

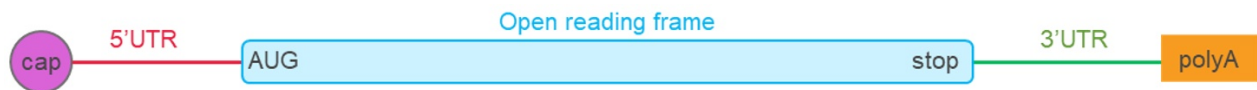


12065

## Description

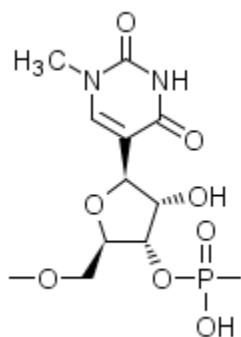
Messenger RNA encoding the full-length SARS-CoV-2 spike glycoprotein

## Schematic of the vector

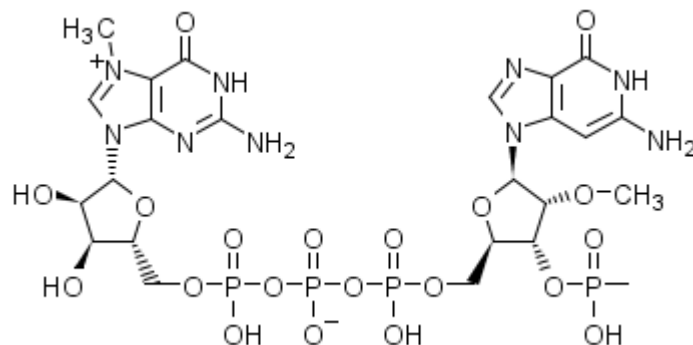


UTR = Untranslated region; Open reading frame of the full-length SARS-CoV-2 spike (S) glycoprotein containing mutations K986P and V987P; poly(A) = polyadenylation tail

$m^1\psi$   
= 1-methyl-3'-pseudouridylyl :



cap  $G^1G^2 =$   
 $m^7G^+-5'-ppp^- -5'-Gm :$



## Table of Features

Element	Description	Position
Cap	5'-cap1 structure (m <sup>7</sup> G-5'-ppp-5'-Gm)	1-2 <b>(bold, underlined)</b>
5'-UTR	The 5'-untranslated region (UTR) is based on the artificial 5'-UTR from Warren et al., 2010 (PMID: 20888316) that confers robust protein expression. It has been further optimized with a GC-rich sequence near the start codon to increase the fidelity of translation initiation at the designated start codon.	3-58 <b>(underlined)</b>
ORF	Codon-optimized sequence encoding full-length SARS-CoV-2 spike (S) glycoprotein, containing mutations K986P and V987P (nucleosides 3014-3019; <b>bold, underlined</b> ) to stabilize the pre-fusion conformation	59-3886
	S glycoprotein signal peptide (leader sequence), which guides translocation of the nascent polypeptide chain into the endoplasmic reticulum.	59 - 94 <b>(bold)</b>
	Stop codons; additional stop codons ensure complete translation termination	3878-3886 <b>(bold)</b>
3'-UTR	The 3'-UTR is derived from the human hemoglobin subunit alpha (HBA1) gene. The 3' UTR has been further optimized by removal of an AUG to eliminate the risk of any aberrant translation.	3887-3996 <b>(underlined)</b>
polyA tail	A 100-nucleotide poly(A)-tail followed by a 5-nucleotide XbaI scar.	3997-4101

