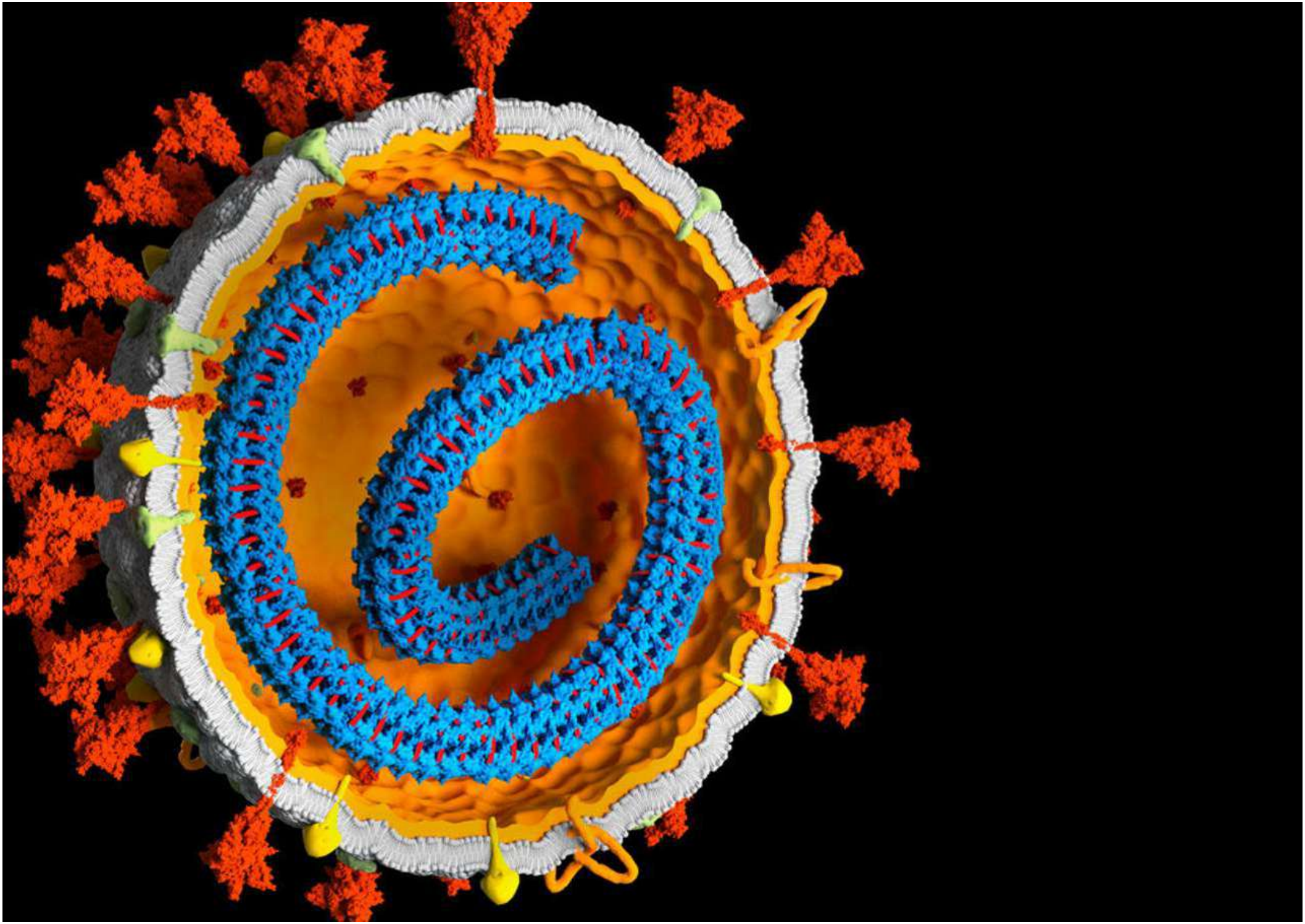
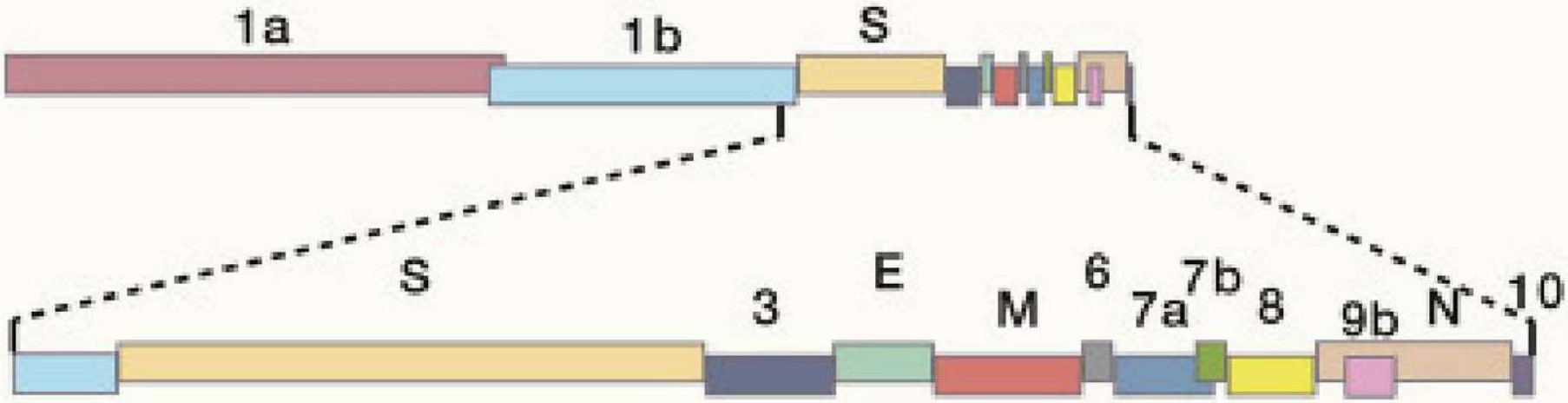
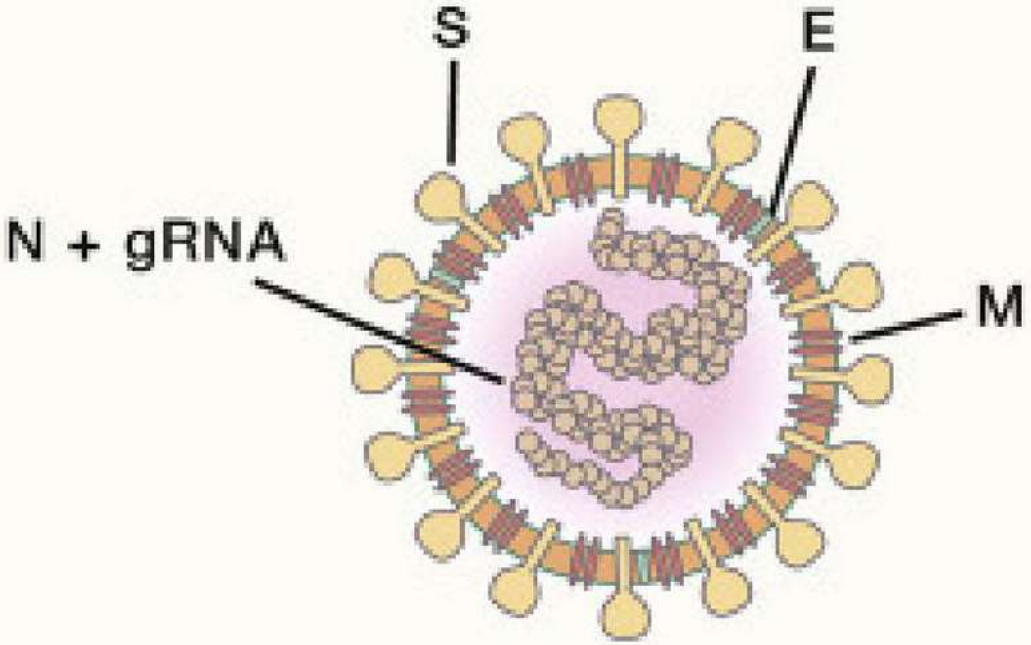






S (Spike): Mediates cell entry, contains majority of neutralizing antibody and T cell epitopes

N (Nucleocapsid protein): Complexes with genomic RNA, contains T cell epitopes and is a major target of the antibody response

M (Matrix protein) and E (Envelope protein): Interact to form membrane and with N during assembly, M contains T cell epitopes

