



10125

Structure



Nucleotide sequence / séquence de nucléotides / Secuencia de nucleotidos

```

GGGGAGAAAG CUUACCAUGG UGCCCCAGGC CCUGCUCUUC GUCCCCGUCG 50
UGGUGUUCCC CCUCUGCUUC GGCAAGUUC CCAUCUACAC CAUCCCCGAC 100
AAGCUGGGGC CGUGGAGCCC CAUCGACAUC CACCACCUGU CCUGCCCCAA 150
CAACCUCGUG GUCGAGGACG AGGGCUGCAC CAACCUGAGC GGGUUCUCCU 200
ACAUGGAGCU GAAGGUGGGC UACAUCAGCG CCAUCAAGAU GAACGGGUUC 250
ACGUGCACCG GCGUGGUCAC CGAGGCGGAG ACCUACACGA ACUUCGUGGG 300
CUACGUGACC ACCACCUUCA AGCGGAAGCA CUUCCGCCCC ACGCCGGACG 350
CCUGCCGGGC CGCCUACAAC UGGAAGAUGG CCGGGGACCC CCGCUACGAG 400
GAGUCCCUCC ACAACCCCUA CCCCAGCUAC CACUGGCUGC GGACCGUCAA 450
GACCACCAAG GAGAGCCUGG UGAUCAUCUC CCCGAGCGUG GCGGACCUUG 500
ACCCCUACGA CCGCUCCUG CACAGCCGGG UCUUCCCCGG CGGGAACUGC 550
UCCGGCGUGG CCGUGAGCUC CACGUACUGC AGCACCAACC ACGACUACAC 600
CAUCUGGAUG CCCGAGAACC CGCGCCUGGG GAUGUCCUGC GACAUUUUCA 650
CCAACAGCCG GGGCAAGCGC GCCUCCAAGG GCAGCGAGAC GUGCGGGUUC 700
GUCGACGAGC GGGGCCUCUA CAAGUCCUG AAGGGGGCCU GCAAGCUGAA 750
GCUCUGCGGC GUGCUGGGCC UGCGCCUCAU GGACGGGACC UGGGUGGCGA 800
UGCAGACCAG CAACGAGACC AAGUGGUGCC CCCCCGGCCA GCUGGUCAAC 850
CUGCACGACU UCCGGAGCGA CGAGAUUGAG CACCUCGUGG UGGAGGAGCU 900
GGUCAAGAAG CGCGAGGAGU GCCUGGACGC CCUCGAGUCC AUCAUGACGA 950
CCAAGAGCGU GUCCUUCGGG CGCCUGAGCC ACCUGCGGAA GCUCGUGCCC 1000
GGGUUCGGCA AGGCCUACAC CAUCUUCAAC AAGACCCUGA UGGAGGCCGA 1050
CGCCCACUAC AAGUCCGUCC GCACGUGGAA CGAGAUCAUC CCGAGCAAGG 1100
GGUGCCUGCG GGUGGGCGGC CGCUGCCACC CCCACGUCAA CGGGGUGUUC 1150
UUCAACGGCA UCAUCCUCGG GCCCGACGGC AACGUGCUGA UCCCCGAGAU 1200
GCAGUCCAGC CUGCUCAGC AGCACAUGGA GCUGCUGGUC UCCAGCGUGA 1250
UCCCGCUCAU GCACCCCUUG GCGGACCCCU CCACCGUGUU CAAGAACGGG 1300
GACGAGGCCG AGGACUUCGU CGAGGUGCAC CUGCCCGACG UGCACGAGCG 1350
GAUCAGCGGC GUCGACCUCG GCCUGCCGAA CUGGGGGAAG UACGUGCUGC 1400
UCUCCGCCGG CGCCUGACC GCCCUGAUGC UGAUCAUCUU CCUCAUGACC 1450
UGCUGGCGCC GGGUGAACCG GAGCGAGCCC ACGCAGCACA ACCUGCGCGG 1500
GACCGGCCGG GAGGUCUCCG UGACCCCGCA GAGCGGGAAG AUCAUCUCCA 1550
GCUGGGAGUC CUACAAGAGC GCGGCGGAGA CCGGGCUGUG AGGACUAGUU 1600
AUAAGACUGA CUAGCCCGAU GGGCCUCCA ACGGGCCUC CUCCCCUCCU 1650
UGCACCGAGA UUAUUAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1700
AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA CCCCCCCCCC 1750
CCCCCCCCC CCCCCAAGG CUCUUUUCAG AGCCACCAGA AUU 1793

```

Element	Position
5'untranslated region (5'-UTR; artificial sequence)	1-16
Position of start codon	17-19
Coding region (open reading frame (ORF) encoding the rabies RAV-G protein)	17-1591
Position of stop codon	1589-1591
3'-untranslated region (3'-UTR design) including the following elements: Human alpha-globin 3'-UTR sequence element:	1592-1793   1614-1657



Poly (A) sequence:	1666-1729	
Poly (C) sequence:	1735-1764	
Histone stem-loop sequence (consensus of metazoan histones):	1765-1788	

**Modified residues** The 5'-terminal 2 nucleotides (underlined in the sequence) comprise the eukaryotic RNA Cap Structure 7mG(ppp)G.

