

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
1 March 2007 (01.03.2007)

PCT

(10) International Publication Number
WO 2007/025008 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01) *C12N 15/86* (2006.01)

(21) International Application Number:
PCT/US2006/033017

(22) International Filing Date: 23 August 2006 (23.08.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/711,149 24 August 2005 (24.08.2005) US

(71) Applicant (for all designated States except US): **THE SCRIPPS RESEARCH INSTITUTE** [US/US]; 10550 North Torrey Pines Road, La Jolla, CA 92037 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **MAURO, Vincent, P.** [CA/US]; 7311 Juncus Court, Sa Diego, CA 92121 (US). **EDELMAN, Gerald, M.** [US/US]; 7428 High Avenue, La Jolla, CA 92037 (US). **ZHOU, Wei** [CN/US]; 9338 Babuata Road, #77, San Diego, CA 92129 (US).

(74) Agent: **HAILE, Lisa, A.**; DLA Piper Rudnick Gray Cary US LLP, 4365 Executive Drive, Suite 1100, San Diego, CA 92121-2133 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: TRANSLATION ENHANCER-ELEMENT DEPENDENT VECTOR SYSTEMS

(57) Abstract: A translation enhancer-driven positive feedback vector system is disclosed which is designed to facilitate identification of a Translational Enhancer Element (TEE) and to provide a means for overexpression of gene products. The system exploits both transcriptional and translational approaches to control the expression levels of genes and/or gene products. Methods are also disclosed for screening libraries of random nucleotide sequences to identify translational elements and for overproduction of proteins, which have uses in both research and industrial environments.

WO 2007/025008 A2

TRANSLATION ENHANCER-ELEMENT DEPENDENT VECTOR SYSTEMS

Field of Invention

[0001] The present invention relates generally to vector constructs, and more specifically to positive feedback vector constructs bearing translational enhancer elements (TEEs) in combination with transcriptional elements and genes encoding transcription factors, where such constructs may be used to identify other TEEs and to modulate levels of protein expression.

Background of the Invention

[0002] Eukaryotic mRNAs can initiate translation by either cap-dependent or cap-independent mechanisms. Presently, the relative contributions of these mechanisms to the proteome are unknown; however, some studies suggest that cap-independent mechanisms may account for the translation of many mRNAs. For some mRNAs, cap-independent translation is facilitated by sequence elements termed internal ribosome entry sites (IRESes). IRESes were first discovered in uncapped picornavirus RNAs and were subsequently identified in other viral and cellular mRNAs from mammals, insects, and yeast. For some mRNAs, IRESes facilitate translation when cap-dependent initiation is less efficient or blocked. Internal initiation also facilitates the translation of particular mRNAs with 5' leaders that are encumbered by numerous upstream AUGs or RNA secondary structures.

[0003] A variety of evidence suggests that different IRESes vary in length, sequence composition, and in their requirements for initiation factors or other trans-acting factors, suggesting that internal initiation of translation occurs by a number of different mechanisms. Some IRESes are modular in composition. For example, an IRES module from the 5' leader of the *Gtx* homeodomain mRNA showed that maximal activity was obtained with sequences of 7 nucleotides. Various lines of evidence suggested that the mechanism underlying the activity of this sequence element involves base pairing to a complementary sequence of 18S rRNA. In another study, a 22-nt IRES was identified in the 5' leader of the *Rbm3* mRNA. In addition, it has been reported that the 5' leader of the thymidine kinase mRNA contains an IRES-element and that the 5' leader of the c-myc mRNA contains two short IRES elements.

[0004] The short size of some IRES/TEE modules suggests that they may be prevalent within mRNA populations.

[0005] Some IRES elements can also function as translation enhancer elements (TEEs), i.e., they can enhance translation in the context of a monocistronic mRNA. However, not all TEEs are IRESes and not all IRESes are TEEs.

SUMMARY OF THE INVENTION

[0006] The present invention describes a series of vectors designed to select for translational enhancer elements and overexpression of proteins of interest, and includes methods for the use of such vectors.

[0007] In one embodiment, a nucleic acid vector including a first construct that includes two or more first transcriptional elements, a first cistron encoding a transcription factor, and one or more first translational enhancer elements (TEEs), where the transcription factor amplifies the transcription of at least one cistron of the first construct is envisaged.

[0010] In a related aspect, the vector may include, but is not limited to, at least one cistron on one or more second constructs including at least one transcriptional unit. In a further related aspect, the first construct and the at least one transcriptional unit of one or more second constructs include transcriptional elements that are targets for the transcription factor encoded by the first cistron.

[0011] In a related aspect, such vectors may encode a gene product that is a reporter protein, a therapeutic protein, an enzyme, an antigen, a structural protein, or an antibody.

[0012] In another aspect, at least one gene product blocks host protein synthesis. In a related aspect, the gene product may include, but is not limited to, NSP3, L-proteinase, or proteinase 2A.

[0013] In one aspect, the vector includes at least one TEE that is resistant to the activity of the product which blocks host protein synthesis, where the vector may contain transcriptional elements including, but not limited to, minimal promoters, regulatable promoters, upstream activating sequences, and bacteriophage RNA polymerase specific promoters.

[0014] In another embodiment, a nucleic acid vector including a first construct which includes a first transcriptional element, a first cistron encoding a first gene product, and a first

translational enhancer element (TEE), wherein the TEE is resistant to an activity of a second gene product which blocks host protein synthesis is envisaged.

[0015] In a related aspect, TEEs include, HCV-IRES, IRESes, and IRES-elements, including, but not limited to, Gtx sequences (e.g., Gtx9-nt, Gtx8-nt, Gtx7-nt). In another related aspect, TEEs may include N-18 random nucleotides which when operably linked to a cistron, increase the amount of protein induced per unit mRNA.

[0016] In one embodiment, a method of identifying a translational enhancer element (TEE) is envisaged including inserting nucleotides from a library of nucleotides into a vector including a first construct which includes two or more first transcriptional elements, and a first cistron encoding a transcription factor, transfecting a cell with the vector, and determining the level of gene product translation from one or more second constructs in the transfected cell, where determining an enhanced level of translation of the gene product in the presence of the inserted nucleotides is indicative of the presence of at least one TEE.

[0017] In a related aspect, the method may further include co-transfecting the cell with one or more second vectors, wherein the second vectors include the one or more second constructs.

[0018] In another embodiment, a method of overexpressing a gene is envisaged including, transfecting a cell with a vector including a first construct including two or more first transcriptional elements, a first cistron encoding a transcription factor, and one or more first translation enhancer elements (TEEs), and expressing a gene product from one or more second constructs including a second TEE, where the resulting level of gene product expressed from the second construct is enhanced in the presence of the first and second TEEs.

[0019] In a related aspect, the method may further include co-transfecting the cell with one or more second vectors including the one or more second constructs. In a further related aspect, the method may further include expressing a gene from a third construct including a third TEE, where the gene from the third construct encodes one or more gene products which block host protein synthesis. In a related aspect, the method may further include co-transfecting the cell with a third vector including the third construct.

[0020] In one embodiment, a method of overexpressing a gene is envisaged including transfecting a cell with a vector including a first construct including at least one first transcriptional element, a first cistron encoding a first gene product, and at least one translational enhancer element (TEE) and expressing a gene from one or more second constructs encoding a protein which blocks host protein synthesis, where the resulting level of gene product expressed from the first construct is enhanced in the presence of the blocking protein.

[0021] Exemplary methods and compositions according to this invention, are described in greater detail below.

BRIEF DESCRIPTION OF THE DRAWINGS

[0022] Figure 1 illustrates a translation enhancer-driven positive feedback vector. A schematic representation of the positive feedback vector is shown along with the various promoter (P1) and transcriptional enhancer (P2) sequences, transcription factor (TF) genes, and a protein of interest. The transcription factor gene and the gene of interest may be on the same plasmid or on different plasmids. For the selection application, a random nucleotide sequence (N), n nucleotides in length is present in the 5' leader of the transcription factor mRNA. A sequence that functions as a translational enhancer element (TEE) will facilitate the translation of this mRNA. The encoded transcription factor will then bind to sites in the promoters of the two genes and increase their transcription.

[0023] Figure 2 illustrates a translation enhancer-driven positive feedback vector with a third protein to block host protein synthesis. The first two genes (transcription factor gene and gene of interest) are the same as in Figure 1 except that all three mRNAs contain a TEE in their 5' leader. The third protein (e.g., the Rotavirus NSP3 protein) will increase the translation of the first two encoded mRNAs by blocking the translation of host mRNAs and reducing the competition from them. The third gene is under the transcriptional control of promoter P3, which is either a constitutive promoter or an inducible promoter. P3 may also include promoter elements P1 and P2. For the selection application, the mRNAs for the gene of interest and third protein will contain a known TEE, while the transcription factor gene will contain a random nucleotide sequence. In this scenario, the TEE is resistant to the activity of the third protein. For a protein production application, all three genes may contain

a known TEE that is resistant to the activity of the third protein. The three genes may be on one, two, or three different plasmids.

[0024] Figure 3 illustrates a protein overexpression vector. The first gene encodes the gene of interest. As shown in Figure 2, each gene contains a TEE in its 5' leader. In this scenario, the TEE is resistant to the activity of the other encoded protein (e.g., NSP3). The two genes may be on one or two different plasmids.

[0025] Figure 4 shows a restriction map for a positive feedback reporting vector (SEQ ID NO: 1). Sequences in bold represent: 1) promoter sequences, including TATA boxes and upstream activating sequences (UAS); 2) GAL4R1-GAL4R4, primer sequences for PCR amplification; 3) ATG and TAA, start and stop translation sequences, respectively; and 4) HRGB1-HRGB4, primer sequences for PCR amplification.

[0026] Figure 5 shows sequences and one-cut enzymes for pUAS-GV16-UAS-EGFP (SEQ ID NO: 2).

DETAILED DESCRIPTION OF THE INVENTION

[0027] Before the present compositions and methods are described, it is understood that this invention is not limited to the particular methodology, protocols, and reagents described as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be described by the appended claims.

[0028] It must be noted that as used herein and in the appended claims, the singular forms "a", "an", and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, reference to "a cell" includes a plurality of such cells, reference to "a protein" includes one or more proteins and equivalents thereof known to those skilled in the art, and so forth.

[0029] Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the methods,

devices, and materials are now described. All publications mentioned herein are incorporated herein by reference for the purpose of describing and disclosing the proteins, nucleic acids, and methodologies which are reported in the publications which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

[0030] As used herein, "translational enhancer element (TEE)," including grammatical variations thereof, means cis-acting sequences that increase the amount of protein induced per unit mRNA. In a related aspect, TEEs include, HCV-IRES, IRESes, and IRES-elements, including, but not limited to, Gtx sequences (e.g., Gtx9-nt, Gtx8-nt, Gtx7-nt). In another related aspect, TEEs may include N-18 random nucleotides which when operably linked to a cistron, increase the amount of protein induced per unit mRNA.

[0031] In a further related aspect, sequences for such elements include, but are not limited to, GenBank accession numbers AX205123 and AX205116 (Gtx IRES element), D17763 (HCV-IRES, 5'-untranslated region).

[0032] As used herein, "cistron" including grammatical variations thereof, means a unit of DNA that codes for a single polypeptide or protein.

[0033] As used herein, "transcriptional unit," including grammatical variations thereof, means the segment of DNA within which the synthesis of RNA occurs.

[0034] As used herein, "nucleotide sequence," "nucleic acid sequence," "nucleic acid," or "polynucleotide," refers to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and unless otherwise limited, encompasses known analogs of natural nucleotides that hybridize to nucleic acids in a manner similar to naturally-occurring nucleotides. Nucleic acid sequences can be, e.g., prokaryotic sequences, eukaryotic mRNA sequences, cDNA sequences from eukaryotic mRNA, genomic DNA sequences from eukaryotic DNA (e.g., mammalian DNA), and synthetic DNA or RNA sequences, but are not limited thereto.

[0035] In a related aspect, synthetic methods for preparing a nucleotide sequence include, for example, the phosphotriester and phosphodiester methods (see Narang et al., *Meth. Enzymol.* 68:90, (1979); U.S. Pat. No. 4,356,270, U.S. Pat. No. 4,458,066, U.S. Pat. No.

4,416,988, U.S. Pat. No. 4,293,652; and Brown et al., *Meth Enzymol* 68:109, (1979), each of which is incorporated herein by reference).

[0036] As used herein, "promoter" including grammatical variations thereof, means a nucleic acid sequence capable of directing transcription. A variety of promoter sequences are known in the art. For example, such elements can include, but are not limited to, TATA-boxes, CCAAT-boxes, bacteriophage RNA polymerase specific promoters (T7: TAATACGACTCACTATAGG; SP6: ATTTAGGTGACACTATAGA; and T3: AATTAACCCTCACTAAAGG), an SP1 site (GGGCGG), and a cyclic AMP response element (TGACGTCA).

