COVID-19 and SARS-COV-2 Research Update

Rick Stevens – Argonne National Laboratory – University of Chicago

Ian Foster, Arvind Ramanathan, Tom Brettin, Austin Clyde, Kyle Chard, Maulik Shukla, Jonathan Ozik, Chick Macal, Yadu Babuji, Carlos Olivares, Alex Partin, Gyorgy Babnigg, Xuefeng Liu, Ben Blaiszik, Ryan Chard, Zhi Hong, Zhuozhag Li, Marcus Schwarting, Logan Ward, Anna Woodward, Shantenu Jha, Kerstin Van Dam, Frank Alexander, Misha Salim, Matteo Turilli, Andre Merzky, Mike Papka, Katherine Riley, Venkat Vishwanath, Carlos Simmerling, Hubertus Van Dam, Gina Tourassi, Dan Stanzione, Mike Norman, and many others > 200
In 2003

The emergence of SARS and the identification of a coronavirus as the causative agent of the disease astounded the coronavirus community, as it was the first definitive association of a coronavirus with a severe disease in humans.
**Virus species**
- Middle East respiratory syndrome-related coronavirus
- Severe acute respiratory syndrome-related coronavirus

**ICTV-CSG**
- MERS-CoV
- SARS-CoV
- SARS-CoV-2

**WHO**
- Middle East respiratory syndrome (MERS)
- Severe acute respiratory syndrome (SARS)
- Coronavirus disease 2019 (COVID-19)

**Year**
- 2012
- 2003
- 2019

△ First name  →  Name origin
COVID-19 CORONAVIRUS OUTBREAK

Last updated: March 09, 2020, 15:01 GMT

https://www.worldometers.info/coronavirus/

Coronavirus Cases: 111,746
Deaths: 3,888
Recovered: 62,722

COVID-19 Fatality Rate by AGE:

*Death Rate = (number of deaths / number of cases) = probability of dying if infected by the virus (%).
This probability differs depending on the age group. The percentages shown below do not have to add up to 100%, as they do NOT represent share of deaths by age group. Rather, it represents, for a person in a given age group, the risk of dying if infected with COVID-19.

<table>
<thead>
<tr>
<th>AGE</th>
<th>DEATH RATE confirmed cases</th>
<th>DEATH RATE all cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>80+ years old</td>
<td>21.9%</td>
<td>14.8%</td>
</tr>
<tr>
<td>70-79 years old</td>
<td>8.0%</td>
<td></td>
</tr>
<tr>
<td>60-69 years old</td>
<td>3.8%</td>
<td></td>
</tr>
<tr>
<td>50-59 years old</td>
<td>1.6%</td>
<td></td>
</tr>
<tr>
<td>40-49 years old</td>
<td>0.4%</td>
<td></td>
</tr>
<tr>
<td>30-39 years old</td>
<td>0.2%</td>
<td></td>
</tr>
<tr>
<td>20-29 years old</td>
<td>0.2%</td>
<td></td>
</tr>
<tr>
<td>10-19 years old</td>
<td>0.2%</td>
<td></td>
</tr>
<tr>
<td>0-9 years old</td>
<td>no fatalities</td>
<td></td>
</tr>
</tbody>
</table>

PRE-EXISTING CONDITION | DEATH RATE confirmed cases | DEATH RATE all cases |
------------------------|---------------------------|---------------------|
Cardiovascular disease  | 13.2%                     | 10.5%               |
Diabetes                | 9.2%                      | 7.3%                |
Chronic respiratory disease | 8.0%                  | 6.3%                |
Hypertension            | 8.4%                      | 6.0%                |
Cancer                  | 7.6%                      | 5.6%                |
no pre-existing conditions |                         | 0.9%                |

view by country

Daily Deaths

Daily Deaths
Deaths per Day
Data as of 0.00 GMT 8

© Daily Deaths
Coronavirus Cases: 2,744,606
Deaths: 191,790
Recovered: 755,390
Understanding the Enemy
Size & Content

Diameter: \( \approx 100 \text{ nm} \)

Volume: \( \approx 10^6 \text{ nm}^3 = 10^3 \text{ fL} \)

