

GENERAL SCHEDULE

December - project initial ideas and development

January to February - work on project, MAKE SURE TO WRITE IN LOGBOOK

- Do more research into background information

Early February to Mid February - finish project

Mid February to late February - register project

March - make presentation

*meetings every week with Dr. Edwards Tuesdays 4:30

December 23rd:

- Started working with Dr. Edwards on his Covid-19 project
- Initial ideas:
 - Labs are closed, need to find online alternatives
 - Pandemic is the most pressing current issue
 - Have already been working with Dr. Edwards exploring various biochemistry topics

Background (rough notes):

- Mutations could make sick people contagious for longer periods of time
- could help viruses survive for longer outside of the body or
- could increase their ability to replicate
- individual tweaks to the virus's genome might function differently on their own than they do as a patchwork of mutations
- severity; or a change in the way the virus interacts with the immune system
- A virus replicates by hijacking its host's cellular machinery to make copies of itself. But the genetic copies accumulate small errors, or mutations.
- Changes that are beneficial to the virus can also drive its spread, leading to a variant that outcompetes other local varieties and may send cases surging.
- For SARS-CoV-2, these mutations—the small errors made naturally when genomes are copied—develop at a steady pace of one or two each month, says Loman, a professor of microbial genomics and bioinformatics at the University of Birmingham.
- The rise of the variants emphasizes the need for greater precautions against the virus
- viral entry into host cells through the spike (S) proteins on the virus' surface

<https://www.nationalgeographic.com/science/2021/01/why-some-coronavirus-variants-are-more-contagious/#close>

- In most cases, the fate of a newly arising mutation is determined by natural selection
- *Mutation* refers to the actual change in sequence: D614G is an aspartic acid-to-glycine substitution at position 614 of the spike glycoprotein. Genomes that differ in sequence are often called *variants*. This term is somewhat less precise because 2 variants can differ by 1 mutation or many. Strictly speaking, a variant is a *strain* when it has a demonstrably different phenotype (eg, a difference in antigenicity, transmissibility, or virulence).
- Concerning outbreaks of SARS-CoV-2 began to emerge on mink farms in the Netherlands and Denmark in late spring and early summer 2020.⁶ Genomic and epidemiologic investigation of an early outbreak in the Netherlands demonstrated human to mink, mink to mink, and mink to human transmission.⁷ Many SARS-CoV-2 sequences from the Netherlands and Danish outbreaks had a Y453F mutation in the receptor binding domain of spike, which might mediate increased binding affinity for mink ACE2 (angiotensin-converting enzyme 2).
- Genomic surveillance of SARS-CoV-2 variants has largely focused on mutations in the spike glycoprotein, which mediates attachment to cells and is a major target of neutralizing antibodies
- There is intense interest in whether mutations in the spike glycoprotein mediate escape from host antibodies and could potentially compromise vaccine effectiveness, since spike is the major viral antigen in the current vaccines.
- Sometimes a mutation that enhances one viral property, such as binding to a receptor, can reduce another property, such as escaping host antibodies.

<https://jamanetwork.com/journals/jama/fullarticle/2775006>

The pandemic of coronavirus disease 2019 (COVID-19) caused by the newly emerged severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has become a global health crisis. SARS-CoV-2 belongs to the *Betacoronavirus* genus of the *Coronaviridae* family. protein of SARS-CoV-2 mediates viral entry into host cells by binding to their shared receptor, angiotensin-converting enzyme 2 (ACE2) through the receptor-binding domain (RBD), which is how it enters the lungs. SARS-CoV-2 primarily replicated in the respiratory tracts. Originated in Wuhan, Hubei Province, Central China. 4 structural proteins. The four structural proteins consist of the spike (S) surface glycoprotein, the membrane (M) protein, the envelope (E) protein and the nucleocapsid (N) protein. The receptor binding domain (RBD) of spike from both these viruses binds ACE2 with high affinity . Because of its role in viral entry, the RBD is a major determinant of cross-species transmission and evolution

