SARS-CoV-2 Infection


European Bioinformatics Institute, New York University Langone Medical Center, Ontario Institute for Cancer Research, Oregon Health and Science University.

The contents of this document may be freely copied and distributed in any media, provided the authors, plus the institutions, are credited, as stated under the terms of Creative Commons Attribution 4.0 International (CC BY 4.0) License. For more information see our license.

26/03/2022


**Introduction**

Reactome is open-source, open access, manually curated and peer-reviewed pathway database. Pathway annotations are authored by expert biologists, in collaboration with Reactome editorial staff and cross-referenced to many bioinformatics databases. A system of evidence tracking ensures that all assertions are backed up by the primary literature. Reactome is used by clinicians, geneticists, genomics researchers, and molecular biologists to interpret the results of high-throughput experimental studies, by bioinformaticians seeking to develop novel algorithms for mining knowledge from genomic studies, and by systems biologists building predictive models of normal and disease variant pathways.

The development of Reactome is supported by grants from the US National Institutes of Health (P41 HG003751), University of Toronto (CFREF Medicine by Design), European Union (EU STRP, EMI-CD), and the European Molecular Biology Laboratory (EBI Industry program).

**Literature references**


Reactome database release: 79

This document contains 6 pathways (see Table of Contents)

https://reactome.org
This pathway, SARS-CoV-2 infection of human cells (COVID-19), was initially generated via electronic inference from the manually curated and reviewed Reactome SARS-CoV-1 (Human SARS coronavirus) infection pathway. The inference process created SARS-CoV-2 events corresponding to each event in the SARS-CoV-1 pathway and populated those events with SARS-CoV-2 protein-containing physical entities based on orthology to SARS-CoV-1 proteins (https://reactome.org/documentation/inferred-events). All of these computationally created events and entities have been reviewed by Reactome curators and modified as appropriate where recently published experimental data indicate the existences of differences between the molecular details of the SARS-CoV-1 and SARS-CoV-2 infection pathways.

SARS-CoV-2 infection begins with the binding of viral S (spike) protein to cell surface angiotensin converting enzyme 2 (ACE2) and endocytosis of the bound virion. Within the endocytic vesicle, host proteases mediate cleavage of S protein into S1 and S2 fragments, leading to S2-mediated fusion of the viral and host endosome membranes and release of the viral capsid into the host cell cytosol. The capsid is uncoated to free the viral genomic RNA, whose cap-dependent translation produces polyprotein pp1a and, by means of a 1â€ʻbase frameshift, polyprotein pp1ab. Autoproteolytic cleavage of pp1a and pp1ab generates 15 or 16 nonstructural proteins (nsps) with various functions. Importantly, the RNA dependent RNA polymerase (RdRP) activity is encoded in nsp12. Nsp3, 4, and 6 induce rearrangement of the cellular endoplasmic reticulum membrane to form cytosolic double membrane vesicles (DMVs) where the viral replication transcription complex is assembled and anchored. With viral genomic RNA as a template, viral replicaseâ€ʻtranscriptase synthesizes a full length negative sense antigenome, which in turn serves as a template for the synthesis of new genomic RNA. The replicaseâ€ʻtranscriptase can also switch template during discontinuous transcription of the genome at transcription regulated sequences to produce a nested set of negativeâ€ʻsense subgenomic (sg) RNAs, which are used as templates for the synthesis of posit-
iveâ€™sense sgRNAs that are translated to generate viral proteins. Finally, viral particle assembly occurs in the ER Golgi intermediate compartment (ERGIC). Viral M protein provides the scaffold for virion morphogenesis (Hartenian et al. 2020; Fung & Liu 2019; Masters 2006).

**Literature references**


**Editions**

<table>
<thead>
<tr>
<th>Date</th>
<th>Action</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020-06-15</td>
<td>Authored</td>
<td>Varusai, TM.</td>
</tr>
<tr>
<td>2020-07-13</td>
<td>Authored</td>
<td>Jassal, B.</td>
</tr>
<tr>
<td>2020-07-22</td>
<td>Authored</td>
<td>D'Eustachio, P., Stephan, R.</td>
</tr>
<tr>
<td>2020-07-28</td>
<td>Authored</td>
<td>Orlic-Milacic, M.</td>
</tr>
<tr>
<td>2020-07-30</td>
<td>Authored</td>
<td>Rothfels, K.</td>
</tr>
<tr>
<td>2020-08-12</td>
<td>Authored</td>
<td>Senff-Ribeiro, A.</td>
</tr>
<tr>
<td>2020-08-14</td>
<td>Authored</td>
<td>Gillespie, ME.</td>
</tr>
<tr>
<td>2020-09-09</td>
<td>Reviewed</td>
<td>Acencio, ML.</td>
</tr>
<tr>
<td>2020-09-09</td>
<td>Edited</td>
<td>Gillespie, ME.</td>
</tr>
<tr>
<td>2020-09-10</td>
<td>Authored</td>
<td>Haw, R.</td>
</tr>
</tbody>
</table>
Attachment and Entry ↗

