The thrombo-inflammation and neuropathology sequence motifs of the SARS-CoV-2 spike protein appear to have been engineered into the virus

Steven Quay, MD, PhD[1]

ABSTRACT

A landmark paper[2] entitled, "Fibrin drives thrombo-inflammation and neuropathology in COVID-19," was published in August 2024 that concluded the mechanism of the thrombotic and neurological symptoms following a SARS-CoV-2 infection, often called "long COVID," is attributable to the binding of fibrin to discrete portions of the spike protein, specifically three N-terminal domains. This paper is a high impact publication with >110,000 views, placing it in the 99th percentile of articles published contemporaneously.

Here I examine the regions of the spike protein that bind to fibrin, fibrinogen, or both. The N-terminus of the spike protein contains the three strongest binding peptides and surprisingly, these regions are also the three insertions in the protein sequence that are unique to SARS-CoV-2 and not found in natural sarbecoviruses. All pre-pandemic sarbecoviruses have either a partial deletion in these regions or have protein amino acid substitutions that are non-conserved and therefore would not support fibrin binding.

In addition, the three inserts also correspond to regions of the spike protein that have been shown previously to have high sequence homology with the HIV gp120 protein. GP120 is an HIV surface protein that binds to a host cell surface receptor on CD4+ T-cells and facilitates cell entry to begin infections. In comparing the immunological and clinical presentation of HIV and COVID-19 patients, the commonality of D-dimer production, CD4+ lymphopenia, neurotropism, and IL-10 expression strongly suggests that these protein sequence homologies are clinically relevant.

A conclusion that the pathophysiology of long COVID, based on the insertion of spike protein motifs with sequence homology that mimic the HIV gp120 protein motif properties, and that these SARS-CoV-2 motifs are not found in the sarbecovirus subgenus strongly suggests that these inserts were design features in the synthetic assembly of SARS-CoV-2.

INTRODUCTION

A recent landmark paper on the mechanism of SARS-CoV-2 pathology in COVID-19 patients was published in August 2024 entitled, "Fibrin drives thrombo-inflammation and neuropathology in COVID-19," written by Ryu et al. A 2021 paper[3] previously described the S1 spike protein interaction with fibrin. In this paper they document that fibrin binds to the SARS-CoV-2 spike protein, forming proinflammatory blood clots that drive systemic thrombo-inflammation and neuropathology in COVID-19. Fibrin, acting through its inflammatory domain, is required for oxidative stress and macrophage activation in the lungs, whereas it suppresses natural killer cells, after SARS-CoV-2 infection. Fibrin promotes

[1] ORCID; Atossa therapeutics, Inc., Seattle, WA

[2] Ryu, J.K., Yan, Z., Montano, M. *et al.* Fibrin drives thromboinflammation and neuropathology in COVID-19. *Nature* **633**, 905–913 (2024). https://doi.org/10.1038/s41586-024-07873-4

[3] Grobbelaar LM, Venter C, Vlok M, Ngoepe M, Laubscher GJ, Lourens PJ, Steenkamp J, Kell DB, Pretorius E. SARS-CoV-2 spike protein S1 induces fibrin(ogen) resistant to fibrinolysis: implications for microclot formation in COVID-19. Biosci Rep. 2021 Aug 27;41(8):BSR20210611. doi: 10.1042/BSR20210611. PMID: 34328172; PMCID: PMC8380922

neuroinflammation and neuronal loss after infection, as well as innate immune activation in the brain and lungs independently of active infection.

On January 23, 2020, Dr. Shi and her colleagues at the WIV published one of the earliest pre-print papers[4] on SARS-CoV-2, comparing it to RaTG13, a virus first collected in 2013 but not fully described in the literature until this paper. Shi notes in this pre-print: "the S genes of nCoV-2019 and RaTG13 S gene are longer than other SARS related-CoVs. The major differences in nCoV-2019 are the three short insertions in the N-terminal domain."

