SARS-CoV-2 Replication Model

Liam Fergusson and Tom Freeman

Abstract

SARS-CoV-2 is a recently emerged zoonotic beta-coronavirus and is the cause of a global pandemic. In response to this pandemic, an international collaborative effort has been established to produce models of the SARS-CoV-2 lifecycle and host responses to infection.

To date, we have focused our efforts on constructing a graphical and computational model of the SARS-CoV-2 lifecycle, covering viral entry into a host cell, the replication cycle of the virus and the release of daughter virions from the host cell. The model has been built in the yEd network editing software using the modified Edinburgh Pathway Notation (mEPN) scheme. The depiction of events described within the model has been produced from information available in the literature and informed by models produced by other members of the Disease Mapping Consortium, most notably models 703 and 704, which have provided guidance for the layout of certain systems covered by the current diagram.

In addition to supporting the graphical representation of biological pathways, the mEPN scheme supports computational modelling without simplification of the graphical elements. The modelling system is based on stochastic Petri nets where, in general, place nodes represent biological entities and transition nodes represent the interactions between them. Activity is represented by the rule-based flow of tokens. Simulations are performed within the tool BioLayout. To date, we have sought only to test the logic of the diagram, ensuring that infection of a cell by a few virus particles, leads to virus release.

Model Description

Cell entry:
SARS-CoV-2 cell entry is reliant on an interaction between the viral spike protein (S) and host cell surface proteases, most notably angiotensin converting enzyme 2 (ACE2)(1) and transmembrane protease, serine 2 (TMPRSS2)(2). These proteases prime the S homotrimer and facilitate viral entry either by endocytosis (ACE2) or membrane fusion, releasing the nucleocapsid into the cytoplasm (TMPRSS2)(3). These pathways are both represented in this model, diverging at the point of ACE2 binding. Cell entry can be inhibited by cell surface protease inhibitors such as camostat mesylate, a serine protease inhibitor(4).

The endocytotic pathway features cleavage of the ACE2:S complex by CTSB or CTSL to release the virion from the endosomal membrane. This then allows the viral membrane (M), envelope (E) and ORF7a proteins to mediate membrane fusion with the endosomal membrane, resulting in the release of the nucleocapsid into the cytoplasm. Both cell entry pathways re-converge at the released nucleocapsid node, as the entry mechanism is assumed to have no bearing on the mechanisms of viral replication beyond this stage in the model.

Viral Replication:
Once released into the cytoplasm, the nucleocapsid protein (N) is cleaved from the viral genomic RNA (gRNA) by the effector caspases 3 and 6, releasing the gRNA and allowing for

the first round of gRNA translation(5). This produces one of two polypeptides (pp) from the non-structural protein (nsp) coding open reading frames (ORFs) 1a & 1b that make up the 5' 2/3rds of the genome, pp1a or pp1ab(6). Translation of pp1ab occurs at approximately 1/3 of the frequency of pp1a, resulting in a lower concentration of its 4 specific nsps. The nsps have a number of roles within viral replication cycle, although they can be largely summarised as modulation of the host defences and formation of the replication/translation complex (RTC). All SARS-CoV-2 nsps have been implicated in the formation of the RTC, however the current model only lists those that form major functional components of the RTC(7). The RTC is anchored within a double membrane vesicle, requisitioned from various host organelles by nsp's 3, 4 and 6 which are also responsible for the anchoring of the RTC to these vesicles27/05/2020 18:12:00. The newly formed RTC then takes up the gRNA(+) of the infective virions and transcribes negative sense RNA templates. This is then translated into further nsps, viral structural proteins and accessory proteins, which go on to form daughter virions and RTCs. The negative sense template dissociates from the gRNA:RTC complex and is taken up by the daughter RTCs, where it is used to transcribe daughter gRNAs for daughter virion assembly and further transcription/translation cycles.

