siRNA-MEDIATED GENE SILENCING WITH VIRAL VECTORS

Claim of Priority

This is a continuation of application U.S. Application Serial No. 10/430,351 filed on May 2, 2003, which is a continuation of U.S. Application Serial No. 10/322,086 filed on December 17, 2002, which is a continuation-in-part application of U.S. Application Serial No. 10/212,322, filed August 5, 2002.

Background of the Invention

Double-stranded RNA (dsRNA) can induce sequence-specific posttranscriptional gene silencing in many organisms by a process known as RNA interference (RNAi). However, in mammalian cells, dsRNA that is 30 base pairs or longer can induce sequence-nonspecific responses that trigger a shut-down of protein synthesis. Recent work suggests that RNA fragments are the sequence-specific mediators of RNAi (Elbashir *et al.*, 2001). Interference of gene expression by these small interfering RNA (siRNA) is now recognized as a naturally occurring strategy for silencing genes in *C. elegans*, *Drosophila*, plants, and in mouse embryonic stem cells, oocytes and early embryos (Cogoni *et al.*, 1994; Baulcombe, 1996; Kennerdell, 1998; Timmons, 1998; Waterhouse *et al.*, 1998; Wianny and Zernicka-Goetz, 2000; Yang *et al.*, 2001; Svoboda *et al.*, 2000). In mammalian cell culture, a siRNA-mediated reduction in gene expression has been accomplished by transfecting cells with synthetic RNA oligonucleotides (Caplan *et al.*, 2001; Elbashir *et al.*, 2001). However, as Bass (2001) notes, various issues regarding the use of siRNA in mammalian cells have yet to be addressed, including effective delivery of siRNA to mammalian cells *in vivo*. Furthermore, if siRNA is to be utilized in *in vivo* therapy, it will be important in many cases to develop methods to express siRNA in tissues in vivo to achieve extended intracellular transcription of the siRNA.

Summary of the Invention

The present invention provides a viral vector containing an expression cassette, wherein the expression cassette contains a pol II promoter operably-

linked to a nucleic acid sequence encoding a small interfering RNA molecule (siRNA) targeted against a gene of interest.

The present invention also provides a viral vector containing an expression cassette, wherein the expression cassette contains a pol II promoter operably-linked to an isolated nucleic acid sequence encoding a first segment, a second segment located immediately 3' of the first segment, and a third segment located immediately 3' of the second segment, wherein the first and third segments are each less than 30 base pairs in length and each more than 10 base pairs in length, and wherein the sequence of the third segment is the complement of the sequence of the first segment, and wherein the isolated nucleic acid sequence functions as a small interfering RNA molecule (siRNA) targeted against a gene of interest.

The present invention further provides a method of reducing the expression of a gene product in a cell by contacting a cell with viral vector containing an expression cassette, wherein the expression cassette contains a pol II promoter operably-linked to an isolated nucleic acid sequence encoding a small interfering RNA molecule (siRNA) targeted against a gene, wherein expression from the targeted gene is reduced.

A method of reducing the expression of a gene product in a cell, comprising contacting a cell with viral vector comprising an expression cassette, wherein the expression cassette comprises an isolated nucleic acid sequence encoding a first segment, a second segment located immediately 3' of the first segment, and a third segment located immediately 3' of the second segment, wherein the first and third segments are each less than 30 base pairs in length and each more than 10 base pairs in length, and wherein the sequence of the third segment is the complement of the sequence of the first segment, and wherein the isolated nucleic acid sequence functions as a small interfering RNA molecule (siRNA) targeted against a gene of interest.

The present invention provides a method of treating a patient by administering to the patient a composition a viral vector described above.

Brief Description of the Figures

This patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

Figure 1. siRNA expressed from CMV promoter constructs and *in vitro* effects. (A) A cartoon of the expression plasmid used for expression of functional siRNA in cells. The CMV promoter was modified to allow close juxtaposition of the hairpin to the transcription initiation site, and a minimal polyadenylation signal containing cassette was constructed immediately 3' of the MCS (mCMV, modified CMV; mpA, minipA). (B, C) Fluorescence photomicrographs of HEK293 cells 72 h after transfection of pEGFPN1 and pCMVβgal (control), or pEGFPN1 and pmCMVsiGFPmpA, respectively. (D) Northern blot evaluation of transcripts harvested from pmCMVsiGFPmpA (lanes 3, 4) and pmCMVsiβgalmpA (lane 2) transfected HEK293 cells. Blots were probed with ^{32}P-labeled sense oligonucleotides. Antisense probes yielded similar results (not shown). Lane 1, ^{32}P-labeled RNA markers. AdsiGFP infected cells also possessed appropriately sized transcripts (not shown). (E) Northern blot for evaluation of target mRNA reduction by siRNA (upper panel). The internal control GAPDH is shown in the lower panel. HEK293 cells were transfected with pEGFPN1 and pmCMVsiGFPmpA, expressing siGFP, or plasmids expressing the control siRNA as indicated. pCMVeGFPx, which expresses siGFPx, contains a large poly(A) cassette from SV40 large T and an unmodified CMV promoter, in contrast to pmCMVsiGFPmpA shown in (A). (F) Western blot with anti-GFP antibodies of cell lysates harvested 72 h after transfection with pEGFPN1 and pCMVsiGFPmpA, or pEGFPN1 and pmCMVsiβglucmpA. (G, H) Fluorescence photomicrographs of HEK293 cells 72 h after transfection of pEGFPN1 and pCMVsiGFPx, or pEGFPN1 and pmCMVsiβglucmpA, respectively. (I, J) siRNA reduces expression from endogenous alleles. Recombinant adenoviruses were generated from pmCMVsiβglucmpA and pmCMVsiGFPmpA and purified. HeLa cells were infected with 25 infectious viruses/cell (MOI = 25) or mock-infected (control)

and cell lysates harvested 72 h later. (I) Northern blot for ß-glucuronidase mRNA levels in Adsißgluc and AdsiGFP transduced cells. GAPDH was used as an internal control for loading. (J) The concentration of β-glucuronidase activity in lysates quantified by a fluorometric assay. Stein, C.S. et al., *J. Virol.* **73**:3424-3429 (1999).

Figure 2. Viral vectors expressing siRNA reduce expression from transgenic and endogenous alleles *in vivo*. Recombinant adenovirus vectors were prepared from the siGFP and siβgluc shuttle plasmids described in Fig. 1. (A) Fluorescence microscopy reveals diminution of eGFP expression *in vivo*. In addition to the siRNA sequences in the E1 region of adenovirus, RFP expression cassettes in E3 facilitate localization of gene transfer. Representative photomicrographs of eGFP (left), RFP (middle), and merged images (right) of coronal sections from mice injected with adenoviruses expressing siGFP (top panels) or siβgluc (bottom panels) demonstrate siRNA specificity in eGFP transgenic mice striata after direct brain injection. (B) Full coronal brain sections (1 mm) harvested from AdsiGFP or Adsiβgluc injected mice were split into hemisections and both ipsilateral (il) and contralateral (cl) portions evaluated by western blot using antibodies to GFP. Actin was used as an internal control for each sample. (C) Tail vein injection of recombinant adenoviruses expressing siβgluc directed against mouse β-glucuronidase (AdsiMuβgluc) reduces endogenous β–glucuronidase RNA as determined by Northern blot in contrast to control-treated (Adsißgal) mice.

Figure 3. siGFP gene transfer reduces Q19-eGFP expression in cell lines. PC12 cells expressing the polyglutamine repeat Q19 fused to eGFP (eGFP-Q19) under tetracycline repression (A, bottom left) were washed and dox-free media added to allow eGFP-Q19 expression (A, top left). Adenoviruses were applied at the indicated multiplicity of infection (MOI) 3 days after dox removal. (A) eGFP fluorescence 3 days after adenovirus-mediated gene transfer of Adsiβgluc (top panels) or AdsiGFP (bottom panels). (B, C) Western blot analysis of cell lysates harvested 3 days after infection at the indicated MOIs demonstrate a dose-dependent decrease in GFP-Q19 protein levels. NV, no virus. Top lanes, eGFP-Q19. Bottom lanes, actin loading

controls. (D) Quantitation of eGFP fluorescence. Data represent mean total area fluorescence ± standard deviation in 4 low power fields/well (3 wells/plate).

Figure 4. siRNA mediated reduction of expanded polyglutamine protein levels and intracellular aggregates. PC12 cells expressing tet-repressible eGFP-Q80 fusion proteins were washed to remove doxycycline and adenovirus vectors expressing siRNA were applied 3 days later. (A-D) Representative punctate eGFP fluorescence of aggregates in mock-infected cells (A), or those infected with 100 MOI of Adsiβgluc (B), AdsiGFPx (C) or Adsiβgal (D). (E) Three days after infection of dox-free eGFP-Q80 PC12 cells with AdsiGFP, aggregate size and number are notably reduced. (F) Western blot analysis of eGFP-Q80 aggregates (arrowhead) and monomer (arrow) following Adsiβgluc or AdsiGFP infection at the indicated MOIs demonstrates dose dependent siGFP-mediated reduction of GFP-Q80 protein levels. (G) Quantification of the total area of fluorescent inclusions measured in 4 independent fields/well 3 days after virus was applied at the indicated MOIs. The data are mean ± standard deviation.

Detailed Description of the Invention

RNA interference is now established as an important biological strategy for gene silencing, but its application to mammalian cells has been limited by nonspecific inhibitory effects of long double-stranded RNA on translation. The present inventors have developed a viral mediated delivery mechanism that results in specific silencing of targeted genes through expression of small interfering RNA (siRNA). The inventors have establish proof of principle by markedly diminishing expression of exogenous and endogenous genes *in vitro* and *in vivo* in brain and liver, and further apply this novel strategy to a model system of a major class of neurodegenerative disorders, the polyglutamine diseases, to show reduced polyglutamine aggregation in cells. This viral mediated strategy is generally useful in reducing expression of target genes in order to model biological processes or to provide therapy for dominant human diseases.

Disclosed herein is a viral-mediated strategy that results in silencing of targeted genes *via* siRNA. Use of this strategy results in markedly diminished *in vitro* and *in vivo* expression of targeted genes. This viral-mediated strategy is

useful in reducing expression of targeted genes in order to model biological processes or to provide therapy for human diseases. For example, this strategy can be applied to a major class of neurodegenerative disorders, the polyglutamine diseases, as is demonstrated by the reduction of polyglutamine aggregation in cells following application of the strategy.

To accomplish intracellular expression of the therapeutic siRNA, an RNA molecule is constructed containing a hairpin sequence (such as a 21-bp hairpin) representing sequences directed against the gene of interest. The siRNA, or a DNA sequence encoding the siRNA, is introduced to the target cell, such as a diseased brain cell. The siRNA reduces target mRNA and gene protein expression.

The construct encoding the therapeutic siRNA is configured such that the the hairpin is immediately contiguous to a pol II promoter. The pol II promoter used in a particular construct is selected from readily available pol II promoters known in the art, depending on whether regulatable, inducible, tissue or cell-specific expression of the siRNA is desired. The construct is introduced into the target cell, such as by injection, allowing for diminished target-gene expression in the cell. It was surprising that a pol II promoter would be effective. While small RNAs with extensive secondary structure are routinely made from Pol III promoters, there is no *a priori* reason to assume that small interfering RNAs could be expressed from pol II promoters. Pol III promoters terminate in a short stretch of Ts (5 or 6), leaving a very small 3' end and allowing stabilization of secondary structure. Polymerase II transcription extends well past the coding and polyadenylation regions, after which the transcript is cleaved. Two adenylation steps occur, leaving a transcript with a tail of up to 200 As. This string of As would of course completely destabilize any small, 21 base pair hairpin. Therefore, in addition to modifying the promoter to minimize sequences between the transcription start site and the hairpin (thereby stabilizing the hairpin), the inventors also extensively modified the polyadenylation sequence to test if a very short polyadenylation could occur. The results, which were not predicted from prior literature, showed that it could.

The present invention provides a viral vector comprising an expression cassette, wherein the expression cassette comprises an isolated nucleic acid

sequence encoding a small interfering RNA molecule (siRNA) targeted against a gene of interest. The siRNA may form hairpin structure comprising a duplex structure and a loop structure. The loop structure may contain from 4 to 10 nucleotides, such as 4, 5 or 6 nucleotides. The duplex is less than 30 nucleotides in length, such as from 19 to 25 nucleotides. The siRNA may further comprises an overhang region. Such an overhang may be a 3' overhang region or a 5' overhang region. The overhang region may be, for example, from 1 to 6 nucleotides in length. The expression cassette may further comprise a pol II promoter, as described herein. Examples of pol II promoters include regulatable promoters and constitutive promoters. For example, the promoter may be a CMV or RSV promoter. The expression cassette may further comprise a polyadenylation signal, such as a synthetic minimal polyadenylation signal. The nucleic acid sequence may further comprise a marker gene. The viral vector of the present invention may be an adenoviral, lentiviral, adeno-associated viral (AAV), poliovirus, herpes simplex virus (HSV) or murine Maloney-based viral vector. The gene of interest may be a gene associated with a condition amenable to siRNA therapy. Examples of such conditions include neurodegenerative diseases, such as a trinucleotide-repeat disease (e.g., polyglutamine repeat disease). Examples of these diseases include Huntington's disease or spinocerebellar ataxia. Alternatively, the gene of interest may encode a ligand for a chemokine involved in the migration of a cancer cell, or a chemokine receptor.

The present invention also provides a viral vector comprising an expression cassette, wherein the expression cassette comprises a pol II promoter operably-linked to an isolated nucleic acid sequence encoding a first segment, a second segment located immediately 3' of the first segment, and a third segment located immediately 3' of the second segment, wherein the first and third segments are each less than 30 base pairs in length and each more than 10 base pairs in length, and wherein the sequence of the third segment is the complement of the sequence of the first segment, and wherein the isolated nucleic acid sequence functions as a small interfering RNA molecule (siRNA) targeted against a gene of interest.

The present invention provides a method of reducing the expression of a gene product in a cell by contacting a cell with a viral vector described above. It also provides a method of treating a patient by administering to the patient a composition comprising a viral vector described above.

The present invention further provides a method of reducing the expression of a gene product in a cell, comprising contacting a cell with viral vector comprising an expression cassette, wherein the expression cassette comprises a pol II promoter operably-linked to an isolated nucleic acid sequence encoding a first segment, a second segment located immediately 3' of the first segment, and a third segment located immediately 3' of the second segment, wherein the first and third segments are each less than 30 base pairs in length and each more than 10 base pairs in length, and wherein the sequence of the third segment is the complement of the sequence of the first segment, and wherein the isolated nucleic acid sequence functions as a small interfering RNA molecule (siRNA) targeted against a gene of interest.

The present method also provides a method of treating a patient, comprising administering to the patient a composition comprising a viral vector, wherein the viral vector comprises an expression cassette, wherein the expression cassette comprises a pol II promoter operably-linked to an isolated nucleic acid sequence encoding a first segment, a second segment located immediately 3' of the first segment, and a third segment located immediately 3' of the second segment, wherein the first and third segments are each less than 30 base pairs in length and each more than 10 base pairs in length, and wherein the sequence of the third segment is the complement of the sequence of the first segment, and wherein the isolated nucleic acid sequence functions as a small interfering RNA molecule (siRNA) targeted against a gene of interest.

I. Definitions

The term "nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form, composed of monomers (nucleotides) containing a sugar, phosphate and a base that is either a purine or pyrimidine. Unless specifically limited, the term encompasses nucleic acids containing known analogs of natural nucleotides that have similar binding

properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also encompasses conservatively modified variants thereof (*e.g.*, degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al.*, (1991); Ohtsuka *et al.*, (1985); Rossolini *et al.*, (1994)).

