

COMPUTING CoVID-19: SUMMER 2020

COMPUTING THE RECEPTOR:

GENETIC ALIGNMENT AND BLAST

OF THE ACE2 GENE

Developers:

Em Ambrosius

June 8, 2020

COMPUTING CoVID-19. COPYRIGHT HELD BY THE NORTH CAROLINA SCHOOL OF
SCIENCE AND MATH, MAY 1, 2020. ALL RIGHTS RESERVED.

INTRODUCTION

SARS-CoV and SARS-CoV-2 (CoVID-19) both use the ACE2 receptor in their infection of the human body, so learning more about the ACE2 protein and gene aids in the understanding of CoVID-19's evolution as well as its infection pathway.

The full name of the ACE2 protein is the angiotensin-converting enzyme 2 which is located on the outer cell membrane for cells in the lungs, arteries, heart, kidney, and intestines. ACE2 is involved in lowering blood pressure so is a promising drug target for many cardiovascular diseases. Most notably, ACE2 acts at the entry point for multiple coronaviruses since it is a transmembrane protein where the S protein in coronaviruses can bind. The image below shows the infection pathway of SARS-CoV-2 including how it binds to a cell through the ACE2 protein.

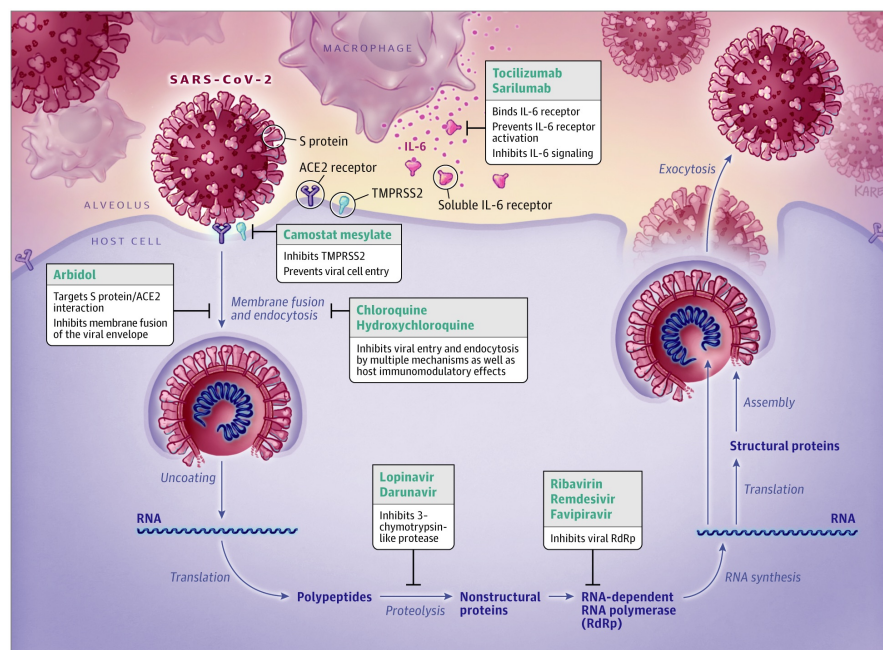


Figure 1: SARS-CoV-2 infection pathway and possible drug targets.

Protein alignments are a method of comparing two or more protein segments in order

to draw conclusions about the relation of organisms and the functionality of their proteins. Other sequence alignments are a commonly used tool in bioinformatics and can be used for DNA, RNA, and protein sequences in order to determine functional, structural, or evolutionary relationships between these sequences. [1]

2

STUDENT ACTIVITY

This activity will utilize the bioinformatics database **UniProt** and **NBCI's BLAST Tool**. UniProt or the Universal Protein Resource stores both protein sequence and annotation data. UniProt uses data from many other bioinformatics databases such as Ensembl and Swiss-Prot. [2] BLAST is the Basic Local Alignment Search Tool made by the National Center for Biotechnology Information. [3]

2.1 PART 1 – INTRODUCTION TO GENETIC ALIGNMENTS

In this first section of the lab, you will get an introduction of the process of manually completing a sequence alignment. For this you need to access the Sequence Alignment Scoring Matrix and the BLAST SciTable documents.

You will perform an alignment on the words PEACE and PIECE using the Needleman-Wunsch algorithm. [4]

This alignment will work with each combination of letters being compared in order to find the best path through the matrix to serve as the alignment. If there is a match between the letters it will have the value of +1 added, and if it is a mismatch or gap between the letters it will have the value of -1. For each square of the matrix 3 scores of the horizontal, vertical, and diagonal scores will be calculated with the largest number being kept. Depending on the direction that the kept score was from, an arrow will be drawn. At the end these arrows will be followed to complete the alignment.

The process of a manual sequence alignment is as follows:

1. With the scoring matrix, first write the words GENES and JEANS with one horizontal

and one vertical outside of the matrix. Make sure that each letter corresponds to one row or column.