[0037] As used herein, "transcriptional element," including grammatical variations thereof, means a cis-acting site on DNA that allows for initiation or stimulation of initiation of transcription, usually through recognition by a transcription factor. For example, such elements exist in motifs including, but not limited to, CACGTG (c-Myc), TGAc/gTc/aA (c-Fos), and t/aGATA (GATA). In a related aspect, the vector may contain one or more transcriptional elements comprising one or more upstream activating sequences (UAS), including but not limited to, the consensus GAGTACTGTCCTCCGAGCG.

[0038] As used herein, "transcription factor," including grammatical variations thereof, means any protein required to initiate or regulate transcription. For example, such factors include, but are not limited to, c-Myc, c-Fos, c-Jun, CREB, cEts, GATA, GAL4, GAL4/Vp16, c-Myb, MyoD, NF- κ B, bacteriophage-specific RNA polymerases, Hif-1, and TRE.

[0039] In a further related aspect, sequences for such factors include, but are not limited to, GenBank accession numbers K02276 (c-Myc), K00650 (c-fos), BC002981 (c-jun), M27691 (CREB), X14798 (cEts), M77810 (GATA), K01486 (GAL4), AY136632 (GAL4/Vp16), M95584 (c-Myb), M84918 (MyoD), 2006293A (NF- κ B), NP 853568 (SP6 RNA polymerase), AAB28111 (T7 RNA polymerase), NP 523301 (T3 RNA polymerase), AF364604 (HIF-1), and X63547 (TRE).

[0040] As used herein, "transcriptional activator regions," including grammatical variations thereof, means protein sequences that, when tethered to DNA near a promoter, activate transcription by contacting targets in the transcriptional machinery (see, e.g.,

Xiangyang et al., Proc Natl Acad Sci USA (2000) 97:1988-1992). Activator regions, characterized by having an excess of acidic amino acid residues, are found in a wide array of eukaryotic activators, including the yeast activators Gal4, GCN4, and the herpesvirus activator VP16.

[0041] As used herein, "construct," including grammatical variations thereof, means nucleic acid sequence elements arranged in a definite pattern of organization such that the expression of genes/gene products that are operably linked to these elements can be predictably controlled. In a related aspect, a wide variety of heterologous sequences may be included in the construct, including, but not limited to, for example, sequences which encode growth factors, cytokines, chemokines, lymphokines, toxins, prodrugs, antibodies, antigens, ribozymes, as well as antisense sequences. In another related aspect, such heterologous sequences encode proteins which can serve as therapeutic modalities.

[0042] As used herein, "vector," including grammatical variations thereof, means the DNA of any transmissible agent (e.g., plasmid or virus) into which a segment of foreign DNA can be spliced in order to introduce the foreign DNA into host cells to promote its replication and/or transcription.

[0043] As disclosed herein, a vector comprising a construct is useful for identifying translational enhancer elements. In one embodiment, the construct is contained in a vector, which generally is an expression vector that contains certain components, but otherwise can vary widely in sequence and in functional element content. The vector also can contain sequences that facilitate recombinant DNA manipulations, including, for example, elements that allow propagation of the vector in a particular host cell (e.g., a bacterial cell, insect cell, yeast cell, or mammalian cell), selection of cells containing the vector (e.g., antibiotic resistance genes for selection in bacterial or mammalian cells), and cloning sites for introduction of reporter genes or the elements to be examined (e.g., restriction endonuclease sites or recombinase recognition sites).

[0044] Preferably, constructs as envisaged provide the advantage that the activity of an oligonucleotide can be examined in the context or milieu of the whole eukaryotic chromosome. A chromosome offers unique and complex regulatory features with respect to the control of gene expression, including translation. As such, it is advantageous to have a

system and method for obtaining regulatory oligonucleotides that function in the context of a chromosome. Thus, a method of the invention can be practiced such that integration of the expression vector into the eukaryotic host cell chromosome occurs, forming a stable construct prior to selection for an expressed reporter molecule.

[0045] A vector comprising a construct as envisaged can be integrated into a chromosome by a variety of methods and under a variety of conditions. Thus, the present invention should not be construed as limited to the exemplified methods. Shotgun transfection, for example, can result in stable integration if selection pressure is maintained upon the transfected cell through several generations of cell division, during which time the transfected nucleic acid construct becomes stably integrated into the cell genome. Directional vectors, which can integrate into a host cell chromosome and form a stable integrant, also can be used. These vectors can be based on targeted homologous recombination, which restricts the site of integration to regions of the chromosome having the homology, and can be based on viral vectors, which can randomly associate with the chromosome and form a stable integrant, or can utilize site specific recombination methods and reagents such as a lox-Cre system and the like.

[0046] Shotgun transfections can be accomplished by a variety of well known methods, including, for example, electroporation, calcium phosphate mediated transfection, DEAE dextran mediated transfection, a biolistic method, a lipofectin method, and the like. For random shotgun transfections, the culture conditions are maintained for several generations of cell division to ensure that a stable integration has resulted and, generally, a selective pressure also is applied. A viral vector based integration method also can be used and provides the advantage that the method is more rapid and establishes a stable integration by the first generation of cell division. A viral vector based integration also provides the advantage that the transfection (infection) can be performed at a low vector:cell ratio, which increases the probability of single copy transfection of the cell. A single copy expression vector in the cell during selection increases the reliability that an observed regulatory activity is due to a particular oligonucleotide, and facilitates isolation of such an oligonucleotides.

[0047] Reference is made herein to techniques commonly known in the art. Guidance in the application of such techniques can be found, e.g., in Ausubel et al. eds., 1995, *Current Protocols In Molecular Biology*, and in Sambrook et al., 1989, *Molecular Cloning: A*

Laboratory Manual, Cold Spring Harbor Laboratory Press, NY, the contents of which are incorporated herein by reference.

[0048] Previous studies have generated synthetic IRESes containing multiple individual IRES elements and showed that this multimerization led to higher, and in some cases exponential, increases in IRES activity. To facilitate the discovery process, a number of methods to screen for IRES elements in mammalian cells and in yeast have been developed. In all of these methods, dicistronic mRNAs containing a library of random nucleotide sequences in the intercistronic sequences (ICS) were expressed in cells, and those cells containing IRES elements were identified on the basis of the expression of the second cistron. The mammalian methods used a fluorescent reporter protein as the second cistron, and positive cells were identified with FACS. However, a limitation of these methods was that the activities of individual IRES elements were relatively low, leading to a large number of false positive cells.

[0049] To circumvent this signal-to-noise problem, the present invention describes a positive feedback based system where one cistron encoding a TEE containing transcription factor triggers a positive feedback loop in which the transcription factor binds to a select sequence in the upstream promoter of the cistron encoding the factor and one or more cistrons encoding an mRNA of interest, thereby increasing the transcription of both mRNAs. More mRNA results in more transcription factor, leading to ever-increasing amounts of both the transcription factor and protein of interest mRNA and the encoded proteins.

[0050] In one embodiment, the vector system is a translation enhancer-driven positive feedback vector (e.g., see Figure 1). In one aspect, a construct expresses two mRNAs: one encoding a protein of interest which may be a reporter protein and the other encoding a transcription factor. The transcription of both mRNAs is driven by minimal promoters but can be enhanced by the expression of the transcription factor via binding sites for the transcription factor that are located in the promoters of both genes. A small amount of transcription factor mRNA is expressed from the minimal promoter but the translation of this mRNA is blocked by an obstacle in the 5' leader of the mRNA encoding the transcription factor. This obstacle may be a stable stem-loop structure and/or upstream AUG initiation codons. The synthesis of this transcription factor is dependent on the presence of a translation enhancer in the mRNA encoding this factor.

[0051] The translational enhancer can be located in the 5' leader of the mRNA, downstream of the inhibitory elements but upstream of the initiation codon, or it may be located in the 3' untranslated region (UTR). In one aspect, the TEE is situated in the 5' leader sequence which is contained within a cistron.

[0052] Utilization of this vector system may require that the encoded transcription factor not be expressed in the cells of interest. For example, a transcription factor (e.g., Gal4/Vp16), that contains the DNA binding domain of the yeast GAL4 transcription factor is suitable for use in mammalian cells because mammalian genes do not appear to be targets of this transcription factor. Other suitable transcription factors include, but are not limited to, those that are expressed endogenously at very low levels, or are absent in the cells of interest. For example, bacteriophage specific- RNA polymerases (e.g., T7, T3, and SP6 RNA polymerases) are suitable for use in both mammalian cells and yeast. The genes as envisaged can be encoded by one or more vectors.

[0053] In one embodiment, the vector system can be used to identify a TEE. An oligonucleotide to be examined for translational activity can be operatively linked to an expressible polynucleotide, which, for example, can encode a reporter molecule. As used herein, the term "operatively linked" means that a regulatory element, which can be a synthetic regulatory oligonucleotide or an oligonucleotide to be examined for such activity, is positioned with respect to a translatable nucleotide sequence such that the regulatory element can affect its regulatory activity. An oligonucleotide having translational enhancer activity generally is positioned within about 1 to 500 nucleotides, particularly within about 1 to 100 nucleotides of a translation start site.

[0054] A library of randomized oligonucleotides to be examined for translational regulatory activity can be provided, and one or more individual members of the library can be cloned into multiple copies of the construct of the vector. The oligonucleotide to be examined for translational regulatory activity is introduced such that it is operatively linked to the minimal promoter element in the construct and, therefore, has the potential to function as a TEE. In this way, a library of different constructs, which can be contained in a vector, is formed, each construct differing in the introduced potential regulatory oligonucleotide sequence.

[0055] Oligonucleotides to be examined for translational regulatory activity can be, for example, cDNA sequences encoding 5' UTRs of cellular mRNAs, including a library of such cDNA molecules. Furthermore, as disclosed herein, TEEs identified according to a method of the invention, including synthetic TEE elements, have been found to be complementary to oligonucleotide sequences of ribosomal RNA, particularly to un-base paired oligonucleotide sequences of rRNA, which are interspersed among double stranded regions that form due to hybridization of self complementary sequences within rRNA. Accordingly, oligonucleotides to be examined for translational regulatory activity, can be designed based on their being complementary to an oligonucleotide sequence of rRNA, particularly to an un-base paired oligonucleotide sequence of rRNA such as a yeast, mouse or human rRNA (e.g., see GenBank Accession Nos. V01335, X00686, X03205, each of which is incorporated herein by reference). In addition, oligonucleotides to be examined for translational regulatory activity can be a library of variegated oligonucleotide sequences (see, for example, U.S. Pat. No. 5,837,500, incorporated herein by reference), which can be based, for example, on a translational enhancer element as disclosed herein or identified using a method of the invention, or on an oligonucleotide sequence complementary to an un-base paired region of a rRNA.

[0056] The oligonucleotides identified herein as having translational regulatory activity provide modules that can be used alone or combined with each other to produce desired activities. For example, concatemers of an identified TEE can vastly increase polypeptide expression from an associated cistron, including concatemers of 2, 5, 10, 20, 35, 50 or 75 copies of a TEE, which independently can be multiple copies of the same or different TEEs, and which can be operatively linked adjacent to each other or separated by spacer nucleotide sequences that can vary from 1 to about 100 nucleotides in length.

[0057] A synthetic translational regulatory element can be identified by screening, for example, a library of oligonucleotides containing a large number of different nucleotide sequences. The oligonucleotides can be variegated oligonucleotide sequences, which are based on but different from a known translational regulatory element, for example, an oligonucleotide complementary to an un-base paired sequence of a rRNA, or can be a random oligonucleotide library. The use of randomized oligonucleotides (e.g., N18) provides the advantage that no prior knowledge is required of the nucleotide sequence, and provides the

additional advantage that completely new regulatory elements can be identified. Methods for making a combinatorial library of nucleotide sequences or a variegated population of nucleotide sequences are well known in the art (see, for example, U.S. Pat. No. 5,837,500; U.S. Pat. No. 5,622,699; U.S. Pat. No. 5,206,347; Scott and Smith, *Science* 249:386-390, 1992; Markland et al., *Gene* 109:13 19, 1991; O'Connell et al., *Proc Natl Acad Sci, USA* 93:5883-5887, 1996; Tuerk and Gold, *Science* 249:505-510, 1990; Gold et al., *Ann Rev Biochem* 64:763-797, 1995; each of which is incorporated herein by reference).

[0058] A synthetic TEE oligonucleotide, which can be obtained using a method of the invention, can increase or decrease the level of translation of an mRNA containing the oligonucleotide. In particular, a TEE oligonucleotide can selectively regulate translation in a context specific manner, depending, for example, on the cell type for expression, the nature of the TEE sequence, or the presence of other effector sequences in the construct.

