

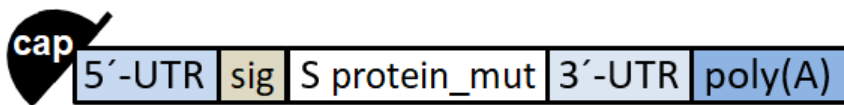


11889

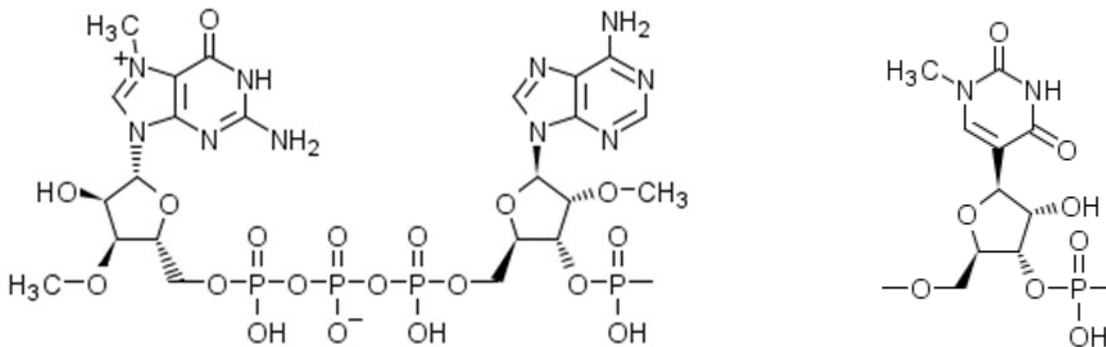
Description

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 G¹A² = 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	55-102



	into the endoplasmic reticulum.	
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