On January 31, 2020, a non-peer reviewed preprint[5], entitled, "Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag," was published. The abstract stated, "we found insertions in the spike glycoprotein (S) which are unique to the 2019-nCoV and are not present in other coronaviruses. Importantly, amino acid residues in all the 4 inserts have identity or similarity to those in the HIV-1 gp120 or HIV-1 Gag. Interestingly, despite the inserts being discontinuous on the primary amino acid sequence, 3D-modelling of the 2019-nCoV suggests that they converge to constitute the receptor binding site. The finding of 4 unique inserts in the 2019-nCoV, all of which have identity /similarity to amino acid residues in key structural proteins of HIV-1 is unlikely to be fortuitous in nature." The article was widely read, with over 760,000 abstract reads and was picked up by 116 news outlets. It was also, apparently, widely discussed within the NIH and NIAID agencies at the director's level.

Within 48 hours the authors withdrew the pre-print stating: "this paper has been withdrawn by its authors. They intend to revise it in response to comments received from the research community on their technical approach and their interpretation of the results." A brief review of the 135 comments received during that time focused on the short length of the inserts, the finding of these inserts in RaTG13, which at the time was believed to be a natural, unmodified, wild collected bat virus and which had been uploaded to GenBank only a few days earlier, but did confirm these sequences were absent from other coronaviruses. There were multiple requests in the comments to look at the probability of all four of the inserts in a combined fashion, but this was not done.

Here I examine the fibrin-spike protein interaction region, specifically three portions of the N-terminus of the spike protein that have the strongest affinity for fibrin, fibrinogen, or both, and find it is closely correlated with the three N-terminal inserts. I also perform a concatenation of the three inserts to allow an estimation of the probability that all three would be found in both SARS-CoV-2 and HIV by chance, as well as the probability that RaTG13, collected six years before and 1100 km from the SARS-CoV-2 outbreak, is related to SARS-CoV-2 by chance.

RESULTS

A protein blast of the inserts identifies homologies with HIV surface gp120 protein

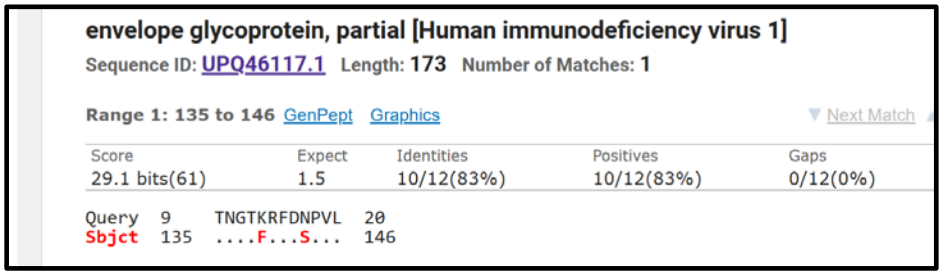
A blast of the first insert, a 20-mer covering both the fibrin binding region and the HIV-like segments (FHAIHVSGTNGTKRFDNPVL), identifies a Chinese HIV-patient sequence with high homology for

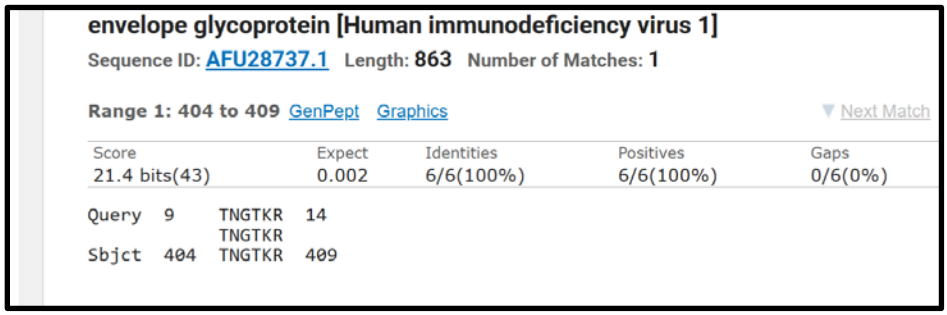
---

[4] Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. Peng Zhou, et. al., bioRxiv 2020.01.22.914952; doi: https://doi.org/10.1101/2020.01.22.914952

