Novel Nucleic Acid Molecules

Field of the invention

This invention relates *inter alia* to novel nucleic acid molecules, vectors containing them, eukaryotic host cells and host cell lines transformed with said vectors and their use in protein production, especially production of proteins containing non-natural amino acids.

Introduction

It has been observed that one orthogonal RS/tRNA pair, naturally evolved in a subset of archaebacteria, (methanogenic archaea bacteria that catabolize methylamines) has specificity for the amino acid pyrrolysine. Pyrrolysine uses a 21st aminoacyl-tRNA synthetase, naturally evolved to be orthogonal to all other amino acids and tRNAs.

Pyrrolysine is a natural amino acid, the only one that is authentically specified by an amber codon. Blight et al., 2004 showed that PylRS and tRNApyl can incorporate Pyrrolysine at amber codons in E. coli. They also showed that the wt PyLRS is naturally promiscuous and can incorporate analogs of lysine.

Yokoyama et al (EP1911840) demonstrated that the PylRS/tRNA system is orthogonal in eukaryotic cells and showed the incorporation of several nnAAs into a target proteins encoded by amber codons in bacterial cells. These authors also identified key amino acid residues in pylRS that form the amino acid binding pocket and function in selecting pyrrolysine over other canonical amino acids. Mutations at this site generated mutants able to recognize and aminoacylate the tRNApyl with AzZ-lys (Yanagisawa 2008).

This orthogonality extends to bacteria and eukaryotic cells.

PylRS is a naturally promiscuous synthetase that has naturally evolved to exclude lysine, but will incorporate lysine analogs without mutation, including azides, alkynes and alkenes, (Yanagisawa et al, 2008; Neumann et al. 2008; Mukai et al., 2008; Nguyen et al., 2009) . The basis of this specificity is dependent on hydrophobic interactions between amino acid residues of the pylRS binding pocket with the pyrrole ring of pyrrolysine that stabilizes and correctly positions the amino acid in the active site of the synthetase (Kavran et al., 2007). This RS/tRNA pair has been introduced via transient transfection into bacterial, yeast and mammalian cells and shown to be effective for incorporation of a number of non-natural amino acids into target proteins.

For instance, EP 1911840 demonstrates incorporation of N-ε-boc-Lysine into a target protein in *E. coli* cells.

The expression of tRNA in eukaryotic cells is carried out by RNA polymerase III ("PolIII"), requiring a type 2 intragenic promoter that contains a bipartite structure, in which two short conserved sequence elements known as an A-box and a B-box are separated by a variable sequence. In contrast to mammalian tRNA genes, prokaryotic tRNA genes are regulated by extragenic promoter elements. As a consequence, many prokaryotic tRNAs do not contain A-box and B- box elements that function as promoters in mammalian cells. This is the case for the *Methanosarcina* derived tRNApyl.

Introduction of a tRNApyl coding sequence into mammalian cells, along with PylRS, did not result in amber suppression of a reporter construct, presumably due to the lack of conserved A- and B- box elements (Mukai, et al., 2008).

Hancock et al (Hancock et al, 2010. JACS) attempted to reconstruct a functional A-Box and B-Box within the tRNApyl gene from *Methanosarcina barkeri*, and failed to demonstrate functional amber suppression of the mutated tRNA genes in yeast, suggesting that any modifications to introduce an A- and B- box elements into the tRNA sequence alters the tRNA architecture, hindering its overall function. Sequence analysis of the *Mm* tRNApyl shows that the B-box has significant homology to a consensus B-box, but the A-box shows significant divergence.

Attempts to find optimal conditions to express suppressor tRNAs in mammalian cells date back to 1982 (Hudziak et al, 1982, Laski et al, 1982) in the context of generating nonsense suppressor tRNAs for gene therapy of genetic diseases where a stop codon caused early termination of transcription of a key protein.

The amount of DNA introduced into cells, the number of copies of the tRNA gene as well as the promoter elements driving its expression have been identified as key elements to reach optimal tRNA level, e.g. to allow efficient activity without causing toxicity to the cell.

The expression of tRNA by processing of a bicistronic construct has been shown to exist in eukaryotic cells. A genomically encoded gene pair for tRNAarg –tRNAasp has been described in yeast cells (Schmidt et al. 1980). This gene is expressed as a single precursor RNA that is posttranscriptionally modified to generate two mature tRNAs in vivo and in oocyte cell extracts

(Schmidt et al. 1980). In this case the tRNA genes are separated by a short spacer region consisting of CTTTGTTTCT (SEQ ID No. 68). This gene arrangement has also been used to express tRNAs that are not normally expressed in yeast (Francis et al 1990). Here, the human *tRNAmet* gene, which contains consensus A and B- box elements, but is not transcribed in yeast, was successfully expressed in yeast cells when placed downstream of a tRNAarg element. The authors show that the distance of the two RNA genes affects transcriptional efficiency. Expression of the exogenous tRNA was achieved when the two tRNAs were separated by a short 10 bp thymidine rich spacer found between the yeast Arg-Asp tRNA pair. Expression of the *tRNAmet* gene was also observed when this spacer length was increased (up to 92bp) of intervening sequence but with lower efficacy (Francis et al., 1990). In another example, Straby (1988) utilized the bicistronic yeast Arg-Asp tRNA gene pair to express the normally transcriptionally silent suppressor tRNAsup (tyr). In one construct, the tRNAsup replaced the tRNAasp gene and retained the short 10 bp thymidine rich spacer found between the yeast Arg-Asp tRNA pair and showed efficient tRNAsup expression.

tRNAs have also been expressed under external promoters such as the U6 promoter (Koukuntla et al., 2013 Mukai et al. 2008), T7 RNA polymerase promoter (Mukai et al. 2008), SNR52 (Wang and Wang, 2012) and as a component of a bicistronic construct regulated by an upstream tRNA gene (Hancock et al.,2010; Francis et al., 1990; Straby 1988; Schmidt et al., 1980; Mukai et al. 2008); as multimers such as the tRNAser (Buvoli et al., 2000)

A variety of extragenic promoters were examined for the expression of tRNApyl by Mukai et al. (see also US 8,168,407) as well as by Hancock et al. These included a human tRNA, e.g. tRNAval or tRNAarg as bicistronic constructs. Whilst the bicistronic system was functional in yeast cells (see Hancock et al), Mukai et al. (2008) reported that in eukaryotic cells placing *tRNApyl* gene downstream to a eukaryotic tRNA (*tRNAval*) gene resulted in suboptimal expression of tRNApyl, which was not ideal as levels of expression were too low for optimal amber suppression in mammalian cells.

Further, Mukai et al. (2008) and Hancock et al. (2010) tested the PolII dependent CMV promoter for the expression of tRNApyl as well as the phage T7 polymerase promoter, and the PolIII/type 3 promoter of the gene for the small nuclear RNA U6. In particular, the authors showed that the U6 promoter linked to the 5' end of the coding sequence for the tRNApyl, was

particularly efficient in causing amber suppression when introduced into eukaryotic cells along with the PylRS, a nnAA and a target including an amber codon.

The promoters for the snRNA gene U6 and H1 have been extensively used for the expression of RNA species in mammalian cells. These promoters are particularly useful as expression tools given their function as extragenic promoters. Furthermore, they enable very high expression of genes that are placed downstream of these elements.

However, U6 and H1 promoters have recently been shown to have deleterious effects in cells transfected with these constructs leading to cytotoxicity that is believed to occur due to the high transcriptional demands of this promoter (Ehlert, et al., 2010) (Giering, et al., 2008) (Stegmeier, et al., 2005). Furthermore, the U6 promoter in particular has been shown to affect the localization of the RNA product, leading to its prevalent accumulation in the nucleus, thus posing a problem for its utilization by cytoplasmic machineries (Paul et al., 2003).

While the use of extragenic promoters described above led to the expression of high levels of tRNApyl, it is likely that a majority of the tRNA expressed is mislocalised in the nuclear compartment and unavailable for aminoacylation in the cytosol. Although efficient amber suppression was observed, it is possible that this was due to the relatively small proportion of tRNA that escaped the nucleus.

Hence, it is desirable to engineer an alternative tRNA promoter to allow for better localization of the tRNA and its availability to the translation machinery.

It is also desirable to engineer a tRNA promoter that obviates the cellular toxic effects that have been reported for external promoters such as U6 and H1 promoter elements which may contribute to low levels of protein expression and loss in cell viability in cell lines stably expressing said tRNA constructs.

The RNA polymerase III (sometimes referred to herein as “pol III”) transcription machinery is devoted to the production of non-protein coding RNAs (ncRNA) of small size in eukaryotes (reviewed by Dieci et al. 2013).

Pol III-transcribed genes are bound and activated by transcription factors that are selectively recognized by this polymerase . These factors act in a few different combinations, as specified by correspondingly differentpromoter architectures. These have been classified into three general classes: Type 1 promoters (such as that used by 5S rRNA) contain intragenic elements

(promoters are found within the coding sequence of the gene) and consist of an internal control region that is subdivided into an A-box, a C-box and an intermediate element (IE). These elements are necessary for transcription, as they are bound by the basal transcription factors TFIIIC and TFIIIA. Type 2 promoters are used by tRNA genes and contain intragenic promoters that consist of conserved elements known as A-box and B-box. These are the binding sites for the transcription factor TFIIIC, which in turn recruits the initiation factor TFIIIB on an upstream region extending up to position -40 with respect to the transcription start site . Type 3 promoters are extragenic, upstream located promoters and consist of distal and proximal sequence elements and a TATA-box. The promoters for the U6 snRNA and RNase P (H1) RNA are members of this promoter class. These promoters are uniquely recognized by the transcription factor SNAPc (PBP or PTF), which in turn recruits a specific variant of TFIIIB. Interestingly PolIII/Type3 promoters have been adapted for the expression of inhibitory RNA and the expression of tRNA lacking A- and B- box elements in eukaryotic cells, for instance U6 promoters in mammalian cells (Mukai, et al., 2008) and the SNR52 promoter used in yeast (Wang and Wang, 2012).

The expression of several ncRNA genes is regulated by promoter elements placed upstream of the transcribed region, similarly to Type 3 promoters, in combination with intragenic promoters, e.g. A-Box and B-Box used by tRNA genes. These can be considered hybrid promoters – called "Type 4" promoters due to the presence of both extragenic and intragenic control elements. Genes regulated by these promoters include the 7SL (SRP) RNA gene, Vault RNA gene, and the Epstein Barr virus coded small RNA (EBER).

7SL RNA is the RNA component of the signal recognition particle (SRP) that mediates co-translational insertion of secretory proteins into the lumen of the endoplasmic reticulum. Its promoter elements were characterized *inter alia* in Englert et al., 2004.

Vault RNAs are small RNAs contained in large cytoplasmic ribonucleoprotein particles involved in macromolecular assembly and transport. Its promoter elements were characterized *inter alia* in Mossink et al., 2002.

EBER gene is a small RNA of unknown function. EBER is encoded by the Epstein-Barr virus genome that is efficiently expressed in mammalian cells and its regulatory promoter elements well characterized (Howe and Shu, Cell 1989). EBER has two known variants- EBER1 and EBER2, EBER2 being expressed at higher levels.

Summary of the Invention

The inventors have engineered novel DNA constructs for the expression of the *tRNApyl* genes in eukaryotic cells, especially mammalian cells, under new and improved promoter systems.

The *tRNApyl* gene sequence possesses two internal regions that resemble an eukaryotic A box and B box, with the B box-like region more closely resembling a functional B box.

Although previous attempts at reconstituting a consensus A- Box and B-box were unsuccessful as reported in Hancock et al (2010) and Mukai et al. (US 8,168,407), the inventors have now surprisingly found that the *tRNApyl* sequence can be altered to enable a functioning intragenic promoter and obtain a tRNApyl able to mediate efficient amber suppression in combination with WT pylRS.

Such new *tRNApyl* gene can be used to generate highly active and stable cell lines for the incorporation of nnAAs into cells.

The inventors have also found that the new modified *tRNApyl* gene containing a functional intragenic promoter element can be further improved by placing them downstream of the 5' regulatory elements of genes expressed under type 4 promoters, thereby reconstituting a functional type 4 promoter element containing both extragenic and intragenic elements.

The inventors have also surprisingly found that the WT *tRNApyl* gene can be expressed under transcriptional control of a *tRNAglu* gene and/or a *tRNAasp* gene, when said *tRNAglu* gene and/or *tRNAasp* gene is placed upstream of the *tRNApyl* gene and altered to lack the transcription termination sequence in order to effectively form a bicistronic message.

DNA constructs bearing tandem repeats of novel *tRNApyl* genes of the invention have shown to lead to increased amber suppression.

Brief Description of the Figures:

Figure 1. DNA constructs of the invention (SEQ ID Nos: 69 – 71)

Figure 2. Putative A and B- box regions of the wild type tRNApyl from *Methanosarcina mazei* (SEQ ID No: 3).

Figure 3.Sequence alignment among WT and mutant M.mazei and M.barkeri *tRNApyl* genes (SEQ ID Nos: 72, 3, 73, & 74, respectively)

Figure 4. tRNA mutants retain function when expressed by the H1 promoter, as demonstrated by amber suppression-dependent GFP expression

Figure 5. Functional tRNA expression shown from use of mutant tRNApyl comprising an intragenic polIII promoter, demonstrated by increased amber suppression-dependent GFP expression

Figure 6. tRNA mutants are functional and retain orthogonality in mammalian cells

Figure 7. Increasing gene dose by increasing the number of genes in the same construct, results in increased amber suppression-dependent GFP expression

Figure 8. type 4 5' promoter elements allow expression of a tRNA containing a functional B-box, as determined by amber suppression-dependent GFP expression

Figure 9. type 4 5' promoter elements and the tRNA B mutant alone enable higher levels of tRNA function than the wild type tRNA.

Figure 10. Increasing gene dose by increasing the number of genes in the same construct, results in increased amber suppression-dependent GFP expression

Figure 11. *tRNApyl* expression driven by upstream *tRNAglu* and *tRNAasp* genes.

Brief description of the sequence listing:

SEQ ID No 1: tRNApyl *Methanosarcina barkeri, WT*

SEQ ID No 2: tRNApyl *Methanosarcina acetivorans, WT*

SEQ ID No 3: tRNApyl *Methanosarcina mazei, WT*

SEQ ID No 4: tRNApyl *Methanococcoides burtonii, WT*

SEQ ID No 5: tRNApyl *Desulfobacterium hafniense, WT*

SEQ ID No 6: Terminator sequence used for WT and mutant A –box and/or B-box tRNApyl constructs

SEQ ID No 7: tRNApyl *Methanosarcina mazei A10G mutation, tRNA-A1*

SEQ ID No 8: tRNApyl *Methanosarcina mazei A10G, A52C mutations, tRNA-AB*

SEQ ID No 9: tRNApyl *Methanosarcina mazei A52C mutation, tRNA-B*

SEQ ID No 10: tRNApyl *Methanosarcina mazei A10G, T14A, A52C mutations, tRNA-A2B*

SEQ ID No 11tRNApyl *Methanosarcina mazei A10G, T14A mutation, tRNA-A2*

SEQ ID No 12: 7SL Promoter 5' region sequence

SEQ ID No 13: Spacer used in 7SL constructs

SEQ ID No 14: 7SL tRNA-WT construct

SEQ ID No 15: 7SL tRNA-B construct

SEQ ID No 16: 7SL tRNA-A1 construct

SEQ ID No 17: 7SL tRNA-AB construct

SEQ ID No 18: 7SL tRNA-A2B construct

SEQ ID No 19: Terminator sequence used in the 7SL, Vault, EBER constructs

SEQ ID No 20: Vault promoter 5' region sequence

SEQ ID No 21: Spacer used for Vault constructs

SEQ ID No 22: Vault tRNA-WT constructs

SEQ ID No 23: Vault tRNA-B construct

SEQ ID No 24: Vault tRNA-AB construct

SEQ ID No 25; Vault tRNA-A2B construct

SEQ ID No 26: EBER2 Promoter 5' region sequence

SEQ ID No 27: Spacer used in EBER constructs

SEQ ID No 28: EBER2 tRNA-WT constructs

SEQ ID No 29: EBER2 tRNA-B construct

SEQ ID No 30: EBER2 tRNA-A1 construct

SEQ ID No 31: EBER2 tRNA-AB construct

SEQ ID No 32: EBER2 tRNA-A2B construct

SEQ ID No 33: H1 promoter sequence

SEQ ID No 34: H1 tRNA-WT construct

SEQ ID No 35: H1 tRNA-B construct

SEQ ID No 36: H1 tRNA-A1 construct

SEQ ID No 37: H1 tRNA-A2 construct

SEQ ID No 38: H1 tRNA-A2B construct

SEQ ID No 39: Terminator sequence used in the H1-tRNA constructs:

SEQ ID No 40: M. Musculus tRNAglu coding sequence

SEQ ID No 41: M. musculus tRNAglu coding sequence with upstream and downstream regions

SEQ ID No 42 : tRNAglu-tRNApyl DNA construct

SEQ ID No 43: 3x tRNAglu-tRNApyl: DNA construct including 3 repeats of the bicistronic construct.