GAGAAΨAAAC ΨAGΨAΨΨCΨΨ CΨGGΨCCCCA CAGACΨCAGA GAGAACCCGC 50
CACCAΨGΨΨC **G**ΨGΨΨCCΨGG ΨGCΨGCΨGCC ΨCΨGGΨGΨCC **A**GCCAGΨGΨG 100
ΨGAACCΨGAC **C**ACCAGAACA **C**AGCΨGCCΨC **C**AGCCΨACAC **C**AACAGCΨΨΨ 150
ACCAGAGGCG ΨGΨACΨACCC **C**GACAAGGΨG ΨΨCAGAΨCCA **G**CGΨGCΨGCA 200
CΨCΨACCCAG **G**ACCΨGΨΨC ΨGCCΨΨCΨΨ **C**AGCAACGΨG **A**CCΨGGΨΨCC 250
ACGCCAΨCCA **C**GΨGΨCCGGC **A**CCAAΨGGCA **C**CAAGAGAΨΨ **C**GACAACCCC 300
GΨGCΨGCCCΨ ΨCAACGACGG **G**ΨGΨACΨΨΨ **G**CCAGCACCG **A**GAAGΨCCAA 350
CAΨCAΨCAGA **G**GCΨGGAΨCΨ ΨC**G**GCACCAC **A**CΨGGACAGC **A**AGACCCAGA 400
GCCΨGCΨGAP **C**GΨGAACAAC **G**CCACCAACG ΨGGΨCAΨCAA **A**GΨGΨGCGAG 450
ΨΨCCAGΨΨCΨ **G**CAACGACCC **C**ΨΨCCΨGGGC **G**ΨCΨACΨACC **A**CAAGAACAA 500
CAAGAGCΨGG **A**ΨGGAAAGCG **A**GΨΨCCGGGΨ **G**ΨACAGCAGC **G**CCAACAACΨ 550
GCACCΨΨCGA **G**ΨACGΨGΨCC **C**AGCCΨΨΨC ΨGAΨGGACCΨ **G**GAAGGCAAG 600
CAGGGCAACΨ ΨCAAGAACCΨ **G**CGCGAGΨΨC **G**ΨGΨΨAAGA **A**CAΨCGACGG 650
CΨACΨΨCAAG **A**ΨCΨACAGCA **A**GCACACCC **Ψ**AΨCAACCΨC **G**ΨGCGGGAPC 700
ΨGCCΨCAGGG **C**ΨΨCΨCΨGCΨ **C**ΨGGAACCC **Ψ**GGΨGGAΨCΨ **G**CCCAΨCGGC 750
AΨCAACAΨCA **C**CCGGΨΨΨCA **G**ACACΨGCΨG **G**CCCΨGCACA **G**AAGCΨACCΨ 800
GACACCΨGGC **G**AΨAGCAGCA **G**CGGAΨGGAC **A**GCΨGGΨGCC **G**CCGCΨΨACΨ 850
AΨGΨGGGCPA **C**CΨGCAGCCΨ **A**GAACCΨΨC **Ψ**GCΨGAAGΨA **C**AACGAGAAC 900
GGCACCAΨCA **C**CGACGCCGΨ **G**GAΨGΨGCPΨ **C**ΨGGAΨCCΨC **Ψ**GAGCGAGAC 950
AAAGΨGCACC **C**ΨGAAGΨCCΨ ΨC**A**CCGΨGGA **A**AAGGGCAΨC **Ψ**ACCAGACCA 1000
GCAACΨΨCCG **G**GΨGCAGCCC **A**CCGAAPCCA **Ψ**CGΨGCGGΨΨ **C**CCCAAPAPC 1050
ACCAAΨCΨGΨ **G**CCCCΨΨCGG **C**GAGGΨGΨΨC **A**AΨGCCACCA **G**AΨΨCGCCΨC 1100
ΨGΨGΨACGCC **Ψ**GGAACCGGA **A**GCGGAΨCAG **C**AAΨΨGCGΨG **G**CCGACΨACΨ 1150
CCGΨGCΨGΨA **C**AACΨCCGCC **A**GCΨΨCAGCA **C**CΨΨCAAGΨG **C**ΨACGGCGΨG 1200
ΨCCCCΨACCA **A**GCΨGAACGA **C**CΨGΨGCΨΨC **A**CAAACGΨGΨ **A**CGCCGACAG 1250
CΨΨCGΨGAPC **C**GGGGAGAPG **A**AGΨGCGGCA **G**AΨΨGCCCCΨ **G**GACAGACAG 1300
GCAAGAPCGC **C**GACΨACAAC **Ψ**ACAAGCΨGC **C**CGACGACΨΨ **C**ACCGGCΨGΨ 1350
GΨGAΨΨGCCΨ **G**GAACAGCAA **C**AACCΨGGAC **Ψ**CCAAAGΨCG **G**CGGCAACΨA 1400
CAAΨΨACCΨG **Ψ**ACCGGCΨGΨ **Ψ**CCGGAAGΨC **C**AAΨCΨGAAG **C**CCΨΨCGAGC 1450
GGGACAΨCΨC **C**ACCGAGAΨC **Ψ**AΨCAGGCCG **G**CAGCACCCC **Ψ**ΨGΨAACGGC 1500