## Sequence / Séquence / Secuencia

**GGGGAAAΨAA** **GAGAGAAAAG** **AAGAGΨAAGA** **AGAAAΨAΨAA** **GACCCCGGCG** 50  
**CCGCCACCAΨ** **GΨΨCGΨGΨΨC** **CΨGGΨGCΨGC** **ΨGCCCCΨGGΨ** **GAGCAGCCAG** 100  
**ΨGCGΨGAACC** ΨGACCACCCG GACCCAGCΨG CCACCAGCCΨ ACACCAACAG 150  
CΨΨCACCCGG GGCΨCΨACΨ ACCCCGACAA GGΨGΨΨCCGG AGCAGCGΨCC 200  
ΨGCACAGCAC CCAGGACCΨG ΨΨCCΨGCCCΨ ΨCΨΨCAGCAA CGΨGACCΨGG 250  
ΨΨCCACGCCA ΨCCACGΨGAG CGGCACCAAC GGCACCAAGC GGΨΨCGACAA 300  
CCCCGΨGCΨG CCCΨΨCAACG ACGGCGΨGΨA CΨΨCGCCAGC ACCGAGAAGA 350  
GCAACAΨCAΨ CCGGGGCΨGG AΨCΨΨCGGCA CCACCCΨGGA CAGCAAGACC 400  
CAGAGCCΨGC ΨGAΨCGΨGAA ΨAACGCCACC AACGΨGGΨGA ΨCAAGGΨGΨG 450  
CGAGΨΨCCAG ΨΨCΨGCAACG ACCCCΨΨCCΨ GGGCGΨGΨAC ΨACCACAAGA 500  
ACAACAAGAG CΨGGAΨGGAG AGCGAGΨΨCC GGGΨGΨACAG CAGCGCCAAC 550  
AACΨGCACCCΨ ΨCGAGΨACGΨ GAGCCAGCCC ΨΨCCΨGAΨGG ACCΨGGAGGG 600

