Improved expression elements

BACKGROUND

The present invention relates to polynucleotides comprising elements conferring improved expression on operably-linked transcription units. These elements are naturally associated with the promoter regions of ribosomal protein genes and, in recombinant DNA constructs, confer high and reproducible levels of gene expression. The present invention also relates to vectors comprising such polynucleotide sequences, host cells comprising such vectors and use of such polynucleotides, vectors or host cells in therapy, for production of recombinant proteins in cell culture and other biotechnological applications.

The current model of chromatin structure in higher eukaryotes postulates that genes are organised in "domains" (Dillon, N. & Grosveld, F. Chromatin domains as potential units of eukaryotic gene function. *Curr. Opin. Genet. Dev. 4,* 260-264 (1994); Higgs, D.R. Do LCRs open chromatin domains? *Cell* 95, 299-302 (1998)) Chromatin domains are envisaged to exist in either a condensed, "closed", transcriptionally silent state, or in a de-condensed, "open" and transcriptionally competent configuration. The establishment of an open chromatin structure characterised by increased DNaseI sensitivity, DNA hypomethylation and histone hyperacetylation, is considered a pre-requisite to the commencement of gene expression.

The open and closed nature of chromatin regions is reflected in the behaviour of transgenes that are randomly integrated into the host cell genome. Identical constructs give different patterns of tissue-specific and development stage-specific expression when integrated at different locations in the mouse genome (Palmiter, R.D. & Brinster, R.L. *Ann. Ref. Genet.* 20, 465-499 (1986); Allen, N.D. *et al. Nature* 333, 852-855 (1988); Bonnerot, C., Grimber, G., Briand, P. & Nicolas, J.F. *Proc. Natl. Acad. Sci. USA* 87:6331-6335 (1990)).

The chromatin domain model of gene organisation suggests that genetic control elements that are able to establish and maintain a transcriptionally competent open chromatin structure should be associated with active regions of the genome.

Locus Control Regions (LCRs) are a class of transcriptional regulatory elements with long-range chromatin remodelling capability. LCRs are functionally defined in transgenic

mice by their ability to confer site-of-integration independent, transgene copy number-dependent, physiological levels of expression on a gene linked in *cis*, especially single copy transgenes Fraser, P. & Grosveld, F. *Curr. Opin. Cell Biol.* 10, 361-365 (1998); Li, Q., Harju, S. & Peterson, K.R. *Trends Genet.* 15: 403-408 (1999). Crucially, such expression is tissue-specific. LCRs are able to obstruct the spread of heterochromatin, prevent PEV (Kioussis, D. & Festenstein, R. *Curr. Opin. Genet. Dev.* 7, 614-619 (1997)) and consist of a series of DNase I hypersensitive (HS) sites which can be located either 5' or 3' of the genes that they regulate (Li, Q., Harju, S. & Peterson, K.R. *Trends Genet.* 15: 403-408 (1999)).

The generation of cultured mammalian cell lines producing high levels of a therapeutic protein product is a major developing industry. Chromatin position effects make it a difficult, time consuming and expensive process. The most commonly used approach to the production of such mammalian "cell factories" relies on gene amplification induced by a combination of a drug resistance gene (e.g., DHFR, glutamine synthetase (Kaufman RJ. *Methods Enzymol* 185, 537–566 (1990)). and the maintenance of stringent selective pressure. The use of vectors containing LCRs from highly expressed gene domains, using cells derived from the appropriate tissue, greatly simplifies the procedure, giving a large proportion of clonal cell lines showing stable high levels of expression (Needham M, Gooding C, Hudson K, Antoniou M, Grosveld F and Hollis M. *Nucleic Acids Res* 20, 997–1003 (1992); Needham M, Egerton M, Millest A, Evans S, Popplewell M, Cerillo G, McPheat J, Monk A, Jack A, Johnstone D and Hollis M. *Protein Expr Purif* 6,124–131 (1995).

However, the tissue-specificity of LCRs, although useful in some circumstances, is also a major limitation for many applications, for instance where no LCR is known for the tissue in which expression is required, or where expression in many, or all, tissues is required.

US 6,689,606 and co-pending patent application WO 00/0539, incorporated by reference herein, describe elements that are responsible, in their natural chromosomal context, for establishing an open chromatin structure across a locus that consists exclusively of ubiquitously expressed, housekeeping genes. These elements are not derived from an LCR and comprise extended methylation-free CpG islands.

In mammalian DNA, the dinucleotide CpG is recognised by a DNA methyltransferase enzyme that methylates cytosine to 5-methylcytosine. However, 5-methylcytosine is unstable and is converted to thymine. As a result, CpG dinucleotides occur far less frequently than one would expect by chance. Some sections of genomic DNA nevertheless do have a frequency of CpG that is closer to that expected, and these sequences are known as "CpG islands". As used herein a "CpG island" is defined as a sequence of DNA, of at least 200bp, that has a GC content of at least 50% and an observed / expected CpG content ratio of at least 0.6 (i.e. a CpG dinucleotide content of at least 60% of that which would be expected by chance) (Gardiner-Green M and Frommer M. *J Mol Biol* 196, 261–282 (1987); Rice P, Longden I and Bleasby A *Trends Genet* 16, 276–277 (2000).

