

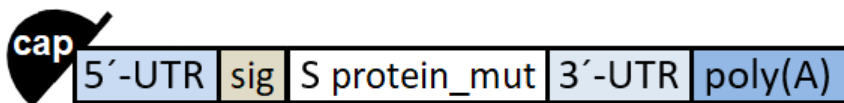


11889

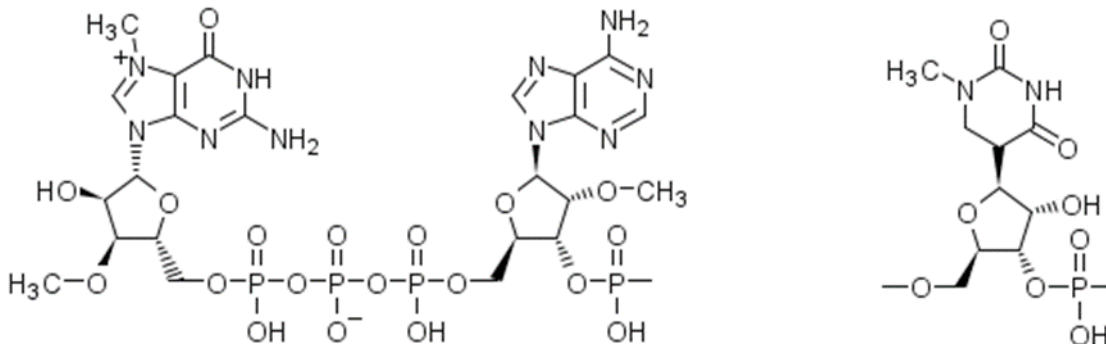
Description

Messenger RNA encoding the full-length SARS-CoV-2 spike glycoprotein.

Schematic



UTR = Untranslated region; sig = extended signal sequence of the S glycoprotein; S protein_mut = S glycoprotein sequence containing mutations K986P and V987P; poly(A) = polyadenylate signal tail.



5'- capping structure

cap G^{1A2} = m⁷G⁺m^{3'}-5'-ppp-5'-Am^{2'}-3'-p-
[m⁷ = 7-CH₃; m^{3'} = 3'-O-CH₃; m^{2'} = 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	3-54



	alpha-globin RNA with an optimized Kozak sequence	
sig	S glycoprotein signal peptide (extended leader sequence), which guides translocation of the nascent polypeptide chain into the endoplasmic reticulum.	55-102
S protein_mut	Codon-optimized sequence encoding full-length SARS-CoV-2 spike (S) glycoprotein containing mutations K986P and V987P to ensure the S glycoprotein remains in an antigenically optimal pre-fusion conformation; stop codons: 3874-3879 (underlined)	103-3879
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.	3880-4174
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.	4175-4284