SEQ ID No 44: Leader sequence for the tRNAglu-pyl constructs

SEQ ID No 45: Terminator sequence used in the tRNAglu-pyl constructs

SEQ ID No 46 : Terminator sequence used in the tRNAglu-pyl constr

SEQ ID No 47: Intergene sequence used in the tRNAglu-pyl constructs

SEQ ID No 48 : M.Musculus tRNAasp coding sequence

SEQ ID No 49 : M.Musculus tRNAasp coding sequence with upstream and downstream regions

SEQ ID No 50 : tRNAasp-pyl DNA construct

SEQ ID No 51: 2x tRNAasp-pyl: DNA construct including 2 repeats of the bicistronic construct

SEQ ID No 52: Leader sequence used in the tRNAasp-pyl construct

SEQ ID No 53: Terminator sequence used in the tRNAasp-pyl construct

SEQ ID No 54: intergene sequence used in the tRNAasp-pyl constructs

SEQ ID No 55: human tRNAglu coding sequence

SEQ ID No 56 : human tRNAasp coding sequence

SEQ ID No 57: 7SL promoter variant: 7SL-1

SEQ ID No 58:7SL promoter variant: 7SL-2

SEQ ID No 59:7SL promoter variant: SL28

SEQ ID No 60: 7SL promoter variant: 7L63

SEQ ID No 61:7SL promoter variant: 7L23

SEQ ID No 62:7SL promoter variant: 7L7

SEQ ID No 63: EBER promoter variant: EBER1

SEQ ID No 64: Vault promoter variant: Hvg3

SEQ ID No 65: Vault promoter variant: Hvg2

SEQ ID No 66: Methanosarcina mazei pylRS amino acid sequence

SEQ ID No 67: Methanosarcina mazei pylRS nucleotide sequence

SEQ ID No 68: short spacer region

SEQ ID No 69: a DNA construct shown in Figure 1

SEQ ID No 70: a DNA construct shown in Figure 1

SEQ ID No 71: a DNA construct shown in Figure 1

SEQ ID No 72: a DNA sequence shown in Figure 3

SEQ ID No 73: a DNA sequence shown in Figure 3

SEQ ID No 74: a DNA sequence shown in Figure 3

SEQ ID No 75: proximal sequence element of the Vault promoter

SEQ ID No 76: SP1 binding site of the EBER promoter

SEQ ID No 77: ATF binding sequence of the EBER promoter

Embodiments of the invention

According to a first aspect of the invention, there is provided a DNA construct which comprises a *tRNApyl* coding sequence and a RNA polymerase III promoter sequence which is capable of acting to express functional tRNAPyl sufficiently to support nonsense suppression (particularly amber suppression) in a eukaryotic expression system.

Suitably, the *tRNApyl* coding sequence of the invention lacks the final three nucleotides, CCA, which are added posttranscriptionally in eukaryotic cells.

Suitably, the DNA construct of the invention comprises a terminator sequence which is downstream of the *tRNApyl* coding sequence. Suitably, a terminator sequence consists of a run of 4 or more thymidine nucleotides 3' of the tRNA coding sequence that functions as a signal for termination of transcription.

Preferred terminator sequences of the invention are identified as SEQ ID Nos 6, 19, 39, 45,46 and 53, especially, SEQ ID No. 6.

In an embodiment, a DNA construct of the invention comprise 2 to 60 repeats of a *tRNApyl* gene. Preferably the DNA construct comprises 2 to 30 repeats, more preferably 8 repeats.

Suitably, the DNA construct of the invention comprises a *tRNApyl gene c*omprising the *tRNApyl* coding sequence and a functional intragenic RNA polymerase III promoter. Said intragenic RNA polymerase III promoter comprises two components – an A box and a B box.

Suitably, the invention provides a DNA construct wherein the *tRNApyl* gene contains mutations in 1 or 2 nucleotide positions in the A box and/or the B box of the intragenic RNA polymerase III promoter.

For example, the invention provides a DNA construct wherein the *tRNApyl* gene contains mutations in 1 or 2 (e.g. 1) nucleotide positions in the B box.

In an embodiment of the invention, the DNA construct comprises a *tRNApyl* gene comprising an intragenic PolIII promoter, e.g. as shown in Figure 1D, which is capable of acting to express functional tRNAPyl sufficiently to support nonsense suppression in a eukaryotic expression system.

In a preferred embodiment of the invention, there is provided a DNA construct wherein the *tRNApyl* gene has or comprises a sequence selected from SEQ ID Nos 7, 8, 9, 10 and 11 derived from *Methanosarcina mazei* (especially SEQ ID Nos 9 or 10) or has or comprises the sequence of an analogous *tRNApyl* gene derived from another bacterial species in which the mutations indicated by bold in SEQ ID Nos 7, 8, 9, 10 or 11 (especially SEQ ID Nos 9 or 10) are made in equivalent positions.

The determination of equivalent positions may be made by conventional alignment programs, e.g. using BLASTN.

Particularly preferred DNA constructs of the present invention comprise a *tRNApyl* gene of SEQ ID Nos 9 or 10, and a terminator sequence downstream of the *tRNApyl* coding sequence, which terminator sequence is SEQ ID No 6.

In a further, preferred embodiment, the invention provides a DNA construct comprising multiple copies of the *tRNApyl* gene comprising an intragenic RNA polymerase III promoter. Preferably the DNA construct comprises 2 to 30 copies, more preferably 8 copies of the *tRNApyl* gene comprising an intragenic RNA polymerase III promoter.

In an embodiment, the invention provides a DNA construct wherein the *tRNApyl* coding sequence is operably linked to a RNA polymerase III promoter wherein said promoter is placed 5' to the t*RNApyl* coding sequence.

Suitably, the 5' promoter and the *tRNApyl* coding sequence are separated by a spacer sequence containing a transcriptional start site.

Preferably, the spacer sequence contains a transcription start site and 4 to 6 nucleotides from the end of the promoter sequence to the start of the tRNA coding sequence. Preferred spacer sequences to be used in the present invention are listed as SEQ ID NOs 13, 21 and 27.

Suitably a DNA construct comprises an intragenic promoter element and a 5' promoter element being a RNA polymerase III promoter element placed 5' to the t*RNApyl* coding sequence such that the combination of said intragenic promoter element and said 5' promoter element constitutes a hybrid promoter,e.g. as shown in Figure 1C.

Suitably, said 5' promoter element is the 5' element of a Type 4 RNA polymerase III promoter. Thus it typically comprises an upstream promoter element typical of the pol III/type 3 class of promoters, for example it comprises a TATA box, (such as one used in a U6 or H1 promoter). Suitably, said intragenic promoter element comprises an element typical of the polIII/type 2 promoters, for example an A-and B-box (e.g. such as used in eukarotic tRNA promoters).

It will be understood that "an element" is not restricted to single nucleotide sequence and can be construed to include more than one nucleotide sequence. Thus for example an A and B box together can represent "an intragenic promoter element". The A and B box may be viewed as separate components of the intragenic promoter element.

Preferably, the 5' promoter element of a Type 4 RNA Polymerase III promoters is selected from 5' promoter elements of the EBER RNA gene promoter, the 7SL RNA gene promoter and the Vault RNA gene promoter, particularly the 7SL RNA gene promoter.

The EBER RNA gene promoter may be, for example EBER variant EBER1 or EBER2, especially EBER2,

The 7SL RNA gene promoter may be, for example, variant 7SL-1, 7SL-2, SL28, 7L63, 7L23 or 7L7.

The Vault RNA gene promoter may be, for example, variant Hvg2 or Hvg3.

Preferably, said 5' promoter element is selected from SEQ ID Nos 12, 20, 26, 57, 58, 59, 60, 61, 62, 63, 64 and 65.

Suitably, the 5' promoter element and the *tRNApyl* coding sequence are separated by a spacer sequence containing a transcriptional start site. Suitably, the spacer sequence contains a transcription start site and 4-6 nucleotides from the end of the promoter sequence to the start of the tRNA coding sequence. Preferred spacer sequences to be used in the present invention are listed as SEQ ID NOs 13, 21 and 27.

Preferably, the Type 4 RNA polymerase III promoters are promoters from eukaryotic genes, more preferably mammalian genes, more preferably, human genes.

Particularly preferred 5' promoter elements are selected from SEQ ID Nos 12, 20 and 26.

In a preferred embodiment of the invention, there is provided a DNA construct which is capable of acting to express functional tRNAPyl sufficiently to support nonsense suppression in a eukaryotic expression system comprising a *tRNApyl* gene comprising a *tRNApyl* coding sequence which contains mutations in 1 or 2 nucleotide positions in the putative A box and/or B box of an intragenic promoter element which is operably linked to a 5'promoter element being an RNA polymerase III promoter placed 5' to the *tRNApyl* gene, and wherein the combination of said intragenic promoter element and said 5' promoter element constitutes a hybrid promoter.

A particularly preferred DNA construct within the context of the present invention comprises a *tRNApyl* gene comprising a *tRNApyl* gene coding sequence and an intragenic promoter element wherein said gene has or comprises a sequence selected from SEQ ID No 7, 8, 9, 10 and11, derived from *Methanosarcina mazei* or has or comprises the sequence of an analogous

tRNApyl gene derived from another bacterial species in which the mutations indicated by bold in SEQ ID Nos 7, 8, 9, 10 and 11 are made in equivalent positions, which is capable of acting to express functional tRNAPyl sufficiently to support nonsense suppression in a eukaryotic expression system, wherein the *tRNApyl* coding sequence is operably linked to a 5' promoter element being an RNA polymerase III promoter element which is placed 5' to the *tRNApyl* coding sequence, separated by a spacer sequence and wherein the combination of said 5' promoter element and intragenic promoter constitutes a hybrid promoter.

Thus, preferred DNA constructs of the invention are selected from SEQ ID NOs 15, 16, 17, 18, 22, 23, 24, 25, 28, 29, 30, 31 and 32 , particularly SEQ ID 15, 16, 17, 18, 23, 25 and 29, especially SEQ ID Nos 15, 18 and 23.

An aspect of the invention provides a multimeric DNA construct comprising multiple copies (e.g. 2 to 60 copies e.g. 2 to 30 copies e.g. 8 copies) of any of the DNA constructs as aforesaid.

Particularly preferred DNA constructs of the invention comprise 8 copies of the *tRNApy*l gene selected from SEQ ID NOs 15, 16, 17, 18, 22, 23, 24, 25, 28, 29, 30, 31 and 32.

Most preferred *tRNApy*l genes repeated multiple times in DNA constructs of the present invention are selected from SEQ ID Nos 15, 18 and 23.

In an alternative embodiment, the invention provides a DNA construct wherein a *tRNApyl* gene is operably linked to a eukaryotic tRNA gene which comprises a RNA polymerase III type 2 promoter and expressed as a bicistronic message e.g. as shown in Figure 1A.

Suitably, the eukaryotic tRNA gene lacks the 3' terminator sequence, thereby allowing expression of the *tRNApyl* gene to occur under the control of the eukaryotic RNA polymerase III type 2 promoter.

Preferably, the eukaryotic tRNA gene operably linked to the *tRNApyl* is *tRNAglu* or *tRNAasp* which lead to high level of expression of the tRNApyl

More preferably, the eukaryotic tRNA gene operably linked to the *tRNApyl* gene is a mammalian *tRNAglu* or a *tRNAas*p , more preferably murine or human.

Thus the invention provides a DNA construct which comprises a eukaryotic tRNAglu gene or a eularyotic *tRNAasp* gene 5' to a *tRNApyl* gene, said *tRNAglu* gene or *tRNAasp* gene being deprived of a termination sequence thus leading to a bicistronic message upon transcription.

The *tRNApyl* gene may be a wildtype gene. Alternatively, the *tRNApyl* gene may contain mutations in 1 or 2 nucleotide positions in the A box and/or the B box e.g. as described above.

Suitably, the first tRNA is separated from the second tRNA by a spacer sequence consisting of a sequence devoid of thymidine nucleotides which constitute a signal for termination of transcription.

Preferred spacer sequences are defined in SEQ ID Nos 47 and 54.

Preferably, the eukaryotic tRNA gene operably linked to the *tRNApyl* gene is a mammalian *tRNAglu* or a *tRNAasp* of SEQ ID No 41, 49, 55 or 56; more preferably, a murine *tRNAglu* or *tRNAasp* of SEQ ID No 41 or 49.

In an embodiment, a murine *tRNAglu* or *tRNAasp* gene is used for expression of the tRNApyl gene in murine cell lines. In an embodiment, a human *tRNAglu* or *tRNAasp* geneis used for expression of the *tRNApyl* gene in murine cell lines

In a particularly preferred embodiment of the invention, it is provided a DNA construct wherein the *tRNApyl* gene is expressed as a bicistronic message of SEQ ID No 42 or SEQ ID No 50.

In a further embodiment, the first tRNA is preceded by a 5' leader sequence that may contain sequence elements such as a TATA box, a transcription start site (TSS), and sequences that regulate the post-transcriptional processing of the pre-tRNA to generate a mature tRNA.

The leader sequence may be selected from untranslated regions upstream of the coding sequence of any mammalian tRNA that is transcribed . Preferably, the leader sequence is selected from SEQ ID Nos 44 and 52. Alternatively, the leader sequence is comprised in the genomic untranslated region 5' to the first tRNA.

Particularly preferred constructs according to the present invention comprise multiple copies (e.g. 2-20 copies) of the bicistronic transcription unit of SEQ ID Nos 42 or 50. Thus, preferred DNA constructs of the present invention comprise, 2, 3, 4, 5, 6, 7 or 8 copies of the transcription unit of SEQ ID Nos 42 or 50. Most preferred DNA constructs according to the present invention are SEQ ID No 43, representing 3 copies of the tRNAglu-pyl bicistronic transcription unit of SEQ ID No 42, and SEQ ID No 51, representing 2 copies of the tRNAasp-pyl bicistronic unit of SEQ ID No 50.

As an aspect of the invention, any of the DNA constructs as aforesaid may comprise a PylRS coding sequence or gene.

According to an aspect of the invention there is provided a eukaryotic host cell which is transformed with a vector comprising a DNA construct as aforesaid.

According to a second aspect of the invention, there is provided a eukaryotic cell line which expresses or is capable of expressing PylRS and tRNApyl, in which tRNApyl is introduced into the cells with DNA constructs of the invention and which functions in said eukaryotic cell line.

According to a third aspect of the invention, there is provided a cell-free expression system wherein a synthesis reaction lysate obtained from from a host cell comprises at least one component required for the synthesis of polypeptides.

Suitably, the synthesis reaction lysate is obtained from bacterial or eukaryotic cells. Preferably, the synthesis reaction lysate is obtained from eukaryotic cells, more preferably, from rabbit reticulocytes or wheat germ.