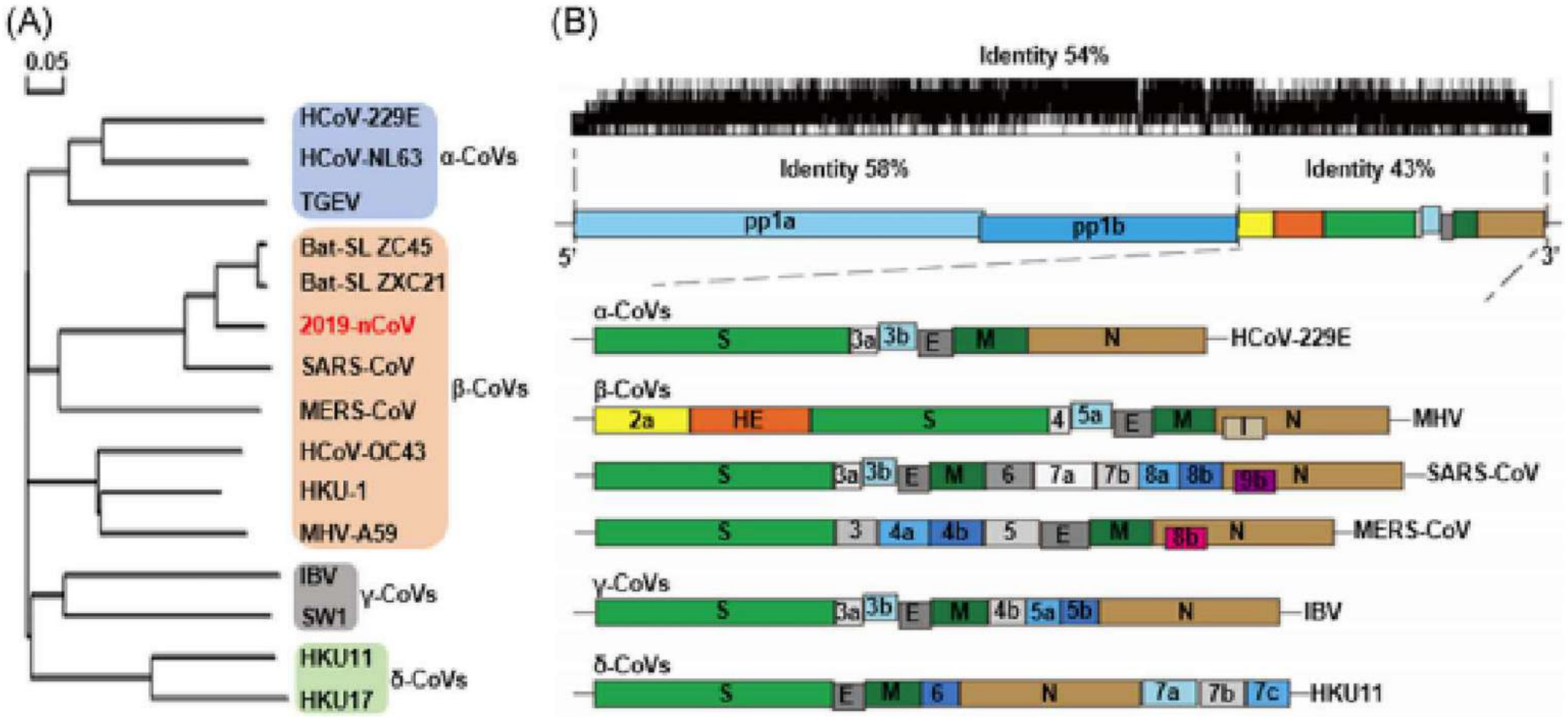
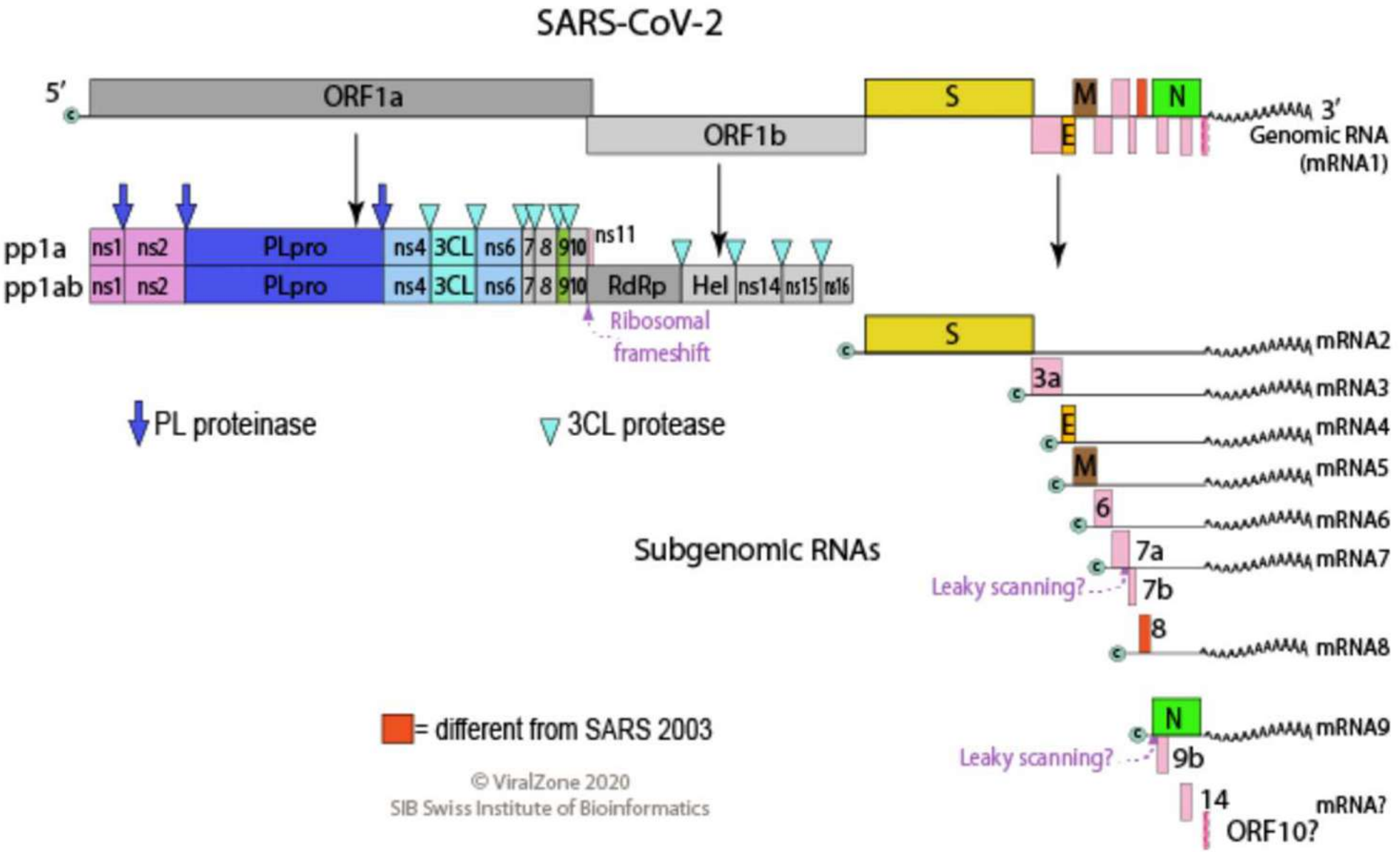


FIGURE 1 The genomic structure and phylogenetic tree of coronaviruses. A, The phylogenetic tree of representative CoVs, with the new coronavirus 2019-nCoV highlighted in red. B, The genome structure of four genera of coronaviruses. Pp1a and pp1b represent the two long polypeptides that are processed into 16 nonstructural proteins. S, E, M, and N indicate the four structural proteins spike, envelope, membrane, and nucleocapsid. 2019-nCoV, 2019 novel coronavirus; CoVs, coronavirus; HE, hemagglutinin-esterase. Viral names: HKU, coronaviruses identified by Hong Kong University; HCoV, human coronavirus; IBV, infectious bronchitis virus; MHV, murine hepatitis virus; TGEV, transmissible gastroenteritis virus



The SARS-CoV-2 genome structure as available on viralzone.expasy.org

nsp1 (180 amino acids):

- Residue 1-180 in orf1ab polyprotein

  MESLVPGFNEKTHVQLSLPVLQVRDVLVRGFGDSVEEVLSEARQHLKDGTCGLVEVEKGVLPQLE
  QPYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGETLGVLVPHVGEIPVAYRKVLLRKNGNKG
  AGGHSYGADLKSFDLGDELGTDPYEDFQENWNTKHSSGVTRELMRELNGG

- Cellular mRNA degradation, inhibiting IFN signaling [https://onlinelibrary.wiley.com/doi/epdf/10.1002/jmv.25681]
- Inhibits host translation by interacting with the 40S ribosomal subunit. The nsp1-40S ribosome complex further induces an endonucleolytic cleavage near the 5'UTR of host mRNAs, targeting them for degradation. Viral mRNAs are not susceptible to nsp1-mediated endonucleolytic RNA cleavage thanks to the presence of a 5'-end leader sequence and are therefore protected from degradation. By suppressing host gene expression, nsp1 facilitates efficient viral gene expression in infected cells and evasion from host immune response. [https://swissmodel.expasy.org/repository/species/2697049]
- Host translation inhibitor nsp1.
- 3D-Structures:
  - No known experimental structure
  - Similar nsp1 from SARS-CoV with known structure: RCSB, 3decision
  - Low quality model from SwissModel

nsp2 (638 amino acids):

- Residue 181-818 in orf1ab polyprotein

  AYTRYVDNNFCGPDGYPLECIKDLLARAGKASCTLSEQLDFIDTKRGVYCCREHEHEIAWYTERS
  EKSYELQTPFEIKLAKKFDTFNGECPNFVFPLNSIIKTIQPRVEKKKLDGFMGRIRSVYPVASPN
  ECNQMCLSTLMKCDHCGETSWQTGDFVKATCEFCGTENLTKEGATTCGYLPQNAVVKIYCPACHN
  SEVGPEHSLAEYHNESGLKTILRKGGRTIAFGGCVFSYVGCHNKCAYWVPRASANIGCNHTGVVG
  EGSEGLNDNLLEILQKEKVNINIVGDFKLNEEIAIILASFSASTSAFVETVKGLDYKAFKQIVES
  CGNFKVTKGKAKKGAWNIGEQKSILSPLYAFASEAARVVRSIFSRTLETAQNSVRVLQKAAITIL
  DGISQYSLRLIDAMMFTSDLATNNLVVMAYITGGVVQLTSQWLTNIFGTVYEKLKPVLDWLEEKF
  KEGVEFLRDGWEIVKFISTCACEIVGGQIVTCAKEIKESVQTFFKLVNKFLALCADSIIIGGAKL
  KALNLGETFVTHSKGLYRKCVKSREETGLLMPLKAPKEIIFLEGETLPTEVLTEEVVLKTGDLQP
  LEQPTSEAVEAPLVGTPVCINGLMLLEIKDTEKYCALAPNMMVTNNTFTLKGG