Sequence / Séquence / Secuencia

GA**G**A**A**Ψ**A**A**A**C Ψ**A**G**Ψ**A**Ψ**Ψ**C**Ψ**Ψ** C**Ψ**G**G**Ψ**C**C**C**C**A** C**A**G**A**C**Ψ**C**A**G**A** G**A**G**A**A**C**C**C**G**C** 50
CA**C**C**A**Ψ**G**Ψ**Ψ****C** **G**Ψ**G**Ψ**Ψ**C**Ψ**G**G** **Ψ**G**C**Ψ**G**C**Ψ**G**C** **Ψ**C**Ψ**G**G**Ψ**G**Ψ**C** **A**G**C**C**A**G**Ψ**G**Ψ**G 100
ΨG**A**A**C**C**Ψ**G**A**C **C**A**C**C**A**G**A**A**C**A **C**A**G**C**Ψ**G**C**C**Ψ**C **C**A**G**C**C**Ψ**A**C**A**C **C**A**A**C**A**G**C**Ψ**Ψ**Ψ 150
AC**C**A**G**A**G**G**C**G **Ψ**G**Ψ**A**C**Ψ**A**C**C**C **C**G**A**C**A**A**G**G**Ψ**G **Ψ**Ψ**C**A**G**A**Ψ**C**C**A **G**C**G**Ψ**G**C**Ψ**G**C**A 200
CΨ**C**Ψ**A**C**C**C**A**G **G**A**C**C**Ψ**G**Ψ**Ψ**C** **Ψ**G**C**C**Ψ**Ψ**Ψ**C**Ψ**Ψ **C**A**G**C**A**A**C**G**Ψ**G **A**C**C**Ψ**G**G**Ψ**Ψ**C** 250
AC**G**C**C**A**Ψ**C**C**A **C**G**Ψ**G**Ψ**C**C**G**G**C **A**C**C**A**A**Ψ**G**G**C**A **C**C**A**A**G**A**G**A**Ψ**Ψ **C**G**A**C**A**A**C**C**C** 300
GΨ**G**C**Ψ**G**C**C**C**Ψ **Ψ**C**A**A**C**G**A**C**G**G **G**G**Ψ**G**Ψ**A**C**Ψ**Ψ**Ψ **G**C**C**A**G**C**A**C**C**G **A**G**A**A**G**Ψ**C**C**A**A 350
CA**Ψ**C**A**Ψ**C**A**G**A **G**G**C**Ψ**G**G**A**Ψ**C**Ψ **Ψ**C**G**G**C**A**C**C**A**C **A**C**Ψ**G**G**A**C**A**G**C **A**A**G**A**C**C**C**A**G**A 400
GC**C**Ψ**G**C**Ψ**G**A**Ψ **C**G**Ψ**G**A**A**C**A**A**C **G**C**C**A**C**C**A**A**C**G **Ψ**G**G**Ψ**C**A**Ψ**C**A**A **A**G**Ψ**G**Ψ**G**C**G**A**G 450
ΨΨ**C**C**A**G**Ψ**Ψ**C**Ψ **G**C**A**A**C**G**A**C**C**C **C**Ψ**Ψ**C**C**Ψ**G**G**G**C **G**Ψ**C**Ψ**A**C**Ψ**A**C**C **A**C**A**A**G**A**A**C**A**A 500
CA**A**A**G**A**C**Ψ**G**G **A**Ψ**G**G**A**A**A**G**C**G **A**G**Ψ**Ψ**C**C**G**G**G**Ψ **G**Ψ**A**C**A**G**C**A**G**C **G**C**C**A**A**C**A**A**C**Ψ 550
GC**A**C**C**Ψ**Ψ**C**G**A **G**Ψ**A**C**G**Ψ**G**Ψ**C** **C**A**G**C**C**Ψ**Ψ**Ψ**C** **Ψ**G**A**Ψ**G**G**A**C**C**Ψ **G**G**A**A**G**G**C**A**A**G 600
CA**G**G**G**C**A**A**C**Ψ **Ψ**C**A**A**G**A**A**C**C**Ψ **G**C**G**C**G**A**G**Ψ**Ψ**C **G**Ψ**G**Ψ**Ψ**A**A**A**G**A **A**C**A**Ψ**C**G**A**C**G**G 650
CΨ**A**C**Ψ**Ψ**C**A**A**G **A**Ψ**C**Ψ**A**C**A**G**C**A **A**G**C**A**C**A**C**C**C** **Ψ**A**Ψ**C**A**A**C**C**Ψ**C **G**Ψ**G**C**G**G**G**A**Ψ**C 700
ΨG**C**C**Ψ**C**A**G**G**G **C**Ψ**Ψ**C**Ψ**C**Ψ**G**C**Ψ **C**Ψ**G**G**A**A**C**C**C** **Ψ**G**G**Ψ**G**G**A**Ψ**C**Ψ **G**C**C**C**A**Ψ**C**G**G**C 750
AΨ**C**A**A**C**A**Ψ**C**A **C**C**C**G**G**Ψ**Ψ**Ψ**C**A **G**A**C**A**C**Ψ**G**C**Ψ**G **G**C**C**C**Ψ**G**C**A**C**A **G**A**A**G**C**Ψ**A**C**C**Ψ 800
GA**C**A**C**Ψ**G**G**C** **G**A**Ψ**A**G**C**A**G**C**A **G**C**G**G**A**Ψ**G**G**A**C **A**G**C**Ψ**G**G**Ψ**G**C** **G**C**C**G**C**Ψ**Ψ**A**C**Ψ 850
AΨ**G**Ψ**G**G**G**C**Ψ**A **C**C**Ψ**G**C**A**G**C**C**Ψ **A**G**A**A**C**C**Ψ**Ψ**C** **Ψ**G**C**Ψ**G**A**A**G**Ψ**A **C**A**A**C**G**A**G**A**A**C 900
GG**C**A**C**C**A**Ψ**C**A **C**C**G**A**C**G**C**G**Ψ** **G**G**A**Ψ**Ψ**G**Ψ**G**C**Ψ **C**Ψ**G**G**A**Ψ**C**C**Ψ**C **Ψ**G**A**G**C**G**A**G**A**C 950
AA**A**G**Ψ**G**C**A**C**C **C**Ψ**G**A**A**G**Ψ**C**C**Ψ **Ψ**C**A**C**C**G**Ψ**G**G**A **A**A**A**G**G**G**C**A**Ψ**C **Ψ**A**C**C**A**G**A**C**C**A 1000
GC**A**A**C**Ψ**Ψ**C**C**G **G**G**Ψ**G**C**A**G**C**C**C **A**C**C**G**A**A**Ψ**C**C**A **Ψ**C**G**Ψ**G**C**G**G**Ψ**Ψ **C**C**C**C**A**A**Ψ**A**Ψ**C 1050
AC**C**A**A**Ψ**C**Ψ**G**Ψ **G**C**C**C**Ψ**Ψ**C**G**G** **C**G**A**G**G**Ψ**G**Ψ**Ψ**C **A**A**Ψ**G**C**C**A**C**C**A **G**A**Ψ**Ψ**C**G**C**C**Ψ**C 1100



