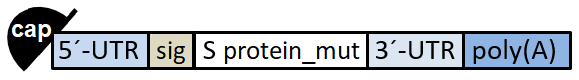
***11889***

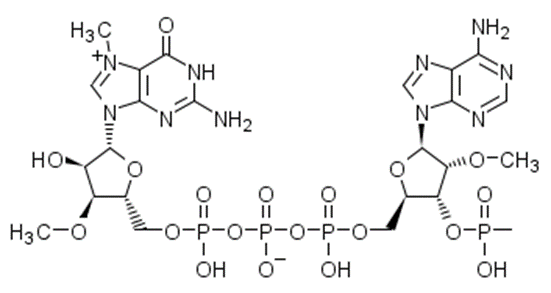
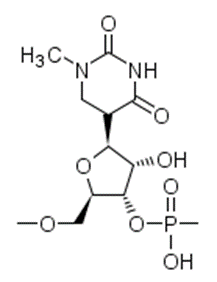
**Description**

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 (m7G+m3'-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. | 55-102 |
| 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**GAAΨAAAC ΨAGΨAΨΨCΨΨ CΨGGΨCCCCA CAGACΨCAGA GAGAACCCGC 50

CACC**AΨGΨΨC GΨGΨΨCCΨGG ΨGCΨGCΨGCC ΨCΨGGΨGΨCC AGCCAGΨGΨG** 100

**ΨGAACCΨGAC CACCAGAACA CAGCΨGCCΨC CAGCCΨACAC CAACAGCΨΨΨ** 150

**ACCAGAGGCG ΨGΨACΨACCC CGACAAGGΨG ΨΨCAGAΨCCA GCGΨGCΨGCA**  200

**CΨCΨACCCAG GACCΨGΨΨCC ΨGCCΨΨΨCΨΨ CAGCAACGΨG ACCΨGGΨΨCC**  250

**ACGCCAΨCCA CGΨGΨCCGGC ACCAAΨGGCA CCAAGAGAΨΨ CGACAACCCC**  300

**GΨGCΨGCCCΨ ΨCAACGACGG GGΨGΨACΨΨΨ GCCAGCACCG AGAAGΨCCAA**  350

**CAΨCAΨCAGA GGCΨGGAΨCΨ ΨCGGCACCAC ACΨGGACAGC AAGACCCAGA**  400

**GCCΨGCΨGAΨ CGΨGAACAAC GCCACCAACG ΨGGΨCAΨCAA AGΨGΨGCGAG** 450

**ΨΨCCAGΨΨCΨ GCAACGACCC CΨΨCCΨGGGC GΨCΨACΨACC ACAAGAACAA** 500

**CAAGAGCΨGG AΨGGAAAGCG AGΨΨCCGGGΨ GΨACAGCAGC GCCAACAACΨ** 550

**GCACCΨΨCGA GΨACGΨGΨCC CAGCCΨΨΨCC ΨGAΨGGACCΨ GGAAGGCAAG** 600

**CAGGGCAACΨ ΨCAAGAACCΨ GCGCGAGΨΨC GΨGΨΨΨAAGA ACAΨCGACGG** 650

**CΨACΨΨCAAG AΨCΨACAGCA AGCACACCCC ΨAΨCAACCΨC GΨGCGGGAΨC** 700

**ΨGCCΨCAGGG CΨΨCΨCΨGCΨ CΨGGAACCCC ΨGGΨGGAΨCΨ GCCCAΨCGGC** 750

**AΨCAACAΨCA CCCGGΨΨΨCA GACACΨGCΨG GCCCΨGCACA GAAGCΨACCΨ** 800

**GACACCΨGGC GAΨAGCAGCA GCGGAΨGGAC AGCΨGGΨGCC GCCGCΨΨACΨ** 850

**AΨGΨGGGCΨA CCΨGCAGCCΨ AGAACCΨΨCC ΨGCΨGAAGΨA CAACGAGAAC** 900

**GGCACCAΨCA CCGACGCCGΨ GGAΨΨGΨGCΨ CΨGGAΨCCΨC ΨGAGCGAGAC** 950

**AAAGΨGCACC CΨGAAGΨCCΨ ΨCACCGΨGGA AAAGGGCAΨC ΨACCAGACCA** 1000

**GCAACΨΨCCG GGΨGCAGCCC ACCGAAΨCCA ΨCGΨGCGGΨΨ CCCCAAΨAΨC** 1050

**ACCAAΨCΨGΨ GCCCCΨΨCGG CGAGGΨGΨΨC AAΨGCCACCA GAΨΨCGCCΨC** 1100

**ΨGΨGΨACGCC ΨGGAACCGGA AGCGGAΨCAG CAAΨΨGCGΨG GCCGACΨACΨ** 1150

**CCGΨGCΨGΨA CAACΨCCGCC AGCΨΨCAGCA CCΨΨCAAGΨG CΨACGGCGΨG** 1200

**ΨCCCCΨACCA AGCΨGAACGA CCΨGΨGCΨΨC ACAAACGΨGΨ ACGCCGACAG** 1250