Methylation-free CpG islands are well-known in the art (Bird *et al* (1985) Cell 40: 91-99, Tazi and Bird (1990) Cell 60: 909-920) and may be defined as CpG islands where a substantial proportion of the cytosine residues are not methylated and which usually extend over the 5' ends of two closely spaced (0.1-3 kb) divergently transcribed genes. These regions of DNA are reported to remain hypomethylated in all tissues throughout development (Wise and Pravtcheva (1999) Genomics 60: 258-271). They are often associated with the 5' ends of ubiquitously expressed genes, as well as an estimated 40% of genes showing a tissue-restricted expression profile (Antequera, F. & Bird, A. *Proc. Natl. Acad. Sci. USA* 90, 1195-11999 (1993); Cross, S.H. & Bird, A.P. *Curr. Opin, Genet. Dev. 5,* 309-314 (1995) and are known to be localised regions of active chromatin (Tazi, J. & Bird, A. *Cell* 60, 909-920 (1990).

An 'extended' methylation-free CpG island is a methylation-free CpG island that extends across a region encompassing more than one transcriptional start site and/or extends for more than 300bp and preferably more than 500bp. The borders of the extended methylation-free CpG island are functionally defined through the use of PCR over the region in combination with restriction endonuclease enzymes whose ability to digest (cut) DNA at their recognition sequence is sensitive to the methylation status of any CpG residues that are present. One such enzyme is *Hpa*II, which recognises and digests at the site CCGG, which is commonly found within CpG islands, but only if the central CG residues are not methylated. Therefore, PCR conducted with *Hpa*II-digested DNA and over a region harbouring *Hpa*II sites, does *not* give an amplification product due to *Hpa*II digestion if the DNA is unmethylated. The PCR will *only* give an amplified product if the DNA is methylated. Therefore, beyond the methylation-free region *Hpa*II will not digest

the DNA a PCR amplified product will be observed thereby defining the boundaries of the "extended methylation-free CpG island".

It has been shown (WO 00/05393) that regions spanning methylation-free CpG islands encompassing dual, divergently transcribed promoters from the human TATA binding protein (*TBP*)/proteosome component-B1 (*PSMBI)* and heterogeneous nuclear ribonucleoprotein A2/B1 (*hnRNPA2*)/heterochromatin protein 1Hsγ (*HP1* $^{Hs\gamma}$) gene loci give reproducible, physiological levels of gene expression and that they are able to prevent a variegated expression pattern and silencing that normally occurs with transgene integration within centromeric heterochromatin.

It is known that methylation-free CpG islands associated with actively transcribing promoters possess the ability to remodel chromatin and are thus thought to be a prime determinant in establishing and maintaining an open domain at housekeeping gene loci (WO 00/05393) and that such elements confer an increased proportion of productive gene delivery events with improvements in the level and stability of transgene expression.

Ribosomes are large RNA and protein complexes responsible for the translation of mRNA into polypeptides. Each ribosome is comprised of 4 ribosomal RNA (rRNA) molecules and large number of ribosomal proteins (currently thought to be 79 in mammalian cells). Ribosomal proteins have functions including facilitation of rRNA folding, protection from cellular ribonucleases, and co-ordinating protein synthesis. Some ribosomal proteins have additional extraribosomal functions (Wool, 1996, TIBS 21: 164–165). Given the structural and functional similarities of ribosomes across species, it is unsurprising that the amino acid sequence conservation of ribosomal proteins is high, and among mammals the sequences of most ribosomal proteins are almost identical (Wool *et al*, 1995, Biochem Cell Biol 73: 933–947.

Two ribosomal proteins appear atypical in that they are expressed in the form of propeptides (carboxy-extension proteins) fused to ubiquitin. Ubiquitin is a highly conserved 76-residue polypeptide involved in a variety of cellular functions, including the regulation of intracellular protein breakdown, cell cycle regulation and stress response (Hershko & Ciechanover, 1992, Annu Rev Biochem 61: 761–807; Coux *et al*, 1996, Annu Rev Biochem 65: 801–847).

Ubiquitin is encoded by two distinct classes of gene. One is a poly-ubiquitin gene encoding a linear polymer of ubiquitin repeats. The other comprises genes encoding natural fusion proteins in which a single ubiquitin molecule is linked to the ribosomal protein rps27A or rpL40 (Finley *et al*, 1989, Nature 338: 394–401; Chan *et al*, 1995, Biochem Biophys Res Commun 215: 682–690; Redman & Burris, 1996, Biochem J 315: 315–321).

The common structural features of ribosomal protein promoters are discussed by Perry (2005, BMC Evolutionary Biology 5: 15). The promoters may be classified according to the nature of the TATA box motifs, number and type of transcription factor binding sites and location of AUG start codons. However, such classification does not appear to predict promoter strength and evidence suggests that several such promoters tested have equivalent transcriptional activity as measured by expression of a linked reporter gene (Hariharan et al, 1989, Genes Dev 3: 1789–800).

US patent 6,063,598 discloses the hamster-ubiquitin/S27a promoter its use to drive high level production of recombinant proteins. However, there is no suggestion of its use to enhance the expression of a gene primarily transcribed from a further promoter (i.e one other than hamster-ubiquitin/S27a promoter).