GΨGGAAGGCΨ	ΨCAACΨGCΨA	CΨΨCCCACΨG	CAGΨCCΨACG	GCΨΨΨCAGCC	1550
CACAAAΨGGC	GΨGGGCΨAΨC	AGCCCΨACAG	AGΨGGΨGGΨG	CΨGAGCΨΨCG	1600
AACΨGCΨGCA	ΨGCCCCΨGCC	ACAGΨGΨGCG	GCCCΨAAGAA	AAGCACCAAΨ	1650
CΨCGΨGAAGA	ACAAAΨGCGΨ	GAACΨΨCAAC	ΨΨCAACGGCC	ΨGACCGGCAC	1700
CGGCΨGΨCΨG	ACAGAGAGCA	ACAAGAAGΨΨ	CCΨGCCAΨΨC	CAGCAGΨΨΨG	1750
GCCGGGAΨAΨ	CGCCGAΨACC	ACAGACGCCG	ΨΨAGAGAΨCC	CCAGACACΨG	1800
GAAAΨCCΨGG	ACAΨCACCCC	ΨΨGCGAGCΨΨC	GGCGGAGΨGΨ	CΨGΨGAΨCAC	1850
CCCΨGGCACC	AACACCAGCA	AΨCAGGΨGGC	AGΨGCΨGΨAC	CAGGACGΨGA	1900
ACΨGΨACCGA	AGΨGCCCGΨG	GCCAΨΨCACG	CCGAΨCAGCΨ	GACACCΨACA	1950
ΨGGCGGGΨGΨ	ACΨCCACCGG	CAGCAAΨGΨG	ΨΨΨCAGACCA	GAGCCGGCΨG	2000
ΨCΨGAΨCGGA	GCCGAGCACG	ΨGAACAAΨAG	CΨACGAGΨGC	GACAΨCCCCA	2050
ΨCGGCGCΨGG	AAΨCΨGCGCC	AGCΨACCAGA	CACAGACAAA	CAGCCCΨCGG	2100
AGAGCCAGAA	GCGΨGGCCAG	CCAGAGCAΨC	AΨΨGCCΨACA	CAAΨGΨCΨCΨ	2150
GGGCGCCGAG	AACAGCΨGG	CCΨACΨCCAA	CAACΨCΨAΨC	GCΨAΨCCCCA	2200
CCAACΨΨCAC	CAΨCAGCGΨG	ACCACAGAGA	ΨCCΨGCCΨGΨ	GΨCCAΨGACC	2250
AAGACCAGCG	ΨGGACΨGCAC	CAΨGΨACAΨC	ΨGCGGCGAΨΨ	CCACCAGΨΨG	2300
CΨCCAACCΨG	CΨGCΨGCAGΨ	ACGGCAGCΨΨ	CΨGCACCCAG	CΨGAAΨAGAG	2350
CCCΨGACAGG	GAΨCGCCGΨG	GAACAGGACA	AGAACACCCA	AGAGGΨGΨΨC	2400
GCCCAAGΨGA	AGCAGAΨCΨA	CAAGACCCCΨ	CCΨAΨCAAGG	ACΨΨCGGCGG	2450
CΨΨCAAΨΨΨC	AGCCAGAΨΨC	ΨGCCCGAΨCC	ΨAGCAAGCCC	AGCAAGCGGA	2500
GCΨΨCAΨCGA	GGACCΨGCΨG	ΨΨCAACAAAG	ΨGACACΨGGC	CGACGCCGGC	2550
ΨΨCAΨCAAGC	AGΨAΨGGCGA	ΨΨGΨCΨGGGC	GACAΨΨGCCG	CCAGGGAΨCΨ	2600
GAΨΨΨGCGCC	CAGAAGΨΨΨA	ACGGACΨGAC	AGΨGCΨGCCΨ	CCΨCΨGCΨGA	2650
CCGAΨGAGAΨ	GAΨCGCCCAG	ΨACACAΨCΨG	CCCΨGCΨGGC	CGGCACAAΨC	2700
ACAAGCGGCΨ	GGACAΨΨΨGG	AGCAGGCGCC	GCΨCΨGCAGA	ΨCCCCΨΨΨGC	2750
ΨAΨGCAGAΨG	GCCΨACCGGΨ	ΨCAACGGCAΨ	CGGAGΨGACC	CAGAAΨGΨGC	2800
ΨGΨACGAGAA	CCAGAAGCΨG	AΨCGCCAACC	AGΨΨCAACAG	CGCCAΨCGGC	2850
AAGAΨCCAGG	ACAGCCΨGAG	CAGCACAGCA	AGCGCCCΨGG	GAAAGCΨGCA	2900
GGACGΨGGΨC	AACCAGAAΨG	CCCAGGCACΨ	GAACACCCΨG	GΨCAAGCAGC	2950
