# **Analysis of Second Vassiliev Measure Scans for SARS-CoV-2 S Proteins**

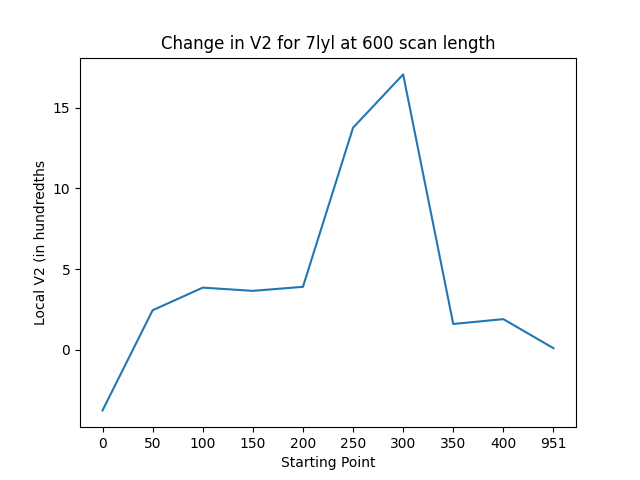
6/23/21

## Introduction

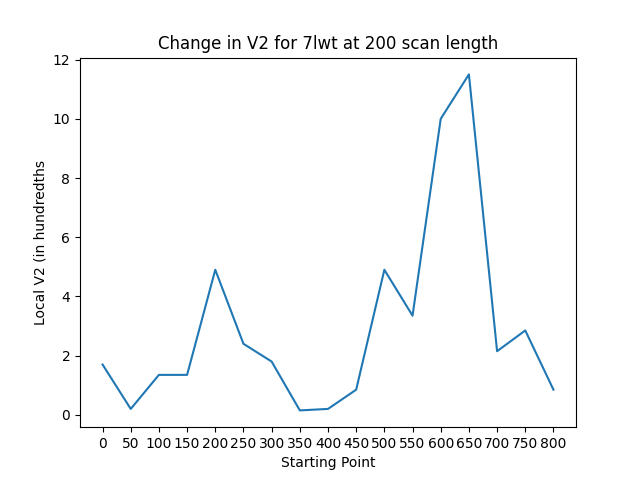
The sections following are an analysis comparing different versions of the spike proteins (S proteins) of SARS-CoV-2, its variants, and SARS-CoV from 2003. These proteins are analyzed with regards to a topological calculation called the second Vassiliev measure (V2).

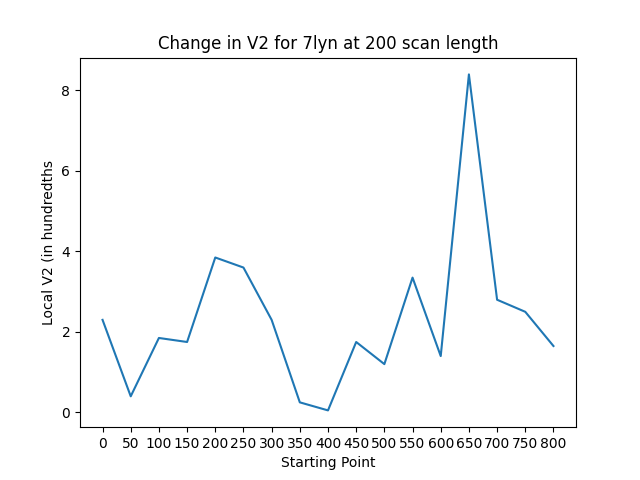
## Alpha and Beta

Both of the figures shown below display the change in V2 as the scans’ starting points increase, both figures using 600 scan lengths. These are for S proteins of the Alpha (7lws) and Beta (7lyl) variants of SARS-CoV-2, both in the RBD-down conformations. As you can see, there is an increase in V2 at the 300 starting point of each plot, which indicates the chain of the protein containing CA atoms at indices 300 to 900.