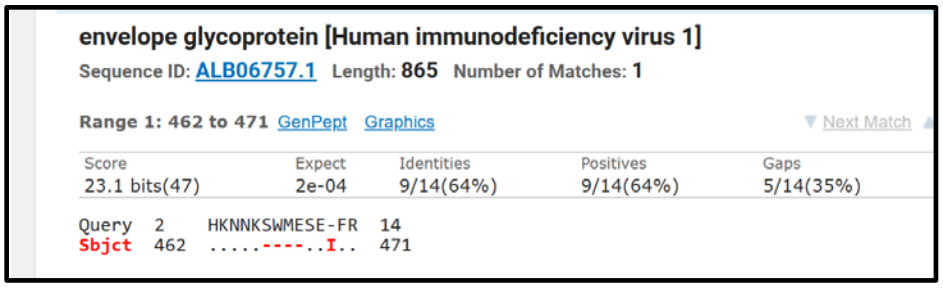
[5] Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag Prashant Pradhan, Ashutosh Kumar Pandey, Akhilesh Mishra, Parul Gupta, Praveen Kumar Tripathi, Manoj Balakrishnan Menon, James Gomes, Perumal Vivekanandan, Bishwajit Kundu. bioRxiv 2020.01.30.927871; doi: https://doi.org/10.1101/2020.01.30.927871

such a short segment. Two kinds of homology can be seen: long partial identity and short complete identity. This maps to the V4 domain of the HIV gp120 protein.

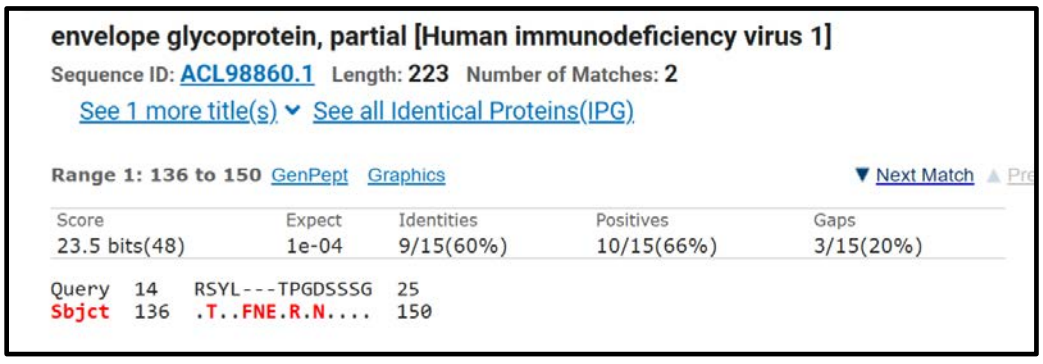




A blast of the second insert, the 15-mer YHKNNKSWMESEFRV, documents similar results. This maps to the V5 domain of the gp120 protein. One sequence has partial identity over 14 of 15 residues, as shown below. This maps to the V5 domain of HIV.

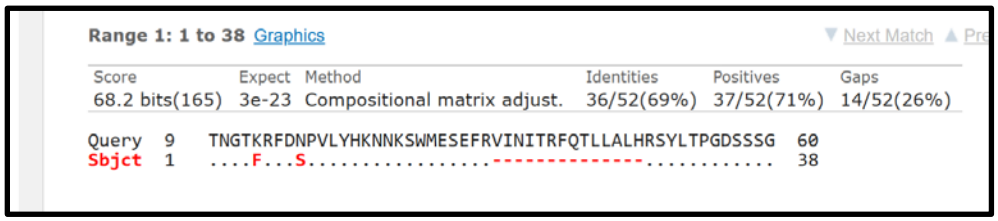


A blast of the third insert, a 25-mer INITRFQTLLALHRSYLTPGDSSSG, shows modest homology with the V1 domain of the HIV gp120 protein.



These discontinuous inserts are brought together in three dimensions in both the gp120 protein of HIV and the spike protein of SARS-CoV-2

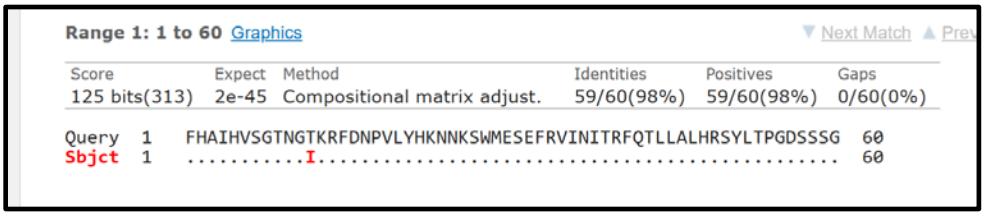
One criticism of comparing these peptide inserts is their short length. However, it is known that the folded viral surface protein of HIV brings these peptides together to form the CD4 receptor recognition domain. Therefore, it is appropriate to form a concatenation of these to examine the collective homology of the three inserts. That analysis is shown here.