[0059] A regulatory element can be of various lengths from a few nucleotides to several hundred nucleotides. Thus, the length of an oligonucleotide in a library of oligonucleotides to be screened can be any length, including oligonucleotides as short as about 6 nucleotides or as long as about 100 nucleotides or more. Generally, the oligonucleotides to be examined are about 6, 12, 18, 30 nucleotides or the like in length. The complexity of the library, i.e., the number of unique members, also can vary, although preferably the library has a high complexity so as to increase the likelihood that regulatory sequences are present. Libraries can be made using any method known in the art, including, for example, using an oligonucleotide synthesizer and standard oligonucleotide synthetic chemistry. Where the oligonucleotides are to be incorporated into a vector, the library complexity depends in part on the size of the expression vector population being used to clone the random library and transfect cells. Thus, a theoretical limitation for the complexity of the library also relates to utilization of the library content by the recipient expression vector and by the transfected cells, as well as by the complexity that can be obtained using a particular method of oligonucleotide synthesis.

[0060] To identify TEEs, libraries of constructs that contain either random nucleotide sequences or cDNA segments of 5' leaders of mRNAs are introduced into cells. Cells containing a construct with a TEE upstream of the transcription factor will trigger the positive feedback mechanism and produce large amounts of both proteins. If one of the proteins is a

reporter protein, its activity can be readily assayed. For example, if the reporter is enhanced green fluorescent protein (EGFP), fluorescence activated cell sorting (FACS) can be used to identify cells expressing the fluorescent protein. In a preferred embodiment of the present invention, the means of detecting the presence of GFP in transfected cells is by fluorescence microscopy and by FACS. However, it will be readily understood by those of skill in the art that other means for detecting the presence of GFP may also be used in the practice of the present invention. The means of detecting GFP or of tracking or monitoring cells which have been transfected with a construct of the present invention may be any means whereby the presence GFP protein is detectable. For example, optical imaging, infrared imaging of gene expression, and flow cytometry may also be used.

[0061] In a related aspect, other reporter proteins can include, but are not limited to, enhanced yellow fluorescent protein (EYFP), enhanced cyan fluorescent protein (ECFP), luciferase, β -galactosidase, β -glucuronidase, alkaline phosphatase, and chloramphenicol acetyltransferase.

[0062] In a related aspect, such a vector can be used to identify sequences that enhance translation by various mechanisms, including but not limited to, cap-independent and cap-dependent mechanisms.

[0063] In one aspect, the vector system as envisaged can be used to overexpress another protein of interest. In a related aspect, the mRNA encoding the transcription factor will contain a known TEE. Such a system may be suitable for the batch production of proteins, where the transcription factor is under the control of a regulatable promoter, so that cells can be grown to a large volume before the positive feed back mechanisms and large scale protein production are induced by, for example, an inducing agent.

[0064] The term "inducing agent" is used to refer to a chemical, biological or physical agent that effects translation from an inducible translational regulatory element. In response to exposure to an inducing agent, translation from the element generally is initiated de novo or is increased above a basal or constitutive level of expression. Such induction can be identified using the methods disclosed herein, including detecting an increased level of a reporter polypeptide encoded by the expressible polynucleotide that is operatively linked to the TEE. An inducing agent can be, for example, a stress condition to which a cell is

exposed, for example, a heat or cold shock, a toxic agent such as a heavy metal ion, or a lack of a nutrient, hormone, growth factor, or the like; or can be exposure to a molecule that affects the growth or differentiation state of a cell such as a hormone or a growth factor. As disclosed herein, the translational regulatory activity of an oligonucleotide can be examined in cells that are exposed to particular conditions or agents, or in cells of a particular cell type, and oligonucleotides that have translational regulatory activity in response to and only under the specified conditions or in a specific cell type can be identified.

[0065] In another embodiment, a vector system is envisaged which expresses a third mRNA that encodes a protein that can increase the translation of the other two encoded mRNAs by decreasing the translation of cellular mRNAs (Figure 2). In one aspect, the expression of the third protein decreases competition from the cellular mRNAs and leads to an increased signal-to-noise ratio in the selection application.

[0066] In a related aspect, the third protein may be a viral protein that blocks translation of host mRNAs and thereby decreases the competition arising from these mRNAs. Viral proteins include, but are not limited to, NSP3, L-proteinase, or proteinase 2A. For example, without being bound to theory, the NSP3 protein blocks cap-dependent translation by binding to the eukaryotic initiation factor 4G (eIF4G) with high affinity, displacing the poly(A) binding protein, disrupting mRNA circulation, and dramatically decreasing efficiency. In this scenario, the three vector encoded mRNAs contain features that prevent their translation from being blocked. For example, where a TEE does not require eIF4G for activity, such an element would be resistant to NSP3 (e.g., Gtx IRES elements and the Cricket paralysis virus IRES). In a related aspect, the third protein can be under the control of the encoded transcription factor or under the control of an inducible promoter.

[0067] In one embodiment, the vector system expresses a gene product or protein of interest and a third protein, where the third protein blocks host protein synthesis (Figure 3). In a related aspect, the gene product or protein of interest will contain features that prevent their translation from being blocked. In another related aspect, the genes are transcribed by either a constitutive promoter or an inducible promoter.

[0068] A kit of the invention is also envisaged. Such a kit can contain a packaging material, for example, a container having a TEE containing oligonucleotide according to the

invention and a label that indicates uses of the oligonucleotide for regulating translation of a polynucleotide in an expression vector or other expression construct. In one embodiment, the system, preferably in kit form, provides an integrating expression vector for use in selecting a TEE oligonucleotide using a method as disclosed herein. Such a kit can contain a packaging material, which comprises a container having an integrating expression vector and a label that indicates uses of the vector for selecting oligonucleotide sequences capable of regulatory function.

[0069] Instructions for use of the packaged components also can be included in a kit of the invention. Such instructions for use generally include a tangible expression describing the components, for example, a TEE containing oligonucleotide, including its concentration and sequence characteristics, and can include a method parameter such as the manner by which the reagent can be utilized for its intended purpose. The reagents, including the oligonucleotide, which can be contained in a vector or operably linked to an expressible polynucleotide, can be provided in solution, as a liquid dispersion, or as a substantially dry powder, for example, in a lyophilized form. The packaging materials can be any materials customarily utilized in kits or systems, for example, materials that facilitate manipulation of the regulatory oligonucleotides and, if present, of the vector, which can be an expression vector. The package can be any type of package, including a solid matrix or material such as glass, plastic (e.g., polyethylene, polypropylene, and polycarbonate), paper, foil, or the like, which can hold within fixed limits a reagent such as a TEE containing oligonucleotide or vector. Thus, for example, a package can be a bottle, vial, plastic and plastic-foil laminated envelope, or the like container used to contain a contemplated reagent. The package also can comprise one or more containers for holding different components of the kit.

[0070] The following examples are intended to illustrate but not limit the invention.

EXAMPLES

[0071] **Example 1. Experimental Procedures/Materials**

[0072] *Construction of Vectors*

[0073] The constructs used in the present invention express one or more cistronic mRNAs that encode a transcription factor, protein of interest and/or a protein blocking host protein

synthesis. Promoters used to drive transcription of the cistronic mRNAs consist of a minimal promoter (TATA box) or regulatable promoter, alone or in combination with one or more other transcriptional elements, including upstream activating sequences (UAS) and bacteriophage specific promoters. For feedback vectors, a first cistron would comprise a gene encoding a transcription factor, which is inserted downstream from a known or unknown TEE. One such construct, containing a known TEE, would comprise a TATA box promoter element, one or more UAS sequences, and one or more *Gtx* modules upstream from a cistron encoding Gal4/Vp16.

[0074] Example 2. Positive Feedback Reporter Vector for Identifying TEE elements

[0075] Construction

[0076] As shown in Figure 4, the promoters used to drive transcription in this example comprise a minimal promoter (TATA box) in combination with four copies of the GAL4 upstream activating sequence (UAS). The first transcription unit encodes the GAL4/VP16 fusion protein and the second transcription unit encodes EGFP. The TEE insertion site (denoted by "N" in Figure 4) contains nucleotides from a library of 18 random nucleotides. The vector backbone is based on plasmid pHRG-B (Promega, Madison, WI). The original *Bam*HI site in pHRG-B was mutated so that both *Eco*RI and *Bam*HI sites in the TEE insertion site were unique. The random N₁₈ fragments are cloned into this reporter vector by using the *Eco*RI and *Bam*HI restriction sites.

[0077] Cell Culture and Transfection Analysis

[0078] Reporter constructs are transfected into Chinese hamster ovary cells (CHO) (2×10^4) by using FuGENE 6™ (Roche, Alameda, CA). Transfection efficiencies are normalized by co-transfection with a *LacZ* reporter gene construct (pCMVβ, Clontech, Mountainview, CA). Cells are harvested 3 days after transfection and sorted by FACS on a FACSVantage SE™ (Becton Dickinson, Franklin Lakes, NJ) (Also, see Owens et al., Proc Natl Acad Sci USA (2004) 101:9590-9594). β-galactosidase assays may be performed by methods known in the art. For example, see Chappell et al., Proc Natl Acad Sci USA (2000) 97:1536-1541.

[0079] Double-stranded oligonucleotides containing N18 sequences are cloned into the TEE assay site of the positive feedback vector by using *Eco*RI and *Bam*HI restriction sites.

Overnight ligations are performed using T4 DNA ligase at 16°C. The resulting ligation mix is transfected into CHO cells, and FACS analysis is performed as above. For each FACS analysis, the first 100,000 cells are analyzed and a sorting window is drawn to select the cells with the highest EGFP expression. DNA is extracted from cells recovered by FACS and PCR reactions are carried-out by using primers to sequences that flank the *EcoRI* and *BamHI* restriction sites. After digestion with both *EcoRI* and *BamHI* restriction enzymes, the resulting fragments are re-cloned into the same amplification vector and retested.

[0080] For determining the number of plasmids per transfected cell, equal amounts of two plasmids are mixed, CMV-EGFP and CMV-enhanced cyan fluorescent protein (CMV-ECFP; Clontech, Mountainview, CA). The cloning vector pBluescript-KS II™ (Stratagene, La Jolla, CA) is used as filler for co-transfection. CHO cells are transfected with the different mixtures and FACS analysis is performed 2 days later to assess the expression of both EGFP and ECFP.

[0081] All of the compositions and methods disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of illustrative embodiments, it will be apparent to those of skill in the art that variations may be applied to the composition, methods and in the steps or in the sequence of steps of the methods described herein without departing from the concept, spirit, and scope of the invention. More specifically, it will be apparent that certain agents which are both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. Although the invention has been described with reference to the above examples, it will be understood that modifications and variations are encompassed within the spirit and scope of the invention.

WHAT IS CLAIMED IS:

1. A nucleic acid vector comprising a first construct comprising two or more first transcriptional elements, a first cistron encoding a transcription factor, and one or more first translational enhancer elements (TEEs), wherein the transcription factor amplifies the transcription of at least one cistron of the first construct.
2. The vector of claim 1, wherein the TEEs are interposed between a transcription start site within the first transcriptional elements and the first cistron.
3. The vector of claim 1, further comprising at least one cistron on one or more second constructs comprising at least one transcriptional unit.
4. The vector of claim 1, wherein the first construct and the at least one transcriptional unit of one or more second constructs comprise transcriptional elements that are targets for the transcription factor encoded by the first cistron.
5. The vector of claim 1, wherein at least one of the sequences comprising the first transcriptional elements is not the target of transcription factors endogenous to a host cell.
6. The vector of claim 1, wherein the transcription factor is not endogenous to a host cell.
7. The vector of claim 3, wherein at least one cistron of the one or more second constructs encodes at least one gene product.
8. The vector of claim 7, wherein the gene product is a reporter protein, therapeutic protein an enzyme, an antigen, an antibody, or a structural protein.
9. The vector of claim 8, wherein the reporter protein is GFP, luciferase, β -galactosidase, β -glucuronidase, alkaline phosphatase, chloramphenicol acetyltransferase, ECFP, EGFP, or EYFP.

10. The vector of claim 7, wherein at least one second construct comprises at least one transcriptional element and one or more TEEs.
11. The vector of claim 10, wherein the at least one second construct comprises one or more TEEs interposed between a transcription start site within the at least one transcriptional element and the cistron encoding the gene product.
12. The vector of claim 7, wherein at least one gene product blocks host protein synthesis.
13. The vector of claim 12, wherein the gene product is NSP3, L-proteinase, or proteinase 2A.
14. The vector of claim 12, wherein at least one of the TEEs is resistant to the activity of the product which blocks host protein synthesis.
15. The vector of claim 1, wherein the two or more transcriptional elements are selected from the group consisting of minimal promoters, regulatable promoters, upstream activating sequences, and bacteriophage RNA polymerase specific promoters.
16. The vector of claim 1, wherein the first cistron encodes GAL4/VP16, GAL4, or a bacteriophage specific RNA polymerase.
17. The vector of claim 1, wherein at least one of the TEEs is an IRES, HCV-IRES or IRES element.
18. The vector of claim 1, wherein the IRES, HCV-IRES or IRES element is a Gtx sequence.
19. The vector of claim 1, wherein at least one of the TEEs is an N18 sequence, wherein N18 is a random oligonucleotide.
20. A nucleic acid vector comprising a first construct comprising a first transcriptional element, a first cistron encoding a first gene product, and a first translational enhancer

element (TEE), wherein the TEE is resistant to an activity of a second gene product which blocks host protein synthesis.