ΨGΨCCΨCCAA	CΨΨCGGCGCC	AΨCAGCΨCΨG	ΨGCΨGAACGA	ΨAΨCCΨGAGC	3000
AGACΨGGACC	CΨCCΨGAGGC	CGAGGΨGCAG	AΨCGACAGAC	ΨGAΨCACAGG	3050
CAGACΨGCAG	AGCCΨCCAGA	CAΨACGΨGAC	CCAGCAGCΨG	AΨCAGAGCCG	3100
CCGAGAΨΨAG	AGCCΨCΨGCC	AAΨCΨGGCCG	CCACCAAGAΨ	GΨCΨGAGΨGΨ	3150
GΨGCΨGGGCC	AGAGCAAGAG	AGΨGGACΨΨΨ	ΨGCGGCAAGG	GCΨACCACCΨ	3200
GAΨGAGCΨΨC	CCΨCAGΨCΨG	CCCCΨCACGG	CGΨGGΨGΨΨΨ	CΨGCACGΨGA	3250
CAΨAΨGΨGCC	CGCΨCAAGAG	AAGAΨΨΨΨCA	CCACCGCΨCC	AGCCAΨCΨGC	3300
CACGACGGCA	AAGCCACΨΨ	ΨCCΨAGAGAA	GGCGΨGΨΨCG	ΨGΨCCAACGG	3350
CACCCAΨΨGG	ΨΨCGΨGACAC	AGCGGAACΨΨ	CΨACGAGCCC	CAGAΨCAΨCA	3400
CCACCAGCAA	CACCΨΨCGΨG	ΨCΨGGCAACΨ	GCGACGΨCΨΨ	GAΨCGGCAΨΨ	3450
GΨGAACAAΨA	CCGΨGΨACGA	CCCΨCΨGCAG	CCCAGCΨGG	ACAGCΨΨCAA	3500
AGAGGAACΨG	GACAAGΨACΨ	ΨΨAAGAACCA	CACAAGCCCC	GACGΨGGACC	3550
ΨGGGCGAΨAΨ	CAGCGGAAΨC	AAΨGCCAGCG	ΨCGΨGAACAΨ	CCAGAAAGAG	3600
AΨCGACCGGC	ΨGAACGAGGΨ	GGCCAAGAAΨ	CΨGAACGAGA	GCCΨGAΨCGA	3650
CCΨGCAAGAA	CΨGGGGAAAGΨ	ACGAGCAGΨA	CAΨCAAGΨGG	CCCΨGGΨACA	3700
ΨCΨGGCΨGGG	CΨΨΨAΨCGCC	GGACΨGAΨΨG	CCAΨCGΨGAΨ	GGΨCACAAΨC	3750
AΨGCΨGΨGΨΨ	GCAΨGACCAG	CΨGCΨGΨAGC	ΨGCCΨGAAGG	GCΨGΨΨGΨAG	3800
CΨGΨGGCAGC	ΨGCΨGCAAGΨ	ΨCGACGAGGA	CGAΨΨCΨGAG	CCCΨGΨCΨGA	3850
AGGGCGΨGAA	ACΨGCACΨAC	ACAΨGAΨGAC	ΨCGAGCΨGGΨ	ACΨGCAΨGCA	3900
CGCAAΨGCΨA	GCΨGCCCCΨΨ	ΨCCCCΨCCΨG	GGΨACCCCGA	GΨCΨCCCCCG	3950
ACCΨCGGGΨC	CCAGGΨAΨGC	ΨCCCACCΨCC	ACCΨGCCCCA	CΨCACCACCΨ	4000
CΨGCΨAGΨΨC	CAGACACCΨC	CCAAGCACGC	AGCAAΨGCAG	CΨCAAAAACGC	4050
ΨΨAGCCΨAGC	CACACCCCCA	CGGGAAACAG	CAGΨGAΨΨAA	CCΨΨΨAGCAA	4100



WHO
International Nonproprietary Names Programme

9/2020

ΨAAACGAAAG	ΨΨΨAACΨAAG	CΨAΨACΨAAC	CCCAGGGΨΨG	GΨCAAΨΨΨCG	4150
ΨGCCAGCCAC	ACCCΨGGAGC	ΨAGCAAAAAA	AAAAAAAAAA	AAAAAAAAAA	4200
AAAAGCAΨAΨ	GACΨAAAAAA	AAAAAAAAAA	AAAAAAAAAA	AAAAAAAAAA	4250
AAAAAAAAAA	AAAAAAAAAA	AAAAAAAAAA	AAAA		4284

Ψ = 1-methyl-3'-pseudouridylyl