

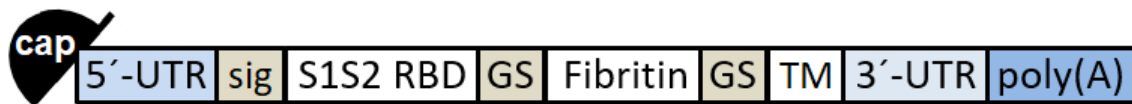


11890

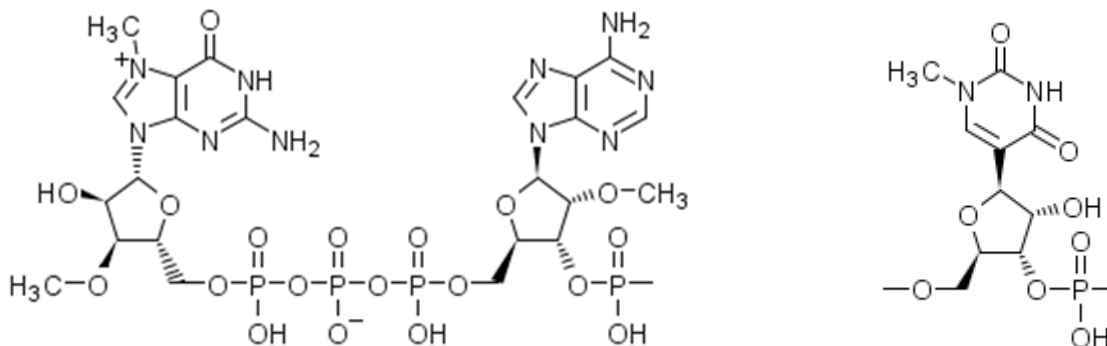
Description

Messenger RNA encoding the receptor binding domain of the SARS-CoV-2 spike (S) glycoprotein connected to the T4 fibrin and the S glycoprotein transmembrane domain.

Schematic



UTR= Untranslated region; sig = extended signal sequence of the S glycoprotein; S1S2 RBD = Receptor Binding Domain of the S glycoprotein; GS = glycine/serine-rich linker; Fibrin = trimerization domain of enterobacteria phage T4 fibrin; TM = transmembrane domain; poly(A) = polyadenylate signal tail.



5'- capping structure

cap G^{1A2} = m⁷G⁺m³-5'-ppp-5'-Am²-3'-p-
[m⁷ = 7-CH₃; m³ = 3'-O-CH₃; m² = 2'-O-CH₃;
-ppp- = -PO₂H-O-PO₂H-O-PO₂H-); -p- = -PO₂H-]

m¹Ψ = 1-methyl-3'-pseudouridylyl



Table of features

Element	Description	Position
cap	A modified 5'-cap1 structure (m ⁷ G+m ³ '-5'-ppp-5'-Am)	1-2
5'-UTR	5'-untranslated region derived from human alpha-globin RNA with an optimized Kozak sequence	3-54
sig	S glycoprotein signal peptide (extended leader sequence), which guides translocation of the nascent polypeptide chain into the endoplasmic reticulum.	55-111
S1S2 RBD	Codon-optimized sequence, encoding the Receptor Binding Domain (RBD) of the of the SARS-CoV-2 spike (S) glycoprotein.	112-717
GS	Sequences that code for short linker peptides consisting of the amino-acids glycine and serine.	718-744 and 826-843
Fibrin	Partial sequence of T4 fibrin (foldon), used as a trimerization domain.	745-825
TM	Transmembrane domain of the S glycoprotein.	844-993
3'-UTR	The 3'-untranslated region comprises two sequence elements derived from the amino-terminal enhancer of split (AES) mRNA and the mitochondrial encoded 12S ribosomal RNA to confer RNA stability and high total protein expression.	994-1288
poly(A)	A 110-nucleotide poly(A)-tail consisting of a stretch of 30 adenosine residues, followed by a 10-nucleotide linker sequence and another 70 adenosine residues.	1289-1398