US application US 2004/0148647 discloses a reporter assay using an expression vector comprising a hamster ubiquitin /S27A promoter functionally linked to a gene for a product of interest and a fluorescent protein reporter. Again, the application only discloses constructs in which transcription of gene of interest is from the hamster-ubiquitin/ S27a promoter itself.

It remains an objective in the field of recombinant gene expression to obtain higher and more reliable levels of expression, particularly for *in vivo* and *ex vivo* therapeutic applications and for in vitro recombinant protein production.

SUMMARY OF THE DISCLOSURE

Throughout the description and claims of this specification, the words "comprise" and "contain" and variations of the words, for example "comprising" and "comprises", means

"including but not limited to", and is not intended to (and does not) exclude other moieties, additives, components, integers or steps.

Throughout the description and claims of this specification, the singular encompasses the plural unless the context otherwise requires. In particular, where the indefinite article is used, the specification is to be understood as contemplating plurality as well as singularity, unless the context requires otherwise.

Features, integers, characteristics, compounds, chemical moieties or groups described in conjunction with a particular aspect, embodiment or example of the invention are to be understood to be applicable to any other aspect, embodiment or example described herein unless incompatible therewith.

Definitions

As used herein, a promoter region is defined as being a genomic nucleotide sequence consisting of a promoter and transcriptional start site together with 5kb of 5' sequence upstream of the transcriptional start site and 500bp 3' sequence downstream of the distal end of the first exon.

A 5' untranslated region means a region 5' of the translational start encoded in the genomic or cDNA sequence. It is taken to include all upstream regulatory elements. A 5' upstream sequence is used to mean sequence 5' to the transcriptional start encoded in the genomic sequence.

As used herein, 'transcribable nucleic acid' means a nucleic acid which, when operably linked to a functional promoter and other regulatory sequences, is capable of being transcribed to give a functional RNA molecule, such as mRNA. Such sequences may comprise open reading frames, which encode translatable polypeptide sequences. Alternatively, the functional RNA may have another function, such as ribosomal RNA, ribozymes or antisense RNA.

'Gene' is commonly taken to means the combination of the coding region of the transcribable nucleic acid, together with the promoter from which it is transcribed and other regulatory sequences such as enhancers and 3' polyadenylation signals. In genomic DNA genes also contain introns. 'Transcription unit' is sometimes used to describe a functional combination of at least a promoter and minimal regulatory

sequences with a transcribable nucleic acid, often derived from cDNA from which the introns have been spliced out. 'Cistron' is defined as a nucleic acid encoding a single polypeptide together with functional initiation and termination signals. 'Transgene' implies a gene that has been transferred from one genome to another, although the term may be more loosely applied to any gene or even transcribable nucleic acid comprised in a recombinant DNA construct such as a vector.

Promoter and enhancer are terms well known in the art and include the following features which are provided by example only, and not by way of limitation. Promoters are 5', *cis*-acting regulatory sequences directly linked to the initiation of transcription. Promoter elements include so-called TATA box and RNA polymerase initiation selection (RIS) sequences which function to select a site of transcription initiation. These sequences also bind polypeptides which function, *inter alia*, to facilitate transcription initiation selection by RNA polymerase.

In simple terms, promoters are directional elements that act to initiate transcription of sequences situated less than 100 (and usually less than 50) nucleotide base pairs (bp) downstream. They contain a number of short consensus nucleotide sequences that act as binding sites for various proteins that participate in the initiation of transcription and the assembly of a multi-subunit complex known as the pre-initiation complex (McKnight and Tjian, 1987, Cell 46: 795–805). In most genes, this occurs at a very widely conserved sequence known as the TATA box (TATAAA) to which the TATA box-binding protein (TBP, a subunit of the general transcription factor TFIID) binds. There follows an ordered assembly of more than ten further transcription factors to finally form the Pol II holoenzyme complex. RNA transcription actually starts at an initiator site about 25–30 bases downstream (Breathnach and Chambon, 1981, Annu Rev Biochem 50: 349–393) to which TBP also binds.

Most functional promoters contain further upstream promoter elements (UPEs), of which the most highly conserved are the CAAT box (CCAAT, the binding site for the transcription factors CBF, C/EBP and NF-1), about 70–200bp upstream, and the GC box (GGGCGG, binding site for the general transcription factor Sp-1) a similar distance upstream. Although basal levels of transcription occur from the TATA box alone, for most promoters at least the CAAT and GC boxes are required for optimal levels of transcription.

Enhancers are sequences that act non-directionally to increase transcription from promoters situated locally but not necessarily immediately adjacent (up to several kilobases away (Kadonaga (2004) Cell 116: 247–257). Enhancers contain short (8–12bp) consensus sequences representing the binding sites for a wide range of transcriptional activator proteins (Ondek *et al*, 1988, Science 236: 1237–1244) including some, such as NF-1 and SP-1 that are also associated with promoter elements. These sequences are often duplicated in tandem or inverted repeats.

In some natural transcription units, including the very active immediate/early gene transcription units of many DNA viruses such as cytomegalovirus, enhancer and promoter elements may be functionally combined into what is effectively one extended upstream element.

