North Carolina School of Science and Mathematics

Computing CoVID-19

Course Review and Recap
Science Department
Summer Research and Innovation Program

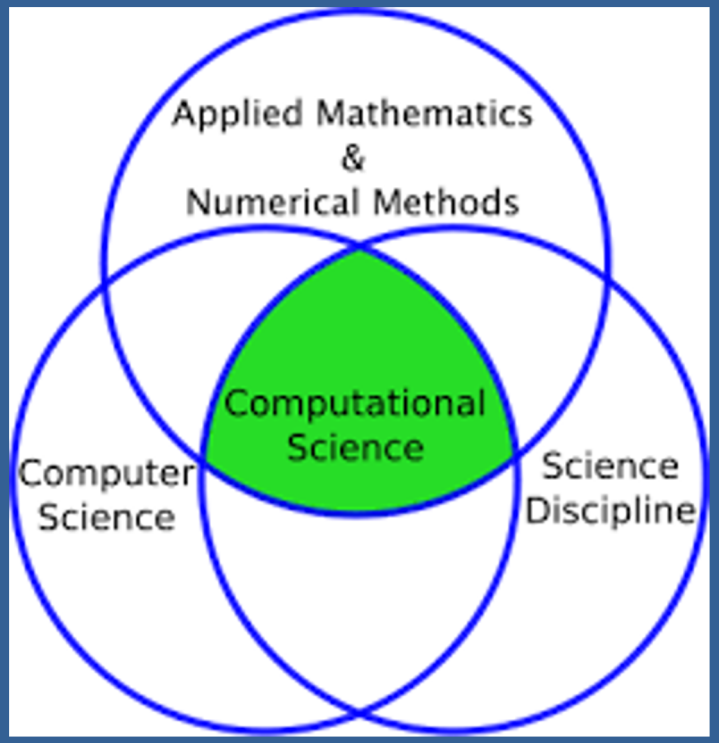
Robert Gotwals, Computational Science Educator
gotwals@ncssm.edu
Em Ambrosius, Teaching Assistant (TA)
ambrosius20e@ncssm.edu
Elise Ray, Teaching Assistant (TA)
ray20e@ncssm.edu
Marielle Rath, Teaching Assistant (TA)
rath20m@ncssm.edu
David Major, Teaching Assistant (TA)
major21d@ncssm.edu

North Carolina School of Science and Mathematics

About Computational Science

An interdisciplinary topic that incorporates and integrates

- All “science” disciplines
  - Physical sciences
  - Life sciences
  - Social sciences
- Computer Science
- Mathematics



North Carolina School of Science and Mathematics

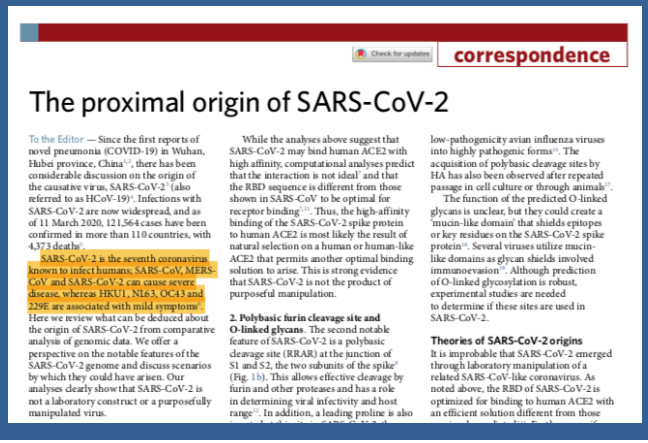
NCSSM Computational Science Program

1. Introduction to Computational Science
2. **Computational Biology / Bioinformatics**
3. Computational Chemistry
4. Computational Physics
5. Nanotechnology and Research
6. Scientific Programming
7. Data Science for Scientists
8. **Computational Medicinal Chemistry**
9. Digital Humanities (residential)
10. Research Experience in Computational Science (spring/summer)
11. Research in Computational Science (residential)

Day 1: Introduction

- Morning
  - Overview of the SARS-CoV-2 virus
  - Journal Article: “The proximal origin of SARS-CoV-2 (Nature Medicine)
- Afternoon
  - Dr. Holden Thorp, Editor-in-Chief, Science magazine
  - Software installs

North Carolina School of Science and Mathematics



correspondence

The proximal origin of SARS-CoV-2

To the Editor — Since the first reports of novel pneumonia (COVID-19) in Wuhan, Hubei province, China[1,2], there has been considerable discussion on the origin of the causative virus, SARS-CoV-2[3] (also referred to as HCoV-19)[4]. Infections with SARS-CoV-2 are now widespread, and as of 11 March 2020, 121,564 cases have been confirmed in more than 110 countries, with 4,373 deaths[5].