21. The vector of claim 20, wherein the vector comprises one or more second constructs which encode the second gene product.
22. The vector of claim 20, wherein the TEE is an IRES, HCV-IRES or IRES element.
23. The vector of claim 22, wherein the IRES, HCV-IRES or IRES element is a Gtx sequence.
24. The vector of claim 20, wherein at least one of the TEEs is an N18 sequence, wherein N18 is a random oligonucleotide.
25. The vector of claim 20, wherein the second gene product is NSP3, L-proteinase, or proteinase 2A.
26. A method of identifying a translational enhancer element (TEE) comprising:
 - a) inserting nucleotides from a library of nucleotides into a vector comprising a first construct comprising two or more first transcriptional elements, and a first cistron encoding a transcription factor;
 - b) transfecting a cell with the vector of step (a);
 - and
 - c) determining the level of gene product translation from one or more second constructs in the cell of step (b),wherein determining an enhanced level of translation of the gene product in the presence of the inserted nucleotides is indicative of the presence of at least one TEE.
27. The method of claim 26, further comprising co-transfecting the cell of step (b) with one or more second vectors comprising the one or more second constructs.
28. The method of claim 26, wherein at least one of the sequences comprising the first transcriptional elements is not the target of transcription factors endogenous to the cell.

29. The method of claim 26, wherein the transcription factor is not endogenous to the cell.
30. The method of claim 26, wherein two or more second constructs comprise at least one transcriptional element and a TEE which are 5'-proximal to at least one cistron contained therein.
31. The method of claim 30, wherein one of the second constructs comprises a cistron encoding a gene product which blocks host protein synthesis.
32. The method of claim 31, wherein the blocking gene product is NSP3, L-proteinase, or proteinase 2A.
33. The method of claim 31, wherein the at least one TEE is an IRES, HCV-IRES or IRES element.
34. The method of claim 33, wherein the IRES, HCV-IRES or IRES element is a Gtx sequence.
35. The method of claim 33, wherein at least one TEE is resistant to the activity of the product which blocks host protein synthesis.
36. The method of claim 26, wherein the two or more transcriptional elements are selected from the group consisting of minimal promoters, regulatable promoters, upstream activating sequences, and bacteriophage RNA polymerase specific promoters.
37. The method of claim 26, wherein the first cistron encodes GAL4/VP16, GAL4, or a bacteriophage specific RNA polymerase.
38. A method of overexpressing a gene comprising:
 - a) transfecting a cell with a vector comprising a first construct comprising two or more first transcriptional elements, a first cistron encoding a transcription factor, and one or more first translation enhancer elements (TEEs); and

b) expressing, in the cell of step (a), a gene product from one or more second constructs comprising a second TEE, wherein the resulting level of gene product expressed from the second construct is enhanced in the presence of the first and second TEEs.

39. The method of claim 38, further comprising co-transfecting the cell of step (a) with one or more second vectors comprising the one or more second constructs.

40. The method of claim 38, further comprising expressing, in the cell of step (a), a gene from a third construct comprising a third TEE, wherein the gene from the third construct encodes one or more gene products which block host protein synthesis.

41. The method of claim 40, further comprising co-transfecting the cell of step (a) with a third vector comprising the third construct.

42. The method of claim 40, wherein the blocking gene product is NSP3, L-proteinase, or proteinase 2A.

43. The method of claim 38, wherein the two or more transcriptional elements are selected from the group consisting of minimal promoters, regulatable promoters, upstream activating sequences, and bacteriophage RNA polymerase specific promoters.

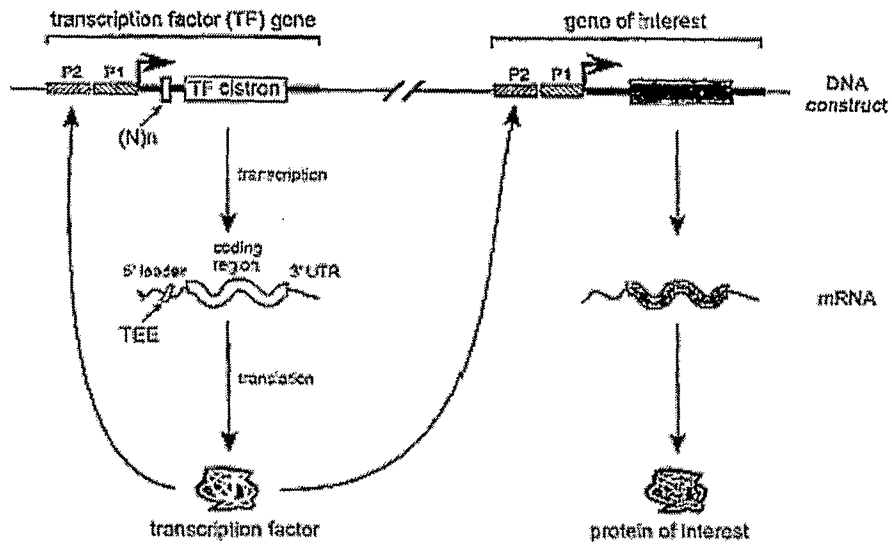
44. The method of claim 38, wherein the first cistron encodes GAL4/VP16, GAL4, or a bacteriophage specific RNA polymerase.

45. The method of claim 33, wherein the at least one TEE is an IRES, HCV-IRES or IRES element.

46. The method of claim 45, wherein the IRES, HCV-IRES or IRES element is a Gtx sequence.

47. The method of claim 33, wherein the at least one TEE is an N18 sequence, wherein N18 is a random oligonucleotide.

48. A method of overexpressing a gene comprising:
a) transfecting a cell with a vector comprising a first construct comprising at least one first transcriptional element, a first cistron encoding a first gene product, and at least one translational enhancer element (TEE); and
b) expressing, in the cell of step (a), a gene from one or more second constructs encoding a protein which blocks host protein synthesis with one or more TEEs,
wherein the resulting level of gene product expressed from the first construct is enhanced in the presence of the blocking protein.
49. The method of claim 48, further comprising co-transfecting the cell of step (a) with a second vector comprising the one or more second constructs.
50. The method of claim 48, wherein the TEE is resistant to the blocking protein.
51. The method of claim 48, wherein the two or more transcriptional elements are selected from the group consisting of minimal promoters and regulatable promoters.
52. The method of claim 48, wherein the at least one TEE is an IRES, HCV-IRES or IRES element.
53. The method of claim 52, wherein the IRES, HCV-IRES or IRES element is a Gtx sequence.
54. The method of claim 48, wherein the at least one TEE is an N18 sequence, wherein N18 is a random oligonucleotide.
55. The method of claim 48, wherein the blocking gene product is NSP3, L-proteinase, or proteinase 2A.



Promoter elements:

- P1: - minimal promoter (TATA)
- regulatable promoter
- P2: - upstream activating sequence (UAS; one to multiple copies)
- T7 promoter
- other transcription elements

Transcription factor genes:

- GAL4, GAL4/VP16
- T7 RNA polymerase
- other transcription factors

Genes of interest:

- fluorescent protein genes (e.g. enhanced green fluorescent protein; EGFP)
- luciferase genes (*Renilla* or *Photinus*)
- therapeutic protein (e.g. monoclonal antibodies)

Figure 1

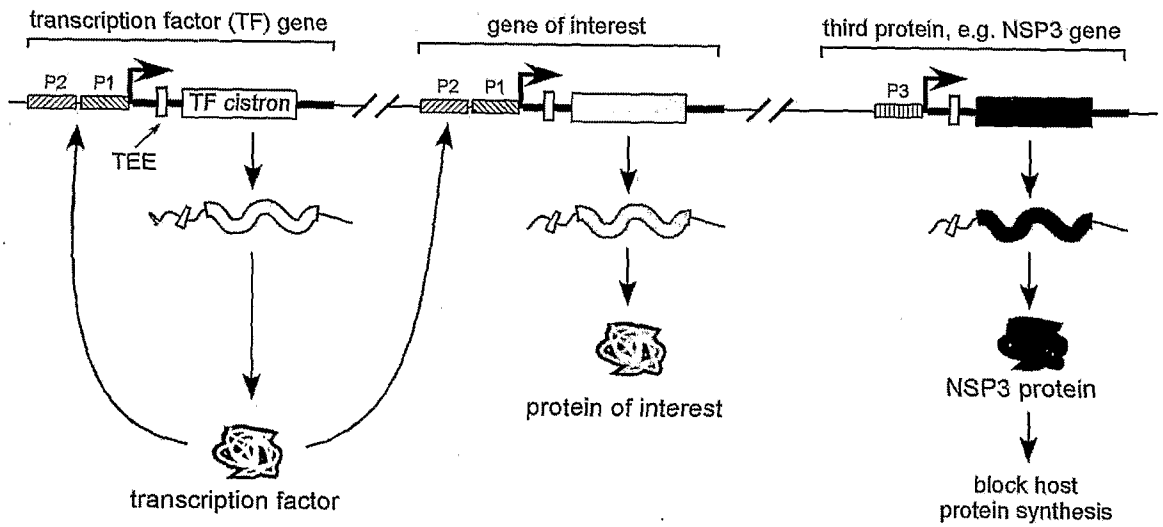


Figure 2

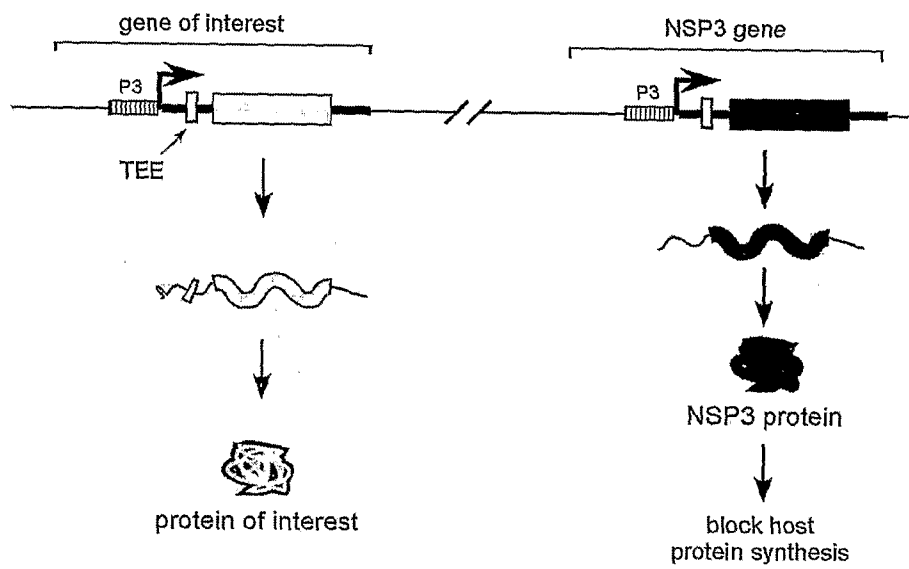


Figure 3

4/12

Sequences and one cut enzymes: pUAS-GV16-UAS-EGFP

Acc65I | KpnI | UAS
 | (ASP718) |
 1 GGTACCGAGCTCAAGCTTCGGAGTACTGTCTCTCCGAGCGGAGTACTGTCTCTCCGAGCGGA 60
 -----+-----+-----+-----+-----+-----+-----+-----+-----+
 CCATGGCTCGAGTTCGAAGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCT

BstBI (CSP45I)
 |
 61 GTACTGTCTCTCCGAGCGGAGTACTGTCTCTCCGTTTCAAAGCGGACCCCTGGCAGGAGGAA 120
 -----+-----+-----+-----+-----+-----+-----+-----+-----+
 CATGACAGGAGGCTCGCCTCATGACAGGAGGCAAGCTTTCGCTGGGGACCGTCTCTCTT

121 GGTACGAGAGCTGTGATAAGAGCCGTATAAAGAGGGTTCGCTCGCAAAGATCTATGG 180
 -----+-----+-----+-----+-----+-----+-----+-----+-----+
 CCAGTCGTCTCGACGACTATTCTCGGCATATTTCTCCCAAGGCGAGCGTTTCTAGATACC