ΨGΨGΨACGCC	ΨGGAACCGGA	AGCGGAYCAG	CAAΨΨGCGΨG	GCCGACΨACΨ	1150
CCGΨGCΨGΨA	CAACΨCCGCC	AGCΨΨCAGCA	CCΨΨCAAGΨG	CΨACGGCGΨG	1200
ΨCCCCΨACCA	AGCΨGAACGA	CCΨGΨGCΨΨC	ACAAACGΨGΨ	ACGCCGACAG	1250
CΨΨCGΨGAΨC	CGGGGAGAYG	AAGΨGCGGCA	GAΨΨGCCCCΨ	GGACAGACAG	1300
GCAAGAYCAGC	CGACΨACAAC	ΨACAAGCΨGC	CCGACGACΨΨ	CACCGGCΨGΨ	1350
GΨGAΨΨGCCΨ	GGAACAGCAA	CAACCΨGGAC	ΨCCAAAGΨCG	GCGGCAACΨA	1400
CAAΨΨACCΨG	ΨACCGGCΨGΨ	ΨCCGGAAGΨC	CAAΨCΨGAAG	CCCΨΨCAGAC	1450
GGGACAYCΨC	CACCGAGAYC	ΨAYCAGGCCG	GCAGCACCCC	ΨΨGΨAACGGC	1500
GΨGGAAGGCΨ	ΨCAACΨGCΨA	CΨΨCCCACΨG	CAGΨCCΨACG	GCΨΨΨCAGCC	1550
CACAAAΨGGC	GΨGGGΨAYΨC	AGCCCΨACAG	AGΨGGΨGGΨG	CΨGAGCΨΨCG	1600
AACΨGCΨGCA	ΨGCCCCΨGCG	ACAGΨGΨGCG	GCCCΨAAGAA	AAGCACCAAY	1650
CΨCGΨGAAGA	ACAAAΨGCGΨ	GAACΨΨCAAC	ΨΨCAACGGCC	ΨGACCGGCAC	1700
CGGCΨGCGΨG	ACAGAGAGCA	ACAAGAAGΨΨ	CCΨGCCAYΨC	CAGCAGΨΨΨG	1750
GCCGGGAYAY	CGCCGAYACC	ACAGACGCCG	ΨΨAGAGAYCC	CCAGACACΨG	1800
GAAAΨCCΨGG	ACAΨCACCCC	ΨΨGCAGCΨΨC	GGCGGAGΨGΨ	CΨGΨGAΨCAC	1850
CCCΨGGCACC	AACACCAGCA	AΨCAGGΨGGC	AGΨGCΨGΨAC	CAGGACGΨGA	1900
ACΨGΨACCGA	AGΨGCCCGΨG	GCCAΨΨCACG	CCGAYCAGCΨ	GACACCΨACA	1950
ΨGGCGGGΨGΨ	ACΨCCACCGG	CAGCAAΨGΨG	ΨΨΨCAGACCA	GAGCCGGCΨG	2000
ΨCΨGAYCAGGA	GCCGAGCACG	ΨGAACAAΨAG	CΨACGAGΨGC	GACAΨCCCCA	2050
ΨCGGCGCΨGG	AAΨCΨGCGCC	AGCΨACCAGA	CACAGACAAA	CAGCCCΨCGG	2100
AGAGCCAGAA	GCGΨGGCCAG	CCAGAGCAYC	AΨΨGCCΨACA	CAAΨGΨCΨCΨ	2150
GGGCGCCGAG	AACAGCΨGG	CCΨACΨCCAA	CAACΨCΨAYC	GCΨAYCCCCA	2200
CCAACΨΨCAC	CAΨCAGCGΨG	ACCACAGAGA	ΨCCΨGCCΨGΨ	GΨCCAΨGACC	2250
AAGACCAGCG	ΨGGACΨGCAC	CAΨGΨACAYC	ΨGCGGCGAYΨ	CCACCAGAYG	2300
CΨCCAACCΨG	CΨGCΨGCAGΨ	ACGGCAGCΨΨ	CΨGCACCCAG	CΨGAAΨAGAG	2350
CCCΨGACAGG	GAΨCGCCGΨG	GAACAGGACA	AGAACACCCA	AGAGGΨGΨΨC	2400