SARS-CoV-2 is the seventh coronavirus known to infect humans; SARS-CoV, MERS-CoV and SARS-CoV-2 can cause severe disease, whereas HKU1, NL63, OC43 and 229E are associated with mild symptoms[6]. Here we review what can be deduced about the origin of SARS-CoV-2 from comparative analysis of genomic data. We offer a perspective on the notable features of the SARS-CoV-2 genome and discuss scenarios by which they could have arisen. Our analyses clearly show that SARS-CoV-2 is not a laboratory construct or a purposefully manipulated virus.

While the analyses above suggest that SARS-CoV-2 may bind human ACE2 with high affinity, computational analyses predict that the interaction is not ideal[7] and that the RBD sequence is different from those shown in SARS-CoV to be optimal for receptor binding[7–11]. Thus, the high-affinity binding of the SARS-CoV-2 spike protein to human ACE2 is most likely the result of natural selection on a human or human-like ACE2 that permits another optimal binding solution to arise. This is strong evidence that SARS-CoV-2 is not the product of purposeful manipulation.

2. Polybasic furin cleavage site and O-linked glycans. The second notable feature of SARS-CoV-2 is a polybasic cleavage site (RRAR) at the junction of S1 and S2, the two subunits of the spike[1] (Fig. 1b). This allows effective cleavage by furin and other proteases and has a role in determining viral infectivity and host range[12]. In addition, a leading proline is also low-pathogenicity avian influenza viruses into highly pathogenic forms[16]. The acquisition of polybasic cleavage sites by HA has also been observed after repeated passage in cell culture or through animals[17].

The function of the predicted O-linked glycans is unclear, but they could create a ‘mucin-like domain’ that shields epitopes or key residues on the SARS-CoV-2 spike protein[18]. Several viruses utilize mucin-like domains as glycan shields involved immunoevasion[19]. Although prediction of O-linked glycosylation is robust, experimental studies are needed to determine if these sites are used in SARS-CoV-2.

