CLAIMS

1. A computer-implemented method for generating an optimized nucleotide sequence, comprising:

   (i) receiving an amino acid sequence, wherein the amino acid sequence encodes a peptide, polypeptide, or protein;

   (ii) receiving a first codon usage table, wherein the first codon usage table comprises a list of amino acids, wherein each amino acid in the table is associated with at least one codon and each codon is associated with a usage frequency;

   (iii) removing from the codon usage table any codons associated with a usage frequency which is less than a threshold frequency;

   (iv) generating a normalized codon usage table by normalising the usage frequencies of the codons not removed in step (iii); and

   (v) generating an optimized nucleotide sequence encoding the amino acid sequence by selecting a codon for each amino acid in the amino acid sequence based on the usage frequency of the one or more codons associated with the amino acid in the normalized codon usage table.

2. The method according to claim 1, wherein normalising comprises:

   (a) distributing the usage frequency of each codon associated with a first amino acid and removed in step (iii) to the remaining codons associated with the first amino acid; and

(b) repeating step (a) for each amino acid to produce the normalized codon usage table,

optionally wherein the usage frequency of the removed codons is distributed:

equally amongst the remaining codons; or

amongst the remaining codons proportionally based on the usage frequency of each remaining codon.

3. The method according to any preceding claim, wherein selecting a codon for each amino acid comprises:

(a) identifying, in the normalized codon usage table, the one or more codons associated with a first amino acid of the amino acid sequence;

(b) selecting a codon associated with the first amino acid, wherein the probability of selecting a certain codon is equal to the usage frequency associated with the codon associated with the first amino acid in the normalized codon usage table; and

(c) repeating steps (a) and (b) until a codon has been selected for each amino acid in the amino acid sequence.

4. The method according to any preceding claim, wherein step (v) is performed a plurality of times to generate a list of optimized nucleotide sequences.

5. The method according to any preceding claim, wherein the threshold frequency is selectable by a user, and/or wherein the threshold frequency is in the range of 5% - 30%, in particular 5%, 10%, or 15%, or 20%, or 25%, or 30%, or, in particular, 10%.

6. The method according to claim 4 or claim 5, further comprising:

screening the list of optimized nucleotide sequences to identify and remove optimized nucleotide sequences failing to meet one or more criteria,

optionally wherein screening the list of optimized nucleotide sequences comprises, for each of the one or more criteria:

determining whether each optimized nucleotide sequence in the list, or most recently updated list, of optimized nucleotide sequences meets the criterion; and

updating the list of optimized nucleotide sequences by removing any nucleotide sequence from the list, or most recently updated list, if the nucleotide sequence does not meet the criterion.

7. The method according to claim 6, wherein determining whether each optimized nucleotide sequence in the list, or most recently updated list, of optimized nucleotide sequences meets the criterion comprises, for each nucleotide sequence:
determining whether a first portion of the nucleotide sequence meets the criterion, and wherein updating the list of optimized nucleotide sequences comprises:

removing the nucleotide sequence if the first portion does not meet the criterion, optionally wherein determining whether each optimized nucleotide sequence in the list, or most recently updated list, of optimized nucleotide sequences meets the criterion further comprises, for each nucleotide sequence:

determining whether one or more additional portions of the nucleotide sequence meets the criterion, wherein the additional portions are non-overlapping with each other and with the first portion, and wherein updating the list of optimized sequences comprises:

removing the nucleotide sequence if any portion does not meet the criterion, optionally wherein determining whether an optimized nucleotide sequence meets the criterion is halted when any portion is determined not to meet the criterion.

8. The method according to claim 7, wherein the first portion and/or the one or more additional portions of the nucleotide sequence comprise a predetermined number of nucleotides, optionally wherein the predetermined number of nucleotides is in the range of: 5 to 300 nucleotides, or 10 to 200 nucleotides, or 15 to 100 nucleotides, or 20 to 50 nucleotides, e.g., 30 nucleotides, e.g., 100 nucleotides.