EcoRI | BamHI
 | |
 181 CATGAATTCNNNNNNNNNNNNNNNNNNNNNGGATCCATGGTGAAGCTGTCTTCTATCGAACAA 240
 -----+-----+-----+-----+-----+-----+-----+-----+-----+
 GTACTTAAGNNNNNNNNNNNNNNNNNNNNCCTAGGTACCACTTCGACAGAAGATAGCTTGTT
 M V K L S S I E Q -

GAL4R1
 241 GCATGCGATATTTGCCGACTTAAAAAGCTCAAGTGCTCCAAAGAAAAACCGAAGTGCGCC 300
 -----+-----+-----+-----+-----+-----+-----+-----+-----+
CGTACGCTATAAACGGCTGAATTTTTCGAGTTCACGAGGTTTCTTTTGGCTTCACGCGG
 A C D I C R L K K L K C S K E K P K C A -

GAL4R2 **GAL4R3**
 301 AAGTGTCTGAAGAACAACCTGGGAGTGTGCTACTCTCCCAAACCAAAGGTCTCCGCTG 360
 -----+-----+-----+-----+-----+-----+-----+-----+-----+
TTACAGACTTCTTGTGACCCTCACAGCGATGAGAGGGTTTTGGTTTTCCAGAGGCGAC
 K C L K N N W E C R Y S P K T K R S P L -

GAL4R4
 361 ACTAGGGCACATCTGACAGAAGTGAATCAAGGCTAGAAAGACTGGAACAGCTATTTCTA 420
 -----+-----+-----+-----+-----+-----+-----+-----+-----+
TGATCCCGTGTAGACTGTCTTACCTTAGTTCCGATCTTTCTGACCTTGTCGATAAAGAT
 T R A H L T E V E S R L E R L E Q L F L -

XhoI
 |
 421 CTGATTTTCTCGAGAAGACCTTGACATGATTTTGAATGGATTTCTTTACAGGATATA 480
 -----+-----+-----+-----+-----+-----+-----+-----+-----+
 GACTAAAAAGGAGCTCTTCTGGAAGTACTAAACTTTTACCTAAGAAATGTCCTATAT
 L I F P R E D L D M I L K M D S L Q D I -

481 AAAGCATTGTTAACAGGATTATTTGTACAAGATAATGTGAATAAAGATGCCGTCACAGAT 540
 -----+-----+-----+-----+-----+-----+-----+-----+-----+
 TTTGTAACAATTGTCCTAATAAACATGTTCTATTACACTTATTTCTACGGCAGTGTCTA
 K A L L T G L F V Q D N V N K D A V T D -

541 AGATTGGCTTCAGTGGAGACTGATATGCCTCTAACATTGAGACAGCATAGAATAAGTGG 600
 -----+-----+-----+-----+-----+-----+-----+-----+-----+
 TCTAACCGAAGTCACCTCTGACTATACGGAGATTGTAAGTCTGTCGTATCTTATTCACGC
 R L A S V E T D M P L T L R Q H R I S A -

Figure 4

5/12

```

ACATCATCATCGGAAGAGAGTAGTAACAAAGGTCAAAGACAGTTGACTGTATCGATTAAA
601 -----+-----+-----+-----+-----+-----+-----+ 660
TGTAGTAGTAGCCTTCTCTCATCATTGTTTCCAGTTTCTGTCAACTGACATAGCTAATTT
T S S S E E S S N K G Q R Q L T V S I K -
GTCGCCCCCGACCGATGTCAGCCTGGGGGACGAGCTCCACTTAGACGGCGAGGACGTG
661 -----+-----+-----+-----+-----+-----+-----+ 720
CAGCGGGGGGGCTGGCTACAGTCGGACCCCTGCTCGAGGTGAATCTGCCGCTCCTGCAC
V A P P T D V S L G D E L H L D G E D V -

GCGATGGCGCATGCCGACGGCTAGACGATTTGATCTGGACATGTTGGGGGACGGGGAT
721 -----+-----+-----+-----+-----+-----+-----+ 780
CGCTACCGCGTACGGCTGCGCGATCTGCTAAAGCTAGACCTGTACAACCCCTGCCCTA
A M A H A D A L D D F D L D M L G D G D -
      SmaI
      XmaI |
      |
TCCCCGGGGCCGGGATTTACCCCCAGACTCCGCCCCCTACGGCGCTCTGGATATGGCC
781 -----+-----+-----+-----+-----+-----+-----+ 840
AGGGGCCCGGCCCTAAATGGGGGTGCTGAGGCGGGGGATGCCGCGAGACCTATACCGG
S P G P G F T P H D S A P Y G A L D M A -
                VP16-FOR
GACTTCGAGTTTGAGCAGATGTTTACCGATGCCCTTGAATTGACGAGTACGGTGGGTAA
841 -----+-----+-----+-----+-----+-----+-----+ 900
CTGAAGCTCAAACCTCGTCTACAAATGGCTACGGGACCTTAAC TGCTCATGCCACCATT
D F E F E Q M F T D A L G I D E Y G G * -
      SacII (KSP1)
      |
CCGCGGGCTAGAGTCGGGGCGGCCGGCGCTTCGAGCAGACATGATAAGATACATTGATG
901 -----+-----+-----+-----+-----+-----+-----+ 960
GGCGCCCGATCTCAGCCCCGCCGGCGGAAGCTCGTCTGTACTATTCTATGTAACTAC

AGTTTGGACAAACCACAAC TAGAATGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTG
961 -----+-----+-----+-----+-----+-----+-----+ 1020
TCAAACCTGTTTGGTGTGATCTTACGTCAC TTTTTTTACGAAATAAACACTTTAAACAC

ATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATT
1021 -----+-----+-----+-----+-----+-----+-----+ 1080
TACGATAACGAAATAAACATTGGTAATATTGACGTTATTTGTTCAATTGTTGTTGTTAA

GCATTCATTTTATGTTTCAGGTT CAGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAA
1081 -----+-----+-----+-----+-----+-----+-----+ 1140
CGTAAGTAAAATACAAAGTCCAAGTCCCCCTCCACACCCTCCAAAAAATTTCGTTCATTT

ACCTCTACAAATGTGGTAAAATCGATAAGGATCGATCCGTCGAGATCTGCGATCTAAGTA
1141 -----+-----+-----+-----+-----+-----+-----+ 1200
TGGAGATGTTTACACCATTTTAGCTATTCTAGCTAGGCAGCTCTAGACGCTAGATTTCAT
      PstI
      |
AGCTTGGCTGCAGGTCGACGGATCGGAGTACTGTCTCCGAGCGGAGTACTGTCTCCGA
1201 -----+-----+-----+-----+-----+-----+-----+ 1260
TCGAACCGACGTCCAGCTGCCTAGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCT
      UAS
GCGGAGTACTGTCTCCGAGCGGAGTACTGTCTCCGAGCGGACCCCTGGCAGGAGGAAG
1261 -----+-----+-----+-----+-----+-----+-----+ 1320
CGCCTCATGACAGGAGGCTCGCCTCATGACAGGAGGCTCGCCTGGGGACCGTCTCTCTC

```

Figure 4 (cont'd)

GTCAGCAGAGCTGCTGATAAGAGCCGTATAAAGAGGGTTCCGCTCATGGCAAGGGGCAGT
1321 -----+-----+-----+-----+-----+-----+-----+ 1380
CAGTCGTCTCGACGACTATTCTCGGCATATTTCTCCAAGGCGAGTACCGTTCCCCGTCA

GGTCTCGGGATCTGAGCTTGGCATTCGGTACTGTGGTAAACCATGGTGAGCAAGGGCG
1381 -----+-----+-----+-----+-----+-----+-----+ 1440
CCAGAGCCCTAGACTCGAACCGTAAGGCCATGACAACCATTTGGTACCACTCGTTCCCGC
M V S K G E -
AGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAACGGCC
1441 -----+-----+-----+-----+-----+-----+-----+ 1500
TCCTCGACAAGTGGCCCCACCACGGGTAGGACCAGCTCGACCTGCCGCTGCATTTGCCGG
E L F T G V V P I L V E L D G D V N G H -

ACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGA
1501 -----+-----+-----+-----+-----+-----+-----+ 1560
TGTTCAAGTCGCACAGGCCGCTCCCGCTCCCGCTACGGTGGATGCCGTTGACTGGGACT
K F S V S G E G E G D A T Y G K L T L K -

AGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGA
1561 -----+-----+-----+-----+-----+-----+-----+ 1620
TCAAGTAGACGTGGTGGCCGTTTCGACGGGCACGGGACCGGGTGGGAGCACTGGTGGGACT
F I C T T G K L P V P W P T L V T T L T -

CCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCAGCACTTCTTCA
1621 -----+-----+-----+-----+-----+-----+-----+ 1680
GGATGCCGCACGTCACGAAGTCGGCGATGGGGCTGGTGTACTTCGTGCTGCTGAAGAAGT
Y G V Q C F S R Y P D H M K Q H D F F K -

AGTCCGCCATGCCCCAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCA
1681 -----+-----+-----+-----+-----+-----+-----+ 1740
TCAGGCGGTACGGGCTTCCGATGCAGGTCTTCGCGTGGTAGAAGAAGTTCCTGCTGCCGT
S A M P E G Y V Q E R T I F F K D D G N -

ACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGC
1741 -----+-----+-----+-----+-----+-----+-----+ 1800
TGATGTTCTGGGCGCGGCTCCACTTCAAGCTCCCGCTGTGGGACCCTTGGCGTAGCTCG
Y K T R A E V K F E G D T L V N R I E L -

TGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAAC
1801 -----+-----+-----+-----+-----+-----+-----+ 1860
ACTTCCCGTAGCTGAAGTTCCTCCTGCCGTTGTAGGACCCCGTGTTCGACCTCATGTTGA
K G I D F K E D G N I L G H K L E Y N Y -

ACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACT
1861 -----+-----+-----+-----+-----+-----+-----+ 1920
TGTTGTCGGTGTTCAGATATAGTACCGGCTGTTTCGTCTTCTTGCCGTAGTTCCACTTGA
N S H N V Y I M A D K Q K N G I K V N F -

TCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGA
1921 -----+-----+-----+-----+-----+-----+-----+ 1980
AGTTCAGGCGGTGTTGTAGCTCCTGCCGTCGCACGTCGAGCGGCTGGTGATGGTCTGCT
K I R H N I E D G S V Q L A D H Y Q Q N -

Figure 4 (cont'd)

7/12

ACACCCCCATCGGGCAGCGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGT
 1981 -----+-----+-----+-----+-----+-----+-----+ 2040
 TGTGGGGGTAGCCGCTGCCGGGCACGACGACGGGCTGTTGGTGATGGACTCGTGGGTCA
 c T P I G D G P V L L P D N H Y L S T Q S -

 CCGCCCTGAGCAAAGACCCCAACGAGAAGCGGATCACATGGTCTGCTGGAGTTCGTGA
 2041 -----+-----+-----+-----+-----+-----+-----+ 2100
 GGCGGGACTCGTTTCTGGGGTTGCTCTTCGCGCTAGTGTACCAGGACGACCTCAAGCACT
 c A L S K D P N E K R D H M V L L E F V T -

 CCGCCCGGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAACGCGTTGACTAAGCTA
 2101 -----+-----+-----+-----+-----+-----+-----+ 2160
 GGCGGGCCCTAGTGAGAGCCGTACCTGCTCGACATGTTTCATTGCGCAACTGATTTCGAT
 c A A G I T L G M D E L Y K * -

 BssHII
 |
 TGGCGCGCACTAGGGGCTAGAGTCGGGGCGGCCGGCCGCTTCGAGCAGACATGATAAGAT
 2161 -----+-----+-----+-----+-----+-----+-----+ 2220
 ACCGCGCGTGATCCCCGATCTCAGCCCCCGCGCCGCGGAAGCTCGTCTGTACTATTCTA

 ACATTGATGAGTTTGGACAAACCACAAC TAGAATGCAGTGAAAAAATGCTTTATTTGTG
 2221 -----+-----+-----+-----+-----+-----+-----+ 2280
 TGTAAC TACTCAAACCTGTTGGTGTGATCTTACGTCAC TTTTTTTACGAAATAAACAC

 AAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACA
 2281 -----+-----+-----+-----+-----+-----+-----+ 2340
 TTTAAACACTACGATAACGAATAAACATTGGTAATATTCGACGTTATTTGTTCAATTGT

 ACAACAATTCGATTCATTTTATGTTTCAGGTTTCAGGGGAGGTGTGGGAGGTTTTTTAAA
 2341 -----+-----+-----+-----+-----+-----+-----+ 2400
 TGTTGTTAACGTAAGTAAAATACAAAGTCCAAGTCCCCTCCACACCTCCAAAAAATTT

 GCAAGTAAAACCTCTACAAATGTGGTAAAATCGATAAGGATCGATCCGTCGACCGATGCC
 2401 -----+-----+-----+-----+-----+-----+-----+ 2460
 CGTTCATTTTGGAGATGTTACACCATTTTAGCTATTCCTAGCTAGGCAGCTGGCTACGG

 CTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGC
 2461 -----+-----+-----+-----+-----+-----+-----+ 2520
 GAACTCTCGGAAGTTGGGTCAGTCGAGGAAGGCCACCCGCGCCCGTACTGATAGCAGCG

 CGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGTGCCGGCAGCGCTCTT
 2521 -----+-----+-----+-----+-----+-----+-----+ 2580
 GCGTGAATACTGACAGAAGAAATAGTACGTTGAGCATCCTGTCCACGGCCGTCGCGAGAA

 CCGCTTCCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCCGCTGCGGCGAGCGGTATCAG
 2581 -----+-----+-----+-----+-----+-----+-----+ 2640
 GGCGAAGGAGCGAGTGACTGAGCGACGCGAGCCAGCAAGCCGACGCCGCTCGCCATAGTC

 CTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACA
 2641 -----+-----+-----+-----+-----+-----+-----+ 2700
 GAGTGAGTTCCGCCATTATGCCAATAGGTGTCTTAGTCCCCTATTGCGTCCTTTCTTGT

Figure 4 (cont'd)

TGTGAGCAAAGGCCAGCAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTT
 2701 -----+-----+-----+-----+-----+-----+-----+ 2760
 AACTCGTTTTCCGGTCGTTTTCCGGTCCTGGCATTTCGGCGCAACGACCGCAAAA

 TCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGC
 2761 -----+-----+-----+-----+-----+-----+-----+ 2820
 AGGTATCCGAGGCGGGGGACTGCTCGTAGTGTTTTAGCTGCGAGTTCAGTCTCCACCG

 GAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCCT
 2821 -----+-----+-----+-----+-----+-----+-----+ 2880
 CTTTGGGCTGTCTGATATTTCTATGGTCCGCAAAGGGGGACCTTCGAGGGAGCACGCGA

 CTCTGTTCCGACCTGCCGCTTACCGGATACCTGTCCGCTTCTCCCTTCGGGAAGCG
 2881 -----+-----+-----+-----+-----+-----+-----+ 2940
 GAGGACAAGGCTGGGACGGCGAATGGCCTATGGACAGGCGGAAAGAGGGAAGCCCTTCGC

 TGGCGCTTCTCATAGCTCAGCTGTAGGTATCTCAGTTCGGTGTAGGTGCTTCGCTCCA
 2941 -----+-----+-----+-----+-----+-----+-----+ 3000
 ACCGCGAAAAGAGTATCGAGTGCACATCCATAGAGTCAAGCCACATCCAGCAAGCGAGGT

 AGCTGGGCTGTGTGCACGAACCCCCGTTTCAGCCCGACCGCTGCGCCTTATCCGGTAACT
 3001 -----+-----+-----+-----+-----+-----+-----+ 3060
 TCGACCCGACACACGTGCTTGGGGGGCAAGTCGGGCTGGCGACCGGAATAGGCCATTGA

 ATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTA
 3061 -----+-----+-----+-----+-----+-----+-----+ 3120
 TAGCAGAACTCAGGTTGGGCCATTCTGTCTGAATAGCGGTGACCGTCTCGGTGACCAT

 ACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTCTTGAAGTGGTGGCCTA
 3121 -----+-----+-----+-----+-----+-----+-----+ 3180
 TGTCTAATCGTCTCGCTCCATACATCCGCCACGATGTCTCAAGAACTTCACCACCGGAT

 ACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCT
 3181 -----+-----+-----+-----+-----+-----+-----+ 3240
 TGATGCCGATGTGATCTTCTGTGCATAAACCATAGACCGGAGACGACTTCGGTCAATGGA

 TCGGAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTT
 3241 -----+-----+-----+-----+-----+-----+-----+ 3300
 AGCCTTTTTCTCAACCATCGAGAAGTGGCGTTTGGTTGGTGGCGACCATCGCCACCAA

 TTTTTGTTTGC AAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGA
 3301 -----+-----+-----+-----+-----+-----+-----+ 3360
 AAAACAAACGTTTCGTCTAATGCGCGTCTTTTTTCTAGAGTCTTCTAGGAAACT

 TCTTTTCTACGGGTCTGACGCTCAGTGGAAACGAAACTCAGTTAAGGGATTTTGGTCA
 3361 -----+-----+-----+-----+-----+-----+-----+ 3420
 AGAAAAGATGCCCCAGACTGCGAGTACCTTGCTTTTGGTGGCAATCCCTAAAACAGT

 TGAGATTATCAAAAAGGATCTTACCTAGATCCTTTTAAATTAATAATGAAGTTTAAAT
 3421 -----+-----+-----+-----+-----+-----+-----+ 3480
 ACTCTAATAGTTTTTCTAGAAAGTGGATCTAGGAAAATTTAATTTTACTTCAAATTTA

Figure 4 (cont'd)

9/12

CAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGG
 3481 -----+-----+-----+-----+-----+-----+-----+ 3540
 GTTAGATTCATATATACTCATTTGAACCAGACTGTCAATGGTTACGAATTAGTCACTCC
 * W H K I L S -

 CACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGT
 3541 -----+-----+-----+-----+-----+-----+-----+ 3600
 GTGGATAGAGTCGCTAGACAGATAAAGCAAGTAGGTATCAACGGACTGAGGGGCAGCACA
 A G I E A I Q R N R E D M T A Q S G T T -

 AGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGTGCAATGATAACCGCGAG
 3601 -----+-----+-----+-----+-----+-----+-----+ 3660
 TCTATTGATGCTATGCCCTCCCGAATGGTAGACCGGGTACGACGTTACTATGGCGCTC
 Y I V V I R S P K G D P G L A A I I G R -

 ACCCAGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGC
 3661 -----+-----+-----+-----+-----+-----+-----+ 3720
 TGGGTGCGAGTGGCCGAGGCTAAATAGTCGTTATTTGGTTCGGTTCGGCCTTCCCGGCTCG
 S G R E G A G S K D A I F W G A P L A S -

 AseI
 |
 GCAGAAGTGGTCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAG
 3721 -----+-----+-----+-----+-----+-----+-----+ 3780
 CGTCTTCACCAGGACGTTGAAATAGGCGGAGGTAGGTACGATAATTAACAACGGCCCTTC
 R L L P G A V K D A E M W D I L Q Q R S -

 CTAGAGTAAGTAGTTCCGCGTAAATAGTTGCGCAACGTTGTTGCCATTGCTACAGGCA
 3781 -----+-----+-----+-----+-----+-----+-----+ 3840
 GATCTCATTCATCAAGCGGTCAATTATCAAACGCGTTGCAACAACGGTAACGATGTCGGT
 A L T L L E G T L L K R L T T A M A V P -

 TCGTGGTGTACGCTCGTCTGTTGGTATGGCTTCATTCAGCTCCGGTTCCTCAACGATCAA
 3841 -----+-----+-----+-----+-----+-----+-----+ 3900
 AGCACCACAGTGCAGCAGCAAACCATAACGAAGTAAGTCGAGGCCAAGGTTGCTAGTT
 M T T D R E D N P I A E N L E P E W R D -

 GCGAGTTACATGATCCCCATGTTGTGCAAAAAGCGGTTAGCTCCTTCGGTCTCCTCCGA
 3901 -----+-----+-----+-----+-----+-----+-----+ 3960
 CCGCTCAATGTACTAGGGGTACAACACGTTTTTTTCGCAATCGAGGAAGCCAGGAGGCT
 L R T V H D G M N H L F A T L E K P G G -

 TCGTTGTCAGAAGTAAGTTGGCCGAGTGTATCACTCATGGTTATGGCAGCACTGCATA
 3961 -----+-----+-----+-----+-----+-----+-----+ 4020
 AGCAACAGTCTTCATTCACCGGCGTCACAATAGTGAGTACCAATACCGTCGTGACGTAT
 I T T L L L N A A T N D S M T I A A S C -

 ATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCA
 4021 -----+-----+-----+-----+-----+-----+-----+ 4080
 TAAGAGAAATGACAGTACGGTAGGCATTTACGAAAAGACACTGACCACTCATGAGTTGGT
 L E R V T M G D T L H K E T V P S Y E V -

 AGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCTCAATACGGG
 4081 -----+-----+-----+-----+-----+-----+-----+ 4140
 TCAGTAAGACTCTTATCACATACGCCGCTGGCTCAACGAGAACGGGCCGAGTTATGCC
 L D N Q S Y H I R R G L Q E Q G A D I R -

Figure 4 (cont'd)

10/12

XmnI

4141 ATAAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGG
 -----+-----+-----+-----+-----+-----+-----+ 4200
 TATTATGGCGCGGTGTATCGTCTTGAAATTTTACAGAGTAGTAACCTTTGCAAGAAGCC
 S L V A G C L L V K F T S M M P F R E E -

 4201 GGCGAAAACCTCTCAAGGATCTTACCGTGTGAGATCCAGTTCGATGTAACCCACTCGTG
 -----+-----+-----+-----+-----+-----+-----+ 4260
 CCGCTTTGAGAGTTCCTAGAATGGCGACAACCTCTAGGTCAAGCTACATTGGGTGAGCAC
 P R F S E L I K G S N L D L E I Y G V R -

 4261 CACCCAACCTGATCTTCAGCATCTTTTACTTTTACCAGCGTTTCTGGGTGAGCAAAAACAG
 -----+-----+-----+-----+-----+-----+-----+ 4320
 GTGGGTGACTAGAAAGTCGTAGAAAATGAAAGTGGTCGCAAAGACCCACTCGTTTTTGTG
 A G L Q D E A D K V K V L T E P H A F V -

 4321 GAAGGCAAAATGCCGCAAAAAGGGAATAAGGGCGACACGGAAATGTGAATACTCATA
 -----+-----+-----+-----+-----+-----+-----+ 4380
 CTTCCGTTTTACGGCGTTTTTCCCTTATTCCCGCTGTGCCTTTACAACCTTATGAGTATG
 P L C F A A F F P I L A V R F H Q I S M -

 4381 TCTTCCTTTTTCAATATATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACA
 -----+-----+-----+-----+-----+-----+-----+ 4440
 AGAAGGAAAAGTTATAATAACTTCGTAAATAGTCCCAATAACAGAGTACTCGCCTATGT

 4441 TATTTGAATGTATTTAGAAAAATAACAAATAGGGGTCCGCGCACATTTCCCCGAAAAG
 -----+-----+-----+-----+-----+-----+-----+ 4500
 ATAAACTTACATAAATCTTTTTATTTGTTTATCCCCAAGCGCGGTGTAAGGGGCTTTTC
 TGCCACCTGACGCGCCTGTAGCGGCGATTAAGCGCGGCGGTGTGGTTACGCGCA
 4501 -----+-----+-----+-----+-----+-----+-----+ 4560
 ACGGTGGACTGCGCGGACATCGCCGCGTAATTCGCGCGCCACACCACCAATGCGCGT

 4561 GCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTGCTTTCTTCCCTTCT
 -----+-----+-----+-----+-----+-----+-----+ 4620
 CGCACTGGCGATGTGAACGGTTCGCGGGATCGCGGCGAGGAAAGCGAAAGAAGGAAGGA

 4621 TTCTGCCACGTTCCCGGCTTTCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGT
 -----+-----+-----+-----+-----+-----+-----+ 4680
 AAGAGCGGTGCAAGCGGCCGAAAGGGGCGAGTTTCGAGATTTAGCCCCGAGGGAAATCCA

 4681 TCCGATTTAGTGCTTTACGGCACCTCGACCCAAAAAAGTATTAGGGTGATGGTTCAC
 -----+-----+-----+-----+-----+-----+-----+ 4740
 AGGCTAAATCACGAAATGCCGTGGAGCTGGGGTTTTTTGAACTAATCCCACTACCAAGTG

 4741 GTAGTGGGCCATCGCCCTGATAGACGGTTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCT
 -----+-----+-----+-----+-----+-----+-----+ 4800
 CATCACCCGGTAGCGGGACTATCTGCCAAAAAGCGGGAAACTGCAACCTCAGGTGCAAGA

 4801 TTAATAGTGGACTCTTGTTCCAAAGTGAACAACACTCAACCTATCTCGGTCTATTCTT
 -----+-----+-----+-----+-----+-----+-----+ 4860
 AATTATCACCTGAGAACAAGTTTTGACCTTGTGTGAGTTGGGATAGAGCCAGATAAGAA

Figure 4 (cont'd)