Even with the absence of the N-terminal 14 amino acids of insert 3, the statistic of identity of 3e-23 is a highly significant result. With most BLAST an E-value below 1e-3 (0.001) is considered statistically significant. Here the very low E-value (e.g., 3e-23) is 20 orders of magnitude smaller than the above statistical threshold and thus strongly suggests that the alignment is not due to random chance but is an extremely strong indicator of homology.

The same concatenation of these inserts with bat virus RaTG13, whose providence is in question, has a homology that is near identity

A similar concatenation was performed with the three inserts and compared to RaTG13. It has been claimed that this virus, collected in 2013 from a bat, was never genetically manipulated in the laboratory. Nonetheless, the comparison to the human SARS-CoV-2, that emerged 1100 km from the RaTG13 collection site and six years in time apart, is highly statistically significant.



There is no known evolutionary mechanism that would simultaneously freeze this protein motif in the RaTG13-related viruses for six years in the wild and also promote its migration of 1100 km in China, to appear in the fall of 2019.

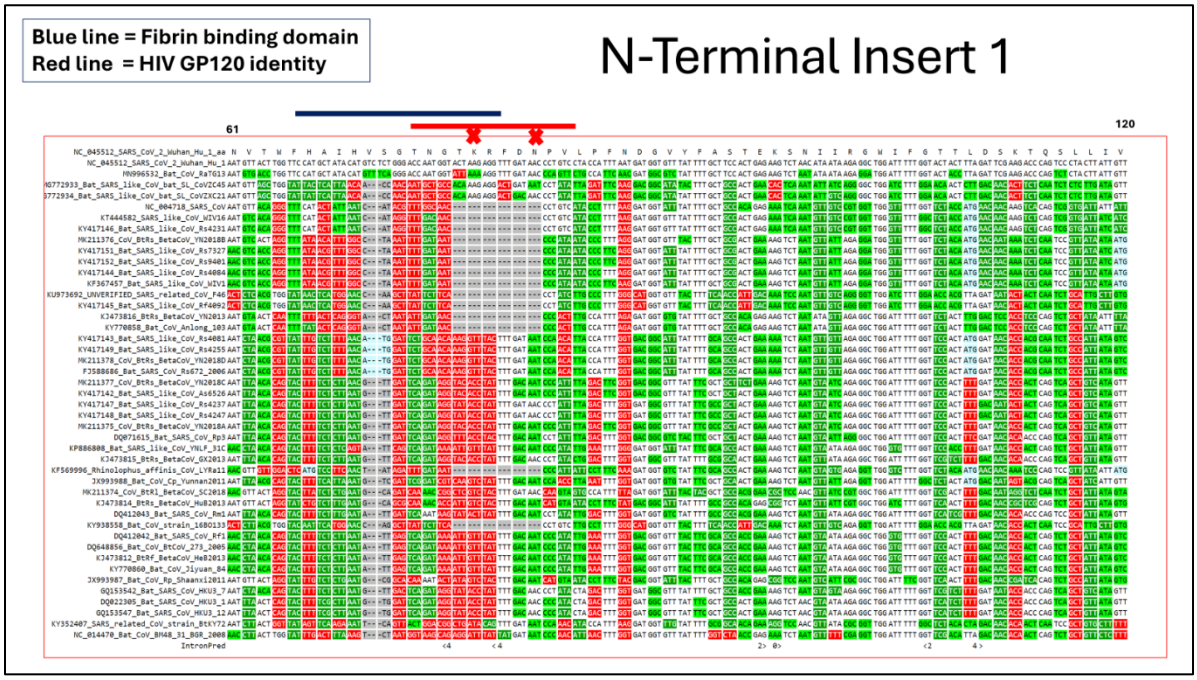
With respect to pre-pandemic bat sarbecoviruses, the three N-terminal spike protein insertions are unique and not found in a panel of sarbecoviruses selected to reflect maximum sequence diversity since the sarbecovirus time to Most Recent Common Ancestor (tMRCA) of about 1200 ADE