9. The method according to any one of claims 6 to 8, wherein a first criterion comprises the nucleotide sequence not containing a termination signal, such that determining and updating comprise:

determining whether each optimized nucleotide sequence in the list, or most recently updated list, of optimized nucleotide sequences contains a termination signal; and

updating the list of optimized nucleotide sequences by removing any nucleotide sequence from the list, or most recently updated list, if the nucleotide sequence contains one or more termination signals.

10. The method according to claim 9, wherein the one or more termination signals has/have the following nucleotide sequence:

a. 5'-X_1AUCUX_2UX_3-3', wherein X_1, X_2 and X_3 are independently selected from A, C, U or G, optionally wherein the one or more termination signals has/have one of the following nucleotide sequences:

UAUCUGUU; and/or
UUUUUU; and/or
AAGCUU; and/or
GAAGAGC; and/or
UCUAGA, or

b. 5'-X_1ATCTX_2TX_3-3', wherein X_1, X_2 and X_3 are independently selected from A, C, T or G, optionally wherein the one or more termination signal has/have one or more of the following nucleotide sequences:

TATCTGTT; and/or
TTTTTT; and/or
AAGCTT; and/or
GAAGAGC; and/or
TCTAGA.

11. The method according to any one of claims 6 to 10, wherein a second criterion comprises the nucleotide sequence having a guanine-cytosine content within a predetermined guanine-cytosine content range, such that determining and updating comprise:

determining the guanine-cytosine content of each of the optimized nucleotide sequences in the list, or most recently updated list, of optimized nucleotide sequences, wherein the guanine-cytosine content of a sequence is the percentage of bases in the nucleotide sequence that are guanine or cytosine;

updating the list of optimized nucleotide sequences by removing any nucleotide sequence from the list, or most recently updated list, if its guanine-cytosine content falls outside the predetermined guanine-cytosine content range,

optionally wherein the predetermined guanine-cytosine content range is selectable by a user and/or wherein the predetermined guanine-cytosine content range is 15% - 75%, or 40% - 60%, or, in particular 30% - 70%.

12. The method according to any one of claims 6 to 11, wherein a third criterion comprises the nucleotide sequence having a codon adaptation index greater than a predetermined codon adaptation index threshold, such that determining and updating comprise:

determining the codon adaptation index of each of the optimized nucleotide sequences in the list, or most recently updated list, of optimized nucleotide sequences, wherein the codon adaptation index of a sequence is a measure of codon usage bias and can be a value between 0 and 1;

updating the list, or most recently updated list, of optimized nucleotide sequences by removing any nucleotide sequence if its codon adaptation index is less than or equal to the predetermined codon adaptation index threshold,

optionally wherein the codon adaptation index threshold is selectable by a user, and/or wherein the codon adaptation index threshold is 0.7, or 0.75, or 0.85, or 0.9, or, in particular, 0.8.

13. The method according to any one of claims 6 to 12, wherein a fourth criterion comprises the nucleotide sequence not containing at least 2, for example 3, adjacent identical codons, such that determining and updating comprise:

determining whether any optimized nucleotide sequence in the list, or most recently updated list, of optimized nucleotide sequences, containing at least 2, for example 3 or more, adjacent identical codons; and

updating the list, or most recently updated list, of optimized nucleotide sequences by removing any nucleotide sequence if it contains at least 2, for example 3 or more, adjacent identical codons, optionally wherein the fourth criterion is applied only in respect of codons whose frequency in the normalized codon usage table is less than an adjacency rarity threshold, wherein the adjacency rarity threshold is between 10 and 50%, for example between 15 and 40 %, for example between 20 and 30%.

14. The method according to any preceding claim, wherein the amino acid sequence is received from a database of amino acid sequences, optionally further comprising requesting the amino acid sequence from the database of amino acid sequences, wherein the amino acid sequence is received in response to the request.

15. The method according to any preceding claim, wherein the first codon usage table is received from a database of codon usage tables, optionally further comprising requesting the first codon usage table from the database of codon usage tables, wherein the first codon usage table is received in response to the request.

16. The method according to any preceding claim, further comprising displaying at least one optimized nucleotide sequence on a screen.

17. A computer program comprising instructions which, when the program is executed by a computer, cause the computer to carry out the method of any preceding claim.