11/12

```

TTGATTTATAAGGGATTTTGCCGATTCGGCCTATTGGTTAAAAAATGAGCTGATTTAAC
4861 -----+-----+-----+-----+-----+-----+-----+ 4920
AACTAAATATTCCCTAAAACGGCTAAAGCCGGATAACCAATTTTTTACTCGACTAAATTG

AAAAATTTAACGCGAATTTAACAAAATATTAACGCTTACAATTTGCCATTGCCATTCA
4921 -----+-----+-----+-----+-----+-----+-----+ 4980
TTTTTAAATGCGCTTAAAATTGTTTTATAATTGCCAATGTTAAACGGTAAGCGGTAAGT

                                HRGB4
GGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCCCA
4981 -----+-----+-----+-----+-----+-----+-----+ 5040
CCGACGCGTTGACAACCCTTCCCCTAGCCACGCCCGGAGAAGCGATAATGCGGTCCGGT

                                NotI
                                |
                                v
                                HRGB3
AGCTACCATGATAAGTAAGTAATATTAAGGTACGGGAGGTACTTGGAGCGGCCGCAATAA
5041 -----+-----+-----+-----+-----+-----+-----+ 5100
TCGATGGTACTATTTCATTCATTATAATTCATGCCCTCCATGAACCTCGCCGGCGTTATT

AATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGATAGTACTAAC
5101 -----+-----+-----+-----+-----+-----+-----+ 5160
TTATAGAAAATAAAAGTAATGTAGACACACAACCAAAAAACACACTTAGCTATCATGATTG

                                HRGB2
ATACGCTCTCCATCAAACAAAACGAAACAAAACAAACTAGCAAATAGGCTGTCCCCAG
5161 -----+-----+-----+-----+-----+-----+-----+ 5220
TATGCGAGAGGTAGTTTTGTTTTGCTTTGTTTTGTTTGTGATCGTTTTTATCCGACAGGGGTC

                                HRGB1
TGCAAGTGCAGGTGCCAGAACATTTCTCTATCGATA
5221 -----+-----+-----+-----+-----+-----+ 5256
ACGTTACAGTCCACGGTCTTGTAAGAGATAGCTAT
    
```

Figure 4 (cont'd)

12/12

Enzymes that do cut and were not excluded:

<u>AarI</u>	<u>Acc65I</u>	<u>AfeI</u>	<u>AhdI</u>	<u>AloI</u>	<u>AlwNI</u>	<u>AseI</u>	<u>BamHI</u>
<u>BmgBI</u>	<u>BsaAI</u>	<u>BsaHI</u>	<u>BsaXI</u>	<u>BseRI</u>	<u>BseYI</u>	<u>BssHI</u>	<u>BstBI</u>
<u>EstEII</u>	<u>DraI</u>	<u>EcoRI</u>	<u>FalI</u>	<u>KpnI</u>	<u>MluI</u>	<u>NotI</u>	<u>PfoI</u>
<u>PstI</u>	<u>SacII</u>	<u>SapI</u>	<u>SmaI</u>	<u>XhoI</u>	<u>XmaI</u>	<u>XmnI</u>	

Enzymes that do not cut:

<u>AatII</u>	<u>AflII</u>	<u>AgeI</u>	<u>AlfI</u>	<u>ApaI</u>	<u>AscI</u>	<u>AsiSI</u>	<u>AvrII</u>
<u>BaeI</u>	<u>BbeI</u>	<u>BbvCI</u>	<u>BclI</u>	<u>BfrBI</u>	<u>BlpI</u>	<u>BmtI</u>	<u>BplI</u>
<u>BsiWI</u>	<u>BsmBI</u>	<u>BspEI</u>	<u>BstAPI</u>	<u>BstXI</u>	<u>BstZ17I</u>	<u>Bsu36I</u>	<u>CspCI</u>
<u>EcoNI</u>	<u>EcoO109I</u>	<u>EcoRV</u>	<u>FspAI</u>	<u>KasI</u>	<u>MscI</u>	<u>NarI</u>	<u>NdeI</u>
<u>NheI</u>	<u>NruI</u>	<u>NsiI</u>	<u>PacI</u>	<u>PasI</u>	<u>PflMI</u>	<u>PmeI</u>	<u>PmlI</u>
<u>PpuMI</u>	<u>PspOMI</u>	<u>PspXI</u>	<u>PsrI</u>	<u>PvuII</u>	<u>RsrII</u>	<u>SanDI</u>	<u>SbfI</u>
<u>SexAI</u>	<u>SfiI</u>	<u>SfoI</u>	<u>SgrAI</u>	<u>SnaBI</u>	<u>SpeI</u>	<u>SrfI</u>	<u>StuI</u>
<u>SwaI</u>	<u>Tth111I</u>	<u>XbaI</u>	<u>XcmI</u>	<u>ZraI</u>			

Enzymes excluded; MinCuts: 1 MaxCuts: 1

<u>AccI</u>	<u>AclI</u>	<u>AcuI</u>	<u>AflIII</u>	<u>AleI</u>	<u>AloI</u>	<u>ApalI</u>	<u>ApoI</u>
<u>AvaI</u>	<u>BanI</u>	<u>BanII</u>	<u>BbsI</u>	<u>BcgI</u>	<u>BcgI</u>	<u>BciVI</u>	<u>BglI</u>
<u>BglII</u>	<u>Bme1580I</u>	<u>BmrI</u>	<u>BpmI</u>	<u>Bpu10I</u>	<u>BpuEI</u>	<u>BsaI</u>	<u>BsaBI</u>
<u>BsaWI</u>	<u>BsaXI</u>	<u>BsqI</u>	<u>BsiEI</u>	<u>BsiHKAI</u>	<u>BsmI</u>	<u>Bsp1286I</u>	<u>BspHI</u>
<u>BspMI</u>	<u>BsrBI</u>	<u>BsrDI</u>	<u>BsrFI</u>	<u>BsrGI</u>	<u>BssSI</u>	<u>BstYI</u>	<u>BtgI</u>
<u>BtgZI</u>	<u>BtsI</u>	<u>ClaI</u>	<u>DraI</u>	<u>DrdI</u>	<u>EaeI</u>	<u>EagI</u>	<u>EarI</u>
<u>EciI</u>	<u>EcoICRI</u>	<u>Eco57MI</u>	<u>FseI</u>	<u>FspI</u>	<u>HaeII</u>	<u>Hin4I</u>	<u>Hin4I</u>
<u>HincII</u>	<u>HindIII</u>	<u>HpaI</u>	<u>MfeI</u>	<u>MmeI</u>	<u>MslI</u>	<u>MspA1I</u>	<u>NaeI</u>
<u>NcoI</u>	<u>NgomIV</u>	<u>NspI</u>	<u>PciI</u>	<u>PpiI</u>	<u>PpiI</u>	<u>PshAI</u>	<u>PsiI</u>
<u>PvuI</u>	<u>SacI</u>	<u>SalI</u>	<u>Scal</u>	<u>SfcI</u>	<u>SmlI</u>	<u>SphI</u>	<u>SspI</u>
<u>StyI</u>	<u>TaqII</u>	<u>TaqII</u>	<u>TatI</u>				

Figure 5

SEQUENCE LISTING

<110> THE SCRIPPS RESEARCH INSTITUTE
 MAURO, Vincent P.
 EDELMAN, Gerald M.
 ZHOU, Wei

<120> TRANSLATION ENHANCER-ELEMENT DEPENDENT VECTOR SYSTEMS

<130> SCRIP1760-1WO

<150> US 60/711,149

<151> 2005-08-24

<160> 3

<170> PatentIn version 3.3

<210> 1

<211> 5256

<212> DNA

<213> Artificial sequence

<220>

<223> Synthetic construct

<220>

<221> misc_feature

<222> (190)..(207)

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

<220>

<221> CDS

<222> (214)..(897)

<220>

<221> CDS

<222> (1425)..(2141)

<400> 1

ggtaccgagc tcaagcttcg gagtactgtc ctccgagcgg agtactgtcc tccgagcggga 60

gtactgtcct ccgagcggag tactgtcctc cgttcgaaag cggaccctg gcaggaggaa 120

ggtcagcaga gctgctgata agagccgtat aaagagggtt ccgctcgcaa agatctatgg 180

catgaattcn nnnnnnnnnn nnnnnnngga tcc atg gtg aag ctg tct tct atc 234
 Met Val Lys Leu Ser Ser Ile
 1 5

gaa caa gca tgc gat att tgc cga ctt aaa aag ctc aag tgc tcc aaa 282
 Glu Gln Ala Cys Asp Ile Cys Arg Leu Lys Lys Leu Lys Cys Ser Lys
 10 15 20

gaa aaa ccg aag tgc gcc aag tgt ctg aag aac aac tgg gag tgt cgc 330
 Glu Lys Pro Lys Cys Ala Lys Cys Leu Lys Asn Asn Trp Glu Cys Arg
 25 30 35

tac tct ccc aaa acc aaa agg tct ccg ctg act agg gca cat ctg aca 378
 Tyr Ser Pro Lys Thr Lys Arg Ser Pro Leu Thr Arg Ala His Leu Thr

40	45	50	55	
gaa gtg gaa tca agg cta gaa aga ctg gaa cag cta ttt cta ctg att				426
Glu Val Glu Ser Arg Leu Glu Arg Leu Glu Gln Leu Phe Leu Leu Ile				
	60	65	70	
ttt cct cga gaa gac ctt gac atg att ttg aaa atg gat tct tta cag				474
Phe Pro Arg Glu Asp Leu Asp Met Ile Leu Lys Met Asp Ser Leu Gln				
	75	80	85	
gat ata aaa gca ttg tta aca gga tta ttt gta caa gat aat gtg aat				522
Asp Ile Lys Ala Leu Leu Thr Gly Leu Phe Val Gln Asp Asn Val Asn				
	90	95	100	
aaa gat gcc gtc aca gat aga ttg gct tca gtg gag act gat atg cct				570
Lys Asp Ala Val Thr Asp Arg Leu Ala Ser Val Glu Thr Asp Met Pro				
	105	110	115	
cta aca ttg aga cag cat aga ata agt gcg aca tca tca tcg gaa gag				618
Leu Thr Leu Arg Gln His Arg Ile Ser Ala Thr Ser Ser Ser Glu Glu				
	120	125	130	135
agt agt aac aaa ggt caa aga cag ttg act gta tcg att aaa gtc gcc				666
Ser Ser Asn Lys Gly Gln Arg Gln Leu Thr Val Ser Ile Lys Val Ala				
	140	145	150	
ccc ccg acc gat gtc agc ctg ggg gac gag ctc cac tta gac ggc gag				714
Pro Pro Thr Asp Val Ser Leu Gly Asp Glu Leu His Leu Asp Gly Glu				
	155	160	165	
gac gtg gcg atg gcg cat gcc gac gcg cta gac gat ttc gat ctg gac				762
Asp Val Ala Met Ala His Ala Asp Ala Leu Asp Asp Phe Asp Leu Asp				
	170	175	180	
atg ttg ggg gac ggg gat tcc ccg ggg ccg gga ttt acc ccc cac gac				810
Met Leu Gly Asp Gly Asp Ser Pro Gly Pro Gly Phe Thr Pro His Asp				
	185	190	195	
tcc gcc ccc tac ggc gct ctg gat atg gcc gac ttc gag ttt gag cag				858
Ser Ala Pro Tyr Gly Ala Leu Asp Met Ala Asp Phe Glu Phe Glu Gln				
	200	205	210	215
atg ttt acc gat gcc ctt gga att gac gag tac ggt ggg taaccgcggg				907
Met Phe Thr Asp Ala Leu Gly Ile Asp Glu Tyr Gly Gly				
	220	225		
ctagagtcgg ggcggccggc cgcttcgagc agacatgata agatacattg atgagtttgg				967
acaaaccaca actagaatgc agtgaaaaaa atgctttatt tgtgaaattt gtgatgctat				1027
tgctttattt gtaaccatta taagctgcaa taaacaagtt aacaacaaca attgcattca				1087
ttttatgttt caggttcagg gggaggtgtg ggaggttttt taaagcaagt aaaacctcta				1147
caaatgtggt aaaatcgata aggatcgatc cgtcgagatc tgcgatctaa gtaagcttgg				1207
ctgcaggtcg acggatcgga gtactgtcct ccgagcggag tactgtcctc cgagcggagt				1267
actgtcctcc gagcggagta ctgtcctccg agcggacccc tggcaggagg aaggtcagca				1327
gagctgctga taagagccgt ataaagaggg ttccgctcat ggcaaggggc agtggctctcg				1387

ggatctgagc ttggcattcc ggtactgttg gtaaacc atg gtg agc aag ggc gag 1442
 Met Val Ser Lys Gly Glu
 230

gag ctg ttc acc ggg gtg gtg ccc atc ctg gtc gag ctg gac ggc gac 1490
 Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val Glu Leu Asp Gly Asp
 235 240 245 250

gta aac ggc cac aag ttc agc gtg tcc ggc gag ggc gag ggc gat gcc 1538
 Val Asn Gly His Lys Phe Ser Val Ser Gly Glu Gly Glu Gly Asp Ala
 255 260 265

