyr Lys Gly Tyr Gln
435 440 445

Val Ala Pro Ala Glu Leu Glu Ser Ile Leu Leu Gln His Pro Asn Ile
450 455 460

Phe Asp Ala Gly Val Ala Gly Leu Pro Asp Asp Asp Ala Gly Glu Leu
465 470 475 480

Pro Ala Ala Asp Val Val Leu Glu His Gly Lys Thr Met Thr Glu Lys
485 490 495

Glu Ile Val Asp Tyr Val Ala Ser Gln Val Thr Thr Ala Lys Lys Leu
500 505 510

Arg Gly Gly Val Val Phe Val Asp Glu Val Pro Lys Gly Leu Thr Gly
515 520 525

Lys Leu Asp Ala Arg Lys Ile Arg Glu Ile Leu Ile Lys Ala Lys Lys
530 535 540

Gly Gly Lys Ile Ala Val
545 550

$\langle 210 \rangle$	41
$\langle 211 \rangle$	30
$\langle 212 \rangle$	DNA
$\langle 213 \rangle$	Artificial Sequence

<220>
<223> INTC9FP intron

<400> 41

acgtcagacg ggcaacttta cagcagctg 30

<210> 42
 <211> 30
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> INTC9RP intron sequence

<400> 42
 acgtcagcgt cgtgtaaagt tgcccgtctg 30

<210> 43
 <211> 32
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Regulatable intron insered in caspase 9 coding sequence

<400> 43
 cagcagacgg gcaactttac acgacgctgc tg 32

<210> 44
 <211> 8008
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> SYNP-CASP9-INT vector with intron construct

<400> 44
 gacggatcgg gagatctccc gatcccctat ggtgcactct cagtacaatc tgctctgatg 60
 ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggctgct gagtagtgcg 120
 cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc 180
 ttagggttag gcgttttgcg ctgcttcgcg agtacattta tattggctca tgtccaatat 240
 gaccgccatg ttgacattga ttattgacta gttattaata gtaatcaatt acgggggtcat 300
 tagttcatag cccatatatg gagttccgcg ttacataact tacggtaaatt ggcccgcctg 360
 gctgaccgcc caacgacccc cgcccattga cgtcaataat gacgtatggt cccatagtaa 420
 cgccaatagg gactttccat tgacgtcaat ggggtggagta ttacggtaa actgcccact 480

pctgb2018052387-seq1.txt

tggcagtaca tcaagtgtat catatgccaa gtccgcccc tattgacgtc aatgacggta	540
aatggcccgct ctggcattat gcccagtaca tgaccttacg ggactttcct acttggcagt	600
acatctacgt attagtcacg gctattacca tggatgatgcg gttttggcag tacaccaatg	660
ggcgtggata gcggtttgac tcacggggat ttccaagtct ccacccatt gacgtcaatg	720
ggagtttggt ttggcaccaa aatcaacggg actttccaaa atgtcgtaat aaccccgccc	780
cgttgacgca aatgggagggt aggcgtgtac ggtgggagggt ctatataagc agagctcggt	840
tagtgaaccg tcagatcctc actctcttcc gcatcgctgt ctgcgagggc cagctgttgg	900
gctcgcggtt gaggacaaac tcttcgcggt ctttcagta ctcttgatc ggaaaccgt	960
cggcctccga acggtactcc gccaccgagg gacctgagcg agtccgcatc gaccggatcg	1020
gaaaacctct cgagaaaggc gtctaaccag tcacagtcgc aaggtaggct gagcaccgtg	1080
gcgggaggca gcgggtggcg gtcgggggtt tttctggcgg aggtgctgct gatgatgtaa	1140
ttaaagtagg cggctctgag acggcggatg gtcgaggtga ggtgtgggtt tagtgaaccg	1200
tcagatcctc actctcttcc gcatcgctgt ctgcgagggc cagctgtcag gcttgagatc	1260
cagctgttgg ggtgagtact ccctctcaaa agcgggcatt acttctgcgc taagattgtc	1320
agtttccaaa aacgaggagg atttgatatt cacctggccc gatctggcca tacacttgag	1380
tgacaatgac atccactttg ctttctctc cacagggtgc cactcccagg tccaagttta	1440
aactttaata cgactcacta taggggcccgc caccaagctt ggtacatgga cgaagcggat	1500
cggcggctcc tgcggcgggtg ccggctgcgg ctggtggaag agctgcaggt ggaccagctc	1560
tgggacgccc tgctgagccg cgagctgttc agggccata tgatcgagga catccagcgg	1620
gcaggctctg gatctcggcg ggatcaggcc aggcagctga tcatagatct ggagactcga	1680
gggagtcagg ctcttccttt gttcatctcc tgcttagagg acacaggcca ggacatgctg	1740
gcttcgtttc tgcgaaactaa caggcaagca gcaaagttgt cgaagccaac cctagaaaac	1800
cttaccaccag tgggtgctcag accagagatt cgcaaaccag aggttctcag accggaaaca	1860
cccagaccag tggacattgg ttctggagga ttcggtgatg tcggtgctct tgagagtttg	1920
aggggaaatg cagatttggc ttacatcctg agcatggagc cctgtggcca ctgcctcatt	1980
atcaacaatg tgaacttctg ccgtgagtcg gggctccgca cccgactgg ctccaacatc	2040