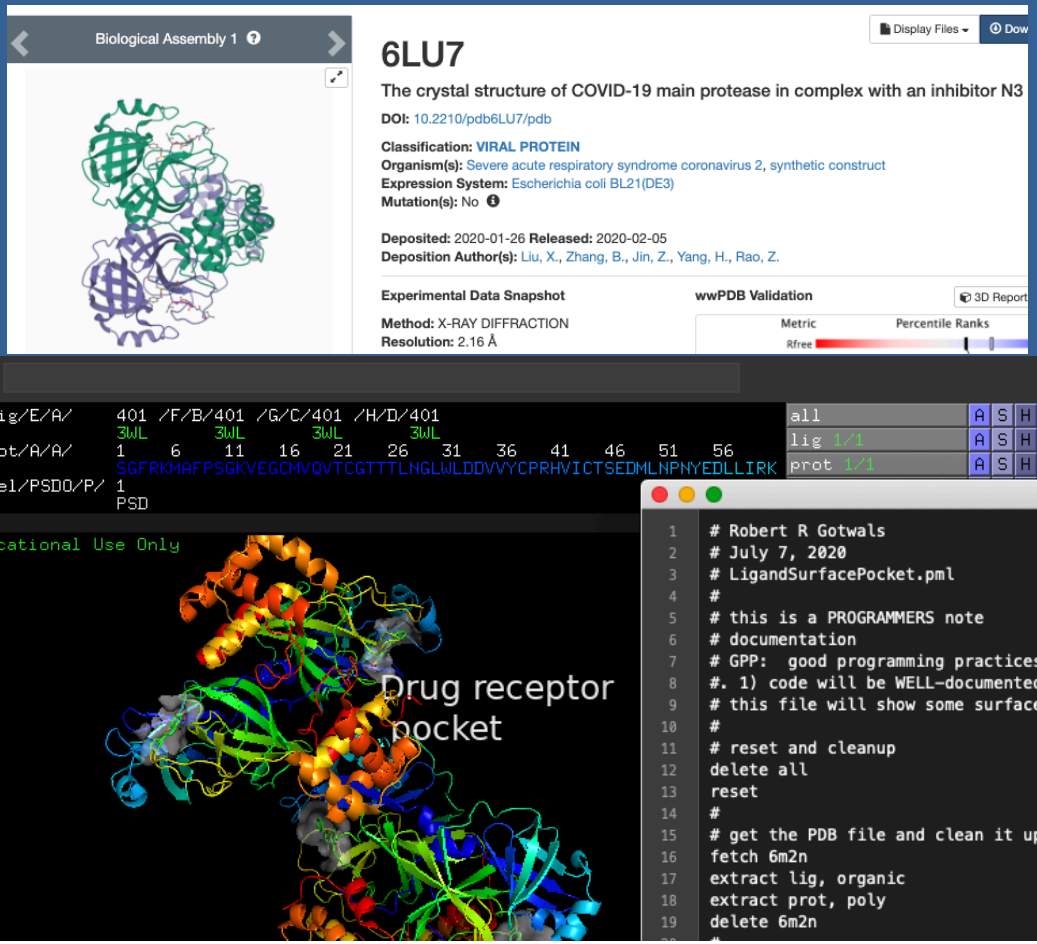
Theories of SARS-CoV-2 origins

It is improbable that SARS-CoV-2 emerged through laboratory manipulation of a related SARS-CoV-like coronavirus. As noted above, the RBD of SARS-CoV-2 is optimized for binding to human ACE2 with an efficient solution different from those

Day 2: Protein Structure Basics

North Carolina School of Science and Mathematics

- Morning: Overview of protein structure
  - Protein Data Bank
- Afternoon: Visualizing Protein Structure
  - PyMOL
  - PyMOL scripting



Day 3: Genetic Alignments and Phylogenetic Trees

North Carolina School of Science and Mathematics

- Morning: Genetic Alignments
  - UniProt
  - BLAST
- Afternoon: Phylogenetic Trees
  - UPGMA
  - Clades
  - Genetic Distances

☐	Q9BYF1	ACE2_HUMAN		**Angiotensin-converting enzyme 2**	**ACE2** UNQ868/PRO1885	Homo sapiens (Human)
☐	Q8R0I0	ACE2_MOUSE		**Angiotensin-converting enzyme 2**	**Ace2**	Mus musculus (Mouse)
☐	Q5EGZ1	ACE2_RAT		**Angiotensin-converting enzyme 2**	**Ace2**	Rattus norvegicus (Rat)

Number of Amino Acid Difference in ACE2 Among Five Organisms

	Human	Mouse	Rat	Cat	Cow
Human		144	141	119	153
Mouse			77	147	158
Rat				148	159
Cat					135
Cow					

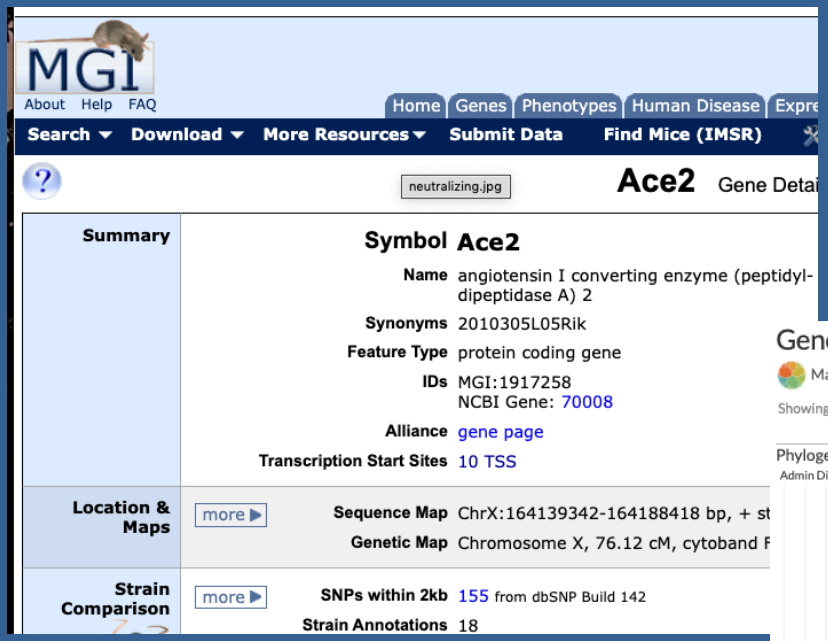
	Human	MR	Cat	Cow
Human		144	119	153
MR			147.5	158.5
Cat				135
Cow				