CAAGCAGGGC	AACΨΨCAAGA	ACCΨGCGGGA	GΨΨCGΨGΨΨC	AAGAACAΨCG	650
ACGGCΨACΨΨ	CAAGAΨCΨAC	AGCAAGCACA	CCCCAAΨCAA	CCΨGGΨGCGG	700
GAΨCΨGCCCC	AGGGCΨΨCΨC	AGCCCΨGGAG	CCCCΨGGΨGG	ACCΨGCCCAΨ	750
CGGCAΨCAAC	AΨCACCCGGΨ	ΨCCAGACCCΨ	GCΨGGCCΨG	CACCGGAGCΨ	800
ACCΨGACCCC	AGGCGACAGC	AGCAGCGGGΨ	GGACAGCAGG	CGCGGCΨGCΨ	850
ΨACΨACGΨGG	GCΨACCΨGCA	GCCCCGGACC	ΨΨCCΨGCΨGA	AGΨACAACGA	900
GAACGGCACC	AΨCACCGACG	CCGΨGGACΨG	CGCCCΨGGAC	CCΨCΨGAGCG	950
AGACCAAGΨG	CACCCΨGAAG	AGCΨΨCACCG	ΨGGAGAAGGG	CAΨCΨACCAG	1000
ACCAGCAACΨ	ΨCCGGGΨGCA	GCCCACCGAG	AGCAΨCGΨGC	GGΨΨCCCCAA	1050
CAΨCACCAAC	CΨGΨGCCCCΨ	ΨCGGCGAGGΨ	GΨΨCAACGCC	ACCCGGΨΨCG	1100
CCAGCGΨGΨA	CGCCΨGGAAC	CGGAAGCGGA	ΨCAGCAACΨG	CGΨGGCCGAC	1150
ΨACAGCGΨGC	ΨGΨACAACAG	CGCCAGCΨΨC	AGCACCCΨΨCA	AGΨGCΨACGG	1200
CGΨGAGCCCC	ACCAAGCΨGA	ACGACCΨGΨG	CΨΨCACCAAC	GΨGΨACGCCG	1250
ACAGCΨΨCGΨ	GAΨCCGΨGGC	GACGAGGΨGC	GGCAGAΨCGC	ACCCGGCCAG	1300
ACAGGCAAGA	ΨCGCCGACΨA	CAACΨACAAG	CΨGCCCGACG	ACΨΨCACCGG	1350
CΨGCGΨGAΨC	GCCΨGGAACA	GCAACAACCCΨ	CGACAGCAAG	GΨGGGCGGCA	1400
ACΨACAACΨA	CCΨGΨACCGG	CΨGΨΨCCGGA	AGAGCAACCCΨ	GAAGCCCΨΨC	1450
GAGCGGGACA	ΨCAGCACCGA	GAΨCΨACCAA	GCCGGCΨCCA	CCCCΨΨGCAA	1500
CGGCGΨGGAG	GGCΨΨCAACΨ	GCΨACΨΨCCC	ΨCΨGCAGAGC	ΨACGGCΨΨCC	1550
AGCCACCAA	CGGCGΨGGGC	ΨACCAGCCΨΨ	ACCCGGGΨGGΨ	GGΨGCΨGAGC	1600
ΨΨCGAGCΨGC	ΨGCACGCCCC	AGCCACCGΨG	ΨGΨGGCCCCA	AGAAGAGCAC	1650
CAACCΨGGΨG	AAGAACAAGΨ	GCGΨGAACΨΨ	CAACΨΨCAAC	GGCCΨΨACCG	1700
GCACCGGCGΨ	GCΨGACCGAG	AGCAACAAGA	AAΨΨCCΨGCC	CΨΨΨCAGCAG	1750
ΨΨCGGCCCGG	ACAΨCGCCGA	CACCACCGAC	GCΨGΨGCGGG	AΨCCCCAGAC	1800
CCΨGGAGAΨC	CΨGGACAΨCA	CCCCΨΨGCAG	CΨΨCGGCGGC	GΨGAGCGΨGA	1850
ΨCACCCCAGG	CACCAACACC	AGCAACCAGG	ΨGGCCGΨGCGΨ	GΨACCAGGAC	1900
GΨGAACΨGCA	CCGAGGΨGCC	CGΨGGCCAΨC	CACGCCGACC	AGCΨGACACC	1950
CACCΨGGCGG	GΨCΨACAGCA	CCGGCAGCAA	CGΨGΨΨCCAG	ACCCGGGCCG	2000
GΨΨGCCΨGAΨ	CGGCGCCGAG	CACGΨGAACA	ACAGCΨACGA	GΨGCGACAΨC	2050
CCCAΨCGGCG	CCGGCAΨCΨG	ΨGCCAGCΨAC	CAGACCCAGA	CCAAΨΨCACC	2100
CCGGAGGGCA	AGGAGCGΨGG	CCAGCCAGAG	CAΨCAΨCGCC	ΨACACCAΨGA	2150
GCCΨGGGCGC	CGAGAACAGC	GΨGGCCΨACA	GCAACAACAG	CAΨCGCCAΨC	2200
CCCACCAACΨ	ΨCACCAΨCAG	CGΨGACCACC	GAGAΨΨCΨGC	CCGΨGAGCAΨ	2250
GACCAAGACC	AGCGΨGGACΨ	GCACCAΨGΨA	CAΨCΨGCGGC	GACAGCACCG	2300
AGΨGCAGCAA	CCΨGCΨGCGΨ	CAGΨACGGCA	GCΨΨCΨGCAC	CCAGCΨGAAC	2350
CGGGCCCCΨGA	CCGGCAΨCGC	CGΨGGAGCAG	GACAAGAACA	CCCAGGAGGΨ	2400
GΨΨCGCCCAG	GΨGAAGCAGA	ΨCΨACAAGAC	CCCΨCCCAΨC	AAGGACΨΨCG	2450
GCGGCΨΨCAA	CΨΨCAGCCAG	AΨCCΨGCCCCG	ACCCAGCAA	GCCCAGCAAG	2500
CGGAGCΨΨCA	ΨCGAGGACCCΨ	GCΨGΨΨCAAC	AAGGΨGACCC	ΨAGCCGACGC	2550
CGGCΨΨCAΨC	AAGCAGΨACG	GCGACΨGCCΨ	CGGCGACAΨA	GCCGCCCGGG	2600
ACCΨGAΨCΨG	CGCCCAGAAG	ΨΨCAACGGCC	ΨGACCGΨGCGΨ	GCCΨCCCCΨG	2650
CΨGACCGACG	AGAΨGAΨCGC	CCAGΨACACC	AGCGCCCΨGΨ	ΨAGCCGGAAC	2700
CAΨCACCGAC	GGCΨGGACΨΨ	ΨCGGCGCΨGG	AGCCGCΨCΨG	CAGAΨCCCCΨ	2750
ΨCGCCAΨGCA	GAΨGGCCΨAC	CGGΨΨCAACG	GCAΨCGGCGΨ	GACCCAGAAC	2800
GΨGCΨGΨACG	AGAACCAGAA	GCΨGAΨCGCC	AACCAGΨΨCA	ACAGCGCCAΨ	2850
CGGCAAGAΨC	CAGGACAGCC	ΨGAGCAGCAC	CGCΨAGCGCC	CΨGGGCAAGC	2900
ΨGCAGGACGΨ	GGΨGAACCAG	AACGCCCAGG	CCCΨGAACAC	CCΨGGΨGAAG	2950
CAGCΨGAGCA	GCAACΨΨCGG	CGCCAΨCAGC	AGCGΨGCΨGA	ACGACAΨCCΨ	3000
GAGCCGGCΨG	GAC <b>CCΨCCC</b> G	AGGCCGAGGΨ	GCAGAΨCGAC	CGGCΨGAΨCA	3050
CΨGGCCGGCΨ	GCAGAGCCΨG	CAGACCΨACG	ΨGACCCAGCA	GCΨGAΨCCGG	3100
GCCGCCGAGA	ΨΨCGGGCCAG	CGCCAACCCΨG	GCCGCCACCA	AGAΨGAGCGA	3150
GΨGCGΨGCGΨ	GGCCAGAGCA	AGCGGGΨGGA	CΨΨCΨGCGGC	AAGGGCΨACC	3200
ACCΨGAΨGAG	CΨΨΨCCCCAG	AGCGCACCCC	ACGGAGΨGGΨ	GΨΨCCΨGCAC	3250
GΨGACCCΨACG	ΨGCCCGCCCA	GGAGAAGAAC	ΨΨCACACC	CCCCAGCCAΨ	3300