pctgb2018052387-seq1.txt

gactgtgaga agttgcggcg tcgcttctcc tcgctgcatt tcatggtgga ggtgaagggc	2100
gacctgactg ccaagaaaat ggtgctggct ttgctggagc tggcgcagca ggaccacggt	2160
gctctggact gctgcgtggt ggtcattctc tctcacggct gtcaggccag ccacctgcag	2220
ttcccagggg ctgtctacgg cacagatgga tgccctgtgt cggtcgagaa gattgtgaac	2280
atcttcaatg ggaccagctg cccagcctg ggagggaagc ccaagctctt tttcatccag	2340
gcctgtggtg gggagcagaa agaccatggg tttgaggtgg cctccacttc ccctgaagac	2400
gagtcccctg gcagtaacct cgagccagat gccaccccg tccaggaagg tttgaggacc	2460
ttcgaccagc tggacgcat atctagtttg cccacacca gtgacatctt tgtgtcctac	2520
tctactttcc caggttttgt ttcctggagg gacccaaga gtggctcctg gtcagacggg	2580
caactttaca cgacgtgga gaccctggac gacatcttg agcagtgggc tcactctgaa	2640
gacctgcagt ccctcctgct tagggctgct aatgctgttt cggtgaaagg gatttataaa	2700
cagatgcctg gttgctttaa tttcctccga aaaaaacttt tctttaaaac atcaggggggt	2760
ggaggctctg tgagcaaggg cgaggagctg ttcaccgggg tggtgcccac cctggctgag	2820
ctggacggcg acgtaaacgg ccacaagttc agcgtgtccg gcgagggcga gggcgatgcc	2880
acctacggca agctgacct gaagtccatc tgcaccaccg gcaagctgcc cgtgccctgg	2940
cccaccctcg tgaccacct gacctacggc gtgcagtgtc tcagccgcta ccccgaccac	3000
atgaagaagc acgacttctt caagtccgcc atgcccgaag gctacgtcca ggagcgcacc	3060
atcttcttca aggacgacgg caactacaag acccgcgccg aggtgaagtt cgagggcgac	3120
accctggtga accgcatcga gctgaagggc atcgacttca aggaggacgg caacatcctg	3180
gggcacaagc tggagtacaa ctacaacagc cacaacgtct atatcatggc cgacaagcag	3240
aagaacggca tcaaggctaa cttcaaggtt cgccacaaca tcgaggacgg cagcgtgcag	3300
ctgcccgaac actaccagca gaacaccccc atcggcgacg gccccgtgct gctgcccgaac	3360
aaccactacc tgagcaccca gtccgccctg agcaaagacc ccaacgagaa gcgcgatcac	3420
atggtcctgc tggagtctgt gaccgccgcc gggatcactc tcggcatgga cgagctgtac	3480
aagtaaactc gagtctagag cggccgccga attcggggcc gtttaaaccc gctgatcagc	3540
ctcgactgtg ctttctagtt gccagccatc tgttgtttgc ccctccccg tgccttcctt	3600