Promoters may be regulated, being responsive to cell type, temperature, metal ions or other factors; or constitutive, giving transcription that is unresponsive to such factors. For many purposes a strong, constitutive promoter giving consistent, high, levels of transcription in many, if not all, cell types is highly advantageous. For many years the enhancer/promoter element driving immediate/early gene expression in human cytomegalovirus has been very widely used for driving such expression of heterologous genes in eukaryotic expression vectors (Foecking & Hoffstetter, 1986, Gene 45: 101–105).

It was hypothesised that promoter regions of ribosomal protein genes might have useful activity in boosting and stabilising expression of linked transgenes and that the regulatory regions from highly expressed genes might be more likely to contain elements that are very effective at maintaining chromatin in a transcriptionally active conformation. The linking of such elements to a heterologous promoter might then generate a more open chromatin environment surrounding that promoter, resulting in increased expression. It will be understood by one of skill in the art that promoters derived from ribosomal protein genes are to be distinguished from ribosomal promoters, which are RNA polymerase Type I dependent promoters from which rRNA is transcribed.

To test the hypothesis, RNA was obtained from exponentially growing CHO-K1 and NS0 cell lines and a microarray analysis against 13,443 murine genes was performed. We limited our analysis to elements having a high CpG island content and a likelihood of bi-directional promoters. Using criteria based on the minimal effective sequence from the

hnRNPA2 regulatory region, approximately 3kb of DNA from a selection of these genes was amplified by PCR from NS0 genomic DNA. These sequences were then cloned into EGFP expression vectors and transfected into CHO-K1 along with hnRNPA2 control versions of the same vector.

It was found that sequences derived from the promoter regions of two ribosomal proteins gave consistently high levels of expression of the heterologous reporter sequence in the assay used. In each case, the promoter region comprised a GC-rich sequence extending from the 5' region upstream of the actual promoter elements well into the first exon and, indeed into the first intron. This GC-rich sequence fulfilled the criteria of being an extended CpG island, as defined herein as extending over 300bp.

Accordingly, the invention provides an isolated polynucleotide comprising

a) an element comprising at least 500 contiguous nucleotides from the promoter region of a ribosomal protein gene
b) a heterologous promoter
c) a transcribable nucleic acid sequence adjacent said heterologous promoter

wherein the transcribable nucleic acid sequence is transcribed from said heterologous promoter and the level of said transcription is enhanced by said element.
an element. Preferably said element comprises more than 1kb and most preferably more than 3kb 5' untranslated sequence from a ribosomal protein gene.

The contiguous nucleotides are selected from the promoter area, which extends from a point 5kb upstream (5' relative to the sense strand) of the transcriptional start site to a point 500bp downstream (3' relative to the sense strand) of the distal (3') end of the first exon.

Preferably said ribosomal protein gene is selected from the list consisting of RPSA, RPS2, RPS3, RPS3A, RPS4, RPS5, RPS6, RPS7, RPS8, RPS9, RPS10, RPS11, RPS12, RPS13, RPS14, RPS15, RPS15A, RPS16, RPS17, RPS18, RPS19, RPS20, RPS21, RPS23, RPS24, RPS25, RPS26, RPS27, RPS27A, RPS28, RPS29, RPS30, RPL3, RPL4, RPL5, RPL6, RPL7, RPL7A, RPL8, RPL9, RPL10, RPL10A, RPL11, RPL12, RPL13, RPL13A, RPL14, RPL15, RPL17, RPL18, RPL18A, RPL19, RPL21, RPL22, RPL23, RPL23A, RPL24, RPL26, RPL27, RPL28, RPL29, RPL30, RPL31, RPL32, RPL34, RPL35, RPL35A, RPL36, RPL36A, RPL37, RPL37A, RPL38, RPL39,

RPL41, RPLP0, RPLP1 and RPLP2 and their orthologues. More preferably it is RPS3 or RPS11.

In one preferred embodiment the element comprises a CpG island, preferably an extended CpG island of at least 300bp and more preferably 500bp. Preferably the CpG island is unmethylated. It is also preferred that said element comprises the promoter from a ribosomal protein gene and from which transcription of the ribosomal protein gene is naturally initiated. Such a promoter is often referred to as the endogenous promoter. In a preferred embodiment, the element further comprises or more exons of said ribosomal protein gene.

It is preferred that the ribosomal protein gene is a mammalian gene, although such genes and their promoters and 5' upstream sequences are highly conserved across species and might alternatively be an insect, nematode or yeast gene. Preferably, however, it is a human or rodent gene and most preferably it is a mouse gene.

In a highly preferred embodiment, the isolated polynucleotide of the invention comprises nucleotides 38 to 3154 of the mouse rps3 nucleotide sequence depicted in SEQ ID NO:1. Alternatively it comprises nucleotides 12 to 3032 of the mouse rps11 nucleotide sequence as listed in SEQ ID NO:2.

In one aspect, in addition to the element described, the polynucleotide further comprises a promoter not naturally associated with said element from a ribosomal protein gene. In this embodiment a heterologous promoter (distinct from the endogenous promoter which may or may not be present in the first element) is situated in an adjacent and operably-linked position downstream from the element containing ribosomal protein gene-derived 5' sequence. In this arrangement, expression directed by this heterologous promoter is enhanced by the effect of the ribosomal protein gene element.

In one embodiment, said promoter is a constitutive promoter, more preferably selected from the list consisting of the cytomegalovirus early/immediate promoter, SV40, EF-1α, Rous sarcoma virus (RSV) LTR, or HIV2 LTR or combinations of sequences derived therefrom. More preferably the promoter is a CMV immediate/early promoter. Most preferably it is the mouse or guinea pig CMV immediate/early promoter.

