Flu Fighters: The Ever-Evolving Need for New Flu Vaccines

We all know the drill: every year, we have to get a new flu shot to ensure our immunity against the influenza virus is renewed. But why do we need it every year? How is each new vaccine different from the last? The goal of this project is to help you understand how vaccinations function and how changes in the sequence and structure of the vaccine target affects vaccine efficacy.

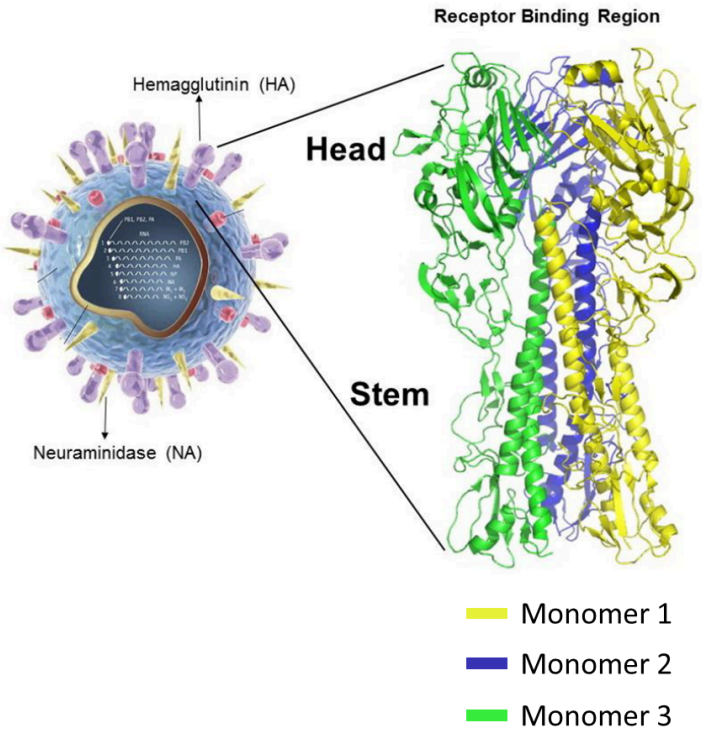
Pre-activity prep:

- Read the background information provided to you to get an understanding of topics relevant to this project
    - Some topics you (the instructor) might want to consider for background readings are the following. Feel free to modify this list and add or remove resources/readings for students as needed based on student level.
    - Intro to vaccines: types of vaccines; vaccine antigens and the role of antibodies in eliciting immunity; herd immunity and the importance of vaccines in public health
    - Intro to the influenza virus: contribution to disease and R0 (thereby eliciting the need for a vaccine); basic structure of the virus
- Complete the pre-activity quiz.

PART 1: INFLUENZA H1

There are four types of influenza viruses: A, B, C, and D. Influenza A and B viruses cause seasonal epidemics of disease in people (known as flu season) almost every winter in the United States. Influenza A viruses are divided into subtypes based on two major proteins extending from the envelope, hemagglutinin (H) and neuraminidase (N). Hemagglutinin mediates virus binding to sialic acid on host cells whereas neuraminidase promotes cleavage of sialic acid to promote viral exit. There are 18 different hemagglutinin subtypes and 11 different neuraminidase subtypes (H1 through H18 and N1 through N11, respectively). Current subtypes of influenza A viruses that routinely circulate in people include H1N1 and H3N2, and the current flu vaccines target these hemagglutinin subtypes specifically. For our purposes, we are going to focus on H1.

The H1 protein is composed of 3 identical monomers (single units) that come together to form the full protein. Each monomer protein, when folded into its final structure, has a “head” and a “stem”. In the amino acid sequence for each monomer, the first ~345 amino acids become the “head”, and the remaining ~221 amino acids fold into the “stem”. The total amino acid sequence for each monomer is therefore approximately 566 amino acids long.



7 amino acid sequences of **H1 monomer 1 variants** have been collected together in an NCBI Collection. Each of these H1 sequences has been used as a vaccine antigen in a flu vaccine.

1. Open the link to view the sequence list.
   H1 Monomer 1 Amino Acid Sequences

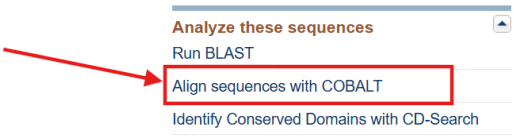
2. Click on the first sequence. This will open a new page with details about that particular H1 protein.
   a. Right under the main title, you will see a "GenBank" Accession number. This is an identifier for the sequence in the NCBI database. In the table below, note down the GenBank Accession number in the corresponding column.
   b. Halfway down the page, there is a "Features" section. This lists important information about the protein and its source virus. Find the /geo_loc_name (indicating the geographical location where the source virus was isolated) and the /collection_date (indicating when the source virus was isolated. Note both things in the table below.
   c. Repeat this for all the H1 sequences, and note down the information in the table below.