18. A data processing system comprising means for carrying out the method of any one of claims 1 to 16.

19. A computer-readable data carrier having stored thereon the computer program of claim 17.

20. A data carrier signal carrying the computer program of claim 17.

21. A method for synthesizing a nucleotide sequence, comprising:

performing the computer-implemented method of any one of claims 1 to 16 to generate at least one optimized nucleotide sequence; and

synthesizing at least one of the generated optimized nucleotide sequences.

22. The method according to claim 21, wherein:

a. the method further comprises inserting the synthesized optimized sequence in a nucleic acid vector for use *in vitro* transcription; and/or

b. the method further comprises inserting one or more termination signals at the 3' end of the synthesized optimized nucleotide sequence, optionally wherein:

i. the one or more termination signals are encoded by the following nucleotide sequence:

5'-X_1ATCTX_2TX_3-3',

wherein X_1, X_2 and X_3 are independently selected from A, C, T or G; and/or

ii. the one or more termination signals are encoded by one or more of the following nucleotide sequences:

TATCTGTT;
TTTTTT;
AAGCTT;
GAAGAGC; and/or

TCTAGA; and/or

iii. more than one termination signal is inserted, and said termination signals are separated by 10 base pairs or fewer, e.g. separated by 5-10 base pairs, further optionally wherein the more than one termination signals are encoded by the following nucleotide sequence: (a) 5'-X_1ATCTX$_2$TX$_3$-(Z_N)- X$_4$ATCTX$_5$TX$_6$-3' or (b) 5'-X_1ATCTX$_2$TX$_3$-(Z_N)-X$_4$ATCTX$_5$TX$_6$-(Z_M)- X$_7$ATCTX$_8$TX$_9$-3', wherein X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8 and X_9 are independently selected from A, C, T o

r G, Z_N represents a spacer sequence of N nucleotides, and Z_M represents a spacer sequence of M nucleotides, each of which are independently selected from A, C, T of G, and wherein N and/or M are independently 10 or fewer; and/or

c. the nucleic acid vector comprises an RNA polymerase promoter operably linked to the optimized nucleotide sequence, optionally wherein the RNA polymerase promoter is a SP6 RNA polymerase promoter or a T7 RNA polymerase promoter; and/or

d. the nucleic acid vector comprises a nucleotide sequence encoding a 5' UTR operably linked to the optimized nucleotide sequence, optionally wherein:

i. the 5' UTR is different to the 5' UTR of a naturally occurring mRNA encoding the amino acid sequence; and/or

ii. the 5' UTR has the nucleotide sequence of SEQ ID NO: 16; and/or

e. the nucleic acid vector comprises a nucleotide sequence encoding a 3' UTR operably linked to the optimized nucleotide sequence, optionally wherein:

i. the 3' UTR is different to the 3' UTR of a naturally occurring mRNA encoding the amino acid sequence; and/or

ii. the 3’ UTR has the nucleotide sequence of SEQ ID NO: 17 or SEQ ID NO: 18; and/or

f. the nucleic acid vector is a plasmid, optionally wherein:

i. the plasmid is linearized before *in vitro* transcription; or

ii. the plasmid is not linearized before *in vitro* transcription, further optionally wherein the plasmid is supercoiled.

23. The method according to claim 21 or claim 22, wherein the method further comprises using at least one of the synthesized optimized nucleotide sequences in *in vitro* transcription to synthesize mRNA, optionally wherein:

a. the mRNA is synthesized by a SP6 RNA polymerase, further optionally wherein:

i. the SP6 RNA polymerase is a naturally occurring SP6 RNA polymerase; or

ii. the SP6 RNA polymerase is a recombinant SP6 RNA polymerase, optionally wherein the SP6 RNA polymerase comprises a tag, optionally wherein the tag is a his-tag; or

b. the mRNA is synthesized by a T7 RNA polymerase; and/or

c. the method further comprises a separate step of capping and/or tailing the synthesized mRNA; or

d. capping and tailing occurs during *in vitro* transcription.