**CΨΨCGΨGAΨC CGGGGAGAΨG AAGΨGCGGCA GAΨΨGCCCCΨ GGACAGACAG** 1300

**GCAAGAΨCGC CGACΨACAAC ΨACAAGCΨGC CCGACGACΨΨ CACCGGCΨGΨ** 1350

**GΨGAΨΨGCCΨ GGAACAGCAA CAACCΨGGAC ΨCCAAAGΨCG GCGGCAACΨA** 1400

**CAAΨΨACCΨG ΨACCGGCΨGΨ ΨCCGGAAGΨC CAAΨCΨGAAG CCCΨΨCGAGC** 1450

**GGGACAΨCΨC CACCGAGAΨC ΨAΨCAGGCCG GCAGCACCCC ΨΨGΨAACGGC** 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

**CGGCGΨGCΨG ACAGAGAGCA ACAAGAAGΨΨ CCΨGCCAΨΨC CAGCAGΨΨΨG** 1750

**GCCGGGAΨAΨ CGCCGAΨACC ACAGACGCCG ΨΨAGAGAΨCC CCAGACACΨG** 1800

**GAAAΨCCΨGG ACAΨCACCCC ΨΨGCAGCΨΨC 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 AACAGCGΨGG CCΨACΨCCAA CAACΨCΨAΨC GCΨAΨCCCCA** 2200

**CCAACΨΨCAC CAΨCAGCGΨG ACCACAGAGA ΨCCΨGCCΨGΨ GΨCCAΨGACC** 2250

**AAGACCAGCG ΨGGACΨGCAC CAΨGΨACAΨC ΨGCGGCGAΨΨ CCACCGAGΨG** 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 AAGAAΨΨΨCA CCACCGCΨCC AGCCAΨCΨGC** 3300

**CACGACGGCA AAGCCCACΨΨ ΨCCΨAGAGAA GGCGΨGΨΨCG ΨGΨCCAACGG** 3350

**CACCCAΨΨGG ΨΨCGΨGACAC AGCGGAACΨΨ CΨACGAGCCC CAGAΨCAΨCA** 3400

**CCACCGACAA CACCΨΨCGΨG ΨCΨGGCAACΨ GCGACGΨCGΨ GAΨCGGCAΨΨ** 3450

**GΨGAACAAΨA CCGΨGΨACGA CCCΨCΨGCAG CCCGAGCΨGG ACAGCΨΨCAA** 3500

**AGAGGAACΨG GACAAGΨACΨ ΨΨAAGAACCA CACAAGCCCC GACGΨGGACC** 3550

**ΨGGGCGAΨAΨ CAGCGGAAΨC AAΨGCCAGCG ΨCGΨGAACAΨ CCAGAAAGAG** 3600

**AΨCGACCGGC ΨGAACGAGGΨ GGCCAAGAAΨ CΨGAACGAGA GCCΨGAΨCGA** 3650

**CCΨGCAAGAA CΨGGGGAAGΨ ACGAGCAGΨA CAΨCAAGΨGG CCCΨGGΨACA** 3700

**ΨCΨGGCΨGGG CΨΨΨAΨCGCC GGACΨGAΨΨG CCAΨCGΨGAΨ GGΨCACAAΨC** 3750

**AΨGCΨGΨGΨΨ GCAΨGACCAG CΨGCΨGΨAGC ΨGCCΨGAAGG GCΨGΨΨGΨAG** 3800

**CΨGΨGGCAGC ΨGCΨGCAAGΨ ΨCGACGAGGA CGAΨΨCΨGAG CCCGΨGCΨGA** 3850

**AGGGCGΨGAA ACΨGCACΨAC ACAΨGAΨGA**C ΨCGAGCΨGGΨ ACΨGCAΨGCA 3900

CGCAAΨGCΨA GCΨGCCCCΨΨ ΨCCCGΨCCΨG GGΨACCCCGA GΨCΨCCCCCG 3950

ACCΨCGGGΨC CCAGGΨAΨGC ΨCCCACCΨCC ACCΨGCCCCA CΨCACCACCΨ 4000

CΨGCΨAGΨΨC CAGACACCΨC CCAAGCACGC AGCAAΨGCAG CΨCAAAACGC 4050

ΨΨAGCCΨAGC CACACCCCCA CGGGAAACAG CAGΨGAΨΨAA CCΨΨΨAGCAA 4100

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