11478

Schematic

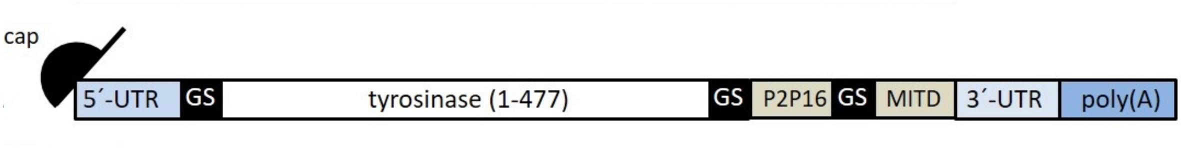


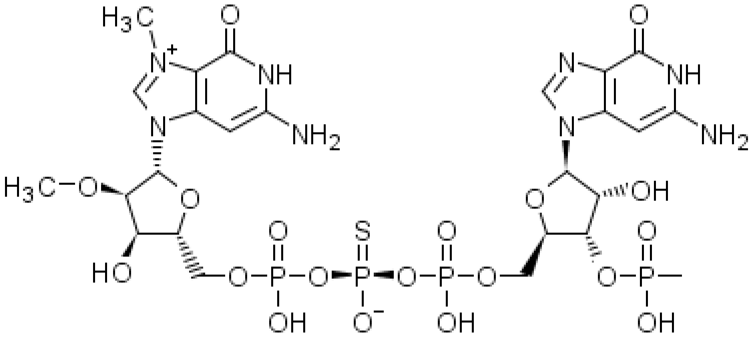
Table of features

Element	Description	Position
5´cap	A specific capping structure at the 5´-end for improved RNA stability and translational efficiency.	1-1
5'-UTR (hAg-Kozak)	5´-UTR sequence derived from the human alpha-globin RNA with an optimized "Kozak sequence".	3-53
Tyrosinase	Codon-optimized sequences encoding the target protein tyrosinase. Note that for tyrosinase only the sequence coding for the first 477 amino acids is used; the endogenous trans-membrane domain of tyrosinase starting with amino acid 478 has been deleted	54-1484
GS	Sequences that code for short GS-rich linker peptides generally consisting of the amino acids glycine and serine (usually GGSGGGGSGG).	1485-1514, 1710-1751
P2P16	Sequence coding for tetanus toxoid-derived helper epitopes to break immunological tolerance.	1515-1709
MITD	MITD corresponds to the transmembrane and cytoplasmic domain of the MHC class I molecule and is used as a fusion-protein tag, to improve antigen processing and presentation.	1752-1922
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.	1923-2239
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.	2240-2349

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$^+$2'*O*Me-5'-p-(*R*)p^S-p-5'-G; *3-53*: 5'UTR; **54-1484** : tyrosinase; **1485-1514** : G$_2$SG$_4$SG$_2$ linker; **1515-1709** : P2P16 tetanus toxoid-derived helper epitopes; **1710-1751** : GS$_2$G$_4$SPG$_3$S$_2$ linker; **1752-1922** : MITD; *1923-2239* : FI element (3' UTR); **2240-2349** : A30L70 polyA tail.