- May play a role in the modulation of host cell survival signaling pathway by interacting with host PHB and PHB2. Indeed, these two proteins play a role in maintaining the functional integrity of the mitochondria and protecting cells from various stresses. [https://swissmodel.expasy.org/repository/species/2697049, https://jvi.asm.org/content/83/19/10314]

- 3D-Structures:

  - No known experimental structure

nsp3 (1945 amino acids):

- Residue 819-2763 in orf1ab polyprotein

APTKVTFGDDTVIEVQGYKSVNITFELDERIDKVLNEKCSAYTVELGTEVNEFACVVADAVIKTL
QPVSELLTPLGIDLDEWSMATYYLFDESGEFKLASHMYCSFYPPDEDEEEGDCEEEEFEPSTQYE
YGTEDDYQGKPLEFGATSAALQPEEEQEEDWLDDDSQQTVGQQDGSEDNQTTTIQTIVEVQPQLE
MELTPVVQTIEVNSFSGYLKLTDNVYIKNADIVEEAKKVKPTVVVNAANVYLKHGGGVAGALNKA
TNNAMQVESDDYIATNGPLKVGGSCVLSGHNLAKHCLHVVGPNVNKGEDIQLLKSAYENFNQHEV
LLAPLLSAGIFGADPIHSLRVCVDTVRTNVYLAVFDKNLYDKLVSSFLEMKSEKQVEQKIAEIPK
EEVKPFITESKPSVEQRKQDDKKIKACVEEVTTTLEETKFLTENLLLYIDINGNLHPDSATLVSD
IDITFLKKDAPYIVGDVVQEGVLTAVVIPTKKAGGTTEMLAKALRKVPTDNYITTYPGQGLNGYT
VEEAKTVLKKCKSAFYILPSIISNEKQEILGTVSWNLREMLAHAEETRKLMPVCVETKAIVSTIQ
RKYKGIKIQEGVVDYGARFYFYTSKTTVASLINTLNDLNETLVTMPLGYVTHGLNLEEAARYMRS
LKVPATVSVSSPDAVTAYNGYLTSSSKTPEEHFIETISLAGSYKDWSYSGQSTQLGIEFLKRGDK
SVYYTSNPTTFHLDGEVITFDNLKTLLSLREVRTIKVFTTVDNINLHTQVVDMSMTYGQQFGPTY
LDGADVTKIKPHNSHEGKTFYVLPNDDTLRVEAFEYYHTTDPSFLGRYMSALNHTKKWKYPQVNG
LTSIKWADNNCYLATALLTLQQIELKFNPPALQDAYYRARAGEAANFCALILAYCNKTVGELGDV
RETMSYLFQHANLDSCKRVLNVVCKTCGQQQTTLKGVEAVMYMGTLSYEQFKKGVQIPCTCGKQA
TKYLVQQESPFVMMSAPPAQYELKHGTFTCASEYTGNYQCGHYKHITSKETLYCIDGALLTKSSE
YKGPITDVFYKENSYTTTIKPVTYKLDGVVCTEIDPKLDNYYKKDNSYFTEQPIDLVPNQPYPNA
SFDNFKFVCDNIKFADDLNQLTGYKKPASRELKVTFFPDLNGDVVAIDYKHYTPSFKKGAKLLHK
PIVWHVNNATNKATYKPNTWCIRCLWSTKPVETSNSFDVLKSEDAQGMDNLACEDLKPVSEEVVE
NPTIQKDVLECNVKTTEVVGDIILKPANNSLKITEEVGHTDLMAAYVDNSSLTIKKPNELSRVLG
LKTLATHGLAAVNSVPWDTIANYAKPFLNKVVSTTTNIVTRCLNRVCTNYMPYFFTLLLQLCTFT
RSTNSRIKASMPTTIAKNTVKSVGKFCLEASFNYLKSPNFSKLINIIIWFLLLSVCLGSLIYSTA
ALGVLMSNLGMPSYCTGYREGYLNSTNVTIATYCTGSIPCSVCLSGLDSLDTYPSLETIQITISS
FKWDLTAFGLVAEWFLAYILFTRFFYVLGLAAIMQLFFSYFAVHFISNSWLMWLIINLVQMAPIS
AMVRMYIFFASFYYVWKSYVHVVDGCNSSTCMMCYKRNRATRVECTTIVNGVRRSFYVYANGGKG
FCKLHNWNCVNCDTFCAGSTFISDEVARDLSLQFKRPINPTDQSSYIVDSVTVKNGSIHLYFDKA
GQKTYERHSLSHFVNLDNLRANNTKGSLPINVIVFDGKSKCEESSAKSASVYYSQLMCQPILLLD
QALVSDVGDSAEVAVKMFDAYVNTFSSTFNVPMEKLKTLVATAEAELAKNVSLDNVLSTFISAAR
QGFVDSDVETKDVVECLKLSHQSDIEVTGDSCNNYMLTYNKVENMTPRDLGACIDCSARHINAQV
AKSHNIALIWNVKDFMSLSEQLRKQIRSAAKKNNLPFKLTCATTRQVVNVVTTKIALKGG

- Alternative name: papain-like proteinase

- Alternative name: papain-like proteinase
- Responsible for the cleavages located at the N-terminus of the replicase polyprotein. In addition, PL-PRO possesses a deubiquitinating/deISGylating activity and processes both 'Lys-48'- and 'Lys-63'-linked polyubiquitin chains from cellular substrates. Participates together with nsp4 in the assembly of virally induced cytoplasmic double-membrane vesicles necessary for viral replication. Antagonises innate immune induction of type I interferon by blocking the phosphorylation, dimerisation and subsequent nuclear translocation of host IRF3. Prevents also host NF-kappa-B signaling. [https://swissmodel.expasy.org/repository/species/2697049]. As indicated on the expasy schema, it's cleaving the polyprotein between nsp1 & 2, nsp2 and itself and itself and nsp4.
- Large, multi-domain transmembrane protein, activities include: [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4369385/ general review on coronaviruses]
    - Ubl1 and Ac domains, interact with N protein
    - ADRP activity, promotes cytokine expression
    - PLPro/Deubiquitinase domain, cleaves viral polyprotein and blocks host innate immune response
    - Ubl2, NAB, G2M, SUD, Y domains, unknown functions
- 3D-Structures
    - One of the domains covered by experimental structures

nsp4 (500 amino acids):

- Residue 2764-3263 in orf1ab polyprotein

  KIVNNWLKQLIKVTLVFLFVAAIFYLITPVHVMSKHTDFSSEIIGYKAIDGGVTRDIAST
  DTCFANKHADFDTWFSQRGGSYTNDKACPLIAAVITREVGFVVPGLPGTILRTTNGDFLH
  FLPRVFSAVGNICYTPSKLIEYTDFATSACVLAAECTIFKDASGKPVPYCYDTNVLEGSV
  AYESLRPDTRYVLMDGSIIQFPNTYLEGSVRVVTTFDSEYCRHGTCERSEAGVCVSTSGR
  WVLNNDYYRSLPGVFCGVDAVNLLTNMFTPLIQPIGALDISASIVAGGIVAIVVTCLAYY
  FMRFRRAFGEYSHVVAFNTLLFLMSFTVLCLTPVYSFLPGVYSVIYLYLTFYLTNDVSFL
  AHIQWMVMFTPLVPFWITIAYIICISTKHFYWFFSNYLKRRVVFNGVSFSTFEEAALCTF
  LLNKEMYLKLRSDVLLPLTQYNRYLALYNKYKYFSGAMDTTSYREAACCHLAKALNDFSN
  SGSDVLYQPPQTSITSAVLQ

- Participates in the assembly of virally-induced cytoplasmic double-membrane vesicles necessary for viral replication [https://swissmodel.expasy.org/repository/species/2697049, https://jvi.asm.org/content/83/19/10314]
- 3D-Structures
  - No known experimental structures yet

nsp5 (306 amino acids):

- Residue 3264-3569 in orf1ab polyprotein

  SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDVVYCPRHVICTSEDMLNPNYEDLLIR
  KSNHNFLVQAGNVQLRVIGHSMQNCVLKLKVDTANPKTPKYKFVRIQPGQTFSVLACYNG
  SPSGVYQCAMRPNFTIKGSFLNGSCGSVGFNIDYDCVSFCYMHHMELPTGVHAGTDLEGN
  FYGPFVDRQTAQAAGTDTTITVNVLAWLYAAVINGDRWFLNRFTTTLNDFNLVAMKYNYE
  PLTQDHVDILGPLSAQTGIAVLDMCASLKELLQNGMNGRTILGSALLEDEFTPFDVVRQC
  SGVTFQ

- Alternative names: 3C-like proteinase, 3CLpro, Mpro (that's the protease a lot of people are after right now)
- Cleaves the C-terminus of replicase polyprotein at 11 sites. Recognizes substrates containing the core sequence [ILMVF]-Q-|-[SGACN]. Also able to bind an ADP-ribose-1'-phosphate (ADRP).
- 3D-Structures:
  - a lot of structures are available with different fragments bound into the active site

nsp6 (290 amino acids):