The graphs for scan lengths of 400 look similar, but with peak V2 measures at the 500 to 900 sections instead – still ending at 900, suggesting that there is a knot in the protein that becomes less of a knot after the 900-atom mark.



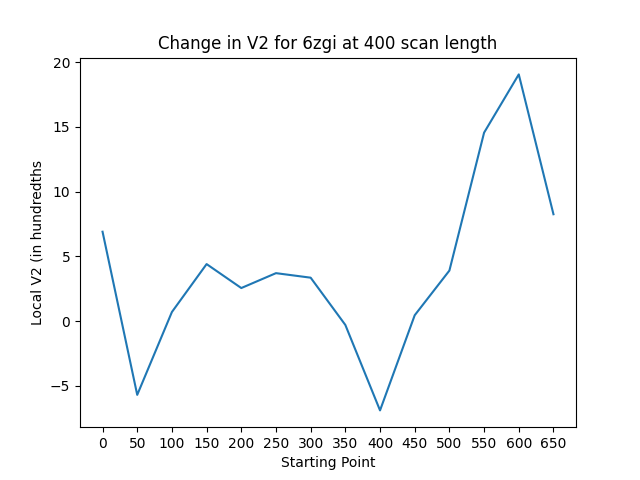
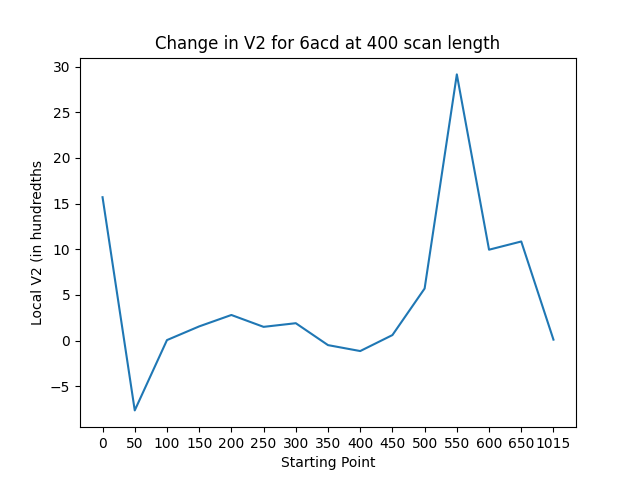
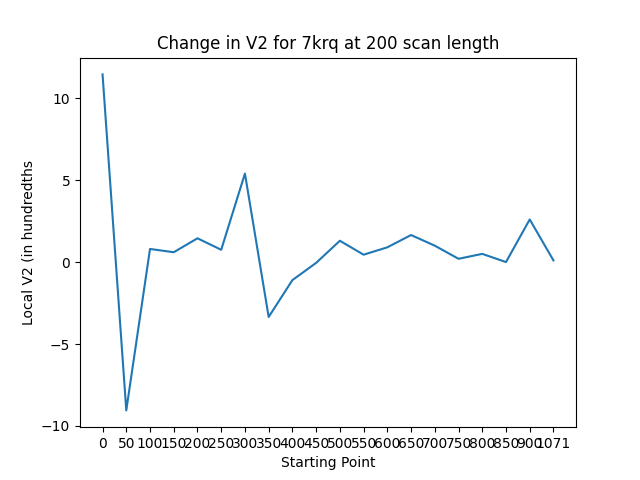


The graphs above are similar, except they are for the same S proteins for Alpha (7lwt) and Beta (7lyn) variants of SARS-CoV-2 but in the RBD-up conformation. Additionally, these graphs show the change in V2 across the entire proteins’ chains using scanning lengths of 200, rather than the 600 shown and 400 mentioned above. Still, in the RBD-up conformation, in 200, 400, and 600 scanning lengths, there is a peak in V2 at the section ending at around 900. In this case, however, the peak is actually from 650 to 850 – the V2 is a much smaller 0.0215 (0.028) at starting point 700 than the 0.115 (0.084) found at the peak for 7lwt (7lyn respectively). Interestingly enough, the same pattern can be found in the 200 scan-length scans of the RBD-down conformation versions of the S protein shown previously, where the peak is at atoms 650 to 850 rather than the expected 700 to 900.