pctgb2018052387-seq1.txt

gaccctggaa ggtgccactc ccactgtcct ttcctaataa aatgaggaaa ttgcatcgca	3660
ttgtctgagt aggtgtcatt ctattctggg ggggtggggtg gggcaggaca gcaaggggga	3720
ggattgggaa gacaatagca ggcatgctgg ggatgcggtg ggctctatgg cttctgaggc	3780
ggaaagaacc agctggggct ctagggggta tccccacgcg ccctgtagcg gcgcattaag	3840
cgcggcgggt gtggtggtta cgcgcagcgt gaccgctaca cttgccagcg ccctagcgcc	3900
cgctcctttc gctttcttcc cttcctttct cgccacgttc gcaggctttc cccgtcaagc	3960
tctaaatcgg gggctccctt tagggttccg atttagtgct ttacggcacc tcgaccccaa	4020
aaaacttgat tagggtgatg gttcacgtag tgggccatcg ccctgataga cggtttttcg	4080
ccctttgacg ttggagtcca cgttctttta tagtggactc ttgttccaaa ctggaacaac	4140
actcaaccct atctcggctc attcttttga tttataaggg attttgccga tttcggccta	4200
ttggttaaaa aatgagctga ttttaacaaa atttaacgcg aattaattct gtggaatgtg	4260
tgtcagttag ggtgtggaaa gtccccaggc tccccagcag gcagaagtat gcaaagcatg	4320
catctcaatt agtcagcaac cagggtgtgga aagtccccag gctccccagc aggcagaagt	4380
atgcaaagca tgcattctaa ttagtcagca accatagtcc cgcccctaac tccgcccac	4440
ccgccccta ctccgcccag ttccgcccac tctccgcccc atggctgact aatttttttt	4500
atztatgcag aggccgaggc cgcctctgcc tctgagctat tccagaagta gtgaggaggc	4560
ttttttggag gcctaggctt ttgcaaaaag ctctcgggag cttgtatatc cattttcgga	4620
tctgatcagc acgtgatgaa aaagcctgaa ctcaccgcga cgtctgtcga gaagtttctg	4680
atcgaaaagt tcgacagcgt ctccgacctg atgcagctct cggagggcga agaattctct	4740
gctttcagct tcgatgtagg agggcgtgga tatgtcctgc gggtaaatag ctgcgccgat	4800
ggtttctaca aagatcggtt tgtttatcgg cactttgcat cggccgcgct cccgattccg	4860
gaagtgcttg acattgggga attcagcgag agcctgacct attgcatctc ccgccgtgca	4920
cagggtgtca cgttgcaaga cctgcctgaa accgaactgc ccgctgttct gcagccggtc	4980
gcggaggcca tggatgcgat cgctgcggcc gatcttagcc agacgagcgg gttcggccca	5040
ttcggaccgc aaggaatcgg tcaatacact acatggcgtg atttcatatg cgcgattgct	5100
gatcccatg tgtatcactg gcaaactgtg atggacgaca ccgtcagtgc gtccgtcgcg	5160

pctgb2018052387-seq1.txt

caggctctcg atgagctgat gctttgggcc gaggactgcc ccgaagtccg gcacctcgtg	5220
cacgcgggatt tcggctccaa caatgtcctg acggacaatg gccgcataac agcggtcatt	5280
gactggagcg aggcgatgtt cggggattcc caatacgagg tcgccaacat cttcttctgg	5340
aggccgtggt tggcttgtat ggagcagcag acgcgctact tcgagcggag gcatccggag	5400
cttgaggat cgccgcggct cggggcgtat atgctccgca ttggtcttga ccaactctat	5460
cagagcttgg ttgacggcaa tttcgatgat gcagcttggg cgcagggtcg atgcgacgca	5520
atcgtccgat ccggagccgg gactgtcggg cgtacacaaa tcgcccgcag aagcgcggcc	5580
gtctggaccg atggctgtgt agaagtactc gccgatagtg gaaaccgacg ccccagcact	5640
cgtccgaggg caaaggaata gcacgtgcta cgagatttcg attccaccgc cgccttctat	5700
gaaaggttgg gcttcggaat cgttttccgg gacgctggct ggatgatcct ccagcgcggg	5760
gatctcatgc tggagttctt cgcccacccc aacttgttta ttgcagctta taatggttac	5820
aaataaagca atagcatcac aaatttcaca aataaagcat ttttttcact gcattctagt	5880
tgtggtttgt ccaaactcat caatgtatct tatcatgtct gtataaccgtc gacctctagc	5940
tagagcttgg cgtaatcatg gtcatactgt tttcctgtgt gaaattgtta tccgctcaca	6000
attccacaca acatacgagc cggaagcata aagtgtaaag cctgggggtgc ctaatgagtg	6060
agctaactca cattaattgc gttgcgctca ctgcccgtt tccagtcggg aaacctgtcg	6120
tgccagctgc attaatgaat cggccaacgc gcggggagag gcggtttgcg tattggg'gcg	6180
tcttccgctt cctcgctcac tgactcgctg cgctcggctg ttcggctgcg gcgagcggta	6240
tcagctcact caaaggcggc aatacggtta tccacagaat caggggataa cgcaggaaaag	6300
aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta aaaaggccgc gttgctggcg	6360
tttttccata ggctccgccc ccctgacgag catcacaaaa atcgacgctc aagtcagagg	6420
tggcgaaacc cgacaggact ataaagatac caggcgtttc cccctggaag ctccctcgtg	6480
cgctctcctg ttccgaccct gccgcttacc ggatacctgt ccgcctttct cccttcggga	6540
agcgtggcgc tttctcatag ctcacgctgt aggtatctca gttcgggtgta ggtcgttcgc	6600
tccaagctgg gctgtgtgca cgaaccccc gttcagcccc accgctgcgc cttatccggt	6660
aactatcgtc ttgagtcmaa cccggtgaaga cacgacttat cgccactggc agcagccact	6720