Preferably, the cell-free expression system is capable of expressing WT PylRS and tRNApyl of the present invention, wherein tRNApyl is introduced into the cells used to obtain the synthesis reaction lysate with DNA constructs of the invention.

Cell-free expression systems suitable for use in the present invention are described for instance in WO201008110, WO2010081111, WO2010083148, incorporated in their entirety herein by reference.

A eukaryotic host cell according of the invention may be transformed with a vector comprising a gene encoding a target protein containing one or more non-natural amino acids encoded by a nonsense codon.

In an embodiment, a eukaryotic cell line expresses or is capable of expressing functional PylRS and functional tRNApyl in which functional tRNApyl expression occurs under the control of a an intragenic promoter element and a 5' promoter element being a RNA polymerase III promoter element placed 5' to the tRNApyl coding sequence such that the combination of said intragenic promoter element and said 5' promoter element constitute a hybrid promoter sufficiently to support nonsense suppression.

Suitably the nonsense suppression is amber suppression.

Suitably the host cell is a mammalian cell.

In a fourth aspect, the invention provides a process for preparing a target protein containing one or more non-natural amino acids encoded by a nonsense codon which comprises expressing said target protein in a eukaryotic cell line which is transformed with a gene encoding a target protein containing one or more non-natural amino acids encoded by a nonsense codon and which is also transformed with genes encoding PylRS and tRNApyl, such that PylRS and tRNApyl are expressed and function in said eukaryotic cell line sufficiently to support nonsense suppression.

Suitably the cell line is transformed with a DNA construct according to other aspects of the invention described herein.

Thus, the nonsense codon utilized in the embodiments of the invention might be an amber codon, an opal codon, or an ochre codon. A particularly preferred nonsense codon is an amber codon.

The invention provides a process for preparing a chemically modified target protein which comprises preparing a target protein according to an above aspect of the invention and chemically modifying the resultant target protein. The protein may be chemically modified via a side chain on a non natural amino acid.

Thus, the non-natural amino acids utilized in embodiments of the present invention, or at least one of them, may comprise an alkyne or azide moiety. These moieties are well adapted for chemical modification e.g. via the Huisgen reaction leading to generation of a triazole linker moiety.

In an fifth aspect of the present invention, there is provided a protein conjugate such as an antibody conjugate prepared using a process according to processes of the present invention in which the protein such as an antibody is conjugated to one or more other moieties selected from protein, drug and PEG moieties via linkers comprising a triazole moiety.

In a sixth aspect the present invention provides a process for increasing the viability of a eukaryotic cell line stably transformed to express PylRS and tRNApyl which comprises transforming the eukaryotic cell line with a tRNApyl gene which is introduced into the cells with DNA constructs of the invention and which functions in said eukaryotic cell line, in order to express a target protein containing one or more non-natural amino acids.

In a seventh aspect the present invention provides a process for increasing the viability of a eukaryotic cell line stably transformed to express a target protein containing one or more non-natural amino acids and to express PylRS and tRNApyl which comprises transforming the eukaryotic cell line with genes encoding PylRS and tRNApyl such that PylRS and tRNApyl are expressed and function sufficiently to support nonsense suppression.

Definitions and abbreviations:

As used herein, "gene" means a locatable region of genomic sequence, corresponding to a unit of inheritance, which is associated with regulatory regions, transcribed regions, and or other functional sequence regions. Thus a gene will include a coding sequence (e.g. a coding sequence for an RNA such as a tRNA) together with regulatory regions including promoter elements.

As used herein, "intragenic promoter" means a promoter contained within the gene sequence.

As used herein, a "transcribed region" includes both a coding sequence, i.e. a sequence which is part of the mature RNA product, and portions of the primary transcript that are removed and thus do not appear in the mature RNA, such as the 5' leader portion of tRNA precursors.

As used herein "promoter element" means an element of a promoter, especially of a hybrid promoter that may have a number of elements (e.g. an intragenic promoter element and a 5' promoter element). As used herein, "Type 4 RNA polymerase III promoter" means an RNA polymerase III promoter having an upsteam promoter element or elements (e.g. typical of the type 3 class of RNA polymerase III promoters, for example a TATA box used in U6 or H1 promoters) and an intragenic promoter element or elements (e.g. typical of the type 2 class of RNA polymerase III promoters, for example an A and B box, for example as used in tRNA promoters).

As used herein, a TATA box is a DNA sequence typically containing the sequence TATAAA or a variant thereof found in the promoter region of certain genes, located upstream of the transcription start site.

As used herein, a "nonsense codon" means an amber, ochre or opal codon.

As used herein "nonsense suppression" means suppression of chain termination by a nonsense codon. .

As used herein, "nnAA" or "non-natural amino acid" means a amino acid which is suitable for incorporation into proteins and is not one of the 20 amino acids naturally coded for by the genetic code, Pyrrolysine and Selenocysteine.

As used herein, "WT" means wild type i.e. a form naturally obtained in nature.

As used herein, "RS" is an abbreviation for "tRNA synthase".

As used herein, "TSS" is an abbreviation for "transcription start site".

As used herein, "CPE" is an abbreviation for "cyclic AMP response element".

As used herein, "PSE" is an abbreviation for "proximal sequence element".

As used herein, "CHO" is an abbreviation for "Chinese hamster ovary".

As used herein a "coding sequence" is the nucleotide sequence of a gene which codes for the gene product (tRNA, RS etc) and does not include non-coding sequences , such as regulatory sequences.

As used herein, there may be transcribed sequences which are part of the primary transcript but not of the mature RNA, and are therefore considered "non-coding".

As used herein, a "DNA construct" means an artificially constructed segment of nucleic acid to be introduced into an expression system. A DNA construct typically includes an expression cassette and may also contain nucleotides belonging to the vector system used to introduce the nucleic acid into the expression system.

As used herein, an "expression cassette" means an artificially constructed segment of nucleic acid to be introduced into an expression system which comprises one or more gene-coding sequences such as cDNA or RNA coding sequences, along with any necessary regulatory regions for optimal expression. An expression cassette does not contain nucleotides belonging to the vector system.

As used herein, a "bicistronic" transcript is a transcript with the potential to code for two final products.

Detailed description

tRNApyl to be expressed in cell lines according to the invention

The *tRNApyl* to be expressed in combination with the PylRS of the present invention has an anticodon and a tertiary structure which are complementary to the amber nonsense codon UAG, in order to function as a suppressor tRNA.

An artificial tRNA could be constructed that is complementary to other nonsense codons such as UGA, opal; UAA, ochre codons in order to function as a suppressor tRNA.

Thus it will be understood that although the present invention is substantially described and exemplified by reference to use of the amber codon for coding the nnAA and with discussion of the concept of amber suppression, the amber codon can be replaced with an another nonsense codon such as opal or ochre codons and would be expected to work in the same way.

However use of amber codon is preferred.

Preferably, the *tRNApyl* of the present invention is a *tRNApyl* derived from one of the following bacterial strains: *Methanosarcina mazei* (SEQ ID No 3), *Methanosarcina barkeri* (SEQ ID No 1), *Desulfitobacterium hafniense* (SEQ ID No 5), M*ethanosarcina acetivorans* (SEQ ID No 2), *Methanosarcina burtonii (SEQ ID No 4), or Methanosarcina thermophila.*

More preferably, the *tRNApyl* of the present invention is a *tRNApyl* derived from *Methanosarcina mazei* (SEQ ID No 3)

Suitably, the *tRNApyl* of use in aspects of the present invention is not a *tRNApyl* derived from *Methanosarcina barkeri* (SEQ ID No 1)

Engineering of *tRNApyl* sequences in order to optimize expression in eukaryotic cell lines has been described in WO2007099854, WO12038706, and Hancock et al (2010) incorporated herein by reference.

WO2007099854 provides *inter alia* DNA constructs comprising a tRNA coding sequence deriving from Archaebacteria, preferably tRNApyl, a transcription terminator sequence placed 3' of said tRNA gene, a promoter sequence that induces transcription by RNA Polymerase II or III such as U1 snRNA promoter or U6 snRNA promoter placed 5' to said tRNApyl coding sequence.

WO12038706 provides *inter alia* DNA constructs comprising *tRNApyl* from *Methanosarcina barkeri* operably linked to a RNA polymerase III promoter, in particular linked to the RNA polymerase III promoter of a yeast tRNAarg gene.

WO12038706 further provides DNA constructs for expression in yeast cells, comprising tRNApyl from *Methanosarcina barkeri* mutated at several internal positions, summarized in Figure 3 . Interestingly, M.Barkeri tRNApyl differs by one nucleotide from M.mazei tRNApyl, corresponding to nucleotide G at position 3 of SEQ ID No. 3

tRNApyl with intragenic polIII promoter

A *tRNApyl* gene included in DNA constructs of the present invention may comprise an intragenic polIII promoter.

Hancock et al (2010) described DNA constructs for expression of tRNApyl in yeast cells, comprising tRNApyl from *Methanosarcina barkeri* mutated at several internal positions, summarized in Figure 3. In yeast, the introduction of a mutation in the B-box resulted in the generation of the consensus B box sequence, and led to very low but detectable levels of the mutant tRNApyl expression. However, expression of the mutant *tRNApyl* did not support amber suppression in the yeast cells. The authors concluded that either the tRNA was improperly folded or processed, or that the mutation abolished synthetase recognition.

Similarly, the authors recreated a near-consensus A box which expression was not detectable and they combined the A-box and B-box mutations which led to detectable expression levels but no amber suppression.

In particular, the mutations generated by Hancock et al in the *M.Barkeri tRNApyl* gene correspond to *M.mazei* position 10, 14, 25, and 52. Interestingly,wild type *M.Barkeri tRNApyl* differs by one nucleotide from M.mazei *tRNApyl*, corresponding to nucleotide G at position 3 of SEQ ID No. 1

Surprisingly, the inventors have now discovered that introduction of certain mutations in the putative A-box and B-box regions of the *M.mazei tRNApyl* gene does result in successful expression of *tRNApyl* in mammalian cells, associated with efficient amber suppression.

The inventors have recreated a consensus B box and a near-consensus A box which function both in combination and separately to allow amber suppression, as discussed in Example 1 and Figures 4-7.

Suitably, the B box mutation is A52C referred to the WT sequence of SEQ ID 3.

Suitably, the A box mutation comprises two mutations which have been analysed together and separately and determined to lead to amber suppression in mammalian cells, corresponding to positions 10 and 14 referred to the WT sequence of SEQ ID No 3.

Exemplary *tRNApyl* mutant genes of SEQ ID Nos 7, 8, 9, 10 or 11 (corresponding to mutated sequences from the *tRNApyl* gene of Methanosarcina mazei) and analogous *tRNApyl* genes derived from other bacterial species (e.g. those mentioned above) in which a mutation (indicated by bold in SEQ ID Nos 7, 8, 9, 10, and 11) is made in an equivalent position. The mutations may, for example, be 1 or 2 nucleotide positions in the putative A box and/or the putative B box. For example a single mutation is made in the putative B box e.g. in the position shown in SEQ ID Nos 7, 8, 9, 10 or 11. For example a single or two mutations are made in the putative A box and a single mutation is made in the B box e.g. in the positions shown in SEQ ID No. 7-11. The putative A and B boxes are shown in Figure 2 and the skilled person can derive the putative A and B boxes regions for analogous *tRNApyl* genes derived from other bacterial species (e.g. those mentioned above).

Suitably, the DNA constructs of the present invention comprise a mutated *tRNApyl* gene followed by a terminator sequence. An exemplary terminator sequence is listed in SEQ ID 6

The proper folding of the tRNA is highly sensitive to any modification of its sequence. Surprisingly, the mutations in equivalent positions in the tRNA from *M.barkeri* and *M.mazei* has led to dramatically different results as shown in WO2012/038706 and in the examples of the present invention. Notably, the *M.barkeri tRNApyl* gene has a different nucleotide at position 3 (G) compared to the *M.mazei* tRNA where at the equivalent position there is an A.

It has been hypothesized that the G at position 3 decreased efficacy of RS aminoacylation, and that G at position 3 affects orthogonality of the *tRNApyl* in yeast. (Hancock 2010; Gundllapalli et al 2008).

Briefly, the recognition of specific tRNAs by the Aminoacyl tRNA synthetases and the tRNA aminoacylation occurs through identity elements specified by the tRNA sequence. The acceptor stem consists of 7 nucleotide pairs formed by the 3' and 5' terminal ends of the tRNA and is important for the recognition of tRNA by enzymes and critical in translation. The tRNApyl of *M. barkeri Fusaro* contains an unusual G3:70U base pair in the acceptor stem region that is not found in other *tRNApyl* genes. This mutation was shown to affect the orthogonality of the tRNA in yeast by allowing misacylation of the tRNApyl by the yeast seryl tRNA synthetase.

As a further comment to the The regulation of expression of polIII dependent genes can vary between yeast and mammalian cells. In both, tRNAs are governed by intragenic A and B-box elements, but extragenic sequences can affect the expression level, transcription start site (TSS) and the processing of pre-tRNA. In yeast, the TSS is frequently found 18-20 bp upstream of the T marking the first base of the A-box (or 10-12 bp upstream of the start of mature tRNA coding sequence). In addition, in yeast the TSS is often surrounded by a core promoter element tCAAca (where the capital letters indicate high conservation).

Surprisingly, the inventors have also found that DNA constructs built by assembling a series of repeats of the mutant tRNA genes of the invention provide enhanced expression and amber suppression, as exemplified in Figure 7 and Example 2

Suitably, DNA constructs of the invention comprise 2 to 60 repeats of a *tRNApyl* gene. Preferably the DNA construct comprises 2 to 30 repeats, more preferably 8 repeats.

tRNApyl with intragenic polIII promoter expressed under 5' promoter elements

The lack of success in producing a functional amber suppressor by enhancing the transcription of a functional tRNApyl by mutation of the A-and B-box sequences within the structural gene, led investigators in the field to engineering of DNA constructs wherein the transcription of the *tRNApyl* gene would be increased using only extragenic sequences, as reported above.

The inventors have now engineered a novel and improved promoter element which combines upstream promoter elements and intragenic promoter elements.

Certain eukaryotic RNA genes are expressed under a hybrid polymerase III promoter, or type 4 promoter, comprising upstream promoter elements typical of the pol III/type 3 class of promoters (eg TATA box, used in U6 or H1 promoter), or also possibly comprising sequences unrelated to type 3 promoter sequences, and intragenic promoter elements typical of the polIII/type 2 promoters (eg A-and B-boxes, used in tRNA promoters).

Suitably, upstream promoter elements utilized in the invention are selected from promoters of 7SL (SRP) RNA gene, Vault RNA gene, Epstein Barr virus coded small RNA (EBER) identified in SEQ ID Nos 12, 20, 26, 57, 58, 59, 60, 61, 62, 63, 64 or 65.

Suitably, upstream promoter elements of the invention are isolated from mammalian genes, preferably human or murine genes, most preferably human as in SEQ ID Nos 12, 20 or 26.

Suitably, the promoter sequences contain functional elements that recruit transcription factors that regulate transcription by PolIII RNA polymerase as shown in Figure 1E (SEQ ID Nos: 69 – 71).

Hence, the 7SL promoter contains an Activating transcription factor (ATF) or cyclic AMP response element (CRE) binding site (tgacgtcac) located 43-51 bp upstream of the transcription start site. The promoter also contains a TATA—like sequence called an ETAB sequence (ttctagtgct) 20-31 bp upstream of TSS.

The Vault promoter contains a proximal sequence element (PSE) located 23-53 bp upstream of the TSS (tgacgtaggtctttctcaccagtca) (SEQ ID No: 75) and is also a variant of the cyclic AMP response element (CRE). In addition the Vault promoter contains a TATA box (tataat) 16-22 bp upstream of the TSS.

The EBER promoter contains three regulatory regions, a SP1 binding site (gcacgcttaaccccgcctaca) (SEQ ID No: 76) 76-55 bp upstream of the TSS, an ATF binding sequence (accgtgacgtagctgttta) (SEQ ID No:77) 55-36 bp upstream of the TSS, and a TATA-like box (tatagag) 28-21 bp upstream of the TSS.

