Introduction and background:

Understanding what host proteins are interacting with the viral proteins can give us insights on the mechanisms of virus infecting the cells and can potentially lead to drug targets. To understand such host-pathogen interactions, there have been many studies that look at protein-protein interactions (PPI) using affinity-purification mass spectrometry (AP-MS). Using the AP-MS method, recently, UCSF scientists have identified the human proteins that are physically associated with each of the 26 SARS-CoV-2 proteins.

Proteins interaction happens via their three-dimensional structures, but the structures of the proteins are not always readily available. However, we can easily access the primary amino acid sequence of all the proteins.

Question:

1. Given primary amino acid sequences of the viral proteins, can we predict their host protein interactors?
2. Would proteins of similar sequences interact with the same host proteins?

Database:

There are a handful of databases on host-pathogen protein-protein interactions, such as Chlamydia, Ebola, HIV, and SARS-CoV-2.

PPI data available here:
https://drive.google.com/file/d/1FO67Ol-JtaR9P9wR5LbyAFrXmWoWTN_q/view?usp=sharing

Analysis:

- Identify sequence similarities among pathogen proteins
- Identify sequence similarities among the host proteins that the same pathogen protein interacts with
    - Is there a more active part of the viral protein sequence that aligns to the host proteins?
- Predict the host proteins for pathogen proteins with similar sequence
- What is the caveat of just using the protein sequence in predicting PPI?