A "nucleic acid fragment" is a portion of a given nucleic acid molecule. Deoxyribonucleic acid (DNA) in the majority of organisms is the genetic material while ribonucleic acid (RNA) is involved in the transfer of information contained within DNA into proteins.

The term "nucleotide sequence" refers to a polymer of DNA or RNA which can be single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases capable of incorporation into DNA or RNA polymers.

The terms “nucleic acid”, “nucleic acid molecule”, “nucleic acid fragment”, “nucleic acid sequence or segment”, or “polynucleotide” are used interchangeably and may also be used interchangeably with gene, cDNA, DNA and RNA encoded by a gene.

The invention encompasses isolated or substantially purified nucleic acid or protein compositions. In the context of the present invention, an "isolated" or "purified" DNA molecule or RNA molecule or an "isolated" or "purified" polypeptide is a DNA molecule, RNA molecule, or polypeptide that exists apart from its native environment and is therefore not a product of nature. An isolated DNA molecule, RNA molecule or polypeptide may exist in a purified form or may exist in a non-native environment such as, for example, a transgenic host cell. For example, an "isolated" or "purified" nucleic acid molecule or protein, or biologically active portion thereof, is substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. In one embodiment, an "isolated" nucleic acid is free of sequences that naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends

of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb, or 0.1 kb of nucleotide sequences that naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. A protein that is substantially free of cellular material includes preparations of protein or polypeptide having less than about 30%, 20%, 10%, or 5% (by dry weight) of contaminating protein. When the protein of the invention, or biologically active portion thereof, is recombinantly produced, preferably culture medium represents less than about 30%, 20%, 10%, or 5% (by dry weight) of chemical precursors or non-protein-of-interest chemicals. Fragments and variants of the disclosed nucleotide sequences and proteins or partial-length proteins encoded thereby are also encompassed by the present invention. By "fragment" or "portion" is meant a full length or less than full length of the nucleotide sequence encoding, or the amino acid sequence of, a polypeptide or protein.

The term "gene" is used broadly to refer to any segment of nucleic acid associated with a biological function. Thus, genes include coding sequences and/or the regulatory sequences required for their expression. For example, "gene" refers to a nucleic acid fragment that expresses mRNA, functional RNA, or specific protein, including regulatory sequences. "Genes" also include nonexpressed DNA segments that, for example, form recognition sequences for other proteins. "Genes" can be obtained from a variety of sources, including cloning from a source of interest or synthesizing from known or predicted sequence information, and may include sequences designed to have desired parameters.

"Naturally occurring" is used to describe an object that can be found in nature as distinct from being artificially produced. For example, a protein or nucleotide sequence present in an organism (including a virus), which can be isolated from a source in nature and which has not been intentionally modified by a person in the laboratory, is naturally occurring.

The term "chimeric" refers to a gene or DNA that contains 1) DNA sequences, including regulatory and coding sequences, that are not found together in nature, or 2) sequences encoding parts of proteins not naturally

adjoined, or 3) parts of promoters that are not naturally adjoined. Accordingly, a chimeric gene may include regulatory sequences and coding sequences that are derived from different sources, or include regulatory sequences and coding sequences derived from the same source, but arranged in a manner different from that found in nature.

A "transgene" refers to a gene that has been introduced into the genome by transformation. Transgenes include, for example, DNA that is either heterologous or homologous to the DNA of a particular cell to be transformed. Additionally, transgenes may include native genes inserted into a non-native organism, or chimeric genes.

The term "endogenous gene" refers to a native gene in its natural location in the genome of an organism.

A "foreign" gene refers to a gene not normally found in the host organism that has been introduced by gene transfer.

The terms "protein," "peptide" and "polypeptide" are used interchangeably herein.

A "variant" of a molecule is a sequence that is substantially similar to the sequence of the native molecule. For nucleotide sequences, variants include those sequences that, because of the degeneracy of the genetic code, encode the identical amino acid sequence of the native protein. Naturally occurring allelic variants such as these can be identified with the use of molecular biology techniques, as, for example, with polymerase chain reaction (PCR) and hybridization techniques. Variant nucleotide sequences also include synthetically derived nucleotide sequences, such as those generated, for example, by using site-directed mutagenesis, which encode the native protein, as well as those that encode a polypeptide having amino acid substitutions. Generally, nucleotide sequence variants of the invention will have at least 40%, 50%, 60%, to 70%, *e.g.*, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, to 79%, generally at least 80%, *e.g.*, 81%-84%, at least 85%, *e.g.*, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, to 98%, sequence identity to the native (endogenous) nucleotide sequence.

"Conservatively modified variations" of a particular nucleic acid sequence refers to those nucleic acid sequences that encode identical or

essentially identical amino acid sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given polypeptide. For instance, the codons CGT, CGC, CGA, CGG, AGA and AGG all encode the amino acid arginine. Thus, at every position where an arginine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded protein. Such nucleic acid variations are "silent variations," which are one species of "conservatively modified variations." Every nucleic acid sequence described herein that encodes a polypeptide also describes every possible silent variation, except where otherwise noted. One of skill in the art will recognize that each codon in a nucleic acid (except ATG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule by standard techniques. Accordingly, each “silent variation” of a nucleic acid that encodes a polypeptide is implicit in each described sequence.

“Recombinant DNA molecule” is a combination of DNA sequences that are joined together using recombinant DNA technology and procedures used to join together DNA sequences as described, for example, in Sambrook and Russell (2001).

The terms “heterologous gene”, “heterologous DNA sequence”, “exogenous DNA sequence”, “heterologous RNA sequence”, “exogenous RNA sequence” or “heterologous nucleic acid” each refer to a sequence that either originates from a source foreign to the particular host cell, or is from the same source but is modified from its original or native form. Thus, a heterologous gene in a host cell includes a gene that is endogenous to the particular host cell but has been modified through, for example, the use of DNA shuffling. The terms also include non-naturally occurring multiple copies of a naturally occurring DNA or RNA sequence. Thus, the terms refer to a DNA or RNA segment that is foreign or heterologous to the cell, or homologous to the cell but in a position within the host cell nucleic acid in which the element is not ordinarily found. Exogenous DNA segments are expressed to yield exogenous polypeptides.

A "homologous" DNA or RNA sequence is a sequence that is naturally associated with a host cell into which it is introduced.

"Wild-type" refers to the normal gene or organism found in nature.

"Genome" refers to the complete genetic material of an organism.

A "vector" is defined to include, *inter alia*, any viral vector, as well as any plasmid, cosmid, phage or binary vector in double or single stranded linear or circular form that may or may not be self transmissible or mobilizable, and that can transform prokaryotic or eukaryotic host either by integration into the cellular genome or exist extrachromosomally (*e.g.*, autonomous replicating plasmid with an origin of replication).

"Expression cassette" as used herein means a nucleic acid sequence capable of directing expression of a particular nucleotide sequence in an appropriate host cell, which may include a promoter operably linked to the nucleotide sequence of interest that may be operably linked to termination signals. It also may include sequences required for proper translation of the nucleotide sequence. The coding region usually codes for a protein of interest but may also code for a functional RNA of interest, for example an antisense RNA, a nontranslated RNA in the sense or antisense direction, or a siRNA. The expression cassette including the nucleotide sequence of interest may be chimeric. The expression cassette may also be one that is naturally occurring but has been obtained in a recombinant form useful for heterologous expression. The expression of the nucleotide sequence in the expression cassette may be under the control of a constitutive promoter or of an regulatable promoter that initiates transcription only when the host cell is exposed to some particular stimulus. In the case of a multicellular organism, the promoter can also be specific to a particular tissue or organ or stage of development.

Such expression cassettes can include a transcriptional initiation region linked to a nucleotide sequence of interest. Such an expression cassette is provided with a plurality of restriction sites for insertion of the gene of interest to be under the transcriptional regulation of the regulatory regions. The expression cassette may additionally contain selectable marker genes.

"Coding sequence" refers to a DNA or RNA sequence that codes for a specific amino acid sequence. It may constitute an "uninterrupted coding sequence", *i.e.*, lacking an intron, such as in a cDNA, or it may include one or more introns bounded by appropriate splice junctions. An "intron" is a sequence

of RNA that is contained in the primary transcript but is removed through cleavage and re-ligation of the RNA within the cell to create the mature mRNA that can be translated into a protein.

The term "open reading frame" (ORF) refers to the sequence between translation initiation and termination codons of a coding sequence. The terms "initiation codon" and "termination codon" refer to a unit of three adjacent nucleotides (a 'codon') in a coding sequence that specifies initiation and chain termination, respectively, of protein synthesis (mRNA translation).

"Functional RNA" refers to sense RNA, antisense RNA, ribozyme RNA, siRNA, or other RNA that may not be translated but yet has an effect on at least one cellular process.

The term "RNA transcript" refers to the product resulting from RNA polymerase catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA" (mRNA) refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a single- or a double-stranded DNA that is complementary to and derived from mRNA.

"Regulatory sequences" and "suitable regulatory sequences" each refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences include enhancers, promoters, translation leader sequences, introns, and polyadenylation signal sequences. They include natural and synthetic sequences as well as sequences that may be a combination of synthetic and natural sequences. As is noted above, the term "suitable regulatory sequences" is not limited to promoters. However, some suitable regulatory sequences useful in the present invention will include, but are not limited to constitutive promoters, tissue-specific promoters, development-specific promoters, regulatable promoters and viral promoters.

Examples of promoters that may be used in the present invention include CMV, RSV, polII and polIII promoters.

"5' non-coding sequence" refers to a nucleotide sequence located 5' (upstream) to the coding sequence. It is present in the fully processed mRNA upstream of the initiation codon and may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency (Turner *et al.*, 1995).

"3' non-coding sequence" refers to nucleotide sequences located 3' (downstream) to a coding sequence and may include polyadenylation signal sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor.

The term "translation leader sequence" refers to that DNA sequence portion of a gene between the promoter and coding sequence that is transcribed into RNA and is present in the fully processed mRNA upstream (5') of the translation start codon. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency.

The term "mature" protein refers to a post-translationally processed polypeptide without its signal peptide. "Precursor" protein refers to the primary product of translation of an mRNA. "Signal peptide" refers to the amino terminal extension of a polypeptide, which is translated in conjunction with the polypeptide forming a precursor peptide and which is required for its entrance into the secretory pathway. The term "signal sequence" refers to a nucleotide sequence that encodes the signal peptide.

"Promoter" refers to a nucleotide sequence, usually upstream (5') to its coding sequence, which directs and/or controls the expression of the coding sequence by providing the recognition for RNA polymerase and other factors required for proper transcription. "Promoter" includes a minimal promoter that is a short DNA sequence comprised of a TATA- box and other sequences that serve to specify the site of transcription initiation, to which regulatory elements are added for control of expression. "Promoter" also refers to a nucleotide sequence that includes a minimal promoter plus regulatory elements that is

capable of controlling the expression of a coding sequence or functional RNA. This type of promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an "enhancer" is a DNA sequence that can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue specificity of a promoter. It is capable of operating in both orientations (normal or flipped), and is capable of functioning even when moved either upstream or downstream from the promoter. Both enhancers and other upstream promoter elements bind sequence-specific DNA-binding proteins that mediate their effects. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even be comprised of synthetic DNA segments. A promoter may also contain DNA sequences that are involved in the binding of protein factors that control the effectiveness of transcription initiation in response to physiological or developmental conditions.

The "initiation site" is the position surrounding the first nucleotide that is part of the transcribed sequence, which is also defined as position +1. With respect to this site all other sequences of the gene and its controlling regions are numbered. Downstream sequences (*i.e.*, further protein encoding sequences in the 3' direction) are denominated positive, while upstream sequences (mostly of the controlling regions in the 5' direction) are denominated negative.

Promoter elements, particularly a TATA element, that are inactive or that have greatly reduced promoter activity in the absence of upstream activation are referred to as "minimal or core promoters." In the presence of a suitable transcription factor, the minimal promoter functions to permit transcription. A "minimal or core promoter" thus consists only of all basal elements needed for transcription initiation, *e.g.*, a TATA box and/or an initiator.

"Constitutive expression" refers to expression using a constitutive or regulated promoter. "Conditional" and "regulated expression" refer to expression controlled by a regulated promoter.

"Operably-linked" refers to the association of nucleic acid sequences on single nucleic acid fragment so that the function of one of the sequences is affected by another. For example, a regulatory DNA sequence is said to be

"operably linked to" or "associated with" a DNA sequence that codes for an RNA or a polypeptide if the two sequences are situated such that the regulatory DNA sequence affects expression of the coding DNA sequence (*i.e.*, that the coding sequence or functional RNA is under the transcriptional control of the promoter). Coding sequences can be operably-linked to regulatory sequences in sense or antisense orientation.

"Expression" refers to the transcription and/or translation of an endogenous gene, heterologous gene or nucleic acid segment, or a transgene in cells. For example, in the case of siRNA constructs, expression may refer to the transcription of the siRNA only. In addition, expression refers to the transcription and stable accumulation of sense (mRNA) or functional RNA. Expression may also refer to the production of protein.

"Altered levels" refers to the level of expression in transgenic cells or organisms that differs from that of normal or untransformed cells or organisms.

"Overexpression" refers to the level of expression in transgenic cells or organisms that exceeds levels of expression in normal or untransformed cells or organisms.

"Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of protein from an endogenous gene or a transgene.

"Transcription stop fragment" refers to nucleotide sequences that contain one or more regulatory signals, such as polyadenylation signal sequences, capable of terminating transcription. Examples include the 3' non-regulatory regions of genes encoding nopaline synthase and the small subunit of ribulose bisphosphate carboxylase.

"Translation stop fragment" refers to nucleotide sequences that contain one or more regulatory signals, such as one or more termination codons in all three frames, capable of terminating translation. Insertion of a translation stop fragment adjacent to or near the initiation codon at the 5' end of the coding sequence will result in no translation or improper translation. Excision of the translation stop fragment by site-specific recombination will leave a site-specific sequence in the coding sequence that does not interfere with proper translation using the initiation codon.

The terms "*cis*-acting sequence" and "*cis*-acting element" refer to DNA or RNA sequences whose functions require them to be on the same molecule. An example of a *cis*-acting sequence on the replicon is the viral replication origin.

The terms "*trans*-acting sequence" and "*trans*-acting element" refer to DNA or RNA sequences whose function does not require them to be on the same molecule.

"Chromosomally-integrated" refers to the integration of a foreign gene or nucleic acid construct into the host DNA by covalent bonds. Where genes are not "chromosomally integrated" they may be "transiently expressed." Transient expression of a gene refers to the expression of a gene that is not integrated into the host chromosome but functions independently, either as part of an autonomously replicating plasmid or expression cassette, for example, or as part of another biological system such as a virus.

The following terms are used to describe the sequence relationships between two or more nucleic acids or polynucleotides: (a) "reference sequence", (b) "comparison window", (c) "sequence identity", (d) "percentage of sequence identity", and (e) "substantial identity".

(a) As used herein, "reference sequence" is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset or the entirety of a specified sequence; for example, as a segment of a full-length cDNA or gene sequence, or the complete cDNA or gene sequence.

(b) As used herein, "comparison window" makes reference to a contiguous and specified segment of a polynucleotide sequence, wherein the polynucleotide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Generally, the comparison window is at least 20 contiguous nucleotides in length, and optionally can be 30, 40, 50, 100, or longer. Those of skill in the art understand that to avoid a high similarity to a reference sequence due to inclusion of gaps in the polynucleotide sequence a gap penalty is typically introduced and is subtracted from the number of matches.

