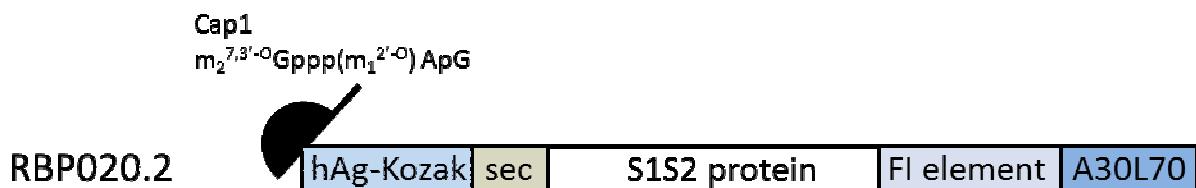


3.2.S.1.2. STRUCTURE

BNT162b2 drug substance is a single-stranded, 5'-capped mRNA that is translated into a protein (the encoded antigen). Figure 3.2.S.1.2-1 illustrates the general structure of the antigen-encoding RNA, which is determined by the respective nucleotide sequence of the DNA used as template for in vitro RNA transcription. In addition to the codon-optimized sequence encoding the antigen, the RNA contains common structural elements optimized for mediating high RNA stability and translational efficiency (5'-cap, 5'-UTR, 3'-UTR, poly(A) - tail; see below). Furthermore, an intrinsic signal peptide (sec) is part of the open reading frame and is translated as an N-terminal peptide. The RNA does not contain any uridines; instead of uridine the modified N1-methylpseudouridine is used in RNA synthesis.

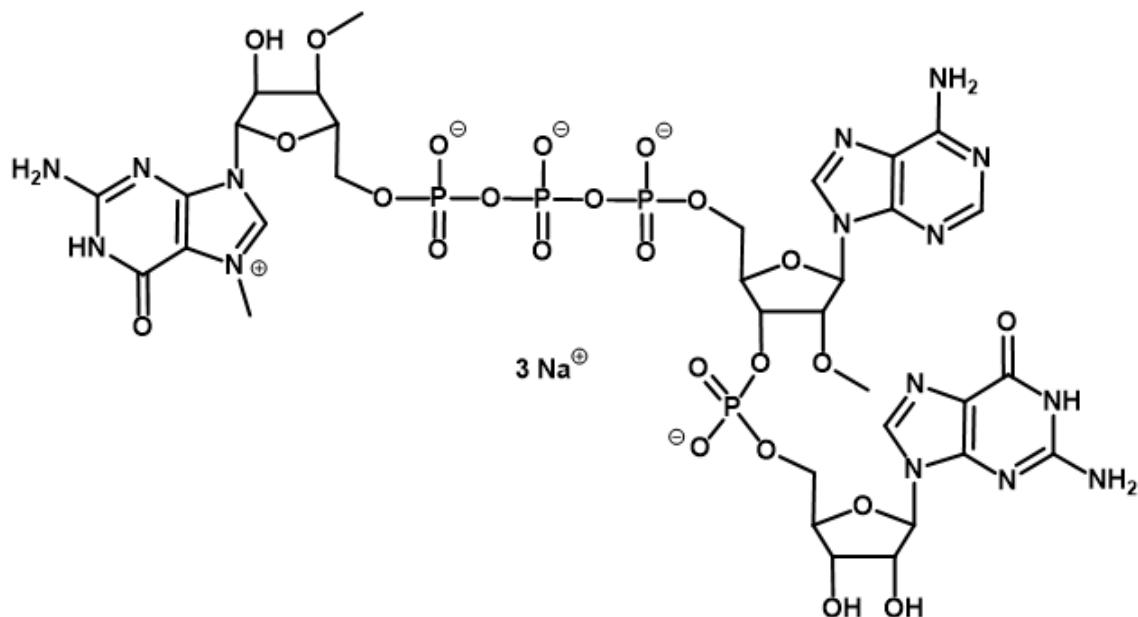
Figure 3.2.S.1.2-1. General structure of the RNA



Schematic illustration of the general structure of the BNT162b2 drug substance with 5'-cap, 5'- and 3'-untranslated regions (hAg-Kozak and FI element, respectively), coding sequence with mutations and intrinsic signal peptide (sec) as well as poly(A)-tail (A30L70). Individual elements are not drawn to scale compared to their respective sequence lengths.

mRNA cap

A cap1 structure $m_2^{7,3'-O}Gppp(m_1^{2'-O})ApG$ [Figure 3.2.S.1.2-2](#) is utilized as specific capping structure at the 5'-end of the RNA drug substance ([Figure 3.2.S.1.2-2](#)).