Location: SARS-CoV-2 Infection

Stable identifier: R-HSA-9694614

Diseases: COVID-19

Inferred from: Attachment and Entry (Homo sapiens)

This COVID-19 event has been created by a combination of computational inference (see https://reactome.org/documentation/inferred-events) from SARS-CoV-1 data and manual curation, as described in the summation for the overall SARS-CoV-2 infection pathway.

Coronavirus replication is initiated by the binding of S protein to the cell surface receptor(s). The S protein is composed of two functional domains, S1 (bulb) which mediates receptor binding and S2 (stalk) which mediates membrane fusion. Specific interaction between S1 and the cognate receptor triggers a drastic conformational change in S2, leading to fusion between the virus envelope and the cellular membrane and release of the viral nucleocapsid into the host cell cytosol. Receptor binding is the major determinant of the host range and tissue tropism for a coronavirus. Some human coronaviruses (HCoVs) have adopted cell surface enzymes as receptors, angiotensin converting enzyme 2 (ACE2) for SARS-CoV-2, SARS-CoV-1, and HCoV NL63. The receptor-bound S protein is activated by cleavage into S1 and S2, mediated by one of two of two host proteases, the endosomal cysteine protease cathepsin L and another trypsin like serine protease. Type II transmembrane serine proteases TMPRSS2 and TMPRSS11D have also been implicated in the activation of S protein of SARS-CoV-2. Host factors may play additional roles in viral entry (not annotated here). Valosin containing protein (VCP) contributes by a poorly understood mechanism to the release of coronavirus from early endosomes. Host factors may also restrict the attachment and entry of HCoV.

Literature references

<table>
<thead>
<tr>
<th>Date</th>
<th>Action</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020-09-09</td>
<td>Reviewed</td>
<td>Acencio, ML.</td>
</tr>
<tr>
<td>2020-09-09</td>
<td>Authored, Edited</td>
<td>Gillespie, ME.</td>
</tr>
</tbody>
</table>
Translation of Replicase and Assembly of the Replication Transcription Complex

Location: SARS-CoV-2 Infection

Stable identifier: R-HSA-9694676

Diseases: COVID-19

Inferred from: Translation of Replicase and Assembly of the Replication Transcription Complex (Homo sapiens)

This COVID-19 pathway has been created by a combination of computational inference from SARS-CoV-1 data (https://reactome.org/documentation/inferred-events) and manual curation, as described in the summation for the overall SARS-CoV-2 infection pathway.

After entry and uncoating, the genomic RNA serves as a transcript to allow cap dependent translation of ORF1a to produce polyprotein pp1a. A slippery sequence and an RNA pseudoknot near the end of ORF1a enable 25-30% of ribosomes to undergo -1 frameshifting, to continue translation of ORF1b to produce a longer polyprotein pp1ab. Autoproteolytic cleavage of pp1a and pp1ab generates 15-16 nonstructural proteins (nsps) with various functions. RNA dependent RNA polymerase (RdRP) activity is encoded in nsp12, and papain like protease (PLPro) and main protease (Mpro) activities are encoded in nsp3 and nsp5, respectively. nsp3, 4, and 6 induce rearrangement of the cellular membrane to form double membrane vesicles (DMVs) where the coronavirus replication transcription complex (RTC) is assembled and anchored.

Programmed ribosomal frameshifting (PRF) may be regulated by viral or host factors in addition to viral RNA secondary structures. For example, PRF in the related arterivirus porcine reproductive and respiratory syndrome virus (PRRSV) is transactivated by the viral protein nsp1, which interacts with the PRF signal via a putative RNA binding motif. A host RNA-binding protein called annexin A2 (ANXA2) binds the pseudoknot structure in the IBV genome. Host factors in the early secretory pathway appear to be involved in DMV formation and RTC assembly: Golgi specific brefeldin A resistance guanine nucleotide exchange factor 1 (GBF1) and its effector ADP ribosylation factor 1 (ARF1) are both required for normal
DMV formation and efficient RNA replication of mouse hepatitis virus (MHV), a prototypic betacoronavirus that infects mice (Fung & Liu 2019).