GenBank Accession Number	Location of Source Virus	Date of Collection of Source Virus
WEY08903.1	Victoria, Australia	2022
WBO08838.1	Wisconsin, USA	2022
QLF80309.1	Hawaii, USA	2019
QFR38303.1	Delaware, USA	2009
QBL89789.1	Michigan, USA	2015
AHG96683.1	Brisbane, Australia	2007
AGQ47728.1	New Caledonia	1999

What can you note about the distribution of the source viruses in terms of location and date of isolation?

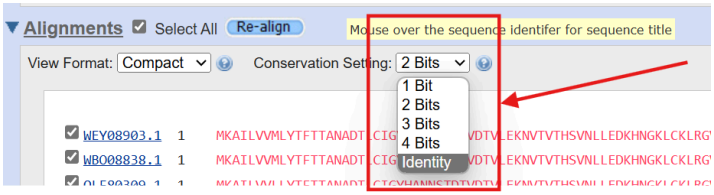
The viruses have been isolated from various locations spread out across the globe, over several different years in the last 20+ years, i.e. they were not found close to each other in terms of either time or location.

3. On the right side of the main H1 Sequences collection page, click on “Align Sequences with COBALT”



4. The new page will have a Query window with all the GenBank Accession numbers already populated.
    a. Delete any FIVE of the GenBank IDs, leaving TWO behind.
    b. Now click the blue “Align” button.

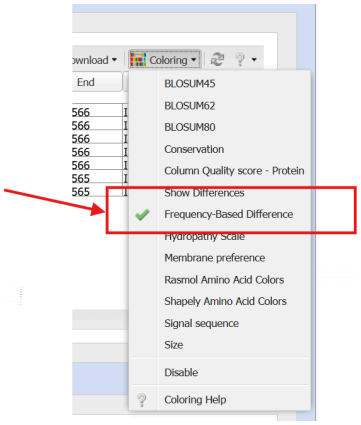
5. The COBALT alignment window will have three embedded sections: “Graphical Overview”, “Descriptions”, and “Alignments”.
    a. In the “Alignments” section, make sure the Conservation Setting dropdown menu is set to *Identity*.



What do you think red letter coloring indicates for the amino acid sequence? What do you think blue letter coloring indicates?

The red colored letters are amino acids conserved between the two proteins, i.e. they are the same in both proteins. The blue letters are amino acids that are different between the two proteins.

    b. In the “Graphical Overview” section, change the Coloring dropdown menu to *Frequency-Based Difference*.



How does what you see now relate to what you see in the "Alignments" section?

The amino acid sequence for each protein is depicted as a long bar, with gray color representing the conserved areas and red areas representing the variable regions between the two proteins.

6. Go back to the main H1 Sequences collection page (linked in Step 1), and click on Align Sequences with COBALT as before. This time, align <u>all seven</u> sequences together.

7. Repeat step 5 for the "Alignments" and "Graphical Overview" sections, in that order.

   What do you notice now about the alignment between the H1 variants?

   All the H1 variants seem to have variability in the same regions or parts of the protein across the 7 sequences.

Through this exercise, you have essentially seen real-time evidence of **antigenic drift**. Antigenic drift in a virus consists of small changes (mutations) in a gene that leads to changes in the protein. Changes associated with antigenic drift can accumulate over time as the viruses replicate, and may eventually result in viruses that are no longer recognized by the antibodies generated against the original protein.

What do you think antigenic drift means in the context of the flu vaccine?

The variable regions between the 7 H1 variants represent antigenic drift. If these variable regions are the places where antibodies bind to neutralize the virus, then changes in these regions probably make the antibody-binding inefficient or weaker. This would lead to the vaccine not working as well anymore against the changed or mutated protein and virus.

Online tools can help scientists visualize what the molecular structure of a protein looks like (for a solved structure) or would potentially look like (for an unsolved structure). We are going to use NCBI iCn3D tools to visualize the H1 structure.

8. Click on the following link for the H1 3D structure in iCn3D:
   H1 3D Structure

9. Use your mouse to play around with rotating the 3D structure.

   What do you think the yellow, green, and blue colors represent?

The three colors represent the three monomers of an H1 protein, with each monomer having a “head” and a “stem”.

10. Rotate around to the yellow-colored region of the H1 protein. This “head” of the H1 protein was aligned to the “head” sequence of another H1 variant, and the differences between the two sequences are highlighted in pink.

    Are the pink-highlighted areas buried within the “head” or more exposed to the outside of the protein structure?

    The pink regions seem to be all on the outermost surface-exposed part of the H1 Monomer 1 head.

Checkpoint Questions:

- Notice in your COBALT alignment of amino acid sequences where the head sequence ends and the stem sequence starts. Which part of the protein seems to have more variability?

  The proteins seem to have the most variability in the head regions, with only a few small variable regions in the stems.