To examine if the thrombin binding motifs and the apparent HIV-mimetic motifs of the N-terminus of SARS-CoV-2 were unique to SARS2 and not found in other sarbecoviruses, a previously curated set of 44 sarbecoviruses was used.[6] These viruses were "selected at evolutionary distances well-suited for identifying protein-coding genes and non-coding purifying selection, spanning ~3 substitutions per 4-fold degenerate site on average (comparable to 29-mammals/12-flies projects)." This provides an

[6] Jungreis, I., Sealfon, R. & Kellis, M. SARS-CoV-2 gene content and COVID-19 mutation impact by comparing 44 Sarbecovirus genomes. Nat Commun 12, 2642 (2021). https://doi.org/10.1038/s41467-021-22905-7

approximation of the evolutionary pallet available to the sarbecoviruses over long periods of time, approaching the tMRCA.

N-Terminal Insert 1: This Text-Figure shows the first N-terminal insertion and the location of the fibrin binding sequence and the HIV-like insertion.

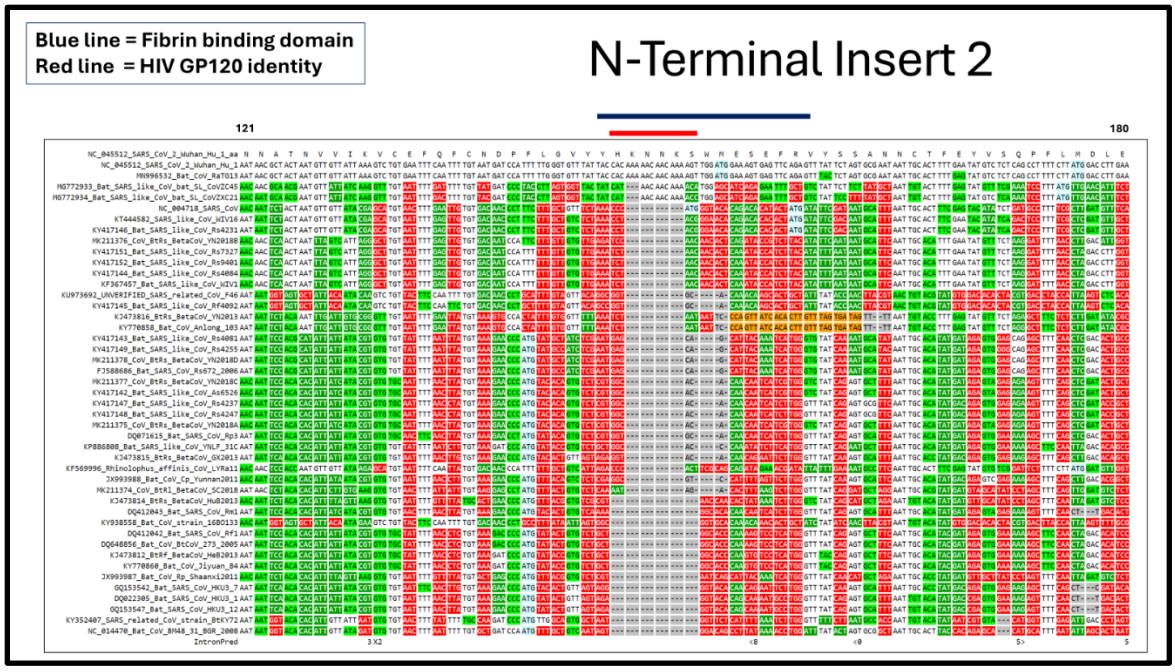


Above the sequences, the blue line indicates the fibrin binding motif and the red line the HIV-like section. The top row is the SARS-CoV-2 peptide sequence. The next row is the genetic sequence with each three nt codon noted. Below SARS-CoV-2 are the 44 sarbecovirus genetic sequence. Codons without color are identical to SARS-CoV-2, light green are synonymous codons, in which the nt changes but the amino acid stays the same. Dark green are non-synonymous changes but where the resulting amino acid is “similar” to the amino acid in SARS-CoV-2. Red codons denote coding for an amino acid that is unrelated to the SARS-CoV-2 amino acid. And blanks denote the related viruses have no codons in those locations.