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7139247/>

In addition, the RBD is the target of the most potent anti-SARS-CoV-2-neutralizing antibodies identified to date. mutation effects on RBD binding and expression are correlated (Figures 5C

and S5C), with residues that deviate from this trend clustering at the ACE2 interface (Figure 5C, cyan points). This correlation between expression and binding is consistent with studies on antibodies, where mutations that improve stability and rigidity accompany increases in binding affinity

<https://www.sciencedirect.com/science/article/pii/S0092867420310035#fig5>

All viruses mutate. How fast depends on several factors. Viruses with an RNA genome tend to mutate faster than viruses with a DNA genome. This is because RNA viruses have less ability to fix errors when their genetic material is copied to make virus particles inside infected cells. So every time a virus replicates, there is the chance of a mutation occurring.

bioserendipity.com/defining-a-new-strain-of-a-virus/

Mutation Y453F also occur in the RBD region and increase binding affinity to ACE2, and have been shown to escape the neutralising effect of a few monoclonal antibodies (mAbs).

One way the body's immune system attacks foreign substances is by making large numbers of antibodies. An antibody is a protein that sticks to a specific protein called an *antigen*. Antibodies circulate throughout the body until they find and attach to the antigen. Once attached, they can force other parts of the immune system to destroy the cells containing the antigen. Researchers can design antibodies that specifically target a certain antigen, such as one found on cancer cells. They can then make many copies of that antibody in the lab. These are known as *monoclonal antibodies* (mAbs or Moabs). Monoclonal antibodies are man-made proteins that act like human antibodies in the immune system. Mab - A neutralizing antibody is an antibody that defends a cell from a pathogen or infectious particle by neutralizing any effect it has biologically.

Y453F has received widespread attention because it has been observed in the context of mink-human infections and it appeared in the widely reported Danish Mink cluster but has not been observed in the UK to date.

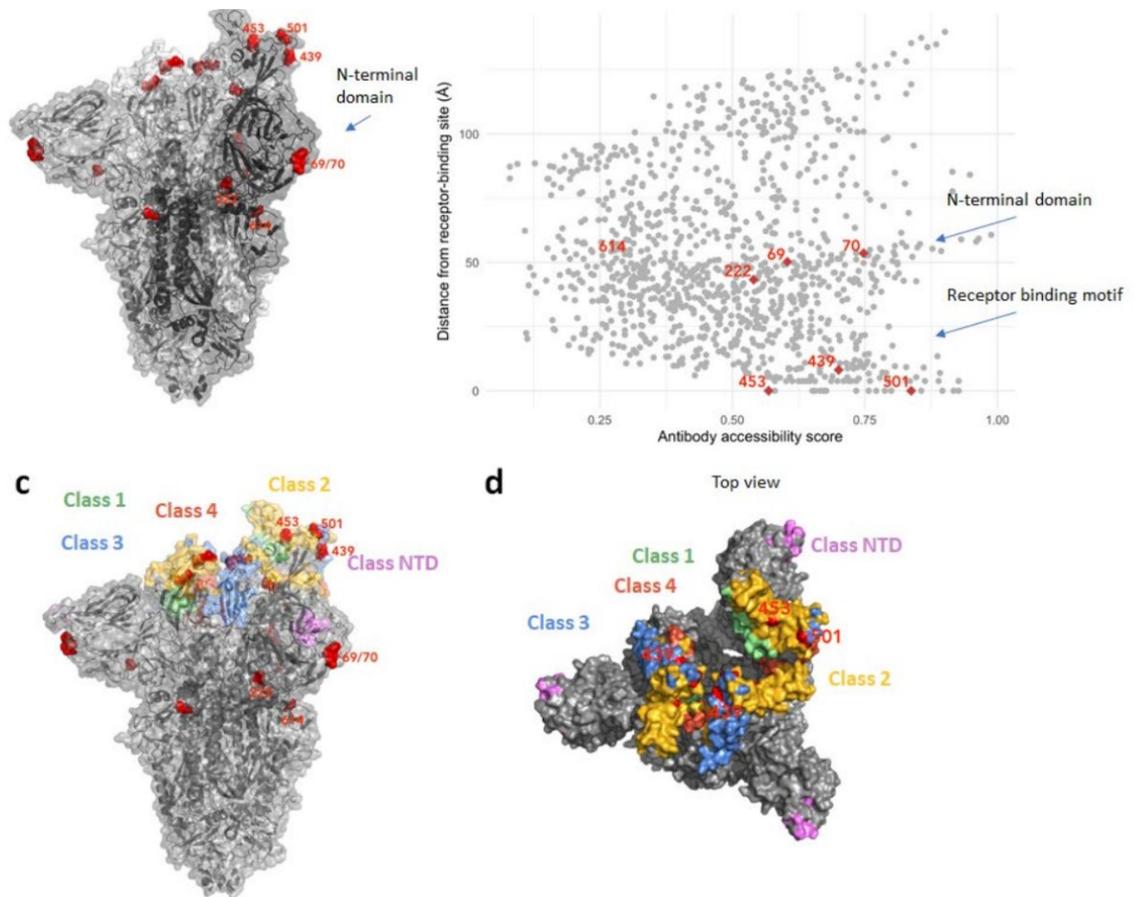
Y453F was identified in the Netherlands and Denmark associated with mink-human infections. Y453F has also been shown in laboratory studies to increase the affinity of Spike protein binding to the ACE2 receptor. This mutation has arisen independently multiple times in several countries. Mutations arise naturally in the SARS-CoV-2 genome as the virus replicates and circulates in the human population. As a result of this on-going process, many thousands of mutations have already arisen in the SARS-CoV-2 genome since the virus emerged in late 2019. The vast majority of mutations have no apparent effect on the virus. Only a very small minority are likely to be important and change the virus in any appreciable way. We pay most attention to