Sequence / Séquence / Secuencia

GG*GGCGAACU AGUAUUCUUC UGGUCCCCAC AGACUCAGAG AGAACCCGCC* 50
*ACC***AUGCUUC UGGCUGUGCU CUAUUGCUUG CUGUGGUCCU UCCAGACUUC** 100
UGCCGGACAC UUUCCCAGAG CAUGUGUGUC AUCCAAAAAC CUCAUGGAGA 150
AAGAAUGCUG UCCUCCUUGG AGUGGUGAUA GAUCCCCAUG UGGACAGCUC 200
UCAGGCAGAG GAUCUUGCCA GAACAUUCUG CUGAGCAAUG CACCUCUUGG 250
CCCACAGUUU CCCUUUACCG GAGUAGAUGA CAGAGAAAGU UGGCCCUCCG 300
UGUUCUACAA CAGAACAUGC CAAUGCAGCG GCAAUUUUAU GGGAUUUAAC 350
UGCGGAAAUU GCAAAUUUGG AUUUUGGGGC CCAAACUGUA CAGAAAGAAG 400
ACUGCUCGUU AGAAGAAACA UCUUUGAUCU GAGCGCACCC GAAAAAGACA 450
AAUUCUUCGC UUAUCUGACU CUCGCCAAAC ACACAAUAAG CAGUGAUUAU 500
GUCAUUCCCA UUGGCACUUA UGGACAGAUG AAAAAUGGCU CAACUCCCAU 550
GUUCAACGAC AUCAACAUAU ACGAUCUGUU UGUGUGGAUG CAUUACUACG 600
UGAGUAUGGA UGCUCUGCUG GGUGGCUCCG AAAUAUGGAG GGAUAUAGAU 650
UUUGCACACG AAGCUCCUGC CUUCCUUCCA UGGCAUAGAC UGUUCCUGCU 700
GAGAUGGGAA CAAGAAAUCC AGAAACUUAC AGGCGAUGAA AACUUCACUA 750
UCCCUUAUUG GGAUUGGAGA GAUGCUGAGA AAUGCGAUAU CUGUACCGAU 800
GAAUACAUGG GUGGUCAACA CCCAACCAAC CCCAAUCUCC UGAGCCCUGC 850
CUCUUUCUUC AGUUCUUGGC AGAUUGUCUG UUCAAGAUUG GAAGAAUACA 900
AUUCCCAUCA GUCCCUGUGU AACGGAACAC CAGAAGGACC UCUGAGAAGA 950
AACCCUGGCA AUCAUGACAA AAGCAGGACA CCUAGACUUC CCUCCUCUGC 1000
CGACGUUGAA UUUUGCCUCU CUCUGACCCA GUACGAAAGC GGUAGCAUGG 1050
ACAAAGCCGC CAAUUUCAGC UUUAGAAAUA CCUUGGAAGG AUUUGCCUCA 1100
CCUCUGACAG GAAUUGCUGA UGCUAGCCAG UCAUCCAUGC ACAACGCUUU 1150

GCACAUCUAU AUGAAUGGAA CCAUGAGUCA GGUUCAGGGA UCUGCCAAUG 1200
AUCCUAUUUU CCUGCUGCAC CAUGCAUUCG UGGACUCUAU CUUUGAGCAG 1250
UGGCUUAGAA GACACAGACC AUUGCAGGAA GUCUAUCCAG AGGCCAACGC 1300
ACCAAUCGGC CAUAAUAGAG AAAGCUACAU GGUACCCUUC AUUCCUCUGU 1350
ACAGAAAUGG AGAUUUCUUC AUCAGCUCCA AAGACCUGGG CUACGAUUAC 1400
UCAUAUCUGC AAGACAGUGA UCCCGAUAGC UUCCAAGACU ACAUUAAGUC 1450
UUAUCUCGAA CAGGCGAGCA GAAUCUGGUC CUGGGGAGGA UCCGGUGGUG 1500
GCGGCAGCGG CGGCAAGAAG CAGUACAUCA AGGCCAACAG CAAGUUCAUC 1550
GGCAUCACCG AGCUGAAGAA GCUGGGAGGG GGCAAACGGG GAGGCGGCAA 1600
AAAGAUGACC AACAGCGUGG ACGACGCCCU GAUCAACAGC ACCAAGAUCU 1650
ACAGCUACUU CCCCAGCGUG AUCAGCAAAG UGAACCAGGG CGCUCAGGGC 1700
AAGAAACUGG GCUCUAGCGG AGGGGGAGGC UCUCCUGGCG GGGGAUCUAG 1750
CAUCGUGGGA AUUGUGGCAG GACUGGCAGU GCUGGCCGUG GUGGUGAUCG 1800
GAGCCGUGGU GGCUACCGUG AUGUGCAGAC GGAAGUCCAG CGGAGGCAAG 1850
GGCGGCAGCU ACAGCCAGGC CGCCAGCUCU GAUAGCGCCC AGGGCAGCGA 1900
CGUGUCACUG ACAGCCUAGU AA*CUCGAGCU GGUACUGCAU GCACGCAAUG* 1950
CUAGCUGCCC CUUUCCCGUC CUGGGUACCC CGAGUCUCCC CCGACCUCGG 2000
GUCCCAGGUA UGCUCCCACC UCCACCUGCC CCACUCACCA CCUCUGCUAG 2050
UUCCAGACAC CUCCCAAGCA CGCAGCAAUG CAGCUCAAAA CGCUUAGCCU 2100
AGCCACACCC CCACGGGAAA CAGCAGUGAU UAACCUUUAG CAAUAAACGA 2150
AAGUUUAACU AAGCUAUACU AACCCCAGGG UUGGUCAAUU UCGUGCCAGC 2200
*CACACCGAGA CCUGGUCCAG AGUCGCUAGC CGCGUCGCU***A AAAAAAAAAA** 2250
AAAAAAAAAA AAAAAAAAAG CAUAUGACUA AAAAAAAAAA AAAAAAAAAA 2300
AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAA 2349