Homework 2: Sequence Alignment

Samuel Kitzerow, kitze012

October 11, 2022

1 Compare your alignments from #2 and #3. In what way did the free start and end gaps improve the alignment?

The free start and end gaps improved the alignment by reducing the restrictions imposed on the alignment by the gap penalties. This allowed the two sequences to better align with each other without more gaps in the middle of the sequence which also improved the alignment score.

2 Based on this annotated image of the full sequence of Pfizer mRNA vaccine, why did it make sense to ignore start gaps and end gaps in your alignment in #3?

It makes sense to ignore the start and end gaps for the full sequence of the Pfizer vaccine because it is the COVID-19 spike protein that we are trying to align it with. Ignoring the start and end gaps will enable to two proteins to align; whereas, not ignoring the gaps will will cause them to dis-align and produce poor results.

3 How many mismatches (counting gaps as a mismatch if any) are there between the real spike protein and the Pfizer version, in the coding portion of the RNA sequences?

There were 998 mismatches and 271 gaps observed between the real spike protein and the Pfizer version with start/end gap penalties enforced. When the start and end gaps where ignored, there was 1004 mismatches and 317 gaps observed between the two sequences (54 start/end gaps were ignored).

4 How many mismatches are there in the two amino acid sequences (counting gaps as a mismatch if any)? What exactly is different between the amino acid sequence of the vaccine and the amino acid sequence of the real spike protein?

There are two mismatches between the amino acid sequences of the vaccine and the real spike protein. No gaps were observed.

5 Describe why your findings in #9 make sense in the context of the article.

The two mismatches observed previously are two Proline amino acids added to the key joints of the vaccine's spike protein. It stabilizes the structure's prefusion shape and prevents the spike protein from spring into an elongated postfusion form. Stabilizing the prefusion found allows the body's immune system to recognize the virus and build a defence against it.

6 Why did the vaccine makers introduce so many synonymous mutations into the vaccine? Why didn't they just copy the spike protein sequence exactly (apart from the amino acids discussed in question 10)?

The vaccine makers introduced many synonymous mutations into the vaccine and didn't flat out copy the spike protein sequence because a protein sequence is not as important as the protein conformation. The vaccine could have the right encoding for the protein but if the shape isn't right then it won't work. Some of the extra mutations help stabilize the shape of the structure.