In the context of the present invention, the combination of upstream promoter elements of the 7SL, EBER, Vault RNAs with the reconstituted intragenic promoter (A- and/or B- box) within the *tRNApyl* gene has proven to improve the amber suppression function of such tRNA, over the *tRNApyl* containing the reconstituted intragenic promoter alone, and over the WT *tRNApyl,* as shown in Figure 8 and example 3.

Suitably, the 7SL promoter of SEQ ID No 12 is combined with the B-box mutation of SEQ ID No 9 , as in the construct of SEQ ID No 15.

Alternatively, the 7SL promoter of SEQ ID No 12 is combined with the A- and B-box mutation of SEQ ID No 10, as in in the construct of SEQ ID No 18.

Preferred DNA constructs of the present invention comprise an upstream promoter element, separated by the transcription start site by a spacer. Suitably, the spacer sequence contains a transcription start site and 4-6 nucleotides from the end of the promoter sequence to the start of the tRNA coding sequence.

DNA constructs of the invention comprise a downstream termination site, placed 3' of the *tRNApyl* coding sequence, wherein said termination site signals termination of polymerase III

transcription through a small poly-thymidine (4-6) sequence. Preferred terminator sequence of the invention is listed as SEQ ID No 19.

tRNApyl expressed under eukaryotic tRNA genes as bicistronic messages

Mukai et al, 2008 and EP1992698 describe how the tRNApyl gene downstream to a eukaryotic tRNA (*tRNAval*) gene led to a suboptimal expression of *tRNApyl*, not ideal as levels of expression were too low for optimal amber suppression in mammalian cells.

The present inventors have surprisingly found DNA constructs wherein the *tRNApyl* coding sequence is operably linked to a eukaryotic tRNA gene such as *tRNAglu* or *tRNAasp* (Figure 1A) lead to optimal expression of the tRNApyl gene product as well as amber suppression, as described in detail in Figure 11 and example 4

Suitably, the *tRNApyl* gene of the present invention is expressed as a bicistronic message in a eukaryotic host cell.

Suitably, the eukaryotic tRNA gene lacks the 3' terminator sequence, thereby allowing expression of the tRNApyl gene to occur under the eukaryotic RNA polymerase III type 2 promoter.

Hence, the first tRNA is separated from the second tRNA by a spacer sequence consisting of a sequence devoid of Thymidine nucleotides which constitute a signal for termination of transcription.

Exemplary spacer sequences are shown in SEQ ID Nos 47 and 54.

Preferably, the eukaryotic tRNA gene operably linked to the *tRNApyl* coding sequence is a mammalian *tRNAglu* or a *tRNAasp* gene, of SEQ ID No 41, 49, 55 or 56; more preferably, a murine *tRNAglu* or *tRNAasp* of SEQ ID No 41 or 49.

The inventors have surprisingly found that expression of the tRNApyl and amber suppression are particularly efficient when the first tRNA gene is preceded by a 5' leader sequence that contains sequence elements such as a TATA box, a transcription start site (TSS), and sequences that regulate the post-transcritpional processing of the pre-tRNA to generate a mature tRNA, collectively referred to as leader sequence.

Hence, the leader sequence is the 5' region of any mammalian tRNA gene that is transcribed to generate a portion of tRNA precursor that will not be part of the mature tRNA.

Particularly preferred leader sequences used in the present invention are disclosed in SEQ ID No 41 and 49 and listed as SEQ ID No 44 (tRNAasp) and SEQ ID No 52 (tRNAglu).

Particularly preferred DNA constructs of the invention are listed as SEQ ID No 42 and SEQ ID No 50.

SEQ ID No 42 consists of the murine *tRNAglu* gene followed by a spacer region (ccccaaacctc) devoid of Thymidines, linked to the *tRNApyl* gene. Thus, the construct of SEQ ID No 42 is transcribed as a single entity that is postrascriptionally processed to generate a mature *tRNApyl*.

SEQ ID No 50 consists of the murine *tRNAasp* gene followed by a spacer region (tagccccacctc) devoid of Thymidines, linked to the *tRNApyl* gene. Thus, the construct of SEQ ID No 50 is transcribed as a single entity that is postrascriptionally processed to generate a mature *tRNApyl*.

Particularly preferred DNA constructs of the present invention comprise multiple copies of the bicistronic transcription unit of SEQ ID Nos 42 and 50.

Thus, particularly preferred DNA constructs comprise, 2, 3, 4, 5, 6, 7, 8, copies of the transcription unit of SEQ ID Nos 42 or 50. Most preferred DNA constructs are SEQ ID No 43, representing 3 copies of the tRNAglu-pyl bicistronic transcription unit of SEQ ID No 42, and SEQ ID No 51, representing 2 copies of the tRNAasp-pyl bicistronic unit of SEQ ID No 50.

The *tRNApyl* genes of the invention can be introduced into eukaryotic cell lines. Suitably the cell lines are mammalian cell lines.

More preferably, the cell line is a CHO cell line, but also may be a HEK293, PERC6, COS-1, HeLa VERO, or mouse hybridoma cell line.

CHO and HEK293 cells lines are particularly suitable.

PylRS

Suitably, the *tRNApyl* genes of the present invention are introduced into eukaryotic cell lines in association with *pyl*RS genes .

As used herein, pylRS relates to an amino acyl tRNA synthetase which will aminoacylate a suitable tRNA molecule with pyrrolysine or a derivative thereof.

The pylRS of the present invention is suitably a Pyrrolysyl-tRNA Synthetase orthogonal in eukaryotic cells which is derived from methanogenic archaea spp.-i.e. it is wildtype in methanogenic archaea spp. or is a mutant thereof.

Preferably, the pylRS of the present invention is a Pyrrolysyl-tRNA Synthetase derived from one of the following: *Methanosarcina mazei, Methanosarcina barkeri, Desulfitobacterium hafniense*, M*ethanosarcina acetivorans, Methanosarcina burtonii , Methanosarcina thermophila, Methanosalsum zhilinae, Methanohalobium evastigatum, Methanohalophilus mahii, Desulfotomaculum gibsoniae, Desulfosporosinus meridei and Desulfotomaculum acetoxidans.*

Most preferably, the pylRS of the present invention is the pyrrolysyl tRNA synthetase (pylRS) derived from *Methanosarcina mazei*

The pylRS of the present invention may be a wild type synthetase.

Alternatively, the pylRS of the present invention may be mutated at one or more positions e.g. in order to increase its catalytic activity and/or to modify its selectivity for substrate amino acids (See for instance Yanagisawa 2008).

Preferably, the pylRS of the present invention may be mutated at position corresponding to Tyr 384 of or its equivalent. Most preferably, Tyr 384 is mutated into Phenylalanine.

In one embodiment, the pylRS of the present invention may be mutated at one or more positions in order to modify its substrate specificity and allow (or improve) incorporation of pyrrolysine analogs

Further mutant PylRS enzymes are described in WO09038195 and in WO2010114615, each document incorporated herein by reference in its entirety.

Incorporation of non-natural amino acids encoded for by nonsense codon

In proteins prepared using cell lines of the invention, one or more nnAAs may be incorporated. Suitably one nnAA is incorporated into a protein chain. In the case of the protein being an antibody, one nnAA may be incorporated into the light chain or the heavy chain or both.

In other embodiments more than one e.g. up to four e.g. two (or perhaps three) nnAAs may be incorporated into a protein chain. Suitably all the incorporated nnAAs are the same.

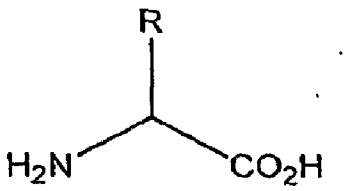
nnAAs are suitably encoded by an amber codon.

Non-natural amino acids that may be encoded by amber codon for incorporation into target proteins

Non-natural amino acids for incorporation into target proteins

The use of non-natural amino acids to allow for conjugating moieties to peptides is disclosed in WO 2007/130453, incorporated herein by reference.

As used herein a "non-natural amino acid" refers to any amino acid, modified amino acid, or amino acid analogue other than selenocysteine, pyrrolysine, and the following twenty alpha-amino acids: alanine, arginine, asparagine, aspartic acid, cysteine, glutamine, glutamic acid, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, valine. The generic structure of an alpha-amino acid is illustrated by Formula I:

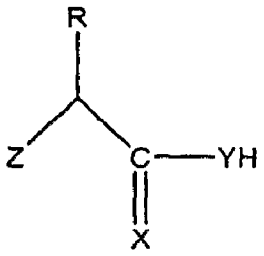


Formula I

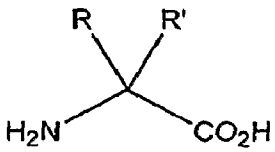
A non-natural amino acid is typically any structure having Formula I wherein the R group is any substituent other than one used in the twenty natural amino acids. See, e.g., any biochemistry text such as Biochemistry by L. Stryer, 3rd ed. 1988, Freeman and Company, New York, for structures of the twenty natural amino acids. Note that the non-natural amino acids disclosed herein may be naturally occurring compounds other than the twenty alpha-amino acids above. Because the non-natural amino acids disclosed herein typically differ from the natural amino acids in side chain only, the non-natural amino acids form amide bonds with other amino acids, e.g., natural or non-natural, in the same manner in which they are formed in naturally occurring proteins. However, the non-natural amino acids have side chain groups that distinguish them from the natural amino acids. For example, R in Formula I optionally comprises an alkyl-, aryl-, aryl halide, vinyl halide, alkyl halide, acetyl, ketone, aziridine, nitrile, nitro, halide, acyl-, keto-, azido-, hydroxyl-, hydrazine, cyano-, halo-, hydrazide, alkenyl, alkynyl, ether, thioether,

epoxide, sulfone, boronic acid, boronate ester, borane, phenylboronic acid, thiol, seleno-, sulfonyl-, borate, boronate, phospho, phosphono, phosphine, heterocyclic-, pyridyl, naphthyl, benzophenone, cycloalkynes such as the constrained ring such as a cyclooctyne, cycloalkenes such as a norbornenes, transcycloalkenes, cyclopropenes, tetrazines, pyrones, , thioester, enone, imine, aldehyde, ester, thioacid, hydroxylamine, amino, carboxylic acid, alpha-keto carboxylic acid, alpha or beta unsaturated acids and amides, glyoxyl amide, or organosilane group, or the like or any combination thereof.

In addition to non-natural amino acids that contain novel side chains, non-natural amino acids also optionally comprise modified backbone structures, e.g., as illustrated by the structures of Formula II and III:



Formula II



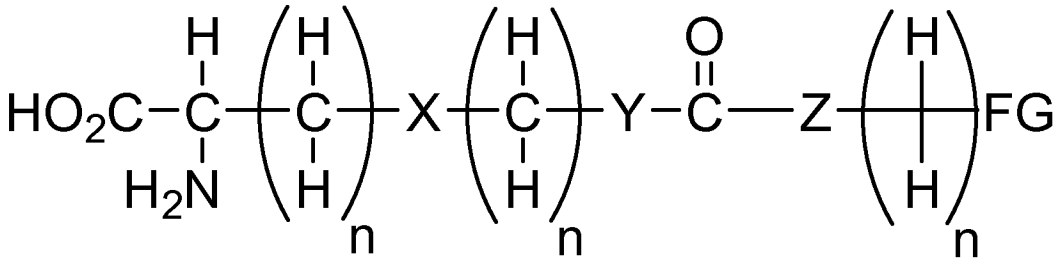
Formula III

wherein Z typically comprises OH, NH.sub.2, SH, NH.sub.2O--, NH--R', R'NH--, R'S--, or S--R'--; X and Y, which may be the same or different, typically comprise S, N, or O, and R and R', which are optionally the same or different, are typically selected from the same list of constituents for the R group described above for the non-natural amino acids having Formula I as well as hydrogen or (CH.sub.2).sub.x or the natural amino acid side chains.

Other examples of amino acid analogs include (but are not limited to) a non-natural analog of a Lysine or Pyrrolysine amino acid which include one of the following functional groups; an alkyl, aryl, acyl, azido, nitrile, halo, hydrazine, hydrazide, hydroxyl, alkenyl, cycloalkenes, alkynl, cycloalkynes, cycloalkynes such as the constrained ring such as a cyclooctyne, cycloalkenes such as a norbornenes, transcycloalkenes, cyclopropenes, aryl halide, vinyl halide, alkyl halide, aziridine, nitro, hydroxyl, ether, , epoxide, vinyl ethers, silyl enol ethers, thiol,

thioether,sulfonamide, sulfonyl, sulfone, seleno, ester, thioacid, boronic acid, boronate ester, borane, phosphono, phosphine, heterocyclic, pyridyl, naphthyl, benzophenone, tetrazines, pyrones, enone, imine, aldehyde, hydroxylamine, keto, thioester, ester, thioacid, organosilane group, amino,, a photoactivatable cross-linker; a spin-labeled amino acid; a fluorescent amino acid;; an amino acid that covalently or noncovalently interacts with another molecule; a metal binding amino acid; a metal-containing amino acid; a radioactive amino acid; a photocaged amino acid; a photoisomerizable amino acid; a biotin or biotin-analogue containing amino acid; a glycosylated or carbohydrate modified amino acid; a keto containing amino acid; an amino acid comprising polyethylene glycol; an amino acid comprising polyether; a heavy atom substituted amino acid; a chemically cleavable or photocleavable amino acid; an amino acid with an elongated side chain; an amino acid containing a toxic group

In an embodiment of the present invention, non natural amino acids (nnAA) of the general structure below are utilized for the production of proteins:



Preferably, the X group can be a methylene group, alkene, arene, oxygen, sulfur, phosphorus, nitrogen, ester, amide, carbonate, carbamate, ether, amine, thioether, alkyne, and heterocycle.

Prefereably, the Y group can be a methylene, alkene, arene, oxygen, sulfur, phosphorus, nitrogen, ester, amide, carbonate, carbamate, ether, amine, thioether, alkyne, and heterocycle.

Preferably, the Z group can be a nitrogen, oxygen, sulfur, phosphorus, α-methyleneamino, α-hydroxylamino, α-methlyleneazido.

The FG group can be an alkyl azide, alkyl-alkyne, alkyl-alkene, alkyl cyclohexene, alkyl cycloalkyne, alkyl aryl halide, aryl halide, amido cycloalkyne, amido cycloalkene, transcycloalkene, cyclopropenes, tetrazines, pyrones, norbornenes, aryl azide, azido, a hydroxyl amine, a hydrazide, a vinyl halide, a aryl halide, a tetrazine, a pyrone, a imine, boronic ester or acid, a cyano, a carbonyl group such as an aldehyde or ketone.

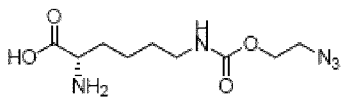
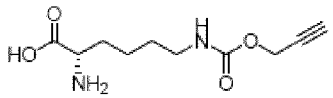
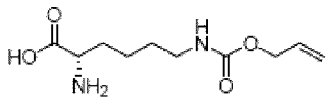
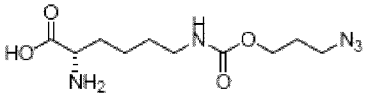
In a preferred embodiment, non natural amino acids (nnAA) of the general structure above can have methylene groups of varying length in which n=1-12. Preferably, non natural amino acids (nnAA) of the general structure above can contain cycloalkanes and aromatic rings as part of the connective structure.