pctgb2018052387-seq1.txt

ggtaacagga ttagcagagc gaggtatgta ggcggtgcta cagagttctt gaagtgggtg	6780
cctaactacg gctacactag aagaacagta tttggatatct gcgctctgct gaagccagtt	6840
accttcggaa aaagagttgg tagctcttga tccggcaaac aaaccaccgc tggtagcgg	6900
ggtttttttg tttgcaagca gcagattacg cgcagaaaaa aaggatctca agaagatcct	6960
ttgatctttt ctacggggtc tgagcgcgga acccctatct gtttatcttt ctaaatacat	7020
tcaaatatgt atccgctcat gaattaattc ttagaaaaac tcatcgagca tcaaataaaa	7080
ctgcaattta ttcatactag gattatcaat accatatctt tgaaaaagcc gtttctgtaa	7140
tgaaggagaa aactcaccga ggcagttcca taggatggca agatcctggg atcgggtctgc	7200
gattccgact cgtccaacat caatacaacc tattaatttc ccctcgtcaa aaataagggt	7260
atcaagttag aaatcaccat gagtgacgac tgaatccggg gagaatggca aaagtttatg	7320
catttctttc cagacttggt caacaggcca gccattacgc tcgtcatcaa aatcactcgc	7380
atcaacaaaa ccgttattca ttcgtgattg cgcctgagcg agacgaaata cgcgatcgtc	7440
gttaaaagga caattacaaa caggaatcga atgcaaccgg cgcaggaaca ctgccagcgc	7500
atcaacaata ttttcacctg aatcaggata ttcttctaata acctggaatg ctgttttccc	7560
agggatcgcg gtggtgagta accatgcatc atcaggagta cggataaaat gcttgatggg	7620
cggaagaggc ataaattccg tcagccagtt tagtctgacc atctcatctg taacatcatt	7680
ggcaacgcta cctttgccat gtttcagaaa caactctggc gcatcgggct tcccatacaa	7740
tcgatagatt gtcgcacctg attgcccagc attatcgcga gcccatctat acccatataa	7800
atcagcatcc atgttggaat ttaatcgcgg cctagagcaa gacgtttccc gttgaatatg	7860
gctcataaca ccccttgat tactgtttat gtaagcagac agttttattg ttcatgacca	7920
aaatccctta acgtgagttt tcgttcact gagcgtcaga ccccgtagaa atccgcgcac	7980
atttccccga aaagtgccac ctgacgtc	8008