Starting with the 11-amino acid HIV-like sequence, TNGTKRFDNPV, the other sarbecoviruses have large deletions or non-homologous amino acids comprising over have of the sequence. RaTG13 is closest with 10 of the 11 amino acids. The fibrin binding motif is N-terminal to the HIV-like sequences but it too shows that the other viruses have a mixture of deletions and non-homologous amino acid substitutions. Again, only RaTG13 has the single non-homologous amino acid. The providence of RaTG13 is not the subject of this report but its potential laboratory genetic manipulation should be kept in mind.

N-Terminal Insert 2:

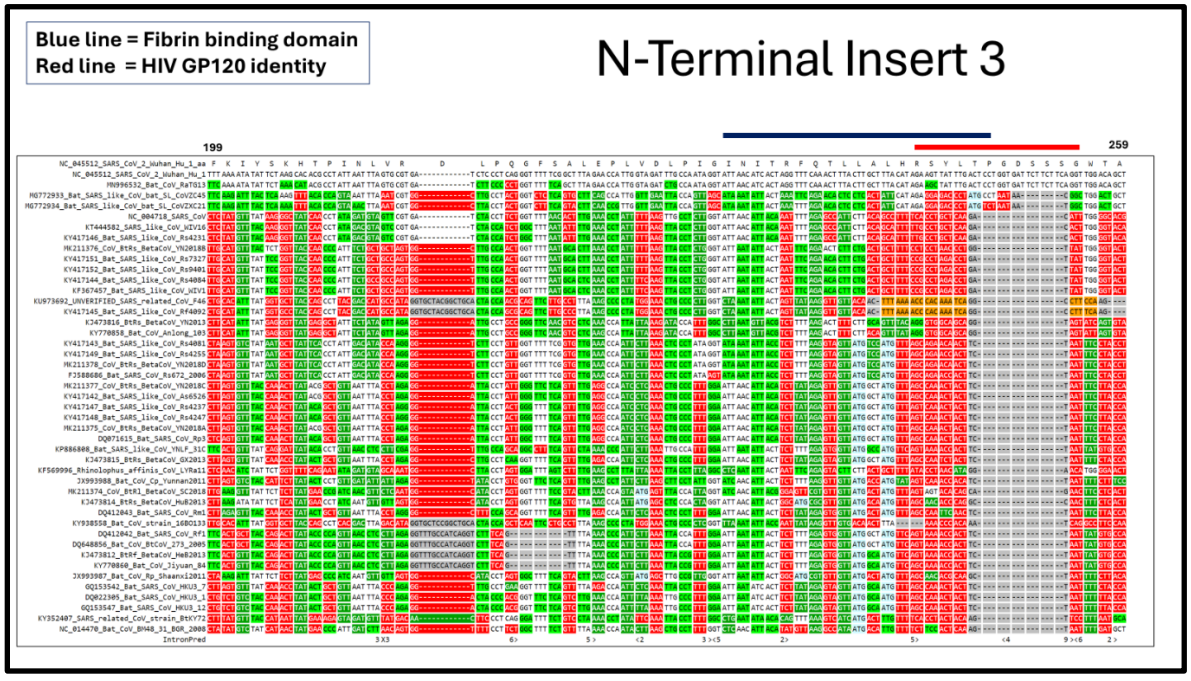
The second N-terminal insert is shown in the below Text-Figure.



Here the HIV-like inserts are smaller and internal to the fibrin binding sequences. As can be seen, over one-half of both amino acid sequences in SARS-CoV-2 are non-homologous in the other viruses or even deleted all together. For this insert the RaTG13 sequence is even more remarkable: it is identical at both the nt and amino acid sequences. None of the other viruses would have fibrin or CD4 cell binding activity.

N-Terminal Insert 3:

The third N-terminal insert is shown in the below Text-Figure:



Once again one sees that, except for RaTG13, the two functional motifs are obliterated by either deletions or non-homologous amino acids. Here RaTG13 stands out again with only a single homologous nt change over 150 contiguous nts, that is, a 99.3% nt homology. As a reminder, SARS-CoV-2 and RaTG13 have a lower overall nt homology of 96.3%, making this region abnormally conserved.

