

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
12 February 2009 (12.02.2009)

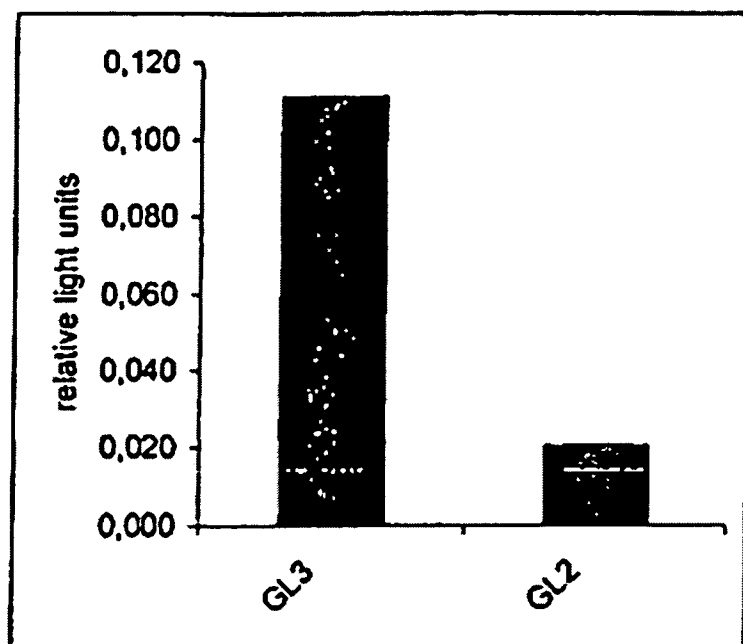
PCT

(10) International Publication Number
WO 2009/019018 A1

- (51) International Patent Classification:
C12N 15/67 (2006.01)
- (21) International Application Number:
PCT/EP2008/006511
- (22) International Filing Date: 4 August 2008 (04.08.2008)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
EP07015280.6 3 August 2007 (03.08.2007) EP
- (71) Applicant (for all designated States except US): **MAX - PLANCK - GESELLSCHAFT ZUR FORDERUNG DER WISSENSCHAFTEN E.V.** [DE/DE]; Hofgarten str. 8, 80539 München (DE).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **SCHWEIGER, Susann** [DE/GB]; 4 Cedar Road, DD5 3BB Dundee (GB). **ARANDA-ORGILLÉS, Beatriz** [ES/DE]; Schwedter Str. 245, 10435 Berlin (DE). **SCHNEIDER, Rainer** [AT/AT]; Prof.-Hans-Hömburg-Strasse 14, 6300 Wörgl (AT). **ROEPCKE, Stefan** [DE/DE]; Handelsgartenweg 7, 78462 Konstanz (DE). **KRAUSE, Sven** [DE/DE]; Am Eikborn 46, 37079 Göttingen (DE).
- (74) Agent: **ARTH, Hans-Lothar**; ABK Patentanwälte, Jasminweg 9, 14052 Berlin (DE).
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:**
- with international search report
 - with sequence listing part of description published separately in electronic form and available upon request from the International Bureau

(54) Title: ENHANCEMENT OF PROTEIN PRODUCTION IN EUKARYOTIC CELLS

Figure 2



(57) Abstract: The present invention relates to the increased expression of polypeptides in eukaryotic cells. A method for increased production is described using an expression vector comprising a polynucleotide encoding a desired polypeptide and a G quartet like motif sequence.

WO 2009/019018 A1

Enhancement of protein production in eukaryotic cells

Specification

5

The present invention relates to the increased expression of polypeptides in eukaryotic cells. A method for increased production is described using an expression vector comprising a polynucleotide encoding a desired polypeptide and a G quartet like motif sequence.

10

Background of the invention

The production of eukaryote proteins in cell systems is an important issue for a wide range of applications in modern biotechnology. Applications range from research
15 into the biological function of proteins to their production as biopharmaceuticals and diagnostic reagents and the development of transgenic animals and plants. However importantly, most of these applications require correct post-translational modification. Production of eukaryote proteins in prokaryotes is limited by the lack of post-translational processing, including appropriate protein folding and secondary
20 modifications such as glycosylation and some phosphorylations. Also, the absence of most cell organelles prohibits the functional analysis of eukaryotic proteins in prokaryotes. Thus many applications, including gene therapy, the fabrication of recombinant protein therapeutics and diagnostics and the production of proteins as well as the production of transgenic plants and animals for in vivo and in vitro studies,
25 depend on proteins produced in eukaryotic cells. Unfortunately, the relatively low productivity of most eukaryotic cell systems is the most significant hindrance.

Efforts have been undertaken during the last few years to improve productivity of eukaryotic protein expression systems. The site of integration of a plasmid into the
30 genome of an acceptor cell line has a major impact on its transcription, for example, integration into heterochromatin results in little or no expression. Several strategies have been developed to overcome this problem. Targeted integration via homologous recombination is one possibility and enzymes with recombinase activity, such as bacteriophage P1 Cre recombinase, lambda phage integrase or yeast Flp recombinase, can be used to enhance the probability of targeted integration. Also
35 due to the strong influence of splicing on translation most modern expression vectors include an intron between the promoter and the cDNA coding sequence. Additionally, engineering of rare tRNA codons into more frequently used tRNA codons encoding identical amino acids, can be used to convert a gene expressed at

low levels into a high-expressing gene. In another strategy host cells have been genetically modified into higher efficiency producers by integrating genes into their genome that make them resistant to influences that reduce viability or growth-promoting proto-oncogenes, cell cycle control genes, growth factors and anti-apoptotic genes.

Guanine quartet sequences and guanine quadruplexes (G-quartet like sequences) are known in the art. For example, Bonnal *et al.*, (2003) Journal of Biological Chemistry, Vol. 278(41) pp. 39330-39336 discloses an expression vector having a internal G-quartet like sequence; WO 99/14346 and WO 2006/022712 disclose G-quartet like sequences and that these sequences are involved in regulation of gene expression; and Aranda-Orgilles, B. (2006) Online @ URL: <http://www.diss.fu-berlin.de/2006/660/indexe.html> discloses the interaction between the MID1 protein and the $\alpha 4$ protein and a G-quartet like sequence. However none of these documents disclose that a G-quartet like sequence can be used in a method to increase protein production in eukaryotic cells as disclosed in the present application.

The present invention surprisingly found that protein production is increased by polypeptide translation from and / or increased mRNA stability of an mRNA carrying G-quartet like sequences. It is the object of the present invention to provide a method and expression vector that optimizes the productivity of eukaryotic expression systems at the translation level.

The object of the present invention is solved by the teaching of the independent claims. Further advantageous features, aspects and details of the invention are evident from the dependent claims, the description, and the examples of the present application.

30

Description of the invention

The present inventors have identified a cytoskeleton-associated ribosomal complex that includes active polyribosomes and components of the translation-controlling mTOR (mammalian target of rapamycin) pathway, such as the MID1 and $\alpha 4$ proteins, which is present in all cell types tested, including yeast and insect cells. mRNA associated to this complex is stabilized, thus significantly increasing their translation. Furthermore, the present inventors found that mRNA containing G-quartet like RNA motifs having a consensus sequence of WGG-N(1-4)-WGG-N(1-4)-WGG-N(1-4)-WGG has a high affinity to the described ribosome/mTOR MID1 complex. Thus the

MID1 complex associates with mRNAs via a sequence motif, which we called MIDAS (MID1 association sequence) G-quartet sequence motif, that corresponds to the consensus sequence above and this association enhances the translation efficiency of these mRNAs several fold as demonstrated in Examples 2 to 6 and Figures 2 to 5 and Figure 6d. Furthermore, the observed effects of the MID1 / $\alpha 4$ protein complex on the synthesis of proteins encoded by RNAs with MIDAS G-quartet sequence motifs has multiple implications for the production of proteins in eukaryotic cell systems that either rely on high productivity or on dose-dependent protein expression.

G-quartet like RNA motifs have been described previously, however the biological function of these structures is still under investigation. They have been variously associated with mRNA turnover and HIV packaging, as well as being proposed to have a regulatory role in cell metabolism.

The present inventors have demonstrated that the MIDAS G-quartet like motifs having a consensus sequence of WGG-N(1-4)-WGG-N(1-4)-WGG-N(1-4)-WGG (SEQ ID NO: 3) mediate and stabilize the binding of mRNAs carrying the G-quartet like sequence to the ribosome complex. Further the present inventors have demonstrated that binding is increased in a linear additive manner by the presence of the MIDAS G-quartet like motif sequences. As described in Example 1 and shown in Figure 1 and Example 6 and Figure 6d as the number of MIDAS G-quartet like motif sequences in an mRNA was increased up to four MIDAS G-quartet like motifs, the amount of the protein production from the mRNA also increased.

The position of the MIDAS G-quartet like motif(s) within the respective mRNA does not seem to play a role, meaning that the MIDAS G-quartet like motif sequence can be located inside or outside (3' or 5') of an open reading frame. When the positioning of essential sequence MIDAS G-quartet like sequence motifs is outside of the open reading frame, this method will not influence the protein sequence or post-translational modifications, guaranteeing unaltered quality of the produced proteins.

Without wishing to be bound by any particular theory the inventors propose that the possible molecular basis of the increase in protein production is both an increase of mRNA stability mediated by proteins interacting with the complex and an induction of translation via the mTOR signalling cascade.

Production of recombinant therapeutic proteins in cultivated mammalian cells has surpassed that in microbial expression systems because of the greater capacity of

eukaryotic cells for proper folding, assembly and post-translational modification of proteins, which has helped to enhance the quality and efficacy of protein products for research and clinical applications. However, the production of complex proteins in mammalian cells remains extremely costly, and consequently there is high demand
5 for novel approaches to lower costs and enhance protein yields. Therefore our present finding that MIDAS motifs introduced into mRNAs markedly enhanced the production of desired proteins will permit considerable improvements in the cost-efficient production of sophisticated biopharmaceuticals.

10 We describe herein a strategy to enhance protein production by increased translation from an mRNA carrying MIDAS G-quartet like sequences 3' or 5' of or within an open reading frame in cell lines and transgenic plants and animals. Furthermore, protein amounts produced in the host cell lines or transgenic plants or animals can be titrated upon the number of G-quartets, preferably up to five incorporated in the vector.

15

Preferably the MIDAS G-quartet motif encodes the following sequence:

NGGN₍₂₋₅₎GGN₍₂₋₅₎GGN₍₂₋₅₎GG (SEQ ID NO: 4)

wherein N is any nucleotide.

20 Further preferably the MIDAS G quartet motif sequence encodes an RNA consensus sequence of:

WGGN₍₁₋₄₎WGGN₍₁₋₄₎WGGN₍₁₋₄₎WGG (SEQ ID NO: 3)

wherein N is any nucleotide and W is A or T.

25 Preferably, if incorporated 3' of an open reading-frame, the RNA motif enhances protein production 2-10 fold as demonstrated in Example 6 and Figures 6b and 6c. Further preferably co-transfection with a plasmid expressing the MID1 or $\alpha 4$ gene additionally increases protein production another 5-fold.