**Literature references**


**Editions**

<table>
<thead>
<tr>
<th>Date</th>
<th>Action</th>
<th>Author/Editor</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020-08-28</td>
<td>Authored, Edited</td>
<td>Stephan, R.</td>
</tr>
<tr>
<td>2020-09-09</td>
<td>Reviewed</td>
<td>Acencio, ML.</td>
</tr>
</tbody>
</table>
SARS-CoV-2 Genome Replication and Transcription

**Location:** SARS-CoV-2 Infection

**Stable identifier:** R-HSA-9694682

**Diseases:** COVID-19

This COVID-19 pathway has been created by a combination of computational inference from SARS-CoV-1 data (https://reactome.org/documentation/inferred-events) and manual curation, as described in the summation for the overall SARS-CoV-2 infection pathway. Specifically, binding of the replication-transcription complex (RTC) to the RNA template and the polymerase activity of nsp12 (Hillen et al. 2020, Wang et al. 2020, Yin et al. 2020), helicase activity of nsp13 (Chen et al. 2020, Ji et al. 2020, Shu et al. 2020), capping activity of nsp16 (Viswanathan et al. 2020), and polyadenylation of SARS-CoV-2 genomic RNA and transcripts (Kim et al. 2020, Ravindra et al. 2020) have been studied directly, and the remaining steps have been inferred from previous studies in SARS-CoV-1 and related coronaviruses.

Using the genomic RNA as a template, the coronavirus replicase synthesizes full-length negative-sense antigenome, which in turn serves as a template for the synthesis of new genomic RNA (Masters 2006). The polymerase can also switch template during discontinuous transcription of the genome at specific sites called transcription-regulated sequences, thereby producing a 5'-nested set of negative-sense sgRNAs, which are used as templates for the synthesis of a 3'-nested set of positive-sense sgRNAs (Masters 2006). Although genome replication/transcription is mainly mediated by the viral replicase and confines in the RTC, the involvement of various additional viral and host factors has been implicated. For instance, coronavirus N protein is known to serve as an RNA chaperone and facilitate template switching (Zúñiga et al. 2007, Zúñiga et al. 2010). Importantly, the N protein of SARS-CoV-1 and mouse hepatitis virus (MHV-JHM) is also phosphorylated by the host glycogen synthase kinase 3 (GSK3), and inhibition of GSK3 was shown to inhibit viral replication in Vero E6 cells infected with SARS-CoV-1 (Wu et al. 2009). Additionally, GSK3-mediated phosphorylation of the MHV-JHM N protein recruits an RNA-binding protein DEAD-box helicase 1 (DDX1), which facilitates template read-through, favoring the synthesis of genomic RNA and longer sgRNAs (Wu et al. 2014). Another RNA-binding protein called heterogeneous nucle-
ar ribonucleoprotein A1 (hnRNPA1) can also bind tightly to SARS-CoV-1 N protein and potentially regulate viral RNA synthesis (Luo et al. 2005). Host RNA-binding proteins could also bind directly to untranslated regions (UTRs) of the coronavirus genome to modulate replication/transcription, such as zinc-finger CCHC-type and RNA-binding motif 1 (ZCRB1) binding to the 5'-UTR of IBV (Tan et al. 2012), mitochondrial aconitase binding to the 3' UTR of MHV (Nanda and Leibowitz 2001), and poly(A)-binding protein (PABP) to the poly(A) tail of bovine coronavirus (Spagnolo and Hogue 2000). For review, please refer to Snijder et al. 2016 and Fung and Liu 2019.

**Literature references**


**Editions**

<table>
<thead>
<tr>
<th>Date</th>
<th>Action</th>
<th>Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020-08-12</td>
<td>Authored</td>
<td>Orlic-Milacic, M., Senff-Ribeiro, A.</td>
</tr>
<tr>
<td>2020-08-18</td>
<td>Edited</td>
<td>Orlic-Milacic, M.</td>
</tr>
<tr>
<td>2020-09-09</td>
<td>Reviewed</td>
<td>Acencio, ML.</td>
</tr>
</tbody>
</table>
Translation of Structural Proteins

Location: SARS-CoV-2 Infection

Stable identifier: R-HSA-9694635

Compartments: endoplasmic reticulum lumen, cytosol

Diseases: COVID-19

Inferred from: Translation of Structural Proteins (Homo sapiens)

This COVID-19 pathway has been created by a combination of computational inference from SARS-CoV-1 data (https://reactome.org/documentation/inferred-events) and manual curation, as described in the summation for the overall SARS-CoV-2 infection pathway.