Methods of alignment of sequences for comparison are well-known in the art. Thus, the determination of percent identity between any two sequences can be accomplished using a mathematical algorithm. Preferred, non-limiting examples of such mathematical algorithms are the algorithm of Myers and Miller (1988); the local homology algorithm of Smith *et al.* (1981); the homology alignment algorithm of Needleman and Wunsch (1970); the search-for-similarity-method of Pearson and Lipman (1988); the algorithm of Karlin and Altschul (1990), modified as in Karlin and Altschul (1993).

Computer implementations of these mathematical algorithms can be utilized for comparison of sequences to determine sequence identity. Such implementations include, but are not limited to: CLUSTAL in the PC/Gene program (available from Intelligenetics, Mountain View, California); the ALIGN program (Version 2.0) and GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Version 8 (available from Genetics Computer Group (GCG), 575 Science Drive, Madison, Wisconsin, USA). Alignments using these programs can be performed using the default parameters. The CLUSTAL program is well described by Higgins *et al.* (1988); Higgins *et al.* (1989); Corpet *et al.* (1988); Huang *et al.* (1992); and Pearson *et al.* (1994). The ALIGN program is based on the algorithm of Myers and Miller, *supra*. The BLAST programs of Altschul *et al.* (1990), are based on the algorithm of Karlin and Altschul *supra*.

Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold. These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty

score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when the cumulative alignment score falls off by the quantity X from its maximum achieved value, the cumulative score goes to zero or below due to the accumulation of one or more negative-scoring residue alignments, or the end of either sequence is reached.

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences. One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a test nucleic acid sequence is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid sequence to the reference nucleic acid sequence is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

To obtain gapped alignments for comparison purposes, Gapped BLAST (in BLAST 2.0) can be utilized as described in Altschul *et al.* (1997). Alternatively, PSI-BLAST (in BLAST 2.0) can be used to perform an iterated search that detects distant relationships between molecules. See Altschul *et al.*, *supra*. When utilizing BLAST, Gapped BLAST, PSI-BLAST, the default parameters of the respective programs (*e.g.* BLASTN for nucleotide sequences, BLASTX for proteins) can be used. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix. See http://www.ncbi.n1m.nih.gov. Alignment may also be performed manually by inspection.

For purposes of the present invention, comparison of nucleotide sequences for determination of percent sequence identity to the promoter sequences disclosed herein is preferably made using the BlastN program (version 1.4.7 or later) with its default parameters or any equivalent program.

By "equivalent program" is intended any sequence comparison program that, for any two sequences in question, generates an alignment having identical nucleotide or amino acid residue matches and an identical percent sequence identity when compared to the corresponding alignment generated by the preferred program.

(c) As used herein, "sequence identity" or "identity" in the context of two nucleic acid or polypeptide sequences makes reference to a specified percentage of residues in the two sequences that are the same when aligned for maximum correspondence over a specified comparison window, as measured by sequence comparison algorithms or by visual inspection. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (*e.g.*, charge or hydrophobicity) and therefore do not change the functional properties of the molecule. When sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences that differ by such conservative substitutions are said to have "sequence similarity" or "similarity." Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, *e.g.*, as implemented in the program PC/GENE (Intelligenetics, Mountain View, California).

(d) As used herein, "percentage of sequence identity" means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid

residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison, and multiplying the result by 100 to yield the percentage of sequence identity.

(e)(i) The term "substantial identity" of polynucleotide sequences means that a polynucleotide comprises a sequence that has at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, or 79%, preferably at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, or 89%, more preferably at least 90%, 91%, 92%, 93%, or 94%, and most preferably at least 95%, 96%, 97%, 98%, or 99% sequence identity, compared to a reference sequence using one of the alignment programs described using standard parameters. One of skill in the art will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning, and the like. Substantial identity of amino acid sequences for these purposes normally means sequence identity of at least 70%, more preferably at least 80%, 90%, and most preferably at least 95%.

Another indication that nucleotide sequences are substantially identical is if two molecules hybridize to each other under stringent conditions. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. However, stringent conditions encompass temperatures in the range of about 1°C to about 20°C, depending upon the desired degree of stringency as otherwise qualified herein. Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides they encode are substantially identical. This may occur, *e.g.*, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. One indication that two nucleic acid sequences are substantially identical is when the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the polypeptide encoded by the second nucleic acid.

(e)(ii) The term "substantial identity" in the context of a peptide indicates that a peptide comprises a sequence with at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, or 79%, preferably 80%, 81%, 82%, 83%, 84%, 85%,

86%, 87%, 88%, or 89%, more preferably at least 90%, 91%, 92%, 93%, or 94%, or even more preferably, 95%, 96%, 97%, 98% or 99%, sequence identity to the reference sequence over a specified comparison window. Preferably, optimal alignment is conducted using the homology alignment algorithm of Needleman and Wunsch (1970). An indication that two peptide sequences are substantially identical is that one peptide is immunologically reactive with antibodies raised against the second peptide. Thus, a peptide is substantially identical to a second peptide, for example, where the two peptides differ only by a conservative substitution.

For sequence comparison, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

As noted above, another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions. The phrase "hybridizing specifically to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex mixture (*e.g.*, total cellular) DNA or RNA. "Bind(s) substantially" refers to complementary hybridization between a probe nucleic acid and a target nucleic acid and embraces minor mismatches that can be accommodated by reducing the stringency of the hybridization media to achieve the desired detection of the target nucleic acid sequence.

"Stringent hybridization conditions" and "stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments such as Southern and Northern hybridizations are sequence dependent, and are different under different environmental parameters. Longer sequences hybridize specifically at higher temperatures. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Specificity is typically the function of post-

hybridization washes, the critical factors being the ionic strength and temperature of the final wash solution. For DNA-DNA hybrids, the T_m can be approximated from the equation of Meinkoth and Wahl (1984); T_m 81.5°C + 16.6 (log M) +0.41 (%GC) - 0.61 (% form) - 500/L; where M is the molarity of monovalent cations, %GC is the percentage of guanosine and cytosine nucleotides in the DNA, % form is the percentage of formamide in the hybridization solution, and L is the length of the hybrid in base pairs. T_m is reduced by about 1°C for each 1% of mismatching; thus, T_m, hybridization, and/or wash conditions can be adjusted to hybridize to sequences of the desired identity. For example, if sequences with >90% identity are sought, the T_m can be decreased 10°C. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence and its complement at a defined ionic strength and pH. However, severely stringent conditions can utilize a hybridization and/or wash at 1, 2, 3, or 4°C lower than the thermal melting point (T_m); moderately stringent conditions can utilize a hybridization and/or wash at 6, 7, 8, 9, or 10°C lower than the thermal melting point (T_m); low stringency conditions can utilize a hybridization and/or wash at 11, 12, 13, 14, 15, or 20°C lower than the thermal melting point (T_m). Using the equation, hybridization and wash compositions, and desired T, those of ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions are inherently described. If the desired degree of mismatching results in a T of less than 45°C (aqueous solution) or 32°C (formamide solution), it is preferred to increase the SSC concentration so that a higher temperature can be used. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993). Generally, highly stringent hybridization and wash conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH.

An example of highly stringent wash conditions is 0.15 M NaCl at 72°C for about 15 minutes. An example of stringent wash conditions is a 0.2X SSC wash at 65°C for 15 minutes (see, Sambrook and Russell, *infra*, for a description of SSC buffer). Often, a high stringency wash is preceded by a low stringency wash to remove background probe signal. An example medium stringency wash for a duplex of, *e.g.*, more than 100 nucleotides, is 1X SSC at 45°C for 15

minutes. An example low stringency wash for a duplex of, *e.g.*, more than 100 nucleotides, is 4-6X SSC at 40°C for 15 minutes. For short probes (*e.g.*, about 10 to 50 nucleotides), stringent conditions typically involve salt concentrations of less than about 1.5 M, more preferably about 0.01 to 1.0 M, Na ion concentration (or other salts) at pH 7.0 to 8.3, and the temperature is typically at least about 30°C and at least about 60°C for long probes (*e.g.*, >50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. In general, a signal to noise ratio of 2X (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization. Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the proteins that they encode are substantially identical. This occurs, *e.g.*, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code.

Very stringent conditions are selected to be equal to the T_m for a particular probe. An example of stringent conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or Northern blot is 50% formamide, *e.g.*, hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.1X SSC at 60 to 65°C. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1M NaCl, 1% SDS (sodium dodecyl sulphate) at 37°C, and a wash in 1X to 2X SSC (20X SSC = 3.0 M NaCl/0.3 M trisodium citrate) at 50 to 55°C. Exemplary moderate stringency conditions include hybridization in 40 to 45% formamide, 1.0 M NaCl, 1% SDS at 37°C, and a wash in 0.5X to 1X SSC at 55 to 60°C.

By "variant" polypeptide is intended a polypeptide derived from the native protein by deletion (also called “truncation”) or addition of one or more amino acids to the N-terminal and/or C-terminal end of the native protein; deletion or addition of one or more amino acids at one or more sites in the native protein; or substitution of one or more amino acids at one or more sites in the native protein. Such variants may results from, for example, genetic polymorphism or from human manipulation. Methods for such manipulations are generally known in the art.

Thus, the polypeptides of the invention may be altered in various ways including amino acid substitutions, deletions, truncations, and insertions. Methods for such manipulations are generally known in the art. For example, amino acid sequence variants of the polypeptides can be prepared by mutations in the DNA. Methods for mutagenesis and nucleotide sequence alterations are well known in the art. See, for example, Kunkel (1985); Kunkel *et al.* (1987); U. S. Patent No. 4,873,192; Walker and Gaastra (1983), and the references cited therein. Guidance as to appropriate amino acid substitutions that do not affect biological activity of the protein of interest may be found in the model of Dayhoff *et al.* (1978). Conservative substitutions, such as exchanging one amino acid with another having similar properties, are preferred.

Thus, the genes and nucleotide sequences of the invention include both the naturally occurring sequences as well as variant forms. Likewise, the polypeptides of the invention encompass both naturally occurring proteins as well as variations and modified forms thereof. Such variants will continue to possess the desired activity. The deletions, insertions, and substitutions of the polypeptide sequence encompassed herein are not expected to produce radical changes in the characteristics of the polypeptide. However, when it is difficult to predict the exact effect of the substitution, deletion, or insertion in advance of doing so, one skilled in the art will appreciate that the effect will be evaluated by routine screening assays.

Individual substitutions deletions or additions that alter, add or delete a single amino acid or a small percentage of amino acids (typically less than 5%, more typically less than 1%) in an encoded sequence are "conservatively modified variations," where the alterations result in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. The following five groups each contain amino acids that are conservative substitutions for one another: Aliphatic: Glycine (G), Alanine (A), Valine (V), Leucine (L), Isoleucine (I); Aromatic: Phenylalanine (F), Tyrosine (Y), Tryptophan (W); Sulfur-containing: Methionine (M), Cysteine (C); Basic: Arginine (R), Lysine (K), Histidine (H); Acidic: Aspartic acid (D), Glutamic acid (E), Asparagine (N), Glutamine (Q). In addition, individual substitutions,

deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids in an encoded sequence are also "conservatively modified variations."

The term "transformation" refers to the transfer of a nucleic acid fragment into the genome of a host cell, resulting in genetically stable inheritance. A "host cell" is a cell that has been transformed, or is capable of transformation, by an exogenous nucleic acid molecule. Host cells containing the transformed nucleic acid fragments are referred to as "transgenic" cells, and organisms comprising transgenic cells are referred to as "transgenic organisms".

"Transformed", "transduced", "transgenic", and "recombinant" refer to a host cell or organism into which a heterologous nucleic acid molecule has been introduced. The nucleic acid molecule can be stably integrated into the genome generally known in the art and are disclosed in Sambrook and Russell, *infra*. See also Innis *et al.* (1995); and Gelfand (1995); and Innis and Gelfand (1999). Known methods of PCR include, but are not limited to, methods using paired primers, nested primers, single specific primers, degenerate primers, gene-specific primers, vector-specific primers, partially mismatched primers, and the like. For example, "transformed," "transformant," and "transgenic" cells have been through the transformation process and contain a foreign gene integrated into their chromosome. The term "untransformed" refers to normal cells that have not been through the transformation process.

A "transgenic" organism is an organism having one or more cells that contain an expression vector.

"Genetically altered cells" denotes cells which have been modified by the introduction of recombinant or heterologous nucleic acids (*e.g.*, one or more DNA constructs or their RNA counterparts) and further includes the progeny of such cells which retain part or all of such genetic modification.

The term "fusion protein" is intended to describe at least two polypeptides, typically from different sources, which are operably linked. With regard to polypeptides, the term operably linked is intended to mean that the two polypeptides are connected in a manner such that each polypeptide can serve its intended function. Typically, the two polypeptides are covalently attached through peptide bonds. The fusion protein is preferably produced by standard

recombinant DNA techniques. For example, a DNA molecule encoding the first polypeptide is ligated to another DNA molecule encoding the second polypeptide, and the resultant hybrid DNA molecule is expressed in a host cell to produce the fusion protein. The DNA molecules are ligated to each other in a 5' to 3' orientation such that, after ligation, the translational frame of the encoded polypeptides is not altered (*i.e.*, the DNA molecules are ligated to each other in-frame).

As used herein, the term "derived" or "directed to" with respect to a nucleotide molecule means that the molecule has complementary sequence identity to a particular molecule of interest.

"Gene silencing" refers to the suppression of gene expression, e.g., transgene, heterologous gene and/or endogenous gene expression. Gene silencing may be mediated through processes that affect transcription and/or through processes that affect post-transcriptional mechanisms. In some embodiments, gene silencing occurs when siRNA initiates the degradation of the mRNA of a gene of interest in a sequence-specific manner *via* RNA interference (for a review, see Brantl, 2002). In some embodiments, gene silencing may be allele-specific. "Allele-specific" gene silencing refers to the specific silencing of one allele of a gene.

"Knock-down," "knock-down technology" refers to a technique of gene silencing in which the expression of a target gene is reduced as compared to the gene expression prior to the introduction of the siRNA, which can lead to the inhibition of production of the target gene product. The term "reduced" is used herein to indicate that the target gene expression is lowered by 1-100%. For example, the expression may be reduced by 10, 20, 30, 40, 50, 60, 70, 80, 90, 95, or even 99%. Knock-down of gene expression can be directed by the use of dsRNAs or siRNAs. For example, "RNA interference (RNAi)," which can involve the use of siRNA, has been successfully applied to knockdown the expression of specific genes in plants, *D. melanogaster, C. elegans*, trypanosomes, planaria, hydra, and several vertebrate species including the mouse and zebrafish. For a review of the mechanisms proposed to mediate RNAi, please refer to Bass *et al.*, 2001, Elbashir *et al.*, 2001 or Brantl 2002.

"RNA interference (RNAi)" is the process of sequence-specific, post-transcriptional gene silencing initiated by siRNA. RNAi is seen in a number of organisms such as *Drosophila*, nematodes, fungi and plants, and is believed to be involved in anti-viral defense, modulation of transposon activity, and regulation of gene expression. During RNAi, siRNA induces degradation of target mRNA with consequent sequence-specific inhibition of gene expression.