30 Also in the present invention the protein production increases according to the increasing number of MIDAS G-quartets present and can be regulated by the number of MIDAS G-quartets incorporated in the vector as demonstrated in Example 6 and Figure 6d. Thus the amount of polypeptide produced is increased as the number of MIDAS G quartet motif sequences is increased. For example, a larger amount of
35 polypeptide is produced when four MIDAS G quartet motif sequences are used that when only one MIDAS G quartet motif sequence is used. Preferably up to five MIDAS G-quartet like motifs shall be used in the same vector.

As used herein "increased production of a polypeptide" is defined as that the amount of a polypeptide produced by a host cell in the presence of the MIDAS G quartet motif sequence is greater than the amount of the polypeptide produced in the absence of the MIDAS G quartet motif sequence.

5

As used herein "polypeptide" is defined as a polymer of amino acids joined by peptide bonds. The polypeptide is encoded by a corresponding polynucleotide sequence.

10 As used herein "polynucleotide sequence" is defined as a DNA sequence capable of encoding a polypeptide.

As used herein "protein" is defined as a macromolecule comprising one or more polypeptide chains. A protein may also comprise non-peptide components, such as
15 carbohydrate groups.

As used herein "expression vector" is defined as a nucleic acid molecule encoding a polynucleotide sequence that is expressed in a host cell. The expression vector can also comprise regulatory sequences, termination sequences and sequence encoding
20 a selection factor(s). The expression vector may be, for example, a plasmid or a viral or retroviral vector, such as the pLenti6-V5, the pLenti6/V5-DEST Gateway™, the pAd/PL-DEST™ the pAd/CMV/V5-DEST™, the BD BaculoGold™, the AcNPV, the pCMV/Blue/OuabainR, the pCMV/OuabainR, the pRK-5, the pAPtag-4, the Gateway™ pCMV•SPORT6, the pcDNA™3.1, the pcDNA™4, the pcDNA™5, the
25 pcDNA™6, the pCEP4, the pCMV, the pDisplay™, the pEF, the pEF1, the pEF4/, the pEF6, the Plasmid pCMV•SPORT, the pREP4, the pSecTag2, the pTracer™-EF, the pUB6, the pVAX1, the pVP22, the pZeoSV2, the pBud, the pIRES, the pCI, the pSI, the pTARGET™, the pFLAG®-MAC, the p3XFLAG-CMV™, the pBICEP-CMV™, the pcDNA, the pCMV-FLAG®, the pFLAG®-CMV, the pRc, the pGS, the pUSEamp, the
30 pGL, the pRL, the pHMGFP, the pEGFP, the pGFP, the pYFP, the pRFP, the pDSred, the pSV-β-Gal, the pCAT®3, the pHTS-MCS, the pcherry, the pHT3, the pStec, the FMHGW, the rTA, the pHR'CMV, the pSNVRU5luc, the pSNVluc, the pTRL2, the pCALwL, the tTS, the psiRNA™, the pBROAD, the pWHERE, the pMONO, the pSELECT, the pVITRO, the pVIVO, the pFUSE-Fc and others.

35

As used herein "host cell" is defined as a cell capable of producing a polypeptide encoded by a polynucleotide sequence carried on an expression vector. The host cell may be, for example, a yeast cell, an insect cell, or a mammalian cell. Examples of specific host cells include the following: any insect cell line of interest, any plant

cell of interest, any mammalian cells of interest or yeast or any transgenic animal cells of interest. For example, Table 1 provides a listing of possible mammalian host cells that could be used with an expression vector containing at least one G quartet motif sequence according to the present invention. However the listing is not exhaustive and the expression vector containing at least one MIDAS G quartet motif sequence according to the present invention could be used in a mammalian host cell that is not listed in Table 1. Example 7 and Figures 7a, 7b, 7c and 7d demonstrate the activity of the MIDAS G quartet motif sequence in a method of increasing protein production in HeLa (Figure 7a), CHO (Figure 7b), COS7 (Figure 7c) and HEK293 (Figure 7d) mammalian cell lines.

Table 1. Mammalian host cells

Cell line	Cell type	Species
143B	Osteosarcoma	Human
15P-1	Sertoli (Testis)	Mouse
16HBE	Epithelial airway	Human
293, 293H, 293F	Kidney	Human
293T	Embryonic kidney	Human
3T3-F442A	Adipocytes	Mouse
A10	Smooth muscle	Rat
A20.J; M12	Lymphoid	Mouse B cell lymphoma
A498; UMRC 6	Kidney carcinoma	Human
A549	Lung	Human carcinoma
AKR	Spleen	Mouse
AR4-2J	Pancreas	Rat
AS52; CHO-DXB11	Ovary	Hamster
AtT-20	Pituitary	Mouse
B16	Melanoma	Mouse
BEAS-2B	Bronchial epithelial cells	Human
BL3.1	Lymphoid	Bovine leukemia
Blastoderm cells		Chicken
C127	Breast Cancer	Mouse
C2C12	Myoblast	Mouse
C2C12	Muscle	Cow
C-33 A	Cervical Cancer	Human
C6	Glioma	Rat
Capan-2	Pancreas cancer	Human
CHO	Ovary (epithelial like)	Chinese Hamster

CHO-K1	Ovary	Hamster
COLO 320 HSR; SW40; 3SW480; Caco-2	Colon	Human cancer
COS1, COS-7	Kidney	Green Monkey
CT60	Ovary	Hamster
D3	Stem cells	Mouse
Dede	Lung	Hamster
F11	Mouse neuroblastoma + Rat DRG neurons	Mouse/Rat
FLK	Kidney	Sheep
FTO-2B	Liver cancer	Rat
GD25-β1	Embryonal fibroblast	Mouse
GH3	Pituitary tumor	Rat
H187; NCI-H146	Lung	Human small cell carcinoma
HBL-100	Transformed Breast	Human
HCT116	Colon	Human cancer
HEK293	Embryonic kidney	Human
HeLa	Cervical Cancer	Human
Hep 3B	Liver cancer	Human
Hep2	Laryngeal Epithelium	Human
HepG2	Hepatoma	Human
Hs 578T; T-47D; MDA-MB-231; ZR-75-1	Breast Cancer	Human
HT-29	Colon	Human cancer
Huh7	Liver cancer	Human
Isikawa	Endometrial cancer	Human
Jurkat	Lymphoid	Human lymphoma
KG-1; KG1a; K562	leukemia	Human
L	Connective Tissue	Mouse
L6E9	Muscle	Rat
LLC	Lung	Mouse carcinoma
LLC-PK	Kidney	Pig
LNCaP; PC-3	Prostate	Human
MCF-7	Breast Adenocarcinoma	Human
MCF7	Breast Cancer	Human
MDCK	Kidney	Canine
MeWo	Melanoma	Human

MOLT-4	Lymphoid	Human leukemia
MRC-5	Lung	Human
MSN610.2; Morris 7777; H35; McA-RH7777	Liver cancer	Rat
Mv 1 Lu	Lung Fibroblasts	Mink
N2A	Neuroblastoma	Mouse
NCI-H23	Lung	Human cancer
Neuro-2a	Neuroblastoma	Mouse
NG108-15	Neurabloma-glioma fusion	Mouse-Rat
NIH 3T3; TA1; Y-2; NIH 3T6; 10T1/2	Fibroblasts	Mouse
NIH-3T3	Fibroblasts	Mouse
NMU	Breast Cancer	Rat
P19; F9	Embryonic carcinoma	Mouse
P388D1	Lymphoid	Mouse lymphoma
PC 12	Pheochromocytoma	Rat
Primary chondrocytes		Chicken
Primary Endothelial	Endothelial	Rabbit
Primary fetal astrocyte	Astrocyte	Rat
Primary fetal islet cells	Pancreas islet	Rat
Primary Fibroblast	Fibroblasts	Human
Primary fibroblasts		Chicken
Primary hematopoietic CD34+	Stem cells	Human
Primary hepatocytes		Chicken
Primary HUAEC	Endothelial	Human
Primary HUVEC	Endothelial	Human
Primary keratinocytes	Skin	Human
Primary Liver	Liver	Dog
Primary Liver	Liver	Mouse
Primary Liver	Liver	Rat
Primary Ovary	Ovary	Mouse
Primary Pituitary	Pituitary	Rat
Primary; G8	Muscle	Mouse
Primary; J774.16; RAW 264.7	Monocytes	Mouse macrophage
R1610	Lung	Hamster

SK-BR-3	Breast Cancer	Human
SK-N-MC	Neuroblastoma	Human
SK-N-SH	Neuroblastoma	Human
SKOV3	Ovary cancer	Human
SK-UT-1	Uterus	Human
Sperm	Sperm	Mouse
SW13	Adrenal carcinoma	Human
Swiss 3T3	Fibroblasts	Mouse
T24	Bladder carcinoma	Human
T84	Colon	Human cancer
THP-1	Macrophages/	Human monocyte
Trachea smooth muscle	Muscle	Cow
U-2 OS	Bone sarcoma	Human
U373	Glioblastoma	Human
U-87; A-172	Glioma	Human
U937	Lymphoid	Human lymphoma
WI-38	Lung	Human

As used herein a "host" is an organism capable of producing a polypeptide encoded by a polynucleotide sequence carried on an expression vector. The host may be, for example, a transgenic animal

Thus the present invention is directed to a method for increasing production of a polypeptide in a eukaryote host cell comprising transforming the eukaryote host cell with an expression vector comprising a polynucleotide sequence that encodes an open reading frame for the polypeptide and at least one and preferably up to five G quartet like motif sequences, culturing the transformed cells under suitable conditions and isolating the expressed polypeptide.

General methods for constructing expression vectors, transforming host cells, culturing host cells and isolating an expressed polypeptide are well known in the art. For example, the commercial products and accompanying protocols available from any of Invitrogen Corporation, InvivoGen, Promega Corporation, or Sigma-Aldrich Co. are suitable to carry out the present invention.

In the method of the present invention the amount of a polypeptide produced in the presence of a G quartet motif sequence is greater than the amount of polypeptide produced in the absence of a G quartet motif sequence.

Preferably the number of G quartet motif sequences is between 1 and 5, more preferably between 3 and 4.

- 5 In the method of the present invention the amount of the polypeptide produced is increased as the number of G quartet motif sequences is increased.

The method of the present invention is suitable for use in an in vitro eukaryote expression system. The method of the present invention is also suitable for use in an
10 in vivo eukaryote expression system.

Preferably in the expression vector the G quartet motif sequence is situated 3' of the polypeptide open reading frame. When the G quartet motif sequence is incorporated 3' of the open reading-frame preferably protein production is increased 2-10 fold.
15

Alternatively in the expression vector the G quartet motif sequence is situated within the polypeptide open reading frame.

Preferably the G-quartet motif encodes the following sequence:

20 $\text{NGGN}_{(2-5)}\text{GGN}_{(2-5)}\text{GGN}_{(2-5)}\text{GG}$ (SEQ ID NO: 4)

wherein N is any nucleotide.

Further preferably the G quartet motif sequence encodes an RNA consensus sequence of:

25 $\text{WGGN}_{(1-4)}\text{WGGN}_{(1-4)}\text{WGGN}_{(1-4)}\text{WGG}$ (SEQ ID NO: 3)

wherein N is any nucleotide and W is A or T.