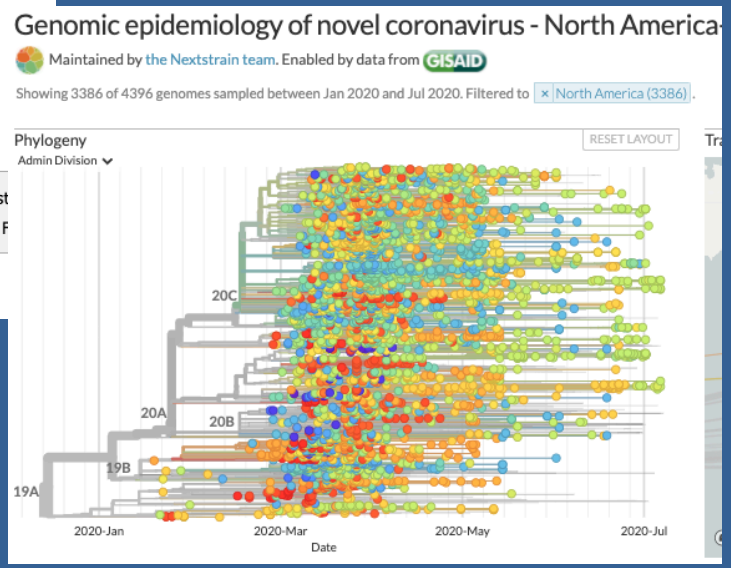
Clade: MR	77
Distance	38.5
Clade:Hcat	144
Distance	72

Day 4: Mouse Strains and Genomic Epidemiology

North Carolina School of Science and Mathematics

- Morning: Mice Studies
  - MGI
- Afternoon: Genomic Epidemiology
  - NextStrain

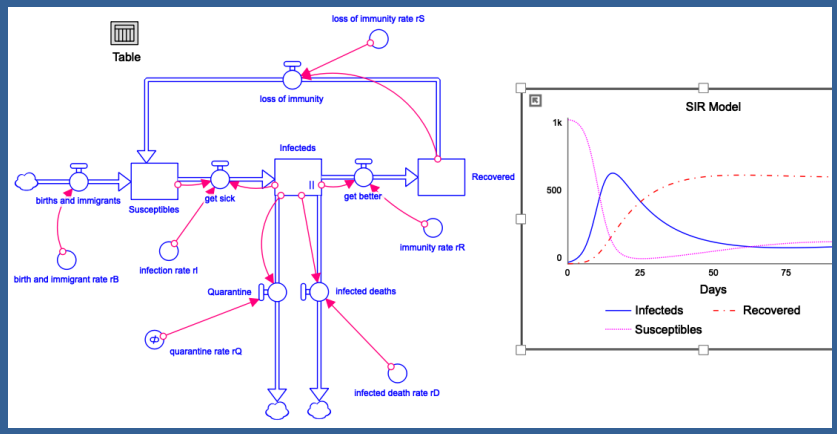


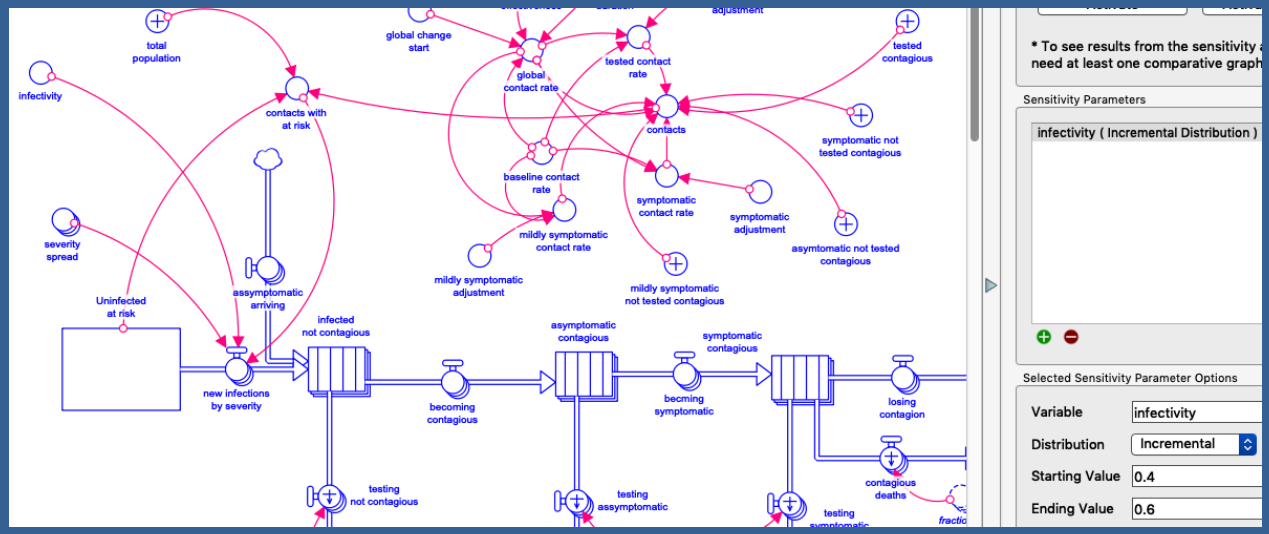


Day 5: Epidemiology

North Carolina School of Science and Mathematics

- Morning: Building an Epidemiology Model
  - MGI: Mouse Genome Informatics
- Afternoon: Modifying an epidemiology model
  - Sensitivity analysis
  - Infectivity samples