<210> 45

<211> 371

<212> PRT

<213> Artificial Sequence

<220>

<223> truncated caspase 9

pctgb2018052387-seql.txt

<400> 45

Met Asp Glu Ala Asp Arg Arg Leu Leu Arg Arg Cys Arg Leu Arg Leu
1 5 10 15

Val Glu Glu Leu Gln Val Asp Gln Leu Trp Asp Ala Leu Leu Ser Arg
20 25 30

Glu Leu Phe Arg Pro His Met Ile Glu Asp Ile Gln Arg Ala Gly Ser
35 40 45

Gly Ser Arg Arg Asp Gln Ala Arg Gln Leu Ile Ile Asp Leu Glu Thr
50 55 60

Arg Gly Ser Gln Ala Leu Pro Leu Phe Ile Ser Cys Leu Glu Asp Thr
65 70 75 80

Gly Gln Asp Met Leu Ala Ser Phe Leu Arg Thr Asn Arg Gln Ala Ala
85 90 95

Lys Leu Ser Lys Pro Thr Leu Glu Asn Leu Thr Pro Val Val Leu Arg
100 105 110

Pro Glu Ile Arg Lys Pro Glu Val Leu Arg Pro Glu Thr Pro Arg Pro
115 120 125

Val Asp Ile Gly Ser Gly Gly Phe Gly Asp Val Gly Ala Leu Glu Ser
130 135 140

Leu Arg Gly Asn Ala Asp Leu Ala Tyr Ile Leu Ser Met Glu Pro Cys
145 150 155 160

Gly His Cys Leu Ile Ile Asn Asn Val Asn Phe Cys Arg Glu Ser Gly
165 170 175

Leu Arg Thr Arg Thr Gly Ser Asn Ile Asp Cys Glu Lys Leu Arg Arg
180 185 190

Arg Phe Ser Ser Leu His Phe Met Val Glu Val Lys Gly Asp Leu Thr

195

200

205

Ala Lys Lys Met Val Leu Ala Leu Leu Glu Leu Ala Gln Gln Asp His
 210 215 220

Gly Ala Leu Asp Cys Cys Val Val Val Ile Leu Ser His Gly Cys Gln
 225 230 235 240

Ala Ser His Leu Gln Phe Pro Gly Ala Val Tyr Gly Thr Asp Gly Cys
 245 250 255

Pro Val Ser Val Glu Lys Ile Val Asn Ile Phe Asn Gly Thr Ser Cys
 260 265 270

Pro Ser Leu Gly Gly Lys Pro Lys Leu Phe Phe Ile Gln Ala Cys Gly
 275 280 285

Gly Glu Gln Lys Asp His Gly Phe Glu Val Ala Ser Thr Ser Pro Glu
 290 295 300

Asp Glu Ser Pro Gly Ser Asn Pro Glu Pro Asp Ala Thr Pro Phe Gln
 305 310 315 320

Glu Gly Leu Arg Thr Phe Asp Gln Leu Asp Ala Ile Ser Ser Leu Pro
 325 330 335

Thr Pro Ser Asp Ile Phe Val Ser Tyr Ser Thr Phe Pro Gly Phe Val
 340 345 350

Ser Trp Arg Asp Pro Lys Ser Gly Ser Trp Ser Ala Gly Asp Pro Gly
 355 360 365

Arg His Leu
 370

<210> 46

<211> 430

<212> PRT

<213> Artificial Sequence

<220>

<223> full length caspase 9

<400> 46

Met Asp Glu Ala Asp Arg Arg Leu Leu Arg Arg Cys Arg Leu Arg Leu
 1 5 10 15

Val Glu Glu Leu Gln Val Asp Gln Leu Trp Asp Ala Leu Leu Ser Arg
 20 25 30

Glu Leu Phe Arg Pro His Met Ile Glu Asp Ile Gln Arg Ala Gly Ser
 35 40 45

Gly Ser Arg Arg Asp Gln Ala Arg Gln Leu Ile Ile Asp Leu Glu Thr
 50 55 60

Arg Gly Ser Gln Ala Leu Pro Leu Phe Ile Ser Cys Leu Glu Asp Thr
 65 70 75 80

Gly Gln Asp Met Leu Ala Ser Phe Leu Arg Thr Asn Arg Gln Ala Ala
 85 90 95

Lys Leu Ser Lys Pro Thr Leu Glu Asn Leu Thr Pro Val Val Leu Arg
 100 105 110

Pro Glu Ile Arg Lys Pro Glu Val Leu Arg Pro Glu Thr Pro Arg Pro
 115 120 125

Val Asp Ile Gly Ser Gly Gly Phe Gly Asp Val Gly Ala Leu Glu Ser
 130 135 140

Leu Arg Gly Asn Ala Asp Leu Ala Tyr Ile Leu Ser Met Glu Pro Cys
 145 150 155 160

Gly His Cys Leu Ile Ile Asn Asn Val Asn Phe Cys Arg Glu Ser Gly
 165 170 175

Leu Arg Thr Arg Thr Gly Ser Asn Ile Asp Cys Glu Lys Leu Arg Arg
 180 185 190

pctgb2018052387-seq1.txt

Arg Phe Ser Ser Leu His Phe Met Val Glu Val Lys Gly Asp Leu Thr
195 200 205

Ala Lys Lys Met Val Leu Ala Leu Leu Glu Leu Ala Gln Gln Asp His
210 215 220

Gly Ala Leu Asp Cys Cys Val Val Val Ile Leu Ser His Gly Cys Gln
225 230 235 240

Ala Ser His Leu Gln Phe Pro Gly Ala Val Tyr Gly Thr Asp Gly Cys
245 250 255

Pro Val Ser Val Glu Lys Ile Val Asn Ile Phe Asn Gly Thr Ser Cys
260 265 270

Pro Ser Leu Gly Gly Lys Pro Lys Leu Phe Phe Ile Gln Ala Cys Gly
275 280 285

Gly Glu Gln Lys Asp His Gly Phe Glu Val Ala Ser Thr Ser Pro Glu
290 295 300

Asp Glu Ser Pro Gly Ser Asn Pro Glu Pro Asp Ala Thr Pro Phe Gln
305 310 315 320

Glu Gly Leu Arg Thr Phe Asp Gln Leu Asp Ala Ile Ser Ser Leu Pro
325 330 335

Thr Pro Ser Asp Ile Phe Val Ser Tyr Ser Thr Phe Pro Gly Phe Val
340 345 350

Ser Trp Arg Asp Pro Lys Ser Gly Ser Trp Ser Ala Asp Gly Gln Leu
355 360 365

Tyr Thr Thr Leu Leu Glu Thr Leu Asp Asp Ile Phe Glu Gln Trp Ala
370 375 380

His Ser Glu Asp Leu Gln Ser Leu Leu Leu Arg Val Ala Asn Ala Val
385 390 395 400

pctgb2018052387-seql.txt

Ser Val Lys Gly Ile Tyr Lys Gln Met Pro Gly Cys Phe Asn Phe Leu
405 410 415

Arg Lys Lys Leu Phe Phe Lys Thr Ser Gly Gly Gly Gly Ser
420 425 430

<210> 47
<211> 154
<212> DNA
<213> Artificial Sequence

<220>
<223> UPR inducible promoter

<400> 47
tgacgtgctg atgatgcgta gctagtagtt gacgtgctga tgatgcgtag ctagtagttg 60
acgtgctgat gatgcgtagc tagtagttga cgtgctgatg atgcgtagct agtagttgac 120
gtgctgatga tgcgtagcta gtagttgacg tgct 154