Further preferably the G quartet motif sequence is:

30 $\text{GGTATCAGGCAAGGATATGG}$ (SEQ ID NO: 2)

Preferably the polypeptide is a human polypeptide. Alternatively the polypeptide is an animal polypeptide. Further preferably the polypeptide is a recombinant polypeptide. Further preferably the polypeptide is a plant polypeptide.

- 35 Additionally the eukaryote cell can be further transformed with an expression vector encoding the MID1 protein. Preferably when the host cell is co-transfected with a plasmid expressing the MID1 gene protein production is increased 5-fold. The host cell can also be further transformed with an expression vector encoding the $\alpha 4$ protein.

Preferably the host cell is any insect cell of interest or any plant cell of interest or any mammalian cell of interest or yeast or any transgenic animal cell of interest.

5

The present invention is also directed to the use of a G quartet motif sequence to increase production of a polypeptide in a eukaryote host cell, wherein the eukaryote host cell is transformed with an expression vector comprising a polynucleotide sequence that encodes an open reading frame for the polypeptide and at least one G quartet motif sequence.

10

In the use of the present invention the amount of a polypeptide produced in the presence of a G quartet motif sequence is greater than the amount of polypeptide produced in the absence of a G quartet motif sequence.

15

Preferably the number of G quartet motif sequences is between 1 and 5, more preferably between 3 and 4.

In the use of the present invention the amount of the polypeptide produced is increased as the number of G quartet motif sequences is increased.

20

The G quartet motif sequence of the present invention is suitable for use in an in vitro eukaryote expression system. The G quartet motif sequence of the present invention is also suitable for use in an in vivo eukaryote expression system.

25

Preferably in the expression vector the G quartet motif sequence is situated 3' of the polypeptide open reading frame. When the G quartet motif sequence is incorporated 3' of the open reading-frame preferably protein production is increased 2-10 fold.

30

Alternatively in the expression vector the G quartet motif sequence is situated within the polypeptide open reading frame.

Preferably the G-quartet motif encodes the following sequence:

NGGN₍₂₋₅₎GGN₍₂₋₅₎GGN₍₂₋₅₎GG (SEQ ID NO: 4)

35

wherein N is any nucleotide.

Further preferably the G quartet motif sequence encodes an RNA consensus sequence of:

WGGN₍₁₋₄₎WGGN₍₁₋₄₎WGGN₍₁₋₄₎WGG (SEQ ID NO: 3)

wherein N is any nucleotide and W is A or T.

Further preferably the G quartet motif sequence is:

GGTATCAGGCAAGGATATGG (SEQ ID NO: 2)

5

Preferably the polypeptide is a human polypeptide. Alternatively the polypeptide is an animal polypeptide. Further preferably the polypeptide is a recombinant polypeptide. Further preferably the polypeptide is a plant polypeptide.

10 Additionally the eukaryote cell can be further transformed with an expression vector encoding the MID1 protein. Preferably when the host cell is co-transfected with a plasmid expressing the MID1 gene protein production is increased 5-fold. The host cell can also be further transformed with an expression vector encoding the $\alpha 4$ protein.

15

Preferably the host cell is any insect cell of interest or any plant cell of interest or any mammalian cell of interest or yeast or any transgenic animal cell of interest.

20 The present invention is further directed to an expression vector comprising a polynucleotide encoding an open reading frame for a polypeptide and at least one G quartet motif sequence, wherein the G quartet motif sequence increases production of the polypeptide in a eukaryote host cell.

25 In a host cell transformed with the expression vector of the present invention the amount of a polypeptide produced in the presence of a G quartet motif sequence is greater than the amount of polypeptide produced in the absence of a G quartet motif sequence.

30 Preferably the number of G quartet motif sequences is between 1 and 5, more preferably between 3 and 4.

In a host cell transformed with the expression vector of the present invention the amount of the polypeptide produced is increased as the number of G quartet motif
35 sequences is increased.

The expression vector of the present invention is suitable for use in an in vitro eukaryote expression system. The expression vector of the present invention is also suitable for use in an in vivo eukaryote expression system.

Preferably in the expression vector the G quartet motif sequence is situated 3' of the polypeptide open reading frame. When the G quartet motif sequence is incorporated 3' of the open reading-frame preferably protein production is increased 2-10 fold.

5

Alternatively in the expression vector the G quartet motif sequence is situated within the polypeptide open reading frame.

Preferably the G-quartet motif encodes the following sequence:

10 NGGN₍₂₋₅₎GGN₍₂₋₅₎GGN₍₂₋₅₎GG (SEQ ID NO: 4)

wherein N is any nucleotide.

Further preferably the G quartet motif sequence encodes an RNA consensus sequence of:

15 WGGN₍₁₋₄₎WGGN₍₁₋₄₎WGGN₍₁₋₄₎WGG (SEQ ID NO: 3)

wherein N is any nucleotide and W is A or T.

Further preferably the G quartet motif sequence is:

20 GGTATCAGGCAAGGATATGG (SEQ ID NO: 2)

Preferably the polypeptide is a human polypeptide. Alternatively the polypeptide is an animal polypeptide. Further preferably the polypeptide is a recombinant polypeptide. Further preferably the polypeptide is a plant polypeptide.

25 Additionally the expression vector further comprises polynucleotide sequence encoding the MID1 protein. Preferably when a host cell is co-transfected with an expression vector expressing the MID1 gene protein production is increased 5-fold. The expression vector can also further comprise polynucleotide sequence encoding the α 4 protein.

30

Preferably the host cell is any insect cell of interest or any plant cell of interest or any mammalian cell of interest or yeast or any transgenic animal cell of interest.

Description of the figures:

Figure 1: EFNB1 mRNAs either lacking the entire 3'UTR (-3'U) or including one G-quartet (+1G), two G-quartets (+2G), three G-quartets (+3G) or four G-quartets (+4G), with antisense transcripts of 3 G-quartet (+3G-AS) and four G-quartet (+4G-AS) mRNAs used as background controls. G-quartets were amplified and *in vitro* transcribed using biotinylated ribonucleotides. Elution fractions of RNA-protein pull-down assays in FLAG-MID1 over-expressing cells were analysed for the presence of FLAG-MID1 with an anti-FLAG antibody.

Figure 2: pGL3 had 6 times higher luciferase activity from the firefly luciferase gene compared to pGL2 as measured in relative light units.

Figure 3: Basal activities of the wild-type pGL3 (GL3) and a pGL3 vector mutated at a core guanine of the predicted G-quartet (GL3mut) as measured in relative light units.

Figure 4: Induction of luciferase activity is observed by co-transfection with wild-type MID1, but not with a mutated (A130T) form of MID1. Induction is seen of the pGL3 wild-type firefly luciferase gene (a, b) but not of a firefly luciferase gene with a mutated G-quartet sequence motif (GL3mut) or the pGL2 firefly luciferase gene.

4a. Firefly luciferase activity recorded from HeLa cells transfected with vectors either containing (GL3) or lacking (GL2) a functional MIDAS G-quartet sequence motif, and co-transfected with empty pCMV vector (-cmv), wild-type-MID1 (-MID1), or mutated, inactive MID1 (-A130T).

4b. Firefly luciferase activity recorded from HeLa cells transfected solely with pGL3 vectors containing either a functional MIDAS G-quartet sequence motif (GL3) or a mutated MIDAS G-quartet sequence motif (GL3mut) within the open reading frame, and also co-expressing either empty pCMV vector (+cmv), wild-type-MID1 (+MID1), or mutated, dysfunctional MID1 (+A130T).

Figure 5: FLAG-MID1 was identified in the elution fraction of an RNA-protein pulldown assay performed with biotinylated firefly luciferase RNA *in-vitro* transcribed from the pGL3 vector (a, b), but not in an assay performed with firefly RNA transcribed from pGL2 (a), with renilla luciferase RNA (a, b) or in an experiment without RNA. Only little binding was observed with RNA transcribed from a pGL3 vector carrying a mutation in the G-quartet structure (b).

5a. RNA-protein pulldown assay: Lysates from FLAG-MID1-overexpressing HeLa cells were incubated with biotinylated RNA *in vitro*-transcribed from either pGL3 (Firefly-luc+ pGL3) or pGL2 (Firefly-luc pGL2) vectors, or from a

renilla vector without MIDAS G-quartet sequence motif (renilla pRL). Binding fractions and lysate were analyzed on a Western blot with an anti-FLAG antibody.

5b. RNA-protein pulldown assay: Lysates from FLAG-MID1 overexpressing HeLa cells were incubated alone (No RNA) or with biotinylated RNA in vitro-transcribed from a pGL3 vector containing either a functional MIDAS G-quartet sequence motif (Firefly) or a mutated MIDAS G-quartet sequence motif (Mutated), or from a renilla vector lacking the MIDAS motif (renilla). Binding fractions and lysate were analysed on a Western blot with an anti-FLAG antibody.

10 Figure 6: MIDAS G-quartet sequence motif sequence increases production of proteins *in vitro* and *in vivo*.

6a. *In vitro* translation of a firefly luciferase gene containing either a functional MIDAS G-quartet sequence motif (GL3) or a mutated MIDAS G-quartet sequence motif (GL3m) motif. Luciferase activities are shown.

15 6b. Luciferase activity in CHO cells expressing a mutated pGL3 vector (GL3m) carrying a functional (GL3m-MIDAS) or a non-functional (GL3m-MIDASm) MIDAS G-quartet sequence motif 3' to the firefly luciferase coding region.

20 6c. Intensities of green fluorescence recorded from CHO cells transfected with GFP-encoding vectors (pEGFP) either containing (MIDAS) or lacking (empty) a MIDAS G-quartet sequence motif 3' to the coding region for the GFP protein, measured by FACS.

25 6d. Production of GFP from pEGFP vectors with no (empty), single (MIDAS), double (2xMIDAS) or triple (3xMIDAS) functional MIDAS G-quartet sequence motif within the 3'UTR of the GFP gene in CHO cells. Fluorescence intensities measured by FACS.

Figure 7: Production of GFP from pIRES-vectors either containing (GFPI-MIDAS) or lacking (GFPI) a functional MIDAS G-quartet sequence motif 3' to the coding region of the GFP protein, in 7a) HeLa, 7b) HEK, 7c) COS7 and 7d) CHO cells. Fluorescence intensities measured by FACS (top panels), GFP RNA levels relative to those of GAPDH, measured by real-time PCR (centre panels), and ratios between fluorescence intensities and mRNA expression (translation efficiencies, bottom panels) are shown.

35

EXAMPLES

Materials and Methods

5 Constructs

pGL3-promoter-, pGL2-promoter- and pRL-CMV vectors were purchased at Promega. *In vitro* mutagenesis experiments were performed on the pGL3-promoter (primers: F: 5'-CCATCTGCCAGGTATCAGACAAGGATATGGGCTCAC-3' (SEQ ID NO: 5),

10 R: GTGAGCCCATATCCTTGTCTGATACCTGGCAGATGG-3' (SEQ ID NO: 6)) using the QuickChange® Site Directed Mutagenesis Kit (Stratagene) according to manufacturers' instructions.