Day 6: Pharmacokinetics

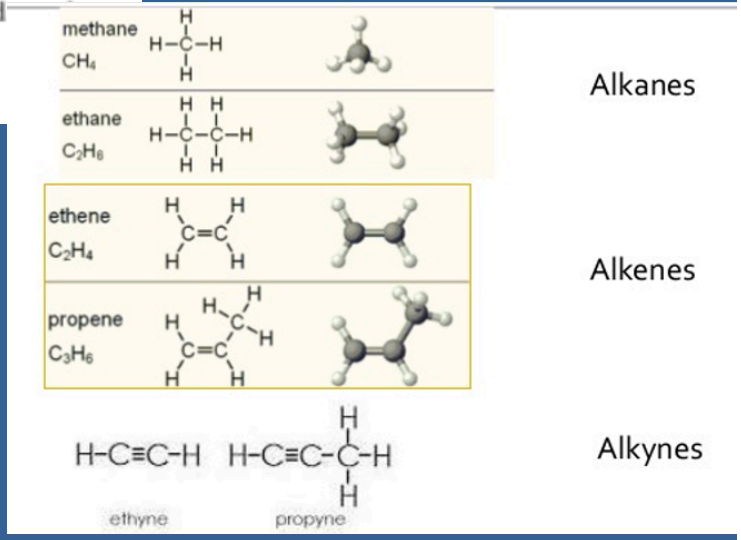
North Carolina School of Science and Mathematics

- Morning: Intro to Pharmacokinetics
  - Journal Article: MedChem of CNS Drugs
- Afternoon: Organic chemistry crash course model
  - Basics of structure and naming
  - Functional groups

drugs. From this analysis, the "Rule of Five" was developed. The "Rule of Five" is so named because all the essential physical properties are parameters of five. According to this rule, a good absorption and permeability is likely if:

- Molecular weight is ≤500
- Oil/water distribution coefficient (LogP) is ≤5
- Hydrogen bond donors ≤5 (expressed as the sum of OHs and NHs)
- Hydrogen bond acceptor ≤10 (expressed as the sum of Ns and Os)