2. The axes of the alignment matrix should already be filled in so that the top-left corner or the matrix has the value of zero and that the first row and column are filled with increasingly negative values from the top-left corner.
3. Start with a comparison of the first two letters and work from left-to-right across the row then moving to the subsequent rows.
4. If the first two letters match, one point is added to each of the scores above, to the left, and the left diagonal values that are already filled in the matrix. If the first two letters are different, one point is subtracted from each of the three scores. The largest of the three scores is kept and an arrow is drawn showing what direction that score came from. Figure 2 shows an example alignment up to this point.

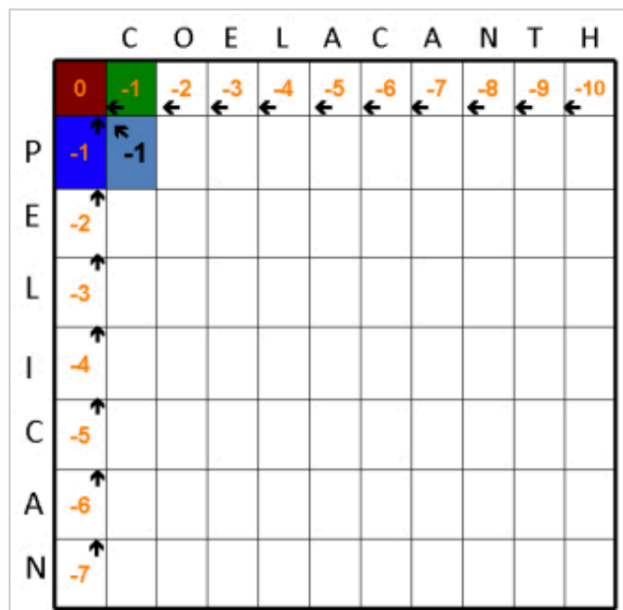


Figure 2: Example Alignment: For the first square one was subtracted from 0, -1, and -1 since C and P are different numbers. This leaves $0-1=-1$ as the largest value coming from the diagonal which is kept in the box and the arrow is drawn going to the diagonal.^[4]

5. Continue this process from left-to-right across the first row and then moving down the rows. Completion of this step should result in every box having an arrow and number.
6. Start from the bottom-right corner of the matrix and highlight that box. Follow the arrow in that box to the next box. Continue following the arrow until there are no more

arrows to follow. Figure 3 shows an example alignment with all of the important boxes highlighted.

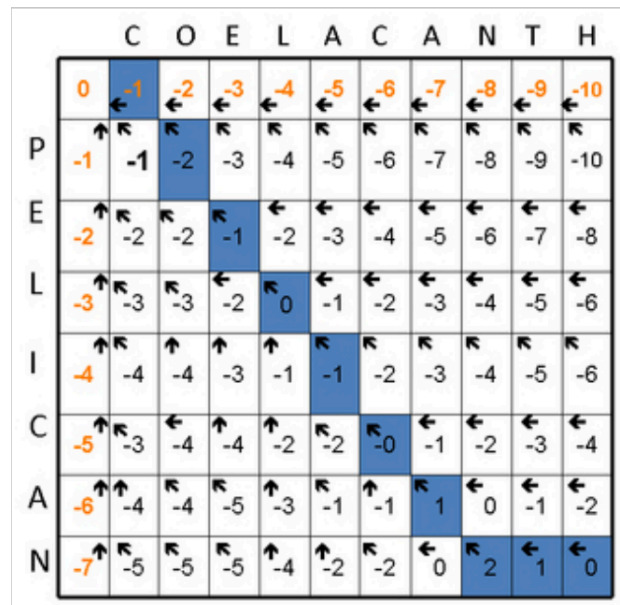


Figure 3: Example Alignment: The boxes have been filled in and highlighted by following the arrows from the bottom-right corner.[4]

- From these highlighted boxes, the alignment can be determined. If the arrow in a highlighted box points diagonally, the letters are either a match or a mismatch and then should be written directly corresponding to each other. If the arrow is pointing to the left, a dash will be added to in the word in written vertically (PELICAN). If there is an arrow pointing upwards, a dash will added in the word written horizontally (COELACANTH). The solution for this example alignment is shown below.

C O E L A C A N T H
- P E L I C A N - -

2.2 PART 2 – HUMAN ACE2 ALIGNMENT

Go to www.uniprot.org[2]. Search in the menu at the top of the page for the ACE2 gene. Find the human gene and click on that entry. This entry shows a ton of information on this gene in humans including its function, cell location, cell expression, interactions, structure, family, and similar proteins. Read through to find the general function and location of this gene to prepare for quiz questions.

