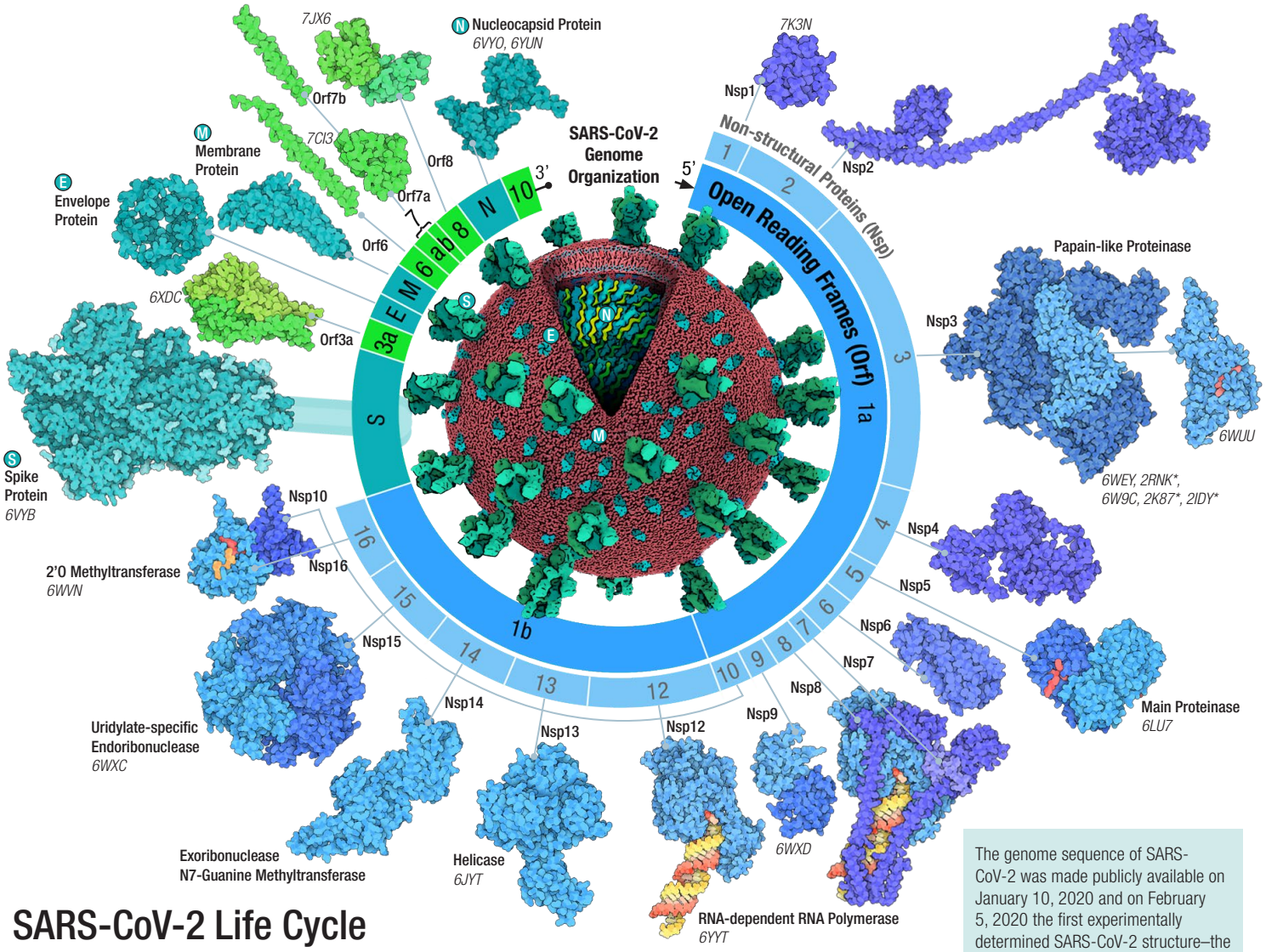
SARS-CoV-2 Genome and Proteins

The genome of SARS-CoV-2 is a single strand of RNA encoding a large collection of proteins that are synthesized by the ribosomal machinery after the virus infects a host cell.

Four **STRUCTURAL PROTEINS (S, E, M, N)** package the **RNA GENOME** within a **LIPID BILAYER MEMBRANE** to form infectious viral particles. Sixteen **NON-STRUCTURAL PROTEINS (Nsp 1-16)** and six **ACCESSORY PROTEINS** facilitate construction and release of new viral particles and support evasion of cell defenses and the host immune system.



SARS-CoV-2 Life Cycle

ENTERING THE CELL

The S-protein (spike protein) supports cell entry during infection by binding to Angiotensin Converting Enzyme 2, which is found on the surface of most human cells.

GENERATING THE REPLICATION/ TRANSLATION COMPLEX (RTC)

Inside the virus, the RNA genome is packaged by the nucleocapsid protein (N). Upon entry into the host cell, the RNA is released and serves as functional mRNA for the host ribosomes. First, the Orf1a and Orf1b are translated resulting in synthesis of two long polyprotein chains, encompassing most of the non-structural proteins (Nsps). Two enzymes (Nsp3 and Nsp5) act as molecular scissors, cleaving the polyproteins into 16 functional Nsp proteins. The Nsps then form the Replication/ Translation Complex (RTC). Some of the Nsp proteins modify the internal cellular membranes to anchor the complex to membranes within the host cell, others interfere with cellular defense mechanisms or the host immune response. At the heart of the RTC is the RNA-dependent RNA polymerase (Nsp12), which functions with several other Nsp proteins and host factors to replicate the viral RNA. It first creates a new viral RNA template, which is later copied multiple times to create new copies of the genomic RNA.

SYNTHESIZING NEW VIRUSES

Some of the new RNA strands are translated by host cell ribosomes to produce the building blocks for assembling new viral particles, including the S, M, E, and N proteins. At this time the accessory proteins are also synthesized. The M proteins interact first with S and E proteins to create scaffolds for the new virions and later with N proteins and the RNA genome to form the nucleocapsid. Finally, new viruses bud from the surface of the cell and are released to infect new cells.

Reference: R. Yadav *et al.* (2021) Role of Structural and Non-Structural Proteins and Therapeutic Targets of SARS-CoV-2 for COVID-19. *Cells* **10**: 821. doi.org/10.3390/cells10040821

Graphic adapted from: J. H. Lubin *et al.* (2021) Evolution of the SARS-CoV-2 proteome in three dimensions (3D) during the first 6 months of the COVID-19 pandemic. *Proteins: Structure, Function, and Bioinformatics.* doi.org/10.1002/prot.26250

The genome sequence of SARS-CoV-2 was made publicly available on January 10, 2020 and on February 5, 2020 the first experimentally determined SARS-CoV-2 structure–the main proteinase–was released into the Protein Data Bank archive. Since then, the number of SARS-COV-2 structures in the PDB has grown exponentially surpassing 1,500 in October 2021. The illustration presents examples of SARS-COV-2 proteins available in the PDB, (shown with the PDB ID, e.g. 6LU7), models developed based on other SARS structures in the PDB (shown with PDB ID and *, e.g. 2RNK*) and additional computational models developed based on the genome and other structural information available. These structures provide insight on the mechanisms of COVID infection and are currently used in vaccination and therapeutics design.

RCSB PDB-101

pdb101.rcsb.org