## SARS-CoV-2 and 2003 SARS-CoV

While the 6zge (SARS-CoV-2 S protein) scans are still being processed, the scans for 6zgi, which is the furin-cleaved version of the S protein in closed conformation, have been processed. In this section I will compare 6zgi to 6acd, which is the RBD-up S protein for SARS-CoV from 2003.

The figures below show a familiar pattern, with 6acd’s V2 peak at 550 to 950 and 6zgi’s V2 peak at 600 to 1000. However, there is a stark difference between these proteins’ plots and those of the variants in the previous section. These plots show a reversal in sign for V2 in both proteins as the starting point progresses from 0 to 50. This is repeated in some other proteins’ scans, as with that of 7krq (a version of the S protein with a common mutation) shown as well.

7/7/2021

**Variant Spike Protein Graphs**

The 600 RBD-Down graph suggests that perhaps the RBD and a little beyond the RBD domain has a different starting point for the different RBD-Down proteins. This is because the peak is at different amino acid residues, and because they start around 350 (where RBD begins). However, when we look at the 400 scan-length graph, the peak’s location around 590 suggests that it is the section of the proteins that ends at amino acids around 950-1000 which is knotted in a way that increases V2 rather than the section starting around 350.

The same pattern can be seen in the graphs for SARS-CoV-2 variants in the RBD-Up conformation. 6xkl, 6zgg, and 6acd all have V2 peaks at around 560, whereas 7lyn, 7lwt, and 7lww have more extreme V2 peaks closer to 600 (using the 400 scan-length). For the 600 scan-length, 7lyn, 7lwt, and 7lww have extreme V2 peaks around 380. 6zgg and 6acd have smaller peaks around 360, whereas 6xkl has a V2 peak at 600 scan-length with the starting amino acid of around 320. Like the aforementioned RBD-Down proteins, this suggests that the interesting mutations are changing the V2 of the sections ending at around 950-1000, rather than around the starting residues of the RBD.

**Wild Spike Protein Graphs**

In the graphs containing 6zgh, the slopes of the graphs at points which include amino acids from 328 to 530 are misleading due to a skip in the data where there are no amino acids from indices 328 to 530. This is mainly visible in the 400 scan-length graph for the wild SARS-CoV-2 spike proteins in various conformations, where the V2 appears to increase around that skip when in reality it is because the line averages from the ordinary V2 around 300 to the V2 peak at around 560, which it shares with 6zgg. This skip is also evident in the 200 scan-length graph, where a stark dip in V2 around amino acid 415 is clearly not included in the 6zgh data due to a lack of amino acid data there.

In the same plot, it is interesting to see that there are actually two peaks for 6zgh and 6zgg, with the second, smaller peak at around amino acid 700 (around 470 for 6zgg on the 600 scan-length graph). The other two conformations, 6zge and 6zgi, do not share this secondary peak, and in the 400 scan-length graph their peaks sit around the 650th amino acid.

An earlier peak is also quite interesting in each of the wild spike proteins’ V2 because it acts the same for every scan-length – particularly with regard to 6zge, 6zgi, and 6zgg. It is around amino acid 220 for the 200 scan-length graph, 170 for 400, and 110 for 600 scan-length. The interesting part of this peak is that for 200 and 400 scan-lengths, there is a specific order for the magnitude of the peak. 6zgi has the biggest V2, then 6zgg, then 6zge – although 6zgg and 6zge are flipped for the 600 scan-length plot. This suggests a particular change in the domain including amino acids in that area from one conformation to the other, in a way that specifically increases V2 for 6zgi and appears to act differently for 6zgh. There is also an extreme dip in V2 right before this small peak, visible in the 400 and 600 scan-length plots.