- Residue 3570-3859 in orf1ab polyprotein

  SAVKRTIKGTHHWLLLTILTSLLVLVQSTQWSLFFFLYENAFLPFAMGIIAMSAFAMMFV
  KHKHAFLCLFLLPSLATVAYFNMVYMPASWVMRIMTWLDMVDTSLSGFKLKDCVMYASAV
  VLLILMTARTVYDDGARRVWTLMNVLTLVYKVYYGNALDQAISMWALIISVTSNYSGVVT
  TVMFLARGIVFMCVEYCPIFFITGNTLQCIMLVYCFLGYFCTCYFGLFCLLNRYFRLTLG
  VYDYLVSTQEFRYMNSQGLLPPKNSIDAFKLNIKLLGVGGKPCIKVATVQ

- Plays a role in the initial induction of autophagosomes from host reticulum endoplasmic. Later, limits the expansion of these phagosomes that are no longer able to deliver viral components to lysosomes.
- 3D-Structures:
  - no known experimental structures

nsp7 (63 amino acids):

- Residue 3860-3942 in orf1ab polyprotein

  SKMSDVKCTSVVLLSVLQQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEKMVSLLSVLLS MQGAVDINKLCEEMLDNRATLQ

- Cofactor with nsp8 and nsp12
- Forms a hexadecamer with nsp8 (8 subunits of each) that may participate in viral replication by acting as a primase. Alternatively, may synthesize substantially longer products than oligonucleotide primers. May be required to activate RNA-synthesizing activity of Pol [https://swissmodel.expasy.org/repository/species/2697049 , https://jvi.asm.org/content/83/19/10314]
- 3D-Structures:
  - no known experimental structures

nsp8 (198 amino acids):

- Residue 3943-4140 in orf1ab polyprotein

  AIASEFSSLPSYAAFATAQEAYEQAVANGDSEVVLKKLKKSLNVAKSEFDRDAAMQRKLEK
  MADQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALNNIINNARDGCVPLNIIPL
  TTAAKLMVVIPDYNTYKNTCDGTTFTYASALWEIQQVVDADSKIVQLSEISMDNSPNLAWP
  LIVTALRANSAVKLQ

- Cofactor with nsp7 and nsp12, primase
- Forms a hexadecamer with nsp7 (8 subunits of each) that may participate in viral replication by acting as a primase. Alternatively, may synthesize substantially longer products than oligonucleotide primers. May be required to activate RNA-synthesizing activity of Pol [https://swissmodel.expasy.org/repository/species/2697049 , https://jvi.asm.org/content/83/19/10314]
- 3D-Structures:
  - no known experimental structures

nsp9 (113 amino acids):

- Residue 4141-4253 in orf1ab polyprotein

  NNELSPVALRQMSCAAGTTQTACTDDNALAYYNTTKGGRFVLALLSDLQDLKWARFPKSDG TGTIYTELEPPCRFVTDTPKGPKVKYLYFIKGLNNLNRGMVLGSLAATVRLQ

- May participate in viral replication by acting as a ssRNA-binding protein.
- Dimerization [https://swissmodel.expasy.org/repository/species/2697049 , https://jvi.asm.org/content/83/19/10314]
- 3D-Structures:
  - experimental structures are known for the full sequence

nsp10 (139 amino acids):

- Residue 4254-4392 in orf1ab polyprotein

  AGNATEVPANSTVLSFCAFAVDAAKAYKDYLASGGQPITNCVKMLCTHTGTGQAITVTPEA NMDQESFGGASCCLYCRCHIDHPNPKGFCDLKGKYVQIPTTCANDPVGFTLKNTVCTVCGM WKGYGCSCDQLREPMLQ

- Plays a pivotal role in viral transcription by stimulating both nsp14 3'-5' exoribonuclease and nsp16 2'-O-methyltransferase activities. Therefore plays an essential role in viral mRNAs cap methylation [https://swissmodel.expasy.org/repository/species/2697049]
- Cofactor for nsp16 and nsp14, forms heterodimer with both and stimulates ExoN and 2-O-MT activity [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4369385/]
- 3D-Structures:
  - experimental structures are available of the full domain

nsp11 (13 amino acids):

- Residue 4393-4405 in orf1a polyprotein
- Becomes nsp12 in orf1ab polyprotein

nsp12 (932 amino acids):

- Residue 4393-5324 in orf1ab polyprotein

SADAQSFLNRVCGVSAARLTPCGTGTSTDVVYRAFDIYNDKVAGFAKFLKTNCCRFQEKDE
DDNLIDSYFVVKRHTFSNYQHEETIYNLLKDCPAVAKHDFFKFRIDGDMVPHISRQRLTKY
TMADLVYALRHFDEGNCDTLKEILVTYNCCDDDYFNKKDWYDFVENPDILRVYANLGERVR
QALLKTVQFCDAMRNAGIVGVLTLDNQDLNGNWYDFGDFIQTTPGSGVPVVDSYYSLLMPI
LTLTRALTAESHVDTDLTKPYIKWDLLKYDFTEERLKLFDRYFKYWDQTYHPNCVNCLDDR
CILHCANFNVLFSTVFPPTSFGPLVRKIFVDGVPFVVSTGYHFRELGVVHNQDVNLHSSRL
SFKELLVYAADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQTVKPGNFNKDFYDFAVSKG
FFKEGSSVELKHFFFAQDGNAAISDYDYYRYNLPTMCDIRQLLFVVEVVDKYFDCYDGGCI
NANQVIVNNLDKSAGFPFNKWGKARLYYDSMSYEDQDALFAYTKRNVIPTITQMNLKYAIS
AKNRARTVAGVSICSTMTNRQFHQKLLKSIAATRGATVVIGTSKFYGGWHNMLKTVYSDVE
NPHLMGWDYPKCDRAMPNMLRIMASLVLARKHTTCCSLSHRFYRLANECAQVLSEMVMCGG
SLYVKPGGTSSGDATTAYANSVFNICQAVTANVNALLSTDGNKIADKYVRNLQHRLYECLY
RNRDVDTDFVNEFYAYLRKHFSMMILSDDAVVCFNSTYASQGLVASIKNFKSVLYYQNNVF
MSEAKCWTETDLTKGPHEFCSQHTMLVKQGDDYVYLPYPDPSRILGAGCFVDDIVKTDGTL
MIERFVSLAIDAYPLTKHPNQEYADVFHLYLQYIRKLHDELTGHMLDMYSVMLTNDNTSRY
WEPEFYEAMYTPHTVLQ

- Alternative names:
  - RNA directed RNA polymerase Pol/RdRp
- Responsible for replication and transcription of the viral RNA genome. [https://swissmodel.expasy.org/repository/species/2697049]
- 3D-Structures:
  - no known experimental structure

nsp13 (601 amino acids):

- Residue 5325-5925 in orf1ab polyprotein

```
AVGACVLCNSQTSLRCGACIRRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGCDVTDVT
QLYLGGMSYYCKSHKPPISFPLCANGQVFGLYKNTCVGSDNVTDFNAIATCDWTNAGDYIL
ANTCTERLKLFAAETLKATEETFKLSYGIATVREVLSDRELHLSWEVGKPRPPLNRNYVFT
GYRVTKNSKVQIGEYTFEKGDYGDAVVYRGTTTYKLNVGDYFVLTSHTVMPLSAPTLVPQE
HYVRITGLYPTLNISDEFSSNVANYQKVGMQKYSTLQGPPGTGKSHFAIGLALYYPSARIV
YTACSHAAVDALCEKALKYLPIDKCSRIIPARARVECFDKFKVNSTLEQYVFCTVNALPET
TADIVVFDEISMATNYDLSVVNARLRAKHYVYIGDPAQLPAPRTLLTKGTLEPEYFNSVCR
LMKTIGPDMFLGTCRRCPAEIVDTVSALVYDNKLKAHKDKSAQCFKMFYKGVITHDVSSAI
NRPQIGVVREFLTRNPAWRKAVFISPYNSQNAVASKILGLPTQTVDSSQGSEYDYVIFTQT
TETAHSCNVNRFNVAITRAKVGILCIMSDRDLYDKLQFTSLEIPRRNVATLQ
```

- Helicase (Hel)
- Multi-functional protein with a zinc-binding domain in N-terminus displaying RNA and DNA duplex-unwinding activities with 5' to 3' polarity. . Activity of helicase is dependent on magnesium. [https://swissmodel.expasy.org/repository/species/2697049]
- 3D-Structures:
  - no known experimental structure

nsp14 (527 amino acids):

- Residue 5926-6452 in orf1ab polyprotein

  AENVTGLFKDCSKVITGLHPTQAPTHLSVDTKFKTEGLCVDIPGIPKDMTYRRLISMMGFK
  MNYQVNGYPNMFITREEAIRHVRAWIGFDVEGCHATREAVGTNLPLQLGFSTGVNLVAVPT
  GYVDTPNNTDFSRVSAKPPPGDQFKHLIPLMYKGLPWNVVRIKIVQMLSDTLKNLSDRVVF
  VLWAHGFELTSMKYFVKIGPERTCCLCDRRATCFSTASDTYACWHHSIGFDYVYNPFMIDV
  QQWGFTGNLQSNHDLYCQVHGNAHVASCDAIMTRCLAVHECFVKRVDWTIEYPIIGDELKI
  NAACRKVQHMVVKAALLADKFPVLHDIGNPKAIKCVPQADVEWKFYDAQPCSDKAYKIEEL
  FYSYATHSDKFTDGVCLFWNCNVDRYPANSIVCRFDTRVLSNLNLPGCDGGSLYVNKHAFH
  TPAFDKSAFVNLKQLPFFYYSDSPCESHGKQVVSDIDYVPLKSATCITRCNLGGAVCRHHA
  NEYRLYLDAYNMMISAGFSLWVYKQFDTYNLWNTFTRLQ