Suitably, nnAAs of the present invention are derived from lysine as listed in the structure tables below. Those that are not commercially available are optionally synthesized as provided in the examples of US 2004/138106 A1 (incorporated herein by reference) or using standard methods known to those of skill in the art. For organic synthesis techniques, see, e.g., Organic Chemistry by Fessendon and Fessendon, (1982, Second Edition, Willard Grant Press, Boston Mass.); Advanced Organic Chemistry by March (Third Edition, 1985, Wiley and Sons, New York); and Advanced Organic Chemistry by Carey and Sundberg (Third Edition, Parts A and B, 1990, Plenum Press, New York), and WO 02/085923, all of which are hereby incorporated by reference.

nnAAs of the invention may be synthesized by published methods. For instance, synthesis of (S)-2-amino-6((prop-2-ynyloxy)carbonylamino)hexanoic acid and S)-2-amino-6((2azidoethoxy)carbonylamino)hexanoic acid is published in WO2010139948 and Nguyen et al. 2009, incorporated herein by reference.

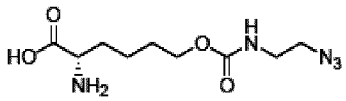
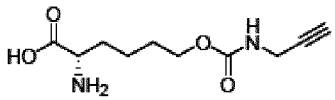
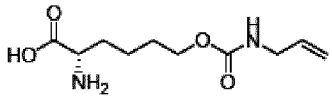
Suitably, nnAAs of the present invention are derived from lysine as shown in the table below as described in part in WO2010139948, incorporated herein by reference.

Lysine based analogs

Structure	Name
	(2S)-2-amino-6-{[(2-azidoethoxy)carbonyl]amino}hexanoic acid
	(2S)-2-amino-6-{[(prop-2-yn-1-yloxy)carbonyl]amino}hexanoic acid
	(2S)-2-amino-6-{[(prop-2-en-1-yloxy)carbonyl]amino}hexanoic acid
	(2S)-2-amino-6-{[(3-azidopropoxy)carbonyl]amino}hexanoic acid

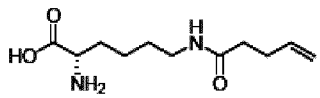
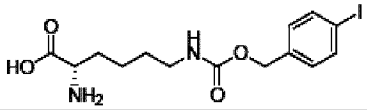
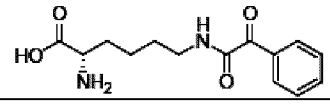
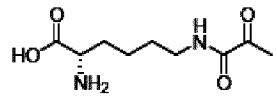
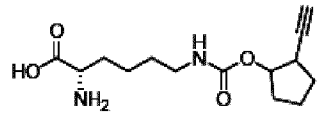
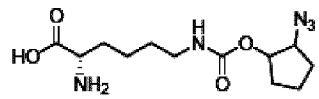
Suitably, nnAAs of the present invention are derived from (2S)-2-amino-6-hydroxyhexanoic acid as listed in the table below.

6-Hydroxy Leucy based analogs

Structure	Name
	(2S)-2-amino-6-{[(2-azidoethyl)carbamoyl]oxy}hexanoic acid
	(2S)-2-amino-6-{[(prop-2-yn-1-yl)carbamoyl]oxy}hexanoic acid
	(2S)-2-amino-6-{[(prop-2-en-1-yl)carbamoyl]oxy}hexanoic acid

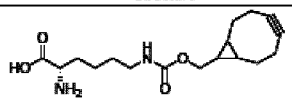
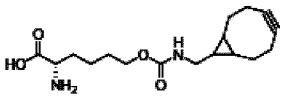
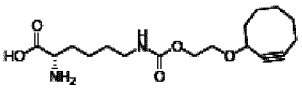
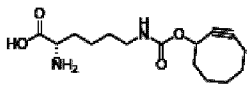
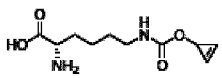
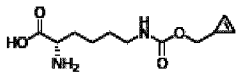
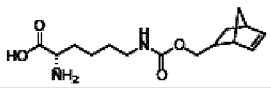
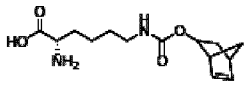
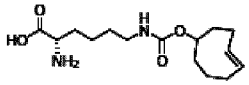
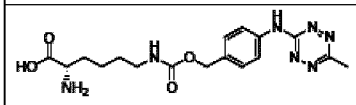
Suitably, nnAAs of the present invention are derived to include different functional groups listed in the table below. Some of which are described in part in WO20110442255A1, incorporated herein by reference.

Pyrrolysine Analogs

Structure	Name
	(2S)-2-amino-6-(pent-4-enamido)hexanoic acid
	(2S)-2-amino-6-({[(4-iodophenyl)methoxy]carbonyl}amino)hexanoic acid
	(S)-2-amino-6((2-oxo-2-phenylacetamide)hexanoic acid
	(S)-2-amino-6((2-oxo-2-propanamide)hexanoic acid
	(2S)-2-amino-6-({[(2-ethynylcyclopentyl)oxy]carbonyl}amino)hexanoic acid
	(2S)-2-amino-6-({[(2-azidocyclopentyl)oxy]carbonyl}amino)hexanoic acid

Suitably, nnAAs of the present invention are derived to include different functional groups suitable for metal free cycloaddition chemistry and are described in part in WO2012104422A1 and listed in the table below.

Pyrrolysine Analogs

Structure	Name
	(2S)-2-amino-6-[({bicyclo[6.1.0]non-4-yn-9-ylmethoxy}carbonyl)amino]hexanoic acid
	(2S)-2-amino-6-[({bicyclo[6.1.0]non-4-yn-9-ylmethyl}carbamoyl)oxy]hexanoic acid
	(2S)-2-amino-6-[[(cyclooct-2-yn-1-yloxy)carbonyl]amino}hexanoic acid
	(2S)-2-amino-6-({[2-(cyclooct-2-yn-1-yloxy)ethoxy]carbonyl}amino)hexanoic acid
	(2S)-2-amino-6-[[(cycloprop-2-en-1-yloxy)carbonyl]amino}hexanoic acid
	(2S)-2-amino-6-[[(cycloprop-2-en-1-ylmethoxy)carbonyl]amino}hexanoic acid
	(2S)-2-amino-6-[({bicyclo[2.2.1]hept-5-en-2-ylmethoxy}carbonyl)amino]hexanoic acid
	(2S)-2-amino-6-[({bicyclo[2.2.1]hept-5-en-2-yloxy}carbonyl)amino]hexanoic acid
	(2S)-2-amino-6-({[(4E)-cyclooct-4-en-1-yloxy]carbonyl}amino)hexanoic acid
	(2S)-2-amino-6-[[({4-[(6-methyl-1,2,4,5-tetrazin-3-yl)amino]phenyl}methoxy)carbonyl]amino}hexanoic acid:

Site specific Conjugation of proteins with incorporated non-natural amino acids

Proteins having incorporated non-natural amino acids using methods according to the invention may be used for the preparation of functionalized protein conjugates. Molecules that may be conjugated to proteins having incorporated non-natural amino acids include (i) other proteins, e.g. antibodies especially monoclonal antibodies and (ii) PEG groups or other groups that may cause half life extension in the system. Moreover these modified proteins can be conjugated to drugs or nucleotides for targeted delivery of these potent compounds.

More details of certain embodiments are given below in the discussion of antibody drug conjugates.

Non-natural amino acids may conveniently contain a unique chemical group permitting conjugation in a targeted fashion without risk of side reaction with other amino acids. For example non-natural amino acids conveniently contain azide or alkyne groups permitting reaction with a molecule to be conjugated which contains a corresponding alkyne or azide group using the Huisgen 1,3-dipolar cycloaddition reaction.

A further aspect of the invention is a process for preparing a chemically modified target protein which comprises preparing a target protein according to the process according to an aspect of the invention and chemically modifying the resultant target protein.

Preferred conjugation chemistries of the invention include reactions which are orthogonal to the natural twenty amino acids. Such reactions do not interact or cause side reactions with the native 20 amino acids, they are specific to the functional groups associated with the reaction. Suitably the necessary functional groups are incorporated into the target protein via the nnAA.

Further, said reactions proceed under conditions which are not destructive to the protein, for instance aqueous conditions, with a pH range which is acceptable to the protein and maintains its solubility, at a temperature which does not lead to deleterious effects upon the protein.

Increasing the stability of the attachment moiety between the protein and the linker can be advantageous. Conventional methods conjugate to the thiol groups of cysteine by reaction with a maleimide forming a thiol ether. The thiol ether can undergo the reverse reaction releasing the linker drug derivative from the antibody. In an embodiment of the invention, the conjugation chemistry employed between an azide and an alkyne results in an aromatic triazole which is significantly more stable, and not as prone to reversibility.

In addition, the product of the reaction, the linkage between protein and payload, ought to be stable, equal to or greater than the stability associated with conventional linkages (amide, thiol ether). Though not an impediment to conjugation, it is often advantageous if the conjugation reactions can be done under native conditions, as this will eliminate an extra refolding processing step.

Preferred chemical conjugations for production of conjugates of the invention include : a 3+2 alkyne-azide cycloaddition; 3+2 dipolar cycloaddition; palladium based couplings including the Heck reaction; Sonogashira reaction; Suzuki reaction; Stille coupling; Hiyama/Denmark reaction; olefin metathesis; Diels-alder reaction; carbonyl condensation with hydrazine,

hydrazide, alkoxy amine or hydroxyl amine; strain promoted cycloadditions, including Strain promoted azide alkyne cycloaddition; metal promoted azide alkyne cycloaddition; electron promoted cycloaddition; fragment extrusion cycloaddition; alkene cycloaddtion followed by a b-elimination reaction.

According to one preferred embodiment, the incorporated amino acid contains an azide or an alkyne group and the process of chemical modification comprises reacting said azide or alkyne group with a reagent comprising an alkyne or azide group. The envisaged reaction is a Huisgen 1,3-dipolar cycloaddition reaction which leads to production of a triazole linkage. The reagent comprising an alkyne or azide group may be a protein (eg an antibody) or a toxin or a cytotoxic drug or a substance suitable for half life extension (eg a PEG group) which carries an alkyne or azide group optionally via a linker.

The alkyne group of use in said reaction is, for example, a terminal alkyne, a cyclooctyne such as a bicyclo[6.1.0]non-4-yne, dibenzocyclooctyne, and difluorocyclooctyne moiety.

In a variant reaction, the incorporated amino acid contains an azide or an alkene group and the process of chemical modification comprises reacting said azide or alkene group with a reagent comprising an alkene or azide group. The reagent comprising an alkene or azide group may be a protein (eg an antibody) or a toxin or a substance suitable for half life extension (eg a PEG group) which carries an alkyne or alkene group optionally via a linker.

When more than one nnAA is incorporated into a target protein (eg an antibody), the chemical modification may be the same or different. For example if two nnAAs are incorporated, one may be modified to be conjugated to a drug moiety and one may be modified to be conjugated to a PEG moiety.

In an embodiment, conjugation chemistry of the invention is used for preparing an antibody drug conjugate.

Target proteins

Target proteins include antibodies.

Antibodies of the invention include full length antibodies and antibody fragments including Fab, Fab2, and single chain antibody fragments (scFvs) directed to TROP-2, SSTR3, B7S1/B7x, PSMA, STEAP2, PSCA, PDGF, RaSL, C35D3, EpCam, TMCC1, VEGF/R, Connexin-30, CA125 (Muc16), Semaphorin-5B, ENPP3, EPHB2, SLC45A3 (PCANAP), ABCC4 (MOAT-1), TSPAN1,

PSGRD-GPCR, GD2, EGFR (Her1), TMEFF2, CD74, CD174 (leY), Muc-1, CD340(Her2), Muc16, GPNMB, Cripto, EphA2, 5T4, Mesothelin, TAG-72, CA9 (IX), a-v-Integrin, FAP, Tim-1, NCAM/CD56, alpha folate receptor, CD44v6, Chondroitin sulfate proteoglycan, CD20, CA55.1, SLC44A4, RON, CD40, HM1.24, CS-1, Beta2 microglobulin, CD56, CD105, CD138, Lewis Y, GRNMP, Tomoregulin, CD33, FAP, CAIX, FasL Receptor, MMPmatrix metallo proteases.

In a preferred embodiment of the invention, antibodies of the invention directed to tumor targets are conjugated to protein moieties selected from the following: immunostimulatory and proapoptotic proteins, particularly Immune stimulators such as IL-1alpha, IL-1beta, other IL-1 family members, any of the interleukins, including but not limited to IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, IL-13, IL-15, IL-17 family, IL-18, IL-21, IL-22, IL-23, IL-28, or costimulatory ligands such as B7.1 and B7.2, TACI. Interferons such as any of the Type I IFN family (IFN alpha and beta and lambda) or the Type II IFN gamma.Hematopoietic growth factors such as GM-CSF. Chemokines including CXCL-1, CXCL-2, CXCL-5, CXCL-6, CXCL-8, CXCL-9, CXCL-10, and CXCL-11, CXCL-13, CCL-2, CCL-3, CCL-4, CCL-5, CCL-21, IP-10, Eotaxin, RANTES, PF4, GRO related peptides, IL-8.Proapoptotic ligands such as those of the TNF superfamily including FasL, TNF, PD-L1.Antimicrobial peptides such as alpha and beta defensins and cathelicidin LL37/hCAP18, histatins, cathepsin G, azurocidin, chymase, eosinophil derived neurotoxin, high mobility group 1 nuclear proteins, HMGB1, lactoferrin.ROS and RNS producing enzymes such as the members of NADPH oxidases (NOXs), nitric oxide synthase NOS, INOS), neutrophil granule proteins including proteases such as elastases and cathepsins, Azurocidin (also known as CAP37 or HBP), myeloperoxidase, perforin, granzymes.

In one embodiment the target protein is an anti-Her-2 antibody.

In one embodiment, the target protein is an anti-IL6 antibody.

In one embodiment, the target protein is an anti-PSMA antibody.

In a preferred embodiment, the anti-PSMA antibody is an scfv.

In a preferred embodiment, FGF21 is modified to contain non natural aminoacid lys-azide or propargyl lysine at position R131 and conjugated to a PEG moiety via a triazole linker.

PEG moieties

Target proteins may be conjugated to PEG moieties. PEG moieties may be incorporated into antibody drug conjugates. The PEG moiety may typically have a molecular weight ranging

between 5kDa and 40 kDa. More preferably, the PEG moiety may have a molecular weight of around 20 kDa. PEG moieties may be straight chain or branched.

Antibody Drug Conjugates (ADCs)

Cell lines according to the invention are particularly useful for production of Antibody Drug Conjugates (recombinant antibody covalently bound by a synthetic linker to a given drug, typically a cytotoxic drug, or else a protein or a PEG group) which are homogeneous nature, in which the number of drugs (or other conjugated molecule) per antibody and position of those drugs upon the antibody are explicitly controlled, whereby monoclonal antibodies containing incorporated non-natural amino acids are obtained and site specifically conjugated to a linker carrying a drug moiety (or other conjugated molecule) through orthogonal chemistry.

Suitably, the present invention provides a process to obtain ADCs including the following steps:

1. Introducing into a stable cell line of the invention one or more plasmids carrying the DNA sequence coding for a full length antibody, whereby a stop codon is introduced at specific positions within the sequence
2. Purify the modified antibody with non natural amino acid (nnAA) installed at desired position(s).
3. React a cytotoxin-linker derivative modified to include a functional group complimentary to the nnAA installed in the antibody with the modified antibody containing a complementary reactive group through an orthogonal chemistry
4. Purify the resulting ADC

Thus, the present invention also provides ADCs whereby the antibody component has been modified to incorporate non natural aminoacids bearing a unique reactive functional group at desired positions, whereby such functional group allows conjugation to a drug moiety (or protein or PEG group).

In an embodiment the present invention provides an antibody conjugate comprising an anti-Her-2 antibody which is conjugated to one or more moieties (e.g. one, two, three or four, preferably one or two, especially one) selected from protein, drug and PEG moieties via linkers comprising a triazole moiety.

In particular, the triazole moiety may be formed by reaction of an azide or alkyne moiety in the side chain of a non-natural amino acid incorporated into the sequence of the anti-Her-2 antibody and an alkyne or azide moiety attached to the protein, drug or PEG moiety.

In one embodiment, the triazole moiety is formed by reaction of an azide or alkyne moiety in the side chain of a non-natural amino acid incorporated into the sequence of the anti-Her-2 antibody and an alkyne or azide moiety attached to the protein, drug or PEG moiety under conditions of Cu(I) catalysis.