Virion assembly and release:
Structural and accessory proteins produced by the RTC cycle, excluding N, are translocated to the rough endoplasmic reticulum (RER) for post-translational modification by the host cell(8). N, however, oligomerises and is taken up by the RTC for processing events not yet defined in this model. N is then released from the RTC and complexes with gRNA produced by the RTC cycle to form daughter nucleocapsids(9). At present, this part of the pathway has been omitted from the model for visual clarity as the purpose of this process is not clear, producing what appears to be a redundant pathway, i.e. N binding then immediate dissociation. Future iterations of this model may include this pathway should its purpose become clear.

The nucleocapsids, structural and accessory proteins are translocated to the ER-Golgi intermediate complex after processing, where M, E, S and ORF7a mediate membrane rearrangement to enclose the nucleocapsid(10)(11)(12). This marks the formation of new virions which are transported to the cell surface for release within a smooth-walled vesicle.

For this section, post-translational modification, membrane rearrangement and nucleocapsid encapsulation, and virion release have been simplified to a single node each due to the complexity of each process. To accurately portray each process would result in a significant increase in the visual complexity of the model, potentially without a notable influence on its behaviour.
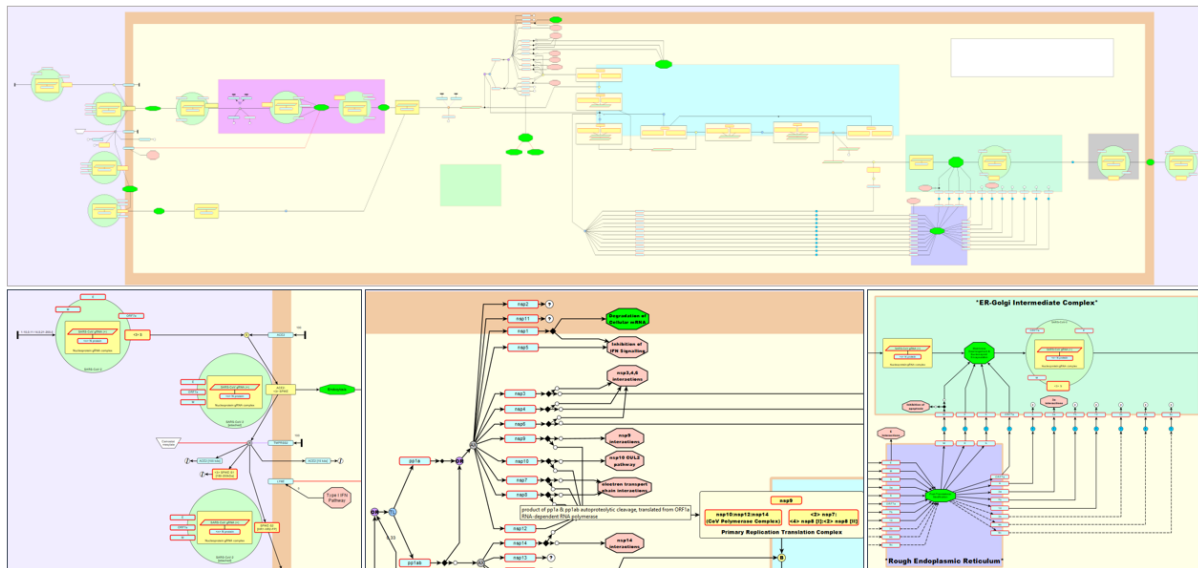
Top; Complete model of virus lifecycle and, bottom; detail of specific areas. Left; Viral cell entry, Centre; pp1a/pp1ab translation and autoproteolytic cleavage into the non-structural proteins, Right; production of structural proteins prior to assembly of new virus particles.

Red pathway nodes are hyperlinked to other Disease Map Consortium Covid-19 related pathways, all host genes are hyperlinked to the NCBI gene database and viral proteins to NCBI protein database, and process nodes include PubMed IDs to work as evidence for given interactions. Editors notes are also included. To open and edit the model use yED (https://www.yworks.com/products/yed).

Computational Modelling
In order to use models drawn using mEPN for modelling a number of parameterisation steps must be undertaken:

1. Firstly, the model must have a bipartite structure, that is to say nodes must alternate between representing entities (places) and processes (transitions). On occasion this means adding spacer nodes, small white circles or small black diamonds, respectively.