- Alternative names:
  - Guanine-N7 methyltransferase
  - ExoN/nsp14
- Enzyme possessing two different activities: an exoribonuclease activity acting on both ssRNA and dsRNA in a 3' to 5' direction and a N7-guanine methyltransferase activity.
- [https://swissmodel.expasy.org/repository/species/2697049]
- 3D-Structures:
  - no known experimental structure

nsp15 (346 amino acids):

- Residue 6453-6798 in orf1ab polyprotein

  SLENVAFNVVNKGHFDGQQGEVPVSIINNTVYTKVDGVDVELFENKTTLPVNVAFELWAKR
  NIKPVPEVKILNNLGVDIAANTVIWDYKRDAPAHISTIGVCSMTDIAKKPTETICAPLTVF
  FDGRVDGQVDLFRNARNGVLITEGSVKGLQPSVGPKQASLNGVTLIGEAVKTQFNYYKKVD
  GVVQQLPETYFTQSRNLQEFKPRSQMEIDFLELAMDEFIERYKLEGYAFEHIVYGDFSHSQ
  LGGLHLLIGLAKRFKESPFELEDFIPMDSTVKNYFITDAQTGSSKCVCSVIDLLLDDFVEI
  IKSQDLSVVSKVVKVTIDYTEISFMLWCKDGHVETFYPKLQ

- Alternative names:
  - Uridylate-specific endoribonuclease
  - NendoU/nsp15
- Mn(2+)-dependent, uridylate-specific enzyme, which leaves 2'-3'-cyclic phosphates 5' to the cleaved bond.[https://swissmodel.expasy.org/repository/species/2697049]
- 3D-Structures:
  - Experimental structures available

nsp16 (298 amino acids):

- Residue 6799-7096 in orf1ab polyprotein

  SSQAWQPGVAMPNLYKMQRMLLEKCDLQNYGDSATLPKGIMMNVAKYTQLCQYLNTLTLAV
  PYNMRVIHFGAGSDKGVAPGTAVLRQWLPTGTLLVDSDLNDFVSDADSTLIGDCATVHTAN
  KWDLIISDMYDPKTKNVTKENDSKEGFFTYICGFIQQKLALGGSVAIKITEHSWNADLYKL
  MGHFAWWTAFVTNVNASSSEAFLIGCNYLGKPREQIDGYVMHANYIFWRNTNPIQLSSYSL
  FDMSKFPLKLRGTAVMSLKEGQINDMILSLLSKGRLIIRENNRVVISSDVLVNN

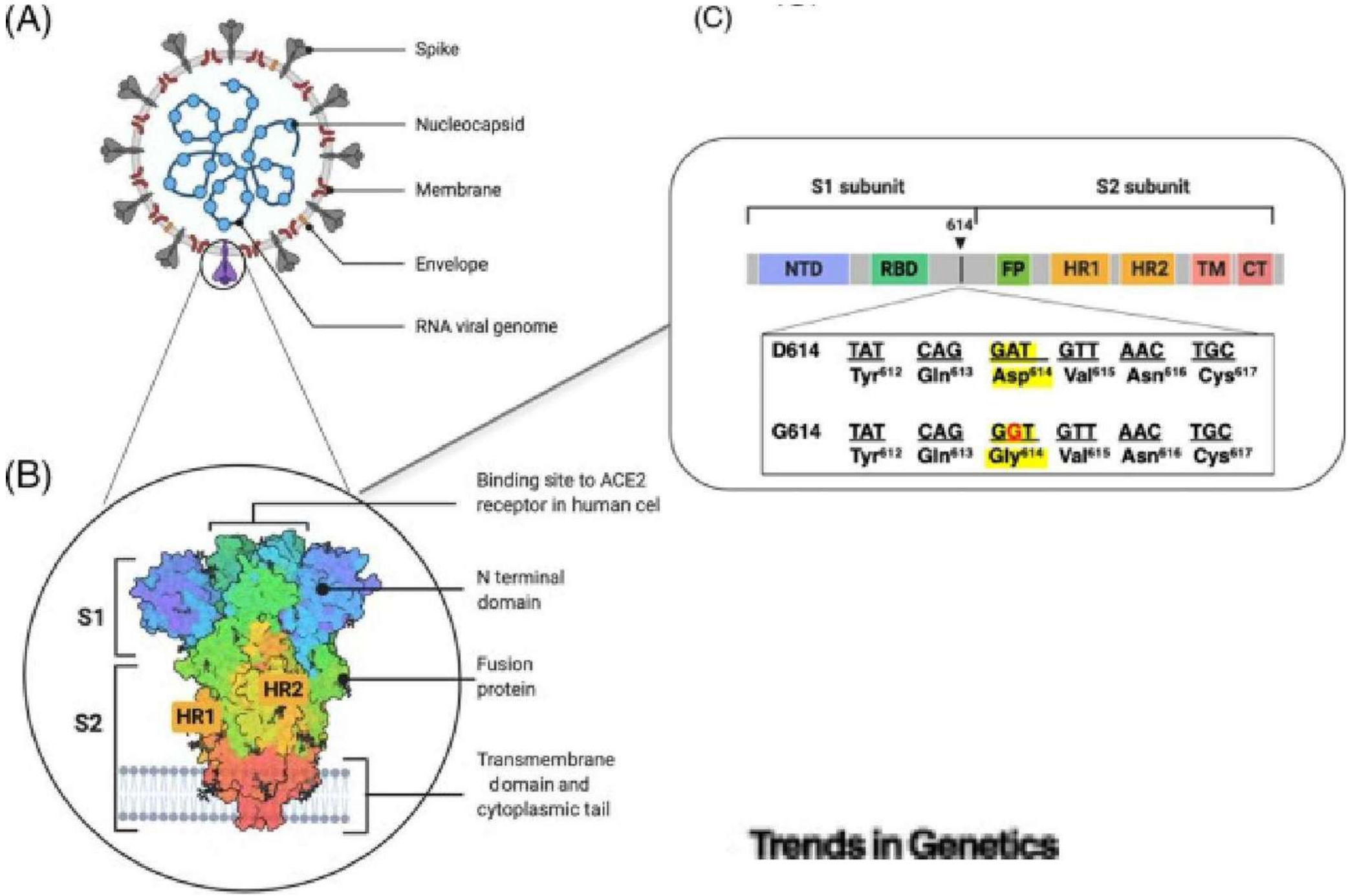
- Alternative names:
  - 2'-O-ribose methyltransferase
- Methyltransferase that mediates mRNA cap 2'-O-ribose methylation to the 5'-cap structure of viral mRNAs. N7-methyl guanosine cap is a prerequisite for binding of nsp16. Therefore plays an essential role in viral mRNAs cap methylation which is essential to evade immune system.
- 2'-O-MTase; avoiding MDA5 recognition, negatively regulating innate immunity
- [https://swissmodel.expasy.org/repository/species/2697049]
- 3D-Structures:
  - experimental structures are available

Spike glycoprotein

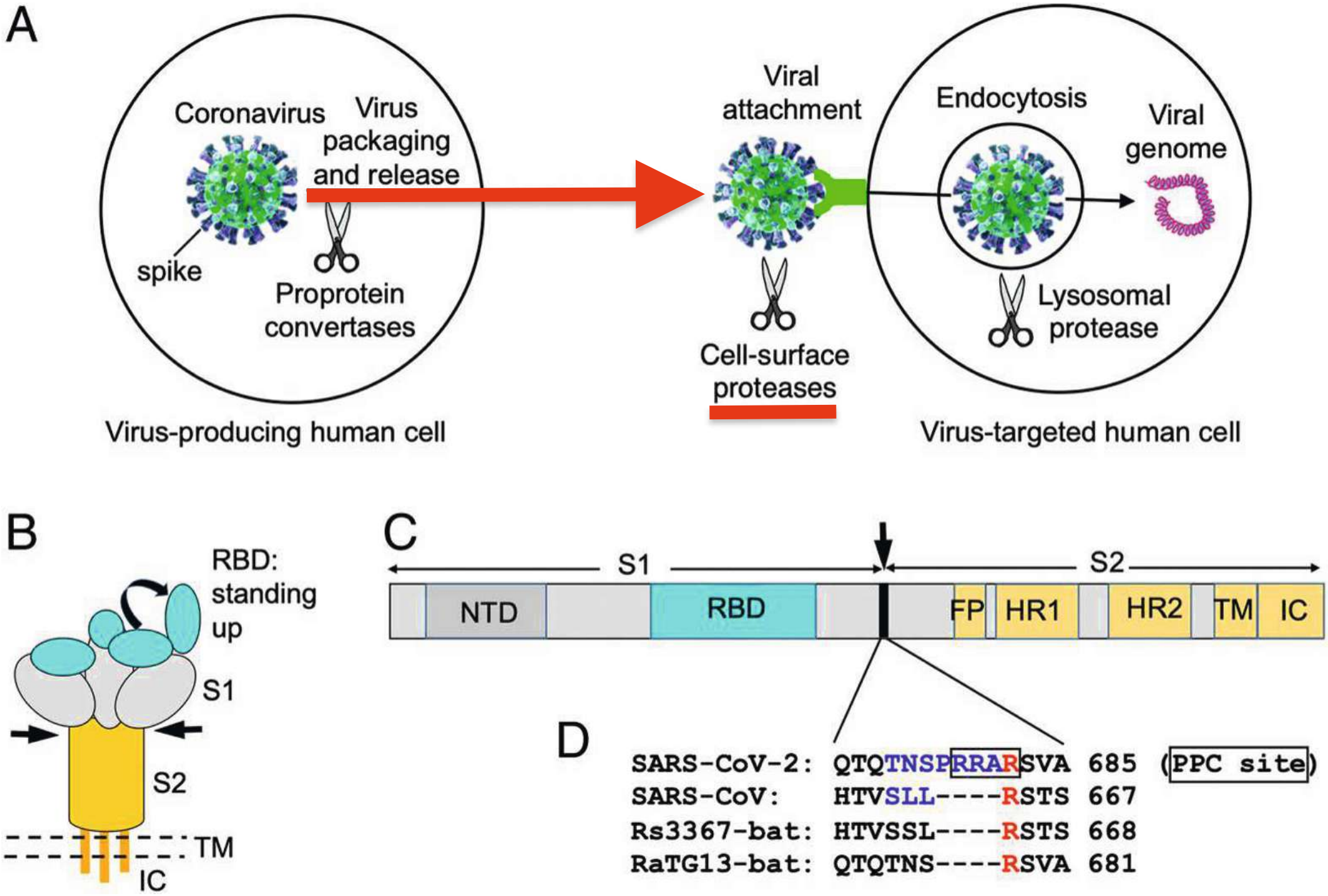
MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSN
VTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNN
ATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQ
GNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLLAL
HRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKS
FTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYS
VLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPD
DFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNC
YFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGT
GVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAV
LYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSYECDIPIGAGIC
ASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMT
KTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPP
IKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQK
FNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNV
LYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALNTLVKQLSSNFGAISSV
LNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQS
KRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSN
GTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNH
TSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGF
IAGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT

- Spike protein S1 (residue 14-685): attaches the virion to the cell membrane by interacting with host receptor, initiating the infection. Binding to human ACE2 and CLEC4M/DC-SIGNR receptors and internalization of the virus into the endosomes of the host cell induces conformational changes in the S glycoprotein. Proteolysis by cathepsin CTSL may unmask the fusion peptide of S2 and activate membranes fusion within endosomes.

- Spike protein S1 (residue 14-685): attaches the virion to the cell membrane by interacting with host receptor, initiating the infection. Binding to human ACE2 and CLEC4M/DC-SIGNR receptors and internalization of the virus into the endosomes of the host cell induces conformational changes in the S glycoprotein. Proteolysis by cathepsin CTSL may unmask the fusion peptide of S2 and activate membranes fusion within endosomes.
- Spike protein S2 (residue 686-1273): mediates fusion of the virion and cellular membranes by acting as a class I viral fusion protein. Under the current model, the protein has at least three conformational states: pre-fusion native state, pre-hairpin intermediate state, and post-fusion hairpin state. During viral and target cell membrane fusion, the coiled coil regions (heptad repeats) assume a trimer-of-hairpins structure, positioning the fusion peptide in close proximity to the C-terminal region of the ectodomain. The formation of this structure appears to drive apposition and subsequent fusion of viral and target cell membranes.
- Spike protein S2' (residue 816-1273): acts as a viral fusion peptide which is unmasked following S2 cleavage occurring upon virus endocytosis.
- https://swissmodel.expasy.org/repository/species/2697049, https://zhanglab.ccmb.med.umich.edu/C-I-TASSER/2019-nCov/, https://www.ncbi.nlm.nih.gov/protein/1791269090



Trends in Genetics


A
Coronavirus
Virus packaging and release
spike
Proprotein convertases
Virus-producing human cell
Viral attachment
Cell-surface proteases
Endocytosis
Viral genome
Lysosomal protease
Virus-targeted human cell
B
RBD: standing up
S1
S2
TM
IC
C
S1
S2
NTD
RBD
FP
HR1
HR2
TM
IC
D
SARS-CoV-2: QTQTNSPRRARSVA 685 (PPC site)
SARS-CoV: HTVSLL----RSTS 667
Rs3367-bat: HTVSSL----RSTS 668
RaTG13-bat: QTQTNS----RSVA 681

Protein 3a

MDLFMRIFTIGTVTLKQGEIKDATPSDFVRATATIPIQASLPFGWLIVGVALLAVFQSASK
IITLKKRWQLALSKGVHFVCNLLLLFVTVYSHLLLVAAGLEAPFLYLYALVYFLQSINFVR
IIMRLWLCWKCRSKNPLLYDANYFLCWHTNCYDYCIPYNSVTSSIVITSGDGTTSPISEHD
YQIGGYTEKWESGVKDCVVLHSYFTSDYYQLYSTQLSTDTGVEHVTFFIYNKIVDEPEEHV
QIHTIDGSSGVVNPVMEPIYDEPTTTTSVPL

- Forms homotetrameric potassium sensitive ion channels (viroporin) and may modulate virus release. Up-regulates expression of fibrinogen subunits FGA, FGB and FGG in host lung epithelial cells. Induces apoptosis in cell culture. Downregulates the type 1 interferon receptor by inducing serine phosphorylation within the IFN alpha- receptor subunit 1 (IFNAR1) degradation motif and increasing IFNAR1 ubiquitination.
- https://www.ncbi.nlm.nih.gov/protein/1791269091
- 3D-Structures:
  - no known experimental structure

ns6 (ORF6):

MFHLVDFQVTIAEILLIIMRTFKVSIWNLDYIINLIIKNLSKSLTENKYSQLDEEQPMEID

- Could be a determinant of virus virulence, since, when expressed in an otherwise attenuated JHM strain of murine coronavirus, it can dramatically increase the lethality of the latter. Seems to stimulate cellular DNA synthesis in vitro.
- https://www.ncbi.nlm.nih.gov/protein/1791269094
- 3D-Structures:
  - no known experimental structure

Protein 7a (ns7):

MKIILFLALITLATCELYHYQECVRGTTVLLKEPCSSGTYEGNSPFHPLADNKFALTCFST
QFAFACPDGVKHVYQLRARSVSPKLFIRQEEVQELYSPIFLIVAAIVFITLCFTLKRKTE

- Non-structural protein which is dispensable for virus replication in cell culture. Suppression of host tetherin activity.
- https://www.ncbi.nlm.nih.gov/protein/1791269095
- 3D-Structures:
  - no known experimental structure

Protein 7b

- NA

ns 8 (ORF8)

```
MKFLVFLGIITTVAAFHQECSLQSCTQHQPYVVDDPCPIHFYSKWYIRVGARKSAPLIELC
VDEAGSKSPIQYIDIGNYTVSCLPFTINCQEPKLGSLVVRCSFYEDFLEYHDVRVVLDFI
```

- May play a role in host-virus interaction.
- https://www.ncbi.nlm.nih.gov/protein/1791269096
- 3D-Structures:
  - no known experimental structure

