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

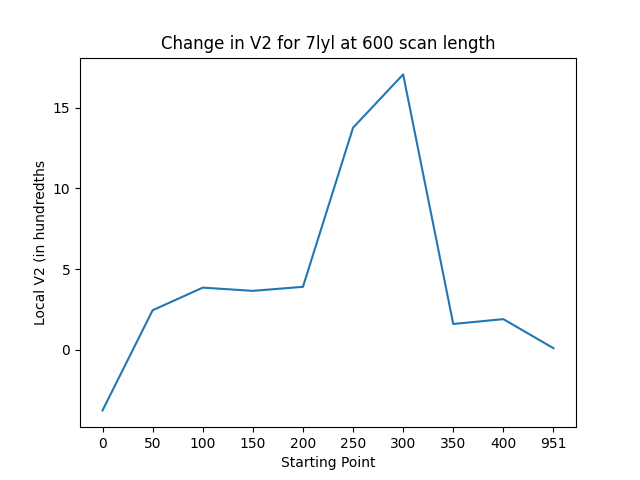
6/23/21

## Introduction

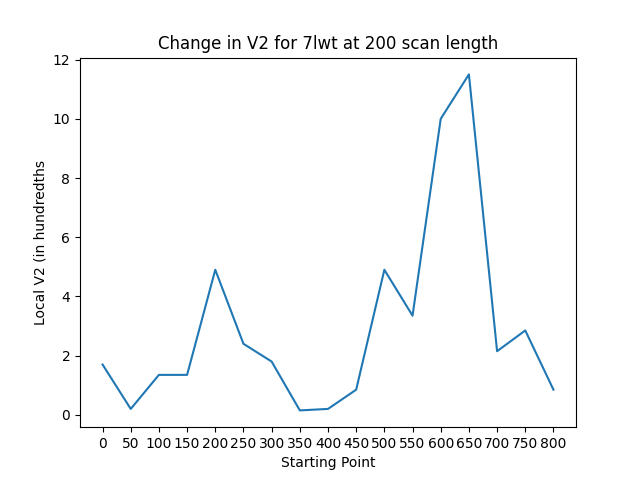
The sections following are an analysis comparing different versions of the spike proteins (S proteins) of SARS-CoV-2, its variants, and SARS-CoV from 2003. These proteins are analyzed with regards to a topological calculation called the second Vassiliev measure (V2).

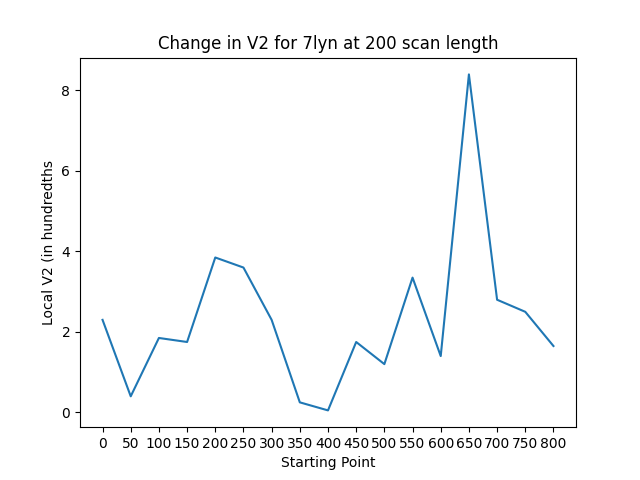
## Alpha and Beta

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.