mutations in the gene that encodes the Spike protein, which is associated with viral entry into cells and it is relevant in the context of immunity and vaccine efficacy.

Mutation is used to describe a change of a nucleotide in the virus RNA genome, a subset of which results in a change in amino acid (sometimes referred to as a substitution or replacement), or a mutation can refer to a deletion or insertion event in the virus genome. By convention an amino change is written N501Y to denote the wildtype (N, asparagine) and replacement amino acid (Y, tyrosine) at site 501 in the amino acid sequence.

When a mutation first occurs, it is difficult to determine whether it is important, some steps to identify that are

1. Of theoretical concern, right now is being determined by laboratory experiments.
2. As vaccines are rolled out, it will be important to sequence SARS-CoV-2 virus from infected people who have been vaccinated, or have had a second episode of COVID-19. The aim is to detect variants that are evading the immune system elicited by infection or vaccination.



https://www.attogene.com/wp-content/uploads/2020/12/Report-1_COG-UK_19-December-2020_SARS-CoV-2-Mutations.pdf

Here we experimentally measure how all amino-acid mutations to the RBD affect expression of folded protein and its affinity for ACE2. The entry receptor for SARS-CoV-1 and SARS-CoV-2 is the human cell-surface protein angiotensin converting enzyme 2 (ACE2), Because of its key role in viral entry, the RBD is a major determinant of cross-species transmission and evolution the RBD is one of the most variable regions in sequence alignments of sarbecoviruses quantitative deep mutational scanning approach

The virus moves down your respiratory tract. That's the airway that includes your mouth, nose, throat, and lungs. Your lower airways have more ACE2 receptors than the rest of your respiratory tract. So COVID-19 is more likely to go deeper than viruses like the common cold.

December 23rd - January 1st

Protein visualization and analysis:

First Glance in Jmol

- visualizes protein and can find salt bridges and H bonds
- can change different views of protein (ball and stick, cartoon)
- choose view salt bridges (under tools)
- change background option
- right click to view bring up menu
 - style > scheme to select different view (ball and stick, cartoon)
- under the after right clicking to go to the menu, go to console
 - select: a (to select chain a)
 - select Thr500 (example) to select a specific residue
- the floating red molecules are water
- atom = O means it's a part of the backbone, atom = OG1 or any other variation means it's on the R group
- 4 angstroms or less means that there is either an H bond or salt bridge (look for oxygens interacting with nitrogens)
- carbons interacting with carbons are H-phobic interactions
- note: interactions can be a salt bridge AND hydrogen bond
- **NCBI and RCSB PDB**
- find PDB files of proteins