Alternatively, said promoter may be a tissue-specific promoter, which directs expression in a limited range of tissues. Such promoters are well-known in the art and include those from β-globin, the κ and λ immunoglobulin light chains, immunoglobulin heavy chain, desmin, tyrosinase, CD2, IL-3, myosin light chain, human melanoma inhibitory activity gene promoter and keratins. In a particularly preferred embodiment the promoter is a tumour-selective promoter, which directs expression preferentially one or more tumour types. Examples of such promoters include those from carcino-embryonic antigen (CEA), prostate-specific antigen (PSA), cyclooxygenase-2 (COX-2), alpha-fetoprotein (AFP), tyrosinase, and T-cell Factors 1–4 (TCF) -based promoters.

The transcribable nucleic acid may encode any useful polypeptide for *in vitro* expression and is preferably selected from the list consisting of an antibody, a functional epitope-binding fragment of an antibody, growth factor, cytokine, protein kinase, soluble receptor, membrane-bound receptor, or blood clotting factor. Alternatively, the transcribable nucleic acid may encode a therapeutic gene of use for *in vivo* or *ex vivo* gene therapy. Such a therapeutic nucleic acid may act by replacing or supplementing the function of a defective gene causing a disease such as cystic fibrosis, thalassaemia, sickle anaemia, Fanconi's anaemia, haemophilia, severe combined immunodeficiency (SCID), phenylketonuria (PKU), alpha-1 antitrypsin deficiency, Duchenne muscular dystrophy, ornithine transcarbamylase deficiency or osteogenesis imperfecta. Alternatively, it may encode a cytotoxic agent or prodrug-converting enzyme selectively expressed in a target cell, such as a malignant cancer cell, in order to kill it. Such applications, and many others, are well-known to those of skill in the art and the relevance of the current invention in enhancing the expression of therapeutic nucleic acids will be clear to such skilled practitioners.

In another aspect, the invention provides a vector comprising the polynucleotide of invention as disclosed above. Preferably said vector is an expression vector adapted for eukaryotic gene expression.

Typically said adaptation includes, by example and not by way of limitation, the provision of transcription control sequences (promoter sequences) which mediate cell/tissue specific expression. Adaptations also include the provision of selectable markers and autonomous replication sequences which both facilitate the maintenance of said vector in either the eukaryotic cell or prokaryotic host. Vectors which are maintained autonomously are referred to as episomal vectors. Episomal vectors are desirable since

they are self-replicating and so persist without the need for integration. Episomal vectors of this type are described in WO98/07876.

Adaptations which facilitate the expression of vector encoded genes include the provision of transcription termination/polyadenylation sequences. This also includes the provision of internal ribosome entry sites (IRES) which function to maximise expression of vector encoded genes arranged in bicistronic or multi-cistronic expression cassettes.

These adaptations are well-known in the art. There is a significant amount of published literature with respect to expression vector construction and recombinant DNA techniques in general. Please see, Sambrook et al (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbour Laboratory, Cold Spring Harbour, NY and references therein; Marston, F (1987) DNA Cloning Techniques: A Practical Approach Vol III IRL Press, Oxford UK; DNA Cloning: F M Ausubel et al, Current Protocols in Molecular Biology, John Wiley & Sons, Inc.(1994).

The vector may be an episomal vector or an integrating vector. Preferably, the vector is a plasmid. Alternatively, the vector may be a virus, such as an adenovirus, adeno-associated virus, a herpesvirus, vaccinia virus, lentivirus or other retrovirus.

Alternatively, such a vector may comprise

a) an element comprising at least 500 contiguous nucleotides from the promoter region of a ribosomal protein gene
b) a heterologous promoter
c) a multiple cloning site

wherein a transcribable nucleic acid sequence inserted into said multiple cloning site is capable of being transcribed from said heterologous promoter and the level of said transcription is enhanced by said element.

In a further aspect the invention provides a host cell comprising the isolated polynucleotide or vector as herein described. Preferably said host cell is a mammalian cell, more preferably selected from the list consisting of CHO, NS0, BHK, HeLa, HepG2.

Also provided by the invention is a method of expressing a polypeptide comprising inserting expression vector comprising the polynucleotide of the invention into an appropriate host cell as herein described and culturing said host cell in suitable

conditions to allow expression. Preferably said polypeptide is a therapeutically useful polypeptide.