In another embodiment, an antibody conjugate comprises an antibody which is conjugated to one or more moieties selected from drug and PEG moieties via linkers comprising a triazole moiety in which the triazole moiety is formed by reaction of an azide moiety in the side chain of a non-natural amino acid incorporated into the sequence of the antibody and an alkyne moiety attached to the drug or PEG moiety and in which the alkyne moiety is a cyclooctyne moiety.

In another embodiment, an antibody conjugate comprises an antibody which is conjugated to one or more moieties selected from drug and PEG moieties via linkers comprising a triazole moiety in which the triazole moiety is formed by reaction of an alkyne moiety in the side chain of a non-natural amino acid incorporated into the sequence of the antibody and an azide moiety attached to the drug or PEG moiety and in which the alkyne moiety is a terminal alkyne, a substituted alkyne or a cyclooctyne moiety.

The cyclooctyne moiety may, for example, be a bicyclo[6.1.0]non-4-yne, dibenzocyclooctyne, and difluorocyclooctyne moiety.

The non-natural amino acid incorporated into the sequence of the antibody is suitably a non-natural amino acid substrate for PylRS, particularly a non natural lysine analog such as (S)-2-amino-6((2-azidoethoxy)carbonylamino)hexanoic acid.

Antibodies

In the present invention ADCs include the use of full length antibodies as well as antibody fragments such as, but not limited to Fab, Fab2, and single chain antibody fragments.

Antibodies suitable for conjugation to cytotoxins include those targeted against : anti-Her2, anti-IL-6, TROP-2, SSTR3, B7S1/B7x, PSMA, STEAP2, PSCA, PDGF, RaSL, C35D3, EpCam, TMCC1, VEGF/R, Connexin-30, CA125 (Muc16), Semaphorin-5B, ENPP3, EPHB2, SLC45A3 (PCANAP), ABCC4 (MOAT-1), TSPAN1, PSGRD-GPCR, GD2, EGFR (Her1), TMEFF2, CD74, CD174 (leY), Muc-1,

CD340(Her2), Muc16, GPNMB, Cripto, EphA2, 5T4, Mesothelin, TAG-72, CA9 (IX), a-v-Integrin, FAP, Tim-1, NCAM/CD56, alpha folate receptor, CD44v6, Chondroitin sulfate proteoglycan, CD20, CA55.1, SLC44A4, RON, CD40, HM1.24, CS-1, Beta2 microglobulin, CD56, CD105, CD138, Lewis Y, GRNMP, Tomoregulin, CD33, FAP, CAIX, FasL Receptor, MMP matrix metallo proteases.

One particular antibody of interest is an anti-Her-2 antibody.
Another particular antibody of interest is an anti-PSMA antibody, especially a scfv.

Said scfv may, for example, be conjugated to an auristatin, paclitaxel, doxorubicin, or amanitin derivative.

Examples of the invention:

Example 1:

Generation of intragenic promoter elements

Site directed mutagenesis was utilized to generate the following mutations in the *tRNApyl* gene: A at position 52 (referred to the WT Methanosarcina mazei tRNApyl sequence, SEQ ID No 3)) was mutated into C; A at position 10 was mutated into G and T at position 14 was mutated into A .

In particular, constructs containing a single mutation A52C (Construct named tRNA-B, SEQ ID No 9), all the mutations described above, namely A52C, A10G and T14A (tRNA-A2B, SEQID No 10) or mutations in the A box alone (A10G and T14A, tRNA-A2, SEQ ID No 11 or A10G, tRNA-A1, SEQ ID No 7) were made

Demonstration of functional expression of tRNApyl with intragenic promoter

To assess the function of the tRNA variants, an in vitro assay was established where cells are transiently transfected with a tRNA construct and a reporter construct encoding GFP containing an amber codon interrupting its reading frame (GFPY40). HEK293 cells stably expressing pylRS were transiently transfected with various tRNA constructs (1ug) and the GFPY40 reporter (1ug). In the absence of elements that support amber suppression, a truncated GFP protein is generated and the cell is not fluorescent. However, the introduction of a functional tRNA and the presence of a nnAA that supports amber suppression allows for the production of full length GFP resulting in fluorescent cells. The level of GFP expression reflects the efficiency of

amber suppression and is quantified by flow cytometry. tRNA genes encoding wild type tRNApyl (WT tRNA), or containing mutations in the B-box (B tRNA – SEQ ID No 9), in both the A and B box (A2B tRNA – SEQ ID No 10), or in the A box (A1 tRNA-SEQ ID No 7; or A2 tRNA SEQ ID No 11), all under control of the H1 promoter were generated by site directed mutagenesis and assayed. The H1 promoter ensured that the variants were expressed and allowed us to examine the functional effects of the mutations introduced into the tRNA genes. In addition a sample transfected with GFPY40 alone was generated to establish background GFP levels. The transfected cells were grown in presence of the aminoacid analogue lys azide (2mM) and incubated for 18 hours. GFP fluorescence was then assayed by flow cytometry and mean fluorescence intensity determined for each sample. Data were analysed with FCS Express and the mean fluorescence intensity of each sample determined using markers that constrain the data set to cells showing fluorescence above negative controls. The data, as shown in figure 4, demonstrates that the incorporation of mutations at the A-box site (H1-A1 tRNA and H1 A2 tRNA) does not affect the activity of the tRNA with respect to its capacity to support amber suppression relative to the wild type H1 WT tRNA. The constructs encoding B tRNA and A2B tRNA support amber suppression, albeit at a lower level than the WT tRNA. The constructs thus are functional and may contain the necessary genetic sequence to enable the expression of the tRNA also in the absence of an upstream promoter element such as the H1 promoter.

To begin to assess whether the tRNA mutations can support expression of the tRNA we assayed tRNA constructs encoding the wild type tRNA in the absence of the H1 promoter. Constructs encoding wild type tRNApyl (WT tRNA) or containing mutations in the B-box (B tRNA – SEQ ID No 9) or in both the A and B box (A2B; A2B tRNA – SEQ ID No 10) were assayed using our in vitro amber suppression assay described above. WT tRNA under control of the H1 promoter (H1-tRNA) was used as a positive control, and in addition one sample was transfected only with the GFPY40 reporter to illustrate background amber suppression in the absence of tRNA. The transfected cells were exposed to 2mM lys-azide nnAA and cells incubated for three days. GFP expression was examined by flow cytometry and mean fluorescence intensity determined. These data were plotted and shown in the Figure 5. The H1-tRNA construct had been shown previously to enable efficient amber suppression (MFI= 1,004,228.00). However, no GFP fluorescence was observed in a sample lacking tRNA (GFPY40; MFI= 65,929) and low levels in samples containing wt tRNA (MFI= 114,478.1). tRNA mutants B tRNA (MFI =410,131.1) and A2B

tRNA (MFI =242,747.9) enabled four-fold and 2.5 fold higher expression of GFP respectively than the wt tRNA construct. This data shows that the B and A2B mutations reconstitute a functional intragenic promoter and that the expressed tRNA is functional as it can mediate the delivery of nnAA to the target protein.

Having established that the B mutation introduced into the pyltRNA sequence enabled expression of the tRNA that supported amber suppression we then asked whether this mutation affected the orthogonality of the tRNA and led to the incorporation of natural amino acids. To do this 3H7 cells were transiently transfected with expression constructs encoding WT-tRNA, B-tRNA, H1-tRNA and H1-B tRNA along with the GFPY40 reporter. Transfected cells were grown in the presence of 2mM lys azide or absence of nnAA as indicated (Figure 6), and GFP expression assessed by flow cytometry after three days. In addition, a sample transfected only with GFPY40 was examined to determine background amber suppression levels in the absence of tRNA. Figure 3 shows that in the presence of lys-azide H1-tRNA (MFI= 347,958), H1-B tRNA (MFI=313,386) and B-tRNA (MFI=232,505) all support robust amber suppression and expression of full length GFP above background levels (MFI=23,487). In the absence of nnAA background levels of GFP expression were observed in cells containing these three constructs (MFI=33,452; 36,217; 42,104 respectively). These data show that B tRNA is orthogonal and does not promote amber suppression levels above background in the absence of nnAA including a construct previously shown to support high levels of expression (H1-tRNA B). A comparison of GFP expression between the WT tRNA (MFI 33,781) and tRNA B (232,505) lacking extragenic promoter elements confirms that the B mutation generates a functional intragenic promoter. Indeed, the B tRNA construct enabled a seven-fold increase in the level of amber suppression dependent expression of GFP over WT tRNA.

Example 2

Copy number and amber suppression

In an effort to increase the efficacy of amber suppression by the tRNA-B construct, we assessed whether increased copy numbers led to an increased activity.

Two copies of the WT, B, or A2B tRNA were cloned in tandem to generate a single vector containing two copies of WT, B, or A2B tRNA. To assess the effect of increased gene doses by these constructs, cells expressing WTpylRS were transiently transfected with the tRNA

constructs and the GFPY40 reporter construct and GFP expression assayed after two days. The data in figure 7 show that B (MFI=67,489) and A2B tRNA (MFI=61,950) can support amber suppression levels above those by WT tRNA (MFI=18,002) confirming the observation that the B and A2B tRNAs contain functional promoter elements. Furthermore, we observed that tandem copies of B (MFI=110,941) and A2B tRNA (MFI=96,158) led to a concomitant increase in the efficacy of amber suppression as determined by GFP expression. Gene dose did not improve WT tRNA amber readthrough (MFI=20,617).

Example 3

Expression of tRNApyl under extragenic promoters and hybrid (Type 4) promoters

The H1 promoter has been shown here and by others to support expression of *tRNApyl*. However, significant deleterious effects have been observed by the use of polIII/Type3 (e.g. H1 and U6) promoters in mammalian cells. Thus, to overcome these deficiencies, a set of extragenic promoters elements, known to govern RNA expression, were assessed for their capacity to express *tRNApyl*. Extragenic promoter elements are known to regulate the expression of human Vault, EBER and 7SL RNA (Englert et al., 2004; Kickhoefer et al., 2003; Howe & Shu 1989). Each of these promoters has been shown to recruit the transcription factor TFIIIc, and not SNAPc (used by U6 and H1) (Moqtaderi et la., 2010) (Felton-Edkins et al (2006) J Biol Chem 281:33871) . Interestingly, efficient transcription by these promoters requires both intragenic and extragenic regulatory sequences (Englert et al., (2004); Kickhoefer et al., (2003); and Howe & Shu (1989)). The identification of a functional tRNApyl containing consensus A and B-boxes opened the door to their use in the expression of the tRNApyl. The promoter elements know to regulate Vault, 7SL, and EBER2 RNA were constructed from previously described sequences (Englert et al., 2004; Kickhoefer et al., 2003; Howe & Shu 1989). We tested the promoters for human Vault and7SL RNA genes and Epstein Barr virus EBER geneas extragenic promoter elements by placing them 5' of genes encoding WT tRNA, B tRNA, A2B tRNA, and additional variants containing A10G and A52C (AB tRNA; SEQ ID No. 8) or containing a single mutation at A10G (A1 tRNA; SEQ ID No. 7).

The activity of each construct was assessed by transient transfection into cells stably expressing WT pylRS with the GFPY40 reporter as described above. For this experiment, 1 ug of a PCR product encoding each of the indicated constructs was used. In each case, cells were exposed to the lys azide nnAA for three days and GFP levels assayed by flow cytometry. Mean

fluorescence intensities were determined for each sample and plotted (Figure 8). A sample transfected only with GFPY40 (no tRNA) was used to establish background amber suppression in the absence of tRNA (no tRNA, MFI= 137,589). In addition, a sample transfected with H1-WT tRNA was to define high expression levels of GFP and function of the tRNA (H1 WT tRNA, MFI= 828,503) and low levels of amber suppression determined by tRNA lacking intragenic promoters (WT tRNA, MFI=117,626). As shown in figure 7, the use of extragenic promoters (EBER, Vault and 7SL) showed improved expression levels of the GFP reporter over WT tRNA with EBER, Vault and 7SL promoters. However, improved performance from each of these promoters required intragenic tRNA elements. Indeed, cells transfected with 7SL WT tRNA showed minimal amber suppression (7SL WT tRNA, MFI= 106,075). However, relative amber suppression efficacy improved with tRNA containing conserved A and B-box elements (7SL tRNA A1, MFI= 114,393; 7SL AB, MFI=210,380; 7SL A2B, MFI= 245,120). Interestingly, the highest amber suppression levels was observed with the tRNA constructs containing only a functional B-box (7SL B tRNA, MFI=265,396). This general trend was also observed with the Vault RNA promoter. Constructs containing a functional B box (Vault B tRNA, MFI=170,838) showed improved performance over those containing both consensus A and B boxes (Vault A2B tRNA, MFI=154,445) and constructs of the wild type tRNA (Vault WT tRNA, MFI=102,415). The EBER promoter functioned best when the tRNA contained conserved A and B-box elements (EBER A2B tRNA, MFI=136,196). This construct outperformed tRNA containing a B mutation (EBER B tRNA, MFI=113,496) and tRNA containing a single B-box and a partially reconstituted A box (EBER AB tRNA, MFI=119,719), and unmodified tRNA (EBER WT tRNA, MFI=99,393). Our data show that the extragenic RNA promoters for Vault snRNA, 7SL RNA and EBER RNA can support expression and function of tRNApyl. The efficacy of amber suppression is dependent on the presence of a functional B-box.

We then tested amber suppression activity to directly compare the levels of tRNA dependent amber suppression between the various functional tRNA expression constructs (Figure 9). In this assay, cells stably expressing WTpylRS were transiently transfected with 1 ug of GFPY40 reporter and 1ug of one of the tRNA expression plasmids (H1-tRNA, Vault B tRNA, 7SL B tRNA, EBER B tRNA, B tRNA or WT tRNA). Figure 9 shows that both 7SL-B tRNA (MFI=167,024) and Vault-B tRNA (MFI=168,298) can support amber suppression activity comparable to that seen with the H1-tRNA (MFI= 241,901) construct, resulting in an improvement of activity compared

to the tRNA-B construct (MFI= 89,869) and WT tRNA (MFI= 25,459), relative to H1-tRNA (Figure 8).

In an effort to increase the amber suppression levels that are supported by the extragenic 7SL promoter with B tRNA we examined the effects of tandem repeats of 7SL-tRNA-B relative to H1 tRNA. To do this, two copies of the 7SL- tRNA-B constructs were cloned into a single vector and amber suppression levels assessed using transient transfections of cells stably expressing the WTpylRS along with GFPY40. Cells were incubated with lys azide for three days and GFP levels quantified by flow cytometry. Figure 10 shows that increased gene copies of 7SL- B tRNA as well as H1-tRNA results in improved amber suppression efficacy and GFP expression levels. In this experiment cell transfected with GFP only were used as a negative control (MFI=0). Cells transfected with a tandem copy of the H1-tRNA (2x H1 tRNA, MFI=629) resulted in a 28% increase in expression of the GFPY40 reporter construct over a single copy of the H1 tRNA (MFI=453). The same effect was observed with tandem repeats of tRNA lacking an external promoter but containing a functional B box. Here tandem repeats of tRNA B (2x tRNA B, MFI 366) showed a 10% improvement over the single gene copy (tRNA B, MFI=329). Tandem copies of the 7SL tRNA B (2x 7SL tRNA B, MFI=568), also showed a 8.5% improvement over the signal observed with a single gene (7SL tRNA B, MFI=520). A construct encoding wild type tRNA lacking a functional B box under control of the 7SL promoter has been previously shown to be less effective than 7SL tRNA B. This former construct was included in this analysis to verify this effect (7SL tRNA WT, MFI=312). Indeed this data confirms previous experiments that the B mutation improves the performance of the 7SL-tRNA construct. These data show that increased gene numbers can improve the efficacy of the suppressor tRNA. Furthermore these data also show that the 7SL promoter is equivalent to the H1 promoter in facilitating the expression of tRNApyl.