<210> 48
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Regulatable intron

<400> 48
cagctgcagc actcagacta cgtgcacctc tgctg 35

<210> 49
<211> 20
<212> RNA
<213> Artificial Sequence

<220>
<223> intron sequence inside cleavage sites

<400> 49
acgggcaacu uuacacgacg 20

<210> 50
<211> 32
<212> RNA

<213> Artificial Sequence

<220>

<223> regulatable intron sequence

<220>

<221> misc_feature

<222> (2)..(2)

<223> n is a, c, g, or u

<220>

<221> misc_feature

<222> (31)..(31)

<223> n is a, c, g, or u

<400> 50

cngcagacgg gcaacuuuac acgacgcugc ng

32

<210> 51

<211> 5

<212> PRT

<213> Artificial Sequence

<220>

<223> truncated EGFP protein - first stop codon to second stop codon

<400> 51

Ala Pro Ser Pro Pro

1

5

<210> 52

<211> 16

<212> PRT

<213> Artificial Sequence

<220>

<223> truncated EGFP protein - from second stop codon to third stop codon

<400> 52

Ala Lys Thr Pro Thr Arg Ser Ala Ile Thr Trp Ser Cys Trp Ser Ser

1

5

10

15

<210> 53

<211> 13

<212> PRT

<213> Artificial Sequence

<220>

<223> truncated EGFP protein from third stop codon to end of protein

<400> 53

Pro	Pro	Pro	Gly	Ser	Leu	Ser	Ala	Trp	Thr	Ser	Cys	Thr
1				5					10			

<210> 54

<211> 47

<212> DNA

<213> Artificial Sequence

<220>

<223> UPR-responsive CRE

<220>

<221> misc_feature

<222> (8)..(8)

<223> optional spacer sequence

<220>

<221> misc_feature

<222> (16)..(16)

<223> optional spacer sequence

<220>

<221> misc_feature

<222> (24)..(24)

<223> optional spacer sequence

<220>

<221> misc_feature

<222> (32)..(32)

<223> optional spacer sequence

<220>

<221> misc_feature

<222> (40)..(40)

<223> optional spacer sequence

<400> 54

tgacgtgntg acgtgntgac gtgntgacgt gntgacgtgn tgacgtg

47

<210> 55

<211> 59

<212> DNA

<213> Artificial Sequence

<220>

<223> UPR-responsive CRE

<220>

<221> misc_feature

<222> (10)..(10)

<223> Optional spacer sequence

<220>

<221> misc_feature

<222> (20)..(20)

<223> Optional spacer sequence

<220>

<221> misc_feature

<222> (30)..(30)

<223> Optional spacer sequence

<220>

<221> misc_feature

<222> (40)..(40)

<223> Optional spacer sequence

<220>

<221> misc_feature

<222> (50)..(50)

<223> Optional spacer sequence

<400> 55

tgacgtgctn tgacgtgctn tgacgtgctn tgacgtgctn tgacgtgctn tgacgtgct 59

<210> 56

<211> 154

<212> DNA

<213> Artificial Sequence

<220>

<223> UPR-responsive CRE

<400> 56

tgacgtgctg atgatgcgta gctagtagtt gacgtgctga tgatgcgtag ctagtagttg 60

acgtgctgat gatgcgtagc tagtagttga cgtgctgatg atgcgtagct agtagttgac 120

gtgctgatga tgcgtagcta gtagttgacg tgct 154

<210> 57

pctgb2018052387-seq1.txt

<211> 255
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> UPR-inducible promoter

<400> 57
 tgacgtgctg atgatgcgta gctagtagtt gacgtgctga tgatgcgtag ctagtagttg 60
 acgtgctgat gatgcgtagc tagtagttga cgtgctgatg atgcgtagct agtagttgac 120
 gtgctgatga tgcgtagcta gtagttgacg tgctgatgat gcgtagctag tagtgagttt 180
 agcgtagctg aggtaccgtc gacgatatcg gatccaggctc tatataagca gagctcgttt 240
 agtgaaccgt cagat 255

<210> 58
 <211> 255
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> UPR-inducible promoter

<400> 58
 tgacgtgctg atgatgcgta gctagtagtt gacgtgctga tgatgcgtag ctagtagttg 60
 acgtgctgat gatgcgtagc tagtagttga cgtgctgatg atgcgtagct agtagttgac 120
 gtgctgatga tgcgtagcta gtagttgacg tgctgatgat gcgtagctag tagtgagttt 180
 agcgtagctg aggtaccgtc gacgatatcg gacgcttcgc atattaaggt gacgcgtgtg 240
 gcctcgaaca ccgag 255