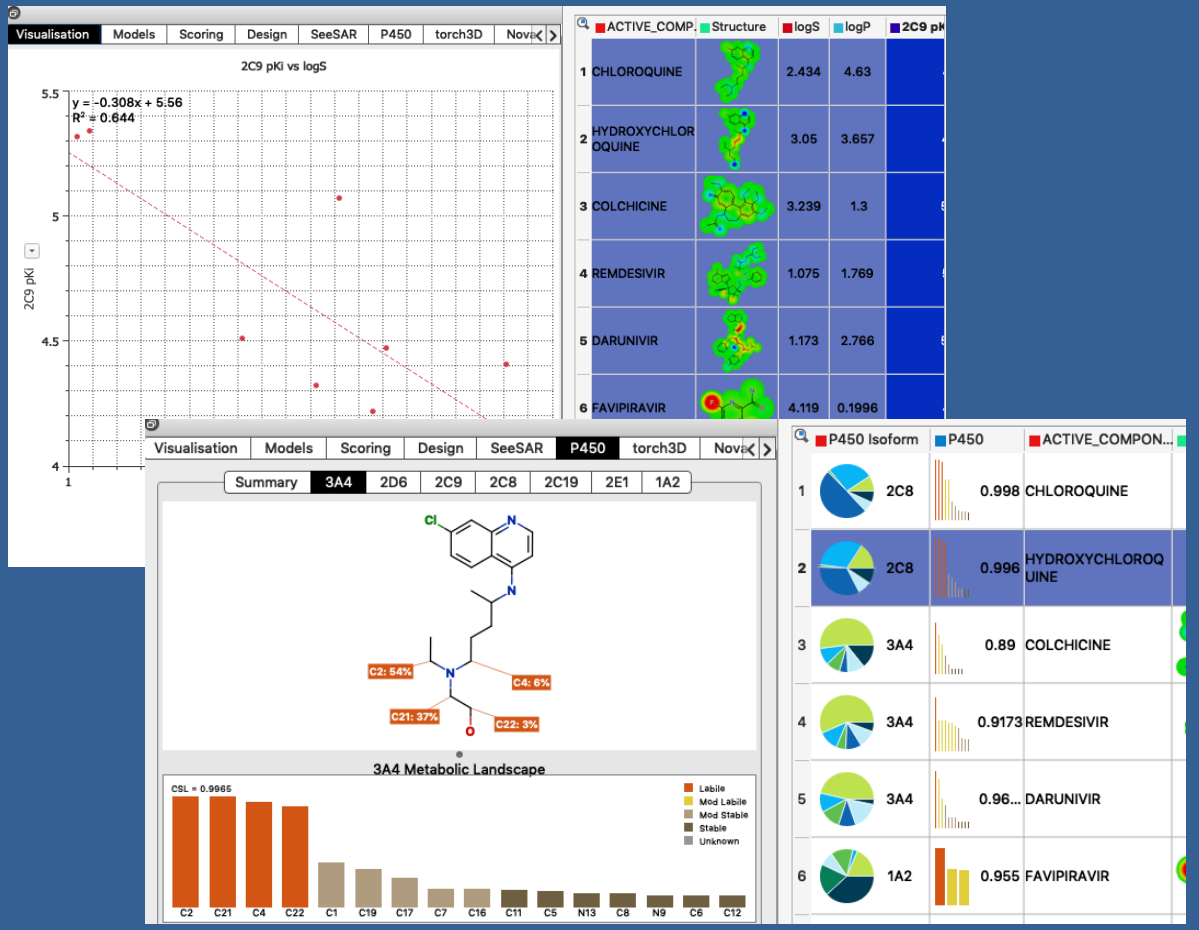
A fifth rule was added later:



Day 7: StarDrop Studies

North Carolina School of Science and Mathematics

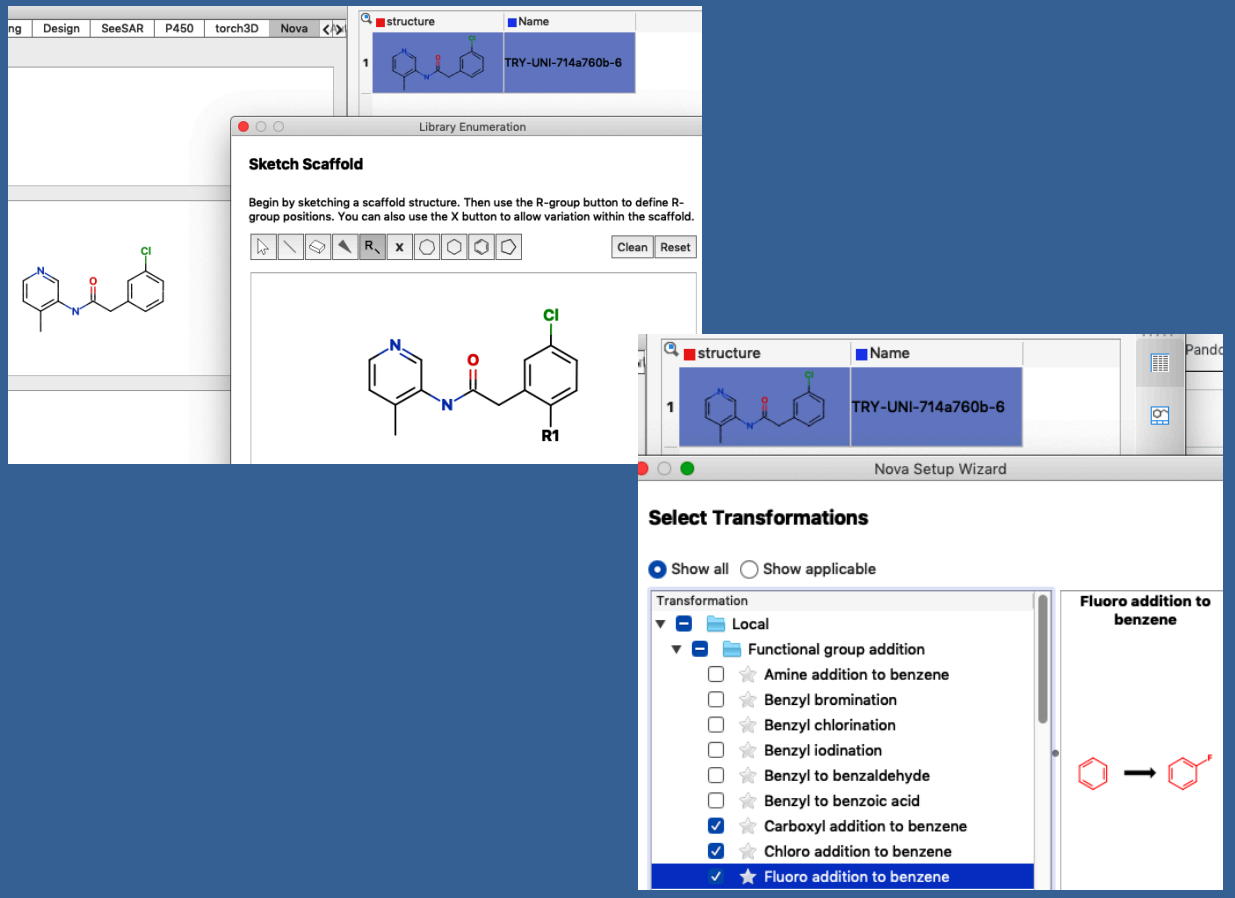
- Morning: ADME studies
  - logP, logS, 2Cp pKi
- Afternoon: Metabolism Studies
  - Cytochrome P450 enzymes
  - Lability vs. Stability