To produce pGL3m-MIDAS and pGL3m-MIDASm, the following 175 bp DNA-fragment including the MIDAS G-quartet sequence motif (underlined):

15 CACGAAATTGCTTCTGGTGGCGCTCCCCTCTCTAAGGAAGTCGGGGA-
AGCGGTTGCCAAGAGGTTCCATCTGCCAGGTATCAGGCAAGGATATGGGCTCA
CTGAGACTACATCAGCTATTCTGATTACACCCGAGGGGGATGATAAACCGGGCG
CGGTCGGTAAAGTTGTTCCA (SEQ ID NO: 7)

or MIDASm (G-nucleotide in italics changed to A):

20 CACGAAATTGCTTCTGGTGGCGCTCCCCTCTCTAAGGAAGTCGGGGA-
AGCGGTTGCCAAGAGGTTCCATCTGCCAGGTATCAGACAAGGATATGGGCTCA
CTGAGACTACATCAGCTATTCTGATTACACCCGAGGGGGATGATAAACCGGGCG
CGGTCGGTAAAGTTGTTCCA (SEQ ID NO: 8)

was cloned into the XbaI restriction site of pGL3m

25 (primers: F: 5'-TGCTCTAGACACGAAATTGCTTCTGGTGGCC-3' (SEQ ID NO: 9),
R: 5'-TGCTCTAGATGGAACAACCTTACCGCGCC-3' (SEQ ID NO: 10)).

MID1 (wild-type, A130T and C145S) cDNA was excised with HindIII and Sall from pEGFP-C1 and ligated into pCMV-Tag2C (Stratagene).

30 To create pEGFP-MIDAS, a 60bp fragment of pGL3 containing the MIDAS G-quartet sequence motif was amplified

(primers: F: 5'-CCGCTCGAGGCCAAGAGGTTCCATCTGCC-3' (SEQ ID NO: 11),
R: 5'-CCAAGCTTTTAGCTGATGTAG-TCTCAGTGAGC-3' (SEQ ID NO: 12))

and cloned between XhoI and Hind III in the pEGFP-C1 vector (Clontech). GFP-2xMIDAS was created by inserting annealed oligos HE

35 (HE1: AGCTTGGTATCAGGCAA-GGATATGGG (SEQ ID NO: 13),
HE2: AATCCCATATCCTTGCCTGATACCA (SEQ ID NO: 14)),

which comprise a core G-quartet sequence with HindIII and EcoRI overhangs into the HindIII and EcoRI restriction sites of GFP-MIDAS. GFP-3MIDAS was created by analogously inserting annealing oligos ES

(ES1: AATTCGGATCAGGCAAGGATATGGG (SEQ ID NO: 15),

ES2: TCGACCCATATCCTTGCCTGATACCG (SEQ ID NO: 16))

with EcoRI and Sall overhangs into the EcoRI and Sall restriction sites of pEGFP-2xMIDAS). To create GFPi-MIDAS, GFP together with the MIDAS G-quartet
5 sequence motif was excised with NheI and EcoRI from GFP-MIDAS and cloned in the MCS of pIRES-DsRed-Express (Clontech).

RNA-protein pull down assays

10 Firefly and Renilla cDNA were amplified by PCR with primers including the T7 promoter sequence from the different vectors. *In vitro* transcriptions were performed using the RiboMAX™ Large scale RNA production system-T7 (Promega), following the manufacturers instructions with some modifications. Briefly, 2-5µg of purified PCR product were transcribed in the presence of 0.75mM Biotin-16-uridine-5'-triphosphate (Roche) and 5mM UTP. Biotinylated RNAs were purified by
15 phenol/chloroform extraction and kept in nuclease free water after ethanol-precipitation.

RNA-Protein binding assay

20 3x10⁷ HeLa cells were lysed in TKM buffer (20mM Tris; 50mM KCl; 5mM MgCl₂) supplemented with proteinase inhibitors (complete mini, Roche), RNase inhibitor (Promega), 1% NP40 and 1mM DTT on ice, using a 27-G needle. The lysate was cleared by centrifugation for 15 min at 12000g

In order to assay proteins for binding to specific RNA, 3 µg of *in vitro* transcribed and biotinylated RNA were incubated with 200 µg of cytosolic extract of HeLa cells in 500
25 µl of TKM buffer for 1 hour at 4°C and, subsequently, with 40 µl of 50% slurry of M-280 streptavidin coated magnetic beads (DYNAL). Beads were washed 3x with TKM buffer for 10 min at 4°C and boiled off in magic mix (15mM Tris, 48% Urea, 8.7% glycerol, 1% SDS) for 10 min at 95°C. Bound proteins were analyzed on western blots with anti-FLAG antibody (Stratagene).

30

Cell culture and transfection

Cos-7, HeLa and HEK293T cells were maintained in DMEM, and CHO cells in DMEM/F12, supplemented with 10% FCS, 2mM L-glutamine and 100 U/ml penicillin/streptomycin. For FACS analysis, one day before transfection, cells were
35 seeded on six well plates. Next day, Cos-7, HEK and CHO cells were transfected with lipofectamine™ 2000 (Invitrogen) and HeLa cells with DreamFect™ (OZ biosciences) and 1,5µg of DNA according to manufactures' instructions. For the luciferase assays, HeLa cells were seeded on 12 well plates and next day transfected with 200ng Firefly, 10 ng Renilla and 400 ng MID1 construct.

Luciferase assay

24h after translation, HeLa cells were harvested in 1x luciferase passive lysis buffer (Promega). Firefly and Renilla luciferase activities were measured using a Centro LB
5 960 luminometer (Berthold technologies) and the Dual Luciferase Assay System (Promega) according to manufacturers' instructions. Firefly light units were normalized to Renilla light units.

In vitro transcription/ translation

10 Specific PCR products of the Firefly luciferase cDNAs were in vitro transcribed with the mMessage mMachine ® T7 kit (Ambion). Capped transcripts were in vitro translated with the Flexi ® Rabbit reticulocytes lysates (Promega) according to manufactures' instructions. After 90 min Firefly light units were measured using the Luciferase Assay Reagent (Promega) in a luminometer.

15

FACS analysis

24h after transfection, cells were trypsinized and fixed with ethanol. FACS was analysis was performed on a BD FACSCalibur™. GFP was excited with a 488nm laser and emission collected using a 530 nm emission filter. A total of 10,000
20 events/experiment were collected for each construct in three independent experiments. Transfection efficiencies with the different constructs were comparable.

Quantitative RT-PCR analysis

RNA from 6 well dishes was isolated using QIAGEN's RNAeasy kit. cDNA was
25 synthesized using the TaqMan reverse transcription reagents kit (Applied Biosystems) with random primers. Primers for real time PCR analysis were designed with the Primer Express ® Software v2.0

(GFP; F: 5'- CTACCTGAGCACCCAGTCCG-3' (SEQ ID NO: 17);

R: 5'- TGATCGCGTTCTCGTTG-3' (SEQ ID NO: 18); GAPDH).

30 For real time PCR analysis, absolute quantification with standard curve using SYBRGreen PCR master mix (Applied Biosystems) was performed on an Applied Biosystems ABI Prism 7900HT Sequence Detection System equipped with SDS software v2.

35

Example 1:

G-quartet motifs bind to a microtubule-associated mRNP

MID1 and $\alpha 4$ are the core of a microtubule-associated mRNP that, in addition to active ribosomes, also assembles G-rich mRNA motifs. In an RNA-protein pull-

down experiment, we found that G-quartet structures with the consensus sequence of WGG-N(1-4)-WGG-N(1-4)-WGG-N(1-4)-WGG (SEQ ID NO: 3) have a particularly high affinity to the mRNP.

5 G-quartet motifs from different mRNAs (EFNB1b, EPBH2b, EFNB2a, EPBH3c) were amplified and *in vitro*-transcribed with biotinylated ribonucleotides. Transcripts were then immobilized on streptavidin-coated beads and loaded with lysates from MID1-FLAG over-expressing HeLa cells. Elution fractions were analysed on SDS gels using the respective antibodies detecting members of the MID1 protein complex and binding
10 of mRNP to the G-quartet structures was demonstrated.

G-quartet motifs bind to a microtubule-associated mRNP in an additive manner

A similar protein-RNA pull-down assay with biotinylated RNA *in vitro* transcribed from a plasmid containing the open reading frame and 3'UTR of ephrin B1 (EFNB1)
15 showed an additive effect of multiple G-quartets on the binding affinity of the respective RNA to the mRNP.

We used EFNB1 mRNAs with differing lengths, including only the open reading frame or the open reading frame and different parts of the 3'UTR with one, two, three or four
20 G-quartets. Antisense transcripts of the two longest mRNAs were used as background controls. G-quartets were amplified and *in vitro* transcribed using biotinylated ribonucleotides. Elution fractions of RNA-protein pull-down experiments with HeLa cells over-expressing FLAG-tagged MID1 were subsequently analysed by Western blot analysis with an anti-FLAG antibody, and showed a linear increase of the binding
25 affinity between mRNA and MID1 proportional to the number of G-quartet structures (see Figure 1).

Example 2:

30 *Comparison of luciferase activity: pGL2 and pGL3*

Luciferase assays showed a 6-fold higher basal activity of firefly luciferase activity from pGL3 vector compared to the pGL2 vector (see Figure 2). Ratios between firefly and renilla luciferase were measured in a dual luciferase vector system.

35 *Alignment between pGL2 and pGL3*

Sequence alignments of pGL2 and pGL3 show identity of the two vectors over large parts of the sequence. Only a few base pair mismatches were detected within the open reading frame of the firefly luciferase. Among these, four base pair exchanges

were found to be located in a G-quartet structure predicted for pGL3, thus eliminating the motif in pGL2.

5 GGGATACGACAAGGATATGG pGL2 (SEQ ID NO: 1)

 GGTATCAGGCAAGGATATGG pGL3 (SEQ ID NO: 2)

Example 3:

10 *Translation decrease in a pGL3 vector with mutated G-quartet*

Next, we mutated one of the core guanines of the G-quartet predicted in the pGL3 firefly luciferase gene such that the amino acid sequence did not change (core guanine at position 9 of the pGL3 motif sequence shown above was mutated to adenine), and checked luciferase activity of the mutated vector pGL3mut. Interestingly, basal activity
15 of the mutated vector went down significantly (see Figure 3).

Example 4:

Increase of translation by co-transfection with MID1

20 To test the hypothesis that the MID1 complex modulates the efficiency of protein production from associated mRNAs, we overexpressed either a FLAG-MID1wt construct or a FLAG-MID1 construct containing inactive MID1 or an empty pCMV vector in HeLa cells and analyzed its influence on the expression of firefly luciferase using a dual luciferase reporter assay. For this assay FLAG-MID1 or pCMV
25 expressing cells were co-transfected with a renilla luciferase vector and one or the other of two different firefly luciferase vectors (firefly pGL3-promoter or firefly pGL2-promoter, respectively). Unexpectedly, we observed a >2-fold MID1-dependent induction of firefly luciferase activity compared with that seen following co-expression with an empty pCMV vector or with a pCMV construct that encodes an inactive MID1
30 protein, but only in the presence of the pGL3, not the pGL2 vector (Fig. 4a). Comparison of the pGL2 and pGL3 vector sequences indicated that all but a few nucleotides within the open reading frame of the firefly luciferase genes were identical, and that some of the changed nucleotides in pGL3 reside within the MIDAS signature sequence of a G-quartet like structure, a putative protein-RNA binding
35 motif, whereas in pGL2 this MIDAS G-quartet like motif is destroyed (see pGL2 and pGL2 motif sequences in Example 2 above).