<210> 59
 <211> 2269
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Construct ATF06-MP-001

<400> 59
 gaagttccta ttccgaagtt cctattctct agaaagtata ggaacttctg acgtgctgat 60
 gatgcgtagc tagtagttga cgtgctgatg atgcgtagct agtagttgac gtgctgatga 120

pctgb2018052387-seq1.txt

tgcgtagcta gtagttgacg tgctgatgat gcgtagctag tagttgacgt gctgatgatg	180
cgtagctagt agttgacgtg ctgatgatgc gtagctagta gtgcagttag cgtagctgag	240
gtaccgtcga cgatatcgga tccttcgcat attaaggtga cgcgtgtggc ctcgaacacc	300
gagcgaccct gcagcgaccc gcttaaattgg aagatgccaa aaacattaag aagggccag	360
cgccattcta cccactcgaa gacgggaccg ccggcgagca gctgcacaaa gccatgaagc	420
gctacgccct ggtgcccggc accatcgctt ttaccgacgc acatatcgag gtggacatta	480
cctacgccga gtacttcgag atgagcgctt ggctggcaga agctatgaag cgctatgggc	540
tgaatacaaa ccatcggatc gtggtgtgca gcgagaatag cttgcagttc ttcatgcccg	600
tggtgggtgc cctgttcatc ggtgtggctg tggccccagc taacgacatc tacaacgagc	660
gcgagctgct gaacagcatg ggcatcagcc agcccaccgt cgtattcgtg agcaagaaag	720
ggctgcaaaa gatcctcaac gtgcaaaaga agctaccgat catacaaaag atcatcatca	780
tggatagcaa gaccgactac cagggttcc aaagcatgta caccttcgtg acttcccatt	840
tgccaccggg cttcaacgag tacgacttcg tgcccagagag cttcgaccgg gacaaaacca	900
tcgccctgat catgaacagt agtggcagta ccggattgcc caaggcgta gccctaccgc	960
accgcaccgc ttgtgtccga ttcagtcatg cccgcgaccc catcttcggc aaccagatca	1020
tccccgacac cgctatcctc agcgtggtgc catttcacca cggcttcggc atgttcacca	1080
cgctgggcta cttgatctgc ggctttcggg tcgtgctcat gtaccgcttc gaggaggagc	1140
tattcttgcg cagcttgcaa gactataaga ttcaatctgc cctgctggtg cccacactat	1200
ttagcttctt cgctaagagc actctcatcg acaagtacga cctaagcaac ttgcacgaga	1260
tcgccagcgg cggggcgccg ctcagcaagg aggtaggtga ggccgtggcc aaacgcttcc	1320
acctaccagg catccgccag ggctacggcc tgacagaaac aaccagcgcc attctgatca	1380
ccccgaagg ggacgacaag cctggcgag taggcaagg ggtgcccttc ttcgaggcta	1440
aggtggtgga cttggacacc ggtaagacac tgggtgtgaa ccagcgggc gagctgtgcg	1500
tccgtggccc catgatcatg agcggctacg ttaacaaccc cgaggctaca aacgctctca	1560
tcgacaagga cggctggctg cacagcggcg acatcgcccta ctgggacgag gacgagcact	1620
tcttcatcgt ggaccggctg aagagcctga tcaaatacaa gggctaccag gtagccccag	1680

pctgb2018052387-seq1.txt

ccgaactgga gagcatcctg ctgcaacacc ccaacatctt cgacgccggg gtcgccggcc	1740
tgcccgacga cgatgccggc gagctgcccg ccgcagtcgt cgtgctggaa cacggtaaaa	1800
ccatgaccga gaaggagatc gtggactatg tggccagcca ggttacaacc gccaagaagc	1860
tgcgcggtgg tgttgtgttc gtggacgagg tgcctaaagg actgaccggc aagttggacg	1920
cccgcaagat ccgcgagatt ctcattaagg ccaagaaggg cggcaagatc gccgtgtaat	1980
gaaagcttgg tctctacgag taatagacgc ccagttgaat tccttcgagc agacatgata	2040
agatacattg atgagtttgg acaaaccaca actagaatgc agtgaaaaaa atgctttatt	2100
tgtgaaatth gtgatgctat tgctttatth gtaaccatta taagctgcaa taaacaagtt	2160
aacaacaaca attgcattca ttttatgtth caggttcagg gggagggtgtg ggaggththt	2220
taaagcaagt aaaacctcta caaatgtggt aaaatcgata aggatccgt	2269

<210> 60
 <211> 2425
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Costruct ATF06-MP-002