In a further aspect the invention provides a pharmaceutical preparation comprising the polynucleotide, vector or host cell as herein described and a pharmaceutically acceptable carrier, excipient, buffer or medium.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a plasmid map of vector rps3- 1005-EGFP (see Example 1)
Figure 2 shows a plasmid map of vector rps11-1005-EGFP (see Example 2)
Figure 3 shows the expression of the EGFP reporter as expressed by various rps3 constructs in CHO-K1 cells analysed by FACS analysis, 8 days post transfection. Figure A shows mean fluorescence, and Figure B a indicates the percentage of cells expressing the reporter gene to a detectable level (% positive cells). See Example 1.
Figure 4 shows reporter gene expression of rps11 constructs in CHO-K1 cells 7 days post-transfection (analysed by FACS). A and C are total counts, B to E are results based on just the expressing cells within the population. Figure A shows the mean fluorescence of cells in the stably selected pool, and Figure B shows percentage (%) positive cells. See Example 2.
Figure 5 shows the expression level of a reporter gene in stably transfected NS0 cells when it is driven by the hCMV promoter with no additional elements present or when either an 8kb hnRNPA2 or a 3kb RPS3 element are placed immediately 5' to the hCMV promoter. Figure A shows the mean fluorescence intensity of the stable pools at 28 days and Figure B shows the percentage (%) positive cells.
Figure 6 shows similar data to Figure 5 for rps11 constructs. Figure A shows the mean fluorescence intensity values for stable pools generated with either hCMV driven constructs or identical constructs with a 8kb hnRNPA2 or 3kb RPS3 elements immediately 5' to the promoter. Figure B shows the percentage of cells within the pool expressing the reporter gene (percentage (%) positive cells).

DETAILED DESCRIPTION

Materials and Methods

Microarray analysis.

Total RNA was extracted from ~80% confluent CHO-K1 cells using RNeasy RNA extraction kit (Qiagen, Crawley, UK) according to the manufacturers protocol. Total RNA (2μg/μl) was subjected to microarray expression analysis using the mouse 70-mer oligonucleotide library (Operon V.1) representing 13,443 known transcripts. The University of Cincinnati, Genomics and Microarray Laboratory undertook Microarray analysis according to referenced protocols (http://microarray.uc.edu).

Gene transcript sequences were ranked according to increasing fluorescence. Since our previous study detailed the *HNRPA2B1/CBX3* loci as a chromatin-remodelling element and conferring benefit to *hCMV* the *HNRPA2* transcript was identified as the baseline expression level. However, using the available microarray analysis the *HNRPA2* transcript was barely detectable. Since the expression level of the *HNRPA2* transcript was minimal, using *HNRPA2* as our reference would have identified 3829 sequences for potential analysis. Therefore, 7 sequences from the top 2% (76 sequences) of the ordered, expressed transcripts were identified according to the criteria of containing a CpG island and one or more putative/known transcriptional start sites (see Table 1). CpG islands position, size and GC: CG ratios were verified using GrailEXP (http://compbio.ornl.gov). Putative/known transcriptional starts sites were identified from NIX blast analysis (http://www.hgmp.mrc.ac.uk) and Ensembl databases (http://www.ensembl.org).

PCR amplification of CpG island-containing fragments.

PCR oligonucleotides were designed to amplify approximately 3kb fragments encompassing the complete CpG island embedded promoter region whilst including approximately 500bp of coding sequence according to known or predicted coding sequence structure (see Table 2).

PCR reactions contained oligonucleotide sets specific each genomic fragment (2pmol of each primer; Table 2). PCR amplification was achieved using the Failsafe™ PCR premixes A-F (Cambio, UK), 1unit *Taq* DNA polymerase (Promega, UK) and 200ng of template DNA. Initial denaturation was 96°C for 2 min, whilst PCR amplification

was carried out for 35 cycles (94°C for 1min, 55-60°C for 1 min, 72°C for 5min). A final extension step (72°C for 10min) was included.

PCR products were gel purified, using GFX DNA purification columns (Amersham, UK) according to the manufacturers protocol, and subjected to TOPO TA cloning® according to the manufacturers protocol (TOPO; Invitrogen, UK). Sense and anti-sense orientations were obtained for each CpG island-containing fragment cloned into TOPO vectors (Invitrogen, UK).

Expression vector construction.

A control expression vector (designated CET1005EGFP, SEQ ID NO:20) was constructed by the insertion of an *hCMV/EGFP/sv40* pA (Nhe*I*/ Age*I* deleted multiple cloning site) from pEGFP-N1 into CET 900 followed by the insertion of the Asc*I* cassette from this vector into the Asc*I* site of CET 1005.

All CpG island fragments were removed from TOPO2.1 (Invitrogen, UK) unless otherwise stated. *Terf2ip* Acc*65I*/Eco*RV* fragment was inserted into Acc*65I*/Swa*I* of 1005. *GAPDH* Spe*I*/Sna*BI* was inserted into Pme*I*/Xba*I* of 1005. *RPS3* Xba*I*/Spe*I* fragment was inserted into Xba*I* of 1005. *RPS11* and *TUBA1* Eco*RI* blunt fragments was removed from TOPO4.0 and TOPO2.1 respectively (Invitrogen, UK) and inserted into Pme*I* of 1005. Finally, *A430106P18Rik* (Eco*RV*) and *2510006D16Rik* (Bst*XI*) fragments were also inserted into Pme*I* of 1005. All CpG island containing fragments were inserted in both sense and anti-sense orientations immediately upstream of the *hCMV* promoter.

Cell lines and transfections.

CHO-K1 cells were grown in HAMS F12 (Invitrogen, Paisley, UK) plus 4500 mg/l L-ananyl-L-glutamine, 10 μg/ml each of penicillin and streptomycin, and 10% (v/v) heat inactivated foetal calf serum (FCS; Invitrogen, Paisley, UK). Transfection was carried out by electroporation using approximately 10^7 cells from 80% confluent cultures and a BioRad Gene Pulser II ™ set to deliver a single pulse of 975μF at 250V. Transfections used 2μg of linearised CET1005EGFP plasmid and equivalent molar quantities for expression vectors of different size. Stably transfected cells were selected and maintained in growth medium containing 12.5μg/ml puromycin sulphate (Sigma, UK).