A "small interfering" or "short interfering RNA" or siRNA is a RNA duplex of nucleotides that is targeted to a gene interest. A "RNA duplex" refers to the structure formed by the complementary pairing between two regions of a RNA molecule. siRNA is "targeted" to a gene in that the nucleotide sequence of the duplex portion of the siRNA is complementary to a nucleotide sequence of the targeted gene. In some embodiments, the length of the duplex of siRNAs is less than 30 nucleotides. In some embodiments, the duplex can be 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11 or 10 nucleotides in length. In some embodiments, the length of the duplex is 19 - 25 nucleotides in length. The RNA duplex portion of the siRNA can be part of a hairpin structure. In addition to the duplex portion, the hairpin structure may contain a loop portion positioned between the two sequences that form the duplex. The loop can vary in length. In some embodiments the loop is 5, 6, 7, 8, 9, 10, 11, 12 or 13 nucleotides in length. The hairpin structure can also contain 3' or 5' overhang portions. In some embodiments, the overhang is a 3' or a 5' overhang 0, 1, 2, 3, 4 or 5 nucleotides in length.

The siRNA can be encoded by a nucleic acid sequence, and the nucleic acid sequence can also include a promoter. The nucleic acid sequence can also include a polyadenylation signal. In some embodiments, the polyadenylation signal is a synthetic minimal polyadelylation signal.

"Treating" as used herein refers to ameliorating at least one symptom of, curing and/or preventing the development of a disease or a condition.

"Neurodegenerative disease" and "neurodegenerative disorder" refer to both hereditary and sporadic conditions that are characterized by nervous system dysfunction, and which may be associated with atrophy of the affected central or peripheral nervous system structures, or loss of function without atrophy. Neurodegenerative diseases and disorders include but are not limited to

amyotrophic lateral sclerosis (ALS), hereditary spastic hemiplegia, primary lateral sclerosis, spinal muscular atrophy, Kennedy's disease, Alzheimer's disease, Parkinson's disease, multiple sclerosis, and repeat expansion neurodegenerative diseases, *e.g.*, diseases associated with expansions of trinucleotide repeats such as polyglutamine (polyQ) repeat diseases, *e.g.*, Huntington's disease (HD), spinocerebellar ataxia (SCA1, SCA2, SCA3, SCA6, SCA7), spinal and bulbar muscular atrophy (SBMA), and dentatorubropallidoluysian atrophy (DRPLA).

II. Nucleic Acid Molecules of the Invention

Sources of nucleotide sequences from which the present nucleic acid molecules can be obtained include any vertebrate, preferably mammalian, cellular source.

As discussed above, the terms "isolated and/or purified" refer to *in vitro* isolation of a nucleic acid, *e.g.*, a DNA or RNA molecule from its natural cellular environment, and from association with other components of the cell, such as nucleic acid or polypeptide, so that it can be sequenced, replicated, and/or expressed. For example, "isolated nucleic acid" is DNA containing less than 300, and more preferably less than 100 sequential nucleotide bases that comprise a DNA sequence that encodes a siRNA that forms a hairpin structure with a duplex 21 base pairs in length, or a variant thereof, that is complementary or hybridizes to a sequence in a gene of interest and remains stably bound under stringent conditions as defined by methods well known in the art, *e.g.*, in Sambrook and Russell, 2001. Thus, the RNA or DNA is "isolated" in that it is free from at least one contaminating nucleic acid with which it is normally associated in the natural source of the RNA or DNA and is preferably substantially free of any other mammalian RNA or DNA. The phrase "free from at least one contaminating source nucleic acid with which it is normally associated" includes the case where the nucleic acid is reintroduced into the source or natural cell but is in a different chromosomal location or is otherwise flanked by nucleic acid sequences not normally found in the source cell, *e.g.*, in a vector or plasmid.

In addition to a DNA sequence encoding a siRNA, the nucleic acid molecules of the invention include double-stranded interfering RNA molecules, which are also useful to inhibit expression of a target gene.

As used herein, the term "recombinant nucleic acid", *e.g.*, "recombinant DNA sequence or segment" refers to a nucleic acid, *e.g.*, to DNA, that has been derived or isolated from any appropriate cellular source, that may be subsequently chemically altered *in vitro*, so that its sequence is not naturally occurring, or corresponds to naturally occurring sequences that are not positioned as they would be positioned in a genome which has not been transformed with exogenous DNA. An example of preselected DNA "derived" from a source, would be a DNA sequence that is identified as a useful fragment within a given organism, and which is then chemically synthesized in essentially pure form. An example of such DNA "isolated" from a source would be a useful DNA sequence that is excised or removed from said source by chemical means, *e.g.*, by the use of restriction endonucleases, so that it can be further manipulated, *e.g.*, amplified, for use in the invention, by the methodology of genetic engineering.

Thus, recovery or isolation of a given fragment of DNA from a restriction digest can employ separation of the digest on polyacrylamide or agarose gel by electrophoresis, identification of the fragment of interest by comparison of its mobility versus that of marker DNA fragments of known molecular weight, removal of the gel section containing the desired fragment, and separation of the gel from DNA. See Lawn *et al.* (1981), and Goeddel *et al.* (1980). Therefore, "recombinant DNA" includes completely synthetic DNA sequences, semi-synthetic DNA sequences, DNA sequences isolated from biological sources, and DNA sequences derived from RNA, as well as mixtures thereof.

Nucleic acid molecules having base pair substitutions (*i.e.*, variants) are prepared by a variety of methods known in the art. These methods include, but are not limited to, isolation from a natural source (in the case of naturally occurring sequence variants) or preparation by oligonucleotide-mediated (or site-directed) mutagenesis, PCR mutagenesis, and cassette mutagenesis of an earlier prepared variant or a non-variant version of the nucleic acid molecule.

Oligonucleotide-mediated mutagenesis is a method for preparing substitution variants. This technique is known in the art as described by Adelman *et al.* (1983). Briefly, nucleic acid encoding a siRNA can be altered by hybridizing an oligonucleotide encoding the desired mutation to a DNA template, where the template is the single-stranded form of a plasmid or bacteriophage containing the unaltered or native gene sequence. After hybridization, a DNA polymerase is used to synthesize an entire second complementary strand of the template that will thus incorporate the oligonucleotide primer, and will code for the selected alteration in the nucleic acid encoding siRNA. Generally, oligonucleotides of at least 25 nucleotides in length are used. An optimal oligonucleotide will have 12 to 15 nucleotides that are completely complementary to the template on either side of the nucleotide(s) coding for the mutation. This ensures that the oligonucleotide will hybridize properly to the single-stranded DNA template molecule. The oligonucleotides are readily synthesized using techniques known in the art such as that described by Crea *et al.* (1978).

The DNA template can be generated by those vectors that are either derived from bacteriophage M13 vectors (the commercially available M13mp18 and M13mp19 vectors are suitable), or those vectors that contain a single-stranded phage origin of replication as described by Viera *et al.* (1987). Thus, the DNA that is to be mutated may be inserted into one of these vectors to generate single-stranded template. Production of the single-stranded template is described in Chapter 3 of Sambrook and Russell, 2001. Alternatively, single-stranded DNA template may be generated by denaturing double-stranded plasmid (or other) DNA using standard techniques.

For alteration of the native DNA sequence (to generate amino acid sequence variants, for example), the oligonucleotide is hybridized to the single-stranded template under suitable hybridization conditions. A DNA polymerizing enzyme, usually the Klenow fragment of DNA polymerase I, is then added to synthesize the complementary strand of the template using the oligonucleotide as a primer for synthesis. A heteroduplex molecule is thus formed such that one strand of DNA encodes the mutated form of the DNA, and the other strand (the original template) encodes the native, unaltered sequence of

the DNA. This heteroduplex molecule is then transformed into a suitable host cell, usually a prokaryote such as *E. coli* JM101. After the cells are grown, they are plated onto agarose plates and screened using the oligonucleotide primer radiolabeled with 32-phosphate to identify the bacterial colonies that contain the mutated DNA. The mutated region is then removed and placed in an appropriate vector, generally an expression vector of the type typically employed for transformation of an appropriate host.

The method described immediately above may be modified such that a homoduplex molecule is created wherein both strands of the plasmid contain the mutations(s). The modifications are as follows: The single-stranded oligonucleotide is annealed to the single-stranded template as described above. A mixture of three deoxyribonucleotides, deoxyriboadenosine (dATP), deoxyriboguanosine (dGTP), and deoxyribothymidine (dTTP), is combined with a modified thiodeoxyribocytosine called dCTP-(*S) (which can be obtained from the Amersham Corporation). This mixture is added to the template-oligonucleotide complex. Upon addition of DNA polymerase to this mixture, a strand of DNA identical to the template except for the mutated bases is generated. In addition, this new strand of DNA will contain dCTP-(*S) instead of dCTP, which serves to protect it from restriction endonuclease digestion.

After the template strand of the double-stranded heteroduplex is nicked with an appropriate restriction enzyme, the template strand can be digested with ExoIII nuclease or another appropriate nuclease past the region that contains the site(s) to be mutagenized. The reaction is then stopped to leave a molecule that is only partially single-stranded. A complete double-stranded DNA homoduplex is then formed using DNA polymerase in the presence of all four deoxyribonucleotide triphosphates, ATP, and DNA ligase. This homoduplex molecule can then be transformed into a suitable host cell such as *E. coli* JM101.

III. Expression Cassettes of the Invention

To prepare expression cassettes, the recombinant DNA sequence or segment may be circular or linear, double-stranded or single-stranded. Generally, the DNA sequence or segment is in the form of chimeric DNA, such

as plasmid DNA or a vector that can also contain coding regions flanked by control sequences that promote the expression of the recombinant DNA present in the resultant transformed cell.

A "chimeric" vector or expression cassette, as used herein, means a vector or cassette including nucleic acid sequences from at least two different species, or has a nucleic acid sequence from the same species that is linked or associated in a manner that does not occur in the "native" or wild type of the species.

Aside from recombinant DNA sequences that serve as transcription units for a peptide, or portions thereof, a portion of the recombinant DNA may be untranscribed, serving a regulatory or a structural function. For example, the recombinant DNA may itself have a promoter that is active in mammalian cells, or may utilize a promoter already present in the genome that is the transformation target. Such promoters include the CMV promoter, as well as the RSV promoter, SV40 late promoter and retroviral LTRs (long terminal repeat elements), although many other promoter elements well known to the art, such as tissue specific promoters or regulatable promoters may be employed in the practice of the invention.

Other elements functional in the host cells, such as introns, enhancers, polyadenylation sequences and the like, may also be a part of the recombinant DNA. Such elements may or may not be necessary for the function of the DNA, but may provide improved expression of the DNA by affecting transcription, stability of the siRNA, or the like. Such elements may be included in the DNA as desired to obtain the optimal performance of the siRNA in the cell.

Control sequences are DNA sequences necessary for the expression of an operably linked coding sequence in a particular host organism. The control sequences that are suitable for prokaryotic cells, for example, include a promoter, and optionally an operator sequence, and a ribosome binding site. Eukaryotic cells are known to utilize promoters, polyadenylation signals, and enhancers.

Operably linked nucleic acids are nucleic acids placed in a functional relationship with another nucleic acid sequence. For example, a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of

the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, operably linked DNA sequences are DNA sequences that are linked are contiguous. However, enhancers do not have to be contiguous. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors or linkers are used in accord with conventional practice.

The recombinant DNA to be introduced into the cells may contain either a selectable marker gene or a reporter gene or both to facilitate identification and selection of transformed cells from the population of cells sought to be transformed. In other embodiments, the selectable marker may be carried on a separate piece of DNA and used in a co-transformation procedure. Both selectable markers and reporter genes may be flanked with appropriate regulatory sequences to enable expression in the host cells. Useful selectable markers are known in the art and include, for example, antibiotic-resistance genes, such as *neo* and the like.

Reporter genes are used for identifying potentially transformed cells and for evaluating the functionality of regulatory sequences. Reporter genes that encode for easily assayable proteins are well known in the art. In general, a reporter gene is a gene that is not present in or expressed by the recipient organism or tissue and that encodes a protein whose expression is manifested by some easily detectable property, *e.g.*, enzymatic activity. For example, reporter genes include the chloramphenicol acetyl transferase gene (cat) from Tn9 of *E. coli* and the luciferase gene from firefly *Photinus pyralis*. Expression of the reporter gene is assayed at a suitable time after the DNA has been introduced into the recipient cells.

The general methods for constructing recombinant DNA that can transform target cells are well known to those skilled in the art, and the same compositions and methods of construction may be utilized to produce the DNA useful herein. For example, Sambrook and Russell, *infra*, provides suitable methods of construction.

The recombinant DNA can be readily introduced into the host cells, *e.g.*, mammalian, bacterial, yeast or insect cells by transfection with an expression vector composed of DNA encoding the siRNA by any procedure useful for the

introduction into a particular cell, *e.g.*, physical or biological methods, to yield a cell having the recombinant DNA stably integrated into its genome or existing as a episomal element, so that the DNA molecules, or sequences of the present invention are expressed by the host cell. Preferably, the DNA is introduced into host cells via a vector. The host cell is preferably of eukaryotic origin, *e.g.*, plant, mammalian, insect, yeast or fungal sources, but host cells of non-eukaryotic origin may also be employed.

Physical methods to introduce a preselected DNA into a host cell include calcium phosphate precipitation, lipofection, particle bombardment, microinjection, electroporation, and the like. Biological methods to introduce the DNA of interest into a host cell include the use of DNA and RNA viral vectors. For mammalian gene therapy, as described hereinbelow, it is desirable to use an efficient means of inserting a copy gene into the host genome. Viral vectors, and especially retroviral vectors, have become the most widely used method for inserting genes into mammalian, *e.g.*, human cells. Other viral vectors can be derived from poxviruses, herpes simplex virus I, adenoviruses and adeno-associated viruses, and the like. See, for example, U.S. Patent Nos. 5,350,674 and 5,585,362.

As discussed above, a “transfected”, “transformed’ or “transduced” host cell or cell line is one in which the genome has been altered or augmented by the presence of at least one heterologous or recombinant nucleic acid sequence. The host cells of the present invention are typically produced by transfection with a DNA sequence in a plasmid expression vector, a viral expression vector, or as an isolated linear DNA sequence. The transfected DNA can become a chromosomally integrated recombinant DNA sequence, which is composed of sequence encoding the siRNA.

To confirm the presence of the recombinant DNA sequence in the host cell, a variety of assays may be performed. Such assays include, for example, "molecular biological" assays well known to those of skill in the art, such as Southern and Northern blotting, RT-PCR and PCR; "biochemical" assays, such as detecting the presence or absence of a particular peptide, *e.g.*, by immunological means (ELISAs and Western blots) or by assays described herein to identify agents falling within the scope of the invention.

To detect and quantitate RNA produced from introduced recombinant DNA segments, RT-PCR may be employed. In this application of PCR, it is first necessary to reverse transcribe RNA into DNA, using enzymes such as reverse transcriptase, and then through the use of conventional PCR techniques amplify the DNA. In most instances PCR techniques, while useful, will not demonstrate integrity of the RNA product. Further information about the nature of the RNA product may be obtained by Northern blotting. This technique demonstrates the presence of an RNA species and gives information about the integrity of that RNA. The presence or absence of an RNA species can also be determined using dot or slot blot Northern hybridizations. These techniques are modifications of Northern blotting and only demonstrate the presence or absence of an RNA species.

While Southern blotting and PCR may be used to detect the recombinant DNA segment in question, they do not provide information as to whether the preselected DNA segment is being expressed. Expression may be evaluated by specifically identifying the peptide products of the introduced recombinant DNA sequences or evaluating the phenotypic changes brought about by the expression of the introduced recombinant DNA segment in the host cell.