PBDePISA - proteins, interfaces, structures and assemblies

- bond strength is measured in kilocalories/mol

SWISS MODEL

- Select user template to make model
- sequence of the template NYL*Y*RL (position 453)
- Use 6m0j xray
- **qmean less than 2 (standard deviation), we are not more than 2 standard deviations out, decent accuracy, can use swiss model**

GetArea

- Can get non polar accessible surface areas
 - larger probe, tends to decrease area
 - smaller probe, tends to increase
1. complex - all exposed non polar atoms
 2. pull apart - what are the exposed non polar atoms of ACE2 and virus
 1. add together
 2. subtract non polar surface area for the complex (all positive)
 3. subtract wild type (mutant minus wild type), positive is more strong binding

January 1st - January 15th

- Made Google Docs of all the mutants
- When downloading PDB files indicate the TEMPLATE used (ex. swiss/xray)
- the HHHHHH tag on the PDB file purifies the protein, it is not actually a part of the sequence
- What are GOL, TAR in FirstGlance in Jmol
-
- Swiss models of all mutations
 - Alanine is weird, insertion/deletion?
 - sequence identity is 97.4%.
 - Check under template: template upload
 - confirm sequence
 - swiss model places alignment of angle to optimal, maybe more accurate after swiss model, pdb people also use computer models to make the amino acids twist in the same orientation
- So many other templates/pdb found, 6m0j is only one of them, how do you know which one is the best model? Shouldn't they all be the same?
- when receptor is not there, virus relaxes a little bit, better to have one without antibody and one without receptor, but receptor is better because the interface changes less
 - 7jmo, found on microsoft word the areas of interaction are not where it binds to the receptor, may be better model
 - Ask Dr. Edwards
 - hydrogen bond (1000 calories) typical HB, 3 angstroms, = 30 square angstroms
 - if our score is about 30 calories per square angstrom
- Notes: use \$ to lock in equation on Excel

- Use Notepad to make notes and OPEN IN WORD
- Aliphatic vs aromatic nonpolar surface areas?
 - Carbons can be described as aliphatic or aromatic (does not have to be a whole compound)
- Complex - virus energy of equation
- The salt bridges and hydrogen bonds are wonky when templates are made in Swiss Model
- To make the process of GetArea faster, don't wait the full 10 buffer seconds, press enter after you enter a probe radii
- the biggest advantage to the swiss models as a template that it relaxes things (imitates sort of in solution)
- On the graph, tryptophan looks to be a lot higher than the others
- Sulfur is actually not nonpolar, there are different sulfur settings in GetArea
 - Getarea was made in 1998, more accurate websites?
 - Even if sulfur was oxidized it would be more polar
- Can we figure out a relative percent binding from our values?
- NAG is natural
- NaN errors - change probe to 0.5001/0.4999
 - Really no explanation for errors except that the software is glitchy
 - Completely random, no pattern
- Y453F - more NP surface area exposed than the wild type, this factor says the virus is higher in energy, more tendency to bind to receptor by hydrophobic interactions
- in solution protein flexes larger than x ray solution, flexes larger more than smaller
- Not saving aromatic and aliphatic anymore, too complicated for now
- calculating non polar area - virus will also relax, didn't relax in calculations (sort of relaxes in swiss model)
- Probe radii of 1.3, 1.35, 1.4, 1.45, 1.5 gave better results
- maybe charged amino acids have ionic interactions (salt bridges), maybe the ones close to are not forming the salt bridges. putting charged amino acids may affect other salt bridges (85 to 12 forms a salt bridge)
- even though enthalpy is favourable entropy may be disfavour able
- try to eliminate random error, averages

January 15th - March 1st

- our calculations can help people determine what to test for
 - Can take researchers in labs several months for mutants of one position
- our models depend on template (ideally they shouldn't), Swiss Model glitch
 - Can't access supercomputers