Figure 3.2.S.1.2-2. 5'-cap analog ($m_2^{7,3'-O}Gppp(m_1^{2'-O})ApG$) for production of RNA containing a cap1 structure

The cap1 structure (i.e., containing a 2'-O-methyl group on the penultimate nucleoside of the 5'-end of the RNA chain) is incorporated into the BNT162b2 drug substance by using a respective cap analog during *in vitro* transcription. For RNAs with modified uridine nucleotides, the cap1 structure is superior to other cap structures, since cap1 is not recognized by cellular factors such as IFIT1¹ and, thus, cap1-dependent translation is not inhibited by competition with eukaryotic translation initiation factor 4E². In the context of IFIT1 expression, mRNAs with a cap1 structure give higher protein expression.

In addition, use of the cap1 structure leads to low amounts of uncapped transcripts³. In general, the T7 Polymerase prefers a guanosine as priming nucleoside with the highest transcription efficiencies as compared to other starting nucleosides⁴. Capping structures with a guanosine moiety compete with GTP for incorporation in the mRNA resulting in uncapped transcripts. The $m_2^{7,3'-O}Gppp(m_1^{2'-O})ApG$ cap analog rescues transcription efficiency from

¹ Habjan M, Hubel P, Lacerda L, et al. Sequestration by IFIT1 Impairs Translation of 2' O-unmethylated Capped RNA. 2013. PLOS Pathog;9(10):e1003663

² Diamond MS. IFIT1: A dual sensor and effector molecule that detects non-2'-O methylated viral RNA and inhibits its translation. 2014. Cytokine Growth Factor Rev;25(5):543-50.

³ Trilink Patent auf CC413 cap. Accessed at <https://patentimages.storage.googleapis.com/4c/83/15/99418d175a3be2/WO2017053297A1.pdf>

⁴ Kuzmine I, Gottlieb PA, Martin CT. Binding of the priming nucleotide in the initiation of transcription by T7 RNA polymerase. 2003. J Biol Chem;278(5):2819-23.

templates starting with adenosines, because the ApG moiety of cap1 allows transcription initiation at the second position, a guanosine, thereby giving mainly capped mRNAs.

Modified Uridine

The RNA does not contain any uridines; instead of uridine the modified N1-methylpseudouridine is used in RNA synthesis. Several reports have demonstrated that such a substitution often strongly enhances translation of *in vitro* transcribed mRNA sequences by reducing its immunogenicity^{5,6,7}. Accordingly, the BNT162b2 drug substance is synthesized in the presence of N1-methylpseudouridine triphosphate ($^{m1}\Psi TP$) instead of uridine triphosphate (UTP).

RNA sequence

The general sequence elements of the BNT162b2 drug substance, as depicted in [Figure 3.2.S.1.2-1](#), are given below. The full sequence is given in [Figure 3.2.S.1.2-3](#).

The vaccine is based on the spike glycoprotein (S) of the SARS-CoV-2 virus. The sequence was chosen based on the sequence for the “Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1”, which was available when the program was initiated:

- GenBank: MN908947.3 (complete genome)
- GenBank: QHD43416.1 (spike surface glycoprotein)

hAg-Kozak (nucleotides 2 to 54): 5'-UTR sequence of the human alpha-globin mRNA with an optimized ‘Kozak sequence’ to increase translational efficiency⁸.

Sec (nucleotides 55 to 102): Sec corresponds to the intrinsic S1S2 protein signal peptide (sec), which guides translocation of the nascent polypeptide chain into the endoplasmic reticulum.

S1S2 protein (nucleotides 103 to 3879): Codon-optimized sequence encoding the spike antigen of SARS-CoV-2. The S1S2 protein or spike glycoprotein is expressed on

⁵ Kariko K, Muramatsu H, Welsh FA, et al. Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability. 2008. Mol Ther;16(11):1833-40.

⁶ Andries O, Mc Cafferty S, De Smedt SC, et al. N(1)-methylpseudouridine-incorporated mRNA outperforms pseudouridine-incorporated mRNA by providing enhanced protein expression and reduced immunogenicity in mammalian cell lines and mice. 2015. J Control Release; 217:337-44.