The instant invention provides a cell expression system for expressing exogenous nucleic acid material in a mammalian recipient. The expression system, also referred to as a "genetically modified cell", comprises a cell and an expression vector for expressing the exogenous nucleic acid material. The genetically modified cells are suitable for administration to a mammalian recipient, where they replace the endogenous cells of the recipient. Thus, the preferred genetically modified cells are non-immortalized and are non-tumorigenic.

According to one embodiment, the cells are transformed or otherwise genetically modified *ex vivo*. The cells are isolated from a mammal (preferably a human), transformed (*i.e.*, transduced or transfected *in vitro*) with a vector for expressing a heterologous (*e.g.*, recombinant) gene encoding the therapeutic agent, and then administered to a mammalian recipient for delivery of the therapeutic agent *in situ*. The mammalian recipient may be a human and the

cells to be modified are autologous cells, *i.e.*, the cells are isolated from the mammalian recipient.

According to another embodiment, the cells are transformed or otherwise genetically modified *in vivo*. The cells from the mammalian recipient are transduced or transfected *in vivo* with a vector containing exogenous nucleic acid material for expressing a heterologous (*e.g.*, recombinant) gene encoding a therapeutic agent and the therapeutic agent is delivered *in situ*.

As used herein, "exogenous nucleic acid material" refers to a nucleic acid or an oligonucleotide, either natural or synthetic, which is not naturally found in the cells; or if it is naturally found in the cells, is modified from its original or native form. Thus, "exogenous nucleic acid material" includes, for example, a non-naturally occurring nucleic acid that can be transcribed into an anti-sense RNA, a siRNA, as well as a "heterologous gene" (*i.e.*, a gene encoding a protein that is not expressed or is expressed at biologically insignificant levels in a naturally-occurring cell of the same type). To illustrate, a synthetic or natural gene encoding human erythropoietin (EPO) would be considered "exogenous nucleic acid material" with respect to human peritoneal mesothelial cells since the latter cells do not naturally express EPO. Still another example of "exogenous nucleic acid material" is the introduction of only part of a gene to create a recombinant gene, such as combining an regulatable promoter with an endogenous coding sequence via homologous recombination.

IV. Promoters of the Invention

As described herein, an expression cassette of the invention contains, *inter alia*, a pol II promoter that is operably linked to a nucleic acid sequence encoding a siRNA. Thus, the pol II promoter, *i.e.*, a RNA polymerase II dependent promoter, initiates the transcription of the siRNA. In another embodiment, the pol II promoter is regulatable.

Three RNA polymerases transcribe nuclear genes in eukaryotes. RNA polymerase II (pol II) synthesizes mRNA, *i.e.*, pol II transcribes the genes that encode proteins. In contrast, RNA polymerase I (pol I) and RNA polymerase III (pol III) transcribe only a limited set of genes, synthesizing RNAs that have structural or catalytic roles. RNA polymerase I makes the large ribosomal RNAs

(rRNA), which are under the control of pol I promoters. RNA polymerase III makes a variety of small, stable RNAs, including the small 5S rRNA and transfer RNAs (tRNA), the transcription of which is under the control of pol III promoters.

As described herein, the inventors unexpectedly discovered that pol II promoters are useful to direct transcription of the siRNA. This was surprising because, as discussed above, pol II promoters are thought to be responsible for transcription of messenger RNA, *i.e.*, relatively long RNAs as compared to RNAs of 30 base pairs or less.

A pol II promoter may be used in its entirety, or a portion or fragment of the promoter sequence may be used in which the portion maintains the promoter activity. As discussed herein, pol II promoters are known to a skilled person in the art and include the promoter of any protein-encoding gene, *e.g.*, an endogenously regulated gene or a constitutively expressed gene. For example, the promoters of genes regulated by cellular physiological events, *e.g.*, heat shock, oxygen levels and/or carbon monoxide levels, *e.g.*, in hypoxia, may be used in the expression cassettes of the invention. In addition, the promoter of any gene regulated by the presence of a pharmacological agent, *e.g.*, tetracycline and derivatives thereof, as well as heavy metal ions and hormones may be employed in the expression cassettes of the invention. In an embodiment of the invention, the pol II promoter can be the CMV promoter or the RSV promoter. In another embodiment, the pol II promoter is the CMV promoter.

As discussed above, a pol II promoter of the invention may be one naturally associated with an endogenously regulated gene or sequence, as may be obtained by isolating the 5′ non-coding sequences located upstream of the coding segment and/or exon. The pol II promoter of the expression cassette can be, for example, the same pol II promoter driving expression of the targeted gene of interest. Alternatively, the nucleic acid sequence encoding the siRNA may be placed under the control of a recombinant or heterologous pol II promoter, which refers to a promoter that is not normally associated with the targeted gene's natural environment. Such promoters include promoters of genes other that the target gene of interest, and promoters isolated from any other eukaryotic cell, and promoters not "naturally occurring," *i.e.*, containing different elements of

different transcriptional regulatory regions, and/or mutations that alter expression. In addition to producing nucleic acid sequences of promoters synthetically, sequences may be produced using recombinant cloning and/or nucleic acid amplification technology, including PCR™, in connection with the compositions disclosed herein (see U.S. Patent 4,683,202, U.S. Patent 5,928,906, each incorporated herein by reference).

In one embodiment, a pol II promoter that effectively directs the expression of the siRNA in the cell type, organelle, and organism chosen for expression will be employed. Those of ordinary skill in the art of molecular biology generally know the use of promoters for protein expression, for example, see Sambrook and Russell (2001), incorporated herein by reference. The promoters employed may be constitutive, tissue-specific, inducible, and/or useful under the appropriate conditions to direct high level expression of the introduced DNA segment, such as is advantageous in the large-scale production of recombinant proteins and/or peptides. The identity of tissue-specific promoters, as well as assays to characterize their activity, is well known to those of ordinary skill in the art.

V. Methods for Introducing the Expression Cassettes of the Invention into Cells

The condition amenable to gene inhibition therapy may be a prophylactic process, *i.e.*, a process for preventing disease or an undesired medical condition. Thus, the instant invention embraces a system for delivering siRNA that has a prophylactic function (*i.e.*, a prophylactic agent) to the mammalian recipient.

The inhibitory nucleic acid material (*e.g.*, an expression cassette encoding siRNA directed to a gene of interest) can be introduced into the cell *ex vivo* or *in vivo* by genetic transfer methods, such as transfection or transduction, to provide a genetically modified cell. Various expression vectors (*i.e.*, vehicles for facilitating delivery of exogenous nucleic acid into a target cell) are known to one of ordinary skill in the art.

As used herein, "transfection of cells" refers to the acquisition by a cell of new nucleic acid material by incorporation of added DNA. Thus, transfection refers to the insertion of nucleic acid into a cell using physical or chemical

methods. Several transfection techniques are known to those of ordinary skill in the art including: calcium phosphate DNA co-precipitation (Methods in Molecular Biology (1991)); DEAE-dextran (supra); electroporation (supra); cationic liposome-mediated transfection (supra); and tungsten particle-facilitated microparticle bombardment (Johnston (1990)). Strontium phosphate DNA co-precipitation (Brash *et al.* (1987)) is also a transfection method.

In contrast, "transduction of cells" refers to the process of transferring nucleic acid into a cell using a DNA or RNA virus. A RNA virus (*i.e.*, a retrovirus) for transferring a nucleic acid into a cell is referred to herein as a transducing chimeric retrovirus. Exogenous nucleic acid material contained within the retrovirus is incorporated into the genome of the transduced cell. A cell that has been transduced with a chimeric DNA virus (*e.g.*, an adenovirus carrying a cDNA encoding a therapeutic agent), will not have the exogenous nucleic acid material incorporated into its genome but will be capable of expressing the exogenous nucleic acid material that is retained extrachromosomally within the cell.

The exogenous nucleic acid material can include the nucleic acid encoding the siRNA together with a promoter to control transcription. The promoter characteristically has a specific nucleotide sequence necessary to initiate transcription. The exogenous nucleic acid material may further include additional sequences (*i.e.*, enhancers) required to obtain the desired gene transcription activity. For the purpose of this discussion an "enhancer" is simply any non-translated DNA sequence that works with the coding sequence (in *cis*) to change the basal transcription level dictated by the promoter. The exogenous nucleic acid material may be introduced into the cell genome immediately downstream from the promoter so that the promoter and coding sequence are operatively linked so as to permit transcription of the coding sequence. An expression vector can include an exogenous promoter element to control transcription of the inserted exogenous gene. Such exogenous promoters include both constitutive and regulatable promoters.

Naturally-occurring constitutive promoters control the expression of essential cell functions. As a result, a nucleic acid sequence under the control of a constitutive promoter is expressed under all conditions of cell growth.

Constitutive promoters include the promoters for the following genes which encode certain constitutive or "housekeeping" functions: hypoxanthine phosphoribosyl transferase (HPRT), dihydrofolate reductase (DHFR) (Scharfmann *et al.* (1991)), adenosine deaminase, phosphoglycerol kinase (PGK), pyruvate kinase, phosphoglycerol mutase, the β-actin promoter (Lai *et al.* (1989)), and other constitutive promoters known to those of skill in the art. In addition, many viral promoters function constitutively in eucaryotic cells. These include: the early and late promoters of SV40; the long terminal repeats (LTRs) of Moloney Leukemia Virus and other retroviruses; and the thymidine kinase promoter of Herpes Simplex Virus, among many others.

Nucleic acid sequences that are under the control of regulatable promoters are expressed only or to a greater degree in the presence of an inducing agent, (*e.g.*, transcription under control of the metallothionein promoter is greatly increased in presence of certain metal ions). Regulatable promoters include responsive elements (REs) that stimulate transcription when their inducing factors are bound. For example, there are REs for serum factors, steroid hormones, retinoic acid and cyclic AMP. Promoters containing a particular RE can be chosen in order to obtain an regulatable response and in some cases, the RE itself may be attached to a different promoter, thereby conferring regulatability to the encoded nucleic acid sequence. Thus, by selecting the appropriate promoter (constitutive versus regulatable; strong versus weak), it is possible to control both the existence and level of expression of a nucleic acid sequence in the genetically modified cell. If the nucleic acid sequence is under the control of an regulatable promoter, delivery of the therapeutic agent *in situ* is triggered by exposing the genetically modified cell *in situ* to conditions for permitting transcription of the nucleic acid sequence, *e.g.*, by intraperitoneal injection of specific inducers of the regulatable promoters which control transcription of the agent. For example, *in situ* expression of a nucleic acid sequence under the control of the metallothionein promoter in genetically modified cells is enhanced by contacting the genetically modified cells with a solution containing the appropriate (*i.e.*, inducing) metal ions *in situ*.

Accordingly, the amount of siRNA generated *in situ* is regulated by controlling such factors as the nature of the promoter used to direct transcription

of the nucleic acid sequence, (*i.e.*, whether the promoter is constitutive or regulatable, strong or weak) and the number of copies of the exogenous nucleic acid sequence encoding a siRNA sequence that are in the cell.

In addition to at least one promoter and at least one heterologous nucleic acid sequence encoding the siRNA, the expression vector may include a selection gene, for example, a neomycin resistance gene, for facilitating selection of cells that have been transfected or transduced with the expression vector.

Cells can also be transfected with two or more expression vectors, at least one vector containing the nucleic acid sequence(s) encoding the siRNA(s), the other vector containing a selection gene. The selection of a suitable promoter, enhancer, selection gene and/or signal sequence is deemed to be within the scope of one of ordinary skill in the art without undue experimentation.

The following discussion is directed to various utilities of the instant invention. For example, the instant invention has utility as an expression system suitable for silencing the expression of gene(s) of interest.

The instant invention also provides various methods for making and using the above-described genetically-modified cells.

The instant invention also provides methods for genetically modifying cells of a mammalian recipient *in vivo*. According to one embodiment, the method comprises introducing an expression vector for expressing a siRNA sequence in cells of the mammalian recipient *in situ* by, for example, injecting the vector into the recipient.

VI. Delivery Vehicles for the Expression Cassettes of the Invention

Delivery of compounds into tissues and across the blood-brain barrier can be limited by the size and biochemical properties of the compounds. Currently, efficient delivery of compounds into cells *in vivo* can be achieved only when the molecules are small (usually less than 600 Daltons). Gene transfer for the correction of inborn errors of metabolism and neurodegenerative diseases of the central nervous system (CNS), and for the treatment of cancer has been accomplished with recombinant adenoviral vectors.

The selection and optimization of a particular expression vector for expressing a specific siRNA in a cell can be accomplished by obtaining the nucleic acid sequence of the siRNA, possibly with one or more appropriate control regions (*e.g.*, promoter, insertion sequence); preparing a vector construct comprising the vector into which is inserted the nucleic acid sequence encoding the siRNA; transfecting or transducing cultured cells *in vitro* with the vector construct; and determining whether the siRNA is present in the cultured cells.

Vectors for cell gene therapy include viruses, such as replication-deficient viruses (described in detail below). Exemplary viral vectors are derived from Harvey Sarcoma virus, ROUS Sarcoma virus, (MPSV), Moloney murine leukemia virus and DNA viruses (*e.g.*, adenovirus) (Ternin (1986)).

Replication-deficient retroviruses are capable of directing synthesis of all virion proteins, but are incapable of making infectious particles. Accordingly, these genetically altered retroviral expression vectors have general utility for high-efficiency transduction of nucleic acid sequences in cultured cells, and specific utility for use in the method of the present invention. Such retroviruses further have utility for the efficient transduction of nucleic acid sequences into cells *in vivo*. Retroviruses have been used extensively for transferring nucleic acid material into cells. Standard protocols for producing replication-deficient retroviruses (including the steps of incorporation of exogenous nucleic acid material into a plasmid, transfection of a packaging cell line with plasmid, production of recombinant retroviruses by the packaging cell line, collection of viral particles from tissue culture media, and infection of the target cells with the viral particles) are provided in Kriegler (1990) and Murray (1991).

An advantage of using retroviruses for gene therapy is that the viruses insert the nucleic acid sequence encoding the siRNA into the host cell genome, thereby permitting the nucleic acid sequence encoding the siRNA to be passed on to the progeny of the cell when it divides. Promoter sequences in the LTR region have been reported to enhance expression of an inserted coding sequence in a variety of cell types (see *e.g.*, Hilberg *et al.* (1987); Holland *et al.* (1987); Valerio *et al.* (1989). Some disadvantages of using a retrovirus expression vector are (1) insertional mutagenesis, *i.e.*, the insertion of the nucleic acid sequence encoding the siRNA into an undesirable position in the target cell

genome which, for example, leads to unregulated cell growth and (2) the need for target cell proliferation in order for the nucleic acid sequence encoding the siRNA carried by the vector to be integrated into the target genome (Miller *et al.* (1990)).

Another viral candidate useful as an expression vector for transformation of cells is the adenovirus, a double-stranded DNA virus. The adenovirus is infective in a wide range of cell types, including, for example, muscle and endothelial cells (Larrick and Burck (1991)). The adenovirus also has been used as an expression vector in muscle cells *in vivo* (Quantin *et al.* (1992)).

Adenoviruses (Ad) are double-stranded linear DNA viruses with a 36 kb genome. Several features of adenovirus have made them useful as transgene delivery vehicles for therapeutic applications, such as facilitating *in vivo* gene delivery. Recombinant adenovirus vectors have been shown to be capable of efficient *in situ* gene transfer to parenchymal cells of various organs, including the lung, brain, pancreas, gallbladder, and liver. This has allowed the use of these vectors in methods for treating inherited genetic diseases, such as cystic fibrosis, where vectors may be delivered to a target organ. In addition, the ability of the adenovirus vector to accomplish *in situ* tumor transduction has allowed the development of a variety of anticancer gene therapy methods for non-disseminated disease. In these methods, vector containment favors tumor cell-specific transduction.

