Q1: NA

Q2: NA

Q3: NA

Q4: We can see from the result that the **SARS reference is part of the mRNA vaccine** (∗), so forcing the sequence to align from the very beginning has a large chance to result in incorrect alignments. Unpenalized start gaps allow the alignment to skip over uninformative regions and start the alignment where it makes the most biological sense. Similarly, unpenalized end gaps enable the alignment to end where it's appropriate.
(∗) Q5 will answer this question.

Q5: From the image, we could see that basically SARS reference is the coding region, meanwhile the vaccine also includes 5' UTR before start codon ATG and 3'UTR after stop codon TGA, so we do not need to match the vaccine to SARS reference from the beginning since there are no reason to match the UTR regions to the reference.

Q6: In the coding portion, there are 1051 mismatches. (1033 "x" and 18 gaps)

Q7: NA

Q8: NA

Q9: There are only 2 mismatches in the middle, shown in the figure, which means in the coding region, there are only 2 amino acids are different between these sequences. (DPPE in vaccine and DKVE in reference)

Q10: As described in the article, "Through a bit of structural biology and persistent protein engineering, McLellan discovered that adding two prolines—the most rigid of the 20 amino acids—to a key joint of a vaccine's spike protein could stabilize the structure's prefusion shape. This 2P mutation worked in preclinical studies of Graham and Moderna's MERS vaccine, so they applied it to Moderna's COVID-19 vaccine." (https://cen.acs.org/pharmaceuticals/vaccines/tiny-tweak-behind-COVID-19/98/i38). It is just the differences we have detected: DPPE in vaccine and DKVE in reference.

Q11: Different species have different preferences for codon usage, which is known as "codon bias." Codon optimization involves changing the codons used in the mRNA sequence to match the preferred codons of the host organism (in this case, humans) for translation. Codon optimization fitting right GC content is critical in mRNA vaccine development to ensure that the mRNA is efficiently and accurately translated into protein, leading to a more effective and safe vaccine.

12: In my coding result, there are no gaps inside the coding region with the affine gap penalty (-10), since the penalty is really large, so when doing alignment, the algorithm tends to avoid gaps in the middle.