⁷ Richner JM, Himansu S, Dowd KA, et al. Modified mRNA Vaccines Protect against Zika Virus Infection. 2017. Cell;168(6):1114-25.e10

⁸ Kozak M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. 1987. Nucleic Acids Res;15(20):8125-48.

BNT162b2
3.2.S.1.2 Structure

membranes. It facilitates recognition by the host cells as well as cellular uptake. The protein sequence contains two proline mutation (K986P and V987P), which ensures an antigenically optimal pre-fusion confirmation (P2 S).

FI element (nucleotides 3880 to 4174): The 3'-UTR is a combination of two sequence elements derived from the “amino terminal enhancer of split” (AES) mRNA (called F) and the mitochondrial encoded 12S ribosomal RNA (called I). These were identified by an *ex vivo* selection process for sequences that confer RNA stability and augment total protein expression⁹.

A30L70 (nucleotides 4175 to 4284): A poly(A)-tail measuring 110 nucleotides in length, consisting of a stretch of 30 adenosine residues, followed by a 10 nucleotide linker sequence and another 70 adenosine residues designed to enhance RNA stability and translational efficiency in dendritic cells¹⁰.

Figure 3.2.S.1.2-3. RNA nucleotide Sequence of the BNT162b2 drug substance:

Nucleotide sequence 5'→3':

```

GAGAAYAAAC YAGYAYYCYY CYGGYCCCCA CAGACYCAGA GAGAACCGC 50
CACCAYGYYC GYGYYYCCYGG YGCYGCYGCC YCYGGYGYCC AGCCAGYGYG 100
YGAACCYGAC CACCAGAACCA CAGCYGCCYC CAGCCYACAC CAACAGCYYY 150
ACCAGAGGCG YGYACYACCC CGACAAGGGY YYCAGAYCCA GCGYGCYGC 200
CYCYACCCAG GACCYGYYYCC YGCCYYCYY CAGCAACGYG ACCYGGYYCC 250
ACGCCAYCCA CGYGYCCGGC ACCAAYGGCA CCAAGAGAYY CGACAACCCC 300
GYGCYGCCY YCAACGACGG GGYGYACYYY GCCAGCACCG AGAACGCCA 350
CAYCAYCAGA GGCYGGAYCY YCGGCACCAC ACYGGACAGC AAGACCCAGA 400
GCCYGCYGA CGYGAACAAAC GCCACCAACG YGGYCAYCAA AGYGYGCGAG 450
YYCCAGYYCY GCAACGACCC CYYCCYGGGC GYCYACYACC ACAAGAACAA 500
CAAGAGCYGG AYGGAAAGCG AGYYCCGGGY GYACAGCAGC GCCAACAAACY 550
GCACCCYYCGA GYACGYGYCC CAGCCYYCC YGAYGGACCC GGAAGGCAAG 600
CAGGGCAACY YCAAGAACCY GCGCGAGYYC GYGYYYAAGA ACAYCGACGG 650
CYACYYCAAG AYCYACAGCA AGCACACCCC YAYCAACCCY GYGCAGGAYC 700
YGCYCAGGG CYYCYCYGCY CYGGAACCCC YGGYGGAYCY GCCCAYCGGC 750
AYCAACAYCA CCCGGYYYCA GACACYGCYGCY GCCCYGCACA GAAGCYACCCY 800
GACACCCYGGC GAYAGCAGCA GCGGAYGGAC AGCYGGYGCC GCCGCYYACY 850
AYGYGGGCYA CCGACGCCGY AGAACCCYCC YGCYGAAGYA CAACCGAGAAC 900
GGCACCCAYCA CCGACGCCGY GGAYYYGYGCY CYGGAYCCYC YGAGCGAGAC 950
AAAGYGCCACC CYGAAGYCCY YCACCGYGGA AAAGGGCAYC YACCAGACCA 1000
GCAACYYCCG GGYGCAGCCC ACCGAAYCCA YCGYGCAGGGY CCCCAAYAYC 1050
ACCAAYCYGY GCCCCYYCGG CGAGGYGYYC AAYGCCACCA GAYYCGCCYC 1100
YGYGYACGCC YGGAACCGGA AGCGGAYCAG CAAYYGCYGYG GCCGACYACY 1150

```

⁹ Orlandini von Niessen AG, Poleganov MA, Rechner C, et al. Improving mRNA-Based Therapeutic Gene Delivery by Expression-Augmenting 3' UTRs Identified by Cellular Library Screening. 2019. Mol Ther;27(4):1-13.