Day 8: Lead Drug Modification

North Carolina School of Science and Mathematics

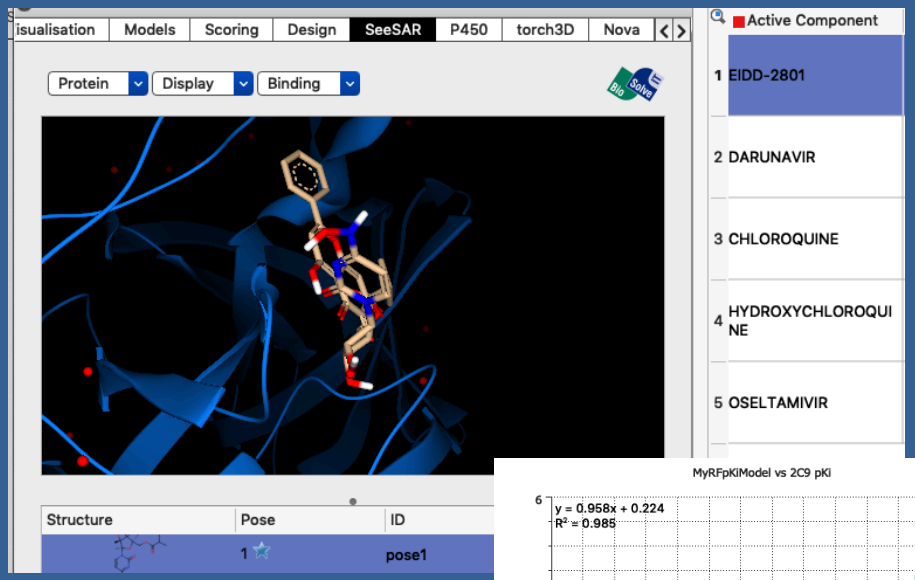
- Morning: R-Group Enumeration
  - Methyl, ethyl, propyl, butyl
- Afternoon: Chemistry Transformations
  - Generation of 575 novel compounds
  - Screening – 7 compounds

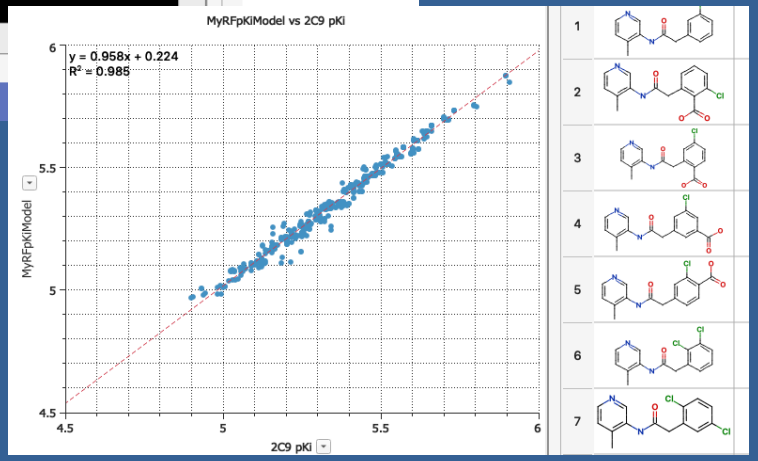


Day 9: Docking and QSAR

North Carolina School of Science and Mathematics

- Morning: Protein-Ligand Docking
  - SeeSAR Tool
  - Five antivirals against PDB ID:6M2N
- Afternoon: QSAR using machine learning
  - AutoModeler tool
  - Goal: generate a random forest model

