

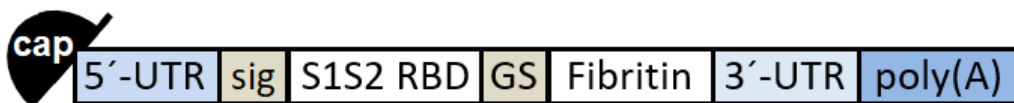


11888

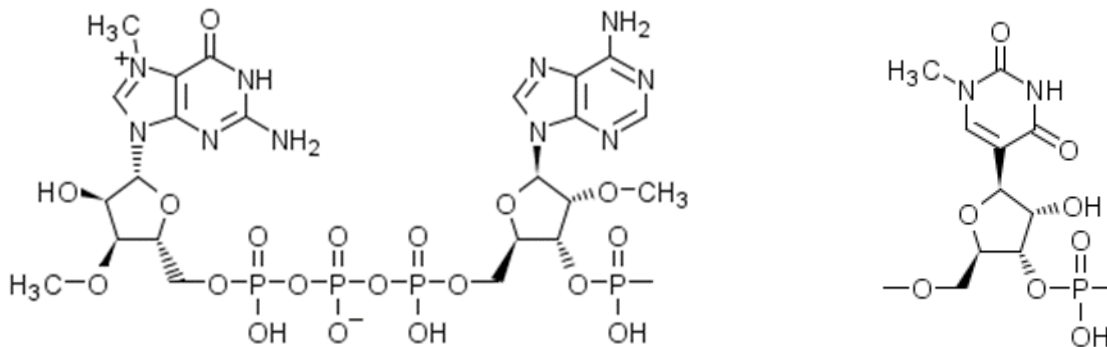
Description

Messenger RNA encoding the receptor binding domain of the SARS-CoV-2 spike glycoprotein fused with the T4 fibrin 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; 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'- | 1-2 |



| | | |
|----------|---|-----------|
| | Am) | |
| 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 |
| S1S2 RBD | Codon-optimized sequence, encoding the Receptor Binding Domain (RBD) of the of the SARS-CoV-2 spike (S) glycoprotein. | 103 - 708 |
| GS | Sequence that codes for a short linker peptide consisting of the amino-acids glycine and serine. | 709-735 |
| Fibritin | Partial sequence of T4 fibritin (foldon), used as a trimerization domain. | 736-858 |
| 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. | 859-1153 |
| 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. | 1154-1263 |

Sequence / Séquence / Secuencia

| | | | | | |
|--------------------|-------------------|-------------------|-------------------|-------------------|------------|
| GAGAAΨAAAC | ΨAGΨAΨΨCΨΨ | CΨGGΨCCCCA | CAGACΨCAGA | GAGAACCCGC | 50 |
| CACCAΨGΨΨΨ | GΨGΨΨΨCΨΨG | ΨGCΨGCΨGCC | ΨCΨΨGΨGΨCΨ | ΨCΨCAGΨGΨG | 100 |
| ΨGGΨGAGAΨΨ | ΨCCAAAΨAΨΨ | ACAAAΨCΨGΨ | GΨCCAΨΨΨGG | AGAAGΨGΨΨΨ | 150 |
| AAΨGCAACAA | GAΨΨΨGCAΨC | ΨGΨGΨAΨGCA | ΨGGAAΨAGAA | AAAGAAΨΨΨC | 200 |
| ΨAAΨΨGΨGΨG | GCΨGAΨΨAΨΨ | CΨGΨGCΨGΨA | ΨAAΨAGΨGCΨ | ΨCΨΨΨΨΨCCA | 250 |
| CAΨΨΨAAAΨG | ΨΨAΨGGAGΨG | ΨCΨCCAACAA | AAΨΨAAAΨGA | ΨΨΨAΨGΨΨΨΨ | 300 |
| ACAAAΨGΨGΨ | AΨGCΨGAΨΨC | ΨΨΨΨGΨGAΨC | AGAGGΨGAΨG | AAGΨGAGACA | 350 |
| GAΨΨGCCCCC | GGACAGACAG | GAAAAAΨΨGC | ΨGAΨΨACAAΨ | ΨACAAACΨGC | 400 |
| CΨGAΨGAΨΨΨ | ΨACAGGAΨGΨ | GΨGAΨΨGCΨΨ | GGAAΨΨCΨAA | ΨAAΨΨΨAGAΨ | 450 |
| ΨCΨAAAAGΨGG | GAGGAAAΨΨA | CAAΨΨAΨCΨG | ΨACAGACΨGΨ | ΨΨAGAAAAΨC | 500 |
| AAAΨCΨGAAA | CCΨΨΨΨGAAA | GAGAΨAΨΨΨC | AACAGAAAΨΨ | ΨAΨCAGGCΨG | 550 |
| GAΨCAACACC | ΨΨGΨAAΨGGA | GΨGGAAGGAΨ | ΨΨAAΨΨGΨΨA | ΨΨΨΨCCAΨΨA | 600 |
| CAGAGCΨAΨG | GAΨΨΨCAGCC | AACCAAΨGGΨ | GΨGGGAΨAΨC | AGCCAΨAΨAG | 650 |
| AGΨGGΨGGΨG | CΨGΨCΨΨΨΨG | AACΨGCΨGCA | ΨGCACCΨGCA | ACAGΨGΨGΨG | 700 |
| GACCΨAAAGG | CΨCCCCCGGC | ΨCCGGCΨCCG | GAΨCΨGGΨΨA | ΨAΨΨCCΨGAA | 750 |
| GCΨCCAAGAG | AΨGGGCAAGC | ΨΨACGΨΨCGΨ | AAAGAΨGGCG | AAΨGGGΨAΨΨ | 800 |
| ACΨΨΨCΨACC | ΨΨΨΨΨAGGCC | GGΨCCCΨGGA | GGΨGCΨGΨΨC | CAGGGCCCCG | 850 |
| GCΨGAΨGACΨ | CGAGCΨGGΨA | CΨGCAΨGCAC | GCAAΨGCΨAG | CΨGCCCCΨΨΨ | 900 |



WHO
International Nonproprietary Names Programme

9/2020

| | | | | | |
|--------------|--------------|--------------|--------------|--------------|------|
| CCCGΨCCΨGG | GΨACCCCGAG | ΨCΨCCCCGA | CCΨCGGGΨCC | CAGGΨAΨGΨ | 950 |
| CCCACCΨCCA | CCΨGCCCCAC | ΨCACCACCΨC | ΨGCΨAGΨΨCC | AGACACCΨCC | 1000 |
| CAAGCACGCA | GCAAΨGCAGC | ΨCAAAACGCΨ | ΨAGCCΨAGCC | ACACCCCCAC | 1050 |
| GGGAAACAGC | AGΨGAΨΨAAC | CΨΨΨAGCAAΨ | AAACGAAAGΨ | ΨΨAACΨAAGC | 1100 |
| ΨAΨACΨAACC | CCAGGGΨΨGG | ΨCAAΨΨΨCGΨ | GCCAGCCACA | CCCΨGGAGCΨ | 1150 |
| AGCAAAAAAAAA | AAAAAAAAAAAA | AAAAAAAAAAAA | AAAGCAΨAΨG | ACΨAAAAAAAA | 1200 |
| AAAAAAAAAAAA | AAAAAAAAAAAA | AAAAAAAAAAAA | AAAAAAAAAAAA | AAAAAAAAAAAA | 1250 |
| AAAAAAAAAAAA | AAA | | | | 1263 |

Ψ = 1-methyl-3'-pseudouridylyl