- swiss models aren't the best models, one error is steric hindrance, could be swiss model hasn't try all possible torsion angles, focuses on common torsion angles, energy minimization
- Appears to be a aromatic aromatic interaction that was not accounted for at all
- Look up nearest neighbor interactions
- Note: standard equivalent radius - 1.4 angstroms, in hydrophobic interaction, if there can fit a water between, then it is not hydrophobic
- mutation get area on the left
- then separate chains of mutation in PDB and put it into excel
- aim for r squared to be up to 0.7
- put in pisa results
- at the very right put in 4 - the length of the hydrogen bond
- add together at the bottom, do for hydrogen bonds and the salt bridges
- place in a separate location
- in alphabetical order, both hydrogen bonds and salt bridges
- mutant minus wild type
- write calculated score with hydrogen bond score and salt bridges
- score2 includes hydrogen bonds and salt bridges
- Y453Y - N501N double mutation is predicted to be stronger binding than wild type
- motifs are generally something that occurs in a very similar fashion in different proteins; whereas a domain is a single protein.

March 1st - 19th

- Incorporate background information in the methods
- Put the hypothesis in the problem section?
- Put improvements in analysis
- to calculate p value : social science statistics pearsons
- Use a flow diagram for presentation?
- False positives/negatives method?

Note: not all calculations and trials could be included as there were too many

6m0j-SWS	1.36	1.372	1.38	1.39	1.4	1.41	1.42	1.428	1.44		dASA	CORR
C	38670.7	38408.8	38146.6	37880.6	37629.5	37380.0	37145.2	36914.6	36705.9	37653.6	83.68611	-0.99876
F	38577.6	38326.0	38053.6	37817.3	37536.7	37290.5	36960.7	36837.5	36588.9	37554.3	-15.5772	0.99889
G	38691.9	38429.4	38148.5	37901.1	37640.6	37394.7	37130.0	36941.7	36693.0	37663.4	93.5583	0.99943
H	38593.8	38330.6	38069.8	37803.9	37553.0	37303.0	37069.7	36839.5	36604.9	37574.2	4.38055	0.99983
I	38610.8	38293	38092.2	37823.1	37560.0	37329.5	37086.6	36934.0	36628.5	37595.3	25.4327	0.99915
L	38614.2	38315.0	38110.2	37839.7	37578.6	37319.2	37081.9	36917.4	36632.3	37600.9	31.0561	0.99985
M	38612.6	38290.4	38086.2	37858.6	37576.3	37328.4	37086.9	36905.6	36621.4	37596.3	26.4033	0.99978
N	38664.5	38340.9	38140.8	37907.8	37627.7	37383.7	37139.9	36913.1	36673.8	37643.6	73.7077	0.99978
Q	38643	38320.3	38118.6	37852.5	37605.1	37364.1	37117.5	36890.8	36678.0	37621.1	51.2305	0.99978
S	38673.2	38421.5	38149.0	37886.6	37631.8	37385.4	37083.0	36932.0	36709.4	37652.4	82.5794	0.99895
T	38588.0	38324.2	38113.6	37891.2	37608.2	37359.7	37117.0	36934.6	36654.3	37621.2	51.3438	0.99864
V	38677.2	38362.8	38123.2	37850.1	37617.2	37366.9	37121.1	36931.2	36650.9	37633.4	63.5361	0.99909
W	38600.0	38303.0	38066.6	37800.8	37548.8	37278.6	37038.7	36839.9	36589.1	37562.8	-7.0411	0.99979
Y	38577.6	38347.3	38052.4	37788.4	37537.6	37326.8	37054.7	36824.1	36620.1	37569.9	0	0.99920