GCCCAAGΨGA	AGCAGAYCΨA	CAAGACCCCΨ	CCΨAYCAAGG	ACΨΨCGGGCG	2450
CΨΨCAAΨΨΨC	AGCCAGAYΨC	ΨGCCCGAYCC	ΨAGCAAGCCC	AGCAAGCGGA	2500
GCΨΨCAYCAG	GGACCΨGCGΨG	ΨΨCAACAAAG	ΨGACACΨGGC	CGACGCCGGC	2550
ΨΨCAYCAAGC	AGΨAYGGCGA	ΨΨGΨCΨGGGC	GACAΨΨGCCG	CCAGGGAYCΨ	2600
GAΨΨΨGCGCC	CAGAAGΨΨΨA	ACGGACΨGAC	AGΨGCΨGCCΨ	CCΨCΨGCΨGA	2650
CCGAYGAGAY	GAΨCGCCAG	ΨACACAYCΨG	CCCΨGCΨGGC	CGGCACAAΨC	2700
ACAAGCGGCΨ	GGACAYΨΨGG	AGCAGGCGCC	GCΨCΨGCAGA	ΨCCCCΨΨΨG	2750
ΨAYGCAGAYG	GCCΨACCGGΨ	ΨCAACGGCAΨ	CGGAGΨGACC	CAGAAΨGΨGC	2800
ΨGΨACGAGAA	CCAGAAGCΨG	AΨCGCCAACC	AGΨΨCAACAG	CGCCAΨCGGC	2850
AAGAYCCAGG	ACAGCCΨGAG	CAGCACAGCA	AGCGCCCΨGG	GAAAGCΨGCA	2900
GGACGΨGGΨC	AACCAGAAΨG	CCCAGGCACΨ	GAACACCCΨG	GΨCAAGCAGC	2950
ΨGΨCCΨCCAA	CΨΨCGGCGCC	AΨCAGCΨCΨG	ΨGCΨGAACGA	ΨAYCCΨGAGC	3000
AGACΨGGACC	CΨCCΨGAGGC	CGAGGΨGCAG	AΨCGACAGAC	ΨGAYCACAGG	3050
CAGACΨGCAG	AGCCΨCCAGA	CAΨACGΨGAC	CCAGCAGCΨG	AΨCAGAGCCG	3100
CCGAGAYΨAG	AGCCΨCΨGCC	AAΨCΨGGCCG	CCACCAAGAY	GΨCΨGAGΨGΨ	3150
GΨGCΨGGGCC	AGAGCAAGAG	AGΨGGACΨΨΨ	ΨGCGGCAAGG	GCΨACCACCΨ	3200
GAΨGAGCΨΨC	CCΨCAGΨCΨG	CCCCΨCACGG	CGΨGGΨGΨΨΨ	CΨGCACGΨGA	3250
CAΨAYGΨGGC	CGCΨCAAGAG	AAGAAΨΨΨCA	CCACCGCΨCC	AGCCAΨCΨGC	3300
CACGACGGCA	AAGCCCACΨΨ	ΨCCΨAGAGAA	GGCGΨGΨΨCG	ΨGΨCCAACGG	3350
CACCCAΨΨGG	ΨΨCGΨGACAC	AGCGGAACΨΨ	CΨACGAGCCC	CAGAYCAYΨCA	3400
CCACCAGCAA	CACCΨΨCGΨG	ΨCΨGGCAACΨ	GCGACGΨCΨG	GAΨCGGCAΨΨ	3450
GΨGAACAAΨA	CCGΨGΨACGA	CCCΨCΨGCAG	CCCAGCΨGG	ACAGCΨΨCAA	3500
AGAGGAACΨG	GACAAGΨAYC	ΨΨAAGAACCA	CACAAGCCCC	GACGΨGGACC	3550
ΨGGGCGAYAY	CAGCGGAAΨC	AAΨGCCAGCG	ΨCGΨGAACAY	CCAGAAAGAG	3600
AΨCGACCGGC	ΨGAACGAGGΨ	GGCCAAGAAΨ	CΨGAACGAGA	GCCΨGAYCAG	3650
CCΨGCAAGAA	CΨGGGGAAGΨ	ACGAGCAGΨA	CAΨCAAGΨGG	CCCΨGGΨACA	3700