Mass: \( \approx 10^3 \text{ MDa} \approx 1 \text{ fg} \)

Spike trimer
Length: \( \approx 10 \text{ nm} \)
Copies per virion: \( \approx 100 \)
(300 monomers, measured for SARS-CoV-1)
Affinity to ACE2
receptor \( K_d \approx 1-30 \text{ nM} \)
primed by TMPRSS2

Membrane protein
\( \approx 2000 \text{ copies} \)
(measured for SARS-CoV-1)

Nucleoprotein
\( \approx 1000 \text{ copies} \)
(measured for SARS-CoV-1)

Envelope protein
\( \approx 20 \text{ copies} \)
(100 monomers, measured for TGEV coronavirus)
Replication Timescales

in tissue-culture

Virion entry into cell: \( \sim 10 \text{ min} \) (measured for SARS-CoV-1)

Eclipse period: \( \sim 10 \text{ hrs} \) (time to make intracellular virions)

Burst size: \( \sim 10^3 \) virions (measured for MHV coronavirus)
Nucleotide identity to SARS-CoV-2

- bat CoV: 96%
- pangolin CoV: 91%
- SARS-CoV-1: 80%
- MERS: 55%
- common cold CoV: 50%

Length: ≈30kb; β-coronavirus with 10-14 ORFs (24-27 proteins)

Evolution rate: $\sim 10^{-3}$ nt$^{-1}$ yr$^{-1}$ (measured for SARS-CoV-1)

Mutation rate: $\sim 10^{-6}$ nt$^{-1}$ cycle$^{-1}$ (measured for MHV coronavirus)
Polyproteins encode many domains per gene

Published online 2014 Dec 29. doi: 10.1016/j.antiviral.2014.12.015
SARS-CoV-2 most likely came from bats perhaps via an intermediate host

SARS-CoV-1 bats ⇒ civet ⇒ humans
MERS-CoV bats ⇒ camel ⇒ humans
"Characteristic" Infection Progression in a Single Patient

Basic reproductive number $R_0$: typically 2-4
Varies further across space and time (Li et al. 2020; Park et al. 2020)
(number of new cases directly generated from a single case)

- Exposure
- Incubation period: ≈5 days (99% ≤ 14 days unless asymptomatic)
- Diagnosis after ≈5 days
- Symptomatic
- Infectious

Case Fatality Rate (ECDC 2020)
≈0.8%-10% (uncorrected)

Infected Fatality Rate
≈0.3%-1.3%

Recovery
- Mild cases: ≈2 weeks
- Severe cases: ≈6 weeks

Inter-individual variability is substantial and not well characterized. The estimates are parameter fits for population median in China and do not describe this variability (Li et al. 2020; He et al. 2020).
Kinetics of infection and clinical deterioration suggest that pathology is from immune over-stimulation.

- Which immune parameters correlate with disease severity?
  - Serum cytokines (IL-6, IL-1β, others)
  - Antibody response
  - T cell response

- Why do some patients develop severe disease and some not?
  - Germline genetics
  - Airway microbiota
  - Pre-existing cross-reactive immunity

Clinical Symptoms:
- Mild constitutional symptoms
- Fever >99.6°F
- Dry cough, diarrhea, headache

Clinical Signs:
- Lymphopenia, increased prothrombin time, increased D-Dimer and LDH (mild)
- Shortness of Breath
- Hypoxia (PaO2/FiO2≤300mmHg)
- ARDS
- SIRS/Shock
- Cardiac failure
- Elevated Inflammatory markers (CRP, LDH, IL-6, D-dimer, ferritin)
- Troponin, NT-proBNP elevation

Potential Therapies:
- Remdesivir, chloroquine, hydroxychloroquine, convalescent plasma transfusions
- Reduce immunosuppression
- Corticosteroids, human immunoglobulin, IL-6 inhibitors, IL-2 inhibitors, JAK inhibitors
Research effort on COVID-19/SARS-COV-2

• Work at the APS and SBC to solve structures and screening small molecules (nsp15), 3CLpro, PLpro, etc.