Quantification of transgene expression

Analysis of cells transfected with *EGFP* reporter constructs was with a Becton-Dickinson

FACScan using the parental CHO-K1 cell line as a background, autofluorescence control.

Table 1. **Sequences analysed**

Locus	Acc. #[a]	Description[b]	CpG island[c]		
			bp	% CG	GC/CG
Terf2ip	AB041557	Telomeric repeat binding factor 2 interacting protein 1 (*TRF2*- interacting telomeric protein Rap1).	968	64.56	0.90
Gapd[d]	M32599	Glyceraldehyde-3-phosphate dehydrogenase	1187	60.50	0.84
RPS3	NM012052	*RPS3* - 40S ribosomal protein S3.	419	60.39	0.87
TUBA1	M13445	Tubulin alpha-1 chain (Alpha-tubulin 1).	850	66.30	0.91
RPS11	AK011207	*RPS11* - 40S ribosomal protein S11.	957	59.71	0.95
A430106P18Rik	AK020778	Expressed sequence tag	982	63.62	0.93
2510006D16Rik	AK010915	Expressed sequence tag	679	67.69	0.75

[a]Genbank Accession

[b]Enseml description (http://www.ensembl.org/)

[c]Grailexp (http://compbio.ornl.gov/grailexp)

[d]Gapd – derived from human sequence

Table 2. **PCR oligonucleotides and amplicon sizes**

Locus	Sense	Antisense	Amplicon
Terf2ip	gtagtttctgacttggaaatgt (SEQ ID NO: 3)	aactgacctgccatgccattc (SEQ ID NO: 4)	2995 bp
Gapd	gagcagtccggtgtcacta (SEQ ID NO: 5)	gcagagaagcagacagttatg (SEQ ID NO: 6)	3096 bp
RPS3	cagagcatcaagtacctgtga (SEQ ID NO: 7)	taaccactaagccatctctcc (SEQ ID NO: 8)	3056 bp
TUBA1	caagaacaaggaagctggcc (SEQ ID NO: 9)	taaaacccacagcactgtaggg (SEQ ID NO: 10)	3049 bp
RPS11	aagactgtttgcctcatgcc (SEQ ID NO: 11)	ggatgacaatggtcctctgc (SEQ ID NO: 12)	3020 bp
A430106P18Rik	atggttgtaggttcacgtcc (SEQ ID NO: 13)	atccctcacattgccaagcc (SEQ ID NO: 14)	3128 bp
2510006D16Rik	acttaagacctgatgcctcc (SEQ ID NO: 15)	gctagcttacataggcagcc (SEQ ID NO: 16)	2997 bp

Example 1: rps3 element driven expression

SEQ ID NO:1 shows the RPS3 cloned sequence (Nucleotides 38 to 3154) ; SEQ ID NO:17 shows the complete plasmid sequence of pRPS3-1005-EGFP; SEQ ID NO:18 shows the complete plasmid sequence of pCET1015-EGFP.

EGFP expression levels, 8 days post-transfection, were investigated, within CHO-K1 pools containing hCMV alone (control construct; plasmid pCET1005-EGFP, linearised with PmeI prior to transfection), constructs containing an 8kb RNPA2 fragment (plasmid pCET1015-EGFP, linearised with *Pme*I prior to transfection) and Rps3 (plasmid pRPS3-1005-EGFP; linearised with *Pme*I prior to transfection).
Pools generated with Rps3 containing constructs show a significant increase in EGFP expression levels compared to control constructs. Addition of the Rps3 sequence upstream of the hCMV promoter resulted in a 5.5- or 1.5-fold increase in mean fluorescence intensity relative to the control or hnRNPA2 element containing constructs respectively (Figure 3A).

The activity of the constructs was investigated in NS0 cells. The increase in mean fluorescence intensity in stable pools when RPS3 element or hnRNPA2 elements are

included in the constructs, compared to the hCMV promoter alone, was 28-fold or 18-fold respectively (Figure 5A).

In both CHO-K1 and NS0 cells, the percentage positive cells was significantly increased with the hnRNPA2 element but this increase was greater with the RPS3 element (Figures 3B and 5B)

Example 2: rps11 element-driven expression

SEQ ID NO:2 shows the RPS11 cloned sequence (nucleotides 12 to 3032); SEQ ID NO:19 shows the complete sequence of pRPS11-1005-EGFP.

Rps11 containing- and control vectors (*Pme*I linearised) were transfected into CHO-K1 and NS0 cell lines and stable pools were generated by puromycin selection. Mean EGFP expression levels were assessed by FACscan analysis.

The addition of the Rps11 element upstream of the hCMV resulted in a 1.2-fold increase in mean EGFP expression levels, in CHO-K1 pools, compared to a construct containing the previously described RNPA2 fragment (Fig 4A).

NS0 cell lines stably transfected with Rps11 containing constructs demonstated a 1.8 and 1.5 fold, (respectively) increase in mean EGFP expression levels compared to hCMV and RNPA2 constructs (Fig 6A).