¹⁰ BioNTech Patent auf STABILISIERUNG VON DNA-SEQUENZEN ZUR POLY(A)SEQUENZ-CODIERUNG. Accessed at <https://data.epo.org/publication-server/pdf-document?pn=3167059&ki=B1&cc=EP&pd=20190626>

BNT162b2

3.2.S.1.2 Structure

CCGYGCYGYA CAACYCCGCC AGCYYCAGCA CCYYCAAGYG CYACGGCGYG 1200
 YCCCCYACCA AGCYGAACGA CCYGYCYYC ACAAACGYGY ACGCCGACAG 1250
 CYYCGYGYAC CGGGGAGAYG AAGYGCGGCA GAYYGCCCY GGACAGACAG 1300
 GCAAGAYCGC CGACYACAAC YACAAGCYGC CCGACGACYY CACCGGYGY 1350
 GYGAYYGCCY GGAACAGCAA CAACCYGGAC YCCAAAGYCG GCGGCAACYA 1400
 CAAYYACCYG YACCGGYGY YCCGGAAGYC CAAYCYGAAG CCCYYCGAGC 1450
 GGGACACYCYC CACCGAGAYC YAYCAGGCCG GCAGCACCCC YYGYAACGGC 1500
 GYGGAAAGGCY YCAACYGCYA CYYCCACAYG CAGYCCYACG GCYYYCAGCC 1550
 CACAAAYGGC GYGGGYAYC AGCCCYACAG AGYGGYGGY CYGAGCYYCG 1600
 AACYGCYGCYCA YGCCCCYGCY ACAGYGYGCG GCCCYAAGAA AAGCACCAAY 1650
 CYCGYGAAGA ACAAAAYGCGY GAACYYCAAC YYCAACAGGCC YGACCGGCAC 1700
 CGCGCGYGYC AGAGAGAGCA ACAAGAAGYY CCYGCCAYYC CAGCAGYYYG 1750
 GCCGGGAYAY CGCCGAYACC ACAGACGCCG YYAGAGAYCC CCAGACACYG 1800
 GAAAYCCYGG ACAYCACCCC YYGCAGCYYC GGCGGAGYGY CYGYGAYCAC 1850
 CCCYGGCACC AACACCAGCA AYCAGGYGGC AGYGCYGYAC CAGGACGYGA 1900
 ACYGYACCAGA AGYGCCCGYGC GCCAYYCACG CCGAYCAGCY GACACCYACA 1950
 YGGCGGGYGY ACYCCACCGG CAGCAAYGYG YYCAGACCCA GAGCCGGCYG 2000
 YCYGAYCGGA GCCGAGCAGC YGAACAAYAG CYACGAGYGC GACAYCCCCA 2050
 YCGGCGYGG AAYCYGCGCC AGCYACCAGA CACAGACAAA CAGCCCYCGG 2100
 AGAGCCAGAA GCGYGGCCAG CCAGAGCAYC AYYGCCYACA CAAYGYCYCY 2150
 GGGCGCCGAG AACAGCGYGG CCYACYCAA CAACYCYAYC GCYAYCCCCA 2200
 CCAACYYCAC CAYCAGCGY ACCACAGAGA YCCYGCCYGY GYCCAYGACC 2250
 AAGACCAGCG YGGACYGCAC CAYGYACAYC YCGGGCGAYY CCACCGAGYG 2300
 CYCCAACCYG CYGCYGCAGY ACGGCAGCYY CYGCACCCAG CYGAAYAGAG 2350
 CCCYGACAGG GAYCGCCGYG GAACAGGACA AGAACACCCA AGAGGYGYYC 2400
 GCCCAAGYGA AGCAGAYCYA CAAGACCCCY CCYAYCAAGG ACYYCGGCAG 2450
 CYYCAAYYYC AGCCAGAYYC YGCCCCGAYCC YAGCAAGCCC AGCAAGCGGA 2500
 GCYYCAYCGA GGACCYGYCYY CAACACAAAG YGACACYGGC CGACGCCGGC 2550
 YYCAYCAAGC AGYAYGGCGA YYGYCYGGC GACAYYGCCG CCAGGGAYCY 2600
 GAYYYGCGCC CAGAAGYYYY ACGGACYGAC AGYGCYGCCY CCYCYGYGA 2650
 CCGAYGAGAY GAYCGCCCAAG YACACAYCYG CCCYGYGYGC CGGCACAAYC 2700
 ACAAGCGGYC GGACAYYYGG AGCAGGCGCC GCYCYGCAGA YCCCCYYYGC 2750