• Participation in the NVBTL working groups (EPI, Testing, Manufacturing, Therapeutics)

• Computational work on five subproblems
  • Antiviral drug screening ⇒ priority compounds for wet lab screening
  • Epidemiology ⇒ transmission and interventions
  • Evolution ⇒ origins, diversity and host-adaptation
  • Vaccine ⇒ epitope analysis and antibody design
  • Host-pathogen interactions / host response ⇒ severity and drugs
Antiviral Drug Screening
<table>
<thead>
<tr>
<th>Protein</th>
<th>Mol. weight (kDa)</th>
<th>Seq. similarity with SARS-CoV</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nsp1</td>
<td>19.8</td>
<td>91.1%</td>
<td>Suppresses host antiviral response</td>
</tr>
<tr>
<td>Nsp2</td>
<td>70.5</td>
<td>82.9%</td>
<td></td>
</tr>
<tr>
<td>Nsp3</td>
<td>217.3</td>
<td>86.5%</td>
<td>Nsp3-Nsp4-Nsp6 complex involved in viral replication</td>
</tr>
<tr>
<td>Nsp4</td>
<td>56.2</td>
<td>90.8%</td>
<td>Nsp3-Nsp4-Nsp6 complex involved in viral replication</td>
</tr>
<tr>
<td>Nsp5</td>
<td>33.8</td>
<td>98.7%</td>
<td>Main protease (3C-like)</td>
</tr>
<tr>
<td>Nsp6</td>
<td>33.0</td>
<td>94.8%</td>
<td>Nsp3-Nsp4-Nsp6 complex involved in viral replication</td>
</tr>
<tr>
<td>Nsp7</td>
<td>9.2</td>
<td>100.0%</td>
<td>Nsp7-Nsp8 complex is part of RNA polymerase</td>
</tr>
<tr>
<td>Nsp8</td>
<td>21.9</td>
<td>99.0%</td>
<td>Nsp7-Nsp8 complex is part of RNA polymerase</td>
</tr>
<tr>
<td>Nsp9</td>
<td>12.4</td>
<td>98.2%</td>
<td>ssRNA binding</td>
</tr>
<tr>
<td>Nsp10</td>
<td>14.8</td>
<td>99.3%</td>
<td>Essential for Nsp16 methyltransferase activity</td>
</tr>
<tr>
<td>Nsp11</td>
<td>1.3</td>
<td>92.3%</td>
<td>Short peptide</td>
</tr>
<tr>
<td>Nsp12</td>
<td>106.7</td>
<td>98.3%</td>
<td>RNA polymerase</td>
</tr>
<tr>
<td>Nsp13</td>
<td>66.9</td>
<td>100.0%</td>
<td>Helicase/triphosphatase</td>
</tr>
<tr>
<td>Nsp14</td>
<td>59.8</td>
<td>98.7%</td>
<td>3′-5′ exonuclease</td>
</tr>
<tr>
<td>Nsp15</td>
<td>38.8</td>
<td>95.7%</td>
<td>Uracil-specific endoribonuclease</td>
</tr>
<tr>
<td>Nsp16</td>
<td>33.3</td>
<td>98.0%</td>
<td>RNA-cap methyltransferase</td>
</tr>
<tr>
<td>S</td>
<td>141.2</td>
<td>87.0%</td>
<td>Spike protein, mediates binding to ACE2</td>
</tr>
<tr>
<td>Orf3a</td>
<td>31.1</td>
<td>85.1%</td>
<td>Activates the NLRP3 inflammasome</td>
</tr>
<tr>
<td>Orf3b</td>
<td>6.5</td>
<td>9.5%</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>8.4</td>
<td>96.1%</td>
<td>Envelope protein, involved in virus morphogenesis and assembly</td>
</tr>
<tr>
<td>M</td>
<td>25.1</td>
<td>96.4%</td>
<td>Membrane glycoprotein, predominant component of the envelope</td>
</tr>
<tr>
<td>Orf6</td>
<td>7.3</td>
<td>85.7%</td>
<td>Type I IFN antagonist</td>
</tr>
<tr>
<td>Orf7a</td>
<td>13.7</td>
<td>90.2%</td>
<td>Virus-induced apoptosis</td>
</tr>
<tr>
<td>Orf7b</td>
<td>5.2</td>
<td>84.1%</td>
<td></td>
</tr>
<tr>
<td>Orf8</td>
<td>13.8</td>
<td>45.3%</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>45.6</td>
<td>94.3%</td>
<td>Nucleocapsid phosphoprotein, binds to RNA genome</td>
</tr>
<tr>
<td>Orf9b</td>
<td>10.8</td>
<td>84.7%</td>
<td>Type I IFN antagonist</td>
</tr>
<tr>
<td>Orf9c</td>
<td>8.0</td>
<td>78.1%</td>
<td></td>
</tr>
<tr>
<td>Orf10</td>
<td>4.4</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>
Structures exist for most of the proteome