Example 4

Expression of tRNApyl as part of a bicistronic message

We also tested the utility of a bicistronic tRNA expression construct for its ability to express tRNApyl. Here, the murine tRNA Glu and tRNA Asp were utilized. In each case, multiple repeats of the constructs were generated to increase activity (3xGlu-pyl and 2x Asp-pyl). Here, 3H7 cells were transiently transfected with 1ug of the reporter construct GFPY40 and 1ug of the indicated tRNA expression vectors (Figure 11) and incubated with lys-azide for three days.

Data were collected by flow cytometry and mean fluorescence intensities for each sample were plotted. We observed that the bicistronic constructs enabled high levels of amber suppression relative to the H1-tRNA positive control. tRNA expression cassettes consisting of three tandem repeats of the murine tRNA glu and pyltRNA pair (3xGlu, MFI=419) and two tandem repeats of the murine tRNA asp-pyl tRNA pair (2x Asp, MFI=419) showed >90% of the signal generated by H1-tRNA (MFI=453) Indicating their ability to efficiently produce functional tRNA. In addition, 7SL tRNA B (MFI=520) also demonstrated efficient expression of the tRNA. The construct consisting of the 7SL WT tRNA (MFI=312) was included as a metric for low efficiency. In addition cell lines containing only the GFPY40 reporter were used as negative controls (MFI=0). This data indicates that the bicistronic construct is an efficient method for the expression of tRNAs lacking functional A and B-box elements and furthermore demonstrates that mammalian cells can process these RNA molecules to generate mature and functional tRNAs.

References:

1. Blight et aL. 2004 Nature. 431 333-335.
2. Buvoli et al. 2000 Mol Cell Biol. 3116–3124.
3. Chen, P., 2009 Agnew Chem Int Ed Engl. 48, 4052-55.
4. Ehlert, et al., 2010 BMC Neuroscience, 11(20), p. 1471
5. Englert et al., 2004 Biochimie 86 (2004) 867–874
6. Felton-Edkins et al (2006) J Biol Chem 281:33871
7. Francis MA et al., 1990 Mol Cell Biol 10(9):4486-94.
8. Giering, et al., 2008 Mol Ther. 16(9):1630-6
9. Hancock et AL. JACS 2010, 132, 14819-24
10. Hecht et al., 1978 JBC 253, 4517-20.
11. Herold et al., 2008 PNAS 105, 18507-12.
12. Howe JG et al., 1989 Cell. 57(5):825-34.
13. Hudziak RM, et al., 1982 Cell 31(1):137-46.
14. Kavran et al., 2007 PNAS 104, 11268-73.
15. Kohrer et al., 2001 PNAS 98, 14310-15;

16. Kohrer et al., Chem & Biol., (2003) 10, 1095-1102;

17. Laski FA et al., 1982 Proc Natl Acad Sci U S A. ;79(19):5813-7.

18. Liebman SW. and Sherman, F. 1976 Genetics, 82, 233-249.

19. Liebman, SW et al., 1976 Genetics 82, 251-272.

20. Liu W. et al. 2007, Nature methods, 4; 239-244.

21. Mossink et al., 2002 294(1-2):225-32

22. Mukai et al 2008 BBRC 371, 818-823

23. Naykova et al. 2003 J Mol. Evol. 57:520-532.

24. Neumann et al. 2008 Nat. Chem. Biol. 4, 232-234.

25. Nguyen et al., 2009 *J. Am. Chem. Soc. 131* (25), pp 8720–8721

26. Nozawa 2009, Nature. 457 1163-67.

27. Paul et al., 2003 Mol. Ther., 7(2), pp. 237-247

28. Pettit et al., 1997 Fortschr. Chem. Org. Naturst 70, 1-79

29. Sakamoto, K. 2002 Nucl. Acid Res. 30, 4692-4699.

30. Schmidt O et al., 1980 Nature 287(5784):750-2.

31. Shan L. et al., J Gene Med.(2006)8, 1400-1406.

32. Senter P. et al., 2003 Blood 102, 1458-65.

33. Stegmeier, et al., 2005 PNAS, 102(37), p. 13212–13217.

34. Straby KB, 1988 Nucleic Acids Res. 16(7):2841-57.

35. Takimoto j. 2009, Mol. Biosystems, 5, 931-34.

36. Wang w. Nature Neuro. 2007, 8; 1063-1072.

37. Ye, S. 2008, JBC 283, 1525-1533.

38. Yanagisawa 2008 Chem & Biol. 15, 1187-1197.

39. Wang et Al, 2011 Aijun Wang, Natalie Winblade Nairn, Marcello Marelli and KennethGrabstein (2012). Protein Engineering with Non-Natural Amino Acids, Protein Engineering, Prof. Pravin Kaumaya (Ed.), ISBN: 978-953-51-0037-9, InTech, Available from:

http://www.intechopen.com/books/protein-engineering/protein-engineering-with-nonnatural-amino-acids

40. Dieci, G., Conti, A., Pagano, A. & Carnevali, D., 2013. Idnetification of RNA Polymerase III-transcribed genes in eukaryotic genomes. *Biochim. Biophys. Acta,* Volume 1829, pp. 296-305..

41. Ehlert, E., Eggers, R., Niclou, S. & Verhaagen, J., 2010. Cellular toxicity following application of adenoassociated viral vector-mediated RNA interference in the nervous system. *BMC Neuroscience,* 11(20), p. 1471.

42. Giering, J., Grimm, D., Storm, T. & Kay, M., 2008. Expression of shRNA from a tissue-specific pol II promoter is an effective and safe RNAi therapeutic.. *Mol.Ther.,* Volume 144, pp. 1630-6.

43. Hancock, S. M., Uprety, R., Deiters, A. & Chin, J. W., 2010. Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair. *JACS,* Volume 132, pp. 14819-24.

44. Paul, C. et al., 2003. Localized Expression of Small RNA Inhibitors. *Mol. Ther.,* 7(2), pp. 237-247.

45. Stegmeier, F. et al., 2005. A lentiviral microRNA-based system for single-copy polymerase II-regulated RNA interference in mammalian cells. *PNAS,* 102(37), p. 13212–13217.

46. Wang, Q. & Wang, L., 2012. Genetic incorporation of unnatural amino acids into proteins in yeast. *Methods Mol. Biol.,* Volume 794, pp. 199-213.

47. Englert, M., Martha, F., Junker, V. & Beier, H., 2004. Novel upstream intragenic control elements for the RNA polymerase III-dependent transcription of human 7SL RNA genes. Biochimie, Volume 86, pp. 867-874.

48. Howe, J. & Shu, M.-D., 1989. Epstein Barr virus small RNA (EBER) Genes: unique transcription units that combine RNA polymerase II and III promoter elements. Cell, Volume 57, pp. 825-834.

49. Kickhoefer, V. et al., 2003. Identification of conserved vault RNA expression elements and a non expressed mouse vault RNA gene. Gene, Volume 309, pp. 65-70.

50. Moqtaderi, Z. et al., 2010. Genomic binding profiles of fucntionally distinct RNA polymerase III transcription complexes in human cells. Nature Struct. & Mol. Biol., 17(5), pp. 635-641.

Sequence listing:

Underlined sequences identify A box and/or B box sequences

Bold characters identify mutated nucleotides

SEQ ID No 1

>tRNApyl Methanosarcina_barkeri, WT; coding sequence

GGAAACCTGATCATGTAGATCGTGGACTTCTAAATCCGCAGCCGGGTAGATTCCCGGGGTTTCCG

SEQ ID No 2

>tRNApyl Methanosarcina acetivorans, WT; coding sequence

GGAAACCTGATCATGTAGATCGAATGGACTCTAAATCCGTTCAGCCGGGTTAGATTCCCGGGGTTTCCG

SEQ ID No 3

>tRNApyl Methanosarcina mazei Go1, WT; coding sequence

GGAAACCTGATCATGTAGATCGAATGGACTCTAAATCCGTTCAGCCGGGTTAGATTCCCGGGGTTTCCG

SEQ ID No 4

>tRNApyl Methanococcoides burtonii DSM, WT; coding sequence

GGAGACTTGATCATGTAGATCGAACGGACTCTAAATCCTTTCAGCCGGGTTAGATTCCCGGAGTTTCCG

SEQ ID No 5

>tRNApyl Desulfobacterium hafniense, WT; coding sequence

GGAAACCTGATCATGTAGATCGTGGACTCTAAATCCGCAGCCGGGTAGATTCCCGGGGTTTCCG

SEQ ID No 6

>Terminator sequence used for WT and mutant A –box and/or B-box tRNApyl constructs

gacaagtgcggttttttt

SEQ ID No 7

> tRNApyl Methanosarcina mazei A10G mutation ("*tRNA-A1*"); coding sequence

ggaaacctg**g**tcatgtagatcgaatggactctaaatccgttcagccgggttagattcccggggtttccg

SEQ ID No 8

tRNApyl Methanosarcina mazei A10G A52C mutation ("*tRNA-AB*"); coding sequence

ggaaacctg**g**tcatgtagatcgaatggactctaaatccgttcagccgggtt**c**gattcccggggtttccg

SEQ ID No 9

tRNApyl Methanosarcina mazei A52C mutation ("*tRNA-B*"); coding sequence

ggaaacctgatcatgtagatcgaatggactctaaatccgttcagccgggtt**c**gattcccggggtttccg

SEQ ID No 10

tRNApyl Methanosarcina mazei A10G, T14A, A52C mutations ("*tRNA-A2B*"); coding sequence

ggaaacctg**g**tca**a**gtagatcgaatggactctaaatccgttcagccgggtt**c**gattcccggggtttccg

SEQ ID No 11

tRNApyl Methanosarcina mazei A52C, T14A, mutations ("*tRNA-A2*"); coding sequence

ggaaacctg**g**tca**a**gtagatcgaatggactctaaatccgttcagccgggttcgattcccggggtttccg

SEQ ID No 12

7SL Promoter 5'region sequence

cccagttgatgacgtcaccataccacagcttctagtgctattctgcgccggtatccgacc

SEQ ID No 13

Spacer used in 7SL constructs

acca

SEQ ID No 14

7SL tRNA-WT construct

cccagttgatgacgtcaccataccacagcttctagtgctattctgcgccggtatccgaccaccaggaaac ctgatcatgtagatcgaatggactctaaatccgttcagccgggttagattcccggggtttccggacaagt gcggttttt

SEQ ID No 15

7SL tRNA-B construct

cccagttgatgacgtcaccataccacagcttctagtgctattctgcgccggtatccgaccaccaggaaac
ctgatcatgtagatcgaatggactctaaatccgttcagccgggttcgattcccggggtttccggacaagt
gcggttttt

SEQ ID No 16

7SL tRNA-A1 construct

cccagttgatgacgtcaccataccacagcttctagtgctattctgcgccggtatccgaccaccaggaaac
ctggtcatgtagatcgaatggactctaaatccgttcagccgggttagattcccggggtttccggacaagt
gcggttttt

SEQ ID No 17

7SL tRNA-AB construct

cccagttgatgacgtcaccataccacagcttctagtgctattctgcgccggtatccgaccaccaggaaac
ctggtcatgtagatcgaatggactctaaatccgttcagccgggttcgattcccggggtttccggacaagt
gcggttttt

SEQ ID No 18

7SL tRNA-A2B construct

Cccagttgatgacgtcaccataccacagcttctagtgctattctgcgccggtatccgaccaccaggaaac
ctggtcaagtagatcgaatggactctaaatccgttcagccgggttcgattcccggggtttccggacaagt
gcggttttt

SEQ ID No 19

Terminator sequence used in the 7SL, Vault, EBER constructs

GACAAGTGCGGTTTTT

SEQ ID No 20

Vault Promoter 5'region sequence

tggtcgctgagtcatgctgaagcactctttttgactcactctttgtgacgtaggtctttctcaccagt
caataaaatataatccgaaagcaacctc

SEQ ID No 21

Spacer used for Vault-tRNA constructs

CCACCA

SEQ ID No 22

Vault tRNA-WT construct

tggtcgctgagtcatgctgaagcactctcttttttgactcactctttgtgacgtaggtctttctcaccagt
caataaaatataatccgaaagcaacctcccaccaggaaacctgatcatgtagatcgaatggactctaaat
ccgttcagccgggttagattcccggggtttccggacaagtgcggttttt

SEQ ID No 23

Vault tRNA-B construct

tggtcgctgagtcatgctgaagcactctcttttttgactcactctttgtgacgtaggtctttctcaccagt
caataaaatataatccgaaagcaacctcccaccaggaaacctgatcatgtagatcgaatggactctaaat
ccgttcagccgggtt**c**gattcccggggtttccggacaagtgcggttttt

SEQ ID No 24

Vault tRNA-AB construct

tggtcgctgagtcatgctgaagcactctcttttttgactcactctttgtgacgtaggtctttctcaccagt
caataaaatataatccgaaagcaacctcccaccaggaaacctg**g**tcatgtagatcgaatggactctaaat
ccgttcagccgggtt**c**gattcccggggtttccggacaagtgcggttttt

SEQ ID No 25

Vault tRNA-A2B construct

Tggtcgctgagtcatgctgaagcactctcttttttgactcactctttgtgacgtaggtctttctcaccagt
caataaaatataatccgaaagcaacctcccaccaggaaacctg**g**tca**a**gtagatcgaatggactctaaat
ccgttcagccgggtt**c**gattcccggggtttccggacaagtgcggttttt

SEQ ID No 26

EBER2 Promoter 5'region sequence

agatgcacgcttaaccccgcctacaaccgtgacgtagctgtttaccagcatgtatagagttacggttcgc
tacatcaaac

SEQ ID No 27

Spacer used in EBER constructs

acca

SEQ ID No 28

EBER2 tRNA-WT construct

agatgcacgcttaaccccgcctacaaccgtgacgtagctgtttaccagcatgtatagagttacggttcgc
tacatcaaacaccaggaaacctgatcatgtagatcgaatggactctaaatccgttcagccgggttagatt
cccggggtttccgggacaagtgcggttttt

SEQ ID No 29

EBER2 tRNA-B construct

agatgcacgcttaaccccgcctacaaccgtgacgtagctgtttaccagcatgtatagagttacggttcgc
tacatcaaacaccaggaaacctgatcatgtagatcgaatggactctaaatccgttcagccgggtt**c**gatt
cccggggtttccgggacaagtgcggttttt

SEQ ID No 30

EBER2 tRNA-A1 construct

agatgcacgcttaaccccgcctacaaccgtgacgtagctgtttaccagcatgtatagagttacggttcgc
tacatcaaacaccaggaaacctg**g**tcatgtagatcgaatggactctaaatccgttcagccgggttagatt
cccggggtttccgggacaagtgcggttttt

SEQ ID No 31

EBER2 tRNA-AB construct

agatgcacgcttaaccccgcctacaaccgtgacgtagctgtttaccagcatgtatagagttacggttcgc
tacatcaaacaccaggaaacctg**g**tcatgtagatcgaatggactctaaatccgttcagccgggtt**c**gatt
cccggggtttccgggacaagtgcggttttt

SEQ ID No 32

EBER2 tRNA-A2B construct

agatgcacgcttaaccccgcctacaaccgtgacgtagctgtttaccagcatgtatagagttacggttcgc
tacatcaaacaccaggaaacctg**g**tca**a**gtagatcgaatggactctaaatccgttcagccgggtt**c**gatt
cccggggtttccgggacaagtgcggttttt

SEQ ID No 33

H1 promoter sequence

Aatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagact
tataagttccctatcagtgatagagacacca

SEQ ID No 34

H1 tRNA-WT construct

aatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagact
tataagttccctatcagtgatagagacaccaggaaacctgatcatgtagatcgaatggactctaaatccg
ttcagccgggttagattccccggggtttccggacaagtgcggttttttt

SEQ ID No 35

H1 tRNA-B construct

Aatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagact
tataagttccctatcagtgatagagacaccaggaaacctgatcatgtagatcgaatggactctaaatccg
ttcagccgggtcagattcccggggtttccggacaagtgcggttttttt