To determine whether the MIDAS G-quartet like structure of pGL3 is required for its ability to enhance pGL3 firefly luciferase activity, a core guanine at position 9 of the

pGL3 motif sequence shown in Example 2 above was mutated to adenine. This left the amino acid sequence of the firefly luciferase intact, and both the mutated and non-mutated codons are common in eukaryotic cells. It was found that transfection with the mutated vector virtually eliminated the induction of luciferase activity caused by co-expression with wild-type MID1, as occurred with the wild-type pGL3 vector. In addition, basal firefly luciferase activity was decreased upon transfection with the mutated pGL3 vector, probably owing to the loss of the positive regulatory influence of endogenous MID1 complex upon translation of the mutated mRNA (Fig. 4b).

Thus co-transfection of dual luciferase vectors (firefly, pGL and, as an internal control, renilla, pRL) with an expression vector containing the MID1 open reading frame showed an upregulation of firefly luciferase activity only when using the wild-type pGL3 vector, but not with the mutated pGL3 vector carrying a single substitution mutation in the G-quartet (Fig. 4b) or the pGL2 vector (Fig. 4a). Co-expression of the two luciferase vectors with a MID1-containing plasmid carrying a missense mutation that leads to dysfunction of the MID1 protein did not result in any induction of luciferase activity. Ratios between firefly and renilla luciferase were measured in a dual luciferase reporter system.

Example 5:

Binding of the pGL3 firefly luciferase mRNA to the MID1 protein complex

It was found that FLAG-MID1 bound specifically to biotin-labeled luciferase mRNA *in vitro*-transcribed from the pGL3-firefly luciferase vector, but not to mRNAs transcribed from pGL2- or pRL-renilla luciferase vectors, using a RNA-protein pull-down assay (Fig. 5a). Specificity of the reaction was verified by omission of RNA from the assay (Fig. 5a). From this data we conclude that, while the MID1 protein binds to the firefly luciferase mRNA derived from the pGL3 vector thereby enhancing pGL3-firefly luciferase activity, there is no binding between the MID1 protein and the pGL2 luciferase mRNA and consequently also no influence of MID1 on the pGL2 luciferase activity.

Also the ability of the mutated pGL3 (having a core guanine at position 9 of the pGL3 motif sequence shown in Example 2 above mutated to adenine) firefly luciferase mRNA to bind to FLAG-MID1 was significantly reduced as compared with that of wild-type mRNA, as shown by RNA-protein pull-down assay (Fig. 5b). Again, renilla mRNA and an experiment without RNA served as background controls.

Thus, specific binding of the MID1-protein complex to the wildtype pGL3 firefly luciferase mRNA (Fig. 5a and 5b) but not to the pGL2 firefly luciferase mRNA (Fig. 5a) nor to a pGL3 firefly luciferase mRNA carrying a single substitution mutation in the G-quartet (Fig. 5b) could be shown in an RNA-protein pulldown assay with biotin-streptavidin immobilized RNA and cell lysate from HeLa cells overexpressing FLAG-MID1. Elution fractions were analysed on a Western blot with an anti-FLAG antibody.

Example 6:

The observed increase in protein production driven by the MIDAS G-quartet sequence motif could either result from increased transcription, RNA stability, and/or translation efficiency. To distinguish among these, *in vitro* transcription/translation from the wild-type and MIDAS-mutated forms of pGL3 was performed. Using rabbit reticulocyte lysates, which contain substantial endogenous MID1 levels (data not shown) a 10-fold increase in protein translation was observed from the RNA containing the intact MIDAS motif vs. its single-site substitution mutated form (Fig. 6a), supporting a major impact of the MIDAS motif on translation efficiency.

We also asked whether the position of the MIDAS G-quartet sequence motif within an mRNA alters its translation-promoting activity. Transferability of the MIDAS motif and its translation-enhancing properties from the open reading frame into the 3' or 5' untranslated regions of mRNAs would be of great value for biotechnological protein production processes. Also, because biotechnological protein production heavily depends on optimal viability of the protein-producing cells, and overexpression of MID1 has a negative effect on cell viability (unpublished data), we decided to study putative MIDAS G-quartet sequence motif dependent enhancements of protein synthesis, which we designated the "MIDAS effect" without ectopic expression of MID1, instead relying upon the activity of the endogenous complex. Endogenous MID1 and its protein interaction partners are ubiquitously expressed in many tissues and cell lines. Intact or single substitution mutated (G to A) MIDAS motifs were cloned 3' to the luciferase coding sequence of a pGL3 vector whose intrinsic MIDAS motif contained an inactivating mutation (pGL3mut). These constructs were transfected into CHO cells, which are commonly used in biotechnological processes, and the resulting luciferase activity measured. The results show a several-fold increase of luciferase activity after transfection of the pGL3mut vector containing a functional MIDAS motif 3' to the luciferase coding sequence, as compared with that seen with a vector carrying a non-functional motif (Fig. 6b).

Furthermore, it was found that a strong MIDAS effect also resulted when the motif was cloned 3' to the coding region of the green fluorescent protein (GFP) gene in a pEGFP-C1 vector (Fig 6c). Together, these results show that the MIDAS motif can increase cellular production of proteins encoded by transfected genes, independent of (i) its position within the mRNA, (ii) the particular protein produced, and (iii) the vector used for protein expression.

Furthermore, we asked whether the MIDAS effect is effectively dose-dependent, such that it can be potentiated by inclusion of multiple MIDAS G-quartet sequence motifs. To test this, we introduced one and two additional MIDAS sequences into the 3'UTR of the GFP-MIDAS vector. Transfection of CHO cells with these constructs containing multiple MIDAS motifs in the 3'UTR of the GFP gene enhanced translation efficiency up to 20 fold in CHO cells (Fig. 6d). Thus, increasing the number of MIDAS motifs can lead to an additive increase in translation efficiency, which seems to be based on additive binding of MID1 to the respective RNA (unpublished results).

Example 7:

To determine whether the enhancement of protein production by MIDAS G-quartet sequence motif is cell type-dependent, and also the relative importance of the MIDAS motif's influence upon mRNA stability and protein translation, we studied the MIDAS effect in the GFP-IRES-vector system in four different cell lines and measured simultaneously the fluorescence intensity of the protein produced, by FACS (top panels of Fig. 7a, 7b, 7c and 7d), and the corresponding GFP mRNA levels relative to those of GAPDH, by real-time PCR (centre panels of Fig. 7a, 7b, 7c and 7d). A MIDAS motif was cloned 3' to the coding region for the GFP protein, and a vector lacking the MIDAS motif was used as control. Strong MIDAS effects upon production of GFP was seen in all cell lines tested (top panels of Fig. 7a - HeLa, Fig. 7b - CHO, Fig. 7c - COS7 and Fig. 7d - HEK293), while its influence on the mRNA levels seemed to play a minor role in the overall effect (centre panels of Fig. 7a, 7b, 7c and 7d). Nevertheless, the strongest MIDAS effect, seen in HEK293 cells (8-fold enhancement) was accompanied by a substantial (50%) increase in GFP mRNA relative to that of GAPDH (centre panels of Fig. 7b), suggesting a synergistic effect potentially involving increased mRNA stability as well as translation. The translation efficiency was measured by the ratios between fluorescence intensities and mRNA expression (bottom panels of Fig. 7a, 7b, 7c and 7d).

Claims

1. A method for increasing production of a polypeptide in a eukaryote host cell comprising:
 - 5 a) transforming the eukaryote host cell with an expression vector comprising a polynucleotide sequence that encodes an open reading frame for the polypeptide and at least one G quartet motif sequence;
 - b) culturing the transformed cells;
 - 10 c) isolating the expressed polypeptide.
2. The method of claim 1, wherein the number of G quartet motif sequences is between 1 and 5.
3. The method of claim 1, wherein the amount of the polypeptide produced is
15 increased as the number of G quartet motif sequences is increased
4. The method of claim 1, wherein in the expression vector the G quartet motif sequence is situated 3' of the polypeptide open reading frame.
- 20 5. The method of claim 1, wherein in the expression vector the G quartet motif sequence is situated within the polypeptide open reading frame.
6. The method of claim 1, wherein the G quartet motif sequence encodes an RNA consensus sequence of:
25 $WGGN_{(1-4)}WGGN_{(1-4)}WGGN_{(1-4)}WGG$ (SEQ ID NO: 3).
7. The method of any one of claims 1 to 6, wherein the polypeptide is a human polypeptide.
- 30 8. The method of any one of claims 1 to 6, wherein the polypeptide is an animal polypeptide.
9. The method of any one of claims 1 to 6, wherein the polypeptide is a plant polypeptide.

10. The method of any one of the previous claims, wherein the polypeptide is a recombinant polypeptide.
- 5 11. The method of any one of the previous claims, wherein the eukaryote cell is further transformed with an expression vector encoding the MID1 protein.
12. The method of any one of the previous claims, wherein the eukaryote cell is further transformed with an expression vector encoding the $\alpha 4$ protein.
- 10 13. Use of a G quartet motif sequence to increase production of a polypeptide in a eukaryote host cell, wherein the eukaryote cell is transformed with an expression vector comprising a polynucleotide sequence that encodes an open reading frame for the polypeptide and at least one G quartet motif sequence.
- 15 14. An expression vector comprising a polynucleotide encoding an open reading frame for a polypeptide and at least one G quartet motif sequence, wherein the G quartet motif sequence increases production of the polypeptide in a eukaryote host cell.
- 20