An increase in percentage positive cells was observed for CHO-K1 cell lines transfected with Rps11 constructs compared to RNPA2 constructs (Fig 4B). Furthermore, an increase in percentage positive cells was observed in NS0 pools transfected with Rps11 constructs compared to both hCMV and RNPA2 (Fig 6B)

Claims

1. An isolated polynucleotide comprising
   a) an element comprising at least 500 contiguous nucleotides from the promoter region of a ribosomal protein gene
   b) a heterologous promoter
   c) a transcribable nucleic acid sequence adjacent said heterologous promoter

   wherein the transcribable nucleic acid sequence is transcribed from said heterologous promoter and the level of said transcription is enhanced by said element.

2. A polynucleotide according to claim 1, wherein said ribosomal protein gene is selected from the list consisting of RPSA, RPS2, RPS3, RPS3A, RPS4, RPS5, RPS6, RPS7, RPS8, RPS9, RPS10, RPS11, RPS12, RPS13, RPS14, RPS15, RPS15A, RPS16, RPS17, RPS18, RPS19, RPS20, RPS21, RPS23, RPS24, RPS25, RPS26, RPS27, RPS28, RPS29, RPS30, RPL3, RPL4, RPL5, RPL6, RPL7, RPL7A, RPL8, RPL9, RPL10, RPL10A, RPL11, RPL12, RPL13, RPL13A, RPL14, RPL15, RPL17, RPL18, RPL18A, RPL19, RPL21, RPL22, RPL23, RPL23A, RPL24, RPL26, RPL27, RPL27A, RPL28, RPL29, RPL30, RPL31, RPL32, RPL34, RPL35, RPL35A, RPL36, RPL36A, RPL37, RPL37ARPL38, RPL39, RPL41, RPLP0, RPLP1, RPLP2.

3. A polynucleotide according to claim 1 or 2, wherein said ribosomal protein gene is RPS3 or RPS11.

4. A polynucleotide according to claims 1 to 3, wherein said element comprises at least 1000 contiguous nucleotides from the promoter region from a ribosomal protein gene.

5. A polynucleotide according to any preceding claim, wherein said element comprises one or more exons of said ribosomal protein gene.

6. A polynucleotide according to any preceding claim, wherein said element comprises a CpG island.

7. A polynucleotide according to claim 5, wherein said element comprises an extended CpG island.

8. A polynucleotide according to claim 6, wherein said element comprises an extended, methylation-free CpG island.

9. A polynucleotide according to any preceding claim, wherein the ribosomal protein gene is a mammalian gene.

10. A polynucleotide according to claim 9, wherein the ribosomal protein gene is a murine gene.

11. A polynucleotide according to claim 10 comprising the nucleotide sequence of SEQ ID NO:1.

12. A polynucleotide according to claim 10 comprising the nucleotide sequence of SEQ ID NO:2.

13. A polynucleotide according to claim 9, wherein the ribosomal protein gene is a human gene.

14. A polynucleotide according to any preceding claim, wherein said heterologous promoter is a constitutive promoter.

15. A polynucleotide according to claim 14, wherein said constitutive promoter is selected from the list consisting of cytomegalovirus early/immediate promoter, SV40, EF-1α, Rous sarcoma virus (RSV) LTR and HIV2 LTR

16. A polynucleotide according to any of claims 1 to 13, wherein said heterologous promoter is a tissue-specific promoter.

17. A polynucleotide according to claim 16, wherein said heterologous promoter is a tumour-selective promoter.

18. A polynucleotide according to claim 17, wherein said promoter is selected from the list consisting of carcino-embryonic antigen (CEA), prostate-specific antigen (PSA), cyclooxygenase-2 (COX-2), alpha-fetoprotein (AFP), tyrosinase, and T-cell Factors 1–4 (TCF) -based promoters.

19. A polynucleotide according to any preceding claim, wherein said transcribable nucleic acid encodes a polypeptide selected from the list consisting of an antibody, a functional epitope-binding fragment of an antibody, growth factor, cytokine, protein kinase, soluble receptor, membrane-bound receptor, blood clotting factor [etc...]

20. A vector comprising the polynucleotide of any preceding claim.

21. A eukaryotic expression vector according to claim 20.

22. A eukaryotic expression vector comprising
    a) an element comprising at least 500 contiguous nucleotides from the promoter region of a ribosomal protein gene
    b) a heterologous promoter
    c) a multiple cloning site
    wherein a transcribable nucleic acid sequence inserted into said multiple cloning site is capable of being transcribed from said heterologous promoter and the level of said transcription is enhanced by said element.

23. A host cell comprising an isolated polynucleotide according to any of claims 1 to 19, or the vector of any of claims 20 to 22.

24. A host cell according to claim 23, wherein is said cell is selected from the list consisting of CHO, NS0, BHK, HeLa, HepG2.

25. A method of expressing a polypeptide comprising inserting expression vector according to either claim 21 or claim 22 into an appropriate host cell and culturing said host cell in suitable conditions to allow expression.

26. Method according to claim 25, wherein said polypeptide is a therapeutically useful polypeptide.

27. A pharmaceutical preparation comprising a polynucleotide according to any of claims 1 to 19, a vector according to any of claims 20 to 22, or a host cell according to either claim 23 or claim 24, and a pharmaceutically acceptable carrier, excipient, buffer or medium.