 YAYGCAGAYG GCCYACCGGY YCAACGGCAY CGGAGYGCACC CAGAAAGYGC 2800
 YGYACGAGAA CCAGAAGCGY AYCGCCAACC AGYYCAACAG CGCCAYCGGC 2850
 AAGAYCCAGG ACAGCCYAGG CAGCACAGCA AGCGCCCCYGG GAAAGCYGCA 2900
 GGACGYGGYC AACCAGAAYG CCCAGGCACY GAACACCCYGC GYCAAGCAGC 2950
 YGYCCYCAA CYYCGCGCC AYCAGCYCYG YGYGAACGA YAYCCYGAGC 3000
 AGACYGGACC CYCCYGAGGC CGAGGYGCAG AYCGACAGAC YGAYCACAGG 3050
 CAGACYGCAG AGCCYCCAGA CAYACCGYGC CCAGCAGCYG AYCAAGGCCG 3100
 CCGAGAAYYAG AGCCYCYGCC AAYCYGGCCG CCACCAAGAY GYCYGAGYGY 3150
 GYGCYGGGCC AGAGCAAGAG AGYGGACYYY YGCGGCAAGG GCYACCACCC 3200
 GAYGAGCYYC CCYCAAGCYCYY CCCCYCACGG CGYGGYGYYY CYGCACGYGA 3250
 CAYAYGYGCC CGCYCAAGAG AAGAAYYYCA CCACCGCYCC AGCCAYCYGC 3300
 CACGACGGCA AAGCCCACYY YCCYAGAGAA GGCGYGYYCG YGYCCAACGG 3350
 CACCCAYYYGG YYCGYGYACAC AGCGGAACYY CYACGAGCCC CAGAYCAYCA 3400
 CCACCGACAA CACCYYCGYGC YCYGGCAACY GCGACGYCGY GAYCGGCAYY 3450
 GYGAACAAAYA CCGYGYACGA CCCYCYGCAG CCCGAGCYGG ACAGCYYCAA 3500
 AGAGGAACCGY GACAAGYACT YYAAGAACCA CACAAGCCCC GACGYGGACC 3550
 YGGGCGAYAY CAGCGGAACY AAYGCCAGCG YCGYGAACAY CCAGAAAGAG 3600
 AYCGACCGGC YGAACGAGGY GGCCAAGAA CYGAACGAGA GCCYGYCGA 3650
 CCYGCAGAA CYGGGGAAAGY ACGAGCAGYA CAYCAAGYGG CCCYGGYACA 3700
 YCYGGCYGGG CYYAYCGCC GGACYGAYY CCAYCGYGY GGYCACAAYC 3750
 AYGCYGYGYY GCAYGACCAAG CYGCYGYAGC YGCCYGAAGG GCYGGYGYAG 3800
 CYGYGGCAGC YGCYGAAGY YCGACGAGGA CGAYYCYGAG CCCGYGYCGA 3850
 AGGGCGYGAAC YCYGCACAC YCAAYGAGAC YCGAGCYGGY ACYGCAYGCA 3900
 CGCAAYGCYA GCGYCCCCYY YCCCGYCCYGC GGYACCCCGA GYCYCCCCCC 3950
 ACCYCGGGYGC CCAGGYAYGC YCCCACCCYCC ACCYGCACCA CYCACCAACCCY 4000

BNT162b2
3.2.S.1.2 Structure

CYGCYAGYYC CAGACACCCYC CCAAGCACGC AGCAAYGCAG CYCAAAACGC 4050
YYAGCCYAGC CACACCCCCA CGGGAAACAG CAGYGAYYAA CCYYYAGCAA 4100
YAAACGAAAG YYYAACYAAAG CYAYACYAAC CCCAGGGYYG GYCAAYYYCG 4150
YGCCAGCCAC ACCCYGGAGC YAGAAAAAAA AAAAAAAAAA AAAAAAAAAA 4200
AAAAGCAYAY GACYAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 4250
AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAA 4284

Sequence length: 4284, which includes G to denote the presence of the 5'-cap analog

G: 1062 C: 1315 A: 1106 Y: 801

A = Adenine; C = Cytosine; G = Guanine; Y = N1-methylpseudouridine