FIGURES

Figure 1

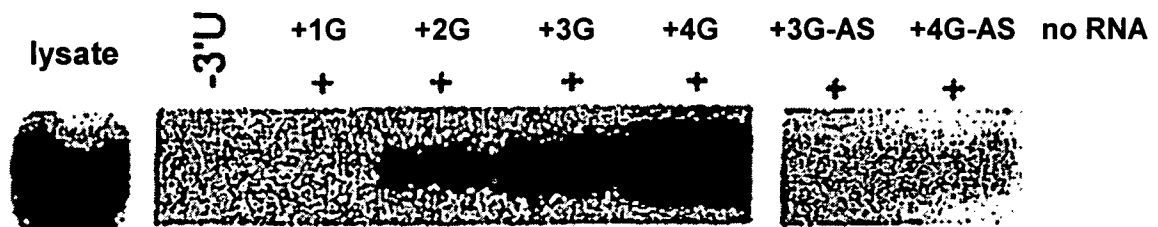


Figure 2

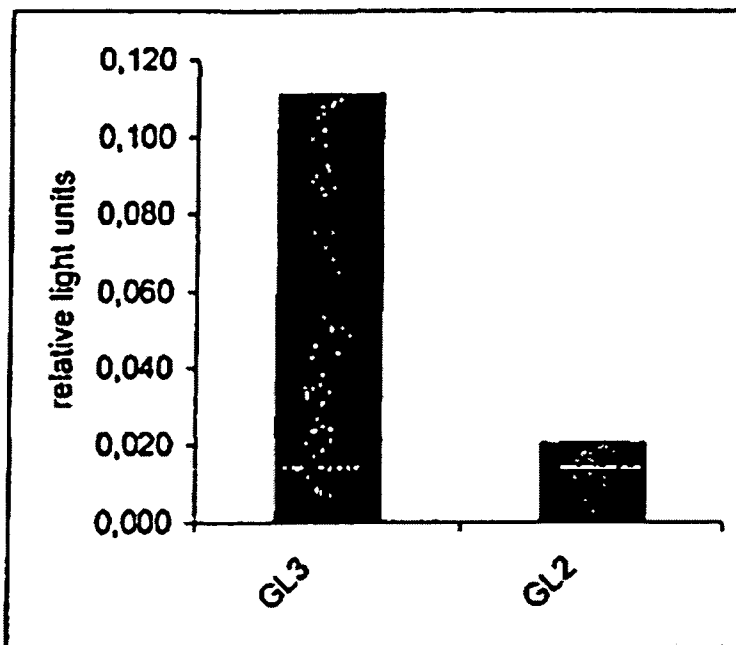


Figure 3

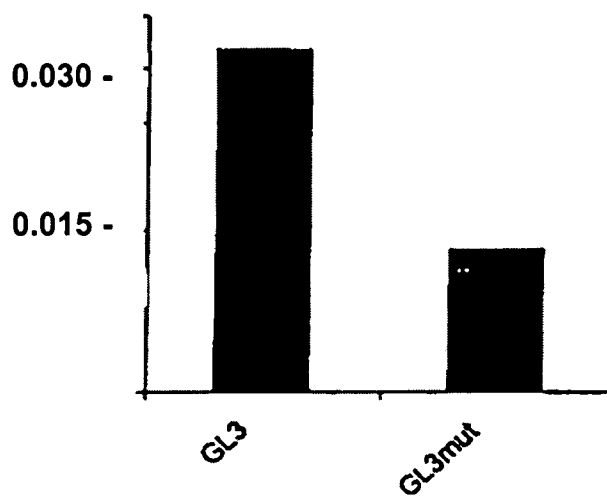


Figure 4a

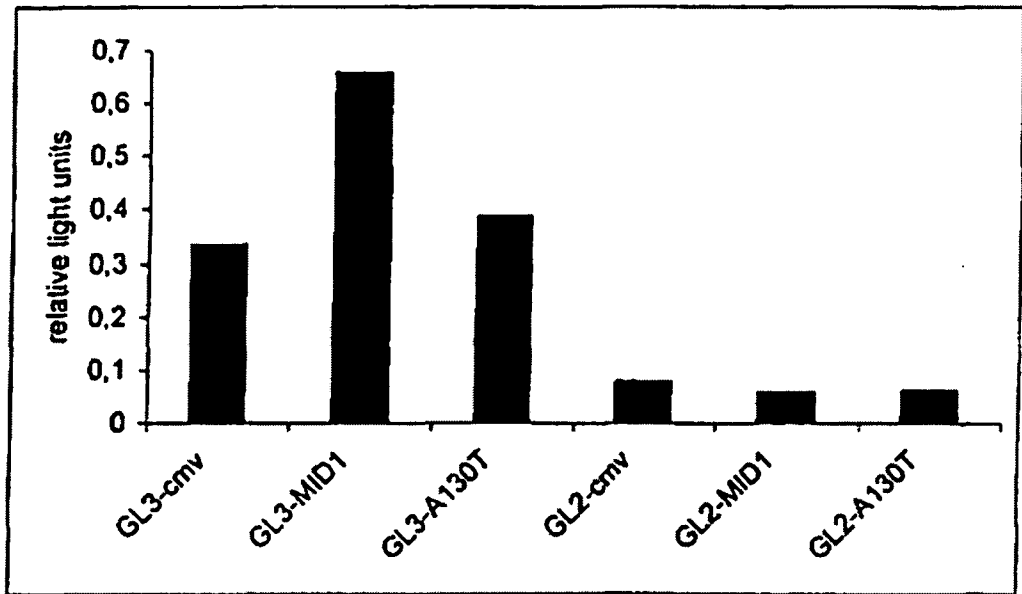


Figure 4b

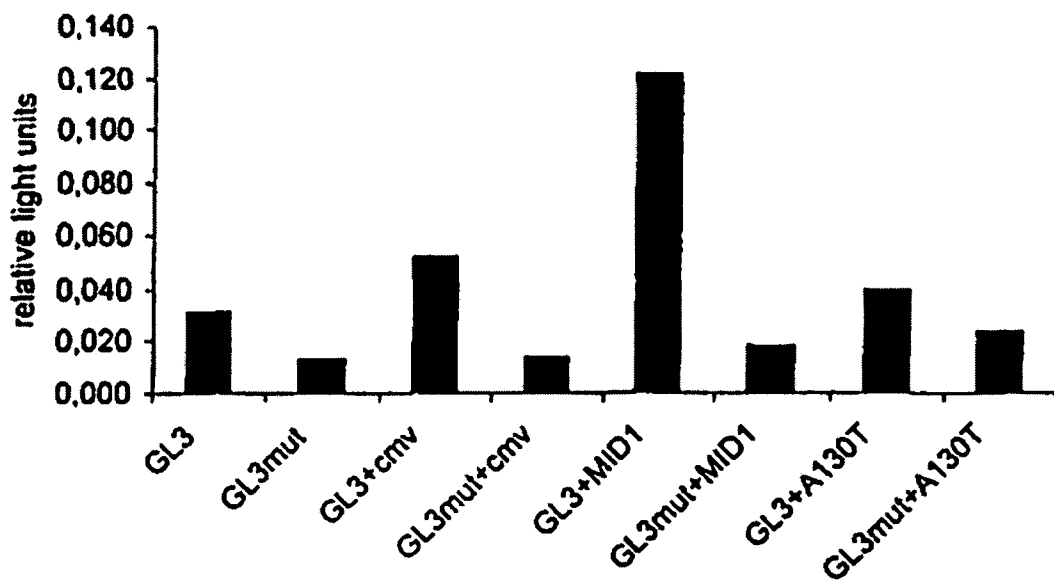


Figure 5a

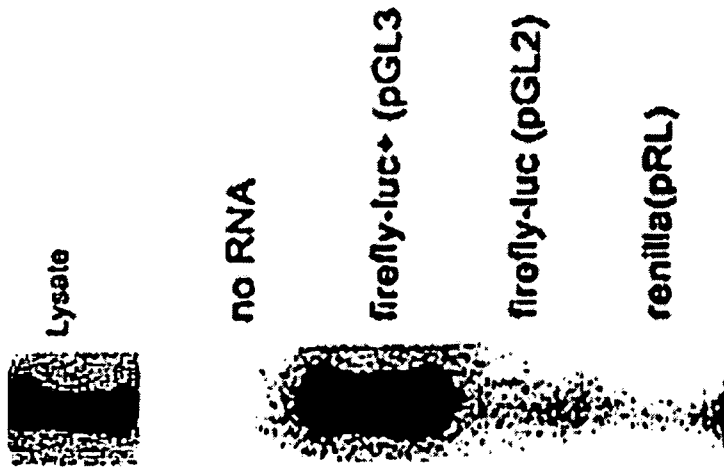


Figure 5b

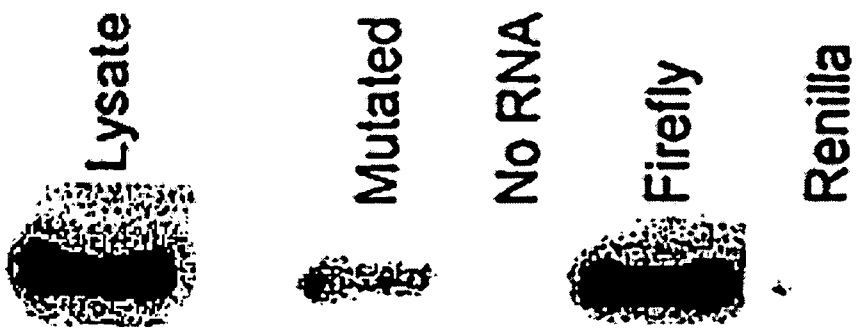


Figure 6a

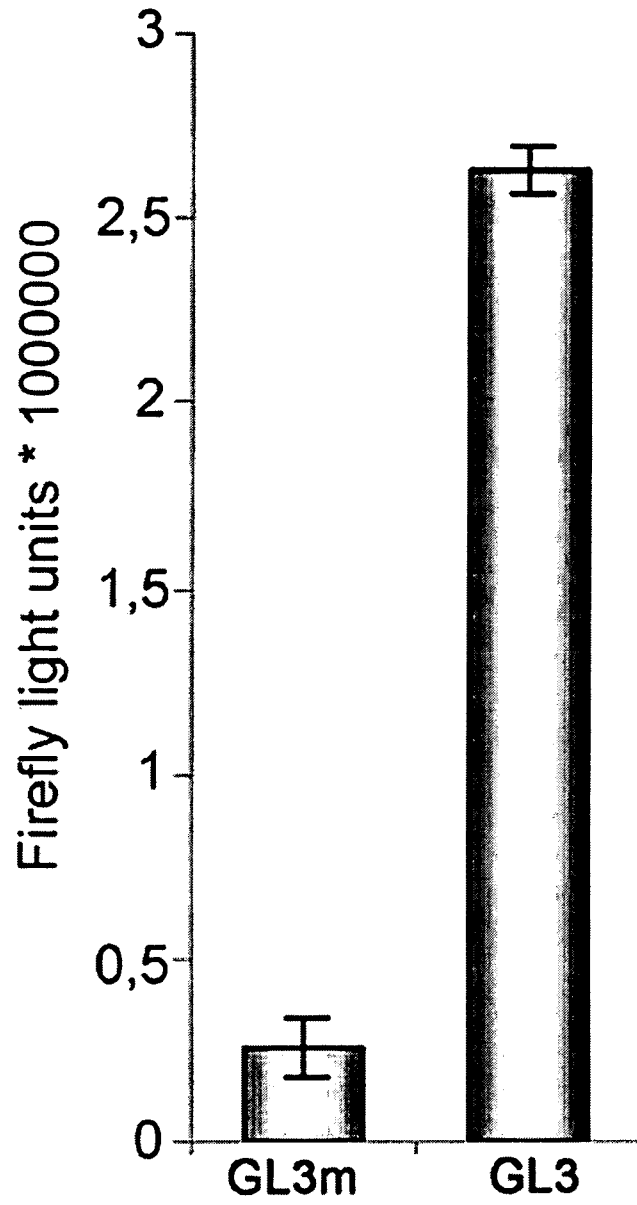


Figure 6b

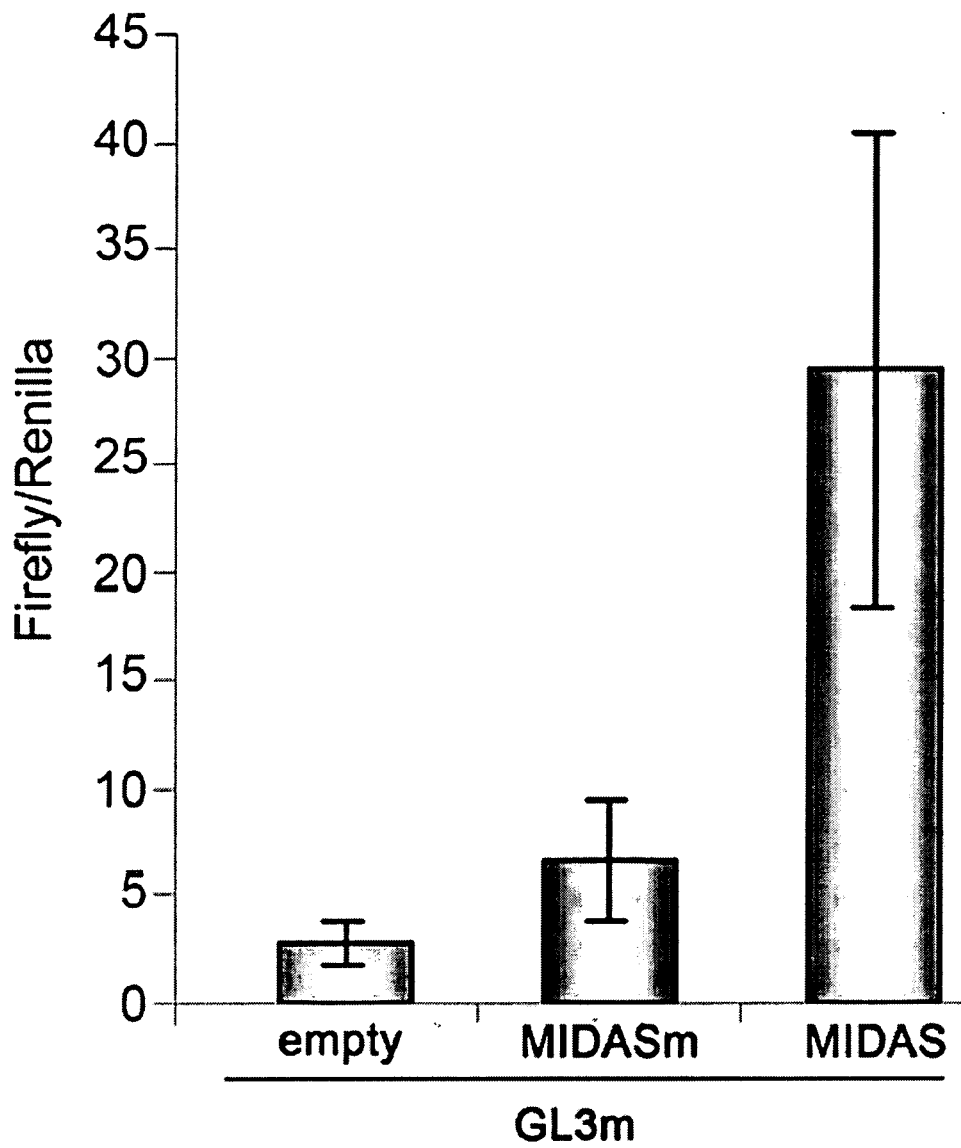


Figure 6c

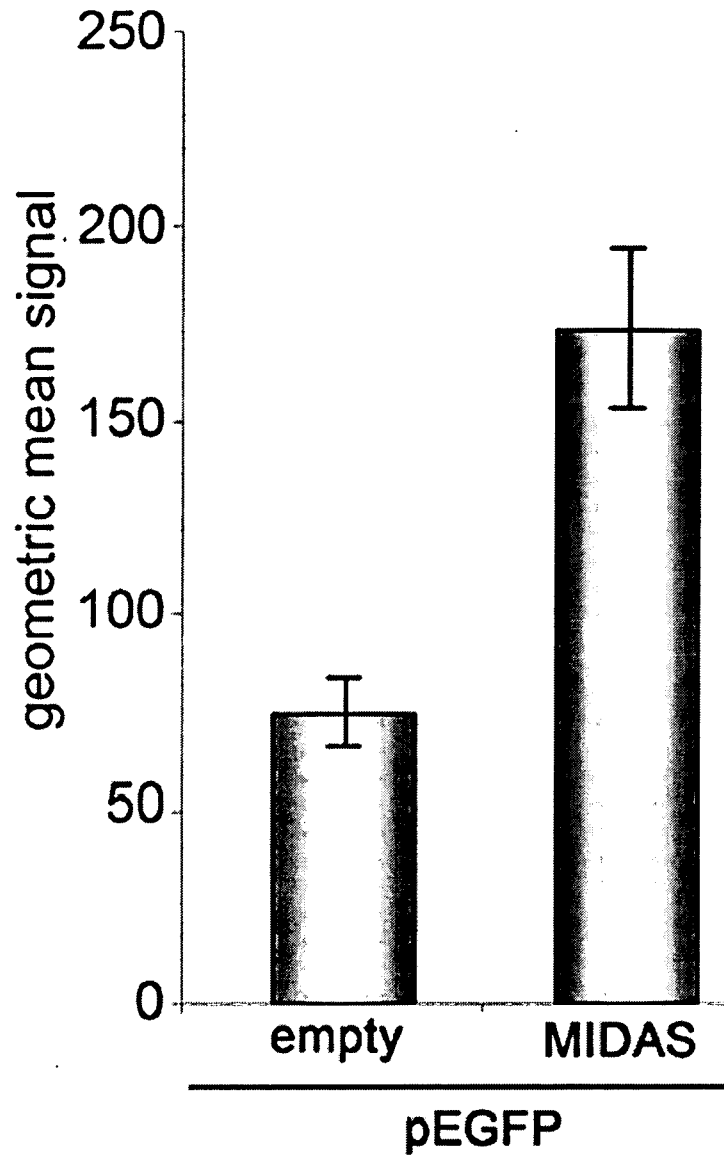


Figure 6d

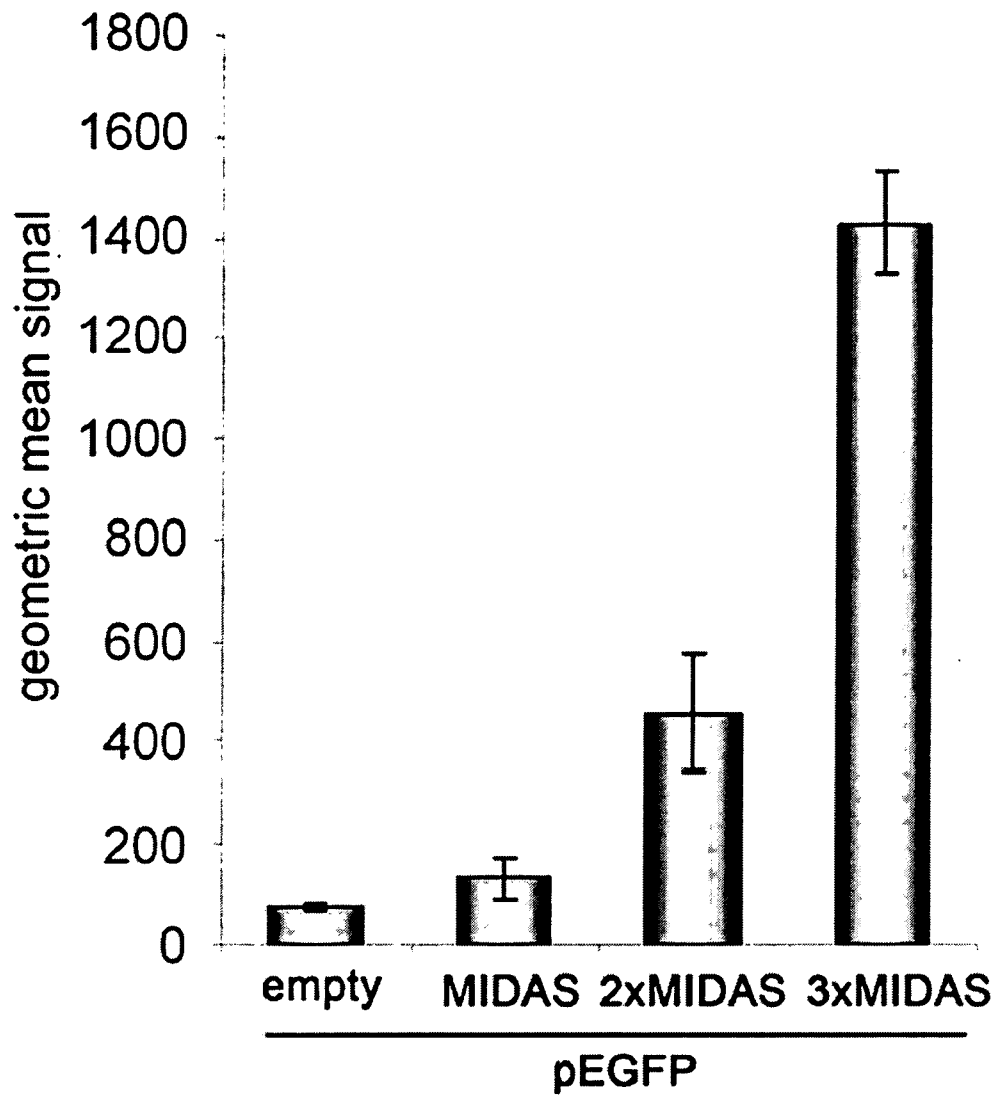


Figure 7a

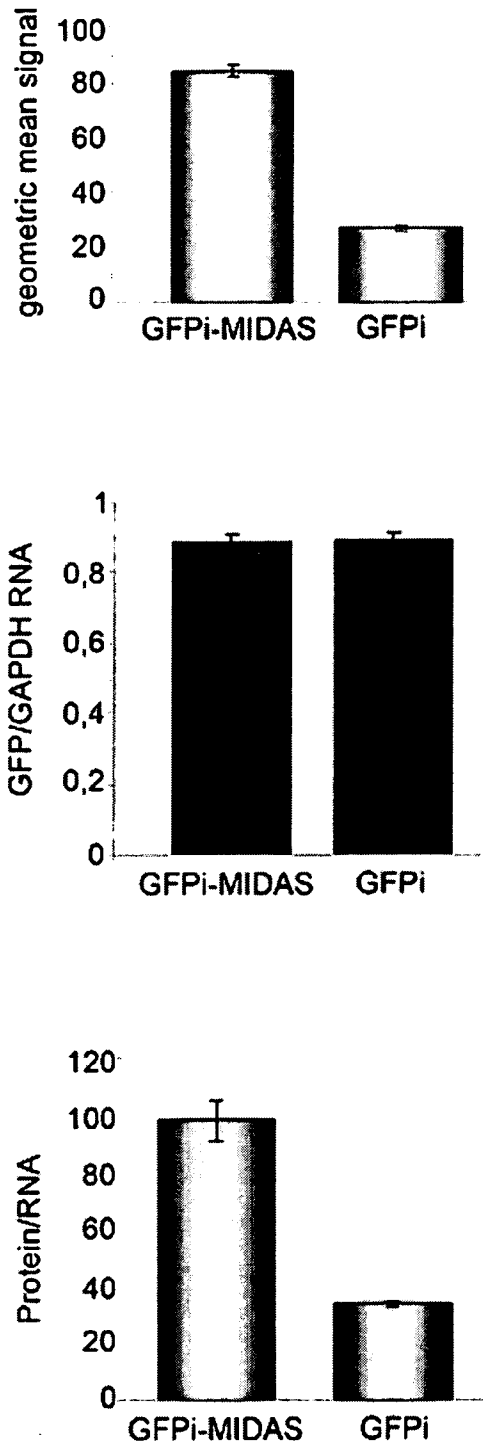


Figure 7b

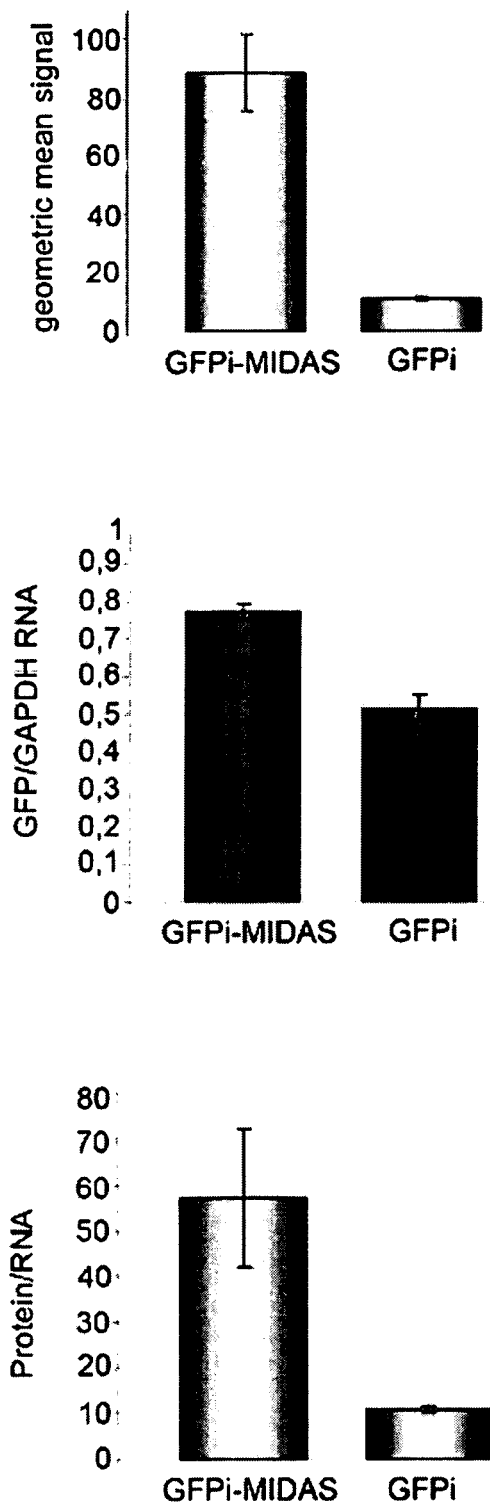


Figure 7c

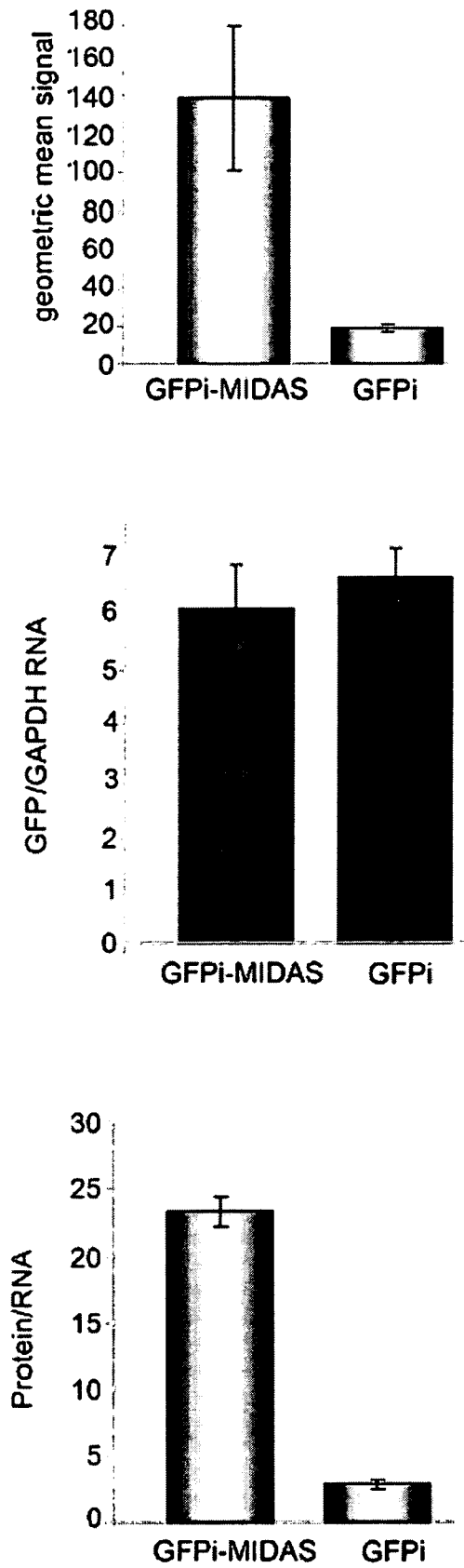
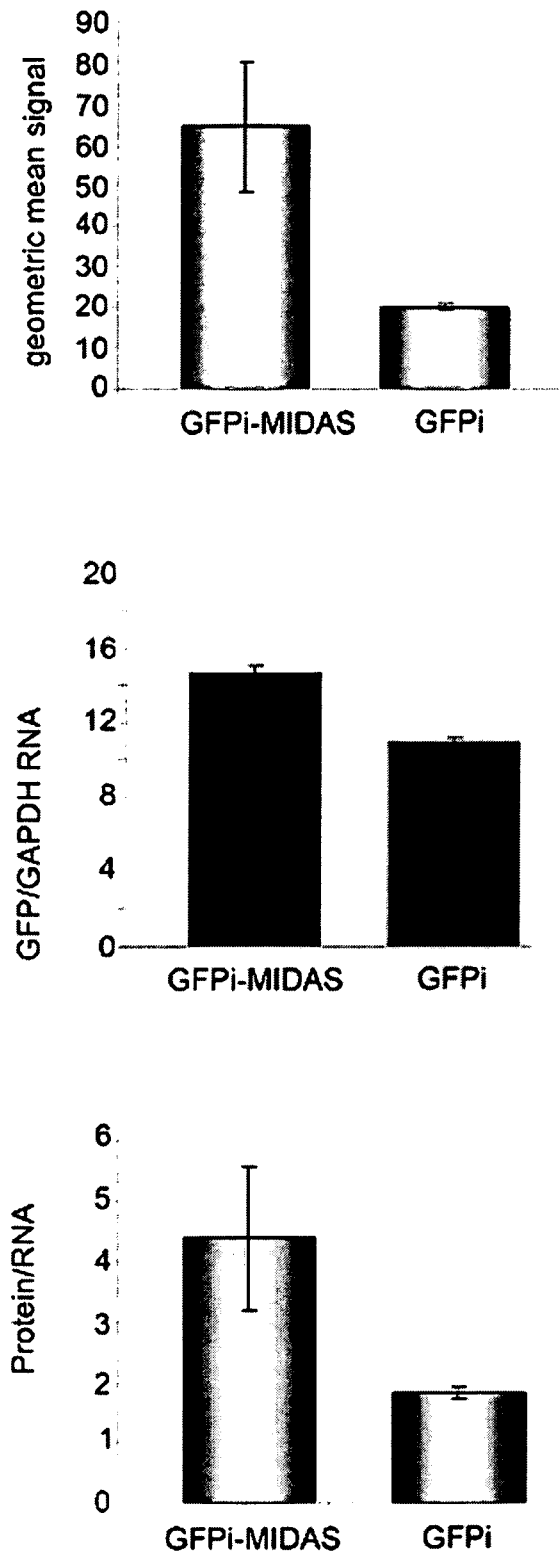


Figure 7d



INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2008/006511

A. CLASSIFICATION OF SUBJECT MATTER

INV. C12N15/67

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, BIOSIS, WPI Data, PAJ