ΨCΨGGCΨGGG	CΨΨAΨCGCC	GGACΨGAΨΨG	CCAΨCGΨGAΨ	GGΨCACAAΨC	3750
AΨGCΨGΨGΨΨ	GCAΨGACCAG	CΨGCΨGΨAGC	ΨGCCΨGAAGG	GCΨGΨΨGΨAG	3800
CΨGΨGGCAGC	ΨGCΨGCAAGΨ	ΨCGACGAGGA	CGAΨΨCΨGAG	CCCGΨGCΨGA	3850
AGGGCGΨGAA	ACΨGCACΨAC	ACAΨGAΨGAC	ΨCGAGCΨGGΨ	ACΨGCAΨGCA	3900
CGCAAΨGCΨA	GCΨGCCCCΨΨ	ΨCCCGΨCCΨG	GGΨACCCCGA	GΨCΨCCCCCG	3950
ACCΨCGGGΨC	CCAGGΨAΨGC	ΨCCCACCCΨCC	ACCΨGCCCCA	CΨCACCACCΨ	4000
CΨGCΨAGΨΨC	CAGACACCΨC	CCAAGCACGC	AGCAAΨGCAG	CΨCAAAAACGC	4050
ΨΨAGCCΨAGC	CACACCCCA	CGGGAAACAG	CAGΨGAΨΨAA	CCΨΨΨAGCAA	4100
ΨAAACGAAAG	ΨΨΨAACΨAAG	CΨAΨACΨAAC	CCCAGGGΨΨG	GΨCAAΨΨΨCG	4150
ΨGCCAGCCAC	ACCCΨGGAGC	ΨAGCAAAAAA	AAAAAAAAAAAA	AAAAAAAAAAAA	4200
AAAAGCAΨAΨ	GACΨAAAAAA	AAAAAAAAAAAA	AAAAAAAAAAAA	AAAAAAAAAAAA	4250
AAAAAAAAAAAA	AAAAAAAAAAAA	AAAAAAAAAAAA	AAAA		4284

Ψ = 1-methyl-3'-pseudouridylyl