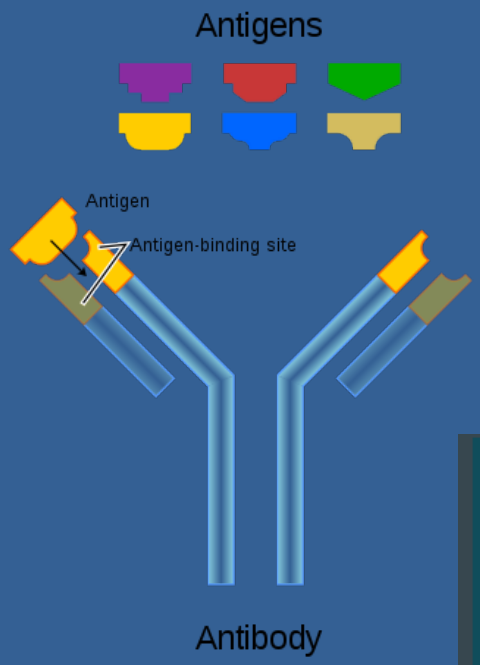


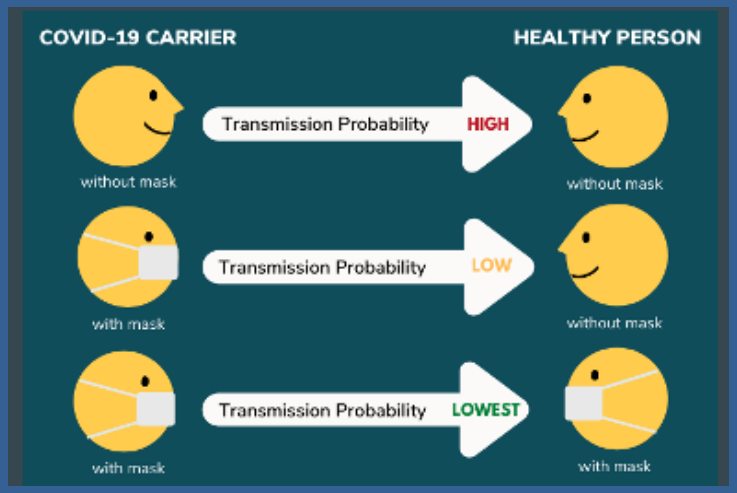


Day 10: Vaccines; Mask Discussion; Wrap-Up

North Carolina School of Science and Mathematics

- Morning: Vaccines and Masks
  - Overview of vaccines
  - Overview of the science of masks
  - Discussion groups
- Afternoon: Wrap-up
  - This talk
  - Course survey





Benefits

North Carolina School of Science and Mathematics

This student participated in a special summer "short-course" entitled "Computing CoVID-19", presented jointly by the Science Department and the Summer Research and Innovation Program (SRIP)at the North Carolina School of Science and Mathematics (NCSSM). This course was designed to use the technologies, techniques, and tools of computational science to study a variety of concept and topics revolving around the SARS-CoV-2 virus and the related pandemic disease of CoVID-19. During the first week of this course, students were introduced to protein structure and the use of protein computing tools and databases, including the Protein Data Bank (PDB), UniProt, and PyMOL. Students explored protein structures, performed BLAST alignments, and created phylogenetic trees from scratch. They also used the NextStrain phylogenetic tree resource to explore the SARS-CoV-2 evolutionary history. Also during Week 1, they built a simple epidemiology model using STELLA, and then modified a complext CoVID-19 model to evaluate infectivity and its impact on total number of infected persons. In Week 2, they were introduced to basic organic chemistry in medicinal chemistry and used a research-grade medicinal chemistry computing tool -- StarDrop -- to study a wide variety of chemical properties of anti-viral drugs and drugs that are being reconsidered for repurposing to address the CoVID-19 epidemic. This course, a nationally-unique offering, was highly rigorous. The student's completion of this course indicates above-average abilities, intellectual curiosity, and work ethic, all characteristics that should be of interest for any university!

North Carolina School of Science and Mathematics

This WAS an extraordinary opportunity – we hope you TOOK every advantage!

North Carolina School of Science and Mathematics

QUESTIONS?
COMMENTS?