Like the retrovirus, the adenovirus genome is adaptable for use as an expression vector for gene therapy, *i.e.*, by removing the genetic information that controls production of the virus itself (Rosenfeld *et al.* (1991)). Because the adenovirus functions in an extrachromosomal fashion, the recombinant adenovirus does not have the theoretical problem of insertional mutagenesis.

Several approaches traditionally have been used to generate the recombinant adenoviruses. One approach involves direct ligation of restriction endonuclease fragments containing a nucleic acid sequence of interest to portions of the adenoviral genome. Alternatively, the nucleic acid sequence of interest may be inserted into a defective adenovirus by homologous recombination results. The desired recombinants are identified by screening individual plaques generated in a lawn of complementation cells.

Most adenovirus vectors are based on the adenovirus type 5 (Ad5) backbone in which an expression cassette containing the nucleic acid sequence of interest has been introduced in place of the early region 1 (E1) or early region 3 (E3). Viruses in which E1 has been deleted are defective for replication and are propagated in human complementation cells (*e.g.*, 293 or 911 cells), which supply the missing gene E1 and pIX in *trans*.

In one embodiment of the present invention, one will desire to generate siRNA in a brain cell or brain tissue. A suitable vector for this application is an FIV vector (Brooks *et al.* (2002); Alisky *et al.* (2000a)) or an AAV vector. For example, one may use AAV5 (Davidson *et al.* (2000); Alisky *et al.* (2000a)). Also, one may apply poliovirus (Bledsoe *et al.* (2000)) or HSV vectors (Alisky *et al.* (2000b)).

Thus, as will be apparent to one of ordinary skill in the art, a variety of suitable viral expression vectors are available for transferring exogenous nucleic acid material into cells. The selection of an appropriate expression vector to express a therapeutic agent for a particular condition amenable to gene silencing therapy and the optimization of the conditions for insertion of the selected expression vector into the cell, are within the scope of one of ordinary skill in the art without the need for undue experimentation.

In another embodiment, the expression vector is in the form of a plasmid, which is transferred into the target cells by one of a variety of methods: physical (*e.g.*, microinjection (Capecchi (1980)), electroporation (Andreason and Evans (1988), scrape loading, microparticle bombardment (Johnston (1990)) or by cellular uptake as a chemical complex (*e.g.*, calcium or strontium co-precipitation, complexation with lipid, complexation with ligand) (Methods in Molecular Biology (1991)). Several commercial products are available for cationic liposome complexation including Lipofectin™ (Gibco-BRL, Gaithersburg, Md.) (Felgner *et al.* (1987)) and Transfectam™ (ProMega, Madison, Wis.) (Behr *et al.* (1989); Loeffler *et al.* (1990)). However, the efficiency of transfection by these methods is highly dependent on the nature of the target cell and accordingly, the conditions for optimal transfection of nucleic acids into cells using the above-mentioned procedures must be optimized. Such

optimization is within the scope of one of ordinary skill in the art without the need for undue experimentation.

VII. Diseases and Conditions Amendable to the Methods of the Invention

In the certain embodiments of the present invention, a mammalian recipient to an expression cassette of the invention has a condition that is amenable to gene silencing therapy. As used herein, "gene silencing therapy" refers to administration to the recipient exogenous nucleic acid material encoding a therapeutic siRNA and subsequent expression of the administered nucleic acid material *in situ*. Thus, the phrase "condition amenable to siRNA therapy" embraces conditions such as genetic diseases (*i.e.*, a disease condition that is attributable to one or more gene defects), acquired pathologies (*i.e.*, a pathological condition that is not attributable to an inborn defect), cancers, neurodegenerative diseases, *e.g.*, trinucleotide repeat disorders, and prophylactic processes (*i.e.*, prevention of a disease or of an undesired medical condition). A gene "associated with a condition" is a gene that is either the cause, or is part of the cause, of the condition to be treated. Examples of such genes include genes associated with a neurodegenerative disease (e.g., a trinucleotide-repeat disease such as a disease associated with polyglutamine repeats, Huntington's disease, and spinocerebellar ataxia), and genes encoding ligands for chemokines involved in the migration of a cancer cells, or chemokine receptor. Also siRNA expressed from viral vectors may be used for in vivo antiviral therapy using the vector systems described.

Accordingly, as used herein, the term "therapeutic siRNA" refers to any siRNA that has a beneficial effect on the recipient. Thus, "therapeutic siRNA " embraces both therapeutic and prophylactic siRNA.

A. Gene defects

A number of diseases caused by gene defects have been identified. For example, this strategy can be applied to a major class of neurodegenerative disorders, the polyglutamine diseases, as is demonstrated by the reduction of polyglutamine aggregation in cells following application of the strategy. The

neurodegenerative disease may be a trinucleotide-repeat disease, such as a disease associated with polyglutamine repeats, Huntington's disease or spinocerebellar ataxia.

B. Acquired pathologies

As used herein, "acquired pathology" refers to a disease or syndrome manifested by an abnormal physiological, biochemical, cellular, structural, or molecular biological state. For example, the disease could be a viral disease, such as hepatitis or AIDs.

C. Cancers

The condition amenable to gene silencing therapy alternatively can be a genetic disorder or an acquired pathology that is manifested by abnormal cell proliferation, *e.g.*, cancer. According to this embodiment, the instant invention is useful for silencing a gene involved in neoplastic activity. The present invention can also be used to inhibit overexpression of one or several genes that impart differentiation. The present invention can be used to treat neuroblastoma, medulloblastoma, or glioblastoma.

D. Neurodegenerative Diseases

Expansions of poly-glutamine tracts in proteins that are expressed in the central nervous system can cause neurodegenerative diseases. Some neurodegenerative diseases are caused by a $(CAG)_n$ repeat that encodes poly-glutamine in a protein include Huntington disease (HD), spinocerebellar ataxia (SCA1, SCA2, SCA3, SCA6, SCA7), spinal and bulbar muscular atrophy (SBMA), and dentatorubropallidoluysian atrophy (DRPLA). In these diseases, the poly-glutamine expansion in a protein confers a novel toxic property upon the protein. Studies indicate that the toxic property is a tendency for the disease protein to misfold and form aggregates within neurons.

HD is also known as Huntington's Chorea, Chronic Progressive Chorea, and Hereditary Chorea. HD is an autosomal dominant genetic disorder characterized by choreiform movements and progressive intellectual deterioration, usually beginning in middle age (35 to 50 yr). The disease affects both sexes equally. The caudate nucleus atrophies, the small-cell population

degenerates, and levels of the neurotransmitters γ-aminobutyric acid (GABA) and substance P decrease. This degeneration results in characteristic "boxcar ventricles" seen on CT scans.

The gene involved in Huntington's disease (IT-15) is located at the end of the short arm of chromosome 4. A mutation occurs in the coding region of this gene and produces an unstable expanded trinucleotide repeat (cytosine-adenosine-guanosine), resulting in a protein with an expanded glutamate sequence. The normal and abnormal functions of this protein (termed huntingtin) are unknown. The abnormal huntingtin protein appears to accumulate in neuronal nuclei of transgenic mice, but the causal relationship of this accumulation to neuronal death is uncertain.

Symptoms and signs develop insidiously. Dementia or psychiatric disturbances, ranging from apathy and irritability to full-blown bipolar or schizophreniform disorder, may precede the movement disorder or develop during its course. Anhedonia or asocial behavior may be the first behavioral manifestation. Motor manifestations include flicking movements of the extremities, a lilting gait, motor impersistence (inability to sustain a motor act, such as tongue protrusion), facial grimacing, ataxia, and dystonia.

Treatment for HD is currently not available. The choreic movements and agitated behaviors may be suppressed, usually only partially, by antipsychotics (*e.g.*, chlorpromazine 100 to 900 mg/day po or haloperidol 10 to 90 mg/day po) or reserpine begun with 0.1 mg/day po and increased until adverse effects of lethargy, hypotension, or parkinsonism occur.

VIII. Dosages, Formulations and Routes of Administration of the Agents of the Invention

The agents of the invention are preferably administered so as to result in a reduction in at least one symptom associated with a disease. The amount administered will vary depending on various factors including, but not limited to, the composition chosen, the particular disease, the weight, the physical condition, and the age of the mammal, and whether prevention or treatment is to be achieved. Such factors can be readily determined by the clinician employing animal models or other test systems which are well known to the art.

Administration of siRNA may be accomplished through the administration of the nucleic acid molecule encoding the siRNA (see, for example, Felgner *et al.*, U.S. Patent No. 5,580,859, Pardoll *et al.* 1995; Stevenson *et al.* 1995; Molling 1997; Donnelly *et al.* 1995; Yang *et al.* II; Abdallah *et al.* 1995). Pharmaceutical formulations, dosages and routes of administration for nucleic acids are generally disclosed, for example, in Felgner *et al.*, *supra.*

The present invention envisions treating a disease, for example, a neurodegenerative disease, in a mammal by the administration of an agent, *e.g.*, a nucleic acid composition, an expression vector, or a viral particle of the invention. Administration of the therapeutic agents in accordance with the present invention may be continuous or intermittent, depending, for example, upon the recipient's physiological condition, whether the purpose of the administration is therapeutic or prophylactic, and other factors known to skilled practitioners. The administration of the agents of the invention may be essentially continuous over a preselected period of time or may be in a series of spaced doses. Both local and systemic administration is contemplated.

One or more suitable unit dosage forms having the therapeutic agent(s) of the invention, which, as discussed below, may optionally be formulated for sustained release (for example using microencapsulation, see WO 94/07529, and U.S. Patent No. 4,962,091 the disclosures of which are incorporated by reference herein), can be administered by a variety of routes including parenteral, including by intravenous and intramuscular routes, as well as by direct injection into the diseased tissue. For example, the therapeutic agent may be directly injected into the brain. Alternatively the therapeutic agent may be introduced intrathecally for brain and spinal cord conditions. In another example, the therapeutic agent may be introduced intramuscularly for viruses that traffic back to affected neurons from muscle, such as AAV, lentivirus and adenovirus. The formulations may, where appropriate, be conveniently presented in discrete unit dosage forms and may be prepared by any of the methods well known to pharmacy. Such methods may include the step of bringing into association the therapeutic agent with liquid carriers, solid matrices, semi-solid carriers, finely

divided solid carriers or combinations thereof, and then, if necessary, introducing or shaping the product into the desired delivery system.

When the therapeutic agents of the invention are prepared for administration, they are preferably combined with a pharmaceutically acceptable carrier, diluent or excipient to form a pharmaceutical formulation, or unit dosage form. The total active ingredients in such formulations include from 0.1 to 99.9% by weight of the formulation. A "pharmaceutically acceptable" is a carrier, dilutent, excipient, and/or salt that is compatible with the other ingredients of the formulation, and not deleterious to the recipient thereof. The active ingredient for administration may be present as a powder or as granules; as a solution, a suspension or an emulsion. Pharmaceutical formulations containing the therapeutic agents of the invention can be prepared by procedures known in the art using well known and readily available ingredients. The therapeutic agents of the invention can also be formulated as solutions appropriate for parenteral administration, for instance by intramuscular, subcutaneous or intravenous routes.

The pharmaceutical formulations of the therapeutic agents of the invention can also take the form of an aqueous or anhydrous solution or dispersion, or alternatively the form of an emulsion or suspension.

Thus, the therapeutic agent may be formulated for parenteral administration (*e.g.*, by injection, for example, bolus injection or continuous infusion) and may be presented in unit dose form in ampules, pre-filled syringes, small volume infusion containers or in multi-dose containers with an added preservative. The active ingredients may take such forms as suspensions, solutions, or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the active ingredients may be in powder form, obtained by aseptic isolation of sterile solid or by lyophilization from solution, for constitution with a suitable vehicle, *e.g.*, sterile, pyrogen-free water, before use.

It will be appreciated that the unit content of active ingredient or ingredients contained in an individual aerosol dose of each dosage form need not in itself constitute an effective amount for treating the particular indication or disease since the necessary effective amount can be reached by administration of

a plurality of dosage units. Moreover, the effective amount may be achieved using less than the dose in the dosage form, either individually, or in a series of administrations.

The pharmaceutical formulations of the present invention may include, as optional ingredients, pharmaceutically acceptable carriers, diluents, solubilizing or emulsifying agents, and salts of the type that are well-known in the art. Specific non-limiting examples of the carriers and/or diluents that are useful in the pharmaceutical formulations of the present invention include water and physiologically acceptable buffered saline solutions such as phosphate buffered saline solutions pH 7.0-8.0. saline solutions and water.

The invention will now be illustrated by the following non-limiting Example.

Example 1

Experimental Protocols

Generation of the expression cassettes and viral vectors. The modified CMV (mCMV) promoter was made by PCR amplification of CMV by primers 5'-AAGGTACCAGATCTTAGTTATTAATAGTAATCAATTACGG-3' (SEQ ID NO:1) and 5'-GAATCGATGCATGCCTCGAGACGGTTCACTAAACCAGCTCTGC-3' (SEQ ID NO:2) with peGFPN1 plasmid (purchased from Clontech, Inc) as template. The mCMV product was cloned into the Kpn*I* and Cla*I* sites of the adenoviral shuttle vector pAd5KnpA, and was named pmCMVknpA. To construct the minimal polyA cassette, the oligonucleotides, 5'-CTAGAACTAGTAATAAAGGATCCTTTATTTTCATTGGATCCGTGTGTT GGTTTTTTGTGTGCGGCCGCG-3' (SEQ ID NO:3) and 5'-TCGACGCGGCCGCACACAAAAAACCAACACACGGATCC AATGAAAATAAAGGATCCTTTATTACTAGTT-3' (SEQ ID NO:4), were used. The oligonucleotides contain Spe*I* and Sal*I* sites at the 5' and 3' ends, respectively. The synthesized polyA cassette was ligated into Spe*I*, Sal*I* digested pmCMVKnpA. The resultant shuttle plasmid, pmCMVmpA was used for construction of head-to-head 21bp hairpins of eGFP (bp 418 to 438), human β-

glucuronidase (bp 649 to 669), mouse β-glucuronidase (bp 646 to 666) or *E. coli* β-galactosidase (bp 1152-1172). The eGFP hairpins were also cloned into the Ad shuttle plasmid containing the commercially available CMV promoter and polyA cassette from SV40 large T antigen (pCMVsiGFPx). Shuttle plasmids were co-transfected into HEK293 cells along with the adenovirus backbones for generation of full-length Ad genomes. Viruses were harvested 6-10 days after transfection and amplified and purified as described (Anderson, R.D., et al., *Gene Ther.* **7**:1034-1038 (2000)).

Northern blotting. Total RNA was isolated from HEK293 cells transfected by plasmids or infected by adenoviruses using TRIZOL®Reagent (Invitrogen™ Life Technologies, Carlsbad, CA) according to the manufacturer's instruction. RNAs (30μg) were separated by electrophoresis on 15% (wt/vol) polyacrylamide-urea gels to detect transcripts, or on 1% agarose-formaldehyde gel for target mRNAs analysis. RNAs were transferred by electroblotting onto hybond N+ membrane (Amersham Pharmacia Biotech). Blots were probed with ^{32}P-labeled sense (5'-CACAAGCTGGAGTACAACTAC-3' (SEQ ID NO:5)) or antisense (5'-GTACTTGTACTCCAGCTTTGTG-3' (SEQ ID NO:6)) oligonucleotides at 37°C for 3h for evaluation of siRNA transcripts, or probed for target mRNAs at 42°C overnight. Blots were washed using standard methods and exposed to film overnight. *In vitro* studies were performed in triplicate with a minimum of two repeats.