<400> 60	
gcgacgtaat acgactcact atagggcgaa ttggcggaag gccgtcaagg ccgcatgcat	60
aataaaatat ctttattttc attacatctg tgtgttggtt ttttgtgtgt gacgtgctga	120
tgatgcgtag ctagtagttg acgtgctgat gatgcgtagc tagtagttga cgtgctgatg	180
atgcgtagct agtagttgac gtgctgatga tgcgtagcta gtagttgacg tgctgatgat	240
gcgtagctag tagttgacgt gctgatgatg cgtagctagt agtgcagtta gcgtagctga	300
ggtaccgtcg acgatatcgg atccaggtct atataagcag agctcgttta gtgaaccgtc	360
agatgccta gatacgccat ccacgtgtt ttgacctca tagaagatcg ccaccatgga	420
agatgcaaaa aacattaaga agggcccagc gccattctac cactcgaag acgggaccgc	480
cggcgagcag ctgcacaaag ccatgaagcg ctacgccctg tgccccggca ccatcgcctt	540
taccgacgca catatcgagg tggacattac ctacgccgag tacttcgaga tgagcgttcg	600
gctggcagaa gctatgaagc gctatgggct gaatacaaac catcggatcg tgggtgtgcag	660

pctgb2018052387-seq1.txt

cgagaatagc ttgcagttct tcatgcccgt gttgggtgcc ctgttcatcg gtgtggctgt	720
ggccccagct aacgacatct acaacgagcg cgagctgctg aacagcatgg gcatcagcca	780
gcccaccgtc gtattcgtga gcaagaaagg gctgcaaaag atcctcaacg tgcaaaagaa	840
gctaccgatc atacaaaaga tcatcatcat ggatagcaag accgactacc agggcttcca	900
aagcatgtac accttcgtga cttcccattt gccacccggc ttcaacgagt acgacttcgt	960
gcccgagagc ttcgaccggg aaaaaacat cgccctgata atgaacagta gtggcagtac	1020
cggattgccc aagggcgtag ccctaccgca ccgcaccgct tgtgtccgat tcagtcatgc	1080
ccgcgacccc atcttcggca accagatcat ccccgacacc gctatcctca gcgtggtgcc	1140
atttcaccac ggcttcggca tggtcaccac gctgggctac ttgatctgcg gctttcgggt	1200
cgtgctcatg taccgcttcg aggaggagct attcttgcgc agcttgcaag actataagat	1260
tcaatctgcc ctgctggtgc ccacactatt tagcttcttc gctaagagca ctctcatcga	1320
caagtacgac ctaagcaact tgcacgagat cgccagcggc ggggcgccgc tcagcaagga	1380
ggtaggtgag gccgtggcca aacgcttcca cctaccaggc atccgccagg gctacggcct	1440
gacagaaaca accagcgcca ttctgatcac cccgaagg gacgacaagc ctggcgcagt	1500
aggcaaggtg gtgcccttct tcgaggctaa ggtggtggac ttggacaccg gtaagacact	1560
gggtgtgaac cagcgcggcg agctgtgcgt ccgtggcccc atgatcatga gcggctacgt	1620
taacaacccc gaggtacaa acgctctcat cgacaaggac ggctggctgc acagcggcga	1680
catcgcctac tgggacgagg acgagcactt cttcatcgtg gaccggctga agagcctgat	1740
caaatacaag ggctaccagg tagccccagc cgaactggag agcatcctgc tgcaacaccc	1800
caacatcttc gacgccgggg tcgccggcct gcccgacgac gatgccggcg agctgccgc	1860
cgcagtcgtc gtgctggaac acggtaaaac catgaccgag aaggagatcg tggactatgt	1920
ggccagccag gttacaaccg ccaagaagct gcgcggtggt gttgtgttcg tggacgaggt	1980
gcctaaagga ctgaccggca agttggacgc ccgcaagatc cgcgagattc tcattaaggc	2040
caagaagggc ggcaagatcg ccgtgtaatg aaagcttggc ctctacgagt aatagacgcc	2100
cagttgaatt ccttcgagca gacatgataa gatacattga tgagtttgga caaaccacaa	2160
ctagaatgca gtgaaaaaaaa tgctttattt gtgaaatttg tgatgctatt gctttatttg	2220

pctgb2018052387-seq1.txt

taaccattat aagctgcaat aaacaagtta acaacaacaa ttgcattcat tttatgtttc	2280
aggttcaggg ggaggtgtgg gaggtttttt aaagcaagta aaacctctac aaatgtggta	2340
aaatcgataa ggatccgtct gggcctcatg ggccttccgc tcaactgcccg ctttccagtc	2400
gggaaacctg tcgtgccagc tgcac	2425

<210> 61
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Spacer sequence

<400> 61	
gatgatgcgt agctagtagt	20