24. The method according to claim 23, wherein:

a. the mRNA is synthesized in a reaction mixture comprising NTPs at a concentration ranging from 1-10 mM each NTP, the DNA template at a concentration ranging from 0.01-0.5 mg/ml, and the SP6 RNA polymerase at a concentration ranging from 0.01-0.1 mg/ml, optionally wherein the reaction mixture comprises NTPs at a concentration of 5 mM each NTP, the DNA template at a concentration of 0.1 mg/ml, and the SP6 RNA polymerase at a concentration of 0.05 mg/ml; and/or

b. the mRNA is synthesized at a temperature ranging from 37-56 °C;

optionally wherein (i) the NTPs are naturally-occurring NTPs; or (ii) the NTPs comprise modified NTPs.

25. The method according to any one of claims 21 to 24, wherein the method further comprises transfecting the synthesized optimized nucleotide sequence into a cell either *in vitro* or *in vivo*, optionally wherein:

a. the expression level of the protein encoded by the synthesized optimized nucleotide sequence in transfected cell is determined; and/or

b. the functional activity of the protein encoded by the synthesized optimized nucleotide sequence is determined.

26. The method according to any one of claims 1 to 16, wherein:

a. the method further comprises synthesizing a reference nucleotide sequence encoding the amino acid sequence and the at least one optimized nucleotide sequence according to the method of any one of claims 36 to 70, and contacting the reference nucleotide sequence and the at least one optimized nucleotide sequence with a separate cell or organism, wherein the cell or organism contacted with the at least one synthesized optimized nucleotide sequence produces an increased yield of the protein encoded by the optimized nucleotide sequence compared to the yield of the protein encoded by the reference nucleotide

sequence produced by the cell or organism contacted with the synthesized reference nucleotide sequence; or

b. the at least one optimized nucleotide sequence, when synthesized, is configured to increase the expression of the protein encoded by the at least one optimized nucleotide sequence compared to the expression of the protein encoded by the reference nucleotide sequence, when synthesized;

optionally wherein the reference nucleotide sequence is (a) a naturally occurring nucleotide sequence encoding the amino acid sequence or (b) a nucleotide sequence encoding the amino acid sequence generated by a method other than the method according to any one of claims 1 to 16.

27. The method of any one of claims 21 to 25, wherein the method further comprises producing a therapeutic composition comprising an mRNA encoding a therapeutic peptide, polypeptide, or protein for use in the delivery to or treatment of a subject,

optionally wherein the mRNA encodes cystic fibrosis transmembrane conductance regulator (CFTR) protein;

further optionally wherein the reference nucleotide sequence is (a) a naturally occurring nucleotide sequence encoding the amino acid sequence or (b) a nucleotide sequence encoding the amino acid sequence generated by a method other than the method according to any one of claims 1 to 16.

28. A synthesized optimized nucleotide sequence generated according to the methods of any one of claims 21 to 24 and 27 for use in therapy.

29. An *in vitro* synthesized nucleic acid, optionally wherein the nucleic acid is mRNA and/or is for use in therapy, comprising an optimized nucleotide sequence consisting of codons associated with a usage frequency which is greater than or equal to 10%; wherein the optimized nucleotide sequence:

(i) does not contain a termination signal having one of the following nucleotide sequences:

5'-X_1AUCUX_2UX_3-3', wherein X_1, X_2 and X_3 are independently selected from A, C, U or G; and 5'-X_1AUCUX_2UX_3-3', wherein X_1, X_2 and X_3 are independently selected from A, C, U or G;

(ii) does not contain any negative cis-regulatory elements and negative repeat elements; and

(iii) has a codon adaptation index greater than 0.8;

wherein, when divided into non-overlapping 30 nucleotide-long portions, each portion of the optimized nucleotide sequence has a guanine cytosine content range of 30% - 70%.

30. The *in vitro* synthesized nucleic acid of claim 29, wherein the optimized nucleotide sequence does not contain a termination signal having one of the following sequences: TATCTGTT; TTTTTT; AAGCTT; GAAGAGC; TCTAGA; UAUCUGUU; UUUUUU; AAGCUU; GAAGAGC; UCUAGA.