***In vivo* studies and tissue analyses.** All animal procedures were approved by the University of Iowa Committee on the Care and Use of Animals. Mice were injected into the tail vein (n = 10 per group) or into the brain (n = 6 per group) as described previously (Stein, C.S., et al., *J. Virol.* **73**:3424-3429 (1999)) with the virus doses indicated. Animals were sacrificed at the noted times and tissues harvested and sections or tissue lysates evaluated for β-glucuronidase expression, eGFP fluorescence, or β-galactosidase activity using established methods (Xia, H. et al., *Nat. Biotechnol.* **19**:640-644 (2001)). Total RNA was harvested from transduced liver using the methods described above.

Cell Lines. PC12 tet off cell lines (Clontech Inc., Palo Alto, CA) were stably transfected with a tetracycline regulatable plasmid into which was cloned

GFPQ19 or GFPQ80 (Chai, Y. et al., *J. Neurosci.* 19:10338-10347 (1999)). For GFP-Q80, clones were selected and clone 29 chosen for regulatable properties and inclusion formation. For GFP-Q19 clone 15 was selected for uniformity of GFP expression following gene expression induction. In all studies 1.5 μg/ml dox was used to repress transcription. All experiments were done in triplicate and were repeated 4 times.

Results and Discussion

To accomplish intracellular expression of siRNA, a 21-bp hairpin representing sequences directed against eGFP was constructed, and its ability to reduce target gene expression in mammalian cells using two distinct constructs was tested. Initially, the siRNA hairpin targeted against eGFP was placed under the control of the CMV promoter and contained a full-length SV-40 polyadenylation (polyA) cassette (pCMVsiGFPx). In the second construct, the hairpin was juxtaposed almost immediate to the CMV transcription start site (within 6 bp) and was followed by a synthetic, minimal polyA cassette (Fig. 1A, pmCMVsiGFPmpA) (Experimental Protocols), because we reasoned that functional siRNA would require minimal to no overhangs (Caplan, N.J., et al., *Proc. Natl. Acad. Sci. U. S. A.* **98**:9742-9747 (2001); Nykänen, A., et al., *Cell* **107**:309-321 (2001)). Co-transfection of pmCMVsiGFPmpA with pEGFPN1 (Clontech Inc) into HEK293 cells markedly reduced eGFP fluorescence (Fig. 1C). pmCMVsiGFPmpA transfection led to the production of an approximately 63 bp RNA specific for eGFP (Fig. 1D), consistent with the predicted size of the siGFP hairpin-containing transcript. Reduction of target mRNA and eGFP protein expression was noted in pmCMVsiGFPmpA-transfected cells only (Fig. 1E, F). In contrast, eGFP RNA, protein and fluorescence levels remained unchanged in cells transfected with pEGFPN1 and pCMVsiGFPx (Fig. 1E, G), pEGFPN1 and pCMVsißglucmpA (Fig. 1E, F, H), or pEGFPN1 and pCMVsißgalmpA, the latter expressing siRNA against *E. coli* ß-galactosidase (Fig. 1E). These data demonstrate the specificity of the expressed siRNAs.

Constructs identical to pmCMVsiGFPmpA except that a spacer of 9, 12 and 21 nucleotides was present between the transcription start site and the 21 bp hairpin were also tested. In each case, there was no silencing of eGFP

expression (data not shown). Together the results indicate that the spacing of the hairpin immediate to the promoter can be important for functional target reduction, a fact supported by recent studies in MCF-7 cells (Brummelkamp, T.R., et al., *Science* **296**:550-553 (2002)).

5 Recombinant adenoviruses were generated from the siGFP (pmCMVsiGFPmpA) and siβgluc (pmCMVsiβglucmpA) plasmids (Xia, H., et al., *Nat. Biotechnol.* **19**:640-644 (2001); Anderson, R.D., et al., *Gene Ther.* **7**:1034-1038 (2000)) to test the hypothesis that virally expressed siRNA allows for diminished gene expression of endogenous targets *in vitro* and *in vivo*. HeLa 10 cells are of human origin and contain moderate levels of the soluble lysosomal enzyme β-glucuronidase. Infection of HeLa cells with viruses expressing siβgluc caused a specific reduction in human ß-glucuronidase mRNA (Fig. 1I) leading to a 60% decrease in β-glucuronidase activity relative to siGFP or control cells (Fig 1J). Optimization of siRNA sequences using methods to refine 15 target mRNA accessible sequences (Lee, N.S., et al., *Nat. Biotechnol.* **19**:500-505 (2002)) could improve further the diminution of ß-glucuronidase transcript and protein levels.

The results in Fig. 1 are consistent with earlier work demonstrating the ability of synthetic 21-bp double stranded RNAs to reduce expression of target 20 genes in mammalian cells following transfection, with the important difference that in the present studies the siRNA was synthesized intracellularly from readily available promoter constructs. The data support the utility of regulatable, tissue or cell-specific promoters for expression of siRNA when suitably modified for close juxtaposition of the hairpin to the transcriptional start site and inclusion of 25 the minimal polyA sequence containing cassette (*see*, Methods above).

To evaluate the ability of virally expressed siRNA to diminish target-gene expression in adult mouse tissues *in vivo*, transgenic mice expressing eGFP (Okabe, M. et al., *FEBS Lett.* **407**:313-319 (1997)) were injected into the striatal region of the brain with 1 x 10^7 infectious units of recombinant adenovirus 30 vectors expressing siGFP or control siβgluc. Viruses also contained a dsRed expression cassette in a distant region of the virus for unequivocal localization of the injection site. Brain sections evaluated 5 days after injection by fluorescence

(Fig. 2A) or western blot assay (Fig. 2B) demonstrated reduced eGFP expression. Decreased eGFP expression was confined to the injected hemisphere (Fig. 2B). The *in vivo* reduction is promising, particularly since transgenically expressed eGFP is a stable protein, making complete reduction in this short time frame unlikely. Moreover, evaluation of eGFP levels was done 5 days after injection, when inflammatory changes induced by the adenovirus vector likely enhance transgenic eGFP expression from the CMV enhancer (Ooboshi, H., et al., *Arterioscler. Thromb. Vasc. Biol.* **17**:1786-1792 (1997)).

It was next tested whether virus mediated siRNA could decrease expression from endogenous alleles *in vivo*. Its ability to decrease β-glucuronidase activity in the murine liver, where endogenous levels of this relatively stable protein are high, was evaluated. Mice were injected via the tail vein with a construct expressing murine-specific siβgluc (AdsiMuβgluc), or the control viruses Adsiβgluc (specific for human β-glucuronidase) or Adsiβgal. Adenoviruses injected into the tail vein transduced hepatocytes as shown previously (Stein, C.S., et al., *J. Virol.* **73**:3424-3429 (1999)). Liver tissue harvested 3 days later showed specific reduction of target ß-glucuronidase RNA in AdsiMußgluc treated mice only (Fig. 2C). Fluorometric enzyme assay of liver lysates confirmed these results, with a 12% decrease in activity from liver harvested from AdsiMuβgluc injected mice relative to Adsiβgal and Adsiβgluc treated ones ($p<0.01$; $n=10$). Interestingly, sequence differences between the murine and human siRNA constructs are limited, with 14 of 21 bp being identical. These results confirm the specificity of virus mediated siRNA, and suggest that allele-specific applications may be possible. Together, the data are the first to demonstrate the utility of siRNA to diminish target gene expression in brain and liver tissue *in vivo*.

One powerful therapeutic application of siRNA is to reduce expression of toxic gene products in dominantly inherited diseases such as the polyglutamine (polyQ) neurodegenerative disorders (Margolis, R.L. & Ross, C.A. *Trends Mol. Med.* **7**:479-482 (2001)). The molecular basis of polyQ diseases is a novel toxic property conferred upon the mutant protein by polyQ expansion. This toxic property is associated with disease protein aggregation. The ability of virally

expressed siRNA to diminish expanded polyQ protein expression in neural PC-12 clonal cell lines was evaluated. Lines were developed that express tetracycline-repressible eGFP-polyglutamine fusion proteins with normal or expanded glutamine of 19 (eGFP-Q19) and 80 (eGFP-Q80) repeats, respectively. Differentiated, eGFP-Q19-expressing PC12 neural cells infected with recombinant adenovirus expressing siGFP demonstrated a specific and dose-dependent decrease in eGFP-Q19 fluorescence (Fig. 3A, C) and protein levels (Fig. 3B). Application of Adsiβgluc as a control had no effect (Fig. 3A-C). Quantitative image analysis of eGFP fluorescence demonstrated that siGFP reduced GFPQ19 expression by greater than 96% and 93% for 100 and 50 MOI respectively, relative to control siRNA (Fig. 3C). The multiplicity of infection (MOI) of 100 required to achieve maximal inhibition of eGFP-Q19 expression results largely from the inability of PC12 cells to be infected by adenovirus-based vectors. This barrier can be overcome using AAV- or lentivirus-based expression systems (Davidson, B.L., et al., *Proc. Natl. Acad. Sci. U. S. A.* **97**:3428-3432 (2000); Brooks, A.I., et al, *Proc. Natl. Acad. Sci. U. S. A.* **99**:6216-6221 (2002)).

To test the impact of siRNA on the size and number of aggregates formed in eGFP-Q80 expressing cells, differentiated PC-12/eGFP-Q80 neural cells were infected with AdsiGFP or Adsiβgluc 3 days after doxycycline removal to induce GFP-Q80 expression. Cells were evaluated 3 days later. In mock-infected control cells (Fig. 4A), aggregates were very large 6 days after induction as reported by others (Chai, Y., et al., *J. Neurosci.* **19**:10338-10347 (1999; Moulder, K.L., et al., *J. Neurosci.* **19**:705-715 (1999)). Large aggregates were also seen in cells infected with Adsiβgluc (Fig. 4B), AdsiGFPx, (Fig. 4C, siRNA expressed from the normal CMV promoter and containing the SV40 large T antigen polyadenylation cassette), or Adsiβgal (Fig. 4D). In contrast, polyQ aggregate formation was significantly reduced in AdsiGFP infected cells (Fig. 4E), with fewer and smaller inclusions and more diffuse eGFP fluorescence. AdsiGFP-mediated reduction in aggregated and monomeric GFP-Q80 was verified by Western blot analysis (Fig. 4F), and quantitation of cellular

fluorescence (Fig. 4G). AdsiGFP caused a dramatic and specific, dose-dependent reduction in eGFP-Q80 expression (Fig. 4F, G).

It was found that transcripts expressed from the modified CMV promoter and containing the minimal polyA cassette were capable of reducing gene expression in both plasmid and viral vector systems (Figs. 1-4). The placement of the hairpin immediate to the transcription start site and use of the minimal polyadenylation cassette was of critical importance. In plants and Drosophila, RNA interference is initiated by the ATP-dependent, processive cleavage of long dsRNA into 21-25 bp double-stranded siRNA, followed by incorporation of siRNA into a RNA-induced silencing complex that recognizes and cleaves the target (Nykänen, A., et al., *Cell* **107**:309-321 (2001); Zamore, PD., et al., *Cell* **101**:25-33 (2000); Bernstein, E., et al., *Nature* **409**:363-366 (2001); Hamilton, A.J. & Baulcombe, D.C. *Science* **286**:950-952 (1999); Hammond, S.M. et al., *Nature* **404**:293-296 (2000)). Viral vectors expressing siRNA are useful in determining if similar mechanisms are involved in target RNA cleavage in mammalian cells *in vivo*.

In summary, these data demonstrate that siRNA expressed from viral vectors *in vitro* and *in vivo* specifically reduce expression of stably expressed plasmids in cells, and endogenous transgenic targets in mice. Importantly, the application of virally expressed siRNA to various target alleles in different cells and tissues *in vitro* and *in vivo* was demonstrated. Finally, the results show that it is possible to reduce polyglutamine protein levels in neurons, which is the cause of at least nine inherited neurodegenerative diseases, with a corresponding decrease in disease protein aggregation. The ability of viral vectors based on adeno-associated virus (Davidson, B.L., et al., *Proc. Natl. Acad. Sci. U. S. A.* **97**:3428-3432 (2000)) and lentiviruses (Brooks, A.I., et al., *Proc. Natl. Acad. Sci. U. S. A.* **99**:6216-6221 (2002)) to efficiently transduce cells in the CNS, coupled with the effectiveness of virally-expressed siRNA demonstrated here, extends the application of siRNA to viral-based therapies and to basic research, including inhibiting novel ESTs to define gene function.

All publications, patents and patent applications are incorporated herein by reference. While in the foregoing specification this invention has been described in relation to certain preferred embodiments thereof, and many details

have been set forth for purposes of illustration, it will be apparent to those skilled in the art that the invention is susceptible to additional embodiments and that certain of the details described herein may be varied considerably without departing from the basic principles of the invention.

References

Adelman *et al.*, DNA, 2, 183 (1983).

Alisky *et al.*, NeuroReport, 11,2669 (2000a).

Alisky *et al.*, Hum Gen Ther, 11,2315 (2000b).

Altschul *et al.*, JMB, 215, 403 (1990).

Altschul *et al.*, Nucleic Acids Res. 25, 3389 (1997).

Anderson *et al.*, Gene Ther., 7, 1034 (2000).

Andreason and Evans, Biotechniques, 6, 650 (1988).

Baulcombe, Plant Mol. Biol., 32, 79 (1996).

Bass, Nature, 411, 428 (2001).

Batzer *et al.*, Nucl. Acids Res.,19, 508 (1991).

Behr *et al.*, Proc. Natl. Acad. Sci. USA, 86, 6982 (1989).

Bledsoe *et al.*, NatBiot, 18, 964 (2000).

Bernstein *et al.*, Nature, 409, 363 (2001).

Brantl, Biochemica and Biophysica Acta, 1575, 15 (2002).

Brash *et al.*, Molec. Cell. Biol., 7, 2031 (1987).

Brooks *et al.*, Proc. Natl. Acad. Sci. U. S. A., 99,6216 (2002).

Brummelkamp *et al.*, Science, 296, 550-553 (2002).

Capecchi, Cell, 22, 479 (1980).

Caplan *et al.*, Proc. Natl. Acad. Sci. U. S. A., 98, 9742 (2001).

Chai *et al.*, J. Neurosci., 19, 10338 (1999).

Cogoni *et al.*, Antonie Van Leeuwenhoek, 65, 205 (1994).

Corpet *et al.*, Nucl. Acids Res., 16, 10881 (1988).

Crea *et al.*, Proc. Natl. Acad. Sci. U.S.A., 75, 5765 (1978).

Davidson *et al.*, Proc. Natl. Acad. Sci. U. S. A., 97, 3428 (2000).

Dayhoff *et al.*, Atlas of Protein Sequence and Structure (Natl. Biomed. Res. Found. 1978)

Elbashir *et al.*, Nature, 411, 494 (2001).

Elbashir *et al.*, Genes and Development, 15, 188 (2001).

Felgner *et al.*, Proc. Natl. Acad. Sci., 84, 7413 (1987).

Gelfand, PCR Strategies, Academic Press (1995).

Goeddel *et al.*, Nucleic Acids Res., 8, 4057 (1980).

Hamilton and Baulcombe, Science, 286, 950 (1999).

Hammond *et al.*, Nature, 404, 293 (2000).

Higgins *et al.*, Gene, 73, 237 (1988).

Hilberg *et al.*, Proc. Natl. Acad. Sci. USA, 84, 5232 (1987).

Higgins *et al.*, CABIOS, 5, 151 (1989).

Holland *et al.*, Proc. Natl. Acad. Sci. USA, 84, 8662 (1987).