These V1, V4, and V5 domains of HIV are hypervariable in length and contribute to immune escape for HIV

In a paper[7] entitled, "HIV-1 and SARS-CoV-2: Patterns in the evolution of two pandemic pathogens," the authors note: "four of the variable loops of the HIV-1 Env protein, (V1, V2, V4, and V5, but not V3) contain hypervariable sections that have an extraordinary capacity to change by insertion and deletion; the variability in these regions is dramatic and plays an important role in neutralizing antibody resistance."

The authors also note there are small, spatially localized clusters of distinct indels found in Spike that are rare but likely to be viable and transmitted, as they often are sampled multiple times. The most interesting of these clusters is in the region between Spike positions 137–148, which corresponds to insert 3, above. While this Spike variable region is much less variable than the hypervariable regions of HIV-1, it shares some features with them: (1) the region overlaps with an exposed loop on Spike, the N3 loop; (2) there are many distinctive patterns of local deletions observed in this region—along with the very frequently observed Δ144 deletion, a spectrum of 24 other distinct deletion patterns are found among 341 different Spike sequences; and finally, (3) it is embedded in the NTD supersite, and so, like Δ144, the other deletions in this region are also likely to impact antibody resistance.

Evidence the HIV-like sequences in SARS-CoV-2 are functional in patients

Both SARS-CoV-2 and HIV infect CD4+ T-cells in patients, leading to IL-10 expression, immunosuppression, lymphocytopenia of CD4+ cells, and a poor clinical outcome

When HIV first appeared on the clinical scene in the 1980s, it presented as opportunistic infections in severally immunocompromised patients.[8] This was traced to a specific and severe lymphopenia of CD4+ cells, via binding to the CD4 surface glycoprotein via the trimeric

In 2023 it was found that SARS-CoV-2 also caused a drop in CD4+ cells by infecting these immune cells via an interaction between the CD4 surface glycoprotein and the S1 region of the Spike Protein.[9]

Both SARS-CoV-2 and HIV produce clinical evidence of fibrinogen activation via D-dimer elevation

If the Gag protein of HIV and the Spike Protein of SARS-CoV-2 have a biochemical interaction with fibrinogen, activation of coagulation would be seen in elevated D-dimer levels in patients infected with these viruses. In fact, in the Strategies for Management of Anti-Retroviral Therapy trial (SMART) study,

---

[7] Fischer, W, et. al., HIV-1 and SARS-CoV-2: Patterns in the evolution of two pandemic pathogens, Cell Host & Microbe, Vol 29, Issue 7, 2021, P 1093-1110, ISSN 1931-3128, https://doi.org/10.1016/j.chom.2021.05.012

[8] Kovacs J, Masur H. HIV related opportunistic infections: still relevant after 25 years of AIDS progress. Vol 26, No. 6, p. 323-4 (2008). https://www.elsevier.es/es-revista-enfermedades-infecciosas-microbiologia-clinica-28-articulo-hiv-related-opportunistic-infections-still-S0213005X08727204

[9] Brunetti NS, Davanzo GG, de Moraes D, Ferrari AJR, Souza GF, Muraro SP, Knittel TL, Boldrini VO, Monteiro LB, Virgílio-da-Silva JV, Profeta GS, Wassano NS, Nunes Santos L, Carregari VC, Dias AHS, Veras FP, Tavares LA, Forato J, Castro IMS, Silva-Costa LC, Palma AC, Mansour E, Ulaf RG, Bernardes AF, Nunes TA, Ribeiro LC, Agrela MV, et al. SARS-CoV-2 uses CD4 to infect T helper lymphocytes. Elife. 2023 Jul 31;12:e84790. doi: 10.7554/eLife.84790. PMID: 37523305; PMCID: PMC10390044.

there was an association with the coagulation activation in the form of elevated D-dimer measurement in those patients who died after an HIV infection.[10] In COVID-19, higher admission and peak D-dimer values were associated with worsening clinical outcomes, specifically with higher rates of intubation and mortality.[11]

This is clinical evidence that the protein sequence identity of the S1 inserts in the Spike Protein also produces the same fibrinogen interaction and clinical pathology.

SARS-CoV-2 but not HIV produce neurological findings via direct thrombin interactions

As noted in reference 2, SARS-CoV-2 produces neurological findings due to its interaction with fibrin. There is less evidence that HIV can bind fibrin or fibrinogen directly. An examination of the peptide motifs of the inserts shows that, at least for inserts 1 and three, the HIV-like sequences and the fibrin binding sequences have only partial overlap. It would not therefore be surprising that HIV did not have this functionality.