ORF10 protein

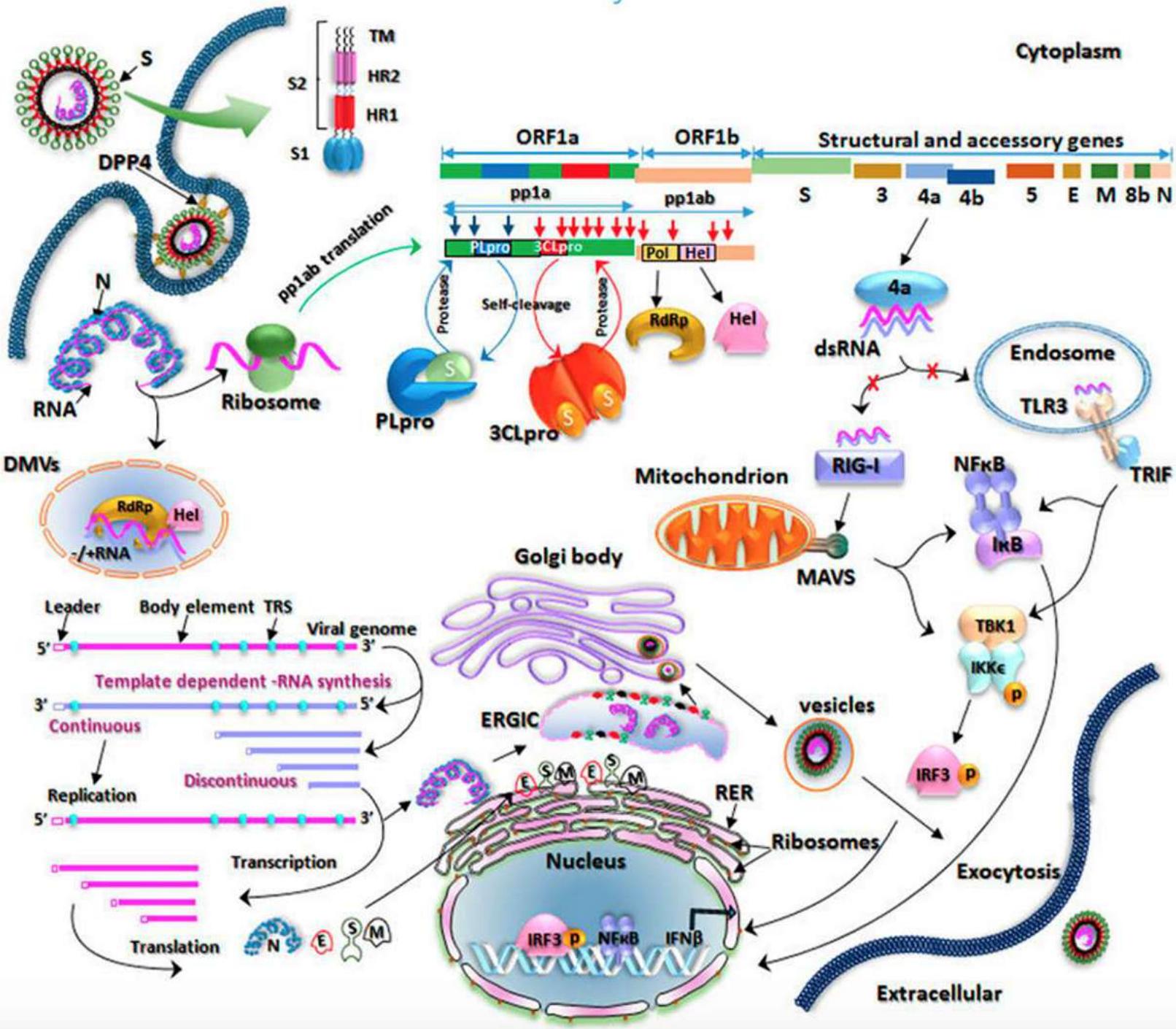
```
MGYINVFAFPFTIYSLLLCRMNSRNYIAQVDVVNFNLT
```

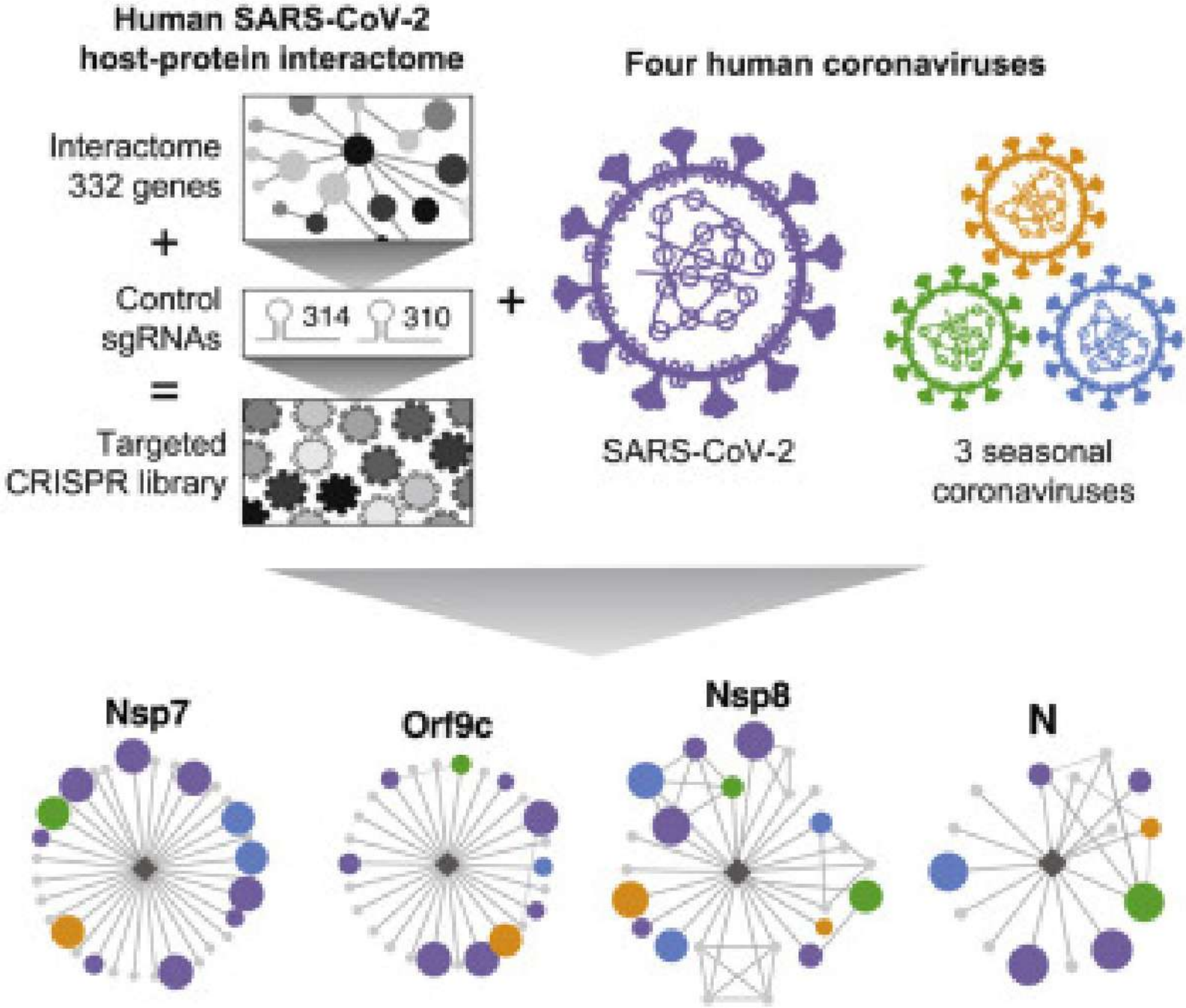
- Dubious/uncertain protein. It is currently unclear whether this region translates into a functional protein. The respective region in related beta-coronaviruses, such as SARS-CoV, does not correspond to translated peptide.
- https://www.ncbi.nlm.nih.gov/protein/1798172433
- 3D-Structures:
  - no known structures

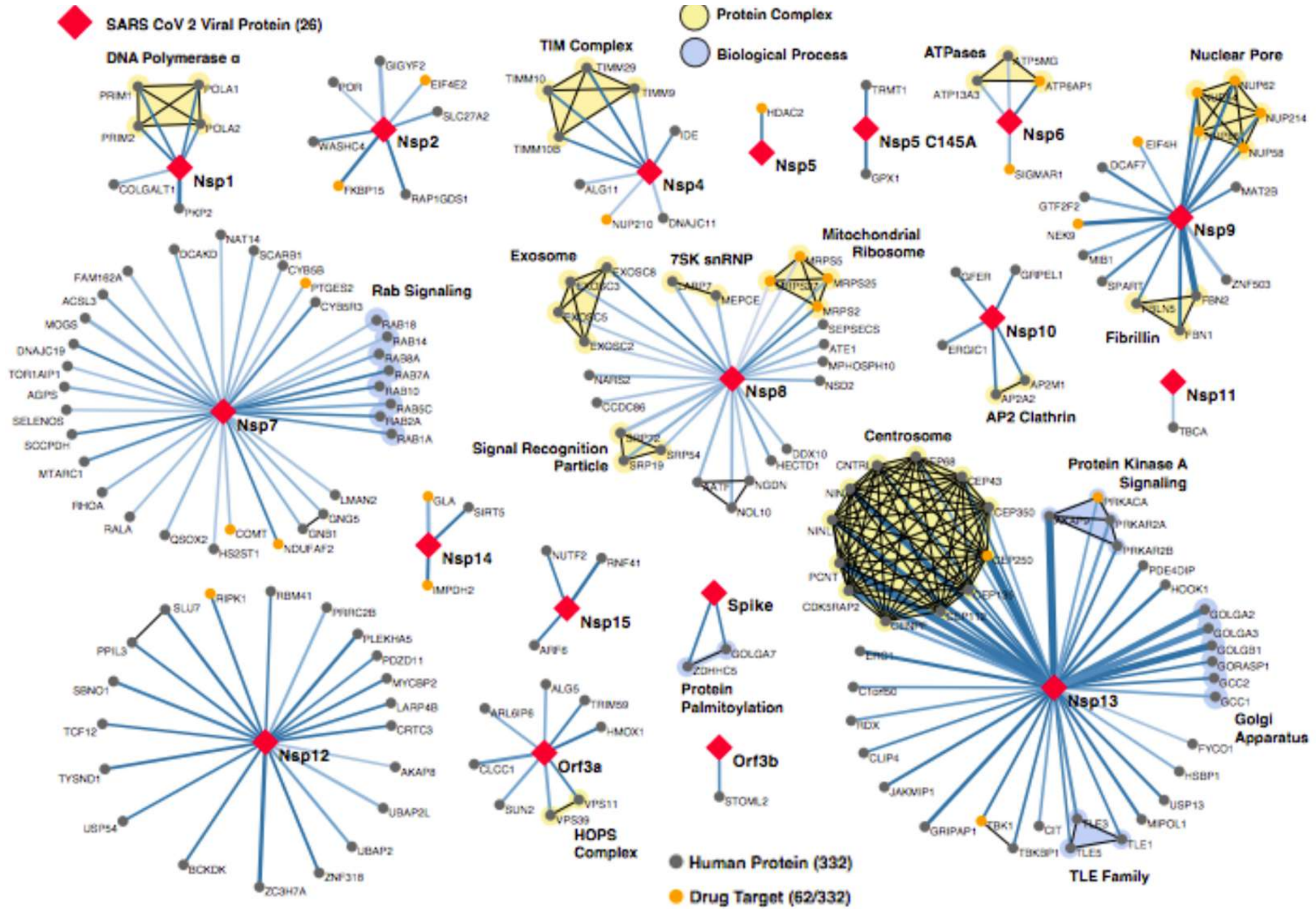
The following proteins have been mentioned as of high interest (and priority) for drug discovery programs for antivirals because of the obvious important function for the virus:

- Mpro (the main protease, nsp5) (structures available)
- papain-like protease (nsp3) (structures available)
- helicase (nsp13) (good models available)
- RNA dependent RNA polymerase (nsp12) (good models available)
- spike glycoprotein (structures available)