2. The initial starting concentrations of components is determined by the number of tokens placed on entities that are required for the system to work. These are added to these entities by placing black rectangular transition nodes immediately upstream of nodes representing these entities and defining the number of tokens they will receive on the edge that connects them. Importantly, on the node representing the infective virus we have added in a small number of tokens (20) over 4-time blocks, mimicking infection of a cell by a few virus particles at a discrete time. For the rest of the model, 100 tokens have been placed on all other necessary entities without time block restrictions. This ensures a constant feed of tokens to these entities throughout a simulation.

3. Amplification or reduction of token flow at specific points can be achieved by annotating the edge following an internal transition node. In this way, we have been able to simulate a reduced production of nsp12-15 proteins (n = 0.33) and an amplification of the daughter RTC complexes, SARS-CoV gRNA(-) and structural

proteins following their production from the primary RTC complex (n = 100) on the current model.

4. In these diagrams the distance tokens have to travel represents time, longer distances requiring more time-blocks for tokens to reach downstream nodes. Therefore, if timings are known 'delay motifs' may be added. For instance, a node representing the translation of a protein represents a process that likely takes longer to complete than a phosphorylation event and, therefore, one might introduce a delay motif (a series of places and transitions) to represent this. So far these have not been added to the model as we do not know the timing of events, e.g. the transcription of the viral genome, production of viral proteins, etc. While the exact values could potentially improve the accuracy of this model, they are not necessary for this model to function. Once these values are defined, they can be easily added to future iterations of this model.

5. For simulations, models saved as .graphml files from yEd (its standard file format) may be imported into our open-source tool BioLayout (http://biolayout.org/) that presents options for performing simulations, e.g. number of time-blocks, no. of runs, nature of stochasticity model as well as providing a format to visualise the results as dynamic visualisations or static plots of token accumulation over time for selected nodes. It is possible to work fluidly between yEd and BioLayout, running simulations and, based on what you observe, editing the model and rerunning the simulation. Run times on this model (100 time-blocks, 50 runs) take less than 1 sec.
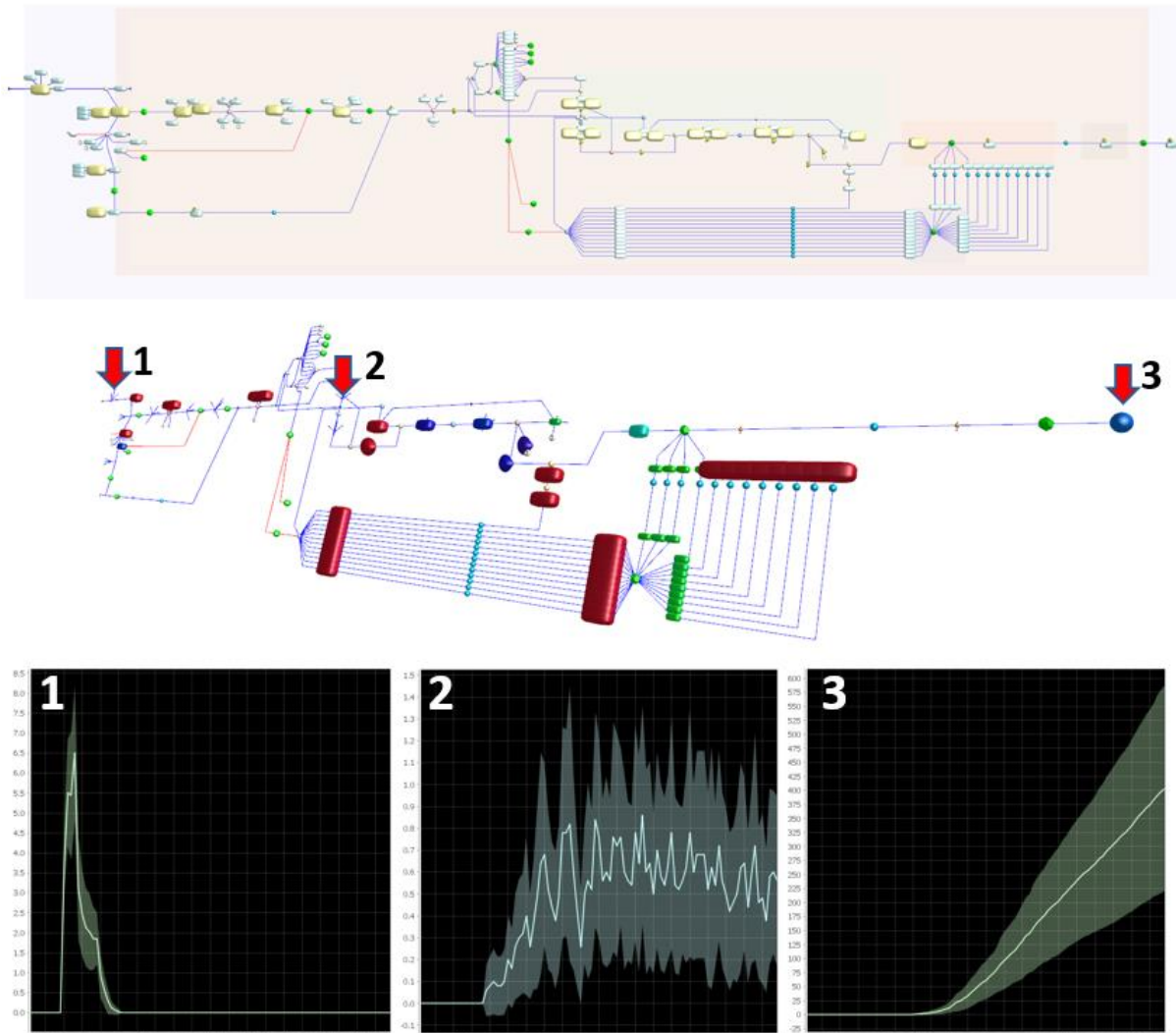
Top; *Full pathway loaded within BioLayout (compartments shown):* Middle; *Mid-run animation with node size/colour reflecting token level, with larger, red nodes approaching a user controlled token value: and* Bottom; *graphs of token flow per time-block, starting with the infective virions (left), primary RTC (center) and viral production (right).*