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BONNAL SOPHIE ET AL: "A single internal ribosome entry site containing a G quartet RNA structure drives fibroblast growth factor 2 gene expression at four alternative translation initiation codons." THE JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 278, no. 41, 10 October 2003 (2003-10-10), pages 39330-39336, XP002465266 ISSN: 0021-9258 page 39335, right-hand column, paragraph 2	1-5, 7-10,14
X	WO 99/14346 A (SEQUITUR INC [US]) 25 March 1999 (1999-03-25) page 5, line 4; claim 25 page 18	1-5, 7-10,14
	----- -/--	

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *&* document member of the same patent family

Date of the actual completion of the international search

7 November 2008

Date of mailing of the international search report

19/11/2008

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Cupido, Marinus

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2008/006511

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2006/022712 A (PTC THERAPEUTICS INC [US]; CAO LIANGXIAN [US]; MEHTA ANURADHA [US]; NA) 2 March 2006 (2006-03-02) page 25	1-5, 7-10,14
X	BEATRIZ ARANDA-ORGILLES: "Characterisation of the MID1/alpha4 multiprotein complex" [Online] April 2006 (2006-04), FREIE UNIVERSITÄT BERLIN, XP002465269 Retrieved from the Internet: URL: http://www.diss.fu-berlin.de/2006/660/indexe.html * Title Page + Chapter 3 - Results * pages 49-88 page 85 - page 88	6,11,12
A	RAMOS A ET AL.: "G-quartet-dependent recognition between the FMRP RGG box and RNA." RNA (NEW YORK, N.Y.) OCT 2003, vol. 9, no. 10, October 2003 (2003-10) pages 1198-1207, XP002465267 ISSN: 1355-8382 page 1199	1-14
A	XU YAN ET AL: "Structural and functional characterizations of the G-quartet and i-motif elements in retinoblastoma susceptibility genes (Rb)." NUCLEIC ACIDS SYMPOSIUM SERIES (2004) 2005, no. 49, 2005, pages 177-178, XP002465268 ISSN: 1746-8272 the whole document	1-14

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No PCT/EP2008/006511

Patent document cited in search report	A	Publication date	Patent family member(s)	Publication date
WO 9914346	A	25-03-1999	AU 9319398 A	05-04-1999
			CA 2304982 A1	25-03-1999
			EP 1021549 A2	26-07-2000
			JP 2002508299 T	19-03-2002
WO 2006022712	A	02-03-2006	CA 2574076 A1	02-03-2006
			EP 1786933 A1	23-05-2007
			JP 2008507271 T	13-03-2008