	SWS-SWS			SWS-SWS			
	SCORE	HB	SB	SCORE-2		Mutant	Score
BIND							
-0.26	-29.9111	0.005	0	-0.892333		C	-0.89233
0.25	8.57444	0	0.21	0.4672333		F	0.46723
-0.92	-16.2133	0.005	0	-0.4814		G	-0.4814
-0.07	9.42	0	0.005	0.2876		H	0.2876
-0.53	-14.8633	-0.08	-0.015	-0.5409		I	-0.5409
-0.17	-14.1972	0.845	0.23	0.6490833		L	0.64908
-0.67	-17.8927	0	0.005	-0.531783		M	-0.53178
-0.82	-22.3344	0.01	0	-0.660033		N	-0.66003
-0.37	-18.3988	0	0	-0.551966		Q	-0.55196
-0.21	-22.6677	0.005	0	-0.675033		S	-0.67503
-0.55	-9.47944	-0.005	0	-0.289383		T	-0.28938
-0.11	-32.5472	0.35	0.12	-0.506416		V	-0.50641
-0.16	-1.26333	0.46	0.165	0.5871		W	0.5871
0	0	0	0	0		Y	0
	-12.9838			-0.389516			
	13.02142			0.3906428			

Mutant	Score
C	-0.89233
F	0.467233
G	-0.4814
H	0.2876
I	-0.5409
L	0.649083
M	-0.53178
N	-0.66003
Q	-0.55196
S	-0.67503
T	-0.28938
V	-0.50641
W	0.5871
Y	0

6m0j Original		6m0j Template						
-31.455	49.474	2.505	-31.324	49.379	2.571	-0.131	0.095	-0.066
-31.359	50.852	2.04	-31.271	50.815	2.1	-0.088	0.037	-0.06
-31.051	50.892	0.548	-30.974	50.958	0.63	-0.077	-0.066	-0.082
-31.921	51.244	-0.251	-31.804	51.445	-0.13	-0.117	-0.201	-0.121
-30.297	51.627	2.826	-30.226	51.645	2.912	-0.071	-0.018	-0.086
-30.882	52.734	3.49	-30.889	52.708	3.594	0.007	0.026	-0.104
-29.822	50.528	0.169	-29.791	50.499	0.176	-0.031	0.029	-0.007
-29.424	50.561	-1.234	-29.378	50.502	-1.228	-0.046	0.059	-0.006
-30.215	49.535	-2.042	-30.186	49.508	-2.043	-0.029	0.027	0.001
-30.926	48.687	-1.5	-30.839	48.629	-1.491	-0.087	0.058	-0.009
-27.931	50.272	-1.393	-27.869	50.277	-1.468	-0.062	-0.005	0.075
-27.717	48.857	-1.32	-27.458	48.918	-1.491	-0.259	-0.061	0.171
-27.13	50.951	-0.3	-27.055	50.942	-0.353	-0.075	0.009	0.053
-30.064	49.605	-3.366	-30.164	49.611	-3.39	0.1	-0.006	0.024
-30.722	48.634	-4.234	-30.772	48.617	-4.271	0.05	0.017	0.037
-30.068	47.257	-4.105	-30.133	47.238	-4.099	0.065	0.019	-0.006
-30.75	46.226	-4.131	-30.835	46.231	-3.986	0.085	-0.005	-0.145
-30.716	49.144	-5.684	-30.717	49.105	-5.719	0.001	0.039	0.035
-31.726	50.282	-5.845	-31.722	50.28	-5.904	-0.004	0.002	0.059
-31.018	48.007	-6.66	-30.956	47.962	-6.74	-0.062	0.045	0.08
-33.16	49.804	-6.052	-33.195	49.85	-6.034	0.035	-0.046	-0.018
-28.745	47.213	-3.951	-28.784	47.161	-3.994	0.039	0.052	0.043
-28.081	45.925	-3.794	-28.042	45.924	-3.798	-0.039	0.001	0.004
-28.514	45.235	-2.507	-28.399	45.19	-2.512	-0.115	0.045	0.005
-28.792	44.03	-2.509	-28.594	43.975	-2.51	-0.198	0.055	0.001
-26.564	46.102	-3.832	-26.499	46.149	-3.863	-0.065	-0.047	0.031
-25.763	44.827	-3.582	-25.706	44.828	-3.644	-0.057	-0.001	0.062
-24.256	45.008	-3.805	-24.186	44.893	-3.774	-0.07	0.115	-0.031