

- 1) Implemented in NeedlemanWunsch.py
Usage can be seen with -h
- 2) Ran with: `python3 NeedlemanWunsch.py -q hw2_files/pfizer_mrna.fna -r hw2_files/sars_spike_protein.fna -o pfizer_sars_wgap.txt -g`
Therefore the data is in pfizer_sars_wgap.txt
- 3) Ran with: `python3 NeedlemanWunsch.py -q hw2_files/pfizer_mrna.fna -r hw2_files/sars_spike_protein.fna -o pfizer_sars_nogap.txt`
Therefore the data is in pfizer_sars_nogap.txt
- 4) Foremost, the alignment score for no start/end gap penalty is 1711 instead of 1075. The alignment places long start and end gaps in the SARS-Cov-2 sequence and in the middle 'section' there are few gaps. For the alignment with start and end gap penalties, the start and end gaps are somewhat present, but with infrequent matches in these areas, and eventually getting into a similar center section with few gaps
- 5) The start and end gaps should be ignored because, as in the image, the encoding region is in the middle, and this is what can be easily aligned with the vaccine. Lacking length and non-encoding regions should not affect the quality of the alignment of the encoding region.
- 6) The original output is trimmed in SARSencoding.txt and found to have:
1033 (x) mismatches, 2780 (|) matches , 18 () gaps
Therefore , there are 1051 mismatches in the alignment.
- 7) Called with `python3 AAconvert.py -s tmpfile.txt -o pfizer_mrna.aa`
And `python3 AAconvert.py -s tmpfile.txt -o sars_spike_protein.aa`
- 8) Called with `python3 NeedlemanWunsch.py -q sars_spike_protein.aa -r pfizer_mrna.aa -o SarsPfizerAAalignment.txt -g` Results are therefore in 'SarsPfizerAAalignment.txt'
It's found that there are 2 mismatches in the AA sequences at AA #986 and #987 where K and V are each replaced with P in the vaccine.
- 9) The article says that a 2P mutation is what is performed in the vaccine. In a 2P mutation, two prolines are added to stabilize the protein. In the vaccine sequence The P mutation is the addition of prolines the article describes.
- 10) The Pfizer vaccine GC content = 57% where the spike protein GC content is =37%.
This was done to increase the stability and melting temperature of the vaccine. [Source](#)
This is the reason for the large number of mutations in the DNA sequence, but few changes in the AA sequence; the vaccine is designed to have a higher GC content