Figure 3. Structurally characterized structural proteins and an ORF of 2019-nCoV. Highlighted in pink are mutations found when aligning the proteins against their homologs from the closest related coronaviruses: 2019-nCoV and human SARS, bat coronavirus, and another bat betacoronavirus BtRF-BetaCoV. Highlighted in yellow are novel protein inserts found in wS.

Figure 2. Structurally characterized non-structural proteins of 2019-nCoV. Highlighted in pink are mutations found when aligning the proteins against their homologs from the closest related coronaviruses: 2019-nCoV and human SARS, bat coronavirus, and another bat betacoronavirus BtRF-BetaCoV. The structurally resolved part of wNsp7 is sequentially identical to its homolog.

*Viruses* 2020, 12(4), 360; [https://doi.org/10.3390/v12040360](https://doi.org/10.3390/v12040360)
HPC/AI is helping discover novel small molecules that can inhibit various virus proteins

- ML/AI approaches are enabling the identification of potential leads that can bind to 8 viral protein targets
- DeepDriveMD helps identify conformational states that bind to specific ligands
- Identified over 30 lead molecules that have been submitted to various open forums for experimental validation
- Collaborations with University of Chicago, Brookhaven National Lab, Frederick National Laboratory and the University of Michigan
HPC/AI is helping discover novel small molecules that can inhibit various virus proteins.

Consensus hits

ADP-ribose-1'-monophosphatase (ADRP) (Nsp3)

Molecular States Classified by ML
Sampling of Top Hits from ML (Enamine_REAL 1.2B) for ADRP-P1

2045 compounds with softmax = 1.0
Predicted docking score < -8.50
Left is BSL2 → BSL3

Vivarium is BSL3
Our HPC- and AI-enabled small molecule filtering pipeline

Chemical Databases
- Enamine
- DrugBank
- BindingDB
- eMolecules
- cureFFI
- MOSES
- ZINC15
- LINCS
- SureChEMBL
- PubChem
- And more...

Canonicalization ➔ Compute Features ➔ ML based filtering

Fingerprinting ➔ Similarity Search

Generate Images ➔ DNN filtering

Computing Resources
Frontera

https://2019-ncovgroup.github.io
First release: 21 sources, 3.9B molecules, 80 TB computed features

https://2019-ncovgroup.github.io
ENAMINE REAL  1.2 billion molecules which comply with “rule of 5“ and Veber criteria: MW≤500, SlogP≤5, HBA≤10, HBD≤5, rotatable bonds≤10, TPSA≤140.

GDB-13 enumerates small organic molecules up to 13 atoms of C, N, O, S and Cl following simple chemical stability and synthetic feasibility rules.

21 sources, 3.9B molecules, 80 TB computed features

https://2019-ncovgroup.github.io
Mining literature for drug discovery and repurposing

- Thousands of papers already published about COVID-19 and similar coronaviruses
- Developing human and machine pipelines to identify, extract drugs (current) and mechanisms (future)
- Identify key molecules for simulation team as starting points
- Build a list of known antiviral molecules and molecules active against SARS/MERS/HKU/SARS-CoV-2
- Use this list and “most-similar” molecules to build confidence in ML and simulation predictions
Mining literature for drug discovery and repurposing

1. Manual Extraction
   - Engaging CELS admin staff
   - Currently have extracted 803 screened molecules and structures from 61 articles and reviews. (top figures)
   - Capacity to extract from ~100 articles

2. Deep Learning (NLP)
   - Team has labeled ~1500 abstracts with drugs in their natural language context in CORD-19 papers (bottom figure)
   - Building named-entity models to enable automated extraction of drugs from entire corpus (~40k articles)
   - Current F1: 82.7 – more validation needed

Resulting data and models will be published openly
Computing at Argonne, Oak Ridge, TACC, SDSC, IU, LRZ, Brookhaven
The COVID-19 High Performance Computing Consortium

Bringing together the Federal government, industry, and academic leaders to provide access to the world’s most powerful high-performance computing resources in support of COVID-19 research.