CΨGCCACGAC	GGCAAGGCC	ACΨΨΨCCCCG	GGAGGGCGΨG	ΨΨCGΨGAGCA	3350
ACGGCACCCA	CΨGGΨΨCGΨG	ACCCAGCGGA	ACΨΨCΨACGA	GCCCCAGAΨC	3400
AΨCACCACCG	ACAACACCΨΨ	CGΨGAGCGGC	AACΨGCGACG	ΨGGΨGAΨCGG	3450
CAΨCGΨGAAC	AACACCGΨΨ	ACGAΨCCCCΨ	GCAGCCCCGAG	CΨGGACAGCΨ	3500
ΨCAAGGAGGA	GCΨGGACAAG	ΨACΨΨCAAGA	AΨCACACCAG	CCCCGACGΨG	3550
GACCΨGGGCG	ACAΨCAGCGG	CAΨCAACGCC	AGCGΨGGΨGA	ACAΨCCAGAA	3600
GGAGAΨCGAΨ	CGGCΨGAACG	AGGΨGGCCAA	GAACCΨGAAC	GAGAGCCΨGA	3650
ΨCGACCΨGCA	GGAGCΨGGGC	AAGΨACGAGC	AGΨACAΨCAA	GΨGGCCCΨGG	3700
ΨACAΨCΨGGC	ΨGGGCΨΨCAΨ	CGCCGGCCΨG	AΨCGCCAΨCG	ΨGAΨGGΨGAC	3750
CAΨCAΨGCΨG	ΨGCΨGCAΨGA	CCAGCΨGCΨG	CAGCΨGCCΨG	AAGGGCΨGΨΨ	3800
GCAGCΨGCGG	CAGCΨGCΨGC	AAGΨΨCGACG	AGGACGACAG	CGAGCCCΨG	3850
CΨGAAGGGCG	ΨGAAGCΨGCA	CΨACACCΨGA	<b>ΨAAΨAGGCΨG</b>	GAGCCΨCGGΨ	3900
<u>GGCCΨAGCΨΨ</u>	<u>CΨΨGCCCCΨΨ</u>	<u>GGGCCΨCCCC</u>	<u>CCAGCCCCΨC</u>	<u>CΨCCCCΨΨCC</u>	3950
<u>ΨGCACCCGΨA</u>	<u>CCCCCGΨGGΨ</u>	<u>CΨΨΨGAAΨAA</u>	<u>AGΨCΨGAGΨG</u>	<u>GGCGGCAAAA</u>	4000
AAAAAAAAAA	AAAAAAAAAA	AAAAAAAAAA	AAAAAAAAAA	AAAAAAAAAA	4050
AAAAAAAAAA	AAAAAAAAAA	AAAAAAAAAA	AAAAAAAAAA	AAAAAAΨCΨA	4100
G					4101