For more details of the mEPN modelling system: O'Hara *et al.*(13), and Livigni *et al*(14)*.*

Outstanding Questions

1. Current diagram shows virus entering by two means: Binding of the SPIKE protein complex and ACE2 leads to endocytosis and; cleavage of the of the SPIKE:ACE2 complex by TMPRSS2 leads to membrane fusion and nucleocapsid release. Are both pathways operating concurrently in the same cells or are they cell type specific? By endocytosis do we mean phagocytosis, e.g. by a macrophage or DC?

2. It is not clear from the literature whether the accessory proteins ORF3b, ORF9b and ORF9c are present in the SARS-CoV-2 genome, as various sources include or exclude these proteins with limited explanation(15). Clarification on this point would be useful.

3. Where available, the kinetics and time frames of interactions within the featured pathways could be included in the current model by the addition of delay motifs etc, thereby increasing the accuracy of the model. Is there any information on, for instance, how long it takes the virus to enter the cell, when is the genome replicated or when are newly synthesised viral proteins first produced?

4. Do we know anything about the relative productivity rates of the virus in an epithelial cell vs. a macrophage/DC? Do we know how long replication takes in these cell types?

5. Following translation of the new N proteins there is some evidence that they combine with the RTC complex before being released prior to their binding to newly synthesised viral RNA. If this does occur what is the purpose of the interaction between N protein and RTC?

6. As the machinery for gRNA and viral protein production are primarily located within the cytoplasm(16), the role of the host nucleus is currently unclear. Does it play a role in viral replication or is it limited to host cell responses to infection?

Further Development

The current model is lacking any representation of the host defence systems to inhibit viral replication or how the viral proteins subvert these pathways. Many of the accessory and non-structural proteins encoded by the SARS-CoV-2 genome play a wide range of roles in both viral replication and host defence modulation. The poorly defined or obscure nature of some of these protein's interactions means that for now we have chosen not represent them. Future iterations of this model will aim to incorporate these systems and we have included some 'markers' to other consortium maps that have sought to model these interactions.