SEQ ID No 36

H1 tRNA-A1 construct

aatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagact
tataagttccctatcagtgatagagacaccaggaaacctggtcatgtagatcgaatggactctaaatccg
ttcagccgggttagattcccggggtttccggacaagtgcggttttttt

SEQ ID No 37

H1 tRNA-A2 construct

aatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagact
tataagttccctatcagtgatagagacaccaggaaacctggtcaagtagatcgaatggactctaaatccg
ttcagccgggttagattcccggggtttccggacaagtgcggttttttt

SEQ ID No 38

H1 tRNA-A2B construct

aatatttgcatgtcgctatgtgttctgggaaatcaccataaacgtgaaatccctatcagtgatagagact
tataagttccctatcagtgatagagacaccaggaaacctggtcaagtagatcgaatggactctaaatccg
ttcagccgggtcagattcccggggtttccggacaagtgcggttttttt

SEQ ID No 39

Terminator sequence used in the H1-tRNA constructs:

gacaagtgcggttttttt

SEQ ID No 40

M.Musculus tRNAglu coding sequence

tccctggtggtctagtggttaggattcggcgctctcaccgccgcggcccgggttcgattctcggtcaggg
aa

SEQ ID No 41

Mus musculus tRNA Glu with 5’ and 3’ non coding regions:

Gcctgtagtagtccgccggagcggaggttccaggcgggcaagcagtgtatctgccacggtggcggagagg
ctgtgtgtctgagtgtctccaggagtcggtagccgtgacctcgggccgacgctccctggtggtctagtgg

ttaggattcggcgctctcaccgccgcggcccgggttcgattctcggtcagggaagcctctcttcttcctc
tctgcaccctcacttccccaaacctcgctcctcttccctcgctcacttttgcccaacaccaacgcacacc

SEQ ID No 42

1xtRNAglu-pyl DNA construct

gccgtgacctcgggccgacgctccctggtggtctagtggttaggattcggcgctctcaccgccgcggccc
gggttcgattctcggtcagggaagcctctcccccaaacctcggaaacctgatcatgtagatcgaatggac
tctaaatccgttcagccgggttagattcccggggtttccgcttcttcctcttttct

SEQ ID No 43

3x tRNAglu-pyl: DNA construct including 3 repeats of the bicistronic construct tRNAglu-tRNApyl

gccgtgacctcgggccgacgctccctggtggtctagtggttaggattcggcgctctcaccgccgcggccc
gggttcgattctcggtcagggaagcctctcccccaaacctcggaaacctgatcatgtagatcgaatggac
tctaaatccgttcagccgggttagattcccggggtttccgcttcttcctcttttctgcaccctcacttca
gctgagccgtgacctcgggccgacgctccctggtggtctagtggttaggattcggcgctctcaccgccgc
ggcccgggttcgattctcggtcagggaagcctctcccccaaacctcggaaacctgatcatgtagatcgaa
tggactctaaatccgttcagccgggttagattcccggggtttccgcttcttcctcttttctgcaccccaa
tattgccgtgacctcgggccgacgctccctggtggtctagtggttaggattcggcgctctcaccgccgcg
gcccgggttcgattctcggtcagggaagcctctcccccaaacctcggaaacctgatcatgtagatcgaat
ggactctaaatccgttcagccgggttagattcccggggtttccgtagcgtccttttggaaagacaagtg
ggtggcgaattcgatatc

SEQ ID No 44

5'Leader sequence for tRNAglu-pyl constructs

gccgtgacctcgggccgacgc

SEQ ID No 45

Terminator sequence used in the tRNAglu-pyl constructs

tagcgtccttttg

SEQ ID No 46

Terminator sequence used in the tRNAglu-pyl constructs

Cttcttcctcttttct

SEQ ID No 47

Intergene sequence used in the tRNAglu-pyl constructs

gcctctcccccaaacctc

SEQ ID No 48

M.Musculus tRNAasp coding sequence

Tcctcgttagtatagtggtgagtatccccgcctgtcacccgcgagaccggcgttccattccccgacgggg
ag

SEQ ID No 49

M. musculus tRNA Asp with upstream and downstream regions

tgggtgagacggaggttgtgggtcgtgtgtgtcgtcgtagacggtcgggcgacggtgcgtcgtagtcggc
gttgtcctcgttagtatagtggtgagtatccccgcctgtcacccgcgagaccggcgttccattccccgac
ggggagacgtagcgtccttttggaaagacaagtgggtggcccgcggg

SEQ ID No 50

1x tRNAasp-pyl DNA construct

ggtgcgtcgtagtcggcgttgtcctcgttagtatagtggtgagtatccccgcctgtcacccgcgagaccg
gcgttccattccccgacggggagacgtagccccacctcggaaacctgatcatgtagatcgaatggactct
aaatccgttcagccgggttagattcccggggtttccg

SEQ ID No 51

2x tRNAasp-pyl: DNA construct including 2 repeats of the bicistronic construct tRNAasp-pyl

tcctcgttagtatagtggtgagtatccccgcctgtcacccgcgagaccggcgttccattccccgacgggg
agacgtagccccacctcggaaacctgatcatgtagatcgaatggactctaaatccgttcagccgggttag
attcccggggtttccgtagcgtccttttggaaagacaaggtgcgtcgtagtcggcgttgtcctcgttag
tatagtggtgagtatccccgcctgtcacccgcgagaccggcgttccattccccgacggggagacgtagcc
ccacctcggaaacctgatcatgtagatcgaatggactctaaatccgttcagccgggttagattcccgggg
tttccgtagcgtccttttt

SEQ ID No 52

Leader sequence used in the 2xAsp-pyl construct

Ggtgcgtcgtagtcggcgttg

SEQ ID No 53

Terminator sequence used in the 2xAsp-pyl construct

tagcgtccttttt

SEQ ID No 54

intergene sequence used in the tRNAasp-pyl constructs

acgtagccccacctc

SEQ ID No 55

human tRNAglu coding sequence

TCCCTGGTGGTCTAGTGGCTAGGATTCGGCGCTTTCACCGCCGCGGCCCGGGTTCGATTCCCGGTCAGGA
AT

SEQ ID No 56

human tRNAasp coding sequence

TCCTTGTTAGTATAGTGGTGAGTGTTTCTGCCTGTCATGTGGAGACTGGAGTTTGAGTCCCCAACAGGGA

SEQ ID No 57

7SL promoter variant: 7SL-1

CGCTCCCCAATGACGTAACTGCCCTGC-AGCTTCTAGTAGCT-TTTCGCAGCGTCTCCGACCG

SEQ ID No 58

7SL promoter variant: 7SL-2

TCTGGTTGCTACCATGTGTAGCC-TGCAAGCCTCTAGCAGCTCTTTTGCAGCGACGCCGACCG

SEQ ID No 59

7SL promoter variant: SL28

TAACATTGCTAATGAGTTTTGGGCACACATTGTCATTGATAACATCTTATCAGGAGACAGGG

SEQ ID No 60

7SL promoter variant: 7L63

ACAGAGGAACTACCATAACAAGAACCAAAGAGAAATGGCAtACcTCAGG

SEQ ID No 61

7SL promoter variant: 7L23

CTCCCTCTCAAAACTCAAGGAAATCTAGTGGGCATGGTGCACAT

SEQ ID No 62

7SL promoter variant: 7L7

AGGTACTACCCAATGGTTTTCTAAAATACTTGCATC-TGCACATGCCATG

SEQ ID No 63

EBER promoter variant: EBER1

GATCCAAACTTTTGTTTTAG GATTTATGCA TCCATTATCC CGCAGTTCCA CCTAAACGGG
GCTTAACGTT GCATCCCAGA AGATGCACGC TTAACCCCGC CTACAACCGT GACGTAGCTG
TTTACCAGCA TGTATAGAGT TACGGTTCGC TACATCAAAC

SEQ ID No 64

Vault promoter variant: Hvg3

aggtcgctgagtcaagctaaagcaccgttgttttgactcactctttgtgacgtaggtctttctcaccagt
caagaaaatacaatccgaaagcaacctctg

SEQ ID No 65

Vault promoter variant: Hvg2

Tggttgctgagtcatgctaaaacactgttgttttgactcactttgtgacgtaggtctttctcaccagtca
agaaaatataatccggagatagcctctg

SEQ ID No 66

Methanosarcina mazei pylRS aa sequence

Mdkkplntlisatglwmsrtgtihkikhhevsrskiyiemacgdhlvvnnsrssrtaralrhhkyrktck
rcrvsdedlnkfltkanedqtsvkvkvvsaptrtkkampksvarapkplenteaaqaqpsgskfspaipv
stqesvsvpasvstsissistgatasalvkgntnpitsmsapvqasapaltksqtdrlevllnpkdeisl
nsgkpfrelesellsrrkkdlqqiyaeerenylgklereitrffvdrgfleikspilipleyiermgidn
dtelskqifrvdknfclrpmlapnlynylrkldralpdpikifeigpcyrkesdgkehleeftmlnfcqm
gsgctrenlesiitdflnhlgidfkivgdscmvygdtldvmhgdlelssavvgpipldrewgidkpwiga
gfglerllkvkhdfknikraarsesyyngistnl

SEQ ID No 67

Methanosarcina mazei pylRS nucleotide sequence

atggataaaaaaccattgaatacgctcattagcgcaactgggctgtggatgagccgtacgggaacgattc
ataaaatcaagcaccacgaagtatctcgtagcaaaatctatatcgagatggcttgcggcgaccatctcgt
ggtaaacaatagcaggtcctcacggaccgcccgtgccttgcgccaccacaaatatcgtaaaacttgtaag

agatgtagagtgagcgacgaggatctgaacaagtttcttacaaaggccaacgaggaccaaaccagcgtca
aagtcaaggttgtgagcgccccaacacgcaccaagaaggccatgcccaagtctgttgcgcgggcaccgaa
acctctggagaatactgaggccgctcaggcccagcccagcggttcaaaattctctcctgccattccagtt
agcactcaagagtcagtcagcgtgcccgcctctgtgtctacatccatcagctctatctccaccggcgcaa
cagcctctgccctggtgaagggtaatacgaaccctatcacgagtatgtccgcacccgtgcaagcaagtgc
tcccgcactcactaaatcccaaacggaccggctggaggtcctgcttaaccctaaggatgaaatcagcctg
aacagtggaaaaccgtttcgagaactggaatccgagctcttaagccggcgaaagaaagatttgcaacaga
tttacgccgaagaacgggaaaattatctgggcaagctggagagagaaatcactaggttctttgtagatag
gggctttctggagattaagagtcccatattgatccctctcgaatacattgagcgtatgggcatcgacaac
gacacagaacttagcaagcagatctttcgggtggacaaaaacttctgcctcaggcctatgctggctccaa
atctgtacaactatcttaggaaactcgaccgggccctgcccgatcccattaaaatcttcgaaattggacc
ttgctatagaaaggagagcgatggcaaggagcacctggaggagtttactatgctcaatttctgtcaaatg
ggctccggctgcacacgtgagaacctcgaatccattataaccgacttcctgaatcacctggggattgatt
tcaagatcgtgggcgactcctgcatggtgtttggtgatacgttggatgtgatgcacggagatttggaatt
gtcaagcgctgtggtaggccccattcctctcgacagggagtggggtattgacaagccctggatcggcgca
ggttttggactggagcgcctgttgaaggttaagcatgacttcaaaaacataaagagagccgcacgcagcg
aatcctattataatggaatcagcactaacttgtaa

Throughout the specification and the claims which follow, unless the context requires otherwise, the word 'comprise', and variations such as 'comprises' and 'comprising', will be understood to imply the inclusion of a stated integer, step, group of integers or group of steps but not to the exclusion of any other integer, step, group of integers or group of steps.

The application of which this description and claims forms part may be used as a basis for priority in respect of any subsequent application. The claims of such subsequent application may be directed to any feature or combination of features described herein. They may take the form of product, composition, process, or use claims and may include, by way of example and without limitation, the following claims:

Claims

1. A DNA construct which comprises a *tRNApyl* gene comprising a *tRNApyl* coding sequence and a functional intragenic RNA polymerase III promoter sequence which is capable of acting to express functional tRNApyl sufficiently to support amber suppression in a eukaryotic expression system; wherein the tRNApyl gene comprises a nucleic acid sequence selected from SEQ ID Nos 7, 8, 9, 10, and 11 derived from *Methanosarcina mazei* or a sequence of an analogous *tRNApyl* gene derived from another bacterial species which contains mutations in 1 or 2 nucleotide positions in a putative A box and/or putative B box of the intragenic RNA polymerase III promoter in positions equivalent to A10G, A52C, T14A, or a combination thereof as shown in SEQ ID Nos 7, 8, 9, 10, or 11.

2. A DNA construct according to claim 1 comprising a terminator sequence which is downstream of the *tRNApyl* coding sequence.

3. A DNA construct according to claim 1 wherein the *tRNApyl* gene has or comprises a sequence of a *tRNApyl* gene derived from *Desulfitobacterium hafniense* (SEQ ID NO: 5), *Methanosarcina acetivorans* (SEQ ID NO: 2), *Methanosarcina burtonii* (SEQ ID NO: 4) or *Methanosarcina thermophila* in which the mutations indicated by bold in SEQ ID Nos 7, 8, 9, 10, or 11 are made in equivalent positions.

4. A DNA construct according to any one of claims 1 to 3 wherein the *tRNApyl* coding sequence is operably linked to a RNA polymerase III promoter wherein said promoter is placed 5' to the *tRNApyl* coding sequence; and optionally the 5' promoter and the *tRNApyl* coding sequence are separated by a spacer sequence containing a transcriptional start site; or wherein the DNA construct comprises an intragenic promoter element and a 5' promoter element being a RNA polymerase Ill promoter element placed 5' to the *tRNApyl* coding sequence such that the combination of said intragenic promoter element and said 5' promoter element constitutes a hybrid promoter; and wherein optionally the 5' promoter element and the *tRNApyl* coding sequence are separated by a spacer sequence containing a transcriptional start site.

5. A DNA construct according to claim 4 wherein said 5' promoter element is the 5' element of a Type 4 RNA polymerase III promoter.

6. A DNA construct according to claim 4 wherein the 5' promoter element of a Type 4 RNA Polymerase Ill promoters is selected from:
5' promoter elements of the EBER RNA gene promoter, the 7SL RNA gene promoter and the Vault RNA gene promoter;
SEQ ID Nos 12, 20, 26, 57, 58, 59, 60, 61, 62, 63, 64 and 65.

7. A DNA construct according to claim any one of claims 3 to 6 wherein the spacer sequence is selected from SEQ ID NOs 13, 21 and 27.

8. A multimeric DNA construct comprising multiple copies (e.g. 2 to 60 copies) of the DNA construct according to any one of claims 1 to 7.

9. A DNA construct according to any one of claims 1 to 8 which further comprises a PylRS coding sequence or gene.

10. A eukaryotic host cell which is transformed with a vector comprising a DNA construct according to any one of claims 1 to 9 and which is optionally transformed with a vector comprising a gene encoding a target protein containing one or more non- natural amino acids encoded by an amber codon.

11. A eukaryotic cell line comprising a DNA construct according to any one of claims 3 to 9 which expresses or is capable of expressing functional PylRS and functional tRNApyl in which functional tRNApyl expression occurs under the control of a an intragenic promoter element and a 5' promoter element being a RNA polymerase III promoter element placed 5' to the tRNApyl coding sequence such that the combination of said intragenic promoter element and said 5' promoter element constitute a hybrid promoter sufficiently to support amber suppression.

12. A eukaryotic host cell according to claim 10 or cell line according to claim 11 wherein the host cell is a mammalian cell or the cell line is a mammalian cell line.

13. A process for preparing a target protein containing one or more non-natural amino acids encoded by an amber codon which comprises expressing said target protein in a eukaryotic cell line according to claim 11, which is transformed with a gene encoding the target protein.