The lifecycle of the virus




Human SARS-CoV-2
host-protein interactome
Four human coronaviruses
Interactome
332 genes
+
Control
sgRNAs
314
310
=
Targeted
CRISPR library
+
SARS-CoV-2
3 seasonal
coronaviruses
Nsp7
Orf9c
Nsp8
N
Functional virus-host interactome networks


SARS CoV 2 Viral Protein (26)
Protein Complex
Biological Process
DNA Polymerase α
PRIM1
POLA1
PRIM2
POLA2
Nsp1
COLGALT1
PKP2
GIGYF2
POR
EIF4E2
SLC27A2
Nsp2
WASHC4
FKBP15
RAP1GDS1
TIM Complex
TIMM29
TIMM10
TIMM9
IDE
TIMM10B
Nsp4
ALG11
NUP210
DNAJC11
HDAC2
Nsp5
TRMT1
Nsp5 C145A
GPX1
ATPases
ATP5MG
ATP13A3
ATP6AP1
Nsp6
SIGMAR1
Nuclear Pore
NUP62
NUP214
NUP58
EIF4H
DCAF7
MAT2B
GTF2F2
NEK9
Nsp9
MIB1
SPART
ZNF503
TSLN5
FBN2
FBN1
Fibrillin
NAT14
DCAKD
SCARB1
CYB5B
FAM162A
PTGES2
ACSL3
CYB5R3
Rab Signaling
MOGS
RAB18
DNAJC19
RAB14
TOR1AIP1
RAB8A
RAB7A
AGPS
RAB10
RAB5C
SELENOS
RAB2A
SCCPDH
RAB1A
MTARC1
Nsp7
RHOA
LMAN2
RALA
GNG5
DNB1
QSOX2
COMT
NDUFAF2
HS2ST1
Exosome
EXOSC8
EXOSC3
EXOSC5
EXOSC2
7SK snRNP
LARP7
MEPCE
Mitochondrial Ribosome
MRPS5
MRPS27
MRPS25
MRPS2
SEPSECS
ATE1
MPHOSPH10
NSD2
NAF1/52
Nsp8
CCDC86
DDX10
HECTD1
Signal Recognition Particle
SRP72
SRP54
SRP19
NGDN
NOL10
GLA
SIRT5
Nsp14
IMPDH2
NUTF2
RNF41
Nsp15
ANP32E
Spike
GOLGA7
ZDHHC5
Protein Palmitoylation
Orf3b
STOML2
ALG5
ARL6IP6
TRIM59
HMOX1
CLCC1
Orf3a
SUN2
VPS11
VPS39
HOPS Complex
C2orf49
CNTRL
CEP68
CEP43
NIN
CEP350
NINL
FOR20
CEP250
PCNT
CEP192
CDK5RAP2
CEP131
Centrosome
Protein Kinase A Signaling
PRKACA
PRKAR2A
PRKAR2B
PDE4DIP
HOOK1
GOLGA2
GOLGA3
GOLGB1
GORASP1
GCC2
GCC1
Golgi Apparatus
FYCO1
HSBP1
USP13
MIPOL1
TLE1
TLE5
TLE3
CIT
TBK1
TBKBP1
GRIPAP1
TLE Family
SLU7
RIPK1
RBM41
PRRC2B
PPIL3
PLEKHA5
PDZD11
SBNO1
MYCBP2
LARP4B
TCF12
CRTC3
Nsp12
AKAP8
TYSND1
UBAP2L
USP54
UBAP2
BCKDK
ZNF318
ZC3H7A
Human Protein (332)
Drug Target (62/332)

Envelope small protein (E protein)

MYSFVSEETGTLIVNSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYSR
VKNLNSSRVPDLLV

- Plays a central role in virus morphogenesis and assembly. Acts as a viroporin and self-assembles in host membranes forming pentameric protein-lipid pores that allow ion transport. Also plays a role in the induction of apoptosis. Activates the host NLRP3 inflammasome, leading to IL-1beta overproduction.
- https://www.ncbi.nlm.nih.gov/protein/1791269092
- 3D-Structures:
  - no known experimental structure

Membrane protein (M protein)

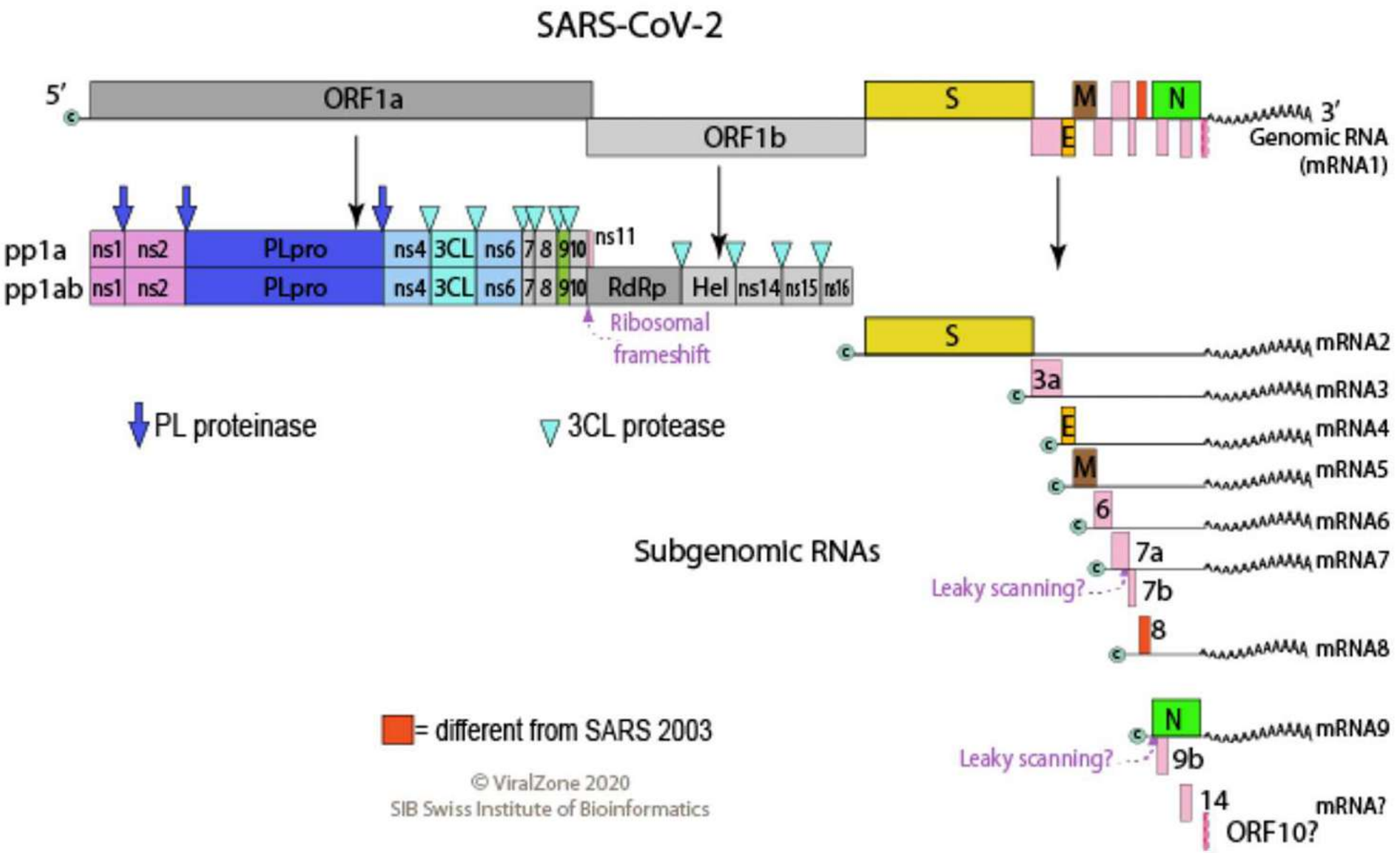
MADSNGTITVEELKKLLEQWNLVIGFLFLTWICLLQFAYANRNRFLYIIKLIFLWLLWPVT
LACFVLAAVYRINWITGGIAIAMACLVGLMWLSYFIASFRLFARTRSMWSFNPETNILLNV
PLHGTILTRPLLESELVIGAVILRGHLRIAGHHLGRCDIKDLPKEITVATSRTLSYYKLGA
SQRVAGDSGFAAYSRYRIGNYKLNTDHSSSSDNIALLVQ

- Component of the viral envelope that plays a central role in virus morphogenesis and assembly via its interactions with other viral protein
- https://www.ncbi.nlm.nih.gov/protein/1791269093
- 3D-Structures:
  - no known experimental structure

N protein

MSDNGPQNQRNAPRITFGGPSDSTGSNQNGERSGARSKQRRPQGLPNNTASWFTALTQHGK
EDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEAGLP
YGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQLPQGTTLPKGFYAEGSRGGSQAS
SRSSSRSRNSSRNSTPGSSRGTSPARMAGNGGDAALALLLLDRLNQLESKMSGKGQQQQGQ
TVTKKSAAEASKKPRQKRTATKAYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYKHWPQIA
QFAPSASAFFGMSRIGMEVTPSGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDAYKTFPPT
EPKKDKKKKADETQALPQRQKKQQTVTLLPAADLDDFSKQLQQSMSSADSTQA

- Packages the positive strand viral genome RNA into a helical ribonucleocapsid (RNP) and plays a fundamental role during virion assembly through its interactions with the viral genome and membrane protein M. Plays an important role in enhancing the efficiency of subgenomic viral RNA transcription as well as viral replication.
- https://www.ncbi.nlm.nih.gov/protein/1798172432
- 3D-Structures:
  - one domain is resolved experimentally



The SARS-CoV-2 genome structure as available on viralzone.expasy.org