Huang *et al.*, CABIOS, 8, 155 (1992).

Innis *et al.*, PCR Protocols, Academic Press (1995).

Innis and Gelfand, PCR Methods Manual, Academic Press (1999).

Johnston,, Nature, 346, 776 (1990).

Kriegler, M. Gene Transfer and Expression, A Laboratory Manual, W.H. Freeman Co, New York, (1990).

Kunkel, Proc. Natl. Acad. Sci. USA, 82, 488 (1985).

Kunkel *et al.*, Meth. Enzymol., 154, 367 (1987).

Karlin and Altschul, Proc. Natl. Acad. Sci. USA, 87, 2264 (1990).

Karlin and Altschul, Proc. Natl. Acad. Sci. USA, 90, 5873 (1993)

Kennerdell and Carthew, Cell, 95, 1017 (1998).

Johnston, Nature, 346, 776 (1990).

Lai *et al.*, Proc. Natl. Acad. Sci. USA, 86, 10006 (1989).

Larrick, J. W. and Burck, K. L., Gene Therapy. Application of Molecular Biology, Elsevier Science Publishing Co., Inc., New York, p. 71-104 (1991).

Lawn *et al.*, Nucleic Acids Res., 9, 6103 (1981).

Lee, N.S., et al., *Nat. Biotechnol.* **19**:500-505 (2002).

Loeffler *et al.*, J. Neurochem., 54, 1812 (1990).

Manche *et al.*, Mol. Cell Biol., 12, 5238 (1992).

Margolis and Ross, Trends Mol. Med., 7, 479 (2001).

Meinkoth and Wahl, Anal. Biochem., 138, 267 (1984).

Methods in Molecular Biology, 7, Gene Transfer and Expression Protocols, Ed. E. J. Murray, Humana Press (1991).

Miller, *et al.*, Mol. Cell. Biol., 10, 4239 (1990).

Minks *et al.*, J. Biol. Chem., 254, 10180 (1979).

Miyagishi, M. & Taira, K. *Nat. Biotechnol.* **19**:497-500 (2002).

Moulder *et al.*, J. Neurosci., 19, 705 (1999).

Murray, E. J., ed. Methods in Molecular Biology, Vol. 7, Humana Press Inc., Clifton, N.J., (1991).

Myers and Miller, CABIOS, 4, 11 (1988).

Needleman and Wunsch, JMB, 48, 443 (1970).

Nykänen *et al.*, Cell, 107, 309 (2001).

Ohtsuka *et al.*, JBC, 260, 2605 (1985).

Okabe *et al.*, FEBS Lett., 407, 313 (1997).

Ooboshi *et al.*, Arterioscler. Thromb. Vasc. Biol., 17, 1786 (1997).

Paul, C.P., et al., *Nat. Biotechnol.* **19**:505-508 (2002).

Pearson *et al.*, Meth. Mol. Biol., 24, 307 (1994).

Pearson and Lipman, Proc. Natl. Acad. Sci. USA, 85, 2444 (1988).

Quantin, B., *et al.*, Proc. Natl. Acad. Sci. USA, 89, 2581 (1992).

Rossolini *et al.*, Mol. Cell. Probes, 8, 91 (1994).

Rosenfeld, M. A., *et al.*, Science, 252, 431 (1991).

Sambrook and Russell, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press Cold Spring Harbor, NY (2001).

Scharfmann *et al.*, Proc. Natl. Acad. Sci. USA, 88, 4626 (1991).

Shipley *et al.*, J. Biol. Chem., 268, 12193 (1993).

Smith *et al.*, Adv. Appl. Math., 2, 482 (1981).

Stein *et al.*, J. Virol., 73, 3424 (1999).

Svoboda *et al.*, Development, 127, 4147 (2000).

Ternin, H., "Retrovirus vectors for gene transfer", in Gene Transfer, Kucherlapati R, Ed., pp 149-187, Plenum, (1986).

Timmons and Fire, Nature, 395, 854 (1998).

Tijssen, Laboratory Techniques in Biochemistry and Molecular Biology Hybridization with Nucleic Acid Probes, part I chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays" Elsevier, New York (1993).

Turner *et al.*, Mol. Biotech., 3, 225 (1995).

Valerio *et al.*, Gene, 84, 419 (1989).

Viera *et al.*, Meth. Enzymol., 153, 3 (1987).

Walker and Gaastra, Techniques in Mol. Biol. (MacMillan Publishing Co. (1983).

Waterhouse *et al.*, Proc. Natl. Acad. Sci. U. S. A., 95, 13959 (1998).

Wianny and Zernicka-Goetz, Nat. Cell Biol., 2, 70 (2000).

Xia *et al.*, Nat. Biotechnol., 19, 640 (2001).

Yang *et al.*, Mol. Cell Biol., 21, 7807 (2001).

Zamore *et al.*, Cell, 101, 25 (2000).

WHAT IS CLAIMED IS:

1. A viral vector comprising an expression cassette, wherein the expression cassette comprises a nucleic acid encoding a small interfering RNA molecule (siRNA) targeted against a gene of interest.

2. A viral vector comprising an expression cassette, wherein the expression cassette comprises a nucleic acid encoding a first segment, a second segment located immediately 3' of the first segment, and a third segment located immediately 3' of the second segment, wherein the first and third segments are each less than 30 base pairs in length and each more than 10 base pairs in length, and wherein the third segment is complementary to the first segment, and wherein a product nucleic acid generated by the expression cassette functions as a small interfering RNA molecule (siRNA) targeted against a gene of interest.

3. The vector of claim 1 or 2, wherein the siRNA forms a hairpin structure comprising a duplex structure and a loop structure.

4. The vector of claim 3, wherein the loop structure contains from 4 to 10 nucleotides.

5. The vector of claim 3, wherein the loop structure contains 4, 5 or 6 nucleotides.

6. The vector of claim 3, wherein the duplex is less than 30 nucleotides in length.

7. The vector of claim 3, wherein the duplex contains from 19 to 25 nucleotides.

8. The vector of claim 1 or 2, wherein the siRNA further comprises an overhang region.

9. The vector of claim 1 or 2, wherein the siRNA further comprises a 3' overhang region, a 5' overhang region, or both 3' and 5' overhang regions.

10. The vector of claim 8 or 9, wherein the overhang region is from 1 to 10 nucleotides in length.

11. The vector of any of the preceding claims, wherein the expression cassette further comprises a promoter.

12. The vector of claim 11, wherein the promoter is a regulatable promoter.

13. The vector of claim 11, wherein the promoter is a constitutive promoter.

14. The vector of claim 11, wherein the promoter is a CMV, RSV, or pol III promoter.

15. The vector of claim 11, wherein the promoter is a CMV or RSV promoter.

16. The vector of claim 11, wherein the promoter is not a pol III promoter.

17. The vector of any of the preceding claims, wherein the expression cassette further comprises a polyadenylation signal.

18. The vector of claim 17, wherein the polyadenylation signal is a synthetic minimal polyadenylation signal.

19. The vector of any of the preceding claims, further comprising a marker gene.

20. The vector of any of the preceding claims, wherein the vector is an adenoviral, lentiviral, adeno-associated viral (AAV), poliovirus, HSV, or murine Maloney-based viral vector.

21. The vector of any of the preceding claims, wherein the vector is an adenoviral vector.

22. The vector of any of the preceding claims, wherein the gene of interest is a gene associated with a condition amenable to siRNA therapy.

23. The vector of clam 22, wherein the condition amenable to siRNA therapy is a neurodegenerative disease.

24. The vector of claim 23, wherein the neurodegenerative disease is a trinucleotide-repeat disease.

25. The vector of claim 24, wherein the trinucleotide-repeat disease is a disease associated with polyglutamine repeats.

26. The vector of claim 25, wherein the trinucleotide-repeat disease is Huntington's disease or spinocerebellar ataxia.

27. The vector of any of the preceding claims, wherein the gene of interest encodes a ligand for a chemokine involved in the migration of a cancer cell, or a chemokine receptor.

28. The vector of claim 11, wherein the promoter is a pol II promoter.

29. The vector of claim 2, wherein the expression cassette further comprises a pol II promoter operably-linked to the nucleic acid encoding the first, second, and third segments.

30. The vector of claim 28 or 29, wherein the pol II promoter is regulatable.

31. The vector of claim 28 or 29, wherein the pol II promoter is a constitutive promoter.

32. The vector of claim 28 or 29, wherein the pol II promoter is a CMV or RSV promoter.

33. The vector of claim 28 or 29, wherein the pol II promoter is a CMV promoter.

34. The vector of claim 28 or 29, wherein the promoter is a RSV promoter.

35. A method of reducing the expression of a gene product in a cell, comprising contacting a cell with a viral vector comprising an expression cassette, wherein the expression cassette comprises a nucleic acid encoding a small interfering RNA molecule (siRNA) targeted against a gene, wherein expression from the targeted gene is reduced.

36. A method of reducing the expression of a gene product in a cell, comprising contacting a cell with viral vector comprising an expression cassette, wherein the expression cassette comprises a nucleic acid encoding a first segment, a second segment located immediately 3' of the first segment, and a third segment located immediately 3' of the second segment, wherein the first and third segments are each less than 30 base pairs in length and each more than 10 base pairs in length, and wherein the third segment is complementary to the first segment, and wherein a product nucleic acid generated by the expression cassette functions as a small interfering RNA molecule (siRNA) targeted against a gene of interest.

37. The method of claim 35 or 36, wherein the siRNA forms a hairpin structure comprising a duplex structure and a loop structure.

38. The method of claim 37, wherein the loop structure contains from 4 to 10 nucleotides.

39. The method of claim 37, wherein the loop structure contains 4, 5 or 6 nucleotides.

40. The method of claim 37, wherein the duplex is less than 30 nucleotides in length.

41. The method of claim 37, wherein the duplex is from 19 to 25 nucleotides in length.

42. The method of claim 35 or 36, wherein the siRNA further comprises an overhang region.

43. The method of claim 35 or 36 wherein the siRNA further comprises a 3' overhang region, a 5' overhang region, or both 3' and 5' overhang regions.

44. The method of claim 42 or 43, wherein the overhang region is from 1 to 10 nucleotides in length.

45. The method of any of claims 35-44, wherein the expression cassette further comprises a promoter.

46. The method of claim 45, wherein the promoter is a regulatable promoter.

47. The method of claim 45, wherein the promoter is a constitutive promoter.

48. The method of claim 45, wherein the promoter is a CMV, RSV, or pol III promoter.

49. The method of claim 45, wherein the promoter is a CMV or RSV promoter.

50. The method of claim 45, wherein the promoter is not a pol III promoter.

51. The method of any of claims 35-50, wherein the expression cassette further comprises a polyadenylation signal.

52. The method of claim 51, wherein the polyadenylation signal is a synthetic minimal poyladenylation signal.

53. The method of any of claims 35-52, wherein the expression cassette further comprises a marker gene.

54. The method of any of claims 35-53, wherein the vector is an adenoviral, lentiviral, adeno-associated viral (AAV), poliovirus, HSV, or murine Maloney-based viral vector.

55. The method of any of claims 35-54, wherein the vector is an adenoviral vector.

56. The method of any of claims 35 or 36, wherein the expression is from a gene associated with a neurodegenerative disease.

57. The method of claim 56, wherein the neurodegenerative disease is a trinucleotide-repeat disease.

58. The method of claim 57, wherein the trinucleotide-repeat disease is a disease associated with polyglutamine repeats.

59. The method of claim 58, wherein the trinucleotide-repeat disease is Huntington's disease or spinocerebellar ataxia.

60. The method of claim 35 or 36, wherein the gene of interest encodes a ligand for a chemokine involved in the migration of a cancer cell, or a chemokine receptor.

61. The method of claim 45, wherein the promoter comprises a pol II promoter

62. The method of claim 36, wherein the expression cassette further comprises a pol II promoter operably-linked to the nucleic acid encoding the first, second and third segments.

63. The method of claim 61 or 62, wherein the pol II promoter is regulatable.

64. The method of claim 61 or 62, wherein the pol II promoter is a constitutive promoter.

65. The method of claim 61 or 62, wherein the pol II promoter is a CMV or RSV promoter.

66. The method of claim 65, wherein the pol II promoter is a CMV promoter.

67. The method of claim 65, wherein the pol II promoter is a RSV promoter.

68. A method of treating a patient, comprising administering to the patient a composition comprising a viral vector, wherein the vector comprises an expression cassette, wherein the expression cassette comprises a nucleic acid encoding a small interfering RNA molecule (siRNA) targeted against a gene, wherein expression from the targeted gene is reduced by the siRNA.

69. A method of treating a patient, comprising administering to the patient a composition comprising a viral vector, wherein the vector comprises an expression cassette, wherein the expression cassette comprises a nucleic acid encoding a first segment, a second segment located immediately 3' of the first segment, and a third segment located immediately 3' of the second segment, wherein the first and third segments are each less than 30 base pairs in length and each more than 10 base pairs in length, and

wherein the third segment is complementary to the first segment, and wherein a product nucleic acid generated by the expression cassette functions as a small interfering RNA molecule (siRNA) targeted against a gene of interest.

70. A method of treating a patient, comprising administering to the patient a composition comprising a vector, wherein the vector comprises an expression cassette, wherein the expression cassette comprises a pol II promoter operably-linked to a nucleic acid encoding a small interfering RNA molecule (siRNA) targeted against a gene, wherein expression from the targeted gene is reduced.

71. A method of treating a patient, comprising administering to the patient a composition comprising a vector, wherein the vector comprises an expression cassette, wherein the expression cassette comprises a pol II promoter operably-linked to a nucleic acid encoding a first segment, a second segment located immediately 3' of the first segment, and a third segment located immediately 3' of the second segment, wherein the first and third segments are each less than 30 base pairs in length and each more than 10 base pairs in length, and wherein the third segment is complementary to the first segment, and wherein a product nucleic acid generated by the expression cassette functions as a small interfering RNA molecule (siRNA) targeted against a gene of interest.

72. A method of performing allele-specific gene silencing comprising administering an expression cassette comprising a pol II promoter operably-linked to a nucleic acid encoding a small interfering RNA molecule (siRNA) targeted against a gene of interest, wherein the siRNA silences only one allele of a gene.

73. A method of performing allele-specific gene silencing in a mammal comprising administering to the mammal a vector comprising an expression cassette, wherein the expression cassette comprises a nucleic

acid encoding a small interfering RNA molecule (siRNA) targeted against a gene of interest, wherein the siRNA silences only one allele of a gene.

74. The method of claim 72 or 73, wherein alleles of the gene differ by seven or fewer base pairs out of 21 base pairs.

75. The method of claims 72 or 73, wherein the gene is a beta-glucuronidase gene.

76. The method of claim 74, wherein the alleles are murine-specific and human-specific alleles of beta-glucuronidse.

77. The method of claim 72 or 73, wherein the expression cassette further comprises a pol II promoter operably-linked to the nucleic acid encoding the siRNA.

78. The method of claims 70-77, wherein the vector is a viral vector.

79. An expression cassette comprising a pol II promoter operably linked to a nucleic acid encoding a small interfering RNA molecule (siRNA) targeted against a gene of interest.

80. An expression cassette comprising a pol II promoter operably linked to a nucleic acid encoding a first segment, a second segment located immediately 3' of the first segment, and a third segment located immediately 3' of the second segment, wherein the first and third segments are each less than 30 base pairs in length and each more than 10 base pairs in length, and wherein the third segment is complementary to the first segment, and wherein a product nucleic acid generated by the expression cassette functions as a small interfering RNA molecule (siRNA) targeted against a gene of interest.