Virus mRNA is translated according to the ribosomal scanning model. It is capped and polyadenylated, with regions of nontranslated sequences on both the 5' and 3' ends. Structural proteins are encoded after the polymerase/replicase genes by mRNAs 2 (Spike protein), 3, 4 (Envelope protein), 5 (Membrane protein), and 9. mRNA 3 and 9 are bicistronic, the proteins 3a and 9a (Nucleocapsid protein) having functions in virus assembly and structure. Translation happens in the ER with the exception of 9a which is translated by cytosolic free ribosomes (Fung and Liu, 2019).

Literature references


Editions

<table>
<thead>
<tr>
<th>Date</th>
<th>Action</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020-07-22</td>
<td>Authored</td>
<td>Stephan, R.</td>
</tr>
<tr>
<td>2020-08-28</td>
<td>Edited</td>
<td>Stephan, R.</td>
</tr>
<tr>
<td>2020-09-09</td>
<td>Reviewed</td>
<td>Acencio, ML.</td>
</tr>
</tbody>
</table>
Virion Assembly and Release

Location: SARS-CoV-2 Infection

Stable identifier: R-HSA-9694322

Diseases: COVID-19

Inferred from: Virion Assembly and Release (Homo sapiens)

This COVID-19 pathway has been created by a combination of computational inference from SARS-CoV-1 data (https://reactome.org/documentation/inferred-events) and manual curation, as described in the summation for the overall SARS-CoV-2 infection pathway.

Based on work done in related coronaviruses, CoV-2 viral assembly is expected to occur at the ERGIC membrane (reviewed in Masters, 2006; Fehr and Perlman, 2015; Fung and Liu, 2019). Membrane protein components of the virus concentrate at the ERGIC membrane but are also found throughout the secretory system including at the plasma membrane. Accumulation at the site of viral assembly has been shown to depend on interaction between retrieval signals in the cytoplasmic tails of viral proteins and host factors such as the COPI coat, and likely involves repeated rounds of anterograde and retrograde traffic (McBride et al, 2007; Ujike et al, 2016; Tan et al, 2004; Tan et al, 2005; reviewed in McBride and Fielding, 2012; Chang et al, 2014).

Viral assembly is initiated by homotypic interactions of M protein (Tseng et al, 2010; Siu et al, 2008). This forms an M-lattice that contributes to the induction of membrane curvature and additionally acts as a scaffold for the recruitment of the other structural components of the virus (Voss et al, 2009). M protein makes interactions with each of the main components of the mature virus, including E, S and N (He et al, 2004; Luo et al, 2006; Siu et al, 2008; reviewed in Masters, 2006). Electron micrographic studies suggest the final size of the mature virus is ~100 nm. The ribonuclear particle is predominantly helical and is packaged with an outer diameter of ~ 16 nm (Neuman et al, 2006; Neuman et al, 2011; reviewed in Chang et al, 2014). These physical constraints suggest a final stoichiometry in the mature virion of 75 S trimers:1200 M proteins:300 N:1 RNA genome (Neuman et al, 2011; reviewed in Chang et al, 2014). Minor
amounts of other viral proteins, including proteins E, 3a and 7a may also be components of the mature virus, although their functions are not well established (reviewed in Schoeman and Fielding, 2019; Liu et al, 2014).

**Literature references**


**Editions**

<table>
<thead>
<tr>
<th>Date</th>
<th>Action</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020-07-30</td>
<td>Authored, Edited</td>
<td>Rothfels, K.</td>
</tr>
<tr>
<td>2020-09-09</td>
<td>Reviewed</td>
<td>Acencio, ML.</td>
</tr>
</tbody>
</table>
Table of Contents

Introduction 1

- SARS-CoV-2 Infection 2
  - Attachment and Entry 4
  - Translation of Replicase and Assembly of the Replication Transcription Complex 6
- SARS-CoV-2 Genome Replication and Transcription 8
  - Translation of Structural Proteins 10
- Virion Assembly and Release 11

Table of Contents 13