WHO
International Nonproprietary Names Programme

7/2020

11480

Schematic

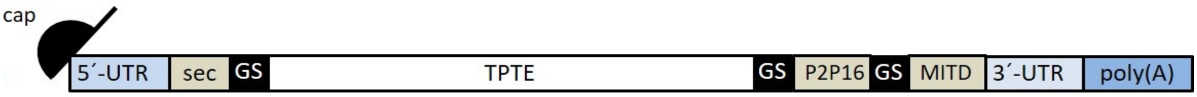


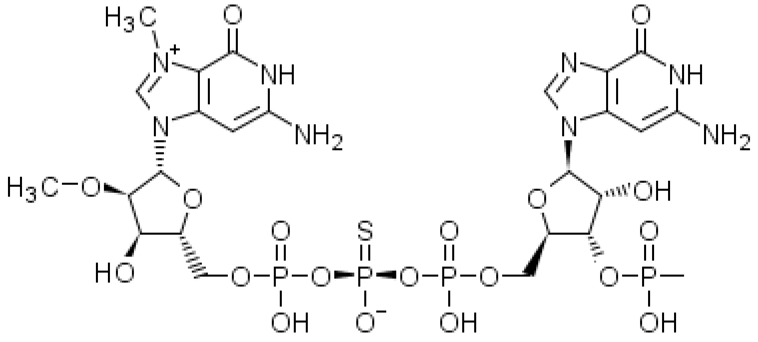
Table of features

Element	Description	Position
5´cap	A specific capping structure at the 5´-end for improved RNA stability and translational efficiency.	1-2
5'-UTR (hAg-Kozak)	The 5´-UTR sequence has been derived from the human alpha-globin RNA with an optimized "Kozak sequence".	3-53
sec	The secretory signal peptide "sec" derived from the sequence encoding the human MHC Class I complex alpha chain "HLA-I, Cw*" is used as a fusion-protein tag, to improve antigen processing and presentation.	54-131
GS	Sequences that code for short GS-rich linker peptides generally consisting of the amino acids glycine and serine (usually GGSGGGGSGG).	132-161, 1812-1841 and 2037-2078
TPTE	Codon-optimized sequences encoding the target protein TPTE.	162-1811
P2P16	Sequence coding for tetanus toxoid-derived helper epitopes to break immunological tolerance.	1842-2036
MITD	MITD corresponds to the transmembrane and cytoplasmic domain of the MHC class I molecule and is used as a fusion-protein tag, improve antigen processing and presentation.	2079-2249
3'-UTR (FI element)	The 3´-UTR is a combination of two sequence elements derived from the AES mRNA (called F) and the mitochondrial encoded 12S ribosomal RNA (called I) to confer RNA stability and highest total protein expression.	2250-2566
Poly(A)	A poly(A)-tail measuring 110 nucleotides in length, consisting of a stretch of 30 adenosine residues, followed by a 10-nucleotide linker sequence and another 70 adenosine residues.	2567-2676

Abbreviations: AES = amino terminal enhancer of split; GS = glycine and serine;

HLA = human leukocyte antigen, MHC = major histocompatibility complex; MITD = MHC class I transmembrane domain; UTR = untranslated region.

5‘- capping structure = G^1G^2: m^7Guo^+2'*O*Me-5'-p-(*R*)p^S-p-5'-G



1-2 : cap m^7Guo^+2'*O*Me-5'-p-(*R*)p^S-p-5'-G; *3-53* : 5'UTR; **54-131** : sec; **132-161** : $G_2SG_4SG_2$ linker; 162-1811 : TPTE; **1812-1841** : $G_2SG_4SG_2$ linker; **1842-2036** : P2P16 tetanustoxoid-derived helper epitopes; **2037-2078** : $GS_2G_4SPG_3S_2$ linker; **2079-2249** : MITD; *2250-2566* : FI element (3' UTR); **2567-2676** : A30L70 polyA tail.

Sequence / Séquence / Secuencia

GG_GGCGAACU AGUAUUCUUC UGGUCCCCAC AGACUCAGAG AGAACCCGCC_ 50
ACC**AUGAGAG UGAUGGCCCC CAGAACCCUG AUCCUGCUGC UGUCUGGCGC** 100
CCUGGCCCUG ACAGAGACAU GGGCCGGAAG CGGCGGCUCU GGAGGAGGCG 150
GCUCCGGAGG CAUGAACGAG AGCCCCGACC CUACAGAUCU GGCCGGCGUG 200
AUCAUCGAGC UGGGACCCAA CGAUAGCCCU CAGACCAGCG AGUUCAAGGG 250
GGCCACAGAG GAAGCCCCUG CCAAAGAGAG CCCCCACACC UCCGAGUUUA 300
AGGGCGCUGC UCGGGUGUCC CCUAUCAGCG AGAGCGUGCU GGCCCGGCUG 350
AGCAAGUUCG AGGUGGAGGA CGCCGAGAAC GUGGCCAGCU ACGACAGCAA 400
GAUCAAGAAA AUCGUGCACA GCAUCGUGUC CAGCUUCGCC UUCGGCCUGU 450
UCGGCGUGUU CCUGGUGCUG CUGGACGUGA CACUGAUCCU GGCCGACCUG 500
AUCUUCACCG ACAGCAAGCU GUACAUCCCC CUGGAAUACC GGUCCAUCAG 550
CCUGGCCAUU GCCCUGUUCU UUCUGAUGGA CGUGCUGCUG CGGGUGUUCG 600
UGGAGCGGCG GCAGCAGUAC UUCAGCGACC UGUUCAACAU CCUGGACACC 650
GCCAUCAUCG UGAUUCUGCU GCUGGUGGAU GUGGUGUACA UCUUCUUCGA 700
CAUCAAGCUG CUGAGAAACA UCCCCCGGUG GACCCAUCUG CUGCGGCUGC 750
UGAGACUGAU CAUCCUGCUG CGGAUCUUCC ACCUGUUCCA CCAGAAGCGG 800
CAGCUGGAAA AGCUGAUCAG ACGGCGGGUG UCCGAGAACA AGCGGCGGUA 850
CACCAGGGAC GGCUUCGACC UGGACCUGAC CUACGUGACC GAGCGGAUCA 900
UUGCCAUGAG CUUCCCCAGC AGCGGCAGAC AGAGCUUCUA CCGGAACCCC 950
AUCAAAGAAG UGGUGCGGUU CCUGGACAAG AAGCACCGGA ACCACUACCG 1000
GGUGUACAAC CUGUGCAGCG AGCGGGCCUA CGACCCCAAG CACUUCCACA 1050
ACCGGGUGGU GCGGAUCAUG AUCGACGACC ACAACGUGCC CACCCUGCAC 1100