References

1. Letko M, Marzi A, Munster V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. Nat Microbiol. 2020 Apr;5(4):562-+.

2. Reinke LM, Spiegel M, Plegge T, Hartleib A, Nehlmeier I, Gierer S, et al. Different residues in the SARS-CoV spike protein determine cleavage and activation by the host cell protease TMPRSS2. Plos One. 2017 Jun 21;12(6):e0179177.

3. Hoffmann M, Kleine-Weber H, Schroeder S, Krueger N, Herrler T, Erichsen S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell. 2020 Apr 16;181(2):271-+.

4. Zhou Y, Vedantham P, Lu K, Agudelo J, Carrion R Jr, Nunneley JW, et al. Protease inhibitors targeting coronavirus and filovirus entry. Antiviral Res. 2015/02/07. 2015 Apr;116:76–84.

5. Diemer C, Schneider M, Seebach J, Quaas J, Frösner G, Schätzl HM, et al. Cell Type-Specific Cleavage of Nucleocapsid Protein by Effector Caspases during SARS Coronavirus Infection. J Mol Biol. 2008 Feb 8;376(1):23–34.

6. Masters PS. The molecular biology of coronaviruses. In: Maramorosch K, Shatkin AJ, editors. Advances in Virus Research, Vol 66. 2006. p. 193-+.

7. Imbert I, Ulferts R, Ziebuhr J, Canard B. SARS Coronavirus Replicative Enzymes: Structures and Mechanisms. Lal SK, editor. Molecular Biology of the Sars-Coronavirus. 2010.

8. Fung TS, Liu DX. Post-translational modifications of coronavirus proteins: roles and function. Future Virol. 2018 Jun;13(6):405–30.

9. Surjit M, Lal SK. The Nucleocapsid Protein of the SARS Coronavirus: Structure, Function and Therapeutic Potential. Lal SK, editor. Molecular Biology of the Sars-Coronavirus. 2010.

10. KLUMPERMAN J, LOCKER J, MEIJER A, HORZINEK M, GEUZE H, ROTTIER P. CORONAVIRUS M-PROTEINS ACCUMULATE IN THE GOLGI-COMPLEX BEYOND THE SITE OF VIRION BUDDING. J Virol. 1994 Oct;68(10):6523–34.

11. Hogue BG, Machamer CE. Coronavirus Structural Proteins and Virus Assembly. Perlman S, Gallagher T, Snijder EJ, editors. Nidoviruses. 2008.

12. Lim KP, Liu DX. The missing link in coronavirus assembly - Retention of the avian coronavirus infectious bronchitis virus envelope protein in the pre-Golgi compartments and physical interaction between the envelope and membrane proteins. J Biol Chem. 2001 May 18;276(20):17515–23.

13. O'Hara L, Livigni A, Theo T, Boyer B, Angus T, Wright D, et al. Modelling the Structure and Dynamics of Biological Pathways. Plos Biol. 2016 Aug;14(8):e1002530.

14. Livigni A, O'Hara L, Polak ME, Angus T, Wright DW, Smith LB, et al. A graphical and computational modeling platform for biological pathways. Nat Protoc. 2018 Apr;13(4):705–22.

15. Xu J, Zhao S, Teng T, Abdalla AE, Zhu W, Xie L, et al. Systematic Comparison of Two Animal-to-Human Transmitted Human Coronaviruses: SARS-CoV-2 and SARS-CoV. Viruses-Basel. 2020 Feb;12(2):244.

16. Snijder EJ, van der Meer Y, Zevenhoven-Dobbe J, Onderwater JJM, van der Meulen J, Koerten HK, et al. Ultrastructure and origin of membrane vesicles associated with the severe acute respiratory syndrome coronavirus replication complex. J Virol. 2006 Jun;80(12):5927–40.

17. Agarwal A, Chen A, Ravindran N, To C, Thuluvath PJ. Gastrointestinal and Liver Manifestations of COVID-19. J Clin Exp Hepatol. 2020 Jun;10(3):263–5.