Go to the "Align" page in UniProt found on the top menu of the home page. It will prompt you to input the identifiers for each protein you are trying to align. Choose the human ACE2 protein and the mouse ACE2 protein (you can find these identifiers in the table which results from searching for the ACE2 gene). Perform an alignment between these two proteins. You should come to a screen that looks similar to the image below.

```

1  MSSSSWLLLSLVAVTAAQSTIEEQAKTFLDKFNHEAEDLFYQSSLASWNYNTNITEENVQ      60
1  MSSSSWLLLSLVAVTTAQSLTEENAKTFLNNFNQEAEDLSYQSSLASWNYNTNITEENAQ      60
   *****:*** **:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
61  NMNNAQDKWSAFLKEQSTLAQMYPLQEIQNLTVKLQLQALQQNGSSVLSEDKSKRLNTIL      120
61  KMSEAAKWSAFYEEQSKTAQSFSLEIQTPIIKRQLQALQQSGSALSADKNKQLNTIL      120
   :*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
121 NTMSTIYSTGKVCNPDNPQECLLLEPGLNEIMANSLDYNERLWAWESWRSEVGKQLRPLY      180
121 NTMSTIYSTGKVCNPKNPQECLLLEPGLDEIMATSTDYNSRLWAWEGWRAEVGKQLRPLY      180
   *****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
181 EEYVVLKNEMARANHYEDYGDYWRGDYEVNGVDGYDYSRGQLIEDVEHTFEEIKPLYEHL      240
181 EEYVVLKNEMARANNYNDYGDYWRGDYEAEGADGYNRYNRNQLIEDVERTFAEIKPLYEHL      240
   *****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
241 HAYVRAKLMNAYPSYISPIGCLPAHLLGDMWGRFWTNLYSLTVPFQKPNIDVTAMVDQ      300
241 HAYVRRKLMNTYPSYISPTGCLPAHLLGDMWGRFWTNLYPLTVPFAQKPNIDVTAMMNQ      300
   *****:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*
301 AWDQRIFKEAEKFFVSVGLPNMTQGFWNSMLTDPGNVQKAVCHPTAWDLGKGFRIILM      360
301 GWDAERIFQAEKFFVSVGLPHMTQGFWNSMLTEPADGRKVCHPTAWDLGHGDFRIKM      360
   .***:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*:*

```

Figure 4: Graphic of an example alignment.

From this page you should be able to answer the following questions. These questions should serve as a guide as to what will be on the quiz. Some answers will require outside research.

2.2.1 QUESTIONS

1. How many amino acids are in each ACE2 protein?
2. What amino acid do both of the proteins start with? How is this significant?
3. What do the notations of "*" ":" and "." mean for the alignment?
4. How many amino acids are identical between the two proteins?
5. How many amino acids are similar between the two proteins?
6. What does it mean that the amino acids are similar?
7. What is CLUSTALO?
8. Do you think these two proteins are similar enough to be orthologs?

This link [here](#) may aid in answering the questions above.^[2]

2.3 PART 3 – BLAST OF HUMAN ACE2

This section of the lab will use NCBI's BLAST tool to find similar genetic sequences to the reference sequence of the ACE2 gene.

1. Go to <https://www.ncbi.nlm.nih.gov/>. Search the top bar for "ACE2." Select the top gene entry of the overall page for the ACE2 gene. This page is a culmination of the information known about the human ACE2 gene.
2. Scroll down to the "NCBI Reference Sequences (RefSeq)" section of the page. This shows the list of annotated sequences of genomic, mRNA, and protein sequences of the ACE2 gene.
3. Click on the "FASTA" link under the genomic sequence. This will take you to the full genetic sequence of the ACE2 gene.
4. The right sidebar allows you to directly choose to "Run BLAST" on this genetic sequence. This link will take you directly to the BLAST website with the intended sequence entered into the query.
5. Nothing needs to be changed from the standard BLAST settings, so scroll down and click the blue BLAST button.
6. The job will begin running and the page will reload as it continues. When it is finished you will have access the list of related genetic sequences to the reference sequence.

From the BLAST results you may need to read more about their scoring methods and what each value associated with related sequences means. Useful information can be found at these links:

[Introduction to NCBI's Bioinformatics Resources: BLAST Results](#) [5]

[Explanation of the BLAST output](#) [6]

[Interpreting BLAST output](#) [7]

REFERENCES

- [1] Wikipedia Contributors. Sequence alignment, May 2020.
- [2] The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*, 47(D1):D506–D515, nov 2018.
- [3] J Zhang and T L Madden. PowerBLAST: a new network BLAST application for interactive or automated sequence analysis and annotation. *Genome research*, 7(6):649–656, jun 1997.
- [4] I Lobo. Basic local alignment search tool (BLAST). *Nature Education*, 1(1), 2008.
- [5] Introduction to ncbi bioinformatics resources: Blast results.
- [6] Explanation of the blast output.
- [7] Interpreting blast output, 2008.