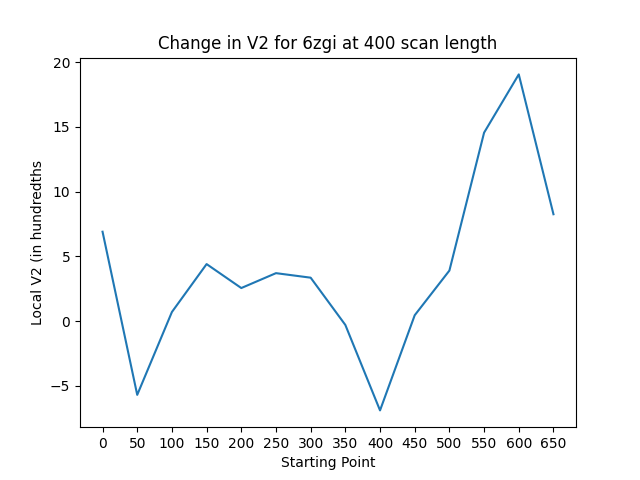
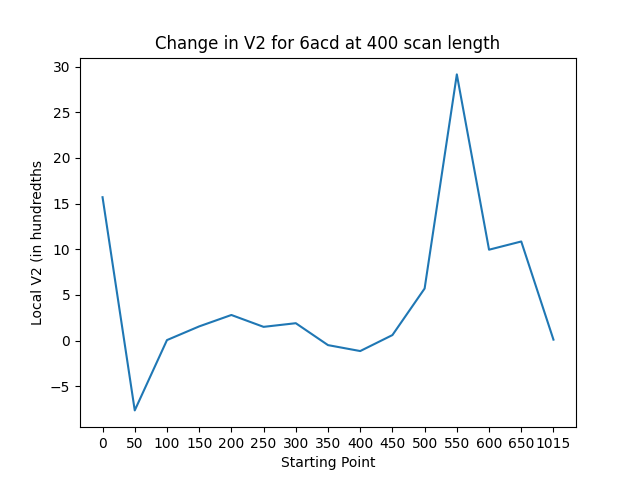
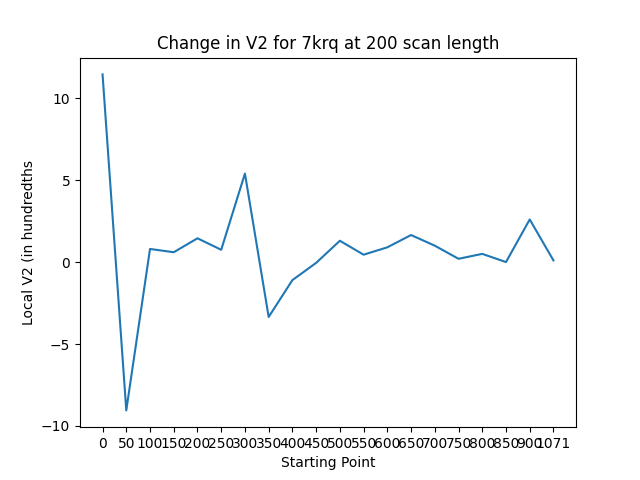


The graphs above are similar, except they are for the same S proteins for Alpha (7lwt) and Beta (7lyn) variants of SARS-CoV-2 but in the RBD-up conformation. Additionally, these graphs show the change in V2 across the entire proteins’ chains using scanning lengths of 200, rather than the 600 shown and 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.

Created 7/7/2021

We have analyzed the second Vassiliev measure (V2) of the SARS-CoV-2 spike (S) protein at different sections using various window-sliding lengths (scan-lengths) of 200, 400 and 600 CA atoms. These scans were at intervals of 50 CA atoms, for example we would scan section 0 to 200 first, then 50 to 250. After we finish through the length of the protein, we scan 0 to 400, then 50 to 450, et cetera. This is to capture the tertiary structure of these individual sections as V2 data to see if we can measure how knotted they are and determine if that relates to the importance and functionality of a given scanned section. 200, 400, and 600 scan-lengths and intervals of 50 were chosen to efficiently compute V2 at as many segments of the proteins as was feasible.

\begin{figure}

\includegraphics[width=0.45\linewidth]{./figData/6zg200.png}\\

\includegraphics[width=0.45\linewidth]{./figData/6zg400.png}\includegraphics[width=0.45\linewidth]{./figData/6zg600.png}

\caption{The second Vassiliev measure of parts of the SARS-CoV-2 spike protein in uncleaved closed (6zge), cleaved closed (6zgi), cleaved open (6zgg), and cleaved intermediate (6zgh) using scanning lengths of 200, 400 and 600. }

\label{fig:rbdupdown}

\end{figure}

First, we analyze the tertiary structure of the SARS-CoV-2 S protein in 4 pre-fusion stages (closed uncleaved, closed cleaved, open cleaved, and intermediate cleaved). These different stages have structural changes that mostly affect the receptor-binding domain (RBD) and N-terminal domain (NTD). Notably, this means that for the closed conformation, the RBD is down, closed off and unable to bind to the ACE2 receptor. In the open conformation, 1 out of the 3 RBDs exposes itself so that it can assist in the binding process, with nearby NTDs moving as well. \cite{Wrobel2020}.

These scans can be seen in Figure \ref{fig:rbdupdown}. The slopes of the graphs at amino acids from indices 328 to 530 are misleading due to a skip in the data where there are no amino acids recorded for the cleaved intermediate conformation of the protein (6zgh). 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 it is actually because the line averages from the ordinary V2 around 300 to the V2 peak at around 560. This blank spot is where the RBD should be and is likely due to the intermediate state of the protein \cite{Gobeil2021} \cite{Xia2021}.

With regard to the visible spike in the value of V2 around amino acid 600 in the scan-length 400 graph and amino acid 360 (with the exception of 6zgh) in the scan-length 600 graph, we can see that they have something in common. They do not share the same starting point, but the segments in all proteins for both 400 and 600 scan-length graphs which give the high V2, $v\_2>0.15$ have an ending point around amino acids 950-1000. This suggests that segments ending at these residues are more knotted. These ending residues align with the HR1 and CH domains, which, along with the rest of S2, are involved in membrane fusion \cite{Xia2021}.