33
Consortium members

41k
GPUs

Active projects

Fighting COVID-19 will require extensive research in areas like bioinformatics, epidemiology, and molecular modeling to understand the threat we’re facing and to develop strategies to address it.

Here are some of our active projects.

See all

Request computing resource for de novo protein therapeutics design simulations to treat the COVID-19...

Discovering molecular targets of the human coronavirus with HPC and AI

Harnessing Large-Scale Quantum-Based DFTB Calculations for a More Accurate Assessment of COVID-19...
Epidemiology
WE ARE MODELING COVID-19 SPREAD AMONG PEOPLE IN CHICAGO

Joint DOE Laboratory Plan for Pandemic Modeling and Analysis Capability

- Argonne, Oak Ridge, Los Alamos and Sandia will collaborate over a 3 month-period to develop an integrated COVID-19 pandemic monitoring, modeling, and analysis capability that will address the key questions about the pandemic.
Modeling PEOPLE WITH ARGONNE’S CityCOVID

- CityCOVID is a city-scale agent-based model
- 2.7M+ individual agents (people)
  - move to/from 1.2M spatially-located places
  - on an hourly basis
  - over a period of a year (8760 hours)
- Each agent has contact with other agents at each place (possible disease transmission)
  - agent has individual behaviors, engages in activities, and responds and adapts:
    - to the disease
    - to public health messaging
    - to public health interventions
- Up to $10^{12}$ (trillion) individual contacts during a yearly simulation
MODELING INDIVIDUAL AGENT DISEASE STATES WITH CITYCOVID

susceptible.to.exposed.probability
seasonality.multiplier
seasonality.peak
age dependent
shielding.scaling

Susceptible → Exposed

Beta(incubation.duration.alpha, incubation.duration.beta) → 2-6 days

Exposed → Severe Infected

Beta(severe.infection.duration.alpha, severe.infection.duration.beta) → 14 - 35 days

Severe Infected → Hospitalization

Beta(infection.duration.alpha, infection.duration.beta) → 7-17 days

Infected → Symptomatic (at home) → Asymptomatic (at home)

Asymptomatic (business as usual)

Recovered

0-14 days per day probability remain asymptomatic
(start.home.isolation.prob)

0-3 days per day probability remain symptomatic
(end.home.isolation.prob)

Deceased

13% age dependent

Jonathan Ozik jozik@anl.gov
Chick Macal macal@anl.gov
Argonne National Laboratory
April 20, 2020
CityCOVID PARAMETER ESTIMATION ON THETA

- CityCOVID is implemented as an MPI application using the Repast HPC ABM toolkit and the Chicago Social Interaction Model (ChiSIM) framework
- Each model is distributed across 256 ranks for efficient execution (each simulated year, at an hourly time step, takes approximately 8-12 minutes to complete for a full city-scale run)
- We are using our large-scale model exploration framework (EMEWS) to implement sequential approximate Bayesian computation (ABC) parameter estimation/calibration workflows, coordinating large ensemble runs (30k+ models) on Theta
CityCOVID MODEL OUTPUTS

- CityCOVID generates projections of epidemiological variables, including COVID-19 exposures and deaths

- CityCOVID enables running policy scenarios, such as those examining the consequences of easing current in-place restrictions
Evolution
- >10,000 viral sequences
- Phylogenetic trees updated daily
- DOE $\Rightarrow$ FEMA, BARDA etc.
- Place, Date
- Trees from WGS, SNPs
- Tracking new mutations
- Capturing significant variants
Supercomputing Focus Areas

- Accelerating development of treatment options
- Learning how epidemics impact critical social services
- Improving understanding of human virus interactions
Questions