Cross-neutralization of SARS-CoV-2 by HIV-1 specific antibodies and polyclonal plasma

A paper[12] entitled, "Cross-neutralization of SARS-CoV-2 by HIV-1 specific broadly neutralizing antibodies and polyclonal plasma," was published in 2021. Cross-reactive epitopes (CREs) are similar epitopes on viruses that are recognized or neutralized by the same antibodies. The S protein of SARS-CoV-2, similar to type I fusion proteins of viruses such as HIV-1 envelope (Env) and influenza hemagglutinin, is heavily glycosylated. In recent years, highly potent and/or broadly neutralizing human monoclonal antibodies (bnAbs) that are generated in chronic HIV-1 infections have been defined. As some of the HIV-1 bnAbs have evolved to recognize the dense viral glycans and cross-reactive epitopes (CREs), we assessed if these bnAbs cross-react with SARS-CoV-2. Several HIV-1 bnAbs showed cross-reactivity with SARS-CoV-2 **while one HIV-1 CD4 binding site bnAb, N6, neutralized SARS-CoV-2.**

This study demonstrates that the gp120 CD4 binding motif and the SARS-CoV-2 HIV-like inserts have three dimensional structures that permit antibodies raised against HIV to recognize and neutralize SARS-CoV-2. The HIV motif recognized by this antibody is the CD4 binding site.

---

[10] Kuller LH, Tracy R, Belloso W, De Wit S, Drummond F, Lane HC, Ledergerber B, Lundgren J, Neuhaus J, Nixon D, Paton NI, Neaton JD; INSIGHT SMART Study Group. Inflammatory and coagulation biomarkers and mortality in patients with HIV infection. PLoS Med. 2008 Oct 21;5(10):e203. doi: 10.1371/journal.pmed.0050203 . PMID: 18942885; PMCID: PMC2570418.

[11] Nemec HM, Ferenczy A, Christie BD 3rd, Ashley DW, Montgomery A. Correlation of D-dimer and Outcomes in COVID-19 Patients. Am Surg. 2022 Sep;88(9):2115-2118. doi: 10.1177/00031348221091940. Epub 2022 Apr 29. PMID: 35487527; PMCID: PMC9066233.

[12] Mishra N, Kumar S, Singh S, Bansal T, Jain N, Saluja S, et al. (2021) Cross-neutralization of SARS-CoV-2 by HIV-1 specific broadly neutralizing antibodies and polyclonal plasma. PLoS Pathog 17(9): e1009958. https://doi.org/10.1371/journal.ppat.1009958

CONCLUSION

This paper highlights an unusual set of facts:

1. SARS-CoV-2 causes neuro-inflammation and its “long COVID” clinical findings through a mechanism whereby fibrin binds to the SARS-CoV-2 spike protein, forming proinflammatory blood clots.
2. Three of the strongest binding motifs identified by Ryu et. al., are in the N-terminus of the SARS-CoV-2 Spike Protein.
3. These motifs are contiguous to the three inserts identified in January 2020 that are not widely found in related-SARS viruses.
4. These inserts are shown to have primary amino acid sequence homology to portions of the HIV gp120 protein that is responsible for CD4 cell receptor binding. While the sequences are individually small, making their probabilities of being randomly significant unlikely, when they are combined as a continuous 60-amino acid sequence, the homology to HIV is highly significant. The combination is justified because the non-contiguous sequences in HIV are none-the-less brought together to form the CD4 cell receptor binding protein.
5. Antibodies from patients with HIV infections have been found which block the HIV CD4 recognition site and neutralize SARS-CoV-2. This demonstrates the three-dimensional homology of these regions in a functionally significant manner.
6. A hypothesis that both HIV and SARS-CoV-2 can infect CD4 cells via this non-ACE2 interaction is supported by abundant clinical findings.
7. HIV does not share the fibrin binding motifs seen in SARS-CoV-2 and, for the most part, evidence of direct HIV binding to fibrin has not been found.
8. RaTG13 shares 59/60 amino acid homology and, for insert 3, a nt homology of 99.3%. Since there are numerous papers suggesting that RaTG13 has undergone laboratory genetic experiments, one much conclude that there is a likelihood that these unusual properties of SARS-CoV-2 did not arise naturally.