acc tac ggc aag ctg acc ctg aag ttc atc tgc acc acc ggc aag ctg 1586
 Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys Thr Thr Gly Lys Leu
 270 275 280

ccc gtg ccc tgg ccc acc ctc gtg acc acc ctg acc tac ggc gtg cag 1634
 Pro Val Pro Trp Pro Thr Leu Val Thr Thr Leu Thr Tyr Gly Val Gln
 285 290 295

tgc ttc agc cgc tac ccc gac cac atg aag cag cac gac ttc ttc aag 1682
 Cys Phe Ser Arg Tyr Pro Asp His Met Lys Gln His Asp Phe Phe Lys
 300 305 310

tcc gcc atg ccc gaa ggc tac gtc cag gag cgc acc atc ttc ttc aag 1730
 Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys
 315 320 325 330

gac gac ggc aac tac aag acc cgc gcc gag gtg aag ttc gag ggc gac 1778
 Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp
 335 340 345

acc ctg gtg aac cgc atc gag ctg aag ggc atc gac ttc aag gag gac 1826
 Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp
 350 355 360

ggc aac atc ctg ggg cac aag ctg gag tac aac tac aac agc cac aac 1874
 Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn
 365 370 375

gtc tat atc atg gcc gac aag cag aag aac ggc atc aag gtg aac ttc 1922
 Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly Ile Lys Val Asn Phe
 380 385 390

aag atc cgc cac aac atc gag gac ggc agc gtg cag ctc gcc gac cac 1970
 Lys Ile Arg His Asn Ile Glu Asp Gly Ser Val Gln Leu Ala Asp His
 395 400 405 410

tac cag cag aac acc ccc atc ggc gac ggc ccc gtg ctg ctg ccc gac 2018
 Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp
 415 420 425

aac cac tac ctg agc acc cag tcc gcc ctg agc aaa gac ccc aac gag 2066
 Asn His Tyr Leu Ser Thr Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu
 430 435 440

aag cgc gat cac atg gtc ctg ctg gag ttc gtg acc gcc gcc ggg atc 2114
 Lys Arg Asp His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile
 445 450 455

act ctc ggc atg gac gag ctg tac aag taacgcggtg actaagctat	2161
Thr Leu Gly Met Asp Glu Leu Tyr Lys	
460	465
ggcgcgcact aggggctaga gtcggggcgg ccggccgctt cgagcagaca tgataagata	2221
cattgatgag tttggacaaa ccacaactag aatgcagtga aaaaaatgct ttatttgtga	2281
aatttgtgat gctattgctt tatttghtaac cattataagc tgcaataaac aagttaacaa	2341
caacaattgc attcatttta tgtttcaggt tcagggggag gtgtgggagg ttttttaaag	2401
caagtaaac ctctacaaat gtggtaaaat cgataaggat cgatccgtcg accgatgccc	2461
ttgagagcct tcaaccagcag cagctccttc cggggggcgc ggggcatgac tatcgtcgcc	2521
gcacttatga ctgtcttctt tatcatgcaa ctcgtaggac aggtgccggc agcgtctctc	2581
cgcttcctcg ctactgact cgctgogctc ggtcgttcgg ctgcggcgag cggtatcagc	2641
tcactcaaag gcggtaatac ggttatccac agaatcaggg gataacgcag gaaagaacat	2701
gtgagcaaaa ggccagcaaa aggccaggaa ccgtaaaaag gccgcggtgc tggcgttttt	2761
ccataggctc cgccccctg acgagcatca caaaaatcga cgctcaagtc agagggtggc	2821
aaaccgaca ggactataaa gataccaggc gtttccccct ggaagctccc tcgtgcgctc	2881
tcctgttccg accctgccgc ttaccggata cctgtccgcc tttctccctt cgggaagcgt	2941
ggcgccttct catagctcac gctgtaggta tctcagttcg gtgtaggctc ttcgctccaa	3001
gctgggctgt gtgcacgaac cccccgttca gcccgaccgc tgcgccttat ccggtaaacta	3061
tcgtcttgag tccaaccggg taagacaoga cttatcgcca ctggcagcag cactggtaa	3121
caggattagc agagcgagggt atgtaggcgg tgctacagag ttcttgaaagt ggtggcctaa	3181
ctacggctac actagaagaa cagtatttgg tatctgcgct ctgctgaagc cagttacctt	3241
cggaaaaaga gttggtagct cttgatccgg caaacaacc accgctggta gcggtggttt	3301
ttttgtttgc aagcagcaga ttacgcgcag aaaaaaagga tctcaagaag atcctttgat	3361
cttttctacg gggctctgacg ctcagtgga cgaaaactca cgttaaggga ttttggcat	3421
gagattatca aaaaggatct tcacctagat ctttttaaat taaaaatgaa gttttaaatc	3481
aatctaaagt atatatgagt aaacttggtc tgacagttac caatgcttaa tcagtgaggc	3541
acctatctca gcgatctgtc tatttoggte atccatagtt gcctgactcc ccgtcgtgta	3601
gataactacg atacgggagg gcttaccatc tggccccagt gctgcaatga taccgcgaga	3661
cccacgctca ccggctccag atttatcagc aataaaccag ccagccggaa gggccgagcg	3721
cagaagtggc cctgcaactt tatccgcctc catccagctc attaattggt gccgggaagc	3781
tagagtaagt agttcgccag ttaatagttt gcgcaacggt gttgccattg ctacaggcat	3841
cgtgggtgtca cgctcgtcgt ttgggtatggc ttcattcagc tccggttccc aacgatcaag	3901

gogagttaca tgatccccc a tgttgtgcaa aaaagcgggt agctccttcg gtccctccgat 3961
cgttgtcaga agtaagttgg ccgcagtggt atcactcatg gttatggcag cactgcataa 4021
ttctcttact gtcatgccat ccgtaagatg cttttctgtg actgggtgagt actcaaccaa 4081
gtcattctga gaatagtga tgcggcgacc gagttgctct tgccccggcg caatacggga 4141
taataccgcg ccacatagca gaactttaaa agtgctcatc attggaaaac gttcttcggg 4201
gcgaaaactc tcaaggatct taccgctggt gagatccagt tcgatgtaac cactcgtgc 4261
acccaactga tottcagcat cttttacttt caccagcgtt tctgggtgag caaaaacagg 4321
aaggcaaaat gccgcaaaaa agggaataag ggcgacacgg aaatggtgaa tactcact 4381
cttccttttt caatattatt gaagcattta tcaggggtat tgtctcatga gcggatacat 4441
at ttgaaatgt at tttagaaaa ataaacaaat aggggttccg cgcacatttc cccgaaaagt 4501
gccacctgac gcgccctgta gcggcgcaat aagcgcggcg ggtgtggtgg ttacgcgcag 4561
cgtgaccgct acacttgcca gcgccctagc gcccgctcct ttcgctttct tcccttcctt 4621
tctcgccacg ttcgcggct ttccccgtca agctctaaat cgggggctcc ctttaggggt 4681
ccgatttagt gctttacggc acctcgacct caaaaaactt gattagggtg atggttcacg 4741
tagtgggcca tcgccctgat agacggtttt tcgcccttg acgttggagt ccacgttctt 4801
taatagtgga ctcttgttcc aaactggaac aacactcaac cctatctcgg tctattcttt 4861
tgatttataa gggattttgc cgatttcggc ctattggtta aaaaatgagc tgatttaaca 4921
aaaatttaac gcgaatttta acaaaatatt aacgcttaca atttgccatt cgccattcag 4981
gctgcgcaac tgttgggaag ggcgatcggg gcgggcctct tcgctattac gccagcccaa 5041
gctaccatga taagtaagta atattaaggt acgggaggta cttggagcgg ccgcaataaa 5101
atatctttat tttcattaca tctgtgtggt ggtttttgt gtgaatcgat agtactaaca 5161
tacgctctcc atcaaaacaa aacgaaacaa aacaaactag caaaataggc tgtccccagt 5221
gcaagtgcag gtgccagaac atttctctat cgata 5256

<210> 2
<211> 467
<212> PRT
<213> Artificial sequence

<220>
<223> Synthetic Construct

<400> 2

Met Val Lys Leu Ser Ser Ile Glu Gln Ala Cys Asp Ile Cys Arg Leu
1 5 10 15

Lys Lys Leu Lys Cys Ser Lys Glu Lys Pro Lys Cys Ala Lys Cys Leu
 20 25 30

Lys Asn Asn Trp Glu Cys Arg Tyr Ser Pro Lys Thr Lys Arg Ser Pro
 35 40 45

Leu Thr Arg Ala His Leu Thr Glu Val Glu Ser Arg Leu Glu Arg Leu
 50 55 60

Glu Gln Leu Phe Leu Leu Ile Phe Pro Arg Glu Asp Leu Asp Met Ile
 65 70 75 80

Leu Lys Met Asp Ser Leu Gln Asp Ile Lys Ala Leu Leu Thr Gly Leu
 85 90 95

Phe Val Gln Asp Asn Val Asn Lys Asp Ala Val Thr Asp Arg Leu Ala
 100 105 110

Ser Val Glu Thr Asp Met Pro Leu Thr Leu Arg Gln His Arg Ile Ser
 115 120 125

Ala Thr Ser Ser Ser Glu Glu Ser Ser Asn Lys Gly Gln Arg Gln Leu
 130 135 140

Thr Val Ser Ile Lys Val Ala Pro Pro Thr Asp Val Ser Leu Gly Asp
 145 150 155 160

Glu Leu His Leu Asp Gly Glu Asp Val Ala Met Ala His Ala Asp Ala
 165 170 175

Leu Asp Asp Phe Asp Leu Asp Met Leu Gly Asp Gly Asp Ser Pro Gly
 180 185 190

Pro Gly Phe Thr Pro His Asp Ser Ala Pro Tyr Gly Ala Leu Asp Met
 195 200 205

Ala Asp Phe Glu Phe Glu Gln Met Phe Thr Asp Ala Leu Gly Ile Asp
 210 215 220

Glu Tyr Gly Gly Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val
 225 230 235 240

Val Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe
 245 250 255

Ser Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr
260 265 270

Leu Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr
275 280 285

Leu Val Thr Thr Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro
290 295 300

Asp His Met Lys Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly
305 310 315 320

Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys
325 330 335

Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile
340 345 350

Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His
355 360 365

Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp
370 375 380

Lys Gln Lys Asn Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile
385 390 395 400

Glu Asp Gly Ser Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro
405 410 415

Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr
420 425 430

Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val
435 440 445

Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu
450 455 460

Leu Tyr Lys
465

<210> 3
<211> 286
<212> PRT
<213> Artificial sequence

<220>

<223> Synthetic construct

<400> 3

Met Ser Ile Gln His Phe Arg Val Ala Leu Ile Pro Phe Phe Ala Ala
 1 5 10 15

Phe Cys Leu Pro Val Phe Ala His Pro Glu Thr Leu Val Lys Val Lys
 20 25 30

Asp Ala Glu Asp Gln Leu Gly Ala Arg Val Gly Tyr Ile Glu Leu Asp
 35 40 45

Leu Asn Ser Gly Lys Ile Leu Glu Ser Phe Arg Pro Glu Glu Arg Phe
 50 55 60

Pro Met Met Ser Thr Phe Lys Val Leu Leu Cys Gly Ala Val Leu Ser
 65 70 75 80

Arg Ile Asp Ala Gly Gln Glu Gln Leu Gly Arg Arg Ile His Tyr Ser
 85 90 95

Gln Asn Asp Leu Val Glu Tyr Ser Pro Val Thr Glu Lys His Leu Thr
 100 105 110

Asp Gly Met Thr Val Arg Glu Leu Cys Ser Ala Ala Ile Thr Met Ser
 115 120 125

Asp Asn Thr Ala Ala Asn Leu Leu Leu Thr Thr Ile Gly Gly Pro Lys
 130 135 140

Glu Leu Thr Ala Phe Leu His Asn Met Gly Asp His Val Thr Arg Leu
 145 150 155 160

Asp Arg Trp Glu Pro Glu Leu Asn Glu Ala Ile Pro Asn Asp Glu Arg
 165 170 175

Asp Thr Thr Met Pro Val Ala Met Ala Thr Thr Leu Arg Lys Leu Leu
 180 185 190

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

Met Glu Ala Asp Lys Val Ala Gly Pro Leu Leu Arg Ser Ala Leu Pro
 210 215 220

Ala Gly Trp Phe Ile Ala Asp Lys Ser Gly Ala Gly Glu Arg Gly Ser
225 230 235 240

Arg Gly Ile Ile Ala Ala Leu Gly Pro Asp Gly Lys Pro Ser Arg Ile
245 250 255

Val Val Ile Tyr Thr Thr Gly Ser Gln Ala Thr Met Asp Glu Arg Asn
260 265 270

Arg Gln Ile Ala Glu Ile Gly Ala Ser Leu Ile Lys His Trp
275 280 285