CAGAUGGUGG UGUUCACCAA AGAAGUGAAC GAGUGGAUGG CCCAGGACCU 1150
GGAAAACAUC GUGGCCAUCC ACUGCAAGGG CGGCACCGAC AGAACCGGCA 1200
CCAUGGUGUG CGCCUUUCUG AUCGCCAGCG AGAUCUGUAG CACCGCCAAA 1250
GAGUCCCUGU ACUACUUCGG CGAGCGGAGA ACCGACAAGA CCCACAGCGA 1300
GAAGUUCCAG GGCGUGGAGA CACCCAGCCA GAAAAGAUAU GUGGCUUACU 1350
UCGCCCAGGU GAAGCACCUG UACAACUGGA ACCUGCCCCC CAGACGGAUU 1400
CUGUUCAUCA AGCACUUCAU CAUCUACAGC AUCCCCAGAU ACGUGCGGGA 1450
CCUGAAGAUC CAGAUCGAGA UGGAAAAGAA AGUGGUGUUC AGCACCAUCU 1500
CCCUGGGCAA GUGCAGCGUG CUGGACAACA UCACCACCGA CAAGAUCCUG 1550
AUCGACGUGU UCGACGGCCU GCCCCUGUAC GACGACGUGA AGGUGCAGUU 1600
CUUCUACAGC AACCUGCCCA CCUACUACGA CAAUUGCAGC UUCUACUUCU 1650
GGCUGCACAC CAGCUUCAUC GAGAACAACA GGCUGUACCU GCCCAAGAAC 1700
GAGCUGGACA ACCUGCACAA GCAGAAGGCC AGAAGAAUCU ACCCCAGCGA 1750
CUUCGCCGUG GAGAUCCUGU UUGGCGAGAA GAUGACCAGC AGCGACGUGG 1800
UGGCCGGCAG C**GGAGGAUCC GGUGGUGGCG GCAGCGGCGG CAAGAAGCAG** 1850
UACAUCAAGG CCAACAGCAA GUUCAUCGGC AUCACCGAGC UGAAGAAGCU 1900
GGGAGGGGGC AAACGGGGAG GCGGCAAAAA GAUGACCAAC AGCGUGGACG 1950
ACGCCCUGAU CAACAGCACC AAGAUCUACA GCUACUUCCC CAGCGUGAUC 2000
AGCAAAGUGA ACCAGGGCGC UCAGGGCAAG AAACUGGGCU CUAGCGGAGG 2050
GGGAGGCUCU CCUGGCGGGG GAUCUAGCAU CGUGGGAAUU GUGGCAGGAC 2100
UGGCAGUGCU GGCCGUGGUG GUGAUCGGAG CCGUGGUGGC UACCGUGAUG 2150
UGCAGACGGA AGUCCAGCGG AGGCAAGGGC GGCAGCUACA GCCAGGCCGC 2200
CAGCUCUGAU AGCGCCCAGG GCAGCGACGU GUCACUGACA GCCUAGUAA*C* 2250
UCGAGCUGGU ACUGCAUGCA CGCAAUGCUA GCUGCCCCUU UCCCGUCCUG 2300
GGUACCCCGA GUCUCCCCCG ACCUCGGGUC CCAGGUAUGC UCCCACCUCC 2350
ACCUGCCCCA CUCACCACCU CUGCUAGUUC CAGACACCUC CCAAGCACGC 2400
AGCAAUGCAG CUCAAAACGC UUAGCCUAGC CACACCCCCA CGGGAAACAG 2450
CAGUGAUUAA CCUUUAGCAA UAAACGAAAG UUUAACUAAG CUAUACUAAC 2500
CCCAGGGUUG GUCAAUUUCG UGCCAGCCAC ACCGAGACCU GGUCCAGAGU 2550
*CGCUAGCCGC GUCGCU***AAAA AAAAAAAAAA AAAAAAAAAA AAAAAAGCAU** 2600
AUGACUAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 2650
AAAAAAAAAA AAAAAAAAAA AAAAAA 2676