Furthermore, we can see in the 600 scan-length graph that the cleaved open conformation protein has a much higher peak at this section than the cleaved and uncleaved closed conformation proteins. This suggests that the change from closed to open, which raises one RBD earlier along the protein, increases either the complexity or tightness of the knot at this segment.

In the same plot, we see that there are 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.

Dips at amino acids 80 and 415

\begin{figure}

\includegraphics[width=0.45\linewidth]{./figData/RBD-Down200.png}\includegraphics[width=0.45\linewidth]{./figData/RBD-Down400.png}\\

\includegraphics[width=0.45\linewidth]{./figData/RBD-Down600.png}

\caption{The second Vassiliev measure of parts of spike proteins for SARS-CoV-2 variants in the closed, RBD down conformation (6zge, wild SARS-CoV-2; 7lws, UK variant; 7lyl, South African variant; 7kdk, D614G mutated) using scanning lengths of 200, 400 and 600. }

\label{fig:closedvariants}

\end{figure}

Next, we compare the tertiary structure of SARS-CoV-2 variants in closed and in open conformation, including the closed uncleaved and open cleaved wild SARS-CoV-2 S (6zge and 6zgg respectively) . Figure \ref{fig:closedvariants} further supports the claim that there is a knotting phenomenon at the segment starting around amino acids 360-600 and ending around amino acids 950-1000. Even the 200 scan-length graph in Figure \ref{fig:closedvariants} has a peak for the UK and South African S proteins (7lws and 7lyl respectively), showing knotting as specific as around 780-980. These lower scan-length peaks are less extreme, however, with values less than 0.1.

Another point of interest is the fact that there is an order of magnitude with respect to the value of V2 for each protein at the peak. For both the 400 and 600 scan-length graphs, 7lyl (SA) has a smaller value than 7lws (UK), which has a much smaller value than 7kdk. 6zge hovers around where 7lws and 7lyl are for both graphs.

This suggests that there is a correlation between the increase in knot complexity or tightness at this segment, which increases V2, and the mutations that alter the domains that are located there. The segments in the larger scan-length graphs cross over crossing from S1 to S2, but it is notable that only the variants have the peak at the much more concentrated sections at residues 780-980, along which the FP, IFP, and HR1 domains are located. The T716I mutation found in the UK variant (7lws and 7lwt) alters the protein along this same segment, and the S982A and D1118H mutations are also found in the same variant at domains HR1 and CD, respectively. These mutations could be the reason for the increased values for V2 in these variant S proteins.

\begin{figure}

\includegraphics[width=0.45\linewidth]{./figData/RBD-Up200.png}\includegraphics[width=0.45\linewidth]{./figData/RBD-Up400.png}\\

\includegraphics[width=0.45\linewidth]{./figData/RBD-Up600.png}

\caption{The second Vassiliev measure of parts of spike proteins for SARS-CoV-2 variants (7lww, Brazilian variant; 7lyn, South African variant; 6acd, SARS-CoV; 7lwt, UK variant; 6xkl, Hexapro; 6zgg, wild SARS-CoV-2) in the open, RBD up conformation using scanning lengths of 200, 400 and 600. }

\label{fig:openvariants}

\end{figure}

Similar patterns can be seen in Figure \ref{fig:openvariants}. 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 in Figure \ref{fig:closedvariants}, this is further evidence for the knotting at segments ending around 950-1000, visible with all three scan-lengths.

Additionally, there is a clear order of magnitude for the V2 values in the Figure \ref{fig:openvariants} peaks ending around 950-1000 as follows, from least to greatest: 6zgg, 6acd, 6xkl, 7lyn, 7lww, 7lwt. This is visible for both the 400 and 600 scan-length graphs, and to a lesser degree the 200 scan-length graph. This suggests that, especially given the fact that the wild proteins 6zge and 6zgg were the shortest peaks for both groups, there is some importance to the magnitude of the V2 values at these peaks – given the fact that the graph with the wild SARS-CoV-2 proteins only has V2 from -0.1 to 0.2, it would appear that the variants have more extreme knotting in the peak segments.

A third point of interest is the fact that for these larger scan-length graphs, there is a visible difference in magnitude between the two groups mentioned thus far. The Figure \ref{fig:closedvariants} peaks hover around 0.15, which 7kdk closer to 0.4. For the RBD-up proteins in Figure \ref{fig:openvariants}, however, the peaks were more diverse and averaged around 0.4. The shortest was around 0.2 (6zgg), while the tallest was close to 0.8 (7lwt). This supports that there is a topological difference between the two conformations of protein which increases V2 as the domains change into the ‘up’ position, as suggested by Figure \ref{fig:rbdupdown}, despite the location of the peaks’ sections being farther along the protein than S1.