- Think about antigenic drift and what it means for vaccine design. If you were to design a vaccine targeting a small section of the H1 protein to generate antibodies, what part(s) would you choose? Why?

  To design a vaccine, I would try to choose the part of the protein that displays the least amount of antigenic drift, so that the vaccine can still effectively target the protein even if small changes occur in other parts. Therefore, the stem portion or generally the conserves portions would be good options for a vaccine target.

- The flu vaccines we use each year target mostly the head regions of the H1 protein. Does that sound like the smart thing to do? Why do you think the vaccines are designed to target the H1 head and not the stem?

  The head region of H1 seems to have the most variability, so at first glance this does not seem to be a good choice. However, it’s possible the stem region is not accessible to the antibodies somehow - perhaps it is buried within the virus envelope as an anchor - and therefore a vaccine would HAVE to target the exposed head region instead.

- Think about the variable regions highlighted in pink in the 3D structure, and how they are mostly surface-exposed. How does that directly affect the longevity and efficacy of a flu vaccine that might be targeting the H1 head regions?

For a protein region to be an effective antibody target, it needs to be surface-exposed so the antibody can access it and bind to it. Therefore, the parts of the H1 protein buried within the head or stem will not work as good vaccine targets. However, it seems that the best surface-exposed areas are also the ones with the most variability. This means the only good choice for flu vaccine targets will keep changing, and therefore the vaccines will not stay effective over time as new H1 variants emerge.

PART 2: INFLUENZA H5

We have talked about antigenic drift, but there is another type of change that can occur in a gene called **antigenic shift**. Antigenic shift is a major abrupt change in the gene (and therefore the protein); in the hemagglutinin protein of a virus like influenza A, it can result in a new subtype of the protein that is significantly different from the original.

H5N1 is a highly pathogenic avian influenza virus that incorporates the hemagglutinin subtype **H5**. This virus is able to spread efficiently among wild birds and poultry in different parts of the world. Some human cases have also been reported globally. While human-to-human transmission of this virus does not seem to happen right now, constant vigilance of influenza viruses in birds and poultry is important - previous influenza pandemics have often been caused by an avian or swine influenza virus mutating or combining with a human influenza virus to become deadly.

Let's look at the H5 amino acid sequence alongside an H1 amino acid sequence.

1. Open the H1 Sequences collection page again (linked in Step 1 of Part 1), and click on Align Sequences with COBALT as before.
2. This time, in the Query window with the GenBank Accession numbers for the seven H1 sequences, ADD the following H5 Sequence Accession number in a new line:
   ABW90135.1
   Click the blue "Align" button.
3. Repeat the steps for the "Alignments" and "Graphical Overview" sections as before.

   Looking at both the sequence alignment and visual alignment of H5 with H1 sequences, what do you notice about H5?

   H5 seems to have even more variability compared to the variability within the H1 variants. This is especially noticeable in the Graphical Overview, there are a lot more red areas for H5 compared to the H1 variants. The increased variability in more regions is a good indication of antigenic shift

Checkpoint Questions:

- Imagine if H5N1 suddenly learned to transmit between humans and started spreading among the human population. Based on what you have learned so far about H1 and H5 amino acid sequences and structures, do you think the current flu vaccines (targeting H1) would work against H5? Why or why not?

  The current flu vaccines target the H1 head surface exposed regions. We have already seen that these regions are prone to antigenic drift. It seems that H5, a result of antigenic shift, is even more different in its sequence (and in many more regions). Therefore, the antibodies generated by the vaccine against H1 would most likely not be effective against H5, and the vaccine would not work well to protect against H5N1.

PART 3: MEASLES

Hemagglutinins are, in fact, surface proteins produced by many different viruses, helping them to attach to host cells and infect them. One such virus is the measles virus, and the measles vaccine is also designed to target the hemagglutinin protein.

1. Open the following collection 4 different measles hemagglutinin amino acid sequences: Measles hemagglutinin sequences

2. Similar to what you did with H1, align the sequences with COBALT.

   What do you notice when you compare the measles virus hemagglutinin variants?

   They seem to have very little variability in amino acid sequence between the variants.

Checkpoint Questions:

- What does the rate of antigenic drift for measles virus hemagglutinin mean for the measles vaccine?

  Since it seems that there is very little antigenic drift at all for the measles hemagglutinin protein, a measles vaccine would only need to be designed once and the antibodies it generates in your body would stay effective against the virus over a long time. Therefore, unlike the influenza vaccine, a single measles vaccine is enough to achieve long-term immunity.

- Suggest some hypotheses for why you think the measles virus hemagglutinin protein displays such low levels of antigenic variation.

  It is possible that any mutations the measles virus makes over time are in proteins that are inside the virus rather than surface proteins. Perhaps mutating the surface hemagglutinin protein makes the measles virus weaker somehow, or more likely to die out, or less effective at infecting cells; if any of these reasons were true, then the virus would avoid mutating that protein.