Sequence / Séquence / Secuencia

```

GAGAAΨAAAC  ΨAGΨAΨCΨΨ  CΨGGΨCCCCA  CAGACΨCAGA  GAGAACCCGC   50
CACCAΨGΨΨΨ  GΨGΨΨΨCΨΨG  ΨGCΨGCΨGCC  ΨCΨΨGΨGΨCΨ  ΨCΨCAGΨGΨG  100
ΨGAAΨΨΨGAC  AGΨGAGAΨΨΨ  CAAAΨAΨΨA  CAAAPCΨGΨG  ΨCCAΨΨΨGGA  150
GAAGΨGΨΨΨA  AΨGCAACAAG  AΨΨΨGCAΨCΨ  GΨGΨAΨGCAΨ  GGAAΨAGAAA  200
AAGAAΨΨΨCΨ  AAΨΨGΨGΨGG  CΨGAΨΨAΨΨC  ΨGΨGCΨGΨAΨ  AAΨAGΨGCΨΨ  250
CΨΨΨΨCCAC  AΨΨΨAAAΨGΨ  ΨAΨGGAGΨGΨ  CΨCCAACAAA  AΨΨAAAΨGAY  300
ΨΨAΨGΨΨΨA  CAAAΨGΨGΨA  ΨGCΨGAΨΨCΨ  ΨΨΨGΨGAΨCA  GAGGΨGAΨGA  350

```



AGΨGAGACAG	AΨΨGCCCCCG	GACAGACAGG	AAAAAΨΨGCΨ	GAΨΨACAAΨΨ	400
ACAAACΨGCC	ΨGAΨGAΨΨΨΨ	ACAGGAΨGΨG	ΨGAΨΨGCΨΨG	GAAΨΨCΨAAΨ	450
AAΨΨΨAGAΨΨ	CΨAAAAGΨGGG	AGGAAAΨΨAC	AAΨΨAΨCΨGΨ	ACAGACΨGΨΨ	500
ΨAGAAAAΨCA	AAΨCΨGAAAC	CΨΨΨΨGAAAG	AGAΨAΨΨΨCA	ACAGAAAΨΨΨ	550
AΨCAGGCΨGG	AΨCAACACCCΨ	ΨGΨAAΨGGAG	ΨGGAAGGAΨΨ	ΨAAΨΨGΨΨAΨ	600
ΨΨΨCCAΨΨAC	AGAGCΨAΨGG	AΨΨΨCAGCCA	ACCAAΨGGΨG	ΨGGGAΨAΨCA	650
GCCAΨAΨAGA	GΨGGΨGGΨGC	ΨGΨCΨΨΨΨGA	ACΨGCΨGCAΨ	GCACCΨGCAA	700
CAGΨGΨGΨGG	ACCΨAAAAGGC	ΨCCCCCGGCΨ	CCGGCΨCCGG	AΨCΨGGΨΨAΨ	750
AΨΨCCΨGAAG	CΨCCAAGAGA	ΨGGGCAAGCΨ	ΨACGΨΨCGΨA	AAGAΨGGCGA	800
AΨGGGΨAΨΨA	CΨΨΨCΨACCCΨ	ΨΨΨΨAAGGAAG	CGGCAGCGGA	ΨCΨGAACAGΨ	850
ACAΨΨAAAΨG	GCCΨΨGGΨAC	AΨΨΨGGCΨΨG	GAΨΨΨAΨΨGC	AGGAΨΨAAΨΨ	900
GCAAΨΨGΨGA	ΨGGΨGACAAΨ	ΨAΨGΨΨAΨGΨ	ΨGΨAΨGACAΨ	CAΨGΨΨGΨΨC	950
ΨΨGΨΨΨAAAA	GGAΨGΨΨGΨΨ	CΨΨGΨGGAAAG	CΨGΨΨGΨΨGA	ΨGACΨCGAGC	1000
ΨGGΨACΨGCA	ΨGCACGCAAΨ	GCΨAGCΨGCC	CCΨΨΨCCCGΨ	CCΨGGGΨACC	1050
CCGAGΨCΨCC	CCCGACCΨCG	GGΨCCCAGGΨ	AΨGCΨCCAC	CΨCCACCΨGC	1100
CCCACΨCACC	ACCΨCΨGCAΨ	GΨΨCCAGACA	CCΨCCCAAGC	ACGCAGCAAΨ	1150
GCAGCΨCAAA	ACGCΨΨAGCC	ΨAGCCACACC	CCCACGGGAA	ACAGCAGΨGA	1200
ΨΨAACCΨΨΨA	GCAAΨAAAACG	AAAGΨΨΨAAC	ΨAAGCΨAΨAC	ΨAACCCACAGG	1250
GΨΨGGΨCAAΨ	ΨΨCGΨGCCAG	CCACACCCΨG	GAGCΨAGCAA	AAAAAAAAAAAA	1300
AAAAAAAAAAAA	AAAAAAAAAAGC	AΨAΨGACΨAA	AAAAAAAAAAAA	AAAAAAAAAAAA	1350
AAAAAAAAAAAA	AAAAAAAAAAAA	AAAAAAAAAAAA	AAAAAAAAAAAA	AAAAAAA	1398

Ψ = 1-methyl-3'-pseudouridylyl